PDF updates and PDF4LHC Benchmarking and Combination.

Robert Thorne

December 3rd 2021

University College London

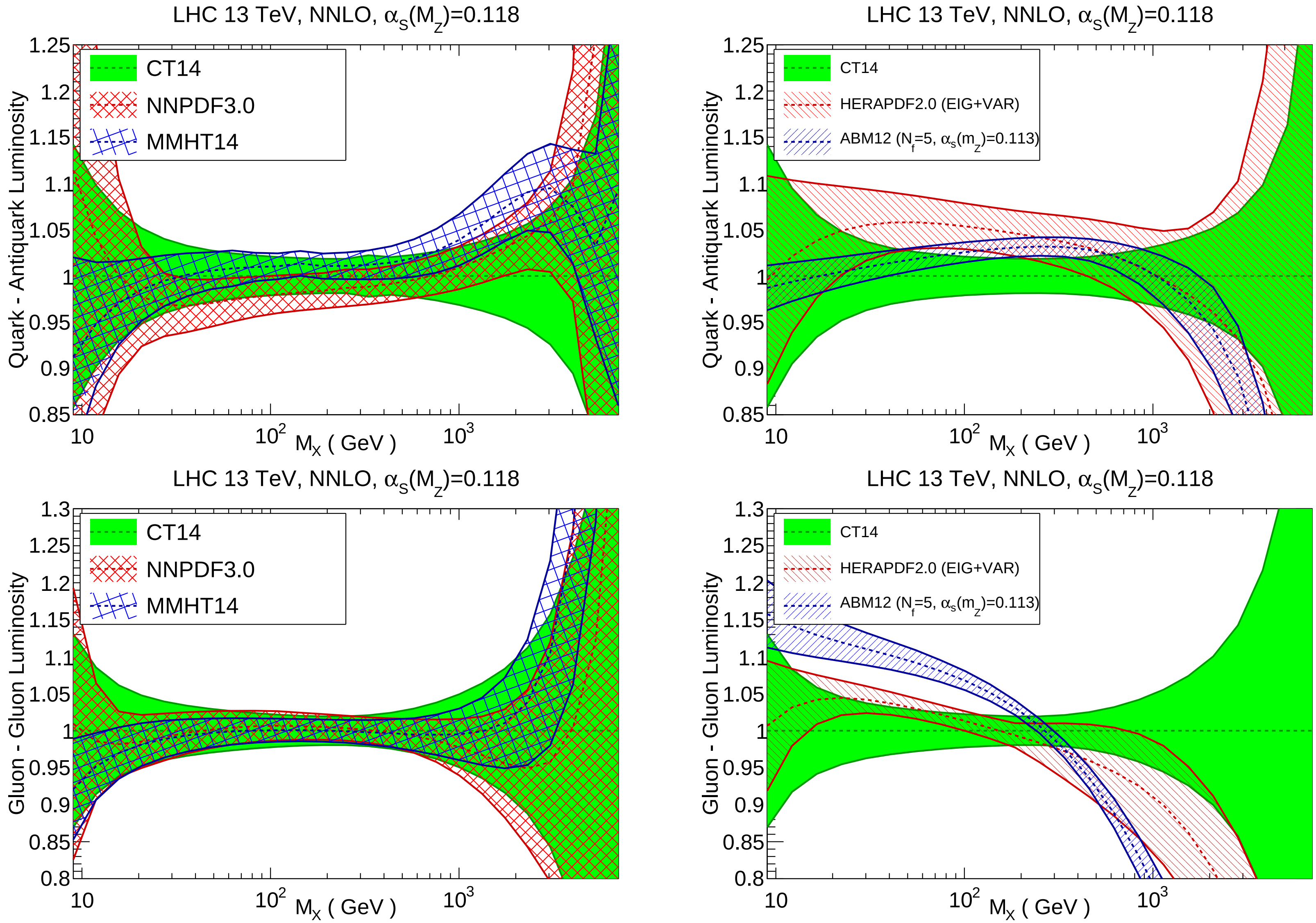
Updates to major PDF sets due to inclusion of new, largely LHC (different amounts and choices for different groups) but also some HERA, Tevatron, data sets. Also significant new procedural changes in each group.

I will discuss PDF benchmarking, outlining a completed exercise by the PDF4LHC Working Group.

Has led to a new PDF4LHC21 combination. Will report ongoing progress.

All discussions based on most recent sets, NNPDF3.1, CT18, MSHT2020, along with ABMP16 and HERAPDF2.0. Brief mention of NNPDF4.0 which postdates the benchmarking and combination work.

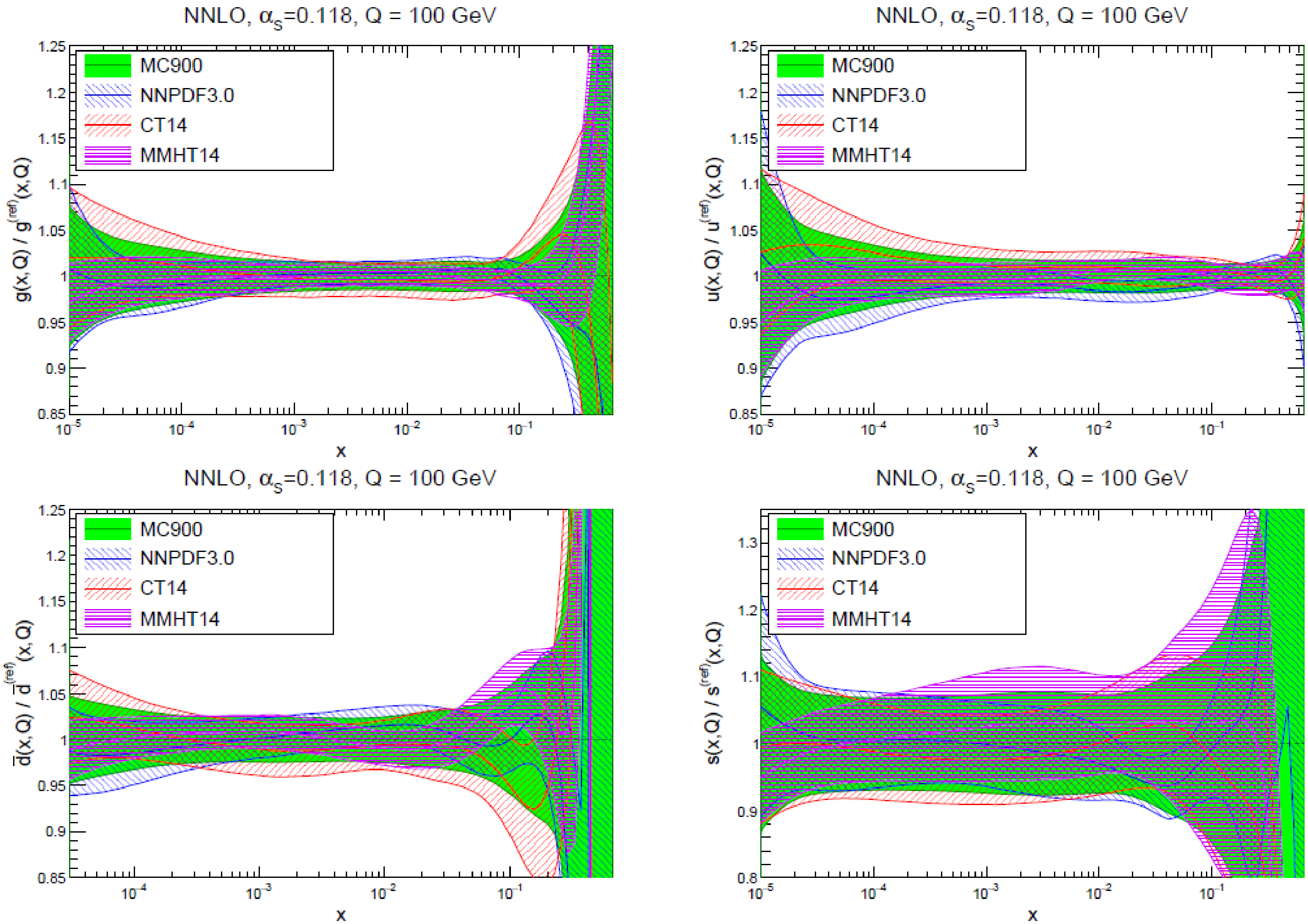
Comparison of state-of-the-art PDFs (late 2015)



Some good agreement between CT14, MMHT2014 and NNPDF3.0.

Some differences in some PDF sets in central values and uncertainty.

Comparison of Combination of CT, MMHT, NNPDF using "Monte Carlo" sets to the Individual PDFs



Works well if PDFs are fairly compatible.

The PDF4LHC Prescription

Perform a Monte Carlo combination of the included PDF sets.

Sets entering into the combination must satisfy requirements, i.e. be compatible for combination. $\alpha_S(M_Z^2) = 0.118$

Deliver a single combined PDF set - either Monte Carlo or Hessian form for combined PDF.

– Monte Carlo - A set of PDF replicas is delivered. The mean is the central value and the standard deviation the uncertainty.

– Hessian - A central set and eigenvectors representing orthogonal sources of uncertainty are delivered. Uncertainty obtained by summing each uncertainty source in quadrature.

Followed by numerous updates.

The fit ingredients

DATA:

- DIS NC/CC inclusive (HERA I+II added, no deuteron data included)
- DIS NC charm production (HERA)
- DIS CC charm production (HERA, NOMAD, CHORUS, NuTeV/CCFR)
- fixed-target DY
- LHC DY distributions (ATLAS, CMS, LHCb)
- t-quark data from the LHC and Tevatron
- deuteron data are excluded

QCD:

- NNLO evolution
- NNLO massless DIS and DY coefficient functions
- NLO+ massive DIS coefficient functions (**FFN scheme**)
  - – NLO + NNLO(approx.) corrections for NC
  - – NNLO CC at Q>> m_c
  - – running mass
- NNLO exclusive DY (FEWZ 3.1)
- NNLO inclusive ttbar production (pole / running mass)
- Relaxed form of (dbar-ubar) at small x

Power corrections in DIS:

- target mass effects
- dynamical twist-4 terms

Alekhin PDF4LHC 2017

NNPDF3.1 released in 2017.

New datasets in NNPDF3.1

Measurement	Data taking	Motivation
Combined HERA inclusive data	Run I+II	quark singlet and gluon
D0 legacy W asymmetries	Run II	quark flavor separation
ATLAS inclusive W, Z rap 7 TeV	2011	strangeness
ATLAS inclusive jets 7 TeV	2011	large-x gluon
ATLAS low-mass Drell-Yan 7 TeV	2010+2011	small-x quarks
ATLAS Z pT 7,8 TeV	2011+2012	medium-x gluon and quarks
ATLAS and CMS tt differential 8 TeV	2012	large-x gluon
CMS Z (pT,y) 2D xsecs 8 TeV	2012	medium-x gluon and quarks
CMS Drell-Yan low+high mass 8 TeV	2012	small-x and large-x quarks
CMS W asymmetry 8 TeV	2012	quark flavor separation
CMS 2.76 TeV jets	2012	medium and large-x gluon
LHCb W,Z rapidity dists 7 TeV	2011	large-x quarks
LHCb W,Z rapidity dists 8 TeV	2012	large-x quarks

4 Juan Rojo DIS2017, Birmingham, 04/04/2017

Rojo DIS2017

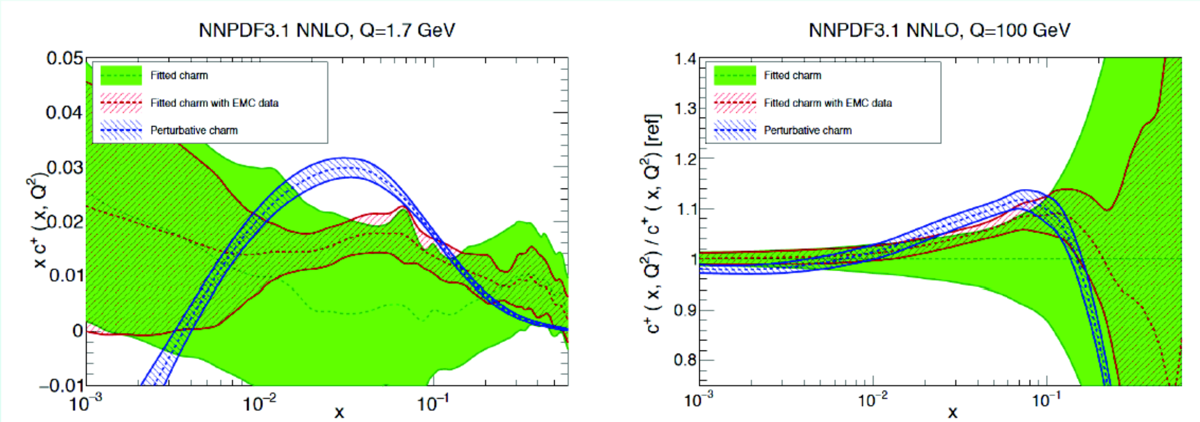
Theory developments

Charm content of proton revisited

- The new LHC experiments provide additional constraints on **non-perturbative charm**
- Including the EMC charm data, we find **evidence for non-perturbative charm at the 1.5 sigma level.** Even without EMC data, **non-perturbative charm bounded < 1.0 % at the 90% CL**

$$[C(Q^2)] \equiv \int_0^1 dx\ \left(xc(x,Q^2) + x\bar{c}(x,Q^2)\right)$$

PDF set	$C(Q = 1.65\ \mathrm{GeV})$	$C(Q = 100\ \mathrm{GeV})$
Perturbative charm	$(0.360 \pm 0.007)\%$	$(3.77 \pm 0.02)\%$
Fitted charm	$(0.45 \pm 0.40)\%$	$(3.8 \pm 0.2)\%$
Fitted charm with EMC data	$(0.52 \pm 0.14)\%$	$(3.86 \pm 0.08)\%$

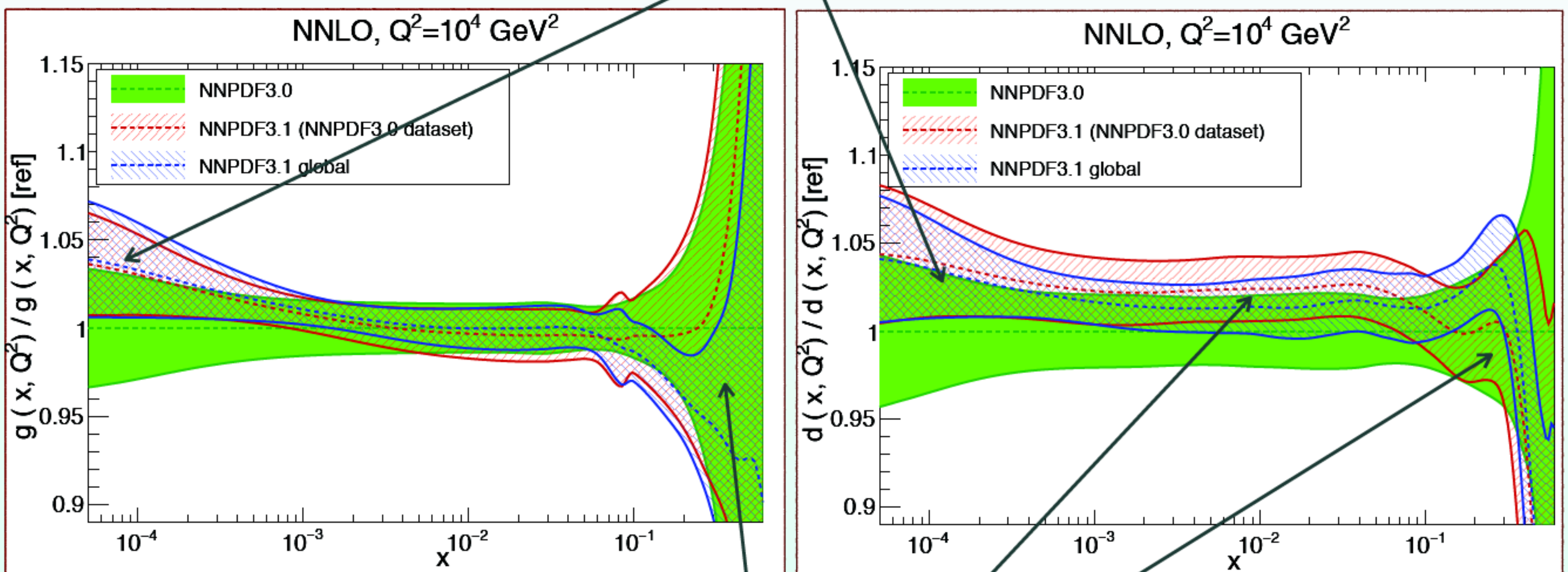


LHC W, Z data prefer lower charm for $0.01 < x < 0.1$.

Rojo, DIS 2017

new data vs new methodology

new methodology (mostly fitting charm)

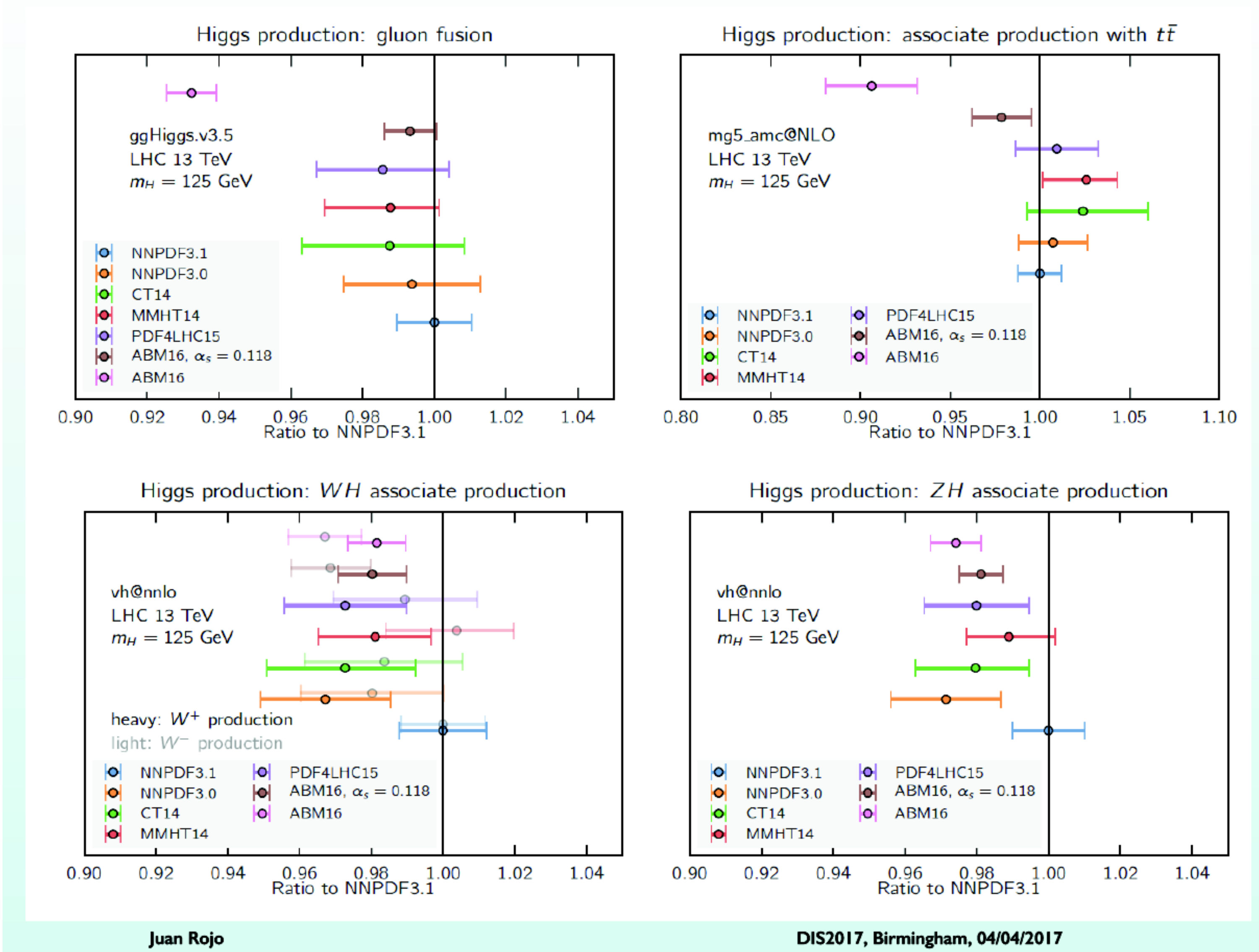


Impact of new data

10

Juan Rojo DIS2017, Birmingham, 04/04/2017

Higgs production cross-sections



Some impact on cross sections.

ATLAS fits to **HERA**, $W, Z/\gamma^*$ and differential $\bar{t}t$ data.

Largest published update to HERA fit.

Table 2: Total and partial χ^2 for data sets entering the PDF fit for simultaneous fits to the p_T^t and y_t spectra, and p_T^t and m_{tt}, with and without using the statistical correlations between the spectra

		lepton+jets spectra			
		p_T^t and y_t with statistical correlations	p_T^t and y_t without statistical correlations	p_T^t and m_{tt} with statistical correlations	p_T^t and m_{tt} without statistical correlations
Total χ^2/NDF		1264 / 1068	1260 / 1068	1290 / 1070	1287 / 1070
Partial χ^2/NDP	HERA	1148 / 1016	1147 / 1016	1162 / 1016	1162 / 1016
Partial χ^2/NDP	ATLAS $W, Z/\gamma^*$	82.7 / 55	83.5 / 55	83.2 / 55	83.1 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	33 / 13	30 / 13	45 / 15	42 / 15

Fits precision $W, Z/\gamma^*$ data and single-differential $\bar{t}t$ data.

