Parton distributions: HERA–Tevatron–LHC

Graeme Watt

CERN PH-TH

Hadron Collider Physics Symposium
Evian, France
16th November 2009

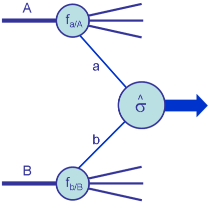
Introduction

- Protons are not elementary particles: made of **partons**.
  $\Rightarrow$ **P**arton **D**istribution **F**unctions (**PDF**s) essential to relate theory to experiment at the LHC (and Tevatron, HERA, ...).
- $f_{a/A}(x, Q^2)$ gives *number density* of partons a in hadron A with momentum fraction x at a hard scale $Q^2 \gg \Lambda^2_{\rm QCD}$.

$$\sigma_{AB} = \sum_{a,b=q,g} \int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b \; f_{a/A}(x_a, Q^2) \; f_{b/B}(x_b, Q^2) \; \hat{\sigma}_{ab}$$

Outline of talk:

1. Global PDF analyses.
2. HERA structure functions: F_2, F_2^c, F_2^b, F_L.
3. Tevatron Z, W, jet data and implications.
4. α_S and SM cross sections at LHC.



Fixed-order collinear factorisation at hadron colliders

- The "standard" pQCD framework: holds up to formally power-suppressed ("higher-twist") terms $\mathcal{O}(\Lambda^2_{\rm QCD}/Q^2)$.
- Expand $\hat{\sigma}_{ab}$, $P_{aa'}$ and β as perturbative series in α_S ($\mu_R = \mu_F = Q$).

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

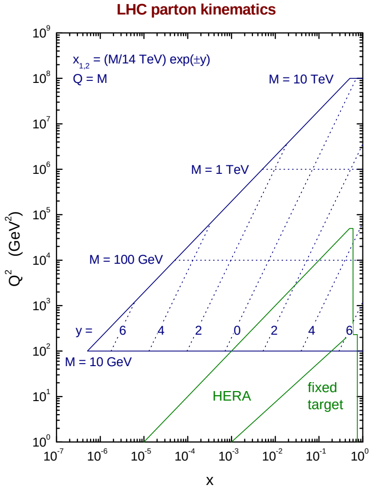
PDF evolution:
$$\frac{\partial f_{a/A}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi} \sum_{a'=q,g} \left[P_{aa'}^{\rm LO} + \alpha_S P_{aa'}^{\rm NLO} + \ldots\right] \otimes f_{a'/A}$$

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 $f_{a/A}(x, Q_0^2)$ and $\alpha_S(M_Z^2)$ from data.
- Structure functions in deep-inelastic scattering (DIS):

$$F_i\left(x_{\rm Bj}, Q^2\right) = \sum_{a=q,g} C_{i,a} \otimes f_{a/A}, \quad C_{i,a} = C_{i,a}^{\rm LO} + \alpha_S C_{i,a}^{\rm NLO} + \ldots$$

From HERA *et al.* to the LHC



- PDFs are **universal**.
- Fit existing data from HERA and fixed-target experiments, together with Tevatron data.
- HERA *ep* (H1, ZEUS).
- Fixed-target experiments:
  ℓp, ℓd
  (BCDMS, NMC, E665, SLAC),
  νN
  (CCFR, NuTeV, CHORUS),
  pp, pd (E866/NuSea).
- Tevatron $p\bar{p}$ (CDF, DØ).
- DGLAP evolution gives PDFs at higher Q^2 for LHC.

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.

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

subject to number- and momentum-sum rule constraints.

2. **Evolve** the PDFs to higher scales $Q^2 > Q_0^2$ using the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) evolution equations.
3. **Convolute** the evolved PDFs with $C_{i,a}$ and $\hat{\sigma}_{ab}$ to calculate theory predictions corresponding to a wide variety of data.
4. **Vary** the input parameters $\{A_a, \Delta_a, \eta_a, \epsilon_a, \gamma_a, \ldots\}$ to minimise

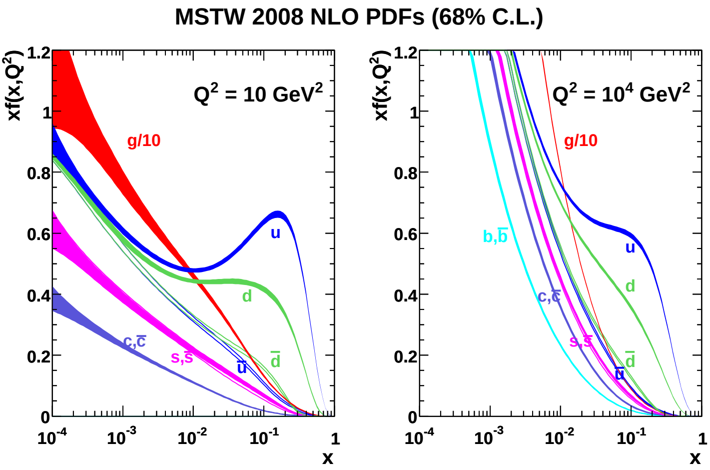
$$\chi^2 = \sum_{i=1}^{N_{\rm pts.}} \left(\frac{\text{Data}_i - \text{Theory}_i}{\text{Error}_i} \right)^2$$

Determination of parton distributions by global analysis

An "industry" for more than 20 years.
Regular updates as new data and theory become available.

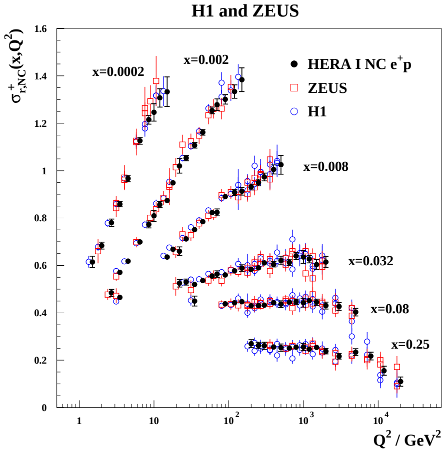
1. First NLO fit: **M**artin+**R**oberts+**S**tirling ('87) + **T**horne ('98).
   Recently, "**MSTW**" = **M**R**ST** − Roberts + G.**W**.
   {MRST 2001 LO, MRST 2004 NLO, MRST 2006 NNLO}
   → MSTW 2008 LO, NLO, NNLO fits [arXiv:0901.0002]
2. Other major group: "**CTEQ**" = **C**oordinated **T**heoretical–**E**xperimental Project on **Q**CD.
   - CTEQ6L1 LO [hep-ph/0201195]
   - CTEQ6.6 NLO [arXiv:0802.0007]
   - CTEQ NNLO?
3. Other groups fitting a restricted range of data with fewer free parameters: S. Alekhin *et al.*, HERA experiments (H1, ZEUS).
4. NNPDF Collaboration (see backup slides).

