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

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 datasets considered are the ATLAS measurements of W^+, W^- and Z boson production in association with jets in pp collisions at $\sqrt{s} = 8$ TeV at the LHC with integrated luminosities of 20.2 fb^{-1} and 19.9 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 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 of similar size as the up-and down-sea quark densities in the range $x \lesssim 0.02$ found by previous ATLAS analyses.

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 four-momentum transfer, 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 better 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 (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 give 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 from mid to high-x comes from fixed-target DIS experiments, as well as $W^\pm$ and Z data 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 $p\bar{p}$ 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].

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 [11]. Differential W and Z/γ^* boson cross sections 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 with respect to the up- and down-quark sea at $x \lesssim 0.05$, in contradiction to previous assumptions based on data of dimuon production from muon-neutrino CC DIS with associated charm-quark production [12]. This observation is supported by the ATLAS measurement of W-boson production in association with a charm quark ($W + c$) at 7 TeV [13], however a recent analysis of CMS $W + c$ 7 and 13 TeV data [14] 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 than can be accessed by W and Z data alone [15]. The tree-level production modes of the production of a vector boson in association with jets, referred to as V + jets, are 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. Both production modes require a higher x and Q^2 in comparison to inclusive production, thereby yielding a complementary data set to the inclusive W, Z measurements.

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 [16, 17] in combination with the previous inclusive W, Z measurement at $\sqrt{s} = 7$ TeV [11] and the combination of 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 [18, 19], and accounts for the systematic correlations between data sets. The resulting PDF set is called *ATLASepWZVjet20*.

2 Input data sets

The final combined $e^{\pm}p$ 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 6×10^{-7}. Data below $x = 10^{-5}$ are not used in this analysis by requiring $Q^2 > 10$ GeV^2, motivated by the previously-observed worse fit quality in the removed kinematic region compared to the rest of the HERA data [1]. This was recently explained by the need for resummation corrections at low x [20], and further motivated by the impact of higher-twist corrections. 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 $Q^2 = 3\,000$ GeV^2.

The ATLAS W, 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 [11]. The $W^{\pm}$ differential cross sections are measured as functions of the W decay lepton pseudorapidity, η_{ℓ}, 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 $Z/\gamma*$ 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 ranges, with an experimental precision of up to 0.4% for central rapidity and 2.3% for forward rapidity. The luminosity of the data set used for the 7 TeV W, Z cross-section measurement is known to within 1.8%. There are a total of 131 sources of correlated systematic uncertainty across the W and Z data sets [11]. 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 the current analysis uses the data before this combination. This choice was made as the uncombined data retain the physical origin of the sources of correlated uncertainties, thereby allowing to correlate sources to 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 [16]. 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 [21, 22] 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 as 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% [16]. There are 50 sources of correlated systematic uncertainty common between 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 on the statistical bin-to-bin correlations in data 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 and a total integrated luminosity of 19.9 fb^{-1}, in the $Z \to e^+e^-$ decay channel [17]. The measurement is performed as a function of the absolute rapidity of inclusive anti-k_t R = 0.4 jets, $|y^{\text{jet}}|$, for several bins of the transverse momentum within 25 GeV < p_T^{jet} < 1050 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 relating to the data and background MC simulation statistics.

The 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 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, 23, 24]. This program interfaces to theoretical calculations directly or uses fast interpolation grids to make theoretical predictions for the considered processes. The program MINUIT [25] is used for the minimisation of the PDF fit. The results are cross-checked with an independent fit framework [26].

For the DIS processes, coefficient functions with massless quarks are calculated at NNLO as implemented in QCDNUM v17-01-13 [27]. The contributions of heavy quarks are calculated in the general-mass variable-flavour-number scheme of Refs. [28–30]. 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/γ^* boson cross sections, the theoretical framework is the same as that used in the ATLASepWZ16 analysis of Ref. [11]. The xFitter package uses outputs from the APPLGRID code [31] interfaced to the MCFM program [32, 33] 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 [34, 35].

Predictions for W + jets and Z + jets production are obtained similarly to the W, Z predictions to NLO in QCD and LO in EW 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_{jetti} program [18] 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, as discussed in the publication of the ATLAS measurement [16]. The bin-by-bin K-factors are derived as the ratio of the NNLO to the NLO calculation from N_{jetti} with the same fiducial selection as the W + jets data, multiplied by the non-perturbative correction. The renormalisation and factorisation scales are set to $\mu_{\text{R/F}} = \sqrt{m_W^2 + \Sigma(p_T^j)^2}$, where the second term in the square root is the scalar sum of the transverse momenta of the jets. More details about the predictions are given in the respective ATLAS publication [16]. In addition to these predictions, NLO EW corrections inclusive of QED radiation effects are computed using Sherpa by the authors of Refs. [36–38] and applied as additional bin-by-bin multiplicative K-factors.

Predictions for Z + jets production to NNLO in QCD and LO in EW are calculated by the authors of Ref. [19], and the K-factor is calculated as the ratio of the NNLO and NLO predictions. 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 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 of the ATLAS measurement [17], 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 v2.2.10 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 up to 20% at high p_T^W.

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, μ_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, 39], $m_c = 1.43$ GeV and $m_b = 4.5$ 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 ATLASepWZ16 fits [1, 11]

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

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, xd_v) and the light antiquark distributions ($x\bar{u}$, $x\bar{d}$, $x\bar{s}$). The gluon distribution is parameterised with the more flexible form

$$xg(x) = A_g x^{B_g} (1-x)^{C_g} P_g(x) - A'_g x^{B'_g} (1-x)^{C'_g},\tag{2}$$

where C'_g is fixed to a value of 25 to suppress negative contributions of 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 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. (1), with $P_{\bar{s}} = 1$ and $B_{\bar{s}} = B_{\bar{d}}$, leaving two free strangeness parameters, $A_{\bar{s}}$ and $C_{\bar{s}}$. 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_{u_v} and D_g. There are 16 free parameters used in total in the central fit.

The agreement of the data with the predictions from a PDF parameterisation is quantified with a χ^2. The definition of the χ^2 without statistical correlations between data points is as follows [1, 11]

$$\chi^2 = \sum_i \frac{[D_i - T_i(1 - \sum_j \gamma_{ij} b_j)]^2}{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i} + \sum_j b_j^2 + \sum_i \log \frac{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i}{\delta_{i,\text{uncor}}^2 D_i^2 + \delta_{i,\text{stat}}^2 D_i^2},\tag{3}$$

where D_i represent the measured data, T_i the corresponding theoretical prediction, $\delta_{i,\text{uncor}}$ and $\delta_{i,\text{stat}}$ are the uncorrelated systematic and statistical uncertainties on D_i, and correlated systematics, described by

γ_{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 systematics. For each data set, the first term gives the *partial* χ^2 and the second term gives the *correlated* χ^2. The third term is a bias correction term arising from the transition of the likelihood to χ^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 χ^2 definition as follows [40]

$$
\begin{aligned}
\chi^2 = \sum_{ik} & \left(D_i - T_i(1 - \sum_j \gamma_{ij} b_j) \right) C_{\text{stat},ik}^{-1}(D_i, D_k) \left(D_k - T_k(1 - \sum_j \gamma_{kj} b_j) \right) \\
& + \sum_j b_j^2 \\
& + \sum_i \log \frac{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i}{\delta_{i,\text{uncor}}^2 D_i^2 + \delta_{i,\text{stat}}^2 D_i^2} ,
\end{aligned}
\tag{4}
$$

in which the first term 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

Throughout this section, the presented ATLASepWZVjet20 PDFs are compared to a equivalent fit without the V + jets data, labelled ATLASepWZ20. Although these PDFs differ from the ATLASepWZ16 analysis by an additional parameter, D_g, a tighter selection criteria of $Q^2 > 10$ GeV^2 and the use of ATLAS 7 TeV W, Z data in which the electron and muon channels have not been 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, Z data with the combination of electron and muon channels provides a fit with very similar central values and uncertainties.

