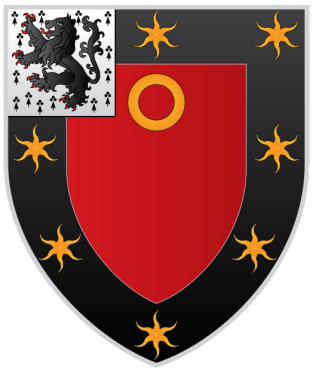
An observation of the electroweak production of a W-boson with two associated jets in a vector-boson fusion topology

James T. Henderson

St. John's College, Oxford



Thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Trinity Term, 2015

An observation of the electroweak production of a W-boson with two associated jets in a vector-boson fusion topology

James T. Henderson
St. John's College, Oxford

Thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

June 2015

Abstract

An observation of Electroweak production of a W-boson in association with two jets at high dijet mass is performed, using $\sqrt{s} = 8$ TeV pp collision data from the ATLAS detector, corresponding to 20.2 fb^{-1} of data. The process is observed to greater than the 5σ level and a fiducial cross section of $\sigma_{\text{Fid}} = 387.9 \pm 20.4$ (stat.) ± 5.4 (lumi.) $^{+34.8}_{-27.3}$ (syst.) $^{+34.9}_{-34.6}$ (theo.) fb is extracted.

Acknowledgements

Over the course of my DPhil. I have been fortunate enough to receive a huge amount of assistance from a great many people. I must thank the STFC for providing the 3.5 years funding to allow me to carry out my work and spend a fantastic 18 months living in Geneva as well as visiting numerous schools and conferences. Thanks too, to St. John's College, who were kind enough to provide me with assistance to visit conferences in New York and Moscow.

Huge thanks to my supervisor Chris Hays for his enduring patience and assistance during my random walk towards completion. Your incredible depth of knowledge and ability to spot questionable results have guided my learning and culminated in an interesting and enlightening analysis. Thanks for your support and the many beer-fuelled, excited physics discussions.

To Tony Weidberg for all your kind support and advice during my time in Oxford and for helping me to get to grips with the SCT, my heartfelt thanks.

The analysis presented in this thesis would not have been possible without the enthusiasm, knowledge and persistence of the Atlas VBF-W group. Alex, Aparajita, Christian, Darren, Julia and Sabine have all guided me to understanding the details of the analysis and sat through my meandering and often misguided presentations, thanks to you all. I would also like to acknowledge Andy Pilkington for his help and explanations on the finer points of VBF.

My learning process throughout my DPhil. has involved a great deal of computing as well as physics. My dubious coding skills have tested the Oxford computing system to its limit but it has performed immaculately and I would like to extend a great thanks to all the support staff for their help and for the impressive system. The administrative staff have also

helped to make light work of an often daunting level of paperwork and always with a smile.

On a personal level, my friends travelling the DPhil. journey with me have been a constant source of inspiration and amusement. There are far too many to name all of you but my kindest thanks to all at St. Johns, Oxford physics and CERN for making this testing process such fun. Special mention must go to those who have been burdened with me as an office mate; to Kate, David, Shaun, and Craig thank you for joining me in ROOT rants, putting up with my silly macros and for all the great memories. I will always have fond, rose-tinted, memories of an excellent cycle with Gerhard and Craig, thanks for the great adventure.

To Mum and Dad, thank you so much for all your support in multiple forms. This would have been impossible without your confidence and belief. My heart-felt thank you for encouraging me to question and investigate. To Sarah, Peter and Calum, thanks for putting up with my aimless physics chat and for the moral support over these years.

Finally thank you to Ali, for giving me the space I needed to complete my DPhil. Your support and encouragement for me to finish my thesis have kept me focused on better days ahead.

Preface

The work presented in this thesis is only possible due to the 3000 people working together in the ATLAS collaboration. I have been able to add my small contribution only because of the huge effort put in by thousands of physicists and engineers to enable not only the incredible success of the proton collisions but also the ability to understand and analyse them. The work in the first half of this thesis, explaining how the data is acquired and cleaned, is therefore dependent on the work of others. I have tried to reference further reading and additional documentation where possible, although unfortunately not all of this reference material is publicly available outside the ATLAS collaboration. I have produced all analysis plots, tables and values shown in the analysis sections but the analysis design was the work of all members of the ATLAS VBF-W group. I explicitly outline below my contributions to each chapter but I had overall responsibility for designing and implementing the code framework used in the analysis for both the creation of the detector level plots and the fits.

Chapter 5 : Using structures from previous analyses, I designed and implemented the multijet normalisation fits.

Chapter 6 : I implemented the uncertainty variations into the analysis and added the PDF interface. I also generated the interference samples for both SHERPA and MADGRAPH and extracted the interference uncertainties.

Chapter 7 : I designed and optimised the centrality variables and I contributed to the design of the remaining cuts.

Chapter 8 : I designed the *Control* region for use in constraining the background, adding to the existing *Validation* region, allowing comparison between corrections from two regions orthogonal to the *Search* region.

Chapter 9 : I implemented the dijet mass fits, pseudo-experiment procedure and the signal extraction.

Chapter 10 : I implemented a procedure for correcting the background modelling from the *Control* region and constraining the errors as shown in the plots. I also carried out the cross section calculations.

In addition to the analysis documented here I qualified for ATLAS authorship working on an analysis of single-event-upsets in the semiconductor tracker [1]. As this work does not easily fit into the discussion of the analysis documented in this thesis, it has been omitted. I also carried out inner detector shifts whilst at CERN and for a few terrifying nights was the only person in the ATLAS control room. Recently I have become involved in the the physics modelling group, updating and expanding the capabilities of the ATLAS Monte Carlo software.

Figures

I have created all figures used in this thesis unless otherwise specified in the figure caption. Any plots with the **ATLAS** label shown have been through the internal ATLAS approval process and are publicly available. All Feynman diagrams shown are created using *JaxoDraw* [2].

Units

Natural units are used throughout this thesis, setting $c = \hbar = 1$. Masses and momenta are therefore given in units of energy where the electron volt (eV) is used.

Contents

Chapter 1

Introduction

Particle physics has enjoyed a golden age in recent years with the highly publicised beginnings of the Large Hadron Collider at CERN. The discovery of the Higgs boson made front page news across the world and was a catalyst for excited discussions about science in pubs and classrooms alike.

Inside the world of particle physics the excitement was greater still due to the possibilities that the Large Hadron Collider presented. The unprecedented energy that was now available could rip up the accepted theories of physics in favour of something no-one had yet even thought of. The working model could be tested to minute precision, potentially right up to its limits, and brand new physics could be observed for the first time.

This thesis investigates whether a process predicted by the current accepted theory is realised in nature. All possible outcomes of asking such a question are thrilling. A positive result would mean observing nature working in a way that no human has ever seen before. On the other hand, finding that this process does not exist would mean that the theory, constructed by the founders of modern particle physics more than 50 years ago, would need radical rethinking. Unbiased viewpoints are vital in science but it helps to be presented with a win-win.

The electroweak production of a W-boson in association with two jets is investigated in

data and compared to the prediction from the Standard Model of particle physics. The 2012 ATLAS data-set with a centre of mass energy of 8 TeV and integrated luminosity of 20.2 fb^{-1} is studied. This process has never before been observed and an observation would be the first confirmation of the production of a charged boson via the fusion of vector bosons.

Chapter 2

Theory

The current accepted theory of particle physics is the Standard Model (SM); a relativistic quantum field theory describing all known particles and their interactions. This chapter gives a short overview of the theory in Section 2.1 and the two important components for this thesis in the Electroweak (EW) force in Section 2.2 and Quantum Chromodynamics (QCD) in Section 2.3. An overview of possible extensions to the current theory is given in Section 2.4. The methods of creating theoretical predictions are discussed in 2.5, and a discussion of the analysis signal and main background is covered in Section 2.6 .

2.1 The Standard Model

The success of the SM in describing almost all the phenomena that we observe in particle physics has led to it being adopted as the benchmark model in the field [3, 4]. Whilst there are unexplained observations (neutrino masses, dark matter, matter-antimatter asymmetry), for any new model to be considered it must first agree with the SM predictions for a great many results. The incredible accuracy of the SM thus far in the measured energy regime leads to the common consensus that it is simply incomplete rather than incorrect. This chapter gives an overview of the SM as it stands; more detail can be found in almost any particle physics text [5–9].

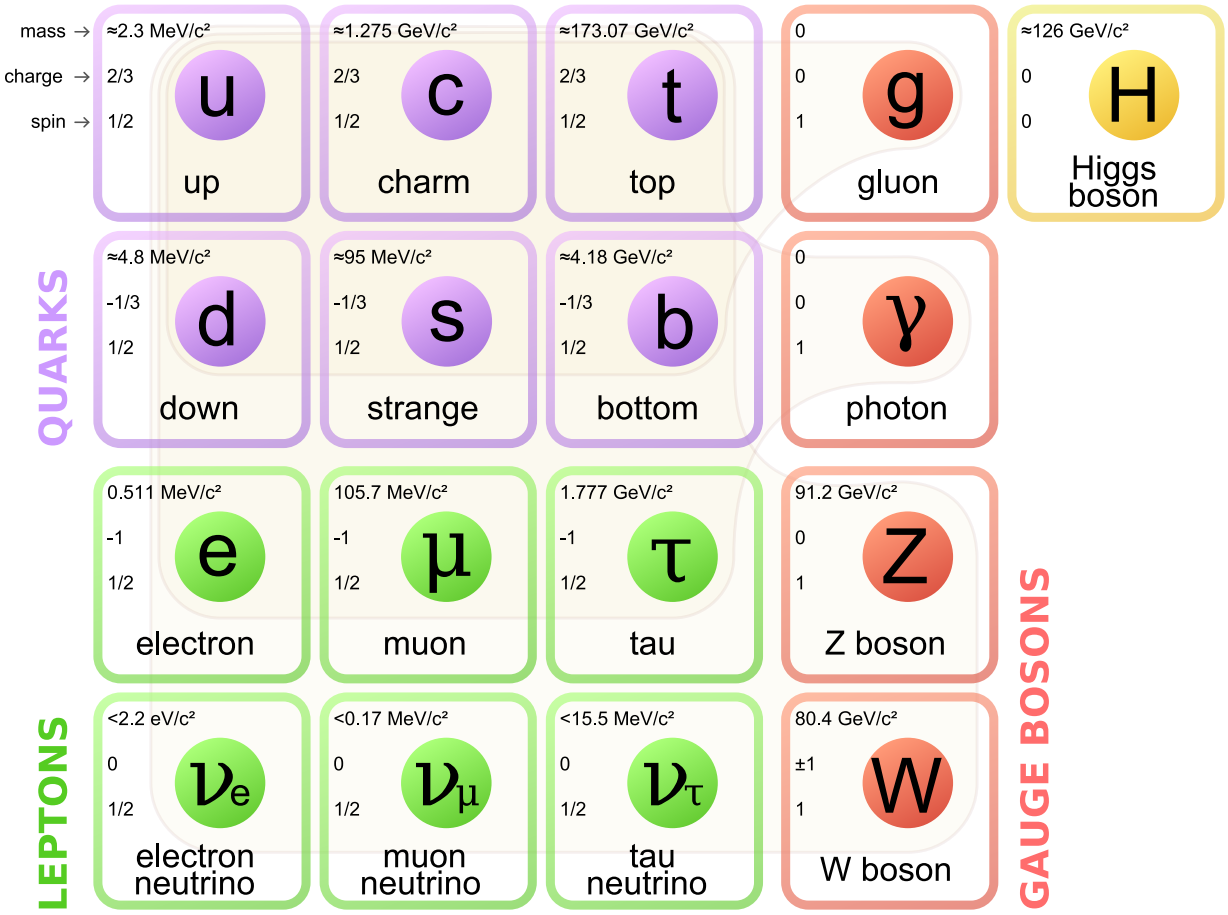


Figure 2.1: The fundamental particles of the SM. The "charge" shown is the electric charge. Taken from [11].

The SM defines the fundamental particles that make up the matter and forces in the universe. These particles are grouped into related types as shown in Figure 2.1. The matter particles are the spin 1/2 fermions: the leptons and quarks. Experimental evidence shows that there are 3 generations of these fermions [10] that appear to be replicas of the same structure with increasing mass. Particles observed to exhibit interactions with a force are said to be "charged" with respect to that force where the mannitude of the charge relates to the strength of the interaction. Each generation consists of an electromagnetically charged lepton (e.g. the electron), an associated neutral neutrino (e.g. the electron neutrino) and two quarks (e.g. the up and down quarks). The fermions have the same charges in every generation, only mass (and hence particle lifetime) is seen to change.

In each generation the charged lepton has an electromagnetic charge of $-1e$ and have masses ranging from 0.511 MeV to 1.777 GeV. Neutrinos are observed to oscillate between

lepton flavours suggesting they have eigenstates with non-zero mass but as yet only experimental upper bounds of these masses have been proven [12]. Both charged leptons and neutrinos carry no colour charge.

The two quarks have electric charge of $\pm\frac{2}{3}e$ for the "up-type" quark and $\mp\frac{1}{3}e$ for the "down-type" quark. They have masses ranging from $\sim$ 2 MeV to 173 GeV and a single QCD charge, with three possible colours.

Each of the fermions have an associated anti-particle* which has the identical mass and spin to the particle but opposite charges. The fermion particles are shown in Table 2.1.

Chirality	Particle			Q (e)	$\lvert I\rvert$	I_3	Y	coloured
	Generation							
	1^{st}	2^{nd}	3^{rd}					
Left-handed	Electron Neutrino ($\nu_{e,L}$)	Muon Neutrino ($\nu_{\mu,L}$)	Tau Neutrino ($\nu_{\tau,L}$)	0	$\frac{1}{2}$	$+\frac{1}{2}$	-1	×
	Electron (e^-_L)	Muon (μ^-_L)	Tau (τ^-_L)	-1	$\frac{1}{2}$	$-\frac{1}{2}$	-1	×
	Up (u_L)	Charm (c_L)	Top (t_L)	$+\frac{2}{3}$	$\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{3}$	✓
	Down (d_L)	Strange (s_L)	Bottom (b_L)	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{3}$	✓
Right-handed	Electron Neutrino ($\nu_{e,R}$)	Muon Neutrino ($\nu_{\mu,R}$)	Tau Neutrino ($\nu_{\tau,R}$)	0	0	0	0	×
	Electron (e^-_R)	Muon (μ^-_R)	Tau (τ^-_R)	-1	0	0	-2	×
	Up (u_R)	Charm (c_R)	Top (t_R)	$+\frac{2}{3}$	0	0	$+\frac{4}{3}$	✓
	Down (d_R)	Strange (s_R)	Bottom (b_R)	$-\frac{1}{3}$	0	0	$-\frac{2}{3}$	✓

Table 2.1: The fermion particles noting their charges for left and right handed chirality. The anti-particles are not shown but have the opposite sign charges from those shown. The right handed neutrinos have been included for completeness but their lack of charge means they do not interact and have not been observed.

The SM is a description of three of the four fundamental forces of nature: the strong (QCD), the weak, and the electromagnetic forces. The locally gauge-invariant mathematical structure associated to these interactions is a quantum field theory with $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry. The subscripts refer to the charges in each case: **C**olour, **L**eft-handed, and weak hypercharge (**Y**). Each symmetry gives rise to spin 1 bosons which couples to particles with the associated symmetry charge.

The remaining piece of the SM is the Higgs boson, which was pivotal in the proof of EW unification and subsequent symmetry breaking. Table 2.2 shows the SM bosons and their properties.

*Anti-particles are denoted by an overline ($u \to \overline{u}$) or an explicit charge change ($e^- \to e^+$).

Particle	Q (e)	coloured	spin	Couples to (associated charge)
W^-	-1	×	1	Weak-isospin (I)
Z	0	×	1	Combination of { weak-isospin (I) & weak-hypercharge (Y)
Photon (γ)	0	×	1	Electric charge (Q)
Gluons (g)	0	✓	1	Colour
Higgs (H)	0	×	0	Mass

Table 2.2: The bosons of the SM and associated charges and coupling charge. Only the W^- has a distinct anti-particle with opposite sign charges.

2.2 The Electroweak Force

The electromagnetic and weak forces have been unified into a single theory in the Glashow-Salam-Weinberg (GSW) model [13, 14].

The observed physical basis of the EW interaction is distinct from the underlying mathematical basis of the symmetries. That is to say the bosons that we observe are superpositions of the "symmetry bosons".

The physical bosons are the photon (A_μ or γ), transmitter of the electromagnetic force with the conserved electromagnetic charge (Q), and the $W^\pm$ and Z-bosons, transmitters of the weak force associated to the isospin charge (I). In terms of the mathematical symmetries, the charge of the $U(1)_Y$ SM symmetry is weak-hypercharge (Y), and the associated boson is denoted B_μ. The charge of the $SU(2)_L$ symmetry is weak-isospin (I) and the three bosons are denoted $\boldsymbol{W}^i_\mu$.

Right handed fermions have no weak-isospin charge and hence are singlets in $SU(2)_L$. Left-handed fermions form $SU(2)_L$ doublets of:

$$L_G = \begin{pmatrix} \nu_{L,G} \\ l_{L,G} \end{pmatrix}, \qquad Q_G = \begin{pmatrix} u_{L,G} \\ d'_{L,G} \end{pmatrix} \tag{2.1}$$

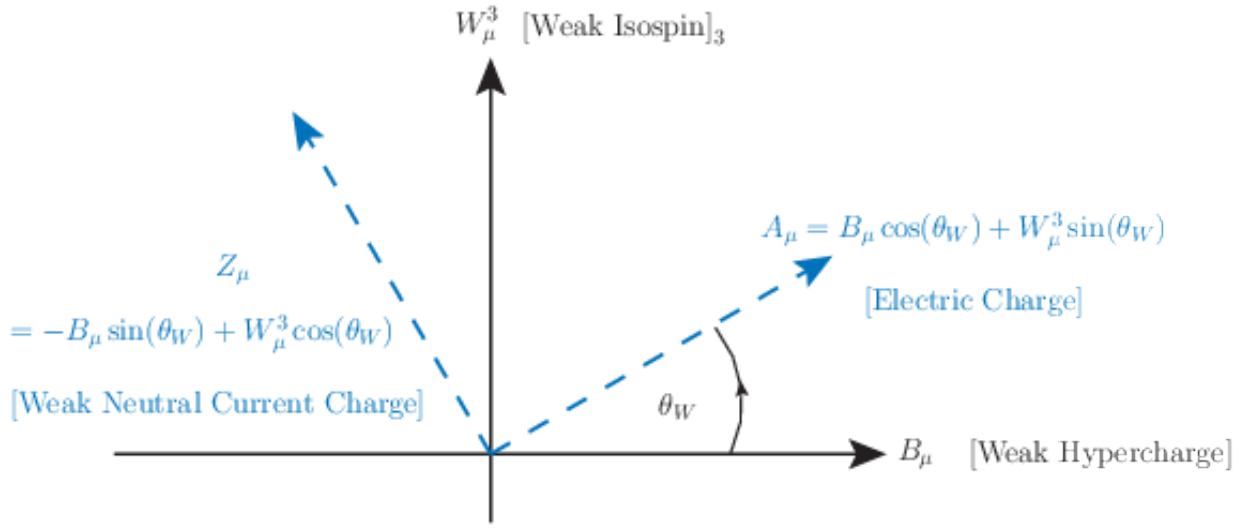


Figure 2.2: The rotation of the underlying symmetries to the physical bosons. The black axes show the underlying symmetries and the blue rotated axes show the physical bosons. The *charge* that each boson couples to is shown in square brackets.

where G denotes the fermion generation, ν denotes the neutrinos, l denotes the charged leptons, u the up-type quarks and d' the Cabibbo rotated down-type quarks[†]. The $SU(2)_L$ charged current is seen to transform one member of these doublets into the other. Two components of $\boldsymbol{W}^i_\mu$ can be identified with the physical W-bosons via the combination:

$$W^\pm_\mu = \frac{1}{\sqrt{2}} \left(W^1_\mu \mp W^2_\mu \right) \tag{2.2}$$

Whilst it is tempting to identify the remaining two bosons, B_μ and W^3_μ, as the A_μ and Z respectively, they are in fact combinations of both. A simple identification of the photon as the B_μ would suggest photons can decay to two left-handed, same-flavour neutrinos but this is not observed. Figure 2.2 shows how the underlying symmetries rotate via the Weinberg angle θ_W to the physical bosons.

The electromagnetic and weak forces are unified through their shared underlying symmetries. Because both left and right-handed fermions couple to weak hypercharge, the

[†]The Cabibbo rotations are a consequence of the mass eigenstates (the physical particles) being a rotation of the weak-isospin eigenstates.

250 Z-boson and the photon couple to both[‡]. None of the W-bosons, Z-boson or photon couple
251 to right-handed neutrinos, as they possess neither weak isospin nor weak hypercharge.

252 The EW force is a non-abelian theory which means the bosons have self-couplings. Self
253 couplings involve two W-bosons and one or two additional bosons. Vertices such as $ZZ\gamma$ or
254 $\gamma\gamma Z$ are forbidden. This can be interpreted as the Z-boson carrying no electric charge so
255 the γ does not couple to it.

256 2.2.1 The Higgs Field

257 Up to this point the model offers no explanation for the experimental observation of the weak
258 boson masses. Considering only fermions with gauge symmetry, all bosons are predicted to
259 be massless. The GSW theory adds a scalar field of a complex scalar weak doublet known as
260 the Higgs field [15–17]. The Higgs mechanism has the effect of hiding the gauge symmetry
261 at the cost of three additional Nambu-Goldstone bosons and a new scalar (spin-0) particle
262 called the Higgs boson. The three Nambu-Goldstone bosons are said to be "eaten" by the
263 weak bosons, which gain a longitudinal polarisation. This extra degree of freedom means
264 the weak bosons do not travel at the speed of light and hence, are said to have a non-zero
265 mass.

266 Recent experimental results have provided a first observation of the Higgs boson [18, 19].
267 The searches utilised, amongst others, the vector-boson fusion (VBF) formation of the Higgs
268 boson, see Section 2.6. The studies of this new boson have resolved its mass and couplings
269 to some SM particles, which as predicted scale with the particle mass. Much further study
270 of this boson is planned for the near future and the ability to search for high purity Higgs
271 boson events will be very important.

[‡]Though with different coupling strengths.

2.2.2 Coupling Constants

At this stage it is worth considering how the coupling strength—directly proportional to the probability for an interaction to occur—changes under different conditions. For the EW force discussed above, the coupling of any boson to a fermion depends on the charges Q and Y, but there is also an overall constant to apply: g_Y for the $U(1)_Y$ symmetry and g_I for the $SU(2)_L$ symmetry. As an example, the coupling for the photon is then:

$$g_\gamma = g_Y Y \cos(\theta_W) + g_I I_3 \sin(\theta_W) \tag{2.3}$$

using the formula depicted in Figure 2.2 and the charges Y and I_3 of the fermions.

The effective couplings vary depending on the energy at which they are being probed. For example, when determining the attraction of an electron to a positron through the electromagnetic boson γ, the effective charge of the electron in question depends on the momentum transfer in the process. An explanation for this are the virtual particles that are created around an electron, thus screening its charge. When we get infinitely close to the electron we experience its bare charge. On moving further away the bare charge is obscured by electron-positron pairs that are created from photons in the presence of the main electron. These extra charges make the electron charge more diffuse as we move away from the bare charge and thus the effective charge reduces. The changes in effective electron charge can be assigned to the coupling constants g in a method known as *renormalisation*. This dependence of the coupling 'constants' of a theory on the probe energy is know as *running couplings*. g_Y is experimentally seen to increase with increasing probe energy or equivalently, decrease as we move away from the particle. The probe energy at which the constants values are taken is known as the *renormalisation scale*, μ_R. A typical choice of μ_R for processes involving the production of a W-boson is the mass of the W-boson.

2.3 The Strong Force

The strong force, or QCD, is the description of the interactions of colour charged particles. The leptons have no colour charge so do not participate in QCD at all. The quarks, however, carry a single colour charge - *red, blue or green* - and the anti-quarks carry an anti-colour. This charge bears no relation to colour as experienced on a macroscopic scale but instead is simply a new quantum number required to allow three fermions to occupy an otherwise identical quantum state [20]. Singly coloured particles have never been directly observed, only *colourless* combinations: either a quark and anti-quark in a *colour, anti-colour* state or three quarks in a mixture of *red, green* and *blue*[§]. There are eight bosons associated to QCD called the gluons, which are massless and carry only colour charge - one colour and one anti-colour each.

The fact that QCD is a universal gauge symmetry with a single charge and massless gluons may appear to make its processes simpler to calculate than those of the EW force, but one simple difference makes it more complex. The coupling constant for QCD, denoted g_c, increases as the probe energy decreases (or equivalently the interaction length increases). The reason for this increase in g_c with distance is the same distribution of virtual particles that causes the electromagnetic screening only this time the bosons involved carry the charge. The number of gluons and the coloured quark causes the effective colour charge of a quark to increase as we move away from it. As coloured particles move away from one another, the rate of increase of the energy of the system increases. This leads to *confinement*, whereby increasing energy pulls two coloured particles apart until the system holds enough energy to create two new particles.

The reversed running of the strong coupling constant means that below the QCD scale $\Lambda_{QCD} \approx 200$ MeV the interactions of coloured particles cannot be calculated with perturb-

[§]In recent years four and five quark particles have been observed although it is still unclear if they are fundamental or orbital states of the two and three quark states [21, 22]

ation theory as the coupling strength:

$$\alpha_S = \frac{g_c^2}{4\pi}, \tag{2.4}$$

is greater than one. The inelastic interactions of proton collisions in a hadron collider can be calculated with perturbation theory, as the probe energy is $\gg \Lambda_{QCD}$. As the outgoing partons themselves are never observed they must be evolved to the physical colourless hadrons using less precise empirical models of low-energy, non-perturbative hadronisation. Models such as the "Lund string model" [23] or cluster models [24].

2.4 Beyond the Standard Model

There are a great many proposals of mathematically viable theories which would extend the SM and provide justifications for unexplained phenomenon. These are known as beyond-the-standard-model (BSM) theories and a large part of modern day experimental particle physics is spent on analyses designed to investigate their predictions. One such proposal, relevant to this analysis, are for additional charged bosons, known as a W's, to exist at masses higher than the W-boson mass. These bosons would be produced extremely rarely due to the high energy required to create a high mass particle. Current lower limits for the W' boson's mass are around 1.7 TeV [25].

There are also proposals for BSM physics which would alter the observed rate at which bosons couple to one another, known as anomalous triple gauge couplings (aTGCs) [26, 27]. These theories can be tested in a model independent manner by measuring the rate at which bosons couple together and comparing to the SM prediction. A significant deviation from the SM prediction would allow the aTGC values to be measured, whereas precise agreement with the SM prediction places upper limits on the aTGCs and excludes BSM models that predict a higher level for the coupling.

2.5 Simulating Theoretical Predictions

The only true test of a theory is to compare its prediction to experimental data. The mathematical formulation of the SM can be transformed into predictions via Monte Carlo (MC) methods. MC computer codes utilise random number techniques to simulate the probabilistic nature of quantum mechanics and integrate over the phase-space by sampling. They calculate a fully differential cross section for a physics process and produce individual events with the four momenta of final-state particles such that the full kinematics are accessible to the user. State-of-the-art EW and QCD predictions are implemented in the codes and free parameters, tuned to the observed data in order to give the most accurate possible predictions. Many texts cover the details of MC generation [28, 29] so only an overview is given here.

The colliding protons have a huge number of possible interactions so rather than simulate all processes we often pick out a specific interaction. To this end, requirements or "cuts" can be applied to the generation in order to force the simulation to use a certain interaction or final state, e.g. require the presence of a W-boson in the interaction. The structure of simulating particle physics event is modular. The "hard-process" defines the high-momenta-transfer kernel of the collision, defined by the ability to use perturbation theory to describe the interaction.

Protons are composite particles formed of three valence quarks (two *up*s and a *down*) and many low-energy gluons and sea quarks and anti-quarks of all flavours. Any quark, anti-quark or gluon in a proton in called a "parton". The hard-process at proton accelerators, such as the Large Hadron Collider (LHC), is initiated by interactions between partons rather than full protons. In simulating collisions it is therefore important to know the relative fractions of partons inside a proton and how these ratios alter with energy. Parton distribution functions (PDFs) describe the proportion of partons that carry a percentage x of the proton's energy. There are a number of PDFs available to use in MC production, most use a parameterisation to fit the amount of each parton and it's evolution with x.

The remaining partons of the colliding protons that are not involved in the hard-process are called the "underlying event" and interact with each other at low energies below the reach of QCD perturbation theory. In this thesis we are concerned with a certain number of hard-processes but the underlying event remains important to consider as it has the ability to produce a large number of strongly interacting particles that can affect the event topology.

Both the hard-process and underlying event produce particles which must be further evolved to be comparable to data. Strongly interacting, high momenta particles in the final state will radiate quarks and/or gluons in a process known as QCD final state radiation (FSR). Similarly it is possible for incoming protons or partons to radiate particles with colour charge; this is QCD initial state radiation (ISR). Isolated coloured particles are not observed, so final-state coloured particles shower a multitude of other coloured particles which interact to form colourless hadrons in a process called "hadronisation". This showering is known as a Parton Shower (PS) and is simulated with a QCD based model. Depending on the lifetime of a given hadron it may then decay into further, lower-energy particles. Photon emission is also possible at any point from electromagnetically charged particles. A schematic of a simulated event is shown in Figure 2.5, noting the various modular sections and how they connect.

On top of the complex nature of a two-parton initiated hard-process, it is possible to have multiple parton interactions where two sets of partons interact. These cases are extremely rare at significant momentum transfer.

2.5.1 α_S Order

MC generators calculate the hard-process to a certain order in perturbation theory in α_S. If the lowest possible order is used for a specific process, the calculation is known as leading order (LO). To achieve improved accuracy from the generators a higher order in the perturbative expansion can be used, although this comes at the (sometimes huge) cost of increased computing time. The current standard in particle physics is to use next to leading order (NLO) calculations, where an additional factor of α_S is involved. This can be viewed

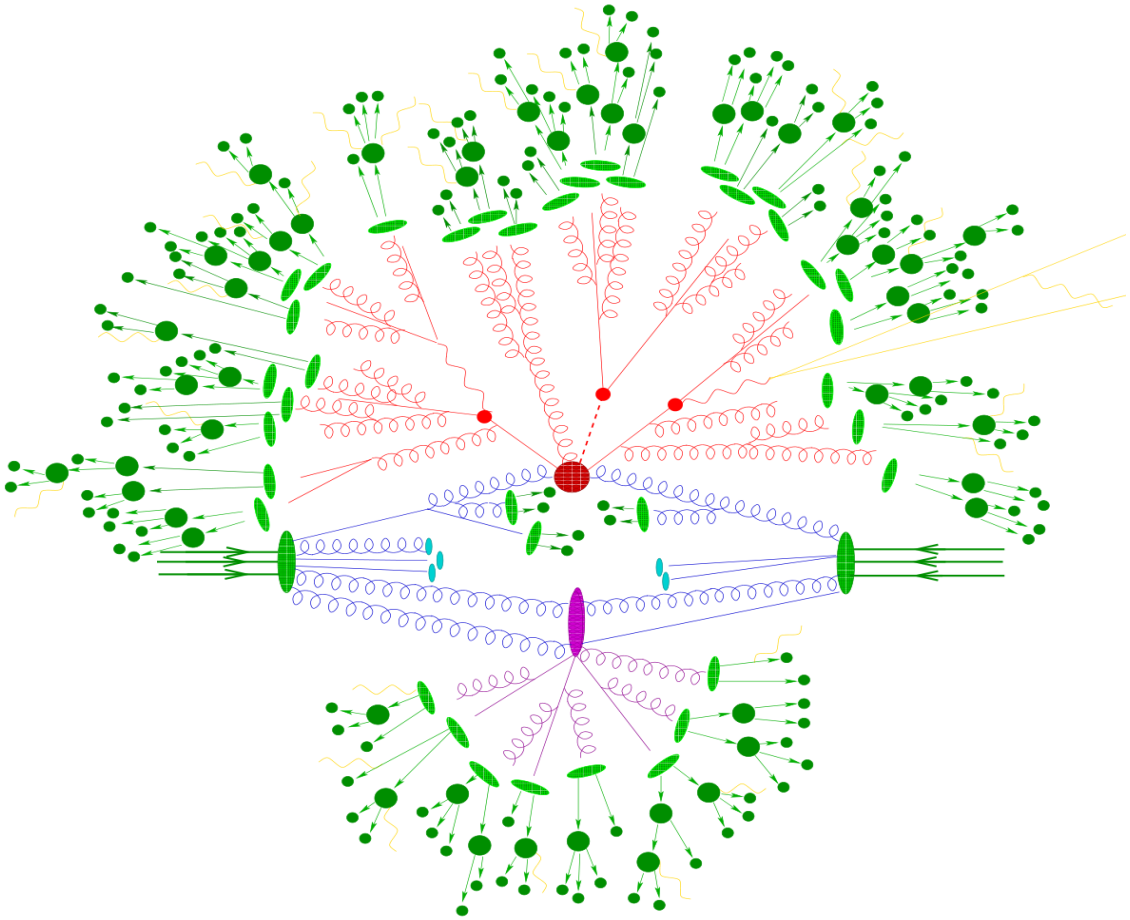
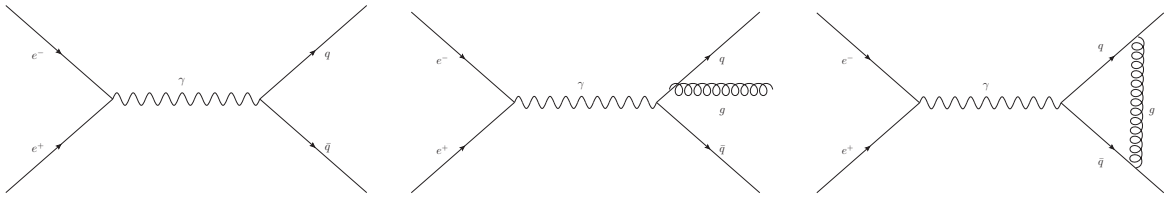


Figure 2.3: Pictorial representation of a proton-proton collision event as simulated by a MC generator. The incoming protons are shown in the green arrows before breaking up into the partons which collide. The hard-interaction is shown in the large red blob. The particles produced in the hard-process decay in the small red blobs. Additional QCD radiation is also shown in red. A secondary parton-parton interaction makes up the purple blob. The light green blobs show the hadronisation of the final-state partons and the dark green shows the decay of these hadrons. Final state, electromagnetic radiation is shown in yellow. Taken from [30].



(a) LO diagram (b) Real NLO emission (c) Virtual NLO

Figure 2.4: Example process showing LO and NLO diagrams to consider.

in terms of Feynman diagrams as in Figure 2.4 for an e^+e^- initial state example.

Calculations based on the LO diagram in Figure 2.4(a) would give predictions accurate to first order, α_S^0 in this case as there are no QCD vertices. An extra emission as shown in Figure 2.4(b) adds a factor of g_c so must be considered if we wish to be accurate to NLO at α_S. This type of diagram is known as an NLO "real" emission as the extra gluon leaves the hard-process. As the final state of the real diagram is distinct from the LO contribution this term does not interfere with the LO. The other type of NLO diagram in Figure 2.4(c) is known as a "virtual" contribution and contributes at NLO by interference with the LO diagram. Denoting the process cross section as σ, Equation 2.5 shows how the contributions combine to form a cross section accurate to NLO (α_S).

$$\begin{aligned}
\sigma_{NLO} &\propto |\text{LO} + \text{Virtual}|^2 + |\text{Real}|^2 \\
&= \left|f_{LO} \cdot g_c^0 + f_{Virtual} \cdot g_c^2\right|^2 + |f_{Real} \cdot g_c|^2 \\
&= f_{LO}^2 + f_{Virtual}^2 \cdot g_c^4 + 2 f_{LO} \cdot f_{Virtual} \cdot g_c^2 + f_{Real}^2 \cdot g_c^2 \\
&= f_{LO}^2 + 4\pi\alpha_S \left(f_{Real}^2 + 2 f_{LO} \cdot f_{Virtual}\right) + \mathcal{O}(\alpha_S^2)
\end{aligned} \tag{2.5}$$

The final line of Equation 2.5 shows how the NLO cross section conceptually changes from the LO cross section. The second term adds the NLO modification showing the real contribution and the virtual-LO interference. The final term is the pure virtual term and is neglected for containing two orders of α_S.

The real emission calculation logarithmically diverges for soft and collinear terms and is

made finite when virtual terms are included. To correct for the remaining logarithmically-enhanced terms in the perturbative expansion, a PS model is used to simulate further quark and gluon emissions from the hard-process.

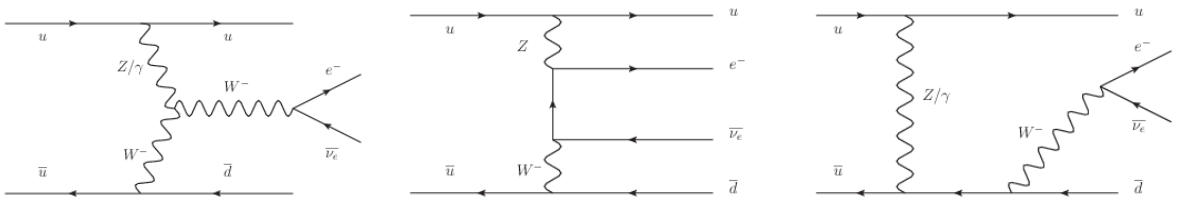
A *matching scale* must be chosen in order to control the jet energy at which the PS takes over the showering from the hard-process calculation. The hard-process is best suited for hard emissions of high momenta, well separated partons. The parton shower works best when considering low momenta, low opening-angle emissions. Recent advances in MC methods have meant the effect of this nonphysical scale on calculations is minimal [31].

The nature of QCD is that low energy calculations cannot be explicitly computed due to the large α_S value excluding perturbation theory [32]. In a hadron collider the total calculation of a process must not only consider the parton interaction but also the extraction of the parton from the proton. The partons inside the proton are able to possess any fraction of the energy of the total proton. The low energy component must be taken into account in simulating interactions. As it is not explicitly calculable, the PDFs model the distribution of partons and energies. The calculation for a parton interaction is thus a convolution of the parton densities from the PDF and the hard-scatter calculation. A *factorisation* scale (μ_F) is introduced to separate the two regimes; where the PDF or hard-scatter controls the parton evolution. This factorisation scale can be thought of as separating the short-distance hard-scatter from the lower energy, proton interactions. As the factorisation scale separating these domains is a mathematical artefact of the calculation rather than a physical phase-change, variations in the scale must be tested to observe their effect on the predictions. The size of the variation gives a sense of the missing terms in the perturbative expansion, so higher order calculations in α_S are expected to have less sensitivity to μ_F.

2.6 W + jets Production

The EW production of a W-boson in association with two jets at high dijet mass, dijet mass (M_{jj}), has never before been observed at a hadron collider. The SM predicts that

this process has a large contribution from the VBF production mechanism and hence is a sensitive test of VBF. The VBF process consists of two incoming quarks, which each radiate a vector-boson that then then fuse to form a single vector-boson. At LO the quarks have no further interaction and simply hadronise to form high-momentum, final-state jets in the forward region of the detector. The resulting fused vector-boson tends to decay centrally, in between the two quark jets, into final state particles. W-bosons decay to hadrons about 67% of the time [9] but identifying hadrons as coming from a W-boson is very challenging at a hadron collider. The high level of hadronic activity in most events means that leptons are much simpler objects to identify as coming from a boson so the leptonic decay channel is used in this analysis. When this boson decays leptonically there is very little hadronic activity in the central region of the detector due to the lack of colour exchange. The Feynman diagram for the case of VBF production of a W-boson is shown in Figure 2.5(a).



(a) VBF diagram (b) "Multiperipheral" (c) "Bremsstrahlung"

Figure 2.5: Subset of EW $W + 2$ jets diagrams making up the signal for this analysis [33].

The VBF diagram is only one of an interfering set of EW $W + 2$ jets diagrams. A subset of these diagrams are shown in Figure 2.5. As all these diagrams interfere, only the squared modulus of their sum is visible and it is impossible to separate one diagram from the others even throuhg kinematic cuts. The signal definition in any VBF search must be the interfering combination of all diagrams of this type.

Any fundamental boson can be produced via VBF, including W, Z, γ or a Higgs boson. Measuring the production rate of each is an important test of the SM to evaluate if the bosons couple to each other in the triple-boson vertex as predicted. Searches for the presence of EW production of a Z-boson have been recently published by both the CMS and ATLAS [34, 35] collaborations at European Centre for Nuclear Research (CERN). Both analyses

were able to confirm the presence of the EW $Z + 2$ jets signal to more than 5σ significance and no significant deviations from the SM predicted cross section were found. A previous ATLAS thesis also investigated W-boson production via VBF [36] finding a significance of 4.89σ.

The VBF mechanism and indeed the EW $W + 2$ jets production provide striking characteristics that can be used to discriminate this signal process from the extremely large backgrounds that dominate the data. The central leptons, high-momentum forward jets and lack of central hadronic activity allow us to distinguish EW $W + 2$ jets from the QCD production of the same final state, as well as large $t\bar{t}$ and other dijet backgrounds. QCD production of $W + 2$ jets forms the identical final state to the signal and occurs at a rate three orders of magnitude larger than the EW production. Not only does this mean the QCD background is very challenging to reduce to a level where the EW production becomes observable but it means that the modelling of the QCD events is pivotal to the entire analysis. A representative sub-set of the QCD $W + 2$ jets, the dominant background, diagrams can be seen in Figure 2.6.

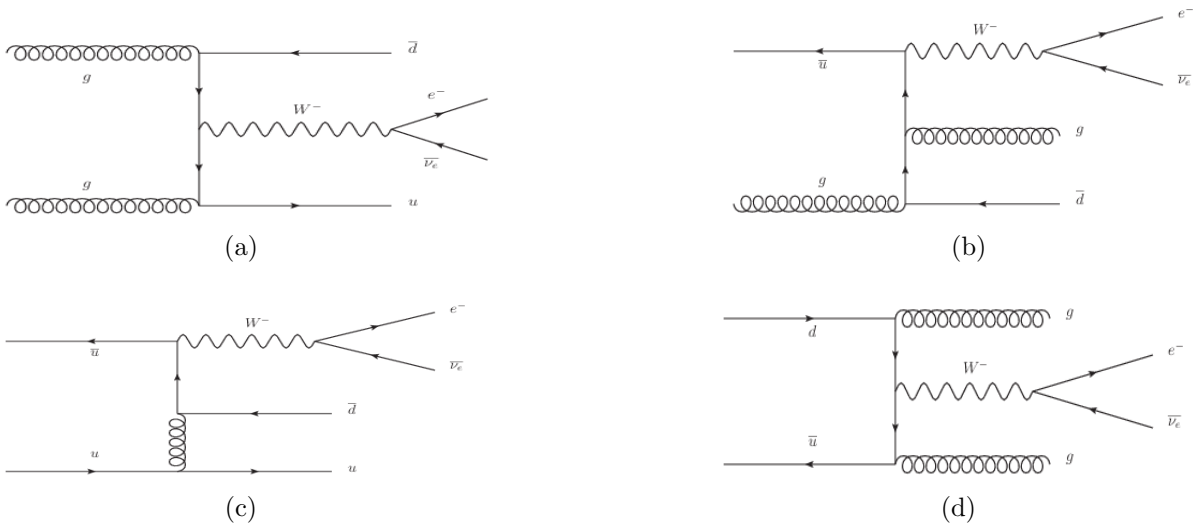
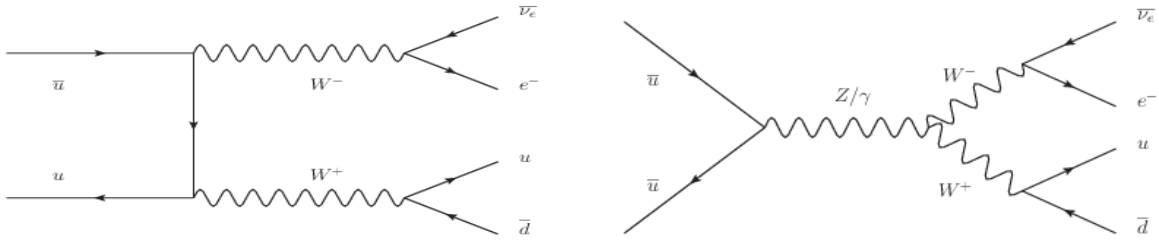


Figure 2.6: Sub-set of QCD $W + 2$ jets diagrams making up the dominant background for this analysis.

The complete set of EW $W + 2$ jets diagrams includes extra diboson diagrams not included in the signal definition. These diagrams mimic the signal final state when one boson decays hadronically and one decays leptonically, the LO diagrams are shown in Figure 2.7.

The presence of these processes has been previously verified [37], whereas our signal has yet to be observed. Therefore, the diboson component is excluded from the signal definition by use of an M_{jj} threshold above the mass of the W and Z-bosons. This cut minimises the diboson contribution in the signal phase space and reduces contributions from interference[¶] between the diboson diagrams and the EW signal diagrams, although this interference is considered in the analysis.



(a) (b)

Figure 2.7: EW W + 2 jets diboson production diagrams, which are excluded from the defined VBF signal and suppressed using an M_{jj} cut.

Whilst it is not the aim of this analysis, the EW production of $W + 2$ jets is sensitive to the presence of new heavy bosons and aTGCs. A new heavy W' boson would be able to be created off-shell in diagram 2.5(a), therefore not requiring the entire energy of its mass in the collision. The effect of this analysis would be a difference between the SM predicted cross section and the extracted cross section. There have also been suggestions for new heavy bosons which do not couple to leptons [38] and so would not be observed in diboson analyses but would be able to replace the t-channel W-bosons in Figure 2.5.

The VBF process also offers a probe of aTGCs which would cause a deviation from the SM predicted cross section [39]. Tests of the aTGCs using the VBF process provide information on the time-like momentum transfer of the couplings which are complementary to the aTGC tests using the diboson processes which probe the space-like momentum transfer.

[¶]The EW $W + 2$ jets diagrams interfere with the QCD $W + 2$ jets and diboson diagrams as they all can have identical external particles and can have the same couplings.

2.7 Summary

The SM is a highly successful theory, accurately predicting a vast number of observed physics phenomena. The SM predictions for EW $W + 2$ jets production can be be tested with data using the sophisticated MC models. The aim of this thesis is to search for EW $W + 2$ jets production. Either this study will observe this process for the first time and investigate the kinematics of VBF production or find a disagreement with the SM and an exciting prediction of new and unexplained physics.

Chapter 3

Experimental Apparatus

The analysis covered in this thesis uses data collected in 2012 by ATLAS (A Toroidal LHC ApparatuS) from proton-proton collisions in the LHC.

The LHC is the largest particle accelerator to date and is designed to collide proton beams at an unprecedented centre of mass energy. Exploring this untapped high-energy regime allows us to test the predictions of the SM under novel conditions. The large data-set collected by the LHC experiments allow precise measurements as well as first observations of low cross section processes. ATLAS is a general purpose particle-physics detector situated on the LHC accelerator.

This Chapter describes the proton acceleration chain in Section 3.1, the performance of the accelerator in Section 3.2 and the ATLAS detector and its sub-components in Section 3.3.

3.1 LHC Overview

The LHC accelerator is located in a 26.7 km circumference tunnel beneath the French-Swiss border near Geneva. This tunnel has an average depth of 100 m and was originally the home of the Large Electron-Positron (LEP) experiment [40], CERN's previous highest energy

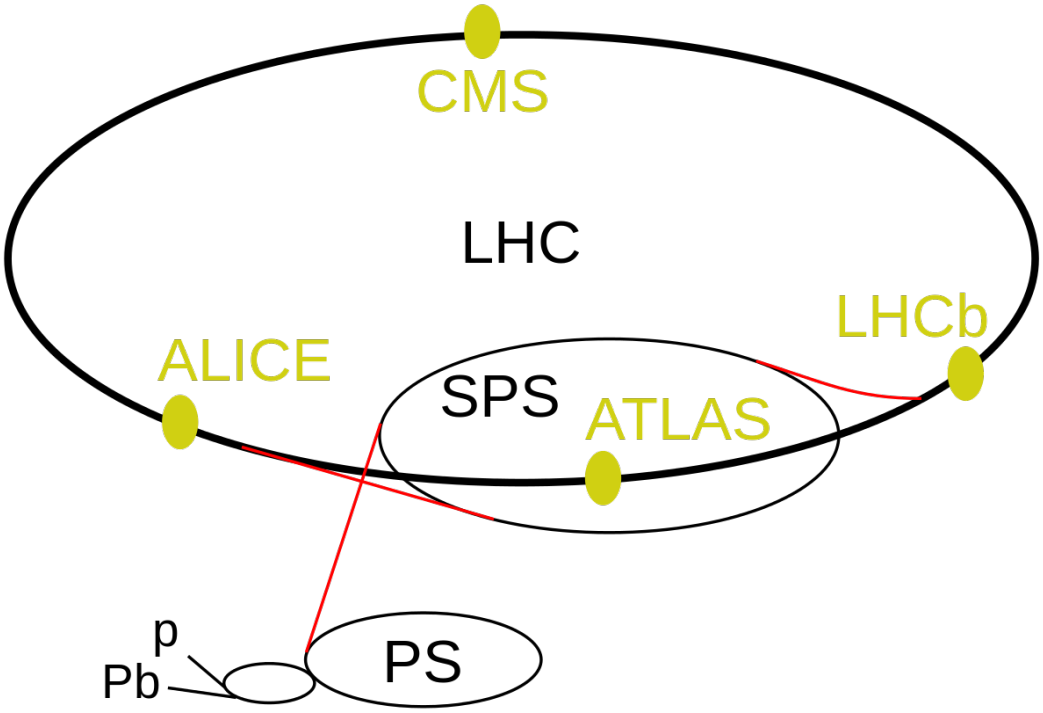


Figure 3.1: The LHC accelerator chain. The path of the protons (or heavy ions) is shown from from the Linac2 to the LHC. Taken from [41].

collider. The LHC is designed to accelerate protons up to a centre-of-mass energy of 14 TeV although, this energy is only expected to be reached around the year 2020. Run 1 consisted of 7 and 8 TeV in 2011-12 and only 8 TeV collisions are used in this thesis.

It is also possible to accelerate beams of lead ions up to 2.76 TeV/nucleon. These "heavy ion" collisions are used to investigate very high nuclear density.

The LHC is last in a series of accelerators that raise the protons to the required energy through incremental increases. The chain starts at the Linac2, where hydrogen atoms are ionised and the protons are accelerated to 50 MeV. The protons are progressively fed through the Proton Synchotron Booster (accelerated to 1.4 GeV), the Proton Synchotron (25 GeV), the Super Proton Synchotron (450 GeV) and finally the LHC itself (7 TeV). A schematic of the accelerator chain is shown in Figure 3.1.

The LHC consists of eight straight sections containing radio-frequency cavities to accelerate the beams and compensate for energy losses and eight arcs containing superconducting

bending magnets. Higher moment magnets are also used to focus the beams and maintain bunch structure. The limited size of the LEP tunnel means the two counter rotating beams must be held inside a twin-bore magnet beam-pipe rather than in separate rings.

Rather than have continuous beams, each LHC ring contains discrete bunches of protons so that the collisions are not continuous and the detectors can identify a particle measurement with a particular bunch crossing. The LHC proton beams are split into 2,808 bunches, each containing roughly 1.5×10^{15} protons. When operating under design conditions, bunches are spaced 25 ns apart and there is a total centre-of-mass energy of 14 TeV in each proton-proton collision. During the 2012 run, when all data used in this thesis was collected, the LHC operated at a centre-of-mass energy of 8 TeV with 1380 bunches spaced by 50 ns.

N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is the relativistic gamma factor, Σ_X and Σ_Y are the widths of the beam in metres in the X and Y directions respectively, and F the geometric luminosity reduction factor due to the crossing angle of the beams at the interaction point.

The rate of collisions R occurring per second is related to the inclusive cross section for a collision $\sigma_{Inclusive}$, and the luminosity $\mathcal{L}$.

$$R = \sigma_{Inclusive}\mathcal{L}. \tag{3.1}$$

The luminosity $\mathcal{L}$ can be calculated from the following beam parameters:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \Sigma_X \Sigma_Y} F. \tag{3.2}$$

The inclusive cross section can be replaced by the cross section for any particular process to give the number of events per second of that process, and the luminosity can be integrated to give the total number of events occurring for a given data-set. The maximum design luminosity of the LHC is 10^{34} cm^{-2}s^{-1}.

Due to phase space considerations and the relatively small size of hard parton collisions

(at 8 TeV this is ~ 20 mb compared to the total ~ 100 mb), the vast majority of the collisions are low momentum transfer interactions and result in low transverse energy particles. To increase the rate of interesting events, it is possible to increase the average number of interactions per bunch crossing. By increasing N_b or decreasing Σ_X and Σ_Y in Equation 3.2, we can produce an average of more than one proton-proton interaction per bunch crossing. This is known as in-time pile-up. By decreasing the spacing between bunches, the bunch crossings occur in quicker succession. This has the disadvantage that particles from other bunch crossings can affect the measurement of particles in a given bunch crossing. This is known as out-of-time pile-up. A certain level of pile-up is allowed so that for the same time collecting data, more interesting events are produced. The LHC was designed to have a nominal pile-up of ~ 35.

There are four points on the LHC ring where the two proton beams can be crossed to produce proton collisions in the experiments. The particle detectors located at these four crossing points are:

LHCb[42]

A single-arm forward spectrometer that specialises in B-hadron physics to study CP-violation and matter-antimatter asymmetry. Located at LHC point 8 near Ferney-Voltaire, France.

ALICE[43]

Focused on the study of heavy ion collisions and investigating the resulting quark-gluon plasma. Located at LHC point 2 in St. Genis, France.

CMS[44]

A general purpose particle detector located at LHC point 5 in Cessy, France.

ATLAS[45, 46]

A general purpose particle detector located at LHC point 1 near to the CERN Meyrin site in Switzerland.

3.1.1 General Purpose Detectors

CMS and ATLAS are both general purpose particle detectors designed to study a wide range of physics. They are similarly designed as cylinders, centred at the collision point and extending along the beam line, allowing for near complete coverage of the particle collision. Both aim to probe the SM at high energies and test for new physical processes. In 2012 both experiments published observations of the famous Higgs boson [18, 19] - a discovery that resulted in Nobel prizes for the architects of theory of electroweak symmetry breaking (see Section 2.2.1).

3.2 Collider Performance

All data used in this analysis were collected in 2012 when the LHC was colliding protons at 8 TeV. This section discusses the performance of the LHC during that time.

It is important to record as many collisions as possible in order to investigate rare physics processes and reduce statistical uncertainties on measurements. The expected number of events scales directly with the delivered integrated luminosity (see Equation 3.1). During the 2012 run the LHC delivered 22.8 fb^{-1} of data; details are shown in Figure 3.2. ATLAS recorded 21.3 fb^{-1} of data and all detector components were operational for 20.2 fb^{-1} of data.

For this data-set, on average more than twenty proton-proton interactions took place in every beam crossing. Figure 3.3 shows the distribution of the average number of interactions per bunch crossing. The symbol μ is used to represent this quantity. The mean value for μ throughout the 2012 run was 20.7.

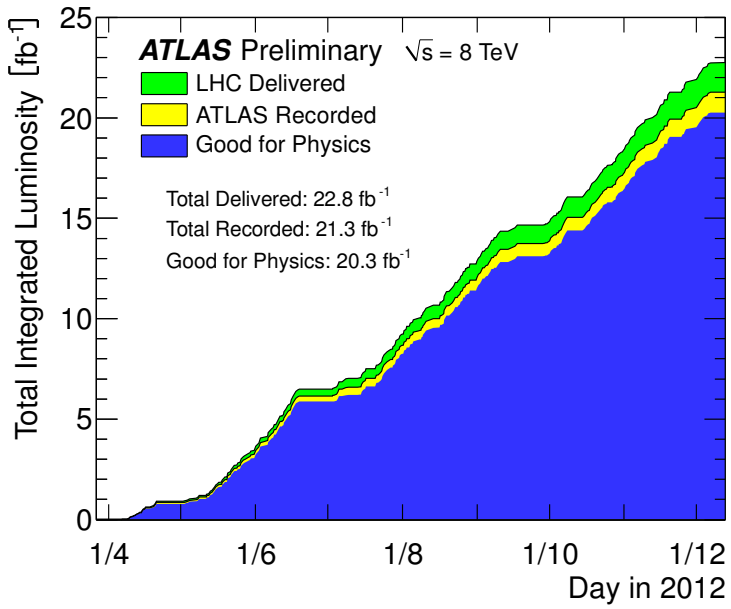


Figure 3.2: Cumulative delivered luminosity in the 2012 LHC run. Taken from [47].

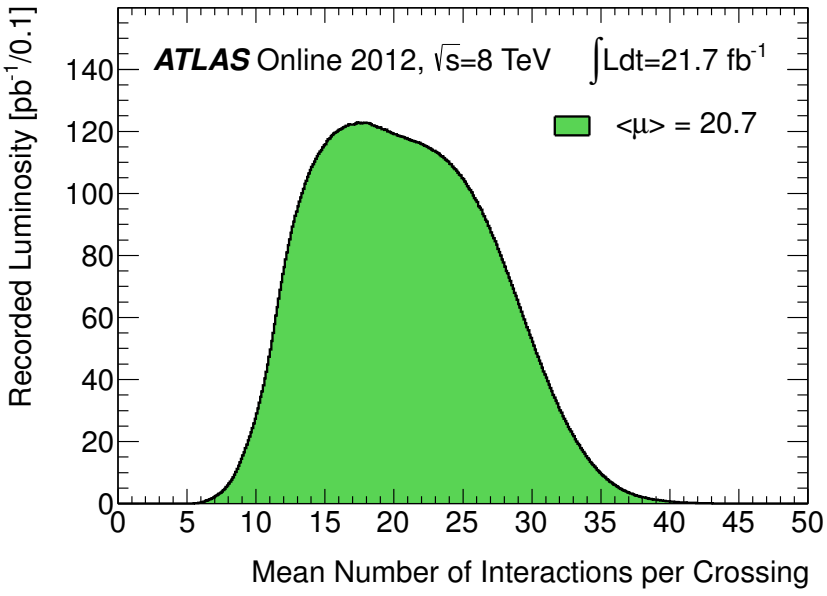


Figure 3.3: The distribution of the mean number of proton-proton interactions per bunch crossing in the 2012 LHC run. Taken from [47].

3.3 The ATLAS Detector

ATLAS is a general purpose particle physics detector with the ability to measure a multitude of physics phenomena. The detector design aims to cover as much of the nominal interaction point as possible in order to reconstruct the entire event kinematics. Sub-detector components are built up in concentric cylinders about the interaction point at the centre of the detector with the beam pipe running through the long axis of the cylinders. The philosophy of a general purpose detector such as ATLAS is to measure a particle's momentum in the inner sub-detector components and then destructively measure its energy in the outer components. There are many descriptions of ATLAS available in the literature [45, 46], this section gives an overview of the components most important to this analysis.

The ATLAS detector weighs 7000 tonnes and measures 44m in length and 25m in diameter. It can identify electrons, muons, τ leptons, photons and jets and infer the presence of neutrinos by measuring the momentum imbalance. The innermost sub-detector, the Inner Detector (ID), measures the momentum of charged particles. Surrounding the ID are two calorimeter systems designed to measure the energies of electromagnetic particles (in the inner one) and of hadronic particles (in the outer one). The outermost sub-detector is a spectrometer to measure high energy muons, which tend to pass through the previous components. A solenoid magnet bends particles in the ID and a toroidal magnet system bends muons in the outer muon spectrometer.