Example of PDFs obtained from global analysis

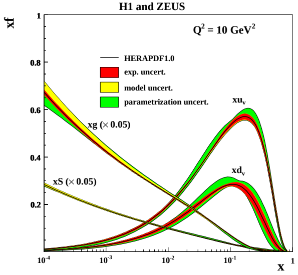


- **Error bands** shown are obtained from propagation of experimental uncertainties on the fitted data points.

Combination of H1 and ZEUS data [arXiv:0911.0884]



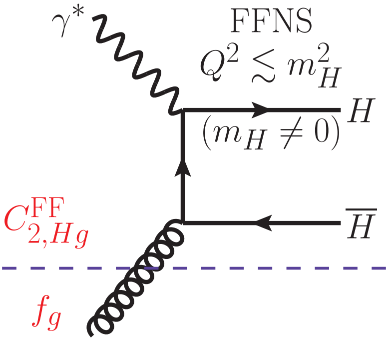
Fit only to HERA data:



- 10 free input PDF parameters (cf. 28 for MSTW 2008).
- Experimental uncertainties using $\Delta\chi^2 = 1$.

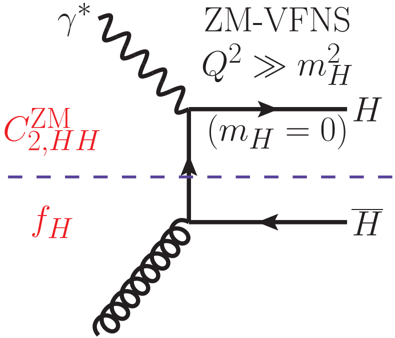
- H1 and ZEUS NC and CC measurements **combined** to improve accuracy: will be used in next generation of global fits.

Heavy quark contribution to DIS structure function F_2



Fixed flavour number scheme

- No heavy quark PDF.
- Includes $\mathcal{O}(m_H^2/Q^2)$ terms.
- No resummation of $\alpha_S \ln(Q^2/m_H^2)$ terms.



Zero-mass variable flavour number scheme

- Use heavy quark PDF.
- Mass dependence neglected.
- Resums $\alpha_S \ln(Q^2/m_H^2)$ terms similar to light quarks.

General-mass variable flavour number scheme (GM-VFNS)

- Interpolate between two well-defined regions.
- FFNS for $Q^2 \leq m_H^2$, ZM-VFNS for $Q^2 \gg m_H^2$.

CTEQ6.1 NLO (ZM-VFNS) $\longrightarrow$ CTEQ6.5 NLO (GM-VFNS)

- 8% increase in W and Z cross sections at LHC.

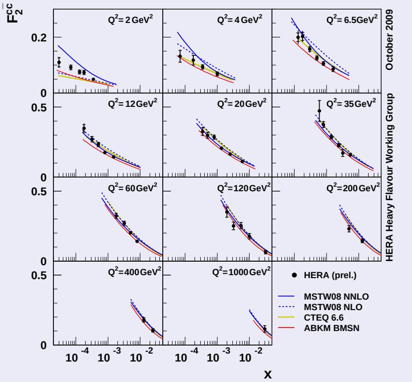
MRST 2004 $\longrightarrow$ MRST 2006 [arXiv:0706.0459]

- The MRST group have used a GM-VFNS since 1998.
- At NNLO, PDFs are discontinuous at $Q^2 = m_H^2$, but neglected in MRST NNLO fits prior to 2006.
- 2004 NNLO $\rightarrow$ 2006 NNLO: 6% increase in $\sigma_{W,Z}$ at LHC.
- Pre-2006 MRST *NNLO* PDF sets should be considered obsolete due to incomplete heavy flavour treatment.

- NNPDF fits (including future NNPDF2.0) still use ZM-VFNS.

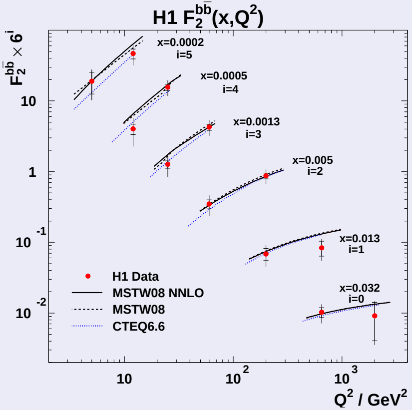
Heavy flavour structure function data

Charm structure function $F_2^{c\bar{c}}$



H1 and ZEUS combination (prel.)

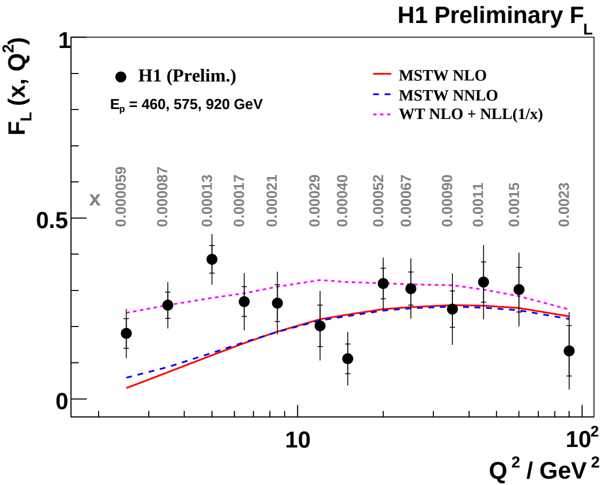
Beauty structure function $F_2^{b\bar{b}}$



[arXiv:0907.2643]

- Good agreement with theoretical predictions using GM-VFNS.

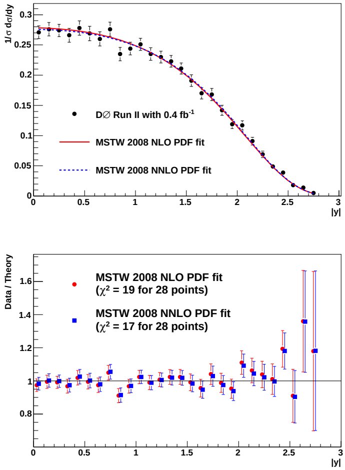
Longitudinal proton structure function at HERA



- NLO and NNLO calculations lower than data at low Q^2.
- Small-x resummation helps [White, Thorne, hep-ph/0611204].

