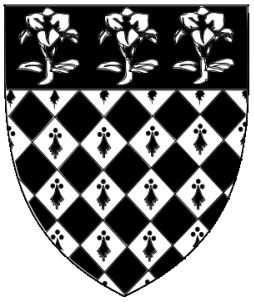
A measurement of the ratio of the $W + 1$ jet and $Z + 1$ jet cross sections using the ATLAS detector at the LHC.

James Christopher Buchanan

Magdalen College, Oxford



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

University of Oxford, Hilary Term 2012

Abstract

The first measurement of the ratio of the W and Z cross sections in association with a single jet, known as the R_{Jets} measurement, is presented. The measurement was performed using 33pb^{-1} of integrated luminosity, recorded during the year 2010 by the ATLAS detector at the LHC. At this time the LHC was operating at a center-of-mass energy of $\sqrt{s} = 7$ TeV. The measurement is made as a function of the threshold on the jet transverse momentum, from 30 GeV up to a value of 200 GeV. The motivation for this measurement is outlined in terms of providing a stringent test of the Standard Model of particle physics, as well as a model independent tool for searching for new physics. Data driven tools are developed to perform the measurement and their performance is discussed. The result of the measurement is compared with the predictions of Next-to-Leading order perturbative QCD and found to be in good agreement over the entire range of transverse momenta considered.

To my friends and family.

Acknowledgements

First and foremost I wish to thank Pierre-Hugues Beauchemin. Despite many important commitments and obligations, such as convening the ATLAS WZ physics group, he has always been willing to put aside what he is doing and discuss my work with me when I have questions or concerns, often for long periods of time or at unusual hours of the day (and night). I have learned so much as a result of his supervision. I would also like to thank his wife Genevieve du Paul, and their children Marine and Raphael for their wonderful company and hospitality. I also wish to thank Georg Viehhauser. Although the work I did with him is not reflected in this thesis, I feel I learned many new skills in the two years I worked for him and I genuinely hope that the result of that work proved valuable. I was very fortunate during my doctorate to have two such supportive and friendly supervisors.

I would also very much like to thank Tony Weidberg, Amanda Cooper-Sarkar, Alan Barr, Claire Gwenlan, Todd Huffman, Richard Nickerson and Roy Wastie for their support, both professionally and personally, over the past four years.

Many people at CERN assisted with the undertaking of various aspects of the work documented herein. Amongst them I would particularly like to thank Stefan Ask, Tanya Sandoval, Maarten Boonekamp, Giovanni Siragusa and Renaud Bruneliere, all of whom invested a significant amount of time assisting me. I also wish to thank everyone in the CERN R_{Jets} group, all of whom have helped me in some respect over the past few years: Jose E. Garcia-Navarro, Gerhard Brandt, Zhijun Liang, Elina Berglund, Rudi Apolle, Thorsten Kuhl, Marisa Sandhoff, Andrew Meade, Sofia Chouridou, Diego Casadei, Jason Nielsen and Benjamin Brau.

I wish to thank the other students I have worked with at Oxford, I have learned a lot from all of them and greatly enjoyed their company at coffee time over the last few years. Alex Pinder gets a mention first as he bribed me with tea and a kit-kat, after that I wish to thank: Sam Whitehead, Caterina Doglioni, Dan Short, Chris Boddy, Aimee Larner, Gemma Wooden, Carly Mattravers, Robert Apsimon, Ryan Buckingham, Nick Ryder, Ellie Davies, Ellie Dobson, Chris Young, Rob King, Adrian Lewis and Sarah Livermore.

Last but by no means least I wish to thank Pavel Avgustinov, Vicki Blackmore, Hugo Martay, Stef Simmons, John Dunn and Alastair Stewart. You folks have basically kept me sane over the last decade.

Contents

List of Figures

List of Tables

Glossary

ΔR The Euclidean distance between two objects in $\eta - \phi$ space.

ATLAS A Toroidal LHC Apparatus. A large general purpose detector at the LHC.

BSM Beyond Standard Model.

CP Symmetry Charge-Parity Symmetry. The symmetry operation of converting matter to anti-matter and performing a spatial inversion.

ECAL The ATLAS Electromagnetic Calorimeter.

EF The ATLAS Event Filter. The last stage of the trigger chain where full detector granularity and offline analysis tools are used.

etcone20 The amount of transverse energy in a cone of $\Delta R < 0.2$ around an electron after the subtraction of the transverse energy of the electron cluster.

HCAL The ATLAS Hadronic Calorimeter. This surrounds the ECAL.

FSR Final State Radiation. Radiation from the outgoing partons after the hard scattering.

GRL The Good Run List. A list of the run periods during which all required subdetectors were operational.

ID The ATLAS Inner Detector. The ID consists of the pixel detector, the SCT and the TRT and is responsible for reconstructing the tracks of charged particles.

ISR Initial State Radiation. Radiation from the incident partons prior to the hard scattering.

JER Jet Energy Resolution.

JES Jet Energy Scale. The response of the ATLAS calorimeters to hadrons is different to its response to electrons. Jets reconstructed at ATLAS are at the EM scale and therefore require calibration to be restored to the hadronic scale. In 2010 this was achieved by a Monte Carlo based numerical inversion process known as the Jet Energy Scale.

JVF The Jet Vertex Fraction. The fraction of the track momentum of a jet which is associated with a primary vertex.

L1 The ATLAS hardware based Level 1 Trigger.

L2 The ATLAS software based Level 2 Trigger.

LAr Liquid Argon.

LEP The Large Electron Positron Collider.

LO Leading Order. This refers to predictions obtained using only the lowest order diagrams in perturbation theory.

MC Monte Carlo. This usually refers to a quantity obtained from the use of simulation.

MCFM Monte Carlo for FeMtobarn Processes. A Next-to-Leading order partonic Monte Carlo generator.

OTx The optical transmitters used to read out the detector.

NLO Next-to-Leading Order. This refers to predictions obtained by the inclusion of the next to lowest order diagrams in perturbation theory. These diagrams typically involve the inclusion of loops or radiative emissions in the final state.

PDF Parton Distribution Function. A function giving the probability density that a particular partonic constituent of the proton will be observed in a collision at a scale Q^2 to have a particular momentum fraction x of the incident proton.

PMT Photomultiplier Tube.

pQCD Perturbative QCD. pQCD refers to predictions obtained from QCD via the use of perturbation theory. Such predictions are only valid at suitably high scales where the strong coupling is small compared with unity.

QED Quantum Electrodynamics. The quantum theory of the electromagnetic interaction, which is mediated by the photon, γ.

QCD Quantum Chromodynamics. One of the four fundamental interactions in nature. The name given to the strong sector of the Standard Model.

R_{Jets} The variable constructed as the ratio of the cross sections for the production of W and Z vector bosons in association with jets.

ROI Region of Interest. The coordinates of a cluster which passes one of the L1 trigger thresholds. This information is passed to the L2 trigger.

SM The Standard Model.

SCT The Semiconductor Tracker. One of the three subdetectors comprising the ATLAS Inner Detector (ID).

SPS The Super Proton Synchrotron. A 450 GeV proton-proton collider at CERN. Now part of the LHC injection system.

SUSY Supersymmetry. A possible extension of the Standard Model which requires the existence of bosonic partners for the fermionic matter particles and vice-versa.

TRT The Transition Radiation Tracker. The outermost subdetector of the ATLAS Inner Detector (ID).

UE Underlying Event. The component of a proton-proton interaction due to the constituent partons not involved in the hard-scattering which must then subsequently organise themselves into colour singlet final states.

Chapter 1

Introduction

The Standard Model of particle physics provides an elegant and precise framework with which three of the four interactions presently known to exist can be understood. All of the data collected by high energy physics experiments to date can be explained in terms of this framework[1]. Its predictions have been tested at LEP and the Tevatron as well as other experiments and found to be in excellent agreement with observed results. Despite these successes there are empirical and theoretical reasons for believing that new physics must exist beyond the Standard Model and that this new physics may manifest itself at the TeV scale. Examples of the former kind include incompatibilities with the observed abundances of dark matter and dark energy, while the hierarchy problem and the strong CP problem are examples of the latter kind.

The construction of the Large Hadron Collider, a proton-proton synchrotron with a nominal center-of-mass energy of 14 TeV, allows the Standard Model to be probed at higher energies and correspondingly smaller distance scales than ever before. These higher center-of-mass energies allow the creation of more massive particles than has previously been possible. If such particles are observed it would give important information on how to construct or refine new physics models to address the problems outlined above. The LHC also affords an opportunity to achieve a better understanding of the Standard Model itself. In particular,

[1] With the key exception of recent neutrino oscillation data, which can be accommodated via straightforwards extensions of the model.

the predictions of perturbative QCD (pQCD) will be able to be compared with measurements in successively higher kinematic regimes. These predictions are sensitive to higher order corrections, both directly and through the running of the coupling constants, and consequently these comparisons will aid in the understanding of the strong force at new scales. Predictions are also sensitive to non-perturbative QCD effects, such as the underlying event, and our current understanding of these effects comes mostly through the use of phenomenological models. Precise measurements of QCD processes at the LHC will aid in the tuning of these models. A solid understanding of QCD at the new scales accessible by the LHC is not only beneficial in its own right, but is crucial in order to reliably determine the existence of new physics at high energies.

The above goals of the LHC physics program make final states which contain vector bosons and jets particularly important, both because they allow a stringent comparison with the predictions of pQCD as described above and also because they are backgrounds to a variety of other processes of interest, both within and beyond the Standard Model, most notably top quark and Higgs boson production. The total cross sections for W and Z/γ^* vector boson production in association with at least n jets, where $n \leq 4$, have been measured at CDF [4] [5] and D0 [6] [7]. The Tevatron experiments also measured differential cross sections as a function of the n^{th} leading jet transverse momentum for each value of n. In both cases the results were compared with the predictions of Next-to-Leading order QCD calculations and observed to be in good agreement. Measurements of this kind have more recently been performed by the ATLAS collaboration where similarly good agreement between data and theory has been observed [8] [9]. Such measurements suffer from significant experimental uncertainties related to the identification and the energy scale of the jets in the event. For instance in the CDF measurement of $\sigma(W(\rightarrow e\nu) + \geq n\text{jets})$ the systematic uncertainty on the cross sections due to the jet energy measurement increases with jet transverse energy from 5% to 20% and is dominant for low values of the jet E_T. In the case of the corresponding D0 measurement the jet energy scale uncertainty contributes between 3% and 16% to the cross section uncertainty, the jet energy resolution contributes between 0.1% and 17% and the jet identification efficiency contributes between 0.3% and 4%. Taken together

the jet systematics are again the dominant contribution to the total systematic uncertainty on the measurement. As a consequence of this, it has recently been suggested that measurements of the total cross sections for the production of W and Z vector bosons in association with jets be augmented by measurements of the ratio of these quantities [10]. Such ratios have been demonstrated to benefit from the reduction of experimental systematics related to the identification and energy scale of the jets as well as theoretical systematics related to the choice of scales and the PDFs. From this point it is then natural to question whether ratios of the differential distributions also have such benefits.

This thesis therefore presents a new measurement, known as the R_{Jets} measurement, which is defined as the ratio of the $W + n$ jet and $Z + n$ jet cross sections when calculated as a function of a variable related to the kinematics of the jets in the event. The aim of this measurement is twofold: Firstly the measurement is a test of the strong sector of the Standard Model. It is constructed specifically to enable comparisons to be made between theoretical predictions and measurements at progressively higher energies. Secondly, it provides a model independent method for searching for new physics beyond the Standard Model at a variety of scales with high precision. The construction of the measured observable as a ratio will be shown to afford a drastic reduction of the systematic uncertainties on the measurement as the dominant sources of uncertainty on the numerator and denominator are highly correlated. Notably, the measurement does not rely on precise knowledge of the integrated luminosity as this quantity completely cancels in the ratio. This allows a more precise comparison with the predictions of pQCD and a more robust variable with which to search for new physics.

In this work the R_{Jets} measurement is performed in the electron channel with exactly one associated jet up to a threshold transverse momentum of 200 GeV. With the data available by the end of 2010 and with this limited kinematic reach there is little expectation of seeing new physics, instead the initial focus is to develop and understand the tools required to perform measurements of this kind as accurately as possible and to obtain an early but still very precise comparison with theory. This measurement was performed by the R_{Jets} group within the ATLAS collaboration and consequently not all of the material presented

is the work of the author. The main instances where this is the case are the following: the $E_{\mathrm{T}}^{\mathrm{miss}}$ based W QCD background subtraction and the Z background subtraction, which are the work of R. Apolle, the electroweak background subtraction, which is the work of M. Sandhoff and the trigger efficiency calculation, which was performed by E. Berglund. The official ATLAS paper documenting the measurement described herein, together with the corresponding measurement in the muon channel, can be found here [11].

After providing an overview of the theoretical framework of the Standard Model in chapter 2 and of the ATLAS detector and how it reconstructs quantities relevant to this measurement in chapter 3, R_{Jets} will be formally defined in chapter 4 and a more technical description of its merits will be given. Chapter 4 also describes the procedure for implementing the R_{Jets} measurement at ATLAS and correcting the observed result for detector effects. Chapter 5 discusses the derivation of a theoretical prediction for R_{Jets} using Next-to-Leading Order perturbative QCD as well as the theoretical uncertainties which must be considered in making such a prediction. The chapters following this detail the various stages of the R_{Jets} measurement and the tools developed for implementing them. Results and conclusions are given in chapter 9.

Chapter 2

The Standard Model of Particle Physics

The Standard Model represents our current best understanding of the interactions of the fundamental constituents of our universe. These constituents are summarised in figure 2.1 and can be grouped into three generations of spin 1/2 particles, each of which contains a pair of quarks with electric charges $+2/3$ (u, c, t) and $-1/3$ (d, s, b) respectively, as well as a pair of leptons with charges of -1 (e, μ, τ) and 0 $(\nu_e, \nu_\mu, \nu_\tau)$. There are also twelve spin one vector bosons responsible for mediating the three forces manifest in nature[1]. The massless electrically neutral photon (γ) can be exchanged between any charged particles resulting in an electromagnetic force. The 8 coloured gluons (g), which are also massless and neutral, may be exchanged between quarks and mediate the strong force. The remaining three vector bosons, the $W^{\pm}$ and Z, are massive and their exchange is responsible for the short range weak force. All fermions may engage in the weak force.

The remainder of this chapter provides a brief overview of the mathematical formalism behind the Standard Model, starting with an explanation of gauge theory and moving on to a qualitative description of spontaneous symmetry breaking and renormalization. These will not be covered in any depth, a full treatment can be found here [13]. Some of the

[1]Gravity has so far not been incorporated into this framework.

