PDF4LHC2021 Benchmarking

Thomas Cridge

University College London

14th April 2021

On behalf of PDF4LHC Collaboration

DIS 2021 Workshop

Outline

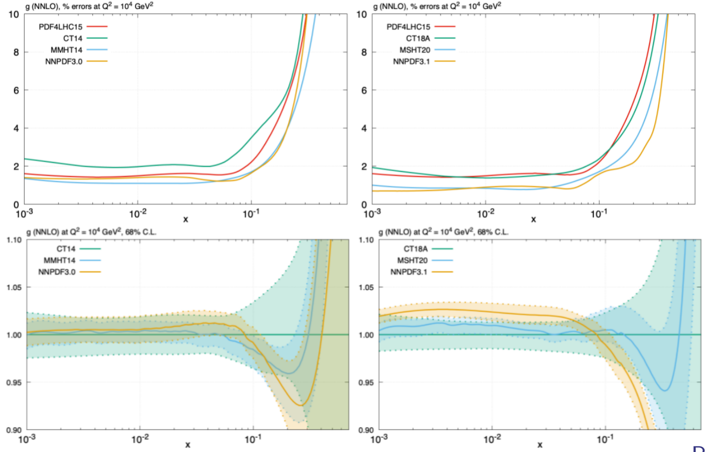
Work undertaken through many useful discussions, many thanks to all members involved.

Introduction - PDF Landscape

- PDF4LHC15 was a 1 year benchmarking exercise of the CT14, MMHT14, NNPDF3.0 PDFs which resulted in a combination set.
- It has now been more than 5 years since the PDF4LHC15 benchmarking exercise.
- Increasing amounts of data coming out of the LHC, greater precision, more channels, more differential $\Rightarrow$ many changes in PDFs.
- In addition there have been theoretical improvements $\Rightarrow$ full NNLO predictions, and methodological improvements (parameterisations, algorithms, etc).
- PDFs now known more accurately and precisely than ever before, but some differences remain $\Rightarrow$ benchmarking needed.
- We consider 3 global PDF fits most recent sets, which include much of the recent datasets: MSHT20, CT18, NNPDF3.1.

Introduction - Changes in PDFs

Gluon



Plots from L. Harland-Lang

- Reduction in PDF uncertainties seen across all 3 groups.
- Central value agreement not as good, some differences emerging.

Note: CT18A shown for ease of comparison, however CT18 is the default set.

PDF Benchmarking: Aim and Approach

- New PDFs CT18, MSHT2020, NNPDF3.1 ⇒ now is a good time to undertake a benchmarking exercise, examine and understand differences ahead of a new PDF4LHC future combination.
- End result of exercise will be a PDF4LHC21 set of PDFs.
- Desire to understand origin of differences:
  - Are they due to variations of experimental input, different theory settings, methodologies? Are these equally valid choices?
- Seek to remove as many differences in input/approach as possible:
  - Common input data - Small subset of datasets ⇒ reduced fits.
  - Common theory settings wherever possible.
  - Examine methodological differences in parallel as much as possible.
- Reduced fits offer *ease of comparison at expense of robustness*.
- To benchmark the reduced fits:
  - Compare PDFs directly to look for areas of difference.
  - Compare χ^2 to determine particular datasets showing differences.
  - Compare cross-sections and point-by-point theory predictions.

PDF Benchmarking: Datasets

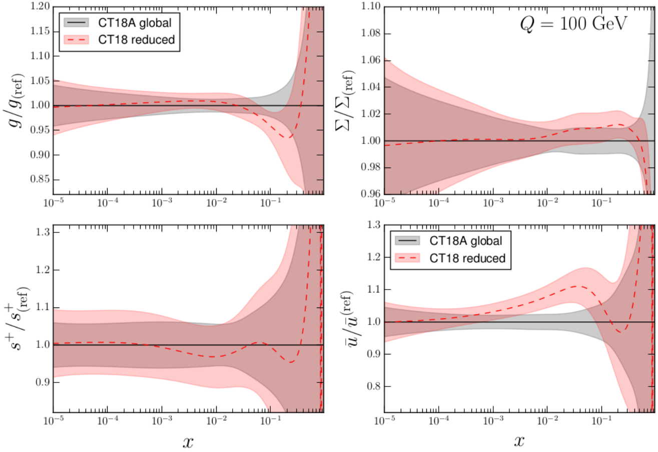
- Chosen subset of datasets fit by all 3 groups in (almost) the same way, list is surprisingly small! Small reduced fit set.
- Take most conservative cuts applied by any group for consistency.
- Ensure enough datasets and a sufficient variety of dataset types are fit to have some (but incomplete) constraints on all PDF flavours.
- Overall list:
  - NMC deuteron to proton ratio in DIS.
  - NuTeV dimuon cross-sections.
  - HERA I+II inclusive cross-sections from DIS.
  - E866 fixed target Drell-Yan ratio pd/pp data.
  - D0 Z rapidity distribution.
  - ATLAS W, Z 7 TeV rapidity distribution, only Z peak and central.
  - CMS 7 TeV W asymmetry.
  - CMS 8 TeV inclusive jet data.
  - LHCb 7, 8 TeV W, Z rapidity distributions.
  - BCDMS proton and deuteron DIS data.

PDF Benchmarking: Theory Settings

- Choose common theory settings for simplicity:
  - Same heavy quark masses ($m_c = 1.4$GeV, $m_b = 4.75$GeV) and $\alpha_S(M_Z^2) = 0.118$.
  - No strangeness asymmetry at input scale: $(s - \bar{s})(Q_0) = 0$.
  - Perturbative charm.
  - Positive definite quark distributions (lack of constraint may allow negative fluctuations).
  - No deuteron or nuclear corrections.
  - Fixed branching ratio for charm hadrons to muons.
  - NNLO corrections for dimuon data.
- Note: These are not the chosen settings for any one group, but rather are a compromise to the least common denominator in each case, *we would not recommend them for a full global fit*.

