

QCD analysis and measurement of W boson production in association with jets at the ATLAS detector

Gavin Pownall

Christ Church

University of Oxford

*A thesis submitted for the degree of
Doctor of Philosophy*

Michaelmas 2020

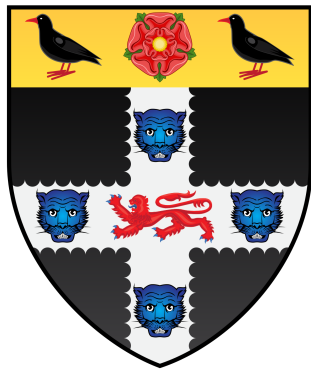
Abstract

The Standard Model of particle physics stands as one of the triumphs of the 20th and 21st century so far. Colliders around the world have been built for the purpose of pushing this theory to its limit, and have consistently found their measurements to be well described. Yet outside of colliders, there is strong evidence that the Standard Model does not completely describe the universe in which we reside. This provides the motivation to continue to push even further, with the hope of resolving these gaping inconsistencies.

At this time, many measurements are no longer limited by the power or experimental precision of our colliders – it is our limited ability to calculate predictions from the Standard Model which is holding us back. To combat this, great effort is being undertaken to measure processes which will help us to calibrate and better understand the finer details of the theory.

The production of a W boson in association with jets in proton-proton collisions serves as a precision test of perturbative quantum chromodynamics, and provides access to the composition of the proton. In this thesis, a measurement of this process using the ATLAS detector at the LHC for proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV and integrated luminosity of 36.1 fb^{-1} is presented. Results are split by the charge of the W boson and shown for jet multiplicities up to five in comparison with calculations to next-to-leading order in quantum chromodynamics scaled to the next-to-next-to-leading order cross section. In addition, the results of a similar measurement at a centre-of-mass energy of $\sqrt{s} = 8$ TeV have been used for the first time in an analysis of the structure of the proton, giving a significantly improved determination of the sea-quark densities at medium-to-high Bjorken x compared to other fits of ATLAS data.

QCD analysis and measurement of W boson production in association with jets at the ATLAS detector



Gavin Pownall
Christ Church
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2020

Abstract

The Standard Model of particle physics stands as one of the triumphs of the 20th and 21st century so far. Colliders around the world have been built for the purpose of pushing this theory to its limit, and have consistently found their measurements to be well described. Yet outside of colliders, there is strong evidence that the Standard Model does not completely describe the universe in which we reside. This provides the motivation to continue to push even further, with the hope of resolving these gaping inconsistencies.

At this time, many measurements are no longer limited by the power or experimental precision of our colliders – it is our limited ability to calculate predictions from the Standard Model which is holding us back. To combat this, great effort is being undertaken to measure processes which will help us to calibrate and better understand the finer details of the theory.

The production of a W boson in association with jets in proton-proton collisions serves as a precision test of perturbative quantum chromodynamics, and provides access to the composition of the proton. In this thesis, a measurement of this process using the ATLAS detector at the LHC for proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV and integrated luminosity of 36.1 fb^{-1} is presented. Results are split by the charge of the W boson and shown for jet multiplicities up to five in comparison with calculations to next-to-leading order in quantum chromodynamics scaled to the next-to-next-to-leading order cross section. In addition, the results of a similar measurement at a centre-of-mass energy of $\sqrt{s} = 8$ TeV have been used for the first time in an analysis of the structure of the proton, giving a significantly improved determination of the sea-quark densities at medium-to-high Bjorken x compared to other fits of ATLAS data.

Contents

List of Figures	ix
1 Introduction	1
2 Theory	5
2.1 The Standard Model of Particle Physics	5
2.1.1 Quantum electrodynamics	8
2.1.2 The Weak force	10
2.1.3 Electroweak theory	13
2.1.4 Quantum chromodynamics	17
2.2 The structure of the proton – Parton Distribution Functions	19
2.3 The production of a W boson in association with jets at a proton-proton collider	23
3 The ATLAS Experiment at the LHC	29
3.1 The Large Hadron Collider	31
3.2 The ATLAS detector	33
3.2.1 The Inner Detector	34
3.2.2 The Calorimeters	37
3.2.3 The Muon Spectrometer	39
3.2.4 The Trigger and Data Acquisition	40
3.3 Radiation damage studies of the SemiConductor Tracker optical links	41
3.3.1 Studies prior to LHC operation	42
3.3.2 Analysis method	43
3.3.3 Results	49
3.3.4 Conclusion	54
4 W + jets reconstruction and analysis strategy	59
4.1 Particle reconstruction	61
4.1.1 Electrons	61
4.1.2 Jets	66
4.1.3 E_T^{miss}	72
4.2 W + jets analysis strategy	73

4.2.1	Event selection	74
4.2.2	Distributions to be analysed	77
4.2.3	Monte Carlo simulations	79
5	Background estimation	81
5.1	The estimation of fake and non-prompt leptons	84
5.1.1	Sources of fake and non-prompt electrons	84
5.1.2	The Matrix Method	85
5.1.3	The measurement of the fake efficiency	86
5.1.4	Results and discussion	91
6	Deconvolution/Unfolding	101
6.1	Iterative Bayesian unfolding	103
6.2	Response matrices	104
6.3	Closure tests	105
6.4	Selecting the number of iterations	105
7	Estimation of uncertainties	111
7.1	Statistical uncertainties	112
7.2	Systematic uncertainties	113
7.2.1	Electron uncertainties	114
7.2.2	Jet uncertainties	115
7.2.3	E_T^{miss} uncertainties	116
7.2.4	Cross-section uncertainties	116
7.2.5	Multijet-estimation uncertainties	117
7.2.6	Breakdown of total uncertainty	117
7.3	Statistical correlations	117
7.3.1	The bootstrap method	119
7.3.2	Results	120
8	W + jets cross section results	125
9	Parton Distribution Functions with V + jets	135
9.1	Introduction	136
9.2	Input data sets	140
9.3	Fit framework	142
9.4	Results	146
9.4.1	Goodness of fit and parton distributions	146
9.4.2	The high- x sea-quark distributions	152
9.4.3	Strange-quark density	154
9.4.4	Comparison with global PDFs	156

10 Summary and conclusion	161
Appendices	
A Photo-current stability	167
B Additional material for the $W + \text{jets}$ analysis at $\sqrt{s} = 13$ TeV	173
C ATLASepWZVjet20 parton distribution functions at the Z -mass scale	187
D Systematic correlations between data sets in the ATLASepWZV-jet20 fit	191
E Cross checks of the theoretical predictions of data used in the ATLASepWZVjet20 fit	199

List of Figures

2.1	The elementary-particle content of the SM. For each particle, the mass, charge and spin is given. The table is split into four sections: the quarks consisting of the up, down, charm, strange, top and bottom particles; the leptons consisting of the electron, muon, tau and their associated neutrino partners; the spin-1 gauge bosons consisting of the gluon, photon, Z and W bosons; and the Higgs boson. [22] . . .	7
2.2	Feynman diagram representing the interaction vertex obtained from the interaction term of the QED Lagrangian.	9
2.3	Feynman diagram representing the interaction vertex obtained from the interaction term of the Weak Lagrangian.	11
2.4	Feynman diagrams depicting the absorption of the higher-order corrections in to the gluon propagator, resulting in the dependence of α_S on the energy scale of the interaction, Q^2	19
2.5	A summary of the measurements of α_S as a function of the energy scale Q across a range of experiments [22].	20
2.6	Feynman diagrams depicting the (a) quark-quark, (b) gluon-quark, (c) quark-gluon and (d) gluon-gluon LO splitting kernels of the DGLAP evolution equations.	23
2.7	A pie chart displaying the branching ratios of the W boson decay. The decay modes themselves are labelled on the perimeter of the chart.	25
2.8	Feynman diagrams of the main production modes, initial-state radiation (a) and quark-gluon fusion (b), of the $W + \text{jets}$ process at a proton-proton collider.	26
2.9	A summary of Standard Model measurements taken by the ATLAS collaboration, ordered by production rate [48]. Each process is labelled on the x -axis. W production has a rate lower than only γ and jets production; those high statistics are also present in $W + \text{jets}$ production.	27
3.1	A graphic showing the CERN accelerator complex including the LINAC2, BOOSTER, PS, SPS, LHC and collision points[58]. . . .	32

3.2	A computer-generated image of the ATLAS detector [50]. The three inner-detector components, the calorimeters and muon chambers, as well as the solenoidal and toroidal magnets, are annotated. The coordinate system is drawn in red arrows, with angles labelled in blue (see text for more details).	34
3.3	A graphic showing how the pseudorapidity, η , varies with the polar angle, θ [59].	35
3.4	A computer-generated image of the ATLAS calorimeter system [50]. The subcomponents are annotated.	38
3.5	The geometry of the SCT, with the regions used in this analysis labelled. The number of modules in each region is also labelled. The SCT is symmetric, with endcap C reflecting endcap A [66].	41
3.6	On-detector VCSELS: Barrels. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018	50
3.7	On-detector VCSELS: Endcap A. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018	51
3.8	On-detector VCSELS: Endcap C. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018	52
3.9	A plot to show the measured reduction in slope efficiency per fb^{-1} per SCT region, against the fluence per SCT region, fitted linearly.	53
3.10	On-detector $p-i-n$ diodes: Barrels. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018.	54
3.11	On-detector $p-i-n$ diodes: Endcap A. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018.	55

3.12	On-detector p-i-n diodes: Endcap C. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018.	56
3.13	A plot to show the measured R_∞ values per SCT region, where the inner error band in the statistical uncertainty and the outer band is the sum in quadrature of the statistical and systematic uncertainties.	57
4.1	A schematic of the electron calorimeter-based isolation criteria. The grid represents the calorimeter cells of the second layer of the ECal and the circle represents the isolation cone. The topological clusters are shown in red, and the yellow rectangle represents the area used to calculate to core energy deposit of the candidate electron.	65
4.2	A sample event with topological clusters grouped into jets using a range of algorithms, including the anti- k_t algorithm used in this analysis [86].	68
4.3	The efficiency of the single-electron trigger with the FCTight isolation criterion shown as a function of the offline electron E_T during Run 2 [97].	76
4.4	The integrated luminosity collected by the ATLAS detector in the years 2015 (a) and 2016 (b). The LHC-delivered integrated luminosity and that collected are displayed separately [99].	78
4.5	An example of a one-loop diagram through which the naïve T-odd asymmetry manifests in the $W + \text{jets}$ cross section.	79
5.1	Plots depicting the real efficiency for each of the signal and validation regions as well as the measured fake efficiencies as a function of each of the variables over which the Multijet estimation is parameterised.	88
5.2	Plots depicting the real efficiency for each of the signal and validation regions as well as the measured fake efficiencies as a function of $\min(\Delta R(e, \text{jet}))$ for exclusive jet multiplicities up to four, and five inclusive.	89
5.3	A diagram depicting the signal, control and validation regions on the $E_T^{\text{miss}} - M_T^W$ plane. These regions are equivalent in all other criteria.	90
5.4	Plots depicting the two-dimensional $\Delta\phi(e, E_T^{\text{miss}}) - M_T^W$ one-jet exclusive (a) fake efficiency evaluated by MG + PY8 $W + \text{jets}$ MC, (b) real efficiency evaluated by MG + PY8 $W + \text{jets}$ MC, (c) fake efficiency evaluated by SHERPA2.2.1 $W + \text{jets}$ MC and (d) real efficiency evaluated by SHERPA2.2.1 $W + \text{jets}$ MC.	92

5.5	A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region. The distributions shown are the (a) jet multiplicity, (b) p_T^W , (c) $x_{T\text{-odd}}$, (d) p_T^{leading} and (e) H_T	93
5.6	A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region for two or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T	94
5.7	A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis. The distributions shown are the (a) jet multiplicity, (b) p_T^W , (c) $x_{T\text{-odd}}$, (d) p_T^{leading} and (e) H_T	95
5.8	A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis for two or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T	96
5.9	The ratio of the estimation of the multijet background to the sum of the same estimation of the multijet background and the MG + PY8 prediction of the $W + \text{jets}$ signal. The selection criteria on the E_T^{miss} and M_T^W variables has been removed. Also shown in red lines are the boundaries defining the criterion of this analysis ($E_T^{\text{miss}} + M_T^W < 60$ GeV and $E_T^{\text{miss}} < 25$ GeV) as well as the additional criterion of the ATLAS $W + \text{jets}$ 8 TeV analysis ($M_T^W < 40$ GeV) at Ref. [100].	99
6.1	Two-dimensional histograms depicting the response matrices of each of the analysed spectra for events with at least one jet.	106
6.2	Closure tests of the unfolding procedure used in this analysis. The MG + PY8 particle-level distribution is shown as black points, the MG + PY8 reconstructed distribution is shown as black lines, and the unfolded MG + PY8 distribution is shown as a series of coloured lines for up to five iterations. In most of the bins shown, the unfolded distributions are overlaid such that only the five-iteration case is visible.	107
6.3	The figures used to determine the number of iterations to use in the iterative Bayesian unfolding procedure. The reconstructed data are shown as a black line and the reweighted reconstructed MC is shown as a series of coloured lines for up to five iterations.	110

7.1	The breakdown of fractional uncertainties for the (a) jet multiplicity, (b) p_T^W , (c) $x_{T\text{-odd}}$, (d) p_T^{leading} and (e) H_T distributions. In each case, the uncertainties are grouped by cross-section variations (including integrated luminosity), b -tagging, pileup, E_T^{miss} , the variation of the fake efficiency in the multijet background estimation, jets and electron. The quadrature sum of each group is displayed separately.	118
7.2	The breakdown of fractional uncertainties for the p_T^W spectrum for inclusive jet multiplicities ranging between one (a) and four (d). In each case, the uncertainties are grouped by cross-section variations (including integrated luminosity), b -tagging, pileup, E_T^{miss} , the variation of the fake efficiency in the multijet background estimation, jets and electron. The quadrature sum of each group is displayed separately.	119
7.3	The correlation matrices evaluated for each of the (a) p_T^W , (b) p_T^{leading} , (c) $x_{T\text{-odd}}$ and (d) H_T distributions with themselves.	122
7.4	The correlation matrices between distributions (a) p_T^W and p_T^{leading} , (b) p_T^W and $x_{T\text{-odd}}$, and (c) p_T^W and H_T	123
7.5	The correlation matrices between distributions (a) p_T^{leading} and $x_{T\text{-odd}}$, (b) p_T^{leading} and H_T , and (c) $x_{T\text{-odd}}$ and H_T	124
8.1	The unfolded cross section for the production of a W boson in association with at least one jet for the N_{jet} observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.	127
8.2	The unfolded cross section for the production of a W boson in association with at least one jet for the p_T^W observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.	128

8.3	The unfolded cross section for the production of a W boson in association with at least one jet for the p_T^{leading} observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.	129
8.4	The unfolded cross section for the production of a W boson in association with at least one jet for the H_T observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.	130
8.5	The unfolded cross section for the production of a W boson in association with at least one jet for the $x_{T\text{-odd}}$ observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.	131
8.6	The unfolded cross section for the production of a W boson in association with jets for the p_T^W observable for at least (a) two jets, (b) three jets, (c) four jets and (d) five jets. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators. . . .	133
9.1	The ratio of the strange to up and down sea distributions ($R_s = (s + \bar{s})/(\bar{u} + \bar{d})$) of the ATLASepWZ16 PDFs evaluated at $Q^2 = 1.9 \text{ GeV}^2$ shown (a) as a function of Bjorken x and (b) evaluated at $x = 0.023$ and compared to a range of global PDF analyses [145].	139
9.2	The concentration of W + jets events on the x - Q^2 plane for (a) zero jets inclusive and (b) one jet inclusive [100].	139
9.3	Examples of the leading-order production modes of the production of a W boson in association with jets.	140

- 9.4 The differential cross-section measurements of (a) $W^+ + \text{jets}$ and (b) $W^- + \text{jets}$ in Ref. [100] (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. (9.3) are allowed to vary to minimise the χ^2 147
- 9.5 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) $25 < p_T^{\text{jet}} < 50$ GeV, (b) $50 < p_T^{\text{jet}} < 100$ GeV, (c) $100 < p_T^{\text{jet}} < 200$ GeV, (d) $200 < p_T^{\text{jet}} < 300$ GeV, (e) $300 < p_T^{\text{jet}} < 400$ GeV and (f) $400 < p_T^{\text{jet}} < 1050$ GeV, where the transverse momentum of inclusive jets is labelled p_T^{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. (9.3) are applied. 148
- 9.6 PDFs at the scale $Q^2 = 1.9 \text{ GeV}^2$ 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) $x\bar{D}$ and (h) $x\bar{D}$ distributions 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. 151

- 9.7 The $x(\bar{d} - \bar{u})$ distribution evaluated at $Q^2 = 1.9 \text{ GeV}^2$ as a function of Bjorken x (a) extracted from the ATLASepWZ20 (green) and ATLASepWZVjet20 (blue) fits with 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. 152
- 9.8 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 systematic uncertainties, and the lowest recorded χ^2 for each line shown, $\chi^2_{\min}$, subtracted. Shown are (a) the total χ^2 and (b) the χ^2 separated into contributions from HERA (solid lines) and ATLAS (dashed lines) data and smoothed. 154
- 9.9 The R_s distribution, evaluated at $Q^2 = 1.9 \text{ GeV}^2$, (a) extracted from the ATLASepWZ20 (green) and ATLASepWZVjet20 (blue) fits with 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. 156
- 9.10 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 [135–138, 175]. 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 [177]. 158
- 9.11 Summary plots of R_s evaluated at (a) $x = 0.023$ and $Q^2 = 1.9 \text{ GeV}^2$, and (b) $x = 0.013$ and $Q^2 = m_Z^2$, for the ATLASepWZVjet20 PDF set in comparison with global PDFs [135–138, 175], 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 [177]. 159

9.12	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 [135–138, 175]. The experimental and total uncertainty bands are plotted separately for the ATLASep-WZVjet20 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 [177].	160
A.1	The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC for a period prior to data taking in 2016.	168
A.2	The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC for a period prior to data taking in 2017.	169
A.3	The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC for a period prior to data taking in 2018.	170
A.4	A histogram of the fitted gradients to the plots of normalised I_{PIN} against time for the years 2016, 2017 and 2018. The RMS is shown on this plot, which is used as a systematic error for the final measurement of R_∞	171
B.1	A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region for three or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T .174	
B.2	A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region for four or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T .175	
B.3	A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis for three or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T	176

- B.4 A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis for four or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T 177
- B.5 A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region for one or more reconstructed jets split in to W^+ and W^- selection criteria. The distributions shown are (a)-(b) the jet multiplicity, (c)-(d) p_T^W and (e)-(f) $x_{T\text{-odd}}$ 178
- B.6 A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the signal region for one or more reconstructed jets split in to W^+ and W^- selection criteria. The distributions shown are (a)-(b) the jet multiplicity, (c)-(d) p_T^W and (e)-(f) $x_{T\text{-odd}}$ 179
- B.7 Closure tests of the unfolding procedure used in this analysis. The MG + PY8 particle-level distribution is shown as black points, the MG + PY8 reconstructed distribution is shown as black lines, and the unfolded MG + PY8 distribution is shown as a series of coloured lines for up to five iterations for two and three jets inclusive. The p_T^W distribution is shown in all subfigures for the W^+ selection criteria in (a) and (b), and the W^- selection criteria in (c) and (d). In most of the bins shown, the unfolded distributions are overlaid such that only the five-iteration case is visible. 180
- B.8 The unfolded cross section for the production of a W boson in association with (a)-(b) at least two jets and (c)-(d) at least three jets for the p_T^W observable split in to the W^+ cross section and the W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators. 181
- B.9 The unfolded cross section for the production of a W boson in association with (a)-(b) at least two jets and (c)-(d) at least three jets for the p_T^{leading} observable split in to the W^+ cross section and the W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators. 182

- B.10 The unfolded cross section for the production of a W boson in association with (a)-(b) at least two jets and (c)-(d) at least three jets for the H_T observable split in to the W^+ cross section and the W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators. 183
- B.11 The unfolded cross section for the production of a W boson in association with jets for the p_T^{leading} observable for at least (a) two jets, (b) three jets, (c) four jets and (d) five jets. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators. . . . 184
- B.12 The unfolded cross section for the production of a W boson in association with jets for the H_T observable for at least (a) two jets, (b) three jets, (c) four jets and (d) five jets. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators. . . . 185
- C.1 PDFs at the scale $Q^2 = m_Z^2$ as a function of Bjorken x obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting W + jets, Z + jets, inclusive W and Z , and HERA data (ATLASepWZVjet20, blue bands), compared with 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 relative uncertainties around the nominal value of each PDF centred on 1 is displayed in the bottom panel in each case. 188
- C.2 The $R_s = (s + \bar{s})/(\bar{u} + \bar{d})$ distribution evaluated at $Q^2 = m_Z^2$ 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 [135–138, 175]. 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 [177]. 189

- C.3 The $x(\bar{d} - \bar{u})$ distribution evaluated at $Q^2 = m_Z^2$ 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 [135–138, 175]. 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 [177]. 190
- D.1 PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive W, Z and HERA data, with a comparison between a fit to the original data and a similar fit to the data with systematics related to unfolding decorrelated. Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties. 195
- D.2 PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive W, Z and HERA data, with a comparison between a fit to the data with systematics related to unfolding decorrelated, and a fit with correlations between $W + \text{jets}$ and $Z + \text{jets}$ accounted for in addition. Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties. 196
- D.3 PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive W, Z and HERA data, with a comparison between a fit to data with systematics related to unfolding decorrelated and correlations between $W + \text{jets}$ and $Z + \text{jets}$ accounted for, and a similar fit with certain systematics in the $V + \text{jets}$ data summed in quadrature in preparation for correlation to 7 TeV W, Z data (ATLASepWZ20C). Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties. 196
- D.4 PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive W, Z and HERA data, with a comparison between the use of combined 7 TeV data and uncombined (where in the uncombined case, correlation to the $V + \text{jets}$ data are also accounted for). Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties. 197

D.5	PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting data with either the 8 TeV jet-scale systematics combined and correlated to the corresponding systematic in the 7 TeV data (epWZVjet_CorrJetScale) or not (ATLASepWZVjet20).	197
D.6	PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting data with the 7 TeV JES systematic approximately decomposed into multiple components to replicate possible models of correlation with the JES systematics present in the 8 TeV data compared to the ATLASepWZVjet20 fit with experimental uncertainty bands only.	198
D.7	PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting data with the 7 TeV luminosity uncertainty decomposed in to 50% correlated between centre-of-mass energies and 50% uncorrelated compared to the ATLASepWZVjet20 fit with experimental uncertainty bands only.	198
E.1	The scale-variation K -factors of the (a) $Z + \text{jets upward}$, (b) $Z + \text{jets downward}$, (c) $W + \text{jets upward}$ and (d) $W + \text{jets downward}$	200
E.2	Ratio plots of the ATLASepWZVjet20 PDF set with experimental uncertainties only compared to the central values of similar fits corrected by the scale-variation K -factors of fig. E.1 for the (a) up valence, (b) down valence, (c) gluon, (d) $\bar{u}$, (e) $\bar{d}$, (f) $\bar{s}$, (g) R_s , (h) r_s and $x(\bar{d} - \bar{u})$ distributions.	201
E.3	Ratio plots of the ATLASepWZVjet20 PDF set with experimental uncertainties only compared to the central values of similar fits corrected by the scale-variation K -factors of fig. E.1 for the (a) up valence, (b) down valence, (c) gluon, (d) $\bar{u}$, (e) $\bar{d}$, (f) $\bar{s}$, (g) R_s , (h) r_s and $x(\bar{d} - \bar{u})$ distributions.	201
E.4	Ratio plots of the ATLASepWZVjet20 PDF set with experimental uncertainties only compared to the central values of similar fits with variations in the prediction of the 7 TeV W, Z data for the (a) up valence, (b) down valence, (c) gluon, (d) $\bar{u}$, (e) $\bar{d}$, (f) $\bar{s}$, (g) R_s , (h) r_s and $x(\bar{d} - \bar{u})$ distributions.	202
E.5	Ratio plots of the ATLASepWZVjet20 PDF set with experimental uncertainties only compared to the central values of similar fits with variations in the prediction of the 7 TeV W, Z data for the (a) up valence, (b) down valence, (c) gluon, (d) $\bar{u}$, (e) $\bar{d}$, (f) $\bar{s}$, (g) R_s , (h) r_s and $x(\bar{d} - \bar{u})$ distributions.	203

Acknowledgements

I would like to firstly thank the Deutsches Elektronen-Synchrotron (DESY) institute of Hamburg for making it all possible, and the staff at the DESY and Oxford institutions whose support consistently proves to be extremely valuable, and who made this whole experience very pleasant and almost problem-free (from my point of view, anyway).

There are a huge number of people with whom I have interacted throughout this degree, who have either provided part of a fantastic social structure or who have made me a better scientist. I cannot hope to name all individuals, lest I forget someone important and regret it forever, but I wish to acknowledge the Oxford students in particular, whose collaborative efforts and social encouragement made my first year thoroughly enjoyable, and the ATLAS PDF forum, the academic support of which was invaluable.

I would like to go back four years and give my deep thanks to Prof. Chris Parkes of the University of Manchester. On the occasion that my time supervised by him for my internship and Masters degree comes up in conversation, I always look forward to saying the same thing: you could be walking to his office, preparing to let him down because you have made little progress that week, and come away over an hour later instead planning your Nobel Prize acceptance speech.

Finally, having saved the best for last, I would like to thank my supervisors Claire Gwenlan and James Ferrando. The support I received was incredibly strong before I even accepted the position. I always felt encouraged to aim high, and that the trust in my abilities and work never wavered. There was always somebody I could go to about anything at all, and they always went the extra mile. I could not be happier with my decision to take on these projects, with these supervisors.

Preface

Unless stated otherwise, the work presented in this thesis was completed by the author as a member of the ATLAS collaboration. It is therefore important to mention that, like any ATLAS publication, this is the culmination of the effort of thousands.

The following is a list of publications the author has contributed to during the time of this D.Phil., including those still to come with their expected title and release date, in reverse-chronological order.

- **Measurement of differential cross sections for the productions of a W boson in association with jets at $\sqrt{s} = 13$ TeV with the ATLAS detector**

This analysis is expected to be published in 2022. The results presented are therefore interim. The author has made significant contributions as detailed in multiple chapters of this thesis by independently writing the framework for the analysis through to an initial measurement result, deconvoluted from detector effects with a wide range of systematic uncertainties. This is intended to be a legacy Run 2 measurement of the ATLAS detector, and as such is a long-term analysis with contributions from multiple analysts and with eventual cross checks and contributions from a range of institutions.

- **QCD analysis of ATLAS Run 1 and Run 2 measurements**

Following the analysis below, for which the author was the primary analyst and editor, a significant group effort by the ATLAS PDF forum has been directed towards a “global” PDF fit of all sensitive ATLAS measurements performed to date. This is expected to be published in the year 2021.

- **Determination of the parton distribution functions of the proton from ATLAS measurements of differential $W^\pm$ and Z boson production in association with jets ([arXiv 2101.05095](#))**

This analysis has been made public and has been submitted to the Journal of High Energy Physics. The author was solely responsible for producing the results, the central values of which were cross-checked by other analysts using the same setup passed through an alternative framework, and acted as the primary contact editor throughout circulation and publication.

- **The Large Hadron-Electron Collider at the HL-LHC ([arXiv 2007.14491](#))**

The author contributed to this design-report update of a potential future experiment “LHeC” through parton distribution function and α_s sensitivity studies, and is listed in the authorship as a result.

- **QCD analysis of ATLAS W -boson production data in association with jets ([ATL-PHYS-PUB-2019-016](#))**

The author of this thesis is the primary analyst and editor of the above ATLAS PUB. note. Subsequent improvements to this analysis are published in the above “Determination of the parton distribution functions of the proton from ATLAS measurements of differential $W^\pm$ and Z boson production in association with jets” analysis.

- **In-situ radiation damage studies of the ATLAS SCT optical links. ([JINST 14 \(2019\) 07, P07014](#))**

The author of this thesis was one of the primary analysts of this publication. The research undertaken by the author is detailed in section 3.3 of this thesis.

1

Introduction

The first written conception of the idea that matter is composed of a series of elementary “building blocks” is usually attributed by historians as far back as the 4th century BC to Leucippus of Abdera and his pupil Democritus; the idea was that there could exist some limit to the number of times we can divide an object before we are left with its indivisible constituents. It was believed that the properties of different types of matter could be attributed to the properties of the atomic matter composing them (for example, water flows as the constituting particles are “slippery”, and lemon tastes bitter as the constituting particles have sharp edges) [1, 2]. With hindsight, these ideas are surprisingly accurate in some ways, and the Ancient Greeks can perhaps be forgiven for attributing atomic properties to their shape as opposed to chemical interactions. Such ideas can be found repeated throughout history in various ways, for example in Newton’s corpuscular theory of light, which suggested that light was composed of incomprehensibly small particles scattering from surfaces [3].

Modern scientific advancements of this same idea were first made in the field of chemistry; firstly by John Dalton in 1808, whose work on stoichiometry led to the conclusion that each element of nature is composed of a unique particle, and that these elements combined to form compounds [4]; and secondly by Dmitri Mendeleev in 1869, who arranged the elements into a *periodic table* in order of

mass and grouped by columns according to similar chemical properties - this table was revolutionary, firstly as gaps were left in various places, only to be filled in one-by-one over time by future discoveries (thus providing testable predictions), but furthermore as the format of the table was suggestive of a not-yet-understood underlying substructure [5].

Subsequent advances were made by Ernest Rutherford in 1911 by scattering α particles off a gold-leaf target. The interpretation of the results of this measurement was that an atom consists of a positively charged nucleus, containing the vast majority of the mass concentrated into a very small fractional volume, surrounded by a set of orbiting negatively charged particles of significantly less mass [6]. Since then, the experimental investigation of these constituent, fundamental building blocks, from which everything is made, and the nature of their interactions, has earned its place as one of the pillars of modern physics, referred to as either *Elementary Particle Physics* or (with foreshadowing of the methodology), *High-Energy Physics*.

Over the subsequent years, it became clear through a large worldwide effort of the scientific community that the nucleus discovered by Rutherford is composed of what we refer to as *protons* and *neutrons*, which themselves contain further substructure called *quarks* and *gluons*, and the negative charge is composed of *electrons*. These particles are a subset of the particles described by the *Standard Model* (SM), the most prevalent model used to describe the nature of the fundamental constituents of the Universe and their interactions.

The SM has so far provided a remarkably accurate description of elementary particle physics. The last piece of the SM, the *Higgs Boson*, was discovered by both the ATLAS and CMS experiments of the Large Hadron Collider (LHC) in 2012 [7, 8], and no discovery beyond the SM has been made with collider data. However, it still has deficiencies. That the SM does not describe for example Dark Matter or explain why the Universe is dominated by matter without a similar presence of antimatter are just a couple of examples of the ways in which the SM fails to describe observations outside of colliders. The hunt for a new, or more complete, model continues.

In the LHC era, precision and discovery are no longer as simple as breaching the high-energy or high-statistics frontiers. Our restricted understanding of the SM itself has become a limiting factor of our measurements - when the precision of our data exceeds the precision of the calculations of our model, we cannot go further. That is, not without the calibration of the model using benchmark processes which better constrain our understanding.

This thesis is mainly focused on the measurement of one such benchmark process: the production of a W boson in association with jets (referred to as “ $W + \text{jets}$ ”), and the SM interpretation of these results. This is a process which is one of the most prevalent at the LHC, with very high statistics providing the potential for fantastic constraining power of the SM phenomena to which it is sensitive.

In chapter 2, background understanding of the theory of the SM is discussed. The relevant components of the SM are given a brief overview in section 2.1, with more detail given to proton structure in section 2.2 and the $W + \text{jets}$ cross section at the LHC in section 2.3. Subsequently in chapter 3, an overview of the LHC and the ATLAS detector is given, with relevant details about the experiment discussed in sections 3.1 and 3.2. Furthermore, an analysis of the radiation damage suffered by the optical links of a subcomponent of the detector performed by the author is discussed in section 3.3.

The physics analyses performed for this thesis are discussed thereafter. The analysis strategy of the measurement of the $W + \text{jets}$ cross section at the ATLAS detector using proton-proton collisions of a centre-of-mass energy of 13 TeV is discussed in chapter 4. Details on the estimation of background contamination is detailed in chapter 5 and the subtraction of this background, along with correction for other detector effects, is described in chapter 6. The composition of the uncertainty of this measurement is described in chapter 7, and the resulting measurement is presented in chapter 8. Additionally, a QCD analysis of a similar measurement of $W + \text{jets}$ at a centre-of-mass energy of 8 TeV is presented in chapter 9.

2

Theory

Contents

2.1	The Standard Model of Particle Physics	5
2.1.1	Quantum electrodynamics	8
2.1.2	The Weak force	10
2.1.3	Electroweak theory	13
2.1.4	Quantum chromodynamics	17
2.2	The structure of the proton – Parton Distribution Functions	19
2.3	The production of a W boson in association with jets at a proton-proton collider	23

In this chapter, an overview of the Standard Model (SM) is presented, giving a summary of the elementary particles and their interactions. Emphasis is given in subsequent sections to the topics connected with the measurements detailed in later chapters: the structure of the proton (Parton Distribution Functions) and the $W + \text{jets}$ cross section at a proton-proton collider.

2.1 The Standard Model of Particle Physics

The SM was formulated in the 1960s and early 1970s [9–14] to represent our current understanding of elementary particles and their interactions. It was motivated by prior measurements, but came with many testable predictions which have been

scrutinised by a number of dedicated experiments since. So far, no measurement with collider data has deviated from its corresponding SM prediction by the discovery benchmark of 5σ (equivalent to a p -value of $\sim 10^{-7}$). On the contrary, all predictions which had not previously been observed have since been discovered at that same benchmark, including the W and Z gauge bosons in 1983 [15–18], the top quark in 1995 [19, 20], the tau neutrino in 2000 [21] and the Higgs Boson in 2012 [7, 8].

The SM postulates that the world around us is formed from just a few different particles described in fig. 2.1 interacting through the manifestation of three fundamental forces: Electromagnetism, the Weak force, and the Strong force. It is these interactions, along with the gravitational force¹, which give rise to the rich structure of the universe - from the binding and decay of atoms and their electron orbitals to the nuclear fusion powering the stars. The elegant structure of the periodic table, and how it gives rise to the chemical reactions which power everything at a macroscopic scale, can all, on a fundamental level, be explained by these theories of interaction, with the particle content of the SM.

The particle content can be separated into two categories: fermions and bosons, where the fermions are the spin-1/2 particles which obey Fermi-Dirac statistics and bosons are integer-spin particles obeying Bose-Einstein statistics. All fermions interact through the Weak force, but only some interact through the Strong force – those which do are labelled “quarks”, while those which do not are labelled “leptons”. Those which are electromagnetically charged also interact through the Electromagnetic force. All interactions occur through the bosons acting as mediators, as discussed in more detail in sections 2.1.1 to 2.1.4. The fermions are further divided into three generations, each generation represented by a column in fig. 2.1. Fermions in a different generation but on the same row in fig. 2.1 share quantum numbers, but differ in mass.

¹The gravitational force is well-known, and the theory through which it appears to present, General Relativity, is thoroughly tested in other studies mostly in fields other than Particle Physics. At this time, the SM does not account for gravitation, however it is expected that its effects for most collider studies, including those in this thesis, are negligible.

mass →	≈2.3 MeV/c ²	≈1.275 GeV/c ²	≈173.07 GeV/c ²	0	≈126 GeV/c ²
charge →	2/3	2/3	2/3	0	0
spin →	1/2	1/2	1/2	1	0
	u up	c charm	t top	g gluon	H Higgs boson
QUARKS					
	≈4.8 MeV/c ²	≈95 MeV/c ²	≈4.18 GeV/c ²	0	
	-1/3	-1/3	-1/3	0	
	1/2	1/2	1/2	1	
	d down	s strange	b bottom	γ photon	
	0.511 MeV/c ²	105.7 MeV/c ²	1.777 GeV/c ²	91.2 GeV/c ²	
	-1	-1	-1	0	
	1/2	1/2	1/2	1	
	e electron	μ muon	τ tau	Z Z boson	
LEPTONS					
	<2.2 eV/c ²	<0.17 MeV/c ²	<15.5 MeV/c ²	80.4 GeV/c ²	
	0	0	0	±1	
	1/2	1/2	1/2	1	
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson	
					GAUGE BOSONS

Figure 2.1: The elementary-particle content of the SM. For each particle, the mass, charge and spin is given. The table is split into four sections: the quarks consisting of the up, down, charm, strange, top and bottom particles; the leptons consisting of the electron, muon, tau and their associated neutrino partners; the spin-1 gauge bosons consisting of the gluon, photon, Z and W bosons; and the Higgs boson. [22]

The SM is a quantum field theory (QFT) through which fundamental particles are understood to be quantum fields. Fermions are described as Dirac fields satisfying the following Lagrangian:

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi, \quad (2.1)$$

where ψ is the field, γ^μ are the Dirac matrices, ∂_μ is a partial derivative with respect to the μ^{th} component of space-time, m is the particle mass, and summation over μ is implied. Interactions within the SM are understood through the additional components of the Lagrangian denoted $\mathcal{L}_{\text{QCD}}$ - the Quantum Chromodynamics (QCD) Lagrangian describing the Strong force - and $\mathcal{L}_{\text{EW}}$ - the Electroweak Lagrangian describing the EM and Weak forces.

According to Noether's theorem, continuous symmetries of a Lagrangian result in a conserved current. The SM Lagrangian is founded on the gauge symme-

try of the group:

$$G_{\text{SM}} = \text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y, \quad (2.2)$$

where (C) represents interaction through the colour charge of fields, (L) represents interaction through only left-handed fields, and ($Y \equiv 2(Q - T_3)$) represents interaction through the weak hypercharge, where Q is the electric charge of the field and T_3 represents the projection of the weak isospin along the arbitrarily-chosen z axis. The first term, $\text{SU}(3)_C$, is a non-Abelian symmetry² referring to the colour symmetry of QCD. Together, the second and third terms, $\text{SU}(2)_L \times \text{U}(1)_Y$, are groups associated with the EW fields referring to the symmetry of the weak isospin and hypercharge, where $\text{U}(1)$ is Abelian but $\text{SU}(2)$ is not.

As they are formulated here, the imposition of these gauge symmetries gives rise to spin-1 massless mediators – the gluon for the $\text{SU}(3)_C$ group, the $W^\pm, Z$ bosons for the $\text{SU}(2)_L$ group and the photon for the $\text{U}(1)_Y$ group. The symmetries and their resulting gauge bosons are discussed in sections 2.1.1 to 2.1.4.

2.1.1 Quantum electrodynamics

Quantum electrodynamics (QED) is the theory which describes electromagnetic interactions, causing particles of opposite-sign charges to attract and same-sign charges to repel. The formulation of QED within the SM starts from the expression of Maxwell's equations in relativistic form, namely:

$$\partial_\mu F^{\mu\nu} = J^\nu, \quad (2.3)$$

$$\text{where } F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu,$$

and J^ν is a conserved current. These equations can be derived from the following Lagrangian using the Euler-Lagrange equations:

$$\mathcal{L}_{\text{Maxwell}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - J^\mu A_\mu. \quad (2.4)$$

This Lagrangian together with the Dirac Lagrangian in eq. (2.1) gives the QED Lagrangian, $\mathcal{L}_{\text{QED}} = \mathcal{L}_{\text{Maxwell}} + \mathcal{L}_{\text{Dirac}}$. The term containing the conserved current

²Non-Abelian here means that the group operations do not commute, ie. $ab \neq ba; a, b \in G$, where G is the group.

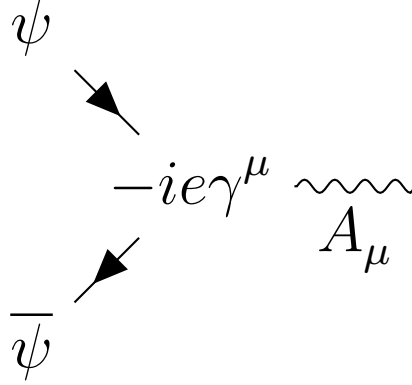


Figure 2.2: Feynman diagram representing the interaction vertex obtained from the interaction term of the QED Lagrangian.

is the interaction term, where a good candidate for the electromagnetic current describing a particle of charge Q and coupling strength e is:

$$J^\mu = Qe\bar{\psi}\gamma^\mu\psi. \quad (2.5)$$

It is worth noting that this interaction term could be formulated instead with the transformation $\partial_\mu \rightarrow D_\mu = \partial_\mu + iQeA_\mu$ in the other terms. Only with the inclusion of all terms, including the interaction term, is this Lagrangian invariant with respect to the local gauge transformations of the $U(1)$ group:

$$\psi(x) \rightarrow \psi'(x) = e^{iQe\alpha(x)}\psi(x), \quad A_\mu(x) \rightarrow A'_\mu(x) = A_\mu(x) + \partial_\mu\alpha(x), \quad (2.6)$$

where $\alpha(x) \in \mathbb{R}$ is the $U(1)$ group generator.

This could also be thought of the other way round: by starting with the field potential term $F^{\mu\nu}F_{\mu\nu}$ and the Dirac term, and requiring that the Lagrangian is invariant under local gauge transformations, we must transform the covariant derivative, thereby introducing an additional term which corresponds to the interaction of the particle field ψ with the field potential A_μ . Such an interaction is shown in fig. 2.2.

It is from this Lagrangian that all electromagnetic interactions can be derived; not just Maxwell's equations, but also those reliant on quantum effects such as Compton scattering.

2.1.2 The Weak force

The Weak force is the mediator of many decay processes in the SM, for example nuclear β -decay. This is the process whereby a neutron will “transform” into a proton, releasing an electron and an anti-neutrino in addition. The underlying particle process is $d \rightarrow u + e^- + \bar{\nu}$, where d represents a down quark, u represents an up quark, e represents an electron and $\bar{\nu}$ represents the anti-neutrino. This process is described similarly to that of QED except that the group is now $SU(2)$, motivated as the particles change flavour at the interaction vertex, and will therefore be considered as “weak isospin” doublets:

$$\chi_q = \begin{bmatrix} u \\ d \end{bmatrix}, \quad \chi_l = \begin{bmatrix} l \\ \nu \end{bmatrix}. \quad (2.7)$$

The generators of the $SU(2)$ group are the Pauli matrices:

$$\sigma^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad (2.8)$$

with the commutation rule $[\frac{\sigma^i}{2}, \frac{\sigma^j}{2}] = i\epsilon_{ijk}\frac{\sigma^k}{2}$, where ϵ_{ijk} is the Levi-Civita tensor.

To impose local gauge symmetry, the covariant derivative is written:

$$\begin{aligned} \partial_\mu \rightarrow D_\mu &= \partial_\mu + i\frac{g_W}{2}W_\mu^i\sigma^i \\ &= \partial_\mu + i\frac{g_W}{2} \begin{bmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{bmatrix}, \end{aligned} \quad (2.9)$$

where the $SU(2)$ gauge fields analogous to A_μ in the QED formulation have been denoted W_μ^i . It is convenient to write the complex gauge fields in the off-diagonal elements as $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)$. One can write the required Lagrangian term which gives rise to β -decay in terms of the 2×2 covariant derivative and the particle doublets:

$$\mathcal{L}_{\text{Weak}} \supset i\chi_l\gamma^\mu D_\mu\chi_l + i\chi_q\gamma^\mu D_\mu\chi_q, \quad (2.10)$$

generating interactions between different-flavour particles from the off-diagonal terms of the covariant derivative of the form $-ig_W\sqrt{2}\bar{\nu}\gamma^\mu W_\mu^+e^-$. Again analogously

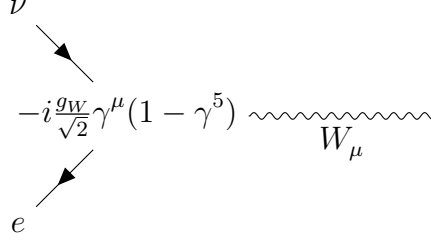


Figure 2.3: Feynman diagram representing the interaction vertex obtained from the interaction term of the Weak Lagrangian.

to the formulation of QED, the interaction vertex is shown in fig. 2.3 and a current can be derived in a similar format:

$$J^\mu = -\frac{g_W}{\sqrt{2}}\bar{\nu}\gamma^\mu e^-. \quad (2.11)$$

However, one piece is missing from this formulation so far which differs from the QED analogy, and that is parity violation. Discovered by Wu et al. [23] in 1957, it is known that the weak interaction maximally violates symmetry under parity transformation³ by means of a vector-minus-axial-vector ($V - A$) structure. In order to impose this on the Weak Lagrangian, the following transformation is made:

$$\gamma^\mu \rightarrow \gamma^\mu(1 - \gamma^5), \quad (2.12)$$

where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. The new charged current weak reaction is then written as:

$$J^\mu = -\frac{g_W}{\sqrt{2}}(\bar{\nu}\gamma^\mu e^- - \bar{\nu}\gamma^\mu\gamma^5 e^-). \quad (2.13)$$

The two terms in this formulation transform under parity in the following ways:

$$\bar{\psi}\gamma^\mu\psi = \begin{bmatrix} \bar{\psi}\gamma^0\psi \\ \bar{\psi}\gamma^i\psi \end{bmatrix} \rightarrow \begin{bmatrix} +\bar{\psi}\gamma^0\psi \\ -\bar{\psi}\gamma^i\psi \end{bmatrix} \quad \text{and} \quad \bar{\psi}\gamma^\mu\gamma^5\psi = \begin{bmatrix} \bar{\psi}\gamma^0\gamma^5\psi \\ \bar{\psi}\gamma^i\gamma^5\psi \end{bmatrix} \rightarrow \begin{bmatrix} +\bar{\psi}\gamma^0\gamma^5\psi \\ +\bar{\psi}\gamma^i\gamma^5\psi \end{bmatrix}, \quad (2.14)$$

ie. like a vector and axial vector, thus fulfilling the $V - A$ requirement. Projection operators can be defined as $P_{L,R} = \frac{1}{2}(1 \mp \gamma^5)$, which act on fermion fields to generate their left- and right-handed components, $P_{L,R}\psi = \psi_{L,R}$. Right-handed fermions have spin pointing in the same direction as their motion through space, whereas left-handed fermions exhibit the opposite. As P_L is present in the interaction vertex,

³Here, “parity transformation” refers to a reversal of the spatial coordinates.

only left-handed fermions and right-handed anti-fermions couple to the W gauge bosons. Under this formulation, we have an $SU(2)$ theory of weak interactions, where the fermion fields are charged under *weak isospin*, which can take the eigenvalues of the third $SU(2)$ generator: $I_3 = 0, \pm\frac{1}{2}$.

Cabibbo-Kobayashi-Maskawa mixing

The quark model of hadrons introduced by Gell-Mann and Zweig [24, 25] explained the spectrum of known mesons and baryons as $q\bar{q}$, qqq and $\bar{q}\bar{q}\bar{q}$ with the up, down and strange quarks. One missing piece of this puzzle was the lack of a flavour-changing neutral current (FCNC) which was subsequently explained by the formulation of the Glashow Iliopoulos Maiani (GIM) mechanism [26], in which the physical d - and s -quarks are a linear superposition of their weak eigenstates d' and s' :

$$\begin{bmatrix} d' \\ s' \end{bmatrix} = \begin{bmatrix} \cos(\theta_C) & \sin(\theta_C) \\ -\sin(\theta_C) & \cos(\theta_C) \end{bmatrix} \begin{bmatrix} d \\ s \end{bmatrix}, \quad (2.15)$$

where θ_C is the *Cabibbo angle*. An up quark interacts only with the d' eigenstate via the weak force. The later-discovered charm quark, c , interacts with the s' eigenstate, thereby interacting more strongly with the strange quark compared to the down.

A later proposal by Kobayashi and Maskawa posited that, in order to account for the existence of CP^4 violation within the weak force, this matrix would need a complex component and therefore a third generation, the bottom (b) and top (t) quarks, where we now have the CKM matrix defining the weak eigenstates:

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}. \quad (2.16)$$

The current best determination of the magnitudes of the CKM matrix elements are [22]:

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97427 & 0.22534 & 0.00351 \\ 0.22520 & 0.97344 & 0.0412 \\ 0.00867 & 0.0404 & 0.999146 \end{bmatrix}. \quad (2.17)$$

⁴C conjugation here refers to the transformation from particle to anti-particle and vice versa.

2.1.3 Electroweak theory

The electroweak (EW) theory is the unification of QED and the Weak force. Not discussed in the previous section was the effect of the gauge field W_μ^3 , which couples fermions to themselves. One might assume that this corresponds to the physical Z boson in fig. 2.1, however were this the case, the Z boson would also couple only to left-handed fermions and right-handed antifermions. This is not the case.

It is possible to generate the Z field by taking a mixture of the W_μ^3 field with the $U(1)$ boson, now denoted B_μ for the purpose of this mixing. Were this the case, the physical A_μ would also be a mixture, derived from the following transformation:

$$\begin{bmatrix} W_\mu^3 \\ B_\mu \end{bmatrix} = \begin{bmatrix} \cos(\theta_W) & \sin(\theta_W) \\ -\sin(\theta_W) & \cos(\theta_W) \end{bmatrix} \begin{bmatrix} Z_\mu \\ A_\mu \end{bmatrix}, \quad (2.18)$$

where θ_W is the *Weinberg angle*, measured by a number of experiments, for example those at the Large Electron-Positron (LEP) collider whose combined efforts gave the result $\sin^2(\theta_W) = 0.23153 \pm 0.00016$ [27].

Under this formulation, the B^0 couples to the *weak hypercharge*, Y , with strength denoted g' , and the W_μ^3 field interacts with the weak isospin, I_3 , with strength denoted g_W . This gives the following terms in the Lagrangian:

$$\begin{aligned} \mathcal{L} \supset & -\bar{\psi} g' Y \gamma^\mu B_\mu \psi - \bar{\psi} g_W I_3 \gamma^\mu W_\mu^3 \psi \\ & = -\bar{\psi} (+g' Y \cos(\theta_W) + g_W I_3 \sin(\theta_W)) \gamma^\mu A_\mu \psi \\ & + -\bar{\psi} (-g' Y \sin(\theta_W) + g_W I_3 \cos(\theta_W)) \gamma^\mu Z_\mu \psi. \end{aligned} \quad (2.19)$$

The term containing A_μ can now be compared directly to the current given in eq. (2.5) for consistency with the original QED formulation:

$$\begin{aligned} \bar{e}_L \gamma^\mu e_L : & \quad Q_e e = \frac{1}{2} g' Y_{e_L} \cos(\theta_W) + I_3^e g_W \sin(\theta_W), \\ \bar{\nu}_L \gamma^\mu \nu_L : & \quad 0 = \frac{1}{2} g' Y_{\nu_L} \cos(\theta_W) + I_3^\nu g_W \sin(\theta_W), \\ \bar{e}_R \gamma^\mu e_R : & \quad Q_e e = \frac{1}{2} g' Y_{e_R} \cos(\theta_W), \\ \bar{\nu}_R \gamma^\mu \nu_R : & \quad 0 = \frac{1}{2} g' Y_{\nu_R} \cos(\theta_W). \end{aligned} \quad (2.20)$$

where the weak isospin is $I_3^e = -1/2$ and $I_3^\nu = 1/2$ for each of the left-handed particles (such that they form an isospin doublet) and 0 for right-handed particles,

and the charge of the neutrinos has been taken to be 0. From these equations, and the condition that $Y_{e_L} = Y_{\nu_L}$ again as they form a doublet, one can derive the weak unification condition giving the relationship between the electromagnetic and weak couplings:

$$e = g_W \sin(\theta_W) = g' \cos(\theta_W), \quad (2.21)$$

and the relationship between the weak hypercharge, the magnetic charge, and the third component of weak isospin:

$$Y = 2(Q - I_3). \quad (2.22)$$

Particle mass: The Higgs mechanism

By now, most of the required information for the “ W ” part of the $W + \text{jets}$ measurement has been covered. The previous discussion gives us a Lagrangian that is missing one final, crucial component: the boson mass.

In a series of independent articles published around 1964, the issue of producing boson masses without breaking gauge invariance was solved through the introduction of the *Higgs mechanism* by Brout, Englert, Guralnik, Hagen, Higgs, and Kibble. In this subsection, the basic principle of the Higgs mechanism and the resulting bosonic masses are discussed.

The idea is to introduce a new complex scalar $\text{SU}(2)$ doublet, labelled Φ , with weak hypercharge 1 and weak isospin $\frac{1}{2}$. The scalar part of the corresponding “Higgs Lagrangian” is given by:

$$\mathcal{L}_{\text{Higgs}} \supset (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi^\dagger \Phi), \quad (2.23)$$

where the first term contains the electroweak covariant derivative discussed in previous sections, thereby (as we shall discover) giving rise to couplings between the electroweak gauge bosons and the new scalar field, and the second term describes the *Higgs potential*. Crucially, gauge symmetry is maintained as this potential is a function of $\Phi^\dagger \Phi$. The precise form of the Higgs potential is unknown, but for

the purpose of introducing mass to the electroweak bosons, a relatively simple quadratic toy potential is assumed:

$$V(\Phi^\dagger\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2. \quad (2.24)$$

This toy potential is chosen as it can have minima (also referred to as vacuum expectation values (VEV)) at values other than 0; for $\mu^2 < 0$, the VEV take values of $|\Phi| = v/2$, where $v = \sqrt{-\mu^2/\lambda}$.

The SU(2) symmetry is maintained, but can be “spontaneously” broken as the vacuum takes values within this minimum. By choosing a value $\Phi = \frac{1}{\sqrt{2}}(0, v)$ and performing a Taylor expansion about its radial and circular degrees of freedom, we obtain the following expression:

$$\Phi(x) = \frac{1}{\sqrt{2}}e^{-iG_a(x)\frac{\sigma_a}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \quad (2.25)$$

where x is the measure of the radial distance in Φ -space, $G_a(x)$ labels the (three) circular degrees of freedom, $h(x)$ corresponds to the expanded radial degree of freedom and σ^a are the usual SU(2) generators.

Subsequently, one can substitute this expression into the covariant derivative term of the Higgs Lagrangian, with the W_μ and B_μ fields which mix to produce the $W_\mu^\pm$, Z_μ and A_μ fields, to write the covariant derivative in its expanded form:

$$\begin{aligned} (D^\mu\Phi)^\dagger(D_\mu\Phi) &= \frac{1}{4}g_W^2(h^2 + 2vh + v^2)W_\mu^+W_\mu^{-,\mu} \\ &+ \frac{1}{8}(g_W^2 + g'^2)(h^2 + 2vh + v^2)Z_\mu Z^\mu \\ &+ \mu^2h^2 - \lambda vh^3 - \frac{\lambda}{4}h^4, \end{aligned} \quad (2.26)$$

where the *Unitary gauge* has been chosen such that the G_a terms are absorbed into the longitudinal components of the W and Z fields. From this expression, one can read off the masses of the bosons:

$$\begin{aligned} m_W &= \frac{v}{2}g, \\ m_Z &= \frac{v}{2}\sqrt{g_W^2 + g'^2} = \frac{m_W}{\cos\theta_W}, \\ m_h &= \sqrt{2}|\mu| = v\sqrt{2\lambda}. \end{aligned} \quad (2.27)$$

Furthermore, that there is no $A^\mu A_\mu$ term generated is highly convenient, given that the photon is known to be massless, and some Higgs self-interaction terms are also generated. These results are consistent with the measurements of the boson masses, $m_W = 80.4$ GeV and $m_Z = 90.2$ GeV, as well as the discovery of the Higgs boson itself in the year 2012 [7, 8]. Subsequent measurements of the properties of the Higgs Boson have provided results consistent with their SM predictions, shown for example in Ref. [28] for the ATLAS Collaboration and [29] for the CMS Collaboration.

Fermionic masses are generated in a similar way from the non-scalar part of the Higgs Lagrangian:

$$\mathcal{L}_{\text{Higgs}} \supset \sum_{f=l,q} y_f (\bar{f}_L \Phi f_R + \bar{f}_R \bar{\Phi} f_L), \quad (2.28)$$

where y_f are matrices with dimensionality of the number of fermion generations (which, in the SM, is three), f_L are the previously introduced left-handed fermionic doublets and f_R are the right-handed singlets. By replacing Φ with the VEV, as before, fermionic mass terms and Higgs-fermion interaction terms are generated as they were for the bosons. The coupling strength is given simply by y_f and the mass for each fermion is given by:

$$m_f = y_f \frac{v}{\sqrt{2}}. \quad (2.29)$$

There are a few notable remarks:

- As there are no right-handed neutrinos in the SM, no interaction or mass terms are generated for the neutrinos. In fact, that multiple experiments have reported the presence of neutrino mass is empirical evidence of physics beyond the SM.
- The y_f matrices are not necessarily diagonal. In order to determine the physical masses of the fermions, one must diagonalise the matrices, giving $M_f = U_f y_f U_f^\dagger$. These are known as the *Yukawa* couplings. These unitary U matrices redefine the fermion fields into their physical states. Applying this throughout all terms in the Lagrangian gives the result that the W bosons

couple to the physical quarks with couplings proportional to $U_u U_d^\dagger$, which gives a 3×3 matrix – this is, in fact, the CKM matrix discussed in section 2.1.2.

- As there is no mixing between lepton generations, the Yukawa couplings of the electron, muon and tau leptons are simply given by y_l .

2.1.4 Quantum chromodynamics

The theory of Strong interactions is described by Quantum Chromodynamics (QCD). It is similar in structure to QED and formulated in a similar way, however the symmetry group through which it is formulated is $SU(3)$. This is motivated by measurements from various experiments; for example, by taking the ratio

$$R = \frac{\sigma(e^+e^- \rightarrow q\bar{q})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)}, \quad (2.30)$$

one can infer from the fact that R is three times higher than otherwise expected that each flavour of quark can itself manifest as a triplet of three sub-flavours, or ‘colours’ [22]. Another way of formulating this is

$$R = N_C \sum_q e_q^2, \quad (2.31)$$

where N_C denotes the number of colours and e_q denotes the charge of quark q . We proceed similarly to the Weak formulation by therefore assuming that strongly-interacting particles can be written as a three-dimensional state.

The QCD Lagrangian is defined as follows:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}G_{\mu\nu}^a G_{\mu\nu}^a + \sum_{f \in u,d,c,s,t,b} (\bar{\psi}(i\gamma^\mu D_\mu - m_f)\psi), \quad (2.32)$$

where the three colour fields are labelled by the a index and the summation is over the quark flavours. The $G_{\mu\nu}^a$ tensor is the gluon field strength tensor, playing an analogous role to that of the $F_{\mu\nu}$ tensor in QED:

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g_s f^{abc} A_\mu^b A_\nu^c, \quad (2.33)$$

and the covariant derivative is transformed in order to maintain local gauge symmetry:

$$D_\mu = \partial_\mu + ig_s t_a A_\mu^a. \quad (2.34)$$

The t_a are the generators of the SU(3) group, and do not commute as the SU(3) group is non-Abelian:

$$[t^a, t^b] = if^{abc}t^c. \quad (2.35)$$

It is for this reason that an additional term containing the f^{abc} tensor is required in the gluon field strength tensor⁵. This leads to a vital feature of QCD: the cross term containing $A^b A^c$ results in the self-coupling of the mediator - unlike in QED where the photon is not charged, the mediator exhibits a colour charge, and so can self-interact.

The conserved current for the QCD interaction vertex can be read from the Lagrangian as follows:

$$J^{\mu,a} = g_S \frac{\Lambda^a}{2} \bar{q} \gamma^\mu q, \quad (2.36)$$

where $\Lambda^a = 2t^a$ are the *Gell-Mann matrices*. The dimensionless variable $\alpha_S = g_S^2/4\pi$ is commonly used to denote the QCD coupling strength.

When calculating the coupling strength, one vital feature of all of the interactions discussed is that the coupling varies with respect to the energy scale of interest (it is said that it is a ‘running’ coupling). This is because loop corrections, such as those shown in fig. 2.4, must be summed to all orders, ie. considering up to an infinite number of loops. The most common method used to deal with this is *renormalisation*, in which the running of the coupling strength is absorbed in to the definition of α_S instead of modifying the strong current itself. The previously-mentioned loops introduce terms which are dependent on the momentum of the propagation gluon – denoted Q^2 – resulting in a coupling which is energy-scale dependent.

Given an initial starting/reference scale μ^2 , α_S can be calculated at one-loop level as the following:

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu^2)}{1 + \beta_0 \alpha_S(\mu^2) \ln(Q^2/\mu^2)}, \quad (2.37)$$

where

$$\beta_0 = \frac{11N_C - 2N_f}{12\pi}, \quad (2.38)$$

⁵In fact, it is required in Abelian groups too, however in this case $f^{abc} = 0$.

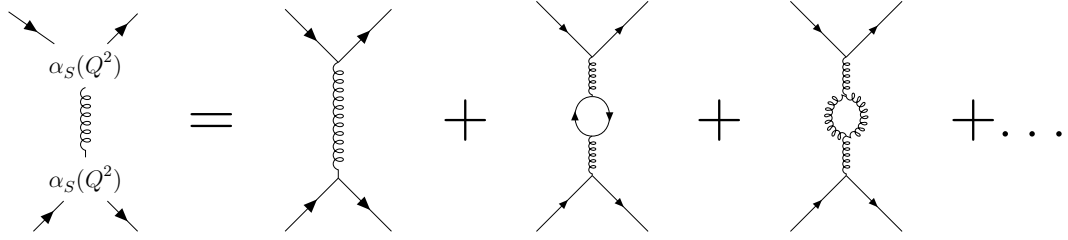


Figure 2.4: Feynman diagrams depicting the absorption of the higher-order corrections into the gluon propagator, resulting in the dependence of α_S on the energy scale of the interaction, Q^2 .

where $N_C = 3$ and $N_f = 6$, resulting in a positive β_0 and therefore an α_S decreasing with increasing Q^2 . This relationship has been confirmed by a number of experiments, as shown in fig. 2.5.

In the limit that $Q^2 \rightarrow \infty$, $\alpha_S \rightarrow 0$. This gives rise to *asymptotic freedom*, the physical consequence of QCD where at sufficiently high energies (for example, in the early universe) the coupling is small enough for quarks and gluons to behave as free particles. On the other hand, as Q^2 decreases to below around 1 GeV, the interaction enters what is referred to as the *non-perturbative regime* where it is too high for a meaningful loop-calculation to a finite order to be performed.

2.2 The structure of the proton – Parton Distribution Functions

The Large Hadron Collider (LHC), around which the measurements of this thesis are based, is a proton-proton colliding machine. Protons are not themselves fundamental particles; the constituents are three *valence quarks* – two up and one down, giving a total charge of +1 – bound together by the Strong interaction. As discussed in section 2.1.4, QCD is the theory describing the strong force, and the mediator is the gluon – the proton could therefore be described (rather unscientifically) as three valence quarks submerged in a gluon sea. In addition, the constituent gluons can themselves split into quark-antiquark pairs, referred to in the context of proton structure as *sea quarks*.

