Determination of the parton distribution functions of the proton from ATLAS measurements of differential $W^\pm$ and $Z$ boson production in association with jets

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

Abstract: This article presents a new set of proton parton distribution functions, ATLASepWZVjet20, produced in an analysis at next-to-next-to-leading order in QCD. The new data sets considered are the measurements of $W^+$ and $W^-$ boson and $Z$ boson production in association with jets in $pp$ collisions at $\sqrt{s} = 8$ TeV performed by the ATLAS experiment at the LHC with integrated luminosities of $20.2 \text{ fb}^{-1}$ and $19.9 \text{ fb}^{-1}$, respectively. The analysis also considers the ATLAS measurements of differential $W^\pm$ and $Z$ boson production at $\sqrt{s} = 7$ TeV with an integrated luminosity of $4.6 \text{ fb}^{-1}$ and deep-inelastic-scattering data from $e^\pm p$ collisions at the HERA accelerator. An improved determination of the sea-quark densities at high Bjorken $x$ is shown, while confirming a strange-quark density similar in size to the up- and down-sea-quark densities in the range $x \lesssim 0.02$ found by previous ATLAS analyses.

Keywords: Hadron-Hadron scattering (experiments), QCD

ArXiv ePrint: 2101.05095
1 Introduction

Precise knowledge of the content of colliding protons, the parton distribution functions (PDFs), is a necessary ingredient for accurate predictions of both Standard Model (SM) and Beyond Standard Model (BSM) cross sections at the Large Hadron Collider (LHC). In order to determine the PDFs to the required precision, data covering a wide range of negative squared four-momentum transfer (denoted by $Q^2$) and Bjorken $x$, the fraction of the proton’s longitudinal momentum carried by the parton initiating the interaction, is required. This is facilitated by combining data from multiple experiments and measurements of various processes to constrain the $x$-dependence and flavour decomposition of the PDFs. While deep inelastic scattering (DIS) data from lepton-hadron collisions typically deliver the best constraints by utilising the lepton as a direct probe of the substructure of the hadron, a hadron-hadron experiment can provide valuable additional insight by introducing new processes which further distinguish contributions from different partons and span kinematic regions at higher $Q^2$.

Precision measurements by the HERA collaborations [1] of neutral current (NC) and charged current (CC) cross sections in $e^\pm p$ scattering constrain PDFs such that the HERA DIS data alone provide sufficient information to determine the PDF set referred to as HERAPDF2.0. However, they do have limitations. For example, they cannot distinguish quark flavour between the down-type sea quarks, $\bar{d}$ and $\bar{s}$. Global PDF analyses [2–5] use a range of data from other experiments together with the HERA data for further constraining power. For example, additional information about quarks and antiquarks at mid- to high-$x$
comes from fixed-target DIS experiments, as well as measurements of \( W \) and \( Z \) boson production from the Tevatron and LHC experiments.

More information about high-\( x \) quarks would be advantageous since a large fraction of the fixed-target DIS data is in a kinematic region where non-perturbative effects, such as those from higher twist, are important and must be computed from phenomenological models [6, 7]. In many PDF analyses, tight cuts are applied to these data to avoid those effects. Furthermore, the interpretation of DIS data using deuteron or heavier nuclei as targets is subject to uncertain nuclear corrections. The \( W^\pm \) asymmetry measurements performed using \( pp \) collisions at Tevatron are free from these uncertainties, but there have historically been tensions between the results of the CDF [8] and DØ [9] collaborations, discussed in further detail by the MSTW group in ref. [10] and the CTEQ group in ref. [11].

Precision measurements from the ATLAS detector at the LHC, together with data from the HERA experiments, have been interpreted previously in a next-to-next-to-leading-order (NNLO) QCD analysis, resulting in the ATLASepWZ16 PDF set [12]. Differential \( W \) and \( Z \) boson\(^1\) cross-section measurements at \( \sqrt{s} = 7 \) TeV were used, thereby allowing the strange content of the sea to be fitted, rather than assumed to be a fixed fraction of the light sea as is required when fitting HERA inclusive data alone. It was found that these additional data were significantly better described by a strange sea unsuppressed relative to the up- and down-quark sea at \( x \lesssim 0.05 \), in contradiction to previous assumptions based on dimuon production data from muon-neutrino CC DIS with associated charm-quark production [13]. The finding of an unsuppressed strange PDF in this kinematic region is supported by the ATLAS measurement of \( W \) boson production in association with a charm quark (\( W + c \)) at 7 TeV [14]; however, a recent analysis of CMS \( W + c \) data at 7 and 13 TeV [15] has found a suppressed strange-quark density relative to the light sea, which is potentially in tension with these ATLAS findings.

Data on the production of a vector boson in association with jets at the LHC provides a novel source of input to PDF determination that is sensitive to partons at higher \( x \) and \( Q^2 \) than can be accessed by \( W \) and \( Z \) boson data alone, thereby yielding a data set complementary to the inclusive \( W, Z \) boson measurements [16]. The tree-level production modes of a vector boson in association with jets (\( V + \) jets) have either quark-antiquark initial states with gluon radiation, or quark-gluon initial states. The process is therefore already sensitive to the gluon density of the proton at leading order in quantum chromodynamics (QCD), while providing constraints on the quark distributions in a similar way to inclusive production of a vector boson.

This paper presents a PDF analysis including data on \( W^\pm + \) jets and \( Z + \) jets production collected in \( pp \) collisions at \( \sqrt{s} = 8 \) TeV by the ATLAS collaboration [17, 18] in combination with the previous inclusive \( W \) and \( Z \) measurements at \( \sqrt{s} = 7 \) TeV [12] and the inclusive combined HERA data [1]. The PDF fit is performed at NNLO in perturbative QCD, made possible by recent theoretical developments for vector-boson production in association with one jet [19, 20], and accounts for the correlation of systematic uncertainties between data sets. The resulting PDF set is called \( ATLASepWZVjet20 \).

---

\(^1\)Where \( Z \) boson production is written, this refers to \( Z/\gamma^* \) boson production.
2 Input data sets

The final combined $e^+p$ cross-section measurements at HERA [1] cover the kinematic range of $Q^2$ from 0.045 GeV$^2$ to 50 000 GeV$^2$ and of Bjorken $x$ from 0.65 down to $6 \times 10^{-7}$. Data below $x = 10^{-5}$ are excluded from this analysis by requiring $Q^2 > 10$ GeV$^2$, motivated by the previously observed poorer fit quality in the excluded kinematic region compared to the rest of the HERA data [1]. Possible explanations for this include the need for resummation corrections at low $x$ [21] and the impact of higher-twist corrections at low $Q^2$. For the final HERA data set, there are 169 correlated sources of uncertainty. Total uncertainties are below 1.5% over the $Q^2$ range of $10 < Q^2 < 500$ GeV$^2$ and below 3% up to $Q^2 = 3000$ GeV$^2$.

The ATLAS $W$ and $Z$ differential cross sections are based on data recorded during $pp$ collisions with $\sqrt{s} = 7$ TeV, and a total integrated luminosity of 4.6 fb$^{-1}$, in the electron and muon boson-decay channels [12]. The $W^\pm$ differential cross sections are measured as functions of the $W$-decay lepton pseudorapidity, $\eta_\ell$, split into $W^+$ and $W^-$ cross sections. The experimental precision is between 0.6% and 1.0%. Double-differential distributions of the dilepton rapidity, $y_{\ell\ell}$, in $Z$ boson decays are measured in three mass ranges: $46 < m_{\ell\ell} < 66$ GeV, $66 < m_{\ell\ell} < 116$ GeV and $116 < m_{\ell\ell} < 150$ GeV in central ($|y_{\ell\ell}| < 2.4$) and forward ($1.2 < |y_{\ell\ell}| < 3.6$) rapidity selections, with an experimental precision of up to 0.4% for central rapidity and 2.3% for forward rapidity. The integrated luminosity of the data set used for the 7 TeV $W$ and $Z$ cross-section measurements is known to within 1.8%. There are a total of 131 sources of correlated systematic uncertainty across the $W$ and $Z$ data sets [12]. These data were used for the ATLASepWZ16 fit in a format in which the measurements of the electron and muon decay channels were combined, whereas for the PDF sets presented in this article the data before this combination is used. This choice was made because the uncombined data retain the physical origin of the sources of correlated uncertainties, thereby allowing these sources to be treated as correlated with those in other data sets.

