News from CTEQ PDF group

Jon Pumplin – Michigan State University
PDF4LHC meeting (CERN, 29 May 2009)

1. CT09: new NLO PDFs using Tevatron run II inclusive jet data. Emphasis on methodology and on gluon at $x > 0.2$. (Most of this talk.)

2. "Data set diagonalization:" Measuring internal compatibility of the global fit by an extension of the Hessian method. (My talk this afternoon.)

3. PDFs for use in LO Monte Carlo simulations. (Joey Huston's talk this afternoon).

4. Simultaneous fit of PDFs and non-perturbative parameters for p_T distributions of DY processes using ResBos — Provides correlation info that is needed for W mass measurements; and has an outside chance of yielding new PDF info. (Talk some other day!)

Testing compatibility of experiments using weighted χ^2

The "Hypothesis testing" criterion $\chi^2 \approx N \pm \sqrt{2\,N}$, where N is the number of data points, represents a minimal requirement for each experiment.

$\mathrm{CDF_I}$ Wt	$\mathrm{CDF_I}$ χ^2	$\mathrm{D0_I}$ Wt	$\mathrm{D0_I}$ χ^2	$\mathrm{CDF_{II}}$ Wt	$\mathrm{CDF_{II}}$ χ^2	$\mathrm{D0_{II}}$ Wt	$\mathrm{D0_{II}}$ χ^2	$\Delta\chi^2$ non-jet
0	55.4	0	115.3	0	99.5	0	134.0	0.0
1	52.6	1	47.0	0	105.6	0	138.3	11.8
0	56.6	0	82.2	1	85.6	1	124.1	6.2
1	52.1	1	59.4	1	88.5	1	121.5	9.6
1	54.8	1	58.8	10	80.3	10	120.0	39.4
10	53.1	10	38.6	1	102.6	1	142.3	21.9
10	51.6	10	49.7	10	82.8	10	120.9	39.6
1	59.6	1	67.5	10	75.2	1	130.9	32.0
1	50.6	1	60.0	1	93.0	10	116.5	20.6

- For $\mathrm{CDF_I}$ (N=33), all fits fall a little outside the expected range 25–41. A couple of the data points appear to be unusually large fluctuations, which cannot be fitted at much better χ^2 using *any* smooth function. Unlike the other jet experiments, $\mathrm{CDF_I}$ has data only at central rapidity, so it is less sensitive to the gluon distribution than the others — in spite of the historic importance of this experiment in changing the view of the gluon at large x! The range of χ^2 for this experiment over the entire series of fits is quite small, and it therefore doesn't have much influence on the global fit.

$\mathrm{CDF_I}$		$\mathrm{D0_I}$		$\mathrm{CDF_{II}}$		$\mathrm{D0_{II}}$		$\Delta\chi^2$
Wt	χ^2	Wt	χ^2	Wt	χ^2	Wt	χ^2	non-jet
0	55.4	0	115.3	0	99.5	0	134.0	0.0
1	52.6	1	47.0	0	105.6	0	138.3	11.8
0	56.6	0	82.2	1	85.6	1	124.1	6.2
1	52.1	1	59.4	1	88.5	1	121.5	9.6
1	54.8	1	58.8	10	80.3	10	120.0	39.4
10	53.1	10	38.6	1	102.6	1	142.3	21.9
10	51.6	10	49.7	10	82.8	10	120.9	39.6
1	59.6	1	67.5	10	75.2	1	130.9	32.0
1	50.6	1	60.0	1	93.0	10	116.5	20.6

- For $\mathrm{D0_I}$ (N=90), the expected range is 77–103. The fact that fits with much lower χ^2 can be obtained suggests that the systematic errors provided with the data were overestimated.

Correlated systematic errors for this experiment were published only in the form of a single covariance matrix, rather than being broken out as individual shifts associated with each source of systematic error, whose magnitudes can be directly tested for plausibility. Systematic errors given in this form can nevertheless be analyzed using Principal Component Analysis — work in progress by Pavel Nadolsky.

$\mathrm{CDF}_{\mathrm{I}}$		$\mathrm{D0}_{\mathrm{I}}$		$\mathrm{CDF}_{\mathrm{II}}$		$\mathrm{D0}_{\mathrm{II}}$		$\Delta\chi^2$
Wt	χ^2	Wt	χ^2	Wt	χ^2	Wt	χ^2	non-jet
0	55.4	0	115.3	0	99.5	0	134.0	0.0
1	52.6	1	47.0	0	105.6	0	138.3	11.8
0	56.6	0	82.2	1	85.6	1	124.1	6.2
1	52.1	1	59.4	1	88.5	1	121.5	9.6
1	54.8	1	58.8	10	80.3	10	120.0	39.4
10	53.1	10	38.6	1	102.6	1	142.3	21.9
10	51.6	10	49.7	10	82.8	10	120.9	39.6
1	59.6	1	67.5	10	75.2	1	130.9	32.0
1	50.6	1	60.0	1	93.0	10	116.5	20.6

- For $\mathrm{CDF}_{\mathrm{II}}$, the expected range is 60–84. The fit gives $\chi^2 = 88$ with all jet weights equal to 1, dropping to 75 for weight 10, $\Longrightarrow$ these data are consistent with theory.

