THE EUROPEAN
PHYSICAL JOURNAL C

CrossMark

The impact of LHC jet data on the MMHT PDF fit at NNLO

L. A. Harland-Lang[1,2,a], **A. D. Martin**[3], **R. S. Thorne**[2]

[1] Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK
[2] Department of Physics and Astronomy, University College London, London WC1E 6BT, UK
[3] Institute for Particle Physics Phenomenology, University of Durham, Durham DH1 3LE, UK

Abstract We investigate the impact of the high precision ATLAS and CMS 7 TeV measurements of inclusive jet production on the MMHT global PDF analysis at next-to-next-to-leading order (NNLO). This is made possible by the recent completion of the long-term project to calculate the NNLO corrections to the hard cross section. We find that a good description of the ATLAS data is not possible with the default treatment of experimental systematic errors, and propose a simplified solution that retains the dominant physical information of the data. We then investigate the fit quality and the impact on the gluon PDF central value and uncertainty when the ATLAS and CMS data are included in a MMHT fit. We consider both common choices for the factorization and renormalization scale, namely the inclusive jet transverse momentum, $p_\perp$, and the leading jet $p_\perp$, as well as the different jet radii for which the ATLAS and CMS data are made available. We find that the impact of these data on the gluon is relatively insensitive to these inputs, in particular the scale choice, while the inclusion of NNLO corrections tends to improve the data description somewhat and has a qualitatively similar though not identical impact on the gluon in comparison to NLO.

1 Introduction

Parton distribution functions (PDFs) are a fundamental input into hadron collider physics for both the theoretical and experimental particle physics communities; see [1] for a recent review. The dominant experimental input in determining PDFs comes from data on deep inelastic scattering (DIS) structure functions, with the final combined HERA Run I + II data set being the most prominent example [2]. By combining data from different processes (neutral and charged current) and targets (protons, deuterons, heavy nuclei) one can obtain much direct information about the quark content

of the proton, while the quark evolution is sensitive to the gluon. Indeed, at small x this evolution is largely driven by the gluon. However, at moderate and high x the proton structure is dominated by the non-singlet valence quark distributions. Then, the evolution is largely decoupled from the gluon and the speed of evolution is determined mostly by the value of the strong coupling constant. There is, of course, some influence from the gluon in the high-x quark evolution; it provides an increase of quarks (and anti-quarks) with Q^2, but this is difficult to decorrelate from the variation with α_S, and also decreases in significance at high x.

In order to obtain the most comprehensive constraints various groups perform global fits to all available data for which precise theoretical calculations are available [3–5]. For example, Drell Yan data in hadron–hadron collisions (both collider and fixed target) provide additional information on the anti-quarks and on the quark flavour decomposition, and improve the overall determination in comparison to DIS data alone [6,7]. A direct constraint on the gluon, particularly at high x, can be obtained from high $p_\perp$ jet production data at hadron colliders.[1] However, until recently the data have been quite limited in precision, while the calculation of the hard cross section has only been available up to next-to-leading order (NLO), with some threshold resummation results also available [10–13]. In our most recent global fit [3] we included data on inclusive jet production as a function of p_T in different rapidity bins from the D0 [14] and CDF [15] experiments at the Tevatron, and from early measurements at both the ATLAS [16,17] and CMS [18] detectors, at 7 TeV. The Tevatron data were generally close to threshold, so that we could reliably include the NNLO approximations obtained from expanding out the threshold resummation of [10]. However, the LHC data extended much

ᵃ e-mail: l.harland-lang@ucl.ac.uk

---

[1] Constraints from differential top quark pair production [8] and the Z boson $p_\perp$ distribution [9] have recently become feasible, in particular due to the relatively recent release of high precision LHC data and NNLO theory calculations for these processes.

further from threshold, where it was clear these approximations break down [12]. Thus in that study we only included the Tevatron jet data in the NNLO fit. We note that at this time other groups used alternative approaches of including jet data at NNLO; see [4,19].

Since this previous study there has been both an increase in the range and precision of LHC jet data, combined with the completion of a very large-scale and long-term project to calculate the NNLO corrections to the hard cross section [20–22]. In this paper we investigate the consequences of both of these new developments for PDF determination, concentrating on the final 7 TeV measurements from the ATLAS [23] and CMS [24] collaborations. We find that neither is quite as straightforward as might be hoped. First, the newer ATLAS jet data are impossible to fit well without some modifications. Second, the variation in NNLO corrections between scale choices and jet radii is potentially quite significant. We therefore examine both of these issues in detail, considering the impact of different choices of scale and jet radius on the fit at both NLO and NNLO, while suggesting a minimal manner in which to improve the fit quality in the case of the ATLAS data that retains the dominant physical constraints implied by the data. We then determine the consequences for both the central values and uncertainties of the gluon PDF obtained at both NLO and NNLO within the MMHT framework. We obtain the very encouraging, and not necessarily expected, result that in practice these are found to very insensitive to any reasonable choices we make in either the treatment of the data or the theory input.[2] We also find in general that the data description is somewhat improved by the inclusion of these NNLO corrections. As mentioned above, we only consider the 7 TeV data, for which the NNLO calculations are currently available, in this study. In fact, already a range of precise jet data from ATLAS and CMS at 8 and 13 TeV [26–29] are available. This study will therefore guide the inclusion of these data in future MMHT fits at NNLO. For example, a similarly poor default description is also present in the ATLAS 8 TeV [26] and 13 TeV [27] data, and so it must be dealt with in any future fit.

The outline of this paper is as follows. In Sect. 2 we describe the theoretical calculation and tools used. In Sect. 3 we describe the issues related to the fit to the ATLAS data and develop a simplified approach to improve the data description. In Sect. 4 we study the fit quality at NLO and NNLO to the ATLAS and CMS 7 TeV jet data, for different choices of jet scale and radius. In Sect. 5 we show the impact of these data on the central value and uncertainty of the gluon PDF. Finally, in Sect. 6 we conclude.

2 Theoretical inputs for jet production

The NLO theoretical predictions for inclusive jet production are calculated using the `NLOJet++` code [30], with the results stored in `APPLgrid` [31] format[3] for fast use in the PDF fit. The theoretical approach used to calculate the NNLO corrections to this is described in [22] (see also [33]). The NNLO to NLO K-factors are provided by the authors for the ATLAS and CMS 7 TeV kinematics, for both jet radii presented by the collaborations and with the renormalization/factorization scale taken as either inclusive jet transverse momentum, $p_\perp^{\rm jet}$, or the maximum jet transverse momentum, $p_\perp^{\rm max}$. These are provided using the NNPDF3.0 set [5], although the K-factors are expected to be largely insensitive to the PDF choice. This therefore allows us to perform a detailed analysis of the impact of these LHC jet data for different jet radii and scale choices.

The predictions are provided with a statistical uncertainty due to the MC integration in the theoretical calculation. We show the predictions for the central ATLAS and CMS rapidity bins for illustration in Fig. 1, taking representative choices of the jet radius and scale. We can see that the MC uncertainties are up to 1% of the K-factors, while the central values are at most 10% from unity. These errors are therefore non-negligible and must be included in the fit. One possibility is to simply include these as an additional bin-by-bin source of uncorrelated uncertainty. However, in general we will expect the K-factors to be smoothly varying functions of the kinematic variables, and therefore this approach is unnecessarily conservative. We instead perform a simple four-parameter fit:

$$\frac{\sigma_{\rm NNLO}^i}{\sigma_{\rm NLO}^i} = \lambda_0^i + \lambda_1^i \log p_\perp + \lambda_2^i \log^2 p_\perp + \lambda_3^i \log^3 p_\perp , \quad (1)$$

where the 'i' labels the specific rapidity region, data set, choice of jet radius and scale. In general this enables a good fit, with $\chi^2/{\rm dof} \sim 1$, and the standard $\Delta\chi^2 = 1$ criterion can be applied to determine the uncertainties associated with the fit. This results in four sources of correlated systematic uncertainty for each rapidity bin, which we then include in the PDF fit. We have explicitly checked that the results are stable with respect to adding in further polynomial terms, and thus we can safely truncate at this order. The best fit curves are shown in Fig. 1 for four representative cases, with the uncertainty band due to the sum in quadrature of the errors in each bin, i.e. omitting correlations, shown for illustration. We can see that biggest difference in the K-factors is due to the choice of scale, which gives quite different trends depending on whether $p_\perp^{\rm max}$ or $p_\perp^{\rm jet}$ is taken. The choice of jet radius also

---

[2] As such, a detailed investigation of the scale dependence as described in [25] may be avoided.

[3] In the case of the ATLAS data with the $p_\perp^{\rm max}$ scale we take this directly from the `APPLgrid` website [32].
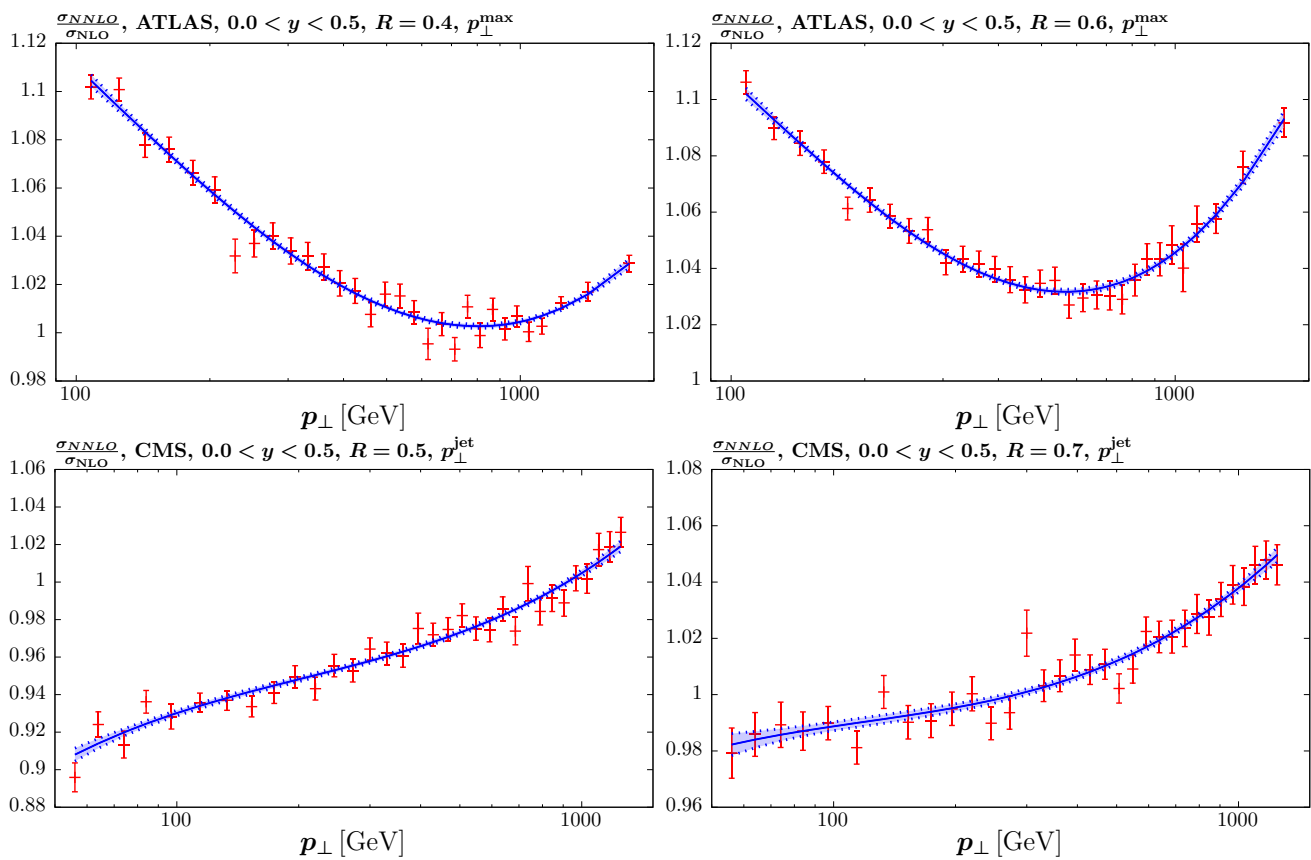
Fig. 1 The calculated NNLO to NLO K-factors [22], including the MC statistical errors, for representative ATLAS and CMS 7 TeV jet kinematics. Also shown is the four-parameter fit to the K-factors described in the text. The uncertainty band is shown for illustration by summing in quadrature the 68% C.L. fit uncertainties in each bin, i.e. omitting correlations. The top left (right) plots show the ATLAS result with the $p_\perp^{\max}$ scale choice and $R = 0.4$ (0.6), while the bottom left (right) plots show the CMS result with the $p_\perp^{\mathrm{jet}}$ scale choice and $R = 0.5$ (0.7). In both cases results for the central rapidity bin are shown

shows some impact, while the ATLAS and CMS results with the same scale choice and comparable (either low or high) jet radii, which are not shown here, show a qualitatively similar trend.

3 Treatment of ATLAS correlated systematic errors

Before considering the general impact of the jet data on the NNLO fit, some care is needed when dealing with the ATLAS data [23]. The discussion follows closely that given in [34], but for completeness we present a summary below. For illustration, we will work at NLO only, using as a baseline PDF the MMHT14 set [3] including the HERA I + II combined data [2], that is, as presented in [35].

The predicted and fit data/theory for the first four jet rapidity bins are shown in Fig. 2, with the shifts due to the correlated systematic uncertainties included and using $R = 0.4$ and $p_\perp^{\mathrm{jet}}$. The description of the data is visibly poor, in particular in the $0.5 < |y_j| < 1.0$ and $1.0 < |y_j| < 1.5$ bins across all $p_\perp$, and with some deterioration in the other bins at high $p_\perp$. A significant contributing factor to this is an essentially systematic offset in the data/theory between these neighbouring rapidity bins, but in opposite directions; as these probe PDF sets of the same flavour in very similar x and Q^2 regions little improvement is possible (or observed) by refitting to these data.

The cause of this appears to lie with the shift allowed by the correlated systematic uncertainties. The ATLAS data contain a large number of individual correlated errors which are generally completely dominant over the (small) statistical errors; for the 'weaker' assumption about error correlations defined in [23] that we take, there are 71 individual sources of systematic error. If we simply assume that all of these uncertainties are completely decorrelated between the six rapidity bins (while remaining fully correlated within the bins) a universally good description is found: in this case, the extra freedom allows the data to shift in order to achieve a
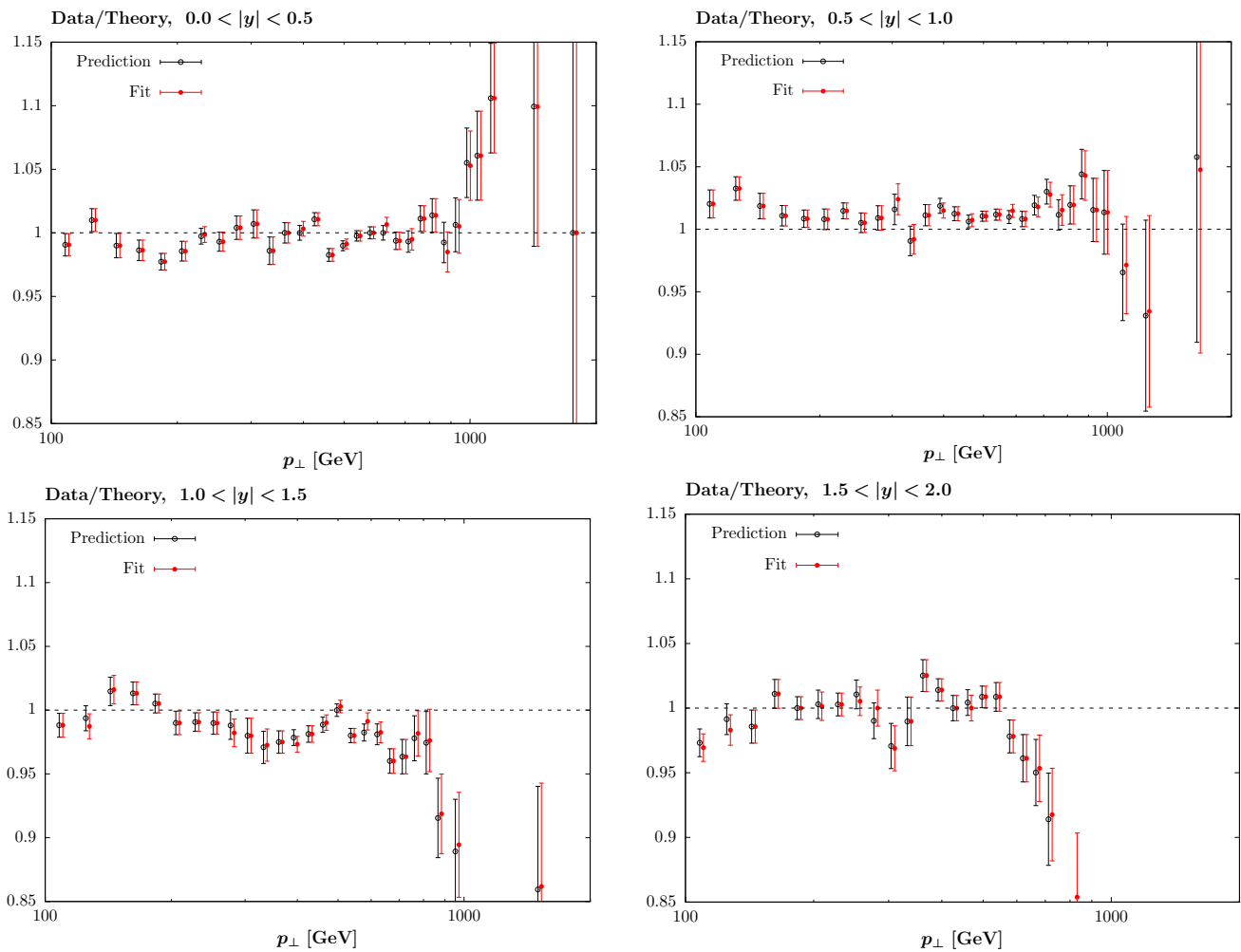
Fig. 2 Comparison of NLO prediction and fit to ATLAS jet data [23] for first four rapidity bins. Data/theory is plotted, with the data already shifted by the systematic uncertainties in order to achieve the best description. The displayed errors are purely statistical

reasonable data/theory description. This is, however, clearly a hugely over-conservative assumption. To be more precise, we examine the size of the shifts r_k for each source of systematic uncertainty by which the theory (or equivalently, data) points are allowed to move, as defined in the χ^2:

$$\chi^2 = \sum_{i=1}^{N_{pts}} \left(\frac{D_i + \sum_{k=1}^{N_{corr}} r_k \sigma_{k,i}^{corr} - T_i}{\sigma_i^{uncorr}} \right) + \sum_{k=1}^{N_{corr}} r_k^2 \,, \qquad (2)$$

where D_i is ith data point, T_i is the theory prediction and σ_i^{uncorr} ($\sigma_{k,i}^{corr}$) are the uncorrelated (correlated) errors. We in particular evaluate the shifts for each of the first four rapidity bins (from 0 to 2.0 in steps of 0.5) individually; including the last two rapidity bins, where the data tend to be less precise, does not affect the conclusions that follow. Any tensions between the different bins may then show up through significantly different r_k values being preferred in the different rapidity bins, in order to achieve good individual fits. In

Fig. 3 we show the average squared sum of the shift differences $(r_i - r_j)^2$ for the four bins. It is clear that for a small subset of the shifts the size of this difference is significantly larger than zero, indicating a large degree of tension.

The three shifts jes21, 45 and 62 as defined in [36], which correspond [37] to the multi-jet balance asymmetry, an in-situ statistical uncertainty and the jet energy scale close by jets, respectively, show particularly large differences. However, for the in-situ statistical uncertainty the correlations are particularly well determined [37], and therefore we omit this from the investigation below. In fact, as shown in [34], decorrelating this source of uncertainty has a less significant impact on the quality of the data description in comparison to the other two sources.

We therefore investigate the impact of decorrelating the systematic uncertainties, jes21 and 62, alone between rapidity bins. We compare to the ATLAS data with $R = 0.4$ and $p_\perp^{jet}$ as the scale choice. The result for the individual
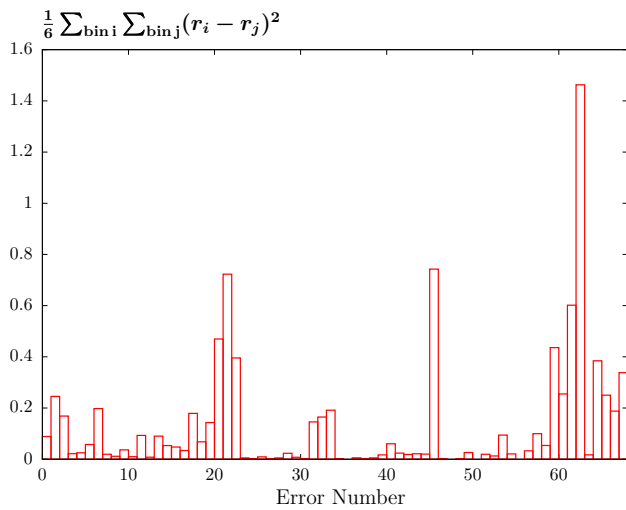
Fig. 3 Average squared sum of the systematic shift differences $(r_i - r_j)^2$ for the first four rapidity bins of the ATLAS 7 TeV jet data [23]

Table 1 χ^2 per number of data points ($N_{pts} = 140$) for fit to ATLAS jets data [23], with the default systematic error treatment ('full') and with certain errors, defined in the text, decorrelated between jet rapidity bins

	Full	21	62	21,62
$\chi^2/N_{pts.}$	2.85	1.58	2.36	1.27

uncertainty sources, as well as the combination, is shown in Table 1, and is found to be dramatic. Simply decorrelating jes21, for example, leads to a reduction of 180 points in χ^2, giving almost a factor of 2 decrease in the $\chi^2/N_{pts.}$ from 2.85 to 1.58. Decorrelating jes62 in addition gives a $\chi^2/N_{pts.}$ of 1.27. The same data/theory comparisons as in Fig. 2, but including this decorrelation of jes21 and jes62, are shown in Fig. 4 and are visibly improved, with the additional freedom allowing the data/theory to shift in the different rapidity bins and achieve a good overall description. While the above analysis only considers the experimental sources of correlated uncertainty, we have also checked that decorrelating the quoted uncertainty associated with the non-perturbative corrections from [23] that we apply leads to some $\sim$ 40 point improvement in the χ^2, which is significantly smaller than for those sources discussed above. In addition, we find that even omitting these corrections entirely has little impact on the fit quality, in other words these appear to be correlated sufficiently with other sources of experimental systematics that their omission does not affect the comparison significantly

We note that this corresponds to a simplified version of the alternative correlation scenarios presented in [26] subsequently to the discussion in [34]. Here, the impact of a more conservative partial decorrelation of various sources of uncertainty (including theoretical uncertainties due to scale choice and variation) in the 8 TeV ATLAS jet data is investigated, and a comparable although somewhat less dramatic improvement in the data description quality is found. However, as we will show below, the effect of our simplified decorrelation model is to improve the fit quality while having a limited effect on the PDFs themselves. Therefore we do not expect the details of the decorrelation model to have a

significant impact on the final result. Thus, while the correlation between systematic errors should clearly be determined by physics considerations and not simply the possibility of improving the theory description of the data, the simplified approach we take is sufficient for our purposes.

4 Fit quality at NNLO

4.1 Individual data sets at NNLO

In Table 2 we show the quality, χ^2, of the prediction and fit to the ATLAS and CMS jet data. For the predictions, we take as a baseline set the fits to the same data set (and using the same theoretical parameters) as MMHT14 [3], but including the final HERA I+II combined data set [2], and excluding all Tevatron jet data. In the latter case the NNLO predictions are not currently publicly available and so these are omitted for consistency. Unless otherwise stated we take $p_\perp^{jet}$ as the factorization/renormalization scale. The NLO (NNLO) results are all made with a fixed value of α_s of 0.120 (0.118), as taken in [3], although the results are insensitive to this precise choice. We first consider the impact of fitting the ATLAS and CMS jet data individually. We show the 'ATLAS' result with the default treatment of systematic errors, with our model of partial error decorrelation (σ_{pd}), and with a full decorrelation of all systematic errors across jet rapidity bins (σ_{fd}). While as discussed above the latter approach is clearly overly conservative, we note that e.g. only fitting the first jet rapidity bin as in [5] implicitly assumes such a decorrelation.

As in the NLO case above, the description and fit of the ATLAS data with the default error treatment is poor, with $\chi^2/N_{pts} \sim 2$ or higher, but this improves to be of order unity when taking our model of partial error decorrelation. If the systematic errors are fully decorrelated between rapidity bins, some further improvement is achieved, giving a value that is somewhat below unity. However, it is clear that the most dramatic change comes from the decorrelation of the first two systematic errors. We also show the comparison for different choices of jet radius, with $R = 0.4$ (0.5) and $R = 0.6$ (0.7) for the ATLAS (CMS) data, which in the following we will label as 'low' and 'high', respectively. Interestingly, with the higher choice of R the quality of the description of the ATLAS data is better, while the change
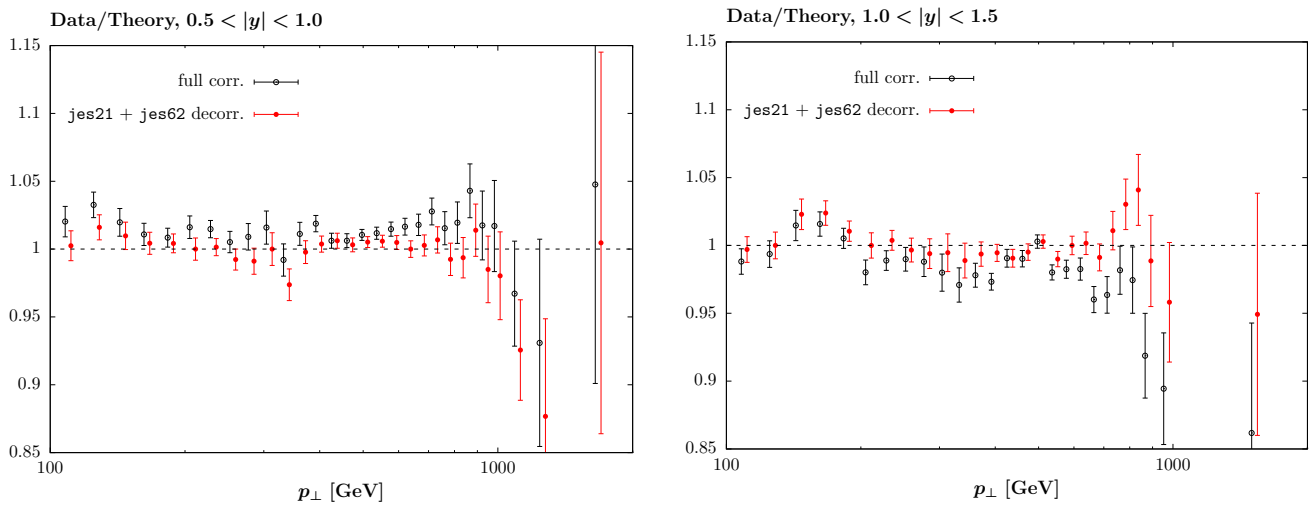
Fig. 4 Data/theory fit as in Fig. 2, for $0.5 < |y_j| < 1.0$ and $1.0 < |y_j| < 1.5$, with and without the labelled systematic errors decorrelated between jet rapidity bins

Table 2 The χ^2 for the ATLAS ($N_{pts} = 140$) and CMS 7 TeV jet data ($N_{pts} = 158$) at NNLO. The quality of the description using the baseline set is shown, while the result of re-fitting to the single jet data set is given in brackets. Results with the different treatments of the ATLAS systematic uncertainties, described in the text, are also shown

	ATLAS	ATLAS, σ_{pd}	ATLAS, σ_{fd}		CMS
$R = 0.4$	350.8 (333.7)	183.1 (170.7)	128.4 (122.2)	$R = 0.5$	191.7 (163.4)
$R = 0.6$	304.0 (264.0)	178.8 (148.9)	128.9 (115.7)	$R = 0.7$	200.1 (175.2)

Table 3 The χ^2 for the combined fit to the ATLAS ($N_{pts} = 140$) and CMS ($N_{pts} = 158$) 7 TeV jet data. The values for the ATLAS and CMS contributions are given, for different choices of jet radius and scale, at NLO and NNLO

	$R_{low}, p_\perp^{jet}$	$R_{low}, p_\perp^{max}$	$R_{high}, p_\perp^{jet}$	$R_{high}, p_\perp^{max}$
ATLAS (NLO)	213.8	190.5	171.5	161.2
ATLAS (NNLO)	172.3	199.3	149.8	152.5
CMS (NLO)	190.3	185.3	195.6	193.3
CMS (NNLO)	177.8	187.0	182.3	185.4

when refitting is significantly increased; for the partial error decorrelation the χ^2 decreases by ~ 30 points, giving a final χ^2/N_{pts} very close to unity. On the other hand, for the full error decorrelation, little difference is seen, which is perhaps unsurprising given the over-estimate in the freedom of the data uncertainties. We also show the χ^2 for the prediction and fit to the CMS jet data. Here the description is fair, and a $\chi^2/N_{pts} \sim 1$ is achieved for both radii after refitting, with a reduction in the χ^2 by ~ 30 points. The fit quality is a little better for the lower choice of jet radius, although the difference is relatively small.

4.2 NNLO vs. NLO fit quality for combined data

In Table 3 we show the results of the combined fit to the ATLAS and CMS data, taking the partial error decorrelation model for the ATLAS data. As above, we show results for low and high jet radii, while we also consider the impact of

the jet scale choice. The χ^2 values for the ATLAS and CMS data sets are given.

A number of observations can be made. First, the quality of the description when fitting both data sets simultaneously is nearly as good as when fitting each individually. This is particularly true for ATLAS, while for CMS there is a little more deterioration. Thus, there is no significant tension between the ATLAS and CMS jet data. The deterioration in the fit quality for the other data sets (not shown here), is relatively mild, with $\Delta\chi^2 \sim 5$ (10) for the low (high) jet radius choice. The biggest deterioration is in the BCDMS deuteron data for the high jet radius choice, which deteriorates by ~ 7 units.

Second, at NNLO some improvement in the fit quality is apparent for most choices of jet radii and scale; such an improvement is also visible upon comparing Tables 1 and 2 for the individual fits. This is more significant for the ATLAS data, where an improvement of up to 40 points in χ^2 can be

Table 4 The χ^2 for the combined NNLO fit to the ATLAS and CMS 7 TeV jet data, excluding and including the calculated NNLO K-factors, and excluding the errors associated with the polynomial fit to the K-factors. The $p_\perp^{\mathrm{jet}}$ factorization/renormalization scale is taken

	NLO theory	NNLO	NNLO (no errors)
ATLAS, R_{low}	215.3	172.3	179.1
ATLAS, R_{high}	159.2	149.8	153.5
CMS, R_{low}	194.2	177.8	182.8
CMS, R_{high}	198.5	182.3	188.8

achieved, while for the CMS data the improvement is at most 10 points. On the other hand, for the low jet radius, and $p_\perp^{\mathrm{max}}$ choice, some slight deterioration in the fit quality is observed. We note that in [34] a deterioration in the fit quality to the ATLAS data when going to NNLO was reported; however, here it was precisely these choices of scale and jet radius that were taken. Following the more detailed study in this work, we can see that this effect is not in general present.

Third, we can see that for the joint fit a clear preference for the higher choice of jet radius is shown at both orders in the ATLAS data, while for the CMS any difference is relatively marginal. Moreover, while at NLO some preference (in particular in the ATLAS data) for the $p_\perp^{\mathrm{max}}$ scale choice is shown, at NNLO this trend is reversed for the low R choice, while for high R essentially no preference is indicated by the fit, with the descriptions of the ATLAS and CMS data being excellent for both scale choices. Thus to achieve the best NNLO fits to these data sets, a higher value of R is preferred, while the result is less sensitive to the choice of scale. As we will show in the following section, this relative insensitivity is also observed in the extracted PDFs, in particular for the gluon.

Finally, it is important to clarify the role played by the NNLO jet production theory, in contrast to the NNLO PDFs,

in leading to the improvement in the fit quality at NNLO. In Table 4 we show the same χ^2 values as before, resulting from the NNLO fit to the combined ATLAS and CMS data, but in addition excluding the NNLO K-factors, i.e. applying NLO theory only to the jet data. We can see that the improvement due to the NNLO corrections in the fit is still present at roughly the same level as before, with some variation in the precise amount. We also show the effect of excluding the correlated errors associated with the K-factor fit described in Sect. 2. This leads to some small increase in the χ^2, as it must, but the trend is unchanged.

5 Impact of LHC jet data on PDFs

5.1 Central values

In this section we investigate the impact of including the jet data on the PDFs. We concentrate on the gluon PDF, as the effect on all quark PDF combinations is significantly smaller. In Fig. 5 we show the impact of including the ATLAS jet data only in the fit, in comparison to the MMHT baseline described in the previous section (i.e. with Tevatron jet data omitted). We show the result with $R = 0.4$ (0.6) in the left (right) figure, with the different treatments of the systematic errors described above. Only the comparison at NNLO is shown here, leaving the comparison to NLO for the combined fit to be presented below. Unless otherwise stated, in what follows we take $p_\perp^{\mathrm{jet}}$ as the choice of scale.

We can see that for both jet radii, despite leading to significantly different fit qualities, the partial decorrelation and default error treatments in fact result in quite similar fits for the gluon PDF, with some softening observed at high x. On the other hand, the full decorrelation of systematic uncertainties leads to a gluon that is qualitatively different, being much less soft at high x, although still consistent within PDF
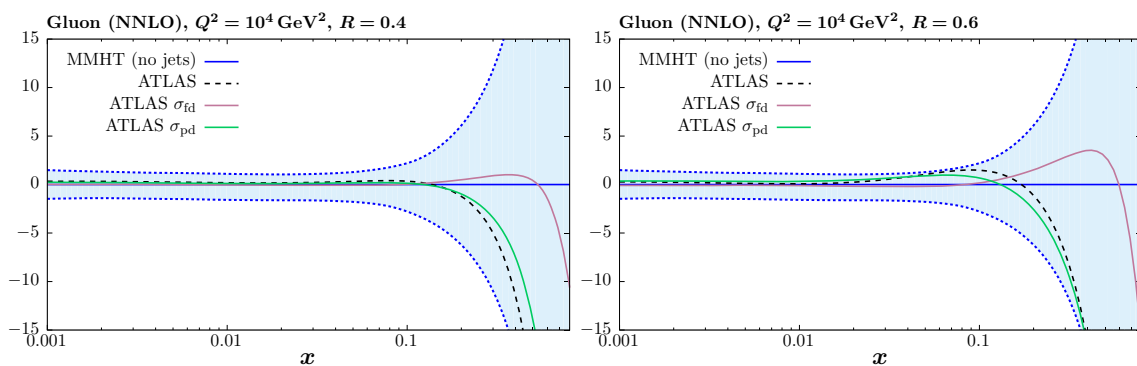

Fig. 5 The impact on the gluon PDF at NNLO of the ATLAS 7 TeV jet data [23], including two alternative treatments of the correlated systematic errors described in the text. The percentage difference in comparison to the baseline fit, with no jet data included, is shown. The jet radii $R = 0.4$ (0.6) ATLAS data is used in the left (right) plot
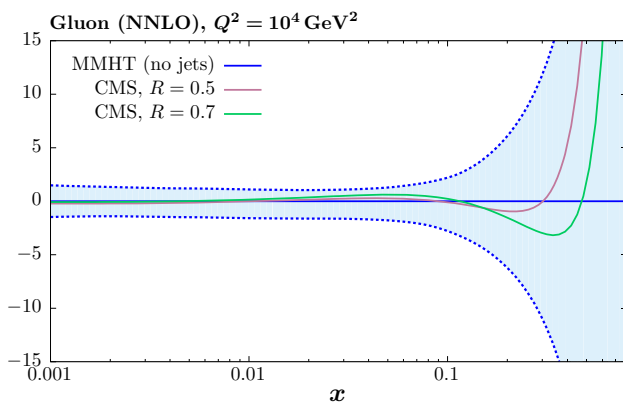
Fig. 6 The impact on the gluon PDF at NNLO of the CMS 7 TeV jet data [24], for two value of jet radii. The percentage difference in comparison to the baseline fit, with no jet data included, is shown

uncertainties. This is perhaps not surprising, as the systematic shifts we determine by profiling with respect to the various correlated uncertainties in (2) have a physical interpretation, giving us the best fit values of the various experimental parameters and a corresponding best fit measurement that is shifted with respect to the default. By treating these sources of uncertainty as uncorrelated across rapidity bins, this connection is largely lost, and in effect an imperfect measurement that is systematically different may be fit. The central value of the extracted gluon may then vary quite significantly. This effect is indeed observed in Fig. 5. Given these results, in what follows we will simply apply our model of partial error decorrelation, although we note that in all cases the results are very similar when taking the default treatment.

It is interesting to observe in Fig. 5 that the difference due to the choice of jet radius is relatively small, and much less than that due to the error treatment, although the higher $R = 0.6$ choice leads to a somewhat softer gluon at high x. In Fig. 6 we show the result of the NNLO fit, including the CMS

jet data only, for both jet radii. Here, the impact on the gluon is relatively flat out to quite high x, where some hardening is observed, albeit within the large PDF uncertainties in this region. As with the ATLAS data, the larger choice of jet radius leads to some softening in the gluon in comparison to the lower choice.

In Fig. 7 we now consider the effect of combined fit to the ATLAS and CMS jet on the gluon. As mentioned above, we take the partial decorrelated treatment of the ATLAS jet data in what follows. We show results for low and high jet radii, i.e. with $R = 0.4$ (0.5) and $R = 0.6$ (0.7) for the ATLAS (CMS) data, respectively. We also show the effect of taking the $p_\perp^{max}$ scale choice in comparison to $p_\perp^{jet}$. The result at NLO (NNLO) is shown in the left (right) panel. The impact of the scale choice on the gluon is quite small, of the same order of or less than that due to the choice of jet radius, although here the difference for the combined fit is also not dramatic. This is not necessarily to be expected, as the difference between the scale choices in the underlying theory prediction is not negligible. In addition, while the qualitative trend in the NLO and NNLO fits is similar, the latter leads to a somewhat softer gluon, which even lies somewhat outside the baseline PDF uncertainty band for the higher jet radius. We can also see that the gluon that results from the combined fit lies closer to the result from the ATLAS then the CMS fit, although these are all consistent within PDF uncertainties. This is consistent with the somewhat larger deterioration observed in the fit quality for the CMS-only case in comparison to the combined fit.

Thus, to summarise the effect of the LHC data and the accompanying theory improvements, in all cases we observe some relative stability in the overall trend of the extracted gluon, with the smallest differences being due to the choice of scale, followed by the choice of jet radius and finally the NLO vs. NNLO difference being largest. We study this result in more detail below.
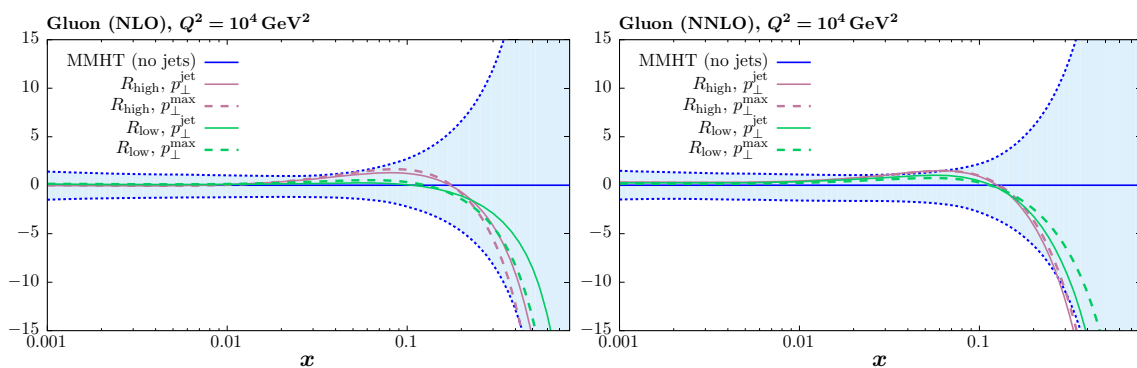

Fig. 7 The impact of the ATLAS [23] and CMS [24] 7 TeV jet data on the gluon PDF at NLO (left) and NNLO (right). The percentage difference in comparison to the baseline fit, with no jet data included, is

shown. Results are given for 'low' and 'high' jet radii described in the text, and for two choices of the factorization scale
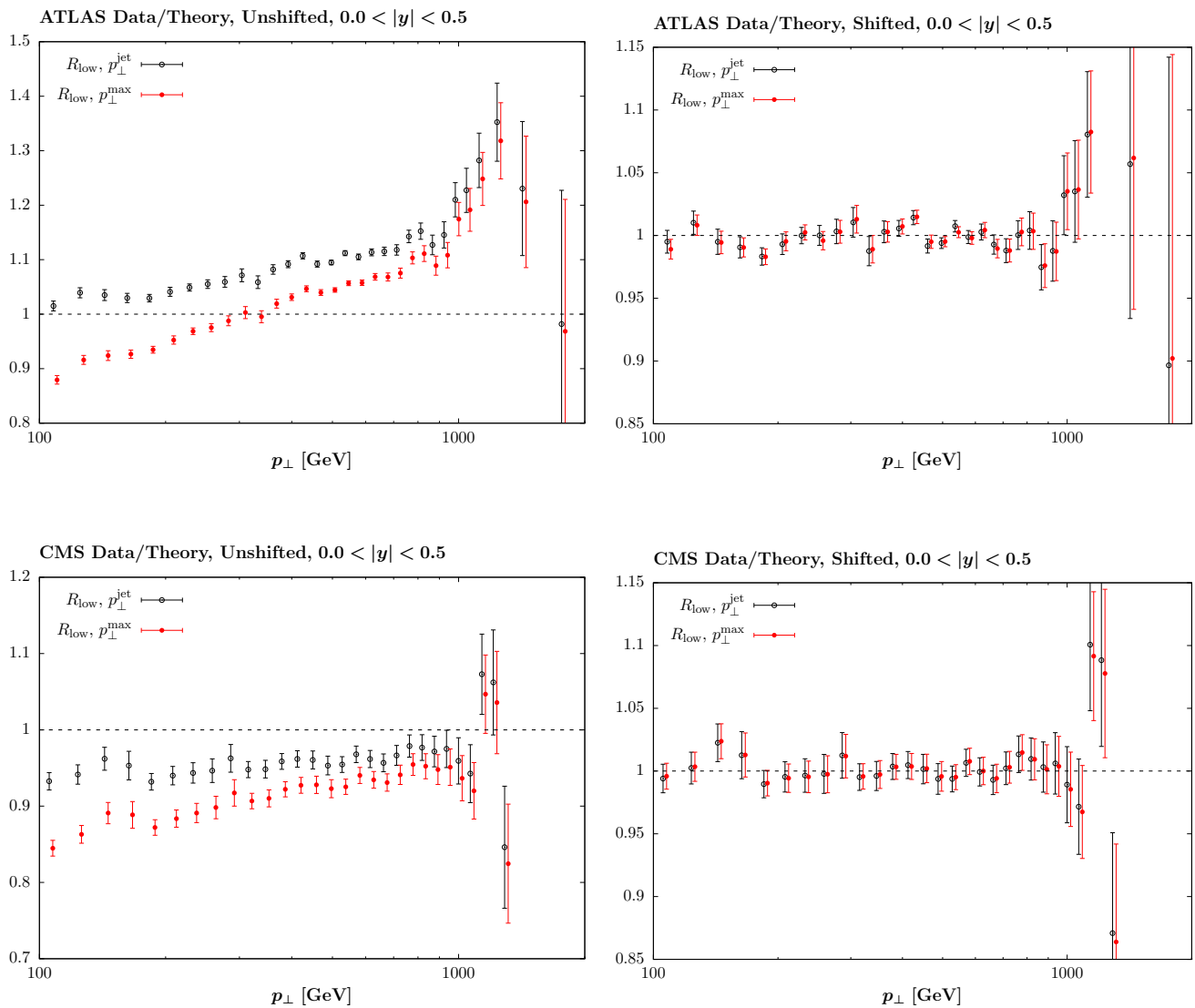
Fig. 8 Data/theory for the NNLO fit including the combined ATLAS and CMS data with the lower R radius choice and for both choices of scale. The comparison in the central rapidity bin for the ATLAS (CMS) data is shown in the upper (lower) plots. The result before and after the inclusion of the shifts due to the correlated systematic errors is shown in the left and right hand plots, respectively. Statistical errors only are shown

To investigate the effect of scale choice further, in Figs. 8 and 9 we show the data/theory for both choices of scale, with lower and higher R choice, respectively. We show results for both the ATLAS and the CMS data, in the central rapidity bin (although similar results are seen at other rapidities). In the left hand plots we show the results prior to including the systematic shift in the correlated errors. We concentrate our discussion below on the lower R choice shown in Fig. 8, as here difference with respect to the two scale choices is more pronounced, but similar conclusions hold for the higher choice, shown in Fig. 9. An approximately 10% difference is observable at lower $p_\perp$, with the $p_\perp^{max}$ choice leading to the larger result, consistent with the findings in [33]. We can see that in both cases the description of the data is poor, highlighting the importance of the systematic experimental uncertainties. However, once the data is allowed to shift by these errors this difference largely disappears, and good description of the data is achieved in all cases. The shift is somewhat larger for the $p_\perp^{max}$ case, with a $\sim 5\,(11)$ point increase in the χ^2 due the shift penalty found for the ATLAS (CMS) data, while for the higher R choice the overall shift penalty is only marginally increased, by 2 points. These findings are consistent with the trends found in Table 3. From Fig. 7 we can see that these results translate into a relative, although not complete, stability in the predicted PDFs. With a further reduction in the size of the systematic experimental uncertainties the differ-
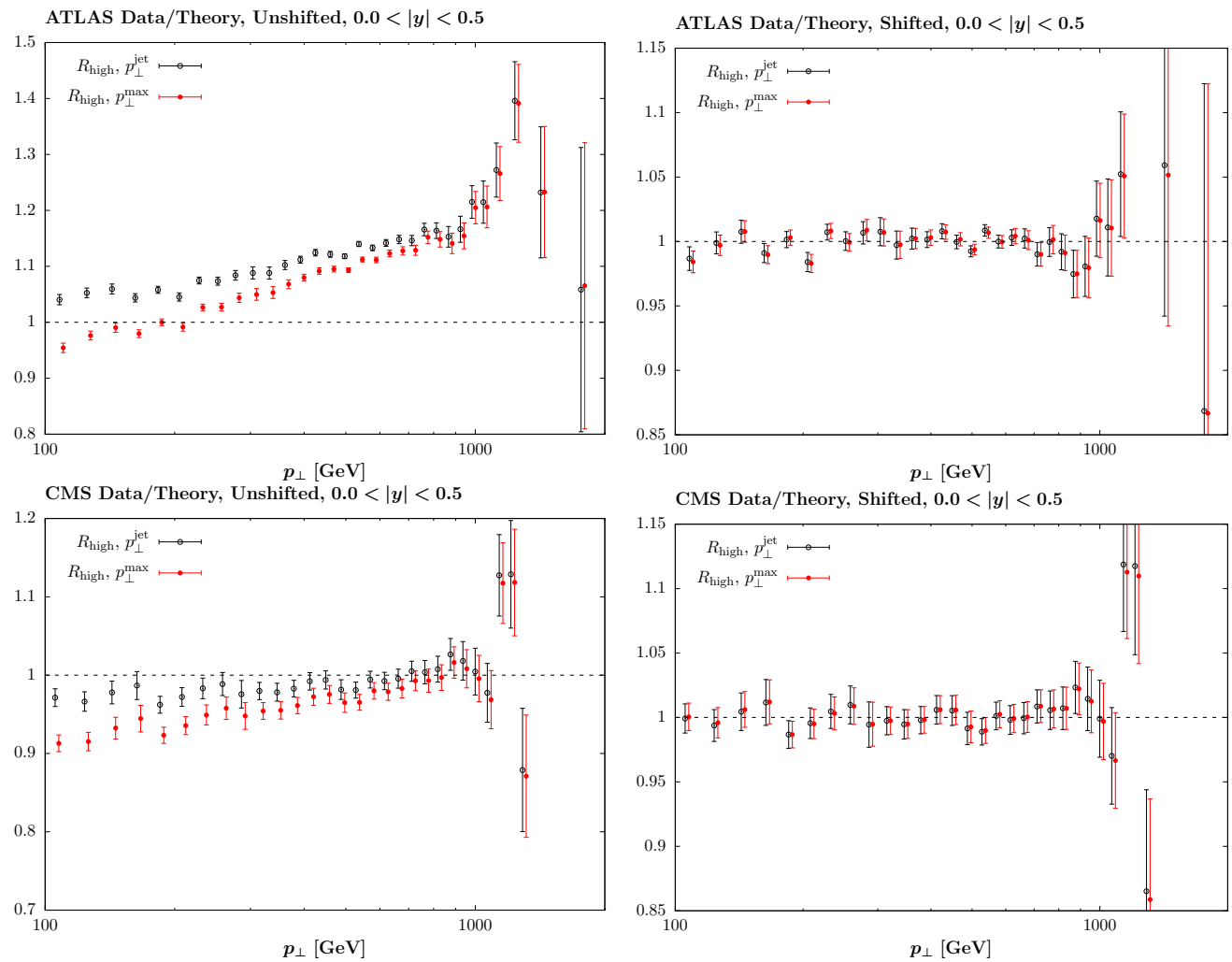
Fig. 9 As in Fig. 8, but with the higher R radius choice

ence may on the other hand become more pronounced, but no significant effect is observed with the 7 TeV data sets.

To investigate the impact of the NNLO corrections further, in Fig. 10 we again show the NNLO gluon resulting from the combined fit, but including the result with the NLO theory applied, i.e. excluding the NNLO K-factors in the fit. In the left (right) panel we show the result with the low (high) choice of jet radius. This therefore shows the impact of the new NNLO theory calculation on the gluon. We can see that in both cases the effect is reasonably small, but not negligible, leading to some additional softening in the gluon at high x. Indeed, for the high jet radius choice, the inclusion of the NNLO theory leads to a central value at high x which lies somewhat outside the uncertainty band of the baseline fit. The effect of using $p_\perp^{\mathrm{max}}$ instead as the scale choice is similar.

Finally, we consider the impact of including the Tevatron jet data [14,15] on the NLO and NNLO fits. As discussed above, for the NNLO case the full calculation is not yet pub-

licly available, and so continue to apply the threshold corrections of [10]. We show in Fig. 11 the result of including the Tevatron data alone, as well as the Tevatron and LHC data. The fit to the LHC jet data (in all cases with R_{high} and $p_\perp^{\mathrm{jet}}$) is also shown for comparison. As the NLO and NNLO cases are qualitatively quite similar, we will only discuss the NNLO fit below. For the fit to the Tevatron data, an increase in the central gluon at higher x is observed, consistent with its impact in the MSTW08 fit [38]. This is in contrast to the LHC data, which we have seen prefers a softer gluon at higher x, although up to $x \gtrsim 0.3$ these are consistent within PDF errors. Nonetheless some tension is observed, and indeed when including both the Tevatron and LHC data into the fit, the description deteriorates by about 10 and 8 points in comparison to the individual fits for the LHC and Tevatron, respectively. The resultant gluon is somewhat harder at high x than the LHC only fit, but still softer than the baseline. For clarity we do not include the PDF uncertainties in this case;
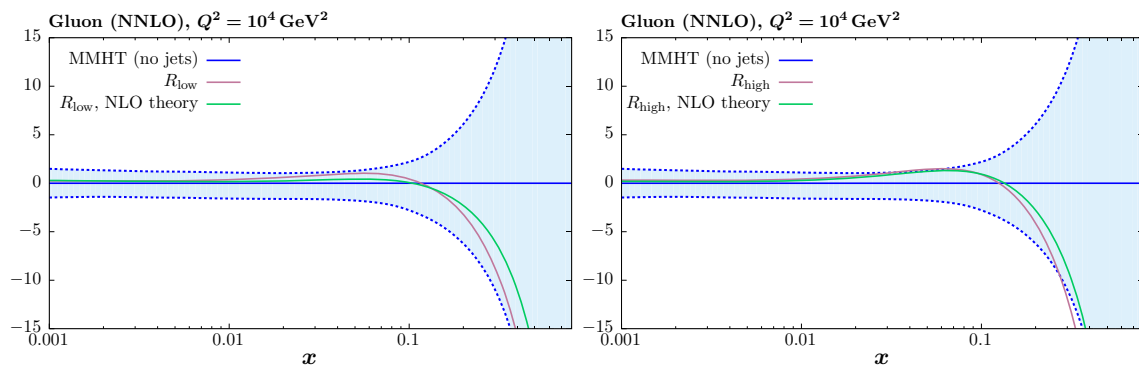
Fig. 10 The impact on the gluon PDF at NNLO of the ATLAS [23] and CMS [24] 7 TeV jet data. The percentage difference in comparison to the baseline fit, with no jet data included, is shown. Results are shown for the default NNLO fit, and to the same fit but using the NLO theory, i.e. with the NNLO K-factors omitted. The left (right) plots correspond to the 'low' and 'high' jet radii described in the text
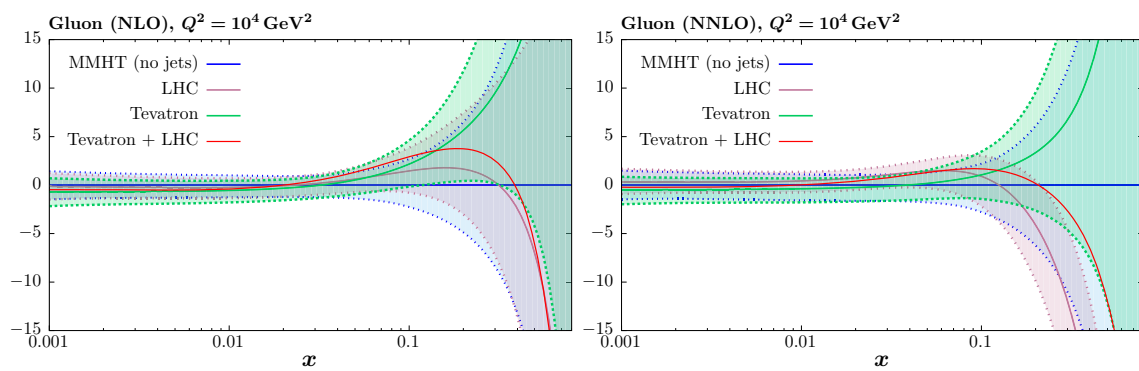

Fig. 11 The impact on the gluon PDF when fitting the ATLAS [23] and CMS [24] 7 TeV jet data and Tevatron data [14,15] individually, as well as including all data sets within the fit. For the LHC case R_{high} and $p_\perp^{jet}$ are taken. The NLO (NNLO) results are shown in the left (right) plots
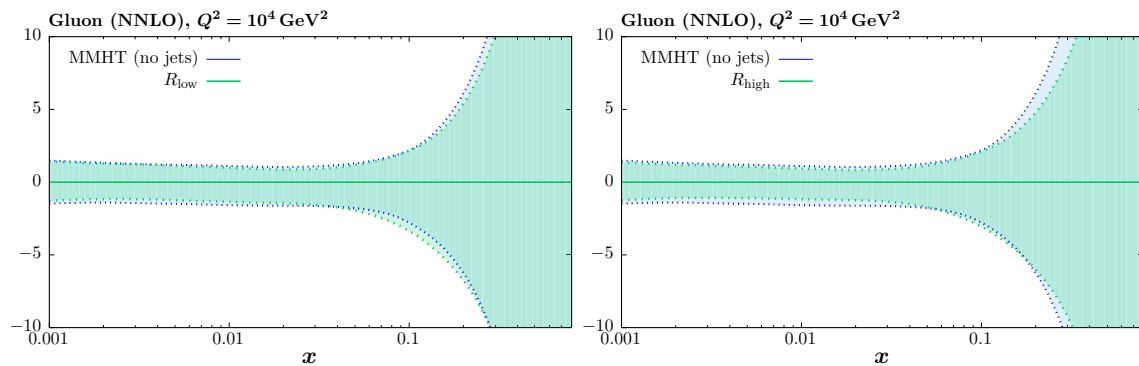
Fig. 12 The impact on the gluon PDF errors at NNLO of including the ATLAS [23] and CMS [24] 7 TeV jet data in the global fit. The percentage errors at 68% C.L. are shown, with the result of the baseline fit, with no jet data included, given for comparison. The left (right) plots correspond to the for 'low' and 'high' jet radii described in the text

these will be shown below. It will be interesting to see how this situation changes when the full NNLO corrections are included for the Tevatron predictions.

5.2 PDF uncertainties

In Fig. 12 we show the impact at NNLO of the ATLAS and CMS jet data on the gluon PDF uncertainty, for the two choices of jet radii. As in the case of the central values, we find that the difference due to the scale choice is minimal, and
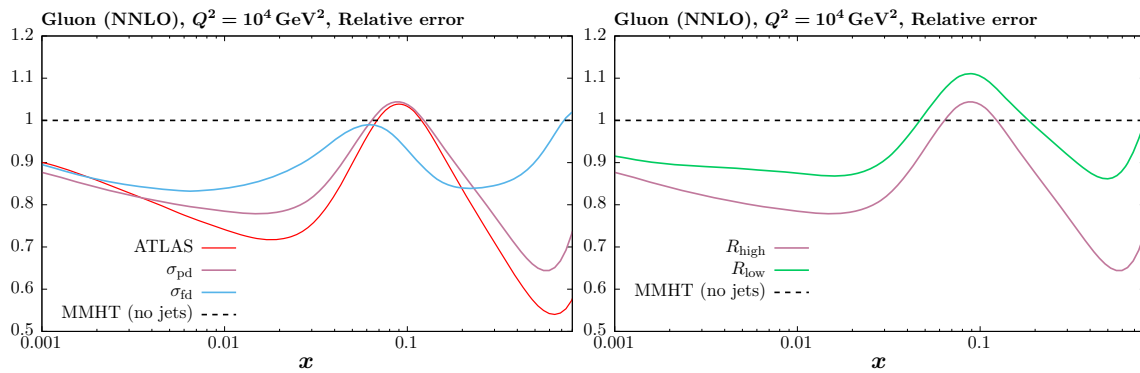
Fig. 13 The impact on the gluon PDF errors at NNLO of including the ATLAS [23] and CMS [24] 7 TeV jet data in the global fit. The ratio of the 68% C.L. errors to the baseline fit is shown for different choices of jet radius and treatment of systematic errors in the ATLAS case
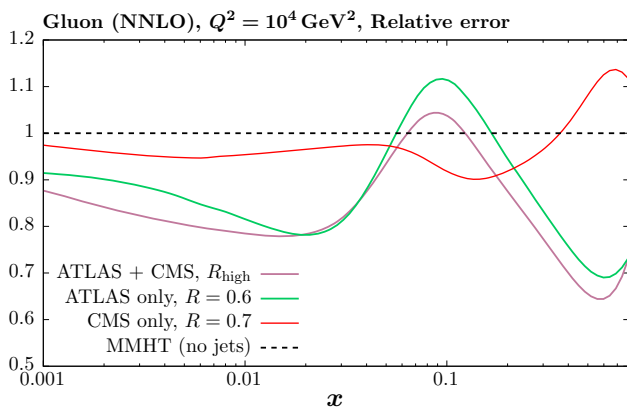

Fig. 14 The impact on the gluon PDF errors at NNLO of including the ATLAS [23] and CMS [24] 7 TeV jet data in the global fit. The ratio of the 68% C.L. errors to the baseline fit is shown. For the LHC case, R_{high} and $p_\perp^{\text{jet}}$ are taken

so we only show results for the $p_\perp^{\text{jet}}$ scale. The overall impact is seen to be moderate, although not negligible. To give a clearer comparison, we show the ratios to the baseline PDF uncertainty in Fig. 13 (left). For the higher R choice, for low and intermediate values of x the error reduction relative to the baseline ranges from $10 - 20\%$, but for the $x \sim 0.05 - 0.2$ there is little reduction and in some regions even a slight increase in the error. At high x there is again a reduction in the uncertainty, although as x approaches 1 and the jet data places little or no constraint, the quantitative result cannot be taken completely literally, as this will depend on the precise choice of PDF parameterisation. For the lower R choice the reduction in the PDF uncertainty is less significant, and the x region where this increases relative to the baseline is wider. In Fig. 13 (right) we should the results for the higher jet radius choice and for different treatments of the ATLAS systematic errors. We can see that the partial decorrelation leads to a similar, although in some places slightly less constraining, impact on the uncertainties across the entire x region in comparison to the default treatment, consistent with the impact

on the central values shown before. On the other hand, for fully decorrelated uncertainties the impact at high x in particular is much less constraining, although in the $x \sim 0.1$ region the uncertainties are in fact somewhat smaller.

In Fig. 14 (we show the impact of fitting the ATLAS and CMS data individually on the PDF uncertainties. We can see that, consistently with the results of the previous section, the impact of the ATLAS is generally larger, in particular at higher x, where the CMS data in fact lead to a somewhat larger uncertainty in comparison to the baseline. Including both the ATLAS and CMS data generally leads to some decrease in the uncertainties in comparison to the individual fits. In Fig. 15 we show the impact of the fits to the LHC and Tevatron data individually, as well as to the combination, at NLO and NNLO. We can see that with the exception of the intermediate $x \sim 0.05 - 0.1$ region at NNLO, the LHC data has a greater impact in reducing the PDF uncertainties. For the combined fit, the relative uncertainties reduce by $\sim 20\%$ across the entire x region at NNLO, while at NLO the reduction in uncertainty is somewhat milder in comparison to the LHC-only fit in the $x \sim 0.01 - 0.1$ region, while in the highest x regions the impact is somewhat larger. We can also see that, with the exception of this very high x region, which will in any case be sensitive to parameterisation effects, the impact of the NNLO fit on the gluon is more significant in comparison to the NLO for all data combinations. Again, it will be interesting to see how this situation changes at NNLO when the full NNLO corrections are included in the Tevatron predictions.

6 Conclusions and outlook

Inclusive jet production data has played a key role in constraining the partonic structure of the proton, and in particular the gluon at higher x, in global PDF fits. The availability of high precision jet data from the LHC combined with the
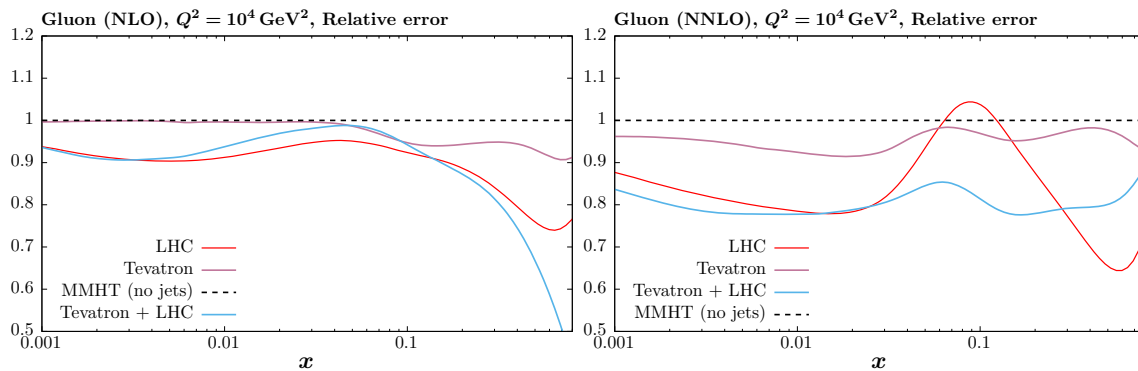
Fig. 15 The impact on the gluon PDF errors of including the LHC [23,24] and Tevatron [14,15] data in the global fit. The ratio of the 68% C.L. errors to the baseline fit is shown and the NLO (NNLO) case is shown in the left (right) figures. For the LHC case, R_{high} and $p_\perp^{\mathrm{jet}}$ are taken

recent release of the NNLO corrections to the hard cross section therefore provides an invaluable tool for high precision PDF constraints.