The ATLAS $W^\pm +$ jets differential cross sections are based on data recorded during $pp$ collisions with $\sqrt{s} = 8$ TeV and a total integrated luminosity of 20.2 fb$^{-1}$, in the electron decay channel only [17]. Each event contains at least one jet with transverse momentum $p_T > 30$ GeV and rapidity $|y| < 4.4$, where jets are defined using the anti-$k_t$ algorithm [22, 23] with a radius parameter $R = 0.4$. The spectrum used is the transverse momentum of the $W$ boson ($p_T^W$), in the range $25 < p_T^W < 800$ GeV, chosen because it provides the most constraining power. This is split into $W^+$ and $W^-$ cross sections, which have large correlations that are fully considered. The experimental uncertainty ranges from 8.2% to 22.1% [17]. There are 50 sources of correlated systematic uncertainty common to the $W^+$ and $W^-$ spectra, as well as three sources of uncorrelated uncertainties related to data statistics, background Monte Carlo (MC) simulation statistics and the statistical uncertainty of the data-driven multijet background estimation. Full information about the statistical bin-to-bin correlations in data, induced by the unfolding process, is available for each $W +$ jets spectrum.

The ATLAS $Z +$ jets double-differential cross sections are also based on data recorded during $pp$ collisions with $\sqrt{s} = 8$ TeV. The total integrated luminosity of this data is
19.9 fb$^{-1}$, and the $Z \rightarrow e^+e^-$ decay channel is used [18]. The measurement is performed as a function of the absolute rapidity of inclusive anti-$k_t$ $R = 0.4$ jets, $|y_{\text{jet}}^\text{abs}|$, for several bins of the transverse momentum within $25 \text{ GeV} < p_T^{\text{jet}} < 1050 \text{ GeV}$. The experimental precision ranges from 4.7% to 37.1%. There are 42 sources of correlated systematic uncertainty and two sources of uncorrelated uncertainty related to the data and background MC simulation statistics.

The integrated luminosity of the data set used for the $W$ + jets and $Z$ + jets cross-section measurements is known to within 1.9%. Systematic uncertainties which contribute significantly, such as from the jet energy scale, are treated as correlated across data sets if they correspond to the same physical source. More details of the correlation model used in this analysis are given in appendix A.

3 Fit framework

This determination of proton PDFs uses the xFitter framework, v2.0.1 [1, 24, 25]. This program interfaces to theoretical calculations directly or uses fast interpolation grids to make theoretical predictions for the considered processes. The program MINUIT [26] is used for the minimisation of the PDF fit. The results are cross-checked with an independent fit framework [27].

For the DIS processes, coefficient functions with massless quarks are calculated at NNLO in QCD as implemented in QCDNUM v17-01-13 [28]. The contributions of heavy quarks are calculated in the general-mass variable-flavour-number scheme of refs. [29–31]. The renormalisation and factorisation scales for the DIS processes are taken as $\mu_r = \mu_f = \sqrt{Q^2}$.

For the differential $W$ and $Z$ boson cross sections, the theoretical framework is the same as that used in the ATLASepWZ16 analysis of ref. [12]. The xFitter package uses outputs from the APPLGRID code [32] interfaced to the MCFM program [33, 34] for fast calculation of the differential cross sections at NLO in QCD and LO in electroweak (EW) couplings. Corrections to higher orders are implemented using a $K$-factor technique, correcting on a bin-by-bin basis from NLO to NNLO in QCD and from LO to NLO for the EW contribution [35, 36].

Predictions for $W$ + jets and $Z$ + jets production are obtained similarly to the $W$ and $Z$ predictions to NLO in QCD and LO in EW couplings by using the APPLGRID code interfaced to the MCFM program. Higher-order corrections are implemented as $K$-factors. For the $W$ + jets data, the $N_{\text{jettip}}$ program [19] is used to calculate and implement corrections to NNLO in QCD, while the non-perturbative hadronisation and underlying event QCD corrections are computed using the SHERPA v.2.2.1 MC simulation$^2$ [39–41]. The bin-by-bin $K$-factors are derived as the ratio of the NNLO to the NLO calculation from $N_{\text{jettip}}$ with the same fiducial selection as the $W$ + jets data, multiplied by the non-perturbative correction. The renormalisation and factorisation scales are set to $\mu_r = \mu_f = \sqrt{m_W^2 + \Sigma(p_T^2)}$, where $m_W$ is the mass of the $W$ boson and the second term in the square root is the scalar sum.

$^2$These non-perturbative corrections account for the collective effect of using jets clustered from showered partons to jets clustered from hadrons with multi-parton-interactions and underlying event simulation enabled [37, 38].
of the squared transverse momenta of the jets. More details about the predictions are given in the respective ATLAS publication [17]. In addition to these predictions, NLO EW corrections inclusive of QED radiation effects are computed using Sherpa v.2.2.10 by the authors of refs. [39–41] and applied as additional bin-by-bin multiplicative $K$-factors.

Predictions for $Z +$ jets production to NNLO in QCD and LO in EW couplings are calculated by the authors of ref. [20], and the $K$-factor is calculated as the ratio of NNLO to NLO predictions. The renormalisation and factorisation scales are set to $\mu_r = \mu_f = \frac{1}{2} \left( \sum p_{T,\text{partons}} + \sqrt{m_{T,\ell \ell}^2 + p_{T,\ell \ell}^2} \right)$ where $m_{T,\ell \ell}$ is the electron-pair invariant mass, $p_{T,\ell \ell}$ 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. Corrections for QED radiation effects and non-perturbative QCD corrections are each calculated using the Sherpa v.1.4.5 MC simulation, as discussed in the publication describing the ATLAS measurement [18], and each provided as a set of bin-by-bin multiplicative $K$-factors. Corrections for NLO EW effects excluding QED radiation are computed using Sherpa v.2.2.10 and applied as additional bin-by-bin $K$-factors. The $K$-factors for both $W +$ jets and $Z +$ jets production are typically within 10% of unity, except for the NLO EW corrections for the $W +$ jets predictions, which are as large as 20% at high $p_T$.

The DGLAP evolution equations of QCD yield the proton PDFs at any value of $Q^2$ given that they are parameterised as functions of $x$ at an initial scale $Q_0^2$. In this analysis, the initial scale is chosen to be $Q_0^2 = 1.9 \text{ GeV}^2$ such that it is below the charm-mass matching scale, $\mu_c^2$, which is set equal to the charm mass, $\mu_c = m_c$. The heavy-quark masses are set to their pole masses as determined by a combined analysis of HERA data on inclusive and heavy-flavour DIS processes [1, 42], $m_c = 1.43 \text{ GeV}$ and $m_b = 4.5 \text{ GeV}$, and the strong coupling constant is fixed to $\alpha_S(m_Z) = 0.118$. These choices follow those of the HERAPDF2.0 fit [1].

The quark distributions at the initial scale are assumed to behave according to the following parameterisation also used by the HERAPDF2.0 and ATLAS$\text{epWZ16}$ fits [1, 12]

$$xq_i(x) = A_i x^{B_i}(1 - x)^{C_i} P_i(x), \tag{3.1}$$

where $P_i(x) = (1 + D_i x + E_i x^2) e^{F_i x}$. The parameterised quark distributions, $xq_i$, are chosen to be the valence-quark distributions ($xu_v, \ \bar{x}d_v$) and the light-antiquark distributions ($x\bar{u}, \ xd, \ x\bar{s}$). The gluon distribution is parameterised with the more flexible form

$$xg(x) = A_g x^{B_g}(1 - x)^{C_g} P_g(x) - A'_g x^{B'_g}(1 - x)^{C'_g},$$

where $C'_g$ is fixed to a value of 25 to suppress negative contributions from the primed term at high $x$, as in ref. [10]. The parameters $A_{u_v}$ and $A_{d_v}$ are constrained using the quark counting rules, and $A_g$ is constrained using the momentum sum rule. The normalisation and slope parameters, $A$ and $B$, of the $\bar{u}$ and $\bar{d}$ PDFs are set equal such that $x\bar{u} = x\bar{d}$ as $x \to 0$. The strange PDF $x\bar{s}$ is parameterised as in eq. (3.1), with $P_s = 1$ and $B_s = B_d$, leaving two free parameters for the strange PDF, $A_s$ and $C_s$. It is assumed that $xs = x\bar{s}$ as the data used are not sufficient to distinguish between the two.
The $D$, $E$ and $F$ terms in the expression $P_i(x)$ are used only if required by the data, following the procedure described in ref. [1]. For the ATLASepWZVjet20 fit, this results in the usage of two additional parameters: $E_{uv}$ and $D_g$. In total, 16 free parameters are used in the central fit.