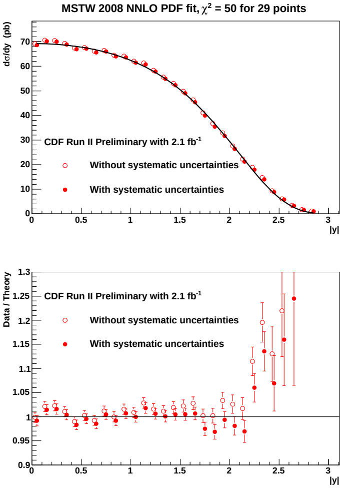
Z/γ^* rapidity distributions from Tevatron Run II

Z/γ* rapidity shape distribution from D∅



[Data: hep-ex/0702025]

Z/γ* rapidity distribution from CDF

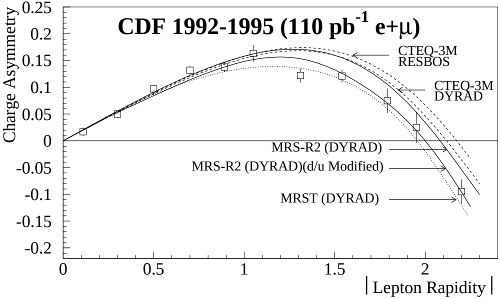


[Data: arXiv:0908.3914]

$W \to \ell\nu$ charge asymmetry from Tevatron Run I

$$A_W(y_W) = \frac{\mathrm{d}\sigma(W^+)/\mathrm{d}y_W - \mathrm{d}\sigma(W^-)/\mathrm{d}y_W}{\mathrm{d}\sigma(W^+)/\mathrm{d}y_W + \mathrm{d}\sigma(W^-)/\mathrm{d}y_W} \approx \frac{u(x_1)d(x_2) - d(x_1)u(x_2)}{u(x_1)d(x_2) + d(x_1)u(x_2)}$$

But **measure** $$A_\ell(\eta_\ell) = \frac{\mathrm{d}\sigma(\ell^+)/\mathrm{d}\eta_\ell - \mathrm{d}\sigma(\ell^-)/\mathrm{d}\eta_\ell}{\mathrm{d}\sigma(\ell^+)/\mathrm{d}\eta_\ell + \mathrm{d}\sigma(\ell^-)/\mathrm{d}\eta_\ell}$$

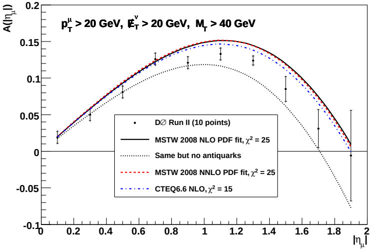


[hep-ex/9809001]

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

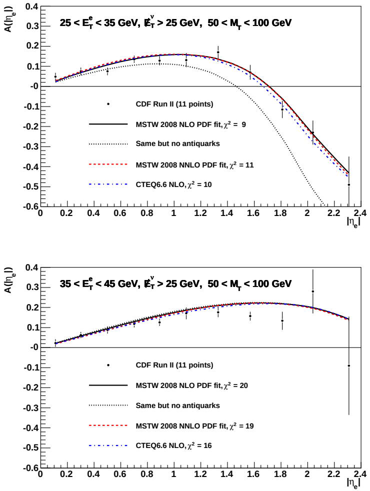
- Run II data in MSTW 2008 fit.
- Mainly constraint on **down** quark.
- Antiquarks important at low p_T^ℓ.

DØ data on lepton charge asymmetry from $W \to \mu\nu$ decays



[Data: arXiv:0709.4254]

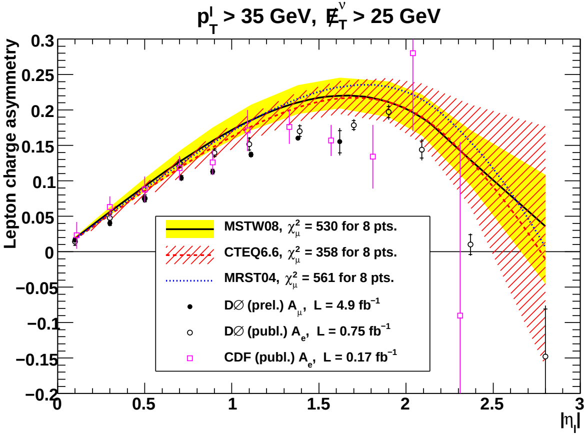
CDF data on lepton charge asymmetry from $W \to e\nu$ decays



[Data: hep-ex/0501023]

Latest DØ data on $W \to \ell\nu$ charge asymmetry

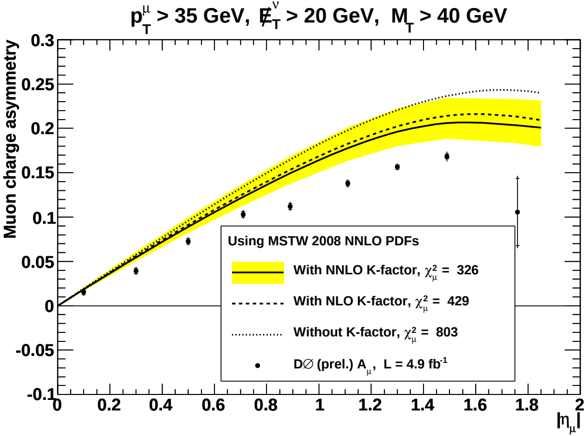
New DØ data: A_e [arXiv:0807.3367] and A_μ [DØ Note 5976-CONF]



- Problems describing new data at NLO, especially for $p_T^\ell > 35$ GeV.

Latest DØ data on $W \to \ell\nu$ charge asymmetry

- Effect of NNLO (or p_T^W-resummation, RESBOS) is small.

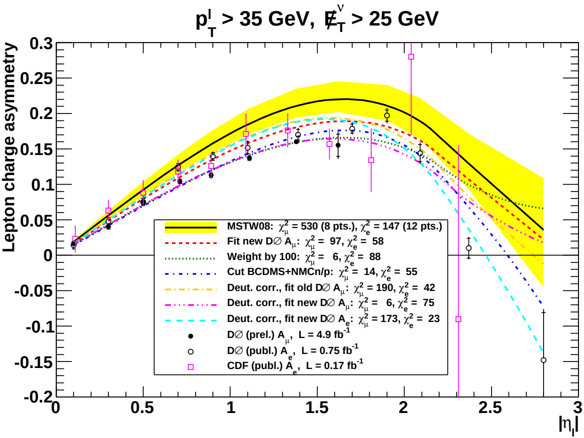


NNLO: Catani, Cieri, Ferrera, de Florian, Grazzini, arXiv:0903.2120

(Previous calculation: Melnikov, Petriello, hep-ph/0609070, FEWZ)

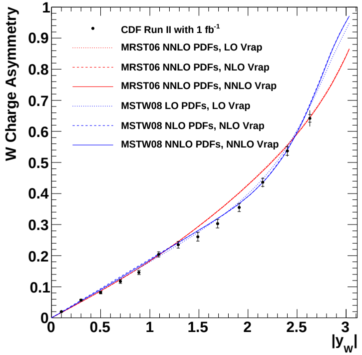
Latest DØ data on $W \to \ell\nu$ charge asymmetry

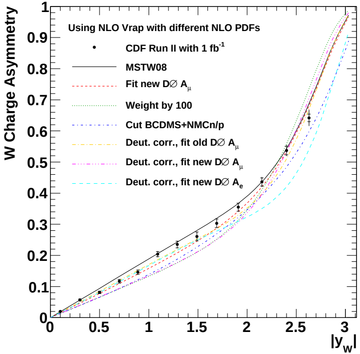
- Can the PDFs be refitted to describe the new data?



