Jet data from hadron colliders and correlated uncertainties

work in progress

Jun Gao, Pavel M. Nadolsky

Southern Methodist University

May 22, 2012

Sensitivity of ATLAS jet data to PDF

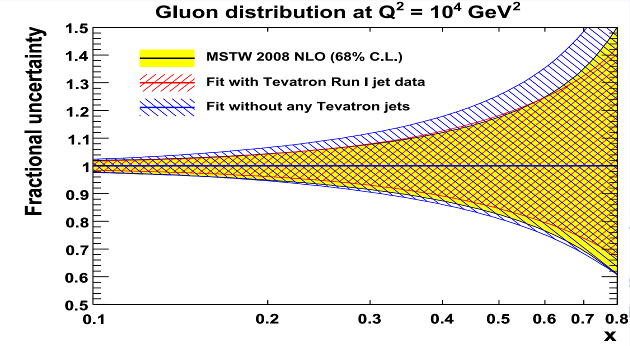
Constraints on PDF by ATLAS jet data (present and future)

Correlated systematic errors of jet data

Conclusions

Jet production at hadron colliders

- Huge event rate and significant uncertainties, test of QCD theory, measurement of QCD coupling constant and PDFs, especially of the gluon PDF in the large x region.



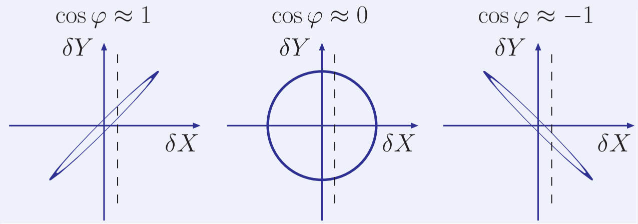
(A. Martin et al., arXiv:0901.0002)

Let's examine the sensitivity of the ATLAS jet data to $g(x, Q)$ using the PDF correlation analysis.

- Compute the PDF correlation cosine *(hep-ph/0101032; arXiv:0802.0007)*,

$$\cos\varphi \equiv \frac{1}{4\Delta X\,\Delta Y}\sum_{i=1}^{N_{PDF\ params}}\left(X_i^{(+)}-X_i^{(-)}\right)\left(Y_i^{(+)}-Y_i^{(-)}\right),$$

to establish which PDF $Y(x,Q)$ contributes most of the PDF uncertainty in the observable X

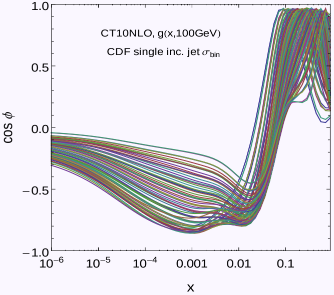


$\cos\varphi \approx \pm 1$: Measurement of X imposes tight constraints on Y

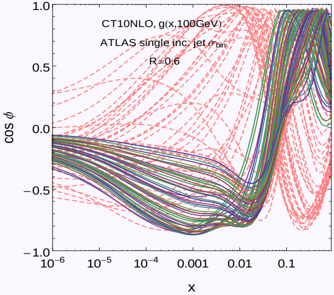
$\cos\varphi \approx 0$: Measurement of X imposes loose constraints on Y

The correlation cosine of jet cross sections and $g(x, Q)$:

p_T range: CDF (> 62 GeV), ATLAS (> 20 GeV)



(arXiv:0807.2204)

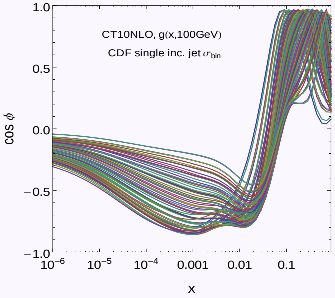


(arXiv:1112.6297)

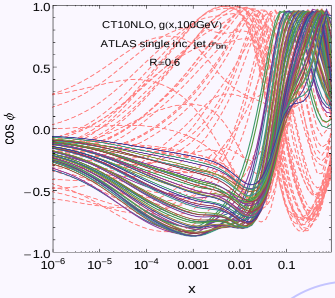
The curves show $\cos\varphi$ between the NLO theory cross sections in experimental p_T^j bins and $g(x, Q = 85 \text{ GeV})$, for various x values in $g(x, Q)$. One curve corresponds to one jet p_T bin. Look for the x value where $\cos\varphi$ is close to ± 1. The data bin is mostly sensitive to $g(x, Q)$ at this x value.

