The CT18 global PDF fit and phenomenology in the HL-LHC era

C.-P. Yuan
Michigan State University
Wu-Ki Tung Endowed Professor

Nov 14, 2019

LPC Workshop @ FNAL

CT = CTEA-TEA = CTEQ – Tung et al. (TEA)
in memory of Prof. Wu-Ki Tung

CTEQ-TEA group

- CTEQ – Tung et al. (TEA)

in memory of Prof. Wu-Ki Tung,
who co-established CTEQ Collaboration in early 90's

- Current members:

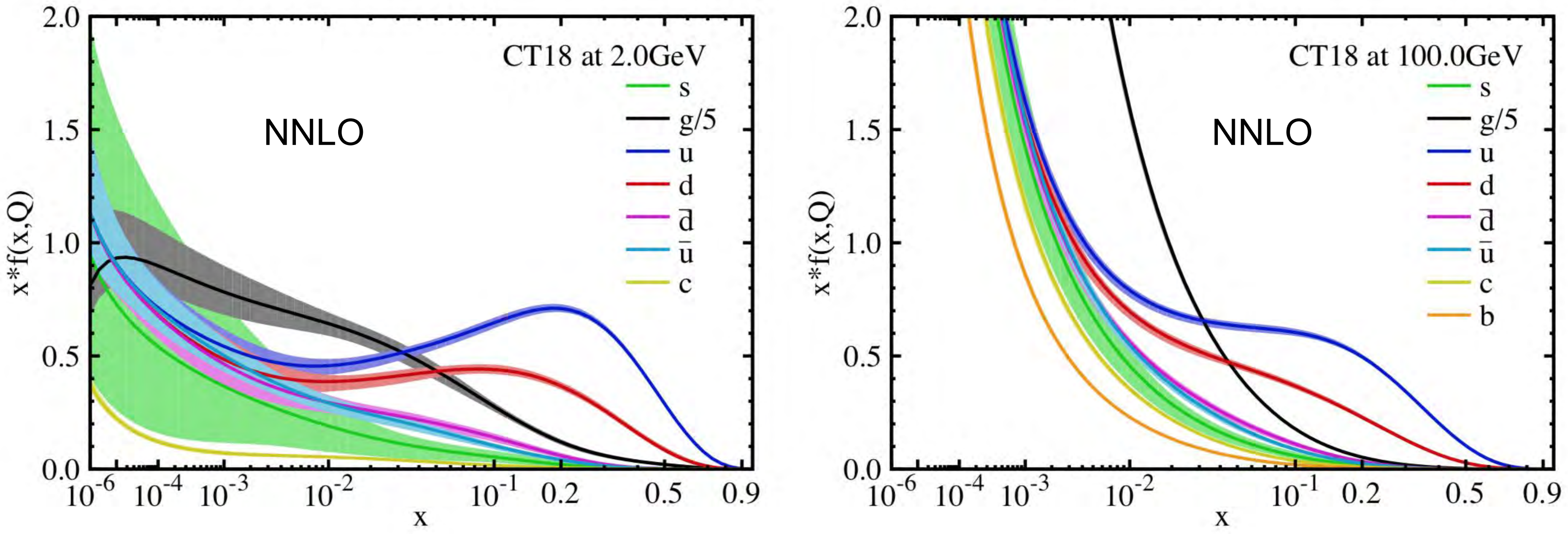
Tie-Jiun Hou (Northeastern U., China), Sayipjamal Dulat, Ibrahim Sitiwaldi (Xinjiang U.), Jun Gao (Shanghai Jiaotong U.), Marco Guzzi (Kennesaw State U.), Tim Hobbs, Pavel Nadolsky, Bo-Ting Wang, Keping Xie (Southern Methodist U.), Joey Huston, Jon Pumplin, Dan Stump, Carl Schmidt, Jan Winter, CPY (Michigan State U.)

CT18 in a nutshell

- Start with CT14-HERA2 (HERAI+II combined data released after publication of CT14)
- Examine a wide range of PDF parameterizations
- Use as much relevant LHC data as possible using applgrid/fastNLO interfaces to data sets, with NNLO/NLO K-factors, or fastNNLO tables in the case of top pair production. **Benchmark the predictions!**
- Examine **QCD scale dependence** in key processes
- Implement **parallelization** of the global PDF fitting to allow for faster turn-around time
- Validate the results using a **strong set of goodness-of-fit tests** *(Kovarik, PN, Soper, arXiv:1905.06957)*
- Use diverse statistical techniques (**PDFSense, ePump, Gaussian variables, Lagrange Multiplier scans**) to examine agreement between experiments

CT18...

- Main product is CT18 (NNLO, NLO, LO)



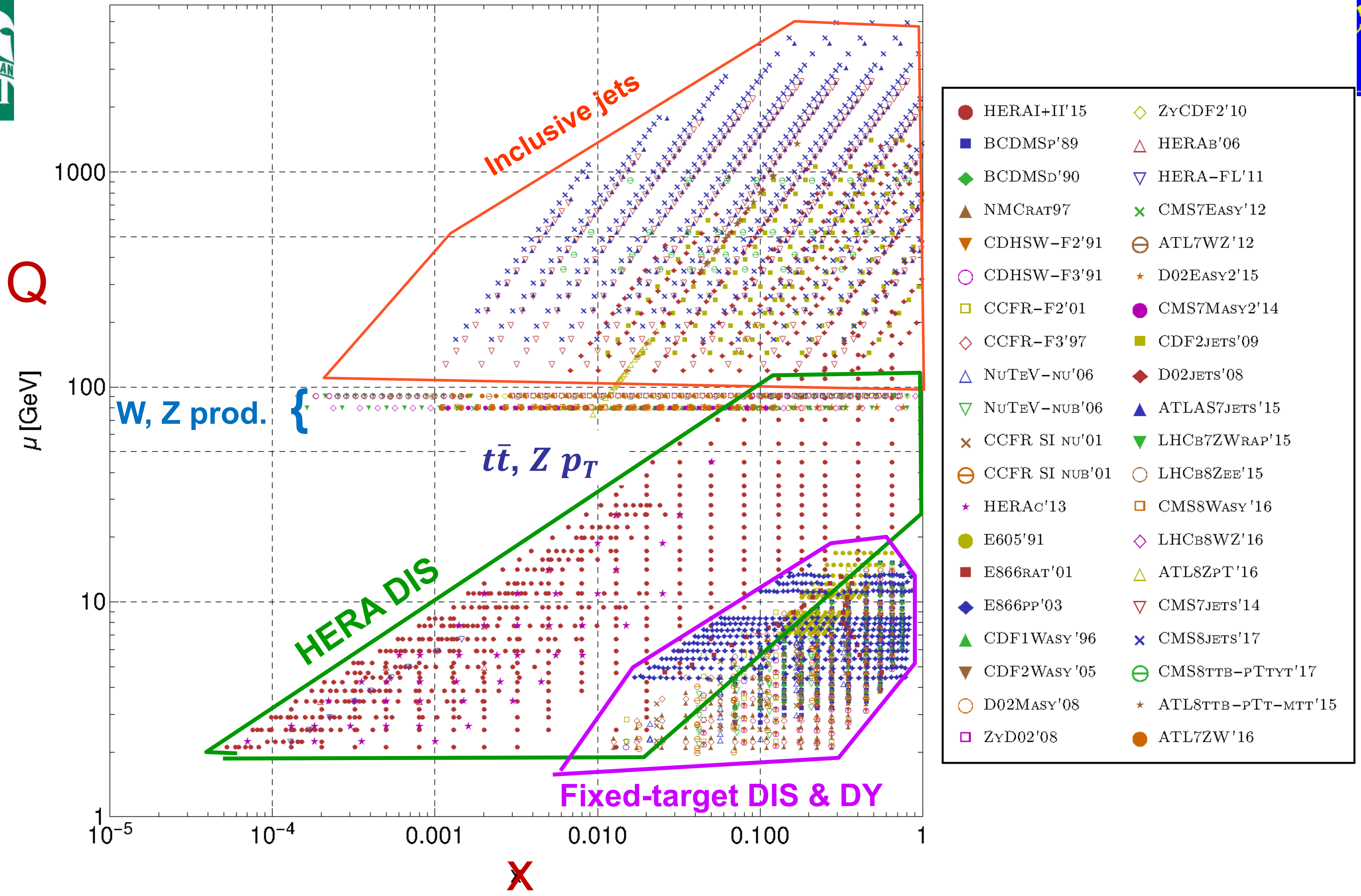
- Including full data set except for ATLAS 7 TeV W/Z, which has a sizeable impact on the global fit (strange quark)

CT18 PDFs available from https://tinyurl.com/ct18pdfs-1

...and family

PDF ensemble	Factorization scale in DIS	ATLAS 7 Z/W data included?	CDHSW $F_2^{p,d}$ data included?	Pole charm mass, GeV
CT18	$\mu^2_{F,DIS} = Q^2$	No	Yes	1.3
CT18X	$\mu^2_{F,DIS} = 0.8^2 \left(Q^2 + \frac{0.3\ GeV^2}{x^{0.3}}\right)$	No	Yes	1.3
CT18A	$\mu^2_{F,DIS} = Q^2$	Yes	Yes	1.3
CT18Z	$\mu^2_{F,DIS} = 0.8^2 \left(Q^2 + \frac{0.3\ GeV^2}{x^{0.3}}\right)$	Yes	No	1.4

Experimental data in CT18 PDF analysis



C T E Q

ID#	Experimental data set		$N_{pt,n}$	χ^2_n	$\chi^2_n/N_{pt,n}$	S_n
160	HERAI+II 1 fb^{-1}, H1 and ZEUS NC and CC $e^{\pm}p$ reduced cross sec. comb.	[10]	1120	1408.7(1377.8)	1.3(1.2)	5.7(5.1)
101	BCDMS F_2^p	[22]	337	373.7(383.8)	1.1(1.1)	1.4(1.8)
102	BCDMS F_2^d	[23]	250	280.4(287.0)	1.1(1.1)	1.3(1.6)
104	NMC F_2^d/F_2^p	[24]	123	125.7(116.2)	1.0(0.9)	0.2(-0.4)
108	CDHSW† F_2^p	[25]	85	85.6(86.8)	1.0(1.0)	0.1(0.2)
109	CDHSW† F_3^p	[25]	96	86.5(85.6)	0.9(0.9)	-0.7(-0.7)
110	CCFR F_2^p	[26]	69	78.8(76.0)	1.1(1.1)	0.9(0.6)
111	CCFR xF_3^p	[27]	86	33.8(31.4)	0.4(0.4)	-5.2(-5.6)
124	NuTeV $\nu\mu\mu$ SIDIS	[28]	38	18.5(30.3)	0.5(0.8)	-2.7(-0.9)
125	NuTeV $\bar{\nu}\mu\mu$ SIDIS	[28]	33	38.5(56.7)	1.2(1.7)	0.7(2.5)
126	CCFR $\nu\mu\mu$ SIDIS	[29]	40	29.9(35.0)	0.7(0.9)	-1.1(-0.5)
127	CCFR $\bar{\nu}\mu\mu$ SIDIS	[29]	38	19.8(18.7)	0.5(0.5)	-2.5(-2.7)
145	H1 σ_r^b	[30]	10	6.8(7.0)	0.7(0.7)	-0.6(-0.6)
147	Combined HERA charm production	[31]	47	58.3(56.4)	1.2(1.2)	1.1(1.0)
169	H1 F_L	[16]	9	17.0(15.4)	1.9(1.7)	1.7(1.4)
201	E605 Drell-Yan process	[32]	119	103.4(102.4)	0.9(0.9)	-1.0(-1.1)
203	E866 Drell-Yan process $\sigma_{pd}/(2\sigma_{pp})$	[33]	15	16.1(17.9)	1.1(1.2)	0.3(0.6)
204	E866 Drell-Yan process $Q^3d^2\sigma_{pp}/(dQdx_F)$	[34]	184	244.4(239.7)	1.3(1.3)	2.9(2.7)
225	CDF Run-1 electron A_{ch}, $p_{T\ell} > 25$ GeV	[35]	11	9.0(9.3)	0.8(0.8)	-0.3(-0.2)
227	CDF Run-2 electron A_{ch}, $p_{T\ell} > 25$ GeV	[36]	11	13.5(13.4)	1.2(1.2)	0.6(0.6)
234	DØ Run-2 muon A_{ch}, $p_{T\ell} > 20$ GeV	[37]	9	9.1(9.0)	1.0(1.0)	0.2(0.1)
260	DØ Run-2 Z rapidity	[38]	28	16.9(18.7)	0.6(0.7)	-1.7(-1.3)
261	CDF Run-2 Z rapidity	[39]	29	48.7(61.1)	1.7(2.1)	2.2(3.3)
266	CMS 7 TeV 4.7 fb^{-1}, muon A_{ch}, $p_{T\ell} > 35$ GeV	[40]	11	7.9(12.2)	0.7(1.1)	-0.6(0.4)
267	CMS 7 TeV 840 pb^{-1}, electron A_{ch}, $p_{T\ell} > 35$ GeV	[41]	11	11.8(16.1)	1.1(1.5)	0.3(1.1)
268	ATLAS 7 TeV 35 pb^{-1} W/Z cross sec., A_{ch}	[42]	41	44.4 50.6)	1.1(1.2)	0.4(1.1)
281	DØ Run-2 9.7 fb^{-1} electron A_{ch}, $p_{T\ell} > 25$ GeV	[43]	13	22.8(20.5)	1.8(1.6)	1.7(1.4)
504	CDF Run-2 inclusive jet production	[44]	72	122.4(117.0)	1.7(1.6)	3.5(3.2)
514	DØ Run-2 inclusive jet production	[45]	110	113.8(115.2)	1.0(1.0)	0.3(0.4)

TABLE I. Data sets included in the CT18(Z) global analysis. The numbers in round brackets are for the CT18Z fit.

Data sets employed in the CT18(Z) analysis. The numbers in round brackets are for the CT18Z fit. $N_{pt,n}$, χ^2 are the number of points and value of χ^2 for the n-th experiment at the global minimum. S_n is the effective Gaussian parameter quantifying agreement with each experiment.

