1. PDF-driven correlations with LHC electroweak precision observables

2. CTEQ-TEA recommendations for LHC Drell-Yan measurements

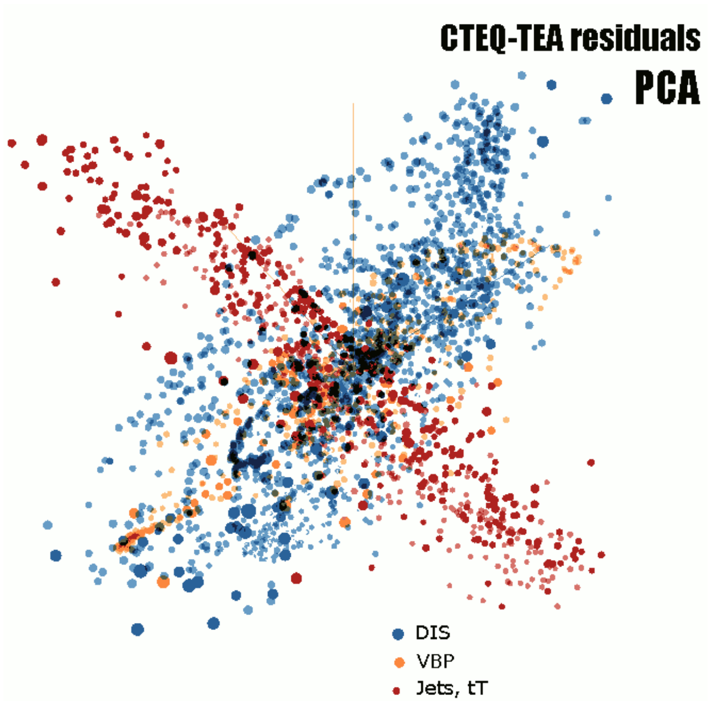
Pavel Nadolsky

Southern Methodist University

In collaboration with CTEQ-TEA group

PDFSense program, article, and figures:
Bo Ting Wang, Tim Hobbs, et al.
arXiv:1803.02777, published in PRD

http://tinyurl.com/PDFSense

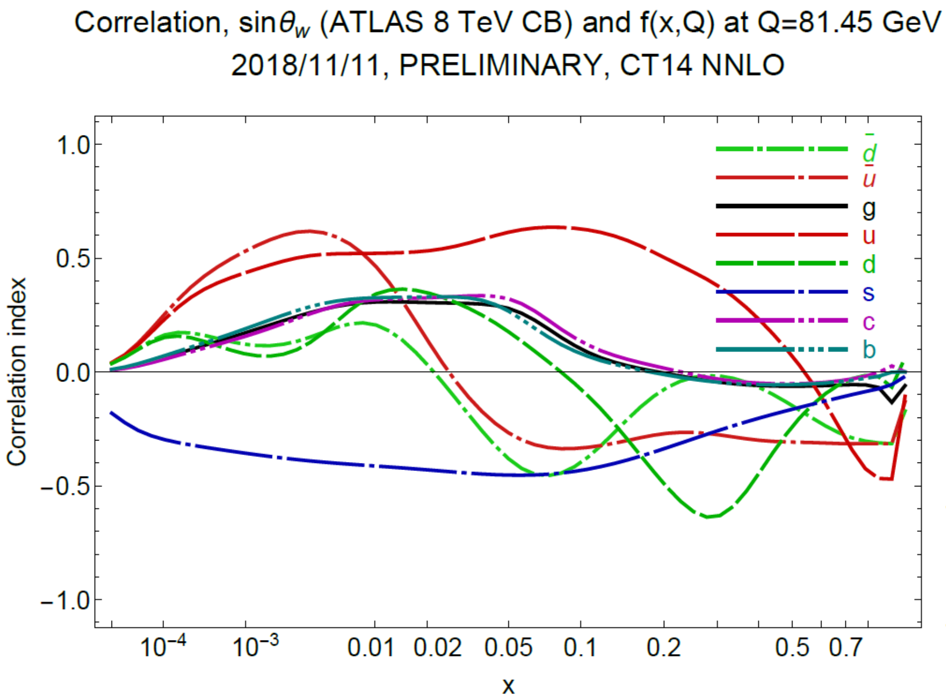


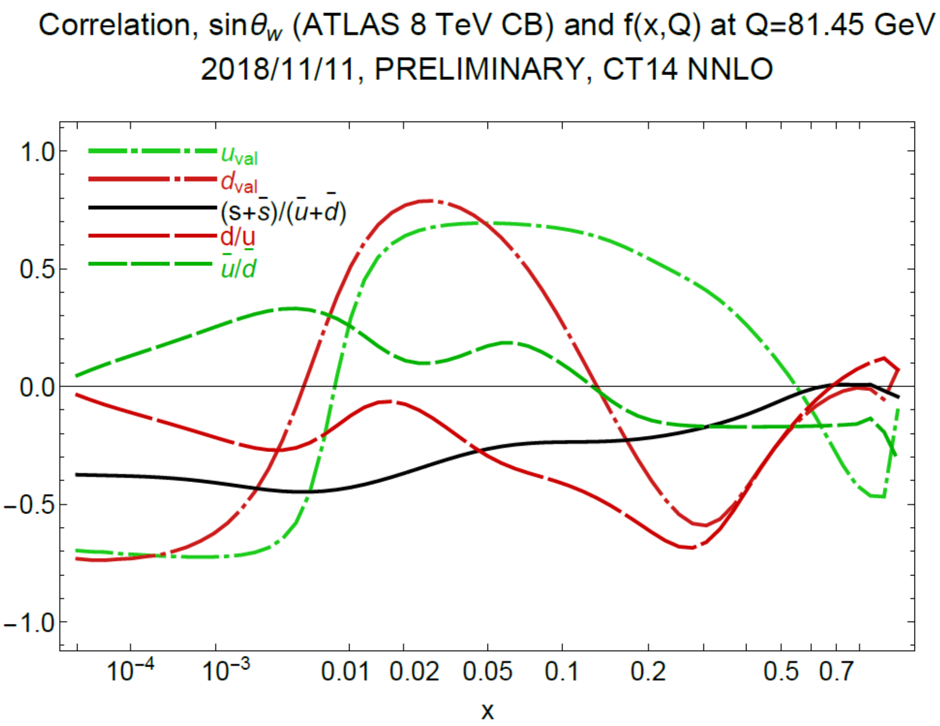
PDF constraints on $\sin^2\theta_w$ for ATLAS 8 TeV

Ongoing study

- Our goal is to identify experiments in the CT14 NNLO analysis that constrain the PDF uncertainty of $\sin^2\theta_w$ at ATLAS 8 TeV
- Available analysis methods:
  - A global fit including pseudodata on s2w – reliable, slow
  - Lagrange multiplier scan – most reliable; even slower; not available for CT14/CT10
  - PDF reweighting/Hessian profiling – ambiguity due to the definition of statistical weights
  - **This talk:** CT14 Hessian analysis of
    - **correlations** C_f of s2w with PDFs; and
    - **sensitivities** S_f of experiments to PDFs affecting s2w

Correlations of $\sin^2\theta_w$ with CT14 NNLO PDFs





Inputs

-s2w values obtained for 56+1 CT14 NNLO error PDFs [from A. Armbruster]

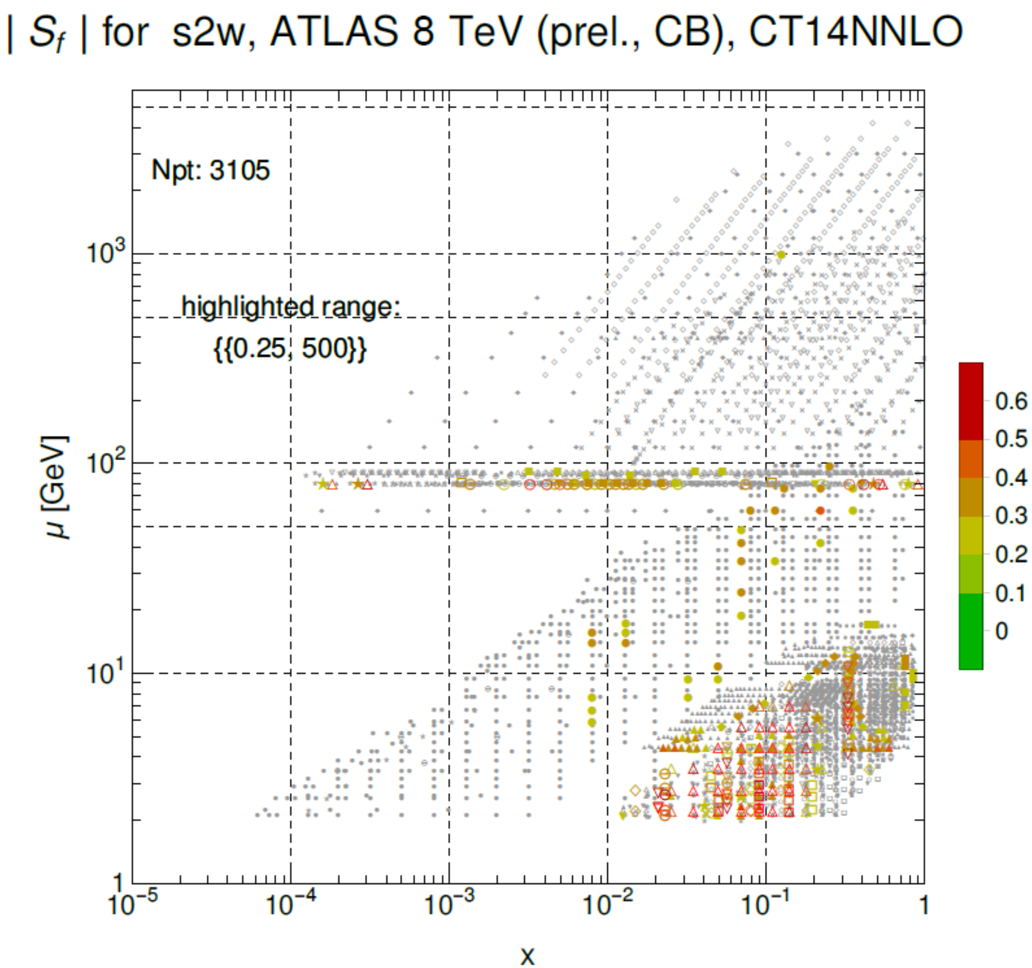
- CT14 NNLO PDF parametrizations

Outputs: $\cos\varphi$ for individual PDFs vs. x at $Q = 81.45$ GeV

$cos\varphi \approx \pm 1$ indicates a large (anti-)correlation of s2w with a given $f_a(x, Q)$

Strongest correlations with u_{val} (u), d_{val}, $\bar{u}$; weak correlations with $\bar{u}$, $\bar{\mathrm{d}}$, $\bar{s}$, g

Sensitivity of CT14 experiments to s2w



No significant difference between CB and CBF s2w samples

Based on the PDFSense analysis, the most sensitive CT14 data sets to s2w are

- **combined HERA1 DIS [most sensitive]**
- CCFR νp DIS $F_{3,2}$
- BCDMS $F_2^{p,d}$
- NMC ep, ed DIS
- CDHSW νA DIS
- NuTeV $\nu A \to \mu\mu X$
- CCFR $\nu A \to \mu\mu X$
- E866 $pp \to \ell^+\ell^- X$
- ATLAS 7 TeV W/Z (35 pb^{-1})
- …

How sensitive is an experiment to a PDF? Can we know it **before** doing the global fit?

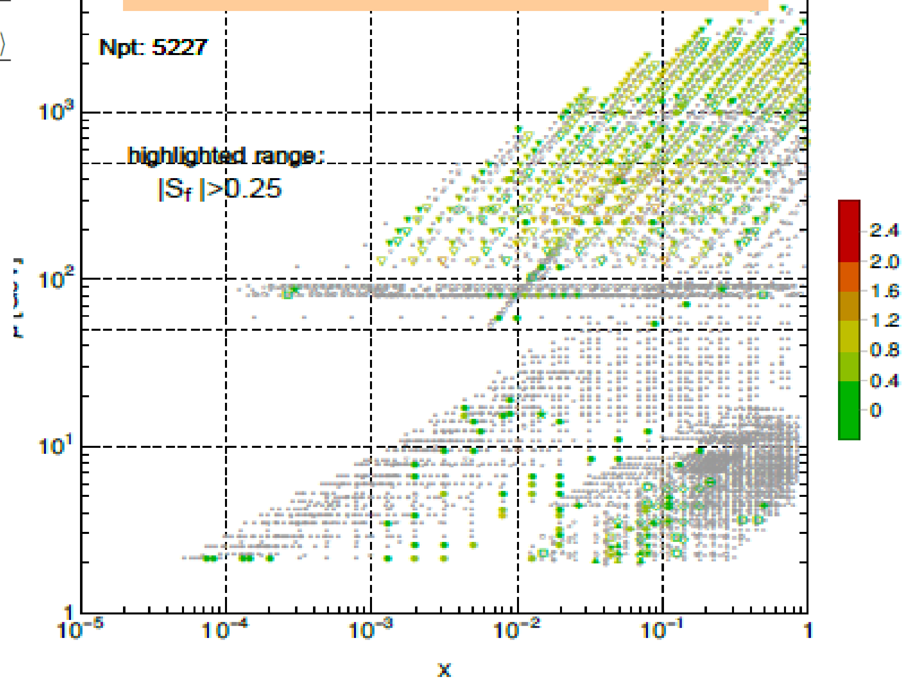
PDFSense estimates…

…ranking of strength of sensitivities of experimental data sets to PDF flavors without (re-)doing the full global fit

