# MSTW2008 Parton Distributions

Robert Thorne

April 28th, 2009

University College London

(In collaboration with A.D. Martin, W.J. Stirling and G. Watt)

(arXiv:0901.0002)

MSTW08 – **New data included.**

NuTeV and CHORUS data on $F_2^{\nu,\bar{\nu}}(x,Q^2)$ and $F_3^{\nu,\bar{\nu}}(x,Q^2)$ replacing CCFR.

NuTeV and CCFR dimuon data included directly. Leads to a direct constraint on $s(x,Q^2)+\bar{s}(x,Q^2)$ and on $s(x,Q^2)-\bar{s}(x,Q^2)$. Affects other partons.

CDFII lepton asymmetry data in two different $E_T$ bins – $25\mathrm{GeV} < E_T < 35\mathrm{GeV}$ and $35\mathrm{GeV} < E_T < 45\mathrm{GeV}$. D0II data for $E_T > 20\mathrm{GeV}$. First data sets only.

CDFII and D0II data on $d\,\sigma(Z)/d\,y$ for $0 < y < 3$.

HERA inclusive jet data (in DIS).

New CDFII and D0II high-$p_T$ jet data.

Direct high-$x$ data on $F_L(x,Q^2)$.

All published (at the time) charm structure function data.

Would like averaged HERA structure function data.

## MSTW08 – **Major changes in theory/approach.**

Implementation of updated heavy flavour VFNS, particularly at NNLO. Already use in MRST06 NNLO distributions, but not in official NLO sets. (Already used a general VFNS since 1998 but change in details.)

Inclusion of NNLO corrections (Anastasiou, Dixon, Melnikov, Petriello) Drell-Yan ( $W, Z$ and $\gamma^*$) data using Vrap and FEWZ.

Change in definition of $\alpha_S$ – same as QCDNUM, Pegasus. No $\Lambda_{QCD}$ parameter.

Improved nuclear corrections, De Florian and Sassot obtained from NLO partons.

Implementation of fastNLO – fast perturbative QCD calculations Kluge, Rabbertz, Wobisch. Allows easy inclusion of new jet data from both Tevatron and HERA.

Change in means of obtaining uncertainties. Still diagonalise covariance matrix of parameters and use 20 (previously 15) orthogonal eigenvectors. Tolerance now determined differently.

# GM-VFNS– **Response to Plenary** (Nadolsky)

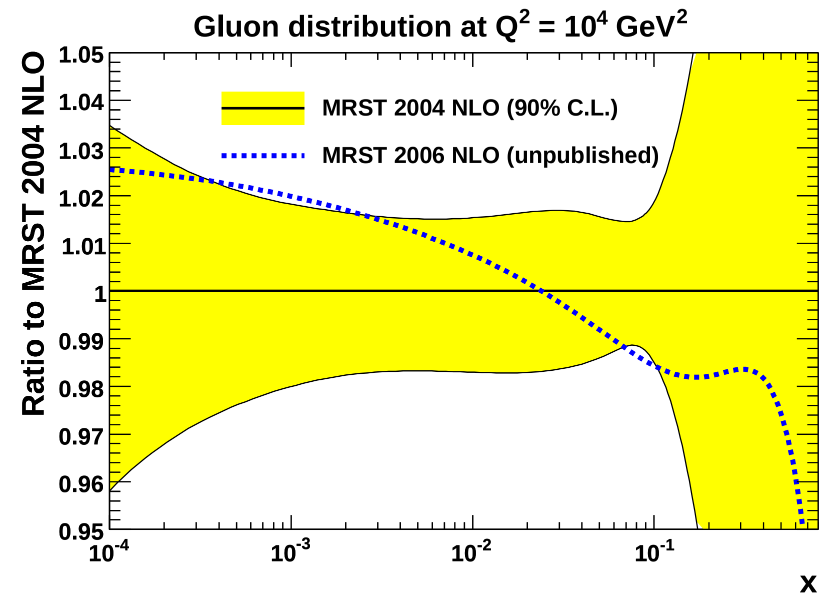
Always used GM-VFNS since it obviously impacts on LHC and Tevatron predictions - compare only to CTEQ5HQ in MRST99.

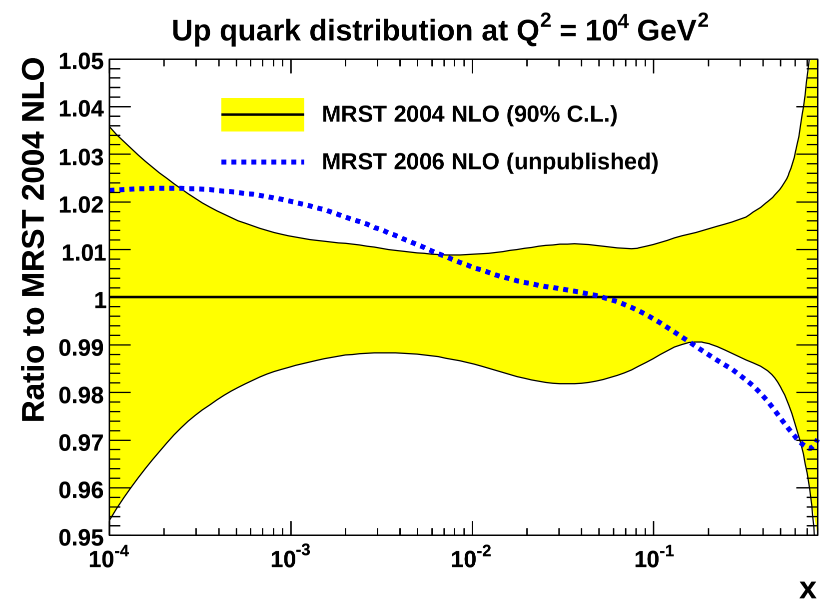
Change in GM-VFNS presented at DIS05 - published Phys.Rev.D73:054019,2006 (cited in CTEQ6.5). Study showed that previous NNLO VFNS (only) too approximate – don't use pre-2006 NNLO.

Pre-2006 NLO different but ok GM-VFNS (more complicated for NNLO). Do not discard pre-2006 NLO sets (except perhaps because they use older data).

Compare MRST2004 to unofficial "MRST2006 NLO" (same data). Changes of up to 2% in PDFs. This is a genuine theory uncertainty.

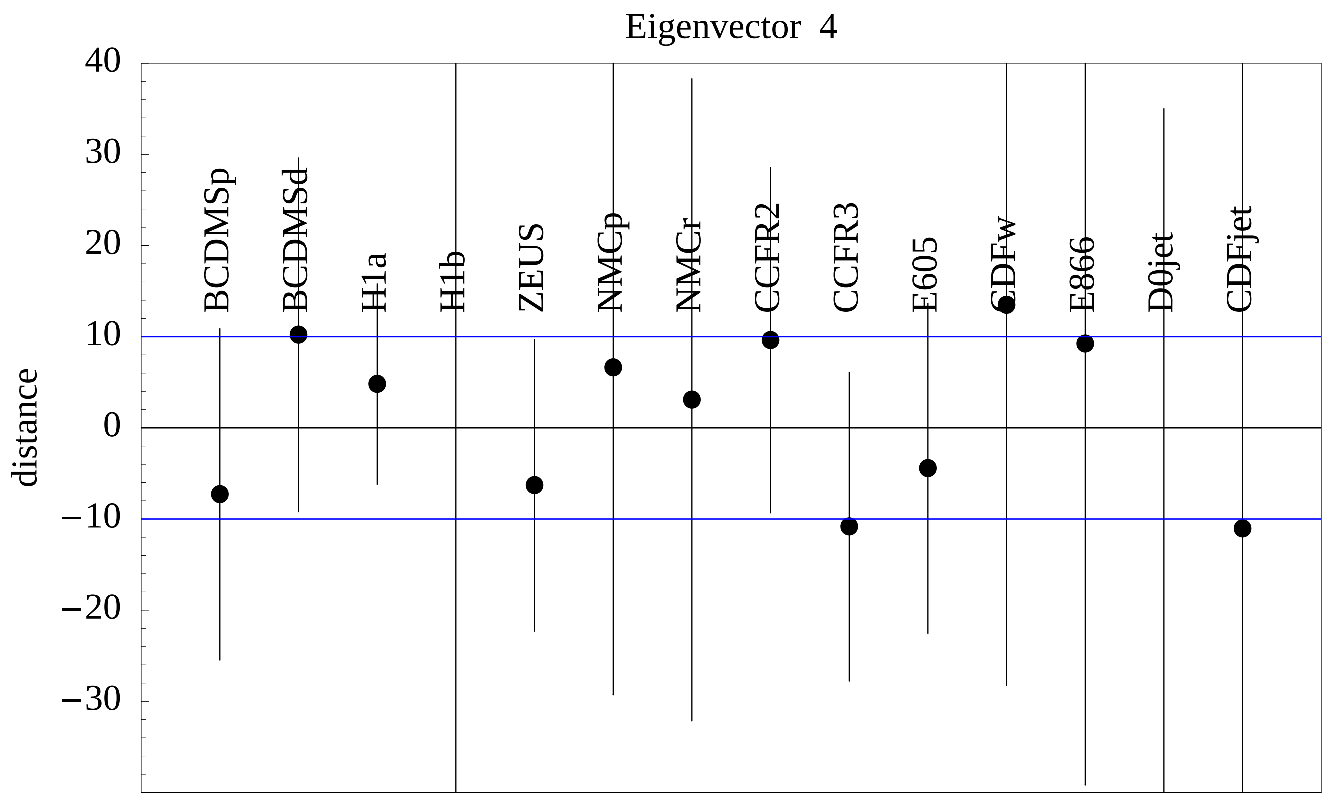
See RT, Tung – arXiv:0809.0714.



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 $\chi^2$ 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

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

- $\xi_{90}$ is the 90th percentile of the $\chi^2$-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

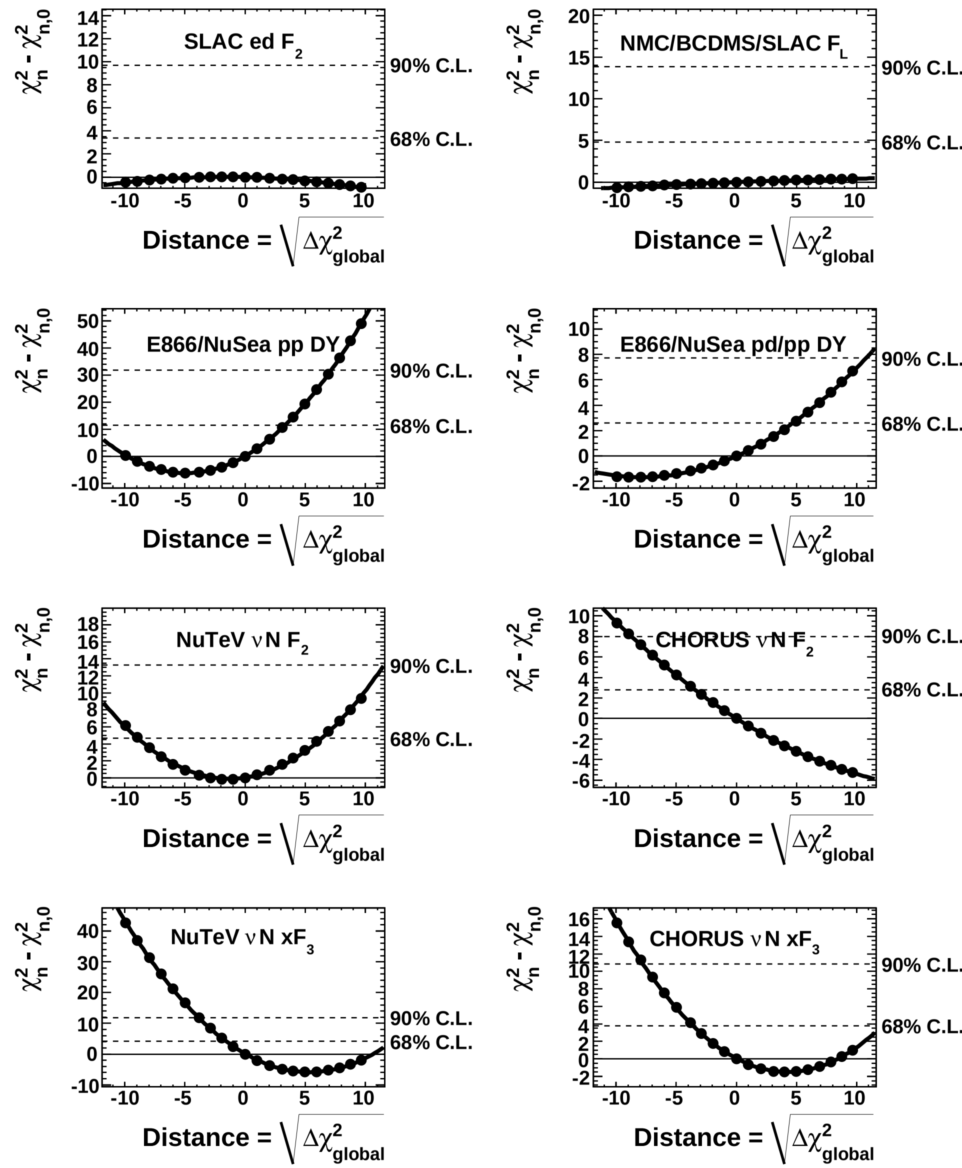
$$f(z; N) = \frac{z^{N/2-1} \, \mathrm{e}^{-z/2}}{2^{N/2} \, \Gamma(N/2)}.$$

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

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

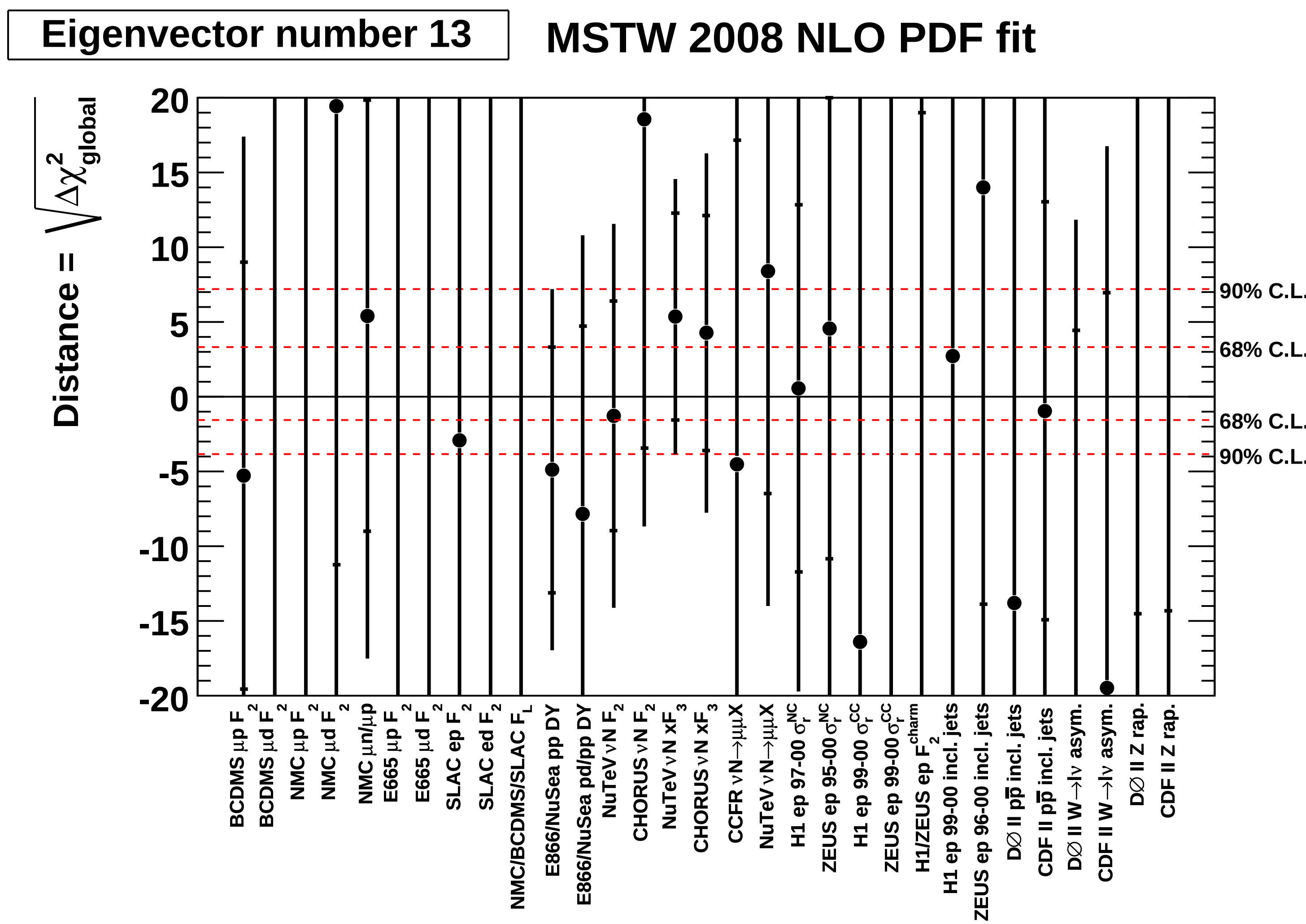
For each determine the point in $\Delta\chi^2_{\rm global}$ at which the appropriate confidence level limit is reached in each direction.