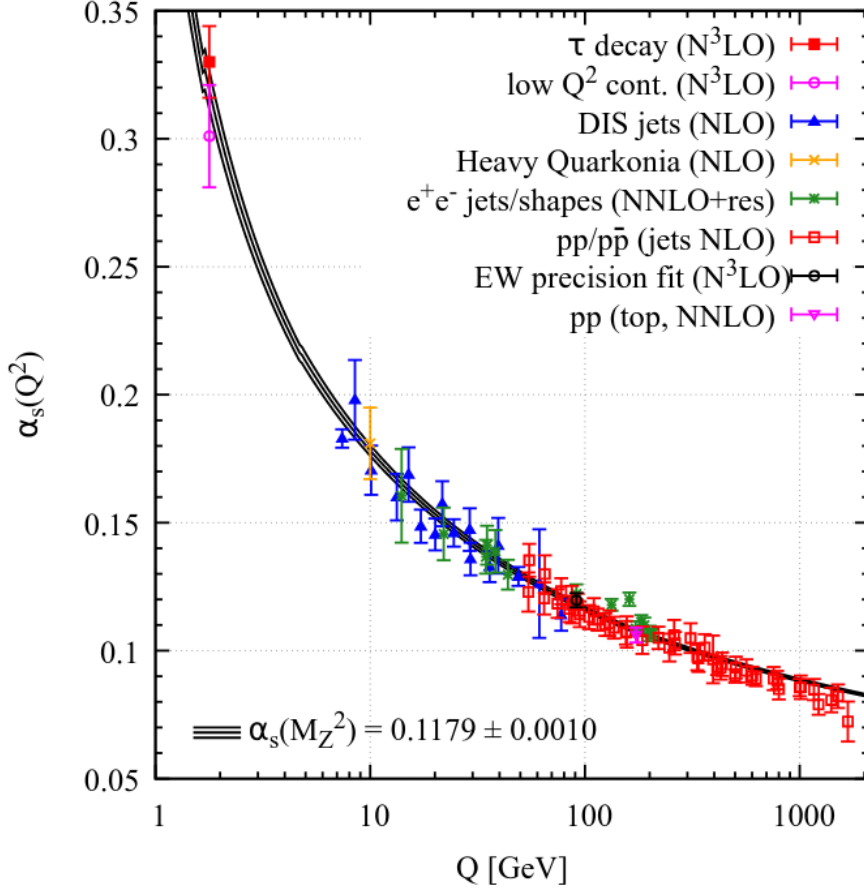


Figure 2.5: A summary of the measurements of α_s as a function of the energy scale Q across a range of experiments [22].

At a high-energy proton-colliding machine, or any high-energy machine in which non-fundamental particles are collided, the “hard” interaction is between the constituents of the colliding particles. For example, in the production of a Z boson, the interaction is not $pp \rightarrow Z$, but $q\bar{q} \rightarrow Z$ (at lowest order), where q denotes a quark.

The cross-section for a process $pp \rightarrow X$ is described in the following way:

$$\sigma_{pp \rightarrow X} = \sum_{i,j=q,\bar{q},g} \int_0^1 dx_i \int_0^1 dx_j f_i(x_i, \mu_F^2) f_j(x_j, \mu_F^2) \sigma_{i,j \rightarrow X}(x_i, x_j, \mu_F, \mu_R), \quad (2.39)$$

where the sum is over each possible combination of interacting *partons*, x_i refers to the fraction of the proton momentum carried by parton i , μ_F is defined as the *factorisation scale* and μ_R is defined as the *renormalisation scale*. The computation of $\sigma_{i,j \rightarrow X}(x_i, x_j)$ leads to divergences similar to those discussed for the running of α_s in section 2.1.4, which themselves come in two flavours:

- Ultraviolet (UV), appearing due to large-momentum loops;
- Infrared (IR), appearing due to either low-momentum emission or the radiation of a massless particle from another massless particle.

As discussed in section 2.1.4, the UV divergences are absorbed into the definition of α_S by the introduction of the renormalisation scale μ_R , which takes the place of Q^2 in our previous formulation. Similarly, the IR divergences can be absorbed into the definition of the *Parton Distribution Functions* (PDFs), $f_i(x, \mu_F^2)$. In this way, the PDFs can be seen as the probability that the associated parton will take part in the interaction given the fraction of the longitudinal proton momentum x required and the factorisation scale of the interaction, here used interchangeably with Q^2 .

A reasonable follow up to this formulation is, given that α_S can be calculated at any Q^2 given a reference scale, can the same be done for PDFs? Were this the case, the PDFs would need only be defined at a single scale, and would be calculable up to any Q^2 – this is an important requirement, given that PDFs are not directly calculable with our current technology and must be measured. The ability to “evolve” PDFs from one scale to another means that all PDF measurements can contribute to a single measurement at one particular scale, giving fantastic precision in comparison to the opposite case where each collision at a different energy scale could rely only on measurements at the same scale to define the structure of the colliding protons.

Fortunately, this is precisely the objective of the Dokshittzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) parton evolution equations [30–32]:

$$Q^2 \frac{\partial f_i(x, Q^2)}{\partial Q^2} = \frac{\alpha_S(Q^2)}{2\pi} \sum_{j=q,g} \int_x^1 \frac{d\xi}{\xi} P_{ij}\left(\frac{x}{\xi}, \alpha_S(Q^2)\right) f_j(\xi, Q^2), \quad (2.40)$$

in which the DGLAP splitting kernels, P_{ij} , represent the probability of a parton j carrying momentum fraction ξ to become parton i with momentum fraction x through the emission of another parton - these splitting kernels are depicted in

fig. 2.6. At leading order (LO), they are as follows:

$$\begin{aligned}
P_{qq}(z) &= C_F \left[\frac{1+z^2}{(1-z)_+} + \frac{3}{2} \delta(1-z) \right], \\
P_{qg}(z) &= T_R [z^2 + (1-z)^2], \\
P_{gq}(z) &= C_F \left[\frac{1+(1-z)^2}{z} \right], \\
P_{gg}(z) &= 2C_A \left[\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right] + \frac{11C_A - 4n_f T_R}{6} \delta(1-z),
\end{aligned} \tag{2.41}$$

where $z = x/\xi$ is the momentum ratio of the parton after the splitting to before, C_F , T_R and C_A are the QCD colour factors with values $4/3$, $1/2$ and 3 , respectively. The ‘+’ subscript indicates the regularisation of divergences through the following means:

$$\int_0^1 dz (g(z))_+ f(z) = \int_0^1 dz g(z) (f(z) - f(1)). \tag{2.42}$$

The δ terms are additional terms arising from virtual corrections to the emission diagrams; their coefficient is obtained using the momentum sum rule:

$$\sum_{i=q,\bar{q},g} \int_0^1 dx (x f_i(x)) = 1, \tag{2.43}$$

and the number sum rules:

$$\begin{aligned}
\int_0^1 dx (f_u(x) - f_{\bar{u}}(x)) &= 2, \\
\int_0^1 dx (f_d(x) - f_{\bar{d}}(x)) &= 1, \\
\int_0^1 dx (f_s(x) - f_{\bar{s}}(x)) &= 0.
\end{aligned} \tag{2.44}$$

corresponding to the requirement of two up, one down and zero strange valence quarks, respectively. The splitting functions have been calculated to NLO by Curci et al. in 1980 [33] and to NNLO by Vogt et al. in 2004 [34].

By assuming that the PDFs factorise from the hard cross section as they do in eq. (2.39)⁶, measurements from a range of processes and experiments can be used in tandem to constrain the proton PDFs. This will be discussed in more detail in chapter 9.

⁶At the time of writing, this is indeed only an assumption, albeit one which is expected to hold true by many. A similar equation for lepton-hadron scattering is proven through the factorisation theorem [35].

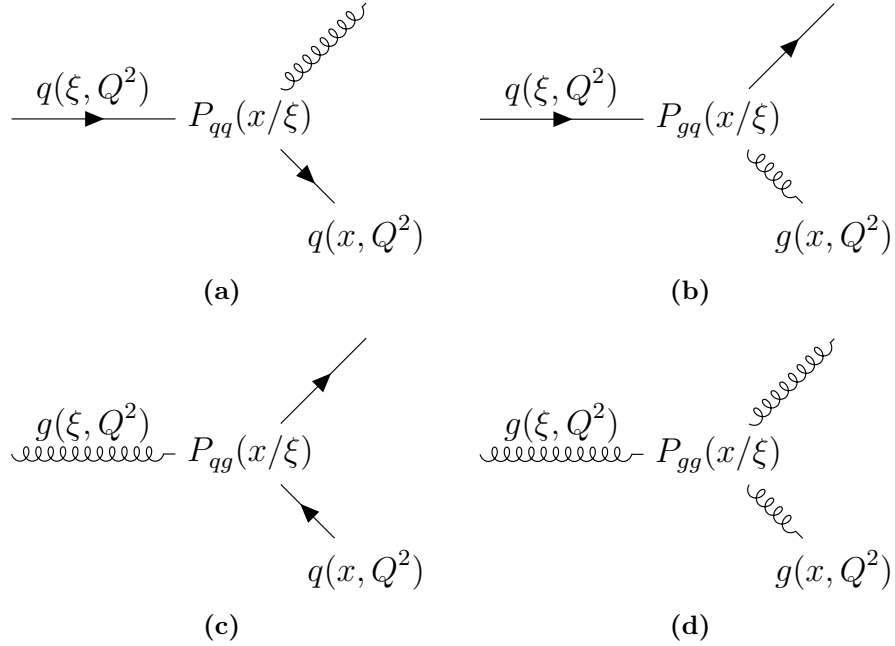


Figure 2.6: Feynman diagrams depicting the (a) quark-quark, (b) gluon-quark, (c) quark-gluon and (d) gluon-gluon LO splitting kernels of the DGLAP evolution equations.

2.3 The production of a W boson in association with jets at a proton-proton collider

The production of a W boson in association with jets – dubbed “ W + jets” from here onwards – contains elements of all theory discussed so far. In essence, the production of the W -boson draws directly from the Electroweak discussion, and the jets are objects driven fundamentally by QCD. Furthermore, the cross-section relies intimately on the PDFs – while it is true that this is the case for all processes, W + jets is one with properties which provide particularly strong constraining powers in PDF determination, as will be discussed towards the latter part of this section.

Firstly, an understanding of W + jets requires an understanding of the “jet” object itself. A jet is always associated to emission of a free quark or gluon. The non-Abelian nature of QCD means that the mediator, the gluon, can self-interact, resulting in a force between two colour-charged particles which is constant regardless of their separation. It therefore quickly becomes energetically favourable for the production of new particles which form colour-neutral states with the previous particles. For this reason, neglecting energy scales for which the previously

mentioned asymptotic freedom occurs, only colour-neutral states are observed in nature - this is known as *colour confinement*.

It is because of this that when a process occurs which emits a free quark or gluon, that particle cannot be observed. It instead undergoes a process denoted as *fragmentation* or *hadronisation*, in which a shower of particles is created; this is the jet object ⁷.

The measurements in this thesis will involve only the leptonic decay of the W boson as this gives a clear signature in the detector. Although the branching ratio of the decay to two quarks is significantly higher as shown in fig. 2.7, the process is far more difficult to distinguish from common backgrounds, to the degree that it would not even be the biggest contributor to the measured data. Furthermore, the kinematics of the W boson would have significantly larger uncertainties from mismeasurement and jet calibration, and the SM calculation would be significantly less accurate.

The most common W + jets production modes at a proton-proton collider are quark-gluon fusion and initial-state radiation (ISR), displayed in fig. 2.8. These production modes are individually highly sensitive to different PDFs in different regions of kinematic space.

Firstly, quark-gluon fusion occurs at the range of Bjorken x in which the valence quarks are most prevalent, where $u_v g \rightarrow W^+ + \text{jet}$ and $d_v g \rightarrow W^- + \text{jet}$. As the gluon is flavour blind, this means that the ratio W^+/W^- can be directly sensitive to the ratio u_v/d_v .

Secondly, the diagram with initial-state radiation shows how the sea quarks can contribute to the production cross-section. Again, the main production modes involve the valence quarks as these are very prevalent at the range of x probed, so

⁷The mediator for the electromagnetic force can also fragment (or as it is usually called, “shower”), and in fact it does do so within a detector at the LHC. The crucial difference between these processes is that EM showers are typically far more collimated than jets, and the initial particle can often be identified. While the weak bosons do also undergo this process, the fragmentation effects are very short lived if they occur at all as the bosons are massive – it is far more likely that they decay either to a lepton pair or a quark pair, which can then of course undergo fragmentation.

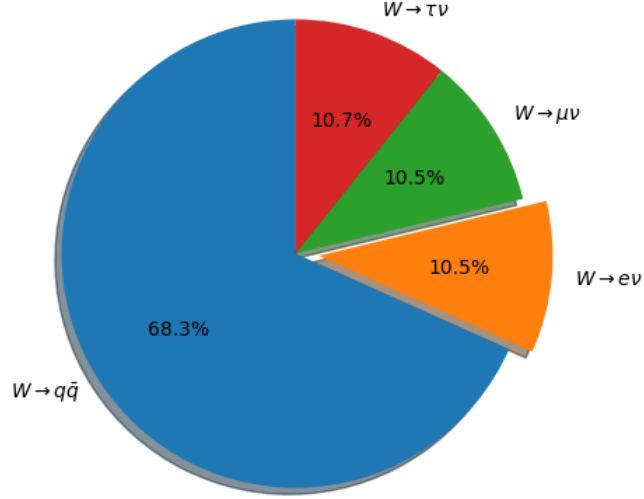


Figure 2.7: A pie chart displaying the branching ratios of the W boson decay. The decay modes themselves are labelled on the perimeter of the chart.

the main contributors will be $u_v\bar{d} \rightarrow W^+$, $u_v\bar{s} \rightarrow W^+$ and $d_v\bar{u} \rightarrow W^-$. The first two are separable via the CKM matrix elements, where

$$\sigma_{u_v\bar{d} \rightarrow W^+} \sim u_v(x)\bar{d}(x)\cos^2(\theta_C) \quad \text{and} \quad \sigma_{u_v\bar{s} \rightarrow W^+} \sim u_v(x)\bar{s}(x)\sin^2(\theta_C). \quad (2.45)$$

It is for these reasons that, although $W + \text{jets}$ alone would be unable to determine PDFs, given a set of data which can determine a prior set of distributions, significant contributions can be made to constrain the valence distributions and separate the strange and down distributions.

In recent years, multiple advancements have been made in the prediction of $W + \text{jets}$ differential cross sections, including calculations to NLO for up to five additional jets [36–38] and for one additional jet at NNLO [39–42]. Additionally, new merging approaches for NLO predictions over a range of jet multiplicities [43–45] and new parton shower approaches [46, 47] have been released. These new approaches to theoretical predictions must be tested to their limit, and $W + \text{jets}$ provides that test as it is one of the most prevalent at the LHC, with statistics second only to QCD jet production and photon production as shown in fig. 2.9, with a very

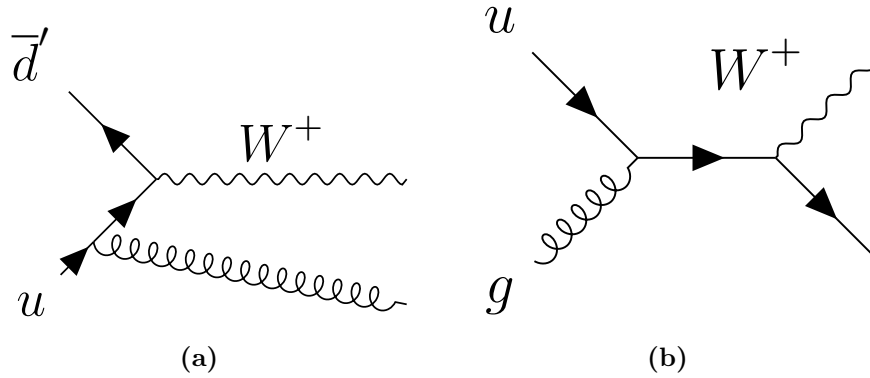


Figure 2.8: Feynman diagrams of the main production modes, initial-state radiation (a) and quark-gluon fusion (b), of the $W + \text{jets}$ process at a proton-proton collider.

clear signature in the detector (as discussed in detail in chapter 4), and without significant potential to be influenced by new physics. As new approaches continue to emerge, it is vital that they can be directly validated by data such as these.

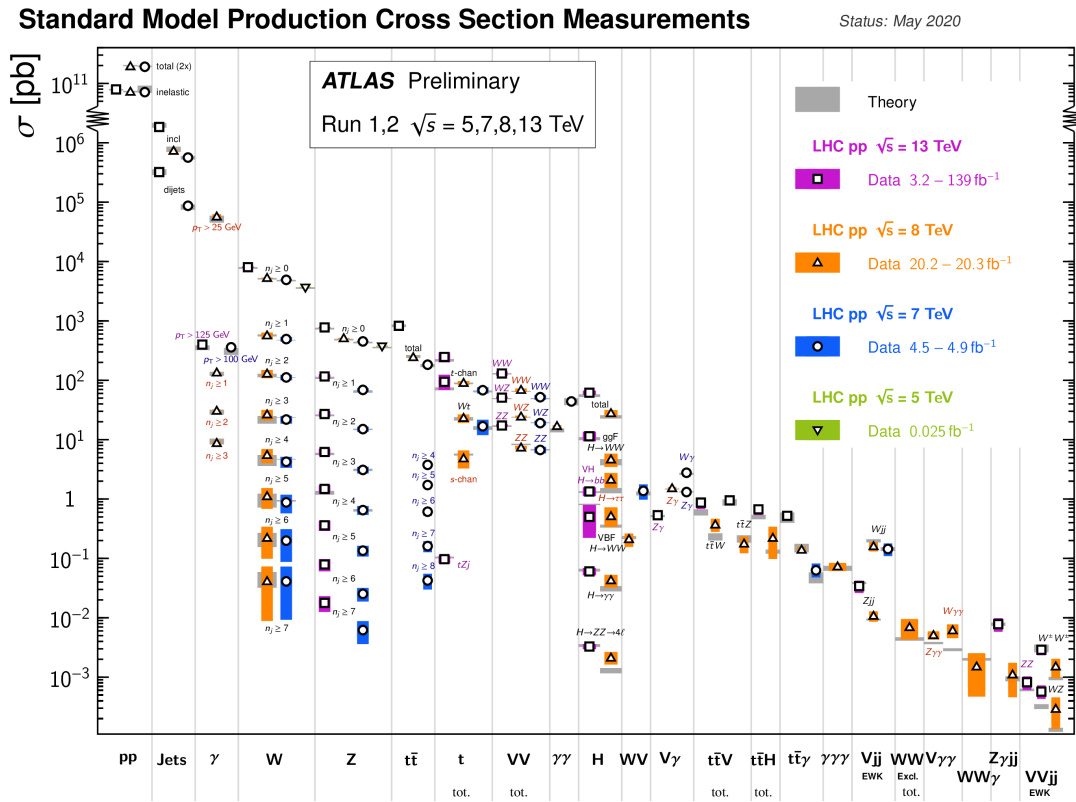


Figure 2.9: A summary of Standard Model measurements taken by the ATLAS collaboration, ordered by production rate [48]. Each process is labelled on the x -axis. W production has a rate lower than only γ and jets production; those high statistics are also present in W + jets production.

3

The ATLAS Experiment at the LHC

Contents

3.1	The Large Hadron Collider	31
3.2	The ATLAS detector	33
3.2.1	The Inner Detector	34
3.2.2	The Calorimeters	37
3.2.3	The Muon Spectrometer	39
3.2.4	The Trigger and Data Acquisition	40
3.3	Radiation damage studies of the SemiConductor Tracker optical links	41
3.3.1	Studies prior to LHC operation	42
3.3.2	Analysis method	43
3.3.3	Results	49
3.3.4	Conclusion	54

The Large Hadron Collider (LHC) [49] is a circular particle accelerator designed to collide proton beams with a centre-of-mass energy of 14 TeV and a luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. At the time of writing this thesis, this luminosity is the highest of any hadron collider, and the centre-of-mass energy is not matched by any other accelerator. Additionally, a range of other ions can be collided at various centre-of-mass energies and luminosities, for example heavy lead (Pb) ions with a centre-of-mass energy of up to 5 TeV per nucleon pair and a peak luminosity of over $10^{27} \text{ cm}^{-2} \text{ s}^{-1}$.

There are, in total, 7 experiments situated on the LHC ring: ATLAS (**A** **T**oroidal **L**HC **A**pparatus) [50], CMS (**C**ompact **M**uon **S**olonoid) [51], LHCb (**L**arge **H**adron **C**ollider **b**eauty) [52], ALICE (**A** **L**arge **I**on **C**ollider **E**xperiment) [53], LHCf (**L**HC **f**orward) [54], TOTEM (**T**OTAL **E**lastic and diffractive cross section **M**easurement) [55] and MoEDAL (**M**onopole and **E**xotics **D**etector **a**t the **L**HC) [56]. Each is designed and can be operated independently of the others¹, with optimisation towards an often specific goal. For example, LHCb is optimised for the measurement of beauty- and charm-meson decays by placement in the forward region relatively close to the beam pipe. It is also one of just two LHC experiments to use a Ring-Imaging Cherenkov (RICH) detector, allowing it to distinguish between charged K - and π -mesons in decay products, and the only experiment to use a Vertex Locator (VELO), a tracking system situated just 6 mm from the collision point which can pick out secondary decay vertices to a significantly higher precision than any other LHC experiment.

ATLAS is one of two “general purpose” detectors situated on the LHC ring; the other is CMS. Both are cylindrical detectors designed with accurate electromagnetic and hadronic calorimeters and high-resolution tracking, with strong magnetic fields, though the precise implementation differs significantly. This allows each experiment to identify and measure the four-momenta of leptons (both electrons and muons), photons and hadronic jets.

As this thesis uses data collected by the ATLAS detector, the following presents an overview of the LHC in section 3.1 and the ATLAS detector in section 3.2. Furthermore, research completed by the author on the radiation damage suffered by the optical links of the ATLAS semiconductor tracker is discussed in section 3.3.

3.1 The Large Hadron Collider

The LHC at the European Organisation for Nuclear Research known as CERN (derived from the name Conseil européen pour la recherche nucléaire), is located

¹Although TOTEM and CMS were originally designed independently, and were operated separately, as of 2015 their operation is coordinated in order to produce combined measurements.

by the city of Geneva, Switzerland. It is effectively a 27 km-long chain of superconducting magnets constructed in the circular tunnel of the former LEP collider at a depth between 45 m and 170 m under the French-Swiss border.

The LHC is a synchrotron designed to produce pp collisions at a centre-of-mass energy of 14 TeV. Protons are not fundamental particles and have substructure, as discussed in detail in section 2.2, so the initial state of each collision is unknown. This makes modelling of events difficult in comparison to, for example, an e^+e^- collision and often introduces large irreducible systematic uncertainties to measurements made; those relevant to the measurement of $W + \text{jets}$ are discussed in chapter 7. However, the upside of using protons is that energy loss due to synchrotron radiation scales with the inverse of the particle mass to the fourth power, allowing protons to reach substantially higher energies than electrons in the same complex for a fraction of the power [57]. In order to reach such unprecedented energy, ionised atoms are passed through a series of accelerators increasing in energy at each step as shown in fig. 3.1. Hydrogen atoms are first ionised in an electric field and the resulting protons are passed through the linear accelerator LINAC2, where they are collected into bunches of approximately 1.15×10^{11} protons and accelerated to 50 MeV. In approximately 17 s, the proton bunches are subsequently accelerated in sequence up to 1.4 GeV by the proton synchrotron called BOOSTER, to 25 GeV by the Proton Synchrotron (PS), and up to 450 GeV by the Super Proton Synchrotron (SPS). Thereafter, the final step is injection into the LHC, where the bunches are further accelerated to their final energies.

When the bunches reach their final energies, they are condensed into a tighter beam and validated for physics. Collisions start when the beams are declared stable, and continue until the luminosity has decreased by approximately 50%, at which point a new fill starts.

The operation of the LHC is divided into several runs, each lasting multiple years, separated by “long shutdowns” for essential repairs and upgrades. The first run, referred to as Run 1, lasted from 2010 to 2012. Due to a failure of the LHC which occurred in 2008 involving a cryogenic quench, Run 1 was operated

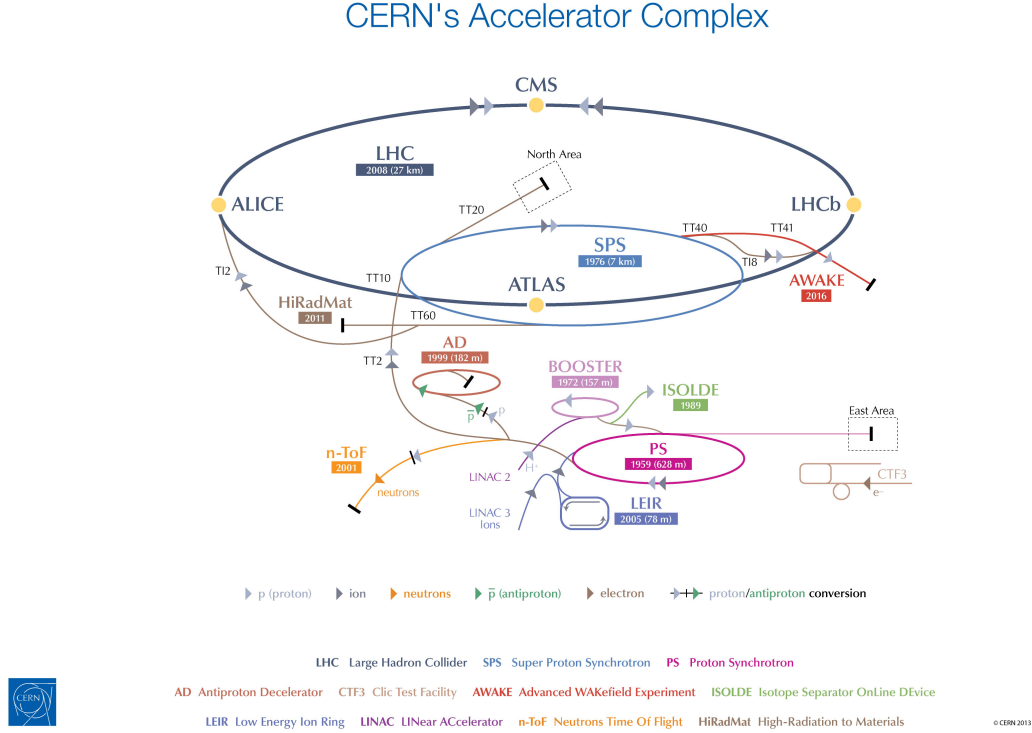


Figure 3.1: A graphic showing the CERN accelerator complex including the LINAC2, BOOSTER, PS, SPS, LHC and collision points[58].

conservatively at 7 TeV and 8 TeV during the years 2010-2011 and 2012, respectively, collecting a total of 26.1 fb^{-1} of data in that time. The results of Run 1 were a huge success, including the discovery of the Higgs boson at both the ATLAS and CMS detectors in 2012 [7, 8].

The second run of the LHC, referred to as Run 2, operated from Spring 2015 until Winter 2018 at a centre-of-mass energy of 13 TeV. An integrated luminosity of 4.2 fb^{-1} , 38.5 fb^{-1} , 40.2 fb^{-1} and 58.5 fb^{-1} was delivered to the ATLAS detector in each respective year. Run 3 is expected to begin in 2022, after a long shutdown in the years 2019–2021 during which crucial repairs, upgrades and test beams are taking place in anticipation of collisions at a centre-of-mass energy of 14 TeV.

3.2 The ATLAS detector

The ATLAS detector is one of two general purpose detectors on the LHC ring, the other being the CMS experiment. It is situated at interaction point 1, and is designed

similarly to many other particle physics collider experiments as a series of concentric cylinders around the beamline. Each layer has its own specific function and is centred on the interaction point. The outer layer of the detector is up to 25 m in diameter and 44 m in length, and all layers together weigh approximately 7000 tons.

The ATLAS detector is composed of four major subsystems, as illustrated in fig. 3.2. These are an inner tracking detector (ID), composed of three sub-detectors as discussed in section 3.2.1; calorimeter systems as discussed in section 3.2.2, and a muon spectrometer described in section 3.2.3. Both the inner detector and the muon spectrometer benefit from an immersive magnetic field, allowing charged particle momenta to be calculated from the curvature of their path through each system. The inner detector is surrounded by a 2 T solenoidal magnet, and the muon spectrometer uses a system of three toroidal magnets ranging from 0.5 to 1 T in field strength. Each of these is designed such that the magnetic field perpendicular to the direction of the beam at all points along the beam pipe is zero, thus there is no interaction between the particle beams and the magnetic fields.

Each component is designed to withstand the high-radiation environment that is the surrounding area of the LHC collisions for a sufficient period of time, at least one Run but ideally multiple. A study of the radiation damage suffered by the optical links which transfer data to and from one of the subcomponents of the ID was performed by the author of this thesis and is reported in section 3.3.

ATLAS uses a right-handed coordinate system, as shown in fig. 3.2, with its origin at the interaction point, the x -axis pointing towards the centre of the LHC ring, the z axis pointing down the beamline and the y axis pointing upward. Cylindrical coordinates (r, ϕ, z) are used in the plane transverse to the beamline, in which ϕ is the azimuthal angle anticlockwise from the x axis and r is the radial distance from the beamline. A longitudinal angle, θ is used to label the angle between a vector and the z axis. The pseudorapidity, defined in terms of the longitudinal angle as $\eta = -\ln \tan(\theta/2)$, is often used as differences in pseudorapidity are approximately lorentz-invariant for highly-relativistic particles; a figure showing the pseudorapidity

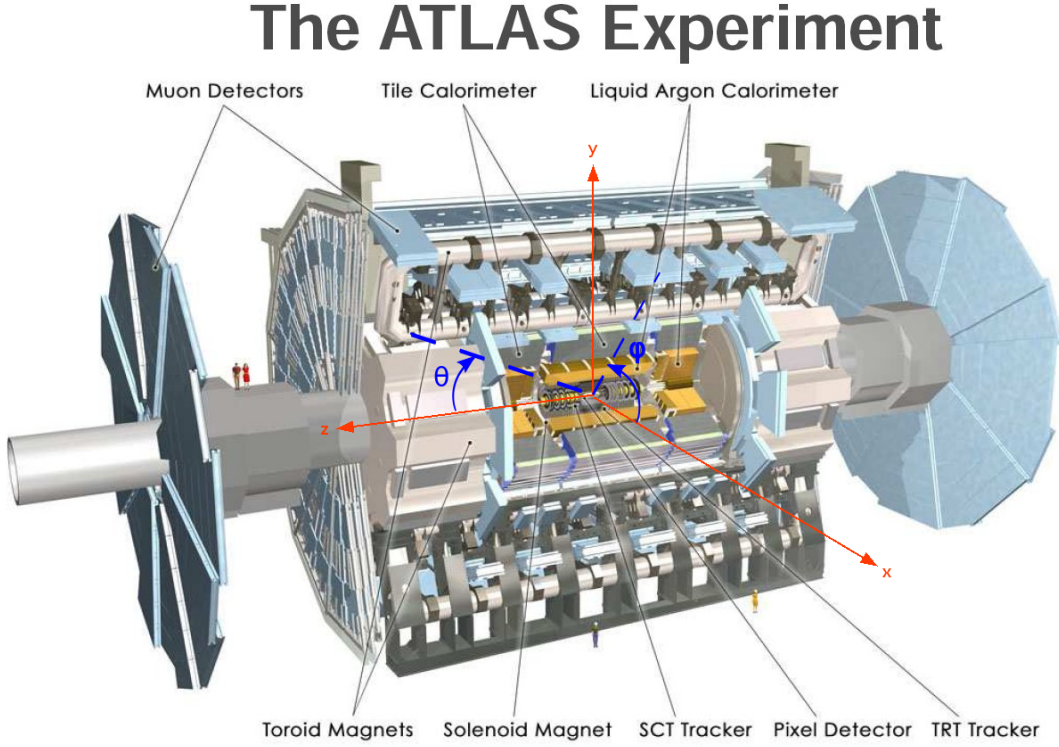


Figure 3.2: A computer-generated image of the ATLAS detector [50]. The three inner-detector components, the calorimeters and muon chambers, as well as the solenoidal and toroidal magnets, are annotated. The coordinate system is drawn in red arrows, with angles labelled in blue (see text for more details).

at certain angles of θ is given in fig. 3.3. The angular separation of two particles emerging from the interaction point is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

3.2.1 The Inner Detector

The ATLAS Inner Detector (ID) [60] provides precision tracking of charged particles between 33.25 mm and 1082 mm from the beam axis with a momentum resolution of $\sigma_{p_T}/p_T = 0.05\% \oplus 1\%$ for charged particles with $p_T > 0.1$ GeV over the pseudorapidity range $|\eta| < 2.5$. This is achieved by three separate subsystems with different ranges of radii from the beam axis: the Pixel Detector, the SemiConductor Tracker (SCT) and the Transition Radiation Tracker (TRT).

The Pixel Detector [61] is the closest subsystem to the beam pipe. As it is so close, a larger angle is covered by a smaller area, which therefore receives a larger particle flux and requires significantly higher granularity than the other subsystems.

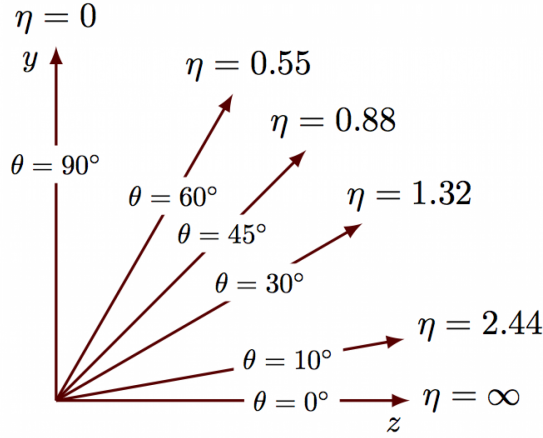


Figure 3.3: A graphic showing how the pseudorapidity, η , varies with the polar angle, θ [59].

This is achieved with four barrels and six end-cap layers (three on each side) of silicon semiconductor pixel sensors. Each pixel is a reverse-biased p-n junction, which produces an electron-hole pair (and therefore an electrical current when the electron and hole each drift to their respective electrodes) when traversed by a charged particle, allowing a track to be traced by connecting the “hits”. There are approximately 80 million $50 \times 400 \mu\text{m}^2$ pixels in total covering pseudorapidity $-2.5 < \eta < 2.5$ in the radial range between 5 and 12 cm. This provides a spatial hit resolution of $10 \mu\text{m}$ in the ϕ plane and $115 \mu\text{m}$ along z and r .

During the first long shutdown between Run 1 and Run 2, a further barrel layer was installed at a radius of 3.3 cm from the beam axis, known as the Insertable B-Layer (IBL). This provides an additional eight million pixels over twelve sectors. The addition of the IBL improves track construction and provides more precise vertex measurement, but its main purpose (for which it was named) is to improve the identification of jets originating from b -quarks [62].

The SemiConductor Tracker (SCT) [63] lies outside of the Pixel detector and is also composed of silicon semiconductor sensors, however instead of pixels the SCT is segmented into strips with a pitch of $80 \mu\text{m}$. The SCT consists of four cylindrical layers in the barrel region, with strips aligned along the beamline, and nine disks, with strips aligned radially outwards, forming the endcaps. For each module, two individual layers of strips are *stereo-paired*, meaning overlaid offset by a small angle,

20 mrad in this case, around the centre of the module. This stereo-pairing results in an improved spatial resolution along the direction of the strips. A hit along a strip has an intrinsic resolution of $17\text{ }\mu\text{m}$ in ϕ and $580\text{ }\mu\text{m}$ in z and r .

Finally, the Transition Radiation Tracker (TRT) [64] is the outermost subcomponent of the ID. The TRT is composed of straw tubes measuring 1.44 m in length and 4 mm in diameter. Each straw is filled with a gas mixture composed of 70% Xe, 27% CO_2 and 3% O_2 . Through the centre of each tube is a gold-coated wire, which is $31\text{ }\mu\text{m}$ in diameter, serving as an anode kept at ground potential. The walls of each straw serve as a cathode, kept at negative potential of approximately -1.5 kV . Charged particles traversing the straw ionise the gas mixture, creating electrons and positive ions. These free charges drift in the electric field, producing a detectable electrical signal proportional to the energy deposited by the particle. The TRT does not provide tracking information in the direction parallel to the straws.

The barrel region consists of 72 layers of straw tubes covering a radius between 0.563 m and 1.066 m aligned along the beam axis. The endcap regions consist of 160 layers of 0.36 m-long tubes which are radially oriented. There are a total of 350,848 straw tubes.

Each layer of straws is separated by a polypropylene radiator which changes the refractive index of the volume and provides discrimination between electrons and heavier charged particles, as lighter particles give a notably larger amount of transition radiation than heavier particles such as mesons or hadrons. The TRT, therefore, provides substantial discriminating power between electrons and heavier particles over the energy range between 1 GeV and 200 GeV.

Track-finding algorithms are subsequently used to combine hits from all three ID subsystems to give a charged particle track. The entire ID is submersed in a 2 T solenoidal magnetic field; this bends the particle track, without affecting the energy or the magnitude of the momentum. The radius of the track in the plane perpendicular to the magnetic field is proportional to the momentum, thus the momentum of the particle transverse to the beamline can be calculated.

3.2.2 The Calorimeters

Most SM particles are, after exiting the ID, subsequently stopped in the ATLAS calorimeters [50], shown in fig. 3.4. These are sampling calorimeters, consisting of alternating layers of dense passive material and an active medium. When passing through the passive material, particles induce a cascade of secondary particles of lower momentum than the preceding particle, referred to as a *particle shower*. This induces a signal in the active medium through ionisation of scintillation proportional to the total released energy. The fractional resolution of a calorimeter is typically expressed as the quadrature sum of three terms:

$$\frac{\sigma_E}{E} = \frac{N}{E} \oplus \frac{S}{\sqrt{E}} \oplus C \quad (3.1)$$

where N is the measurement of noise due to background and electronics, which is independent from the energy measurement; S parameterises the poissonian stochastic uncertainty proportional to $\sqrt{E}$ arising from the random-sampling nature of each measurement, and C is a constant term reflecting non-uniformities in the detector.

The ATLAS calorimeters are split into two regions: the electronic calorimeter (ECal) and hadronic calorimeter (HCal). Both have an acceptance of $|\eta| < 3.2$, except for a gap in the electronic calorimeter between the barrels and endcaps within the range $1.37 < |\eta| < 1.52$.

The ECal is designed for the detection and measurement of electrons and photons. The passive material is composed of lead, a heavy absorber material with which both charged and neutral incoming particles interact, while the active material is composed of liquid argon (LAr), chosen as it reacts in the presence of charged particles, is radiation-hard and stable over time. The argon is held in a liquid state through cryostats, cooling it to 89 K. The lead induces bremsstrahlung radiation in passing electrons and pair production in photons, thus resulting in an electromagnetic shower which ionises the LAr. An electrical signal is produced by the subsequent drift of the liberated electrons towards electrodes, which can subsequently be processed by readout electronics. The ECal is finely segmented,

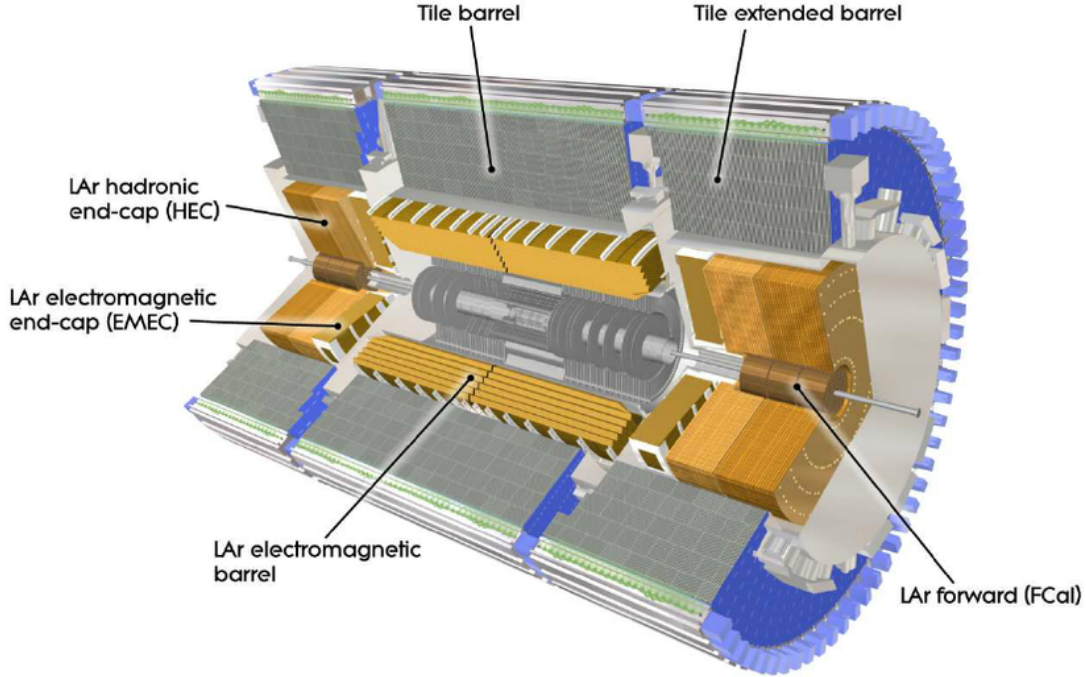


Figure 3.4: A computer-generated image of the ATLAS calorimeter system [50]. The subcomponents are annotated.

with the dimensions of each cell being $\Delta\eta \times \Delta\phi = 0.003 \times 0.025$, allowing the detector to distinguish a single photon from most $\pi^0 \rightarrow \gamma\gamma$ decays, in which the photons could otherwise appear to be merged within the detector. The total thickness of the ECal is between 22 and 33 radiation-lengths, ensuring that the energy of electrons and photons is almost completely contained.

The stochastic and constant terms discussed in eq. (3.1) were measured to be:

$$\frac{\sigma_E}{E} \sim \frac{10\%}{\sqrt{E}} \oplus 0.7\% \quad (3.2)$$

The HCal, on the other hand, is optimised for the detection and measurement of hadrons. It is subdivided into three parts: the Tile hadronic calorimeter (TileCal), the LAr hadronic end-cap calorimeter (HEC) and the LAr forward calorimeter (FCal). The LAr calorimeters are of a very similar structure to that used in the ECal, however the HEC uses copper as the passive medium and the FCal uses copper in the first layer and tungsten for the others. Conversely, the TileCal makes use of steel as an absorbing material with plastic polystyrene scintillator tiles as

the active medium. The tiles are connected to the photomultiplier tubes (PMTs), which measure emitted photons from the scintillation via wavelength-shifting fibres. There are a total of 9852 PMTs servicing the TileCal.

The stochastic and constant terms were measured in test beams to be:

$$\begin{aligned}\frac{\sigma_E}{E} &\sim \frac{53\%}{\sqrt{E}} \oplus 3\%, \\ \frac{\sigma_E}{E} &\sim \frac{71\%}{\sqrt{E}} \oplus 1.5\%, \\ \frac{\sigma_E}{E} &\sim \frac{94\%}{\sqrt{E}} \oplus 7.5\%,\end{aligned}\tag{3.3}$$

for the TileCal, HEC and FCal, respectively.

3.2.3 The Muon Spectrometer

The Muon Spectrometer (MS) is the outermost sub-detector of ATLAS. Surrounding the calorimeters entirely, it is devoted to tracking the position of any charged particle which passes through the calorimeters unperturbed; this particle is assumed to be a muon. The curvature of the track is subsequently used similarly as in the ID to measure charge and momentum as a large air-core toroid magnet system produces a magnetic field of strength varying between 0.5 and 1 T throughout the subsystem.

The MS consists of approximately one million channels spread over three concentric cylinders in the barrel region and two wheels covering the end-cap on either side. The innermost and outermost barrels are at a distance of five and ten metres from the beampipe, respectively. The system is designed to measure the momentum of muons above 5 GeV and provide a resolution of 3% at 100 GeV. Coverage extends to $|\eta| = 2.7$.

The muon chambers consist of two types: the Monitor Drift Tubes (MDT) and Cathode Strip Chambers (CSC). The MDTs cover the majority of the MS, consisting of 3 cm diameter drift tubes which contain a mixture of 93% Argon and 7% CO₂. These operate similarly to the passive material of the calorimeters, where passing charged particles ionise the gas and the drift of these ions induces a signal. This drift is facilitated by a single tungsten-rhenium wire in each tube operating at a voltage of 3 kV. The typical spatial hit resolution of a single tube is between 50

and 100 μm , depending on whether three or four layers of tubes are used, which varies throughout the sub-detector. On the other hand, the CSCs are composed of multi-wire proportional chambers with orthogonal planar cathodes. These have a higher radiation tolerance and can resolve a higher occupancy than the MDTs, making them ideal for regions with larger particle flux. They are therefore placed in the forward region $2 < |\eta| < 2.7$ and the innermost barrel.

3.2.4 The Trigger and Data Acquisition

A single bunch-crossing at the ATLAS detector translates to approximately 1 MB of data in total across all of the subdetectors. Given that bunches cross every 25 ns at the ATLAS interaction point, this provides 40 TB of data each second. With present technology, it is simply not possible to process this amount of data within a reasonable time frame such that the processors are available for the next round of data-taking at the next collision.

However, even the most common processes measured at the ATLAS physics program are at least five orders of magnitude below the total proton-proton collision cross-section, as shown in fig. 2.9. It is, therefore, essential that ATLAS employs a *trigger* system, designed to significantly reduce the amount of data processed and stored by neglecting events *online*, meaning concurrent with data-taking, to a manageable few MB/s.

The ATLAS trigger consists of a hardware-based trigger (denoted L1, or Level-1) and a software-based trigger (named the High-Level Trigger, denoted HLT). For the L1, coarse granularity in the calorimeters or muon spectrometer is used to determine whether the event should be kept. Essentially, events only proceed to further processing if there is a hit in a calorimeter or muon spectrometer, implying the production of a high-energy lepton, photon or jet; the signal close to the hit is denoted as a “region of interest” for the following step. This reduces the event rate from the LHC bunch crossing rate to approximately 100 kHz. The L1 processing time is approximately 2.5 ns, though it is worth noting that 1 ns of this is taken up by signal propagation from the detector to the off-detector electronics within the cavern.

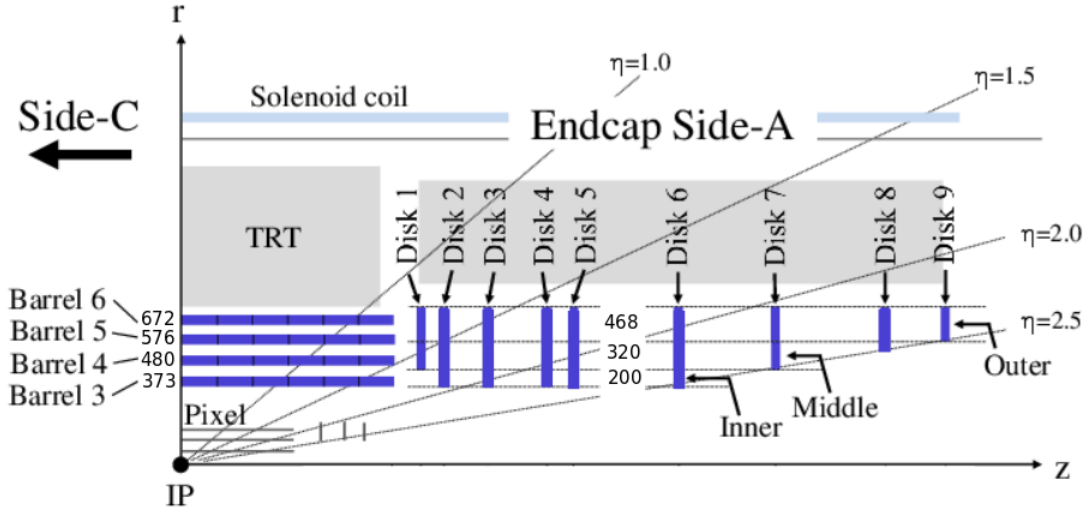


Figure 3.5: The geometry of the SCT, with the regions used in this analysis labelled. The number of modules in each region is also labelled. The SCT is symmetric, with endcap C reflecting endcap A [66].

The regions of interest formed at L1 are subsequently processed in the HLT phase, in which sophisticated selection algorithms use full granularity detector information in either the region of interest or the whole event. This reduces the rate by a further two orders of magnitude to 1 kHz with a processing time of approximately 200 μ s. Events passing both the L1 and HLT triggers are stored for further *offline* processing using the ATLAS Athena framework [65].

3.3 Radiation damage studies of the SemiConductor Tracker optical links

The transfer of data and commands between the SCT (the geometry of which is shown in fig. 3.5) and the offline processes is done via optical links. AlGaAs Vertical-Cavity Surface-Emitting Lasers (VCSELs) output 850 nm light through a step-index multi-mode optical fibre, which is received by a *p-i-n* silicon photodiode at the other end. There is one link from the data-acquisition (DAQ) crates to each module of the subdetector, and two links in the opposite direction. The current

produced by the photodiode is recorded and stored in an offline database. The optical links are described in further detail in reference [67].

This section details the analysis of this current from the 15 February 2015 to 24 October 2018. The aim is to search for evidence of radiation damage, and the nature thereof, in both the on-detector VCSELs and the on-detector *p-i-n* diodes. A summary of previous analyses will be given in section 3.3.1. The methods used in this analysis to extract the key variables are given in section 3.3.2, and the results are given in section 3.3.3. These studies were published in the Journal of Instrumentation [68].

3.3.1 Studies prior to LHC operation

VCSEL studies

A sample of 20 VCSELs from the wafer used for the SCT was irradiated using protons with a kinetic energy of 30 MeV [69], and a further sample from the same wafer was irradiated with protons of kinetic energy 24 GeV [70]. This was done with fluences of $2 \times 10^{14} \text{ cm}^{-2}$ and $1.09 \times 10^{14} \text{ cm}^{-2}$, respectively.

Assuming that the damage in a GaAs lattice scales with Non-Ionising Energy Loss (NIEL) [71], these fluences can be compared to each other and also the expected fluence during SCT operation in-situ at the LHC. For example, the 30 MeV proton irradiation is estimated to be equivalent to $\sim 1000 \text{ fb}^{-1}$ of integrated LHC luminosity.

At the end of irradiation, the slope efficiency (equal to the gradient of a linear fit to the optical power versus the laser drive current) decreased to a value of $(73 \pm 4)\%$ of the pre-irradiation efficiency. However, these devices also showed strong beneficial annealing, dependent on the VCSEL bias current. After one week of operation at a bias current of 10mA, the slope efficiency recovered to $(88 \pm 4)\%$ of the pre-irradiation value [69]. There are no theoretical models which can predict this annealing, but detailed phenomenological models can extrapolate short exposures and annealing periods to much longer exposure periods, such as those at the HL-LHC [72]. Compatible results were observed with the 24 GeV protons [70].

***p-i-n* photodiode studies**

Similar studies of *p-i-n* diodes were performed using protons and neutrons at a range of energies [73, 74], with the resulting fluences converted to equivalent fluences of 1MeV neutrons using the NIEL hypothesis [71]. A rapid initial decrease in responsivity was observed up to a fluence of $\sim 2.5 \times 10^{13}$ 1MeV n_{eq}/cm^2 , but only a very small further decrease thereafter². Note that for the SCT region with highest fluence, this is equivalent to $\sim 80 \text{ fb}^{-1}$ of integrated luminosity, and for the region of lowest fluence this is equivalent to $\sim 160 \text{ fb}^{-1}$. For a fluence of 3×10^{14} 1MeV n_{eq}/cm^2 , the output current (also referred to as the “responsivity”) of the photodiodes decreased to 71% of pre-irradiation values.

3.3.2 Analysis method

The results of this analysis are highly dependent on the fluence of highly-energetic particles through the component. At the ATLAS detector, this varies on a component-by-component basis, both because the detector is cylindrical, and therefore not spherically symmetric, but more importantly because the nature of proton-proton collisions is that the distribution of the transverse momentum of particles is non-trivial. For this analysis, the FLUKA particle transport code [75] was used to perform simulations of LHC collisions, thereby predicting the fluence per module of the SCT.

To provide statistically-significant insight, the SCT modules are grouped into “regions”. For the purpose of the following analysis, each barrel is considered as a region. The endcaps suffer a large variation in fluence over a single disk and over a single ring (referred to as *Inner*, *Middle* and *Outer* in fig. 3.5). Therefore, in order to reduce the spread of fluence over a single measurement, modules in the endcaps are grouped into numerically-labelled regions outlined in table 3.1. The range in fluence for each region in units of 1MeV n_{eq}/cm^2 is 1.52–1.89, 1.89–2.44 and 2.43–2.86 for regions 1, 2 and 3, respectively.

²Here, the unit “1MeV n_{eq}/cm^2 ” is a unit of fluence calculated for comparability between experiments. One unit signifies that the fluence is equivalent to a single neutron of energy 1 MeV per square centimetre.

Region	Disk number	Disk ring
1	1	outer
	2	outer
	3	outer
	4	outer
	5	outer
	6	outer
2	1	middle
	2	middle
	3	middle
	4	middle
	5	middle
	6	middle
3	8	outer
	7	middle
	2	inner
	3	inner
	4	inner
	9	outer
	5	inner
	6	inner
	8	middle

Table 3.1: The groupings of disk regions used in this analysis, ordered by fluence or each disk number and region.

VCSELs

Data are sent from the SCT to the off-detector systems using a Non Return to Zero (NRZ) coding. When the link is sending a ‘0’ or a ‘1’, the drive current is $\sim 1\text{mA}$ or $\sim 10\text{mA}$, respectively. It is therefore expected that the on-detector VCSELs receive a long period of beneficial annealing.

The light output is monitored in-situ using the photocurrents produced by the off-detector *p-i-n* diodes as a proxy. These instruments are not expected to degrade over this time period, so any degradation in this current is assumed to be due to the radiation damage of the on-detector VCSELs.

The monitoring of the *p-i-n* current, I_{pin} , is done by performing special scans periodically, in which a fixed pattern of data is sent to the counting room by each link and the total photo-current is measured. The total number of VCSEL channels per SCT region is described in table 3.2.

Table 3.2: The number of VCSEL channels per SCT region used in the analysis, where the regions are shown in fig. 3.5 and the endcap regions are grouped according to table 3.1. The number of active VCSELs varied by up to four depending on scan; only the minimum number has been quoted.

Region	Active VCSELs	Total VCSELs
Barrel 3	731	768
4	933	960
5	1130	1152
6	1299	1344
A-side region 1	556	704
region 2	450	608
region 3	424	664
C-side region 1	644	704
region 2	444	608
region 3	370	664

The current was normalised to the current at the start of the run on a per-channel basis, and the mean taken over the channels per region. The RMS deviation about the mean was also computed, and the uncertainty on the mean was taken by the central limit theorem as $\text{RMS}/\sqrt{N}$, where N is the number of channels in that region.

There were some outliers, caused by problems with the procedure to measure the photo-currents; these were removed by excluding values which were more than 4 standard deviations from the mean and recomputing. This procedure to remove outliers was iterated twice.

Subsequently, these measurements were fitted linearly against integrated luminosity and the gradient taken as the reduction in slope efficiency per fb^{-1} . The FLUKA particle transport code provided predictions of the fluence per region of the SCT, given in table 3.3. An overall fluence-independent measurement of reduction in slope efficiency for comparison to test beam studies was taken by plotting the gradient measurement against the fluence per region, fitting a linear relationship to this and extracting the gradient, giving a reduction in slope efficiency per $10^{11} \times 1\text{MeV}n_{eq}/\text{cm}^2$.

Table 3.3: The fluence for each region of the SCT analysed in the analysis, as estimated by the FLUKA particle transport code [75]. The regions referred to are shown in fig. 3.5 and the endcap regions are grouped according to table 3.1.

Region	Fluence ($10^{11} \times 1\text{MeV}n_{eq}/\text{cm}^2/\text{fb}^{-1}$)
Barrel 3	2.52
4	2.00
5	1.67
6	1.44
Endcap region 1	1.69
region 2	2.11
region 3	2.68

***p-i-n* photodiodes**

The responsivity of the on-detector *p-i-n* diodes is measured continuously while the detector is in operation by measuring the photo-currents. There is a pedestal current, I_{PED} , a small ($\sim 0.02\text{mA}$) current produced by the *p-i-n* diodes even when the VCSELs are off, which can be different for each module. This is subtracted from the measured value, and any measurements with $I_{PED} > I_{PIN}$ are excluded.

Data points are recorded only when the I_{PIN} value changes by more than 0.01mA . These recordings are averaged over a day by taking a mean, weighted by time:

$$\langle I_{PIN} \rangle = \frac{\sum_i \Delta_i I(t_i)}{\sum_i \Delta_i} \quad (3.4)$$

where $\Delta_i = t_{i+1} - t_i$ is the time interval of measurement i .

A number of optical links were swapped between data-taking in 2016 and 2017, in order to optimise bandwidth. These have been excluded from the analysis. Table 3.4 shows the number of modules per region used in this study.

The measured photocurrents were normalised per channel to the start of 2016 data-taking. These values were then averaged across the SCT regions as given in table 3.4. There are cases for which there is a small discontinuity over the technical stops between 2016 and 2017, and between 2017 and 2018; in these cases, a renormalisation correction is applied for continuity of the data. The uncertainty on each point is calculated as the $\text{RMS}/\sqrt{N}$.

The studies discussed in section 3.3.1 imply that a significant proportion of the damage would have occurred at the beginning of operation, ie. during the years of

Table 3.4: The number of p - i - n channels per SCT region used in this analysis, after swapped modules have been excluded.

Region	Active modules	Total modules
Barrel 3	373	384
4	472	480
5	554	576
6	664	672
A-side region 1	294	352
region 2	198	304
region 3	189	344
C-side region 1	333	352
region 2	185	304
region 3	164	344

Region	Effective integrated luminosity /fb ⁻¹
Barrel 3	23.39
4	23.10
5	22.95
6	22.69
Endcap region 1	23.04
2	22.51
3	22.43

Table 3.5: The effective 13 TeV-equivalent integrated luminosity of each region of the SCT used in this analysis.

2012, 2011, 2012 and 2015. However, data taken during this time are not studied as large temperature fluctuations occurred throughout. To account for this, the data are shifted by an effective integrated luminosity for each region, calculated as

$$L_{\text{eff}} = \frac{F(\sqrt{s})}{F(13 \text{ TeV})} L, \quad (3.5)$$

where L is the raw integrated luminosity and $F(\sqrt{s})$ is the fluence over that luminosity period, calculated as a function of the centre-of-mass energy.

For example, for the region labelled “Barrel 3”, the fluence and integrated luminosity for the year 2011 was $1.65 \times 10^{11} \text{ 1MeV} n_{eq}/\text{cm}^2/\text{fb}^{-1}$ and 5.5 fb^{-1} , respectively, and the fluence during the 13 TeV period is given in table 3.3, giving an additional effective integrated luminosity of 3.60 fb^{-1} . The effective integrated luminosity during the years 2011-2015 for each region is given in table 3.5.

The measurements are subsequently fitted against integrated luminosity in order to extrapolate to an asymptotic level of degradation as discussed in section 3.3.1. There are three conditions on this fit:

1. $I = I(L > 0)$,
2. $I(\infty) = R_\infty$,
3. $\frac{dI}{dL} < 0$.

The first condition is that the current is an analytic function of integrated luminosity, L , and that we are only interested in how the function behaves at $L > 0$. The second is the asymptotic degradation condition - it is expected that the photodiodes will not degrade beyond a certain point, defined as R_∞ , as seen in previous studies. Finally, the third condition is that we do not expect recovery at any point in the process. These conditions are encapsulated by a fitted function with a single condition:

$$I(L) = R_\infty + R_0 e^{f(0) - f(L)},$$

where $\frac{df}{dL} > 0$. (3.6)

Beyond this single remaining condition, the functional form of $f(L)$ is arbitrary. As a result, the aim of the analysis is to find functions which improve the goodness-of-fit, relative to the simplest possible form $f(L) = L$, and combine them linearly. Final functions which give the lowest χ^2 with the lowest number of free parameters are preferred. From each of these functions, the total degradation extrapolated over all years of data-taking (also referred to as the “Extrapolated R_∞ ”) is calculated as:

$$R_\infty^{\text{eff}} = \frac{R_\infty}{R_\infty + R_0}. \quad (3.7)$$

The “central” function provides the quoted value of R_∞^{eff} , and the alternative suitable function with the largest deviation in R_∞^{eff} provides a parameterisation uncertainty, defined as the difference between the two.

In order to estimate a systematic uncertainty pertaining to the stability of the optical links over this time period, the *p-i-n* diode photo-currents were measured

before the start of each run, and a linear fit performed, in order to determine the stability of the system. The stability is given by the gradient of each fit, and the RMS value was used as a contribution towards a systematic error on the final result. As this gradient is in units of per time, it is multiplied by the time over which measurements are taken. Discussion and results of this are given in appendix A.

3.3.3 Results

On-detector VCSELs

The resulting plots with linear fits superimposed are given in figs. 3.6 to 3.8, with the results summarised by table 3.6. The normalised overall slope efficiency reduction rate is given by the gradient of the linear fit shown in fig. 3.9, as described in section 3.3.2.

Table 3.6: Resulting fitted gradients for the on-detector VCSELs for the four barrel layers and three endcap rings for data from 2016 to 2018. The gradients are also normalised to fluences using the expected fluence from table 3.3.

Region	y-intercept	Gradient/ 10^{-4}	Reduced χ^2
Barrel 3	0.998 ± 0.001	-3.03 ± 0.14	3.5/7
4	0.999 ± 0.001	-1.76 ± 0.12	4.6/7
5	1.0003 ± 0.0009	-1.60 ± 0.10	1.9/7
6	0.9995 ± 0.0009	-1.07 ± 0.10	2.8/7
A side region 1	0.996 ± 0.001	-2.09 ± 0.16	14.4/7
2	1.000 ± 0.001	-1.81 ± 0.16	3.3/7
3	0.998 ± 0.001	-2.12 ± 0.15	4.5/7
C side region 1	0.998 ± 0.001	-2.00 ± 0.16	1.1/7
2	1.000 ± 0.001	-1.58 ± 0.17	2.0/7
3	0.999 ± 0.002	-2.16 ± 0.19	3.2/7

Combining the fitted gradients in this way gives an overall reduction rate of $(0.9 \pm 0.1)\%/10^{13} \times 1\text{MeV}n_{eq}/\text{cm}^2$, with a χ^2/nDoF of 66/8, as shown in fig. 3.9. The intercept is highly compatible with zero, implying that there is no change in output power when there is no fluence, as expected. Although this χ^2 does suggest an incompatibility, it is likely to result from an underestimation of uncertainties - that the regions have a non-zero spatial volume means the fluence itself has an associated uncertainty which, though efforts have been made to reduce this through

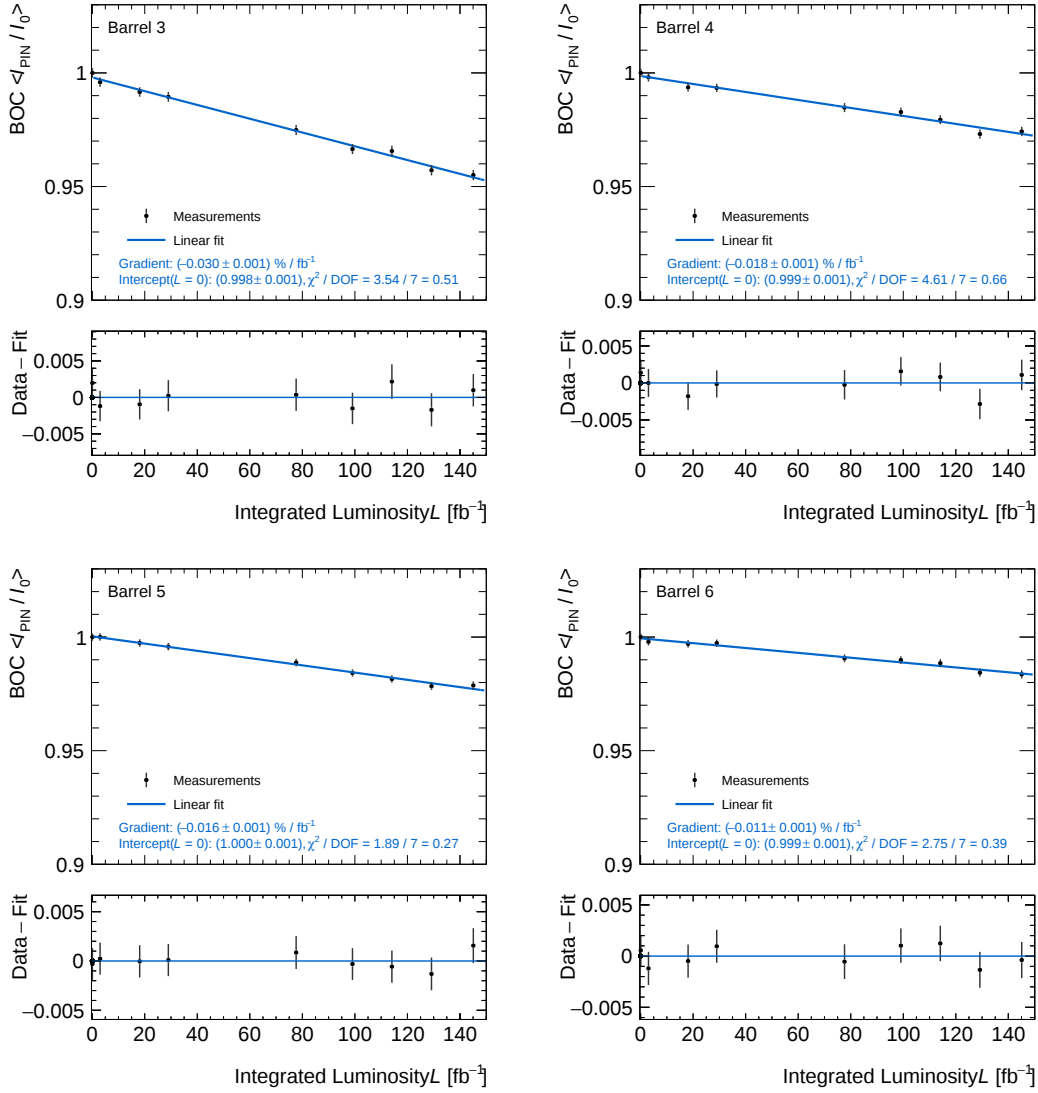


Figure 3.6: On-detector VCSELS: Barrels. The measured mean value of the photocurrents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018 .

the grouping of modules in to regions of similar fluence, is non-zero. The same regions of the A-side and C-side endcaps show consistent degradation, though they are incompatible with the best-fit line, again implying that the discrepancy is likely to be systematic.