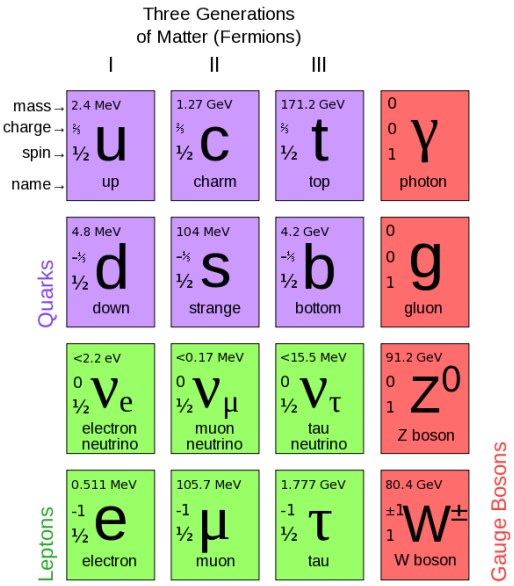


Figure 2.1: A diagram showing the particles comprising the Standard Model. For each of the spin-half particles present in this diagram there is a corresponding anti-particle with opposite sign charges. This diagram is reproduced from [12].

experimental validations of the Standard Model will also be discussed. Following this there will be a discussion of the limitations of the Standard Model and some candidate scenarios for new physics beyond the Standard Model will be introduced which offer potential explanations for these deficiencies and which may be detectable using the R_{Jets} measurement at the LHC. Finally, the phenomenology relevant to the measurement being presented will be introduced. This will involve a description of the structure of collision events at pp colliders as well as a discussion of the particular final states of interest, those involving W and Z vector bosons in association with jets, made in the context of the goals outlined in the introduction.

2.1 Formal Structure

2.1.1 Gauge Invariance

After QED was established as the quantum theory of electromagnetism [14] [15], it was observed that the structure of the interactions between charged fermions and the photon could be obtained by demanding that the Dirac Lagrangian describing the fermions be

invariant under local $U(1)$ transformations of the fermion fields [16] [17]. In order to achieve this local invariance it was necessary to introduce an additional spin one vector field which also transformed in a defined way under the $U(1)$ transformation, the photon field. The way the photon field enters the Lagrangian is constrained by Lorentz invariance and the $U(1)$ gauge invariance to give the correct form for the interaction with the fermion field.

Yang and Mills provided a generalised formulation for the construction of an interacting field theory based on invariance of the Lagrangian under the action of a non-abelian group $G = SU(N)$ with the aim of providing a similar construction for the strong force [18]. In this work an N dimensional multiplet of Dirac spinor fields ψ_i ($i\epsilon\{1, ..., N\}$) was introduced which transforms according to the fundamental representation of G:

$$\psi \rightarrow \psi' = U\psi \tag{2.1}$$

where U is a determinant 1, $N \times N$ unitary matrix. It was then observed that the Dirac Lagrangian for the multiplet of fermions

$$\mathscr{L} = \bar{\psi}_i(i\gamma^\mu\partial_\mu - m)\psi_i \tag{2.2}$$

could be made invariant under the action of G via the introduction of a new set of $N^2 - 1$ spin one vector fields, A^a_μ, which transform according to the equation

$$A_\mu \rightarrow A'_\mu = UA_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1} \tag{2.3}$$

$$A_\mu = \Sigma_a A^a_\mu T^a \tag{2.4}$$

followed by the replacement of the partial derivative in the Lagrangian with a gauge covariant derivative defined as:

$$D_\mu = I_N \partial_\mu - igT^a A^a_\mu. \tag{2.5}$$

In the above equations I_N is the N-dimensional identity, T^a is one of the generators of the group G, of which there are $N^2 - 1$, and g is a coupling constant which determines the strength of the interaction between the fermions and the A_μ fields. The generators T^a form a representation of the Lie Algebra of G and consequently satisfy the following relation:

$$[T^a, T^b] = i f^{ab}_c T^c \tag{2.6}$$

Where f^{ab}_c are the structure constants of the group G.

The newly acquired gauge fields must also be equipped with kinetic terms of their own which are quadratic in the field derivative and the simplest way of including such a term into the Lagrangian which respects the necessary symmetries is the following:

$$\mathscr{L}_{Yang-Mills} = -\frac{1}{4} F^{\mu,\nu,a} F^a_{\mu,\nu} \tag{2.7}$$

$$F^a_{\mu,\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^a_{bc} A^b_\mu A^c_\nu \tag{2.8}$$

This Lagrangian term will generate cubic and quartic terms in the gauge fields whenever the structure constants do not vanish and consequently gauge bosons of non-abelian field theories self-interact. Note that it is not possible to directly introduce mass terms for the gauge bosons as these terms, which are quadratic in the gauge fields, would not be invariant under the gauge symmetry. Introducing such a term destroys the perturbative renormalizability of the theory and more seriously leads to violations of unitarity in high energy scattering processes [13]. The fact that non-abelian gauge theories do not permit the inclusion of masses for the gauge bosons directly was originally seen as a major obstacle to the use of such theories in constructing realistic physical models, however the development of spontaneous symmetry

Representation	Particle	Hypercharge Y	$SU(2)_L$	$SU(3)_C$	$Q = T_3 + Y/2$
ψ_1	ν_L	-1	**2**, $T_3 = +1/2$	**1**	0
	e_L	-1	**2**, $T_3 = -1/2$	**1**	-1
ψ_2	e_R	-2	**1**	**1**	-1
ψ_3	u_L	$+1/3$	**2**, $T_3 = +1/2$	**3**	$+2/3$
	d_L	$+1/3$	**2**, $T_3 = -1/2$	**3**	$-1/3$
ψ_4	u_R	$+4/3$	**1**	**3**	$+2/3$
ψ_5	d_R	$-2/3$	**1**	**3**	$-1/3$

Table 2.1: This table shows the assignment of the standard model fermions to representations of the gauge group for one generation. The assignments for the second and third generations are the same. **1**, **2** and **3** indicate that a particle transforms under a singlet, doublet and triplet representation of the associated group respectively. The double lines separate particles according to the $SU(2)$ representation to which they belong, so for instance e_L and ν_L form a left-handed chiral $SU(2)$ doublet. Q is the electric charge.

breaking, which introduces masses while preserving the gauge invariance of the Lagrangian, overcame these concerns. It is now known that any self-consistent field theory of vector bosons which is renormalizable and unitary must have a gauge symmetry [19].

2.1.2 The Standard Model Lagrangian

The Standard Model is a gauge theory built upon the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$ where 'C' indicates the $SU(3)$ subgroup acts on the colour degrees of freedom of quarks, 'L' indicates the $SU(2)$ group acts on left-handed chiral multiplets and 'Y' denotes hypercharge. As described above, in order to ensure invariance of a fermion Lagrangian under the action of this group, a set of gauge fields must be introduced: a single field B_μ to accommodate the $U(1)$ hypercharge invariance, a triplet of fields W^a_μ for the weak isospin $SU(2)$ subgroup and an octet of fields G^a_μ for colour $SU(3)$. It will transpire that while the octet of coloured fields can instantly be associated with the gluons responsible for mediating the strong force, the other four fields cannot be associated directly with the photon and the W and Z bosons for reasons that will become apparent during the discussion of spontaneous symmetry breaking.

Prior to presenting the Lagrangian corresponding to this theory it is necessary to describe the matter content of the theory and its assignment to irreducible representations of the

group. This assignment is listed in table 2.1 for a single family of quarks and leptons, it is the same for the three families. The first thing to note is that the left handed and right handed components of the fermions are assigned to different representations of the weak isospin subgroup. Such a structure is required if one wants a theory in which parity is violated. Numerous empirical observations have been made of the non-conservation of parity in nature, most famously in the decay of Cobalt 60 in 1957 [20] and in the decays of kaons to pions as well as others [21]. A consequence of this is that it is no longer possible to directly introduce mass terms for the fermions as these terms would couple the left and right handed components of the matter fields and since these components no longer transform under the same representation of the group, such a term is no longer gauge invariant. This malady will be cured via the same mechanism that restores masses to the gauge bosons.

The Lagrangian for the Standard Model consists of four primary terms[2]:

$$\mathscr{L}_{SM} = \mathscr{L}_{Yang-Mills} + \mathscr{L}_{Dirac} + \mathscr{L}_{Higgs} + \mathscr{L}_{Yukawa} \tag{2.9}$$

The first term is the Yang-Mills kinetic term for the gauge fields. This consists of a sum of three terms of the form of equation 2.7 constructed from each of the three sets of gauge fields according to equation 2.8. In the case of the B_μ, the group is abelian so the structure constants are zero and the field has no self-coupling. For the W_μ and G_μ fields this is not the case. The most important consequence of this will be explored in section 2.1.3.

The second term is the Dirac kinetic term for the fermion fields. Having substituted the usual partial derivative for a covariant one to ensure the Lagrangian is gauge invariant, this term now also generates interactions between the fermions and the gauge fields. The form of this term is as follows:

[2]Gauge fixing terms and Ghost terms are also required to perform calculations but these will not be discussed here.

$$\mathscr{L}_{Dirac} = \Sigma_{r\epsilon\{\text{ representations }\}}\Sigma_{i\epsilon\{\text{ generations }\}}(\bar{\psi}_r^i i\gamma^\mu D_{\mu,r}\psi_r^i)$$
$$D_{\mu,r} = \partial_\mu - ig_Y\frac{Y_r}{2}B_\mu - ig_L T^a_{su(2),r}W^a_\mu - ig_s T^b_{su(3),r}G^b_\mu \tag{2.10}$$

In the above expression g_Y, g_L and g_s are the coupling constants corresponding to the $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$ groups respectively. $T^a_{su(2),r}$ are the generators of $SU(2)$ in the representation r. For $SU(2)$ singlets the generators are simply 0 while for doublet representations they are given by 1/2 of the corresponding Pauli matrices. Similarly $T^b_{su(3),r}$ are the generators of $SU(3)$, which are again 0 for colour singlets and equal to half the corresponding Gell-Mann matrices in the triplet representation. Y_r is the hypercharge of the representation r.

Spontaneous Symmetry Breaking

A discussion of the last two terms of equation 2.9 necessitates a discourse on spontaneous symmetry breaking. As was mentioned above, imposing a gauge symmetry and assigning left and right components of the fermion fields to separate representations of the group removes the possibility of directly introducing mass terms for either the gauge fields or the fermion fields as any such term would violate the symmetry. However we observe in nature that several of these fields are massive [22] [23] and consequently a mechanism for giving masses to these fields is required. In order that this mechanism does not remove the unitarity and renormalizability of the theory it must not destroy the gauge symmetry of the Lagrangian.

Spontaneous symmetry breaking achieves this by maintaining the symmetry of the Lagrangian but requiring that the vacuum state $|0\rangle$ is not invariant under the action of the gauge group [24] [25] [26]. Mathematically this implies that a subspace of the generators of the gauge group do not annihilate the vacuum state $T|0\rangle \neq 0$. The generators T for which this is true are said to be *broken*. This situation can be constructed by the inclusion of an additional field ϕ which has non-trivial group transformation properties and a non-zero

vacuum expectation value. The simplest way to achieve this within the Standard Model as presented so far is to posit that ϕ is a hypercharge $+1$ $SU(2)$ doublet of complex-valued Lorentz scalars with a Lagrangian

$$\mathscr{L}_{Higgs} = (D^\mu\phi)^\dagger(D_\mu\phi) - m^2|\phi|^2 + \lambda|\phi|^4 = (D^\mu\phi)^\dagger(D_\mu\phi) + V(|\phi|) \tag{2.11}$$

Where D_μ respects the same definition as above, although here ϕ does not couple to the strong sector as it is an $SU(3)$ singlet. This Lagrangian respects the gauge symmetry, however, for positive values of λ and m^2 the potential $V(|\phi|)$ has a minimum at a value given by $|\phi|^2 = m^2/2\lambda = v^2$. This corresponds to an infinite number of field configurations related by $SU(2)$ transformations of ϕ. Nature chooses one of these states to be the vacuum and this is what is meant by the term 'spontaneous breaking'. Given that there exists an $SU(2)$ group of transformations which act non-trivially on the vacuum, we have 3 broken generators and therefore expect a residual $U(1)$ gauge symmetry which will eventually be identified with the $U(1)$ of electromagnetism.

It can be shown that three of the four components of ϕ can be absorbed by an appropriate gauge transformation. As a result, the field ϕ can be expanded around the vacuum expectation value v in terms of a single physical field H, the Higgs field.

$$\phi = \begin{pmatrix} 0 \\ v + H \end{pmatrix} \tag{2.12}$$

To see how the introduction of this field generates masses for the gauge bosons we consider the kinetic term for the field ϕ. Inserting the above expression for ϕ expanded about its vacuum expectation value and retaining only the terms quadratic in the gauge fields one finds (derivation in appendix A):

$$(D^{\mu}\phi)^{\dagger}(D_{\mu}\phi) = \frac{v^2}{4}(g_L^2(W^{1,\mu}W_{\mu}^1 + W^{2,\mu}W_{\mu}^2) + (B^{\mu}, W^{3,\mu})\begin{pmatrix} g_Y^2 & -g_Y g_L \\ -g_Y g_L & g_L^2 \end{pmatrix}\begin{pmatrix} B_{\mu} \\ W_{\mu}^3 \end{pmatrix}) + \cdots \tag{2.13}$$

It is immediately clear from the appearance of bi-linear terms in the Lagrangian that the two gauge fields W^1 and W^2 have acquired a mass $m_W^2 = \frac{(vg_L)^2}{4}$. The matrix of coefficients of bi-linear terms in the remaining two fields is both real and symmetric and consequently can be diagonalised via an appropriately chosen rotation of the fields:

$$A_{\mu} = B_{\mu}\cos\theta_W + W_{\mu}^3\sin\theta_W \tag{2.14}$$

$$Z_{\mu} = -B_{\mu}\sin\theta_W + W_{\mu}^3\cos\theta_W \tag{2.15}$$

On performing the diagonalisation (appendix A), one finds that the eigenvalue corresponding to the field A_{μ} is zero while the eigenvalue corresponding to the Z_{μ} field is $(g_Y^2 + g_L^2)$ resulting in a mass for this field of $m_Z^2 = \frac{v^2(g_Y^2+g_L^2)}{4}$. $tan(\theta_W)$ is found to be equal to g_Y/g_L.