The correlation cosine of jet cross sections and $g(x, Q)$:

p_T range: CDF (> 62 GeV), ATLAS (> 20 GeV)



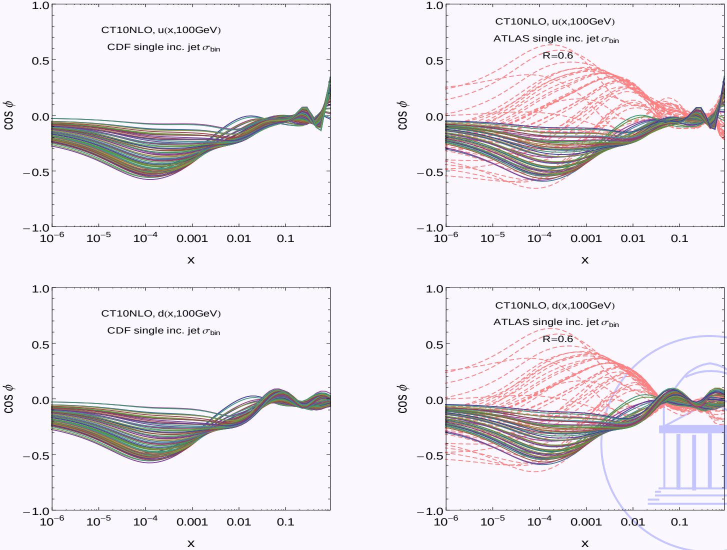
(arXiv:0807.2204)



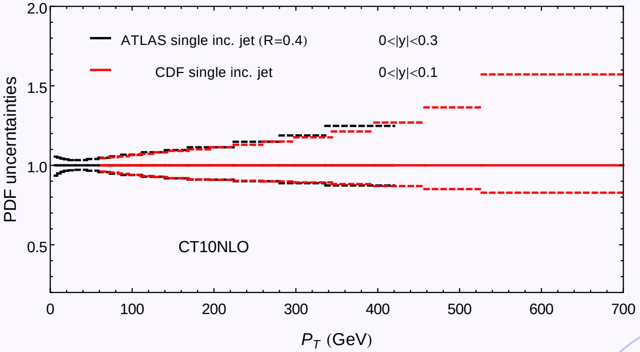
(arXiv:1112.6297)

The CDF (ATLAS) jet measurement are mostly sensitive to the gluon PDF with $x \gtrsim 0.1$ (0.01). In the ATLAS jet data, the bins with $p_T^j < 200$ GeV (pink dashed curves) probe $g(x, Q)$ in a wider range than CDF.

The correlation cosine of jet cross sections and $u(x,Q)$, $d(x,Q)$:

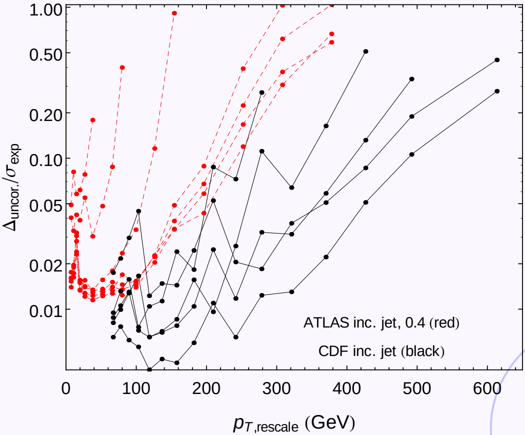


PDF uncertainties of single inclusive jet cross sections at NLO:



Overall features of the PDF uncertainty of the ATLAS jet data are very similar to the CDF jet data PDF uncertainty, if p_T^j of ATLAS data is rescaled down by 1.96/7 (similar x range and flavor composition). And the PDF error is $> 20\%$ at $p_T^j > 400\ \mathrm{GeV}$.

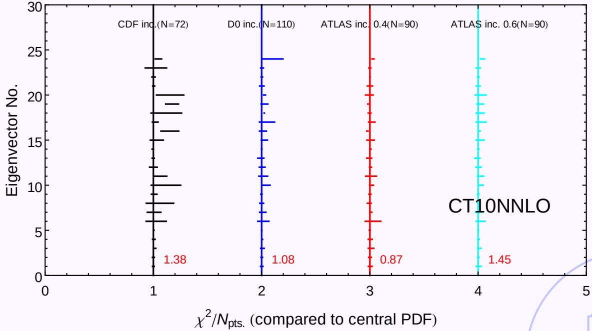
- Comparison of relative uncorrelated errors from CDF and ATLAS inclusive jet data (jet p_T rescaled by 1.96/7 for ATLAS data):



ATLAS data have less exp. points in the kinematic region that are sensitive to gluon PDF with $x \gtrsim 0.1$, and also larger uncorrelated errors in that region.