- ... Not both DØ A_μ and A_e simultaneously.

W charge asymmetry from Tevatron Run II

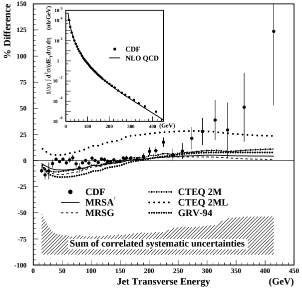


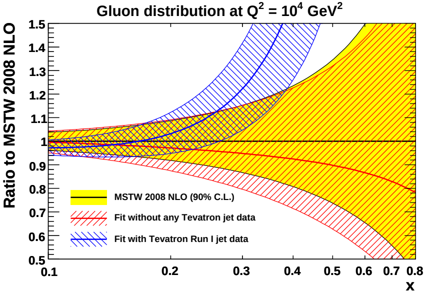


- Data: arXiv:0901.2169
- VRAP: Anastasiou, Dixon, Melnikov, Petriello, hep-ph/0312266

- MSTW08 good description (better than MRST06).
- Modified fits to new DØ A_ℓ tend to **undershoot** CDF A_W.

Impact of Tevatron Run II inclusive jet production data





- Initial Tevatron Run I jet data showed an **excess** at high E_T, later **accommodated** by refitting gluon distribution.
- Run I data included in recent PDF fits up to MRST 2006 (and current CTEQ6.6).
- MSTW 2008 is first PDF fit to include Run II jet data: preference for **smaller** gluon distribution at high x.
- Similar findings by CTEQ [arXiv:0904.2424].

Tension between Run I and Run II inclusive jet data

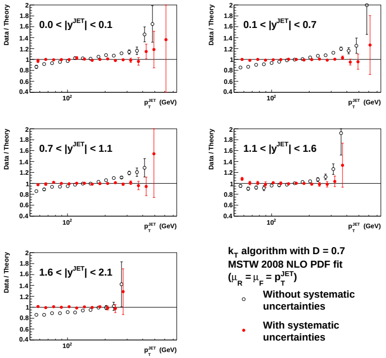
Highlighted numbers indicate χ^2 values for data sets explicitly included in various NLO global fits:

CDFI (33 pts.)	DØI (90 pts.)	CDFII(k_T) (76 pts.)	DØII (110 pts.)	$\Delta\chi^2_{\text{non-jet}}$ (2513 pts.)	$\alpha_S(M_Z^2)$
53	119	64	117	0	0.1197
51	**48**	132	180	9	0.1214
56	110	**56**	**114**	2	0.1202
53	**85**	**68**	**117**	1	0.1204

- Fit to Run I jets ⇒ description of Run II jets bad.
- Fit to Run II jets ⇒ description of Run I jets bad.
- Fit neither ⇒ similar description as fitting Run II only.
- **Summary**: Some inconsistency between Run I and Run II jets. Run II jets slightly more consistent with rest of data.

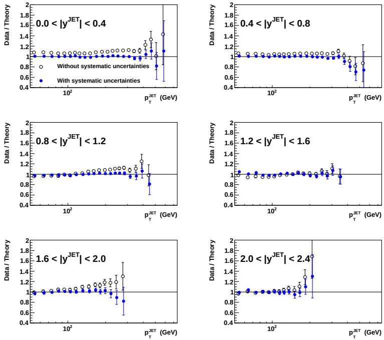
Description of Tevatron Run II inclusive jet data

CDF Run II inclusive jet data, χ^2 = 56 for 76 pts.



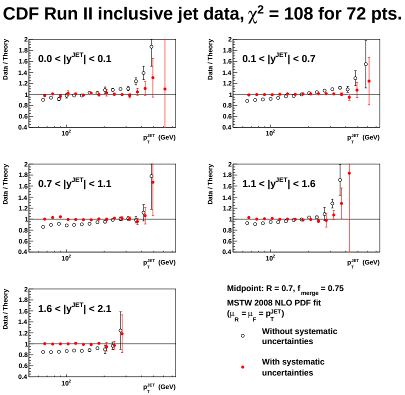
[Data: hep-ex/0701051]

DØ Run II inclusive jet data (cone, R = 0.7)
MSTW 2008 NLO PDF fit ($\mu_R = \mu_F = p_T^{JET}$), χ^2 = 114 for 110 pts.

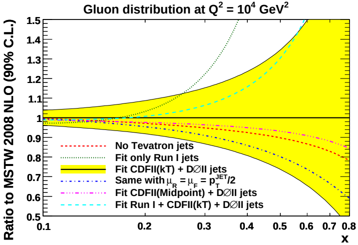


[Data: arXiv:0802.2400]

Effect of CDF Run II jet data using Midpoint algorithm

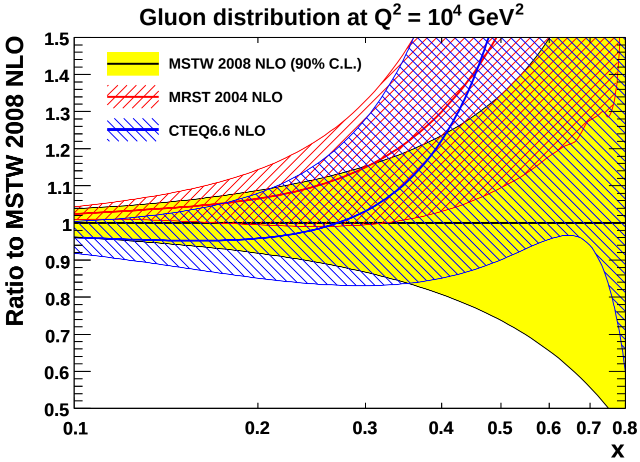


[Data: arXiv:0807.2204]



- Only slight change in gluon if replace CDF(k_T) by CDF(**Midpoint**) data.
- Scale choice $\mu_R = \mu_F = p_T/2$ gives smaller gluon, but within uncertainties.

