Parton Distributions Functions and Uncertainties

Robert Thorne

July 13th, 2010

University College London

Thanks to Alan Martin, James Stirling and Graeme Watt

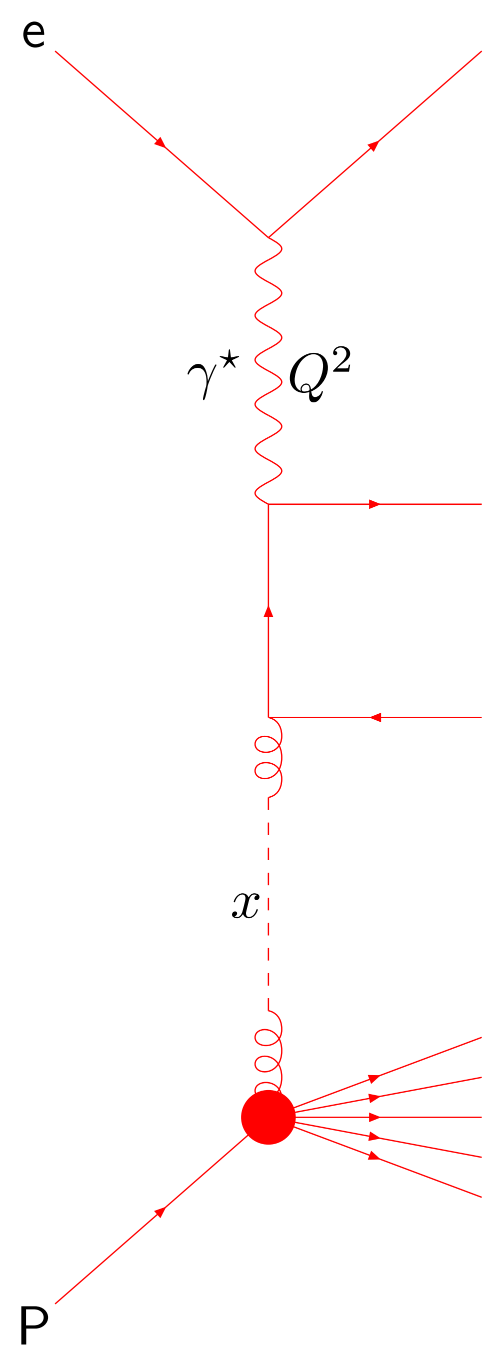
Strong force makes it difficult to perform analytic calculations of scattering processes involving hadronic particles.

The weakening of $\alpha_S(\mu^2)$ at higher scales $\longrightarrow$ the **Factorization Theorem**.

Hadron scattering with an electron factorizes.

Q^2 – Scale of scattering

$x = \frac{Q^2}{2m\nu}$ – Momentum fraction of Parton (ν=energy transfer)



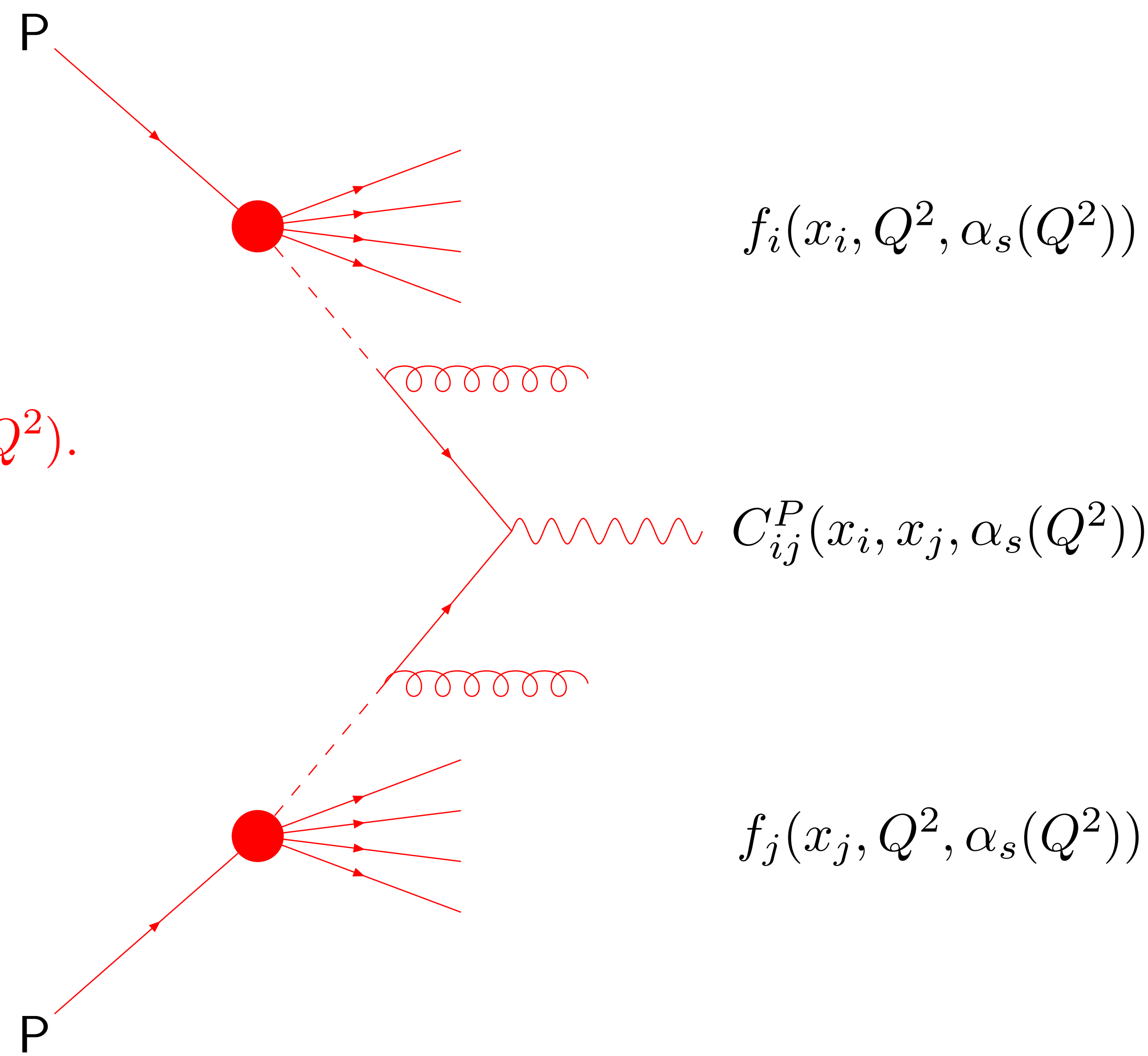
perturbative
calculable
coefficient function
$C_i^P(x, \alpha_s(Q^2))$

nonperturbative
incalculable
parton distribution
$f_i(x, Q^2, \alpha_s(Q^2))$

The coefficient functions $C_i^P(x, \alpha_s(Q^2))$ are process dependent (new physics) but are calculable as a power-series in $\alpha_s(Q^2)$.

$$C_i^P(x, \alpha_s(Q^2)) = \sum_k C_i^{P,k}(x)\alpha_s^k(Q^2).$$

Since the parton distributions $f_i(x, Q^2, \alpha_s(Q^2))$ are process-independent, i.e. **universal**, and evolution with scale is calculable, once they have been measured at one experiment, one can predict many other scattering processes.



General procedure.

Start parton evolution at low scale $Q_0^2 \sim 1\mathrm{GeV}^2$. In principle 11 different partons to consider.

$$u, \bar{u}, \quad d, \bar{d}, \quad s, \bar{s}, \quad c, \bar{c}, \quad b, \bar{b}, \quad g$$

$m_c, m_b \gg \Lambda_{\mathrm{QCD}}$ so heavy parton distributions determined perturbatively. Leaves 7 independent combinations, or 6 if we assume $s = \bar{s}$ – starting not to.

$$u_V = u - \bar{u}, \quad d_V = d - \bar{d}, \quad \mathrm{sea} = 2 * (\bar{u} + \bar{d} + \bar{s}), \quad s + \bar{s} \quad \bar{d} - \bar{u}, \quad g.$$

Input partons parametrised as, e.g. (MRST/MSTW)

$$xf(x, Q_0^2) = (1-x)^\eta (1 + \epsilon x^{0.5} + \gamma x) x^\delta.$$

Evolve partons upwards using LO, NLO (or NNLO) DGLAP equations.

$$\frac{df_i(x, Q^2, \alpha_s(Q^2))}{d \ln Q^2} = \sum_j P_{ij}(x, \alpha_s(Q^2)) \otimes f_j(x, Q^2, \alpha_s(Q^2))$$

Fit data for scales above $2 - 10\mathrm{GeV}^2$. Need many different types of experiment for full determination.

- Lepton-proton collider HERA – (DIS) $\rightarrow$ small-x quarks. Also gluons from evolution, and $F_L(x, Q^2)$. Also, jets $\rightarrow$ moderate-x gluon.

- Fixed target DIS – higher x – leptons (BCDMS, NMC, ...) $\rightarrow$ up quark (proton) or down quark (deuterium) and neutrinos (CHORUS, NuTeV, CCFR) $\rightarrow$ valence or singlet combinations.

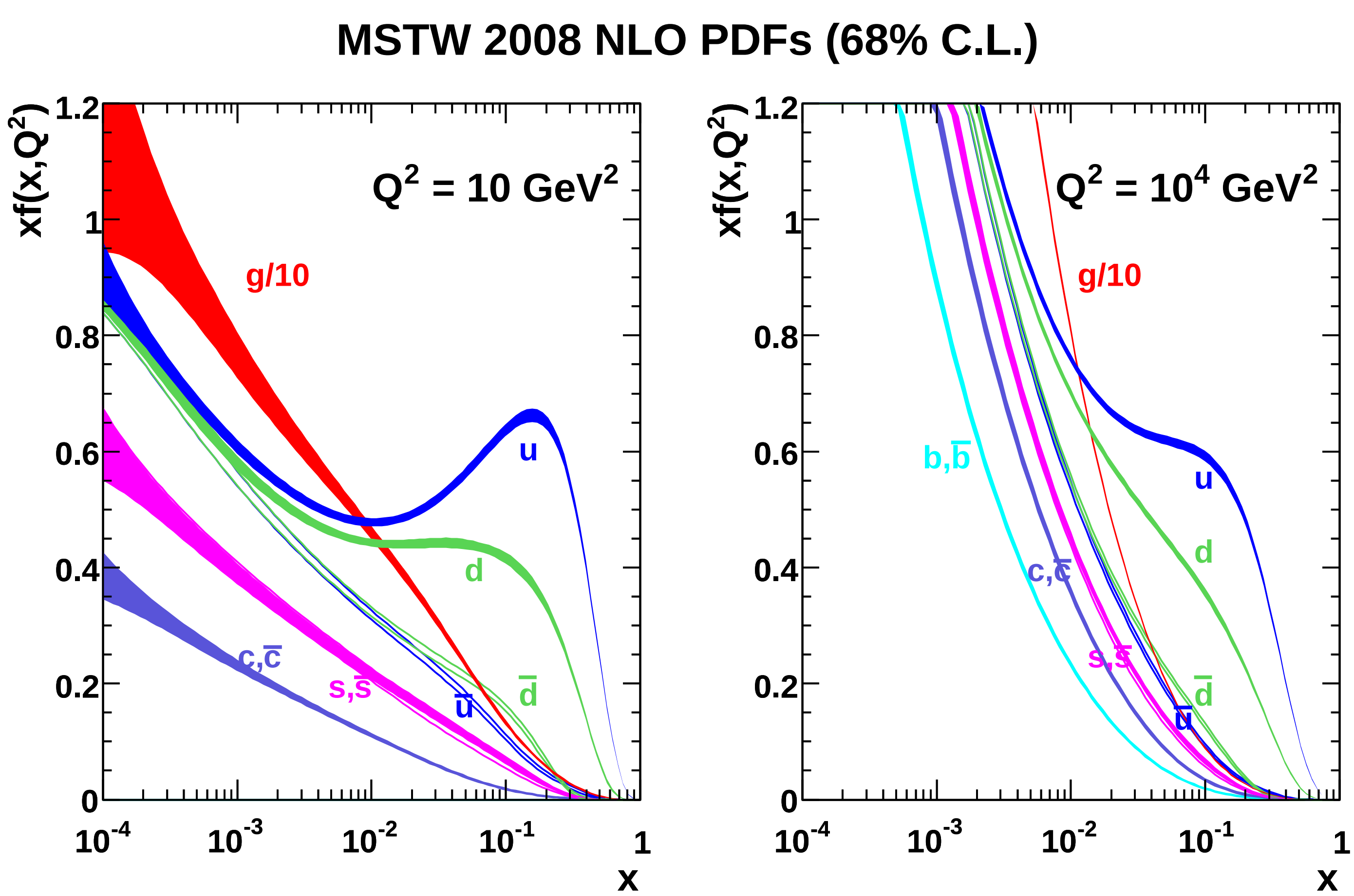
- Di-muon production in neutrino DIS – strange quarks and neutrino-antineutrino comparison $\rightarrow$ asymmetry .

- Drell-Yan production of dileptons – quark-antiquark annihilation (E605, E866) – high-x sea quarks. Deuterium target – $\bar{u}/\bar{d}$ asymmetry.

- High-p_T jets at colliders (Tevatron) – high-x gluon distribution.

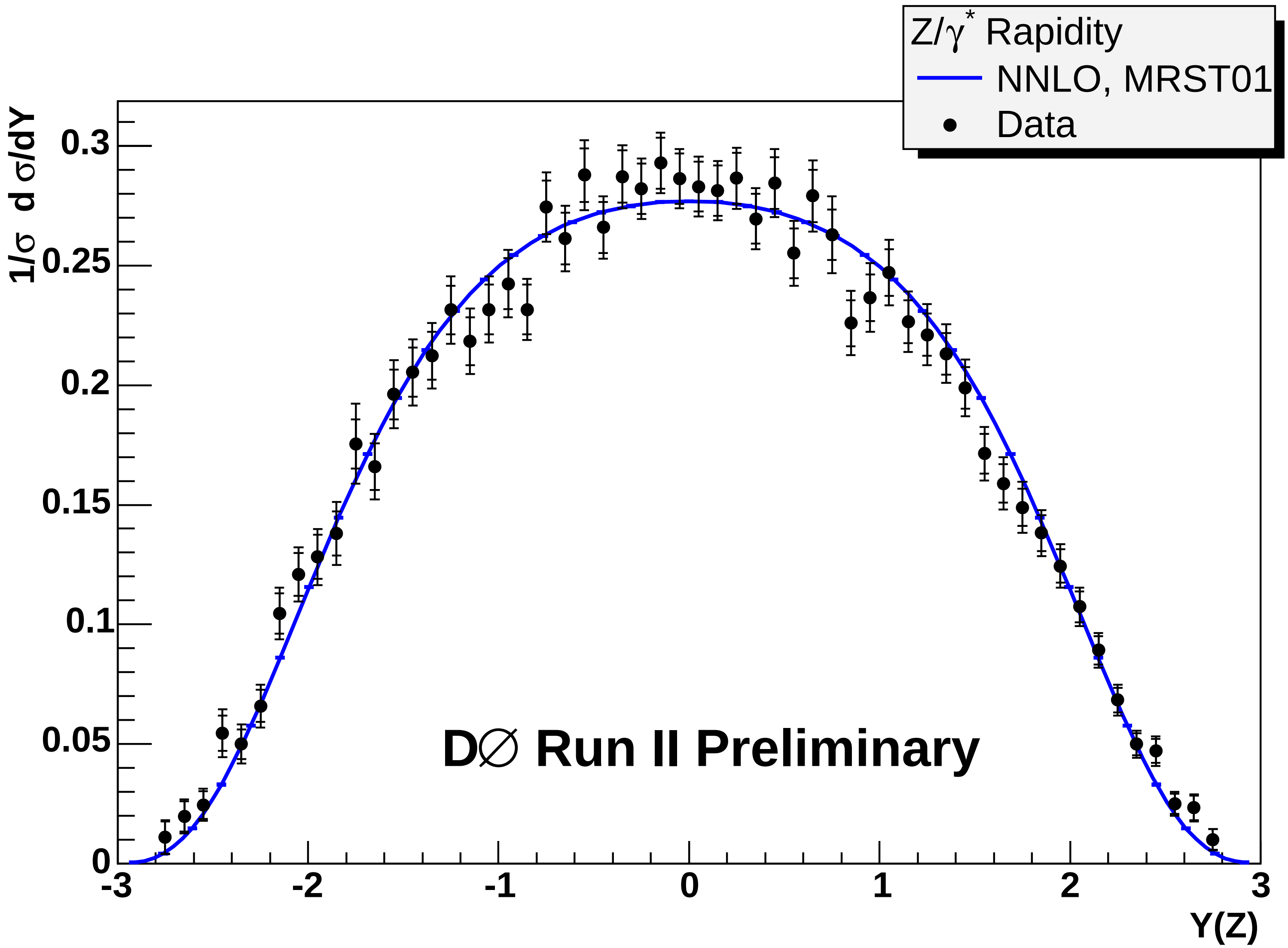
- W and Z production at colliders (Tevatron) – different quark contributions to DIS.

This procedure is generally successful and is part of a large-scale, ongoing project. Results in partons of the form shown.



Various choices of PDF – MSTW, CTEQ, NNPDF, Alekhin, HERA, H1, Jimenez-Delgado *et al* *etc.*. All LHC cross-sections rely on our understanding of these partons.

Excellent predictive power – comparison of MRST prediction for Z rapidity distribution with preliminary data.



Parton Fits and Uncertainties. Two main approaches.

Parton parameterization and Hessian (Error Matrix) approach first used by H1 and ZEUS, and extended by CTEQ.

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

The Hessian matrix H is related to the covariance matrix of the parameters by

$$C_{ij}(a) = \Delta\chi^2 (H^{-1})_{ij}.$$

We can then use the standard formula for linear error propagation.

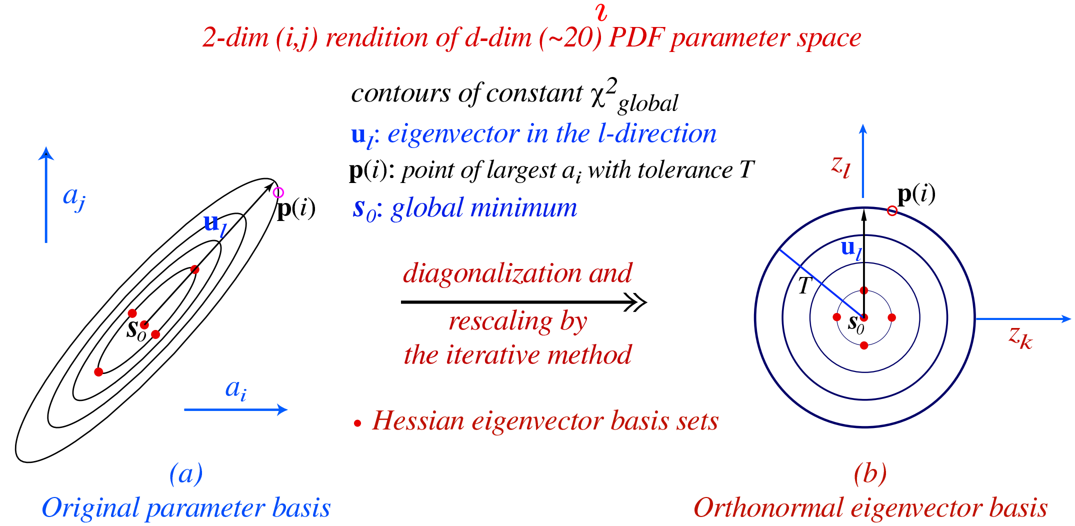
$$(\Delta F)^2 = \Delta\chi^2 \sum_{i,j} \frac{\partial F}{\partial a_i}(H)^{-1}_{ij}\frac{\partial F}{\partial a_j},$$

This is now the most common approach (sometimes $Offset\ method$).

Problematic due to extreme variations in $\Delta\chi^2$ in different directions in parameter space.

Solved by finding and rescaling eigenvectors of H leading to diagonal form

$$\Delta\chi^2 = \sum_i z_i^2$$



Implemented by CTEQ, then others. Uncertainty on physical quantity then given by

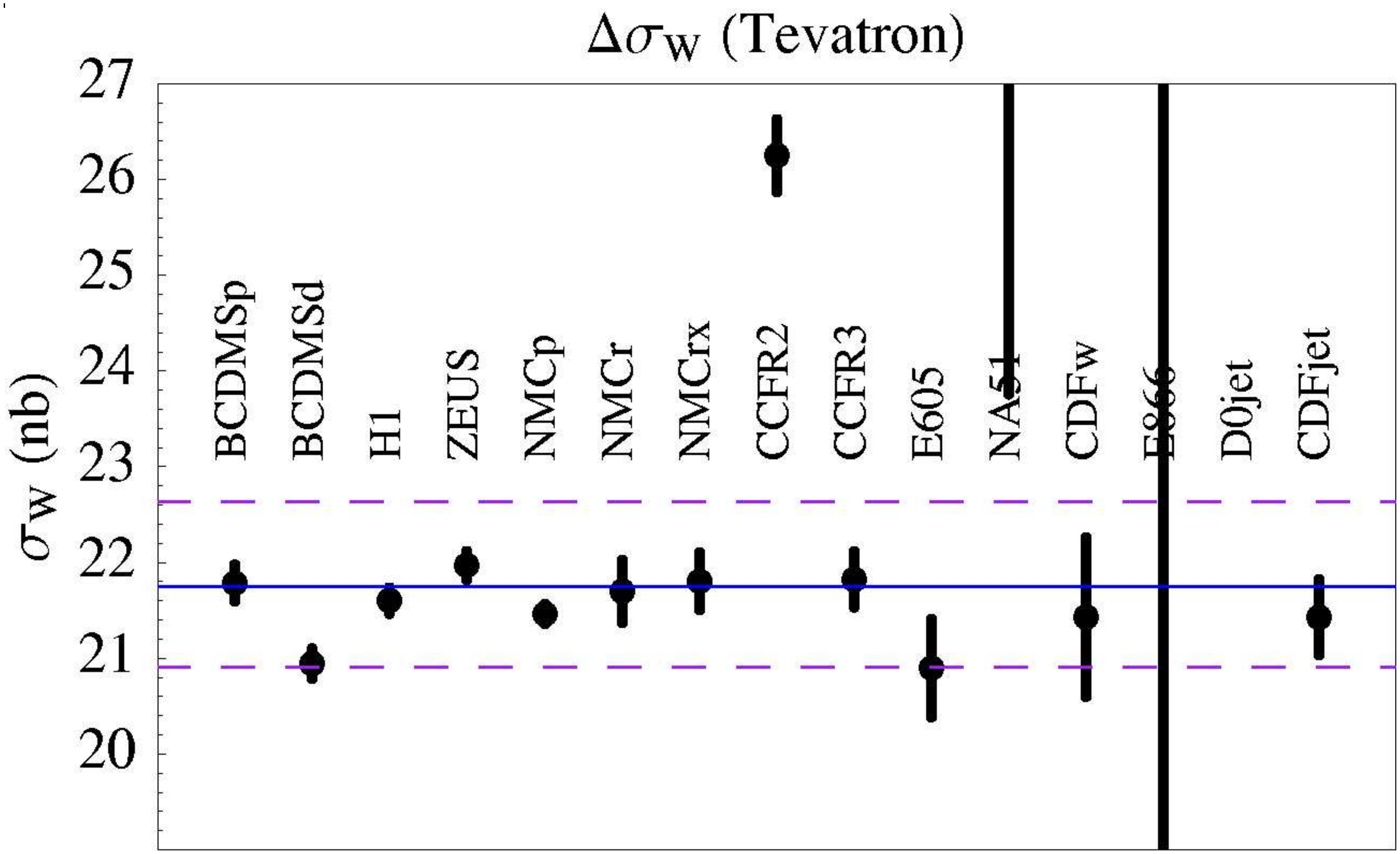
$$(\Delta F) = 1/2\sqrt{\sum_i \left(F(S_i^{(+)}) - F(S_i^{(-)})\right)^2},$$

where $S_i^{(+)}$ and $S_i^{(-)}$ are PDF sets displaced along eigenvector direction.

Question of choosing "correct" $\Delta\chi^2$ given complication of errors in full fit and sometimes conflicting data sets.

CTEQ use $\Delta\chi^2 \sim 40$ and MRST/MSTW use more complicated approach – results in $\Delta\chi^2 \sim 5-20$, for one σ. Other fits less global, keep to $\Delta\chi^2 = 1$.

The inappropriateness of using $\Delta\chi^2 = 1$ when including a large number of sometimes conflicting data sets is shown by examining the best value of σ_W and its uncertainty using $\Delta\chi^2 = 1$ for individual data sets as obtained by CTEQ using Lagrange Multiplier technique.



Also from comparison of partons.

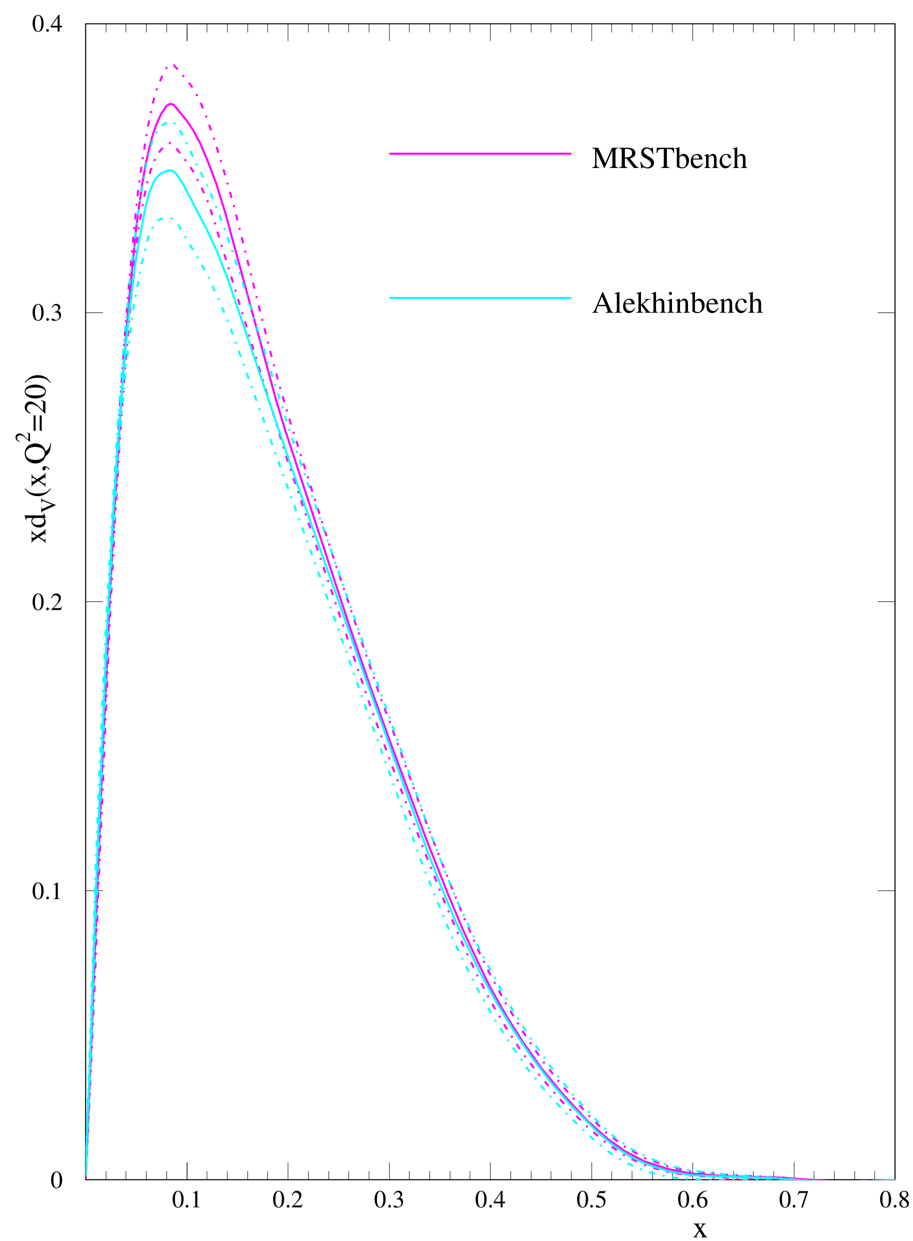
Exercise for $HERA-LHC$ meeting. Fit proton and deuteron structure function data from H1, ZEUS, NMC and BCDMS, for $Q^2 > 9\mathrm{GeV}^2$ using $ZM-VFNS$ and same form of parton inputs at same $Q_0^2 = 1\mathrm{GeV}^2$.

Very conservative fit.

Compare rigorous treatment of all systematic errors (Alekhin) with simple quadratures approach (MRST), both with $\Delta\chi^2 = 1$.

$\rightarrow$ some difference in central values (other possible reasons) and similar errors.

Fairly consistent.

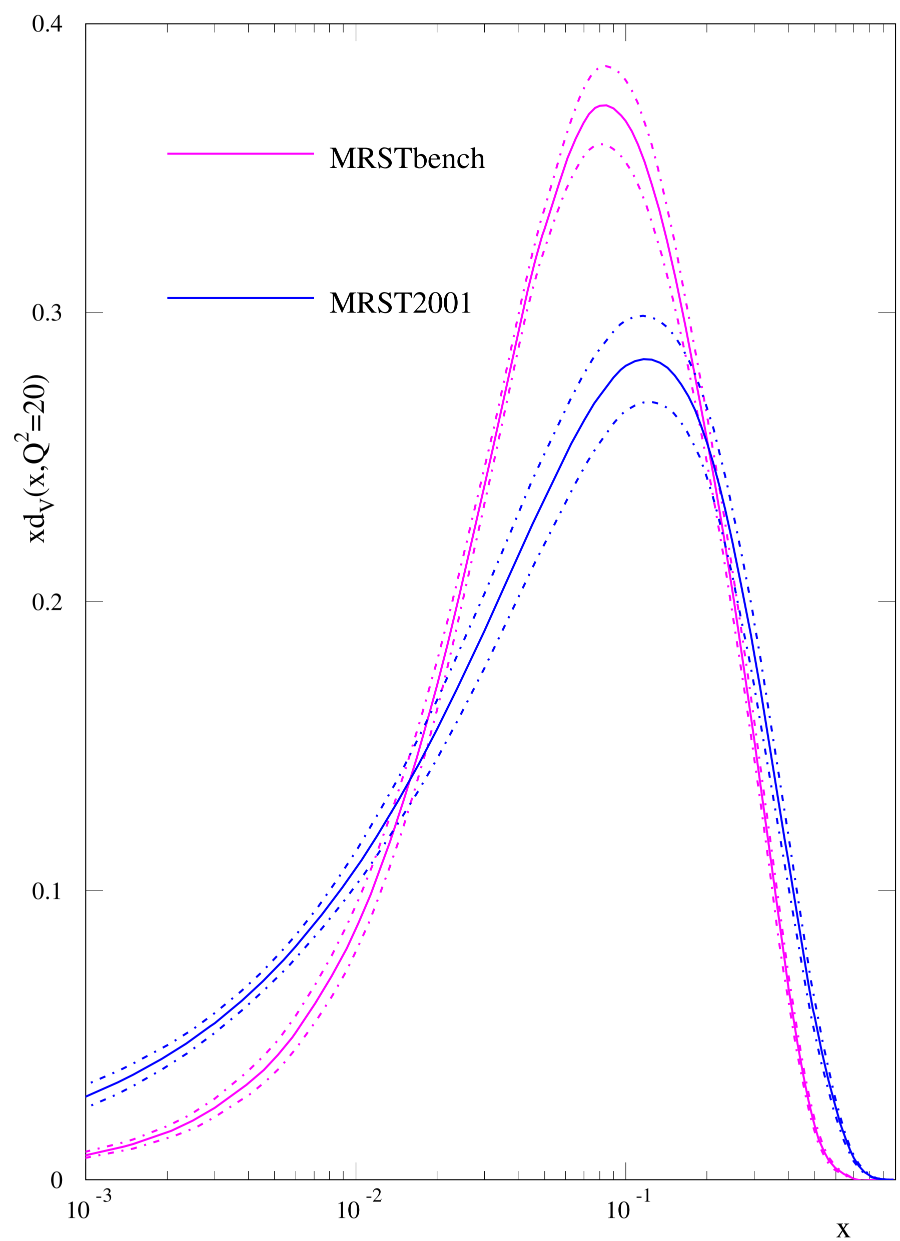


However, how do partons from very conservative, structure function only data compare to global partons?