Having performed the above redefinition of the gauge fields in the Higgs sector of the Standard Model it is also necessary to apply it elsewhere in the Lagrangian, notably in the interaction term between the gauge fields and the fermions, $\mathscr{L}_{Dirac}$. The terms which involve B_{μ} and W_{μ}^3 can be isolated as follows:

$$\mathscr{L}_{nc} = g_Y\bar{\psi}\gamma^{\mu}\frac{Y}{2}\psi B_{\mu} + g_L\bar{\psi}\gamma^{\mu}T^3\psi W_{\mu}^3 \tag{2.16}$$

Applying the rotation of the gauge fields discussed above, this can be rewritten as:

$$\mathscr{L}_{nc} = \bar{\psi}\gamma^{\mu}(g_L\sin\theta_W T^3 + g_Y\cos\theta_W\frac{Y}{2})\psi A_{\mu} + \bar{\psi}\gamma^{\mu}(g_L\cos\theta_W T^3 - g_Y\sin\theta_W\frac{Y}{2})\psi Z_{\mu} \tag{2.17}$$

Recalling that $\tan\theta_W = g_Y/g_L \Rightarrow g_L \sin\theta_W = g_Y \cos\theta_W = e$, a constant, we can see that the term in the A_μ field further reduces to:

$$e\bar{\psi}\gamma^\mu(T^3 + \frac{Y}{2})\psi A_\mu \tag{2.18}$$

Associating the operator $T^3 + \frac{Y}{2}$ with the electric charge operator Q we recover the standard form for the interaction between a fermion of charge Qe with the photon:

$$e\bar{\psi}\gamma^\mu Q\psi A_\mu \tag{2.19}$$

We can now similarly redefine the other gauge boson fields W^1 and W^2 in order to diagonalise the electric charge operator, Q, giving fields $W^\pm = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu)$ which have electric charges plus and minus 1 respectively.

We have now recovered the physical gauge boson fields. The three degrees of freedom in the original field ϕ, which were removed by a gauge transformation, have been absorbed into the longitudinal degrees of freedom of the three gauge fields $W^\pm$ and Z which have now acquired masses related to the vacuum expectation value of the ϕ field as above. We have been left with one massless field, the photon, corresponding to the unbroken U(1) symmetry of electromagnetism and we have gained one additional massive scalar field H, which has a mass[3] given by $2\lambda v^2$. The charged vector bosons couple left handed quarks and leptons within a single generation.

Yukawa Couplings

Although we now have masses for the gauge bosons, the Higgs mechanism as presented above has still not afforded masses to the fermions in the theory. Also there is currently no mechanism present for inter-generational mixings of the quarks, which would result in hadrons comprised of the lightest quarks in each generation being stable. This too is not

[3]This is determined by inserting the expression for ϕ in equation 2.12 into the potential in equation 2.11 and finding the coefficient of the H^2 term.

observed in nature, these hadrons decay via weak interactions to lighter states. The last term in equation 2.9 rectifies both of these problems. $\mathscr{L}_{Yukawa}$ comprises a set of terms which respect all of the required symmetries and couple left and right handed fermion fields between different generations with the field ϕ via a set of matrices h:

$$\mathscr{L}_{Yukawa} = -\bar{\psi}_1^i h_l^{ij} \psi_2^j \phi - \bar{\psi}_3^i h_u^{ij} \psi_4^j \tilde{\phi} - \bar{\psi}_3^i h_d^{ij} \psi_5^j \phi + \text{h.c.} \tag{2.20}$$

When the electroweak gauge symmetry is broken, these terms can be expanded about the vacuum expectation of the field ϕ and will result in mass terms for the fermions as well as a set of couplings to the Higgs. In the case of the leptons, the matrices h can simply be diagonalised without affecting the rest of the Lagrangian, however for the quarks this is not the case. Redefining the fields so as to be mass eigenstates according to $\mathscr{L}_{Yukawa}$ changes the structure of the interaction term resulting in quarks from different generations interacting via charged currents according to a unitary matrix known as the Cabbibo-Kobayashi-Maskawa matrix [27]. The CKM matrix can be written as a relationship between the mass eigenstates of the (d, s, b) quarks and those linear combinations which couple directly with the $W^{\pm}$ vector bosons. Equation 2.21 gives the current measured values of the magnitudes of the elements of the CKM matrix [28].

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97428 \pm 0.00015 & 0.2254 \pm 0.0007 & 0.00347^{+0.00016}_{-0.00012} \\ 0.2252 \pm 0.0007 & 0.97345^{+0.00015}_{-0.00016} & 0.0410^{+0.0011}_{-0.0007} \\ 0.00862^{+0.00026}_{-0.00020} & 0.0403^{+0.0011}_{-0.0007} & 0.999152^{+0.000030}_{-0.000045} \end{pmatrix} \tag{2.21}$$

With the addition of the Yukawa term the formulation of the Standard Model is complete. Both the fermions and the gauge bosons have acquired masses directly related to the vacuum expectation value of the field ϕ and all of the allowed interactions are derivable directly from the Lagrangian. The values of the free parameters in the theory must be determined empirically.

2.1.3 Renormalization and Running Couplings

A good description of the shapes of the distributions of most quantities observable at hadron colliders requires predictions at an accuracy beyond that afforded by the leading order term in perturbation theory. However when one attempts to calculate higher order predictions using the Standard Model it is immediately clear that several contributions to the matrix element for a given process are formally divergent. These divergences can be categorized as *infrared* or *ultraviolet* in origin. The former arise in QCD parton emission processes in the limit where the parton is emitted soft or collinear to its parent, as well as in various QCD loop diagrams in the limit that the energy of the virtual particle tends to zero. The Kinoshita-Lee-Nauenberg [29] [30] and Bloch-Nordsieck theorems [31] prove that these divergences cancel between diagrams for well defined cross sections[4]. The latter arise from loop diagrams in the limit where the unconstrained momentum of a circulating virtual particle tends to infinity. These divergences do not cancel and require a process known as *renormalization* be carried out in order to render them finite [32]. Renormalization qualitatively involves absorbing these divergences into the bare parameters of the Lagrangian by fixing the parameters to a value obtained by measurement at a particular scale. The most striking consequence of this procedure is that the couplings of the three different interactions: g_Y, g_L and g_s acquire a dependence on the scale of the process under consideration, Q^2 [32]. The evolution of each parameter with scale is governed by the renormalization group flow equation as follows:

$$\frac{\partial g}{\partial \ln(Q^2)} = \beta(g) \tag{2.22}$$

where β can be evaluated as a perturbative series in the coupling g. Values of the couplings are therefore quoted at a particular scale Q^2. In the context of the LHC this is usually chosen to be M_Z, the mass of the Z boson.

The electromagnetic coupling has a value small compared to unity at low scales with

[4]It is not true that all of the infrared divergences cancel in this way. Soft and collinear splittings of the initial state partons give rise to divergences for which it is not the case, these are instead factorized into the parton distributions functions (see section 2.4).

$\alpha_{em} = e^2/4\pi \simeq 1/137$ and increases gradually to a value of $\simeq 1/127$ at the Z peak. By contrast the strong coupling is smaller at high scales, having a value of $\alpha_s = 0.1184$ at the Z peak and increasing rapidly as Q^2 decreases [3]. The consequence of this is that QCD processes become non-perturbative below scales ~ 1 GeV. The origin of this opposing behaviour is rooted in the non-abelian character of the strong interaction and the contributions to the perturbative expansion of β from gluon loop corrections. The corresponding diagrams are not allowed in QED as the photon cannot self-interact.

The strong dependence of the QCD coupling on the scale of the process under consideration and the marked change in the resulting phenomenology that this can cause is one of the key reasons why it is important to test our understanding of the strong sector of the standard model over a range of different scales.

2.2 Experimental Validations of the Standard Model

During its development the Standard Model has made several successful predictions which have been validated at a succession of different high energy physics experiments. Prior to the discovery of the J/ψ meson independently at the Brookhaven National Laboratory [33] and the Stanford Linear Accelerator Center [34] in 1974, the existence of the charm quark was predicted by Glashow, Illiopoulos and Maiani [35] in order that the model could accommodate the empirical observation that flavour changing neutral currents were suppressed. The existence of the third generation of quarks was also predicted shortly before the observation of the J/ψ to provide an explanation for observed CP violation in kaon decays [36]. The bottom quark was discovered shortly afterward in 1977 at Fermilab [37] although the top quark was not discovered until 1995 at the Tevatron due to its large mass [38] [39].

While charged current weak interactions had been observed prior to Glashow, Weinberg and Salam introducing electroweak unification, weak neutral currents had not and their existence was therefore also a prediction of the theory. These were first observed in the Gargamelle bubble chamber at CERN in 1973 [40]. The weak vector bosons were subse-

quently discovered by the UA1 and UA2 experiments at the 450 GeV Super Proton Synchrotron (SPS) at CERN in 1983 [22] [41] [23] [42]. The masses and widths of these particles were then measured precisely by the LEP experiments and found to be consistent with the Standard Model [43]. Several experiments, including the ALEPH experiment at LEP, have measured the behaviour of the QCD coupling at different scales and found its behaviour to be in agreement with the predictions of pQCD [44]. Precision measurements of the decay width of the Z also allowed a determination of the number of neutrinos, which was found to be 2.982 ± 0.013, again consistent with the Standard Model [45].

Other experiments such as Babar and Belle have tested the extent of CP violation in b quark systems and obtained results consistent with the predictions of the Standard Model [46] [47]. They have also made precise measurements of particular elements of the CKM matrix.

2.3 Beyond the Standard Model

2.3.1 Problems with the Standard Model

Although the Standard Model has made many successful predictions, there are still several issues which remain unresolved. First and foremost, the Higgs boson has not yet been empirically observed. While the λ parameter and consequently the mass of the Higgs are unknown, there is an upper bound on the Higgs mass ~ 1.4TeV imposed by unitarity considerations in various vector boson scattering processes. Consequently the Higgs, if it exists, should be accessible at LHC energies (see chapter 3).

The second perceived problem is theoretical in nature but still pertains to the Higgs boson. When one attempts to calculate higher order corrections to the squared Higgs mass one finds that these corrections are quadratic in the cutoff Λ_c (The energy scale at which the field theory is no longer expected to be valid). In the absence of any new physics this cutoff would be the Planck scale [48]. Consequently, for the Higgs to have a physical mass

at the scale required for the Standard Model to retain internal consistency, the bare mass must cancel these corrections to an extremely high level of precision. This 'fine-tuning' is considered undesirable in a theory. This problem is known as the hierarchy problem as it is related to the large difference between the electroweak and Planck scales.

There are other problems related to the agreement between the Standard Model of particle physics and the Standard Model of cosmology. These should ultimately agree but at present the Standard Model as described here cannot account for the observed abundance of dark matter in the universe or correctly predict the value of the cosmological constant. An obvious related problem is the lack of any account of quantum gravity within the Standard Model. There are other aesthetic concerns such as the sheer number of free parameters within the theory (19).

2.3.2 Supersymmetry

Supersymmetry offers one possible remedy for the hierarchy problem discussed above. SUSY is constructed from the only consistent extension of the Poincare algebra (The algebra of the group of Lorentz transformations and spacetime translations) to include internal symmetries [48]. The generators of these additional symmetries mix fermionic and bosonic states and consequently, if SUSY exists and is unbroken, one would expect a fermionic partner for each boson and vice-versa with the same quantum numbers. As this is not observed, SUSY would have to be broken at low energies and the various models introduced for achieving this give rise to different phenomenologies. If this breaking mechanism gives rise to superpartner masses below the TeV scale, SUSY has some very appealing theoretical features. Firstly it resolves the problem of the quadratically diverging Higgs mass through a cancellation in the quantum corrections arising from the corresponding bosonic and fermionic degrees of freedom. Secondly it allows for a unification of the three gauge couplings at higher energy which would be compelling evidence for a more fundamental theory of nature within a larger unifying gauge group equipped with a single coupling.

SUSY scenarios also frequently require conservation of a quantity known as R parity.

This is a multiplicative quantum number which is 1 for Standard Model particles and -1 for their superpartners. Imposing this requirement has the consequence that the lightest supersymmetric particle is stable and therefore provides a natural dark matter candidate, thus addressing another of the concerns described above.

Phenomenologically one expects SUSY processes to pair produce superpartners which then undergo cascade decays producing Standard Model particles until the lightest SUSY partner is produced. Experimentally this produces signatures with leptons, jets and large missing transverse momentum.

2.3.3 Leptoquarks

Several Grand Unified Theories have been proposed which attempt to embed the Standard Model gauge group into a larger group, providing some breaking mechanism which is responsible for the larger symmetry not being apparent at those energies thus far probed. Several examples of such models, for instance the Pati-Salam model built on $SU(4)_C \times SU(2)_L \times SU(2)_R$ [49], the Georgi-Glashow $SU(5)$ theory [50] and its embedding into $SO(10)$ include particles which have both quark and lepton quantum numbers and can therefore couple quarks and leptons either within or between generations. Such particles could potentially be pair produced in proton-proton collisions and would subsequently have decay modes directly to quarks and leptons, leading to final states containing leptons and jets.

There are several other classes of BSM models, most notably those involving large extra dimensions, however these do not typically have phenomenologies to which the measurement presented in the remainder of this work would be sensitive and they are consequently not discussed here. The ways in which the above BSM physics models may manifest themselves in the R_{Jets} measurement will be discussed in chapter 4.

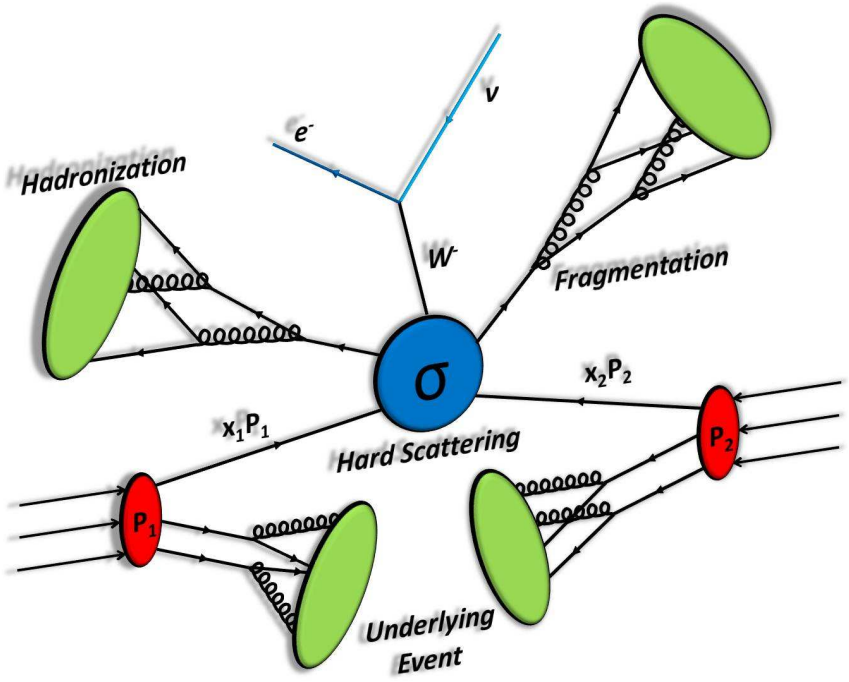


Figure 2.2: A schematic of the evolution of a proton-proton collision producing a W+jets final state.