Figure 3.4 shows a cutaway view of the ATLAS detector with its sub-detector components which are discussed in the rest of this chapter.

3.3.1 Coordinate System

ATLAS conventionally uses a right-handed coordinate system with the origin at the nominal interaction point at the centre of the detector. The Z-axis is defined along the beam line with the positive direction anti-clockwise when viewing the LHC ring from above. The Y-axis points straight up, away from the centre of the earth, and the X-axis points to the centre of

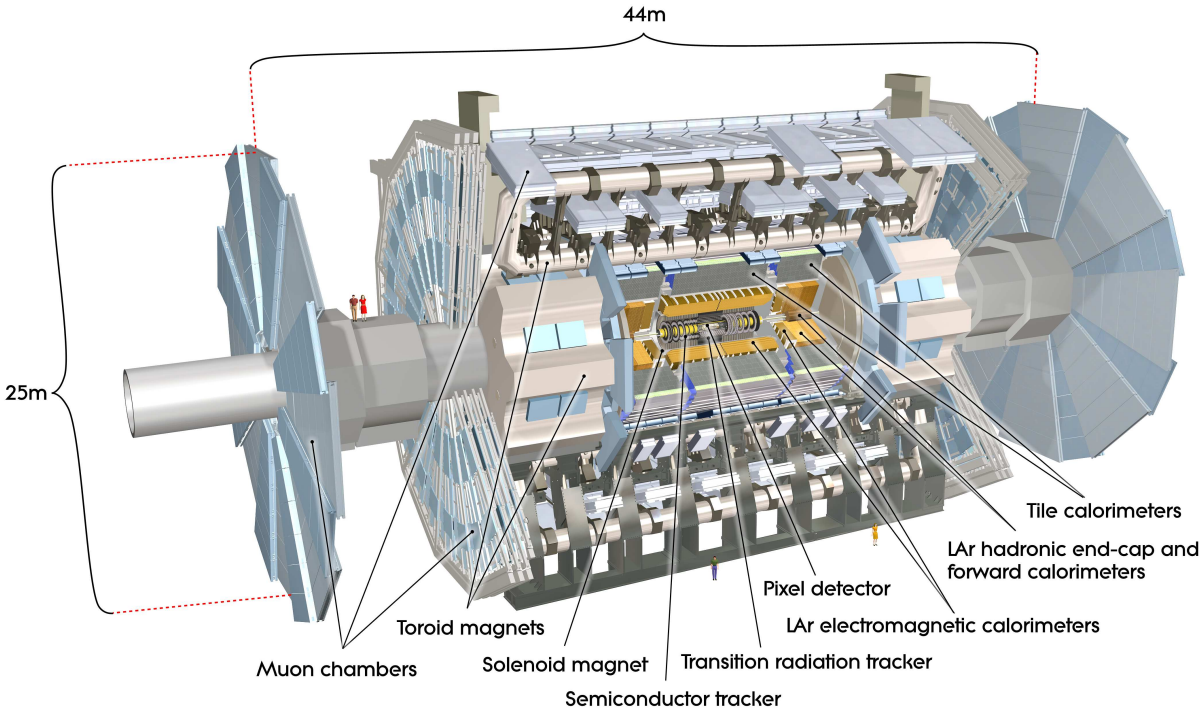


Figure 3.4: Cutaway of the ATLAS detector showing the sub-detectors. Taken from [45].

the LHC ring. The X-Y plane is often referred to as the transverse plane in which transverse momentum (p_T), transverse energy (E_T), and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) are defined. Distances between objects in the transverse plane are denoted R. In order to describe particle trajectories from the interaction point, it is useful to define a polar coordinate system. The azimuthal angle ϕ is defined from the Y-axis in the direction of the X-axis in the transverse plane. The polar angle θ is measured from the Z-axis in the direction of the Y-axis. An alternative metric of a particle's position in the θ plane is rapidity, notable due to *differences* in rapidity being Lorentz-invariant, which is defined as:

$$y = \frac{1}{2} \ln \frac{E + p_Z}{E - p_Z}, \tag{3.3}$$

where E is the particle's energy and p_Z is the particle's momentum along the Z-axis. As the vast majority of particles observed in ATLAS have an energy considerably higher than their mass, we can approximate most particles as massless. This simplifies the rapidity equation by equating a particle's energy with its momentum, resulting in the pseudo-rapidity η, defined

as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{3.4}$$

Units of pseudo-rapidity are not uniform in θ, covering a large θ region near the Y-axis and small θ regions near the Z-axis. Pseudo-rapidity has the appealing property that the flux of particles produced in ATLAS is roughly uniform across a unit of pseudo-rapidity.

The angular separation between objects is defined as:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \text{ with } \Delta\phi \in [0, \pi), \tag{3.5}$$

where $\Delta\phi$ is the azimuthal distance between the objects and $\Delta\eta$ is the pseudo-rapidity distance.

Particle tracks are reconstructed using the inner subdetectors of ATLAS where a track represents the helixial path of a charged particle passing througt a magnetic field. These tracks are parameterised by 5 variables: the inverse transverse momentum ($1/p_{\mathrm{T}}$), the azimuthal angle (ϕ), the polar angle ($\cot\theta$), the transverse impact parameter (d_0) and the longitudinal impact parameter ($z_o \times \sin\theta$). The impact parameters quantify a track's closest approach to a vertex. The impact parameters are calculated by extrapolation of the initial direction of the track, considering the effect of the magnetic field. The transverse impact parameter d_0 is signed; positive if the track curves towards the vertex at the point of closest approach and negative is it curves away [48, 49].

In order to maintain the vacuum of the LHC a beam pipe runs through the centre of the ATLAS detector. The beam pipe is 38 m long with a 5.8 mm inner diameter and is made of 0.8 mm thick beryllium at the interaction point. This construction enables a strong beam pipe that maintains the vacuum but presents the outgoing collision particles with minimal material to traverse.

A summary of the measurement precision from each subdetector is given in Table 3.1.

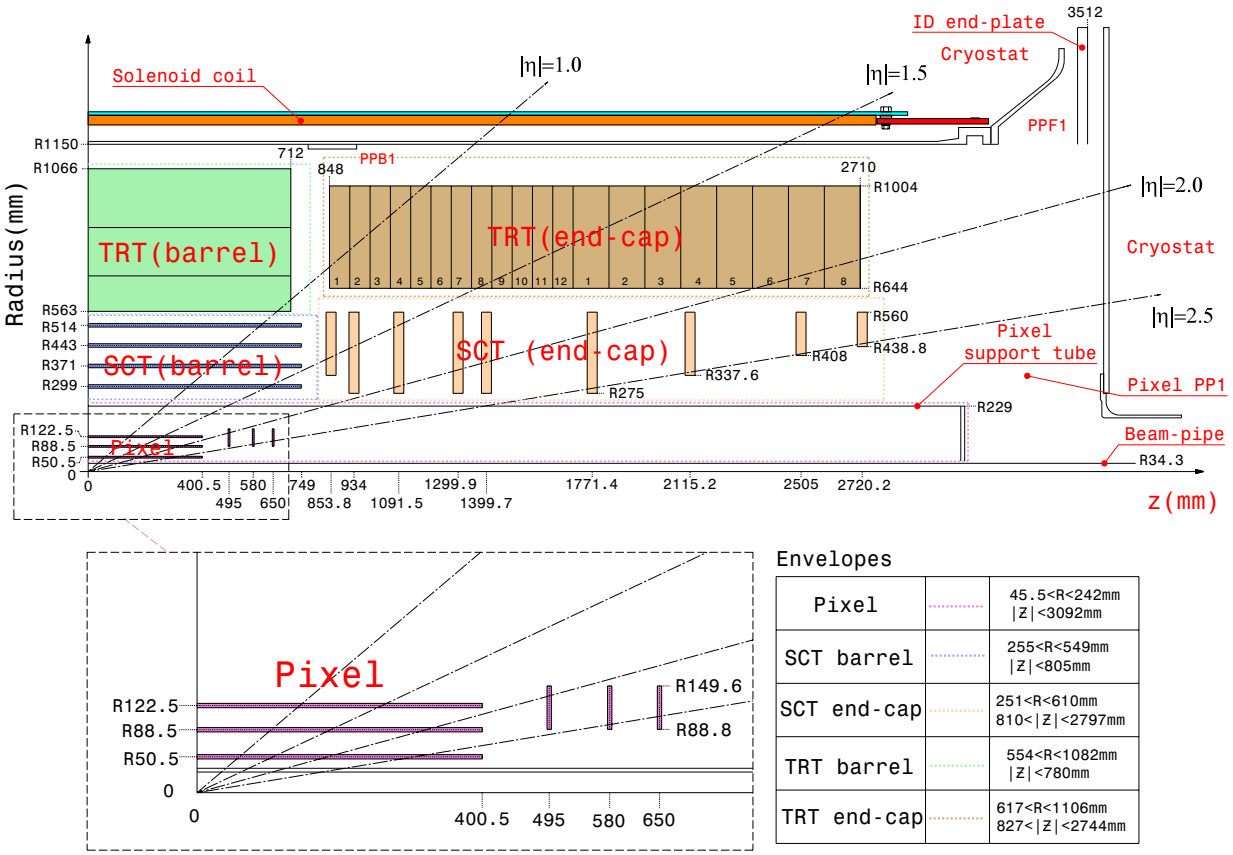


Figure 3.5: Plan view of a quarter section of the ID. Taken from [45].

Further details on each subdetector can be found in the following sections.

Subdetector	Resolution
Inner detector	$\frac{\sigma_{p_T}}{p_T} = 0.03\% \times p_T \oplus 1.5\%$
Electromagnetic calorimeter	$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%$
Hadronic calorimeter	$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$
Muon spectrometer	$\frac{\sigma_{p_T}}{p_T} = 4\%$ at $p_T = 100$ GeV

Table 3.1: The performance of each subdetector of the ATLAS experiment in the central region. More details of each subdetector can be found in the relevant section of the text.

3.3.2 Inner Detector

This section describes the three inner-most sub-detectors of ATLAS, which together make up the ID, as shown in Figure 3.5. These three components are contained within the 2 Tesla solenoid magnet so that charged particles exhibit a curved path in the ϕ direction. The aim of the ID is to provide precise measurements of a particle's position at several radii without

affecting its flight path. The momentum of charged particles is calculated from the curvature of the track. The extrapolation of this track back towards the Z-axis allows the position of the vertex where the particle originated to be reconstructed. The knowledge of this vertex is useful to see if the particle came from a primary vertex—a proton-proton interaction from the beams—or from a secondary vertex—a decay of a particle that travelled a measurable distance before decaying. Multiple primary vertices are possible in a single beam-crossing due to pile-up. The vertex with the highest scalar sum p_{T} of the tracks originating from it is considered to be the vertex of interest for that event and the other primary vertices are considered pile-up. Secondary vertices occur when hadrons with b or c quarks are produced and travel some distance before decaying. Other hadrons typically decay either within the measurement resolution of the primary vertex or within the detector (if at all).

The three ID sub-components are contained within a cylinder of about 5.3 m in length (along the beam line) and 2.3 m in diameter. Tracks with p_{T} as low as 400 MeV in the range $|\eta| < 2.5$ can be reconstructed with high efficiency. The central rapidity region of each detector component consists of a number of open ended cylindrical barrels. The forward components are a number of disks in the transverse plane facing the interaction point and called 'end-caps'.

The Silicon Pixel Detector

As the inner-most component of ATLAS, the pixel detector experiences the highest density of particles; thus the granularity must be extremely fine to resolve individual tracks. This high flux also causes issues with radiation damage to the materials used to construct the sub-detector. The basic unit of the pixel detector is a pixel of (R-ϕ) $\times$ Z dimensions 50 $\times$ 400 μm. 47,232 pixels are printed on a module and 48 modules make up a pixel detector stave. There are 1,702 staves making approximately 80.4 million readout channels in total. There are three concentric barrel layers around the interaction point covering $|Z| < 400.5$ mm and extending from 50.5 mm to 122.5 mm in R. Further out in the Z direction, there are three end-cap disks extending to 650 mm. The inner-most barrel layer is also known

as the 'b-layer' as the resolution it provides is crucial in location secondary vertices from B-hadrons. The pixel detector measures tracks from particles with $|\eta| < 2.5$ with an intrinsic precision of 10 μm in (R-ϕ) and 115 μm in Z in the central region or R in the forward region.

The Semiconductor Tracker

The Semiconductor Tracker (SCT) is located just outside the pixel detector from 299 to 514 mm in R. The particle density experienced by the SCT is high enough to require high resolution particle tracking, although less so than for the pixel detector. The basic unit is a 12 cm long silicon micro-strip, which detects passing charged particles by ionisation. There are 768 strips on each side of an SCT module. Each side is at a relative angle of 40 mrad to improve resolution. There are four barrel layers covering $|Z| < 853$ mm and nine end-cap disks extending to $Z = 2{,}720.2$ mm, allowing tracking coverage for $|\eta| < 2.5$. The intrinsic precision for the space-points of the SCT is 17 μm in (R-ϕ) and 580 μm in Z in the central region and R in the forward region.

The Transition Radiation Tracker

The outermost component of the ID is the Transition Radiation Tracker (TRT). Rather than track particles by ionisation, as with the previous two sub-components, the TRT uses transition radiation information. The detector is made from straw drift tube tracking detectors filled with a Xenon gas mixture (70% Xe, 27% CO_2, 3% O_2) interleaved with polypropylene fibres/foils acting as transition radiation material. The electric field of a charged particle is different in materials with different di-electric constants. Therefore, when traversing the boundary between such materials it emits radiation. The intensity of the radiation is proportional to the Lorentz factor of the particle, $\gamma = E/m$. This mass dependency allows electrons to be identified by the TRT.

The barrel extends from $R = 554$ to R= 1,082 mm and out to $|Z| = 749$ mm. The straws are aligned along the beam-line and split at $\eta = 0$. The barrel straws do not provide any Z

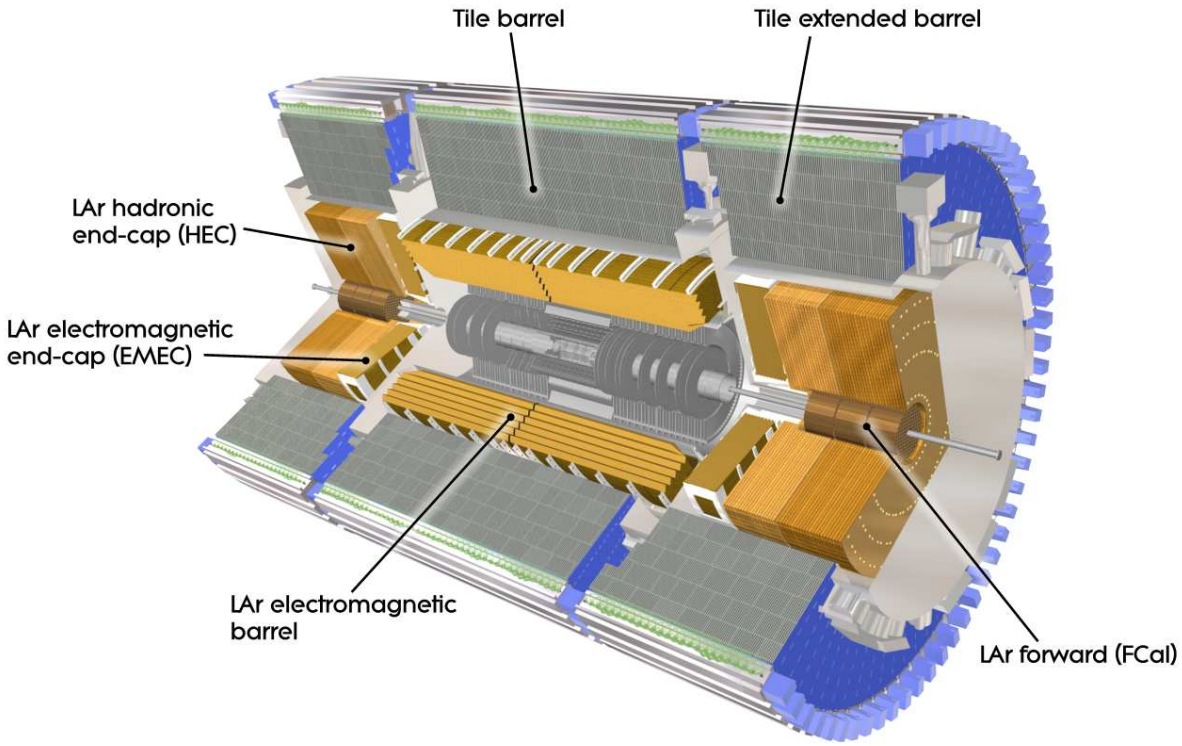


Figure 3.6: Cut-away view of the ATLAS calorimeter system, taken from [45].

measurement but give an (R-ϕ) position to a precision of 130 μm.

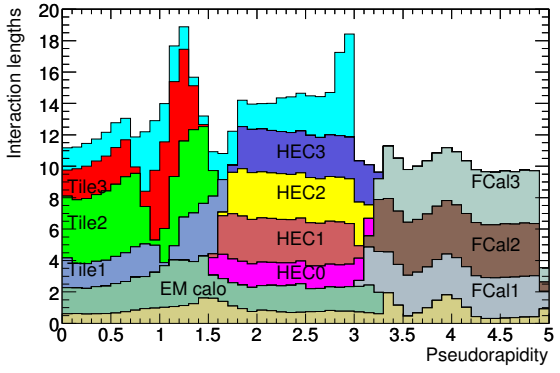
The end-cap section extends from $|Z| = 853.8$ mm to $|Z| = 2,710$ mm, extending coverage to $|\eta| < 2$. The end-cap straws are aligned radially and give a (Z-ϕ) position with a precision of 130 μm but do not give any R information.

In total the TRT has approximately 351,000 read-out channels.

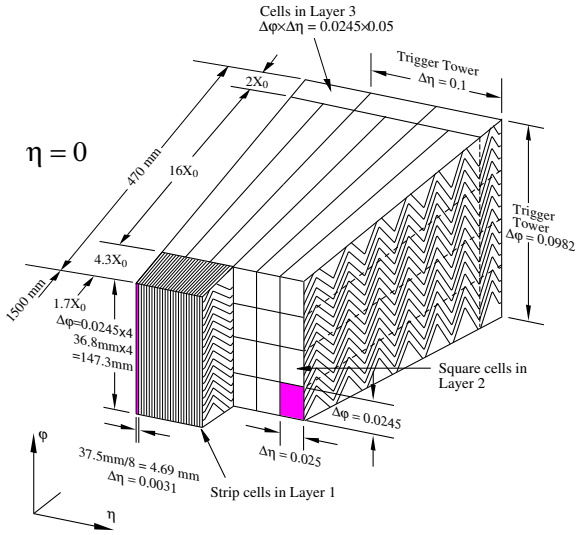
3.3.3 Calorimeters

The ATLAS calorimeters are located outside the ID and solenoid systems. The calorimeters are designed to destructively measure all particle's energies (excluding muons and neutrinos). The inner sub-components are designed to measure electromagnetic particles, primarily electrons and photons, and the outer sub-components are designed to measure the showers of hadronic particles. The precision of the ID complements the energy measurements but the calorimeters extend further forward than the ID, covering $|\eta| < 4.9$.

In order to provide maximum containment for particle showers, the calorimeter must extend for as many interaction lengths* as possible in R-Z. The total amount of material upstream and inside of the calorimeters is shown in Figure 3.7(a) as a function of $|\eta|$.



(a) Cumulative amount of material, in units of interaction length, as a function of $|\eta|$



(b) Geometry of the barrel electromagnetic calorimeter. The three layers in R are shown.

Figure 3.7: Layer of the ATLAS calorimeters, both taken from [45].

The Electromagnetic Calorimeter

Between the solenoid magnet and the Electromagnetic Calorimeter (EM Calo) there is a thin liquid argon layer acting as a pre-sampler in the range $|\eta| < 1.8$. This allows for the correction of energy losses upstream of the calorimeters for central particles.

The innermost calorimeter is designed to capture the showers of electromagnetic particles. The barrel extends from $|\eta| < 1.475$ and the end-caps $1.375 < |\eta| < 3.2$ (separated into an two coaxial wheels). The EM Calo contains 22 to 30 radiation lengths (akin to interaction lengths but for electrons and photons) so the vast majority of electromagnetic particle showers are fully captured by this calorimeter.

The design is a sampling calorimeter using liquid argon as the active detector material, chosen for its intrinsic radiation hardness and its linear response, and lead as the absorber

*An interaction length is the mean path length required to reduce the relative energy of a hadron by $1/e$ or 0.368.

material. Figure 3.7(b) shows the accordion geometry of the lead plates and the segmentation of the EM Calo into three layers. The granularity decreases as a particle moves through the three layers, with the fine-grained innermost layer providing high position resolution of photons making the distinction of $\pi^0 \to \gamma\gamma$ from a single photon possible. The longitudinal evolution of the EM shower gives information about the incident particle and allows discrimination between electrons or photons and hadronic objects.

Electrons in the region $1.37 < |\eta| < 1.52$ are usually excluded from physics analysis due to the large amount of upstream material that reduces the resolution of the EM Calo energy measurement. This η range is known as "the crack" region and affects electron identification.

The Hadronic Calorimeter

The next sub-detector after the EM Calo is the hadronic calorimeter. With a depth of 9.7 interaction lengths at $\eta = 0$, it is designed to capture the vast majority of hadronic showers. Any leakage of the showers would result in mismeasured energies and an incorrect missing momentum measurement, as well as misidentified muons in the muon spectrometer. The hadronic calorimeter is a sampling calorimeter which uses plastic scintillator tiles as the active medium and steel as the absorber material in the barrel region for $|\eta| < 1.7$. The use of steel doubles as a return yoke for the solenoid magnet. As with the EM Calo, there are three layers in the barrel. The inner two layers have read-out cells of dimensions $\Delta\eta \times \Delta\phi \approx 0.1 \times 0.1$, and the outer layer has dimensions 0.2×0.1.

The end-caps cover the range $1.5 < |\eta| < 3.2$. To combat the higher radiation dose in the forward region the end-caps use liquid-argon as the active material and copper plates as the absorber material. The cell size for $|\eta| < 2.5$ is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, and 0.2×0.2 for more forward cells.

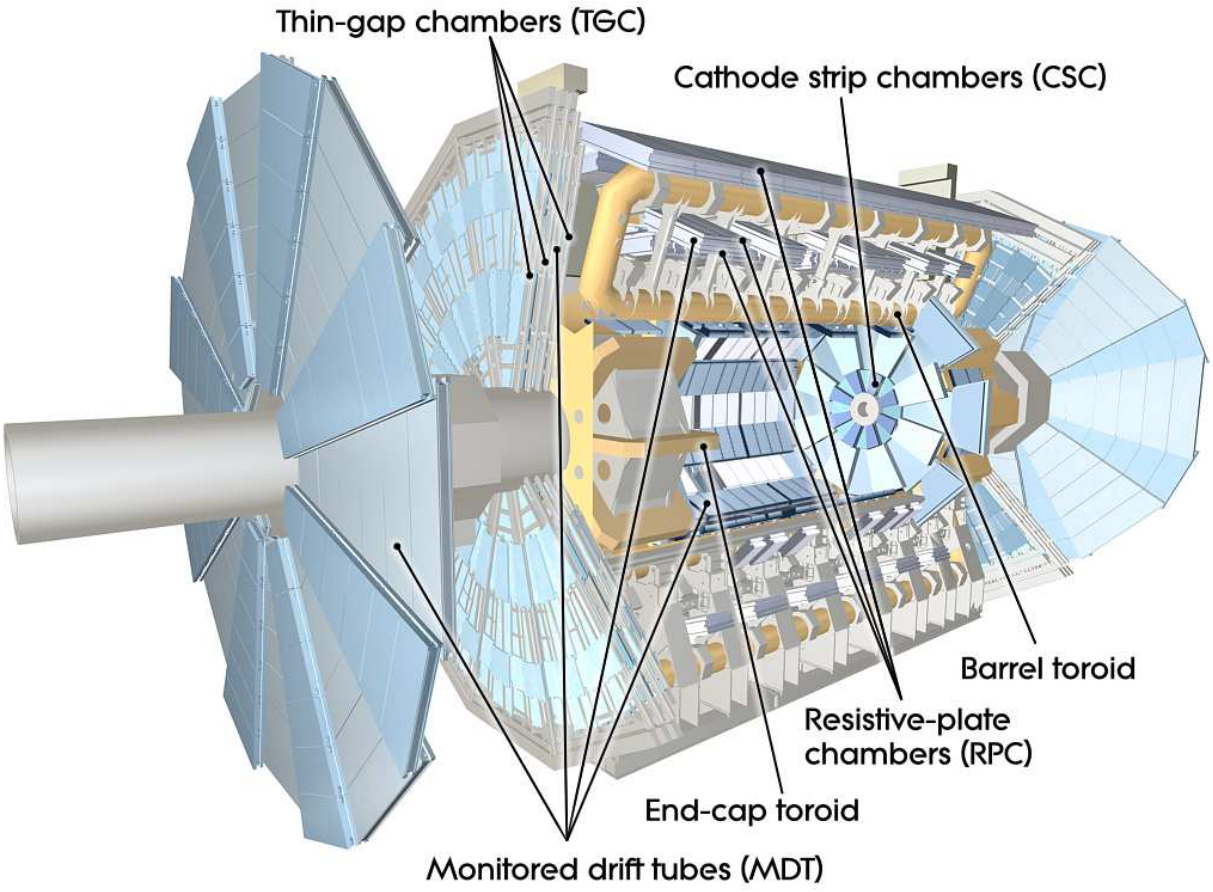


Figure 3.8: Cut-away view of the ATLAS muon spectrometer system, taken from [45].

The Forward Calorimeters

To capture the energy of high-$|\eta|$ particles, the forward calorimeters extend the coverage of ATLAS from $|\eta| = 2.9$ to $|\eta| = 4.9$. They consist of a single electromagnetic layer followed by two hadronic layers. Capturing the energy of these forward particles improves the missing momentum measurement and allows forward particles, such as those produced in a VBF process, to be captured.

3.3.4 The Muon Spectrometer

At the outer edge of the calorimeters the only remaining observable particles are typically muons. The muons interact minimally with the upstream detector components, as they are minimum ionising particles. The muon spectrometer coverage extends to $|\eta| = 2.7$, although

triggering only occurs up to $|\eta| = 2.5$ and there is a small gap at $\eta = 0$ to allow services to access the ID.

The muon system is designed to have a stand alone precision of 10% on a muon of $p_{\mathrm{T}} =$ 1 TeV, and to measure muons down to $p_{\mathrm{T}} = 3$ GeV. A toroidal magnet system surrounds the muon system, emitting a magnetic field between 0.5 and 1 T which causes the muons to curve in the *Y-Z* plane. This allows the charge of the muon to be identified for muons with transverse momentum up to ~ 3 TeV. Information from the muon system is used in the fast triggering (see Section 3.3.5) so it must be available extremely quickly. More accurate measurements take slightly longer and rely on multiple detector components. The various components of the muon spectrometer are explained below.

Trigger Chambers

Resistive Plate Chambers (RPC)

These chambers give a very fast space-point measurement for $|\eta| < 1.05$ when a muon ionises the (97% $C_2H_2F_4$, 5% Iso-C_4H_{10}, 0.3% SF_6) gas held between two charged, resistive plates.

Thin Gap Chambers (TGC)

These multi-wire proportional chambers provide an azimuthal coordinate position in the range $1.05 < |\eta| < 2.7$ when a muon passes through.

Precision Chambers

Monitored Drift Tubes (MDT)

Proportional drift tubes cover the range $|\eta| < 2.7$ and provide tracking to a precision of ~ 80 μm in the (*R-Z*) direction.

Cathode Strip Chambers (CSC)

Multi-wire proportional chambers with cathode strips and anode wires provides

(R-ϕ) coordinates for tracking in the forward region ($2.0 < |\eta| < 2.7$) to a precision of ~ 60 μm.

3.3.5 Triggers

The data collected by all the ATLAS sub-detectors is read out and stored to disk for later analysis. In 2012 the proton bunches were separated by 50 ns, corresponding to a crossing frequency of 20 MHz. The average data size of an event was 1.3 MByte and the recording infrastructure is only capable of writing out data at $\sim$ 300 MBytes per second, so the potential incoming rate of 26 TBytes per second must be reduced. A three level trigger system is used in ATLAS to reduce the incoming rate to a manageable 200 Hz, a reduction of order 10^5. The aim of the trigger system is therefore to make an extremely fast decision as to whether an event is interesting enough to be written out to disk or should be ignored and the data lost forever [50]. The three trigger levels are designated Level 1 (L1), Level 2 (L2) and Event Filter (EF), each reducing the incoming rate from the previous step using increasingly detailed information about the event. The L2 and EF are collectively known as the High Level Trigger (HLT).

A coarse overview of the event is formed using fast information from the calorimeters and the muon system to form primitive electrons, photons, jets[†], muons, τ-leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$. Each trigger step compares properties of these objects to some predefined thresholds to determine if the event should be kept. In general, events with high p_{T} objects are kept and those with low p_{T} objects are dropped.

The L1 trigger reduces the incoming rate of 20 MHz to $\sim$ 75 kHz and must process each event in $\sim 2\mu$s. As speed is of the essence, the L1 system is located as close as possible to the detector and is an entirely hardware based system.

The HLT triggers are processed in computing farms and are both off-detector, software based systems.

[†]See Section 4.4 for the definition of jets.

The data for the event is only read out from pipeline memory buffers if a positive decision comes from the L1 trigger. In that case the L2 system looks at a Region Of Interest (RoI), seeded by L1, in a specific η, ϕ region. The full detector information is read out in that RoI (only about 2% of the total event data) and L2 has an average of 40 ms to make a decision on that event and reduce the rate to 3.5 kHz.

The EF has 4 s per event to make a decision with the advantage of access to more complete event information. In the EF the rate is reduced to the final 200 Hz.

After events are written out, they are classified by 'stream' depending on some defining particles, e.g. a high p_{T} electron would put the event into the `egamma` stream, and then written to disk for offline analysis.

3.3.6 Luminosity Detectors

As can be seen in Equation 3.1, the measurement of any cross section relies on accurate knowledge of the integrated luminosity. To calculate the luminosity there are two small detectors at large Z values.

The luminosity measurement using Cerenkov integrating detector (LUCID) [51] is located at ± 17 m and $R \approx 10$ cm from the ATLAS interaction point. It detects inelastic proton-proton collisions in the forward region, allowing an online instantaneous luminosity measurement.

The absolute luminosity for ATLAS (ALFA) detector [52] measures very forward particles, located at $|Z| = \pm 280$ m from the ATLAS interaction point. The total cross section can be calculated by measuring the very small elastic scattering angles of protons.

LUCID and ALFA measurements combine to give a value of the total integrated luminosity and in 2012 the uncertainty on this measurement was 2.8%.

3.3.7 Data Processing

As the condition of the beams alter with time, the events are tagged with a time-stamp so they can be grouped into similar blocks. The smallest block is known as a 'lumiblock' and typically last 30 minutes unless the beam or detector conditions alter significantly within that time. Data is also grouped by the 'run number' where each fill of the LHC from the accelerator complex has a unique number. A run can last from 10 minutes to 14 hours. Runs are aborted when the beam becomes unstable or the luminosity drops below acceptable levels. Multiple runs are combined into 'periods'. Each period is given an associated letter (e.g. `periodB`) and contains runs with proton beams of similar properties.

The data are stored in various formats, from being read off the detector electronics through to being accessible for physics analysis. The format used in this analysis was the 'Ntuple' stage which allows access to groups of pre-computed objects which have passed loose selection requirements. The 8 TeV `NTUP_COMMON` Ntuples were used through this thesis.

The data are stored in computing farms at CERN but are also replicated to multiple world-wide sites and are accessible over the LHC Computing Grid [53]. This setup allows the analyser to send the analysis code to the data rather than download the (often huge) data-sets to local storage.

3.3.8 Simulation

The physics models that are to be compared to data must be on the same footing as the data collected from ATLAS. To this end, the physics models are used to generate MC events from proton-proton collisions modelling those produced in ATLAS under the 2012 running conditions. These events are then passed through a simulation of the detector using the GEANT [54] code resulting in a simulated output from ATLAS exactly like that used for data. From this stage the MC events follow the same route as the collision data events and are made into Ntuples for analysis. This careful procedure ensures the physics model and collision data have experienced an analysis chain as similar as possible such that event

875 characteristics should be altered (if at all) in exactly the same manner.

Chapter 4

Object Reconstruction

It may seem a trivial question but when asking "What particles did ATLAS just see?", there is often no clear answer. Particle types are identified by their interactions with the detector components. Often subtle differences in these interactions can distinguish what particle type created the track or shower in question. Detector information is fed into reconstruction algorithms that process a variety of characteristics in order to determine the most likely particles to have been produced in the collision hard-process. The aim of the reconstruction is to transform the raw data into final state kinematics for analysis use. An overview of the tracking algorithms common to multiple final-state particles is given in Section 4.1. The rest of this chapter discusses how we identify electrons (Section 4.2), muons (Section 4.3), jets (Section 4.4) and $E_{\mathrm{T}}^{\mathrm{miss}}$ (Section 4.5). Overlapping final-state particles are discussed in Section 4.6 and the definitions of the truth selection, applied to MC samples, is discussed in Section 4.7.

Heavy particles that can decay to lighter ones often have very short lifetimes. A particle will decay before it can reach the detecting components of ATLAS if the lifetime is smaller than $\sim 2 \times 10^{-10}$ s in the lab frame. Only a few particles are stable enough to be observed in ATLAS, simplifying the required number of search algorithms. Identification algorithms use known characteristics of particles to identify their detector signatures.

4.1 Tracking And Vertexing

There are various strategies to reconstruct tracks from charged particle ID hits (see Section 3.3.2) [55]. For each event, there are many hits in the ID caused by low energy particles or electronic noise that have to be filtered out from high momentum particles' signals. Sophisticated algorithms process the high volume of hits in all three ID components efficiently in order to find the tracks of high p_{T} particles. The hits in the ID are transformed into three dimensional "space-points". This is a trivial process in the pixel detector, where the space-point is simply the pixel coordinate. The front side of an SCT strip is only able to record a ϕ and R coordinate with no distinction in the Z direction. The back side of an SCT module has a small stereo angle, so a hit in both sides can be used to determine a Z coordinate.

The primary algorithm for track reconstruction is the *inside-out* sequence [56], which seeds a track with space-points from the pixel detector and SCT. The track is built by iteratively adding space-points and recalculating the quality of the track fit (using a χ^2 metric). Only space-points within a region-of-interest, defined by the a pt dependent cone around the track's current direction, are considered in this reconstruction. If the track passes certain quality requirements the algorithm extends the track to the TRT in search of further space-points to add to the track. This sequence finds the majority of high p_{T} tracks which originate from a primary vertex. The related *outside-in* sequence then carries out this process in reverse, looking for remaining, unused TRT space-points to produce track seeds and extrapolate back to the SCT and pixel detector. This sequence is designed to find tracks missed by the *inside-out* sequence, as well as tracks that come from a secondary vertex, i.e. a vertex that is significantly displaced from the nominal interaction point. This procedure is able to find tracks with $p_{\mathrm{T}} > 400$ MeV.

Finally, a vertex finding algorithm is run to associate tracks to vertices where extrapolated tracks converge [57]. This stage also incorporates finding secondary and photon conversion* vertices.

*Photon conversion occurs when a photon interacts with the detector to form an electron and positron.

4.2 Electrons

Electrons do not decay and are in the final-state of many analyses. The detector signature for an electron is a track in the ID matched with an energy deposit in the EM Calo. The requirement to have an ID track in the identification means electrons can only be found in the central region, $|\eta| < 2.5$.

4.2.1 Trigger

The tracks for potential electrons are found as explained in Section 4.1. A detailed discussion of the electron trigger is given in [58]. The energy deposited in the EM Calo is clustered by the L1 trigger into *towers* by summing over R in $\Delta\eta \times \Delta\phi \approx 0.1 \times 0.1$ blocks over all calorimeter layers [59]. Local maxima are found using a 'sliding window' algorithm, scanning across the towers in 4×4, η-ϕ windows. If the central 2×2 towers have a combined energy above the threshold of the trigger, the L1 electron requirements are satisfied. The efficiency for finding an electron with this procedure is 99% for $E_{\mathrm{T}}>$ 7 GeV, increasing to 99.9% for $E_{\mathrm{T}}>$ 45 GeV.

The L2 electron trigger is seeded by the the highest energy cell in the RoI found by the L1 trigger. The L2 trigger then recaluclates the electron track using the the inside-out procedure and a RoI from the L1, allowing a more accurate track to be founddue to strciter requirements on track quality. Due to strict time constraints the L2 trigger is restricted to energy measurements from only the second layer of the EM Calo. The cluster position is calculated as an energy weighted average 3×7 (5×5) cell position, when the cluster is in the barrel, $|\eta| < 1.4$ (end-cap, $1.4 < |\eta| < 2.47$), centred on the seed. The tracks available to the L2 trigger are limited to those pointing towards the RoI in η-ϕ coordinates and originating from the primary vertex. The trigger is passed if the p_{T} of the electron candidate is above the L2 trigger threshold.

The EF electron trigger has considerably more time to make a decision and so is able to use offline-like algorithms to reconstruct the electron candidate. Calorimeter information in

L1 Trigger	L2 Trigger	Event Filter
L1_EM18VH	L2_e24vhi_medium1	EF_e24vhi_medium1
L1_EM30	L2_e60_medium1	EF_e60_medium1

Table 4.1: Electron trigger flow used in this analysis.

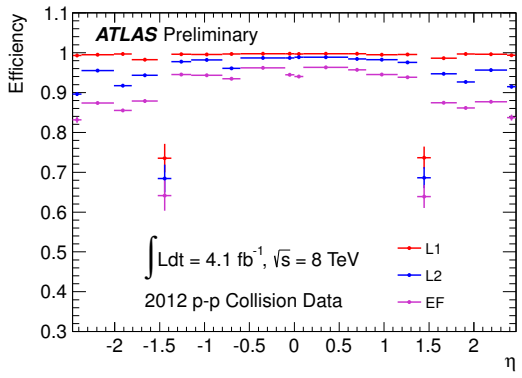
an area slightly larger than the L2 RoI is retrieved and a sliding window algorithm builds energy clusters to which all offline based corrections are applied. The EF tracks use an adapted version of the offline reconstruction and are limited to those tracks available in the RoI. The EF trigger is passed if the electron candidate again passes the trigger p_{T} threshold and the event data is then written out to storage.

Throughout the entire 8 TeV ATLAS data-set two triggers were in operation for electrons with no pre-scale applied[†]. These triggers are `EF_e24vhi_medium1` and `EF_e60_medium1`, described below. In the analysis, events are selected by applying a logical OR between these two triggers. Table 4.1 shows the corresponding L1, L2 and EF triggers used for the electrons in this analysis.

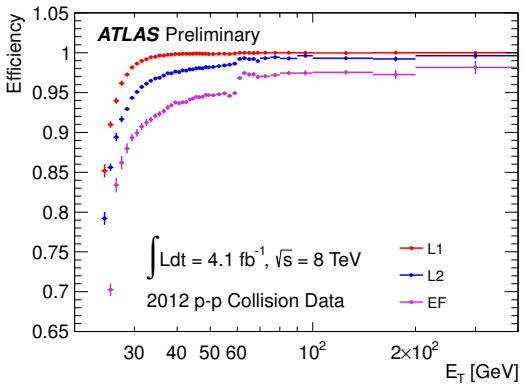
The `EF_e24vhi_medium1` trigger requires an electron candidate with uncalibrated p_{T} of at least 24 GeV at the L2 and EF stages. The `i` in the trigger name denotes an additional requirement that the electron must be isolated, such that the electron track contributes at least 90% of the summed track p_{T} in a cone of $\Delta R < 0.2$. The `v` denotes that the threshold (24 GeV) varies in η. This variation accounts for regions of the detector where the p_{T} measurement is less precise and hence the threshold is lowered to account for mismeasurement. The variation is less the 2 GeV. The `h` denotes the veto of electron candidates with hadronic leakage. Any electron with energy ≥ 1 GeV in the hadronic calorimeter fails the filter requirements. Finally, the `medium1` refers to set of identification cuts which are loosely similar to the offline identification described in Section 4.2.2. The identification at trigger level applies, amongst others, requirements on the lateral width of the energy deposit and the number of space-points in the ID.

The `EF_e60_medium1` trigger requires an electron of uncalibrated p_{T} greater than 60 GeV

[†]A pre-scale of N means only 1 in every N triggered events is stored. Pre-scales are used for high rate triggers, where storing all events would exceed the available bandwidth.



(a) Electron trigger efficiency as a function of electron η.



(b) Electron trigger efficiency as a function of electron E_{T}.

Figure 4.1: Electron trigger efficiencies for an OR between the two triggers used in this analysis. Taken from [60].

but does not require the isolation and hadronic leakage requirements mentioned above for the low-p_{T} trigger. The justification for dropping the isolation requirement is that the p_{T} distribution of jets misreconstructed as electrons falls rapidly as p_{T} increases so an electron candidate with $p_{\mathrm{T}} > 60$ GeV is more likely to be an electron.

The application of an OR of the two triggers increases the efficiency of high p_{T} electrons whilst requiring the lower p_{T} electrons to be isolated. The efficiency of the triggers is studied using "tag-and-probe" methods using $Z \to ee$ and $J/\psi \to ee$ data. One electron passes the trigger and is designated the "tag". The second electron from the Z or J/ψ decay, is designated the "probe" if the two electrons have a combined mass consistent with the parent Z or J/ψ. The efficiency of the trigger is the fraction of probes that pass the trigger. Figure 4.1(a) shows the efficiency for an electron to pass an OR between the two electron triggers as a function of electron η and Figure 4.1(b) shows the same efficiency as a function of E_{T}. The higher level triggers include in efficiencies from the upstream levels.

4.2.2 Offline Reconstruction

The online reconstruction ensures a high efficiency for selecting electrons but allows many misreconstructed or non-prompt electrons to pass the selection. Prompt electrons are produced in the hard-scatter and are what we aim to select in this analysis. 0.025×0.025

energy towers in η-ϕ space are formed in the EM Calo using a sliding window algorithm. Each of the three layers of the EM accordion calorimeter is re-clustered separately, giving information on the longitudinal shower shape. If a tower is found to have $p_{\mathrm{T}} > 5$ GeV and matches an ID track, an electron candidate is formed. To account for the increased track bending after bremsstrahlung the matching in ϕ, the bending plane of the solenoid magnet, is asymmetric. If the track is matched to $|\Delta\eta| < 0.05$ and $|\Delta\phi| < 0.1(0.05)$ of the centre of the cluster it is bending towards (away from), it is considered matched to the cluster. The electron candidate four-momentum is the found from the combined track and cluster, corrected for average expected energy losses.

In order to reduce the background from non-prompt electrons, further offline identification criteria are derived by the "`egamma`" performance group. The criteria come in three grades of increasing prompt electron purity but reducing signal efficiency named *loose*, *medium* and *tight* [61]. As we require high signal purity in our selection, the *tight* requirements are applied in this analysis. The full set of requirements are given in [61], a sub-set for the tight identification is outlined here:

- Low impact parameter ($|d_0| < 5$ mm).
- High number of ID hits (≥ 1 Pixel, ≥ 7 SCT).
- Low ΔR between track and cluster.
- b-layer hit to suppress conversion electrons.

The identification requirements are validated using MC and "tag-and-probe" methods using $Z \to ee$ and $J/\psi \to ee$ data. The efficiency of the identification is then the fraction of probes that pass the cuts.

To further reduce the contamination from non-prompt electrons from heavy-flavour-quark decays, isolation requirements are applied. Heavy flavour decays tend to have a high particle multiplicity concentrated in a small ΔR area. The isolation cuts ensure that the electron is the most energetic, or highest momentum, object in the vicinity.

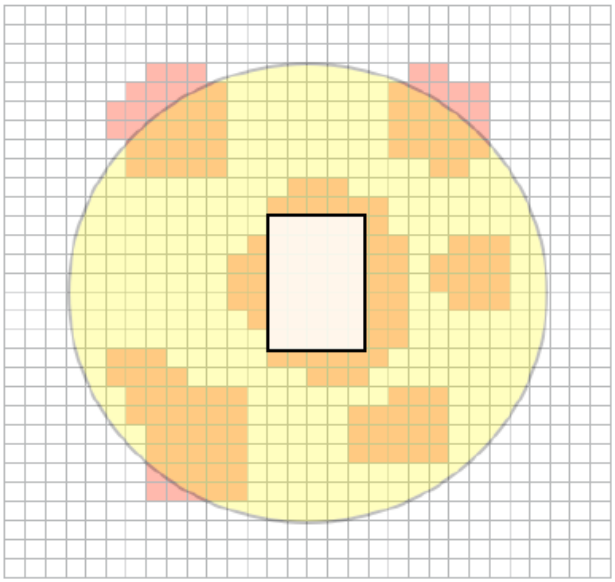


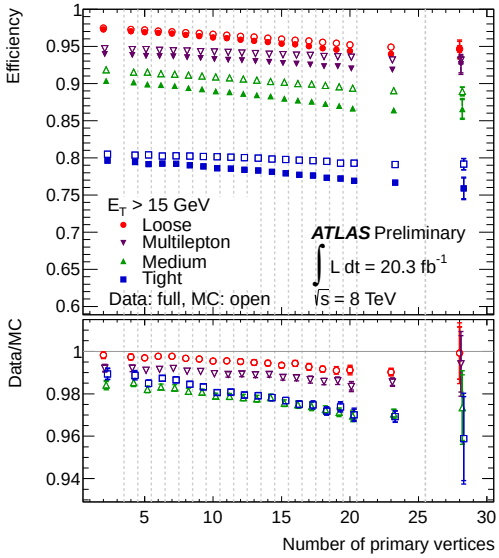
Figure 4.2: Electron calorimeter isolation showing the grid of calorimeter cells. The yellow circle shows a $\Delta R < 0.4$ cone around the electron cluster. The central $\Delta\eta \times \Delta\phi = 5 \times 7$ is excluded and the remaining yellow cell's energy is summed. Taken from [62].

Track isolation - The momenta of all tracks, excluding the electron, within $\Delta R < 0.2$ around the electron track are summed to give p_{T}^{Cone20}. In this analysis p_{T}^{Cone20} is required to be smaller than 5% of the electron's momentum.

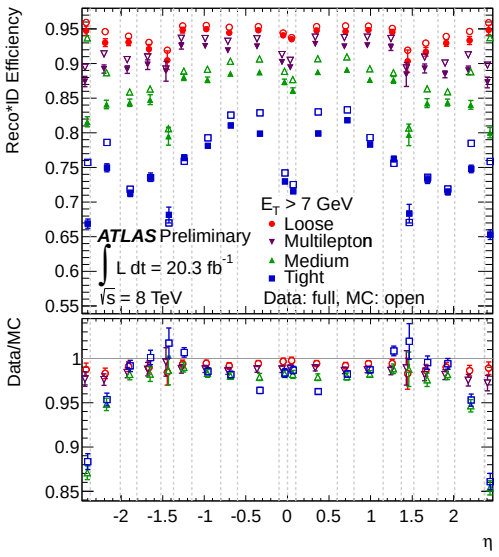
Calorimeter isolation - The EM Calo and hadronic calorimeter energy towers are summed in a $\Delta R < 0.2$ cone around the electron cluster. The central $\Delta\eta \times \Delta\phi = 5 \times 7$ central cells are excluded and the remaining energy is summed to give E_{T}^{Cone20}. This analysis requires E_{T}^{Cone20} to be below 15% of the electron cluster energy. See Figure 4.2.

Electron candidates are required to be associated to the primary vertex by cuts on the impact parameters. The transverse impact parameter d_0 is required to be within 3 standard deviations of the primary vertex and the longitudinal impact parameter is required to be $|z_0 \sin(\theta)| < 0.5$ mm.

At each step of the electron selection the efficiency must be found in order to compare data with MC [59]. The tag-and-probe method, outlined above, is used to calculate reconstruction and identification efficiency factors which are applied to the MC. The efficiency of the probe electron is found at each stage and scale factors are found by comparing the



(a) Electron identification efficiency in data and MC compared to the number of primary vertices.



(b) Electron identification × reconstruction efficiency as a function of the probe electron η.

Figure 4.3: Electron identification and reconstruction efficiency factors. The "Tight" offline requirements are applied in this analysis. Taken from [59].

data and MC efficiencies. The scale factors adjust the electron p_{T} by $< 5\%$ for electrons above 50 GeV. Figure 4.3(a) shows the electron identification efficiency differentially in the number of primary vertices. Figure 4.3(b) shows the combination of the identification and reconstruction efficiencies plotted as a function of the probe electron η.

4.2.3 Calibration

The electron energy calibration is described in detail in [56]. The standard EM Calo electron calibration is derived by probing the response energy when firing electrons of a known energy into a calorimeter module [63]. MC studies are used to investigate the expected calorimeter energy tower response differentially in p_{T} and η of the electrons. These correction factors are then applied to the observed electron.

A data-driven approach is used to validate the correction. $Z \to e^+e^-$ and $J/\psi \to e^+e^-$

events are used to ensure the line-shapes of the resonance matches between data and MC. Small additional correction factors are applied to MC electrons to ensure MC and data electrons are compared on an equal footing [61]. The MC electron energies are also smeared to match the energy resolution observed in data.

4.2.4 Analysis Electrons

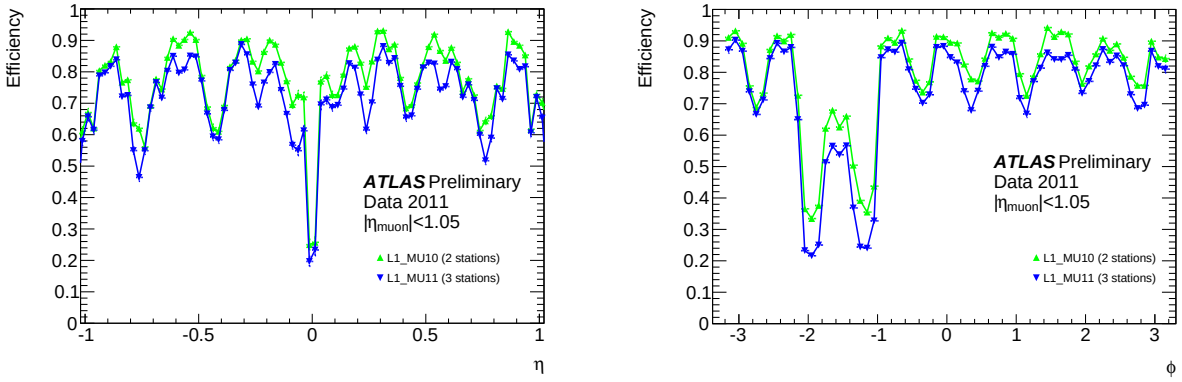
The electrons used in this analysis are required to pass all the above selection criteria. Additional cuts are applied at the analysis level. The electron is required to have $p_T > 25$ GeV in order to suppress any turn-on effects from the 24 GeV trigger. No cut is needed around the 60 GeV trigger as the vast majority of 60 GeV electrons will also pass the 24 GeV trigger and turn on effects will be minimal. Electron $|\eta|$ must be < 2.47 and not fall within the crack region ($1.37 < |\eta| < 1.52$) that suffers from reduced efficiency.

4.3 Muons

Muons are minimally ionising particles so they pass through ATLAS without losing much energy. They are the only particles to typically deposit their energy in the muon spectrometer, the outer component of ATLAS. This allows the muons to be cleanly observed with fewer backgrounds than electrons. The similarity to electrons in terms of the electroweak force (known as lepton-universality) mean the same kinematics apply to both when considering W-boson decays. Therefore, by adding the final state with a muon to this analysis we approximately double the expected event yield without altering the kinematic selection.

4.3.1 Trigger

The muon system uses the fast response trigger chambers in the Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) to seed the L1 trigger [64]. A muon track can be seeded by spatially and temporally clustered hits in three layers of the RPC or TGC. If a



(a) L1 efficiency as a function of muon η.

(b) L1 efficiency as a function of muon ϕ.

Figure 4.4: 2011 L1 muon efficiencies for the barrel region as a function of muon η and ϕ. The muon spectrometer hole location is shown by the efficiency drops at $\eta = 0$ and $\phi = -1, -2$. Taken from [65].

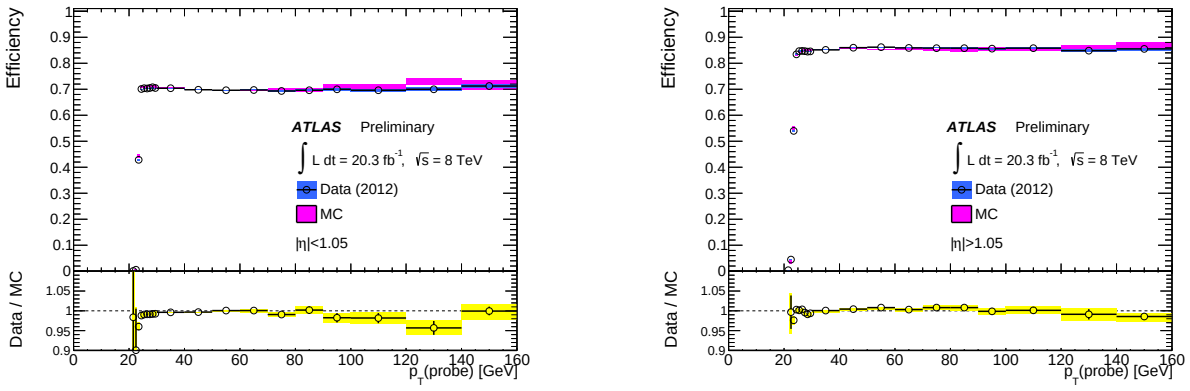
track with sufficient momenta is found at L1, a RoI is sent to the L2 trigger for further investigation. The RoI has dimensions $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ (0.03×0.03) in the RPC (TGC). Whilst these L1 trigger chambers cover $\sim 99\%$ of the end-cap geometry, they only cover $\sim 80\%$ of the barrel region ($|\eta| < 1.5$) due to a hole around $\eta = 0$ allowing services into the ID. This hole is split into two locations at $\phi = -1$ and $\phi = -2$. Unfortunately there are no plots available from the ATLAS muon performance group for 2012 data demonstrating the hole location. Figure 4.4 shows plots from 2011 L1 trigger performance studies showing the trigger efficiencies for the barrel region where the drops in efficiency show the hole location.

The L2 trigger then constructs a track from three-dimensional space-points in the Muon Drift Tubes (MDT). This MDT track is then combined with the ID track with the smallest ΔR to form a L2 Combined Muon (CB). The muon candidate only passes the L2 requirements if the track $p_{\mathrm{T}} > 24$ GeV.

The Event Filter has the full Muon Spectrometer (MS) information to evaluate the muon candidate. The MS track is formed using the full MS hit information. This MS track is then matched to the closest ID track to form the full muon candidate. This method is known as the "outside-in" algorithm, as it is seeded by the MS. An "inside-out" algorithm is then applied to remaining ID tracks, extrapolating to the MS. If a matching, unused MS track is

L1 Trigger	L2 Trigger	Event Filter
L1_MU15	L2_mu24_tight	EF_mu24i_tight
L1_MU15	L2_mu36_tight	EF_mu36_tight

Table 4.2: Muon trigger flow used in this analysis.



(a) MS barrel region.

(b) MS end-cap regions.

Figure 4.5: Muon trigger efficiency for passing either the EF_mu24i_tight or EF_mu36_tight triggers. Both taken from [64].

found, a new muon candidate is formed. Both algorithms form the offline CB.

In a similar manner to the electron trigger choice, the muon channel analysis applies an OR between one low-p_{T} isolated and one high-p_{T} non-isolated trigger. The lower p_{T} trigger is named EF_mu24i_tight and the higher p_{T} version is EF_mu36_tight, requiring muons above 24 and 36 GeV respectively. The lower p_{T} trigger also requires the muon track to be isolated such that ${p_{\mathrm{T}}}^{Cone20}$ is less than 12% of the muon track p_{T} [64]. The full trigger menu for both muon triggers is shown in Table 4.2.

As with the electron triggers, a tag-and-probe method corrects the modelling of the muon trigger efficiency using data. In this case $Z \rightarrow \mu^{+}\mu^{-}$ events are used, where one muon is tagged and the other used as a probe. The trigger scale factors are derived as the data-to-MC ratio of the probe efficiency, and are applied to the MC samples in this analysis. Figure 4.5(a) shows the efficiency in the barrel region for the two triggers used in this analysis, and Figure 4.5(b) shows the same for the end-caps.

4.3.2 Offline Reconstruction

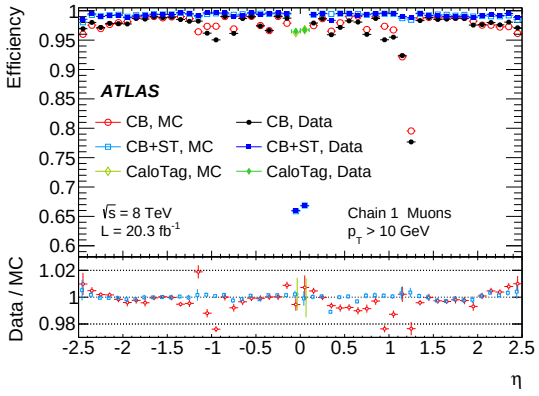
The offline reconstruction of muons is able to use all the information that ATLAS records. There are various algorithms or "chains" used for different types of analyses; this analysis exclusively uses the "chain 1" type muons [66]. This chain 1 reconstruction begins with the MS tracks. These tracks are then matched to the ID tracks using a statistical combination of track parameters considering the covariance matrices. The track is then required to pass the following criteria:

- ≥ 1 Pixel hit,
- ≥ 5 SCT hits,
- ≤ 2 missing silicon hits, where the hits are expected, and
- ≥ 9 TRT hits if the track traverses the full TRT ($|\eta| < 1.9$).

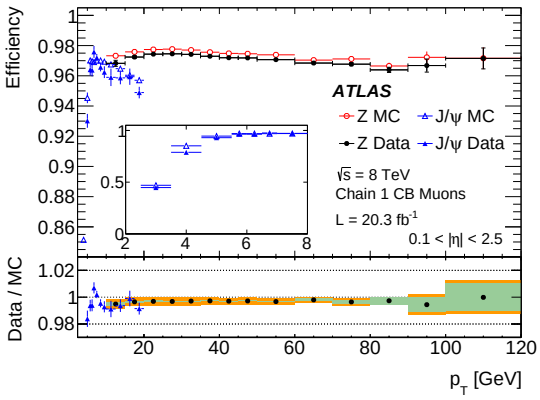
If all the above requirements are satisfied, the candidate is determined to be a muon and passed to the analysis code.

4.3.3 Calibration

The efficiency of the offline reconstruction cuts must be evaluated in both data and MC. As with the electron reconstruction efficiencies, a tag-and-probe method is used. $Z \to \mu\overline{\mu}$ and $J/\psi \to \mu\overline{\mu}$ events are found in data by tagging one muon to pass the above cuts and requiring a second, probe muon, with the muon pair satisfying a mass requirement. The probe muon is tested to see if it passes the reconstruction cuts, and the efficiency of the reconstruction is the ratio of passing candidates to total probes. The same procedure is evaluated using MC samples and the data-to-MC efficiency ratio is applied to the analysis MC as a correction factor. Figure 4.6(a) shows the reconstruction efficiency as a function of the probe muon η. Efficiency drops can be observed around η of 0, where the MS has a gap to allow ID services in, and around η of 1.4 where the MS has some missing chambers for



(a) As a function of the probe muon η.