The level of agreement of the data with the predictions from a PDF parameterisation is quantified with a $\chi^2$. The definition of the $\chi^2$ without statistical correlations between data points is as follows [1, 12]

$$
\chi^2 = \sum_i \left( \frac{D_i - T_i \left( 1 - \sum_j \gamma_{ij} b_j \right)}{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i} \right)^2 + \sum_j b_j^2 + \sum_i \log \frac{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i}{\delta_{i,\text{uncor}}^2 D_i^2 + \delta_{i,\text{stat}}^2 D_i^2},
$$

(3.2)

where $D_i$ represent the measured data, $T_i$ represent the corresponding theoretical prediction, $\delta_{i,\text{uncor}}$ and $\delta_{i,\text{stat}}$ are the uncorrelated systematic and statistical uncertainties in $D_i$, and correlated systematic uncertainties, described by $\gamma_{ij}$, are accounted for using the nuisance parameters $b_j$. The summation over $i$ runs over all data points and the summation over $j$ runs over all sources of correlated systematic uncertainties. For each data set, the first term gives the partial $\chi^2$ and the second term gives the correlated $\chi^2$. The third term is a bias correction term arising from the transition of the likelihood to $\chi^2$ when the scaling of errors is applied, referred to as the log penalty. For the $W + jets$ data, the bin-to-bin statistical correlations are significant in contrast to the other data sets and incorporated into the $\chi^2$ definition as follows

$$
\chi^2 = \sum_{ik} \left( \frac{D_i - T_i \left( 1 - \sum_j \gamma_{ij} b_j \right) C_{\text{stat},ik}(D_i, D_k) \left( D_k - T_k \left( 1 - \sum_j \gamma_{kj} b_j \right) \right)}{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i} \right)^2 + \sum_j b_j^2 + \sum_i \log \frac{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i}{\delta_{i,\text{uncor}}^2 D_i^2 + \delta_{i,\text{stat}}^2 D_i^2},
$$

(3.3)

in which the first term of eq. (3.2) has been replaced with one which takes into account the diagonal and off-diagonal elements of the data statistical covariance matrix between bins $i$ and $k$, $C_{\text{stat},ik}$.

4 Results

In this section, the ATLASepWZVjet20 PDF set is presented and compared with an equivalent fit performed without the $V + jets$ data, where the latter is named the ATLASepWZ20 PDF set. These PDFs differ from the ATLASepWZ16 analysis by an additional parameter, $D_g$, a tighter selection criterion of $Q^2 > 10$ GeV$^2$ and the use of ATLAS 7 TeV $W$ and $Z$ data in which the electron and muon channels are not combined. The result is very similar except for a larger total uncertainty resulting from the use of more parameterisation variations. It was verified that the use of 7 TeV $W$ and $Z$ data with the electron and muon channels combined provides a fit with very similar central values and uncertainties.
Figure 1. The differential cross-section measurements of (a) $W^+ + \text{jets}$ and (b) $W^- + \text{jets}$ in ref. [17] (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 predetermined PDFs resulting from the fits ATLASepWZ20 (red lines) and ATLASepWZVjet20 (blue lines). The solid lines show the predictions without shifts of the systematic uncertainties, while for the dashed lines the $b_j$ parameters associated with the experimental systematic uncertainties as shown in eq. (3.3) are allowed to vary to minimise the $\chi^2$. The ratios of each set of predictions to the data are shown in the bottom panel in each case.

Rather than being intended to supersede the ATLASepWZ16 PDF, the ATLASepWZ20 fit is provided to allow a more meaningful comparison with the ATLASepWZVjet20 fit by having a PDF set that differs only in the addition of the $V + \text{jets}$ data.

4.1 Goodness of fit and parton distributions

Figures 1 to 3 show a comparison of the $W + \text{jets}$ and $Z + \text{jets}$ differential cross-section measurements with the predictions of the ATLASepWZ20 and ATLASepWZVjet20 fits. Adding the $V + \text{jets}$ data to the fit improves the $W + \text{jets}$ description significantly, particularly in the $W^+$ spectrum, where agreement with data improves by approximately 20% at high $p_T^W$. The difference in partial $\chi^2$ between the predictions of the ATLASepWZ20 and ATLASepWZVjet20 PDF sets for the $W + \text{jets}$ and the $Z + \text{jets}$ data is 32 and 7 units, respectively.

The total $\chi^2$ per degree of freedom ($\chi^2/NDF$) for the ATLASepWZVjet20 fit, along with the partial $\chi^2$ per data point ($\chi^2/NDP$) and correlated $\chi^2$ for each data set entering the fit, is given in table 1. The partial $\chi^2$ for the HERA and ATLAS inclusive $W$ and $Z$ data in the ATLASepWZVjet20 fit is similar to those obtained in the ATLASepWZ20 fit, not showing any tension between these data and the $V + \text{jets}$ data. The partial $\chi^2$ of the
Figure 2. The differential cross-section measurements of $Z + \text{jets}$ as a function of the absolute rapidity of inclusive jets, $|y_{\text{jet}}|$, in bins of (a) $25 < p_{\text{T}}^{\text{jet}} < 50$ GeV, (b) $50 < p_{\text{T}}^{\text{jet}} < 100$ GeV, (c) $100 < p_{\text{T}}^{\text{jet}} < 200$ GeV and (d) $200 < p_{\text{T}}^{\text{jet}} < 300$ GeV, where the transverse momentum of inclusive jets is labelled $p_{\text{T}}^{\text{jet}}$. 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 to the predictions computed with the PDFs resulting from the fits ATLASepWZ20 (red lines) and ATLASepWZVjet20 (blue lines). The solid lines show the predictions without shifts of the systematic uncertainties, while for the dashed lines the shifts with fitted $b_j$ parameters as shown in eq. (3.3) are applied. The ratios of each set of predictions to the data are shown in the bottom panel in each case.
Figure 3. The differential cross-section measurements of $Z +\text{jets}$ as a function of the absolute rapidity of inclusive jets, $|y_{\text{jet}}|$, in bins of (a) $300 < p^\text{jet}_T < 400$ GeV and (b) $400 < p^\text{jet}_T < 1050$ GeV, where the transverse momentum of inclusive jets is labelled $p^\text{jet}_T$. 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 to the predictions computed with the PDFs resulting from the fits ATLASepWZ20 (red lines) and ATLASepWZVjet20 (blue lines). The solid lines show the predictions without shifts of the systematic uncertainties, while for the dashed lines the shifts with fitted $b_j$ parameters as shown in eq. (3.3) are applied. The ratios of each set of predictions to the data are shown in the bottom panel in each case.

| Fit                          | ATLASepWZVjet20 |
|------------------------------|-----------------|
| Total $\chi^2$/NDF          | 1460/1198       |
| HERA partial $\chi^2$/NDP   | 1132/1016       |
| HERA correlated $\chi^2$    | 50              |
| HERA log penalty $\chi^2$   | -12             |
| ATLAS $W, Z$ partial $\chi^2$/NDP | 113/105        |
| ATLAS $W +\text{jets}$ partial $\chi^2$/NDP | 25/30           |
| ATLAS $Z +\text{jets}$ partial $\chi^2$/NDP | 82/63           |
| ATLAS correlated $\chi^2$   | 65              |
| ATLAS log penalty $\chi^2$  | 6               |

Table 1. $\chi^2$ values split by partial, correlated and log penalty for each data set entering the ATLASepWZVjet20 fit. The partial component of the $\chi^2$ for each data set is shown compared with the number of data points of that data set (NDP).
$W + \text{jets}$ and $Z + \text{jets}$ data is reasonable, and neither the HERA nor ATLAS correlated $\chi^2$ is observed to increase significantly with the inclusion of this data.

Additional uncertainties in the PDFs are estimated and classified as either model or parameterisation uncertainties. Model uncertainties comprise variations of the charm-quark mass ($m_c$) and bottom-quark mass ($m_b$), variations of the minimum $Q^2$ cut, $Q^2_{\text{min}}$, and the starting scale at which the PDFs are parameterised, $Q^2_0$. The variation in charm-quark mass and starting scale are performed simultaneously to fulfil the condition $Q^2_0 < m_c^2$ such that the charm PDF is calculated perturbatively. Each of these variations follow that of the ATLASepWZ16 analysis [12]. The parameterisation uncertainties are estimated through variations which include a single further parameter in the polynomial $P_i(x)$ or relaxed constraints on the low-$x$ sea quarks. In each variation, listed with its respective total $\chi^2$ per degree of freedom in table 2, the uncertainty is calculated as the difference between the alternative extracted PDF and the nominal PDF at each value of $x$ and $Q^2$. Whereas the model variations are treated independently and the model uncertainty is calculated as the sum in quadrature of the variations, the parameterisation uncertainty is taken as the envelope of the parameterisation variations. The total uncertainty is calculated as the sum in quadrature of the experimental, model and parameterisation uncertainties. While the total uncertainty does give an estimate of the total variability of the fit, only the experimental uncertainty is interpretable similarly to a statistical standard deviation.