4.1 Goodness of fit and parton distributions

Figures 1 and 2 display a comparison of the W + jets and Z + jets differential cross section measurements with the predictions for each of the ATLASepWZ20 and ATLASepWZVjet20 fits. Adding the V + jets data to the fit, giving the ATLASepWZVjet20 PDFs, improves the W + jets description significantly, particularly in the W^+ spectrum where the agreement improves by approximately 20% at high p_{T}^W. The difference in partial χ^2 between the predictions of the ATLASepWZ20 and ATLASepWZVjet20 PDF sets for the W + jets and the Z + jets data is 32 and 7 units, respectively.

The total χ^2 per degree of freedom (NDF) for the ATLASepWZVjet20 fit, along with the partial χ^2 per data point (NDP) and correlated χ^2 for each data set entering the fit, is given in Table 1. The partial χ^2 for the HERA and ATLAS inclusive W, Z data in fits including the additional data are similar to those obtained in fits without these additional data, demonstrating that there is no tension between these sets. The partial χ^2 of the W + jets and Z + jets data is reasonable, and neither the HERA nor ATLAS correlated χ^2 is observed to increase significantly with the inclusion of this data.

Additional uncertainties in the PDFs are estimated and classified as model and parameterisation uncertainties. Model uncertainties comprise variations of the charm-quark mass (m_c) and bottom-quark mass (m_b),
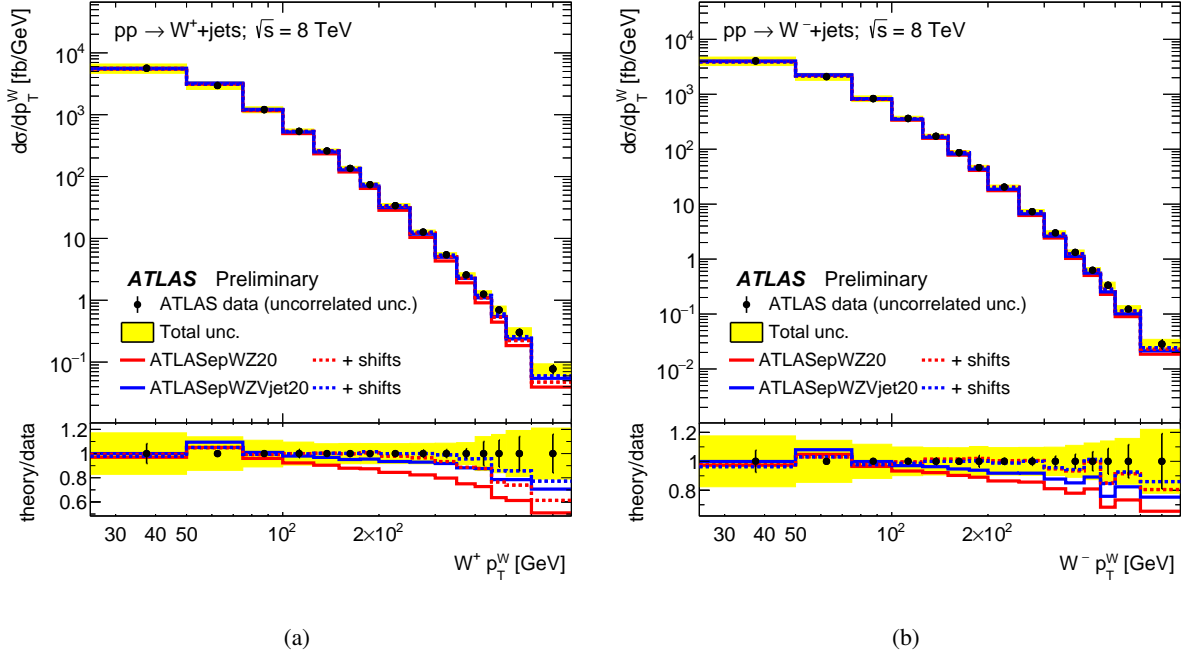
Figure 1: The differential cross sections of (a) W^+ + jets and (b) W^- + jets of Ref. [16] (black points) as a function of the transverse momentum of the W boson, p_T^W. The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties are shown as a yellow band. The cross sections are compared to 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. (4) are allowed to vary to minimise the χ^2.

Table 1: χ^2 values split by partial, correlated and log penalty for each data set entering the ATLASepWZVjet20 fit. The partial component of the χ^2 for each data set is shown compared to the number of data points of that data set (NDP).

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

variations of the minimum Q^2 cut, Q^2_{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 fulfill the condition $Q^2_0 < m_c^2$ such that the charm PDF is calculated perturbatively. Each of these variations follow that of the
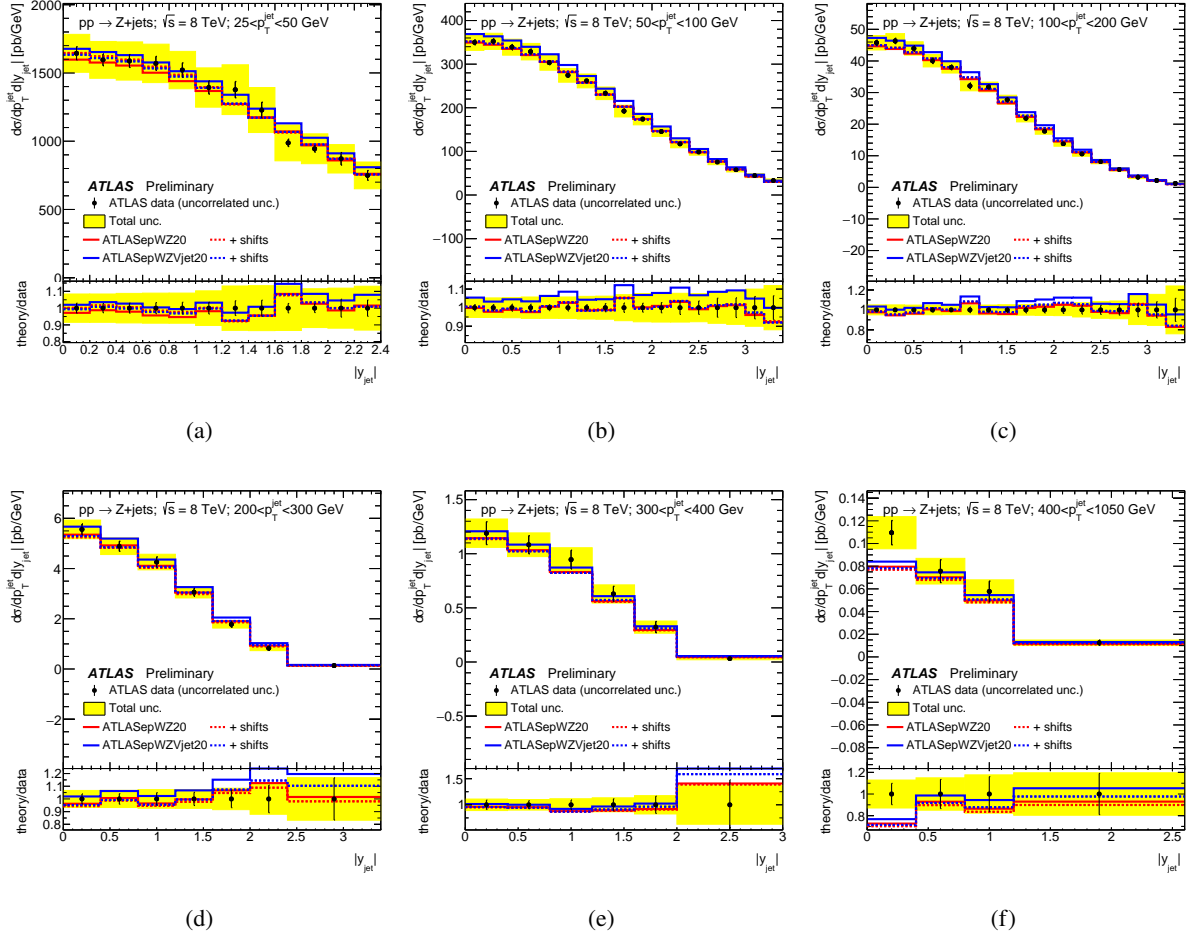
Figure 2: The differential cross sections of Z + jets as a function of the absolute rapidity of inclusive jets, $|y^{\text{jet}}|$, in bins of (a) $25 < p_{\text{T}}^{\text{jet}} < 50$ GeV, (b) $50 < p_{\text{T}}^{\text{jet}} < 100$ GeV, (c) $100 < p_{\text{T}}^{\text{jet}} < 200$ GeV, (d) $200 < p_{\text{T}}^{\text{jet}} < 300$ GeV, (e) $300 < p_{\text{T}}^{\text{jet}} < 400$ GeV and (f) $400 < p_{\text{T}}^{\text{jet}} < 1050$ GeV, respectively, 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. (4) are applied.