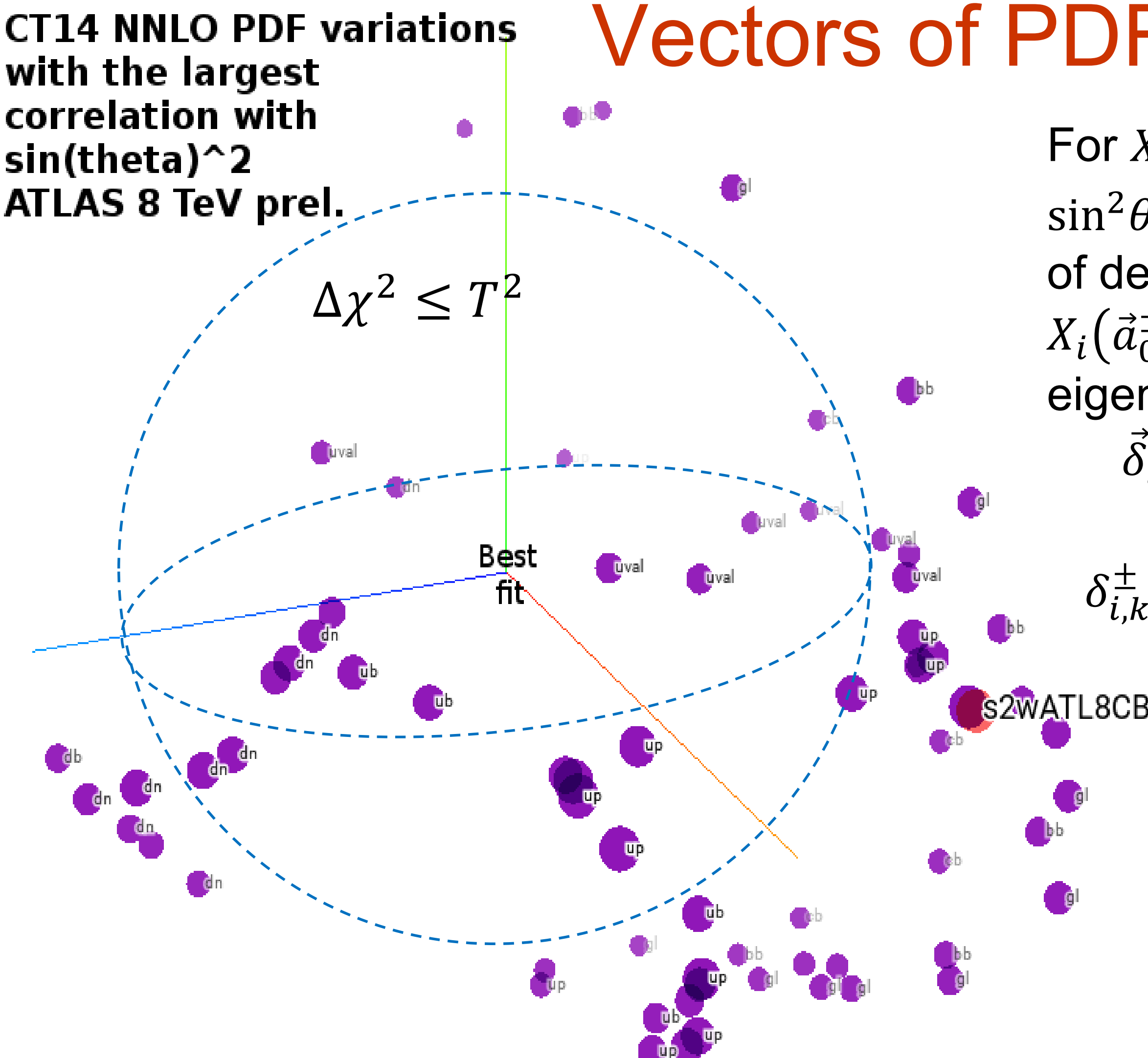
No.	Exp. ID	N_d	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar d}^E \rvert$	$\langle \lvert S_{\bar d}^E \rvert \rangle$	$\lvert S_{\bar u}^E \rvert$	$\langle \lvert S_{\bar u}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
			Rankings													
1	160	1120.	620.	0.0922	B		A	3	A	3	A	3	B		C	
2	**545**	185	232.	0.209	C	3	C	3	B	2					C	3
3	111	86	218.	0.423	C	1	C	1		3	B	1	C	2		
4	**542**	158	194.	0.204	C	3	C	3	B	2					C	3
5	101	337	184.	0.0909			C		C		B	3	C			
6	104	123	169.	0.229	C	2					C	2	B	2		
7	102	250	141.	0.0938	C				C	3	C	3	C	3		
8	109	96	115.	0.199	C	2	C	2		3	C	2	C	3		
9	201	119	113.	0.158	C	2	C	2				3				
10	204	184	103.	0.0935		3	C	3			C	3				
11	110	69	89.3	0.216		3		3	C	2		3		2		3
12	108	85	82.4	0.161		3		3		3		3	C	3		
13	538	133	66.2	0.0829					C	3						
14	124	38	58.9	0.258		3		3				3		3	C	1
15	127	38	49.4	0.217		3		3				3		3	C	1
16	**544**	140	48.7	0.058						3						
17	126	40	48.	0.2		3		3				3		3	C	1
18	**250**	42	41.5	0.165		3		3		3		3		2		
19	268	41	39.6	0.161		3		3				3		3		3
20	**249**	33	39.2	0.198		2		3				3		2		3
21	514	110	36.8	0.0557						3						
22	125	33	36.7	0.185		3		3				3		3		2

…kinematical distributions of sensitivities to the PDFs in the $\{x, \mu\}$ plane

Sensitivity to the PDF error on $\sigma(pp \to H^0 X)$ at 14 TeV



Vectors of PDF uncertainties



For $X_i(\vec{a}_i^{\pm}) \equiv f_{A_i}(x_i, Q_i; \vec{a}_i^{\pm})$ or $\sin^2\theta_w(\vec{a}_i^{\pm})$, construct a vector $\vec{\delta}_i$ of deviations from the best fit $X_i(\vec{a}_0^{\pm})$ for 2N Hessian eigenvectors.

$$\vec{\delta}_i = \{\delta_{i,1}^{+}, \delta_{i,1}^{-}, \ldots, \delta_{i,N}^{+}, \delta_{i,N}^{-}\}$$

$$[\mathrm{N} = 28 \text{ for CT14 NNLO}]$$

$$\delta_{i,k}^{\pm} \equiv \left(X_i(\vec{a}_k^{\pm}) - X_i(\vec{a}_0)\right) / X_i(\vec{a}_0)$$

A 3-dim projection of 56-dim PDF vectors for $f_{A_i}(x_i, Q_i)$ with the smallest angular distance from the $\sin^2\theta_w(\vec{a}_i^{\pm})$ vector; $10^{-5} \le x_i \le 0.8$; $Q_i = 100$ GeV

Tolerance hypersphere in the PDF space

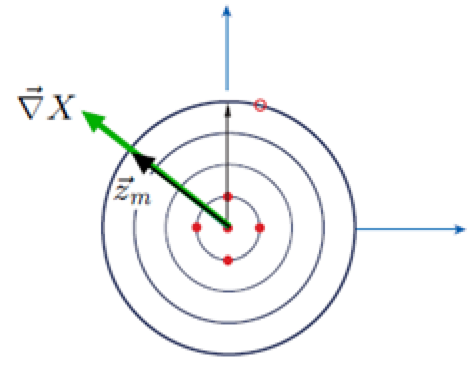
2-dim (i,j) rendition of N-dim (26) PDF parameter space

Hessian method: Pumplin et al., 2001

A symmetric PDF error for a physical observable X is given by

$$\Delta X = \vec{\nabla} X \cdot \vec{z}_m = \left|\vec{\nabla} X\right|$$

$$= \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)^2}$$

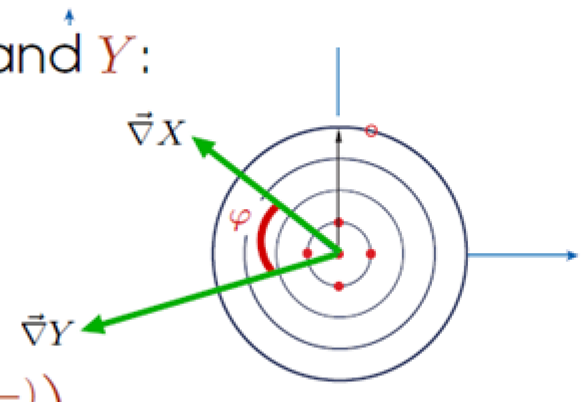


(b)
Orthonormal eigenvector basis

Correlation cosine for observables X and Y:

$$\cos\varphi = \frac{\vec{\nabla} X \cdot \vec{\nabla} Y}{\Delta X \Delta Y} =$$

$$\frac{1}{4\Delta X\,\Delta Y}\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right)$$



(b)
Orthonormal eigenvector basis

Vectors of data residuals

For every data point i, construct a vector of residuals $r_i(\vec{a}_k^{\pm})$ for 2N Hessian eigenvectors. $\mathrm{k} = 1, \ldots, N$, with $N = 28$ for CT14 NNLO:

$$\vec{\delta}_i = \{\delta_{i,1}^{+}, \delta_{i,1}^{-}, \ldots, \delta_{i,N}^{+}, \delta_{i,N}^{-}\} \qquad [\mathrm{N} = 28]$$

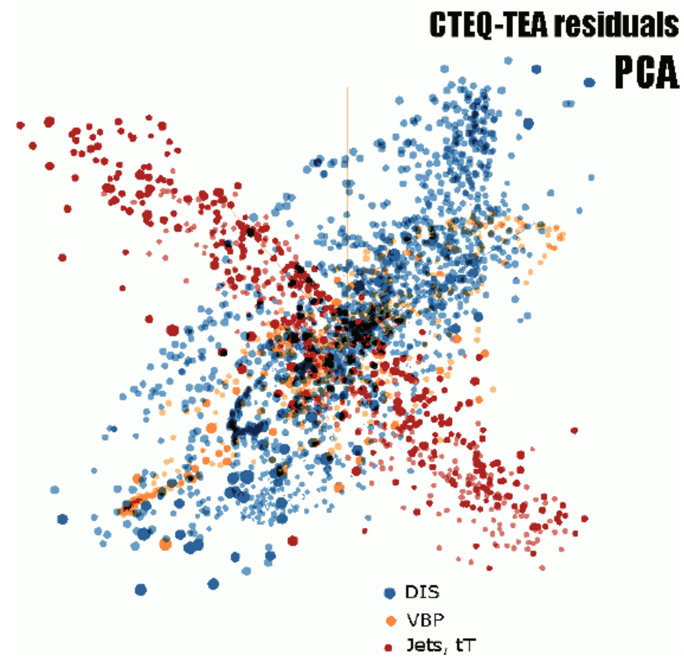
$$\delta_{i,k}^{\pm} \equiv \left(r_i(\vec{a}_k^{\pm}) - r_i(\vec{a}_0)\right) / \langle r_0 \rangle_E$$

-- a 56-dim vector normalized to $\langle r_0 \rangle_E$, the root-mean-squared residual for the experiment E for the central fit $\vec{a}_0$

$$\langle r_0 \rangle_E \equiv \sqrt{\frac{1}{N_{pt}} \sum_{i=1}^{N_{pt}} r_i^2(\vec{a}_0)} \approx \sqrt{\frac{\chi_E^2\,(\vec{a}_0)}{N_{pt}}}$$

$\langle r_0 \rangle_E \approx 1$ in a good fit to E

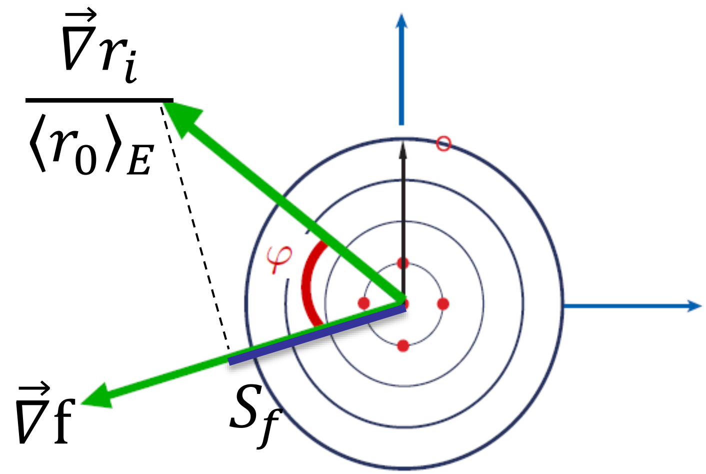
r_i is defined in the backup



The TensorFlow Embedding Projector (http://projector.tensorflow.org) represents CT14HERA2 $\vec{\delta}_i$ vectors by their 10 principal components indicated by scatter points. A sample 3-dim. projection of the 56-dim. manifold is shown above. A symmetric 28-dim. representation can be alternatively used.

Correlation C_f and sensitivity S_f

The relation of data point i on the PDF dependence of f can be estimated by:



- $C_f \equiv \mathrm{Corr}[\rho_i(\vec{a})), f(\vec{a})] = cos\varphi$

$\vec{\rho}_i \equiv \vec{\nabla} r_i / \langle r_0 \rangle_E$ -- gradient of r_i normalized to the r.m.s. average residual in expt E;

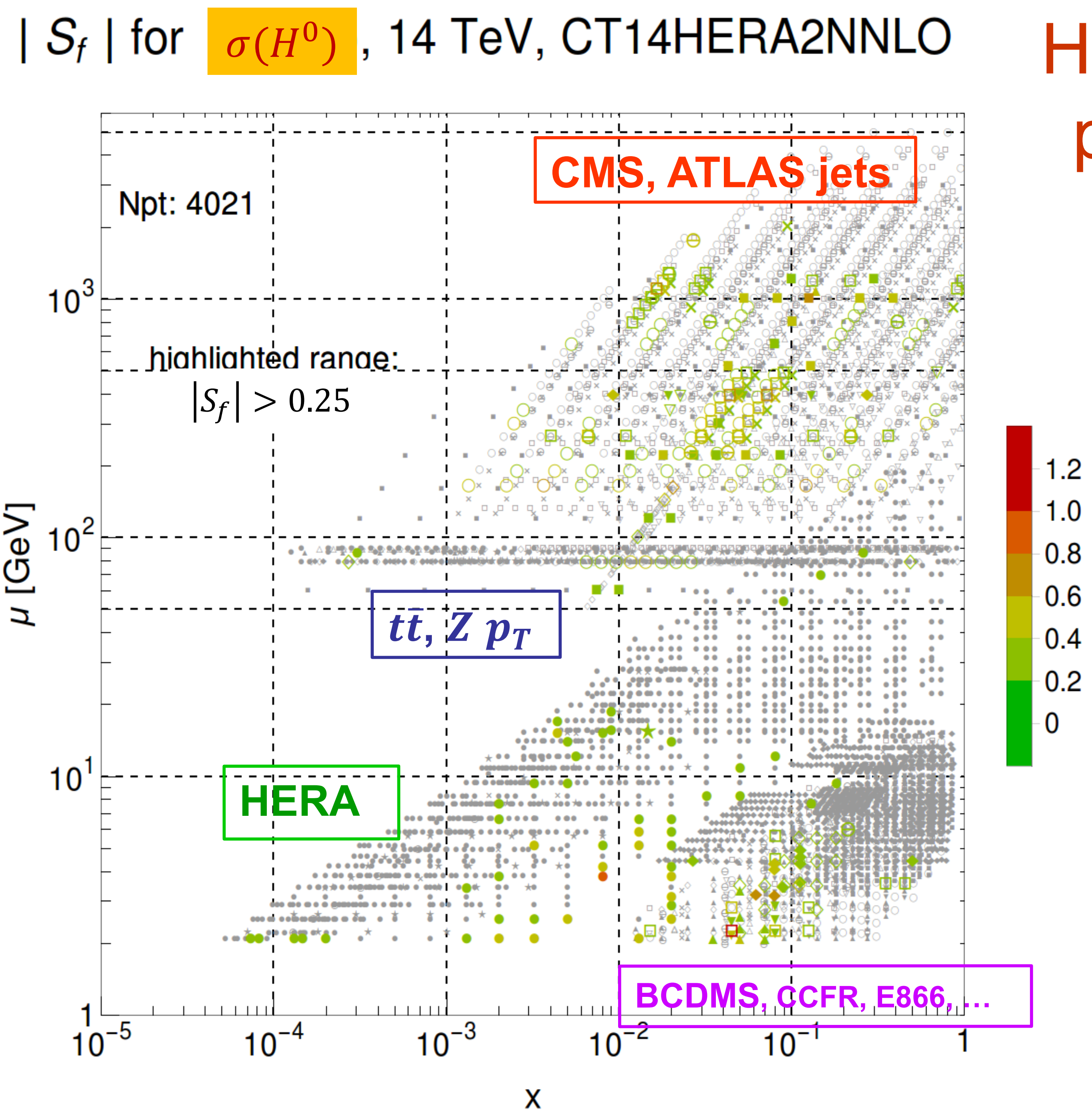
$$\left(\vec{\nabla} r_i\right)_k = \left(r_i(\vec{a}_k^+) - r_i(\vec{a}_k^-)\right)/2$$

C_f is **independent** of the experimental and PDF uncertainties. In the figures, take $\left|C_f\right| \gtrsim 0.7$ to indicate a large correlation.

- $S_f \equiv |\vec{\rho}_i| cos\varphi = C_f \frac{\Delta r_i}{\langle r_0 \rangle_E}$ -- projection of $\vec{\rho}_i(\vec{a})$ on $\vec{\nabla} f$

S_f is proportional to $\cos\varphi$ and the ratio of the PDF uncertainty to the experimental uncertainty. We can sum $|S_f|$.
In the figures, take $\left|S_f\right| > 0.25$ to be significant.

