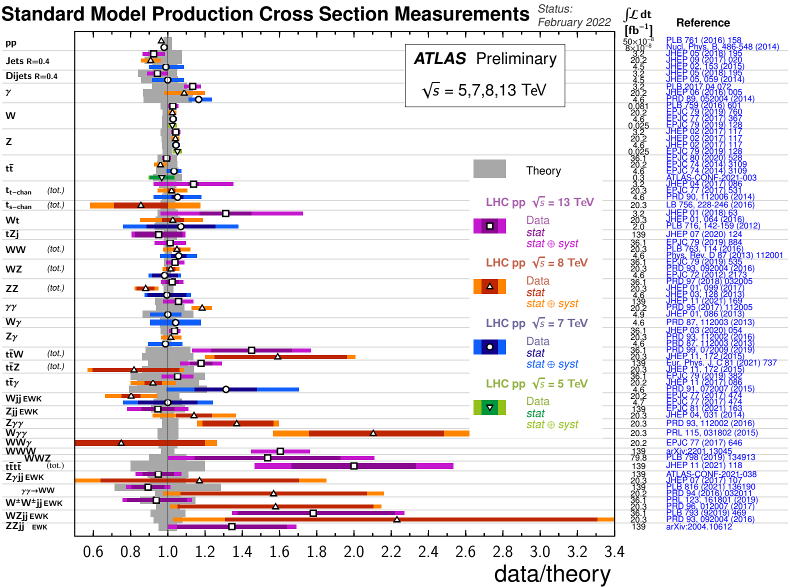
The PDF4LHC21 combination of global PDF fits for the LHC Run III

based on R.D. Ball *et al.*, [2203.05506], accepted for publication in J. Phys. G

Emanuele R. Nocera
School of Physics and Astronomy, The University of Edinburgh
on behalf of the PDF4LHC21 Combination Group

29$^{\text{th}}$ June 2022

Physics at the LHC as Precision Physics



[Plot from ATLAS Collaboration web page]

PDF determination in a nutshell

THEORY: Factorisation of physical observables

$$\sigma(Q^2,\tau,\{k\}) = \sum_{ij}\int_\tau^1 \frac{dz}{z}\mathcal{L}_{ij}(z,Q^2)\hat{\sigma}_{ij}\left(\frac{\tau}{z},\alpha_s(Q^2),\{k\}\right) \qquad \tau = M^2/s$$

$$\mathcal{L}_{ij}(z,Q^2) = \int_x^1 \frac{dx}{x} f_i^{h_1}(x,Q^2) f_j^{h_2}\left(\frac{z}{x},Q^2\right)$$

DATA: A set of experimental measurements

inclusive/differential cross-sections for a variety of ℓp and $p\bar{p}$ production processes

METHODOLOGY: A PDF parametrisation and an optimisation algorithm

$$xf_i(x,Q_0^2) = A_{f_i}\, x^{a_{f_i}}\, (1-x)^{b_{f_i}}\, \mathscr{F}(x,\{c_{f_i}\})$$

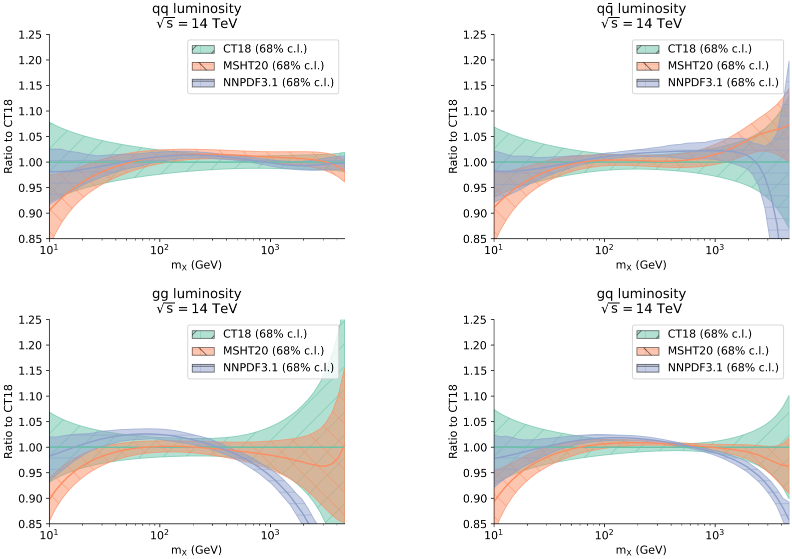
A prescription to represent uncertainties into PDF uncertainties

$$E[\mathcal{O}] = \int \mathcal{D}f\mathcal{P}(f|data)\mathcal{O}(f) \qquad V[\mathcal{O}] = \int \mathcal{D}f\mathcal{P}(f|data)[\mathcal{O}(f)-E[\mathcal{O}]]^2$$

Monte Carlo: bootstrap $\mathcal{P}(f|data)$ Hessian: project $\mathcal{P}(f|data)$

CT18	MSHT20	~~NNPDF4.0~~ NNPDF3.1
[PRD 103 (2021) 014013]	[EPJC 81 (2021) 341]	[~~EPJC 82 (2022) 428;~~ 77 (2017) 663]

Comparison of CT18, MSHT20 and NNPDF3.1



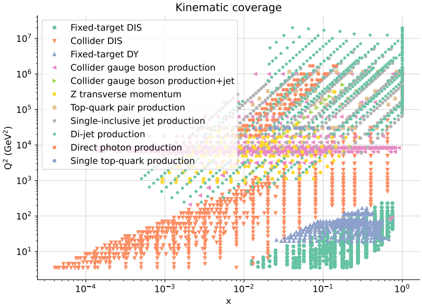
Where do differences come from? **BENCHMARK**

Is there a way to take them into account in a single PDF set? **COMBINATION**

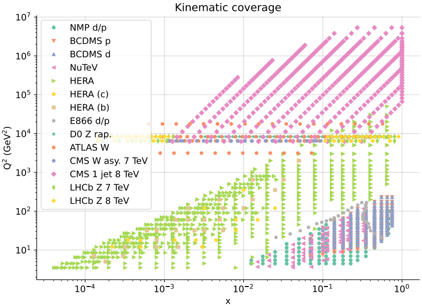
1. Benchmark

Reduced fits — common settings

Global data set($N_{\rm dat} \sim 5000$)



Reduced data set ($N_{\rm dat} \sim 2000$)



Same reduced data set

Same heavy quark mass values: $m_c = 1.4$ GeV; $m_b = 4.75$ GeV; $\alpha_s(M_Z) = 0.118$