- For $\mathrm{D0}_{\mathrm{II}}$, the expected range is 95–125. The fit gives $\chi^2 = 121$ with all jet weights equal to 1, dropping to 116 for weight 10, again consistent with theory.

$\mathrm{CDF_I}$		$\mathrm{D0_I}$		$\mathrm{CDF_{II}}$		$\mathrm{D0_{II}}$		$\Delta\chi^2$ non-jet
Wt	χ^2	Wt	χ^2	Wt	χ^2	Wt	χ^2	
0	55.4	0	115.3	0	99.5	0	134.0	0.0
1	52.6	1	47.0	0	105.6	0	138.3	11.8
0	56.6	0	82.2	1	85.6	1	124.1	6.2
1	52.1	1	59.4	1	88.5	1	121.5	9.6
1	54.8	1	58.8	10	80.3	10	120.0	39.4
10	53.1	10	38.6	1	102.6	1	142.3	21.9
10	51.6	10	49.7	10	82.8	10	120.9	39.6
1	59.6	1	67.5	10	75.2	1	130.9	32.0
1	50.6	1	60.0	1	93.0	10	116.5	20.6

The question of whether the four inclusive jet experiments are consistent with each other in the fit can be studied by looking at how increased weight for some of them affects the χ^2 for the others.

- The two Run II experiments are fairly consistent with each other, since when $\mathrm{CDF_{II}}$ is assigned weight 10, its χ^2 is not strongly dependent on whether $\mathrm{D0_{II}}$ is assigned weight 1 or 10; and similarly when $\mathrm{D0_{II}}$ is assigned weight 10, its χ^2 is not strongly dependent on whether $\mathrm{CDF_{II}}$ is assigned weight 1 or 10. However, in each case there is a small increase in χ^2 for one of the experiments when the weight for the other is increased, which suggests a bit of tension between them. We will see this more clearly this afternoon, using a more powerful method of analysis.

$\mathrm{CDF}_{\mathrm{I}}$		$\mathrm{D0}_{\mathrm{I}}$		$\mathrm{CDF}_{\mathrm{II}}$		$\mathrm{D0}_{\mathrm{II}}$		$\Delta\chi^2$
Wt	χ^2	Wt	χ^2	Wt	χ^2	Wt	χ^2	non-jet
0	55.4	0	115.3	0	99.5	0	134.0	0.0
1	52.6	1	47.0	0	105.6	0	138.3	11.8
0	56.6	0	82.2	1	85.6	1	124.1	6.2
1	52.1	1	59.4	1	88.5	1	121.5	9.6
1	54.8	1	58.8	10	80.3	10	120.0	39.4
10	53.1	10	38.6	1	102.6	1	142.3	21.9
10	51.6	10	49.7	10	82.8	10	120.9	39.6
1	59.6	1	67.5	10	75.2	1	130.9	32.0
1	50.6	1	60.0	1	93.0	10	116.5	20.6

- The Run I and Run II measurements are fairly consistent with each other, e.g. if the Run II experiments are assigned weight 10, then raising the weight for Run I data from 1 to 10 improves the fits to Run I as it must, while making very little change in the χ^2 for the Run II and non-jet experiments.

- When only the Run I data, or only $\mathrm{D0}_{\mathrm{I}}$ are assigned weight 10, $\chi^2(\mathrm{D0}_{\mathrm{I}})$ is reduced quite a bit at the expense of a significant increase in $\chi^2(\mathrm{CDF}_{\mathrm{II}})$ and $\chi^2(\mathrm{D0}_{\mathrm{II}})$. This way of looking at it suggests a small "tension" between Run I and Run II; but it may be related to the same details of the systematic error treatment in D0 Run I that allows χ^2/N to become small for that experiment. Again will look at this in a better way in this afternoon's talk.

Gluon parametrizations

CT09 uses

$$g(x,\mu_0) = a_0\, x^{a_1}\, (1-x)^{a_2}\, \text{exp}(a_3 x + a_4 x^2 + a_5 \sqrt{x})$$

with quartic penalty in χ^2 to force $0.5 < a_2 < 10$.

CTEQ6.6 used less flexible form: $a_5 = 0$, $a_2 = 4$.

A still less flexible form has been used at HERA:

$$g(x,\mu_0) = a_0\, x^{a_1}\, (1-x)^{a_2}\, (1 + a_3 x)$$

That form is too restrictive, as is seen when one attempts to fit the Tevatron jet data with it:

$$a_0\, x^{a_1}\, (1-x)^{a_2}\, \exp(a_3 x + a_4 x^2 + a_5 \sqrt{x})$$

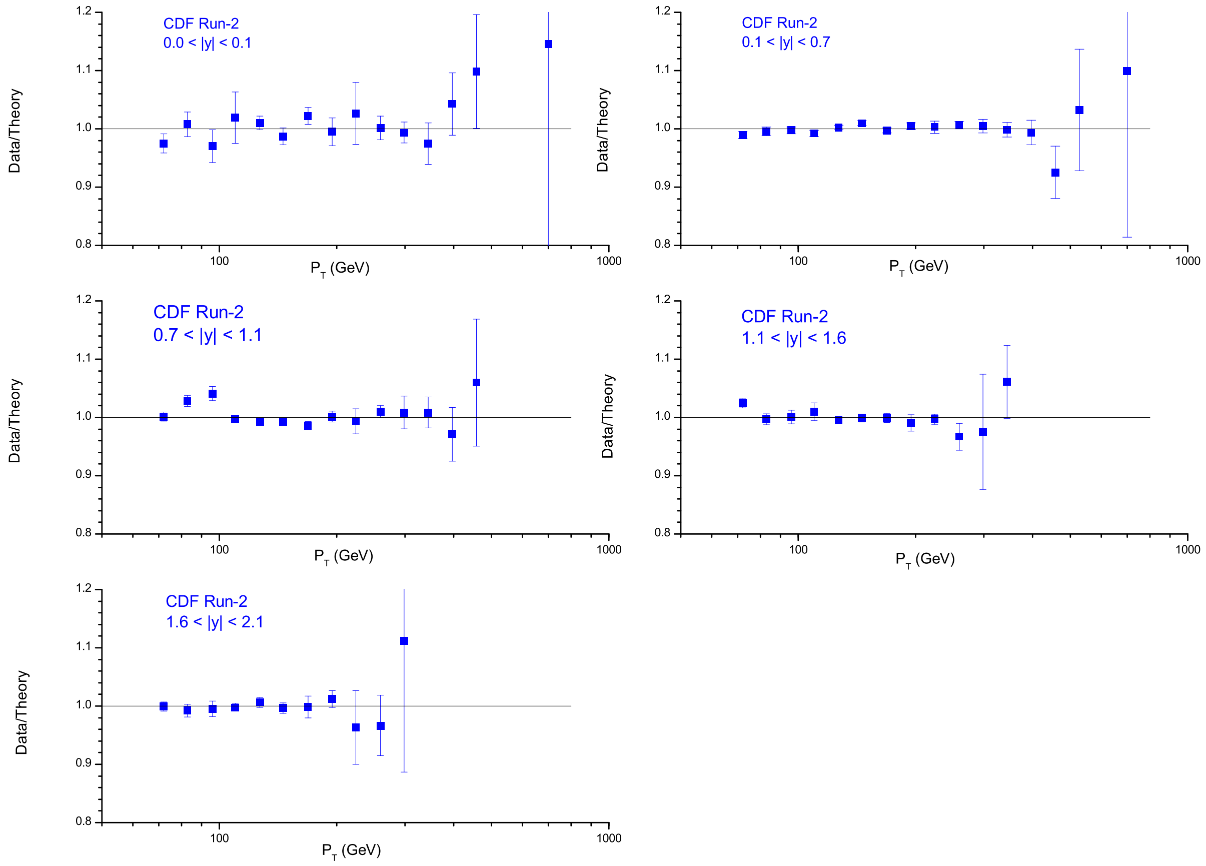
$\mathrm{CDF_I}$		$\mathrm{D0_I}$		$\mathrm{CDF_{II}}$		$\mathrm{D0_{II}}$		$\Delta\chi^2$
Wt	χ^2	Wt	χ^2	Wt	χ^2	Wt	χ^2	non-jet
0	55.4	0	115.3	0	99.5	0	134.0	0.0
1	52.6	1	47.0	0	105.6	0	138.3	11.8
0	56.6	0	82.2	1	85.6	1	124.1	6.2
1	52.1	1	59.4	1	88.5	1	121.5	9.6
1	54.8	1	58.8	10	80.3	10	120.0	39.4
10	53.1	10	38.6	1	102.6	1	142.3	21.9
10	51.6	10	49.7	10	82.8	10	120.9	39.6

$$a_0\, x^{a_1}\, (1-x)^{a_2}\, (1 + a_3 x)$$

$\mathrm{CDF_I}$		$\mathrm{D0_I}$		$\mathrm{CDF_{II}}$		$\mathrm{D0_{II}}$		$\Delta\chi^2$
Wt	χ^2	Wt	χ^2	Wt	χ^2	Wt	χ^2	non-jet
0	57.8	0	175.4	0	139.5	0	180.9	4.3
1	54.5	1	137.9	0	148.5	0	165.8	91.4
0	64.0	0	143.7	1	102.4	1	151.4	27.8
1	53.9	1	128.4	1	105.4	1	143.0	110.7
10	53.5	10	75.6	1	115.8	1	139.6	299.2
1	68.5	1	77.8	10	75.2	10	131.7	163.9
10	54.4	10	67.7	10	72.8	10	131.8	259.0

- The simple parametrization works fine if no jet data: $\Delta\chi^2 = 4.3$. It therefore passes the Dangerous criterion that new parameters don't need to be added if they don't lower χ^2 by much. This criterion would lead to a predicted uncertainty range from the no-jet data which would not encompass the jet data; but the above table shows that with a sufficiently flexible parametrization, the no-jets prediction *should* encompass the jet data.

