PDF4LHC21

J. Huston

Michigan State University

for PDF4LHC benchmarking group

arXiv:2203.05506

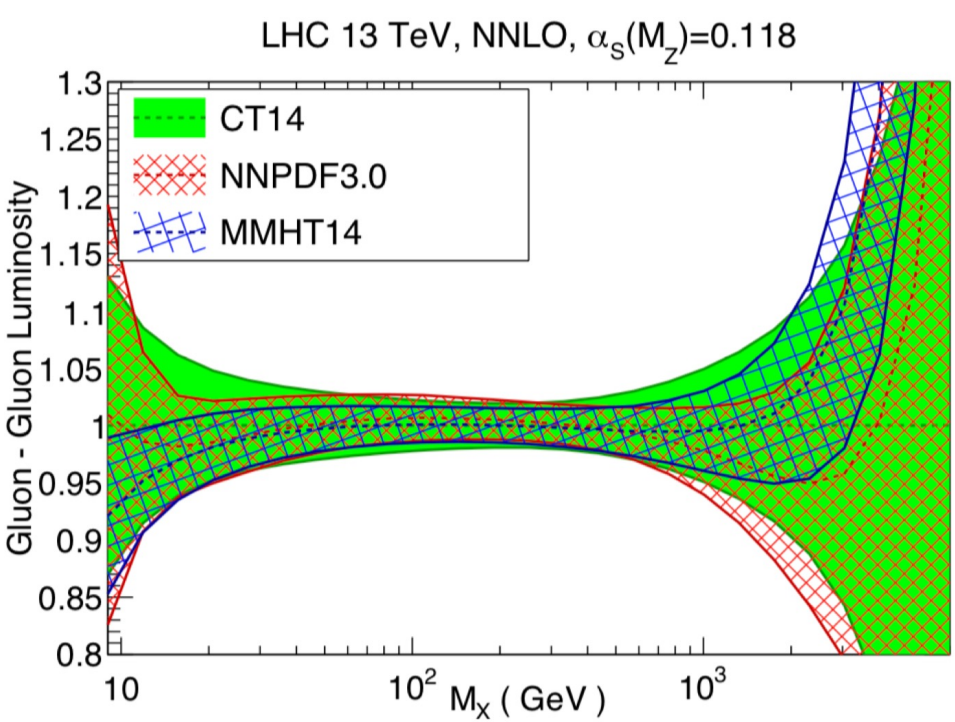
The PDF4LHC21 combination of global PDF fits for the LHC Run III

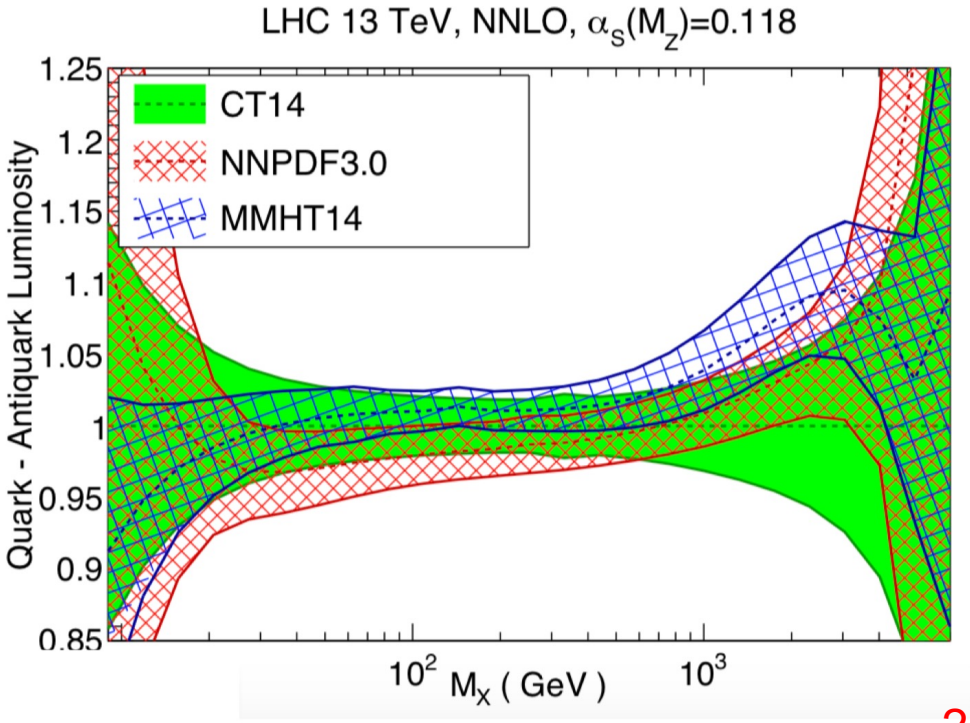
The **PDF4LHC Working Group:**

Richard D. Ball[1], Jon Butterworth[2], Amanda M. Cooper-Sarkar[3], Aurore Courtoy[4], Thomas Cridge[2], Albert De Roeck[5], Joel Feltesse[6], Stefano Forte[7], Francesco Giuli[5], Claire Gwenlan[2], Lucian A. Harland-Lang[8], T. J. Hobbs[9,10], Tie-Jiun Hou[11], Joey Huston[12], Ronan McNulty[13], Pavel M. Nadolsky[14], Emanuele R. Nocera[1], Tanjona R. Rabemananjara[15,16], Juan Rojo[15,16], Robert S. Thorne[2], Keping Xie[17], and C.-P. Yuan[12]

PDF4LHC15

- combination of CT14, MMHT2014, NNPDF3.0
  - 1 year benchmarking exercise comparison of above PDFs
  - 300 Monte Carlo replicas generated for each of the above PDFs
  - condensed to Hessian sets with from 30-100 members for distribution to users with central PDFs and error PDFs representing the three published PDFs
  - good (too good?) agreement for gluon-gluon luminosity->see qqbar





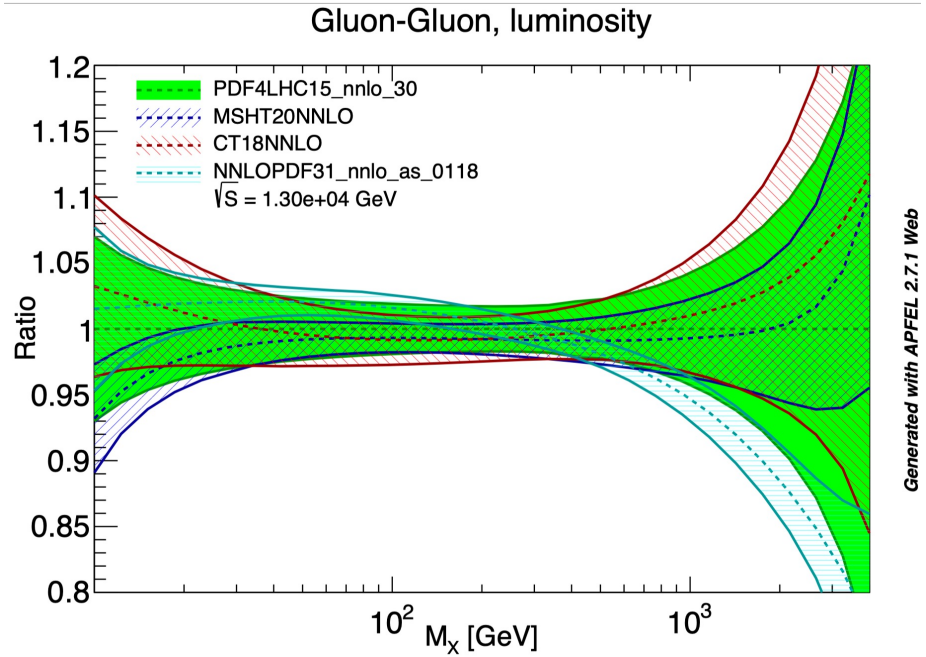
…in the meantime

- New critical data sets from the LHC on Drell-Yan, top, jets, W/Z+jets
- NNLO predictions available for all of above allowing this data to be included in PDF fits
- New NNLO PDFs available (CT18, MSHT20, NNPDF3.1*) that make use of this LHC data
  - additional technical improvements to the PDF fits
- These PDF sets (or rather a slight modification to CT18 (CT18') and a larger modification to NNPDF3.1 (NNPDF3.1') used for the construction of PDF4LHC21