The impact of theoretical uncertainties in the $V + \text{jets}$ predictions on the fit results is cross-checked. Variations of the NNLO QCD calculations are defined from the variations of factorisation and renormalisation scales by factors of two up and down and taking the envelope of these predictions. In the fit, the corresponding $K$-factors are varied for the $W + \text{jets}$ and $Z + \text{jets}$ prediction upward and downward both simultaneously and individually. Each of these variations results in PDFs well within the experimental uncertainties of the nominal ATLASepWZjet20 set.

Figures 4 and 5 shows the ATLASepWZVjet20 PDFs overlaid with the ATLASepWZ20 PDFs, each evaluated at the starting scale $Q^2_0$, for comparison. The experimental and total uncertainties are displayed separately in each case. The ATLASepWZVjet20 $x\bar{d}$ distribution is notably higher in the range $x \gtrsim 0.02$ compared to the ATLASepWZ20 fit. In contrast, the $x\bar{s}$ distribution of the ATLASepWZVjet20 fit in the same region is lower. Together, the differences observed between the ATLASepWZ20 and ATLASepWZVjet20 PDFs allow for an increase in the $W^+$ cross section, as depicted in figure 1, while keeping the total down-type sea $x\bar{D} = x\bar{d} + x\bar{s}$ distribution almost unchanged up to $x \sim 0.1$. Additionally, the $d_v$ distribution is reduced in the ATLASepWZVjet20 fit at high $x$ and increased at low $x$, compensating for the changes in the other PDFs and resulting in an $xD = xd_v + x\bar{d} + x\bar{s}$ distribution which is similar at high $x$. The up-type quark and gluon distributions are similar between the two fits.

4.2 The high-$x$ sea-quark distributions

The difference between the $x\bar{d}$ and $x\bar{u}$ PDFs at high $x$ has been a topic of debate over the recent decades. A measurement by the E866 collaboration of the Drell-Yan cross-section ratios from an 800 GeV proton beam incident on liquid hydrogen and deuterium targets
Table 2. Total $\chi^2$/NDF for each parameterisation and model variation contributing to the parameterisation and model uncertainties, respectively, of the ATLASepWZVjet20 fit. Where a $D$, $E$ or $F$ parameter is referred to, this means that the respective parameter is not constrained to zero in that variation. Where two $A$ or $B$ parameters are referred to in an inequality, this means that the respective two parameters are free to vary independently of each other in a fit.
Figure 4. PDFs multiplied by Bjorken $x$ at the scale $Q^2 = 1.9 \text{GeV}^2$ as a function of Bjorken $x$ obtained for the (a)–(b) valence quarks and (c)–(d) up and down sea quarks when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive $W$ and $Z$, and HERA data (ATLASepWZVjet20, blue bands), compared with a similar fit without $W + \text{jets}$ or $Z + \text{jets}$ data (ATLASepWZ20, green bands). Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties. The relative uncertainties around the nominal value of each PDF centred on 1 is displayed in the bottom panel in each case.
Figure 5. PDFs multiplied by Bjorken $x$ at the scale $Q^2 = 1.9 \text{ GeV}^2$ as a function of Bjorken $x$ obtained for the (a) strange sea quark, (b) gluon, (c) the total of the down-type quarks and (d) the total of the anti-down-type quarks when fitting $W^+ + \text{jets}$, $Z^+ + \text{jets}$, inclusive $W$ and $Z$, and HERA data (ATLASepWZVjet20, blue bands), compared with a similar fit without $W^+ + \text{jets}$ or $Z^+ + \text{jets}$ data (ATLASepWZ20, green bands). Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties. The relative uncertainties around the nominal value of each PDF centred on 1 is displayed in the bottom panel in each case.
found the proton $x(\bar{d} - \bar{u})$ distribution to be positive at high $x$, peaking at $x(\bar{d} - \bar{u}) \sim 0.04$ at $x \sim 0.1$ [13]. In contrast, the ATLASepWZ16 PDF set gives a negative central distribution with its lowest value at $x(\bar{d} - \bar{u}) \sim -0.035$ for $x \sim 0.1$, although the uncertainties are such that it is compatible with zero within two standard deviations.

The $x(\bar{d} - \bar{u})$ distribution as function of Bjorken $x$ at $Q^2 = 1.9$ GeV$^2$ is shown in figure 6, with a comparison between ATLASepWZVjet20 and ATLASepWZ20 displaying the direct effect of the $V +$ jets data, and with the experimental, model and parameterisation uncertainties plotted separately. The impact of the $V +$ jets data is to place significant constraints on the total uncertainty at high $x$, with an overall positive distribution of central values driven by the increase in the high-$x$ $\bar{d}$ distribution, as discussed in section 4.1.

To understand the effect of the different data sets on the high-$x$ $\bar{d}$ distribution, a scan of $\chi^2$ is performed through the parameter controlling the behaviour in this region, $C_{\bar{d}}$. A high $C_{\bar{d}}$ value of $\sim 10$ corresponds to a lower $\bar{d}$ distribution at high $x$, as exhibited by the ATLASepWZ20 fit. Conversely, a low $C_{\bar{d}}$ value of $\sim 2$ corresponds to the higher $\bar{d}$ distribution at high $x$ as exhibited by the ATLASepWZVjet20 fit.

In figure 7a, this scan is shown for each of the presented PDF fits, where the $\chi^2$ is evaluated as a function of the scanned parameter, $C_{\bar{d}}$. At each point, all other parameters (including nuisance parameters associated with experimental uncertainties) are re-fitted and the minimum $\chi^2$ of the scan, $\chi^2_{\text{min}}$, is subtracted for comparison between fits.

The $\chi^2$ of the ATLASepWZ20 fit is smallest at a value of $C_{\bar{d}} = 10 \pm 1$, whereas the $\chi^2$ of the ATLASepWZVjet20 fit is smallest at a lower $C_{\bar{d}} = 1.6 \pm 0.3$, corresponding to a higher $\bar{d}$ distribution at $x \gtrsim 0.1$ consistent with the PDFs presented in section 4.1. Another shallow
minimum is observed for the ATLASepWZ20 fit at \( C_d \sim 3 \), corresponding to a solution similar to that of the ATLASepWZVjet20 fit; however, it exhibits a \( \chi^2 \) approximately two units larger than in the best fit. The ATLASepWZVjet20 fit fails to converge for values of \( C_d \gtrsim 12 \) and no second minimum is observed.

In figure 7b, these \( \chi^2 \) distributions are decomposed into contributions from the HERA and ATLAS data. These contributions include the partial, correlated and log penalty \( \chi^2 \), which are discussed in section 3. In each fit, the ATLAS data favour a low \( C_d \), including in the ATLASepWZ20 fit, where the overall result is a higher \( C_d \). Similarly, the HERA data favour the higher \( C_d \) value exhibited by the ATLASepWZ20 fit. The \( V + \text{jets} \) data provide sufficient constraining power in addition to the inclusive \( W \) and \( Z \) data to dominate the result and tightly constrain the \( C_d \) parameter to a low value, while the ATLASepWZ20 fit lacks the necessary information.

### 4.3 Strange-quark density

The fraction of the strange-quark density in the proton can be characterised by the quantity \( R_s \), defined as the ratio

\[
R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}},
\]

which uses the sum of \( \bar{u} \) and \( \bar{d} \) as a reference point for the strange-sea density.
Table 3. Fitted values of $R_s = (s + \bar{s})/(\bar{u} + \bar{d})$, evaluated at $x = 0.023$ and $Q^2 = 1.9\text{GeV}^2$, for each of the investigated fits compared with the ATLASepWZ16 result.

| Fit                  | $R_s$ | Experimental | Model   | Parameterisation |
|----------------------|-------|--------------|---------|------------------|
| ATLASepWZ16          | 1.13  | 0.05         | 0.03    | $^{+0.01}_{-0.06}$ |
| ATLASepWZ20          | 1.13  | 0.06         | 0.03    | $^{+0.09}_{-0.17}$ |
| ATLASepWZVjet20      | 0.99  | 0.04         | $^{+0.05}_{-0.06}$ | $^{+0.14}_{-0.05}$ |

Before the first LHC precision $W$ and $Z$ boson data, it was widely assumed, motivated by previous analyses of dimuon production in neutrino scattering [43–47], that the strange sea-quark density is suppressed equally for all $x$ relative to the up and down sea over the full range of $x$. Best fits to this neutrino scattering data resulted in a value of $R_s \sim 0.5$ at $Q^2 = 1.9\text{GeV}^2$ [2–4, 48].