## MSTW 2008 NLO PDF fit
## Eigenvector number 13

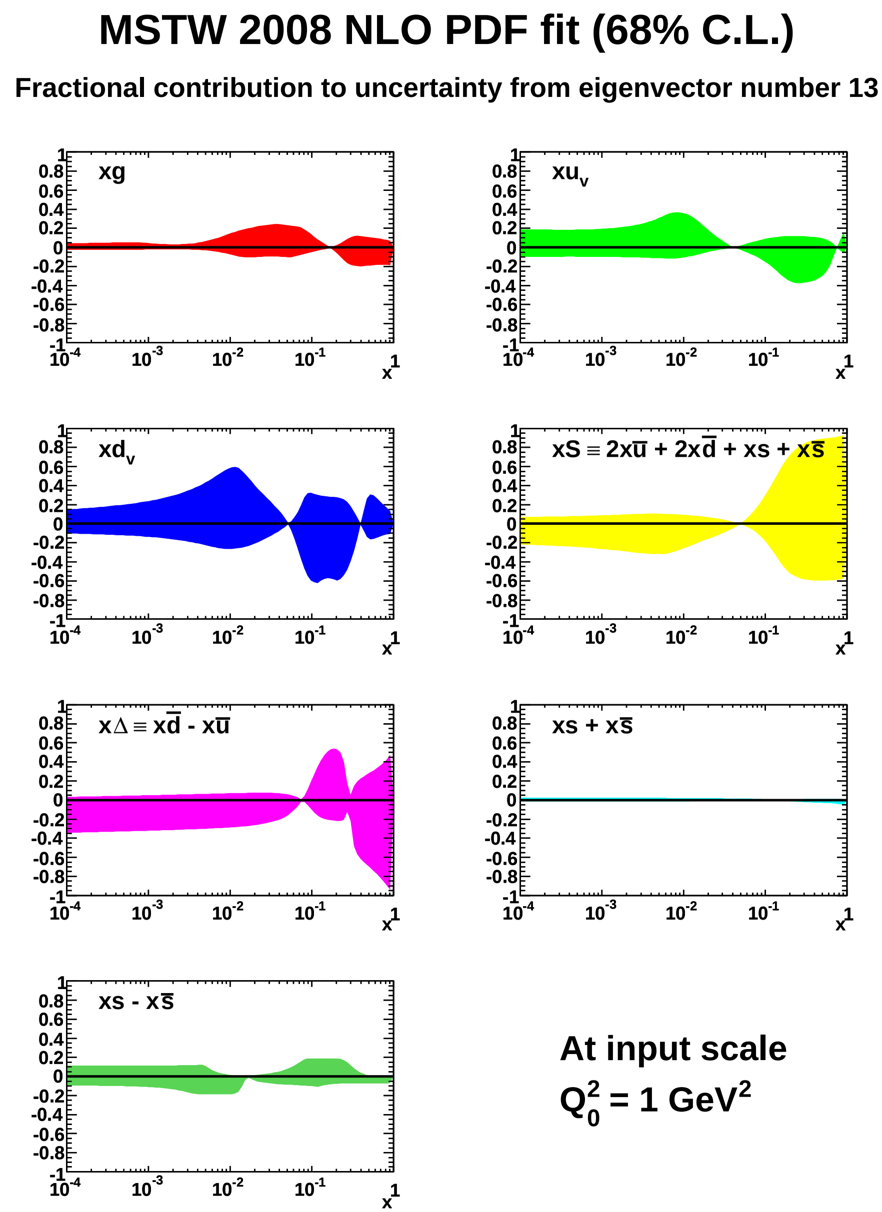


Plot this for all data sets for a given eigenvector.

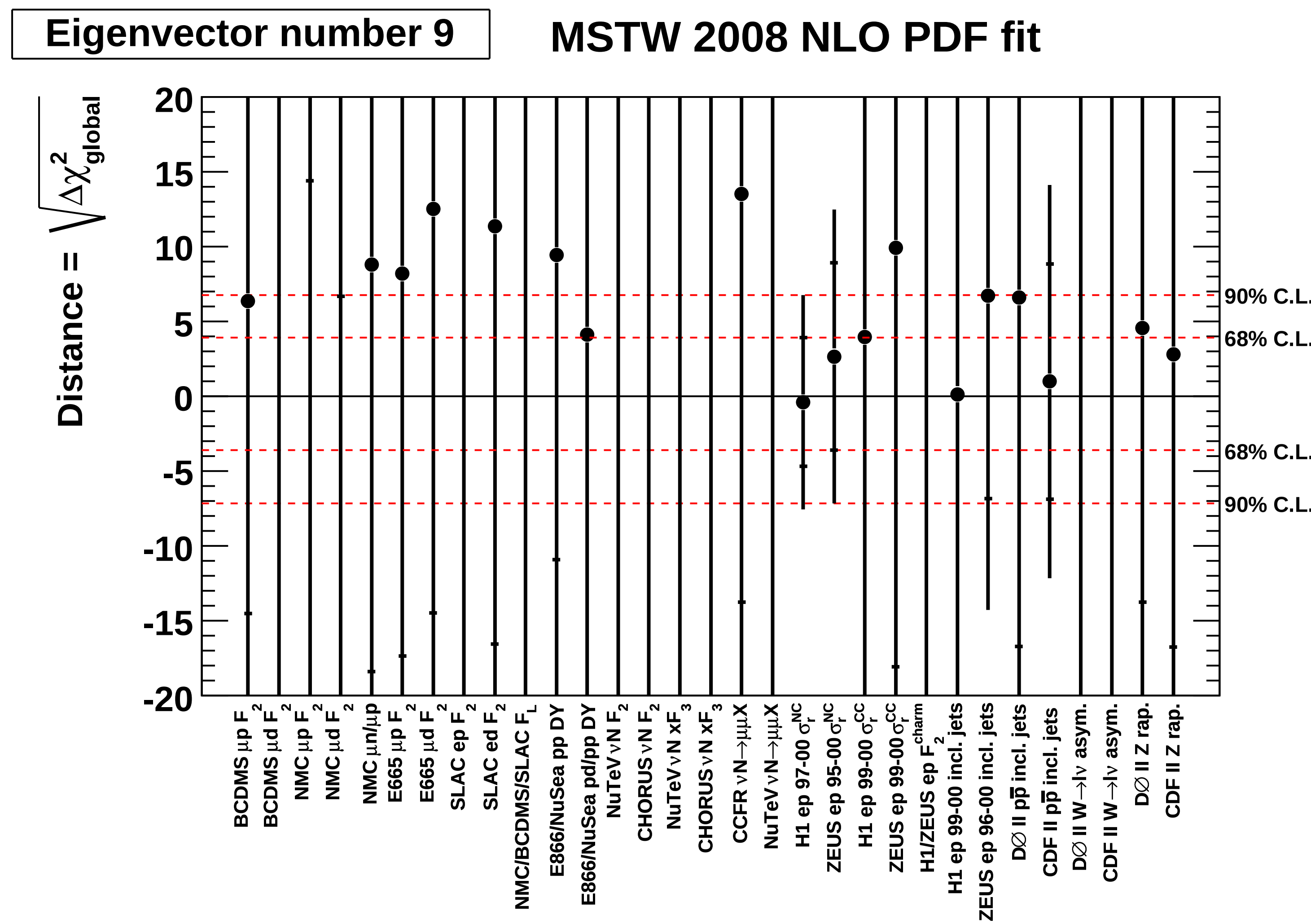
Eigenvector 13 constrained in one direction by E866 Drell-Yan data and in the other direction by NuTeV $F_3^p(x, Q^2)$ data . In this case the best fits for the two sets are highly inconsistent. $\Delta\chi^2 = 100$ well outside $90\%$ confidence level for each.



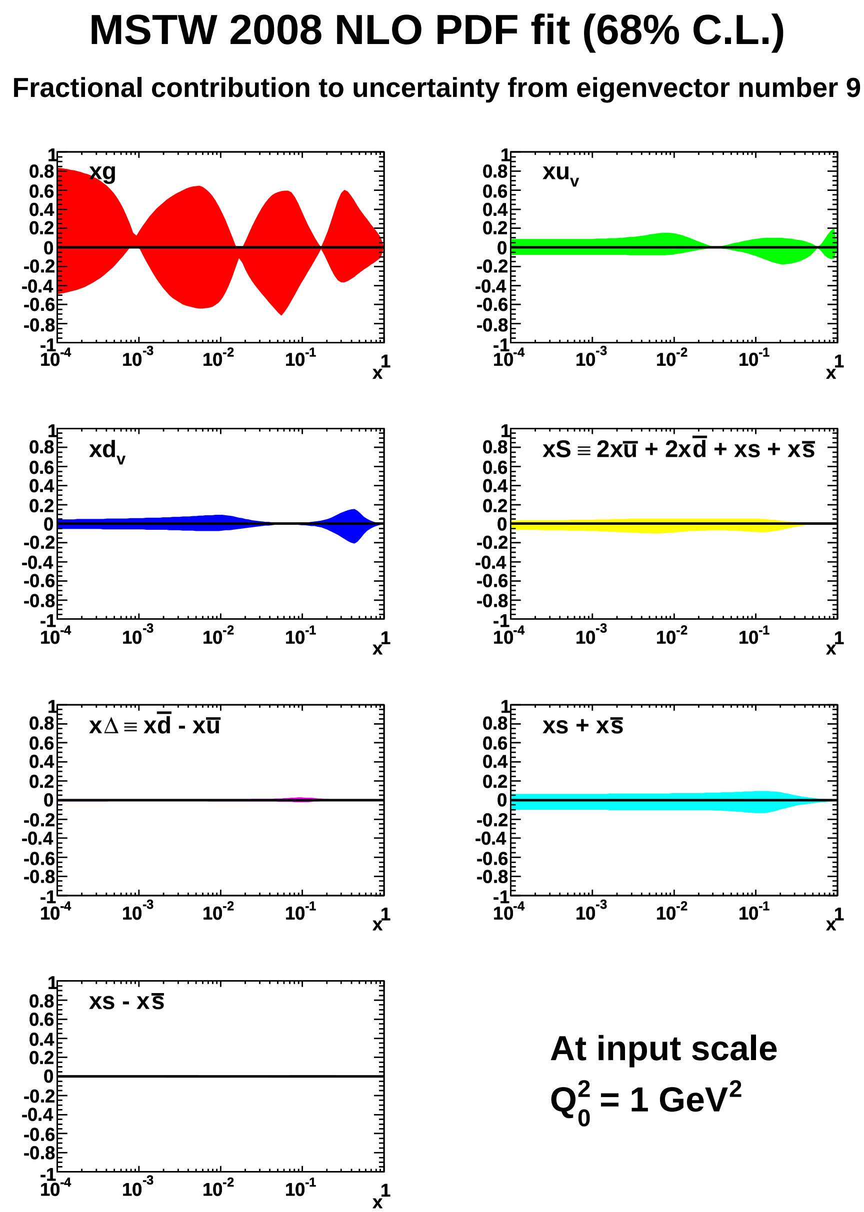
This eigenvector contributes most to the high-$x$ sea quark uncertainty, but also a variety of other quarks.



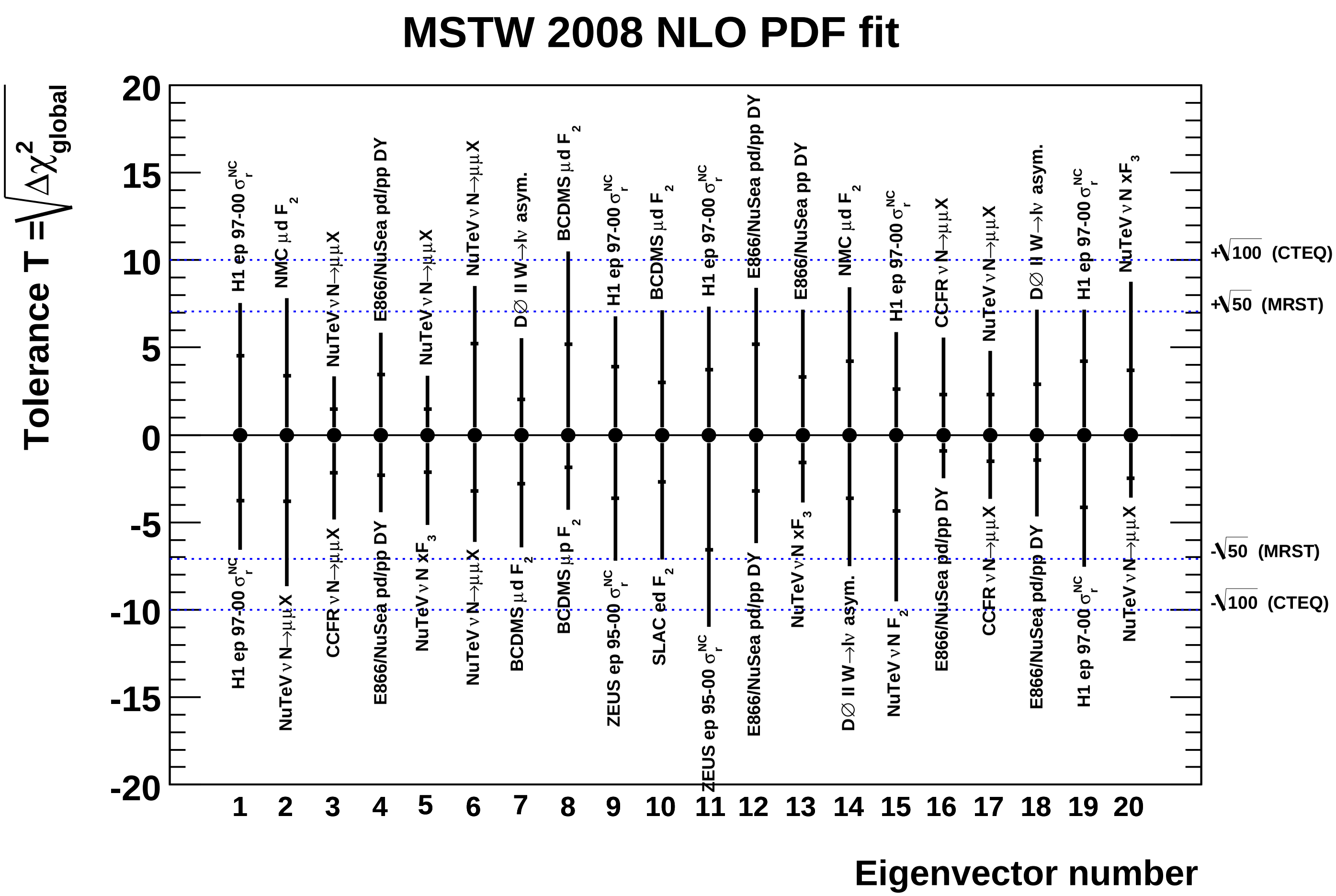
As a simpler example, eigenvector 9 constrained most by H1 and ZEUS data on $F_2^p(x, Q^2)$. 90% confidence limit determining by H1 in *up* direction and ZEUS in *down* direction. Both $\Delta\chi^2 \approx 50$.



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.




In practice should give a conservative estimation of uncertainties.

## Normalisation Uncertainties

Previously (and still for CTEQ) the normalization of each data set was determined by the best fit – and then fixed.

Technical difficulties in including this feature in uncertainties.

CTEQ argue that this is part of the reason for large tolerance. Really seems to be a separate issue.

Now implement procedure of allowing normalisations of all sets to vary while performing scans over eigenvectors.

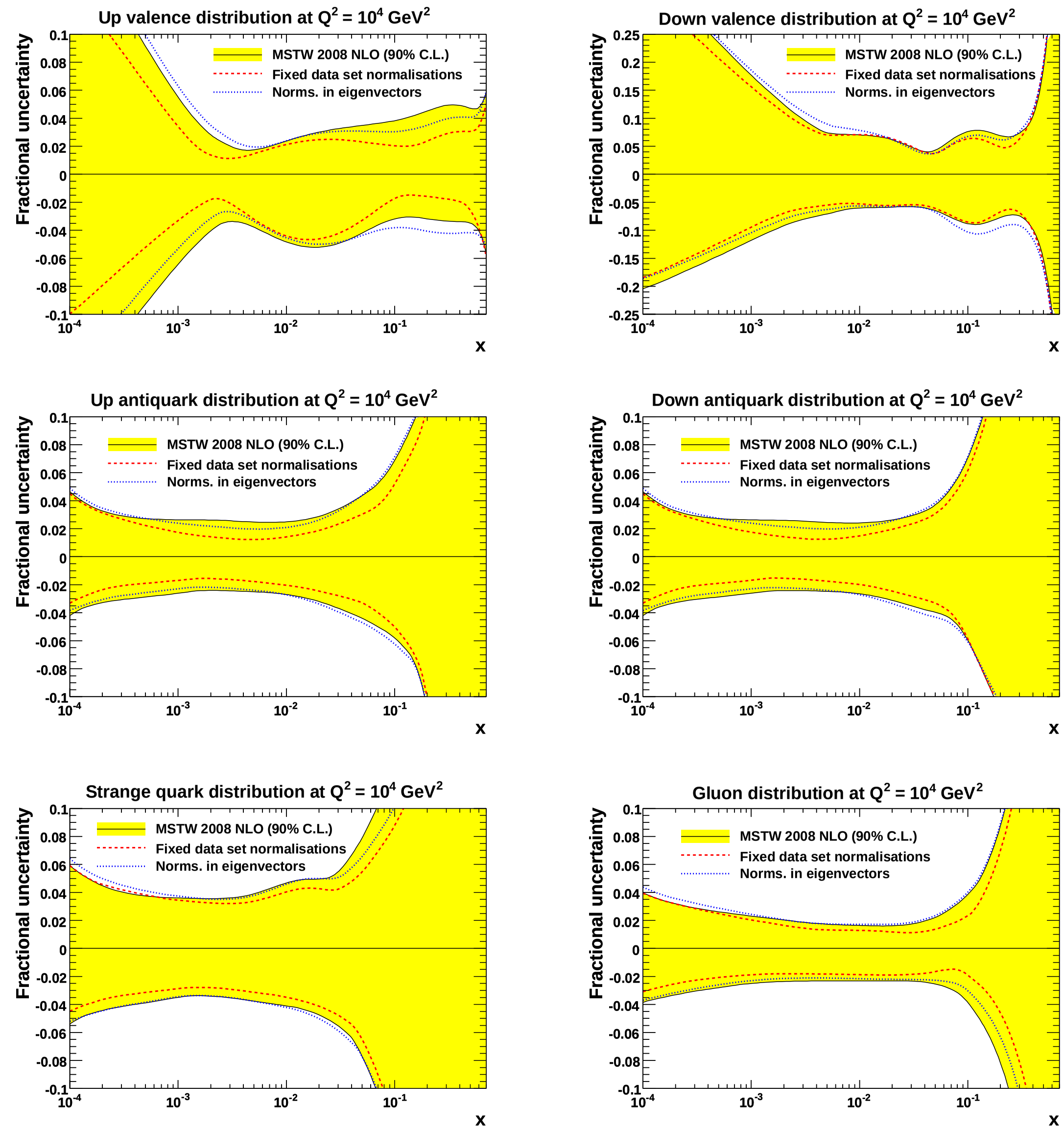
In practice normalization uncertainty $\sim 1.5\%$, for all partons.

Largely independent of $x$ (except high-$x$ gluon where only a small part of total uncertainty).

However now well-determined. Would be difficult to account for in tolerance for eigenvectors since some very sensitive (size of quarks) others insensitive ($\bar{u} - \bar{d}$ determined from ratios).