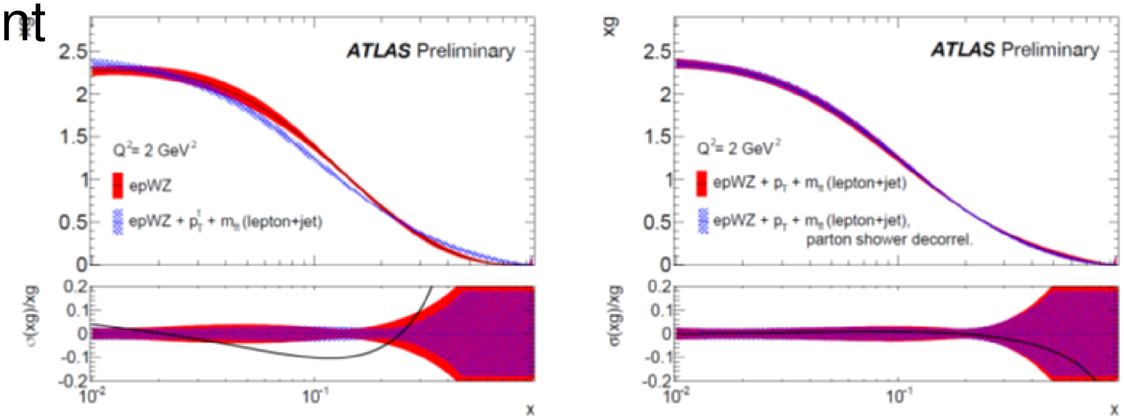
Difficult to simultaneously fit distributions (seen by other groups).

Table 4: Total and partial χ^2 for data sets entering the PDF fit in simultaneous fits to the p_T^t and y_t spectra, and the p_T^t and m_{tt} spectra, with the 2-point uncertainties; ISR/FSR, hard scattering model and parton shower model, decorrelated between the spectra. For the fit to the p_T^t and m_{tt} spectra, the result of decorrelating only the parton shower model uncertainty is also shown.

		lepton+jets spectra		
		p_T^t and y_t decorrelate 2-point uncertainties	p_T^t and m_{tt} decorrelate 2-point uncertainties	p_T^t and m_{tt} decorrelate parton-shower model uncertainty
Total χ^2/NDF		1259 / 1068	1247 / 1070	1248 / 1070
Partial χ^2/NDP	HERA	1147 / 1016	1154 / 1016	1153 / 1016
Partial χ^2/NDP	ATLAS $W, Z/\gamma^*$	83.9 / 55	81.9 / 55	81.6 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	27.8 / 13	11.5 / 15	14.1 / 15

Helped by imposing decorrelation between Monte Carlo dependent error sources.

Impact on high-x gluon.



(a) (b)

CT18 PDFs.

Major update from the CT group.

ID#	Experimental data set		$N_{pt,E}$	χ^2_E	$\chi^2_E/N_{pt,E}$	S_E
245	LHCb 7 TeV 1.0 fb^{-1} W/Z forward rapidity cross sec.	[87]	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.	[88]	17	17.7 (18.0)	1.0 (1.1)	0.2 (0.3)
248‡	ATLAS 7 TeV 4.6 fb^{-1}, W/Z combined cross sec.	[38]	34	287.3 (88.7)	8.4 (2.6)	13.7 (4.8)
249	CMS 8 TeV 18.8 fb^{-1} muon charge asymmetry A_{ch}	[89]	11	11.4 (12.1)	1.0 (1.1)	0.2 (0.4)
250	LHCb 8 TeV 2.0fb^{-1} W/Z cross sec.	[90]	34	73.7 (59.4)	2.1 (1.7)	3.7 (2.6)
253	ATLAS 8 TeV 20.3 fb^{-1}, Z p_T cross sec.	[91]	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)	[92]	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., $R = 0.6$	[9]	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)	[93]	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 and y cross sec.	[94]	16	18.9 (19.1)	1.2 (1.2)	0.6 (0.6)
580	ATLAS 8 TeV 20.3 fb^{-1}, $t\bar{t}$ p^t_T and $m_{t\bar{t}}$ abs. spectrum	[95]	15	9.4 (10.7)	0.6 (0.7)	-1.1 (-0.8)

TABLE II: Like Table I, for newly-included LHC measurements. The ATLAS 7 TeV W/Z data (4.6 fb^{-1}), labeled by ‡, are included in the CT18A and CT18Z global fits, but not in CT18 and CT18X.

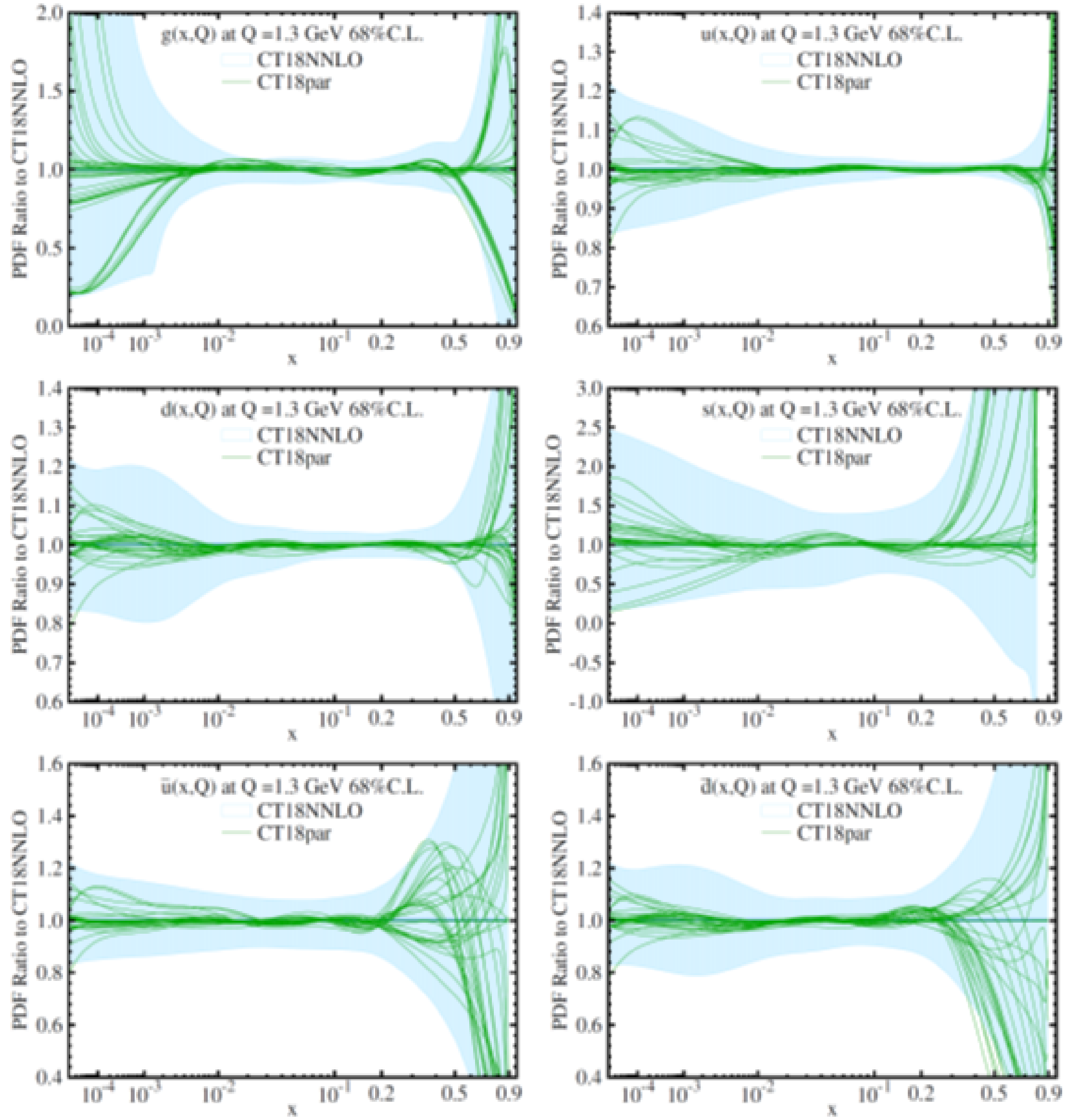
First inclusion of a wide variety of precision LHC data.

Some sets omitted due to tensions (notably ATLAS precision 7 TeV $W, Z/\gamma^*$ data) or lack of impact on the fit.

Sets CT18A, CT18X, CT18Z include pecision ATLAS data, small-x modification or both.

Use 26 parameters, but use various different choices and investigate PDF changes.

Require consistency with χ^2 tolerance chosen.



Generally consistency with CT14HERA2.

More changes when compared to CT18Z
- includes ATLAS precision 7 TeV $W, Z/\gamma^*$ and small-x changes.

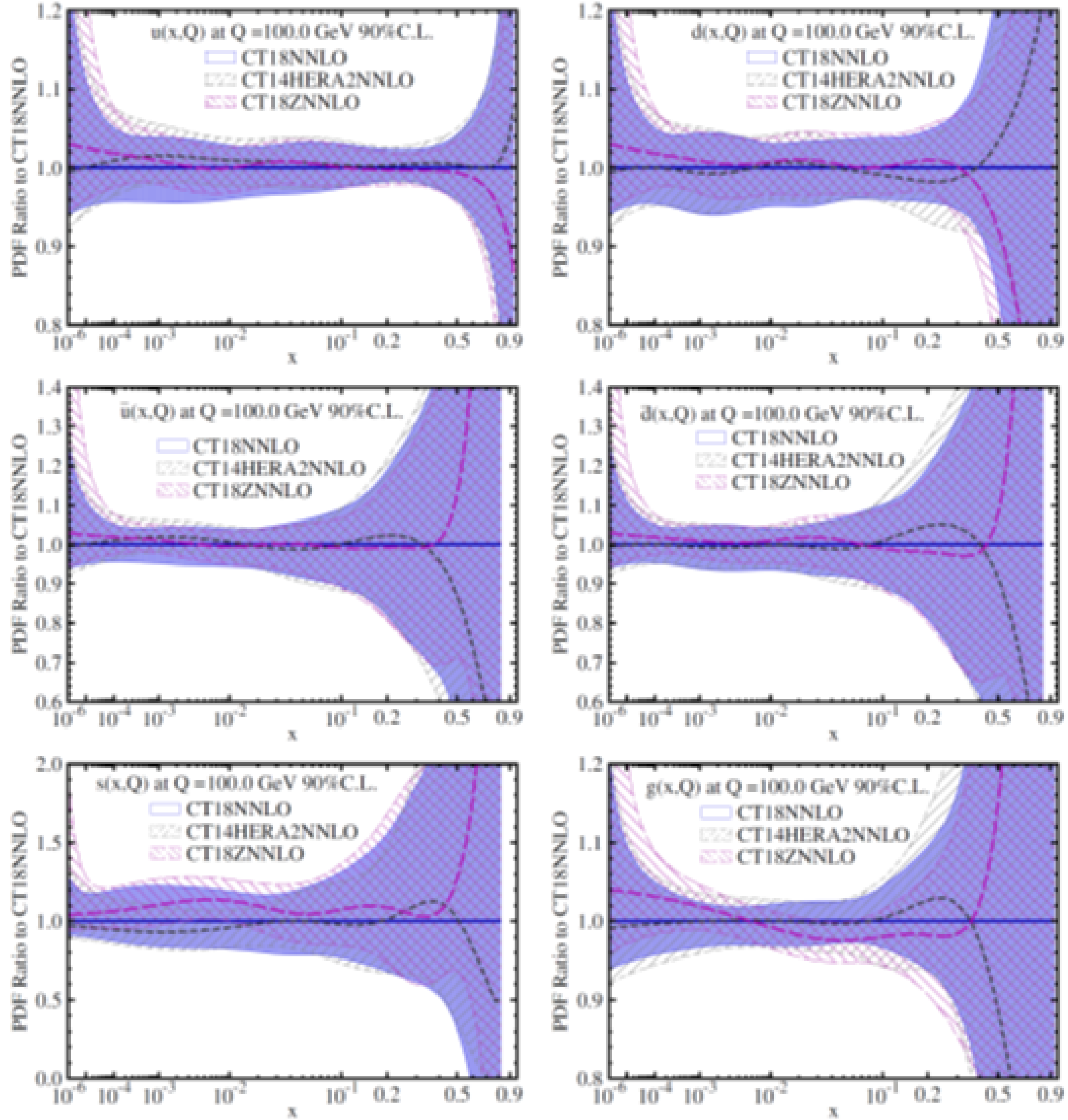
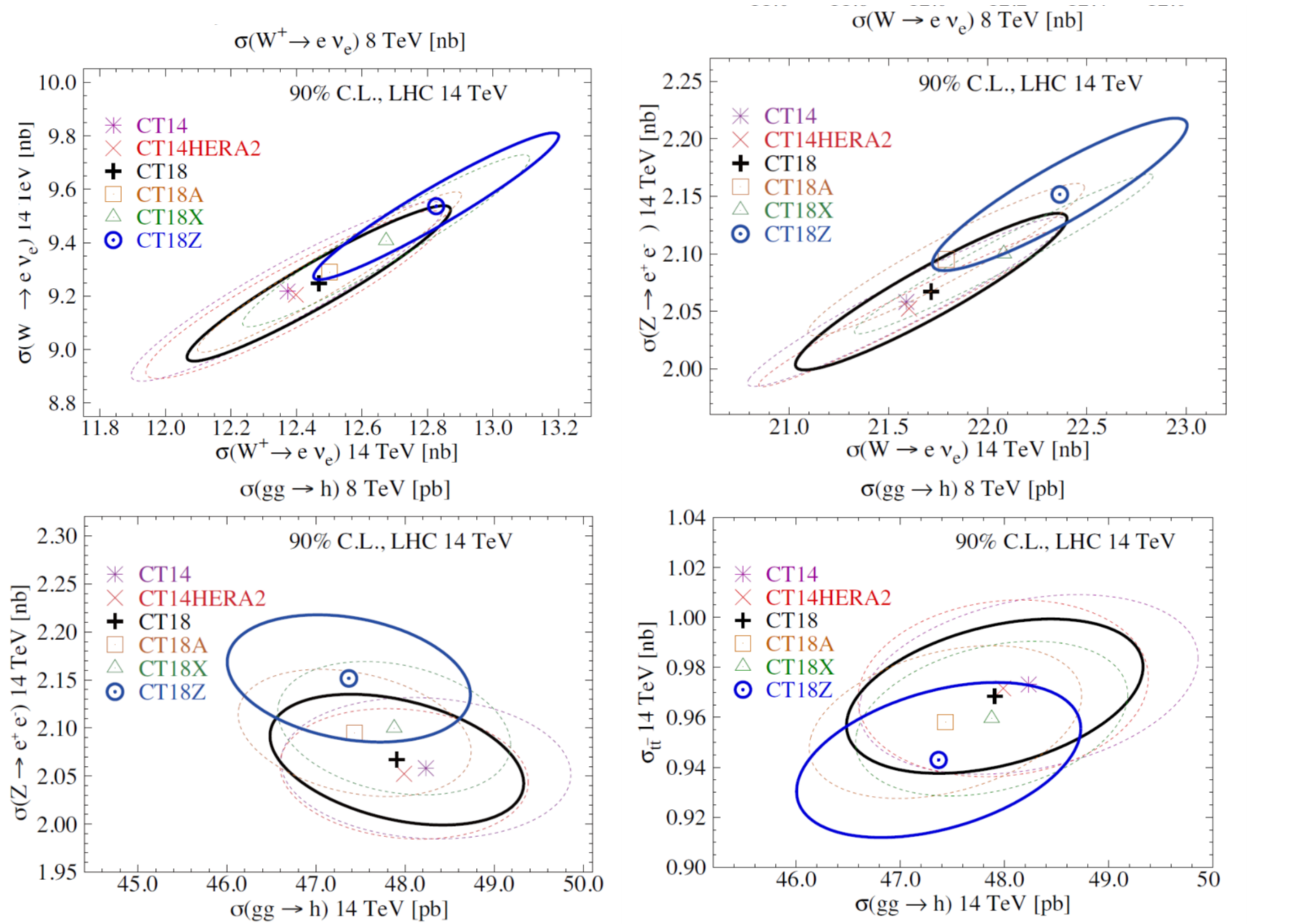
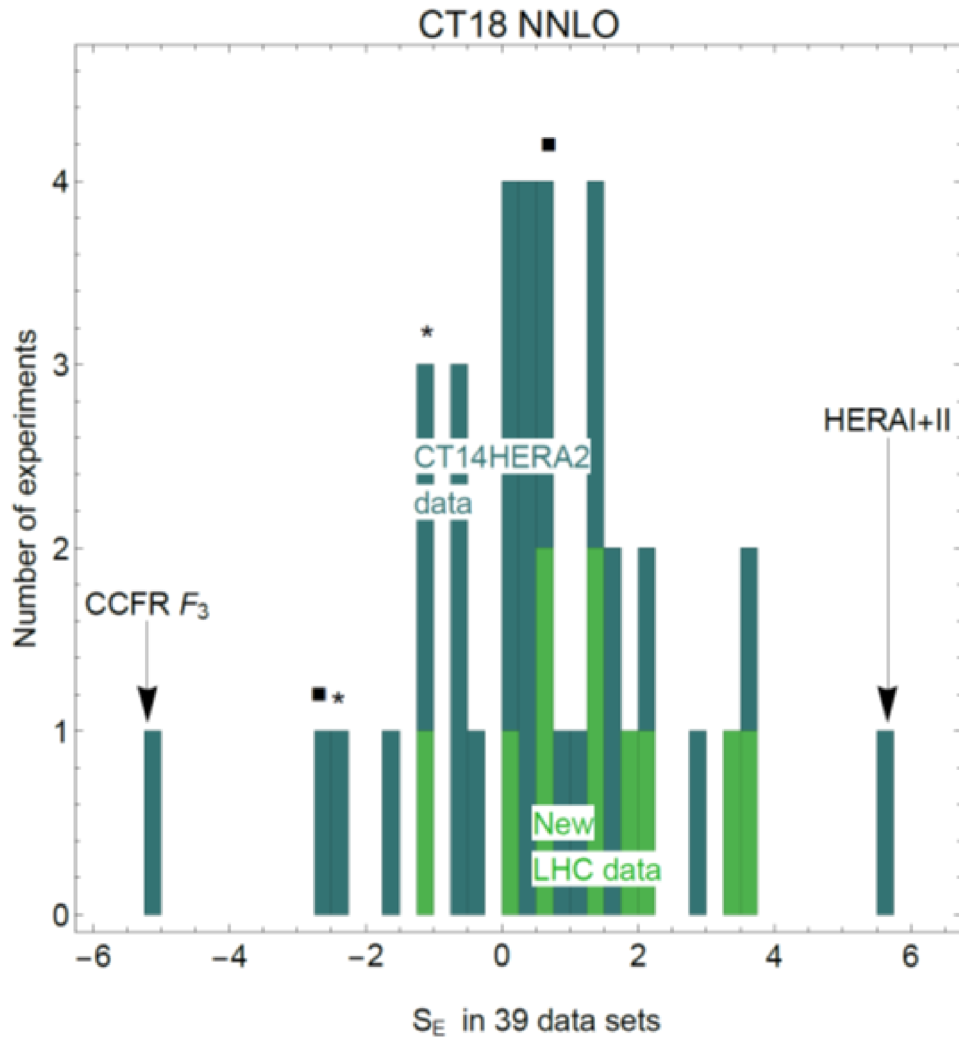


FIG. 7: A comparison of 90% C.L. PDF uncertainties from CT18 (violet solid), CT14$_{HERAII}$(gray short-dashed), and CT18Z (magenta long-dashed) NNLO ensembles at $Q = 100$ GeV. The uncertainty bands are normalized to the central CT18 NNLO PDFs.

Changes in some Cross sections.



Investigation of fit quality.



More variation than expected under standard statistical rules.

MSHT PDFs - major update.

Theoretical Procedures

As before use a general mass variable flavour scheme based on the Thorne-Roberts scheme, using the "optimal" choice of parameters for smoothness near threshold.

Use deuteron and heavy nuclear corrections. Former fit using 4 parameter model, as in MMHT14 and latter use same corrections (arXiv:1112.6324) as MMHT14 with additional penalty-free freedom of order 1%.

Fit data with systematics uncertainties using either nuisance parameters if possible (preferred method) or with the correlation matrix provided. Now also use statistical correlations whenever provided.

Fit to absolute cross sections in preference to normalized if both available.

Extension of parameterisation.

General parameterisation used $A(1-x)^\eta x^\delta(1+\sum_{i=1}^n a_i T_i(1-2x^{\frac{1}{2}}))$, where $T_i(1-2x^{\frac{1}{2}})$ are Chebyshev polynomials.

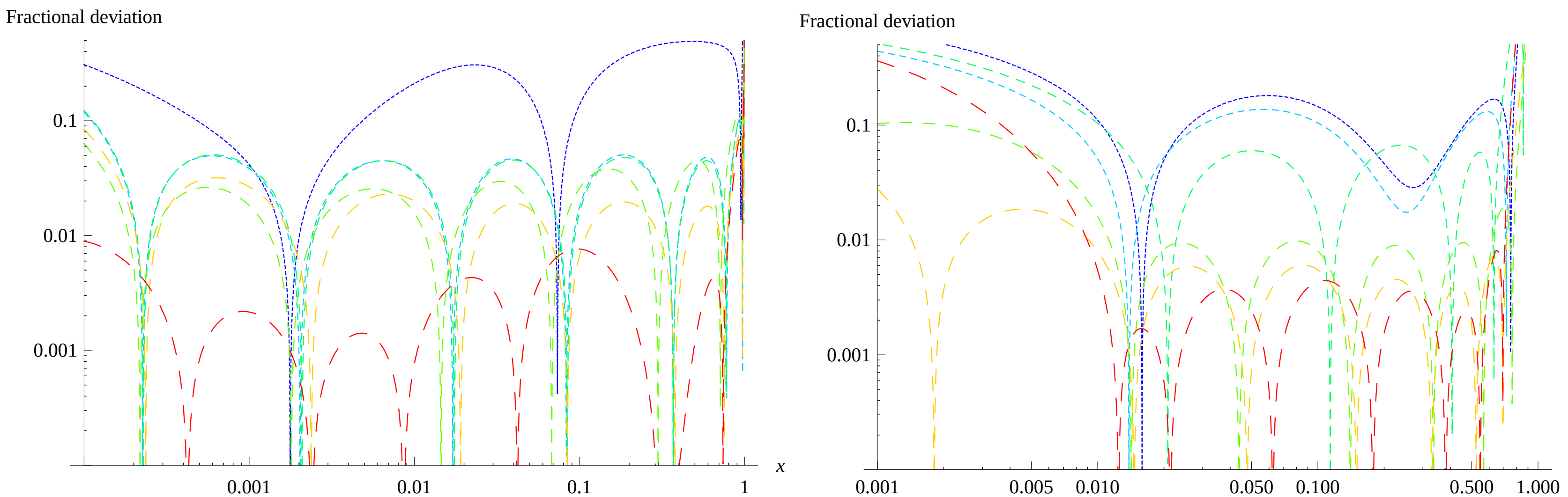


Illustration of precision possible with increasing n, sea-like (left) and valence-like (right) (where pseudo-data for $x > 0.01$). Using $n = 6$ would lead to much better than 1% precision.

