Determination of the parton distribution functions of the proton using diverse ATLAS data from pp collisions at $\sqrt{s} = 7, 8$ and $13\text{ TeV}$

ATLAS Collaboration

DOI
10.1140/epjc/s10052-022-10217-z

Publication date
2022

Document Version
Final published version

Published in
European Physical Journal C

License
CC BY

Citation for published version (APA):
ATLAS Collaboration (2022). Determination of the parton distribution functions of the proton using diverse ATLAS data from pp collisions at $\sqrt{s} = 7, 8$ and $13\text{ TeV}$. European Physical Journal C, 82(5), [438]. https://doi.org/10.1140/epjc/s10052-022-10217-z
Determination of the parton distribution functions of the proton using diverse ATLAS data from $pp$ collisions at $\sqrt{s} = 7, 8$ and 13 TeV

ATLAS Collaboration

CERN, 1211 Geneva 23, Switzerland

Received: 22 December 2021 / Accepted: 12 March 2022 / Published online: 13 May 2022
© CERN for the benefit of the ATLAS collaboration 2022

Abstract This paper presents an analysis at next-to-next-to-leading order in the theory of quantum chromodynamics for the determination of a new set of proton parton distribution functions using diverse measurements in $pp$ collisions at $\sqrt{s} = 7, 8$ and 13 TeV, performed by the ATLAS experiment at the Large Hadron Collider, together with deep inelastic scattering data from $ep$ collisions at the HERA collider. The ATLAS data sets considered are differential cross-section measurements of inclusive $W^\pm$ and $Z/\gamma^*$ boson production, $W^\pm$ and $Z$ boson production in association with jets, $t\bar{t}$ production, inclusive jet production and direct photon production. In the analysis, particular attention is paid to the correlation of systematic uncertainties within and between the various ATLAS data sets and to the impact of model, theoretical and parameterisation uncertainties. The resulting set of parton distribution functions is called ATLASpdf21.

Contents

1 Introduction .................................. 1
2 Input data sets ............................... 2
  2.1 Description of data sets ................. 2
  2.2 Correlations of uncertainties within and between data sets ....... 5
3 Theoretical framework ..................... 8
  3.1 NNLO QCD and NLO electroweak predictions for each data set .... 8
  3.2 Scale uncertainties and sensitivity to NNLO code 10
4 Fit methodology ........................... 11
5 Results ..................................... 12
  5.1 Comparison of PDFs with and without inclusion of correlations between data sets ... 12
  5.2 Impact of each data set ................. 16
  5.3 Model, theoretical and parameterisation uncertainties ....... 21
6 Consideration of $\chi^2$ tolerance and comparison with global fits .... 28
  6.1 $\chi^2$ tolerance ....................... 28
  6.2 Comparison of ATLASpdf21 with global PDFs 30
7 Conclusion .................................. 38
Appendix ..................................... 39
A Scale uncertainties in $W$, $Z$ inclusive data .......... 39
B Comparison of the impact of inclusive jet data at different centre-of-mass energies ........ 41
C Goodness of the fit: comparison with data sets included in the fit .... 41
D Comparison with extra data sets not included in the fit 51
References .................................. 54

1 Introduction

Precise knowledge of the content of protons, the parton distribution functions (PDFs), is a necessary ingredient for accurate predictions of both Standard Model (SM) and beyond-the-SM (BSM) cross sections at the Large Hadron Collider (LHC). Searching for BSM effects in the deviations of SM parameters from their predicted SM values [1,2] requires better knowledge of PDFs in the kinematic regions where they are already best known [3,4]. This means that many effects which were previously considered negligible must now be examined seriously. In this paper, particular attention is paid to some of these effects, including theoretical scale uncertainties and the correlations of experimental systematic uncertainties between, as well as within, data sets.

With present theoretical knowledge, PDFs can be determined at up to next-to-next-leading order (NNLO) in perturbative quantum chromodynamics (QCD). In order to determine the PDFs to the required precision, a wide coverage of the scale, $Q^2$, and $x$, the fraction of the proton’s longitudinal momentum carried by the parton participating in the initial interaction, is required. This is facilitated by combining new inputs with older data from various experiments and
by measurement of different processes to constrain different regions of phase space, which span the kinematic region: $10^{-5} \lesssim x \lesssim 1$ and $1 \lesssim Q^2 \lesssim 10^6$ GeV$^2$.

Knowledge of the PDFs of the proton, in the kinematic region relevant for the LHC, comes mainly from the precision deep inelastic scattering (DIS) data from $e\bar{p}$ collisions at the HERA collider, which covers a broad range of $Q^2$ and $x$. The PDF set HERAPDF2.0 was determined from HERA data alone [5] using information from $e^+\bar{p}$ neutral-current (NC) and charged-current (CC) processes. The lower-$Q^2$ NC data constrain the low-$x$ sea-quark distribution but are not able to distinguish between quark flavours in the sea at low $x$, or between the down-type quarks, $\bar{d}$ and $\bar{s}$ at any $x$. The difference between the NC $e^+\bar{p}$ and $e^-\bar{p}$ cross sections at high $Q^2$, together with the high-$Q^2$ CC data, constrains the valence distributions. The $Q^2$ dependence measured in the data constrains the gluon distribution. Additional information about the gluon comes from using HERA cross sections measured at various centre-of-mass energies ($\sqrt{s}$) through the contribution of the longitudinal structure function [6–8].

Diverse ATLAS data sets can be used in addition to the HERA data to constrain the PDFs better. In Ref. [9], high precision measurements of the inclusive differential $W^\pm$ and $Z/\gamma^*$ boson cross sections at 7 TeV were added to the HERA data, resulting in the PDF set ATLASepWZ16, which improved on the HERAPDF2.0 set in various respects. Firstly, the strange content of the sea was determined rather than assumed to be a fixed fraction of the light sea. Indeed, compared to previous determinations, the strange sea was found to be enhanced at low $x$. Secondly, the accuracy of the valence quark distributions for $x < 0.1$ was improved.

The effect of adding differential $t\bar{t}$ distributions, in the lepton+jets and dilepton channels at 8 TeV, to the HERA data and the inclusive $W, Z/\gamma^*$ data was studied in Ref. [10]. These $t\bar{t}$ data are complementary to the $W, Z/\gamma^*$ data in their power to constrain PDFs because they are sensitive to the high-$x$ gluon distribution. The resulting PDF set was called ATLASepWZTop18. In Ref. [11], data on the production of $W$ and $Z$ bosons in association with jets ($V+$jets) were added to the HERA data and inclusive $W, Z/\gamma^*$ data, resulting in the ATLASepWZVjets20 PDF set. The $V+$jets data are sensitive to partons at higher $x$ than can be accessed by inclusive $W$ and $Z/\gamma^*$ data and, in particular, they constrain the $\bar{d}$ and $\bar{s}$ quarks at higher $x$.

It would clearly be advantageous to combine ATLAS $W, Z/\gamma^*$ data, $t\bar{t}$ data and $V+$jets data in a single QCD fit. This is done in the present paper, and additional ATLAS data sets are also included. This paper presents the first comprehensive and comparative NNLO perturbative QCD analysis of a number of ATLAS data sets with potential sensitivity to parton distributions. The additional ATLAS data sets are described in the following. Firstly, the data on $W$ [12] production and $Z/\gamma^*$ [13] production with 20.2 fb$^{-1}$ at 8 TeV are added, providing further constraints on the valence PDFs and on the composition of the light-quark sea. Secondly, the direct photon production differential cross sections with 20.2 fb$^{-1}$ at 8 TeV and 3.2 fb$^{-1}$ at 13 TeV are added in the form of their ratios [14], whereby many systematic uncertainties cancel out. Thirdly, the $t\bar{t}$ differential cross sections in the lepton+jets channel from 3.2 fb$^{-1}$ at 13 TeV [15] are added and, fourthly, inclusive jet production cross sections with 4.5 fb$^{-1}$ at 7 TeV [16], 3.2 fb$^{-1}$ at 8 TeV [17] and 3.2 fb$^{-1}$ at 13 TeV [18] are considered. These direct photon data, $t\bar{t}$ data and inclusive jet data all have impact on the gluon PDF at medium to high $x$.

All ATLAS data sets considered in this study have full information about correlated systematic uncertainties. The analysis considers systematic uncertainty correlations between, as well as within, data sets. This is important now that several of the input data sets have systematic uncertainties deriving from jet measurements, since these uncertainties are larger than those from the lepton measurements. Theoretical uncertainties, such as scale uncertainties, are also considered, including their correlations between data sets. The resultant PDF set is called ATLASpdf21.

The structure of the paper is as follows. Section 2 presents the input data sets. Section 3 describes the theoretical framework of the fit. Section 4 presents the fitting methodology, including the definition of the fit $\chi^2$ and details of the parameterisation. Section 5 presents ATLASpdf21 PDFs and then compares them with a fit in which correlations between data sets are not implemented. The impact of each data set is considered and then model uncertainties and further theoretical uncertainties, including scale uncertainties, are considered. Section 6 presents a study of the $\chi^2$ tolerance and a comparison with other modern PDF sets. Section 7 gives the summary and conclusions. Appendix A gives a more detailed exposition of the impact of scale uncertainties for the inclusive $W, Z/\gamma^*$ data. The impacts of inclusive jet data at different centre-of-mass energies are compared in Appendix B. Appendices C and D compare the ATLASpdf21 fit predictions with the ATLAS input data sets and with data sets from other experiments, respectively.

## 2 Input data sets

### 2.1 Description of data sets

The combined $e^\pm p$ NC cross-section measurements of H1 and ZEUS [5] cover a kinematic range of $Q^2$, defined as the negative four-momentum transfer squared, from 0.045 GeV$^2$ to 50,000 GeV$^2$ and of Bjorken-$x$, $x_{Bj}$, which is equal to the fraction of the proton’s longitudinal momentum carried by the struck parton at leading order (LO) in QCD, from $6 \times 10^{-7}$ to 0.65. The CC cross-section measurements cover
a kinematic region $Q^2 \sim 300$ GeV$^2$ to beyond $10^4$ GeV$^2$ and of $x$ from $\sim 0.65$ down to $\sim 10^{-2}$. Low- $x$ data, below $x_B^W = 10^{-5}$, are excluded from this analysis by requiring $Q^2 > 10$ GeV$^2$, motivated by the poorer fit observed in this region compared to the rest of HERA data [5], which may reflect the need for resummation corrections at low $x$ [19,20]. Full information about correlated systematic uncertainties is provided. There are 169 sources of correlated systematic uncertainty. Total uncertainties are below 1.5% over the $Q^2$ range of $3 < Q^2 < 500$ GeV$^2$ and remain below 3% up to $Q^2 = 3000$ GeV$^2$.

The ATLAS $W$, $Z/\gamma^*$ differential cross-section measurements at $\sqrt{s} = 7$ TeV based on an integrated luminosity of 4.6 fb$^{-1}$ are used [9]. A combination of the electron and muon decay channels is used. They access a kinematic range whereby the scale is identified with the measured boson mass range, and the $x$ range is determined by the scale, the proton beam energy and the measured rapidity ($y$) ranges, such that $Q^2 = m^2$ and $x = (Q/\sqrt{s}) e^{y}$. This gives an $x$ range of $0.001 < x < 0.1$. The $W/\gamma^*$ differential cross sections were measured as a function of the $W$ decay lepton pseudorapidity, $\eta_l$, with an experimental precision of 0.6%–1.0%. The $Z/\gamma^*$ boson rapidity distribution, $y_Z$, was measured in three mass ranges: 46 < $m_{\ell\ell}$ < 66 GeV; 66 < $m_{\ell\ell}$ < 116 GeV and 116 < $m_{\ell\ell}$ < 136 GeV, with the two decay leptons in central-central (c-c) and central-forward (c-f) rapidity ranges, and experimental uncertainties as low as 0.4% for c-c rapidity and 2.3% for c-f rapidity. There are 131 sources of correlated systematic uncertainty. The low-mass off-peak $Z/\gamma^*$ data are not used in the present analysis because they are subject to much larger corrections for non-fixed order, parton shower $p_T$-resummation effects than the other $W$, $Z/\gamma^*$ data [21]. These low-mass data had little impact on the ATLASepWZ16 PDFs.

ATLAS has published differential cross-sections and forward-backward asymmetries for $Z/\gamma^*$ production at $\sqrt{s} = 8$ TeV, based on 20.2 fb$^{-1}$ of data [13]. The measurement is presented in terms of three variables for both the c-c and c-f regions of the dilepton rapidity, $y_{\ell\ell}$. For the c-c region, there are 12 bins of absolute rapidity ranging from 0.0 to 2.4, in intervals of 0.2, divided into 7 bins of the dilepton mass $m_{\ell\ell}$ ranging from 46 to 200 GeV and 6 bins of the polar angle, $\theta$, of the decay lepton in the Collins–Soper frame [22]. For the c-f region, there are 5 bins of absolute rapidity ranging from 1.2 to 3.6 and 5 mass bins ranging from 66 to 150 GeV in the same 6 bins of the Collins–Soper angle. The dilepton data were extracted in both the dielectron and dimuon channels up to $|y_{\ell\ell}| = 2.4$, which is extended to $|y_{\ell\ell}| = 3.6$ in the dielectron channel. The dimuon and dilepton data are combined for the c-c region. Measurements are used as cross sections within the acceptance of the lepton fiducial cuts with respect to the $(m_{\ell\ell}, y_{\ell\ell}, \cos \theta^*)$ bin is greater than 95%, such that the leptons are constructed with high accuracy and the cross sections are known to $\sim0.5%$ accuracy. There are 184 cross-section data points which fulfil these requirements and all of them lie in the c-c region of dilepton rapidity. These are used for the central fit. The selection of the bins in $m_{\ell\ell}, y_{\ell\ell}$ and $\theta$ ensures that the NNLO predictions for these high acceptance bins are not highly sensitive to fiducial cuts applied to leptons [23]. The uncertainties of the high-$|y_{\ell\ell}|$ data range from 2% to 33% and the combined central rapidity data reaches a precision ranging from 0.4%, for the $Z$-mass-peak bins at $|y_{\ell\ell}| < 1.0$, to 1.8% for the high-mass bins at $|y_{\ell\ell}| = 2.4$. There are 330 sources of correlated systematic uncertainty. The remaining lower-acceptance data are used only in cross-checks. They are converted to 188 data points for forward-backward asymmetries across the Collins–Soper angle in order to minimise sensitivity to acceptance cuts. These data lie in both the c-c and c-f dilepton rapidity regions.

The ATLAS inclusive differential cross-sections, from 20.2 fb$^{-1}$ of data at $\sqrt{s} = 8$ TeV [12], are measured as a function of muon pseudorapidity, $\eta_\mu$. Data are provided for $W^+$ and $W^-$ cross sections separately as well as for the $W$-asymmetry. The separate $W^+$ and $W^-$ cross sections are used for the fit, rather than the $W$ asymmetry. The precision of these cross-section measurements varies from 0.8% to 1.5% as a function of $\eta_\mu$. The pseudorapidity covers the same central range as the $\sqrt{s} = 7$ TeV inclusive $W$ data. There are 41 sources of correlated systematic uncertainty.

The ATLAS differential cross-section measurements of $W$ production in association with jets ($W$+jets) are based on data recorded during $pp$ collisions at $\sqrt{s} = 8$ TeV, and a total integrated luminosity of 20.2 fb$^{-1}$, in the electron decay channel only. Each event contains at least one jet with transverse momentum $p_T > 30$ TeV and rapidity $|y| < 4.4$. The data are split into $W^+$ and $W^-$ cross sections. There are 51 sources of correlated systematic uncertainty. As explained in Ref. [11] the differential spectra of the transverse momenta of the $W^+$ and $W^-$ bosons $p_T^W$ are fitted in the range $25 < p_T^W < 400$ GeV. The data precision ranges from 10 to 25% from low to high $p_T^W$ [24].