Same strong coupling value: $\alpha_s(M_Z) = 0.118$

No strangeness asymmetry at input scale: $(s - \bar{s})(Q_0) = 0$

Charm perturbatively generated

Positive-definite quark distributions

No deuteron or nuclear corrections to analyse eN cross sections

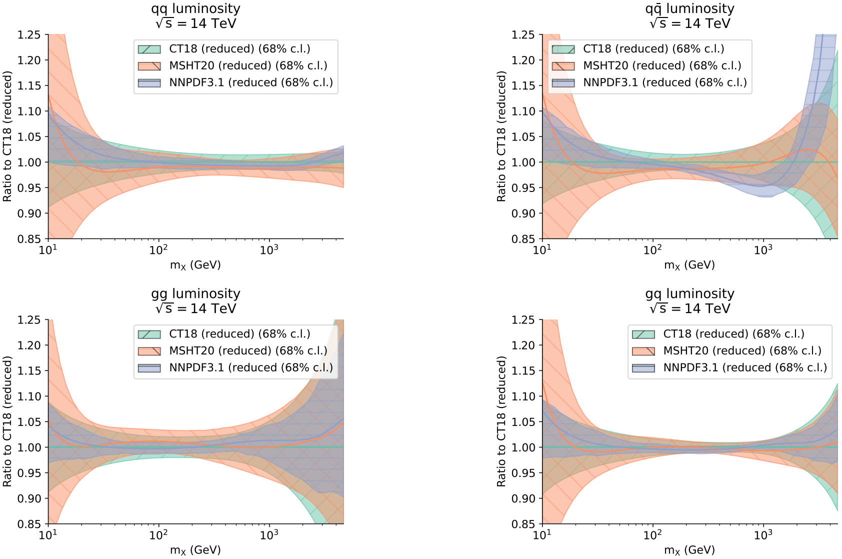
Fixed branching ratio for charm hadrons to muons; NNLO corrections for dimuon data

Reduced fits — fit quality

Dataset	$N_{\rm dat}$	$\chi^2/N_{\rm dat}$		
		CT18	MSHT20	NNPDF3.1
BCDMS F_2^p	329/163††/325†	1.06	1.00	1.21
BCDMS F_2^d	246/151††/244†	1.06	0.88	1.10
NMC F_2^d/F_2^p	118/117†	0.93	0.93	0.90
NuTeV dimuon $\nu+\bar{\nu}$	38+33	0.79	0.83	1.22
HERAI+II	1120	1.23	1.20	1.22
E866 $\sigma_{pd}/(2\sigma_{pp})$	15	1.24	0.80	0.43
LHCb 7 TeV & 8TeV W,Z	29+30	1.15	1.17	1.44
LHCb 8 TeV $Z\to ee$	17	1.35	1.43	1.57
ATLAS 7 TeV W,Z (2016)	34	1.96	1.79	2.33
D0 Z rapidity	28	0.56	0.58	0.62
CMS 7 TeV electron $A_{\rm ch}$	11	1.47	1.52	0.76
ATLAS 7 TeV W,Z(2011)	30	1.03	0.93	1.01
CMS 8TeV incl. jet	185/174††	1.03	1.39	1.30
Total $N_{\rm dat}$	—	2263	1991	2256
Total $\chi^2/N_{\rm pt}$	—	1.14	1.15	1.20

Overall general fair agreement of fit quality across the three reduced fits

Reduced fits — luminositites



Very good agreement within uncertainties; similar size of uncertainties in the data region
Remaining differences reflect methodological choices

2. Combination

The PDF4LHC21 combination

Input: CT18′, MSHT20 and NNPDF3.1′

common, fixed value of strong coupling and quark masses:
$\alpha_s(M_Z) = 0.118$, $m_c = 1.4$ GeV, $m_b = 4.75$ GeV, $m_t = 172.5$ GeV

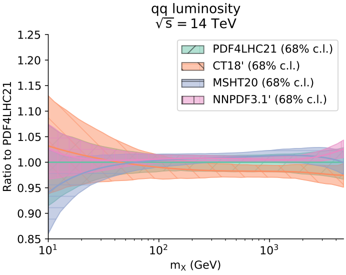
all three parton sets are accurate to NNLO in the strong coupling
all three parton sets incorporate NNLO charm-quark mass corrections relevant to CCFR/NuTeV
other theoretical and methodological details are as in the published parton sets

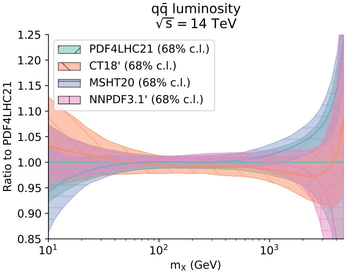
NNPDF3.1′ data set has been extended to resemble CT18 and MSHT20
(ATLAS W, Z 7 TeV; CMS single-inclusive jets 8 TeV; CMS 2D top pair 8 TeV)
CT18′ is as CT18, but it incorporates NNLO charm-quark mass corrections and $m_c = 1.4$ GeV

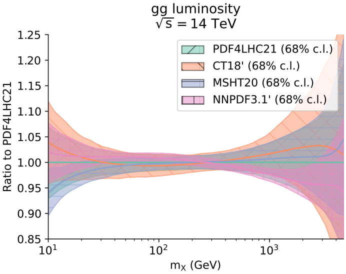
Output: the PDF4LHC21 combination

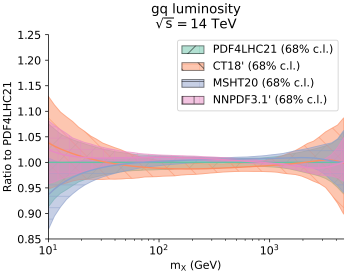
the combination closely follows the procedure adopted for the PDF4LHC15 combination
generate 300 Monte Carlo replicas for each of the three parton sets and collate them
CT18 use the CTEQ-TEA algorithm [JHEP 1703 (2017) 099]
MSHT20 use the Thorne-Watt algorithm [JHEP 1208 (2012) 05]
compression and conversion: 100 Monte Carlo replicas and 40 Hessian eigenvectors
compression: CMC algorithm [Eur.Phys.J. C75 (2015) 474]
conversion: META-PDF algorithm [JHEP 1407 (2014) 035]; [mc2h algorithm [Eur.Phys.J. C81 (2021) 530]]