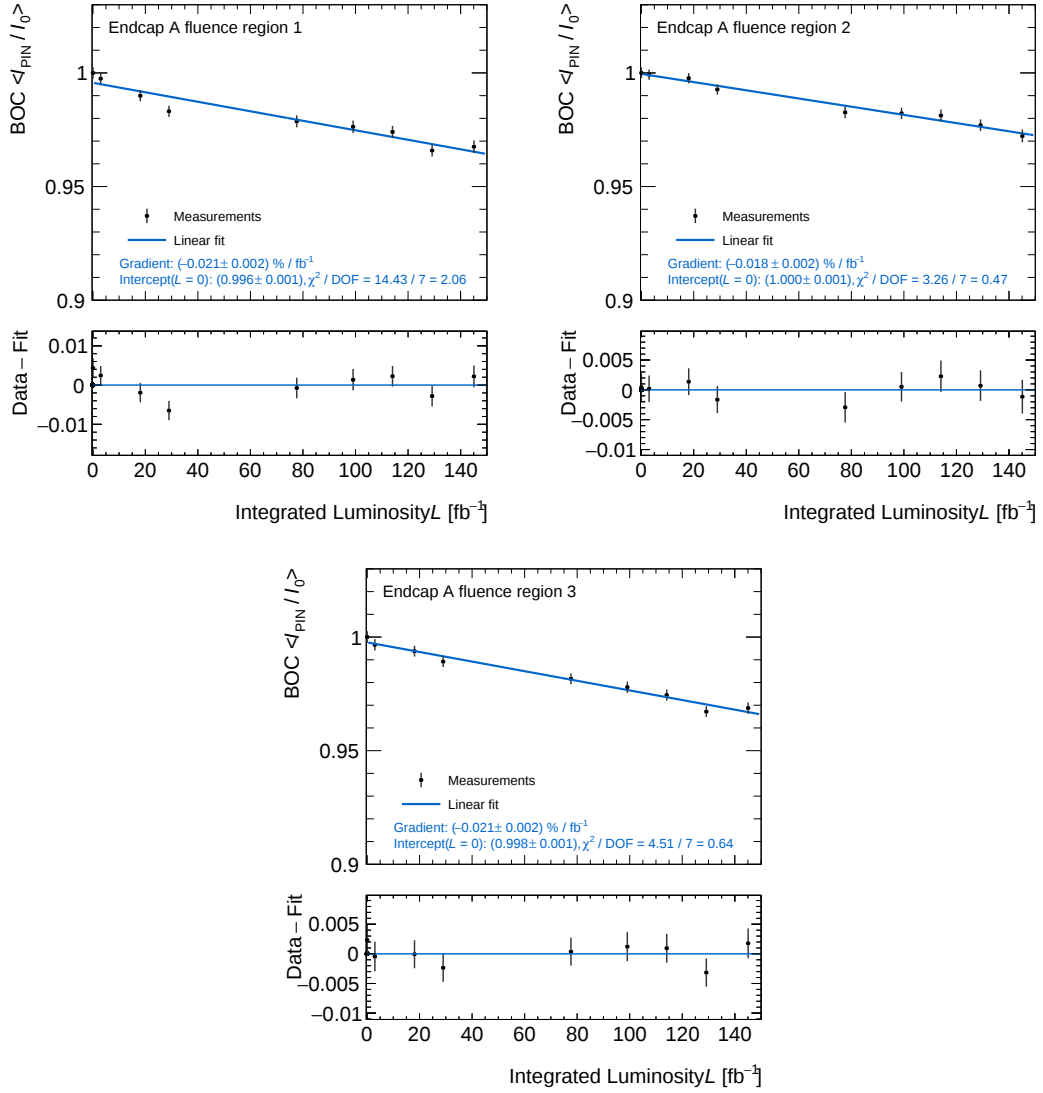


Figure 3.7: On-detector VCSELs: Endcap A. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018 .

On-detector *p-i-n* photodiodes

The central fitting function used was

$$f(L) = \frac{L}{L_0} - \frac{A}{B + L^2} \quad (3.8)$$

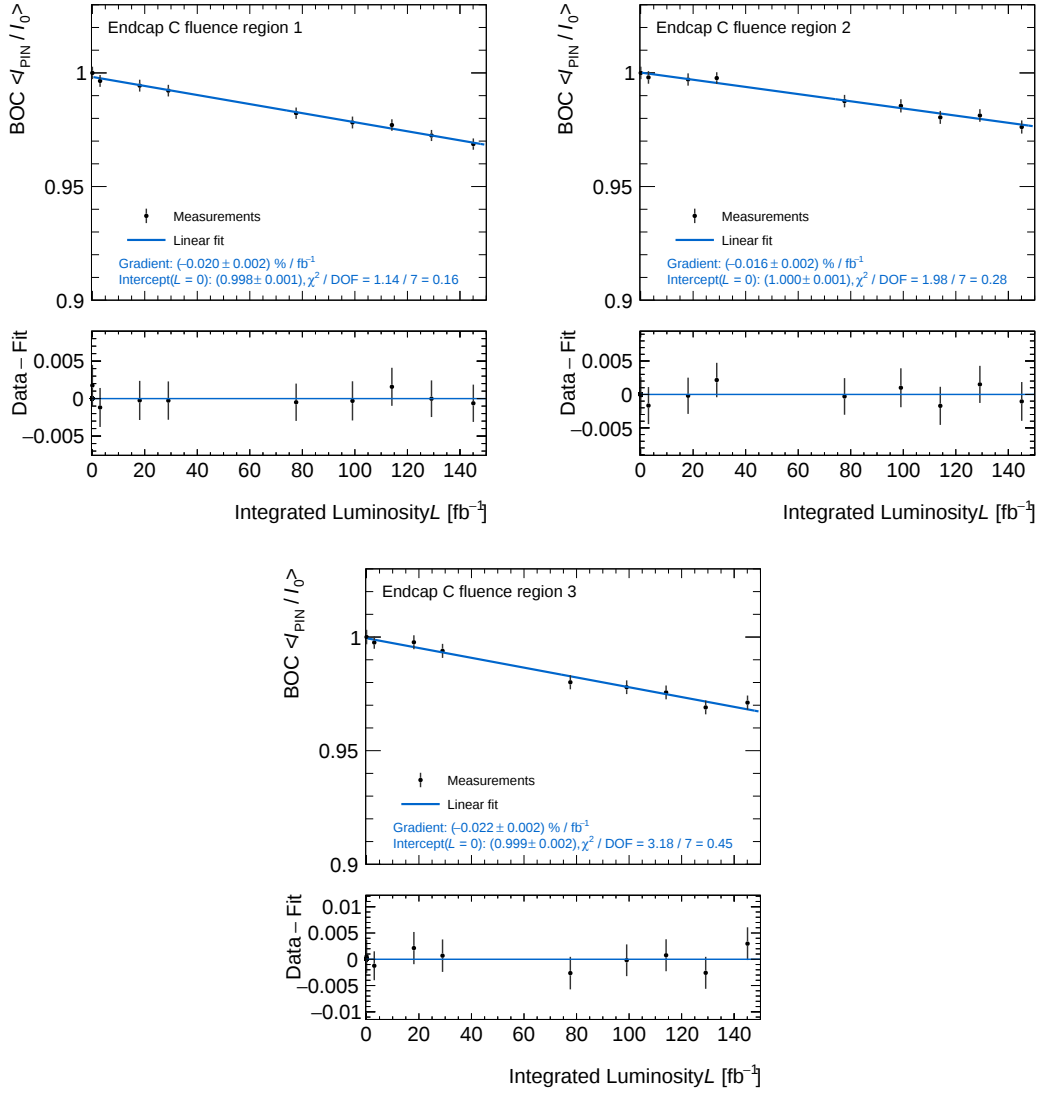


Figure 3.8: On-detector VCSELs: Endcap C. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018 .

and the additional functions used to calculate a parameterisation uncertainty were:

$$f_1(L) = \frac{L}{L_0} + \left(\frac{L}{L_1}\right)^{0.75} + \left(\frac{L}{L_2}\right)^{0.5} \quad (3.9)$$

$$f_2(L) = \left(\frac{L}{L_0}\right)^m$$

where all variables other than L are fitted. The resulting plots with the central fit superimposed are given in figs. 3.10 to 3.12. Table 3.7 summarises the central fit results and a summary plot is given in fig. 3.13.

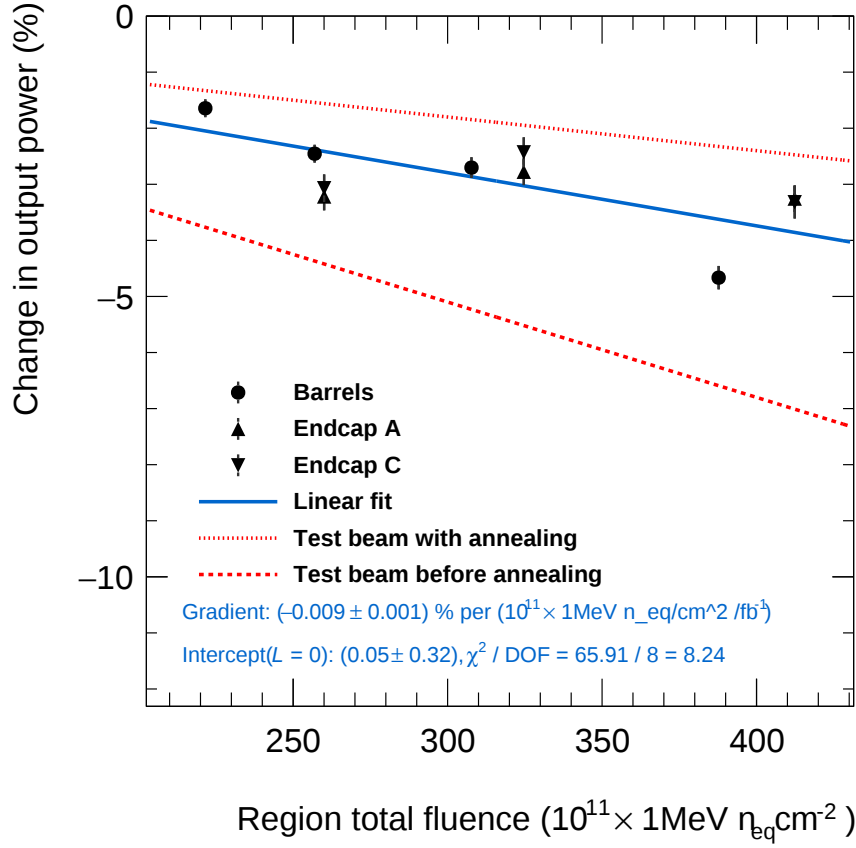


Figure 3.9: A plot to show the measured reduction in slope efficiency per fb^{-1} per SCT region, against the fluence per SCT region, fitted linearly.

The regions which have undergone the lowest fluence suffer from the largest uncertainties as the data do not yet represent the asymptotic behaviour of the degradation. In those regions for which the data have reached their asymptote, the parameterisation uncertainty represents the lack of knowledge of the degradation between the start of degradation and the recorded data.

It is not known how the parameterisation uncertainty is correlated between measurements; treating it as either fully correlated or fully uncorrelated yields results of 0.791 ± 0.001 (stat.) ± 0.059 (param.) and 0.796 ± 0.014 (stat. and param.), respectively. The uncertainty due to the stability is 0.083 for each region and this is one-to-one correlated between each measurement. With all of these uncertainties included, the results are consistent with the value of 71% measured by the previous studies of section 3.3.1.

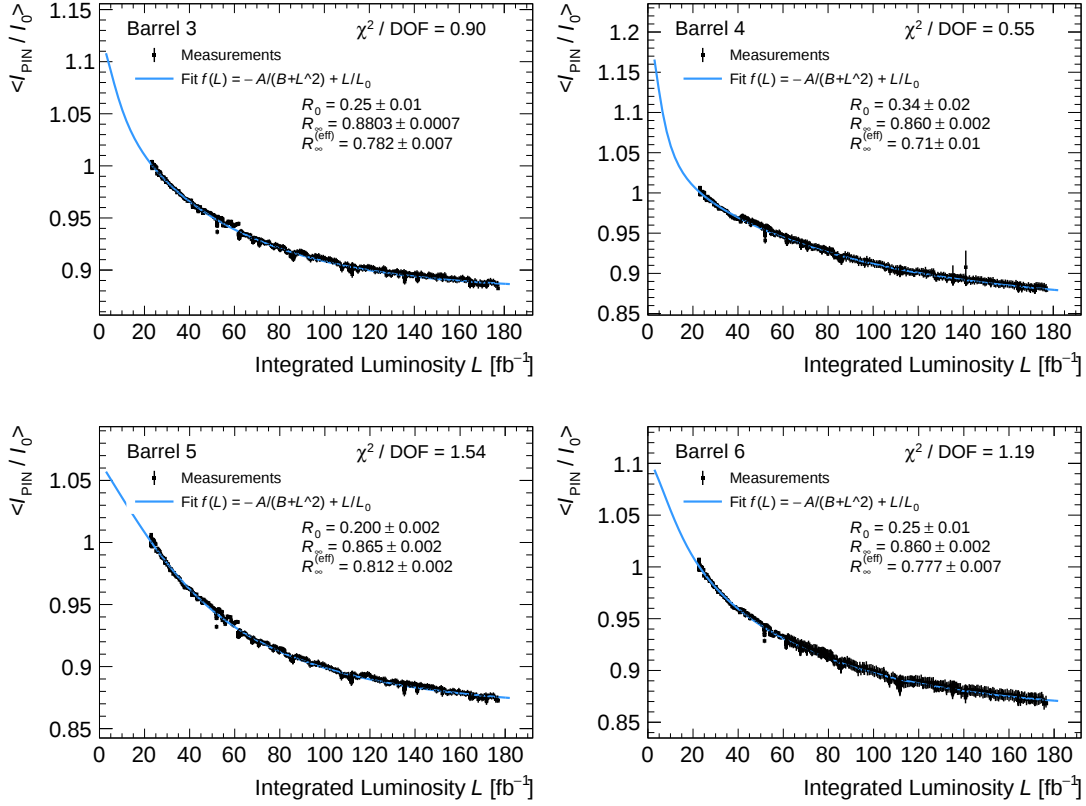


Figure 3.10: On-detector p - i - n diodes: Barrels. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018.

3.3.4 Conclusion

In conclusion, both sections of this analysis are in qualitative agreement with the test beam studies. The VCSELs undergo a linear decline in efficiency, and the p - i - n photodiodes undergo a sharp decline before reaching an asymptotic point of no further degradation.

During the test beam analysis, with one week of annealing at 10 mA, the VCSELs degraded at a rate of $(0.6 \pm 0.1)\%$ per $10^{13} \times 1\text{MeV}n_{eq}/\text{cm}^2$. On the other hand, the VCSELs in situ degraded at a rate of $(0.9 \pm 0.1)\%$ in the same units, despite an equivalent of 18 weeks annealing which should result in a lower overall rate. This inconsistency is not currently understood.

During the test-beam analysis the p - i - n photodiodes degraded to 71% of their

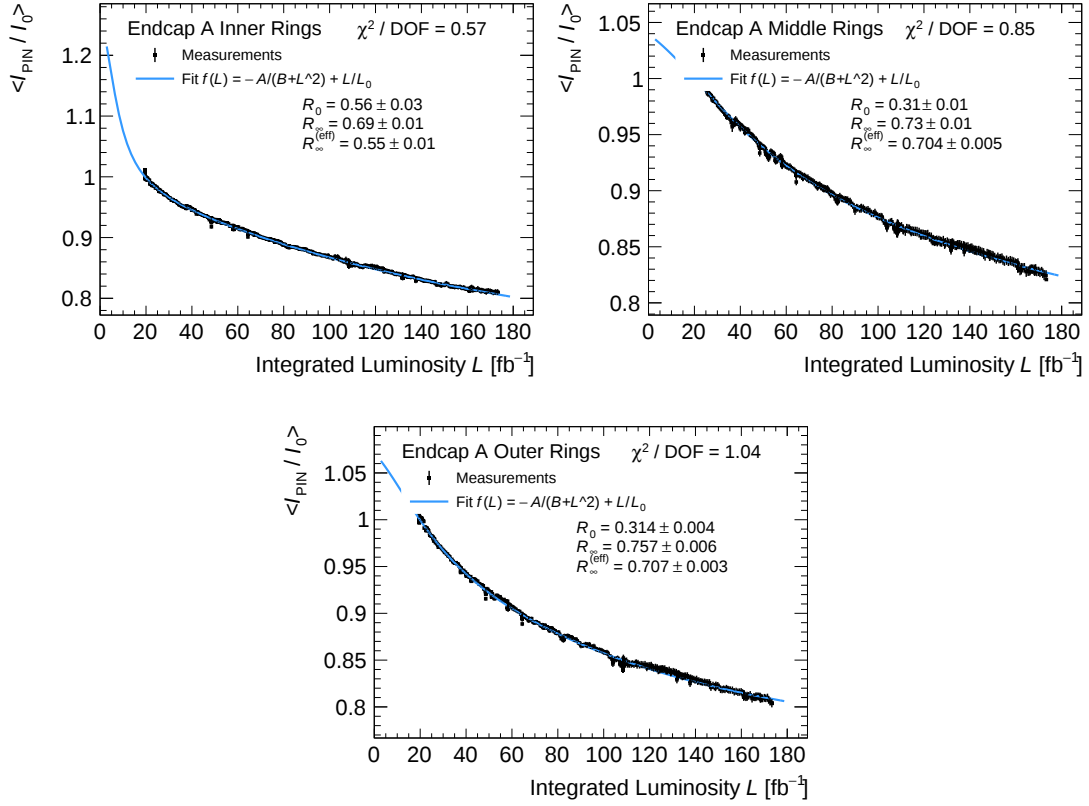


Figure 3.11: On-detector p - i - n diodes: Endcap A. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018.

previous efficiency, and in-situ degraded to an average of $(79 \pm 10)\%$, where the parameterisation uncertainty is conservatively treated as fully correlated. It is hypothesised that the cause of this degradation is charge trapping removing the slow signal from the undepleted n region of the photodiode. The largest source of uncertainty was the stability of the current over the many months of data-taking. It is estimated that, with either a lower level of instability or a higher rate of luminosity, this rate dependence can be investigated further in-situ.

Overall, with these results, it is expected that these components will survive the full run of the LHC in their current configuration. It is also expected that the annealing of the VCSELs can be further improved by running at 20mA as opposed to 10mA, should it be required.

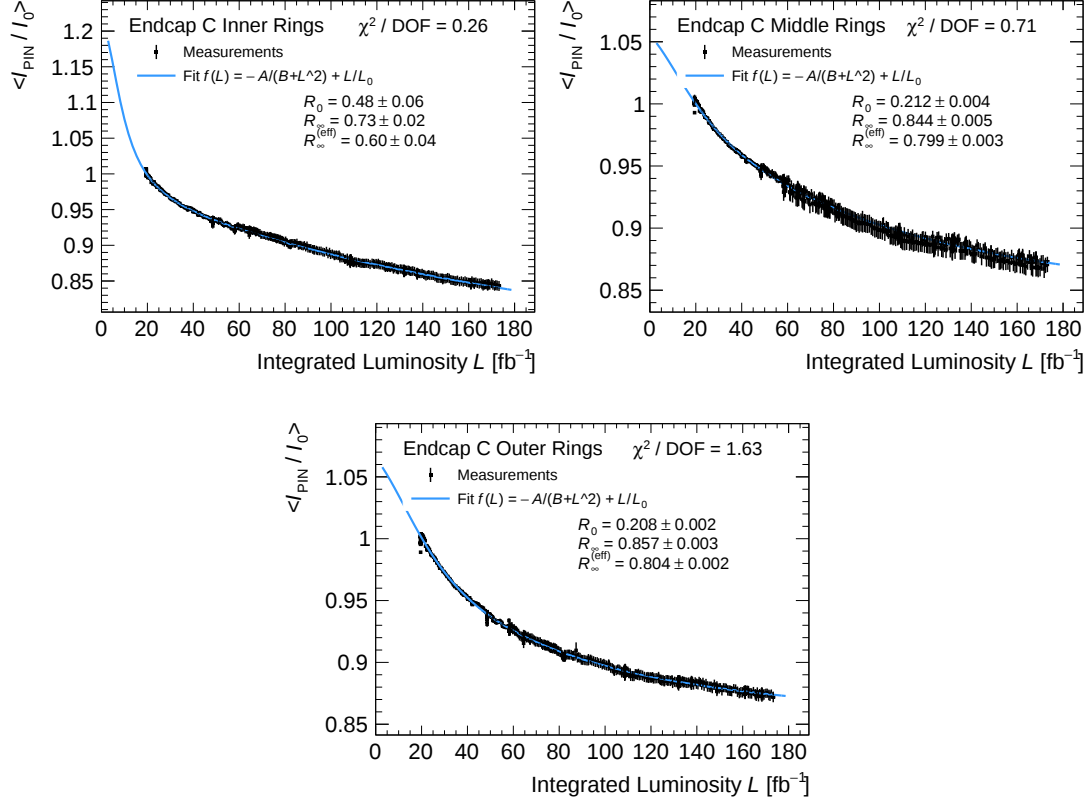


Figure 3.12: On-detector p - i - n diodes: Endcap C. The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC from 3 May 2016 to 30 October 2016, 5 May 2017 to 17 November 2017 and 17 April 2018 to 24 October 2018.

Table 3.7: Resulting fitted R_∞^{eff} values for the on-detector p - i - n diodes for the four barrel layers and three endcap regions using data from 2016 to 2018.

Region	R_∞^{eff}	Stat. unc.	Param. unc.
Barrel 3	0.782	0.007	0.046
Barrel 4	0.717	0.012	0.108
Barrel 5	0.812	0.002	0.017
Barrel 6	0.777	0.007	0.055
A side region 1	0.636	0.014	0.085
region 2	0.77	0.007	0.069
region 3	0.571	0.009	0.303
C side region 1	0.791	0.002	0.083
region 2	0.791	0.008	0.124
region 3	0.654	0.028	0.174

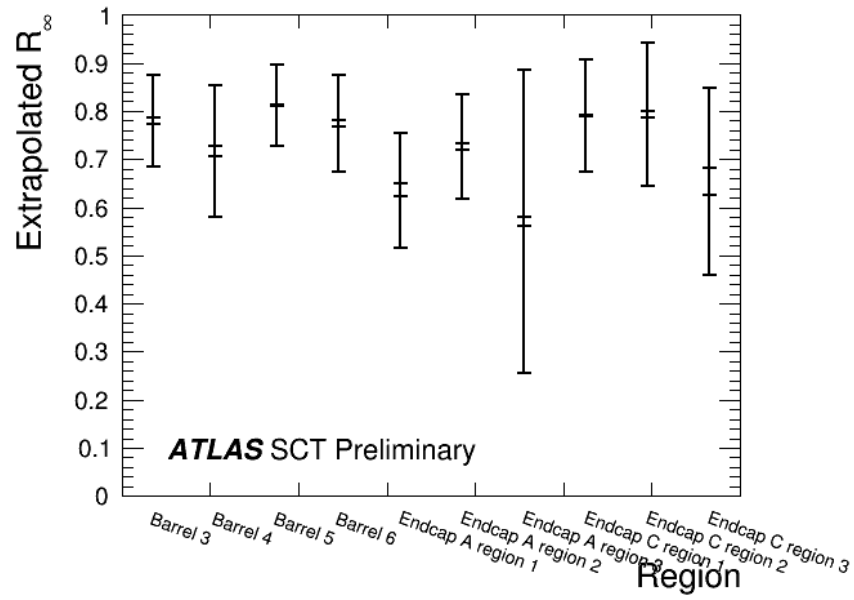


Figure 3.13: A plot to show the measured R_∞ values per SCT region, where the inner error band in the statistical uncertainty and the outer band is the sum in quadrature of the statistical and systematic uncertainties.

4

W + jets reconstruction and analysis strategy

Contents

4.1	Particle reconstruction	61
4.1.1	Electrons	61
4.1.2	Jets	66
4.1.3	E_T^{miss}	72
4.2	W + jets analysis strategy	73
4.2.1	Event selection	74
4.2.2	Distributions to be analysed	77
4.2.3	Monte Carlo simulations	79

The measurement of cross sections associated with the production of a W boson in association with jets can be summarised in the following way:

- Collect data associated with each collision at the ATLAS detector,
- Impose selection criteria on the data such that as many background events as possible are removed while maintaining high statistics of the W + jets signal process,
- Summarise the data in the form of histograms depicting the number of events collected within a number of ranges for a number of kinematic observables,

and compare to MC simulation of the SM prediction,

- Subtract the background contributions and correct for detector effects such as the acceptance rate and smearing due to finite measurement resolution.

In this chapter, the strategy of the first two bullet points are discussed in detail, while the latter two are discussed in chapters 5 and 6, with an overview of the uncertainties of the measurement discussed in chapter 7 and the results presented in chapter 8.

Particles produced in each pp collision are registered as electronic signals by the ATLAS detector in the various subcomponents discussed in section 3.2. These arrays of signals are each combined across the entire detector to *reconstruct* the event, including the objects within and their kinematic variables. Dedicated performance groups within the ATLAS collaboration attempt to identify objects and label them as electrons, muons, jets, etc.; these identification criteria are considered as a trade-off between maximising the selection efficiency and the purity of the resulting sample, performed individually for each label.

Subsequently, the properties of an object are measured such that they correspond to the initial particle. Losses in energy and momentum through interaction with the detector are compensated for through *calibration*, using a combination of simulation and data in-situ. Furthermore, uncertainties associated with these measurements are estimated and propagated through to the resulting analysis, as discussed in more detail in chapter 7. Once these measurements and their uncertainties are taken, they are managed centrally within the ATLAS software framework [65], ensuring consistent treatment of objects and their properties throughout the collaboration.

The reconstruction of objects associated with this analysis is discussed in section 4.1. Additionally, the strategy and selection criteria of the $W + \text{jets}$ analysis are discussed in section 4.2 and the simulations of the SM prediction to which the data will be compared are discussed in section 4.2.3.

4.1 Particle reconstruction

The reconstruction, identification and calibration, as well as classification into subtypes, for electrons [76, 77], jets [78–80] and E_T^{miss} [81] are discussed in the following sections 4.1.1 to 4.1.3, respectively.

4.1.1 Electrons

Electrons and their antimatter conjugate, positrons, are leptons, each with a mass of 511 keV and a negative and positive electrical charge, respectively. When produced in a collision at the ATLAS detector at sufficiently high transverse momentum, they leave a curved track in the ID and deposit energy in the ECal in the form of a relatively narrow EM shower. Along with the transition-radiation signals in the TRT, these components are used to identify and label electron/positron production.

A clean sample of electrons is required in order to understand the response of the detector to these objects and measure the efficiency of any selection criterion. This sample is typically obtained from the production of Z bosons and J/ψ mesons which then decay to an electron-positron pair. This is because it is a relatively common process at the LHC and the backgrounds are relatively low - for example, the rate at which an electron is faked in the detector (meaning that an object which is not an electron mimicks the signal of an electron, as discussed in more detail in chapter 5) is $\epsilon \ll 1$, and the rate at which two electrons are faked is ϵ^2 . Furthermore, looser criteria for electrons can be sampled by using the so-called “Tag and Probe” method, in which one “Tag” electron passing a high selection threshold is required in the event to reduce background contamination and the “Probe” electron is studied.

Reconstruction

Electrons lose a significant amount of their energy due to bremsstrahlung radiation resulting from interaction with the traversed material. The radiated photon can subsequently convert to an electron-positron pair, both of which can also interact with the material. These particles are usually collimated and are therefore reconstructed as part of a single electromagnetic cluster. This occurs throughout

the detector and beam pipe, meaning that it is possible to produce and match multiple tracks to the same cluster, all originating from the same primary electron.

The reconstruction of the primary electron candidate follows a three-step procedure:

- **Seed-cluster reconstruction:** As noted in section 3.2.2, the ECal is segmented into $\Delta\eta \times \Delta\phi = 0.003 \times 0.025$ cells. These are grouped into a grid of 200×256 elements of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ corresponding to the granularity of the second layer of the ECal. Subsequently, a *sliding window* [82] algorithm with a window of size 3×5 elements searches for clusters with transverse energy $E_T > 2.5$ GeV. If two seed-cluster candidates are found within an area of 5×9 units in $\Delta\eta \times \Delta\phi$, one is chosen such that it either has 10% higher E_T than the other candidate or contains the highest E_T central element, and the other is removed. The reconstruction efficiency of this algorithm ranges between 65% at $E_T = 4.5$ GeV and more than 99% above $E_T = 15$ GeV, based on simulation.
- **Track reconstruction:** In order to reconstruct a track in the ID, clusters are assembled from a number of hits within each sub-detector [83]. From these clusters, three-dimensional measurements in space are created and tracks are built using pattern recognition and fitting algorithms, allowing for up to 30% loss in energy due to interaction at each layer of silicon. The good-quality track (ie. at least one pixel hit and seven silicon strip hits) which most closely matches the seed cluster is chosen as the primary electron track.
- **Electron-candidate reconstruction:** The procedure is completed by matching the track candidate to the seed cluster and determining the overall cluster size. Reconstructed clusters are formed around the seed clusters using an extended window of size 3×7 in the region $|\eta| < 1.37$ or 5×5 in the region $1.52 < |\eta| < 2.47$, optimised to contain the energy deposition including

radiation from the electron candidate. Above $E_T = 15$ GeV, the electron-reconstruction efficiency through this whole procedure, with a track of good quality, ranges from 97% to 99%.

Identification

Identification is the process after reconstruction which attempts to further separate genuine *prompt* electrons, originating from the collision vertex, from other objects mimicking the signature, such as narrow hadronic jets, electrons from photon conversions and electrons from the semi-leptonic decay of heavy flavour hadrons, for example. In the analysis discussed in this thesis, a sophisticated *likelihood* (LH) criterion for identification is used alongside a more straightforward cut-based *isolation* criterion.

The electron LH is built from the products for signal, L_S , or background, L_B , of n probability density functions, each labelled P :

$$L_S(\vec{x}) = \prod_{i=1}^n P_{S,i}(x_i), \quad L_B(\vec{x}) = \prod_{i=1}^n P_{B,i}(x_i), \quad (4.1)$$

where $\vec{x}$ is the vector of quantities parameterising the probability density functions; examples include the track-cluster matching (ie. the difference in $\Delta\eta$ and $\Delta\phi$ between the track and cluster) and the transition radiation in the TRT. For each electron candidate, the discriminant is defined:

$$d_L = \frac{L_S}{L_S + L_B} \quad (4.2)$$

on which the subsequent LH identification is defined. Four fixed values of the discriminant are used to define operating points referred to as *VeryLoose*, *Loose*, *Medium* and *Tight* in increasing order of tightness of selection. The benefit of using a tighter criterion is that background contamination is reduced further, however there is a cost in electron identification efficiency to pay. The efficiency of the Tight criterion ranges from 55% at $E_T = 4.5$ GeV to 90% at $E_T > 100$ GeV, whereas the efficiency of the Loose criterion ranges from 85% at $E_T = 20$ GeV to 96% at $E_T > 100$ GeV. On the other hand, the rejection of background improves by a factor of five from the Loose to the Tight criteria.

The electron isolation is used in addition to remove electron candidates whose energy deposit is not sufficiently concentrated. Variables are constructed which quantify the level of activity within the local vicinity of the object, and cut-based criteria are used to remove background while maintaining a high acceptance of prompt electrons. A schematic of the quantification of the calorimeter isolation variable is given in fig. 4.1. The transverse energy of topological clusters within a particular radius are summed to give $E_T^{\text{isol,raw}}$ and the “core” energy of the electron is calculated using a rectangle of cells $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$. Furthermore, the leakage of the electron’s energy outside of the rectangle is estimated as E_T^{leakage} and the deposits due to pile-up in the detector are estimated as $E_T^{\text{pile-up}}$. The calorimeter isolation variable is given as

$$E_T^{\text{isol}} = E_T^{\text{isol,raw}} - E_T^{\text{core}} - E_T^{\text{leakage}} - E_T^{\text{pile-up}}. \quad (4.3)$$

The track-based isolation variable, p_T^{isol} is built in a similar way using transverse momenta, where p_T^{core} instead corresponds to tracks within a $\Delta\eta \times \Delta\phi = 0.05 \times 0.1$ window of the candidate track.

From these, multiple possible selection criteria can be built similarly to those defined for the LH discriminant, for example using $\Delta R = 0.2$:

- *FCLoose*: Fixed-Cut Loose criterion $E_T^{\text{isol}}/p_T < 0.20$ and $p_T^{\text{isol}}/p_T < 0.15$,
- *FCTight*: Fixed-Cut Tight criterion $E_T^{\text{isol}}/p_T < 0.06$ and $p_T^{\text{isol}}/p_T < 0.06$.

Calibration

The kinematics of an electron within a detector are usually mismeasured mostly due to energy loss. In order to ensure a good correspondence between the initial particle and the measurement, the signal is *calibrated*. This is conducted using the following five-step procedure with a combination of data and MC simulation:

- Estimate the energy from the deposit in the calorimeter using the properties of the shower and knowledge of the material between the collision vertex and the calorimeter deposit. This step uses a multivariate regression algorithm

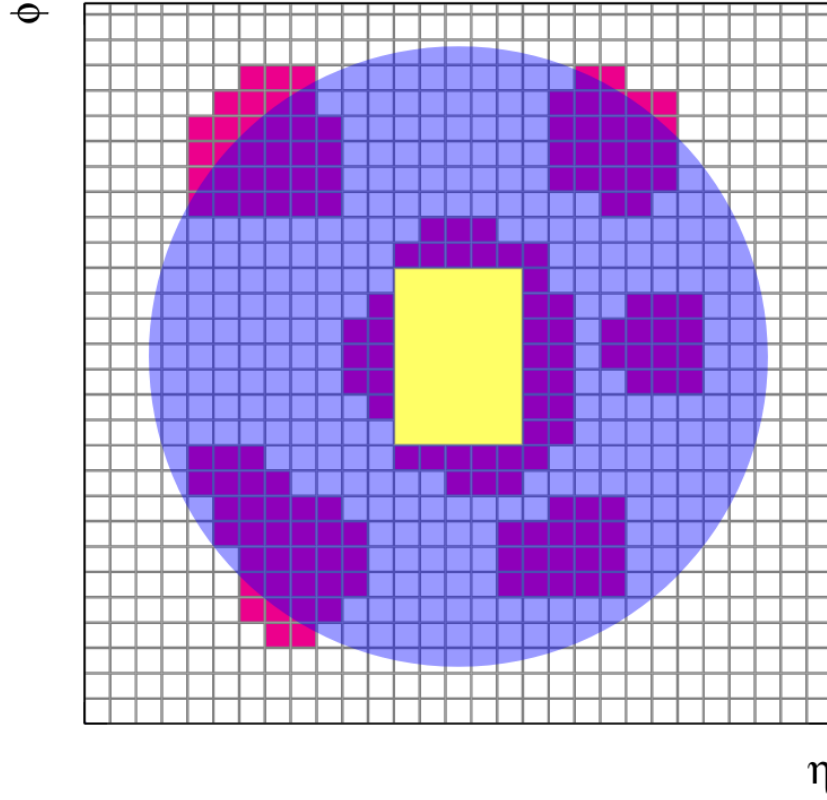


Figure 4.1: A schematic of the electron calorimeter-based isolation criteria. The grid represents the calorimeter cells of the second layer of the ECal and the circle represents the isolation cone. The topological clusters are shown in red, and the yellow rectangle represents the area used to calculate to core energy deposit of the candidate electron.

trained on samples of simulated events, but requires an accurate description of the detector.

- Adjust the relative energy scales of different layers of the ECal using studies of electron showers, correcting the data before the calculation of the energy of the electron.
- Correct for residual local non-uniformities in the calorimeter response, including geometric effects particularly prevalent at the boundaries between calorimeter cells. This is studied using Z -boson decays by comparing the calorimeter energy to the track momentum separately for electrons and positrons.

- Adjust the overall energy scale in the data similarly to the previous step using a large sample of Z -boson decays. At this stage, simulated samples are corrected to fit the data.
- Cross-check the results by comparing simulation and data using independent samples of data, for example $J/\psi \rightarrow ee$ for low-energy electrons.

4.1.2 Jets

A jet is the object traversing the ATLAS detector originating from a free quark or gluon in the final state of the collision process, as discussed briefly in section 2.3. A spray or collimated shower of hadrons is observed as signals in the detector; the main purpose of the jet reconstruction and calibration is to group these hadrons in to a series of jets, providing an estimation of the kinematics and characteristics of the initial particle in each case.

Reconstruction

The *calorimeter jets* are reconstructed with the anti- k_t algorithm [84] through the FASTJET software package [85] using the radius parameter $R = 0.4$. This algorithm is chosen as standard in the ATLAS collaboration as it exhibits experimentally and theoretically desirable properties, such as collinear and infrared safety as well as high computational performance.

The input to the anti- k_t algorithm are *topo-clusters* – topological groups of calorimeter cells of energy deposits where noise has been suppressed – which are first clustered in to local groups and then into jets. Cluster formation begins with the identification of *seed cells*, where the energy deposited is $E > 4\sigma$, and σ is the noise defined as the sum in quadrature of the expected electronic noise and pile-up noise. Neighbouring cells are grouped iteratively with the seed cell if their energy meets the condition $E > 2\sigma$. Finally, calorimeter cells adjacent to the group are included if $E > 0$. A splitting algorithm is implemented in addition to separate the resulting topo-clusters if there are multiple local maxima with energy greater than 500 MeV.

The formed topo-clusters are initially calibrated using the local calibration weighting (LCW) method. The mass of the topo-clusters are set to zero, classified as electromagnetic or hadronic, and then corrected for the different response efficiency of the calorimeter for pions and electrons in simulation. In addition, energy loss from deposits outside of the cluster and inactive areas of the detector are corrected for. The energy of each topo-cluster is equal to the sum of the contained calibrated calorimeter cell energies.

The anti- k_t algorithm groups topo-clusters iteratively into a single jet based on their distance until a stopping criterion is met. The distance between clusters i and j is defined as:

$$d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right) \times \frac{\Delta R_{ij}^2}{R^2} \quad (4.4)$$

where $p_{T,i}$ is the transverse momentum of cluster i , $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the squared spatial distance between the clusters in rapidity-azimuthal angle space, and R is the aforementioned radius parameter set to 0.4 for the purposes of the analysis in this thesis. While this distance is smaller than what is defined as the *beam distance*, $d_{iB} = \frac{1}{p_{T,i}^2}$, the two objects are combined into a single object, which is reinserted to the next iteration of the algorithm. When no two objects meet this condition, the algorithm stops and the set of objects is returned and declared as jets.

The physical interpretation of the anti- k_t algorithm is that the most energetic clusters collect the softer clusters as long as they are within the radius parameter R . This is the most-used method for jet reconstruction in the ATLAS and CMS collaborations and is implemented natively to the ATLAS software framework. The resulting jets are almost circular in $y - \phi$ space [86], especially compared to alternative combination algorithms such as the k_t [87–89] and Cambridge/Aachen [90] algorithms, and shown in fig. 4.2. Jets in Monte Carlo (MC) simulation are reconstructed in the same way as data using these algorithms. Truth jets used in the unfolding procedure described in subsequent sections are defined similarly,

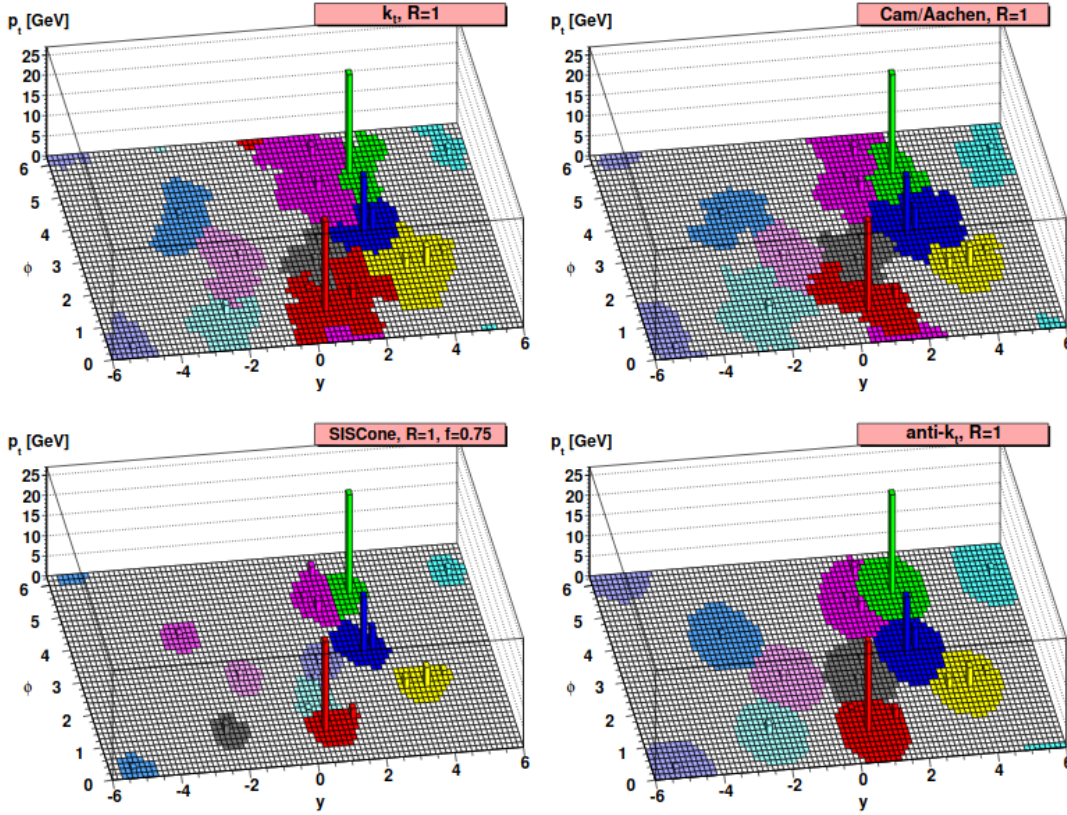


Figure 4.2: A sample event with topological clusters grouped into jets using a range of algorithms, including the anti- k_t algorithm used in this analysis [86].

however the stable¹ truth particles in the final state are used instead of topo-clusters as the input to the anti- k_t algorithm.

Calibration

Although the reconstruction of jets is designed to suppress the effects of detector pile-up and other nuisances such as inactive cells, these effects along with some others still result in a difference between observation and simulation which are corrected for through further calibration. Two variables are measured and calibrated individually: the *jet-energy scale* (JES) and the *jet-energy resolution* (JER), corresponding to the mean and the standard deviation of the jet response distribution, respectively.

The JES calibration proceeds in the following way:

¹Here, “stable” refers to particles which do not decay before they pass through the ATLAS detector.

- **Origin correction:** In the reconstruction of jets, it is assumed that all objects forming the topo-clusters originate from the geometrical centre of the ATLAS detector. In this step, the four-momentum of the jets is recalculated such that they originate from the hard-scatter primary vertex, while maintaining the jet energy. The procedure improves the resolution in η from 0.06 to 0.045 at a jet p_T of 20 GeV and from 0.03 to 0.006 above 200 GeV [91].
- **Jet area-based pile-up correction:** Pile-up can provide excess energy to the jets as particles from other collisions contribute. This correction is formed from two components. Firstly, an area-based method is used to subtract the per-event pile-up contribution to the p_T of each jet parameterised by the area of the jet. The contribution is calculated from the median p_T density of the jets, $\rho = p_T/A$ where A is the area of the jet. In this case only, the k_t clustering algorithm is used as it clusters soft-energy deposits first, thereby capturing pileup in the event. Additional residual pile-up is captured using the established linear dependence on the number of primary vertices, N_{PV} , and the average number of interactions per bunch crossing, μ , such that:

$$p_T^{\text{corr}} = p_T^{\text{reco}} - \rho \times A - \alpha \times (N_{PV} - 1) - \beta \times \mu \quad (4.5)$$

where the linear coefficients α and β are determined from MC simulation and found to be independent of one another.

- **Absolute MC-based calibration:** The absolute JES calibration uses MC simulation to derive a function $f_{\text{calib}}(E_{\text{LCW}}^{\text{jet}}, \eta)$ which corrects the four-momentum of the reconstructed jet to the particle-level energy scale:

$$E_{\text{LCW}+\text{JES}}^{\text{jet}} = \frac{E_{\text{LCW}}^{\text{jet}}}{f_{\text{calib}}(E_{\text{LCW}}^{\text{jet}}, \eta)} \quad (4.6)$$

The dependence of f on η allows this step to simultaneously correct for the η -dependence of the reconstructed JES.

- **Global sequential calibration:** This step of the calibration aims to reduce flavour dependence and energy leakage effects, particularly prominent in high- p_T jets not fully contained in the calorimeter. Dependencies of the JES on longitudinal and transverse features must be observed and accounted for in order to provide an accurate representation. The most notable difference in particle composition and shower shape is between quark- and gluon-initiated jets, where quark-initiated jets will exhibit a relatively higher fraction of the p_T from just a few hadrons, which themselves are more likely to penetrate further into the calorimeter and beyond, whereas gluon-initiated jets contain more particles of softer p_T spread over a wider area. These dependencies are parameterised by five variables in total, including two variables which make use of the tracking system and one which makes use of tracks in the muon spectrometer as a proxy for particles punching beyond the calorimeter. More details are given in Ref. [78].
- **In-situ JES calibration:** The final stage of JES calibration is to account for differences in the jet response between data and MC simulation. Reference particles which balance the jet p_T are used to directly measure the JES in both data and MC *in-situ*. An example of a reference particle could be a Z boson, a photon, or even a second jet, employed for intercalibration from the well-measured central region to forward pseudorapidities. The double ratio

$$\frac{R_{\text{data}}}{R_{\text{MC}}} = \frac{\langle p_T^{\text{jet}}/p_T^{\text{ref}} \rangle_{\text{data}}}{\langle p_T^{\text{jet}}/p_T^{\text{ref}} \rangle_{\text{MC}}} \quad (4.7)$$

is measured, defining the correction to jet four-momentum applied to data. The correction is calculated as a function of both jet p_T and η for the η -intercalibration.

The JER is the precision of the jet energy measurement, and can be parameterised similarly to any signal within the calorimeter:

$$\frac{\sigma(p_T)}{p_T} = \frac{N}{p_T} \oplus \frac{S}{\sqrt{p_T}} \oplus C \quad (4.8)$$

where N is the noise term denoting the effect of pile-up and noise within the electronics, S is the stochastic term and C is the “constant” term independent of p_T due to interaction with the passive material of the calorimeter. The noise term is measured directly in MC simulation, where the noise from the electronics is measured separately in events with no pile-up. The stochastic and constant terms are measured using direct-balance techniques similar to those described for the JES measurement, for example using a second jet as a balance.

Pile-up suppression - the Jet-Vertex Tagger

The jet reconstruction and calibration goes some way to suppress pile-up jets, however further suppression has been a crucial component of many analyses of the ATLAS detector. The Jet-Vertex Tagger (JVT) [92] is a multivariate construction on which a selection criterion can be placed to suppress pile-up jets in the event.

The Jet-Vertex Fraction (JVF) is a variable defined as the ratio between the scalar p_T sum of the tracks associated to the jet originating from the interaction primary vertex and the scalar p_T sum of all associated tracks. It was shown that pile-up jets can be suppressed effectively by imposing a selection criterion on this variable [93], however the resulting efficiency of this selection is dependent on the number of reconstructed primary vertices (N_{Vtx}) in the detector. This was used quite effectively by itself for a number of Run 1 ATLAS analyses. The JVT uses two features similar to JVF:

$$\text{corrJVF} = \frac{p_T^{\text{HS}}}{p_T^{\text{HS}} + p_T^{\text{PU,corr}}}, \quad (4.9)$$

where p_T^{HS} is the usual numerator of the JVF definition and $p_T^{\text{PU,corr}}$ is a measure of the pile-up p_T of the jet relative to average pile-up activity in the event; and

$$R_{p_T} = \frac{p_T^{\text{HS}}}{p_T^{\text{jet}}}. \quad (4.10)$$

The efficiency of the JVT algorithm is approximately 90% for jets $20 < p_T < 30$ GeV independent of the number of primary vertices, whereas the JVF selection varies between 95% at $N_{\text{Vtx}} = 6$ and 75% at $N_{\text{Vtx}} = 30$. The significant increase

in pile-up between Run 1 and Run 2, where the mean number of interactions per bunch crossing is $\langle\mu\rangle = 20.7$ and 33.7 , respectively, makes this algorithm a necessity for Run 2 analyses such as the one reported in this thesis.

4.1.3 E_T^{miss}

Neutrinos are weakly interacting particles, and therefore do not induce a signal in any component of the ATLAS detector. Instead, their production is inferred from the law of conservation of momentum: if there is no momentum transverse to the beams initially, then the total momentum transverse to the beams in the final state must also sum to zero. Any “missing” transverse momentum can therefore be assumed to be carried by one or more particles which have not been detected.

The missing transverse momentum is therefore defined as the negative vectorial sum of the transverse momenta of all particles in a collision event:

$$\vec{E}_T^{\text{miss}} = - \sum_{i \in \text{hardobjects}} \vec{p}_{T,i} - \sum_{i \in \text{softsignals}} \vec{p}_{T,i} \quad (4.11)$$

where the objects have been separated in to two categories: the “hard” objects, referring to reconstructed and calibrated objects such as electrons and jets, and the “soft” objects, referring to unidentified but reconstructed charge-particle tracks associated with the collision vertex.

The scalar E_T^{miss} refers to the magnitude of $\vec{E}_T^{\text{miss}}$, and the azimuthal angle is defined:

$$\phi^{\text{miss}} = \arctan\left(\frac{E_{T,y}^{\text{miss}}}{E_{T,x}^{\text{miss}}}\right) \quad (4.12)$$

The calculation of E_T^{miss} depends intimately on the calibration of the hard objects within the event - for example, criteria imposed on the electron could determine whether the transverse momentum of an object is calibrated as an electron or treated instead as part of the soft term. For the calculation used in this analysis, objects are classified based on the criteria in table 4.1. E_T^{miss} is recalculated for different definitions of electron identification criteria.

Object	Criterion
e	$p_T > 10 \text{ GeV}$
jet	$(\eta < 4.5 \text{ and } p_T > 60 \text{ GeV}) \text{ or } (2.4 < \eta < 4.5 \text{ and } p_T > 20 \text{ GeV}) \text{ or } (\eta < 2.4 \text{ and } p_T > 20 \text{ GeV and JVT} > 0.59)$
soft	$p_T > 400 \text{ MeV and } d_0 < 1.5 \text{ mm and } z_0 \sin(\theta) < 1.5 \text{ mm and } \Delta R(\text{track, cluster}) > 0.05$

Table 4.1: Overview of the criteria for the objects in the E_T^{miss} calculation of the $W + \text{jets}$ analysis presented in this thesis, where JVT is defined in section 4.1.2, d_0 refers to the transverse impact parameter (the minimum transverse distance between the track and the collision point) and z_0 refers to the longitudinal impact parameter.

4.2 $W + \text{jets}$ analysis strategy

The aim of this analysis is to take measurements of the $W + \text{jets}$ cross section at a proton-proton collider using the ATLAS detector. Differential distributions which have been identified to be sensitive to proton PDFs or other features of QCD are presented. The analysis is executed for the charge-independent production of the W boson, as well as the W^+ and W^- cross sections, the difference between which providing additional insights to the structure of the proton. Furthermore, the distributions are measured individually for a range of inclusive jet multiplicities², highlighting the strengths of the various current SM predictions at low jet multiplicities and up to which number of jets these break down.

The W boson is identified via the leptonic $W \rightarrow e\nu$ decay, ie. the selection criteria at its most basic is a single electron, the presence of E_T^{miss} signifying the neutrino, and at least one jet. The charge of the boson is identified by the charge of the electron. Events passing the selection criteria detailed below are used to fill histograms of the aforementioned differential distributions, where the binning of the histogram is selected such that each bin is as narrow as possible for a high-precision measurement with detailed structure, but wide enough to maintain a high statistical precision and reduce the migration of events between bins as

²Here, the term “inclusive” refers to events containing either that number of jets or more, for example. inclusive $N_{\text{jets}} = 3$ means events in which there were three or more jets, whereas exclusive $N_{\text{jets}} = 3$ means events in which there were three jets only.

discussed in chapter 6. These distributions are compared to the *total SM prediction*, which is the sum of the expected $W + \text{jets}$ signal and the various backgrounds discussed in detail in chapter 5, in order to establish the accuracy of - or highlight potential issues in - the modelling.

Subsequently, the histograms are scaled by bin width in order to give the number of events per unit of the horizontal axis, and *unfolded*. In this process, described in more detail in chapter 6, the particle-level distributions are inferred from the detector-reconstructed distributions. This way, theoretical predictions can be compared directly to the measurement without the need to model the detector.

4.2.1 Event selection

The selection criteria with short descriptions for this analysis is shown in table 4.2, and more details are given throughout this subsection.

The preselection is based on recommendations provided by combined-performance groups in the ATLAS collaboration. In the years 2015-16, as shown in fig. 4.4, approximately 42.7 fb^{-1} were delivered by the LHC. Inefficiencies in the ATLAS data-acquisition system, along with the time “lost” while ATLAS moves to full operation, result in a small loss in the total integrated luminosity recorded down to approximately 39.5 fb^{-1} . Finally, issues during data-taking can occur, such as the malfunctioning of a component of the detector, such that a proportion of the data are not “Good for Physics”. Meta-data defining whether the recorded block of luminosity is Good for Physics are stored in “Good Run Lists” [94] by the ATLAS collaboration - the total integrated luminosity declared as such being approximately 36.2 fb^{-1} . Only data which are within these Good Run Lists is used for the $W + \text{jets}$ analysis.

Additional restrictions on the Good for Physics data are imposed such that the events are of good quality. Firstly, events for which the data are corrupted, for example due to incomplete readout or errors within, are rejected. Additionally, the collision point (referred to as the *primary vertex*) must be well-measured, as discussed in detail in Refs. [95, 96]. Tracks with $p_T > 400 \text{ MeV}$, $|\eta| < 2.5$,

more than nine hits in the ID if $|\eta| < 1.65$ or more than eleven hits in the ID if $|\eta| > 1.65$, at least one hit in the first two pixel layers, no pixel holes and at most one SCT hole³ can be used to reconstruct a primary vertex, and at least three tracks associated to a primary vertex are required.

The triggers [97, 98] used for the analysis are the following:

- **Year 2015:** Events must pass `e24_lhmedium_L1EM20VH` or `e60_lhmedium` or `e120_lhloose`.
- **Year 2016:** Events must pass `e26_lhtight_nod0_ivarloose` or `e60_lhmedium_nod0` or `e140_lhloose_nod0`.

The interpretation of these labels is made in the following way:

- The “e” label at the start reads that this trigger requires a single electron. The associated number is the required p_T .
- The next label, for example “lhtight”, defines the requirement on the quality of the electron. All of these labels are associated with the likelihood criterion discussed in more detail in section 4.1.1.
- The label “L1EM20VH” indicates that events in which there was deposit in the hadronic calorimeter behind the electron of $E_T > 20$ GeV are removed.
- Finally, in the 2016 triggers, the label “nod0” indicates that the d_0 impact parameter was not used, and the label “ivarloose” indicates the isolation requirement that the measured p_T within a cone radius $R = 0.2$ centred on the electron (excluding the electron itself) is less than 10% of the p_T of the electron.

Electrons are chosen such that they pass the criterion $p_T^e > 35$ GeV and $|\eta^e| < 2.47$. This p_T threshold is chosen as it is the point at which the ratio of the trigger efficiency in data and MC begins to plateau, as shown in fig. 4.3, therefore reducing the size of the associated systematic uncertainties. In addition to these

³Here, a “hole” refers to a measurement which is expected but not observed.

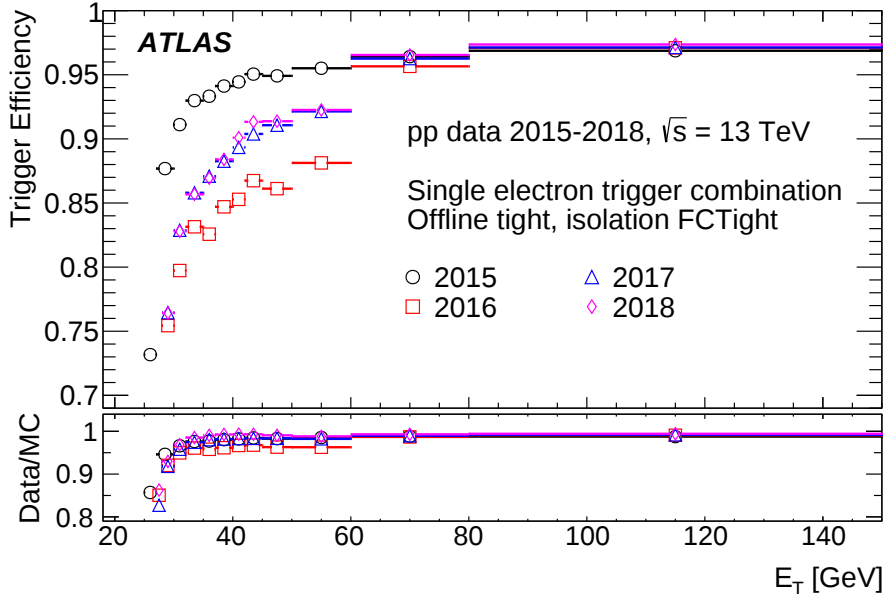


Figure 4.3: The efficiency of the single-electron trigger with the FCTight isolation criterion shown as a function of the offline electron E_T during Run 2 [97].

criteria, electrons within the transition region between the barrel and endcaps of the ECal, $1.37 < |\eta^e| < 1.52$, are rejected. Where possible, for example in the electron isolation and likelihood criteria, the tightest working point is chosen in order to reduce systematic uncertainties – the cost of a lower selection efficiency is worth the gain as the analysis is not statistically limited. Events with a second electron $p_T > 25$ GeV are rejected as this is not expected in $W + \text{jets}$ events but is present in Z -boson decays.

Additional criteria for background suppression include a selection on $E_T^{\text{miss}} > 25$ GeV. This is an optimal selection as leptonic W decays produce a neutrino, but also because the multijet and Z -boson processes, which form significant components of the backgrounds to this analysis as discussed in chapter 5, typically result in low E_T^{miss} . Similarly, the criterion $M_T^W + E_T^{\text{miss}} > 60$ GeV where

$$M_T^W = \sqrt{2p_T^e E_T^{\text{miss}} (1 - \cos(\Delta\phi(e, E_T^{\text{miss}})))} \quad (4.13)$$

is the W -boson *transverse mass*, provides a further suppression of the multijet background and is discussed in more detail in chapter 5.

Finally, a selection on the jets is applied. At least one jet must be present in the event, and it must be sufficiently distinct from the electron such that

	Criterion	Description
Preselection	Good Run List	
	No data corruption	
Leptons	$p_T^e > 35 \text{ GeV}$	Reduce syst. uncs.
	$ \eta^e < 2.47$ (excl. transition)	Detector acceptance
	Tight LH	Reduce misidentification
	FCTight isolation	Reduce misidentification
	No second e : $p_T > 25 \text{ GeV}$, $ \eta < 2.47$	Suppress $Z \rightarrow ee$
	No muon: $p_T > 25 \text{ GeV}$, $ \eta < 2.47$	Suppress backgrounds
W -boson	$M_T^W + E_T^{\text{miss}} > 60 \text{ GeV}$ and $E_T^{\text{miss}} > 25 \text{ GeV}$	Suppress multijet
Jets	At least one	Signal criterion
	JVT > 0.59	Suppress pile-up jets
	$\min(\Delta R(e, \text{jet})) > 0.4$	Remove electrons in jets
	$p_T > 30 \text{ GeV}$	Reduce systematic uncs.
	b -tagging veto	Reduce top background

Table 4.2: A table detailing the event selection criteria of the $W + \text{jets}$ analysis.

$\min(\Delta R(e, \text{jet})) > 0.4$. Additionally, pile-up jets are suppressed with the criterion that $\text{JVT} > 0.59$, as discussed in section 4.1.2, and the p_T of each must be greater than 20 GeV. Events containing jets failing the “loose” criterion are rejected as they distort the measurement of E_T^{miss} . Finally, the ATLAS collaboration implements multivariate analyses to “ b -tag” jets which are likely to have originated from a b quark [79]. As the $W + \text{jets}$ process is very unlikely to produce a b -jet, whereas the top-induced background processes are expected to contain at least one, events containing jets which are b -tagged are rejected.

4.2.2 Distributions to be analysed

In addition to the jet multiplicity, N_{jet} , three spectra (which were analysed in the similar 8 TeV analysis [100]) are evaluated differentially as they are sensitive to higher-order terms and PDFs:

- The transverse momentum of the W boson, p_T^W , evaluated on an event-by-event basis as $p_T^W = \sqrt{(p_T^e)^2 + (E_T^{\text{miss}})^2 + 2p_T^e E_T^{\text{miss}} \cos(\Delta\phi(e, E_T^{\text{miss}}))}$,
- The scalar sum of the transverse momenta of the electron, jets and E_T^{miss} (H_T),

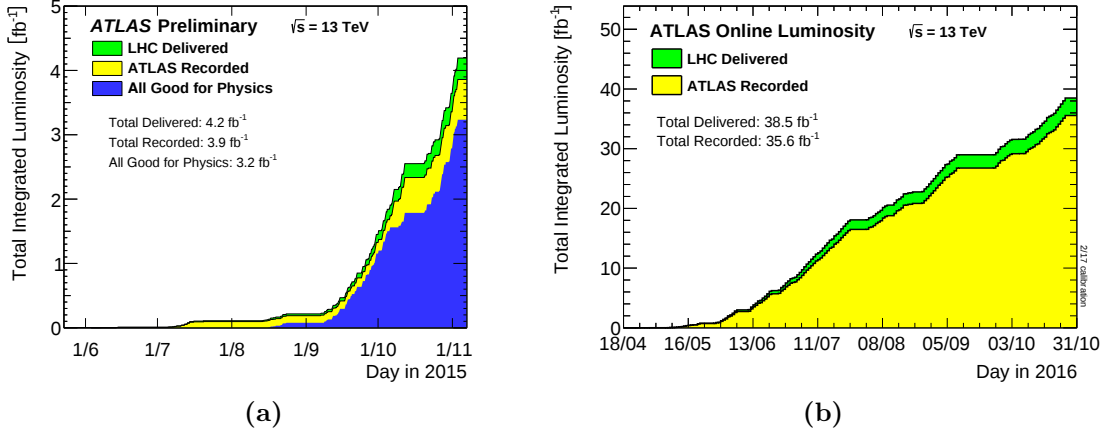


Figure 4.4: The integrated luminosity collected by the ATLAS detector in the years 2015 (a) and 2016 (b). The LHC-delivered integrated luminosity and that collected are displayed separately [99].