The ATLAS differential cross-section measurements of $Z$ production in association with jets [25] are based on data recorded during $pp$ collisions at $\sqrt{s} = 8$ TeV, and a total integrated luminosity of 20.2 fb$^{-1}$, in the electron decay channel only. The measurement was performed as a function of the
absolute rapidity of inclusive jets, $|y^{\text{jet}}|$ for several bins of the transverse momentum within $25 \text{ GeV} < p_T^{\text{jet}} < 1050 \text{ GeV}$. There are 42 sources of correlated systematic uncertainty. The data precision ranges from $\sim 10\%$ to $\sim 30\%$ from low $|y^{\text{jet}}|$ and $p_T^{\text{jet}}$ to high $|y^{\text{jet}}|$ and $p_T^{\text{jet}}$. The $t\bar{t}$ differential cross sections were measured at $\sqrt{s} = 8 \text{ TeV}$ using 20.2 fb$^{-1}$ of data in the lepton+jets [26] and dilepton [27] decay channels. The $t\bar{t}$ cross sections are provided as both normalised and absolute spectra. The absolute spectra are used in the present study since they carry extra information about the normalisation of the cross-sections, which is not input anywhere else in the fit. As detailed in Ref. [10], in the lepton+jets channel the differential spectra that are used are the mass of the $t\bar{t}$ pair, $m_{t\bar{t}}$, and the average top-quark transverse momentum, $p_T^{t\bar{t}}$. Each of these spectra has full information about systematic and statistical bin-to-bin correlations, including statistical correlations between the spectra. There are 55 sources of correlated systematic uncertainty in common between the different spectra. The precision of these lepton+jets data is $\sim 15\%$ for $p_T^{t\bar{t}}$ and ranges from $\sim 10\%$ to $\sim 20\%$ from low to high $m_{t\bar{t}}$. In the dilepton channel, the spectrum for the rapidity of the $t\bar{t}$ pair, $y_{t\bar{t}}$, is used. The precision of these data ranges from $\sim 5\%$ to $\sim 15\%$ from low to high $|y_{t\bar{t}}|$. For these dilepton channel data the correlations are provided as a total covariance matrix. Further studies of these top-quark data are presented in Ref. [10].

The $t\bar{t}$ differential cross sections were measured at $\sqrt{s} = 13 \text{ TeV}$ using 36 fb$^{-1}$ of data in the lepton+jets [15] channel. Similarly to the $t\bar{t}$ data at 8 TeV, the absolute spectra are used. For the $t\bar{t}$ data at 13 TeV, two complementary topologies are measured; the ‘resolved’ and ‘boosted’ topologies, which differ in that the decay products of the hadronically decaying top quark are well separated for the resolved case and collimated for the boosted case. The ‘boosted’ topology reaches higher absolute rapidity of the $t\bar{t}$ pair. The differential spectra used for the PDF fit are the mass of the $t\bar{t}$ pair, $m_{t\bar{t}}$, the average $t\bar{t}$-transverse momentum, $p_T^{t\bar{t}}$, the average top-quark rapidity, $y_t$, and the boosted rapidity of the $t\bar{t}$ pair, $y^b_{t\bar{t}}$. These spectra were found to be the most sensitive to the gluon PDF in Ref. [28]. Each of these spectra has full information about systematic and statistical bin-to-bin correlations, including statistical correlations between the spectra.

There are 88 sources of correlated systematic uncertainty in common between the different spectra and statistical correlations between the spectra are implemented. Double differential spectra in $p_T^{t\bar{t}}$ and $m_{t\bar{t}}$ were also studied. The precision of these lepton+jets data ranges from $15\%$ to $30\%$ for $p_T^{t\bar{t}}$ and from $15\%$ to $22\%$ for $m_{t\bar{t}}$, with the largest uncertainties in the lowest and highest bins. For both of the rapidity spectra the total uncertainty is $\sim 15\%$.

The ratio of the cross sections for inclusive isolated-photon production at $\sqrt{s} = 8$ and 13 TeV comes from integrated luminosities of 20.2 fb$^{-1}$ and 3.2 fb$^{-1}$ respectively. The data are presented as a function of photon transverse energy, $E_T^\gamma$, for $E_T^\gamma > 125 \text{ GeV}$ in bins of photon pseudorapidity, $\eta^\gamma$, for the central region $|\eta^\gamma| < 2.37$. Details of the isolation criteria are given in Ref. [14]. Evaluation of experimental uncertainties in the ratio data takes into account the correlations between the data at the two different centre-of-mass energies. The dominant source of correlated uncertainty is the photon energy scale. The uncertainty due to this source is much lower when considering the ratio of cross sections. Nevertheless, 25 correlated uncertainty components from this source remain, and there are 21 further uncertainty components which are treated as correlated or uncorrelated as specified in Ref. [14]. It should be noted that for the present fits the combined luminosity uncertainty at 8 and 13 TeV is not used, so the separate luminosity uncertainties may be correlated with those of other data at these centre-of-mass energies. The experimental precision of the ratios ranges from $\sim 4\%$ at low $E_T^\gamma$ and $|\eta^\gamma|$ to $\sim 10\%$ at the largest $E_T^\gamma$ and $|\eta^\gamma|$.

The ATLAS measurements of inclusive jet cross sections at $\sqrt{s} = 7 \text{ TeV}$ [16], 8 TeV [17] and 13 TeV [18] were considered. The data are presented as a function of jet $p_T^{\text{jet}}$, in six bins of absolute rapidity from $|y^{\text{jet}}| = 0.0$ to $|y^{\text{jet}}| = 3.0$. Jets were reconstructed using the anti-$k_T$ algorithm with jet radius parameter $R = 0.4$ and $R = 0.6$ for data at 7 and 8 TeV, and with $R = 0.4$ for data at 13 TeV. Statistical correlation matrices between bins are provided for all jet data sets. However, it is not possible to fit inclusive jet production data at different centre-of-mass energies simultaneously, because the full experimental systematic uncertainty correlations between these data sets have not been fully specified. The data at 8 TeV are selected for the central fit, firstly because a better understanding of correlated systematic uncertainties has been achieved for these data than for the data at 7 TeV, and secondly because they are available for $R = 0.6$, unlike for the data at 13 TeV, and a larger $R$ value is considered more reliable theoretically [29]. Details of the data set are therefore given only for the jet data at 8 TeV. The data at 7 and 13 TeV have similar features, the main difference being in the $p_T^{\text{jet}}$ range probed. These data sets, including the alternative choice of jet radius, are used.
in cross-checks. At 8 TeV there are 171 jet data points from 20.2 fb−1 of integrated luminosity. The fitted $p_T^{\text{jet}}$ range is from 85 GeV to 2.5 TeV. The dominant uncertainties come from the estimation of the jet energy scale and jet energy resolution. Extensive work [17] has been done to understand correlated sources of uncertainty. There are 320 sources of correlated systematic uncertainty, in addition to the luminosity uncertainty, and the jet energy scale’s $\eta$ intercalibration uncertainty is split into 250 sources. The total uncertainty in the central $p_T^{\text{jet}}$ range, $300 < p_T^{\text{jet}} < 600$ GeV, is $\sim 5\%$ for $|y^{\text{jet}}| < 0.5$, rising to $\sim 10\%$ for $|y^{\text{jet}}|$ in the range 2.5–3.0. For low $p_T^{\text{jet}}$, $85 < p_T^{\text{jet}} < 300$ GeV, the uncertainty is $\sim 15\%$ and for high $p_T^{\text{jet}}$, $600 < p_T^{\text{jet}} < 2000$ GeV, it is $\sim 50\%$.

All the ATLAS input data sets for the QCD fit are summarised in Table 1.

### Table 1

| Data set                      | $\sqrt{s}$ (TeV) | Luminosity (fb−1) | Decay channel                      | Observables entering the fit                           |
|-------------------------------|------------------|-------------------|------------------------------------|--------------------------------------------------------|
| Inclusive $W, Z/\gamma^{\ast}$ [9] | 7                | 4.6               | $e, \mu$ combined                  | $\eta_\ell (W)$, $y_Z (Z)$                            |
| Inclusive $Z/\gamma^{\ast}$ [13] | 8                | 20.2              | $e, \mu$ combined                  | $\cos \theta^* \text{ in bins of } y_{\ell\ell}, m_{\ell\ell}$ |
| Inclusive $W$ [12]            | 8                | 20.2              | $\mu$                              | $\eta_\mu$                                             |
| $W^{\pm} + \text{jets}$ [24]  | 8                | 20.2              | $e$                                | $p_T^W$                                               |
| $Z + \text{jets}$ [25]        | 8                | 20.2              | $e$                                | $p_T^{\text{jet}}$ in bins of $|y^{\text{jet}}|$    |
| $t\bar{t}$ [26,27]           | 8                | 20.2              | lepton + jets, dilepton $m_{\ell\ell}, p_T^{\perp}, y_{\ell\ell}$ |
| $t\bar{t}$ [15]              | 13               | 36                | lepton + jets                      | $m_{\ell\ell}, p_T^{\perp}, y_{\ell\ell}$          |
| Inclusive isolated $\gamma$ [14] | 8, 13            | 20.2, 3.2         | --                                 | $E_T^{\gamma}$ in bins of $y'$                         |
| Inclusive jets [16–18]       | 7, 8, 13         | 4.5, 20.2, 3.2    | --                                 | $p_T^{\text{jet}}$ in bins of $|y^{\text{jet}}|$    |

2.2 Correlations of uncertainties within and between data sets

The correlated systematic uncertainties are applied within each data set with the following small number of exceptions. Firstly, for the $W + \text{jets}$, the $Z + \text{jets}$ and the inclusive jet spectra at 8 TeV, the systematic uncertainty due to the unfolding procedure is treated as being uncorrelated, both within and between these spectra. As shown in Ref. [11], this affects the $\chi^2$ of the fits to $V + \text{jets}$, but has little impact on the fitted PDFs. Similarly, the parton shower systematic uncertainty is decorrelated between the $p_T^{\perp}$ and $m_{ij}$ spectra in the $t\bar{t}$ lepton + jets channel, as done in Ref. [10]. It was established that this decorrelation has a minimal effect on the PDFs, while reducing the fit $\chi^2$ to acceptable levels. This decorrelation and the aforementioned decorrelation of the unfolding systematic uncertainty in $V + \text{jets}$ and inclusive jet data, can be justified because the systematic uncertainties concerned are evaluated from the difference of two Monte Carlo estimates, and thus do not represent well-behaved Gaussian uncertainties. Similar conclusions were reached in a recent study in the MMHT framework [30]. Thirdly, in the inclusive jet data at 8 TeV further decorrelations of such systematic uncertainties, derived from the difference of two Monte Carlo estimates, are considered following Ref. [17]. The experimental systematic uncertainties for the jet energy scale (JES) such as the ‘Flavour Response’, ‘Multi-Jet Balance Fragmentation’, ‘Pile-up Rho Topology’, and the ‘Non-Perturbative Correction’ uncertainty, are not considered completely correlated between all rapidity bins. Instead, they are split into two or three components as a function of rapidity and $p_T^{\text{jet}}$ as specified in the various splitting options described in the Appendix of Ref. [17]. For the central fit, the preferred set of splitting options for $R = 0.6$ is used, in which the JES ‘Flavour Response’ is split into three components (see Table 6 of Ref. [17]). In the present paper, this is called ‘Decorrelation Scenario 2’ and it is chosen because it is one of the two preferred options as determined in the analysis in Ref. [17]. Alternative decorrelation scenarios are also considered in Sect. 5.3.1.

5 The ‘Flavour Response’ is the systematic uncertainty due to the response difference between quark- and gluon-induced jets, the ‘Multi-Jet Balance Fragmentation’ represents the jet fragmentation uncertainty in the multi-jet $p_T^{\perp}$ balance and the ‘Pile-up Rho Topology’ takes into account the uncertainty in the density $\rho$ of pile-up activity in a given event.

6 The decorrelations used at next-to-leading order (NLO) in Ref. [17], include decorrelations of systematic uncertainties due to scale choice. These are not applied in the present NNLO analysis because scale uncertainties are much smaller. NNLO scale uncertainties for the inclusive jets are studied in Sect. 5.3.1.
Correlations of systematic uncertainties between data sets are explained below. The luminosity uncertainties are considered fully correlated for all data sets at the same centre-of-mass energy. For the data sets considered in this analysis, systematic uncertainties involving electron and muon measurements are small (<1%), whereas systematic uncertainties involving the jet measurements can be much larger ($O(10\%)$). Moreover, the high-precision inclusive $W$, $Z/\gamma^*$ differential cross-section measurements at 7 TeV and the inclusive $Z/\gamma^*$ triple differential cross-section measurements at 8 TeV both had the electron and muon channel data combined and thus the identities of the systematic uncertainty sources are lost in the combination procedure such that the analysis cannot correlate them with muon and electron uncertainties in other data sets (or between the inclusive $Z/\gamma^*$ measurements at 7 and 8 TeV). Thus, correlations of lepton uncertainties between these inclusive $W$ and $Z$ measurements and the other data sets are neglected. This is not a significant limitation in the study of the effect of correlations between data sets since the lepton uncertainties are small compared to the jet uncertainties, as shown in the recent study of the impact of $V+\text{jets}$ data on PDF fits [11]. This study reverted to the use of the separate electron and muon channel data for the inclusive $W$, $Z/\gamma^*$ data at 7 TeV in order to correlate the electron systematic uncertainties with those of the $W+\text{jets}$ data and the $Z+\text{jets}$ data at 8 TeV in the electron channel. These studies produce PDFs which are barely different from those in which the electron/muon combined data are used and lepton systematic uncertainty correlations are not applied between the data sets. Indeed, in Sect. 5.2 of the present analysis, it is shown that the impact of the $V+\text{jets}$ data within the present analysis, which uses combined electron/muon data for both inclusive $W$, $Z/\gamma^*$ data at 7 TeV and inclusive $Z/\gamma^*$ data at 8 TeV, is very similar to the impact of these data in Ref. [11], where uncombined electron and muon data are used. Thus, the use of the more accurate combined electron and muon data is preferred.

In this paper, the correlations of the much larger jet-measurement uncertainties are considered. The $W+\text{jets}$ data and the $Z+\text{jets}$ data at 8 TeV have common sources of jet systematic uncertainties, which are listed in Table 2. These sources are considered 100% correlated. Since the $t\bar{t}$ cross sections at both 8 and 13 TeV consider data in the lepton + jets channel, there are some common sources of systematic uncertainty, due to the jet measurements, between these data and the $V+\text{jets}$ data. These are also listed in Table 2 and are considered 100% correlated between these data sets. This table also lists correlations of some sources of smaller systematic uncertainties, in the muon measurements and diboson, single-top and $Z+\text{jets}$ backgrounds, which can be correlated between the $t\bar{t}$ measurements. It should be noted that detailed correlations between the $t\bar{t}$ data at 8 TeV in the dilepton channel and other data sets cannot be applied because the correlations for the dilepton channel are supplied as a covariance matrix and the separate individual sources of uncertainty cannot be separated. This does not have a significant effect on the fit since the dilepton data itself has only a small impact on the fit, as seen in Ref. [10].

Correlations of jet systematic uncertainties between the inclusive jet data and the $V+\text{jets}$ data and $t\bar{t}$ data in the lepton + jets channel have been identified and are listed in Table 2. There is a choice to be made in the implementation of these inter-data-set correlations. The JES ‘Flavour Response’ and ‘Pile-up Rho Topology’ are part of the Jet Decorrelation Scenario 2 chosen for the central fit, and thus they have been split into three components according to jet rapidity and $p_T$ [17]. Consequently, there are no longer corresponding systematic sources in the $V+\text{jets}$ data and $t\bar{t}$ data. The choice made for the central fit is to correlate only the remaining six sources (in the right-hand column of Table 2) between the data sets. The alternative choice of correlating these two systematic sources with the other data sets but not including them in the jet decorrelation scenario was also investigated. The effect on the resulting PDFs is negligible.

There is a further caveat on how the correlations between the inclusive jet data set and the $V+\text{jets}$ and $t\bar{t}$ data sets are applied, namely that the last two data sets have radius $R = 0.4$ jets, and hence the systematic uncertainties may not be fully correlated. Checks were made using 100% correlation and no correlation, yielding little difference between the resultant PDFs. For the central fit a correlation of 100% is used.

The systematic uncertainties of the inclusive jet data at different beam energies are correlated with each other, but understanding these correlations in detail is non-trivial. In the present study, these data sets are fitted separately and results are compared. As already stated the data at 8 TeV are used for the central fit.