For most PDFs $n = 4$ default for MMHT2014 – 36 parameters.

Now extend to $n = 6$ – total of 51 parton parameters.

When determining uncertainties go from 25 eigenvector pairs to 32.

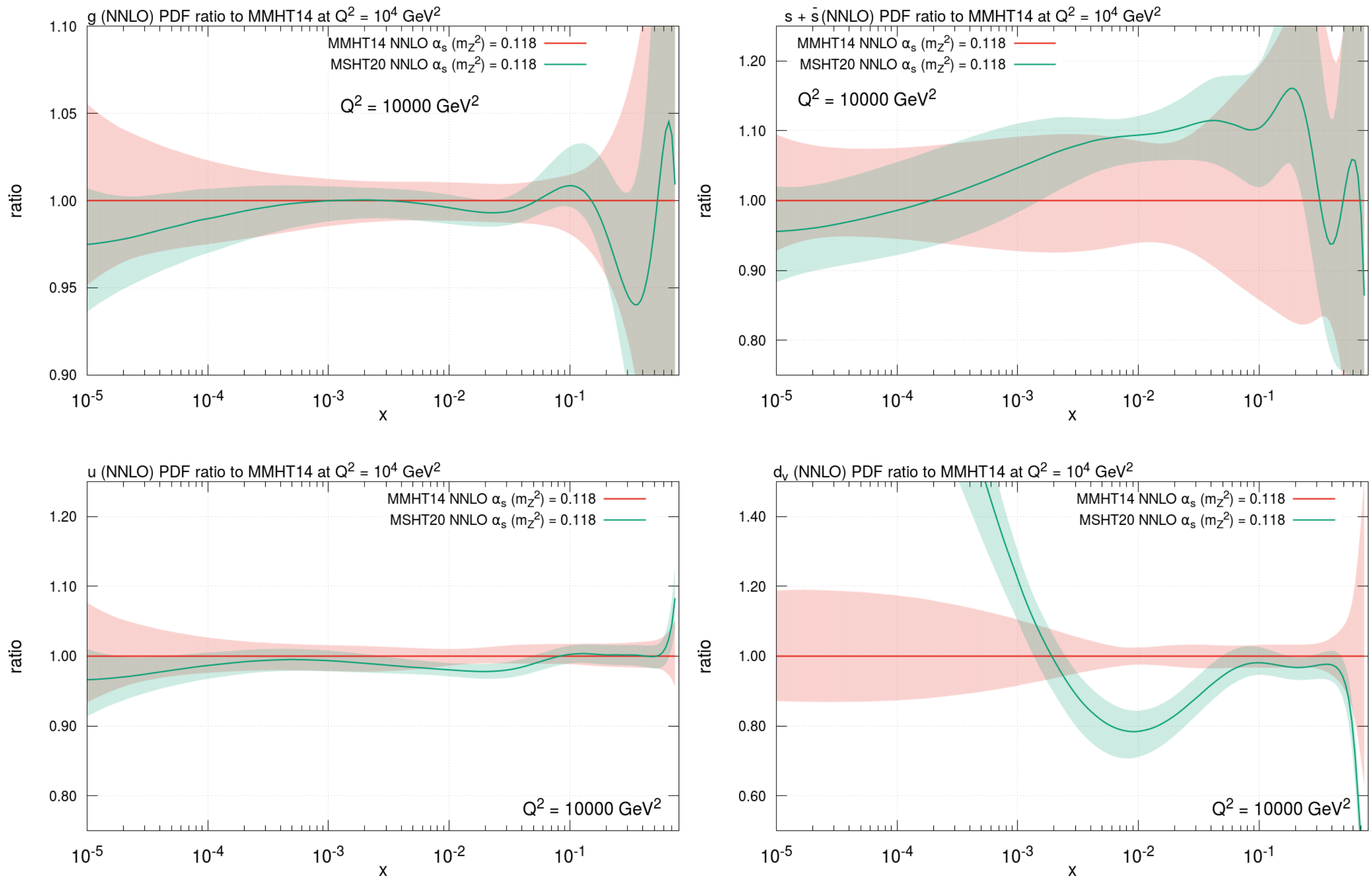
New LHC data fit.

Include all our recent LHC data updates in the fit at NNLO (for default $\alpha_S(M_Z^2) = 0.118$).

	no. points	NNLO χ^2
D0 W asymmetry	14	12.0
$\sigma_{t\bar{t}}$ Tevatron +CMS+ATLAS $7, 8\mathrm{TeV}$	17	14.5
LHCb 7+8 TeV $W + Z$	67	99.4
LHCb 8 TeV e	17	26.2
CMS 8 TeV W	22	12.7
ATLAS 7 TeV jets $R = 0.6$	140	221.6
CMS 7 TeV $W + c$	10	8.6
ATLAS 7 TeV W, Z	61	116.6
CMS 7 TeV jets $R = 0.7$	158	175.8
ATLAS 8 TeV $Z p_T$	104	188.5
CMS 8 TeV jets	174	261.3
ATLAS 8 TeV $t\bar{t} \rightarrow l + \mathrm{j}$ single-diff	25	25.6
ATLAS 8 TeV $t\bar{t} \rightarrow l^+ l^-$ single-diff	5	3.5
ATLAS 8 TeV high-mass Drell-Yan	48	56.7
ATLAS 8 TeV $W^{+,-} + \mathrm{jet}$	32	18.1
CMS 8 TeV $(d\sigma_{t\bar{t}}/dp_{T,t} dy_t)/\sigma_{t\bar{t}}$	15	22.5
ATLAS 8 TeV W^+, W^-	22	57.4
CMS 2.76 TeV jets	81	102.9
CMS 8 TeV $t\bar{t}\ y_t$ distribution	9	13.2
ATLAS 8 TeV double differential Z	59	85.6
total	4363	5122

Fit quality generally good. Relatively poor χ^2 values for some sets all observed by other groups.

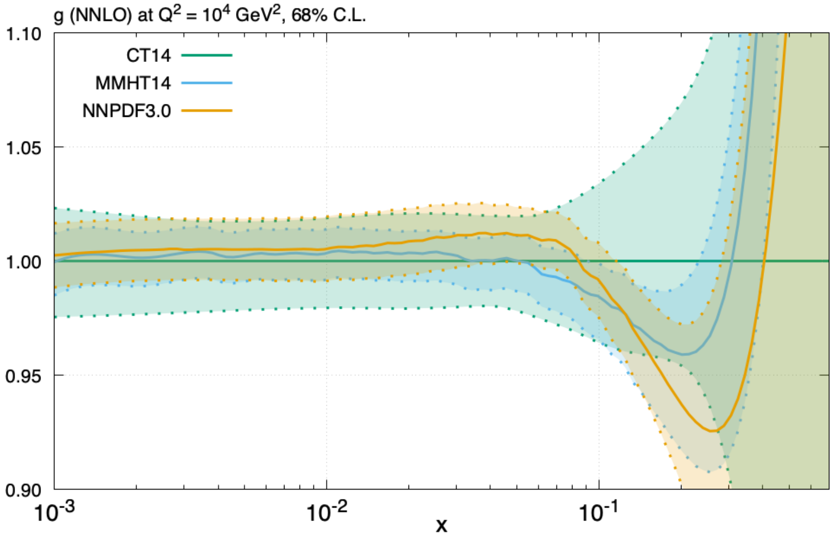
Changes in MSHT PDFs

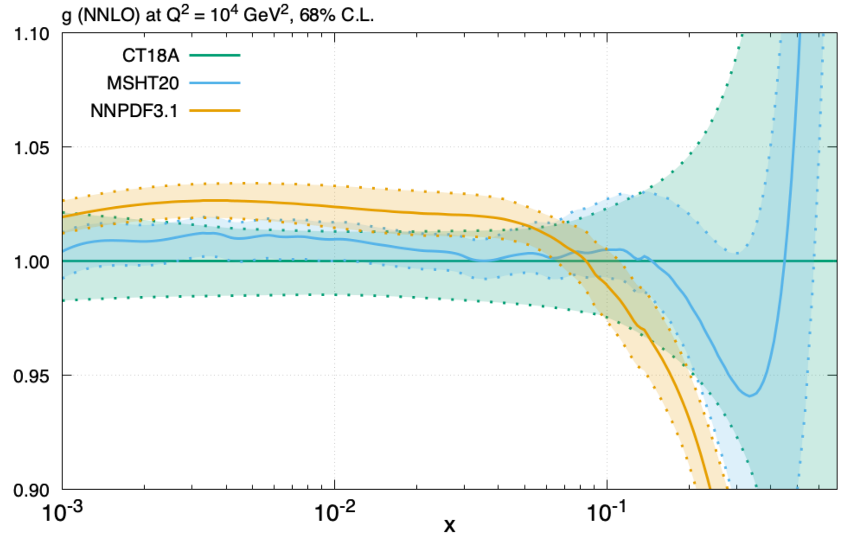


Most significant in d_V (parameterisation and new LHC data) and strange quark. Also $(\bar{d}-\bar{u}) \to (\bar{d}/\bar{u})$ leading to uncertainty increase at small x.

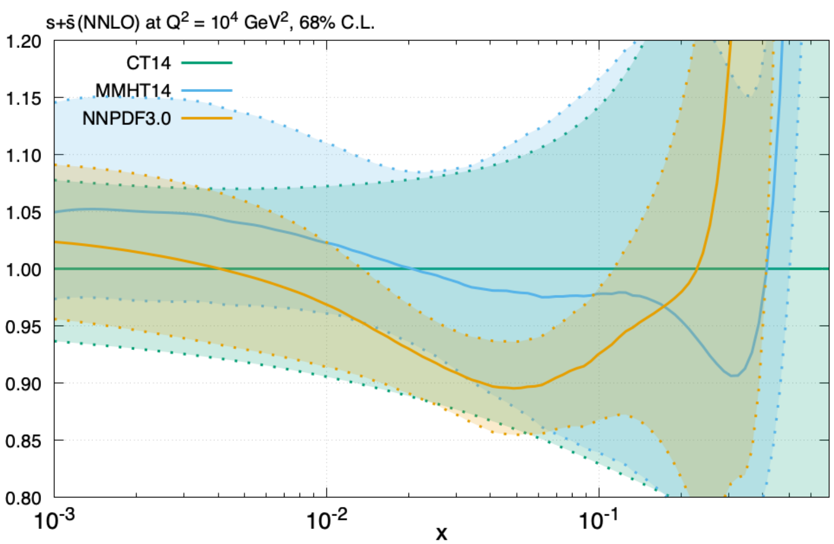
Changes in PDF uncertainties downwards but also in central values.

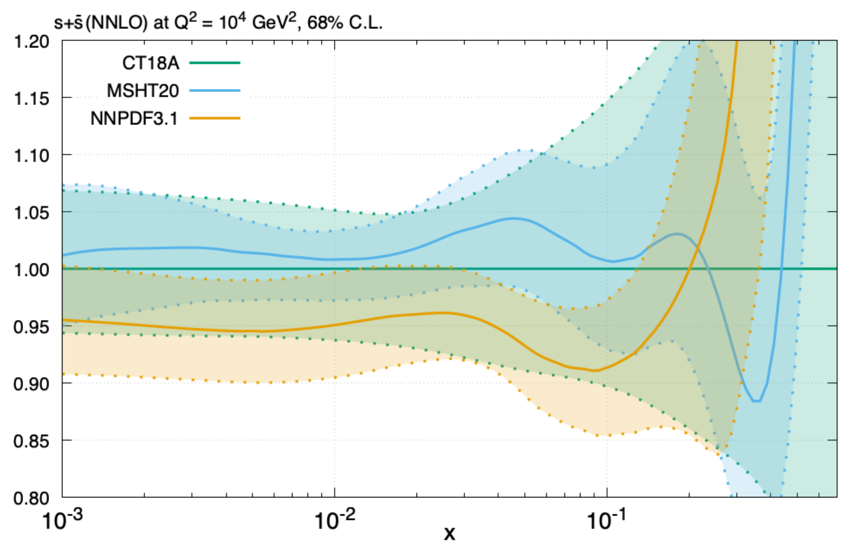
Gluon



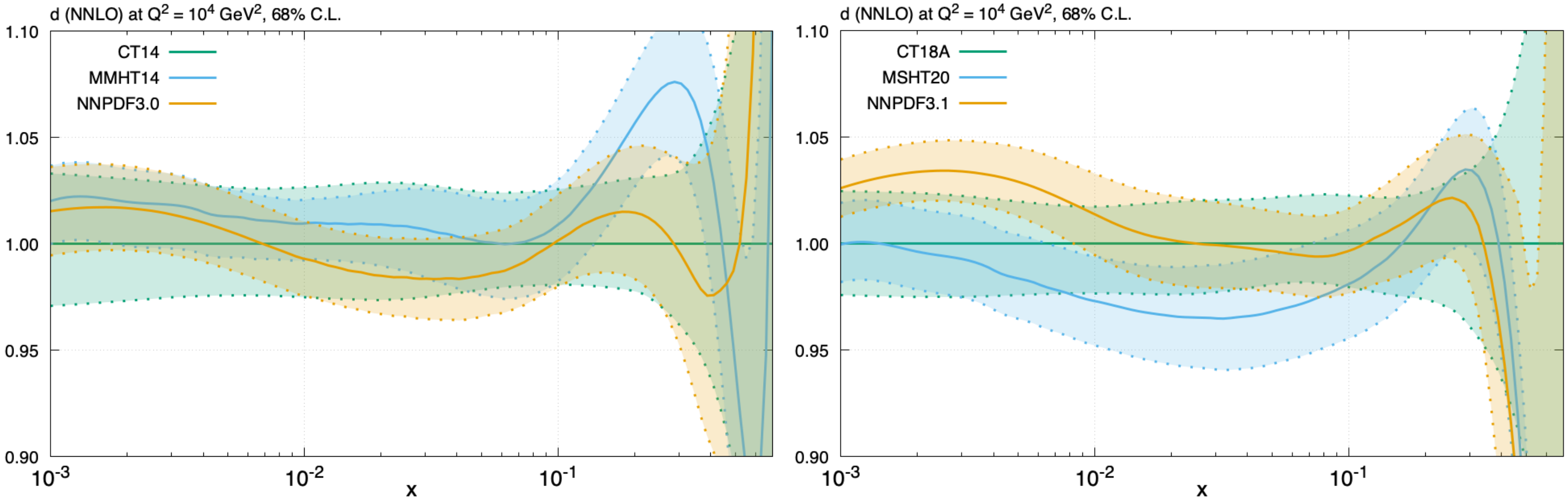


Strange



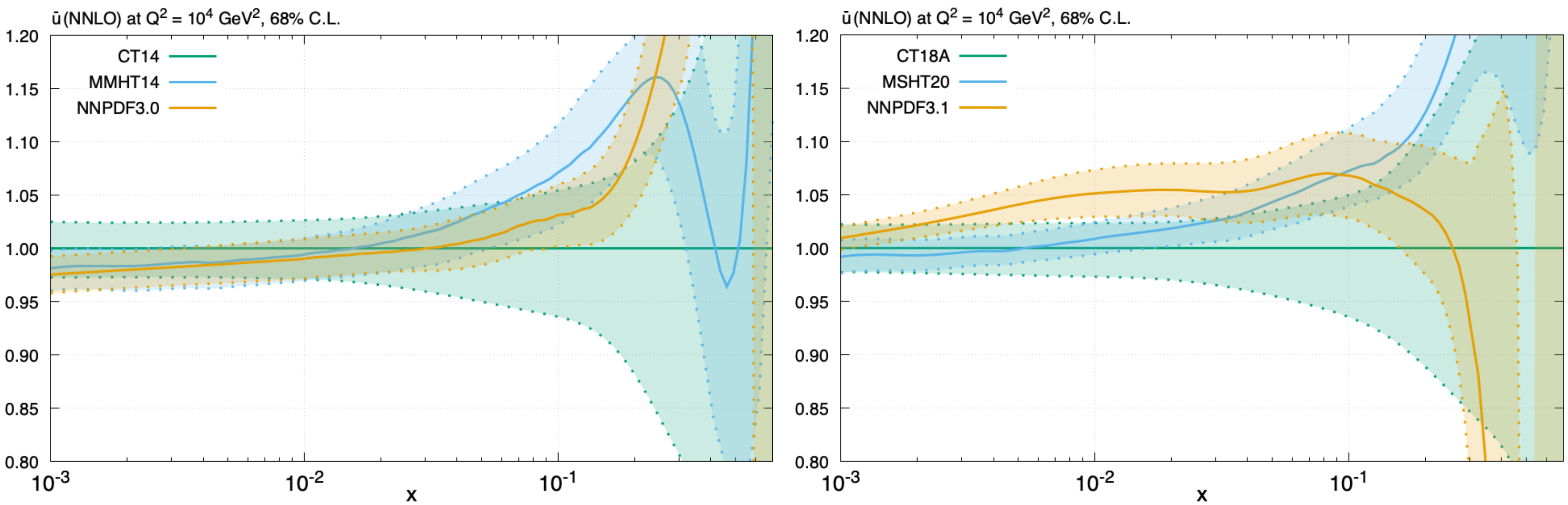


Down



d (NNLO) at Q² = 10⁴ GeV², 68% C.L.
CT14
MMHT14
NNPDF3.0
d (NNLO) at Q² = 10⁴ GeV², 68% C.L.
CT18A
MSHT20
NNPDF3.1
1.10
1.05
1.00
0.95
0.90
10⁻³
10⁻²
10⁻¹
x

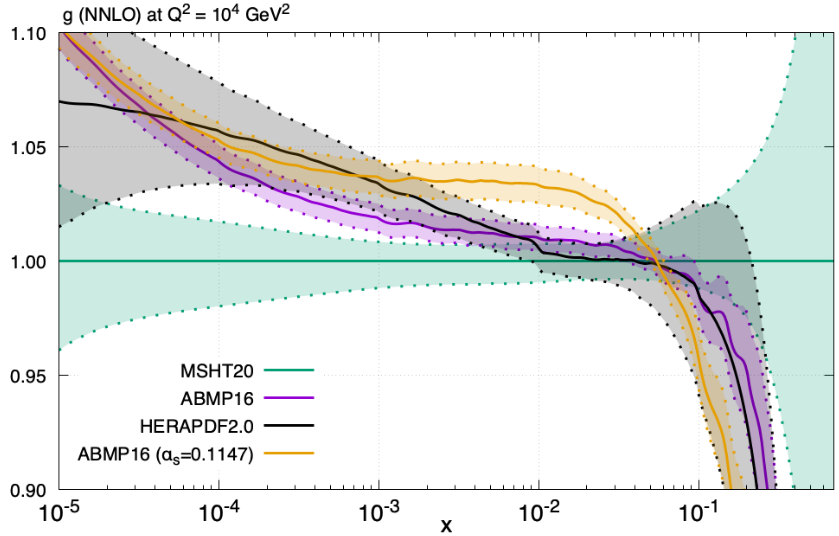
Anti-up

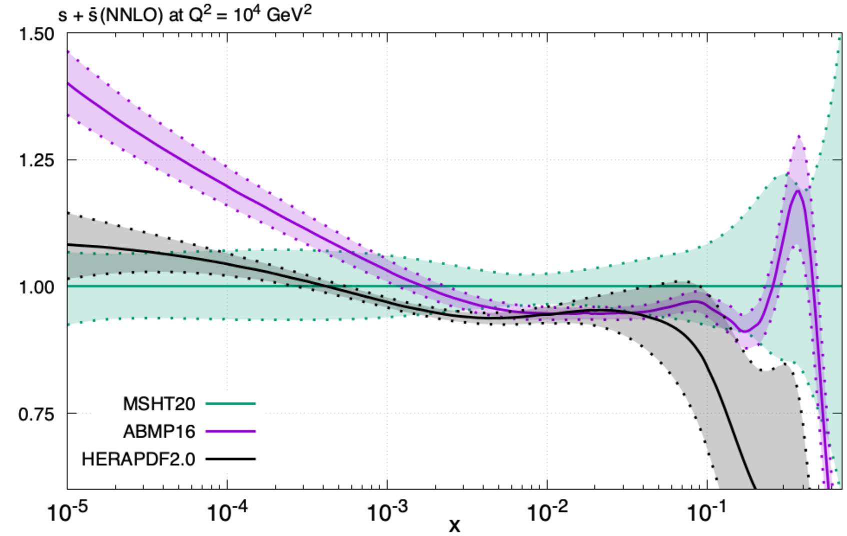


ū (NNLO) at Q² = 10⁴ GeV², 68% C.L.
CT14
MMHT14
NNPDF3.0
ū (NNLO) at Q² = 10⁴ GeV², 68% C.L.
CT18A
MSHT20
NNPDF3.1
1.20
1.15
1.10
1.05
1.00
0.95
0.90
0.85
0.80
10⁻³
10⁻²
10⁻¹
x

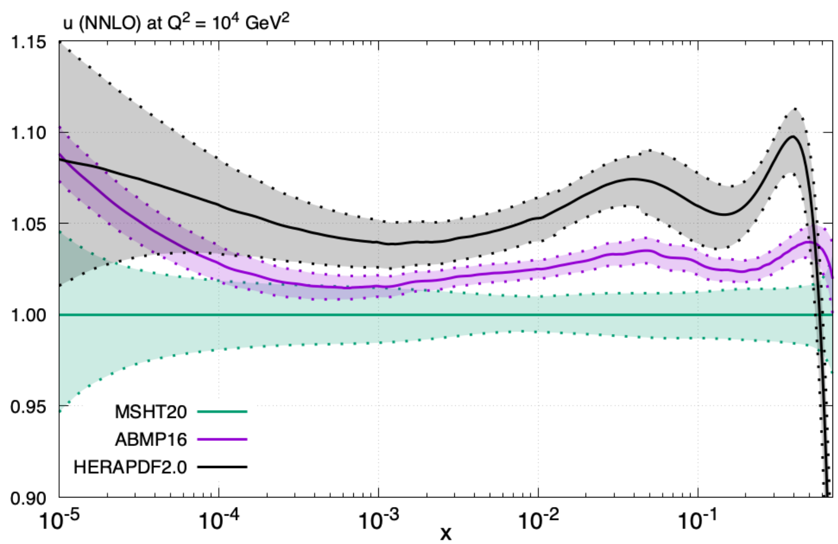
Comparisons to other PDFs. Some big differences.

