Eur. Phys. J. C (2022) 82:90
https://doi.org/10.1140/epjc/s10052-022-10028-2

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

QED parton distribution functions in the MSHT20 fit

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

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

Received: 17 November 2021 / Accepted: 13 January 2022

Abstract We present the MSHT20qed set of parton distribution functions (PDFs). These are obtained from the MSHT20 global analysis via a refit including QED corrections to the DGLAP evolution at $\mathcal{O}(\alpha)$, $\mathcal{O}(\alpha\alpha_S)$ and $\mathcal{O}(\alpha^2)$, and containing the photon PDF of the proton. As in the previous MMHT15qed study we use an input distribution for the photon that is derived from the LUXqed formulation, and find good consistency for the photon PDF with that of MMHT15qed, as well as with other recent sets. We also present a set of QED corrected neutron PDFs and accompanying photon distribution, and provide the photon PDF of the nucleons separated into elastic and inelastic contributions. We assess the general expectations for the impact of photon-initiated (PI) corrections to processes entering PDF fits, and review the effect of QED corrections on the other partons and on the fit quality, where electroweak corrections (including PI production) are appropriately added to the cross sections wherever possible. We explore the phenomenological implications of this set by comparing to a variety of benchmark cross sections, finding small but significant corrections due to the inclusion of QED effects in the PDFs.

1 Introduction

The level of precision aimed for at the Large Hadron Collider (LHC) is now reaching the point where the inclusion of electroweak (EW) corrections in theoretical predictions is often becoming mandatory. This requires that EW corrections are applied not only to the partonic cross sections but also to the corresponding Parton Distribution Functions (PDFs), with QED corrections forming a part of this. These can be included by supplementing the DGLAP [1–5] evolution of the PDFs to include QED parton splittings. This automatically results in the photon becoming a constituent parton of the proton, leading to photon-initiated (PI) sub-processes entering as corrections to the leading QCD cross section for processes such as Drell–Yan [6], EW boson–boson scattering [7] and Higgs production with an associated EW boson [8]. Additionally, semi-exclusive [9,10] and exclusive PI production of states with EW couplings has significant potential as a probe of SM and BSM physics.

The inclusion of QED corrections in DGLAP, and the corresponding photon PDF, has a long history. MRST provided the first publicly available QED set [11], using DGLAP splitting kernels at $\mathcal{O}(\alpha)$ in QED and modelling the input photon as arising radiatively from the quarks below input. Subsequent sets either used similar phenomenological models [12], or constrained the photon by utilising the rather limited sensitivity of DIS and Drell–Yan data via PI final states [13,14]. Both approaches lead to photon PDF uncertainties of at least 10% and often more. Moreover, the distinction between the elastic and inelastic photon emission was rarely considered. In [9,15,16] it was shown how a more accurate determination of the photon distribution at input could be found by using the experimentally well determined elastic form factors of the proton. More generally, as discussed long ago in e.g. [17], the contributions from elastic and inelastic emission to the photon PDF are directly related to the corresponding structure functions, $F_{1,2}^{el}$, $F_{1,2}^{inel}$; this idea has been revisited over the years in [18–21]. However, it has recently been placed within a rigorous and precise theoretical framework by the LUXqed group [22,23], who consequently provided a publicly available photon PDF with uncertainties determined by those on the structure functions used as input, i.e. at the level of a few percent. Additionally, QED DGLAP splitting kernels have now been calculated to $\mathcal{O}(\alpha\alpha_S)$ [24] and $\mathcal{O}(\alpha^2)$ [25]. These are implicit in the LUXqed approach, and are easily implemented in DGLAP evolution codes.

Hence, it is now possible to be far more precise and confident about the effects of QED modified partons, the photon distribution and their impact on cross section calculations. The first global PDF set including a photon distribution based

[a] e-mail: lucian.harland-lang@physics.ox.ac.uk (corresponding author)

Published online: 29 January 2022

on the LUXqed approach was produced by the NNPDF group [26]. This was soon followed by QED corrected PDFs based on the MMHT14 PDFs (in practice also including the final HERA combined cross section data, so more similar to the slightly modified PDFs in [27]), which followed a LUXqed-inspired approach [28] that we will summarise in the following section. More recently the CT group has also produced PDFs with QED corrections and a LUXqed-inspired photon distribution [29]. The photon distributions in these sets now all have uncertainties of a few percent, and are all broadly consistent with each other, representing a huge improvement in the knowledge of the photon content of the proton. On the other hand, some care must be taken when claiming equivalently high precision in the corresponding PI cross sections, which in the above studies are only calculated at LO in α; these will therefore have significantly larger scale variation uncertainties than the percent level uncertainty due to the photon PDF. As we will discuss in the following section, of the processes entering global PDF fits, the PI contributions to off-peak lepton pair production are by far the most dominant ones. With this in mind, when calculating the PI corrections in this case we will follow the approach of [30,31], which applies the structure function (SF) approach to directly calculate the dominant PI contribution to lepton pair production away from the Z peak. This provides percent level precision in the cross section prediction here, bypassing the issue of large LO scale variations. For other processes, a standard EW K-factor approach can be taken (or fast interpolation grids, as presented in [32]), although in the majority of cases the impact of PI production is found to be marginal at the current level of precision.

The outline of the paper is as follows. In Sect. 2.1 our procedure for including the QED effects within the MSHT framework is briefly described; here we start with the latest MSHT20 [33] QCD-only partons. The approach is very similar to that adopted in [28], but involves a minor change to the photon splitting function, the inclusion of more PI initiated contributions to fitted data sets, and the application of the SF approach to calculate these. In Sect. 2.2 we assess the general expectations for the impact of photon-initiated (PI) corrections to processes entering PDF fits such as MSHT20. In Sect. 3 we present the fit quality of the QED corrected global fit. As before it is found that QED corrections cause a slight deterioration in fit quality, and we discuss this. In Sect. 4 we describe the impact on the QCD partons as well as presenting the corresponding photon PDF. We compare to the PDFs without the inclusion of QED effects, as well as to the MMHT15qed set and the results of other groups. We also examine the phenomenological consequences of QED corrections by comparing a number of benchmark cross sections to those obtained using the QCD-only PDFs. This leads to small, but not insignificant changes. In Sect. 5 we present the differences between the neutron PDF up and down quarks, demonstrating the effects of QED-induced isospin violation, and between the proton and neutron photon PDFs. In Sect. 6 we describe the availability of the PDFs. As well as the conventional set of QED altered PDFs, we provide grids for the photon PDF separated into its elastic and inelastic components, and a consistent set of QED corrected neutron PDFs. Finally, in Sect. 7, we conclude.

2 Including QED effects in the MSHT framework

2.1 Modifications to PDFs

The treatment of QED corrections follows that of the MMHT15qed set, outlined in [28]. In particular, the photon PDF is calculated using a suitable modification of the LUXqed formula [22,23]:

$$
\begin{aligned}
x\gamma(x,Q_0^2) = {} & \frac{1}{2\pi\alpha(Q_0^2)}\int_x^1 \frac{dz}{z}\left\{\int_{\frac{x^2m_p^2}{1-z}}^{Q_0^2}\frac{dQ^2}{Q^2}\alpha^2(Q^2)\right. \\
& \times\left[\left(zP_{\gamma,q}(z)+\frac{2x^2m_p^2}{Q^2}\right)F_2(x/z,Q^2)\right. \\
& \left.\left.-z^2F_L(x/z,Q^2)\right]-\alpha^2(Q_0^2)\left(z^2+\ln(1-z)zP_{\gamma,q}(z)\right.\right. \\
& \left.\left.-\frac{2x^2m_p^2z}{Q_0^2}\right)F_2(x/z,Q_0^2)\right\}, \qquad (1)
\end{aligned}
$$

at input scale $Q_0 = 1$ GeV. The structure functions $F_{2,L}$ receive contributions from both elastic and inelastic photon emission, and are precisely determined using experimental data on lepton–proton scattering. In more detail, the elastic structure functions are determined from a fit by the A1 collaboration [34], and the inelastic via the HERMES GD11-P fit [35] and data from the CLAS collaboration [36], in the continuum and resonance regions, respectively. Uncertainties on the photon are included due to several sources (see [28]) for further details): the experimental uncertainty on the elastic structure functions, the value of the R ratio used to determine the inelastic F_L, the value of the threshold W between which the HERMES and CLAS data are used for the inelastic structure functions, the experimental uncertainty on the CLAS data for the resonance region, the uncertainty in the HERMES fit for the continuum region, and the modelling of renormalon corrections to the quark evolution. Each of these is included as a separate error eigenvector pair, while in addition the standard PDF uncertainties due to the fit of the QCD partons as in the MSHT20 set [33] are included; these give 32 eigenvectors, and hence in total there are 38 eigenvectors for the MSHT20qed set.[1]

[1] We in fact find that a more stable set of eigenvectors is found by fixing the 3rd rather than the 4th Chebyshev polynomial associated with the

QED corrections to DGLAP evolution are again included up $O(\alpha^2)$ and $O(\alpha\alpha_S)$. The treatment of these is therefore broadly the same as in MMHT15qed, with however one exception. Namely, we now choose to include leptonic loop contributions to the photon–photon splitting function, which at $O(\alpha)$ is proportional to the sum

$$\sum_i e_i^2 = N_C \sum_q^{n_F} e_q^2 + \sum_l^{n_L} e_l^2. \tag{2}$$

In the MMHT15qed fit, the second term was omitted, as the inclusion of this strictly implies that we must include lepton PDFs (see [37,38] for further studies), which in principle enter due to splittings of the form $\gamma \to l\bar{l}$. As these are not present in the MSHT framework the inclusion of lepton loops in $P_{\gamma\gamma}$ would in particular lead to some small amount of violation of the momentum sum rule, due to absence of $l \to l\gamma$ splitting contributions.[2] However, as discussed in [28] (see Fig. 24 of that paper) the impact of such lepton loops is not completely negligible on the photon PDF itself, reducing it by up to 2% at LHC scales; the $P_{\gamma\gamma}$ splitting function leads to a reduction in the photon as it undergoes DGLAP evolution, and leptonic loop contributions increase this effect. Indeed the lack of these contributions, which are included in both the NNPDF31luxqed [26], and CT18qed, CT18lux [29] sets, can be seen in Fig. 10 of [29] to potentially lead to a rather systematic offset of the MMHT15qed photon PDF with respect to these sets. Moreover, as discussed in [30] the factorization scale of the photon PDF is directly tied to the scale at which one should evaluate the renormalization scale of α associated with the coupling of the photon to the production subprocess. Given the running of α will of course in general include lepton loops, their absence in $P_{\gamma\gamma}$ is potentially inconsistent. For these reasons, we now choose to include them. We will comment on their impact in the following sections.