The measurement of the direct-photon production ratio already considered correlations between the data at 8 TeV and 13 TeV. The photon energy scale is the largest correlated systematic uncertainty between the two measurements. There are no further important correlations with the other data sets. The luminosity uncertainties of the data at 8 TeV and 13 TeV are not combined for the present study. Instead, the 8 TeV luminosity is correlated with that of the other 8 TeV data sets and the 13 TeV luminosity is correlated with that of the other 13 TeV data sets.
Table 2 Systematic uncertainties that are correlated between the \( W \) + jets data at 8 TeV, \( Z \) + jets data at 8 TeV, \( t \bar{t} \) lepton + jets data at 8 TeV, \( t \bar{t} \) lepton + jets data at 13 TeV and inclusive jets data at 8 and 13 TeV are listed. The names of the systematic uncertainties are those found in the HEPData entries [31]. Entries in the same row are taken as 100% correlated for the \( V \) + jets and \( t \bar{t} \) lepton + jets data, which all have jet radius \( R = 0.4 \). Different degrees of correlation are considered for the inclusive jet data at \( R = 0.6 \), because of the differing choice of jet radius. Where entries are omitted, that systematic uncertainty does not exist for that data set (denoted by ‘–’). The luminosity uncertainty of data sets at the same centre-of-mass energy are also fully correlated. The JES ‘Flavour Response’ and JES ‘Pile-up Rho topology’ are considered fully correlated with other data sets only for cross-checks. They are not correlated for the central fit because they are part of the Decorrelation Scenario 2 which is applied to the inclusive jet measurements, as explained in the text. For this reason they are marked with the symbol ‘∗’.

| Systematic uncertainty |     | 8 TeV \( W \) + jets | 8 TeV \( Z \) + jets | 8 TeV \( t \bar{t} \) lepton + jets | 13 TeV \( t \bar{t} \) lepton + jets | 8 TeV inclusive jets |
|------------------------|-----|----------------------|----------------------|---------------------------------|----------------------|----------------------|
| Jet flavour response   | JetScaleFlav2 | Flavor Response | flavres-jes | JET29NP JET Flavour Response | syst JES Flavour Response∗ |
| Jet flavour composition| JetScaleFlav1Known | Flavor Comp | flavcomp-jes | JET29NP JET Flavour Composition | syst JES Flavour Comp |
| Jet punchthrough       | JetScalepunchT | Punch Through | punch-jes | – | syst JES PunchThrough MC15 |
| Jet scale              | JetScalePileup2 | PU OffsetMu | pileoffmu-jes | – | syst JES Pileup MuOffset |
|                        | – | PU Rho | pileoffrho-jes | JET29NP JET Pileup RhoTopology | syst JES Pileup Rho topology∗ |
|                        | JetScalePileup1 | PU OffsetNPV | pileoffnpv-jes | JET29NP JET Pileup OffsetNPV | syst JES Pileup NPV Offset |
|                        | – | PU PrTerm | pileoffPrTerm | JET29NP JET Pileup PrTerm | syst JES Pileup Pr Term |
| Jet JVF selection      | JetJVFcut | JVF | jetvxfac | – | syst JES Zjets JVF |
| B-tagged jet scale     | – | btag-jes | JET29NP JET BJES Response | – | – |
| Jet resolution         | – | jeten-res | JET JER SINGLE NP | – | – |
| Muon scale             | – | – | mup-scale | MUON SCALE | – |
| Muon resolution        | – | – | muonms-res | MUON MS | – |
| Muon identification    | – | – | muoid-res | MUON ID | – |
| Diboson cross section  | – | – | dibos-xsec | Diboson xsec | – |
| \( Z \) + jets cross section | – | – | zjet-xsec | Zjets xsec | – |
| Single-\( t \) cross section | – | – | singletop-xsec | st xsec | – |
3 Theoretical framework

3.1 NNLO QCD and NLO electroweak predictions for each data set

The present analysis uses the xFitter framework [32–34]. This program interfaces to theoretical calculations directly or uses fast interpolation grids to make theoretical predictions for the considered processes. The program MINUIT [35] is used for the minimisation. Each step is cross-checked with an independent fit program [36]. This section describes how these predictions are obtained for each data set.

For the DIS processes the light-quark coefficient functions are calculated to NNLO in QCD theory as implemented in QCDNUM [37]. The contributions of charm and bottom quarks are calculated in the general-mass variable-flavour-number scheme of Refs. [38,39], known as the optimised TRVFN scheme.

The renormalisation and factorisation scales for the DIS processes are taken to be the conventional choices, \( \mu_r = \mu_f = \sqrt{Q^2} \), where \( Q^2 \) is the negative four-momentum transfer squared as already stated. For the DIS data electroweak (EW) effects are already unfolded to leading order, such that LO-EW corrections need not be applied, apart from the running of \( \alpha \). NLO-EW corrections are not well defined and no such corrections are applied to the DIS data.

In the present analysis, the photon PDF within the proton is not accounted for and thus data sets with substantial sensitivity to the photon PDF, such as very high-mass Drell–Yan data, are excluded. It is estimated [40] that the photon takes only \( \sim 0.3\% \) of the momentum of the proton in our kinematic range.

The DIS processes are the only ones for which fast NNLO QCD predictions can be made analytically, such that they can be used in an iterative fit. For the LHC processes, NNLO calculations are too time-consuming and thus fast interpolation grids are used. However, currently only the \( t\bar{t} \) production processes in the lepton + jets channel have grids available for NNLO calculations. For the remainder of the data sets the grids are available only at NLO and \( K \)-factors are used to correct the QCD predictions for the differential cross sections from NLO to NNLO. These \( K \)-factors are calculated from the ratio of the NNLO to NLO cross sections, with the same cuts as the data, using a fixed input PDF. These \( K \)-factors have very little dependence on the choice of input PDF for this calculation, and iteration using the output PDF of the fit is not necessary. The calculations used to construct the interpolation grids are usually to leading order (LO) in the EW part of the calculation. Correction from LO to NLO in EW predictions is also applied by a \( K \)-factor technique. For some data sets, this is applied together with the QCD predictions’ correction from NLO to NNLO while for others it is a separate calculation. Details for each data set are given below and summarised in Table 3. For data sets measuring

Table 3 Summary of code used for NLO, NNLO QCD and LO, NLO EW predictions for ATLAS data as applied in the QCD fit. For most data sets, predictions are provided at NLO in QCD and LO in EW in the form of fast interpolation grids and are corrected to NNLO QCD and NLO EW by \( K \)-factors. For the \( t\bar{t} \) lepton + jets channel the grids are calculated at NNLO in QCD directly, so no entry appears in the column ‘NLO QCD code’.

| Data set          | NLO QCD code | NNLO QCD code | LO EW code | NLO EW code |
|-------------------|--------------|---------------|------------|-------------|
| Inclusive         | MCFM         | MG5_aMC@NLO2.6.4 | N_{jet}    | Ref. [52]   |
| W ±jets [24]      | MCFM         | MG5_aMC@NLO2.6.4 | N_{jet}    | Ref. [52]   |
| Z ± jets [25]     | Ref. [56]    | Ref. [56]     | MCFM       | Ref. [56]   |
| Inclusive isolated γ [14] | MCFM       | MG5_aMC@NLO2.6.4 | N_{jet}    | Ref. [59]   |
| Inclusive jets [16–18] | MCFM       | MG5_aMC@NLO2.6.4 | N_{jet}    | Ref. [64]   |

\( \odot \) Springer
final-state jets, there are also non-perturbative corrections to account for hadronisation. These are also applied using a \( K \)-factor technique as specified below.\(^7\)

The \( W \) and \( Z \) inclusive cross sections at 7 TeV are calculated at fixed order, to NNLO in QCD theory and to NLO in EW theory, as described in Ref. [9]. The results obtained from DYNNLO 1.5 [41,42] and FEWZ 3.1.b2 [43,44] are compared. A small difference between the NNLO predictions of FEWZ and DYNNLO of up to \( \sim 1\% \) is observed for the \( W \)- and \( Z \)-peak data. The sensitivity of NNLO programs to the fiducial cuts applied to the data was highlighted again recently [23]. This is considered further in Sect. 3.2. The scales for the Drell–Yan (DY) processes are taken to be the decay dilepton invariant mass appropriate to the centre of the mass bin in the NC case and the \( W \)-boson mass in the CC case. The xFitter package uses the APPLGRID code [45] interfaced to the MCFM program [46,47] for fast calculation of the differential \( W \) and \( Z / \gamma^* \) boson cross sections at NLO in QCD theory and LO in EW theory, and a \( K \)-factor technique is used to correct the NLO QCD predictions to NNLO and the LO EW predictions to NLO. These \( K \)-factors are within 1–2\% of unity for the \( W^\pm \) and \( Z / \gamma^* \) in the c-c dilepton rapidity region and within \( \sim 4\% \) of unity in the c-f dilepton rapidity region. The high-mass band of \( Z / \gamma^* \) production is also subject to background from photon-induced dilepton production, which was estimated using the MRST2004QED photon PDF [48] and subtracted from the data. Compared to the signal, the size of this background is \((1.5 \pm 0.5)\% \).

The QCD predictions for the triple differential distributions of \( Z / \gamma^* \) production at 8 TeV are made to NLO by MCFM interfaced to APPLGRID, and \( K \)-factors are applied for NNLO QCD and NLO EW effects using NNLOJET. These \( K \)-factors are within \( \sim 3\% \) of unity. The renormalisation and factorisation scales for these data are taken to be the decay dilepton invariant mass appropriate to the centre of the mass bin. For the present study, Particle Data Group (PDG) values [49] are used for the electroweak parameters in the \( G_{\mu} \) scheme, in particular the value of \( \sin^2 \theta_W \) is \( 0.23127 \).

The QCD predictions for the \( W \) cross sections at 8 TeV are calculated to NLO and the EW predictions to LO using MG5\_@MCFM via AMCATNLO 2.6.4 interfaced to APPLGRID via AMC\_FAST 1.3.0, and the NNLO QCD + NLO EW \( K \)-factors are calculated using DYNNLO 1.5. These \( K \)-factors are within 1%–2% of unity. The renormalisation and factorisation scales for these data are taken to be the \( W \)-boson mass.

The predictions for \( W + \) jets production at 8 TeV are obtained at fixed order, up to NNLO in QCD theory and at LO in EW theory, using the N\_jetti program [50]. Outputs from the APPLGRID code are used for fast calculation at NLO in QCD theory, and \( K \)-factors from the aforementioned N\_jetti prediction are used to correct this calculation to NNLO. The renormalisation and factorisation scales are set to \( \sqrt{m_W^2 + \Sigma(p_T^{\text{jet}})^2} \), where the second term in the square root is a sum over the transverse momentum of each jet. Additionally, all non-perturbative corrections, such as for hadronisation effects, are included in the \( K \)-factors. Further \( K \)-factors are applied for NLO EW corrections as computed by the authors of SHERPA [51]. As in the previous ATLAS analysis of \( V + \) jets data [11], the lowest \( p_T^{\text{jet}} \) bin, \( p_T^{\text{jet}} < 25 \) GeV, is not used, since the calculation is effectively only at NLO for such low \( p_T \).

The NNLO QCD predictions for \( Z + \) jets production at 8 TeV were calculated by the authors of Ref. [52]. The renormalisation and factorisation scales are set to \( \mu_r = \mu_F = \frac{1}{2} \sqrt{m_{\ell\ell}^2 + p_{T,\ell\ell}^2 + \Sigma p_{T,\text{partons}}} \), where \( m_{\ell\ell} \) is the invariant mass of the electron pair, \( p_{T,\ell\ell} \) is the transverse momentum of the electron pair and \( \Sigma p_{T,\text{partons}} \) is the sum of the transverse momenta of the outgoing partons. Four sets of \( K \)-factors are provided: the first corrects the NLO QCD predictions to NNLO, the second gives corrections for non-perturbative effects, the third corrects from LO to NLO in QED and the fourth applies NLO EW corrections (excluding QED radiation) as computed by SHERPA 2.2.10. The \( K \)-factors for both \( W + \) jets and \( Z + \) jets production are typically within 10% of unity. Uncertainties in the non-perturbative corrections are supplied for the \( Z + \) jets predictions and these are applied as correlated systematic uncertainties. Their impact is very small. Such uncertainties are not supplied for the \( W + \) jets predictions.

The NNLO QCD predictions for \( t \bar{t} \) production at 8 TeV were calculated by the authors of Ref. [53] and are available in the form of fast interpolation grids, APPLGRID [45] or FASTNLO [54,55], for the data in the lepton+jets channel. The predictions for \( m_t \) are given for renormalisation and factorisation scales equal to \( H_T / 4 \), where \( H_T = \sqrt{m_t^2 + (p_T^t)^2} + \sqrt{m_t^2 + (p_T^t)^2} \), and the predictions for \( p_T^t \) are given for scales equal to \( m_t / 2 \), where \( m_t = \sqrt{m_t^2 + (p_T^t)^2} \) and \( m_t = 173.3 \) GeV is the pole mass. Non-perturbative corrections and their uncertainties are supplied with the data and these uncertainties are applied as correlated systematic uncertainties. For the dilepton channel, APPLGRID is interfaced to MCFM to produce NLO grids, and a \( K \)-factor technique is used to correct NLO predictions to NNLO, using \( K \)-factors from Ref. [28]. The scales used for the predictions are equal to \( H_T / 4 \). The NNLO/NLO \( K \)-factors are \( \sim 7\% \) above unity and constant for \( y_t \). The calculations are for zero width of the top mass. NLO EW corrections for the spectra are also considered, using the additional \( K \)-factors given in Ref. [56].

\(^7\)Note that the ATLAS data are all at much higher scales and the charm and bottom quarks are treated in the 5-flavour zero-mass variable flavour number scheme, for which the top is an additional heavy quark.
For the 8 TeV data, these corrections can be up to 1% for \( y_{||} \) and \( y_{\perp} \), up to 2% for \( m_{\perp} \), and up to 4% for \( p_T \).

The NNLO predictions for \( t\bar{t} \) production at 13 TeV have been calculated by the authors of Ref. [53] in the form of fast interpolation grids [45,54,55] for the data in the lepton+jets channel [57]. The predictions for \( m_{\perp}, y_{||} \) and \( y_{\perp} \) are given for renormalisation and factorisation scales equal to \( m_T/2 \), just as for the 8 TeV \( t\bar{t} \) data. The only difference is that at 13 TeV the value \( m_t = 172.5 \text{ GeV} \) is used for the top-quark mass, because this is the value used for the interpolation grids. The small difference between the masses used for the central fit for 8 and 13 TeV \( t\bar{t} \) data is studied as part of the model uncertainties and found to be negligible (see Sect. 5.3.1). NLO EW corrections for the spectra are also considered, using \( K \)-factors taken from Ref. [56]. For the 13 TeV data, these corrections can be up to 1% for \( y_{||} \), up to 0.5% for \( y_{\perp} \), up to 2.5% for \( m_{\perp} \), and up to 5% for \( p_T \). Non-perturbative corrections and their uncertainties are again supplied with the data and these uncertainties are applied as correlated systematic uncertainties.

The NLO QCD predictions for direct photon production were taken from MCFM interfaced to APPLGRID, and \( K \)-factors from Ref. [58] were applied to obtain NNLO QCD predictions. Resummation of electroweak effects is also applied as in Ref. [59]. The corrections from these \( K \)-factors are 2%–15% and 0.5%–10% for the separate 8 and 13 TeV data sets, respectively, but they almost cancel out in the ratio, leaving corrections of less than 2%. Finally, a photon isolation criterion is applied to the predictions to follow that applied to the data as closely as possible (see Ref. [60]). The renormalisation and factorisation scales are set to \( E_T^\gamma \).

The NLO QCD predictions for inclusive jet production at 7, 8 and 13 TeV are made with NLO\textsc{jet++} interfaced to APPLGRID. The renormalisation and factorisation scale choices \( \mu_r = \mu_f = p_T^{\text{max}} \) and \( \mu_r = \mu_f = p_T^{\text{jet}} \) are both available. For the 7 and 8 TeV data, predictions are available for jet radii \( R = 0.4 \) and 0.6, while for the 13 TeV data they are available for \( R = 0.4 \) only, matching the data. The \( K \)-factors to convert NLO QCD predictions to NNLO have been calculated using NNLO\textsc{jet} [61]. These \( K \)-factors are rather different for the two scale choices. For the 8 TeV data and scale \( p_T^{\text{max}} \) they range from \( \sim 10\% \) at low \( p_T \) to a few percent at high \( p_T \) for low \( |y^{\text{jet}}| \), but the high-\( p_T \) corrections can become slightly negative at high \( |y^{\text{jet}}| \). For the scale \( p_T^{\text{jet}} \), they range from \( \sim 5\% \) at low \( p_T \) to a few percent positive at high \( p_T \) for low \( |y^{\text{jet}}| \), but become increasingly negative for all \( p_T \) at high \( |y^{\text{jet}}| \). The \( K \)-factors for the 7 TeV data are very similar, and those for the 13 TeV data follow the same trends while being a bit larger. The choice of scale for the central fit is \( \mu_r = \mu_f = p_T^{\text{jet}} \), since this is preferred theoretically [29], but the alternative scale choice \( \mu_r = \mu_f = p_T^{\text{max}} \) and other scale variations are considered in Sect. 5.3.1.

The precision of the \( K \)-factors, in the predictions for inclusive jet data, depends on the statistical precision of the NNLO\textsc{jet} calculations which evaluated the NLO and NNLO cross sections. For the 8 TeV inclusive jet data, the resulting \( K \)-factors suffer from some statistical fluctuations that can be as large as \( \sim 1\% \). 8 For the central fit, these \( K \)-factors are smoothed and the estimated uncertainty in this procedure is applied as being 60% correlated and 40% uncorrelated between \( p_T \) bins within the same rapidity bin. Alternative treatments are considered in Sect. 5.3.1.

The QCD inclusive jet predictions are corrected for NLO electroweak effects as detailed in Ref. [17]. This correction can reach more than 10% for the highest \( p_T^{\text{jet}} \) in the lowest \( |y^{\text{jet}}| \) bin, but decreases rapidly as \( |y^{\text{jet}}| \) increases. It is less than 3% for \( |y^{\text{jet}}| > 1.0 \).