Higgs boson production



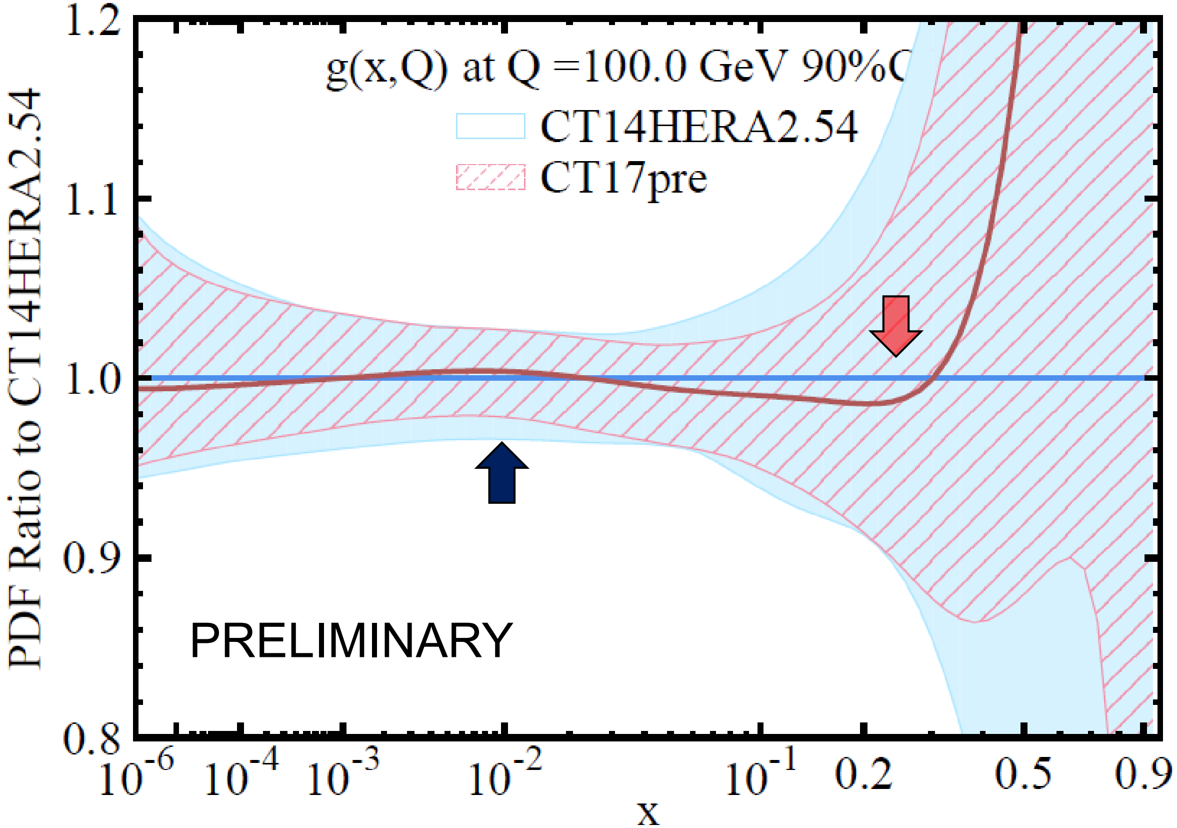
HERA DIS still has the dominant sensitivity!

CMS 8 TeV jets is the next expt. after HERA sensitive to $\sigma_H(14$ TeV); jet scale uncertainty dampens $|S_f|$ for jets

Good correlations C_f with some points in E866, BCDMS, CCFR, CMS WASY, $Z\ p_T$ and $t\bar{t}$ production; but not as many points with high $|S_f|$ in these processes

Gluon PDF before and after including the LHC data

[CT14HERA2 vs. CT17pre NNLO]



$x \approx 0.01$: $g(x, Q)$ mildly increases within the uncertainty

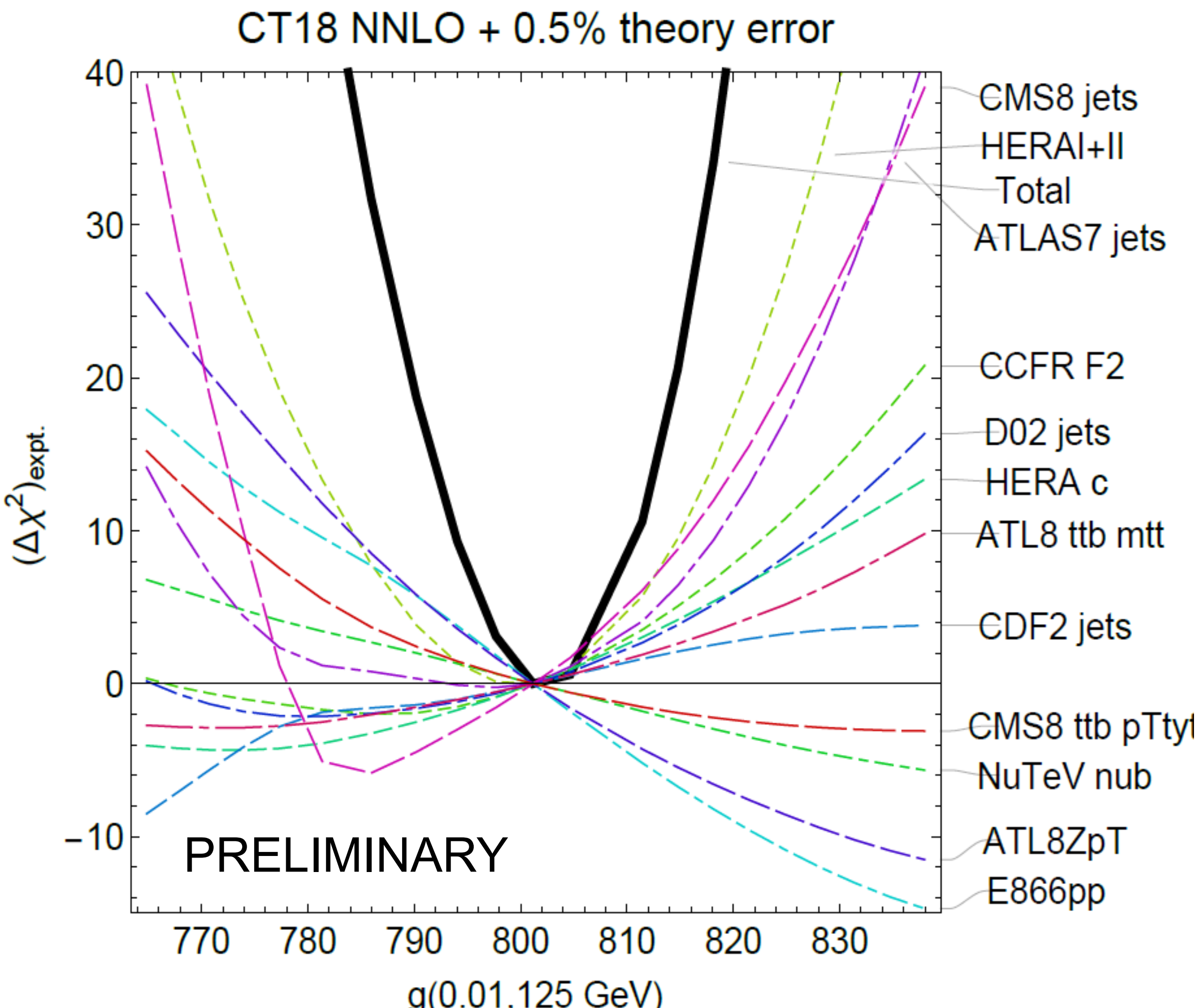
$\Rightarrow$ slightly larger Higgs production rates at 14 TeV

Minor reduction in the gluon PDF uncertainty

$0.05 \lesssim x \lesssim 0.3$: $g(x, Q)$ mildly decreases; lower gg luminosities for $M_X > 700$ GeV

After the fit

Which experiments constrain the gluon? $x = 0.01, Q = 125$ GeV [Higgs region]



The LM scans broadly confirm S_f estimates

HERAI+II, ATLAS7 jets, CMS8 jets impose the tightest constraints; are in agreement

E866, ATLAS 8 Z p_T prefer higher gluon

After the fit

Rankings of experiments most sensitive to $g(0.01, 125\ \text{GeV})$

24

$d/u(x=0.1, \mu=1.3\ \text{GeV})$			$g(x=0.01, \mu=125\ \text{GeV})$		
PDFSENSE		LM scan	PDFSENSE		LM scan
CT14HERA2	CT18pre	CT18pre	CT14HERA2	CT18pre	CT18pre
HERAI+II'15	NMCrat'97	NMCrat'97	HERAI+II'15	HERAI+II'15	HERAI+II'15
BCDMSp'89	HERAI+II'15	CCFR-F3'97	CMS8jets'17	CMS8jets'17	CMS8jets'17
NMCrat'97	BCDMSp'89	HERAI+II'15	CMS7jets'14	CMS7jets'14	ATL8ZpT'16
CCFR-F3'97	CCFR-F3'97	BCDMSd'90	ATLAS7jets'15	E866pp'03	E866pp'03
E866pp'03	BCDMSd'90	BCDMSp'89	E866pp'03	ATLAS7jets'15	ATLAS7jets'15
BCDMSd'90	E605'91	CDHSW-F3'91	BCDMSd'90	BCDMSd'90	CCFR-F2'01
CDHSW-F3'91	E866pp'03	E866rat'01	CCFR-F3'97	BCDMSp'89	D02jets'08
CMS8jets'17	E866rat'01	CMS7Masy2'14	D02jets'08	D02jets'08	HERAc'13
E866rat'01	CMS8jets'17	NuTeV-nu'06	NMCrat'97	NMCrat'97	NuTeV-nub'06
LHCb8WZ'16	CDHSW-F3'91	CMS8jets'17	BCDMSp'89	CDHSW-F2'91	CCFR-F3'97

TABLE I: We list the top 10 experiments predicted to drive knowledge of the d/u PDF ratio and of the gluon distribution in the Higgs region according to PDFSENSE and LM scans. For both, we list the PDFSENSE evaluations based both on the CT14HERA2 fit and on a preliminary CT18pre fit in the first and second columns on either side of the double-line partition.

PDFSense identifies the most sensitive experiments with high confidence and in accord with other methods such as the LM scans. It works the best when the uncertainties are nearly Gaussian, and experimental constraints agree among themselves [arXiv:1803.02777, v.3]

CTEQ-TEA recommendations for LHC DY measurements DRAFT, 2018-11-13, page 1

1. CT18 NNLO or CT14HERA2 NNLO
2. CT18 fits find contradictory preferences for strangeness $x \geq 10^{-3}$ between fitted (SI)DIS experiments, on one hand, and some LHC experiments, especially ATLAS W/Z production measurements and to some extent LHCb W/Z measurements. Benchmarking of LHC measurements and theoretical predictions, as well as new (SI)DIS experiments can be highly effective for resolving these tensions.
3. CT18 NNLO uses NNLO predictions from FEWZ for … and NNLO/NNLL Resbos for ...; NNLOJET K-factors for inclusive jet production, fastNNLO tables for tT productions. Parton shower effects are very limited, especially when NNLO predictions are used.
4. Alternative candidate fits of the CT18 NNLO analysis estimate the QCD scale and numerical uncertainties in high-p_T Z production. In our opinion, NNLO theoretical uncertainties are under good control in the fitted region $50 < p_{TZ} < 150$ GeV of the high-p_T Z production data in the CT18 NNLO analysis.
5. The photon PDFs do not significantly affect the inclusive QCD observables included in the CT18 NNLO analysis.

CTEQ-TEA recommendations for LHC DY measurements DRAFT, 2018-11-13, page 2

6. When it is relevant, QCD predictions using CT18/CT14 PDFs must use the SACOT-chi scheme and the same charm and bottom mass values as those used to fit the CT18 PDFs. For the LHC observables with all scales much larger than the c, b masses, the S-ACOT-chi hard cross section coincides with the zero-mass MSbar hard cross section. On the other hand, the mass effects may be relevant in W/Z p_T distributions in c, b channels at $p_T^2 \lesssim m_{c,b}^2$. A comprehensive study of the power-suppressed/intrinsic/fitted charm distribution is published in **JHEP 1802 (2018) 059 / arXiv:1707.00657.** CTEQ-TEA does not see it mandatory to use the fitted charm parametrizations throughout. The PDFs with fitted charm such as CT14 IC or NNPDF3.1 do not provide a better theoretical framework than the standard CT14 PDFs. A large part of the fitted charm PDF may arise from twist-4 contributions that are unique to low-Q DIS.
7. The TMD effects are negligible in the recent CTEQ-TEA analyses.
8. No, various kinds of parametrization and methodological uncertainties are accounted for in the CTEQ-TEA PDF errors and are studied regularly as a part of the CTEQ-TEA analysis.

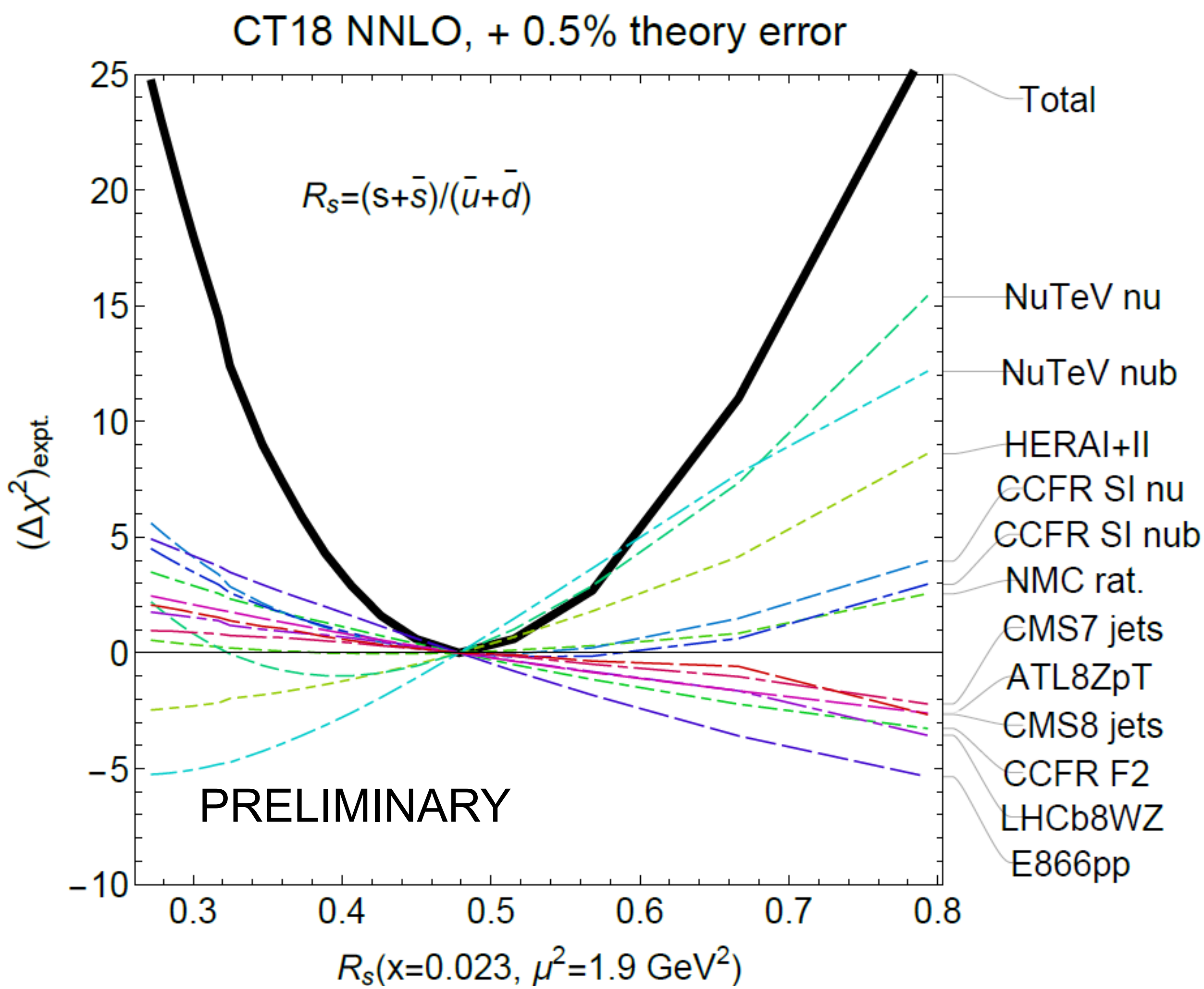
CTEQ-TEA recommendations for LHC DY measurements DRAFT, 2018-11-13, page 3

9. As of 2018, we do not recommend to fit the PDFs only to the LHC or DY data. The most significant constraints arise from other experiments, such as fixed-target DIS. It is ok to perform this type of study with a reduced number of data sets as a benchmarking exercise among the PDF groups, but the resulting PDFs will not be as accurate/precise as the global PDF fits.
10. To a great degree, the important uncertainties, those due to the experimental errors of the datasets included in the fit, are already completely correlated. Correlation of other issues, such as parameterizations/scale choices can be studied.
11. If the PDF sets include the data, but do not agree with the data, and the other PDF sets do, then it is crucial to understand the source of the disagreement.
12. If the measurements do not have clearly defined systematic errors (in the modern sense), then it is justified to not use them in a global PDF fit. If the data sets are in strong tension with the other data sets used in a global fit, then they can be excluded. Of course, this happens on a case-by-case basis.