Constructing PDF4LHC21



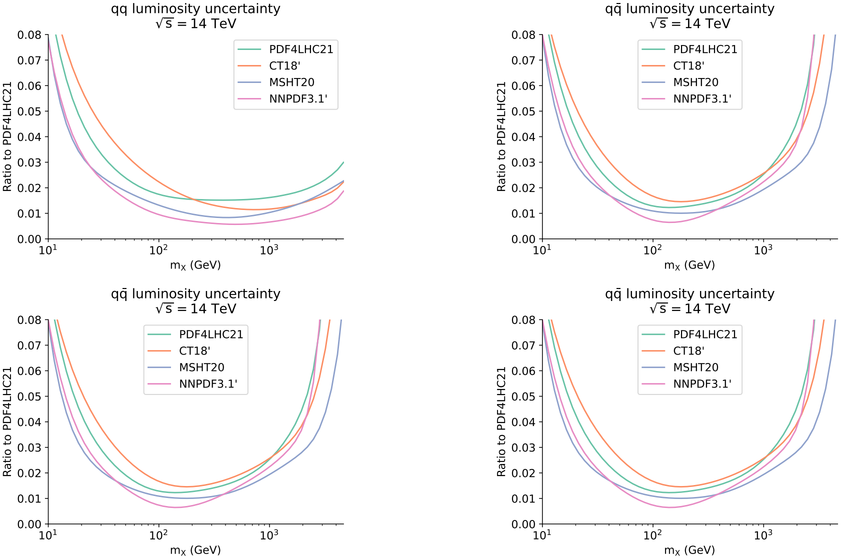






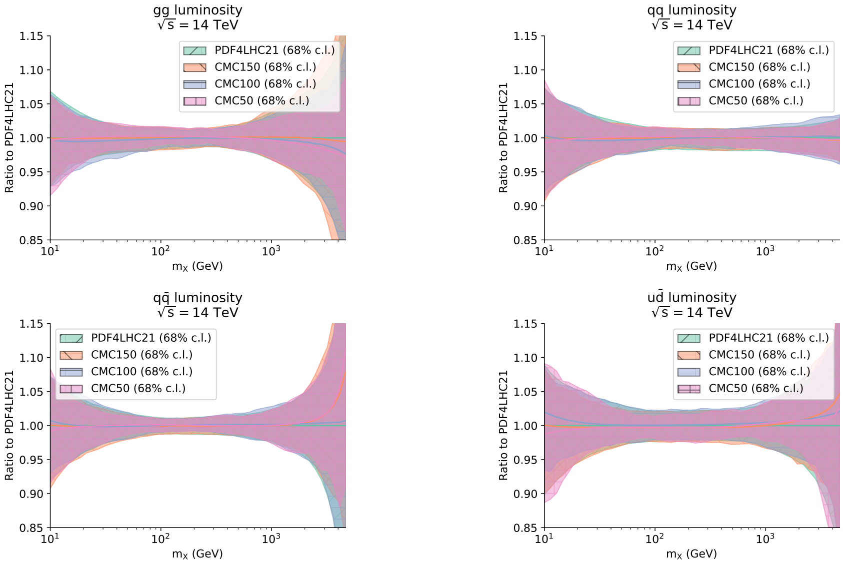
The PDF4LHC21 combination captures the features of the constituent PDF sets
The PDF4LHC21 provides a conservative estimate of PDF uncertainties

Constructing PDF4LHC21



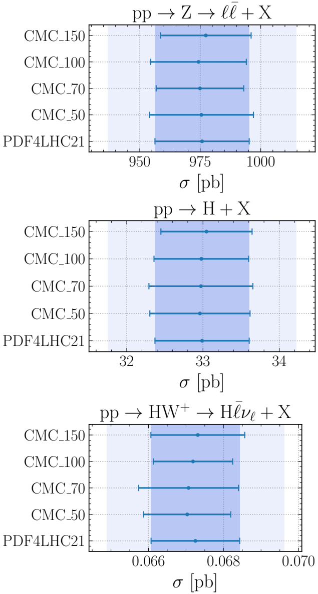
The PDF4LHC21 combination captures the features of the constituent PDF sets
The PDF4LHC21 provides a conservative estimate of PDF uncertainties

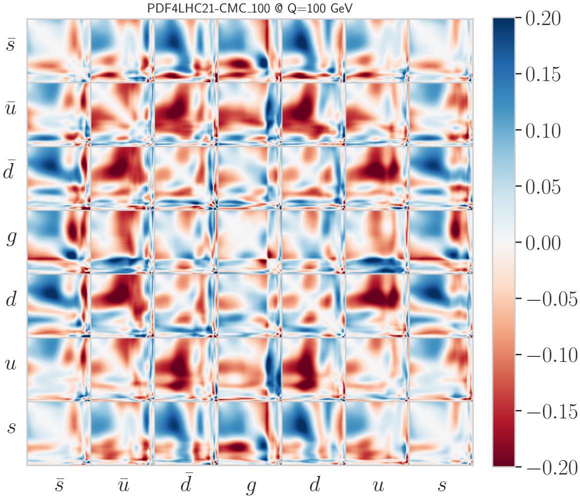
Compressing PDF4LHC21



Compression to 100 Monte Carlo replicas is a good compromise between 50 and 150

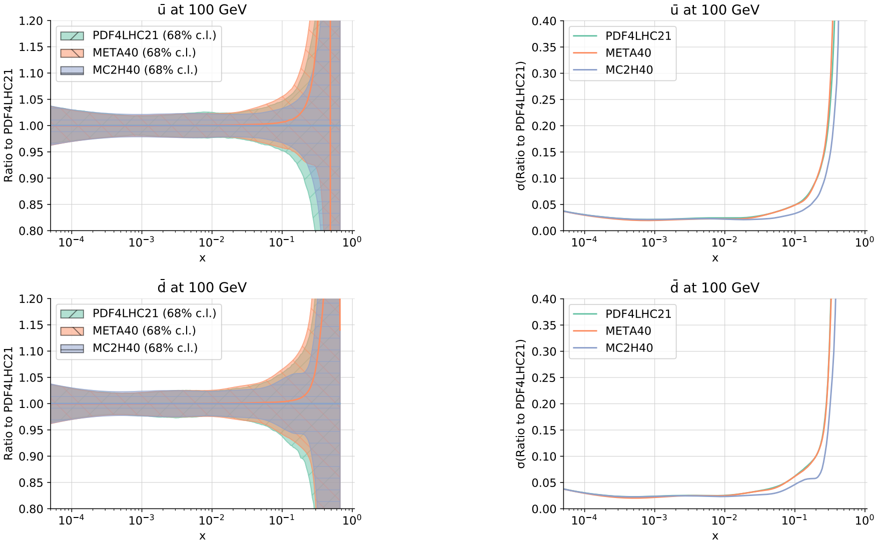
Compressing PDF4LHC21





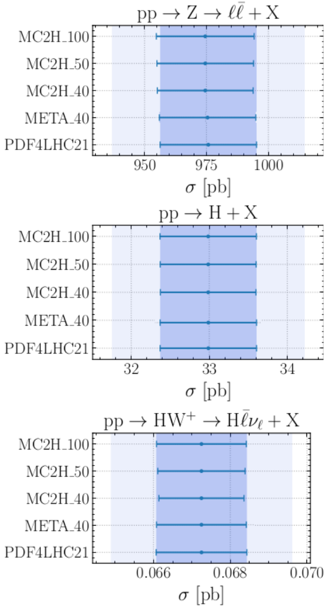
Less than 100 MC replicas may distort uncertainties for key LHC processes and PDF correlations a little too much

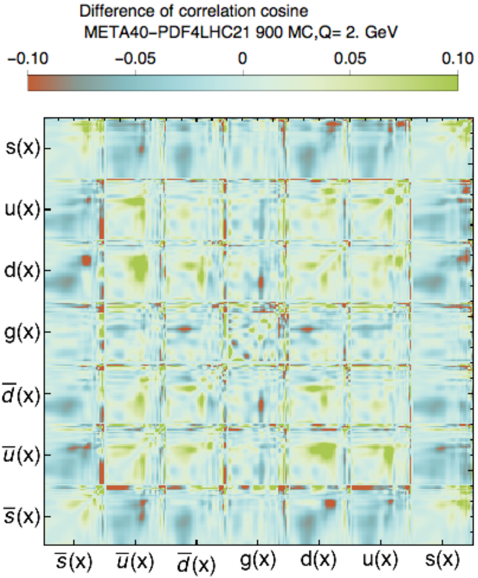
Converting PDF4LHC21



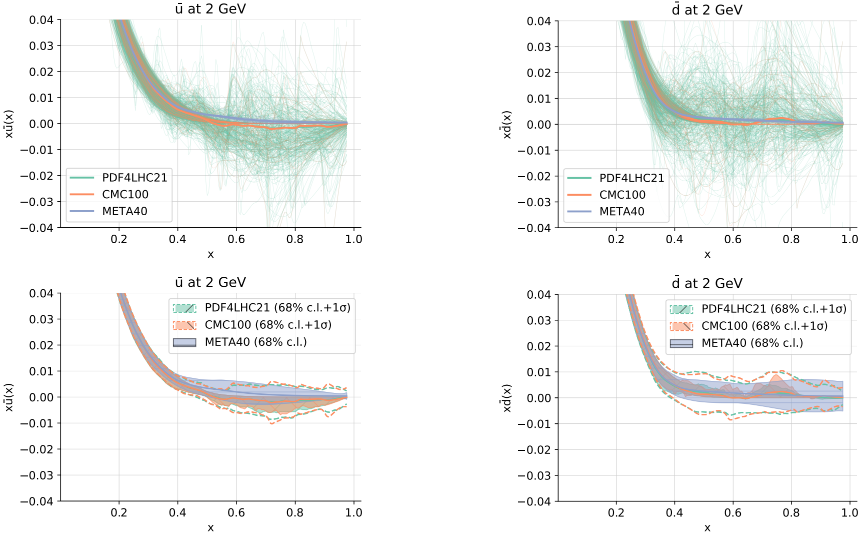
Given a small number of eigenvectors (40), the META pdf algorithm is slightly more accurate than MC2H at large x, therefore it is chosen as default