Central CT09 fit to CDF run II data



CDF provides 24 systematic error parameters (many parameters because some apply only to specific rapidity intervals). The data shown are shifted according to the fitted values of those parameters:

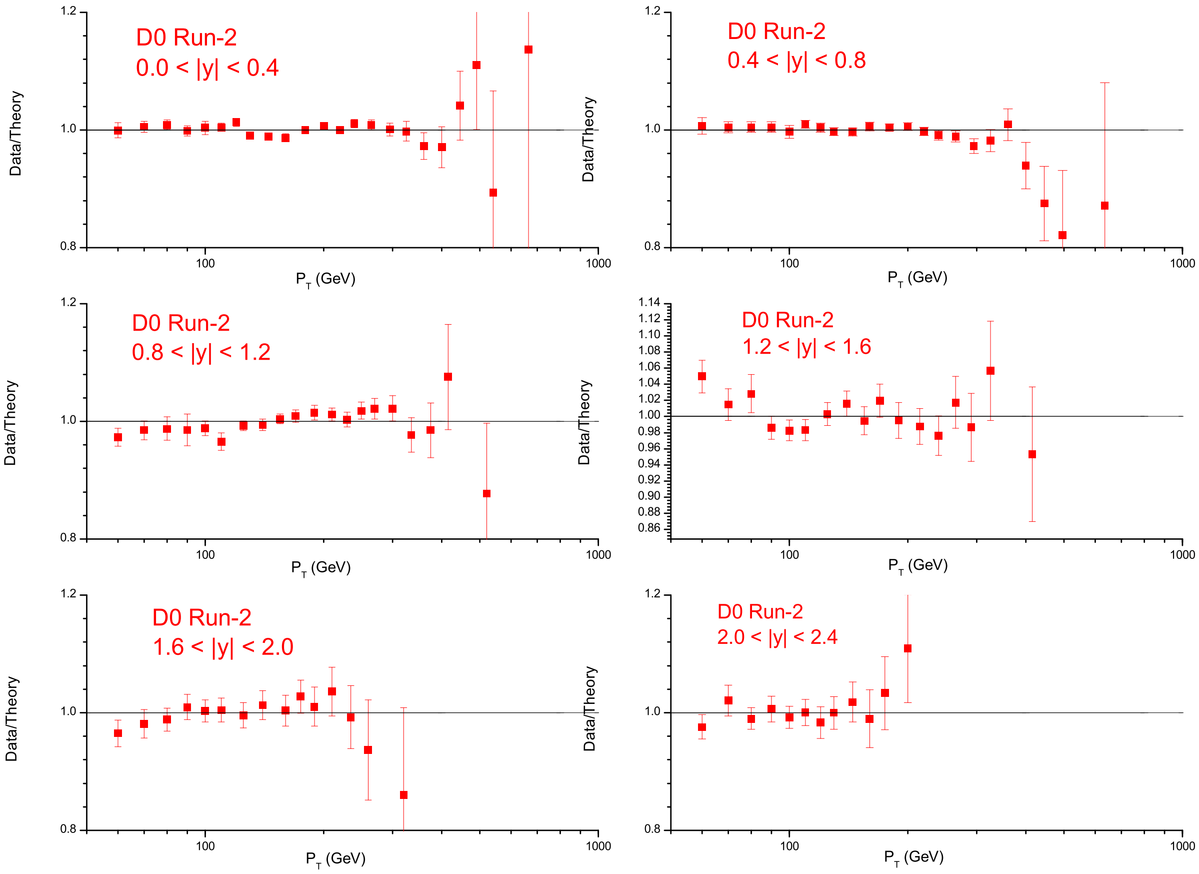
−0.1 −1.0 −0.3 −1.0 0.7 −0.2 0.8 −0.7 −0.7 −0.9 0.1 0.6 1.0 −0.3 −0.3 0.5 −1.2 0.4 0.9 0.0 −1.3 0.1 −0.1 −0.3

All of the shifts are less than or of order 1, as they should be. The fitted normalization factor is also reasonable.

The data table in the published CDF run II jet paper contained some errors. CDF is in the process of publishing an erratum in the arXive and in PRD.

(Watch for a similar update in the systematic errors for the CDF Z^0 rapidity distribution.)