2.4 Proton-Proton Collisions

The LHC is a proton-proton collider and consequently an understanding of the structure of proton-proton interactions is vital for the interpretation of data obtained by ATLAS. As protons are bound states of quarks, this structure is determined predominantly by the behaviour of the strong interaction at different scales.

Figure 2.2 gives a diagramatic representation of the development of a proton-proton collision. The two protons are shown as red ellipses, their structure prior to the collision is itself governed by low energy QCD processes and therefore cannot be understood perturbatively. Instead, our understanding of the proton is encoded into a series of functions known as parton distribution functions (PDFs), which give probability densities for finding a particular constituent of the proton (known collectively as *partons*), with a particular fraction of the incident proton's momentum x, when probed at a particular energy scale Q^2. These functions are obtained empirically from global fits to several sets of data from the study of

many different processes [51].

Typically during such a collision, one parton from each of the colliding protons participates in a hard scattering process at a large energy scale Q^2. This hard scattering process can be understood using perturbative QCD whenever the scale of the interaction is much larger than the scale at which the strong coupling constant is close to unity. For collisions producing final states of interest at the LHC this is usually the case. The hard scattering process is represented in the above diagram by a blue circle.

The products of this interaction, if they are themselves strongly interacting, will proceed to undergo fragmentation. This involves the emission of gluons and the pair production of further quarks, gradually decreasing the relative energies of the constituents of the shower from the original hard scale Q^2 down to the scale Λ_{QCD} where the strong force becomes non-perturbative again. Once this is the case, the partons present in the shower order themselves into colour singlet final states through a process known as *hadronization*, this is represented by green ellipses in figure 2.2. Again this process cannot be formally understood using QCD and is instead understood through the use of phenomenological models. Once colour singlet hadrons have been formed from the parton shower, any unstable mesons or baryons will decay producing the stable final state of the interaction.

There is one further complication to this picture which is due to the remaining partons from the original protons in the collision. These too undergo fragmentation and hadronization leading to a further non-perturbative component to the interaction known as the *underlying event*. This too is only currently understood through the use of phenomenological models tuned to data.

This factorization of the structure of the proton-proton collision into separate stages allows a mechanism for the calculation of cross sections whereby the hard-subprocess cross section of interest is convoluted with the parton density functions as follows [52]:

$$\sigma(\Omega) = \Sigma_{a\epsilon A, b\epsilon B} \int_{x_a, x_b, \Omega} f_{a/A}(x_a, \mu_F^2) f_{b/B}(x_b, \mu_F^2) \hat{\sigma}(\mu_R, \mu_F, x_a, x_b, \Omega') dx_a dx_b d\Omega' \qquad (2.23)$$

In the above equation x_a and x_b are the fractions of the incident proton momenta carried by partons a and b respectively, $\hat{\sigma}$ is the hard scattering cross section for the corresponding values of the fractional momenta, evaluated to some fixed order in perturbation theory and the function $f_i(x_i, \mu_F^2)$ is the PDF for parton i to carry a momentum fraction x_i when probed at the scale μ_F^2. The summations are carried out over all parton types present in the colliding protons and the integration is over the final state phase space of interest, denoted here by Ω. This expression is dependent on two scales, μ_R and μ_F, known as the renormalization and factorization scales respectively. The origin of these scales and their consequences in terms of the accuracy of numerical calculations are discussed in depth in section 5.2.

2.5 W and Z + Jets Final States

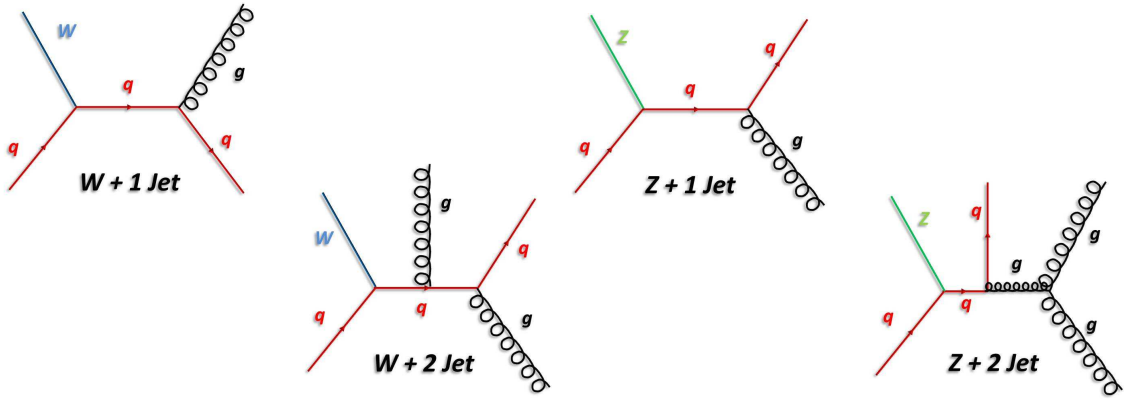


Figure 2.3: A selection of Feynman diagrams producing W and Z+jets final states. Time increases towards the top of the page.

Of particular relevance to the work presented here are those interactions where a weak vector boson is produced in the final state. As discussed in section 2.2, these final states have been extensively studied at previous colliders, their properties have been precisely measured and been found to be in agreement with the Standard Model. They produce experimental

signatures which are relatively easy to identify: an isolated lepton with missing transverse momentum from the neutrino in the case of the W, and two isolated leptons from the same generation in the case of the Z. Both of these facts make such final states excellent standard candles to use both for improving the current understanding of QCD and also for searching for new physics. There are of course possible decays to quarks in each case which result in pairs of energetic sprays of hadrons known as *jets*[5], but these final states are difficult to distinguish from QCD final states.

Interactions are also possible which produce weak vector bosons but also radiate high transverse momentum partons in addition and these interactions will give rise to final states with both the vector boson decay products and QCD jets. Each additional jet of similar energy comes with a cost $\sim 1/10$ in terms of cross section as determined by the value of α_s. It is these events where a vector boson is produced in association with one or more QCD jets which are the final states from which the R_{Jets} measurement is constructed. The events can be clearly tagged using the decay products of the vector boson and the structure can then be studied in terms of the kinematics of the associated jet(s). These final states are also sensitive to contributions from the new physics scenarios discussed in the previous sections. A selection of possible diagrams giving rise to vector boson + jets final states are shown in figure 2.3. The possible decays of the weak vector bosons and their measured branching ratios are listed in table 2.2 [3].

To gain an idea of what the numerical value of R_{Jets} should be according to the Standard Model, the inclusive W to Z ratio can first be considered. One can then ask what differences are expected once an additional jet is introduced. The partonic cross sections for the production of an on-shell W or Z vector boson are as follows [53]:

[5]A jet will be defined formally in section 3.3.2, for now we will simply understand them to be collimated showers of hadrons.

W Boson	
Decay Mode	Branching Fraction (Γ_i/Γ)
eν	10.75 ± 0.13 %
$\mu\nu$	10.57 ± 0.15 %
$\tau\nu$	11.25 ± 0.20 %
Hadrons	67.60 ± 0.27 %
Z Boson	
Decay Mode	Branching Fraction (Γ_i/Γ)
e^+e^-	3.363 ± 0.004 %
$\mu^+\mu^-$	3.366 ± 0.007 %
$\tau^+\tau^-$	3.367 ± 0.008 %
Invisible ($\nu\nu$)	20.00 ± 0.06 %
Hadrons	69.91 ± 0.06 %

Table 2.2: W and Z boson decay modes and branching fractions [3].

$$\hat{\sigma}^{q\bar{q'}\to W} = \frac{\pi}{3}\sqrt{2}G_F M_W^2 |V_{qq'}|^2 \delta(\hat{s} - M_W^2) \tag{2.24}$$

$$\hat{\sigma}^{q\bar{q}\to Z} = \frac{\pi}{3}\sqrt{2}G_F M_Z^2 (V_q^2 + A_q^2) \delta(\hat{s} - M_Z^2) \tag{2.25}$$

where $V_{qq'}$ is the appropriate element of the CKM matrix, and V_q and A_q are the couplings of the Z boson to quarks:

$$V_q = T_q^3 - 2Q_q \sin^2\theta_W, \tag{2.26}$$

$$A_q = T_q^3 \tag{2.27}$$

The Z couplings therefore take the values $(V_u, A_u) = (0.191, 0.5)$ and $(V_d, A_d) = (-0.345, -0.5)$ for up and down type quarks respectively. The total cross sections for each of these processes would be obtained by convoluting these subprocess cross sections with appropriate PDFs, as per equation 2.23, and summing over the contributing flavour combinations. For the purposes of this simple illustration, only the valence u and d quarks are considered and taken as having PDF values in the ratio 2:1. The expression for the ratio is then given by:

$$R = \frac{M_W^2 |V_{ud}|^2}{M_Z^2(\frac{2}{3}(V_u^2 + A_u^2) + \frac{1}{3}(V_d^2 + A_d^2))} \tag{2.28}$$

which yields a value of 2.34. If the ratio is measured with respect to a particular final state, as it will be in this case, the ratio of branching fractions to this final state must also be included. The ratio of the branching fractions for $W \rightarrow e\nu$ and $Z \rightarrow ee$ is 3.23, giving an overall value of $R = 7.55$. This calculation is approximate[6] and the value would be affected by any fiducial selections applied on the final state objects, but it illustrates how the couplings, masses, branching ratios and PDFs enter in principle and these factors are expected to carry over to the one jet case. The value of R_{Jets} is expected to be similar to the value presented here in magnitude and to be somewhat larger at low values of the jet transverse momentum due to the lower mass of the W with respect to the Z. As the jet transverse momentum increases, R_{Jets} should decrease before becoming flat once the scale of the interaction is much greater than the electroweak scale.

[6]The 2010 ATLAS measurement of the inclusive W/Z ratio yielded a value of 11.7 ± 0.9 (stat.) ± 0.4 (sys.) [54].

Chapter 3

The ATLAS Detector at the LHC

3.1 The Large Hadron Collider

In order to address the questions raised in the previous chapter concerning the domain of validity of the Standard Model and the possibility of various extensions to it, the Large Hadron Collider has been constructed at CERN near Geneva. The LHC is a twenty seven kilometre circumference proton-proton synchrotron with a design center-of-mass energy of 14 TeV. Under nominal running conditions the LHC would circulate two opposing beams of protons, consisting of a total of 2808 bunches each with up to 10^{11} protons, at a revolution frequency of roughly 11 kiloHertz. This would result in a design luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, with collisions occurring at four nominal interaction points every 25ns [55]. At this luminosity the average number of proton-proton interactions per bunch crossing is approximately 26. Any interaction giving rise to a signature of interest will therefore be accompanied by activity due to the additional 25 soft interactions in that bunch crossing. This additional activity (and interchangeably the interactions that give rise to it) is known as in-time pile-up. It is also the case that under the running conditions described above, the time resolution of several detector elements is larger than the time between successive bunch crossings. This can lead to additional 'out-of-time' pile-up, where activity from surrounding bunch crossings is recorded as belonging to the crossing of interest. Pile-up of both kinds can significantly

complicate the interpretation of data recorded at the LHC.

The four interaction points coincide with the four main experiments of the LHC: ATLAS and CMS, which are both general purpose physics experiments, LHCb which was specifically designed to study CP violation in the decays of b quark systems and ALICE, which was built primarily to understand the properties of quark-gluon plasmas resulting from collisions of lead ions in the LHC. A diagram of the LHC together with its injection system and the four experiments listed above, is shown in figure 3.1.

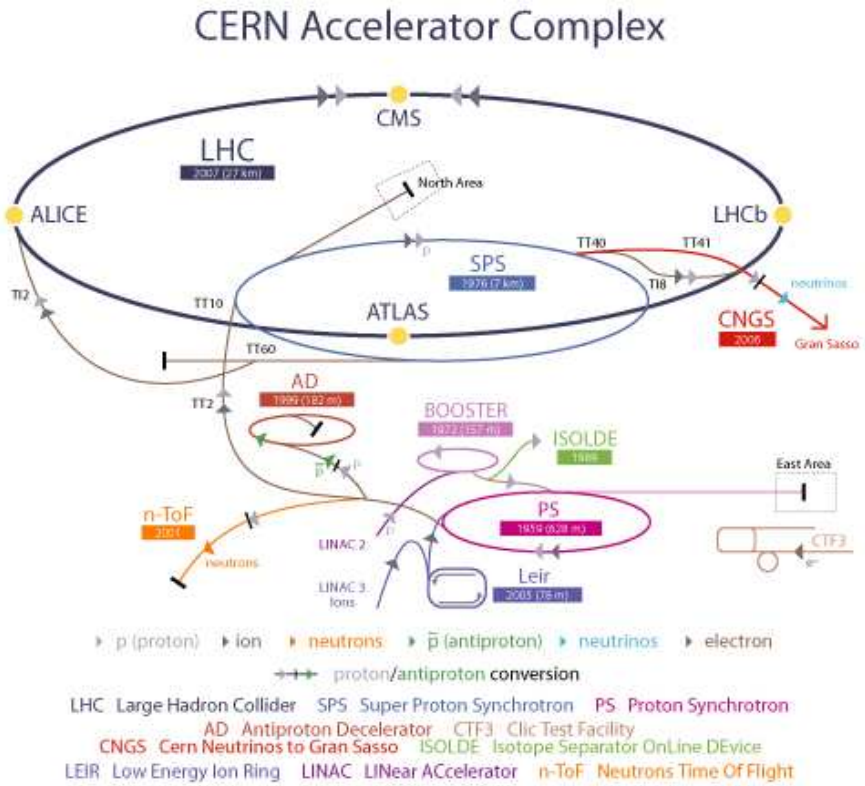


Figure 3.1: A schematic of the layout of the Large Hadron Collider. This figure shows the layout of the LHC and depicts the locations of the four main experiments as well as the injection system [56].