Converting PDF4LHC21





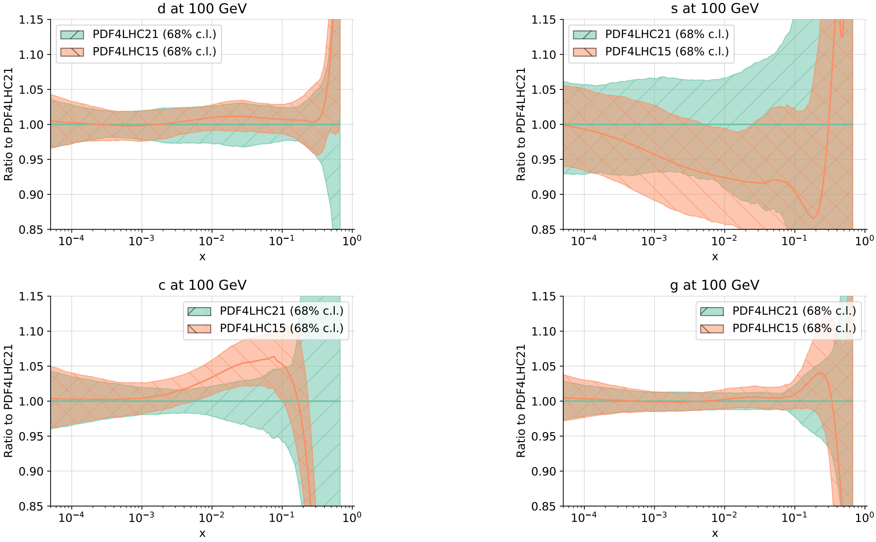
Converting PDF4LHC21



Stretch the meta-PDF parametrisation to enforce PDF positivity at large x
Non-Gaussianity of the replica distribution in the large-x region is lost with meta PDFs

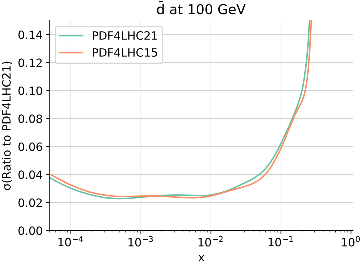
3. Comparison with PDF4LHC15

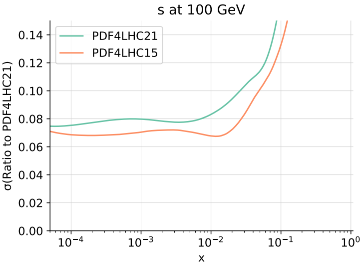
Comparing PDF4LHC21 and PDF4LHC15

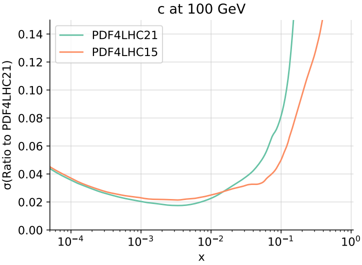


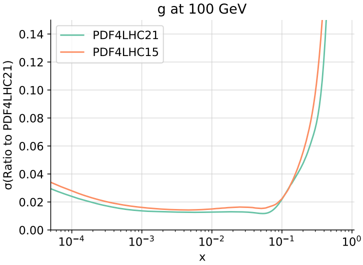
Good agreement between PDF4LHC21 and PDF4LHC15; differences due to new data or theory
Uncertainties increase where spread across the three input sets has increased

Comparing PDF4LHC21 and PDF4LHC15



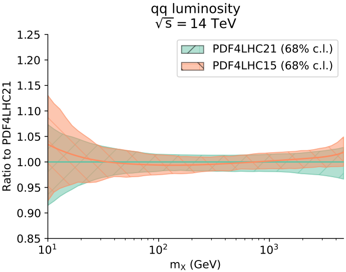


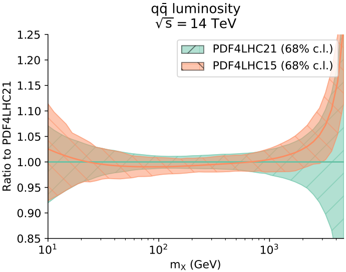


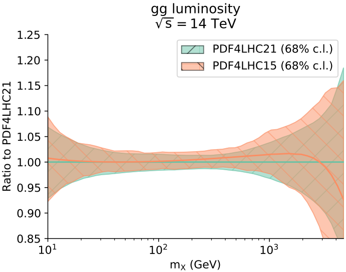


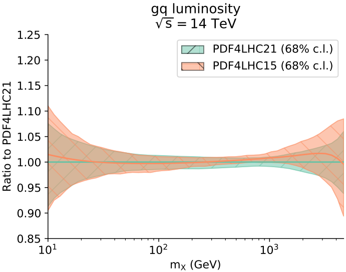
Good agreement between PDF4LHC21 and PDF4LHC15; differences due to new data or theory
Uncertainties increase where spread across the three input sets has increased

Comparing PDF4LHC21 and PDF4LHC15



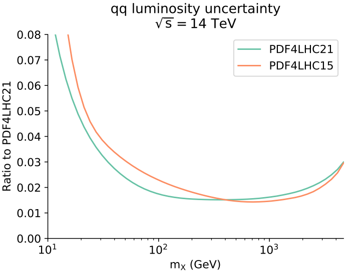


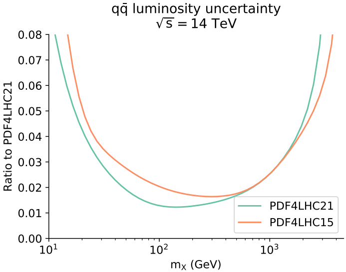


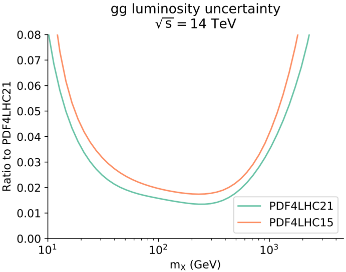


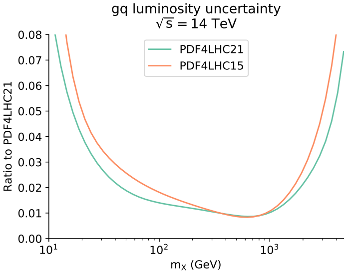
Good agreement of central values
Uncertainties are generally reduced in PDF4LHC21 w.r.t. PDF4LHC15