Comparison of full uncertainty and that from ignoring normalization uncertainties (other than in best fit).

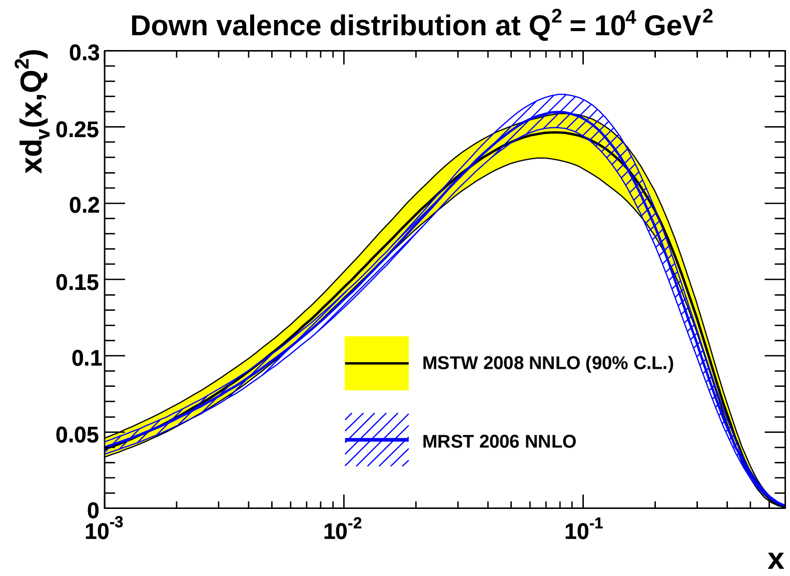
Also shown (blue), alternative approach for including normalization uncertainties (more efficient in fit – less useful).

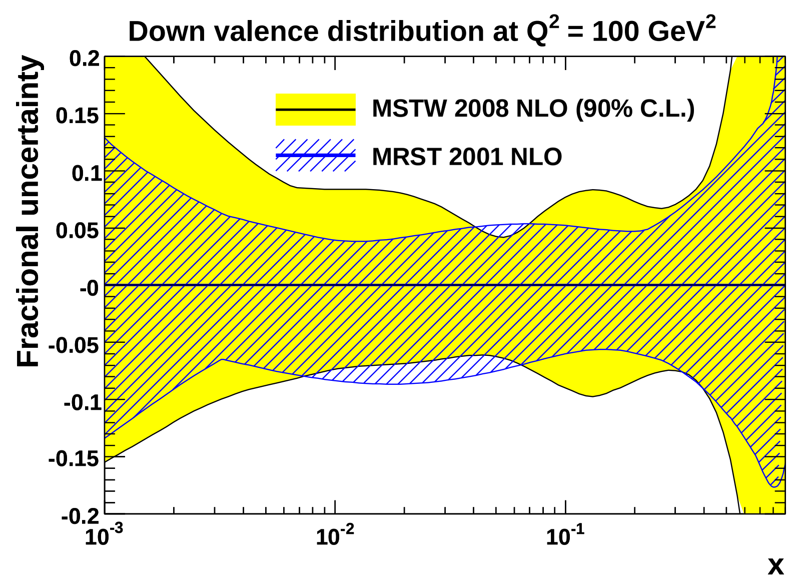


Overall $d_V(x,Q^2)$ now chooses a different type of shape.

Mainly changed by new *Tevatron* $W$-asymmetry data and new neutrino structure function data.

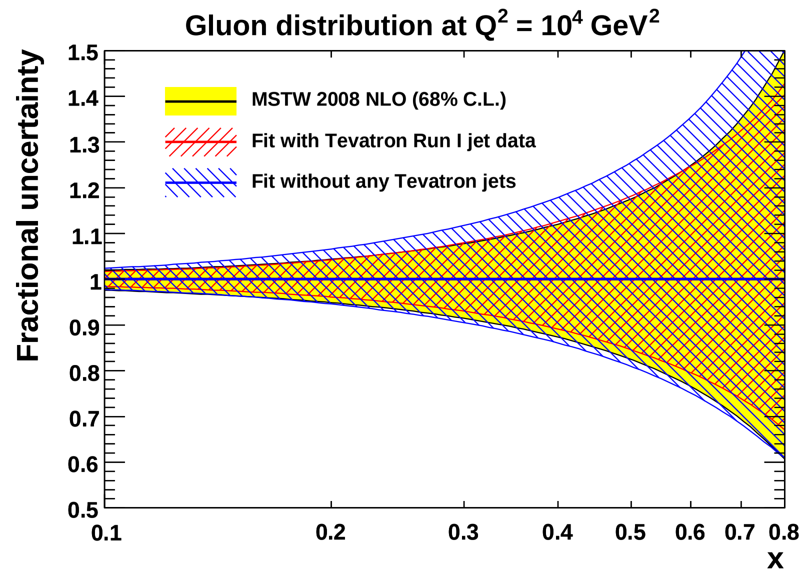
Uncertainty growing more quickly as $x \rightarrow 0$ and $x \rightarrow 1$ than before due to better parameterisation in determining uncertainty eigenvectors.



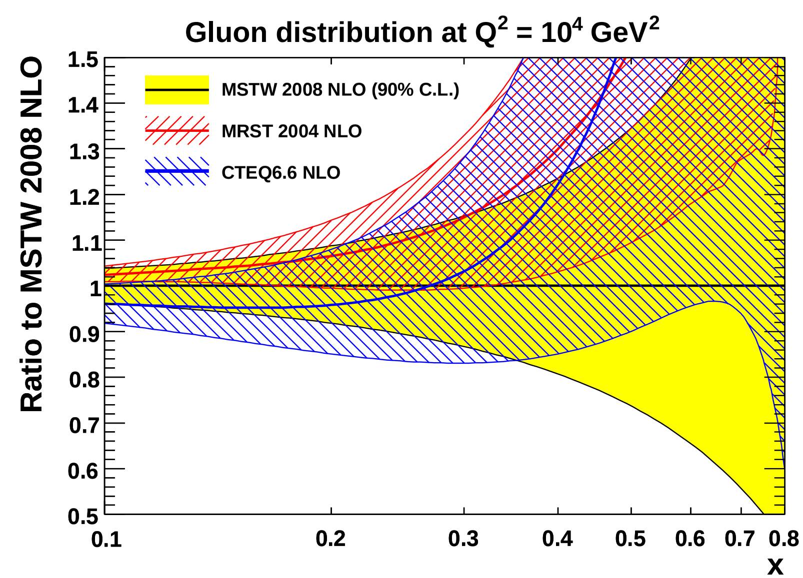



Even though one data set constrains each eigenvector limit, doesn't mean others do not contribute.

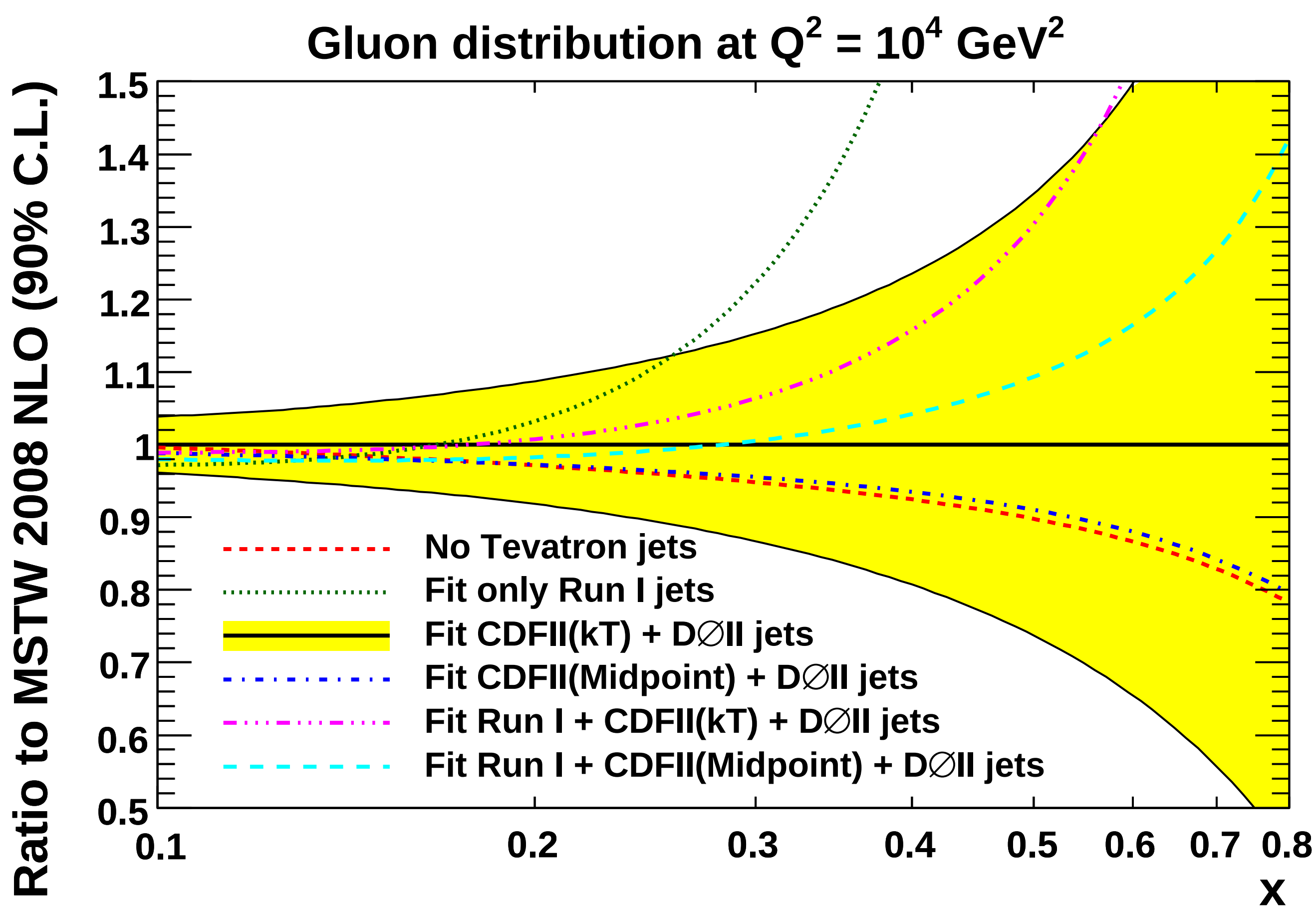
Tevatron jets are never main constraint but impact significantly on uncertainty.




Inclusion of Run II Tevatron jet data instead of Run I softens gluon at high $x$ significantly.

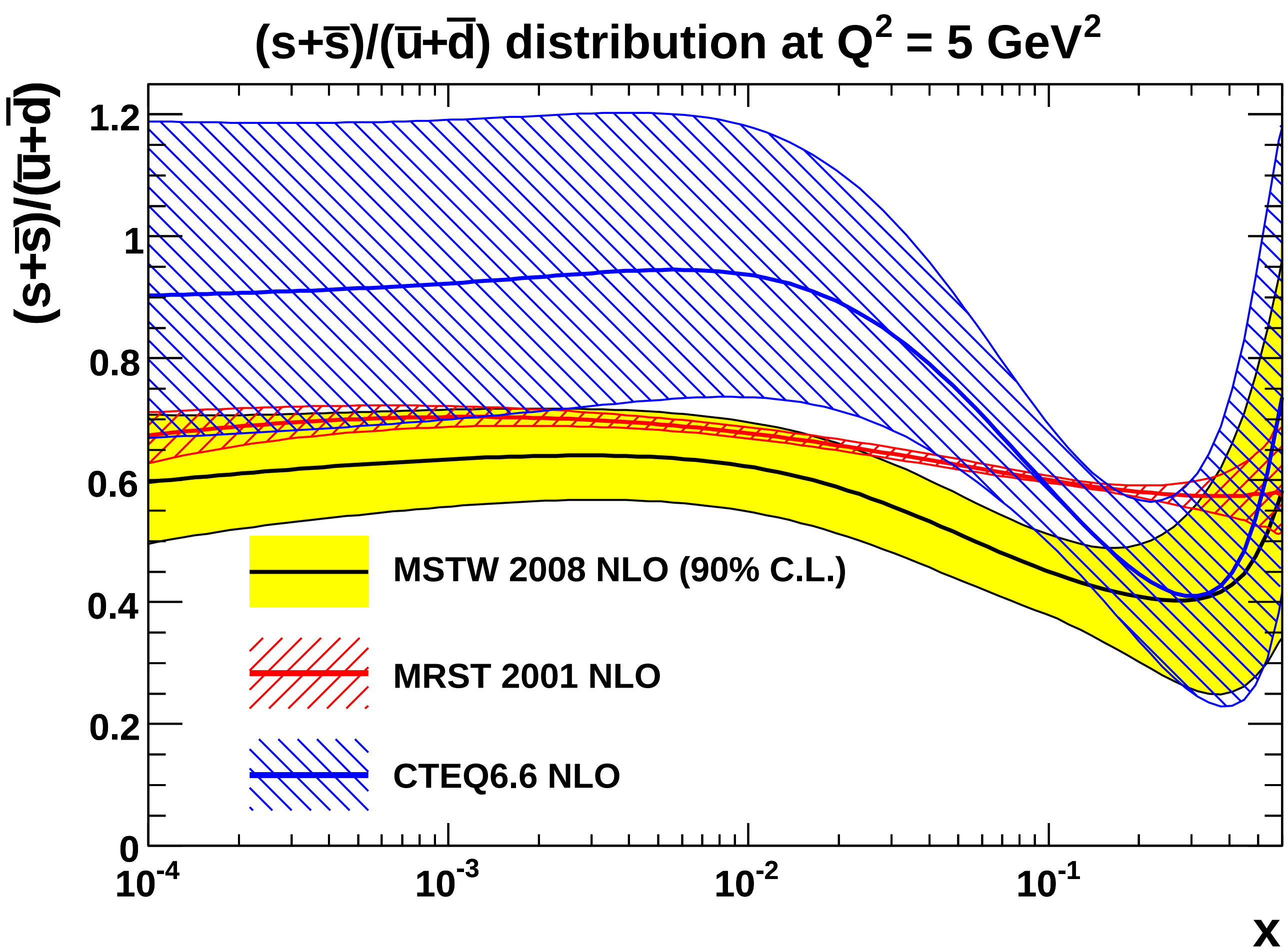


Unlike CT09G we do not find good consistency between fits to Run I and Run II jet data – though 90% uncertainty bands overlap.




Little change in conclusions if CDFII $k_T$-algorithm or midpoint algorithm data used.

Direct fit to $s, \bar{s}$ from dimuon data leads to significant uncertainty increase compared to assumption of fixed fraction of sea.




Significant difference to CTEQ fitting to same data. At small $x$ assume shape of input sea quarks is the same (consistent with mass suppression) whereas CTEQ have different parameterisation.

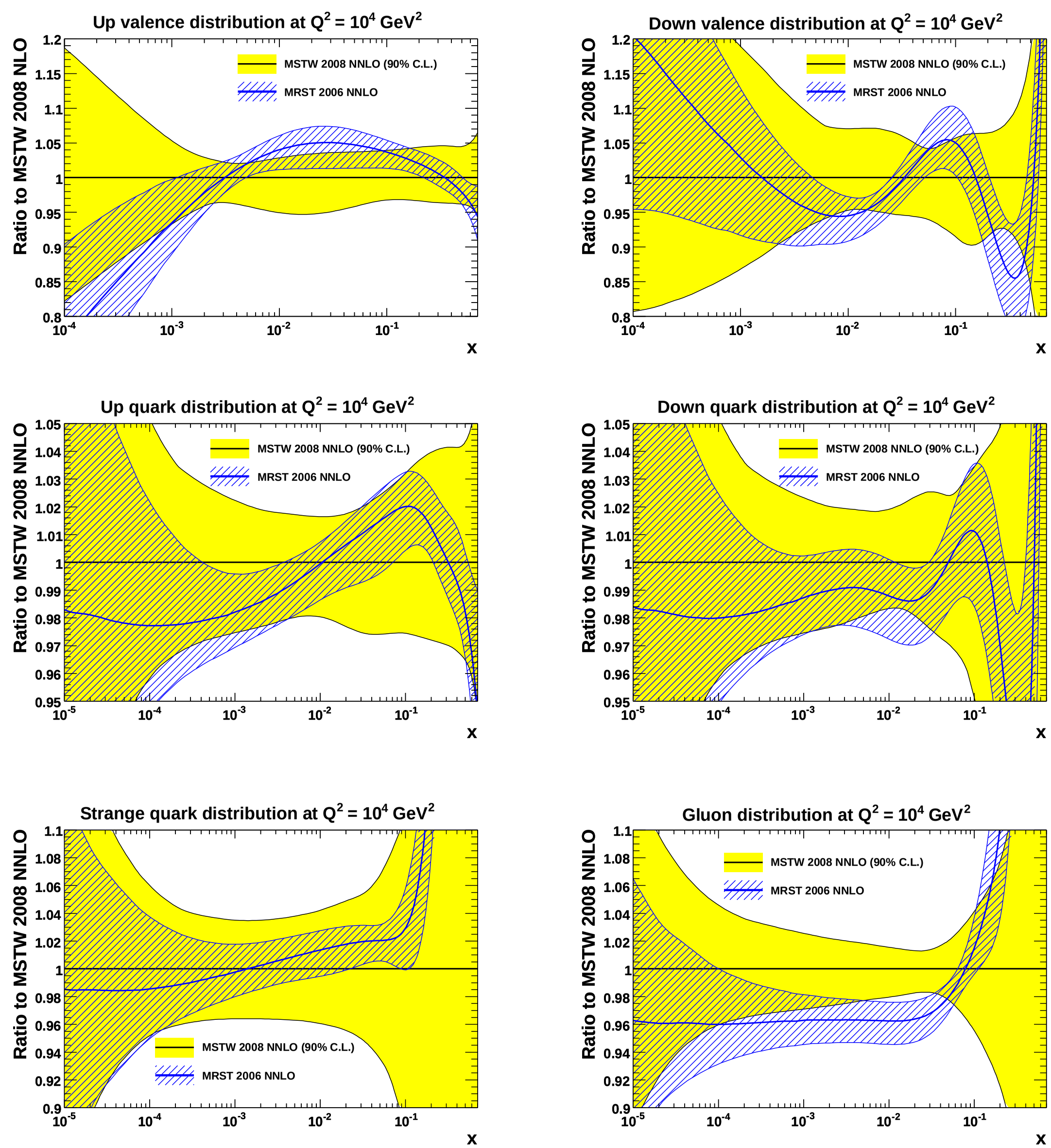
Comparison to MRST2006 PDFs.