The work presented here utilises the data recorded with the ATLAS detector over the course of the year 2010. During 2010 the LHC was not operating at its full design specification, instead maintaining a fixed center-of-mass energy of 7 TeV for the full year[1]. The luminosity was also in the process of being ramped up during the course of the year, reaching a peak instantaneous luminosity of 2.1×10^{32} prior to the winter technical shutdown.

[1]Note that this still constituted a world record for the highest energy particle collisions yet achieved in an accelerator.

This corresponded to a peak mean number of interactions per bunch crossing of just under 3.8. The total integrated luminosity accumulated by ATLAS and the mean number of interactions per bunch crossing are shown as a function of time during 2010 in figure 3.2.

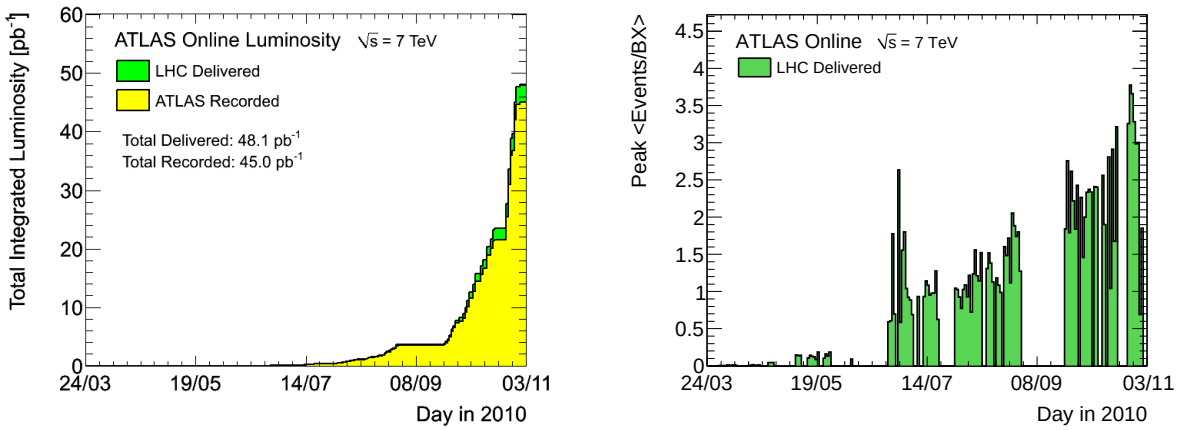


Figure 3.2: The total integrated luminosity delivered to and recorded by ATLAS (left) and the mean number of interactions per bunch crossing (right) during 2010 [1].

3.2 The ATLAS Detector

ATLAS is a large general purpose detector which has been specifically designed to meet the challenges presented by the unprecedented energy and luminosity of the LHC. As described above, under nominal conditions each bunch crossing will result in an average of around 26 proton-proton interactions, leading to a densely populated final state with of the order of 1000 particles produced. This environment calls for a high granularity detector with near hermetic coverage, good radiation hardness and excellent triggering in order to reduce the amount of recorded data to a manageable level [57].

The ATLAS detector is depicted in figure 3.3 and comprises a series of sub-detectors. The innermost volume of the detector is dedicated to reconstructing the tracks of charged particles and consists of a pixel detector nearest to the beamline, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). These three subsystems are collectively referred to as the inner detector (ID) and are enclosed in a solenoid which provides a field of 2 Tesla, allowing momentum and charge determination from the curvature of any recon-

structed tracks. Outside the tracking volume are the calorimeters. The innermost is the electromagnetic calorimeter (ECAL), which is a liquid argon sampling calorimeter with lead absorbers and is encased in a cryostat which it shares with the solenoid magnet. Outside of this is the hadronic tile calorimeter (HCAL), which comprises a series of steel absorber plates interspersed with plastic scintillators in the barrel region and employs liquid argon technology, similar to that used for the ECAL, in the endcaps. The calorimeters are surrounded by the muon spectrometer. The measurement presented here is performed in the electron channel and consequently the muon system is not discussed in detail. For a thorough exposition of all the ATLAS detector subsystems, refer to [58].

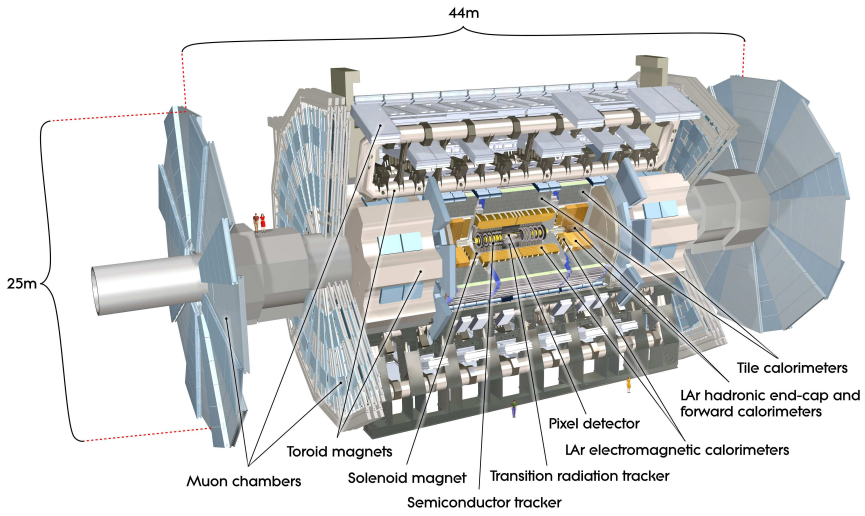


Figure 3.3: A schematic of the layout of the ATLAS detector showing the different subdetectors.

ATLAS employs two main coordinate systems for describing the kinematics of events and the positions of the components of the detector. In both cases the nominal interaction point is taken as origin. The first is a right handed Cartesian coordinate system with the positive z axis aligned with the beamline and oriented so as to point in the anti-clockwise direction of the LHC, the x axis pointing towards the center of the LHC ring and the y axis oriented vertically upwards. The second set, which is more commonly used when describing the kinematics of events, is related to the spherical polar coordinate system. This system comprises the set of coordinates (R, η, ϕ), where R is the radial distance from the beamline, ϕ is the azimuthal angle taken in the anticlockwise sense around the positive z direction

from the x axis and η is the pseudorapidity, which is defined in terms of the polar angle taken from the positive z direction as follows:

$$\eta = -\ln(\tan\frac{\theta}{2}) \tag{3.1}$$

Use of the pseudorapidity is advantageous, as differences in η are invariant under Lorentz boosts along the beam axis for massless particles. At various points, use of the variable ΔR will also be made with reference to two objects, somewhat counter-intuitively this refers to the Euclidean distance between the two objects in $\eta - \phi$ space and does not refer to radial separations in the above coordinate system.

3.2.1 The ATLAS Inner Detector

A schematic of the ATLAS inner detector is shown in figure 3.4. From this diagram it can be seen that the tracking volume covers the pseudorapidity range $|\eta| < 2.5$ and extends out to a radius of 1.150m. The inner parts of the tracking volume employ silicon semiconductor technology to reconstruct a set of precisely measured space points on the path of a charged particle, while at higher radii the TRT uses straw tubes to measure on average 36 values of $R - \phi$ along the particle's track. This information significantly improves the momentum resolution of the reconstructed track. The TRT however only extends to a pseudorapidity of 2.0. The solenoid magnet can also be seen in the diagram just behind the inner wall of the EM cryostat.

The Pixel Detector

As shown in figure 3.4, the ATLAS pixel detector consists of three barrel modules positioned at radii of 50.5mm, 88.5mm and 122.5mm from the beamline together with six endcap rings, three on either side of the barrel module, positioned at longitudinal distances of 495mm, 580mm and 650mm from the nominal interaction point respectively. The barrel modules

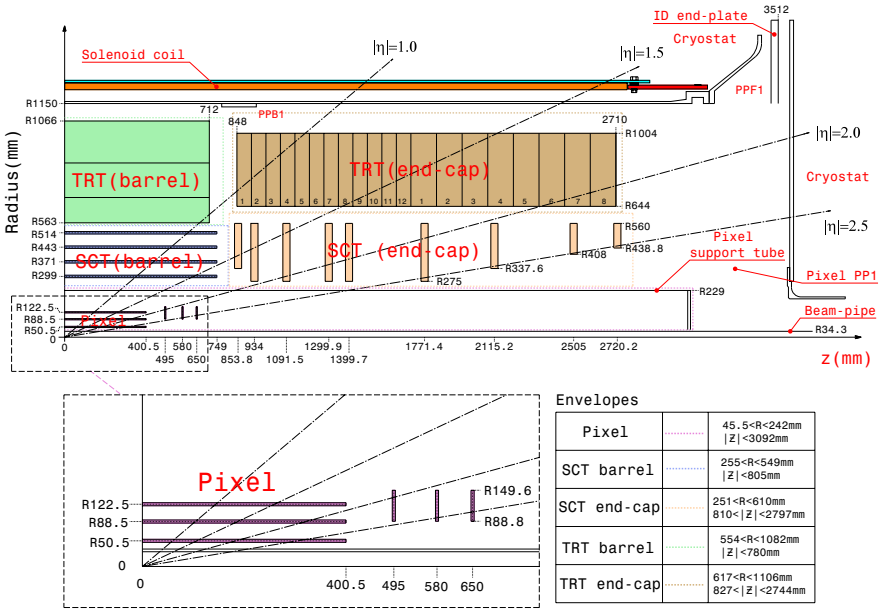


Figure 3.4: In this diagram one can see all the components of the ATLAS inner detector. Closest to the beampipe are the three pixel barrel layers, which are surrounded on each side by 3 pixel endcap rings. The pixel detector is itself surrounded by the semiconductor tracker (SCT) which comprises 4 barrel sections and 9 end cap sections on each side. The transition radiation tracker (TRT) extends from this point out to the electromagnetic calorimeter covering pseudorapidities out to 2.0.

consist of a total of 112 staves equipped with 13 identical pixel modules each, while the endcap rings consist of a total of 48 sectors each mounted with six modules, three on each side of each sector. This gives a total of 1744 pixel modules.

The pixel modules themselves consist of a flexible polyimide circuit board mounted on top of a layer of 47232 silicon pixels, each of which is individually bump bonded to a readout channel of one of 16 front end readout chips which underlay the active silicon layer[2]. Each pixel has dimensions of $50 \times 400\mu$m giving intrinsic spatial resolutions of 10μm in $R-\phi$ and 115μm in z (R) respectively in the barrel (endcap).

The Semiconductor Tracker (SCT)

The intermediate radii in the inner detector volume are occupied by the semiconductor tracker. This sub-detector has a similar structure to the pixel detector, comprising 4 barrel

[2]While most pixels have an individual readout channel, a small subset of the pixels on each sensor are ganged into groups of four.

modules, located between 299 mm and 563 mm in radius, and 9 rings per endcap, providing a minimum of 4 precision space-points per track within a pseudorapidity of 2.5.

The barrel modules consist of two pairs of daisy-chained 6cm long silicon strip detectors mounted onto a thermal pyrolitic graphite (TPG) baseboard, each with a stereo rotation of ± 20 mrad with respect to the center of the module in order to facilitate a measurement of the longitudinal position of any hit. The 780 strips per module each have a pitch of 80 μm [3]. Taking into account the stereo rotations of the strips, this gives intrinsic resolutions of 17 μm in $R-\phi$ and 580 μm in z. In order to achieve the required performance after irradiation, the sensor modules must be held at a temperature of -7 degrees Celsius with less than a 5 degree variation in order to keep leakage currents at an acceptable level and prevent thermal runaway. This requirement is met via the use of a C_3F_8 evaporative cooling system running at approximately -25 degrees. The sensors are thermally coupled to the cooling system via beryllia end facings on the TPG baseboards. These facings are mounted on aluminum blocks which are themselves soldered to the Cu/Ni cooling pipes.

The broad structure of the endcap modules is the same although the strips run radially and consequently have an average pitch of 80 μm. There are three types of endcap module depending on whether the module is to be mounted on an inner, middle or outer ring. The endcaps provide an intrinsic resolution of 17 μm in $R-\phi$ and 580 μm in R. Diagrams of a barrel SCT module and a middle endcap module are shown in figure 3.5.

The Transition Radiation Tracker (TRT)

Additional $R-\phi$ information is provided at the outer radii of the inner detector by the transition radiation tracker. The barrel TRT consists of up to 73 layers of straw tubes divided into 3 coaxial rings, each of which is itself constructed from 32 modules. The straws themselves are 4 mm in diameter and comprise a 31 μm gold plated tungsten anode surrounded by a cathode constructed from two layers of 35 μm thick polyimide film, coated

[3] The two strips which define the edge of the sensor are held at bias potential, there are therefore only 778 active strips per module. The sensors are oriented at 11 degrees with respect to the barrel tangent and consequently overlap in order to accommodate edge effects.

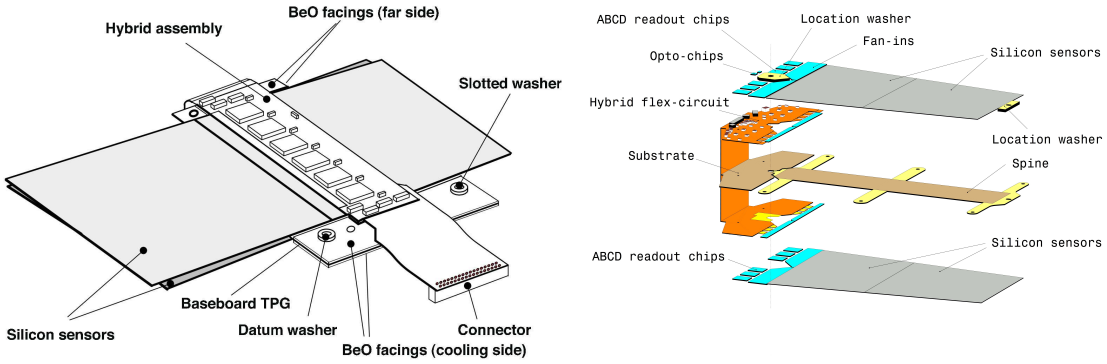


Figure 3.5: The left hand diagram shows a typical barrel SCT module. The silicon sensors with their relative rotation can be seen, as well as the TPG baseboard with its beryllia facings which interface with the cooling blocks. The right hand diagram shows an exploded view of a middle endcap module. The structure is similar, with the main exception that the flex hybrid on which the readout electronics are mounted is at the end of the module as opposed to being mounted over the top of the sensors as in the case of the barrel module.

with aluminum. The anode is kept grounded while the cathode is held at a fixed voltage of -1530 V. The straws are filled with a gas mixture of 70% xenon, 27% carbon dioxide and 3% oxygen. The resolution on the single straw position measurement (each straw affords a measurement of the radius of the interaction point from the anode) is set by the uncertainty on the drift time of the electrons and has been determined to be 130 μm. Any charged particle within the pseudorapidity range $|\eta| < 2.0$ will cross a minimum of 36 straws on passing through the TRT. The resultant position information significantly reduces the resolution on the measurement of the track momentum[4]. In the endcaps the TRT consists of two sets of wheels each of which consists of 8 layers of 768 radially oriented straws with each layer rotated by 3/8 of the straw separation angle. The first set contains 12 such wheels with each of the 8 layers separated by 8 mm, while the second set contains 8 wheels with each layer separated by 15 mm.