The QCD analysis of the inclusive $W$ and $Z$ measurements by ATLAS which formed the ATLASepWZ16 PDF set led to the observation that strangeness is unsuppressed at low $x$ ($\lesssim 0.023$) for $Q^2 = 1.9\text{GeV}^2$. This was the case for the ATLASepWZ16 fit for every parameterisation variation used. Furthermore, a Hessian profiling exercise of the global PDFs MMHT14 [4] and CT14 [49] demonstrated that the data constrain and increase the ratio of the strange to the total up and down sea [12]. Although profiling the PDFs does not necessarily give the same result as including the data in a fit, this effect is indeed found when the data is added to the CT18 fit, resulting in the CT18A set of PDFs [3].

It is therefore of particular interest to check the impact of the new $V+\text{jets}$ data on the strange-quark density.

The $R_s$ distribution plotted as a function of $x$ evaluated at $Q^2 = 1.9\text{GeV}^2$ is shown in figure 8, with a comparison between ATLASepWZVjet20 and ATLASepWZ20 showing the direct effect of the $V+\text{jets}$ data, and with the experimental, model and parameterisation uncertainties of ATLASepWZVjet20 shown separately. The effect of the $V+\text{jets}$ data is most significant in the kinematic region $x > 0.02$, where the uncertainty is significantly reduced. Whereas the $R_s$ distribution of the ATLASepWZ20 PDFs maintained an unsuppressed strange-quark density over a wide range in $x$, the ATLASepWZVjet20 PDFs exhibit an $R_s$ distribution falling from near-unity at $x \sim 0.01$ to approximately 0.5 at $x = 0.1$, driven by the increase in the high-$x \bar{d}$ PDF and the complementary decrease in the high-$x \bar{s}$ PDF shown in section 4.1. At low $x \lesssim 0.023$ and $Q^2 = 1.9\text{GeV}^2$, the fit with the $V+\text{jets}$ data maintains an unsuppressed strange-quark density compatible with the ATLASepWZ16 fit. Fitted values of $R_s$, evaluated at $x = 0.023$ and $Q^2 = 1.9\text{GeV}^2$, are given in table 3.

---

$x = 0.013$ is evaluated at $Q^2 = m_Z^2$ as this corresponds to the mean $x$ of inclusive $Z$ production at $\sqrt{s} = 7\text{TeV}$, $\langle x_Z \rangle = m_Z/2E_p$. At the scale $Q^2 = 1.9\text{GeV}^2$, this corresponds to $x = 0.023$ through DGLAP evolution.
4.4 Comparison with global PDFs

The ATLASepWZVjet20 $R_s$ distribution is shown in figure 9 in comparison with the global PDF sets ABMP16 [2], CT18, CT18A [3], MMHT14 [4] and NNPDF3.1 [5]. An additional comparison in the figures is made with a recent update of the NNPDF3.1 fit with some additional data, including the full ATLAS 7 TeV inclusive $W$ and $Z$ data set, labelled NNPDF3.1_strange [48]. Tension between the ATLASepWZVjet20 fit and the global analyses is reduced compared to the ATLASepWZ16 and ATLASepWZ20 PDF sets, but persists to multiple standard deviations in the range $10^{-2} \lesssim x \lesssim 10^{-1}$ for the global analyses which do not use the full ATLAS 7 TeV data set. This is highlighted in summary plots of $R_s$ evaluated at $x = 0.023$, $Q^2 = 1.9\,\text{GeV}^2$ and at $x = 0.013$, $Q^2 = m_Z^2$ in figure 10. Better agreement is observed with the CT18A PDF set, which includes both the data used in the CT18 fit and the ATLAS 7 TeV data, although tension remains with the NNPDF3.1_strange PDF set, which also uses this data. At high $x \gtrsim 0.02$, the $R_s$ distribution of the ATLASepWZVjet20 fit falls more steeply than the $R_s$ distribution in global analyses and is approximately zero at $x \gtrsim 0.2$. The uncertainty of the NNPDF sets are large at $x > 0.3$ where the data give no constraint.

In figure 11 the extracted $x(\bar{d} - \bar{u})$ distribution at $Q^2 = 1.9\,\text{GeV}^2$ is shown in comparison with the results of the latest global PDF sets, all of which use E866 data. The ATLASepWZVjet20 PDF set is consistent with these global PDF sets up to $x \sim 0.1$, but deviates from them for $x > 0.1$, where the $W +$ jets and $Z +$ jets data are most sensitive and demonstrate a preference for a higher $x d$ distribution as discussed in section 4.1. A new result from the SeaQuest/E906 collaboration has recently become available [51], which may also be in tension with the E866 data. Whereas the $R_s$ distribution of the CT18A

\footnote{All these comparisons are done only at NNLO, and not at NLO.}
Figure 9. The $R_s = (s + \bar{s})/(\bar{u} + \bar{d})$ distribution evaluated at $Q^2 = 1.9\text{ GeV}^2$ as a function of Bjorken $x$, for the ATLASepWZVjet20 PDF set in comparison with global PDFs (a) ABMP16 and CT18, (b) MMHT14 and NNPDF3.1, and in additional comparisons with (c) CT18 and CT18A, and (d) NNPDF3.1 and NNPDF3.1\_strange [2–5, 48]. The experimental and total uncertainty bands are plotted separately for the ATLASepWZVjet20 results. Each global PDF set is taken at $\alpha_S(m_Z) = 0.1180$ except for ABMP16 which uses the fitted value $\alpha_S(m_Z) = 0.1147$. All global PDF uncertainty bands are at 68% confidence level, evaluated for the CT18 PDFs through scaling by 1.645 as recommended by the PDF4LHC group [50].

and NNPDF3.1\_strange fits is affected by the ATLAS data and the tension between these and the ATLAS fits is reduced, as shown in figures 9 and 10, this is not replicated in the $x(\bar{d} - \bar{u})$ distribution in either case.

5 Conclusion

This paper presents the impact of measurements, performed by the ATLAS experiment at the LHC, of vector-boson production in association with at least one jet on the parton distribution functions of the proton, resulting in a new ATLASepWZVjet20 PDF set. The $V^+\text{ jets}$ data was obtained from $pp$ collisions at $\sqrt{s} = 8\text{ TeV}$ corresponding to approximately
Figure 10. Summary plots of $R_s$ evaluated at (a) $x = 0.023$ and $Q_s^2 = 1.9\text{ GeV}^2$, and (b) $x = 0.013$ and $Q_s^2 = m_Z^2$, for the ATLASepWZVjet20 PDF set in comparison with global PDFs [2–5, 48], and the ATLASepWZ20 and ATLASepWZ16 sets. The experimental, model and parameterisation uncertainty bands are plotted separately for the ATLASepWZVjet20 results. Each global PDF set is taken at $\alpha_S(m_Z) = 0.1180$ except for ABMP16 which uses the fitted value $\alpha_S(m_Z) = 0.1147$. All uncertainty bands are at 68% confidence level, evaluated for the CT18 PDFs through scaling by 1.645 as recommended by the PDF4LHC group [50].

20 fb$^{-1}$ of integrated luminosity. The data were fitted along with the data sets used for the previous ATLASepWZ16 fit, i.e. the full combined inclusive data set from HERA and the ATLAS inclusive $W$ and $Z$ production data recorded at $\sqrt{s} = 7\text{ TeV}$. For the new ATLASepWZVjet20 PDF set, correlations between all significant systematic uncertainties across different data sets were considered.

The resulting PDF set is similar to the ATLASepWZ16 set for the up-type quarks and gluon. The down and strange sea-quark distributions exhibit significantly smaller experimental and parameterisation uncertainties at high Bjorken $x$. As a result, the ratio of the strange-quark to light-quark densities, $R_s$, is better constrained and found to fall more steeply at high $x$. The $x(\bar{d} - \bar{u})$ difference is positive, in better agreement with the global PDF analyses which use E866 Drell-Yan data up to $x \sim 0.1$ but differs at higher values of $x$ by up to two standard deviations. At low $x \lesssim 0.023$, the fit confirms the unsuppressed strange PDF as observed in the ATLASepWZ16 PDF set, while it maintains a positive $x(\bar{d} - \bar{u})$ distribution at high $x$.