2.2 Including photon-initiated production in a global PDF fit

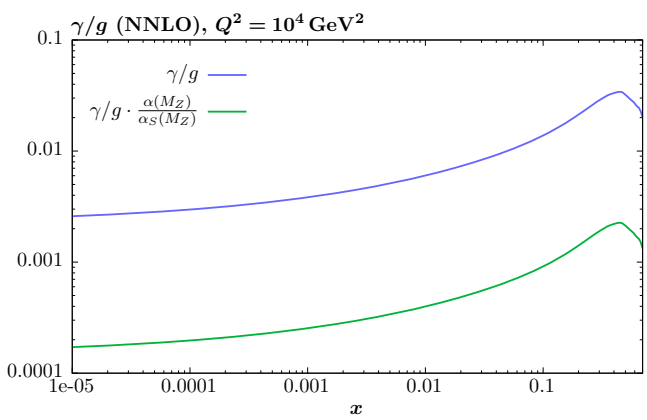
In addition to modifying the DGLAP evolution of the QCD partons, QED corrections will also enter directly into the calculation of the cross section predictions entering the fit. This will include PI production, which forms a subset of the broader class of EW corrections. It is therefore useful to assess on general grounds the size that such PI contributions are expected to have for processes that enter current global PDF fits.

As discussed above, the PI channel can play an important role in the production of objects with EW couplings, such as EW boson–boson scattering and Higgs boson production with an associated EW boson; however, these are not currently included in global PDF fits. Of the processes entering the global fits, such as MSHT20, one might broadly consider DIS, inclusive jets, $t\bar{t}$ and lepton pair production as receiving potentially relevant PI contributions. For the production of strongly interacting particles in hadron–hadron collisions such as $t\bar{t}$ and jets, the leading PI contribution comes from replacing an initial-state gluon with a photon, for all relevant diagrams, e.g. for $gg \to t\bar{t}$ we instead have $\gamma g \to t\bar{t}$. To estimate the expected suppression in this case in Fig. 1 we show the ratio of the photon to gluon PDFs at a representative scale $Q^2 = 10^4$ GeV^2 (although the precise choice does not effect the results significantly). We can see that photon PDF is suppressed by $\sim$ 2 orders of magnitude or more, with the exception of the largest x values where the suppression is a little less, though still significant. This is in particular a significantly larger suppression than one might rather naively expect from simple $\sim \alpha/\alpha_S$ scaling in the corresponding PDFs. In the $g\gamma$-initiated process, we will receive an additional suppression $\sim \alpha/\alpha_S$ due to the γq coupling, and hence we also show the same ratio, but weighted by this factor. We can see that the expected correction enters at the per mille level, and hence in practice may be expected to enter at roughly the same level as N^{3}LO QCD corrections or even less, rather than the NNLO order one might naively expect from the fact that $\alpha_S(M_Z)^2 \sim \alpha(M_Z)$.

Now in reality, the true result will of course require a full calculation, accounting for the colour factors, differing diagrams that enter and the fractional quark charge; the former effects will broadly speaking enhance the PI contribution, whereas the latter will suppress it. This therefore should, and can, be verified by explicit calculation. This is achieved in e.g. [26,39] for the case of $t\bar{t}$ production, and in fact the level of suppression is found to be larger than that expected from Fig. 1 (left), by roughly a further order of magnitude. For jet production, we can expect the suppression to be even greater, due to the presence of pure gluonic channels, which of course receive no leading PI contribution, although the heavy top quark mass in the $t\bar{t}$ case makes the comparison a little less direct. For DIS, the same argument applies as above, but now it is the NLO in QCD gluon-initiated diagram that receives the leading correction as above, and hence the suppression is expected to be an $O(\alpha_S)$ more. We therefore conclude that the impact of PI contributions to these processes is in general not expected to be significant at the NNLO QCD precision

Footnote 1 continued
down valence, but otherwise the same parameters are fixed as in the previous MSHT20 PDFs when generating the eigenvectors.

[2] In fact the MMHT15qed PDFs had a momentum violation of +0.00008% due to contributions to the photon from inelastic and higher twist sources above Q_0^2. The effect from lepton splitting from photons is in the opposite direction, and of comparable, though slightly smaller size, so in total momentum is more closely conserved in MSHT20qed than in MMHT15qed.



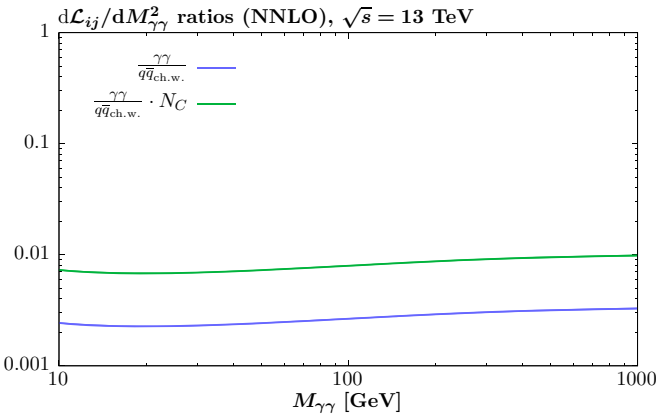


Fig. 1 (Left) The ratio of the γ to the g distributions at $Q^2 = 10^4\,\mathrm{GeV}^2$. In addition the same ratio, but weighted by $\alpha(M_Z)/\alpha_S(M_Z)$, is shown. (Right) Ratio of the $\gamma\gamma$ to the charge weighted $q\overline{q}$ luminosity at 13 TeV. Also shown is the same ratio, weighted by inverse of the LO QCD DY colour factor. In both plots the PDFs result from NNLO fits to the MSHT20 dataset, with QED effects included

level,[3] although this should of course be verified by explicit calculation, given it may not apply uniformly in all kinematic regions and for all processes.

We are therefore left to consider lepton pair production, for which the final state is of course not strongly interacting, and hence the arguments above do not apply. Here, as is well known, the t-channel $\gamma\gamma \to l^+l^-$ process can be of much greater phenomenological relevance. To demonstrate this, in Fig. 1 (right) we show the ratio of the $\gamma\gamma$ partonic luminosity to the e_q^2 charge weighted $q\overline{q}$ luminosity, relevant for $q\overline{q} \to \gamma^*$ DY production; see e.g. [41] for a definition of these. We can see that this ratio is at the level of a few per mille, however once we divide by the LO QCD DY colour factor, $1/N_C$, this is at the level of 1%, as can be seen in the figure. When we also account for the t-channel enhancement of the $\gamma\gamma$ cross section:

$$\frac{|\mathcal{M}(\gamma\gamma \to l^+l^-)|^2}{|\mathcal{M}(q\overline{q} \to l^+l^-)|^2} \propto \frac{\hat{s}^2}{\hat{u}\hat{t}} \geq 4, \tag{3}$$

where $\hat{s}, \hat{t}, \hat{u}$ are the usual partonic Mandlestam variables, we can expect this to enter at the $\sim 5-10\%$ level, which indeed it is seen to [26,31,42]. On the other hand, in the Z peak region the DY process receives a significant resonant enhancement, and hence the above conclusions do not hold; as demonstrated in [26,31,42] in this region the relative contribution from the $\gamma\gamma \to l^+l^-$ subprocess is at the per mille level. We note that there are in addition $\gamma \to q\overline{q}$ splittings that can play a role here, that is from the mixed $\gamma q(\overline{q})$ initial state. However, here we are once again in the situation described above for case of $t\bar{t}$ and jet production, and so again will expect per mille level corrections from this; in fact as the corresponding $g \to q\overline{q}$ diagram now only enters at NLO in QCD we might expect the suppression to be larger still, although this will depend on e.g. the impact of colour factors and fractional quark charges. In e.g. [43,44] results are presented for this channel using the MRST2004QED [11] set, and are indeed found to enter at the level of a few per mille. While this set is certainly now outdated, as seen in [23] it lies within $\sim 50\%$ of the LUXqed photon PDF, and hence the true result will also be of this order. As discussed further below, an additional point to note is that, for the current highest precision on-peak Z data entering the fit, from the ATLAS collaboration, the t-channel $\gamma\gamma \to l^+l^-$ PI contribution is in any case subtracted directly at the data level. We finally note that although the inclusive lepton pair signal has been discussed above, very similar conclusions will hold for measurements of the lepton pair $p_\perp$ distribution, or production in association with jets, which also enter PDF fits.

We therefore conclude that, of the processes that currently enter global PDF fits, lepton pair production away from the Z peak region receives by far the largest PI contribution, which as discussed above can enter at the 5–10% level. With this in mind, to calculate this we use the `SFGen` Monte Carlo implementation of the Structure Function (SF) approach discussed in [31,42]. This automatically provides a percent-level precision calculation of the lepton pair production cross section in this region, without the significant scale variation uncertainties that are present in the LO collinear calculation. The underlying experimental inputs for the proton structure functions are precisely the same as in the current study for the photon PDF, and the dominant theoretical uncertainties are due to these. In practice the SF prediction will depend on the underlying QCD parton set used to calculate the inelastic structure functions in the high Q^2 region (as is the case for the photon PDF), and so an iterative procedure should be

[3] A similar argument can be applied to the case of e.g. isolated photon production, which is included in the NNPDF4.0 fit [40].

adopted in the fit, where the PI cross section is recalculated using the QCD partons from each iteration of the fit, until convergence is reached. We have performed such a procedure here, although in reality we find this converges after one iteration, and we have confirmed that calculating the PI cross section with the MSHT20 QCD partons, without such an iteration, gives an almost identical result in terms of the fit quality and the extracted PDF set. However, this may not remain true in future fits, where more PI corrections may be included in phenomenologically relevant regions.