Central CT09 fit to D0 run II data

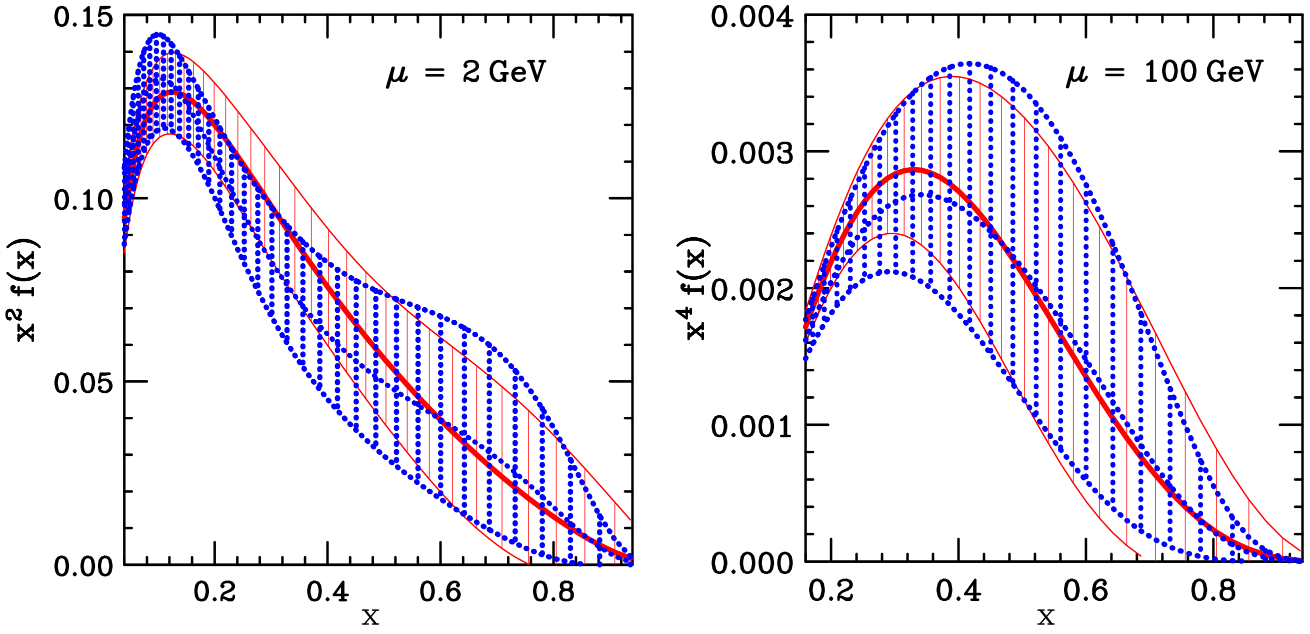


D0 provides 23 systematic error parameters. The data shown are shifted according to the fitted values of those parameters:

−0.6 −1.6 0.1 0.2 −0.9 0.1 −0.5 0.1 1.1 −0.4 1.1 −1.1 −0.5 0.4 −1.6 −0.2 −1.9 0.5 0.3 0.2 1.7 −1.1 −0.1

Several of these are a bit larger than 1. That may be acceptable — given that the systematic errors are generally based on estimates — but it would be a useful precaution to see what happens if a constraint is applied to force all of these parameters to be less than 1.0