Acknowledgments

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,
Figure 11. The $x(\bar{d} - \bar{u})$ distribution evaluated at $Q^2 = 1.9 \text{ GeV}^2$ as a function of Bjorken $x$, for the ATLASepWZVjet20 PDF set in comparison with global PDFs (a) ABMP16 and CT18, (b) MMHT14 and NNPDF3.1, and in additional comparisons with (c) CT18 and CT18A, and (d) NNPDF3.1 and NNPDF3.1_strange [2–5, 48]. The experimental and total uncertainty bands are plotted separately for the ATLASepWZVjet20 result. Each global PDF set is taken at $\alpha_S(m_Z) = 0.1180$ except for ABMP16 which uses the fitted value $\alpha_S(m_Z) = 0.1147$. All global PDF uncertainty bands are at 68% confidence level, evaluated for the CT18 PDFs through scaling by 1.645 as recommended by the PDF4LHC group [50].
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 (U.K.) and BNL (U.S.A.), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. [52].

A Correlations between data sets

The correlation model used for the ATLAS data is summarised in table 4, where the labels used are the same as those in the HEPData entries of the respective ATLAS $W + \text{jets}$ [17, 53] and $Z + \text{jets}$ [18, 54] publications.

Systematic uncertainties related to the jet energy scale, jet energy resolution, rejection of jets from pile-up (JVF), missing transverse momentum ($E_T^{\text{miss}}$) scale, $E_T^{\text{miss}}$ resolution, electron energy scale and electron energy resolution are taken to have a one-to-one correlation between the 8 TeV data sets. Additionally, systematic uncertainties related to electron efficiency scale factors (trigger, reconstruction and isolation) as well as luminosity are considered fully correlated between data sets of the same centre-of-mass energy, but are uncorrelated between 7 TeV and 8 TeV data. The systematic uncertainty related to the top-background cross section in $W + \text{jets}$ data is taken as correlated with the top-quark pair production cross-section uncertainty in the $Z + \text{jets}$ data as this is the largest of the top-quark-related background contributions in both data sets. Similarly, the systematic uncertainty related to the total diboson cross section in the $W + \text{jets}$ data is taken as correlated with the systematic uncertainty in the $Z + \text{jets}$ data related to the production of two $W$ bosons, as this is the highest background contribution. In contrast, systematic uncertainties related to the multijet background are estimated independently in each measurement, and are therefore left uncorrelated.

The two systematic uncertainties in each of the $W + \text{jets}$ and $Z + \text{jets}$ spectra related to unfolding (labelled in HEPData as “UnfoldOtherGen” and “UnfoldReweight” in the $W + \text{jets}$ data, and “ATL_unfold_Data” and “ATL_unfold_MC” in the $Z + \text{jets}$ data)\footnote{“Unfolding” here refers to the procedure used to correct reconstructed data to particle-level cross sections.} are fully decorrelated between spectra and bins within a single spectrum (in addition to the aforementioned statistical uncertainties) as they contain a large statistical component in both data sets owing to MC simulation statistics. Treating this source of uncertainty as
| Systematic Uncertainty | 7 TeV inclusive $W, Z$ | 8 TeV $W$ + jets | 8 TeV $Z$ + jets |
|------------------------|------------------------|------------------|------------------|
| Jet energy scale [55]  | * JetScaleEff1 ATL_JESP1 |                 |                 |
|                        | * JetScaleEff2 ATL_JESP2 |                 |                 |
|                        | * JetScaleEff3 ATL_JESP3 |                 |                 |
|                        | * JetScaleEff4 ATL_JESP4 |                 |                 |
|                        | * JetScaleEff5 ATL_JESP5 |                 |                 |
|                        | * JetScaleEff6 ATL_JESP6 |                 |                 |
|                        | * JetScaleEta1 ATL_JESP7 |                 |                 |
|                        | * JetScaleEta2 ATL_JESP8 |                 |                 |
|                        | * JetScaleHighPt ATL_JESP9 |               |                 |
|                        | JetScaleMC ATL_JESP10 |                  |                 |
| Jet punchthrough [55]  | — JetScalepunchT ATL_PunchThrough |               |                 |
| Jet resolution [55]    | JetRes JetResolution10 ATL_JER |             |                 |
| Jet flavour composition [55] | — JetScaleFlav1Known ATL_Flavor_Comp |    |                 |
| Jet flavour response [55] | — JetScaleFlav2 ATL_Flavor_Response |            |                 |
| Pile-up jet rejection (JVF) [56] | — JetJVFcut ATL_JVF |           |                 |
| $E_T^{miss}$ scale [57] | MetScaleWen METScale — |                |                 |
| $E_T^{miss}$ resolution [57] | MetRes METResLong — |                | METResTrans — |
| Electron energy scale [58] | * EIScaleZee ATL_ElecEnZee |          |                 |
| Electron trigger efficiency [59] | * EISFTTrigger ATL_Trig |              |                 |
| Electron reconstruction efficiency [60, 61] | * EISFReco ATL_RecoEff |            |                 |
| Electron identification efficiency [60, 61] | * EISFId ATL_IDEff |              |                 |
| Luminosity [62, 63] | * LumiUncert ATL_lumi_2012_8TeV |             |                 |
| WW background cross section [64] | XsecDibos ATL_WW_xs |               |                 |
| Top background cross section [65] | * XsecTop ATL_ttbar_xs |             |                 |

**Table 4.** Correlation model for the systematic uncertainties of the ATLAS measurements of $W$ + jets and $Z$ + jets at 8 TeV and inclusive $W$ and $Z$ at 7 TeV. Each row corresponds to one source of systematic uncertainty treated as fully correlated both within and across data sets. The respective ATLAS publication describing each of these sources in detail is given. Sources in different rows are uncorrelated with each other. Each source is reported with the label of the systematic uncertainty used in the respective data set. Where entries are omitted, that systematic uncertainty either does not exist for that data set (denoted by a ‘—’) or it was left decorrelated from the others (denoted by a :*). For the row of source “$E_T^{miss}$ resolution”, where the “MetRes” label in one column corresponds to “MetResLong” and “MetResTrans” labels in the other, this indicates that the “MetResLong” and “MetResTrans” uncertainties were combined in quadrature for one-to-one correlation. Additional uncertainties not reported are treated as uncorrelated between different data sets.
correlated between all $W + \text{jets}$ bins, for example, increases the $\chi^2$ by approximately 200 for 30 data points, despite insignificant changes to the resulting PDFs.

Two systematic uncertainties in the $W + \text{jets}$ data set related the $E_T^{\text{miss}}$ resolution are summed into a single component for one-to-one correlation with the $E_T^{\text{miss}}$ resolution systematic uncertainty in the 7 TeV data set. A further ten systematic uncertainties in the $V + \text{jets}$ data correspond to a single component related to the jet energy scale in the 7 TeV data; it is preferred to keep the more detailed model of ten sources in the 8 TeV data as they have a significant impact on the $V + \text{jets}$ measurements and the impact of the single source in the 7 TeV data is small.

Additionally, two systematic uncertainties related to pile-up dependence of the jet energy scale, one systematic uncertainty related to jet energy resolution and one further related to the $E_T^{\text{miss}}$ scale, are also correlated between 7 TeV and 8 TeV data sets, for a total of five correlated components. Cross-checks have been performed demonstrating that alternative models, for example partially correlating the luminosity uncertainties between 7 and 8 TeV data or leaving all systematic uncertainties uncorrelated between 7 and 8 TeV data and using combined 7 TeV $W$ and $Z$ data, provide similar resulting PDFs.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] H1 and ZEUS collaborations, Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data, Eur. Phys. J. C 75 (2015) 580 [arXiv:1506.06042] [inSPIRE].

[2] S.I. Alekhin, J. Blümlein, S. Moch and R. Plačakyte, Parton distribution functions, $\alpha_s$, and heavy-quark masses for LHC Run II, Phys. Rev. D 96 (2017) 014011 [arXiv:1701.05838] [inSPIRE].

[3] T.-J. Hou et al., New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC, Phys. Rev. D 103 (2021) 014013 [arXiv:1912.10053] [inSPIRE].

[4] L.A. Harland-Lang, A.D. Martin, P. Motylinski and R.S. Thorne, Parton distributions in the LHC era: MMHT 2014 PDFs, Eur. Phys. J. C 75 (2015) 204 [arXiv:1412.3989] [inSPIRE].

[5] NNPDF collaboration, Parton distributions from high-precision collider data, Eur. Phys. J. C 77 (2017) 663 [arXiv:1706.00428] [inSPIRE].

[6] M. Botje, A QCD analysis of HERA and fixed target structure function data, Eur. Phys. J. C 14 (2000) 285 [hep-ph/9912439] [inSPIRE].

[7] S.I. Alekhin, S.A. Kulagin and S. Liuti, Isospin dependence of power corrections in deep inelastic scattering, Phys. Rev. D 69 (2004) 114009 [hep-ph/0304210] [inSPIRE].