Comparing PDF4LHC21 and PDF4LHC15



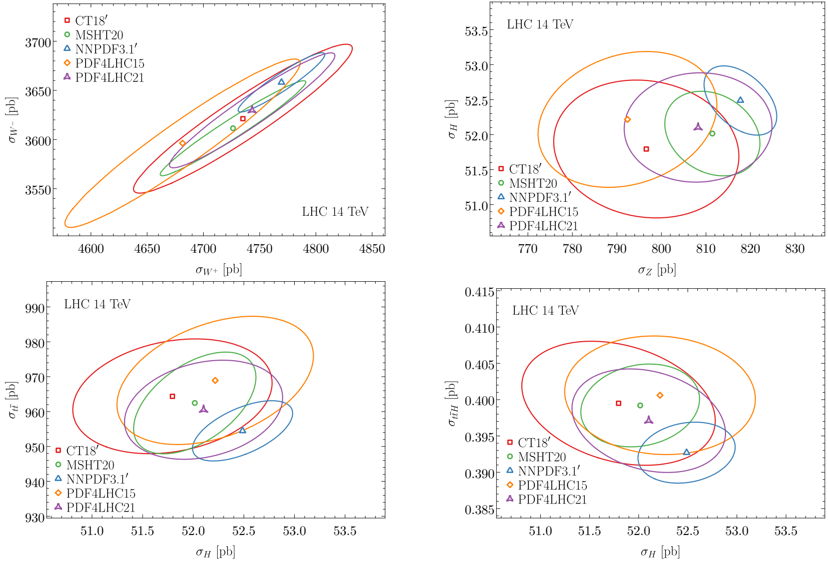






Good agreement of central values
Uncertainties are generally reduced in PDF4LHC21 w.r.t. PDF4LHC15

Comparing PDF4LHC21 and PDF4LHC15



Error ellipses of PDF4LHC21 systematically reduced w.r.t. PDF4LHC15
Good proxy for the constituent PDF sets

4. Conclusion

Summary and delivery

The features of the three partons sets entering the PDF4LHC21 conbination are well understood
Excellent compatibility between PDF4LHC21 and PDF4LHC15, slight increase in precision
Extensive tests of the statistical features of the Monte Carlo and Hessian parton sets

LHAPDF6 grid name	Pert. order	$n_f^{\max}$	ErrorType	$N_{\rm mem}$	$\alpha_s(m_Z^2)$
PDF4LHC21_MC	NNLO	5	replicas	101	0.118
PDF4LHC21_40	NNLO	5	symmhessian	41	0.118
PDF4LHC21_MC_PDFAS	NNLO	5	replicas+as	103	mem 0:100 → 0.118 mem 101 → 0.117 mem 102 → 0.119
PDF4LHC21_40_PDFAS	NNLO	5	symmhessian+as	43	mem 0:40 → 0.118 mem 41 → 0.117 mem 42 → 0.119
PDF4LHC21_MC_NF4	NNLO	4	replicas	101	0.118
PDF4LHC21_40_NF4	NNLO	4	symmhessian	41	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	43	mem 0:40 → 0.118 mem 41 → 0.117 mem 42 → 0.119

All grids are available from LHAPDF: https://lhapdf.hepforge.org/

Usage and Recommendations

Guidance largely follows PDF4LHC15, read the fine print at
https://www.hep.ucl.ac.uk/pdf4lhc/
where guidelines are discussed and additional PDF set variants are provided

Case	Recommendation	Rationale
Comparison between data and theory for **SM measurements**	**Individual sets** (and use as many sets as possible)	If measurements have potential to constrain PDFs, then it is best to compare with individual sets, particularly given the high precision of some measurements. The same applies to the determination of (SM) parameters.
Searches for **BSM phenomena** or study of **low-precision SM measurements**	Use **PDF4LHC21_mc** or **PDF4LHC21 40**	Reduces computational burden and provides estimates of central values/uncertainties that agree with the three input PDF sets. One may wish to consider individual PDF sets if particular sensitivity to PDFs or PDF uncertainties is required. Monte Carlo set PDF4LHC21_mc - Reproduces also non-Gaussian aspects of baseline 900 replica set, however can go negative at very large x . Non-Gaussian features more likely in extrapolation regions so MC set may be beneficial here. Hessian set PDF4LHC21_40 - Advantage when speed is desirable as 40 members, Positivity in $x \to 1$ limit also may be beneficial for some applications.
Theoretical computations	Use **PDF4LHC21_mc** or **PDF4LHC21 40**	The PDF4LHC21 combination includes information from all three input global fits and combines PDF uncertainty before theoretical calculation is done. Its uncertainty is moderately conservative and encloses the predictions of all three groups.

Citation policy: cite PDF4LHC21 document and individual PDF sets

Usage and Recommendations

Guidance largely follows PDF4LHC15, read the fine print at
https://www.hep.ucl.ac.uk/pdf4lhc/
where guidelines are discussed and additional PDF set variants are provided

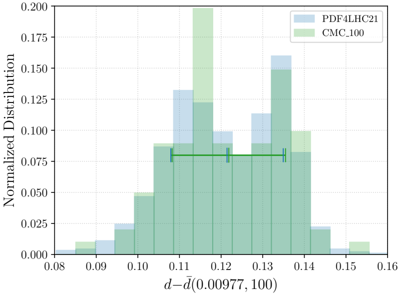
Case	Recommendation	Rationale
Comparison between data and theory for **SM measurements**	**Individual sets** (and use as many sets as possible)	If measurements have potential to constrain PDFs, then it is best to compare with individual sets, particularly given the high precision of some measurements. The same applies to the determination of (SM) parameters.
Searches for **BSM phenomena** or study of **low-precision SM measurements**	Use **PDF4LHC21_mc** or **PDF4LHC21 40**	Reduces computational burden and provides estimates of central values/uncertainties that agree with the three input PDF sets. One may wish to consider individual PDF sets if particular sensitivity to PDFs or PDF uncertainties is required. Monte Carlo set PDF4LHC21_mc - Reproduces also non-Gaussian aspects of baseline 900 replica set, however can go negative at very large x . Non-Gaussian features more likely in extrapolation regions so MC set may be beneficial here. Hessian set PDF4LHC21_40 - Advantage when speed is desirable as 40 members, Positivity in $x \to 1$ limit also may be beneficial for some applications.
Theoretical computations	Use **PDF4LHC21_mc** or **PDF4LHC21 40**	The PDF4LHC21 combination includes information from all three input global fits and combines PDF uncertainty before theoretical calculation is done. Its uncertainty is moderately conservative and encloses the predictions of all three groups.

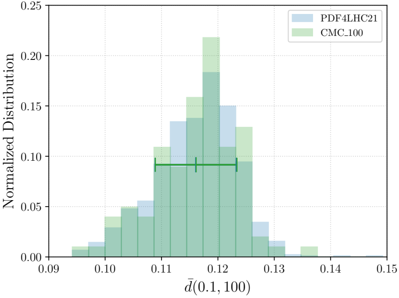
Citation policy: cite PDF4LHC21 document and individual PDF sets

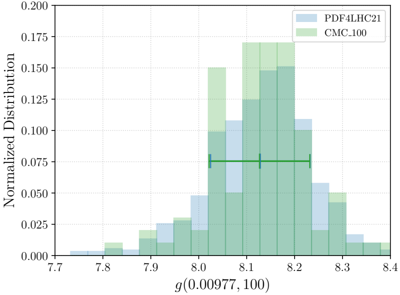
Thank you

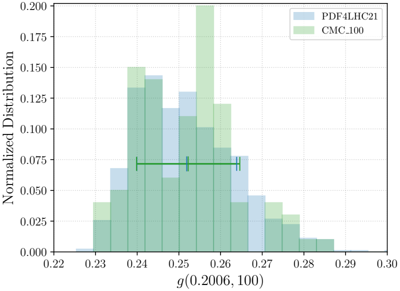
A. Appendix [Most of the slides are by courtesy of Thomas Cridge]