In order to compare fixed-order QCD calculations with the measured inclusive jet cross sections, corrections for non-perturbative effects must also be applied. These are derived using LO Monte Carlo event generators. There are significant differences between the corrections derived with different event generators, \textsc{Pythia} [62] and \textsc{Herwig} [63], and different sets of tuned parameter values. The corrections can be up to \( \sim 10\% \) at low \( p_T^{\text{jet}} \). They are different for \( R = 0.4 \) and \( R = 0.6 \) due to differing interplay of the underlying-event and hadronisation corrections, and for this reason fits to data with both jet radii are investigated. Values of the non-perturbative correction are taken at the centre of the uncertainty band, which is constructed as the envelope of all such corrections considered, and the systematic uncertainty of each correction value, correlated in \( p_T^{\text{jet}} \) and \( y^{\text{jet}} \), is taken from the spread of the band (see Ref. [17] for details).

Fits were performed using 7, 8 and 13 TeV inclusive jet data separately, using both choices for the jet radius, \( R = 0.4 \) and \( R = 0.6 \), when available, and both choices for the scales, \( p_T^{\text{max}} \) and \( p_T^{\text{jet}} \), but the final choice for the central fit is 8 TeV jets with \( R = 0.6 \) and scales \( \mu_r = \mu_f = p_T^{\text{jet}} \), as discussed in Sect. 5.3.1. Table 3 details the code used for NNLO/NLO QCD corrections and NLO/LO EW corrections to ATLAS data in the QCD fit, for ease of reference.

3.2 Scale uncertainties and sensitivity to NNLO code

For some of the ATLAS data sets considered, the precision of the data set is so high that it is comparable to the size of the NNLO scale uncertainties. This is the case for the ATLAS \( W \) and \( Z/\gamma^* \) inclusive data sets at both 7 and 8 TeV, for which the total experimental uncertainty and the scale uncertainties

---

8 For the other processes considered in this section the statistical precision of the \( K \)-factors is much higher and further consideration of this source of uncertainty is not necessary.
both approach \( \sim 0.5\% \). In this case, the scale uncertainties are considered as additional theoretical uncertainties which are added to the \( \chi^2 \) calculation in the same way as experimental systematic uncertainties (see Sect. 4). These scale uncertainties are evaluated as follows. The \( K \)-factors are evaluated for separate changes of the renormalisation and factorisation scales by factors of 2 and 0.5. The magnitude of the \( K \)-factor difference is symmetrised as \((K[\mu_r(2)] - K[\mu_r(0.5)])/2\) and \((K[\mu_t(2)] - K[\mu_t(0.5)])/2\) and its sign is preserved as positive if the upward variation of \( \mu_r \) or \( \mu_t \) makes the \( K \)-factor increase and negative if it makes the \( K \)-factor decrease. For these renormalisation and factorisation scale changes, the ratios of the signed \( K \)-factor changes to the \( K \)-factors for the nominal scales are calculated to obtain the fractional scale uncertainties. These can then be used in the fit like any other systematic uncertainty, as suggested in Ref. [65].

Due to the similarity of the \( W \) and \( Z \) processes, both the renormalisation scale and factorisation scale are considered correlated within the \( W, Z \) data sets at 7 TeV and between the \( W \) and \( Z \) data sets at 8 TeV. They are also considered to be correlated between the \( W \) and \( Z \) data sets at 7 and 8 TeV for the central fit. Studies of alternative approaches for the scale uncertainties are presented in Appendix A.

Scale uncertainties are accounted for in the central fit only for the \( W \) and \( Z \) inclusive data sets, because the scale uncertainties are comparable to the experimental uncertainties only in this case. For the other ATLAS data sets the experimental uncertainties are larger and scale uncertainties are treated by repeating the fit with varied scales as discussed in Sect. 5.3.1.

As remarked earlier, there is a small difference of up to \( \sim 1\% \) between the NNLO predictions of FEWZ and DYNNLO. A fractional uncertainty can be defined as the ratio of the signed difference between the FEWZ and DYNNLO predictions to the FEWZ prediction, and applied similarly to the scale uncertainties. The inclusion of such a theoretical uncertainty makes no further difference to the fits, provided the scale uncertainties are applied.

Recently, next-to-next-to-next-to-leading-order (N3LO) calculations for the total cross section of the Drell–Yan process have become available [66]. It is notable that the N3LO result does not quite lie within the scale uncertainties of the NNLO result. The N3LO correction to the central value is \( \sim 2\% \) compared to NNLO scale uncertainties of \( \sim 0.5\% \). Full accounting for N3LO corrections, as well as \( p_T \) resummation effects which affect the prediction of \( W \) and \( Z \) fiducial cross sections at the level of \( \sim 0.5 - 1\% \) [23], are beyond the scope of the current fixed-order-NNLO QCD fit.

### 4 Fit methodology

The DGLAP [67–69] evolution equations yield the PDFs at any value of \( Q^2 \) given that they are parameterised as a function of \( x \) at a starting scale \( Q_0^2 \). In the present analysis, this scale is chosen to be \( Q_0^2 = 1.9 \text{ GeV}^2 \) so that it is below the charm mass threshold \( m_c^2 \). A recent combination of HERA heavy-quark data [70] has stimulated a re-analysis of the optimal values of the heavy-quark masses and their variations as used in the optimised TRVFN scheme [39]. The resulting values are \( m_c = 1.41 \text{ GeV} \) and \( m_b = 4.2 \text{ GeV} \) [71], and these are used for the central fit as applied to the DIS data. A minimum \( Q^2 \) cut of \( Q_{\min}^2 = 10.0 \text{ GeV}^2 \) is imposed on the HERA data. All these assumptions are varied in the consideration of model uncertainties, as shown in Sect. 5.3.1. The strong coupling constant is fixed to \( \alpha_s(m_Z) = 0.118 \).

The quark distributions at the starting scale are represented by the generic form

\[
x q_i(x) = A_i x^{B_i} (1 - x)^{C_i} P_i(x),
\]

where \( P_i(x) = (1 + D_i x + E_i x^2 + F_i x^3) \). The parameterised quark distributions, \( x q_i \), are chosen to be the valence quark distributions \((x u, x d, x s)\) and the light anti-quark distributions \((\bar{x} u, \bar{x} d, \bar{x} s)\). The gluon distribution is parameterised with the more flexible form

\[
x g(x) = A_g x^{B_g} (1 - x)^{C_g} P_g(x) - A'_g x^{B'_g} (1 - x)^{C'_g},
\]

where \( C'_g \) is set to 25 to suppress negative contributions at high \( x \). The parameters \( A_{u_\ell} \) and \( A_{d_\ell} \) are fixed using the valence-quark number sum rule, and \( A_{s_\ell} \) is fixed using the momentum sum rule. The normalisation and slope parameters, \( A \) and \( B \), of the light-quark sea, \( x \bar{u}, x \bar{d}, x \bar{s} \), are all independent of each other, such that there is no constraint on \( x \bar{d} - x \bar{u} \), or on \( \bar{s}/(\bar{d} + \bar{u}) \), either in shape or in normalisation as \( x \to 0 \). By default it is assumed that \( x s = \bar{x} s \). The \( D, E \) and \( F \) terms in the polynomial expansion \( P_i(x) \) are used only if required by the data, following the procedure described in Ref. [33], whereby parameters are added only if the \( \chi^2 \) of the fit decreases significantly. This leads to a 21-parameter fit with: \( P_{u_\ell}(x) = 1 + D_{u_\ell} x + E_{u_\ell} x^2, P_{d_\ell}(x) = 1 + D_{d_\ell} x \) and \( P_{g}(x) = 1 + D_g x, \) with \( P_{s}(x) = 1 \) for the light-sea PDFs. It has been established that adding further parameters beyond the point at which there is little further change in \( \chi^2 \) (called the point of ‘saturation’ of the \( \chi^2 \) ) only serves to fit noise in the data [72]. Thus, addition of further parameters to the central fit is not considered. However, parameterisation uncertainties are considered in Sect. 5.3.2.

A cross-check was made using Chebyshev polynomials rather than ordinary polynomials for the \( P_i(x) \) terms. There is no improvement in \( \chi^2 \) for a comparable 21-parameter fit using the Chebyshev polynomials and the extracted PDFs lie within the total uncertainty bands of the ordinary polynomial fit.

The level of agreement between the data and the predictions from a PDF parameterisation is quantified by the \( \chi^2 \) per
degree of freedom ($\chi^2$/NDF, for NDF degrees of freedom). The definition of the $\chi^2$ is

$$\chi^2 = \sum_{i,k} \left( D_i - T_i (1 - \sum_j \gamma_{i,j} b_j) \right)^2 
\times C^{-1}_{\text{stat,uncor},ik} (D_i, D_k) \left( D_k - T_k (1 - \sum_j \gamma_{k,j} b_j) \right) + \sum_j \log \frac{\delta^2_{\text{uncor}} T_i^2 + \delta^2_{\text{stat}} D_i T_i}{\delta^2_{\text{uncor}} D_i^2 + \delta^2_{\text{stat}} D_i^2} + \sum_j b_j^2$$

where $D_i$ represent the measured data, $T_i$ the corresponding theoretical prediction, $\delta_{\text{uncor}}$ and $\delta_{\text{stat}}$ are the uncorrelated systematic uncertainties and the statistical uncertainties of $D_i$, and the correlated systematic uncertainties, described by $\gamma_{i,j}$, are accounted for using the nuisance parameters $b_j$. The quantity $C_{\text{stat,uncor},ik}$ is a covariance matrix for both the statistical and uncorrelated systematic uncertainties. Summations over $i$ and $k$ run over all data points and summation over $j$ runs over all sources of correlated systematic uncertainty. For each data set, the first term gives the main contribution to the partial $\chi^2$ of the data set and the second term is a small bias correction term, referred to as the log penalty, which arises because the diagonal term of the matrix, $C$, is given by $C_{ii} = \delta^2_{\text{uncor}} T_i^2 + \delta^2_{\text{stat}} D_i T_i$, with different weighting for statistical uncertainties and uncorrelated systematic uncertainties. This form of the $\chi^2$ is used as standard in HERA and ATLAS PDF fits [5,9,10]. In reporting the $\chi^2$ for the fits, the first and second terms together give the partial $\chi^2$ of the data set. The third term gives the correlated $\chi^2$ arising from the penalty for the nuisance parameters describing correlations of systematic uncertainties within and between data sets. The experimental uncertainties of the fit are first set using the usual tolerance, $T = 1$, where $T^2 = \Delta \chi^2 = 1$. The use of an enhanced tolerance is considered in Sect. 6.

5 Results

In this section the ATLASpdf21 PDF set is presented. The impact of variations of the central choice of fit settings and parameterisation is discussed in Sect. 5.3. Table 4 gives the total $\chi^2$ per degree of freedom, $\chi^2$/NDF, of the fit using all data sets and the $\chi^2$ per data point, $\text{partial} \chi^2$/NDF for NDF data points, of each data set. The correlated terms (term 3) are shown separated into groups with common systematic correlations. In order to evaluate the separate contributions of the data sets to this correlated term, the fit is run with its final parameters fixed for each data set separately. These values follow the total correlated terms in brackets, in order of their appearance in the table. The quality of the fit to the HERA data is $\chi^2$/NDF = 1.14, comparable to that of HERAPDF2.0, so that there is no tension between the ATLAS data and HERA data. The quality of the fit to the ATLAS $W, Z$ data is $\chi^2$/NDF = 1.44, and to the ATLAS $V$+jets, $t\bar{t}$ and inclusive jet data it is $\chi^2$/NDF = 1.40. The quality of fit to the ATLAS direct photon data is $\chi^2$/NDF = 0.7. These $\chi^2$ values are comparable to those obtained by the global PDF fits for similar data sets, but indicate a need to consider the appropriate $\chi^2$ tolerance of the fit. Both of these points are discussed further in Sect. 6.

5.1 Comparison of PDFs with and without inclusion of correlations between data sets

The ATLASpdf21 PDFs at $Q^2 = 1.9$ GeV$^2$ (the starting scale) are shown in Figs. 1 and 2. Only experimental uncertainties with tolerance $T = 1$ are shown for the comparisons throughout this section. Full uncertainties including model and parameterisation variations are considered for the ATLASpdf21 fit in Sect. 5.3.

Figure 2 also shows the ratio $R_s = x(s + \bar{s})/(u + \bar{u})$ and the difference $x(d - \bar{u})$. In this section, all PDFs are presented with uncertainties evaluated using the standard $\chi^2$ tolerance.

| Table 4 $\chi^2$ contributions for the all data sets entering the PDF fit. The partial $\chi^2$ for the individual data sets are given with respect to the number of data points (NDF). They represent the addition of terms 1 and 2 in Eq. (1). The correlated terms (term 3) are shown separated into groups with common systematic correlations. The total value of the correlated term for these groups is also split into the addition of the separate contributions in the order in which they are given in the table |
| --- |
| Total $\chi^2$/NDF |
| HERA $\chi^2$/NDF |
| HERA correlated term |
| ATLAS $W, Z$ 7 TeV $\chi^2$/NDF |
| ATLAS $Z/\gamma^*$ 8 TeV $\chi^2$/NDF |
| ATLAS $W$ 8 TeV $\chi^2$/NDF |
| ATLAS $W$ and $Z/\gamma^*$ 7 and 8 TeV correlated term |
| ATLAS direct $\gamma$ 13/8 TeV $\chi^2$/NDF |
| ATLAS direct $\gamma$ 13/8 TeV correlated term |
| ATLAS $V+$jets 8 TeV $\chi^2$/NDF |
| ATLAS $t\bar{t}$ 8 TeV $\chi^2$/NDF |
| ATLAS $t\bar{t}$ 13 TeV $\chi^2$/NDF |
| ATLAS inclusive jets 8 TeV $\chi^2$/NDF |
| ATLAS $V+$jets 8 TeV and $t\bar{t}$ + jets 8,13 TeV and $R = 0.6$ inclusive jets 8 TeV correlated term |
| 2010/1620 |
| 1112/1016 |
| 50 |
| 68/55 |
| 208/184 |
| 31/22 |
| 71 = (38 + 33) |
| 27/47 |
| 6 |
| 105/93 |
| 13/20 |
| 25/29 |
| 207/171 |
| 87 = (16 + 9 + 21 + 41) |
Fig. 1 ATLASpdf21 PDFs comparing those extracted from a fit in which correlations of systematic uncertainties between data sets are applied, with those extracted from a fit in which only the luminosity uncertainties for each centre-of-mass energy are correlated between data sets. Only experimental uncertainties are shown, evaluated with tolerance \( T = 1 \). Top left: \( xu \). Top right: \( xd \). Bottom left: \( x\bar{u} \). Bottom right: \( x\bar{d} \). The lower panels show the comparison as a ratio to the default ATLASpdf21 PDF.

\[ T^2 = \Delta \chi^2 = 1. \] Enhanced tolerance is considered in Sect. 6. Also shown in these figures are the PDFs from a fit in which correlations of systematic uncertainties between the \( V + \text{jets} \) (lepton + jets) and inclusive jets data sets are not applied, with the exception of the luminosities for the data at the same centre-of-mass energy.\(^9\) The systematic uncertainties which are correlated are specified in Table 2 and their treatment for the central fit is discussed in Sect. 2.2.

\(^9\) The \( W + \text{jets} \) and \( Z + \text{jets} \) data sets are always considered as a single data set with correlations applied and these correlations are maintained in the present exercise. Consideration of the impact of correlations between the \( W + \text{jets} \) and \( Z + \text{jets} \) data sets was studied in Ref. [11].

The differences between the extracted PDFs with and without inter-data-set correlations are mostly in the \( d \)-type sector and can reach \(~\sim 20\%\) for \( x\bar{d} \) at large \( x \). The PDF difference \( x(\bar{d} - \bar{u}) \) and ratio \( R_s \), shown in the bottom half of Fig. 2, are particularly sensitive to the use of these correlations.
Fig. 2 ATLASpdf21 PDFs comparing those extracted from a fit in which correlations of systematic uncertainties between data sets are applied, with those extracted from a fit in which only the luminosity uncertainties for each centre-of-mass energy are correlated between data sets. Only experimental uncertainties are shown, evaluated with tolerance $T = 1$. Top left: $x\bar{s}$. Top right: $xg$. Bottom left: $x(d - \bar{u})$. Bottom right: $R_s$. For the $x\bar{s}$ and $xg$ plots, the lower panels show the comparison as a ratio to the default ATLASpdf21 PDF.

It is important to examine these differences at a scale relevant for precision LHC physics and this is illustrated in Figs. 3 and 4 where the ratios of the PDFs with and without the correlations are shown at $Q^2 = 10,000$ GeV$^2$. Although these differences are not large compared to current experimental precision, they can nevertheless be important if $O(1\%)$ accuracy is to be a realistic goal for PDFs [73].