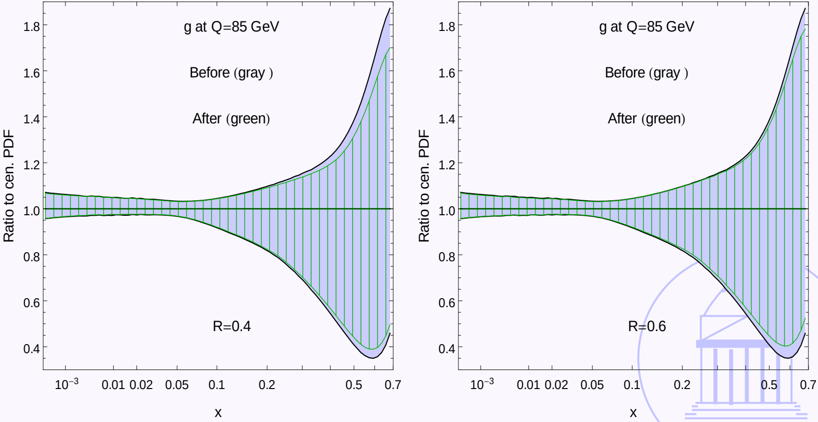
2010 ATLAS jet data in the global QCD analysis

- Before including the 2010 ATLAS jet data: χ^2 variations along different eigenvector directions (CT10 NNLO only use Tevatron inc. jet data, based on the Hessian method). *(arXiv:1007.2241)*



The χ^2 variations of the ATLAS data are much smaller than the ones of the CDF data included in the global fit, indicating that constraints on the gluon PDF after including the ATLAS jet data will not improve much.

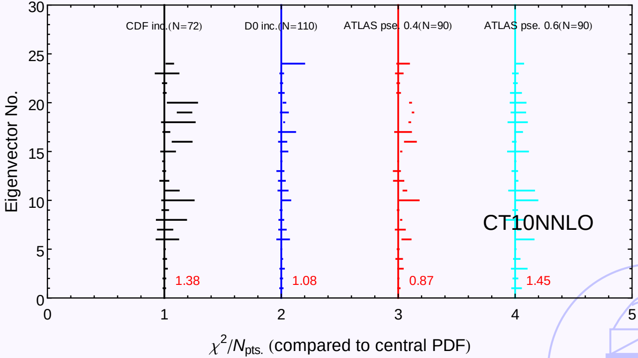
- CT10 NNLO fits before and after including the 2010 ATLAS jet data: no significant improvement in the error band (left: R=0.4, right: R=0.6).



90% C.L. PDF uncertainties normalized to the central PDF

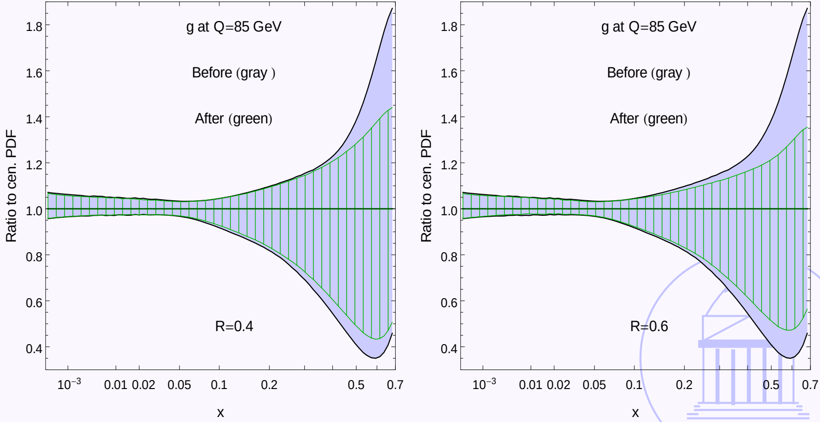
Potential impact of the future ATLAS data

- If the statistical errors of the ATLAS data are reduced by a factor of 10, and the JES errors are reduced by a factor of 2:



χ^2 of the pseudo-data shows larger sensitivity to $g(x, Q)$, indicating possible considerable improvements in the constraints of the gluon PDF.

- After including the pseudo-data: show significant reduction of the gluon PDF uncertainties in region $x \gtrsim 0.1$ (left: R=0.4, right: R=0.6).