[8] CDF collaboration, Direct Measurement of the $W$ Production Charge Asymmetry in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. Lett. 102 (2009) 181801 [arXiv:0901.2169] [inSPIRE].
[9] D0 collaboration, Measurement of the electron charge asymmetry in $p\bar{p} \to W + X \to e\nu + X$ decays in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. D 91 (2015) 032007 [Erratum ibid. 91 (2015) 079901] [arXiv:1412.2862] [inSPIRE].

[10] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, Parton distributions for the LHC, Eur. Phys. J. C 63 (2009) 189 [arXiv:0901.0002] [inSPIRE].

[11] J. Gao et al., CT10 next-to-next-to-leading order global analysis of QCD, Phys. Rev. D 89 (2014) 033009 [arXiv:1302.6246] [inSPIRE].

[12] ATLAS collaboration, Precision measurement and interpretation of inclusive $W^+, W^-$ and $Z/\gamma^*$ production cross sections with the ATLAS detector, Eur. Phys. J. C 77 (2017) 367 [arXiv:1612.03016] [inSPIRE].

[13] NUSea collaboration, Improved measurement of the $d/\bar{u}$ asymmetry in the nucleon sea, Phys. Rev. D 64 (2001) 052002 [hep-ex/0103030] [inSPIRE].

[14] ATLAS collaboration, Measurement of the production of a $W$ boson in association with a charm quark in $pp$ collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, JHEP 05 (2014) 068 [arXiv:1402.6263] [inSPIRE].

[15] CMS collaboration, Measurement of associated production of a $W$ boson and a charm quark in proton-proton collisions at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C 79 (2019) 269 [arXiv:1811.10021] [inSPIRE].

[16] S.A. Malik and G. Watt, Ratios of $W$ and $Z$ Cross Sections at Large Boson $p_T$ as a Constraint on PDFs and Background to New Physics, JHEP 02 (2014) 025 [arXiv:1304.2424] [inSPIRE].

[17] ATLAS collaboration, Measurement of differential cross sections and $W^+/W^-$ cross-section ratios for $W$ boson production in association with jets at $\sqrt{s} = 8$ TeV with the ATLAS detector, JHEP 05 (2018) 077 [Erratum JHEP 10 (2020) 048] [arXiv:1711.03296] [inSPIRE].

[18] ATLAS collaboration, Measurement of the inclusive cross-section for the production of jets in association with a $Z$ boson in proton-proton collisions at 8 TeV using the ATLAS detector, Eur. Phys. J. C 79 (2019) 847 [arXiv:1907.06728] [inSPIRE].

[19] R. Boughezal, C. Focke, X. Liu and F. Petriello, $W$-boson production in association with a jet at next-to-next-to-leading order in perturbative QCD, Phys. Rev. Lett. 115 (2015) 062002 [arXiv:1504.02131] [inSPIRE].

[20] A. Gehrmann-De Ridder, T. Gehrmann, E.W.N. Glover, A. Huss and T.A. Morgan, Precise QCD predictions for the production of a $Z$ boson in association with a hadronic jet, Phys. Rev. Lett. 117 (2016) 022001 [arXiv:1507.02850] [inSPIRE].

[21] R.D. Ball, V. Bertone, M. Bonvini, S. Marzani, J. Rojo and L. Rottoli, Parton distributions with small-$x$ resummation: evidence for BFKL dynamics in HERA data, Eur. Phys. J. C 78 (2018) 321 [arXiv:1710.05935] [inSPIRE].

[22] M. Cacciari, G.P. Salam and G. Soyez, The anti-$k_T$ jet clustering algorithm, JHEP 04 (2008) 063 [arXiv:0802.1189] [inSPIRE].

[23] M. Cacciari, G.P. Salam and G. Soyez, FastJet User Manual, Eur. Phys. J. C 72 (2012) 1896 [arXiv:1111.6097] [inSPIRE].

[24] S.I. Alekhin et al., HERAFitter, Eur. Phys. J. C 75 (2015) 304 [arXiv:1410.4412] [inSPIRE].

[25] H1 collaboration, A Precision Measurement of the Inclusive $ep$ Scattering Cross Section at HERA, Eur. Phys. J. C 64 (2009) 561 [arXiv:0904.3513] [inSPIRE].
[26] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, *Comput. Phys. Commun.* **10** (1975) 343 [arXiv:SPIRE].

[27] ZEUS collaboration, *A ZEUS next-to-leading-order QCD analysis of data on deep inelastic scattering*, *Phys. Rev. D* **67** (2003) 012007 [hep-ex/0208023] [arXiv:SPIRE].

[28] M. Botje, *QCDNUM: Fast QCD Evolution and Convolution*, *Comput. Phys. Commun.* **182** (2011) 490 [arXiv:1005.1481] [arXiv:SPIRE].

[29] R.S. Thorne and R.G. Roberts, *An Ordered analysis of heavy flavor production in deep inelastic scattering*, *Phys. Rev. D* **67** (2003) 012007 [hep-ph/0208023] [arXiv:SPIRE].

[30] R.S. Thorne, *A Variable-flavor number scheme for NNLO*, *Phys. Rev. D* **73** (2006) 054019 [hep-ph/0601245] [arXiv:SPIRE].

[31] R.S. Thorne, *Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections*, *Phys. Rev. D* **86** (2012) 074017 [arXiv:1201.6180] [arXiv:SPIRE].

[32] T. Carli et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, *Eur. Phys. J. C* **66** (2010) 503 [arXiv:0911.2985] [arXiv:SPIRE].

[33] J.M. Campbell and R.K. Ellis, *An Update on vector boson pair production at hadron colliders*, *Phys. Rev. D* **60** (1999) 113006 [hep-ph/9905386] [arXiv:SPIRE].

[34] J.M. Campbell and R.K. Ellis, *MCFM for the Tevatron and the LHC*, *Nucl. Phys. B Proc. Suppl.* **205–206** (2010) 10 [arXiv:1007.3492] [arXiv:SPIRE].

[35] S. Catani and M. Grazzini, *An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC*, *Phys. Rev. Lett.* **98** (2007) 222002 [hep-ph/0703012] [arXiv:SPIRE].

[36] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO*, *Phys. Rev. Lett.* **103** (2009) 082001 [arXiv:0903.2120] [arXiv:SPIRE].

[37] M. Dasgupta, L. Magnea and G.P. Salam, *Non-perturbative QCD effects in jets at hadron colliders*, *JHEP* **02** (2008) 055 [arXiv:0712.3014] [arXiv:SPIRE].

[38] C. Buttar et al., *Standard Model Handles and Candles Working Group: Tools and Jets Summary Report*, in proceedings of the 5th Les Houches Workshop on Physics at TeV Colliders, 11–29 June 2007, [arXiv:0803.0678] [arXiv:SPIRE].

[39] SHERPA collaboration, *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034 [arXiv:1905.09127] [arXiv:SPIRE].

[40] M. Schönherr, *An automated subtraction of NLO EW infrared divergences*, *Eur. Phys. J. C* **78** (2018) 119 [arXiv:1712.07975] [arXiv:SPIRE].

[41] S. Kallweit, J.M. Lindert, P. Maierhofer, S. Pozzorini and M. Schönherr, *NLO QCD + EW predictions for V + jets including off-shell vector-boson decays and multijet merging*, *JHEP* **04** (2016) 021 [arXiv:1511.08692] [arXiv:SPIRE].

[42] H1 and ZEUS collaborations, *Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic ep scattering at HERA*, *Eur. Phys. J. C* **78** (2018) 473 [arXiv:1804.01019] [arXiv:SPIRE].
[43] D. Mason et al., *Measurement of the Nucleon Strange-Antistrange Asymmetry at Next-to-Leading Order in QCD from NuTeV Dimuon Data*, *Phys. Rev. Lett.* 99 (2007) 192001 [inSPIRE].

[44] S.I. Alekhin et al., *Determination of Strange Sea Quark Distributions from Fixed-target and Collider Data*, *Phys. Rev. D* 91 (2015) 094002 [arXiv:1404.6469] [inSPIRE].

[45] A. Kayis-Topaksu et al., *Measurement of charm production in neutrino charged-current interactions*, *New J. Phys.* 13 (2011) 093002 [arXiv:1107.0613] [inSPIRE].

[46] NOMAD collaboration, *A Precision Measurement of Charm Dimuon Production in Neutrino Interactions from the NOMAD Experiment*, *Nucl. Phys. B* 876 (2013) 339 [arXiv:1308.4750] [inSPIRE].