CTEQ-TEA recommendations for LHC DY measurements DRAFT, 2018-11-13, page 4

13. The Hessian and MC approaches are complementary. In recent years, the PDF groups have gained a great deal of experience in converting between Hessian and MC replica PDFs, strengthening the understanding of both. The Hessian PDFs are sufficient for the majority of estimates of PDF uncertainty in the case of sufficient experimental constraints. The MC error PDFs are useful in the case of weak experimental constraints or persistent non-Gaussian effects.
14. Conceptual foundations of PDF reweighting have not been explored sufficiently, which may result in its spurious applications. This area needs additional exploration before PDF reweighting can be safely used in high-stake situations such as in item 11.

Effect of LHC data on strangeness: the CT18pre fit



Some tension between NuTeV, CCFR dimuon production, HERAI+II (preferring $R_s < 0.6$);

and vector boson production at the LHC and Tevatron (preferring $R_s > 0.6$)

However, still large uncertainties

CT14 PDFs with HERA1+2 (=HERA2) combination

Phys.Rev. D95 (2017) 034003

Separate the four HERA2 DIS processes;
(Q_{cut} = 2 GeV)

	N_{pts}	$\chi^2_{red.} / N_{pts}$
NC e^+p	880	1.11
CC e^+p	39	1.10
NC e^-p	159	1.45
CC e^-p	42	1.52
totals		
[reduced χ^2] /N	1120	1.17
χ^2 / N	1120	1.25
R^2 / N	1120	0.08

e^+p data are fitted fine

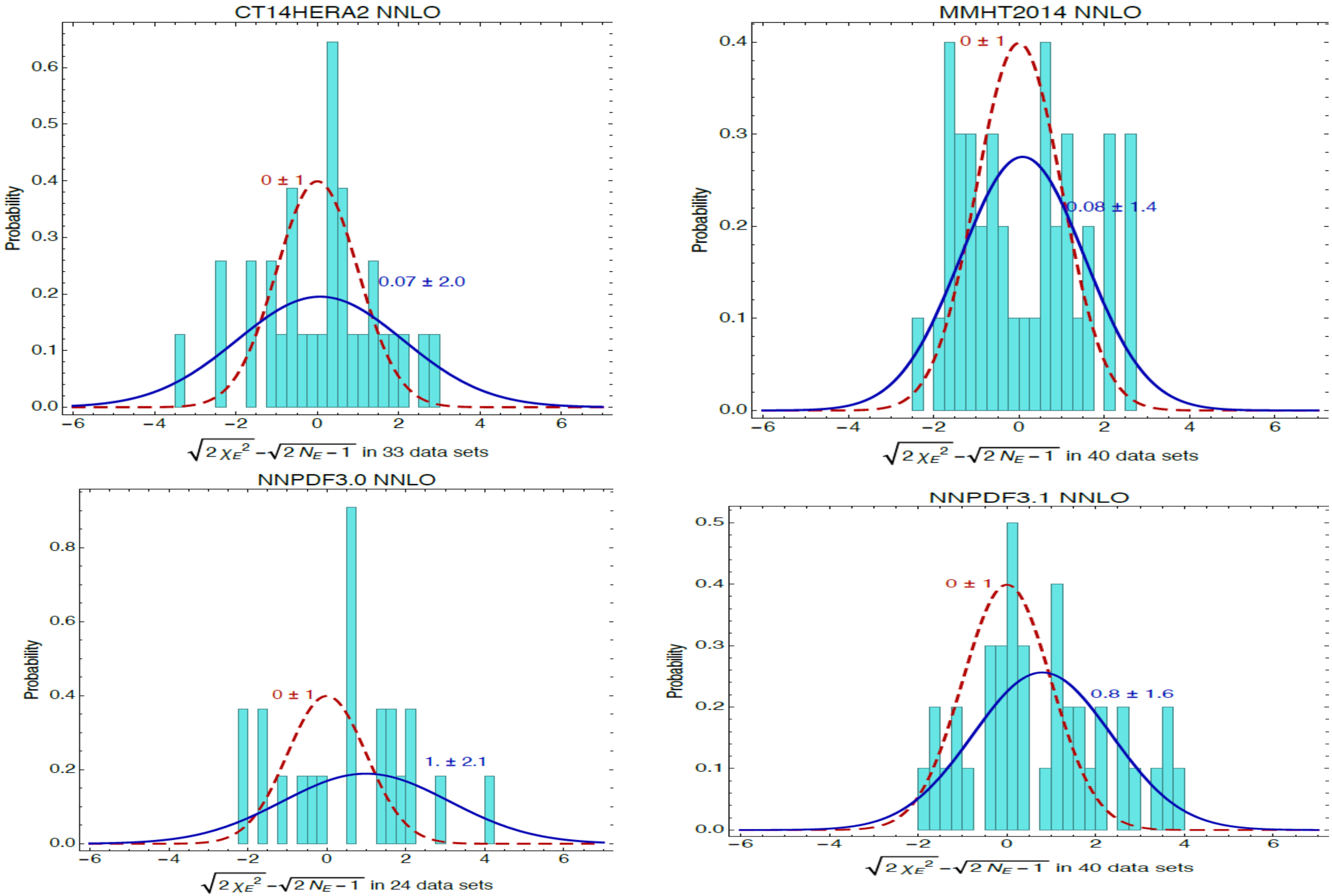
e^-p data are fitted poorly

reduced χ^2 values

χ^2 = [reduced χ^2] + R^2

The quadratic penalty for 162 systematic errors = 87.5

Tolerance of ≥ 2 is required to reconcile experiments in all global PDF fits



Extra details

Experiments in the CT14 HERA2 fit

ID#	Experimental dataset		N_d
101	BCDMS F_2^p	[47]	337
102	BCDMS F_2^d	[48]	250
104	NMC F_2^d/F_2^p	[49]	123
108	CDHSW F_2^p	[50]	85
109	CDHSW F_3^p	[50]	96
110	CCFR F_2^p	[51]	69
111	CCFR xF_3^p	[52]	86
124	NuTeV $\nu\mu\mu$ SIDIS	[40]	38
125	NuTeV $\bar{\nu}\mu\mu$ SIDIS	[40]	33
126	CCFR $\nu\mu\mu$ SIDIS	[41]	40
127	CCFR $\bar{\nu}\mu\mu$ SIDIS	[41]	38
145	H1 σ_r^b (57.4 pb^{-1})	[53][54]	10
147	Combined HERA charm production (1.504 fb^{-1})	[39]	47
160	HERA1+2 Combined NC and CC DIS (1 fb^{-1})	[6]	1120
169	H1 F_L (121.6 pb^{-1})	[55]	9

ID#	Experimental dataset		N_d
201	E605 DY	[56]	119
203	E866 DY, $\sigma_{pd}/(2\sigma_{pp})$	[57]	15
204	E866 DY, $Q^3 d^2\sigma_{pp}/(dQdx_F)$	[58]	184
225	CDF Run-1 $A_e(\eta^e)$ (110 pb^{-1})	[59]	11
227	CDF Run-2 $A_e(\eta^e)$ (170 pb^{-1})	[60]	11
234	DØ Run-2 $A_\mu(\eta^\mu)$ (0.3 fb^{-1})	[61]	9
240	LHCb 7 TeV W/Z muon forward-η Xsec (35 pb^{-1})	[62]	14
241	LHCb 7 TeV W $A_\mu(\eta^\mu)$ (35 pb^{-1})	[62]	5
260	DØ Run-2 Z $d\sigma/dy_Z$ (0.4 fb^{-1})	[63]	28
266	CMS 7 TeV $A_\mu(\eta)$ (4.7 fb^{-1})	[64]	11
267	CMS 7 TeV $A_e(\eta)$ (0.840 fb^{-1})	[65]	11
268	ATLAS 7 TeV W/Z Xsec, $A_\mu(\eta)$ (35 pb^{-1})	[66]	41
281	DØ Run-2 $A_e(\eta)$ (9.7 fb^{-1})	[67]	13
504	CDF Run-2 incl. jet $(d^2\sigma/dp_T^j dy_j)$ (1.13 fb^{-1})	[36]	72
514	DØ Run-2 incl. jet $(d^2\sigma/dp_T^j dy_j)$ (0.7 fb^{-1})	[37]	110
535	ATLAS 7 TeV incl. jet $(d^2\sigma/dp_T^j dy_j)$ (35 pb^{-1})	[68]	90
538	CMS 7 TeV incl. jet $(d^2\sigma/dp_T^j dy_j)$ (5 fb^{-1})	[69]	133

Candidate experiments in the CTEQ-TEA fit

ID#	Experimental dataset		N_d
245	LHCb 7 TeV Z/W muon forward-η Xsec (1.0 fb^{-1})	[70]	33
246	LHCb 8 TeV Z electron forward-η $d\sigma/dy_Z$ (2.0 fb^{-1})	[71]	17
247	ATLAS 7 TeV $d\sigma/dp_T^Z$ (4.7 fb^{-1})	[72]	8
249	CMS 8 TeV W muon, Xsec, $A_\mu(\eta^\mu)$ (18.8 fb^{-1})	[73]	33
250	LHCb 8 TeV W/Z muon, Xsec, $A_\mu(\eta^\mu)$ (2.0 fb^{-1})	[74]	42
252	ATLAS 8 TeV Z $(d^2\sigma/d\|y\|_{ll} dm_{ll})$ (20.3 fb^{-1})	[75]	48
253	ATLAS 8 TeV $(d^2\sigma/dp_T^Z dm_{ll})$ (20.3 fb^{-1})	[76]	45
542	CMS 7 TeV incl. jet, R=0.7, $(d^2\sigma/dp_T^j dy_j)$ (5 fb^{-1})	[34]	158
544	ATLAS 7 TeV incl. jet, R=0.6, $(d^2\sigma/dp_T^j dy_j)$ (4.5 fb^{-1})	[33]	140
545	CMS 8 TeV incl. jet, R=0.7, $(d^2\sigma/dp_T^j dy_j)$ (19.7 fb^{-1})	[35]	185
565	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dp_T^t$ (20.3 fb^{-1})	[38]	8
566	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dy_{<t/\bar{t}>}$ (20.3 fb^{-1})	[38]	5
567	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dm_{t\bar{t}}$ (20.3 fb^{-1})	[38]	7
568	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dy_{t\bar{t}}$ (20.3 fb^{-1})	[38]	5

N_d is the number of data points

A shifted residual r_i

$r_i(\vec{a}) = \frac{T_i(\vec{a}) - D_i^{sh}(\vec{a})}{s_i}$ are N_{pt} **shifted residuals** for point i, PDF parameters $\vec{a}$

$\bar{\lambda}_\alpha(\vec{a})$ are N_λ **optimized nuisance parameters** (dependent on $\vec{a}$)

The $\chi^2(\vec{a})$ for experiment E is

$$\chi^2(\vec{a}) = \sum_{i=1}^{N_{pt}} r_i^2(\vec{a}) + \sum_{\alpha=1}^{N_\lambda} \bar{\lambda}_\alpha^2(\vec{a}) \approx \sum_{i=1}^{N_{pt}} r_i^2(\vec{a})$$

$T_i(\vec{a})$ is the theory prediction for PDF parameters $\vec{a}$

D_i^{sh} is the data value **including the optimal systematic shift**

$$D_i^{sh}(\vec{a}) = D_i - \sum_{\alpha=1}^{N_\lambda} \beta_{i\alpha} \bar{\lambda}_\alpha(\vec{a})$$

s_i is the uncorrelated error

$r_i(\vec{a})$ and $\bar{\lambda}_\alpha(\vec{a})$ are tabulated or extracted from the cov. matrix ⇒ backup slides

Finding shifted residuals r_i from the covariance matrix

The CTEQ-TEA fit returns tables of $r_i(\vec{a})$ and $\bar{\lambda}_\alpha(\vec{a})$ for every i and α

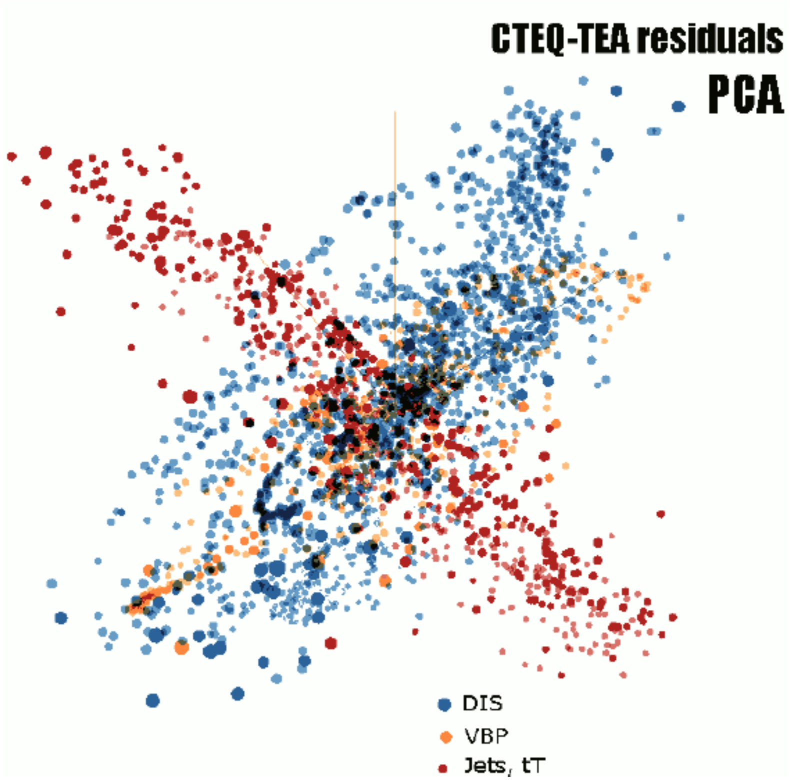
Alternatively, they can be found from the covariance matrix:

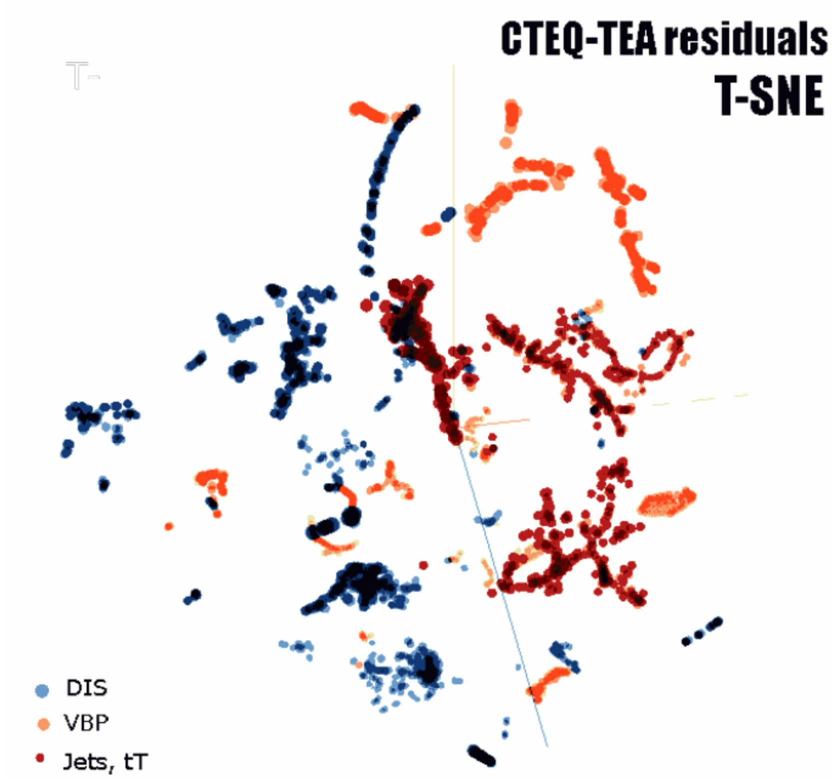
$$r_i(\vec{a}) = s_i \sum_{j=1}^{N_{pt}} (\mathrm{cov}^{-1})_{ij} \, (T_j(\vec{a}) - D_j), \qquad \bar{\lambda}_\alpha(\vec{a}) = \sum_{i,j=1}^{N_{pt}} (\mathrm{cov}^{-1})_{ij} \frac{\beta_{i\alpha}}{s_i} \frac{(T_j(\vec{a}) - D_j)}{s_j}$$

TensorFlow Embedding Projector

http://projector.tensorflow.org

Reads 2 .tsv files with $\vec{\nabla} r_i / \langle r_0 \rangle_E$ vectors and metadata (descriptions of data points)





Principal Component Analysis (PCA) visualizes the 56-dim. manifold by reducing it to 10 dimensions (à la META PDFs)

t-distributed stochastic neighbor embedding (**t-SNE**) sorts $\vec{\nabla} r_i / \langle r_0 \rangle_E$ vectors according to their similarity

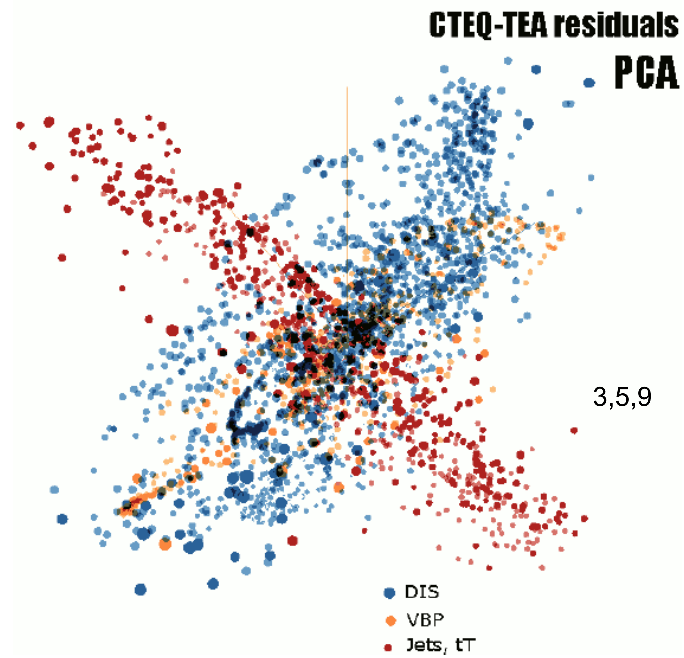
Manifolds of data residuals

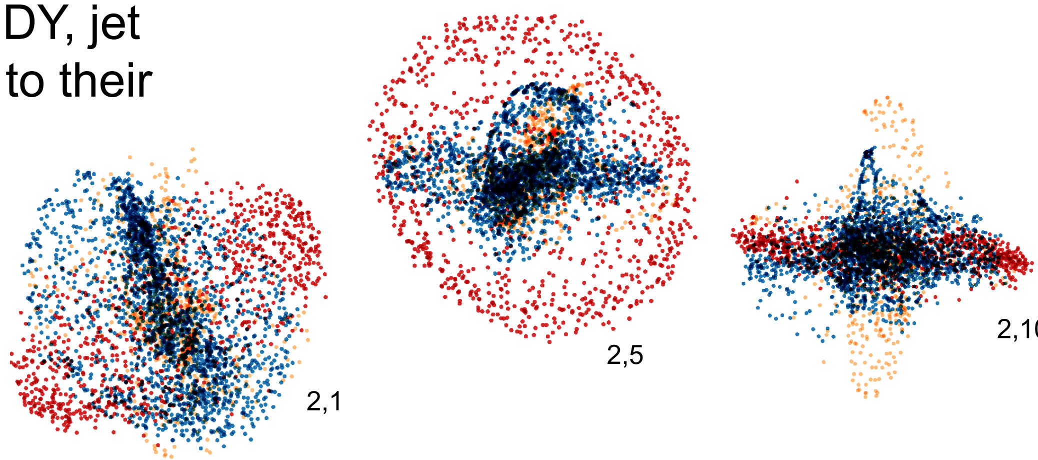
The 2N-dimensional distribution of $\vec{\delta}_i$ is easy to analyze with data-mining tools…

…to sort the fitted data points according to their PDF dependence (expressed by lengths and directions of $\vec{\delta}_i$);

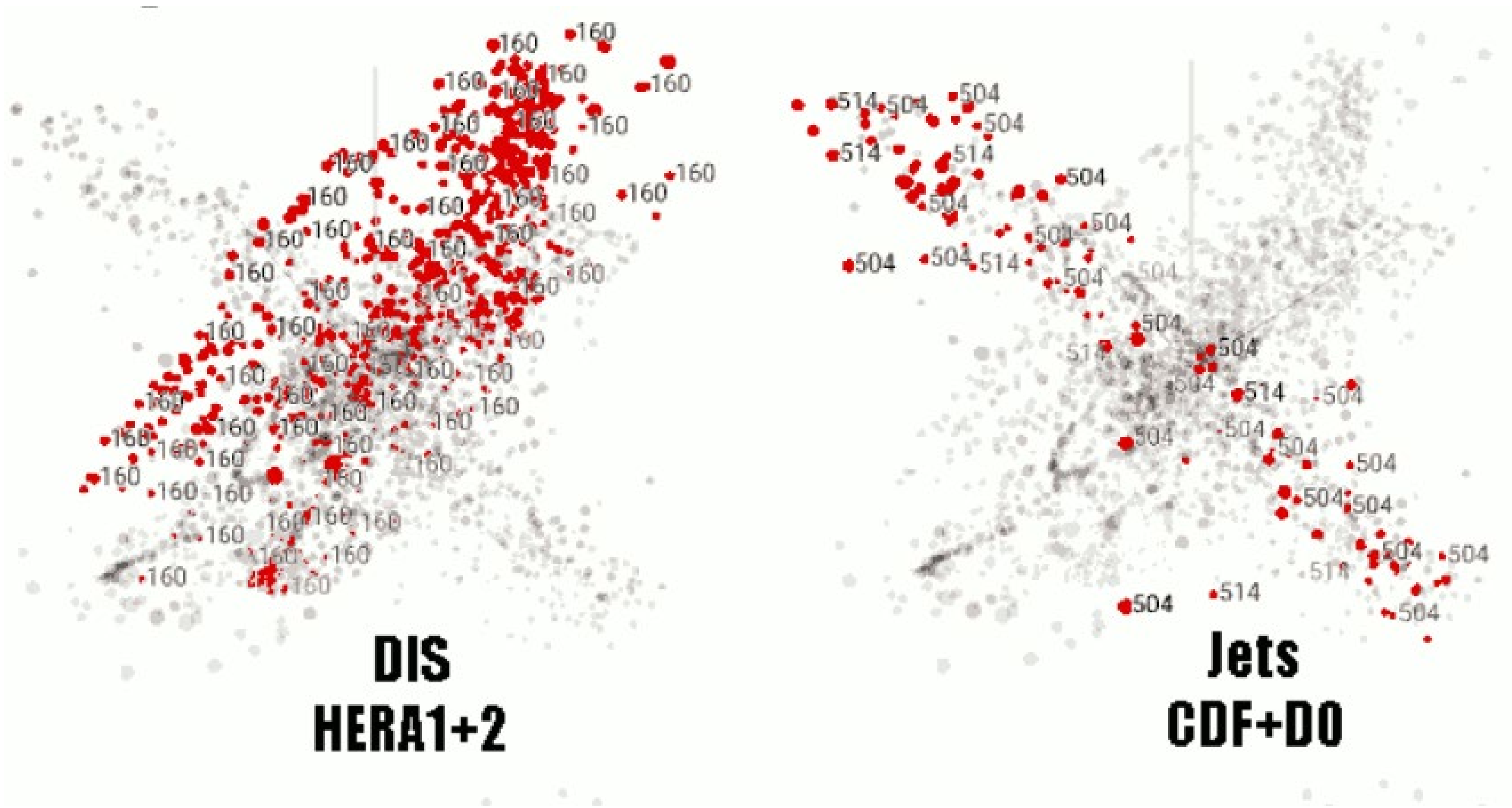
…to identify high-value data points (having long $\vec{\delta}_i$ that point away from the rest of vectors).

Some projections separate DIS, DY, jet and $t\bar{t}$ data residuals according to their PDF dependence.

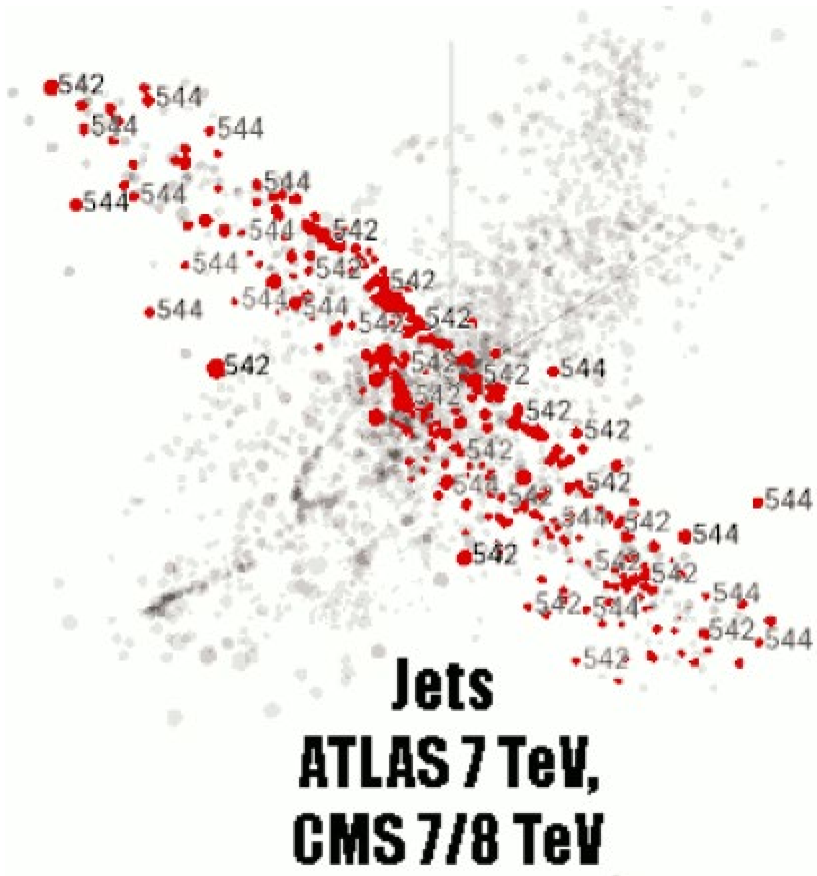


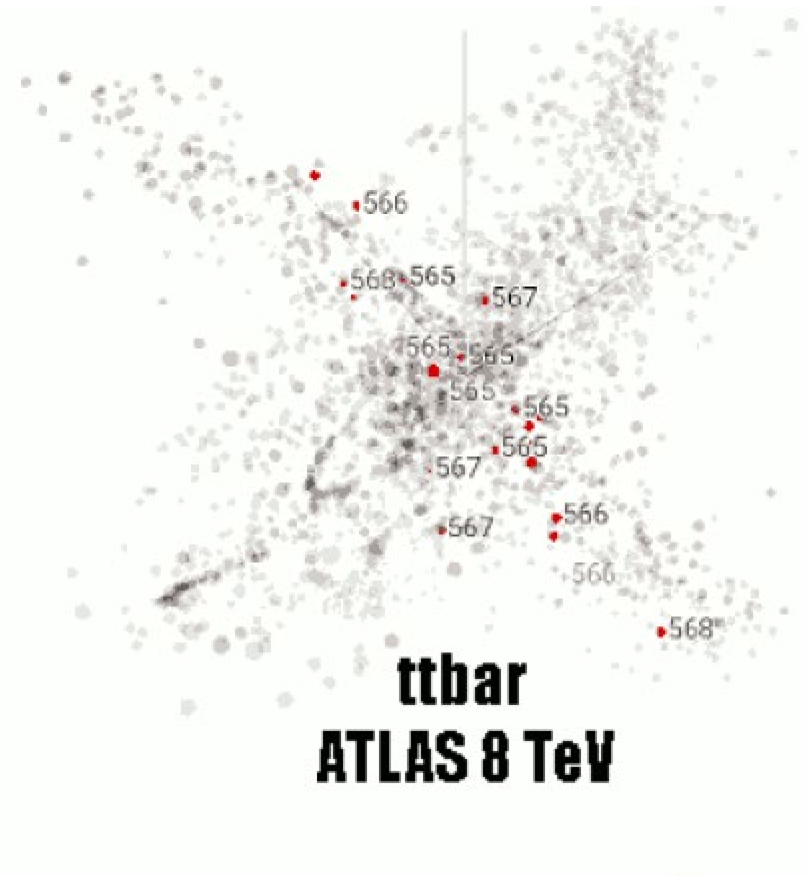


CTEQ-TEA residuals PCA



CTEQ-TEA residuals PCA

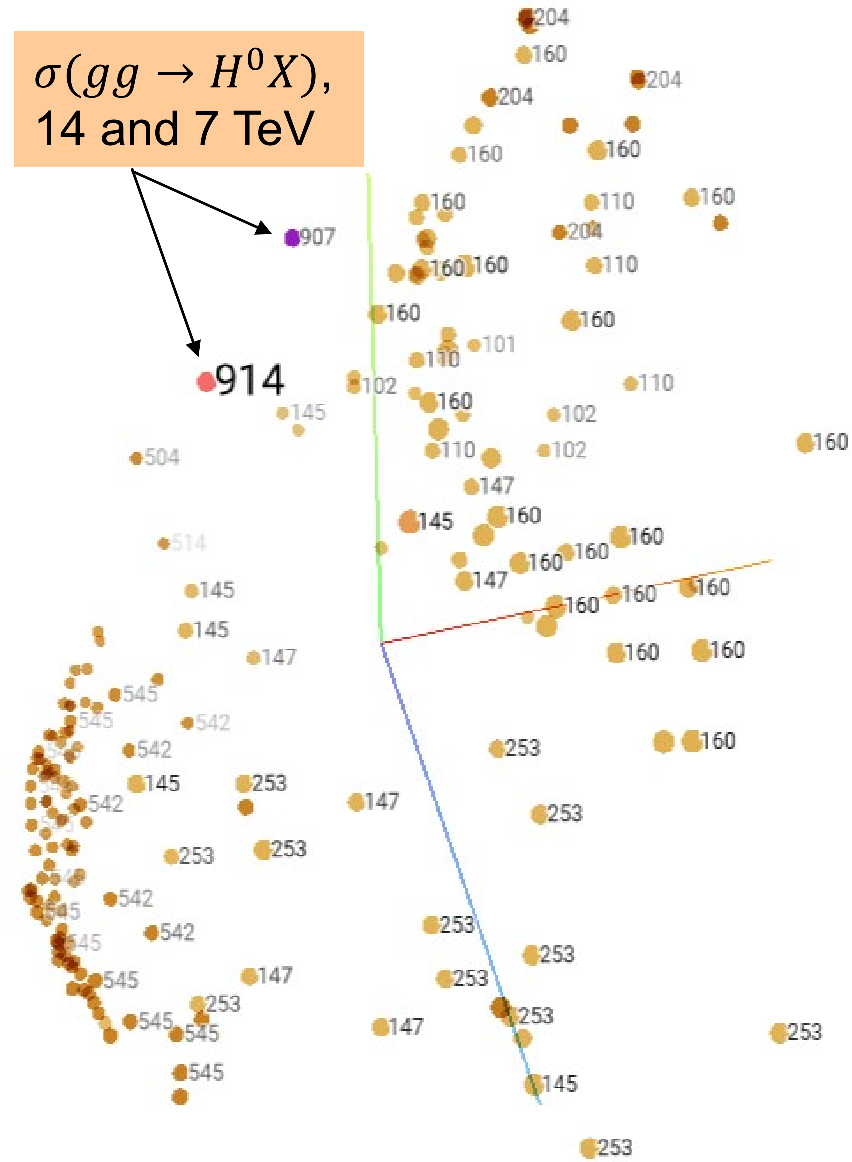




A PDF-dependent quantity f, such as the Higgs cross section at 7 or 14 TeV (ID=907, 914), defines a direction $\vec{\delta}_f$ in the (2)N-dim space.