90% C.L. PDF uncertainties normalized to the central PDF

Implementation of corr. syst. errors in the PDF analysis

- Log-likelihood χ^2 function with correlated systematic error *(hep-ph/0201195)*

$$\chi^2(\{a\},\{\lambda\}) = \sum_{k=1}^{N_{pts.}} \frac{1}{s_k^2}\left(D_k - T_k(\{a\}) - \sum_{\alpha=1}^{N_{cor.}} \lambda_\alpha \beta_{k\alpha}\right)^2 + \sum_{\alpha=1}^{N_{cor.}} \lambda_\alpha^2$$

- Analytical minimization of χ^2 with respect to the systematic shifts λ_α

$$\chi^2(\{a\},\{\lambda_0(a)\}) = d^T(\mathcal{I} + \beta\beta^T)^{-1} d, \quad d_k = D_k - T_k$$

  Theory data differences and β are normalized by the uncorrelated errors.

- Singular values decomposition method (SVD) can be used to find the principal component of the correlated systematic errors. Basic ideal is to rotate into some diagonal basis, thus can compare different combinations of correlated errors directly.

A new tool for the Correlation Error Matrix Analysis (CEMA)

- A mathematica package for general purpose comparisons of theory with correlated data (not only jets production).
- Standalone computation of χ^2 with experimental or theoretical corr. errors using an external theory calculation (FASTNLO, etc.).
- SVD analysis of the correlation matrix. Elimination of corr. error combinations with negligible effect on the PDFs (can be done outside of the fit).
- Various plots: histograms of best-fit syst. shifts, plots of theory vs. shifted data, a variety of statistical comparisons, plots of singular values of the correlation matrix.
- The CEMA package is publicly available by request.

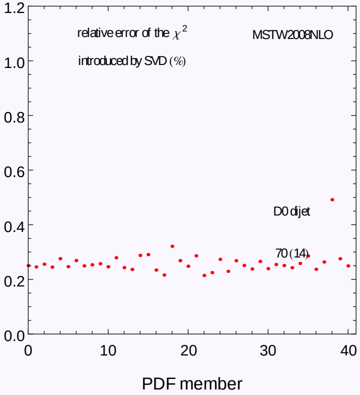
- $\chi^2/N_{pts.}$ of jet data ($\mu_R = \mu_F = p_{T,ind/ave}$, reduced lum. error):

$\chi^2/N_{pts.}$	CDF inc.	D0 inc.	D0 di.	ATL (0.4)	ATL (0.6)
$N_{pts.}(N_{cor.})$	72(25)	110(23)	71(70)	90(88)	90(88)
MSTW08	1.38	1.09	2.29	**0.94**	**0.83**
CT10	1.79	1.43	4.07	**0.98**	**0.98**
CT12(pre.)	1.38	1.08	2.68	**0.83**	**0.76**
ABKM09	2.02	1.31	1.47	**0.84**	**0.80**
HERAPDF1.5	3.08	1.73	2.09	**0.93**	**0.87**

The latest jet measurements provide $70 - 88$ sources of corr. syst. errors (compared to < 26 sources in earlier Tevatron measurements). With corr. sys. errors included, the ATLAS inc. jet data loses the power to discriminate between PDF sets.

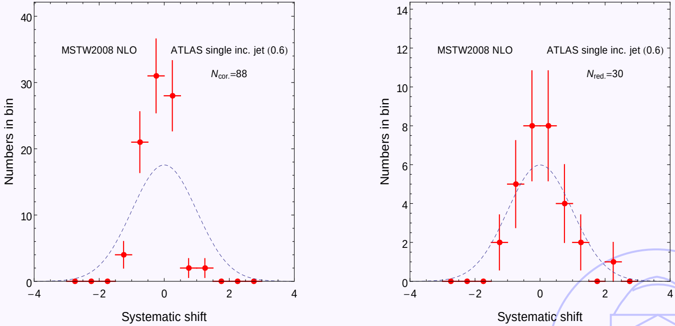
- One example of SVD reduction: D0 dijet data($N_{pts.} = 71, N_{cor.} = 70$)

(arXiv:1002.4594)



Only 14 combinations of the corr. sys. errors out of 70 make non-negligible contributions to the calculations of χ^2 of D0 dijet data, which indicates strong correlations or hierarchy structure of the errors.

- Distributions of the fitted systematic shifts λ_α. In principle, the true values of λ_α should follow a standard Gaussian distribution. But not always for the fitted values. Imaging that one correlated sys. error i is very small, thus the fitted value of λ_i would always be around 0.