Reduced Fits: CT18 changes - central values

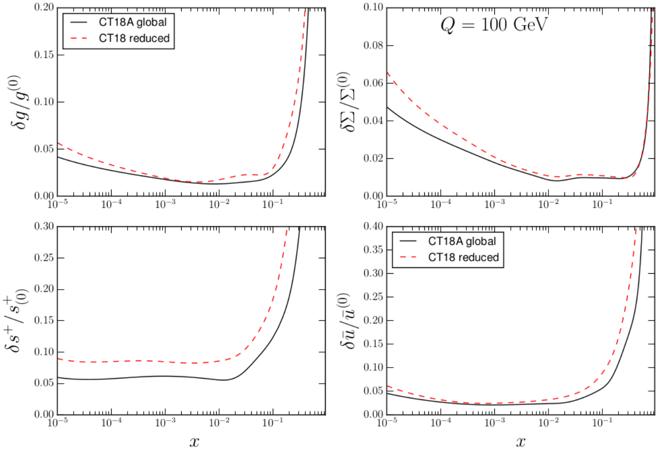
- Current Status:



- Good compatibility with some increase in $\bar{u}$, and change in high x gluon shape. Some changes in flavour decomposition.

Reduced Fits: CT18 changes - uncertainties

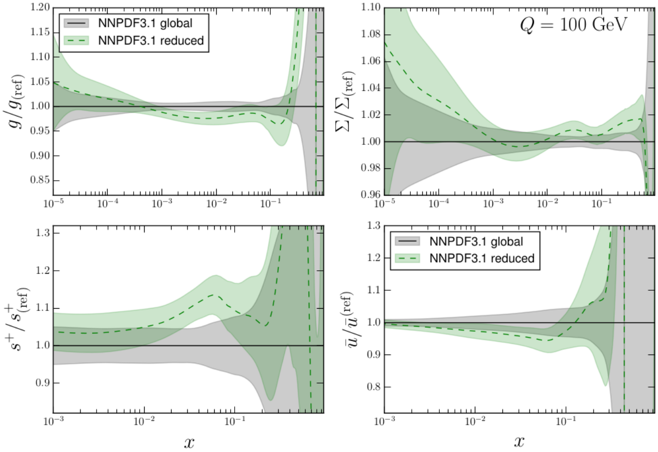
- Current Status:



- Some increase in *nominal* PDF uncertainties, particularly at low x.

Reduced Fits: NNPDF3.1 changes - central values

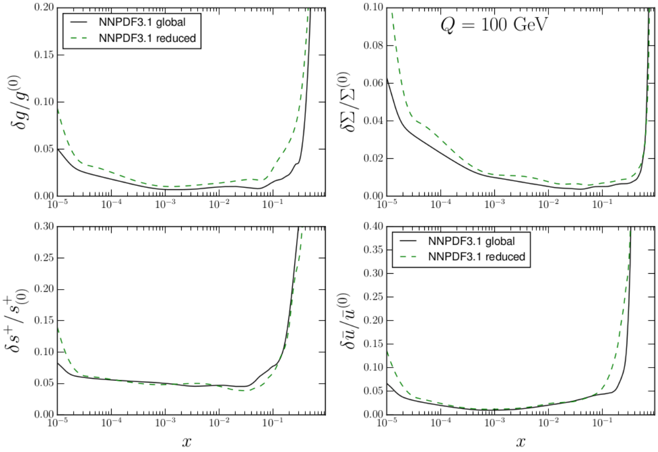
- Current Status:



- Good compatibility, changes in strangeness (see later) and change in large x gluon (removal of top data, addition of CMS 8 TeV jet).

Reduced Fits: NNPDF3.1 changes - uncertainties

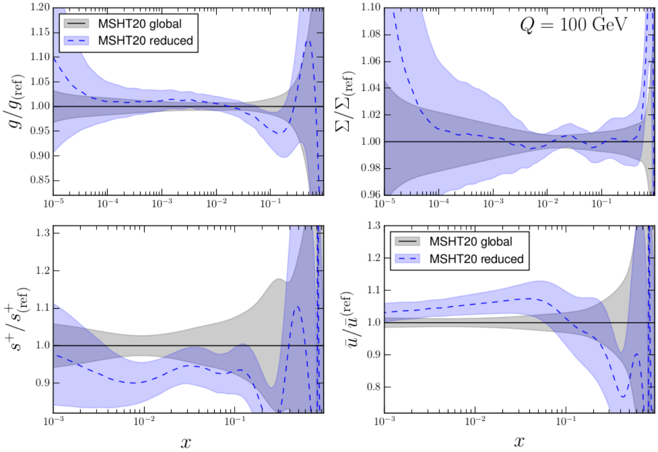
- Current Status:



- Generally slightly increased uncertainties, particularly for the gluon.

Reduced Fits: MSHT20 changes - central values

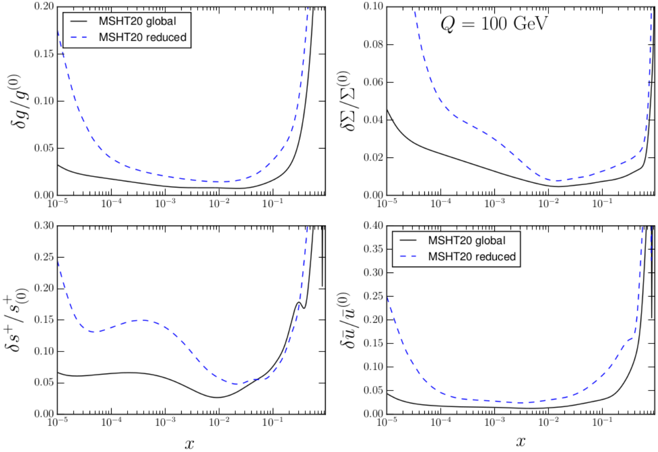
- Current Status:



- Good compatibility, changes in strangeness (removal of 8 TeV ATLAS W, Z data), flavour decomposition and large x gluon.