The 3-dim projection on the right shows 300 vectors $\vec{\delta}_i$ of the CT14HERA2 global set whose directions are closest to $\vec{\delta}_f(\sigma(H^0))$. **These vectors are given by the experiments:**

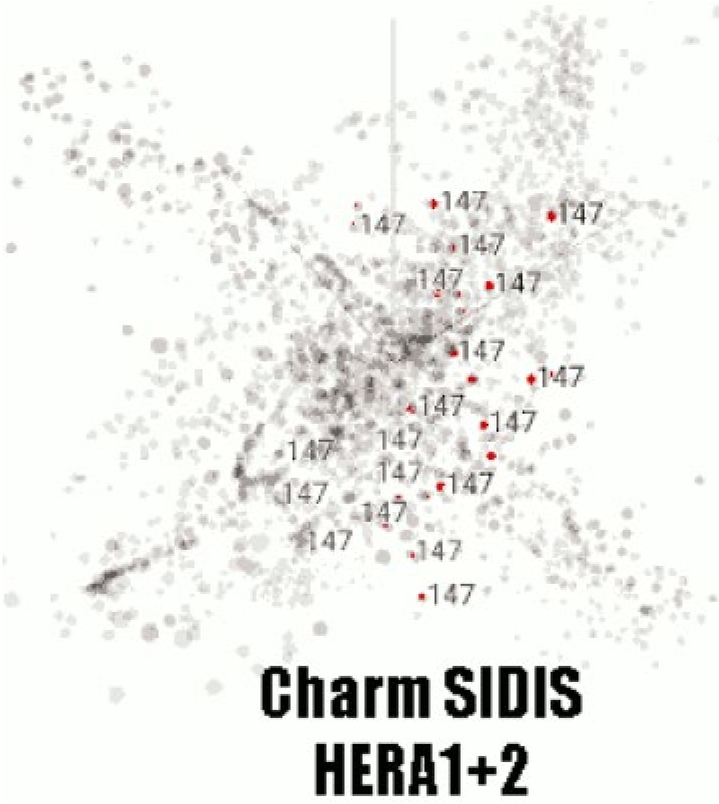
160=HERA I+II; 101, 102=BCDMS;
110=CCFR F2p; 147, 145=HERA I+II c, b;
204=E866 σ_{pp}; 253=Z p_T 8 TeV; 542, 545=CMS jets 7, 8 TeV; 504, 514=Tevatron jets

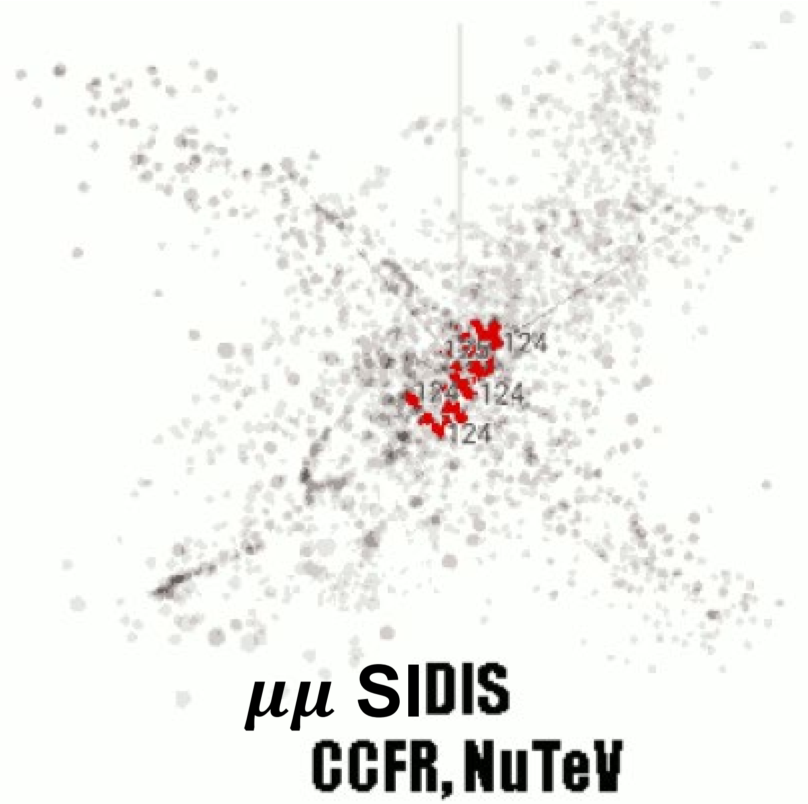


The net constraint of the i-th point on $\sigma(H)$, including systematic errors, is quantified by the projection of $\vec{\delta}_i$ on $\vec{\boldsymbol{\delta}}_f[\sigma(H)]$, called the sensitivity $S_{f,i}$.

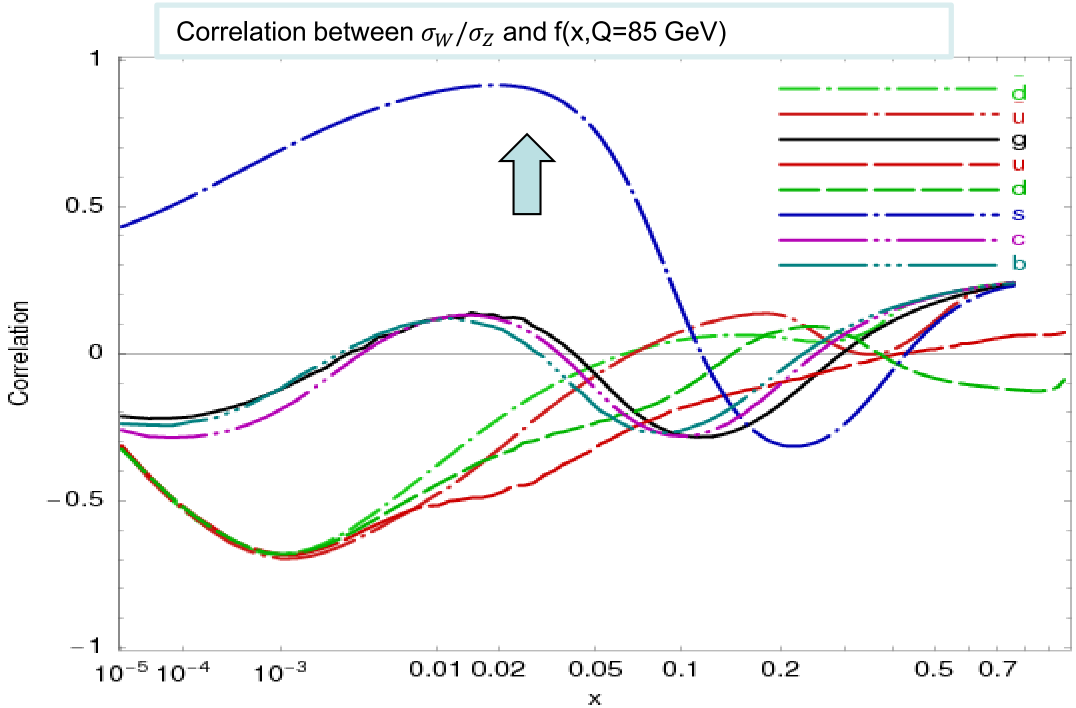
Sensitivity of expt E = sum of $S_{f,i}$ over data points in E

CTEQ-TEA residuals PCA

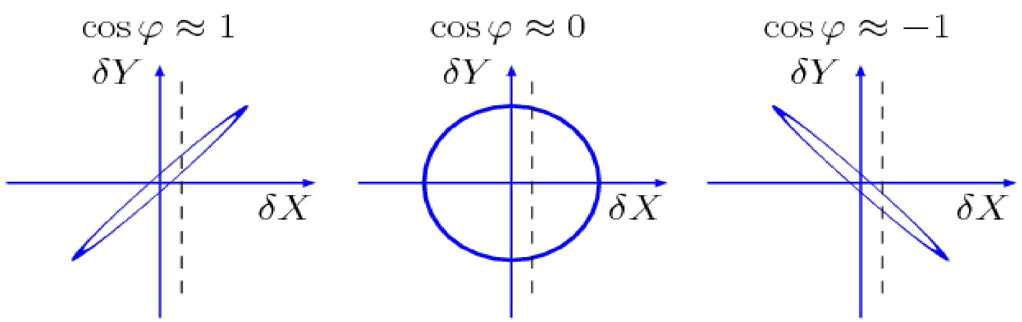




Correlations carry useful, but limited information



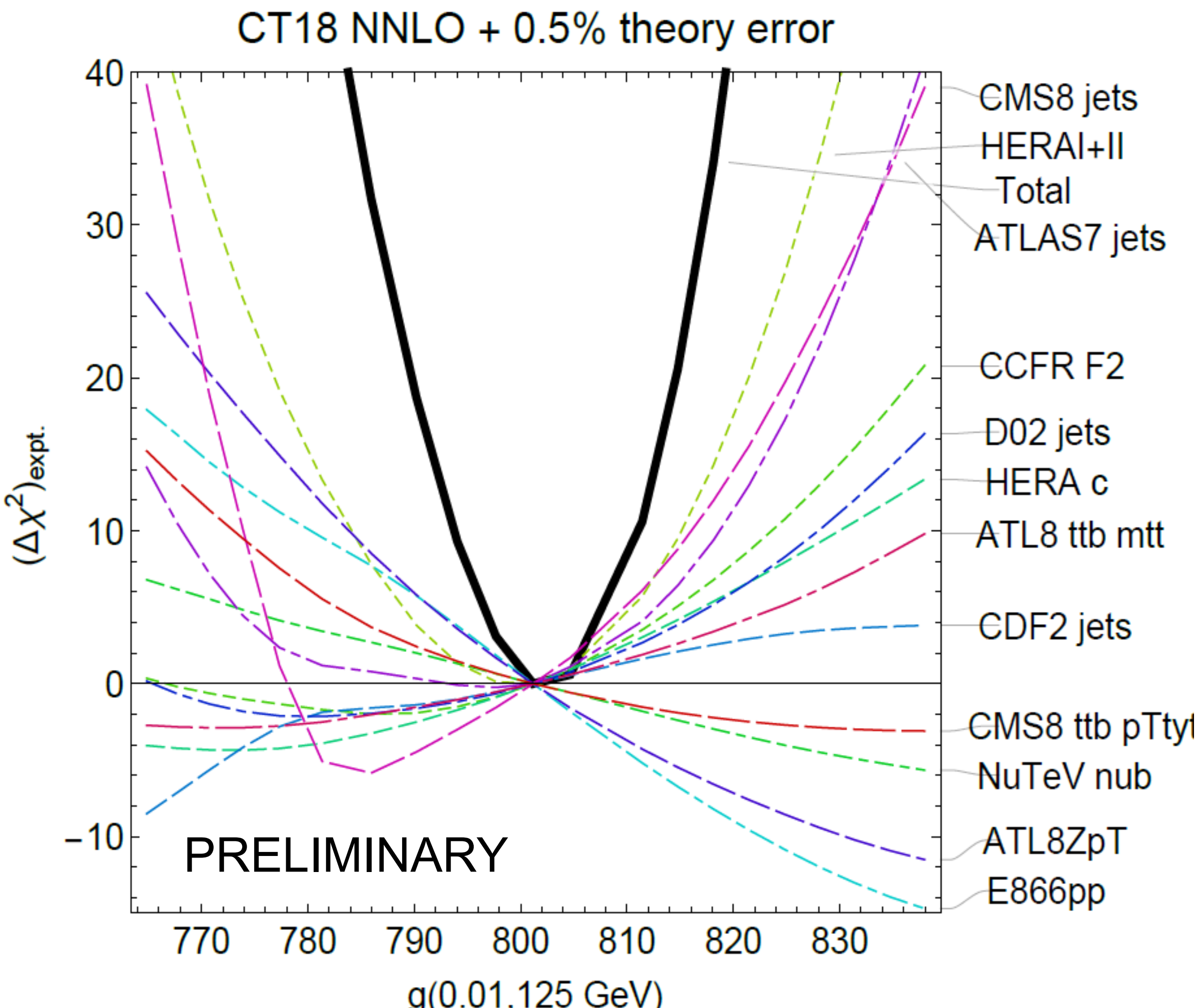
CTEQ6.6 [arXiv:0802.0007]: $\cos\varphi > 0.7$ shows that the ratio σ_W/σ_Z at the LHC must be sensitive to the strange PDF $s(x,Q)$



$\cos\varphi \approx \pm 1$ suggests that a measurement of X **may** impose tight constraints on Y

But, Corr[X,Y] between **theory** cross sections X and Y does not tell us about **experimental** uncertainties

Which experiments constrain the gluon? $x = 0.01, Q = 125$ GeV [Higgs region]



A Lagrange multiplier scan [Stump et al., hep-ph/0101151] of

$$\Delta\chi^2 = \chi^2(g) - \chi^2_{best-fit}$$

for all (black line) and individual (colored lines) experiments

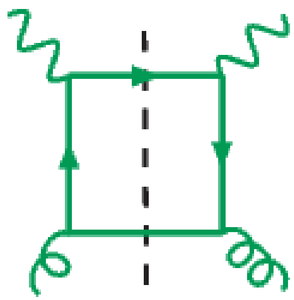
Best-fit $g(0.01, 125\text{GeV}) = 806$

After the fit

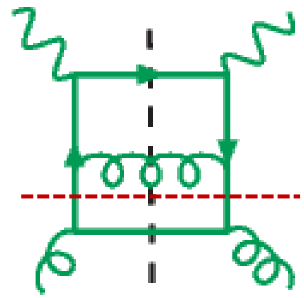
A twist-4 contribution in HERA DIS charm production ($\subset$ "intrinsic charm")

[arXiv:1707.00065]

Twist-2
$\gamma^* g \to c\bar{c}$



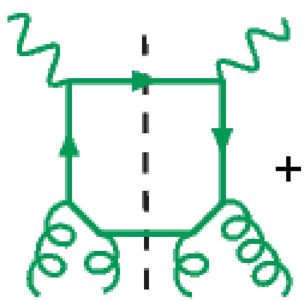
Order $\alpha_s(Q)$



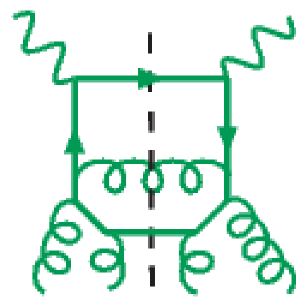
$\alpha_s^2(Q) \cdot \ln(Q^2/m_c^2)$

A ladder; must be resummed in $c(x, Q)$ in the $N_f = 4$ scheme at $Q^2 \gg m_c^2$; e.g., in the ACOT scheme

Twist-4
$\gamma^* (gg) \to c\bar{c}$



+…



$\Lambda \lesssim 1$ GeV

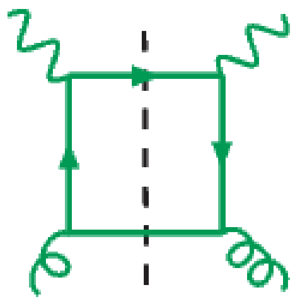
$\alpha_s^2(Q) \cdot (\Lambda^2/Q^2)$
or $\alpha_s^2(Q) \cdot (\Lambda^2/m_c^2)$

$\alpha_s^3(Q) \cdot (\Lambda^2/m_c^2) \ln(Q^2/m_c^2)$

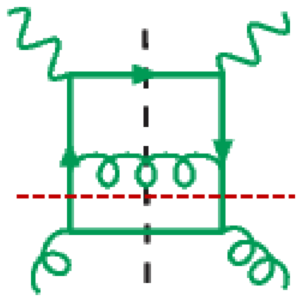
A twist-4 contribution in HERA DIS charm production ($\subset$ "intrinsic charm")

[arXiv:1707.00065]

Twist-2
$\gamma^* g \to c\bar{c}$



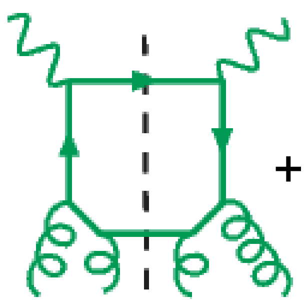
Order $\alpha_s(Q)$



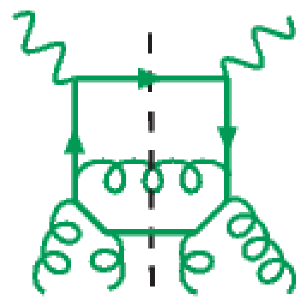
$\alpha_s^2(Q) \cdot \ln(Q^2/m_c^2)$

A ladder; must be resummed in $c(x, Q)$ in the $N_f = 4$ scheme at $Q^2 \gg m_c^2$; e.g., in the ACOT scheme

Twist-4
$\gamma^*(gg) \to c\bar{c}$



+…



$\Lambda \lesssim 1$ GeV

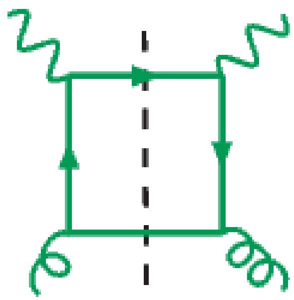
$\alpha_s^2(Q) \cdot (\Lambda^2/Q^2)$
or $\alpha_s^2(Q) \cdot (\Lambda^2/m_c^2)$

$\alpha_s^3(Q) \cdot (\Lambda^2/m_c^2) \ln(Q^2/m_c^2)$

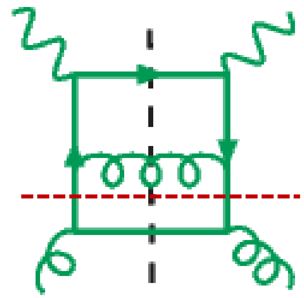
Can be of order ~10% of the twist-2 α_s^2 term

A twist-4 contribution in HERA DIS charm production ($\subset$ "intrinsic charm") [arXiv:1707.00065]

Twist-2
$\gamma^* g \to c\bar{c}$



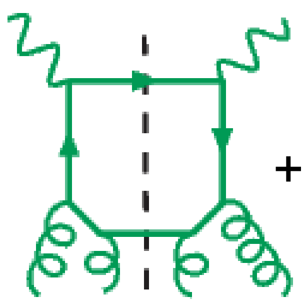
Order $\alpha_s(Q)$



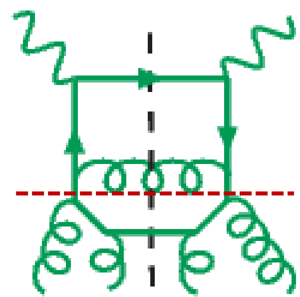
$\alpha_s^2(Q) \cdot \ln(Q^2/m_c^2)$

A ladder; must be resummed in $c(x,Q)$ in the $N_f = 4$ scheme at $Q^2 \gg m_c^2$; e.g., in the ACOT scheme

Twist-4
$\gamma^*(gg) \to c\bar{c}$



+…



$\Lambda \lesssim 1$ GeV

$\alpha_s^2(Q) \cdot (\Lambda^2/Q^2)$
or $\alpha_s^2(Q) \cdot (\Lambda^2/m_c^2)$

Can be of order ~10% of the twist-2 α_s^2 term

$\alpha_s^3(Q) \cdot (\Lambda^2/m_c^2) \ln(Q^2/m_c^2)$

The ladder subgraphs can be resummed as a part of $c(x,Q)$ in the $N_f = 4$ scheme at $Q^2 \gg m_c^2 > \Lambda^2$;

contribute to the boundary condition for $c(x,Q_0)$ at $Q_0 \approx m_c$;

obey twist-2 DGLAP equations.

CT14 IC study clarifies important questions

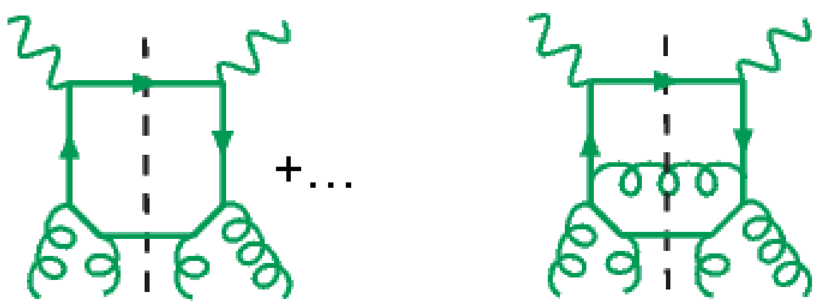
What are phenomenological constraints on the “intrinsic charm” from the global QCD data?

⇒ The CT14 charm PDFs allow a “nonperturbative” component carrying a total momentum fraction $\langle x_{IC} \rangle = 1 - 2\%$ in DIS at $Q \approx m_c$.

Can we estimate its impact on the LHC predictions?

Yes, based on the simplest approximation of the “nonperturbative” charm contribution. In most cases, the estimated impact is less than the net CT14 PDF uncertainty.

Twist-4
$\gamma^*(gg) \to c\bar{c}$



$\alpha_s^2(Q) \cdot (\Lambda^2/Q^2)$
or $\alpha_s^2(Q) \cdot (\Lambda^2/m_c^2)$

$\alpha_s^3(Q) \cdot (\Lambda^2/m_c^2) \ln(Q^2/m_c^2)$

[arXiv:1707.00065]

Note:
“intrinsic charm” ≠ “fitted charm”

PDF fits may include a ``fitted charm'' PDF

``Fitted charm'' = ``higher-twist charm''
+ other (possibly not universal)
higher O(α_s) / higher power terms

QCD factorization theorem for DIS structure function $F(x, Q)$ [Collins, 1998]:

All α_s orders:

$$F(x,Q) = \sum_{a=0}^{N_f} \int_x^1 \frac{d\xi}{\xi} \mathcal{C}_a \left(\frac{x}{\xi}, \frac{Q}{\mu}, \frac{m_c}{\mu}; \alpha(\mu) \right) f_{a/p}(\xi, \mu) + \mathcal{O}(\Lambda^2/m_c^2, \Lambda^2/Q^2).$$

The PDF fits implement this formula up to (N)NLO ($N_{ord} = 1$ or 2):

PDF fits:

$$F(x,Q) = \sum_{a=0}^{N_f} \int_x^1 \frac{d\xi}{\xi} \mathcal{C}_a^{(N_{ord})} \left(\frac{x}{\xi}, \frac{Q}{\mu}, \frac{m_c}{\mu}; \alpha(\mu) \right) f_{a/p}^{(N_{ord})}(\xi, \mu).$$

The perturbative charm PDF component cancels at $Q \approx m_c$ up to a higher order

The 'fitted charm component' may approximate for missing terms of orders α_s^p with $p > N_{ord}$, or Λ^2/m_c^2, or Λ^2/Q^2 -- generally process-dependent

Dependence on the switching scale (no IC)

If the "fitted charm" is purely twist-2, we expect its effect to vanish for a sufficiently high α_s order of the calculation.

This is analogous to the reduction in the dependence on the switching scale μ_c from 3FS to 4FS, when the α_s order increases for a fixed Q_0 and m_c, as demonstrated recently by the xFitter group

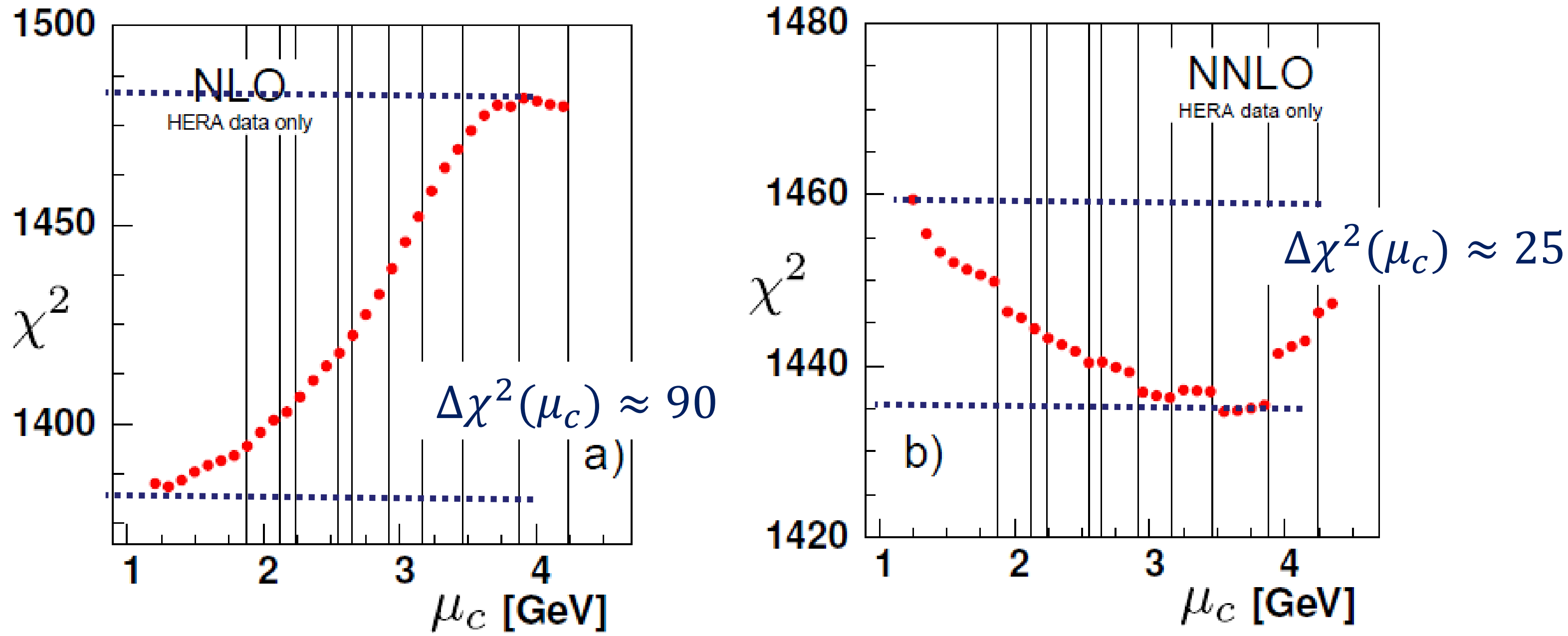


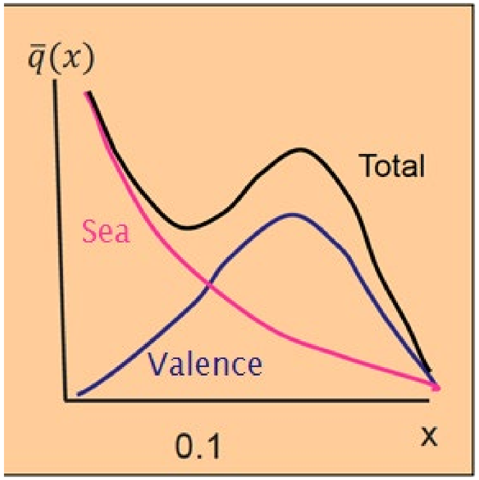
Fig. 5 χ^2 vs. the charm matching scale μ_c at a) NLO and b) NNLO for all data sets. The bin boundaries for the HERA data set "HERA1+2 NCep 920" are indicated by the vertical lines.

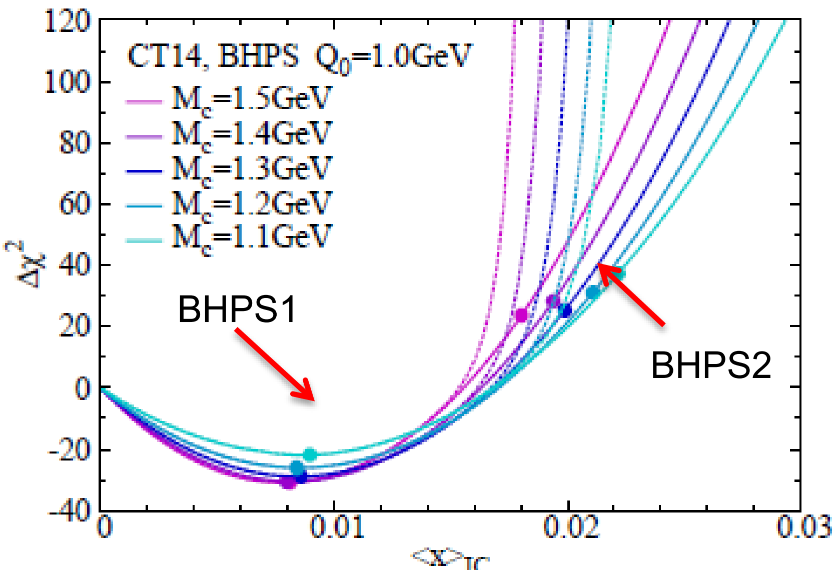
Bertone et al. (xFitter), arXiv:1707.05343

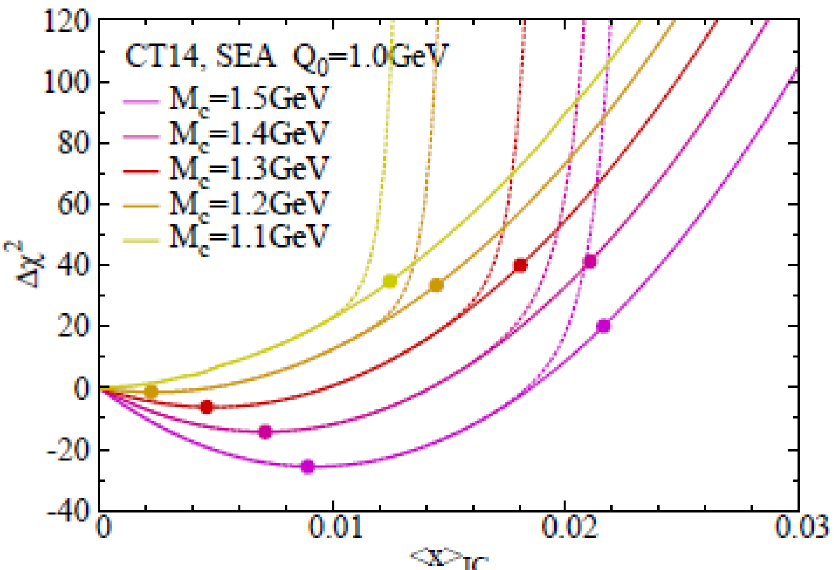
Dependence of $\Delta\chi^2$ on the IC momentum fraction $\langle x \rangle_{IC}$

In contrast, a twist-4 "IC" contribution will not decrease when going from $N^k LO$ to $N^{k+1} LO$.