The TRT is also afforded some electron identification capability by the inclusion of 19 μm diameter polypropylene fibers between the barrel straws and 15 μm thick radiator foils between successive layers in the endcaps. The passage of electrons through these materials results in the creation of low energy transition radiation photons which have a much larger

[4]In the transition region between the barrel and endcap the minimum number of straws traversed drops to 22.

absorption cross section in the xenon gas than a minimum ionising particle would, resulting in a significantly enhanced signal. The TRT therefore operates two separate hit thresholds to distinguish between minimum ionising particles and transition radiation due to the passage of an electron.

3.2.2 Calorimetry

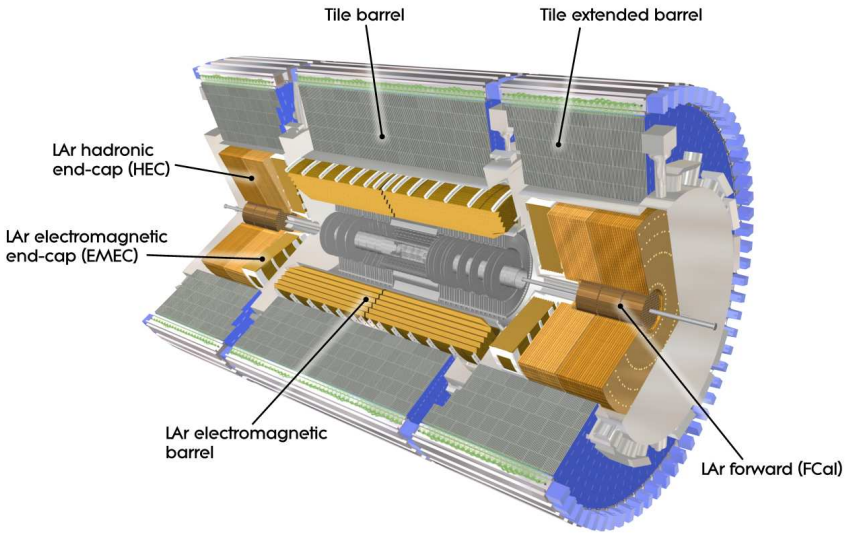


Figure 3.6: A diagram showing the ATLAS calorimetry. The electromagnetic calorimeter is shown in gold surrounding the inner detector. The hadronic calorimeter is shown in blue and surrounds the ECAL.

The Electromagnetic Calorimeter

The ATLAS calorimeters are shown in figure 3.6. The electromagnetic calorimeter (ECAL) is a sampling calorimeter constructed from a set of accordion shaped lead absorber plates immersed in liquid Argon (LAr). It consists of a barrel module which shares its cryostat with the ID solenoid and two endcaps which are housed in separate cryostats along with the hadronic endcaps and the forward detectors. In both the barrel and endcap calorimeters the lead plates lie along planes of constant ϕ, however in the barrel, the direction of propagation of the accordion waves is radial, whereas in the endcaps it is axial to best suit the momentum direction of incident particles. Interspersed between the lead absorber plates are sets of three

copper electrodes separated by polyimide insulating sheets. The two outer electrodes are held at high voltage and signals are read out from the central electrode via capacitive coupling.

As a particle passes through the ECAL, the presence of the absorbers results in the development of an electromagnetic cascade via Bremsstrahlung and the consequent pair production of electrons. These cascades lead to ionisation of the active liquid argon medium and result in a measurable current through the copper electrodes from which the energy of the incident particle can be deduced.

The barrel module is constructed from two half barrels each of which is 3.2 m is length and extends from 1.4 m to 2 m in radius. It covers the pseudorapidity region $|\eta| < 1.475$ and contains 2048 absorber plates in total. The barrel section is divided into three radial layers, the first of which is the strip layer and has an extremely high granularity in η. This layer allows the calorimeter to distinguish between isolated electrons or photons and pairs of photons produced from pion decays. The second layer is where the majority of the energy of an electromagnetic cluster is deposited, the cells in this layer have dimensions $\Delta\eta \times \Delta\phi = 0.025 \times 0.0245$. The third layer is coarser and measures the tail of electromagnetic showers. The segmentation of the ECAL in the barrel region is shown in figure 3.7. The depth of the barrel ECAL as a function of pseudorapidity, expressed in terms of radiation lengths, is shown on the left hand side of figure 3.8.

The EM endcaps each consist of two further wheels with a similar structure. The outer wheel, which covers the pseudorapidity range $1.375 < |\eta| < 2.5$ is also segmented into three layers in depth and contains 768 absorber plates, while the inner wheel, which covers $2.5 < |\eta| < 3.2$ is only divided into two sections longitudinally and contains 256 absorbers. The depth of the EM endcap calorimeters, similarly expressed in terms of radiation lengths, is shown on the right hand side of figure 3.8.

In order to understand the amount of energy lost by electrons and photons prior to reaching the ECAL, a liquid Argon presampler has been placed in front of the calorimeter over the pseudorapidity range $|\eta| < 1.8$.

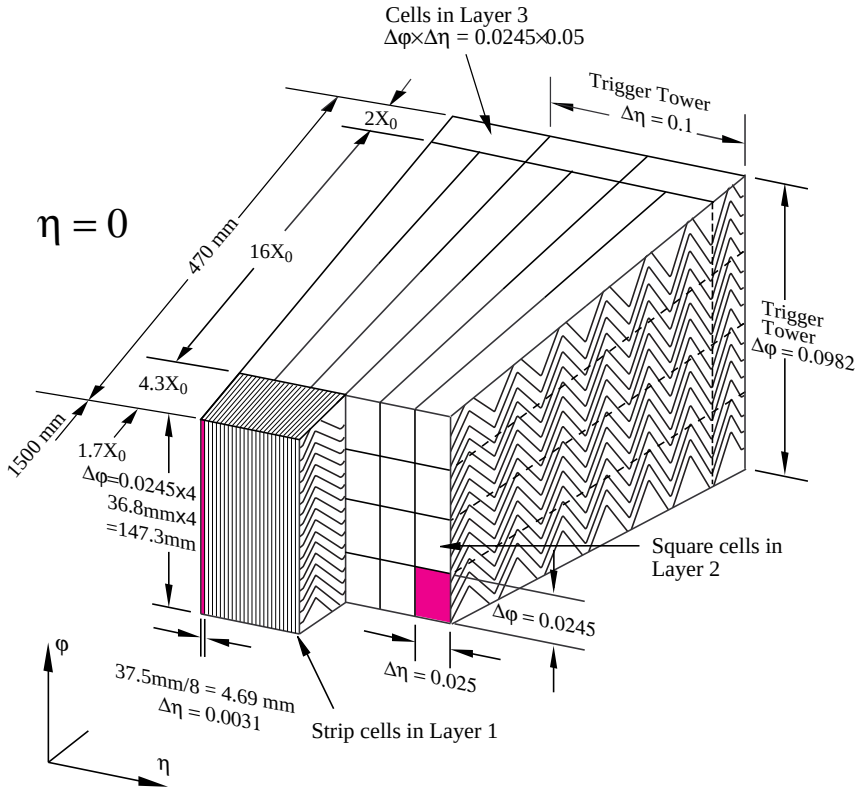


Figure 3.7: A diagram showing the longitudinal segmentation of the barrel ECAL. The accordion structure of the lead absorber plates is clearly visible as are the dimensions of the three longitudinal layers into which the absorbers are divided. The segmentation is angled so as to remain pseudo-projective.

The Hadronic Calorimeter

The hadronic calorimeter (HCAL), like its electromagnetic counterpart, is divided into barrel and endcap regions. The barrel region is itself further divided into a central barrel calorimeter of 5.8 m in length, and two extended barrel calorimeters which are a further 2.6 m long, one on each side of the central barrel. Each of these three sections is constructed from 64 wedge shaped modules subtending an azimuthal angle of $\Delta\phi \sim 0.1$. The inner and outer radii of the barrel hadronic calorimeter are 2.28 m and 4.25 m respectively.

Each of the barrel modules is constructed from a staggered arrangement of steel absorber plates interspersed with polystyrene scintillator tiles. As with the ECAL, interactions in the absorber sheets lead to the development of hadronic cascades. As ionising particles in the showers interact with the scintillator medium they induce the emission of ultraviolet scintillation light. This light is then collected in 1mm diameter wavelength shifting fibers

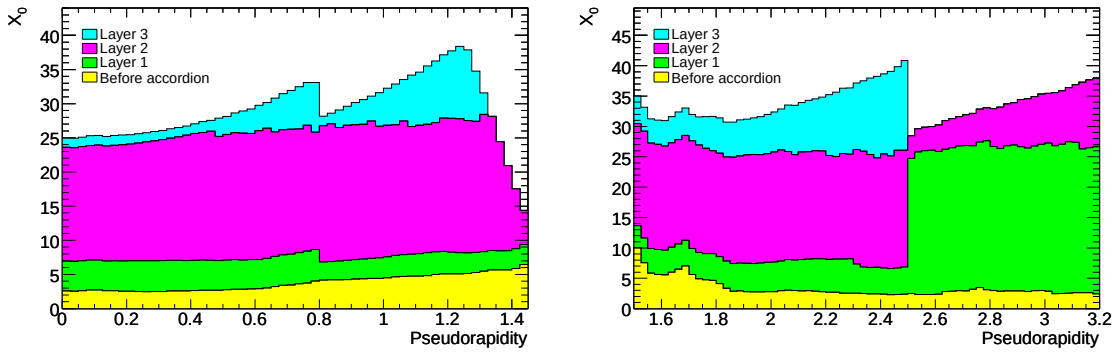


Figure 3.8: Diagrams showing the depth of the ATLAS Electromagnetic Calorimeter in units of the radiation length. The plots show the amount of material in each accordion layer as well as in front of the calorimeter for the barrel (left) and endcap (right) regions respectively.

which are arranged to press against the sides of the scintillator tiles. The fibers increase the wavelength of the collected light into the visible via interactions with dopant species in the polystyrene and deliver it to photomultipliers mounted in the outer support structure of the HCAL. The fibers occupy the gaps between the 64 modules which are designed to be 1.5 mm in width at the inner radius of the HCAL. The scintillator tiles are ganged together into the photomultipliers to form an almost projective geometry in η segmented into three radial layers. The inner two layers have dimensions $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, while the outer layer has coarser dimensions of 0.2×0.1.

Due to the increased hit rate and correspondingly higher rate of radiation damage as one approaches the forward region of the detector, the hadronic endcaps employ liquid argon technology, similar to the ECAL, instead of the scintillation technology used in the barrel. The hadronic endcaps share a cryostat with the electromagnetic endcaps and the forward calorimeters and consist of two wheels per endcap, each of which is divided into 32 wedge shaped modules and together cover a pseudorapidity range $1.5 < |\eta| < 3.2$. The front wheel is constructed from 24 copper plates oriented normal to the beamline, each of which is 25 mm thick, while the outer wheel is constructed from 16 plates each of which is 50 mm thick. The 8.5 mm thick gaps between the copper absorber plates are instrumented with three electrodes held at high voltage. The central electrode is used for signal readout and is divided into a series of pads with dimensions $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for $|\eta| < 2.5$ and $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ at higher $|\eta|$. Figure 3.9 shows diagrams of single modules in both

the barrel tile calorimeter and the endcap liquid argon calorimeter. Figure 3.10 shows the amount of material in each region of the hadronic calorimeter in units of the interaction length. The amounts of material in and in front of the EM calorimeter are also shown.

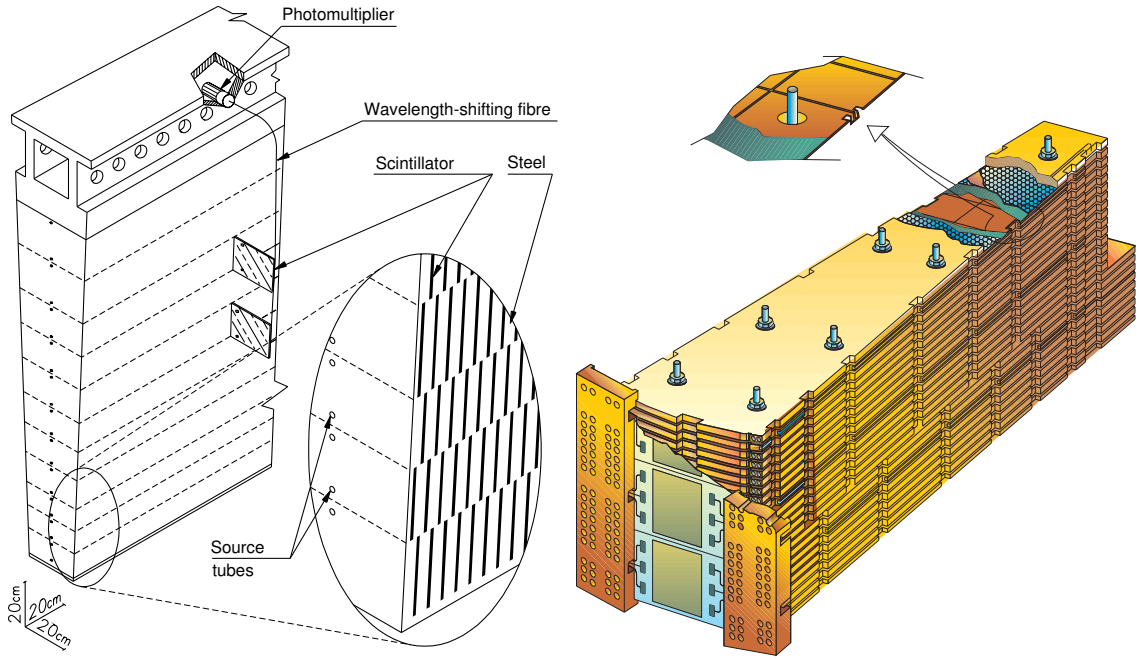


Figure 3.9: The left hand diagram shows a typical module in the barrel tile calorimeter. The pattern of absorbers and scintillation tiles is shown as well as the path of a wavelength shifting fiber, which connects to a PMT mounted at the outer end of the module support structure. The right hand diagram shows a view of an endcap module. The copper absorption plates are shown as well as a cutaway view of one of the readout electrodes with its etched pad structure.