Reduced Fits: MSHT20 changes - uncertainties

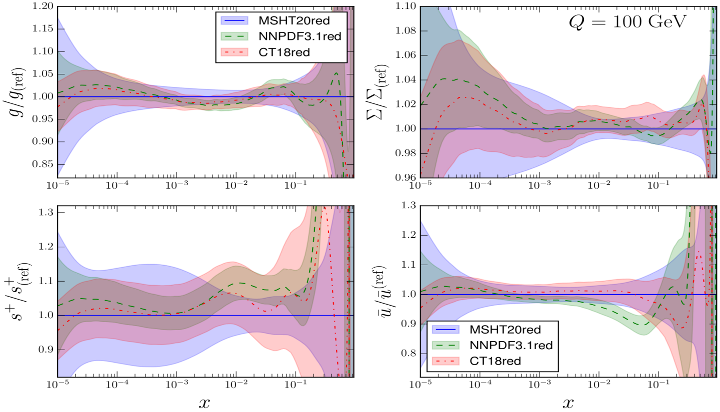
- Current Status:



- General marked increase in uncertainties of reduced fit, particularly outside of regions where there are data.

Reduced Fits PDF Comparison - central values

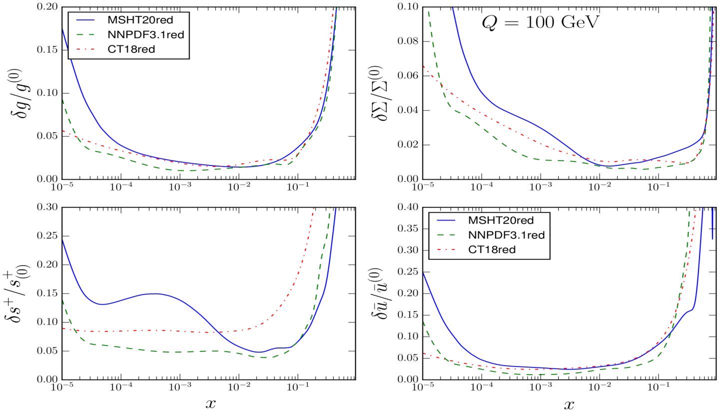
- Current Status:



- Good general agreement within uncertainties, perhaps with the exception of high x flavour decomposition of NNPDF.
- Nonetheless, strangeness and flavour decomposition improved through benchmarking (NuTeV - later). *Note this is without the $t\bar{t}$ added.

Reduced Fits PDF Comparison - uncertainties

- Current Status:



- Similar size uncertainties in data regions, MSHT generally larger errors where constraints lacking in reduced fit.
- Parallel study into differences in uncertainty bands ongoing.

*Note this is without the $t\bar{t}$ added.

PDF4LHC15 in Predictions Datasets χ^2 Comparison

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

Table from T. Hobbs

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††/325†	1.35	1.2	1.51
102	BCDMS F_2^d	246/151††/244†	0.97	1.27	1.24
104	NMC F_2^d/F_2^p	118/117†	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††	1.53	1.89	1.78
Total	N_{pt}	—	2263	1991	2256
Total	χ^2/N_{pt}	—	1.31	1.36	1.62

- 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.

Reduced Fits Datasets χ^2 Comparison

- Current status:

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

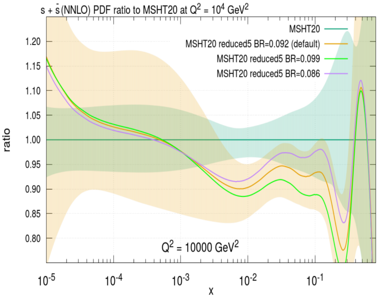
- Similar overall quality of fit in χ^2/N.
- Differences remaining in some datasets:
  - NuTeV agreement improved but difference remains, seen in $s + \bar{s}$.
  - Some potential differences in ATLAS 7 TeV W, Z and LHCb?
  - Some differences in NNPDF fit quality to small datasets.

Table from T. Hobbs

Flavour Decomposition - Strangeness and NuTeV

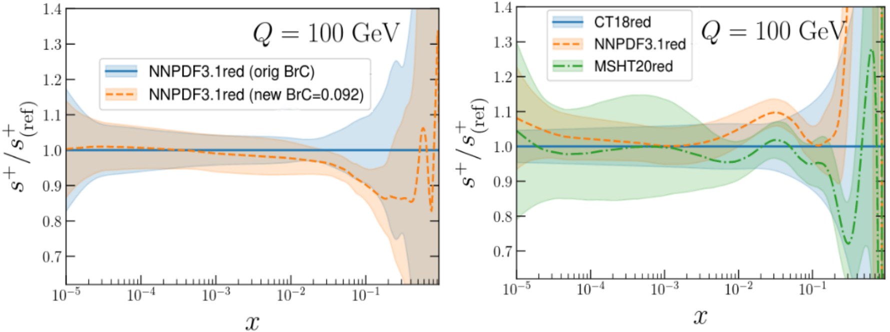
- One of the main differences between the first reduced sets was in the flavour decomposition and strangeness.
- NuTeV dimuon data key driver of this, complicated dataset:
  - Requires knowledge of charm $\rightarrow$ hadrons branching ratio (BR).
  - Non-isoscalar nature of target.
  - Prefers non-zero strangeness asymmetry.
  - Acceptance corrections required.
- BR($c \rightarrow \mu$) anti-correlated with total strangeness, 3 groups have different default values:
  - NNPDF 0.086 ± 0.05
  - MSHT 0.092 ± 0.1 variable.
  - CT 0.099, normalisation uncertainty.
- MSHT20 reduced fit χ^2/N:

Dataset /BR	0.086	0.092	0.099
NuTeV Dimuon	58.8/71	49.6/71	68.5/71
ATLAS 7 TeV W, Z	60.8/34	65.1/34	57.1/34



Flavour Decomposition - Strangeness and NuTeV

- Setting all variables the same in all 3 fits - same Dimuon BR fixed at 0.092, all treat non-isoscalarity, same acceptance corrections.
- NNPDF strangeness reduced as expected, CT strangeness increases.