In addition to the ATLAS high mass DY data, we now also include the PI contribution to the CMS 7 TeV double differential DY data [45],[4] which extends above and below the Z peak region. There are also ATLAS measurements of W, Z production at 7 and 8 TeV [47,48], which extend below and above the Z peak region, as well as the 8 TeV measurement of the Z boson (or more precisely, dilepton) $p_\perp$ distribution. In these cases we could in principle include the PI contribution via the SF approach, however as discussed in [33], for these ATLAS measurements this contribution is already subtracted from the data. We note that the procedure for doing this is theory-dependent and indeed in some cases uses rather outdated photon PDF sets/calculations. It is therefore certainly recommended that such subtractions are not made for future datasets. However, as a result of this, we do not include PI contributions for these processes here. We note that in [33] we did include the PI contribution to the ATLAS 8 TeV DY data [48], but as discussed above this should not have been done, given the data is subtracted. This has now been removed, and indeed doing so improves the fit quality, as we will see. However the impact on the PDFs is generally very small, and hence while for completeness in Sect. 4 we will compare against a QCD-only set resulting from a fit with this correction applied, in reality for public use one can take the MSHT20nnlo set as a baseline for comparison, and indeed for wider use; we therefore do not make this QCD-only set available for release.

Finally, we note that one could alternatively include a calculation of the PI contribution to lepton pair production in collinear factorization that goes beyond LO in α, in order to improve the precision of the large LO scale variation uncertainty. However, as discussed in [31], even at NLO the scale variation in this case can remain larger than the PDF uncertainty. Therefore, given the rather large size of these contributions away from the Z peak region, we consider the use of the SF calculation as the more appropriate choice here.

On the other hand, when considering the broader class of PI observables that currently enter a global PDF fit, for example lepton pair production in the Z peak region (where initial-state $\gamma \to q\bar{q}$ splittings become relevant), $t\bar{t}$ or jet production, one can more straightforwardly make use of collinear factorization. Indeed, in practice this becomes essential, as here the PI contribution will in general form a subset of the larger class of QED/EW corrections. These can more broadly be relevant at the NNLO QCD precision level, even if the dominant contribution from these often comes from EW Sudakov logarithms that are therefore unrelated to pure QED corrections and PI production. The conclusion in this case is that, for the processes such as those considered above, where PI production enters at the per mille level, it is certainly appropriate to include them as part of the broader NLO EW corrections in a standard K-factor manner, provided a photon PDF based on the LUXqed approach is used to calculate these K-factors. If this is the case, then the difference due to the precise choice of photon PDF will be at the level of a few percent of a per mille correction, i.e. clearly negligible and in any case significantly less than the scale variation uncertainty in a LO PI calculation.

In the MSHT20 fit, we include EW corrections for a range of processes. In more detail, for inclusive jet production we include these for all 7 and 8 TeV LHC datasets [49–51] using K-factors evaluated from the calculation of [52] (see also [53]). These do not include QED corrections, and therefore PI production, arguing that the dominant contribution is from the pure weak corrections (a distinction that can be made in a gauge invariant way in this case), due to their Sudakov logarithmic enhancement; the size of the overall EW corrections, which is driven by this, can be as large as $\sim 10\%$ at the highest jet $p_\perp$ values, though it is generally rather less.

For the ATLAS 8 TeV Z $p_\perp$ data we apply the same EW K-factors as used in the ATLAS analysis [54], which are derived from the calculation of [55]. These include mixed γq PI production, and are found to enter at the per mille level and be significantly smaller than the other EW corrections. However, these make use of the now outdated MRST2004QED [11] set, and hence in principle it should not be relied upon. In practice, as discussed above this set lies within $\sim 50\%$ of the LUXqed photon PDF. Hence any update to account for this would only modify the eventual EW K-factor at the per mille level, and possibly less. We therefore for simplicity continue to make use of these K-factors, which correctly account for the dominant EW correction. In future analyses a photon PDF based on the LUXqed formalism should be used to calculate such K-factors (as in e.g. [56]), although in reality as discussed above the precise choice will not matter. The total size of EW corrections is as large as $\sim 20\%$ at high $p_\perp^{ll}$, though is generally less than this [54].

For the ATLAS high precision W, Z data [47] we apply the same EW K-factors as used in the ATLAS analysis. These also include mixed γq corrections, in this case derived from the MCSANC generator [57,58], but again using the MRST2004QED set. However, once again these are observed to enter at the per mille level, and hence we can safely apply

[4] We recall [33] that the 8 TeV measurement is not included in the fit, due to apparent issues in the data [46].

these, even if future calculations should use an updated photon PDF set. The total size of the EW corrections is $\sim 0.5\%$ at intermediate and high masses, but $\sim$ 6% in the lowest mass region [47]. For the ATLAS 8 TeV high mass DY [6] we include NLO EW corrections via the K-factors described in [14], which are calculated using `FEWZ` [59] (this publicly available tool can provide EW corrections for all such cases). These have the LO $\gamma\gamma \to l^+l^-$ PI contribution explicitly subtracted; this $\gamma\gamma \to l^+l^-$ component will instead be calculated within the SF approach. The impact of other NLO EW corrections is then introduced by weighting the pure QCD calculation by the EW K-factor. The breakdown between NNLO QCD and NLO EW is not given, but the combined K-factor is at the percent level in all regions.

For all other DY datasets, we do not include EW corrections, judging the impact of these to be smaller than the precision of the data. With the exception of the CMS 7 TeV double differential DY data [45], these remaining datasets do not extend beyond the Z peak region, and hence the impact of EW corrections will be small. For the CMS data, the experimental uncertainties are sizeable at higher masses, where EW corrections will be most important. Moreover, their impact is considered across all mass bins in [45] and found to have a negligible impact within experimental uncertainties. We therefore do not include these in this case.

Finally, for the ATLAS and CMS single differential top quark pair production data [60,61] we use the EW K-factors calculated in [62] (based on the earlier study in [39]). These include the γg initiated channel, calculated using the LUXqed photon PDF [22]. However, entirely consistently with the discussion above, this contribution is found to be negligible.

3 Fit quality

In this section we describe the changes in fit quality in the NNLO global fit from the addition of QED effects, with a fixed value of $\alpha_S(M_Z) = 0.118$, as this corresponds to the default fit value. Allowing this to be free gives a marginal improvement in the fit quality, by about 1 point in χ^2, and a preferred value of $\alpha_S(M_Z) = 0.1176$, that is the same at the quoted level of precision to the case of the pure QCD fit. Table 1 provides the comparison for the non-LHC datasets, whilst Table 2 shows the LHC datasets. The final column of each table highlights the main changes in the fit qualities in terms of the χ^2 of each dataset. Overall there is a worsening of the fit quality upon inclusion of QED effects of approximately 24 points in χ^2. In the previous MMHT2015qed fit [28], the fit quality was also found to deteriorate, but by a rather smaller amount of $\sim$ 7 points, albeit for rather fewer data points. Indeed, there are now substantially more datasets included in the global fit, with in particular various high precision LHC datasets being included. Some of the changes in χ^2 are similar to before, with the BCDMS and HERA e^-p data fit quality deteriorating by similar amounts to the MMHT2015qed case, whilst the NuTeV F_2 data improves by a comparable amount to that observed in MMHT2015qed. In the BCDMS case this is due to $q \to q\gamma$ emission, which leads to a quicker high-x quark evolution, i.e. mimicking a slightly larger value of α_S, which the BCDMS data is known to disfavour [106].

The new datasets in MSHT20 which show the largest sensitivity to the presence of QED effects in the fit are the ATLAS 8 TeV Z p_T, CMS 8 TeV inclusive jets and ATLAS 7 TeV inclusive jets datasets, with the first two of these worsening by 10 and 5.5 points in χ^2 and the last improving by 3.5 points. All three of these datasets are precise and sensitive to the shape of the gluon in the high x region, and given this is altered by the addition of QED effects in the refit (see Fig. 4 (right) later) it is perhaps unsurprising that their fit qualities alter. For the inclusive jet datasets, given the relatively small changes in the χ^2 and their sensitivity to a range of x values, it is difficult to be precise about the causes of the improvement in the ATLAS 7 TeV jets and the worsening of the CMS 8 TeV jets. Nonetheless, we have seen previously that these datasets pull on the high x gluon in different ways and display some tension, therefore it is consistent at least that one worsens in fit quality whilst the other improves. As for the ATLAS 8 TeV Z p_T data, this has already been demonstrated to be in tension with many of the datasets sensitive to the high x gluon in the MSHT20 fit [33]. This tension clearly worsens upon the inclusion of the QED effects, and indeed if this dataset is removed the deterioration in fit quality when including QED effects reduces to 12 points in χ^2, more similar to the effect of the inclusion of QED effects observed in MMHT2015qed.

Further changes in fit quality tend to be small and spread out across the datasets, with several non-LHC datasets improving marginally, while many of the newer LHC datasets worsen in fit quality very slightly upon the addition of QED effects, including the ATLAS 7 and 8 TeV high precision W, Z measurements. We note that the fit quality for the QCD fit to the ATLAS 8 TeV DY data [48] is $\chi^2 \sim 76$, which is $\sim$ 10 points better than the value reported in [33]. As discussed in the previous section, this is due to the removal of the PI contribution, which was incorrectly included in the MSHT20 fit, given this is subtracted from the data.

We note that, as discussed in the previous section, explicit PI contributions are only included for two datasets, namely the ATLAS 8 TeV high mass DY [6] and the CMS 7 TeV double differential DY [45]. The size of the PI contributions to these are as much as 5% of the QCD DY prediction, depending on the mass and rapidity region, and hence it is interesting to investigate the impact this has on the fit quality, in addition to that due to QED corrections to DGLAP evolution.