8

ATLASepWZ16 analysis [11]. The parameterisation uncertainties are estimated through variations which include a single further parameter in the polynomial $P_i(x)$ or relaxed constraints of the low-x sea quarks. In each variation, listed with its respective total χ^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 a measure of the total variability of the fit, only the experimental uncertainty is interpretable as a statistical standard deviation.

The effect of theoretical uncertainties in the V + jets predictions on the fit results is cross-checked. Variations of the NNLO QCD calculations are defined from the envelope of 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 + jets and Z + jets prediction upward and downward both simultaneously and individually. Each of these variations result in PDFs well within the experimental uncertainties of the nominal ATLASepWZVjet20 set.

Figure 3 show the ATLASepWZVjet20 PDFs overlaid with the ATLASepWZ20 PDFs, each evaluated at the starting scale Q_0^2, for comparison. The experimental and total uncertainties are displayed separately in each case, where the total uncertainty is calculated as the sum in quadrature of the experimental, model and parameterisation uncertainties.

The ATLASepWZVjet20 $x\bar{d}$ distribution is notably higher in the range $x \gtrsim 0.02$ compared to the ATLASepWZ20 fit. Similarly, the $x\bar{s}$ distribution of the ATLASepWZVjet20 fit in the same region is lower. Together, these changes allow for an increase in the W^+ cross-section, as depicted in Fig. 1, while maintaining a highly consistent $x\bar{D} = x\bar{d} + x\bar{s}$ distribution up to $x \sim 0.1$. Additionally, the d_v distribution is reduced at high x and increased at low x, compensating for the changes in the other PDFs and maintaining the $xD = xd_v + x\bar{d} + x\bar{s}$ distribution consistent on the overall x range. 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 ratio of Drell-Yan yields from an 800 GeV proton beam incident on liquid hydrogen and deuterium target 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 $x \sim 0.1$ [12]. 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 $x \sim 0.1$, though the uncertainties are such that it is compatible with zero within two standard deviations.

The $x(\bar{d} - \bar{u})$ distribution as function of x at $Q^2 = 1.9$ GeV^2 is shown in Fig. 4, 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 effect of the V + jets data is to provide significant constraints to 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 $x\bar{d}$ distribution, a scan of χ^2 is performed
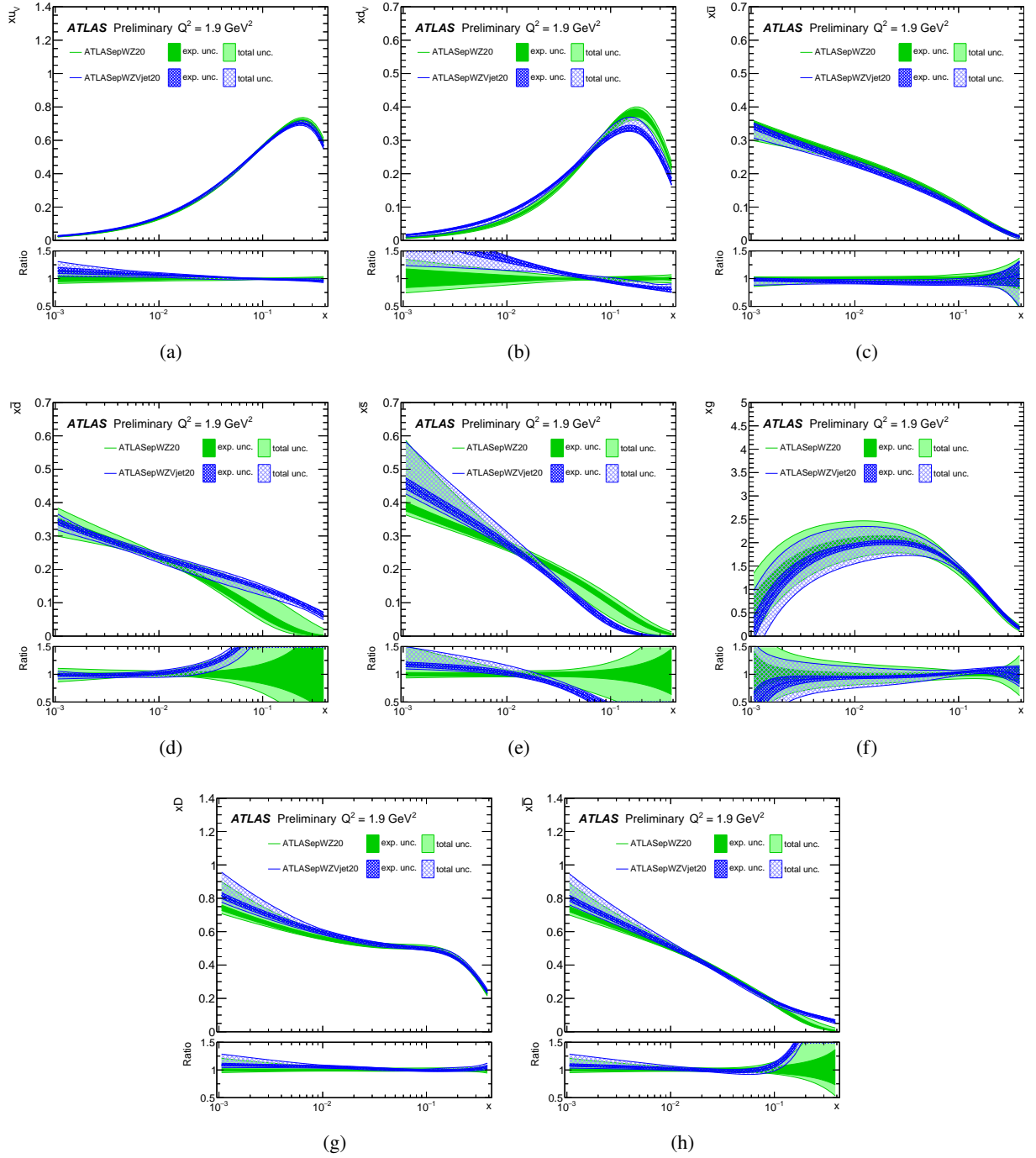
Figure 3: PDFs at the scale $Q^2 = 1.9$ GeV^2 as a function of Bjorken x obtained for the (a)-(b) valence quarks, (c)-(d) up and down sea quarks, (e) strange sea quark, (f) gluon, (g) xD and (h) $x\bar{D}$ distributions when fitting W + jets, Z + jets, inclusive W, Z and HERA data (ATLASepWZVjet20, blue bands), compared to a similar fit without W + jets or Z + 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 ratio of each to the ATLASepWZ20 PDF set is displayed in the bottom panel in each case.