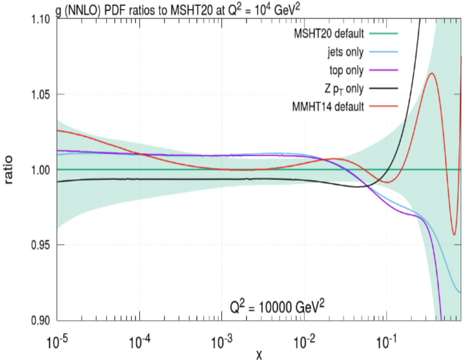
- Better strangeness agreement, certainly in data region, now largely within uncertainties between all 3 groups.
- Also aids reduction in flavour decomposition differences.

Plots from J. Rojo

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.



- 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?*

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

- Comes differential in 4 variables with statistical and systematic 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.
- NNPDF however able to fit all 4 distributions well individually †.

Benchmarking:

- Start by adding this to the reduced fit, first check theory predictions for PDF4LHC15 read in (no fitting):
  - Data agree and theory agrees to better than 1%.
  - All groups χ^2 in agreement and follow same pattern:

Distribution/N	MSHT	CT	NNPDF
$p_t^T/8$	3.0	3.1	3.4
$y_t/5$	10.6	10.1	9.5
$y_{tt}/5$	17.6	15.3	16.2
$m_{tt}/7$	4.3	4.2	4.1

  - Differences in global fits likely not from $t\bar{t}$ theory implementations.

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

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

- *What happens when this dataset is added to the reduced fits?*
- Two cases considered - "uncorrelated" (all systematic and statistical correlations between distributions turned off) and "correlated" (including all correlations, produces a very poor fit):

Distribution/N	$p_t^T/8$	$y_t/5$	$y_{tt}/5$	$m_{tt}/7$	Total
MSHT uncorrelated	3.9	8.4	12.5	6.4	31.2
NNPDF uncorrelated	7.2	3.9	5.1	2.5	18.7
CT uncorrelated	3.4	12.9	17.3	6.1	39.7
MSHT correlated	-	-	-	-	130.6
NNPDF correlated	-	-	-	-	122.7
MSHT decorrelated	-	-	-	-	35.3

- MSHT observe usual pattern as in global fits, p_t^T and m_{tt} can be fit but y_t, y_{tt} struggle, although better than in full fit. Awful fit if all correlations included, can fit with parton shower decorrelation.
- CT see usual global fit pattern also, poor fits to rapidities y_t, y_{tt}.
- NNPDF however able to fit rapidity distributions in uncorrelated case, yet correlated case similar to MSHT.

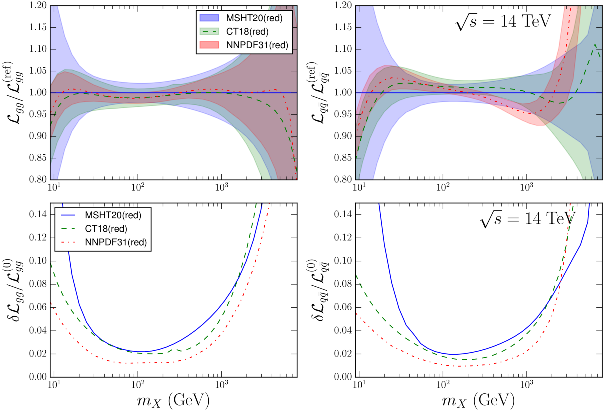
Benchmarking ATLAS 8 TeV $t\bar{t}$ lepton+jets

Preliminary!

- Potential explanations lie in other datasets included - tensions?
- NNPDF-4.0, will have much more jet data, sees similar issues as MSHT, CT, ATLAS for this dataset.
- Useful to consider different jet datasets as well as CMS 8 TeV jets.
- Potential further explanation is division of training and validation data in NNPDF - training fraction 50% but for small datasets such a division is unfeasible - all data in training.
- Potentially double-weights small datasets - e.g. ATLAS $t\bar{t}$.
- May also explain NNPDF better fit of E866 DYratio data and CMS W charge asymmetry data (15 and 11 points respectively):

Dataset	MSHT uncorrelated	NNPDF uncorrelated	MSHT uncorrelated double weight
Total	2314.1	2731.4	2313.3
χ^2/N	1.15	1.20	1.15
DYratio (15)	9.5	5.2	9.2
CMS W asym. (11)	14.2	8.2	10.2
p_t^T (8)	3.8	7.2	4.2
y_t (5)	8.4	4.3	5.8
y_{tt} (5)	12.5	5.7	7.4
m_{tt} (7)	6.4	2.4	6.5
$t\bar{t}$ total	31.2	19.6	23.9

Reduced Fits: Current Status Summary*



- Very good agreement in gluon-gluon, quark-quark and quark-gluon luminosities. (Latter two in backup slides).
- Differences in quark-antiquark luminosity, still some flavour decomposition differences, although within MSHT uncertainties.

*Note this is without the $t\bar{t}$ added.

Conclusions and Future Work

- Fitting reduced datasets checks consistency between different groups, provides an environment to analyse origin of differences.
- Good overall consistency is now observed in the reduced fits between the three groups involved, particularly in luminosities.
- Sources of differences in strangeness in reduced fits largely identified, currently analysing high x gluon region of interest.
- Some differences in flavour decomposition and uncertainties of reduced fits remain, mainly outside of data regions.
- Several possibilities for further investigations:
  - Continue analysis of high x gluon differences, consider different jet datasets - and their effects on both the ATLAS $t\bar{t}$ and the gluon.
  - Ongoing effort to consider different uncertainties.
  - Expand reduced fits towards global fits.
- Overall very good progress towards benchmarking the global fits and beyond that eventual combination $\Rightarrow$ PDF4LHC21.

Many thanks to all those involved in this work/discussions, special thanks to T. Hobbs, T.-J. Hou, L. Harland-Lang, P. Nadolsky, E. Nocera, J. Rojo, R. Thorne for providing tables/plots/fits.