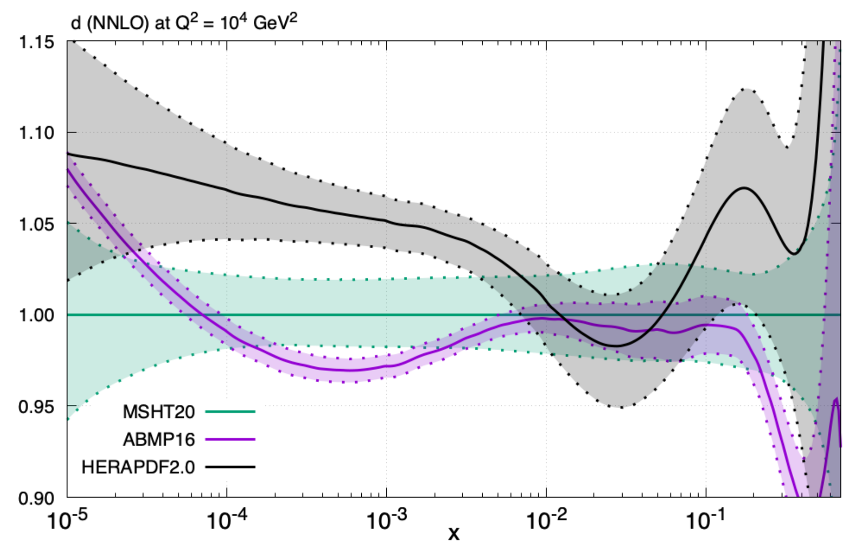
Gluon and Strange





Up and Down





New PDF4LHC Benchmarking exercise

Fit to a subset of data such that all groups (CT, MSHT, NNPDF) fit it (very largely) the same way.

Make definition flexible enough that a decent set of constraints on all PDF flavours and combinations is achieved.

Use most conservative cuts applied by any group – avoid most questionable kinematic regions.

Overall list is surprisingly small.

- NMC deuteron to proton ratio in DIS.

- NuTeV dimuon cross sections.

- HERA I + II inclusive cross sections from DIS.

- E866 fixed target Drell-Yan pd/pp data.

- D0 Z rapidity distribution.

- ATLAS W, Z $7\ \mathrm{TeV}$ rapidity distributions, only Z peak and not forward rapidity.

- CMS $7\ \mathrm{TeV}$ W asymmetry.

- CMS $8\ \mathrm{TeV}$ inclusive jet data.

- LHCb $7, 8\ \mathrm{TeV}$ W, Z rapidity distributions.

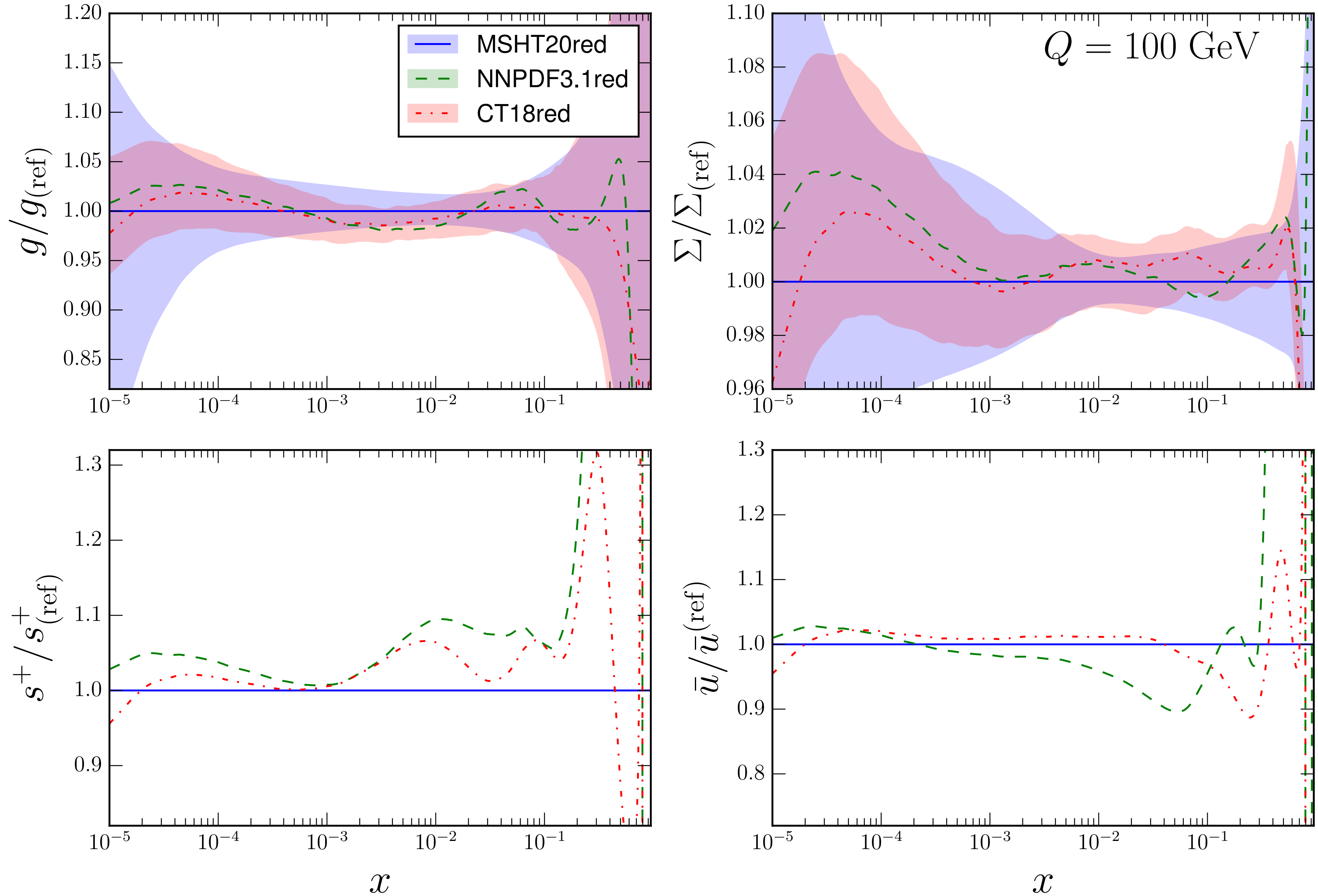
- BCDMS proton and deuteron DIS (MSHT use averaged data).

Set as many theoretical procedural choices the same as possible.

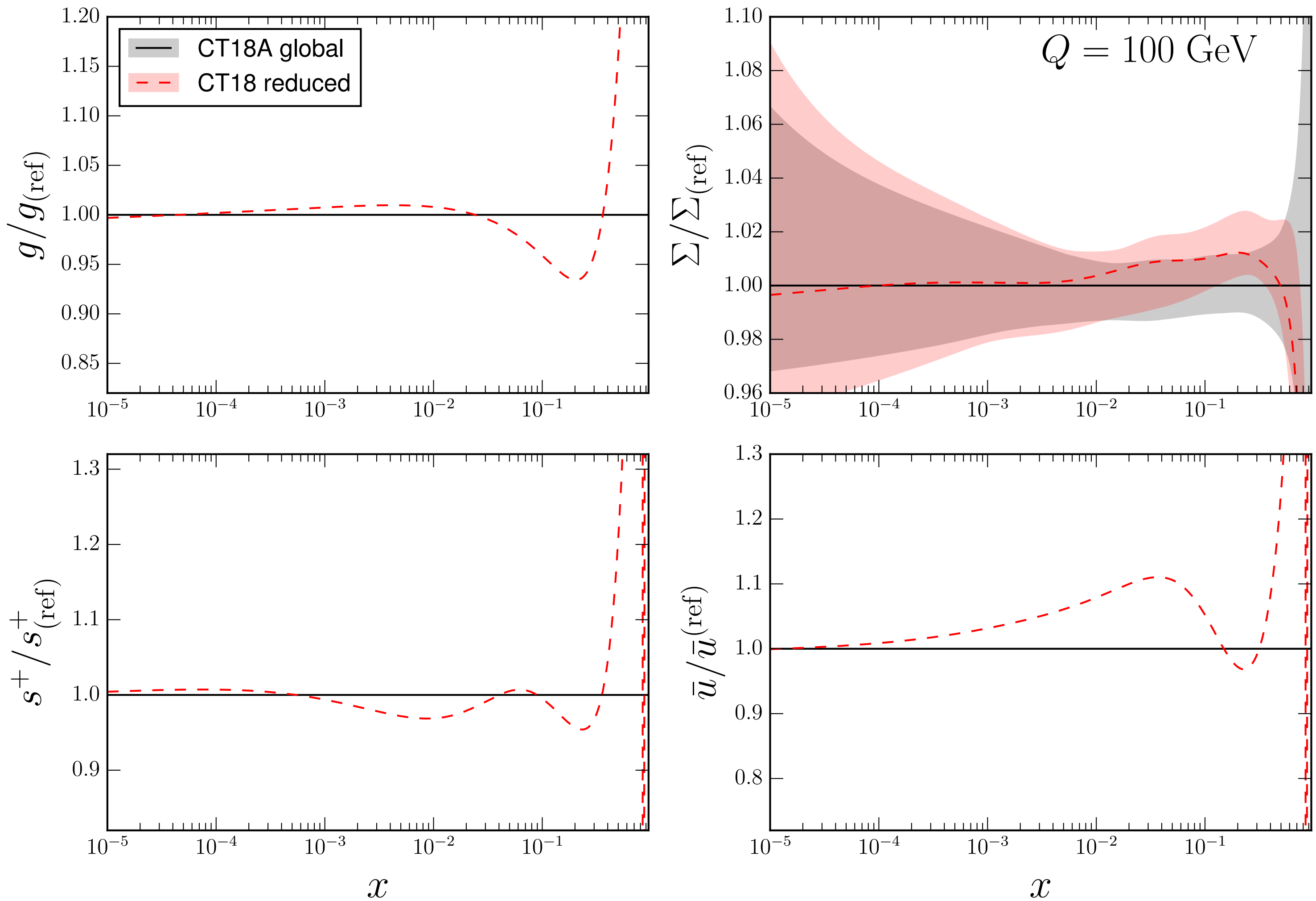
- $s - \bar{s} = 0$
- Perturbative charm.
- Positive definite quark distributions (lack of constraints allow negative fluctuations).
- Common values of $\alpha_S(M_Z^2) = 0.118$, $m_c = 1.4\mathrm{GeV}$ and $m_b = 4.75\mathrm{GeV}$.
- No deuteron or nuclear corrections.
- Fixed branching ratio for charm hadrons $\rightarrow$ muons.
- NNLO corrections for dimuon data.

Note this is for simplicity and these are not fully the defaults of any group, or indeed, always recommended practice in a global fit.

PDF comparison in reduced fits

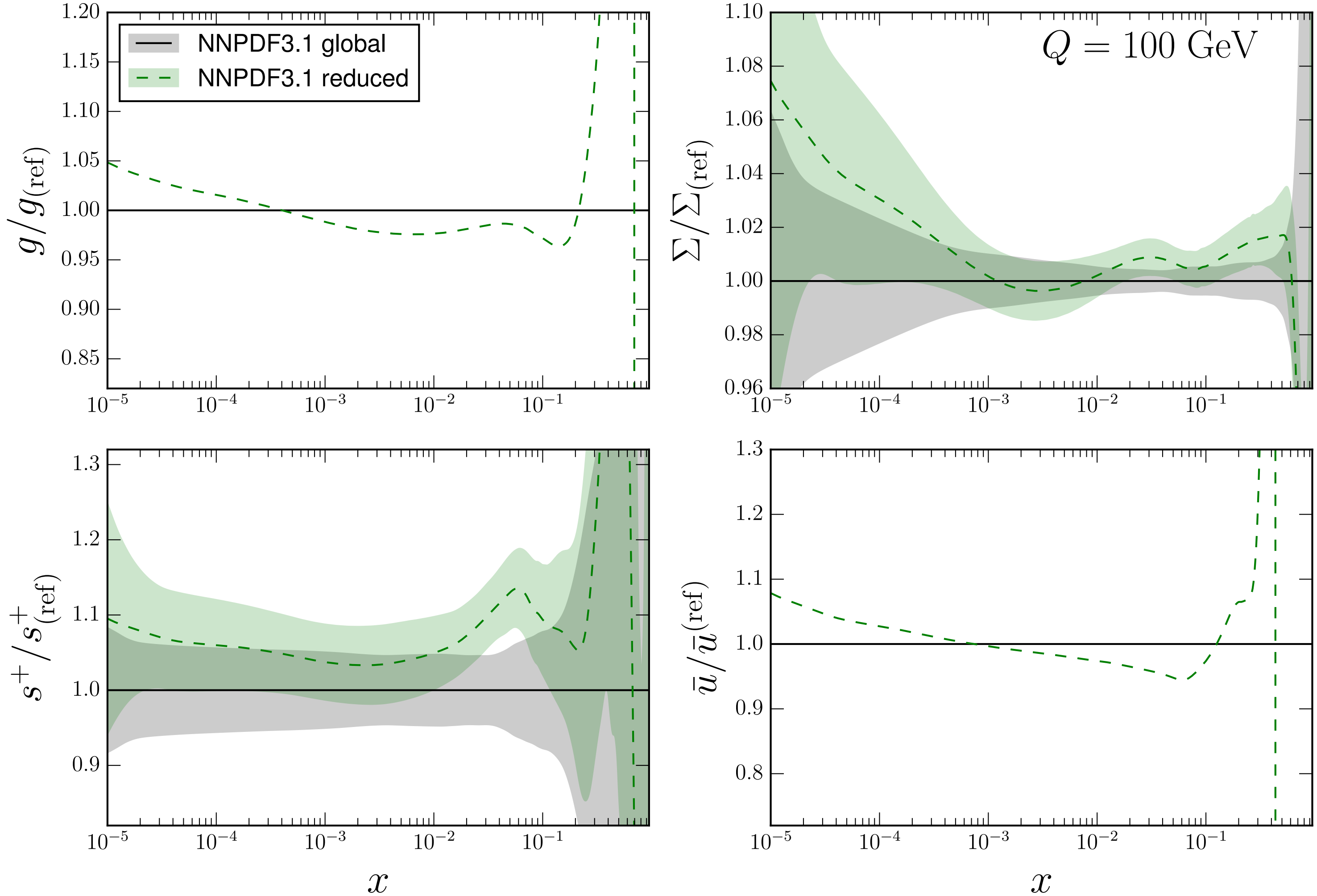


CT18 changes



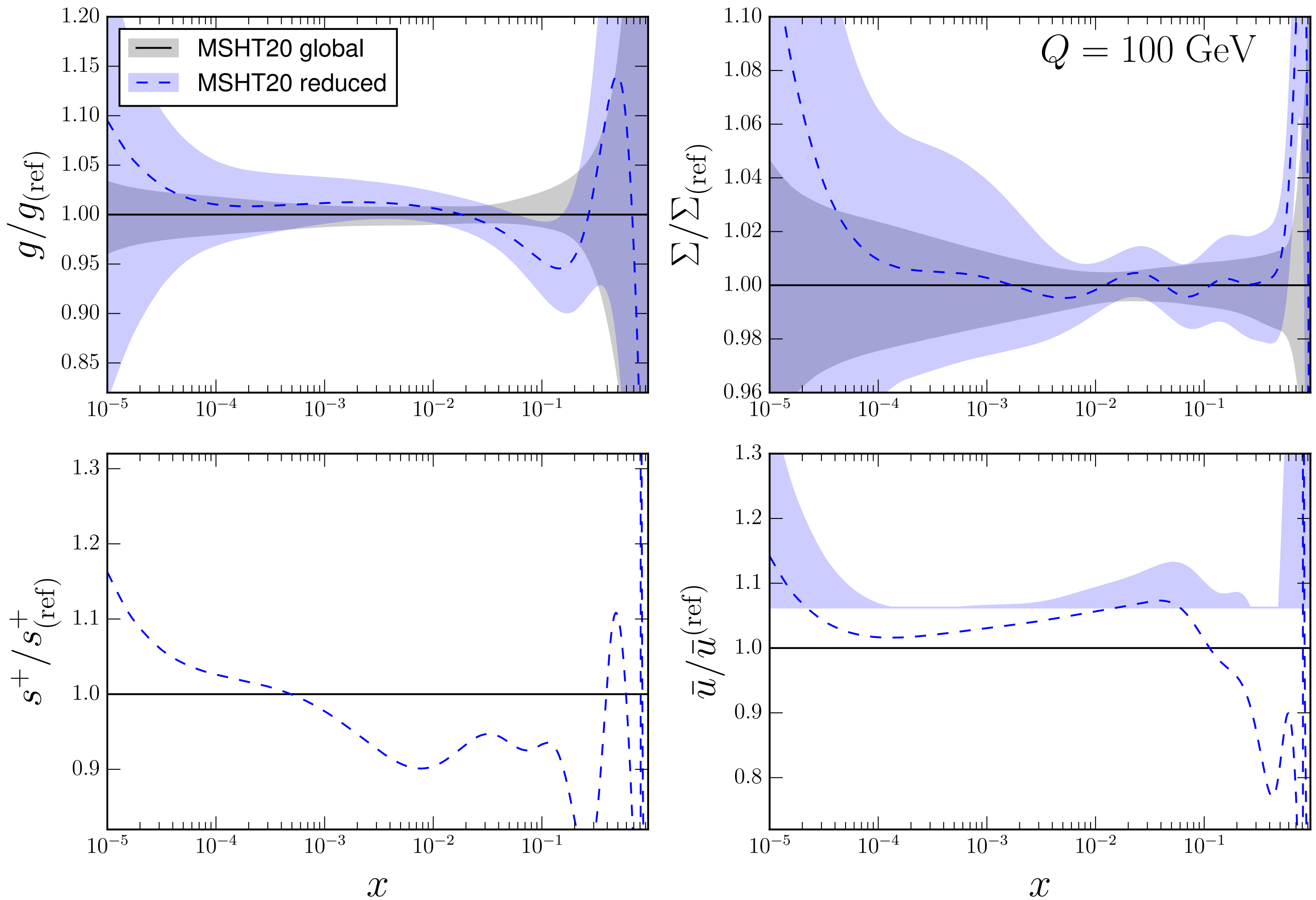
Plots from T. Cridge.

NNPDF3.1 changes



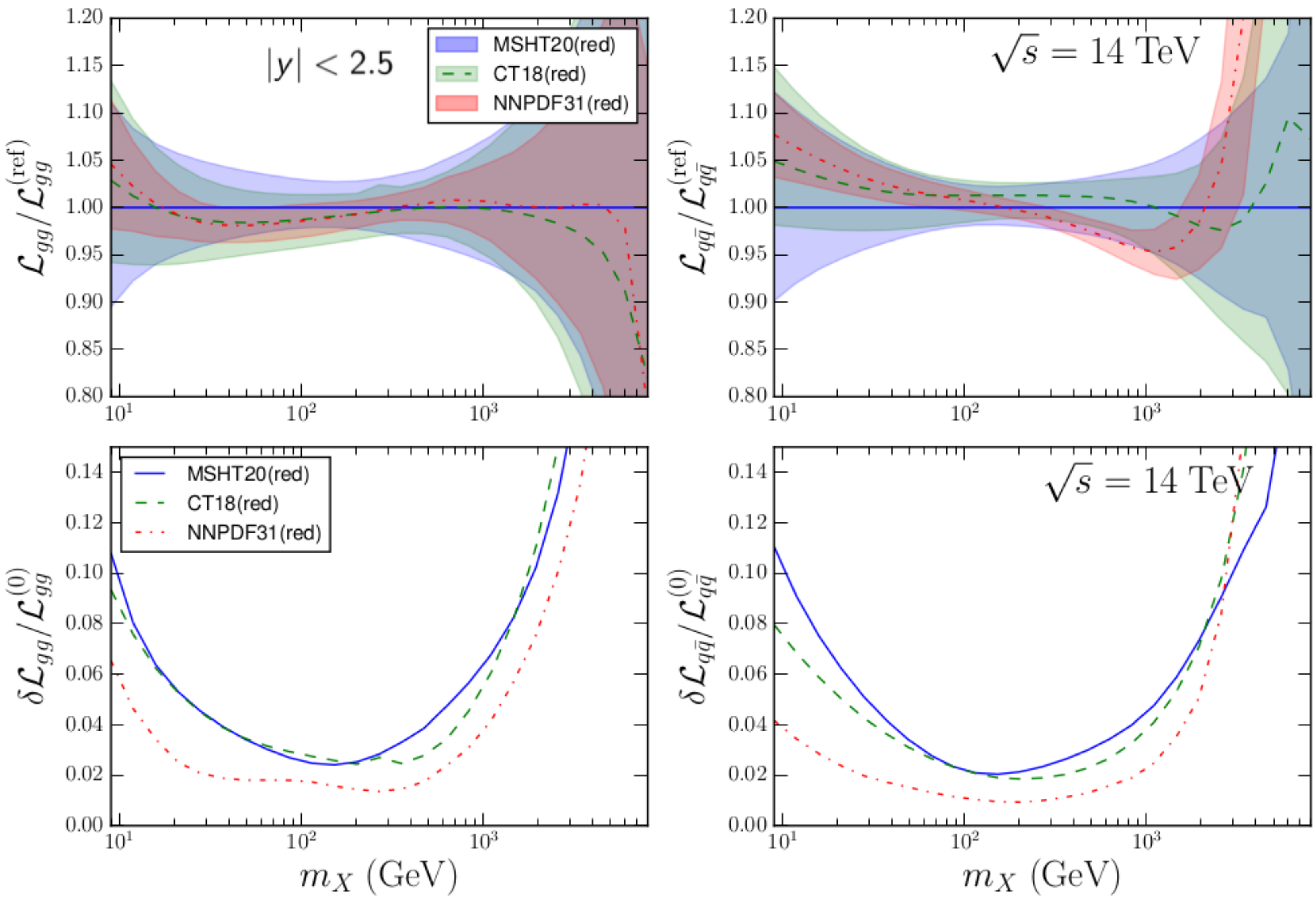
Plots from T. Cridge.