Compressing PDF4LHC21

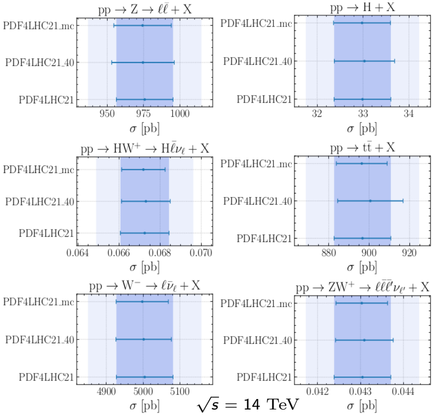


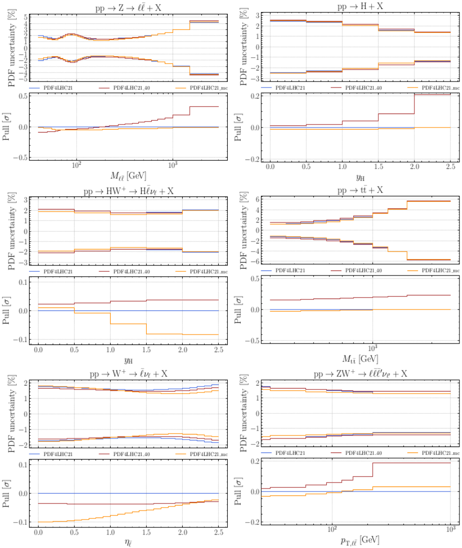






Comparison with baseline 900 set: σ, $d\sigma/d\mathcal{O}$

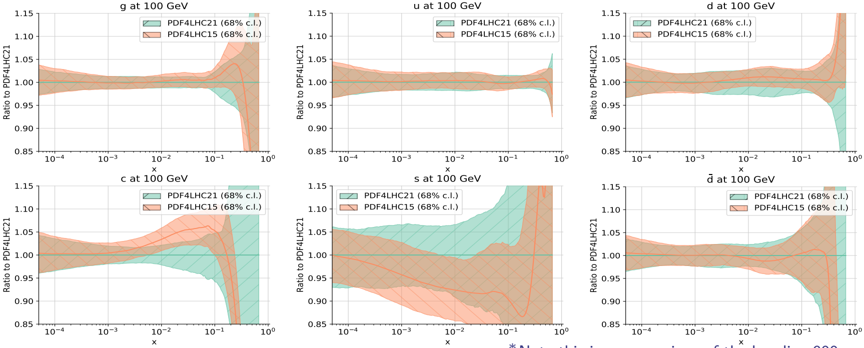




- Very good agreement of baseline 900 replica set with MC 100 replica, Hessian 40 member sets.

N.B. Can have small differences for Hessian 40 set as positivity imposed at large x (backup).

PDF4LHC21 vs PDF4LHC15*: PDF Central Values

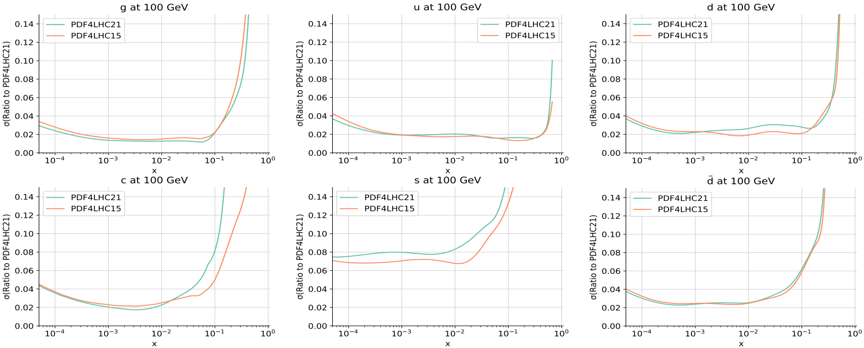


*Note this is a comparison of the baseline 900 replica sets.

- Consistent for all flavours and x values.
- Remarkable agreement for $u, d, \bar{d}, \bar{u}$ and g for $x \lesssim 0.1$.
- High x gluon differs due to new data, lowered but within errorbands.
- Strange quark notably raised for $x \gtrsim 10^{-3}$ due to ATLAS high precision W, Z data in NNPDF3.1' and MSHT20.
- Charm raised at (very) high x due to NNPDF3.1' fitted charm.

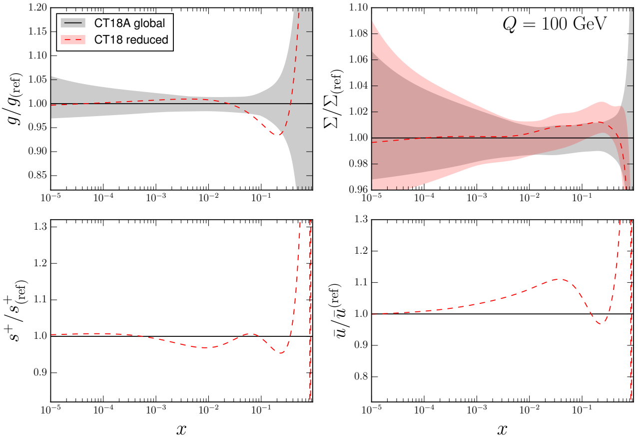
In PDF4LHC15 all groups had perturbative charm.

PDF4LHC21 vs PDF4LHC15: PDF Uncertainties



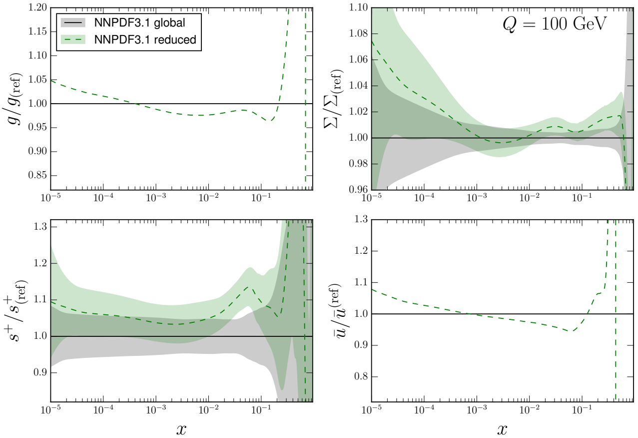
- PDF errorbands similar, reduced in some places, raised in others.
- Gluon errorband reduced across all x even though individual groups disagreement increased because individual groups' errorbands reduced.
- Uncertainties increase where disagreement between three input sets have worsened, e.g. for strangeness or for charm at $x \gtrsim 10^{-2}$.
- s disagreement affects d PDF at $x \sim 10^{-2}$ increasing its uncertainty.