These plots also allow a comparison of the size of the current experimental uncertainty of the PDFs with and without inclusion of inter-data-set correlations. The sizes are generally similar but the uncertainty in the high-$x$ $d$, strange and gluon PDFs is larger if such correlations are not taken into account.
Fig. 3 Ratios of ATLASpdf21 PDFs extracted from a fit including correlations of systematic uncertainties between data sets to those extracted from a fit in which only the luminosity uncertainties for each centre-of-mass energy are correlated between data sets, at scale $Q^2 = 10,000 \text{ GeV}^2$. Only experimental uncertainties are shown, evaluated with tolerance $T = 1$. Top left: $x_u$. Top right: $x_d$. Bottom left: $x \bar{u}$. Bottom right: $x \bar{d}$.
5.2 Impact of each data set

In this section the impact of each data set is considered. Only experimental uncertainties with tolerance $T = 1$ are shown for these comparisons. Full uncertainties including model and parameterisation variations are considered for the ATLASpdf21 fit in Sect. 5.3.

5.2.1 Impact of $W, Z$ inclusive data

Figure 5 shows the ratio $R_s$ for the ATLASpdf21 fit and compared with a fit in which the inclusive $W, Z$ data at 7 and 8 TeV are removed (left-hand plot), as well as to a fit in which only $W, Z$ data at 8 TeV are removed (right-hand plot). It is clear that without $W, Z$ inclusive data the ratio $R_s$ cannot be

---

Fig. 4 Ratios of ATLASpdf21 PDFs extracted from a fit including correlations of systematic uncertainties between data sets to those extracted from a fit in which only the luminosity uncertainties for each centre-of-mass energy are correlated between data sets, at scale $Q^2 = 10,000$ GeV$^2$. Only experimental uncertainties are shown, evaluated with tolerance $T = 1$. Left: $x_s$. Right: $x_g$. 

---

Fig. 5 The PDF ratio $R_s = x(s + \bar{s})/x(\bar{u} + \bar{d})$ from ATLASpdf21 compared with $R_s$ for fits not including some of the $W, Z$ data sets. Only experimental uncertainties are shown, evaluated with tolerance $T = 1$. Left: not including $W, Z$ data at both 7 and 8 TeV. Right: not including $W, Z$ data at 8 TeV.
Fig. 6 Relative uncertainties in ATLASpdf21 $x_d$ and $x_g$ compared with fits not including some of the $W, Z$ data sets. Only experimental uncertainties are shown, evaluated with tolerance $T = 1$. Top: $x_d$ uncertainties, (left) not including inclusive $W, Z$ data at both 7 and 8 TeV, (right) not including inclusive $W, Z$ data at 8 TeV. Bottom: $x_g$ uncertainties, (left) not including inclusive $W, Z$ data at both 7 and 8 TeV, (right) not including inclusive $W, Z$ data at 8 TeV.

Determined reliably. Once $W, Z$ data at 7 TeV are input the determination improves considerably, but the inclusive $W, Z$ data at 8 TeV still add information.

In contrast, the valence and gluon PDFs are still reasonably well determined without any $W, Z$ data but the input of these data decreases their uncertainties significantly, as illustrated for the $x_d$ and $x_g$ PDFs on the left-hand side of Fig. 6.

On the right-hand side of Fig. 6 the decrease in the uncertainties of the $x_d$ and $x_g$ PDFs from removing only the $W, Z$ data taken at 8 TeV is illustrated, showing that the major decrease comes from retaining the $W, Z$ data taken at 7 TeV.

However, the $W, Z$ data taken at 8 TeV have a major role to play in ensuring that $x\bar{u} \sim x\bar{d}$ holds at low $x$, even though this constraint is not imposed. Without them, one observes $x\bar{d} < x\bar{u}$ at low $x$, as seen in Fig. 7. These data also somewhat reduce the low-$x$ strange distribution and harden the high-$x$ strange distribution, while softening the high-$x$ $x\bar{d}$ distribution, as also shown in Fig. 7.

There is mild tension between the $W, Z$ data at 8 TeV and the $W, Z$ data at 7 TeV. The partial $\chi^2$/NDP for the $W, Z$ data at 7 TeV decreases from 68/55 to 50/55 if the $W, Z$ data at 8 TeV are excluded from the fit, and the partial $\chi^2$/NDP for the $W, Z$ data at 8 TeV decreases from 239/206.
to 222/206 if the W, Z data at 7 TeV are excluded from the fit. These increases in $\chi^2$ are most pronounced for the 7 TeV c-c data around the Z mass-peak (66–116 GeV) and for the mass bins around the Z peak in 8 TeV data. As already remarked, theoretical scale uncertainties for W, Z data at both 7 and 8 TeV are added to the fit uncertainties. If these uncertainties are not added the tension between W, Z data at 7 and 8 TeV increases. The partial $\chi^2$/NDP for W, Z data at 7 TeV increases to 80/55 and the partial $\chi^2$/NDP for W, Z data at 8 TeV increases to 268/206 if both data sets are included in the fit and scale uncertainties are not applied. The differences between the PDFs are not large, whether or not scale uncertainties are applied, compared to the current experimental precision. However, consideration of such theoretical uncertainties is important if accuracy to 1% is the ultimate goal for the PDFs. N^3LO calculations indicate that higher-order corrections are likely to be larger than our current estimate of scale uncertainties. A further study of scale uncertainties is given in Appendix A.

5.2.2 Impact of V + jets data

The impact of the V+jets data is shown in Figs. 8 and 9. There are significant changes in the $xd$ and $xs$ PDF shapes such that the high-$x$ $xs$ and $xd$ PDFs become softer and harder, respectively, with the input of V+jets data. Because of the change in the $xs$ shape, the difference $x(\bar{d} - \bar{u})$ is also strongly affected. This is shown in Fig. 8. This figure also shows that, without the V+jets data, there is little information about the ratio $R_s$ at high $x$. The changes in Fig. 8 are large because the V+jets data resolve a double minimum in
parameter space such that the fit now prefers a hard $x\bar{d}$ and soft $x\bar{s}$ at high $x$, whereas it previously had an additional minimum with $x\bar{d} \sim x\bar{s}$ at high $x$, which was marginally preferred. These results are similar to those already seen and fully explained in the ATLASepWZVjets20 PDF analysis [11] and the double minimum may be seen in Figure 5 of that paper. A fit with only HERA data, or with HERA plus ATLAS $W, Z$ data at 7 TeV (ATLASepWZ20), prefers the minimum with $x\bar{d} \sim x\bar{s}$ at high $x$, but once ATLAS $V +$ jets data at 8 TeV are added to the fit the double minimum with a hard $x\bar{d}$ and soft $x\bar{s}$ at high $x$ is preferred and the previous minimum disappears. In Figure 3 of Ref. [11], it can also be seen that for ATLASepWZ20 (the fit without $V+ $jets data) the full uncertainties, including model and parameterisation variations, are very large because they cover both minima, whereas the full uncertainties of the ATLASepWZVjets20 fit (including $V +$ jets data) are much smaller because there is no double minimum. There is thus no strong inconsistency between the fits with or without the $V+$jets data because of the large uncertainties of the fit without the $V +$ jets data. This also applies in the present analysis, but in Fig. 8, only experimental uncertainties are shown. The full uncertainties for the fit without $V +$ jets data in our present analysis are of little interest and are not presented. Model and parametrisation uncertainties for the ATLASpdf21 fit will be discussed in Sect. 5.3.
These $V + \text{jets}$ data also effect a modest change in the $x_{d_{u}}$ shape and the gluon PDF shape as shown in Fig. 9. The $u$-type quarks are not strongly affected by these data and thus they are not shown. There is no tension between the $V + \text{jets}$ data at 8 TeV and other data sets in the fit, since all data are fitted well at the minimum chosen by these $V + \text{jets}$ data.

5.2.3 Impact of $t\bar{t}$ data

The impact of the $t\bar{t}$ data is shown in the top half of Fig. 10. The high-$x$ gluon distribution is mildly softened when the $t\bar{t}$ data are added to the fit. This effect is opposite to the one observed in the ATLASepWZtop18 fit. This is because more data which harden the gluon PDF, in particular the $V + \text{jets}$ and inclusive jet data, are included in the present fit. The more significant effect is in the uncertainties of the high-$x$ gluon distribution, which are reduced. There is no significant tension between the $t\bar{t}$ data and the other data in the fit. Figure 10 (bottom half) also shows the impact of removing only the $t\bar{t}$ data at 13 TeV (left) or only the $t\bar{t}$ data at 8 TeV (right). It is clear that the data at 8 TeV have the stronger impact on the shape of the $x_g$ PDF but both data sets contribute to a modest reduction in the uncertainties.
5.2.4 Impact of photon data and inclusive jet data

There is little impact from the addition of the direct-photon production ratio data apart from a marginal softening of the high- \( x \) gluon distribution as shown in Fig. 11 (left). However, it is notable that these data can now be well fitted at NNLO in QCD, given that they have been excluded from PDF fits for the last 20 years because of poor fits to lower-energy data [60, 75]. There is minimal tension with other data sets.

The principal impact of the inclusive jet data is on the gluon PDF. The main effect is a considerable decrease in high- \( x \) gluon uncertainties, with a mild hardening of the gluon PDF at high \( x \), as shown in Fig. 11 (right). There is minimal tension with other data sets. As explained earlier in Sect. 2, the central ATLASpdf21 fit includes only the inclusive jet data at 8 TeV, with \( R = 0.6 \). The full uncertainties of the fit, see Sect. 5.3.1, include the difference between the choices \( R = 0.6 \) and \( R = 0.4 \). The impact of using inclusive jet data at 7 TeV and at 13 TeV, instead of at 8 TeV, is explored in Appendix B.

5.3 Model, theoretical and parameterisation uncertainties

The consideration of additional uncertainties affecting the PDFs is presented in this section. These are classified and
pole masses, $m_t$ NNLO predictions for the $t$-value of the top-quark mass. The interpolation grids for the

For $m_t = 172.5$ GeV. The effect of this change in central value is negligible, as illustrated in Fig. 12. A cross-check was performed using the double differential $p_T^t$ and $m_t\bar{t}$ distributions at 13 TeV, for which predictions are available for $m_t = 171.0, 172.5, 174.0$ GeV and the effect was found to be negligible. The 13 TeV $t\bar{t}$ data have a smaller impact in the fit than the 8 TeV $t\bar{t}$ data, so only the $m_t$ variations of the 8 TeV data set are input to the final model uncertainty.

A further potential model uncertainty comes from the treatment of the jet systematic uncertainties. Alternative decorrelation scenarios are considered as follows: the alternative option in which the JES Flavour Response is split into only two components rather than three, called “Decorrelation scenario I”; complete decorrelation of the Jet Flavour Response between rapidity bins, called “FR decorrelated”; and no decorrelation, called “Fully correlated”. Table 5 gives the total $\chi^2$/NDP for the jets (including all three terms of Eq. (1)) for alternative correlation scenarios, showing that the difference between full correlations and various choices of decorrelation can have a significant effect on the $\chi^2$. However, the difference between the ATLASpdf21 gluon PDFs obtained using jet data with these different correlation scenarios is relatively small compared to the model uncertainties considered above, e.g. the variation of $Q^2_0$. There are also no changes in the PDF uncertainties as a result of using different correlation scenarios. Hence, these variations are not considered as a source of significant uncertainty.

A further source of uncertainty for the inclusive jet data comes from the choice of jet radius $R$. The difference between the PDFs due to a different choice of jet radius for the 8 TeV jets is shown in Fig. 12. The only significant change is in the gluon PDF. This small difference between the PDFs

labeled here as either model, theoretical or parameterisation uncertainties.

5.3.1 Model and theoretical uncertainties

The class of model uncertainties includes effects due to variations of the heavy-quark masses input to the TRVFN heavy-quark-mass scheme for the inclusive DIS calculations, the minimum $Q^2$ cut on the HERA inclusive DIS data and the value of the starting scale for evolution. The minimum $Q^2$ cut was varied in the range $7.5 < Q^2_{\text{min}} < 12.5$ GeV$^2$ and the starting scale was varied in the range $1.6 < Q^2_0 < 2.2$ GeV$^2$. In the inclusive DIS calculations, the heavy-quark masses were varied in the ranges $1.37 < m_c < 1.45$ GeV and $4.1 < m_b < 4.3$ GeV [71]. The variations of $m_c$ and $Q^2_0$ are coupled since the requirement $Q^2_0 < m_c^2$ must be met.

For this reason the upward variation of $m_c$ and the downward variation of $Q^2_0$ are symmetrised. Figure 12 illustrates the effect of these variations on the gluon distribution since this is the PDF most sensitive to these changes. The impact of the choice of $m_c$ and $m_b$ is modest. The effect of the variation of the $Q^2_{\text{min}}$ cut is larger but still within the experimental uncertainties. The largest of these uncertainties is due to the choice of $Q^2_0$, which gives a gluon PDF differing from the central fit by $\sim 2\sigma$ for $x \sim 0.1$.

An additional model uncertainty comes from the assumed value of the top-quark mass. The interpolation grids for the NNLO predictions for the $t\bar{t}$ data at 8 TeV are available for pole masses, $m_t = 172.5, 173.3, 175.0$ GeV. The smaller and larger values are used to estimate an asymmetric model uncertainty around the central value, shown in Fig. 12. For the $t\bar{t}$ 1-D distributions at 13 TeV the grids are only available

Fig. 11 ATLASpdf21 $xg$ PDF compared with $xg$ for fits not including various data sets. Only experimental uncertainties are shown, evaluated with tolerance $T = 1$. Left: not including the direct-photon production ratio data taken at 13 and 8 TeV. Right: not including inclusive jet data at 8 TeV.
Fig. 12 Impact of model assumption variations on the ATLASpdf21 gluon PDF illustrated in their ratio to the $xg$ distribution for the central choice. Uncertainties of the central fit are experimental, evaluated with tolerance $T = 1$. Top left: variation of $m_c$ and $m_b$ values. The $m_c$ upward variation is shown and this is symmetrised. Top right: impact of variations of $Q_{\text{min}}^2$ and $Q_0^2$ on the ATLASpdf21 gluon PDF. The $Q_0^2$ downward variation is shown and it is symmetrised. Bottom: impact of the variation of $m_t$ and of the choice of jet radius for the inclusive jets on the ATLASpdf21 gluon PDF.

Table 5 $\chi^2$ contributions for the inclusive jet data set at 8 TeV with $R = 0.6$, for different correlation scenarios, as explained in the text. The $\chi^2$ values given here represent the addition of all terms in Eq. (1).

| Jets 8 TeV $R = 0.6$ | Fully correlated | FR decorrelated | Decorrelation scenario 1 | Decorrelation scenario 2 (default) |
|----------------------|------------------|-----------------|---------------------------|-----------------------------------|
| $\chi^2$/NDP        | 289/171          | 227/171         | 250/171                   | 248/171                           |

extracted using jet data at 8 TeV for $R = 0.4$ and $R = 0.6$ is considered as an extra uncertainty because, although it is clear that the choice $R = 0.6$ is theoretically favoured [29], this paper also inputs $t\bar{t}$ data from the lepton+jets channel and $V^+$ jets data, and for these data sets only $R = 0.4$ jets are available.

The effect of using jet production data at different centre-of-mass energies of 7, 8 and 13 TeV is shown in Appendix B. This is not considered as an additional uncertainty but is presented as a cross-check that the choice of centre-of-mass energy for the jet production data does not have a significant influence on the PDFs extracted.
All model uncertainties are added in quadrature to form a total model uncertainty.

Theoretical uncertainties include the scale uncertainties of the predictions. The largest of these are the scale uncertainties for the inclusive W and Z data which are considered in detail in Sect. 3.2. Since these are treated as correlated systematic uncertainties in the fit \( \chi^2 \) they are actually already included in what has been labelled as the experimental uncertainty of the fit.

The effect of scale uncertainties for the V+jets data was studied in Ref. [11] and was found to be negligible. The effect of scale uncertainties for the \( \tau \bar{\tau} \) lepton + jets data was studied for the present paper and was also found to be negligible. The direct-photon production ratio data are relatively insensitive to scale variations because 100% correlation of the scale uncertainties is assumed between the 8 and 13 TeV data. Since the fit is relatively insensitive to these data, alternative assumptions for these correlations were not pursued.

The scale uncertainty for the inclusive jets at 8 TeV requires further consideration. These K-factors were supplied for two choices of scale: \( \mu_r = \mu_f = p_T^{\text{max}} \) or \( \mu_r = \mu_f = p_T^{\text{jet}} \). A scale related to \( p_T^{\text{jet}} \) is favoured over \( p_T^{\text{max}} \) [29], as already mentioned. However, this difference in scale choice is now investigated. A comparison of the gluon PDFs for these two scales is shown in Fig. 13 (left), since this is the PDF most sensitive to this change. It can be seen that there is no significant difference between the resulting gluon PDFs. Changes in the nuisance parameter values for the correlated systematic uncertainties absorb the change in the predictions for the two scales. The \( \chi^2 \) values for these fits are given in Table 6. The only significant change in \( \chi^2 \) comes from the inclusive jets, not from other data sets.