- The transverse momentum of the most energetic (referred to as “leading”) jet (p_T^{leading}).

Furthermore, $W + \text{jet}$ measurements are sensitive to the *naïve T -odd asymmetry*, described in detail at Ref. [101]. This is a so-far unobserved effect of corrections to the cross section at one- and higher-loop levels expected in the current formulation of the SM, as shown in fig. 4.5. The asymmetry manifests through a kinematic variable referred to in this thesis as the T -odd x :

$$x_{T\text{-odd}} = p_l^\perp / (m_W/2) \quad \text{where} \quad p_l^\perp = \frac{\vec{p}_{p_1} \times \vec{p}_T^W \cdot \vec{p}_e}{|\vec{p}_{p_1} \times \vec{p}_T^W|}, \quad (4.14)$$

where $\vec{p}_{p_1}$ is the momentum of the proton moving in the positive z direction as defined in chapter 3. This can be interpreted as the component of the electron transverse momentum which is perpendicular to the transverse momentum of the W boson. The differential cross section of this observable is measured in addition to those defined above, in preparation for a more complete analysis at a later date.

4.2.3 Monte Carlo simulations

The SM predictions of measurements in particle physics are often produced using Monte Carlo (MC) event generators. These generators can be used firstly to calculate cross sections and expected distributions of observable quantities, but also

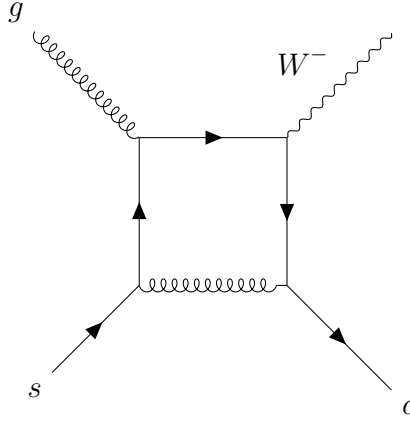


Figure 4.5: An example of a one-loop diagram through which the naïve T-odd asymmetry manifests in the $W + \text{jets}$ cross section.

to understand the relationships between these quantities and the effect of alternative selection criteria. Further, when interfaced to a model of the ATLAS detector, the response of the detector itself can be modelled, including any mismeasurements.

In this analysis, MC simulations are used for a variety of purposes discussed throughout chapters 5 to 7. In this section, the generators used are summarised.

The $W \rightarrow e\nu + \text{jets}$ process is simulated using two generators, which are contrasted throughout the analysis. Each are described in detail in Ref. [102].

- The first is MADGRAPH5_aMC@NLO (v2.3.2.p1) [45] interfaced to PYTHIA 8 (v8.210) [103], where the former provides predictions of event kinematics at LO for up to four jets from the matrix element and the latter is used for the parton showering, hadronisation and underlying event, based on the “A14” (ATLAS 2014) set of “tuned” generator parameters [104]. This simulation uses the NNPDF2.3 NLO PDFs for the hard scattering and the NNPDF2.3 LO PDFs for the parton shower [105].
- The second is SHERPA (v2.2.1) [106, 107], which provides the full event simulation for up to two jets at NLO and four jets at LO. A dedicated set of tuned generator parameters is developed by the SHERPA authors for this version. This simulation uses the NNPDF3.0 NNLO PDFs [108].

In later sections, including plots and figures, these simulated data are referred to as “MG + PY8” and “SHERPA2.2.1” for conciseness.

Various aspects of the detector response are sufficiently difficult to model that the ATLAS collaboration calibrates the MC simulation by comparison between the prediction and data. In this analysis, this includes firstly the effect of pileup and secondly the effect of non-trivial selection criteria, such as the JVT cut and the electron trigger, reconstruction efficiency, identification and isolation. For each of these, a “scale factor”, defined as the ratio of efficiencies $SF = \epsilon_{\text{data}}/\epsilon_{\text{MC}}$, is calculated as a function of the transverse momentum and rapidity of the object, and each event is reweighted according to this factor [109, 110].

Finally, a “luminosity weight” is applied to the MC prediction such that its overall normalisation is representative of the data set to which it is compared:

$$w_{\text{Lumi}} = \frac{L \times \sigma}{\sum w}, \quad (4.15)$$

where L is the integrated luminosity of the data set, σ is the NNLO cross section of the modelled process, and the sum in the denominator is over the overall weight of each of the simulated events.

5

Background estimation

Contents

5.1	The estimation of fake and non-prompt leptons	84
5.1.1	Sources of fake and non-prompt electrons	84
5.1.2	The Matrix Method	85
5.1.3	The measurement of the fake efficiency	86
5.1.4	Results and discussion	91

There are processes which can mimic the $W + \text{jets}$ signature forming what are referred to as “background” contributions to the observed data. This is either because a process results in the same signature or there is a mismeasurement. These backgrounds are reduced through the selection criteria discussed in section 4.2.1, and that which remains is estimated through a combination of MC simulation and “data-driven” methods. The estimated background can then be subtracted, leaving the reconstructed $W + \text{jets}$ cross-section.

The following processes contribute to the background:

- **Multijet production:** Events containing jets only can pass the selection criteria if one jet mimics the signature of the lepton, referred to as *faking* the lepton. This mismeasurement can furthermore give rise to fake E_T^{miss} . This background is reduced through the tight trigger and isolation requirements,

as well as further selection on E_T^{miss} and M_T^W . Although the probabilities of a fake lepton passing the selection criteria are small, the large cross section of multijet production yields a significant contribution to the background. This is the most dominant background at low jet multiplicities. It is difficult to simulate this background, so a data-driven method described in section 5.1 is used to estimate it.

- **Z-boson production:** In the leptonic channel, Z bosons decay to two same-flavour opposite-charge leptons, therefore contributing to the background of this analysis if one lepton is missed in a $Z \rightarrow \ell\ell + \text{jets}$ decay. This mismeasurement can also give rise to E_T^{miss} . This background is modelled using the same SHERPA setup as the $W + \text{jets}$ SHERPA signal.
- **Top-quark production:** Top quarks decay almost always through the $t \rightarrow Wb$ channel, where b refers to a bottom quark which can manifest as a b -jet. Thus, the decay products are $W + \text{jets}$, mimicking the signal of this analysis. There are two processes considered in this analysis which produce top quarks: pair production ($t\bar{t}$) and “single-top”. The $t\bar{t}$ process exhibits the higher cross section of the two; in this case, should one top decay leptonically and the other hadronically, this process at leading order will mimic a $W + \text{jets}$ decay with four jets, two of which originate from b -quarks. As a result, this is the dominant background at higher jet multiplicities. Similarly to the Z -boson background, if both top quarks decay leptonically, this can also mimic a $W + \text{jets}$ decay if one lepton is missed. The single-top process involves the production of a top quark in association with either a b quark or a W boson. Both of these can mimic the $W + \text{jets}$ signal directly, either through leptonic decay of the top quark in the former or hadronic decay in the latter.

These processes are modelled using MC simulation. The POWHEG-BOX v2 [111–114] generator at NLO in QCD with the NNPDF3.0NLO [108] PDF set is interfaced to PYTHIA (v8.230) [115] to model the parton shower, hadronisation, and underlying event, with parameters set according to the A14 tune [104]

and using the NNPDF2.3 LO set of PDFs [105]. The decays of bottom and charm hadrons are performed by EVTGEN v1.6.0 [116].

- **τ production:** ($W \rightarrow \tau\nu$) + jets forms a background to this analysis as one of the decay modes of the τ particle is $\tau \rightarrow (e/\mu)\nu\nu$. As both of the neutrinos give rise only to E_T^{miss} in the detector, they give the same signature as a single neutrino. Similarly, ($Z \rightarrow \tau^+\tau^-$) + jets forms a background if either both τ particles decay leptonically and one lepton is missed, or one of the τ particles decays hadronically giving rise to a process which at leading order mimics W + jets with a jet multiplicity of two. This background is modelled using the same SHERPA setup as the W + jets SHERPA signal and Z + jets background.
- **Diboson production:** Diboson production is considered in three forms in this analysis: WW , WZ , and ZZ . Each can mimic the W + jets signal for the same reasons stated above, particularly if one boson decays leptonically while the other decays hadronically. It is suppressed relative to the other background components as the production of two bosons exhibits a smaller cross section. The POWHEG-BOX v2 [112–114] generator is used to generate the WW , WZ and ZZ [117] processes at NLO-accuracy in QCD. Events are interfaced to PYTHIA (v. 8.186) [103] for the modelling of the parton shower, hadronisation, and underlying event, with parameters set according to the AZNLO tune [118]. The CT10 PDF set [119] is used for the hard-scattering processes and the CTEQ6L1 PDF set [120] is used for the parton shower.

These predictions, along with W + jets estimated through MC, are compared to data in section 5.1.4 for a variety of kinematic distributions and jet multiplicities.

5.1 The estimation of fake and non-prompt leptons

This background can be classified in to two categories: *fake* leptons, where objects such as jets are incorrectly reconstructed as leptons; and *non-prompt* leptons, where

the reconstructed lepton is real but it does not originate from the collision point. This section presents the data-driven Matrix Method, which is used to estimate this background, and presents the results of this estimation in the electron channel.

5.1.1 Sources of fake and non-prompt electrons

The reconstruction of an electron depends on the reconstructed track in the inner detector which is matched to a deposition of energy within the EM calorimeter, as described in section 4.1.1. A similar signature is given by hadrons from light-quark jets (u , d or s) or low-energy gluon jets. These are usually distinguished through the likelihood and isolation criteria, discussed in section 4.1.1, which capitalise on the differences between the signature of the light jets and that of the electrons; whereas the light jets exhibit a more diffuse shower deposited in both the EM and hadronic calorimeter, the shower of an electron is typically narrow and fully contained within the EM calorimeter. Although these criteria do suppress the background substantially, its high cross section provides a substantial contamination of events nevertheless.

Non-prompt electrons can be produced in multiple ways. Firstly, heavy-quark jets can give rise to a lepton embedded within the jet if the initial quark exhibits a semi-leptonic decay such as $b \rightarrow (W \rightarrow e\nu)$. The aforementioned criteria, along with the requirement that $\min(\Delta R(e, \text{jet})) > 0.4$, reduce this background by rejecting electrons with high activity within the vicinity of the candidate. Additionally, non-prompt electrons can be produced through the pair-production of photons, $\gamma \rightarrow e^+e^-$, resulting in two tracks and deposits which can, if collimated enough, be interpreted by the detector as a single track and deposit.

5.1.2 The Matrix Method

The Matrix Method is a data-driven background estimation method introduced relatively recently to the ATLAS experiment and so far implemented mainly in top-quark measurements, such as those in Refs. [121–123] with details of the application for final states with one or two leptons given in Ref. [124]. This is

the first time that the method has been applied to an analysis of the production of an electroweak boson with the ATLAS detector.

The proportion of faked electrons (where “faked” refers now and henceforth to both fake and non-prompt electrons for conciseness) is estimated by extrapolation from a loose selection criteria to the tight selection criteria used as the signal region of the analysis. The loose criteria is such that it is the same in terms of kinematics as the tight criteria, but the isolation requirement is removed and a loose LH criterion is used. As a result, the tight selection is a subset of the loose selection, and the loose selection contains a higher proportion of events with a fake electron. The numbers of tight and loose events, N_t and N_l respectively, are related to the numbers of events containing real and faked electrons in the loose sample, N_r and N_f respectively, by the following matrix equation:

$$\begin{bmatrix} N_t \\ N_l \end{bmatrix} = \begin{bmatrix} \epsilon_r & \epsilon_f \\ 1 & 1 \end{bmatrix} \begin{bmatrix} N_r \\ N_f \end{bmatrix}, \quad (5.1)$$

where $\epsilon_{r,f}$ refers to the real and fake efficiencies, the probability of an event in the loose criteria also passing the tight criteria for real and fake electrons respectively. The number of fake events passing the tight selection is:

$$\begin{aligned} N_t^{\text{Fake}} &= N_t^f \\ &= \epsilon_f N_f. \end{aligned} \quad (5.2)$$

N_f is calculated by inverting the matrix in eq. (5.1):

$$\begin{bmatrix} N_r \\ N_f \end{bmatrix} = \frac{1}{\epsilon_r - \epsilon_f} \begin{bmatrix} 1 & -\epsilon_f \\ -1 & \epsilon_r \end{bmatrix} \begin{bmatrix} N_t \\ N_l \end{bmatrix}. \quad (5.3)$$

Therefore:

$$N_t^f = \epsilon_f \frac{1}{\epsilon_r - \epsilon_f} (\epsilon_r N_l - N_t). \quad (5.4)$$

In general, these efficiencies are themselves dependent on the topology of the event, for example different regions of the detector at different energies can exhibit different effectiveness at identifying fake electrons. In order to account for this, the fake contamination is calculated on a weighted event-by-event basis, ie.

$$N_t^f = \sum_{i=1}^{N_l} w_i,$$

where:

$$w_i = \frac{\epsilon_f}{\epsilon_r - \epsilon_f}(\epsilon_r - \delta_i), \quad (5.5)$$

where $\delta_i = 1$ if the event is tight, else 0. The efficiencies are parameterised by the following equation:

$$\epsilon(\vec{x}; \vec{y}) = \epsilon(\vec{x}) \prod_{j=1}^n \frac{\epsilon(\vec{x}; y_j)}{\epsilon(\vec{x})}, \quad (5.6)$$

where $\vec{x}$ is the set of variables through which the efficiencies are considered separately, and $\vec{y}$ is the set of n variables through which the efficiencies are parameterised.

The efficiencies are measured in dedicated analyses whose selection criteria differ from the signal region of this analysis. The real efficiency has been measured by the ATLAS collaboration using the Tag And Probe method discussed in chapter 4. The ratio of the number of tight probe electrons to the number of loose probe electrons, where tight is a subset of loose, is the real efficiency ϵ_r . This measurement is propagated to the analysis by the use of the measured scale factors in MC simulation, as discussed in section 4.2.3 [110]. The measurement of the fake efficiency for this analysis is discussed in section 5.1.3.

5.1.3 The measurement of the fake efficiency

The fake efficiency, ϵ_f , is measured in a dedicated “control region” for events with $E_T^{\text{miss}} < 25$ GeV as this is where the contamination of the background is expected to be largest. The efficiencies are parameterised through the following variables:

- The transverse momentum of the electron, p_T^e ;
- The pseudorapidity of the electron, η_e ;
- The transverse momentum of the leading jet, p_T^{leading} ;
- The transverse mass of the electron- E_T^{miss} pair, M_T^W ;
- The angular distance between the electron and the nearest jet, $\min(\Delta R(e, \text{jet}))$;
- The angular distance between the transverse momentum of the electron and E_T^{miss} , $\Delta\phi(e, E_T^{\text{miss}})$.

For each variable, the fake efficiency is measured by taking measurements of each in data for both the loose and tight selection criteria, subtracting the $W + \text{jets}$ and each background MC, and dividing the tight histogram by the loose:

$$\epsilon_f = \frac{N_{\text{tight}}^{\text{data}} - N_{\text{tight}}^{\text{MC}}}{N_{\text{loose}}^{\text{data}} - N_{\text{loose}}^{\text{MC}}}, \quad (5.7)$$

where N is the number of events and “MC” refers to all MC simulation including the $W + \text{jets}$ signal. The resulting efficiency plots are shown in fig. 5.1.

The topology of the fake background depends on the number of reconstructed jets in the event. This is depicted in fig. 5.2, where the efficiencies for a number of jet multiplicities for the $\min(\Delta R(e, \text{jet}))$ variable as an example are displayed and both the shape and normalisation of the distributions differ. As a result, each jet multiplicity is considered separately for the fake-contamination estimation.

The implementation of the Matrix Method described so far is the one used for the aforementioned top-quark analyses. For this $W + \text{jets}$ analysis, a “validation region” is introduced with the aim of establishing the legitimacy of the model in another region of phase space differing from the signal region where the multijet background contamination is high. The criterion for each region, where it differs from the signal region, is shown in table 5.1 and depicted in fig. 5.3. The real efficiency is taken separately for the signal and validation regions.

Criterion	Signal	Control	Validation
$E_{\text{T}}^{\text{miss}} + M_{\text{T}}^W$	$> 60 \text{ GeV}$	-	$< 60 \text{ GeV}$
$E_{\text{T}}^{\text{miss}}$	$> 25 \text{ GeV}$	$< 25 \text{ GeV}$	$> 25 \text{ GeV}$

Table 5.1: The selection criterion on $E_{\text{T}}^{\text{miss}}$ and $(E_{\text{T}}^{\text{miss}} + M_{\text{T}}^W)$ for the signal, control and validation regions. Where an entry is missing, no selection criterion is applied. These regions do not differ in any other selection criterion.

With this initial implementation, significant inconsistencies were found with data in the validation region. The cause of this was the variables M_{T}^W and $\Delta\phi(e, E_{\text{T}}^{\text{miss}})$, which contain bins for which the fake efficiency differs significantly from the mean. This can be problematic for events which inhabit these bins in both distributions,

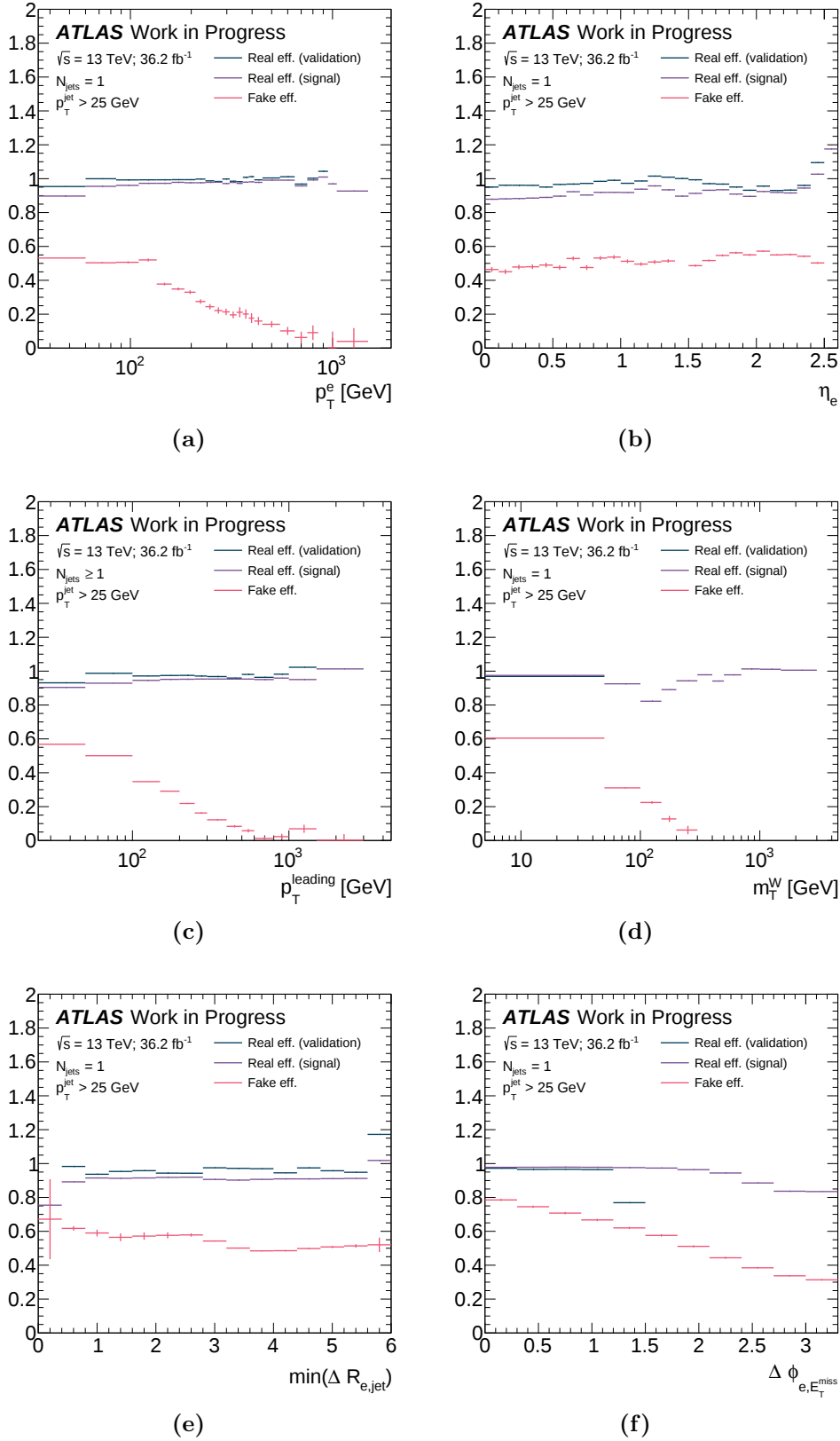


Figure 5.1: Plots depicting the real efficiency for each of the signal and validation regions as well as the measured fake efficiencies as a function of each of the variables over which the Multijet estimation is parameterised.

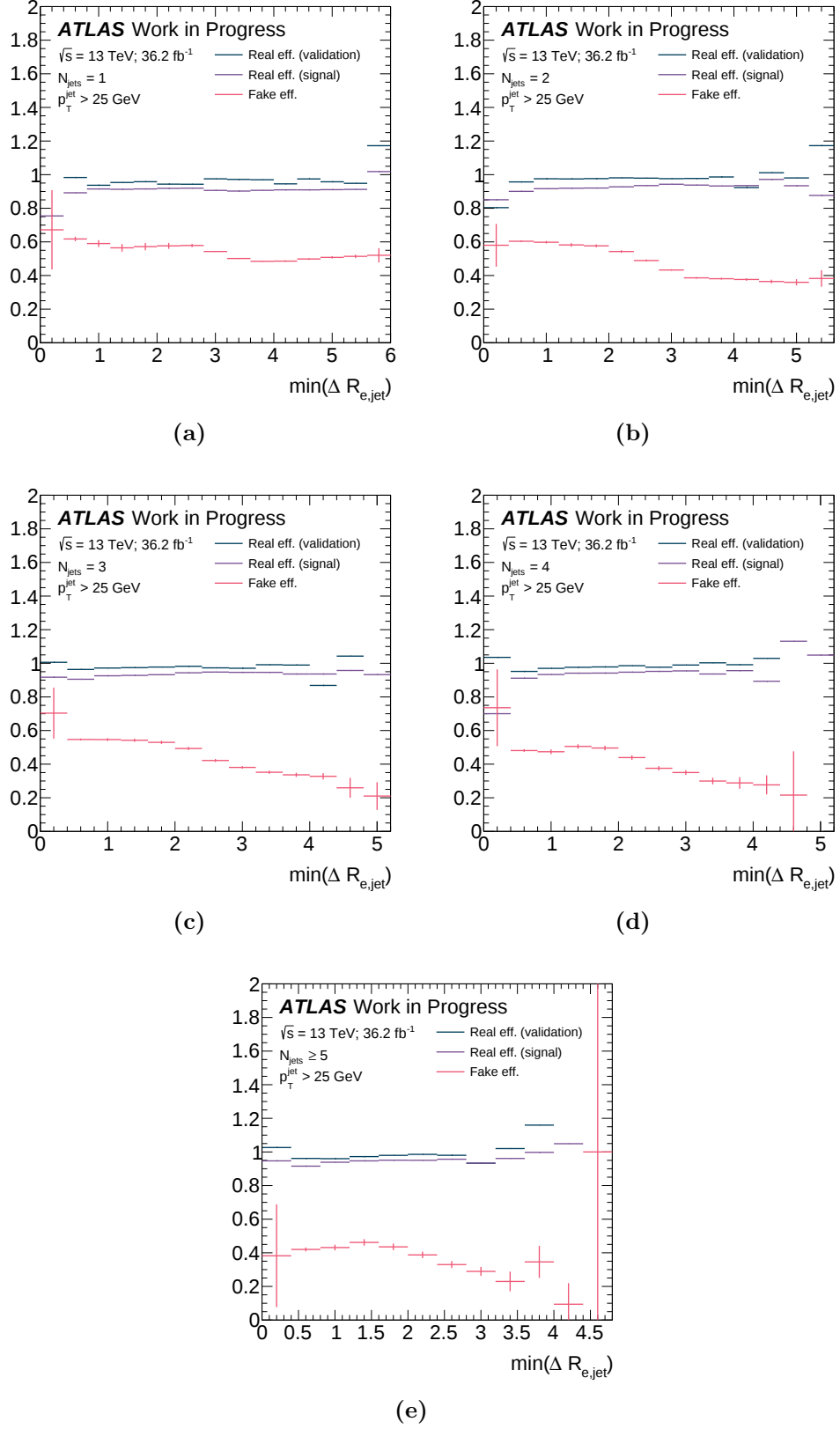


Figure 5.2: Plots depicting the real efficiency for each of the signal and validation regions as well as the measured fake efficiencies as a function of $\min(\Delta R(e, \text{jet}))$ for exclusive jet multiplicities up to four, and five inclusive.

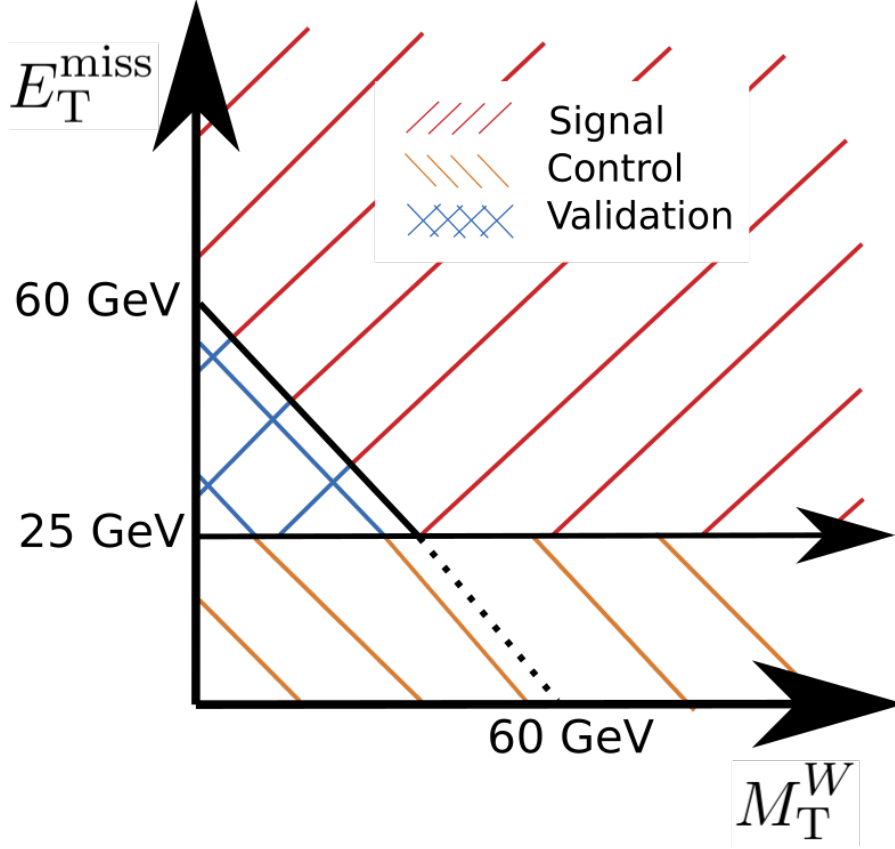


Figure 5.3: A diagram depicting the signal, control and validation regions on the $E_T^{\text{miss}} - M_T^W$ plane. These regions are equivalent in all other criteria.

for example an event i with properties $M_T^W = 10$ GeV, $\Delta\phi(e, E_T^{\text{miss}}) = 0.1$ and a single jet would exhibit a fake efficiency parameterised through these observables of:

$$\begin{aligned}
 \epsilon_f(M_T^W, \Delta\phi(e, E_T^{\text{miss}}); N_{\text{jet}}) &= \epsilon_f(N_{\text{jet}}) \frac{\epsilon_f(M_T^W; N_{\text{jet}})}{\epsilon_f(N_{\text{jet}})} \frac{\epsilon_f(\Delta\phi(e, E_T^{\text{miss}}); N_{\text{jet}})}{\epsilon_f(N_{\text{jet}})} \\
 &\sim 0.52 \times \frac{0.71}{0.52} \times \frac{0.80}{0.52} \\
 &= 1.09,
 \end{aligned} \tag{5.8}$$

which is certainly an overestimation as the data passing the tight criteria are a subset of that passing the loose, so this fraction must be below one. As the real efficiency does not deviate so much in this way, this also results in negative weights for loose events and positive weights for tight events, thereby giving an estimation of a negative number of events in some bins.

By using an efficiency parameterised through a two-dimensional histogram of these observables, the description of the data in both the validation and signal

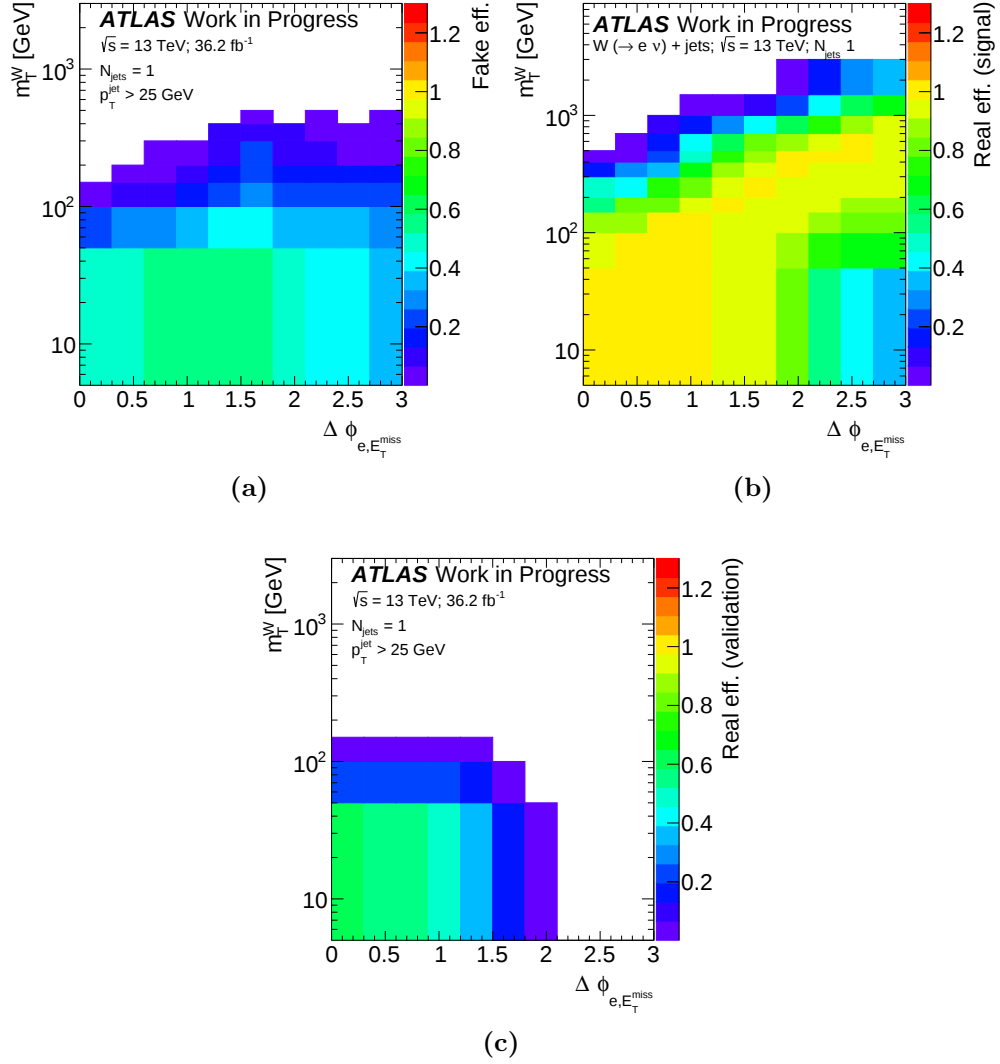


Figure 5.4: Plots depicting the two-dimensional $\Delta\phi(e, E_T^{\text{miss}})$ – M_T^W one-jet exclusive (a) fake efficiency evaluated by MG + PY8 W +jets MC, (b) real efficiency evaluated by MG + PY8 W +jets MC, (c) fake efficiency evaluated by SHERPA2.2.1 W +jets MC and (d) real efficiency evaluated by SHERPA2.2.1 W +jets MC.

regions improved. A sample of the two-dimensional efficiencies is shown in fig. 5.4. In this case, the resulting fake efficiency for an event with properties $M_T^W = 10$ GeV, $\Delta\phi(e, E_T^{\text{miss}}) = 0.1$ is 0.80. Furthermore, the correlation between these particular observables is now accounted for. On the other hand, the statistical precision of the measured efficiency is reduced as the events are diluted over a higher number of bins.

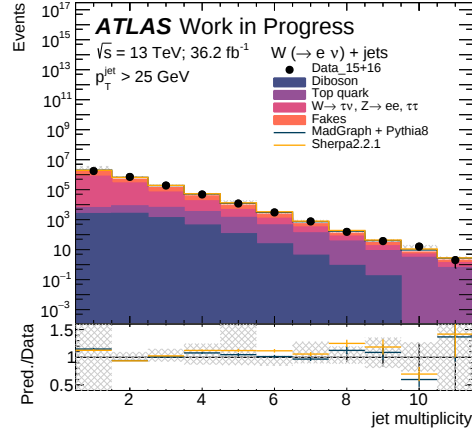
5.1.4 Results and discussion

The resulting comparison between data and estimation in the validation and signal regions is shown for up to two jets inclusive in figs. 5.5 to 5.8 for the following observables:

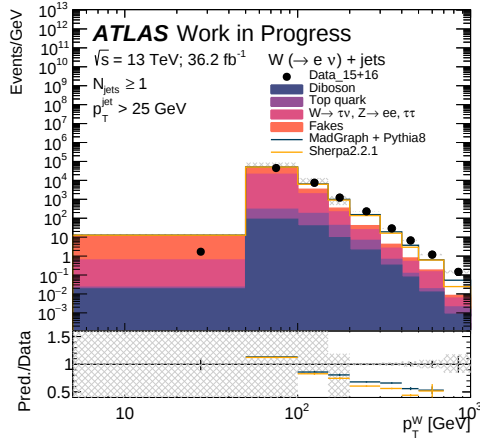
- N_{jet} : the jet multiplicity,
- $p_{\text{T}}^W = \sqrt{(p_{\text{T}}^e)^2 + (E_{\text{T}}^{\text{miss}})^2 + 2p_{\text{T}}^e E_{\text{T}}^{\text{miss}} \cos(\Delta\phi(e, E_{\text{T}}^{\text{miss}}))}$: the transverse momentum of the W boson,
- $x_{\text{T-odd}}$: the x variable of the T-odd asymmetry measurement discussed in section 4.2.2,
- $p_{\text{T}}^{\text{leading}}$: the transverse momentum of the leading jet,
- H_{T} : the scalar sum of the p_{T} of all reconstructed objects and $E_{\text{T}}^{\text{miss}}$.

Additionally, the composition of the data for each jet multiplicity is given in tables 5.2 and 5.3 for the signal and validation regions, respectively. Statistical and systematic uncertainties as discussed in chapter 7 are included in the plots of the distributions measured in the signal region. In the validation region plots, only statistical uncertainties and the uncertainty propagated from the statistical uncertainty on the measured fake efficiency and systematic uncertainties on the real efficiencies are shown. Additional figures with measurements split by the charge of the electron and for up to an inclusive jet multiplicity of four are shown in appendix B.

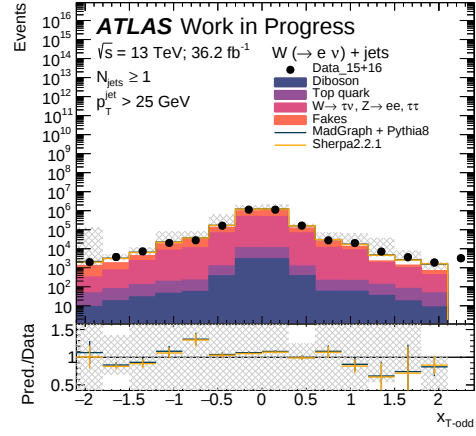
Deviations from the data in some regions of phase space are clearly noticeable in the validation region. Multiple causes may be responsible, for example at high p_{T}^W and $p_{\text{T}}^{\text{leading}}$, this is where the prediction is dominated by the $W + \text{jets}$ signal and a similar deviation is seen in the signal region, so it may be the case that the MC modelling is responsible rather than the multijet modelling. More deviations are seen at lower energy scales, particularly in the $p_{\text{T}}^{\text{leading}}$ and H_{T} spectra, however they are typically within the statistical uncertainty of the fake efficiency measurement.



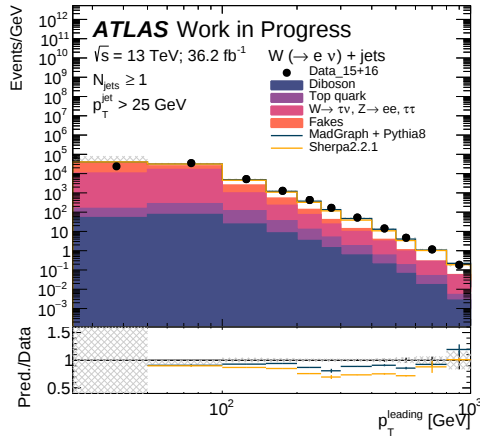
(a)



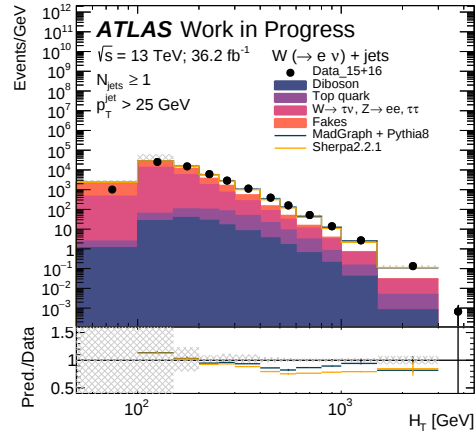
(b)



(c)



(d)



(e)

Figure 5.5: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region. The distributions shown are the (a) jet multiplicity, (b) p_T^W , (c) $x_{T\text{-odd}}$, (d) p_T^{leading} and (e) H_T .

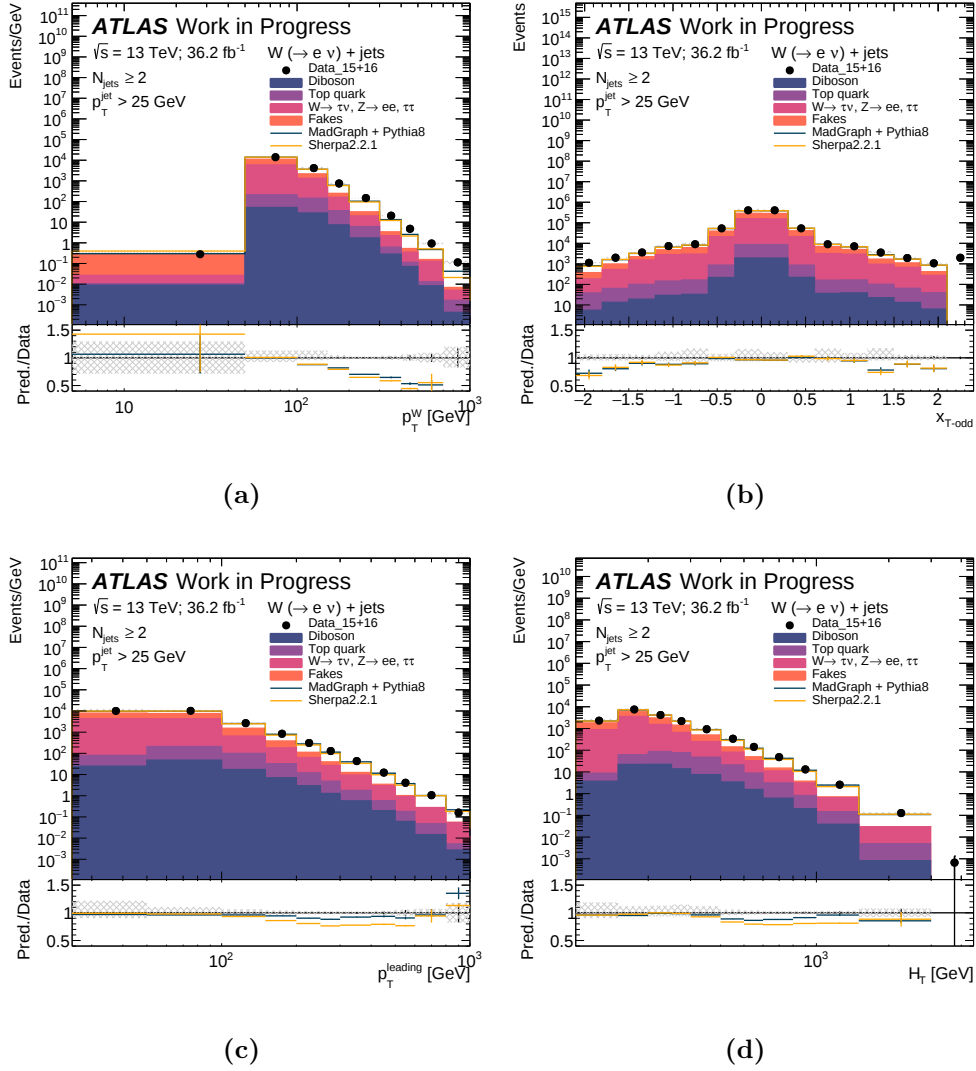
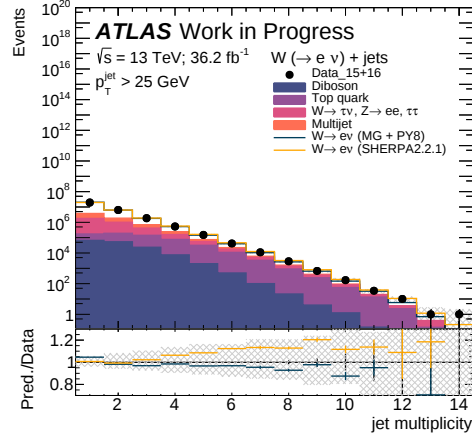


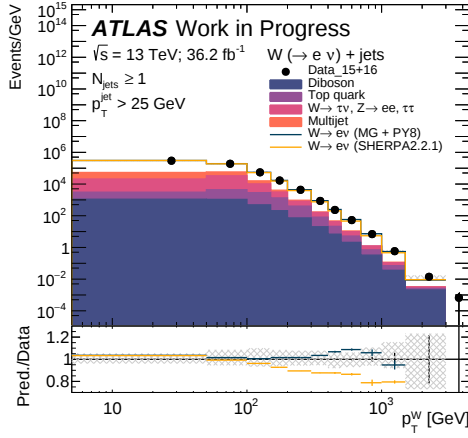
Figure 5.6: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region for two or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T .

With the additional integrated luminosity from the data collected in years 2017 and 2018, further studies could have a strong effect on this analysis as all statistical uncertainties, including those on the fake efficiency, will be dramatically reduced and the observables driving the fake efficiency could be grouped further - for example, in to a three-dimensional $\min(\Delta R(e, \text{jet})), M_T^W$ and $\Delta\phi(e, E_T^{\text{miss}})$ efficiency, if it improves the model.

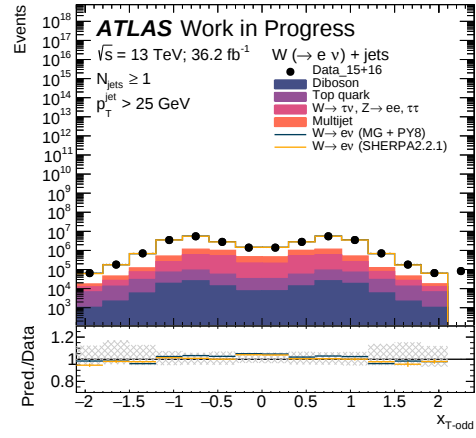
The estimation of the MC-simulated backgrounds could also be validated through



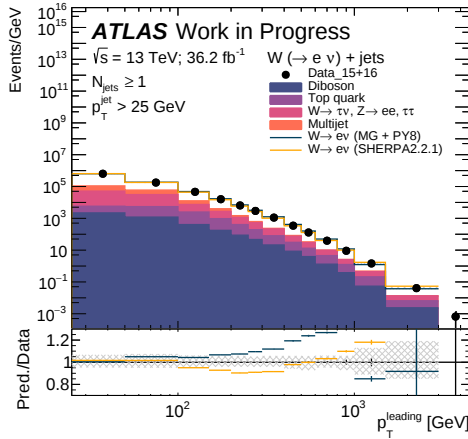
(a)



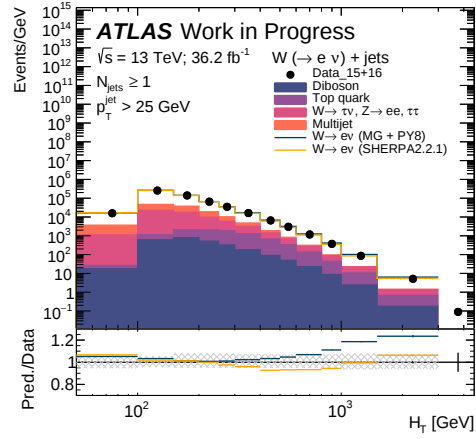
(b)



(c)



(d)



(e)

Figure 5.7: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis. The distributions shown are the (a) jet multiplicity, (b) p_T^W , (c) $x_{T\text{-odd}}$, (d) p_T^{leading} and (e) H_T .

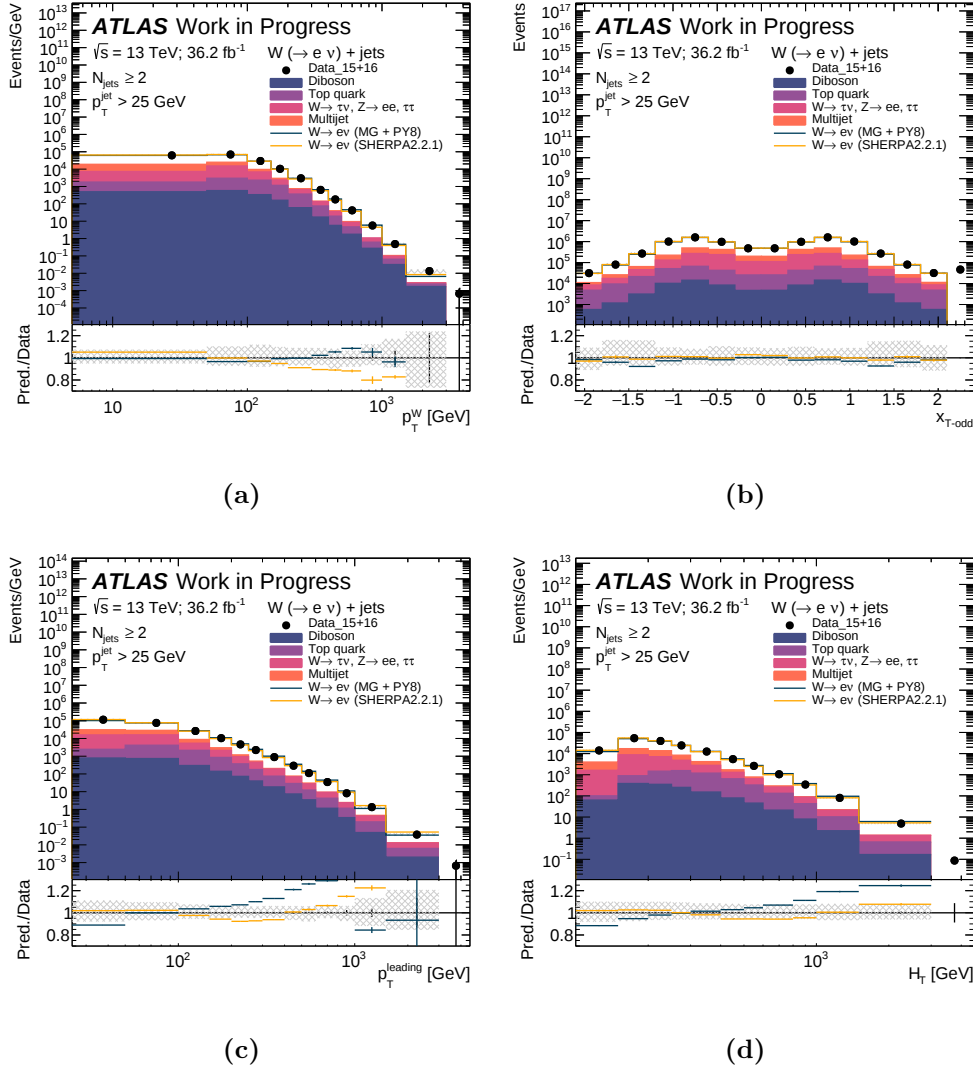


Figure 5.8: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis for two or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T .

dedicated selection criteria. For example, the top background could be validated in a sample which requires at least one b -tagged jet. The normalisation of the top background could subsequently be fit to data in this top-validation region, and that scale factor applied to the background of the signal region. This is most important for higher-jet multiplicities, for example at four jets where top-decays constitute 13% of the background, and higher where the top background is the most dominant. This will be implemented in later stages of the analysis.

N_{jets}	$W \rightarrow e\nu$ (MG + PY8)	$W \rightarrow e\nu$ (SHERPA2.2.1)	Diboson	Top quark	$W \rightarrow \tau\nu, Z \rightarrow ee, \tau\tau$	Multijet
1	0.852	0.813	0.003	0.005	0.083	0.093
2	0.688	0.699	0.008	0.022	0.129	0.148
3	0.592	0.645	0.013	0.064	0.150	0.153
4	0.526	0.605	0.014	0.130	0.155	0.172
5	0.470	0.588	0.013	0.198	0.160	0.136
6	0.426	0.581	0.012	0.253	0.166	0.111
7	0.397	0.575	0.009	0.295	0.169	0.062
8	0.360	0.560	0.007	0.311	0.172	0.056
9	0.368	0.594	0.006	0.356	0.181	0.030
10	0.315	0.556	0.007	0.325	0.169	0.053
11	0.367	0.553	0.004	0.391	0.166	0.035

Table 5.2: The composition of the total Standard-Model prediction of the data passing the signal criteria of the $W + \text{jets}$ 13 TeV analysis for each jet multiplicity. Each entry is taken as a ratio to the total number of events in reconstructed data.

N_{jets}	$W \rightarrow e\nu$ (MG + PY8)	$W \rightarrow e\nu$ (SHERPA2.2.1)	Diboson	Top quark	$W \rightarrow \tau\nu, Z \rightarrow ee, \tau\tau$	Multijet
1.0	0.204	0.180	0.001	0.002	0.449	0.492
2.0	0.242	0.237	0.004	0.008	0.384	0.303
3.0	0.263	0.282	0.007	0.027	0.332	0.377
4.0	0.273	0.318	0.009	0.063	0.307	0.426
5.0	0.270	0.344	0.010	0.109	0.279	0.379
6.0	0.252	0.356	0.008	0.140	0.260	0.354
7.0	0.264	0.355	0.006	0.169	0.254	0.274
8.0	0.312	0.438	0.006	0.238	0.279	0.290
9.0	0.284	0.382	0.005	0.227	0.260	0.311
10.0	0.166	0.264	0.000	0.190	0.128	0.114
11.0	0.384	0.434	0.000	0.320	0.354	0.310

Table 5.3: The composition of the total Standard-Model prediction of the data passing the validation criteria of the $W + \text{jets}$ 13 TeV analysis for each jet multiplicity. Each entry is taken as a ratio to the total number of events in data.

In the signal region, the MG + PY8 prediction appears to display a good consistency with data at all jet multiplicities as each individual data point is consistent within statistical uncertainties. The same generator also shows good agreement with data in the differential spectra except at high $p_{\text{T}}^{\text{leading}}$ and H_{T} , where a large deviation well outside of the total uncertainty is observed. The SHERPA2.2.1 prediction appears to exhibit a systematic over-prediction at high jet multiplicities, but better agreement across the differential spectra in comparison to the MG + PY8 prediction. Both generators give similar results for the $x_{\text{T-odd}}$ spectrum.

The ratio of the estimation of the multijet background to the sum of the same estimation of the multijet background and the MG + PY8 prediction of the $W + \text{jets}$ signal is shown in fig. 5.9 for the $E_{\text{T}}^{\text{miss}} - M_{\text{T}}^W$ plane, where the selection criteria on those observables has been removed. This plot shows that the multijet events

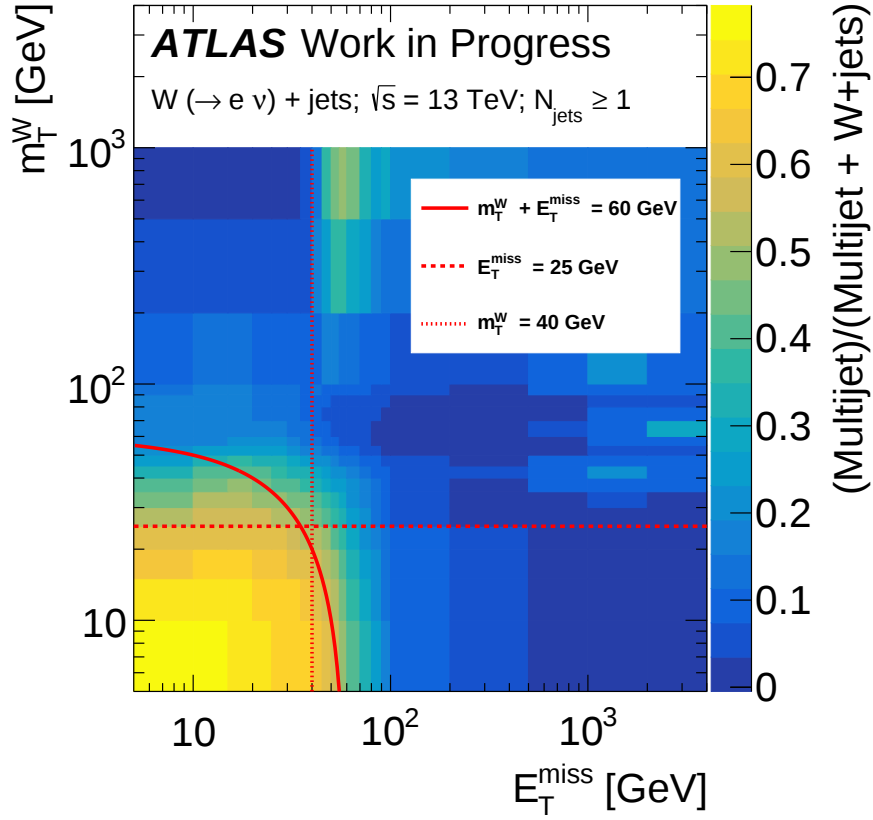


Figure 5.9: The ratio of the estimation of the multijet background to the sum of the same estimation of the multijet background and the MG + PY8 prediction of the $W + \text{jets}$ signal. The selection criteria on the E_T^{miss} and M_T^W variables has been removed. Also shown in red lines are the boundaries defining the criterion of this analysis ($E_T^{\text{miss}} + M_T^W < 60 \text{ GeV}$ and $E_T^{\text{miss}} < 25 \text{ GeV}$) as well as the additional criterion of the ATLAS $W + \text{jets}$ 8 TeV analysis ($M_T^W < 40 \text{ GeV}$) at Ref. [100].

typically exhibit the property that $E_T^{\text{miss}} + M_T^W < 60 \text{ GeV}$, and the $W + \text{jets}$ signal does not, thereby giving justification for this criterion. One advantage of the Matrix Method is that it allows for this kind of E_T^{miss} -dependent criterion to be established outside of the definition of the control region, whereas other methods such as the Template Method used in the ATLAS $W + \text{jets}$ 8 TeV analysis [100] could only make the additional criterion that $M_T^W > 40 \text{ GeV}$ in both the signal and control regions, shown also in fig. 5.9. This provides both flexibility in the definition of the signal region and enhanced statistics in the control region. Further analysis will involve the use of the Template Method in a restricted signal region as a cross-check.

6

Deconvolution/Unfolding

Contents

6.1	Iterative Bayesian unfolding	103
6.2	Response matrices	104
6.3	Closure tests	105
6.4	Selecting the number of iterations	105

A significant restriction to the comparison between particle-physics theory and experiment is the modelling of the detector. Cross sections can be calculated directly, and can be done so as a function of various observables, but to account for the detector one must model this theory on an event-by-event basis, requiring an understanding of the correlations between all observables. The particle-level final state must be propagated through various processes such as hadronisation and showering, and finally, propagated through a full simulation of the ATLAS detector. This process is known as the *folding* of the prediction from the particle-level estimation to the reconstructed distribution. Alternatively, it can be referred to as *convolution*, as the particle-level distribution is convoluted with the simulation of detector effects to return the reconstructed distribution. It is not sufficient to simply understand how a particular set of kinematics is reconstructed in the detector as the convolution is probabilistic in nature.

Criterion	
Leptons	$p_T^e > 35 \text{ GeV}$
	$ \eta^e < 2.47$ (excl. transition)
	No second e : $p_T > 25 \text{ GeV}$, $ \eta < 2.47$
	No muon: $p_T > 25 \text{ GeV}$, $ \eta < 2.47$
W -boson	$M_T^W + E_T^{\text{miss}} > 60 \text{ GeV}$ and $E_T^{\text{miss}} > 25 \text{ GeV}$
Jets	At least one
	$\min(\Delta R(e, \text{jet})) > 0.4$
	$p_T > 30 \text{ GeV}$

Table 6.1: A table detailing the fiducial selection criteria of the $W + \text{jets}$ analysis.

In order to remove the necessity of this extra complication and requirement of computational power, the opposite operation, known as *unfolding* or *deconvolution*, is performed to estimate the particle-level distributions from the reconstructed observations. The method used in this analysis is iterative Bayesian unfolding, as described by D’Agostini [125, 126]. The probabilistic nature of the unfolding process is accommodated for using Bayes’ theorem, resulting in additional uncertainties on the unfolded cross sections. Distributions can be unfolded only to a restricted phase space similar to the reconstructed one – this is referred to as the *fiducial* phase space, described in table 6.1. This selection is similar to that discussed in section 4.2 except that the detector-level variables such as the likelihood and isolation criteria are removed. Notably, the b -jet veto is removed, meaning that this analysis is extrapolating from measurements of light-flavour to heavy-flavour jets; this follows the precedent of previous ATLAS analyses such as that of Ref. [100], but studies are ongoing within both the analysis group and the ATLAS collaboration as a whole to determine whether this is reliable. The particle-level electrons are *dressed*, meaning that radiative QED effects are accounted for in the calculation of its kinematics.

The iterative Bayesian unfolding method is described in detail in section 6.1. The response matrices and particle-level measurements required for the unfolding are presented in section 6.2, with closure tests presented in section 6.3 and the optimisation of the number of iterations presented in section 6.4. The calculation

of systematic uncertainties is described in chapter 7, and the unfolded distributions are presented with their full uncertainties in chapter 8.

6.1 Iterative Bayesian unfolding

The number of particle-level events occupying bin i in a histogram, $N_i^{(T)}$ ¹, is extrapolated from the number of reconstructed events in bin j , $N_j^{(R)}$, by the following:

$$N_i^{(T)} = \sum_j \mathcal{M}_{i,j} N_j^{(R)}, \quad (6.1)$$

where $\mathcal{M}_{i,j} = P(T_i|R_j)/\epsilon_i$, $P(T_i|R_j)$ is the probability that the particle-level value falls in bin T_i given that it was reconstructed in bin R_j , and ϵ_i is the efficiency of an event in bin i being reconstructed. This probability can be calculated using the application of Bayes' theorem:

$$P(T_i|R_j) = \frac{P(R_j|T_i)P(T_i)}{\sum_i P(R_j|T_i)P(T_i)}, \quad (6.2)$$

where the denominator can be considered as a normalisation factor rather than directly calculated.

$P(T_i)$ is referred to in Bayesian statistics as the *prior*, denoting prior knowledge or assumptions on the probability of the event occupying the T_i bin. $P(R_j|T_i)$ is referred to as the *response matrix* which, along with the efficiency, can be determined from MC simulation.

The drawback of the Bayesian approach is that there is a reliance on the prior; this is an arbitrary element of the calculation and precisely something which should not influence the result as the purpose of the unfolded distribution is to be compared with alternative priors. This is where the iterative nature of the method is implemented. The unfolding process is repeated multiple times, each time using the unfolded distribution from the previous iteration as the new prior. Upon each iteration, the influence of the original prior is reduced; typically, only two or three iterations are required before it is sufficiently diminished. Convergence

¹The superscript (T), signifying “Truth”, is used to describe particle-level distributions to avoid confusion with the symbol P used when referring to probabilities.

to the preferred unfolded distribution occurs in a smaller number of iterations if the original prior is close to the solution, therefore the true distribution of signal MC is used, though in principle any prior can be used. Although one drawback is resolved, another is introduced as the response matrix is statistically limited, so the statistical uncertainties on the prior and resulting unfolded distribution increase with each iteration - as a result, the number of iterations must be kept as low as possible. The error propagation is discussed briefly in chapter 7 and described in detail in Ref. [127].

The iterative Bayesian unfolding method is implemented using the ROOUNFOLD software package [128]. The requirement is a reconstructed distribution, similar to those in figs. 5.5 to 5.8 with the background estimation subtracted; a particle-level distribution (which is also used as the initial prior), and a response matrix.

For the remainder of this analysis, the unfolding is performed using efficiencies and response matrices derived from the the MG + PY8 MC described in section 4.2.3. This is because this sample has been produced using a higher number of events, thereby improving the statistical precision of the response and therefore the final result.

6.2 Response matrices

The response matrices are calculated from MC simulation by filling a two-dimensional histogram - the vertical axis being the particle-level value and the horizontal the reconstructed value - with the events. The response matrices for a range of distributions are given in fig. 6.1. The binning of the histograms is chosen such that these matrices are almost diagonal - if the off-diagonal components were large, the histograms would not accurately represent the precision of the detector and the uncertainties of the unfolded distributions would suffer significantly.

The $x_{T\text{-odd}}$ response matrix shows some interesting properties. Firstly, the effect of the smearing is larger at larger absolute values. Secondly, a large proportion of the events reconstructed outside of the $|x_{T\text{-odd}}| < 1$ range actually come from events within the same range at particle level. This explains why such a large

proportion of reconstructed events appear outside of this range, even though there is no leading-order contribution.

6.3 Closure tests

The most simple test of the performance of the unfolding procedure is to assess the *closure* of the method. This means that the response matrix is constructed from a particular MC, and subsequently the reconstructed distribution from that same MC is unfolded. The agreement between the original prior and the unfolded distribution should be very close to perfect, with the potential for only slight losses due to limited statistics in the reconstructed distribution and response matrices. The result of this closure test is shown in fig. 6.2. In each case, the unfolded distribution overlays the particle-level distribution very closely, as expected. There is a minor deviation from the particle-level distribution at large absolute values $x_{T\text{-odd}}$ caused by the limited statistics of the response matrix and large migration between bins. Similar results for higher jet multiplicities and distributions split by the charge of the electron are shown in appendix B.

6.4 Selecting the number of iterations

As discussed in section 6.1, the number of iterations, n_{iter} , to be used in the iterative Bayesian unfolding technique is selected through a trade-off of minimising bias and uncertainty, where the former is reduced and the latter increased with each iteration. As such, it is important to attempt to quantify each and compare the result at each iteration.

The following method is used to determine the bias in this analysis:

1. Unfold the data and store the resulting unfolded distributions for each iteration up to the highest number of iterations to be analysed.
2. Take the ratio of the unfolded distribution to the particle-level MC-simulated distribution used in the unfolding process, and fit this with a smooth distribution, $f(x)$.