Generally larger uncertainties from normalization uncertainties and more flexible parameterisations. Tolerance for eigenvectors generally smaller.

Change in shapes of gluon due to new Tevatron jet data, in strange from fitting dimuon data and in down from Tevatron vector boson production and neutrino data.

Other changes partly due to lower $\alpha_S(M_Z^2)$, e.g. at NNLO goes from $0.119 \rightarrow 0.117$.

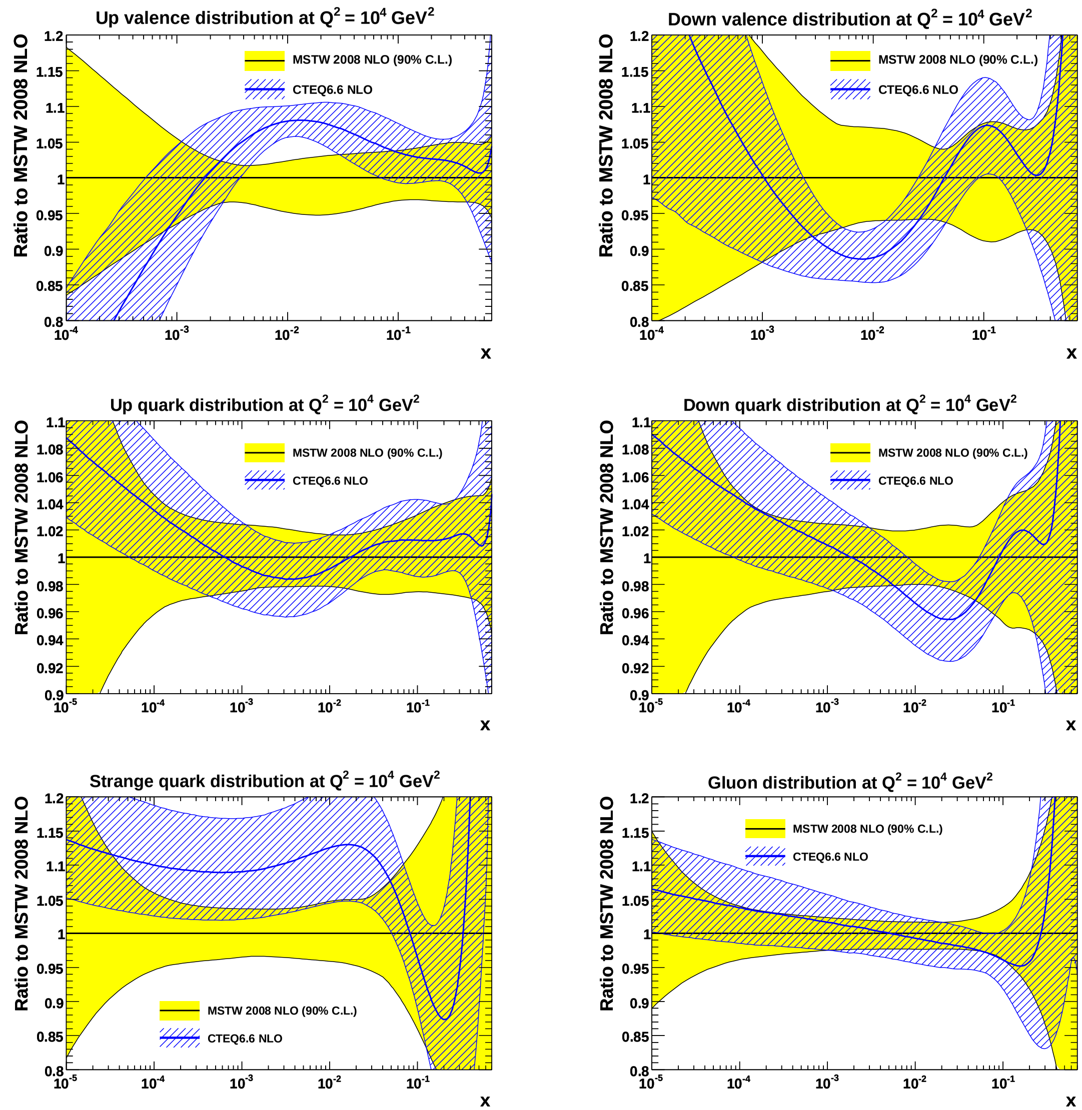
Uncertainties due to $\alpha_S$ variation to be studied (not accounted for by anyone in uncertainty sets).



Comparison to CTEQ6.6 at NLO.

CTEQ strange now larger at small-$x$.

CTEQ gluon larger at high-$x$ (Run I jets) and small-$x$ (input parameterization). Feeds into quark shape.



Comparison of uncertainties to NLO Neural Network PDF set.

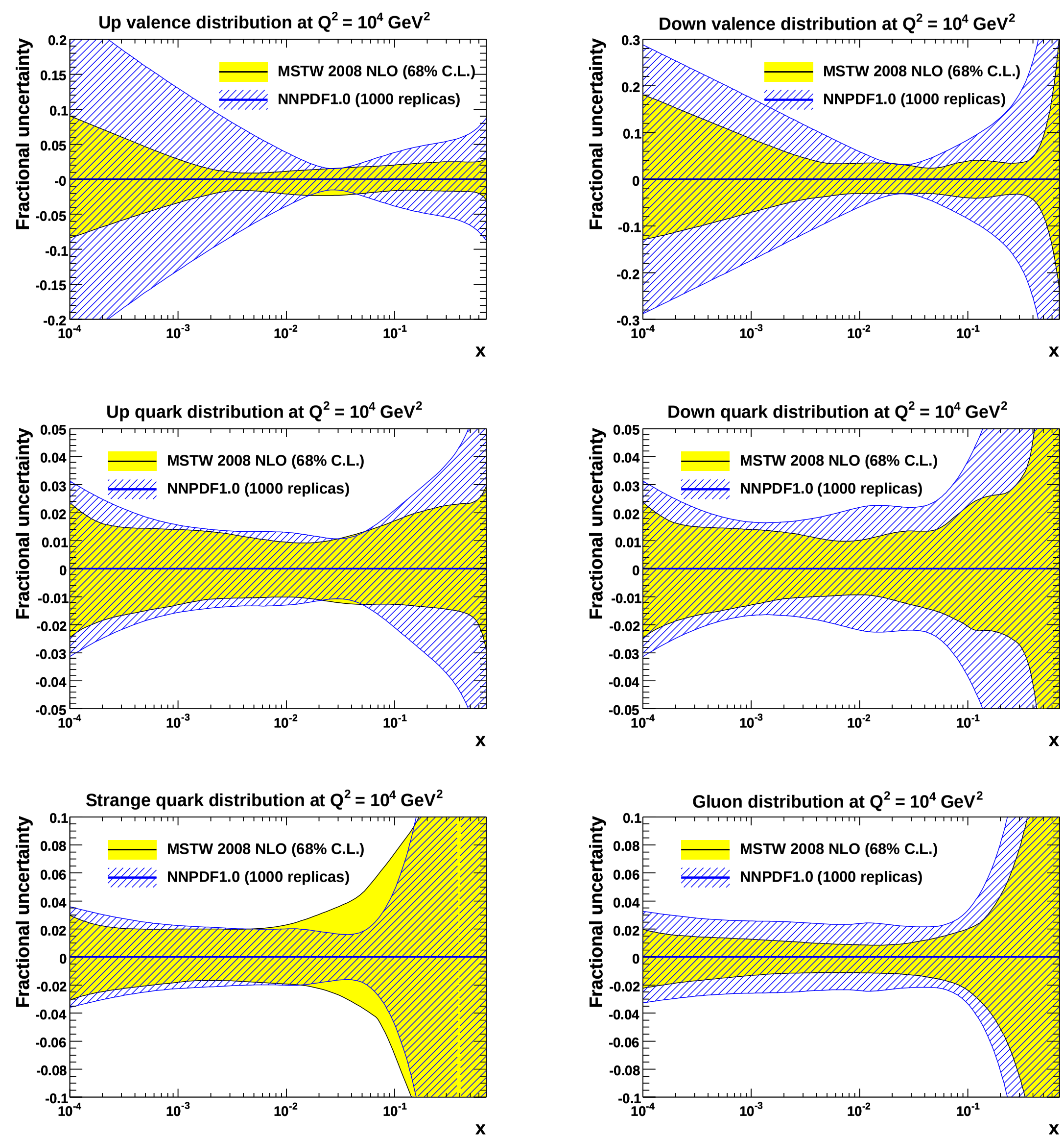
Often comparable despite input flexibility in NNPDF.

Remember they use only DIS data.

Extra constraints from Tevatron W,Z and high-$p_T$ jet data, Drell-Yan data *etc.*

Particularly important for quark flavour decomposition and impacts on gluon.

Currently differ somewhat in some central values.



# Kinematics at the LHC

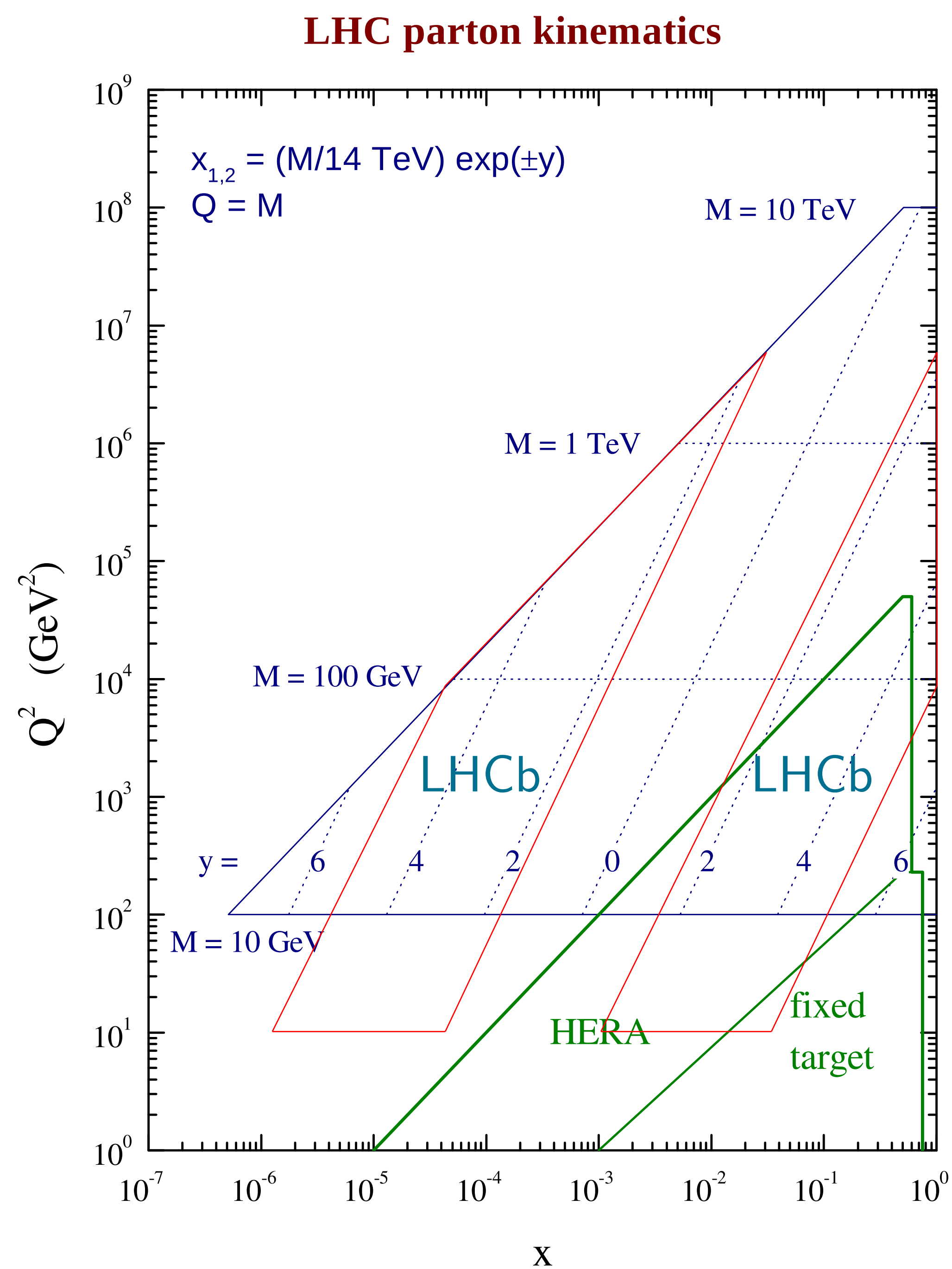
Overall – new kinematic regime.

High scale and small-$x$ parton distributions are vital for understanding processes at the LHC.

LHCb in particular probes regions of intrinsic (experimental) uncertainty

Possible problems at small $x$ due to extra powers of $\ln(1/x)$.

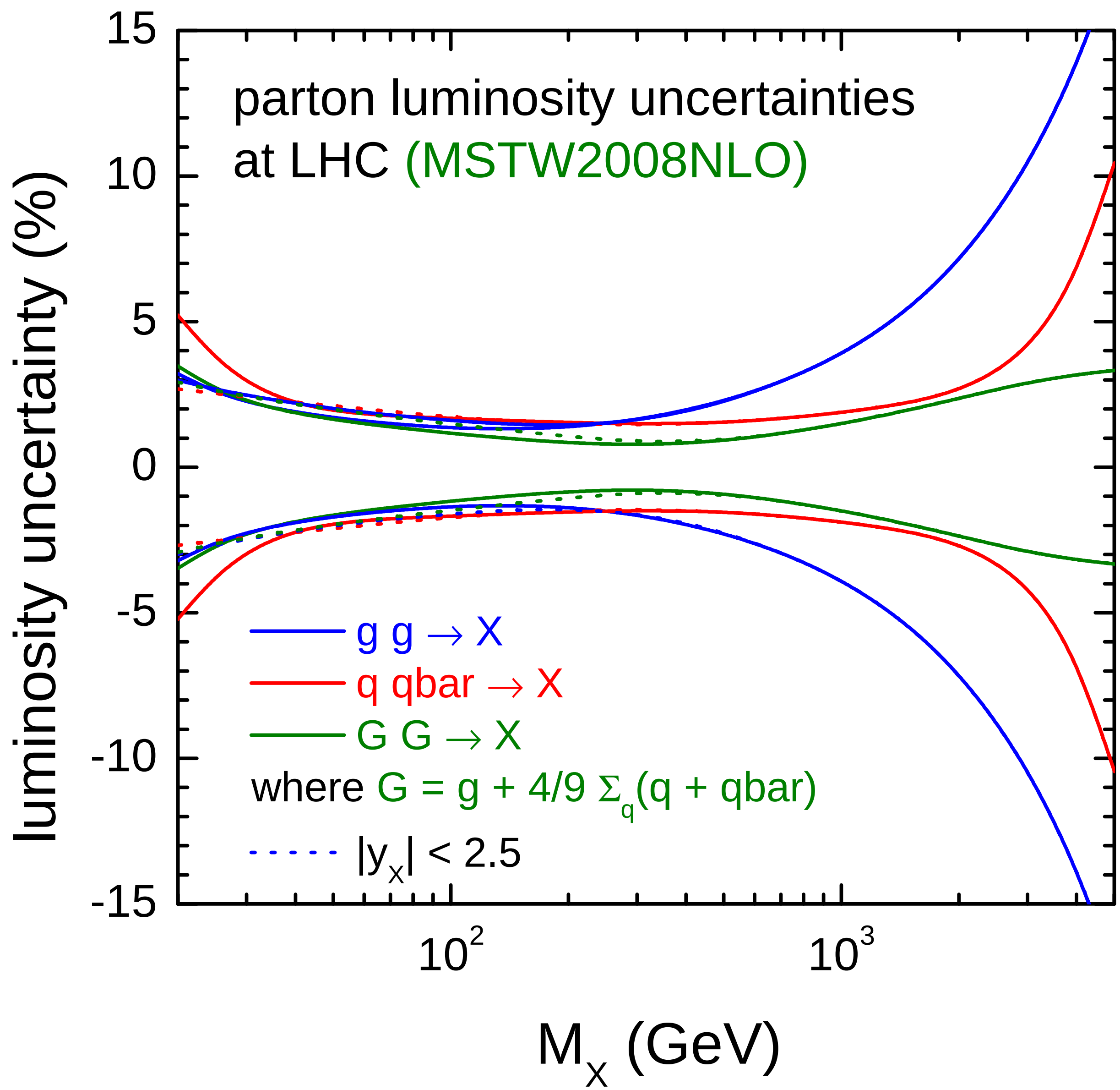
Some predictions potentially unstable, e.g. high-rapidity (all LHCb physics) and/or lowish mass final states at LHC.



# Parton Luminosity Uncertainties

Uncertainties on parton luminosities, i.e. of fundamental rates for creation processes, are optimum for standard model particle production.