3.2.3 The ATLAS Trigger System

The bunch crossing rate in ATLAS at the LHC design luminosity is 40.08 MHz and each event produces on average $\sim$ 1.3 MB of data. If all the data were stored this would result in data being written to tape at a rate of roughly 52 TBs^{-1} which is unsustainable both in terms of read-out and storage of the raw data, but also in terms of the resources required to actually analyse the stored data. Consequently a trigger system is required which maintains a high rejection rate whilst still being efficient for processes of interest. Trigger chains have been developed which search for high P_T electrons, photons, jets, muons and taus as well as large missing transverse momentum or total momentum. The triggers used in the

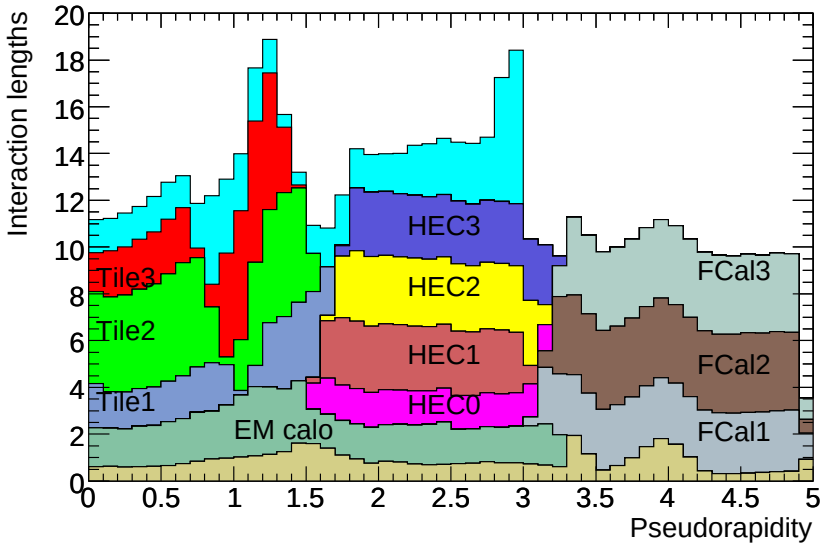


Figure 3.10: A diagram showing the amount of material in the different layers of the tile and endcap hadronic calorimeters in units of the interaction length. The diagram also shows the material depths of the electromagnetic calorimeter, the forward calorimeters and the total amount of active material in front of the muon system (pale blue).

analysis presented here rely on the identification of events containing high P_T electrons and consequently the other trigger chains will not be described in any detail.

The ATLAS trigger is divided into three main parts, the Level 1 trigger is hardware based and reduces the offline rate to 75 kHz in 2.1 μs. For the identification of electrons the L1 trigger uses reduced granularity information from the calorimeters. Cells are grouped into trigger towers of dimension $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ and the energies within these towers are summed. A sliding window algorithm is then run over the trigger towers which searches for two by two squares where any pair of adjacent cells within the square have a total energy which passes one of a particular set of programmable thresholds. To avoid possible double counting, the identified cluster is required to be a local maximum of the total energy within the four trigger towers (including the contributions from the hadronic calorimeter cells behind the ECAL).

The software-based Level 2 trigger receives the coordinates of the clusters identified by the L1 trigger, referred to more generally as Regions of Interest (RoI). The L2 trigger has access to the full granularity of the detector within the RoI and applies more stringent cuts,

reducing the rate to less than 3.5 kHz in a further 40 ms. The final stage of the ATLAS trigger is the event filter (EF) which applies offline analysis techniques to the full event and results in a final acceptance rate of 200 Hz in 4 s. The two triggers employed in this analysis are described in section 4.2 along with the other event selections.

3.3 Object Reconstruction in ATLAS

3.3.1 Electron Reconstruction

The reconstruction of central electrons in ATLAS ($|\eta| < 2.47$) is seeded by a cluster with a transverse energy of greater than 2.5 GeV. Clusters are located using a 3×5 cell sliding window algorithm where the individual cells have dimensions of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ to match the dimensions of the cells in the middle layer of the ECAL. Once a cluster has been located satisfying the transverse energy requirement, it is required to loosely match a track extrapolated from the inner detector volume. If multiple tracks are matched, those with silicon hits are prioritised, otherwise the track with the smallest Euclidean distance in (η, ϕ) space is chosen.

The energy of reconstructed electrons is then obtained by the further use of a sliding window algorithm with dimensions 3×7 and 5×5 in the barrel and endcaps respectively. The final energy is constructed from four components: the total energy in the sliding window; the estimated energy lost in front of the ECAL, calculated using the measurement from the LAr presampler when available; the estimated lateral leakage of the shower outside of the cluster and the estimated longitudinal leakage into the hadronic calorimeter. These contributions are obtained using parameterisations in terms of the energy in the sliding window obtained from Monte Carlo. The electron four momentum is then calculated using the measured energy from the above prescription and the η and ϕ values taken from the best matched track at the vertex. Electrons outside of the central region are reconstructed using topoclusters (see section 3.3.2) and have no track requirements imposed. Such electrons are not used in this

work and are not discussed further.

These candidate electrons are then subjected to a cut based identification procedure optimised in 10 bins of cluster η and 11 bins of cluster E_T. Three baseline sets of identification criteria have been defined: *loose*, *medium* and *tight*. The cuts involved in these criteria are fully specified in [59] and are only summarised here. The loose criteria place constraints on the amount of hadronic leakage of the cluster, defined in terms of the ratio between the first HCAL sampling energy and the energy of the electromagnetic cluster, as well as placing requirements on the lateral shower shape of the cluster in the second sampling of the EM calorimeter. The medium criteria additionally require a good track by demanding minimum numbers of hits in the different silicon detectors as well as requiring an improved agreement between the extrapolated track and the cluster barycenter. The medium criteria also exploit the measured shower shape in the strip layer of the ECAL. The tight criteria offer the largest background rejection by requiring a certain number of high threshold TRT hits, an even more stringent track quality and a hit in the inner pixel layer. The tight criteria additionally require an agreement of the cluster and track in terms of E/P and require that the track impact parameter be less than 1mm with respect to a primary vertex.

The different baseline identification criteria are expected to give background rejections of the order of 500, 5000 and 50000 respectively. The efficiencies of the different criteria in terms of signal vary with η and E_T and are discussed in depth in chapter 7.

3.3.2 Jet Reconstruction

As described in section 2.4, it is not possible to observe single quarks or gluons created in hard processes. Final state radiation and hadronisation lead to broad showers of hadrons being produced which leave large depositions of energy in the ATLAS calorimeters. In order to simply characterise these depositions and compare them with the results of simulations at different levels (matrix element, parton shower, post hadronisation and after full detector simulation) a prescription for clustering the depositions and reconstructing a composite four momentum is required, this prescription is known as a *jet algorithm*. The jet algorithm used

must be meaningful at all levels of calculation and must also satisfy a constraint known as *infrared safety* in order to be theoretically well defined [60]. This constraint requires that the final state produced by the jet definition be robust against collinear splittings of any of the input objects (be they partons from a matrix element calculation, final state hadrons or cells in the detector) and also against the addition of a further soft (low P_T) input object anywhere in the phase space. If these conditions are not met, the final state produced by the algorithm would be sensitive to small changes in the inputs, which is extremely undesirable. A further desirable feature of a jet definition is that the final state produced be similar at successive stages of simulation.

The class of jet definitions in common use within the ATLAS collaboration are known as sequential recombination algorithms [61]. These algorithms progressively cluster the input objects, which henceforth are referred to as pseudojets, according to a distance measure between the pseudojets. The other main class of jet definitions are cone algorithms but, with the exception of the SIScone algorithm, these are frequently infrared unsafe and so are not used [62]. The clustering prescription employed is as follows:

- For each pseudojet calculate the value of $d_i = p_T^{2n}$.
- For each pair of pseudojets calculate the distance defined by $d_{ij} = (\frac{\Delta R_{ij}}{R})^2 min(p_{T,i}^{2n}, p_{T,j}^{2n})$
- If one of the values of d_i is the smallest of the combined set (d_i, d_{ij}), remove the corresponding pseudojet and consider it a jet.
- If one of the values of d_{ij} is the smallest of the combined set, merge the two corresponding pseudojets into one pseudojet with a four momentum given by the sum of the four momenta of the two constituents.
- Repeat this prescription until all there are no remaining pseudojets.

So long as the final set of jets has a minimum transverse momentum imposed, this prescription will satisfy the condition of infrared safety. The value R in the distance measure is free and can be chosen by the user, it characterises the size of the jets that will be

produced by the algorithm. The value of n specifies one of three separate sub-categories of recombination algorithm:

- $n = 1$ corresponds to the k_T or Durham algorithm. This algorithm effectively inverts a P_T ordered parton shower, preferentially clustering around the lowest P_T object.
- $n = 0$ corresponds to the Cambridge-Aachen algorithm, which effectively inverts an angular ordered shower, clustering the nearest pseudojets first [63].
- $n = -1$ corresponds to the anti-k_T algorithm, which is the most widely used in ATLAS at present and will be utilised for the R_{Jets} measurement. It preferentially clusters objects close to the hardest remaining pseudojet first and consequently has a tendency to produce jets which are close to circular [60].

Prior to applying the jet clustering algorithm on calorimeter cells, a noise suppression procedure is implemented [64]. This procedure is seeded by calorimeter cells with energies greater than $4\sigma_{noise}$. For any given seed cell all immediate neighbour cells are added and are then considered as seeds for further cluster expansion if they have an energy greater than $2\sigma_{noise}$. Once this process has terminated, a splitting and merging step takes place to deal with any cells belonging to two clusters. The clusters formed by this sequence are known as *(4,2,0) topoclusters* and are in close correspondence with the energy depositions of individual hadrons. It is these objects which are input to the jet clustering algorithm.

Once the jet clustering algorithm has been run, the output jets still need to be calibrated due to the fact that the response of the calorimeters to hadrons is different to their response to electrons[5]. In order to correct the energy of the jets to the hadronic scale, a Monte Carlo based numerical inversion procedure has been developed. This is simply an energy and η dependent scale factor which, when applied to the jets in an event, restores the average jet energy to its expected value. It does not improve the resolution of the jets. The uncertainty on this procedure varies both with jet energy and detector region, having a value below

[5]Calorimeters which do not correct for the different response of electrons and hadrons are known as non-compensating.

2.5% for high energy, central jets ($|\eta| < 0.8$, 60 GeV $\leq P_T^{jet} <$ 800 GeV) and a value of up to 14% for low energy, forward jets ($3.2 \leq |\eta| < 4.5$, 20 GeV $\leq P_T^{jet} <$ 30 GeV) [65]. A full treatment of the jet energy scale correction (JES) and its associated uncertainties can be found here [66]. Other more sophisticated schemes for calibrating jets which will in principle improve the resolution are also being developed although these were not used for 2010 analyses.

3.3.3 Missing Transverse Energy Reconstruction

W and Z bosons have decay modes to neutrinos which propagate through the detector without interacting leading to an imbalance in the transverse momentum measured. As discussed in section 2.3, beyond Standard Model scenarios also frequently involve decays giving rise to large missing transverse momentum (E_T^{miss}), and the presence of large E_T^{miss} is therefore a widely used signal criterion in BSM searches. It is consequently important that ATLAS is capable of consistently reconstructing the E_T^{miss} in an event. Three algorithms for reconstructing the $E_{\mathrm{T}}^{\mathrm{miss}}$ are used during the course of this work, they are described in the following subsections.

LocHadTopo

The LocHadTopo $E_{\mathrm{T}}^{\mathrm{miss}}$ definition is conceptually the simplest and is constructed from topoclusters (see previous section) in the region $|\eta| < 4.5$. These clusters are then subjected to a calibration scheme known as 'Local Hadronic Calibration' which utilises properties of the clusters to correct the cluster energy to the hadronic scale. The transverse momentum of each cluster is then constructed using the calibrated energy and the angular coordinates of the cluster. The total calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ is then determined from the energies of the cells in the topoclusters as follows:

$$E_x^{miss} = -\Sigma_{i\epsilon\{topocells\}} \frac{E_i cos(\phi)}{cosh(\eta)}, E_y^{miss} = -\Sigma_{i\epsilon\{topocells\}} \frac{E_i sin(\phi)}{cosh(\eta)} \tag{3.2}$$

$$E_T^{miss} = \sqrt{E_x^{miss,2} + E_y^{miss,2}}, \phi^{miss} = arctan(\frac{E_y^{miss}}{E_x^{miss}}) \tag{3.3}$$

In events with muons further terms are added to accommodate their contribution to the $E_{\mathrm{T}}^{\mathrm{miss}}$ [67].

HadRecoil

This $E_{\mathrm{T}}^{\mathrm{miss}}$ definition is constructed from the LocHadTopo $E_{\mathrm{T}}^{\mathrm{miss}}$ definition above but with some modifications to improve the treatment of leptons in events with vector bosons [68]. After calibration of the topoclusters as described above, any topocells within a cone of ΔR <0.2 of a reconstructed lepton are removed from consideration. Instead of contributing the combined momentum of these cells to the above summation, the momentum components of the reconstructed lepton are used. This replacement procedure suffers from the complication that in addition to energy deposited by the lepton, energy due to the underlying event (UE) will also be removed from the summation and unlike the lepton energy, this contribution is not replaced by default. To correct for this, a cone of identical dimension is randomly selected at the same value of η as the removed lepton, and the transverse momentum within this cone is rotated so as to be aligned with the lepton direction and then included in the summation. Additional requirements are imposed on the distance of the UE replacement cone from both the leptons in the event and the direction of the hadronic recoil.

RefFinal

The RefFinal $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithm attempts to classify topoclusters according to the high-P_T objects to which they belong and then calibrates them according to this classification. Clusters are associated with high P_T objects in the following order: electrons, photons, hadronically decaying tau leptons, hard and soft jets and muons. Any topoclusters left unaccounted for are then subject to separate calibration. The total $E_{\mathrm{T}}^{\mathrm{miss}}$ is then calculated from the negative sum of all of the independently calibrated contributions. Details of the exact methods

for assigning clusters to these categories and the calibrations used for the different cluster classifications can be found in [69].