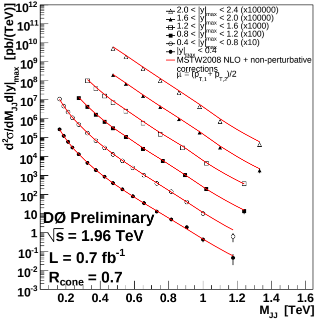
New high-x gluon distribution compared to previous sets



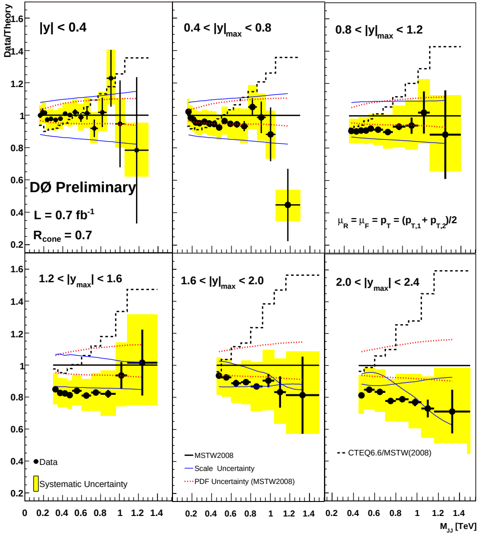
- Smaller high-x gluon than previous MRST and CTEQ fits.

Description of DØ dijet mass spectrum

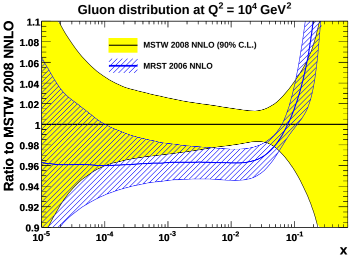
[DØ Note 5919-CONF]



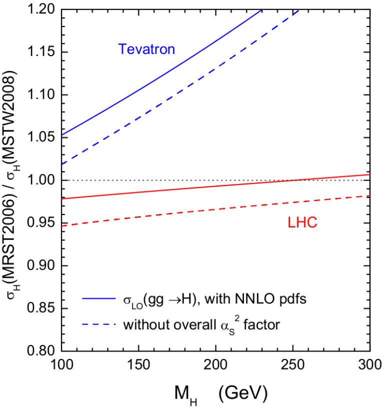
- Data favour **less** gluon at high x (MSTW 2008 over CTEQ6.6).



Implications of new PDFs for Higgs cross sections

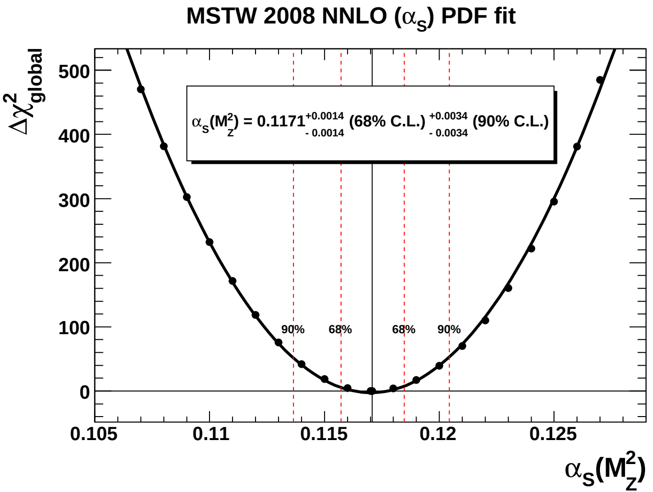


- NNLO trend similar to NLO (N.B. exact NNLO jet cross section unavailable, use threshold corrections).
- $\alpha_S(M_Z^2) = 0.1191$ (2006) $\rightarrow$ 0.1171 (MSTW 2008)



- Higgs cross sections **smaller** at Tevatron with 2008 PDFs.
- Used in Tevatron exclusion results (March 2009).

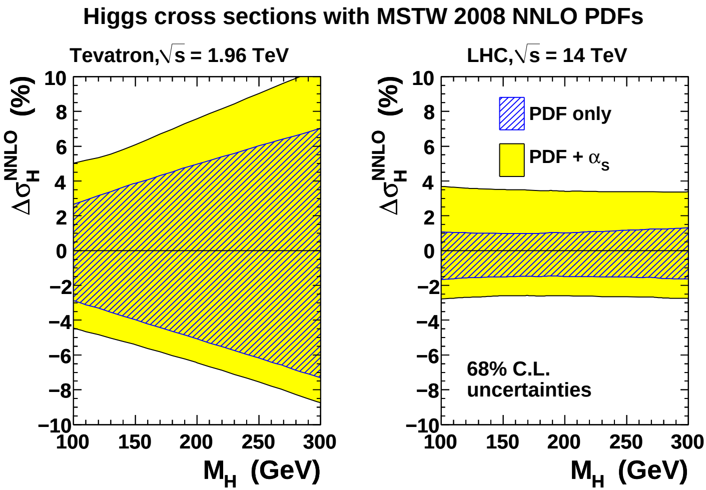
Determination of $\alpha_S(M_Z^2)$ from NNLO global PDF analysis



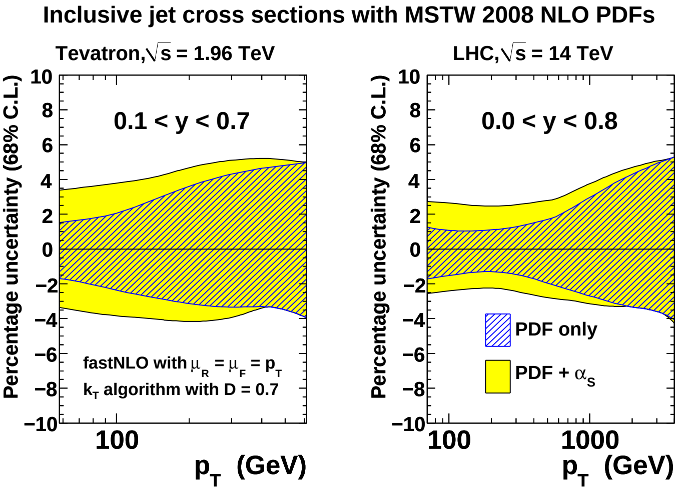
- Additional theory uncertainty ($\lesssim |\mathrm{NNLO} - \mathrm{NLO}| = 0.003$).
- cf. PDG world average value of $\alpha_S(M_Z^2) = 0.1176 \pm 0.002$.