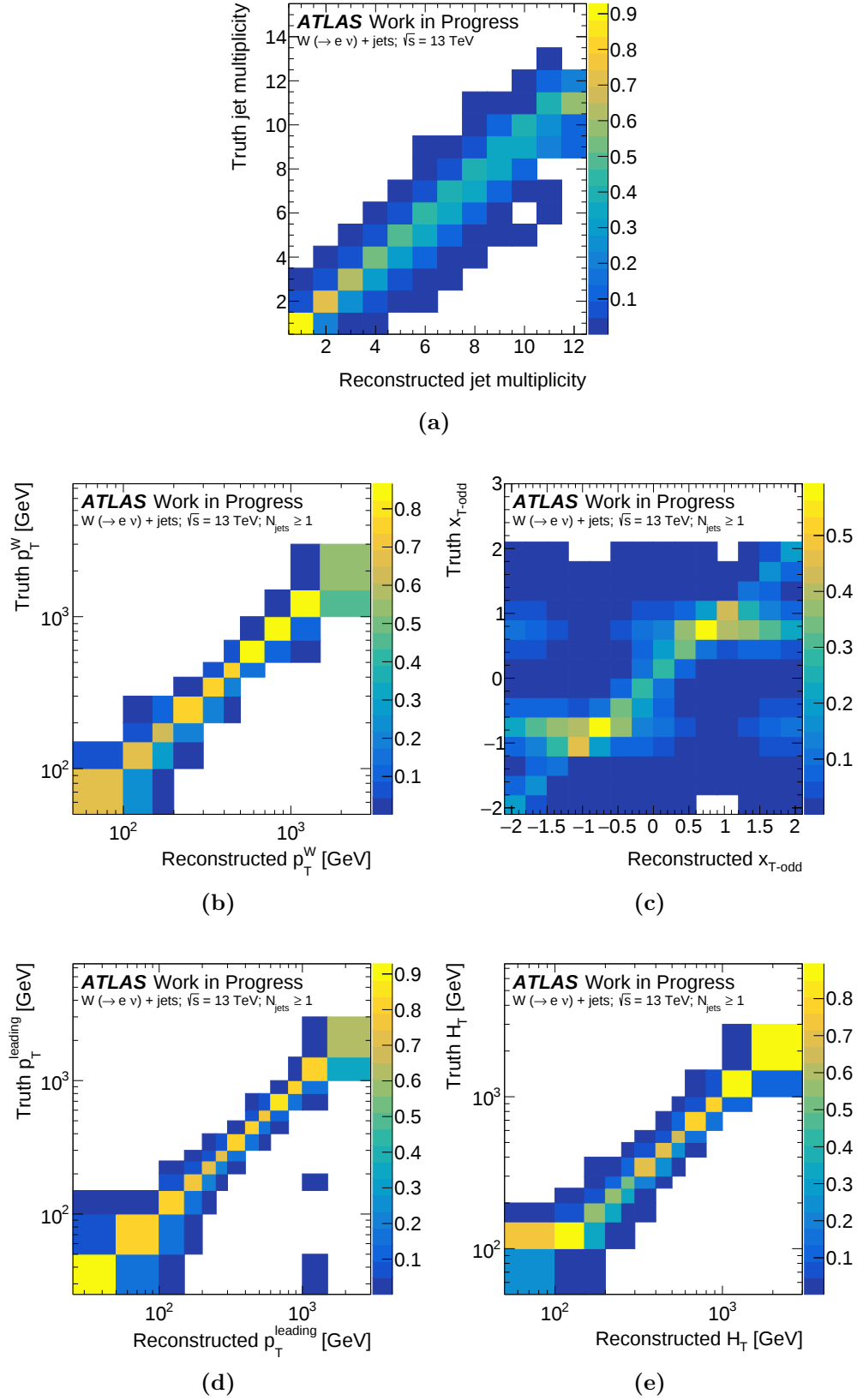
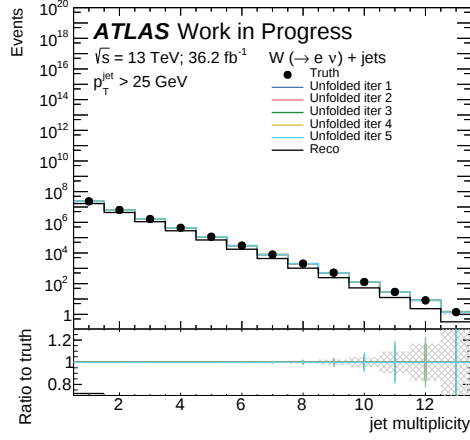
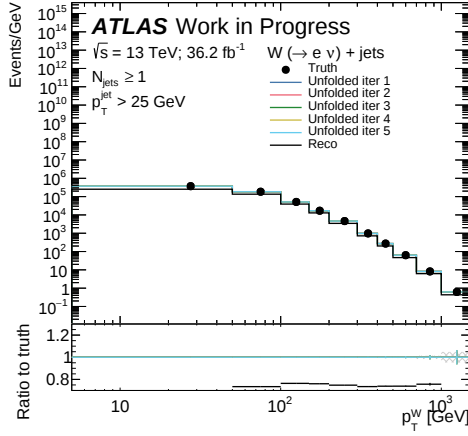


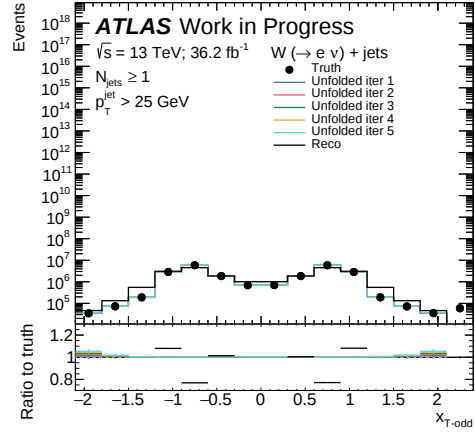
Figure 6.1: Two-dimensional histograms depicting the response matrices of each of the analysed spectra for events with at least one jet.



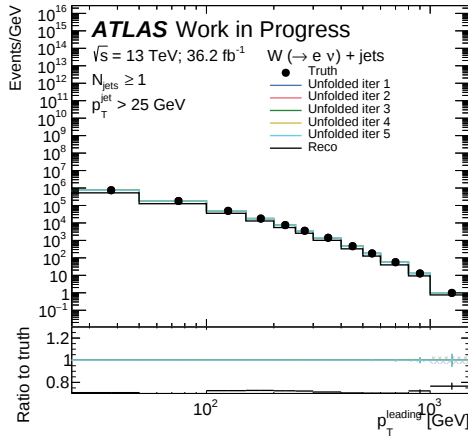
(a)



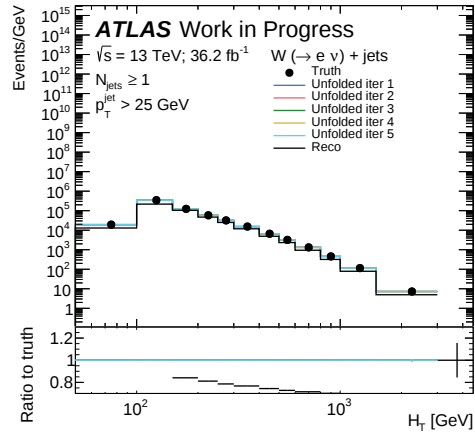
(b)



(c)



(d)



(e)

Figure 6.2: Closure tests of the unfolding procedure used in this analysis. The MG + PY8 particle-level distribution is shown as black points, the MG + PY8 reconstructed distribution is shown as black lines, and the unfolded MG + PY8 distribution is shown as a series of coloured lines for up to five iterations. In most of the bins shown, the unfolded distributions are overlaid such that only the five-iteration case is visible.

3. Reproduce the reconstructed MC distribution by applying a weight of $f(x_{\text{Particlelevel}})$ to each event.

If the unfolded distribution is unbiased, then re-folding (as it is done above) should reproduce the original distribution. It therefore follows that the bias can be quantified at each iteration by the difference between the data and the reweighted MC distribution.

For each of the evaluated distributions, the measurement (with background subtracted) is compared to the reweighted reconstructed MC evaluated at each unfolding iteration in fig. 6.3. The additional uncertainty of each iteration is estimated by varying the ratio by its uncertainty in each bin, refitting, and considering the difference between the new reweighted distribution and original as an additional uncorrelated uncertainty. The χ^2 per data point at each iteration for each of the distributions is summarised in table 6.2.

By requiring that the χ^2 per data point is one for sufficient convergence, the following is recognised:

- The N_{jet} distribution is almost converged after two iterations, with clear agreement after three. Subsequently, the agreement between the data and reweighted MC continues to improve only at a much-reduced rate.
- The p_{T}^W and $p_{\text{T}}^{\text{leading}}$ distributions show a similar convergence between two and three iterations when the lowest bin is excluded. Although the reweighted distributions in the lowest bin do appear to converge, they do not converge to the position of the same bin in the reconstructed data.
- The H_{T} , $x_{\text{T-odd}}$ and p_{T}^W (with the more narrow histogram bins) distributions do not converge to one as quickly.

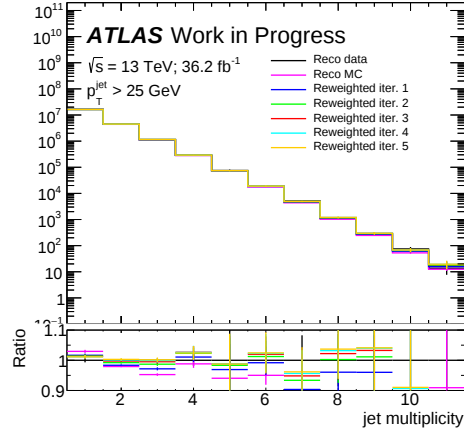
Notably, those distributions which do not appear to converge also suffer from response matrices with large migration between bins. This can be seen clearly by comparison between the p_{T}^W distribution with the histogram binning of this analysis, and the same distribution for which the width of the bins is halved. The migration

	N_{jet}	p_{T}^W	$p_{\text{T}}^{\text{leading}}$	H_{T}	$x_{\text{T-odd}}$	p_{T}^W (narrow binning)
Reconstructed MC	16.4	17.7	483.6	195.7	35.6	76.0
Reweighted MC iter. 1	5.1	2.7	17.6	8.1	24.1	12.7
Reweighted MC iter. 2	1.2	1.7	2.4	2.1	19.4	8.3
Reweighted MC iter. 3	0.5	1.1	1.2	1.4	15.4	5.7
Reweighted MC iter. 4	0.4	0.7	0.9	1.3	12.3	4.3
Reweighted MC iter. 5	0.4	0.4	0.8	1.3	10.0	3.5

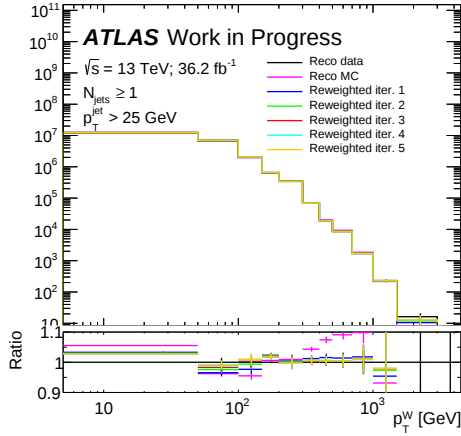
Table 6.2: The χ^2 per data point of the reconstructed MC and the reweighted MC against the reconstructed data for a range of distributions up to five iterations of the iterative Bayesian unfolding procedure, where the lowest bin is excluded.

between bins in the latter case is above 30% across the whole spectrum. Similarly, the H_{T} and $x_{\text{T-odd}}$ distributions suffer a larger migration than the $p_{\text{T}}^{\text{leading}}$ and p_{T}^W spectra. It could be that the statistical correlation between bins is not accounted for in the χ^2 test, thereby giving overestimated results.

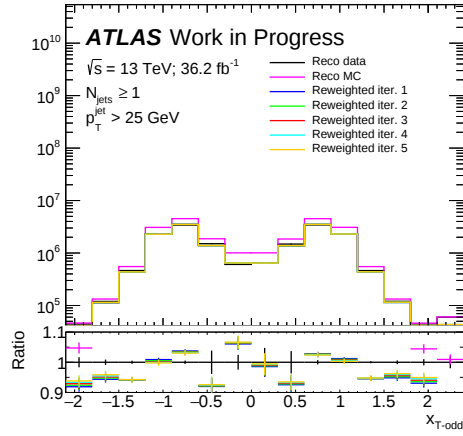
As the p_{T}^W and $p_{\text{T}}^{\text{leading}}$ distributions would also exhibit a slightly overestimated χ^2 , the number of iterations is chosen to be three, where the χ^2 per data point is 1.1 and 1.2, respectively.



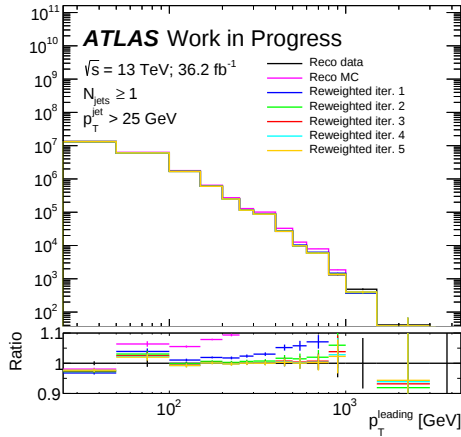
(a)



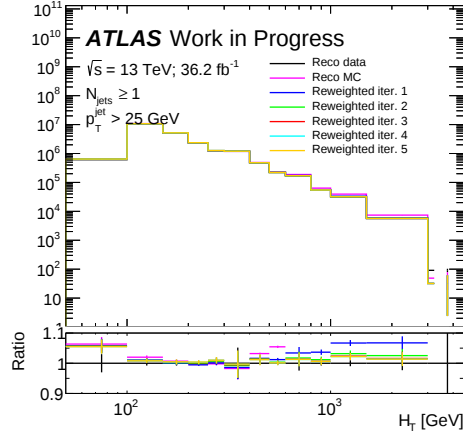
(b)



(c)



(d)



(e)

Figure 6.3: The figures used to determine the number of iterations to use in the iterative Bayesian unfolding procedure. The reconstructed data are shown as a black line and the reweighted reconstructed MC is shown as a series of coloured lines for up to five iterations.

7

Estimation of uncertainties

Contents

7.1	Statistical uncertainties	112
7.2	Systematic uncertainties	113
7.2.1	Electron uncertainties	114
7.2.2	Jet uncertainties	115
7.2.3	E_T^{miss} uncertainties	116
7.2.4	Cross-section uncertainties	116
7.2.5	Multijet-estimation uncertainties	117
7.2.6	Breakdown of total uncertainty	117
7.3	Statistical correlations	117
7.3.1	The bootstrap method	119
7.3.2	Results	120

The estimation of uncertainties is vital for any quantitative measurement. It allows one to declare their confidence that their model suitably represents the data, whether that model be the null hypothesis or some other expectation. If we declare that we have discovered a new phenomenon, we must state quantitatively how confident we are in that result. In this chapter, the method of the estimation of uncertainties for the $W + \text{jets}$ differential cross section measurement is described. The focus is the uncertainties forming the analysis at a centre-of-mass energy of 13 TeV, however a lot of the ideas are applicable to the QCD analysis of the measurements at 7 and 8 TeV discussed in chapter 9.

A range of uncertainties are evaluated and considered together as orthogonal components of an overall total uncertainty. Below, under separate section headers, systematic and statistical components of uncertainty are discussed separately.

7.1 Statistical uncertainties

The statistical uncertainties of this analysis are those which allow us to make a probabilistic statement on the number of events collected. In this analysis, they are related to:

- the statistical precision of the data collected by the ATLAS detector;
- the statistical precision of the background modelling;
- the additional uncertainty implied when a response matrix of limited statistical precision is used for unfolding.

For the data statistics, the distribution of events in each measurement bin is expected to be Poissonian, so the uncertainty of bin i takes the form:

$$\delta_{i,\text{stat}} = \sqrt{N_i}. \quad (7.1)$$

As discussed in chapter 4, the MC is weighted on an event-by-event basis in order to correct for effects which are difficult to model and the distribution is scaled to match the integrated luminosity of the data collected. Similarly, the Matrix Method discussed in chapter 5 involves calculating a weight for each event in order to parameterise the multijet contamination. In these cases, the statistical uncertainty takes a modified form:

$$\delta_{i,\text{MCstat}} = \sqrt{\sum_j^{N_i} w_j^2}, \quad (7.2)$$

where the sum runs over each event j with weight w_j in bin i . This is derived by considering the probability distribution in each bin as the convolution of probability distributions for each event in that bin, each taking a Poissonian form with mean and standard deviation w_j .

Before unfolding, the expected background is subtracted from the data. The statistical uncertainty of this reduced distribution is the quadrature sum of the data uncertainty and the uncertainties of each background component. The uncertainty of the signal MC is not included in the uncertainty of the cross-section measurement, however it is included in a similar way as above through quadrature sum when performing a comparison between the model and the measurement.

Statistical uncertainty of unfolding

As mentioned in chapter 6, the statistical uncertainty of the response matrix and truth prior when unfolding gives rise to a component in addition to the reconstructed uncertainty which is propagated to the unfolded distribution. This is discussed in detail in Ref. [127] and summarised here.

Using the same notation as before, the error propagation matrix is computed as:

$$\frac{\partial N_i^{(T)}}{\partial N_j^{(R)}} = M_{i,j} + \sum_k^{n_R} M_{i,k} N_k^{(R)} \left(\frac{1}{N_i^{(T_0)}} \frac{\partial N_i^{(T_0)}}{\partial N_j^{(R)}} - \sum_l^{n_T} \frac{\epsilon_l}{N_l^{(T_0)}} \frac{\partial N_l^{(T_0)}}{\partial N_j^{(R)}} M_{l,k} \right), \quad (7.3)$$

where n_R is the number of bins in the reconstructed distribution and n_T the same in the particle level (or Truth) distribution. $\frac{\partial N_i^{(T_0)}}{\partial N_j^{(R)}}$ is the error propagation matrix of the previous iteration (0 for the first iteration), so it can be seen from this that the values in the matrix grow with each iteration. The uncertainties themselves are taken from the diagonal elements of the covariance matrix:

$$V(N_k^{(T)}, N_l^{(T)}) = \sum_{i,j=1}^{n_T} \frac{\partial N_k^{(T)}}{\partial N_i^{(R)}} V(N_k^{(R)}, N_l^{(R)}) \frac{\partial N_l^{(T)}}{\partial N_j^{(R)}}. \quad (7.4)$$

7.2 Systematic uncertainties

Recall that chapter 4 discusses the criteria of a $W + \text{jets}$ event, and that the requirement is the reconstruction of an electron, E_T^{miss} and jets. In each case, the measurement is calibrated to give the best possible accuracy of the object kinematics; however, each of these calibrations is based on data, which itself contains statistical uncertainties similar to those discussed above, as well as additional estimated modelling uncertainties which must be propagated through to the final result.

The physical source of each uncertainty is discussed below. Their treatment through to the result is fundamentally the same: for each uncertainty, the analysis is repeated in full with each event reevaluated by varying the kinematic variables or event weights according to that uncertainty, and the difference between this measurement and the nominal measurement is taken as the propagated uncertainty. For example, the uncertainty on the integrated luminosity of the data used for this measurement is 2.1% [99]. To evaluate the effect of this on the results, the analysis is repeated in full with the integrated luminosity varied upward (and downward, separately) by 2.1%. The estimated background subtracted from the data before the unfolding procedure is therefore 2.1% higher (or lower), and when the histograms are divided by integrated luminosity to give a cross section this result is also varied by 2.1%. The resulting histograms deviate from the nominal measurement by slightly more than 2.1%, and this is the reported integrated luminosity uncertainty. Another systematic uncertainty, such as in the measurement of E_T^{miss} , gives different values on the kinematic variables themselves, as opposed simply to alternative normalisations - the distributions themselves could change, as could the number of events passing the selection criteria, both in a non-trivial way.

These systematic uncertainties contain very valuable information. If the correlations between spectra are known, those spectra can be used together in a statistical analysis. For example, if one were to measure a ratio, such as the ratio W^+/W^- , each systematic variation would be divided separately, and many (for example, the integrated luminosity uncertainty, in which both are varied by approximately 2.1%) would give the same result as the nominal measurement. Thus, some uncertainties are cancelled.

7.2.1 Electron uncertainties

Electron uncertainties correspond to uncertainties arising from the reconstruction, identification, trigger and isolation efficiencies as well as the resolution of the energy/momentum measurements. Like many of the uncertainties, these are evaluated in MC simulations and applied to the data. Notably, as the measurement

of the real efficiency used for the matrix method discussed in chapter 5 comes directly from the electron scale factors from MC simulation, it is these systematic uncertainties which correspond to the uncertainty on the real efficiency, thereby contributing to the multijet-background uncertainty. The electron uncertainties comprise of the following components:

- 1 related to the electron resolution,
- 2 related to the electron scale,
- 4 related to scale factors applied to MC simulation - these correspond to the trigger, reconstruction, identification and isolation.

7.2.2 Jet uncertainties

The calibration of jets gives the largest contribution to the systematic uncertainty of this measurement. They are evaluated similarly to the electron uncertainties, and comprise of the following components:

- 17 components corresponding to the fit parameters of the machine-learning algorithm used to determine whether a jet is b -tagged;
- 4 components corresponding to the correction for pileup;
- 8 components related to the nuisance parameters of the jet resolution correction;
- 16 components related to the nuisance parameters of the jet energy scale correction, related to the detector modelling, heavy-flavour scale, statistical uncertainties and jet-modelling uncertainties;
- 5 components related to the η -dependence of the non-closure of jet modelling;
- 5 additional components related to punch-through (particles exiting the calorimeter), the response of a single particle at high p_T , the flavour composition of the jet, the response to the flavour of the detector and the JVT criterion.

This gives a total of 55 components, each with a separate up/down variation. More details of the jet uncertainties are given in Ref. [80], with details of the b -tagging uncertainties given in Ref. [129]. Uncertainties arising from the extrapolation to jets originating from b quarks in the fiducial phase space are not included.

Many analyses evaluate the relative size of each of these uncertainties and remove those which are insignificant from the final result for conciseness - this will be done in this analysis when the full Run 2 data is included.

7.2.3 E_T^{miss} uncertainties

There are a total of three uncertainties related to E_T^{miss} used in this analysis. Each are related to the measurement of the soft term described in section 4.1.3 as systematic uncertainties relating to the hard objects are already evaluated. The first component is measured by varying the soft-term energy scale upward and downward by its uncertainty. The second and third components correspond to the resolution of the soft term either parallel to the direction of the hard term or perpendicular, measured by taking the difference between data and the prediction of Z -boson decay measurements where the E_T^{miss} is expected to be zero. More details of this evaluation are given in Ref. [81].

7.2.4 Cross-section uncertainties

As is already described, the integrated luminosity of the data used for this measurement is known to within 2.1%, and is therefore varied in this analysis in order to propagate a “luminosity uncertainty” to the final result. Similarly, the cross sections of each of the background processes are known only to within a certain confidence interval. This is 5% for each of the background processes, following the evaluation of Refs. [102, 130, 131].

7.2.5 Multijet-estimation uncertainties

In addition to the reevaluation of the multijet background for each systematic variation, uncertainties corresponding to the measurement of the fake efficiency

performed for this analysis must be propagated to the final result.

Ideally, this would involve varying the efficiency in each bin of each spectrum through which this background is parameterised separately, with full consideration of the correlations between spectra, and producing a new measurement of the background. However, this is very computationally expensive and would provide a large number of systematic uncertainties, each contributing only to a small number of bins if contributing at all. Instead, the conservative approach of varying all efficiencies upward/downward simultaneously is taken, giving just a single additional systematic uncertainty.

7.2.6 Breakdown of total uncertainty

The breakdown of the total uncertainty for the measured distributions is shown in fig. 7.1. The largest sources of uncertainty are related to the jet in the majority of bins in all distributions. Many sources of uncertainty show a strong dependence on the p_T of the objects within the event, peaking at higher values. The uncertainties generally display a tendency to increase with higher jet multiplicity, as shown by fig. 7.1(a) and the plots in fig. 7.2, where the breakdown of uncertainties is shown for the p_T^W spectrum over a range of jet multiplicities.

7.3 Statistical correlations

In addition to the systematic correlations between distributions, statistical correlations can be of equal or greater importance when the same data is used to evaluate both. For example, in events with one jet, the jet and W boson tend to balance each other (ie. $p_T^{\text{leading}} = p_T^W$). If both of these spectra were produced from the same data, and they were fit simultaneously in a QCD analysis without information on their correlations, the data would effectively be double-counted and the results biased. In this section, the correlations between spectra are evaluated and presented.

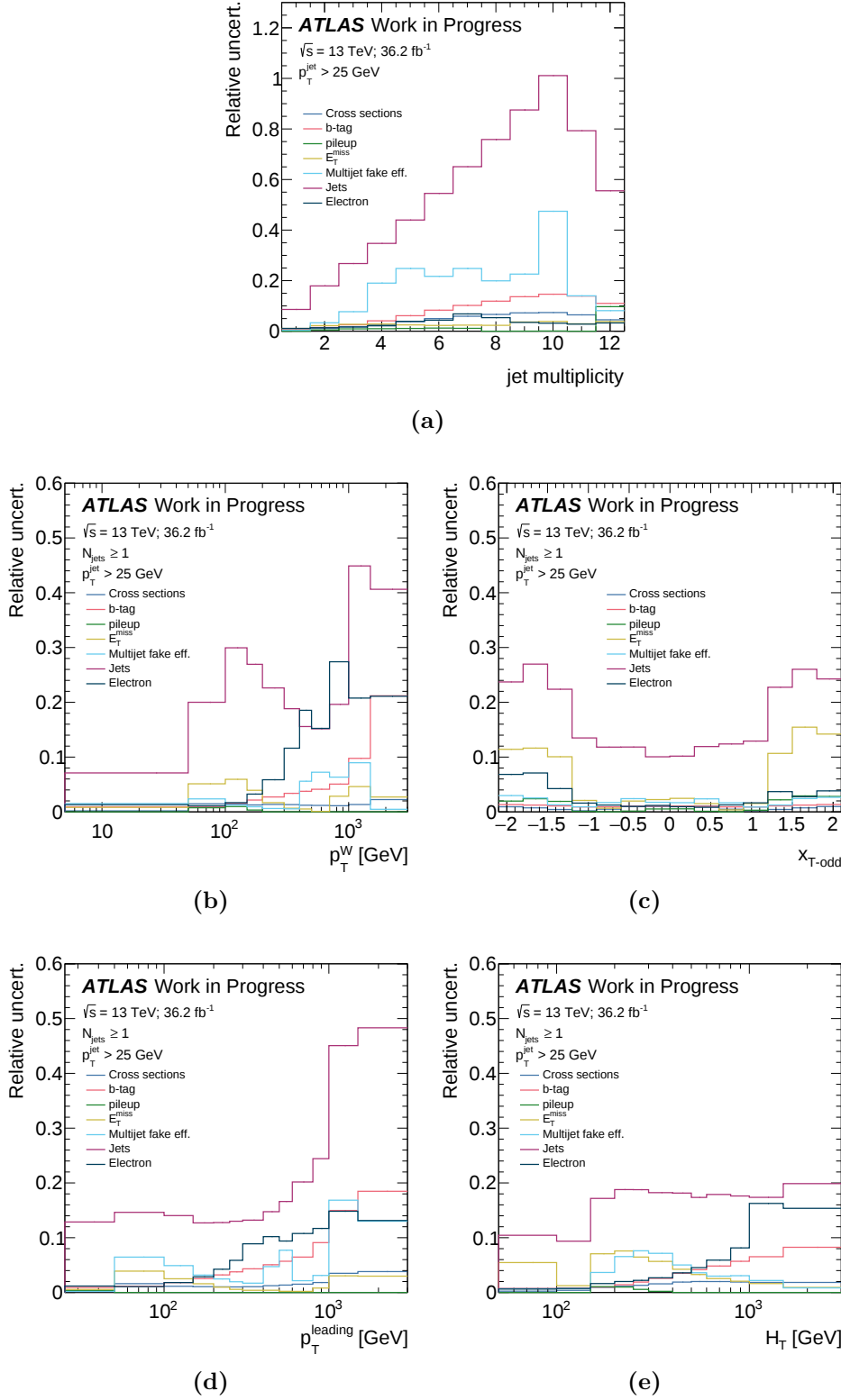


Figure 7.1: The breakdown of fractional uncertainties for the (a) jet multiplicity, (b) p_T^W , (c) $x_{T\text{-odd}}$, (d) p_T^{leading} and (e) H_T distributions. In each case, the uncertainties are grouped by cross-section variations (including integrated luminosity), b -tagging, pileup, E_T^{miss} , the variation of the fake efficiency in the multijet background estimation, jets and electron. The quadrature sum of each group is displayed separately.

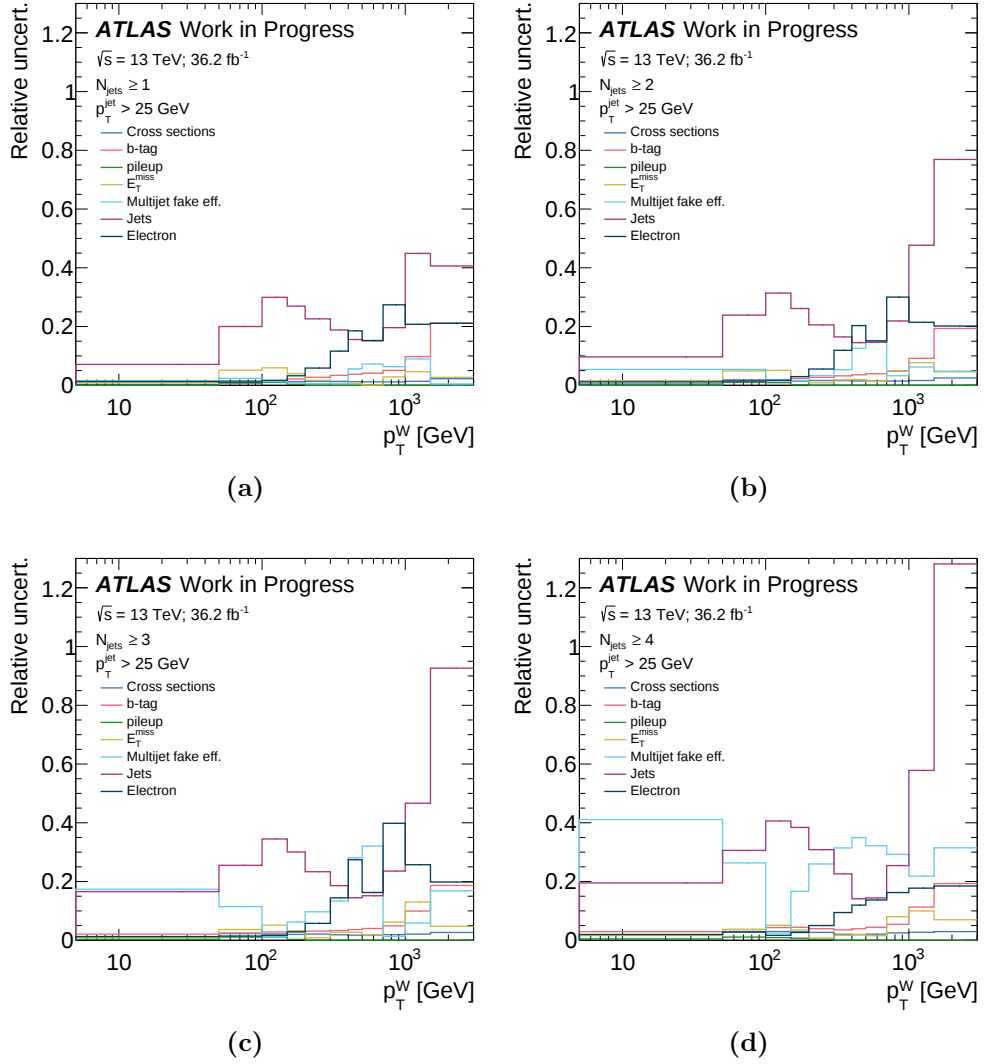


Figure 7.2: The breakdown of fractional uncertainties for the p_T^W spectrum for inclusive jet multiplicities ranging between one (a) and four (d). In each case, the uncertainties are grouped by cross-section variations (including integrated luminosity), b -tagging, pileup, E_T^{miss} , the variation of the fake efficiency in the multijet background estimation, jets and electron. The quadrature sum of each group is displayed separately.

7.3.1 The bootstrap method

The “bootstrap” method is a form of statistical analysis used to calculate any non-trivial statistical distribution. It is implemented here as the correlation between bins of histograms produced from events cannot be calculated from the histograms themselves - the information on how a bin in histogram A is related to another bin in histogram B is lost in the process of filling the histograms.

Sample	Weights
Data	[1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1]
Pseudo-data 1	[3, 1, 2, 0, 1, 0, 2, 1, 1, 0]
2	[1, 3, 0, 1, 0, 0, 0, 1, 3, 2]
3	[2, 2, 2, 1, 0, 0, 0, 0, 0, 1]
4	[1, 2, 2, 0, 1, 1, 0, 0, 2, 0]
5	[1, 0, 2, 1, 1, 0, 1, 0, 1, 4]
6	[2, 0, 0, 0, 1, 2, 0, 3, 0, 1]

Table 7.1: An example of pseudo-data weights produced for a measurement of ten events using the bootstrap method.

The method involves repeating the measurement a number of times, n , using a different set of pseudo-data each time. The pseudo-data is produced by varying the number of times each event is recorded by a Poisson distribution of mean one. A simple example is shown in table 7.1. Given that the weights are produced on an event-by-event basis, the replicated spectra are fully synchronised. Accordingly, one can proceed to calculate the statistical properties of that measurement - including the mean, standard deviation, and of course correlations. The statistical correlations are evaluated bin-by-bin in the following way:

$$C_{i,j}^{A,B} = \frac{\frac{1}{N} \sum_{k=1}^N (M_i^{A,k} - \mu_i^A)(M_j^{B,k} - \mu_j^B)}{\sigma_i^A \sigma_j^B} \quad (7.5)$$

where $C_{i,j}^{A,B}$ is the correlation between bins i and j of spectra A and B , μ_i^A and σ_i^A are the mean and standard deviation of bin i in spectrum A , respectively; and $M_i^{A,k}$ is the measurement of pseudo-data k in bin i of spectrum A . The number of replicas used in this analysis has been set to $N = 10^4$.

7.3.2 Results

The most simple sanity check one can perform to establish that the correlations are calculated correctly is to show the correlation matrix between a spectrum and itself. The diagonal elements should be one by definition, with minor fluctuations in the off-diagonal elements owing to the limited number of pseudo-data samples. This is shown in fig. 7.3 for each of the analysed distributions. In each case,

the diagonal elements of the matrix are at unity and all off-diagonal elements are close to zero, as expected.

As there are four differential distributions, there are a total of six correlation matrices per jet multiplicity. The matrices for an inclusive jet multiplicity of one are shown in figs. 7.4 and 7.5.

The p_T^W distribution exhibits a clear positive correlation with both the p_T^{leading} and H_T distributions, as do the p_T^{leading} and H_T distributions with each other. While this does limit the additional constraining power which can be obtained by using multiple spectra simultaneously in an analysis such as that of chapter 9, most individual bins are correlated only up to a value of approximately 0.4, implying that there is still a lot of potential for using these measurements together.

The $x_{T\text{-odd}}$ distribution exhibits some correlation with the other spectra for low- p_T events in the range $0.5 < x_{T\text{-odd}} < 1.2$. This is the range of values expected to be most sensitive to the naïve T-odd asymmetry as described in Ref. [101], implying that it is typically lower-energy interactions which will contribute to that measurement.

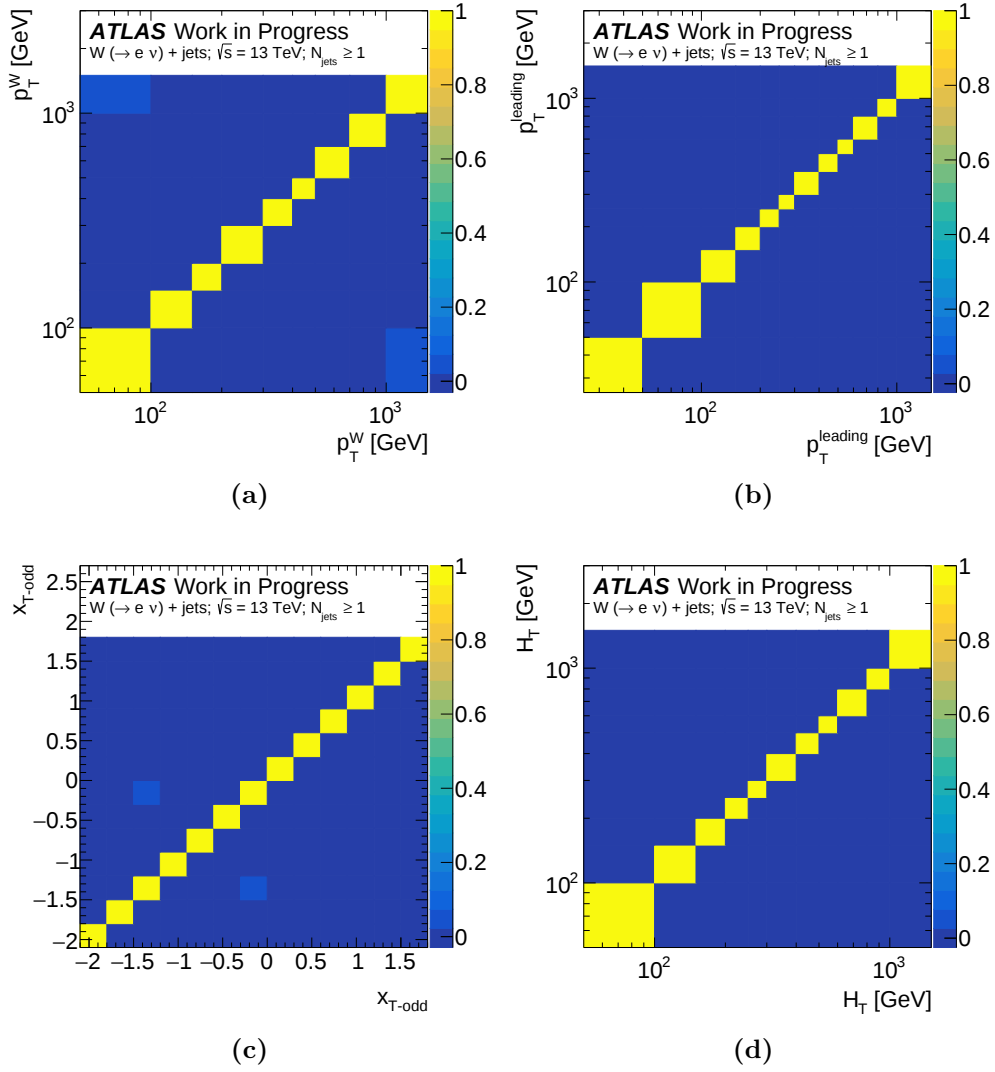


Figure 7.3: The correlation matrices evaluated for each of the (a) p_T^W , (b) p_T^{leading} , (c) $x_{T\text{-odd}}$ and (d) H_T distributions with themselves.

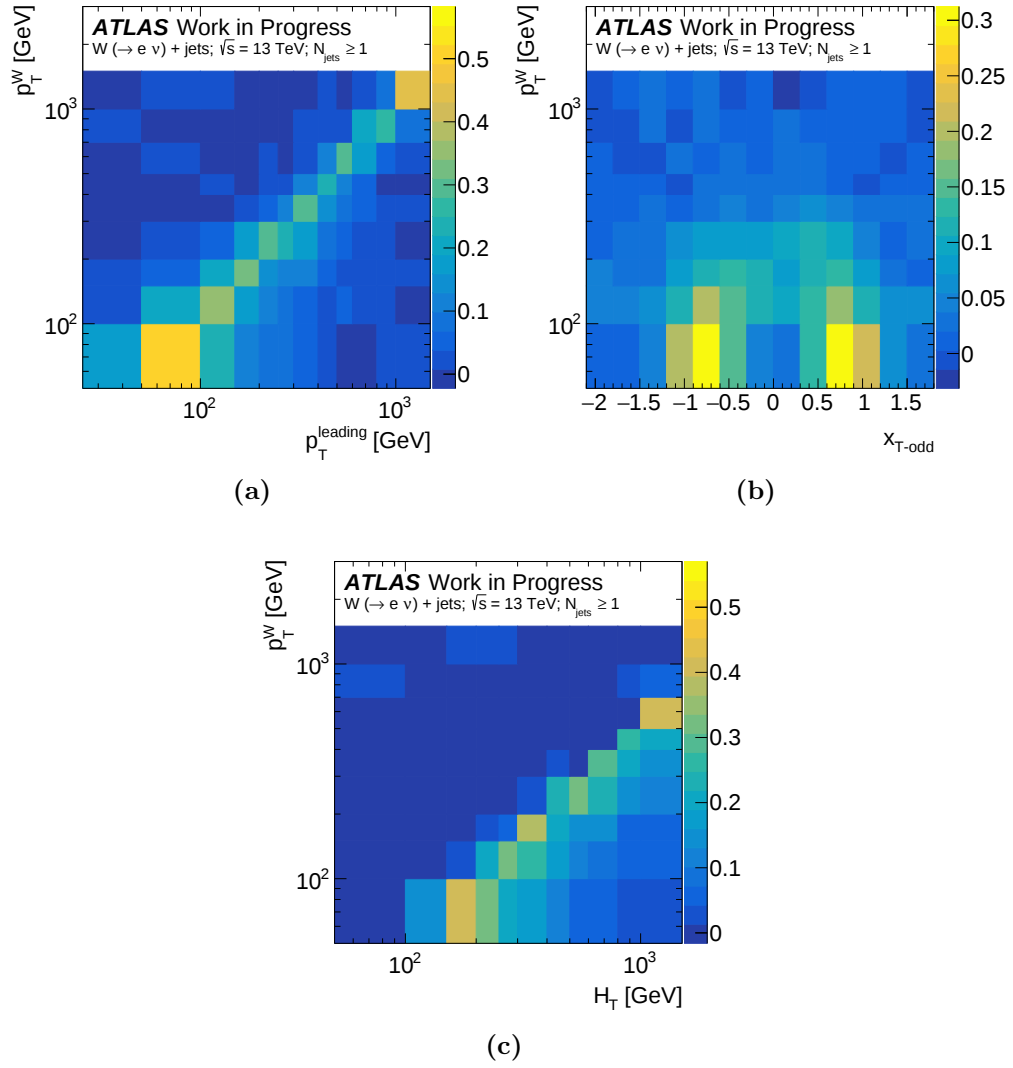


Figure 7.4: The correlation matrices between distributions (a) p_T^W and p_T^{leading} , (b) p_T^W and $x_{T\text{-odd}}$, and (c) p_T^W and H_T .

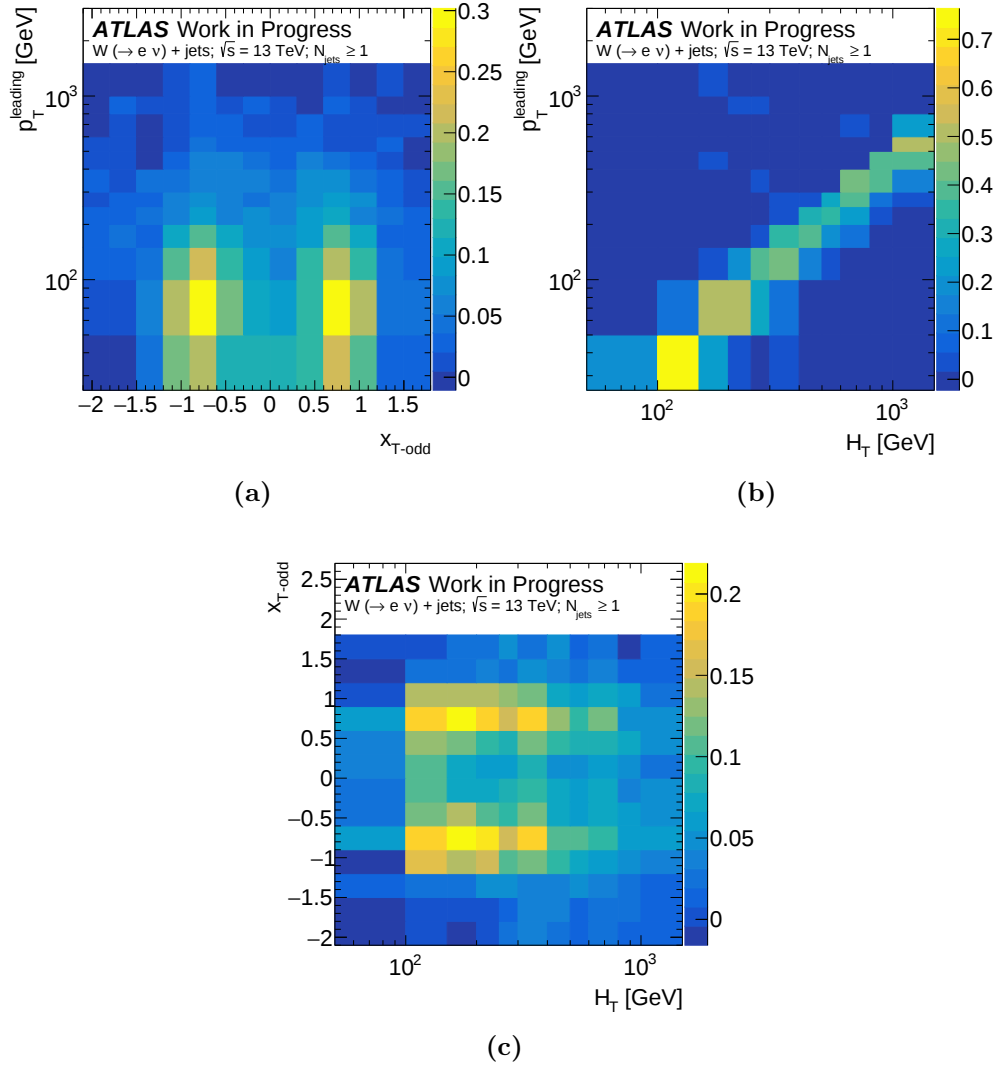


Figure 7.5: The correlation matrices between distributions (a) $p_{\text{T}}^{\text{leading}}$ and $x_{\text{T-odd}}$, (b) $p_{\text{T}}^{\text{leading}}$ and H_{T} , and (c) $x_{\text{T-odd}}$ and H_{T} .

8

W + jets cross section results

Finally, everything in the previous chapters culminates in a measurement of the differential cross section of the production of a W boson in association with jets in pp collisions at $\sqrt{s} = 13$ TeV. This is performed in the kinematic space most sensitive to these events as a function of multiple observables, as discussed in chapter 4, with the background contributions estimated as in chapter 5, results deconvoluted from the effect of the detector, as discussed in chapter 6, and a wide range of uncertainties as discussed in chapter 7.

Each of the observables evaluated are shown for the charge-independent distributions and split in to contributions from W^+ and W^- separately, along with the ratio of the two, in figs. 8.1 to 8.5.

The results for the charge-independent cross section are similar to those presented in the comparison between data and the full SM prediction in chapter 5. For high jet multiplicities, the SHERPA2.2.1 MC simulation tends to overestimate the cross section, whereas the MG + PY8 simulation tends to underestimate. Both generators deviate from the data in regions of all measured spectra outside of the statistical uncertainties, though they are consistently within systematic uncertainties except for at high $p_{\text{T}}^{\text{leading}}$ where the MG + PY8 simulation significantly overestimates the cross section, as it did in the comparison between reconstructed data and MC simulation in chapter 5. The charge-independent cross sections for each inclusive

and exclusive jet multiplicity are given in tables 8.1 and 8.2 for up to eleven jets. At each jet multiplicity, the precision of the measurement of the cross section is limited mainly by the size of the systematic uncertainties.

In all cases, a similar deviation from the data that is seen in the charge-independent cross sections is observed in both the W^+ and W^- cross section for both generators, giving a W^+/W^- ratio which is in good agreement between the generators. These highly consistent ratio predictions emerge also at higher jet multiplicities, as shown for example in fig. 8.6 for the p_T^W spectrum up to five inclusive jets. Cancellation of the systematic uncertainties results in a ratio for which the statistical uncertainty plays a far greater role than in the individual cross sections themselves. Compared to the data, for the N_{jet} and $x_{T\text{-odd}}$ distributions, there is good agreement within total uncertainties across the whole spectrum. However, in the p_T^W , p_T^{leading} and H_T spectra, there is a systematic overestimation outside of total uncertainties for values greater than approximately 200 GeV. This may have strong implications in terms of constraints on the structure of the proton as both generators use the NNPDF PDF set (albeit different versions). It is possible that, at high- x , the ratio of the up and down valence quarks, which strongly contributes to this distribution, is too high to accurately describe this data, but further analysis is required to determine the precise cause of this discrepancy.

Additional unfolded distributions with full uncertainties for higher jet multiplicities are given in appendix B. The use of the full Run 2 data collected from the years 2015–2018 will provide a total of approximately 140 fb^{-1} of pp collision data. As the ratios are statistically limited, this will provide great improvements to the measurements. Further, a similar measurement in the $W \rightarrow \mu\nu$ decay channel is expected, providing extra constraining power.

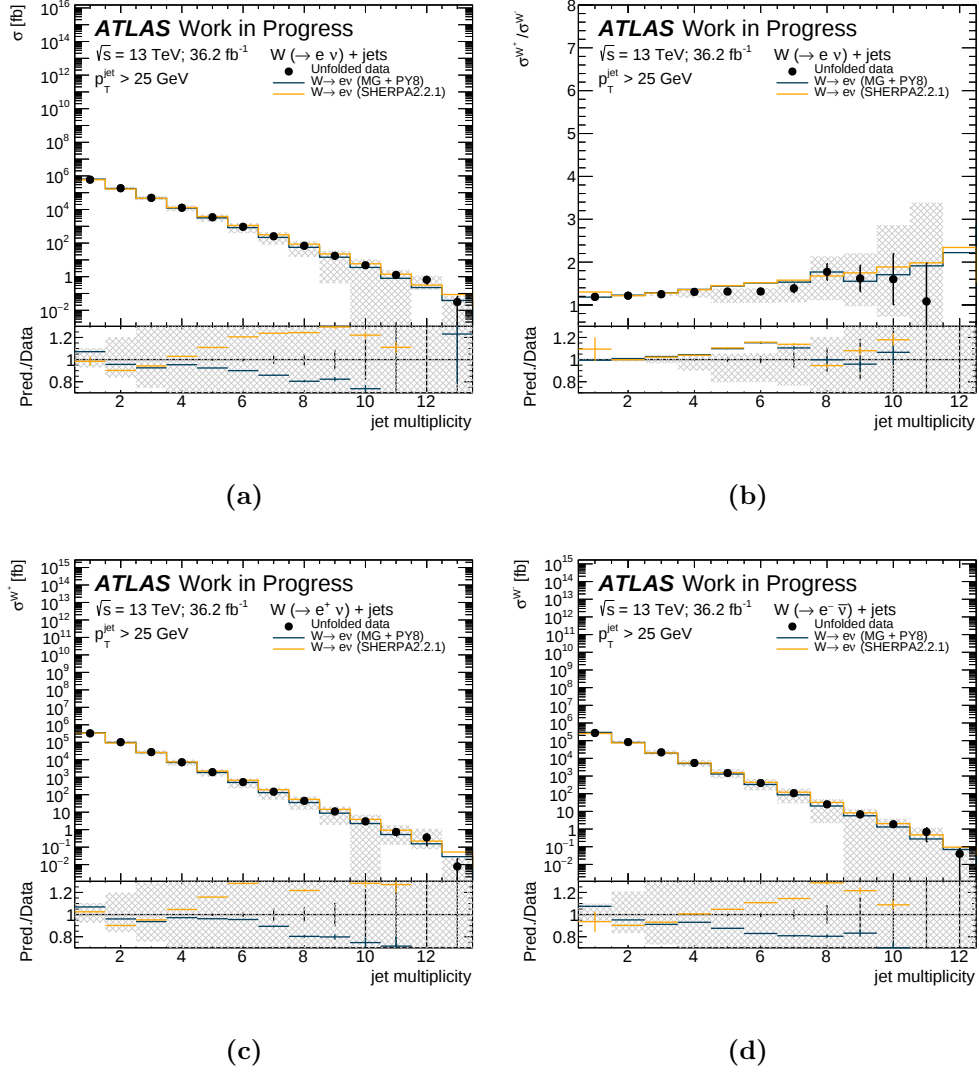


Figure 8.1: The unfolded cross section for the production of a W boson in association with at least one jet for the N_{jet} observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.

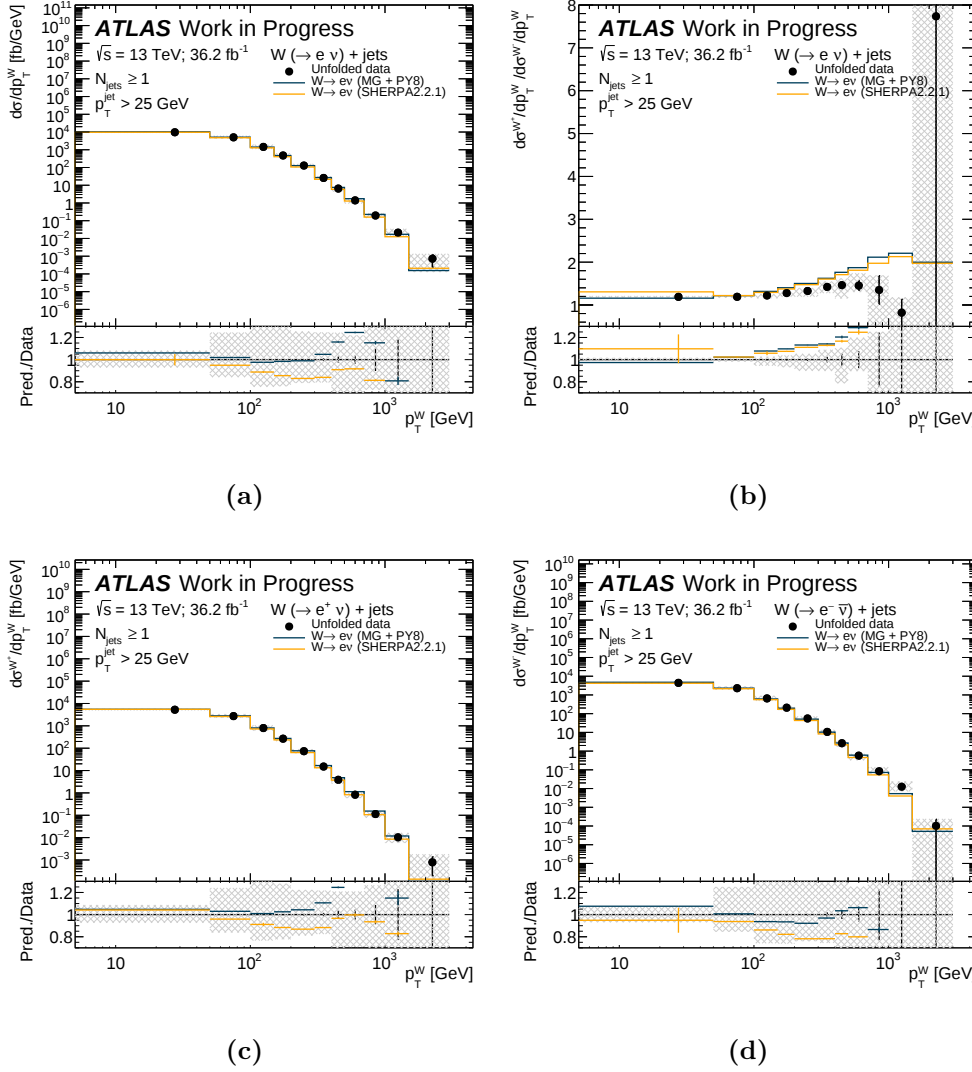


Figure 8.2: The unfolded cross section for the production of a W boson in association with at least one jet for the p_T^W observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.

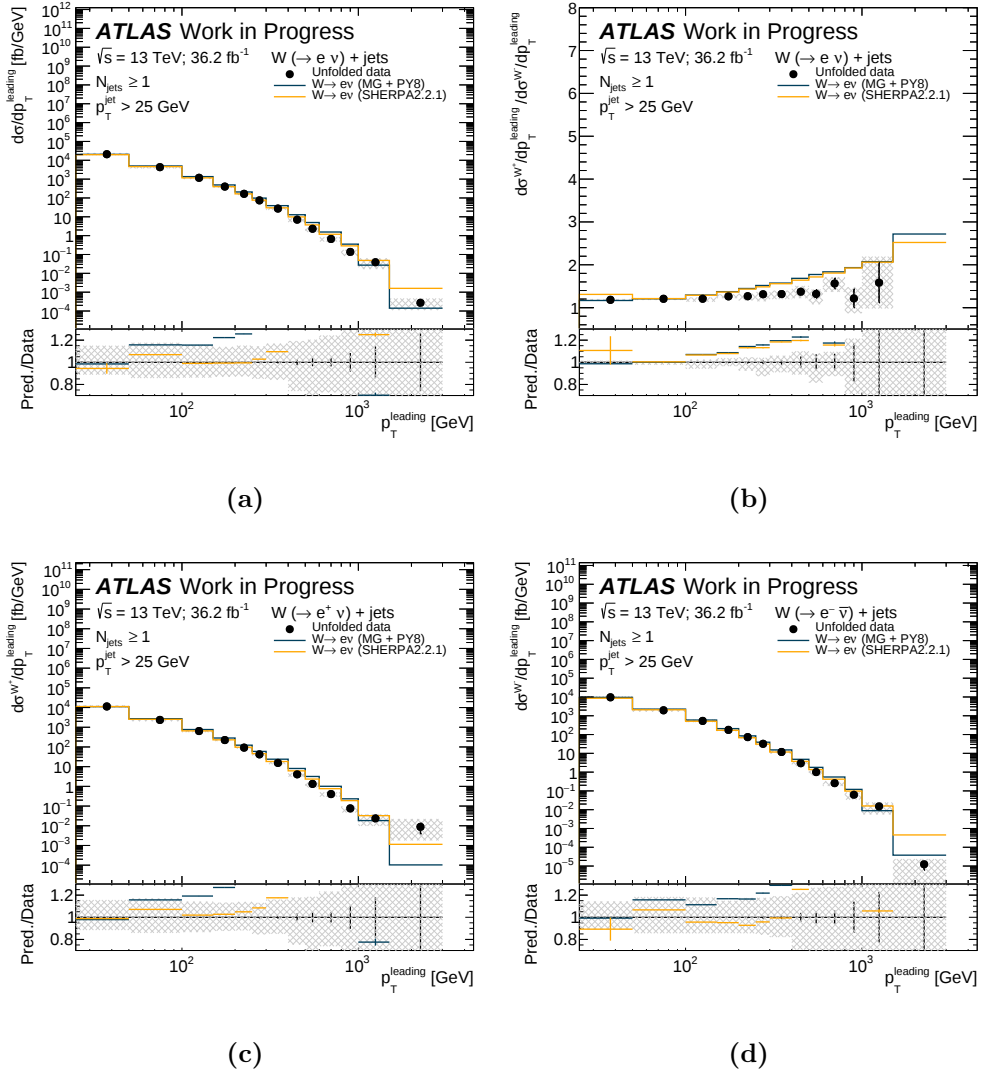


Figure 8.3: The unfolded cross section for the production of a W boson in association with at least one jet for the p_T^{leading} observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.

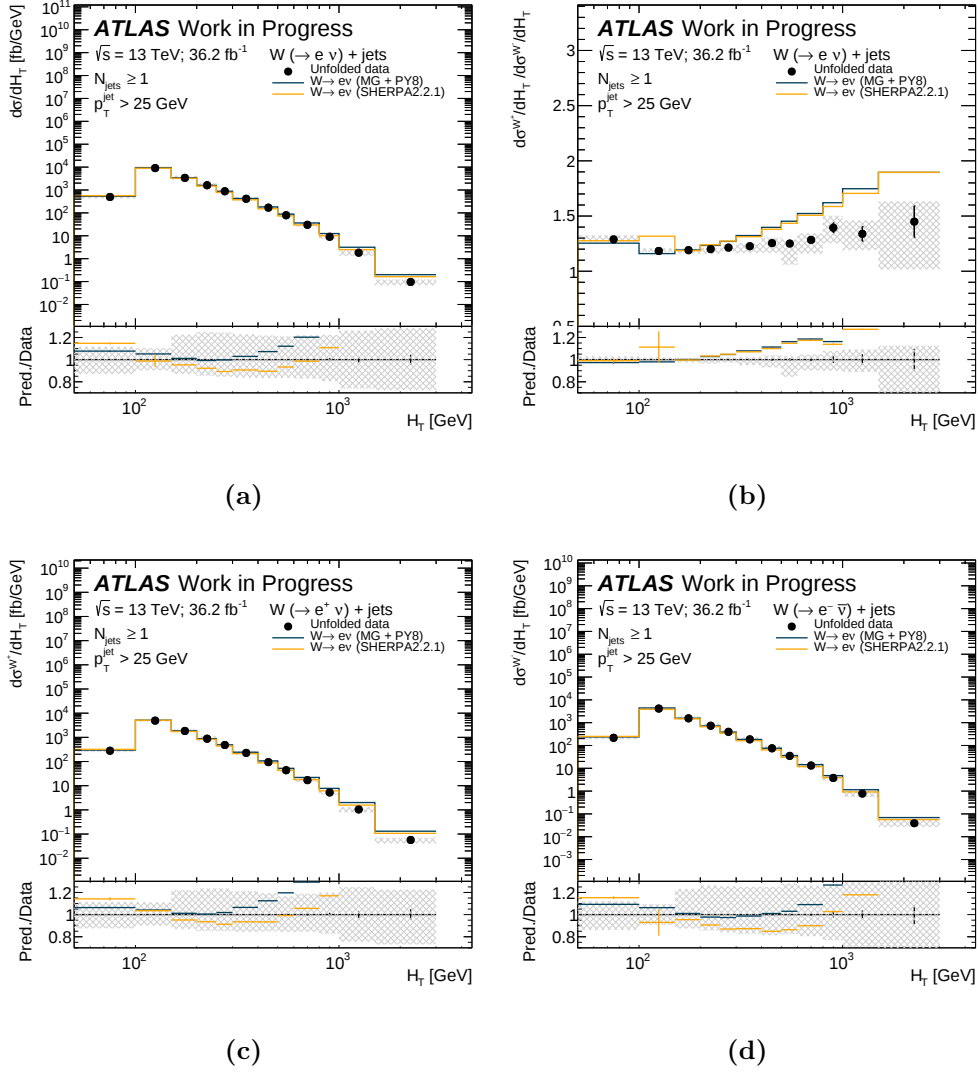


Figure 8.4: The unfolded cross section for the production of a W boson in association with at least one jet for the H_T observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.