Table 1 The values of $\chi^2/N_{\rm pts}$ for the non-LHC data sets included in a NNLO fit to the MSHT20 dataset, with and without QED corrections. The difference in χ^2/Npts is also shown explicitly, for the cases that the magnitude is larger than 1 point; negative values indicate a better fit quality in the QED case

Data set	QCD	QED	Change
BCDMS μp F_2 [63]	178.8/163	182.5/163	(+3.7)
BCDMS μd F_2 [63]	145.6/151	146.8/151	(+1.2)
NMC μp F_2 [64]	124.0/123	124.9/123	–
NMC μd F_2 [64]	112.4/123	113.1/123	–
NMC $\mu n/\mu p$ [65]	130.4/148	128.9/148	(−1.5)
E665 μp F_2 [66]	65.0/53	65.0/53	–
E665 μd F_2 [66]	59.9/53	59.7/53	–
SLAC ep F_2 [67,68]	32.3/37	32.4/37	–
SLAC ed F_2 [67,68]	22.9/38	23.0/38	–
NMC/BCDMS/SLAC/HERA F_L [63,64,68–71]	68.4/57	68.2/57	–
E866/NuSea pp DY [72]	225.8/184	226.0/184	–
E866/NuSea pd/pp DY [73]	9.5/15	8.8/15	–
NuTeV νN F_2 [74]	38.2/53	37.2/53	(−1.0)
CHORUS νN F_2 [75]	30.3/42	29.4/42	–
NuTeV νN xF_3 [74]	30.9/42	30.5/42	
CHORUS νN xF_3 [75]	18.4/28	18.4/28	–
CCFR $\nu N \to \mu\mu X$ [76]	68.1/86	68.4/86	–
NuTeV $\nu N \to \mu\mu X$ [76]	57.5/84	56.7/84	(−1.0)
HERA e^+p CC [77]	50.2/39	50.9/39	–
HERA e^-p CC [77]	70.3/42	72.2/42	(+1.9)
HERA e^+p NC 820 GeV [77]	89.9/75	90.1/75	–
HERA e^+p NC 920 GeV [77]	510.7/402	511.2/402	–
HERA e^-p NC 460 GeV [77]	247.6/209	248.0/209	–
HERA e^-p NC 575 GeV [77]	262.2/259	262.8/259	–
HERA e^-p NC 920 GeV [77]	243.9/159	244.8/159	–
HERA ep $F_2^{\rm charm}$ [78]	132.6/79	131.9/79	–
DØ II $p\bar{p}$ incl. jets [79]	120.3/110	119.5/110	–
CDF II $p\bar{p}$ incl. jets [80]	60.1/76	60.9/76	–
CDF II W asym. [81]	18.9/13	18.5/13	–
DØ II $W \to \nu e$ asym. [82]	33.5/12	33.5/12	–
DØ II $W \to \nu\mu$ asym. [83]	17.8/10	17.6/10	–
DØ II Z rap. [84]	16.3/28	16.4/28	–
CDF II Z rap. [85]	37.1/28	37.2/28	–
DØ W asym. [86]	12.8/14	11.3/14	(−1.5)

We have therefore repeated the fit with the PI components for these datasets excluded, as well as with them included but using LO collinear factorization (with $\mu_F = \mu_R = m_{ll}$), as opposed to the SF approach. For the ATLAS data, we find that including the PI component via LO collinear factorization, or even excluding the PI component entirely, has a very mild impact, at the level of less than 1 point in χ^2. For the CMS data, on the other hand, we find that excluding the PI component leads to a deterioration of the fit quality by $\sim$ 10 points in χ^2. Interestingly, a very similar level of deterioration is seen if we instead include the PI component via LO collinear factorization. Therefore, we can conclude that there is a preference for the SF approach here, although the impact on the final PDF fit will be very small and some caution is needed in this interpretation, given as discussed above other EW corrections are not included in this case. If the PI component were included at NLO in collinear factorization, we would on the other hand expect a closer matching to the SF result, and therefore potentially a similar level of improvement.

4 Impact on PDFs and benchmark cross sections

We now present the impact of the inclusion of QED effects on the MSHT PDFs. Many of the effects mimic those seen

Table 2 The values of $\chi^2/N_{\rm pts}$ for the LHC data sets included in a NNLO fit to the MSHT20 dataset, with and without QED corrections. The difference in χ^2/Npts is also shown explicitly, for the cases that the magnitude is larger than 1 point; negative values indicate a better fit quality in the QED case. The total χ^2 value corresponds to the sum of the individual values shown in Tables 1 and 2

Data set	QCD	QED	Change
ATLAS W^+, W^-, Z [87]	29.9/30	29.7/30	–
CMS W asym. $p_T > 35$ GeV [88]	8.2/11	8.0/11	–
CMS asym. $p_T > 25, 30$ GeV [89]	7.4/24	7.5/24	–
LHCb $Z \to e^+e^-$ [90]	22.3/9	22.6/9	–
LHCb W asym. $p_T > 20$ GeV [91]	12.4/10	12.1/10	–
CMS $Z \to e^+e^-$ [92]	18.0/35	18.0/35	–
ATLAS High-mass Drell–Yan [93]	18.6/13	19.1/13	–
CMS double diff. Drell–Yan [45]	144.8/132	145.4/132	–
Tevatron, ATLAS, CMS $\sigma_{t\bar{t}}$ [94,95]	14.5/17	14.4/17	–
LHCb 2015 W, Z [96,97]	101.4/67	100.6/67	–
LHCb 8 TeV $Z \to ee$ [98]	26.2/17	25.8/17	–
CMS 8 TeV W [99]	12.6/22	14.0/22	(+1.4)
ATLAS 7 TeV jets [49]	221.3/140	217.8/140	(−3.5)
CMS 7 TeV $W + c$ [100]	8.3/10	7.8/10	–
ATLAS 7 TeV high precision W, Z [47]	117.3/61	119.4/61	(+2.1)
CMS 7 TeV jets [50]	176.9/158	176.4/158	–
CMS 8 TeV jets [51]	262.4/174	267.9/174	(+5.5)
CMS 2.76 TeV jet [101]	102.4/81	102.8/81	–
ATLAS 8 TeV Z p_T [54]	190.8/104	200.8/104	(+10.0)
ATLAS 8 TeV single diff $t\bar{t}$ [60]	25.8/25	26.8/25	(+1.0)
ATLAS 8 TeV single diff $t\bar{t}$ dilepton [102]	3.3/5	3.7/5	–
CMS 8 TeV double differential $t\bar{t}$ [103]	22.3/15	22.1/15	–
CMS 8 TeV single differential $t\bar{t}$ [61]	13.0/9	13.3 /9	–
ATLAS 8 TeV High-mass Drell–Yan [6]	56.9/48	57.8/48	–
ATLAS 8 TeV W [104]	57.8/22	59.2/22	(+1.4)
ATLAS 8 TeV W + jets [105]	18.7/30	19.1/30	–
ATLAS 8 TeV double differential Z [48]	75.9/59	77.5/59	(+1.6)
Total	5111.8/4363	5136.1/4363	(+24.3)

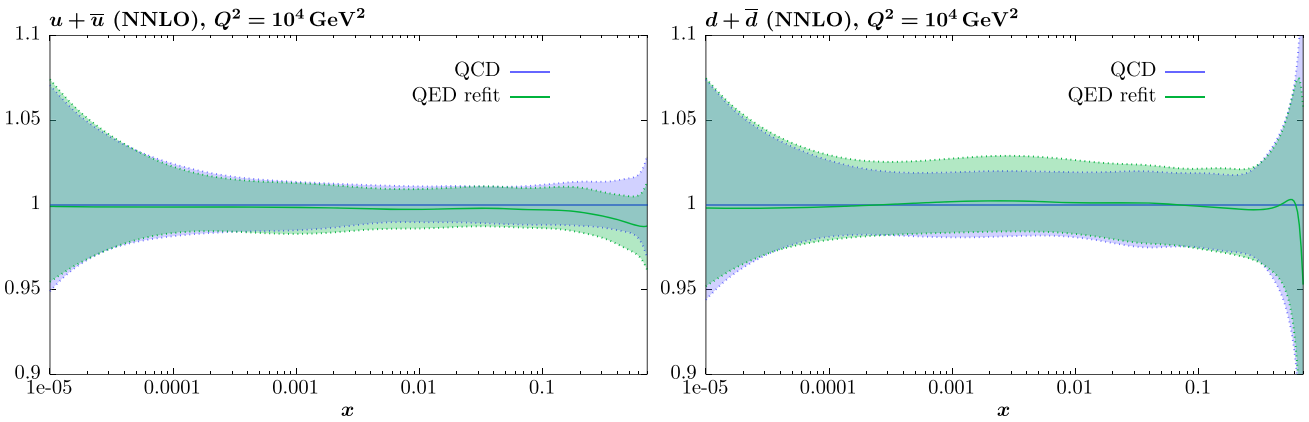


Fig. 2 The ratio of the $u + \overline{u}$ and $d + \overline{d}$ distributions (with uncertainties) at $Q^2 = 10^4$ GeV^2, resulting from NNLO fits to the MSHT20 dataset, with QED effects included to that without

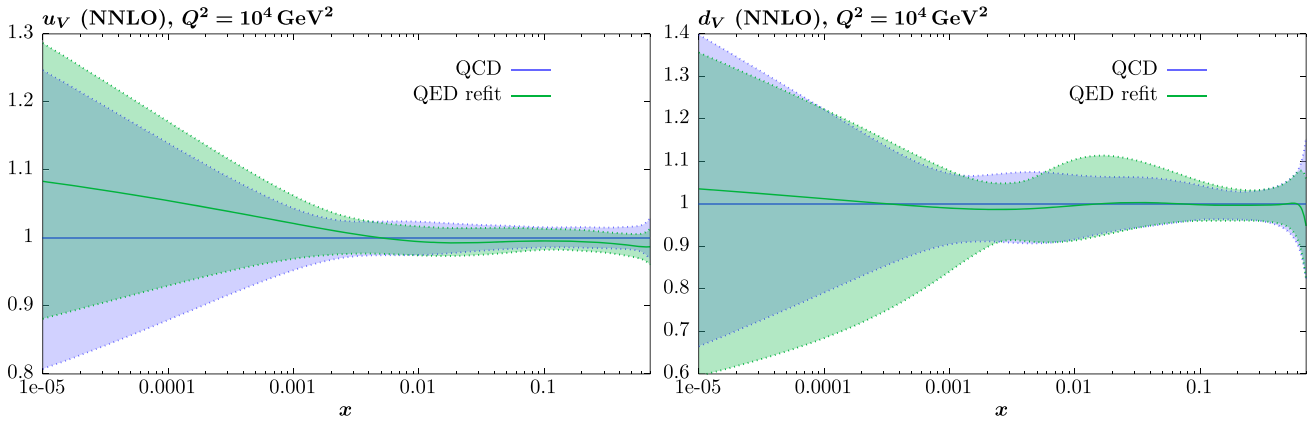


Fig. 3 The ratio of the u_V and d_V distributions (with uncertainties) at $Q^2 = 10^4$ GeV^2, resulting from NNLO fits to the MSHT20 dataset, with QED effects included to that without

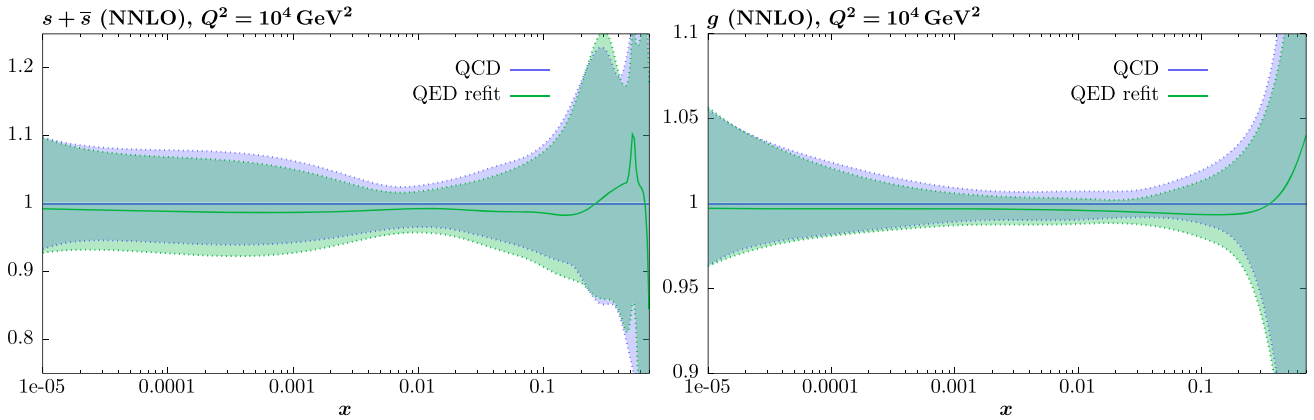


Fig. 4 The ratio of the $s + \bar{s}$ and g distributions (with uncertainties) at $Q^2 = 10^4$ GeV^2, resulting from NNLO fits to the MSHT20 dataset, with QED effects included to that without

in the MMHT15qed set, and therefore we show only a selection here. In all cases, these correspond to the scale $Q^2 = 10^4$ GeV^2.