*NNPDF4.0 came out too late to be included in the combination

PDF4LHC21

- new PDFs CT18, MSHT2020, NNPDF3.1, containing large amount of LHC data
- some new/different techniques, i.e. fitted charm for NNPDF3.1



consistency with PDF4LHC15, a bit more of a spread of the gg uncertainty bands than for the 2015 combination; some of gg fusion Higgs uncertainty will be due to spread of central values

- exercise: start with a reduced data set large enough to provide constraints, small enough that resulting PDFs should be similar for the different groups
  - add more data sets, ttbar, jets … and finally use the full fit PDFs

1.5 years in the making; many Friday meetings; many details in the paper left out of this talk

Aside: uncertainties

- PDF uncertainties depend first on the experimental uncertainties of the data
- Data from two measurements, or even from within the same measurement, can both be very precise, but the result of adding both to the PDF fit can be an increase in the PDF uncertainty (or more likely) a smaller decrease in uncertainty than expected) if the data are in tension with each other
- The resultant PDF uncertainty relies on the definition of a tolerance, i.e. what is a significant increase from the global minimum χ^2, i.e. PDF uncertainty can be adjusted by changing the tolerance
- $\Delta\chi^2=1$ is not applicable for ~4000 data points from different experiments
- NB: CT (Tier 2) and MSHT (dynamic tolerance) have introduced criteria to restrict the pull of data sets that disagree with global fit
- NB: all groups see tensions; the relevant χ^2 values show that the fits are not good

Reduced data set fits

- Diverse enough to provide information for all PDFs
- Sparse enough that uncertainties should be very similar for all 3 PDFs
- Origins of differences of PDFs
  - due to variations of experimental input, treatment of systematic errors, different theory choices, fitting methodologies?
  - so for benchmarking, use common theory settings (i.e. perturbative charm, m_{charm}=1.4 GeV, s=sbar at input scale, $\alpha_s(m_Z)$=0.118, positive-definite PDFs, no deuteron or nuclear corrections…)

Dataset	Reference	Dataset	Reference
BCDMS proton, deuteron DIS	[155, 156]	LHCb 8 TeV $Z \to ee$	[62]
NMC deuteron to proton ratio DIS	[157]	ATLAS 7 TeV high precision W, Z (2016)	[63]
NuTeV νN dimuon	[158]	D0 Z rapidity	[159]
HERA I+II inclusive DIS	[60]	CMS 7 TeV electron asymmetry	[160]
E866 Drell-Yan ratio pd/pp DIS	[161]	ATLAS 7 TeV W, Z rapidity (2011)	[149]
LHCb 7, 8 TeV W, Z rapidity	[61, 65]	CMS 8 TeV inclusive jet	[69]

Table 3.1. The measurements included in the initial round of PDF fits to a reduced dataset, together with the corresponding publication reference. This dataset is chosen as the largest subset of data fit by CT18, MHST20, and NNPDF3.1 in an (almost) identical manner.

Reduced fits

- Central values agree reasonably well
- …as do uncertainties at higher x
- There are some differences, for example at low x for the gluon distribution; this is a region nominally not well constrained by data

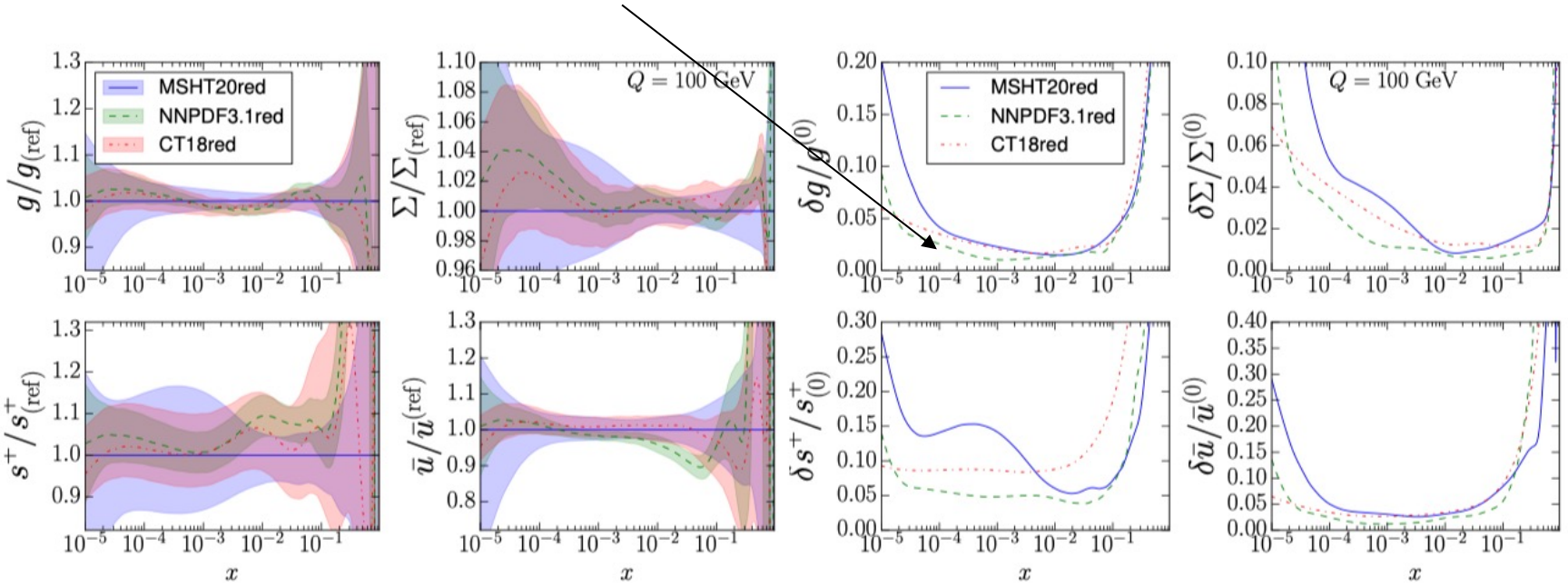


Figure 3.4. Comparison between the reduced PDF fits from the three groups, in the same format as in Fig. 3.1. For the three groups, PDF errors correspond to 1σ intervals. In the left panels, PDFs are displayed normalised to the central value of the MSHT20 reduced PDF set.