MICHIGAN STATE

C T E Q

Only in CT18Z

ID#	Experimental data set		$N_{pt,n}$	χ^2_n	$\chi^2_n/N_{pt,n}$	S_n
245	LHCb 7 TeV 1.0 fb^{-1} W/Z forward rapidity cross sec.	[46]	33	53.8 (39.9)	1.6 (1.2)	2.2 (0.9)
246	LHCb 8 TeV 2.0 fb^{-1} $Z \to e^-e^+$ forward rapidity cross. sec.	[47]	17	25.8 (23.0)	1.5 (1.4)	1.4 (1.0)
248	ATLAS‡ 7 TeV 4.6 fb^{-1}, W/Z combined cross sec.	[19]	34	287.3 (88.7)	8.4 (2.6)	13.7 (4.8)
249	CMS 8 TeV 18.8 fb^{-1} W cross sec. and A_{ch}	[48]	11	11.4 (12.1)	1.0 (1.1)	0.2 (0.4)
250	LHCb 8 TeV 2.0fb^{-1} W/Z cross sec.	[49]	34	73.7 (59.4)	2.1 (1.7)	3.7 (2.6)
251	ATLAS 8 TeV 20.3 fb^{-1} single diff. high-mass cross sec.	[50]	12	20.4 (25.6)	1.7 (2.1)	1.6 (2.3)
253	ATLAS 8 TeV 20.3 fb^{-1}, Z p_T cross sec.	[51]	27	30.2 (28.3)	1.1 (1.0)	0.5 (0.3)
542	CMS 7 TeV 5 fb^{-1}, single incl. jet cross sec., $R = 0.7$ (extended in y)	[52]	158	194.7 (188.6)	1.2 (1.2)	2.0 (1.7)
544	ATLAS 7 TeV 4.5 fb^{-1}, single incl. jet cross sec., $R = 0.6$	[53]	140	202.7 (203.0)	1.4 (1.5)	3.3 (3.4)
545	CMS 8 TeV 19.7 fb^{-1}, single incl. jet cross sec., $R = 0.7$, (extended in y)	[54]	185	210.3 (207.6)	1.1 (1.1)	1.3 (1.2)
573	CMS 8 TeV 19.7 fb^{-1}, $t\bar{t}$ norm. double-diff. top p_T & y cross sec.	[55]	16	18.9 (19.1)	1.2 (1.2)	0.6 (0.6)
580	ATLAS 8 TeV 20.3^{-1}, $t\bar{t}$ p_T^t and $m_{t\bar{t}}$ abs. spectrum	[56]	15	9.4 (10.7)	0.6 (0.7)	-1.1 (-0.8)

TABLE II. High precision LHC measurements employed in the CT18(Z) analysis. The numbers in round brackets are for the CT18Z fit. $N_{pt,n}$, χ^2_n are the number of points and value of χ^2 for the n-th experiment at the global minimum. S_n is the effective Gaussian parameter [18, 57, 58] quantifying agreement with each experiment. The ATLAS data, labelled by ‡, are included in the CT18Z global fit, but not in CT18. O

Treatment of new LHC data

- Include processes that have a sensitivity for the PDFs of interest, and for which NNLO predictions are available.
- Include as large a rapidity interval for the jet data as possible
  - for ATLAS this involves using the ATLAS de-correlation model, rather than using a single rapidity interval. Using a single rapidity interval may result in selection bias. The result is a worse χ^2 due to the remaining tensions in the ATLAS jet data, and a reduced sensitivity compared to the CMS jet data.
  - the use of only a single jet rapidity interval provides incomplete information
- Use multiple t-tbar observables, possible using experimentally provided statistical correlations.
  - and for CMS, using the double differential calculation from Mitov et al
  - again, some of the observables are in tension with each other.
- NB: previous data (including CMS 7 TeV W,Z data) continue having an impact on global fits and tend to dilute the impact of new data

Theory input

CTEQ

Obs.	Expt.	fast table	NLO code	K-factors	R,F scales
Inclusive jet	ATL 7 CMS 7/8	APPLgrid fastNLO	NLOJet++	NNLOJet	p_T, p_T^1
p_T^Z	ATL 8	APPLgrid	MCFM	NNLOJet	$\sqrt{Q^2 + p_{T,Z}^2}$
W/Z rapidity W asymmetry	LHCb 7/8 ATL 7 CMS 8	APPLgrid	MCFM/aMCfast	FEWZ/MCFM	$M_{W,Z}$
DY (low,high mass)	ATL 7/8 CMS 8	APPLgrid	MCFM/aMCfast	FEWZ/MCFM	Q_{ll}
$t\bar{t}$	ATL 8 CMS 8	fastNNLO			$\frac{H_T}{4}, \frac{m_T}{2}$

When justified, a small Monte-Carlo error (typically 0.5%) added for NNLO/NLO K-factors, such as

- ATLAS 7 TeV and CMS 7/8 TeV jet production, and
- ATLAS 8 TeV high-p_T Z production

to account for numerical uncertainties in the MC integration of NNLO cross sections.

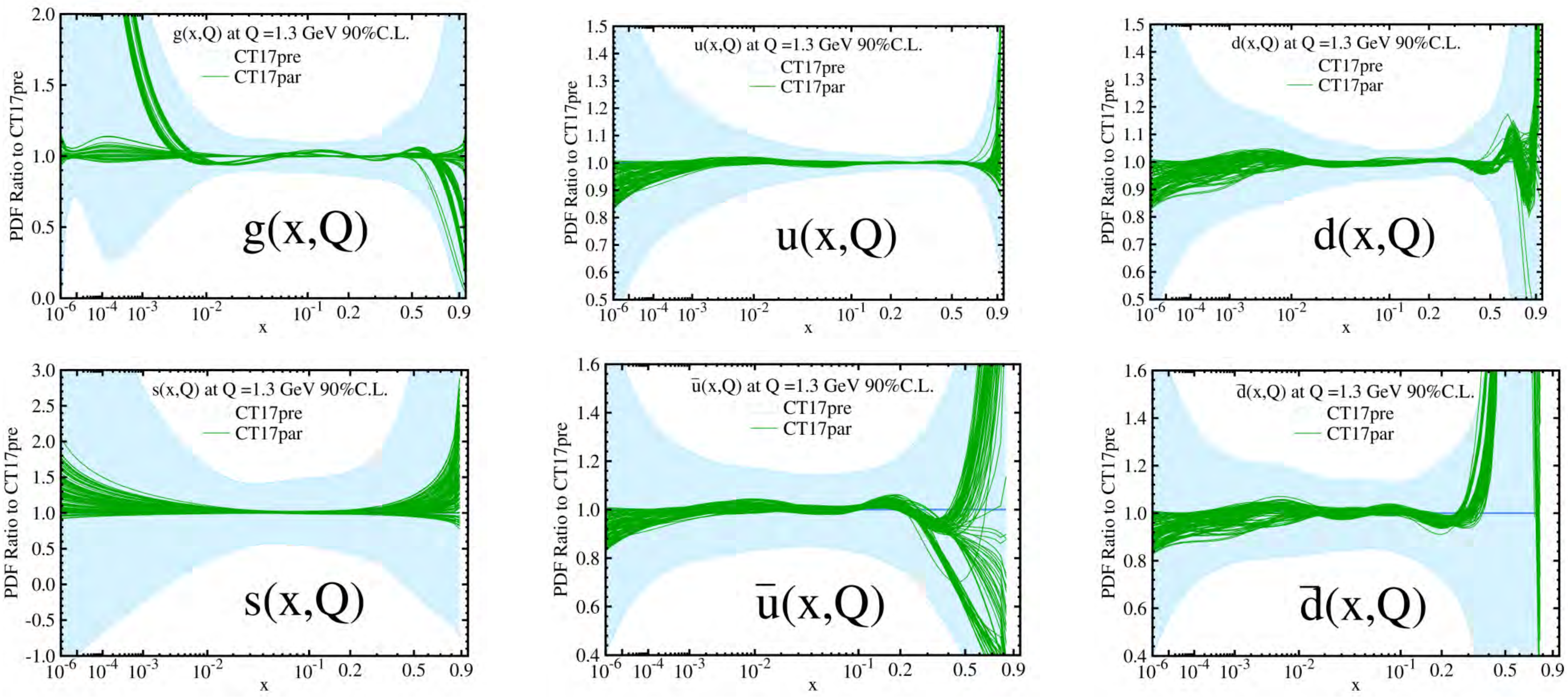
Theoretical calculations for vector boson production

ID	Obs.	Expt.	fast table	NLO code	K-factors	$\mu_{R,F}$
245 246 250 249	$y_{\mu\mu}, \eta_\mu$ y_{ee} $y_{\mu\mu}, \eta_\mu$ $A(\mu)$	LHCb7ZW LHCb8Z LHC8ZW CMS8W	APPLgrid	MCFM/aMCfast	MCFM/FEWZ	$M_{Z,W}$
253	p_T^{ll}	ATL8Z	APPLgrid	MCFM	NNLOJet	M_T^{ll}
201 203 204	$\sqrt{\tau}, y$ $\sigma_{pd}/\sigma_{pp}, x_F$ Q, x_F	E605 E866 E866	CTEQ		FEWZ	Q_{ll}
225 227 234 281	$A(e)$ $A(e)$ $A(\mu)$ $A(e)$	CDF1Z CDF2W DØ2W DØ2W	CTEQ		ResBos	Q_{ll} M_W
260 261	y_{ll} y_{ll}	DØ2 CDF2	CTEQ		VRAP	Q_{ll}
266 267 268	$A(\mu)$ $A(e)$ $y_{ll}, \eta_l, A(l)$	CMS7W CMS7W ATL7ZW(2012)	CTEQ		ResBos	M_W $M_{Z,W}$
248	y_{ll}, η_l	ATL7ZW(2016)	APPLgrid	MCFM/aMCfast	MCFM/FEWZ	$M_{Z,W}$

Other theoretical issues

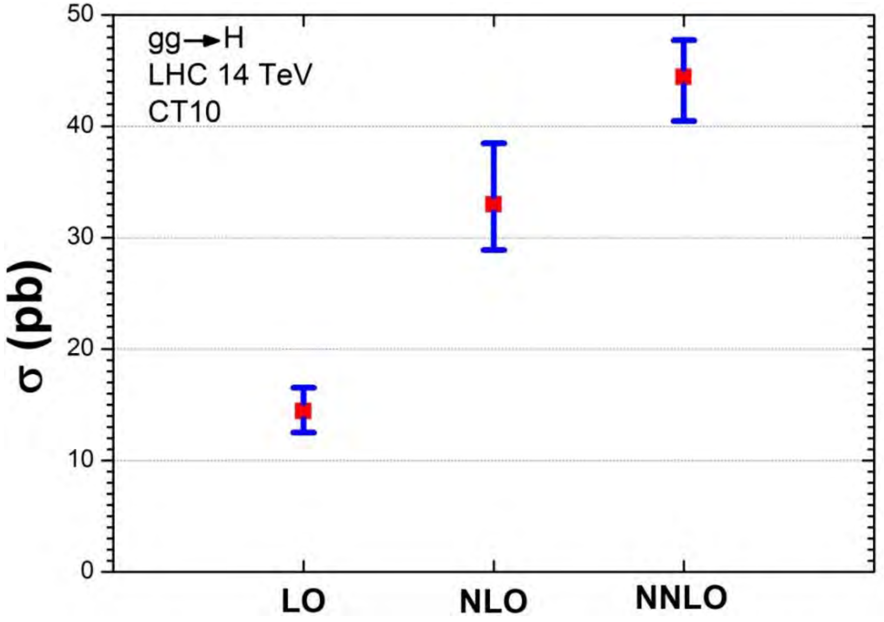
- We have been using the “canonical” factorization and renormalization scales which provide the best fit to the data or stabilize high order correction. (E.g., inclusive pT in jet and HT/4 in t-tbar productions.)
- We have considered the impact of non-perturbative PDF functional forms to PDF errors.
- We have not included theoretical errors due to higher order corrections. Varying the factorization or renormalization scales by a factor of 2 does not always provide the correct information about not-yet included higher order corrections.
- Sizable (of the order of 1%) difference is found in various NNLO calculations for W and Z productions at the LHC, while the statistical error of precision W and Z data is about 0.1%.

Explore various non-perturbative parametrization forms of PDFs



- CT18 – sample result of exploring various non-perturbative parametrization forms.
- There is no data to constrain very large or very small x region.

Fixed order gg → H cross sections



arXiv: 1310.7601

HNNLO:

arXiv: hep-ph/0703012

CT10 NNLO PDFs:

arXiv: 1302.6246

- Vary the scale by a factor of 2 around M_H=125 GeV, using the HNNLO code, with CT10 NNLO PDFs.
- Cross sections (pb) with scale uncertainties:

$$\mathrm{LO}\left(14.44^{+14.5\%}_{-13.4\%}\right) \rightarrow \mathrm{NLO}\left(33.00^{+16.6\%}_{-12.4\%}\right) \rightarrow \mathrm{NNLO}\left(44.41^{+7.45\%}_{-8.89\%}\right)$$

- Very large corrections in going from

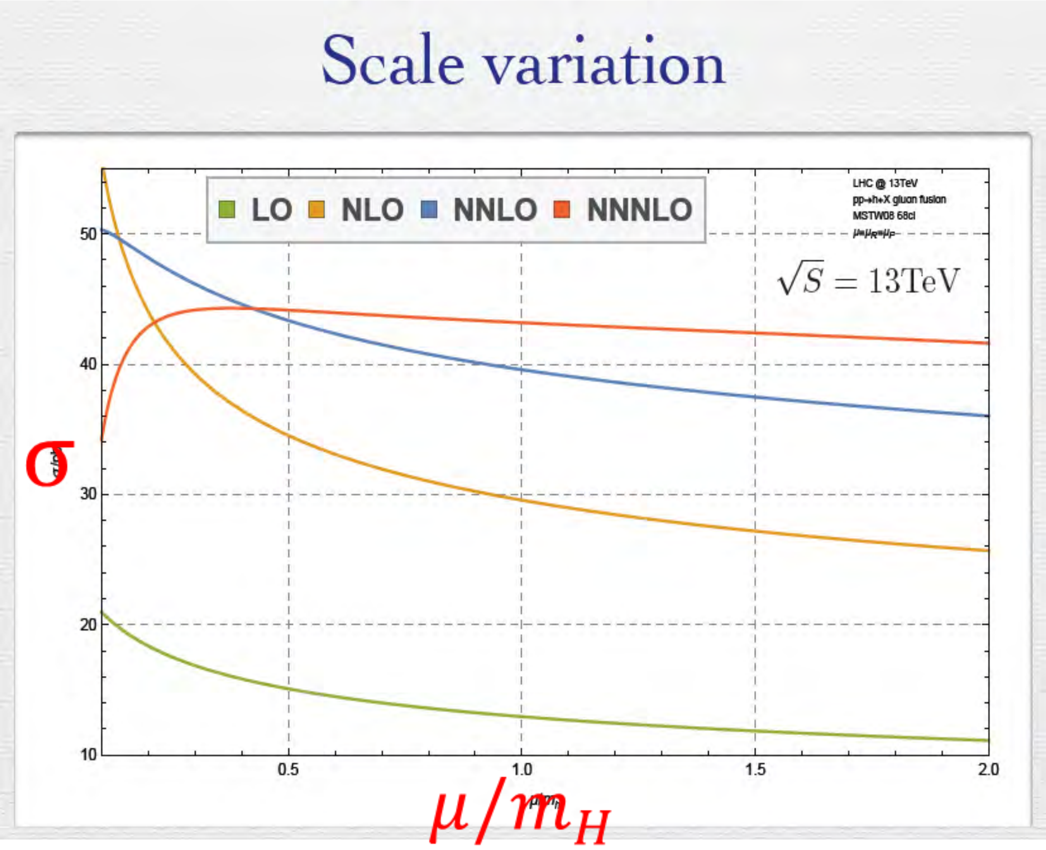
$$\mathrm{LO}\left(K = 1.00^{+0.15}_{-0.13}\right) \rightarrow \mathrm{NLO}\left(K = 2.28^{+0.38}_{-0.28}\right) \rightarrow \mathrm{NNLO}\left(K = 3.07^{+0.23}_{-0.27}\right)$$

- Soft dynamics alone cannot describe NLO or NNLO results.
  Hard gluon dynamics is important.

gg → H cross section at NNNLO (N3LO) in QCD

$$\sigma = 48.58\,\text{pb}^{+2.22\,\text{pb}\,(+4.56\%)}_{-3.27\,\text{pb}\,(-6.72\%)}\ (\text{theory}) \pm 1.56\,\text{pb}\,(3.20\%)\ (\text{PDF}+\alpha_s)$$