We begin with the up and down singlet distributions, $u + \bar{u}$ and $d + \bar{d}$, in Fig. 2. At high x these may be expected to show a reduction in the PDFs due to $q \to q + \gamma$ emission. This reduces the quark singlet momenta, and correspondingly increases the photon PDF, with the effect being most pronounced at high x. This can be clearly seen in the up singlet distribution in Fig. 2 (left), although the changes are $O(1\%)$ and well within the uncertainty bands. However, the effect on the down singlet (right) is minimal due to its smaller charge, and is largely removed upon refitting.

The impact on the valence quarks is shown in Fig. 3. The same $q \to q + \gamma$ emission as before plays the dominant role here; this mimics the impact of QCD DGLAP on the valence quarks, due to gluon emission, i.e. both the quarks and antiquarks are shifted to lower x, and hence the valence difference tends to reduce at intermediate to high x. This effect is visible in the up valence, which is reduced at intermediate to high x, with a corresponding increase observed at lower x, due to the valence sum rule (though the size of the impact at the very lowest values of x is to some extent driven by extrapolation). Again, the impact on the down valence is rather milder.

Finally, the effects of the inclusion of QED effects on the total strangeness, $s + \bar{s}$, and on the gluon are illustrated in Fig. 4. The strangeness will be sensitive to the same photon emission effect as the up and down quarks, and indeed some reduction is observed in the high x region, though not at the highest values of x. However, this reduction extends to intermediate and low values of x. This is due to the addition of the photon PDF, which carries a fraction of the proton momentum, and hence requires a reduction in the size of the other PDFs in order to satisfy the momentum sum rule. As the strangeness is rather less well constrained than the up and down singlets, it is more affected by this. A similar reduction is observed for the gluon across a broad region, apart from at the very highest values of x, where some increase is seen, an effect which seems common to other analyses [26,29].

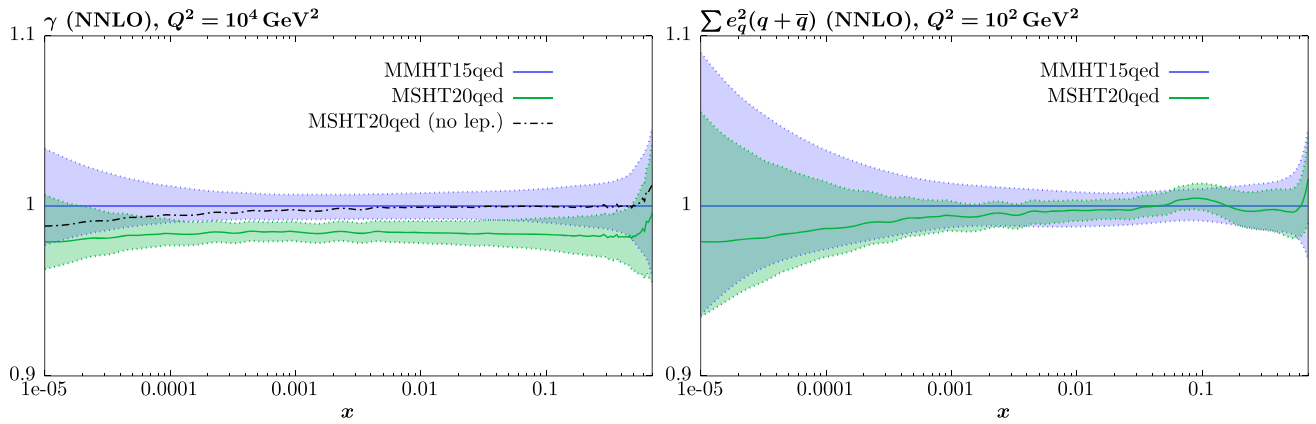


Fig. 5 The ratio of the (left) photon and (right) charge-weighted singlet distributions (with uncertainties) at $Q^2 = 10^4$ GeV^2, resulting from fits to the MSHT20 dataset, to the MMHT15qed case

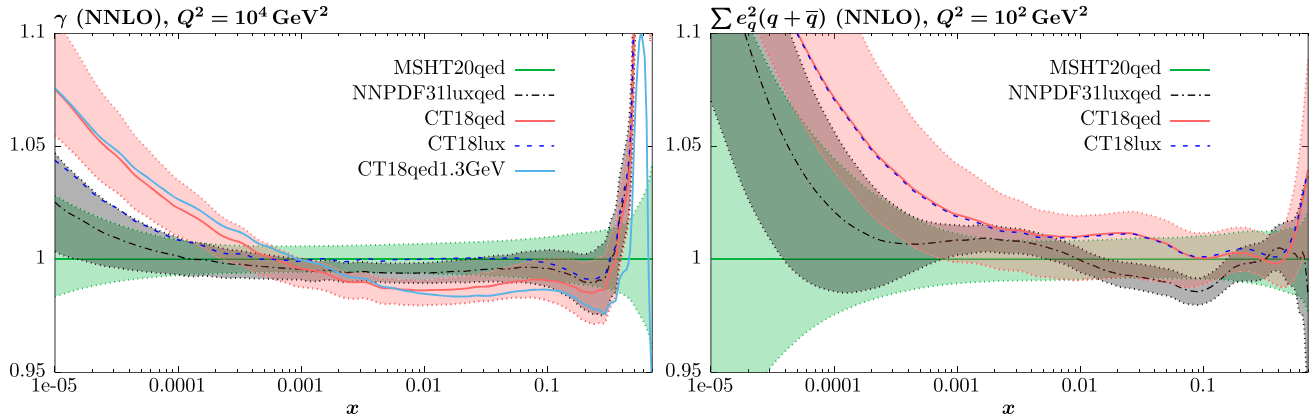


Fig. 6 The ratio of the (left) photon and (right) charge-weighted singlet distributions for various PDF sets at $Q^2 = 10^4$ GeV^2

In Fig. 5 (left) we show the change in the photon PDF with respect to the MMHT15qed case. We can see that, as expected from the discussion in Sect. 2, the photon is now $O(2\%)$ lower across the entire x region. This is almost entirely due to the impact of lepton loops in $P_{\gamma\gamma}$; to demonstrate this we also show the result of performing a fit to the same MSHT20 dataset, but with leptonic loops excluded (i.e. following the procedure of MMHT15qed). We can see that in this case the photon is very similar to the MMHT15qed, apart from at rather lower x, where it is somewhat reduced. In Fig. 5 (right) we show the charge-weighted quark singlet, which is somewhat lower at intermediate to low x, reflecting the difference in the MMHT15 and MSHT20 QCD-only PDFs. This difference will drive the reduction in the photon at low x, due to the reduced contribution from $q \to q+\gamma$ emission. We note that the impact of including lepton loops in $P_{\gamma\gamma}$ on all other partons is very minor, and is for that reason not shown here.

In Fig. 6 we compare the MSHT20qed photon PDF with other results in the literature, namely the NNPDF31luxqed [26], and CT18qed, CT18lux [29] sets. These all apply the same basic LUXqed approach as outlined in [22,23] and used for the MSHT set, but differ in the specifics of the implementation, as well as the underlying QCD partons. In more detail, the CT18qed set applies a similar modification to us, namely applying the LUXqed formula for the photon at input scale Q_0, before evolving with standard QED DGLAP. On the other hand, NNPDF31luxqed and CT18lux apply the LUXqed formula at higher scales, see [26,29] for more details. We can see that for intermediate to reasonably high values of x the agreement between the sets is good, as we might expect. At low x the CT and NNPDF photons lie somewhat above MSHT, which from Fig. 6 (right) we can see is largely driven by the difference in the charge weighted quark singlet PDFs, via their impact on the photon through DGLAP evolution. At the highest values $x \gtrsim 0.5$, on the other hand, the MSHT photon is lower than the other results. In [29] it is argued that the MSHT 'Q_0' approach tends to lead to a lower photon at high x in comparison to the high scale approach, due to the difference in treatment of non-leading twist contributions to $F_2(x, Q^2)$ above Q_0^2, and hence this could explain the difference with respect to

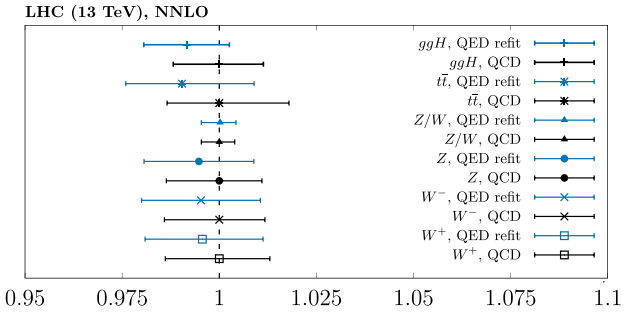


Fig. 7 Benchmark cross sections obtained with NNLO fits to the MSHT20 dataset, with QED effects included to that without. Results are normalized to the central value of the QCD only fit

the NNPDF31luxqed and CT18lux sets. We can see that the CT18qed set, which applies the same basic 'Q_0' methodology as MSHT, remains higher than MSHT (though indeed smaller than CT18lux, even if this is not evident in the plot), but there is better agreement with CT18qed1.3GeV, which has a more similar starting scale to MSHT, whereas the default CT18qed uses $Q_0 = 3$ GeV. We note that in [29] it was observed that the MMHT15qed set was similar in size to CT18qed. However, as discussed above this excluded leptonic loop contributions to $P_{\gamma\gamma}$ (which are included in CT18), which as observed in Fig. 24 of [28] reduce the photon PDF most prominently at high x.