Results: Gluon distributions and uncertainties

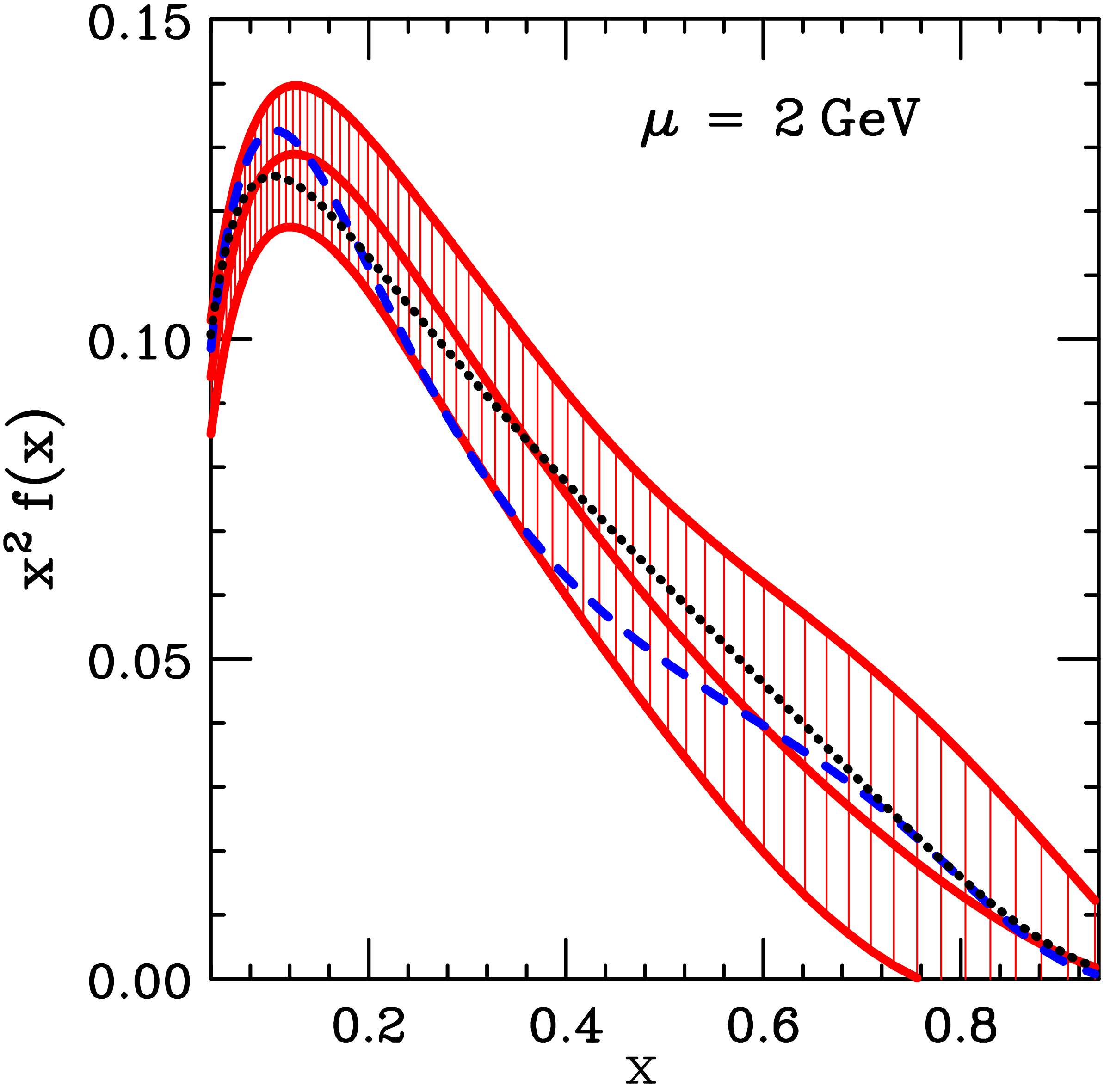


CT09 (red) and CTEQ6.6 (blue).

- Including substantially more jet data has not reduced the uncertainty very much! (Why not?? — More freedom in the parametrization + small conflicts between the data sets.)

- The change at $x = 0.3$ is as large as the old 90 % confidence limit. Hence the $\Delta\chi^2 = 100$ criterion was not overly conservative.

History of CTEQ gluon at large x

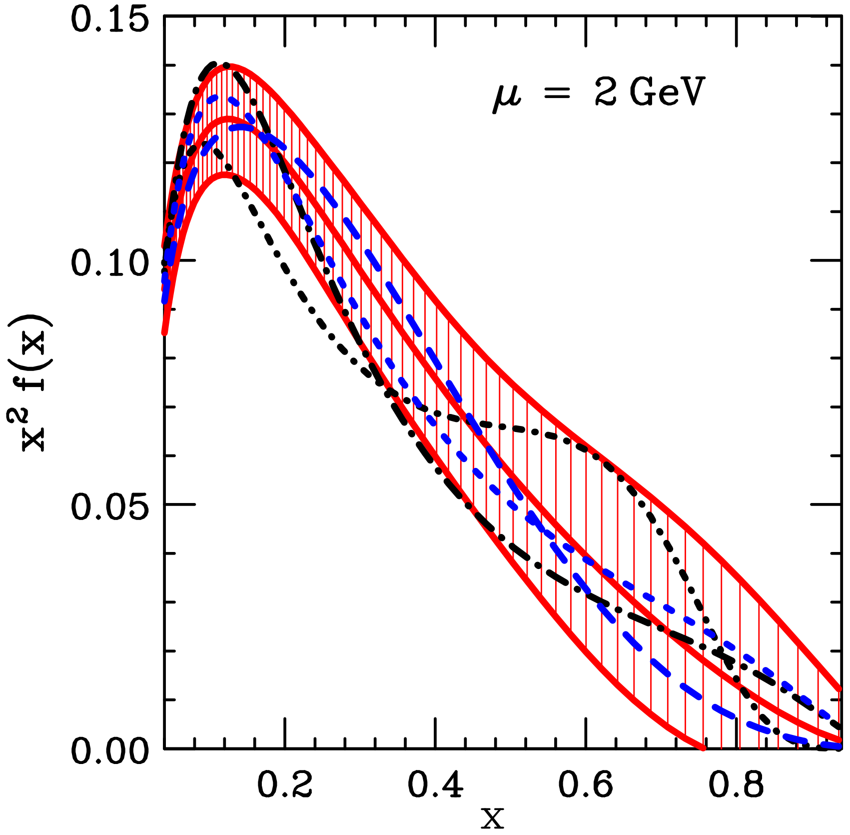


red: CT09

blue: CTEQ6.6

black: CTEQ6.1 (Zero Mass scheme for c, b)