PDF luminosities: no rapidity restriction

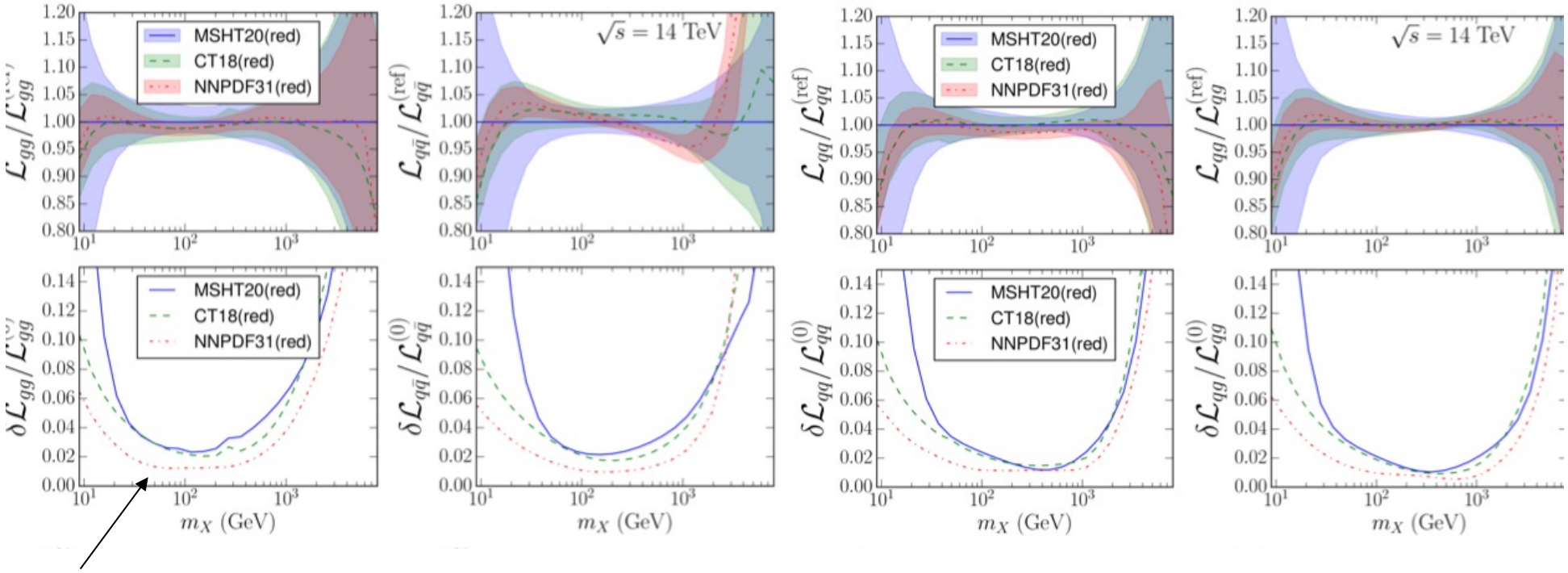


Figure 3.5. Comparison of the partonic luminosities between the CT18, MSHT20, and NNPDF3.1 reduced fits at $\sqrt{s} = 14$ TeV as a function of the invariant mass of the produced final state m_X. From left to right we show the gluon-gluon, quark-antiquark, quark-quark and quark-gluon luminosities, normalised to the central value of the MSHT20 prediction, together with the associated 1σ relative PDF uncertainties. The upper panels display the luminosities evaluated without any restriction on the final-state rapidity y_X, while the bottom panels instead account for a rapidity cut of $|y_X| < 2.5$ which restricts the produced final state to lie within the ATLAS/CMS central acceptance region.

$|y_X<2.5|$

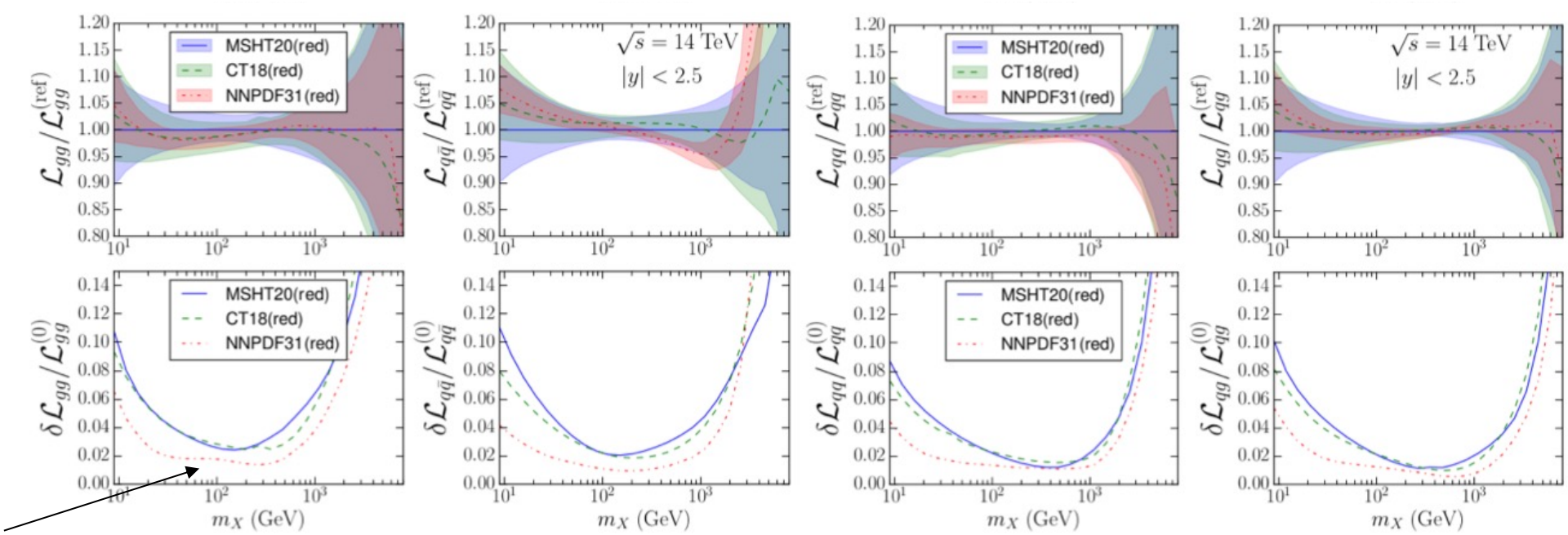


Figure 3.5. Comparison of the partonic luminosities between the CT18, MSHT20, and NNPDF3.1 reduced fits at $\sqrt{s} = 14$ TeV as a function of the invariant mass of the produced final state m_X. From left to right we show the gluon-gluon, quark-antiquark, quark-quark and quark-gluon luminosities, normalised to the central value of the MSHT20 prediction, together with the associated 1σ relative PDF uncertainties. The upper panels display the luminosities evaluated without any restriction on the final-state rapidity y_X, while the bottom panels instead account for a rapidity cut of $|y_X| < 2.5$ which restricts the produced final state to lie within the ATLAS/CMS central acceptance region.