Next, scale variations \( \mu_r = \mu_f = 2p_T^{\text{jet}} \) and \( \mu_r = \mu_f = p_T^{\text{jet}}/2 \) are considered. The effect on the gluon PDF is shown in Fig. 13 (right). The \( \chi^2 \) values for these fits are given in Table 6. Again, the only significant change in \( \chi^2 \) comes from the inclusive jets, not from other data sets.

Finally, a cross-check is performed in which the K-factors are not smoothed but their statistical uncertainties are used as an additional uncorrelated uncertainty. The \( \chi^2 \) value for this fit is also given in Table 6. It is lower than that for smoothed K-factors because the statistical uncertainties of

---

**Table 6** \( \chi^2 / \text{NDF} \) for the total fit and the contribution \( \chi^2 / \text{NDF} \) for the inclusive jet data, for various scale choices, and treatment of K-factors and their uncertainties. The first row represents the default values used in the central ATLASpdf21 fit.

| Scale choice | \( \chi^2 / \text{NDF} \) for jets | Treatment of K-factors |
|--------------|----------------------------------|------------------------|
| \( p_T^{\text{jet}} \) scale | 2010/1620 248/171 | Smoothed |
| \( p_T^{\text{max}} \) scale | 2019/1620 257/171 | Smoothed |
| \( 2p_T^{\text{jet}} \) scale | 2022/1620 272/171 | Smoothed |
| \( p_T^{\text{jet}}/2 \) scale | 1991/1620 228/171 | Smoothed |
| Unsmoothed | 1983/1620 223/171 | |

Fig. 13 Ratios of the ATLASpdf21 gluon PDFs for different inclusive jet scale choices. Left: \( p_T^{\text{max}} \) over \( p_T^{\text{jet}} \) (default). Right: \( 2p_T^{\text{jet}} \) and \( p_T^{\text{jet}}/2 \) over \( p_T^{\text{jet}} \) (default). Uncertainties of the central fit are experimental, evaluated with tolerance \( T = 1 \).
the unsmoothed $K$-factors are $\sim 1\%$. However, the resulting PDFs are very similar to those obtained using smoothed $K$-factors. Since none of the scale variation effects produced significant changes in the PDFs, no further theoretical uncertainty is added for this source.

Thus the theoretical uncertainties considered so far are either already included in the experimental uncertainties of the fit, or they are negligible. It should be noted that the data are also sensitive to the value of $\alpha_s(m_Z)$, which affects the shape of the gluon PDF. The correlation between $\alpha_s(m_Z)$ and the gluon PDF shape is specified by the DGLAP formalism. A determination of $\alpha_s(m_Z)$ is beyond the scope of the current paper, since a correct NNLO treatment requires variation of the $K$-factor calculations with $\alpha_s(m_Z)$, or direct NNLO grids for all processes. This is left for future work. The conventional value $\alpha_s(m_Z) = 0.118$ is used, in line with the value used by the global fitting groups, CT [76], MSHT [77] and NNPDF [78].

5.3.2 Parameterisation uncertainties

The optimal number of parameters was determined by ‘saturation’ of the $\chi^2$ as explained in Sect. 4. However, the effect on the PDFs of adding extra parameters is investigated. Although there is no significant further decrease in $\chi^2$, some small shape changes are observed when adding an $F_u$ term to the $u$-valence PDF and/or a $D_{\bar{d}}$ term to the $x\bar{d}$ PDF. Figure 14 shows the impact of adding both of these as free parameters on the $xu$, $xd$, $\bar{x}u$ and $x\bar{d}$ PDFs, which are the PDFs which show the largest variations. The total

![Image](https://example.com/image.png)

**Fig. 14** Impact of adding $F_u$ and $D_{\bar{d}}$ as free parameters on the valence PDFs in comparison with the central ATLASpdf21 21-parameter fit. Uncertainties of the central fit are experimental, evaluated with tolerance $T = 1$. Top left: $xu$. Top right: $xd$. Bottom left: $\bar{x}u$. Bottom right: $x\bar{d}$
parameterisation uncertainty is the envelope of these $D$ and $F$ parameterisation variations. Model and parameterisation variations, and their effect on the total $\chi^2$ of the fit, are summarised in Table 7.

### 5.3.3 Imposing a maximum $Q^2$ cut on the input data

A further theoretical uncertainty stems from the possibility that subtle effects of BSM physics may be present in high-scale data. If these are included in a PDF fit, then the estimates of background to further higher-scale physics from new data could be distorted. A fit is performed in which the maximum scale for events accepted from each process is 500 GeV, or $Q^2 = 250,000$ GeV$^2$. This removes 82 points from the fit: 63 from the inclusive jets data, 6 from the $V+$jets data, 11 from the direct-photon ratio data, and one each from the $tt$, $p_T$ spectra at 8 and 13 TeV. The $\chi^2$/NDF value of the fit improves from 2010/1620 to 1870/1538, with the largest improvement coming from the inclusive jets, for which the $\chi^2$/NDF value improves from 248/171 to 136/108.

Figure 15 compares the PDFs at the scale of $Q^2 = 10,000$ GeV$^2$, with a linear scale in $x$ so as to emphasise the differences at high $x$. The shapes of the PDFs and their uncertainties are not strongly affected by the cut, and a maximum $Q^2$ cut is not considered further.

### 5.4 Combination of uncertainties

The PDFs are now presented with experimental uncertainties from the fit and with additional model and parameterisation uncertainties as discussed in this section. These model and parameterisation uncertainties are added in quadrature to the experimental uncertainty. It should be noted that the scale uncertainties for the inclusive $W$, $Z$ data appear as part of the experimental uncertainty, whereas the scale uncertainties for other data sets are negligible. Figure 16 shows the valence distributions and the $x\bar{u}$ and $x\bar{d}$ sea distributions, and Fig. 17 shows the $x\bar{s}$ and $xg$ distributions and the $(\bar{d} - \bar{u})$ distribution. The experimental uncertainties are much larger for the $d$-type sector than for the $u$-type sector. The larger fractional uncertainty of the $x\bar{s}$ PDF in comparison with the $xd$ PDF reflects its smaller absolute size as $x \to 1$. The gluon PDF is well determined for $0.01 < x < 0.3$ but not at low or high $x$. The model uncertainties in all PDFs are moderate, being largest in the gluon PDF. The parameterisation uncertainty of $xd$ affects both $xd$ and $xd_v$ since $xd_{sea} = xd$ and the sea and valence $d$-quarks act together. The $xd_v$ distribution is affected at both low and high $x$ because of the valence-quark number sum rule. Similarly, the parameterisation uncertainty of $xu_v$ is seen at both low and high $x$. The parameterisation uncertainty of the gluon PDF comes from the other PDFs through the momentum sum rule.

Figure 18 shows the $R_s$ distribution and its uncertainties as a function of $x$, at $Q^2 = 1.9$ GeV$^2$. Figure 19 shows the value of $R_s$, at $Q^2 = 1.9$ GeV$^2$ and $x = 0.023$ and at $Q^2 = m_Z^2$ and $x = 0.013$, compared with other PDFs. For the ATLASpdf21 fit, the value of $R_s$ at this low scale has decreased relative to previous ATLAS fits. The ATLASpdf21 central value of $R_s \sim 0.8$ is due firstly to the input of the $V+$jets data, as seen in the change from ATLASepWZVjets20 ($R_s \sim 1.15$) to ATLASepWZVjets20 ($R_s \sim 1.0$), secondly to the new freedom in the low-$x$ parameterisation, whereby the ATLASpdf21 central parameterisation corresponds to the low end of the uncertainty band of the ATLASepWZVjets20 result ($R_s \sim 0.85$), and thirdly to the input of the $W$, $Z$ data at 8 TeV. On the other hand, $R_s$ has increased for some of the other PDFs, e.g. MSHT20 [77] compared with MMHT14 [79], and CT18A [76] compared with CT14 [80]. This increase is due to the input of the ATLAS precision inclusive $W$, $Z$ data at 7 TeV. Thus, the recent evaluations of this ratio of strangeness to light quarks are now in fair agreement. However, the strangeness is still substantially larger at low $x$ than the older values which estimated $R_s \sim 0.5$. This illustration of the ratio $R_s$ at $x = 0.023$ should be supplemented by a comparison of the full shape of the $R_s$ distribution in Sect. 6.

| Central $\chi^2$/NDF | 2010/1620 |
|----------------------|-----------|
| **Model variations** |           |
| $Q^2_{min} = 12.5$ GeV$^2$ | 1947/1571 |
| $Q^2_{max} = 7.5$ GeV$^2$ | 2076/1660 |
| $m_c = 1.45$ GeV (sym) | 2025/1620 |
| $m_b = 1.6$ GeV$^2$ (sym) | 2018/1620 |
| $m_b = 4.3$ GeV | 2016/1620 |
| $m_b = 4.1$ GeV | 2014/1620 |
| $m_t = 175.0$ GeV | 2063/1620 |
| $m_t = 172.5$ GeV | 2018/1620 |
| $R = 0.4$ | 2080/1620 |
| **Parameter variations** |           |
| $F_{a, \alpha}$, $g_{\alpha}$ | 2007/1620 |
Fig. 15 Ratios of the PDFs from a fit in which a maximum scale cut of 500 GeV is imposed, to the central ATLASpdf21 at the scale $Q^2 = 10,000$ GeV$^2$. Both fits are shown with just experimental uncertainties, evaluated with tolerance $T = 1$. Top left: $xu_v$. Top right: $xd_v$. Middle left: $x\bar{u}$. Middle right: $xd$. Bottom left: $xs$. Bottom right: $xg$. Here the $x$-scale is linear.
6 Consideration of $\chi^2$ tolerance and comparison with global fits

6.1 $\chi^2$ tolerance

The experimental uncertainties presented for the fitted PDFs were evaluated by the usual criterion of parameter setting, $\Delta \chi^2 = 1$. However, global PDF fitting groups such as CT and MSHT use enhanced $\chi^2$ tolerances, $\Delta \chi^2 = T^2$, where $T^2$ is a dynamic tolerance ranging from 10 to 100.\textsuperscript{10} The justification for this comes from consideration of the consistency, or relative inconsistency, of input data sets [84]. In the present case, whereas the HERA data is a single consistent set of data with common correlated uncertainties, the ATLAS data are more disparate and a study to find an appropriate tolerance was performed.

One way to estimate a suitable tolerance for a fit including the ATLAS data is to consider the dynamic tolerance procedure of the MSHT group, as first introduced in the MSTW paper [84]. Briefly restated, the method is as follows. The inverse of the Hessian matrix of the fit is diagonalised to give the eigenvector combinations of the parameters of the fit. The eigenvalues are the squares of the uncertainties in these eigenvector combinations of the parameters. The parameter displacements from the global minimum can then be expanded in a basis of rescaled eigenvectors,

$$e_{ik} = \sqrt{\lambda_k} v_{ik},$$

where $\lambda_k$ is the $k$-th eigenvalue and $v_{ik}$ is the $i$-th component of the $k$-th orthonormal eigenvector. Pairs of eigenvector PDFs have parameters given by

$$a_i(k^\pm) = a_i^0 \pm t e_{ik},$$

where $a_i^0$ are the central parameters, and $t = 1$ should give the usual change in $\chi^2$ of $T^2 = \Delta \chi^2 = 1$. In the quadratic approximation, $t = T$, and this is reasonably well obeyed for the present ATLASpdf21 21-parameter fit, at least up to $t = 5$.

Having determined the eigenvectors, each of these is treated as a different ‘hypothesis’ for a theory which can fit the data, and each of them is checked to see if it can describe each of the data sets within the hypothesis-testing criterion, $\chi^2_n \lesssim N + \sqrt{2N}$, as $t$ is increased and decreased along the eigenvector, where $\chi^2_n$ is the $\chi^2$ for a particular data set and $N$ is the number of data points for that set. As soon as $\chi^2_n$ is

\textsuperscript{10} NNPDF uses a completely different method for estimation of PDF uncertainties, but the magnitude of their uncertainties is comparable to those of MSHT and CT.
goes above this limit, this sets the maximum value of \( t \), and hence of \( T \), that can be tolerated for this eigenvector in this direction, for this data set. Each eigenvector will have a most constraining data set, for each direction, and in principle the values of the tolerances \( T \) for the most constraining data set can be different for each eigenvector and each direction, hence the term 'dynamic'.

This method is made more sophisticated in a number of ways. Firstly, the hypothesis-testing criterion chosen is not exactly the simple \( \chi^2_n < N + \sqrt{2N} \), but is \( \chi^2_n < \xi_{68} \), where \( \xi_{68} \) is given by \( \int_{\xi_{68}}^{0} d\chi^2 P_N(\chi^2) = 0.68 \) and \( P_N(\chi^2) \) is the usual probability density function for the \( \chi^2 \) distribution with \( N \) degrees of freedom. Thus each data set should be described within its 68% confidence level. Secondly, the criterion imposed is not actually \( \chi^2_n < \xi_{68} \), but \( \chi^2_n < (\chi^2_{n,0}/\xi_{50}) \xi_{68} \), to allow for the fact that the central fit \( \chi^2_{n,0} \) for the data set may not be perfect, where perfect would imply \( \chi^2_{n,0} = \xi_{50} \).

For the ATLASpdf21 fit the most constraining data sets for the eigenvectors are the \( W, Z \) data at 7 TeV or the \( Z/\gamma^* \) triple differential cross-section data at 8 TeV (and very occa-
These values are all fairly similar, so a single value can be used and, since the tolerance should not be set too large whereby a data set is far outside its 68% CL limit, this suggests the simple choice of $T = 3$.

In the above procedure to evaluate the dynamic tolerance, the fact that the data sets are not fitted perfectly was taken into account. One may estimate how much of the increased tolerance comes from this part of the procedure by increasing the uncertainties of each data set by $\chi^2/\text{NDP}$ for those data, such that $\chi^2/\text{NDP} \sim 1$ is achieved in the fit. This will clearly increase the uncertainties of the output PDFs. However, a naive application of this procedure would increase all the experimental uncertainties and this seems somewhat unnecessary for the statistical uncertainties, which are well known. Hence, only the systematic uncertainties are increased. Starting with the $\chi^2/\text{NDP}$ scaling factors, the fit is iterated, adjusting these factors at each iteration, until a $\chi^2/\text{NDP} \sim 1$ for each data set is achieved. After three iterations there is no longer any significant change and the scaling factors range from 1.0 to 1.6.

It is non-trivial that for these reweightings of the data the PDF shapes hardly differ from those of an unweighted fit. This shows that the fit is stable for modest reweighting of the input data sets. The uncertainties of the PDFs have increased, but only by a factor which corresponds to a tolerance of up to $T \sim 1.5$. Thus the greater part of the dynamic tolerance estimate of $T = 3$ comes from the application of the hypothesis-testing criterion to the $\chi^2$ for the eigenvectors of the fit.

The tolerance $T = 3$ is applied for the final estimation of PDF uncertainty; however, ATLASpdf21 PDFs with both choices of tolerance, $T = 1$ and $T = 3$, will be made available on the LHAPDF PDF repository [85]. Having fixed the tolerance for the experimental uncertainties of the PDFs, the full uncertainties, including experimental, model and parametrisation uncertainties are compared for the choices $T = 3$ and $T = 1$ in Fig. 20. When an enhanced tolerance such as this is applied to the experimental uncertainties, the model and parametrisation uncertainties are far less significant.

Comparisons of the differential cross-section measurements used as input to the fit with the predictions of ATLASpdf21 fit, with $T = 3$ and full uncertainties, are shown in Appendix C.