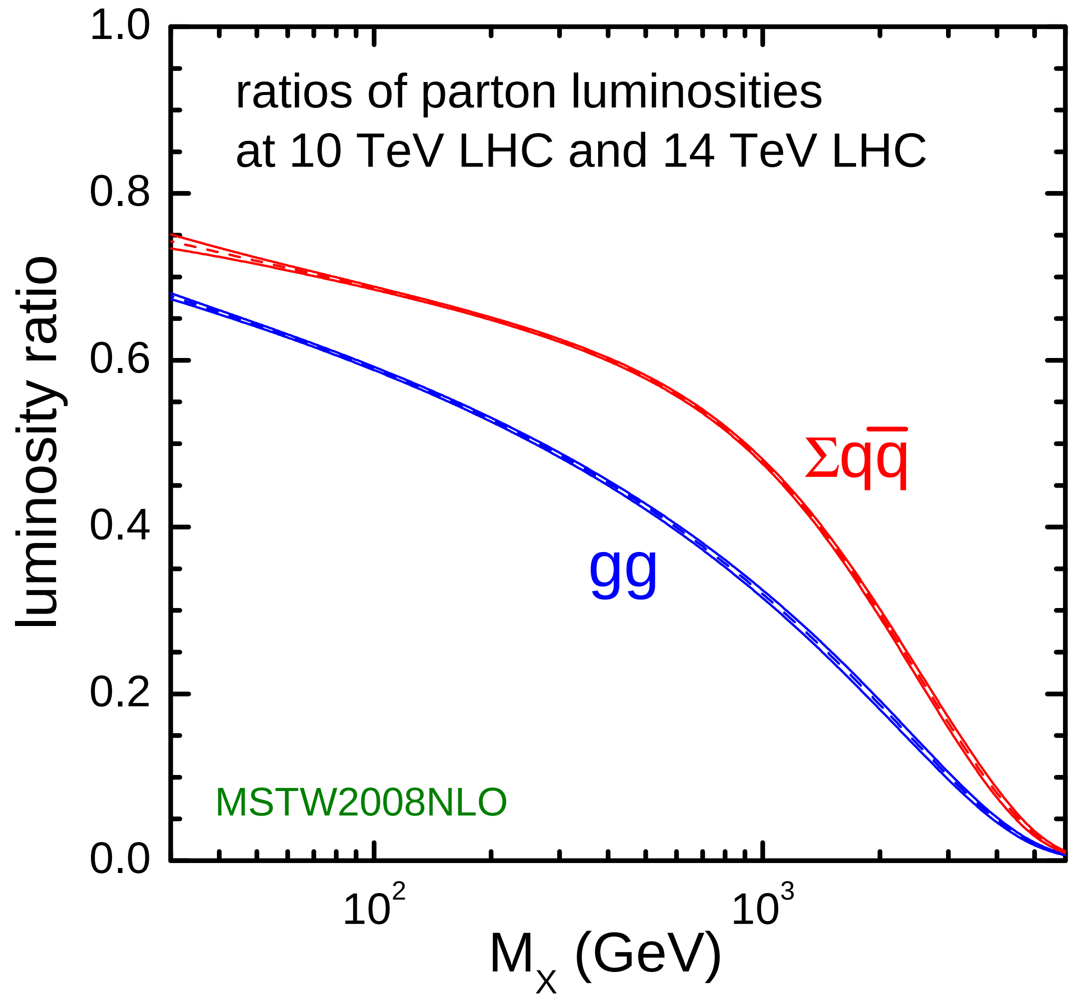
Start to worsen at highest masses where sensitive to large-$x$ PDFs.



# Initial Running

Of course, will be starting the LHC running at 10 TeV rather than the full 14 TeV.

Roughly $60 - 70\%$ the full cross-sections for most standard model (including light Higgs) processes.

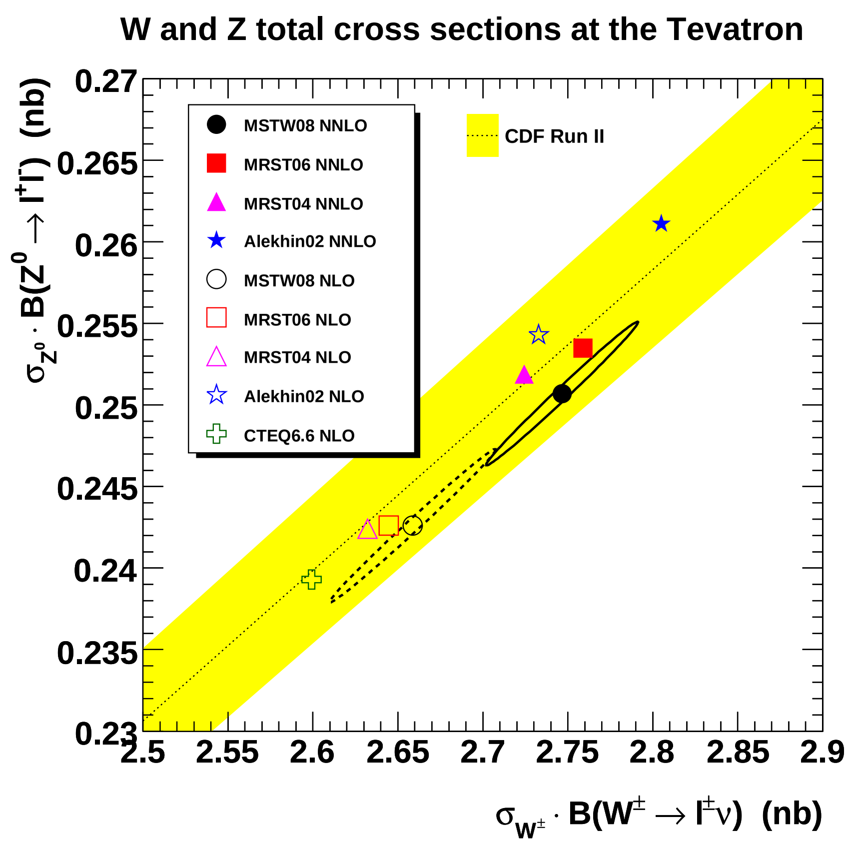


# Predictions – Cross-sections

Predictions for W and Z cross-sections for Tevatron with common fixed order QCD and vector boson width effects, and common branching ratios.

Fairly significant change from NLO to NNLO mainly due to hard cross-section correction.

Other than this reasonable agreement in predictions.

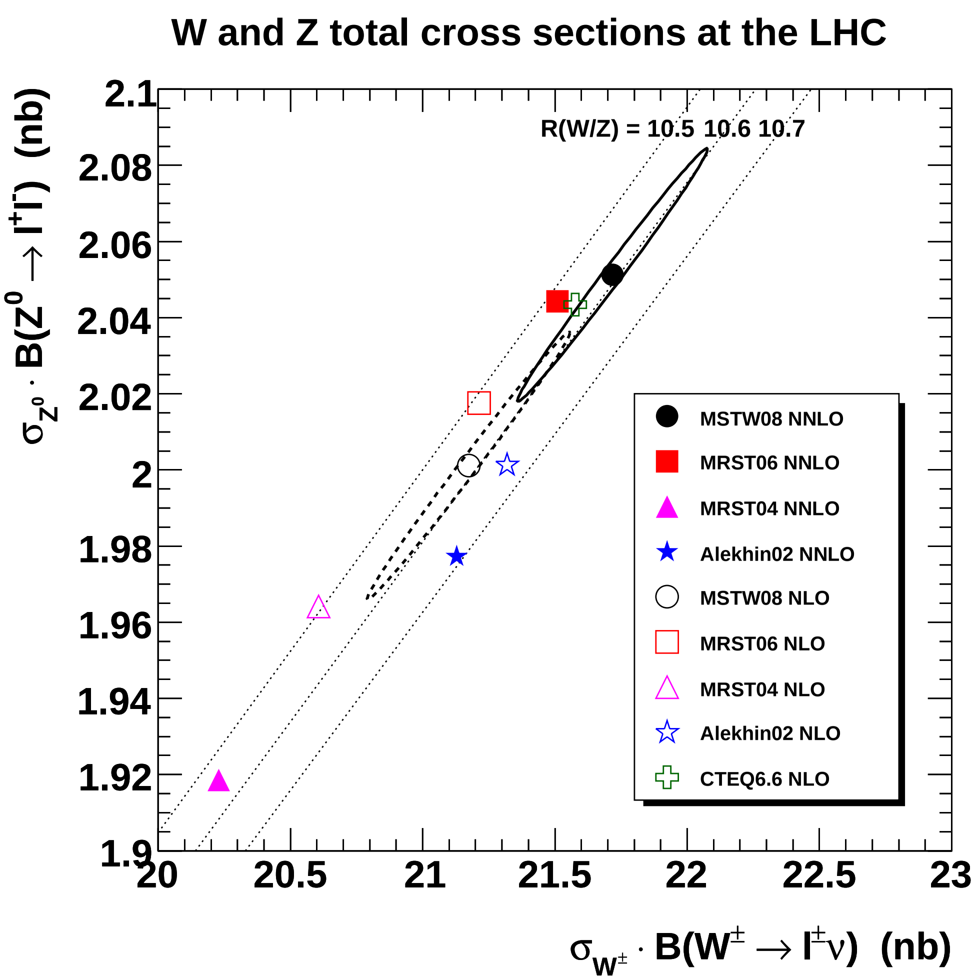


Predictions for W and Z cross-sections for LHC with common fixed order QCD and vector boson width effects, and common branching ratios.

Increases from MRST2006 compared to MRST2004 due to changes due to improved (NLO) or completed (NNLO) heavy flavour prescription.

Virtually no change from MRST2006 → MRST2008. Ratio changes due to change in strange distribution.

Reasonable agreement at NLO with CTEQ6.6, but systematic difference mirrors shape of gluon/quarks.

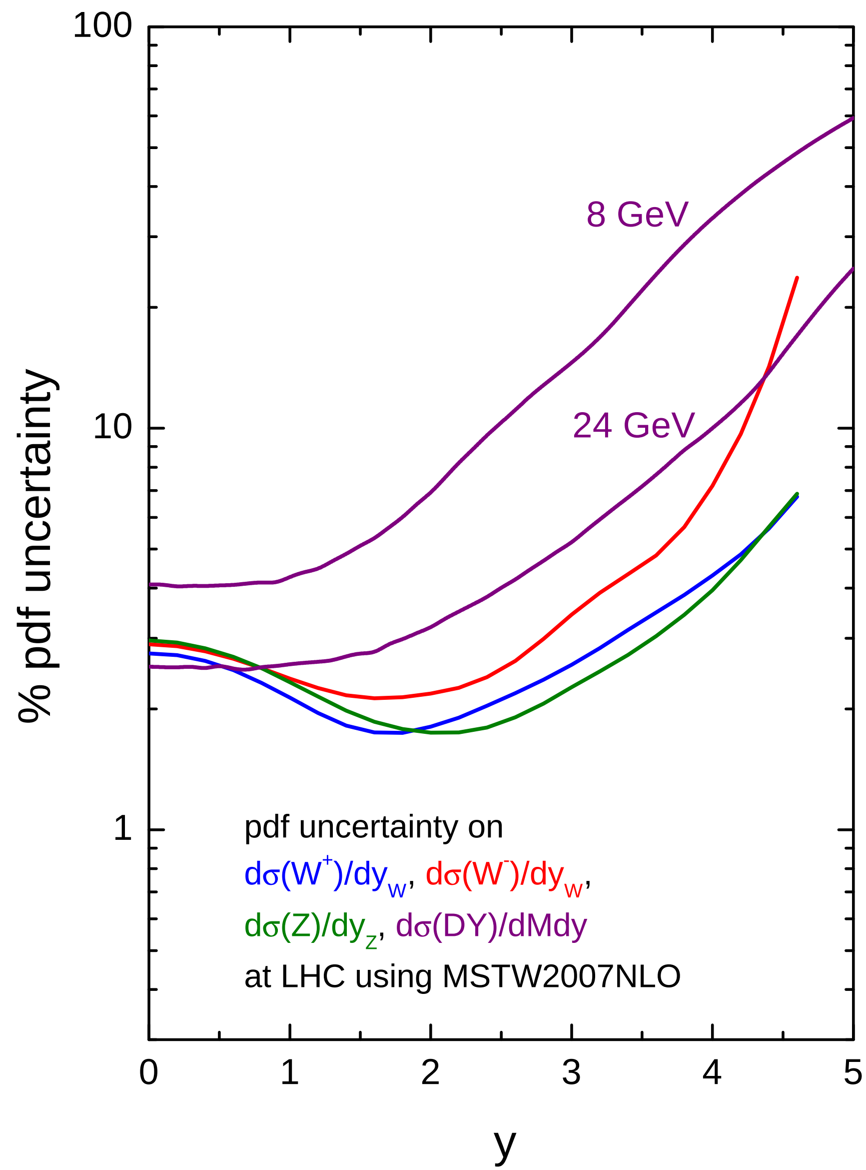


# Uncertainty – more details

Uncertainty on $\sigma(Z)$ and $\sigma(W^+)$ grows at high rapidity.

Uncertainty on $\sigma(W^-)$ grows more quickly at very high $y$ – depends on less well-known down quark.

Uncertainty on $\sigma(\gamma^\star)$ is greatest as $y$ increases. Depends on partons at very small $x$.



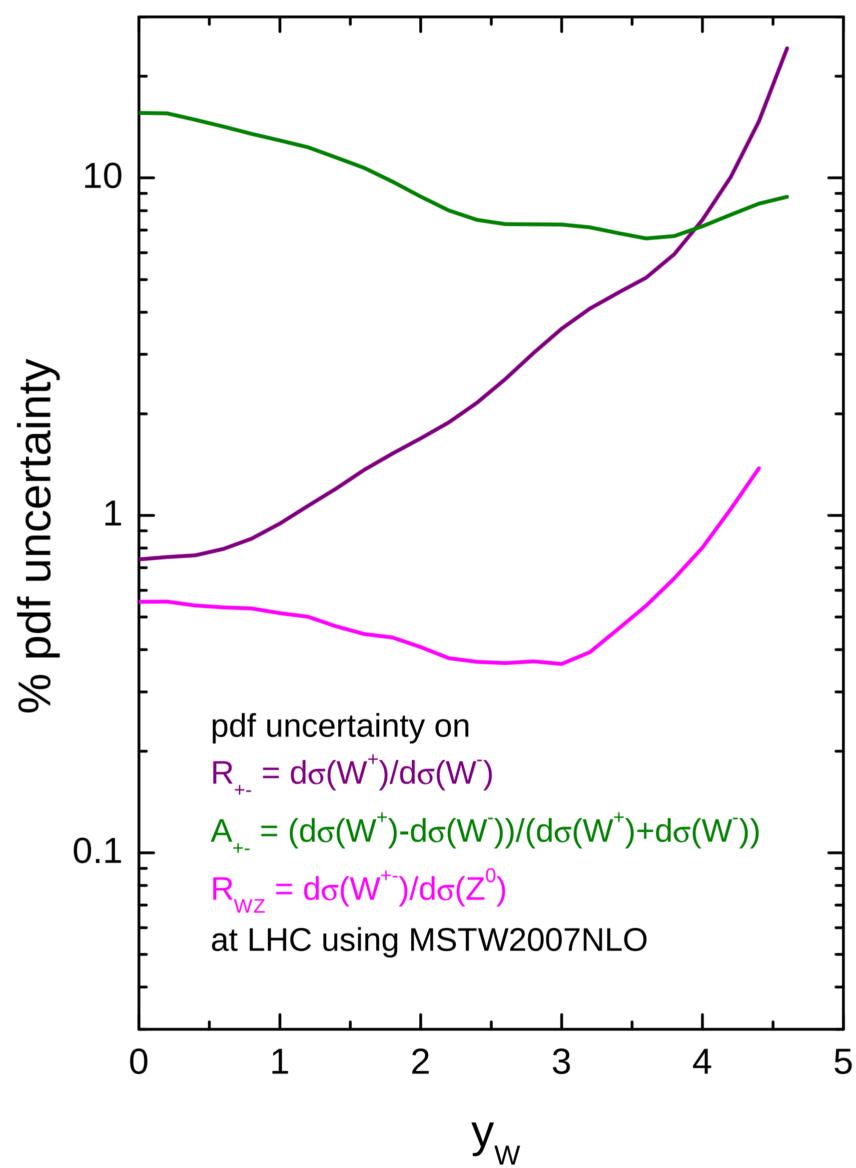
More information from ratios including $\sigma(Z)$, $\sigma(W^-)$ and $\sigma(W^+)$.

Cleaner experimentally.

Uncertainty on $A_W$ large even just from experimental sources.

But $y = 0$ is $x_1 = x_2 = 0.006$ – range of extrapolation of valence quarks. Differences in different PDF extractions.

One of most useful inputs to PDFS with very little data.



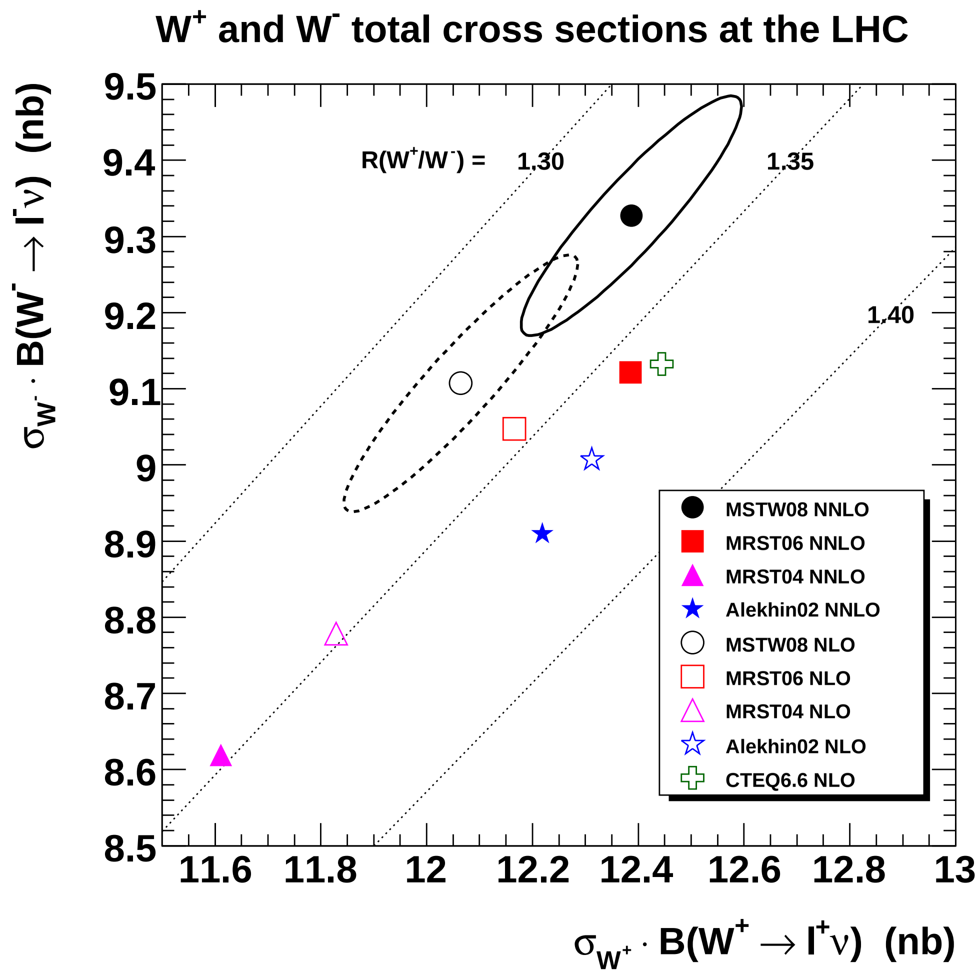

Extremely small uncertainty on ratios $W/Z$ and $W^+/W^-$.