Compare to MRST01 partons with uncertainty from $\Delta\chi^2 = 50$.

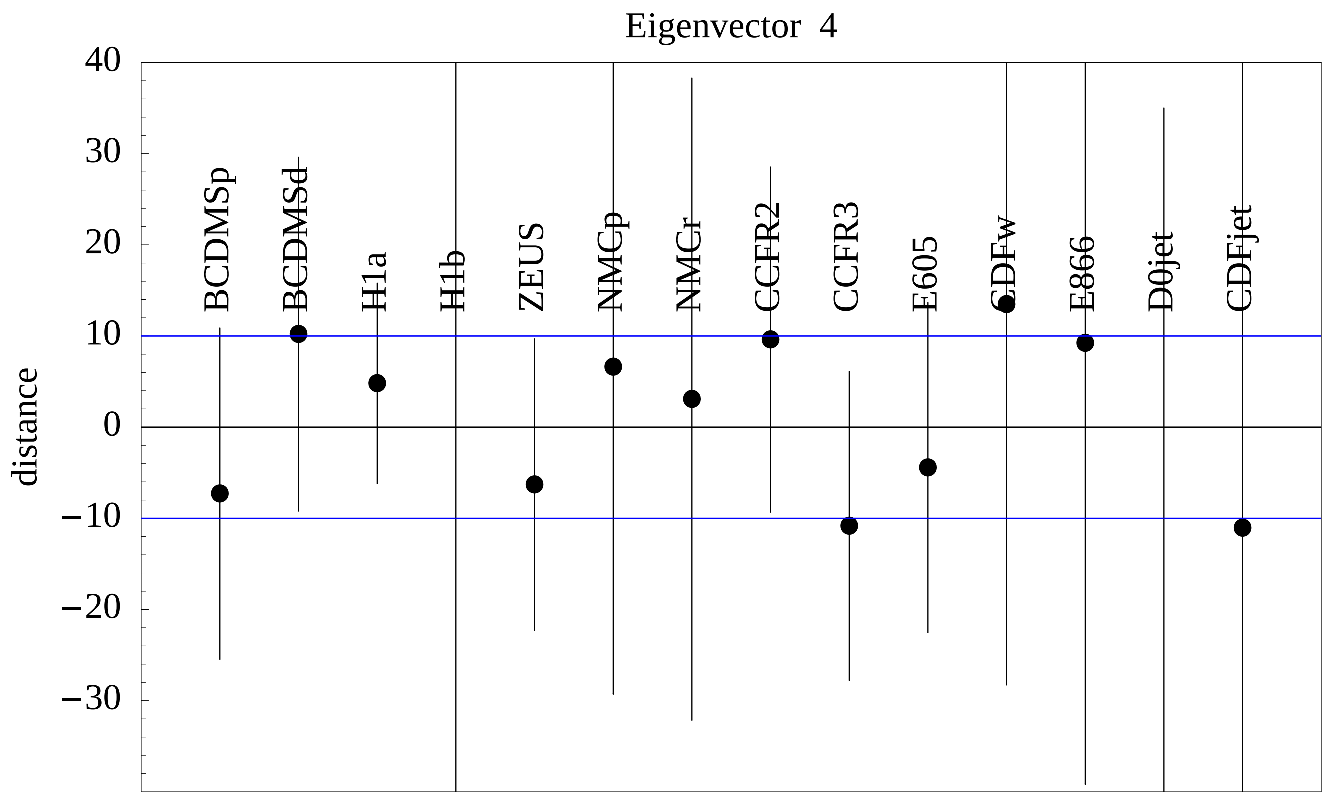
Enormous difference in central values.

Errors similar.



Previous reasoning, allow $\Delta\chi^2$ to take a value such that every data set remains roughly within its 90% confidence limit compared to the χ^2 at best global fit.

These limits shown for CTEQ6 eigenvector 4 as function of $T = \sqrt{\Delta\chi^2}$. Some sets somewhat outside 90% confidence limits for $T = 10$



Using similar sort of reasoning MRST used $\Delta\chi^2 \sim 50$ for 90% confidence level on partons. Still same basic idea but more sophisticated.

Explained below (Watt DIS08)

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

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

- 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

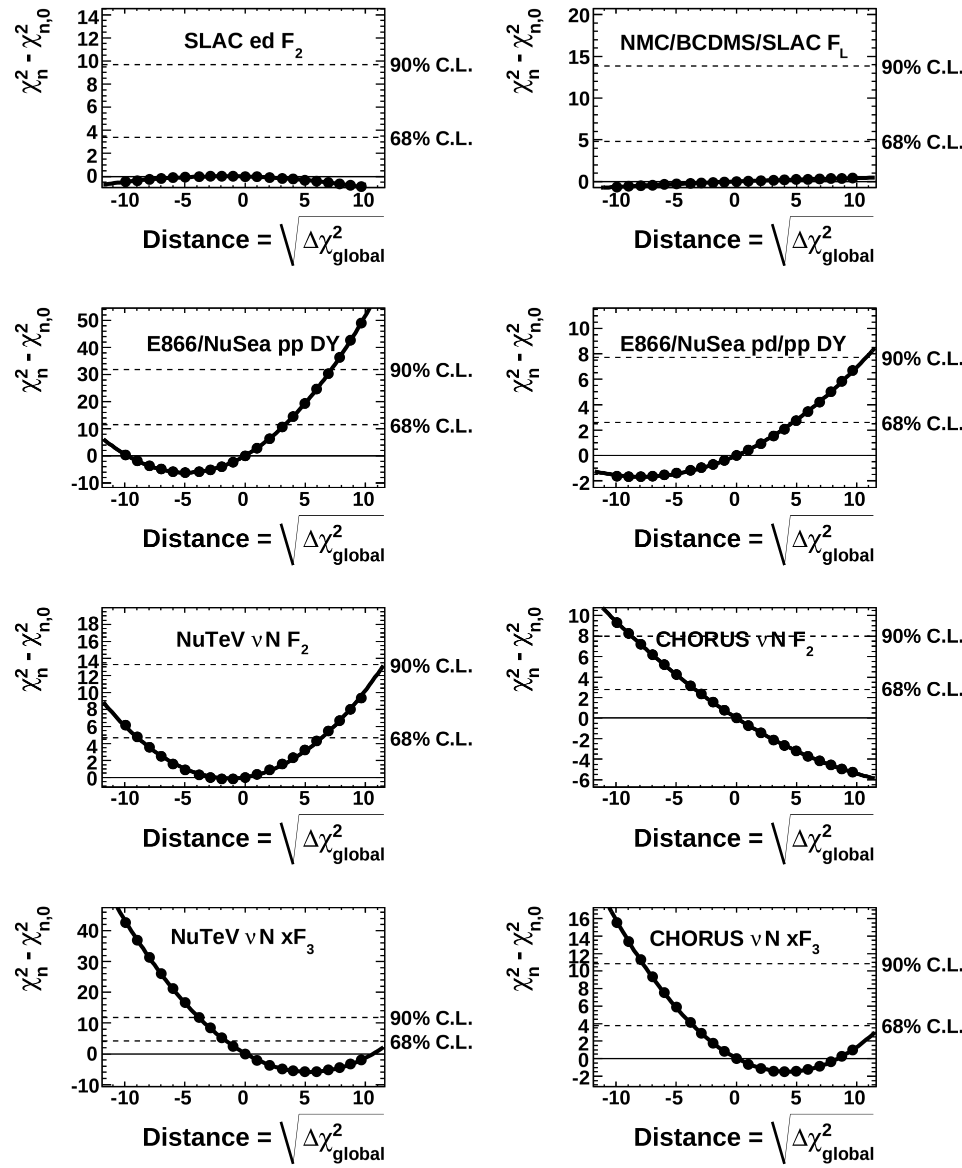
$$f(z;N) = \frac{z^{N/2-1}\,\mathrm{e}^{-z/2}}{2^{N/2}\,\Gamma(N/2)}.$$

- $\xi_{50} \simeq N_n$ 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.

For eigenvector 13, for example, the change in χ^2 for the most sensitive data sets is shown.

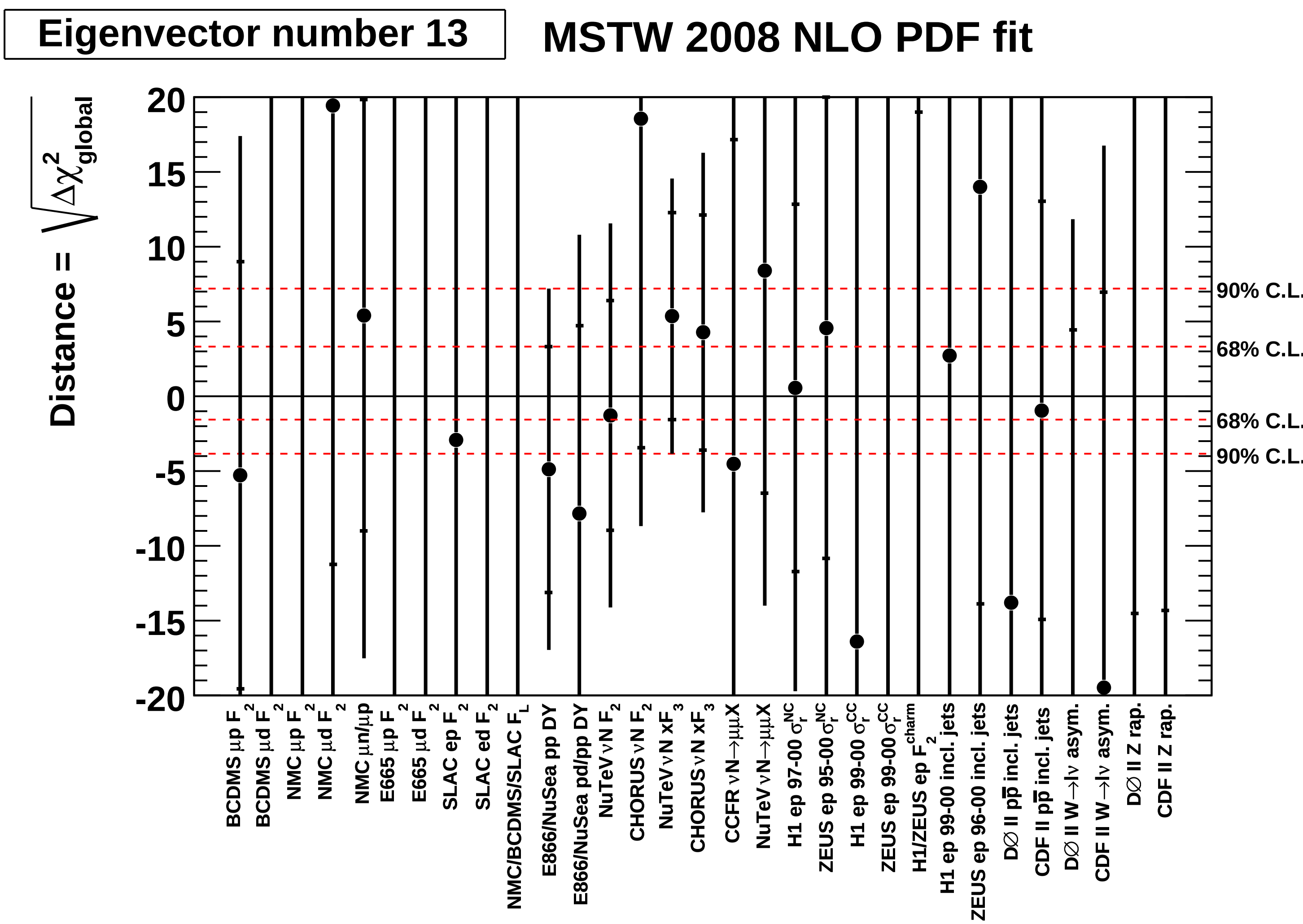
For each determine the point in $\Delta\chi^2_{\rm global}$ at which the appropriate confidence level limit is reached in each direction.

MSTW 2008 NLO PDF fit
Eigenvector number 13

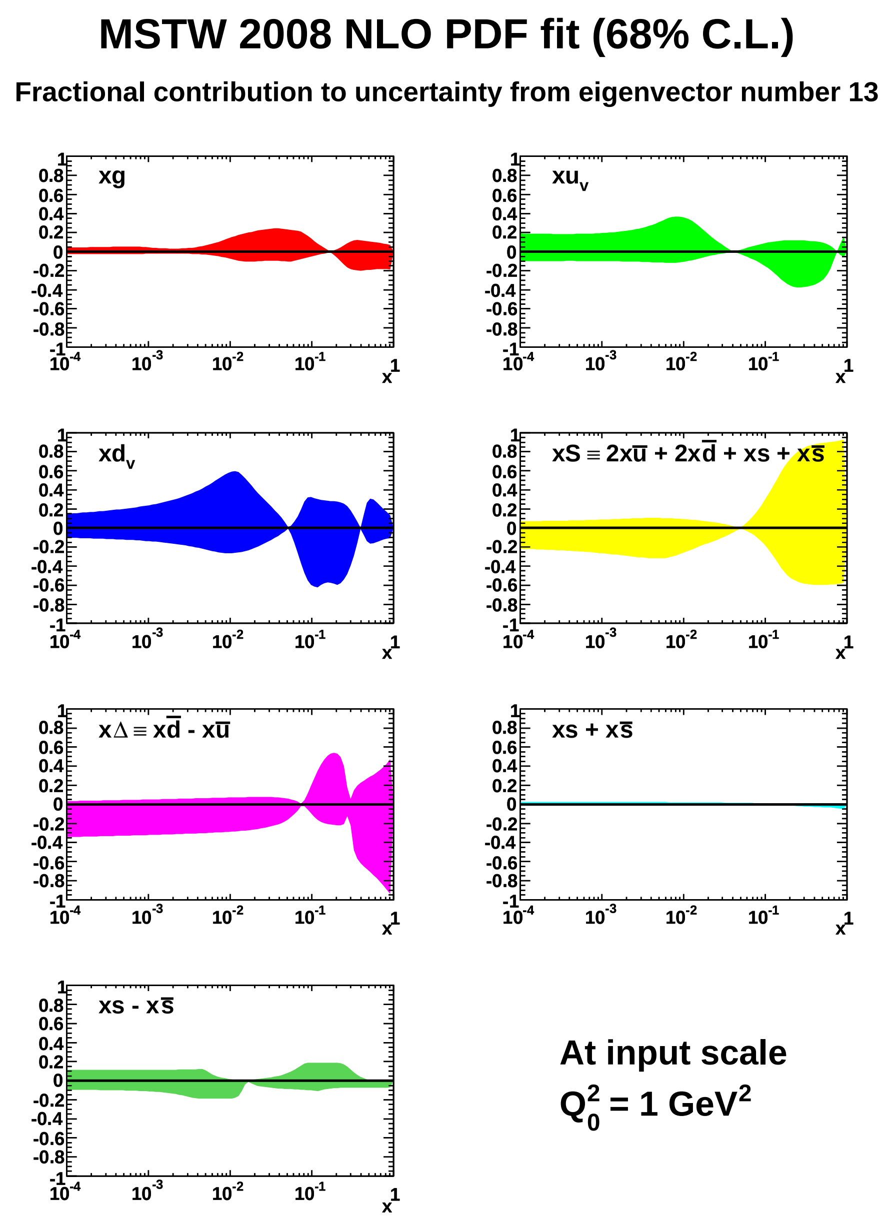


Plot this for all data sets for a given eigenvector.

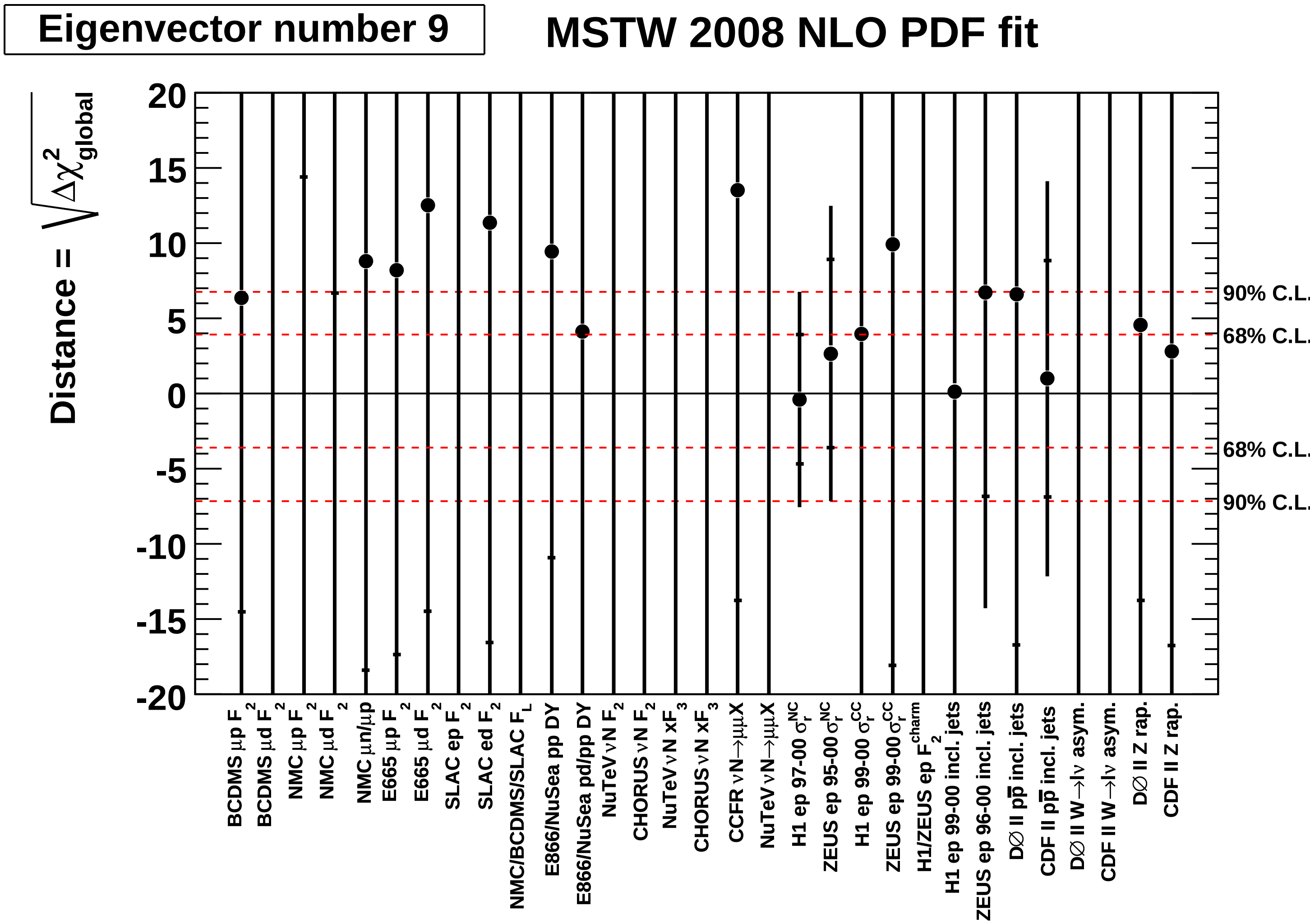
Eigenvector 13 constrained in one direction by E866 Drell-Yan data and in the other direction by NuTeV $F_3^p(x,Q^2)$ data . In this case the best fits for the two sets are highly inconsistent. $\Delta\chi^2 = 100$ well outside 90% confidence level for each.



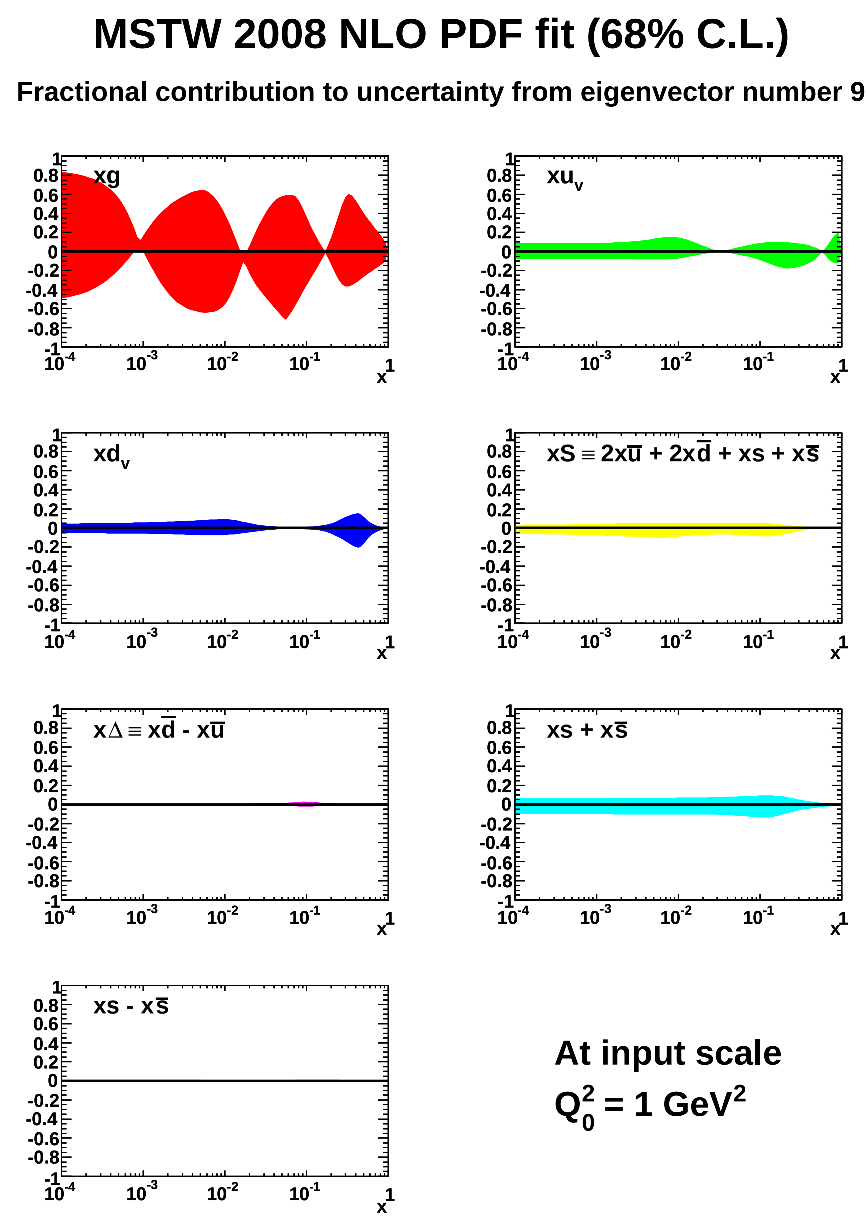
This eigenvector contributes most to the high-x sea quark uncertainty, but also a variety of other quarks.



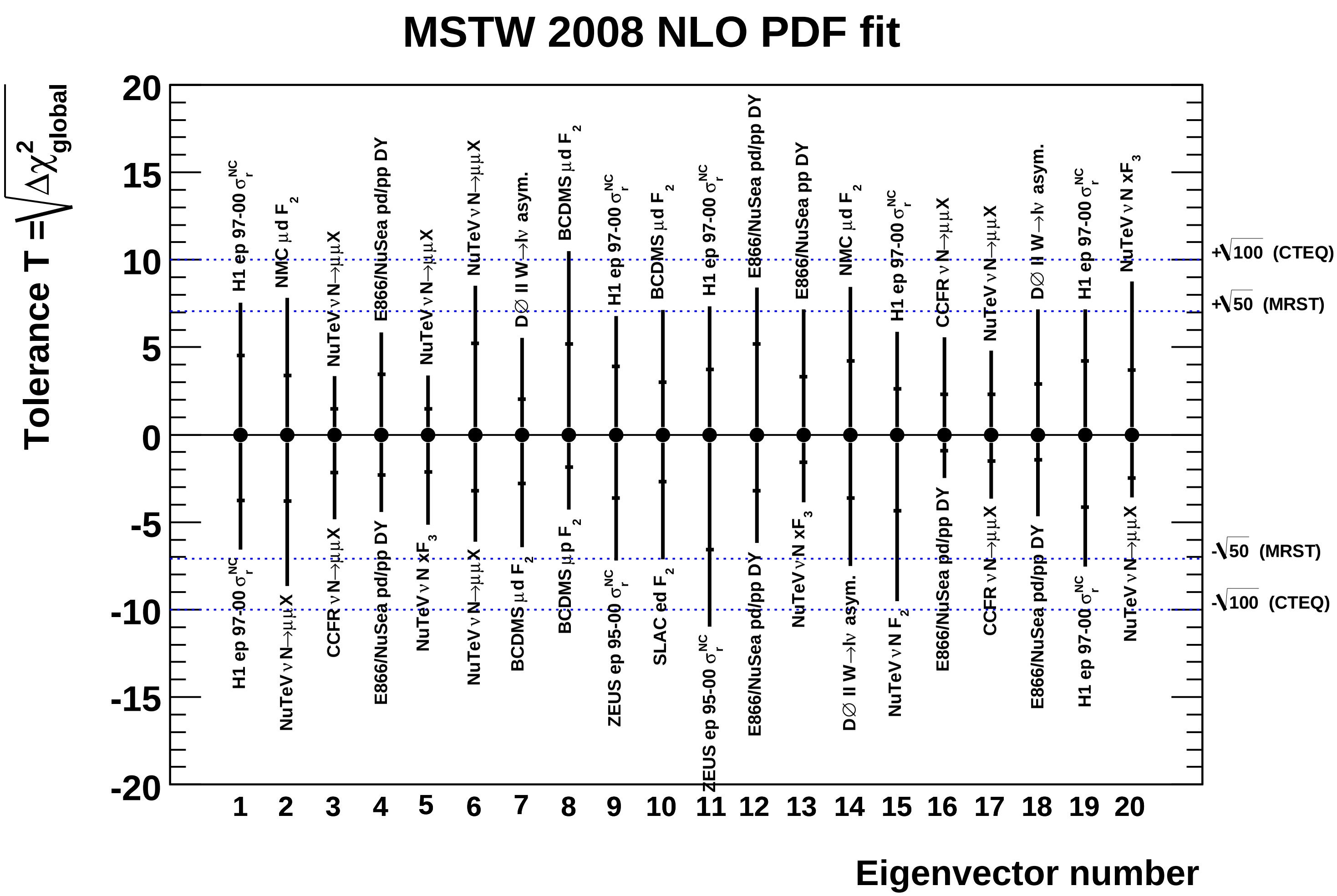
As a simpler example, eigenvector 9 constrained most by H1 and ZEUS data on $F_2^p(x,Q^2)$. 90% confidence limit determining by ZEUS in up direction and H1 in $down$ direction. Both $\Delta\chi^2 \approx 50$.



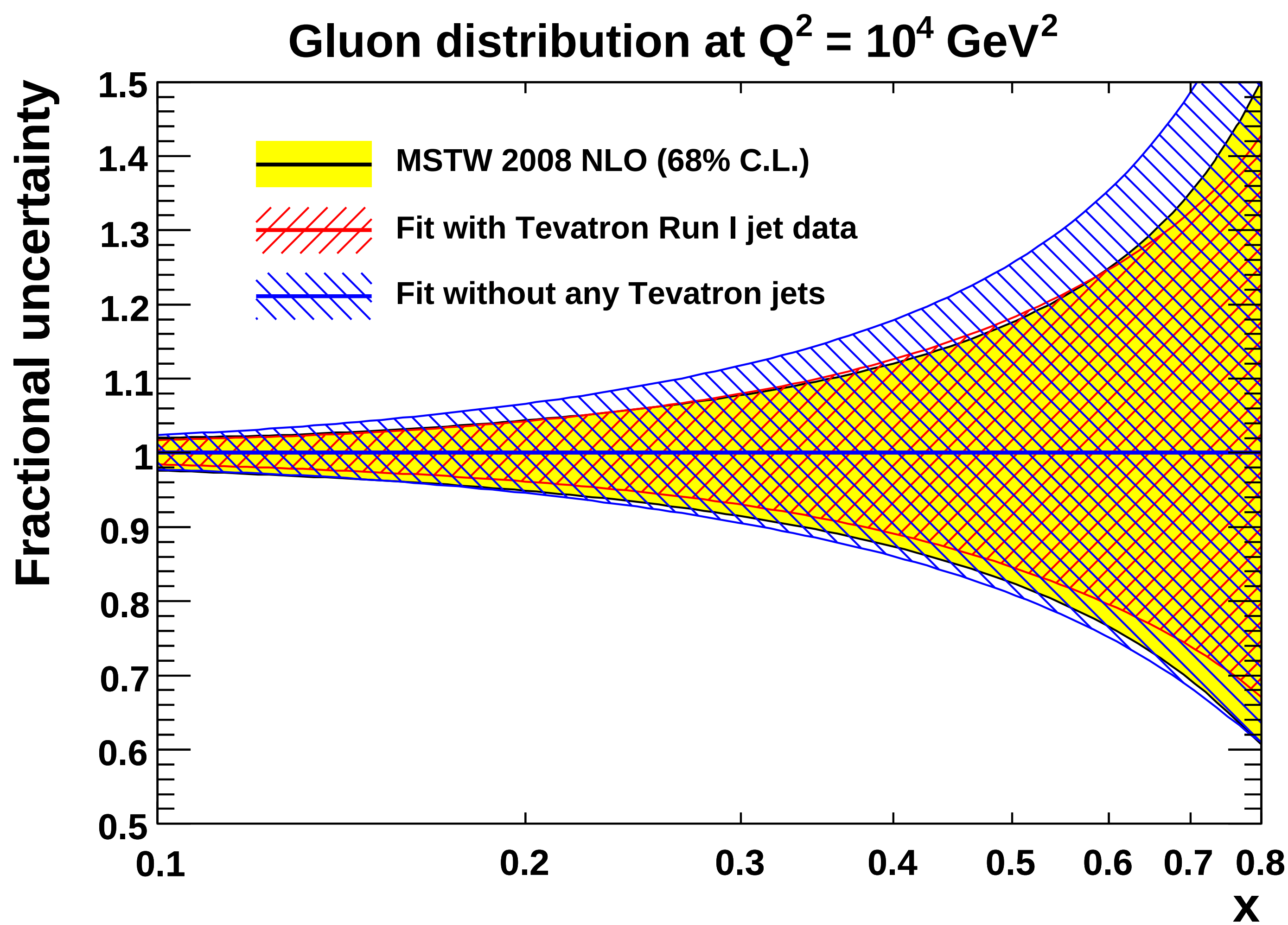
Not surprising this eigenvector contributes most to the gluon uncertainty.



Approach repeated for all 20 eigenvectors to determine uncertainty on each. On average $\Delta\chi^2 = 40$ for 90% and $\Delta\chi^2 = 15$ for $1-\sigma$, but large variations, and asymmetries.



Even though one data set constrains each eigenvector limit, doesn't mean others do not contribute.



Normalisation Uncertainties

Previously the normalization of each data set was determined by the best fit – and then fixed.