Backup Slides

Introduction - New Datasets (MSHT20)

LHCb W, Z data at high rapidity

CMS W+c

Precision DY data

⇒ Flavour Decomposition

LHC Jet, Zp_T, $t\bar{t}$ data

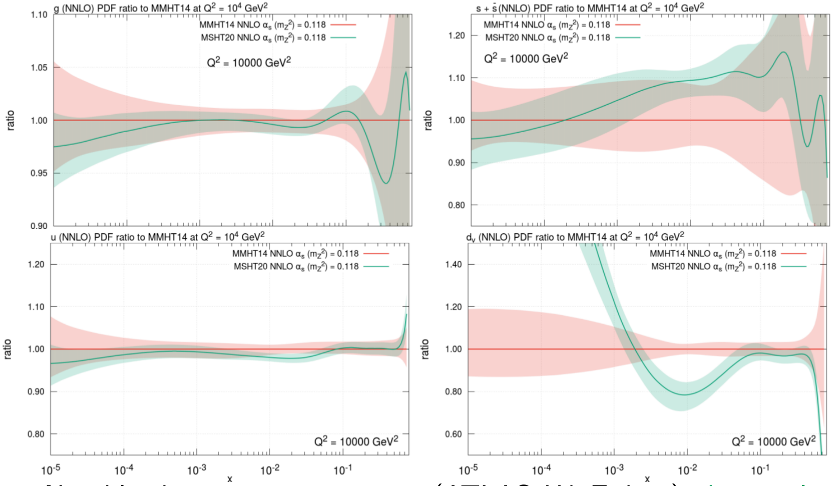
⇒ High x gluon

Data set	Points	NLO χ^2/N_{pts}	NNLO χ^2/N_{pts}
DØ W asymmetry	14	0.94 (2.53)	0.86 (14.7)
$\sigma_{t\bar{t}}$ [93]- [94]	17	1.34 (1.39)	0.85 (0.87)
LHCb 7+8 TeV $W + Z$ [95, 96]	67	1.71 (2.35)	1.48 (1.55)
LHCb 8 TeV $Z \to ee$ [97]	17	2.29 (2.89)	1.54 (1.78)
CMS 8 TeV W [98]	22	1.05 (1.79)	0.58 (1.30)
CMS 7 TeV $W + c$ [99]	10	0.82 (0.85)	0.86 (0.84)
ATLAS 7 TeV jets $R = 0.6$ [18]	140	1.62 (1.59)	1.59 (1.68)
ATLAS 7 TeV $W + Z$ [20]	61	5.00 (7.62)	1.91 (5.58)
CMS 7 TeV jets $R = 0.7$ [100]	158	1.27 (1.32)	1.11 (1.17)
ATLAS 8 TeV Z p_T [75]	104	2.26 (2.31)	1.81 (1.59)
CMS 8 TeV jets $R = 0.7$ [101]	174	1.64 (1.73)	1.50 (1.59)
ATLAS 8 TeV $t\bar{t} \to l + j$ sd [102]	25	1.56 (1.50)	1.02 (1.15)
ATLAS 8 TeV $t\bar{t} \to l^+l^-$ sd [103]	5	0.94 (0.82)	0.68 (1.11)
ATLAS 8 TeV high-mass DY [73]	48	1.79 (1.99)	1.18 (1.26)
ATLAS 8 TeV W^+W^-+ jets [104]	30	1.13 (1.13)	0.60 (0.57)
CMS 8 TeV $(d\sigma_{\bar{t}t}/dp_{T,t}dy_t)/\sigma_{t\bar{t}}$ [105]	15	2.19 (2.20)	1.50 (1.48)
ATLAS 8 TeV W^+W^- [106]	22	3.85 (13.9)	2.61 (5.25)
CMS 2.76 TeV jets [107]	81	1.53 (1.59)	1.27 (1.39)
CMS 8 TeV $\sigma_{\bar{t}t}/dy_t$ [108]	9	1.43 (1.02)	1.47 (2.14)
ATLAS 8 TeV double differential Z [74]	59	2.67 (3.26)	1.45 (5.16)
Total, LHC data in MSHT20	1328	1.79 (2.18)	1.33 (1.77)
Total, non-LHC data in MSHT20	3035	1.13 (1.18)	1.10 (1.18)
Total, all data	4363	1.33 (1.48)	1.17 (1.36)

- Lots of new information constraining PDFs.

MSHT20, 2012.04684

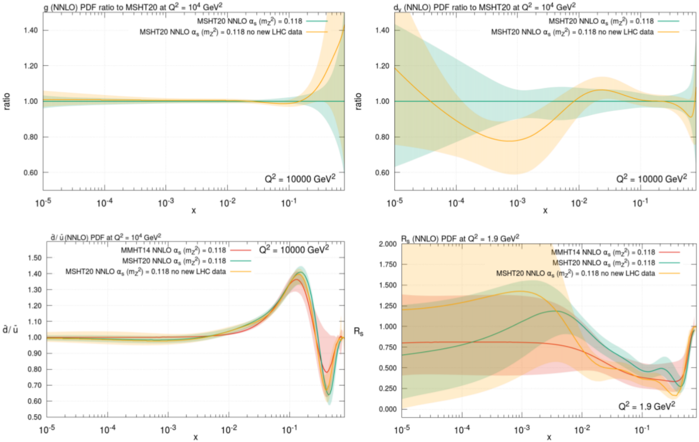
Introduction - Changes in PDFs: MSHT20



- Notable changes in strangeness (ATLAS W, Z data), down valence (new data and parameterisation), gluon (new jets, top, Zp_T data).

More details in R. Thorne's MSHT20 talk.

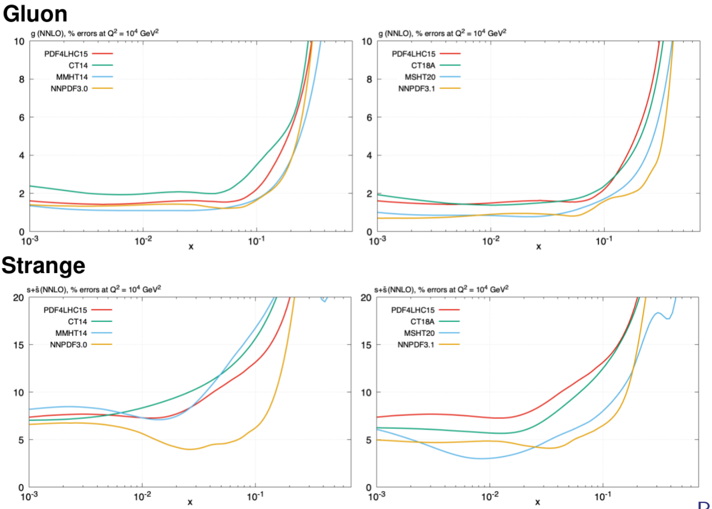
Effect of new LHC data in MSHT20



Main effect on details of flavour, i.e. d_V shape, increase in strange quark for $0.001 < x < 0.3$ and $\bar{d}, \bar{u}$ details, though also partially from parameterisation change. Decrease in high-x gluon.

*MSHT20 2012.04684. Slide from R. Thorne

Introduction - Changes in PDFs: Uncertainties

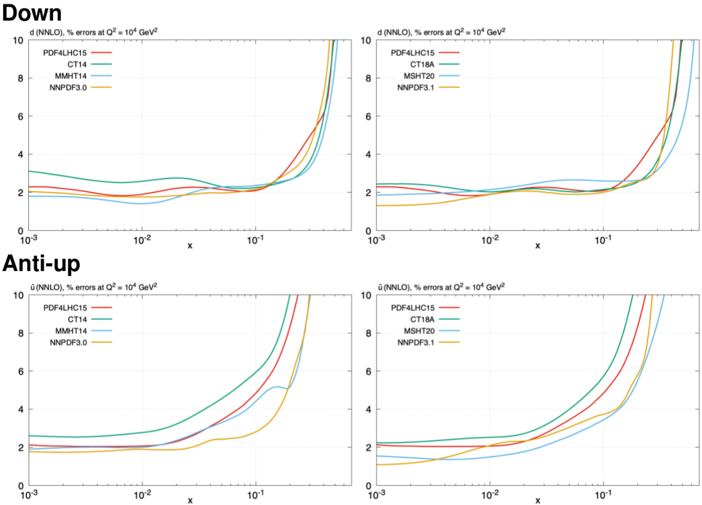


Plots from L. Harland-Lang

- Reduction in PDF uncertainties seen across all 3 groups.

Note: CT18A shown for ease of comparison, however CT18 is the default set.

Introduction - Changes in PDFs: Uncertainties



Plots from L. Harland-Lang

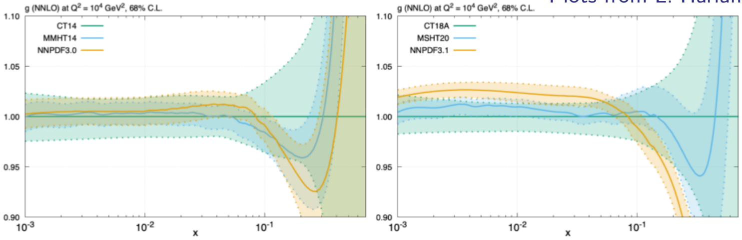
- Reduction in PDF uncertainties seen across all 3 groups.

Note: CT18A shown for ease of comparison, however CT18 is the default set.

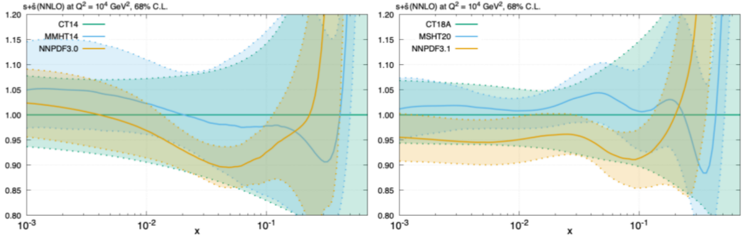
Introduction - Changes in PDFs: Central Values

Gluon

Plots from L. Harland-Lang



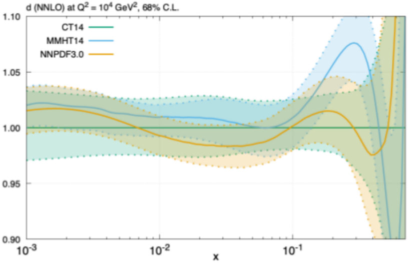
Strange

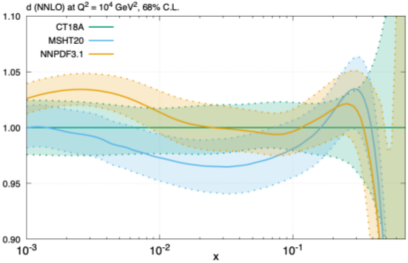


- Central value agreement not as good, some differences emerging.

Introduction - Changes in PDFs: Central Values

Down





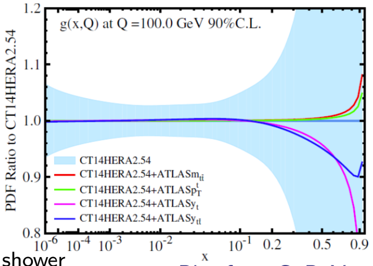
- Central value agreement not as good, some differences emerging.
- In summary:
  - Large amount of progress since the last PDF4LHC combination on experimental, theoretical and methodological fronts.
  - Some differences emerging between the 3 sets.

$\Rightarrow$ now is a good time to undertake a benchmarking exercise ahead of a new PDF4LHC future combination.