[47] NuTeV collaboration, *Precise Measurement of Dimuon Production Cross-Sections in νµFe and ¯νµFe Deep Inelastic Scattering at the Tevatron*, *Phys. Rev. D* 64 (2001) 112006 [hep-ex/0102049] [inSPIRE].

[48] F. Faura, S. Iranipour, E.R. Nocera, J. Rojo and M. Ubiali, *The Strangest Proton?*, *Eur. Phys. J. C* 80 (2020) 1168 [arXiv:2009.00014] [inSPIRE].

[49] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev. D* 93 (2016) 033006 [arXiv:1506.07443] [inSPIRE].

[50] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* 43 (2016) 023001 [arXiv:1510.03865] [inSPIRE].

[51] SeaQuest collaboration, *The asymmetry of antimatter in the proton*, *Nature* 590 (2021) 561 [arXiv:2103.04024] [inSPIRE].

[52] ATLAS collaboration, *ATLAS Computing Acknowledgements*, ATL-SOFT-PUB-2020-001 (2020).

[53] ATLAS collaboration, *Measurement of differential cross sections and W+/W− cross-section ratios for W boson production in association with jets at √s = 8 TeV with the ATLAS detector*, HEPData record (2018) [https://doi.org/10.17182/hepdata.80076].

[54] ATLAS collaboration, *Measurement of the inclusive cross-section for the production of jets in association with a Z boson in proton-proton collisions at 8 TeV using the ATLAS detector*, HEPData record (2019) [https://doi.org/10.17182/hepdata.90953].

[55] ATLAS collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at √s = 7 TeV with the ATLAS detector*, *Eur. Phys. J. C* 75 (2015) 17 [arXiv:1406.0076] [inSPIRE].

[56] ATLAS collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at √s = 8 TeV using the ATLAS detector*, *Eur. Phys. J. C* 76 (2016) 581 [arXiv:1510.03823] [inSPIRE].

[57] ATLAS collaboration, *Performance of algorithms that reconstruct missing transverse momentum in √s = 8 TeV proton-proton collisions in the ATLAS detector*, *Eur. Phys. J. C* 77 (2017) 241 [arXiv:1609.09324] [inSPIRE].

[58] ATLAS collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, *Eur. Phys. J. C* 74 (2014) 3071 [arXiv:1407.5063] [inSPIRE].

[59] ATLAS collaboration, *Performance of the ATLAS Electron and Photon Trigger in p-p Collisions at √s = 7 TeV in 2011*, ATLAS-CONF-2012-048 (2012).
[60] ATLAS collaboration, Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data, *Eur. Phys. J. C* **74** (2014) 2941 [arXiv:1404.2240] [insPIRE].

[61] ATLAS collaboration, Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data, *Eur. Phys. J. C* **77** (2017) 195 [arXiv:1612.01456] [insPIRE].

[62] ATLAS collaboration, Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC, *Eur. Phys. J. C* **73** (2013) 2518 [arXiv:1302.4393] [insPIRE].

[63] ATLAS collaboration, Luminosity determination in pp collisions at $\sqrt{s}=8$ TeV using the ATLAS detector at the LHC, *Eur. Phys. J. C* **76** (2016) 653 [arXiv:1608.03953] [insPIRE].

[64] ATLAS collaboration, Multi-Boson Simulation for 13 TeV ATLAS Analyses, ATL-PHYS-PUB-2017-005 (2017).

[65] ATLAS collaboration, Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO, ATL-PHYS-PUB-2017-007 (2017).
1 Department of Physics, University of Adelaide, Adelaide, Australia
2 Physics Department, SUNY Albany, Albany, NY, U.S.A.
3 Department of Physics, University of Alberta, Edmonton, AB, Canada
4 (a) Department of Physics, Ankara University, Ankara; (b) Istanbul Aydin University, Application and Research Center for Advanced Studies, Istanbul; (c) Division of Physics, TOBB University of Economics and Technology, Ankara, Turkey
5 LAPP, Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS/IN2P3, Annecy, France
6 High Energy Physics Division, Argonne National Laboratory, Argonne, IL, U.S.A.
7 Department of Physics, University of Arizona, Tucson, AZ, U.S.A.
8 Department of Physics, University of Texas at Arlington, Arlington, TX, U.S.A.
9 Physics Department, National and Kapodistrian University of Athens, Athens, Greece
10 Department of Physics, University of Texas at Austin, Austin, TX, U.S.A.
11 (a) Bahcesehir University, Faculty of Engineering and Natural Sciences, Istanbul; (b) Istanbul Bilgi University, Faculty of Engineering and Natural Sciences, Istanbul; (c) Department of Physics, Bogazici University, Istanbul; (d) Department of Physics Engineering, Gaziantep University, Gaziantep, Turkey
12 Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
13 Institut de Física d’Altes Energies (IFAE), Barcelona Institute of Science and Technology, Barcelona, Spain
14 (a) Institute of High Energy Physics, Chinese Academy of Sciences, Beijing; (b) Physics Department, Tsinghua University, Beijing; (c) Department of Physics, Nanjing University, Nanjing; (d) University of Chinese Academy of Science (UCAS), Beijing, China
15 Institute of Physics, University of Belgrade, Belgrade, Serbia
16 Department for Physics and Technology, University of Bergen, Bergen, Norway
17 Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley, CA, U.S.A.
18 Institut für Physik, Humboldt Universität zu Berlin, Berlin, Germany
19 Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern, Switzerland
20 School of Physics and Astronomy, University of Birmingham, Birmingham, U.K.
21 (a) Facultad de Ciencias y Centro de Investigaciones, Universidad Antonio Nariño, Bogotá; (b) Departamento de Física, Universidad Nacional de Colombia, Bogotá, Colombia, Colombia
22 (a) INFN Bologna and Universita’ di Bologna, Dipartimento di Fisica; (b) INFN Sezione di Bologna, Italy
23 Physikalisches Institut, Universität Bonn, Bonn, Germany
24 Department of Physics, Boston University, Boston, MA, U.S.A.
25 Department of Physics, Brandeis University, Waltham, MA, U.S.A.
26 (a) Transilvania University of Brasov, Brasov; (b) Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest; (c) Department of Physics, Alexandru Ioan Cuza University of Iasi, Iasi; (d) National Institute for Research and Development of Isotopic and Molecular Technologies, Physics Department, Cluj-Napoca; (e) University Politehnica Bucharest, Bucharest; (f) West University in Timisoara, Timisoara, Romania
27 (a) Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava; (b) Department of Subnuclear Physics, Institute of Experimental Physics of the Slovak Academy of Sciences, Kosice, Slovak Republic
28 Physics Department, Brookhaven National Laboratory, Upton, NY, U.S.A.
29 Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
30 California State University, CA, U.S.A.
31 Cavendish Laboratory, University of Cambridge, Cambridge, U.K.
32 (a) Department of Physics, University of Cape Town, Cape Town; (b) iThemba Labs, Western Cape; (c) Department of Mechanical Engineering Science, University of Johannesburg, Johannesburg; (d) University of South Africa, Department of Physics, Pretoria; (e) School of Physics, University of the Witwatersrand, Johannesburg, South Africa
Also at Department of Physics, California State University, Sacramento, U.S.A.
Also at Department of Physics, King’s College London, London, U.K.
Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg, Russia
Also at Department of Physics, University of Fribourg, Fribourg, Switzerland
Also at Dipartimento di Matematica, Informatica e Fisica, Università di Udine, Udine, Italy
Also at Faculty of Physics, M.V. Lomonosov Moscow State University, Moscow, Russia
Also at Faculty of Physics, Sofia University, ‘St. Kliment Ohridski’, Sofia, Bulgaria
Also at Giessen University, Faculty of Engineering, Giessen, Turkey
Also at Graduate School of Science, Osaka University, Osaka, Japan
Also at Hellenic Open University, Patras, Greece
Also at Institucio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona, Spain
Also at Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany
Also at Institute for Nuclear Research and Nuclear Energy (INRNE) of the Bulgarian Academy of Sciences, Sofia, Bulgaria
Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary
Also at Institute of Particle Physics (IPP), Canada
Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
Also at Instituto de Fisica Teorica, IFT-UAM/CSIC, Madrid, Spain
Also at Istanbul University, Department of Physics, Istanbul, Turkey
Also at Joint Institute for Nuclear Research, Dubna, Russia
Also at Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia
Also at National Research Nuclear University MEPhI, Moscow, Russia
Also at Physics Department, An-Najah National University, Nablus, Palestine
Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany
Also at The City College of New York, New York, NY, U.S.A.
Also at TRIUMF, Vancouver, BC, Canada
Also at Universita di Napoli Parthenope, Napoli, Italy
Also at University of Chinese Academy of Sciences (UCAS), Beijing, China
* Deceased