Impact of α_S on SM Higgs uncertainty versus M_H

- **Correlation** between PDF and α_S uncertainties in cross section calculations [MSTW, arXiv:0905.3531].



Impact of α_S on inclusive jet uncertainty versus p_T



- Mostly gluon-initiated at low p_T $\Rightarrow$ correlated with α_S.
- Mostly quark-initiated at high p_T $\Rightarrow$ anticorrelated with α_S.

W and Z total cross sections at the LHC

- Potential "**standard candle**" for luminosity determination.
- NNLO total cross sections with "PDF+α_S" uncertainties using MSTW 2008 NNLO PDFs:

LHC	$B_{\ell\nu} \cdot \sigma_W$ (nb)	$B_{\ell^+\ell^-} \cdot \sigma_Z$ (nb)	R_{WZ}
$\sqrt{s} = 7$ TeV	$10.47^{+0.27}_{-0.20}$ $\left(^{+2.5\%}_{-1.9\%}\right)$	$0.958^{+0.024}_{-0.018}$ $\left(^{+2.5\%}_{-1.9\%}\right)$	$10.92^{+0.03}_{-0.02}$ $\left(^{+0.3\%}_{-0.2\%}\right)$
$\sqrt{s} = 10$ TeV	$15.35^{+0.39}_{-0.31}$ $\left(^{+2.6\%}_{-2.0\%}\right)$	$1.429^{+0.037}_{-0.027}$ $\left(^{+2.6\%}_{-1.9\%}\right)$	$10.74^{+0.03}_{-0.03}$ $\left(^{+0.3\%}_{-0.3\%}\right)$
$\sqrt{s} = 14$ TeV	$21.72^{+0.56}_{-0.48}$ $\left(^{+2.6\%}_{-2.2\%}\right)$	$2.051^{+0.053}_{-0.043}$ $\left(^{+2.6\%}_{-2.1\%}\right)$	$10.59^{+0.03}_{-0.03}$ $\left(^{+0.3\%}_{-0.3\%}\right)$

LHC	$B_{\ell\nu} \cdot \sigma_{W^+}$ (nb)	$B_{\ell\nu} \cdot \sigma_{W^-}$ (nb)	$R_{\pm}$
$\sqrt{s} = 7$ TeV	$6.16^{+0.16}_{-0.12}$ $\left(^{+2.6\%}_{-2.0\%}\right)$	$4.31^{+0.11}_{-0.08}$ $\left(^{+2.5\%}_{-2.0\%}\right)$	$1.429^{+0.015}_{-0.012}$ $\left(^{+1.1\%}_{-0.8\%}\right)$
$\sqrt{s} = 10$ TeV	$8.88^{+0.23}_{-0.19}$ $\left(^{+2.6\%}_{-2.1\%}\right)$	$6.47^{+0.16}_{-0.13}$ $\left(^{+2.5\%}_{-2.0\%}\right)$	$1.373^{+0.013}_{-0.010}$ $\left(^{+0.9\%}_{-0.7\%}\right)$
$\sqrt{s} = 14$ TeV	$12.39^{+0.32}_{-0.28}$ $\left(^{+2.6\%}_{-2.3\%}\right)$	$9.33^{+0.24}_{-0.20}$ $\left(^{+2.6\%}_{-2.1\%}\right)$	$1.328^{+0.011}_{-0.009}$ $\left(^{+0.9\%}_{-0.7\%}\right)$

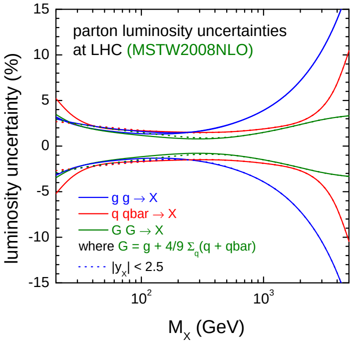
- Additional uncertainty from scale variation less than 1%.

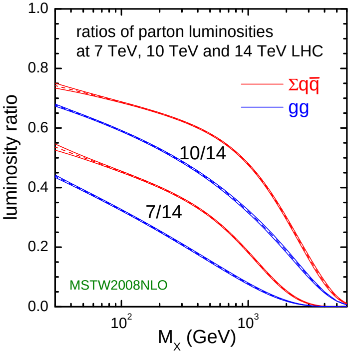
Parton luminosity functions at the LHC

If $\hat{\sigma}_{ab} = C_{ab}\,\delta\left(\hat{s} - M_X^2\right)$, with $\hat{s} = x_a\,x_b\,s$, then

$$\sigma_{AB} = \sum_{a,b}\int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b\; f_{a/A}(x_a, M_X^2)\; f_{b/B}(x_b, M_X^2)\;\hat{\sigma}_{ab} = \sum_{a,b} C_{ab}\,\frac{\partial\mathcal{L}_{ab}}{\partial M_X^2}$$

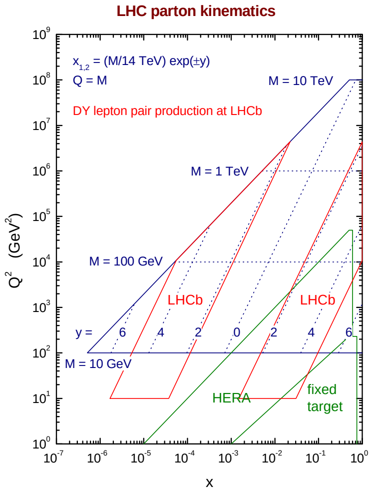
$$\boxed{\frac{\partial\mathcal{L}_{ab}}{\partial M_X^2} = \int_\tau^1 \frac{\mathrm{d}x}{x}\; f_{a/A}(x, M_X^2)\; f_{b/B}(\tau/x, M_X^2), \qquad \tau = \frac{M_X^2}{s}}$$

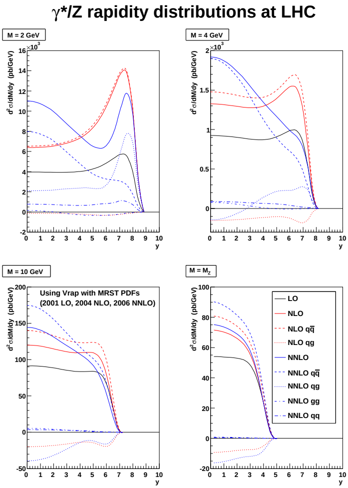




Precision measurements at high rapidity from LHCb

[MSTW, arXiv:0808.1847; R. McNulty, arXiv:0810.2550]





Summary

- **P**arton **D**istribution **F**unctions (**PDF**s) are a non-negotiable input to all theory predictions at hadron colliders.
- **HERA** averaged cross sections reduce uncertainties and will be an important input to future global PDF analyses.
- **Tevatron** Run II W, Z data provide PDF constraints, but problems describing new $W \to \ell\nu$ charge asymmetry data.
- **Tevatron** Run II jets prefer smaller high-x gluon than Run I: impact on Higgs cross sections at Tevatron.
- Now possible to consistently calculate combined "PDF+α_S" uncertainty on hadronic cross sections.