Technical difficulties in including this feature in uncertainties.

Now implement procedure of allowing normalisations of all sets to vary in best fit and scan over eigenvectors, with penalty term for each set

$$\chi^2_{\mathcal{N}} = \left(\frac{1-\mathcal{N}}{\sigma^{\mathcal{N}}}\right)^4$$

Quartic penalty avoids very large deviations. Still shift down at LO (fit failure) and slightly at NLO.

Rescale errors with normalization to avoid bias (D'Agostini).

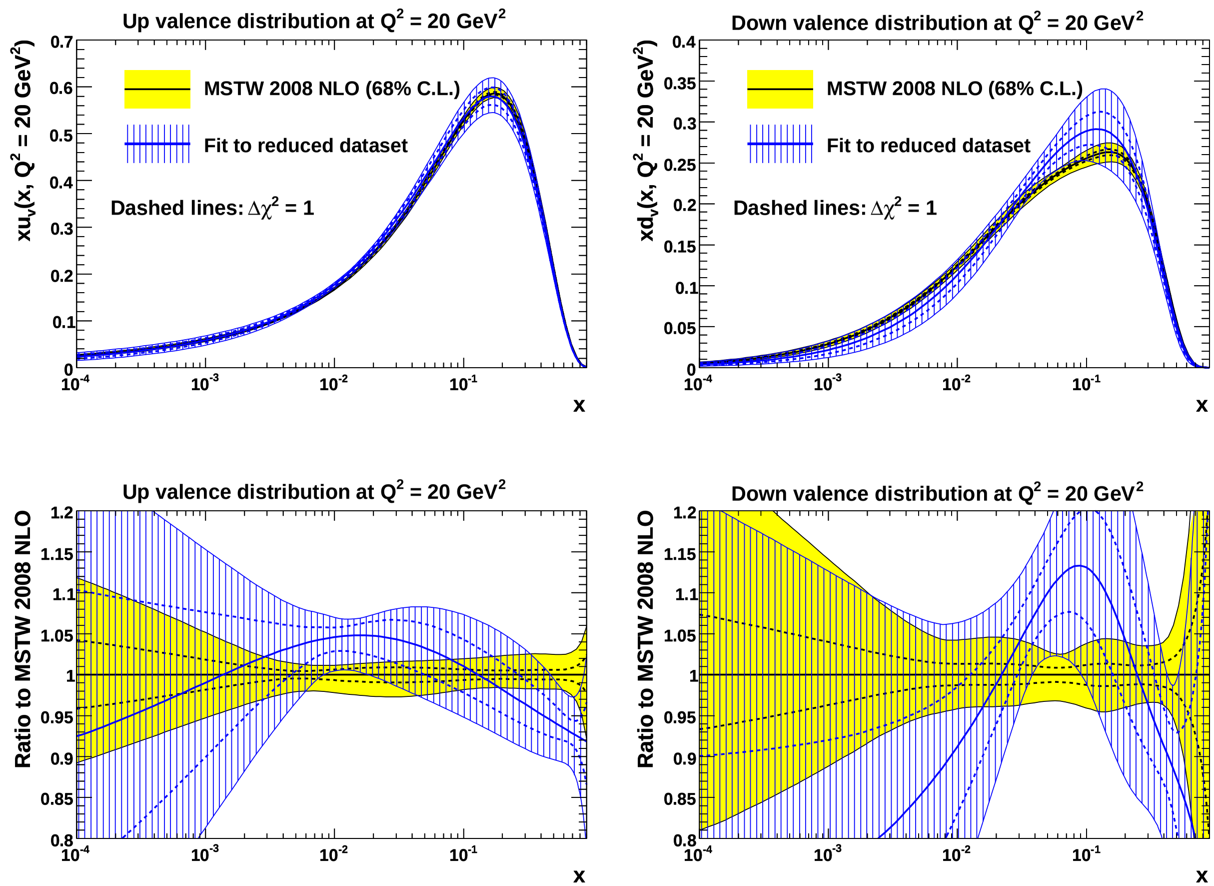
Data set	$\sigma_n^{\mathcal{N}}$	LO	NLO	NNLO
BCDMS μp F_2 [32]	3%	0.9667	0.9644	0.9678
BCDMS μd F_2 [102]	3%	0.9667	0.9644	0.9678
NMC μp F_2 [33]	2%	1.0083	0.9982	0.9999
NMC μd F_2 [33]	2%	1.0083	0.9982	0.9999
NMC $\mu n/\mu p$ [103]	—	1	1	1
E665 μp F_2 [104]	1.85%	1.0146	1.0052	1.0024
E665 μd F_2 [104]	1.85%	1.0146	1.0052	1.0024
SLAC ep F_2 [105, 106]	1.9%	1.0227	1.0125	1.0078
SLAC ed F_2 [105, 106]	1.9%	1.0227	1.0125	1.0078
NMC/BCDMS/SLAC F_L [32–34]	—	1	1	1
E866/NuSea pp DY [107]	6.5%	1.0629	1.0086	1.0868
E866/NuSea pd/pp DY [108]	—	1	1	1
NuTeV νN F_2 [37]	2.1%	0.9987	0.9997	0.9992
CHORUS νN F_2 [38]	2.1%	0.9987	0.9997	0.9992
NuTeV νN xF_3 [37]	2.1%	0.9987	0.9997	0.9992
CHORUS νN xF_3 [38]	2.1%	0.9987	0.9997	0.9992
CCFR $\nu N \to \mu\mu X$ [39]	2.1%	0.9987	0.9997	0.9992
NuTeV $\nu N \to \mu\mu X$ [39]	2.1%	0.9987	0.9997	0.9992
H1 MB 99 e^+p NC [31]	1.3%	0.9861	1.0098	1.0090
H1 MB 97 e^+p NC [109]	1.5%	0.9863	0.9921	0.9953
H1 low Q^2 96–97 e^+p NC [109]	1.7%	1.0029	1.0095	1.0172
H1 high Q^2 98–99 e^-p NC [110]	1.8%	0.9782	0.9851	0.9860
H1 high Q^2 99–00 e^+p NC [35]	1.5%	0.9762	0.9834	0.9842
ZEUS SVX 95 e^+p NC [111]	1.5%	0.9944	0.9948	1.0004
ZEUS 96–97 e^+p NC [112]	2%	0.9735	0.9811	0.9871
ZEUS 98–99 e^-p NC [113]	1.8%	0.9771	0.9855	0.9862
ZEUS 99–00 e^+p NC [114]	2.5%	0.9656	0.9761	0.9762
H1 99–00 e^+p CC [35]	1.5%	0.9762	0.9834	0.9842
ZEUS 99–00 e^+p CC [36]	2.5%	0.9656	0.9761	0.9762
H1/ZEUS ep F_2^{charm} [41–47]	—	1	1	1
H1 99–00 e^+p incl. jets [59]	1.5%	0.9762	0.9834	—
ZEUS 96–97 e^+p incl. jets [57]	2%	0.9735	0.9811	—
ZEUS 98–00 $e^\pm p$ incl. jets [58]	2.5%	0.9656	0.9761	—
DØ II $p\bar{p}$ incl. jets [56]	6.1%	0.9353	1.0596	1.0759
CDF II $p\bar{p}$ incl. jets [54]	5.8%	0.8779	0.9646	0.9900
CDF II $W \to \ell\nu$ asym. [48]	—	1	1	1
DØ II $W \to \ell\nu$ asym. [49]	—	1	1	1
DØ II Z rap. [53]	—	1	1	1
CDF II Z rap. [52]	5.8%	0.8779	0.9646	0.9900

In practice should give a conservative estimation of uncertainties.

Can investigate by repeating HERA-LHC Workshop exercise of obtaining PDFs by fitting to DIS data with conservative cuts only.

Comparison of normal and benchmark sets shown.

Latter have greater uncertainty. Compatibility using *dynamical tolerance* uncertainty approach, but not using $\Delta\chi^2 = 1$.

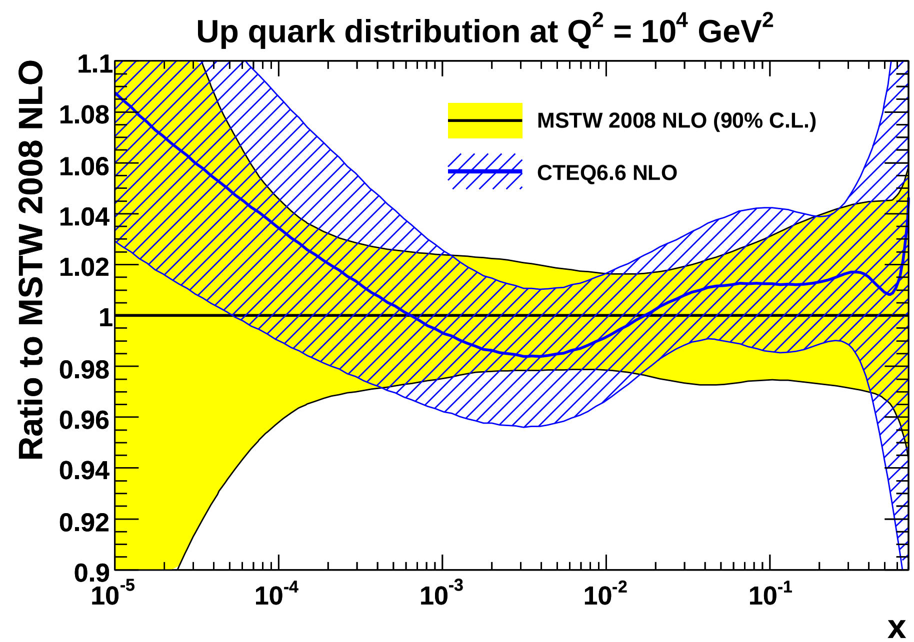


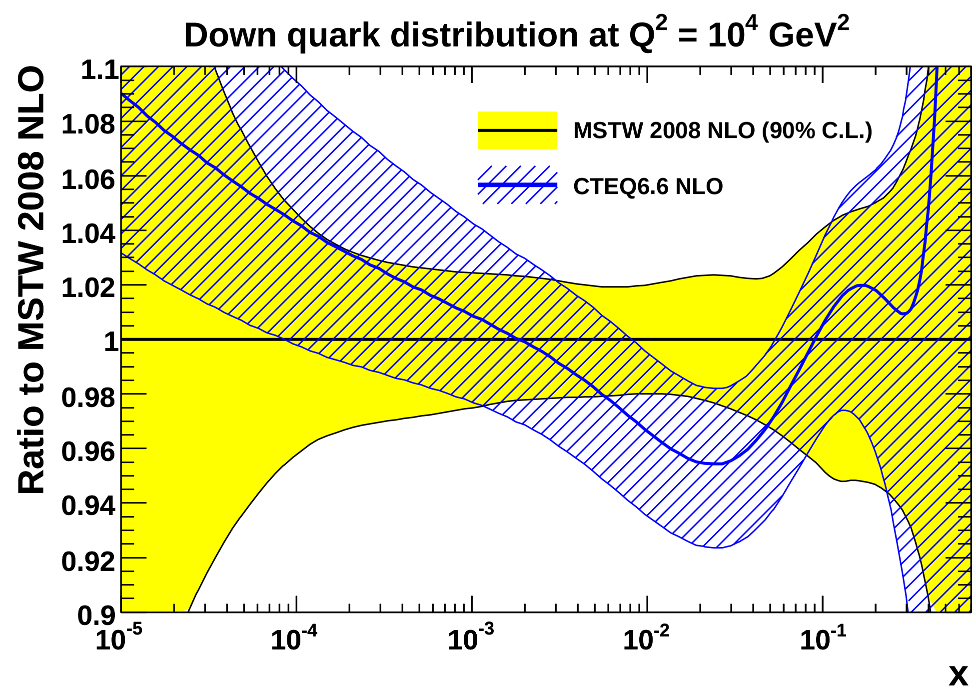
Still lack of compatibility some places, e.g high-x gluon.

χ^2 for benchmark data $458/589$ in reduced fit $\rightarrow 526/589$ within global fit.

Uncertainty on MSTW u and d distributions, along with CTEQ6.6.

Only reasonable agreement between groups despite inflated tolerance.





Different PDF sets

- MSTW08 – fit all previous types of data. Most up-to-date Tevatron jet data. Not most recent HERA combination of data. PDFs at LO, NLO and NNLO.

- CTEQ6.6 – very similar. Not quite as up-to-date on Tevatron data. PDFs at NLO.

- NNPDF2.0 – include all above except HERA jet data (not strongest constraint) and heavy flavour structure functions. Include HERA combined data. PDFs at NLO.

- HERAPDF2.0 – based entirely on HERA inclusive structure functions, neutral and charged current. Use combined data. PDFs at LO, NLO.

- ABKM09 – fit to DIS and fixed target Drell-Yan data. PDFs at NLO and NNLO.

- GJR08 – fit to DIS, fixed target Drell-Yan and Tevatron jet data. PDFs at NLO and NNLO.

Determination of best fit and uncertainties

All but NNPDF minimise χ^2 and define eigenvectors of parameter combinations expanding about best fit.

- MSTW08 – 20 eigenvectors. Due to incompatibility of different sets and (perhaps to some extent) parameterisation inflexibility (little direct evidence for this) have inflated $\Delta\chi^2$ of $5-20$ for eigenvectors.

- CTEQ6.6 – 22 eigenvectors. Inflated $\Delta\chi^2$ of 50 for 1 sigma for eigenvectors (no normalization uncertainties in CTEQ6.6).

- HERAPDF2.0 – 9 eigenvectors. Use "$\Delta\chi^2 = 1$". Additional model and parameterisation uncertainties.

- ABKM09 – 21 parton parameters. Use $\Delta\chi^2 = 1$. Also α_S, m_c, m_b.

- GJR08 – 20 parton parameters and α_S. Use $\Delta\chi^2 \approx 20$. Impose strong theory constraint on input form of PDFs.

Perhaps surprisingly all get rather similar uncertainties for PDFs cross-sections.

Neural Network group (Ball *et al.*) limit parameterization dependence. Leads to alternative approach to "best fit" and uncertainties.

First part of approach, no longer perturb about best fit. Construct a set of Monte Carlo replicas $F_{i,p}^{art,k}$ of the original data set $F_{i,p}^{exp,(k)}$.

- PDFs ARE FITTED TO DATA REPLICAS
- REPLICAS FLUCTUATE ABOUT CENTRAL DATA:
$$F_{i,p}^{(art)(k)} = S_{p,N}^{(k)} F_{i,p}^{\mathrm{exp}} \left(1 + r_p^{(k)} \sigma_p^{\mathrm{stat}} + \sum_{j=1}^{N_{\mathrm{sys}}} r_{p,j}^{(k)} \sigma_{p,j}^{\mathrm{sys}}\right)$$

Where $r_p^{(k)}$ are random numbers following Gaussian distribution, and $S_{p,N}^{(k)}$ is the analogous normalization shift of the of the replica depending on $1 + r_{p,n}^{(k)} \sigma_p^{norm}$.

Hence, include information about measurements and errors in distribution of $F_{i,p}^{art,(k)}$.

Fit to the replicas of the data obtaining a set of PDF replicas $q_i^{(net)(k)}$ (follows Giele *et al.*)

Mean μ_O and deviation σ_O of observable O then given by

$$\mu_O = \frac{1}{N_{rep}} \sum_1^{N_{rep}} O[q_i^{(net)(k)}], \quad \sigma_O^2 = \frac{1}{N_{rep}} \sum_1^{N_{rep}} (O[q_i^{(net)(k)}] - \mu_O)^2.$$

NNPDF approach additionally (largely) eliminates parameterisation dependence by using a neural net which undergoes a series of evolutions (mutations via genetic algorithm) to find the best fit, rather than a fixed parameterisation.

In effect is a much larger sets of parameters – ~ 37 per distribution.

Includes pre-processing exponents as $x \to 1$ and $x \to 0$ to aid convergence of fit,

$$f(x, Q_0^2) = A(1-x)^m x^{-n} NN(x)$$

where n, m are in fairly narrow ranges, so overall behaviour guided at these extremes where data constraints vanish.

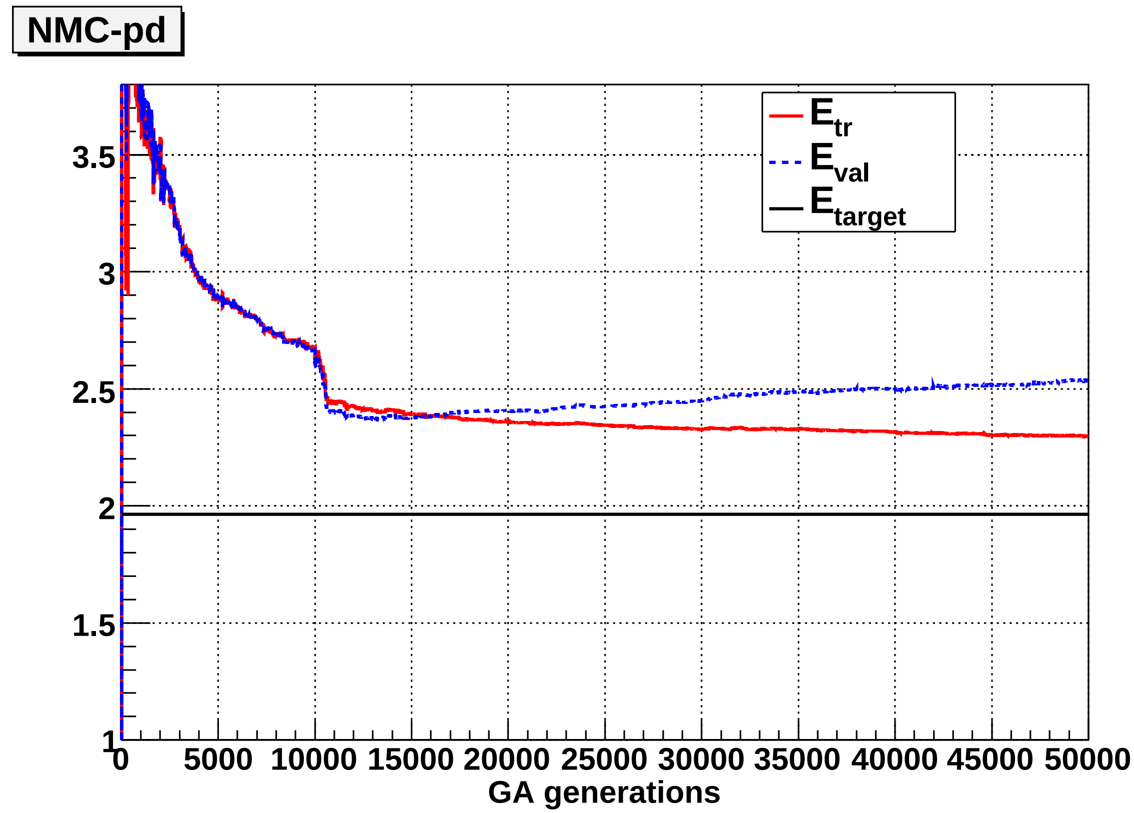
Data included constantly increasing. Recently NNPDF2.0, first global fit of this type.

Freedom in parameterisation means best fit to all data would tend to reproduce data fluctuations (as far as this is possible). Must guard against this.

Split data sets randomly into equal size *training* and *validation* sets.

Fit until quality of fit to validation set starts to go up, even though training set still (hopefully slowly) improving.

Criterion for stopping the fit not simply value of error function (analogous to χ^2) for full global data set, but split into different data sets.



In earlier versions weighted error function for different data sets in early stages to try to give all sets a similar quality fit.

Difficult to know when to stop (analogous to variable $\Delta\chi^2$ in other approaches?).

ARE WE CONSTRAINED BY THE FUNCTIONAL FORM?
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
$\langle\chi^2\rangle_{\mathrm{rep}}$	2.79 ± 0.24	1.65 ± 0.20	$\sim 1.6 \pm 0.2$	2.43 ± 0.13	1.29 ± 0.06
$\langle\chi^2_{\mathrm{tr}}\rangle_{\mathrm{rep}}$	2.76	1.59	$\sim$1.6	2.40	1.27
$\langle\chi^2_{\mathrm{val}}\rangle_{\mathrm{rep}}$	2.80	1.61	$\sim$1.6	2.47	1.30
$\langle\sigma^{\mathrm{dat}}\rangle$	0.039	0.035	$\sim$0.03	0.032	0.019

χ^2 OF THE GLOBAL FIT DECREASES A LOT!

Some evidence not at best fit in previous versions (Forte, DESY Oct 2009).

Quality of global fit for both training and validation decreasing significantly after stopping point.

Fluctuations in sets smaller with longer stopping.

Statistical behaviour (arguably) more like expected for longer stopping?

WHERE IS THE UNCERTAINTY COMING FROM?

WHEN THE BEST FIT IS NOT AT THE MINIMUM

	STANDARD STOPPING			FIXED LONG	
	REPLICAS	CENTRAL VALUE	FIXED PARTITION	REPLICAS	CENTRAL VALUE
χ^2	1.32	1.32	1.35	1.18	1.19
$\langle\chi^2\rangle_{\rm rep}$	2.79 ± 0.24	1.65 ± 0.20	1.60 ± 0.19	2.43 ± 0.13	1.29 ± 0.06
$\langle\sigma^{\rm dat}\rangle$	0.39	0.35	0.28	0.32	0.19

- FIT QUALITY:
  - "FUNCTIONAL" UNCERTAINTY SUPPRESSED IN OVERLEARNING FITS: $\Rightarrow \langle\sigma^{\rm dat}\rangle \approx 0.2 \Rightarrow$ "DATA" UNCERTAINTY
  - FLUCTUATION OF $\langle\chi^2\rangle_{\rm rep}$ FOR OVERLEARNING FIT STATISTICAL: $\sigma = \sqrt{\frac{2}{N_{\rm dat}}} \approx 0.05$

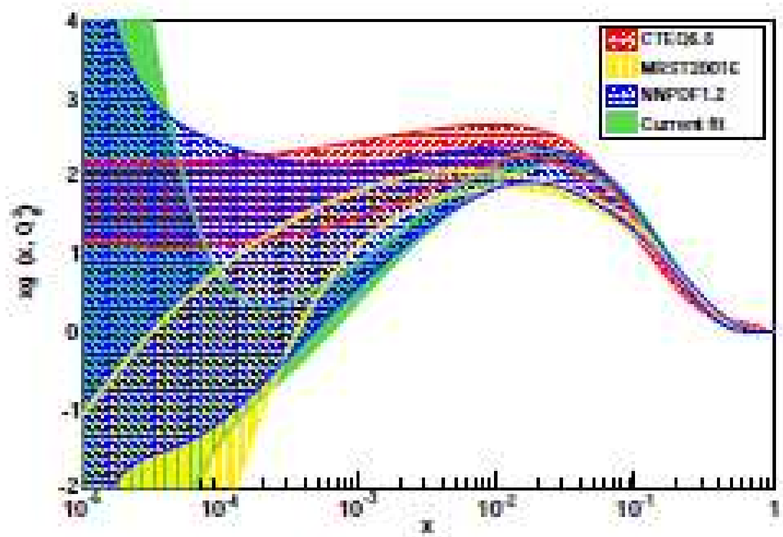
Number of data points ~ 3000 – $\sqrt{2/N_{data}} \sim 0.025$.

Uncertainty from "overlearnt" fits (green) was (normally) rather smaller than default (blue).

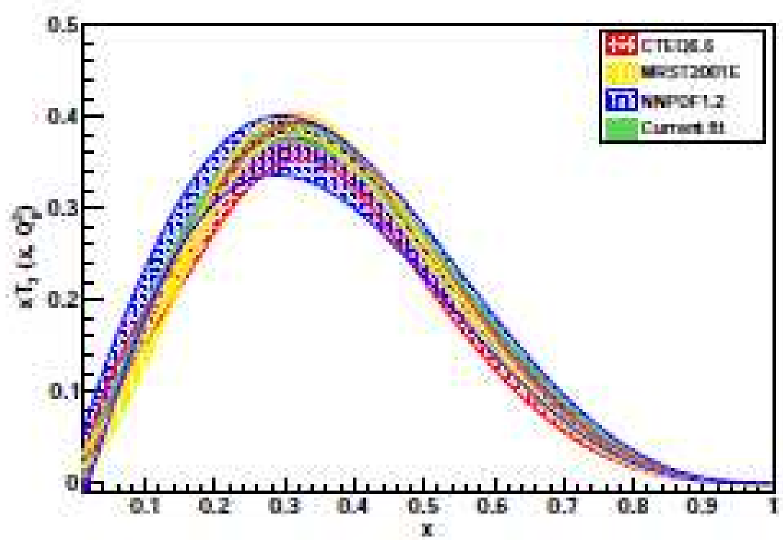
Arguable if lack of smoothness becomes a problem.

→ significant improvements in NNPDF2.0.

GLUON



TRIPLET



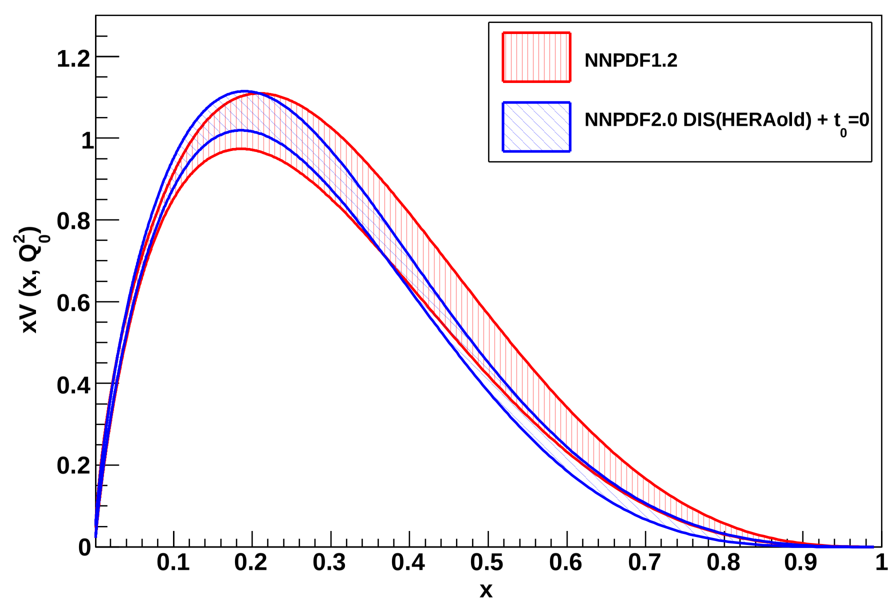
Weighted training in early stages according to a target (determined iteratively), so stopping for global fit more in line with individual sets.

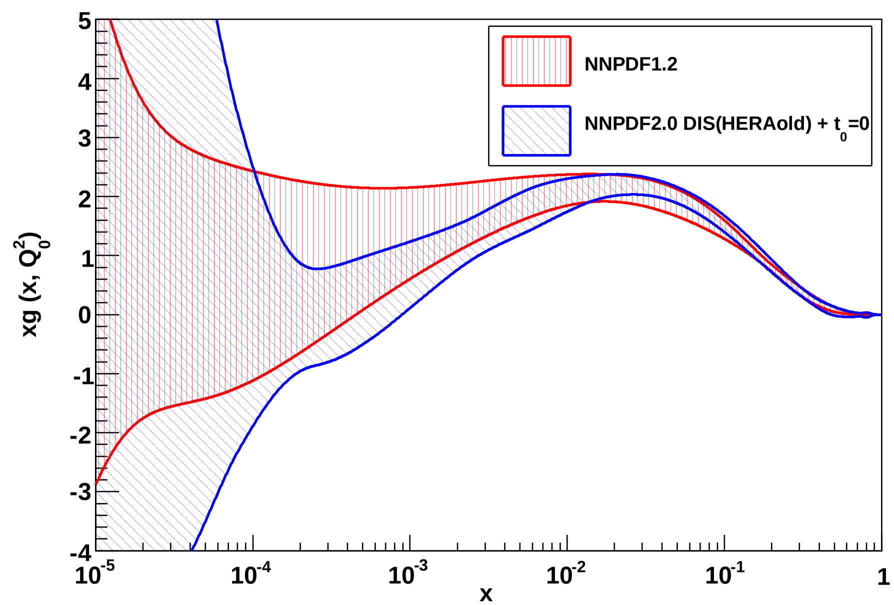
Criterion for increase in fit to validation sets relative to decrease in training sets made more strict.

Significant reductions (usually) in uncertainty in latest version, and changed central values, just due to change in stopping and fitting procedures.

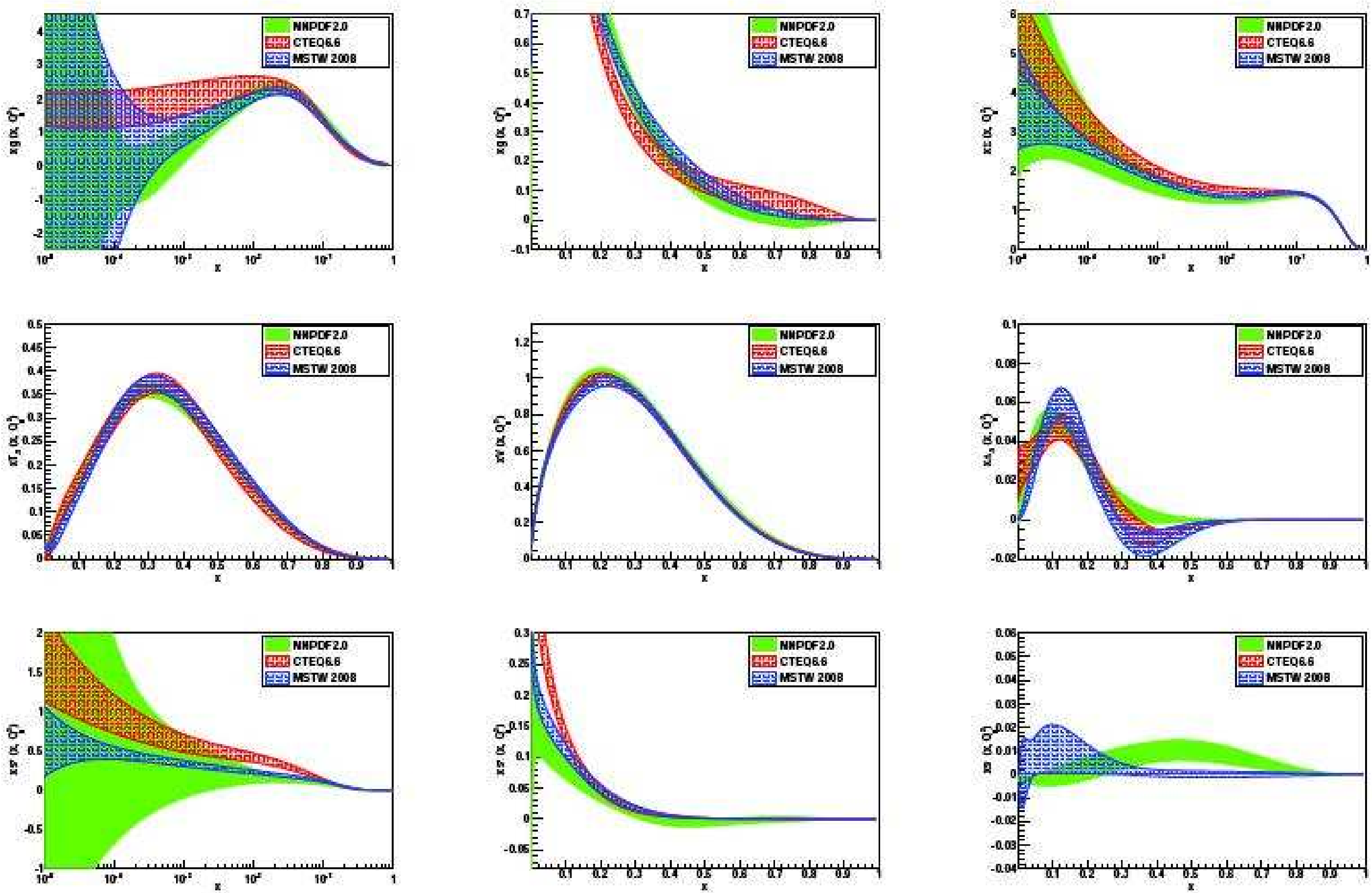
I would suggest uncertainty now more analogous to smaller "$\Delta\chi^2$", but actual value very difficult to ascertain. Fluctuations in error function (and χ^2) still arguably a bit larger than naively expected.