(b) As a function of the probe muon p_{T}.

Figure 4.6: Muon reconstruction efficiency found from the tag-and-probe method. The muons used in this analysis are denoted "CB". Taken from [64].

this data-set. Figure 4.6(b) shows the reconstruction efficiency as a function of the probe muon p_{T}.

The muon momentum calibration is performed in a similar manner to the electron calibration. Calibration correction factors are derived from tag-and-probe methods [66]. The MC muon momentum is scaled and smeared (each $\sim 1\%$ effect of p_{T}) to match the momentum scale and resolution observed in data.

4.3.4 Analysis Muons

In addition to the previously mentioned muon requirements, the analysis applies several further cuts to ensure high quality prompt muons, at the recommendation of the ATLAS "EW" group. The transverse impact parameter is required to be within 3 standard deviations of the primary vertex ($|d_0/\sigma(d_0)| < 3$) and the longitudinal impact parameter ($|z_0 \sin(\theta)|$) is required to be less than 5mm from the primary vertex. The muon p_{T} has a slightly higher offline cut of 25 GeV, suppressing trigger turn on effects and bringing the muon channel in line with the electron channel. Additionally, isolation requirements are applied:

- ${E_{\mathrm{T}}}^{Cone30}/E_{\mathrm{T}} < (0.014 * p_{\mathrm{T}}) - 0.15$;
- ${E_{\mathrm{T}}}^{Cone30}/E_{\mathrm{T}} < 0.2$;

- ${p_{\mathrm{T}}}^{Cone30}/p_{\mathrm{T}} < (0.1 * p_{\mathrm{T}}) - 0.105$; and
- ${p_{\mathrm{T}}}^{Cone30}/p_{\mathrm{T}} < 0.15$.

It has been observed [67] that the additional offline cuts do not require scale factors to bring data and MC into agreement, so none are applied.

4.4 Jets

Strongly charged particles from the hard-process radiate and fragment into a high multiplicity of final-state particles. In most physics analyses we measure the final state in order to gain insight into the hard-process. We therefore need to have algorithms designed to relate the final state particles to the hard-process partons. The final-state remnants of partons are a collimated spray of energetic hadrons, designated "jets". A nice overview is available in [68]. There have been many jet algorithms proposed over the years, from a simple cone around the most energetic hadron to more sophisticated sequential processes. This analysis uses the "anti-k_t" type jets with a radius parameter of 0.4 [69].

4.4.1 Jet Algorithms

Jet definitions vary in two capacities: the conditions to combine two particles into a single jet and the recombination scheme used to find the resulting momentum of the jet. The evolution of jet definitions has been driven by the aim that jets should be simple to implement and yield consistent and intuitive results. Consistency here relates to how jets in an event change if additional low energy QCD or collinear emission is added. Certain jet definitions will return different numbers of jets in an event depending on whether they consider this extra emission or not. Algorithms that are immune to this additional emission are called infra-red safe. Intuitively a jet forms a cone like shape in the final-state with all particles collimated within a certain ΔR. Some jet definitions perform well in terms of stability over extra emissions but return un-intuitive jet shapes, quite distinct from a cone.

The FastJet software package [70] is used to combine all reconstructed particles in an event into anti-k_t jets. The anti-k_t jet algorithm returns infrared safe jets which are approximately conical in shape. The algorithm calculates a distance metric between all pairs of particles, d_{ij} for particles i and j, as well as a measure of how far each particle i is from the beam line, $d_{i,B}$:

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{-2}, p_{\mathrm{T},j}^{-2}\right)\left(\frac{\Delta R_{ij}}{R}\right)^2, \tag{4.1}$$

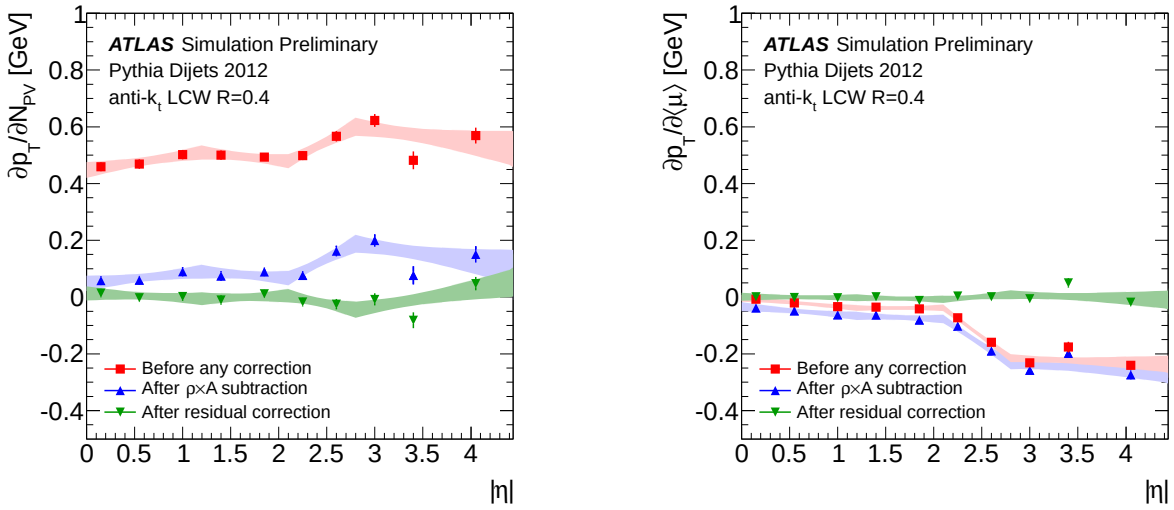
$$d_{i,B} = p_{\mathrm{T},i}^{-2}. \tag{4.2}$$

The algorithm iteratively combines particles i and j if their d_{ij} value is smaller than $d_{i,B}$. If the combined jet k has $d_{k,B} < d_{kj}$ for all j then the jet is removed from further clusterings. The algorithm begins with the smallest d_{ij} value so the "seeds" for the algorithm are the highest p_{T} particles which add lower p_{T} particles.

4.4.2 Reconstruction

ATLAS jets are constructed from energy deposits in the calorimeters ($|\eta| < 4.4$) and associated to charged particles tracks in the ID where possible ($|\eta| < 2.5$) [71, 72]. The main difficulty in creating the energy towers from the calorimeter measurements is detector noise. The energy deposits are combined into "topo-clusters". The process is seeded by a calorimeter cell with significant a signal to noise ratio (S/N > 4). Adjacent cells with a lower signal to noise threshold of S/N > 2 are then iteratively added to the tower. Finally, one extra layer of adjacent cells is combined into the tower if S/N > 0[‡]. The topo-clusters have energy and direction defined by a weighted sum of the component energy deposits.

[‡]A negative S/N value is possible due to signal shapes to reduce pile-up effects.



(a) Jet p_{T} dependence on the number of primary vertices in an event, NPV. The correction reduces in-time pile-up mismeasurements.

(b) Jet p_{T} dependence on the average number of primary vertices in an event, $\langle\mu\rangle$. The correction reduces out-of-time pile-up mismeasurements.

Figure 4.7: Change in measured jet p_{T} with differing pile up conditions as a function of $|\eta|$, and the effect of corrections. Taken from [74].

4.4.3 Calibration

The towers are clustered into jets using the anti-k_t algorithm. A calibration, called local cell signal weighting (LCW) [73], is applied to the cells to reduce fluctuations in the energy response of the calorimeter[§]. A series of four further calibration steps are then applied to each jet:

Pile-up correction [74] An estimate is made of the jet's area and the expected contributions from noise, underlying event and pile-up are subtracted from the jet's energy. This correction is dependent on the calculated number of primary vertices in the event (NPV, a measure of in-time pile-up) as well as the average number of expected collisions ($\langle\mu\rangle$, a measure of out-of-time pile-up). Figure 4.7 shows the effects of NPV and $\langle\mu\rangle$ on jet p_{T}.

Origin correction The direction of the jet is adjusted such that it points back to the primary vertex instead of the centre of the detector.

[§]Response is defined as (measured energy) / (energy at production).

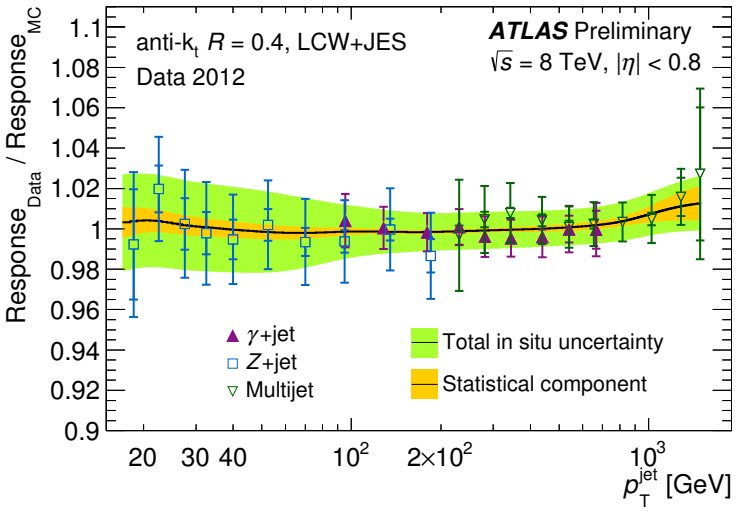


Figure 4.8: Jet response ratio of the data to the MC as a function of p_{T}. The outer band indicates the total uncertainty resulting from the combination of in-situ techniques, while the inner yellow band shows the fraction purely from statistical uncertainties. Taken from [76].

MC scale adjustment The energy and η of the jet are adjusted to the hadron-level so as to be compared to MC. The correction factors are derived from MC using the ratio of a jet's energy or η before and after detector simulation.

Residual in-situ corrections (data only) [75] Correction factors are applied to account for remaining data to MC disagreements. The corrections are derived by balancing the p_{T} between jets and well measured reference objects such as leptonically decaying Z bosons, photons or other jets. Figure 4.8 shows the results of these correction factors.

4.4.4 Analysis Jets

In addition to the above cuts, jets used in the analysis are required to pass further quality checks. Any event with a jet found to be pointing to a faulty calorimeter section is vetoed in both data and MC. This removes events where some of the energy is not reconstructed and incorrectly assigned to missing momentum (see Section 4.5). If the jet has $|\eta| < 2.5$ and associated tracks then the probability that the jet originated from the primary vertex can be checked. The jet vertex fraction (JVF) is a p_{T}-weighted measure of the fraction of tracks

associated to a jet that came from the primary vertex:

$$\mathrm{JVF}(\mathrm{jet}_i, \mathrm{PV}_j) = \frac{\sum_k p_{\mathrm{T}}(\mathrm{track}_k^{\mathrm{jet_i}}, \mathrm{PV}_j)}{\sum_n \sum_l p_{\mathrm{T}}(\mathrm{track}_l^{\mathrm{jet_i}}, \mathrm{PV}_n)}. \tag{4.3}$$

To reject jets that originated from secondary vertices a requirement of JVF > 0.75 is applied to jets with $|\eta| < 2.5$. Finally, jets are required to have $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 4.4$ to be used in the analysis.

4.4.5 b-tagging

If a jet is is within the ID acceptance of $|\eta| < 2.5$ it can be tagged as originating from a b-quark using characteristics of the tracks via the use of impact parameter variables. B-hadrons have a lifetime large enough to travel $\mathcal{O}(1\mathrm{mm})$ before decaying. This slight displacement of a secondary vertex from the primary vertex indicates the presence of a B-hadron, and the resulting jet from the secondary vertex can be tagged as a "b-jet". A multi-variate algorithm has been developed to tag every fiducial jet with a probability that it originated from a B-hadron [77].

Using MC the tagger has been shown to tag 80% of b-jets. Removing b-jets from the analysis allows a large reduction in top quark backgrounds due to the decay mode of $t \to Wb$ with $\approx 100\%$ branching fraction. As with the other object selection cuts, efficiency factors have been determined from data and MC comparisons [78].

4.5 Missing Transverse Energy

Particles which do not interact with matter can often be crucial in identifying a physics process. Neutrinos and proposed BSM particles have sufficiently low interaction cross section so as to leave almost no trace of their passing through the detector. Their presence can however be inferred by calculating the momentum imbalance in the final-state objects [79, 80].

This calculation is slightly complicated at a hadron collider as the initial Z momentum is not known since the partons collide rather than the protons. The colliding partons do however have negligible transverse momentum which must be conserved in the final state. The transverse momentum of the invisible particles, ${p_{\mathrm{T}}}^{miss}$, is the negative sum of the transverse momenta from all observed particles i, as inferred from momentum conservation:

$$\vec{p_{\mathrm{T}}}^{\,miss} = -\sum_i \vec{p_{\mathrm{T}i}}, \tag{4.4}$$

where $\vec{p_{\mathrm{T}}}^{\,miss}$ is a two-dimensional vector in the X-Y plane. Tracks can only be reconstructed out to the acceptance of the ID, $|\eta| < 2.5$. To better study the invisible particles, the transverse energy is considered instead of transverse momentum, allowing the calorimeters to be used to $|\eta| < 4.4$. In the case of particles with mass significantly lower than the LHC energy, such as neutrinos, the momentum and energy are identical. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as

$$\vec{E_T}^{\,miss} = -\sum_i \vec{E_{\mathrm{T}i}}. \tag{4.5}$$

Rather than simply sum the calorimeter towers with their default calibrations, we can select particles we have classified as certain types and apply their specialised calibrations in order to find a more accurate $E_{\mathrm{T}}^{\mathrm{miss}}$ value. For example, electrons are calibrated to the specialised electron energy scale and p_{T} corrections before entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. The $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation then becomes:

$$\vec{E_T}^{\,miss} = \vec{E_T}^{\,miss,e} + \vec{E_T}^{\,miss,\mu} + \vec{E_T}^{\,miss,\tau} + \vec{E_T}^{\,miss,\gamma} + \vec{E_T}^{\,miss,\text{ Jets}} + \vec{E_T}^{\,miss,\text{ Soft terms}}. \tag{4.6}$$

The first five terms simply represent the calibrated versions of Equation 4.5 for electrons, muons, τ leptons, photons and jets respectively. The photons and τ leptons have similar

calibration techniques to those described above for electrons and muons but do not feature heavily in this analysis so are not described here. The photon calibration is described in [56] and the τ leptons in [81]. The final term in Equation 4.6, denoted "Soft terms", adds in calorimeter towers that have not been assigned to any previous object. These unassigned towers can be caused by particles that fail the selection requirements or have small p_{T}.

4.6 Overlapping Objects

Reconstructed objects that overlap one another in η-ϕ space cause ambiguities as to what type of particle was the source. To resolve the ambiguities a priorities list is enacted that removes all overlapping objects that are not at the top of the list. For this analysis the overlap removal eliminates jets that fall within $\Delta R < 0.3$ of an electron or muon. The jet is not available to the analysis jet collection but is still considered in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation.

4.7 Truth Selection

It is often useful to investigate the MC predictions for a particular process before they have been passed through the detector simulation. This allows characteristics of the process to be identified (e.g. lepton rapidity) without the need for a complicated detector simulation. Different MC generators can produce the events at different points in the hadronisation process; some do not include intermediate particles, such as W-bosons, in the event record. Therefore, to compare predictions from different generators it is important to have a common definition of the level at which the events should be compared and what exactly is meant by a jet or lepton at this level.

Leptons

The leptons in the final state of the MC event record are known as "bare leptons". To make them more comparable to data, all final state photons in a ΔR cone of < 0.1 are summed

into the 4-vector of the lepton. After this photon recombination, the leptons are known as "dressed leptons". The truth level studies in this analysis use the dressed leptons.

$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$

The truth $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the 4-vector sum of all neutrinos in the final state. Whilst this means the truth $E_{\mathrm{T}}^{\mathrm{miss}}$ does not exactly represent the neutrino that comes from the hard-scatter W-boson in our EW $W + 2$ jets signal process, it gives a more accurate representation of the W-boson that would be reconstructed after detector simulation. The p_{T} of the hardest neutrino tends to be $\sim 100\times$ that of the next hardest, so the difference is at the percent level.

Jets

Truth jets are constructed using the FastJet package using the anti-k_t algorithm described in Section 4.4. These jets include all final state particles except those which are not observed in the calorimeter at reconstruction level, namely the muons and neutrinos. The reconstruction level overlap removal rules are applied at truth level so high p_{T} charged leptons will cause nearby jets to be removed, just as for reconstruction level.

Jets are tagged as b-jets and removed if a B-hadron with $p_{\mathrm{T}} > 5$ GeV is found within $\Delta R < 0.3$ of the jet. As for reconstruction level any b-jet found in the event causes the event to be vetoed.

Chapter 5

Modelling

In order to compare the observed data to the theoretical predictions, MC samples were generated as statistical representations of the fully differential calculation. In addition to the variety of MC generated samples used in this analysis, a completely data-driven technique was used to estimate the multijets background. This chapter describes the samples used to model the signal and background processes relevant to this analysis. Section 5.1 describes the MC modelling of the signal and backgrounds and Section 5.2 describes the data-driven methods used to model the multijets background.

The aim of this analysis is to select EW $W + 2$ jets events from the data and analyse their properties. This selection is not able to pick out only the signal events howveer as there are multiple physics processes that can produce the same detector signature as the signal. We therefore must consdier these other processes using theoretical predictions of their kinematics. The most advanced (highest-order) predictions have been used wherever possible. The multijets process is impracticle to simulate with MC methods and so a data-driven method is used to generate a representation of this background.

5.1 Monte Carlo Modelling

This section describes the MC models used in the analysis. A summary of the MC samples is given in Table 5.1 noting the generation techniques and features of each.

5.1.1 Signal Modelling

EW $W \to e\nu_e/\mu\nu_\mu$ + 2 Jets

The EW $W + 2$ jets (electron and muon channels) signal is modelled with the EW Wjj implementation [33] in the PowHeg generator [82–84]. Only the signal diagrams of the types shown in Figures 2.5 are included in this signal sample; diboson contributions are considered as background. The PowHeg code models the QCD interactions of this signal process to NLO but, the EW interactions are at LO which means the emission of a high p_{T} photon will not be well modelled. All PowHeg samples in this analysis use the NLO CT10 PDF [85] and are interfaced to Pythia8 [86] for the parton shower and hadronisation. It has recently been suggested that the NLO EW corrections in the similar EW $Z + 2$ jets process could contribute negatively to around 5% at high M_{jj} [87] but an MC model of these corrections is not currently available.

A previous thesis documenting this analysis at 7 TeV [36] utilised the LO Sherpa 1.4.1 MC code [30] to generate samples for the EW signal. These samples were created again for this 8 TeV analysis before being superseded by the NLO PowHeg samples. The Sherpa signal samples are used to cross-check the results of the signal extraction.

Further signal cross-checks were performed comparing to the VBFNLO 2.7.0 code [88–90] which is able to output cross sections at both LO and NLO but only able to generate LO events at parton level. VBFNLO is used as the theoretical prediction for the extracted signal cross section.

EW $W \to \tau\nu_\tau$ + 2 Jets

The EW production of $W + 2$ jets in the τ channel is included in the signal definition for both electron and muon channels and is modelled by the SHERPA 1.4.1 with the CT10 PDF. As this process is expected to contribute $\mathcal{O}(1\%)$ of the signal in the measurement region, it was not updated to POWHEG.

5.1.2 QCD $W + 2$ jets Modelling

The QCD production of $W + 2$ jets presents a large, irreducible background in the search for EW $W + 2$ jets. Therefore, the modelling of this background must be accurate in order to extract the signal component. The POWHEG generator provides events accurate to NLO in QCD interactions for this background [91]. As in the signal sample, the POWHEG modelling of this background is accurate to NLO in QCD and uses the CT10 PDF. All leptonically decaying $W +$ jets channels (e, μ and τ) are included.

Due to the POWHEG sample only becoming available in the later stages of this analysis, most of the analysis was developed using a SHERPA version 1.4.1 sample of this same background. This sample again used the CT10 PDF but is only accurate to LO in QCD emissions for inclusive W production. The MENLOPS prescription [92] is used to include LO real emission diagrams for up to 5 partons. A merging scale is applied to avoid the low energy divergences.

The POWHEG modelling of this background demonstrates a marked improvement (when compared to data) over SHERPA, especially in the M_{jj} distribution, due to the consideration of NLO QCD emissions. As the POWHEG modelling of this process has at least two partons from the hard-scatter, it omits the modelling of W + 0 or + 1 jets. It is possible that a W + 0/1 jet hard-scatter process could occur in the same bunch crossing as a pile-up collision that results in further jets. This process is extremely rare and omitted from the modelling when using the POWHEG QCD $W + 2$ jets samples. To ensure the missing process is not important, detector level plots are shown in Appendix A for using the SHERPA QCD $W + 2$

jets samples in place of the POWHEG. The SHERPA samples include the W +0/1 jet events as well as the pile-up collisions.

The POWHEG $W + 2$ jets normalisation is determined as part of the signal extraction fit (see Chapter 9). To minimise uncertainties, a data based correction is applied to the M_{jj} distribution, as described in Chapter 8.

At a late stage in this analysis a problem was found with the POWHEG QCD $W +$ 2 jets samples and the decay angles between the W-boson and the leptons. This issue is explained in Appendix B and is found to have negligible affect (0.6%) on the fit variable and the analysis results.

5.1.3 $Z + 2$ jets

$Z + 2$ jets production can mimic the signal final state if one of the charged leptons from the Z-boson decay is not reconstructed. The detector signature is then one high momentum lepton with $E_{\mathrm{T}}^{\mathrm{miss}}$ from the missing lepton.

QCD $Z + 2$ jets Modelling

As for the $W + 2$ jets process, the jets in $Z + 2$ jets events can be induced by either QCD or EW processes. The QCD $Z + 2$ jets MC samples are generated at NLO with POWHEG and PYTHIA8, using the CT10 PDF set. The theoretical uncertainty on the cross section is taken from [93].

5.1.4 EW $Z + 2$ jets Modelling

The EW $Z + 2$ jets samples are taken from the ATLAS VBF Z-boson analysis [35]. They are generated at LO with SHERPA using the CT10 PDF. The theoretical uncertainty is taken from the ATLAS paper [35].

5.1.5 Diboson

Diboson production, either WW, WZ or ZZ (see Figure 2.7), can reproduce the signal signature as follows. If a WW or WZ event has one W decaying leptonically and the other boson decaying hadronically, the final state consists of two jets, a charged lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$. ZZ events produce the signal final state when one Z decays hadronically and the other leptonically and one of the charged leptons is not reconstructed.

All diboson samples used in this analysis are generated at LO using the Herwig 6 generator [94] and the CTEQ6L1 PDF [95, 96]. The theoretical uncertainty is taken from [97].

5.1.6 Top Pair Production

A pair of produced top quarks will decay with $\approx 100\%$ branching fraction, into a W-boson and a b-quark. If one W-boson decays leptonically and the other hadronically then we have a final state with a charged lepton, $E_{\mathrm{T}}^{\mathrm{miss}}$, and four jets. This decay chain can result in top pair events producing the signal final state and must be accounted for in the analysis.

The $t\bar{t}$ process is modelled with the PowHeg generator at NLO using the CTEQ6L1 PDF. A k-factor of 1.199 is applied to the cross section to match the next-to-next-to-leading order (NNLO) prediction [98]. The uncertainties on this sample are taken from [99].

Accurate modelling of $t\bar{t}$ is of great importance in regions of high jet multiplicity so an alternative sample was also investigated and the resulting affect on the signal extraction taken as an uncertainty. The alternative sample is modelled at NLO with Mc@Nlo using the CT10 PDF. Again, a k-factor is applied to bring the cross section to the NNLO prediction.

5.1.7 Single Top Production

A single top quark can be produced in the t-channel ($qg \to t\bar{b}q'$), s-channel ($q\bar{q} \to t\bar{b}$) or in association with W-boson. The contribution of single-top events to the signal phase space is

much smaller than for the top pair production due to the smaller cross section.

The t-channel single top samples are simulated with the AcerMC generator [100] with the CTEQ6L1 PDF at LO. The s-channel contribution is generated at NLO with MC@NLO using the CT10 PDF. The sample normalisation and uncertainties are from [101].

The Wt production process is generated at NLO with MC@NLO using the CT10 PDF. The theoretical uncertainty is taken from [102].

5.2 Data-Driven Multijets Template

By far the most probable outcome of proton-proton collisions at the LHC is multijet production due to the high energies of the collisions and the comparatively large strength of the strong force. This process—designated "multijets"—results in a multitude of strongly-charged particles being produced which give rise to jets in the final state. The jets can be misreconstructed as an electron or muon and any energy mismeasurement or loss of particles outside the detector acceptance will contribute to the missing transverse energy. This process is therefore a background to the search for EW $W + 2$ jets production. The $E_{\mathrm{T}}^{\mathrm{miss}}$ in these multijets events is peaked at low values. The p_{T} of the final state jets is also peaked at low values due to the preference for low momentum exchange in the hard-process. As the final state signature of multijets events tends to have low total momenta it is rare that a multijets event can mimic our EW $W + 2$ jets signal, but due to the huge number of collisions that produce multijets events, it is a background that must be considered.

Whilst it is possible to produce MC events for this multijets background, the acceptance into the signal phase-space is so small that a huge number of events would need to be generated to achieve reasonable statistical uncertainty. Therefore, practically speaking it is easier to construct a data-driven template for the multijets background rather than use MC. The data-driven multijets templates are constructed by considering typical detector signatures for a jet to be misreconstructed as a lepton. Selection cuts are then constructed which attempt to capture only events where this misreconstruction occurs whilst being or-

Physics Process	Generator	ATLAS Data Set ID	PDF	QCD order	Cross section error (%)	Default
EW $W + 2$ jets (e, Signal)	PowHeg + Pythia8	185847 & 185848	CT10	NLO	-	✓
EW $W + 2$ jets (μ, Signal)	PowHeg + Pythia8	185849 & 185856	CT10	NLO	-	✓
EW $W + 2$ jets (τ, Signal)	Sherpa	129917	CT10	LO	-	✓
EW $W + 2$ jets (e, Signal)	Sherpa	129915	CT10	LO	-	×
EW $W + 2$ jets (μ, Signal)	Sherpa	129916	CT10	LO	-	×
QCD $W + 2$ jets (e)	PowHeg + Pythia8	185836 & 185837	CT10	NLO	-	✓
QCD $W + 2$ jets (μ)	PowHeg + Pythia8	185696 & 185703	CT10	NLO	-	✓
QCD $W + 2$ jets (τ)	PowHeg + Pythia8	147805 & 147802	CT10	NLO	-	✓
QCD $W +$ jets (e)	Sherpa	129929 & 147774	CT10	LO	-	×
QCD $W +$ jets (μ)	Sherpa	129930 & 147775	CT10	LO	-	×
QCD $W +$ jets (τ)	Sherpa	147776	CT10	LO	-	×
Diboson ZZ, WZ, WW	Herwig	105986, 105987 & 105985	CTEQ6L1	LO	6 (NLO)	✓
EW $Z + 2$ jets (e)	Sherpa	129921	CT10	LO	12 (NLO)	✓
EW $Z + 2$ jets (μ)	Sherpa	129922	CT10	LO	12 (NLO)	✓
QCD $Z + 2$ jets (e)	PowHeg + Pythia8	147806	CT10	NLO	6 (NLO)	✓
QCD $Z + 2$ jets (μ)	PowHeg + Pythia8	147807	CT10	NLO	6 (NLO)	✓
QCD $Z + 2$ jets (τ)	PowHeg + Pythia8	147808	CT10	NLO	6 (NLO)	✓
$t\bar{t}$	PowHeg + Pythia8	110404	CTEQ6L1	NLO	10 (NNLO)	✓
$t\bar{t}$	Mc@Nlo + Herwig++	105200	CT10	NLO	10 (NNLO)	×
t-channel Single top (e)	AcerMC	117360	CTEQ6L1	LO	5 (NLO)	✓
t-channel Single top (μ)	AcerMC	117361	CTEQ6L1	LO	5 (NLO)	✓
t-channel Single top (τ)	AcerMC	117362	CTEQ6L1	LO	5 (NLO)	✓
s-channel Single top (e)	Mc@Nlo	108343	CT10	NLO	5 (NLO)	✓
s-channel Single top (μ)	Mc@Nlo	108344	CT10	NLO	5 (NLO)	✓
s-channel Single top (τ)	Mc@Nlo	108345	CT10	NLO	5 (NLO)	✓
Wt	Mc@Nlo	108346	CT10	NLO	7 (NLO)	✓
Multijets	Data-driven	-	-	-	See Section 5.2	✓

Table 5.1: MC samples summary. Where "cross section error" has been left blank the normalisation error us determined from the signal yield fit. The "Default" column shows the default samples for a process with a "✓" and alternative samples for cross-checks with a "×".

thogonal to the analysis selection. This selection is applied to all data events to produce distributions dominated by multijet events. It is hypothesised that the shapes in this orthogonal, or "multijets", selection are similar to the distribution shapes, in all variables, that the multijets background form in the signal selection. This approach was demonstrated to be valid in [103]. In this way we produce a template for the multijets background. The multijets-selection is chosen via the inversion of a charged-lepton based cut. This means the variable of most imporatance to this analysis, the M_{jj}, should be unaffected by this reversal as the M_{jj} in a mutlijet event should be uncorreltated to the the manner in which the jet fakes the precense of the charged lepton.

A small number of events from MC samples discussed in Section 5.1 will pass the multijets selection. These events fail the signal selection by definition, as the multijets and signal selections are orthogonal. Therefore, we know these MC events should not be in the signal phase space so the multijets selection is thus contaminated by a small number of events from the MC processes. The multijets selection is therefore run on the MC samples discussed in Section 5.1 to find events which we know are from non-multijets origins but pass the multijets selection. These events are subtracted from the multijets events so we are left with multijet sample containing events which is not modelled by any of the MC samples. The subtraction changes the integral of the multijets selection by $< 2\%$.

In addition to the kinematic distributions, the normalisation of the multijets contribution must be determined. Although the shapes are comparable between the multijets and signal selections, the number of events passing the multijets selection does not a-priori determine the number of multijets events that pass the signal selection. A fit is therefore performed to extract a normalisation factor for the multijets templates. The fits use variables that are expected to have shape differences between multijets and the other processes. The $E_{\mathrm{T}}^{\mathrm{miss}}$ shapes show the multijets peaking at low values due to there being no $E_{\mathrm{T}}^{\mathrm{miss}}$ from high p_{T} neutrinos, the other processes have a real W-boson and hence a high p_{T} neutrino. The lack of W-boson in the multijets can also be seen in the W transverse mass (M_{T}) distribution (M_{T} is defined in Equation 7.2), so it can also be used to fit for the normalisation. The QCD W

+ jets normalisation is allowed to float freely in the fits as its normalisation is not precisely known. The resulting QCD W + jets scale factor is compared between the multijets and the signal extraction fit (see Chapter 9). All other samples are normalised to their predicted SM cross sections in this procedure as they have all been measured in previous studies and are relativly minor contributions.

The normalisation factor for the multijets template may differ depending on the phase space being considered. Ideally the fit would use a phase space as close to the measurement region as possible in order to get the scale factor most appropriate to the signal extraction. As the analysis aim is to measure the contribution from the EW $W + 2$ jets signal in data, it cannot be assumed to the signal lies at its SM predicted cross section. Therefore, we cannot assume the signal template exactly represents the signal contribution in data. Instead we must find a phase space that is close to the measurement region such that the multijets kinematics are similar but where any change in the signal contribution will have minimal effect on the multijets scale factor. The chosen phase space is identical to the measurement region only with the M_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ cuts removed so as to increase the multijets contribution and improve the statistical precision of the fit. To ensure this phase space does not make the multijets scale factor dependant on the signal contribution, the signal size is doubled and halved and a negligible effect is observed on the multijets scale factor.

In each fit used to determine the multijets normalisation a metric of the fit quality is given by the χ^2 divided by the number of degrees of freedom (NDF) before and after the fit. The χ^2 is given by:

$$\chi^2 = \sum_{i=0}^{Bins} \frac{(P_i - D_i)^2}{(\Delta P_i)^2 + (\Delta D_i)^2}, \tag{5.1}$$

where P_i represents the number of predicted events in bin i and D_i is the number of data events. ΔP_i and ΔD_i are the statistical errors on P_i and D_i respectively.

5.2.1 Electron Channel

The signal in ATLAS for an electron is similar to that for a jet, the main difference being the TRT identification of an electron and the calorimeter shower profile. It is possible for a jet to mimic these signals, but the jet will not generally be concentrated in a single particle but rather spread across multiple collimated particles. Therefore, hadronic jets can be identified by the number of charged particles in the vicinity of the electron candidate. This is quantified by the track isolation variable defined in Section 4.2.2. The multijets selection in the electron channel is shown in Table 5.2.

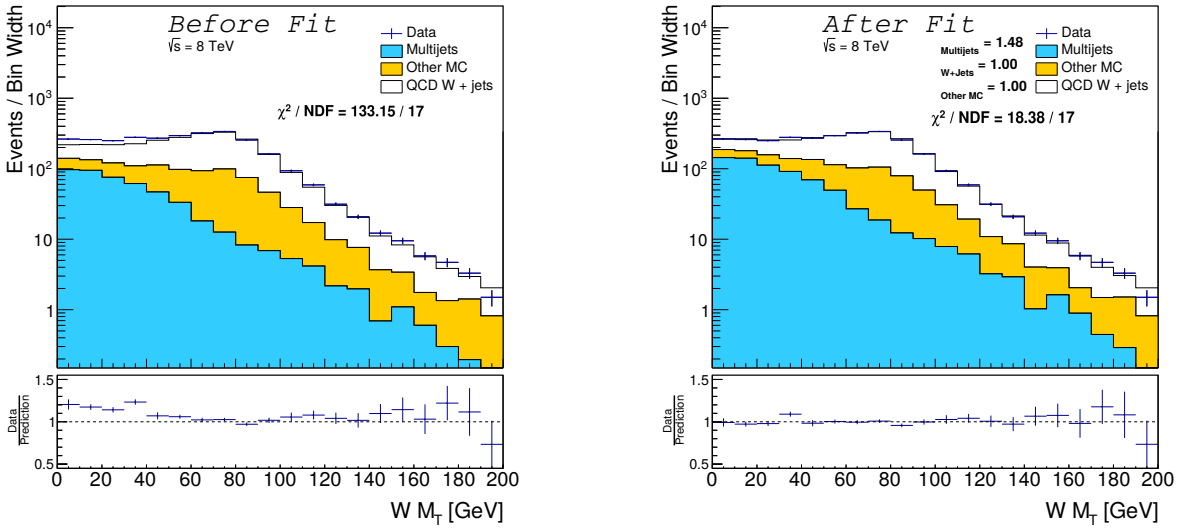
Electron Cut	Nominal Selection	Multijets Selection
Identification requirements	"tight"	"medium" or "tight"
Track Isolation	< 0.05	$0.05 <$ isolation < 0.09

Table 5.2: Electron multijets template definition compared to the nominal selection.

The loosened identification requirement captures more jets misreconstructed as electrons to increase the number of events in the resulting template. The altered track isolation cut means the multijets electrons have more, higher p_{T} tracks around them than the nominal electrons. The isolation cut is made orthogonal to the nominal selection to ensure no overlap and extends out to close to the trigger upper bound of 0.1. The selection upper bound of 0.09 is purposefully away from the trigger threshold to avoid any edge effects caused by uncalibrated p_{T} values being used at the trigger level and calibrated p_{T} values used at the multijets selection level.

Figure 5.1 shows the M_{T} distribution for the electron channel before and after the multijets fit, where the fit is applied across the whole visible range of $0 \rightarrow 200$ GeV. Other fit ranges are considered to estimate an uncertainty on the multijets scale factor. The improvement in the data to MC agreement is observed in the ratio plot of Figure 5.1(b).

To quantify an uncertainty on the multijets template, all detector level uncertainties are varied by one standard deviation in all MC samples and the fit re-done. Table 5.3 shows the multijets scale factors for the uncertainty variations and the corresponding QCD $W + 2$ jets scale factors. The QCD $W + 2$ jets normalisation uncertainty is evaluated in the signal



(a) Before the fit is applied.

(b) After the fit.

Figure 5.1: The fit in W M_{T} for the multijets normalisation factor in the electron channel.

extraction so are not propagated from this table, but the scale factors serve as a cross-check of the fit as the results are consistent with those found in the signal extraction.

The M_{T} fit range is altered to $30 \rightarrow 150$, $20 \rightarrow 100$ and $0 \rightarrow 80$ GeV and the largest resulting change is used in the uncertainty estimate. The fit variable is also changed to $E_{\mathrm{T}}^{\mathrm{miss}}$ from M_{T} and fit in the range $0 \rightarrow 300$ GeV. The change in fit variable is a more substantial test of the kinematic modelling. The largest change from the range and variable fits is added is quadrature with the detector systematic uncertainties to give the final multijets uncertainty. Table 5.4 shows the combination of all uncertainties for the electron channel multijets template.

The effect of doubling and halving the signal contribution is 8% on the multijets scale factor. This effect is small compared to the leading error and demonstrates that the signal contribution does not affect the fit for the multijets scale factor.

5.2.2 Muon Channel

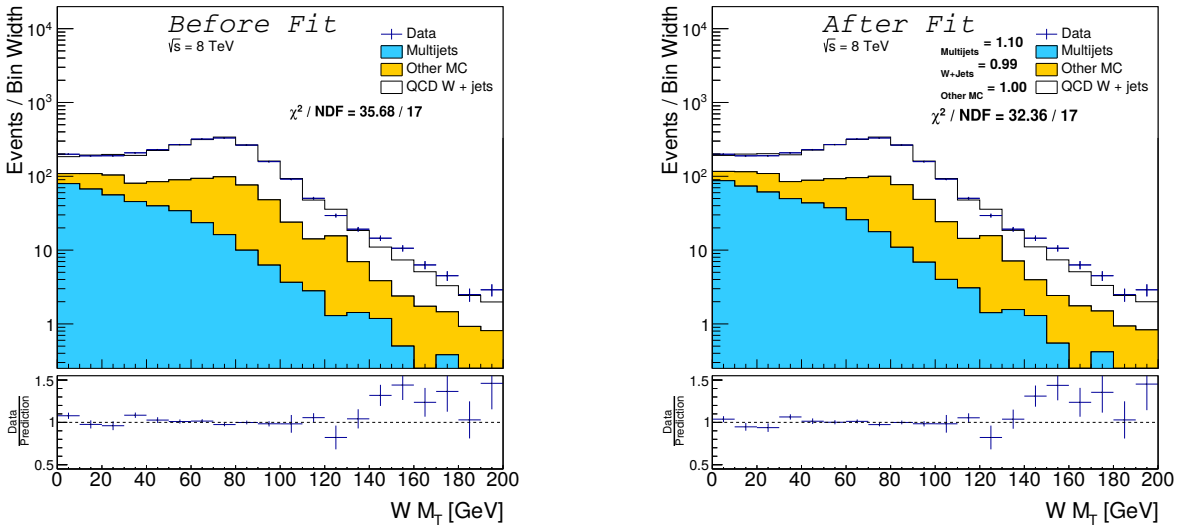
In the same manner as the electron channel above, the muon multijets template was formed by creating a multijets selection orthogonal to the measurement event selection. The muon

Systematic	Multijets Scale Factor	QCD W + 2 Jets Scale Factor
JES up	1.46	0.99
JES down	1.53	1.01
JER	1.47	1.00
JES EtaIntercalibration Modelling up	1.49	0.96
JES EtaIntercalibration Modelling down	1.52	1.02
JES EtaIntercalibration StatAndMethod up	1.48	0.99
JES EtaIntercalibration StatAndMethod down	1.49	1.00
JES MuOffset up	1.49	1.00
JES MuOffset down	1.48	1.00
JES NPVOffsetTerm up	1.49	1.00
JES NPVOffsetTerm down	1.48	1.00
JES PileupPtTerm up	1.48	1.00
JES PileupPtTerm down	1.48	1.00
JES PileupRhoTopology up	1.48	0.99
JES PileupRhoTopology down	1.49	1.01
MultiJES FlavorComp up	1.48	0.97
MultiJES FlavorComp down	1.52	1.02
MultiJES FlavorResponse up	1.52	1.02
MultiJES FlavorResponse down	1.47	0.97
BTag up	1.50	1.07
BTag down	1.47	0.95
JVF up	1.48	1.00
JVF down	1.48	1.00
MET ResoSoftTerms up	1.48	1.00
MET ResoSoftTerms down	1.49	1.00
MET ScaleSoftTerms up	1.48	1.00
MET ScaleSoftTerms down	1.49	1.00
Ele EnergySmearing up	1.48	1.00
Ele EnergySmearing down	1.49	1.00
Ele EnergyScaling up	1.49	1.00
Ele EnergyScaling down	1.49	1.00
Ele TriggerSF up	1.48	0.99
Ele TriggerSF down	1.48	1.00
Ele tightSF up	1.48	0.99
Ele tightSF down	1.49	1.01
Ele recoSF up	1.48	1.00
Ele recoSF down	1.48	1.00
PDF CT10 NLO up	1.47	0.97
PDF CT10 NLO down	1.50	1.03
Nominal	1.48	1.00
Quadrature Error Sum (%)	6.12	11.95

Table 5.3: Detector uncertainty effects on the electron channel multijets scale factors. The QCD $W + 2$ jets scale factors are also shown. The uncertainties are defined in Chapter 6.

Source	% variation	% uncertainty
Detector Systematics	6.12	6.12
Fit Range	16.9	33.8
Fit Variable	33.8	
Total	-	34.3

Table 5.4: Uncertainty sources and combination for the electron channel multijets template.



(a) Before the fit is applied.

(b) After the fit.

Figure 5.2: The fit in W M_{T} for the multijets normalisation factor in the muon channel.

and electron templates differ since the way in which a jet can fake each lepton differs. The muon multijets selection is shown in Table 5.5.

Electron Cut	Nominal Selection	Multijets Selection
Impact parameter significance $(d_0/\sigma(d_0))$	< 3	-
Track Isolation	< 0.15	$0.15 <$ isolation < 0.35

Table 5.5: Muon multijets template definition compared to the nominal selection.

The removal of the impact parameter requirement allows the selection to capture heavy-flavour hadrons displaced from the primary vertex before decaying to a muon. The orthogonal and looser track isolation cut allows the selection to find jets with muons inside.

Fits are performed using the same procedure as for the electron channel with the same M_{T} $0 \rightarrow 200$ GeV default fit range. Figures 5.2 show the muon M_{T} distribution before and after the multijets fit.

The detector uncertainties are evaluated as in the electron channel, and shown in Table 5.6, and added in quadrature with the largest variation from the fit range and fit variable tests. The range and variable tests are identical to those from the electron channel. The results of the muon channel multijets uncertainties are shown in Table 5.7.

Doubling and halving the signal contribution has an 11% effect on the multijets scale factor.

5.2.3 Multijets summary

The multijets scale factors for the analysis and their uncertainties are shown in Table 5.8. The uncertainties are symmeterised so they both increase and decrease the multijets contribution.

Systematic	Multijets Scale Factor	QCD W + 2 Jets Scale Factor
JES up	1.07	0.97
JES down	1.09	1.02
JER	1.12	0.99
JES EtaIntercalibration Modelling up	1.05	0.97
JES EtaIntercalibration Modelling down	1.10	1.02
JES EtaIntercalibration StatAndMethod up	1.09	0.99
JES EtaIntercalibration StatAndMethod down	1.09	1.00
JES MuOffset up	1.09	1.00
JES MuOffset down	1.09	0.99
JES NPVOffsetTerm up	1.08	1.00
JES NPVOffsetTerm down	1.09	0.99
JES PileupPtTerm up	1.09	0.99
JES PileupPtTerm down	1.09	1.00
JES PileupRhoTopology up	1.09	0.99
JES PileupRhoTopology down	1.08	1.01
MultiJES FlavorComp up	1.07	0.97
MultiJES FlavorComp down	1.09	1.02
MultiJES FlavorResponse up	1.09	1.02
MultiJES FlavorResponse down	1.07	0.97
BTag up	1.14	1.05
BTag down	1.10	0.94
JVF up	1.09	0.99
JVF down	1.10	1.00
MET ResoSoftTerms up	1.09	1.00
MET ResoSoftTerms down	1.10	0.99
MET ScaleSoftTerms up	1.09	1.00
MET ScaleSoftTerms down	1.10	0.99
Muon Trigger up	1.09	0.97
Muon Trigger down	1.10	1.02
Muon Efficiency up	1.09	0.99
Muon Efficiency down	1.10	1.00
Muon Momentum Scale up	1.10	0.99
Muon Momentum Scale down	1.09	1.00
Muon Momentum ID up	1.10	0.99
Muon Momentum ID down	1.09	0.99
Muon Momentum MS up	1.09	1.00
Muon Momentum MS down	1.10	0.99
PDF CT10 NLO up	1.08	0.96
PDF CT10 NLO down	1.11	1.03
Nominal	1.10	0.99
Quadrature Error Sum (%)	8.33	12.11

Table 5.6: MC uncertainty effects on muon channel multijets scale factors. The QCD $W+$ 2 jets scale factors are also shown.

Source	% variation	% uncertainty
Detector Systematics	7.0	7.0
Fit Range	22.7	47.3
Fit Variable	47.3	
Total	-	47.8

Table 5.7: Uncertainty sources and combination for the muon channel multijets template.

Channel	Nominal multijets scale factor	Uncertainty (%)
Electron	1.48	34.3
Muon	1.10	47.3

Table 5.8: Multijets scale factors and uncertainties for both channels.

Chapter 6

Measurement Uncertainties

Any scientific measurement must carry an associated assessment of the precision of the result. All experimental measurements have uncertainties which arise from various sources. Every measurement will have a statistical error, representing the random Poisson nature of statistics, which can be reduced by increasing the number of events. Systematic errors arise from two sources in experimental particle physics: first from uncertainties related to detector measurements and second, from uncertainties on MC theoretical predictions. An overview of statistical errors is given in Section 6.1, the detector systematic uncertainties are discussed in Section 6.2, and the theoretical uncertainties are discussed in Section 6.3. The interference between the signal and backgrounds processes is also considered as an uncertainty and is discussed in Section 6.4.

The uncertainties affect both the shape and normalisation of the M_{jj} distribution prediction (used in the signal extraction fits) from each MC / data-driven sample. The uncertainties are all evaluated by repeating the analysis with a single uncertainty varied (e.g. jet energy scale increased by 1 standard deviation). This repeated analysis produces a predicted M_{jj} distribution which is altered from the nominal predictions. The effect of these variations are evealuted in Chapter 9. The motivation for each source of uncertainty is explained in this chapter.

6.1 Statistical Errors

A hypothetical, ideal experiment would have an infinite number of events with which to perform a measurement. The finite yields in any real experiment mean the measured data are only a sub-sample of a hypothetical infinite event data-set. Hence, as we only have a random slice of this "complete" data-set, the mean of any distribution in the sub-sample will only approximate the mean of the full sample. The statistical uncertainty is an error that represents the expected difference between the measured value in the finite sample and that in the total sample.

Particle physics experiments use Poisson statistical errors since they represent the probability distributions of discrete experimental outcomes. The use of the Poisson distribution produces an uncertainty relating to the difference between and expected and observed yield. Detailed discussions are available in [104, 105]. In the limit of large statistics the uncertainties are:

Data $\sqrt{N}$ when N events are measured.

MC $\sqrt{\sum_N w_N^2}$ for N MC events each with weight w_N. MC events can be weighted to increase statistics in a phase space of interest while maintaining the correct overall shape of distributions. Note this formula returns the $\sqrt{N}$ error if events are "unweighted" ($w_N = 1$).

6.2 Experimental Uncertainties

The modelling of the detector will carry uncertainties due to resolution effects and efficiencies. These uncertainties are discussed below for electrons in Section 6.2.1, muons in Section 6.2.2, jets in Section 6.2.3 and $E_{\mathrm{T}}^{\mathrm{miss}}$ in Section 6.2.4. The uncertainty on the integrated luminosity is discussed in Section 6.2.5.

6.2.1 Electron Uncertainties

The electron calibration has two types of associated uncertainty: the energy calibration error and the efficiency scale-factors error, see Section 4.2. These uncertainties and the prescription used to propagate them to the analysis are taken from the central ATLAS electron recommendations [106]. The analysis is repeated with each correction factor altered by $\pm 1\sigma$ to quantify the effect on the signal extraction. The electron systematics are found to have $\sim 1\%$ effect.

The energy calibration has two uncertainties arising from the energy scale and resolution. The method used to derive these calibrations has uncertainties from: event statistics, generator modelling of samples, and the amount of material the electron has traversed. These variations are carried through to the electrons resulting in the uncertainties `Ele_EnergyScaling` and `Ele_EnergySmearing`.

The electron efficiency factors are found from the tag-and-probe method. The efficiency factors that account for the electron trigger, identification and reconstruction each have an associated uncertainty, respectively denoted: `Ele_TriggerSF`, `Ele_tightSF` and `Ele_recoSF`.

6.2.2 Muon Uncertainties

Uncertainties on the muon calibrations (see Section 4.3) are derived in the same manner as for the electrons above. Once again, the uncertainties come from the central working group's recommendations [107] and are individually propagated to the signal extraction by repeating the analysis with the correction factors altered by $\pm 1\sigma$.

The calibration of the muon momentum is driven by separate measurements from the ID and the MS. Each of these measurements has an associated uncertainty from the momentum resolution, denoted: `Muon_Momentum_ID` and `Muon_Momentum_MS` respectively, where the 1σ error is found by the muon momentum resolution width observed in data. The method used to combine the ID and MS measurements to the analysis momentum scale also has an error, `Muon_Momentum_Scale`, where the 1σ error is found from a tag-and-probe method.

The efficiency scale factors that are applied to the muon are also altered by $\pm 1\sigma$. The trigger efficiency is denoted `Muon_Trigger` and the identification and reconstruction efficiency errors are combined by the working group and denoted `Muon_Efficiency`.

The total effect of the muon systematics on the signal extraction is $\sim 2\%$.

6.2.3 Jet Uncertainties

Detailed analyses of the detector uncertainties on the ATLAS jets are given in [108, 109]. These uncertainties can be split into three types: uncertainties on the jet energy caused by the resolution of the detector; errors on the jet energy scale propagated from the methods used to derive the scale correction, and finally efficiency uncertainties caused by the JVF cut. The overall contribution to the signal extraction uncertainty from the jet errors is $\sim 15\%$, although the constraints applied in Chapter 8 reduce this to $\sim 9\%$.

Jet Energy Resolution

Jet energy resolution (`JER`) represents the random event to event error that a jet of a certain truth p_{T} will be reconstructed with a different p_{T} given that all other calibrations are correct. Data dijet events can be used to find the p_{T} balance $({p_{\mathrm{T}}}^{jet}/{p_{\mathrm{T}}}^{ref})$ where one jet is tagged as the reference jet (ref) and the other jet's p_{T} is probed. The width of the p_{T} balance distribution is related to the jet energy resolution and the uncertainty is determined by the precision to which the MC models this resolution.

Jet Energy Scale

There are multiple contributions to the jet energy scale uncertainty. The largest ones for this analysis have been separated out and the remaining errors have been added in quadrature and classified as a single uncertainty.

There are two errors associated with the extrapolation of the jet energy scale meas-

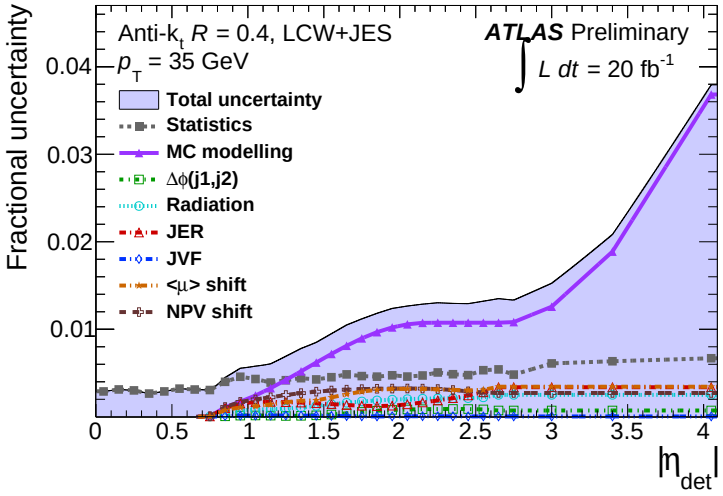


Figure 6.1: Summary of uncertainties on the inter-calibration as a function of the jet $|\eta|$. The largest component in the high $|\eta|$ region is the uncertainty from differences between Pythia8 and Herwig++ modelling. Taken from [72].

urement, from $|\eta| < 0.8$ to the forward region $|\eta| > 0.8$. These two systematics are named `JES_EtaIntercalibration_Modelling` and `JES_EtaIntercalibration_StatAndMethod`. The two contributions are from the MC model used to determine the uncertainty and from the statistics and the method respectively. The contributing uncertainties to the total `JES_EtaIntercalibration` uncertainty, for jets with $p_{\mathrm{T}} = 35$ GeV, are shown in Figure 6.1.

The pile up correction applied to the jet energy scale has four associated uncertainties. These arise from uncertainties on the NPV and $\langle\mu\rangle$ values and from the error on the jet area and p_{T}. These uncertainties shift the p_{T} and E consistently up or down for all jets. These are referred to as `JES_NPVOffsetTerm`, `JES_MuOffset`, `JES_PileupPtTerm` and `JES_PileupRhoTopology`.

The detector response to gluon and quark jets differ. This can be verified by investigating MC jets and extrapolating back to the original parton. The extra colour charge of the gluon means gluon induced jets tend to be wider with higher particle multiplicities than quark jets [110]. As the jet energy scale calibration is produced centrally by the calibration group in ATLAS, it must be applicable for all analyses. Therefore, no prior knowledge is assumed about the relative contributions of quark and gluon jets in an analysis and the default

assumption of 1:1 is used. The `JES_FlavourComp` and `JES_FlavourResponse` represent the uncertainties in altering this quark to gluon jet composition to 100% and 0% gluon induced jets and the associated detector response.

The remaining 50 jet energy scale uncertainties arise from the in-situ calibrations derived from $Z+$ jet balance, $\gamma+$ jet balance, and multi-jet balance events. They are expected to have minimal effect on this analysis and have been summed in quadrature and symmeterised, rather than simply summed, to avoid accidental cancellations resulting in an artificially small uncertainty. These are simply referred to as the `JES` uncertainty in this analysis.

Jet Vertex Fraction Uncertainty

The impact of rejecting jets with the JVF cut (see Section 4.4.4) is to reduce the number of selected jets from pile-up interactions. The value of the JVF cut is chosen to reject the highest number of pile-up jets whilst still ensuring a high fraction of the hard-scatter jets are passed to the analysis. The efficiency of the cut is varied by the centrally provided ATLAS JVF tool [111] by $\pm 1\sigma$. The variations are derived from an in-situ method using Z + jets events. The uncertainty is denoted `JVF` and alters the final signal extraction by $< 1\%$.

b-tagging Uncertainty

Events are rejected if they are tagged as containing a b-jet. The threshold for a jet to be tagged comes from a multi-variate combination of variables describing the jet, see Section 4.4.5. The efficiency of the tagger is varied both up and down by 1σ and the analysis repeated each time. The 1σ variations are derived in [78] using a likelihood fit to $t\bar{t}$ events. The `BTag` uncertainties have $\sim$ a 3% effect on the signal extraction.

6.2.4 Missing Transverse Energy Uncertainties

The analysis uses the event $E_{\mathrm{T}}^{\mathrm{miss}}$ in both the selection cut and in the multijets normalisation fits. The $E_{\mathrm{T}}^{\mathrm{miss}}$ in an event is reconstructed using the negative sum of the observed analysis

objects, see Section 4.5. By altering the analysis objects by their uncertainties (e.g. jets with energy increased by the uncertainty on the jete energy scale), the effects are propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation accounting for the correlation between the two effects. The $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties are therefore taken care of in the other analysis uncertainties. The exception to the above is for analysis objects that are not accounted for outside the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. An additional Soft term covers all objects not included as specific analysis objects. The Soft term's momentum scale and resolution have uncertainties calculated from $Z \to l\bar{l}$ events, where no real $E_{\mathrm{T}}^{\mathrm{miss}}$ is present, and $W \to l\nu$ events. The `MET_ScaleSoftTerms` and `MET_ResoSoftTerms` have a $< 1\%$ effect on the signal extraction.

6.2.5 Luminosity Uncertainty

The integrated luminosity is used to normalise the MC predictions to the amount of collected data, see Section 3.1. The integrated luminosity is calculated from the LUCID experiment [51] as well as "van der Meer" scans in ATLAS where the beams are progressively scanned across one another along each axis in the transverse planes. This allows a relationship to be calculated between the number of proton-proton interactions and the delivered luminosity. The total integrated luminosity is then determined by summing over all proton collisions. The uncertainty on this total integrated luminosity value is documented in [112] and is 2.8%.

6.3 Theoretical Uncertainties

This section documents the uncertainties that originate from the theoretical predictions used in the analysis. These are mostly related to approximations in the mathematical methods used to produce the MC sample predictions.

The full set of theory errors described below are only explicitly considered for the EW and QCD $W + 2$ jets processes. The theoretical errors on other processes are combined into cross section errors as discussed in Section 5.1. These errors are denoted by the process

name followed by an "XS" to denoted the cross section, for example QCD ZJets XS.

It should be noted that the renormalisation and factorisation scales and parton shower uncertainties are not strictly 1σ errors but rather are only the adopted procedure for quantifying these errors.

6.3.1 Renormalisation Scale

The scale at which effective particle charges or masses are calculated is known as the renormalisation scale μ_R, see Section 2.2.2. In MC calculations μ_R is set to the typical scale of the process being considered. For example, in the process: $e^-e^+ \to Z/\gamma^* \to e^-e^+$ a good scale choice is the Z-boson mass as this is the scale at which the couplings are probed. The value of μ_R is not fixed by the physics of the process but is instead an artefact of the calculation prescription. The standard prescription for assessing the error related to the chosen value of μ_R is to re-do the analysis a scale of $2\mu_R$ and then again with $\mu_R/2$. This variation gives a rough idea of the size of the missing contributions from the omitted orders in perturbation theory.

The calculation of a cross section for a particular process depends on μ_R if the calculation is done to any finite order. The variation of μ_R in one calculation is uncorrelated to another process. Therefore, normalisation scale uncertainties are considered independently for different processes.

Only the EW and QCD $W + 2$ jets renormalisation uncertainties in the M_{jj} distribution are explicitly considered in this analysis. They are denoted Signal μ_R and WJets μ_R respectively.

6.3.2 Factorisation Scale

The factorisation scale, μ_F, is the scale at which the hard matrix element process is factorised from the evolution of the parton distributions. As with μ_R, if we were able to calculate a

cross section to all orders in perturbation theory then the result would be independent of the scale choice. The usual approach again is to double and halve the chosen value of μ_F and repeat the analysis to observe the effect on the signal measurement.

The factorisation uncertainty errors are denoted `Signal` μ_F for the signal process and `WJets` μ_F for the QCD $W + 2$ jets process. These are again treated as uncorrelated with one another as well as uncorrelated with the renormalisation errors. The μ_F errors give a combined error of $\sim 7\%$ on the signal extraction*.

6.3.3 Parton Shower

The parton shower modelling is subject to multiple approximations or nonphysical procedures in the calculation. For example, the order in which ISR is produced can be via either p_{T} (highest first) or by $|\eta|$ (smallest first). The non-perturbative models used for the underlying event and hadronisation are based on experimental observation and can vary between generators. As it is very difficult to quantify the 1σ errors for all shower choices, an alternative prescription is to simply shower the same set of events with a different shower code. The nominal signal and QCD $W +$ jets background are showered with Pythia8 [113] so the parton shower uncertainties are found by re-doing the analysis using the same PowHeg events showered with Herwig++ [114]. The main difference between Pythia8 and Herwig++ is the order in which the partons are showered; Pythia8 showers the highest p_{T} parton first and Herwig++ showers the highest $|\eta|$ parton first.

The parton shower uncertainties are denoted `Signal PartonShower` for the EW signal process and `WJets PartonShower` for the QCD $W + 2$ jets process. The parton shower changes are treated as uncorrelated between the EW signal process and QCD $W + 2$ jets process. The signal extraction is changed by $\sim 9\%$ with these parton shower variations.

*Although this effect is reduced by constraining the QCD $W + 2$ jets distribution.

6.3.4 PDF Uncertainties

The particles that initiated the hard-scatter are drawn from the PDF. As the exact parton content of the proton cannot be derived from first principles, it is constructed from experimental measurements. There are uncertainties on the momentum distribution of the partons expressed as a fraction x of the proton's momentum. The PDF groups supply PDF functions with procedures for calculating the uncertainties on the contribution of each parton at each x value. This analysis determines uncertainties based on the PDF4LHC recommendations [115] using the LHAPDF prescription [116].

As the PDFs relate to real observables, like the fraction of *up* quarks with a certain momentum x, any change must be considered across all MC samples. The PDF changes are therefore completely correlated, and investigated simultaneously, across the EW signal and QCD $W + 2$ jets background samples. The other background samples have a small contribution so the PDF uncertainty for them is folded into their cross section uncertainty.

The CT10 collaboration provide 52 error PDF sets consisting of 26 up and down 1σ variations in the parameterisation. The analysis is repeated using each of these PDF sets and the variations in the signal extraction combined using the "Hessian combination". The extracted signal is found for each of the 52 error sets and combined as shown in Equations 6.1 and 6.2 to give the up and down errors respectively. We have used the notation S_i for the extracted signal for using the PDF error set i; S_0 denotes the nominal result.

$$\Delta S_0^{\mathrm{PDF}+} = \frac{1}{C_{90}} \sqrt{\sum_{i=1}^{26} \left[\max\left(S_{2i-1} - S_0, S_{2i} - S_0, 0\right)\right]^2}, \tag{6.1}$$

$$\Delta S_0^{\mathrm{PDF}-} = \frac{1}{C_{90}} \sqrt{\sum_{i=1}^{26} \left[\max\left(S_0 - S_{2i-1}, S_0 - S_{2i}, 0\right)\right]^2}, \tag{6.2}$$

where $C_{90} = 1.64485$ and accounts for the error sets being provided at the 90% confidence level when we wish to use 1σ errors (68% confidence level). The PDF error sets also provide

an error associated with the value of α_S. The analysis is repeated twice more using the nominal CT10 PDF only with the value of α_S set to 0.117 and then to 0.119, where the nominal value is 0.118[†]. The world-average measured value of α_S has a 1σ uncertainty of ± 0.0012 [9] but the PDF sets are only provided at discrete α_S values. The signal extraction results from the α_S PDF variations are then modified by a factor of $C_{59} = 0.82389$ to get the 1σ error as shown in Equation 6.3.

$$\Delta S_0^{\alpha_S^{\pm}} = \frac{1}{C_{59}} \left(S_{\alpha_S^{\pm}} - S_0 \right) \tag{6.3}$$

The error set results, $\Delta S_0^{\mathrm{PDF}\pm}$, are then summed in quadrature with the α_S results, $\Delta S_0^{\alpha_S^{\pm}}$, to give the total PDF uncertainty at the 1σ level.

6.4 Interference

The picture we have used thus far is that the hard-scatter can occur through a Feynman diagram to produce the particles that form the final state. However, these diagrams represent probability amplitudes and not individual processes. The QCD and EW production of $W+$ 2 jets have the identical external particles. Thus, the two processes must be combined at the amplitude level. One must square the amplitude to obtain the probability so these diagrams can interfere. This can result in the total cross section for a process being significantly different than the sum of the cross section of the individual diagrams. Mathematically the amplitude is expressed as the matrix element M_{fi} to go from the initial state i to the final state f:

$$M_{fi} = M_1 + M_2 + M_3 + \dots \,, \tag{6.4}$$

where M_n represents the matrix element for diagram n. The cross section is proportional to

[†]All α_S values are given at the scale of the Z-boson mass.

$|M_{fi}|^2$ so:

$$\sigma_{fi} \propto |M_1|^2 + |M_2|^2 + |M_3|^2 + M_1^* M_2 + M_1^* M_3 + ... \tag{6.5}$$

In certain processes the interference terms ($M_1^* M_2$ etc.) may be large enough to alter σ_{fi} significantly from $\sum_n |M_n|^2$. A particular M_n term can dominate in a particular region of phase space, so the interference contribution can depend on the phase space under consideration. Sizeable interference requires that at least two M_n terms also have a sizeable contribution in that region of phase space.

QCD - EW Interference

In this analysis the largest concern about the interference is that between the QCD and EW production of $W + 2$ jets. If the interference is large, we cannot separate the processes and must consider them together. If the interference is small, compared to each individual squared matrix element, we can use MC templates separately for each production mechanism and consider the interference as an uncertainty on the result.

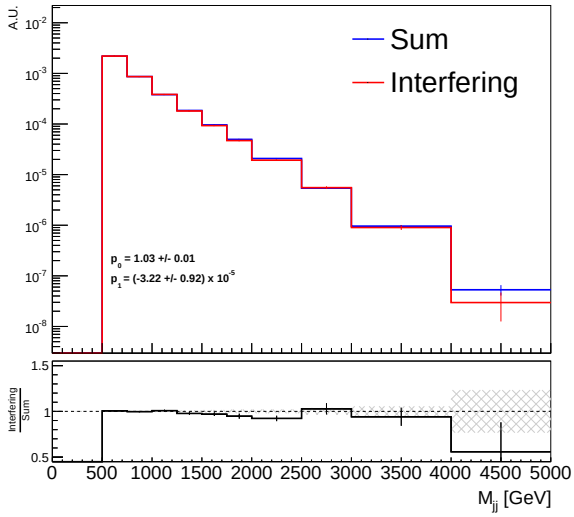
To investigate the size of the interference terms, three electron channel $W + 2$ jets samples were simulated with the MADGRAPH generator [117, 118].

Interfering All QCD and EW diagrams are considered at once allowing the generator to include the interference.

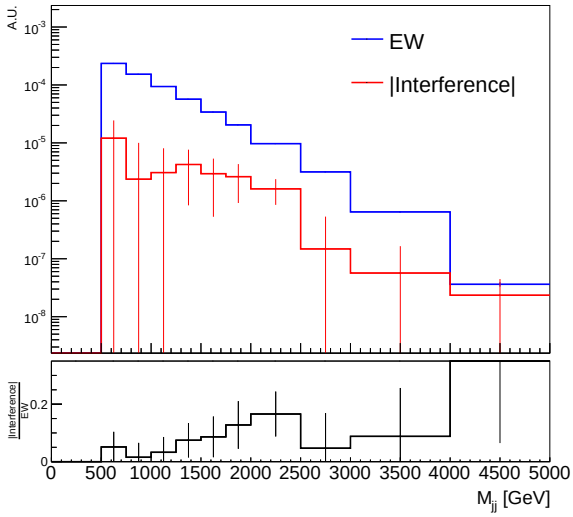
QCD Only QCD production diagrams are considered. This sample had exactly 2 EW vertices, omitting the signal contribution.

EW Only EW production (signal) and diboson diagrams are considered. This sample had no QCD vertices, omitting the QCD $W + 2$ jets background.

Hence, the diboson diagrams are included in the EW sample, which differs from the signal definition in the analysis, however the diboson fraction is expected to be 0.5%. Due to



(a) The interfering sample compared to the sum of QCD and EW. The size of the interference is observed in the ratio plot. The p_0 and p_1 values show the results of a straight line fit to the ratio.



(b) The absolute value of the interference compared to the EW $W+$ jets.

Figure 6.2: The MADGRAPH interference studies showing the difference between the interfering sample and the sum of EW and QCD (left) and the absolute value of the interference compared to the EW (right). The destructive interference can be seen on the left plot.

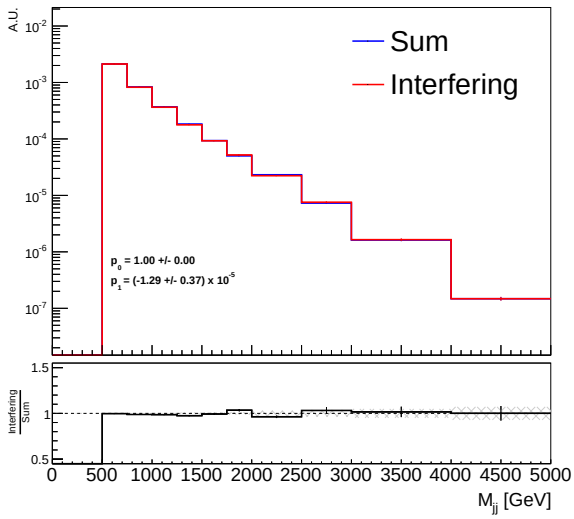
practical constraints all samples were generated at LO. At this point, it is worth defining some naming conventions:

$$\text{Sum} = \text{QCD} + \text{EW}, \tag{6.6}$$

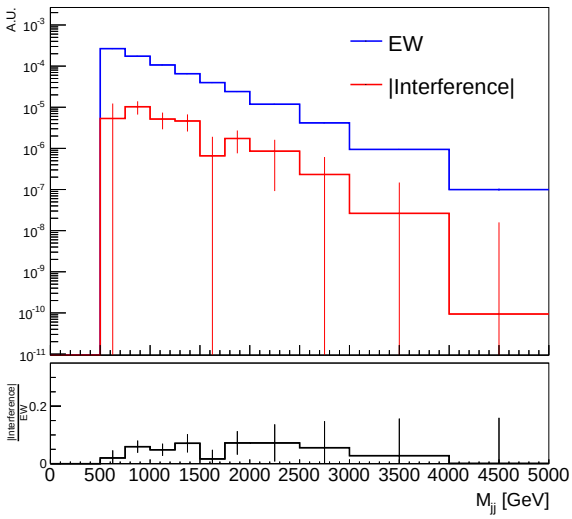
$$\text{Interference} = \text{Interfering} - \text{Sum}. \tag{6.7}$$

If there were no interference at all, the sum of the QCD and EW samples would match exactly the interfering sample. The difference between the QCD and EW sum and the interfering sample gives the size of the interference term. Considering a histogram of the samples in M_{jj} allows us to differentially view the interference in this variable. The interference histograms are shown differentially in M_{jj} in Figure 6.2.

The prevalence for the interfering and QCD samples to have low M_{jj} events mean the limited statistics of these samples cause fluctuations above ~ 2 TeV in Figure 6.2(a). Slightly



(a) The interfering sample compared to the sum of QCD and EW. The size of the interference is observed in the ratio plot. The p_0 and p_1 values show the results of a straight line fit to the ratio.



(b) The absolute value of the interference compared to the EW $W+$ jets.

Figure 6.3: The SHERPA interference studies showing the difference between the interfering sample and the sum of EW and QCD (left) and the absolute value of the interference compared to the EW (right).

below this point the sum is greater than the interfering sample showing that the interference is destructive. Figure 6.2(b) shows the absolute value of the interference (on a log plot we cannot show negative values) compared to the EW $W + 2$ jets process. Below 2 TeV the interference is $< 10\%$ of the EW process but increases as M_{jj} increases.

To cross-check the MADGRAPH results, SHERPA was used to generate the same set of samples and the studies repeated. Figure 6.3 shows the same set of plots as above for the SHERPA results where the interference is seen to be smaller, though the slopes of the interfering/sum ratios are comparable to $< 2\sigma$. Comparing these p_1 values shows that MADGRAPH has roughly twice as much interference as SHERPA. The ratio of the SHERPA p_1 value to the MADGRAPH p_1 is 2.49 ± 0.99. Figures 6.2(b) and 6.3(b) show that statistics are limited in both studies but due to practical constraints it was not possible to generate more events.

Recent VBF analyses [34, 35] also investigated this interference effect (although for VBF Z in both cases) and found the interference caused $\sim 7\%$ systematic on the EW signal

extraction. The theoretical study by Denner et al. [87] suggests that when performing these studies at NLO (in both QCD and EW) the interference becomes negligible.

In this analysis, the interference uncertainty is taken from MADGRAPH as it is the generator that shows the largest effect on the extracted EW signal.

Diboson - VBF interference

Another source of interference that affects the EW signal is that between the signal and the diboson processes. The two samples used in the signal extraction have been generated separately which assumes negligible interference between them. Physics arguments suggest the interference in this case should indeed be negligible as the dibosons M_{jj} spectrum is sharply peaked about the W and Z-boson masses, both of which is well below the M_{jj} *Search* region cut of 500 GeV. In this case the samples, generated with SHERPA, are:

Interfering All EW signal and diboson diagrams are considered at once, including the interference. This sample considers the same diagrams as the EW sample in the QCD - EW interference studies above.

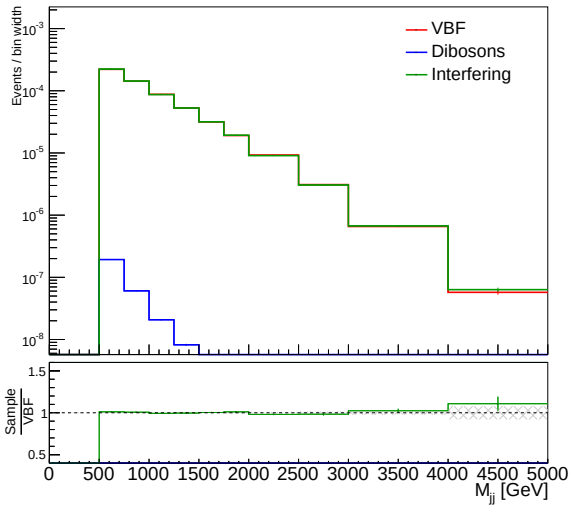
VBF Only the signal diagrams are considered[‡].

Diboson Only the diboson diagrams are considered.

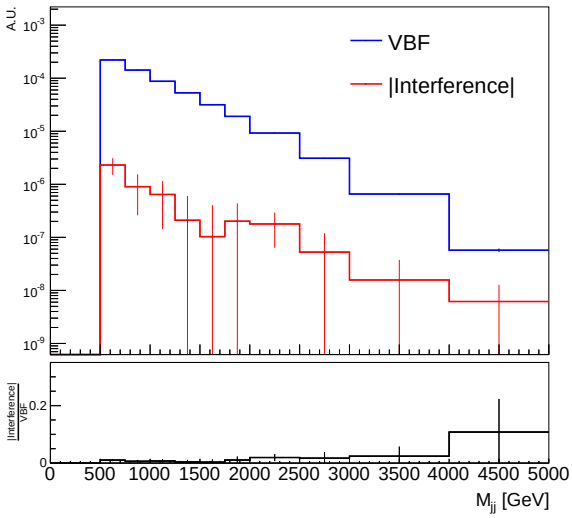
Once again the sum is defined as the VBF + Diboson components and the interference is the difference between the interfering and sum samples. Figure 6.4 shows the three components on the left where the diboson component is shown to be at most $\mathcal{O}(10^3)$ smaller than the VBF sample. The interference compared to the VBF signal component is shown in Figure 6.4(b) and is seen to be minuscule in comparison.

No systematic is assigned for the VBF - diboson interference as it is seen to be negligible.

[‡]This sample is named VBF so as to differentiate it from the EW sample in Section 6.4, which includes diboson diagrams.



(a) The three components for the VBF - diboson interference studies. The tiny size of the diboson component and flat ratio show minimal interference.



(b) The absolute value of the interference compared to the VBF sample.

Figure 6.4: The Sherpa VBF-diboson interference studies showing the difference between the three components (left) and the absolute value of the interference compared to the VBF sample (right). The interference observed on the right plot is expected to be mostly statistical jitter in the samples rather than true interference.

Chapter 7

Event Selection

This chapter describes the selection that is applied to the ATLAS events in order to pick out $W+$ jets events and maximise the relative contribution of the EW $W+$ jets component. The aim is to reduce the number of background events as much as possible whilst maintaining the highest number of signal events. The cuts are applied to kinematic variables such as jet p_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{jj}, that exhibit differences between the EW $W+2$ jets signal and the backgrounds. The selection cuts are therefore optimised to maximise the value of "signal significance", defined as:

$$\text{Significance} = \frac{\mathrm{S}}{\sqrt{\mathrm{S}+\mathrm{B}}}, \tag{7.1}$$

where S represents the number of EW $W+2$ jets signal events and B represents the number of background events.

7.1 Data Sample

The analysis uses the data sample collected in 2012 with the LHC running proton-proton collisions at a centre of mass energy of 8 TeV. The collisions were spaced 50 nanoseconds apart for the entire run. A good runs list (GRL) is applied to the data in order to mask

events collected with un-stable proton beams or when ATLAS was not in a stable, data-taking state, ensuring accurate measurements from all components. On average there were 20.7 proton-proton interactions per bunch crossing and a maximum instantaneous luminosity of 7.7 $\times 10^{33}$ cm^{-2}s^{-1} [64]. The total luminosity collected—after application of the GRL—in this data sample is 20.281 fb^{-1}.

For each lepton channel, events that pass the trigger are analysed after requiring a set of selection cuts consistent with characteristics of an EW $W + 2$ jets event. If an event is significantly more likely to result from a background process than the signal process, it is discarded from the analysis. The data events that pass this selection are then compared to the SM prediction to check the modelling and then fit for the signal contribution. We therefore require some knowledge about the EW $W + 2$ jets signal characteristics. The MC sample for the signal is compared to those of the backgrounds to exploit differences between the two. The event selection has two aims: to remove as many background events as possible to enhance the signal purity, and to retain as many signal EW $W + 2$ jets events as possible to achieve a statistically significant measurement.

7.2 Cuts

This section discusses the selection cuts that are applied in order to maximise the purity of the EW $W +$ jets events in the data.

7.2.1 *W*-boson Selection

The triggers used to select events are designed to find the charged leptons from the W-boson decay. As there are multiple ways to produce charged leptons, further cuts are applied to ensure a W-boson did in fact produce the identified lepton. A leptonic W decay produces a neutrino so we remove events with $E_{\mathrm{T}}^{\mathrm{miss}}$ smaller than 25 GeV. The charged lepton p_{T} is required to be above 25 GeV due to the trigger thresholds (see Chapter 4) so this $E_{\mathrm{T}}^{\mathrm{miss}}$ cut

brings the neutrino into the same energy range as the charged lepton. This cut also removes charged leptons from non-W backgrounds. The W-boson transverse mass*, defined as:

$$WM_T = \sqrt{2E_{T,1}E_{T,2}\left[1 - \cos\left(\Delta\phi\right)\right]}, \tag{7.2}$$

where $E_{T,1}$ is the energy in the transverse plane of the charged lepton and $E_{T,2}$ is the same for the neutrino (equal to the $E_{\mathrm{T}}^{\mathrm{miss}}$). $\Delta\phi$ is the azimuthal angle between the charged lepton and the neutrino. WM_T is required to be above 40 GeV, again removing non-W backgrounds.

To reduced the contribution from mismeasured events containing a Z-boson, events are vetoed if a second, high p_{T} (> 25 GeV) lepton is found. This veto also ensures that events which are identified in the electron channel measurement phase space are not also considered in the muon channel.

7.2.2 Number of Jets

Jets are defined in Section 4.4 with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 4.4$. The number of jets, N_{Jets}, in a QCD $W+$ jets event is a steeply falling distribution. The signal process has two jets from the hard-scatter, see Figure 2.5. Therefore, removing events with fewer than two jets can significantly increase the signal purity of the data. Figure 7.1 shows the unity normalised shapes for the EW $W + 2$ jets signal and QCD $W + 2$ jets background for the N_{Jets} distribution. The analysis rejects events with fewer than two jets.

7.2.3 Dijet Mass

The invariant mass of the two highest p_{T} jets is known as the dijet mass, M_{jj}. This is the most discriminating variable between EW $W+$ jets and the main background of QCD W + jets. Figure 7.2(a) shows a truth level comparison of the EW and QCD $W+$ jets M_{jj}. A

*We are unable to reconstruct the mass of the W-boson as we cannot infer the longitudinal momentum of the neutrino. The transverse mass is the boson mass calculated using only momentum in the transverse direction.

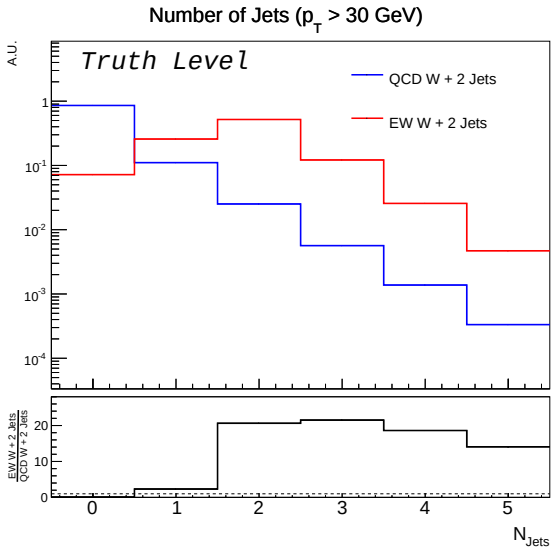


Figure 7.1: Unity normalised N_{Jets} shapes for EW and QCD $W + 2$ jets at truth level. Only jets with $p_{\mathrm{T}} > 30$ GeV are considered.

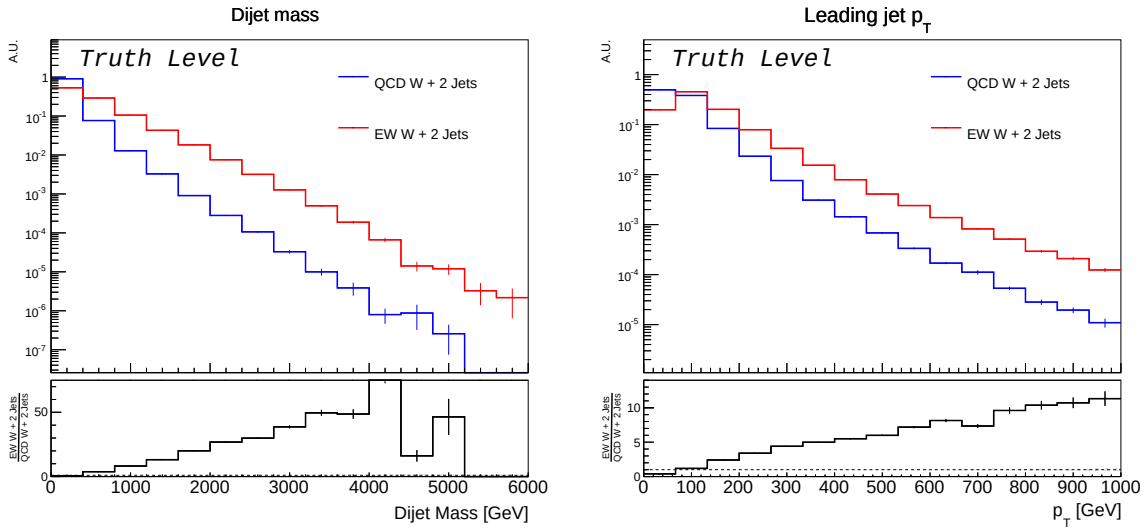
high M_{jj} cut would increase the signal purity but a compromise must be struck in order to maintain a high signal yield. An optimisation is done to maximise the significance as defined by Equation 7.1. The optimised cut for M_{jj} was found to be $M_{jj} > 500$ GeV.

7.2.4 Jet p_{T}

Since the LO signal process is quark induced and the two jets are produced from the hard-scatter rather than from QCD radiation, the signal tends to have higher p_{T} jets than the QCD $W +$ jets background. Figure 7.2(b) shows the unity normalised p_{T} distributions for the highest p_{T} jet in each event for the EW and QCD $W + 2$ jets processes. Rejecting events that have low-p_{T} jets, preferentially discards background processes, but too high a cut will remove a large numbers of signal events. Events are required to have a leading jet with $p_{\mathrm{T}} > 80$ GeV and a sub-leading jet with $p_{\mathrm{T}} > 60$ GeV.

7.2.5 Centrality

At truth level the EW $W +$ jets signal template has a strong relationship between the rapidity of the W-boson and the two tagged jets, as well as the rapidity position of any



(a) Unity normalised M_{jj} shapes for EW and QCD $W+$ jets at truth level.

(b) Unity normalised leading jet p_{T} shapes for EW and QCD $W+$ jets at truth level.

Figure 7.2: Discriminating variables between the EW $W+2$ jets signal and the QCD $W+2$ jets background.

further hadronic activity relative to the tagged jets. The EW signal process in induced by quarks at tree level so the high p_{T} events tend to be initiated by two valence quarks. It follows that the quarks have roughly similar momentum and hence, the tagged jets tend to have the high rapidity gap and the boson is produced centrally. The QCD $W+2$ jets background process can be induced by two gluons, two quarks or a gluon and a quark. This allows there to be a larger imbalance in the momenta of the incoming particles and tends to result in jets with a much smaller rapidity gap and the boson produced at higher rapidities. This relationship allows signal events to be discriminated from various background processes, including QCD $W+$ jets production.

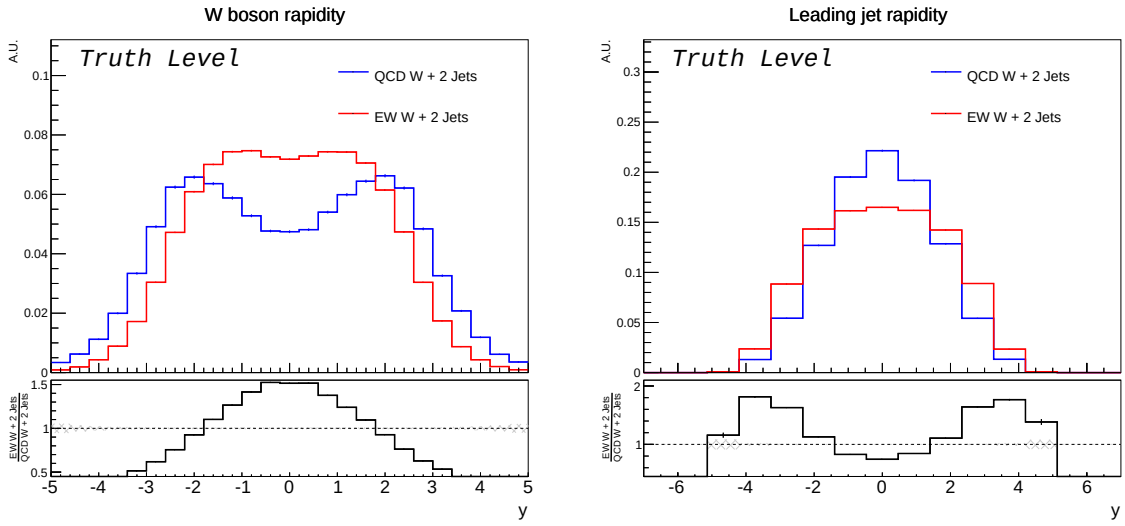
Lepton Centrality

The MC predictions of the W-boson kinematics can be probed with a view to finding discriminating variables between EW and QCD production of $W+$ jets. The W-boson in EW production is produced much more centrally than for QCD production, see Figure 7.3(a). The combination of the W-boson rapidity with the preference for EW produced jets to be more forward than the QCD produced jets, see Figure 7.3(b), leads many VBF analyses to

quantify the position of the boson in the rapidity gap formed by the tagging jets [34, 119]. This combination of variables in known as the *Zeppenfeld variable*† [120] with the structure:

$$\text{Zeppenfeld variable} = y_{boson} - \frac{y_{Jet_1} + y_{Jet_2}}{2}. \tag{7.3}$$

which quantifies the distance of the W-boson from the centre of the jet rapidity gap.



(a) The W-boson rapidity

(b) The rapidity of the leading jet

Figure 7.3: Truth level, unity-normalised rapidity distributions for EW and QCD $W + 2$ jets processes.

In this analysis we modify the Zeppenfeld variable to include information about the size of the jet rapidity gap. This change normalises the Zeppenfeld variable to the size of the jet rapidity gap such that the position within the gap is measured relative to the size of the gap. High p_{T} jets will have a smaller gap but can produce as good discrimination as events with lower p_{T} jets, where the jet gap is wider. This modified variable is known as centrality and is defined as:

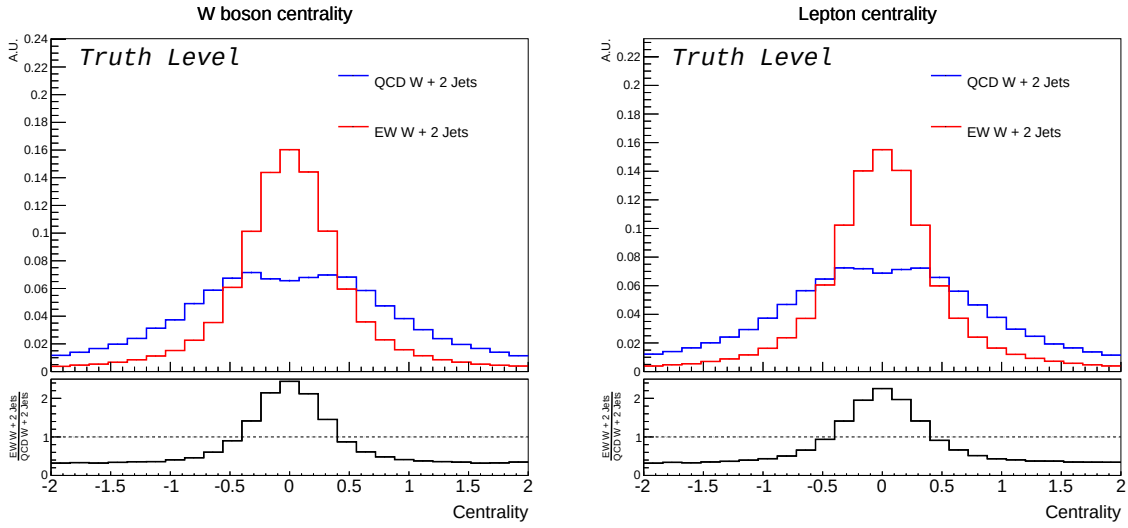
$$\mathcal{C}_{boson} = \frac{y_{boson} - \frac{y_{Jet_1} + y_{Jet_2}}{2}}{|y_{Jet_1} + y_{Jet_2}|} \tag{7.4}$$

The normalisation in centrality has the added benefit of giving the variable a simple topo-

†The Zeppenfeld variable was originally designed to investigate the position of jets in the tagged jet pseudo-rapidity gap. Instead, rapidity is used in this analysis as the boson has a non-zero mass.

logical interpretation. If $|\mathcal{C}_{\text{boson}}|$ is below 0.5, the boson is inside the jet rapidity gap; if it is above 0.5 then the boson is outside the gap.

The rapidity of the W-boson is not observable at reconstruction level due to the neutrino not being detected, so instead the centrality of the charged lepton[‡] ($\mathcal{C}_{Lepton}$) is investigated and found to have a similar shape in both EW signal and QCD background processes, to the boson centrality, see Figure 7.4.



(a) The W-boson centrality at truth level

(b) The lepton centrality at truth level

Figure 7.4: Centrality variables showing EW and QCD W + 2 jets distributions. The centrality correlations are maintained from the boson to the lepton.

Figure 7.4 shows that both the boson and the charged lepton from its decay tend to be very central in the jet rapidity gap for the EW signal. In the QCD W + jets background they are seen to be much more spread out in the gap and peak around ± 0.5, i.e. around the tagging jets. The distributions are symmetric so the absolute value centrality is used in the analysis. The optimisation studies show that rejecting events with $|\mathcal{C}_{lepton}| > 0.4$ gives the highest signal significance.

In summary, the charged lepton is produced in the centre of the tagged jet rapidity gap in signal EW W + 2 jets events and produced close to the tagging jets in the QCD W + 2 jets background.

[‡]Simply replacing y_{boson} with y_{lepton} in Equation 7.4.

Jet Centrality

It is also possible to use the centrality of any additional (non-tagged) jets in the event to discriminate between the EW signal and backgrounds with higher levels of hadronic activity. The jet centrality variable ($\mathcal{C}_{Jet}$) is designed to give a measure of a jet's position in the rapidity gap between the two tagging jets. The low levels of hadronic activity and lack of colour-flow between the jets at tree level in the EW $W + 2$ jets process means that non-tagged jets have a significant probability to result from FSR. The collinear divergence associated with gluon radiation leads FSR jets to have rapidity values very close to those of the two tagging jets which corresponds to jet centrality values peaked near $\pm$ 0.5. The backgrounds that have higher levels of QCD activity are more likely to have a third jet (or even more jets) originating from the hard-scatter rather than soft FSR. This leads to the non-tagged jets being more central in the tagged jet rapidity gap, which corresponds to jet centrality values close to zero.

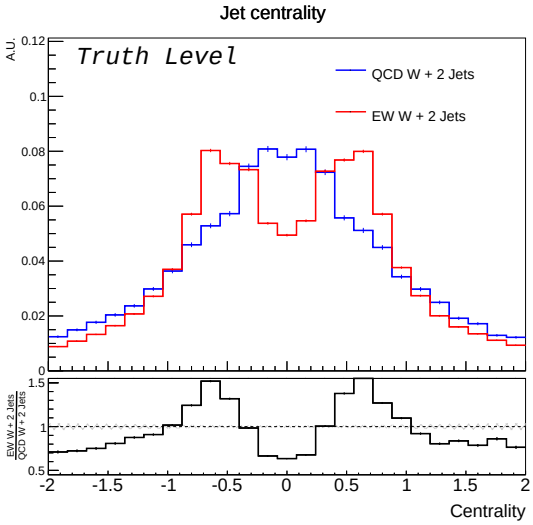


Figure 7.5: Unity normalised centrality for all non-tagging jets in EW and QCD $W +$ jets at truth level.

The truth level $\mathcal{C}_{jet}$ distributions for the EW signal compared to the QCD background are shown in Figure 7.5. The EW signal shape is formed of two peaks near the tagging jets ($\pm$0.5) and the QCD background shape is peaked in the centre of the tagged jet rapidity gap. The shape of the EW signal is modified at the peaks as the jets are of size $\Delta R = 0.4$,

so any non-tagged jet that is closer than this to a tagged jet will be "eaten" by the tagged jet and not appear in the distribution.

7.2.6 Phase Space Definitions

Using the variables outlined above we can split the events into various phase spaces to attempt to either maximise the purity of the signal or create control regions close to the measurement region but with low signal fraction. Table 7.1 summarises the phase space definitions used in this analysis.

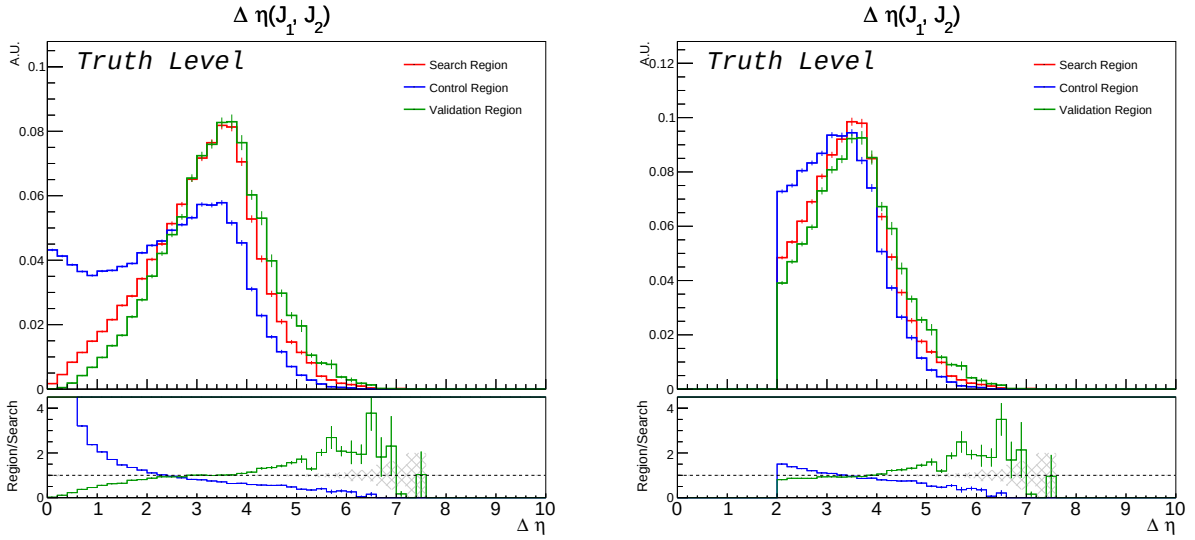
	Inclusive	*High* p_{T}	*High Mass*	*Search*	*Control*	*Validation*
N_{Jets}	≥ 2					
Jet p_{T}	> 30 GeV	> 80 GeV (Jet_1), > 60 GeV (Jet_2)				
M_{jj}	-		> 500 GeV			
$\Delta\eta_{(j1,j2)}$	-		> 2			
$\mathcal{C}_{Jet}$	-			> 0.4		< 0.4
$\mathcal{C}_{Lepton}$	-			< 0.4	> 0.4	< 0.4

Table 7.1: Definitions of the phase spaces used in this analysis. The cut values are chosen to maximise signal significance, see Equation 7.1.

7.3 Phase Space Comparisons

The *Control* and *Validation* regions are used in this analysis to correct the modelling of the QCD $W + 2$ jets background process, as discussed in Chapter 8. The modelling in one region can only be used to correct another region if they have similar kinematics. This section compared the kinematics of the *Search*, *Control* and *Validation* regions for the QCD $W + 2$ jets process.

The *Control* region differs from the *Search* region by the reversal of the $\mathcal{C}_{Lepton}$ cut. Forcing the charged lepton to be away from the centre of the jet rapidity gap affects both the lepton and the tagged jet η distributions. The *Search* region $\mathcal{C}_{Lepton}$ requirement of $|\mathcal{C}_{Lepton}| < 0.4$ forced the tagged jets to be separated and so reversing this cut allows the jets to become



(a) Without the $\Delta\eta(J_1, J_2) > 2$ cut.

(b) With the $\Delta\eta(J_1, J_2) > 2$ cut.

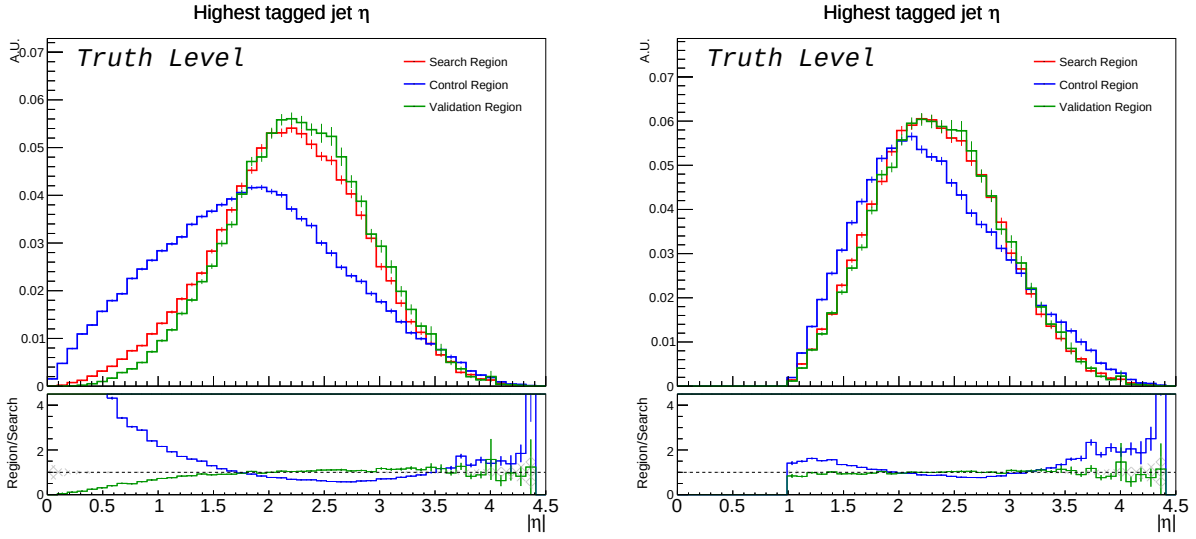
Figure 7.6: Unity normalised $\Delta\eta(J_1, J_2)$ distributions for QCD $W + 2$ jets process in the *Search*, *Control* and *Validation* regions without and with the $\Delta\eta(J_1, J_2) > 2$ cut. Whilst the distirbutions are still different after the cut, we have removed the region of largest disagreement.

close together. Figure 7.6 shows the $\Delta\eta$ between the tagged jets for the *Search*, *Control* and *Validation* regions, without and with the $\Delta\eta(J_1, J_2) > 2$ cut.

The *Control* region is shown to have a component of low $\Delta\eta(J_1, J_2)$ events that are not present in the *Search* region. This is the justification for the $\Delta\eta(J_1, J_2) > 2$ cut as shown in Table 7.1. The *Validation* region has the $\mathcal{C}_{Jet}$ cut reversed from the *Search* region as so is required to have a jet inside the tagged jet rapidity gap. This forces the $\Delta\eta(J_1, J_2)$ distribution to be wider for the *Validation* region as compared to the *Search* region . The application of the $\Delta\eta(J_1, J_2)$ cut removes the area where the three phase spaces diverge the most.

The effect of this cut on other distributions is also checked. Figure 7.7 shows the unity normalised shapes for the $|\eta|$ of the most forward tagged jet with and without the $\Delta\eta(J_1, J_2)$ cut. The *Control* region again has different kinematics to the *Search* region without the cut, but after the cut is applied the distributions become similar.

There are distributions that show increased disagreement after the cut is applied, e.g. the rapidity of the *W*-boson. The *Control* region requires the charged lepton to be forward so



(a) Without the $\Delta\eta(J_1, J_2) > 2$ cut.

(b) With the $\Delta\eta(J_1, J_2) > 2$ cut.

Figure 7.7: The unity normalised shapes of the $|\eta|$ of the most forward tagged jet in the QCD $W + 2$ jets process.

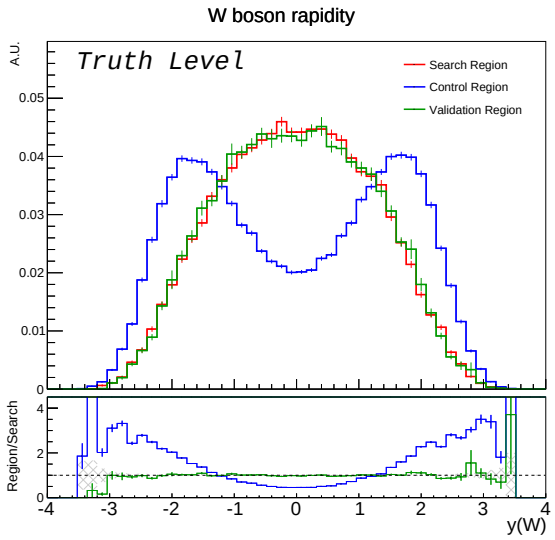
the W-boson is also seen to be peaked at high rapidity values in Figure 7.8. The application of the $\Delta\eta(J_1, J_2)$ cut forces the jets to be wider apart and thus, the W-boson becomes even more forward than before. This increased difference is acceptable as we are using the *Control* region to correct the M_{jj} modelling of the QCD $W + 2$ jets process so it is preferable that the jet distributions agree across the regions even at the cost of the lepton based distributions.

The M_{jj} distributions are shown in Figure 7.9 for the three phase spaces both with and without the $\Delta\eta(J_1, J_2)$ cut. The effect of the cut is small in this distribution but it is shown to slightly improve the *Control* and *Search* region shape agreement.

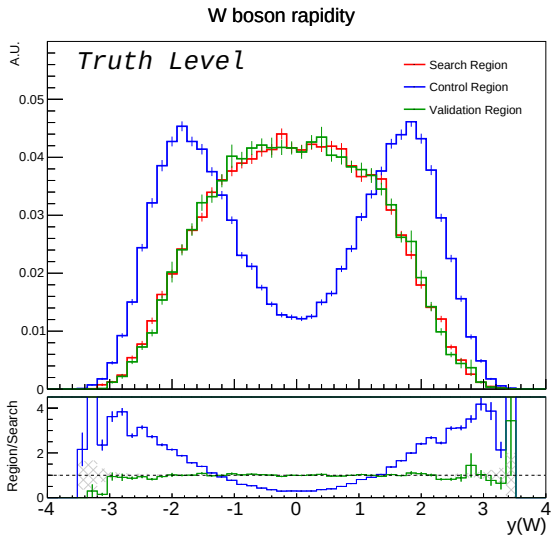
The application of the $\Delta\eta(J_1, J_2) > 2$ cut is justified by improving agreement between the *Control*, *Validation* and *Search* regions.

7.4 Pre-fit Distributions

To validate the modelling of the MC templates the shapes of kinematic distributions are compared to data. The plots shown in this section are the summed electron and muon channels in the *Search* region where all template predictions have not been modified in any

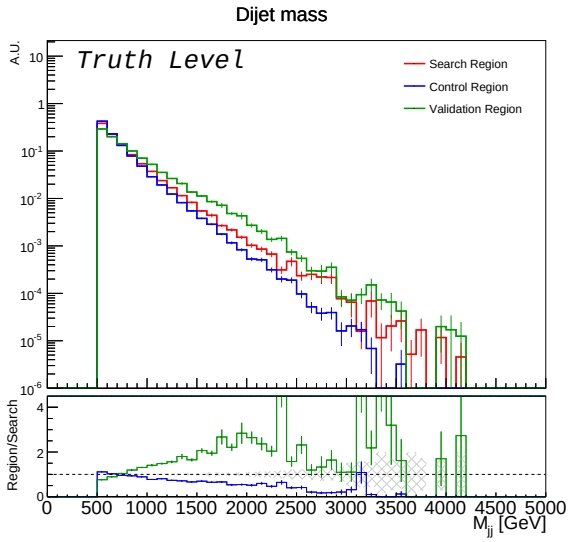


(a) Without the $\Delta\eta(J_1, J_2) > 2$ cut.

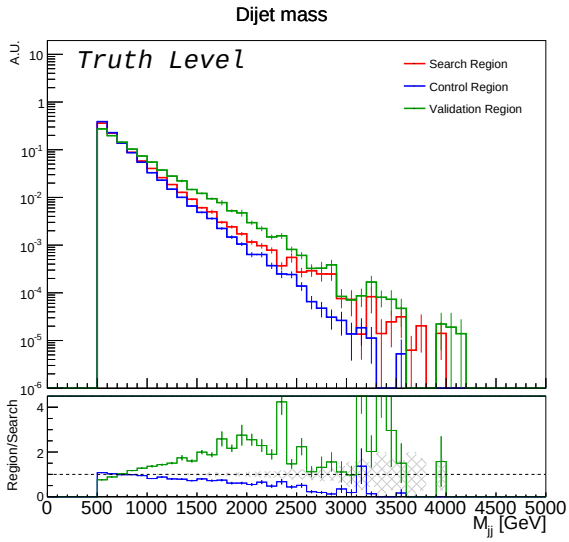


(b) With the $\Delta\eta(J_1, J_2) > 2$ cut.

Figure 7.8: The unity normalised shapes of the rapidity of the W-boson in the QCD $W + 2$ jets process. These distibutions demonstrate that the $\Delta\eta(J_1, J_2)$ cut makes some distibutions disagreement increase. This is seen as an acceptable compramise as we are now extrapolating in the leptons inly and not the jets, i.e. in the weak force and not the strong.



(a) Without the $\Delta\eta(J_1, J_2) > 2$ cut.



(b) With the $\Delta\eta(J_1, J_2) > 2$ cut.

Figure 7.9: The unity normalised shapes of the M_{jj}in the QCD $W + 2$ jets process.

way from the generator predictions.

Physics processes are grouped in the legends of the reconstruction plots as follows:

EW $W \to l\nu$ – The signal EW W + jets process, summed over both lepton channels ($e + \mu$). If a single channel is shown, the entry will be explicit (e.g. EW $W \to \mu\nu$).

QCD $W \to l\nu$ – The QCD W + jets background process. Again, $l\nu$ denotes both channels summed.

Top – The single top, Wt and $t\bar{t}$ background processes.

Other – Other background processes: EW and QCD Z + jets and dibosons.

Multijets – The high jet multiplicity background, see Section 5.2.

Please note; in the plots and tables l and ν refer to the charged leptons, e and μ, and neutrinos, ν_e and ν_μ respectively. The τ channel contributions are separated.

The detector uncertainties are shown on the following plots along with the statistical uncertainty of the prediction. These sources are uncorrelated so are added in quadrature and shown on the summed predictions on the main plots as a hatched grey band. In the ratio plots the detector uncertainties are shown as a yellow band and the combined detector and statistical uncertainty on the prediction is shown as a green band. The statistical uncertainty on the data is shown as the vertical black line about the data points. **Please note** that no theoretical uncertainties are shown on the detector-level plots.

Figure 7.12 shows reconstruction-level distributions for jet-based variables. The jet p_{T} is well modelled for the leading, sub-leading and third jets. The jet η distributions have only slight disagreement between data and prediction at $|\eta| > 3.5$, although the differences are within the systematic bands. The jet centrality distribution, shown in Figure 10.1(f) requiring the cut of > 0.4, agrees between data and prediction. An increased fraction of top

processes is expected and observed in the three jet distributions.

The lepton distributions are shown in Figure 7.13. The lepton p_{T} shows data-to-prediction agreement at low p_{T}, but the MC has a harder distribution than is observed in data at high p_{T}. This disagreement is common across both lepton channels and is thought to stem from a common mismodelling of the W-boson p_{T} as the same shape is observed in the ratio plot for $E_{\mathrm{T}}^{\mathrm{miss}}$[§]. As the data are seen to have a softer W-boson p_{T} spectrum than the MC prediction, is it possible the disagreement comes from the lack of EW corrections modelled in the prediction. A high p_{T} photon may be radiated from the charged lepton where the MC samples only model this emission well for the soft photon case. The addition of EW corrections (when the become available) may therefore improve agreement between data and prediction in lepton and W-boson p_{T} distributions. The lepton η and W-boson transverse mass distributions are well modelled. The lepton centrality distribution shows the contribution of the EW signal peaked at low values and the QCD $W + 2$ jets background peaking around 0.5, as expected from the truth studies.

Remaining important distributions are shown in Figure 7.14 including the fit variable, M_{jj} which shows a slope in the ratio plot where the prediction overestimates the data at high M_{jj}. As the high M_{jj} region is the most sensitive to the EW $W + 2$ jets signal contribution, this data-prediction disagreement suggests the EW $W + 2$ jets signal contribution will be scaled down from the SM prediction in the signal extraction. The $\Delta\eta$ (J_1, J_2) plot shows that the M_{jj} disagreement is not localised to any one $\Delta\eta$ region as there is good agreement across the whole ratio plot. The mismodelling at high M_{jj} values is a region where the EW $W + 2$ jets signal dominates, suggesting the signal extraction fit will scale down the signal contribution. Figure 7.14(c) shows the minimum ΔR between the third jet and the tagged jets, demonstrating the localisation of the EW signal at low values where the third jet is radiated from a tagging jet and a broader distributions from the backgrounds. The p_{T}

[§]A known problem in the POWHEG $W + 2$ jets samples is shown to not be responsible for this mismodelling in Appendix B.2.

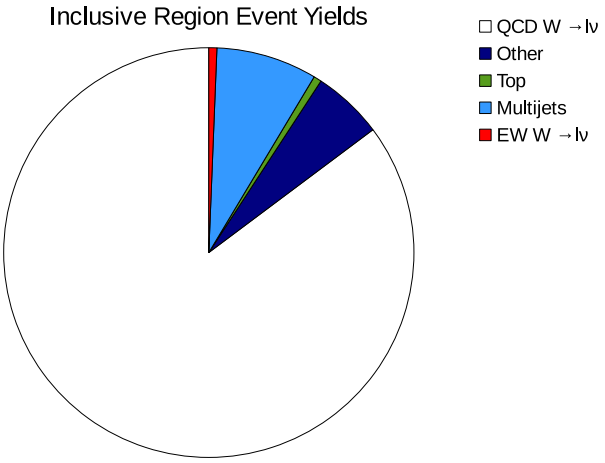


Figure 7.10: Contributions to the *Inclusive* region from different physics processes.

balance describes the percentage imbalance of the total event p_{T} and is defined:

$$p_{\mathrm{T}}\text{ balance} = \frac{\sqrt{\sum_{i=X,Y}\left[p_i^{\mathrm{Lep}} + E_{\mathrm{i}}^{\mathrm{miss}} + \sum_{j=0}^{N_{Jets}}\left(p_i^j\right)\right]^2}}{p_T^{\mathrm{Lep}} + E_{\mathrm{T}}^{\mathrm{miss}} + \sum_{j=0}^{N_{Jets}}\left(p_T^j\right)}, \tag{7.5}$$

where p_i denotes the momentum in direction i of a particle and Lep refers to the charged lepton from the W-boson decay. This is shown in Figure 7.14(d) with an increased fraction of EW signal at low p_{T} balance due to limited QCD radiation, whereas in background events the spread is wider due to objects that are not reconstructed. The N_{Jets} plot shows the prevalence of the signal in the 2-jet bin and increased contribution from top processes in the higher jet multiplicity bins. There is an overestimation from the prediction in the higher jet multiplicity bins but the 2-jet bin shows excellent agreement with data. Figure 7.14(f) shows the p_{T} of the W-boson divided by the leading jet p_{T}. The ratio shows a trend for the MC to be shifted to higher values than the data, this is caused by the same trend on the W-boson p_{T}.

Pie charts show the contributions from different processes to the *Inclusive* region and the *Search* region respectively in Figures 7.10 and 7.11. Both charts are summed over lepton channels. The effect of the selection cuts can be seen from these charts, showing the increase in signal fraction.

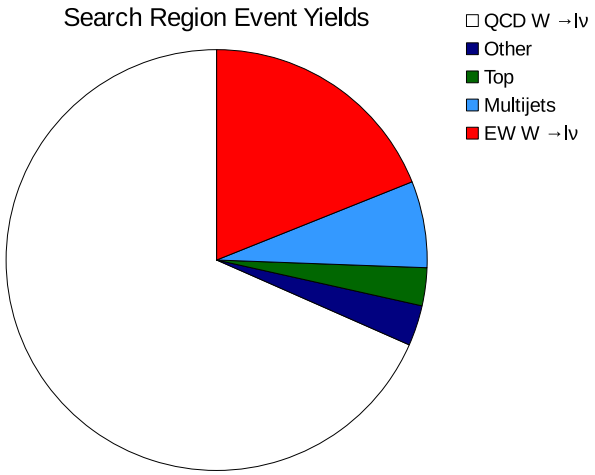
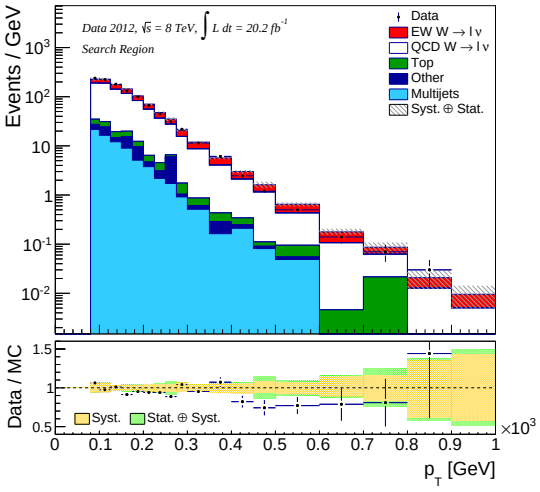


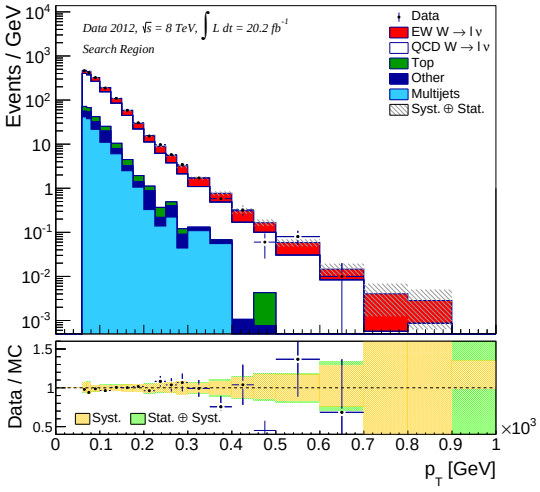
Figure 7.11: Contributions to the *Search* region from different physics processes.

7.4.1 Pre-fit Event Yields

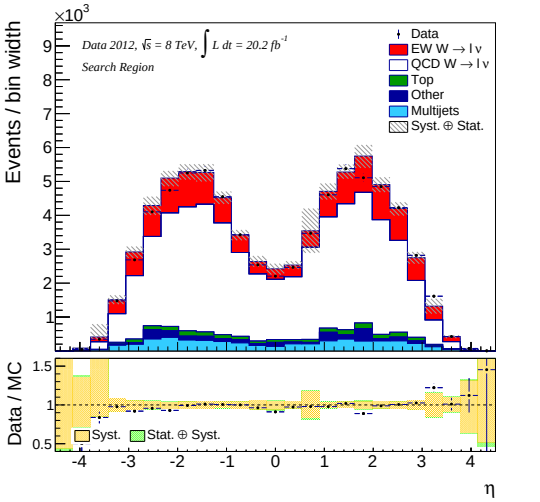
Yields tables are shown for the separate channels, electron in Table 7.2 and muon in Table 7.3 and for the sum of the channels in Table 7.4. The MC event yields have been rounded to the nearest integer. The MC event yields have been normalised to the SM, generator predicted cross sections.



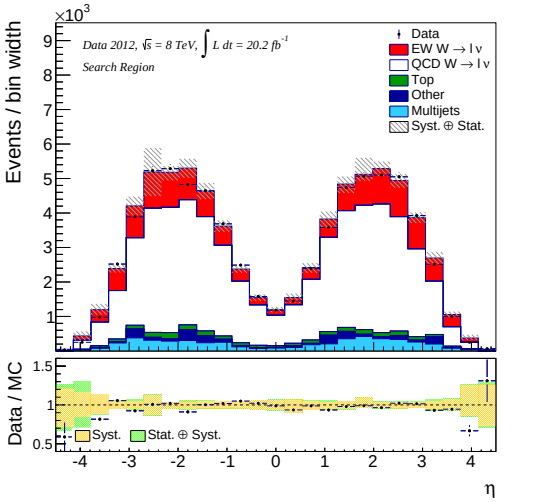
(a) Leading jet p_{T}



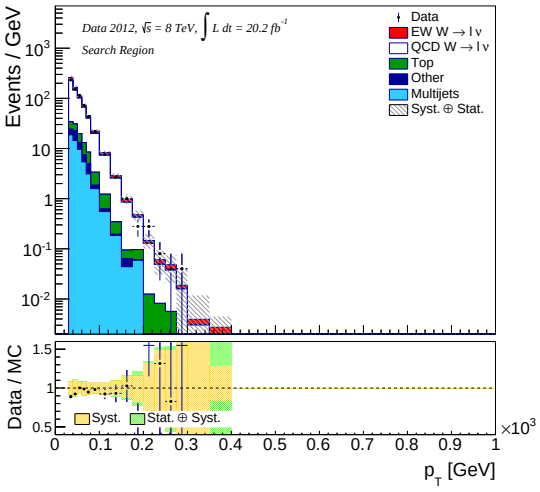
(b) Sub-leading jet p_{T}



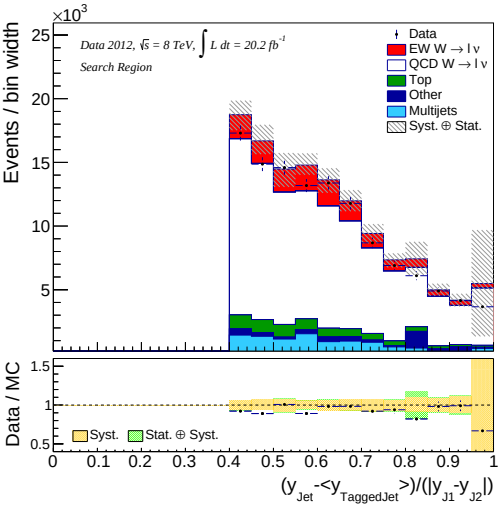
(c) Leading jet η



(d) Sub-leading jet η

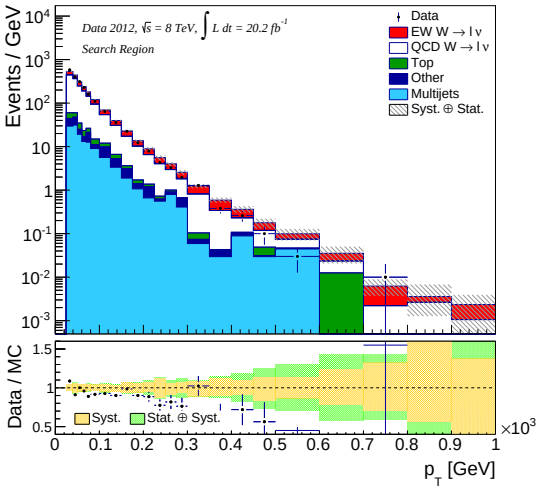


(e) Third jet p_{T}

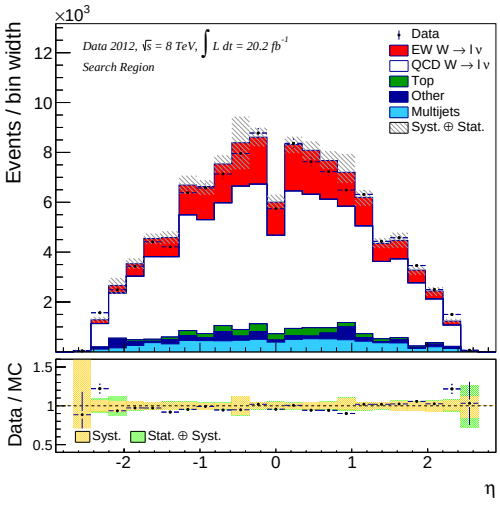


(f) Jet centrality

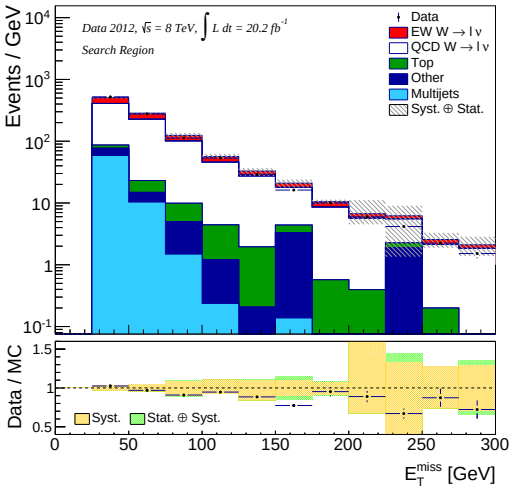
Figure 7.12: Reconstruction level jet distributions summed over both lepton channels in the *Search* region.



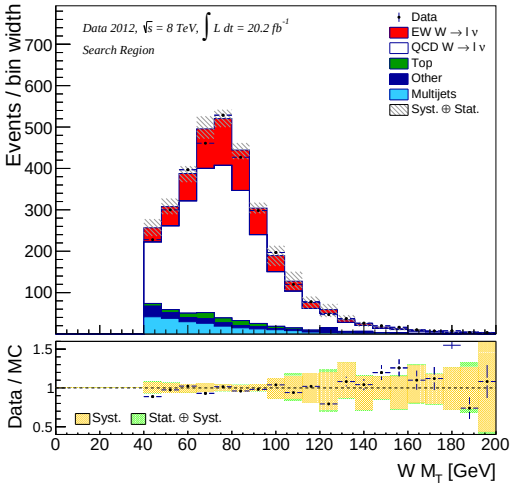
(a) Lepton p_{T}



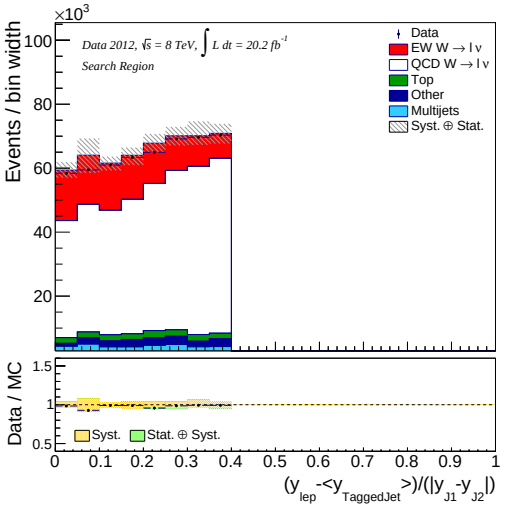
(b) Lepton η



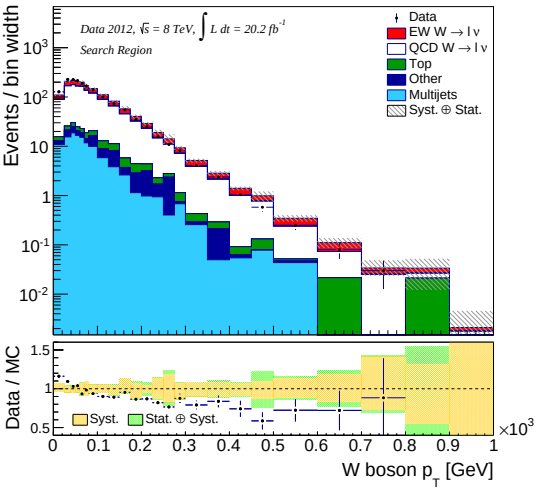
(c) $E_{\mathrm{T}}^{\mathrm{miss}}$



(d) W-boson transverse mass

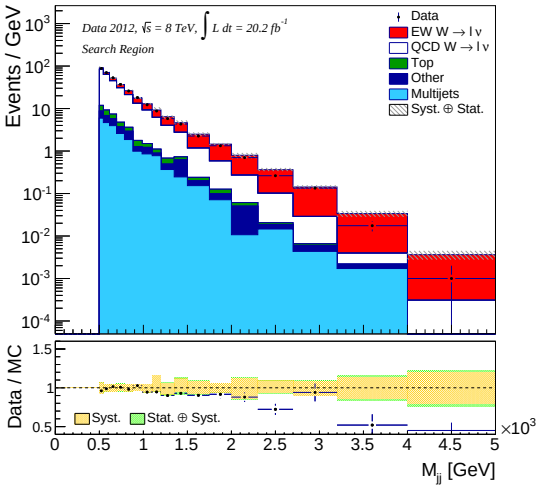


(e) Lepton centrality

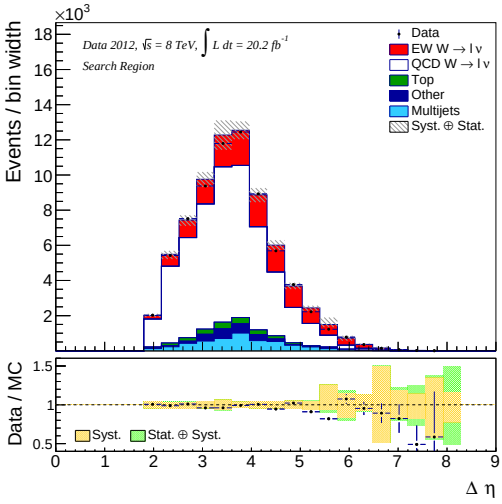


(f) W-boson p_{T}

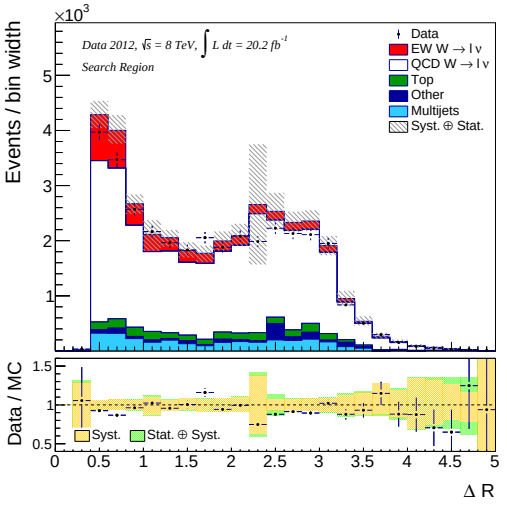
Figure 7.13: Reconstruction level lepton distributions summed over both lepton channels in the *Search* region.



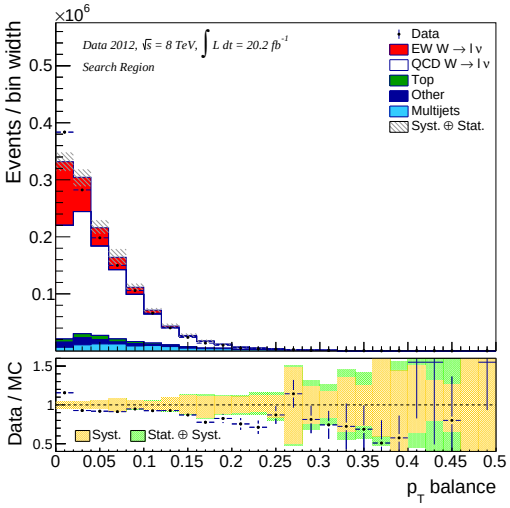
(a) M_{jj}



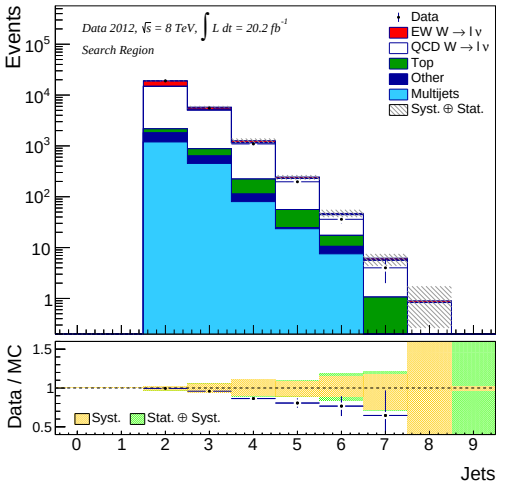
(b) $\Delta\eta(J_1, J_2)$



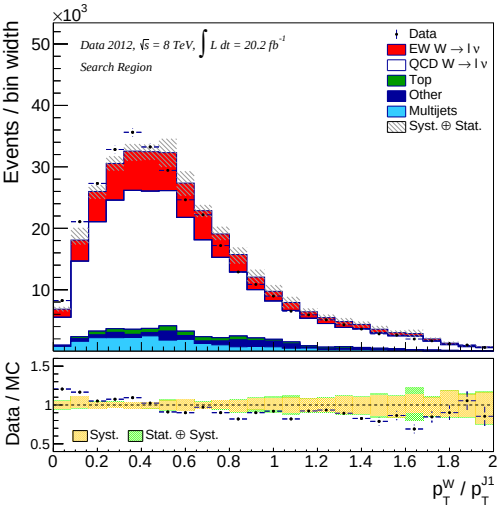
(c) Minimum ΔR (J_3, J_{Tagged})



(d) p_{T} balance



(e) Number of jets



(f) p_T^W / p_T^{J1}

Figure 7.14: Reconstruction level distributions summed over both lepton channels in the *Search* region.

Process \ Region	Inclusive	High p_T	High mass	Search	Control	Validation
QCD $W(\to e\nu)$ + 2 jets	1,868,600	262,674	27,580	8,804	8,929	4,886
QCD $W(\to \tau\nu)$ + 2 jets	32,164	2,595	366	142	156	9
ZZ	192	26	3	1	1	1
WZ	2,188	279	30	13	4	9
WW	9,615	1,037	103	49	20	24
Multijets	241,966	29,110	3,444	1,008	1,166	607
QCD $Z(\to e\bar{e})$ + 2 jets	98,697	8,192	608	215	154	132
QCD $Z(\to \tau\bar{\tau})$ + 2 jets	4,277	387	32	6	18	7
$t\bar{t}$	36,576	11,998	865	210	114	398
Single top	13,533	2,946	400	211	77	84
EW $Z(\to e\bar{e})$ + 2 jets	249	134	54	28	9	14
EW $W(\to e\nu)$ + 2 jets	14,135	6,796	3,375	2,491	564	254
Data	2,187,780	318,634	35,269	12,714	10,916	6,015
Total Prediction	2,322,190	326,174	36,859	13,177	11,214	6,425
Data / Prediction	0.94	0.98	0.96	0.96	0.97	0.94
Signal / Total Prediction	0.01	0.02	0.09	0.19	0.05	0.04
Signal / $\sqrt{\text{Background}}$	9.30	12.03	18.44	24.09	5.47	3.23

Table 7.2: Pre-fit event yields for the electron channel. The statistical uncertainty on the "Data / Prediction" ratio is $< 2\%$ in all regions.

Process \ Region	Inclusive	High p_T	High mass	Search	Control	Validation
QCD $W(\to \mu\nu)$ + 2 jets	2,093,129	283,770	30,776	9,239	10,682	5,032
QCD $W(\to \tau\nu)$ + 2 jets	41,119	3,359	234	111	0	28
ZZ	182	18	1	1	0	1
WZ	2,571	294	31	15	5	7
WW	11,634	1,197	124	54	26	31
Multijets	127,723	13,521	1,711	752	310	466
QCD $Z(\to \mu\overline{\mu})$ + 2 jets	50,601	3,453	411	127	137	93
QCD $Z(\to \tau\overline{\tau})$ + 2 jets	6,077	429	37	21	0	10
$t\bar{t}$	38,785	12,672	877	195	123	377
Single top	14,654	3,032	361	148	88	95
EW $Z(\to \mu\overline{\mu})$ + 2 jets	262	125	54	35	13	5
EW $W(\to \mu\nu)$ + 2 jets	15,430	7,227	3,529	2,521	671	264
Data	2,392,305	337,791	37,799	13,101	12,709	5,956
Total Prediction	2,402,167	329,097	38,146	13,219	12,053	6,409
Data / Prediction	1.00	1.03	0.99	0.99	1.05	0.93
Signal / Total Prediction	0.01	0.02	0.09	0.19	0.06	0.04
Signal / $\sqrt{\text{Background}}$	9.99	12.74	18.97	24.37	6.28	3.37

Table 7.3: Pre-fit event yields for the muon channel. The statistical uncertainty on the "Data / Prediction" ratio is $< 3\%$ in all regions.

Process \ Region	Inclusive	High p_T	High mass	Search	Control	Validation
QCD $W(\to l\nu)$ + 2 jets	3,961,729	546,444	58,356	18,043	19,611	9,918
QCD $W(\to \tau\nu)$ + 2 jets	73,283	5,954	600	253	156	37
ZZ	374	44	4	2	1	2
WZ	4,759	573	61	28	9	16
WW	21,249	2,234	227	103	46	55
Multijets	369,689	42,631	5,155	1,760	1,476	1,073
QCD $Z(\to l\bar{l})$ + 2 jets	149,298	11,645	1,019	342	291	225
QCD $Z(\to \tau\bar{\tau})$ + 2 jets	10,354	816	69	27	18	17
$t\bar{t}$	75,361	24,670	1,742	405	237	775
Single top	28,187	5,978	761	359	165	179
EW $Z(\to l\bar{l})$ + 2 jets	511	259	108	63	22	19
EW $W(\to l\nu)$ + 2 jets	29,565	14,023	6,904	5,012	1,235	518
Data	4,580,085	656,425	73,068	25,815	23,625	11,971
Total Prediction	4,724,359	655,271	75,006	26,397	23,267	12,834
Data / Prediction	0.97	1.00	0.97	0.98	1.02	0.93
Signal / Total Prediction	0.01	0.02	0.09	0.19	0.05	0.04
Signal / $\sqrt{\text{Background}}$	13.64	17.51	26.46	34.27	8.32	4.67

Table 7.4: Pre-fit event yields for combined electron and muon channels. The statistical uncertainty on the "Data / Prediction" ratio is $< 3\%$ in all regions.

Chapter 8

Control Region

The cuts applied in the event selection significantly reduce the contribution from the backgrounds and increase the purity of the EW signal in the *Search* region. Despite this, there is still a sizeable contribution from the dominant QCD W + jets background. The signal extraction in this analysis relies on a fit in the M_{jj} variable where we observe discrimination between the EW $W + 2$ jets signal and the QCD $W + 2$ jets background. As the *Search* region is dominated by QCD $W + 2$ jets we rely on the modelling of this background to be accurate in M_{jj}. This chapter describes the use of a *Control* region to ensure the shape of this background is modelled well. This procedure was also applied in [35, 121].

The large contribution of QCD W + jets at low M_{jj} is useful to fix its normalisation. With this normalisation, the high M_{jj} region quantifies the amount of EW W + jets signal. In the case where the QCD W + jets M_{jj} template shape is biased towards high M_{jj}, signal yield will be biased low. Conversely, if the QCD W + jets M_{jj} shape is biased to low M_{jj}, the signal yield will be biased high. The shape of the QCD W + jets M_{jj} template spectrum is therefore crucial to this analysis.

The choice of the *Control* region definition is important so as to achieve similar particle kinematics to the *Search* region to allow a small extrapolation of the observables. This analysis uses a *Control* region defined by the reversal of a lepton based cut which corresponds to an extrapolation in the weak force coupling constant. A *Validation* region is also defined

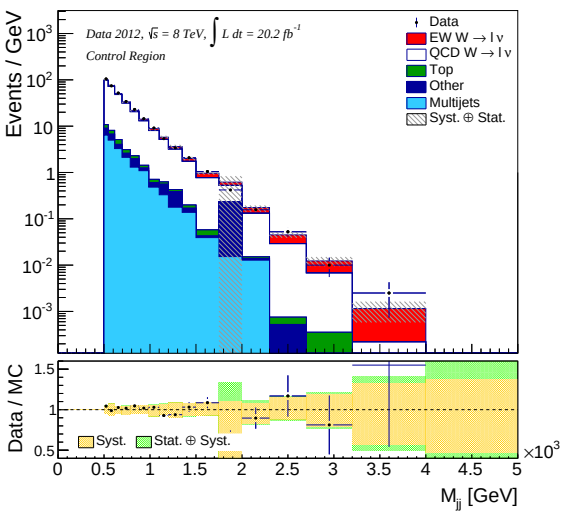
via the reversal of a jet based cut, corresponding to an extrapolation in α_S. This *Validation* region is used to validate the *Control* region correction procedure and provide a cross-check of the analysis results.

8.1 Control Region Correction

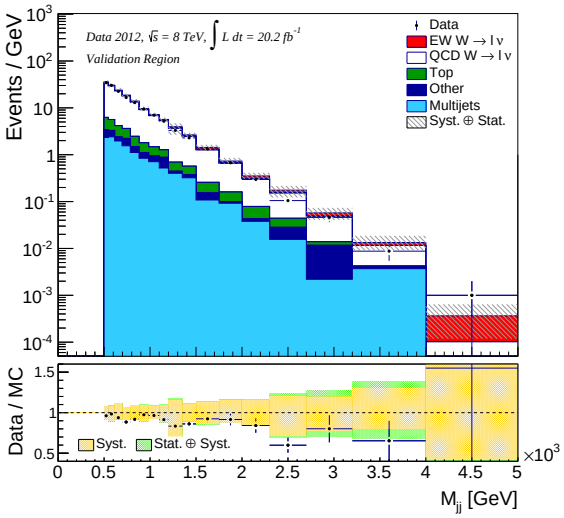
The QCD $W+$ jets M_{jj} shape can be determined from data to correct possible mismodelling in the MC. The procedure to correct the QCD $W+$ jets M_{jj} shape is to find a phase space with a high yield and purity of QCD $W+$ jets events (designated the *Control* region), subtract all other predictions from the data in this phase space, and modify the QCD $W+$ jets template shape to match the subtracted data shape. This procedure benefits if the QCD $W+$ jets process has similar kinematics in the *Control* and *Search* regions; if the M_{jj} shapes are different, the correction only partially corrects for the possible mismodelling.

We cannot compare the QCD $W+$ jets shape to data in the *Search* region where the signal contribution is measured. Instead, we use phase spaces close to the *Search* region where a signal cut is reversed. The chosen reversed cut is the lepton centrality to create the *Control* region. A second phase space is defined as the *Validation* region when reversing the jet centrality cut, see Table 7.1. The *Control* region is used as the default correction phase space and the *Validation* region is used as a cross-check of the method. The M_{jj} spectrum is shown in the *Control* region in Figure 8.1(a) and in the *Validation* region in Figure 8.1(b). Figure 8.1(a) shows a large data-to-prediction disagreement in the bin around 2 TeV, this is due to a single highly weighted MC event in the "Other" backgrounds sample.

The variables used to form the *Control* and *Validation* regions are $\mathcal{C}_{Lepton}$ and $\mathcal{C}_{Jet}$ respectively. The $|\mathcal{C}_{Lepton}|$ distribution is shown in Figure 8.2(a) for the *High Mass* region and for the *Control* region in Figure 8.2(b). The $|\mathcal{C}_{Jet}|$ distributions are shown in Figure 8.3(a) for the *High Mass* region and in Figure 8.3(b) for the *Validation* region. In Figure 8.3(b) a tail can be seen for non-tagged jets with $|\mathcal{C}_{Jet}| > 0.4$, this is because the plot shows the $|\mathcal{C}_{Jet}|$ values for **all** non-tagged jets. This tail corresponds to events where one non-tagged jet has



(a) *Control* region M_{jj} distribution.



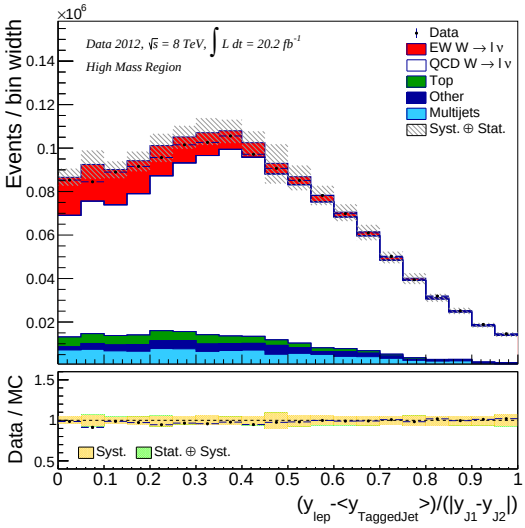
(b) *Validation* region M_{jj} distribution.

Figure 8.1: *Control* and *Validation* region M_{jj} distributions before the QCD W + jets correction is applied.

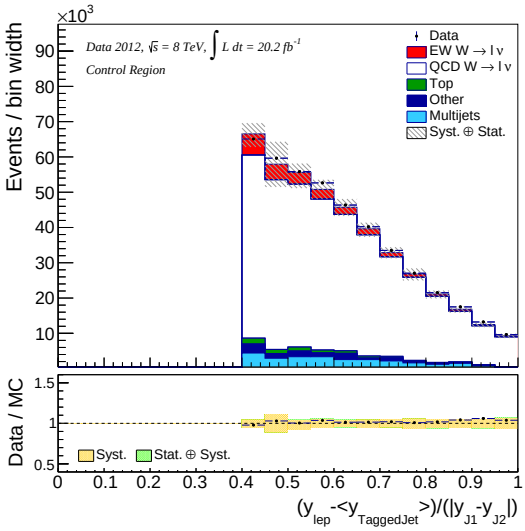
$|\mathcal{C}_{Jet}| < 0.4$ and hence the event fulfils the *Validation* region requirement, but additional jets can have higher $\mathcal{C}_{Jet}$ values.

Correction functions for the QCD W+jets distributions are derived from these two phase spaces. Considering both the electron and muon channels together, predictions from other processes in the *Control* region are subtracted from the data to create a "data-driven template". The W-boson decay exhibits "ltpon universality" meaning the physics are identical between the electron and muon channels. This template is divided by the QCD W + jets MC template to give a ratio of QCD W + 2 jets events in data to the MC prediction for QCD W + 2 jets that can be used to correct the modelling of this prediction. This procedure is desingned to correct any systematic bias in the QCD W + 2 jets modelling so the ratio is fit with a second order polynomial function. This function is applied in the *Search* region to correct the QCD W + jets modelling. The ratio plots and fitted functions are shown in Figure 8.9 for M_{jj} in both the *Control* and *Validation* regions.

The *Control* region has a flatter correction function than the *Validation* region, although it has large statistical errors above $M_{jj} \approx 2$ TeV due to the softer M_{jj} spectrum in the *Control* region (as seen in Figure 8.1). The values and errors of the second order polynomial

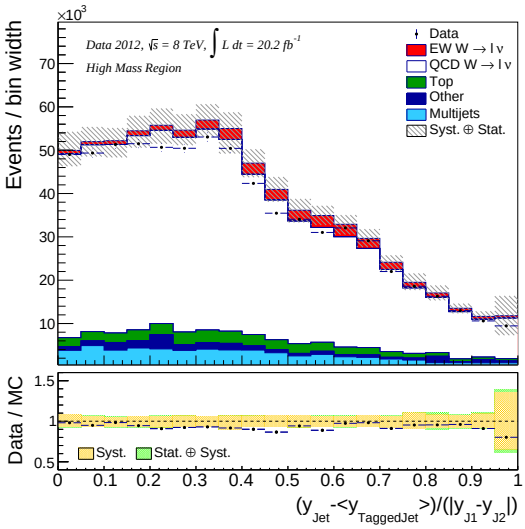


(a) *High Mass* region.

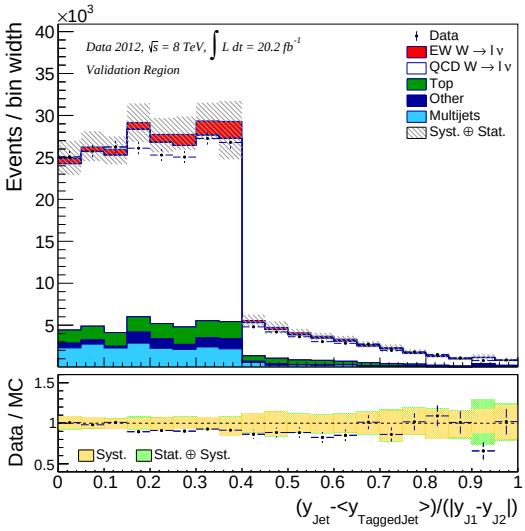


(b) *Control* region.

Figure 8.2: The lepton centrality distribution.

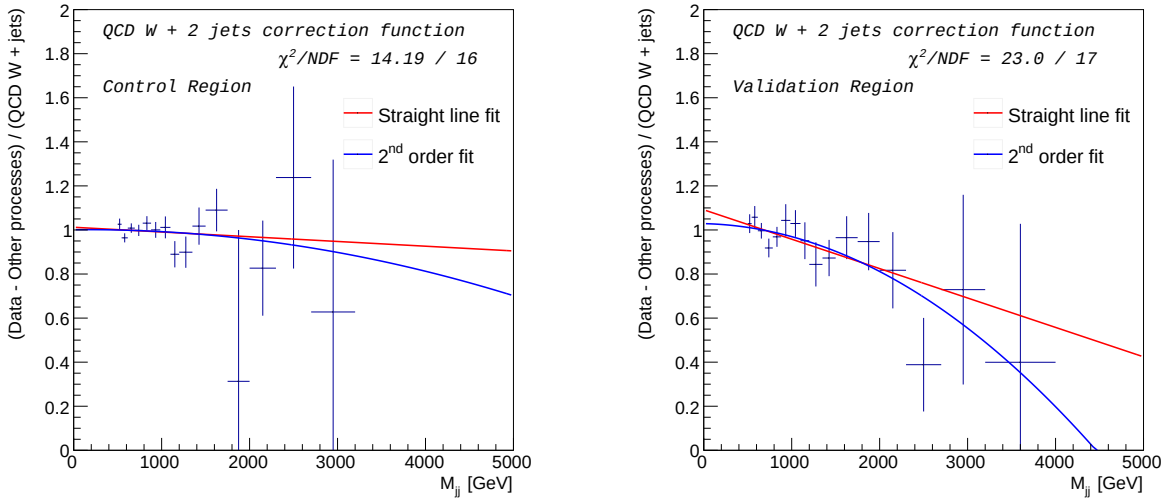


(a) *High Mass* region.



(b) *Validation* region.

Figure 8.3: The jet centrality distribution. The plots show $|\mathcal{C}_{Jet}|$ for **all** non-tagged jets.



(a) *Control* region.

(b) *Validation* region.

Figure 8.4: *Control* and *Validation* region QCD W + jets correction functions for M_{jj} from a fit to (data - other processes) / (QCD W + jets). First and second-order polynomial fit results are shown in red and blue respectively. The error bars are statistical only. The χ^2 values show the consistency of the ratio points with a flat line (i.e. no correction being required).

Phase Space	p_0	Δp_0	p_1 (GeV^{-1})	Δp_1 (GeV^{-1})	p_2 (GeV^{-2})	Δp_2 (GeV^{-2})
Control	1.021	0.08	0.11×10^{-4}	1.7×10^{-4}	-1.6×10^{-8}	8.3×10^{-8}
Validation	0.969	0.08	-0.38×10^{-4}	1.5×10^{-4}	-3.6×10^{-8}	5.8×10^{-8}

Table 8.1: Second-order polynomial correction function parameters and errors from *Validation* and *Control* regions.

correction functions are shown in Table 8.1.

The individual parameters in both functions are in agreement within the statistical errors. The parameters are correlated so this does not translate into statistical consistency between the functions. This can be better checked by fitting with a straight line. The parameters and errors for such fits are shown in Table 8.2; the p_1 values differ by 1.8σ.

Phase Space	p_0	Δp_0	p_1 (GeV^{-1})	Δp_1 (GeV^{-1})
Control	1.034	0.03	-0.20×10^{-4}	0.44×10^{-4}
Validation	1.015	0.04	-1.27×10^{-4}	0.41×10^{-4}

Table 8.2: Cross-check, first-order polynomial correction function parameters and errors from the *Validation* and *Control* regions.

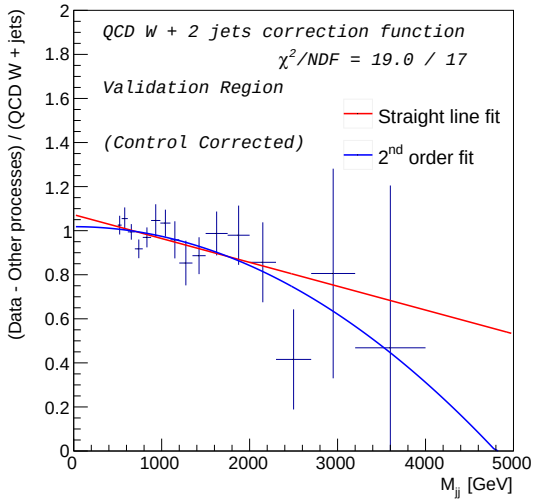
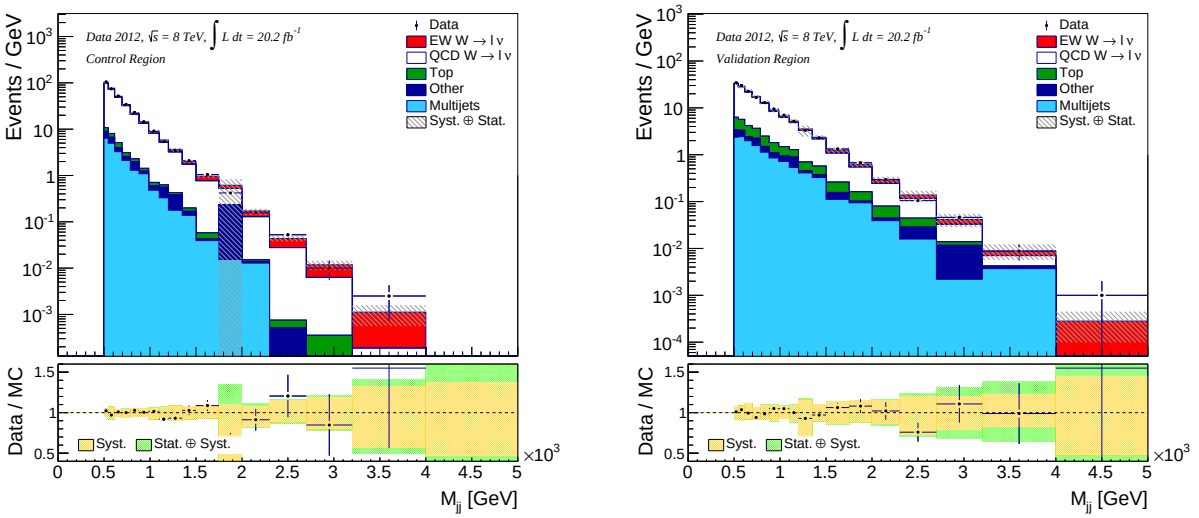


Figure 8.5: *Validation* region (data - other processes) / (QCD $W+$ jets) M_{jj} distribution after the *Control* region second-order correction function has been applied to QCD $W+$ jets. First and second-order polynomial fit results are shown in red and blue respectively. The error bars are statistical only.

The second-order correction function from the *Control* region is applied to the QCD $W+$ jets in the *Validation* region to cross-check the consistency of the correction functions. Figure 8.5 shows the ratio of subtracted data to QCD $W+$ jets in the *Validation* region after the *Control* region constraint has been applied to the QCD $W+$ jets. First and second-order polynomial fits to the ratio are shown. The individual parameters of the second-order fit are all consistent with a flat line. The first-order fit gives a gradient which is 1.9σ from flat, $p_1 = (-9.833 \pm 5.17) \times 10^{-5}$.

The QCD $W+$ jets distributions in the *Search* region can be corrected using the above procedure. The *Control* and *Validation* region M_{jj} distributions after the second-order corrections from their respective regions are applied are shown in Figure 8.6.

The correction function from each phase space is applied to the QCD $W+$ jets M_{jj} distribution in the other phase space and the results are shown in Figure 8.7. The distributions are approximately flat, though there is some mismodelling of the normalisation in each region. The QCD $W+2$ jets normalisation is fit for in the signal extraction, so mismodelling of the normalisation in the *Control* or *Validation* region does not affect the result. Therefore, the normalisation of the QCD $W+$ jets is allowed to float and the distributions are again



(a) *Control* region.

(b) *Validation* region.

Figure 8.6: M_{jj} distributions in the *Control* and *Validation* regions after applying the second-order QCD W + jets corrections derived from the respective region. The Data/MC agreement is shown to be improved as compared to the un-corrected case shown in Figure 8.1.

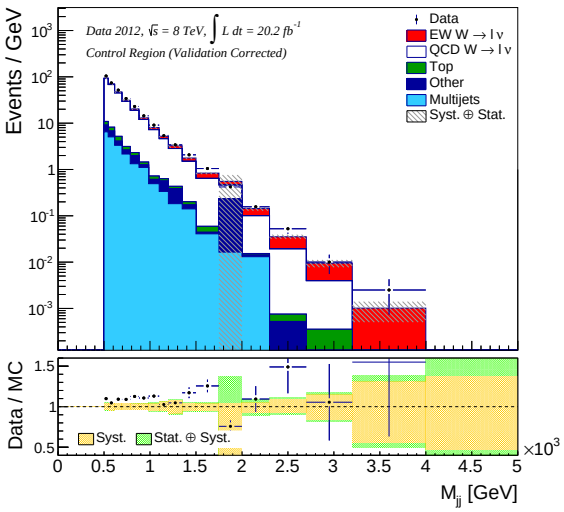
shown in Figure 8.8 where the QCD W + jets has been scaled up by 9.2% in the *Control* region and down by 10.2% in the *Validation* region.

The second order *Control* region correction function is used as the nominal correction and the other functions are used to cross-check the results.

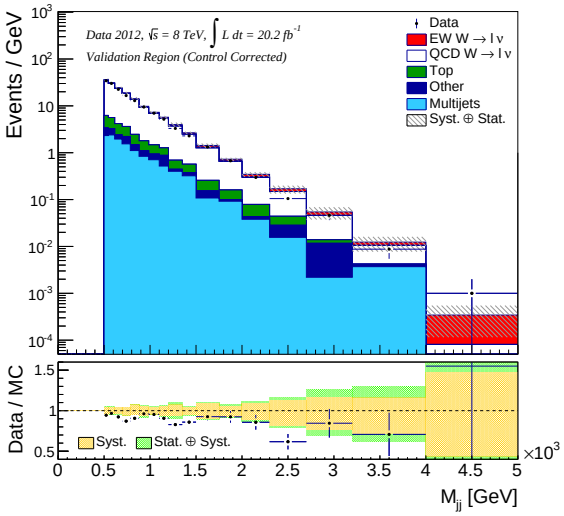
8.2 Control Region Constraints

Alongside correcting possible mismodelling of the QCD W + jets background, the *Control* region correction also reduces the uncertainties on this process. The residual uncertainties in the *Search* region are uncertainties on the extrapolation of the shape constraint from the *Control* region to the *Search* region.

The procedure to constrain the QCD W + jets uncertainties is complex so an example case is given below to demonstrate the principle. Consider a case with a single uncertainty, for example the jet energy scale. If the ATLAS jet energy scale had a systematic bias such that the M_{jj} distribution were incorrectly reconstructed, this bias would be present in both



(a) *Control* region.



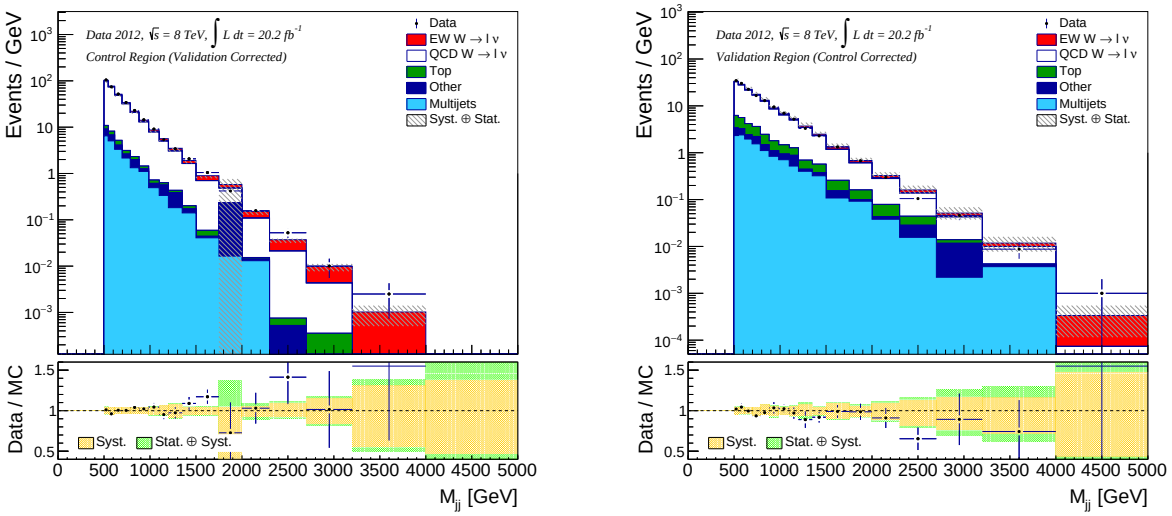
(b) *Validation* region.

Figure 8.7: M_{jj} distributions in the *Control* and *Validation* regions after applying the second-order QCD W + jets corrections derived from the other region.

the *Search* and *Control* regions. The correction function will mitigate the bias completely in the *Control* region. Applying the correction to the *Search* region corrects the bias to the extent that the source of the bias has a common effect on the M_{jj} distributions. The residual uncertainty is the difference in the effect of the jet energy scale in the two regions.

To calculate the residual uncertainty, a double ratio is used. The effect in the *Search* region is given by the M_{jj} distribution calculated with the systematic raised by 1σ divided by the nominal distribution. The same ratio is formed in the *Control* region and the ratio of the effect in the *Search* region to the effect in the *Control* region gives the reduced systematic uncertainty.

The calculation of the double ratio is demonstrated in Equation 8.1, showing the effect of the uncertainty on M_{jj} in the *Search* region ($E^{\text{Search}}_{M_{jj}}$), the effect in the *Control* region



(a) *Control* region.

(b) *Validation* region.

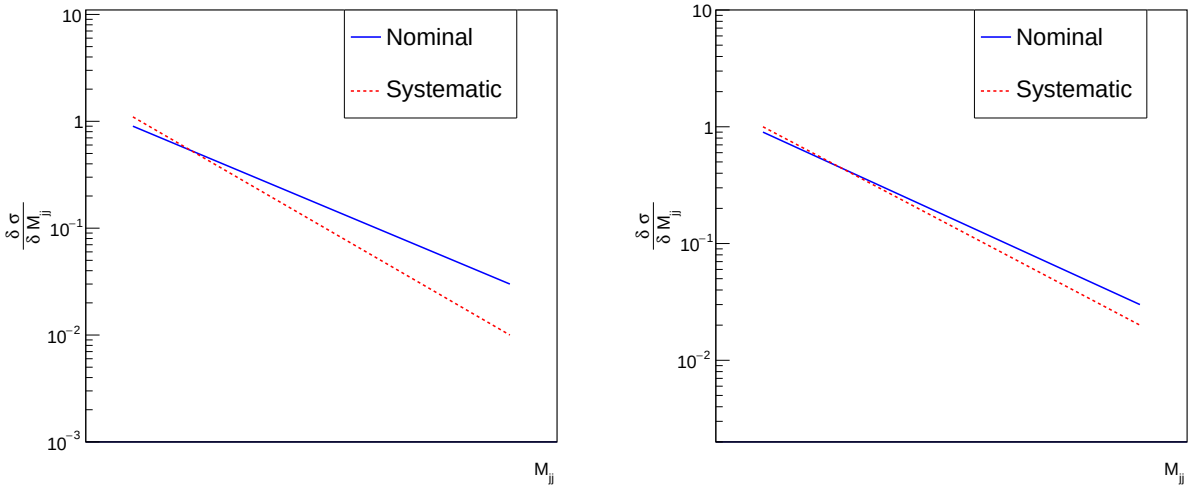
Figure 8.8: M_{jj} distributions after the second order QCD W + jets corrections are applied from the other region. The QCD W + jets normalisation has been adjusted to match the data.

($E^{\text{Control}}_{M_{jj}}$), and finally, the constrained effect in the *Search* region ($E^{\text{Search}}_{M_{jj},\,\text{Constrained}}$).

$$
\begin{aligned}
E^{\text{Search}}_{M_{jj}} &= \frac{M^{\text{Search}}_{jj,\text{Systematically Varied}}}{M^{\text{Search}}_{jj,\text{Nominal}}}, \\
E^{\text{Control}}_{M_{jj}} &= \frac{M^{\text{Control}}_{jj,\text{Systematically Varied}}}{M^{\text{Control}}_{jj,\text{Nominal}}}, \\
E^{\text{Search}}_{M_{jj},\,\text{Constrained}} &= \frac{E^{\text{Search}}_{M_{jj}}}{E^{\text{Control}}_{M_{jj}}}.
\end{aligned}
\tag{8.1}
$$

8.2.1 Control Region Choice

It is important to note that the choice of *Control* region affects the constraints on the QCD W + jets uncertainties. If the *Control* region is affected by the uncertainties in the same manner as the *Search* region, the uncertainties will be reduced. However, if the *Control* and *Search* regions are affected in opposite ways by an uncertainty, the resulting constrained uncertainties will actually increase. Therefore it is important to choose a *Control* region



(a) *Search* region. (b) *Control* region.

Figure 8.9: The differential cross section in M_{jj} and effect of an example uncertainty compared to the nominal distribution, in the *Search* and *Control* regions.

where the particle kinematics match those in the *Search* region.

Figure 8.9 shows how an example uncertainty may affect the *Search* and *Control* regions. The systematic affects the two regions slightly differently but in magnitude rather than in direction. The resulting constrained uncertainty, from Equation 8.1, is shown in Figure 8.10. In the case where the uncertainty affects the two phase spaces oppositely, the constrained systematic is larger than the original, as seen in Figure 8.11.

8.3 Analysis Effects

In this analysis the *Control* and *Validation* regions both have similar kinematics to the *Search* region. It is therefore expected that the correction from either region will reduce the uncertainties on the *Search* region. The correction and constraints procedure is applied to the *Search* region in Figure 8.12 and compared to the original *Search* region distribution. The signal yield shown is that of the SM prediction. The fit for the signal yield will modify the slope of the data/prediction ratio.

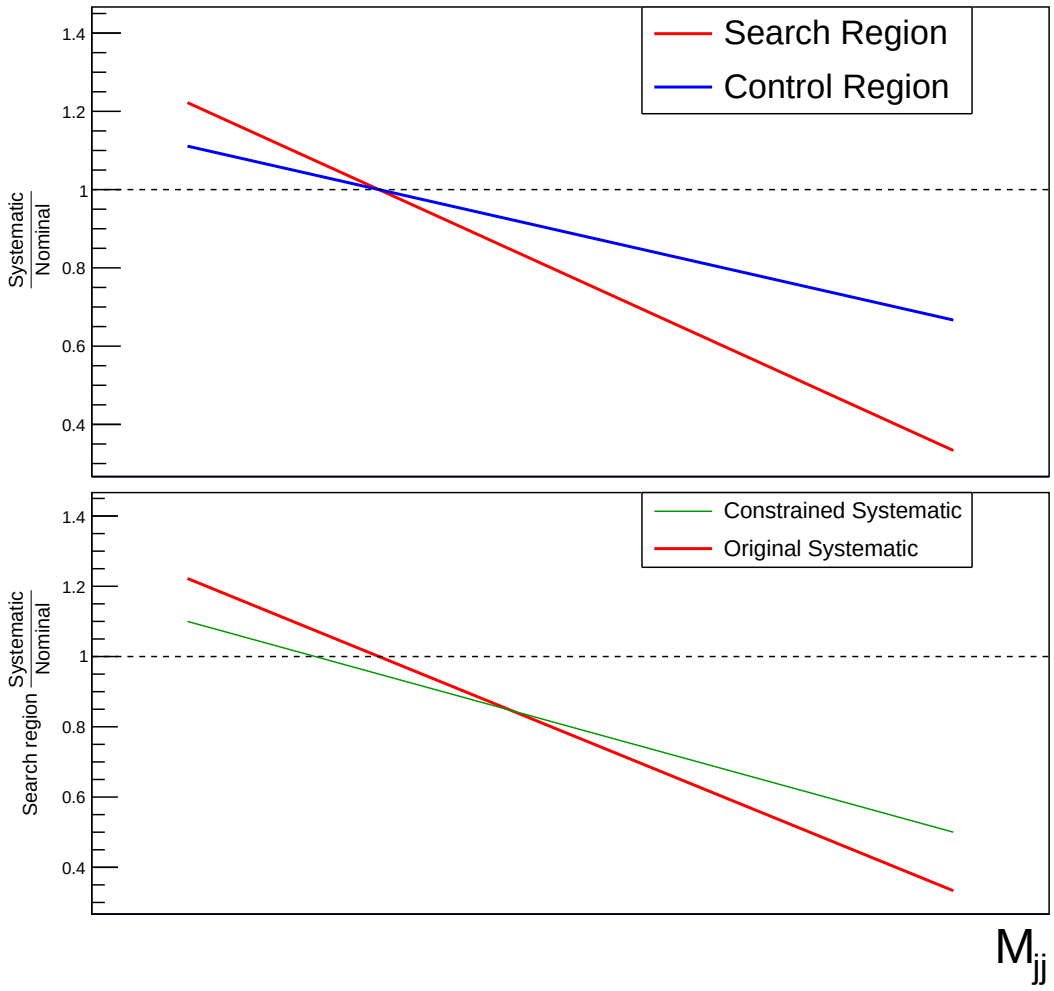


Figure 8.10: The resulting uncertainty effects in the *Search* and *Control* regions for an example systematic uncertainty that affects the two regions in similar ways (top panel) and the constrained systematic uncertainty compared to the original (bottom panel).

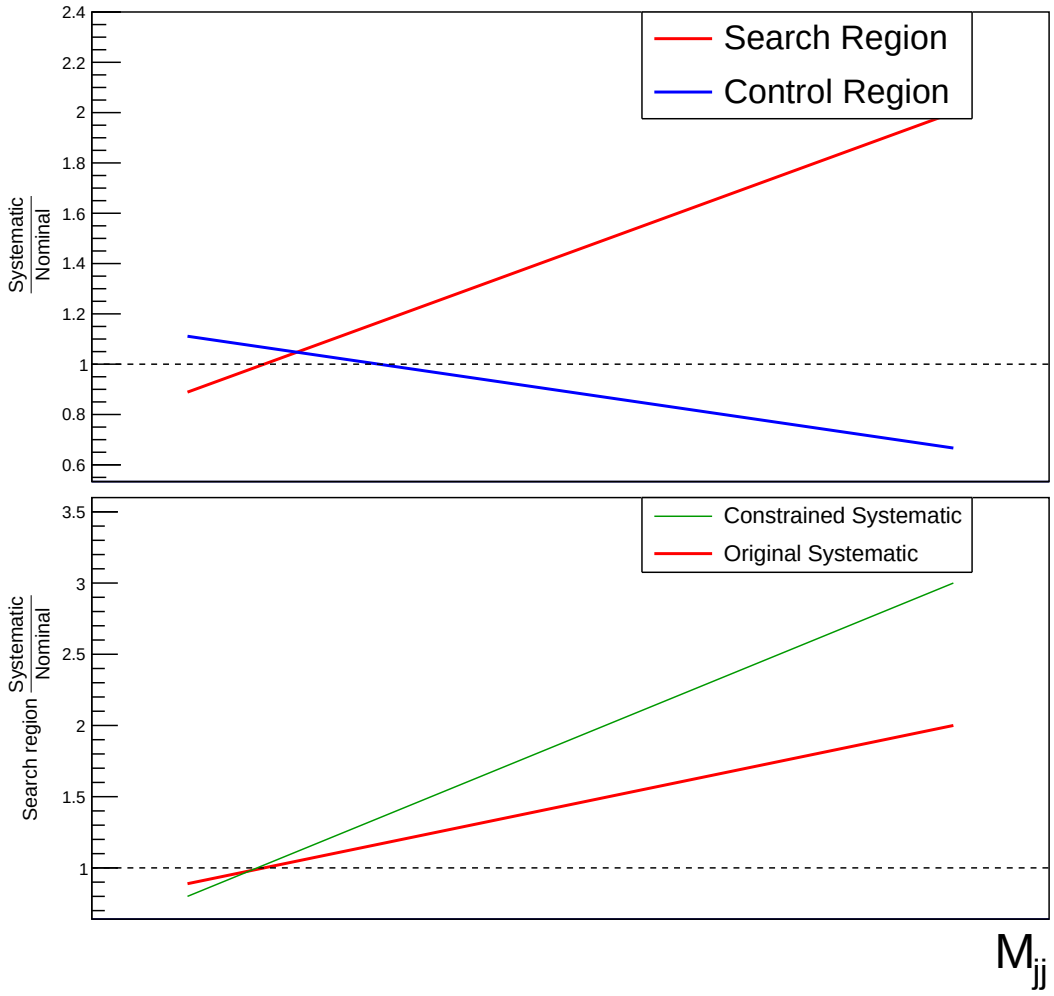
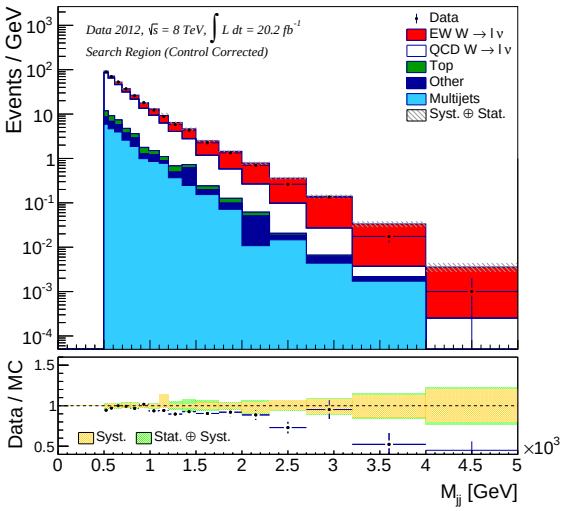
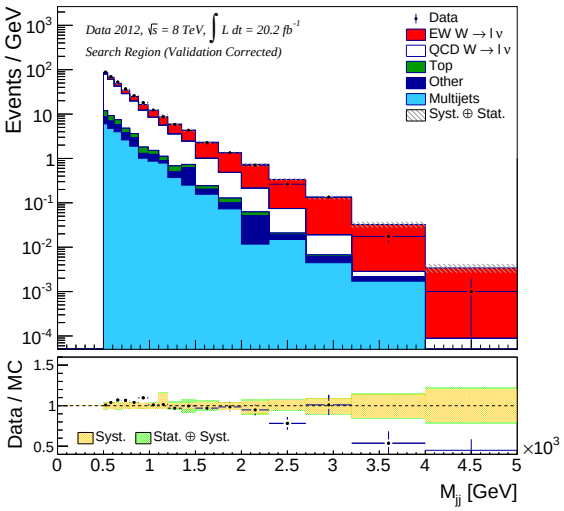


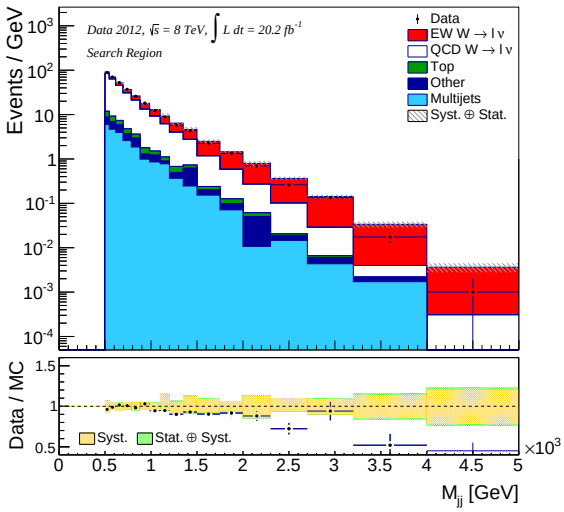
Figure 8.11: The resulting uncertainty effects in the *Search* and *Control* regions for an example systematic uncertainty that affect the two regions differently (top panel) and the constrained systematic uncertainty compared to the original (bottom panel).



(a) *Search* region M_{jj} distribution after *Control* region correction and constraints.



(b) *Search* region M_{jj} distribution after *Validation* region correction and constraints.



(c) The original *Search* region M_{jj} distribution.

Figure 8.12: The effect of the *Control* and *Validation* region corrections and constraints on the *Search* region M_{jj} distribution. The yellow bands in the ratio plots show the reduction achieved from the QCD W + jets uncertainty constraints in Figures 8.12(a) and 8.12(b) compared to the original band in Figure 8.12(c). The signal process is shown normalised to the SM prediction.

Chapter 9

Signal Extraction

This chapter describes the methods developed to extract and measure the EW signal from the data.

9.1 Fitting Procedure

To extract the signal contribution from the data, a fit is performed in M_{jj}, the variable with the best discrimination between the signal and background. The fit procedure uses the RooFit package [122] to maximise a log-likelihood function using `MINUIT` [123, 124]. A brief overview of the maximum log-likelihood method is given here.

Given the low signal purity after selection we must use statistical methods to extract the signal. The Poisson distribution is a discrete probability distribution that describes the number of observed events. The distribution is defined by:

$$f(n;\rho) = \frac{\rho^n e^{-\rho}}{n!}, \tag{9.1}$$

where $f(n;\rho)$ is the Poisson probability to observe n events, given a mean of ρ expected events.

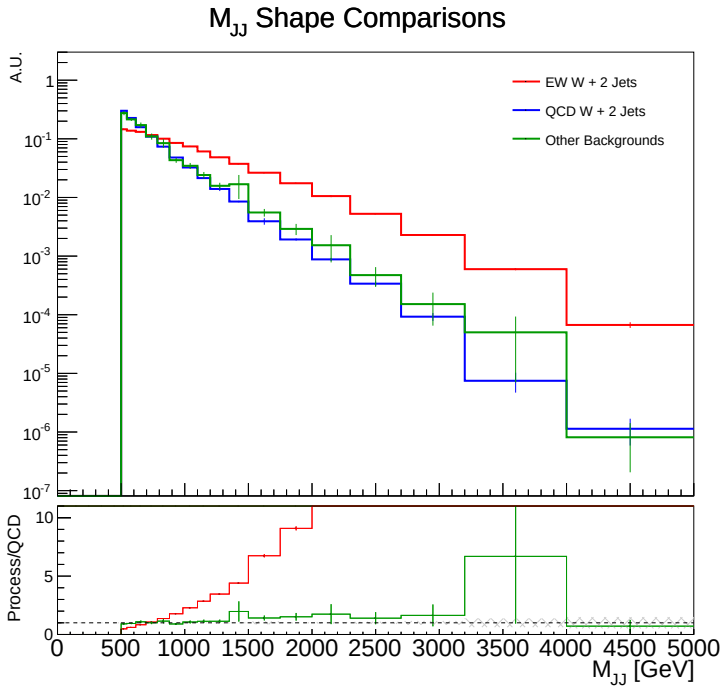


Figure 9.1: Unity normalised M_{jj} shapes that are used in the fit procedure.

The unity normalised M_{jj} distributions are shown in Figure 9.1 for the EW signal, QCD $W+$ jets background and other backgrounds.

The fits in this analysis are used to extract three parameters of interest, the best fit number of events for:

N_{EW} (The EW $W+$ jets signal),
N_{QCD} (The QCD $W+$ jets background), and
N_B (The other backgrounds).

The total number of events in the fitted results, N, is thus,

$$N = N_{EW} + N_{QCD} + N_B. \tag{9.2}$$

Since we have differential information for each component, we perform the fit in bins with index i, with total content n_i. The "likelihood" $\mathcal{L}$ is defined:

$$\mathcal{L}\left(N_{EW}, N_{QCD}, N_B\right) = \prod_{i=0}^{Bins} \frac{n_i^{\mathcal{D}_i} e^{-n_i}}{\mathcal{D}_i!}. \tag{9.3}$$

The likelihood is interpreted as the probability the observed, binned data ($\mathcal{D}_i$) could arise given a model with binned content n_i. The fit determines the maximum value of the likelihood by varying the parameters N_{EW}, N_{QCD} and N_B. In practise it is more convenient to deal with a sum rather than a product so the natural log of Equation 9.3 is taken:

$$\ln \mathcal{L} = \sum_{i=0}^{Bins} \left[\mathcal{D}_i \ln n_i - n_i - \ln \left(\mathcal{D}_i ! \right) \right]. \tag{9.4}$$

This last step is valid because the natural logarithm is a monotonic function, so the maximum of $\mathcal{L}$ is equivalent to the maximum of $\ln \mathcal{L}$ and the minimum of $(-\ln \mathcal{L})$. As we are only concerned with maximising $\ln \mathcal{L}$ relative to changes in n_i, the final term in Equation 9.4 can be ignored as it is independent of n_i. Finally, a term is added to the likelihood that constrains N_B to a gaussian, centred at the predicted cross section with a width of the cross section error σ_B. This step constrains the "other backgrounds" to their cross section as otherwise the similar M_{jj} shapes of the "other backgrounds" and the QCD W + jets (see Figure 9.1) allow little distinction. A penalty term is added to $(-\ln \mathcal{L})$ such that N_B is allowed to float in the fit but any movement away from the predicted number of events N'_B increases the value of $(-\ln \mathcal{L})$, disfavouring this fit. The function that is minimised in the fit is then:

$$-\ln \mathcal{L} = \sum_{i=0}^{Bins} \left(n_i - \mathcal{D}_i \ln n_i \right) + \frac{\left(N_B - N'_B \right)^2}{2\sigma_B^2}. \tag{9.5}$$

The uncertainty on this fitting procedure is calculated using the profile likelihood ratio λ. For example, when considering the error on N_{EW}:

$$\lambda = \frac{\mathcal{L}\left(N_{EW}, N_{QCD}^{Max}, N_B^{Max} \right)}{\mathcal{L}_{Max}}, \tag{9.6}$$

where N_{QCD}^{Max} and N_B^{Max} are the values that maximise $\mathcal{L}$ for the particular N_{EW} under consideration and $\mathcal{L}_{Max}$ is the maximum possible value for $\mathcal{L}$ when all parameters are varied. The denominator is effectively providing a normalisation. The standard 1σ error band is

Generator	Cross section (fb)	Scale Error (fb)	PDF Error (fb)	Total Error (fb)	Total Error (%)
VBFNLO	221.37	+2.63 -1.17	+3.82 -4.22	+4.64 -4.38	+2.10 -2.00
POWHEG	229.49	+3.44 -3.21	+2.11 -2.14	+4.04 -3.86	+1.76 -1.68

Table 9.1: Predicted EW $W + 2$ jets fiducial cross sections and associated errors for a single lepton channel.

found when the value of the likelihood changes by one half, $\Delta\lambda = 1/2$. This uncertainty is denoted "fit-stat." in the rest of this thesis. Examples of expected fit-stat. uncertainties can be seen in Figure 9.2.

It is often useful to quantify the fit results relative to the SM prediction. The change in the signal is denoted:

$$\mu_{EW} = \frac{N_{EW}^{\text{Fit}}}{N_{EW}^{\text{SM}}}, \tag{9.7}$$

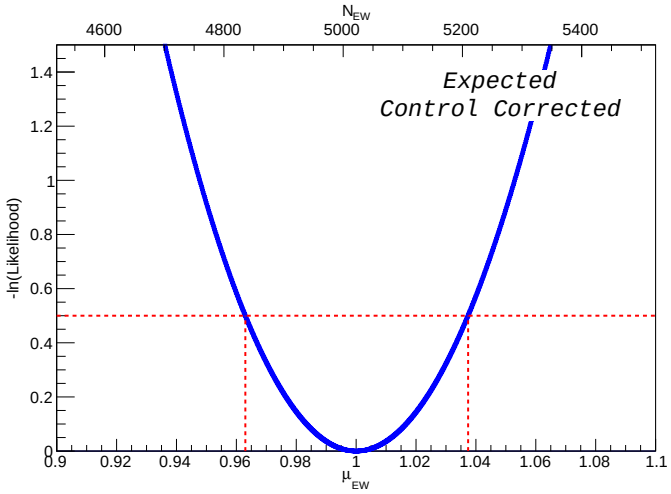
with values of μ_{QCD} and μ_B defined in the same manner. It should be noted that the fractional uncertainty on μ_{EW} is found with:

$$\frac{\Delta\mu_{EW}}{\mu_{EW}} = \sqrt{\left(\frac{\Delta N_{EW}^{\text{Fit}}}{N_{EW}^{\text{Fit}}}\right)^2 + \left(\frac{\Delta N_{EW}^{\text{SM}}}{N_{EW}^{\text{SM}}}\right)^2}, \tag{9.8}$$

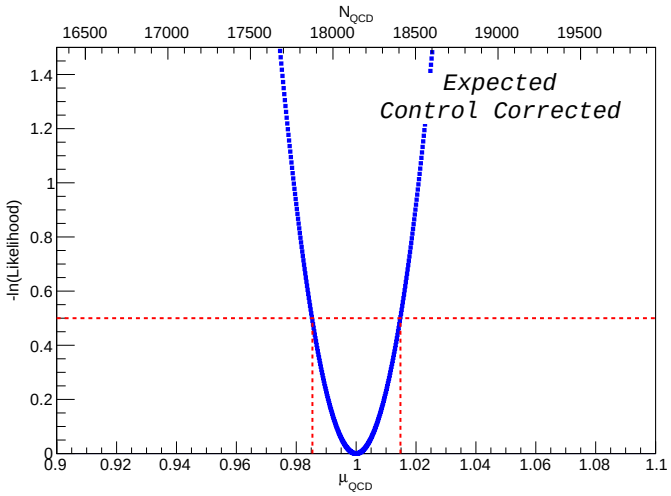
where the value $\Delta N_{EW}^{\text{Fit}}$ is found from the fits and $\Delta N_{EW}^{\text{SM}}$ comes from the theoretical uncertainty on the EW $W + 2$ jets cross section. The error on N_{EW}^{SM} is found using the signal POWHEG samples and compared to the errors found from VBFNLO calculations. Table 9.1 shows the theoretical uncertainties on the fiducial signal cross section.

The error on the predicted fiducial cross section is taken as $+2.1\%,-2\%$ from the VBFNLO prediction as it shows the largest error. This error is always added in quadrature with the $\Delta N_{EW}^{\text{After fit}}$ fractional error when quoting uncertainties on μ_{EW}.

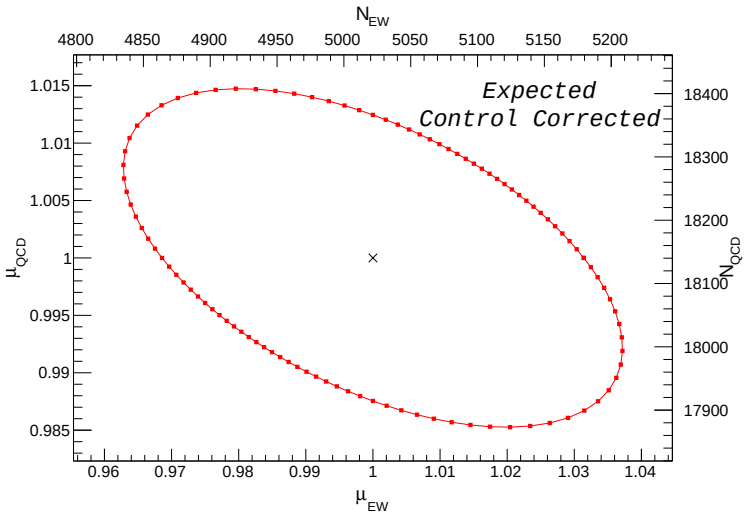
The fits are performed using six input histograms representing the M_{jj} spectra of the electron and muon channels for the EW signal, QCD $W +$ jets background and the other combined backgrounds. These template histograms are normalised to cross sections and the



(a) Expected μ_{EW} fit-stat. error



(b) Expected μ_{QCD} fit-stat. error



(c) Expected correlations for μ_{EW} and μ_{QCD} with 1σ fit-stat. error

Figure 9.2: Expected fitting results using pseudo-data for the EW signal and QCD background using the *Control* region. The blue curves show the $(-\ln \mathcal{L})$ function. The red lines show the $\ln(\lambda) = 0.5$ fit stat. error. As expected N_{EW} and N_{QCD} show a negative correlation maintaining the normalisation in the fits.

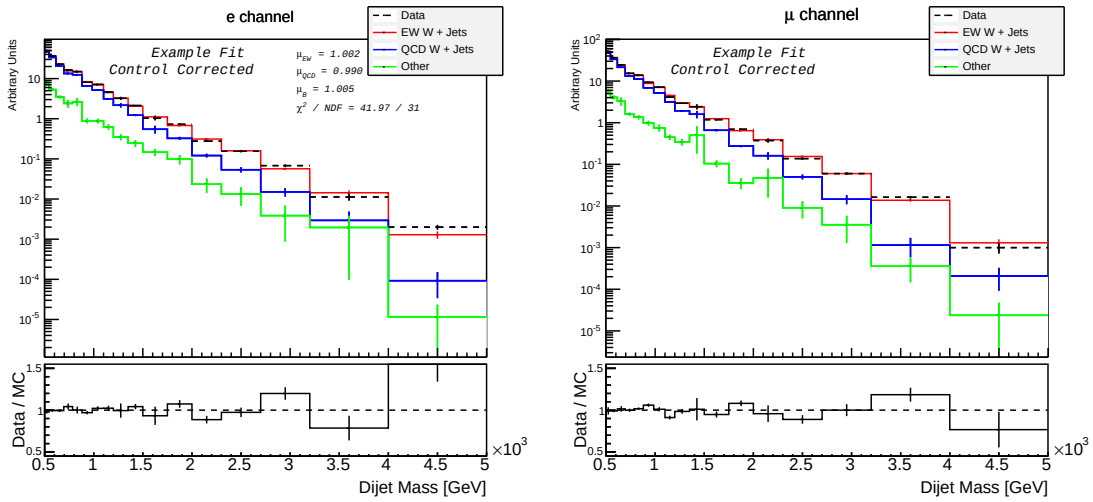


Figure 9.3: Example fit results when fitting nominal templates to pseudo-data created from themselves.

luminosity of the data sample. The fit is performed simultaneously in both lepton channels stipulating that the scale factors in both channels must be the same for each component (e.g. $\mu_{EW}^{\text{Electron}} = \mu_{EW}^{\text{Muon}}$). We justify this requirement by the SM stipulation of lepton universality under the weak force, in effect expecting the cross-section of the electron and muon channels to be identical. The results of an example fit where pseudo-data are created from the nominal input templates are shown in Figure 9.3. The fit passes the basic closure test of returning scale factors close to 1.0 for all templates when fitting to Poisson fluctuated pseudo-data.

9.1.1 Pseudo-Data Sets

The analysis in this chapter relies on pseudo-data-sets used to extract expected uncertainties and predictions on the signal extraction. This section describes these pseudo-data-sets and how they are used.

Throughout the signal extraction process it is useful to quantify how the predicted changes in QCD W + jets, EW W + jets and other backgrounds template shapes, due to uncertainties, can affect the amount of signal that is extracted from the final fit. The affect of the uncertainties is found by repeating the analysis for each of the measurement uncertainties mentioned in Chapter 6 to produce modified histograms. A pseudo-data-set is

constructed by adding the predicted templates together. This can then be used to fit the nominal samples and to view variations in extracted scale parameters from the fit. Unless specifically stated, throughout this thesis varied samples are put into the pseudo-data-sets and the nominal templates are fitted to them. This construction was used so as to mimic the final signal extraction, where the nominal templates will be fitted to the data.

9.2 Fit Uncertainties

This section describes the effect of the experimental and theoretical uncertainties on the signal extraction.

9.2.1 Fit Constraints

As discussed in Section 8, a re-weighting is applied to the QCD W + jets background from the *Control* region in order to correct the mismodelling. This *Control* region reduces the uncertainties on the QCD W + jets shape. The *Control* region constraint is applied for all fit results unless explicitly stated otherwise*.

9.2.2 Statistical Uncertainties

The finite number of events in the input templates results is an uncertainty on the fit results. The statistics uncertainties are completely uncorrelated between samples and the statistics of each bin in a template histogram are uncorrelated with the next bin. Each sample is therefore considered separately. The bin content of each bin is drawn from a Gaussian distribution with mean equal to the nominal bin content and width of the statistical error in that bin. A pseudo-data-set is constructed from this statistically varied template and the other nominal templates, the *Search* region templates are corrected by the *Control* region correction, and the nominal templates are fitted to the corrected templates. This process is repeated 1,000

*Cross-checks are also performed using the *Validation* region in place of the *Control* region.

times and the width of the resulting distribution of μ values gives the standard error from the statistics of the template under consideration.

The statistical errors are shown separately for the signal events and for the combined QCD $W+$ jets and other background templates.

9.2.3 Experimental Uncertainties

Systematic uncertainties originating from detector reconstruction, such as the jet energy scale, are completely correlated across signal and background as well as between electron and muon channels. To quantify their effect on the signal extraction, the 1σ varied templates are summed in each channel to create a pseudo-data sample for each systematic, the templates are corrected using the *Control* region and the nominal templates fitted to the corrected *Search* region templates.

9.2.4 Theoretical Uncertainties

Only the theoretical errors from the EW and QCD $W+2$ jets processes are explicitly considered. The small contribution of the "other backgrounds" in the *Search* region mean we neglect their shape changes due to theoretical uncertainties but theoretical uncertainties on their cross sections are considered. The exception to this is the $t\bar{t}$ background where the above errors are considered as well as replacing the default POWHEG $t\bar{t}$ sample with one produced by MC@NLO. This additional check is carried out as the $t\bar{t}$ background is relatively large and the modelling os such events shows some generator differences.

The errors arising from theoretical sources are uncorrelated between samples, for example, a change in the factorisation scale of the EW signal should not affect the QCD $W+$ jets in any way as the scale is not a physical property but a mathematical artefact representing the lack of higher order corrections to a specific process. Therefore, to gauge the standard errors for each theoretical uncertainty we alter one sample at a time. The bin-to-bin changes are correlated due to changes in normalisation and to the bin migrations of events. The

procedure entails creating pseudo-data for one altered template, applying the *Control* region constraints, and fitting with the nominal templates to find the resulting scale factors.

PDF Uncertainties

PDF errors account for physical differences in the initial state, e.g. a higher percentage of strange quarks in the proton at a certain momentum fraction, so the errors are applied to all samples at once to find the standard error. The default PDF for the EW signal and QCD W + jets background is CT10. The CT10 collaboration provides sets of error PDFs that allow us to find the standard error using a Hessian combination (see Chapter 6). The value of α_S is also varied and the resulting uncertainty is summed in quadrature with the Hessian combination to get the CT10 standard error. As a cross-check the templates are also re-weighted to the NNPDF and MSTW2008NLO nominal PDFs but the resulting difference is observed to be much smaller than the CT10 variations.

9.3 Background Exclusion

To achieve an observation of a new process we must investigate the possibility that background fluctuations led to non-zero signal in the fit. To gauge the probability of this happening, a background exclusion text is performed where pseudo-data is created from the background processes only, and the background and signal templates are fitted to it. In each pseudo-experiment the background templates are gaussian-fluctuated by each systematic uncertainty using a different set of random numbers for each experiment. The resulting templates are then Poisson-fluctuated to mimic the statistical fluctuations in the data. Equation 9.9 shows how the bin content C_i for bin i is altered for a given pseudo-experiment:

$$C_i^{\text{Fluctuated}} = Poisson\left(C_i^{\text{Nominal}} + \sum_{U=0}^{\text{Uncertainties}} R_i^U \delta_i^U\right), \tag{9.9}$$

where U is summed over all uncertainties. R_i^U is the random number for bin i and uncertainty

U; this could be the same or across bins or different samples depending on whether U is correlated or uncorrelated. δ_i^U is the standard error for U in bin i.

Each pseudo-experiment returns a set of scale factors for the templates. The spread in the μ_{EW} values tells us how much signal was fit in the pseudo-data-sets that did not contain any signal. The mean of this distribution is expected to be close to 0 whereas the means for μ_{QCD} and μ_B are expected to be 1, as these templates were included in the pseudo-data construction. To quantify the expected ability of the analysis to exclude the background-only case we count the number of pseudo-experiments that gave a μ_{EW} equal to or greater than the expected value of μ_{EW}, which by definition we expected to equal 1.0. The fraction of pseudo-experiments that return a signal scale factor ≥ 1.0 is designated the p_0 value:

$$p_0 = \frac{N^{\text{Pseudos}}_{\mu_{EW} \geq 1.0}}{N^{\text{Pseudos}}_{Total}}. \tag{9.10}$$

The p_0 value can then be converted into a significance z_0 using:

$$z_0 = \sqrt{2}\text{erf}^{-1}\left(1.0 - p_0\right), \tag{9.11}$$

where erf^{-1} is the inverse error function. Evaluating Equation 9.11 we need $\mathcal{O}\left(10^7\right)$ pseudo-experiments to achieve the 5σ significance required for an observation. Results for the pseudo-experiments are shown in Figure 9.4. The spread of μ_{EW} does not get close to $\mu_{EW} = 1.0$ so we can only say that the background exclusion significance is larger than some value.

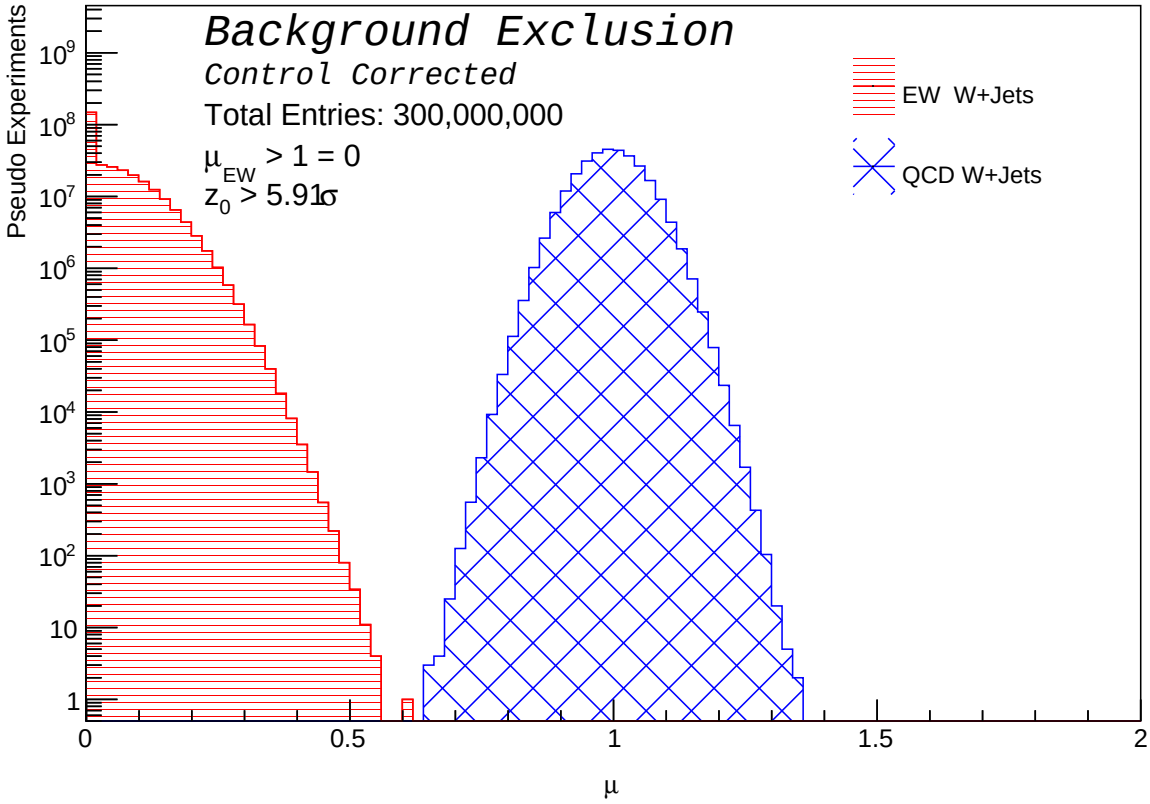


Figure 9.4: μ values for EW signal and QCD background from pseudo-experiments with no signal in the pseudo-data using the *Control* regionto correct and constrain the QCD $W + 2$ jets.

9.4 Signal Measurement

If the EW $W + 2$ jets signal is present in the data, we must quantify the precision with which we are able to measure it. A similar procedure to the background exclusion techniques in Section 9.3 is used, only this time we include the signal in the pseudo-data. This procedure uses only the template predictions (MC and the multijets data-driven sample) and no data is used from the *Search* or *Control* regions.

The effects of all the uncertainties on μ_{EW} are shown in Figure 9.5. The horizontal bands show the change in the μ_{EW} value for each systematic contribution; the quadrature sum of all contributions is shown at the bottom of the figure. The systematic uncertainties when using the unconstrained QCD $W + 2$ jets templates are shown in red. The improvement from constraining using the *Control* region is shown in blue, and using the *Validation* region is shown in green. The values of the bands plotted in Figure 9.5 are explicitly shown in Table 9.2 for the *Control* region case.

Figure 9.5 shows that the smallest total uncertainty on μ_{EW} comes when applying the *Control* region constraint, justifying our use of it as the default region with which to constrain the QCD $W +$ jets templates. The same uncertainties are shown in Figure 9.6 and Table 9.3 for their effects on μ_{QCD}. Again the smallest total uncertainties occur in the case of the *Control* region constraint.

To ensure we are correctly considering the correlations between the uncertainties a cross-check of the total uncertainty is performed with pseudo-experiments by allowing simultaneous variations of all uncertainties. The pseudo-data-set is constructed starting from the nominal templates[†] and adding a systematic shift for each uncertainty to the bin content of the templates. The same gaussian random numbers are used across bins and templates where appropriate to account for correlated uncertainties.

In this manner, a pseudo-data-set is created which represents the templates varied about all systematics. The nominal templates are then fit to this pseudo-data-set and the scale

[†]This time including the signal contribution.

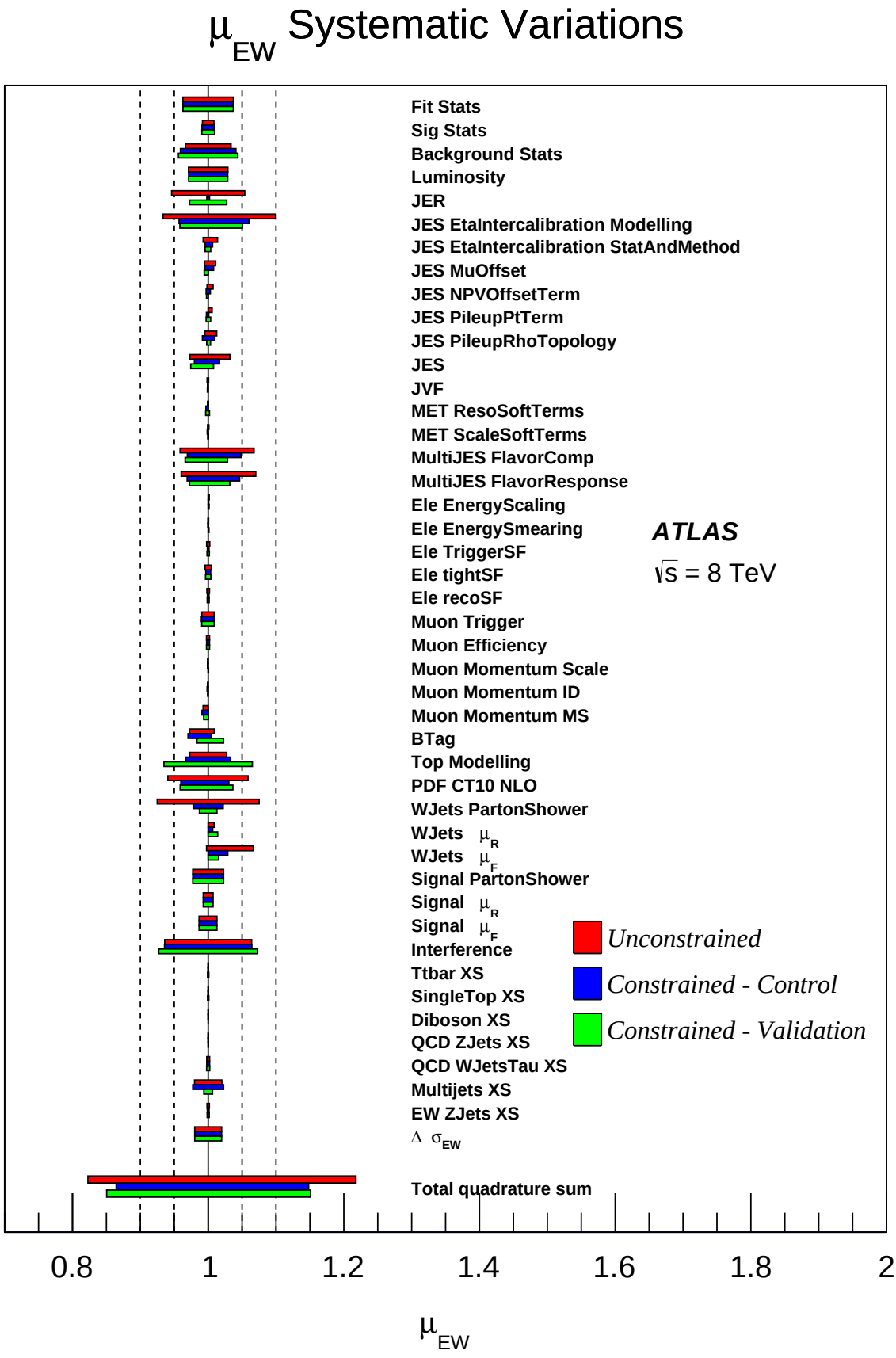


Figure 9.5: Expected uncertainties on μ_{EW} from detector and theoretical effects. The unconstrained case is compared to the constraints from the *Control* and *Validation* regions.

EW Systematic Variations		
Systematic	Up variation	Down variation
Fit Stats	±3.72	
Sig Stats	±0.91	
Background Stats	±4.08	
Luminosity	±2.88	
Interference	±6.44	
JER	±0.20	
JES EtaIntercalibration Modelling	6.03	-4.29
JES EtaIntercalibration StatAndMethod	0.65	-0.44
JES MuOffset	-0.50	0.80
JES NPVOffsetTerm	-0.30	0.32
JES PileupPtTerm	0.08	-0.26
JES PileupRhoTopology	0.98	-0.85
JES	1.67	-2.03
JVF	0.00	-0.09
MET ResoSoftTerms	-0.33	0.01
MET ScaleSoftTerms	-0.14	0.01
MultiJES FlavorComp	4.84	-3.07
MultiJES FlavorResponse	-3.10	4.63
Ele EnergyScaling	0.00	0.08
Ele EnergySmearing	-0.05	0.01
Ele TriggerSF	0.12	-0.12
Ele tightSF	0.36	-0.35
Ele recoSF	0.12	-0.12
Muon Trigger	0.94	-1.00
Muon Efficiency	0.19	-0.23
Muon Momentum Scale	0.00	-0.10
Muon Momentum ID	0.00	-0.16
Muon Momentum MS	-0.91	0.02
BTag	-2.97	0.43
PDF CT10 NLO	3.05	-4.03
Top Modelling	±3.32	
Ttbar XS	±0.04	
SingleTop XS	±0.04	
Diboson XS	±0.03	
QCD ZJets XS	±0.03	
QCD WJetsTau XS	±0.21	
Multijets XS	±2.27	
EW ZJets XS	±0.13	
WJets PartonShower	±2.21	
WJets μ_R	0.00	0.67
WJets μ_F	0.00	2.89
Signal PartonShower	±2.27	
Signal μ_R	-0.73	0.72
Signal μ_F	-1.34	1.29
$\Delta\sigma_{EW}$	-2.10	2.00
Total quadrature sum	+14.81, −13.56	

Table 9.2: Expected uncertainty effects on μ_{EW} using *Control* region constraints.

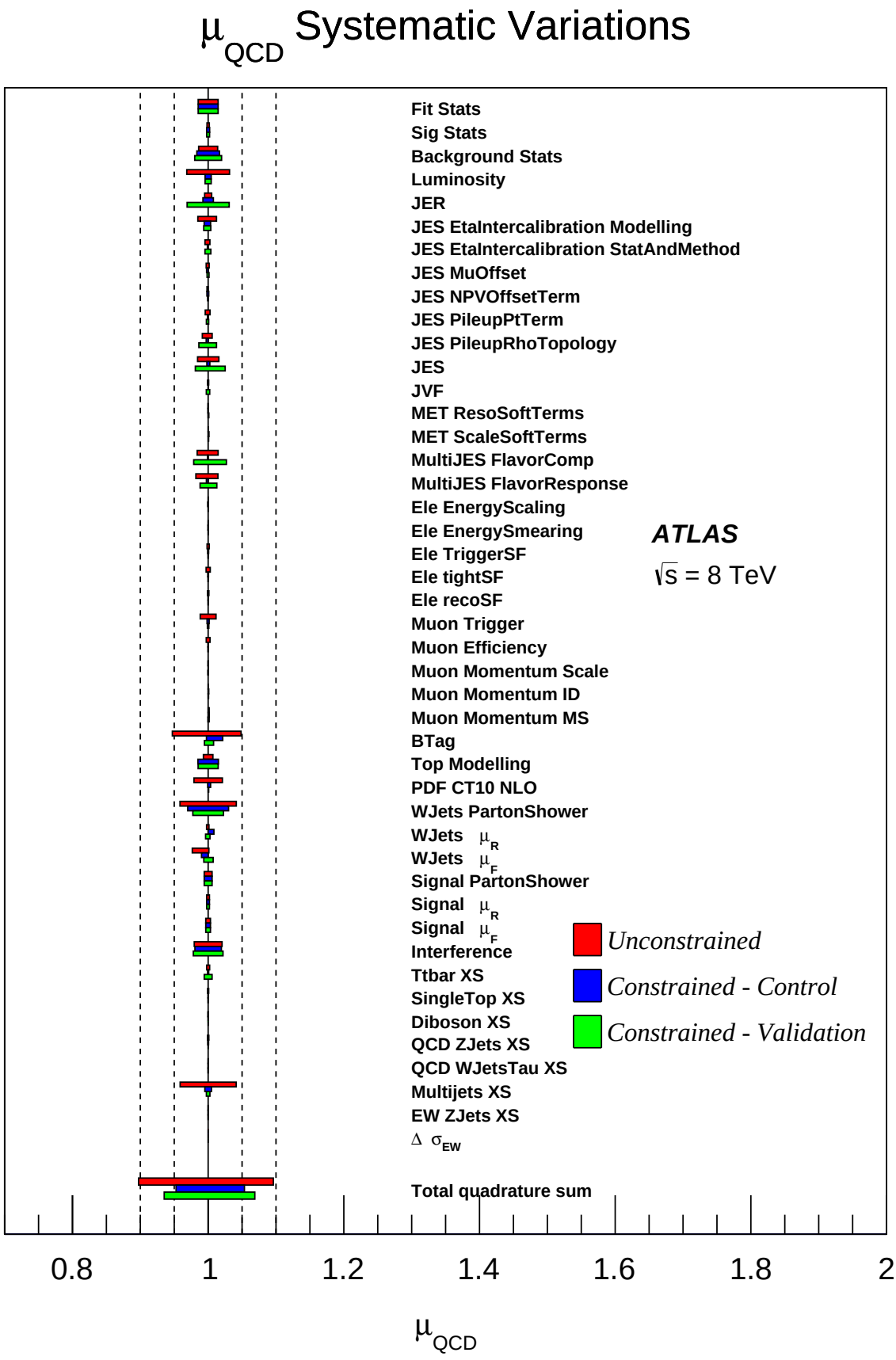


Figure 9.6: Uncertainties on μ_{QCD} from detector and theoretical effects. The unconstrained case is compared to the constraints from the *Control* and *Validation* regions.

QCD Systematic Variations		
Systematic	Up variation	Down variation
Fit Stats	±1.47	
Sig Stats	±0.22	
Background Stats	±1.72	
Luminosity	±0.45	
Interference	±1.95	
JER	±0.77	
JES EtaIntercalibration Modelling	-0.59	0.36
JES EtaIntercalibration StatAndMethod	-0.11	0.00
JES MuOffset	0.03	-0.23
JES NPVOffsetTerm	0.09	-0.18
JES PileupPtTerm	0.02	-0.09
JES PileupRhoTopology	0.00	-0.29
JES	0.25	-0.19
JVF	0.00	0.01
MET ResoSoftTerms	-0.02	0.00
MET ScaleSoftTerms	0.12	0.00
MultiJES FlavorComp	0.00	-0.08
MultiJES FlavorResponse	0.00	-0.26
Ele EnergyScaling	0.00	-0.05
Ele EnergySmearing	0.02	-0.03
Ele TriggerSF	-0.07	0.07
Ele tightSF	-0.05	0.05
Ele recoSF	-0.01	0.01
Muon Trigger	0.15	-0.16
Muon Efficiency	0.01	-0.02
Muon Momentum Scale	0.00	0.03
Muon Momentum ID	0.00	0.09
Muon Momentum MS	0.16	0.00
BTag	-0.25	2.15
PDF CT10 NLO	-0.06	0.36
Top Modelling	±1.51	
Ttbar XS	±0.08	
SingleTop XS	±0.02	
Diboson XS	±0.01	
QCD ZJets XS	±0.06	
QCD WJetsTau XS	±0.01	
Multijets XS	±0.49	
EW ZJets XS	±0.01	
WJets PartonShower	±3.01	
WJets μ_R	0.00	0.85
WJets μ_F	0.00	-0.98
Signal PartonShower	±0.58	
Signal μ_R	0.19	-0.21
Signal μ_F	0.34	-0.35
Total quadrature sum	$+6.20, -5.67$	

Table 9.3: Expected uncertainty effects on μ_{QCD} using *Control* region constraints.

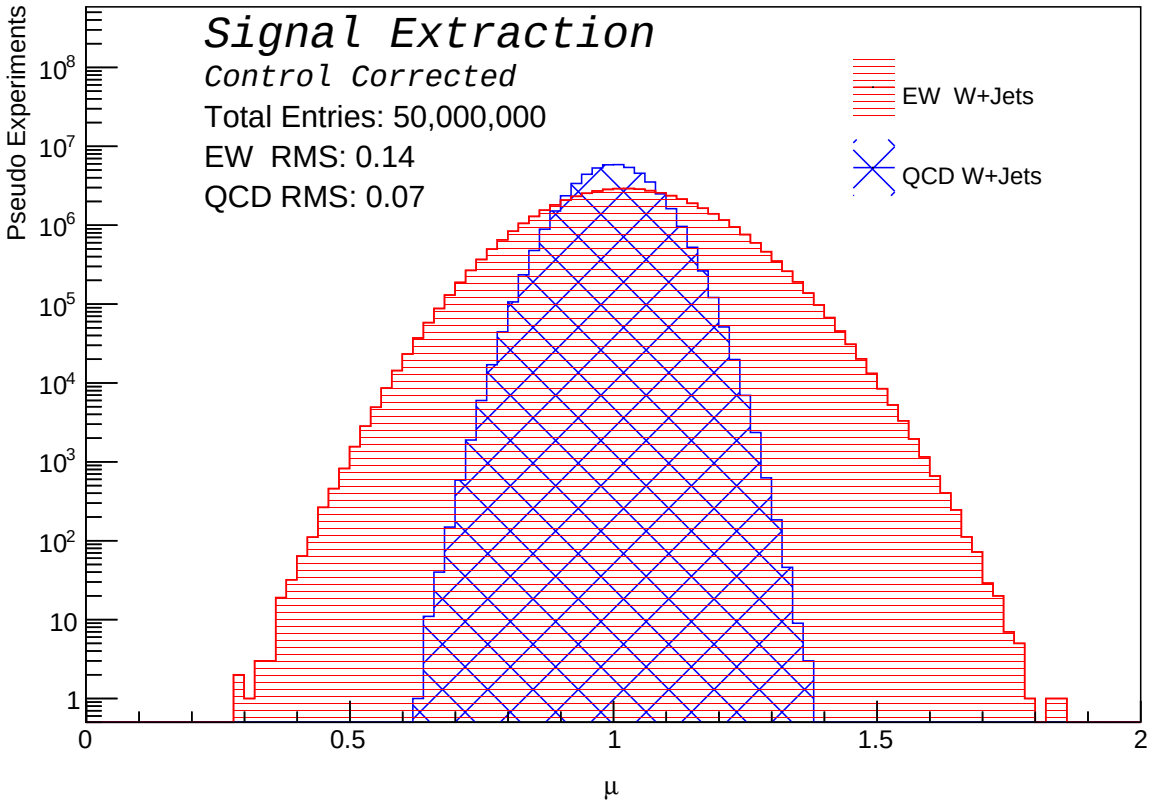


Figure 9.7: μ values for EW signal and QCD background from pseudo-experiments using the *Control* region.

factors are recorded. This procedure is repeated a large number of times to build up distributions of both μ_{EW} and μ_{QCD}. The widths of these distributions are the 1σ error on each scale factor and are equal to the precision to which the analysis can measure each process. Figure 9.7 shows the μ_{EW} and μ_{QCD} distributions from these pseudo-experiments when using the constraints from the *Control* region. The total uncertainties from Tables 9.2 and 9.3 can be seen to agree with the widths (RMSs) of the μ distributions in Figure 9.7.

Table 9.4 shows the final uncertainty on the μ_{EW} measurement from the various fits. The

Fit Channel / Constraint Region	Both	Electron	Muon
Control	13.7	15.7	14.1
Validation	14.7	16.8	15.2
No constraint	19.1	21.3	19.2

Table 9.4: Uncertainties on extracted μ_{EW}, derived from the pseudo-experiment procedure. The various QCD $W + 2$ jets constraints are shown, as well as the errors from fitting in both channels compared to a single lepton channel.

Control region constrained case shows the smallest measurement error.

9.5 Signal Strength Measurement

The nominal M_{jj} templates are fitted to the ATLAS data to measure the observed contribution from the EW $W + 2$ jets and QCD $W + 2$ jets processes. The results of the fit using the *Control* region constraint are shown in Figure 9.8, the *Validation* region constraint in Figure 9.9, and no constraint in Figure 9.10.

Table 9.5 shows the scale factors returned from all the fits. In each case a metric of the fit quality is given by the χ^2 divided by the number of degrees of freedom (NDF) in the fit, where χ^2 is calculated as in Equation 5.1.

The fit was performed in all constraint cases and for the lepton channels individually and combined. The μ_{EW} values and their errors (from Table 9.4) are shown in Figure 9.11. The corresponding set of μ_{QCD} observed values are shown in Figure 9.12. The extracted signal yields are consistent within uncertainties for all cases.

9.6 Alternative Signal Sample

To provide a cross-check to the μ_{EW} results found in this chapter, the POWHEG signal sample was replaced with the SHERPA generated samples and the procedure repeated. The SHERPA samples are at LO and so have larger theoretical errors than the NLO POWHEG samples. Table 9.6 shows the fit results using the SHERPA signal samples.

When comparing the $N_{EW}^{\text{After fit}}$ we see they are similar between Tables 9.5 and 9.6 demonstrating the POWHEG and SHERPA signal sample fits give similar results. The difference between the POWHEG and SHERPA *Control* region constrained results for $N_{EW}^{\text{After fit}}$ is $< 4\%$ which is well within the 1σ error of the measurement. The μ_{EW} values for the *Control* region constrained fits across both channels differ between the tables as SHERPA predicts fewer sig-

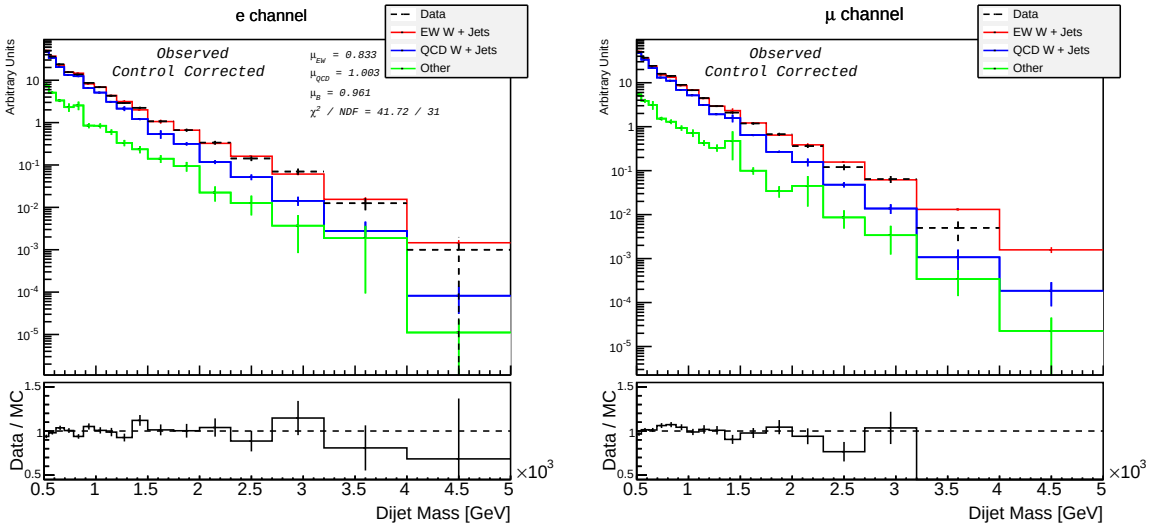


Figure 9.8: The results of the fit to ATLAS data when constraining the QCD $W + 2$ jets distributions using the *Control* region. The fit is performed in both channels simultaneously and the μ values are required to be the same across both channels.

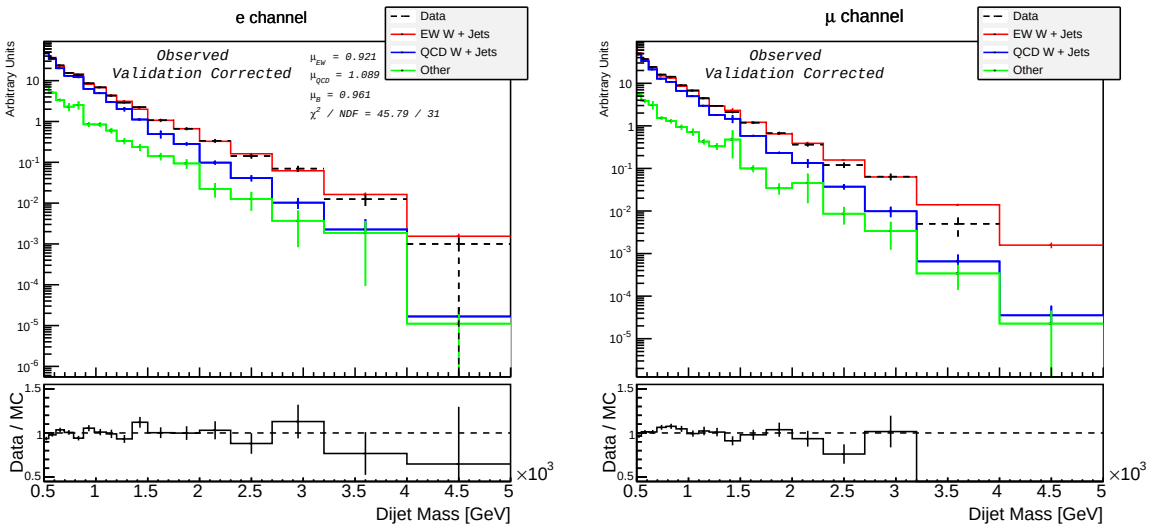


Figure 9.9: The results of the fit to ATLAS data when constraining the QCD $W + 2$ jets using the *Validation* region distributions.

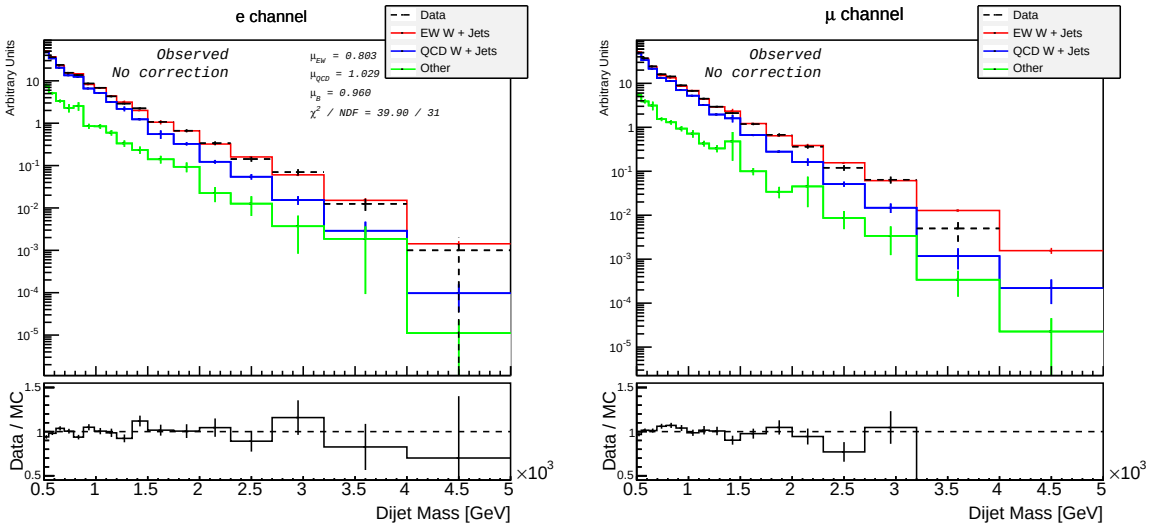


Figure 9.10: The results of the fit to ATLAS data when not constraining the QCD $W+2$ jets distributions.

Constraint Region	Channel	μ_{EW}	N_{EW}^{Fit}	μ_{QCD}	μ_B	$\frac{\chi^2}{NDF}$
Control	**Both**	**0.833**	**4,183**	**1.003**	**0.961**	**41.7 / 31**
	Electron	0.868	2,1780	1.020	0.989	18.2 / 14
	Muon	0.798	2,003	0.982	0.989	18.5 / 14
Validation	**Both**	**0.922**	**4,628**	**1.089**	**0.961**	**45.8 / 31**
	Electron	0.944	2,371	1.056	0.988	15.4 / 14
	Muon	0.896	2,250	1.102	0.991	23.8 / 14
No constraint	**Both**	**0.803**	**4,031**	**1.029**	**0.960**	**39.9 / 31**
	Electron	0.861	2,162	0.990	0.988	15.5 / 14
	Muon	0.744	1,867	1.056	0.989	18.1 / 14

Table 9.5: Fit results from various regions used to constrain the QCD $W+2$ jets background. The χ^2/NDF is shown to give a metric of the fit quality.

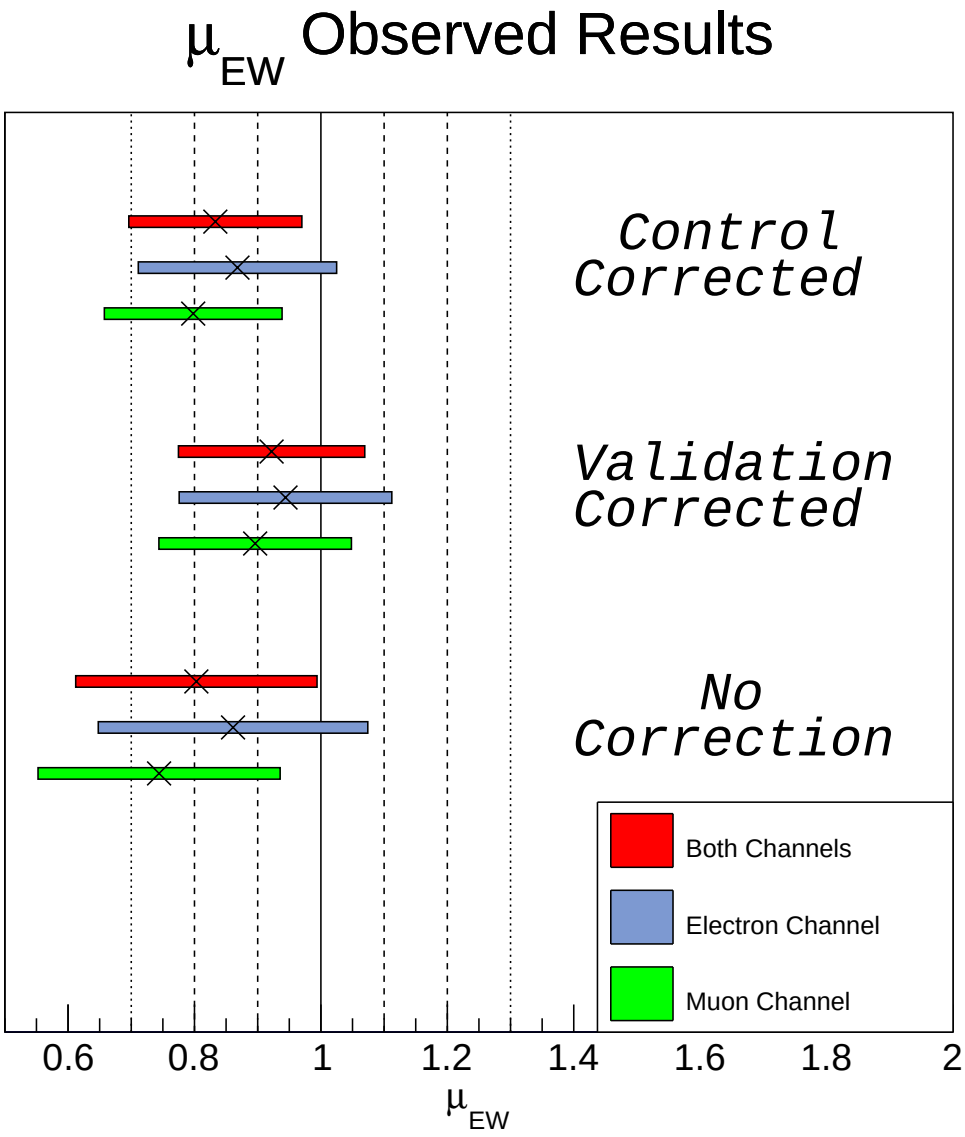


Figure 9.11: The results of the fits to ATLAS data for μ_{EW}. The single channel fits are compared to simultaneous fits in both channels. Constraining the QCD $W + 2$ jets distribution can be seen to reduce the errors.

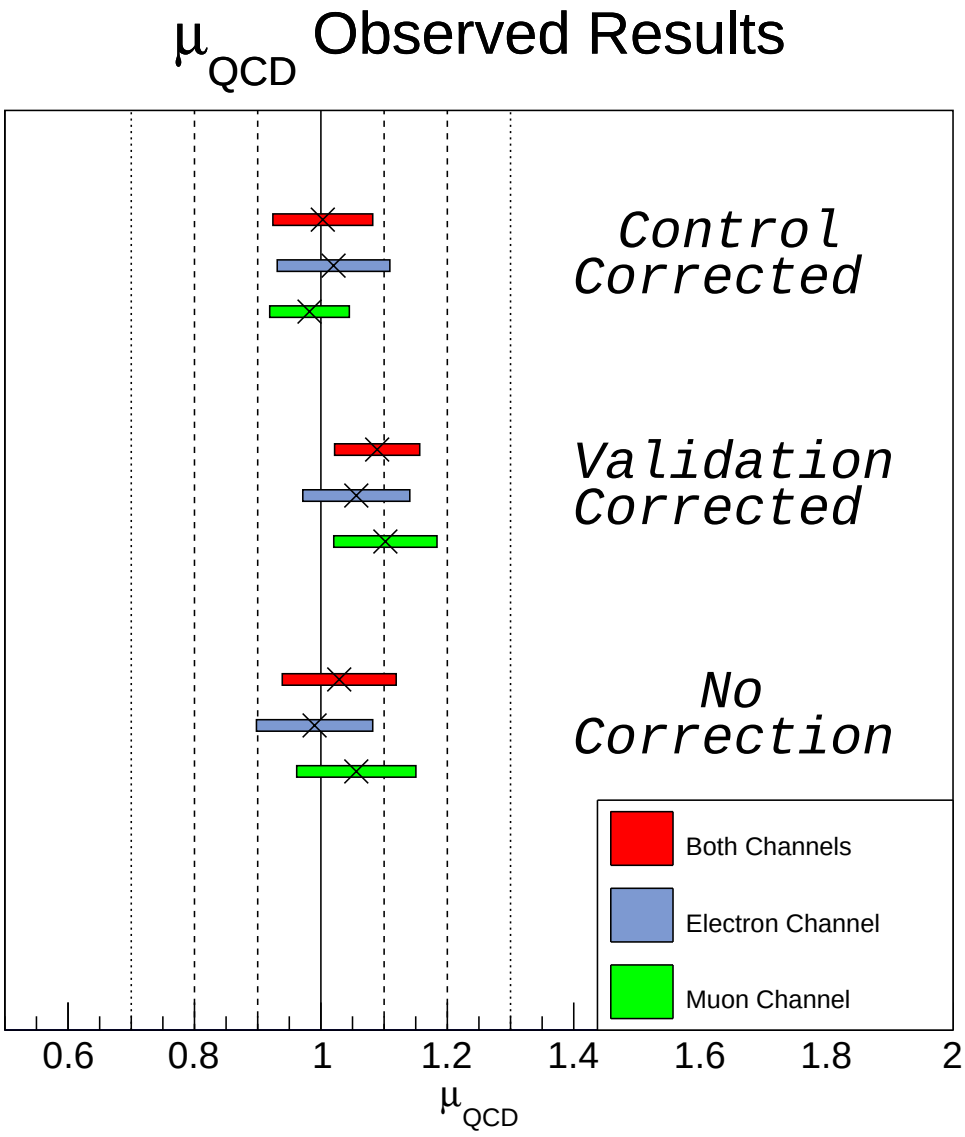


Figure 9.12: The results of the fits to ATLAS data for μ_{QCD}. The single channel fits are compared to simultaneous fits in both channels. Constraining the QCD $W + 2$ jets distribution can be seen to reduce the errors.

Constraint Region	Channel	μ_{EW}	N_{EW}^{Fit}	μ_{QCD}	μ_B	$\frac{\chi^2}{NDF}$
Control	**Both**	**0.962**	**4,345**	**0.988**	**0.959**	**39.5 / 31**
	Electron	1.010	2,281	1.002	0.988	17.2 / 14
	Muon	0.912	2,059	0.970	0.987	17.1 / 14
Validation	**Both**	**1.058**	**4,782**	**1.068**	**0.959**	**42.4 / 31**
	Electron	1.086	2,452	1.033	0.987	16.1 / 14
	Muon	1.027	2,318	1.082	0.989	20.7 / 14
No constraint	**Both**	**0.951**	**4,296**	**1.015**	**0.959**	**39.1 / 31**
	Electron	1.017	2,297	0.976	0.988	16.0 / 14
	Muon	0.883	1,995	1.042	0.988	17.2 / 14

Table 9.6: Fit results when using the SHERPA signal samples. The χ^2/NDF is shown to give a metric of the fit quality.

nal events (4,517, with $\mu_{EW} = 1$) than POWHEG (5,022). This difference is caused by the different orders of the samples, the SHERPA being LO and the POWHEG is NLO.

The errors on the SHERPA fits are $+18.6, -16.4\%$ for the *Control* region constrained fits in both lepton channels.

Chapter 10

Results

This chapter discusses the results of the analysis and shows the effect of the normalisation factors extracted from the fit for the signal yield in Section 10.1. The results are shown in a high EW $W + 2$ jets purity *Confirmation* region in Section 10.2. The cross section for EW $W + 2$ jets is then calculated in Section 10.3. The results are put in context in Section 10.4.

10.1 Post-fit Distributions

The extracted μ_{EW} and μ_{QCD} from Chapter 9 are applied to the reconstructed distributions. The QCD $W + 2$ jets MC events are re-weighted using the *Control* region M_{jj} correction function as discussed in Chapter 8. The uncertainties are on the M_{jj} distribution are also constrained by this procedure.

Figure 10.1 shows jet distributions for the *Search* region. The distributions are altered from those shown in Figure 7.12 by the application of the EW and QCD $W + 2$ jets scale factors extracted from the fit in Chapter 9. Good data-to-MC agreement is observed in all distributions. Appendix C shows χ^2/NDF values for the data-prediction comparison and how these are affected by the fit scale factors.

Figure 10.2 shows good agreement is also observed for the lepton-based distributions.

The lepton and W-boson p_{T} distributions show the same disagreement that was observed in the pre-fit distributions in Figure 7.13. The lepton centrality in Figure 10.2(e) shows good agreement between data and the prediction in a region with small uncertainties and relatively high signal fraction.

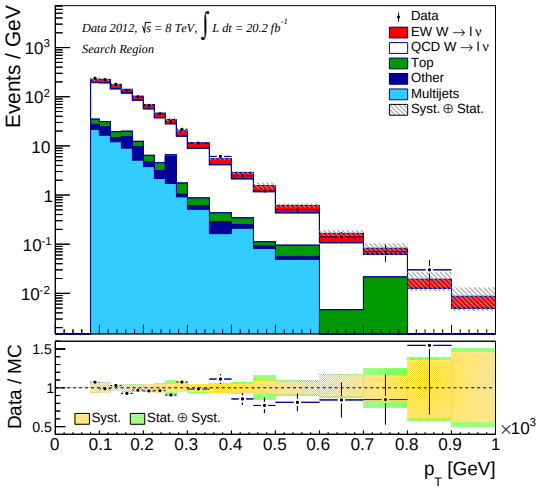
The remaining distributions shown in Figure 10.3 show good modelling in the high-statistics regions. The M_{jj} distribution in Figure 10.3(a) shows markedly improved data-MC agreement in the tail of the distribution as a result of the signal fit scaling down the EW signal contribution. The N_{Jets} plot in Figure 10.3(e) shows an overestimation of the data for high jet multiplicities, as it is not significantly affected by the fit. The individual lepton channel distributions for the variables sensitive to EW $W + 2$ jets production, M_{jj} and $\mathcal{C}_{Lepton}$, are shown in Figures 10.4 and 10.5.

The post-fit yield tables are shown in Tables 10.1, 10.2, and 10.3 for the electron, muon, and combined channels respectively. The tables differ to those shown in the Section 7.4.1 by the 2nd order polynominal applied to the QCD $W + 2$ jets background and the fit scale factors applied to the EW and QCD $W + 2$ jets processes.

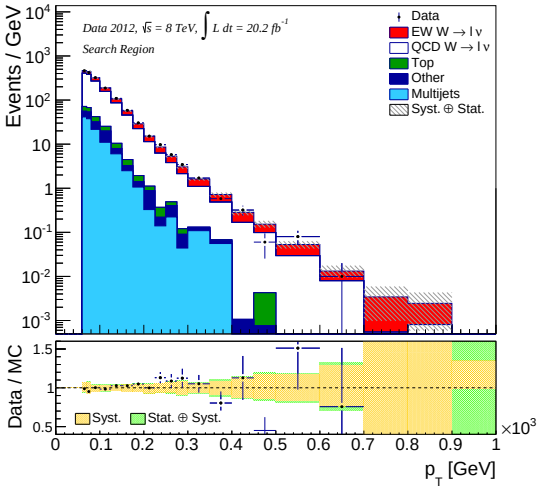
10.2 Confirmation Region

In order to compare the analysis results to data with minimal contribution from background processes, an additional phase space—the *Confirmation* region—is defined which isolates the EW $W + 2$ jets process and maximises its purity. The *Confirmation* region is exactly equal to that of the *Search* region (see Table 7.1) except with the M_{jj} cut increased to $M_{jj} > 1.5$ TeV. This additional cut allows us to view distributions with the EW $W + 2$ jets signal dominating as it now contributes 56% of the events (see Figure 10.6). The *Confirmation* region event yields are shown in Table 10.4.

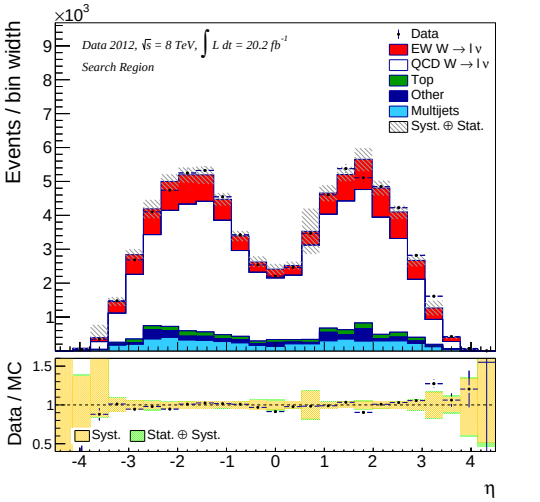
The *Confirmation* region has fewer events than the *Search* region so the errors are increased but still small enough to test the agreement between data and the prediction. The M_{jj} spectrum is not fit in this phase space because there is no QCD $W + 2$ jets dominated



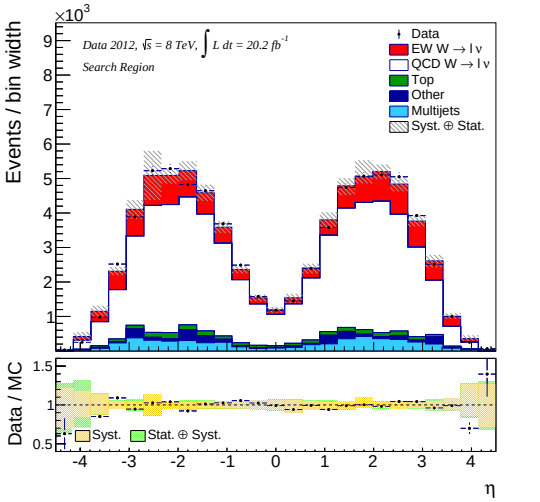
(a) Leading Jet p_{T}



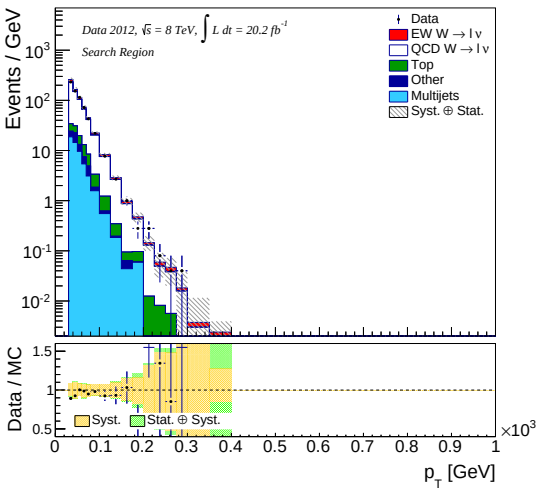
(b) Sub-leading jet p_{T}



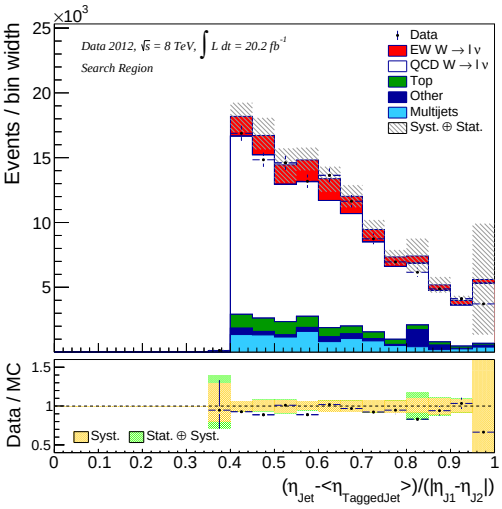
(c) Leading Jet η



(d) Sub-Leading Jet η

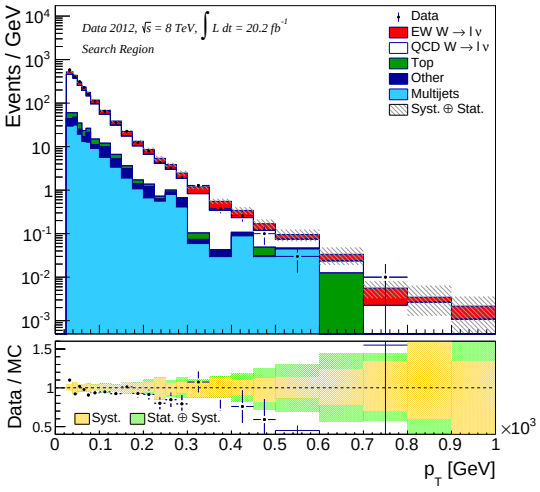


(e) Third jet p_{T}

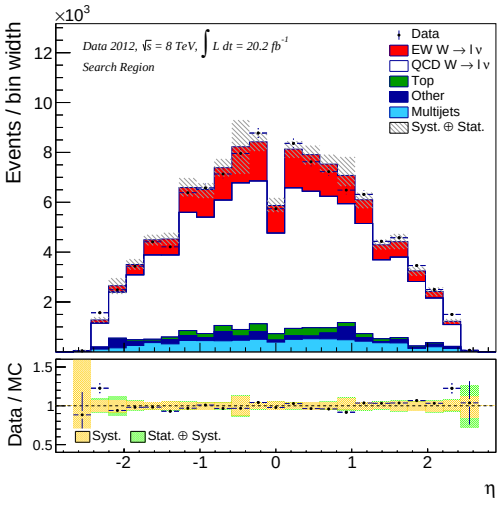


(f) Jet Centrality

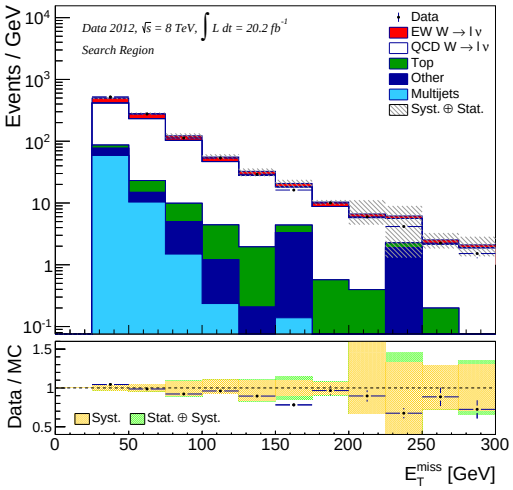
Figure 10.1: Reconstruction-level jet distributions summed over lepton channels in the *Search* region. The EW signal and QCD $W + 2$ jets processes have been scaled by the fit normalisation and the latter also has the *Control* region correction.



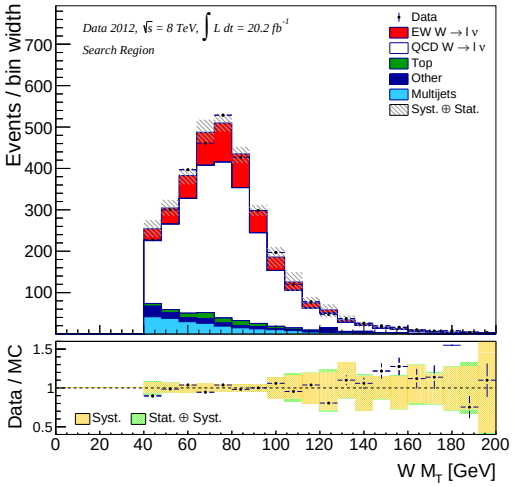
(a) Lepton p_{T}



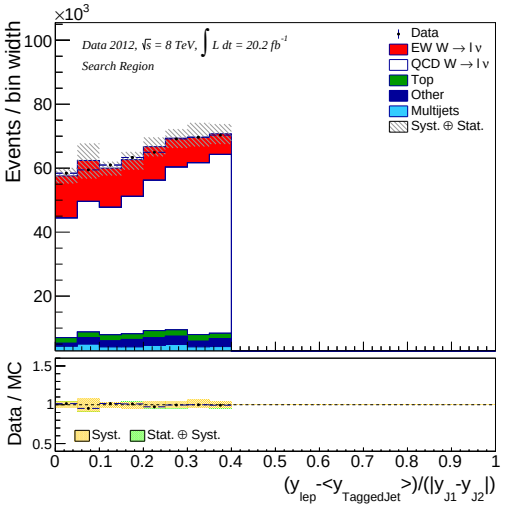
(b) Lepton η



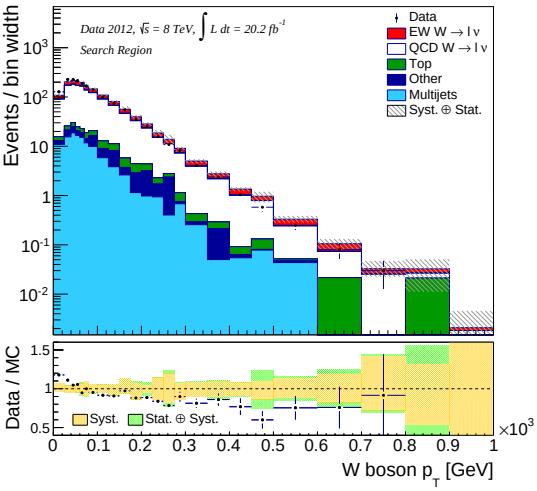
(c) $E_{\mathrm{T}}^{\mathrm{miss}}$



(d) W-boson transverse mass

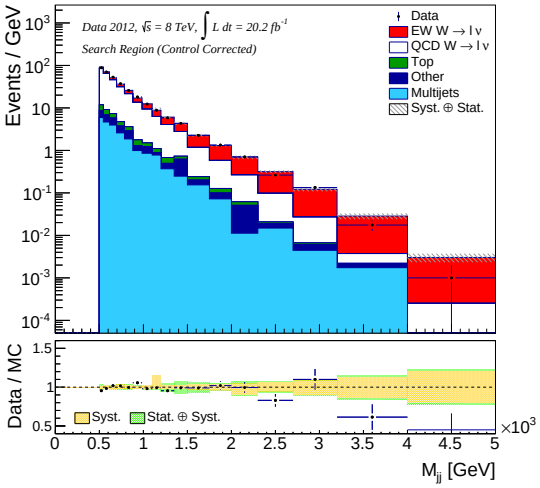


(e) Lepton Centrality

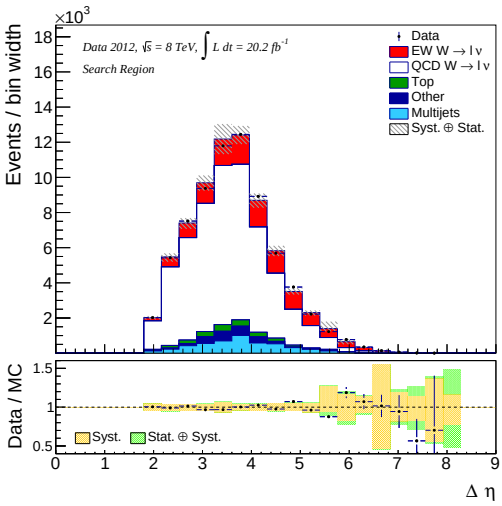


(f) W-boson p_{T}

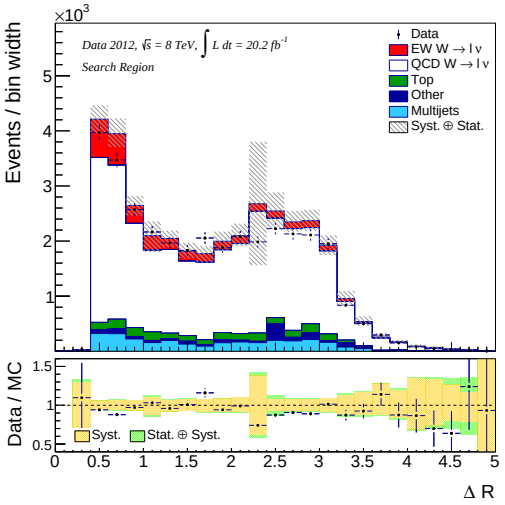
Figure 10.2: Reconstruction-level lepton distributions summed over lepton channels in the *Search* region. The EW signal and QCD $W + 2$ jets processes have been scaled by the fit normalisation and the latter also has the *Control* region correction.



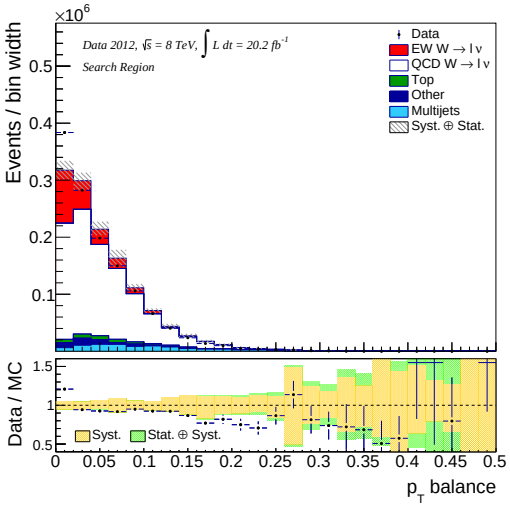
(a) M_{jj}



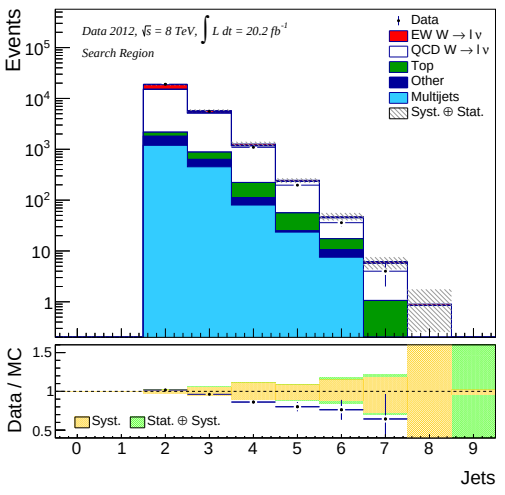
(b) $\Delta\eta(J_1, J_2)$



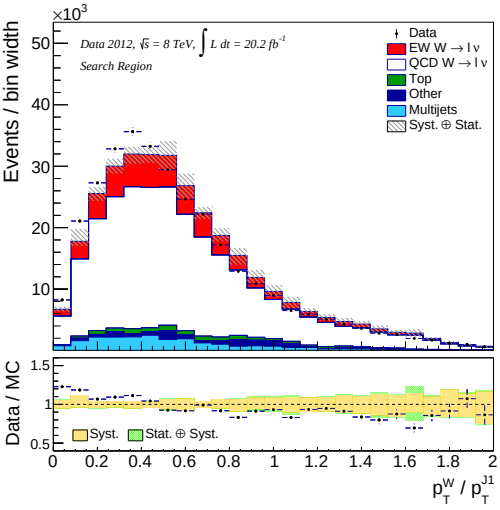
(c) Minimum ΔR (J_3, J_{Tagged})



(d) p_{T} balance

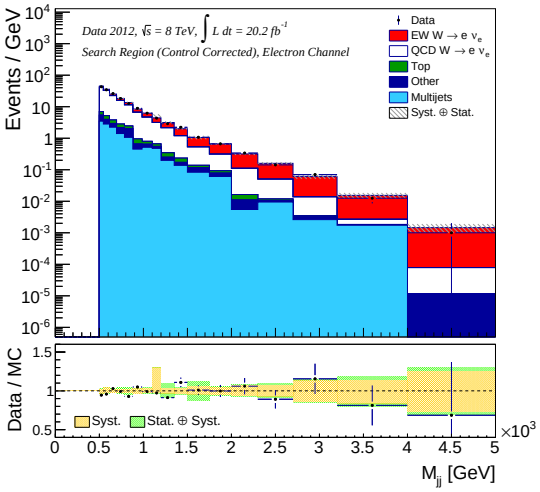


(e) Number of Jets

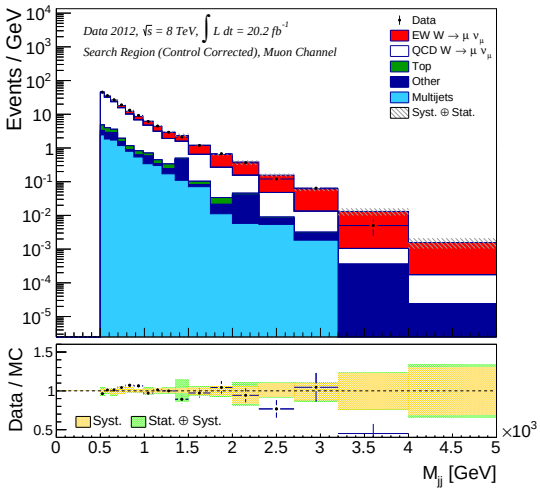


(f) p_T^W / p_T^{J1}

Figure 10.3: Reconstruction-level distributions summed over lepton channels in the *Search* region. The EW signal and QCD $W + 2$ jets processes have been scaled by the fit normalisation and the latter also has the *Control* region correction. The M_{jj} systematic uncertainties have been constrained.

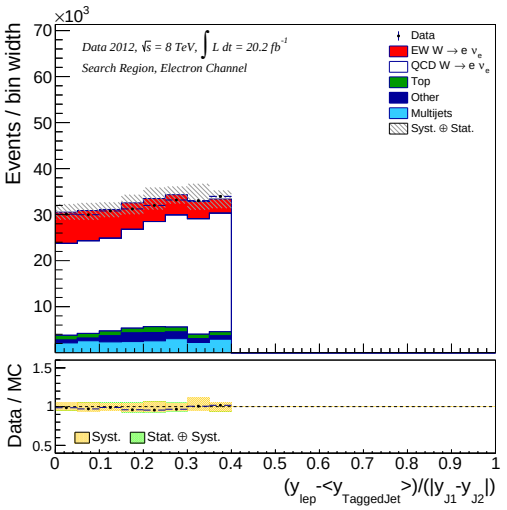


(a) Electron channel M_{jj}

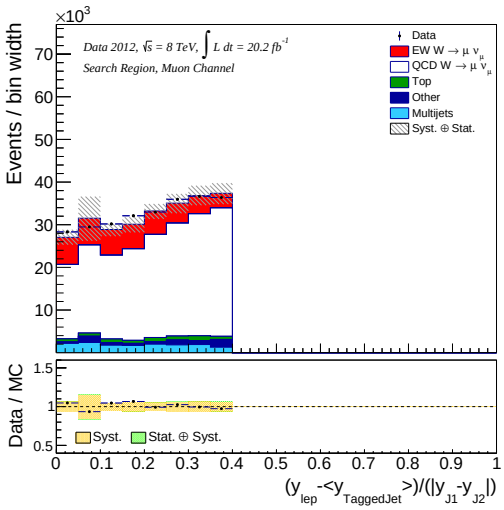


(b) Muon channel M_{jj}

Figure 10.4: The M_{jj} distributions for the individual lepton channels. The EW signal and QCD $W + 2$ jets processes have been scaled by the fit normalisation and the latter also has the *Control* region correction and uncertainty constraints applied.



(a) Electron channel $|\mathcal{C}_{Lepton}|$



(b) Muon channel $|\mathcal{C}_{Lepton}|$

Figure 10.5: The $|\mathcal{C}_{Lepton}|$ distributions for the individual lepton channels. The EW signal and QCD $W + 2$ jets processes have been scaled by the fit normalisation and the latter also has the *Control* region correction.

Process \ Region	Inclusive	High p_T	High mass	Search	Control	Validation
QCD $W(\to e\nu)$ + 2 jets	1,915,542	269,298	28,184	8,998	9,133	4,985
QCD $W(\to \tau\nu)$ + 2 jets	32,260	2,602	367	142	156	9
ZZ	192	26	3	1	1	1
WZ	2,188	279	30	13	4	9
WW	9,615	1,037	103	49	20	24
Multijets	240,967	28,974	3,437	1,008	1,163	605
QCD $Z(\to e^+e^-)$ + 2 jets	98,697	8,192	608	215	154	132
QCD $Z(\to \tau^+\tau^-)$ + 2 jets	4,277	387	32	6	18	7
$t\bar{t}$	36,576	11,998	865	210	114	398
Single top	13,533	2,946	400	211	77	84
EW $Z(\to e^+e^-)$ + 2 jets	249	134	54	28	9	14
EW $W(\to e\nu)$ + 2 jets	11,774	5,661	2,811	2,075	470	211
Data	2,187,780	318,634	35,269	12,714	10,916	6,015
Total Prediction	2,365,869	331,535	36,893	12,956	11,320	6,480
Data / Prediction	0.92	0.96	0.96	0.98	0.96	0.93
Signal / Total Prediction	0	0.02	0.08	0.16	0.04	0.03
Signal / $\sqrt{\text{Background}}$	7.67	9.92	15.23	19.89	4.51	2.67

Table 10.1: Scaled event yields for the electron channel. The EW and QCD W + 2 jets have been scaled by the fit results and the QCD W + 2 jets has been corrected to data using the *Control* region.

Process \ Region	Inclusive	High p_T	High mass	Search	Control	Validation
QCD $W(\to \mu\nu)$ + 2 jets	2,145,823	290,922	31,448	9,443	10,925	5,133
QCD $W(\to \tau\nu)$ + 2 jets	41,243	3,370	234	111	0	28
ZZ	182	18	1	1	0	1
WZ	2,571	294	31	15	5	7
WW	11,634	1,197	124	54	26	31
Multijets	127,408	13,469	1,708	751	308	466
QCD $Z(\to \mu^+\mu^-)$ + 2 jets	50,601	3,453	411	127	137	93
QCD $Z(\to \tau^+\tau^-)$ + 2 jets	6,077	429	37	21	0	10
$t\bar{t}$	38,785	12,672	877	195	123	377
Single top	14,654	3,032	361	148	88	95
EW $Z(\to \mu^+\mu^-)$ + 2 jets	262	125	54	35	13	5
EW $W(\to \mu\nu)$ + 2 jets	12,853	6,020	2,940	2,100	559	220
Data	2,392,305	337,791	37,799	13,101	12,709	5,956
Total Prediction	2,452,093	335,000	38,226	13,001	12,183	6,465
Data / Prediction	0.98	1.01	0.99	1.01	1.04	0.92
Signal / Total Prediction	0.01	0.02	0.08	0.16	0.05	0.03
Signal / $\sqrt{\text{Background}}$	8.23	10.5	15.65	20.11	5.18	2.78

Table 10.2: Scaled event yields for the muon channel. The EW and QCD W + 2 jets have been scaled by the fit results and the QCD W + 2 jets has been corrected to data using the *Control* region.

Process \ Region	Inclusive	High p_T	High mass	Search	Control	Validation
QCD $W(\rightarrow l\nu)$ + 2 jets	4,061,365	560,220	59,632	18,441	20,058	10,118
QCD $W(\rightarrow \tau\nu)$ + 2 jets	73,503	5,972	601	253	156	37
ZZ	374	44	4	2	1	2
WZ	4,759	573	61	28	9	16
WW	21,249	2,234	227	103	46	55
Multijets	368,375	42,443	5,145	1,759	1,471	1,071
QCD $Z(\rightarrow l^+l^-)$ + 2 jets	149,298	11,645	1,019	342	291	225
QCD $Z(\rightarrow \tau^+\tau^-)$ + 2 jets	10,354	816	69	27	18	17
$t\bar{t}$	75,361	24,670	1,742	405	237	775
Single top	28,187	5,978	761	359	165	179
EW $Z(\rightarrow l^+l^-)$ + 2 jets	511	259	108	63	22	19
EW $W(\rightarrow l\nu)$ + 2 jets	24,627	11,681	5,751	4,175	1,029	431
Data	4,580,085	656,425	73,068	25,815	23,625	11,971
Total Prediction	4,817,963	666,535	75,120	25,957	23,503	12,945
Data / Prediction	0.95	0.98	0.97	0.99	1.01	0.92
Signal / Total Prediction	0.01	0.02	0.08	0.16	0.04	0.03
Signal / $\sqrt{\text{Background}}$	11.25	14.43	21.84	28.29	6.86	3.85

Table 10.3: Scaled event yields for the both lepton channels combined. The EW and QCD W + 2 jets have been scaled by the fit results and the QCD W + 2 jets has been corrected to data using the *Control* region.

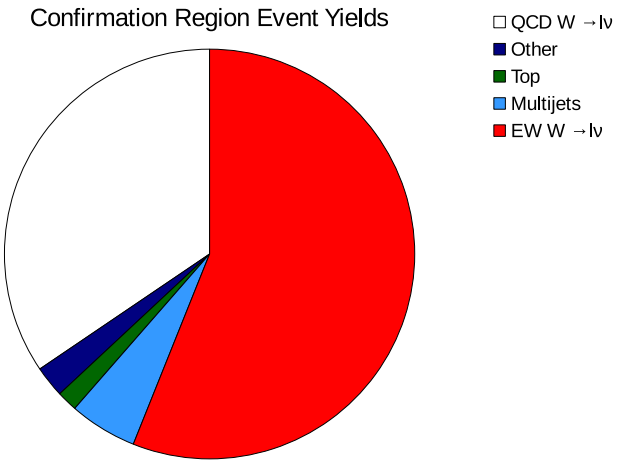


Figure 10.6: Contributions to the *Confirmation* region from different physics processes.

Process \ Channel	Electron	Muon	Both
QCD $W(\to l\nu)$ + 2 jets	205	250	455
QCD $W(\to \tau\nu)$ + 2 jets	0	0	0
ZZ	0	0	0
WZ	1	1	2
WW	2	2	4
Multijets	44	25	69
QCD $Z(\to l^+l^-)$ + 2 jets	9	9	18
QCD $Z(\to \tau^+\tau^-)$ + 2 jets	0	0	0
$t\bar{t}$	6	4	10
Single top	5	6	11
EW $Z(\to l^+l^-)$ + 2 jets	4	5	9
EW $W(\to l\nu)$ + 2 jets	358	383	741
Data	637	657	1,294
Total Prediction	633	686	1,319
Data / Prediction	1.01	0.96	0.98
Signal / Total Prediction	0.57	0.56	0.56
Signal / $\sqrt{\text{Background}}$	21.57	22.03	30.82

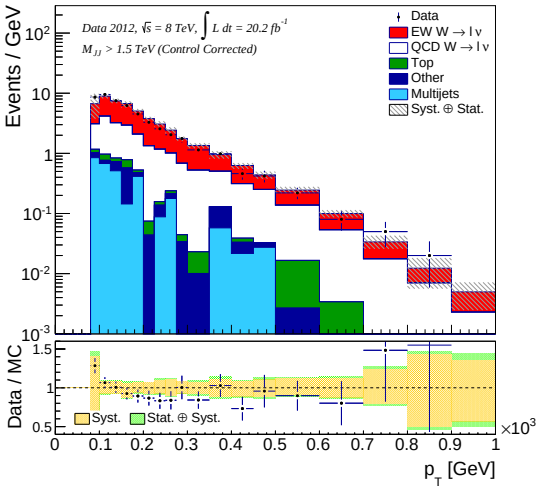
Table 10.4: Event yields for the *Confirmation* region. The QCD W + 2 jets are corrected to the *Control* region fits and the μ_{QCD} from the signal extraction is applied. The EW W + 2 jets process has been scaled by $\mu_{EW} = 0.833$ from the signal extraction.

region to constrain its normalisation resulting in $\Delta\mu_{QCD} = 40\%$. However, if the QCD W + 2 jets contribution is constrained to the normalisation extracted from the *Search* region fits, a fit in the *Confirmation* region returns $\mu_{EW} = 0.790 \pm 0.24$. The fit results are therefore consistent between the *Search* region and the *Confirmation* region as expected as one is a sub-set of the other. The possibility remains that both regions are biased in the same manner or a consistent systematic offset affects the two. The detector-level distributions are shown for the *Confirmation* region for the jet variables in Figure 10.7, the lepton variables in Figure 10.8, and the remaining distributions of interest in Figure 10.9.

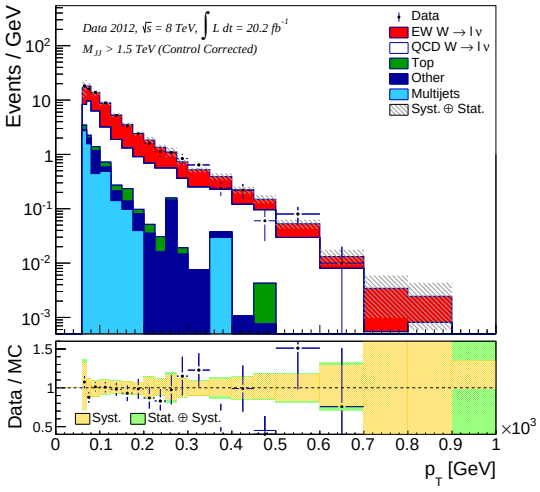
The jet variables shown in Figure 10.7 show that after the high M_{jj} cut we see good agreement between data and prediction. The leading jet p_{T} shows some slope in the ratio plot, where the data starts above the MC at low p_{T} and then goes below at about 150 GeV, although this disagreement in almost always within the uncertainties. The remaining jet distributions show agreement within errors. Figure 10.7(e) shows the number of jets within the η gap formed between the two tagged jets. For this high M_{jj} cut we see very few additional jets. This plot also shows the overall normalisation in the *Confirmation* region is matched very well between data and MC.

The lepton variables in Figure 10.8 also show good agreement, even improving on that seen in the *Search* region for the $E_{\mathrm{T}}^{\mathrm{miss}}$ in Figure 10.8(c). The W-boson and charged lepton p_{T} distributions still show a trend in the ratio plot for the prediction to overestimate the data at high p_{T} although this discrepancy is reduced as compared to the *Search* region. The lepton centrality distribution shown in Figure 10.8(e) shows excellent agreement for the scaled signal and background.

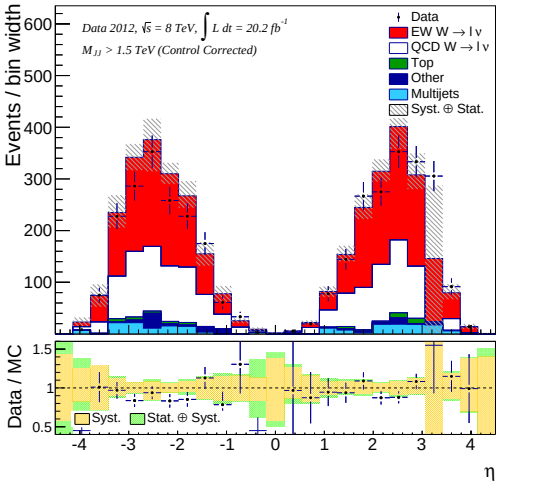
The M_{jj} plot in Figure 10.9 is unremarkable as this is a sub sample of the *Search* region M_{jj} plot but is included for completeness here. Figure 10.9(b) shows the $\Delta\eta$ between the two tagged jets where again good agreement between data and SM prediction is observed. The minimum ΔR between the third jet and the tagging jets is shown in Figure 10.9(c), where the peak is in the towards the low ΔR region that corresponds to FSR. The p_{T} balance distribution is not well modelled in the *Search* region even with the fit scale factors applied



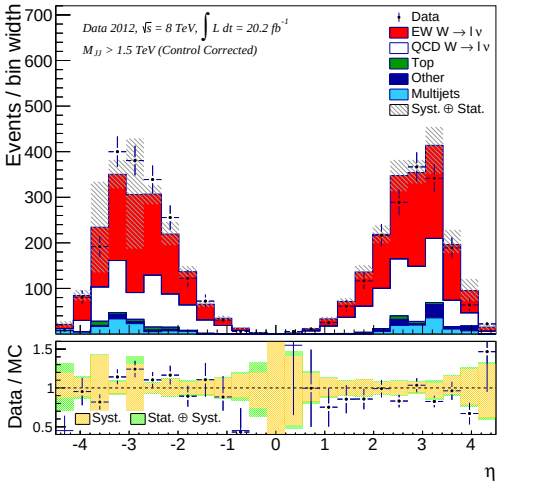
(a) Leading Jet p_{T}



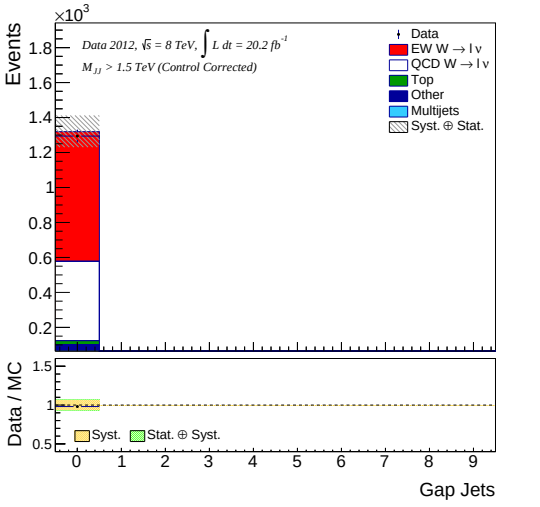
(b) Sub-leading jet p_{T}



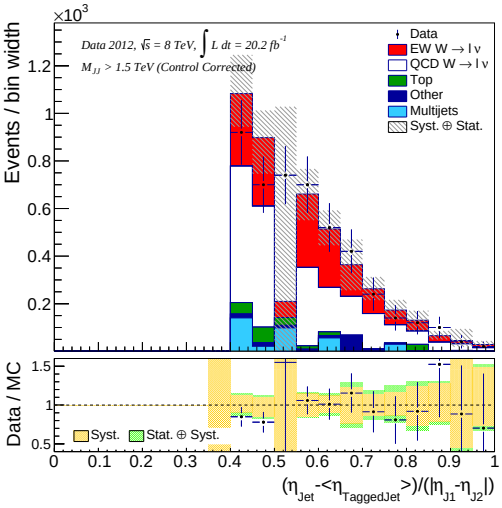
(c) Leading Jet η



(d) Sub-Leading Jet η

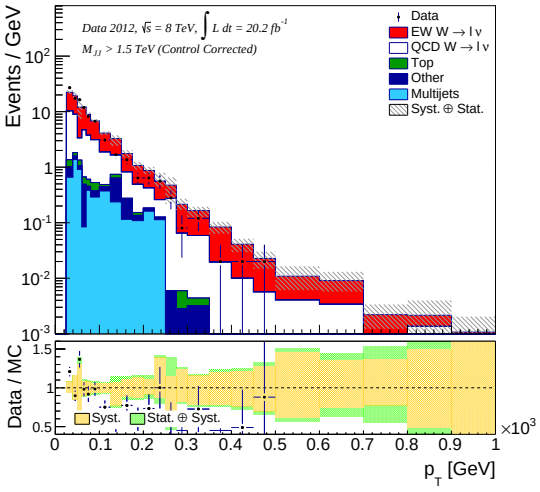


(e) N_{Jets} in the tagged jet η gap

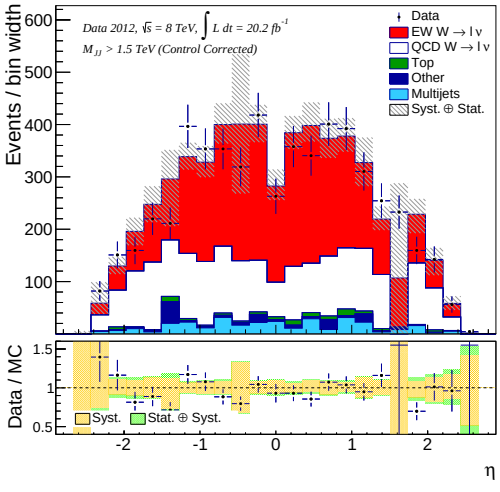


(f) Jet Centrality

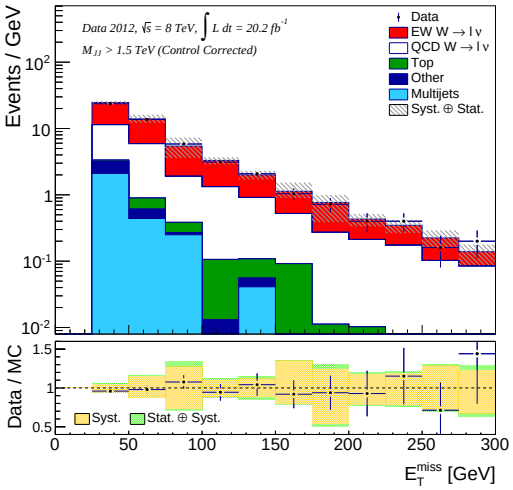
Figure 10.7: Reconstruction-level jet distributions summed over lepton channels in the *Confirmation* region. The EW signal and QCD $W + 2$ jets have been scaled by the fit normalisation results and the latter also has the *Control* region correction.



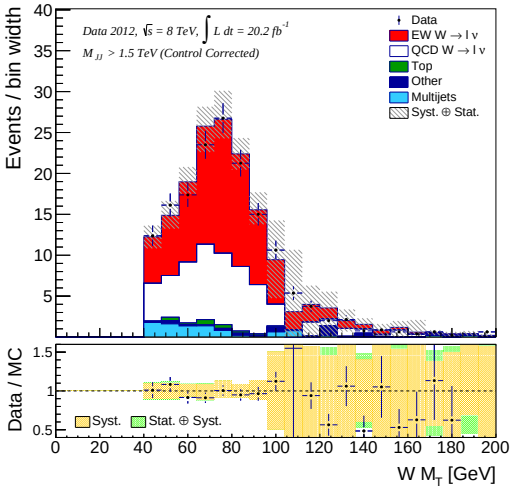
(a) Lepton p_{T}



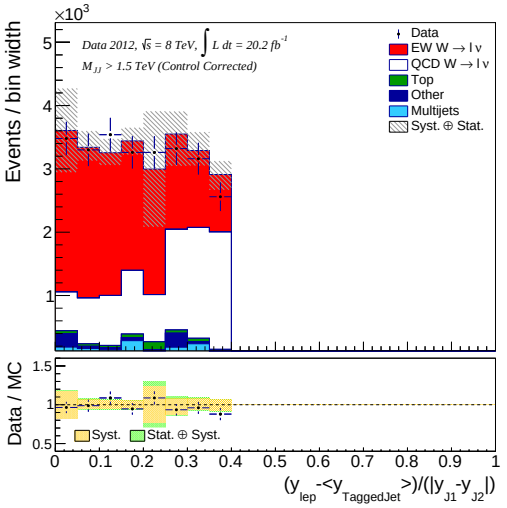
(b) Lepton η



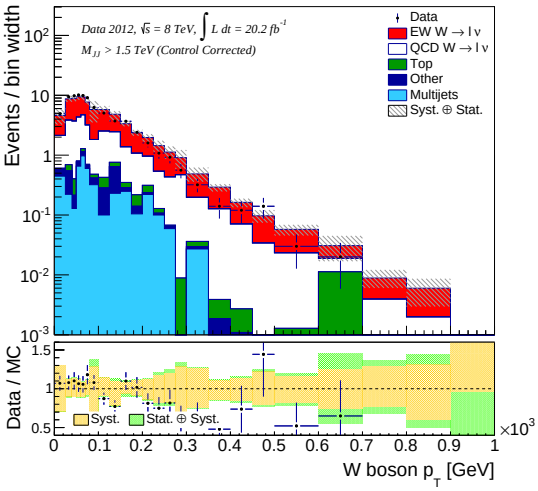
(c) $E_{\mathrm{T}}^{\mathrm{miss}}$



(d) W-boson transverse mass

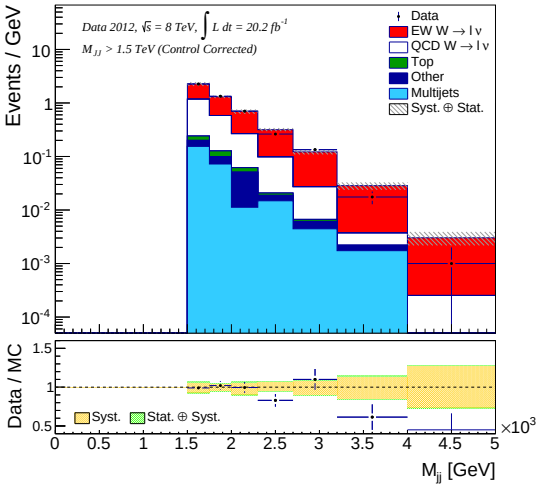


(e) Lepton Centrality

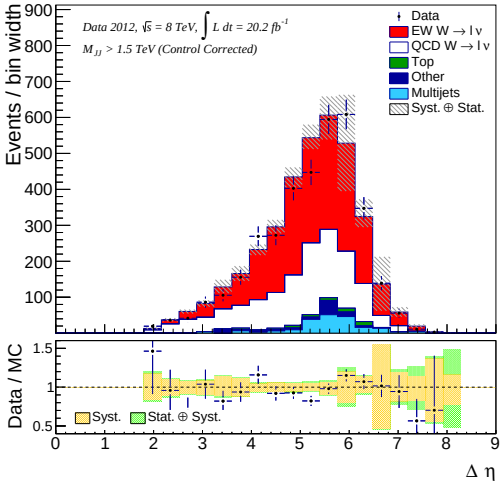


(f) W-boson p_{T}

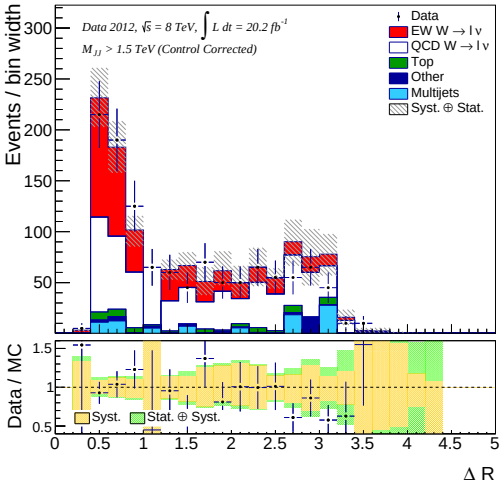
Figure 10.8: Reconstruction-level lepton distributions summed over lepton channels in the *Confirmation* region. The EW signal and QCD W + 2 jets have been scaled by the fit normalisation results and the latter also has the *Control* region correction.



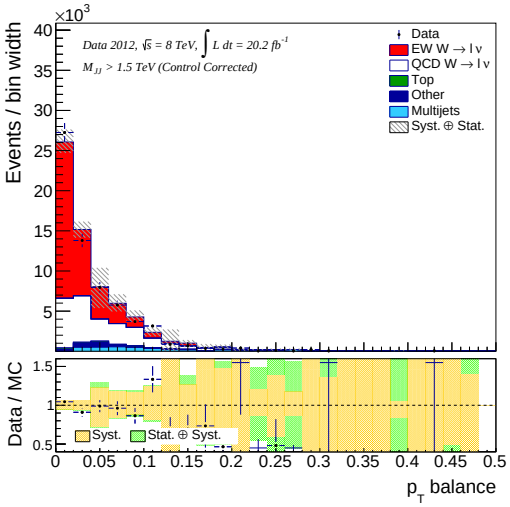
(a) M_{jj}



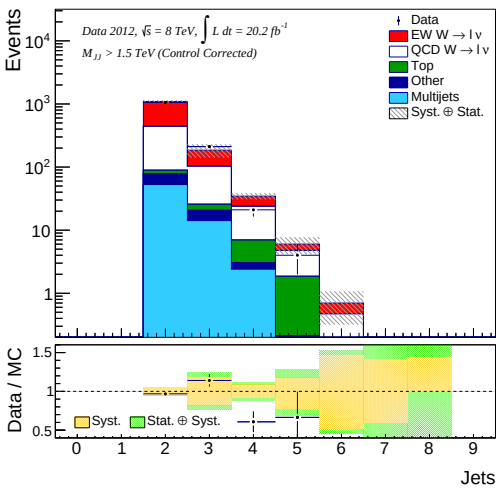
(b) $\Delta\eta(J_1, J_2)$



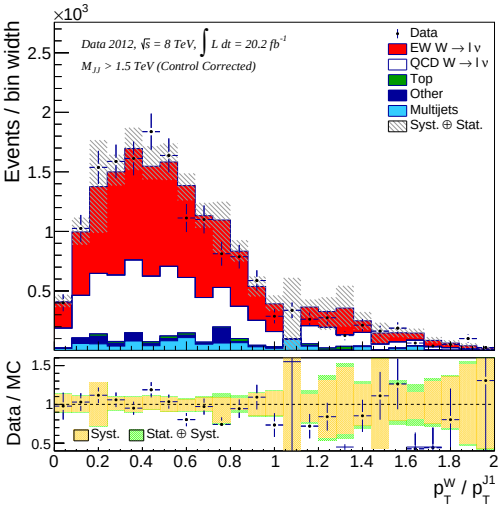
(c) Minimum ΔR (J_3, J_{Tagged})



(d) p_{T} balance



(e) Number of Jets



(f) p_T^W / p_T^{J1}

Figure 10.9: Reconstruction-level distributions summed over lepton channels in the *Confirmation* region. The EW signal and QCD $W + 2$ jets have been scaled by the fit normalisation results and the latter also has the *Control* region correction. The M_{jj} systematic uncertainties have been constrained.

(see Figure 10.3(d)), but in the *Confirmation* region this distribution is well modelled around the peak, as shown in Figure 10.9(d). This improvement is most likely caused by the large reduction in QCD $W + 2$ jets events because the QCD colour-flow makes modelling the exact p_{T} balance difficult and sensitive to higher order corrections. The N_{Jets} plot in Figure 10.9(e) shows the high M_{jj} cut has reduced the previously observed trend in N_{Jets} for high jet multiplicities to be overestimated by the MC, although this is affected by lower statistics in high-jet multiplicity bins.

10.3 Cross Section Calculation

This section discusses the extraction of cross sections for the EW $W + 2$ jets process. A fiducial cross section is calculated in Section 10.3.1 and the total EW $W + 2$ jets cross section is calculated in Section 10.3.2.

10.3.1 Fiducial Cross Section

The visible cross section for the EW $W + 2$ jets process is:

$$\sigma_{\mathrm{Vis}} = \frac{N_{EW}^{\mathrm{Fit}}}{\mathcal{L}}. \tag{10.1}$$

The fiducial cross section corrects the visible cross section for detector response and is defined as:

$$\sigma_{\mathrm{Fid}} = \frac{N_{EW}^{\mathrm{Fit}}}{C.\mathcal{L}}, \tag{10.2}$$

where C accounts for the EW signal events that are not reconstructed by ATLAS but are

Channel	W-boson charge	C
Both	$W^{\pm}$	**0.539**
	W^{+}	0.534
	W^{-}	0.548
Electron	$W^{\pm}$	0.540
	W^{+}	0.536
	W^{-}	0.547
Muon	$W^{\pm}$	0.538
	W^{+}	0.532
	W^{-}	0.548

Table 10.5: C factors for the EW $W + 2$ jets split by lepton channel and W-boson charge. The $W^{\pm}$ factors are a cross section weighted combination of the individual W^{+} and W^{-} factors.

present in the truth-level fiducial phase space (see Section 4.7). C is defined as:

$$C = \frac{N_{EW}^{\text{reco, fid}}}{N_{EW}^{\text{truth, fid}}}. \tag{10.3}$$

The C factors do not depend on the QCD $W + 2$ jets correction are derived for the EW $W + 2$ jets signal from the PowHeg signal samples. Table 10.5 shows the break down of the C factors for the EW $W + 2$ jets PowHeg samples.

As the detector systematics are included in ΔN_{EW}^{reco}, the fractional uncertainty on C is accounted for. The resulting fiducial cross section, for EW $W + 2$ jets (e and μ channels) for the *Search* region as defined in Table 7.1 is:

$$\begin{aligned} \sigma_{\text{Fid}} =& 387.9 \ \pm 20.4 \text{ (stat.)} \pm 5.4 \text{ (lumi.)} \ ^{+34.8}_{-27.3} \text{ (syst.)} \ ^{+34.9}_{-34.6} \text{ (theo.)} \ \text{ fb}, \\ =& 387.9 \ ^{+53.6}_{-48.8} \text{ (total)} \ \text{ fb}, \end{aligned} \tag{10.4}$$

where the QCD $W + 2$ jets has been corrected using the *Control* region.

Table 10.6 shows the fiducial cross section for separate and combined lepton channels and the corresponding errors.

Channel	Cross Section (fb)	Statistical Uncertainty (fb)	Luminosity Uncertainty (fb)	Systematic Uncertainty (fb)	Theoretical Uncertainty (fb)
Both	387.9	$\pm$20.4	$\pm$5.4	$+34.8$ -27.3	$+34.9$ -34.6
Electron	202.7	$\pm$13.6	$\pm$5.3	$+22.1$ -18.1	$+18.1$ -17.7
Muon	184.3	$\pm$14.2	$\pm$5.3	$+14.0$ -14.6	$+16.8$ -16.9

Table 10.6: The extracted fiducial cross section results for individual lepton channels and the combined channel measurements, all using the *Control* region correction.

Channel	Cross Section (fb)	Scale Uncertainty (fb)	PDF Uncertainty (fb)	Generator Uncertainty (fb)
Both	458.98	$+5.26$ -2.34	$+7.64$ -8.44	$\pm$16.24
Electron or Muon	229.49	$+2.63$ -1.17	$+3.82$ -4.22	$\pm$8.12

Table 10.7: Predicted EW $W + 2$ jets fiducial cross sections and associated errors from POWHEG. The generator uncertainty is the POWHEG to VBFNLO difference.

The extracted, fiducial cross sections shown in Table 10.6 are compared to the predicted fiducial cross sections from VBFNLO and POWHEG in Table 10.7. The "Both" fit shown in Table 10.6 is done which considering the prediction from both lepton channels to all the data rather than the sum of the single channel fits. The single channel fit results are shown to be consistent with the "Both" fit result.

10.3.2 Total Cross Section

The total cross section extrapolates the fiducial cross section found above, to as inclusive a phase space as possible. It is defined:

$$\sigma_{\text{Tot}} = \frac{N_{EW}^{\text{Fit}}}{A.C.\mathcal{L}}, \tag{10.5}$$

A	Scale Uncertainty (%)	PDF Uncertainty (%)	Total Uncertainty (%)
0.029	$+1.3$ -0.9	$+2.4$ -2.7	$+2.7$ -2.8

Table 10.8: A factors for the EW $W+2$ jets inclusive cross section, derived using VBFNLO.

where A is the acceptance factor, defined as:

$$A = \frac{N_{EW}^{\text{truth, fid}}}{N_{EW}^{\text{truth, tot}}}. \tag{10.6}$$

N_{EW}^{truth} values are not affected by detector systematic uncertainties at all so the only uncertainties to consider for A are theoretical. The predicted, inclusive cross section summed over lepton channels, is found from VBFNLO to be:

$$\begin{aligned} \sigma_{EW}^{\text{Inclusive, Predicted}} &= 15.209\ ^{+0.09}_{-0.11}\ (\text{scale})\ ^{+0.26}_{-0.29}\ (\text{PDF})\ \text{pb}, \\ &= 15.209\ ^{+0.28}_{-0.31}\ pb. \end{aligned} \tag{10.7}$$

Table 10.8 shows the A factor and associated uncertainties calculated using VBFNLO which provides that cross sections summed over $W^{\pm}$. The values are calculated using the electron channel but assumed to be identical for the muon channel.

The total, extracted cross section for the EW $W+2$ jets process, decaying through the combined electron and muon channels is found to be:

$$\begin{aligned} \sigma_{EW}^{\text{Inclusive, Extracted}} &= 13.375\ \pm 0.70\ (\text{stat.}) \pm 0.18\ (\text{lumi.})\ ^{+1.20}_{-0.94}\ (\text{syst.})\ ^{+1.26}_{-1.25}\ (\text{theo.})\ \ pb, \\ &= 13.375\ ^{+1.88}_{-1.72}\ (\text{total})\ \ pb. \end{aligned} \tag{10.8}$$

The extracted total cross section is consistent with the predicted total cross section at

Analysis	Predicted $\sigma_{EW}^{\text{Fiducial}}$ fb	Extracted $\sigma_{EW}^{\text{Fiducial}}$ fb	μ_{EW} (%)	ΔN_{EW} (%)
ATLAS VBF W $\sqrt{s} = 8$ TeV, 2012	459	388	83	$+15$ -14
ATLAS VBF W $\sqrt{s} = 7$ TeV, 2011	425	325	77	$+26$ -27
CMS VBF Z $\sqrt{s} = 8$ TeV, 2012	208	174	84	$+20$ -20
ATLAS VBF Z $\sqrt{s} = 8$ TeV, 2012	46.1	54.7	119	$+23$ -25
CMS VBF W $\sqrt{s} = 8$ TeV, 2012	500	420	84	$+20.6$ -20.6

Table 10.9: Summary of LHC MS VBF type analyses results.

the 1σ level.

10.4 Reflection

The analysis has found the EW $W + 2$ jets process to occur at 83.3 $^{+14.8}_{-13.6}$ % of the rate predicted by the SM. This puts the SM prediction at just over 1σ away from the observed value. This low measurement qualitatively agrees with the 77 $^{+26}_{-27}$ % rate observed in [36] for the same process. The CMS analysis for the EW $Z + 2$ jets process found the rate to be 84 $\pm$20 % [34]. The ATLAS measurement of the EW $Z + 2$ jets production found a rate 119 $^{+23}_{-25}$ % that predicted by the SM [35]. All measurements are consistent with the SM at about the 1σ level*. These results are summarised in Figure 10.10 and Table 10.9. As the statistical component of the error in these analyses is considerable, although in the current analysis it is sub-leading, further analyses of VBF W and Z production at the LHC in Run 2 are expected to provide more accurate measurements.

A tendency is observed for the exclusive muon-channel measurements to result in a lower cross section than the electron-channel in ATLAS W and Z measurements though the

*Between the submission and examination of this thesis the CMS collaboration published a measurement of the EW production of $W + 2$ jets, the same signal process as searched for in this analysis [125]. The measured rate was found to be 84 $\pm$20.6 % of the SM prediction. This result has been added to Figure 10.10 and Table 10.9

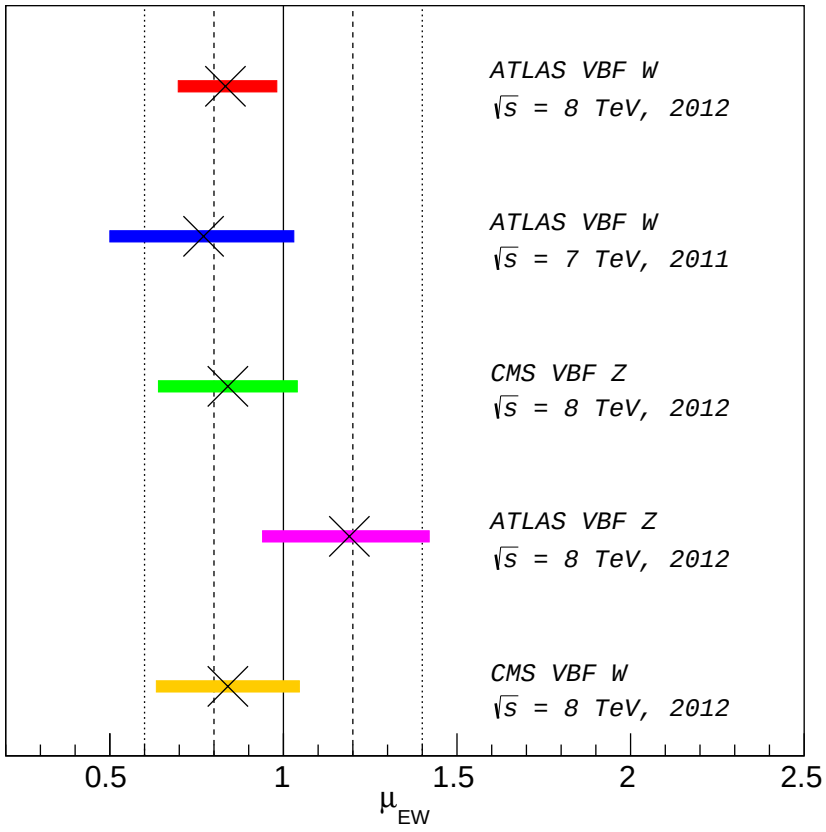


Figure 10.10: μ_{EW} values for VBF analyses from the LHC. The bands show the quoted 1σ uncertainties and the crosses show the central values.

consistency is within 1.1σ [35, 36]. This difference is not observed by the CMS analysis [34].

Future studies of higher energy collisions may be able to use an increased M_{jj} cut and tighter $\mathcal{C}_{\text{Lepton}}$ cut to isolate a high statistics, high purity EW $W + 2$ jets event sample, with reduced uncertainties for a cross section measurement. If the disagreement with the SM prediction observed in this analysis were to persist, it could point towards BSM physics affecting the production cross section. It is possible that a heavy W' boson, as hinted at in [126], could reduce the signal cross-section in this analysis from the SM prediction. The VBF diagram (Figure 2.5(a)) contributes negativly to the interfering sum of EW $W + 2$ jets processes. This is confirmed by calculating the EW $W + 2$ jets cross-section with the triple boson coupling rates set to zero, resulting in the predicted cross-section increases by a factor of 4. If a heavy W' boson did exist, it would be present in the t-channel exchange of Figure 2.5(a) and suppressed in the other EW $W + 2$ jets diagrams due to its high mass. The result could be a reduction in the EW $W + 2$ jets cross-section, as is observed by this, and other, analyses.

The techniques used in this analysis have shown a high M_{jj} cut and small $\mathcal{C}_{\text{Lepton}}$ value can be used to isolate the EW $W + 2$ jets process. This information can be used in future VBF studies and may be especially useful for Higgs boson analyses. Studies in the upgraded LHC runs will have some sensitivity to the $H \to \mu^+\mu^-$ decay (see [127] for a previous analysis) where the topological $\mathcal{C}_{\text{Lepton}}$ will be useful in reducing backgrounds as well as identifying the irreducible background contribution from VBF $Z \to \mu^+\mu^-$ which could be calibrated using VBF $Z \to e^+e^-$ events where the Higgs contribution is negligible. A truth level study using SHERPA is able to show, in Figure 10.11, the centrality of the Higgs boson is more sharply peaked than that of a Z-boson, most likely due to the higher mass of the Higgs. A fit to boson centrality after a high M_{jj} cut could therefore be used to measure the contribution from the Higgs boson to EW $\mu^+\mu^- + 2$ jets production.

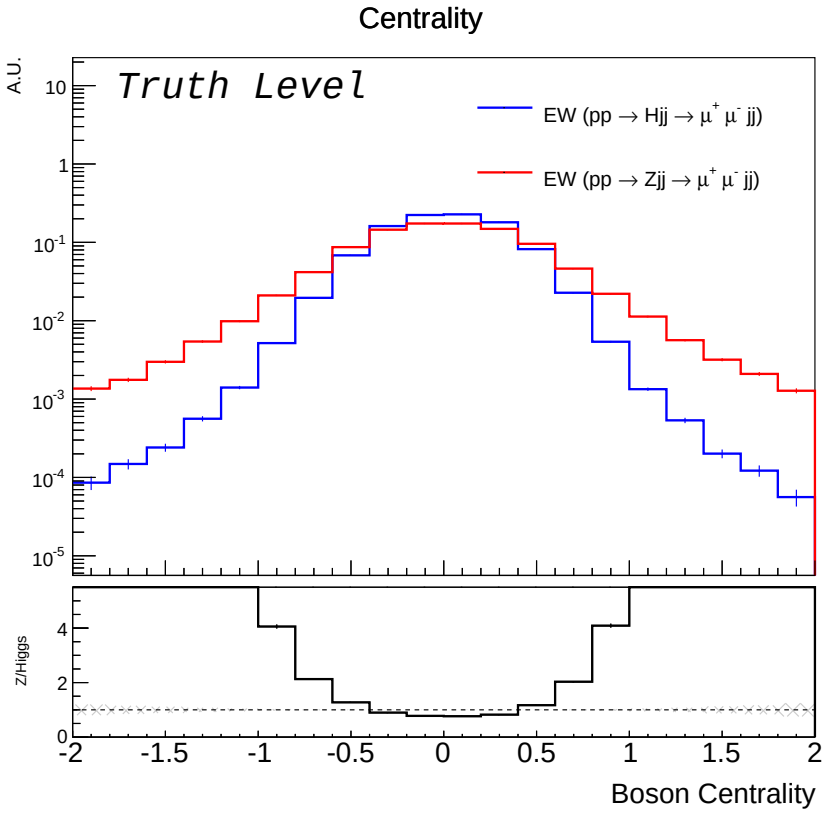


Figure 10.11: Unity normalised, truth level boson centrality distributions for $H \to \mu^+\mu^-$ and $Z \to \mu^+\mu^-$ EW prodcution in association with two jets. The Higgs distribution is observed to be more peaked about zero than the Z-boson case.