PDFs for use in LO Monte Carlo event generators

- Which PDFs (f^X, $X = \mathrm{LO}, \mathrm{NLO}, \ldots$) to use if only a LO $\hat{\sigma}$ is available? Define the "truth" to be $f^{\mathrm{NLO}} \otimes \hat{\sigma}^{\mathrm{NLO}}$, and

$$\mathrm{K}(X) = \frac{f^{\mathrm{NLO}} \otimes \hat{\sigma}^{\mathrm{NLO}}}{f^{X} \otimes \hat{\sigma}^{\mathrm{LO}}}$$

- A. Sherstnev and R. Thorne have studied modified LO PDFs (LO* and LO**) which give K(LO*) or K(LO**) much closer to 1 than either K(LO) or K(NLO) for a variety of processes.

  LO* [arXiv:0711.2473]: LO PDF fit with violation of momentum-sum rule and NLO α_S.

  LO** [arXiv:0807.2132]: same but also modified α_S scale in PDF evolution, similar to in parton shower.

  Based on MRST 2006 analysis: will be updated soon.

- CTEQ [arXiv:0910.4183]: similar idea with NLO "pseudodata".

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

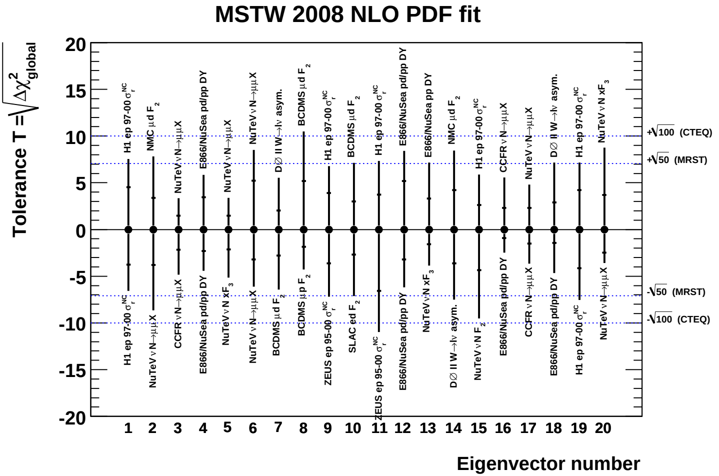
Parameter-fitting criterion

- $T^2 = 1$ for 68% (1-σ) C.L., $T^2 = 2.71$ for 90% C.L.
- **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)

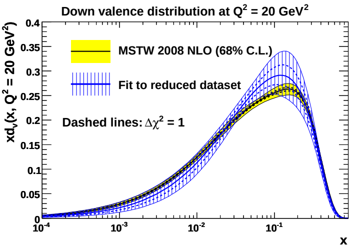
- Much weaker: treat PDF sets obtained from eigenvectors of covariance matrix as alternative hypotheses.
- Determine T^2 from the criterion that each data set should be described within its 90% C.L. limit. Very roughly, a "good" fit has $\chi^2 \simeq N_{\rm pts.} \pm \sqrt{2N_{\rm pts.}}$ for each data set.
- **CTEQ:** $T^2 = 100$ for 90% C.L. limit, **MRST:** $T^2 = 50$.

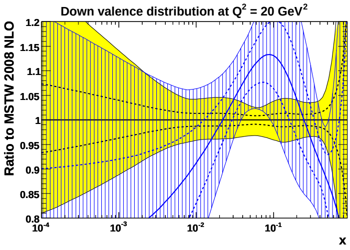
Dynamic tolerance: different for each eigenvector



- Outer (inner) error bars give tolerance for 90% (68%) C.L.

Test of dynamic tolerance: fit to reduced dataset





- Fit to reduced dataset comprising 589 DIS data points, cf. 2699 data points in global fit.
- Errors given by $T^2 = 1$ don't overlap $\Rightarrow$ inconsistent data sets included in global fit.
- **Dynamic tolerance** $T^2 > 1$ accommodates mildly inconsistent data sets.
- J. Pumplin [arXiv:0909.0268]: significant tension from BCDMS/NMC supports $T^2 \approx 10$ for 90% C.L. uncertainties.
- **Issues:**
  1. $T^2 > 1$ not rigorous?
  2. Dependence on input parameterisation?

Alternative approach: NNPDF Collaboration

NNPDF Collaboration: R. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. Latorre, A. Piccione, J. Rojo, M. Ubiali

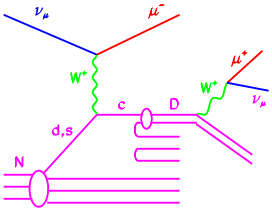
MSTW approach [arXiv:0901.0002] (CTEQ similar)

Parameterisation	$xf_{a/p} \sim A_a\, x^{\Delta_a}\, (1-x)^{\eta_a}\, (1 + \epsilon_a \sqrt{x} + \gamma_a\, x)$
Minimisation	Non-linear least-squares (Marquardt method)
Error propagation	Hessian method with dynamical tolerance
Application	Use best-fit and 40 eigenvector PDF sets

NNPDF approach [arXiv:0808.1231, arXiv:0906.1958]

Parameterisation	Neural network (37 free parameters per PDF)
Minimisation	Genetic algorithm (stop before overlearning)
Error propagation	Generate $N_{\rm rep} \sim \mathcal{O}(1000)$ MC data replicas
Application	Calculate average and s.d. over $N_{\rm rep}$ PDF sets

Illustration: NuTeV/CCFR dimuon cross sections



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

- ν_μ and $\bar{\nu}_\mu$ cross sections constrain s and $\bar{s}$, respectively, for $0.01 \lesssim x \lesssim 0.2$.