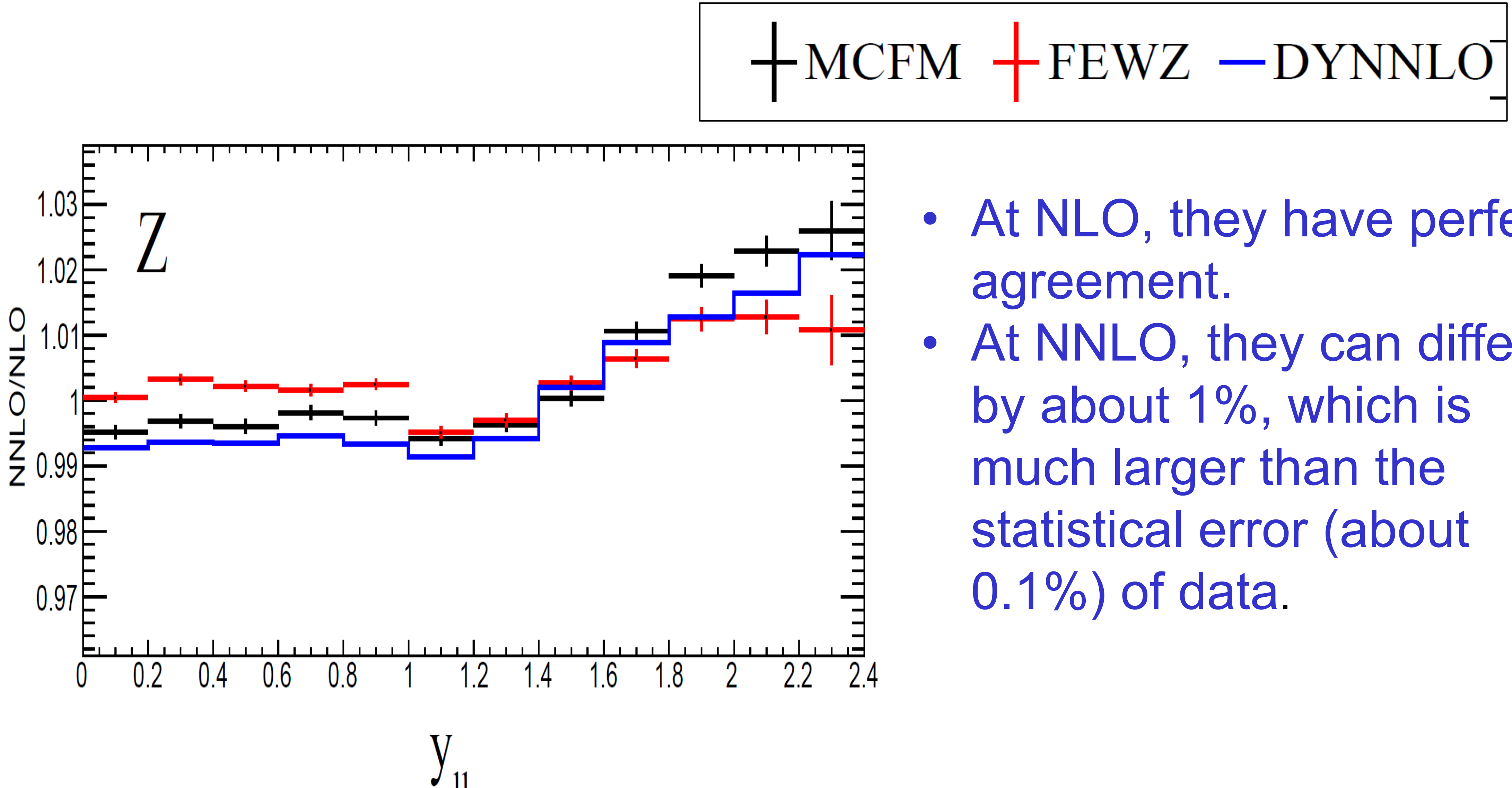
At 13 TeV LHC



arXiv:1602.00695

- The scale variation is clearly reduced at N3LO
- NNLO and NNNLO agree at $\mu = 0.5\, m_H$

ATLAS 7 TeV W and Z data vs. various NNLO calculations



- At NLO, they have perfect agreement.
- At NNLO, they can differ by about 1%, which is much larger than the statistical error (about 0.1%) of data.

We have also asked:

Which of 30+ eligible LHC experiments provide promising constraints on the CTEQ-TEA PDFs?

Do the LHC experiments agree among themselves and with other experiments?

A **consistent** answer emerges from a powerful combination of four methods:

1. **Lagrange multiplier scans** } slow, most accurate
2. **PDFSense and L_2 sensitivity**
3. **ePump** **[Schmidt, Pumplin, Yuan, PRD 98, 094005]**
4. **Effective Gaussian variables**

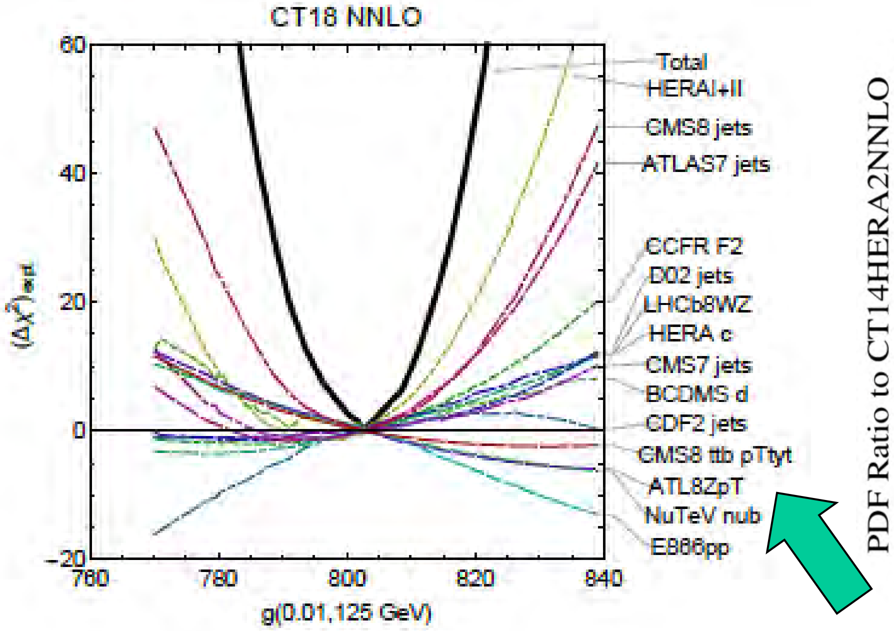
} Fast approximations (methods 2–4)

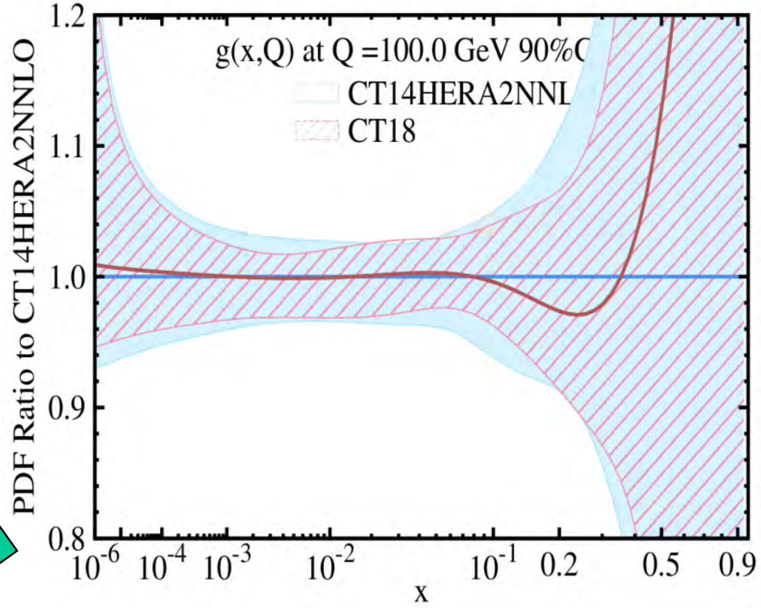
H.-L. Lai et al., arXiv:1007.2241

The LM scan technique was introduced in **Stump et al., Phys. Rev. D65 (2001) 014012**

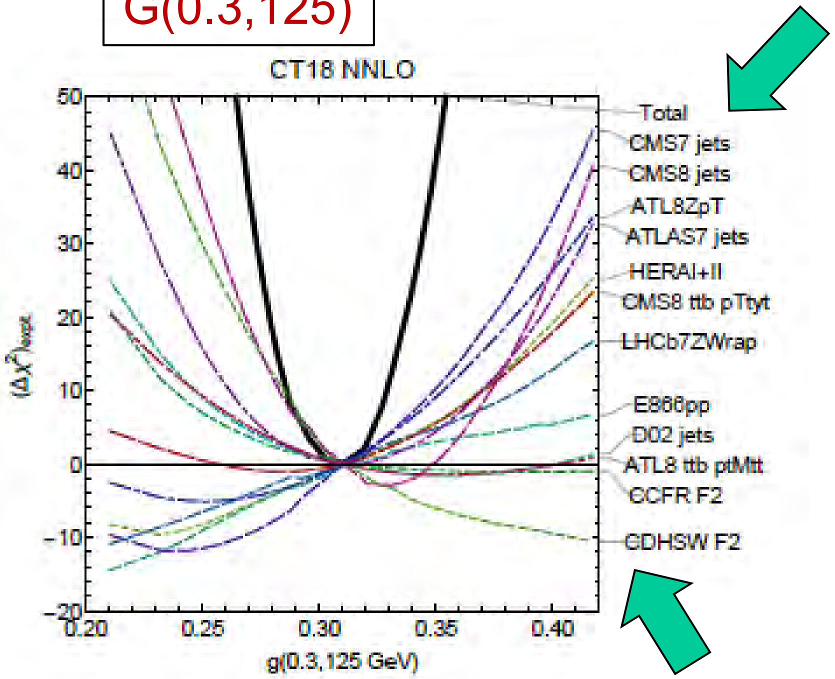
Preview of CT18 PDFs (g-PDF)

G(0.01,125)





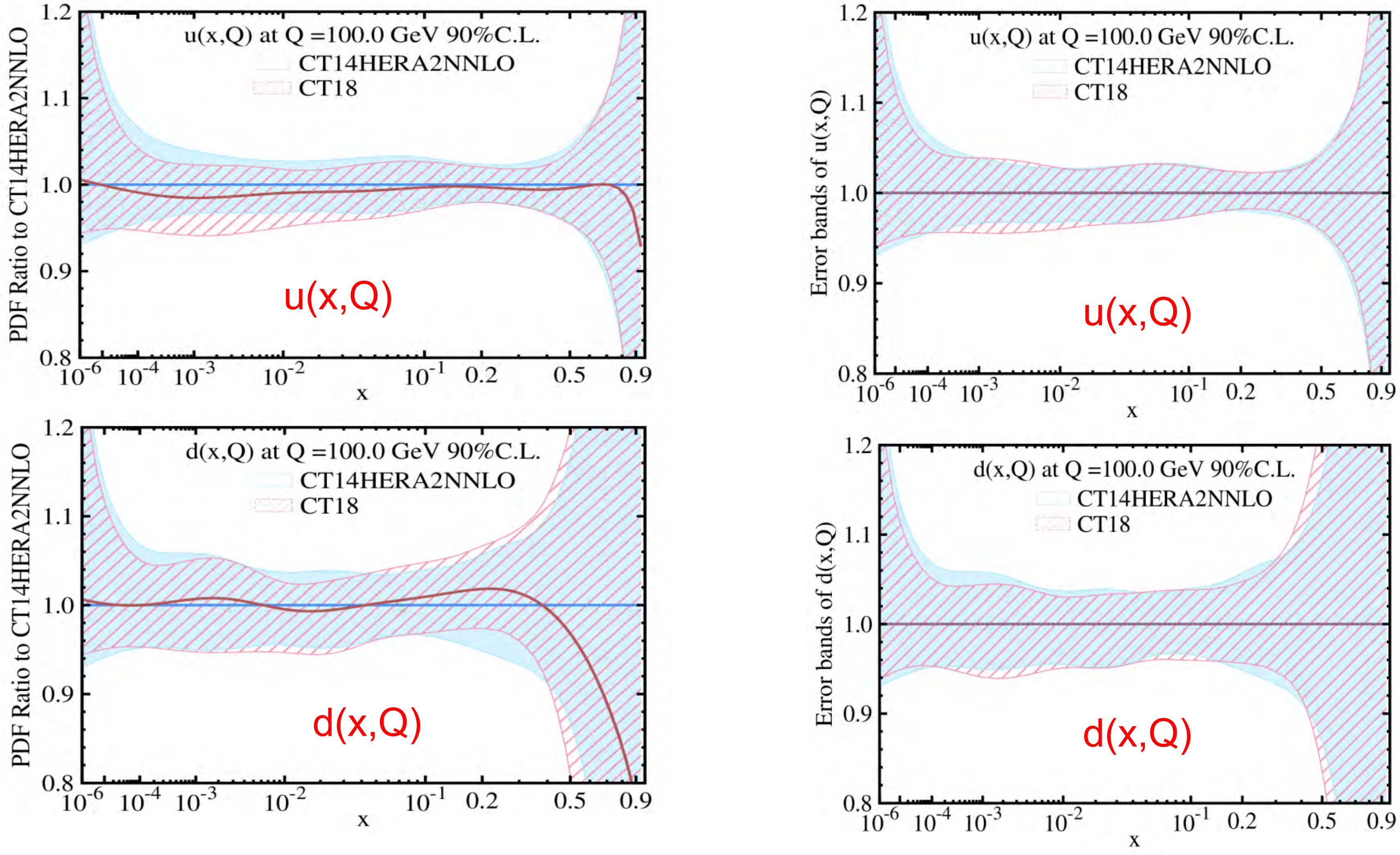
G(0.3,125)



Lagrange Multiplier Scans

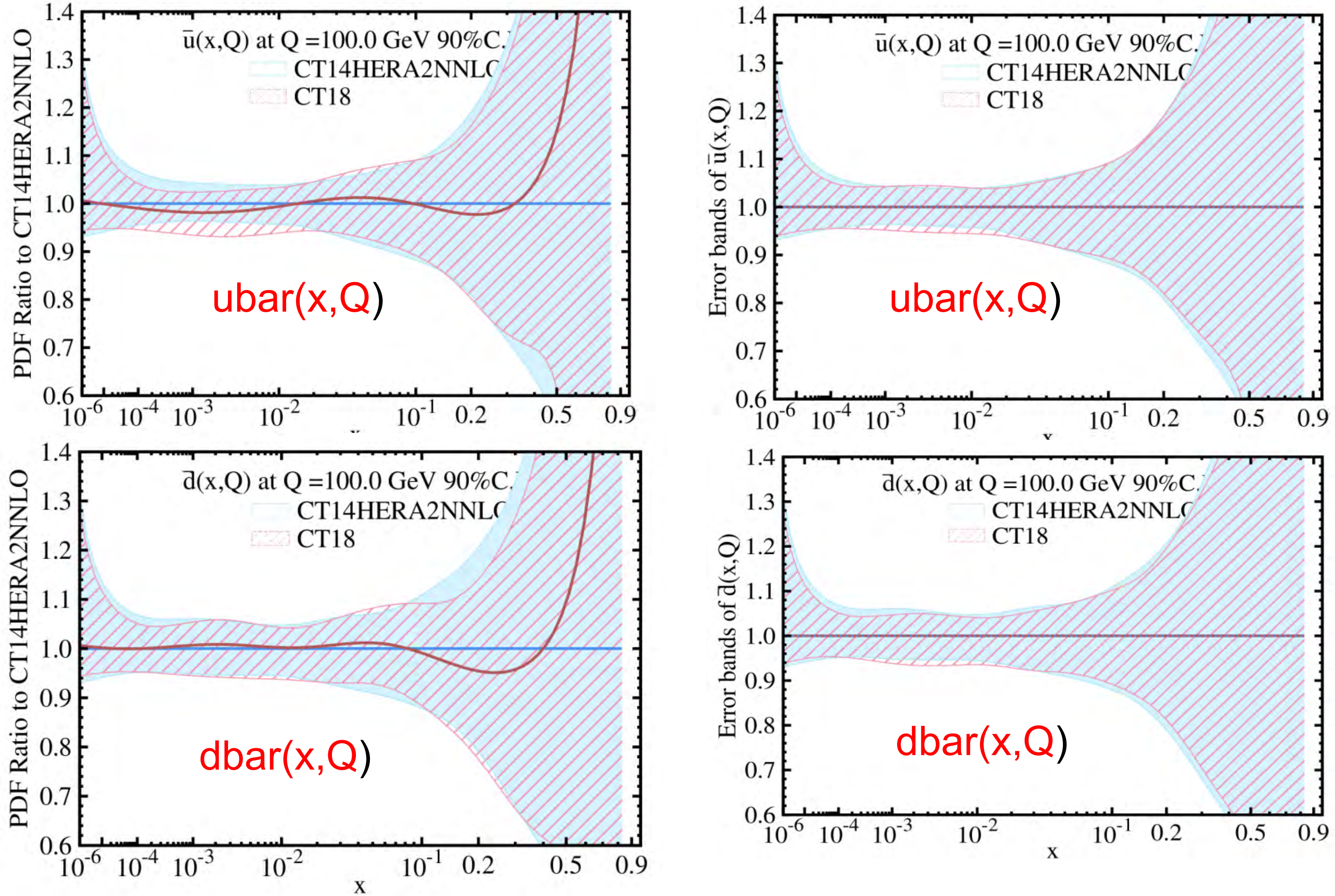
- At x around 0.01, ATLAS8 Z pT data prefer a slightly larger gluon PDF.
- At x around 0.3, competing with the CDHSW F2 and Tevatron jet data, which prefer larger gluon, the ATLAS7 jet, CMS7 jet and ATLAS8 Z pT data prefer a smaller gluon; some tension found in CMS7 and CMS8 jet data.
- The gluon PDF as x → 1 is parametrization form dependent.