Finally, in Fig. 7 we show results for a range of benchmark cross sections, namely Higgs boson production in gluon fusion, top quark pair, and W, Z production. To calculate the cross section at NNLO in QCD perturbation theory we use the same procedure as described in [33]. That is, we use LO electroweak perturbation theory, with the qqW and qqZ couplings defined by

$$g_W^2 = G_F M_W^2/\sqrt{2}, \quad g_Z^2 = G_F M_Z^2\sqrt{2}, \tag{4}$$

and other electroweak parameters as in [107]. We take the Higgs mass to be $m_H = 125$ GeV and the top pole mass is $m_t = 172.5$ GeV. For the $t\bar{t}$ cross section we use `top++` [108]. We note that in all of these cases the impact of PI production is very small, and hence is not included; that is, all changes result from the impact on the quark and gluon PDFs as a result of the refit including QED effects.

The results are plotted as ratio of the central value for the pure QCD case. For the Higgs and top quark cases we can see that the QED fit results in central cross section values that are $\sim 1\%$ lower. Although these shifts, which are driven by the lower gluon PDF at intermediate to high x values seen in Fig. 4, are only moderate in size, we note that they are smaller than but of a similar order to the PDF uncertainty on the cross section, and hence are certainly significant in impact with respect to this. The W, Z cross section are similarly reduced in the QED case, albeit by a somewhat smaller amount. This is driven by the reduced quark PDFs in the QED fit. For the Z to W ratio on the other hand, which is relatively insensitive to such normalization effects, the impact is seen to be rather small and well within uncertainties. We can in addition see that the PDF uncertainties in the QED and QCD cases are rather comparable, reflecting the similar uncertainties in the PDFs themselves.

5 Breakdown between elastic/inelastic components and neutron PDFs

As in [28] we provide the individual elastic, $\gamma^{\text{el.}}(x, Q^2)$, and inelastic, $\gamma^{\text{inel.}}(x, Q^2)$, photon PDF components for our latest fit, with $\gamma(x, Q^2) = \gamma^{\text{el.}}(x, Q^2) + \gamma^{\text{inel.}}(x, Q^2)$. This separation is achieved in exactly the same way as described in [28], to which we refer the reader for details. It can for example be useful when making predictions for exclusive and semi-exclusive PI production [9,10], although in this case care must be taken to also include the survival factor probability of no additional particle production due to multi-particle interactions (MPI), see e.g. [109]. The fractional contribution from these components to the total at different scales is shown in Fig. 8. We can see that at $Q^2 = 10^4$ GeV^2 the inelastic component is dominant until very high x. At the lower scale of $Q^2 = 10^2$ GeV^2 on the other hand the relative contribution from the elastic component is somewhat larger, due to the shorter evolution length for (inelastic) $q \to q\gamma$ splitting. The breakdown is very close indeed to that in the MMHT15qed set, which is not shown for clarity.

We in addition now provide publicly available neutron PDF sets, again following the approach described in the MMHT15qed study [28]. In particular, QED effects are expected to violate the pure assumption of isospin symmetry, for which $d_V(n) = u_V(p)$ and $u_V(n) = d_V(p)$. These will modify the distributions at the input scale $Q_0 = 1$ GeV, as well as then explicitly in the QED corrected DGLAP evolution to higher scales. The ratio of the neutron down and up valence quarks, at the input scale, to their isospin symmetry partners is shown in Fig. 9. We can see that the effect of isospin violation is small, at the 1% level around the peaks of the valence distributions. These results are comparable to the MMHT15qed case, though interestingly the impact of isospin violation on the $u_V(n)/d_V(p)$ ratio is significantly less at low and high x, which is most likely a result of the rather different proton down valence in the MSHT20 fit with respect to the MMHT14 case, due to the more flexible parameterisation as well as the impact of new data in the fit, see [33] for further discussion. The same comparison, but at $Q^2 = 10^4$ GeV^2, is shown in Fig. 10. Broadly, we can see that at high x the neutron d_V (u_V) is enhanced (suppressed) with respect to the proton u_V (d_V), due to the lower (higher) electric charge of the corresponding neutron PDFs, and hence less (more) significant QED radiation effects, which tend to reduce the

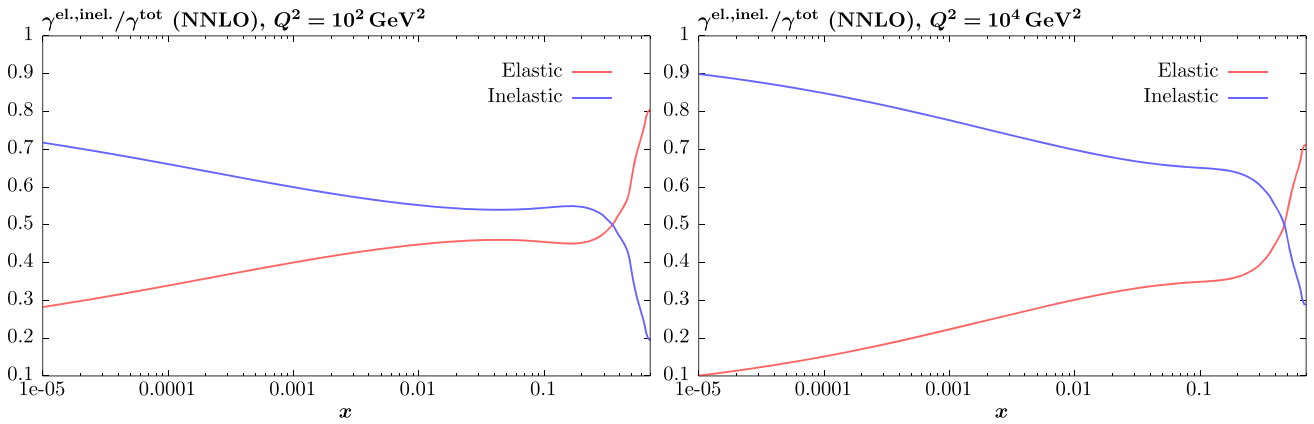


Fig. 8 Ratio of the inelastic and elastic components to the total (given by the sum of both) photon. Results at $Q^2 = 10^2(10^4)$ GeV^2 are shown in the left (right) plots

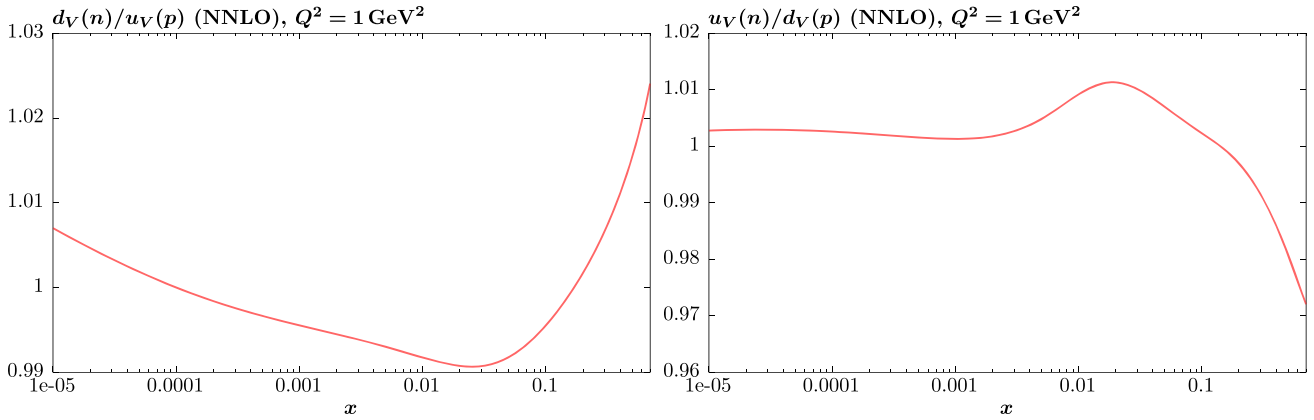


Fig. 9 Ratio of neutron down (up) valence to the proton isospin partner, at $Q^2 = 1$ GeV^2, shown in the left (right) plots

valence distribution in this region. As expected from the form of the prescription for the input neutron PDFs, this trend is already present in the distributions at input, and then is clearly enhanced by the effect of QED DGLAP as we evolve to $Q^2 = 10^4$ GeV^2.

Finally, in Fig. 11 (left) we show the ratio of the photon PDF in the neutron to the proton case. We can see that the neutron's photon PDF is rather lower than that in the proton, due in part to the significantly smaller elastic component of the photon in this case, but also the suppression in the charge-weighted singlet quark PDF at higher x (Fig. 11 (right)), and hence the smaller inelastic photon component that this will generate. At low x the sea quarks dominate and this ratio tends to unity. Hence, at low x the suppression of the photon PDF in the neutron is observed to be less significant, though this is also due to the smaller relative elastic component in the proton case at low x, as seen in Fig. 8.