Is there a definitive set of stopping criteria?

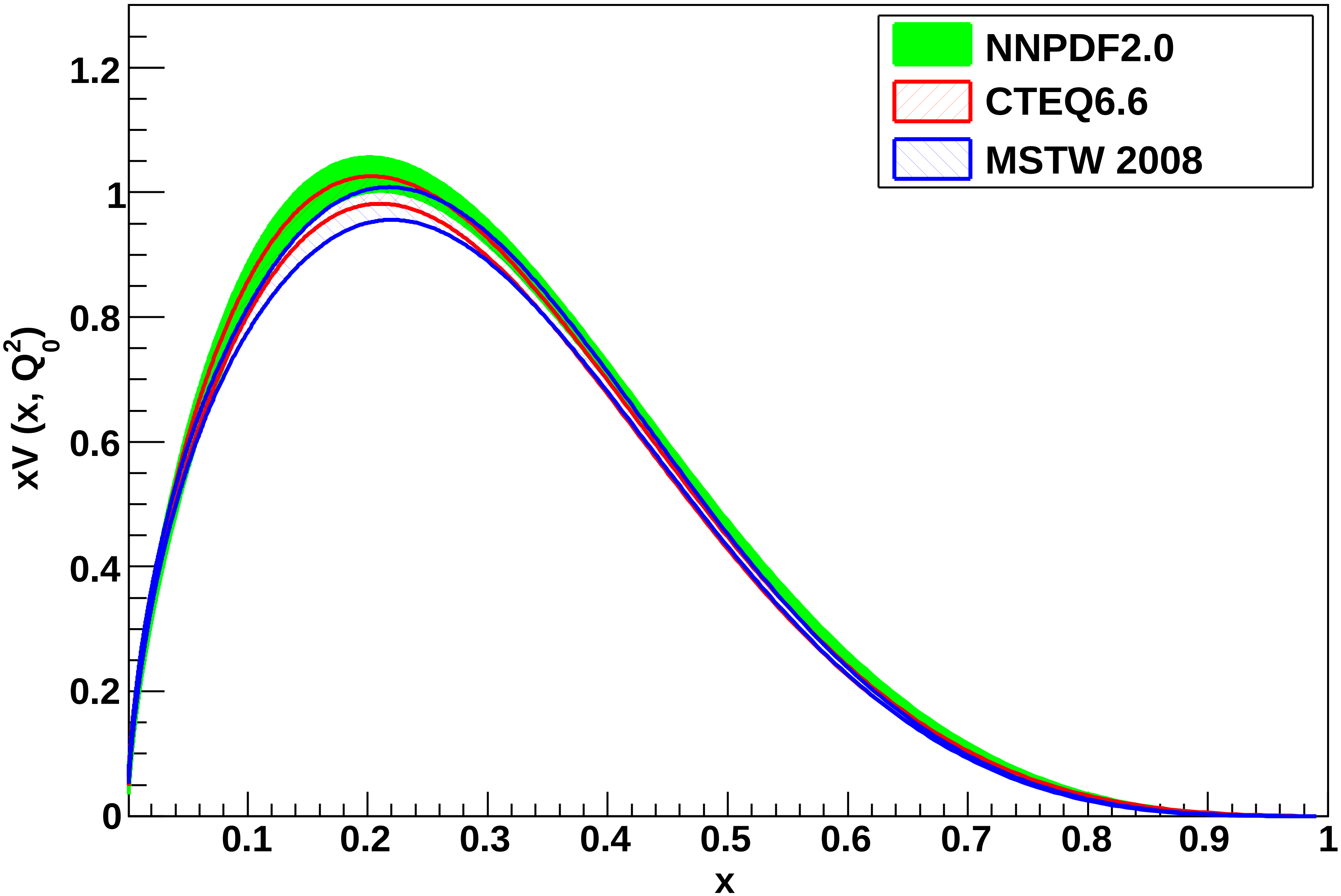




Also reductions in uncertainty due to inclusion of new data. (Improved treatment of normalisations generally increases uncertainty slightly.)

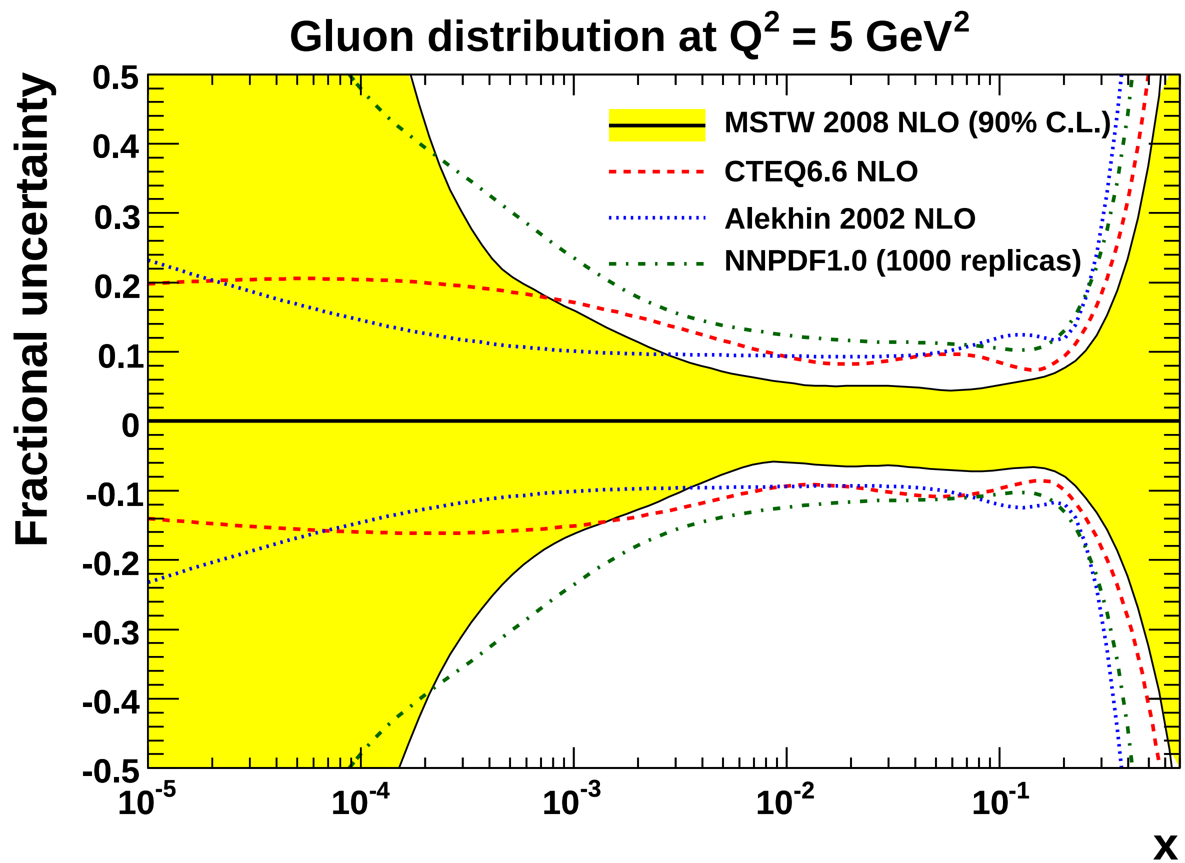


NNPDF uncertainties pretty similar to other groups, with some particular exceptions.



Uncertainties on valence quarks not notably different to other groups at all.

Gluon Parameterisation - small x – different parameterisations lead to very different uncertainty for small x gluon.



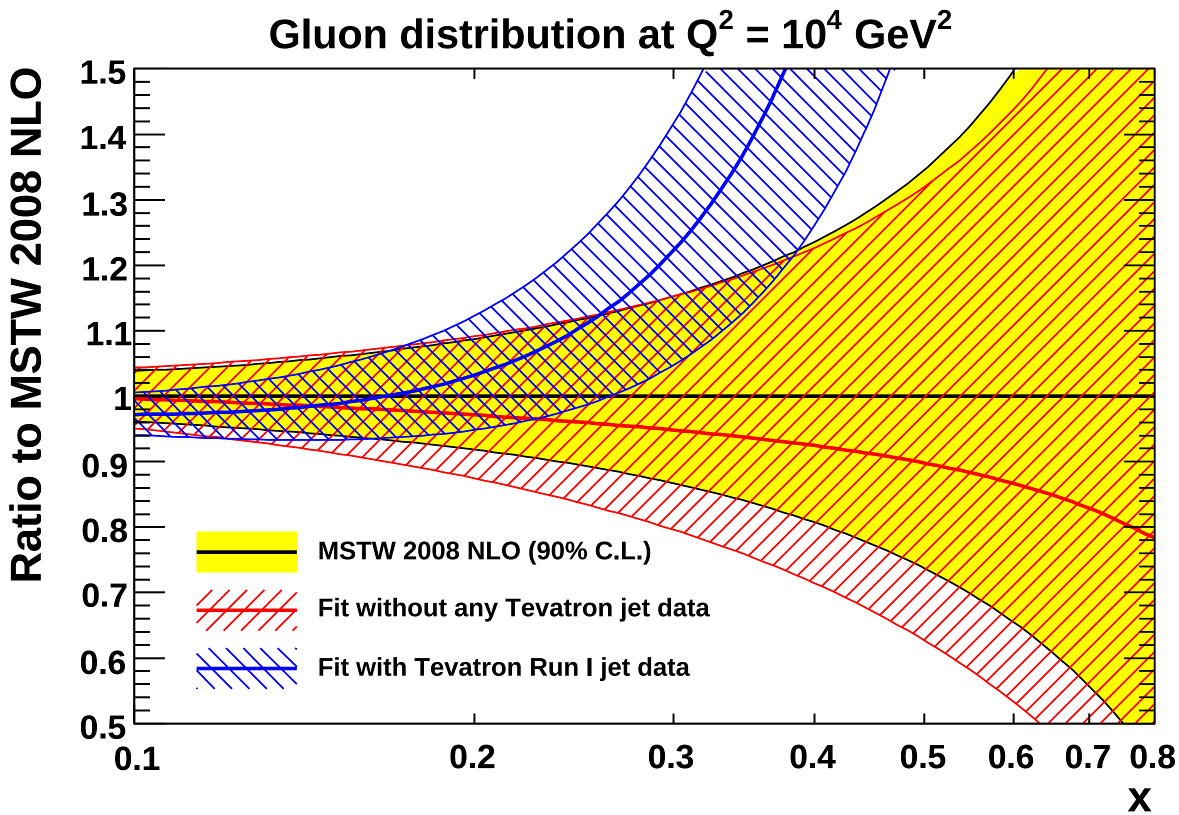
Most assume single power x^λ at input $\rightarrow$ limited uncertainty. If input at low Q^2 λ positive and small-x input gluon *fine-tuned* to ~ 0. Artificially small uncertainty.

If $g(x) \propto x^{\lambda \pm \Delta\lambda}$ then $\Delta g(x) = \Delta\lambda \ln(1/x) * g(x)$.

MRST/MSTW and NNPDF more flexible (can be negative) $\rightarrow$ rapid expansion of uncertainty where data runs out.

Gluon Distribution - large x.

Constrained indirectly, but quite accurately, by DIS data, and directly by Tevatron high-p_T jets, now **Run I** and **Run II** available. *Slightly* confusing picture.



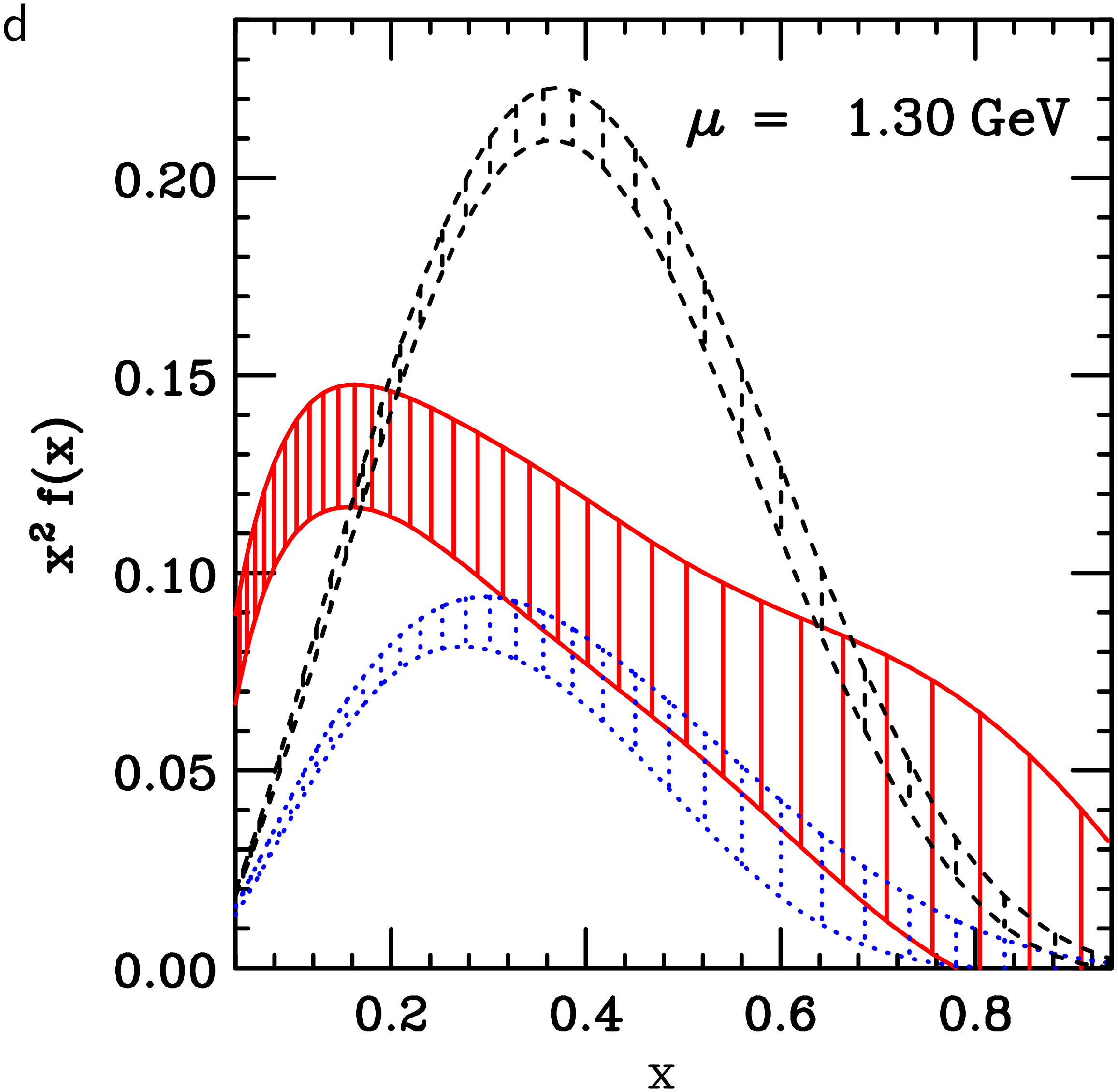
Fit by MSTW and CTEQ and now also NNPDF. Former found gluon much softer for **Run II**. Fits not very consistent between runs.

CTEQ find more compatibility between **Run I** and **Run II** fits.

Generally high-x PDFs parameterised so will behave like $(1-x)^\eta$ as $x \to 1$. More flexibility in CTEQ.

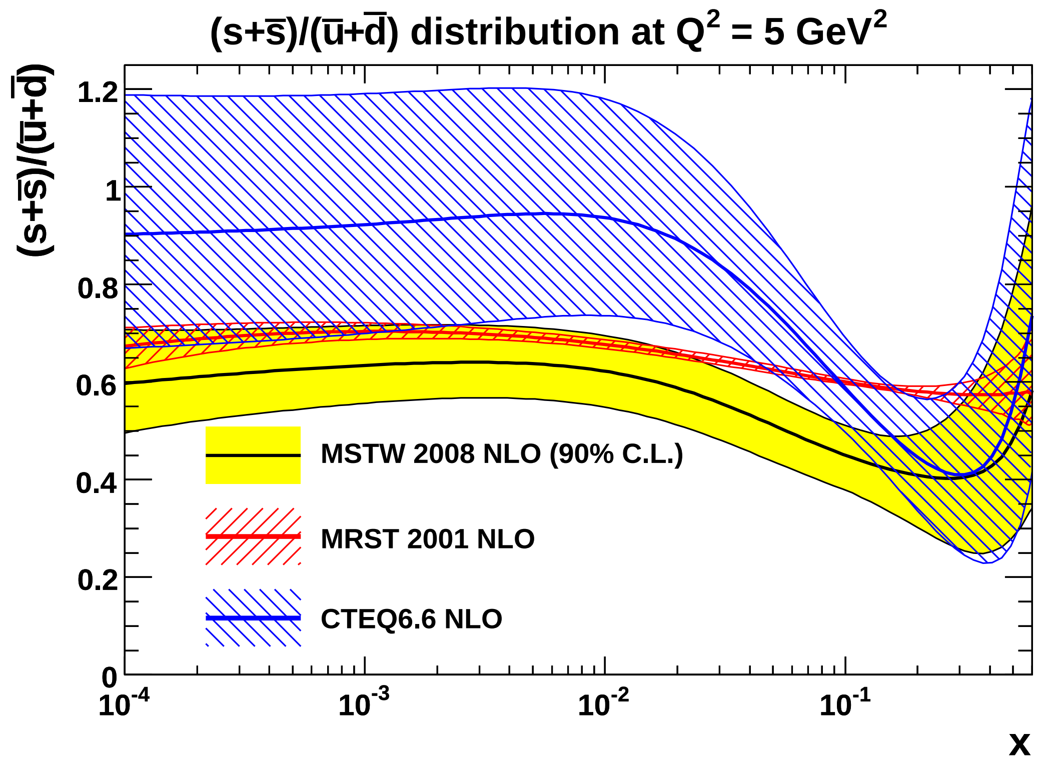
Very hard high-x gluon distribution (more-so even than NNPDF uncertainties).

However, is gluon, which is radiated from quarks, harder than the up valence distribution for $x \to 1$?



Strange Quarks

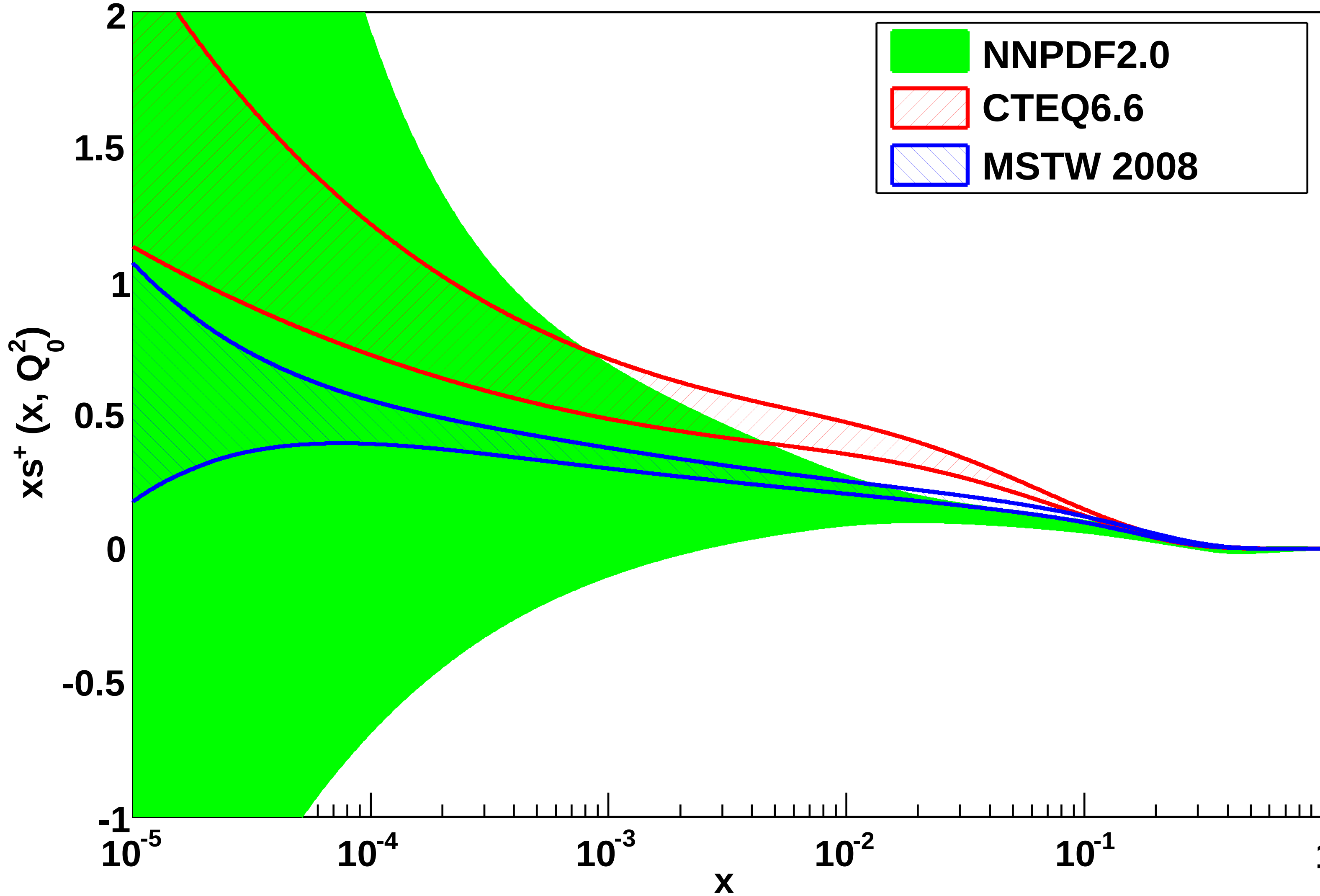
Direct fit to $s, \bar{s}$ from dimuon data leads to significant uncertainty increase compared to assumption of fixed fraction of sea used until recently. Constraint for $x \geq 0.01$.



MSTW assumes shape of strange given by theory assumption that suppression of form of massive quarks. Significantly different to CTEQ fitting to same data assuming only same small-x power for strange as light quarks.

Difference in region of data! Effect of nuclear corrections and/or heavy quark treatment?

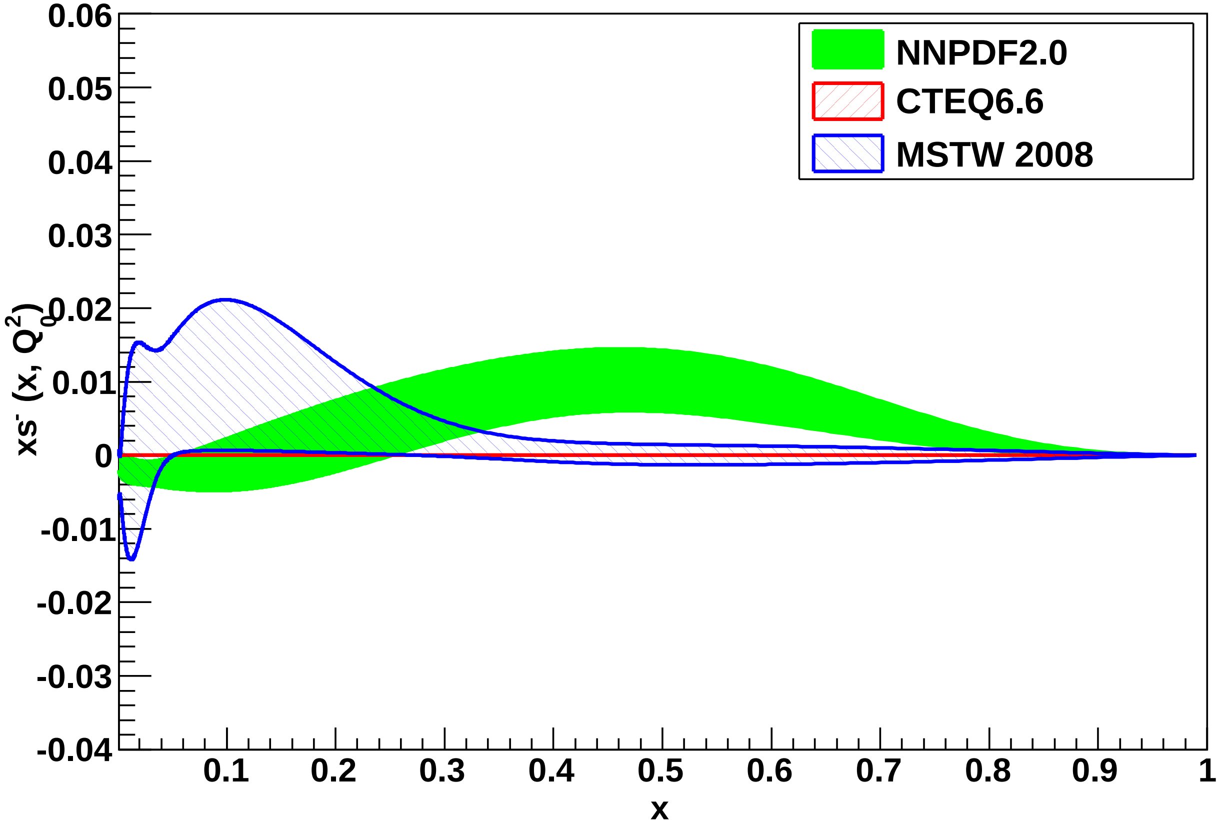
NNPDF2.0, which includes dimuon data, have no theoretical constraint on strange quark distribution at all at small x.



Overestimate of uncertainty? Impacts on small-x light quarks.

Strange asymmetry.

Most recent sets obtain $s - \bar{s}$ for first time from differences in $\nu, \bar{\nu}$ dimuon production.



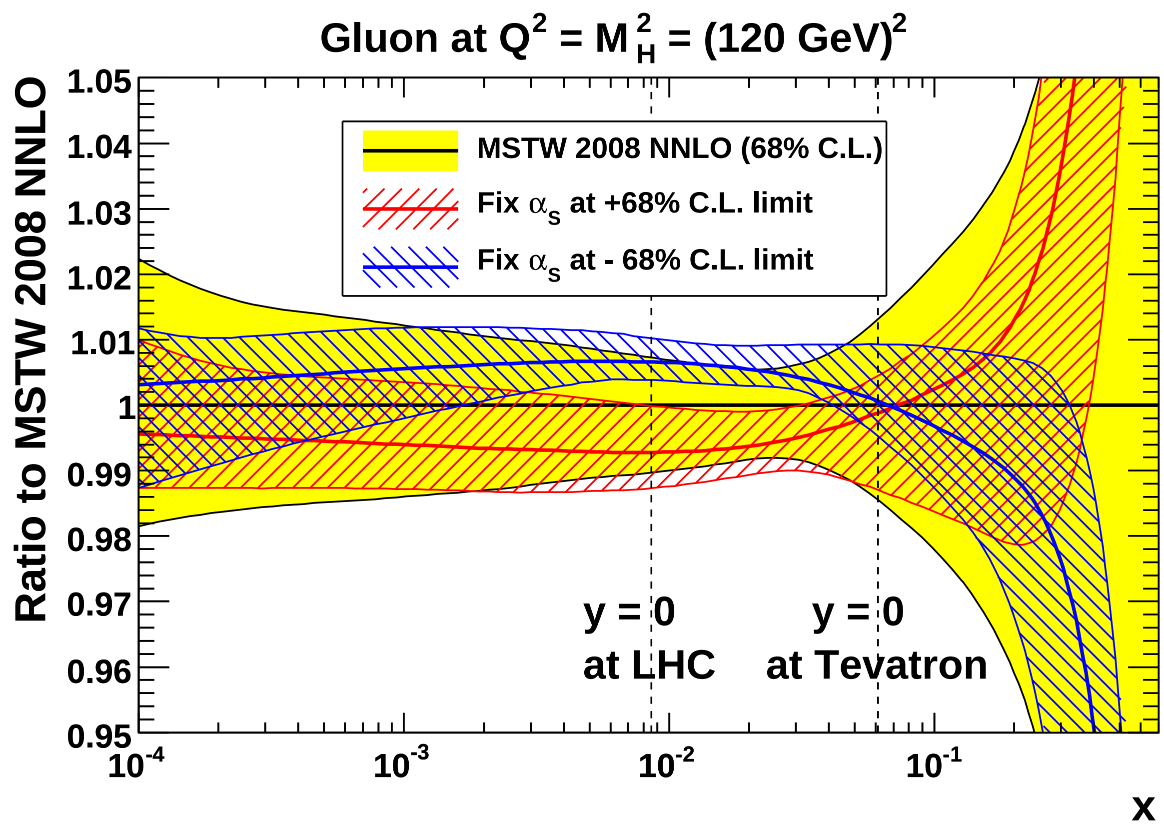
All tend towards positive momentum asymmetry, but all fairly consistent with zero, or with enough to remove (or seriously) reduce NuTeV anomaly on $\sin^2\theta_W$.

In fact NNPDF now smallest uncertainty on this by some way (no data above $x = 0.2$).

PDF correlation with α_S.

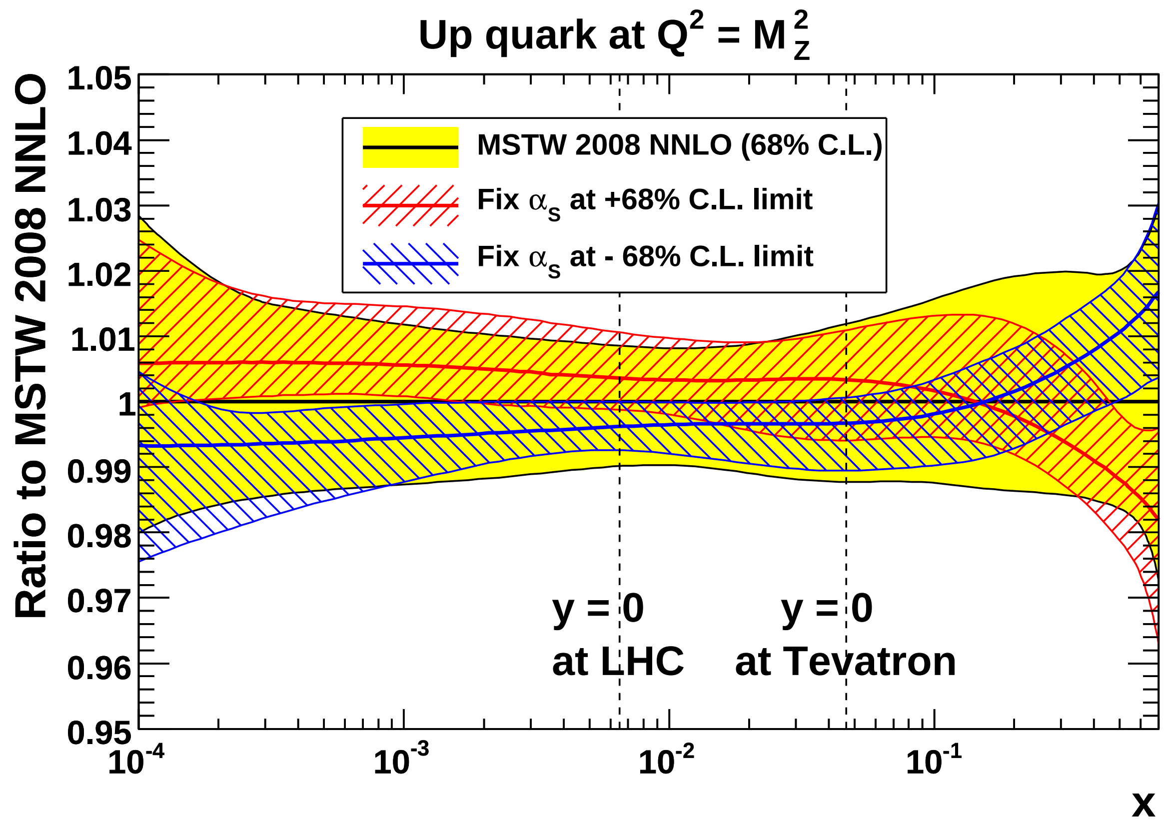
Can also look at PDF changes and uncertainties at different $\alpha_S(M_Z^2)$. Latter usually only for one fixed $\alpha_S(M_Z^2)$. Can be determined from fit, e.g. $\alpha_S(M_Z^2) = 0.1202^{+0.0012}_{-0.0015}$ at NLO and $\alpha_S(M_Z^2) = 0.1171^{+0.0014}_{-0.0014}$ at NNLO from MSTW.