Preview of CT18 (u-PDF and d-PDF)



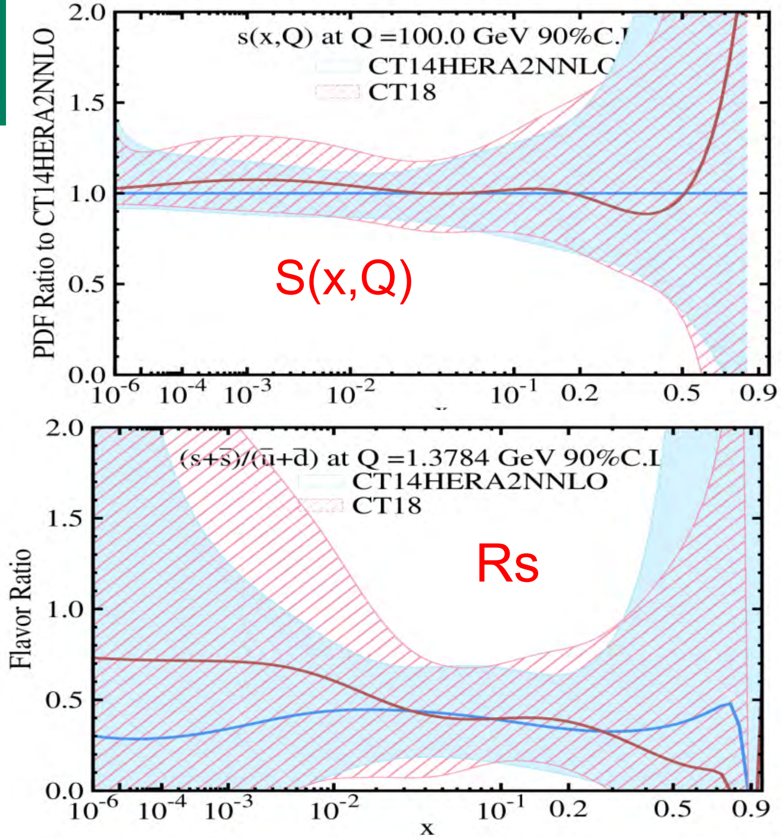
- Some changes on u and d at small x, and d around 0.2; mainly come from LHCb W and Z rapidity data, at 7 and 8 TeV.

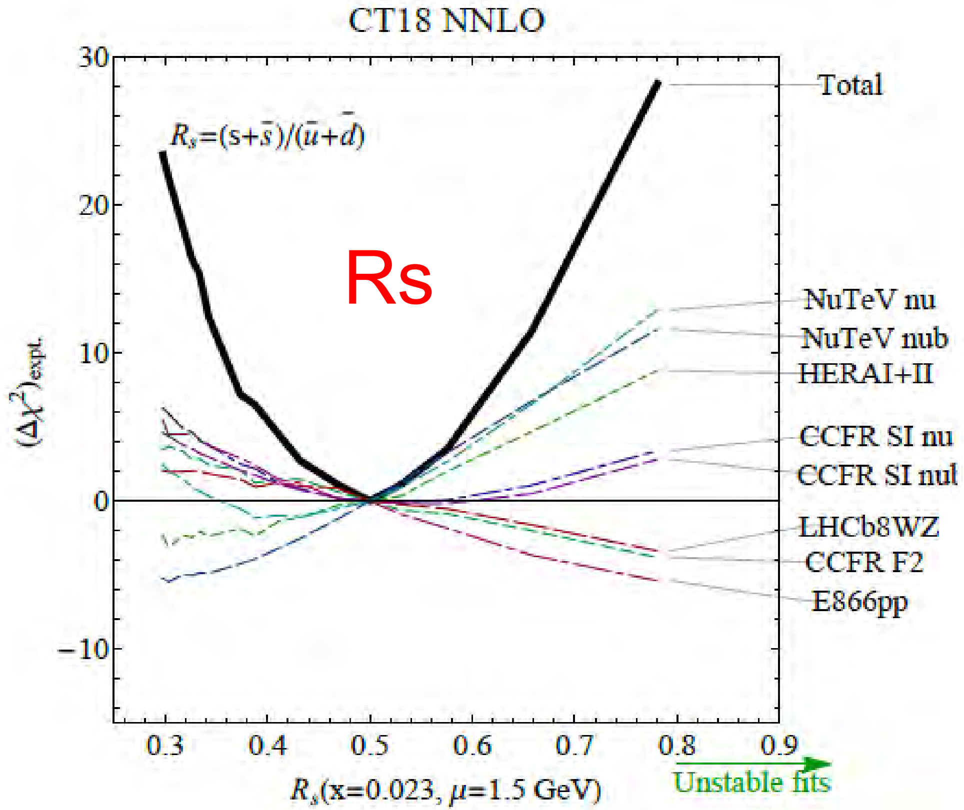
Preview of CT18 (ubar and dbar PDF)



- Minor changes on ubar and dbar PDFs at small x region mainly come from LHCb W and Z rapidity data, at 7 and 8 TeV.
- The behavior of ubar and dbar PDFs, as $x \to 1$, is parametrization form dependent.

Rs=(s+sbar)/(ubar+dbar)





- LHCb W and Z (7,8 TeV) data prefer a larger s-PDF in the small-x region.
- NuTeV dimuon data strongly prefer a smaller Rs value, while the LHCb WZ data prefer a slightly larger Rs value.
- Rs (CT18)= 0.5 ± 0.3 for x = 0.023 and Q^2 = 1.9 GeV^2 .

  (Compare to ATLAS with $R_s = 1.13^{+0.08}_{-0.13}$)

u and d PDFs in CT14HERA2

arXiv: 1907.12177
ePump study

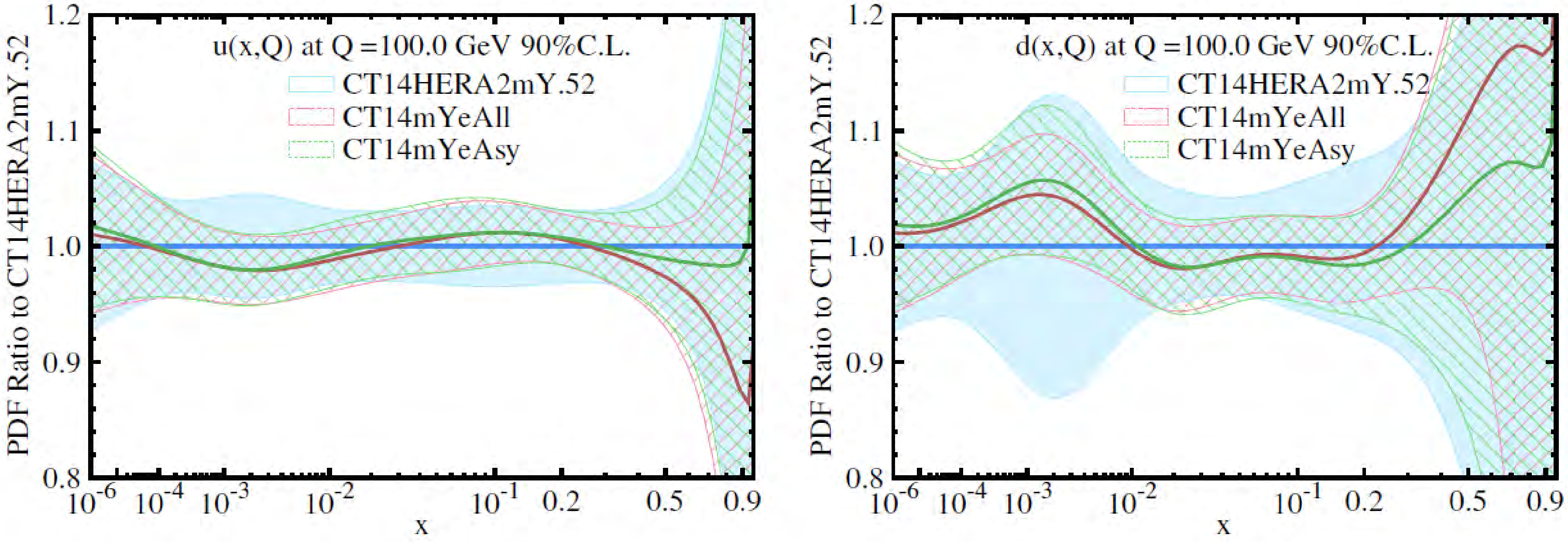
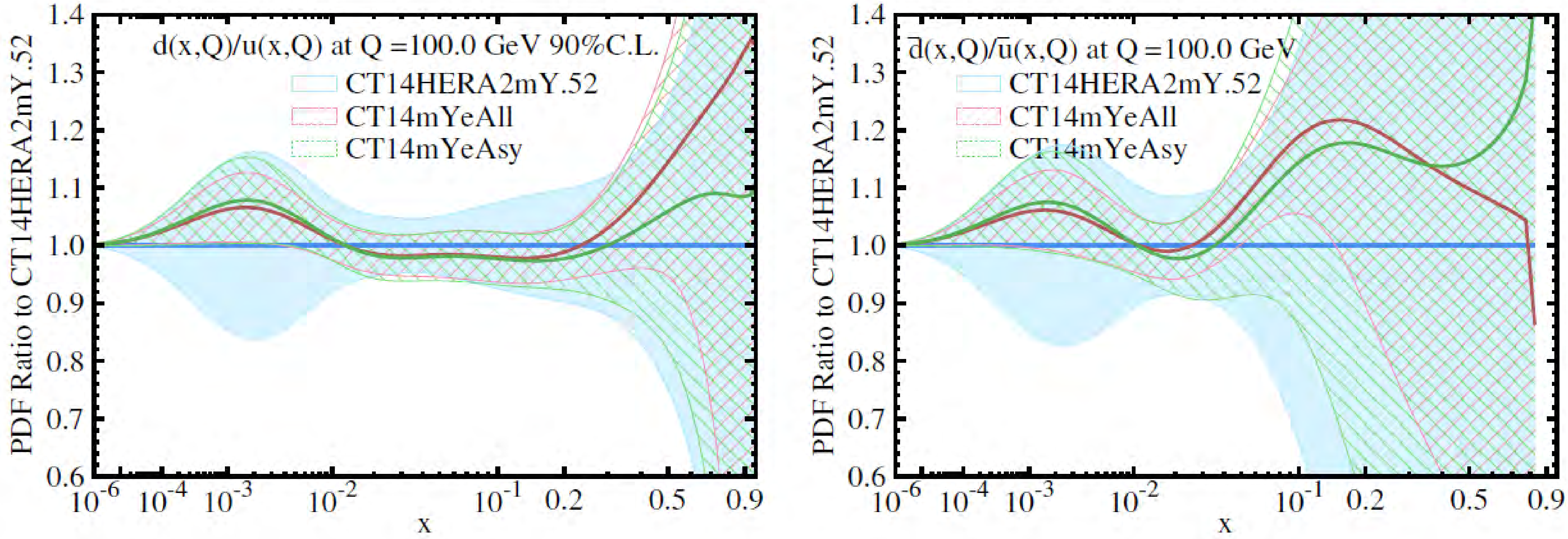


FIG. 9: Comparison of CT14mYeAsy and CT14mYeAll for u and d PDFs at $Q = 100$ GeV. CT14mYeAsy is obtained by adding CMS 7 TeV μ asymmetry data, CMS 7 TeV electron asymmetry data, ATLAS 7 TeV WZ data and DØ Run2 μ asymmetry data to CT14HERA2mY, using `ePump`. The PDF ratios are over the best-fit of CT14HERA2mY.

d/u and dbar/ubar PDFs in CT14HERA2



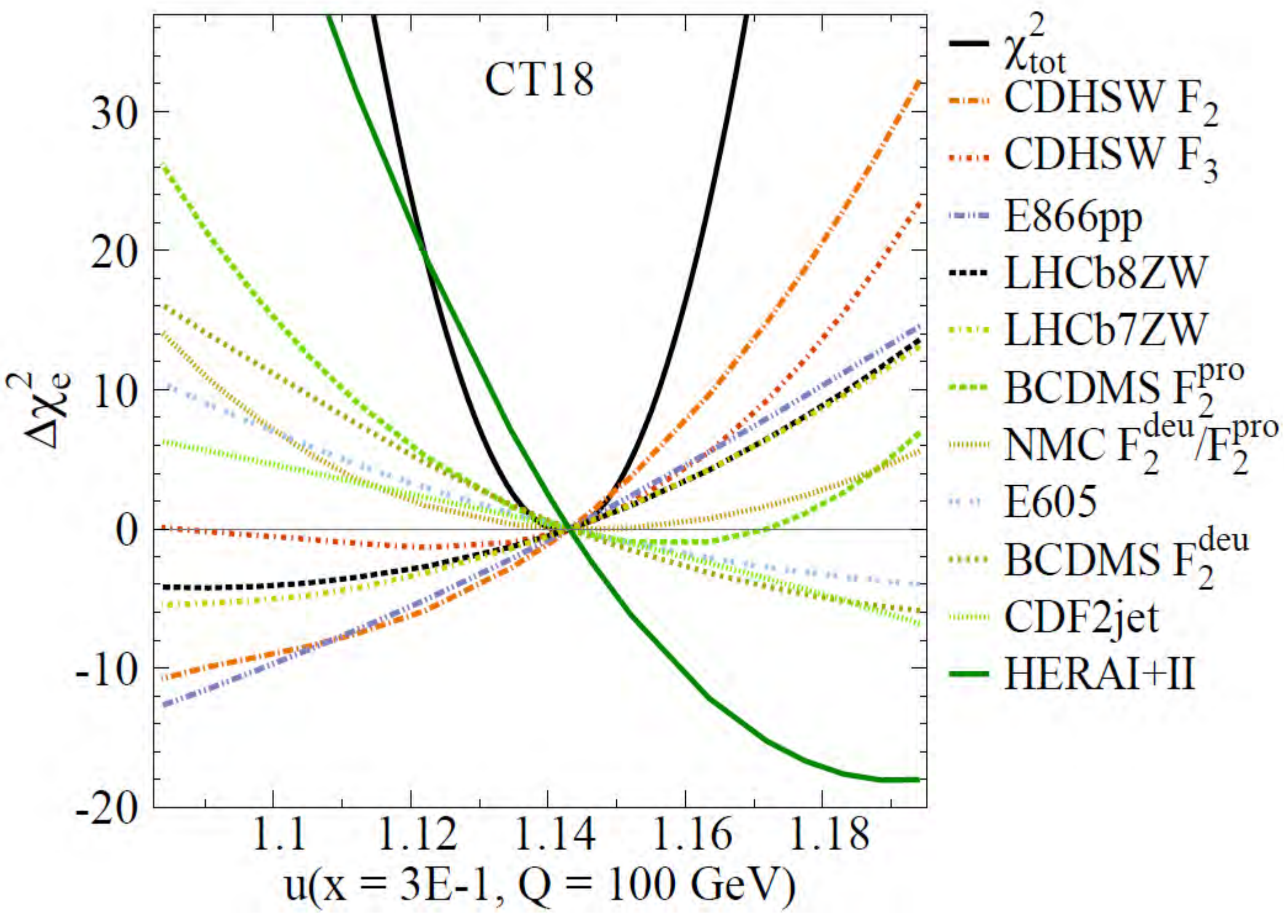
CT14mYeAsy is obtained by adding CMS 7 TeV μ asymmetry data, CMS 7 TeV electron asymmetry data, ATLAS 7 TeV WZ data and DØ Run2 μ asymmetry data to CT14HERA2mY, using `ePump`. The PDF ratios are over the best-fit of CT14HERA2mY.

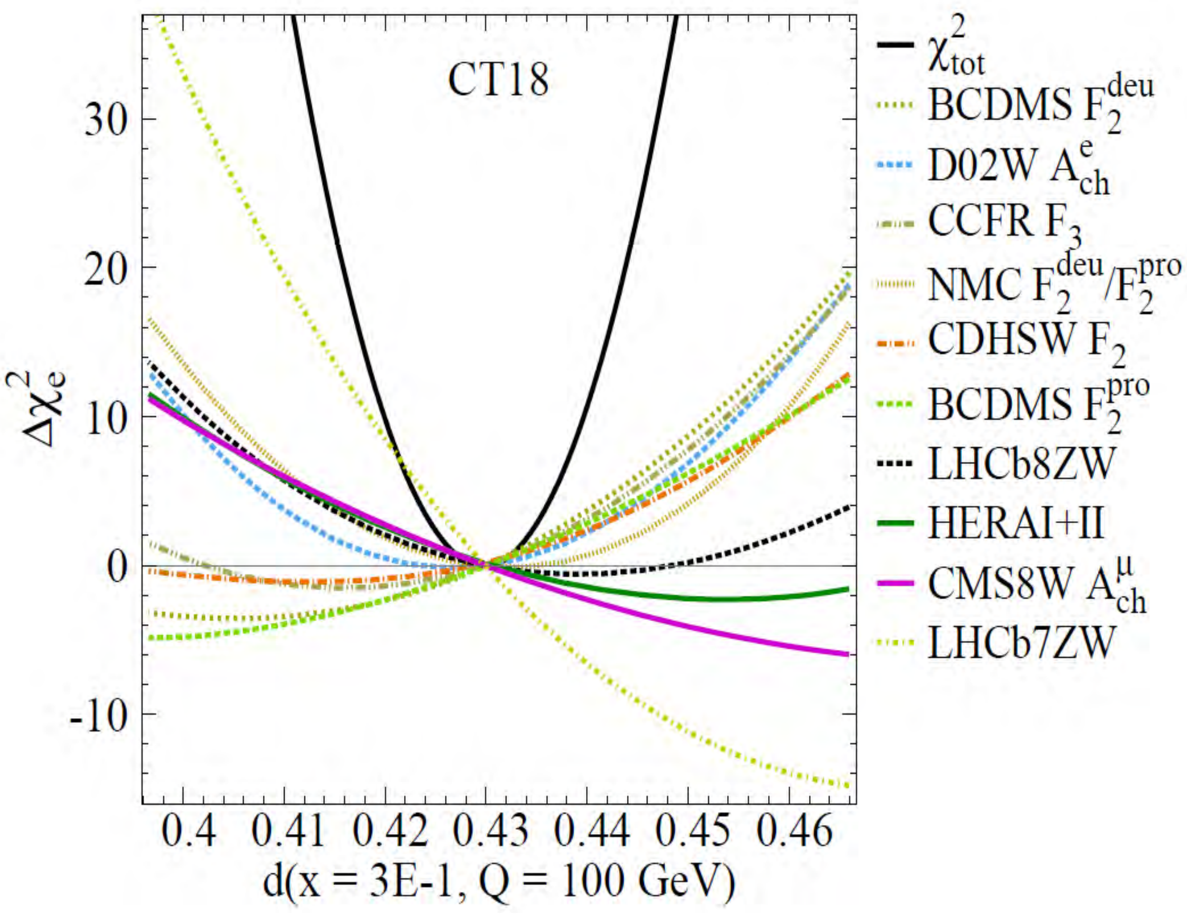
Focus on PDF constraints at large x

For this Workshop –

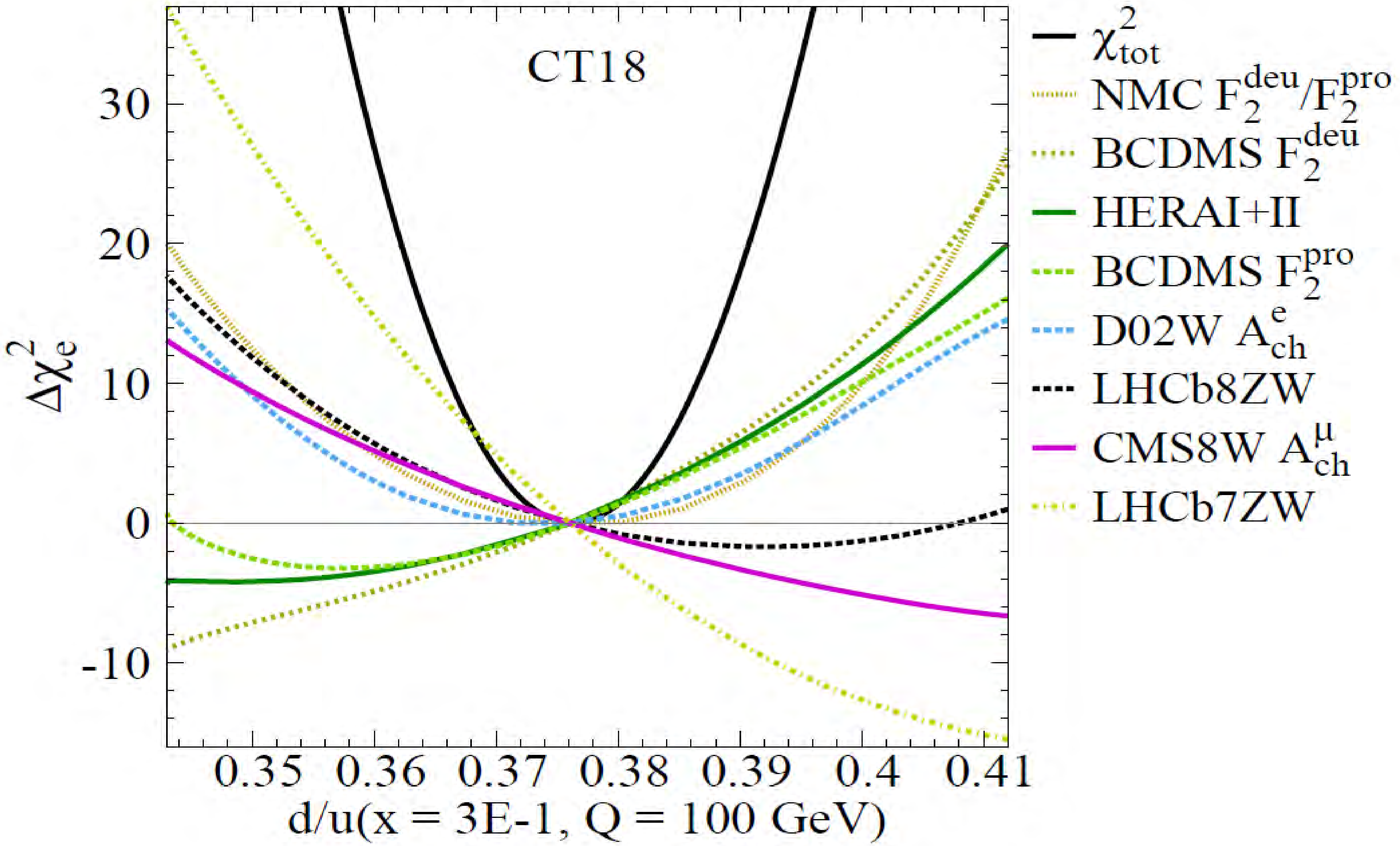
LPC Workshop on Physics Connections between the LHC and EIC

u and d at x=0.3, Q=100 GeV

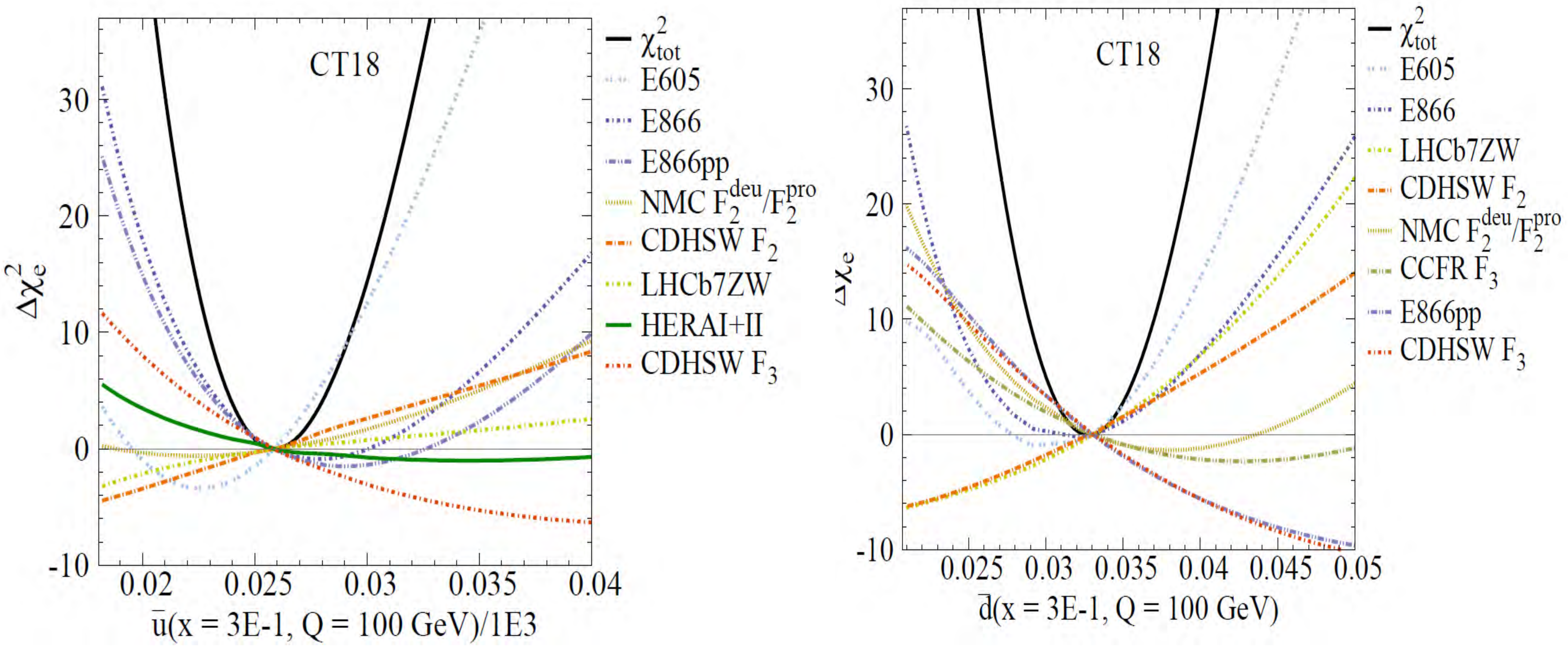




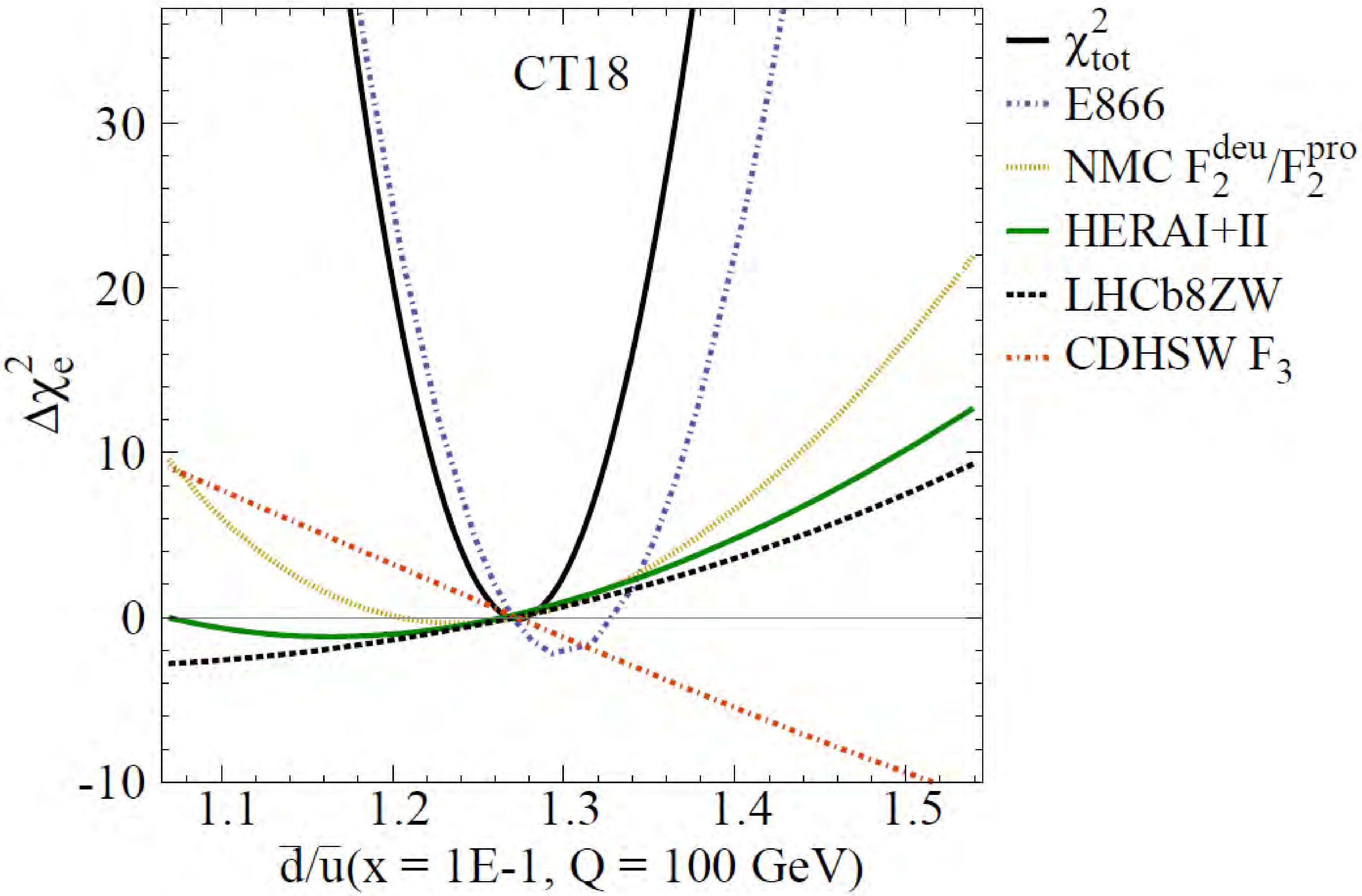
d/u at x=0.3, Q=100 GeV



ubar and dbar at x=0.3, Q=100 GeV



dbar/ubar at x=0.1, Q=100 GeV



CT18Z
LHC data treatment

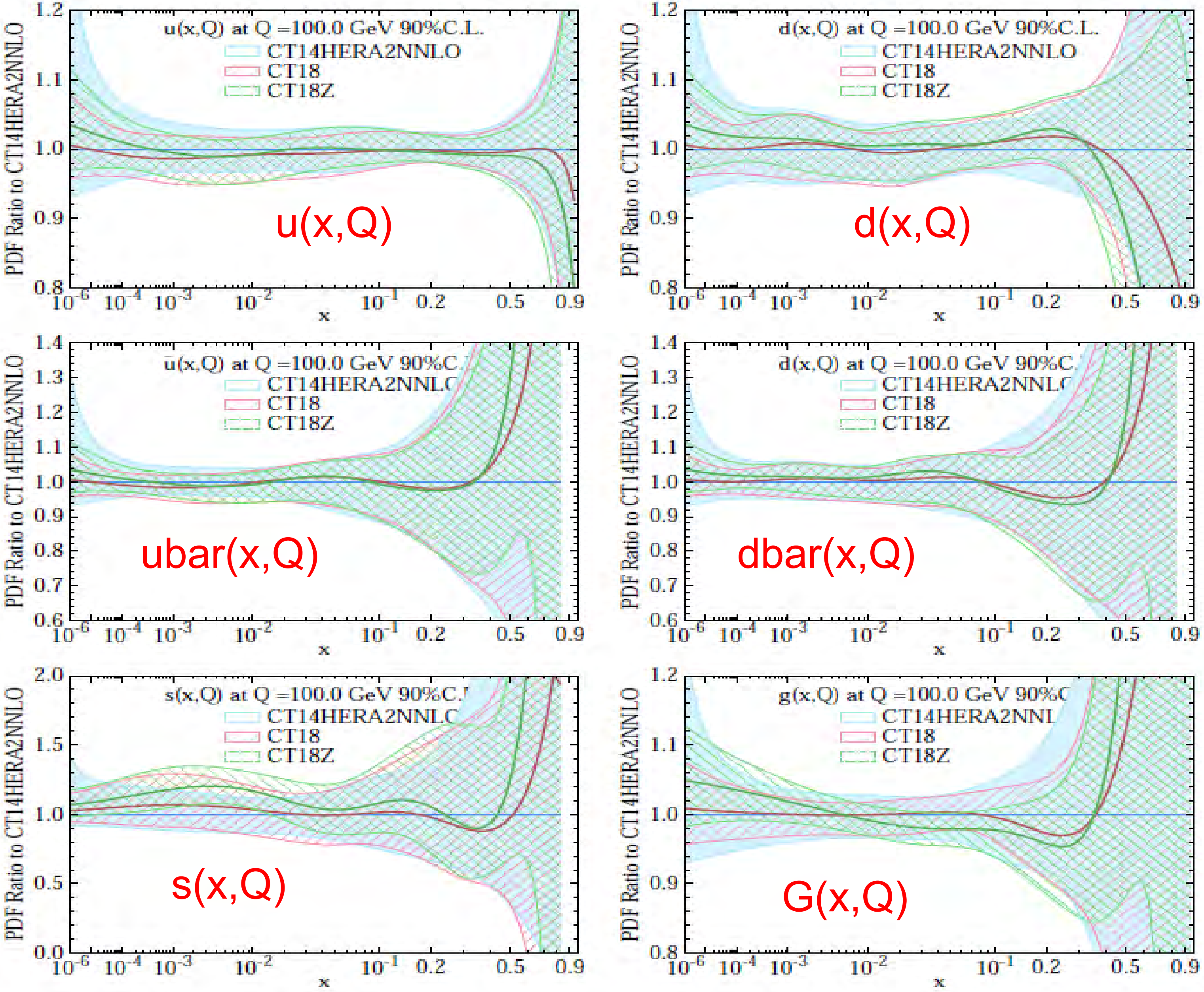
- Start with CT18 data set
- Add in ATLAS 7 TeV *W* and *Z* rapidity data (arXiv:1612.03016; 4.6 1/fb); large chi^2/d.o.f ~ 2.1
- Remove CDHSW data
- Use a special x-dependent factorization scale $\mu_{DIS,x}$ at NNLO calculation.
- CT18Z uses a combination of $\mu_{DIS,x}$ (preferred by DIS) and an increased m_c^{pole} = 1.4 GeV (preferred by LHC vector boson production, disfavored by DIS)

→ CT18Z PDFs

CT18Z vs.CT18 PDFs

u and d increase at small-x

s increases at small-x

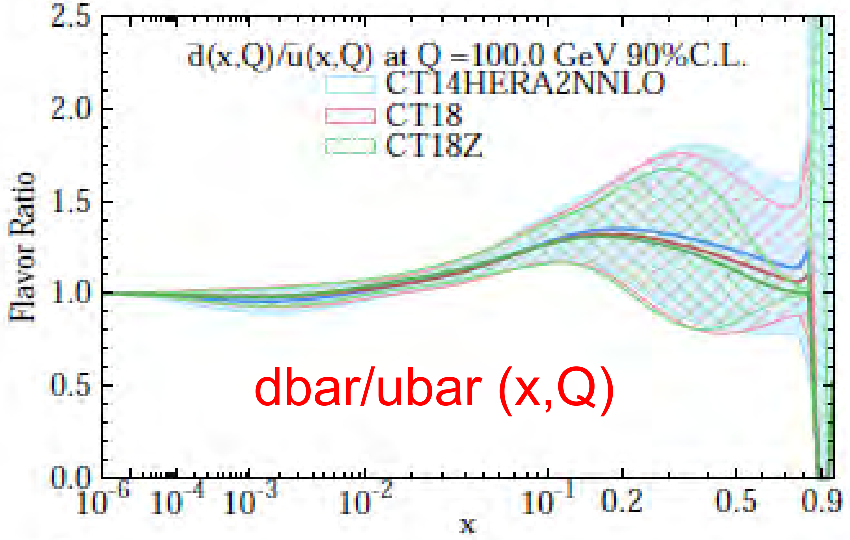


d increases at $x \sim 0.2 - 0.3$

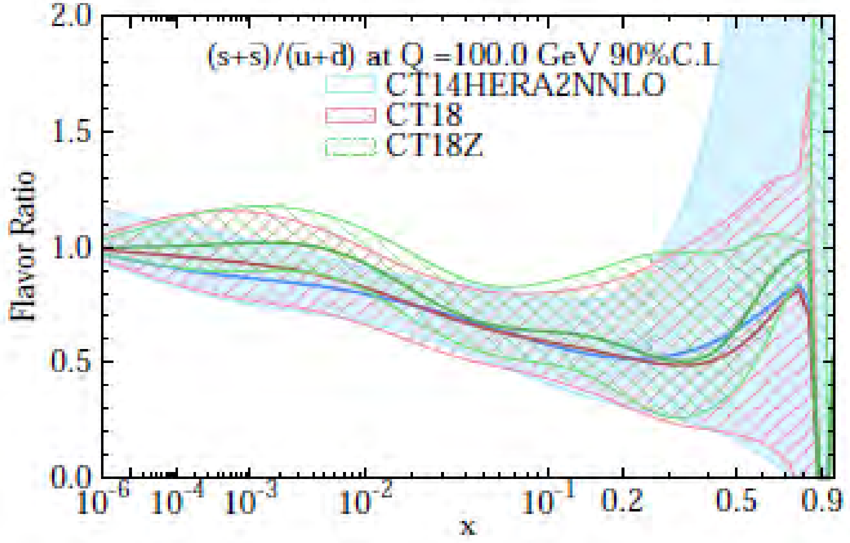
Q=100 GeV; at 90%CL

G increases at small-x, and decreases at $x \sim 0.01 - 0.3$

CT18Z vs.CT18 PDFs

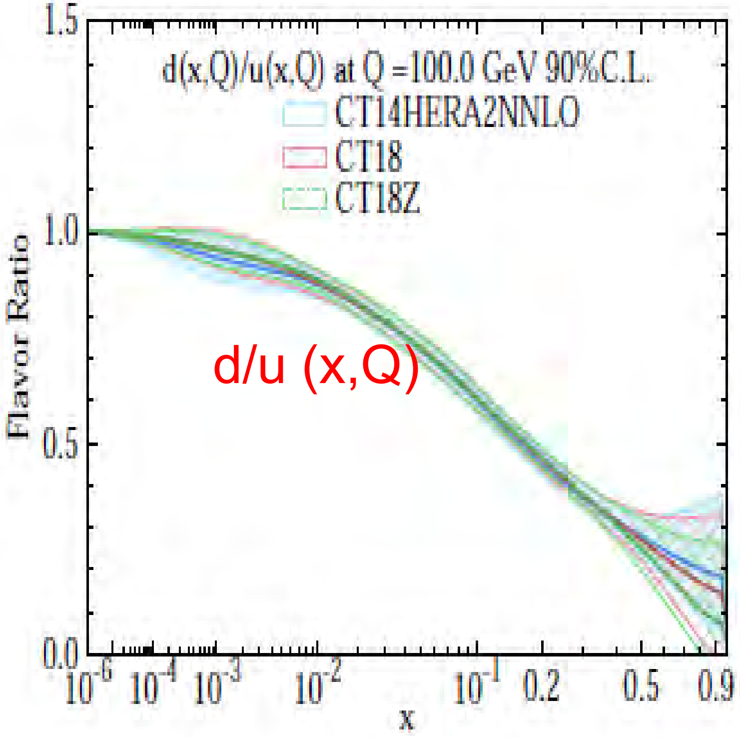


dbar/ubar (x,Q)



(s+sbar)/(ubar+dbar) (x,Q)

Rs increases at small-x

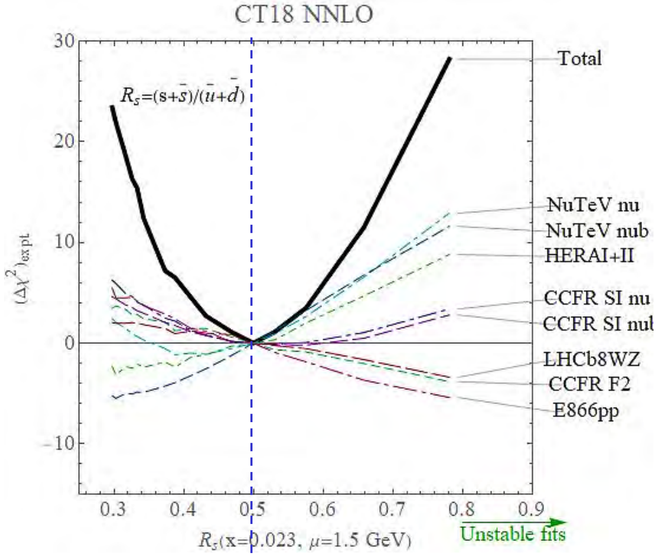


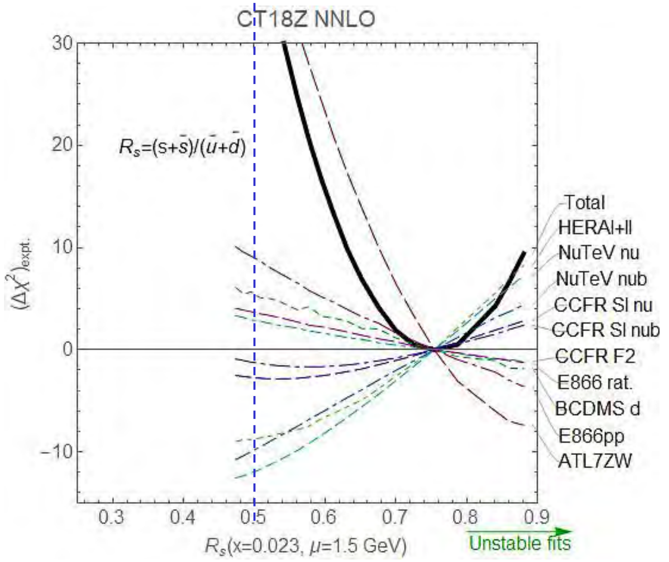
d/u (x,Q)

d/u decreases at large-x

Q=100 GeV; at 90%CL

Lagrange Multiplier scan: $\mathrm{R_s}(x = 0.023, \mu = 1.5\ \mathrm{GeV})$

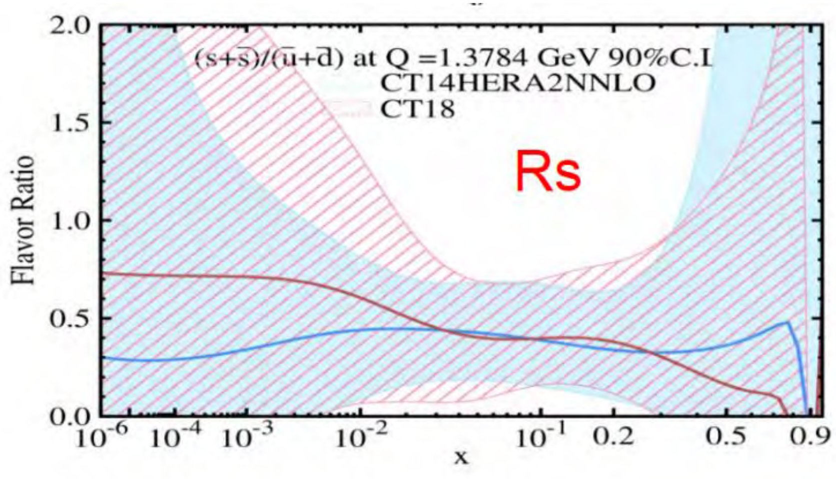




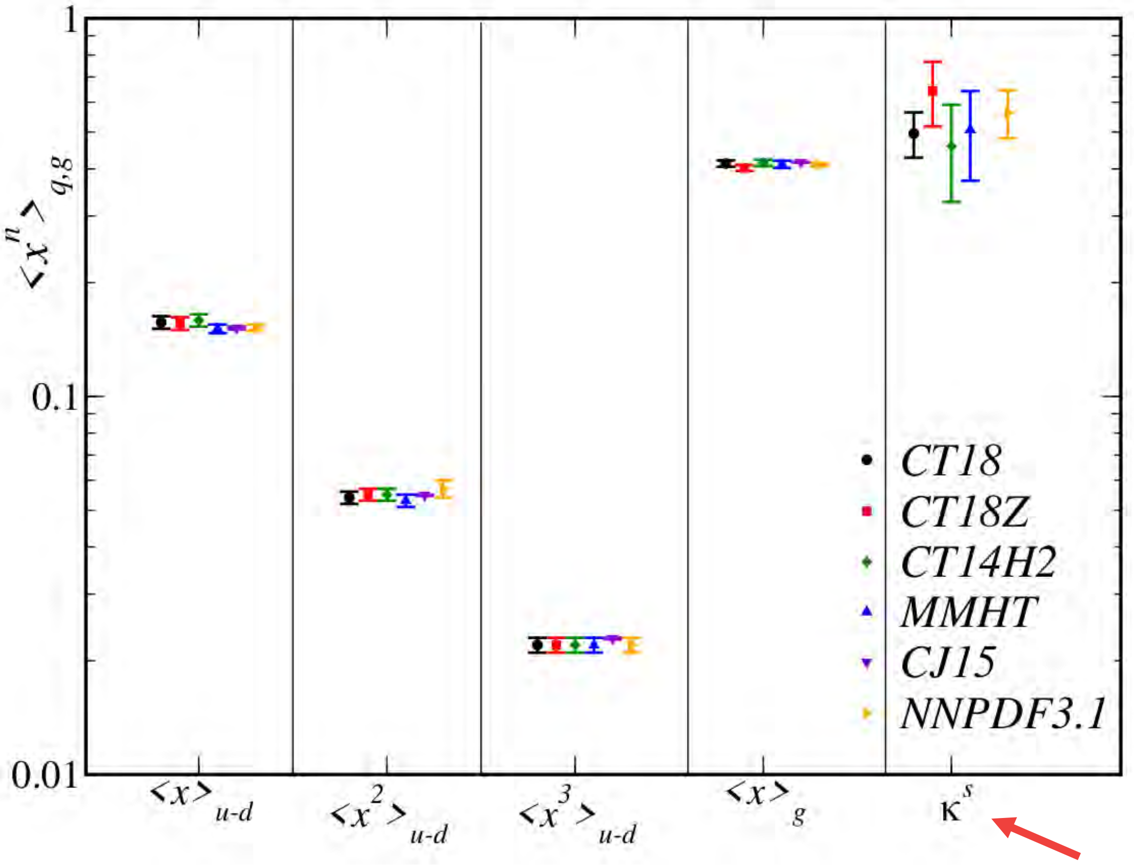
The CT18Z strangeness is increased primarily as a result of including the ATLAS 7 TeV W/Z production data (not in CT18), as well as because of using the DIS saturation scale and $m_c^{pole} = 1.4$ GeV

In either CT18 or CT18Z fit, observe instability in the fits for $R_s > 1$ at $x = 0.01 - 0.1$

Compare to



PDF moments from CT18



$$\langle x^n \rangle_{q,g} = \int dx\, x^n f_{q,g}(x, \mu = 2\,\mathrm{GeV})$$

- progress in lattice QCD, which can evaluate PDF Mellin moments, suggests a possible future synergy with QCD global analyses

PDF-Lattice whitepaper – Lin et al., arXiv:1711.07916.

PDFSense analysis – Hobbs et al., arXiv:1904.00022.

→ good agreement among phenom. predictions of isovector, gluon moments!