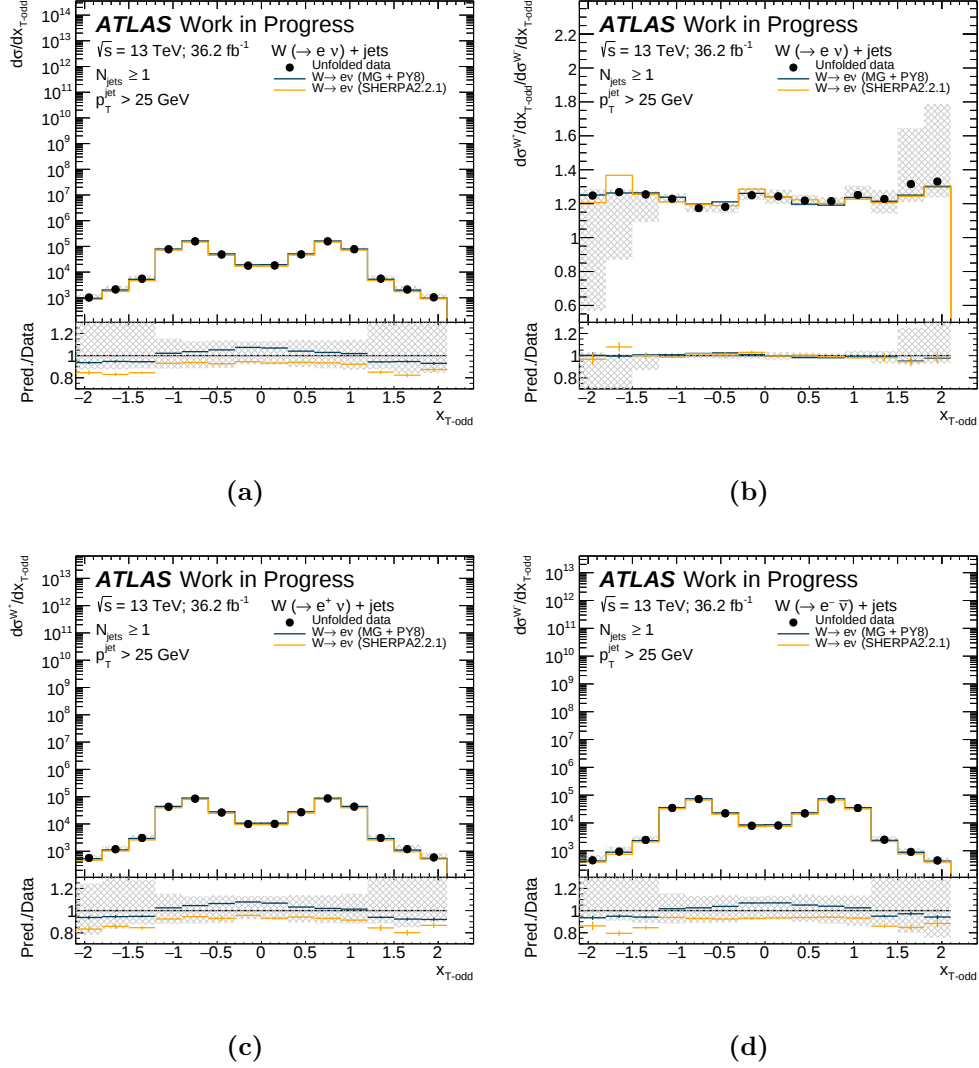


Figure 8.5: The unfolded cross section for the production of a W boson in association with at least one jet for the x_{T-odd} observable for the (a) change-independent measurement, (b) ratio of the W^+ and W^- cross sections, (c) W^+ cross section and (d) W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.

Jet multiplicity	Inclusive cross section [fb]	Stat. unc. (%)	Syst. unc. up (%)	Syst. unc. down (%)
1	849194.871	0.1	12.6	12.6
2	251194.434	0.1	22.0	21.7
3	66582.403	0.2	32.2	31.4
4	17387.434	0.4	43.9	42.3
5	4712.779	0.8	54.5	52.9
6	1276.441	1.4	63.6	64.9
7	348.178	3.0	74.2	76.1
8	93.720	4.4	83.3	87.3
9	24.290	7.7	93.5	102.3
10	6.770	15.3	98.6	103.3
11	1.953	28.4	70.7	92.3

Table 8.1: The measured charge-independent cross section of the production of a W boson in association with jets in pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV for inclusive jet multiplicities up to eleven.

Jet multiplicity	Exclusive cross section [fb]	Stat. unc.	Syst. unc. up	Syst. unc. down
1	598000.438	0.1	8.8	8.9
2	184612.031	0.1	18.5	18.3
3	49194.969	0.2	28.2	27.6
4	12674.655	0.5	40.1	38.3
5	3436.338	1.0	51.2	48.6
6	928.263	1.6	59.7	60.7
7	254.458	3.7	71.0	72.0
8	69.430	5.3	79.8	82.2
9	17.520	8.9	91.8	102.0
10	4.817	18.2	113.0	110.4
11	1.267	36.9	82.8	106.8

Table 8.2: The measured charge-independent cross section of the production of a W boson in association with jets in pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV for exclusive jet multiplicities up to eleven.

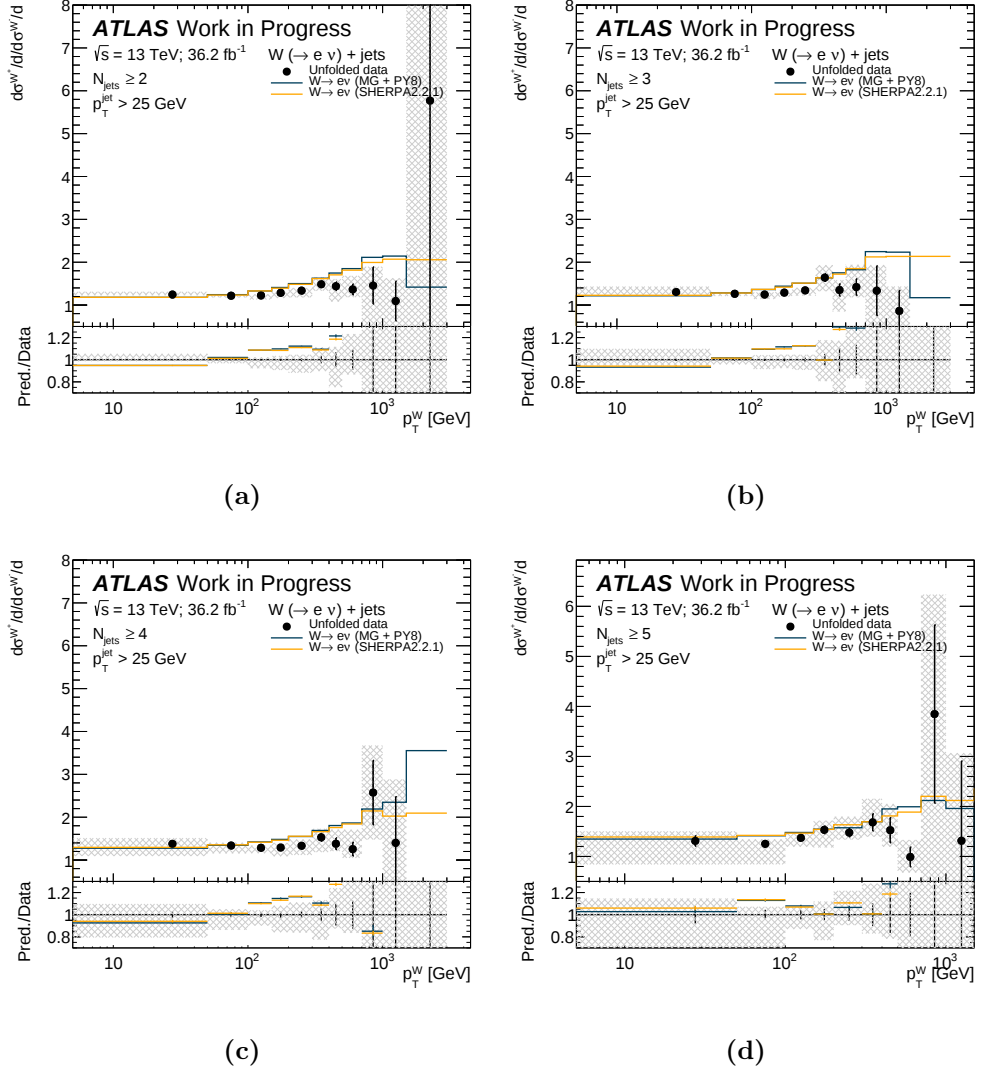


Figure 8.6: The unfolded cross section for the production of a W boson in association with jets for the p_T^W observable for at least (a) two jets, (b) three jets, (c) four jets and (d) five jets. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.

9

Parton Distribution Functions with $V +$ jets

Contents

9.1	Introduction	136
9.2	Input data sets	140
9.3	Fit framework	142
9.4	Results	146
9.4.1	Goodness of fit and parton distributions	146
9.4.2	The high- x sea-quark distributions	152
9.4.3	Strange-quark density	154
9.4.4	Comparison with global PDFs	156

The previous chapters have shown an analysis in which the $W +$ jets cross section was measured for pp collisions at $\sqrt{s} = 13$ TeV at the ATLAS detector. A similar measurement was performed already at $\sqrt{s} = 8$ TeV [100], along with a double-differential measurement of the production of a Z boson in association with jets (“ $Z +$ jets”) at the same centre-of-mass energy [132]. In this section, a NNLO QCD analysis is presented in which the PDFs of the proton are measured using this data alongside precision measurements of W and Z boson production at $\sqrt{s} = 7$ TeV and deep-inelastic-scattering (DIS) data from $e^\pm p$ collisions at the HERA accelerator. An improved determination of the sea-quark densities at medium-to-high Bjorken x

is shown in comparison to previous ATLAS PDF sets, 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. In addition, the use of a χ^2 scan is implemented for the first time in an ATLAS PDF fit, giving new insight on old data and establishing the consistency between alternative ATLAS data sets.

This analysis was performed by the author of this thesis in its entirety, though cross checks of the central distributions have been made by other analysts using alternative frameworks. The author also served as the writer and primary editor for the publication [133]. This chapter follows the content of the publication closely.

9.1 Introduction

As discussed in section 2.2, precise knowledge of the content of colliding protons, the PDFs, is a necessary ingredient for accurate predictions of both SM and BSM cross sections at the 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 better constrain the x -dependence and flavour decomposition of the PDFs. While 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 [134] of neutral current (NC) and charged current (CC) cross sections in $e^\pm p$ scattering constrain PDFs in a number of ways. The low Q^2 NC cross sections give precise information about the low- x sea quarks and antiquarks, while information about valence quark PDFs in the high- x range is provided by the difference between the NC e^+p and e^-p cross sections and CC scattering, both at high Q^2 . The HERA inclusive DIS data alone

provide sufficient information to give the PDF set referred to as HERAPDF2.0 [134]. 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 [135–138] 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 $W^\pm$ 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 typically used for these constraints is in a kinematic region where non-perturbative effects are important and must be computed from phenomenological models [139, 140], such as higher-twist contributions which are visualised as ladders of recombining gluons [141]. 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 [142] and DØ [143] collaborations, discussed in further detail by the MSTW group in Ref. [144].

Precision measurements from the ATLAS detector at the LHC, together with data from the HERA experiments, have been interpreted previously in a NNLO QCD analysis, resulting in the ATLASepWZ16 PDF set [145]. Differential W and Z/γ^* boson 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 [146]. Figures from the publication depicting the key results are given in fig. 9.1.

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 [147]. However, a recent analysis of CMS $W + c$ 7 and 13 TeV data [148] 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 data alone [149], as shown by fig. 9.2. The tree-level production modes of the production of a vector boson in association with jets, referred to as $V + \text{jets}$, are either quark–antiquark initial states with gluon radiation, or quark–gluon initial states, depicted in fig. 9.3. The process is therefore already sensitive to the gluon density of the proton at lowest order in 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 data set with additional constraining power on top of the inclusive W, Z measurements.

This chapter presents a PDF analysis including data on $W^\pm + \text{jets}$ and $Z + \text{jets}$ production collected in pp collisions at $\sqrt{s} = 8$ TeV by the ATLAS collaboration [100, 132] in combination with the previous inclusive W, Z measurement at $\sqrt{s} = 7$ TeV [145] and the combination of inclusive combined HERA data [134]. 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 [150, 151], and accounts for the systematic correlations between data sets. The resulting PDF set is called *ATLASepWZVjet20*.

9.2 Input data sets

The final combined $e^\pm p$ cross-section measurements at HERA [134] cover the kinematic range of Q^2 from 0.045 GeV² to 50 000 GeV² and of Bjorken x from 0.65 down to 6×10^{-7} . Data below $x = 10^{-5}$ are excluded from this analysis by requiring $Q^2 > 10$ GeV², motivated by the previously observed poorer fit quality in the

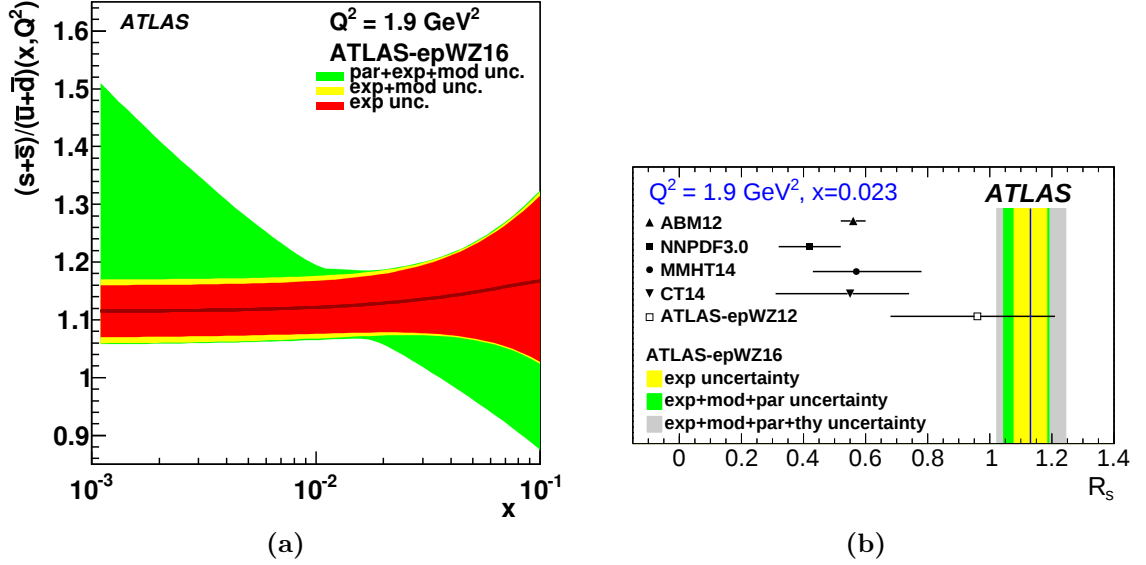


Figure 9.1: The ratio of the strange to up and down sea distributions ($R_s = (s+\bar{s})/(\bar{u}+\bar{d})$) of the ATLASepWZ16 PDFs evaluated at $Q^2 = 1.9 \text{ GeV}^2$ shown (a) as a function of Bjorken x and (b) evaluated at $x = 0.023$ and compared to a range of global PDF analyses [145].

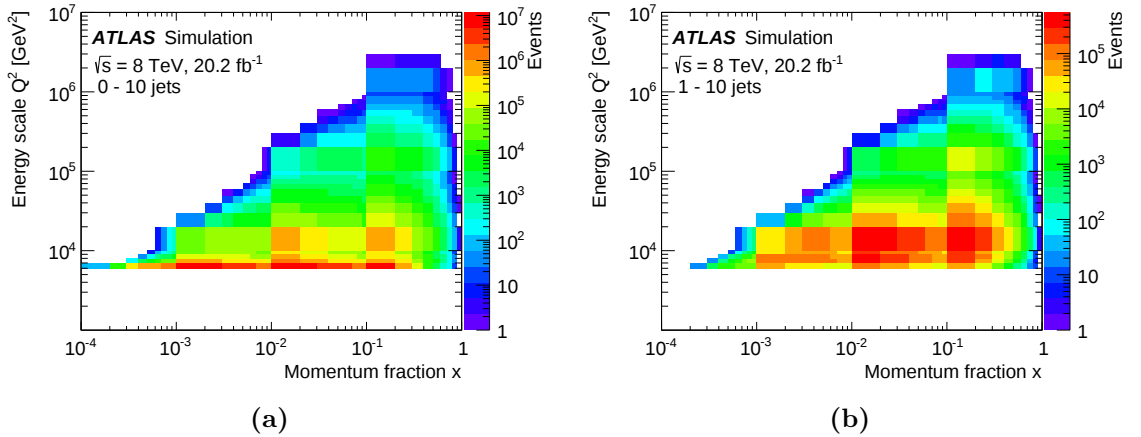


Figure 9.2: The concentration of $W + \text{jets}$ events on the x - Q^2 plane for (a) zero jets inclusive and (b) one jet inclusive [100].

excluded kinematic region compared to the rest of the HERA data [134]. Possible explanations for this include the need for resummation corrections at low x [152] 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 $10 < Q^2 < 500 \text{ GeV}^2$ and below 3% up to $Q^2 = 3000 \text{ GeV}^2$.

The ATLAS W and Z differential cross sections are based on data recorded

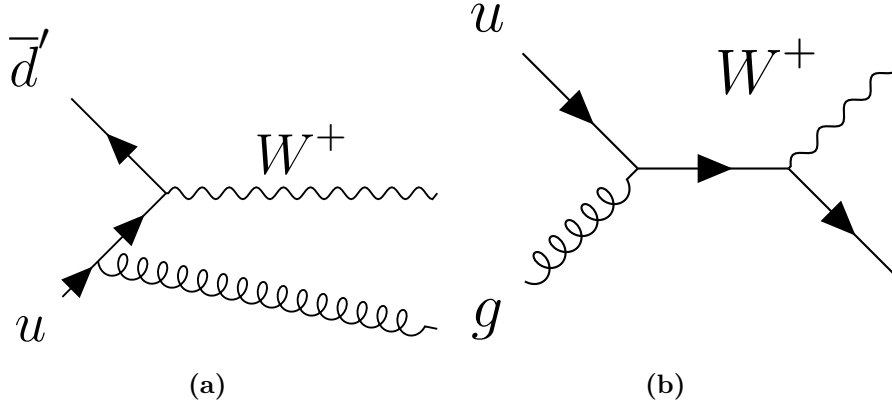


Figure 9.3: Examples of the leading-order production modes of the production of a W boson in association with jets.

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 [145]. 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 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 [145]. 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 $W^\pm + \text{jets}$ 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 [100]. 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 [84, 85] 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% [100]. 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 $W + \text{jets}$ spectrum.

The ATLAS $Z + \text{jets}$ 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 [132]. 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 $25 \text{ GeV} < p_T^{\text{jet}} < 1050$ 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 related to the data and background MC simulation statistics.

The integrated luminosity of the data set used for the $W + \text{jets}$ and $Z + \text{jets}$ 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 and cross checks of alternative models are given in appendix D.

9.3 Fit framework

This determination of proton PDFs uses the xFitter framework, v2.0.1 [134, 153, 154]. This program interfaces to theoretical calculations directly or uses fast

interpolation grids to make theoretical predictions for the considered processes. The program MINUIT [155] is used for the minimisation of the PDF fit. The results are cross-checked with an independent fit framework [156].

For the DIS processes, coefficient functions with massless quarks are calculated at NNLO as implemented in QCDNUM v17-01-13 [157]. The contributions of heavy quarks are calculated in the general-mass variable-flavour-number scheme of Refs. [158–160]. The renormalisation and factorisation scales for the DIS processes are taken as $\mu_R = \mu_F = \sqrt{Q^2}$.

For the proton-proton collision processes, state-of-the-art fast calculations are performed only to NLO in QCD and LO in EW. In order to compare higher-order predictions to the data, these calculations are corrected using multiplicative K -factors. These are calculated as a function of the fitted spectra, ie. there is a different K -factor for each bin, but are independent of the input PDFs.

For the differential W and Z boson cross sections, the theoretical framework is the same as that used in the ATLASepWZ16 analysis of Ref. [145]. The xFitter package uses outputs from the APPLGRID code [161] interfaced to the MCFM program [162, 163] 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 the 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 [164, 165].

Predictions for $W + \text{jets}$ and $Z + \text{jets}$ 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 $W + \text{jets}$ data, the N_{jetti} program [150] 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 [107, 166, 167]. 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 $W + \text{jets}$ data, multiplied by the non-perturbative correction. The renormalisation and factorisation scales are set to $\mu_R = \mu_F = \sqrt{m_W^2 + \Sigma(p_T^j)^2}$,

where m_W is the mass of the W boson and the second term in the square root is the scalar sum of the squared transverse momenta of the jets. More details about the predictions are given in the respective ATLAS publication [100]. 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. [107, 166, 167] and applied as additional bin-by-bin multiplicative K -factors.

Predictions for $Z + \text{jets}$ production to NNLO in QCD and LO in EW couplings are calculated by the authors of Ref. [151], and the K -factor is calculated as the ratio of NNLO to NLO predictions. The renormalisation and factorisation scales are set to $\mu_R = \mu_F = \frac{1}{2}(\sum p_{T,\text{partons}} + \sqrt{m_{\ell\ell}^2 + p_{T,\ell\ell}^2})$ where $m_{\ell\ell}$ is the electron-pair invariant mass, $p_{T,\ell\ell}$ is the transverse momentum of the electron pair and $\sum p_{T,\text{partons}}$ 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 [132], 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 $W + \text{jets}$ and $Z + \text{jets}$ production are typically within 10% of unity, except for the NLO EW corrections for the $W + \text{jets}$ predictions, which are as large as 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 [134, 168], $m_c = 1.43 \text{ GeV}$ 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 [134].

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 [134, 145]

$$xq_i(x) = A_i x^{B_i} (1-x)^{C_i} P_i(x), \quad (9.1)$$

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

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

where C'_g is fixed to a value of 25 to suppress negative contributions from the primed term at high x , as in Ref. [144]. 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 $x \rightarrow 0$. The strange PDF $x\bar{s}$ is parameterised as in eq. (9.1), with $P_{\bar{s}} = 1$ and $B_{\bar{s}} = B_{\bar{d}}$, leaving two free parameters for the strange PDF, $A_{\bar{s}}$ 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. [134]. For the ATLASepWZVjet20 fit, this results in the usage of two additional parameters: E_{u_v} 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 χ^2 . The definition of the χ^2 without statistical correlations between data points is as follows [134, 145]

$$\chi^2 = \sum_i \frac{[D_i - T_i(1 - \sum_j \gamma_{ij} b_j)]^2}{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i} + \sum_j b_j^2 + \sum_i \log \frac{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i}{\delta_{i,\text{uncor}}^2 D_i^2 + \delta_{i,\text{stat}}^2 D_i^2}, \quad (9.2)$$

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 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 systematic uncertainties. 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 $W + \text{jets}$ 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 [169]

$$\begin{aligned} \chi^2 = & \sum_{ik} \left(D_i - T_i \left(1 - \sum_j \gamma_{ij} b_j \right) \right) C_{\text{stat},ik}^{-1}(D_i, D_k) \left(D_k - T_k \left(1 - \sum_j \gamma_{kj} b_j \right) \right) \\ & + \sum_j b_j^2 \\ & + \sum_i \log \frac{\delta_{i,\text{uncor}}^2 T_i^2 + \delta_{i,\text{stat}}^2 D_i T_i}{\delta_{i,\text{uncor}}^2 D_i^2 + \delta_{i,\text{stat}}^2 D_i^2}, \end{aligned} \quad (9.3)$$

in which the first term of eq. (9.2) 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 , $C_{\text{stat},ik}$.

9.4 Results

In this section, the ATLASepWZVjet20 PDF set is presented and compared with an equivalent fit performed without the $V + \text{jets}$ 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 $Q^2 > 10 \text{ GeV}^2$ 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.

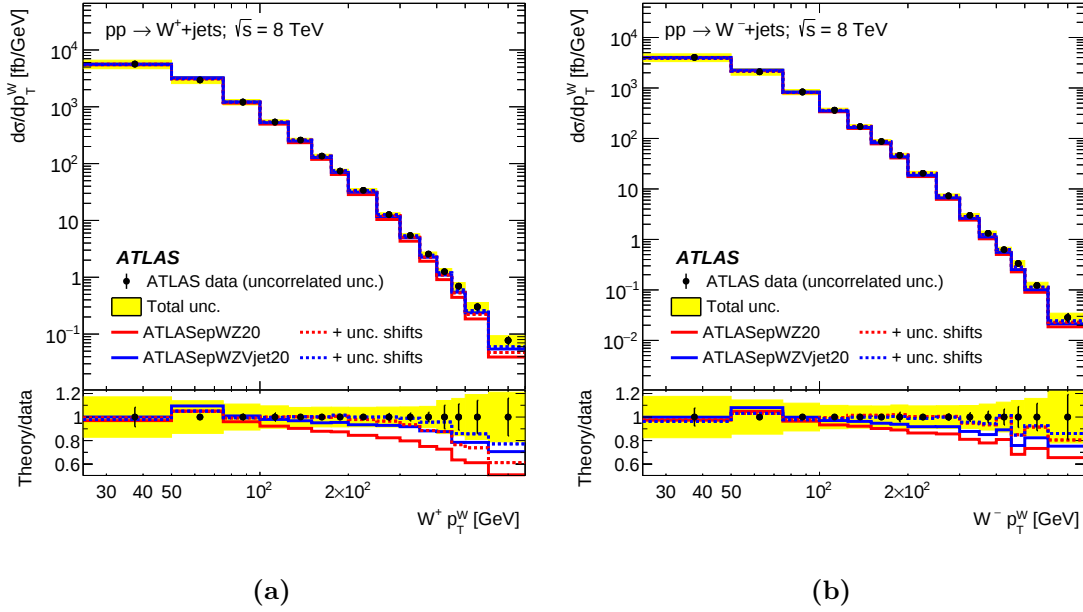


Figure 9.4: The differential cross-section measurements of (a) W^+ + jets and (b) W^- + jets in Ref. [100] (black points) as a function of the transverse momentum of the W boson, p_T^W . The bin-to-bin uncorrelated part of the data uncertainties is shown as black error bars, while the total uncertainties 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. (9.3) are allowed to vary to minimise the χ^2 .

9.4.1 Goodness of fit and parton distributions

Figures 9.4 and 9.5 show a comparison of the W + jets and Z + jets differential cross-section measurements with the predictions of the ATLASepWZ20 and ATLASepWZVjet20 fits. Adding the V + jets data to the fit improves the W + 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 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 (χ^2/NDF) for the ATLASepWZVjet20 fit, along with the partial χ^2 per data point (χ^2/NDP) and correlated χ^2 for each data set entering the fit, is given in table 9.1. The partial χ^2 for the HERA and ATLAS inclusive W and Z data in the ATLASepWZVjet20 fit is similar to those

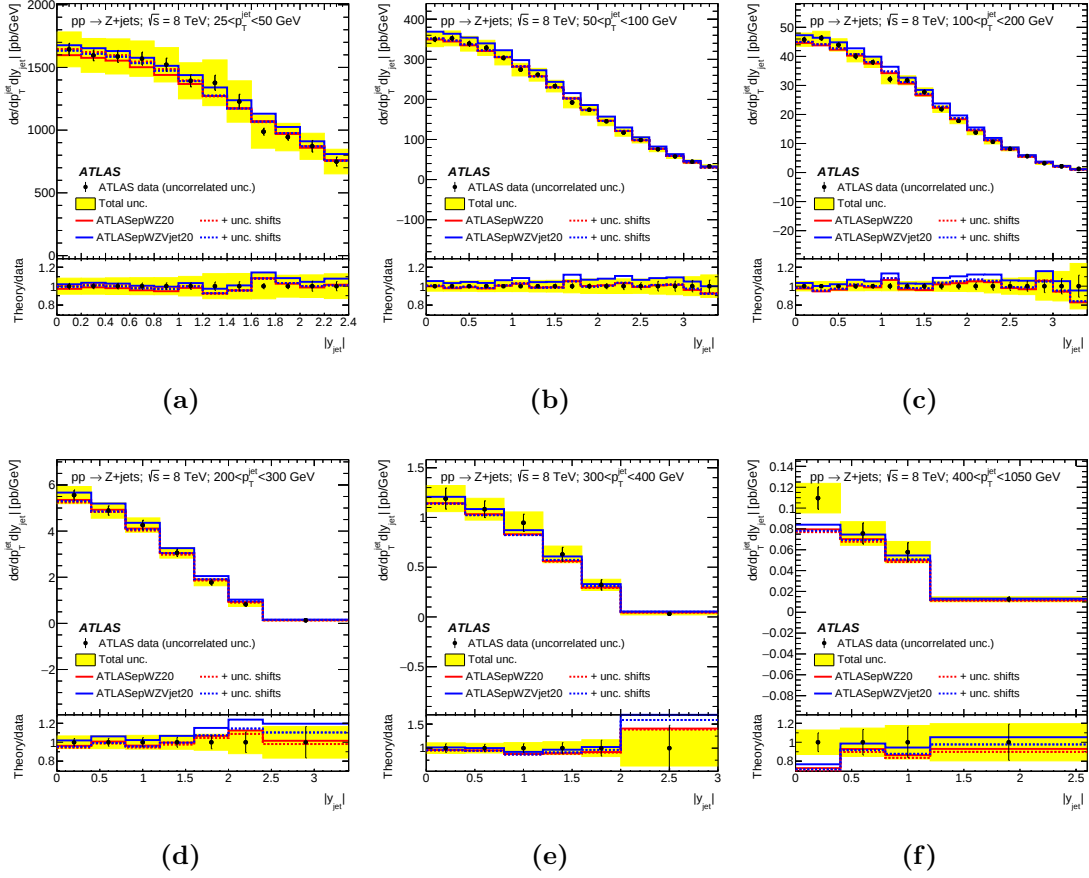


Figure 9.5: The differential cross-section measurements of $Z + \text{jets}$ as a function of the absolute rapidity of inclusive jets, $|y^{\text{jet}}|$, in bins of (a) $25 < p_T^{\text{jet}} < 50$ GeV, (b) $50 < p_T^{\text{jet}} < 100$ GeV, (c) $100 < p_T^{\text{jet}} < 200$ GeV, (d) $200 < p_T^{\text{jet}} < 300$ GeV, (e) $300 < p_T^{\text{jet}} < 400$ GeV and (f) $400 < p_T^{\text{jet}} < 1050$ GeV, where the transverse momentum of inclusive jets is labelled p_T^{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. (9.3) are applied.

obtained in the ATLASepWZ20 fit, demonstrating that there is no tension between these data and the $V + \text{jets}$ data. The partial χ^2 of the $W + \text{jets}$ and $Z + \text{jets}$ 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 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

Fit	ATLASepWZ20	ATLASepWZVjet20
Total χ^2/NDF	1316 / 1105	1460 / 1198
HERA partial χ^2/NDP	1114 / 1016	1132 / 1016
HERA correlated χ^2	49	50
HERA log penalty χ^2	-11	-12
ATLAS W, Z partial χ^2/NDP	121/105	113/105
ATLAS W + jets partial χ^2/NDP	0/0	25/30
ATLAS Z + jets partial χ^2/NDP	0/0	82/63
ATLAS correlated χ^2	41	65
ATLAS log penalty χ^2	1	6

Table 9.1: 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).

Q^2 cut, $Q_{\min}^2$, and the starting scale at which the PDFs are parameterised, Q_0^2 . The variation in charm-quark mass and starting scale are performed simultaneously to fulfil the condition $Q_0^2 < m_c^2$ such that the charm PDF is calculated perturbatively. Each of these variations follow that of the ATLASepWZ16 analysis [145]. 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 χ^2 per degree of freedom in table 9.2, 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 as a statistical standard deviation.

The impact 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 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 ATLASepWZVjet20 set. These, and other similar cross checks, are shown in appendix E.

	ATLASepWZ20	ATLASepWZVjet20
Nominal χ^2/NDF	1316 / 1105	1460 / 1198
Parameter variations		
$A_{\bar{u}} \neq A_{\bar{d}}$	1313 / 1104	1458 / 1197
$A_{\bar{u}} \neq A_{\bar{d}} \& B_{\bar{u}} \neq B_{\bar{d}}$	1313 / 1103	1454 / 1196
$B_{\bar{s}} \neq B_{\bar{d}}$	1313 / 1104	1459 / 1197
$B_{\bar{u}} \neq B_{\bar{d}}$	1313 / 1104	1459 / 1197
$D_{\bar{d}}$	1316 / 1104	1459 / 1197
D_{d_v}	1316 / 1104	1460 / 1197
$D_{\bar{s}}$	1316 / 1104	1460 / 1197
D_{u_v}	1312 / 1104	1457 / 1197
$E_{\bar{d}}$	1315 / 1104	1459 / 1197
$E_{\bar{s}}$	1313 / 1104	1460 / 1197
$E_{\bar{u}}$	1315 / 1104	1459 / 1197
F_{d_v}	1315 / 1104	1460 / 1197
$F_{\bar{s}}$	1310 / 1104	1460 / 1197
$F_{\bar{u}}$	1313 / 1104	1458 / 1197
F_{u_v}	1313 / 1104	1456 / 1197
Model variations		
$Q_{\min}^2 = 12.5 \text{ GeV}^2$	1245 / 1056	1393 / 1149
$Q_{\min}^2 = 7.5 \text{ GeV}^2$	1374 / 1145	1529 / 1238
$Q_0^2 = 2.2 \text{ GeV}^2$ and $m_c = 1.49\text{GeV}$	1315 / 1105	1465 / 1198
$Q_0^2 = 1.6 \text{ GeV}^2$ and $m_c = 1.37\text{GeV}$	1318 / 1105	1458 / 1198
$\alpha_s(m_Z) = 0.120$	1317 / 1105	1463 / 1198
$\alpha_s(m_Z) = 0.116$	1314 / 1105	1458 / 1198
$m_b = 4.75 \text{ GeV}$	1316 / 1105	1461 / 1198
$m_b = 4.25 \text{ GeV}$	1315 / 1105	1458 / 1198

Table 9.2: 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.

Figure 9.6 shows 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. 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, these changes allow for an increase in the W^+ cross section, as depicted in Figure 9.4, 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.

9.4.2 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 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 $x \sim 0.1$ [146]. 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 $Q^2 = 1.9 \text{ GeV}^2$ is shown in Figure 9.7, with a comparison between ATLASepWZVjet20 and ATLASepWZ20 displaying the direct effect of the $V + \text{jets}$ data, and with the experimental, model and parameterisation uncertainties plotted separately. The impact of the $V + \text{jets}$ 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 9.4.1. A similar comparison is shown at $Q^2 = m_Z^2$ in appendix C.

To understand the effect of the different data sets on the high- x $x\bar{d}$ distribution, a scan of χ^2 is performed through the parameter controlling the behaviour in this region, $C_{\bar{d}}$ where $x\bar{d} \propto (1 - x)^{C_{\bar{d}}}$.¹ A high $C_{\bar{d}}$ value of ~ 10 corresponds to a

¹The 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.

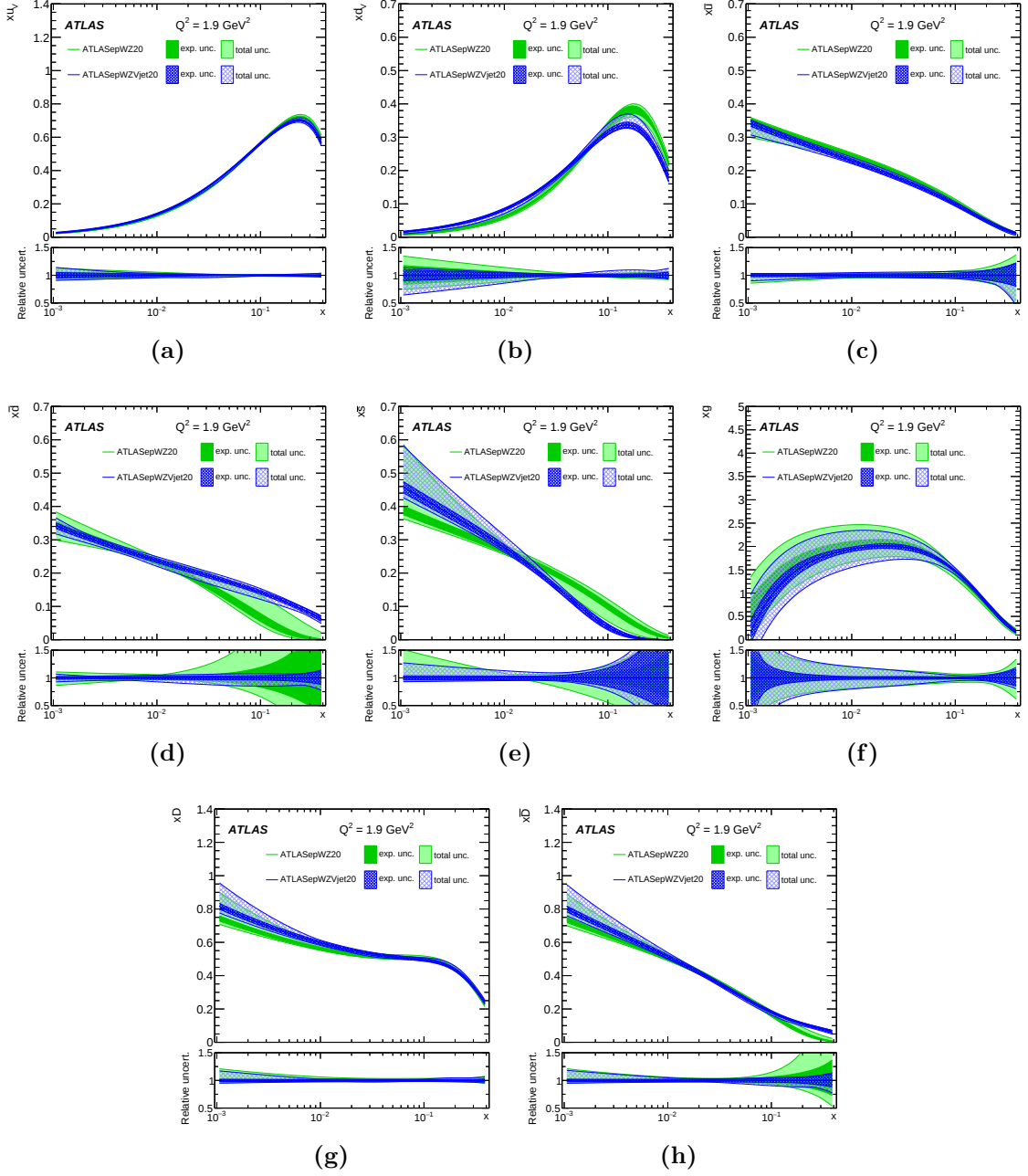


Figure 9.6: PDFs at the scale $Q^2 = 1.9 \text{ GeV}^2$ as a function of Bjorken x obtained for the (a)-(b) valence quarks, (c)-(d) up and down sea quarks, (e) strange sea quark, (f) gluon, (g) x_D and (h) $x_{\bar{D}}$ distributions when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive W and 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.

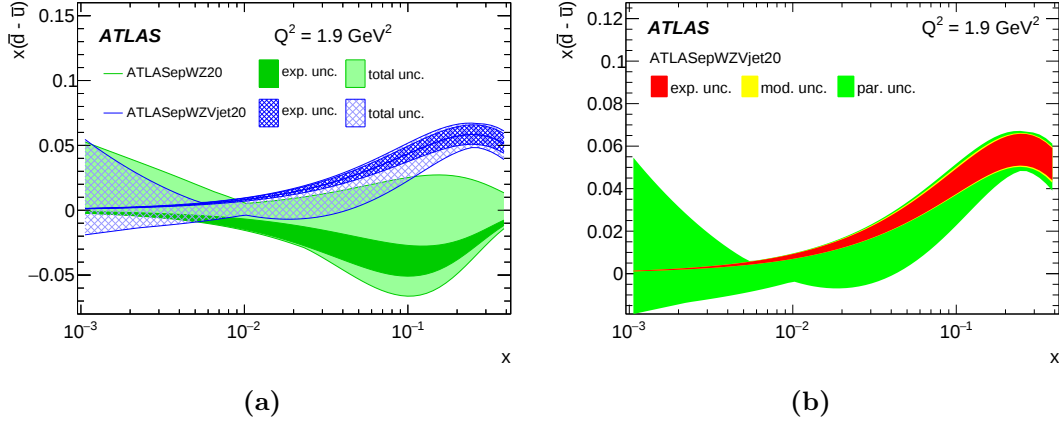


Figure 9.7: The $x(\bar{d} - \bar{u})$ distribution evaluated at $Q^2 = 1.9 \text{ GeV}^2$ as a function of Bjorken x (a) extracted from the ATLASepWZ20 (green) and ATLASepWZVjet20 (blue) fits with experimental and total uncertainties plotted separately, and (b) extracted from the ATLASepWZVjet20 fit only with experimental, model and parameterisation uncertainties shown separately in red, yellow and green, respectively.

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 Figure 9.8a, 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 re-fitted 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 smallest at a value of $C_{\bar{d}} = 10 \pm 1$, whereas the χ^2 of the ATLASepWZVjet20 fit is smallest at a lower $C_{\bar{d}} = 1.6 \pm 0.3$, corresponding to a higher $x\bar{d}$ distribution at $x \gtrsim 0.1$ consistent with the PDFs presented in section 9.4.1. Another shallow minimum is observed for the ATLASepWZ20 fit at $C_{\bar{d}} \sim 3$, corresponding to a solution similar to that of the ATLASepWZVjet20 fit; however, it exhibits a χ^2 approximately two units larger than in the best fit. The ATLASepWZVjet20 fit fails to converge for values of $C_{\bar{d}} \gtrsim 12$ and no second minimum is observed.

In Figure 9.8b, these χ^2 distributions are decomposed into contributions from the HERA and ATLAS data. These contributions include the partial, correlated

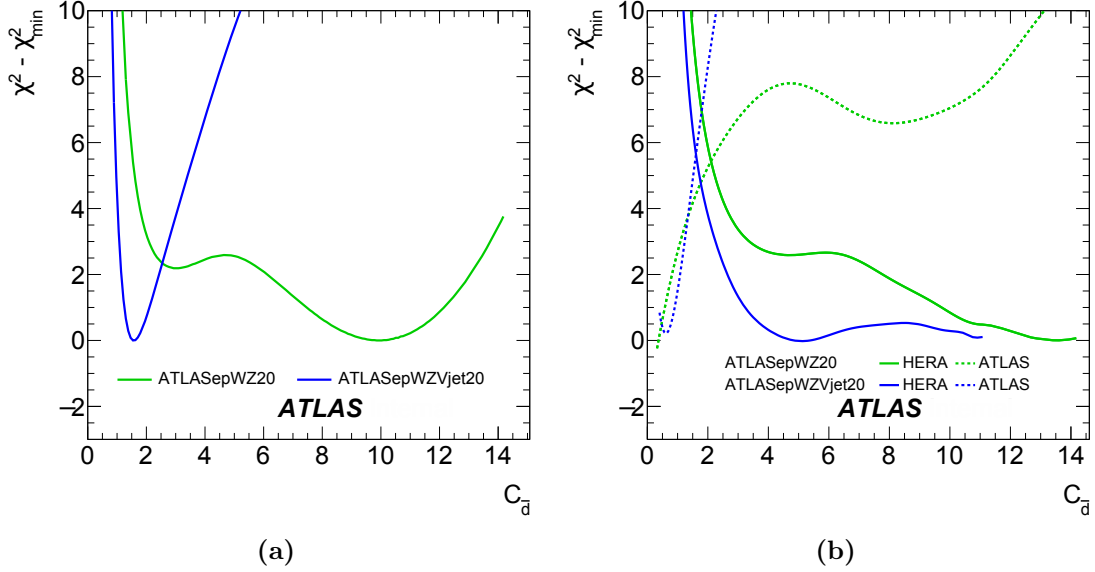


Figure 9.8: The χ^2 of the ATLASepWZ20 (green line) and ATLASepWZVjet20 (blue line) fits recorded as a function of the $C_{\bar{d}}$ fit parameter that determines the high- x behaviour of the $x\bar{d}$ PDF with $x\bar{d} \propto (1-x)^{C_{\bar{d}}}$. At each point, all other parameters are fitted along with the nuisance parameters corresponding to experimental systematic uncertainties, and the lowest recorded χ^2 for each line shown, $\chi_{\min}^2$, subtracted. Shown are (a) the total χ^2 and (b) the χ^2 separated into contributions from HERA (solid lines) and ATLAS (dashed lines) data and smoothed.

and log penalty χ^2 , which are discussed in section 9.3. 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 $V + \text{jets}$ data provide sufficient constraining power in addition to the inclusive W and Z data to dominate the result and tightly constrain the $C_{\bar{d}}$ parameter to a low value, while the ATLASepWZ20 fit lacks the necessary information.

9.4.3 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.

Before the first LHC precision W and Z boson data, it was widely assumed, motivated by previous analyses of dimuon production in neutrino scattering [170–174], 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 $Q^2 = 1.9 \text{ GeV}^2$ [135–137, 175].

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 $Q^2 = 1.9 \text{ GeV}^2$.² This was the case for the ATLASepWZ16 fit for every parameterisation variation used. Furthermore, a Hessian profiling exercise of the global PDFs MMHT14 [137] and CT14 [176] demonstrated that the data constrain and increase the ratio of the strange to the total up and down sea [145]. 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 [136]. 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 9.9, 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 9.4.1. 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

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

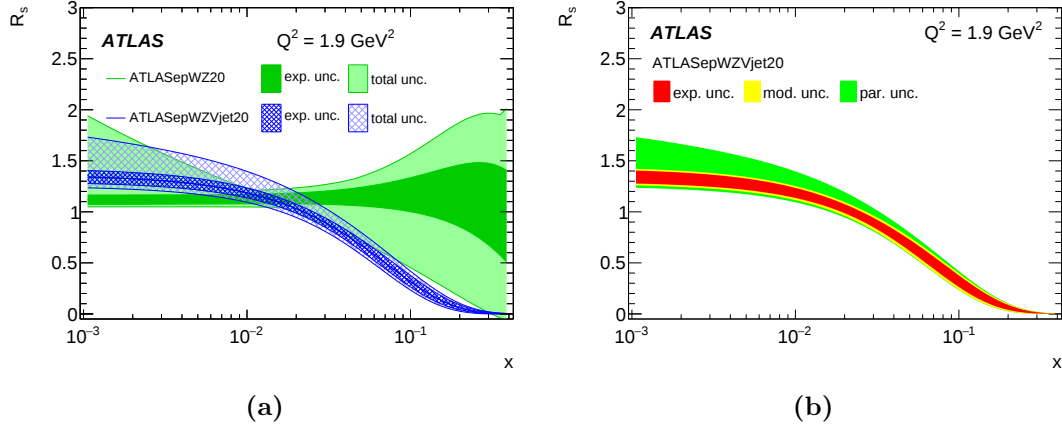


Figure 9.9: The R_s distribution, evaluated at $Q^2 = 1.9 \text{ GeV}^2$, (a) extracted from the ATLASepWZ20 (green) and ATLASepWZVjet20 (blue) fits with experimental and total uncertainties plotted separately, and (b) extracted from the ATLASepWZVjet20 fit only with experimental, model and parameterisation uncertainties shown separately in red, yellow and green, respectively.

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 9.3 and a similar comparison between the ATLASepWZ20 and ATLASepWZVjet20 PDF sets is shown at $Q^2 = m_Z^2$ in appendix C.

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

Table 9.3: 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.

9.4.4 Comparison with global PDFs

The ATLASepWZVjet20 R_s distribution evaluated at $Q^2 = 1.9 \text{ GeV}^2$ is shown in Figure 9.10 in comparison with the global PDF sets ABMP16 [135], CT18, CT18A [136], MMHT14 [137] and NNPDF3.1 [138]. 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 data set labelled NNPDF3.1_strange [175]. Similar comparisons are shown evaluated at $Q^2 = m_Z^2$ in appendix C. 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 9.11. 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 $x \gtrsim 0.2$.

In Figure 9.12 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 $W + \text{jets}$ and $Z + \text{jets}$ data are most sensitive and demonstrate a preference for a higher $x\bar{d}$ distribution as discussed in section 9.4.1. 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 Figures 9.10 and 9.11, this is not replicated in the $x(\bar{d} - \bar{u})$ distribution in either case.

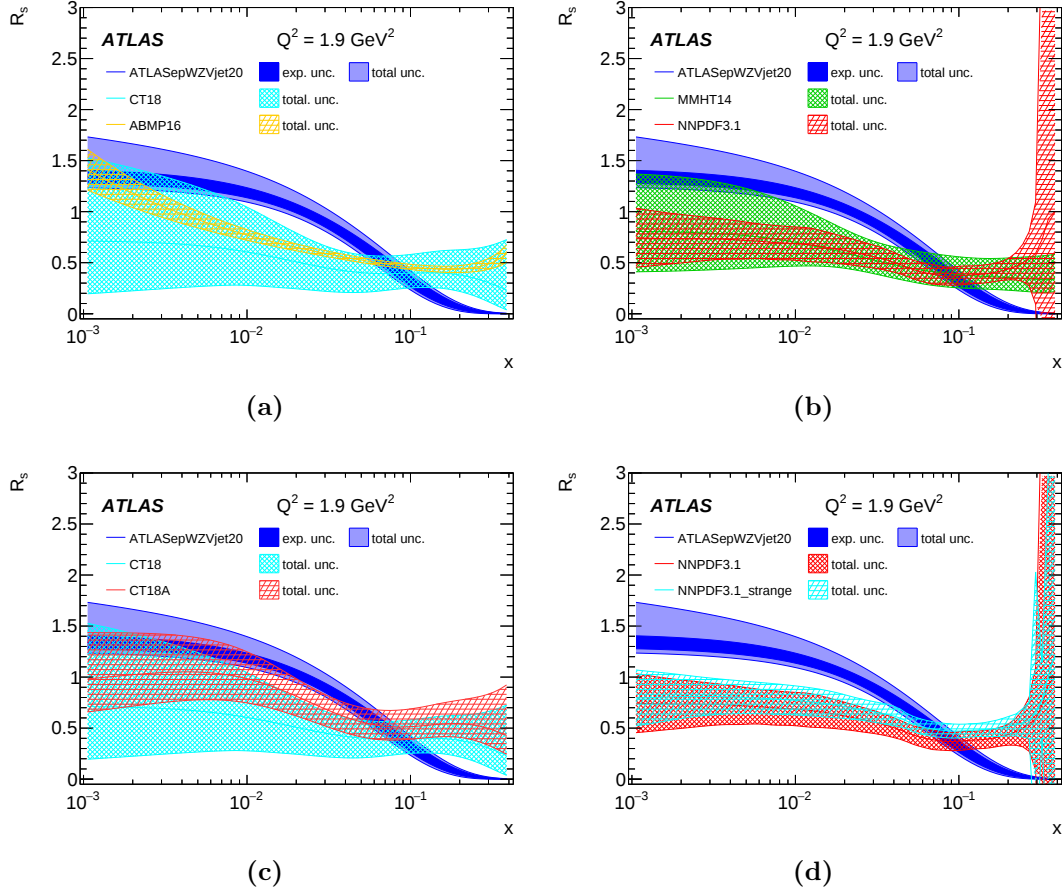


Figure 9.10: The $R_s = (s + \bar{s})/(\bar{u} + \bar{d})$ distribution evaluated at $Q^2 = 1.9 \text{ GeV}^2$ as a function of Bjorken x , for the ATLASepWZVjet20 PDF set in comparison with global PDFs (a) ABMP16 and CT18, (b) MMHT14 and NNPDF3.1, and in additional comparisons with (c) CT18 and CT18A, and (d) NNPDF3.1 and NNPDF3.1_strange [135–138, 175]. 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 [177].

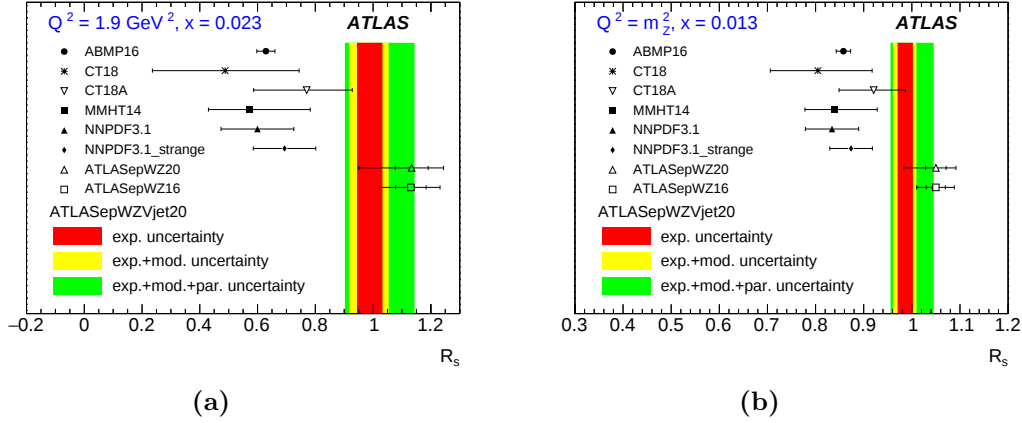


Figure 9.11: Summary plots of R_s evaluated at (a) $x = 0.023$ and $Q^2 = 1.9 \text{ GeV}^2$, and (b) $x = 0.013$ and $Q^2 = m_Z^2$, for the ATLASepWZVjet20 PDF set in comparison with global PDFs [135–138, 175], and the ATLASepWZ16 and ATLASepWZ20 sets. The experimental, model and parameterisation uncertainty 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 [177].

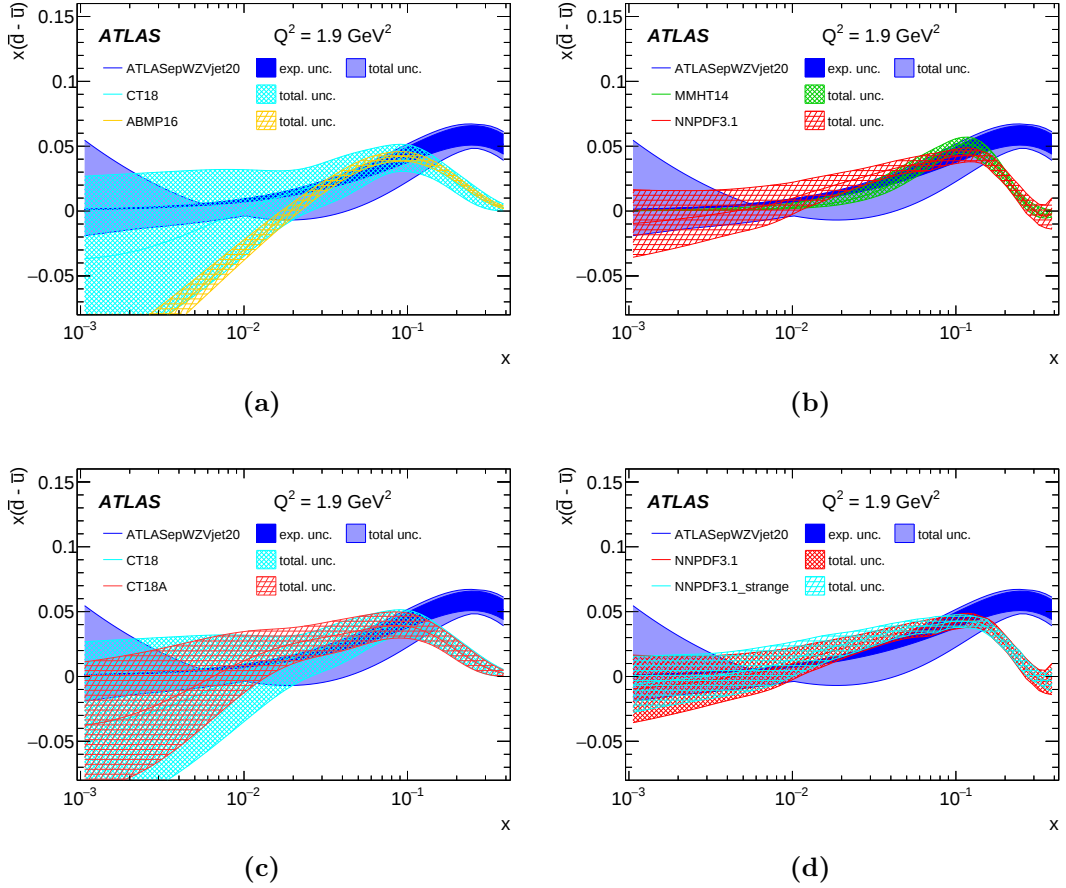


Figure 9.12: The $x(\bar{d} - \bar{u})$ distribution evaluated at $Q^2 = 1.9 \text{ GeV}^2$ as a function of Bjorken x , for the ATLASepWZVjet20 PDF set in comparison with global PDFs (a) ABMP16 and CT18, (b) MMHT14 and NNPDF3.1, and in additional comparisons with (c) CT18 and CT18A, and (d) NNPDF3.1 and NNPDF3.1_strange [135–138, 175]. 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 [177].

10

Summary and conclusion

The production of W bosons is one of the most abundant processes at a proton-proton collider such as the LHC. In association with jets, this is one of the most powerful tests of our understanding of quantum chromodynamics that we can currently access. Higher-order calculations and modelling techniques can be stretched to their limit by the high statistical precision up to high energy scales. Furthermore, the charge-asymmetric initial state of the proton-proton collisions means that we expect a charge-asymmetric final state – the precise values of the ratio of the production of W^+ and W^- bosons gives access to the structure of the proton, with greatly reduced contributions from systematic uncertainties and differences between modelling techniques. The inclusion of jets in the final state serves to improve further our knowledge of the structure of the proton, probing higher Bjorken x and higher negative squared four-momentum transfer, Q^2 .

In this thesis, a measurement of the production of W bosons in association with jets at the ATLAS detector was presented using 36.2 fb^{-1} of proton-proton collision data at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. The contribution of background processes was estimated through both monte-carlo simulation and data-driven methods. It was found that multijet production was dominant up to four jets, and the production of top quarks dominant beyond. These background contributions were subtracted, before the iterative Bayesian unfolding method was used to deconvolute

the effect of the detector from the reconstructed distributions. This deconvolution ultimately serves the purpose of allowing calculations of the $W + \text{jets}$ process to be easily compared to the data, without any requirement to propagate the final state particles through hadronisation, showering or a simulation of the ATLAS detector.

The cross sections were measured firstly as a function of the number of jets, and secondly differentially as a function of four observables: p_T^W , p_T^{leading} , H_T and $x_{T\text{-odd}}$. For each of these, the charge-independent measurement was performed, as well as the individual W^+ and W^- measurements. The ratio of the latter two was also taken. In each case, a large range of systematic uncertainties was evaluated, and statistical correlations between the spectra evaluated through the bootstrap method. Comparisons were made to predictions from the MADGRAPH5_aMC@NLO (v2.3.2.p1) interfaced to PYTHIA 8 (v8.210) and SHERPA v2.2.1 generators, and good agreement to the cross sections within systematic uncertainties was found, although there was a systematic over-prediction of the W^+/W^- ratio for both generators in the p_T^W , p_T^{leading} and H_T spectra. This may have interesting implications for the structure of the proton.

Finally, a determination of the proton PDFs has been performed using similar measurement of $W + \text{jets}$ production completed by the ATLAS collaboration at a centre-of-mass energy of $\sqrt{s} = 8 \text{ TeV}$. Significant improvements to the determination of the sea-quark densities at high Bjorken x have been realised, and the previously-unused χ^2 -scan method has allowed this data to provide new insight on already-analysed data. A new PDF set, labelled ATLASepWZVjet20, has been released to the public domain by this analysis, and new insights on the predictive power of these PDFs will surely come forward in the future as analysts begin to include it within their research. This study, and others like it, will allow our measurements to breach new boundaries, both for the precise determination of Standard Model parameters (such as boson masses) and for the search for new physics at higher energy scales.

The 13 TeV measurement will provide further improvements on the 8 TeV measurements. Not only will the final data set include approximately 140 fb^{-1} of pp data, setting tremendous benchmarks on statistical precision, but a range

of new analysis techniques will be implemented and the muon-decay channel will be included in the analysis, thereby doubling again the precision. Furthermore, lessons learned from the PDF-fitting exercise – such as the importance of statistical correlations – have already been considered, and the next iteration of this analysis will provide a benchmark legacy measurement from the ATLAS detector.

Beyond Run 2, there is substantial potential for the future of QCD analysis and particularly PDF determination. For example, the high-luminosity LHC (HL-LHC) aims to provide a further factor of 10 in integrated luminosity, with the potential not only to improve our statistical precision but a number of systematic uncertainties in addition. There is also the possibility of a Large Hadron-electron Collider (LHeC). This project aims to run concurrent with the HL-LHC with the centre-of-mass energy required to take measurements at Bjorken x as low as 10^{-6} and the luminosity required to approximately replicate the entire HERA program for each day of operation (the details are discussed in Refs. [178] and [179]). Studies such as that in Ref. [180] suggest that both would perform similarly for cross-section determination of most LHC processes, and that it is studies such as those presented in this thesis which would provide the best determination of high- x PDFs and improve the precision of any analysis which uses them. However, the LHeC would provide a well-understood and synergised set of measurements with the ability to push many facets of our QCD calculations to their limit, including but not limited to log- x resummation, the difference between the s and $\bar{s}$ proton PDFs, heavy-ion PDFs and the validation of (or challenge to) hadron-hadron collision factorisation. There remains a huge breadth of what can still be discovered and pushing these boundaries can only give more insight of the phenomenology of quantum chromodynamics. Over the next decade, experimentalists will continue to expand our knowledge and pin down our understanding of the universe in which we live, and with all of these analyses together we could discover what exactly exists beyond the Standard Model.

Appendices

A

Photo-current stability

Three periods were considered to measure stability, prior to 2016, 2017 and 2018 *pp*-collisions, respectively. In each case, a region was chosen to be sufficiently long for a reliable measurement with stable temperature readings. The small variations in these normalised currents is likely due to fluctuations in the VCSEL fibre coupled power. Therefore, the modules are grouped according to readout crates as opposed to SCT geometrical region. The results for 2016, 2017 and 2018 are shown in figs. A.1 to A.3, respectively. The resulting gradients are summarised in a histogram in fig. A.4.

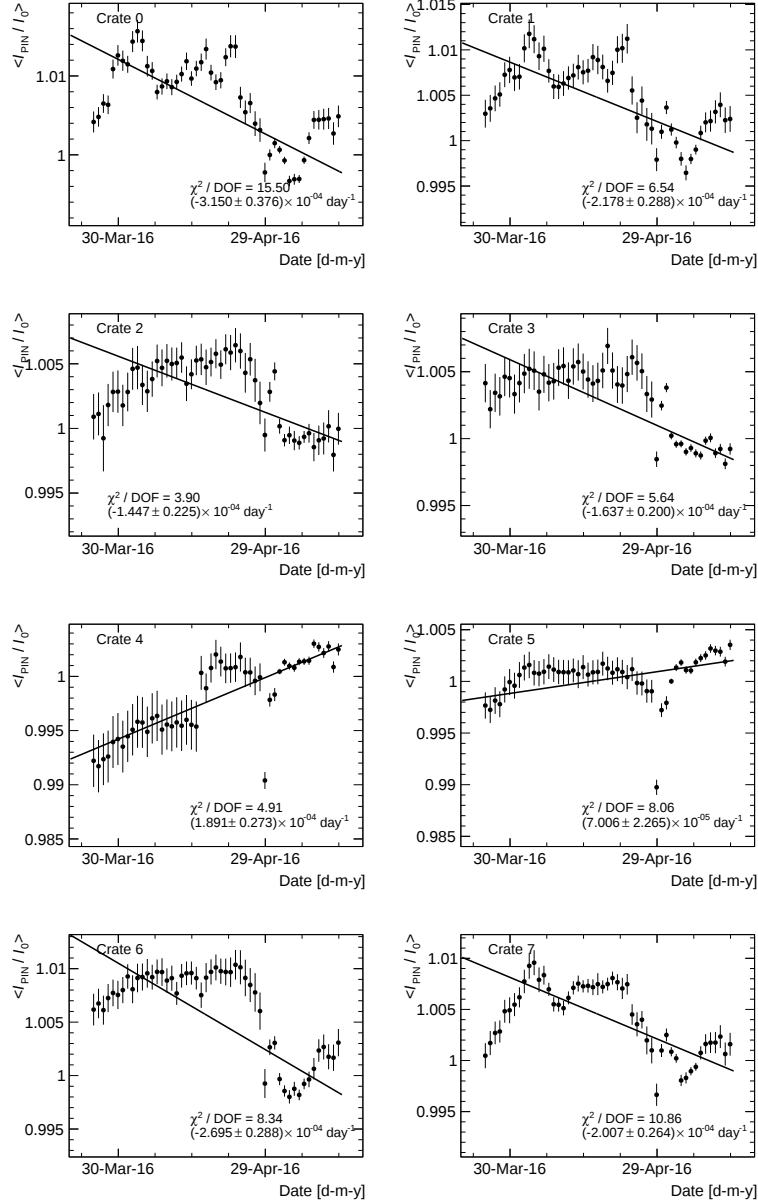


Figure A.1: The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC for a period prior to data taking in 2016.