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

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

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

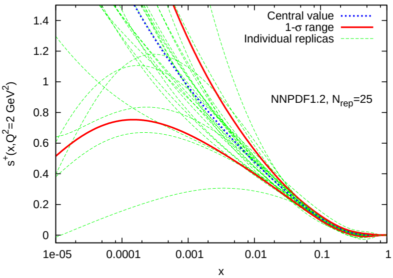
$$xs^+(x, Q_0^2) \equiv xs(x, Q_0^2) + x\bar{s}(x, Q_0^2) = A_+\,(1-x)^{\eta_+}\,xS(x, Q_0^2),$$
$$xs^-(x, Q_0^2) \equiv xs(x, Q_0^2) - x\bar{s}(x, Q_0^2) = A_-\,x^{0.2}\,(1-x)^{\eta_-}\,(1-x/x_0).$$

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

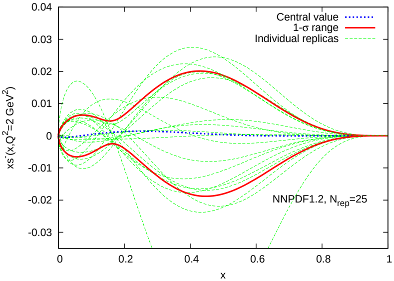
First fits from NNPDF Collaboration

- NNPDF**1.0**: fit only DIS structure function data.
  Fix $s = \bar{s} = (\bar{u} + \bar{d})/4$ at $Q_0^2 = 2$ GeV^2.
- NNPDF**1.1**: free strangeness but no νN dimuon data.
- NNPDF**1.2**: free strangeness and add νN dimuon data.

$s^+ \equiv s + \bar{s}$ at $Q^2 = 2$ GeV^2:



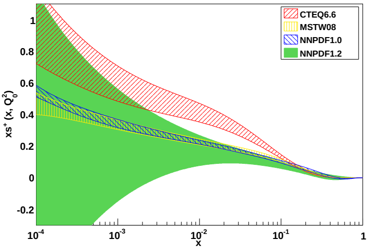
$xs^- \equiv xs - x\bar{s}$ at $Q^2 = 2$ GeV^2:



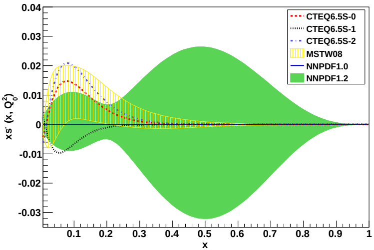
- Data only constrain $0.01 \lesssim x \lesssim 0.2$.

NNPDFs compared to "standard" PDFs

$xs^+ \equiv xs + x\bar{s}$ at $Q^2 = 2$ GeV^2:



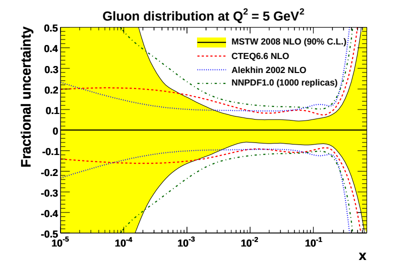
$xs^- \equiv xs - x\bar{s}$ at $Q^2 = 2$ GeV^2:



- NNPDF uncertainties much larger in regions of no data.
- MSTW use a relatively restrictive input parameterisation:
  1. Restrict small-x s^+ to be (mass-suppressed) fraction of $\bar{u}, \bar{d}$.
  2. No reason to expect very large asymmetry s^- at large x.
- CTEQ s^+ disagrees even in region of data: not understood.

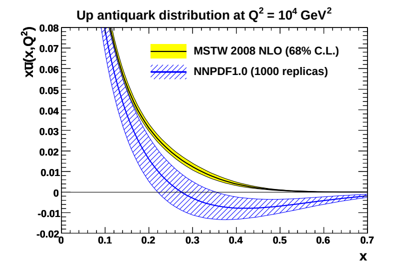
NNPDFs compared to "standard" PDFs

Gluon fractional uncertainty



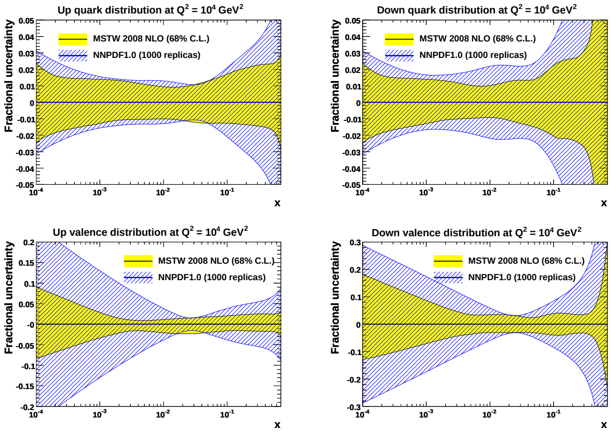
- MSTW small-x gluon: $xg(x, Q_0^2) \sim A_g\, x^{\delta_g} + A_{g'}\, x^{\delta_{g'}}$
- CTEQ, Alekhin: $xg(x, Q_0^2) \sim A_g\, x^{\delta_g}$.

Up antiquark at large x



- NNPDF1.0 negative by $\sim$ 2-σ at large $x \sim 0.5$.
- No E866/NuSea Drell–Yan data to constrain.

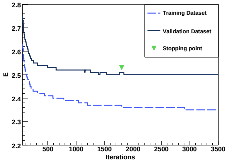
Uncertainties for MSTW 2008 and NNPDF1.0



- NNPDF1.0 uncertainties generally larger, but also less data.
- Fully global fit with Tevatron data (NNPDF2.0) in progress.

Potential issues with stopping criterion in NNPDF fits

- Partition each data set into "training" and "validation" to avoid overlearning. Stop fit when $\chi^2_{\rm val}$ increases. Complications due to each data set having a different training length.



Talk by S. Forte, PDF4LHC, DESY, 23rd October 2009:

REMOVE STOPPING: OVERLEARNING FIT

PERFORM A FIT WITH A FIXED, VERY LARGE NUMBER OF GA GENERATIONS:
25000 gens. (AVERAGE 1000 gens. FOR STANDARD FIT)

	STANDARD STOPPING			FIXED LONG	
	REPLICAS	CENTRAL VALUE	FIXED PARTITION	REPLICAS	CENTRAL VALUE
χ^2	1.32	1.32	$\sim$1.3	1.18	1.19
$\langle\chi^2\rangle_{\rm rep}$	2.79 ± 0.24	1.65 ± 0.20	$\sim 1.6 \pm 0.2$	2.43 ± 0.13	1.29 ± 0.06
$\langle\chi^2_{\rm tr}\rangle_{\rm rep}$	2.76	1.59	$\sim$1.6	2.40	1.27
$\langle\chi^2_{\rm val}\rangle_{\rm rep}$	2.80	1.61	$\sim$1.6	2.47	1.30
$\langle\sigma^{\rm dat}\rangle$	0.039	0.035	$\sim$0.03	0.032	0.019

χ^2 OF THE GLOBAL FIT DECREASES A LOT!
IS IT REALLY OVERLEARNING?

- $\langle\chi^2_{\rm val}\rangle_{\rm rep}$ continues decreasing after standard stopping.
- Stopping point has significant influence on PDF uncertainties.