6 PDF availability

We provide the MSHT20 PDFs in the `LHAPDF` format [110]:
http://lhapdf.hepforge.org/
as well as on the repository:
http://www.hep.ucl.ac.uk/msht/
We present NNLO eigenvector sets of PDFs at the default value of $\alpha_S(M_Z^2) = 0.118$:
`MSHT20qed_nnlo`
but not at NLO, as it is only at NNLO QCD level of precision that QED corrections become relevant. We also provide equivalent PDF sets, but with only the elastic or inelastic components of the photon output (see [28] for discussion of how this is achieved):
`MSHT20qed_nnlo_elastic`
`MSHT20qed_nnlo_inelastic`
Finally, we also present neutron PDF sets, including the

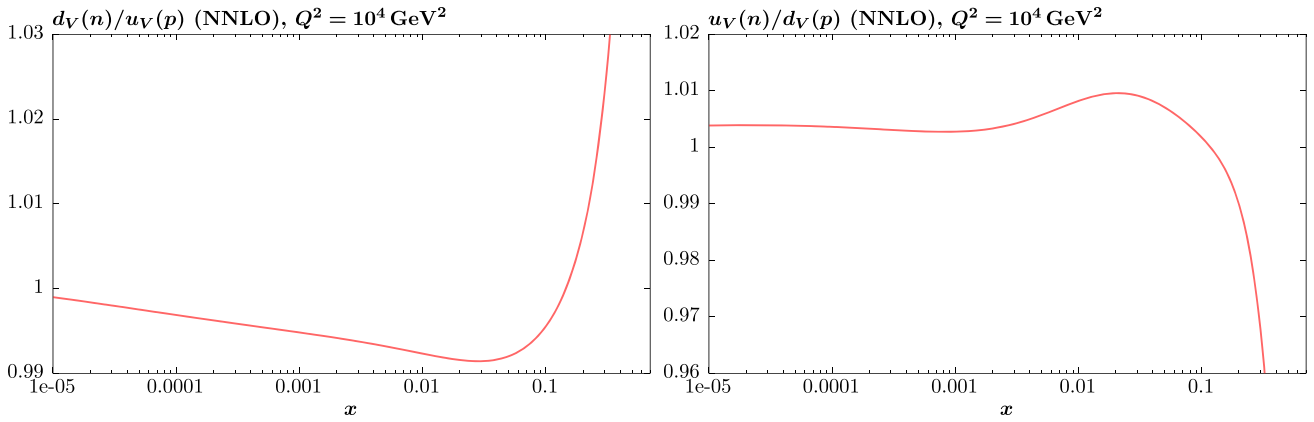


Fig. 10 As in Fig. 9, but at $Q^2 = 10^4$ GeV^2

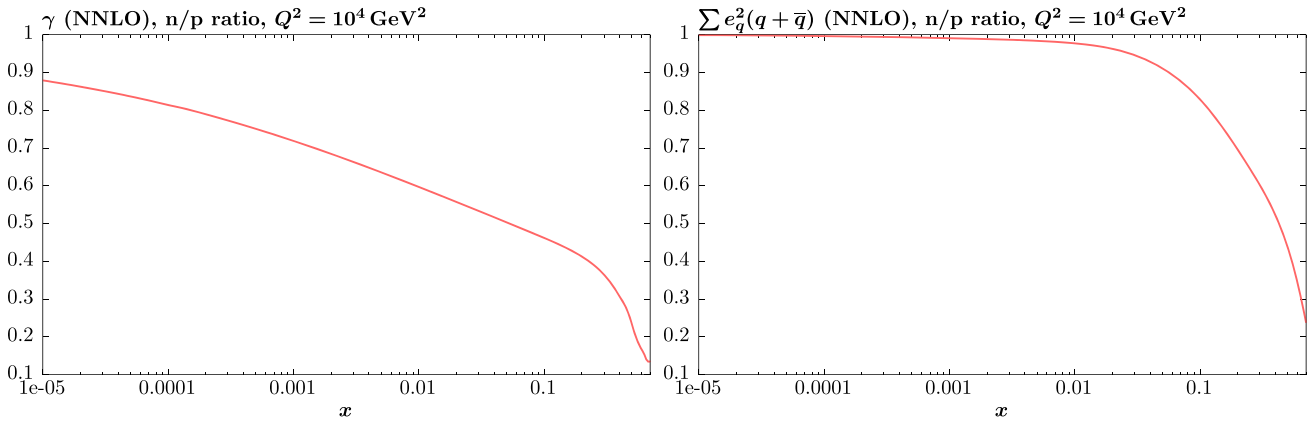


Fig. 11 Ratio of neutron photon (charge-weighted singlet) PDFs to the proton case, at $Q^2 = 10^4$ GeV^2, shown in the left (right) plots

elastic and inelastic photon components:

```
MSHT20qed_nnlo_neutron
MSHT20qed_nnlo_neutron_elastic
MSHT20qed_nnlo_neutron_inelastic
```

In all cases these contain 38 eigenvectors, of which 6 are due to uncertainties in the determination of the photon PDF.

7 Conclusions

In this paper, we have presented the MSHT20qed NNLO PDF set. This closely follows the MSHT20 global analysis [33], but includes QED corrections to the PDF evolution, and a corresponding photon PDF of the proton. The photon PDF is calculated at input following the LUXqed approach, such that percent level PDF uncertainties in the photon PDF are achieved, and a full refit is performed. From this, we have made available a the fully consistent set of QED corrected partons.

We have presented a detailed overview of the expectations for the relevance of photon-initiated (PI) production in processes that enter current global PDF fits. We have seen that in most cases the PI contribution, calculated using a suitable photon PDF set based on the LUXqed approach, such as MSHT20qed, is found to be very small, entering at the per mille level. These can therefore where necessary be safely included via standard NLO EW K-factors, as part of the broader class of EW corrections, provided a photon PDF set consistent with the LUXqed approach is used. In general however, their impact on the PDF fit is expected to be marginal at the current level of precision, and these are significantly suppressed with respect to other dominant EW corrections, and in particular those due to the presence of Sudakov EW enhancements.

On the other hand, the contribution from the PI subprocess to lepton pair production below and above the Z peak at the LHC can be at the level of $\sim 10\%$ of the standard DY contribution. This is therefore rather distinct from the PI corrections to other processes in the fit, and certainly an accurate and precise account of these is mandatory. We have achieved this by making of the structure function (SF) calculation, which provides percent level precision in the underlying cross section for this process.

As in the previous MMHT15qed study [28], we observe some deterioration in the fit quality with respect to a QCD-only fit. However, this effect is relatively mild, and certainly the inclusion of not only PI channels in the fit, where relevant, but also the QED corrections to the PDF evolution of the QCD partons, and their subsequent impact on the PDF sets through refitting, will provide a more accurate result. We have compared our results to a pure QCD-only fit and have found that the impact on the quark and gluon PDFs can be non-negligible, though it is always currently within the PDF uncertainties of the QCD set. It should be noted though that there is in principle no requirement that this should be the case, given that the PDF uncertainties are of a distinct origin to such differences. This is also seen when considering a range of benchmark cross sections: predictions for Higgs boson production via gluon fusion, W, Z production and $t\bar{t}$ production at the 13 TeV LHC change at the level of 1%, which is smaller than, but in some cases comparable to the underlying PDF uncertainty.

We provide a NNLO error set for public use, as well as the breakdown between the photon elastic and inelastic components, and the corresponding neutron PDF set, including QED driven isospin violation. Such QED-corrected PDF sets will play a key role in future LHC precision phenomenology.

Acknowledgements We would like to thank numerous members of the PDF4LHC committee and working group for useful conversations. T. C. and R. S. T. thank the Science and Technology Facilities Council (STFC) for support via grant awards ST/P000274/1 and ST/T000856/1. L. H. L. thanks STFC for support via grant award ST/L000377/1.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: There is no data to be deposited.]

Funded by SCOAP[3].

References

1. G. Altarelli, G. Parisi, Nucl. Phys. B **126**, 298 (1977)
2. Y.L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977)
3. Y.L. Dokshitzer, Zh. Eksp, Teor. Fiz. **73**, 1216 (1977)
4. V.N. Gribov, L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972)
5. V.N. Gribov, L.N. Lipatov, Yad. Fiz. **15**, 781 (1972)
6. ATLAS, G. Aad et al., JHEP **08**, 009 (2016). arXiv:1606.01736
7. A. Bierweiler, T. Kasprzik, J.H. Kühn, S. Uccirati, JHEP **11**, 093 (2012). arXiv:1208.3147
8. A. Denner, S. Dittmaier, S. Kallweit, A. Muck, JHEP **03**, 075 (2012). arXiv:1112.5142
9. L.A. Harland-Lang, V.A. Khoze, M.G. Ryskin, Eur. Phys. J. C **76**, 255 (2016). arXiv:1601.03772
10. L.A. Harland-Lang, M. Tasevsky, V.A. Khoze, M.G. Ryskin, Eur. Phys. J. C **80**, 925 (2020). arXiv:2007.12704
11. A.D. Martin, R.G. Roberts, W.J. Stirling, R.S. Thorne, Eur. Phys. J. C **39**, 155 (2005). arXiv:hep-ph/0411040
12. C. Schmidt, J. Pumplin, D. Stump, C.P. Yuan, Phys. Rev. D **93**, 114015 (2016). arXiv:1509.02905
13. NNPDF, R.D. Ball et al., Nucl. Phys. B **877**, 290 (2013). arXiv:1308.0598
14. xFitter Developers' Team, F. Giuli et al., Eur. Phys. J. C **77**, 400 (2017). arXiv:1701.08553
15. A.D. Martin, M.G. Ryskin, Eur. Phys. J. C **74**, 3040 (2014). arXiv:1406.2118
16. L.A. Harland-Lang, V.A. Khoze, M.G. Ryskin, Phys. Rev. D **94**, 074008 (2016). arXiv:1607.04635
17. V.M. Budnev, I.F. Ginzburg, G.V. Meledin, V.G. Serbo, Phys. Rep. **15**, 181 (1975)
18. H. Anlauf, H.D. Dahmen, P. Manakos, T. Mannel, T. Ohl, Comput. Phys. Commun. **70**, 97 (1992)
19. J. Blumlein, G. Levman, H. Spiesberger, J. Phys. G **19**, 1695 (1993)
20. A. Mukherjee, C. Pisano, Eur. Phys. J. C **30**, 477 (2003). arXiv:hep-ph/0306275
21. M. Łuszczak, W. Schäfer, A. Szczurek, Phys. Rev. D **93**, 074018 (2016). arXiv:1510.00294
22. A.V. Manohar, P. Nason, G.P. Salam, G. Zanderighi, JHEP **12**, 046 (2017). arXiv:1708.01256
23. A. Manohar, P. Nason, G.P. Salam, G. Zanderighi, Phys. Rev. Lett. **117**, 242002 (2016). arXiv:1607.04266
24. D. de Florian, G.F.R. Sborlini, G. Rodrigo, Eur. Phys. J. C **76**, 282 (2016). arXiv:1512.00612
25. D. de Florian, G.F.R. Sborlini, G. Rodrigo, JHEP **10**, 056 (2016). arXiv:1606.02887
26. NNPDF, V. Bertone, S. Carrazza, N.P. Hartland, J. Rojo, SciPost Phys. **5**, 008 (2018). arXiv:1712.07053
27. L.A. Harland-Lang, A.D. Martin, P. Motylinski, R.S. Thorne, Eur. Phys. J. C **76**, 186 (2016). arXiv:1601.03413
28. L.A. Harland-Lang, A.D. Martin, R. Nathvani, R.S. Thorne, Eur. Phys. J. C **79**, 811 (2019). arXiv:1907.02750
29. CTEQ-TEA, K. Xie et al. (2021). arXiv:2106.10299
30. L.A. Harland-Lang, V.A. Khoze, M.G. Ryskin, Phys. Lett. B **761**, 20 (2016). arXiv:1605.04935
31. L.A. Harland-Lang, Phys. Rev. D **104**, 073002 (2021). arXiv:2101.04127
32. S. Carrazza, E.R. Nocera, C. Schwan, M. Zaro, JHEP **12**, 108 (2020). arXiv:2008.12789
33. S. Bailey, T. Cridge, L.A. Harland-Lang, A.D. Martin, R.S. Thorne, Eur. Phys. J. C **81**, 341 (2021). arXiv:2012.04684
34. A1, J.C. Bernauer et al., Phys. Rev. C **90**, 015206 (2014). arXiv:1307.6227
35. HERMES, A. Airapetian et al., JHEP **05**, 126 (2011). arXiv:1103.5704
36. CLAS, M. Osipenko et al., Phys. Rev. D **67**, 092001 (2003). arXiv:hep-ph/0301204
37. V. Bertone, S. Carrazza, D. Pagani, M. Zaro, JHEP **11**, 194 (2015). arXiv:1508.07002
38. L. Buonocore, P. Nason, F. Tramontano, G. Zanderighi, JHEP **08**, 019 (2020). arXiv:2005.06477
39. D. Pagani, I. Tsinikos, M. Zaro, Eur. Phys. J. C **76**, 479 (2016). arXiv:1606.01915
40. R.D. Ball et al. (2021). arXiv:2109.02653