As in the ATLAS single inc. jet case, for the full correlation matrix, most of the systematic shifts concentrate around 0. While using the SVD reduced correlation matrix, the distributions are closer to a standard Gaussian.

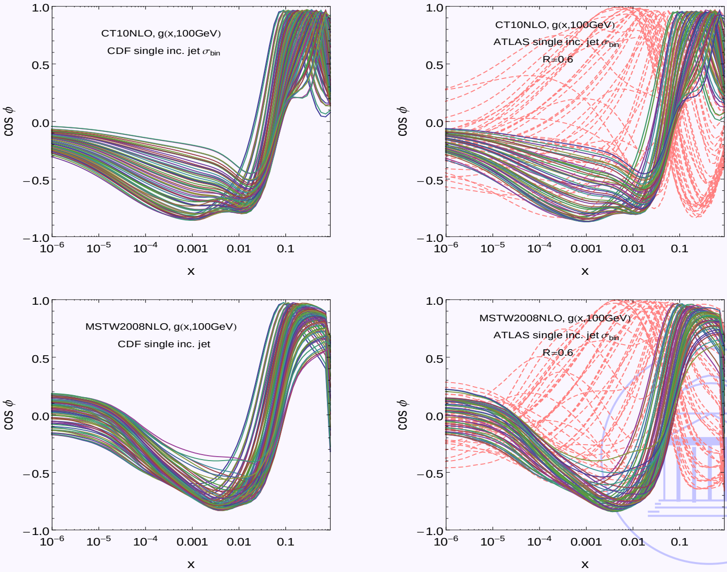
Conclusions

- Jet cross sections in kinematic regions of ATLAS measurement are sensitive to the gluon PDF uncertainties at large x values $x \gtrsim 0.1$, and also the gluon PDF uncertainties at medium x values $x \sim 0.01$, in combination with the light quark PDF uncertainties.
- Due to the large systematic and statistical errors, including current 2010 ATLAS single inc. jet data in the CT10 global fit does not significantly improve the constraints on the gluon PDF.
- Based on a toy model study, future ATLAS jet data will tighten the constraints on the gluon PDF.
- We developed a mathematica package CEMA for general purpose comparisons of theory with correlated data. The code is publicly available by request.

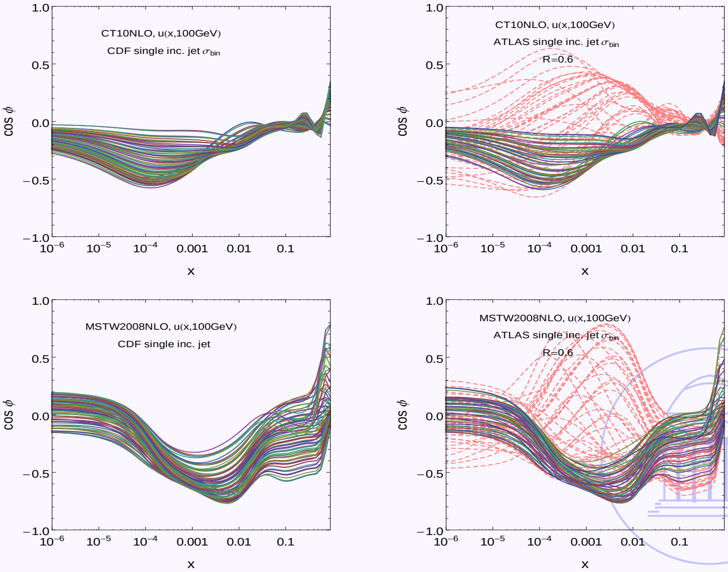
Backup

Exp.	observables	$N_{pts.}$	$N_{cor.}$	kinematics	lum.(fb^{-1})	ref.
Run2 CDF	inc.(cone)	72	25	62-700, 2.1	1.13	0807.2204
Run2 D0	inc.(cone)	110	23	50-635, 2.4	0.70	0802.2400
Run2 D0	dijet(cone)	71	70	150-1500, 2.4	0.70	1002.4594
ATLAS(7TeV)	inc.(0.4)	90	88	20-1500, 4.4	0.037	1112.6297
ATLAS(7TeV)	inc.(0.6)	90	88	20-1500, 4.4	0.037	1112.6297