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. $\approx 0.8\%$.

Significantly more difference in this from other PDFs, including MRST.

Again very interesting for early data.



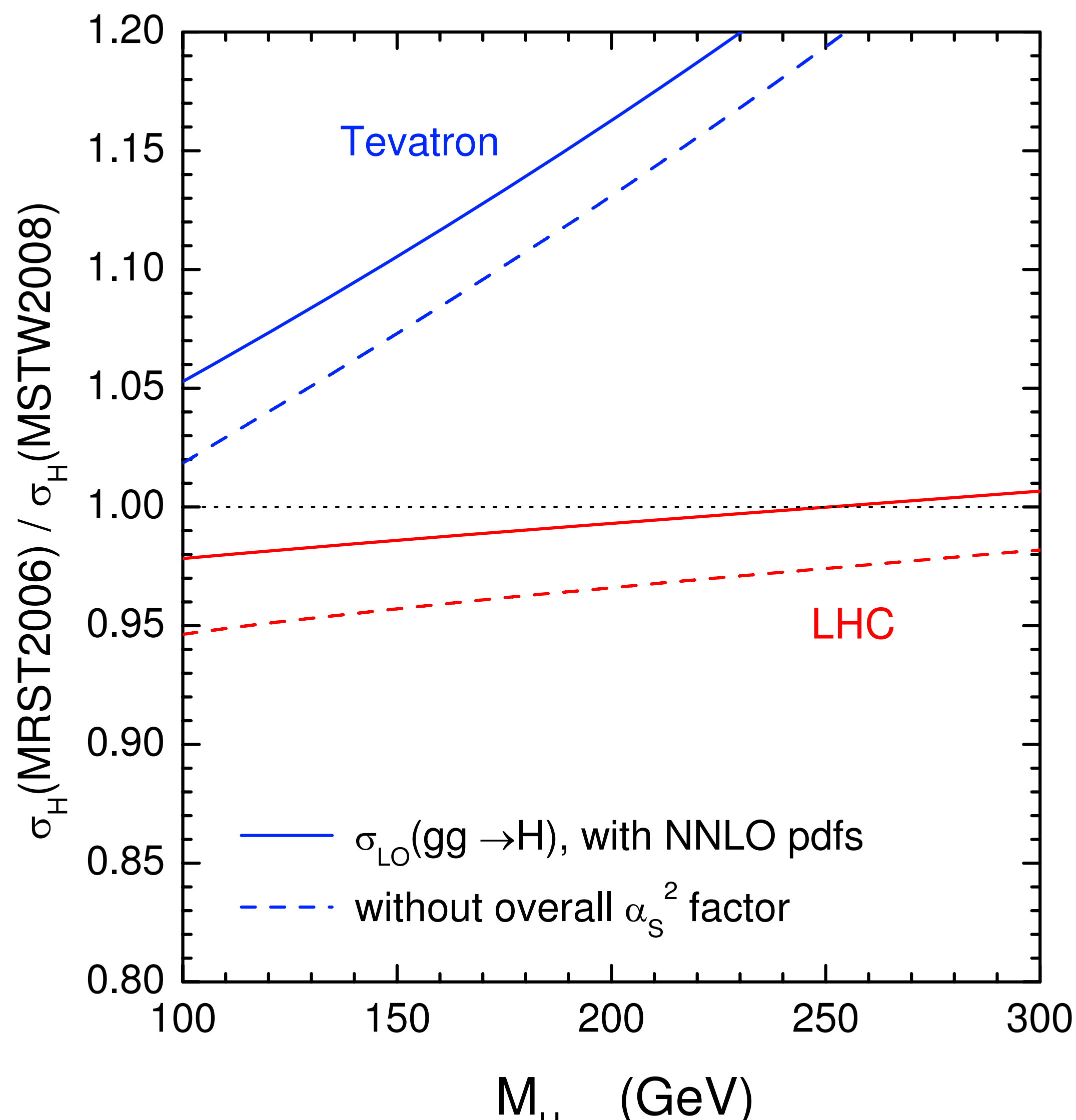
# Higgs predictions

Also consider implications for new physics from updated PDFs.

In MSTW2008 larger small-$x$ gluon, smaller high-$x$ gluon.

$\rightarrow$ decrease in Higgs rate at Tevatron, slight increase at LHC (de Florian *et al*, Anastasiou *et al*).

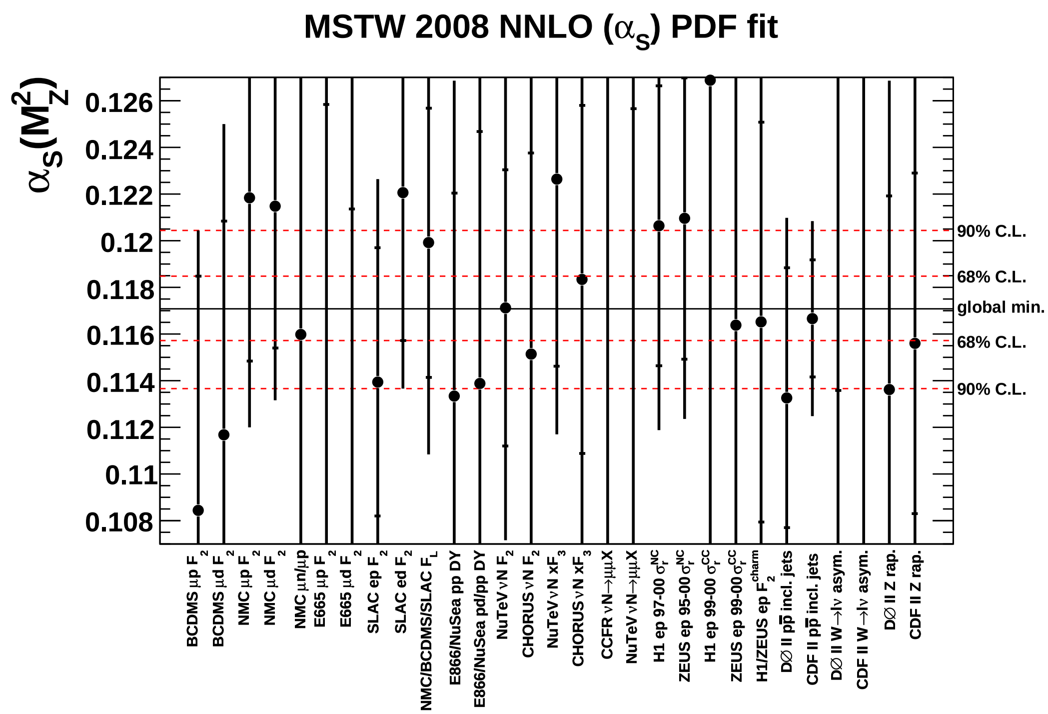
But largely affected by change in central value of coupling – down 2% – anticorrelated with small-$x$ gluon. Increases Tevatron effect, limits LHC effect.

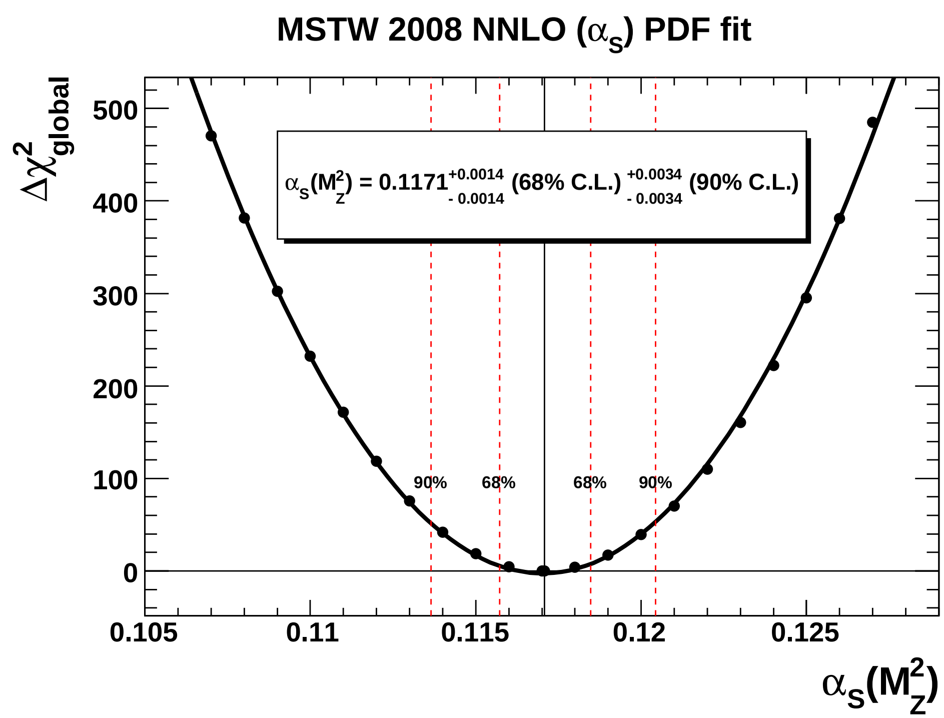



Shows importance of correlations of PDFs and $\alpha_S(M_Z^2)$. Not usually considered.

## Preliminary – Uncertainties on PDFS and $\alpha_S(M_Z^2)$.

Determine uncertainties by scanning over fit quality to different data sets, as with eigenvectors.




Leads to competitive looking uncertainties.

Can also look at PDF changes and uncertainties at limits of $\alpha_S(M_Z^2)$ using same technique.

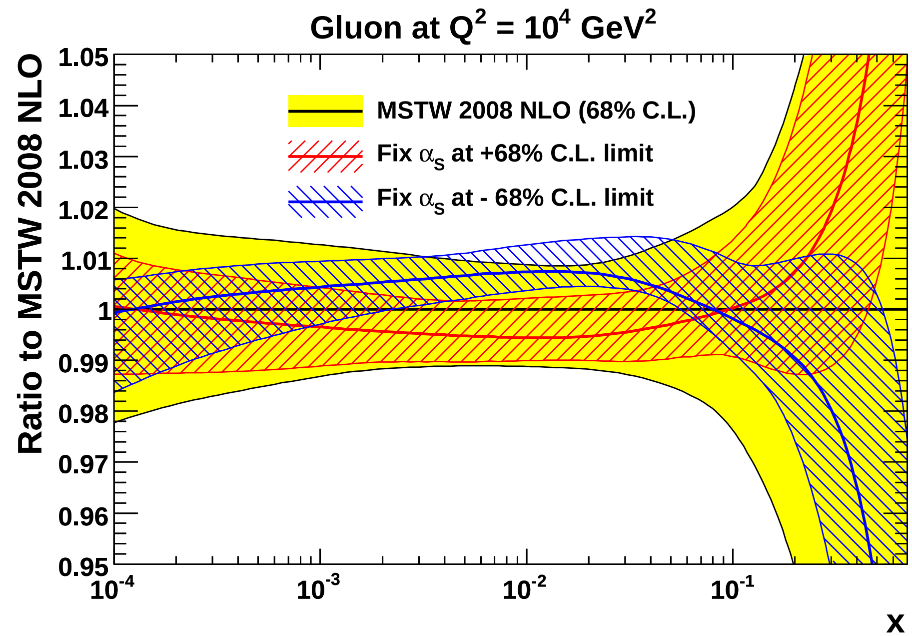
PDF uncertainties reduced since quality of fit already worse than best fit.

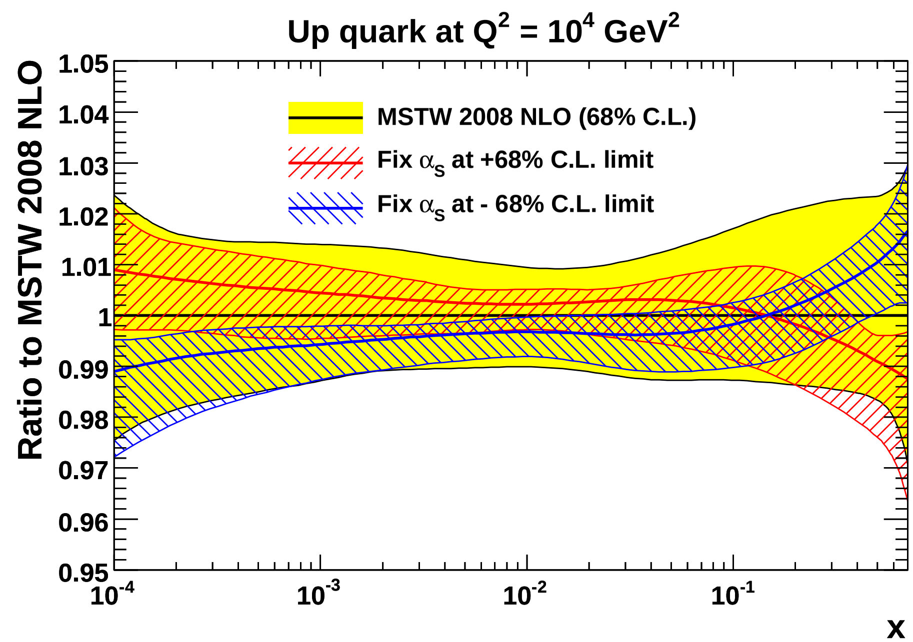
Expected gluon–$\alpha_S(M_Z^2)$ small–$x$ anti-correlation. Variation of gluon largely within central uncertainty band.

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.

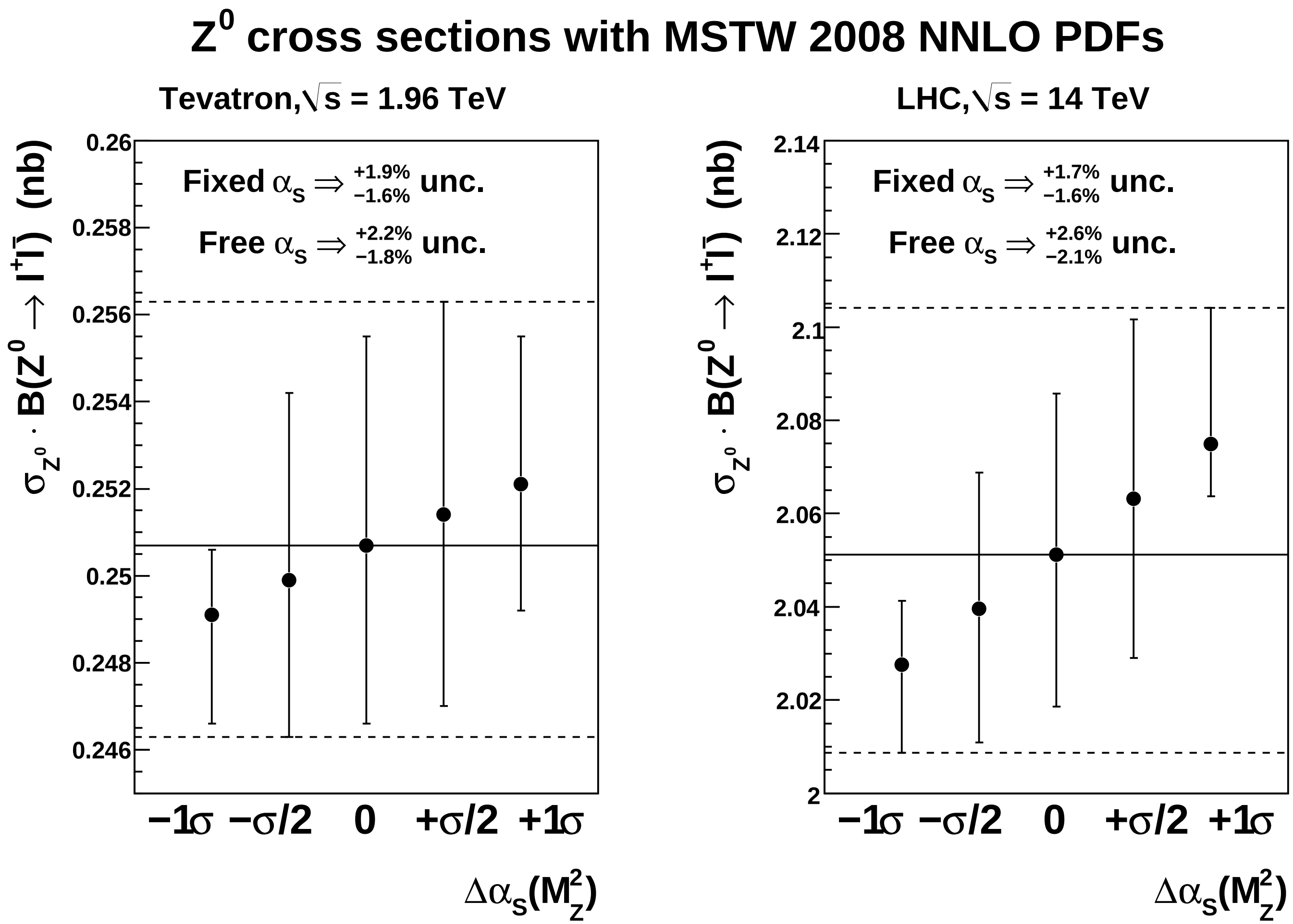
Changes within, or close to within, initial uncertainties.

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



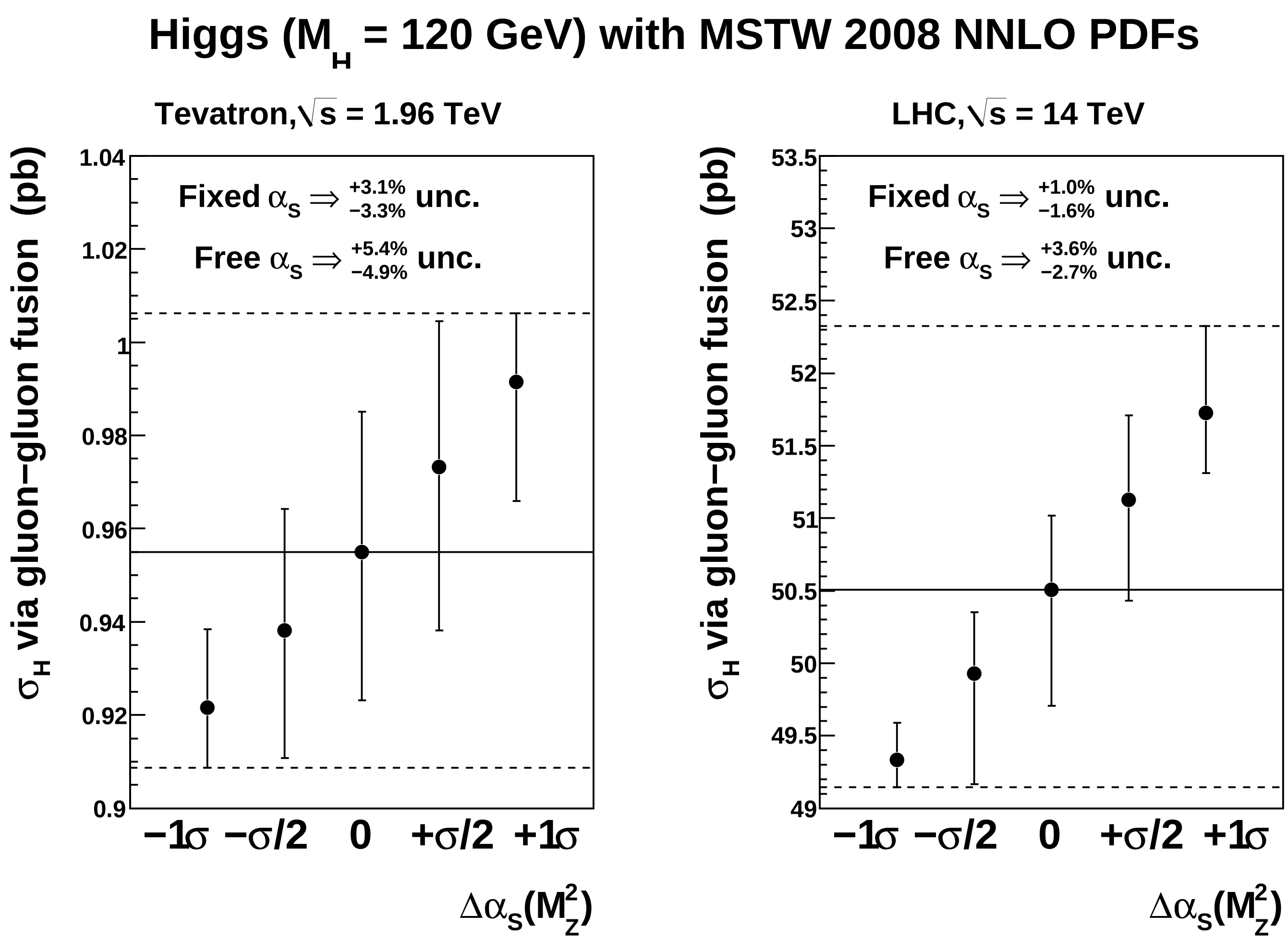



Predictions at NNLO for $Z$ production for different allowed $\alpha_S(M_Z^2)$ values and their uncertainties.



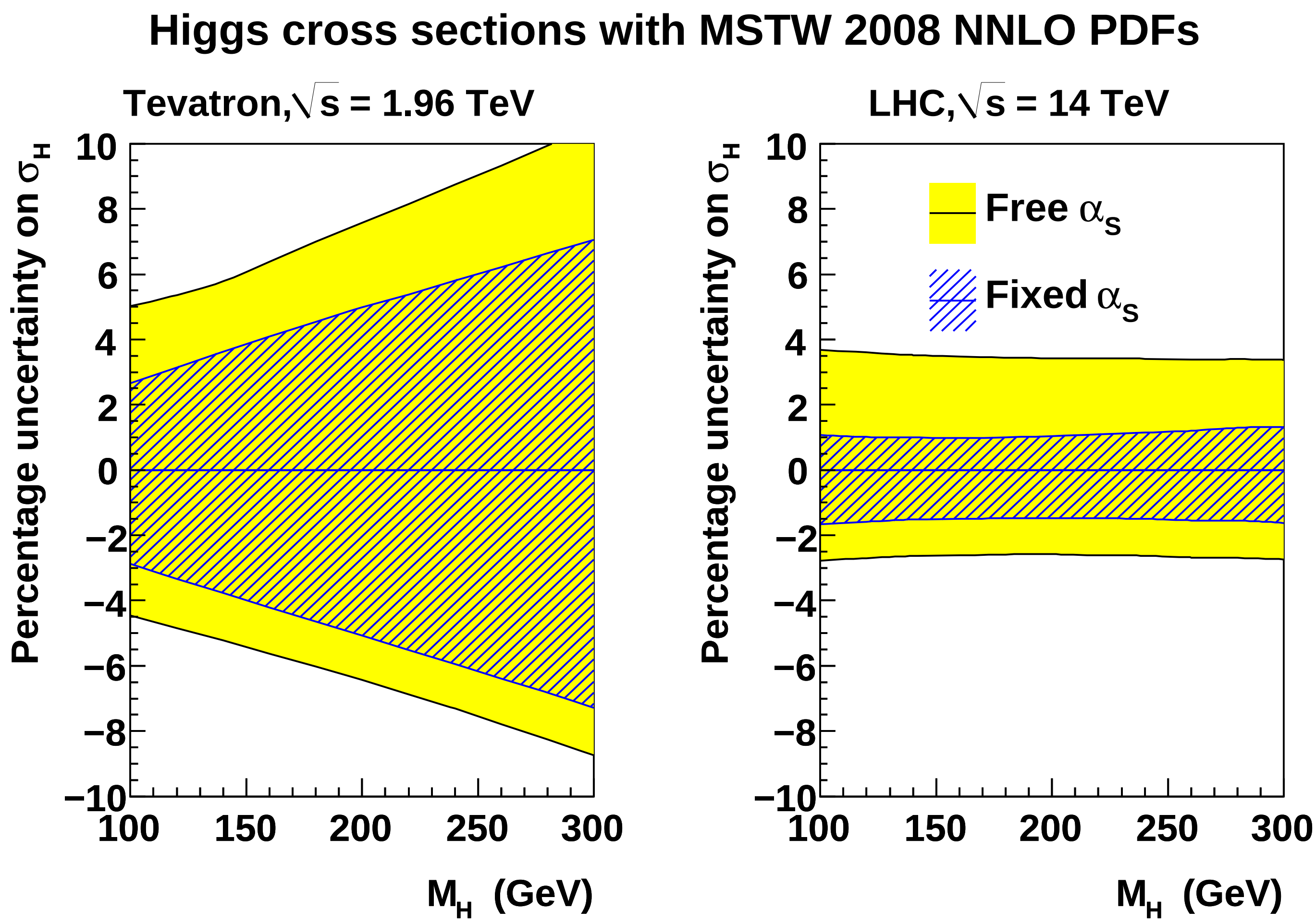

Total uncertainty envelope of set of uncertainties. Increases by up to $50\%$ at LHC. Largely due to effect of PDFs.

Predictions at NNLO 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.

Can extend study to a range of Higgs masses. See same features, i.e. intrinsic PDF uncertainty most important at Tevatron and increases with mass (high-$x$ gluon). Roughly constant at LHC with coupling effect large.




Consistently more PDF-$\alpha_S$ anti-correlation in downwards direction.

## Conclusions

New algorithm for uncertainties. No fixed $\Delta\chi^2$ – on average $\Delta\chi^2 \sim 40$ for $90\%$ confidence limit $\Delta\chi^2 \sim 16$ for 1-$\sigma$. Additional normalization uncertainty $\sim 1\%$.

Inclusion of a lot of new data. Dimuon data fitted directly. Important constraint on strange. New uncertainties on $s + \bar{s}$ feed into other partons.

Tevatron $W, Z$ data important constraint on quarks – constraining for $d_V$ and to some extent $\bar{d}$. Slightly different shape for $d_V(x, Q^2)$.

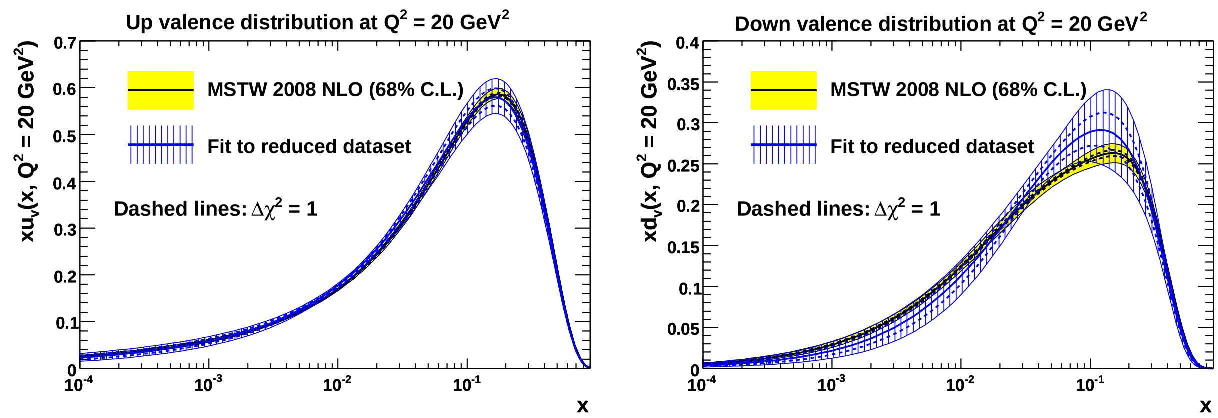
HERA and Tevatron jets now fit using fastNLO. New run II CDF and D0 jet data included. Smaller high-$x$ gluon, larger small-$x$ gluon – impacts on Higgs predictions.

Change in best fit values of $\alpha_S(M_Z^2)$. At NLO $0.120$ compared to $0.121$ in unofficial MRST06 set. At NNLO $0.117$ compared to $0.119$ in MRST06 set. Uncertainty approach now being applied with additional $\alpha_S(M_Z^2)$ variation.

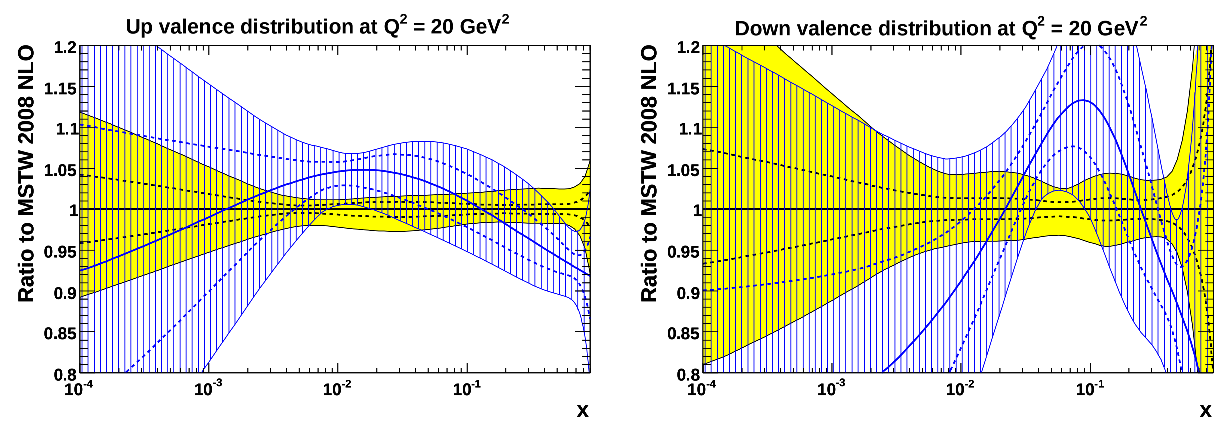
Have full updated NLO and NNLO partons complete with experimental uncertainties. Very precise predictions for $W$ and $Z$ results at the LHC. Total cross-sections and central rapidity $1-2\%$, though asymmetries can add constraints to PDFs.

Full study of correlated PDF and $\alpha_S$ uncertainty leads to uncertainties at LHC at least about $1.3$, and up to $3-4$ times bigger. Correlations vitally important.

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.

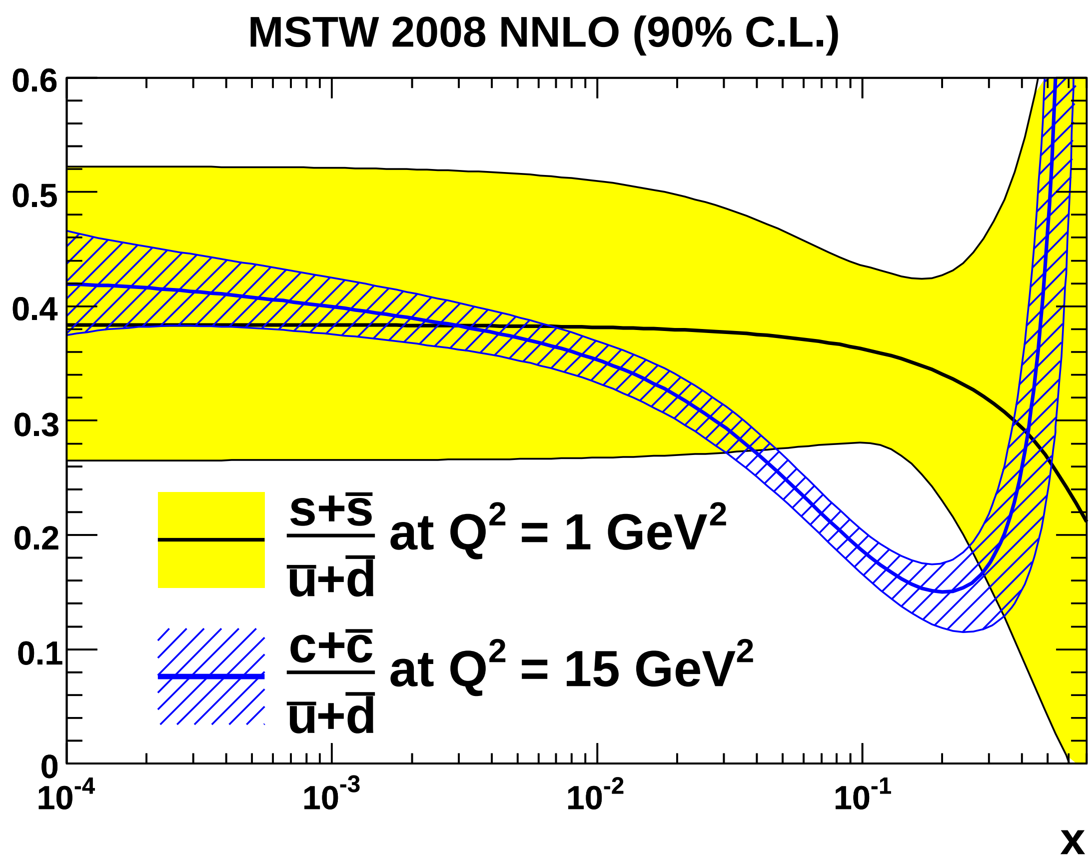



Latter have greater uncertainty. Compatibility using *dynamical tolerance* uncertainty approach, not using $\Delta\chi^2 = 1$.

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

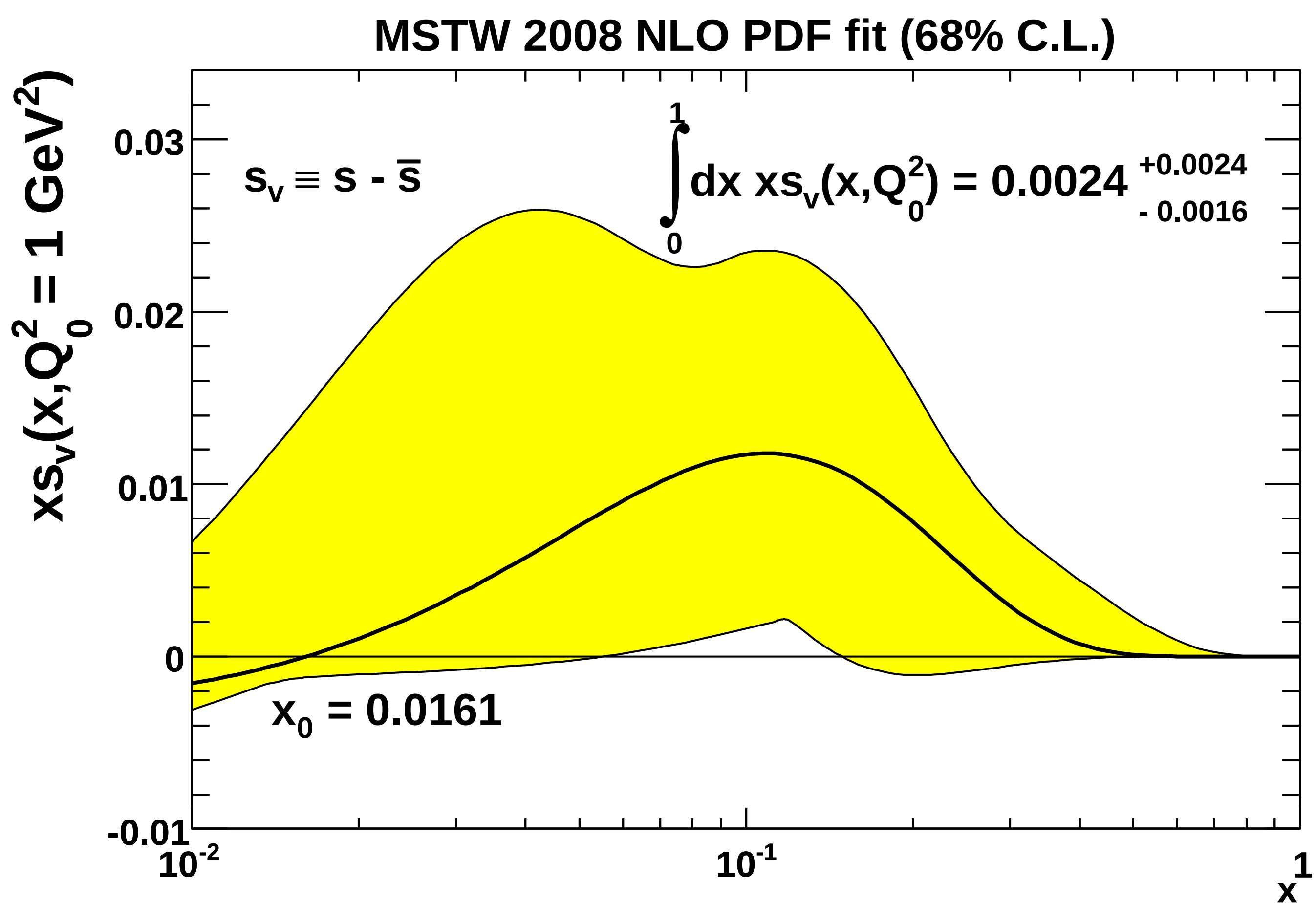
Strange itself has some non-insignificant mass, and this should qualitatively lead to suppression compared to light sea quarks up and down.