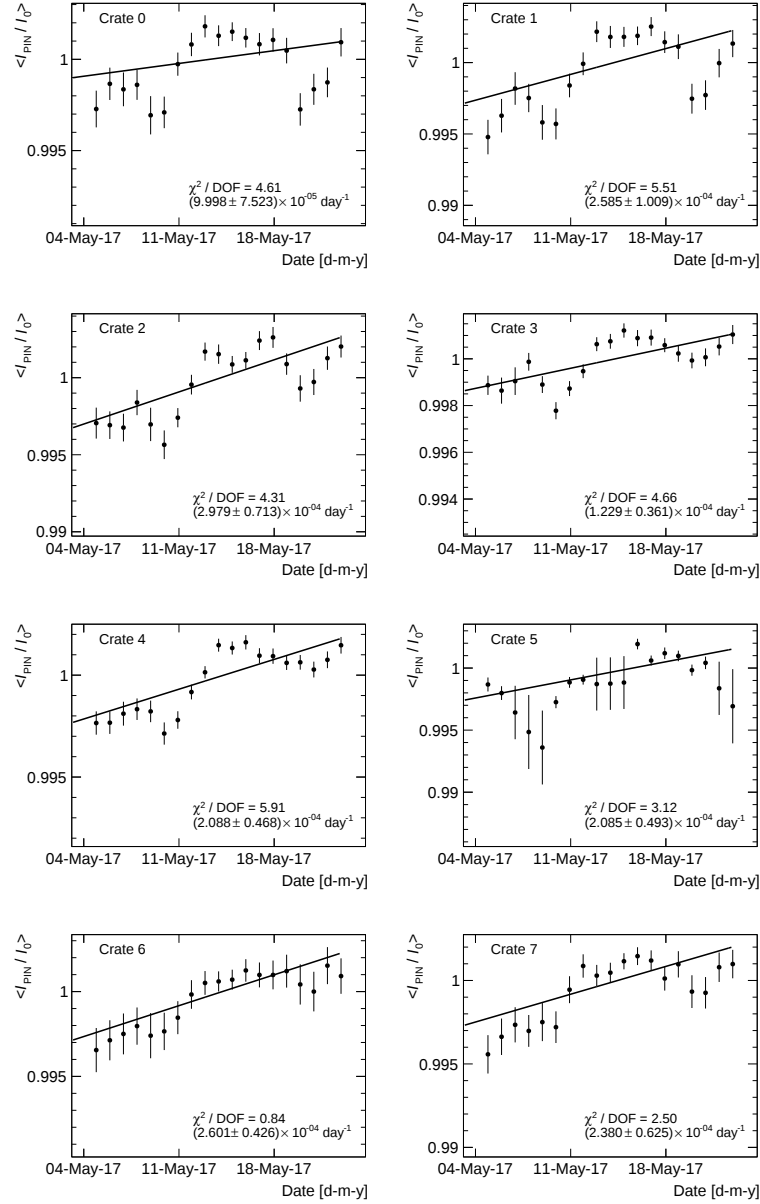


Figure A.2: The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC for a period prior to data taking in 2017.

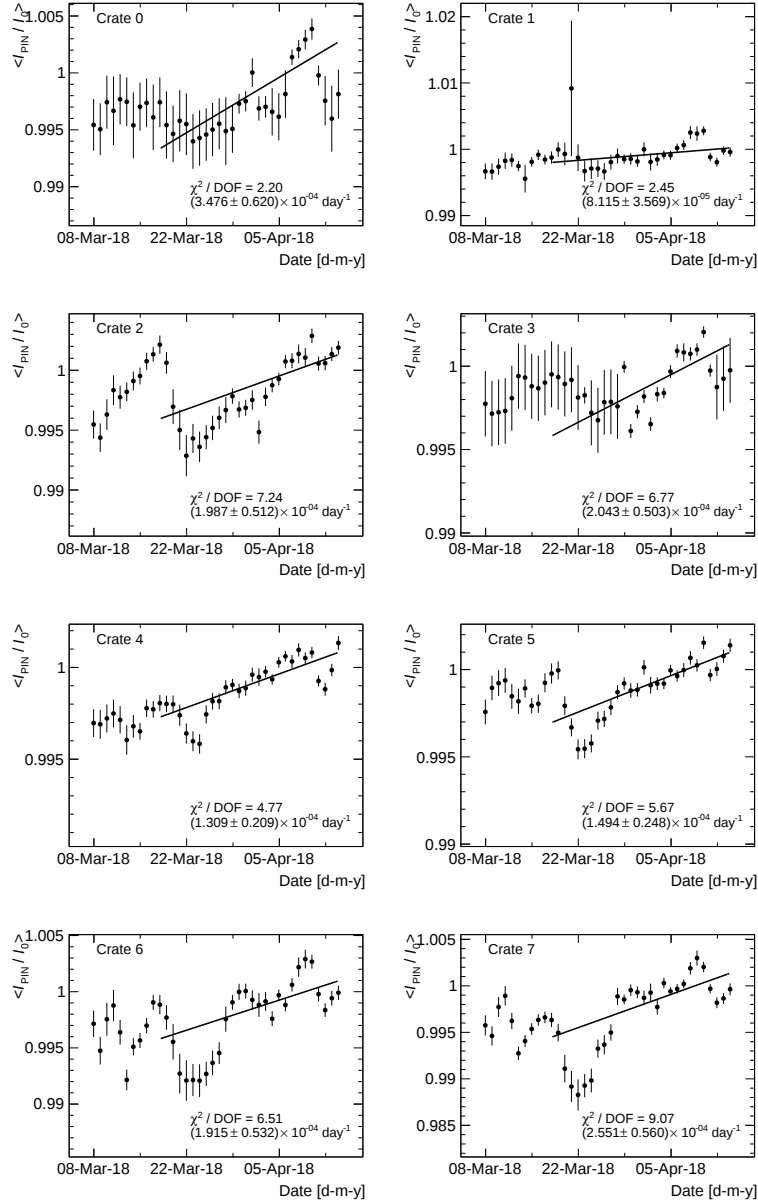


Figure A.3: The measured mean value of the photo-currents, normalised to data taken on 10 May 2016, for each scan point correlated against total integrated luminosity delivered by the LHC for a period prior to data taking in 2018.

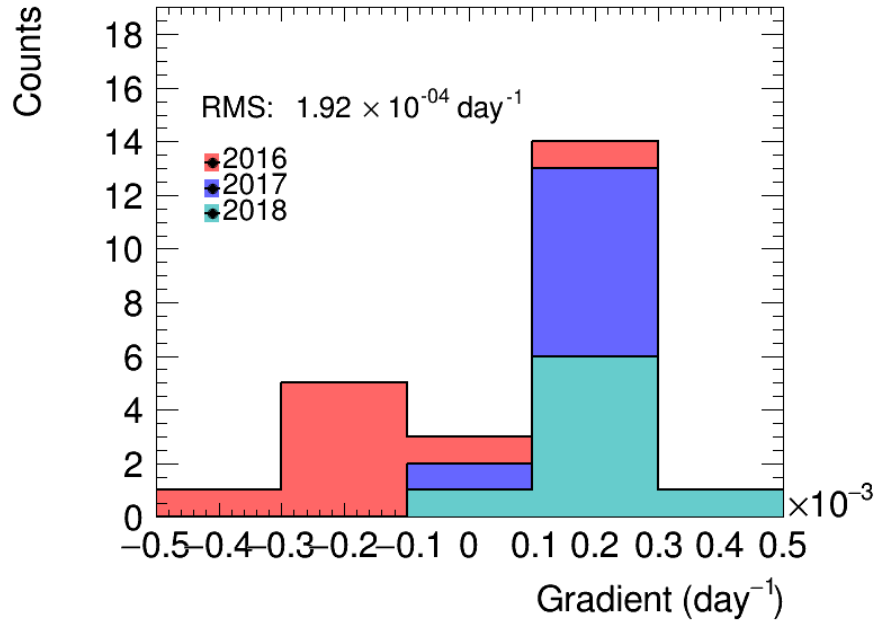


Figure A.4: A histogram of the fitted gradients to the plots of normalised I_{PIN} against time for the years 2016, 2017 and 2018. The RMS is shown on this plot, which is used as a systematic error for the final measurement of R_{∞} .

B

Additional material for the $W + \text{jets}$ analysis at $\sqrt{s} = 13 \text{ TeV}$

In the main body of this thesis, the discussion has centred for brevity around the production of at least either one or two jets in association with a W boson. Additionally, the comparison of data collected at the ATLAS detector with the full SM prediction is shown only for the charge-independent distributions. In this appendix, additional material for the results of this analysis are shown.

Firstly, a comparison between the measured data and the total SM prediction for each of the analysed distributions is shown for three and four inclusive jets for the validation region in figs. B.1 and B.2 and for the signal region in figs. B.3 and B.4. The predictions considered maintain a similar accuracy to their prediction of events with one jet inclusive, except for the SHERPA2.2.1 which for higher jet multiplicities tends to overestimate the $p_{\text{T}}^{\text{leading}}$ distribution at high values, similarly to the MG + PY8 prediction.

Additional comparisons between the measured data and the total SM prediction for the W^+ and W^- selection criteria are shown for the validation and signal regions in figs. B.5 and B.6 for the jet multiplicity, p_{T}^W and $x_{\text{T-odd}}$ distributions. As observed in chapter 8, the ratio of the predictions to the data tend to follow the

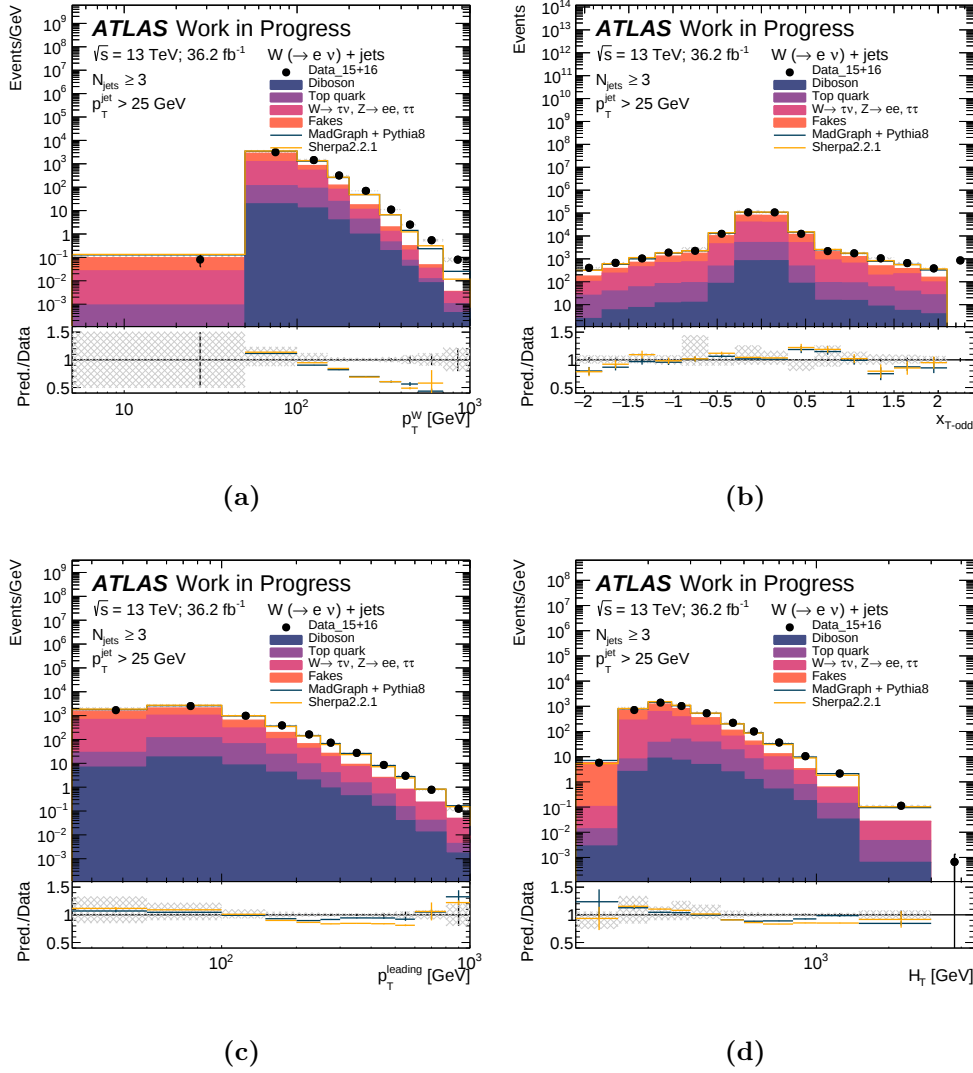


Figure B.1: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region for three or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T .

same trend as they do for the charge-independent distributions. Similar results are observed for the other analysed spectra.

The results of the closure test for the p_T^W spectrum for two and three jets inclusive are shown separately for the W^+ and W^- distributions in fig. B.7. As was the case in chapter 6, the unfolded distributions overlay the truth distribution at all iterations. The same is observed for the other analysed distributions for all inclusive jet multiplicities considered.

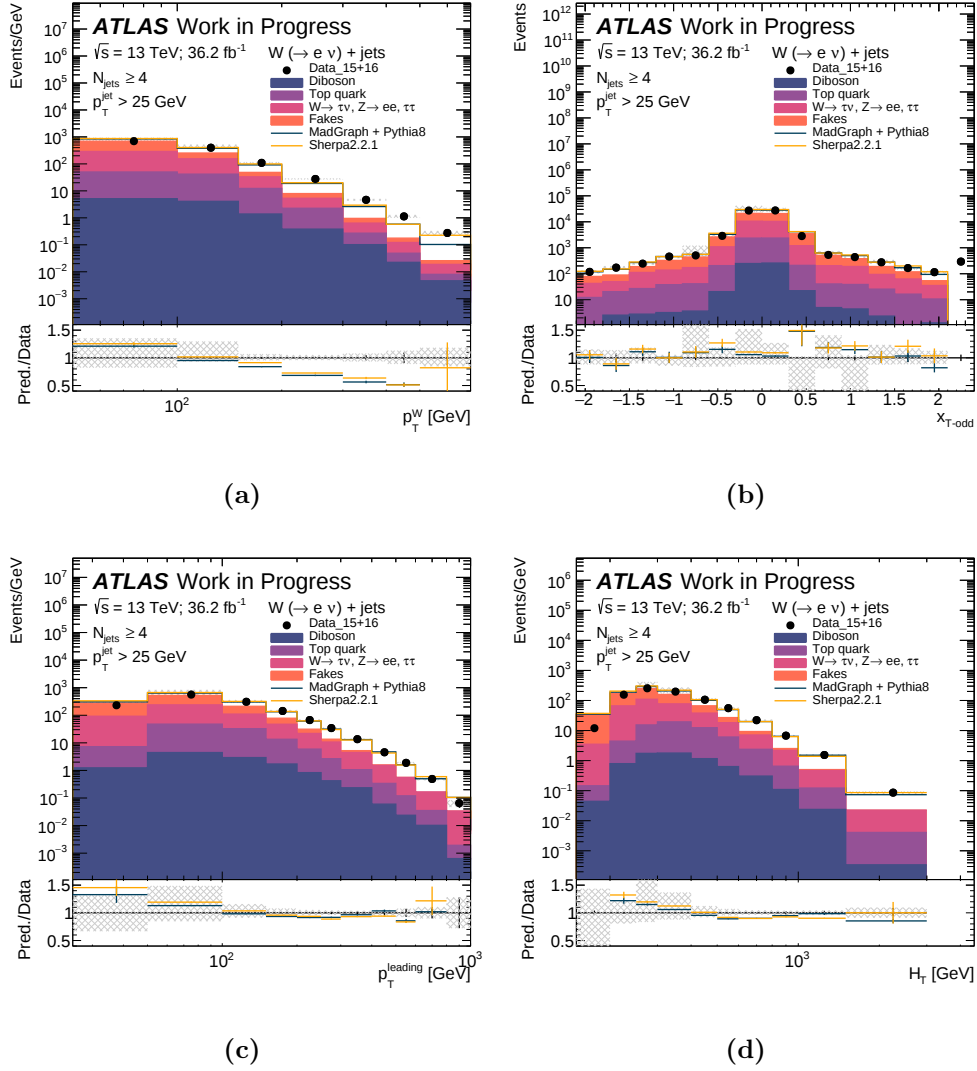


Figure B.2: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region for four or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T .

Finally, the unfolded charge-dependent distributions are shown for the p_T^W , p_T^{leading} and H_T spectra for two and three jets inclusive in figs. B.8 to B.10, with the W^+/W^- ratio shown for the p_T^{leading} and H_T spectra in figs. B.11 and B.12 for between two and five jets inclusive.

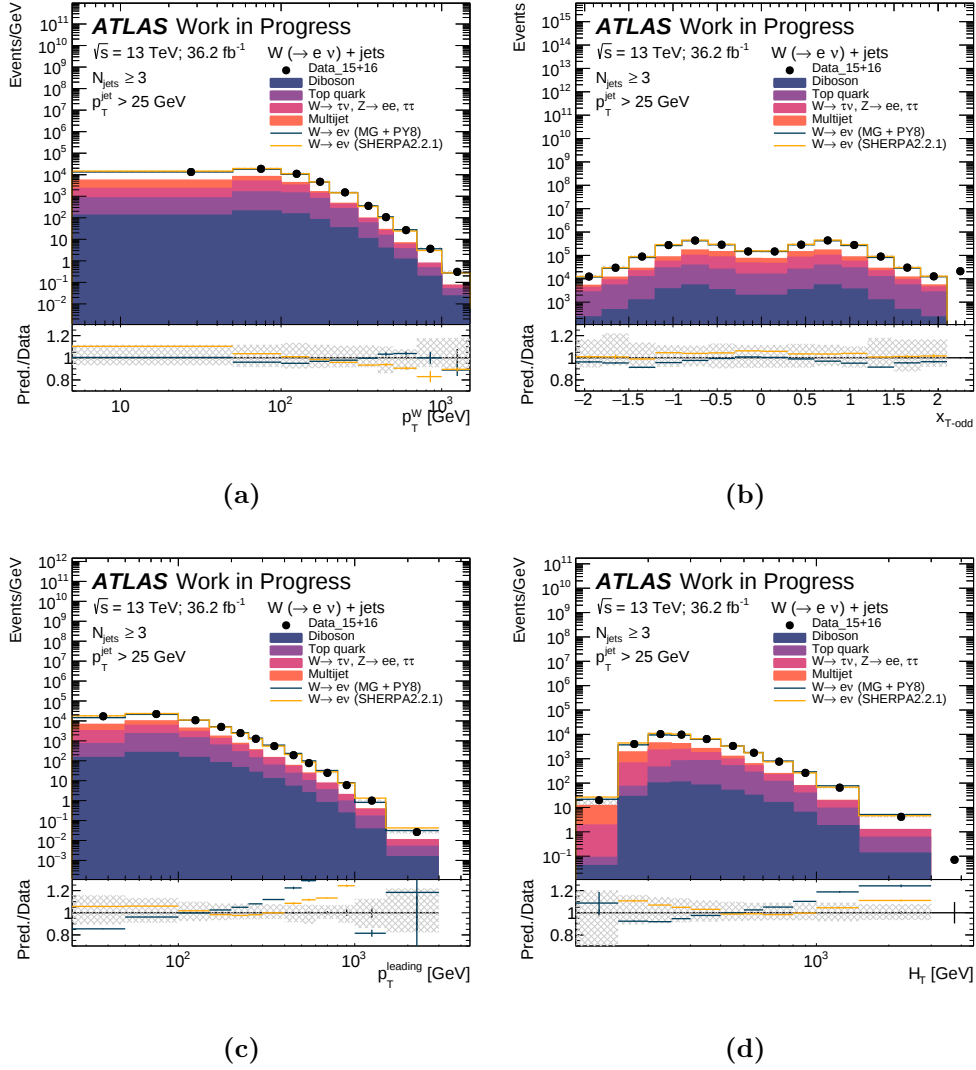


Figure B.3: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis for three or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T .

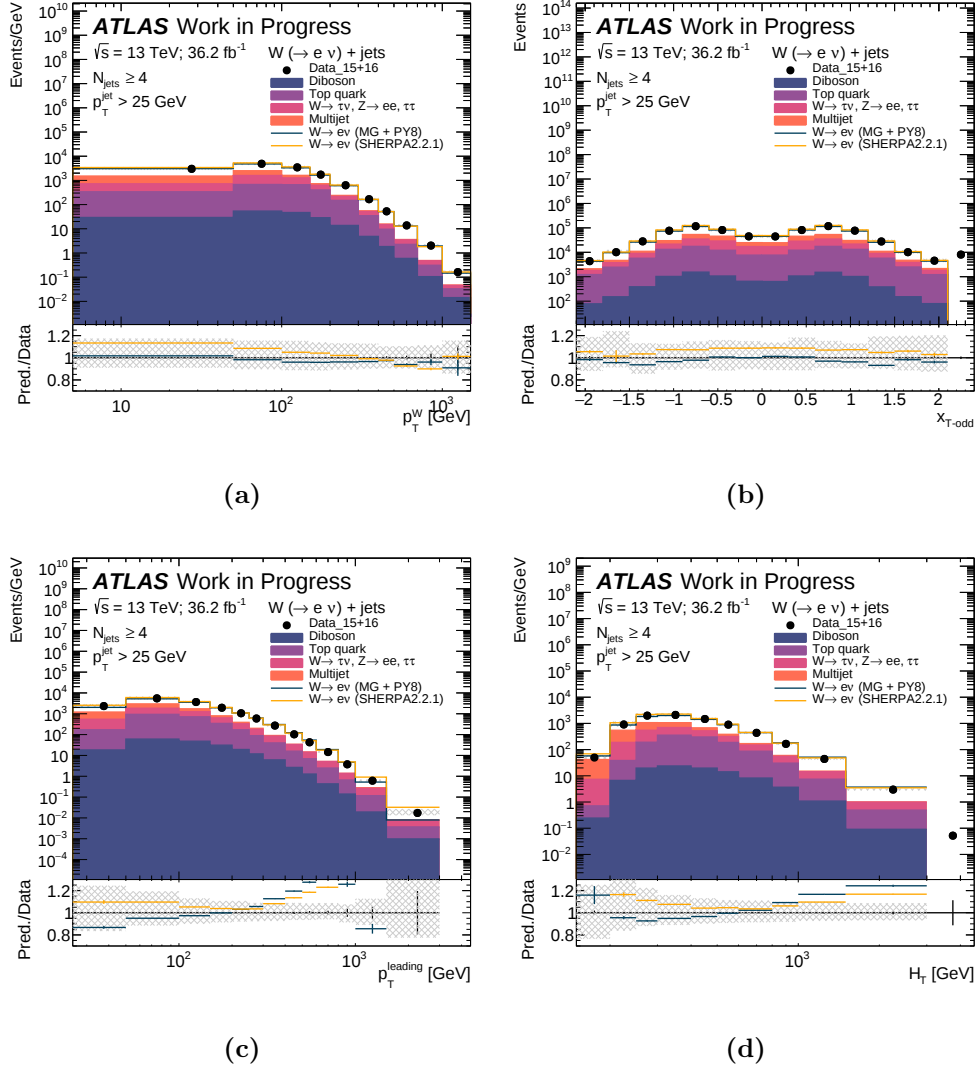


Figure B.4: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis for four or more reconstructed jets. The distributions shown are (a) p_T^W , (b) $x_{T\text{-odd}}$, (c) p_T^{leading} and (d) H_T .

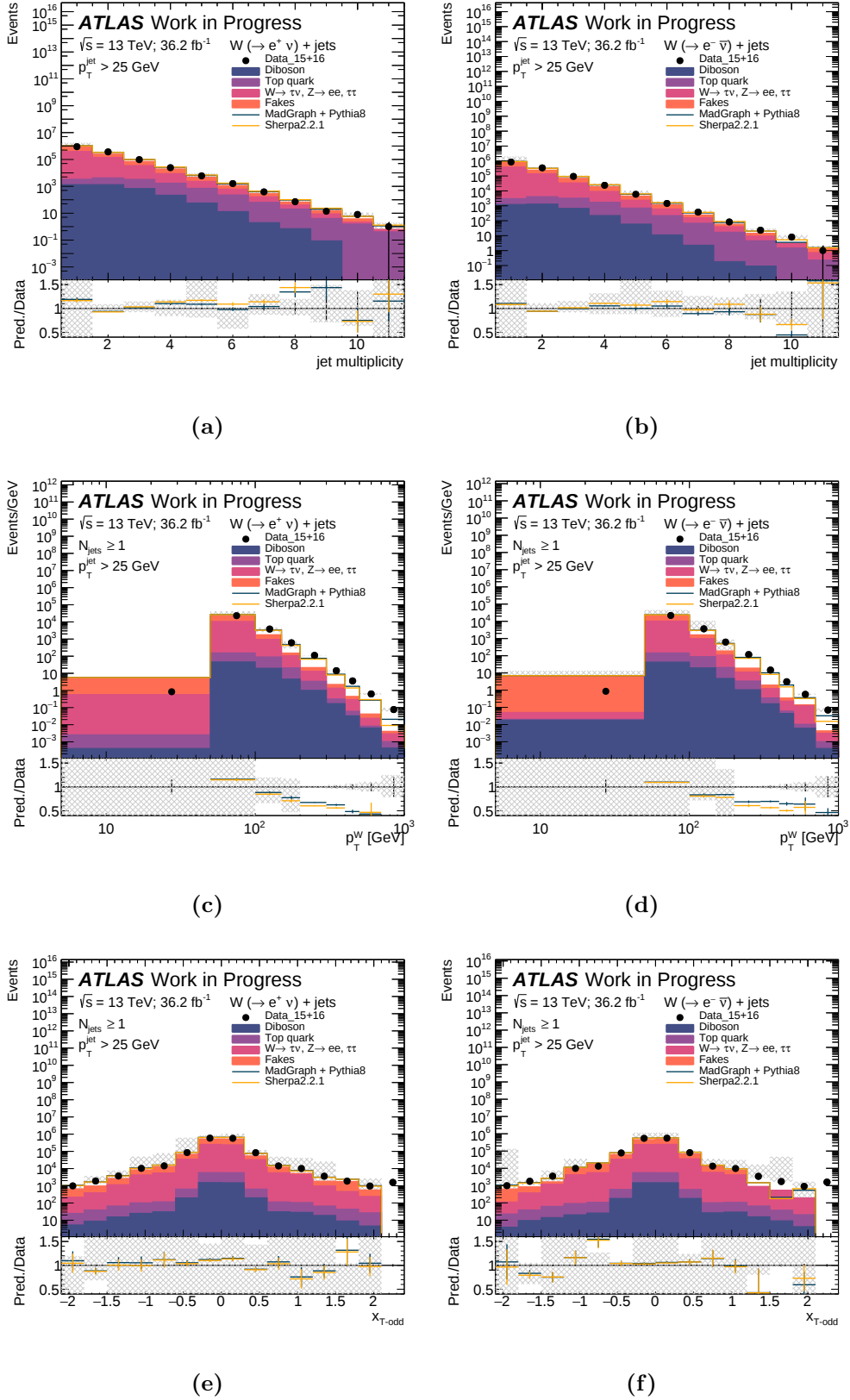


Figure B.5: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the validation region for one or more reconstructed jets split in to W^+ and W^- selection criteria. The distributions shown are (a)-(b) the jet multiplicity, (c)-(d) p_T^W and (e)-(f) $x_{T\text{-odd}}$.

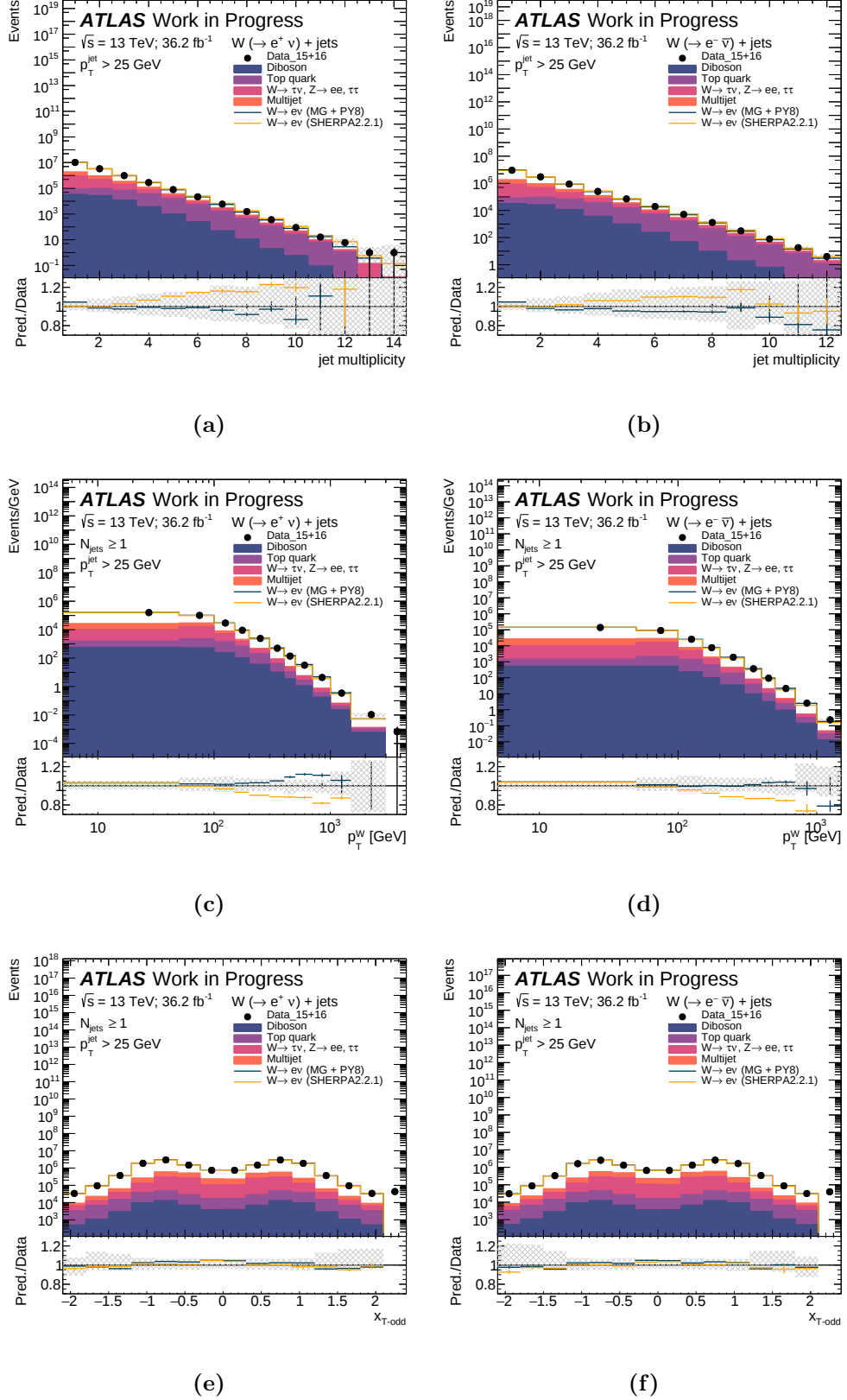


Figure B.6: A comparison between the recorded data passing the signal criteria and the total Standard-Model prediction for the $W + \text{jets}$ 13 TeV analysis in the signal region for one or more reconstructed jets split in to W^+ and W^- selection criteria. The distributions shown are (a)-(b) the jet multiplicity, (c)-(d) p_T^W and (e)-(f) $x_{T\text{-odd}}$.

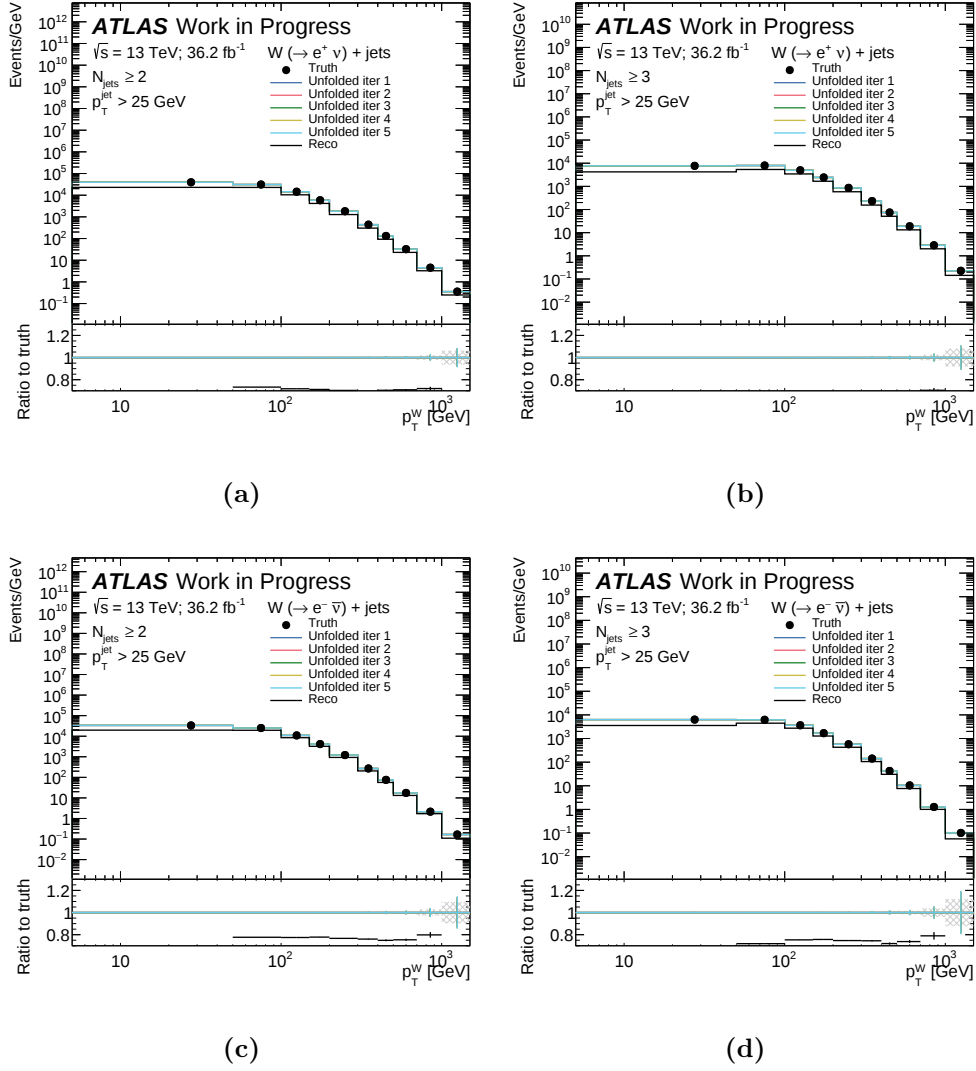


Figure B.7: Closure tests of the unfolding procedure used in this analysis. The MG + PY8 particle-level distribution is shown as black points, the MG + PY8 reconstructed distribution is shown as black lines, and the unfolded MG + PY8 distribution is shown as a series of coloured lines for up to five iterations for two and three jets inclusive. The p_T^W distribution is shown in all subfigures for the W^+ selection criteria in (a) and (b), and the W^- selection criteria in (c) and (d). In most of the bins shown, the unfolded distributions are overlaid such that only the five-iteration case is visible.

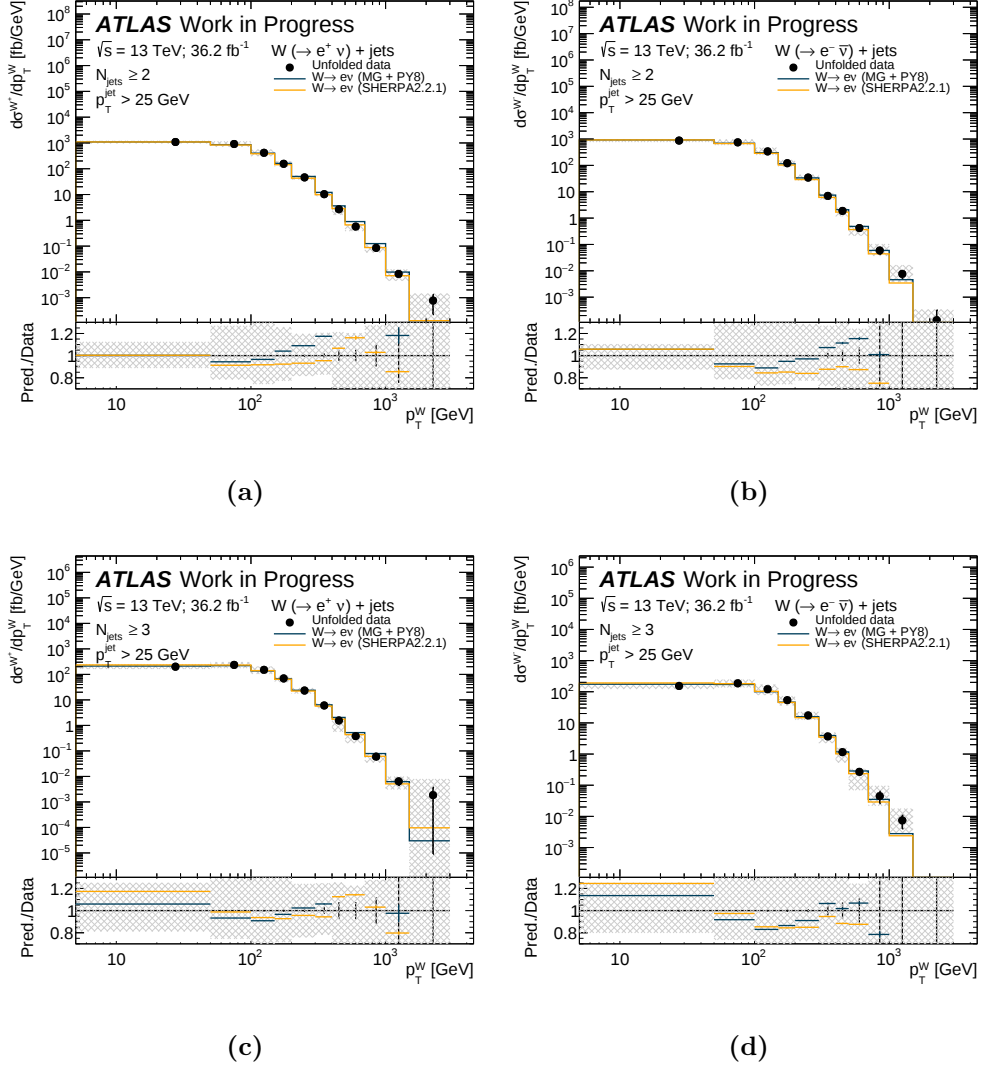


Figure B.8: The unfolded cross section for the production of a W boson in association with (a)-(b) at least two jets and (c)-(d) at least three jets for the p_T^W observable split in to the W^+ cross section and the W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.

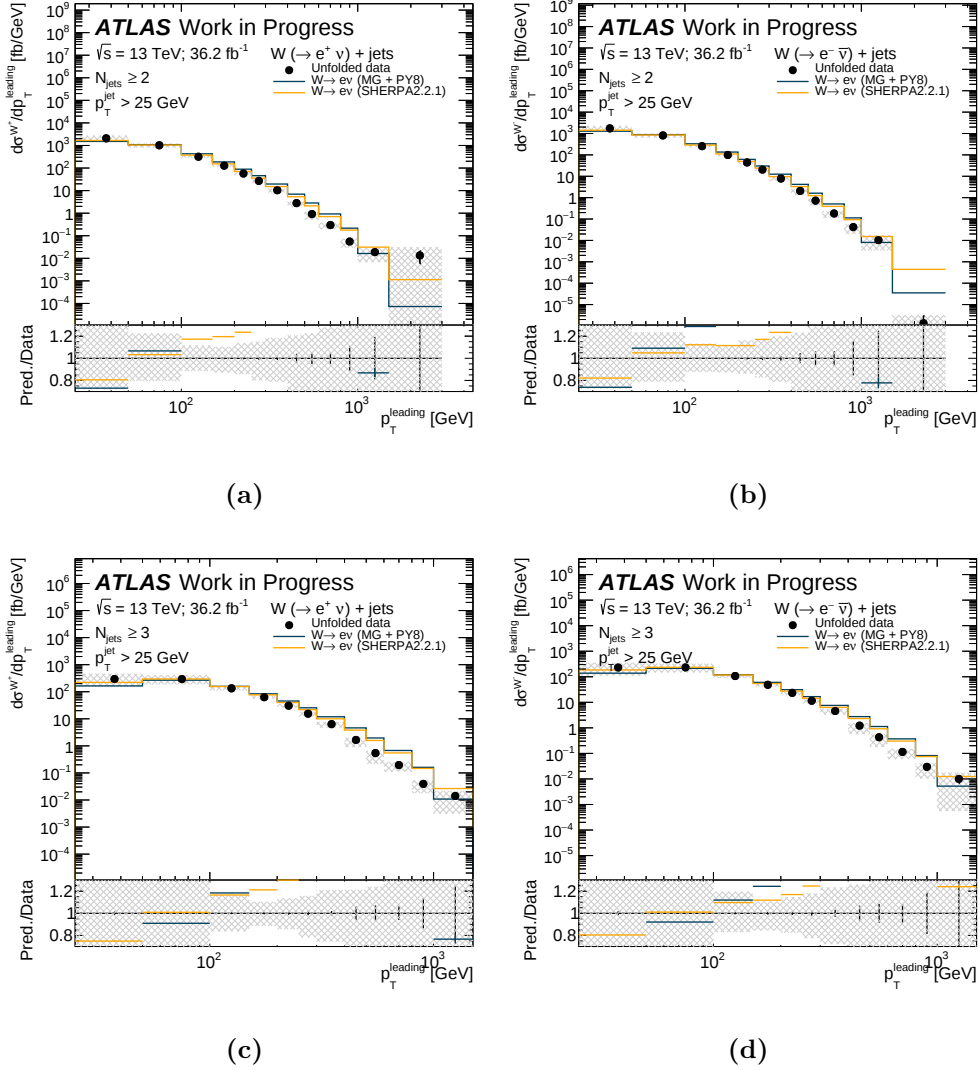


Figure B.9: The unfolded cross section for the production of a W boson in association with (a)-(b) at least two jets and (c)-(d) at least three jets for the $p_{\text{T}}^{\text{leading}}$ observable split in to the W^+ cross section and the W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.

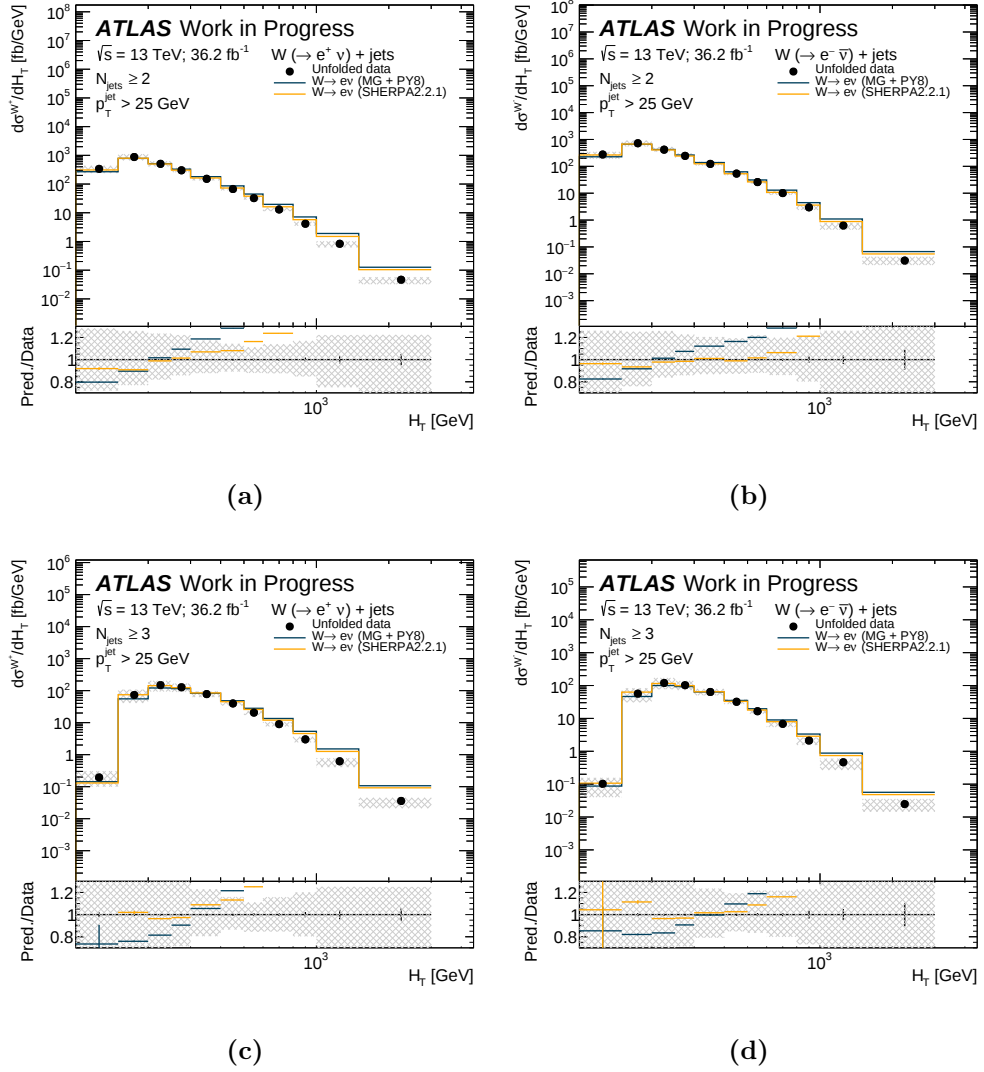
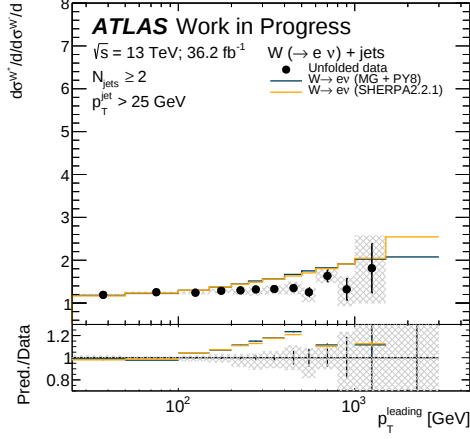
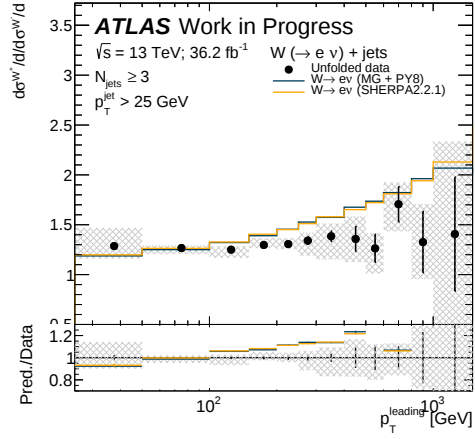


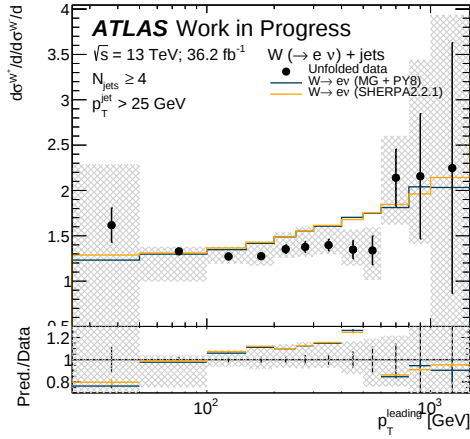
Figure B.10: The unfolded cross section for the production of a W boson in association with (a)-(b) at least two jets and (c)-(d) at least three jets for the H_T observable split in to the W^+ cross section and the W^- cross section. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.



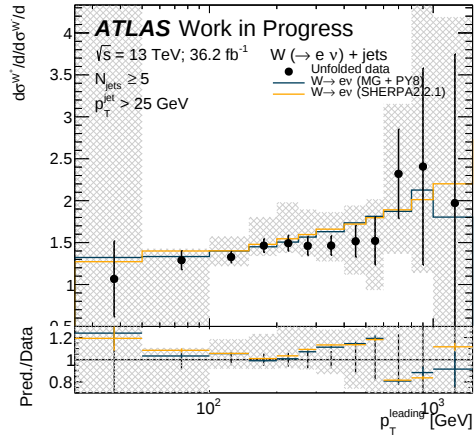
(a)



(b)



(c)



(d)

Figure B.11: The unfolded cross section for the production of a W boson in association with jets for the $p_{\text{T}}^{\text{leading}}$ observable for at least (a) two jets, (b) three jets, (c) four jets and (d) five jets. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.

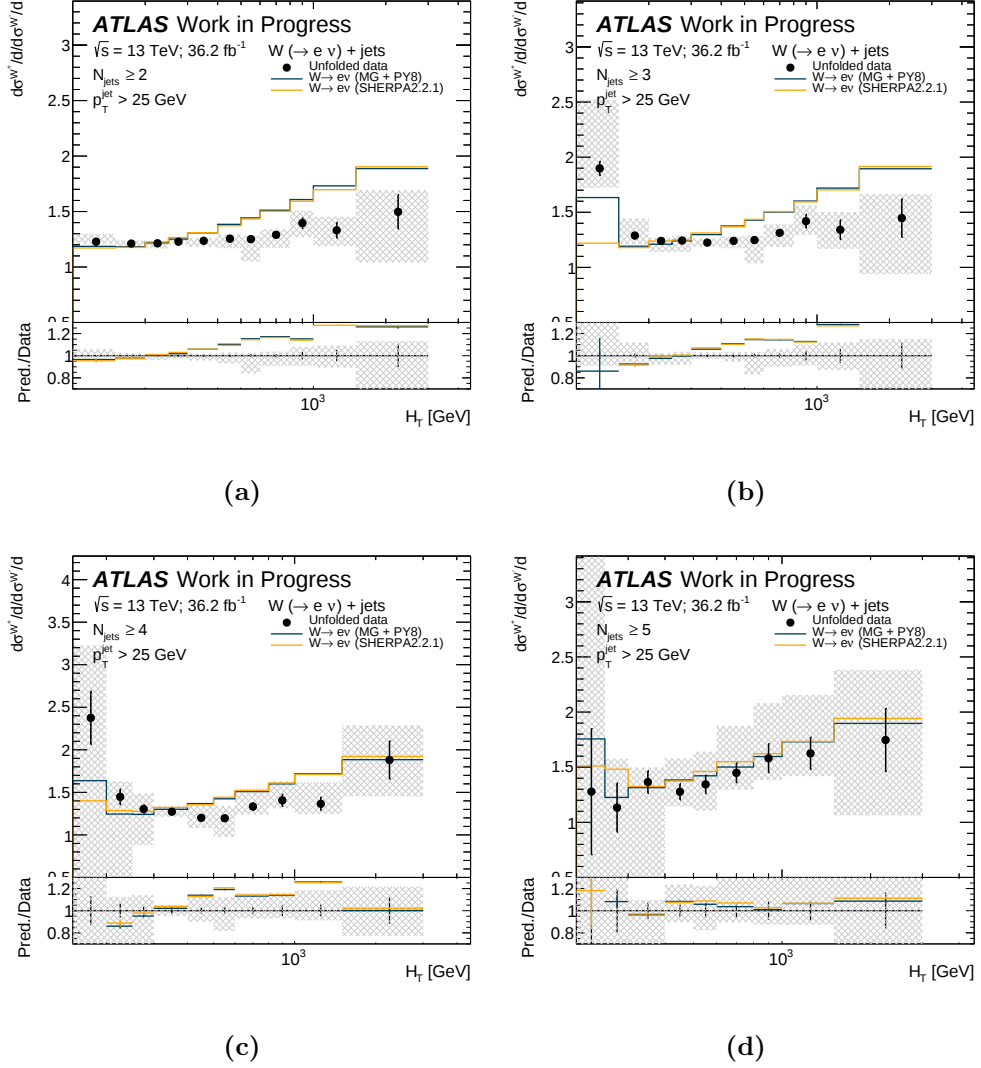


Figure B.12: The unfolded cross section for the production of a W boson in association with jets for the H_T observable for at least (a) two jets, (b) three jets, (c) four jets and (d) five jets. Statistical uncertainties are shown as the vertical black line and the total uncertainty is shown as a gray hatched band. The results are compared to particle-level distributions of the MG + PY8 and SHERPA2.2.1 generators.



ATLASepWZVjet20 parton distribution functions at the Z -mass scale

In chapter 9, the majority of the figures and tables displayed PDFs evaluated at the scale at which the nominal distributions are fit in the ATLASepWZVjet20 PDFs, $Q^2 = 1.9 \text{ GeV}^2$ (also referred to as the “starting scale”). This choice is useful as it allows one to see the precise distribution resulting from the fitted parameters, while at other choices the distributions are convoluted with the DGLAP equations described in section 2.2. Additionally, one of the important results of the ATLASepWZ16 fit is the unsuppressed strange PDF at low x and low Q^2 .

In this appendix, the PDFs are shown evaluated at $Q^2 = m_Z^2$, the scale similar to the renormalisation and factorisation scales of the ATLAS data. In fig. C.1, a comparison between the ATLASepWZ20 and ATLASepWZVjet20 fits is shown for the R_s and $x(\bar{d} - \bar{u})$ distributions. The results are similar to those at the starting scale, with a falling R_s distribution and positive $x(\bar{d} - \bar{u})$ distribution for the ATLASepWZVjet20 fit. Similarly, in fig. C.3, a comparison is shown between the ATLASepWZVjet20 fit and the global PDF sets discussed in chapter 9 for the $x(\bar{d} - \bar{u})$ distribution, and the results are similar to those at the starting scale.

In fig. C.2, a comparison is shown between the ATLASepWZVjet20 fit and the global PDF sets discussed in chapter 9 for the R_s distribution. At this scale, the

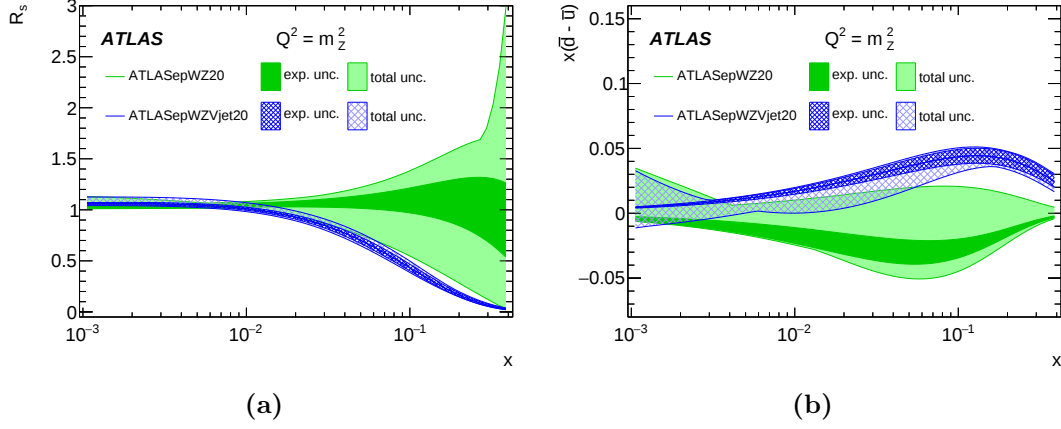


Figure C.1: PDFs at the scale $Q^2 = m_Z^2$ as a function of Bjorken x obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive W and Z , and HERA data (ATLASepWZVjet20, blue bands), compared with a similar fit without $W + \text{jets}$ 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.

global PDFs exhibit a relatively unsuppressed strange PDF, and the CT18A PDF set is consistent with the results of the ATLASepWZVjet20 fit in the range $x \lesssim 0.07$. However, at high $x \gtrsim 0.1$, the discrepancy observed between the ATLASepWZVjet20 fit and the global PDF sets is more significant than that observed at the starting scale.

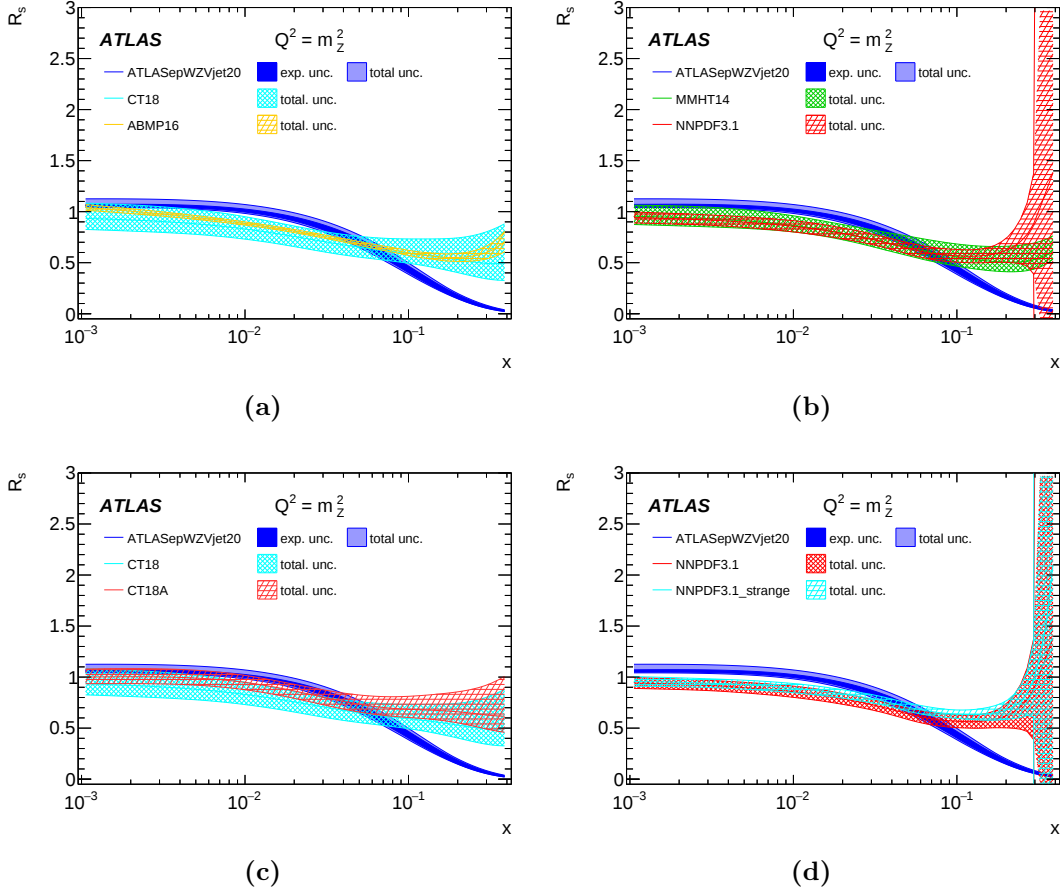


Figure C.2: The $R_s = (s + \bar{s})/(\bar{u} + \bar{d})$ distribution evaluated at $Q^2 = m_Z^2$ as a function of Bjorken x , for the ATLASepWZVjet20 PDF set in comparison with global PDFs (a) ABMP16 and CT18, (b) MMHT14 and NNPDF3.1, and in additional comparisons with (c) CT18 and CT18A, and (d) NNPDF3.1 and NNPDF3.1_strange [135–138, 175]. 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 [177].

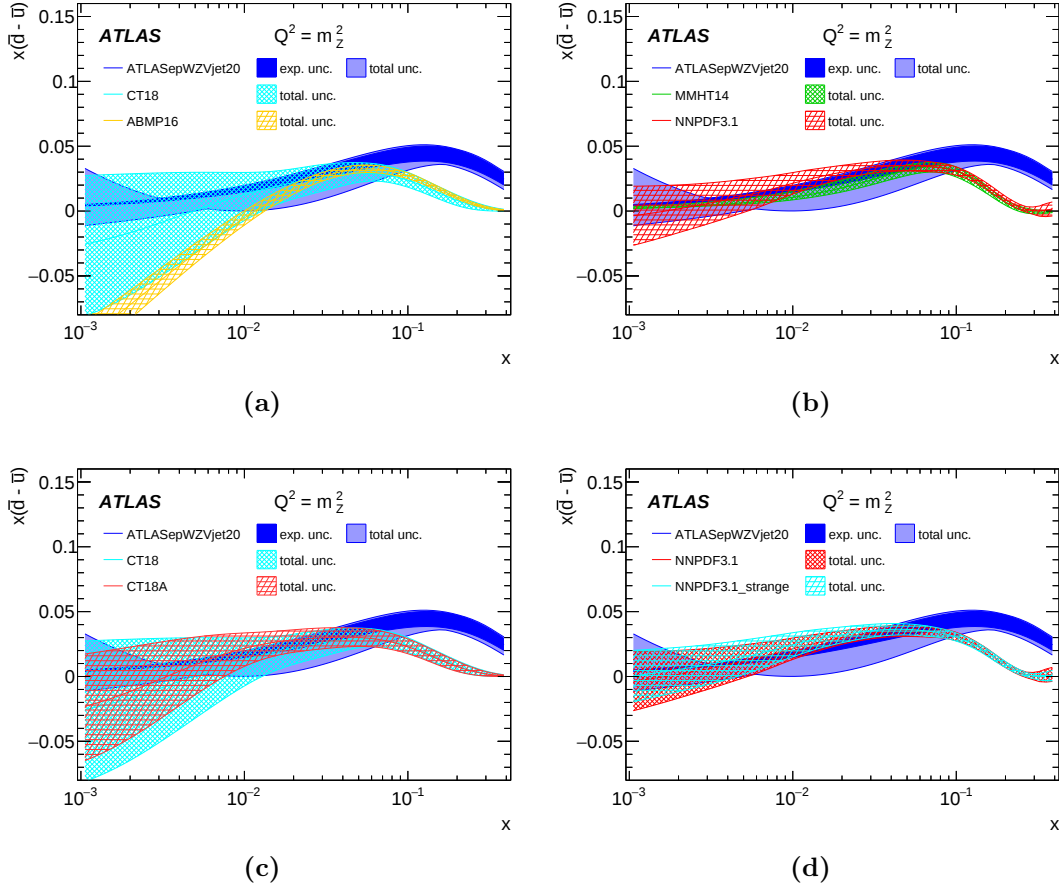


Figure C.3: The $x(\bar{d} - \bar{u})$ distribution evaluated at $Q^2 = m_Z^2$ as a function of Bjorken x , for the ATLASepWZVjet20 PDF set in comparison with global PDFs (a) ABMP16 and CT18, (b) MMHT14 and NNPDF3.1, and in additional comparisons with (c) CT18 and CT18A, and (d) NNPDF3.1 and NNPDF3.1_strange [135–138, 175]. 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 [177].

D

Systematic correlations between data sets in the ATLASepWZVjet20 fit

The correlation model used for the ATLAS data is summarised in table D.1, where the labels used are the same as those in the HEPData entries of the respective ATLAS $W + \text{jets}$ [100, 181] and $Z + \text{jets}$ [132, 182] publications.