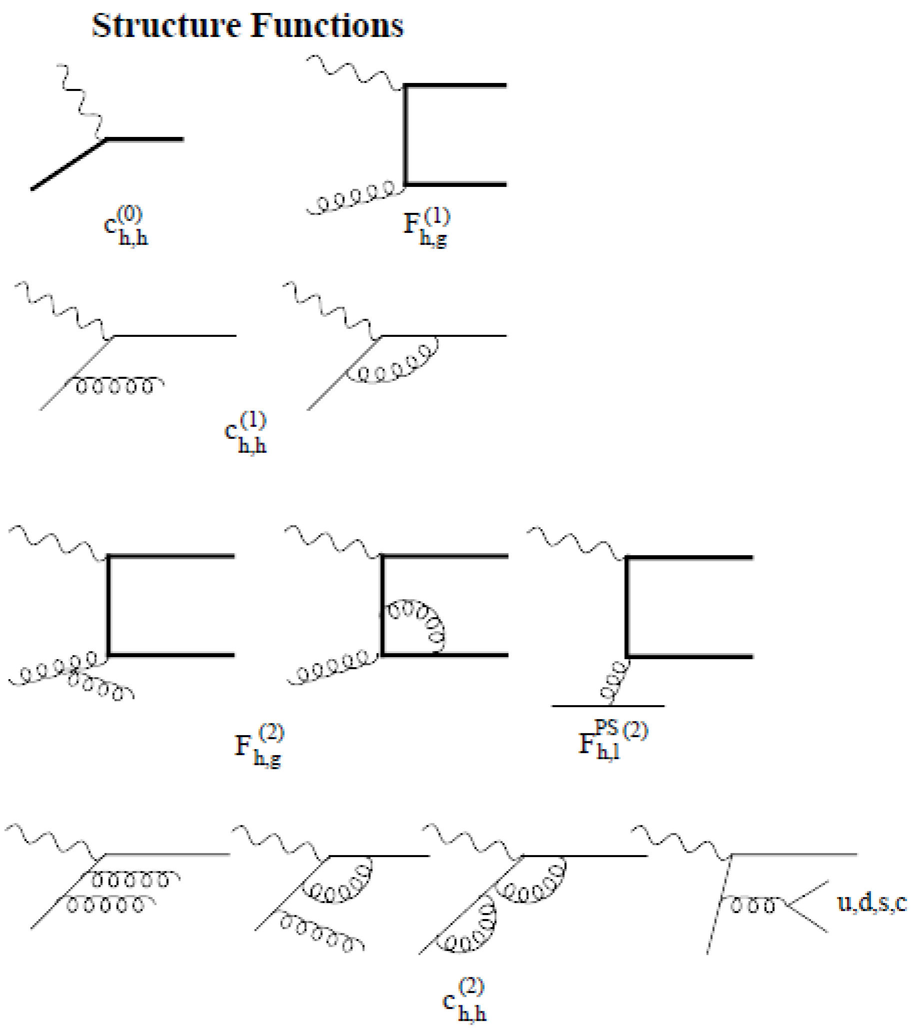
Depending on its dynamical origin, the IC charm takes a variety of shapes, e.g., a "sea-like" (SEA) or "valence-like" form. The Brodsky-Hoyer-Peterson-Sakai form (BHPS) predicts a "valence-like" $c(x, Q_0)$ peaked at $x \sim 0.2$. A sea-like form is monotonic in x.

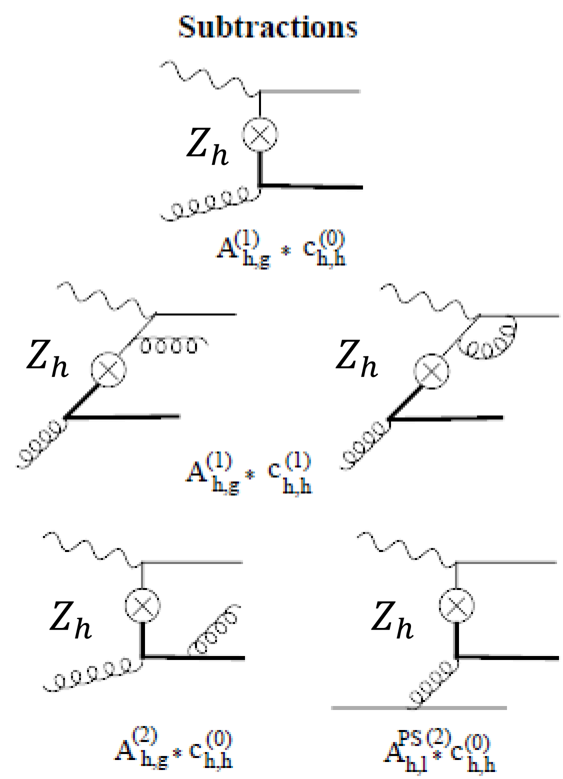






Twist-2: factorization for DIS in S-ACOT-χ scheme up to NNLO



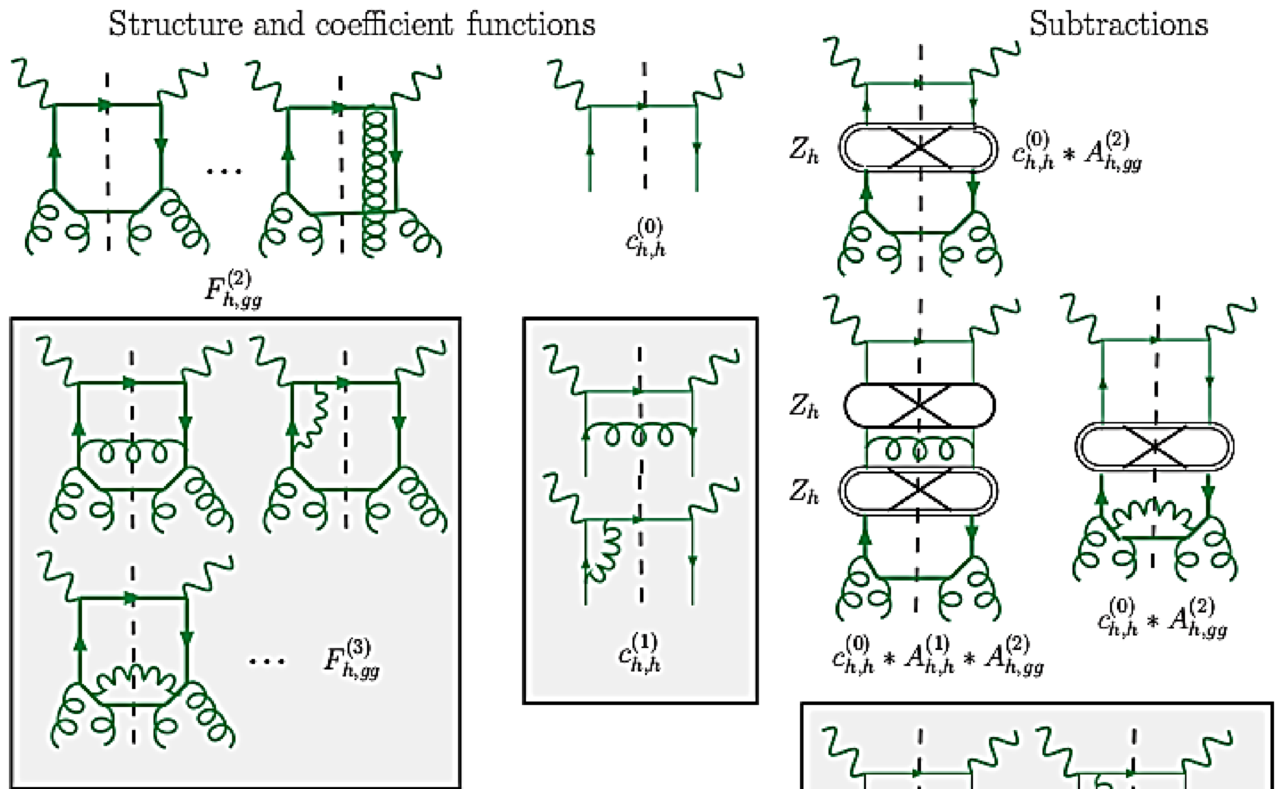


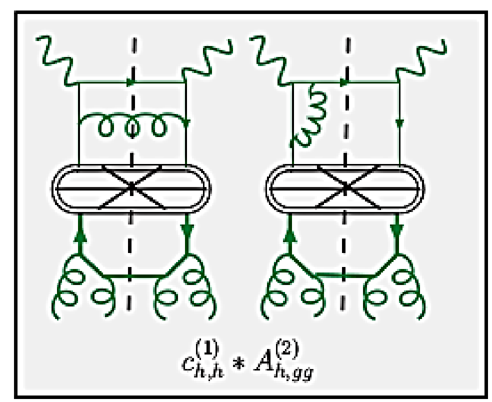
Leading-power radiative contributions to neutral-current DIS charm production in the CTEQ-TEA NNLO analysis,
from Guzzi et al., arXiv:1108.5112

ACOT-like factorization for twist-4 charm contributions (an example)

α_s^2 :

α_s^3 :



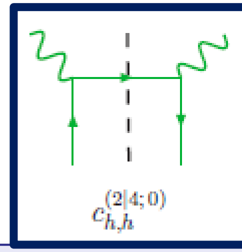


The IC terms can be factorized in the ACOT method into

- universal twist-4 nonperturbative functions $f_{gg}(\xi_1, \xi_2, Q)$, etc.

- process-dependent coefficient functions $c_{h,h}^{(k)}, C_{h,gg}^{(k)}$, etc.

Intrinsic charm contributions, practical implementation



In the absence of full computation, we (and other groups) make the simplest approximation:

$$F_{IC}(x, Q_0) = [c_{h,h}^{(2|4;0)} \otimes f_{c/p}^{IC}](x, Q_0)$$

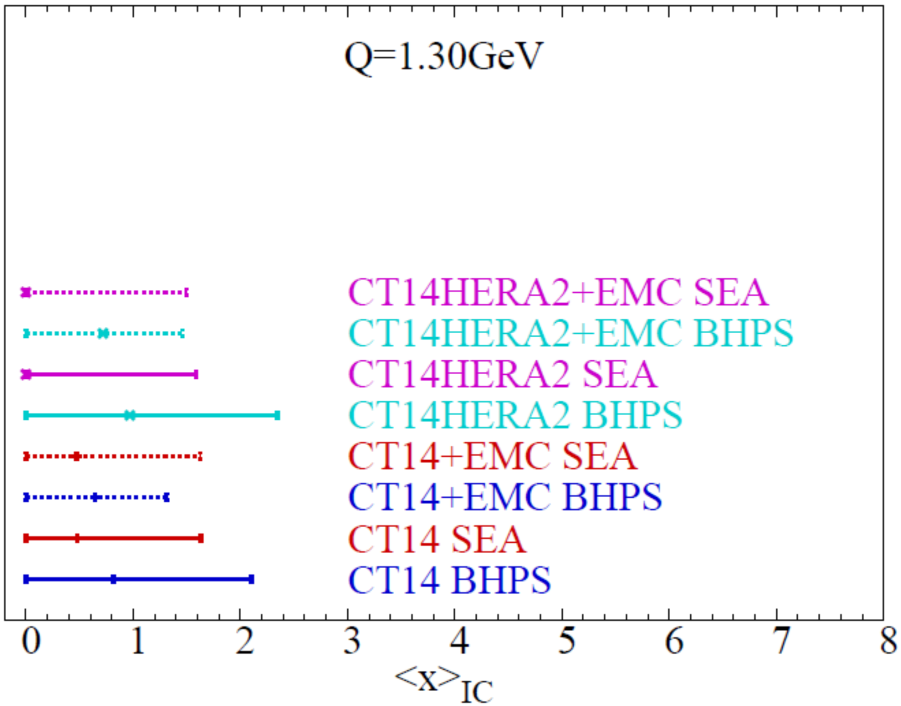
$\boldsymbol{c_{h,h}^{(2|4;0)}}$ **is the twist-2 charm DIS coefficient function** introduced to factorize the $O(\alpha_s^0)$ twist-4 term; depends on the heavy-quark scheme

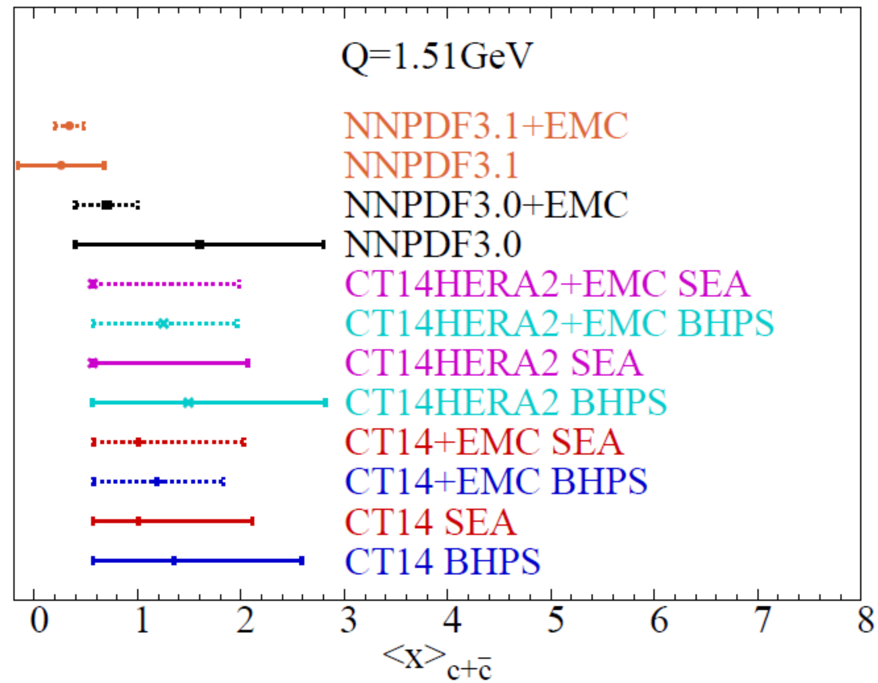
CT14 IC: $c_{h,h}^{(2|4;0)}$ is defined to be equal to $c_{h,h}^{(0)}$ in the S-ACOT-χ scheme

$\boldsymbol{f_{c/p}^{IC}(\xi, Q_0)}$ **is a nonperturbative charm parametrization**:

CT14 IC: $f_{c/p}^{IC}(\xi, Q_0)$ is a "**valence-like**" or a "**sea-like**" function, combined with the to the perturbative charm $f_{c/p}^{pert}$ from $g \to c\bar{c}$ splittings

Allowed $c + \bar{c}$ momentum fractions





Sources of differences	CT14 IC	NNPDF3.x
α_s order	NNLO only	NLO, NNLO
Settings	90% c.l. from Lagrange multiplier scan $Q_0 = m_c^{pole} = 1.3$ GeV	Symmetric. 68% c.l. from Monte-Carlo sampling, $Q_0 = m_c^{pole} = 1.51$ GeV
LHC 8 TeV W, Z	Under validation; mild tension with HERA DIS data	Included; strong effect despite a smallish data sample
1983 EMC F_{2c} data included?	Only as a cross check (unknown syst. effects in EMC data)	Optional, strong effect on the PDF error