PDF uncertainties reduced since quality of fit already worse than best fit.



Expected gluon–$\alpha_S(M_Z^2)$ small–x anti-correlation $\rightarrow$ high-x correlation from sum rule.

Gluon feeds into evolution of quarks, but change in $\alpha_S(M_Z^2)$ just outweighs gluon change, i.e. larger $\alpha_S(M_Z^2) \rightarrow$ slightly more evolution.

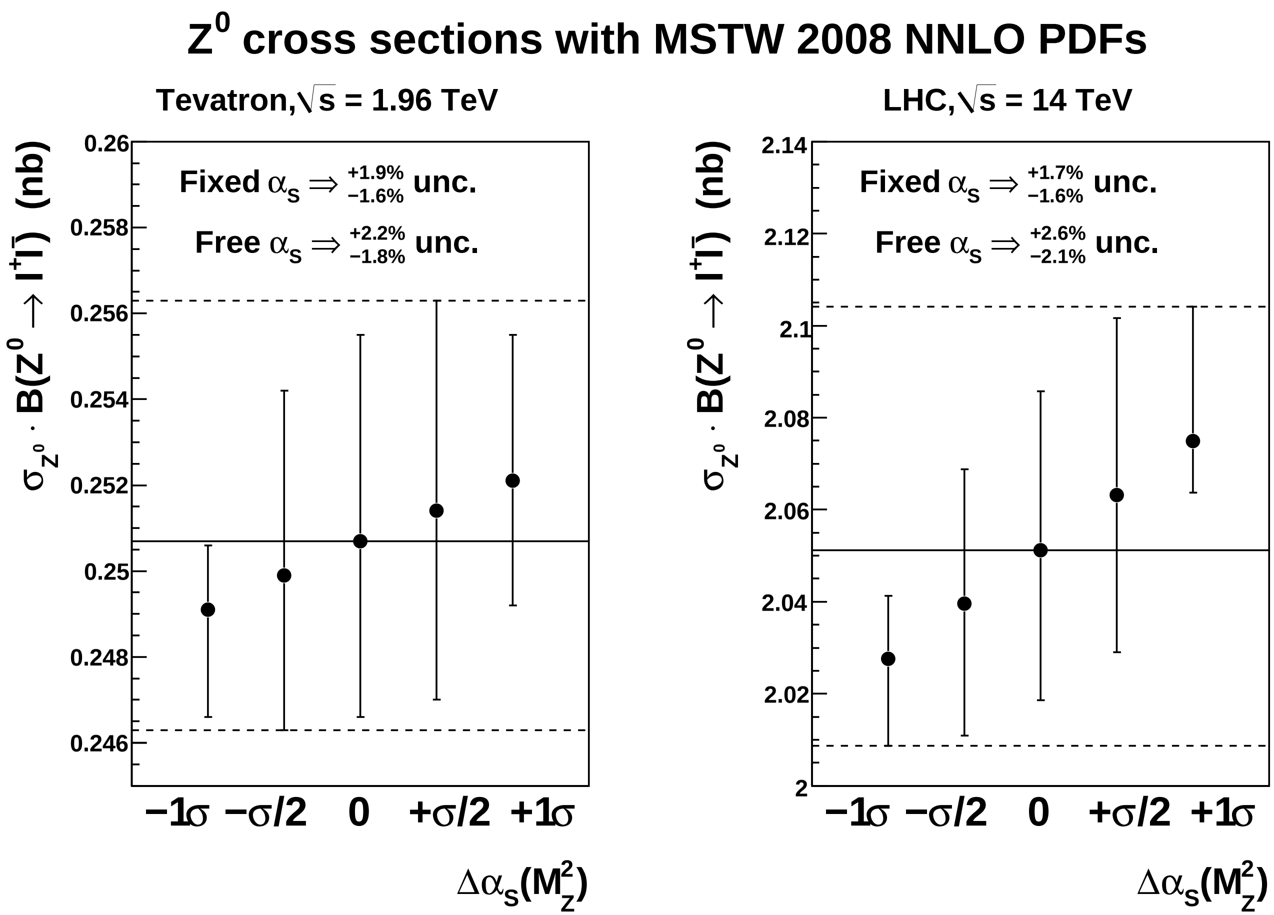


Strong anti-correlation at high-x due to evolution and positive coefficient functions.

Quarks roughly opposite to gluons.

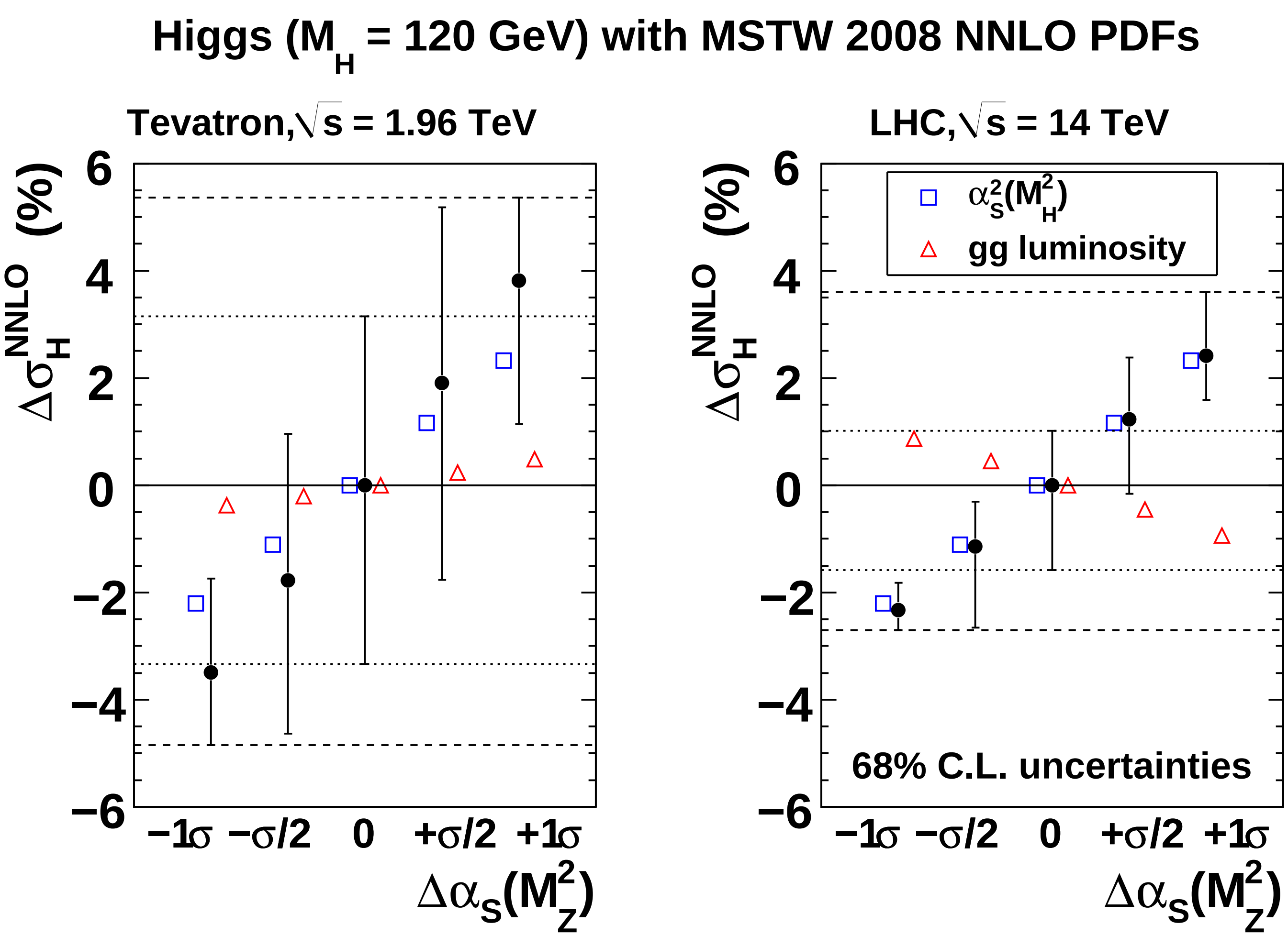
Additional uncertainty from $\alpha_S(M_Z^2)$ variation for quantities depends on how PDFs and coupling are correlated.

NNLO predictions for Z production for allowed $\alpha_S(M_Z^2)$ values and their uncertainties.



Total uncertainty envelope of set of uncertainties. Increases by up to 50% at LHC. Largely due to effect of PDFs.

NNLO predictions for Higgs ($120\mathrm{GeV}$) production for different allowed $\alpha_S(M_Z^2)$ values and their uncertainties.



Increases by a factor of $2-3$ (up more than down) at LHC. Direct $\alpha_S(M_Z^2)$ dependence mitigated somewhat by anti-correlated small-x gluon (asymmetry feature of *minor* problems in fit to HERA data). At Tevatron intrinsic gluon uncertainty dominates.

Other Sources of Uncertainty

It is vital to consider theoretical/assumption-dependent uncertainties:

- Methods of determining "best fit" and uncertainties.
- Underlying assumptions in procedure, e.g. parameterisations and data used.
- Treatment of heavy flavours.
- PDF and α_S correlations.

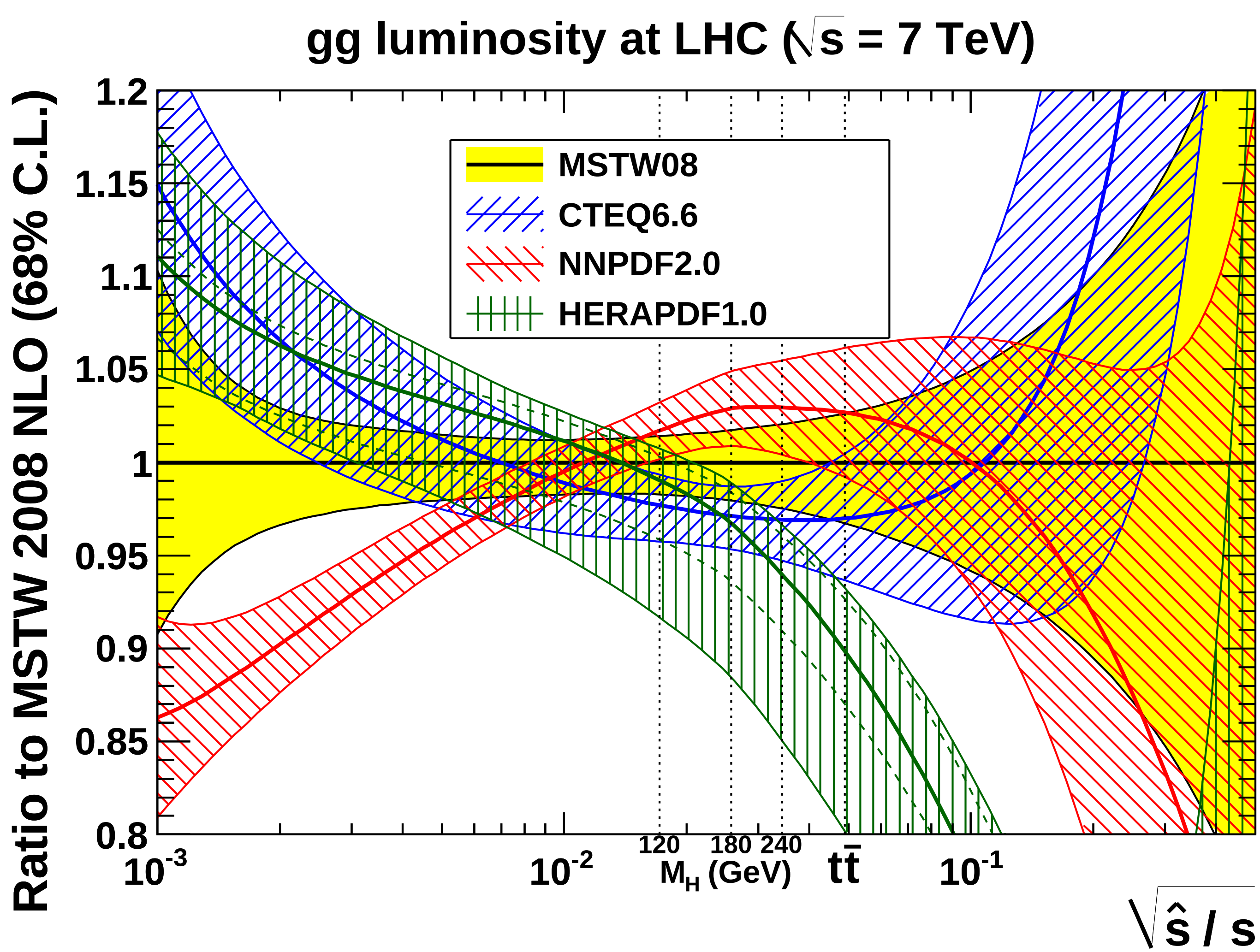
Responsible for differences between groups for extraction of fixed-order PDFs.

Also other sources which (mainly) lead to inaccuracies common to all fixed-order extractions.

- QED and Weak (comparable to NNLO ?) ($\alpha_s^3 \sim \alpha$). Sometime enhancements.
- Standard higher orders (NNLO – may sets available here.)
- Resummations, e.g. small x ($\alpha_s^n \ln^{n-1}(1/x)$), or large x ($\alpha_s^n \ln^{2n-1}(1-x)$)
- low Q^2 (higher twist), saturation

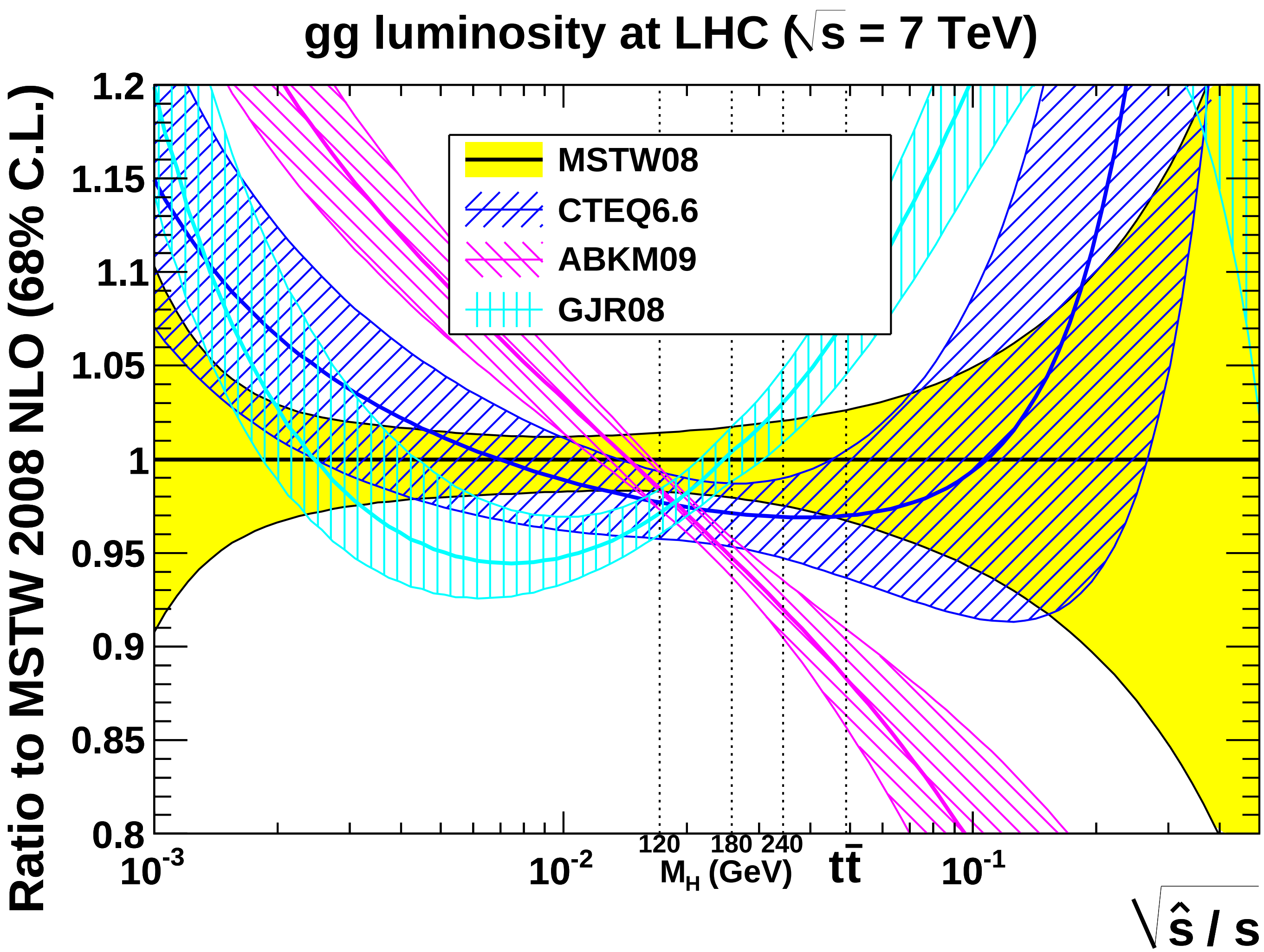
In fact probably does lead to some of the difference it PDFs observed.

Predictions by various groups - parton luminosities – NLO. Plots by G. Watt.

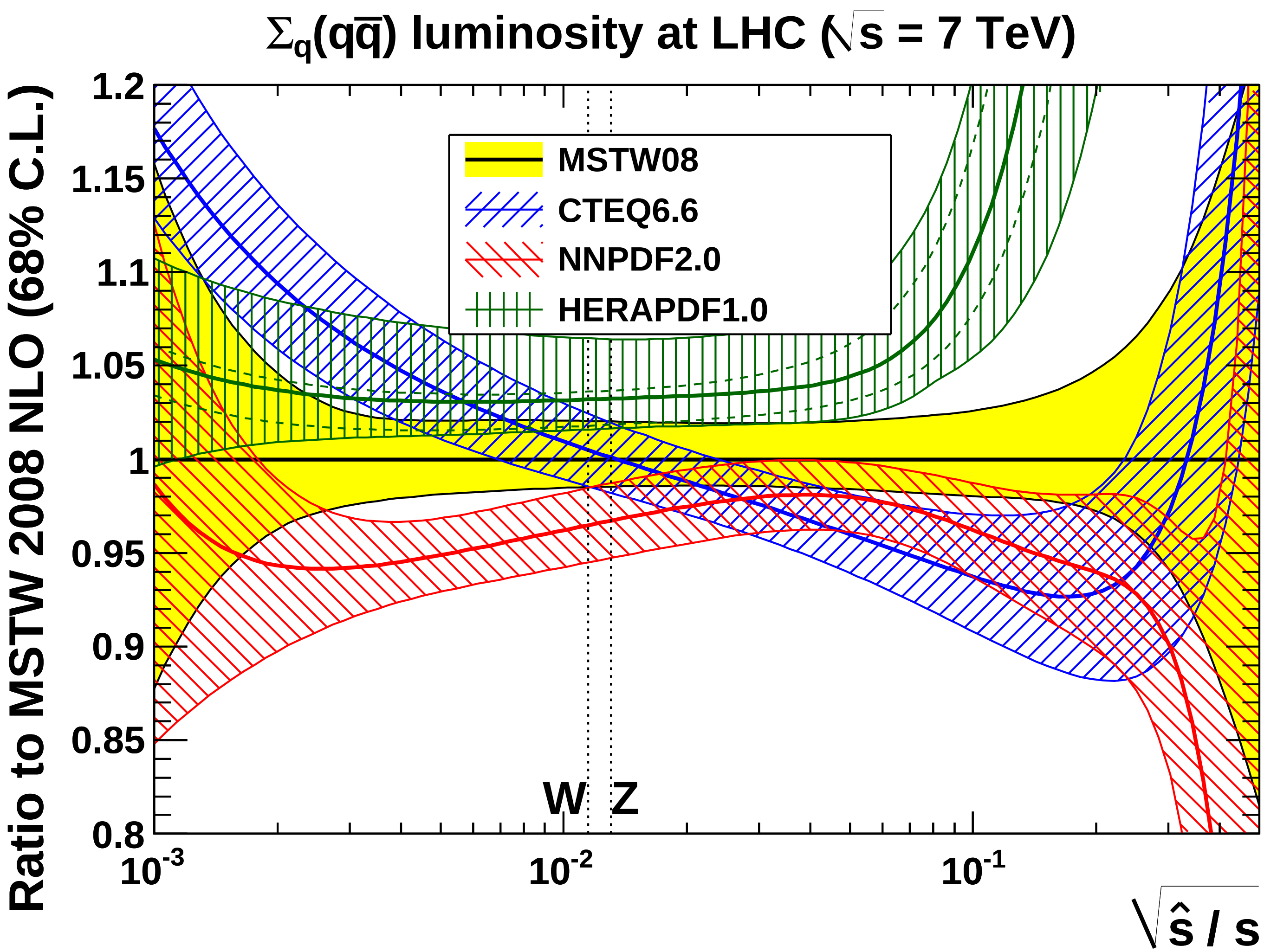


Cross-section for $t\bar{t}$ almost identical in PDF terms to $450\mathrm{GeV}$ Higgs.

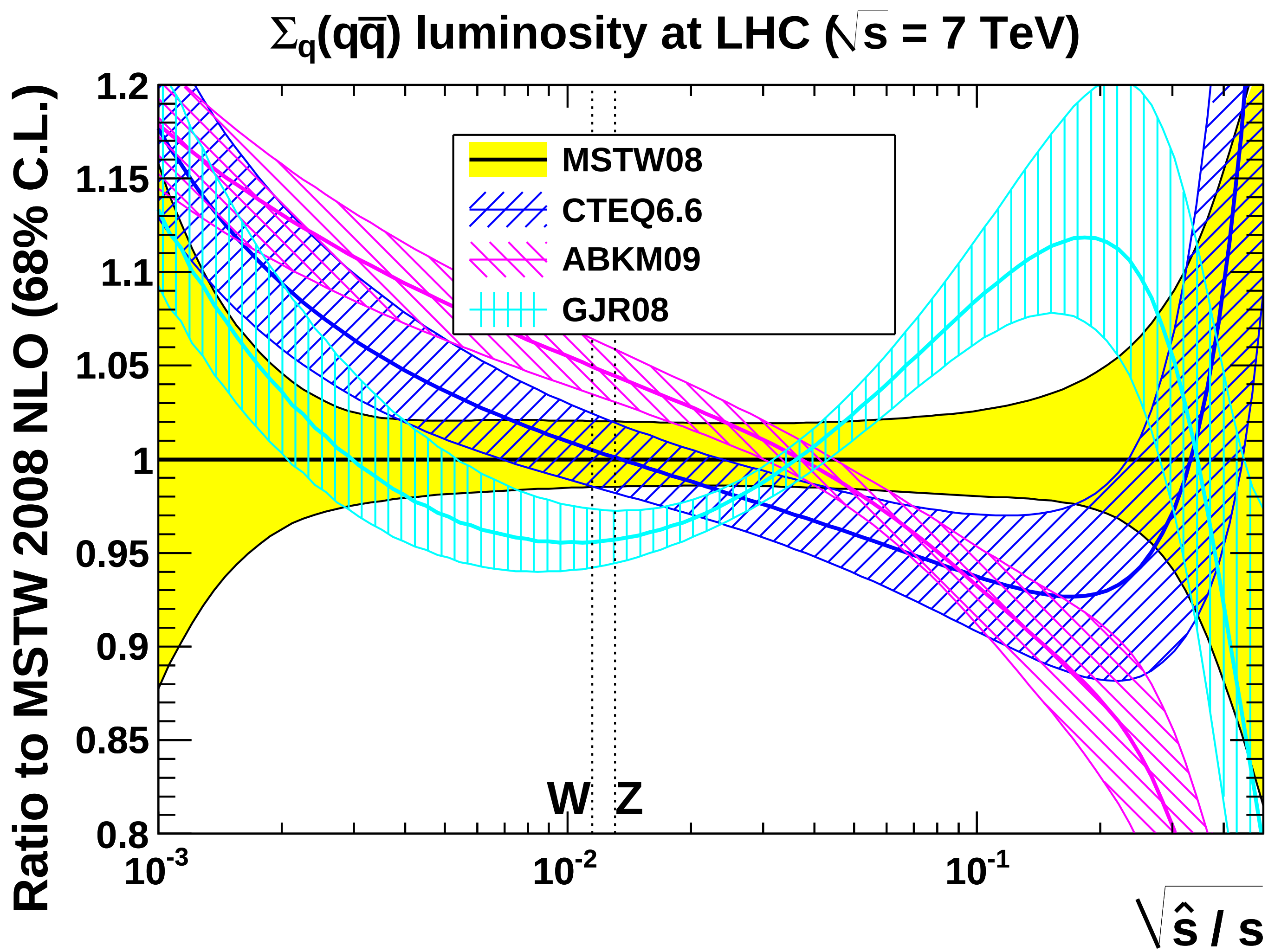
Also $H + t\bar{t}$ at $\sqrt{\hat{s}/s} \sim 0.1$.



Clearly some distinct variation between groups. Much can be understood in terms of previous differences in approaches.



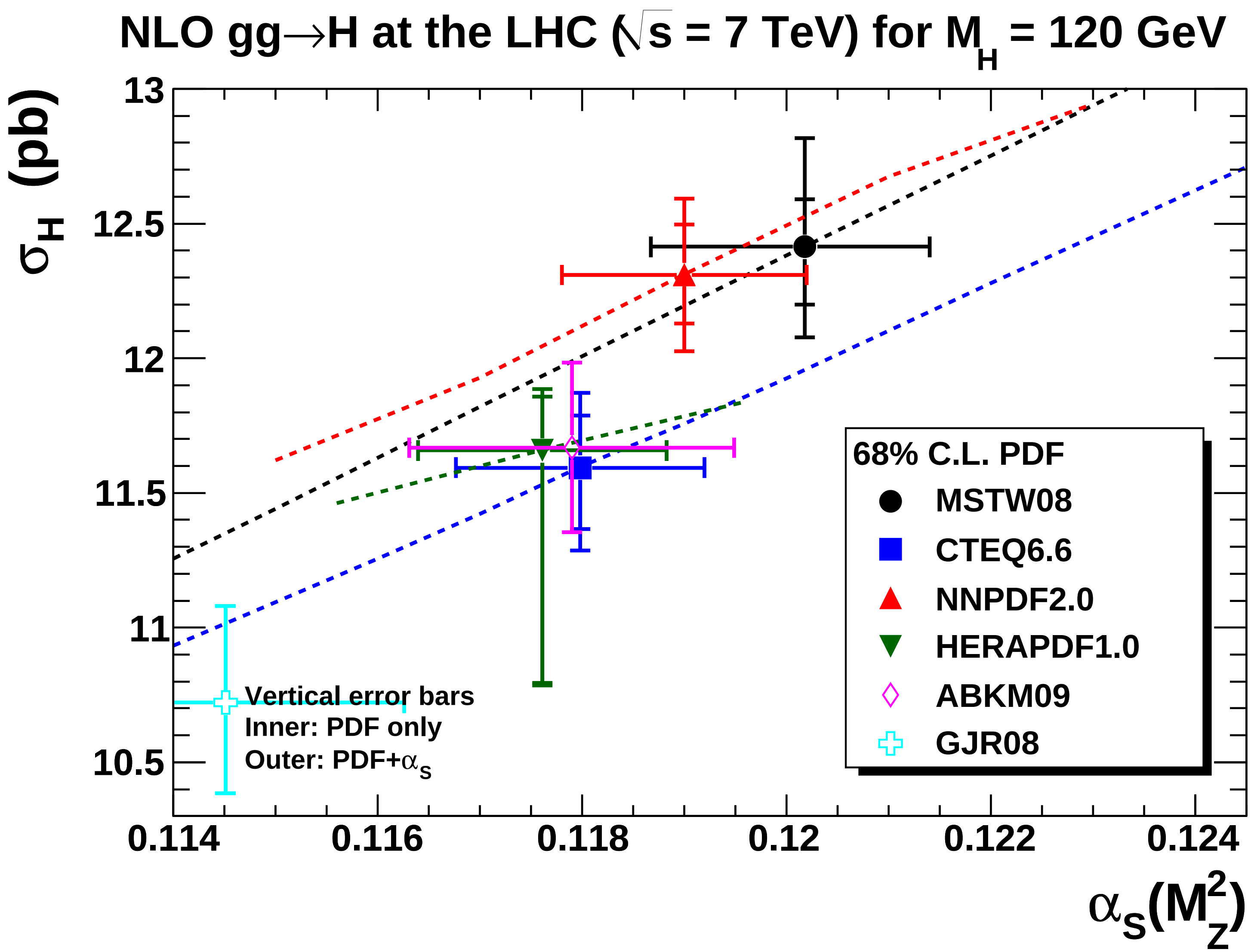
Many of the same general features for quark-antiquark luminosity. Some differences mainly at higher x.



Canonical example W, Z production, but higher $\hat{s}/s$ relevant for WH or vector boson fusion.