Plots from L. Harland-Lang

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

- MSHT*, find difficulties fitting all 4 distributions - $m_{tt}, y_t, y_{tt}, p_t^T$ - simultaneously. CT find same and fit only p_t^T and m_{tt} together.
- MSHT, CT+, ATLAS− cannot get good fit to y_t or y_{tt} individually.
- NNPDF however able to fit all 4 distributions well individually †.
- Different pulls observed for m_{tt}, p_t^T relative to y_t, y_{tt}:
- CT, MSHT decorrelate parton shower systematic to obtain reasonable fit to p_t^T and m_{tt} for former or all 4 for latter:



Plot from C.-P. Yuan

p_T	0.53
y_t	3.12
y_{tt}	3.51
M_{tt}	0.70
$p_T + M_{tt}$	5.73
Combined	7.00

Decorrelate parton shower (within and between) →

Distribution	p.s. correlated	p.s. decorrelated
Combined	7.00	1.80
$p_\perp^t + M_{tt}$	5.73	0.66

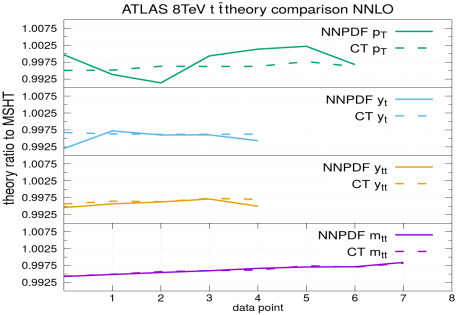
* S. Bailey & L.Harland-Lang 1909.10541.
† Czakon et al 1611.08609.
+ Kadir et al 2003.13740.
− ATL-PHYS-PUB-2018-017.

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

- Start by adding this to the reduced fit, first check theory predictions for PDF4LHC15 read in (no fitting).
- Differences noted in data treatment due to shifting (MSHT) to centre of asymmetric errors, differences in theory due to inclusion (MSHT) or not (CT,NNPDF) of EW corrections.
- Upon removal of these differences, data agree and theory agrees to better than 1%.
- All groups χ^2 in agreement and follow same pattern:

Distribution/N	MSHT	CT	NNPDF
$p_t^T/8$	3.0	3.1	3.4
$y_t/5$	10.6	10.1	9.5
$y_{tt}/5$	17.6	15.3	16.2
$m_{tt}/7$	4.3	4.2	4.1

- Differences in global fits likely not from $t\bar{t}$ theory implementations.



ATLAS 8 TeV multi-differential $t\bar{t}$ lepton+jets: MSHT20*

- MSHT observe the rapidity y_t and y_{tt} distributions have very poor fit quality even when fit alone.
- Moreover, fitting the p_t^T and m_{tt} together or all 4 datasets combined results also in a very poor fit:

p_T	0.53
y_t	3.12
y_{tt}	3.51
M_{tt}	0.70
$p_T + M_{tt}$	5.73
Combined	7.00

Decorrelate parton shower (within and between) →

Distribution	p.s. correlated	p.s. decorrelated
Combined	7.00	1.80
$p_\perp^t + M_{tt}$	5.73	0.66

- Tensions exists between shifts required for large systematics of the different distributions, particularly parton shower uncertainty (and ISR/FSR and hard scattering systematics).
- Two-point systematic evaluated using 2 Monte Carlo generators, assuming any correlation factor determined applies fully correlated way across all bins and distributions is a strong assumption.

* S. Bailey & L.Harland-Lang 1909.10541 and MSHT20 2012.04684.

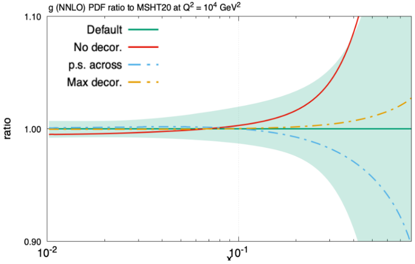
ATLAS 8 TeV multi-differential $t\bar{t}$ lepton+jets: MSHT20*

- Assumption of full correlation of parton shower systematic can be relaxed, then a resonable fit is possible.
- CT decorrelate this systematic between distributions and fit the p_t^T and m_{tt} combination only by default †.
- MSHT do this decorrelation between all 4 distributions and also split it into 2 sources varying smoothly within each distribution:

$$\beta_i^{(1)} = \cos\left[\pi\left(\frac{y_{tt,i} - y_{tt,\min}}{y_{tt,\max} - y_{tt,\min}}\right)\right]\beta_i^{\rm tot}, \qquad \beta_i^{(2)} = \sin\left[\pi\left(\frac{y_{tt,i} - y_{tt,\min}}{y_{tt,\max} - y_{tt,\min}}\right)\right]\beta_i^{\rm tot}.$$

- Then a reasonable fit is possible, e.g. in MSHT20:

Baseline	No decor.	parton shower across	Max decor.
1.04	6.84	1.69	0.81

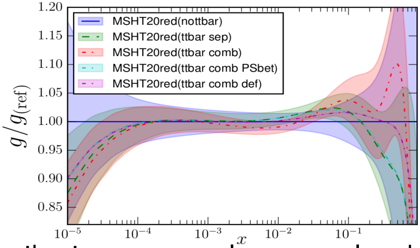


* S. Bailey & L.Harland-Lang 1909.10541 and MSHT20 2012.04684.

† T.-J. Hou et al, CT18 1912.10053.

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

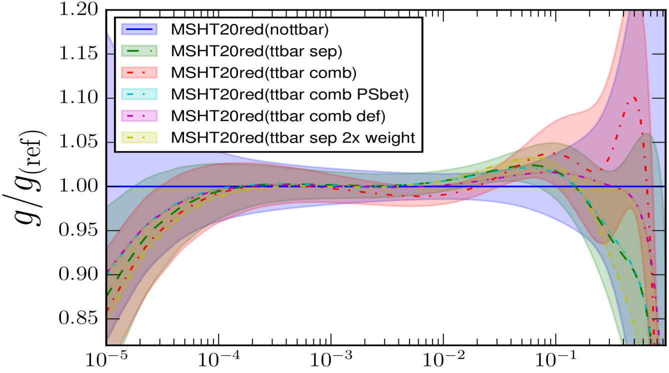
- What effect does the inclusion of this data in the reduced fit have on the gluon?



- Fitting all 4 distributions separately, uncorrelated ⇒ gluon moves down at high x, driven by the rapidity data.
- Applying correlations ⇒ gluon raised and shape altered at high x .
- Decorrelating parton shower between distributions ⇒ reverts the gluon to shape obtained when all 4 separately uncorrelated fitted.
- Additionally decorrelating within distributions ⇒ moves gluon closer to fit without $t\bar{t}$ data as its constraining power is reduced.
- Overall, gluon shape moves in direction of global fit gluon.

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

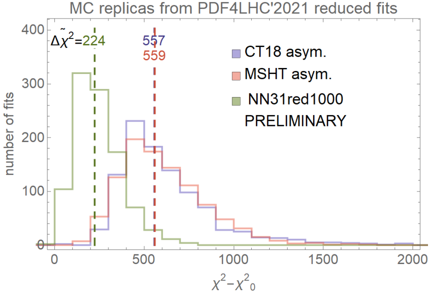
- What effect does the inclusion of this data in the reduced fit have on the gluon?

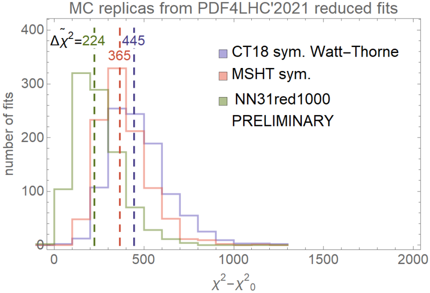


- Double weighting (yellow) pulls gluon further in direction of rapidity pull (lower at high x) as expected.

Reduced Fits χ^2 replica distributions

- As well as the overall PDFs, can analyse χ^2.
- Can use CT/MSHT reduced fit eigenvectors to generate replicas.

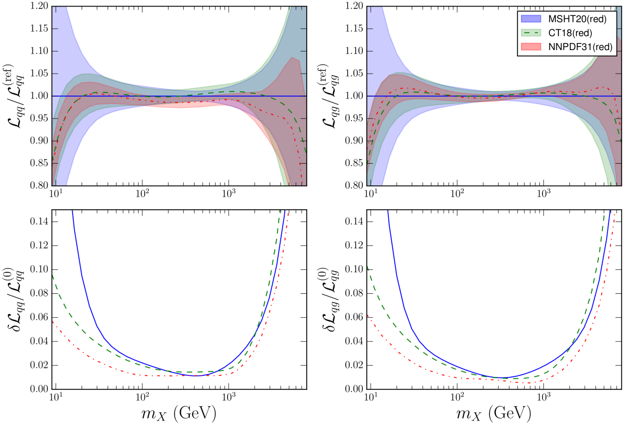




- Overall distributions of reduced fits replicas similar between CT and MSHT, particularly non-symmetrised versions.
- Symmetrised versions in better agreement with NNPDF but still different.
- Some limited qualitative agreement at least for symmetrised case.

Plots from P. Nadolsky

Reduced Fits: Current Status Summary*



- Very good agreement in the gluon-gluon, quark-quark and quark-gluon luminosities.
- Differences in quark-antiquark luminosity, still some flavour decomposition differences, although within MSHT uncertainties.

*Note this is without the $t\bar{t}$ added.

Deuteron and Nuclear Corrections in MSHT20

- Several older DIS datasets use deuteron or heavy nuclear targets.
- Deuteron data required to fully separate u, d at moderate-large x.
- Heavy nuclear data, via C.C. scattering, required for more constraints on flavour decomposition and strange (dimuon data).
- Deuteron correction is 4-parameter prefactor to usual average of p and n:

$$F^d(x,Q^2) = c(x)\left[F^p(x,Q^2)+F^n(x,Q^2)\right]/2,$$

$$\begin{aligned} c(x) &= (1+0.01N)\,[1+0.01c_1\ln^2(x_p/x)], & x<x_p, \\ c(x) &= (1+0.01N)\,[1+0.01c_2\ln^2(x/x_p)+0.01c_3\ln^{20}(x/x_p)], & x>x_p, \end{aligned}$$

- Nuclear correction is prefactor*: *de Florian et al arXiv:1112.6324.

$$f^A(x,Q^2) = R_f(x,Q^2,A)\; f(x,Q^2).$$

- This is multiplied by a 3-parameter modification function to allow penalty-free change in shape and/or normalisation.
- Both deuteron and nuclear corrections prefer modifications of 1%.

More details on all of this in MMHT14 1412.3989, MSHT20 2012.04684.

Future Work

- Several possibilities for further investigations:
  - Continue analysis of high x gluon differences, consider different jet datasets - and their effects on both the ATLAS $t\bar{t}$ and the gluon.
    $\Rightarrow$ Of interest to both reduced fit benchmarking and global fits.
  - Further work perhaps to identify source of flavour decomposition differences?
  - Ongoing effort to consider different uncertainties - CT implementing MSHT dynamic tolerance to analyse reduced fit uncertainty differences. Could repeat for NNPDF by generating Hessian set.
  - Expand reduced fits towards global fits to see how any remaining differences alter.
- Once this is done we can consider a combination of the global fit PDFs $\Rightarrow$ PDF4LHC21.

Conclusions

- Fitting reduced datasets is a good mechanism to check consistency between different groups, provides an environment where any differences in treatment are more relevant to the overall PDFs.
- Allows analysis of origin of differences ⇒ are they due to different, equally valid choices in theoretical, methodological or data approaches.
- Good overall consistency found between the three groups involved, particularly in luminosities.
- Sources of differences in strangeness in reduced fits largely identified, currently analysing high x gluon region of interest.
- Some differences in flavour decomposition and uncertainties remain, mainly outside of data regions.
- Overall very good progress towards benchmarking the global fits and beyond that eventual combination.

Many thanks to all those involved in this work/discussions, special thanks to T. Hobbs, T.-J. Hou, L. Harland-Lang, P. Nadolsky, E. Nocera, J. Rojo, R. Thorne for providing tables/plots/fits.