MSHT20 changes



Plots from T. Cridge.

PDF luminosity comparison



Good agreement in “main” luminosities. Plots from T. Cridge.

Theory comparisons

Also compare quality of fit obtained to “identical” data by each group.

ID	Expt.	N_{pt}	χ^2/N_{pt} (CT)	χ^2/N_{pt} (MSHT)	χ^2/N_{pt} (NNPDF)
101	BCDMS F_2^p	$329/163^{\dagger\dagger}/325^{\dagger}$	1.06	1.00	1.21
102	BCDMS F_2^d	$246/151^{\dagger\dagger}/244^{\dagger}$	1.06	0.88	1.10
104	NMC F_2^d/F_2^p	$118/117^{\dagger}$	0.93	0.93	0.90
124+125	NuTeV $\nu\mu\mu + \bar{\nu}\mu\mu$	38+33	0.79	0.83	1.22
160	HERAI+II	1120	1.23	1.20	1.22
203	E866 $\sigma_{pd}/(2\sigma_{pp})$	15	1.24	0.80	0.43
245+250	LHCb 7TeV& 8TeV W,Z	29+30	1.15	1.17	1.44
246	LHCb 8TeV $Z \to ee$	17	1.35	1.43	1.57
248	ATLAS 7TeV W,Z(2016)	34	1.96	1.79	2.33
260	D0 Z rapidity	28	0.56	0.58	0.62
267	CMS 7TeV eletron A_{ch}	11	1.47	1.52	0.76
269	ATLAS 7TeV W,Z(2011)	30	1.03	0.93	1.01
545	CMS 8TeV incl. jet	$185/174^{\dagger\dagger}$	1.03	1.39	1.30
Total	N_{pt}	—	2263	1991	2256
Total	χ^2/N_{pt}	—	1.14	1.15	1.20

Table from T. Hobbs.

Do the same when using the same PDFs (PDF4LHC15) as input.

ID	Expt.	N_{pt}	χ^2/N_{pt} (CT)	χ^2/N_{pt} (MSHT)	χ^2/N_{pt} (NNPDF)
101	BCDMS F_2^p	$329/163^{\dagger\dagger}/325^{\dagger}$	1.35	1.2	1.51
102	BCDMS F_2^d	$246/151^{\dagger\dagger}/244^{\dagger}$	0.97	1.27	1.24
104	NMC F_2^d/F_2^p	$118/117^{\dagger}$	0.92	0.93	0.94
124+125	NuTeV $\nu\mu\mu + \bar{\nu}\mu\mu$	38+33	0.75	0.73	0.84
160	HERAI+II	1120	1.27	1.24	1.74
203	E866 $\sigma_{pd}/(2\sigma_{pp})$	15	0.45	0.54	0.59
245+250	LHCb 7TeV& 8TeV W,Z	29+30	1.5	1.34	1.76
246	LHCb 8TeV $Z \to ee$	17	1.35	1.65	1.25
248	ATLAS 7TeV W,Z(2016)	34	6.71	7.46	6.51
260	D0 Z rapidity	28	0.61	0.58	0.61
267	CMS 7TeV eletron A_{ch}	11	0.45	0.5	0.73
269	ATLAS 7TeV W,Z(2011)	30	1.21	1.23	1.31
545	CMS 8TeV incl. jet	$185/174^{\dagger\dagger}$	1.53	1.89	1.78
Total	N_{pt}	—	2263	1991	2256
Total	χ^2/N_{pt}	—	1.31	1.36	1.62

Table from T. Hobbs.

In order to investigate differences will look in detail at theoretical predictions from each group with same input PDFs.

Some differences expected from e.g. choice of heavy flavour scheme.

Investigation of differences when fitting differential $\bar{t}t$ data.

ATLAS, MSHT, CT all found difficultly simultaneously fitting distributions. Also, cannot get good fit to $y_t, y_{\bar{t}t}$ distributions.

NNPDF fit all distributions well individually. Pattern largely maintained when $\bar{t}t$ data added to reduced fit.

However, NNPDF do not split small data sets into training validation sets, MSHT see improvement when mimicking this.

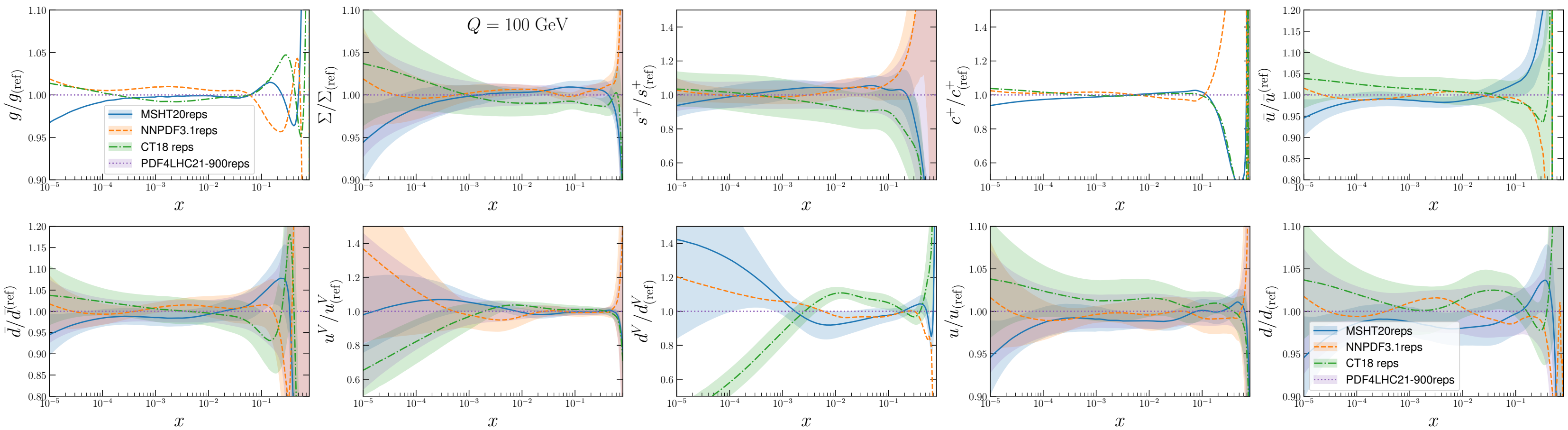
Also fit to top data depends on type and amount of jet data – NNPDF4.0 more similar to other groups.

Dataset (N)	MSHT reduced (default CMS8j)	NNPDF reduced (default CMS8j)	MSHT reduced (CMS7j)	MSHT reduced (AT7j)	MSHT reduced (no jets)	MSHT reduced (CMS8j, double weight $t\bar{t}$)
χ^2/N	1.15	1.20	1.11	1.17	1.12	1.15
p_t^T (8)	3.8	7.2	4.0	4.6	4.5	4.2
y_t (5)	8.4	4.3	6.4	5.5	5.2	5.8
y_{tt} (5)	12.5	5.7	7.2	5.2	6.6	7.4
m_{tt} (7)	6.4	2.4	6.4	6.4	7.4	6.5
$t\bar{t}$ total	31.2	19.6	24.0	21.6	23.8	23.9

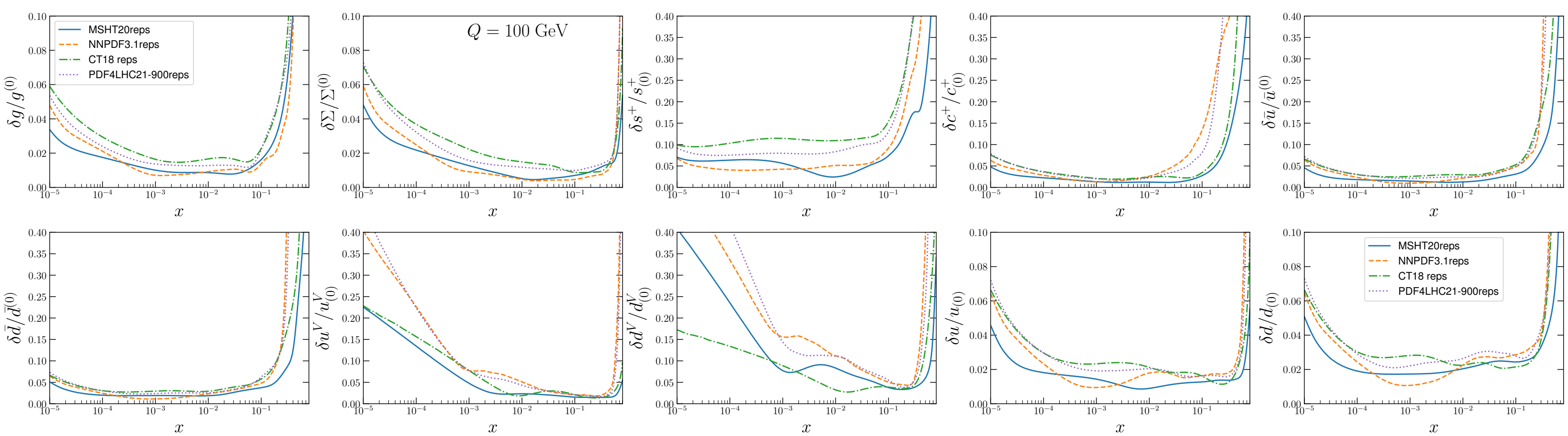
Hence, this issue essentially now understood, though tensions between top and jet data for gluon.

PDF4LHC combination - preliminary. Plots by T. Cridge.

300 replicas from each group (NNPDF3.1 slightly updated) with common $\alpha_S(M_Z^2), m_c, m_b$ produce baseline for combined set.

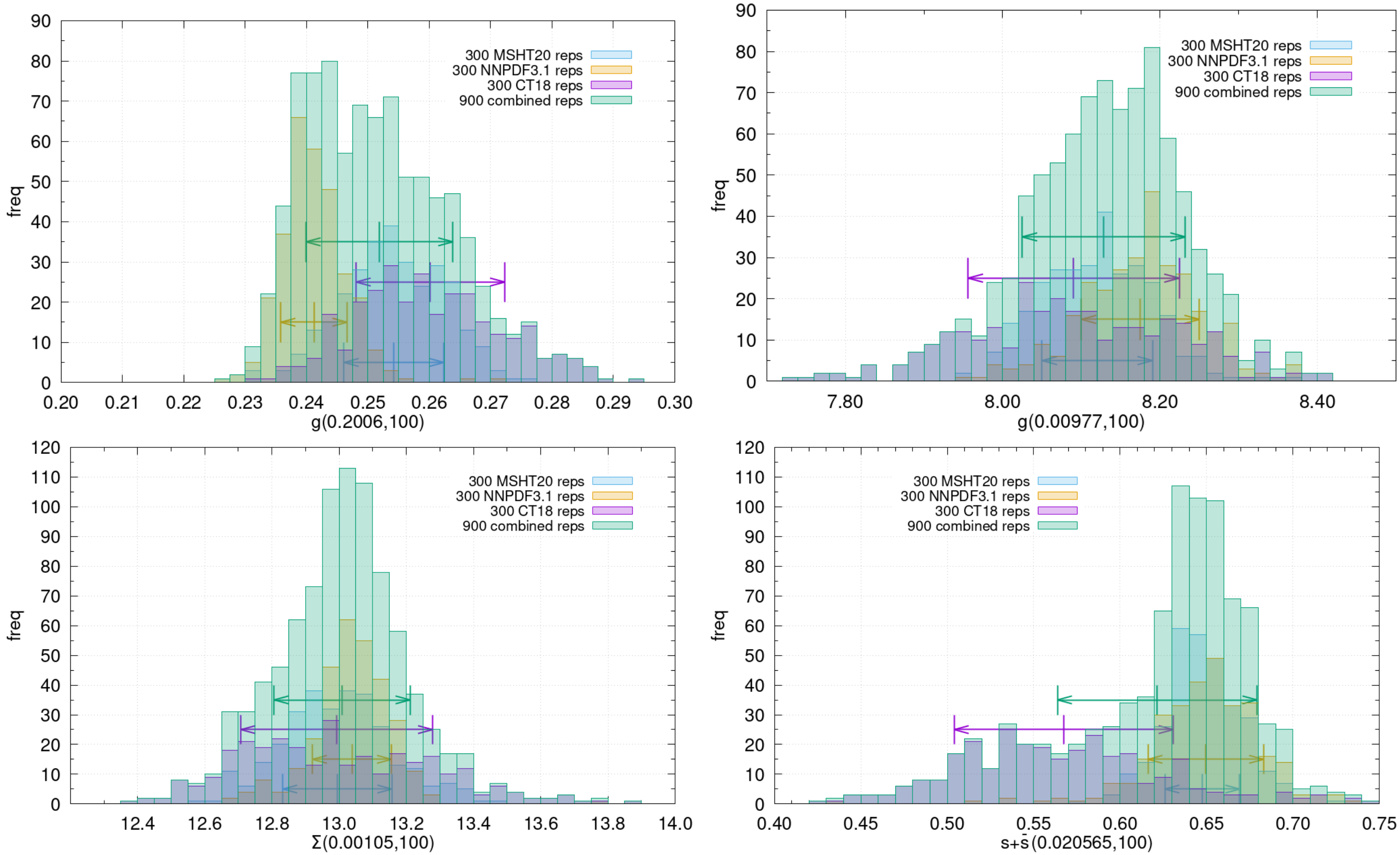


Generally good agreement between groups up to 1 sigma.



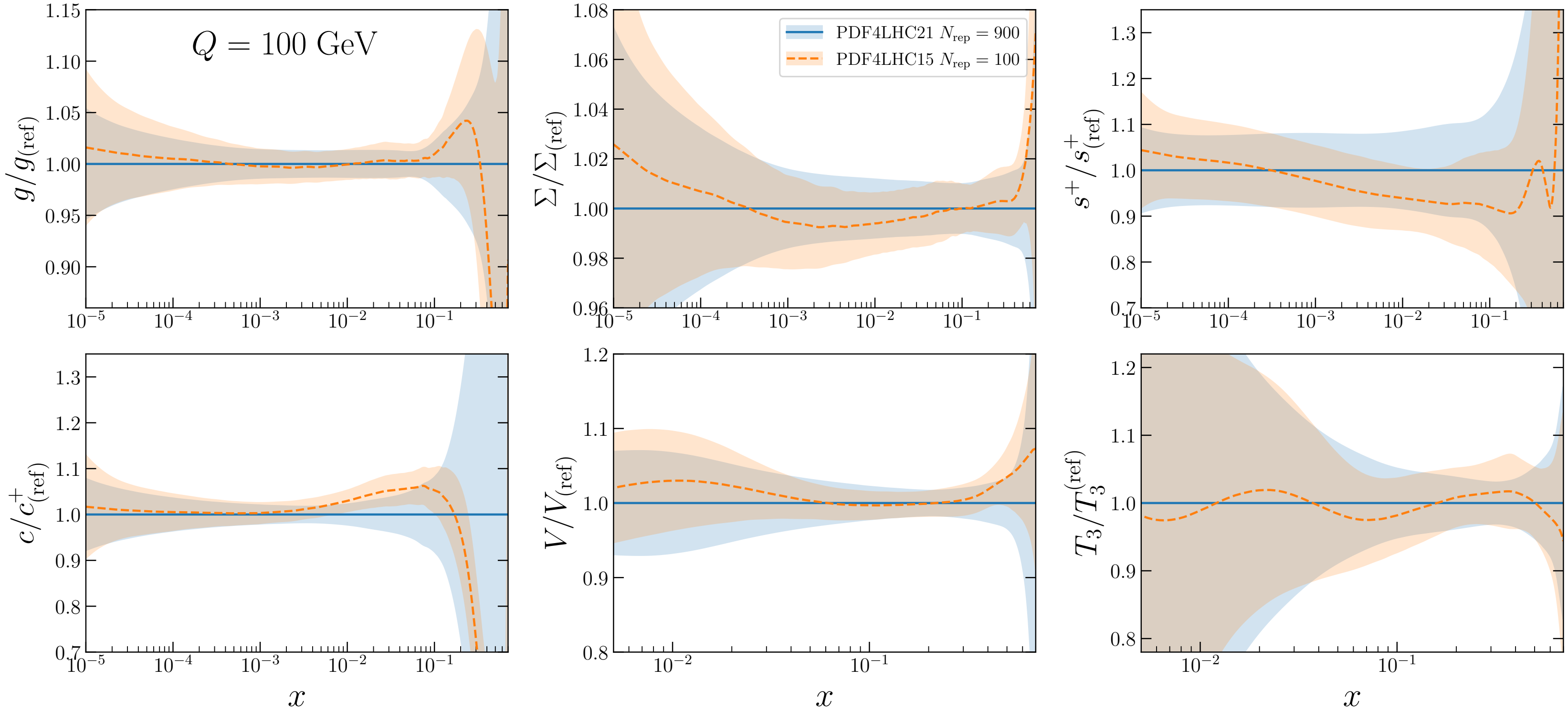
Uncertainty dominated by largest uncertainty or by spread.

Some illustration of distribution at particular points of good or bad agreement.



Plots by T. Cridge. Tails in some distributions.

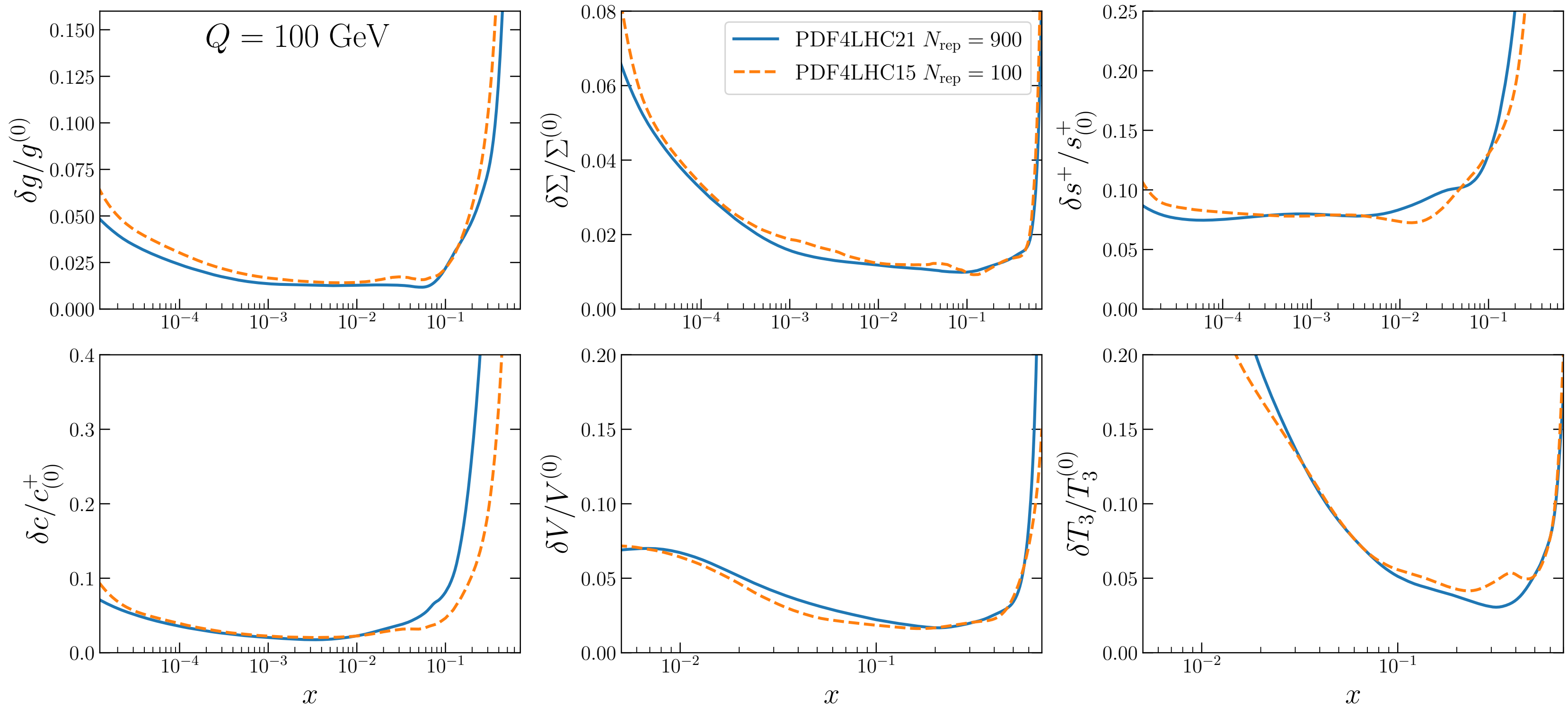
Comparison between PDF4LHC21 and PDF4LHC15 combinations.



Plot by J. Rojo.

Good agreement. Changes in strange, charm understood.