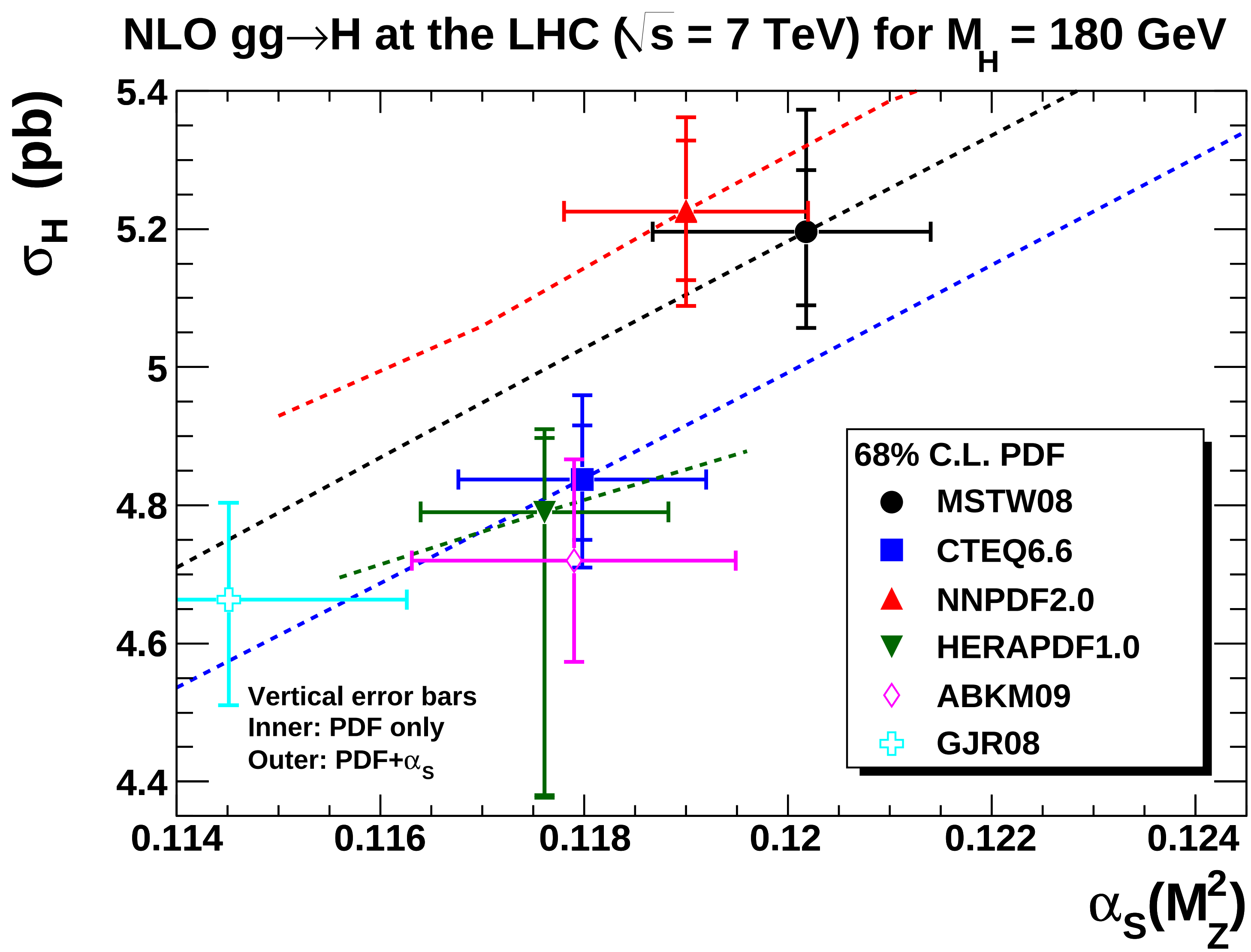
All plots and more at http://projects.hepforge.org/mstwpdf/pdf4lhc

Variations in Cross-Section Predictions – NLO

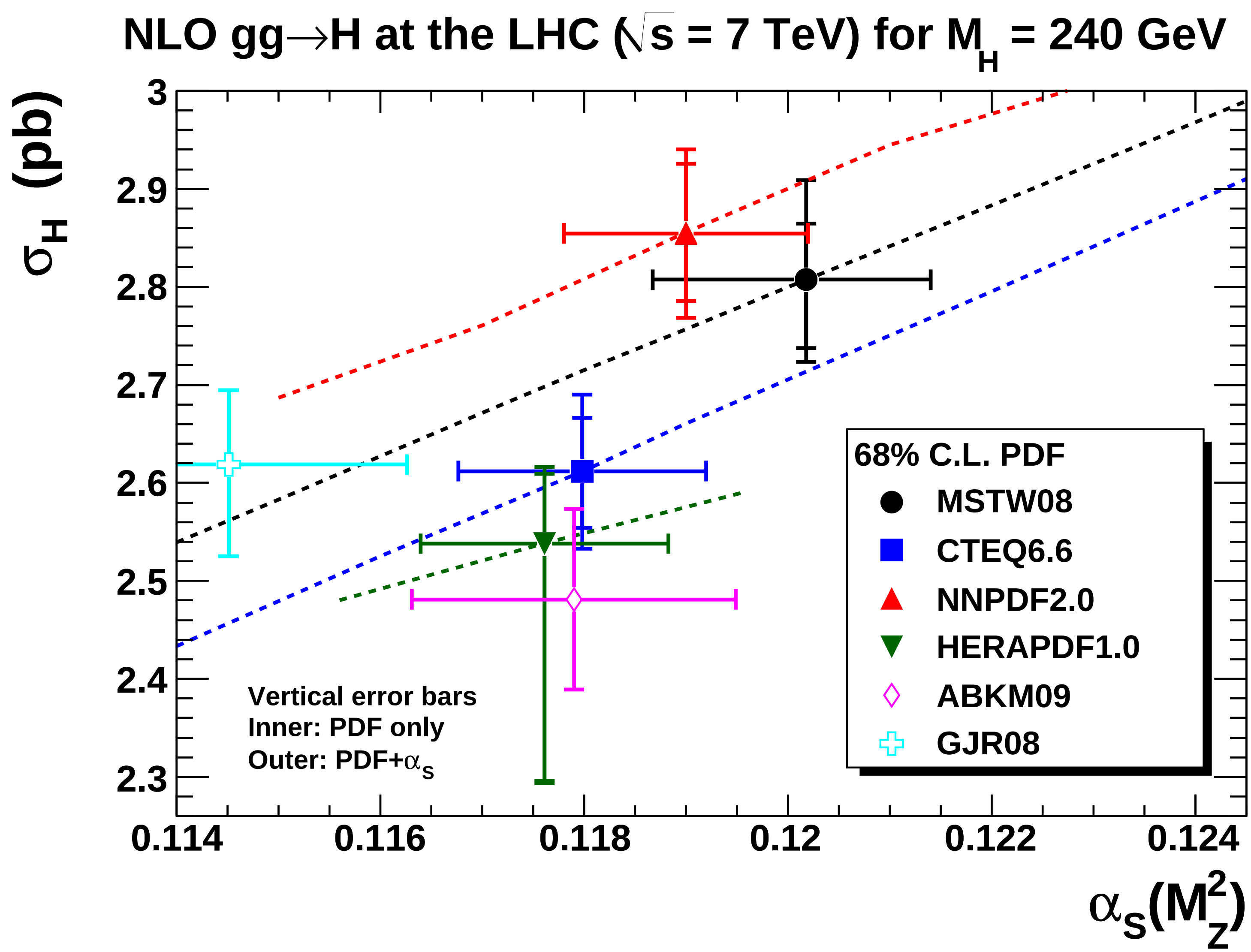


Dotted lines show how central PDF predictions vary with $\alpha_S(M_Z^2)$.

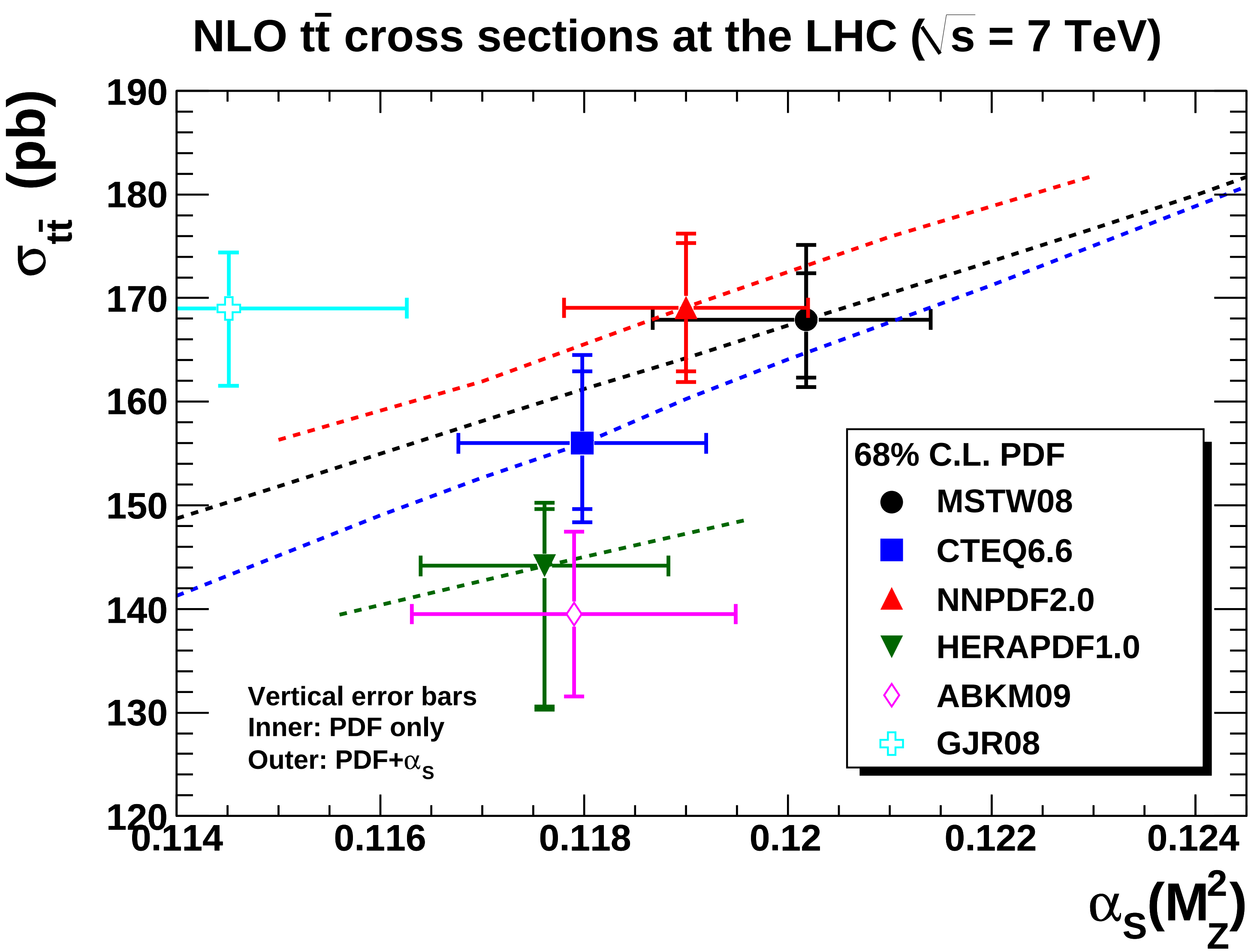
Again plots by G Watt using PDF4LHC benchmark criteria.



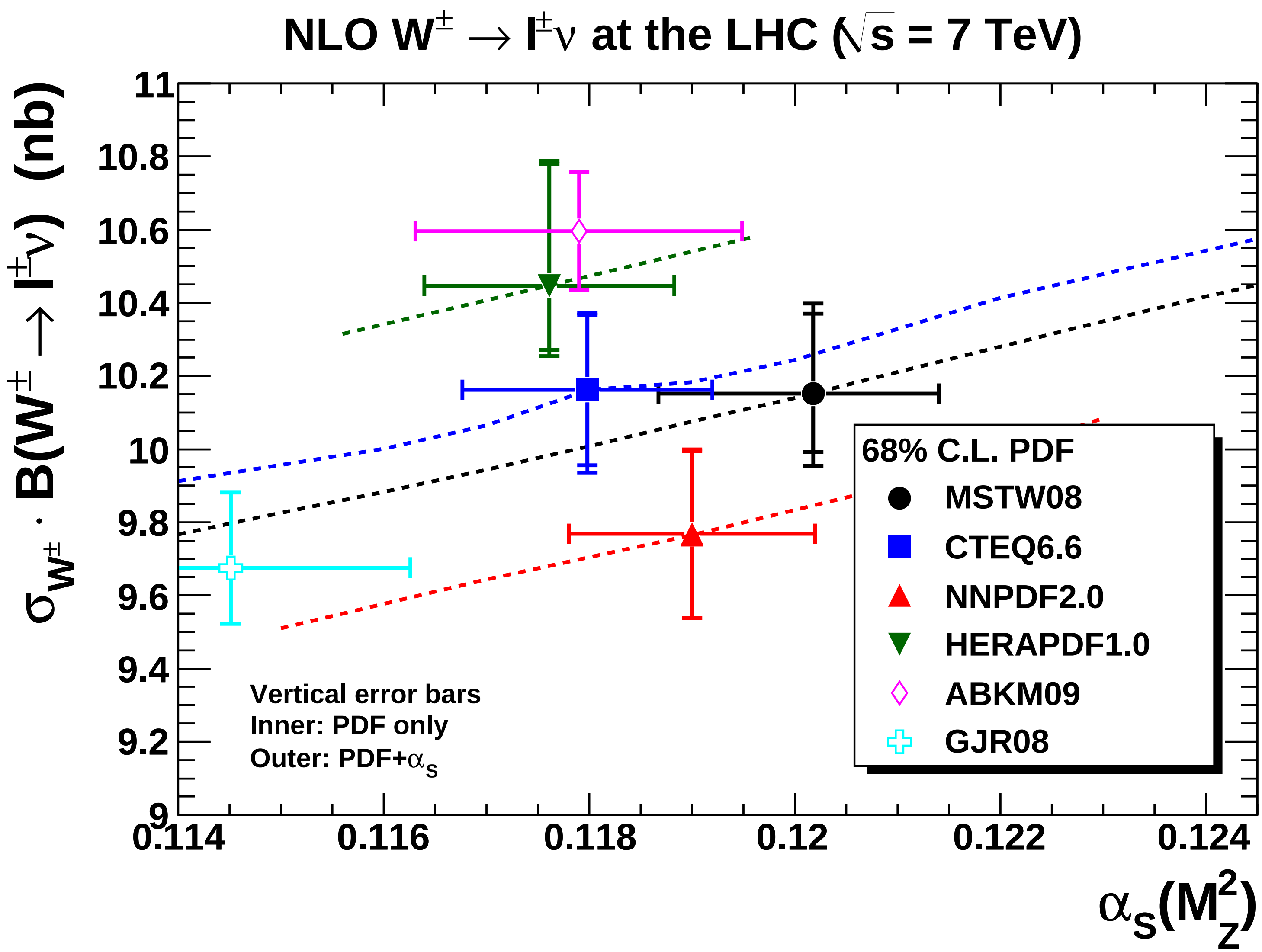
Clearly much more variation in predictions than uncertainties claimed by individual groups.



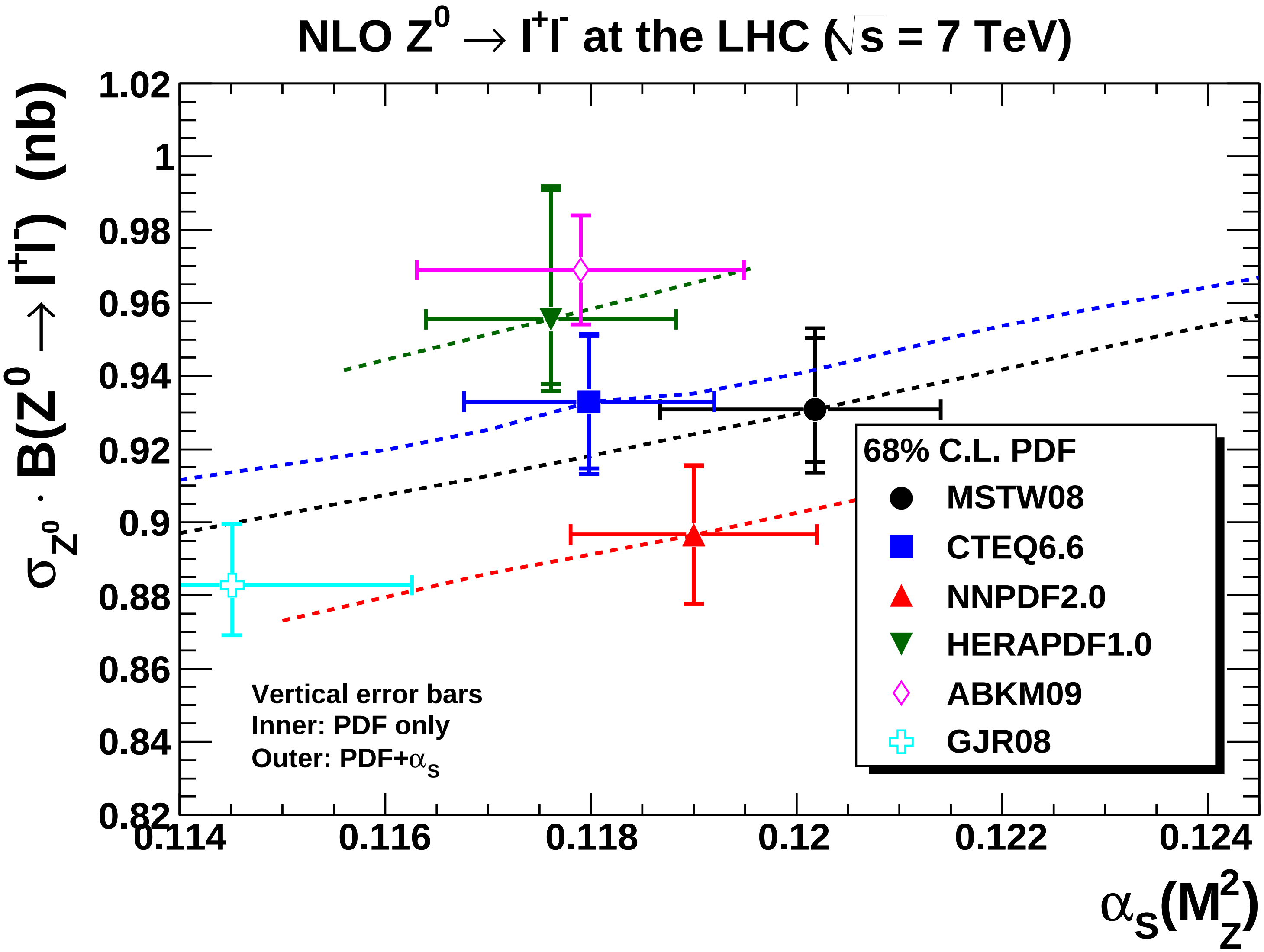
Excluding GJR08 amount of difference due to $\alpha_S(M_Z^2)$ variations $3-4\%$.



CTEQ6.6 now heading back towards MSTW08 and NNPDF2.0.

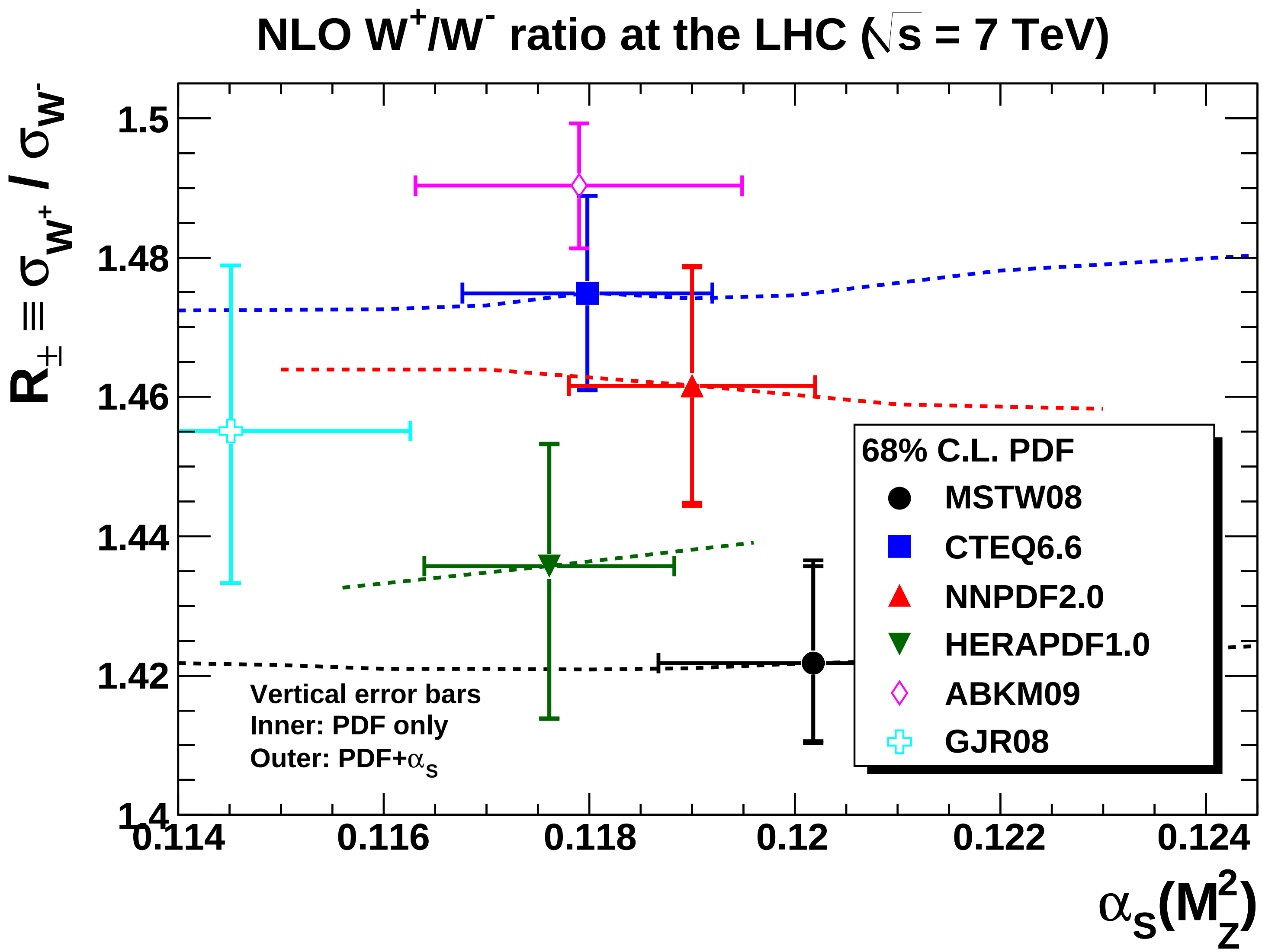


$W^+ + W^-$ cross-section. $\alpha_S(M_Z^2)$ dependence now more due to PDF variation with $\alpha_S(M_Z^2)$.



Again variations somewhat bigger than individual uncertainties.

Roughly similar variation for $\hat{s}$ up to a few times higher.



Quite a variation in ratio. Shows variations in flavour and quark-antiquark decompositions.

All plots and more at http://projects.hepforge.org/mstwpdf/pdf4lhc

Deviations In predictions clearly much more than uncertainty claimed by each.

In some cases clear reason why central values differ, e.g. lack of some constraining data, though uncertainties then do not reflect true uncertainty.

Sometimes no good understanding, or due to difference in procedure which is simply a matter of disagreement, e.g. gluon parameterisation at small x affects predicted Higgs cross-section.

What is true uncertainty. Task asked of PDF4LHC group.

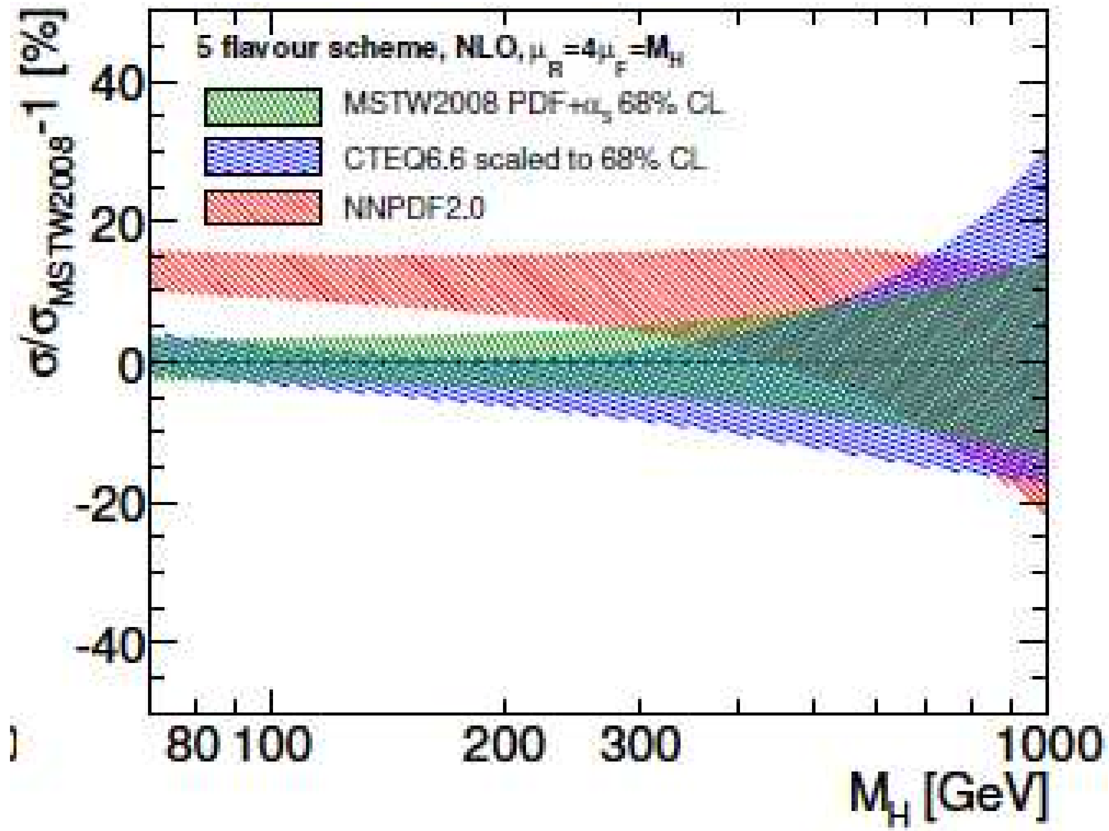
Interim recommendation take envelope of $global$ sets, MSTW, CTEQ NNPDF (check other sets) and take central point as uncertainty.

Not very satisfactory, but not clear what would be an improvement, especially as a general rule.

Usually not a big disagreement, and factor of about 2 expansion of MSTW uncertainty.

Sometimes rather worse than this for special case, e.g. Warsinsky at recent Higgs-LHC working group meeting.

m_b values bring CTEQ and MSTW together but exaggerate NNPDF difference.



Conclusions

One can determine the parton distributions and predict cross-sections at the LHC, and the fit quality using NLO or NNLO QCD is fairly good.

Various ways of looking at uncertainties due to errors on data. All give roughly the same value for uncertainties on PDFs and predictions – $\sim 1-5\%$ for most LHC quantities.

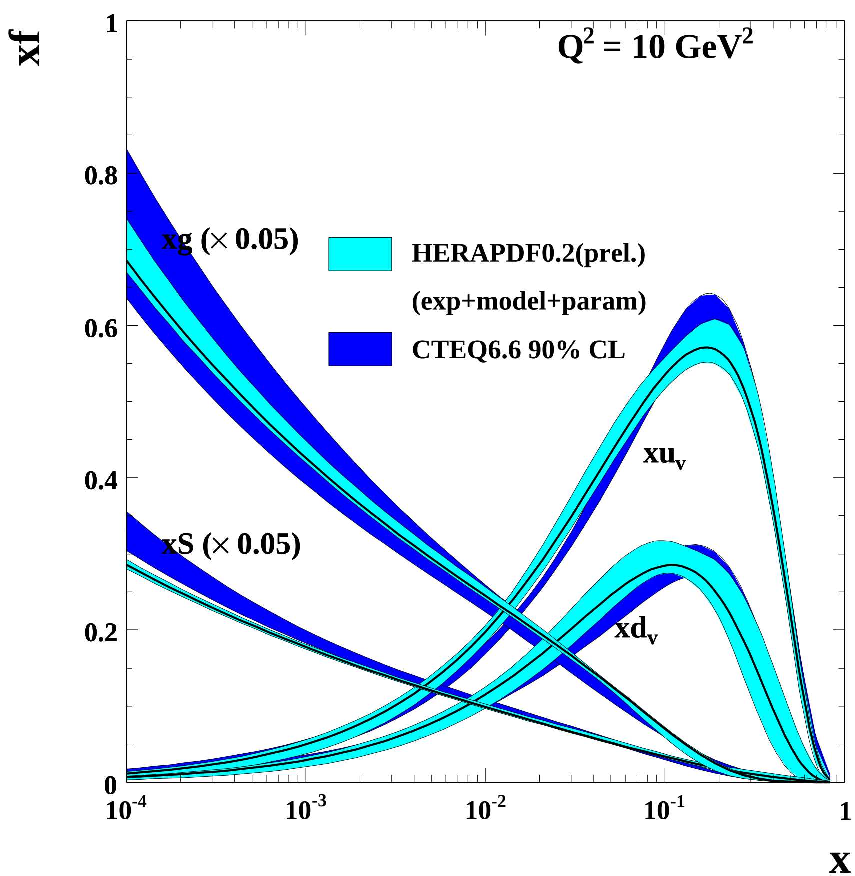
All should be, if anything, an overestimate, i.e. inflated tolerance, or missing data sets which would have an effect, and uncertainty undoubtedly related to the choice of stopping in NNPDF.

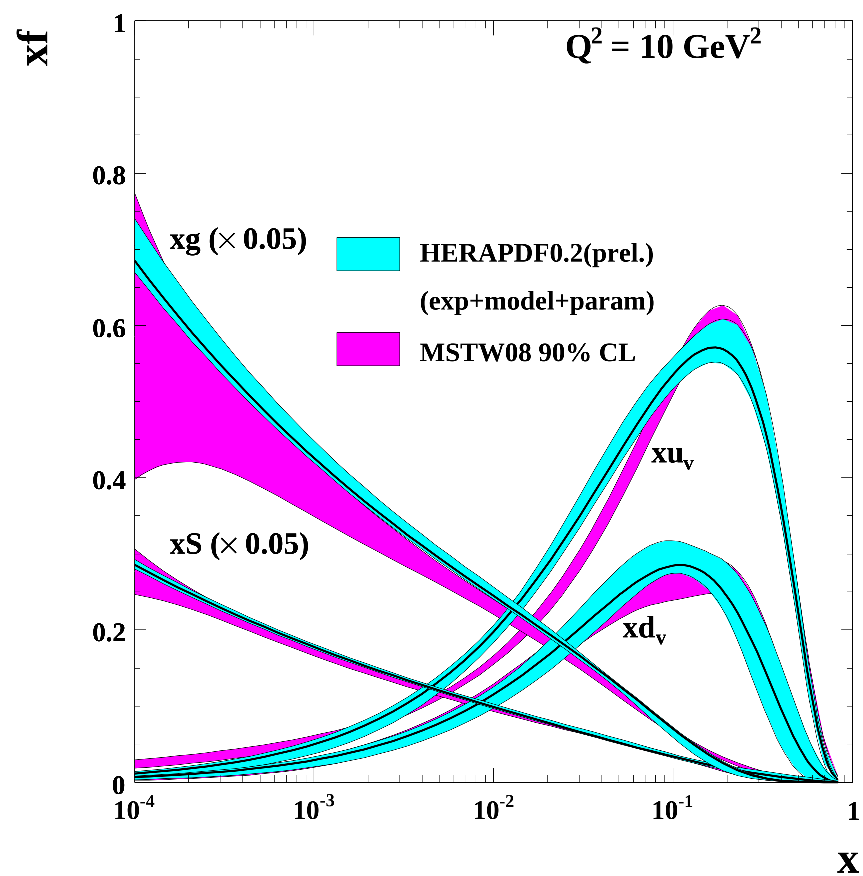
Effects from input assumptions e.g. selection of data fitted, cuts and input parameterisation can shift central values of predictions significantly. Different groups do not always agree very well despite "generous" uncertainties.

Some improvements if effects of heavy flavour treatments and α_S accounted for. α_S and PDFs correlated. Now being dealt with properly for in general. Reduces but does not remove discrepancies. Errors from higher orders/resummation potentially large. Imperfect theory used to fit data.

Extraction of PDFs from existing data and use for LHC far from a straightforward procedure. Lots of theoretical issues to consider for real precision. Relatively few cases where Standard Model discrepancies will not require some significant input from PDF physics to determine real significance.