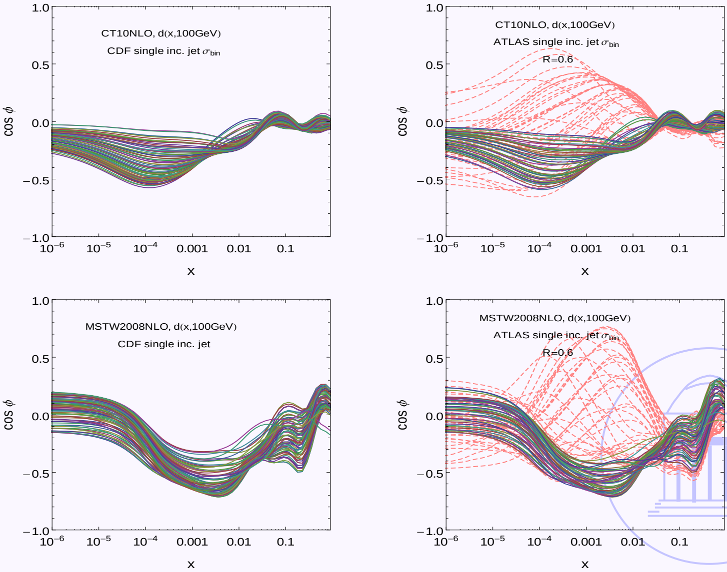
Backup


CT10NLO, g(x,100GeV)
CDF single inc. jet σ_bin
CT10NLO, g(x,100GeV)
ATLAS single inc. jet σ_bin
R=0.6
MSTW2008NLO, g(x,100GeV)
CDF single inc. jet
MSTW2008NLO, g(x,100GeV)
ATLAS single inc. jet σ_bin
R=0.6
cos ϕ
x
1.0
0.5
0.0
−0.5
−1.0
10−6
10−5
10−4
0.001
0.01
0.1

Backup

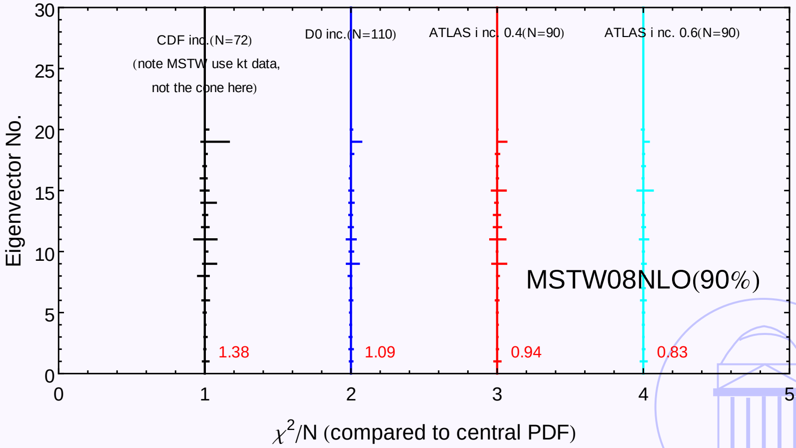

CT10NLO, u(x,100GeV)
CDF single inc. jet σ_{bin}
CT10NLO, u(x,100GeV)
ATLAS single inc. jet σ_{bin}
R=0.6
MSTW2008NLO, u(x,100GeV)
CDF single inc. jet
MSTW2008NLO, u(x,100GeV)
ATLAS single inc. jet σ_{bin}
R=0.6
cos ϕ
x

Backup

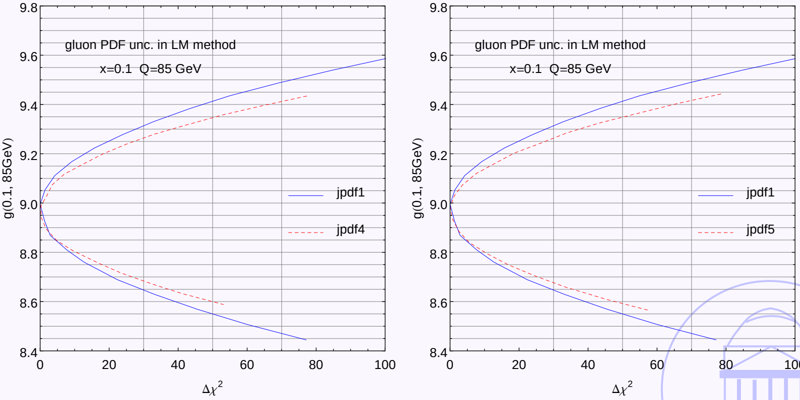


CT10NLO, d(x,100GeV)
CDF single inc. jet $\sigma_{\rm bin}$
CT10NLO, d(x,100GeV)
ATLAS single inc. jet $\sigma_{\rm bin}$
R=0.6
MSTW2008NLO, d(x,100GeV)
CDF single inc. jet
MSTW2008NLO, d(x,100GeV)
ATLAS single inc. jet $\sigma_{\rm bin}$
R=0.6
cos ϕ
x
1.0
0.5
0.0
−0.5
−1.0
10^{-6}
10^{-5}
10^{-4}
0.001
0.01
0.1

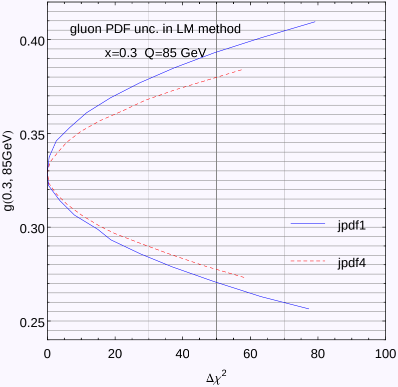
Backup

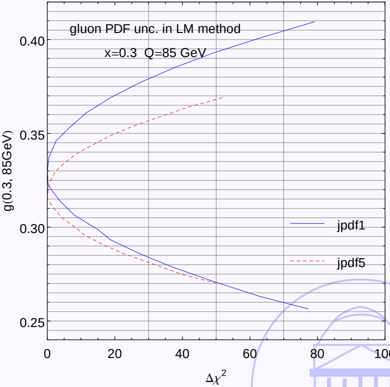

CDF inc.(N=72)
(note MSTW use kt data,
not the cone here)
D0 inc.(N=110)
ATLAS i nc. 0.4(N=90)
ATLAS i nc. 0.6(N=90)
MSTW08NLO(90%)
1.38
1.09
0.94
0.83
Eigenvector No.
χ^2/N (compared to central PDF)
0
5
10
15
20
25
30
1
2
3
4
5

Backup


gluon PDF unc. in LM method
x=0.1 Q=85 GeV
g(0.1, 85GeV)
$\Delta\chi^2$
jpdf1
jpdf4
gluon PDF unc. in LM method
x=0.1 Q=85 GeV
g(0.1, 85GeV)
$\Delta\chi^2$
jpdf1
jpdf5

Backup


gluon PDF unc. in LM method
x=0.3 Q=85 GeV
g(0.3, 85GeV)
0.25
0.30
0.35
0.40
0
20
40
60
80
100
$\Delta\chi^2$
jpdf1
jpdf4


gluon PDF unc. in LM method
x=0.3 Q=85 GeV
g(0.3, 85GeV)
0.25
0.30
0.35
0.40
0
20
40
60
80
100
$\Delta\chi^2$
jpdf1
jpdf5

Backup

- SVD method can be used to find the principal component of the correlated systematic errors.

$$\beta = U^T \beta' V, \quad \beta' = (\mathrm{DiagonalMatrix}[\sigma_i, \{i = 1, N_{cor.}\}], 0)^T$$

Correlations between different sys. errors, or errors with small contributions will show up as having $\sigma_i << 1$. Under the new basis,

$$\chi^2(\{a\}, \{\lambda_0(a)\}) = \sum_{i=1,N_{cor.}} d_i'^2/(1+\sigma_i^2) + \sum_{i=N_{cor.}+1,N_{pts.}} d_i'^2$$

Especially we can reduce the error matrix by dropping errors with $\sigma_i << 1$.

Backup

- Using SVD method to analysis χ^2 variations. Starting point:

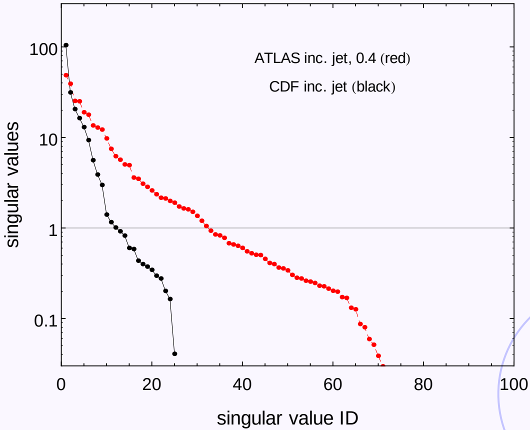
$$\chi^2(\{a\},\{\lambda_0(a)\}) = \sum_{i=1}^{N_{cor.}} d_i'^2/(1+\sigma_i^2) + \sum_{i=N_{cor.}+1}^{N_{pts.}} d_i'^2$$

Imaging that if there is no cor. errors, thus the χ^2 variation propagated through the change of theoretical predictions would be $\sim 2\delta_i/u_i$, where δ_i is the relative change of the theoretical cross sections, u_i is the uncorrelated errors normalized by theoretical cross sections. While including correlated systematic errors, it changes to $\sim 2\delta_i'/(u_i'(1+\sigma_i^2)^{1/2})$, where δ_i' and u_i' are the effective values under the SVD basis, σ_i are the singular values.

- Thus we identified two factors relevant to the χ^2 variations apart from the theoretical cross sections: relative uncorrelated errors, and singular values of the correlation matrix.

Backup

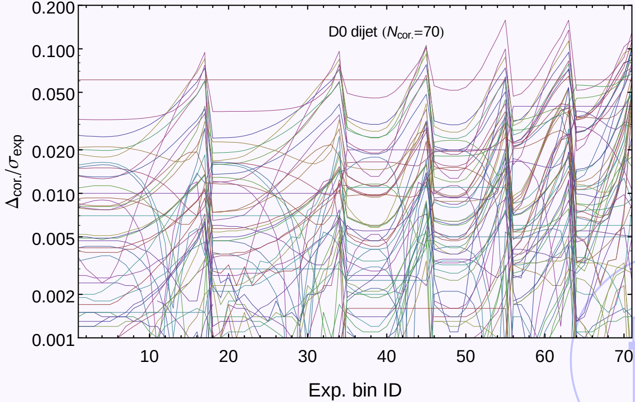
- Comparison of singular values of correlation matrix from CDF and ATLAS inclusive jet data:



ATLAS data have more larger singular values, 32 out of 88, than CDF data, 12 out of 25.

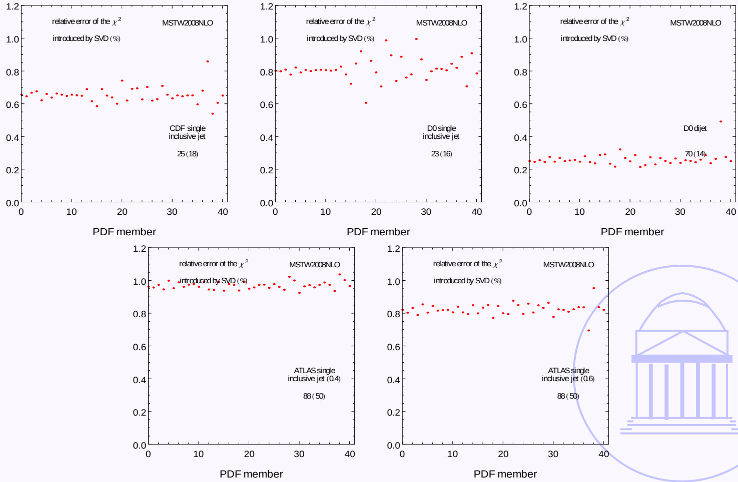
Backup

- Plot of the D0 dijet correlated errors



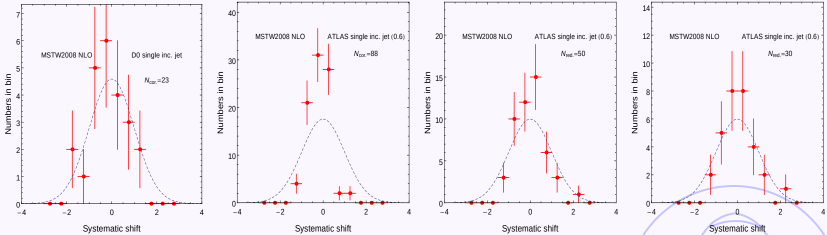
Backup

- Relative errors of χ^2 introduced by the SVD method when keeping $N_{red.}$ new correlated sys. errors instead of original $N_{cor.}$ ones.



Backup

- Another way to see the quality of the fit: Distributions of the fitted systematic shifts λ_α. In principle, the true values of λ_α should follow a standard Gaussian distribution. But not always for the fitted values. Imaging that one correlated sys. error i is very small, thus the fitted value of λ_i would always be around 0.



As in the ATLAS single inc. jet case, for the full correlation matrix, systematic shifts corresponding to smaller errors couldn't be fitted well. While using the SVD reduced correlation matrix, we only keep the well fitted values of λ_α, thus the distributions are closer to standard Gaussian.