Chapter 11

Conclusions

This thesis has presented the first observation of the EW production of the $W + 2$ jets final state. The process is observed with a significance of $> 5\sigma$ using the ATLAS 2012 data-set amounting to 20.2 fb^{-1}.

The analysis uses a *Control* region to correct the modelling of the dominant background contribution. The correction procedure also constrains the systematic uncertainties of the background allowing a more precise signal cross section to be extracted. The signal extraction is repeated using an alternative control region as well as omitting this correction and the results are shown to be consistent with the nominal *Control* region.

The fiducial cross section for EW production of $W + 2$ jets where the W-boson decays leptonically in both the electron and muon channels, is found to be:

$$\begin{aligned} \sigma_{\text{Fid}} =& 387.9 \ \pm 20.4 \text{ (stat.)} \pm 5.4 \text{ (lumi.)} \ ^{+34.8}_{-27.3} \text{ (syst.)} \ ^{+34.9}_{-34.6} \text{ (theo.)} \ \text{ fb}, \\ =& 387.9 \ ^{+53.6}_{-48.8} \text{ (total)} \ \text{ fb}, \end{aligned} \tag{11.1}$$

where the SM prediction from VBFNLO is 443 $^{+5.3}_{-2.3}$ (scale) $^{+7.6}_{-8.4}$ (PDF) fb. The fiducial

cross section is extrapolated to an inclusive cross section of:

$$\begin{aligned}\sigma_{Total} =& 13.38 \ \pm 0.70 \text{ (stat.)} \pm 0.18 \text{ (lumi.)} \ ^{+1.20}_{-0.94} \text{ (syst.)} \ ^{+1.26}_{-1.25} \text{ (theo.)} \ \text{ pb,} \\ =& 13.38 \ ^{+1.88}_{-1.72} \text{ (total)} \ \text{ pb,}\end{aligned} \tag{11.2}$$

where the SM prediction from VBFNLO is $15.21 \ ^{+0.09}_{-0.11}$ (scale) $^{+0.26}_{-0.29}$ (PDF) pb.

In the upcoming LHC Run 2 at $\sqrt{s} = 13$ TeV, the fiducial cross section for EW $W + 2$ jets is expected to increase by a factor of 2.9 as compared to the $\sqrt{s} = 8$ TeV cross section. If this analysis were to be repeated on the expected 30 fb^{-1} of data in Run 2, the statistical uncertainty would be reduced by a factor 2 as compared the current uncertainty.

The dominant analysis uncertainties are the jet energy scale at high $|\eta|$ and the theoretical uncertainty on the QCD-EW $W + 2$ jets interference. The former will hopefully be reduced during the LHC Run 2 where the expertise gained in Run 1 can be built upon. The interference term could be constrained using higher statistics samples that were unfeasible due to time constraints in this analysis, but the main obstacle to reducing the interference uncertainty is the current inability to calculate it at NLO. An accurate prediction of the interference could be incorporated into the fit procedure rather than considered as a systematic, removing the current 6.5% uncertainty. The measurement error also has a significant contribution from the jet flavour composition uncertainty which can be reduced by quantifying the relative fraction of gluon-to-quark jets in the signal.

Appendix A

Sherpa QCD Detector Plots

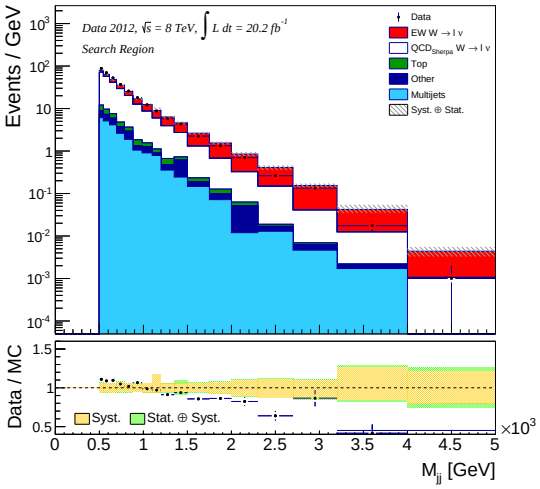
It is possible for a W-boson event with associated 0 or 1 jets to occur in the same bunch crossing as a pile-up event that provides an additional jet or two. This combined process could fulfil the *Search* region requirements of W-boson + 2 jets, although it would be extremely rare for a pile-up jet to have $p_{\mathrm{T}} > 60$ GeV and be able to combine with the hard-scatter jet or a second pile-up jet to produce $M_{jj} > 500$ GeV.

The W + 0/1 jet events are not included in the PowHeg QCD $W + 2$ jets modelling as the MC produces at least two partons in the hard-scatter. To ensure the combination of W + 0/1 jet events with pile-up is not important to this analysis, the PowHeg QCD W + 2 jets sample is replaced by the Sherpa QCD $W +$ jets sample. The Sherpa samples includes $W + 0 \rightarrow 5$ jets events at the matrix element level, although it is not used as the nominal QCD $W +$ jets sample due to fewer statistics than the PowHeg and PowHeg has NLO accuracy for $W + 2$ jets.

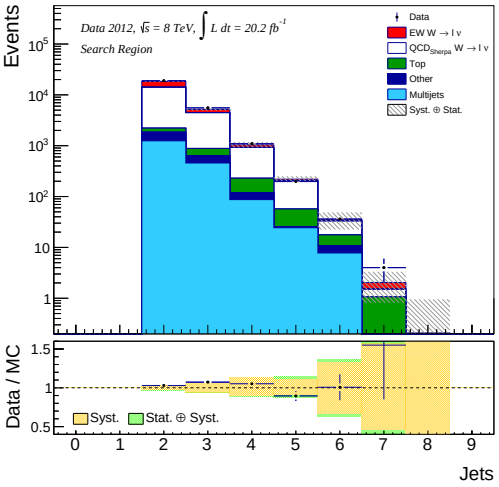
The pre-fit PowHeg QCD $W + 2$ jets detector distributions show some data-prediction disagreement at low W-boson p_{T} (see Figure 7.13(f)) and low p_{T} balance (see Figure 7.14(d)). A low p_{T} W-boson event alongside some pile-up jets would have a low p_{T} balance value and so it is possible the Sherpa QCD $W +$ jets sample will improve the data-prediction agreement in these variables.

The detector plots are shown in the remainder of this chapter for the *Search* region, using the SHERPA samples in place of the POWHEG QCD W + jets. The plots show the "out-of-the-box" MC with no corrections, constraints or fit scale factors applied to any sample. The data-prediction agreement will therefore improve if the *Control* region correction or fit scale factors were applied, but we use these plots to investigate qualitative changes between POWHEG and SHERPA QCD W + jets.

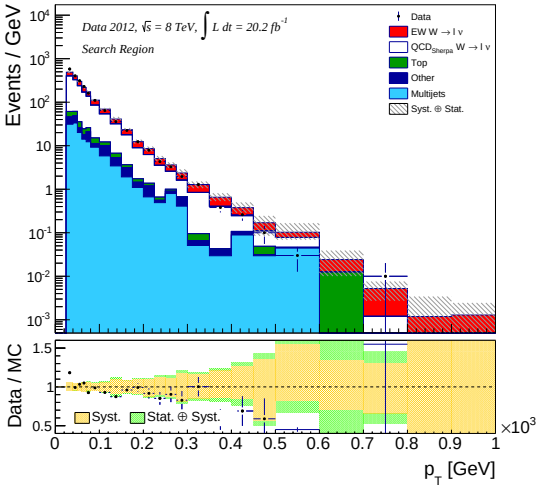
Distributions that show mismodelling in the POWHEG QCD W + 2 jets detector plots are shown in Figure A.1 for the SHERPA QCD W + jets. The M_{jj} distribution shown in Figure A.1(a) shows a steeper data-prediction ratio slope than that observed in the POWHEG case but this could be cancelled by applying the *Control* region correction. The N_{Jets} plot shown in Figure A.1(b) shows better data-prediction agreement for high jet multiplicities than the POWHEG case suggesting that SHERPA is modelling the parton shower better than POWHEG + PYTHIA8. The lepton p_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$ and W-boson p_{T} distributions show a similar level of data-prediction ratio slopes to the POWHEG case. The p_{T} balance distribution shown in Figure A.1(f) shows the same low value disagreement to using the POWHEG QCD W + 2 jets sample. These disagreements cannot be removed by applying the *Control* region correction or fit scale factors. It is therefore concluded that the p_{T} balance and W-boson p_{T} disagreements observed when using the POWHEG QCD W + 2 jets sample are not caused by the lack of W + 0/1 jet events.



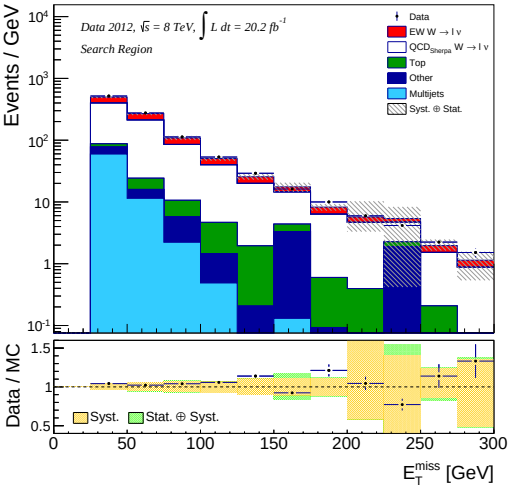
(a) M_{jj}



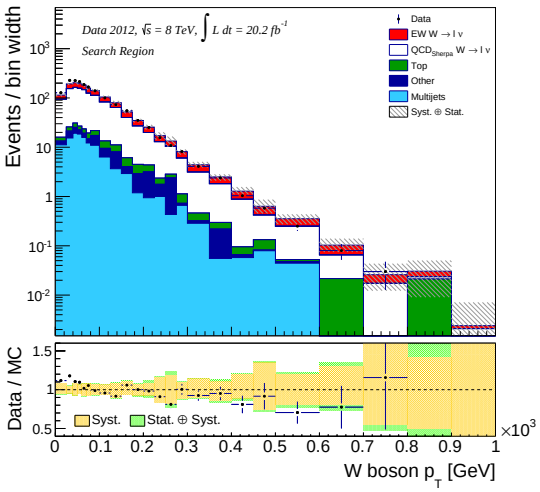
(b) Number of jets



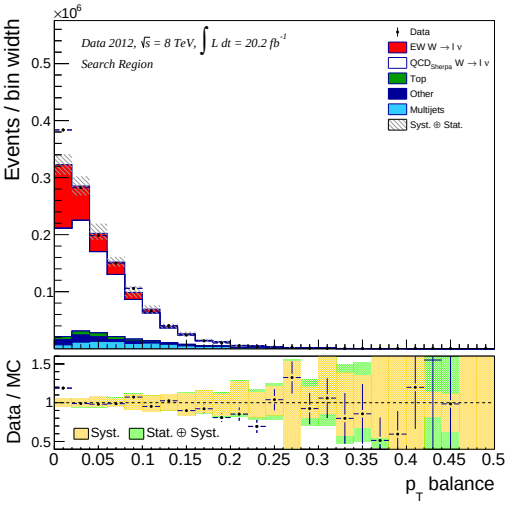
(c) Lepton p_{T}



(d) $E_{\mathrm{T}}^{\mathrm{miss}}$



(e) W-boson p_{T}



(f) p_{T} balance

Figure A.1: Reconstruction level lepton distributions summed over both lepton channels in the *Search* region using the Sherpa QCD W + jets.

Appendix B

PowHeg Polarisation Issues

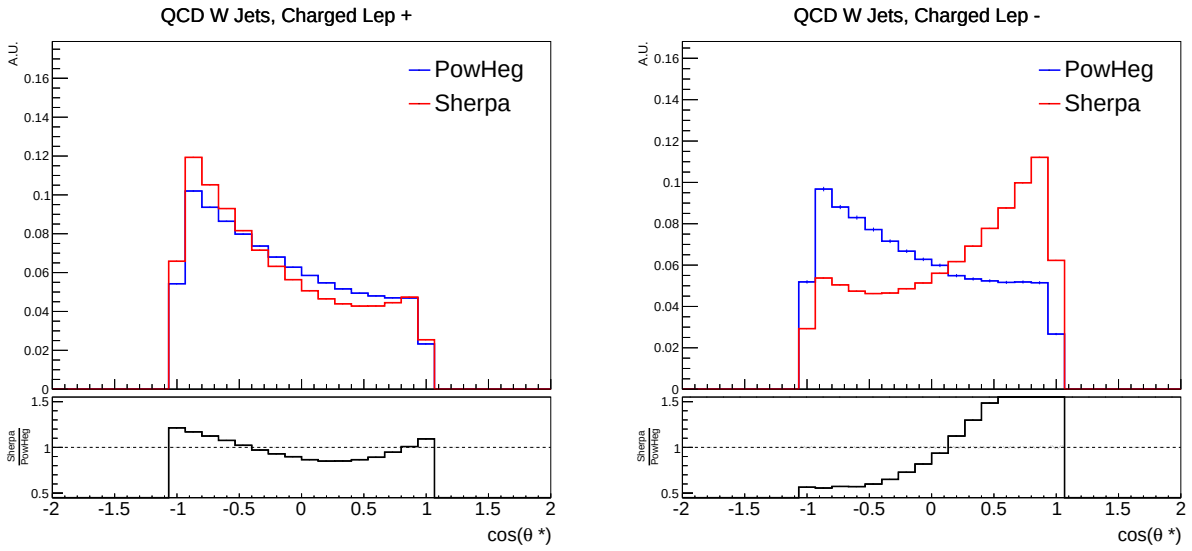
At a late stage in the analysis, another analysis group started to use the same QCD PowHeg $W + 2$ jets samples used in this analysis. While investigating the angle between the W-boson and the charged lepton decay they discovered PowHeg was modelling this incorrectly. This bug has now been corrected in the PowHeg code but there was not sufficient time to re-generate these MC samples. This appendix discusses the bug and its effect on the analysis.

The variable used in these investigations is designated $\cos(\theta^*)$ where θ^* is the angle between the charged lepton and the W-boson when the lepton is boosted in the rest frame of the W-boson. It is calculated as follows:

$$\cos(\theta^*) = \frac{p_x^{l'} p_x^W + p_y^{l'} p_y^W + p_z^{l'} p_z^W}{|p^{l'}||p^W|}, \tag{B.1}$$

where l' denotes the charged lepton 4-vector after being boosted in the W-boson rest frame, W denotes the W-boson, p denotes the magnitude of a particle's momentum and p_i denotes the momentum along the axis i.

In the case of a W^+ boson, the boosted, positively-charged lepton, with $p^{l'}$, tends to decay in the opposite direction to the boson's trajectory (p^W, in the lab-frame), leading to a peak in the $\cos(\theta^*)$ distribution at -1. The opposite is true of the W^- case where the negatively charged lepton tends to decay in the same direction as the boson's trajectory and



(a) W^+ (b) W^-

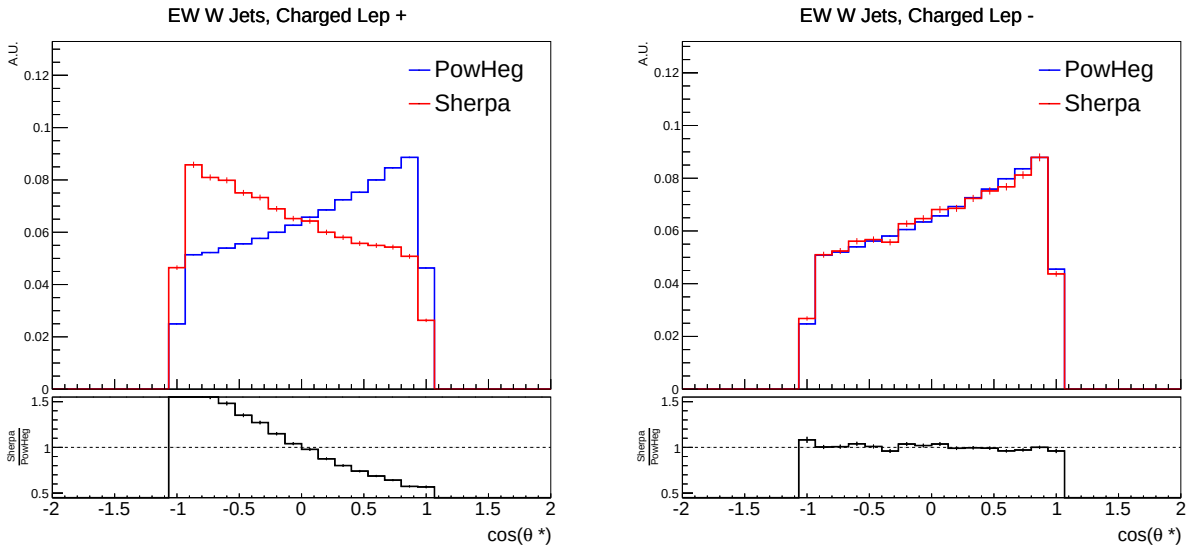
Figure B.1: QCD $W + 2$ jets $\cos(\theta^*)$ distributions for PowHeg and Sherpa muon channel, separated by W-boson charge.

thus $\cos(\theta^*)$ peaks at $+1$.

B.0.1 QCD $W + 2$ jets

Figure B.1(a) shows the $\cos(\theta^*)$ distribution for QCD $W^+ + 2$ jets production where PowHeg is compared to Sherpa and the distribution peaks at -1 in both generators. Figure B.1(b) shows that whilst Sherpa correctly decays the negative lepton in the same direction as the W^- boson, PowHeg does not. The PowHeg distribution is the same shape as that seen for the W^+ case.

This difference is observed in both the charged lepton and neutrino decay angles where PowHeg is decaying a W^- boson as if it were a W^+. The difference is only observed in the muon channel; the electron channel has the correct decay angles for both the W^+ and W^-, and Sherpa and PowHeg agree on these shapes.



(a) W^+ (b) W^-

Figure B.2: EW $W + 2$ jets $\cos(\theta^*)$ distributions for PowHeg and Sherpa muon channel, separated by W-boson charge.

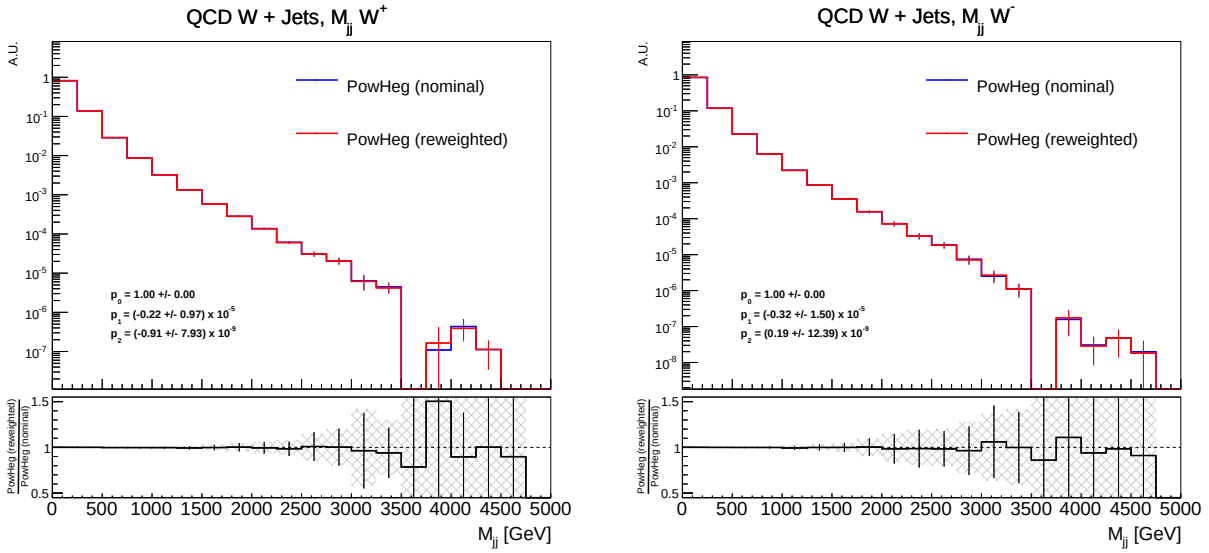
B.0.2 EW $W + 2$ jets

In the EW $W + 2$ jets signal, the faulty PowHeg decay was observed in both the electron and muon channels for the W^+ case. Figure B.2 show the $\cos(\theta^*)$ distributions for the electron channel as an example.

B.1 Affect On Dijet Mass

The most important consideration for this analysis was the effect this bug would have on the M_{jj} distribution. If the bug were to cause a mismodelling of the M_{jj} distribution then the signal extraction fits could be biased. As the W-boson decay is largely independent of the hard-scatter, the faulty decay angle was not thought to have a large effect on the dijet mass. To quantify this effect, the PowHeg events were re-weighted to the Sherpa $\cos(\theta^*)$ distributions and the M_{jj} distribution was compared before and after re-weighting. Figure B.3 shows the M_{jj} shapes for the QCD $W + 2$ jets and the effect of the re-weighting.

Only a subset of the total PowHeg QCD $W + 2$ jets events were used in this cross-check



(a) W^+ (b) W^-

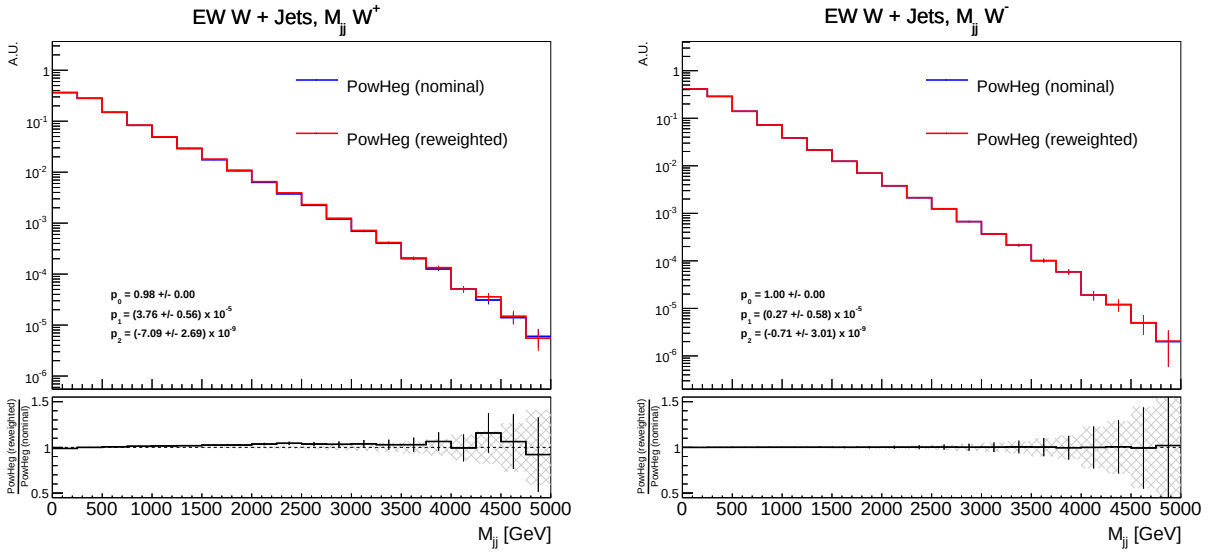
Figure B.3: QCD $W + 2$ jets M_{jj} distributions for PowHeg muon channel before and after correcting to the Sherpa $\cos(\theta^*)$ shapes. The distributions are shown separated by W-boson charge to compare to Figure B.1. The parameters of a 2nd order polynomial fit to the ratio are shown along with their errors.

so the shapes do not match those used in the detector level distributions. A second order polynomial:

$$\frac{M_{jj}^{\text{Re-weighted}}}{M_{jj}^{\text{Nominal}}} = p_0 + p_1 M_{jj} + p_2 M_{jj}^2, \tag{B.2}$$

was fitted to the ratio of the re-weighted to nominal M_{jj} shapes and the parameters are errors are shown in the plots. The effect of the re-weighing in Figure B.3 is seen to be minimal and never outside the statistical error of the events. The fitted p_1 and p_2 parameters for the re-weighting function are seen to be smaller than their errors. It was therefore concluded that the $\cos(\theta^*)$ bug did not affect the QCD $W +$ jets M_{jj} shapes and the signal extraction was insensitive to this bug.

The same re-weighting was done for the EW $W + 2$ jets PowHeg samples and the effects are shown in Figure B.4. The re-weighting has no effect on the W^- M_{jj} shapes, as the re-weighting as shown by the ratio plot of Figure B.2(b) is very flat and the fit to the ratio shows the p_1 and p_2 values are smaller than their errors. Figure B.4(a) does show



(a) W^+ (b) W^-

Figure B.4: EW $W + 2$ jets M_{jj} distributions for POWHEG muon channel before and after correcting to the SHERPA $\cos(\theta^*)$ shapes. The distributions are shown separated by W-boson charge to compare to Figure B.2. The parameters of a 2nd order polynomial fit to the ratio are shown along with their errors.

some effect from the re-weighting are the corrected M_{jj} distribution is shifted to higher M_{jj} than the faulty POWHEG and this time p_1 and p_2 are larger than the errors. The analysis therefore may be dependent on the EW $W +$ jets POWHEG bug.

B.1.1 Effect On Signal Extraction

The nominal signal extraction fit was repeated using the POWHEG EW and QCD $W + 2$ jets samples after re-weighting them to the SHERPA $\cos(\theta^*)$ shapes. The number of extracted N_{EW} events was found to change by 0.6% for the *Control* region corrected fits in both lepton channels. This is much smaller than the total uncertainty of the analysis and so the POWHEG bug in the lepton decay angles is shown to have a negligible effect on the analysis. Table B.1 shows the effect on the extracted number of EW signal events as compared to the nominal fits.

Constraint Region	Channel	Nominal N_{EW}	Re-weighted N_{EW}	% Difference
Control	Both	4,183	4,158	0.6
	Electron	2,180	2,167	0.6
	Muon	2,003	1,992	0.6
Validation	Both	4,628	4,594	0.7
	Electron	2,371	2,353	0.7
	Muon	2,250	2,233	0.8
No constraint	Both	4,031	3,987	1.1
	Electron	2,162	2,139	1.0
	Muon	1,867	1,846	1.2

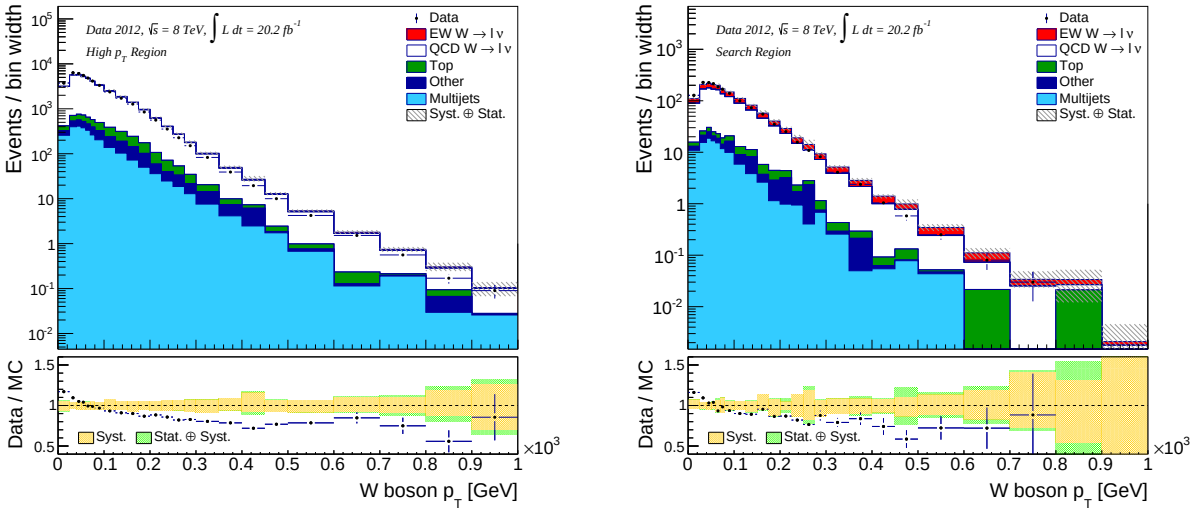
Table B.1: Fit results using the re-weighted POWHEG M_{jj} spectra for both the EW and QCD $W + 2$ jets samples from the SHERPA $\cos(\theta^*)$ re-weighting.

B.2 *W*-boson p_{T}

Given the mismodelling of the *W*-boson p_{T} spectrum it is worth checking whether it is related to the POWHEG bug in the $W + 2$ jets samples. The *W*-boson p_{T} distribution is shown in Figure B.5 for the *High* p_{T} region where we have high statistics, and for the *Search* region were the same data-MC trend is evident.

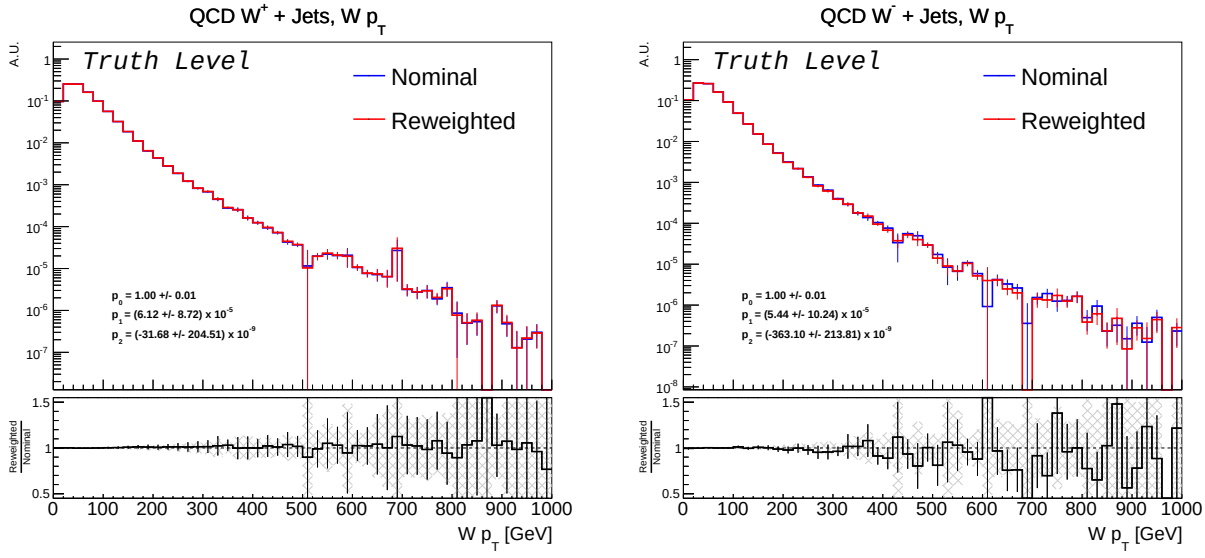
Considering the kinematics of the process, there is no obvious reason that the leptons decaying in the wrong direction should alter the reconstructed *W*-boson p_{T}. The angle between the charged lepton and neutrino is modelled well as evidenced by the good modelling of the *W*-boson transverse mass. The QCD $W + 2$ jets sample is re-weighted to the SHERPA $\cos(\theta^*)$ shape and the *W*-boson p_{T} spectrum compared before and after this re-weighting. The *W*-boson charge-separated distributions are shown in Figure B.6.

The re-weighting is observed to have very little effect on the *W*-boson p_{T} distribution, especially in the low p_{T} region where the data-MC mismodelling is observed. It is therefore concluded that the POWHEG $W + 2$ jets $\cos(\theta^*)$ bug is not responsible for the *W*-boson p_{T} mismodelling or for the charged lepton p_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ mismodelling.



(a) *High* p_{T} region. (b) *Search* region.

Figure B.5: The *W*-boson p_{T} distributions showing the mismodelling of MC underestimating data at low p_{T} and overestimating data at high p_{T}.



(a) W^+ (b) W^-

Figure B.6: The PowHeg QCD $W + 2$ jets *W*-boson p_{T} spectrum before and after correcting to the Sherpa $\cos(\theta^*)$ shape.

Appendix C

Distribution Quality

To quantify the improvement in the detector-level distributions that is achieved from the signal fits, the χ^2/NDF values are found using Equation 5.1. Table C.1 shows the χ^2/NDF values for various distributions in the *Search* region, both pre and post-fit, and the post-fit *Confirmation* region. Only the statistical errors are considered in the calculations. The ratios of χ^2/NDF values demonstrate that the data-prediction agreement improves with the signal fit and that the post-fit *Confirmation* region has good data modelling.

The boson p_{T} is observed to be have a high χ^2/NDF value in all measured regions. This is responsible for the high values for lepton p_{T}, $p_{\mathrm{T}}^{W}/p_{\mathrm{T}}^{J_1}$, and p_{T} balance. The N_{Jets} distribution also demonstrates poor modelling specifically an overestimation at high jet multiplicities.

The NDF values in Table C.1 are found from the number of bins in the distributions with either data or prediction content. The post-fit *Search* region and *Confirmation* region M_{jj} values have three degrees of freedom subtracted for the three extracted scale factors in the fit.

Distribution \ Region	*Search* (Pre-fit)	*Search* (Post-fit)	*Confirmation* (Post-fit)	$\frac{\text{Search (Post-fit)}}{\text{Search (Pre-fit)}}$	$\frac{\textit{Confirmation}\text{ (Post-fit)}}{\text{Search (Post-fit)}}$
Leading jet p_{T}	51 / 17	40 / 17	14 / 17	0.80	0.37
Sub-leading jet p_{T}	22 / 17	20 / 17	16 / 17	0.96	0.80
Third jet p_{T}	17 / 14	16 / 14	12 / 11	0.94	1.02
Leading jet η	41 / 24	37 / 24	25 / 22	0.90	0.73
Sub-leading jet η	50 / 25	43 / 25	27 / 23	0.87	0.68
Third jet η	34 / 25	33 / 25	27 / 20	0.97	1.02
$\mathcal{C}_{Jet}$	24 / 13	21 / 13	7 / 13	0.92	0.35
Lepton p_{T}	103 / 20	87 / 20	56 / 18	0.86	0.71
Lepton η	43 / 23	36 / 23	21 / 22	0.83	0.63
$E_{\mathrm{T}}^{\mathrm{miss}}$	34 / 12	32 / 12	2 / 12	0.98	0.08
W-boson M_{T}	49 / 21	48 / 21	13 / 20	0.98	0.30
$\mathcal{C}_{Lepton}$	7 / 8	3 / 8	4 / 8	0.47	1.28
W-boson p_{T}	154 / 22	144 / 22	32 / 21	0.94	0.23
M_{jj}	49 / 18	22 / 15	13 / 5	0.54	1.75
$\Delta\eta(J_1, J_2)$	24 / 17	15 / 17	16 / 17	0.67	1.05
Minimum $\Delta R(J_3, J_{Tagged})$	39 / 25	36 / 25	10 / 17	0.93	0.44
p_{T} balance	174 / 25	219 / 25	24 / 18	1.27	0.16
N_{Jets}	34 / 6	36 / 6	8 / 4	1.08	0.35
$p_{\mathrm{T}}^{W}/p_{\mathrm{T}}^{J_1}$	153 / 26	155 / 26	39 / 26	1.02	0.25
Average	2.98	2.84	1.25	0.89	0.64

Table C.1: χ^2/NDF values for the detector level distributions. The *Search* (Pre-fit) values corresponds to the plots in Section 7.4. The *Search* (Post-fit) values come from the plots in Section 10.1 and the *Confirmation* values come from the post-fit plots in Section 10.2. The last two columns show the ratio of *Search* post-fit to pre-fit χ^2/NDF and post-fit *Confirmation* region to post-fit *Search* region χ^2/NDF, respectively.

Glossary of abbreviations

ALFA absolute luminosity for ATLAS

aTGC anomalous triple gauge coupling

BSM beyond-the-standard-model

CB Combined Muon

CERN European Centre for Nuclear Research

EM Calo Electromagnetic Calorimeter

EW Electroweak

FSR final state radiation

GRL good runs list

gsw Glashow-Salam-Weinberg

HLT High Level Trigger

ID Inner Detector

ISR initial state radiation

JVF jet vertex fraction

LEP Large Electron-Positron

LHC Large Hadron Collider

LO leading order

LUCID luminosity measurement using Cerenkov integrating detector

MC Monte Carlo

MDT Muon Drift Tubes

M_{jj} dijet mass

MS Muon Spectrometer

NLO next to leading order

NNLO next-to-next-to-leading order

PDF parton distribution function

PS Parton Shower

QCD Quantum Chromodynamics

RoI Region Of Interest

RPC Resistive Plate Chambers

SCT Semiconductor Tracker

SM Standard Model

TGC Thin Gap Chambers

TRT Transition Radiation Tracker

VBF vector-boson fusion

Appendix D

Bibliography

[1] A. Dafinca, J. Henderson, and A. R. Weidberg, "Single event upset studies using the ATLAS SCT", *Journal of Instrumentation* **9** no. 01, (2014) C01050. http://stacks.iop.org/1748-0221/9/i=01/a=C01050.

[2] D. Binosi and L. Theußl, "Jaxodraw: A graphical user interface for drawing feynman diagrams.", *Computer Physics Communications* **161** no. 1-2, (2004) 76–86. http://dblp.uni-trier.de/db/journals/cphysics/cphysics161.html#BinosiT04.

[3] T. Sumida, "Standard model results from ATLAS", *Nuclear Physics B - Proceedings Supplements* **219–220** no. 0, (2011) 201 – 208. http://www.sciencedirect.com/science/article/pii/S0920563211008371. The Proceedings of the 5th Joint International Hadron Structure '11 Conference.

[4] V. Oreshkin, "Standard model and QCD results from ATLAS and CMS", *Nuclear Physics B - Proceedings Supplements* **245** no. 0, (2013) 94 – 100. http://www.sciencedirect.com/science/article/pii/S0920563213006051. The Proceedings of the 7th Joint International Hadron Structure'13 Conference.

[5] T. Cheng and L. Li, *Gauge theory of elementry particle physics.* Oxford University Press, 15th ed., 1954.

[6] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory.* Westview, 1995.

[7] F. Halzen and A. Martin, *Quarks & Leptons.* John Wiley & Sons, 1984.

[8] I. Aitchison and A. Hey, *Gauge Theories in Particle Physics.* Adam Hilger LTD, 1982.

[9] **Particle Data Group** Collaboration, K. Olive *et al.*, "Review of Particle Physics", *Chin.Phys.* **C38** (2014) 090001.

[10] "Precision electroweak measurements on the Z resonance", *Physics Reports* **427** no. 5–6, (2006) 257 – 454. http://www.sciencedirect.com/science/article/pii/S0370157305005119.

[11] U. S. D. o. E. Fermilab, Office of Science, "Wikipedia, standard model article", http://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg. Accessed on 27/05/15.

[12] T. Kajita, "Discovery of neutrino oscillations", *Reports on Progress in Physics* **69** no. 6, (2006) 1607. http://stacks.iop.org/0034-4885/69/i=6/a=R01.

[13] S. Glashow, "Partial Symmetries of Weak Interactions", *Nucl.Phys.* **22** (1961) 579–588.

[14] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries", *Phys.Rev.* **127** (1962) 965–970.

[15] P. Higgs, "Broken symmetries, massless particles and gauge fields", *Physics Letters* **12** no. 2, (1964) 132 – 133. http://www.sciencedirect.com/science/article/pii/0031916364911369.

[16] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323. http://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[17] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles", *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587. http://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[18] **ATLAS** Collaboration, "An update to the combined search for the Standard Model Higgs boson with the ATLAS detector at the LHC using up to 4.9 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV", Tech. Rep. ATLAS-CONF-2012-019, CERN, Geneva, Mar, 2012. https://cds.cern.ch/record/1430033.

[19] **CMS** Collaboration, S. Chatrchyan *et al.*, "Combined results of searches for the standard model Higgs boson in *pp* collisions at $\sqrt{s} = 7$ TeV", *Phys.Lett.* **B710** (2012) 26–48, arXiv:1202.1488 [hep-ex].

[20] V. E. e. a. Barnes, "Observation of a hyperon with strangeness minus three", *Phys. Rev. Lett.* **12** (Feb, 1964) 204–206. http://link.aps.org/doi/10.1103/PhysRevLett.12.204.

[21] **LHCb** Collaboration, R. Aaij *et al.*, "Observation of the resonant character of the $Z(4430)^-$ state", *Phys. Rev. Lett.* **112** no. 22, (2014) 222002, arXiv:1404.1903 [hep-ex].

[22] **LHCb** Collaboration, R. Aaij *et al.*, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays", *Phys. Rev. Lett.* **115** (2015) 072001, arXiv:1507.03414 [hep-ex].

[23] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton fragmentation and string dynamics", *Physics Reports* **97** no. 2–3, (1983) 31 – 145. http://www.sciencedirect.com/science/article/pii/0370157383900807.

[24] D. Amati and G. Veneziano, "Preconfinement as a property of perturbative qcd", *Physics Letters B* **83** no. 1, (1979) 87 – 92. http://www.sciencedirect.com/science/article/pii/0370269379908967.

[25] "Search for $W' \to t\bar{b}$ in the lepton plus jets final state in proton–proton collisions at a centre-of-mass energy of $\sqrt{s} = 8$ TeV with the ATLAS detector ", *Physics Letters B* **743** no. 0, (2015) 235 – 255. http://www.sciencedirect.com/science/article/pii/S0370269315001422.

[26] E. Accomando and A. Kaiser, "Electroweak corrections and anomalous triple gauge-boson couplings in W^+W^- and $W^\pm Z$ production at the LHC", *Phys.Rev.* **D73** (2006) 093006, arXiv:hep-ph/0511088 [hep-ph].

[27] **ATLAS** Collaboration, "Measurement of W^+W^- production in pp collisions at $\sqrt{s}$=7 TeV with the ATLAS detector and limits on anomalous WWZ and $WW\gamma$ couplings", *Phys. Rev. D* **87** (Jun, 2013) 112001. http://link.aps.org/doi/10.1103/PhysRevD.87.112001.

[28] S. Weinzierl, "Introduction to Monte Carlo methods", arXiv:hep-ph/0006269 [hep-ph].

[29] M. D. Palma and G. Iaselli, "A statistical approach to hadronic shower generation", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **235** no. 1, (1985) 70 – 73. http://www.sciencedirect.com/science/article/pii/0168900285902475.

[30] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, *et al.*, "Event generation with SHERPA 1.1", *JHEP* **0902** (2009) 007, arXiv:0811.4622 [hep-ph].

[31] K. Hamilton, P. Nason, and G. Zanderighi, "MINLO: Multi-Scale Improved NLO", *JHEP* **1210** (2012) 155, arXiv:1206.3572 [hep-ph].

[32] R. Devenish and A. Cooper-Sarkar, *Deep Inelastic Scattering*. Oxford University Press, 2004.

[33] F. Schissler and D. Zeppenfeld, "Parton shower effects on W and Z production via vector boson fusion at NLO QCD", *Journal of High Energy Physics* **2013** no. 4, (2013) . http://dx.doi.org/10.1007/JHEP04%282013%29057.

[34] **CMS** Collaboration, V. Khachatryan *et al.*, "Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$", *Eur.Phys.J.* **C75** no. 2, (2015) 66, arXiv:1410.3153 [hep-ex].

[35] **ATLAS** Collaboration, "Measurement of the electroweak production of dijets in association with a Z-boson and distributions sensitive to vector boson fusion in proton-proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector", *JHEP* **1404** (2014) 031, arXiv:1401.7610 [hep-ex].

[36] R. King, *A First Measurement of Electroweak Production of a W Boson in Association with Two Jets with the ATLAS Detector*. PhD thesis, Oxford U., Jul, 2013. https://cds.cern.ch/record/1697448. Presented 23 Oct 2013.

[37] **ATLAS** Collaboration, "Search for resonant diboson production in the $WW/WZ \to \ell\nu jj$ decay channels with the ATLAS detector at $\sqrt{s}$=7 TeV", *Phys. Rev. D* **87** (Jun, 2013) 112006. http://link.aps.org/doi/10.1103/PhysRevD.87.112006.

[38] R. Torre, "Limits on Leptophobic W' after 1 fb^{-1} of LHC Data: A Lesson on parton Level Simulations", arXiv:1109.0890 [hep-ph].

[39] U. Baur and D. Zeppenfeld, "Measuring three vector boson couplings in $qq \to qqW$ at the SSC", arXiv:hep-ph/9309227 [hep-ph].

[40] *LEP design report*. CERN, Geneva, 1984. https://cds.cern.ch/record/102083. Copies shelved as reports in LEP, PS and SPS libraries.

[41] "Wikipedia LHC article", https://en.wikipedia.org/wiki/Large_Hadron_Collider. Accessed on 14/06/15.

[42] **LHCb** Collaboration, J. Alves, A. Augusto *et al.*, "The LHCb detector at the LHC", *JINST* **3** (2008) S08005.

[43] **ALICE** Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** no. 08, (2008) S08002. http://stacks.iop.org/1748-0221/3/i=08/a=S08002.

[44] **CMS** Collaboration, S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004.

[45] L. Evans and P. Bryant, "LHC machine", *Journal of Instrumentation* **3** no. 08, (2008) S08001. http://stacks.iop.org/1748-0221/3/i=08/a=S08001.

[46] **ATLAS** Collaboration, "Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics", arXiv:0901.0512 [hep-ex].

[47] **ATLAS** Collaboration, "TWiki > AtlasPublic > LuminosityPublicResults", https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults. Accessed: 23/02/2015.

[48] **ATLAS** Collaboration, "The atlas experiment at the cern large hadron collider", *Journal of Instrumentation* **3** no. 08, (2008) S08003. http://stacks.iop.org/1748-0221/3/i=08/a=S08003.

[49] P. F. Åkesson, T. Atkinson, M. J. Costa, M. Elsing, S. Fleischmann, A. N. Gaponenko, W. Liebig, E. Moyse, A. Salzburger, and M. Siebel, "ATLAS Tracking Event Data Model", Tech. Rep. ATL-SOFT-PUB-2006-004. ATL-COM-SOFT-2006-005. CERN-ATL-COM-SOFT-2006-005, CERN, Geneva, Jul, 2006. https://cds.cern.ch/record/973401.

[50] **ATLAS** Collaboration, "Performance of the ATLAS Trigger System in 2010", *Eur.Phys.J.* **C72** (2012) 1849, arXiv:1110.1530 [hep-ex].

[51] **ATLAS LUCID** Collaboration, F. Lasagni Manghi, "The LUCID detector",. https://cds.cern.ch/record/2018182.

[52] A. G. Brandt, "ATLAS Forward Derectors", Tech. Rep. ATL-COM-LUM-2010-024, CERN, Geneva, Jun, 2010. https://cds.cern.ch/record/1274041. June 30, 2010.

[53] I. Bird, "Computing for the large hadron collider", *Annual Review of Nuclear and Particle Science* **61** no. 1, (2011) 99–118. http://dx.doi.org/10.1146/annurev-nucl-102010-130059.

[54] S. A. et. al, "Geant4 - a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** no. 3, (2003) 250 – 303. http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[55] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse, and A. Salzburger, "The new ATLAS track reconstruction (NEWT)", *Journal of Physics: Conference Series* **119** no. 3, (2008) 032014. http://stacks.iop.org/1742-6596/119/i=3/a=032014.

[56] **ATLAS** Collaboration, "Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data", *Eur.Phys.J.* **C74** no. 10, (2014) 3071, arXiv:1407.5063 [hep-ex].

[57] **ATLAS** Collaboration, "Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s}$ =7 TeV in the ATLAS experiment", Tech. Rep. ATLAS-CONF-2010-069, CERN, Geneva, Jul, 2010. http://cds.cern.ch/record/1281344.

[58] **ATLAS** Collaboration, "Performance of the ATLAS Electron and Photon Trigger in p-p Collisions at $\sqrt{s} = 7$ TeV in 2011", Tech. Rep. ATLAS-CONF-2012-048, CERN, Geneva, May, 2012. http://cds.cern.ch/record/1450089.

[59] "Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data", Tech. Rep. ATLAS-CONF-2014-032, CERN, Geneva, Jun, 2014. https://cds.cern.ch/record/1706245.

[60] M. Wielers, R. Mantifel, A. Tricoli, and P. Bell, "Single Electron Trigger Performance Plots", Tech. Rep. ATL-COM-DAQ-2012-146, CERN, Geneva, Jun, 2012. https://cds.cern.ch/record/1456795.

[61] **ATLAS** Collaboration, "Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data", *Eur.Phys.J.* **C72** (2012) 1909, arXiv:1110.3174 [hep-ex].

[62] S. Laplace and J. de Vivie, "Calorimeter isolation and pile-up", Tech. Rep. ATL-COM-PHYS-2012-467, CERN, Geneva, May, 2012. https://cds.cern.ch/record/1444890.

[63] M. Aharrouche *et al.*, "Energy linearity and resolution of the ATLAS electromagnetic barrel calorimeter in an electron test-beam", *Nucl. Instrum. Meth.* **A568** (Dec., 2006) 601–623, physics/0608012. http://arxiv.org/abs/physics/0608012.

[64] **ATLAS** Collaboration, "Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s} = 8$ TeV", arXiv:1408.3179 [hep-ex].

[65] **ATLAS** Collaboration, "Muon Trigger Public Results", 2012. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonTriggerPublicResults.

[66] **ATLAS** Collaboration, "Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data", *The European Physical Journal C* **74** no. 11, (2014) . http://dx.doi.org/10.1140/epjc/s10052-014-3130-x.

[67] **ATLAS** Collaboration, "Measurement of the inclusive $W^{\pm}$ and Z/gamma cross sections in the electron and muon decay channels in *pp* collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Phys.Rev.* **D85** (2012) 072004, arXiv:1109.5141 [hep-ex].

[68] G. P. Salam, "Towards Jetography", *Eur.Phys.J.* **C67** (2010) 637–686, arXiv:0906.1833 [hep-ph].

[69] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k_t jet clustering algorithm", *JHEP* **0804** (2008) 063, arXiv:0802.1189 [hep-ph].

[70] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur.Phys.J.* **C72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[71] **ATLAS** Collaboration, "Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV", *The European Physical Journal C* **73** no. 3, (2013) . http://dx.doi.org/10.1140/epjc/s10052-013-2304-2.

[72] "Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s} = 8\ TeV$", Tech. Rep. ATLAS-CONF-2015-017, CERN, Geneva, Apr, 2015. http://cds.cern.ch/record/2008678.

[73] **ATLAS** Collaboration, "Local Hadronic Calibration", Tech. Rep. ATL-LARG-PUB-2009-001, CERN, Geneva, Jun, 2008. https://cds.cern.ch/record/1112035.

[74] "Pile-up subtraction and suppression for jets in ATLAS", Tech. Rep. ATLAS-CONF-2013-083, CERN, Geneva, Aug, 2013. http://cds.cern.ch/record/1570994.

[75] "Jet global sequential corrections with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", Tech. Rep. ATLAS-CONF-2015-002, CERN, Geneva, Mar, 2015. http://cds.cern.ch/record/2001682.

[76] **ATLAS** Collaboration, "ATLAS jet etmiss uncertainties TWIKI", https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2013JESUncertainty. Accessed: 16/05/15.

[77] **ATLAS** Collaboration, "Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data", Tech. Rep. ATLAS-CONF-2011-102, CERN, Geneva, Jul, 2011. http://cds.cern.ch/record/1369219.

[78] "Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment", Tech. Rep. ATLAS-CONF-2014-004, CERN, Geneva, Feb, 2014. http://cds.cern.ch/record/1664335.

[79] **ATLAS** Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV with atlas", *The European Physical Journal C* **72** no. 1, (2012) . http://dx.doi.org/10.1140/epjc/s10052-011-1844-6.

[80] "Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV", Tech. Rep. ATLAS-CONF-2013-082, CERN, Geneva, Aug, 2013. https://cds.cern.ch/record/1570993.

[81] "Determination of the tau energy scale and the associated systematic uncertainty in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector at the LHC in 2012", Tech. Rep. ATLAS-CONF-2013-044, CERN, Geneva, Apr, 2013. http://cds.cern.ch/record/1544036.

[82] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **1006** (2010) 043, arXiv:1002.2581 [hep-ph].

[83] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **0711** (2007) 070, arXiv:0709.2092 [hep-ph].

[84] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **0411** (2004) 040, arXiv:hep-ph/0409146 [hep-ph].

[85] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, *et al.*, "New parton distributions for collider physics", *Phys.Rev.* **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[86] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput.Phys.Commun.* **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[87] A. Denner, L. Hofer, A. Scharf, and S. Uccirati, "Electroweak corrections to lepton pair production in association with two hard jets at the LHC", arXiv:1411.0916 [hep-ph].

[88] K. Arnold, M. Bahr, G. Bozzi, F. Campanario, C. Englert, *et al.*, "VBFNLO: A Parton level Monte Carlo for processes with electroweak bosons", *Comput.Phys.Commun.* **180** (2009) 1661–1670, arXiv:0811.4559 [hep-ph].

[89] K. Arnold, J. Bellm, G. Bozzi, M. Brieg, F. Campanario, *et al.*, "VBFNLO: A Parton Level Monte Carlo for Processes with Electroweak Bosons – Manual for Version 2.5.0", arXiv:1107.4038 [hep-ph].

[90] J. Baglio, J. Bellm, F. Campanario, B. Feigl, J. Frank, *et al.*, "Release Note - VBFNLO 2.7.0", arXiv:1404.3940 [hep-ph].

[91] J. M. Campbell, R. K. Ellis, P. Nason, and G. Zanderighi, "W and Z bosons in association with two jets using the POWHEG method", *JHEP* **1308** (2013) 005, arXiv:1303.5447 [hep-ph].

[92] S. Hoche, F. Krauss, M. Schonherr, and F. Siegert, "NLO matrix elements and truncated showers", *JHEP* **1108** (2011) 123, arXiv:1009.1127 [hep-ph].

[93] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, "High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO", *Phys.Rev.* **D69** (2004) 094008, arXiv:hep-ph/0312266 [hep-ph].

[94] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, *et al.*, "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)", *JHEP* **0101** (2001) 010, arXiv:hep-ph/0011363 [hep-ph].

[95] J. Pumplin, D. Stump, J. Huston, H. Lai, P. M. Nadolsky, *et al.*, "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **0207** (2002) 012, arXiv:hep-ph/0201195 [hep-ph].

[96] P. M. Nadolsky *et al.*, "Implications of CTEQ global analysis for collider observables", *Phys.Rev.* **D78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[97] J. M. Campbell and R. K. Ellis, "An Update on vector boson pair production at hadron colliders", *Phys.Rev.* **D60** (1999) 113006, arXiv:hep-ph/9905386 [hep-ph].

[98] **ATLAS** Collaboration, "Top working group recommendations", https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TopMC12DiTopSamples. Accessed on 26/05/15.

[99] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation", *Phys.Lett.* **B710** (2012) 612–622, arXiv:1111.5869 [hep-ph].

[100] B. Kersevan and E. Richter-Was, "The Monte Carlo event generator AcerMC and package AcerDET",. https://cds.cern.ch/record/941506.

[101] N. Kidonakis, "Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production", *Phys.Rev.* **D83** (2011) 091503, arXiv:1103.2792 [hep-ph].

[102] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W or H boson", *Phys.Rev.* **D82** (2010) 054018, arXiv:1005.4451 [hep-ph].

[103] **ATLAS** Collaboration, "Measurements of the w production cross sections in association with jets with the atlas detector", *The European Physical Journal C* **75** no. 2, (2015) . http://dx.doi.org/10.1140/epjc/s10052-015-3262-7.

[104] L. Lyons, "Statistics for nuclear and particle physicists", 1986.

[105] F. E. James, *Statistical Methods in Experimental Physics; 2nd ed.* World Scientific, Singapore, 2006. https://cds.cern.ch/record/1019859.

[106] **ATLAS** Collaboration, "ATLAS electron guidelines TWIKI", https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaCalibrationGEO20. Accessed on 09/06/15.

[107] **ATLAS** Collaboration, "ATLAS muon guidelines TWIKI", https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MCPAnalysisGuidelinesData2012. Accessed on 09/06/15.

[108] **ATLAS** Collaboration, "Determination of the jet energy scale and resolution at ATLAS using Z/γ-jet events in data at $\sqrt{s} = 8$ TeV", Tech. Rep. ATL-COM-PHYS-2014-791, CERN, Geneva, Jul, 2014. https://cds.cern.ch/record/1741697.

[109] **ATLAS** Collaboration, "ATLAS 2012 jet uncertainties TWIKI", https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/JetUncertainties2012Final. Accessed: 17/05/15.

[110] **ATLAS** Collaboration, "Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$TeV with the ATLAS detector", *The European Physical Journal C* **74** no. 8, (2014) . http://dx.doi.org/10.1140/epjc/s10052-014-3023-z.

[111] **ATLAS** Collaboration, "ATLAS jet vertex fraction tool TWIKI", https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JVFUncertaintyTool. Accessed on 09/06/15.

[112] **ATLAS** Collaboration, "Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC", *The European Physical Journal C* **73** no. 8, (2013) . http://dx.doi.org/10.1140/epjc/s10052-013-2518-3.

[113] T. Sjöstrand, P. Edén, C. Friberg, L. Lönnblad, G. Miu, *et al.*, "High-energy physics event generation with PYTHIA 6.1", *Comput.Phys.Commun.* **135** (2001) 238–259, arXiv:hep-ph/0010017.

[114] M. Bahr *et al.*, "Herwig++ Physics and Manual", *Eur.Phys.J.* **C58** (2008) 639–707, arXiv:0803.0883 [hep-ph].

[115] S. Alekhin, S. Alioli, R. D. Ball, V. Bertone, J. Blumlein, *et al.*, "The PDF4LHC Working Group Interim Report", arXiv:1101.0536 [hep-ph].

[116] M. Whalley, D. Bourilkov, and R. Group, "The Les Houches accord PDFs (LHAPDF) and LHAGLUE", arXiv:hep-ph/0508110 [hep-ph].

[117] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128, arXiv:1106.0522 [hep-ph].

[118] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **0206** (2002) 029, arXiv:hep-ph/0204244 [hep-ph].

[119] P. Govoni and C. Mariotti, "Z Production via Vector Boson Fusion at LHC", arXiv:1001.4357 [hep-ph].

[120] D. L. Rainwater, R. Szalapski, and D. Zeppenfeld, "Probing color singlet exchange in Z + two jet events at the CERN LHC", *Phys.Rev.* **D54** (1996) 6680–6689, arXiv:hep-ph/9605444 [hep-ph].

[121] C. Gutschow, *First observation of electroweak Z boson plus two jet production.* PhD thesis, University College London, 2014.

[122] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", *eConf* **C0303241** (2003) MOLT007, arXiv:physics/0306116 [physics].

[123] F. James, "MINUIT", http://seal.web.cern.ch/seal/snapshot/work-packages/mathlibs/minuit/.

[124] F. James, *Statistical Methods in Experimental Physics.* second ed., 2006.

[125] **CMS Collaboration** Collaboration, "Measurement of the cross section of the electroweak production of a W boson with two jets in pp collisions at sqrt(s) = 8TeV", Tech. Rep. CMS-PAS-SMP-13-012, CERN, Geneva, 2015. http://cds.cern.ch/record/2033044.

[126] **ATLAS** Collaboration, G. Aad *et al.*, "Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", arXiv:1506.00962 [hep-ex].

[127] "Search for the Standard Model Higgs boson decay to $\mu^+\mu^-$ with the ATLAS detector", *Physics Letters B* **738** no. 0, (2014) 68 – 86. http://www.sciencedirect.com/science/article/pii/S0370269314006583.