HERA fits are only to HERA data, but averaged H1/ZEUS data (reduced correlated errors) not yet used by others so far (in $NNPDF2.0$ - effect slightly unclear).





More consistent data sets $\rightarrow$ $\Delta\chi^2 = 1$ for uncertainties. Nevertheless similar to CTEQ/MSTW.

Significant differences in central values sometimes, and in shape of small-x gluon uncertainty.

Parameterisations

MSTW predictions for W+ and W- cross-sections for LHC with common fixed order QCD and vector boson width effects, and common branching ratios.

Quoted uncertainty for ratio very small, i.e. $\approx 0.8\%$. Prediction sensitive to u and d quarks.

$$\frac{\sigma(W^+)}{\sigma(W^-)} \approx \frac{u(x)\bar{d}(x)}{d(x)\bar{u}(x)} \approx \frac{u(x)}{d(x)},$$

If $\bar{u}(x) \to \bar{d}(x), x \to 0$, which data implies, and most parameterisations assume.

Fit includes most recent neutrino DIS and Tevatron vector boson data. Uncertainties should account for this.

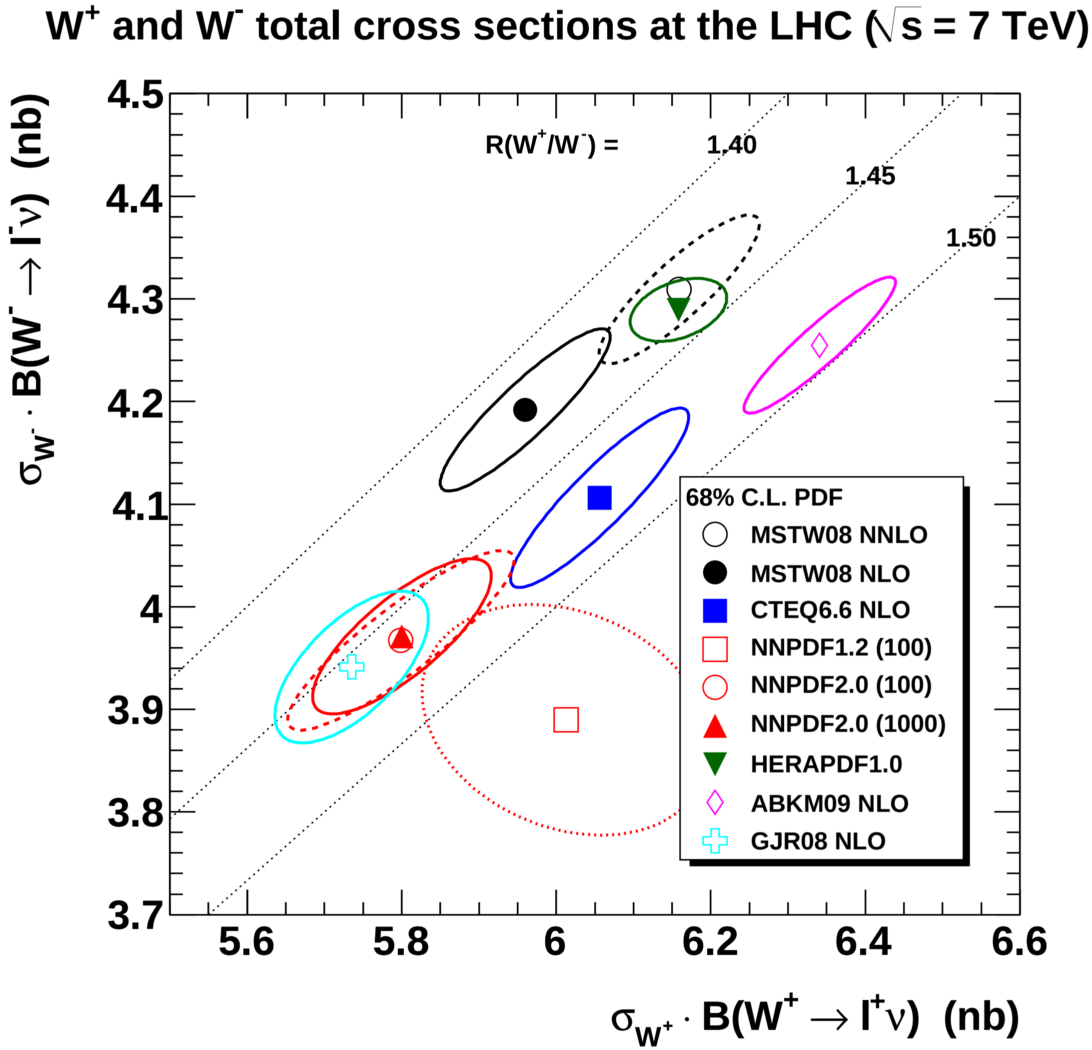
Significantly more difference than uncertainty from other PDFs, including MRST – (effect noted for W-asymmetry by Cooper-Sarkar). Very interesting for early data.

Again comparing comparing more groups even get even more discrepancies between them.

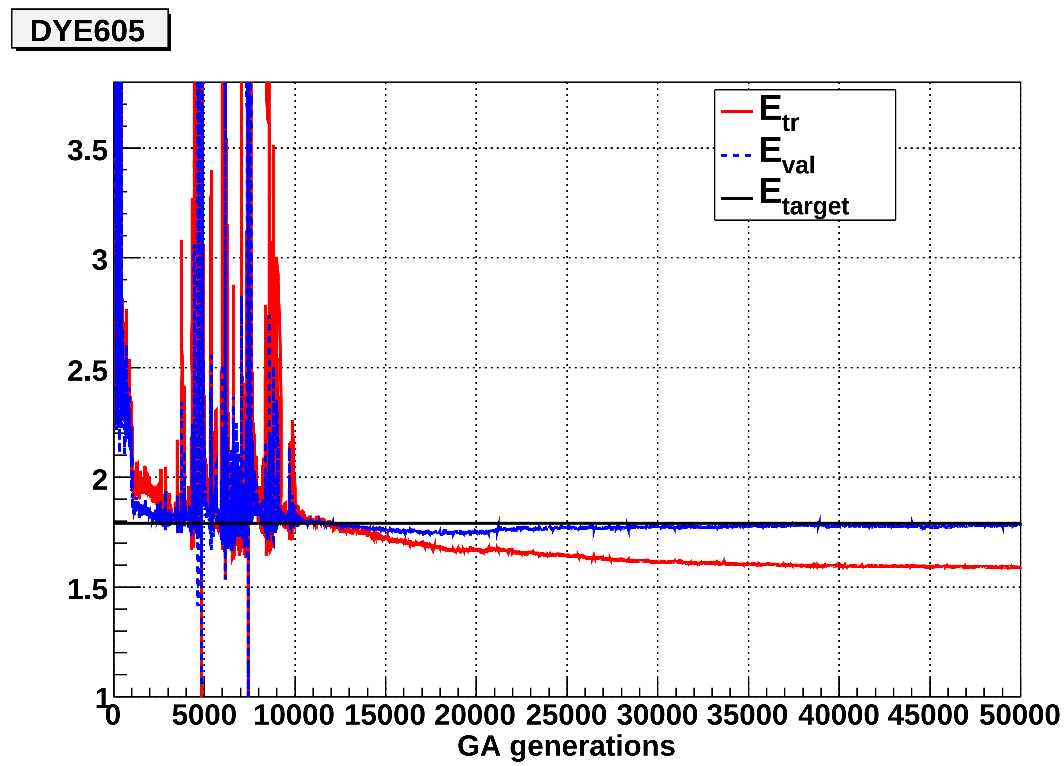
NLO $\rightarrow$ NNLO hardly affects ratio.

Some of the differences not well understood.

NNPDF band shrinks dramatically with new data.



Difficult to know when fit to validation set has started increasing significantly for some sets.



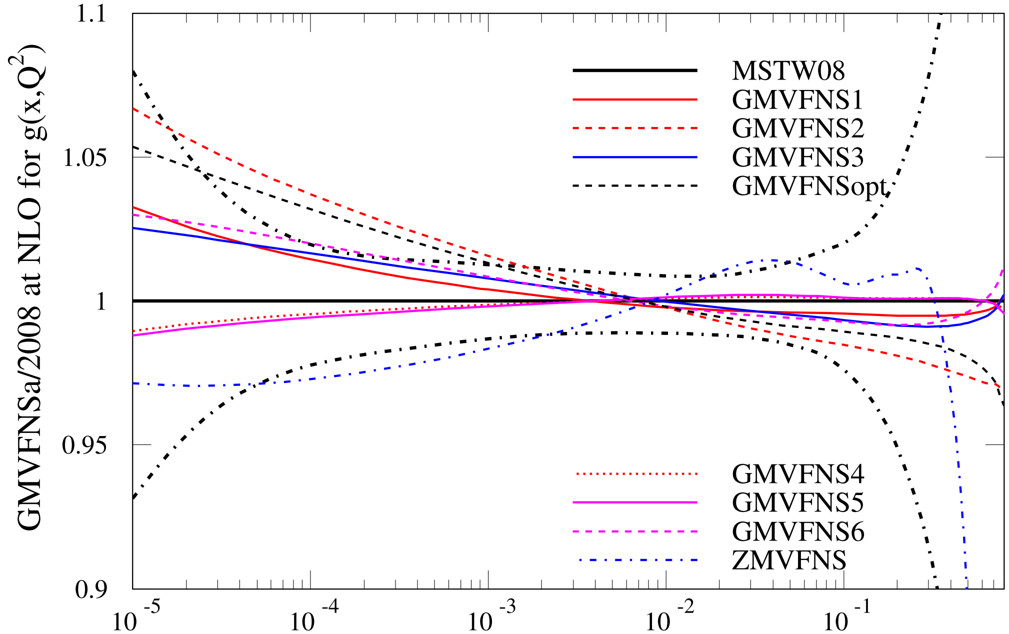
Variations in partons extracted from global fit due to different choices of GM-VFNS at NLO.

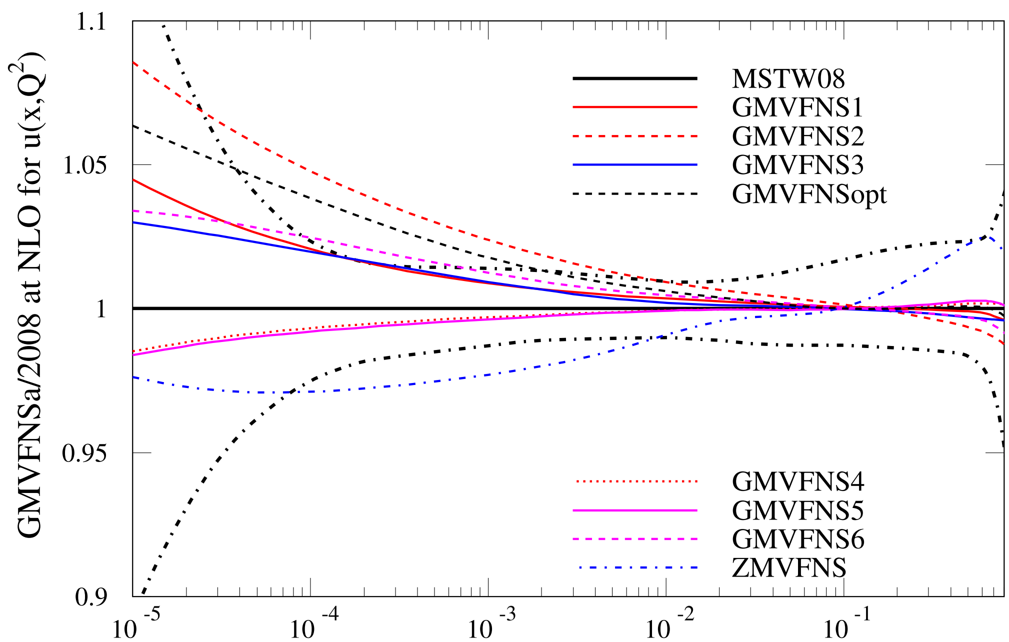
Initial χ^2 can change by 250.

Converges to at most about 15 of original.

Better fit for GMVFNS1, GMVFNS3 and GMVFNS6.

Some changes in PDFs large compared to one-sigma *uncertainty*.





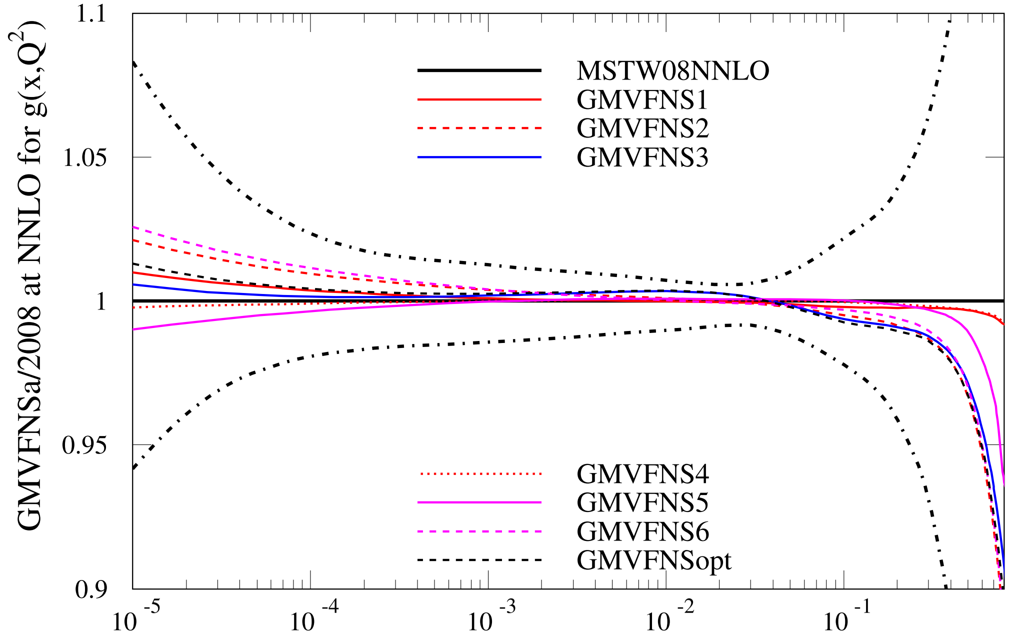
Variations in partons extracted from global fit due to different choices of GM-VFNS at NNLO.

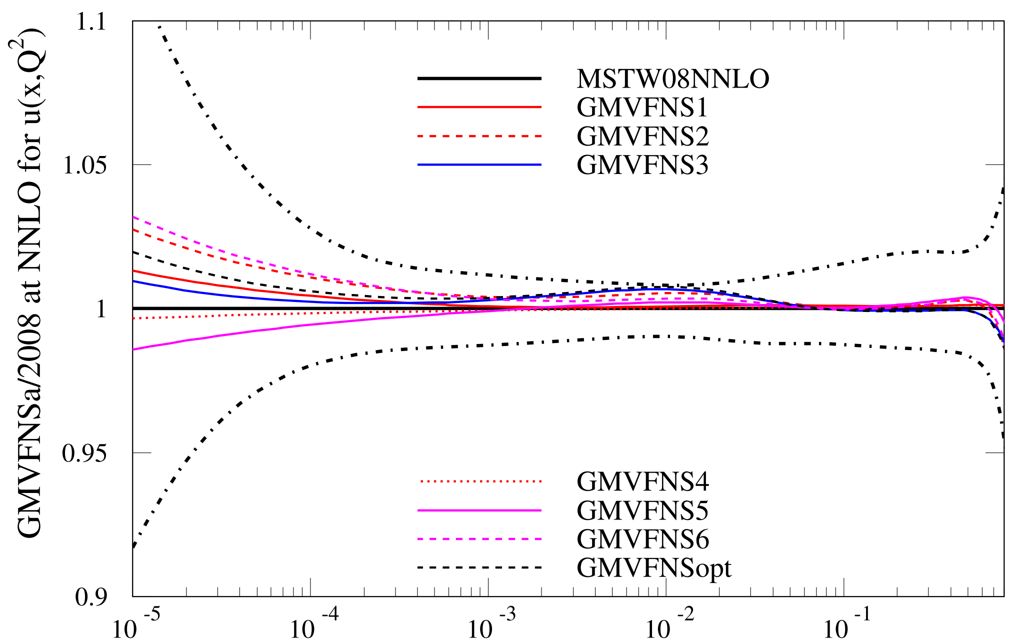
Initial changes in $\chi^2 < 20$.

Converge to about 10. None a marked improvement.

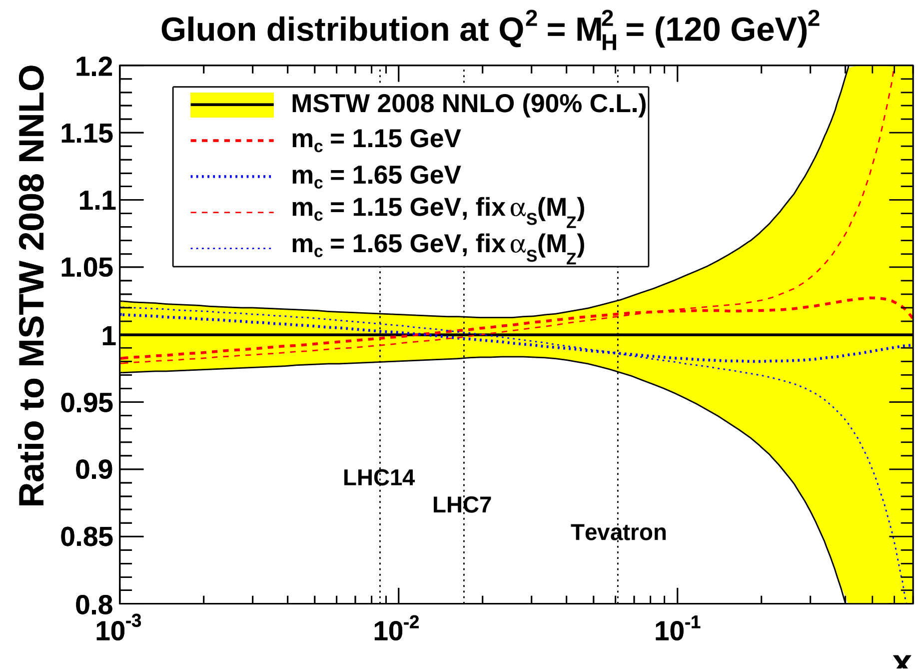
At worst changes approach $uncertainty$.

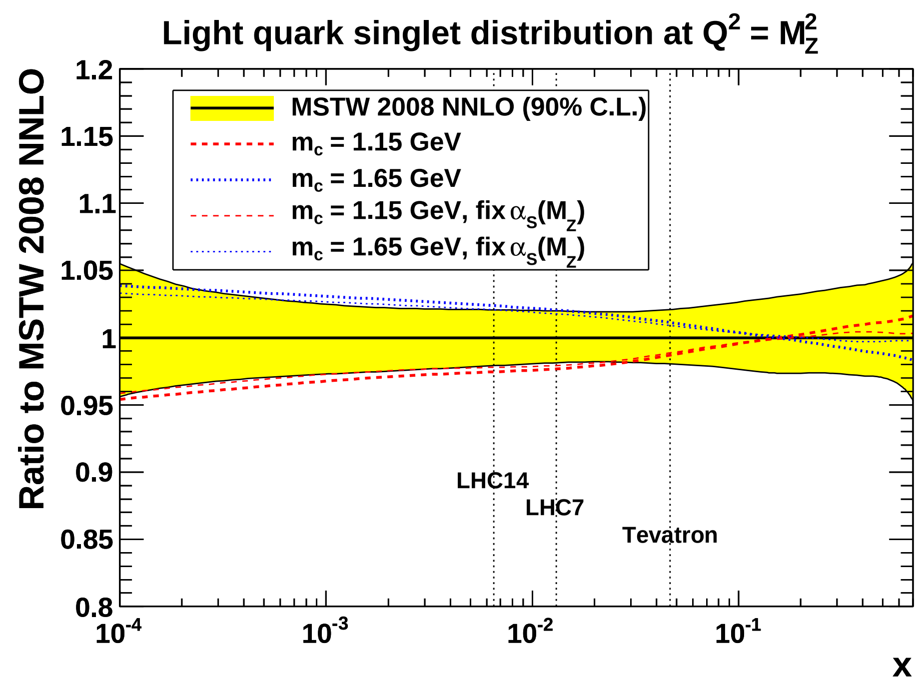
Biggest variation in high-x gluon, which has large uncertainty.



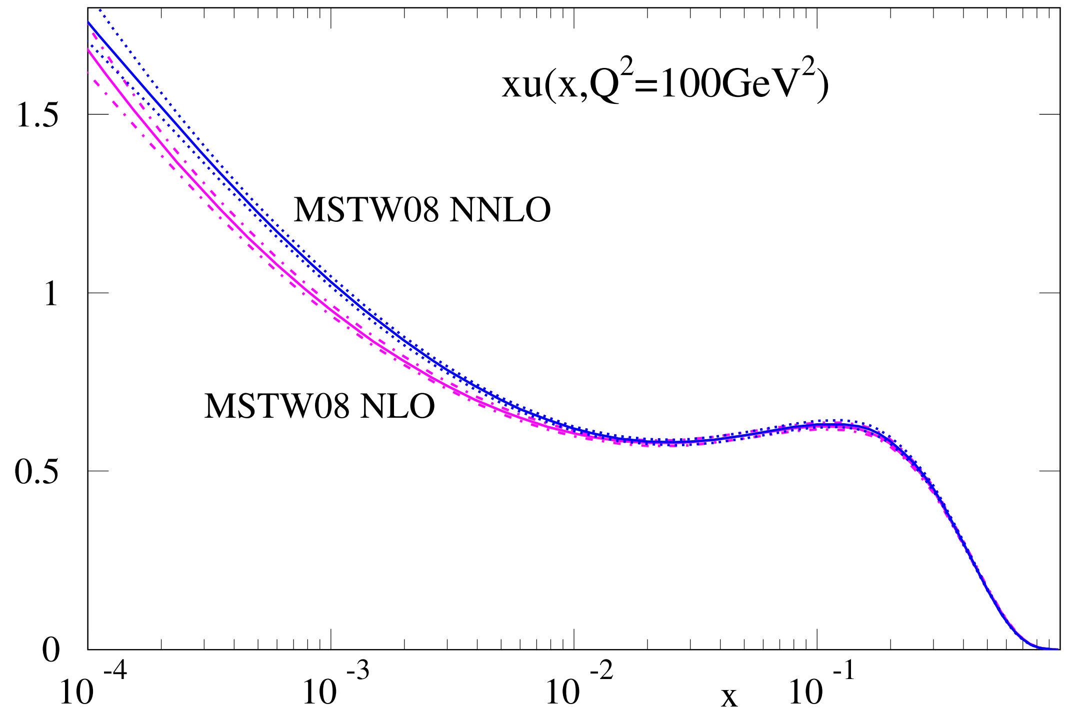


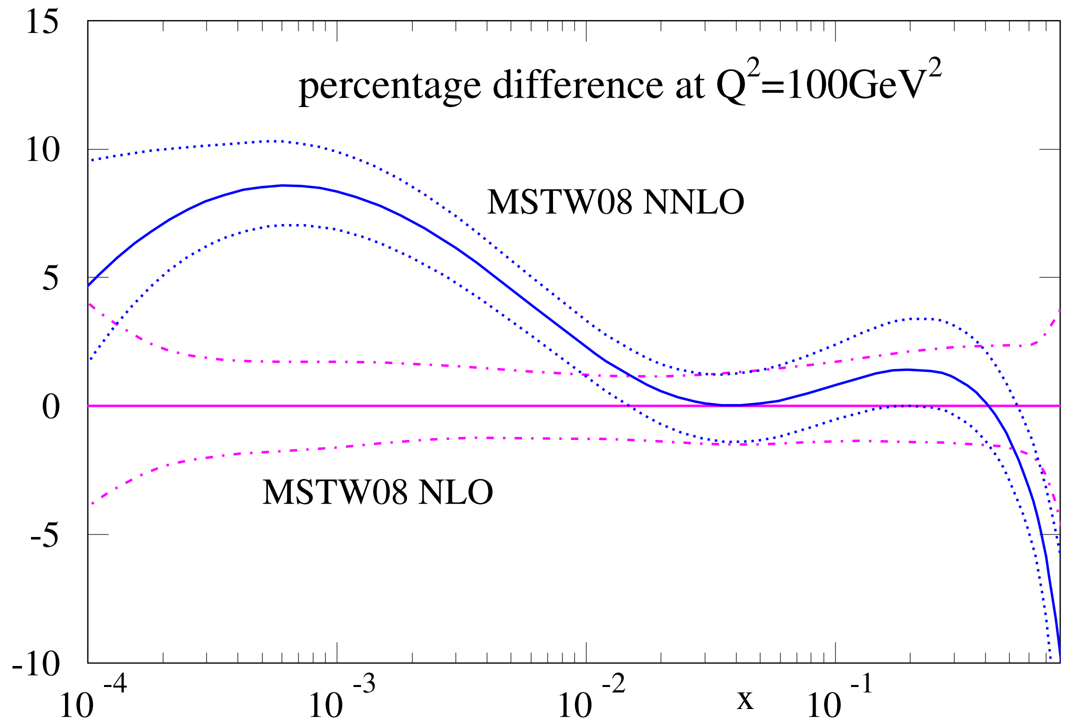
Ratio of partons when m_c is varied either with or without varying α_S





Systematic difference between PDF defined at NLO and at NNLO.





Parameterisation dependence reason for inflated $\Delta\chi^2 = 100$ Tolerance?

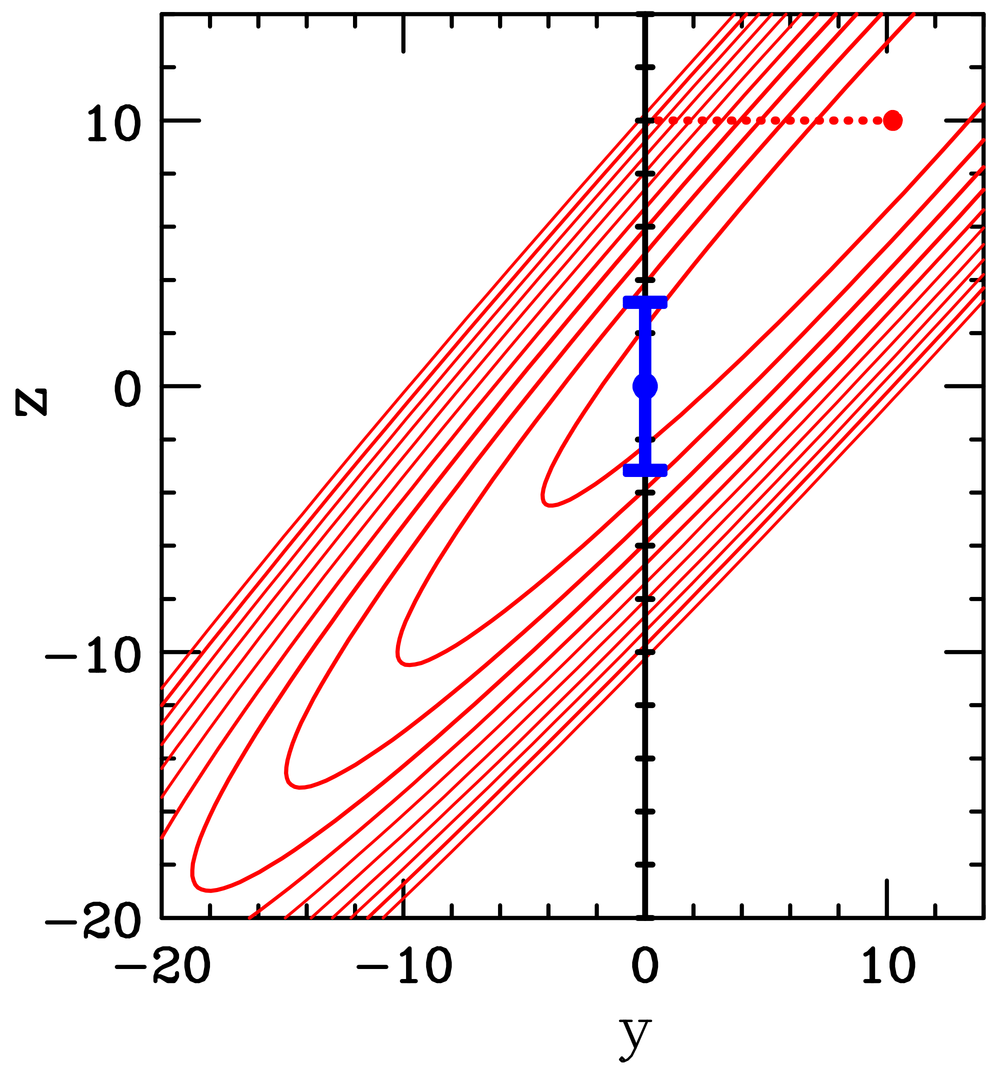
Proposal by Pumplin that this may be the case.

Simple model, fit depends on one parameter, min at $z = 0$ and $\chi^2 = z^2$.

Add second parameter y, could get χ^2 profile as shown.

For long narrow ellipse can get shift in best fit z such that value corresponds to χ^2 in original model with magnitude much greater than improvement in best fit quality.

$\propto (1/R - R)^2$, where R is ratio of minor to major axis.



Why this doesn't apply to global fits

1. In MSTW/MRST and CTEQ global fits there are more free parameters in obtaining best fit minimum than in determining eigenvectors. Even if correct argument, doesn't directly apply since main effect – in change of minimum – already accounted for.

2. This very elliptical profile only occurs if two of the parameters are very correlated. This is in fact why we do not leave all our parameters free in eigenvectors. Along major axis very flat direction always suddenly turns up due to quartic and higher terms in χ^2 distribution. Two parameters compensate almost exactly near minimum, then compensation suddenly breaks. Argument based on quadratic terms breaks down.

3. If z and y highly correlated a large change in z is likely not a large change in a PDF distribution (explaining small improvement is χ^2).

4. If a new parameter is introduced which is not highly correlated with one already there the R is not small and change in old parameters in new best fit is commensurate with the improvement in χ^2

Basic arguments seem to be validated by a variety of checks.

Parameterisation used in MSTW fits. Only those 20 in red appear in eigenvectors.