Reduced Fits: CT18 reduced fit vs CT18A global fit



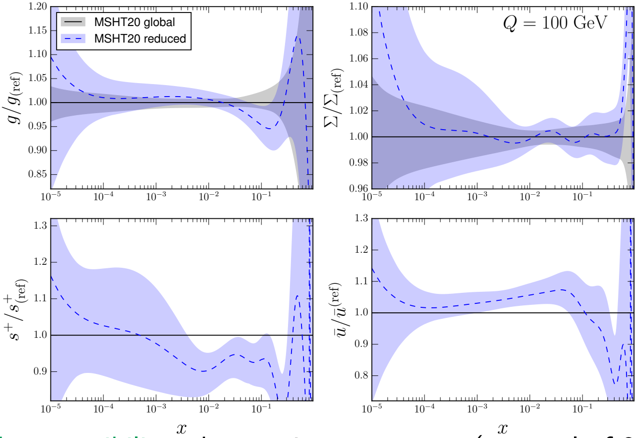
- Good compatibility with change in high x gluon shape and some increase in $\bar{u}$. Some changes in flavour decomposition.
- Some increase in *nominal* PDF uncertainties, particularly at low x.

Reduced Fits: NNPDF reduced fit vs NNPDF3.1 global



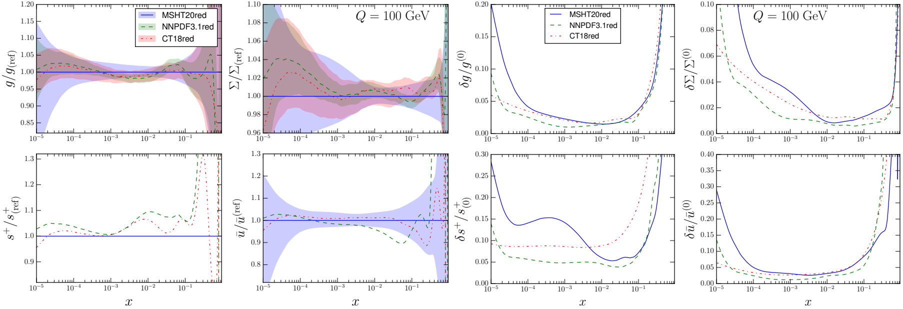
- Good compatibility, changes in strangeness (see later) and change in large x gluon (removal of top data, addition of CMS 8 TeV jet).
- Generally slightly increased uncertainties, particularly at low x.

Reduced Fits: MSHT reduced fit vs MSHT20 global fit



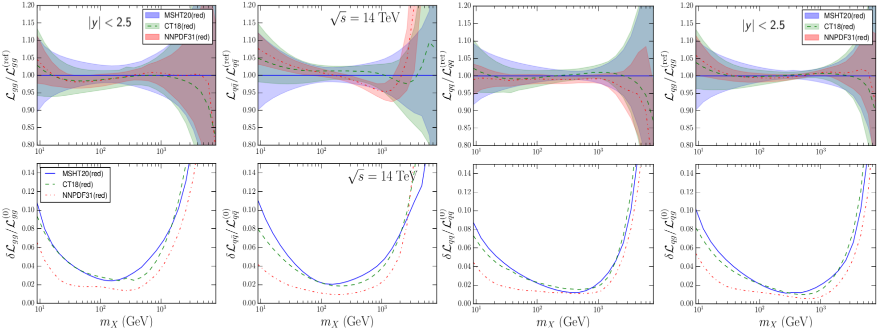
- Good compatibility, changes in strangeness (removal of 8 TeV ATLAS W, Z data), flavour decomposition and large x gluon.
- Marked increase in uncertainties of reduced fit, particularly outside of regions where there are data.

Reduced Fits PDF Comparison



- Very good agreement within uncertainties, including gluon.
- Similar size uncertainties in data regions, differences outside this, parallel study into differences in uncertainty bands ongoing.
- Agreement much improved relative to global PDFs.
- Same data and theory settings $\rightarrow$ consistent PDFs. Smaller remaining differences, e.g. in errors, reflect methodological choices.

Reduced Fits: Luminosity comparison



- Very good agreement in luminosities, gg agrees across whole of m_X.
- Differences in uncertainties, particularly at low masses and in gg.
- Same data and theory settings $\rightarrow$ consistent PDFs. Reduced fits well understood, benchmarking successful!
- Benchmarking with reduced fits has shown valid differences between PDFs from data, theory, methodology $\Rightarrow$ should enter combination.

PDF4LHC15 in Predictions Datasets χ^2 Comparison

- First make predictions with PDF4LHC15 PDFs, identifies any differences in theory/data between groups with fixed PDFs.

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 electron 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

PDF4LHC21 reduced fit dataset χ^2/N_{pt} with PDF4LHC15 PDF inputs, i.e. before fitting, ††MSHT †NNPDF.

- Similar overall quality of fit for MSHT and CT in χ^2/N, NNPDF significantly larger χ^2/N.
- Differences in some datasets:
  - Difference in NNPDF HERA χ^2 - flavour scheme, disappears in fit.

Table from T. Hobbs

Reduced Fits Datasets χ^2 Comparison

ID	Expt.	N_{pts}	χ^2/N_{pts} (CT)	χ^2/N_{pts} (MSHT)	χ^2/N_{pts} (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 electron 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_{pts}	—	2263	1991	2256
Total	χ^2/N_{pts}	—	1.14	1.15	1.20

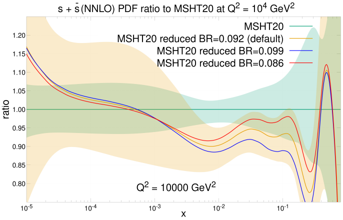
PDF4LHC21 reduced fit dataset χ^2/N_{pts} after fitting, ††MSHT †NNPDF.

Table from T. Hobbs

- Similar overall quality of fit in χ^2/N.
- Differences remaining in some datasets (as expected), investigated in benchmarking (backup slides) ⇒ reflect theory settings and methodological choices.
- Differences remaining in some datasets:
  - NuTeV agreement improved but difference remains, seen in $s + \bar{s}$.
  - Some differences in NNPDF fit quality to small datasets.

Flavour Decomposition - Strangeness and NuTeV

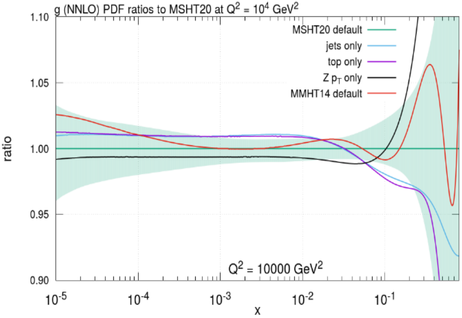
- One of the main differences between the first reduced sets was in the flavour decomposition and strangeness.
- NuTeV dimuon data key driver of this, complicated dataset:
  - Requires knowledge of charm hadron $\rightarrow$ muon branching ratio (BR).
  - Non-isoscalar nature of target.
  - Prefers non-zero strangeness asymmetry.
  - Acceptance corrections required.
- BR($c \rightarrow \mu$) anti-correlated with strangeness, 3 groups have different values:
  - NNPDF 0.087 ± 0.005
  - MSHT 0.092 ± 0.01 variable.
  - CT 0.099, normalisation uncertainty.



- Choose same BR fixed at 0.092 $\Rightarrow$ better strangeness agreement, largely within uncertainties between all 3 groups.
- Also aids reduction in flavour decomposition differences.

High x gluon

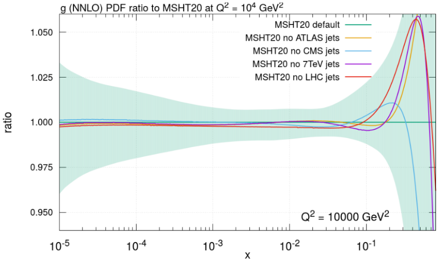
- High x gluon of interest to both reduced and global fits.
- 3 main datasets play a role here - jet data, top data, Zp_T data, different pulls:
- Not straightforward to fit some of them:
    - Difficulties fitting all bins.
    - Possible tensions.
    - Issue of correlated systematics.