In this paper, we have presented a detailed study of the impact of LHC jet data on a PDF fit within the MMHT global fitting framework, at NNLO. We have observed that to reliably perform such a study, certain issues require a careful treatment. Namely we have had to address the choice of jet scale and radius, and the fact that a satisfactory description of the systematics dominated ATLAS data cannot by default be achieved across the full kinematic region. After analysing the structure of the systematic shifts induced in describing the ATLAS data, we have determined a straightforward and minimal method to improve the fit quality; by decorrelating two sources of systematic uncertainty in rapidity, a greatly improved description is achieved. Crucially, despite this change in the fit quality, we have shown that this only has a relatively small impact on the determination of the gluon itself in comparison to the default treatment. The result of our minimal approach should then be in line with a more complete consideration of different decorrelation scenarios permitted by experimental considerations. This suggests that despite this question of the default fit quality, these data can still be reliably included in a PDF fit. On the other hand, we have found that decorrelating all sources of uncertainty in rapidity, in essence the approach that is assumed if only one rapidity bin is fitted, leads to larger shifts. Some caution in applying such a procedure therefore appears to be warranted.

We have then presented the fit quality at NLO and NNLO when the ATLAS and CMS jet data are included in a MMHT fit, for both the inclusive and leading jet $p_\perp$ scale choices, and different values of the jet radius R. We find that some improvement is in general achieved when going to NNLO, with the exception of the $p_\perp^{\mathrm{jet}}$ and lower R choice, where there is a slight deterioration. The impact on the gluon PDF is qualitatively similar between orders. Although the theory predictions are quite different at lower jet $p_\perp$ when consider-

ing the two scale choices, we find that the fit quality including a proper treatment of the experimental systematics is in fact similar. Moreover, the impact on the gluon itself is very stable between the choices. This suggests that at least for the data sets under consideration in this paper, the effect of the choice of jet scale on PDF determination may not be as significant at NNLO as has sometimes previously been assumed.

In terms of the jet radius, the ATLAS data in particular has shown some preference of the larger ($R = 0.6$) choice, although again the impact on the gluon is relatively stable in comparison to the smaller choice. In all cases the jet data are found to consistently prefer a somewhat softer gluon at high x and a harder gluon in the intermediate x region, with in general some $\sim$ 10–20% *relative* reduction in the PDF uncertainty.

Thus, in this paper we have shown that LHC jet data may be reliably included in to global PDF fits at NNLO, while addressing in a minimal way the issue related to achieving a good description of the high precision, systematics dominated, ATLAS data across the whole kinematic region. We have only considered the 7 TeV data, for which the NNLO calculations are available. In future global fits, we will take our partially decorrelated treatment of the experimental systematic errors for these data sets. However, in the future we intend to confirm if the above conclusions hold in the case of the 8 and 13 TeV jet data from the LHC.[4] Moreover, this issue related to the description of the ATLAS data may become increasingly relevant in the high precision LHC era, and may warrant a more detailed study in the future of both the experimental and theoretical sources of uncertainty.

---

[4] We note that the Run-II data sets are by default made available for a single common value of jet radius, $R = 0.4$, preventing any comparison of different jet radii. Interestingly, we have seen here that a better description of the ATLAS data in particular may be achieved for a higher choice of jet radius, $R = 0.6$.

References

1. J. Gao, L. Harland-Lang, J. Rojo (2017). arXiv:1709.04922
2. ZEUS, H1, H. Abramowicz et al., Eur. Phys. J. **C75**, 580 (2015), arXiv:1506.06042
3. L.A. Harland-Lang, A.D. Martin, P. Motylinski, R.S. Thorne, Eur. Phys. J. C **75**, 204 (2015). arXiv:1412.3989
4. S. Dulat et al., Phys. Rev. D **93**, 033006 (2016). arXiv:1506.07443
5. NNPDF, R.D. Ball et al., Eur. Phys. J. **C77**, 663 (2017). arXiv:1706.00428
6. A. Accardi, L.T. Brady, W. Melnitchouk, J.F. Owens, N. Sato, Phys. Rev. D **93**, 114017 (2016). arXiv:1602.03154
7. S. Alekhin, J. Blümlein, S. Moch, R. Placakyte, Phys. Rev. D **96**, 014011 (2017). arXiv:1701.05838
8. M. Czakon, N.P. Hartland, A. Mitov, E.R. Nocera, J. Rojo, JHEP **04**, 044 (2017). arXiv:1611.08609
9. R. Boughezal, A. Guffanti, F. Petriello, M. Ubiali, JHEP **07**, 130 (2017). arXiv:1705.00343
10. N. Kidonakis, J.F. Owens, Phys. Rev. D **63**, 054019 (2001). arXiv:hep-ph/0007268
11. M.C. Kumar, S.-O. Moch, Phys. Lett. B **730**, 122 (2014). arXiv:1309.5311
12. D. de Florian, P. Hinderer, A. Mukherjee, F. Ringer, W. Vogelsang, Phys. Rev. Lett. **112**, 082001 (2014). arXiv:1310.7192
13. X. Liu, S.-O. Moch, F. Ringer (2017). arXiv:1708.04641
14. D0, V.M. Abazov et al., Phys. Rev. **D85**, 052006 (2012). arXiv:1110.3771
15. CDF, A. Abulencia et al., Phys. Rev. **D75**, 092006 (2007). arXiv:hep-ex/0701051 **(Erratum: Phys. Rev. D75, 119901 (2007))**
16. ATLAS, G. Aad et al., Phys. Rev. D **86**, 014022 (2012). arXiv:1112.6297
17. ATLAS, G. Aad et al., Eur. Phys. J. **C73**, 2509 (2013). arXiv:1304.4739
18. CMS, S. Chatrchyan et al., Phys. Rev. **D87**, 112002 (2013), arXiv:1212.6660 **(Erratum: Phys. Rev.D87,no.11,119902(2013))**
19. NNPDF, R.D. Ball et al., JHEP **1504**, 040 (2015). arXiv:1410.8849
20. A. Gehrmann-De Ridder, T. Gehrmann, E.W.N. Glover, J. Pires, Phys. Rev. Lett. **110**, 162003 (2013). arXiv:1301.7310
21. J. Currie, A. Gehrmann-De Ridder, E.W.N. Glover, J. Pires, JHEP **01**, 110 (2014). arXiv:1310.3993
22. J. Currie, E.W.N. Glover, J. Pires, Phys. Rev. Lett. **118**, 072002 (2017). arXiv:1611.01460
23. ATLAS, G. Aad et al., JHEP **02**, 153 (2015). arXiv:1410.8857 **(Erratum: JHEP09,141(2015))**
24. CMS, S. Chatrchyan et al., Phys. Rev. D **90**, 072006 (2014). arXiv:1406.0324
25. A.D. Martin, M.G. Ryskin, Eur. Phys. J. C **77**, 218 (2017). arXiv:1702.01663
26. ATLAS, M. Aaboud et al., JHEP **09**, 020 (2017). arXiv:1706.03192
27. ATLAS, M. Aaboud et al., (2017). arXiv:1711.02692
28. CMS, V. Khachatryan et al., JHEP **03**, 156 (2017). arXiv:1609.05331
29. CMS, V. Khachatryan et al., Eur. Phys. J. **C76**, 451 (2016). arXiv:1605.04436
30. Z. Nagy, Phys. Rev. D **68**, 094002 (2003). arXiv:hep-ph/0307268
31. T. Carli et al., Eur. Phys. J. C **66**, 503 (2010). arXiv:0911.2985
32. https://applgrid.hepforge.org. Accessed 21 Mar 2018
33. J. Currie et al., Acta Phys. Pol. B **48**, 955 (2017). arXiv:1704.00923
34. L.A. Harland-Lang, R. Nathvani, R.S. Thorne, A.D. Martin, Acta Phys. Pol. B **48**, 1011 (2017). arXiv:1704.00162
35. L.A. Harland-Lang, A.D. Martin, P. Motylinski, R.S. Thorne, Eur. Phys. J. C **76**, 186 (2016). arXiv:1601.03413
36. ATLAS, G. Aad et al., Eur. Phys. J. **C75**, 17 (2015). arXiv:1406.0076
37. U. Blumenschein, C. Gwenlan, B Malaescu, M. Schott **(private communication)**
38. A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, Eur. Phys. J. C **63**, 189 (2009). arXiv:0901.0002