PDF4LHC21: Update on the benchmarking of the CT, MSHT and NNPDF global PDF fits

Thomas Cridge[1] on behalf of the PDF4LHC21 combination group.

1 Department of Physics and Astronomy, University College London, London, WC1E 6BT, UK
t.cridge@ucl.ac.uk

July 30, 2021

*Proceedings for the XXVIII International Workshop
on Deep-Inelastic Scattering and Related Subjects,
Stony Brook University, New York, USA, 12-16 April 2021*

Abstract

There have been recent updates to the three global PDF fits (CT, MSHT and NNPDF), all
adding large amounts of data from the LHC, and this has resulted in significant changes
to the global PDFs. Given the impact that the new PDFs will have on physics comparisons
at the LHC, it is crucial to perform a benchmarking among the PDFs, similar in spirit to
that which was carried out for PDF4LHC15, widely used for LHC physics. In this article
we detail a benchmarking comparison of three global PDF sets - CT18, MSHT20 and
NNPDF3.1 - and their similarities and differences that have been observed. The end
result of this study will be a new PDF4LHC21 ensemble of combined PDFs suitable for a
wide range of LHC applications.

Introduction

The knowledge and determination of the parton distribution functions (PDFs) of the proton is a key requirement for the physics program at the Large Hadron Collider (LHC) and future colliders. Both the central values of the PDFs and their uncertainties play an important role in many analyses. As a result, the accurate and precise evaluation of PDFs from available experimental data is of paramount importance. Over the past few years there have been many developments in this area, with the inclusion of processes at NNLO in PDF fits, a great variety of new data from the LHC now included and continuing developments of the PDF fitting methodologies. Consequently, the PDFs are now known more accurately and precisely than ever before, with the global fitting groups having recently provided major updates and new PDF sets including these advances [1–3].

However, with these advances, differences have begun to emerge in both the central values and the uncertainties of the PDFs extracted by different groups. As the experimental and PDF uncertainties continue to reduce, such differences may become a key component of the uncertainties unless they can be understood, particularly in the context of a new PDF4LHC combination set. Moreover, with statistical precision increasing for many measurements, challenges have arisen in fitting some of the datasets available, and differences in the fitting methodology (e.g. in the treatment of correlated systematics) may have effects on the PDFs. Given therefore that it has been several years since the last major benchmarking effort of global PDF fits in PDF4LHC15 [4] (itself the last in a line of several efforts including [5–8]), the time is ripe for a study of these differences. We present such a benchmarking exercise in these proceedings. The need is made all the more pressing by the upcoming LHC Run III, which will provide yet further information constraining the PDFs and also utilise the PDFs as a key input in physics analyses.

In these proceedings we describe the progress made so far in benchmarking three of the global PDF fits: CT, MSHT and NNPDF, with our aims being first to identify differences where they arise and then to understand the origin of these differences through their detailed examination. Ultimately these comparisons will lead to an improved understanding of the differences in the PDFs and enable the production of a new PDF4LHC combination ensemble, to be reported in a future publication. This will be a combination of the most recent versions of the global fits of the three groups: CT18, MSHT20 and NNPDF3.1[1] in the style of PDF4LHC15, reflecting the current knowledge of the PDF central values, their uncertainties and spread for use in the LHC at Run III and beyond, as well as wherever else is appropriate.

Approach to Benchmarking

In order to determine any differences in the PDFs, we seek to begin with a baseline of PDF fits from the three groups where as many differences in input and approach are removed as possible, this allows the benchmarking to focus upon salient differences. We therefore choose both a common input set of datasets to include and common theory settings. This produces the so-called *"reduced fits"*, offering ease of comparison at the expense of the usual robustness offered by global fits.

We begin with the choice of input data, to set up the reduced fits we keep only the largest subset of data fit by all three groups in an (almost) identical manner. Furthermore, we apply the most conservative cuts made by any group, i.e. $Q^2 > 4\text{GeV}^2$ and $W^2 > 15\text{GeV}^2$. Given numerous subtle differences between the groups, the final list of common data for the reduced

---

[1]Specifically, it will be an updated and modified version of NNPDF3.1, similar to that used in the NNPDF study of the strangeness [9].

Dataset	Reference	Dataset	Reference
BCDMS proton, deuteron DIS	[10]	LHCb 8TeV $Z \to ee$	[11]
NMC deuteron to proton ratio DIS	[12]	ATLAS 7 TeV high precision W, Z (2016)	[13]
NuTeV νN dimuon	[14]	D0 Z rapidity	[15]
HERA I+II inclusive DIS	[16] [17]	CMS 7 TeV electron asymmetry	[18]
E866 Drell-Yan ratio pd/pp DIS	[19]	ATLAS 7 TeV W, Z rapidity (2011)	[20]
LHCb 7, 8 TeV W, Z rapidity	[21]	CMS 8 TeV inclusive jet	[22]

Table 1: Datasets included in the initial baseline reduced fits.

fit is rather small and shown in Table 1. This reduced fit set of datasets also satisfies the competing requirement of being sufficiently large and varied so as to provide some constraints on all the PDF combinations and their uncertainties. The constraints placed by this reduced fit dataset will necessarily be significantly more limited than in the usual full global fits, but we trade the robustness of the global fits for the ease of comparison of the reduced fits for the purposes of benchmarking.

With the differences in the input data now removed, we also choose common theory settings wherever possible[2] including: the same heavy quark masses and strong coupling ($m_c^{\text{pole}} = 1.4\text{GeV}$, $m_b^{\text{pole}} = 4.75\text{GeV}$ and $\alpha_S(M_Z^2) = 0.118$); no strangeness asymmetry at the input scale, i.e. $(s - \bar{s})(Q_0) = 0$; perturbative charm; positive definite quark distributions; no deuteron or nuclear corrections; and for the dimuon data we take a fixed branching ratio for charm hadrons to muons $B(D \to \mu)$ (this will be elucidated further later in this article in the section describing the dedicated study of the reduced fit strangeness) and NNLO corrections.

At this point it is worth explicitly noting that neither the reduced fit dataset, nor common theory settings for the reduced fit, correspond to the data or theory settings used by any group, rather they are a compromise to the least common denominator in each case. Indeed, some of the choices made are known to be sub-optimal but we adopt them for the purposes of the benchmarking exercise only. Even when these differences in input and theory are removed, methodological differences remain, including the choice of the General-Mass Variable-Flavour-Number scheme and also the treatment of the PDF uncertainties. These will not be described here but the former has been compared previously in [7,23,24] whilst the latter will be outlined in our future full publication of the benchmarking exercise.

Comparison of Reduced and Global Fits

At this stage, before performing the benchmarking of the reduced fits, it is useful to compare the reduced fits produced by each of the three groups with their default global fits. We show this comparison for CT only for brevity in Figure 1. Overall good compatibility is observed between the central values in the ratio of the CT18 reduced fit to the CT18A global fit[3], with changes in the high-x gluon shape resulting from the reduced amount of jet and other data relevant in this region in the reduced fit. The singlet and strangeness are both compatible within the uncertainties of the global fit, whilst there is an increase in the anti-up PDF at intermediate x, which signals a change in the flavour decomposition in the reduced fit. Such changes are not unexpected given the significantly altered, and specifically reduced, dataset of the reduced fit. In the right-hand plot of Figure 1 the size of the uncertainty bands are compared, clearly there is some increase in the nominal PDF uncertainties of the reduced fit, particularly at low x. Generally, an increase in the PDF uncertainties (as evaluated in an

---

[2]However each group still uses their own version of the theoretical calculation at the appropriate order in QCD.

[3]CT18A is compared against as this includes the ATLAS 7 TeV W, Z data, which is included in the reduced fit.

identical manner to the global fit) is to be expected given the significantly reduced amount of data constraining the PDFs. Similar differences in central values and uncertainties are seen in comparing the MSHT and NNPDF reduced fits with their respective global fits, albeit MSHT sees a reduced strangeness relative to their global fit due to the exclusion of the ATLAS 8 TeV W, Z data [25, 26] from the reduced fit. All reduced fits are compared in the next section where the benchmarking of the reduced fits is performed.
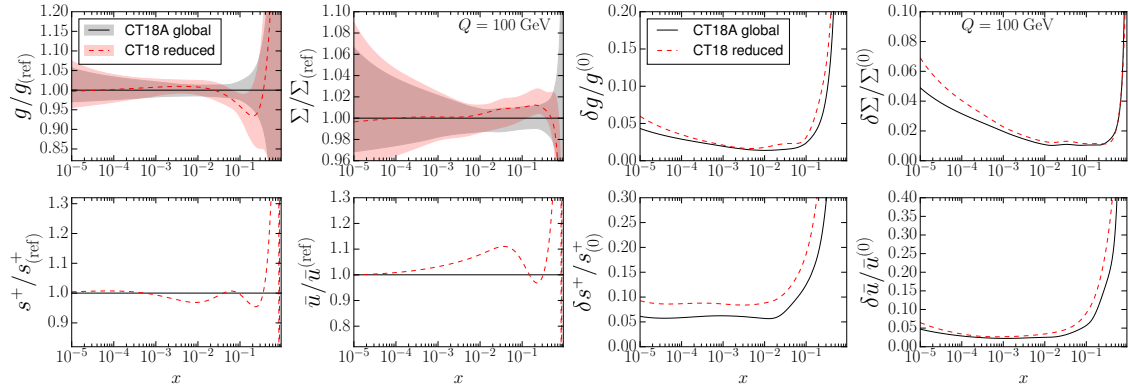
Figure 1: Comparison of the CT reduced fit relative to the CT18A global fit, the ratio of the central values is shown in the left plots and the uncertainties are compared in the right plots. The gluon, singlet (Σ), total strangeness ($s^+ = s + \bar{s}$) and anti-up ($\bar{u}$) are compared. The baseline for the ratio is the CT18A global fit.

Benchmarking of Reduced Fits

Now that we have a gauge of the differences between the reduced fits and the global fits, we begin the benchmarking of the reduced fits. Firstly we compare the reduced fit PDFs themselves directly against one another at the level of both central values and uncertainties, as presented in Figure 2. There is good general agreement between the three reduced fits within uncertainties for the majority of the PDFs over most of the x range. Starting with the gluon, all three groups agree (almost) within uncertainties over the entirety of the x range, this confirms that changes in the high-x gluon shape relative to the global fits are driven by the datasets included. The singlet is also in very good agreement for all x. The strangeness is also largely consistent, albeit the NNPDF3.1 reduced fit is notably high around $10^{-2} \lesssim x \lesssim 10^{-1}$ and outside of the MSHT20 reduced fit uncertainty bands here (although within the CT18 reduced fit uncertainty bands). This difference in the strangeness will be examined later in the dedicated studies on the strangeness of the reduced fits. The NNPDF $\bar{u}$ is lower than both MSHT and CT in the same region, signalling a difference in the high-x flavour decomposition. The uncertainties of the three reduced fits are also similar in size in data regions, with the MSHT reduced fit having larger uncertainties outside of these regions, i.e. where constraints are lacking in the reduced fit - particularly at low x. A parallel study into differences in the uncertainty bands, both the methodologies of their evaluation and differences in the reduced fit uncertainties, is ongoing and will be reported in the future.

In order to further uncover any differences in the reduced fits we may determine the dataset by dataset fit qualities, as given by the χ^2/N_{pts}. Before calculating the fit qualities for the reduced fits, it is useful to compare them before fitting, i.e. with fixed (PDF4LHC15 NNLO) PDFs. This allows any differences of theory or data between the groups to be identified before the complication of fitting is added. Overall, for most of the reduced fit datasets there
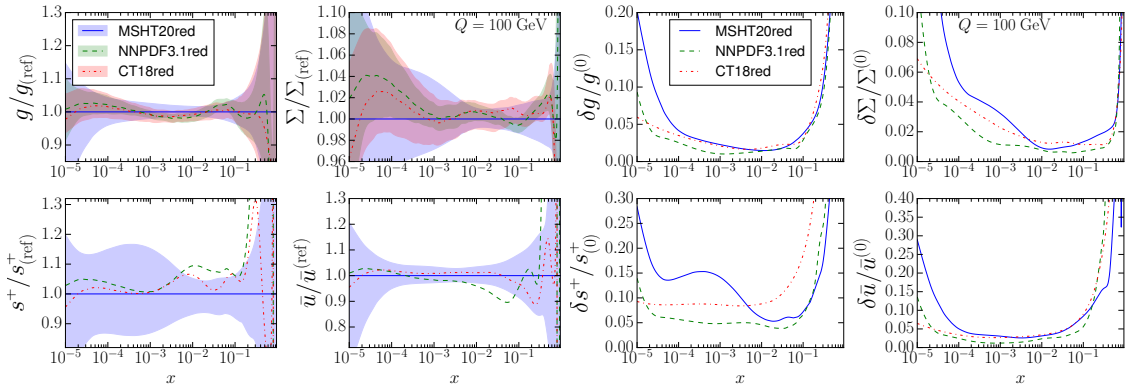
Figure 2: Comparison of the reduced fits of all 3 groups (CT, NNPDF, MSHT), ratio of central values (left) and uncertainties (right), with the baseline for the ratio taken as the MSHT20 reduced fit.

is good agreement observed in the size of the χ^2/N_{pts} before fitting, reflecting agreement in the data/theory comparison for the three groups. Therefore we can progress to comparing the reduced fits' fit qualities after fitting, this is shown in Table 2. Very good overall agreement is observed between the three reduced fits with overall fit qualities in terms of χ^2/N_{pts} of 1.14, 1.15, 1.20 for the CT, MSHT and NNPDF reduced fits respectively. Moreover, this global agreement follows on into the dataset by dataset comparisons with the majority showing good agreement. There are nonetheless some differences, the NuTeV dimuon data shows a noticeable difference in fit quality between CT and MSHT which each observe 0.8 per point, and NNPDF which observes 1.2 per point. This is consistent with the increased strangeness noted earlier in the NNPDF reduced fit in the region $10^{-2} \lesssim x \lesssim 10^{-1}$ as the dimuon data favours lower strangeness here. However this occurs without improvement of the ATLAS 7 TeV W, Z data which pulls the strangeness upwards in this x region. The origin of this difference in the strangeness will be investigated in the first section of the dedicated studies later in this article. An additional difference observed is in the NNPDF fit quality to smaller datasets, such as the E866 Drell-Yan ratio data or the CMS 7 TeV electron asymmetry. For the former there is a known difference between CT and MSHT due to the parameterisation of the $\bar{d}, \bar{u}$ ratio or difference in this high-x region ([1], [2]). For the latter, CT and MSHT both obtain 1.5 per point whereas NNPDF obtains half of this at 0.76, again this will be analysed briefly later in the context of the effect of different weights in the PDF fits in the dedicated study of the high-x gluon.

Dedicated Studies

Overall we have now seen that the reduced fits of the three groups are in good agreement in both the PDFs and fit qualities of most datasets. Nonetheless we highlighted some differences which we will now scrutinise alongside other areas of interest to the global PDF fits.

Strangeness and NuTeV

One of the main differences emerging both in the PDFs and the dataset χ^2 for the reduced fits was in the strangeness, $s + \bar{s}$. The main datasets driving this in the reduced fits and the global fits are the NuTeV dimuon data and the ATLAS 7 TeV W, Z data ([1–3], [9, 27, 28]), with the former preferring reduced strangeness at intermediate x and the latter favouring enhanced strangeness in this region. The NuTeV dataset has several complications, requiring

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[††]/325[†]	1.06	1.00	1.21
102	BCDMS F_2^d	246/151[††]/244[†]	1.06	0.88	1.10
104	NMC F_2^d/F_2^p	118/117[†]	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 $Z \rightarrow ee$	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[††]	1.03	1.39	1.30
Total	N_{pts}	—	2263	1991	2256
Total	χ^2/N_{pts}	—	1.14	1.15	1.20

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

treatment of the non-isoscalar nature of the iron target, acceptance corrections and knowledge of the charm hadrons to muons branching ratio [29], BR(c → μ). The first two of these are treated consistently between the three fits, the last is needed as this process involves the scattering of a muon neutrino off an iron nucleon, this directly produces a muon but also results in the production of a charm hadron which decays to produce a further muon (hence dimuon). By default the three groups use different values for this branching ratio, NNPDF take a PDG value of 0.087 ± 0.005 [30], MSHT take a value of 0.092 ± 0.01 taken from direct measurements [31], [32][4] and further allow the central value to vary with a penalty within this range, and CT take the value of 0.099 used by NuTeV itself [33] with a normalisation uncertainty. However, the total strangeness obtained in the data region $10^{-2} \lesssim x \lesssim 0.4$ is anti-correlated with the branching ratio as it maps from the total strangeness to the NuTeV predictions in the fit. Therefore a smaller branching ratio allows one to fit the same NuTeV data whilst retaining larger strangeness and vice versa. The effect of taking the different branching ratios of the three groups within the MSHT reduced fit is shown in Figure 3, where an increase/decrease in the total strangeness relative to the default value of approximately 5% is observed as the branching ratio is altered. Upon removal of this difference, the NNPDF strangeness reduces whilst the CT strangeness increases as expected, bringing the reduced fits into closer agreement. Given the lack of constraints in a reduced fit, this improved strangeness agreement also enables a reduction in differences in the flavour decomposition.

High-x gluon, top and inclusive jet data

Another area of interest for the reduced fits and global fits is the high-x gluon. We saw in the comparison of the reduced fit gluon in Figure 2 that the gluon was consistent between the three reduced fits. However, this contrasts the situation for the global fits where notable differences are present depending on the datasets included, their treatment and relative weights. This is further complicated by various issues in fitting many of these datasets including difficulties fitting all bins, possible tensions within and between dataset types and issues of correlated systematics. Therefore this is an obvious region that would benefit from benchmarking. There are three main dataset types which play a role in this region, all largely coming from the LHC in the past few years, these are jet data, top data and Z p_T data.

The ATLAS 8 TeV multi-differential $t\bar{t}$ lepton + jets data [34] is one such dataset. This dataset has been studied by all three groups [1], [2], [35–42] and comes differential in four

---

[4]Taking the value from direct measurements avoids a potential circularity in its determination and use.
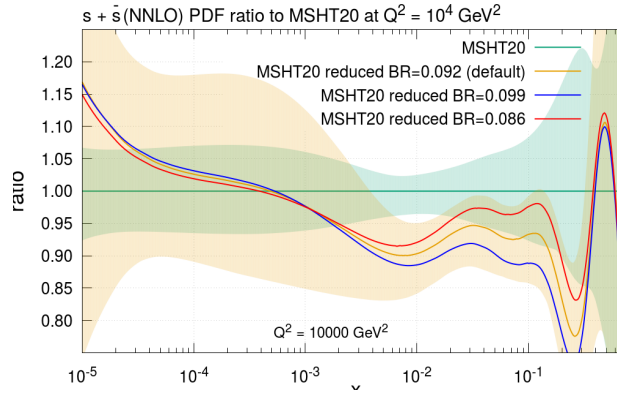
Figure 3: Effect of the charm hadrons to muon branching ratio (BR) $B(D \to \mu)$ on the total strangeness of the MSHT reduced fit, relative to the MSHT20 global fit baseline.

variables: m_{tt}, y_t, y_{tt} and p_t^T, with statistical and systematic correlations provided to enable all four distributions to be fitted together. However several groups, including CT and MSHT, observed difficulties in fitting all four distributions simultaneously, achieving $\chi^2/N_{pts} \gtrsim 5$. Moreover both groups and ATLAS found issues fitting the rapidity distributions even individually [43, 44]. On the other hand, NNPDF3.0 were able to fit all four distributions individually [39][5]. In order to analyse the differences arising here we therefore add this to our reduced fit baseline, this is shown in Table 3. First we check the theory predictions and data for fixed (PDF4LHC15) PDFs, this verifies that the data and theory are in agreement between the three groups. Moreover, at this stage all three groups observe the same pattern in the individual distributions' χ^2, with all groups unable to obtain a good fit to the rapidity distributions.

After fitting however, differences emerge. If all four distributions are added but without any correlations (statistical or systematic) between the distributions - the "uncorrelated case"[6] - then whilst CT and MSHT retain the same behaviour, NNPDF are now able to fit the rapidity distributions well, albeit at the expense of a worsening of the fit quality to the p_t^T. On the other hand, if all statistical and systematic correlations between the four distributions are retained - the "correlated case" - then both MSHT and NNPDF again observe similar behaviour with neither reduced fit able to fit the 4 distributions together, as is seen in the global fits.

There are two potential explanations for these differences. The first is a methodological difference between the PDF fitting groups, NNPDF being based on a neural network, must divide each dataset into training and validation to prevent over-fitting. However, such a split is infeasible for small datasets and so all data are placed in the training, this potentially up-weights smaller datasets, such as the ATLAS $t\bar{t}$ rapidity distributions. To investigate this further a MSHT reduced fit was performed in which small datasets were double-weighted to attempt to approximately mimic this. This reveals a notable improvement in the fit to the ATLAS $t\bar{t}$ y_t and y_{tt} distributions, seen in the fourth row of Table 3, matching the qualitative pattern seen in NNPDF. In addition, beyond the effect on the ATLAS $t\bar{t}$ rapidity distributions, this may also be relevant for some of the differences in fit quality seen in the benchmarking of the reduced fits in Table 2, where it was remarked that the NNPDF reduced fit was able to fit smaller datasets better than the MSHT or CT reduced fits (with a slight trade-off for the fit quality of larger data sets). An example of such a dataset is the CMS 7 TeV electron asymmetry data, when double-weighted its fit quality improves markedly in the MSHT reduced fit - from $\chi^2/N_{pts} \sim 1.3$ to 0.9, more consistent with the $\sim$ 0.75 in the NNPDF reduced fit and so potentially explaining

---

[5]However, the statistical and systematic correlations were not available at the time of the NNPDF3.0 analysis.

[6]This "uncorrelated" case is not how this data should be added to the fit, however it enables the fit qualities of the individual distributions to be observed which offers useful information for understanding the differences.

Distribution/N_{pts}	Group	$p_t^T/8$	$y_t/5$	$y_{tt}/5$	$m_{tt}/7$	Total/25
PDF4LHC15 read in default (CMS 8 TeV jets)	MSHT	3.0	10.6	17.6	4.3	35.5
	CT	3.1	10.1	15.3	4.2	32.7
	NNPDF	3.4	9.5	16.2	4.2	33.3
Uncorrelated case default (CMS 8 TeV jets)	MSHT	3.8	8.4	12.5	6.4	31.2
	CT	3.4	12.9	17.3	6.1	39.7
	NNPDF	7.2	3.9	5.1	2.5	18.7
Correlated case default (CMS 8 TeV jets)	MSHT	-	-	-	-	130.6
	NNPDF	-	-	-	-	122.7
Uncorrelated case double weight (CMS 8 TeV jets)	MSHT	4.2	5.8	7.4	6.5	23.9
Uncorrelated case (no LHC jets)	MSHT	4.5	5.2	6.6	7.4	23.8
Uncorrelated case (CMS 7 TeV jets)	MSHT	4.0	6.4	7.2	6.4	24.0
Uncorrelated case (ATLAS 7 TeV jets)	MSHT	4.6	5.5	5.2	6.4	21.6

Table 3: χ^2 for the ATLAS 8 TeV $t\bar{t}$ lepton + jets distributions when added to the reduced fit, before fitting with PDF4LHC15 PDFs read in, after fitting but including all 4 distributions in an "uncorrelated" manner and after fitting maintaining all statistical and systematic correlations between the distributions. The effect of double weighting small datasets and of the inclusion of different jet datasets in the reduced fit is also shown, with the LHC jet dataset included in brackets.

this difference in the reduced fit comparison.

Nonetheless, for such differences to potentially play a role it suggests they may be changing the balance of pulls between different datasets on the PDFs, therefore possible tensions between datasets in this high-x gluon region are a further explanation. In order to investigate this, reduced fits were run in which the LHC jet data included was varied from the default CMS 8 TeV jets, the effects of this upon the MSHT reduced fit are also shown in Table 3. It is clear that there is tension between the CMS 8 TeV jet data and ATLAS 8 TeV multi-differential $t\bar{t}$ data, once the former is removed the y_t, y_{tt} fit qualities improve significantly to the extent they are now able to be fit well. This is perhaps unsurprising given it is known that the CMS 8 TeV jet dataset tends to pull the gluon up at high x [1, 2, 42], whilst the $t\bar{t}$ rapidity distributions favour a lower gluon in this region [38]. It is therefore broadly consistent with the pulls observed in the full global fits. Similar effects are also observed upon replacement of the CMS 8 TeV jets data in the default reduced fits with the CMS 7 TeV jets or ATLAS 7 TeV jets data, which both favour a reduced gluon in the region of interest for the $t\bar{t}$. As a result these are more consistent with the ATLAS 8 TeV lepton+jet $t\bar{t}$ rapidity distributions, which are then well fit. This demonstrates the effects of tensions in the high-x gluon region on the fit quality of the ATLAS 8 TeV $t\bar{t}$ lepton + jets data. Moreover such tensions are also present within the jet datasets themselves, with the CMS 8 TeV jets χ^2 worsening upon inclusion of the CMS or ATLAS 7 TeV jets data and vice versa (not shown). This provides a potential answer to long-standing questions about the different behaviour seen for this $t\bar{t}$ data by the different global fitting groups, with each group investigating this data originally in the context of different baseline fits, and so with different jet datasets and weights of data involved.

Reduced fit Luminosity Comparison

Overall it should be noted therefore that whilst some differences exist between the reduced fits, these have been reduced significantly in the process of the benchmarking. We have therefore seen that, provided the same data and theory settings are adopted (as far as possible), the

resulting PDFs agree well within errors within the data regions, the benchmarking can therefore be considered a success. This is emphasised by comparing the reduced fit luminosities. Here we present a comparison of the gluon-gluon, quark-antiquark, quark-quark and quark-gluon luminosities for the three reduced fits, including both the ratios of their central values and their uncertainties, shown in Figure 4. There is excellent agreement between the reduced fits at the level of the parton-parton luminosities, with the important gluon-gluon luminosities agreeing within uncertainties across the entire m_X range. Similar very good agreement is also seen in the quark-quark and quark-gluon luminosities. The agreement of the quark-antiquark luminosity is also good, however the NNPDF reduced fit luminosity is on the edge of the uncertainty bands for a small portion of the m_X range, this perhaps signals some slight remaining difference in the quark flavour decomposition. Nonetheless, the overall agreement is very good, indicating the reduced fits are well understood and a good baseline for benchmarking. Moreover, the uncertainties in the luminosities are also similar between all three groups, particularly in the most constrained regions where data is available.
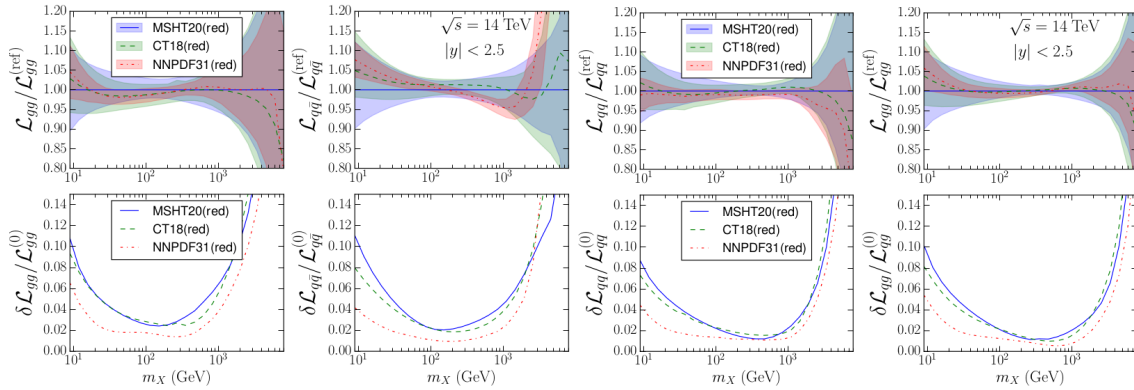
Figure 4: Comparison of the luminosities of the 3 reduced fits for gluon-gluon, quark-antiquark, quark-quark and quark-gluon (going from left to right). The absolute luminosities relative to that of the MSHT reduced fit are shown in the upper plots, whilst the uncertainties in the luminosities are shown in the lower plots. A rapidity cut of $|y| < 2.5$ is applied for the luminosities.

Conclusion

In conclusion, we have demonstrated the benchmarking of the PDFs produced by the three main global fitting groups: CT, MSHT and NNPDF. This is a vital exercise required to understand the differences arising in the global fits and to provide a combination PDF set for use at the LHC in Run III and beyond: PDF4LHC21. This article has served to provide an update on this effort. Benchmarking has been achieved through the use of "reduced fits" to a small subset of datasets treated in (nearly) identical ways by all three groups and with common theory settings as far as possible. These reduced fits were compared to the global fits and then used to identify and understand any differences. This comparison has been carried out both at the level of the PDFs themselves, through their central values and uncertainties, and through the comparison of the χ^2 fit qualities obtained for different datasets. Specific disagreements observed in the reduced fits were studied in more detail, in particular concerning differences in the strangeness of the reduced fit PDFs and in the high-x gluon, the latter being of particular interest to global fits. The reduced fit environment has enabled such differences to be largely resolved. In particular the question of the fit quality to different distributions of the

ATLAS 8 TeV $t\bar{t}$ lepton + jets data has been determined to be a reflection of tensions between datasets in the high-x gluon region and the different weights assigned to them in the different global fit environments. As a result of the benchmarking, the reduced fit PDFs are now in very good agreement, as emphasised in the comparison of the PDF luminosities, and so the benchmarking can be considered a success.

The work presented in this article will continue within the PDF4LHC combination group, with the final aim of producing a benchmarking paper and a new PDF4LHC combination set, both key components of future phenomenological programs at the LHC.

Acknowledgements

T.C. would like to acknowledge the work towards this article by the whole of the PDF4LHC21 combination group and wishes to thank all of those involved in this work and many useful discussions, including in the wider PDF4LHC working group on PDFs and related issues. In particular, special thanks are given to A. Cooper-Sarkar, L. Harland-Lang, T. Hobbs, T.-J. Hou, J. Huston, P. Nadolsky, E. Nocera, J. Rojo and R. Thorne for providing tables, plots, fits, comments and other contributions used in this work.

Funding information T. C. would like to thank the Science and Technology Facilities Council (STFC) for support via grant awards ST/P000274/1 and ST/T000856/1. Other contributors to this article would like to acknowledge support from STFC via grant award ST/L000377/1, from the Leverhulme Trust, the U.S. Department of Energy under Grant No. DE-SC0010129, the U.S. National Science Foundation under Grant No. PHY-2013791, the U.S. Department of Energy under Grant No. DE-AC02-07CH11359 and the U.S. National Science Foundation under Grant No. 2012165.

References

[1] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin and R. S. Thorne, *Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs*, Eur. Phys. J. C **81**(4), 341 (2021), doi:10.1140/epjc/s10052-021-09057-0, 2012.04684.

[2] T.-J. Hou *et al.*, *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC* (2019), 1912.10053.

[3] R. D. Ball *et al.*, *Parton distributions from high-precision collider data*, Eur. Phys. J. C **77**(10), 663 (2017), doi:10.1140/epjc/s10052-017-5199-5, 1706.00428.

[4] J. Butterworth *et al.*, *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43**, 023001 (2016), doi:10.1088/0954-3899/43/2/023001, 1510.03865.

[5] S. Alekhin *et al.*, *The PDF4LHC Working Group Interim Report* (2011), 1101.0536.

[6] R. D. Ball *et al.*, *Parton Distribution Benchmarking with LHC Data*, JHEP **04**, 125 (2013), doi:10.1007/JHEP04(2013)125, 1211.5142.

[7] T. Binoth *et al.*, *The SM and NLO Multileg Working Group: Summary report*, In *6th Les Houches Workshop on Physics at TeV Colliders*, p. 110 (2010), 1003.1241.

[8] J. Rojo *et al.*, *The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II*, J. Phys. G **42**, 103103 (2015), doi:10.1088/0954-3899/42/10/103103, 1507.00556.

[9] F. Faura, S. Iranipour, E. R. Nocera, J. Rojo and M. Ubiali, *The Strangest Proton?*, Eur. Phys. J. C **80**(12), 1168 (2020), doi:10.1140/epjc/s10052-020-08749-3, 2009.00014.

[10] A. C. Benvenuti *et al.*, *A High Statistics Measurement of the Proton Structure Functions F(2) (x, Q**2) and R from Deep Inelastic Muon Scattering at High Q**2*, Phys. Lett. **B223**, 485 (1989), doi:10.1016/0370-2693(89)91637-7.

[11] R. Aaij *et al.*, *Measurement of the cross-section for $Z \rightarrow e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV*, JHEP **02**, 106 (2013), doi:10.1007/JHEP02(2013)106, 1212.4620.

[12] M. Arneodo *et al.*, *Accurate measurement of F2(d) / F2(p) and R**d - R**p*, Nucl. Phys. **B487**, 3 (1997), doi:10.1016/S0550-3213(96)00673-6, hep-ex/9611022.

[13] M. Aaboud *et al.*, *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector*, Eur. Phys. J. C **77**(6), 367 (2017), doi:10.1140/epjc/s10052-017-4911-9, 1612.03016.

[14] M. Goncharov *et al.*, *Precise Measurement of Dimuon Production Cross-Sections in ν_μ Fe and $\bar{\nu}_\mu$ Fe Deep Inelastic Scattering at the Tevatron.*, Phys. Rev. **D64**, 112006 (2001), doi:10.1103/PhysRevD.64.112006, hep-ex/0102049.

[15] V. M. Abazov *et al.*, *Measurement of the shape of the boson rapidity distribution for $p\bar{p} \rightarrow Z/gamma^* \rightarrow e^+e^- + X$ events produced at $\sqrt{s}$ of 1.96-TeV*, Phys. Rev. **D76**, 012003 (2007), doi:10.1103/PhysRevD.76.012003, hep-ex/0702025.

[16] F. Aaron *et al.*, *Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA*, JHEP **01**, 109 (2010), doi:10.1007/JHEP01(2010)109, 0911.0884.

[17] H. Abramowicz *et al.*, *Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data*, Eur. Phys. J. C **75**(12), 580 (2015), doi:10.1140/epjc/s10052-015-3710-4, 1506.06042.

[18] S. Chatrchyan *et al.*, *Measurement of the Electron Charge Asymmetry in Inclusive W Production in pp Collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. Lett. **109**, 111806 (2012), doi:10.1103/PhysRevLett.109.111806, 1206.2598.

[19] R. S. Towell *et al.*, *Improved measurement of the anti-d / anti-u asymmetry in the nucleon sea*, Phys. Rev. **D64**, 052002 (2001), doi:10.1103/PhysRevD.64.052002, hep-ex/0103030.

[20] G. Aad *et al.*, *Measurement of the inclusive $W^\pm$ and Z/gamma cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Rev. D **85**, 072004 (2012), doi:10.1103/PhysRevD.85.072004, 1109.5141.

[21] R. Aaij *et al.*, *Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV*, JHEP **06**, 058 (2012), doi:10.1007/JHEP06(2012)058, 1204.1620.

[22] V. Khachatryan *et al.*, *Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV*, JHEP **03**, 156 (2017), doi:10.1007/JHEP03(2017)156, 1609.05331.

[23] M. Guzzi, P. M. Nadolsky, H.-L. Lai and C. P. Yuan, *General-Mass Treatment for Deep Inelastic Scattering at Two-Loop Accuracy*, Phys. Rev. D **86**, 053005 (2012), doi:10.1103/PhysRevD.86.053005, 1108.5112.

[24] J. R. Andersen *et al.*, *Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group Report* (2014), 1405.1067.

[25] M. Aaboud *et al.*, *Measurement of the Drell-Yan triple-differential cross section in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **12**, 059 (2017), doi:10.1007/JHEP12(2017)059, 1710.05167.

[26] G. Aad *et al.*, *Measurement of the cross-section and charge asymmetry of W bosons produced in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **79**(9), 760 (2019), doi:10.1140/epjc/s10052-019-7199-0, 1904.05631.

[27] M. Aaboud *et al.*, *Precision measurement and interpretation of inclusive W^+, W^- and Z/γ^* production cross sections with the ATLAS detector*, Eur. Phys. J. C **77**(6), 367 (2017), doi:10.1140/epjc/s10052-017-4911-9, 1612.03016.

[28] R. S. Thorne, L. A. Harland-Lang and A. D. Martin, *Inclusion of new LHC data in MMHT PDFs*, PoS **DIS2017**, 202 (2018), doi:10.22323/1.297.0202, 1708.00047.

[29] D. A. Mason, *Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon data*, Ph.D. thesis, Oregon U., doi:10.2172/879078 (2006).

[30] P. Zyla *et al.*, *Review of Particle Physics*, Prog. Theor. Exp. Phys. **8**(8), 083C01 (2020), doi:10.1093/ptep/ptaa104.

[31] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. C **75**(5), 204 (2015), doi:10.1140/epjc/s10052-015-3397-6, 1412.3989.

[32] T. Bolton, *Determining the CKM parameter V(cd) from neutrino N charm production* (1997), hep-ex/9708014.

[33] D. Mason *et al.*, *Measurement of the Nucleon Strange-Antistrange Asymmetry at Next-to-Leading Order in QCD from NuTeV Dimuon Data*, Phys. Rev. Lett. **99**, 192001 (2007), doi:10.1103/PhysRevLett.99.192001.

[34] G. Aad *et al.*, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76**(10), 538 (2016), doi:10.1140/epjc/s10052-016-4366-4, 1511.04716.

[35] S. Bailey and L. Harland-Lang, *Differential Top Quark Pair Production at the LHC: Challenges for PDF Fits*, Eur. Phys. J. **C80**(1), 60 (2020), doi:10.1140/epjc/s10052-020-7633-3, 1909.10541.

[36] O. Amat *et al.*, *Impact of LHC top-quark pair measurements to CTEQ-TEA PDF analysis* (2019), 1908.06441.

[37] T.-J. Hou, Z. Yu, S. Dulat, C. Schmidt and C. P. Yuan, *Updating and optimizing error parton distribution function sets in the Hessian approach. II.*, Phys. Rev. D **100**(11), 114024 (2019), doi:10.1103/PhysRevD.100.114024, 1907.12177.

[38] M. Kadir, A. Ablat, S. Dulat, T.-J. Hou and I. Sitiwaldi, *The impact of ATLAS and CMS single differential top-quark pair measurements at $\sqrt{s} = 8$ TeV on CTEQ-TEA PDFs*, Chin. Phys. C **45**(2), 023111 (2021), doi:10.1088/1674-1137/abce10, 2003.13740.

[39] M. Czakon, N. P. Hartland, A. Mitov, E. R. Nocera and J. Rojo, *Pinning down the large-x gluon with NNLO top-quark pair differential distributions*, JHEP **04**, 044 (2017), doi:10.1007/JHEP04(2017)044, 1611.08609.

[40] L. A. Harland-Lang, A. D. Martin and R. S. Thorne, *The Impact of LHC Jet Data on the MMHT PDF Fit at NNLO*, Eur. Phys. J. C **78**(3), 248 (2018), doi:10.1140/epjc/s10052-018-5710-7, 1711.05757.

[41] R. S. Thorne, S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin and R. Nathvani, *Updates of PDFs using the MMHT framework*, PoS **DIS2019**, 036 (2019), doi:10.22323/1.352.0036, 1907.08147.

[42] R. Abdul Khalek *et al.*, *Phenomenology of NNLO jet production at the LHC and its impact on parton distributions*, Eur. Phys. J. C **80**(8), 797 (2020), doi:10.1140/epjc/s10052-020-8328-5, 2005.11327.

[43] *Determination of the parton distribution functions of the proton from ATLAS measurements of differential W and Z/γ^* and $t\bar{t}$ cross sections*, Tech. Rep. ATL-PHYS-PUB-2018-017, CERN, Geneva (2018).

[44] S. Amoroso *et al.*, *Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report*, In *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches* (2020), 2003.01700.