Preferences of each jet experiment

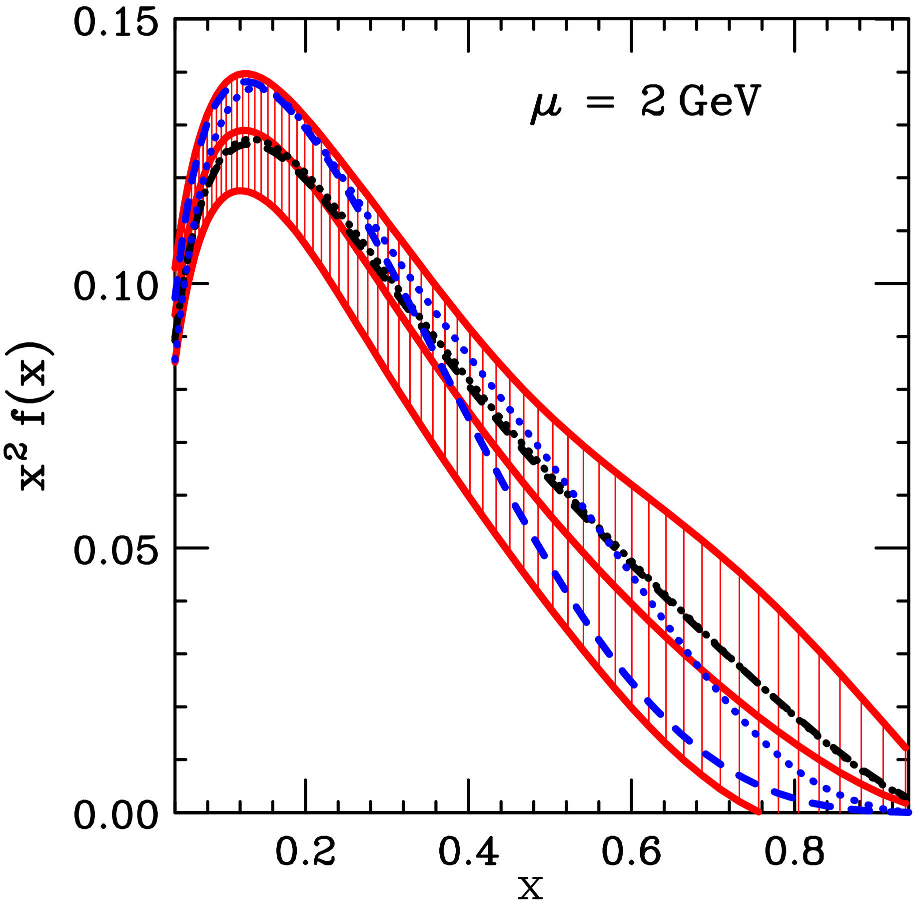


Blue long dash: wgt 10 for CDF_{II}; 1 for others
Blue short dash: wgt 10 for D0_{II}; 1 for others
Black long dash dot: wgt 10 for CDF_{I}; 0 for others
Black short dash dot: wgt 10 for D0_{I}; 0 for others

- Differences between the experiments is responsible for much of the uncertainty.

- CDF_{I} is the least constraining data set because it is only at central rapidity; but it stood out when it was the only jet measurement because it prefers the HJ bump form.

Comparison with MSTW



Red shaded: CT09

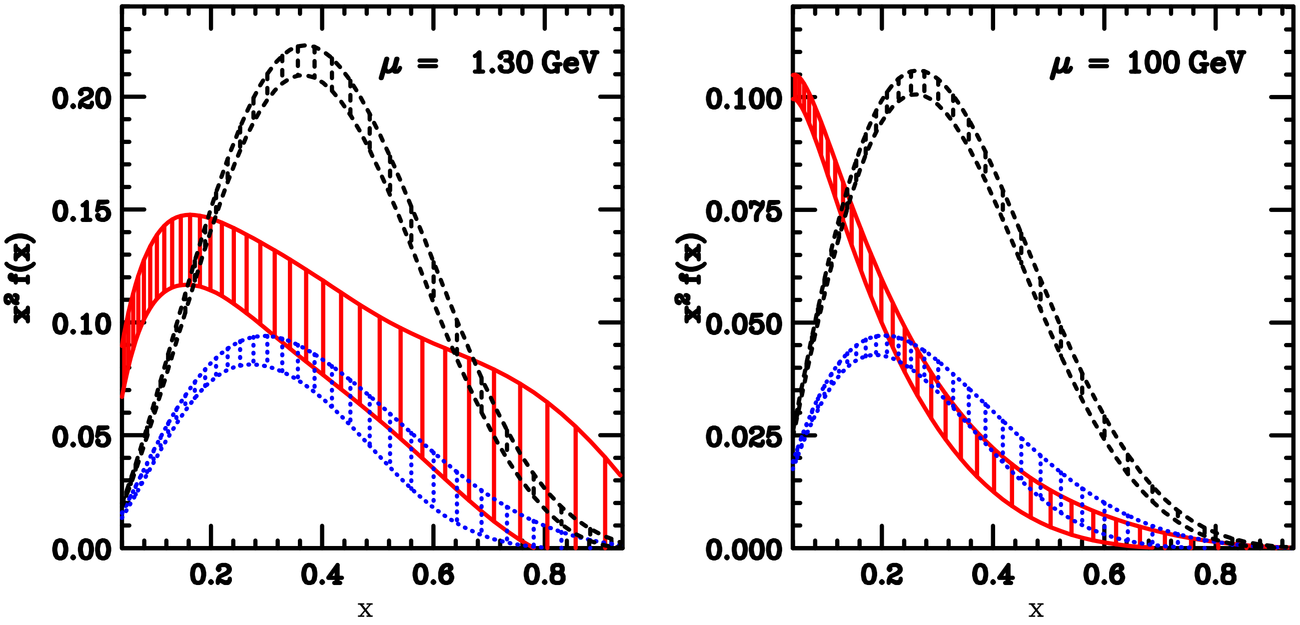
Black: like CT09 but $\alpha_s(m_Z) = 0.12018$ (MSTW)

Blue dashed: MSTW2008NLO

Blue dotted: MSTW2004NLO

- Larger $\alpha_s \Rightarrow$ smaller quark $\Rightarrow$ larger gluon
- Once again, differences in methodology (CTEQ v. MSTW) are as large as the estimated error.
- Both MSTW2008 and MSTW2004 are within our error estimates — though perhaps MSTW2008 is outside the error estimate from MSTW2004.