When $c$ and $b$ turn on they evolve like massless quarks, but always lag behind. $\rightarrow$ some suppression at all $x$ for finite $Q^2$.

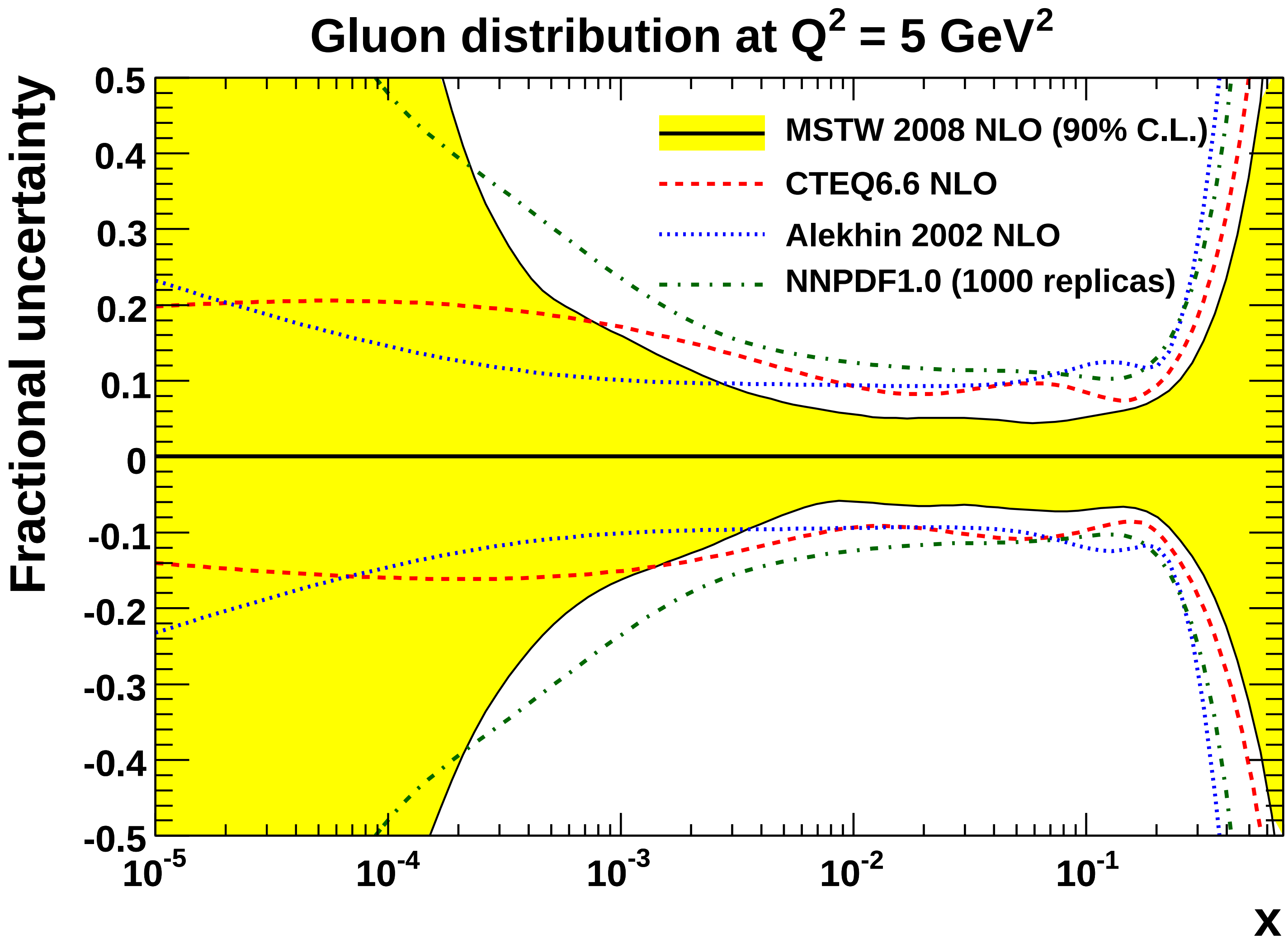



$c + \bar{c}$ evolved through $\sim 7 - 8$ times input scale similar to $s + \bar{s}$ at $Q^2 = 1\mathrm{GeV}^2$. Do not expect exact correspondence, but very good except $c + \bar{c}$ more suppressed at $x \sim 0.1$. (Implication for $s + \bar{s}$ from recent HERMES $K^{\pm}$ data).

Some evidence for positive strangeness momentum asymmetry, but only just more than one-sigma.

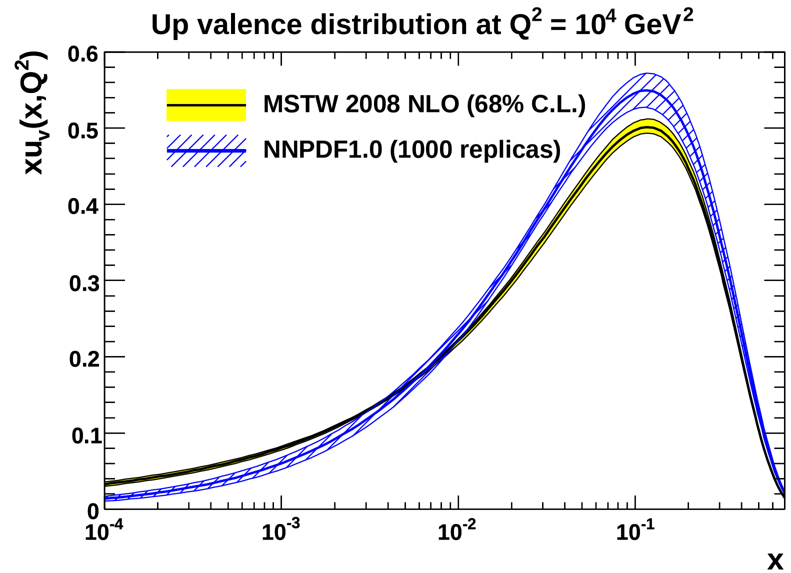


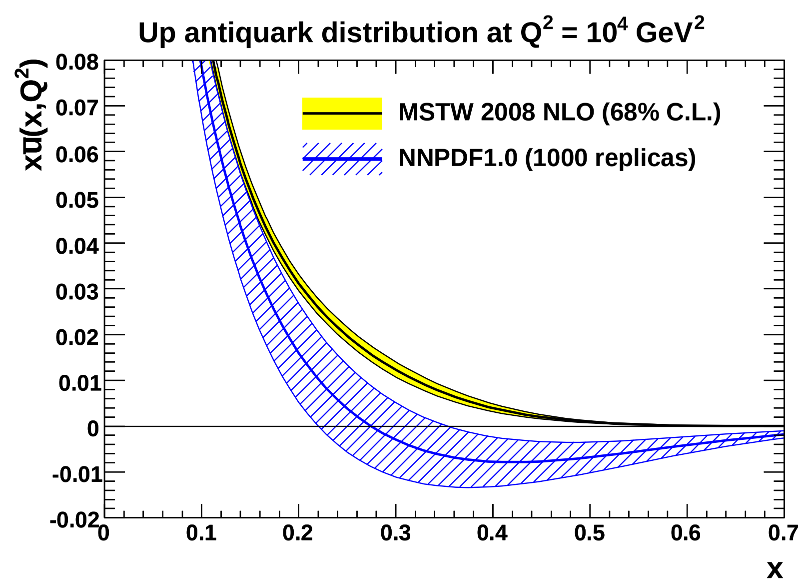
Note that different parameterisations lead to very different types of uncertainty, particularly on small $x$ gluon.



Note that MSTW2008 and NNPDF sets sometimes differ significantly in central values though.

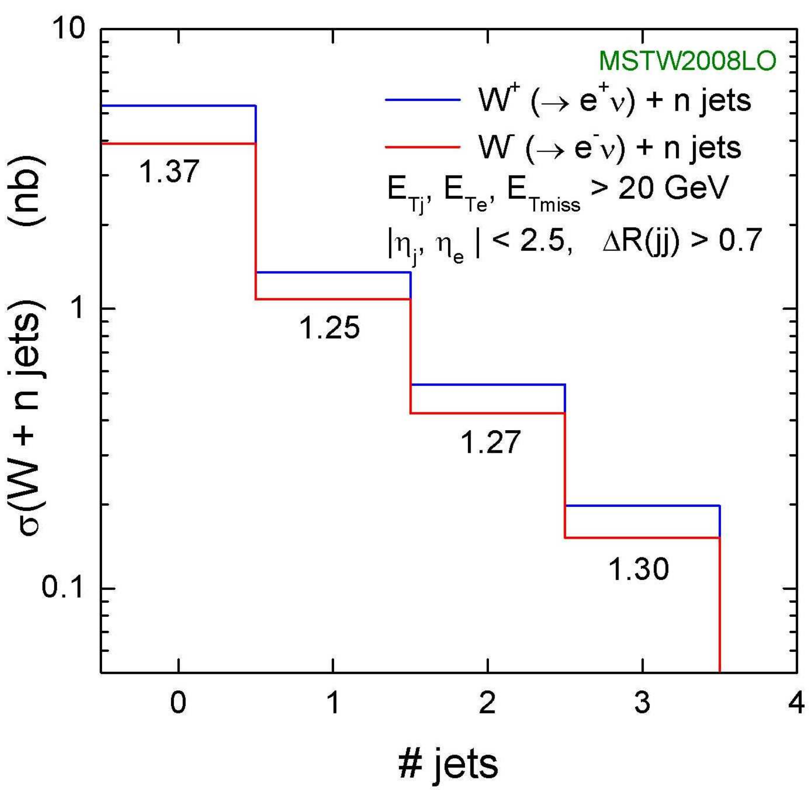
Clearly seen for up valence and anti-up distributions.

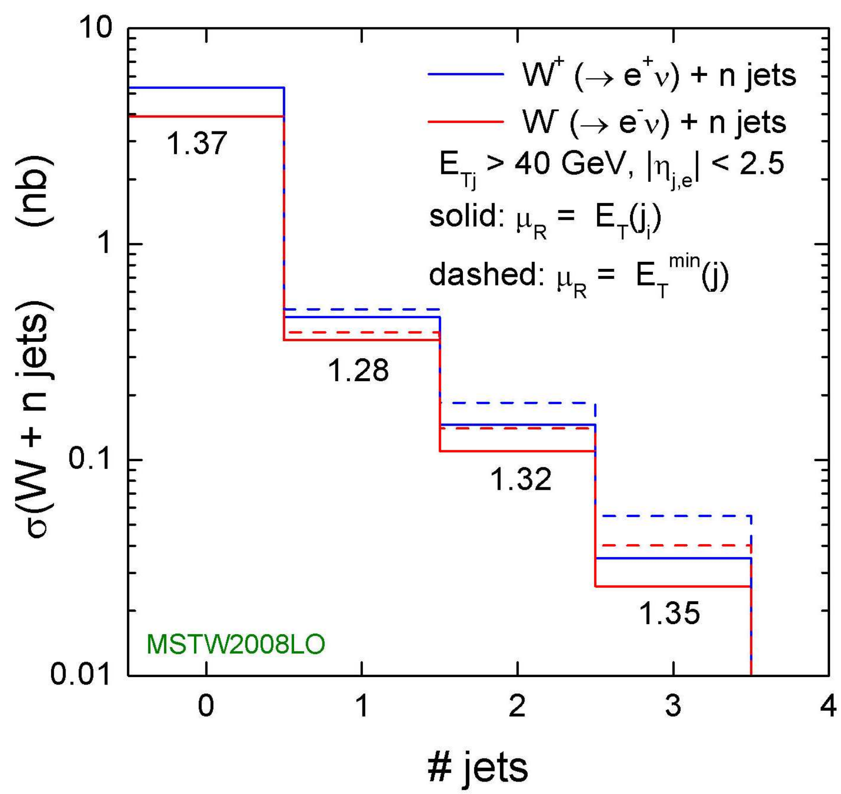


Also information about difference in $u$ and $d$ distributions from ratio of $W^+ + n\,\mathrm{jets}$ to $W^- + n\,\mathrm{jets}$.

Fairly insensitive to scale choice of calculation and some dependence on cuts.



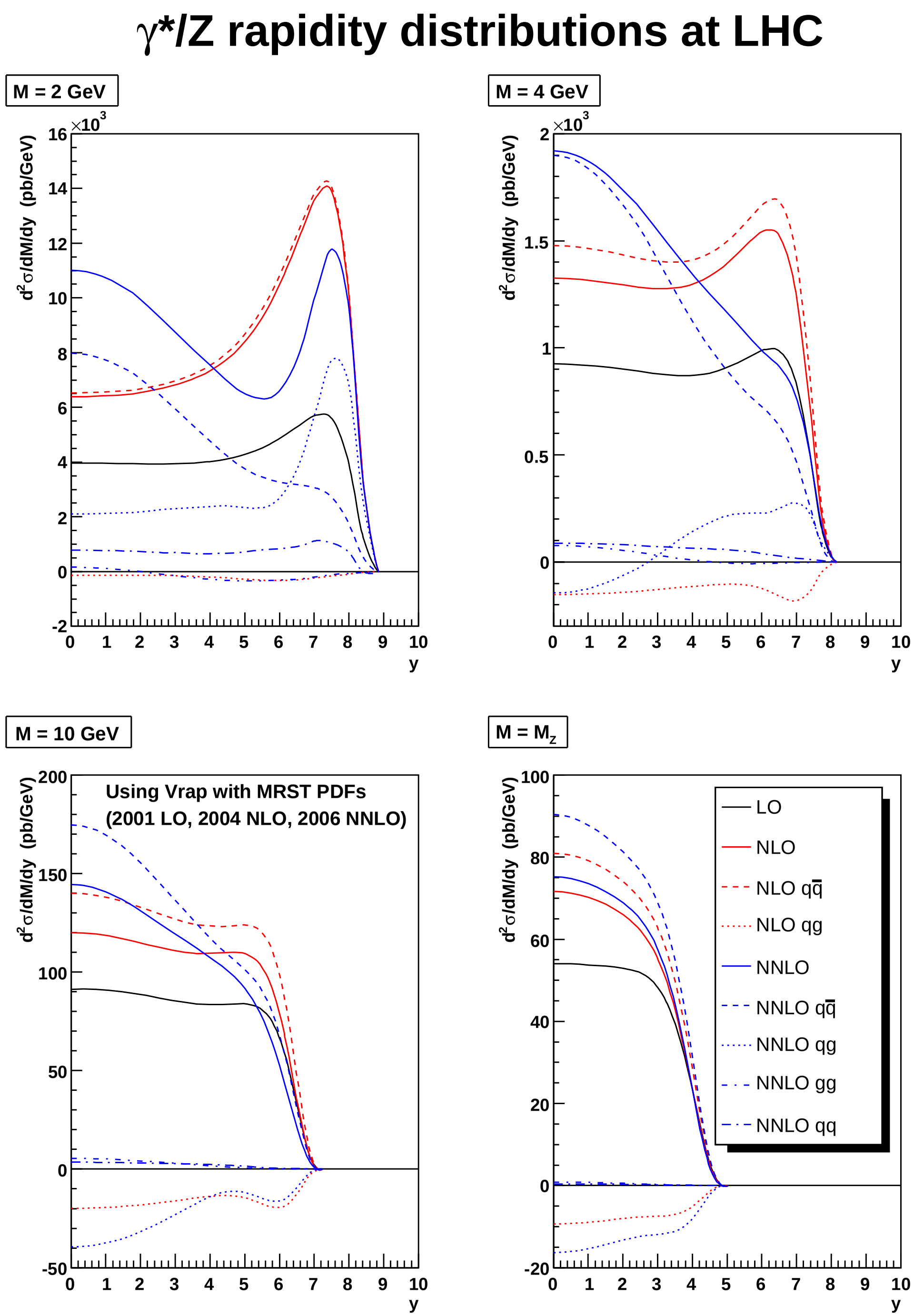

Absolute rates give information on gluon as well, but not so easy to understand in earliest data.

# Perturbative Stability at the LHC

Now have QCD calculations at LO, NLO and NNLO in the coupling constant $\alpha_S$ for $Z, W$ and $\gamma^\star$ production (Anastasiou, Dixon, Melnikov, Petriello).

Good stability in predictions for e.g. $Z$ and $\gamma^\star$ cross-sections for very high virtuality.

Becomes worse at lower scales where $\alpha_S$ larger and large $\ln(s/M^2)$ terms appear in expansion (equivalent to $\ln(1/x)$) and large threshold $\ln(1-x)$ terms.



Quality of fit to various combinations of jet data by MSTW.

| CDFI [179] (33 pts.) | DØI [176] (90 pts.) | CDFII [54] (76 pts.) | DØII [56] (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 |

Quality of fit to various combinations of jet data by CTEQ with restricted gluon parameterization.

| $\text{CDF}_{\text{I}}$ (33 pts) | | $\text{D0}_{\text{I}}$ (90 pts) | | $\text{CDF}_{\text{II}}$ (72 pts) | | $\text{D0}_{\text{II}}$ (110 pts) | | $\Delta\chi^2$ |
|---|---|---|---|---|---|---|---|---|
| Wt | $\chi^2$ | Wt | $\chi^2$ | Wt | $\chi^2$ | Wt | $\chi^2$ | non-jet |
| 0 | 55.8 | 0 | 145.9 | 0 | 120.6 | 0 | 155.2 | 2.0 |
| 1 | 53.2 | 1 | 124.0 | 0 | 118.2 | 0 | 148.6 | 7.7 |
| 0 | 58.6 | 0 | 121.3 | 1 | 98.1 | 1 | 137.8 | 16.8 |
| 1 | 54.5 | 1 | 108.8 | 1 | 95.5 | 1 | 134.2 | 25.8 |
| 10 | 54.1 | 10 | 75.7 | 0 | 142.0 | 0 | 152.1 | 184.3 |
| 10 | 51.9 | 10 | 74.0 | 1 | 101.6 | 1 | 134.6 | 185.5 |
| 0 | 67.1 | 0 | 75.3 | 10 | 77.3 | 10 | 126.1 | 114.6 |
| 1 | 60.3 | 1 | 74.1 | 10 | 77.1 | 10 | 125.8 | 119.3 |
| 10 | 51.7 | 10 | 64.5 | 10 | 76.2 | 10 | 126.1 | 204.3 |