CT18

Experimental data set E		$N_{\rm pt}$	$\chi^2/N_{\rm pt}$	S
LHCb 7 TeV 1.0 fb^{-1} W/Z forward rapidity	[61]	33	1.63 (1.21)	2.3 (0.9)
LHCb 8 TeV 2.0 fb^{-1} $Z \to e^-e^+$ forward rapidity	[62]	17	1.04 (1.06)	0.2 (0.3)
ATLAS 7 TeV 4.6 fb^{-1}, W/Z combined‡	[63]	34	8.45 (2.61)	16 (5.1)
CMS 8 TeV 18.8 fb^{-1} muon charge asymmetry A_{ch}	[64]	11	1.04 (1.10)	0.2 (0.3)
LHCb 8 TeV 2.0 fb^{-1} W/Z cross sec.	[65]	34	2.17 (1.75)	4.0 (2.7)
ATLAS 8 TeV 20.3 fb^{-1}, Z p_T cross sec.	[66]	27	1.12 (1.05)	0.5 (0.2)
CMS 7 TeV 5 fb^{-1}, single incl. jets, $R = 0.7$	[67]	158	1.23 (1.19)	2.0 (1.7)
ATLAS 7 TeV 4.5 fb^{-1}, single incl. jets, $R = 0.6$	[68]	140	1.45 (1.45)	3.4 (3.4)
CMS 8 TeV 19.7 fb^{-1}, single incl. jets, $R = 0.7$, (extended)	[69]	185	1.14 (1.12)	1.3 (1.2)
CMS 8 TeV 19.7 fb^{-1}, $t\bar{t}$ norm. double-diff. top p_T and y	[70]	16	1.18 (1.19)	0.6 (0.6)
ATLAS 8 TeV 20.3 fb^{-1}, $t\bar{t}$ p_T^t and $m_{t\bar{t}}$ abs. spectrum	[71]	15	0.63 (0.71)	-1.1 (-0.8)

Table 2.1. Numbers of points, χ^2/N_{pt}, and the effective Gaussian variables for the newly added LHC measurements in the CT18 and CT18Z NNLO fits. The ATLAS 7 TeV W/Z data (4.6 fb^{-1}), labelled by ‡, are included in the CT18A and CT18Z global fits, but not in CT18 and CT18X.

Spartyness, a variable that describes the goodness of fit, taking into account the number data points; expect S to be in the range of -1 to 1.

If S>>1, that means the data is poorly fit; if S<<1, that means the fit is too good, and possibly the errors are overestimated

$$S_E = \sqrt{2\chi^2_E} - \sqrt{2N_{\rm pt,E} - 1}$$

MSHT20

Experimental data set	N_{pt}	$\chi^2/N_{\rm pt}$	S
D0 W asymmetry [106]	14	0.86	-0.3
$\sigma_{t\bar{t}}$ Tevatron +CMS+ATLAS 7, 8 TeV [107]- [108]	17	0.85	-0.4
LHCb 7+8 TeV $W + Z$ [61, 62]	67	1.48	2.6
LHCb 8 TeV e [65]	17	1.54	1.5
CMS 8 TeV W [64]	22	0.58	-1.5
ATLAS 7 TeV jets $R = 0.6$ [68]	140	1.59	4.4
CMS 7 TeV $W + c$ [102]	10	0.86	-0.2
ATLAS 7 TeV W, Z [63]	61	1.91	4.3
CMS 7 TeV jets $R = 0.7$ [67]	158	1.11	1.0
ATLAS 8 TeV Zp_T [66]	104	1.81	5.0
CMS 8 TeV jets [69]	174	1.50	4.2
ATLAS 8 TeV $t\bar{t} \to l + $j single-diff [71]	25	1.02	0.1
ATLAS 8 TeV $t\bar{t} \to l^+l^-$ single-diff [109]	5	0.68	-0.4
ATLAS 8 TeV high-mass Drell-Yan [110]	48	1.18	0.9
ATLAS 8 TeV $W^{+,-}$ + jet [111]	32	0.60	-1.7
CMS 8 TeV $(d\sigma_{t\bar{t}}/dp_{T,t}dy_t)/\sigma_{t\bar{t}}$ [70]	15	1.50	1.3
ATLAS 8 TeV W^+, W^- [100]	22	2.61	4.2
CMS 2.76 TeV jets [112]	81	1.27	1.7
CMS 8 TeV $t\bar{t}$ y_t distribution [113]	9	1.47	1.0
ATLAS 8 TeV double differential Z [99]	59	1.45	2.3

Note the trouble fitting the ATLAS W/Z data

Table 2.2. Numbers of points, fit qualities $\chi^2/N_{\rm pt}$ and S values for new collider data added to the NNLO MSHT20 fit.

Definitions for CT18'/NNPDF3.1'

- CT18->CT18': m_c=1.4 GeV,m_b=4.75 GeV
- NNPDF3.1->NNPDF3.1': same as above plus some additions to the data set (in some ways NNPDF3.1' is a transition from 3.1-> 4.0)
- No MSHT20' since the above are the heavy quark mass values they normally use

	NNPDF3.1 [15]			NNPDF3.1′		
Experimental data set	$N_{\rm pt}$	$\chi^2/N_{\rm pt}$	S	$N_{\rm pt}$	$\chi^2/N_{\rm pt}$	S
D0 W electron asymmetry [121]	8	2.70	+2.70	11	3.07	+3.64
D0 W muon asymmetry [122]	9	1.56	+1.18	9	1.58	+1.21
ATLAS low-mass DY 7 TeV [123]	6	0.90	−0.03	6	0.89	−0.05
ATLAS W,Z 7 TeV [63]	34	2.14	+3.88	61	1.99	+4.58
ATLAS Z p_T 8 TeV (p_T, $m_{\ell\ell}$) [66]	44	0.93	−0.28	44	0.94	−0.23
ATLAS Z p_T 8 TeV (p_T, y_Z) [66]	48	0.94	−0.25	48	0.95	−0.20
ATLAS single-inclusive jets 7 TeV ($R = 0.6$) [68]	31	1.07	+0.33	140	1.25	+2.00
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8, 13 TeV [124,125]	3	0.86	+0.04	3	0.95	+0.15
ATLAS $t\bar{t}$ ℓ+jets 8 TeV ($1/\sigma\ d\sigma/dy_t$) [71]	9	1.45	+0.99	4	3.56	+2.69
CMS W rapidity 8 TeV [64]	22	1.01	+0.11	22	1.03	+0.17
CMS Z p_T 8 TeV [126]	28	1.32	+1.18	28	1.34	+1.25
CMS single-inclusive jets 2.76 TeV [112]	81	1.03	+0.23	—	—	—
CMS single-inclusive jets 8 TeV [69]	—	—	—	185	1.30	+2.72
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8, 13 TeV [127,128]	3	0.20	−1.14	3	0.18	−1.20
CMS $t\bar{t}$ ℓ+jets 8 TeV ($1/\sigma\ d\sigma/dy_{t\bar{t}}$) [113]	9	0.94	−0.01	9	1.67	+1.36
CMS $t\bar{t}$ 2D 2ℓ 8 TeV ($1/\sigma\ {\rm d}\sigma/{\rm d}y_t dm_{t\bar{t}}$) [70]	—	—	—	16	0.81	−0.48
LHCb $W,Z \to \mu$ 7 TeV [61]	29	1.76	+1.55	29	1.96	+3.11
LHCb $W,Z \to \mu$ 8 TeV [65]	30	1.37	+1.39	30	1.36	+1.35

Note the trouble fitting the ATLAS W/Z data

important addition

Table 2.3. The numbers of points, $\chi^2/N_{\rm pt}$ and S values for new collider data in the NNPDF3.1 fit [15] and in the NNPDF3.1′ fit variant adopted in the present combination. The Tevatron and LHC data sets already included in NNPDF3.0 are kept in NNPDF3.1, but not necessarily in NNPDF3.1′. These are not indicated in the table. Note that, despite the number of LHC data points is larger in NNPDF3.1′ than that in NNPDF3.1, the total number of data points in the two analyses is similar, mainly because the Tevatron single-inclusive jet measurements (not indicated in the table) are no longer included in NNPDF3.1′. See text for details.

Combination

- Generate 300 MC replicas of each of the 3 PDFs and combine

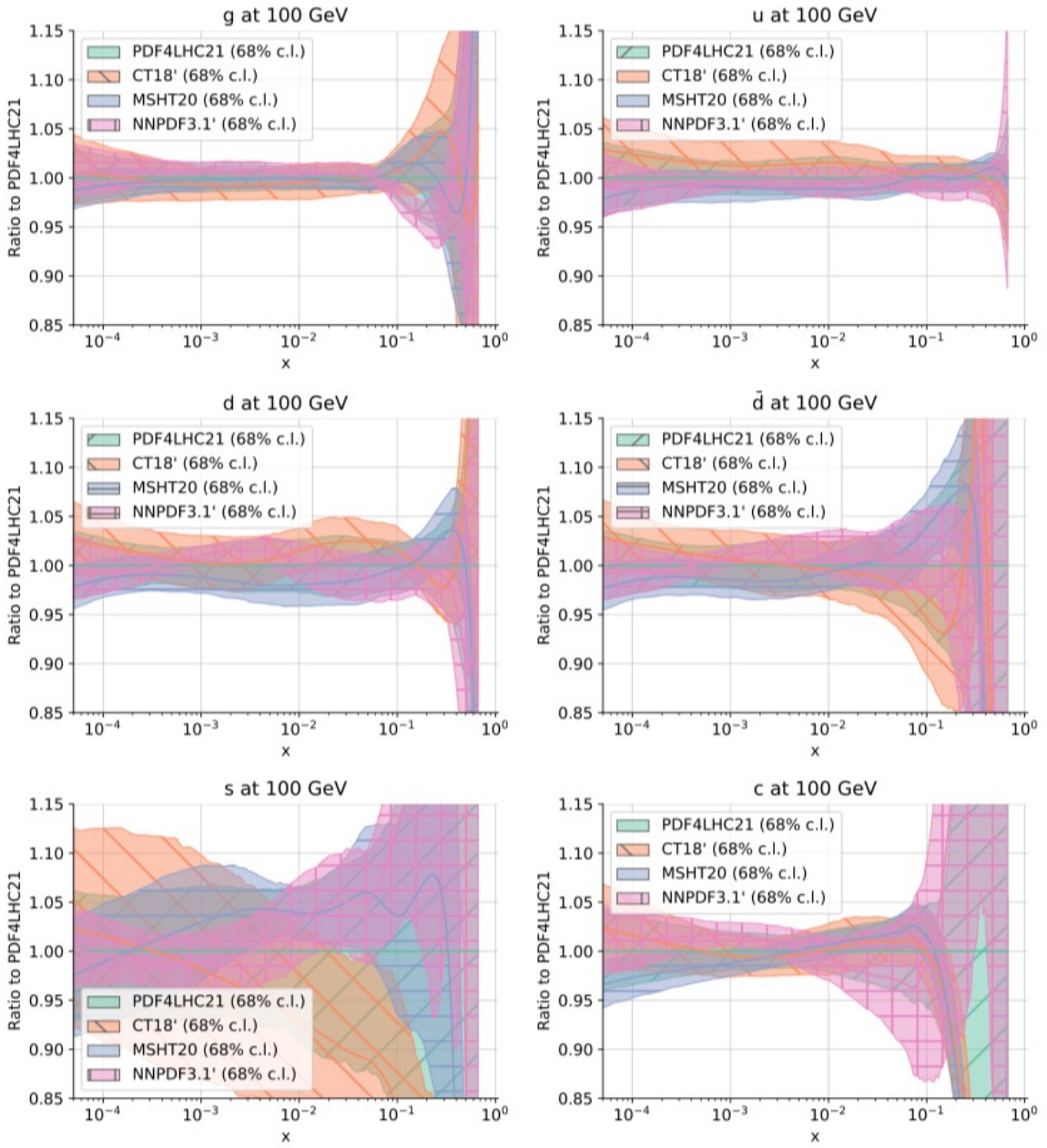


Figure 4.1. Comparison of the PDF4LHC21 combination (composed by $N_{\rm rep} = 900$ replicas) with the three constituent sets at $Q = 100$ GeV, normalised to the central value of the former and with their respective 68%CL uncertainty bands. In the case of the Hessian sets (CT18′ and MSHT20) we display their Monte Carlo representation composed by $N_{\rm rep} = 300$ replicas generated according to Eq. (4.3). The NNPDF3.1′ band is also constituted by $N_{\rm rep} = 300$ (native) replicas.

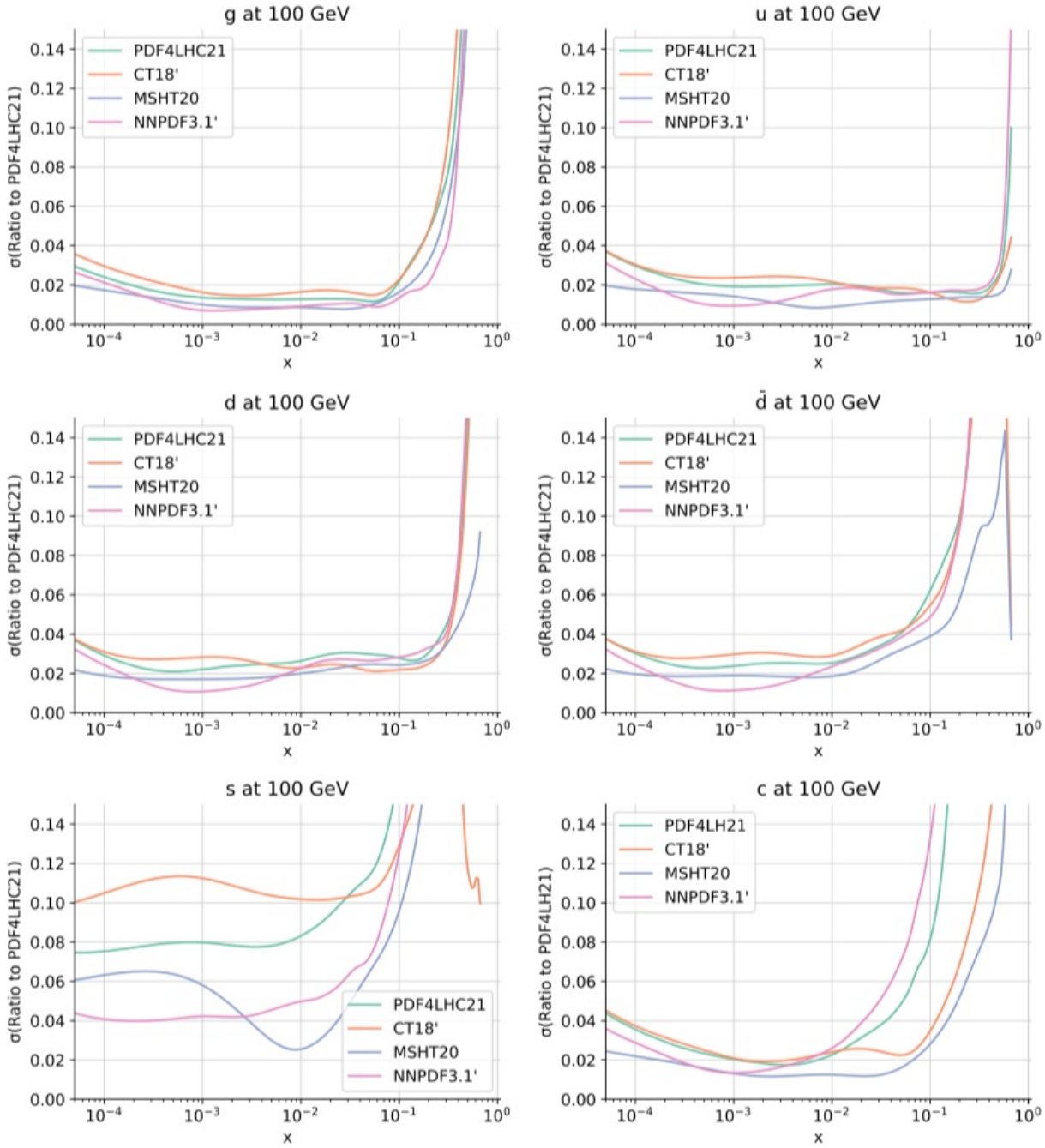


Figure 4.2. Same as Fig. 4.1 now showing the relative PDF 68% CL uncertainties (normalised to the PDF4LHC21 central value) of the four PDF sets.

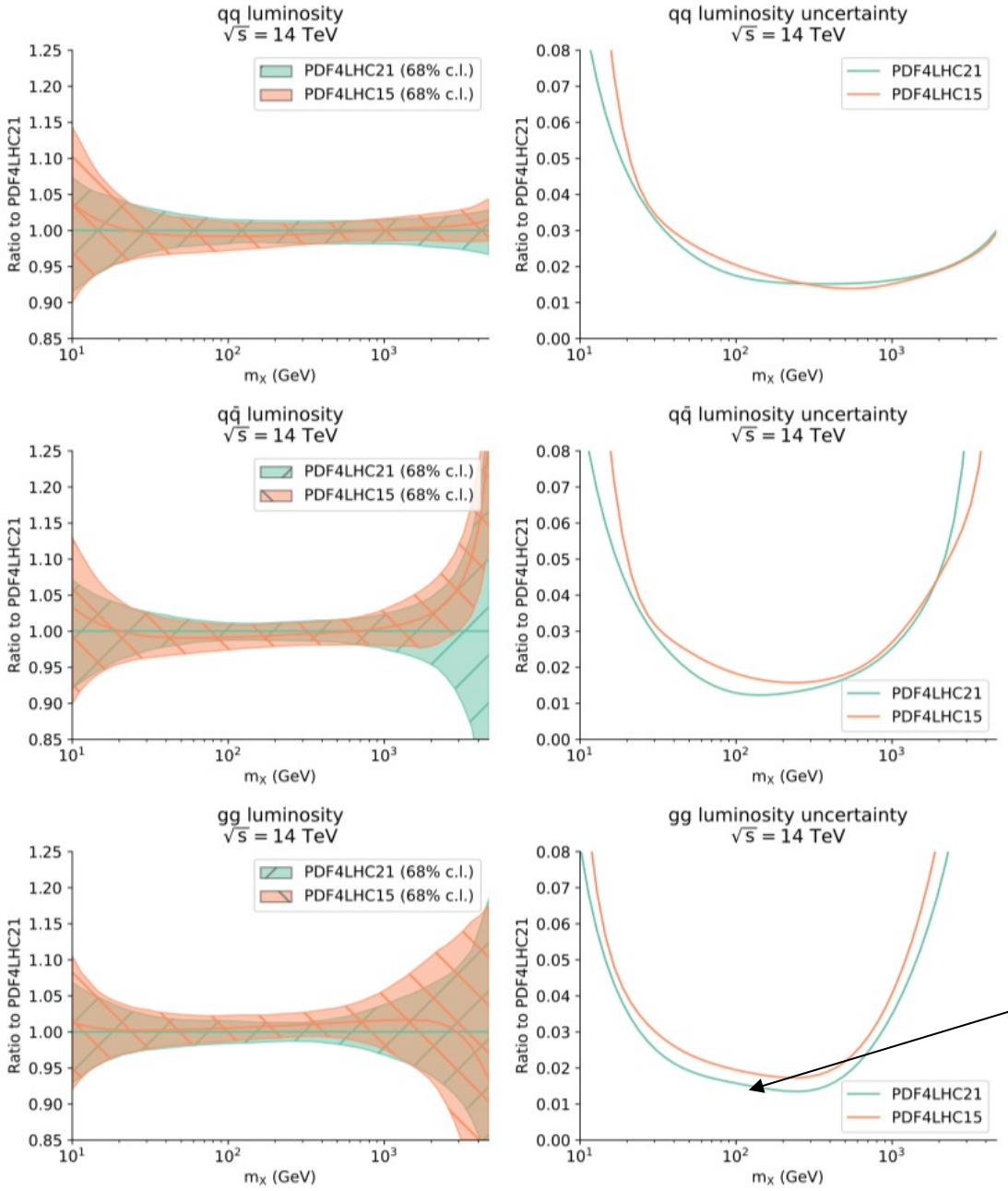


Figure 4.11. Comparison of the partonic luminosities at $\sqrt{s} = 14$ TeV between PDF4LHC15 and PDF4LHC21. In both cases, the original sets with $N_{\rm rep} = 900$ have been used. Results are shown for the quark-quark, quark-antiquark, and gluon-gluon luminosities as a function of the final state invariant mass m_X, and normalised to the central value of the PDF4LHC21 prediction. The right panels display the corresponding 68% CL relative uncertainties.

reduction in uncertainty for gg fusion

_mc and _40

- For distribution, compress 900 MC replicas to 100 replica set and to 40 member Hessian set using same techniques as PDF4LHC15

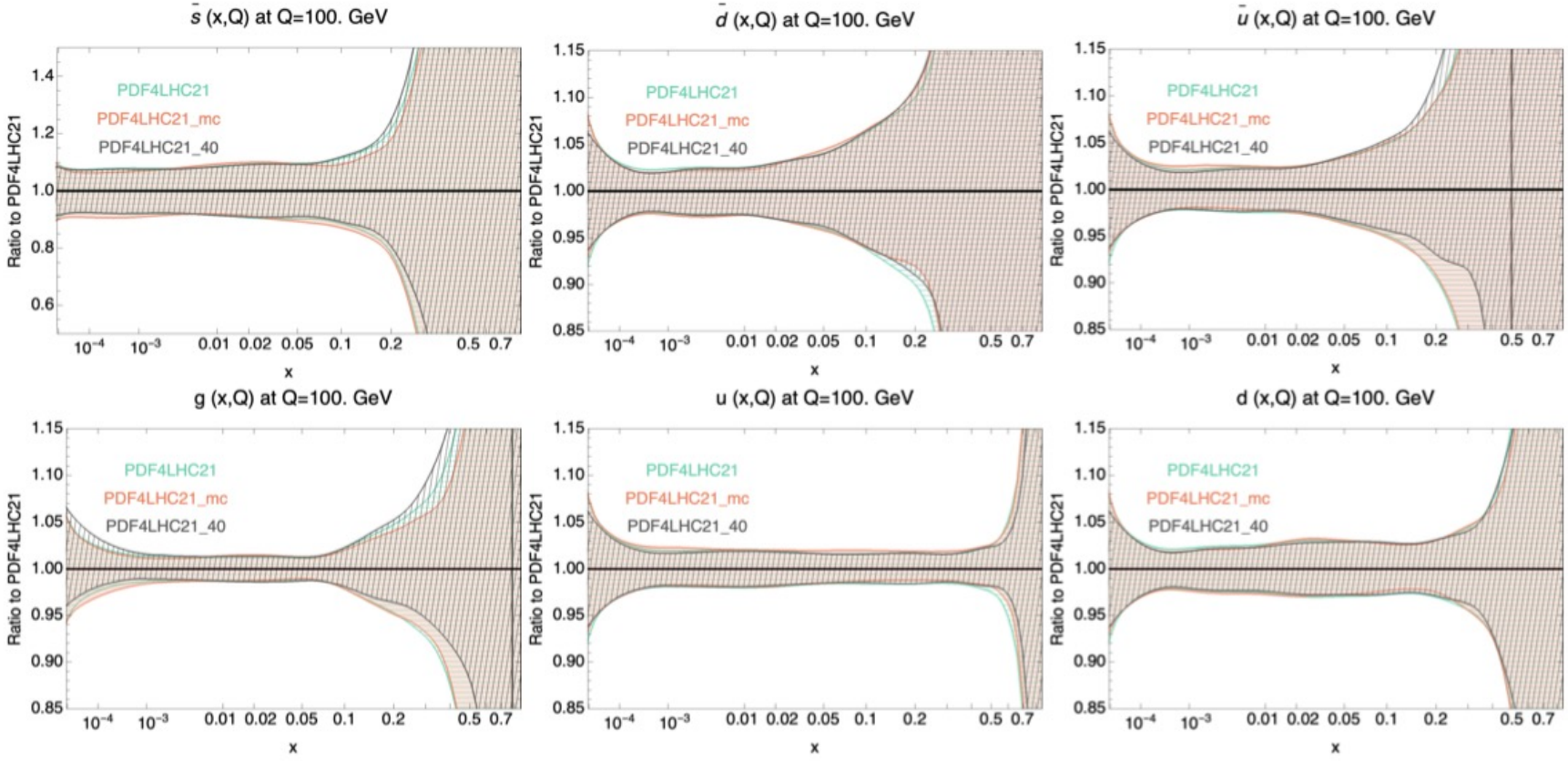


Figure 4.7. Comparison of the $N_{\rm rep} = 900$ PDF4LHC21 set with the corresponding MC compressed set, PDF4LHC21_MC, and its Hessian representation with $N_{\rm eig} = 40$ eigenvectors, PDF4LHC21_40. Results are displayed normalised to the average PDF of the $N_{\rm rep} = 900$ set. We show the results for the anti-strange, anti-down, anti-up, gluon and the up, down quark PDFs at $Q = 100$ GeV.

- Note that the Hessian 40 member set is positive definite, while the MC 100 set can go negative at high x/mass

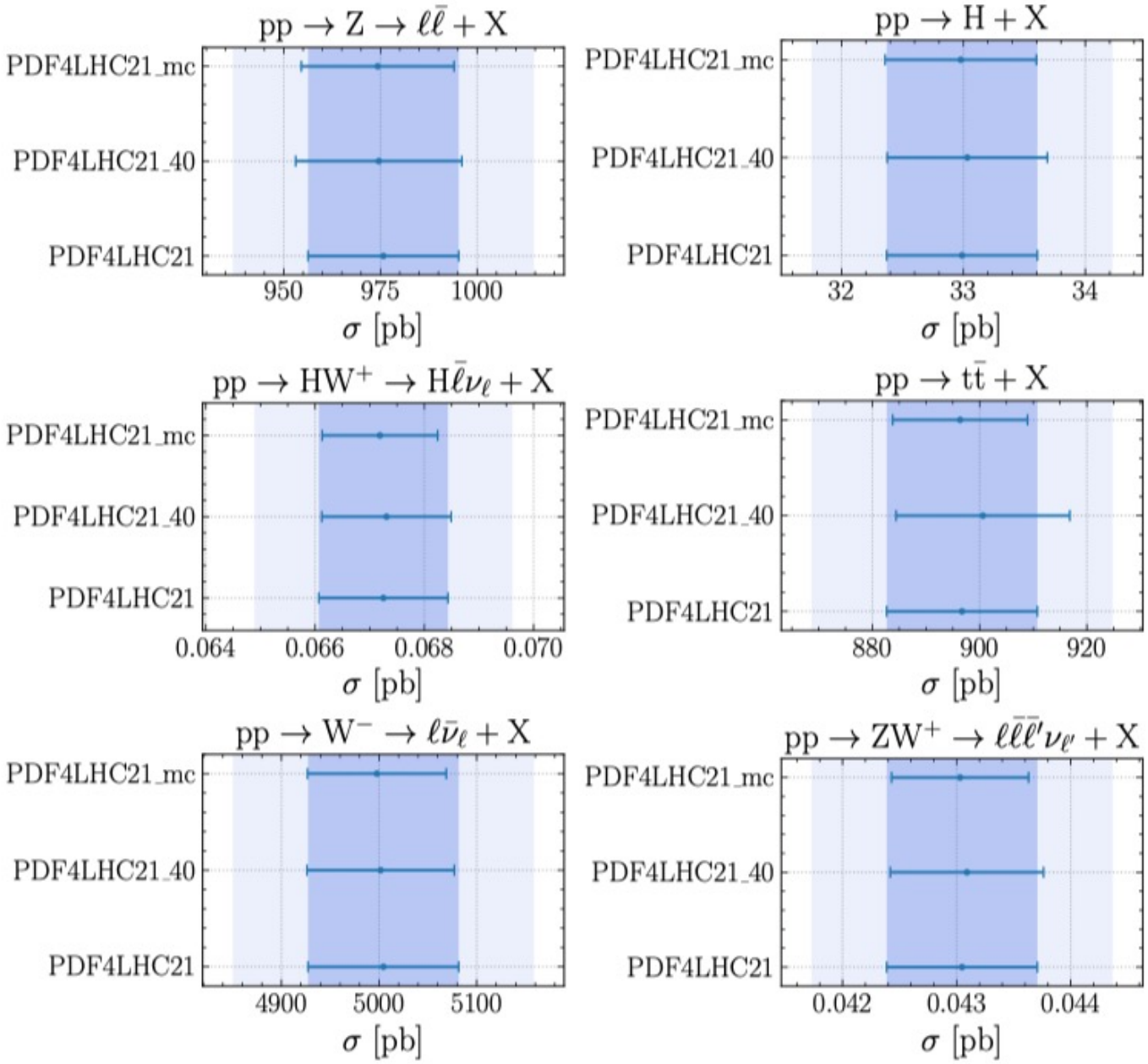


Figure 5.1. The fiducial cross-sections for representative LHC processes at $\sqrt{s} = 14$ TeV, comparing the predictions based on PDF4LHC21 ($N_{\rm rep} = 900$) with the corresponding Monte Carlo compressed set ($N_{\rm rep} = 100$) and the Hessian reduction ($N_{\rm eig} = 40$) using the META-PDF approach. The darker (lighter) band indicates the 68% CL (95% CL) uncertainties associated to the original PDF4LHC21 combination. See text for more details.