- Global fit is a balance between these different pulls.
- MSHT, CT, NNPDF observe differences in the relative importance of these datasets and the quality of their individual fits
  - *does the same hold in reduced fits and can we understand this better in this context?*

High x gluon - Jet tensions

- Not only tensions between different dataset types at high x, also tensions within dataset types, e.g. between different jet measurements.
- ATLAS 7 TeV jets pulls gluon down at high x, whereas CMS jets (mainly 8 TeV) pull gluon up.
- Global fit is a balance between these different pulls and those of Zp_T, $t\bar{t}$ datasets here.



† MSHT20, TC, S. Bailey, L. Harland-Lang, A. Martin, R. Thorne 2012.04684

ATLAS 8 TeV multi-differential $t\bar{t}$ lepton+jets

- Comes differential in 4 variables with correlations - $m_{tt}, y_t, y_{tt}, p_t^T$.
- MSHT*, CT+ difficulties fitting all 4 distributions simultaneously.
- MSHT, CT, ATLAS− cannot get good fit to y_t or y_{tt} individually.
- NNPDF3.0 however able to fit all 4 distributions well individually†.

Benchmarking:

- Adding to reduced fit, what happens?

Distribution/N	$p_t^T/8$	$y_t/5$	$y_{tt}/5$	$m_{tt}/7$	Total
MSHT PDF4LHC15 in	3.0	10.6	17.6	4.3	35.5
NNPDF PDF4LHC15 in	3.4	9.5	16.2	4.1	33.2
CT PDF4LHC15 in	3.1	10.1	15.3	4.2	32.7
MSHT fit uncorrelated	3.8	8.4	12.5	6.4	31.2
CT fit uncorrelated	3.4	12.9	17.3	6.1	39.7
NNPDF fit uncorrelated	7.2	3.9	5.1	2.5	18.7
MSHT fit correlated	-	-	-	-	130.6
NNPDF fit correlated	-	-	-	-	122.7
MSHT fit decorrelated	-	-	-	-	35.3

Before Fitting

All groups χ^2 in agreement, same pattern - poor χ^2 for rapidity data.

After Fitting (Uncorrelated)

MSHT and CT see poor fits to rapidities y_t, y_{tt} but NNPDF see good fits to rapidities, as in global fits.

After Fitting (Correlated)

MSHT and NNPDF both see very poor fit to all 4 distributions with correlations, as in global fits.

- Same behaviour as in global fits after fitting....

* S. Bailey & L.Harland-Lang 1909.10541. + Kadir et al 2003.13740.
† Czakon et al 1611.08609. − ATL-PHYS-PUB-2018-017.

Benchmarking ATLAS 8 TeV $t\bar{t}$ lepton+jets

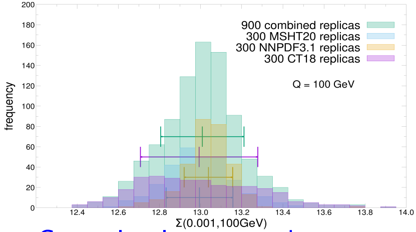
- How can we explain these differences in global and reduced fits?
- Global fits have different fit environments - different weights and other datasets included, tensions may affect fit quality for this dataset:
  - NNPDF3.0 had little jet data - perhaps tensions cause issues in y_t, y_{tt}. NNPDF4.0 sees similar behaviour to other groups.
  - NNPDF reduced fit up-weights this dataset by putting all data in training (as small dataset) - perhaps up-weighting causes difference.
- Investigate weights and tensions in reduced fit environment:

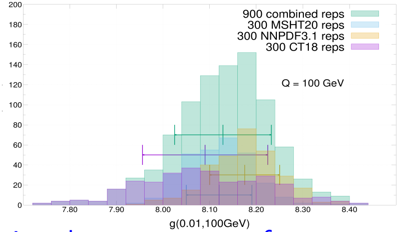
Dataset (N)	MSHT reduced (default CMS8j)	NNPDF reduced (default CMS8j)	MSHT reduced (CMS7j)	MSHT reduced (AT7j)	MSHT reduced (no jets)	MSHT reduced (CMS8j, double weight $t\bar{t}$)
χ^2/N	1.15	1.20	1.11	1.17	1.12	1.15
p_t^T (8)	3.8	7.2	4.0	4.6	4.5	4.2
y_t (5)	8.4	4.3	6.4	5.5	5.2	5.8
y_{tt} (5)	12.5	5.7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

- Weights and tensions with other datasets notably affect fit quality, removing these differences $\Rightarrow$ similar behaviour can be observed.

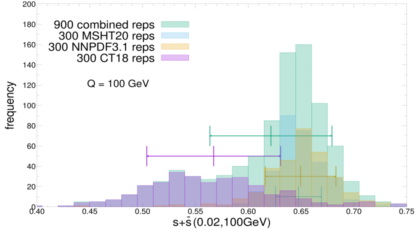
Global Fits Specific Comparisons: PDF4LHC21 input replicas

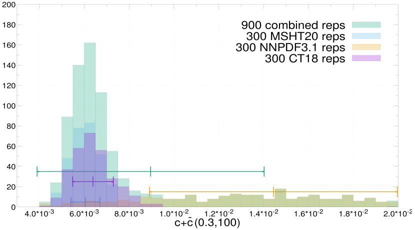
- Central value is average of those of the 3 global fits input.
- Central values agree closely ⇒ uncertainty is average of 3 groups:





- Central values spread ⇒ uncertainty has component from spread.





- Combination has expected properties in central values and errors.

Replica generation:

- The PDF4LHC21 baseline combination is a set of 900 replicas, constituted of 300 replicas from CT18', MSHT20 and NNPDF3.1'.
- CT18' and MSHT20 must therefore be transformed into Monte Carlo representations to generate their 300 replicas.
- Existing methods already available - basic idea is to sample probability distribution described by the eigenvectors randomly whilst preserving the central value as the average of the replicas.
- Watt-Thorne Method (MSHT20):

$$\mathcal{F}^{(k)} = \mathcal{F}\left(S_0\right) + \frac{1}{2}\sum_{j=1}^{N_{\rm eig}} \left[\mathcal{F}\left(S_i^{(+)}\right) - \mathcal{F}\left(S_i^{(-)}\right)\right] R_j^{(k)}\,, \qquad k = 1\,\ldots, N_{\rm rep}$$

- CT (Hou et al) Method (CT18'):

$$X^{(k)} = X(S_0) + \sum_{i=1}^{N_{\rm eig}} \left(\frac{X\left(S_i^{(+)}\right) - X\left(S_i^{(-)}\right)}{2} R_i^{(k)} + \frac{X\left(S_i^{(+)}\right) + X\left(S_i^{(-)}\right) - 2X\left(S_0\right)}{2} \left(R_i^{(k)}\right)^2 \right) + \Delta.$$

PDF4LHC21 and NNPDF4.0:

- NNPDF4.0 appeared relatively late in the PDF4LHC21 benchmarking/combination effort, therefore now included.
- Instead NNPDF3.1' (aka NNPDF3.1.1) is included which is intermediate between NNPDF3.1 and NNPDF4.0.
- Comparison of NNPDF3.1', NNPDF4.0 and PDF4LHC21 PDFs: