Thanks to Alan Martin, James Stirling and Graeme Watt

University College London

July 13th, 2010

Robert Thorne

Parton Distributions, Functions and Uncertainties
Strong forces make it difficult to perform analytic calculations of scattering processes involving hadronic particles. The weakening of $\mathcal{O}(\alpha^2)$ at higher scales involves nonperturbative parton distributions and perturbative coefficient functions. Calculating these nonperturbative parts can be challenging. Hadron scattering with an electron factorizes via the Factorization Theorem: Hadron scattering with an electron factorizes.
The coefficient functions \( C_{ij}(x) \) are process-dependent (new physics) but are calculable as a power-series in \( s(Q^2) \). Since the parton distributions \( f_i(x;Q^2;\alpha_s(Q^2)) \) are process-independent, i.e. universal, and evolution with scale evolution is calculable once they have been measured at one experiment, one can predict many other scattering processes. Once they have been measured at one experiment, one can predict many other scattering processes.
General procedure.

Start part one evolution at low scale \( Q^2 \lesssim 1 \text{GeV} \). In principle 11 different partons to consider. 

\( u; \bar{u}; d; \bar{d}; s; \bar{s}; c; \bar{c}; b; \bar{b}; g \)

\( m_c; m_b \approx Q_{\text{CD}} \ll m_W, m_t \)

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

\[ x(f_x(\mu; Q^2) = x(1 - x) \cdot \left( 1 + \alpha x \right)^5 \]

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

\[ ((\hat{Q} \Delta)_{sQ}, \hat{Q} x f) \otimes ((\hat{Q} \Delta)_{sQ}, x f) \propto \frac{\hat{Q} \Delta \mathcal{H} p}{((\hat{Q} \Delta)_{sQ}, \hat{Q} x f) f p} \]

Independent combinations, or hence assume \( s = s \) starting not to.

Consider "\( n - p s + s \)"\( (s + p + n) \hat{z} = \text{sea} \)\( p - p = \Lambda p \)\( n - n = \Lambda n \)

In principle II different partons to start parton evolution at low scale.
Fit data for scales above $2 \times 10^{5}\mathrm{GeV}^2$.
This procedure is generally successful and is part of a large-scale, ongoing project.
Excellent predictive power – comparison of MRST prediction for rapidity distribution with preliminary data.
Parton Fits and Uncertainties

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

Problematic due to extreme variations in parameter space.

This is now the most common approach (sometimes offset method).

\[ \chi^2 = \sum \chi \nabla = \chi (H \nabla) \]

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

\[ \chi^2 = \sum \chi \nabla = \chi (H \nabla) \]

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

\[ (\chi^2 - \chi_t) (\chi^2 - \chi_t) \chi_t H \chi_t = \chi \nabla = \chi^2 - \chi \]

Parton Fits and Uncertainties. Two main approaches.
Solved by finding and rescaling eigenvectors of leading to diagonal form $H$. Use more complicated approach – results in $\chi^2 \approx 1$. Other fits less global, keep to 0.5 – 20, for one or more complicated approaches, use $\chi^2 \approx 40$ and MRST/MSTW sometimes conflicting data sets.

Given complication of errors in full fit and question of choosing "correct" $\chi^2$, $\chi^2$ given complication or error in physical quantity then given by

$$\chi^2 = \sum \left( \frac{1}{\sigma_i^2} (F_i(S^+) - F_i(S^-)) \right)^2$$

where $S^+$ and $S^-$ are PDF sets displaced along eigenvector direction.

Implement by CTEQ, then others. Uncertainty on physical quantity then given by $\chi^2$.

CTEQ use $\chi^2 \approx 40$ and MRST/MSTW use complicated approach – results in $\chi^2 \approx 5$ – 20. Question of choosing "correct" $\chi^2$.
The inappropriateness of using $\chi^2 = 1$ for individual data sets as obtained by CTET using Lagrange Multiplier technique is shown by examining the best value of $\sigma_W$ and its uncertainty $\sigma_W$ (Tevatron) for conflicting data sets is sometimes large number of sometimes inappropriate of using $\chi^2 = 1$ when including a large number of sometimes.
Fairly consistent.

Errors: (other possible reasons) and similar some difference in central values

\[ \chi^2 \Delta \]

\[ \chi^2 \Delta = 1 \]

\( \chi^2 \Delta \) with simple quadratures approach with 1 systematic error (Alekhin). Compare rigorous treatment of

Very conservative fit.

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

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

Errors similar.

Enormous difference in central values.

However, how do partons from very conservative, structure function only MRSTbench data compare to global partons?
Using similar sort of reasoning, MST used $\chi^2 \approx 50$ for 90% confidence level on partons. Still same basic idea but more sophisticated.

Previous reasoning, allow some sets somewhat outside 90% confidence limits for $\chi^2$. Some sets within its 90% confidence limit compared to the best global fit. Previous reasoning allow a value such that every data set remains roughly
Similarly for the 68% C.L. region, the most probable value of the $\chi^2$-distribution is the 90th percentile of the $\chi^2$-distribution with $n$ d.o.f., i.e.,

$$\frac{\left(\frac{\chi}{N}\right)^2}{\chi^2 - \theta^{n-1}} \sim \left(\frac{\chi}{N}\right)^2$$

where the probability density function is

$$\int_{\chi_0^{50}}^{\chi_0} \frac{d\chi}{\chi^2 - \theta} = (\chi_0 \chi_{\chi^2})_{\chi = \chi_0}^0$$

$\chi_0^{50}$ is the 90th percentile of the $\chi^2$-distribution with $n$ d.o.f., i.e.,

$$\left(\frac{\chi}{N}\right)^2 \geq \frac{\chi_0^{50}}{0.05}$$

Define 90% C.L. region for each data set $n$ with $N$ data points (Watt DIS08)
Eigenvector number 13

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

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

MSTW 2008 NLO PDF fit
Plot this for all datasets for a given eigenvector.

Constrained in one direction by Drell-Yan data and in the other direction by NuTeV data. In this case the best fits for the two sets are highly inconsistent.

$\chi^2 = 100$ well outside 90% C.L. in this case. The best fits for the two sets are

Eigenvector 13 constrained in one direction by E866 Drell-Yan data and in the other direction by NuTeV data.

Eigenvector number 13

Global 2008 NLO PDF fit

Distance $= \sqrt{\Delta \chi^2}$
The eigenvector contributes most to the high-$x$ sea quark uncertainty, but also a variety of other quarks. This eigenvector originates from eigenvector number 13.

$Q^2 = 1 \text{ GeV}^2$

At input scale

Fractional contribution to uncertainty from eigenvector number 13

MSTW 2008 NLO PDF fit (68% C.L.)
As a simpler example, eigenvector 9 is constrained most by ZEUS data on \( \nu F^p \mu \). 90\% confidence limit determining by ZEUS in up direction and H1 in down direction. Both \( \chi^2 \) for 17.

\[
\chi^2 = \sum \frac{(O - E)^2}{E}
\]

H1 ep 97-00 incl. jets
ZEUS ep 96-00 incl. jets
CDF II N MC-II
CDF II Zrap.
CDF II W - H asym.

Distance = \[ \sqrt{\chi^2_{\text{global}}} \]

MSTW 2008 NLO PDF fit

As a simpler example, eigenvector 9 constrained most by H1 and ZEUS data on \( \nu F^p \mu \), 90\% confidence limit determining by ZEUS in up direction and H1 in down direction.
Not surprising this eigenvector contributes most to the gluon uncertainty.
Approach repeated for all 20 eigenvectors to determine uncertainty on each. On average, \( \sum \Delta \chi^2 = 40 \% \) for 90\% and \( \sum \Delta \chi^2 = 15 \) for I - 0, but large variations, and asymmetries.
Eventhoughonedatasetconstrainseacheigenvectorlimit,doesn'tmeanothersdo.
Normalisation Uncertainties

Previously, the normalization of each data set was determined by the best fit and then fixed. This procedure was flawed due to technical difficulties in including this feature in uncertainties.

Now, we implement a procedure of allowing normalizations of all sets to vary in best fit and scan over eigenvectors, with penalty term for each set variance vs LO and slightly at NLO.

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

Rescale errors with normalization to avoid bias.

$\chi^2 = \frac{\sum (N^o - N^p)^2}{N-1}$

In this way, we avoid biasing the results with normalization.

Technical difficulties in including this feature in uncertainties.
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. Inpracticeshouldgiveaconservativeestimationofuncertainties. CaninvestigatebyrepeatingHERA-LHCWorkshopexerciseofobtainingPDFsbyfittingtoDISdatawithconservativecutsonly.

\[ \chi^2 \] for benchmark data 458/589 in reduced fit \( \rightarrow 526/589 \) within global fit. Still lack of compatibility some places, e.g. high-\( x \) gluon. Latter have greater uncertainty. Compatibility using dynamical tolerance uncertainty approach, but not using \( \Delta \chi^2 = 1 \). Comparison of normal and benchmark sets shown. Latter have greater uncertainty.
Groups despite initial tolerance. Only reasonable agreement between CTEQ6.6 and MSTW distributions, along with uncertainty on MSTW

\[ p \]
Different PDF sets

- **CTEQ6.6**
  - Not quite as up-to-date on Tevatron data.
  - Include all above except HERA jet data (not strongest constraint) and heavy flavour structure functions. Include HERA combined data. PDFs at NLO and NNLO.

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

- **NNPDF2.0**
  - Include all above except HERA jet data (not strongest constraint) and HERA combined data. PDFs at NLO and NNLO.

- **MSTW08**
  - Fit all previous types of data. Most up-to-date Tevatron jet data. Not NNPDF2.0.

- **ABKM09**
  - Fit to DIS and fixed target Drell-Yan data. PDFs at NLO and NNLO.

- **CTEQ6.6**
  - Very similar. Not quite as up-to-date on Tevatron data.

- **HERAPDF2.0**
  - Based entirely on HERA inclusive structure functions, neutral and charged current. Use combined data. PDFs at LO, 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 and NNLO.

- **CTEQ6.6**
  - Not quite as up-to-date on Tevatron data. PDFs at NLO and NNLO.

- **HERAPDF2.0**
  - Based entirely on HERA inclusive structure functions, neutral and charged current. Use combined data. PDFs at LO, 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 and NNLO.

- **CTEQ6.6**
  - Very similar. Not quite as up-to-date on Tevatron data.

- **HERAPDF2.0**
  - Based entirely on HERA inclusive structure functions, neutral and charged current. Use combined data. PDFs at LO, 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 and NNLO.
Perhaps surprisingly all get rather similar uncertainties for PDFs cross-sections.

Constraint on input form of PDFs. Impose strong theory additional model and parameterisation uncertainties. Use $\chi^2$ for 20 eigenvectors.

- CTEQ6.6: 9 eigenvectors. Inflated $\chi^2$ of 50 for $\sigma$ and $m_c, m_b$.

Perhappssurprisingly all get rather similar uncertainties for PDFs cross-sections.

Constraint on input form of PDFs. Impose strong theory additional model and parameterisation uncertainties. Use $\chi^2$ for 20 eigenvectors.

- CTEQ6.6: 9 eigenvectors. Inflated $\chi^2$ of 50 for $\sigma$ and $m_c, m_b$.

- HERAPDF2.0: 21 parton parameters. Use $\chi^2$ for 1. Also $\alpha_s, m_c, m_b$.

- MSTW08: 20 parton parameters. Use $\chi^2$ for 20 eigenvectors.

- NNPDF: minimise $\chi^2$ and define eigenvectors of parameter combinations.
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 of the original dataset $F_{\text{exp};(k)}$. Where $r(k)_p$ are random numbers following Gaussian distribution, and $S(k)_p;N$ is the analogous normalization shift of the replica depending on $1 + r(k)_p;\sigma$. Hence, include information about measurements and errors in distribution of $F_{\text{art};(k)}$. by $q_{(\text{net})}(k)_i$. Mean and deviation of observable $O$ then given by $O \pm \sigma$. 

\[
\left( O_n - \left[ \sum_{d \in \text{data}} \frac{N}{I} \right] \right) \sum_{d \in \text{data}} \frac{N}{I} = O \pm \left[ \sum_{d \in \text{data}} \frac{N}{I} \right] \]

Fit to the replicas of the data obtaining a set of PDF replicas $q_{(\text{net})}(k)_i$. (follows Giele et al.)

Replacing fluctuates about central data.
Fluctuations (as far as this is possible). Must guard against this.

Freedom in parameterisation means best fit to all data would tend to reproduce data
Data included consistently increasing. Recently NNPDF2.0, first global fit of this type.

where data constraints vanish.

where $\mu, m$ are in fairly narrow ranges, so overall behaviour guided at these extremes

$$(x) NN - x_w(x - 1)A = \langle 0 \rangle, x \rangle_f$$

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

In effect is a much larger sets of parameters $- 37$ per distribution.

algorithm) to find the best fit, rather than a fixed parametrisation.

using a neural net which undergoes a series of evolutions (mutations) via genetic
approach additionally (largely) eliminates parameterisation dependence by NNPDF
to give all sets a similar quality fit.

In earlier versions weighted error function for different data sets in early stages to try

into different data sets.

full global data set, but split

function (analogous to \( \chi^2 \)) for

not simply value of error

Criterion for stopping the fit

(hopefully slowly) improving.

even though training set still

validation set starts to go up,

Fit until quality of fit to

validation sets.

Split data sets randomly

0 5000 10000 15000 20000 25000 30000 35000 40000 45000 50000
0 1.5 2 2.5 3 3.5

GA generations

NMC-pd

Split datasets randomly into equalsize training and
Fluctuations in sets smaller with longer stopping point. Quality of global fit for both training and validation decreasing significantly after stopping point. Some evidence not at best fit in previous versions (Forte, DESY Oct 2009).

<table>
<thead>
<tr>
<th>$\chi^2$ of the global fit decreases a lot!</th>
<th>$\chi^2_{\text{part}}$ of the global fit increases a lot!</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.019</td>
<td>0.032</td>
</tr>
<tr>
<td>0.039</td>
<td>0.038</td>
</tr>
<tr>
<td>0.036</td>
<td>0.037</td>
</tr>
<tr>
<td>2.97</td>
<td>3.27</td>
</tr>
<tr>
<td>3.07</td>
<td>3.37</td>
</tr>
<tr>
<td>3.17</td>
<td>3.47</td>
</tr>
<tr>
<td>3.27</td>
<td>3.57</td>
</tr>
<tr>
<td>3.37</td>
<td>3.67</td>
</tr>
<tr>
<td>3.47</td>
<td>3.77</td>
</tr>
<tr>
<td>3.57</td>
<td>3.87</td>
</tr>
<tr>
<td>3.67</td>
<td>3.97</td>
</tr>
<tr>
<td>3.77</td>
<td>4.07</td>
</tr>
<tr>
<td>3.87</td>
<td>4.17</td>
</tr>
<tr>
<td>3.97</td>
<td>4.27</td>
</tr>
<tr>
<td>4.07</td>
<td>4.37</td>
</tr>
<tr>
<td>4.17</td>
<td>4.47</td>
</tr>
<tr>
<td>4.27</td>
<td>4.57</td>
</tr>
<tr>
<td>4.37</td>
<td>4.67</td>
</tr>
<tr>
<td>4.47</td>
<td>4.77</td>
</tr>
<tr>
<td>4.57</td>
<td>4.87</td>
</tr>
<tr>
<td>4.67</td>
<td>4.97</td>
</tr>
<tr>
<td>4.77</td>
<td>5.07</td>
</tr>
<tr>
<td>4.87</td>
<td>5.17</td>
</tr>
<tr>
<td>4.97</td>
<td>5.27</td>
</tr>
<tr>
<td>5.07</td>
<td>5.37</td>
</tr>
<tr>
<td>5.17</td>
<td>5.47</td>
</tr>
<tr>
<td>5.27</td>
<td>5.57</td>
</tr>
<tr>
<td>5.37</td>
<td>5.67</td>
</tr>
<tr>
<td>5.47</td>
<td>5.77</td>
</tr>
<tr>
<td>5.57</td>
<td>5.87</td>
</tr>
<tr>
<td>5.67</td>
<td>5.97</td>
</tr>
<tr>
<td>5.77</td>
<td>6.07</td>
</tr>
<tr>
<td>5.87</td>
<td>6.17</td>
</tr>
<tr>
<td>5.97</td>
<td>6.27</td>
</tr>
<tr>
<td>6.07</td>
<td>6.37</td>
</tr>
<tr>
<td>6.17</td>
<td>6.47</td>
</tr>
<tr>
<td>6.27</td>
<td>6.57</td>
</tr>
<tr>
<td>6.37</td>
<td>6.67</td>
</tr>
<tr>
<td>6.47</td>
<td>6.77</td>
</tr>
<tr>
<td>6.57</td>
<td>6.87</td>
</tr>
<tr>
<td>6.67</td>
<td>6.97</td>
</tr>
<tr>
<td>6.77</td>
<td>7.07</td>
</tr>
<tr>
<td>6.87</td>
<td>7.17</td>
</tr>
<tr>
<td>6.97</td>
<td>7.27</td>
</tr>
<tr>
<td>7.07</td>
<td>7.37</td>
</tr>
<tr>
<td>7.17</td>
<td>7.47</td>
</tr>
<tr>
<td>7.27</td>
<td>7.57</td>
</tr>
<tr>
<td>7.37</td>
<td>7.67</td>
</tr>
<tr>
<td>7.47</td>
<td>7.77</td>
</tr>
<tr>
<td>7.57</td>
<td>7.87</td>
</tr>
<tr>
<td>7.67</td>
<td>7.97</td>
</tr>
<tr>
<td>7.77</td>
<td>8.07</td>
</tr>
<tr>
<td>7.87</td>
<td>8.17</td>
</tr>
<tr>
<td>7.97</td>
<td>8.27</td>
</tr>
<tr>
<td>8.07</td>
<td>8.37</td>
</tr>
<tr>
<td>8.17</td>
<td>8.47</td>
</tr>
<tr>
<td>8.27</td>
<td>8.57</td>
</tr>
<tr>
<td>8.37</td>
<td>8.67</td>
</tr>
<tr>
<td>8.47</td>
<td>8.77</td>
</tr>
<tr>
<td>8.57</td>
<td>8.87</td>
</tr>
<tr>
<td>8.67</td>
<td>8.97</td>
</tr>
<tr>
<td>8.77</td>
<td>9.07</td>
</tr>
<tr>
<td>8.87</td>
<td>9.17</td>
</tr>
<tr>
<td>8.97</td>
<td>9.27</td>
</tr>
<tr>
<td>9.07</td>
<td>9.37</td>
</tr>
<tr>
<td>9.17</td>
<td>9.47</td>
</tr>
<tr>
<td>9.27</td>
<td>9.57</td>
</tr>
<tr>
<td>9.37</td>
<td>9.67</td>
</tr>
<tr>
<td>9.47</td>
<td>9.77</td>
</tr>
<tr>
<td>9.57</td>
<td>9.87</td>
</tr>
<tr>
<td>9.67</td>
<td>9.97</td>
</tr>
<tr>
<td>9.77</td>
<td>10.07</td>
</tr>
<tr>
<td>9.87</td>
<td>10.17</td>
</tr>
</tbody>
</table>

25000 Gens (averages 1000 Gens, for standard fit) Perform a fit with a fixed, very large number of CA generations:

**ARE WE CONSTRAINED BY THE FUNCTIONAL FORM?**

Difficult to know when to stop (analogous to variable in other approaches).
Number of data points $\sim 3000 \sim \frac{2\sqrt{N_{\text{data}}}}{\sqrt{2}} \sim 0.025$. 

\[
E \approx 0.05 \sqrt{\frac{N_{\text{data}}}{2}} = 0
\]

**Fluctuation of Overlapping Fit Statistics:**

Data uncertainty $\Leftrightarrow \approx \langle \chi^2 \rangle \Leftrightarrow$ Functional uncertainty suppressed in overlapping fits.

<table>
<thead>
<tr>
<th>$\langle \chi^2 \rangle$</th>
<th>1.19</th>
<th>0.32</th>
<th>0.28</th>
<th>0.35</th>
<th>0.39</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Value</td>
<td>1.29 ± 0.13</td>
<td>1.43 ± 0.19</td>
<td>1.60 ± 0.20</td>
<td>1.65 ± 0.24</td>
<td>2.79 ± 0.24</td>
</tr>
<tr>
<td>Fixed Value</td>
<td>1.19</td>
<td>1.18</td>
<td>1.35</td>
<td>1.32</td>
<td>1.32</td>
</tr>
<tr>
<td>Fixed Long</td>
<td>Replace Central Value</td>
<td>Replace Partition</td>
<td>Replace Central Value</td>
<td>Replace Partition</td>
<td>Replace Central Value</td>
</tr>
</tbody>
</table>

**When the best fit is not at the minimum**

Where is the uncertainty coming from?

Statistical behaviour (arguably) more like expected for longer stopping?
Uncertainty from \( \overline{\text{fits}} \) (green) was (normally) rather smaller than default (blue) \( \overline{\text{fits}} \) (green) was (normally) rather smaller than default (blue). Arguable if lack of smoothness becomes a problem. Arguable if lack of smoothness becomes a problem. Arguable if lack of smoothness becomes a problem. Arguable if lack of smoothness becomes a problem.
Is there a definitive set of stopping criteria?

Arguably a bit larger than naively expected. Fluctuations in error function (and \( \chi^2 \)) still actual value very difficult to ascertain. 

I would suggest uncertainty now more analogous to smaller “\( \chi^2 \)”, but significant reductions (usually) in central values, just due to change in uncertainty in latest version, and changed stopping and fitting procedures. 

Significant reductions made more strict. 

Criterion for increase in fit to validation sets relative to decrease in training sets criterion for increase in fit to validation individual sets. 

Weighted training in early stages according to a target (determined iteratively), so
NNPDF uncertainties pretty similar to other groups, with some particular exceptions.

Also reductions in uncertainty due to inclusion of new data. (Improved treatment of normalisations generally increases uncertainty slightly.)
Uncertainties on valence quarks not notably different to other groups at all.
Gluon parameterisation: different parameterisations lead to very different uncertainty for small $x$ gluon.

Most assume a single power $x^\alpha$ at input $x$. If input at low $Q^2$, fine-tuned to $\alpha \approx 0$. Artificially small uncertainty. If input at low $Q^2$ quark, tuned to $\alpha \approx 0$. Rapid expansion of uncertainty where data runs out.

MRST/MSTW and NNPDF more flexible (can be negative) fractional uncertainty.

If $g(x) / x \beta * (x/1) \ln(y \ln(1/x)) = (x) \ln(1/x) \ln(1/x)$, then $\ln(1/x) \ln(1/x) \propto (x) \beta$.

Banff - July 2010
Find more compatibility between Run I and Run II fits.

Fits not very consistent between runs.

Former found Gluon much softer for Run II.

Fits not very consistent between runs.

CTEQ and MSTW and CTEQ and now also NNPDF. Former found Gluon much softer for Run II.

Slightly confusing picture.

High-$p_T$ jets, now Run II and Run I available. Slightly confusing picture.

CTEQ and MSTW and CTEQ and now also NNPDF. Former found Gluon much softer for Run II.

Slightly confusing picture.

Gluon Distribution at $Q^2 = 1.04$ GeV$^2$
Generally high-$x$ PDFs parameterised so will behave like $\mu(x) - (1 - x)\mu_{CTEQ}$ 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 $x \to 1$. More so eventhan NNPDF.
StrangeQuarks

Direct fitting to dimuon data leads to significant uncertainty increase compared to assumption of fixed fraction of sea used until recently. Constraint for $x < 0.01$. Direct fit to $s^+$ from dimuon data leads to significant uncertainty increase compared to same small-$x$ power for strange as light quarks.

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 strange small-$x$ power. MSTW assumes shape of strange given by theory assumption that suppression of form...
NNPDF 2.0, which includes dimuon data, have no theoretical constraint on strange quark distribution at all at small $x$.

Overestimate of uncertainty? Impact on small-$x$ light quarks.

| MSTW 2008 | CTEQ 6.6 | NNPDF 2.0 |
Most recent sets obtain $s - s$ for first time from differences in dimuon production. $\Delta^2\nu\nu$ Strange asymmetry.

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

NNPDF tends towards positive momentum asymmetry, but all fairly consistent with zero, or

$\sin^2 \theta_W$ anomaly on NuTeV.
PDF uncertainties reduced since quality of fit already worse than best fit.

PDF correlation with $x_s$.

Expected gluon–$x$ small–$x$ anti-correlation from sum rule.
Quarks roughly opposite to gluons.

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

Gluon feeds into evolution of quarks, but change in $\langle \frac{Z_c W}{x} \rangle_{\alpha S}$ slightly more evolution, i.e. larger change for gluons, just outweighs gluon change, i.e. large change for quarks, but change in $\langle \frac{Z_c W}{x} \rangle_{\alpha S}$ still small; gluons' $\frac{Z_c W}{x}$ feed into evolution of quarks.
Total uncertainty envelope of set of uncertainties increases by up to 50% at LHC. Largely due to effect of PDFs.

\[ \sigma_{\beta}^{Z}(M_{Z}^{MSTW}) \]

NNLO predictions for production for allowed \( \frac{Z}{\alpha} \) and their uncertainties. And couplings are correlated.

\[ \sigma_{\beta}^{Z}(M_{Z}^{MSTW}) \]

Additional uncertainty from \( \alpha \) variation for quantities depends on how PDFs vary.
NNLO predictions for Higgs (120 GeV) production for different allowed values of $M_{Z'}$ and their uncertainties.

At Tevatron (HERA data), the intrinsic gluon uncertainty is mitigated somewhat by anti-correlated small-$x$ gluon asymmetry features at minor problems in fit to HERA data. At Tevatron intrinsic gluon uncertainty dominates.

Increases by a factor of 2–3 (up more than down) at LHC. Direct $\sigma$ dependence differs.

Banff, July 2010.
Responsible for differences between groups for extraction of fixed-order PDFs.

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

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

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

\[ (x - 1)_{1-u}^{s_0} \] large or \[ (x/1)_{1-u}^{s_0} \] small.

Resummations, e.g. small \( x \) engineering orders (NNLO – may sets available here).

Standard higher orders (NNLO – comparable to NNLO ?). Sometimes enhancements.

\[ (\frac{x}{3})^\alpha \] QED and weak (comparable to NNLO ?).
Predictions by various groups-parton luminosities NLO.

Plots by G. Watt.

Also $H + \ell^+\ell^-$ at $\sqrt{s} \sim 0.1$.

Cross-section for $\ell^+\ell^-$ almost identical in PDF terms to 450GeV Higgs.

Banff – July 2010
Clearsly 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 $x$. 
Canoncial example $W,Z$ production, but higher $s/s$ relevant for vector boson fusion.

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

MSTW08 NLO (68% C.L.)

$\frac{\sigma}{s_\text{flux}}$ luminosity at LHC ($\sqrt{s} = 7 \text{ TeV}$)
Variations in Cross-Section Predictions

NLO

$\sigma_0(\frac{Z^2}{M_\omega})$

Dotted lines show how central PDF predictions vary with $\sigma_0(\frac{Z^2}{M_\omega})$.

Again plots by C Watt using PDF4LHC benchmark criteria.

68% C.L. PDF

MSTW08

CTEQ6.6

NNPDF2.0

HERAPDF1.0

ABKM09

GJR08

NLO $gg \rightarrow H$ at the LHC ($\sqrt{s} = 7$ TeV) for $M_H = 120$ GeV

NLO - NLO Predictions – NLO
Groups. Clearly much more variation in predictions than uncertainties claimed by individual

\[
\frac{Z}{M_s}\alpha
\]

\(\sigma_{H}^{\text{NLO gg}}\) at the LHC \((\sqrt{s} = 7\ \text{TeV})\) for \(M = 180\ \text{GeV}\)
Excluding GJR08 amount of difference due to $\alpha_S(M_Z^2)$ variations is $3 - 4\%$.
CTEQ6.6 now heading back towards MSTW08 and NNPDF2.0.
dependence now more due to PDF variation with\[ (Z_\nu)_{\alpha} \]cross-section. The\[ (Z_\nu)_{\alpha} \]dependence now more due to PDF variation with $M^+_\nu + M^-_\nu$.
Roughly similar variation for up to a few times higher.

Again variations somewhat bigger than individual uncertainties.
Quite a variation in ratio. Shows variations in flavour and quark-antiquark decompositions.

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

\[ x_s (M_Z^2) \]

\[ R = \frac{\sigma_{W^+}}{\sigma_W} \]

NLO W+/W ratio at the LHC (\( \sqrt{s} = 7 \) TeV)
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 parametrisation at small $x$ affects predicted Higgs cross-section.

General rule: Not very satisfactory, but not clear what would be an improvement, especially as a factor of about 2 expansion of MSTW uncertainty.

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

What is true uncertainty? Task asked of PDF4LHC Group.

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

Not very satisfactory, but not clear what would be an improvement, especially as a factor of about 2 expansion of MSTW uncertainty.
Sometimes rather worse than this for special case, e.g. Warsinsky at LHC Working Group meeting.

\[ m_b \] and CTQ values bring together but NNPDF exaggerate 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 - 1 - 5% for most LHC quantities.

Errors from higher orders/resummation potentially large.

Imperfect theory used to fit data. Various ways of looking at uncertainties due to errors on data. 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 parametrisation can shift central values of predictions significantly. Different groups do not always agree very well despite "generous" uncertainties.

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 $S$ accounted for. $S$ is correlated. Now being dealt with properly for in general. Reduces but does not remove discrepancies. Errors from higher orders/resummation potentially large.

and PDFs correlated. New being dealt with properly for in general. Reduces but does not remove discrepancies. Errors from higher orders/resummation potentially large.

Spatial variability and PDFs correlated. New being dealt with properly for in general. Reduces but does not remove discrepancies. Errors from higher orders/resummation potentially large.

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

CTEQ/MSTW.

More consistent data sets for uncertainties. Nevertheless similar to CTEQ/MSTW.

Banff - July 2010
Parameterisations

MSTW predictions for $W^+$ and $W^-$ cross-sections for LHC with common QCD and vector boson width effects, and common branching ratios. Quoted uncertainty for ratio very small, i.e. 0.8%. Predictions sensitive to and $p$ quarks. Predictions for $W^+$ and $W^-$ cross-sections for LHC with common fixed order and vector boson width effects, and common branching ratios. Very interesting for early data. Very interesting for $W$-asymmetry by Cooper-Sarkar. Significant more difference than uncertainty from other PDFs, including MRST. Should account for this. Fit includes most recent neutrino DIS and Tevatron vector boson data. Uncertainties significantly more difference than uncertainty from other PDFs, including MRST. QCD and vector boson width effects, and common branching ratios.
With new data, the NNPDF band shrinks dramatically and some of the differences are not well understood. Hardly affects the ratio, some groups even get even more discrepancies between them. Again comparing more
It is difficult to know when validation set has started increasing significantly for some sets.
Variations in partons extracted from global fits due to different choices of GM-VFNS. Some changes in PDFs large compared to one-sigma uncertainty.

Better fit for GM-VFNS1, GM-VFNS3 and GM-VFNS6. Initial $\chi^2$ can change by 250.

Converges to at most about 15 of original.
Variations in partons extracted from global fit due to different choices of GM-VFNS at NNLO.

Comparing different GM-VFNS at NNLO, the biggest variation in high-$x$ gluon is seen, which has large uncertainty. At worst changes approach uncertainty.

Initial changes in $\chi^2 > 20$. None a marked improvement. Converge to about 10.

GM-VFNS at NNLO.
Light quark singlet distribution at $Q^2 = M_Z^2$

Gluon distribution at $Q^2 = M_Z^2$ (120 GeV)

Ratio of partons when $m_c$ is varied either with or without varying $\alpha_s$. $\alpha_s$
Systematic difference between PDF defined at NLO and at NNLO.

Percentage difference at $Q^2 = 100 \text{GeV}^2$.
Parameterisation depends on one parameter, \( z \) at \( z = 0 \) such that value corresponds to \( \chi^2_{\text{best fit}} \) in original model with magnitude much greater than improvement in best fit quality. For long narrow ellipse, can get shift in profile as shown.

Add second parameter \( y \), could get \( \chi^2_{\text{best fit}} \) with magnitude much greater than improvement in best fit quality.

Proposal by Pumplin that this may be parameterisation dependence reason for inflated tolerance.

\[ \chi^2 \leq 100 \text{ tolerance} \]
Basic arguments seem to be validated by a variety of checks.

Why this doesn’t apply to global fits

1. In MSTW/MRST and CTEQ, there are more free parameters in new best fit than in old parameters. Fit obtained with one already

\( \chi^2 \) distribution (explaining small improvement is not likely, not a large change in a major axis. Very flat direction always slightly turns up due to quartic and higher terms. Along major axis, very elliptical profile only occurs if two of the parameters are very correlated.

2. This very elliptical profile only occurs if two of the parameters are very correlated. Directly apply since main effect – in change of minimum – already accounted for.

3. If \( z \) and \( y \) are highly correlated, a large change in \( z \) is likely not to change much PDF distribution (explaining small improvement is not likely, not a large change in a major axis. Very flat direction always slightly turns up due to quartic and higher terms. Along major axis, very elliptical profile only occurs if two of the parameters are very correlated.

4. If a new parameter is introduced, which is not highly correlated with one already

\( \chi^2 \) distribution (explaining small improvement is not likely, not a large change in a major axis. Very flat direction always slightly turns up due to quartic and higher terms. Along major axis, very elliptical profile only occurs if two of the parameters are very correlated.
Parameterisation used in MSTW fits. Only those 20 in red appear in eigenvectors.
expected from change in PDF, not well correlated to relative size of change in parameters. Size of change in $d\sqrt{\chi^2}$ biggest than shown. Worst change $1.8$.

Main change in PDFs in valence quarks by $2-3$ times quoted uncertainty. $\chi^2$ expected from change in PDF twice that for global uncertainty. Need these parameters not in eigenvectors equal to zero, or if this is not sensible a round off. Need these parameters not in eigenvectors equal to zero.

Try checking by setting all the parameters not in eigenvectors equal to zero.

Try checking by setting all the parameters not in eigenvectors equal to zero.
Try removing parameter which is not highly correlated, i.e., one in eigenvector definition.

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

This time change in PDF pretty much what deterioration in fit quality corresponds to \( \chi^2 \) uncertainty.

Again relevant eigenvectors suggest

\( \chi^2 \) is about twice the usual magnitude is about the most variation about 1.8 uncertainty.

And in terms of term in PDF shown.

Biggest change in PDF at most variation about 1.8 uncertainty.

Usual magnitude is about twice the term in PDF shown.

Set every eigenvector parameter free with usual eigenvector parameters.

\( \chi^2 \) is 30 worse.

\( \chi^2 \) for \( \Lambda \) is 30 worse.

\( \chi^2 \) for \( \Lambda \) is 30 worse.

\( \chi^2 \) for \( \Lambda \) is 30 worse.

\( \chi^2 \) for \( \Lambda \) is 30 worse.

\( \chi^2 \) for \( \Lambda \) is 30 worse.
Also tried adding terms \( x^2 \) to polynomial in two valence parameters.

Fit quality improved by 2 units.

Change in partons negligible.
at $x = 0.5$, with no new free parameters.

Recall study in first MRST2001, correction of this lead to automatic increase in uncertainty of about 50% in $\chi^2$ global analysis as the $e^+$ and $e^-$ cross sections are varied at HERA.

However, used non-optimum choice of parameters in eigenvectors for parameters in eigenvectors.

Excellent agreement between two for $F_{CC}^Z (e^- p)$ and 10% for $F_{CC}^Z (e^+ p)$.

Uncertainty using Hessian approach was 2%.

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

For charged current HERA structure functions at $x = 0.5$ (Red curve - fixed parameters and same $\chi^2$ for both, 22 parameters and same parameters and Lagrange multiplier with paper comparing the Hessian approach with MRST uncertainties).
about the amount expected.

For gluon increases uncertainty by

Extra parameter in eigenvectors

Uncertainty using Hessian approach

is 1.2% for W and 3% for Higgs.

Using 3 parameters lead to

Slightly smaller in latter case.

Lagrangian multiplier method for

Tevatron.

W and 120GeV Higgs at the

χ² increase in global analysis as the W and H cross sections are varied at the TEVATRON.
address and the problem solved. If this was clearly more than 10% moderate increase in uncertainty. Also looked at uncertainties on

Very similar and latter could be slightly smaller. About 10% smaller.

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

Ban®{July 201078
Not looking for anyway 0 \chi^2 < 100 anyway

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

Majority of eigenvectors correspond to $\chi^2/\nu \sim 2 - 3$. 

$\chi^2 = 100$ anyway
reason for large tolerance?

Not applied by CTEQ. Part of the

significance. Increases uncertainties on partons

Use of normalization uncertainties
determined from ratios.

$p - n$ (quarks) others insensitive size of
tolerance for eigenvectors –

Difficult to account for in

$1.7\%$, for all partons.

Normalization uncertainty $1 - I$

Uncertainties (except in best fit).

Comparison of full uncertainty
AT THE I LEVEL

VALUE & STRANGE PDG SPF APPRECIATED

GET SMALLER WEIGHT

EXPRTS WITH LARGER Syst. (FIXED TARGET) →
DIAGONAL PDG REWEIGETS EXPERIMENTS

CORREL. OF TWO PDG UNCHANGED

CORREL. PDG MUCH LOWER.

REPEAT THE PDG NEGLICIBLE ALL CORRELATIONS (A. DONELL)

THE IMPACT OF CORRELATED UNCERTAINTIES