6.2 Comparison of ATLASpdf21 with global PDFs

In Figs. 21, 22 and 23, the ATLASpdf21 PDFs are compared with the HERAPDF2.0 [5], CT18, CT18A [76], NNPDF3.1 [78], MSHT20 [77] and ABMP16 [81] PDFs. All figures in this section show ATLASpdf21 PDFs with full uncertainties, such that experimental uncertainties are evaluated with $T = 3$ and model and parametrisation uncertainties are also included. It is observed that the addition of the ATLAS data sets to the HERA data brings the PDFs much closer to those of the global fits than to HERAPDF2.0, particularly for $x d_v$, $x \bar{d}$ and $x \bar{u}$. The $x u_v$ and $x d_v$ distributions of the ATLASpdf21 fit are in good agreement with CT18, CT18A, MSHT20 and ABMP16, whereas NNPDF3.1 is an outlier. The $x \bar{d}$ distribution agrees reasonably well with all the other PDFs. The $x \bar{d}$ distribution is in fair agreement with CT, MSHT and NNPDF, with ABMP presenting as somewhat of an outlier. However, it should be noted that $x \bar{d}$ from the ATLASpdf21 fit is somewhat higher at high $x$. This is further illustrated by the $x (\bar{d} - \bar{s})$ distribution, which changes from negative at high $x$ when fitting only HERA data in HERAPDF2.0, to positive at high $x$ for ATLASpdf21, in agreement with the global PDFs. However, the ATLASpdf21

11 Note that ABMP has much smaller uncertainties due to their use of tolerance $T = 1$. 
Fig. 20 ATLASpdf21 fits comparing the full uncertainties (experimental, model, parameterisation) for tolerance values \( T = 1 \) and \( T = 3 \). Top left: \( xu_v \). Top right: \( xd_v \). Middle left: \( x\bar{u} \). Middle right: \( x\bar{d} \). Bottom left: \( x\bar{s} \). Bottom right: \( xg \).
Fig. 21 ATLASpdf21 $xu_v$ and $xd_v$ distributions with full uncertainties (experimental $T = 3$, model, parameterisation) compared with other PDFs. Left: CT18, CT18A, HERAPDF2.0. Right: MSHT20, NNPDF3.1, ABMP16

The $x\bar{d} - \bar{u}$ distribution is more positive than the global PDFs for $x > 0.3$, in better agreement with the newer E906 Drell–Yan data [86], rather than the E866 data which is in the global fits. This is explored further in Appendix D. The central value of the $x\bar{s}$ distribution is larger than those in NNPDF3.1 and CT18 for $x \sim 0.02$, but is in good agreement with CT18A and MSHT20. The gluon distribution agrees best with NNPDF3.1, but is in reasonable agreement with the other global PDFs. The distributions of $R_g$ for all the PDFs illustrated are now in broad agreement but there are differences in detail. The uncertainties of HERAPDF2.0 and CT18 are larger than the others because they do not include ATLAS

inclusive $W, Z$ data at 7 TeV. The uncertainties of ABMP16 are smaller for the usual reason that they use tolerance $T = 1$. The input of ATLAS inclusive $W, Z$ data at 7 TeV is the only difference between the CT18 and CT18A analyses. After input of these data the CT18A $R_g$ PDF ratio has smaller uncertainties and a larger central value of $R_g$ at low $x$ than CT18. CT18A is in good agreement with ATLASpdf21, for both central value and uncertainty, over the full $x$ range illustrated. The PDF analyses of MSHT20 and NNPDF3.1 also use ATLAS inclusive $W, Z$ data at 7 TeV and have a similar size of uncertainty and level of agreement with ATLASpdf21. In addition, the MSHT20 analysis uses ATLAS inclusive $W, Z$ data at 8 TeV, just as the ATLASpdf21 analysis does. Note that NNPDF have updated their study of strangeness

12 The ABMP16 gluon PDF corresponds to $\alpha_s(m_Z) = 0.1145$, a lower value than in the other PDFs illustrated, which all use $\alpha_s(m_Z) = 0.118$.  

13 In the case of HERAPDF2.0 the uncertainty in the strange PDF is not measured but estimated.
Fig. 22 ATLASpdf21 $x\bar{u}$, $xd$ and $x(d - u)$ distributions with full uncertainties (experimental $T = 3$, model, parameterisation) compared with other PDFs. Left: CT18, CT18A, HERAPDF2.0. Right: MSHT20, NNPDF3.1, ABMP16
Fig. 23 ATLASpdf21 $x\bar{s}$, $xg$ and $R_s$ distributions with full uncertainties (experimental $T = 3$, model, parameterisation) compared with other PDFs. Left: CT18, CT18A, HERAPDF2.0. Right: MSHT20, NNPDF3.1, ABM16
recently to NNPDF3.1_strange \cite{83}, which has a somewhat larger strangeness at low $x$ than NNPDF3.1. The modern PDFs which use ATLAS inclusive $W$, $Z$ data, including ATLASpdf21 itself, all agree that strangeness is not suppressed strongly at low $x$, but there is substantial suppression at high $x$.

In summary, the ATLASpdf21 fit now agrees with the global fits as well as they agree with each other. Thus ATLAS data seem able to replicate most of the features that the fixed-target DIS and DY data plus the Tevatron data brought to the global PDFs, see Appendix D. Using only the HERA and ATLAS data allows a more rigorous treatment of correlated systematic uncertainties and, in particular, of correlations between data sets.

For the global fits considered here, the $\chi^2$ values for the data sets included in the ATLASpdf21 fit are HERAPDF2.0: 2262, CT18: 2135, CT18A: 2133, MSHT20: 2218, NNPDF3.1: 2109, compared with ATLASpdf21: 2010. Although the global PDFs from CT, MSHT and NNPDF have more flexible parameterisations, the $\chi^2$ value of the ATLASpdf21 fit is better for the data sets considered. This is a further indication that the ATLASpdf21 parameterisation is sufficiently flexible.

Figure 24 shows the uncertainties of the ATLASpdf21 fit, with the enhanced tolerance $T = 3$, compared with CT18A and MSHT20 PDFs. These two PDFs are chosen because they both use the ATLAS $W$, $Z$ inclusive cross-section data at 7 TeV and they both evaluate uncertainties in a similar way to the present analysis. In general, the uncertainties of CT18A are somewhat larger than those of MSHT20, even though both represent 68\% CL uncertainties. For the $x\bar{u}$ and $x\bar{d}$ sea distributions at low $x$, the uncertainties of ATLASpdf21 lie between those of CT18A and MSHT20, becoming larger at high $x$, $x > 0.1$, reflecting the absence of fixed-target DY data. For the $x\bar{s}$ sea distribution there is a similar pattern, with the low $x$ uncertainties of the ATLAS PDF being comparable to those of CT18A. The uncertainty of the ATLASpdf21 gluon distribution at $x < 2 \times 10^{-3}$ is very similar to that of both CT18A and MSHT20, but the high $x$ gluon uncertainty is somewhat larger for the ATLASpdf21 fit, reflecting the absence of Tevatron jet data. In the valence sector, the $xu_v$ uncertainties of the ATLASpdf21 fit are comparable to those of CT18A at low $x$, but larger at high $x$, reflecting the absence of Tevatron $W$, $Z$ data. For $xd_v$, the uncertainties are larger for all $x$ except $x \sim 0.003–0.008$, reflecting the fact that there is less information about the $d$-quark than the $u$-quark in the absence of deuterium target data. Appendix D shows comparisons between the ATLASpdf21 predictions and some of these older data sets which, while not fitted, are generally well described.

In order to focus more on the high $x$ regime, which is important in searches for new physics, the comparisons between the ATLASpdf21 fit, with the enhanced tolerance of $T = 3$, and the CT18A and MSHT20 PDFs, are repeated with a linear-$x$ and log-$y$ scale in Fig. 25. Some discrepancies, not evident before, appear for $x \geq 0.7$. In this high $x$ region there is no data to determine any of the PDFs. Nevertheless, there is agreement within $\sim 2\sigma$. 
Fig. 24 ATLASpdf21 fit with full uncertainties (experimental $T = 3$, model, parameterisation) compared with CT18A, MSHT20 shown as a ratio with each distribution centred on unity. Top left: $x u_v$. Top right: $x d_v$. Middle left: $x \bar{u}$. Middle right: $x \bar{d}$. Bottom left: $x g$. Bottom right: $x \bar{s}$.
Fig. 25 ATLASpdf21 fit with full uncertainties (experimental $T = 3$, model, parameterisation) compared with global PDFs: CT18A and MSHT20, focusing on high $x$ behaviour. Top left: $xu$. Top right: $xd$. Middle left: $x\bar{u}$. Middle right: $x\bar{d}$. Bottom left: $x\bar{s}$. Bottom right: $xg$.
7 Conclusion

A PDF fit is presented using selected proton–proton collision data sets recorded by the ATLAS experiment at the LHC together with the final combined inclusive HERA data in order to assess the impact of the ATLAS data with a minimum of other input data. The resulting PDF set is called ATLASpdf21. The addition of ATLAS inclusive W and Z data to HERA data allows a fit to be made without any constraints imposed on the relationship between $x\bar{u}$, $x\bar{d}$ and $xx$, nevertheless, $x\bar{d} \sim xx$ is observed at low $x$. The strangeness at low $x$ is found not to be as suppressed as found in earlier PDFs such as CT14, MMHT14 and NNPDF3.0, but to be in agreement with modern PDFs such as CT18A, MSHT20 and NNPDF3.1_strange which use these ATLAS inclusive $W$, $Z$ data. This increase in the low $x$ strange was first seen in the ATLASepWZ12 and ATLASepWZ16 PDF analyses, but is now somewhat moderated by further ATLAS data, and the freedom of the parametrisation at low $x$. These inclusive $W$, $Z$ data also exhibit sensitivity to the valence quark distributions and the gluon distribution, considerably reducing their uncertainty. The ATLAS $V+\text{jets}$ data constrain the light-quark sea at higher $x$, such that strangeness is strongly suppressed at high $x$, and impact the high $x$ gluon distribution. The $t\bar{t}$ data serve to moderately reduce the uncertainty in the high $x$ gluon distribution. Ratios of direct photon production at different proton–proton collision energies have only a mild impact on the gluon PDF, but it is notable that they may now be reliably fitted to NNLO in QCD. The jet production data are sensitive to the gluon distribution at medium to high $x$ and they reduce its uncertainty considerably.

The role of scale uncertainties is considered for all data sets and these theoretical uncertainties are implemented in the fit for the inclusive $W$, $Z$ data, where they are most impactful. The effects of such scale uncertainties are small. A study is made of excluding high-scale data (with a scale above 500 GeV), which may contain subtle effects of physics beyond the Standard Model, from the fit. The resulting PDFs are not strongly affected by this restriction of the fitted kinematic region.

It is observed that the addition of the ATLAS data sets to the HERA data brings the PDFs much closer to the global PDFs of MSHT, CT and NNPDF than to HERAPDF2.0. The ATLASpdf21 PDFs agree with these global fits as well as they agree with each other. Thus, ATLAS data seem to be able to replicate many of the features that the fixed-target deep inelastic scattering and Drell–Yan data plus the Tevatron Drell–Yan data bring to the global PDFs. Using only the HERA and ATLAS data allows a detailed treatment of correlated systematic uncertainties.

Correlations of systematic uncertainties within and between ATLAS data sets are considered, with particular emphasis on the larger systematic uncertainties appertaining to the jet energy scale. Information about correlations between data sets is provided such that the major correlations could be considered by the global fitting groups CT, MSHT and NNPDF in future fits. The effects of these correlations are relatively small, $O(\sim 1\%)$, at the scale $Q^2 \sim 10,000\text{GeV}^2$ considered for precision physics at the LHC, but large enough to be considered in future precision data analyses such as the measurement of $m_W$ and $\sin^2\theta_W$, given that such a level of accuracy is now desired to resolve effects of physics beyond the Standard Model in the measurements of Standard Model parameters.

Acknowledgements We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently. We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSSI, Greece; RGC and Hong Kong SAR, China; IFI and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MEIN, Poland; FCT, Portugal; MNE/IFA, Romania; JINR; MES of Russia and NRC KI, Russian Federation; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DITP, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; COST, ERC, ERDF, Horizon 2020 and Marie Sklodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; Norwegian Financial Mechanism 2014–2021, Norway; NCN and NAWA, Poland; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [87].

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors’ comment: All ATLAS scientific output is published in journals, and preliminary results are made available in Conference Notes. All are openly available, without restriction on use by external parties beyond copyright law and the standard conditions agreed by CERN. Data associated with journal publications...]
are also made available: tables and data from plots (e.g. cross section values, likelihood profiles, selection efficiencies, cross section limits, ...) are stored in appropriate repositories such as HEPDATA (http://hepdata.cedar.ac.uk/). ATLAS also strives to make additional material related to the paper available that allows a reinterpretation of the data in the context of new theoretical models. For example, an extended encapsulation of the analysis is often provided for measurements in the framework of RIVET (http://rivet.hepforge.org/). This information is taken from the ATLAS Data Access Policy, which is a public document that can be downloaded from http://opendata.cern.ch/record/413 [opendata.cern.ch].

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.
Funded by SCOAP3.

Appendix

A Scale uncertainties in W, Z inclusive data

In this appendix, different choices for the treatment of the scale uncertainties in inclusive W, Z data are considered. For these data sets the experimental uncertainties are comparable to the scale uncertainties of NNLO predictions (~0.5%). These theoretical uncertainties can therefore have some significant impact on the PDF fit. For the central fit, these theoretical uncertainties are applied in the same way as correlated systematic uncertainties. Changes in renormalisation and factorisation scales are taken to be correlated between W and Z data and to be correlated between the 7 and 8 TeV data sets. Two alternatives are considered here: (i) the scale uncertainties are not correlated between the 7 and 8 TeV data sets and (ii) scale uncertainties are not applied at all.

The results of fits for these two cases, compared with the central fit, are shown as a ratio at a scale $Q^2 = 10,000$ GeV$^2$, relevant for LHC physics, in Fig. 26. If should be noted that the central fit is shown with just experimental uncertainties, evaluated with tolerance $T = 1$, so that its uncertainties are directly comparable with those of the fit without scale uncertainties. The uncertainties are very similar in size. The differences between the shapes of the PDFs are not large, but they can be important if the desired accuracy of the PDFs is $O(\sim 1\%)$. The difference between the cases where the scale uncertainties are applied as being correlated or uncorrelated between the 7 and 8 TeV inclusive W, Z data sets is shown by the green line in these figures and it can be seen that it is generally a smaller effect.
Fig. 26 ATLASpdf21, showing the ratios of a fit not including theoretical scale uncertainties in the inclusive $W, Z$ data to the central fit which does include these uncertainties, at the scale $Q^2 = 10,000$ GeV$^2$. Also shown is the result of a fit in which the scale uncertainties are applied but not correlated between the 7 and 8 TeV data. All three fits are shown with just experimental uncertainties, evaluated with tolerance $T = 1$. Top left: $xu$, Top right: $xd$. Middle left: $x\bar{u}$. Middle right: $x\bar{d}$. Bottom left: $xs$. Bottom right: $xg$. 
Fig. 27 Comparison showing the ratio of the ATLASpdf21 gluon PDF and the $R_s$ PDF ratio, using inclusive jet data at 8 TeV with $R = 0.6$, to the gluon PDF and $R_s$ PDF ratio for fits using various different jet production data sets at 7, 8 and 13 TeV, with differing choices of jet radius. The data sets in these plots are for scale choice $p_T^{\text{jet}}$. Uncertainties of the central fit are full uncertainties: experimental, evaluated with tolerance $T = 3$, plus model and parameterisation uncertainties.

**B Comparison of the impact of inclusive jet data at different centre-of-mass energies**

It is not possible to fit inclusive jet production data at different centre-of-mass energies simultaneously, because the full experimental systematic uncertainty correlations between these data sets are not known. The inclusive jet production data at 8 TeV with $R = 0.6$ were selected for input to the central fit. In this appendix, fits using the inclusive jet production data at 7 or 13 TeV instead of the data at 8 TeV are compared with the central fit. The data at 7 TeV shown here were extracted for $R = 0.6$ whereas the data at 7 TeV were extracted only for $R = 0.4$. The gluon PDF and $R_s$ PDF ratio using these jet production data sets at 7 and 13 TeV are shown in their ratio to the central fit results in Fig. 27. The scale choice was $p_T^{\text{jet}}$ for all the jet production data sets included in this figure. Since the effect of using $R = 0.4$ or $R = 0.6$ is very similar, as illustrated for the jet production data at 8 TeV, the differences between the PDFs are dominated by the change in centre-of-mass energy. The PDFs using the jet production data at 8 TeV lie between those of the jet production data at 7 TeV and the jet production data at 13 TeV. However, these differences are not significant compared to the full uncertainties of the PDFs evaluated with $T = 3$.

**C Goodness of the fit: comparison with data sets included in the fit**

Figures 28, 29, 30, 31, 32, 33, 34, 35, 36 and 37 show comparisons of the various ATLAS differential cross-section measurements used in the ATLASpdf21 fit, together with the predictions of this fit. Further details are provided in the figure captions.
Fig. 28 The differential cross-section measurements of (left) $W^-$ and (right) $W^+$ bosons in Ref. [9] (black points) as a function of their absolute rapidity, $|y_W|$. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T = 3$) + model + parameterisation) of the fit prediction.
Fig. 29 The differential cross section measurements of $Z$ bosons for (left) central-central and (right) central-forward events in Ref. [9] (black points) as a function of their absolute rapidity, $|y_z|$. The top row shows differential cross sections in the $66 < m_{\ell\ell} < 116$ GeV mass range, while the bottom row shows the $116 < m_{\ell\ell} < 150$ GeV mass range. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T = 3$) + model + parameterisation) of the fit prediction. It is interesting to observe that the N$^3$LO cross section for $Z$ bosons is expected to be $\sim 2\%$ lower than the NNLO cross section [74], which would bring the data into agreement with theory without need of shifts of systematic uncertainties.
Fig. 30 The triple differential cross-section measurements of $Z/\gamma^*$ at 8 TeV in Ref. [13] (black points), in the central-central rapidity region of the dilepton pair. The six plots show data in bins of the dilepton mass $m_{\ell\ell}$ (66–80–91–102–116–150–200 GeV). Within each plot the absolute rapidity increases from left to right in 12 steps of 0.2, ranging from 0.0 to 2.4, although the full rapidity range is not accessed for every mass region. Nearby data points show variation with the Collins–Soper angle at the same rapidity. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T=3$) + model + parameterisation) of the fit prediction. It is interesting to observe that the N$^3$LO cross section for $Z$ bosons is expected to be $\sim$2% lower than the NNLO cross section [74], which would bring data into agreement with theory without need of shifts of systematic uncertainties. The $Z$ mass-peak data are in the top right and middle left plots.
Fig. 31 The differential cross-section measurements of (left) \( W^- \) and (right) \( W^+ \) bosons at 8 TeV in Ref. [12] (black points) as a function of their absolute rapidity, |\( y_W \)|. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the \( b_j \) parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the \( \chi^2 \). The red band represents the full uncertainty (experimental (evaluated with \( T = 3 \) ) + model + parameterisation) of the fit prediction.