→ constraints are significantly weaker for moments of the light quark sea distributions, e.g., the strangeness suppression ratio, $\kappa^s \equiv \langle x \rangle_{s+\bar{s}} / \langle x \rangle_{\bar{u}+\bar{d}}$

CT18 moment results

PDF moment	CT18	CT18Z	CT14H2
$\langle x \rangle_{u^+-d^+}$	0.157 ± 0.006	0.156 ± 0.006	0.159 ± 0.006
$\langle x^2 \rangle_{u^--d^-}$	0.054 ± 0.002	0.055 ± 0.002	0.055 ± 0.002
$\langle x^3 \rangle_{u^+-d^+}$	0.022 ± 0.001	0.022 ± 0.001	0.022 ± 0.001
$\langle x \rangle_g$	0.413 ± 0.008	0.402 ± 0.007	0.415 ± 0.008
κ^s	0.496 ± 0.068	0.643 ± 0.125	0.459 ± 0.132

$\kappa^s(\mu = 2\,GeV)$

$$\kappa^s(\mu = 2\,\mathrm{GeV}) = 0.795 \pm 0.079\,(stat) \pm 0.053\,(sys)$$

- First Lattice calculation by χQCD arXiv: 1901.07526
- It agrees with CT18Z result within 1σ level.

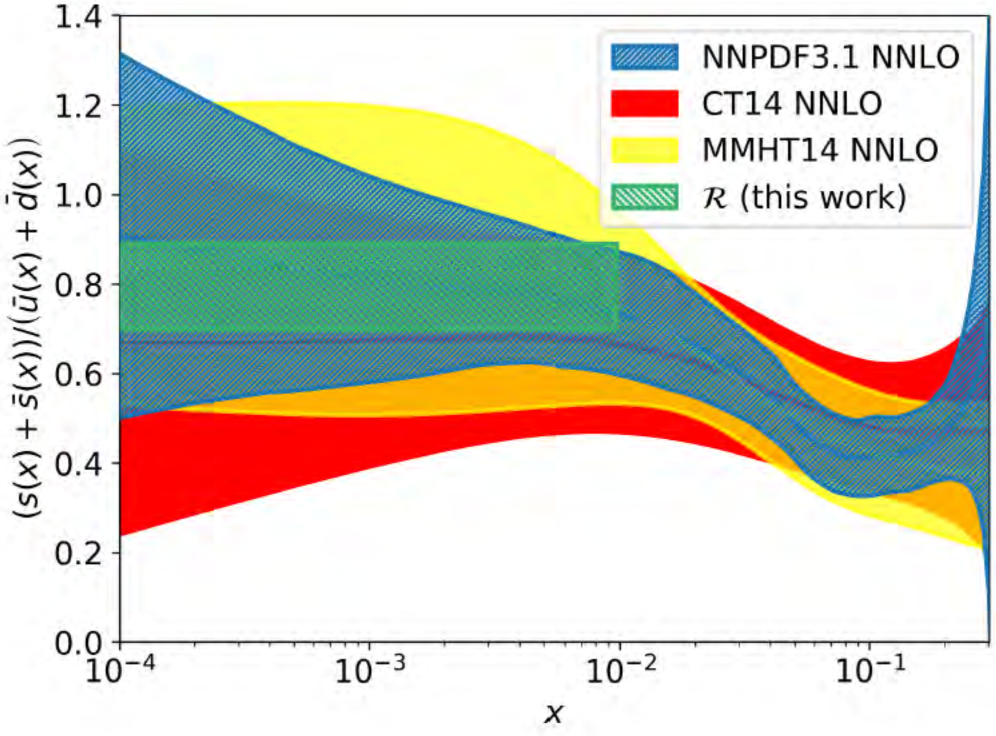


Figure 4. The global fitting results of $(s(x)+\bar{s}(x))\,/\,(\bar{u}(x)+\bar{d}(x))$ at $Q^2 = 4\ \mathrm{GeV}^2$. The green band shows our result under the assumption that the ratio is a constant for small x up to $x = 10^{-2}$.

PDF Profiling method

To study the potential of future (pseudo) data to constrain PDFs, in turn to determine electroweak parameters

xFitter vs. ePump

arXiv: 1907.12177
ePump study

Updating and Optimizing Error PDFs with ePump (error PDF Update Method Package)

Carl Schmidt
Michigan State University

In Collaboration with Jon Pumplin, C.-P. Yuan,
Tie-Jiun Hou, Sayipjamal Dulat, Zhite Yu,

July 16, 2019
QCD@LHC 2019, Buffalo, NY

Motivation for ePump

- UpdatePDFs: With many data sets and NNLO calculations, global fitting can be time consuming.
  - Need for fast and efficient method to estimate effects of new data before doing global fit.
  - Can estimate effects of different data set choices in real time.
  - Can be done without full global fitting machinery.
- OptimizePDFs: Experimental analyses may require many MC calculations, using PDF error sets. Again, it's time consuming.
  - Optimize Hessian error PDFs to the observables, so irrelevant error PDFs may be discarded, while PDF-dependence is still maintained to desired precision.

Hessian Updating

C T E Q

- PDF parametrization $f(x,Q;\mathbf{z})$: (parameters $\mathbf{z}$)

  best-fit: $f^0 = f(x,Q;\mathbf{0})$, error PDFs: $f^{\pm i} = f(x,Q;\pm\mathbf{e}^i)$

- Updated Chi-square function :

$$\Delta\chi^2(\mathbf{z}) = \Delta\chi^2_{\text{old}}(\mathbf{z}) + \left(X^E_\alpha - X_\alpha(\mathbf{z})\right) C^{-1}_{\alpha\beta} \left(X^E_\beta - X_\beta(\mathbf{z})\right)$$

- Hessian approximation :

$$\Delta\chi^2_{\text{old}}(\mathbf{z}) = T^2\mathbf{z}^2 \quad (T = \text{tolerance parameter})$$

$$X_\alpha(\mathbf{z}) = X_\alpha(\mathbf{0}) + \Delta X_\alpha \cdot \mathbf{z} \quad \text{with} \quad \Delta X^i_\alpha = \tfrac{1}{2}\left(X_\alpha(+\mathbf{e}^i) - X_\alpha(-\mathbf{e}^i)\right)$$

- Minimize to find new best fit:

$$\mathbf{z}^0_{\text{new}} = (\mathbf{1} + \mathbf{M})^{-1}\mathbf{A} \quad \text{with} \quad A^i = \tfrac{1}{T^2}\left(X^E_\alpha - X_\alpha(\mathbf{0})\right) C^{-1}_{\alpha\beta}\,\Delta X^i_\beta, \quad M^{ij} = \tfrac{1}{T^2}\Delta X^i_\alpha C^{-1}_{\alpha\beta}\,\Delta X^j_\beta$$

Paukkunen and Zurita, JHEP 1412 (2014) 100

Updated PDF set

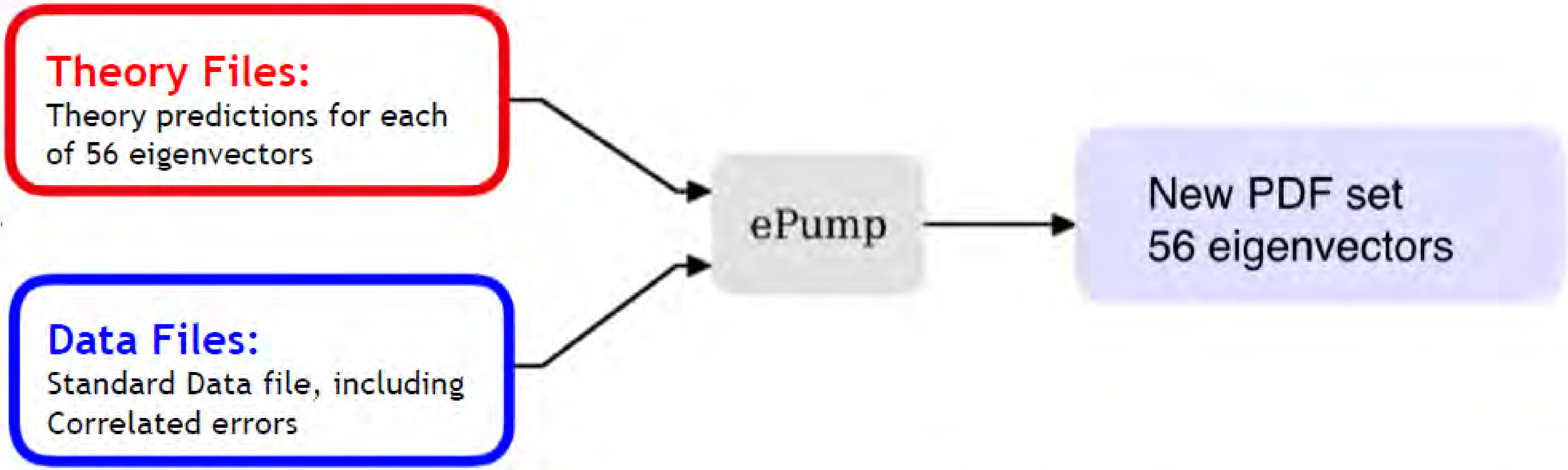
- New best-fit PDF : $f^0_{\text{new}} = f^0 + \Delta f \cdot \mathbf{z}$
- New error PDFs : $f^{\pm(r)} = f^0_{\text{new}} \pm \frac{1}{\sqrt{1+\lambda^{(r)}}} \Delta f \cdot \mathbf{U}^{(r)}$

  where $\lambda^{(r)}$ and $\mathbf{U}^{(r)}$ are the eigenvalues and eigenvectors of matrix $\mathbf{M}$

- Extensions :
  - Best choices for Δf within the linear approximation
  - Dynamical tolerances : $\pm \mathbf{e}^i \Rightarrow \pm \left(T^{\pm i} / T \right) \mathbf{e}^i$
  - Inclusion of diagonal quadratic terms in expansion of $X_\alpha(\mathbf{z})$
  - Direct update of other observables :

$$Y^0_{\text{new}} = Y^0 + \Delta Y \cdot \mathbf{z} \quad , \quad \left|\Delta Y\right| = \Delta Y \cdot (1+\mathbf{M})^{-1} \cdot \Delta Y$$

How to use ePump

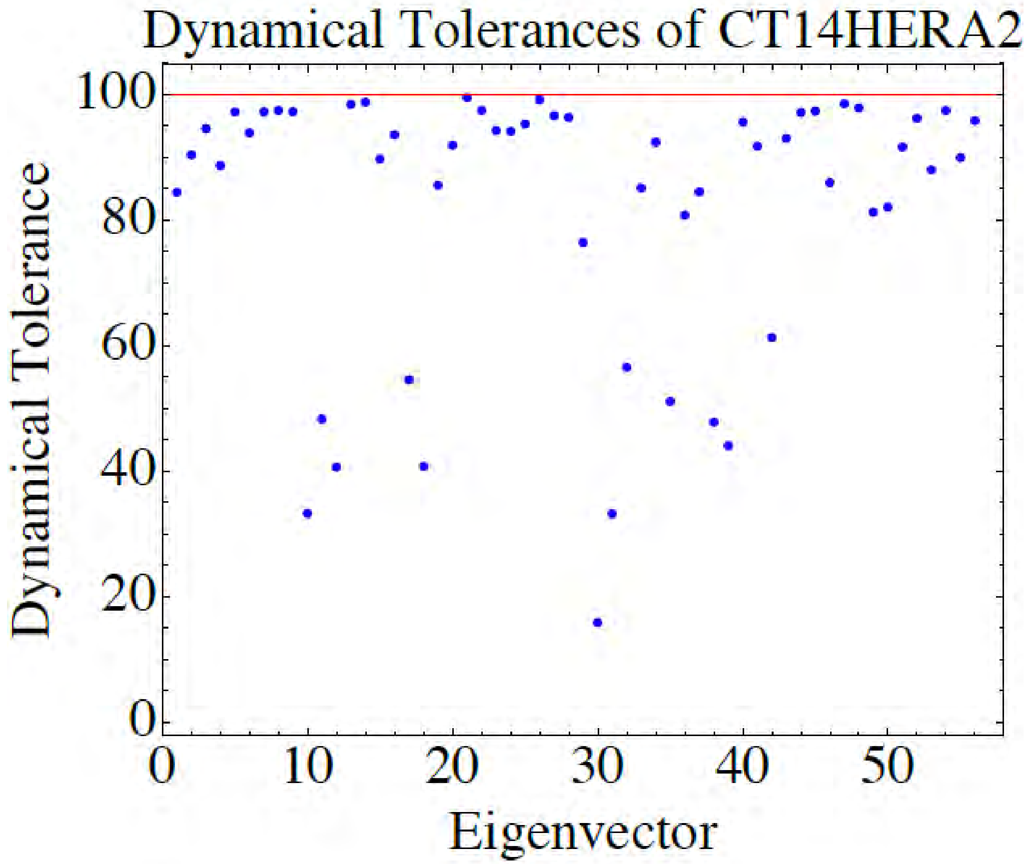


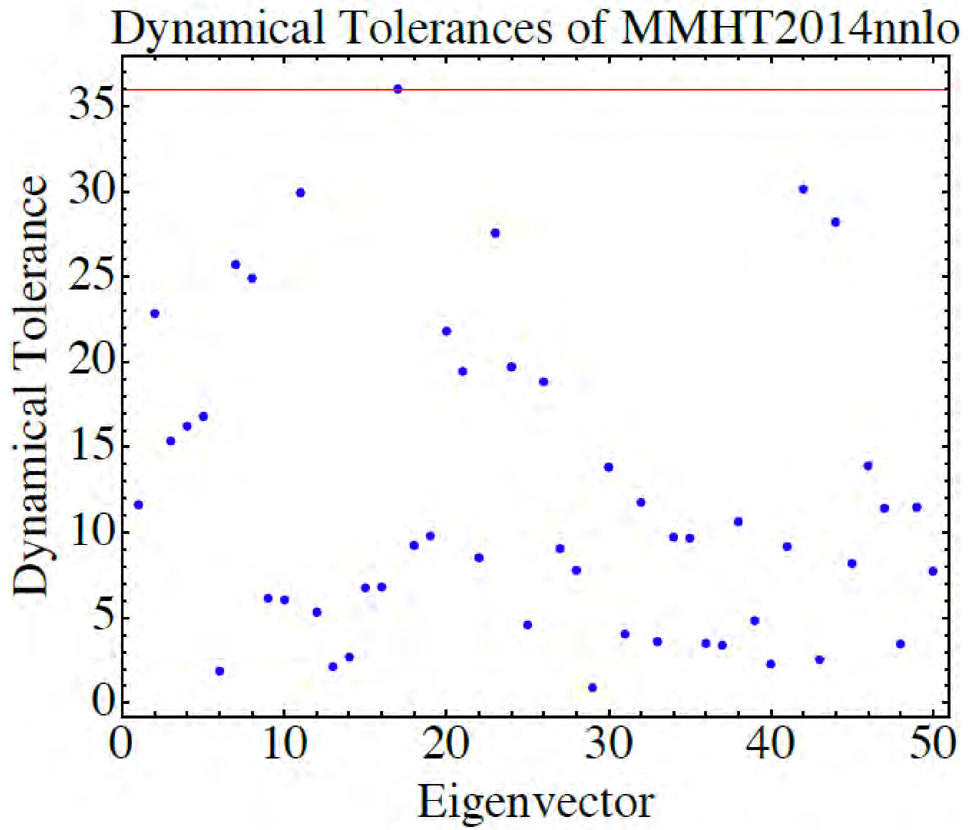
(Auxiliary Theory Files may also be included to update predictions for observables not included in fit.)

Dynamical Tolerance (T^2)

For CT14HERA2, it is given at 90% CL.
To get values at 68% CL, divide by $(1.645)^2=2.7$

For MMHT2104, it is given at 68% CL.
It is different from T=1.