Table 2: Total χ^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.

Nominal χ^2/NDF	1460 / 1198
Parameter variations	
$A_{\bar{u}} \neq A_{\bar{d}}$	1458 / 1197
$A_{\bar{u}} \neq A_{\bar{d}}$ & $B_{\bar{u}} \neq B_{\bar{d}}$	1454 / 1196
$B_{\bar{s}} \neq B_{\bar{d}}$	1459 / 1197
$B_{\bar{u}} \neq B_{\bar{d}}$	1459 / 1197
$D_{\bar{d}}$	1459 / 1197
D_{d_v}	1460 / 1197
$D_{\bar{s}}$	1460 / 1197
D_{u_v}	1457 / 1197
$E_{\bar{d}}$	1459 / 1197
$E_{\bar{s}}$	1460 / 1197
$E_{\bar{u}}$	1459 / 1197
F_{d_v}	1460 / 1197
$F_{\bar{s}}$	1460 / 1197
$F_{\bar{u}}$	1458 / 1197
F_{u_v}	1456 / 1197
Model variations	
$Q^2_{\min} = 12.5 \text{GeV}^2$	1393 / 1149
$Q^2_{\min} = 7.5 \text{GeV}^2$	1529 / 1238
$Q^2_0 = 2.2 \text{GeV}^2$ and $m_c = 1.49 \text{GeV}$	1465 / 1198
$Q^2_0 = 1.6 \text{GeV}^2$ and $m_c = 1.37 \text{GeV}$	1458 / 1198
$\alpha_s(m_Z) = 0.120$	1463 / 1198
$\alpha_s(m_Z) = 0.116$	1458 / 1198
$m_b = 4.75 \text{GeV}$	1461 / 1198
$m_b = 4.25 \text{GeV}$	1458 / 1198

through the parameter controlling the behaviour in this region, $C_{\bar{d}}$. [1] A high $C_{\bar{d}}$ value of ~ 10 corresponds to a lower $x\bar{d}$ distribution at high-x, as exhibited by the ATLASepWZ20 fit. Conversely, a low $C_{\bar{d}}$ value of ~ 2 corresponds to the higher $x\bar{d}$ distribution at high-x as exhibited by the ATLASepWZVjet20 fit.

In Fig. 5(a), this scan is shown for each of the presented PDF fits, where the χ^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 refitted and the minimum χ^2 of the scan, $\chi^2_{\min}$, is subtracted for comparison between fits.

The χ^2 of the ATLASepWZ20 fit is minimal at a value of $C_{\bar{d}} = 10 \pm 1$, whereas the χ^2 of the ATLASepWZVjet20 fit is minimal at a relatively lower $C_{\bar{d}} = 1.6 \pm 0.3$, corresponding to a higher $C_{\bar{d}}$ 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 $C_{\bar{d}} \sim 3$, corresponding to a solution similar to that of the

---

[1] The other main contributor to the difference between the fits, $C_{\bar{s}}$, could equally be considered and would provide a similar insight as these two parameters are highly correlated.
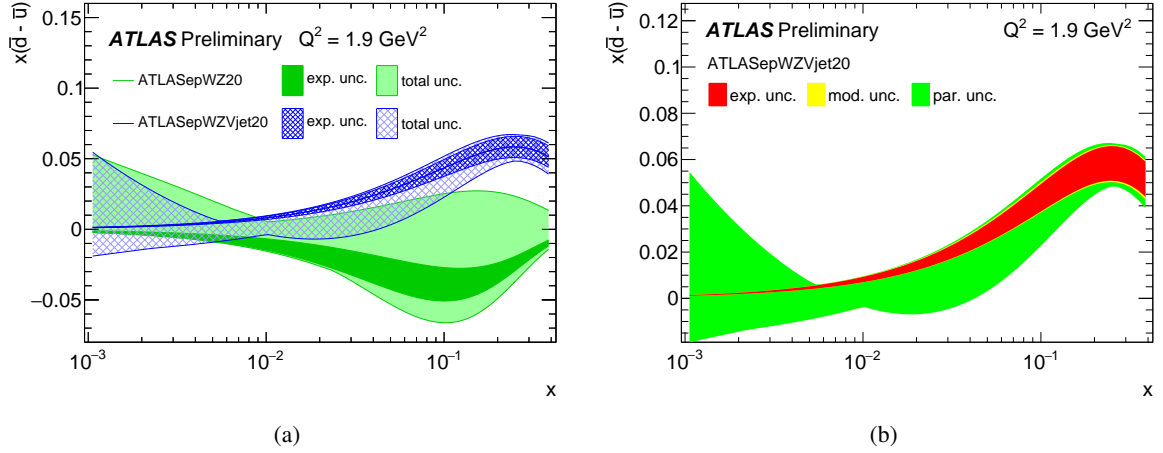
Figure 4: The $x(\bar{d} - \bar{u})$ distribution evaluated at $Q^2 = 1.9$ GeV^2 as a function of Bjorken x extracted from the ATLASepWZ20 (green) and ATLASepWZVjet20 (blue) fits with (a) experimental and total uncertainties plotted separately, and (b) extracted from the ATLASepWZVjet20 fit only with experimental, model and parameterisation uncertainties shown separately in red, yellow and green, respectively.

ATLASepWZVjet20 fit, however it exhibits a χ^2 approximately two units larger compared to the best fit. The ATLASepWZVjet20 fit fails to converge for values of $C_{\bar{d}} \gtrsim 12$ and no second minimum is observed.

In Fig. 5(b), these χ^2 distributions are decomposed into contributions from the HERA and ATLAS data. These contributions include the partial, correlated and log penalty χ^2, which are discussed in Section 3. In each fit, the ATLAS data favour a low $C_{\bar{d}}$, including in the ATLASepWZ20 fit where the overall result is a higher $C_{\bar{d}}$. Similarly, the HERA data favour the higher $C_{\bar{d}}$ value exhibited by the ATLASepWZ20 fit. The V + jets data provide sufficient constraining power in addition to the inclusive W, Z data to dominate the result and tightly constrain the $C_{\bar{d}}$ 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

$$R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}} \tag{5}$$

which uses the sum of $\bar{u}$ and $\bar{d}$ as reference for the strange-sea density.

Before the first LHC precision W, Z boson data, it was widely assumed, motivated by previous analyses of dimuon production in neutrino scattering [41–45], that the strange sea-quark density is suppressed for all x equally relative to the up and down sea over the full range of x probed, corresponding to $R_s \sim 0.5$ at $Q^2 = 1.9$ GeV^2.

The QCD analysis of the inclusive W and Z measurements by ATLAS which formed the ATLASepWZ16
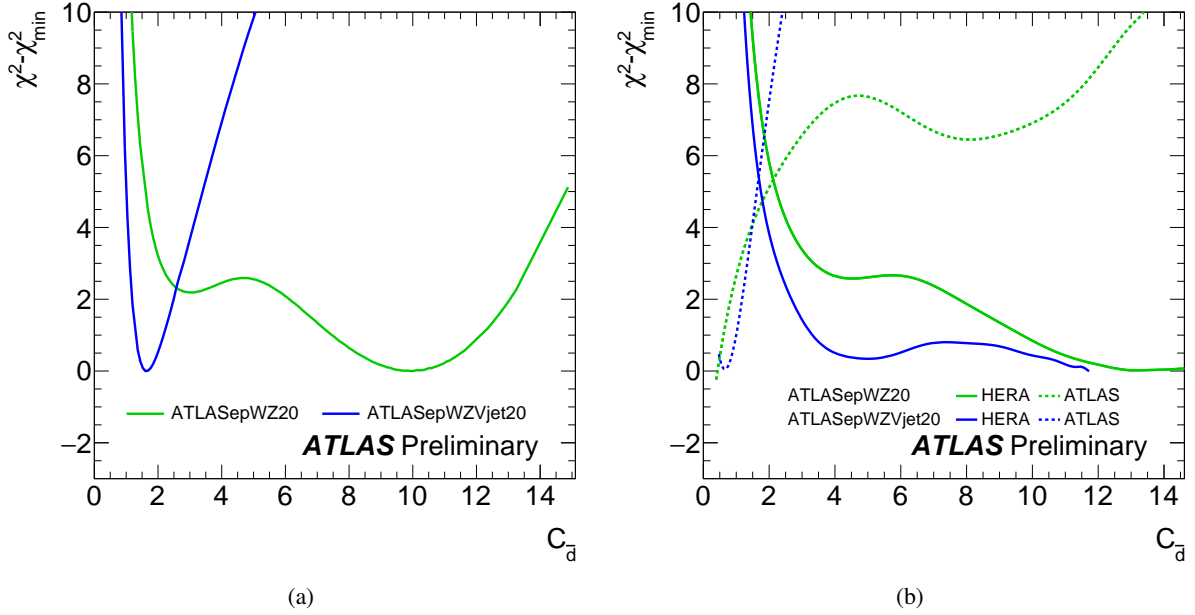
Figure 5: The χ^2 of the ATLASepWZ20 (green line) and ATLASepWZVjet20 (blue line) fits recorded as a function of the $C_{\bar{d}}$ fit parameter that determines the high-x behaviour of the $x\bar{d}$ PDF with $x\bar{d} \propto (1-x)^{C_{\bar{d}}}$. At each point, all other parameters are fitted along with the nuisance parameters corresponding to experimental systematics, and the lowest recorded χ^2, $\chi^2_{\min}$, subtracted. The total χ^2 is shown in Fig. (a), whereas the χ^2 in Fig. (b) is separated into contributions from HERA (solid lines) and ATLAS (dashed lines) data and smoothed.

PDF set led to the observation that strangeness is unsuppressed at low x ($\lesssim 0.023)^2$ for $Q^2 = 1.9$ GeV^2. This was the case for the ATLASepWZ16 fit for every variation of parameterisation used. Furthermore, a Hessian profiling exercise of the global PDFs MMHT14 [4] and CT14 [46] demonstrated that the data constrains and increases the ratio of the strange to up and down sea [11]. 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 $V+$ jets data on the strange-quark density.

The R_s distribution plotted as a function of x evaluated at $Q^2 = 1.9$ GeV^2 is shown in Fig. 6, with a comparison between ATLASepWZVjet20 and ATLASepWZ20 showing the direct effect of the $V+$ jets data, and with the experimental, model and parameterisation uncertainties of ATLASepWZVjet20 shown separately. The effect of the $V+$ 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 distribution 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 $Q^2 = 1.9$ GeV^2, the fit with the $V+$ jets data maintains an unsuppressed strange-quark density compatible with the ATLASepWZ16 fit. Fitted values of R_s, evaluated at $x = 0.023$ and $Q^2 = 1.9$ GeV^2, are given in Table 3, respectively.

---

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

Table 3: Fitted values of R_s, evaluated at $x = 0.023$ and $Q^2 = 1.9$ GeV^2, for each of the investigated fits compared to the ATLASepWZ16 result.

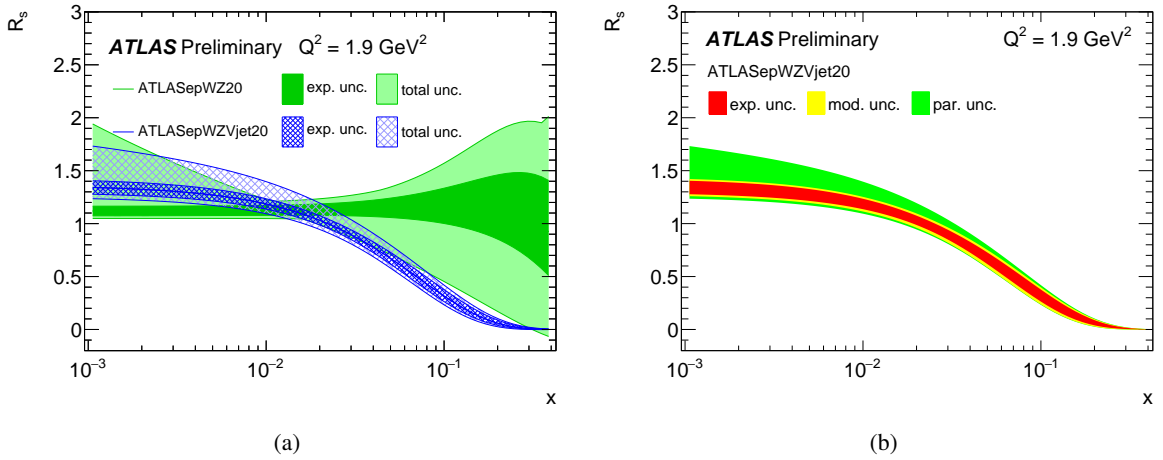| Fit	R_s	Uncertainties		
		Experimental	Model	Parameterisation
ATLASepWZ16	1.13	0.05	0.03	$+0.01$ -0.06
ATLASepWZ20	1.13	0.06	0.03	$+0.09$ -0.17
ATLASepWZVjet20	0.99	0.04	$+0.05$ -0.06	$+0.14$ -0.05

Figure 6: The R_s distribution, evaluated at $Q^2 = 1.9$ GeV^2, extracted from the ATLASepWZ20 (green) and ATLASepWZVjet20 (blue) fits with (a) experimental and total uncertainties plotted separately, and (b) extracted from the ATLASepWZVjet20 fit only with experimental, model and parameterisation uncertainties shown separately in red, yellow and green, respectively.

4.4 Comparison to global PDFs

A comparison between the ATLASepWZVjet20 R_s distribution over a range in x is shown in Fig. 7 in comparison to the global PDF sets ABMP16 [2], CT18, CT18A [3], MMHT14 [4], NNPDF3.1 [5] and the recent update of the NNPDF3.1 fit with some additional data including the full ATLAS 7 TeV data set labelled NNPDF3.1_strange [47]. 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 $10^{-2} \lesssim x \lesssim 10^{-1}$ with those which do not use the full ATLAS 7 TeV data set. This is highlighted in summary plots of R_s evaluated at $x = 0.023$, $Q^2 = 1.9$ GeV^2 and $x = 0.013$, $Q^2 = m_Z^2$ in Fig. 8. However, better agreement is observed with the CT18A PDF set, which includes the data used in the CT18 fit with the addition of ATLAS 7 TeV data, though 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 to zero, at a steeper rate than that of the global analyses.

In Fig. 9 the extracted $x(\bar{d} - \bar{u})$ distribution at $Q^2 = 1.9$ GeV^2 is shown in comparison to the results of the latest global PDF sets, all of which use E866 data. The ATLASepWZVjet20 PDF set shows consistency
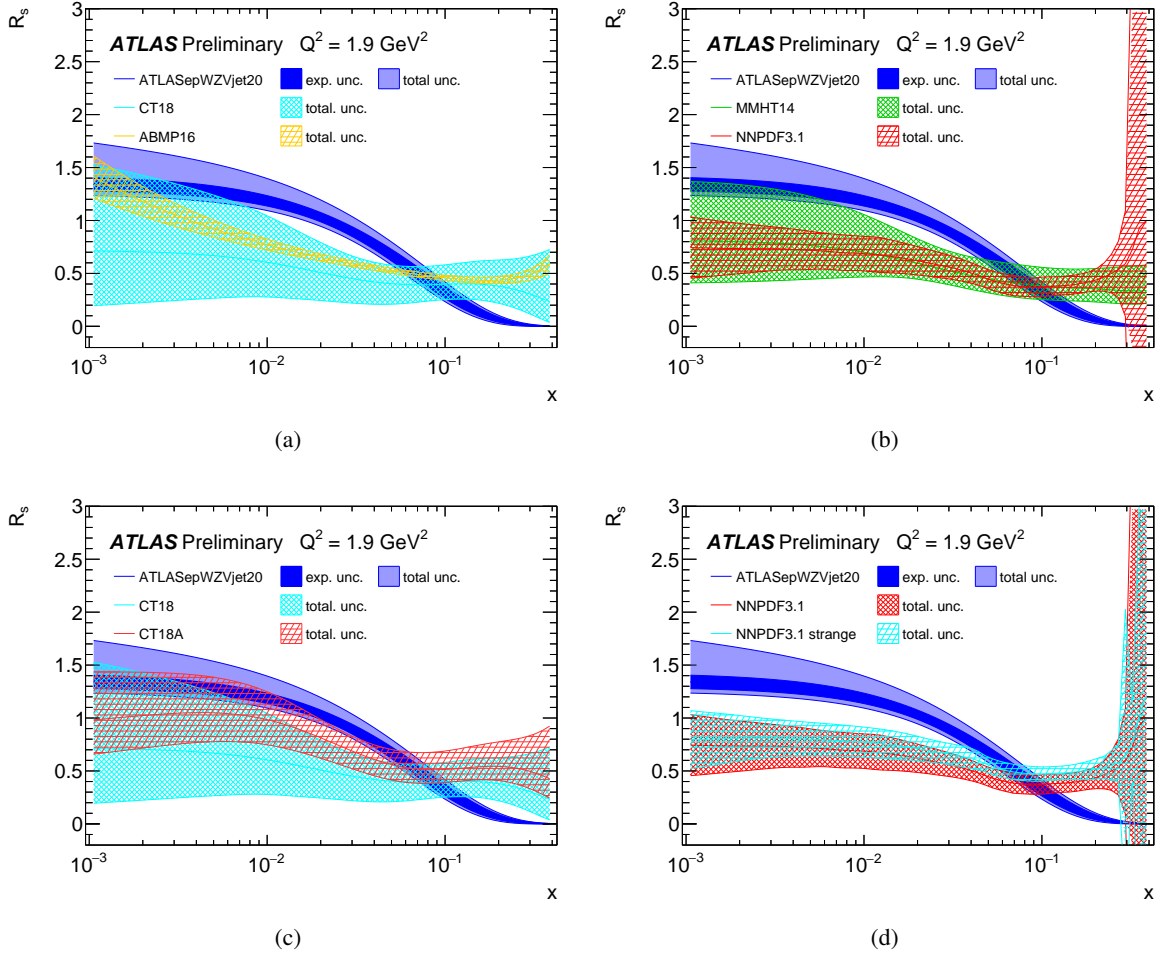
14

(a)

(b)

(c)

(d)

Figure 7: The $R_s = (s + \bar{s})/(\bar{u} + \bar{d})$ distribution evaluated at $Q^2 = 1.9$ GeV^2 as a function of Bjorken x, for the ATLASepWZVjet20 PDF set in comparison to global PDFs ABMP16 and CT18 in Fig. (a), and MMHT14 and NNPDF3.1 in Fig. (b), with additional comparisons between CT18 and CT18A (c), NNPDF3.1 and NNPDF3.1_strange (d) [2–5, 47]. 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 [48].
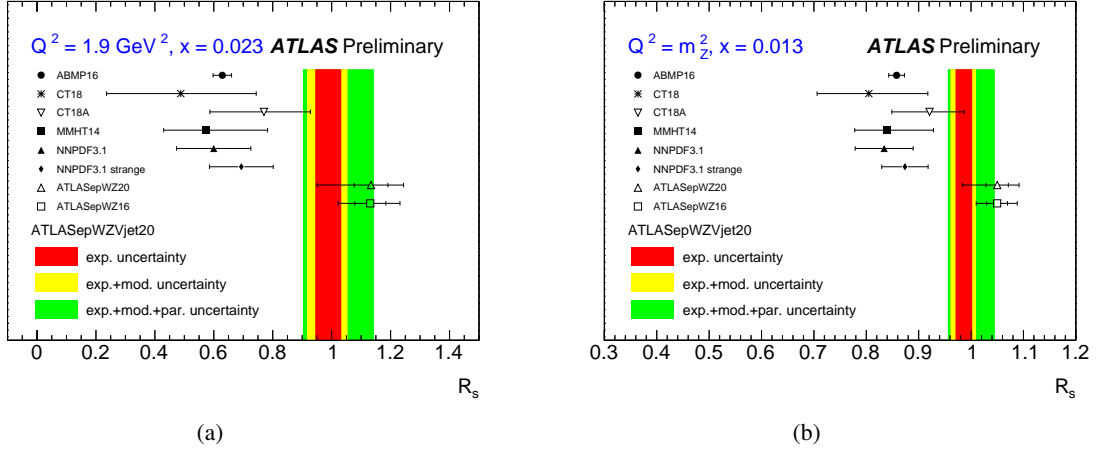
(a)   (b)

Figure 8: Summary plots of R_s evaluated at (a) $x = 0.023$ and $Q^2 = 1.9$ GeV^2, and (b) $x = 0.013$ and $Q^2 = m_Z^2$, for the ATLASepWZVjet20 PDF set in comparison to global PDFs [2–5, 47], and the ATLASepWZ16 and ATLASepWZ20 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 [48].

with these global PDF sets up to $x \sim 0.1$, but deviates in the range $0.1 < x < 0.3$ where the W + jets and Z + jets data are most sensitive and demonstrate a preference for a higher $x\bar{d}$ distribution as discussed in Section 4.1. Whereas the R_s distribution of the CT18A 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 Figs. 7 and 8, this is not replicated in the $x(\bar{d} - \bar{u})$ distribution in either case.
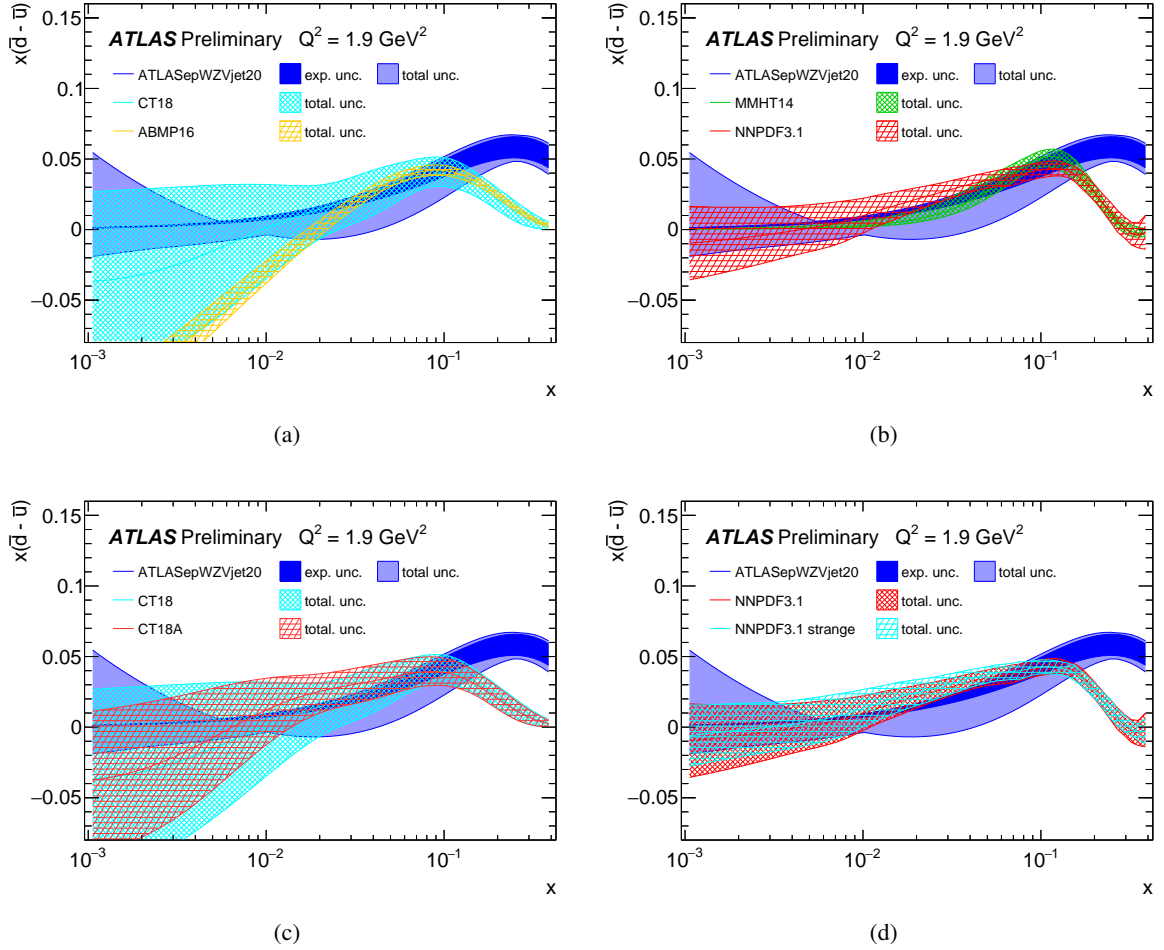
Figure 9: The $x(\bar{d} - \bar{u})$ distribution evaluated at $Q^2 = 1.9\ \text{GeV}^2$ as a function of Bjorken x, for the ATLASepWZVjet20 PDF set in comparison to global PDFs ABMP16 and CT18 in Fig. (a), and MMHT14 and NNPDF3.1 in Fig. (b), with additional comparisons between CT18 and CT18A (c), NNPDF3.1 and NNPDF3.1_strange (d) [2–5, 47]. 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 [48].

5 Conclusion

This paper presents the impact of the ATLAS data 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 + jets data has been measured in pp-collisions at $\sqrt{s}$ = 8 TeV corresponding to approximately 20 fb^{-1} of integrated luminosity. The data were fitted together 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 TeV. For the new ATLASepWZVjet20 PDF set, all significant systematic correlations between 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 to light quarks, R_s, is better constrained, falling 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 differing at higher values by up to two standard deviations. At low $x \lesssim 0.023$, the fit shows consistency with an unsuppressed strange PDF as observed in the ATLASepWZ16 PDF set, while maintaining a positive $x(\bar{d} - \bar{u})$ distribution at high x.

Appendix

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 respective ATLAS publications.

Table 4: Correlation model for the systematic uncertainties of the ATLAS measurements of W and Z + jets at 8 TeV and inclusive W, Z at 7 TeV. Entries in different rows are uncorrelated with each other. Entries within a row are fully correlated. Each entry is given the label of the systematic uncertainty used in the respective data set. Where entries are omitted, that systematic either does not exist for that dataset (denoted by a '-') or it was left decorrelated from the others (denoted by a '*'). Where a single row in one column corresponds to multiple rows in others, those systematics were combined in quadrature for one-to-one correlation.

Systematic	7 TeV inclusive W, Z	8 TeV W + jets	8 TeV Z + jets
		JetScaleEff1	ATL_JESP1
		JetScaleEff2	ATL_JESP2
		JetScaleEff3	ATL_JESP3
		JetScaleEff4	ATL_JESP4
		JetScaleEff5	ATL_JESP5
	*	JetScaleEff6	ATL_JESP6
Jet scale		JetScaleEta1	ATL_JESP7
		JetScaleEta2	ATL_JESP8
		JetScaleHighPt	ATL_JESP9
		JetScaleMC	ATL_JESP10
	JetScaleNPV	JetScalePileup1	ATL_PU_OffsetNPV
	JetScaleMu	JetScalePileup2	ATL_PU_OffsetMu
Jet punchthrough	-	JetScalepunchT	ATL_PunchThrough
Jet resolution	JetRes	JetResolution10	ATL_JER
Jet flavour composition	-	JetScaleFlav1Known	ATL_Flavor_Comp
Jet flavour response	-	JetScaleFlav2	ATL_Flavor_Response
Jet JVF selection	-	JetJVFcut	ATL_JVF
E_T^{miss} scale	MetScaleWen	METScale	-
E_T^{miss} resolution	MetRes	METResLong	-
		METResTrans	-
Electron scale	*	ElScaleZee	ATL_ElecEnZee
Electron trigger scale factor	*	ElSFTrigger	ATL_Trig
Electron reconstruction scale factor	*	ElSFReco	ATL_RecEff
Electron ID scale factor	*	ElSFId	ATL_IDEff
Luminosity	*	LumiUncert	ATL_lumi_2012_8TeV
WW background cross section	*	XsecDibos	ATL_WW_xs
Top background cross section	*	XsecTop	ATL_ttbar_xs

Systematics related to jet scale, jet resolution, JVT, missing transverse energy (E_T^{miss}) scale, E_T^{miss} resolution, and electron scale are assumed to have a one-to-one correlation between the 8 TeV data sets. Additionally, systematics related to scale factors (trigger, reconstruction and isolation) as well as luminosity are also assumed to exhibit a one-to-one correlation between datasets of the same centre-of-mass energy, but are uncorrelated between 7 TeV and 8 TeV data. There is one systematic related to the top-background

cross-section in W + jets data; as $t\bar{t}$ is the largest contributor to this background, this systematic is given one-to-one correlation with the $t\bar{t}$ cross section systematic in the Z + jets data, which is itself at least an order of magnitude larger than the single-top cross-section systematics in each bin. Similarly, the systematic related to the total diboson cross section in the W + jets data is given a one-to-one correlation with the systematic in the Z + jets data related to the production of two W bosons, as this is the highest background contribution; the additional diboson systematics in the Z + jets data related to WZ and ZZ production are left uncorrelated from W + jets data. In contrast, systematics related to the multijet background are estimated independently in each measurement, and are therefore left uncorrelated.

For each of the above sets of correlations, it has been confirmed that the effect on the resulting PDFs is not significant.

The two systematic uncertainties in each of the W + jets and Z + jets spectra related to unfolding[3] 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 correlated between all W + jets bins, for example, increases the χ^2 by approximately 200 for 30 data points, despite insignificant changes to the resulting PDFs.

Two systematics in the W + jets data set related the E_T^{miss} resolution are summed to a single component for one-to-one correlation with the E_T^{miss} resolution systematic in the 7 TeV data set. A further ten systematics in the V + jets data correspond to a single component related to jet scale in the 7 TeV data; these could be summed to a single component in the same way, but have not been as the 7 TeV systematic is relatively small and the constraining power would be reduced by breaking information about ten correlations in return for one.

Additionally, two systematics related to pileup-dependence of the jet energy scale, one systematic related to jet resolution and one further related to the E_T^{miss} energy 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 systematics uncorrelated between 7 and 8 TeV data and using combined 7 TeV W, Z data, provide similar results.

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 (cit. on pp. 2–5).

[2] S. Alekhin, J. Blümlein, S. Moch and R. Plačakyte, *Parton distribution functions, α_s, and heavy-quark masses for LHC Run II*, Phys. Rev. D **96** (2017) 014011, arXiv: 1701.05838 (cit. on pp. 2, 14–17).

[3] T.-J. Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, (2019), arXiv: 1912.10053 [hep-ph] (cit. on pp. 2, 13–17).

[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 (cit. on pp. 2, 13–17).

[5] The NNPDF Collaboration, *Parton distributions from high-precision collider data*, Eur. Phys. J. C **77** (2017) 663, arXiv: 1706.00428 (cit. on pp. 2, 14–17).

---

[3] "Unfolding" here refers to the procedure used to correct reconstructed data to particle-level cross sections.

[6] M. Botje, *A QCD analysis of HERA and fixed target structure function data*, Eur. Phys. J. C **14** (2000) 285, arXiv: `hep-ph/9912439` (cit. on p. 2).

[7] S. Alekhin, S. A. Kulagin and S. Liuti, *Isospin dependence of power corrections in deep inelastic scattering*, Phys. Rev. D **69** (2004) 114009, arXiv: `hep-ph/0304210` (cit. on p. 2).

[8] CDF Collaboration, *Direct Measurement of the W Production Charge Asymmetry in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 TeV*, Phys. Lett. **102** (2009) 181801, arXiv: `0901.2169 [hep-ex]` (cit. on p. 2).

[9] D0 Collaboration, *Measurement of the electron charge asymmetry in $p\bar{p} \rightarrow W + X \rightarrow e\nu + X$ decays in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV*, Phys. Rev. D **91** (2015) 032007, arXiv: `1412.2862 [hep-ex]` (cit. on p. 2).

[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 [hep-ph]` (cit. on pp. 2, 5).

[11] ATLAS Collaboration, *Precision measurement and interpretation of inclusive W^+, W^- and Z/γ^* production cross sections with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 367, arXiv: `1612.03016 [hep-ex]` (cit. on pp. 2–5, 9, 13).

[12] FNAL E866/NuSea Collaboration, *Improved measurement of the $\bar{d}/\bar{u}$ asymmetry in the nucleon sea*, Phys. Rev. D **64** (2001) 052002, arXiv: `hep-ex/0103030` (cit. on pp. 2, 9).

[13] 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 [hep-ex]` (cit. on p. 2).

[14] 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 [hep-ex]` (cit. on p. 2).

[15] S. Alam 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 **14** (2014) 25, arXiv: `1304.2424` (cit. on p. 2).

[16] 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, arXiv: `1711.03296 [hep-ex]` (cit. on pp. 3, 4, 7).

[17] 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 [hep-ex]` (cit. on pp. 3–5).

[18] 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 [hep-ph]` (cit. on pp. 3, 4).

[19] 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 [hep-ph]` (cit. on pp. 3, 4).

[20] R. D. Ball et al., *Parton distributions with small-x resummation: evidence for BFKL dynamics in HERA data*, Eur. Phys. J. C **78** (2018) 321, arXiv: `1710.05935` (cit. on p. 3).

[21] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: `0802.1189 [hep-ph]` (cit. on p. 3).

[22] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: `1111.6097 [hep-ph]` (cit. on p. 3).

[23] S. Alekhin et al., *HERAFitter. Open source QCD fit project*, Eur. Phys. J. C **75**, 304 (2015) 304, arXiv: `1410.4412 [hep-ph]` (cit. on p. 4).

[24] H1 Collaboration, *A precision measurement of the inclusive ep scattering cross section at HERA*, Eur. Phys. J. C **64** (2009) 561, arXiv: `0904.3513` (cit. on p. 4).

[25] 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 (cit. on p. 4).

[26] ZEUS Collaboration, *ZEUS next-to-leading-order QCD analysis of data on deep inelastic scattering*, Phys. Rev. D **67** (2003) 012007, arXiv: `hep-ex/0208023` (cit. on p. 4).

[27] M. Botje, *QCDNUM: Fast QCD evolution and convolution*, Comp. Phys. Comm. **182** (2011) 490, arXiv: `1005.1481` (cit. on p. 4).

[28] R. S. Thorne and R. G. Roberts, *Ordered analysis of heavy flavor production in deep-inelastic scattering*, Phys. Rev. D **57** (1998) 6871, arXiv: `hep-ph/9709442` (cit. on p. 4).

[29] R. S. Thorne, *Variable-flavour number scheme for next-to-next-to-leading order*, Phys. Rev. D **73** (2006) 054019, arXiv: `hep-ph/0601245` (cit. on p. 4).

[30] 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` (cit. on p. 4).

[31] 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` (cit. on p. 4).

[32] J. M. Campbell and R. K. Ellis, *Update on vector boson pair production at hadron colliders*, Phys. Rev. D **60** (1999) 113006, arXiv: `hep-ph/9905386` (cit. on p. 4).

[33] J. M. Campbell and R. K. Ellis, *MCFM for the Tevatron and the LHC*, Nucl. Phys. Proc. Suppl. **10** (2010) 205, arXiv: `1007.3492` (cit. on p. 4).

[34] S. Catani and M. Grazzini, *Next-to-Next-to-Leading-Order Subtraction Formalism in Hadron Collisions and its Application to Higgs-Boson Production at the Large Hadron Collider*, Phys. Rev. Lett. **98** (2007) 222002, arXiv: `hep-ph/0703012` (cit. on p. 4).

[35] S. Catani et. al., *Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order*, Phys. Rev. Lett. **103** (2009) 082001, arXiv: `0903.2120` (cit. on p. 4).

[36] E. Bothmann et al., *Event Generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: `1905.09127 [hep-ph]` (cit. on p. 4).

[37] M. Schönherr, *An automated subtraction of NLO EW infrared divergences*, Eur. Phys. J. C **78** (2018) 119, arXiv: `1712.07975 [hep-ph]` (cit. on p. 4).

[38] 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 [hep-ph]` (cit. on p. 4).

[39] 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 [hep-ex]` (cit. on p. 5).

[40] ATLAS Collaboration, *Determination of the parton distribution functions of the proton from ATLAS measurements of differential W and Z/γ^* boson and $t\bar{t}$ cross sections*, ATL-PHYS-PUB-2018-017, 2018, URL: `https://cds.cern.ch/record/2633819` (cit. on p. 6).

[41] 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 (cit. on p. 12).

[42] S. Alekhin et al., *Determination of strange sea quark distributions from fixed-target and collider data*, Phys. Rev. D **91** (2015) 094002, arXiv: `1404.6469 [hep-ph]` (cit. on p. 12).

[43] A. Kayis-Topaksu et al., *Measurement of charm production in neutrino charged-current interactions*, New J. Phys. **13** (2011) 093002, arXiv: `1107.0613 [hep-ex]` (cit. on p. 12).

[44] O. Samoylov et al., *A precision measurement of charm dimuon production in neutrino interactions from the NOMAD experiment*, Nucl. Phys. B **876** (2013) 339, arXiv: `1308.4750 [hep-ex]` (cit. on p. 12).

[45] M. Goncharov et al., *Precise measurement of dimuon production cross sections in ν_μ Fe and $\bar{\nu}_\mu$ Fe deep inelastic scattering at the Fermilab Tevatron.*, Phys. Rev. D **64** (2001) 112006, arXiv: `hep-ex/0102049` (cit. on p. 12).

[46] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. D **93** (2016) 033006, arXiv: `1506.07443 [hep-ph]` (cit. on p. 13).

[47] F. Faura, S. Iranipour, E. R. Nocera, J. Rojo and M. Ubiali, *The Strangest Proton?*, (2020), arXiv: `2009.00014 [hep-ph]` (cit. on pp. 14–17).

[48] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: `1510.03865 [hep-ph]` (cit. on pp. 15–17).