Fig. 32 The differential cross-section measurements of (left) \( W^- + \) jets and (right) \( W^+ + \) jets at 8 TeV in Ref. [24] (black points) as a function of the transverse momentum of the W boson, \( p_T^W \). The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the \( b_j \) parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the \( \chi^2 \). The red band represents the full uncertainty (experimental (evaluated with \( T = 3 \) ) + model + parameterisation) of the fit prediction.
Fig. 33 The differential cross-section measurements of $Z + \text{jets}$ at 8 TeV in Ref. [25] (black points) as a function of the absolute rapidity of inclusive jets, $|y|$, in bins of $p_{T}^{\text{jet}}$, the transverse momentum of the inclusive jets. Top left: $25 < p_{T}^{\text{jet}} < 50$ GeV. Top right: $50 < p_{T}^{\text{jet}} < 100$ GeV. Middle left: $100 < p_{T}^{\text{jet}} < 200$ GeV. Middle right: $200 < p_{T}^{\text{jet}} < 300$ GeV. Bottom left: $300 < p_{T}^{\text{jet}} < 400$ GeV. Bottom right: $400 < p_{T}^{\text{jet}} < 1050$ GeV. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_{j}$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T = 3$) + model + parameterisation) of the fit prediction.
Fig. 34  Top: the differential cross-section measurements of $t\bar{t}$ at 8 TeV as a function of (left) the average top momentum, $p_t$, and (right) the invariant mass of the $t\bar{t}$ system, $m_{t\bar{t}}$, in Ref. [26] (black points) in the lepton + jets decay channel. Bottom: the differential cross-section measurements of $t\bar{t}$ at 8 TeV as a function of the absolute rapidity of the $t\bar{t}$ pair, $|y_{t\bar{t}}|$, in Ref. [27] (black points) in the dilepton decay channel. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band.

The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T=3$) + model + parameterisation) of the fit prediction.
Fig. 35 The differential cross-section measurements of $t\bar{t}$ at 13 TeV in Ref. [15] (black points) as functions of the (top left) average top transverse momentum $p_T$, (top right) invariant mass of the $t\bar{t}$ pair $m_{t\bar{t}}$, (bottom left) average top absolute rapidity $|y_T|$ and (bottom right) absolute boosted rapidity of the $t\bar{t}$ pair, $|y^b_{t\bar{t}}|$. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T = 3$) + model + parameterisation) of the fit prediction.
Fig. 36 The ratios of the cross sections for inclusive isolated-photon production at 8 and 13 TeV in Ref. [14] (black points) as functions of the photon transverse energy, $E_T^\gamma$, for $E_T^\gamma > 125$ GeV in bins of photon absolute pseudorapidity, $|\eta^\gamma|$. Top left: $|\eta^\gamma| < 0.6$. Top right: $0.6 < |\eta^\gamma| < 1.37$. Bottom left: $1.56 < |\eta^\gamma| < 1.81$. Bottom right: $1.81 < |\eta^\gamma| < 2.37$. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T = 3$) + model + parameterisation) of the fit prediction.
Fig. 37 The differential cross-section measurements of inclusive jet production at 8 TeV in Ref. [17] (black points), for $R = 0.6$, as a function of the jet $p_T^\text{jet}$, in six bins of absolute rapidity, $|y^{\text{jet}}|$. Top left: $|y^{\text{jet}}| < 0.5$. Top right: $0.5 < |y^{\text{jet}}| < 1.0$. Middle left: $1.0 < |y^{\text{jet}}| < 1.5$. Middle right: $1.5 < |y^{\text{jet}}| < 2.0$. Bottom left: $2.0 < |y^{\text{jet}}| < 2.5$. Bottom right: $2.5 < |y^{\text{jet}}| < 3.0$. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed for the scale choice $p_T^\text{jet}$, with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T = 3$) + model + parameterisation) of the fit prediction.
Fig. 38 The differential cross-section measurements of (left) $Z$ and (right) $W$ bosons in Refs. [88,89] (black points) as a function of the absolute rapidity of the boson, $|y_Z|$ or $|y_W|$. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T = 3$) + model + parameterisation) of the fit prediction.

D Comparison with extra data sets not included in the fit

The ATLASpdf21 fit does not include data from the Tevatron or from fixed-target DY data. In this appendix the predictions of the ATLASpdf21 fit for some of these data sets, which were found to be most impactful in the global fits, are explored and found to be satisfactory.

In Figs. 38 and 39 the ATLASpdf21 fit is shown in comparison with Tevatron $W$ and $Z$ data.

The $\chi^2$/NDF values for the CDF data are 31/28 for the $Z$ data and 35/13 for the $W$-asymmetry data.

The $\chi^2$/NDF values for the D0 data are 23/28 for the $Z$ data, 25/13 for the $W$-electron asymmetry data and 13/10 for the $W$-muon asymmetry data. The ATLASpdf21 fit therefore provides a fair description, $\chi^2$/NDF = 126/92, of these Tevatron data, which mostly influence the high-$x$ valence quarks.
Fig. 39 The differential cross-section measurements of (left) $Z$, (right) $W$ (electron decay channel) and (bottom) $W$ (muon decay channel) in Refs. [90–92] (black points) as a function of the absolute rapidity of the $Z$ boson or absolute pseudorapidity of the decay lepton. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T=3$) + model + parameterisation) of the fit prediction.

It is also interesting to consider the description of the fixed-target Drell–Yan data from E866 and E906, since these are uniquely able to constrain the difference $x(\bar{d} - \bar{u})$ at high $x$. Figure 40 compares the predictions of the ATLASpdf21 fit with the E866 $pD/pp$ data [93] in three mass regions. The $\chi^2$/NDF values are 9.6/10, 14/14 and 21/15 for the low-, intermediate- and high-mass regions, respectively. Thus, the description of the E866 data, which mostly give information about high-$x$ sea quarks, is good. However, the high-mass region, which covers larger Bjorken-$x$, is not as well fitted as the lower-mass regions. The ATLASpdf21 fit is in better agreement with the new data from E906 [86] in the high-mass region, as seen in Fig. 41, which compares ATLASpdf21 and other PDFs with the $x\bar{d}/x\bar{u}$ ratios extracted from E866 and E906.
Fig. 40 $\sigma^{D}/2\sigma^{PP}$ from Ref. [93] (black points) in the (top left) low-mass, (top right) intermediate-mass and (bottom) high-mass regions as a function of $x_2$. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared with the predictions computed with the PDFs resulting from the ATLASpdf21 fit. The solid line shows the predictions without shifts of the systematic uncertainties, while for the dashed line the $b_j$ parameters associated with the experimental systematic uncertainties as shown in Eq. (1) are allowed to vary to minimise the $\chi^2$. The red band represents the full uncertainty (experimental (evaluated with $T=3$) + model + parameterisation) of the fit prediction.
References

1. P. Azzi et al., Report from Working Group 1: Standard Model Physics at the HL-LHC and HE-LHC. CERN Yellow Rep. Monogr. 7, 1 (2019). https://doi.org/10.23731/CYRM-2019-007.
2. M. Cepeda et al., Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC. CERN Yellow Rep. Monogr. 7, 221 (2019). https://doi.org/10.23731/CYRM-2019-007.221.
3. J. Gao, L. Harland-Lang, J. Rojo, The structure of the proton in the LHC precision era. Phys. Rep. 742, 1 (2018). https://doi.org/10.1016/j.physrep.2018.03.002.
4. J.J. Ethier, E.R. Nocera, Parton distributions in nucleons and nuclei. Annu. Rev. Nucl. Part. Sci. 70, 43 (2020). https://doi.org/10.1146/annurev-nucl-011720-042725.
5. A.M. Cooper-Sarkar et al., Measurement of the longitudinal structure function and the small x gluon density of the proton. Z. Phys. C 39, 281 (1988). https://doi.org/10.1007/BF01551005.
6. H1 Collaboration, Measurement of inclusive ep cross sections at high $Q^2$ at $\sqrt{s} = 225$ and 252 GeV and of the longitudinal proton structure function $F_2$ at HERA. Eur. Phys. J. C 74, 2814 (2014). https://doi.org/10.1140/epjc/s10052-014-2814-6.
7. ATLAS Collaboration, Determination of the parton distribution functions of the proton from ATLAS measurements of differential W and Z boson cross sections. Eur. Phys. J. C 79, 760 (2019). https://doi.org/10.1140/epjc/s10052-019-7199-0.
8. ATLAS Collaboration, Measurement of the parton distribution functions of the proton from ATLAS measurements of differential W and Z boson production in association with jets. JHEP 07, 223 (2021). https://doi.org/10.1007/JHEP07(2021)223.
9. ATLAS Collaboration, Determination of the parton distribution functions of the proton from ATLAS measurements of differential $W^\pm$ and $Z$ boson production in association with jets. JHEP 07, 223 (2021). https://doi.org/10.1007/JHEP07(2021)223.
10. ATLAS Collaboration, Determination of the parton distribution functions of the proton from ATLAS measurements of differential $W^\pm$ and $Z$ boson production in association with jets. JHEP 07, 223 (2021). https://doi.org/10.1007/JHEP07(2021)223.
65. R. Abdul Khalek et al., Parton distributions with theory uncertainties; general formalism and first phenomenological studies. Eur. Phys. J. C 79, 931 (2019). https://doi.org/10.1140/epjc/s10052-019-7401-4. arXiv:1906.10698 [hep-ph]

66. C. Duhr, F. Dulat, B. Mistlberger, Drell–Yan cross section to third order in the strong coupling constant. Phys. Rev. Lett. 125, 172001 (2020). https://doi.org/10.1103/PhysRevLett.125.172001. arXiv:2001.07717 [hep-ph]

67. V.N. Gribov, L.N. Lipatov, Deep inelastic ep scattering in perturbation theory. Sov. J. Nucl. Phys. 15, 438 (1972)

68. G. Altarelli, G. Parisi, Asymptotic freedom in parton language. Nucl. Phys. B 126, 298 (1977). https://doi.org/10.1016/0550-3213(77)90384-4

69. Y.L. Dokshitzer, Calculation of the structure functions for deep inelastic scattering and e+ e− annihilation by perturbation theory in quantum chromodynamics. Sov. Phys. JETP 46, 641 (1977)

70. H. Abramowicz et al., Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic ep scattering at HERA. Eur. Phys. J. C 78, 473 (2018). https://doi.org/10.1140/epjc/s10052-018-5848-4. arXiv:1804.01019 [hep-ex]

71. I. Abt et al., Impact of jet-production data on the next-to-next-to-leading-order determination of HERAPDF2.0 parton distributions. Eur. Phys. J. C 82(3), 243 (2022). arXiv:2112.01120 [hep-ex]

72. K. Kovcheg et al., Hadronic structure in high-energy collisions. Rev. Mod. Phys. 92, 045003 (2020). https://doi.org/10.1103/RevModPhys.92.045003. arXiv:1905.06957 [hep-ph]

73. R.D. Ball et al., The path to proton structure at one-percent accuracy. Phys. Rev. D 102.181801. arXiv:0901.2169 [hep-ex]

74. X. Chen et al., Di-lepton rapidity distribution in Drell–Yan production. Phys. Rev. D 103, 091102 (2021). https://doi.org/10.1103/PhysRevD.103.091102. arXiv:1703/PhysRevD.76.012003. arXiv:0808.3914 [hep-ex]

75. D. d’Enterria, J. Rojo, Quantitative constraints on the gluon distribution function in the proton from collider isolated-photon data. Nucl. Phys. B 860, 311 (2012). https://doi.org/10.1016/j.nuclphysb.2012.03.003. arXiv:1202.1762 [hep-ph]

76. T.-J. Hou et al., New CTEQ global analysis of quantum chromodynamics. Phys. Rev. D 105, 014011 (2021). https://doi.org/10.1103/PhysRevD.105.014011. arXiv:1903.00014 [hep-ph]

77. A. Martin et al., Parton distributions for the LHC Run II. Eur. Phys. J. C 63, 189 (2009). https://doi.org/10.1140/epjc/s10052-009-1072-5. arXiv:0901.0002 [hep-ph]

78. F. Faura, S. Iranipour, E.R. Nocera, J. Rojo, M. Ubiali, The strangest proton? Eur. Phys. J. C 80, 1148 (2020). https://doi.org/10.1140/epjc/s10052-020-08749-3. arXiv:2009.00014 [hep-ph]

79. A. Martin et al., Parton distributions for the LHC. Eur. Phys. J. C 590, 561 (2021). https://doi.org/10.1038/s41586-021-03282-z. arXiv:2103.04024 [hep-ph]

80. S. Dulat et al., New parton distribution functions from a global analysis of quantum chromodynamics. Phys. Rev. D 93, 033006 (2016). https://doi.org/10.1103/PhysRevD.93.033006. arXiv:1506.07443 [hep-ph]

81. S. Alekhi, J. Blümlein, S. Moch, R. Placakyte, Parton distribution functions, αs, and heavy-quark masses for LHC Run II. Phys. Rev. D 96, 014011 (2017). https://doi.org/10.1103/PhysRevD.96.014011. arXiv:1701.05838 [hep-ph]

82. R.D. Ball et al., Parton distributions for the LHC Run II. JHEP 04, 040 (2015). https://doi.org/10.1007/JHEP04(2015)040. arXiv:1410.8849 [hep-ph]

83. F. Faura, S. Iranipour, E.R. Nocera, J. Rojo, M. Ubiali, The strangest proton? Eur. Phys. J. C 80, 1148 (2020). https://doi.org/10.1140/epjc/s10052-020-08749-3. arXiv:2009.00014 [hep-ph]

84. A. Martin et al., Parton distributions for the LHC. Eur. Phys. J. C 590, 561 (2021). https://doi.org/10.1038/s41586-021-03282-z. arXiv:2103.04024 [hep-ph]

85. LHAPDF6 website. https://lhapdf.hepforge.org

86. J. Dove et al., The asymmetry of antimatter in the proton. Nature 590, 561 (2021). https://doi.org/10.1038/s41586-021-03282-z. arXiv:2103.04024 [hep-ph]

87. ATLAS Collaboration, ATLAS Computing Acknowledgements, ATL-SOFT-PUB-2021-003. https://cds.cern.ch/record/2776662

88. CDF Collaboration, Direct measurement of the W production charge asymmetry in p+p collisions at √s = 1.96 TeV. Phys. Rev. Lett. 102, 181801 (2009). https://doi.org/10.1103/PhysRevLett.102.181801. arXiv:0901.2169 [hep-ex]

89. CDF Collaboration, Measurement of dσ/dy of Drell-Yan e+ e− pairs in the Z mass region from p+p collisions at √s = 1.96 TeV. Phys. Lett. B 692, 232 (2010). https://doi.org/10.1016/j.physletb.2010.06.043. arXiv:0908.3914 [hep-ex]

90. D0 Collaboration, Measurement of the shape of the boson rapidity distribution for p→ Z/γ∗ → e+e− + X events produced at √s = 1.96 TeV. Phys. Rev. D 76, 012003 (2007). https://doi.org/10.1103/PhysRevD.76.012003. arXiv:hep-ex/0702025

91. D0 Collaboration, Measurement of the muon charge asymmetry in p+p → W→μ+X events at √s = 1.96 TeV. Phys. Rev. D 88, 091102 (2013). https://doi.org/10.1103/PhysRevD.88.091102. arXiv:1309.2591 [hep-ex]

92. D0 Collaboration, Measurement of the electron charge asymmetry in p+p → W→e+X decays in p+p collisions at √s = 1.96 TeV. Phys. Rev. D 91, 032007 (2015). https://doi.org/10.1103/PhysRevD.91.032007. arXiv:1412.2862 [hep-ex] [Erratum: Phys. Rev. D 91, 079901 (2015). https://doi.org/10.1103/PhysRevD.91.079901]

93. R.S. Towell et al., Improved measurement of the anti-d / anti-u asymmetry in the nucleon sea. Phys. Rev. D 64, 052002 (2001). https://doi.org/10.1103/PhysRevD.64.052002. arXiv:hep-ex/0103030