A case study, using CT14HERA2

- Start with a global fit (CT14HERA2mDimu) without including the 4 (NuTeV and CCFR) di-muon data set, which dominate the s-PDF determination.
- Include those 4 di-moun data sets as “new” data, and use ePump to update the above PDF sets, which yields CT14mDimu.
- Using Dynamical Tolerance (dyn.tol) reproduces CT14HERA2 PDFs.
- Using T=1, which is equivalent to give these 4 data sets a weight of about 100 (instead of 1) at 90% CL, leads to a too strong constraint on s-PDF, hence a much smaller PDF error band.

s-PDF in CT14HERA2

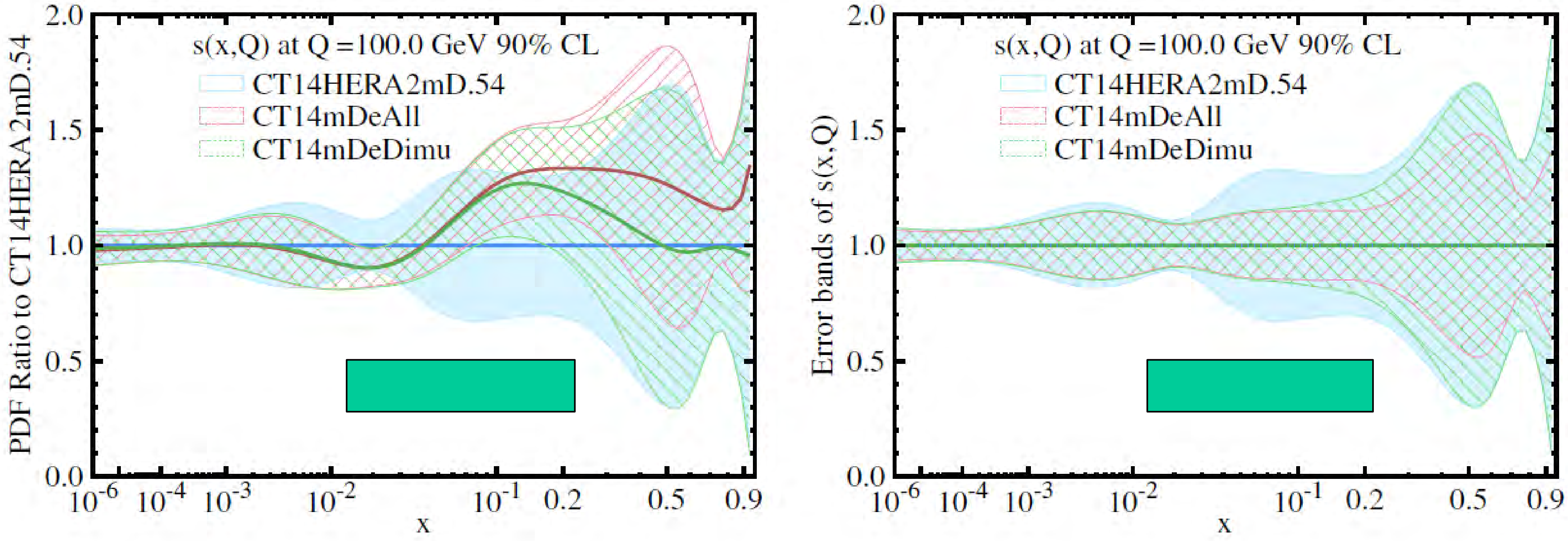
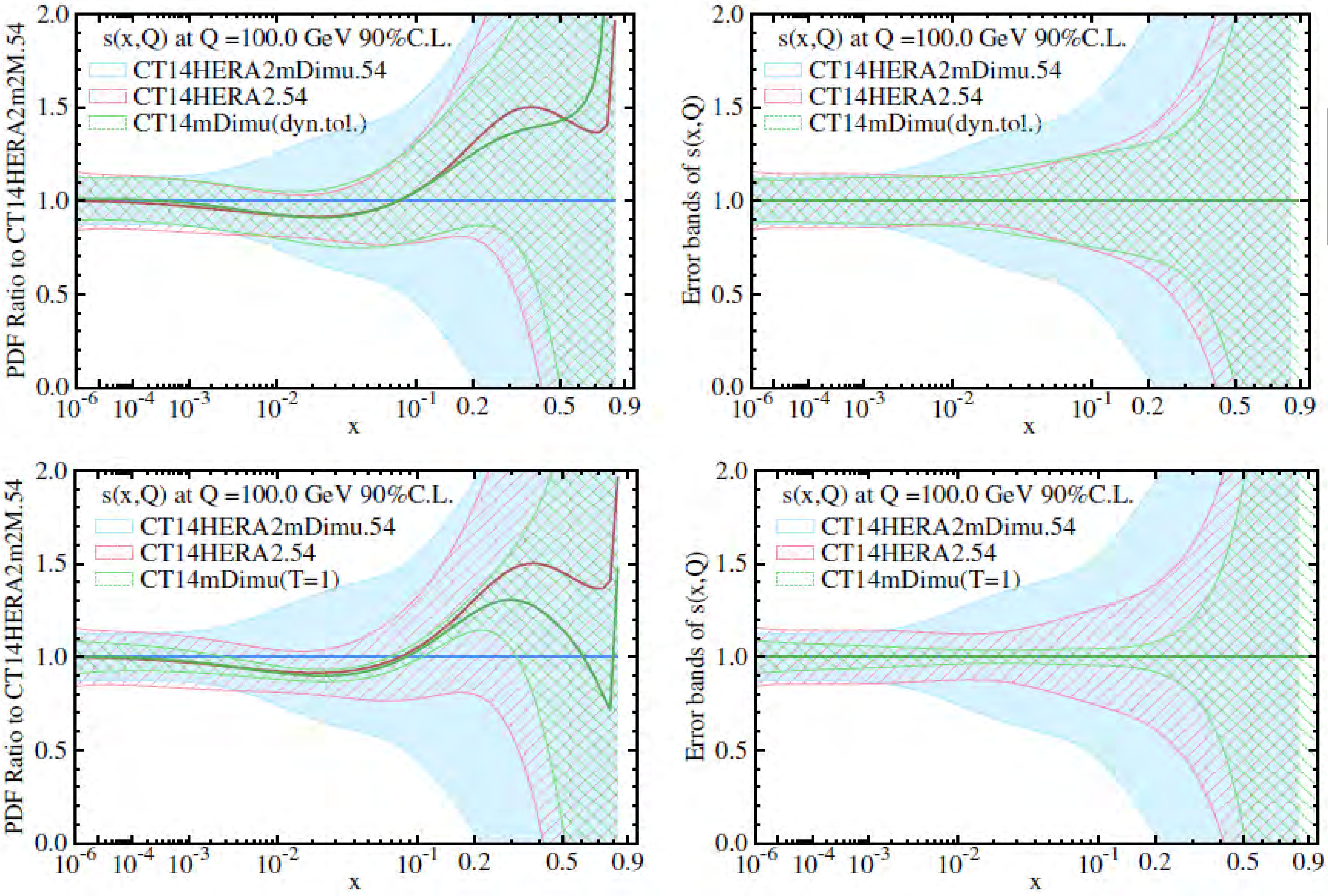


FIG. 16: Comparison of ePump-updated s-PDF, at $Q = 100$ GeV. CT14mDeDimu is obtained by adding only the DIS charged current dimuon data (NuTeV [18], and CCFR [19]) to CT14HERA2mD with ePump.

Using Dynamical Tolerance vs. T=1



Using
Dynamical Tolerance

Using T=1,
equivalent to xFitter approach.
It leads to a much smaller PDF error band.

Summary: I

- A new CT18 PDF analysis is ready released to public.
- The CT18 PDF uncertainty is mildly reduced at NNLO compared to the CT14 PDF uncertainty.
- 700+ data points from 12 new LHC data sets. The LHC constraints on the CT18 PDFs are weaken by some inconsistencies between the LHC data sets and the pre-LHC data sets.
- HERA DIS and fixed-target experiments deliver key constraints on CT18 PDFs.
- We observe some impact on PDFs from ATLAS and CMS incl. jet data, ATLAS, CMS, LHCb W/Z data and ATLAS 8 TeV Z pT data. LHC top quark pair data provides a similar impact to g-PDF as incl. jet data, but cannot reduce g-PDF errors as strong as incl. jet data due to its much smaller number of data points.
- ATLAS 7 TeV W and Z rapidity data is included in the CT18Z PDF analysis, in which NNLO DIS cross sections with an x-dependent factorization scale, behaving like NNLO+NNLx resummed ones, are incorporated in CT18Z PDFs.

Summary: II

- PDF profiling has to use the consistent Tolerance criteria given by a specific PDF global analysis.
- When using CT PDFs, one should use dynamical tolerance. Using T=1 is equivalent to giving the pseudo data set a weight of close to 37 (instead of 1) at the 68% CL.
- T=1 cannot be used to CT, MMHT PDFs, neither LHC4PDF15 set. Otherwise, it will overstate the impact of the pseudo data set to constrain PDFs.

Backup slides

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

reduced χ^2 values

χ^2 = [reduced χ^2] + R^2

The quadratic penalty for 162 systematic errors = 87.5

e^+p data are fitted fine

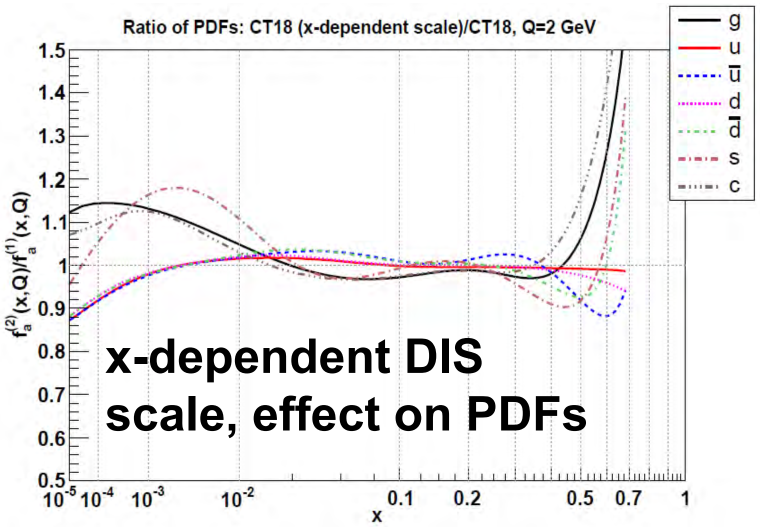
e^-p data are fitted poorly

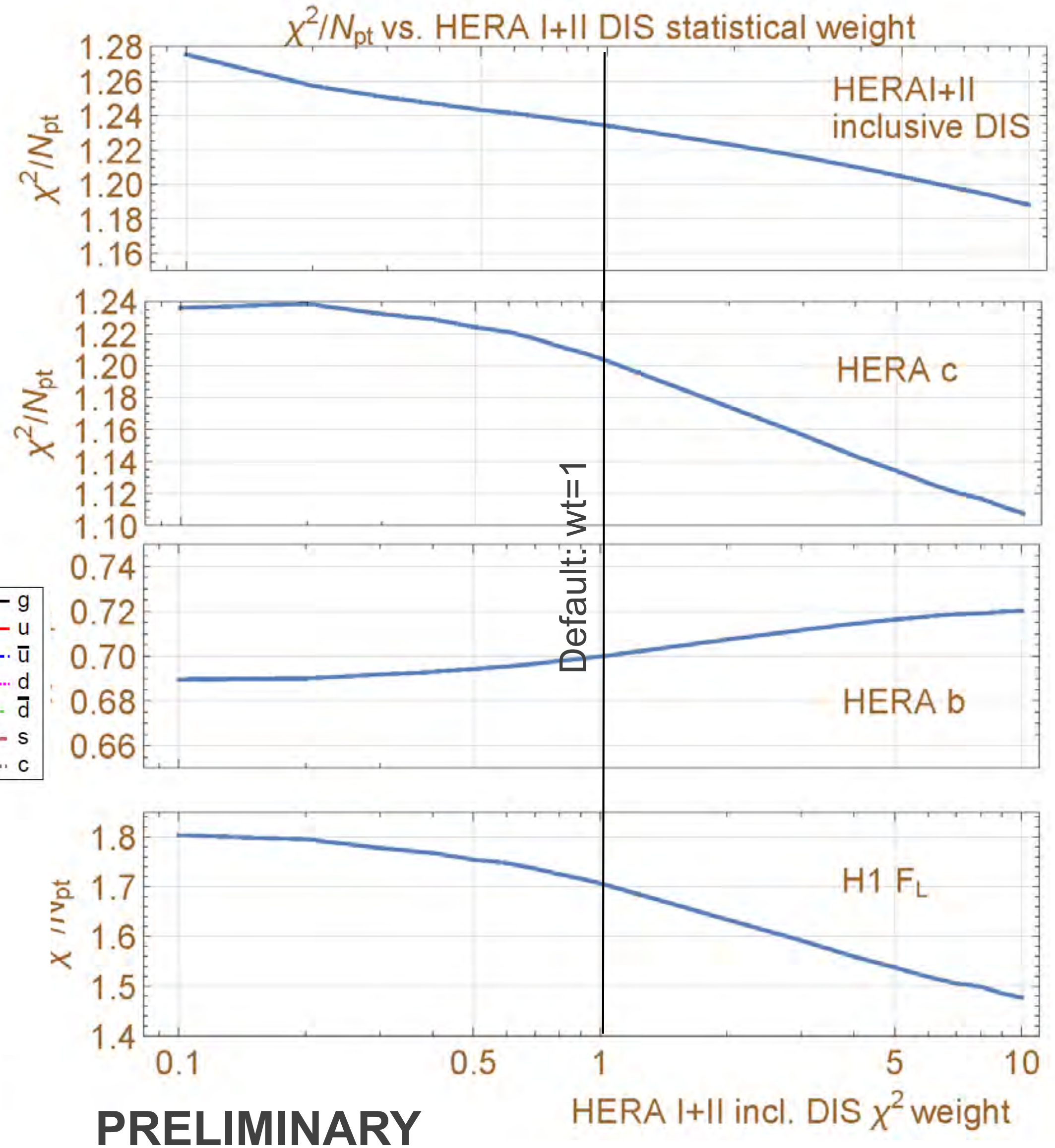
Fair (not perfect) agreement; can be mildly improved by the QCD scale choice

CT18X and Z: a special factorization scale in DIS

The CT18Z fits uses a $\mu_{DIS,X}$ scale that reproduces many features of NNLO-NLLx fits with $\ln(1/x)$ resummation by the NNPDF [arXiv:1710.05935] and xFitter [1802.0064] groups.

$$\mu^2_{DIS,X} = 0.8^2\left(Q^2 + \frac{0.3\ GeV^2}{x^{0.3}}\right)$$





PRELIMINARY

CT14: parametrization forms

- CT14 relaxes restrictions on several PDF combinations that were enforced in CT10. [These combinations were not constrained by the pre-LHC data.]
  - The assumptions $\frac{\bar{d}(x,Q_0)}{\bar{u}(x,Q_0)} \to 1$, $u_v(x,Q_0) \sim d_v(x,Q_0) \propto x^{A_{1v}}$ with $A_{1v} \approx -\frac{1}{2}$ at $x < 10^{-3}$ are relaxed once LHC W/Z data are included
  - CT14 parametrization for $s(x,Q)$ includes extra parameters
- Candidate CT14 fits have 30-35 free parameters
- In general, $\mathrm{f_a}(\mathrm{x}, \mathrm{Q_0}) = \mathrm{A}x^{a_1}(1-\mathrm{x})^{\mathrm{a_2}}\mathrm{P_a}(\mathrm{x})$
- CT10 assumed $P_a(x) = \exp(a_0 + a_3\sqrt{x} + a_4 x + a_5\, x^2)$
  - exponential form conveniently enforces positive definite behavior
  - but power law behaviors from a_1 and a_2 may not dominate
- In CT14, $\mathrm{P_a}(\mathrm{x}) = \mathrm{G_a}(\mathrm{x})\mathrm{F_a}(\mathrm{z})$, where $\mathrm{G_a}(\mathrm{x})$ is a smooth factor
  - $\mathrm{z} = 1 - 1(1-\sqrt{\mathrm{x}})^{\mathrm{a_3}}$ preserves desired Regge-like behavior at low x and high x (with $\mathrm{a_3}$>0)
- Express $F_a(z)$ as a linear combination of Bernstein polynomials:

$$z^4, 4z^3(1-z), 6z^2(1-z)^2, 4z(1-z)^3, (1-z)^4$$

  - each basis polynomial has a single peak, with peaks at different values of z; reduces correlations among parameters