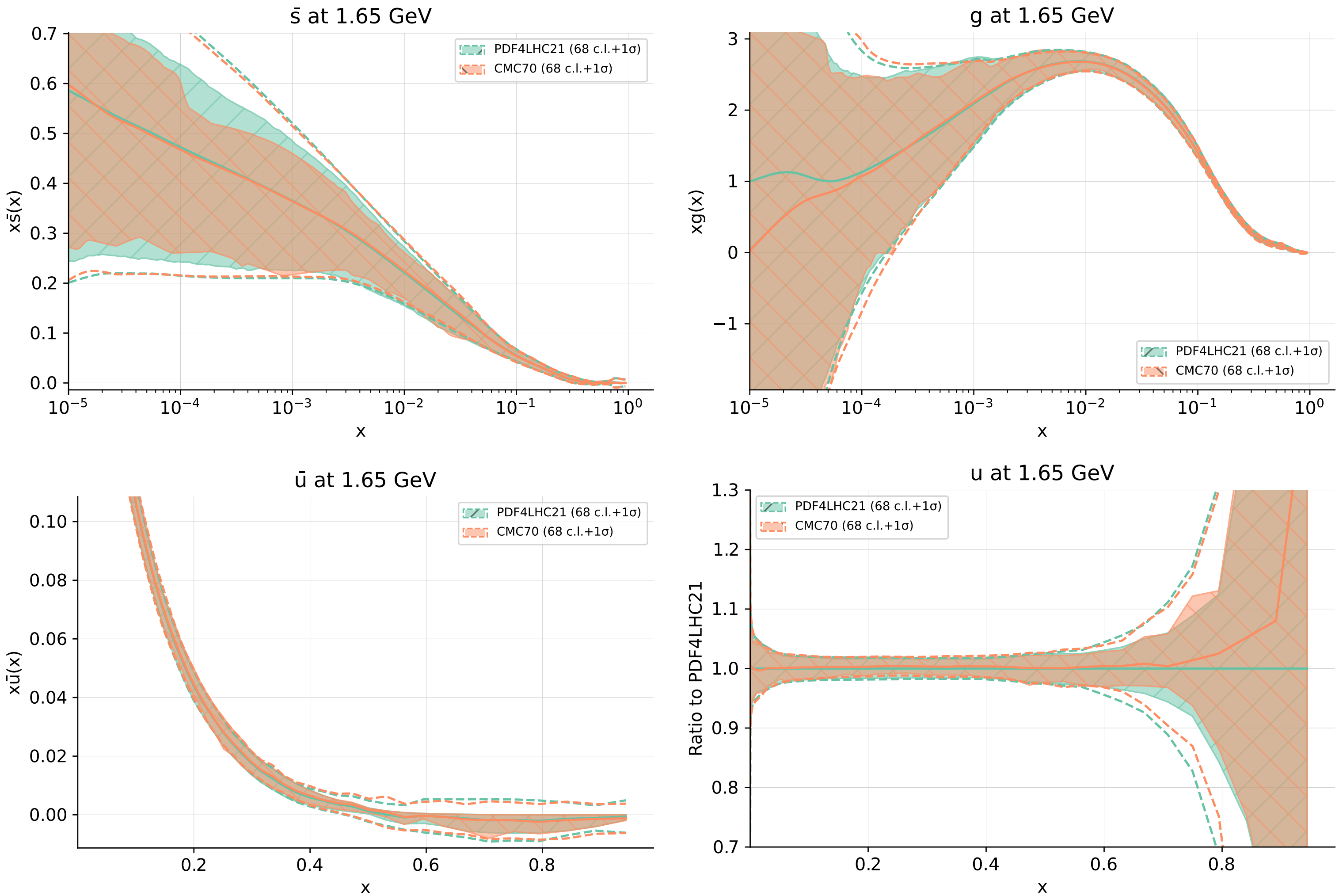
Comparison between uncertainty for PDF4LHC21 and PDF4LHC15 combinations.



Plot by J. Rojo.

Uncertainties comparable. Balance between change in individual uncertainty changes and spreads.

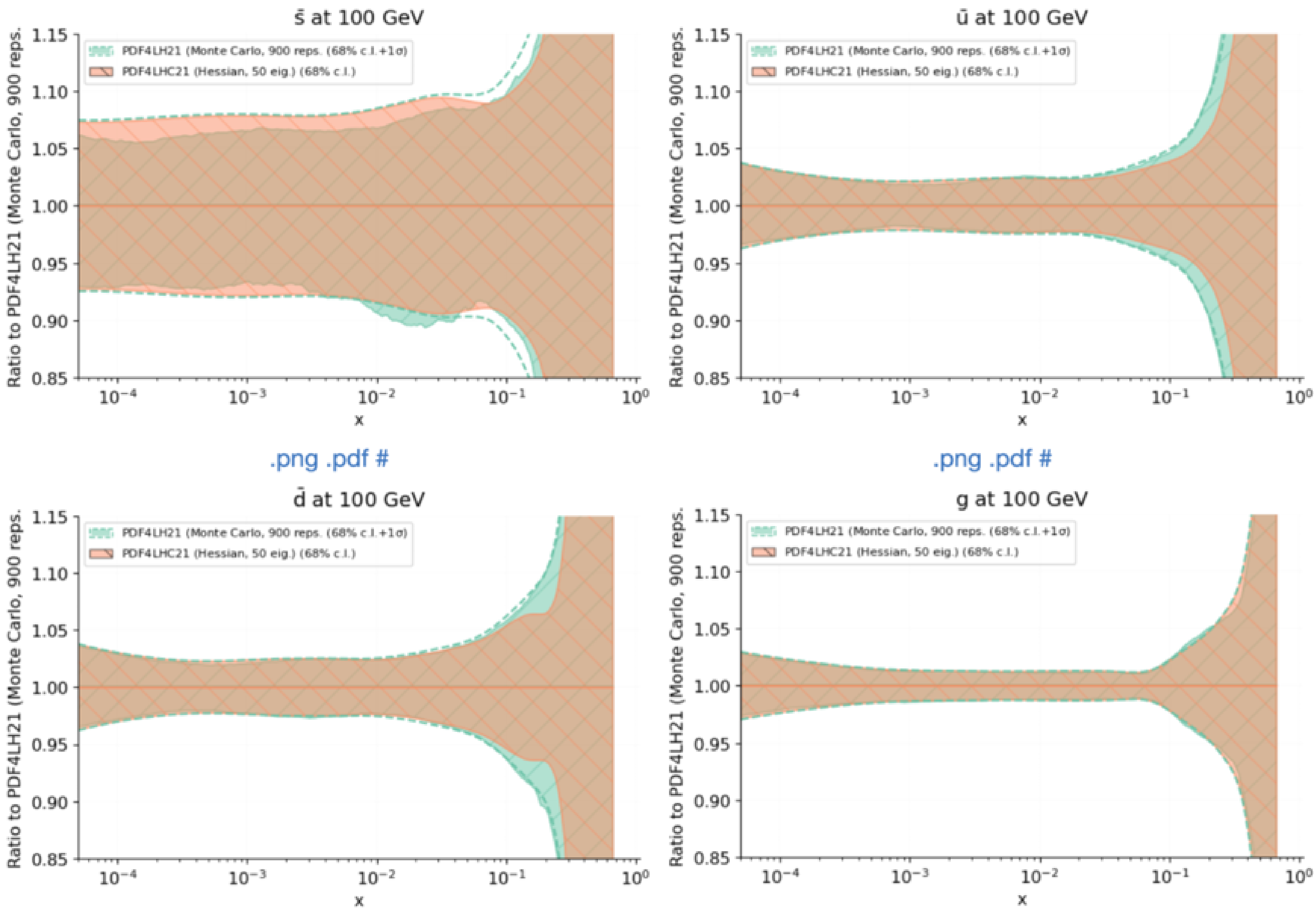
Representation using compressed replica set - preliminary.



Plot by J. Rojo.

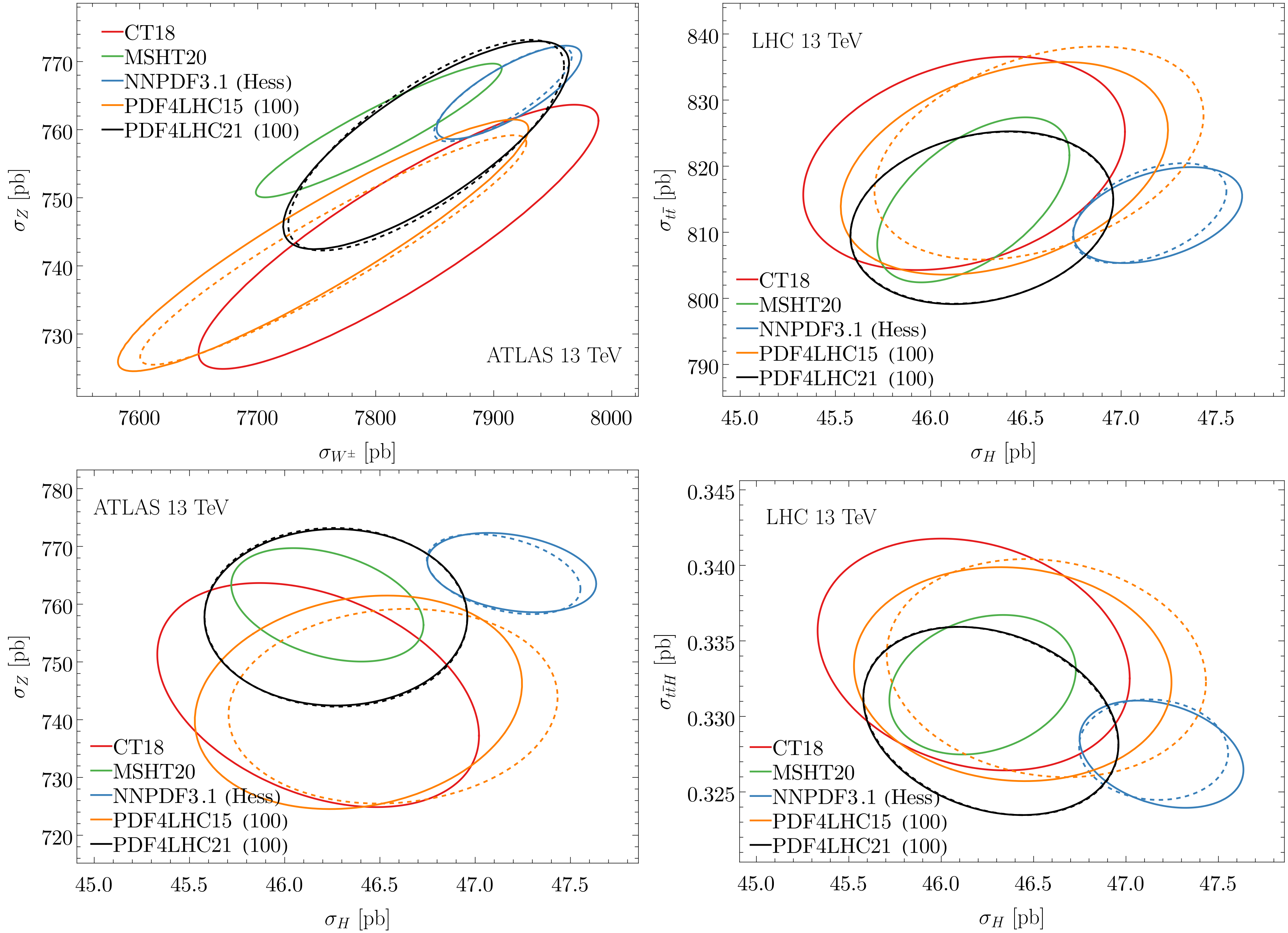
Representation using Hessian replica set - preliminary.

Hessian reduction



much better with 50 eigenvectors, probably **sufficient for phenomenology?**

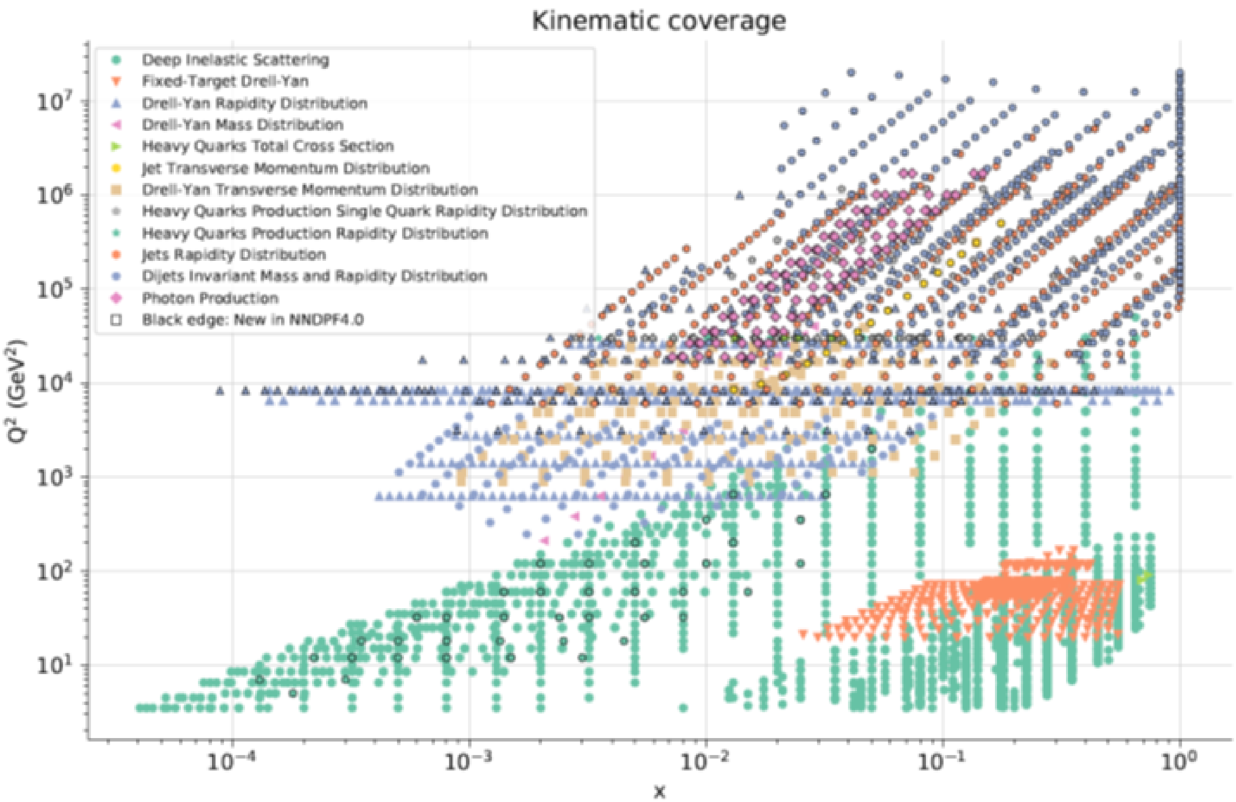
Correlation plots for PDF4LHCHessian sets and individual PDFs.



Plot by K. Xie - **preliminary**.

Subsequent Update – NNPDF4.0

Experimental data in NNPDF4.0



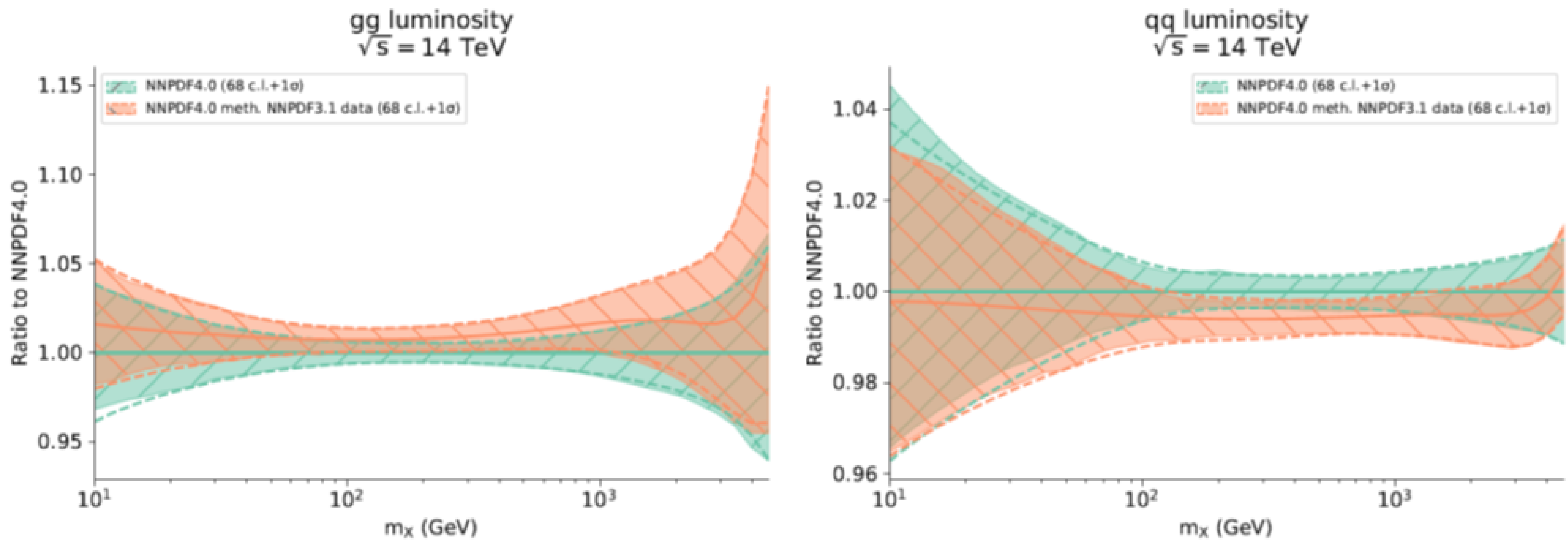
New processes:

- direct photon
- single top
- dijets
- W+jet
- DIS jet

Theoretical improvement

Nuclear uncertainties are included

Impact of the new data

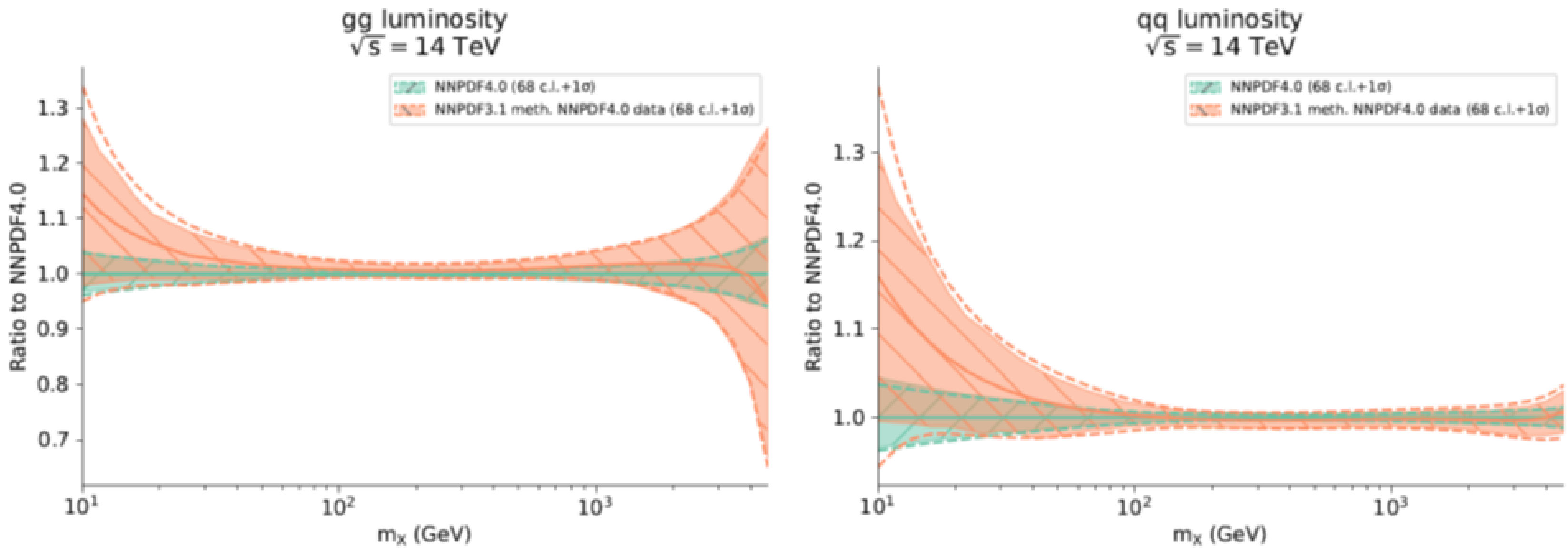


Individual datasets have a limited impact, but collectively they result in:

- Moderate reduction of PDF uncertainties
- Shifts in central value at the one-sigma level

New data shifts central values.

Impact of the new fitting methodology

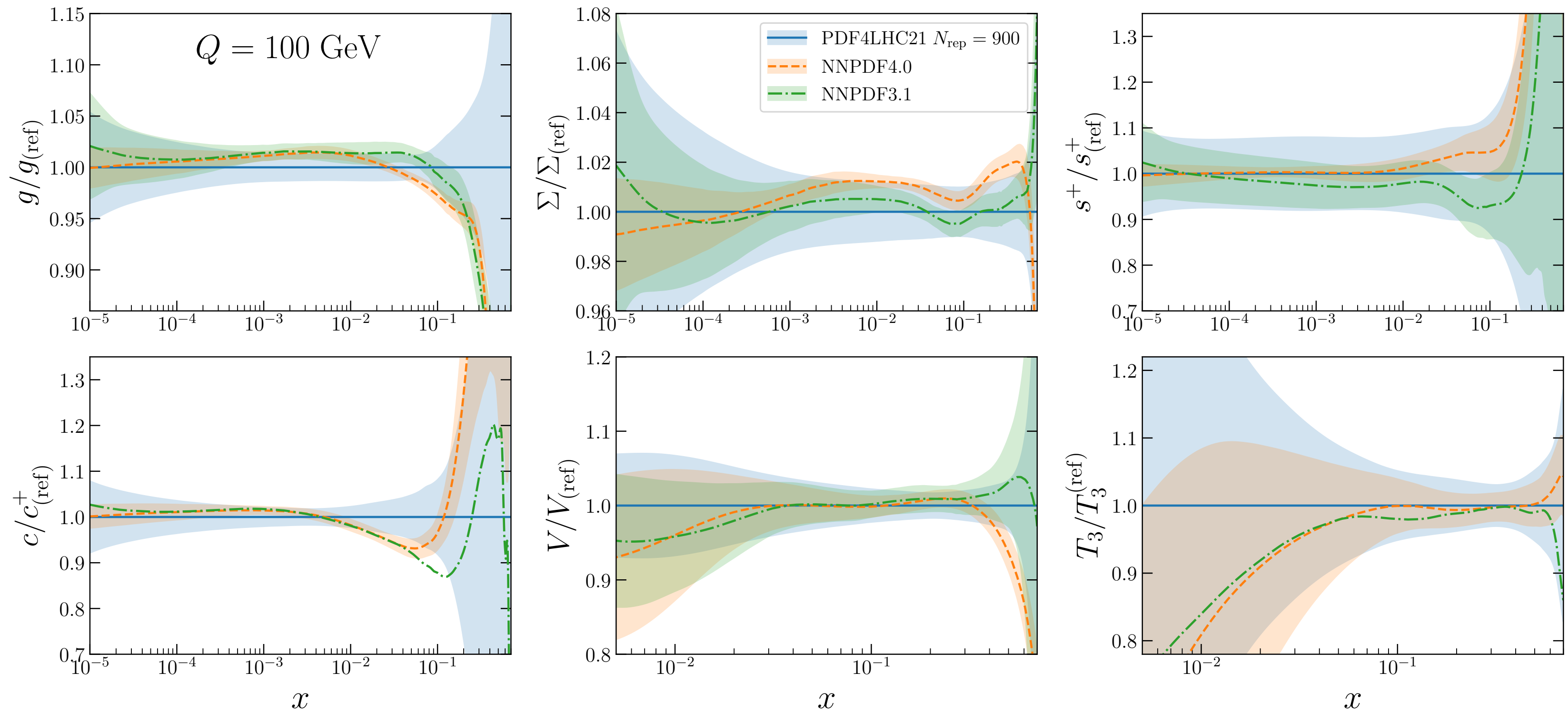


- Significant reduction of PDF uncertainties
- Good agreement between the central values

PDF uncertainties are validated using closure tests and future tests
Validation tests successful for both NNPDF4.0 and NNPDF3.1

Big reduction in uncertainty from methodology.

NNPDF4.0 compared to NNPDF3.1 and PDF4LHC21. J. Rojo.



Consistent but smaller uncertainty. Await updates on other PDFs where data impacts and possibly procedural impacts will be seen.

Conclusions

Many updates from PDF groups in the recent past. Include large but varying amounts of LHC data – starting to have a very significant impact on PDF extractions.

Theory catching up for precision data, e.g NNLO jets, differential top, $Z, W p_T$ . . . More data leads to possible improvements in parameterisation.

Uncertainties generally come down, but not always – in some regions just more realistic.

Agreement between groups generally better, more so in details of flavour separation. Agreement not always better in more general PDF, e.g. gluon.

Benchmarking exercise gives good agreement between groups on most generally constrained quantities, i.e gluon, singlet quarks. Differences in some PDFs depending on flavour separation.

Underlines (clarifies) importance of fitting maximal range of data types to get true constraints. Some PDFs, in particular details of flavour separation, constrained partially, but largely compatibly, by a wide variety of data types.

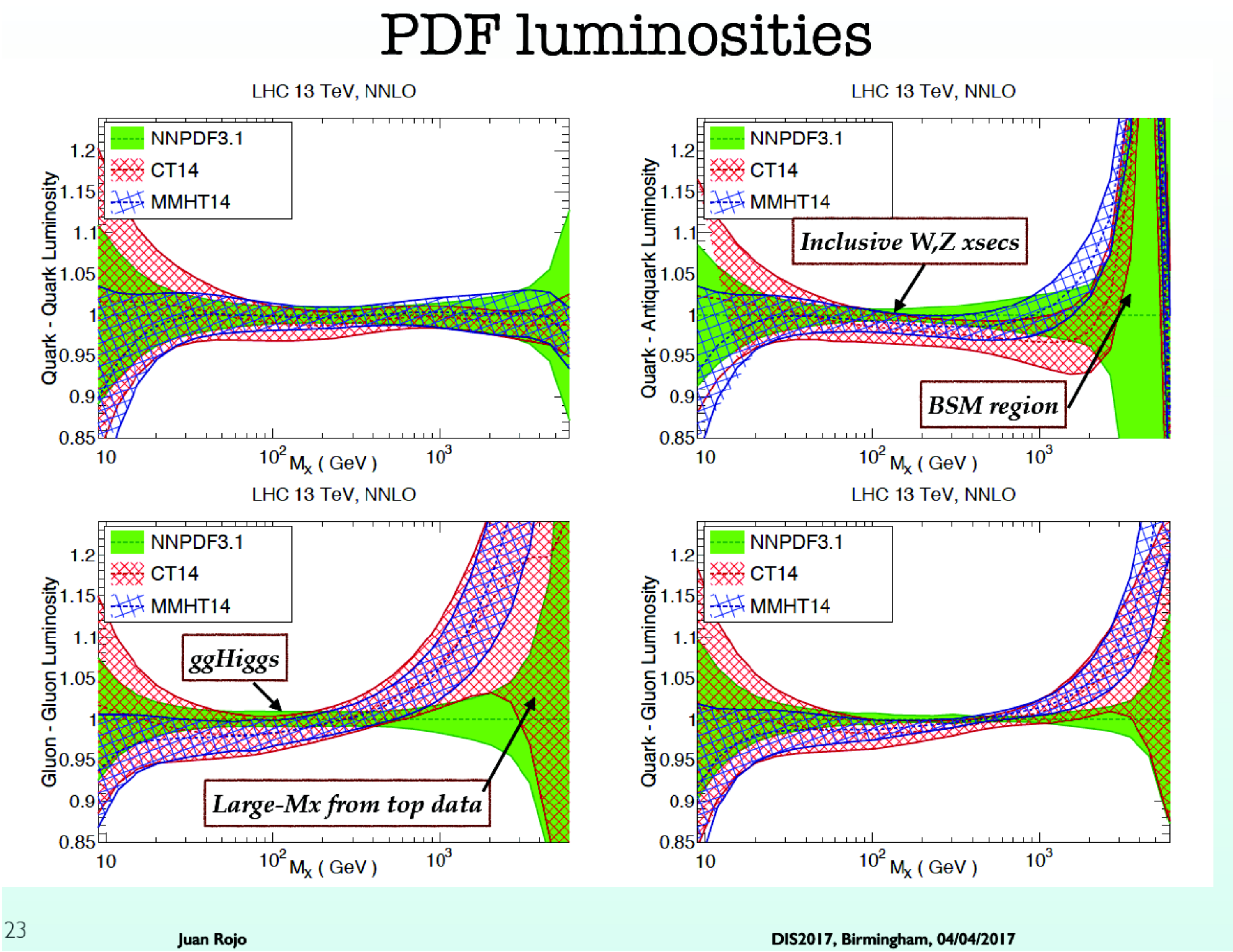
Following success of benchmarking a new combination of CT18, MSHT20 and slightly updated NNPDF3.1 underway as PDF4LHC21.

Showing good compatibility with PDF4LHC15. Despite some more variation between groups uncertainties come down slightly.

Spans range of central values and uncertainties, reflecting information from both.

Deciding on details of presentation - compressed replicas, Hessian sets. Input welcome.

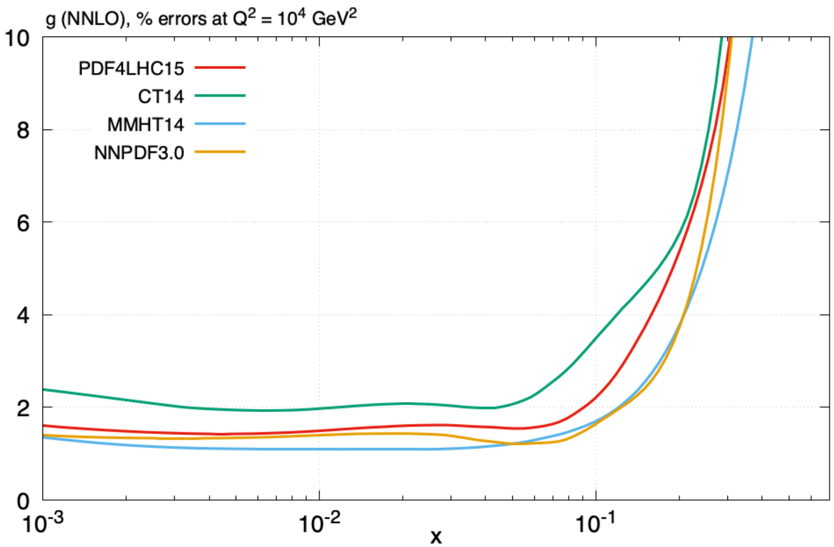
Back-up

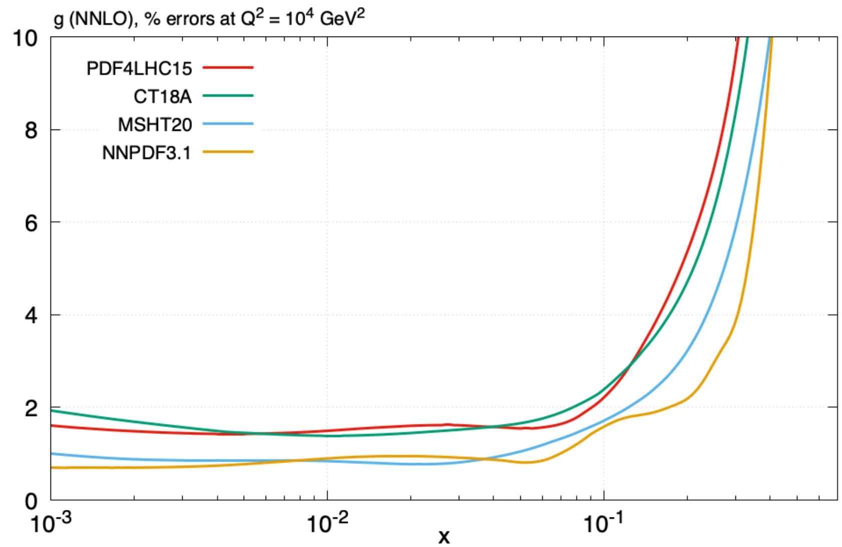


Still good agreement with CT14 and MMHT14 but change in gluon shape and quark increase.

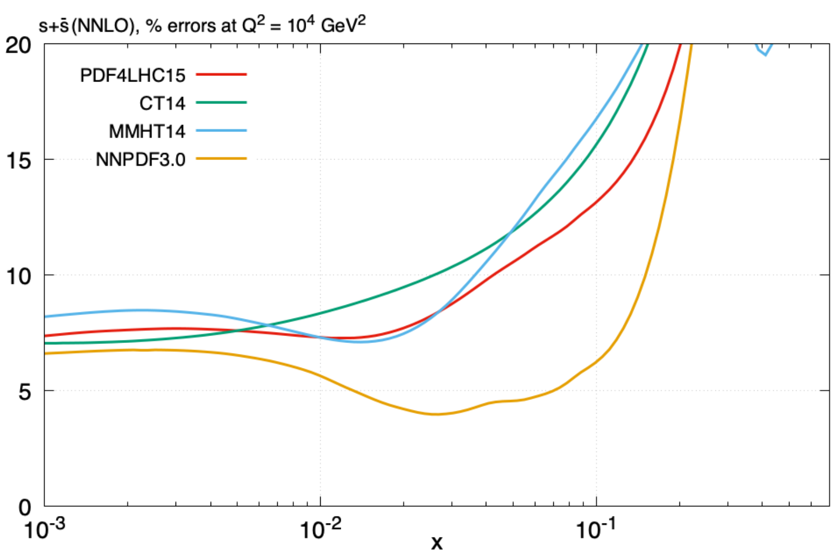
Changes in PDF uncertainties. Largely a decrease.

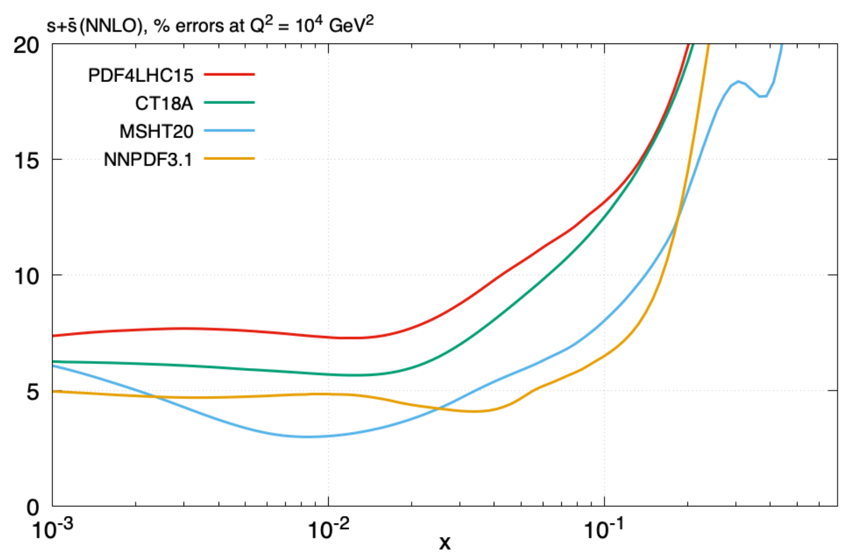
Gluon





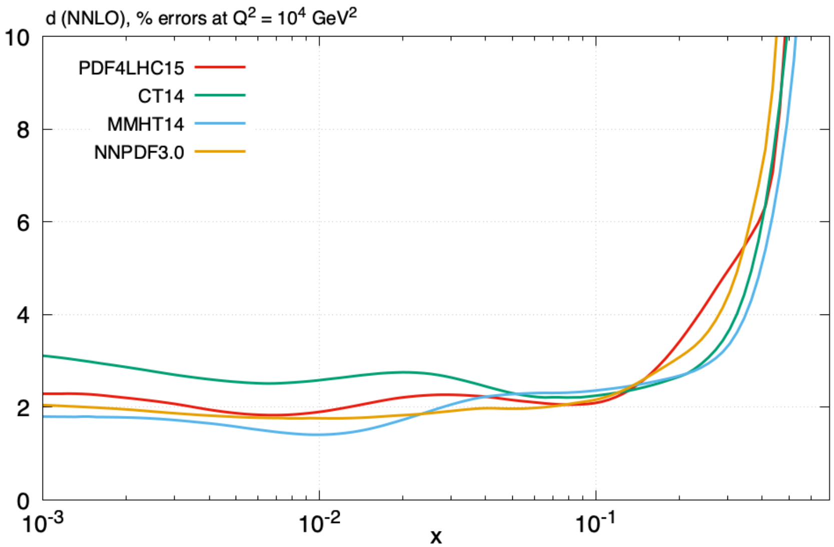
Strange

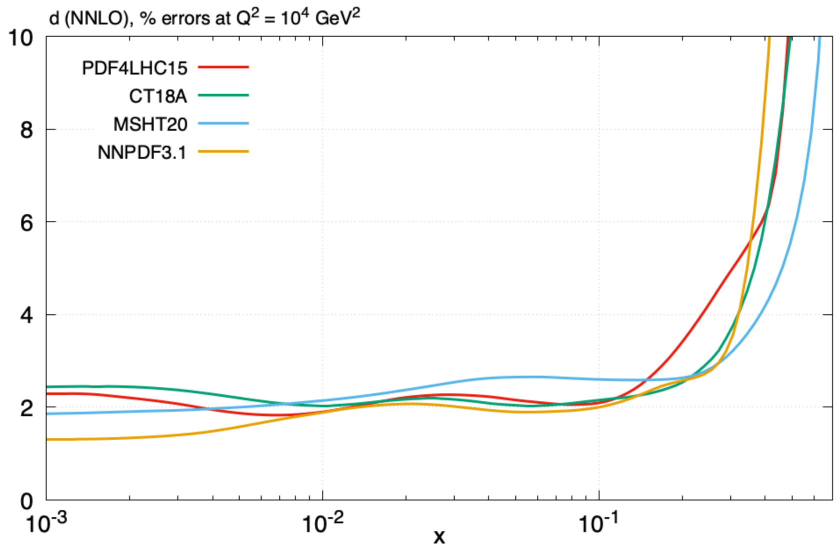




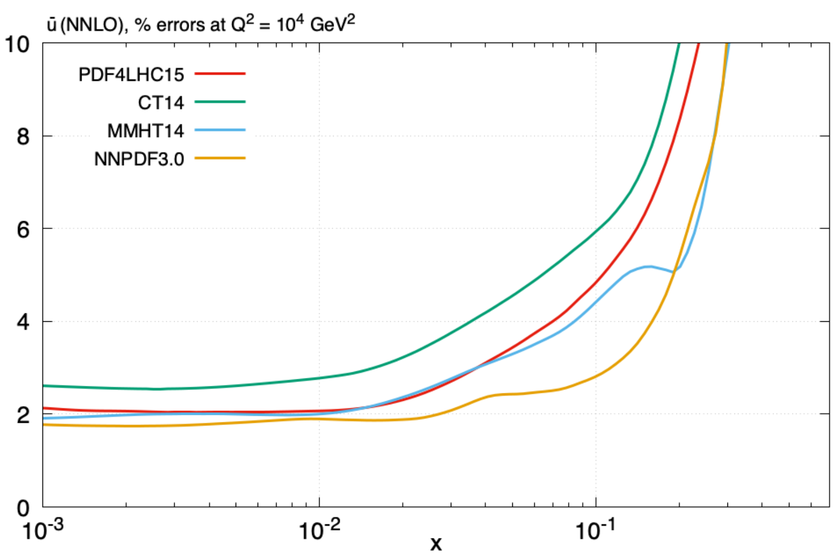
Changes in PDF uncertainties. Largely a decrease.

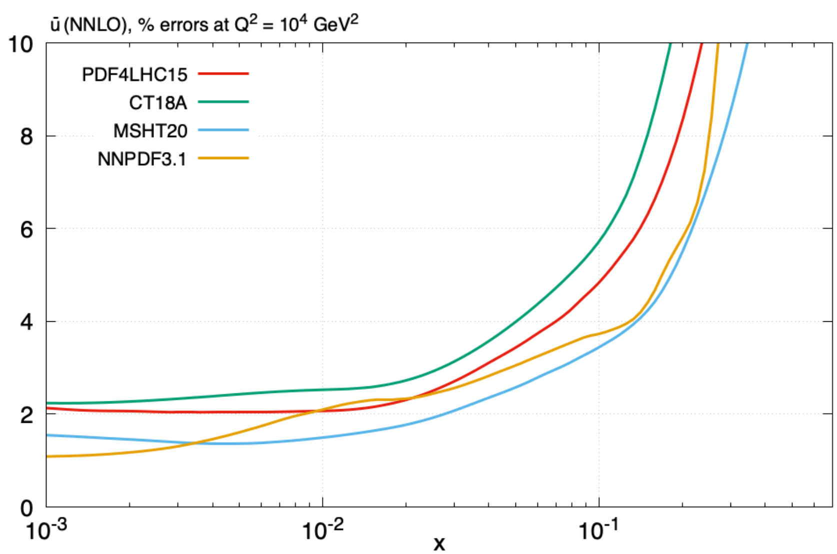
Down





Anti-up





Now extend parameters of PDFs using $n=6$:

Parameterise $(\bar{d}/\bar{u})$ instead of $(\bar{d}-\bar{u})$. $(\bar{d}/\bar{u}) \to \mathrm{constant}$ as $x \to 0$.

$u_v(x,Q_0^2) = A_u(1-x)^{\eta_u}x^{\delta_u}(1+\sum_{i=1}^{6} a_{i,u}T_i(1-2x^{\frac{1}{2}}))$; A_u fixed by $\int_0^1 u_v\,dx = 2$

$d_v(x,Q_0^2) = A_d(1-x)^{\eta_d}x^{\delta_d}(1+\sum_{i=1}^{6} a_{i,d}T_i(1-2x^{\frac{1}{2}}))$; A_d fixed by $\int_0^1 d_v\,dx = 1$

$sea(x,Q_0^2) = A_S(1-x)^{\eta_S}x^{\delta_S}(1+\sum_{i=1}^{6} a_{i,S}T_i(1-2x^{\frac{1}{2}}))$;

$s^+(x,Q_0^2) = A_s(1-x)^{\eta_s}x^{\delta_S}(1+\sum_{i=1}^{6} a_{i,s}T_i(1-2x^{\frac{1}{2}}))$; ($a_{i,s} \neq a_{i,S}, i=5,6$)

$(\bar{d}/\bar{u})(x,Q_0^2) = A_{\mathrm{rat}}(1-x)^{\eta_{\mathrm{rat}}}(1+\sum_{i=1}^{6} a_{i,\mathrm{rat}}T_i(1-2x^{\frac{1}{2}}))$;

$g(x,Q_0^2) = A_g(1-x)^{\eta_g}x^{\delta_g}(1+\sum_{i=1}^{4} a_{i,g}T_i(1-2x^{\frac{1}{2}})) - A_{g_-}(1-x)^{\eta_{g-}}x^{\delta_{g-}}$;

$s^-(x,Q_0^2) = A_{s_-}(1-x)^{\eta_{s-}}(1-x_o/x)x^{\delta_{s-}}$. x_0 fixed by $\int_0^1 s^-\,dx = 0$, δ_{s-} fixed.

Change of to a maximum of 51 parton parameters.

When determining uncertainties go from 25 eigenvector pairs to 32- one extra parameter for each PDF and two for $s+\bar{s}$.

New LHC data fit.

Extremely high precision data on W, Z at 7 TeV from ATLAS, and high precision $W^{+/-}$ data and double differential $8\ \mathrm{TeV}\, Z$ data at 8 TeV.

CMS 8 TeV precise data on the $W^{+,-}$ rapidity distribution.

LHCb data at 7 and $8\ \mathrm{TeV}$ on W, Z rapidity distributions at higher rapidity.

$W + c\ \mathrm{jets}$ data at $7\ TeV$ from CMS.

ATLAS high mass Drell Yan data at 8 TeV.

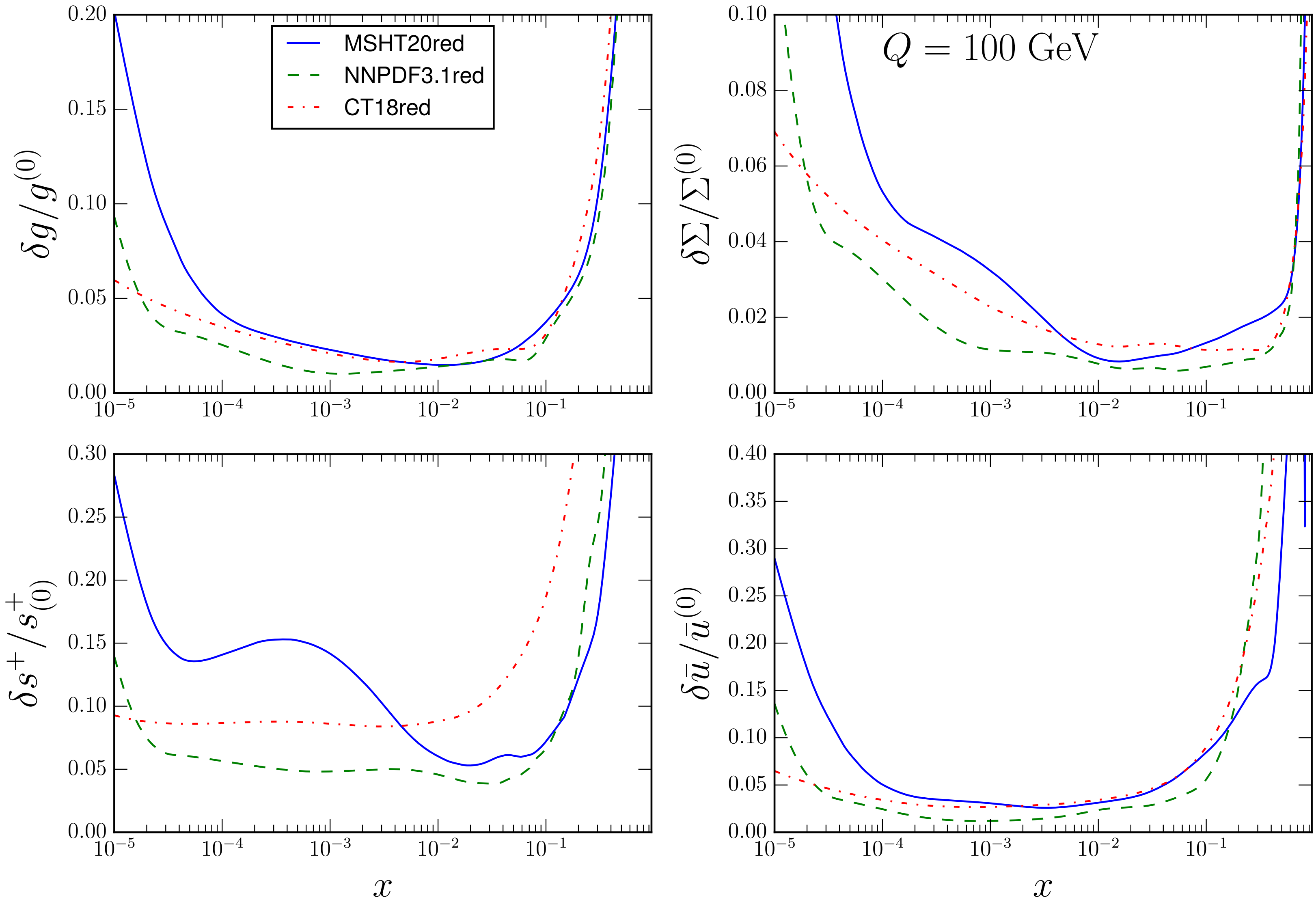
ATLAS data on $W^{+/-} +$ jets at 8 TeV.

$Z\ p_T$ distributions at 8 TeV.

New data on $\sigma_{t\bar{t}}$ at 8 TeV plus ATLAS single differential distributions in $p_{T,t}, M_{t\bar{t}}, y_t, y_{t\bar{t}}$ and CMS double differential distributions in $p_{T,t}, y_t$ both at 8 TeV.

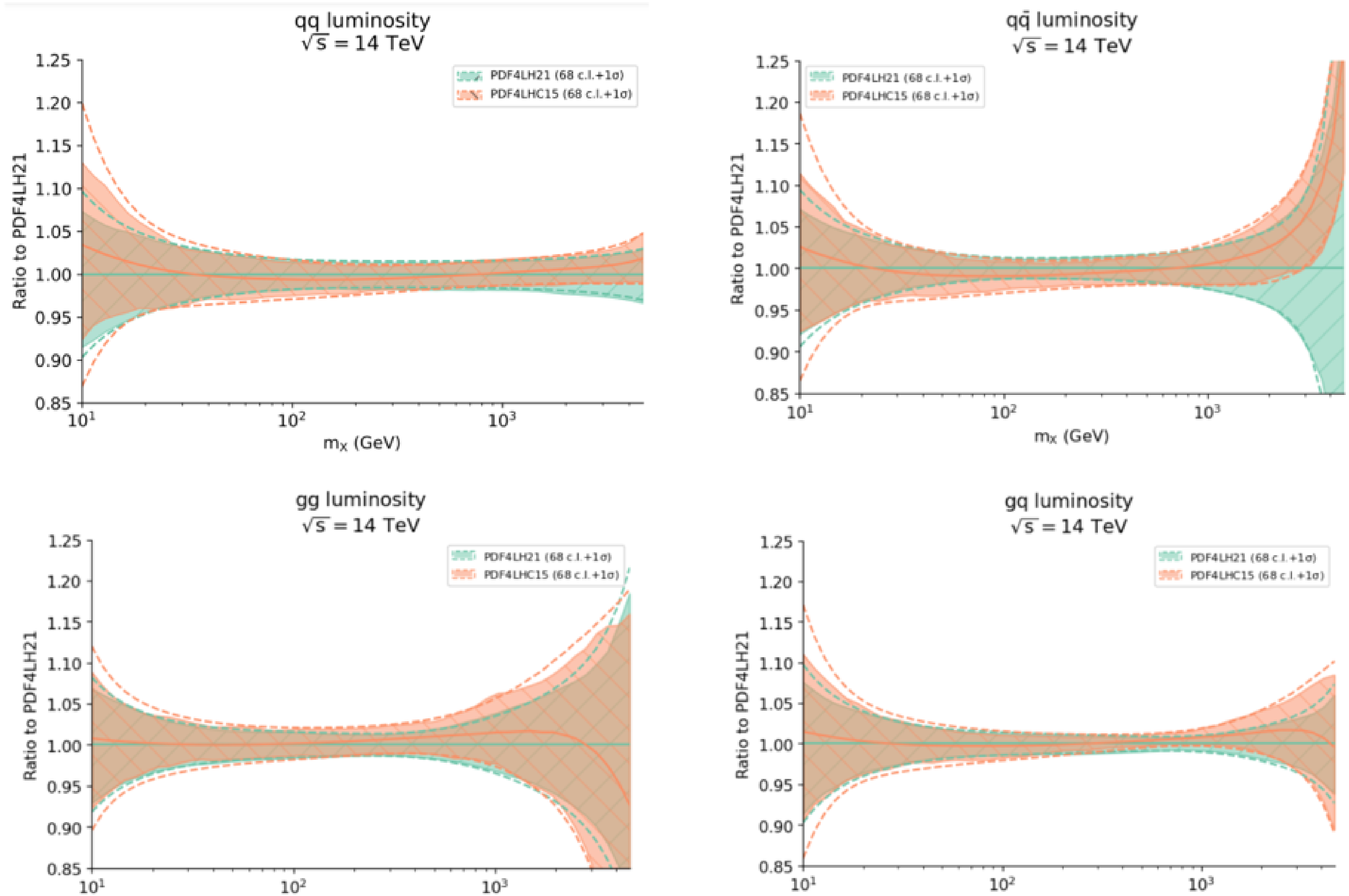
Inclusive jet data from ATLAS at 7 TeV and CMS at 2.76, 7 and 8 TeV.

PDF comparison in reduced fits



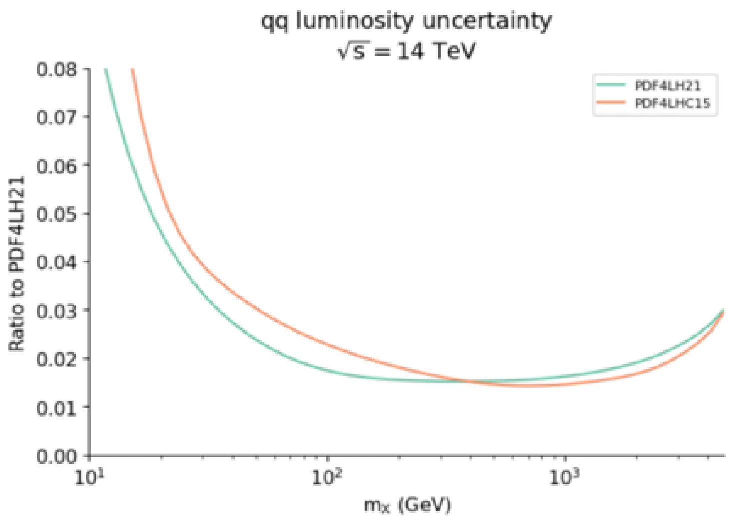
Plots from T. Cridge.

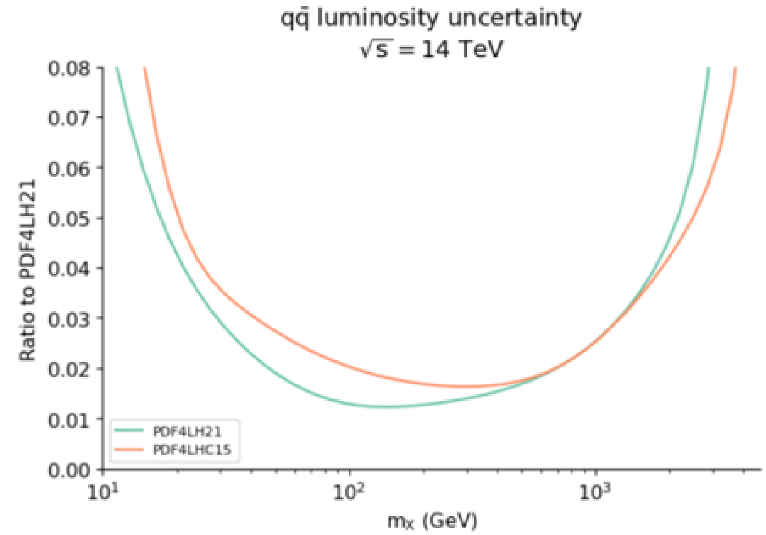
PDF4LHC15 vs PDF4LHC21

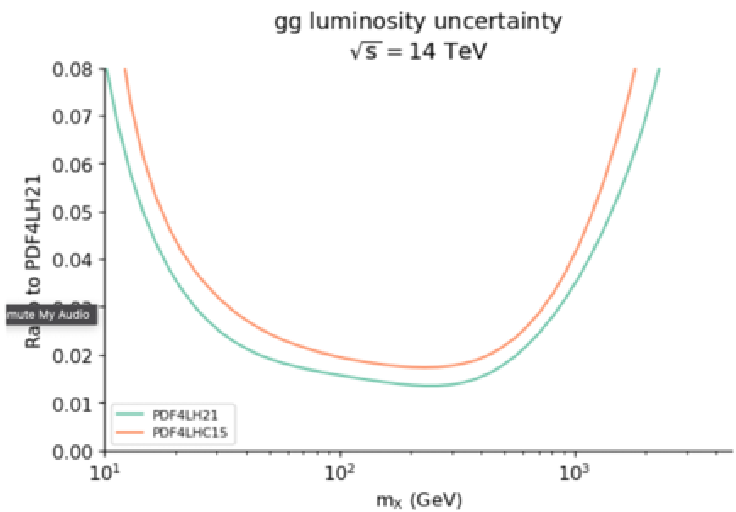


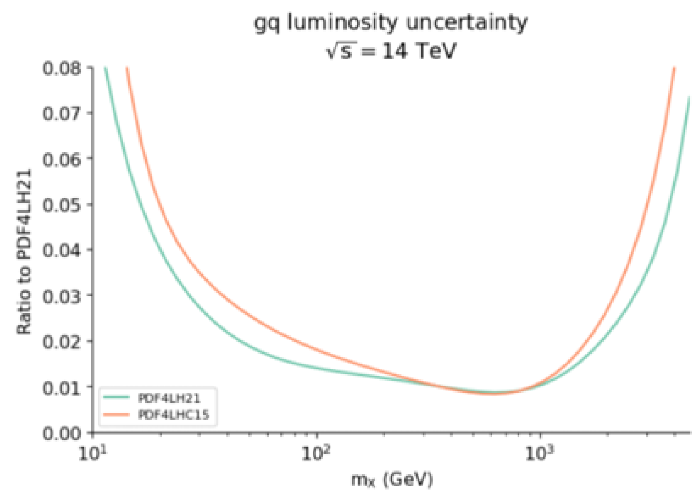
3

PDF4LHC15 vs PDF4LHC21





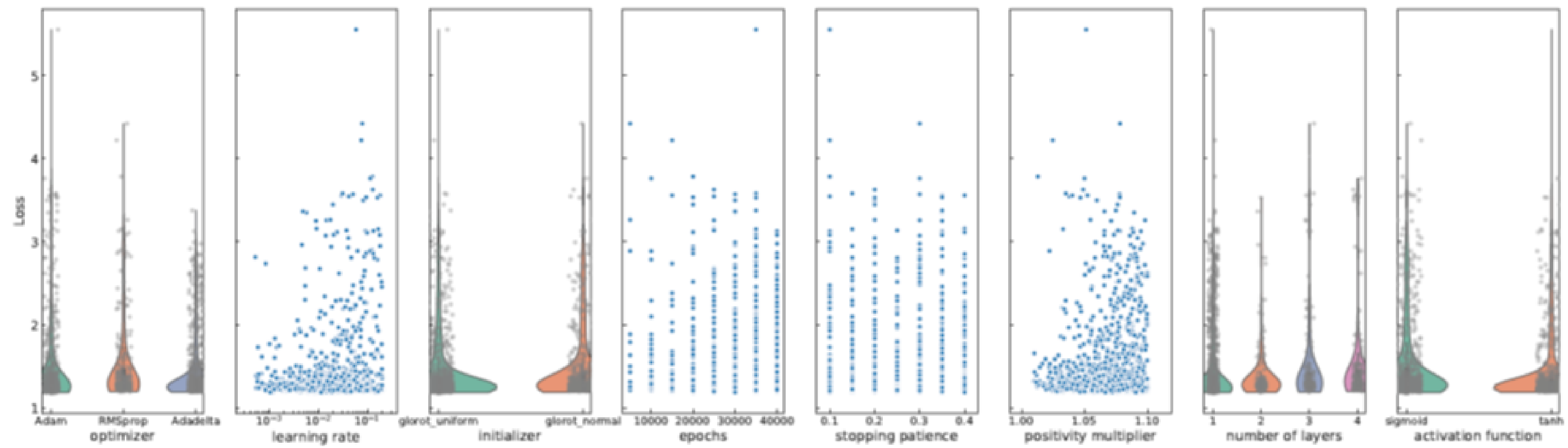




Automated model selection

NNPDF aims to minimize sources of bias in the PDF:

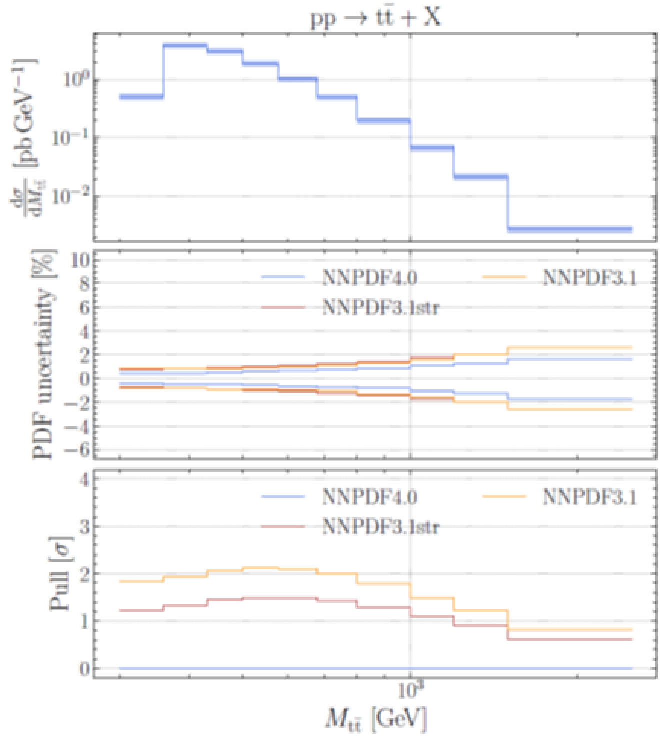
- Functional form → Neural Network
- Model parameters → **Hyperoptimization**

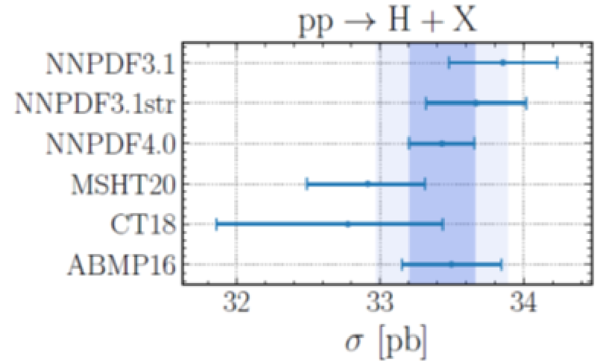


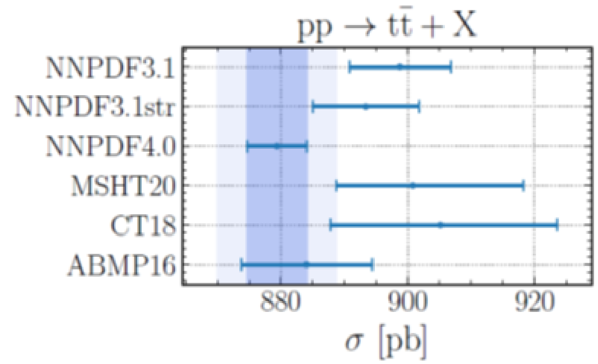
Scan over thousands of hyperparameter combinations and select the best one
k-fold cross-validation: used to define the reward function based on a **test dataset**

Implications for phenomenology

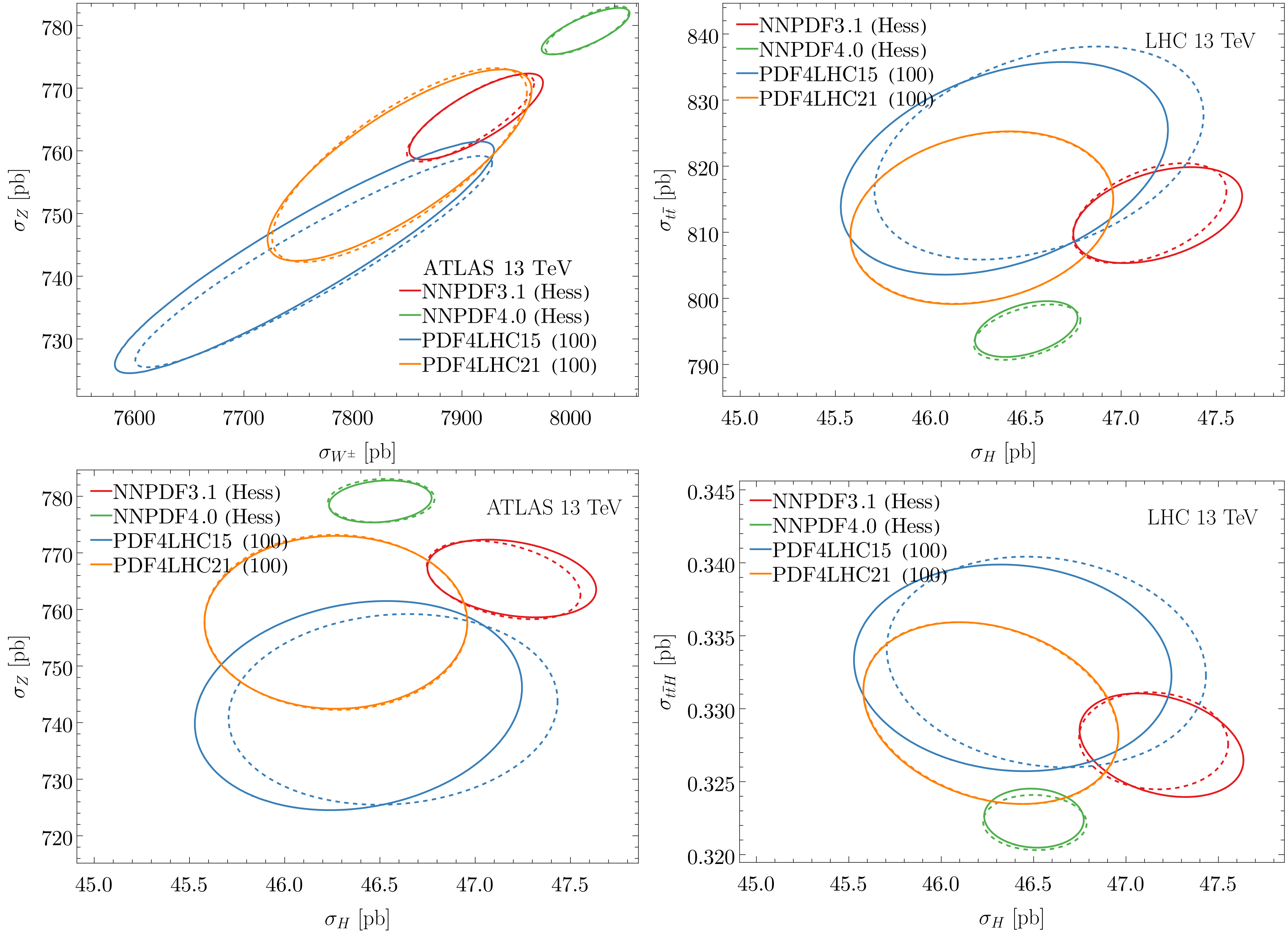
Reduced luminosity uncertainties → Reduced uncertainty at the level of observables







Correlation plots for PDF4LHCHessian sets and individual PDFs.



Plot by K. Xie - **preliminary**.