- It can be useful to look at 2-D ellipses comparing cross sections

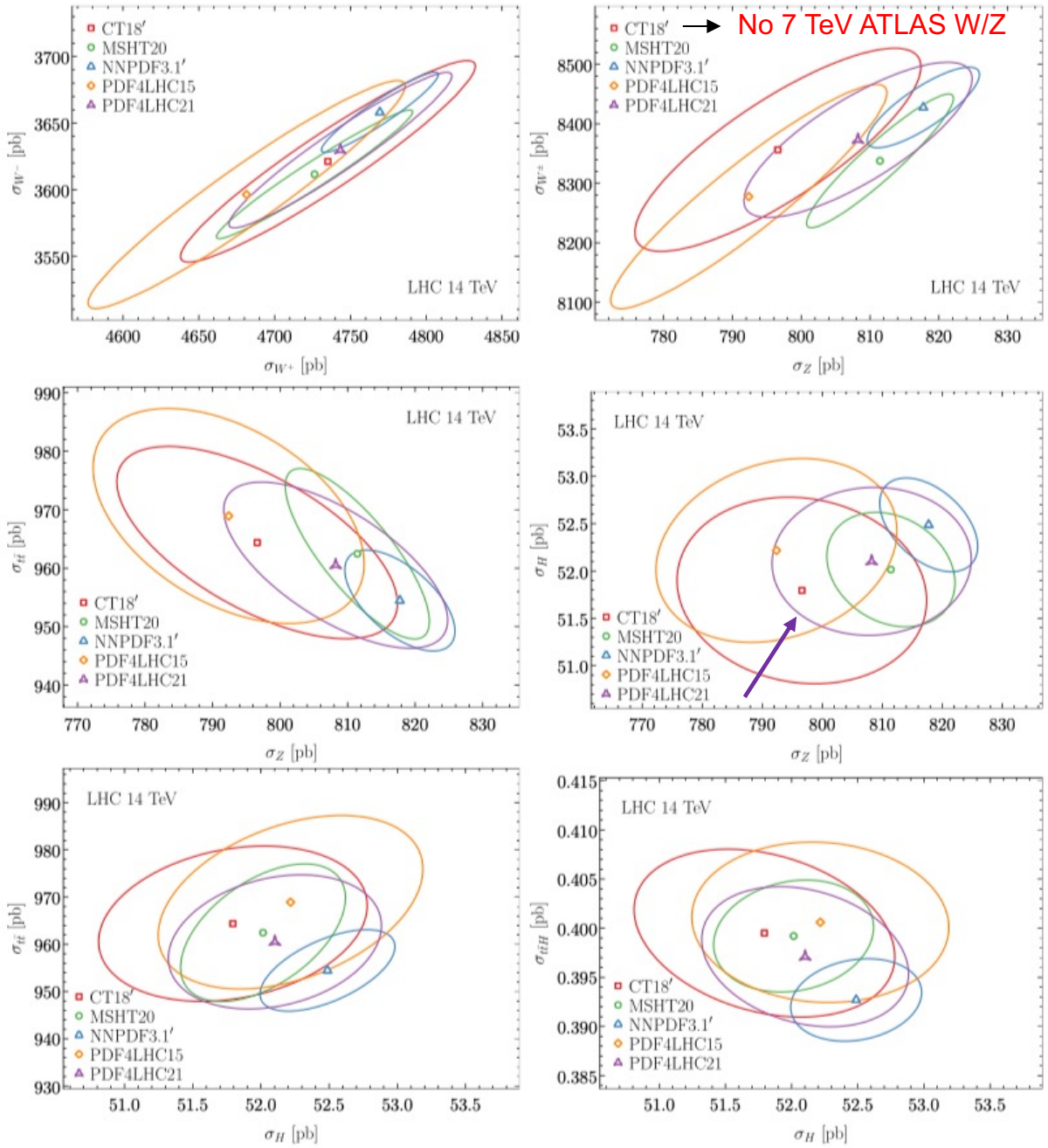


Figure 5.2. The 1σ ellipses for pairs of inclusive cross sections among $W^{\pm}, Z, t\bar{t}, H, t\bar{t}H$ production at the LHC 14 TeV. The $W^{\pm}/Z$ cross sections are defined in the ATLAS 13 TeV fiducial volume [170], while others correspond to the full phase space. See text for details of the theory calculations.

The white smoke has dissipated.
The PDFs have been sent to LHAPDF.

LHAPDF6 grid name	Pert. order	$n_f^{\max}$	ErrorType	N_{mem}	$\alpha_s(m_Z^2)$
PDF4LHC21_mc	NNLO	5	replicas	100	0.118
PDF4LHC21_40	NNLO	5	symmhessian	40	0.118
PDF4LHC21_mc_pdfas	NNLO	5	replicas+as	102	mem 0:100 → 0.118 mem 101 → 0.117 mem 102 → 0.119
PDF4LHC21_40_pdfas	NNLO	5	symmhessian+as	42	mem 0:40 → 0.118 mem 41 → 0.117 mem 42 → 0.119
PDF4LHC21_mc_nf4	NNLO	4	replicas	100	0.118
PDF4LHC21_40_nf4	NNLO	4	symmhessian	40	0.118
PDF4LHC21_mc_pdfas_nf4	NNLO	4	replicas+as	102	mem 0:100 → 0.118 mem 101 → 0.117 mem 102 → 0.119
PDF4LHC21_40_pdfas_nf4	NNLO	4	symmhessian+as	42	mem 0:40 → 0.118 mem 41 → 0.117 mem 42 → 0.119

Table 6.1. Summary of the combined PDF4LHC21 sets presented in this work and made available from LHAPDF6. All these sets are based on NNLO QCD theory and a variable-flavour-number scheme with either $n_f^{\max} = 5$ or $n_f^{\max} = 4$ as maximum number of active flavours. Recall that in the LHAPDF6 grids there is always a zeroth member, so that the total number of PDF members in a given set is always $N_{\text{mem}} + 1$. See text for usage instructions.

…recently added to LHAPDF

93000	PDF4LHC21_mc	(dir)	(tar.gz)	(info)
93100	PDF4LHC21_40	(dir)	(tar.gz)	(info)
93200	PDF4LHC21_mc_pdfas	(dir)	(tar.gz)	(info)
93300	PDF4LHC21_40_pdfas	(dir)	(tar.gz)	(info)
93400	PDF4LHC21_mc_nf4	(dir)	(tar.gz)	(info)
93500	PDF4LHC21_40_nf4	(dir)	(tar.gz)	(info)
93600	PDF4LHC21_mc_pdfas_nf4	(dir)	(tar.gz)	(info)
93700	PDF4LHC21_40_pdfas_nf4	(dir)	(tar.gz)	(info)

Summary

- The PDFs are now available on LHAPDF
  - they can be quite useful, just as those from PDF4LHC15 were, for example in Monte Carlo generation (for Run 3); for comparison to data, the individual PDFs are most often the best to use
- The paper has been submitted to Journal of Physics G
- We will next turn our attention (back) to a followup paper, trying to dive in a bit deeper into a better understanding of how each PDF fit works, what the relative sensitivities are to different data sets, etc
  - see also talk by Pavel Nadolsky at the APS meeting yesterday
  - https://smu.box.com/s/an6abaoh8pbvd7z8vcmbz2h4ye23dc50

Extras

High x gluon

- Of great interest for both SM physics and searches
- 3 main datasets sensitive to high x gluon: jet data, top data, Z p_T data
- Tensions between data sets, tensions within data sets
  - correlated systematics important
- Consider ATLAS 8 TeV top datasets: m_{tt},y_t,y_{tt},p_T^{top}
  - MSHT, CT, ATLAS cannot get good fit to all correlated distributions together, or to y_t, y_{tt} separately, in either reduced fit or full global fit
  - NNPDF able to fit rapidity distributions if systematics for each observable are de-correlated; for correlated case find same large chisquares as the other PDFs
- Theory predictions check out, i.e. common theory used by all groups
- NB: top data sets have a low number of data points; NNPDF cannot divide into training and validation, so all data in training
  - small data sets are effectively double-weighted (e.g. E866, CMS W charge asymmetry)
  - NNPDF4.0 includes more jet data than NNPDF3.1, sees similar issues as CT, MSHT, ATLAS