A challenge for non-perturbative theory



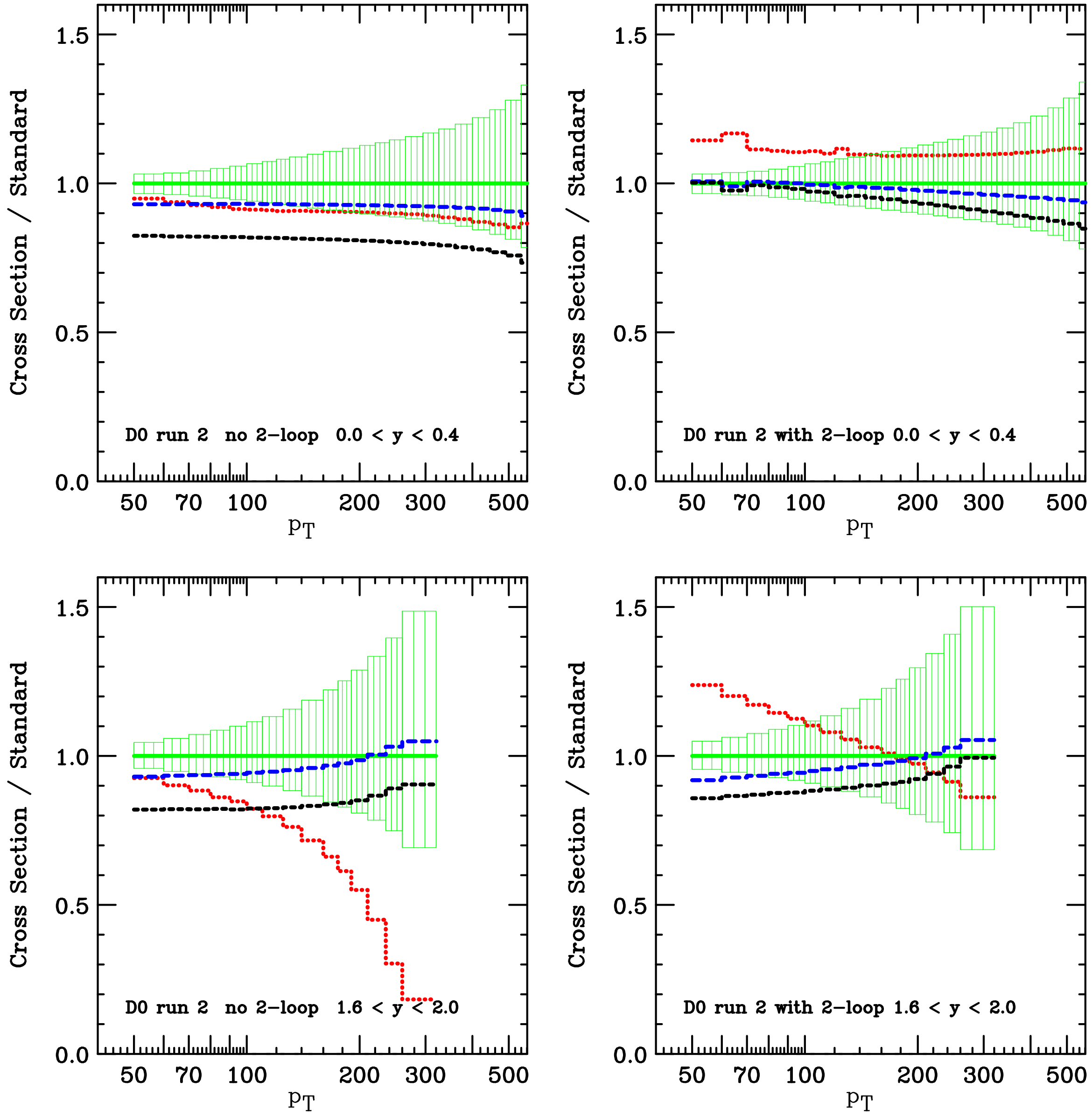
Gluon (red), u quark (black), and d quark (blue) distributions.

Can “valence-like” gluon at small μ be ruled out on theoretical grounds??

(At large μ, valence quarks dominate at large x no matter what PDF assumptions are made.)

Extra Slides

Effect of Scale choice and “2-loop” correction



Effect of scale choice on predicted cross section: $\mu = 2\,p_T$ (black), p_T (blue), $p_T/2$ (green), $p_T/4$ (red), relative to our Standard Choice ($\mu = p_T/2$, no “two-loop” correction). Right panels include the “two-loop” resummation correction. Uncertainty bands from PDFs are shown for comparison.

Does MSTW use the two-loop correction?