Systematic uncertainties related to the jet energy scale, jet energy resolution, rejection of jets from pile-up (JVF), missing transverse momentum (E_T^{miss}) scale, E_T^{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 integrated 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

Systematic Uncertainty	7 TeV inclusive W, Z	8 TeV $W + \text{jets}$	8 TeV $Z + \text{jets}$
Jet energy scale [91]	*	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
		JetScaleNPV	ATL_PU_OffsetNPV
		JetScaleMu	ATL_PU_OffsetMu
Jet punchthrough [91]	-	JetScalepunchT	ATL_PunchThrough
Jet resolution [91]	JetRes	JetResolution10	ATL_JER
Jet flavour composition [91]	-	JetScaleFlav1Known	ATL_Flavor_Comp
Jet flavour response [91]	-	JetScaleFlav2	ATL_Flavor_Response
Pile-up jet rejection (JVF) [109]	-	JetJVFcut	ATL_JVF
E_T^{miss} scale [183]	MetScaleWen	METScale	-
E_T^{miss} resolution [183]	MetRes	METResLong	-
		METResTrans	-
Electron energy scale [184]	*	ElScaleZee	ATL_ElecEnZee
Electron trigger efficiency [185]	*	ElSFTrigger	ATL_Trig
Electron reconstruction efficiency [186, 187]	*	ElSFReco	ATL_RecEff
Electron identification efficiency [186, 187]	*	ElSFId	ATL_IDEff
Luminosity [188, 189]	*	LumiUncert	ATL_lumi_2012.8TeV
WW background cross section [130]	*	XsecDibos	ATL_WW_xs
Top background cross section [131]	*	XsecTop	ATL_ttbar_xs

Table D.1: Correlation model for the systematic uncertainties of the ATLAS measurements of $W + \text{jets}$ and $Z + \text{jets}$ 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.

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)¹ are fully decorrelated between spectra and bins within a single spectrum

¹“Unfolding” here refers to the procedure used to correct reconstructed data to particle-level cross sections.

(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 $W + \text{jets}$ bins, for example, increases the χ^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 to the E_T^{miss} resolution are summed into a single component for one-to-one correlation with the E_T^{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^{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.

To summarise, four changes are made to the original systematics of the $W + \text{jets}$ and/or $Z + \text{jets}$ data in this analysis. The effect of each change on the resulting R_s and $x(\bar{d} - \bar{u})$ distributions is shown in the following order, starting from a fit to the unaltered data:

1. Decorrelating the systematics related to unfolding of the data between spectra and bins within a spectrum, shown in figure fig. D.1.
2. Correlation between the ATLAS $W + \text{jets}$ and $Z + \text{jets}$ data sets, shown in figure fig. D.2.

3. In preparation for correlation to 7 TeV data, summing in quadrature the two systematics related to E_T^{miss} resolution in the $W + \text{jets}$ data, shown in fig. D.3.
4. Finally, a switch from combined to uncombined 7 TeV inclusive W, Z data, and correlation between the systematics related to jets and E_T^{miss} in the 7 TeV and 8 TeV data sets, shown in fig. D.4. This comparison is to the final fit, ATLASepWZVjet20, used in subsequent sections.

In each case, the change in the central distribution is well within experimental uncertainties. In particular, by combining E_T^{miss} systematics in 8 TeV data and correlating to the 7 TeV data by switching from combined to uncombined electron and muon channels, there is insignificant change in both the central position and total uncertainties.

In addition, in the correlation model given in section 9.2, ten 8 TeV jet energy scale (JES) systematics could also have been combined in quadrature for a one-to-one correlation with the same systematic in the 7 TeV data. This decision was not taken as the JES systematic in the 7 TeV is relatively minor, and constraining power could be lost by breaking ten correlations in return for one. In Figure D.5, the PDFs extracted from the ATLASepWZVjet20 fit are compared to the PDFs from a similar fit with the combination of the aforementioned JES systematics, which are subsequently correlated with the corresponding systematic in the 7 TeV data. Similarly, recommendations from the ATLAS JetEtmiss combined performance group for the following potential approximate compositions of the 7 TeV JetScale have been considered and comparisons to the nominal ATLASepWZVjet20 fit are shown in fig. D.6:

- JES_model1: 80% JetScaleEff1 and 20% JetScaleEff2.
- JES_model2: 40% JetScaleEta1 and 60% uncorrelated between 7 TeV and 8 TeV data.
- JES_model3: 40% JetScaleEta2 and 60% uncorrelated between 7 TeV and 8 TeV data.

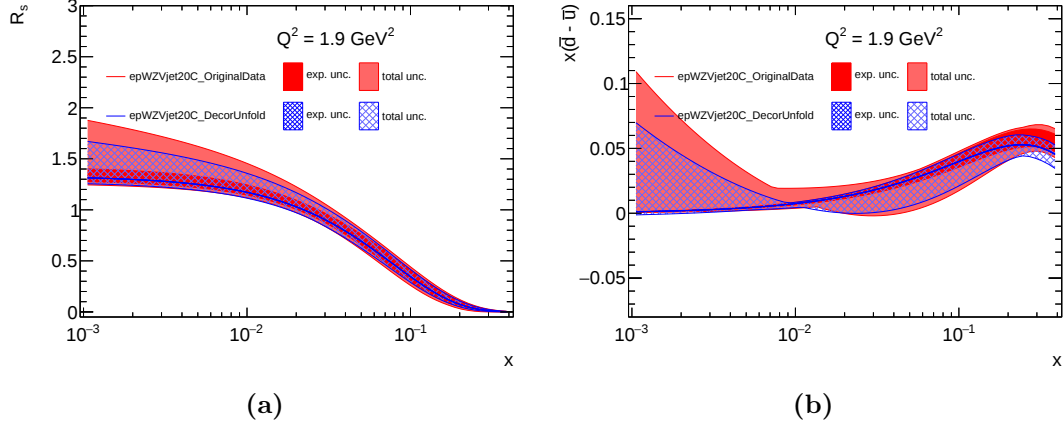


Figure D.1: PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive W, Z and HERA data, with a comparison between a fit to the original data and a similar fit to the data with systematics related to unfolding decorrelated. Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties.

Comparisons between these variations and the nominal ATLASepWZVjet20 fit with only experimental uncertainties plotted are shown in fig. D.6. The results are highly consistent and well within experimental uncertainties.

Finally, in fig. D.7, a fit with the luminosity uncertainty of the 7 TeV and 8 TeV data is decomposed in to two equal-part systematics, in which one is fully correlated between centre-of-mass energies and the other is correlated only between data of the same centre-of-mass energy, is compared to the nominal ATLASepWZVjet20 fit. The resulting PDFs are well within experimental uncertainties.

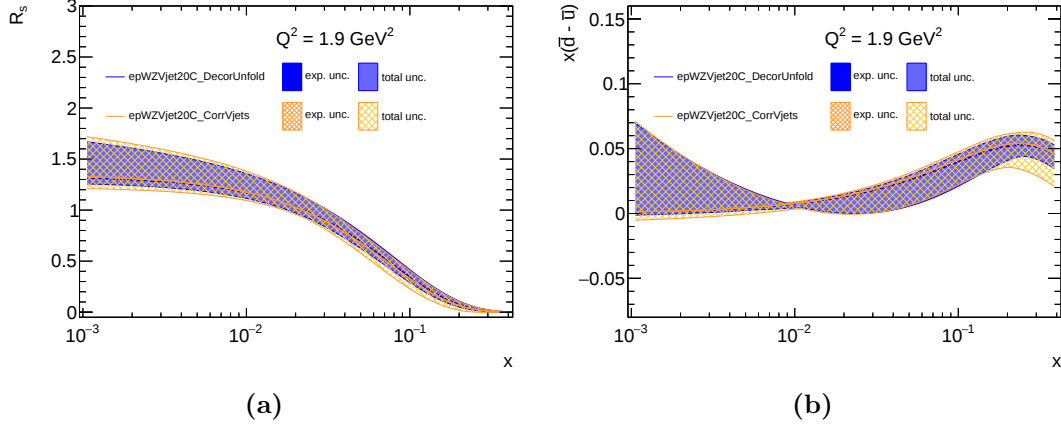


Figure D.2: PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive W, Z and HERA data, with a comparison between a fit to the data with systematics related to unfolding decorrelated, and a fit with correlations between $W + \text{jets}$ and $Z + \text{jets}$ accounted for in addition. Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties.

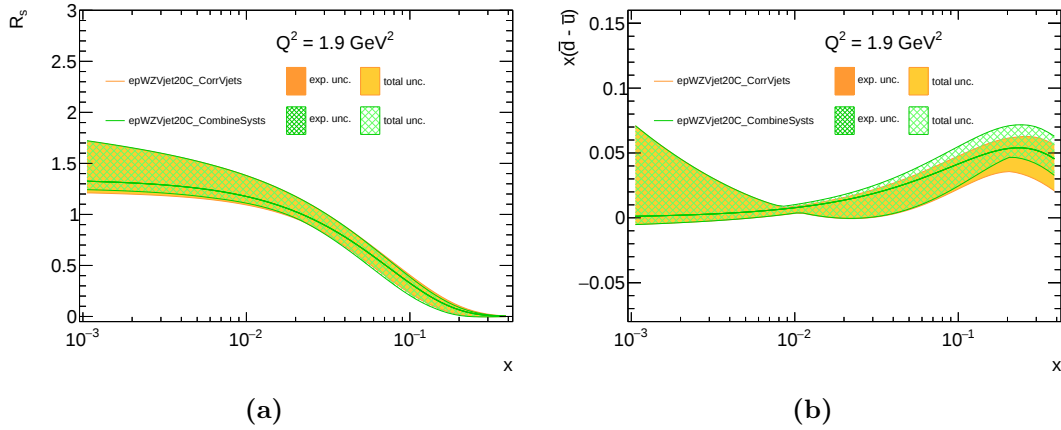


Figure D.3: PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting $W + \text{jets}$, $Z + \text{jets}$, inclusive W, Z and HERA data, with a comparison between a fit to data with systematics related to unfolding decorrelated and correlations between $W + \text{jets}$ and $Z + \text{jets}$ accounted for, and a similar fit with certain systematics in the $V + \text{jets}$ data summed in quadrature in preparation for correlation to 7 TeV W, Z data (ATLASepWZ20C). Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties.

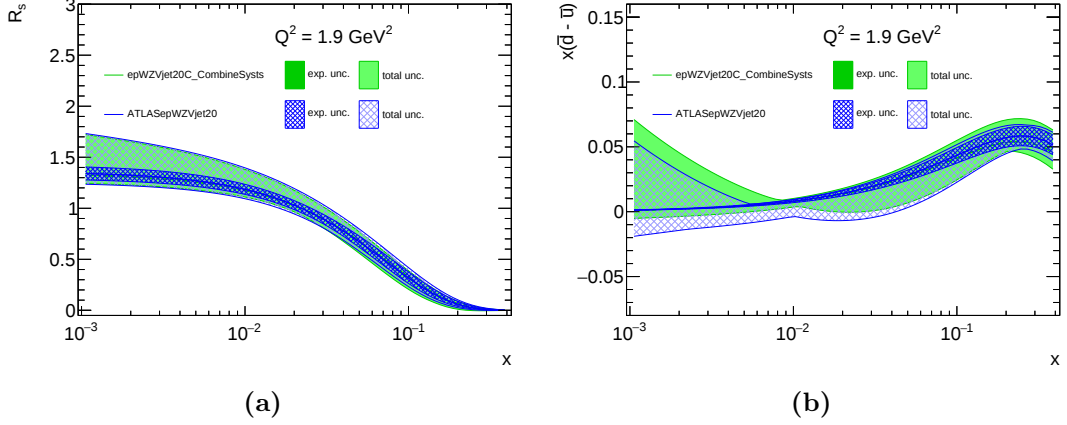


Figure D.4: PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting W + jets, Z + jets, inclusive W, Z and HERA data, with a comparison between the use of combined 7 TeV data and uncombined (where in the uncombined case, correlation to the V + jets data are also accounted for). Inner error bands indicate the experimental uncertainty, while outer error bands indicate the total uncertainty, including parameterisation and model uncertainties.

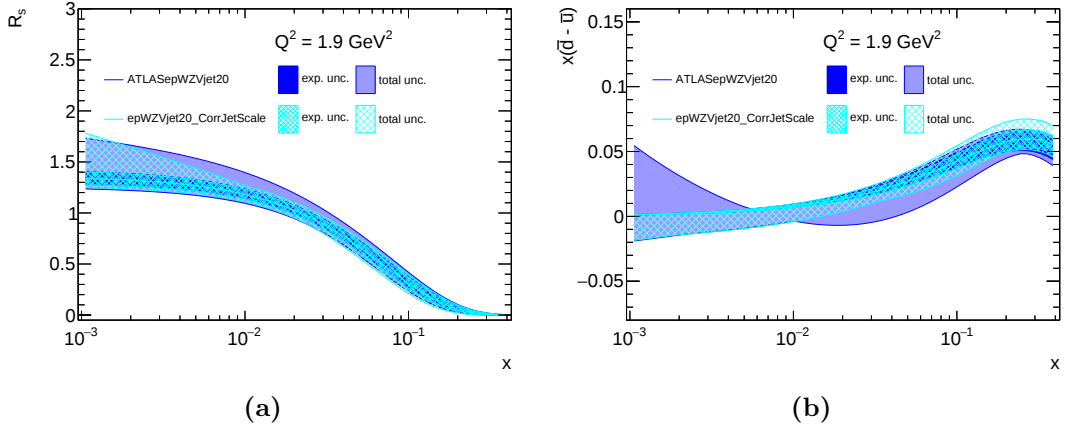


Figure D.5: PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting data with either the 8 TeV jet-scale systematics combined and correlated to the corresponding systematic in the 7 TeV data ($\text{epWZVjet_CorrJetScale}$) or not (ATLASepWZVjet20).

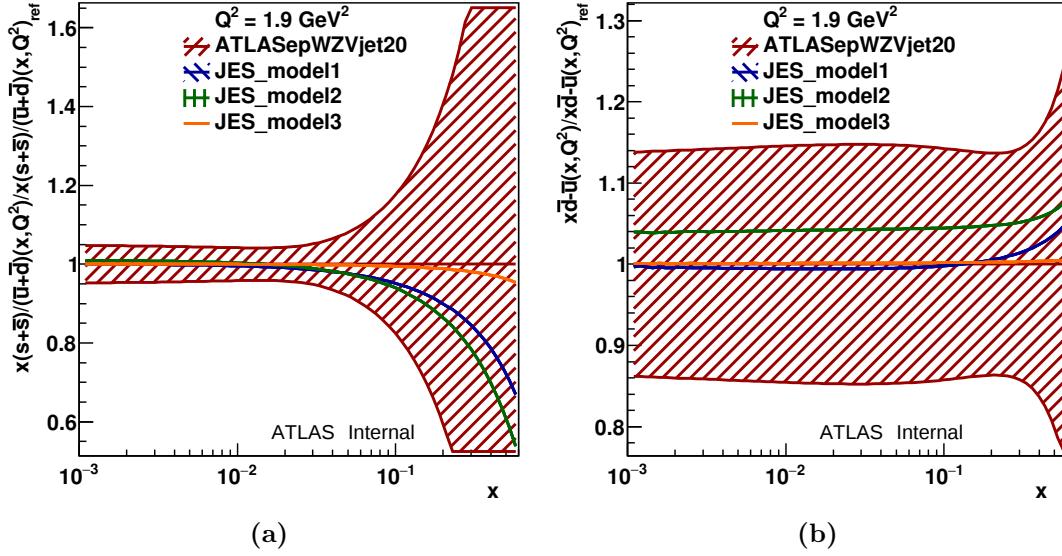


Figure D.6: PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting data with the 7 TeV JES systematic approximately decomposed into multiple components to replicate possible models of correlation with the JES systematics present in the 8 TeV data compared to the ATLASepWZVjet20 fit with experimental uncertainty bands only.

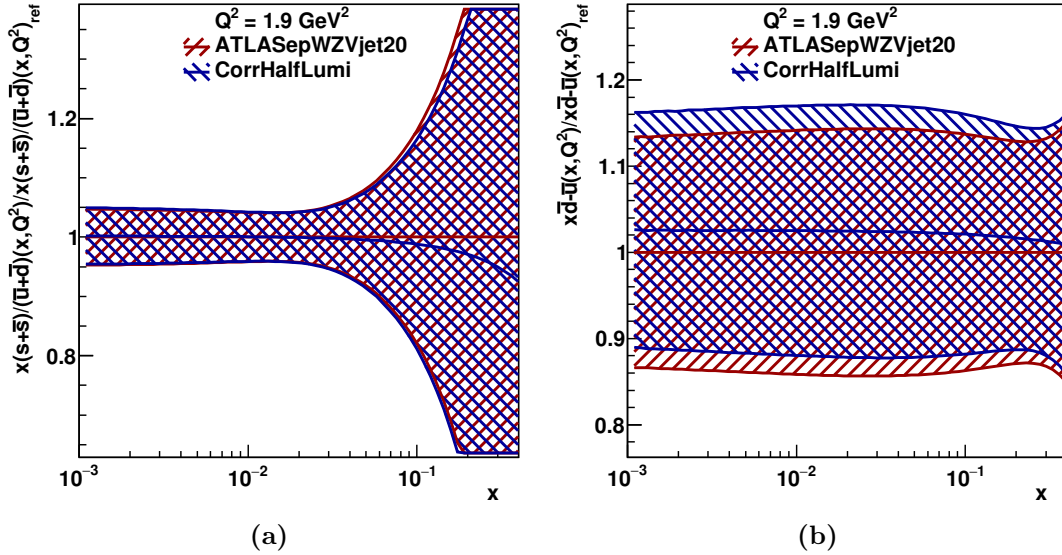


Figure D.7: PDFs obtained for the (a) R_s and (b) $x(\bar{d} - \bar{u})$ distributions when fitting data with the 7 TeV luminosity uncertainty decomposed in to 50% correlated between centre-of-mass energies and 50% uncorrelated compared to the ATLASepWZVjet20 fit with experimental uncertainty bands only.

E

Cross checks of the theoretical predictions of data used in the ATLASepWZVjet20 fit

The scale uncertainties of the 8 TeV $V + \text{jets}$ data used in the ATLASepWZVjet20 fit are available in the following way:

- $W + \text{jets}$: A three-point variation is performed where the factorisation and renormalisation scales take half and double their nominal values simultaneously for calculation of the NNLO cross-section. The uncertainty in the data is given as an asymmetric envelope of these variations, where the largest deviation upward or downward is taken for each point as the respective upward or downward uncertainty.
- $Z + \text{jets}$: As for the $W + \text{jets}$ case, but the variation is seven-point where the factorisation and renormalisation scales take half and double their nominal values individually, but one is never more than twice the other.

In both cases, these variations are taken as additional upward and downward K -factors, shown in fig. E.1.

This study is limited as, for each point, it is not known which variation gave the uncertainty, whether upward or downward. Furthermore, in the case where multiple deviations are in the same direction, it is not known which was the largest

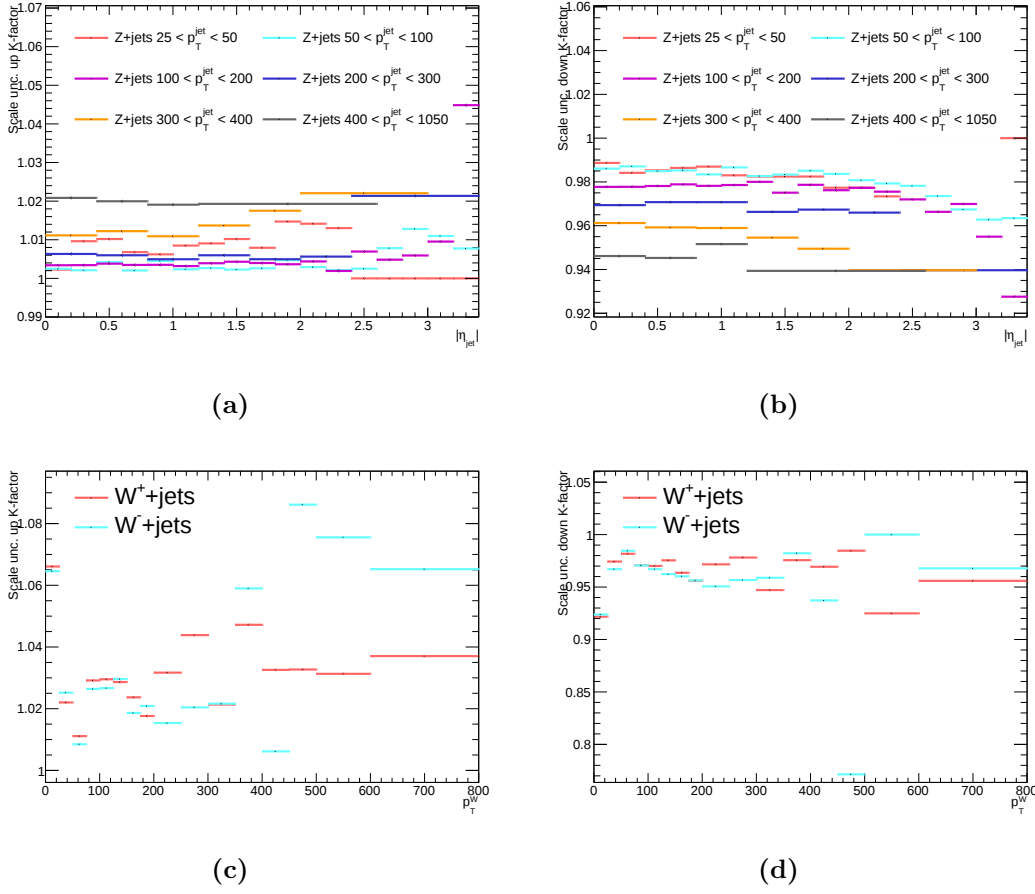


Figure E.1: The scale-variation K -factors of the (a) $Z + \text{jets}$ upward, (b) $Z + \text{jets}$ downward, (c) $W + \text{jets}$ upward and (d) $W + \text{jets}$ downward.

and therefore which gave the uncertainty. In addition, the $W + \text{jets}$ uncertainties exhibit large fluctuations which appear statistical in nature. For these reasons, the variations in scale are not included in the analysis.

Plots comparing the central values of the extracted PDFs using these variations are shown as a ratio to the ATLASepWZVjet20 PDFs with experimental uncertainties only in figs. E.2 and E.3, where the $W + \text{jets}$ and $Z + \text{jets}$ predictions have been varied together and separately, respectively. The effect of these variations is small, contained well inside the experimental uncertainties for all distributions.

In contrast, in the assessment of the ATLASepWZ16 PDF set, multiple cross-checks using alternative theoretical predictions were possible, including coherent variation of renormalisation and factorisation scales, variations in the electroweak theory and the use of predictions by FEWZ to produce NNLO K -factors as opposed

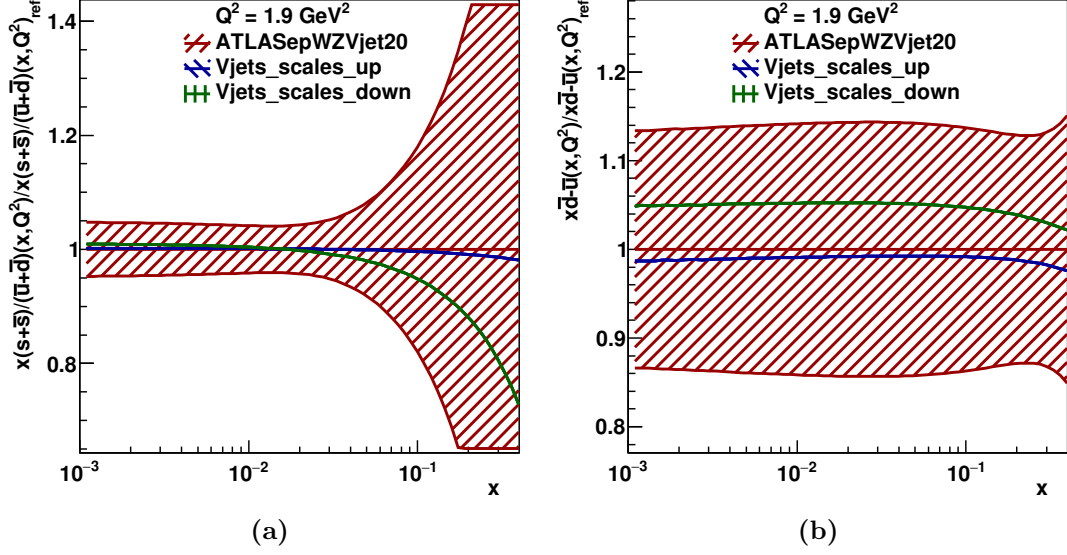


Figure E.2: Ratio plots of the ATLASepWZVjet20 PDF set with experimental uncertainties only compared to the central values of similar fits corrected by the scale-variation K -factors of fig. E.1 for the (a) up valence, (b) down valence, (c) gluon, (d) $\bar{u}$, (e) $\bar{d}$, (f) $\bar{s}$, (g) R_s , (h) r_s and $x(\bar{d} - \bar{u})$ distributions.

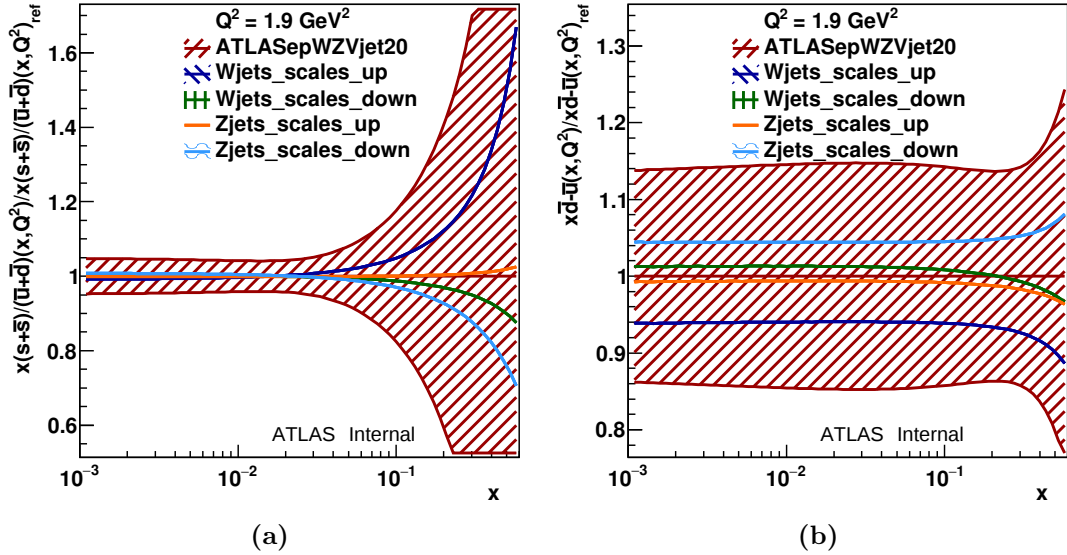


Figure E.3: Ratio plots of the ATLASepWZVjet20 PDF set with experimental uncertainties only compared to the central values of similar fits corrected by the scale-variation K -factors of fig. E.1 for the (a) up valence, (b) down valence, (c) gluon, (d) $\bar{u}$, (e) $\bar{d}$, (f) $\bar{s}$, (g) R_s , (h) r_s and $x(\bar{d} - \bar{u})$ distributions.

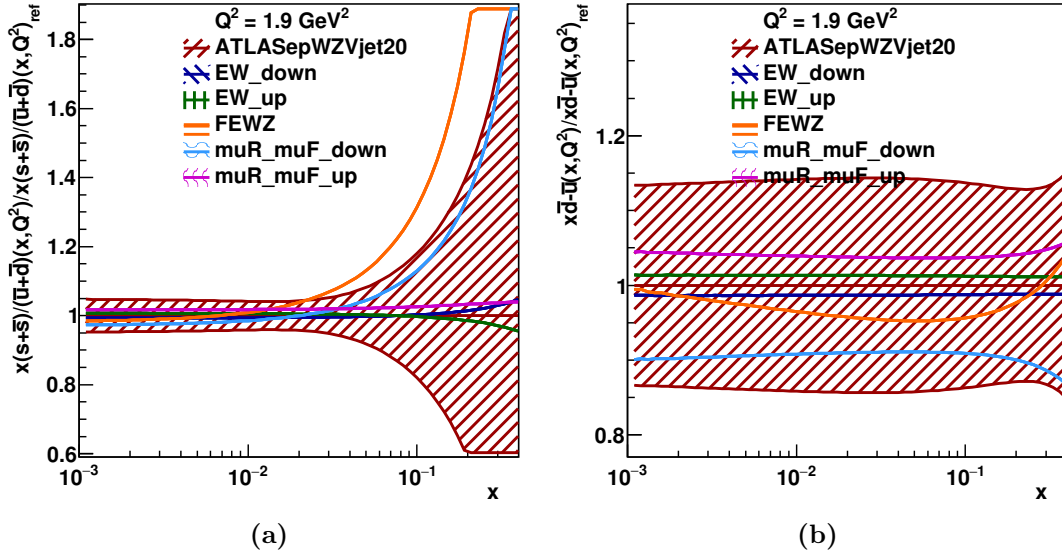


Figure E.4: Ratio plots of the ATLASepWZVjet20 PDF set with experimental uncertainties only compared to the central values of similar fits with variations in the prediction of the 7 TeV W, Z data for the (a) up valence, (b) down valence, (c) gluon, (d) $\bar{u}$, (e) $\bar{d}$, (f) $\bar{s}$, (g) R_s , (h) r_s and $x(\bar{d} - \bar{u})$ distributions.

to DYNNLO. Detailed discussion of these variations are in the ATLAS publication [145]; ultimately, the result of the unsuppressed strange sea at low- x was maintained irrespective of these variations.

The aim of the study is to assess the impact of the 8 TeV $V + \text{jets}$ data on the current state of the ATLAS PDFs. For this reason, the various cross-checks of the W, Z theory, and evaluation of resulting uncertainties, is not performed as part of the main analysis. However, in this section, the resulting PDFs with these variations applied are presented in comparison to the ATLASepWZVjet20 fit in figs. E.4 and E.5.

PDFs evaluated with the use of the FEWZ predictions exhibit a similar effect as in the ATLASepWZ16 fit, where an upward deviation in the strange distribution is observed at high- x and this is propagated to the strange ratios, R_s and r_s . These results are highly consistent with the key observations of the ATLASepWZVjet20 analysis, that the strange sea remains unsuppressed at low- x with a falling ratio at high- x and a positive $\bar{d} - \bar{u}$ distribution. Other variations do not exhibit a significant effect.

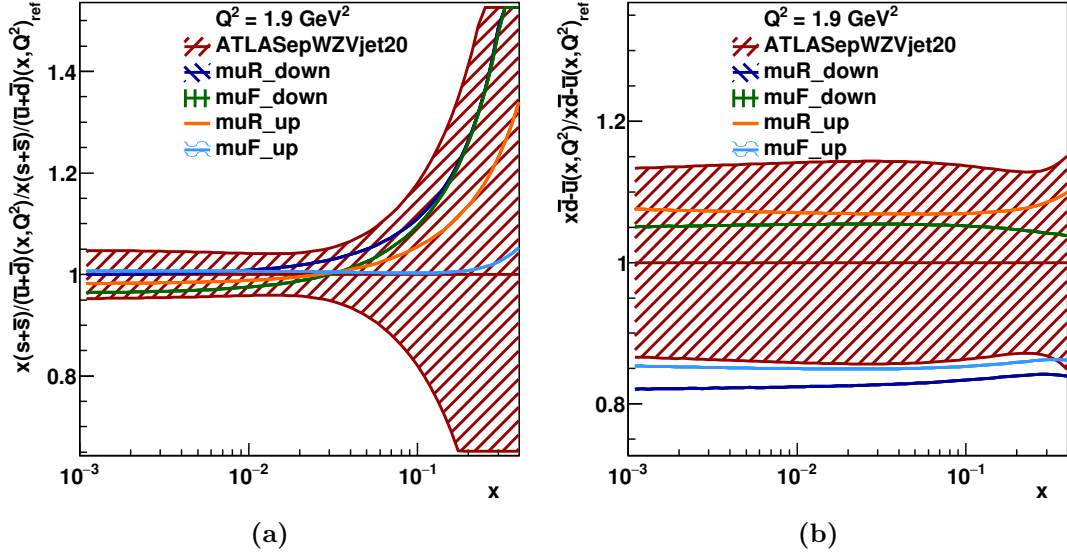


Figure E.5: Ratio plots of the ATLASepWZVjet20 PDF set with experimental uncertainties only compared to the central values of similar fits with variations in the prediction of the 7 TeV W, Z data for the (a) up valence, (b) down valence, (c) gluon, (d) $\bar{u}$, (e) $\bar{d}$, (f) $\bar{s}$, (g) R_s , (h) r_s and $x(\bar{d} - \bar{u})$ distributions.

References

- [1] T.C.C.W. Taylor et al. *The Atomists, Leucippus and Democritus: Fragments : a Text and Translation with a Commentary*. Phoenix pre-Socratics. University of Toronto Press, 1999. URL: <https://books.google.ch/books?id=bIqVFkbD8QYC>.
- [2] A. Kenny. *Ancient philosophy*. New history of Western philosophy. Clarendon Press, 2004. URL: <https://books.google.ch/books?id=AJ8WAQAAMAAJ>.
- [3] Golden Nyambuya, A Dube, and Godwin Musosi. “Salvaging Newton’s 313 Year Old Corpuscular Theory of Light”. In: (Sept. 2017).
- [4] Alan J. Rocke. “In Search of El Dorado: John Dalton and the Origins of the Atomic Theory”. In: *Social Research* 72.1 (2005), pp. 125–158. URL: <http://www.jstor.org/stable/40972005>.
- [5] B.W. Pfennig. *Principles of Inorganic Chemistry*. Wiley, 2015. URL: <https://books.google.ch/books?id=pxX1BgAAQBAJ>.
- [6] E. Rutherford. “The scattering of alpha and beta particles by matter and the structure of the atom”. In: *Philosophical Magazine* 21 (1911), pp. 669–688.
- [7] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), p. 1. arXiv: 1207.7214 [hep-ex].
- [8] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), p. 30. arXiv: 1207.7235 [hep-ex].
- [9] S. Glashow. “Partial-symmetries of weak interactions”. In: *Nucl. Phys.* 22.4 (1961), pp. 579–588.
- [10] S. Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266.
- [11] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323.
- [12] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509.
- [13] David J. Gross and Frank Wilczek. “Asymptotically Free Gauge Theories. I”. In: *Phys. Rev. D* 8 (10 Nov. 1973), pp. 3633–3652. URL: <https://link.aps.org/doi/10.1103/PhysRevD.8.3633>.
- [14] David J. Gross and Frank Wilczek. “Asymptotically free gauge theories. II”. In: *Phys. Rev. D* 9 (4 Feb. 1974), pp. 980–993. URL: <https://link.aps.org/doi/10.1103/PhysRevD.9.980>.

- [15] UA1 Collaboration. “Experimental observation of isolated large transverse energy electrons with associated missing energy at $s=540$ GeV”. In: *Physics Letters B* 122.1 (1983), pp. 103–116.
- [16] AU2 Collaboration. “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider”. In: *Physics Letters B* 122.5 (1983), pp. 476–485.
- [17] UA1 Collaboration. “Experimental observation of lepton pairs of invariant mass around 95 GeV/c² at the CERN SPS collider”. In: *Physics Letters B* 126.5 (1983), pp. 398–410.
- [18] AU2 Collaboration. “Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN pp collider”. In: *Physics Letters B* 129.1 (1983), pp. 130–140.
- [19] CDF Collaboration. “Observation of top quark production in $\bar{p}p$ collisions”. In: *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631. arXiv: [hep-ex/9503002](#).
- [20] D0 Collaboration. “Observation of the top quark”. In: *Phys. Rev. Lett.* 74 (1995), pp. 2632–2637. arXiv: [hep-ex/9503003](#).
- [21] DONUT Collaboration. “Observation of tau neutrino interactions”. In: *Phys. Lett. B* 504 (2001), pp. 218–224. arXiv: [hep-ex/0012035](#).
- [22] Particle Data Group. “Review of Particle Physics”. In: *Chinese Physics C* 40 (2016), p. 100001.
- [23] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415.
- [24] M. Gell-Mann. “A schematic model of baryons and mesons”. In: *Physics Letters* 8.3 (1964), pp. 214–215.
- [25] G Zweig. “An SU₃ model for strong interaction symmetry and its breaking; Version 2”. In: CERN-TH-412 (Feb. 1964), 80 p. URL: <http://cds.cern.ch/record/570209>.
- [26] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. In: *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. URL: <https://link.aps.org/doi/10.1103/PhysRevD.2.1285>.
- [27] The ALEPH Collaboration et al. “Precision Electroweak Measurements on the Z Resonance”. In: (Sept. 2005). arXiv: [hep-ex/0509008](#) [[hep-ex](#)].
- [28] ATLAS Collaboration. “Combined measurements of Higgs boson production and decay using up to 80 fb⁻¹ of proton–proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment”. In: *Phys. Rev. D* 101 (2020), p. 012002. arXiv: [1909.02845](#) [[hep-ex](#)].
- [29] CMS Collaboration. “Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 421. arXiv: [1809.10733](#) [[hep-ex](#)].
- [30] V.N. Gribov and L.N. Lipatov. “Deep inelastic e p scattering in perturbation theory”. In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.
- [31] G. Altarelli and G. Parisi. “Asymptotic freedom in parton language”. In: *Nuclear Physics B* 126.2 (1977), pp. 298–318.

- [32] Yuri L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and $e^+ e^-$ Annihilation by Perturbation Theory in Quantum Chromodynamics.” In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.
- [33] G. Curci, W. Furmanski, and R. Petronzio. “Evolution of parton densities beyond leading order: The non-singlet case”. In: *Nuclear Physics B* 175.1 (1980), pp. 27–92.
- [34] A. Vogt, S. Moch, and J. A. M. Vermaseren. “The three-loop splitting functions in QCD: the singlet case”. In: *Nuclear Physics B* 691.1 (July 2004), pp. 129–181.
- [35] John C. Collins, Davison E. Soper, and George F. Sterman. “Factorization of Hard Processes in QCD”. In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. arXiv: hep-ph/0409313.
- [36] C.F. Berger et al. “Next-to-Leading Order QCD Predictions for $W+3$ -Jet Distributions at Hadron Colliders”. In: *Phys. Rev. D* 80 (2009), p. 074036. arXiv: 0907.1984 [hep-ph].
- [37] C.F. Berger et al. “Precise Predictions for $W + 4$ Jet Production at the Large Hadron Collider”. In: *Phys. Rev. Lett.* 106 (2011), p. 092001. arXiv: 1009.2338 [hep-ph].
- [38] Z. Bern et al. “Next-to-Leading Order $W + 5$ -Jet Production at the LHC”. In: *Phys. Rev. D* 88.1 (2013), p. 014025. arXiv: 1304.1253 [hep-ph].
- [39] Radja Boughezal et al. “ W -boson production in association with a jet at next-to-next-to-leading order in perturbative QCD”. In: *Phys. Rev. Lett.* 115.6 (2015), p. 062002. arXiv: 1504.02131 [hep-ph].
- [40] Radja Boughezal, Xiaohui Liu, and Frank Petriello. “ W -boson plus jet differential distributions at NNLO in QCD”. In: *Phys. Rev. D* 94.11 (2016), p. 113009. arXiv: 1602.06965 [hep-ph].
- [41] A. Gehrmann-De Ridder et al. “Precise QCD predictions for the production of a Z boson in association with a hadronic jet”. In: *Phys. Rev. Lett.* 117.2 (2016), p. 022001. arXiv: 1507.02850 [hep-ph].
- [42] Aude Gehrmann-De Ridder et al. “NNLO QCD Corrections to W +jet Production in NNLOJET”. In: *PoS LL2018* (2018), p. 041. arXiv: 1807.09113 [hep-ph].
- [43] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. “MINLO: Multi-Scale Improved NLO”. In: *JHEP* 10 (2012), p. 155. arXiv: 1206.3572 [hep-ph].
- [44] Rikkert Frederix and Stefano Frixione. “Merging meets matching in MC@NLO”. In: *JHEP* 12 (2012), p. 061. arXiv: 1209.6215 [hep-ph].
- [45] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. arXiv: 1405.0301 [hep-ph].
- [46] Stefan Höche and Stefan Prestel. “The midpoint between dipole and parton showers”. In: *Eur. Phys. J. C* 75.9 (2015), p. 461. arXiv: 1506.05057 [hep-ph].
- [47] Walter T. Giele, David A. Kosower, and Peter Z. Skands. “A simple shower and matching algorithm”. In: *Phys. Rev. D* 78 (2008), p. 014026. arXiv: 0707.3652 [hep-ph].

- [48] ATLAS Collaboration. *Standard Model Summary Plots Spring 2020*. ATL-PHYS-PUB-2020-010. 2020. URL: <http://cds.cern.ch/record/2718937>.
- [49] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001.
- [50] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08003.
- [51] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004.
- [52] The LHCb Collaboration. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008), S08005.
- [53] The ALICE Collaboration. “The ALICE Experiment at the CERN LHC”. In: *JINST* 3 (2008), S08002.
- [54] The LHCf Collaboration. “The LHCf Detector at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08006.
- [55] The TOTEM Collaboration. “The TOTEM Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08007.
- [56] The MoEDAL Collaboration. “Technical Design Report of the MoEDAL Experiment”. In: (2009).
- [57] K. Wille. *The physics of particle accelerators: An introduction*. 2000.
- [58] Fabienne Marcastel. “CERN’s Accelerator Complex. La chaîne des accélérateurs du CERN”. In: (Oct. 2013). General Photo. URL: <https://cds.cern.ch/record/1621583>.
- [59] *Pseudorapidity*. URL: <https://en.wikipedia.org/wiki/Pseudorapidity%7D>.
- [60] ATLAS Collaboration. *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/331063>.
- [61] G. Aad et al. “ATLAS pixel detector electronics and sensors”. In: *JINST* 3 (2008), P07007.
- [62] ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report*. ATLAS-TDR-19; CERN-LHCC-2010-013. 2010. URL: <https://cds.cern.ch/record/1291633>.
- [63] ATLAS Collaboration. “Operation and performance of the ATLAS semiconductor tracker”. In: *JINST* 9 (2014), P08009. arXiv: 1404.7473 [hep-ex].
- [64] ATLAS Collaboration. “Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties”. In: *JINST* 12 (2017), P05002. arXiv: 1702.06473 [hep-ex].
- [65] ATLAS Collaboration. “ATLAS Computing: Technical Design Report”. In: (2005). URL: <https://cds.cern.ch/record/837738>.
- [66] The ATLAS collaboration. “Operation and performance of the ATLAS SemiConductor tracker”. In: *Journal of Instrumentation* 9.08 (2014), P08009. URL: <http://stacks.iop.org/1748-0221/9/i=08/a=P08009>.

- [67] A Abdesselam and et al. “The optical links of the ATLAS SemiConductor Tracker”. In: *Journal of Instrumentation* 2.09 (2007), P09003. URL: <http://stacks.iop.org/1748-0221/2/i=09/a=P09003>.
- [68] I. Dawson et al. “In situ radiation damage studies of optoelectronics in the ATLAS SemiConductor Tracker”. In: *Journal of Instrumentation* 14.07 (July 2019), P07014–P07014. URL: <https://doi.org/10.1088%2F1748-0221%2F14%2F07%2Fp07014>.
- [69] J Beringer et al. “Radiation hardness and lifetime studies of LEDs and VCSELs for the optical readout of the ATLAS SCT”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 435.3 (1999), pp. 375–392. URL: <http://www.sciencedirect.com/science/article/pii/S0168900299005707>.
- [70] P.K Teng et al. “Radiation hardness and lifetime studies of the VCSELs for the ATLAS SemiConductor Tracker”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 497.2 (2003), pp. 294–304. URL: <http://www.sciencedirect.com/science/article/pii/S0168900202019228>.
- [71] G. Lindstroem, M. Moll, and E. Fretwurst. “Radiation hardness of silicon detectors – a challenge from high-energy physics”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 426.1 (1999), pp. 1–15. URL: <http://www.sciencedirect.com/science/article/pii/S0168900298014624>.
- [72] P Stejskal et al. “Modelling radiation-effects and annealing in semiconductor lasers for use in future particle physics experiments”. In: *Journal of Instrumentation* 6.12 (2011), p. C12045. URL: <http://stacks.iop.org/1748-0221/6/i=12/a=C12045>.
- [73] J.D Dowell et al. “Irradiation tests of photodiodes for the ATLAS SCT readout”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 424.2 (1999), pp. 483–494. URL: <http://www.sciencedirect.com/science/article/pii/S0168900298014156>.
- [74] D.G. Charlton et al. “Radiation hardness and lifetime studies of photodiodes for the optical readout of the ATLAS semiconductor tracker”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 456.3 (2001), pp. 300–309. URL: <http://www.sciencedirect.com/science/article/pii/S0168900200006665>.
- [75] A Ferrari et al. *FLUKA: A multi-particle transport code (program version 2005)*. CERN Yellow Reports: Monographs. Geneva: CERN, 2005. URL: <http://cds.cern.ch/record/898301>.
- [76] ATLAS Collaboration. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 639. arXiv: 1902.04655 [hep-ex].

- [77] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data”. In: *JINST* 14 (2019), P03017. arXiv: 1812.03848 [hep-ex].
- [78] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. D* 96 (2017), p. 072002. arXiv: 1703.09665 [hep-ex].
- [79] ATLAS Collaboration. “Measurements of b -jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV”. In: *JHEP* 08 (2018), p. 089. arXiv: 1805.01845 [hep-ex].
- [80] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$, TeV with the ATLAS detector”. In: (2020). arXiv: 2007.02645 [hep-ex].
- [81] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 78 (2018), p. 903. arXiv: 1802.08168 [hep-ex].
- [82] Walter Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*. ATL-LARG-PUB-2008-002. 2008. URL: <https://cds.cern.ch/record/1099735>.
- [83] ATLAS Collaboration. “Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2”. In: *Eur. Phys. J. C* 77 (2017), p. 673. arXiv: 1704.07983 [hep-ex].
- [84] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti- k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. arXiv: 0802.1189 [hep-ph].
- [85] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “FastJet user manual”. In: *Eur. Phys. J. C* 72 (2012), p. 1896. arXiv: 1111.6097 [hep-ph].
- [86] Gavin P. Salam. “Towards Jetography”. In: *Eur. Phys. J. C* 67 (2010), pp. 637–686. arXiv: 0906.1833 [hep-ph].
- [87] S. Catani et al. “New clustering algorithm for multijet cross sections in $e+e$ -annihilation”. In: *Physics Letters B* 269.3 (1991), pp. 432–438. URL: <http://www.sciencedirect.com/science/article/pii/037026939190196W>.
- [88] S. Catani et al. “Longitudinally-invariant $k_{\perp}$ -clustering algorithms for hadron-hadron collisions”. In: *Nuclear Physics B* 406.1 (1993), pp. 187–224. URL: <http://www.sciencedirect.com/science/article/pii/055032139390166M>.
- [89] Stephen D. Ellis and Davison E. Soper. “Successive combination jet algorithm for hadron collisions”. In: *Phys. Rev. D* 48 (1993), pp. 3160–3166. arXiv: hep-ph/9305266.
- [90] Yuri L. Dokshitzer et al. “Better jet clustering algorithms”. In: *JHEP* 08 (1997), p. 001. arXiv: hep-ph/9707323.
- [91] ATLAS Collaboration. “Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 75 (2015), p. 17. arXiv: 1406.0076 [hep-ex].
- [92] ATLAS Collaboration. *Tagging and suppression of pileup jets*. ATL-PHYS-PUB-2014-001. 2014. URL: <https://cds.cern.ch/record/1643929>.

- [93] ATLAS Collaboration. *Pile-up subtraction and suppression for jets in ATLAS*. ATLAS-CONF-2013-083. 2013. URL: <https://cds.cern.ch/record/1570994>.
- [94] M Baak et al. *Data Quality Status Flags and Good Run Lists for Physics Analysis in ATLAS*. Tech. rep. ATL-COM-GEN-2009-015. Geneva: CERN, Mar. 2009. URL: <https://cds.cern.ch/record/1168026>.
- [95] ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. In: *Eur. Phys. J. C* 77 (2017), p. 332. arXiv: 1611.10235 [hep-ex].
- [96] F. Meloni. “Primary vertex reconstruction with the ATLAS detector”. In: *Journal of Instrumentation* 11.12 (Dec. 2016), pp. C12060–C12060. URL: <https://doi.org/10.1088%2F1748-0221%2F11%2F12%2Fc12060>.
- [97] ATLAS Collaboration. “Performance of electron and photon triggers in ATLAS during LHC Run 2”. In: *Eur. Phys. J. C* 80 (2020), p. 47. arXiv: 1909.00761 [hep-ex].
- [98] Aranzazu Ruiz-Martinez and ATLAS Collaboration. *The Run-2 ATLAS Trigger System*. Tech. rep. ATL-DAQ-PROC-2016-003. Geneva: CERN, Feb. 2016. URL: <https://cds.cern.ch/record/2133909>.
- [99] ATLAS Collaboration. *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. ATLAS-CONF-2019-021. 2019. URL: <https://cds.cern.ch/record/2677054>.
- [100] 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”. In: *JHEP* 05 (2018), p. 077. arXiv: 1711.03296 [hep-ex].
- [101] Rikkert Frederix et al. “T-odd Asymmetry in W +jet Events at the LHC”. In: *Phys. Rev. Lett.* 113 (2014), p. 152001. arXiv: 1407.1016 [hep-ph].
- [102] ATLAS Collaboration. *ATLAS simulation of boson plus jets processes in Run 2*. ATL-PHYS-PUB-2017-006. 2017. URL: <https://cds.cern.ch/record/2261937>.
- [103] T. Sjöstrand, S. Mrenna, and P. Skands. “A brief introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. arXiv: 0710.3820 [hep-ph].
- [104] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. 2014. URL: <https://cds.cern.ch/record/1966419>.
- [105] Richard D. Ball et al. “Parton distributions with LHC data”. In: *Nucl. Phys. B* 867 (2013), p. 244. arXiv: 1207.1303 [hep-ph].
- [106] Marek Schönherr and Frank Krauss. “Soft photon radiation in particle decays in SHERPA”. In: *JHEP* 12 (2008), p. 018. arXiv: 0810.5071 [hep-ph].
- [107] Enrico Bothmann et al. “Event Generation with Sherpa 2.2”. In: *SciPost Phys.* 7.3 (2019), p. 034. arXiv: 1905.09127 [hep-ph].
- [108] Richard D. Ball et al. “Parton distributions for the LHC run II”. In: *JHEP* 04 (2015), p. 040. arXiv: 1410.8849 [hep-ph].

- [109] ATLAS Collaboration. “Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”. In: *Eur. Phys. J. C* 76 (2016), p. 581. arXiv: 1510.03823 [hep-ex].
- [110] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data”. In: *JINST* 14 (2019), P12006. arXiv: 1908.00005 [hep-ex].
- [111] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”. In: *JHEP* 09 (2007), p. 126. arXiv: 0707.3088 [hep-ph].
- [112] Paolo Nason. “A new method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 11 (2004), p. 040. arXiv: hep-ph/0409146.
- [113] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with parton shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. arXiv: 0709.2092 [hep-ph].
- [114] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043. arXiv: 1002.2581 [hep-ph].
- [115] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), p. 159. arXiv: 1410.3012 [hep-ph].
- [116] D. J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152.
- [117] Paolo Nason and Giulia Zanderighi. “ W^+W^- , WZ and ZZ production in the POWHEG-BOX-V2”. In: *Eur. Phys. J. C* 74.1 (2014), p. 2702. arXiv: 1311.1365 [hep-ph].
- [118] ATLAS Collaboration. “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. In: *JHEP* 09 (2014), p. 145. arXiv: 1406.3660 [hep-ex].
- [119] H.-L. Lai et al. “New parton distributions for collider physics”. In: *Phys. Rev. D* 82 (2010), p. 074024. arXiv: 1007.2241 [hep-ph].
- [120] J. Pumplin et al. “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis”. In: *JHEP* 07 (2002), p. 012. arXiv: hep-ph/0201195.
- [121] ATLAS Collaboration. “Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment”. In: (2020). arXiv: 2006.13076 [hep-ex].
- [122] ATLAS Collaboration. “Measurement of single top-quark production in association with a W boson in the single-lepton channel at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: (2020). arXiv: 2007.01554 [hep-ex].
- [123] ATLAS Collaboration. “Measurements of top-quark pair differential and double-differential cross-sections in the ℓ +jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector”. In: *Eur. Phys. J. C* 79 (2019), p. 1028. arXiv: 1908.07305 [hep-ex].

- [124] ATLAS Collaboration. *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*. ATLAS-CONF-2014-058. 2014. URL: <https://cds.cern.ch/record/1951336>.
- [125] G. D’Agostini. “A multidimensional unfolding method based on Bayes’ theorem”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 362.2 (1995), pp. 487–498. URL: <http://www.sciencedirect.com/science/article/pii/016890029500274X>.
- [126] G. D’Agostini. “Improved iterative Bayesian unfolding”. In: *arXiv e-prints* (Oct. 2010). arXiv: 1010.0632.
- [127] Tim Adye. *Unfolding algorithms and tests using RooUnfold*. 2011. arXiv: 1105.1160 [physics.data-an].
- [128] *Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding: CERN, Geneva, Switzerland 17 - 20 Jan 2011. PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*. CERN. Geneva: CERN, 2011. URL: <https://cds.cern.ch/record/1306523>.
- [129] ATLAS Collaboration. “ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 970. arXiv: 1907.05120 [hep-ex].
- [130] ATLAS Collaboration. *Multi-Boson Simulation for 13 TeV ATLAS Analyses*. ATL-PHYS-PUB-2017-005. 2017. URL: <https://cds.cern.ch/record/2261933>.
- [131] ATLAS Collaboration. *Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO*. ATL-PHYS-PUB-2017-007. 2017. URL: <https://cds.cern.ch/record/2261938>.
- [132] 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”. In: *Eur. Phys. J. C* 79 (2019), p. 847. arXiv: 1907.06728 [hep-ex].
- [133] ATLAS Collaboration. *Determination of the parton distribution functions of the proton from ATLAS measurements of differential $W^\pm$ and Z boson production in association with jets*. 2021. arXiv: 2101.05095 [hep-ex].
- [134] H1 and ZEUS Collaborations. “Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data”. In: *Eur. Phys. J. C* 75.12 (2015), p. 580. arXiv: 1506.06042.
- [135] S. Alekhin et al. “Parton distribution functions, α_s , and heavy-quark masses for LHC Run II”. In: *Phys. Rev. D* 96 (July 2017), p. 014011. arXiv: 1701.05838.
- [136] Tie-Jiun Hou et al. “New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC”. In: (Dec. 2019). arXiv: 1912.10053 [hep-ph].
- [137] L. A. Harland-Lang et al. “Parton distributions in the LHC era: MMHT 2014 PDFs”. In: *Eur. Phys. J. C* 75 (May 2015), p. 204. arXiv: 1412.3989.

- [138] The NNPDF Collaboration. “Parton distributions from high-precision collider data”. In: *Eur. Phys. J. C* 77.10 (Oct. 2017), p. 663. arXiv: 1706.00428.
- [139] M. Botje. “A QCD analysis of HERA and fixed target structure function data”. In: *Eur. Phys. J. C* 14 (2000), pp. 285–297. arXiv: [hep-ph/9912439](#).
- [140] S.I. Alekhin, Sergey A. Kulagin, and S. Liuti. “Isospin dependence of power corrections in deep inelastic scattering”. In: *Phys. Rev. D* 69 (2004), p. 114009. arXiv: [hep-ph/0304210](#).
- [141] I. Abt et al. “Study of HERA ep data at low Q^2 and low x_{Bj} and the need for higher-twist corrections to standard perturbative QCD fits”. In: *Phys. Rev. D* 94.3 (2016), p. 034032. arXiv: 1604.02299 [[hep-ph](#)].
- [142] CDF Collaboration. “Direct Measurement of the W Production Charge Asymmetry in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV”. In: *Phys. Lett.* 102 (May 2009), p. 181801. arXiv: 0901.2169 [[hep-ex](#)].
- [143] 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”. In: *Phys. Rev. D* 91.3 (2015), p. 032007. arXiv: 1412.2862 [[hep-ex](#)].
- [144] A. D. Martin et al. “Parton distributions for the LHC”. In: *Eur. Phys. J. C* 63 (Sept. 2009), pp. 189–285. arXiv: 0901.0002 [[hep-ph](#)].
- [145] ATLAS Collaboration. “Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector”. In: *Eur. Phys. J. C* 77 (2017), p. 367. arXiv: 1612.03016 [[hep-ex](#)].
- [146] FNAL E866/NuSea Collaboration. “Improved measurement of the $\bar{d}/\bar{u}$ asymmetry in the nucleon sea”. In: *Phys. Rev. D* 64 (Aug. 2001), p. 052002. arXiv: [hep-ex/0103030](#).
- [147] 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”. In: *JHEP* 05 (2014), p. 068. arXiv: 1402.6263 [[hep-ex](#)].
- [148] CMS Collaboration. “Measurement of associated production of a W boson and a charm quark in proton–proton collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 269. arXiv: 1811.10021 [[hep-ex](#)].
- [149] Sarah Alam Malik and Graeme Watt. “Ratios of W and Z cross sections at large boson p_T as a constraint on PDFs and background to new physics”. In: *JHEP* 14.2 (Feb. 2014), p. 25. arXiv: 1304.2424.
- [150] Radja Boughezal et al. “ W -boson production in association with a jet at next-to-next-to-leading order in perturbative QCD”. In: *Phys. Rev. Lett.* 115.6 (2015), p. 062002. arXiv: 1504.02131 [[hep-ph](#)].
- [151] A. Gehrmann-De Ridder et al. “Precise QCD predictions for the production of a Z boson in association with a hadronic jet”. In: *Phys. Rev. Lett.* 117.2 (2016), p. 022001. arXiv: 1507.02850 [[hep-ph](#)].
- [152] Richard D. Ball et al. “Parton distributions with small- x resummation: evidence for BFKL dynamics in HERA data”. In: *Eur. Phys. J. C* 78.4 (Apr. 2018), p. 321. arXiv: 1710.05935.
- [153] S. Alekhin et al. “HERAFitter. Open source QCD fit project”. In: *Eur. Phys. J. C* 75, 304 (July 2015), p. 304. arXiv: 1410.4412 [[hep-ph](#)].

- [154] H1 Collaboration. “A precision measurement of the inclusive ep scattering cross section at HERA”. In: *Eur. Phys. J. C* 64 (2009), pp. 561–587. arXiv: 0904.3513.
- [155] F. James and M. Roos. “Minuit - a system for function minimization and analysis of the parameter errors and correlations”. In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367.
- [156] ZEUS Collaboration. “ZEUS next-to-leading-order QCD analysis of data on deep inelastic scattering”. In: *Phys. Rev. D* 67 (2003), p. 012007. arXiv: hep-ex/0208023.
- [157] M. Botje. “QCDNUM: Fast QCD evolution and convolution”. In: *Comp. Phys. Comm.* 182 (2011), p. 490. arXiv: 1005.1481.
- [158] R. S. Thorne and R. G. Roberts. “Ordered analysis of heavy flavor production in deep-inelastic scattering”. In: *Phys. Rev. D* 57 (1998), p. 6871. arXiv: hep-ph/9709442.
- [159] R. S. Thorne. “Variable-flavour number scheme for next-to-next-to-leading order”. In: *Phys. Rev. D* 73 (2006), p. 054019. arXiv: hep-ph/0601245.
- [160] R. S. Thorne. “Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections”. In: *Phys. Rev. D* 86 (2012), p. 074017. arXiv: 1201.6180.
- [161] T. Carli et al. “A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project”. In: *Eur. Phys. J. C* 66 (2010), p. 503. arXiv: 0911.2985.
- [162] J. M. Campbell and R. K. Ellis. “Update on vector boson pair production at hadron colliders”. In: *Phys. Rev. D* 60 (1999), p. 113006. arXiv: hep-ph/9905386.
- [163] J. M. Campbell and R. K. Ellis. “MCFM for the Tevatron and the LHC”. In: *Nucl. Phys. Proc. Suppl.* 10 (2010), pp. 205–206. arXiv: 1007.3492.
- [164] Stefano Catani and Massimiliano Grazzini. “Next-to-Next-to-Leading-Order Subtraction Formalism in Hadron Collisions and its Application to Higgs-Boson Production at the Large Hadron Collider”. In: *Phys. Rev. Lett.* 98 (May 2007), p. 222002. arXiv: hep-ph/0703012.
- [165] S. Catani et al. “Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order”. In: *Phys. Rev. Lett.* 103 (2009), p. 082001. arXiv: 0903.2120.
- [166] Marek Schönherr. “An automated subtraction of NLO EW infrared divergences”. In: *Eur. Phys. J. C* 78.2 (2018), p. 119. arXiv: 1712.07975 [hep-ph].
- [167] Stefan Kallweit et al. “NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging”. In: *JHEP* 04 (2016), p. 021. arXiv: 1511.08692 [hep-ph].
- [168] H1 and ZEUS Collaborations. “Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic ep scattering at HERA”. In: *Eur. Phys. J. C* 78.6 (2018), p. 473. arXiv: 1804.01019 [hep-ex].
- [169] 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>.

- [170] D. Mason et al. “Measurement of the Nucleon Strange-Antistrange Asymmetry at Next-to-Leading Order in QCD from NuTeV Dimuon Data”. In: *Phys. Rev. Lett.* 99 (Nov. 2007), p. 192001.
- [171] S. Alekhin et al. “Determination of strange sea quark distributions from fixed-target and collider data”. In: *Phys. Rev. D* 91.9 (2015), p. 094002. arXiv: 1404.6469 [hep-ph].
- [172] A. Kayis-Topaksu et al. “Measurement of charm production in neutrino charged-current interactions”. In: *New J. Phys.* 13 (2011), p. 093002. arXiv: 1107.0613 [hep-ex].
- [173] O. Samoylov et al. “A precision measurement of charm dimuon production in neutrino interactions from the NOMAD experiment”. In: *Nucl. Phys. B* 876 (2013), pp. 339–375. arXiv: 1308.4750 [hep-ex].
- [174] 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.” In: *Phys. Rev. D* 64 (2001), p. 112006. arXiv: hep-ex/0102049.
- [175] Ferran Faura et al. “The Strangest Proton?” In: (Aug. 2020). arXiv: 2009.00014 [hep-ph].
- [176] Sayipjamal Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. In: *Phys. Rev. D* 93.3 (2016), p. 033006. arXiv: 1506.07443 [hep-ph].
- [177] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *J. Phys. G* 43 (2016), p. 023001. arXiv: 1510.03865 [hep-ph].
- [178] J. L. Abelleira Fernandez et al. “A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector”. In: *J. Phys. G* 39 (2012), p. 075001. arXiv: 1206.2913 [physics.acc-ph].
- [179] P. Agostini et al. “The Large Hadron-Electron Collider at the HL-LHC”. In: (July 2020). arXiv: 2007.14491 [hep-ex].
- [180] Rabah Abdul Khalek et al. “Probing Proton Structure at the Large Hadron electron Collider”. In: *SciPost Phys.* 7.4 (2019), p. 051. arXiv: 1906.10127 [hep-ph].
- [181] 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*. HEPData record. 2018. URL: <https://doi.org/10.17182/hepdata.80076>.
- [182] 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. URL: <https://doi.org/10.17182/hepdata.90953>.
- [183] ATLAS Collaboration. “Performance of algorithms that reconstruct missing transverse momentum in $\sqrt{s} = 8$ TeV proton–proton collisions in the ATLAS detector”. In: *Eur. Phys. J. C* 77 (2017), p. 241. arXiv: 1609.09324 [hep-ex].
- [184] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data”. In: *Eur. Phys. J. C* 74 (2014), p. 3071. arXiv: 1407.5063 [hep-ex].

- [185] ATLAS Collaboration. *Performance of the ATLAS Electron and Photon Trigger in pp Collisions at $\sqrt{s} = 7$ TeV in 2011*. ATLAS-CONF-2012-048. 2012. URL: <https://cds.cern.ch/record/1450089>.
- [186] ATLAS Collaboration. “Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data”. In: *Eur. Phys. J. C* 74 (2014), p. 2941. arXiv: 1404.2240 [hep-ex].
- [187] ATLAS Collaboration. “Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data”. In: *Eur. Phys. J. C* 77 (2017), p. 195. arXiv: 1612.01456 [hep-ex].
- [188] ATLAS Collaboration. “Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC”. In: *Eur. Phys. J. C* 73 (2013), p. 2518. arXiv: 1302.4393 [hep-ex].
- [189] ATLAS Collaboration. “Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC”. In: *Eur. Phys. J. C* 76 (2016), p. 653. arXiv: 1608.03953 [hep-ex].