41. J.M. Campbell, J.W. Huston, W.J. Stirling, Rep. Prog. Phys. **70**, 89 (2007). arXiv:hep-ph/0611148
42. L.A. Harland-Lang, JHEP **03**, 128 (2020). arXiv:1910.10178
43. A.B. Arbuzov, R.R. Sadykov, J. Exp. Theor. Phys. **106**, 488 (2008). arXiv:0707.0423
44. S. Dittmaier, M. Huber, JHEP **01**, 060 (2010). arXiv:0911.2329
45. CMS, S. Chatrchyan et al., JHEP **12**, 030 (2013). arXiv:1310.7291
46. CMS, V. Khachatryan et al., Eur. Phys. J. C **75**, 147 (2015). arXiv:1412.1115
47. ATLAS, M. Aaboud et al., Eur. Phys. J. C **77**, 367 (2017). arXiv:1612.03016
48. ATLAS, M. Aaboud et al., JHEP **12**, 059 (2017). arXiv:1710.05167
49. ATLAS, G. Aad et al., JHEP **02**, 153 (2015). arXiv:1410.8857 [Erratum: JHEP **09**, 141 (2015)]
50. CMS, S. Chatrchyan et al., Phys. Rev. D **90**, 072006 (2014). arXiv:1406.0324
51. CMS, V. Khachatryan et al., JHEP **03**, 156 (2017). arXiv:1609.05331
52. S. Dittmaier, A. Huss, C. Speckner, JHEP **11**, 095 (2012). arXiv:1210.0438
53. R. Frederix et al., JHEP **04**, 076 (2017). arXiv:1612.06548
54. ATLAS, G. Aad et al., Eur. Phys. J. C **76**, 291 (2016). arXiv:1512.02192
55. A. Denner, S. Dittmaier, T. Kasprzik, A. Muck, JHEP **06**, 069 (2011). arXiv:1103.0914
56. J.M. Lindert et al., Eur. Phys. J. C **77**, 829 (2017). arXiv:1705.04664
57. S.G. Bondarenko, A.A. Sapronov, Comput. Phys. Commun. **184**, 2343 (2013). arXiv:1301.3687
58. A. Arbuzov et al., JETP Lett. **103**, 131 (2016). arXiv:1509.03052
59. Y. Li, F. Petriello, Phys. Rev. D **86**, 094034 (2012). arXiv:1208.5967
60. ATLAS, G. Aad et al., Eur. Phys. J. C **76**, 538 (2016). arXiv:1511.04716
61. CMS, V. Khachatryan et al., Eur. Phys. J. C **75**, 542 (2015). arXiv:1505.04480
62. M. Czakon et al., JHEP **10**, 186 (2017). arXiv:1705.04105
63. BCDMS, A.C. Benvenuti et al., Phys. Lett. B **223**, 485 (1989)
64. NMC, M. Arneodo et al., Nucl. Phys. B **483**, 3 (1997). arXiv:hep-ph/9610231
65. NMC, M. Arneodo et al., Nucl. Phys. B **487**, 3 (1997). arXiv:hep-ex/9611022
66. E665, M.R. Adams et al., Phys. Rev. D **54**, 3006 (1996)
67. L.W. Whitlow, E.M. Riordan, S. Dasu, S. Rock, A. Bodek, Phys. Lett. B **282**, 475 (1992)
68. L.W. Whitlow, S. Rock, A. Bodek, E.M. Riordan, S. Dasu, Phys. Lett. B **250**, 193 (1990)
69. H1, F.D. Aaron et al., Phys. Lett. B **665**, 139 (2008). arXiv:0805.2809
70. H1, F.D. Aaron et al., Eur. Phys. J. C **71**, 1579 (2011). arXiv:1012.4355
71. ZEUS, S. Chekanov et al., Phys. Lett. B **682**, 8 (2009). arXiv:0904.1092
72. J.C. Webb, Measurement of continuum dimuon production in 800-GeV/C proton nucleon collisions. PhD thesis, New Mexico State University (2003). arXiv:hep-ex/0301031
73. NuSea, R.S. Towell et al., Phys. Rev. D **64**, 052002 (2001). arXiv:hep-ex/0103030
74. NuTeV, M. Tzanov et al., Phys. Rev. D **74**, 012008 (2006). arXiv:hep-ex/0509010
75. CHORUS, G. Onengut et al., Phys. Lett. B **632**, 65 (2006)
76. NuTeV, M. Goncharov et al., Phys. Rev. D **64**, 112006 (2001). arXiv:hep-ex/0102049
77. H1, ZEUS, F. Aaron et al., JHEP **01**, 109 (2010). arXiv:0911.0884
78. H1, ZEUS, H. Abramowicz et al., Eur. Phys. J. C **78**, 473 (2018). arXiv:1804.01019
79. D0, V.M. Abazov et al., Phys. Rev. D **85**, 052006 (2012). arXiv:1110.3771
80. CDF, A. Abulencia et al., Phys. Rev. D **75**, 092006 (2007). arXiv:hep-ex/0701051 [Erratum: Phys. Rev. D **75**, 119901 (2007)]
81. CDF, T. Aaltonen et al., Phys. Rev. Lett. **102**, 181801 (2009). arXiv:0901.2169
82. D0, V.M. Abazov et al., Phys. Rev. Lett. **101**, 211801 (2008). arXiv:0807.3367
83. D0, V.M. Abazov et al., Phys. Rev. D **88**, 091102 (2013). arXiv:1309.2591
84. D0, V.M. Abazov et al., Phys. Rev. D **76**, 012003 (2007). arXiv:hep-ex/0702025
85. CDF, T.A. Aaltonen et al., Phys. Lett. B **692**, 232 (2010). arXiv:0908.3914
86. D0, V.M. Abazov et al., Phys. Rev. Lett. **112**, 151803 (2014). arXiv:1312.2895 [Erratum: Phys. Rev. Lett. **114**, 049901 (2015)]
87. ATLAS, G. Aad et al., Phys. Rev. D **85**, 072004 (2012). arXiv:1109.5141
88. CMS, S. Chatrchyan et al., Phys. Rev. Lett. **109**, 111806 (2012). arXiv:1206.2598
89. CMS, S. Chatrchyan et al., JHEP **04**, 050 (2011). arXiv:1103.3470
90. LHCb, R. Aaij et al., JHEP **02**, 106 (2013). arXiv:1212.4620
91. LHCb, R. Aaij et al., JHEP **06**, 058 (2012). arXiv:1204.1620
92. CMS, S. Chatrchyan et al., Phys. Rev. D **85**, 032002 (2012). arXiv:1110.4973
93. ATLAS, G. Aad et al., Phys. Lett. B **725**, 223 (2013). arXiv:1305.4192
94. CDF, D0, T.A. Aaltonen et al., Phys. Rev. D **89**, 072001 (2014). arXiv:1309.7570
95. CMS, S. Chatrchyan et al., JHEP **02**, 024 (2014). arXiv:1312.7582 [Erratum: JHEP **02**, 102 (2014)]
96. LHCb, R. Aaij et al., JHEP **08**, 039 (2015). arXiv:1505.07024
97. LHCb, R. Aaij et al., JHEP **01**, 155 (2016). arXiv:1511.08039
98. LHCb, R. Aaij et al., JHEP **05**, 109 (2015). arXiv:1503.00963
99. CMS, V. Khachatryan et al., Eur. Phys. J. C **76**, 469 (2016). arXiv:1603.01803
100. CMS, S. Chatrchyan et al., JHEP **02**, 013 (2014). arXiv:1310.1138
101. CMS, V. Khachatryan et al., Eur. Phys. J. C **76**, 265 (2016). arXiv:1512.06212
102. ATLAS, M. Aaboud et al., Phys. Rev. D **94**, 092003 (2016). arXiv:1607.07281
103. CMS, A.M. Sirunyan et al., Eur. Phys. J. C **77**, 459 (2017). arXiv:1703.01630
104. ATLAS, G. Aad et al., Eur. Phys. J. C **79**, 760 (2019). arXiv:1904.05631
105. ATLAS, M. Aaboud et al., JHEP **05**, 077 (2018). arXiv:1711.03296
106. T. Cridge, L.A. Harland-Lang, A.D. Martin, R.S. Thorne, An investigation of the α_S and heavy quark mass dependence in the MSHT20 global PDF analysis. Eur. Phys. J. C **81**(8), 744 (2021). https://doi.org/10.1140/epjc/s10052-021-09533-7
107. A.D. Martin, W. Stirling, R.S. Thorne, G. Watt, Eur. Phys. J. C **63**, 189 (2009). arXiv:0901.0002
108. M. Czakon, P. Fiedler, A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013). arXiv:1303.6254
109. V.A. Khoze, A.D. Martin, M.G. Ryskin, Eur. Phys. J. C **73**, 2503 (2013). arXiv:1306.2149
110. A. Buckley et al., Eur. Phys. J. C **75**, 132 (2015). arXiv:1412.7420