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

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

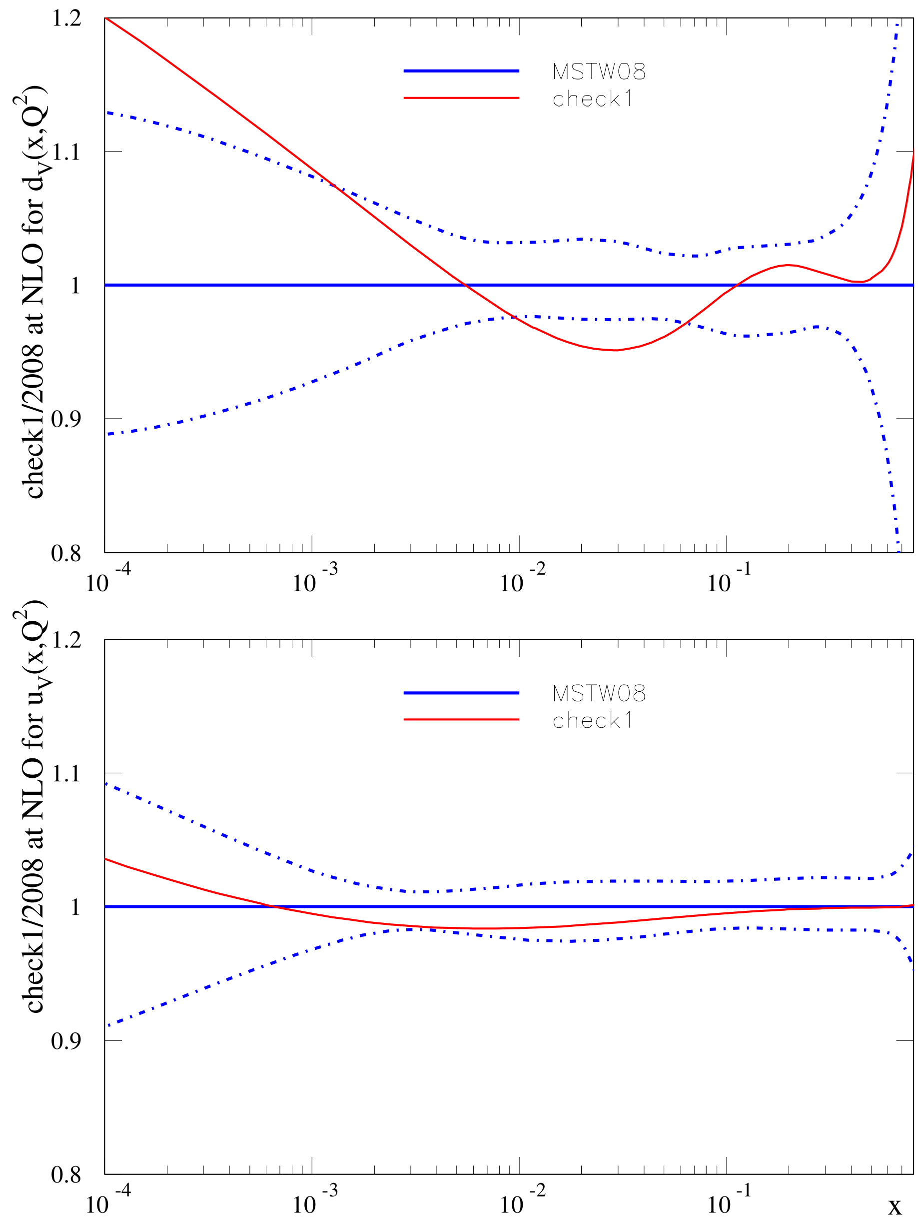
Of others only A_v, A_d, A_g and x_0 fixed by sum rules and $\delta_{s\_}$ fixed due to total correlation.

Try checking by setting all the parameters not in eigenvectors equal to zero, or if this is not sensible a round value. Fit quality awful. Need these parameters in global fit.

Move each by about 10% and refit. Like having y away from best fit. Refit 8 worse. Some parameters left free move by $2-3$ times quoted uncertainty.

Main change in PDFs in valence quarks shown. Worst change 1.8 bigger than uncertainty in d_V. Size of change in PDF not well correlated to relative size of change in parameters.

Trace to eigenvectors 14 and 18 in direction such that $\Delta\chi^2 \sim 10$ for uncertainty. – change in PDF twice that expected from change in χ^2 for global fit.



Try removing parameter which is not highly correlated, i.e. one in eigenvector definition.

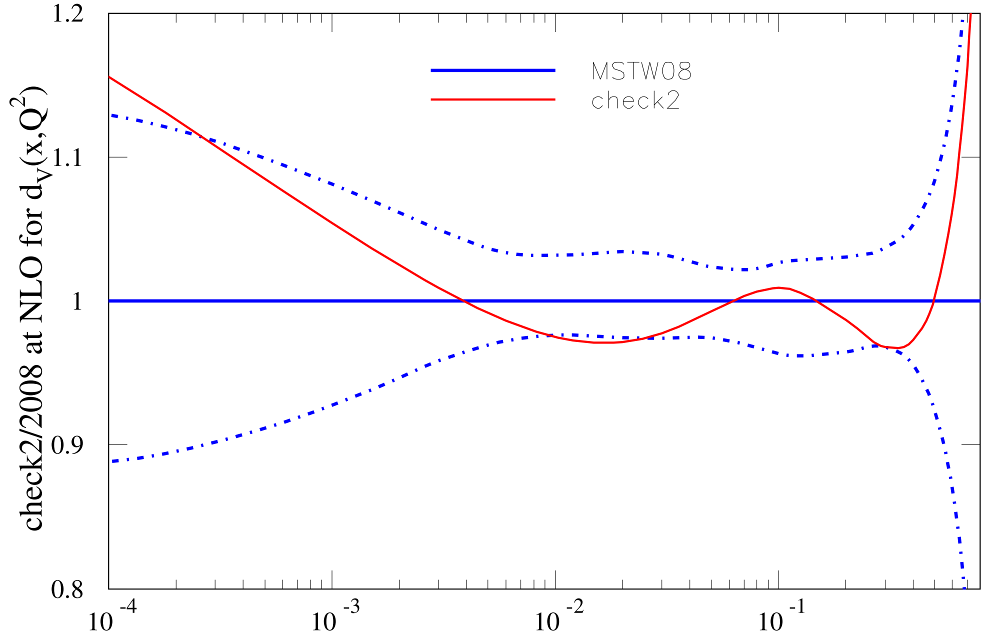
Set $\epsilon x^{0.5}$ term in u_V to zero. Usual magnitude is about twice the uncertainty.

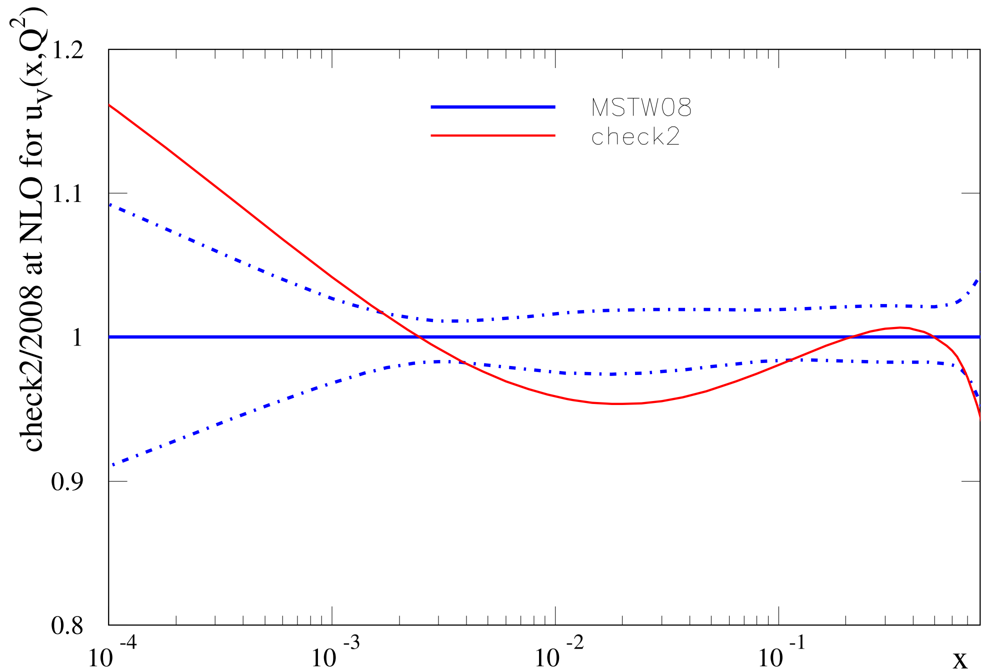
Refit with usual eigenvector parameters free. χ^2 is 30 worse.

Biggest change in PDFs shown. At most variation about 1.8 uncertainty, in u_V.

Again relevant eigenvectors suggest uncertainty corresponds to $\Delta\chi^2 \sim 10$.

This time change in PDF pretty much what deterioration in fit quality suggests.

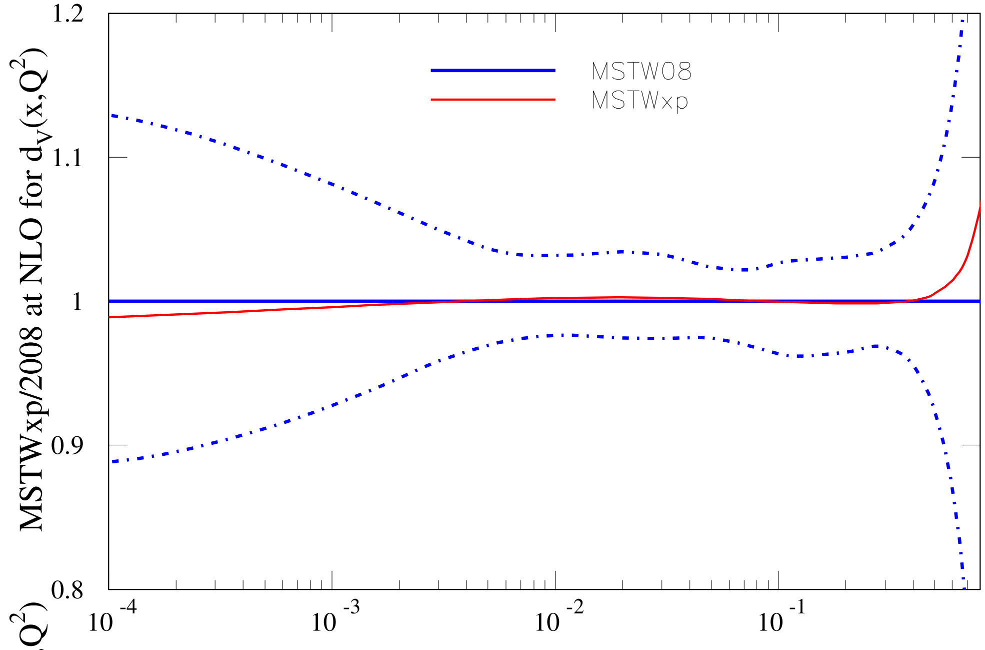


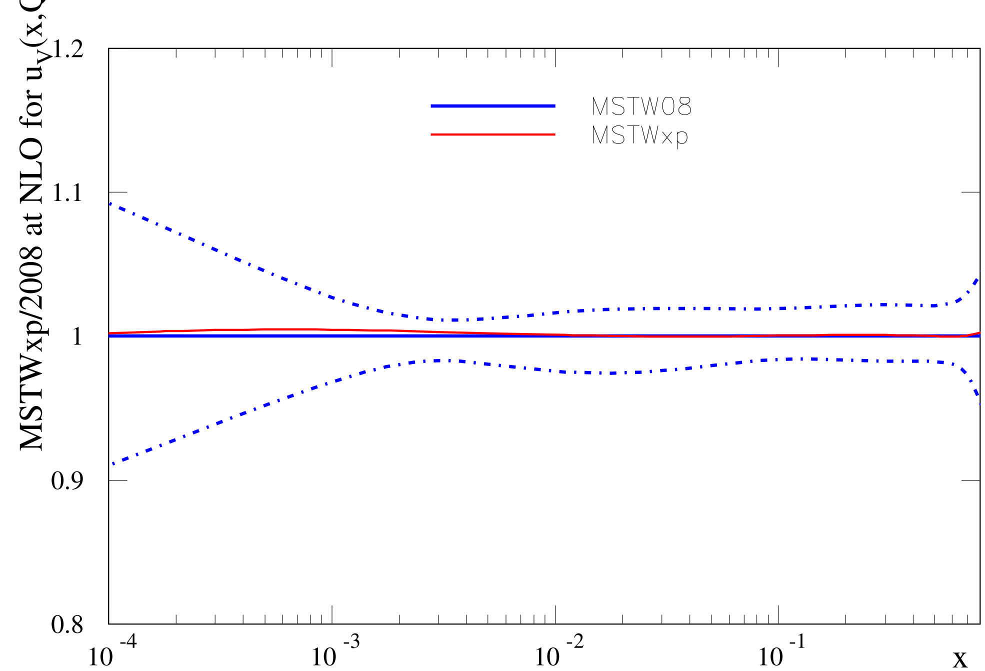


Also tried adding x^2 terms to polynomial in two valence parameterisations.

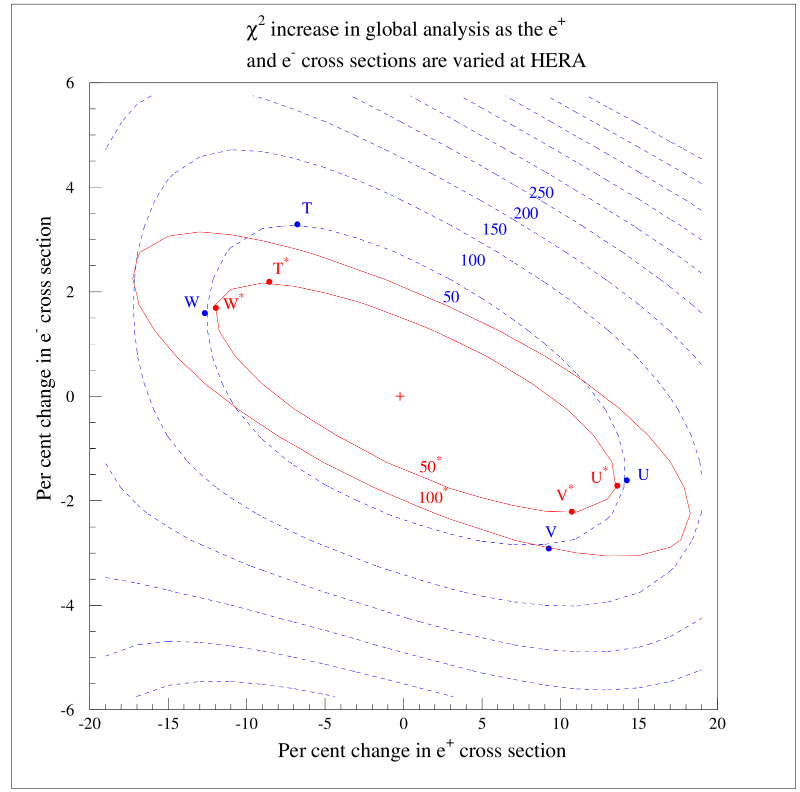
Fit quality improved by 2 units.

Change in partons negligible.





Recall study in first MRST uncertainties paper comparing the Hessian approach with 15 parameters and Lagrange multiplier with 22 parameters and same $\Delta\chi^2 = 50$ for both.



Plot shown for Lagrange multiplier method for charged current HERA structure functions at $x = 0.5$ (Red curve – fixed α_S).

Uncertainty using Hessian approach was 2% for $F_2^{CC}(e^-p)$ and 10% for $F_2^{CC}(e^+p)$.

Excellent agreement between two for $F_2^{CC}(e^-p)$.

Factor of up to 50% too low for $F_2^{CC}(e^+p)$.

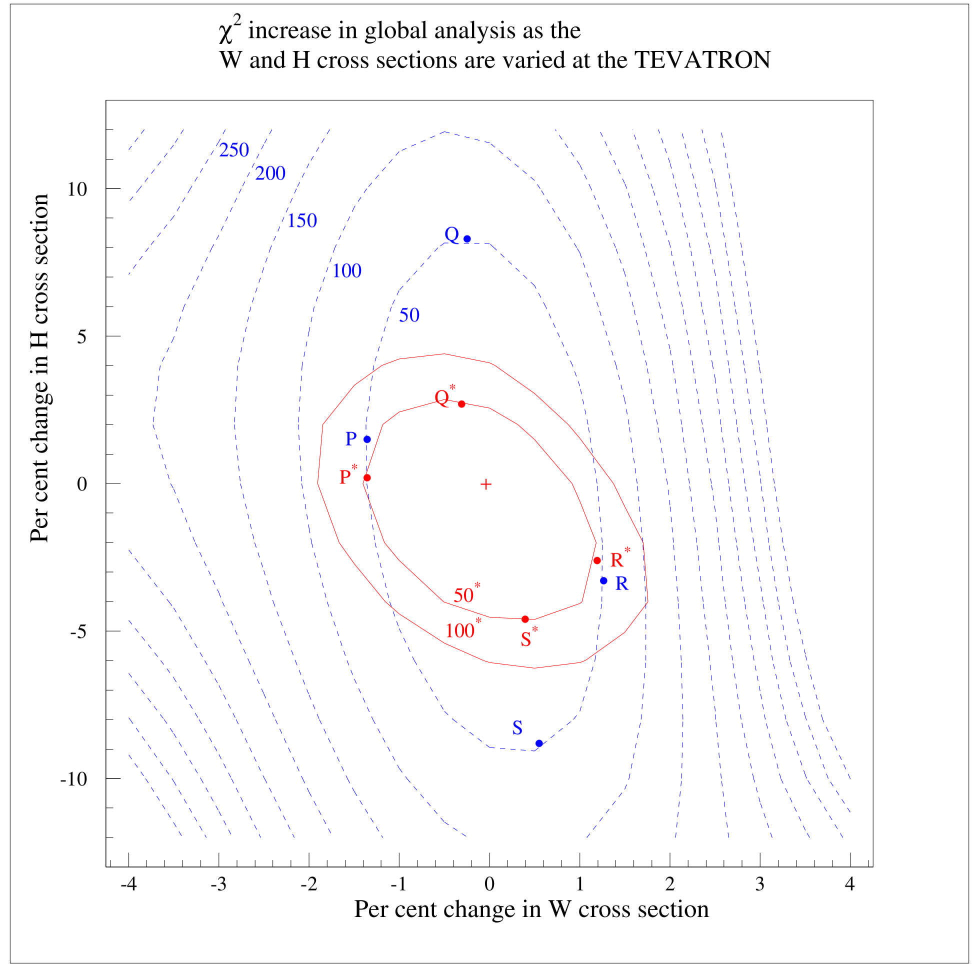
However, used non-optimum choice of parameters in eigenvectors for $d_V(x,Q^2)$ in MRST2001. Correction of this lead to automatic increase in uncertainty of about 50% at $x = 0.5$, with no new free parameters.

Lagrange multiplier method for W and $120\mathrm{GeV}$ Higgs at the Tevatron.

Uncertainty using Hessian approach is 1.2% for W and 3% for Higgs.

Slightly smaller in latter case. Using 3 parameters lead to narrowing uncertainty in gluon at $x \approx 0.2$ – affects Tevatron uncertainty.

Extra parameter in eigenvectors for gluon increases uncertainty by about the amount expected.



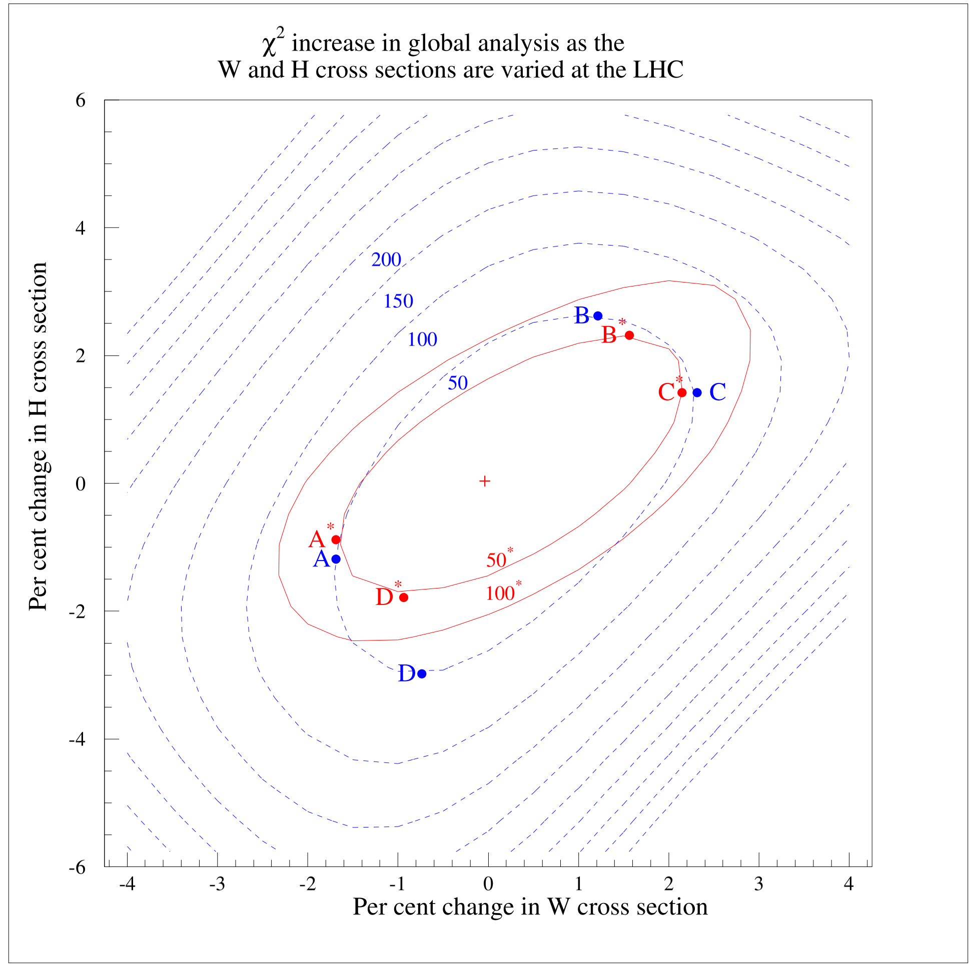
Lagrange multiplier method for W and $120\mathrm{GeV}$ Higgs at the LHC.

Uncertainty using Hessian approach about 10% smaller.

Also looked at uncertainties on moments of u–d using Hessian and Lagrange multiplier approaches. Very similar and latter could be slightly smaller.

In all cases introduction of extra parameters in the Lagrange multiplier method led to at most a moderate increase in uncertainty.

If this was clearly more than 10% the limitations in parameters were addressed and the problem solved.



Not looking for $\Delta\chi^2 = 100$ anyway

```
# iEigen x68plusMin x68minusMax
1 4.53655 -3.76623 (ZEUS ep 95-00 #sigma_{r}^{NC}, H1 ep 97-00 #sigma_{r}^{NC})
2 3.38422 -3.79217 (NMC #mud F_{2}, NuTeV #nuN#rightarrow#mu#muX)
3 1.46292 -2.17007 (NuTeV #nuN#rightarrow#mu#muX, CCFR #nuN#rightarrow#mu#muX)
4 3.45159 -2.31949 (NMC #mun/#mup, E866/NuSea pd/pp DY)
5 1.49487 -2.12523 (NuTeV #nuN#rightarrow#mu#muX, NuTeV #nuN xF_{3})
6 5.2242 -3.21227 (H1 ep 97-00 #sigma_{r}^{NC}, NuTeV #nuN#rightarrow#mu#muX)
7 2.03521 -2.78497 (D#oslash II W#rightarrowl#nu asym., BCDMS #mud F_{2})
8 5.20184 -1.84172 (NuTeV #nuN F_{2}, BCDMS #mup F_{2})
9 3.89046 -3.63201 (H1 ep 97-00 #sigma_{r}^{NC}, ZEUS ep 95-00 #sigma_{r}^{NC})
10 2.99034 -2.67972 (D#oslash II W#rightarrowl#nu asym., SLAC ed F_{2})
11 3.74202 -6.58278 (H1 ep 97-00 #sigma_{r}^{NC}, ZEUS ep 95-00 #sigma_{r}^{NC})
12 5.18993 -3.20527 (SLAC ep F_{2}, BCDMS #mup F_{2})
13 3.32487 -1.57418 (E866/NuSea pp DY, NuTeV #nuN xF_{3})
14 4.21973 -3.62346 (NMC #mud F_{2}, D#oslash II W#rightarrowl#nu asym.)
15 2.63335 -4.3632 (H1 ep 97-00 #sigma_{r}^{NC}, NuTeV #nuN F_{2})
16 2.32169 -0.925389 (CCFR #nuN#rightarrow#mu#muX, E866/NuSea pd/pp DY)
17 2.31104 -1.51795 (NuTeV #nuN#rightarrow#mu#muX, CCFR #nuN#rightarrow#mu#muX)
18 2.88709 -1.42061 (D#oslash II W#rightarrowl#nu asym., E866/NuSea pd/pp DY)
19 4.20991 -4.1461 (H1 ep 97-00 #sigma_{r}^{NC}, CDF II p#bar{p} incl. jets )
20 3.70876 -2.47281 (NuTeV #nuN#rightarrow#mu#muX, NuTeV #nuN#rightarrow#mu#muX)
```

Majority of eigenvectors correspond to $\sqrt{\Delta\chi^2} \sim 2-3$.

More types of data and weaker cuts than CTEQ. Even more discrepancy?

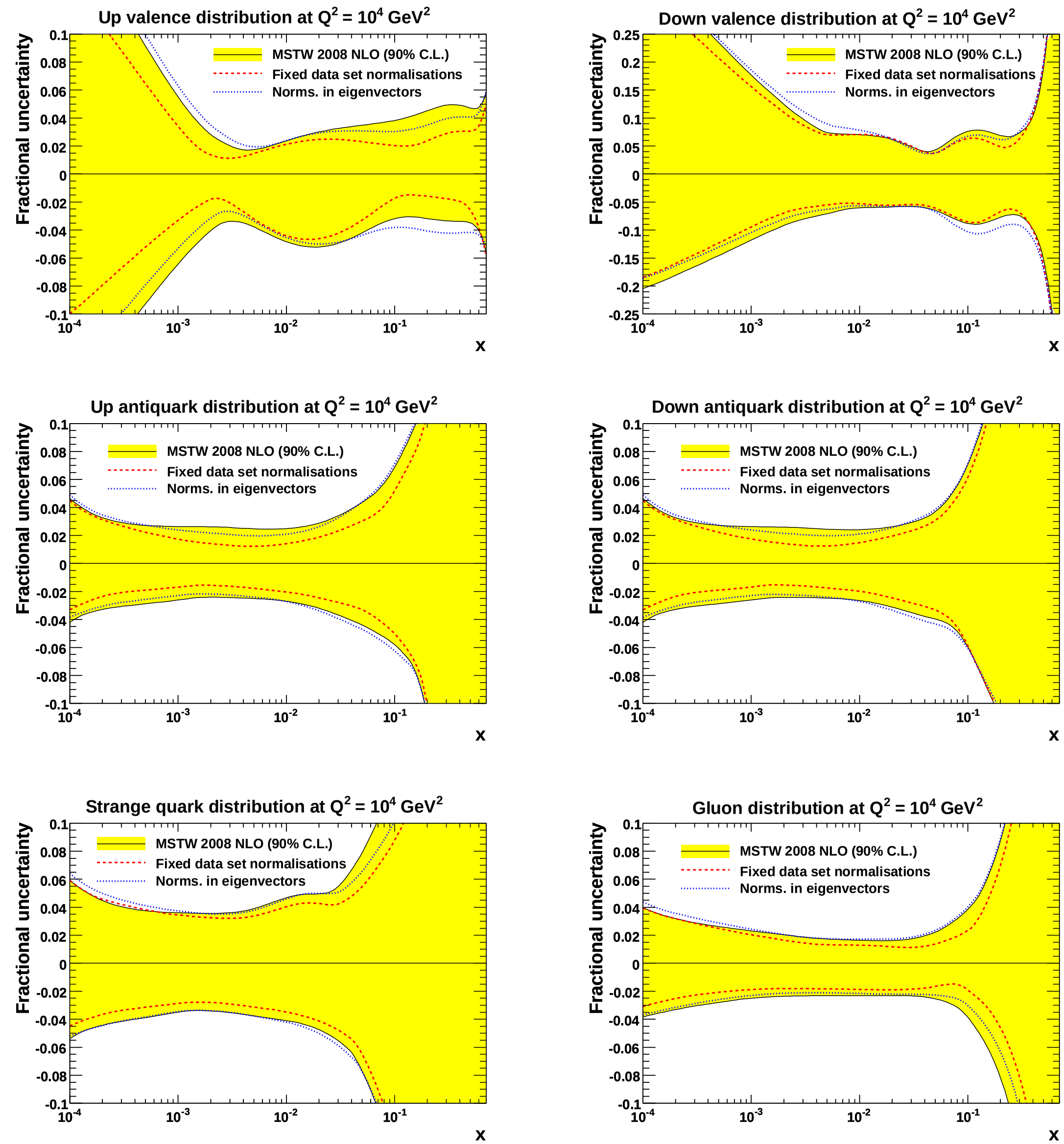
Comparison of full uncertainty and that from no normalization uncertainties (except in best fit).

Normalization uncertainty $\sim 1-1.5\%$, for all partons.

Difficult to account for in tolerance for eigenvectors – some very sensitive (size of quarks) others insensitive ($\bar{u}-\bar{d}$ determined from ratios).

Use of normalisation uncertainties increases uncertainties on partons significantly.

Not applied by CTEQ. Part of the reason for large tolerance?



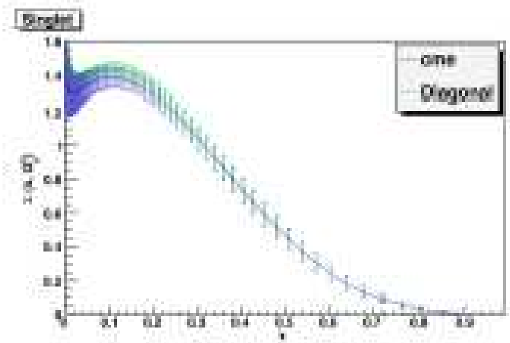
THE IMPACT OF CORRELATED UNCERTAINTIES

REPEAT THE FIT NEGLECTING ALL CORRELATIONS (A.Donati)

		CME fit		Diagonal fit	
Experiment	Set	χ^2_{diag}	χ^2_{corr}	χ^2_{diag}	χ^2_{corr}
TOT (all exp)		0.988	1.323	0.844	1.321
NMC-pd		1.965	1.457	1.167	1.155
NMC		1.006	1.659	1.078	1.76
SLAC		0.836	1.185	1.008	1.406
	SLACp	1.018	1.307	1.132	1.525
	SLACd	0.651	0.912	0.882	1.275
BCDMS		0.777	1.646	0.552	1.604
	BCDMSp	0.873	1.808	0.617	1.703
	BCDMSd	0.648	1.296	0.465	1.23
ZEUS		0.770	1.055	0.742	1.048
	Z97lowQ2	0.474	1.294	0.434	1.367
	Z97NC	0.718	1.125	0.669	1.106
	Z97CC	0.912	0.800	1.021	0.894
	Z02NC	0.798	0.767	0.763	0.733
	Z02CC	0.619	0.592	0.593	0.569
	Z03NC	0.975	1.104	0.907	1.012
	Z03CC	1.131	1.001	1.259	1.115
H1		1.020	1.053	0.997	1.028
	H197mb	0.861	1.298	0.877	1.33
	H197lowQ2	0.666	0.948	0.774	0.97
	H197NC	1.071	0.903	0.986	0.852
	H197CC	0.758	0.764	0.831	0.824
	H199NC	1.229	1.109	1.171	1.068
	H199CC	0.621	0.646	0.644	0.668
	H199NChy	0.333	0.361	0.326	0.353
	H100NC	1.208	1.172	1.120	1.102
	H100CC	1.122	1.013	1.311	1.146
CHORUS		1.018	1.380	0.745	1.392
	CHORUSnu	1.082	1.449	0.628	1.403
	CHORUSnb	0.954	1.178	0.861	1.254
FLH108		0.984	1.729	0.946	1.7
NTVDMN		0.869	0.692	1.094	0.984
	NTVnuDMN	1.061	0.763	0.445	0.421
	NTVnbDMN	0.667	0.660	1.774	1.618
ZEUS-H2		1.392	1.509	1.373	1.512
	Z06NC	1.691	1.495	1.667	1.472
	Z06CC	0.664	1.230	0.659	1.252

- DIAGONAL χ^2 OF DIAGONAL FIT MUCH LOWER, CORREL. χ^2 OF TWO FITS UNCHANGED
- DIAGONAL FIT REWEIGHTS EXPERIMENTS $\Rightarrow$ EXPTS WITH LARGER SYST. (FIXED TARGET) GET SMALLER WEIGHT
- VALENCE & STRANGE PDFS AFFECTED AT THE $\frac{1}{4}\sigma$ LEVEL

SINGLET



STRANGE