Chapter 4

R_{Jets} Measurement Design

4.1 Motivation and Definition

One of the first priorities with any new accelerator and detector is a rediscovery of the Standard Model. ATLAS has already performed measurements of the cross sections of the W and Z bosons [67] and the consistency of the obtained results with theoretical predictions provided early confidence in both the Monte Carlo tools and experimental techniques available for performing analyses involving vector bosons. These measurements, as well as measurements of the W asymmetry, have even allowed new constraints to be placed on parton distribution functions [70]. A measurement of the ratio of the W and Z cross sections was also performed. This measurement had the advantage that some sources of systematic uncertainty, for instance uncertainties in the energy scale of the leptons in the event, produced correlated variations in the numerator and denominator, and consequently the resulting systematic uncertainties on the ratio were reduced with respect to those on either of the cross sections. The ratio will therefore constitute a more stringent test of the Standard Model than either of the cross sections once the dominant source of uncertainty is systematic.

ATLAS has since performed measurements of the cross sections for production of vector bosons in association with jets and following the inclusive measurements it is natural to now consider ratios of these quantities. It is expected that such ratios, when suitably constructed,

will benefit from a similar reduction in the systematic uncertainties associated with the modelling of jets.

The quantity $R^{(n)}_{Jets}$ is therefore defined as follows:

$$R^{(n)}_{Jets}(x) = \frac{\sigma(W + nJets)(x)}{\sigma(Z + nJets)(x)} \tag{4.1}$$

In the above equation x indicates a kinematic variable constructed from the jets in the event. In the one jet case this could simply be the transverse momentum of the jet. An obvious generalisation to higher jet multiplicities would be to consider the scalar sum of the momenta of the jets.

As outlined in the introduction, there are two primary objectives which motivate the above definition of the $R^{(n)}_{Jets}$ variable. The first is to construct a measurement which allows a robust test of the Standard Model over a range of scales. The construction of the $R^{(n)}_{Jets}$ variable as defined above means that for any given value of x, the kinematics of the jets in W and Z events are extremely similar. The only differences are due to selection bias and the mass difference of the vector bosons, and the impact of both of these reduces as x increases. The similarity of the available phase space of the jets between the two types of event means that systematic variations of the way ATLAS reconstructs jets result in very similar variations of the cross sections in the numerator and denominator. The effect on $R^{(n)}_{Jets}$ is therefore considerably smaller than the effect of such variations on either of the component cross sections. This will be studied concretely later in this chapter. A similar logic applies to systematics related to the leptons in the event, just as it did in the inclusive case, although the degree of correlation should be smaller as a result of the differences in the structure of the W and Z decays. As described in section 2.5, the structure of the decays of the electroweak bosons are well understood and the products of these decays allow the relevant final states to identified relatively easily. The reduced uncertainties on $R^{(n)}_{Jets}$ with respect to the cross sections from which it is constructed allow for a more stringent comparison with the predictions of QCD as a function of the variable x. These predictions

in general require a good understanding of the behaviour of higher order QCD effects as well as several aspects of its non-perturbative behaviour, and consequently if the measurement uncertainties can be sufficiently reduced it may be possible to provide constraints on various aspects of QCD modelling with this measurement.

The second goal is to provide a model independent means for conducting searches for physics beyond the Standard Model. Any BSM scenario which allows for the production of massive particles which decay via cascades to final states with leptons, missing transverse momentum and jets in a ratio different from that predicted by the Standard Model would result in deviations between the measured and predicted values of $R^{(n)}_{Jets}$ at high x values characteristic of the scale of the new physics. Were such a deviation observed, it would then be possible to use the measured value of $R^{(n)}_{Jets}$ to provide constraints on candidate new physics models. The BSM models presented in section 2.3 are both potentially observable in this way. Figure 4.1 shows the effect on $R^{(2)}_{Jets}$ resulting from the addition of two example new physics scenarios as an illustration of what might be observed.

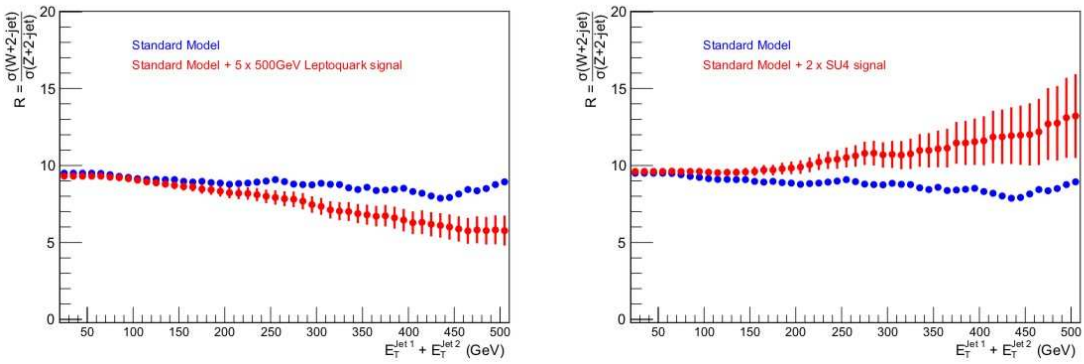


Figure 4.1: The effect on $R^{(2)}_{Jets}$ of two potential BSM scenarios. The left hand plot shows the deviation from the Standard Model expectation caused by the presence of a 500 GeV leptoquark. The right hand plot shows the effect of the SU4 SUSY benchmark point. The effects are exaggerated, the contributions from BSM physics being scaled by 5 in the case of the leptoquark signal and 2 in the case of the SUSY signal in order to emphasize the effect on R_{Jets}.

Each different value of n in equation 4.1 corresponds to a different event topology and will be sensitive to different new physics scenarios. Considering measurements of $R^{(n)}_{Jets}$ for different values of n should therefore allow strong constraints to be placed on any potential new physics.

It is important to note that these two objectives are not incompatible. It is possible to distinguish between deviations of the measured and predicted values of $R^{(n)}_{Jets}$ due solely to an improper modelling of QCD and due to potential new physics. This differentiation can be achieved in practice via the application of upper limits on the transverse and invariant masses of the W and Z bosons respectively. If these limits are chosen close to the electroweak scale, they would suppress the impact of any new physics on the measured value of $R^{(n)}_{Jets}$, allowing for an unbiased comparison with pQCD. In the presence of new physics one would expect a deviation between measured and predicted values of $R^{(n)}_{Jets}$ to appear at high x values once these cuts were relaxed. Continued good agreement at low values of x then gives confidence that the measurement techniques being employed are sound. Any significant deviation observed at low x would indicate a potential problem in the procedure for performing the measurement, either in calculating the backgrounds, the detector unfolding corrections or in evaluating their systematic uncertainties.

The rest of this work is devoted to the first measurement of this kind carried out, which corresponds to $n = 1$ in equation 4.1. The variable x was chosen to be the threshold on the leading jet transverse momentum. Thus all events where there is exactly one jet with a transverse momentum greater than x will contribute to the value obtained in a given bin. The measurement is effectively focusing in on successively higher regions of the available phase space. This affords increased statistics in any given bin but also results in strong bin by bin statistical correlations, these correlations should be considered when interpreting results.

It will be demonstrated in chapter 5 that for the $n = 1$ case considered here, the main systematic uncertainties on the Standard Model theoretical predictions of vector boson plus jets processes are also strongly correlated between W and Z events. Consequently predictions can be obtained with considerably smaller uncertainties for $R^{(1)}_{Jets}$ than they can be for either of the cross sections. In the context of using $R^{(1)}_{Jets}$ as a tool for understanding the Standard Model this allows for a more stringent comparison between measurement and theory, however it also means that the measurement uncertainties on $R^{(1)}_{Jets}$ will need to be extremely small

before the result can make any useful contribution to improving the current understanding of QCD. Conversely, when used to search for beyond Standard Model physics, this is extremely advantageous as a given relative deviation of the ratio from its theoretical prediction would be more significant than the same relative deviation on one of the cross sections. In this sense the $R^{(1)}_{Jets}$ measurement is more sensitive to new physics than either of the cross sections that comprise it. It may not be true that this cancellation of theoretical systematics continues to higher jet multiplicities. However it is clear that whether it does or not there is still benefit in performing the measurement. A cost in terms of constraining the modelling of QCD comes with a benefit in terms of new physics searches and vice-versa.

For the rest of this work, the superscript n will be dropped and the variable R_{Jets} will be implicitly understood to refer to the $n = 1$ case with the choice of x described above.

4.2 Event Selections

This section lists and describes the criteria used to select W and Z events produced in association with jets within ATLAS. These criteria are largely standard for all 2010 analyses conducted by ATLAS [8] [9], however there are some selections which are specifically tailored to this measurement and these will be described in detail. The reconstruction level selections are listed fully in table 4.1.

Only events which are present on the common W/Z Good Run List (GRL) were used. This ensures that all requisite components of the detector were functioning at the time of the interaction. Application of the GRL resulted in a total integrated luminosity of 33 pb^{-1} for use in this analysis. Further to this, the events were required to have a primary vertex consistent with the nominal interaction point ($|z|$ <150 mm) with at least three associated tracks. This requirement reduces contamination from non-collision backgrounds.

Events from run periods A to E3 inclusive were required to pass the L1_EM14 trigger, which requires a single electromagnetic cluster be identified with greater than or equal to 14 GeV at Level 1. From period E4 onwards, this trigger was prescaled and consequently the

trigger EF_e15_medium was used, which requires a 15 GeV electron passing medium quality requirements be identified by the event filter.

For the purposes of this analysis, electrons are defined experimentally as having a transverse energy greater than 20 GeV in order to avoid regions where the reconstruction and identification efficiencies are rapidly varying or poorly understood. The transverse energy is calculated using the cluster energy measured by the calorimeter and the direction measured using the inner detector track thus: $E_T^e = E_{cluster}/cosh(\eta_{track})$. The cluster pseudorapidity is required to satisfy $|\eta| < 2.47$ in order to maximally exploit the phase space afforded by the tracking volume whilst avoiding edge effects. Electrons which are reconstructed within the crack region between the barrel and endcaps suffer from a degraded reconstruction efficiency as this region is poorly instrumented and consequently the region $1.37 < |\eta_{cluster}| < 1.52$ is discarded. Electrons reconstructed in regions suffering from dead optical transmitters (dead OTx) are similarly discarded.

W events are required to have one electron according to the above definition which passes tight identification criteria as defined in section 3.3.1. Z events are required to have one tight electron and a second electron which is at least medium and is of opposite sign to the first. The quality requirements are chosen to optimise the cancellation of systematics in the ratio.

In addition to the electron requirements above, W events are required to satisfy the conditions that the missing transverse momentum be greater than 25 GeV and the transverse mass, as defined by $M_T = \sqrt{2P_T^\nu P_T^e(1 - cos(\phi_e - \phi_\nu))}$, be greater than 40 GeV. These selections reduce the contamination from QCD background, which is predominantly at lower values of these variables. Of the three missing transverse momentum definitions described in section 3.3.3, `LocHadTopo` was chosen for use in this analysis.

Z events are additionally required to have an invariant mass in the region $71 < M_{\ell\ell} < 111$ GeV. This requirement is not only beneficial in terms of background rejection, but is required theoretically in order to provide a definition of what constitutes a Z event, as at lower invariant mass values virtual photon exchange produces identical final states and these processes interfere.

The jet definition chosen for use within this analysis was anti-k_T with a clustering radius of $\Delta R < 0.4$. The constituents over which the jet clustering algorithm was run were the topological clusters described in section 3.3.2. These jets are referred to as *anti-k_T4 TopoJets*. The jets are required to have a transverse momentum greater than 30 GeV after calibration using the jet energy scale, as described in the section referenced above. The jets were also required to have $|\eta|$ less than 2.8 due to the fact that in 2010 the jet calibration was poorly understood in the forward calorimeter and also in the transition region between the forward calorimeter and the endcap calorimeters. Each of the jets found are further required to pass a set of quality requirements to ensure minimal degradation of the missing transverse momentum in the event. These quality criteria remove jets which are probably due to noisy calorimeter cells or are located in problematic detector regions.

This prescription for identification of jets will frequently identify the clusters due to electrons in the event as jets and consequently a prescription for electron-jet overlap removal is required. Ideally one would wish to remove any clusters associated with leptons in the event prior to running the clustering algorithm, but at the time of writing this had not yet been implemented so a more coarse approach was adopted whereby jets separated by less than 0.2 in $\eta - \phi$ space from a lepton are removed from consideration. Further to this, events which have an electron and jet separated by less than 0.6 after overlap removal are rejected, in order to exclude events where the presence of a jet may degrade the efficiencies for reconstruction and identification of an electron and vice-versa.

In W and Z + jets analyses we are particularly interested in events where the jets are produced in association with the vector boson, i.e. they result from the same interaction as opposed to being produced in a secondary pile-up interaction. In order to reject jets due to pile-up interactions, it was required that those jets within the tracking volume have at least 75% of their track momenta associated with a primary vertex. The fraction of the track momentum of a jet which is associated with a primary vertex is known as the *jet vertex fraction* and is abbreviated to JVF, thus this requirement stipulates that $|JVF| > 0.75$.

General Selections
Events pass Common WZ Good Run List
Primary Vertex Selection : N_{vtx} >1 with N_{trk} >3
Jet Definition
AntiK_T4 TopoJets, P_T^{EMJES} >30 GeV, $\|\eta\|$ <2.8
Overlap Removal: Only count jets with $\Delta R_{\ell,j}$ >0.2
Jet-Lepton Isolation: Remove events with $\Delta R_{\ell,j}$ <0.6 after applying overlap removal
Jet Cleaning: Reject events with jets failing standard 2010 quality requirements
JVF Cut: Only count jets with $\|JVF\| > 0.75$
Exclusivity: Require exactly one jet passing the above selections
Electron Definition
$E_T = E_{cluster}/cosh(\eta_{trk}) > 20$ GeV
$\|\eta_{cluster}\| < 2.47$, Crack ($1.37 < \|\eta_{cluster}\| < 1.52$) excluded
Passes Dead OTx Selections
Trigger Requirements
Period A-E3 Inclusive: L1_EM14
Period E4-I2 Inclusive: EF_e15_medium
W Event Selections - $W \rightarrow e\nu + j$
One Tight electron
No Second Medium Electron
`LocHadTopo` Missing Transverse Momentum > 25 GeV
Transverse Mass $M_T > 40$ GeV
Z Event Selections - $Z \rightarrow e^+e^- + j$
One Tight electron and a second electron at least Medium
Opposite Sign : $Q_{e1}Q_{e2} = -1$
Dilepton Invariant Mass : 71 GeV $< M_{\ell\ell} <$ 111 GeV

Table 4.1: A summary of the reconstruction level selections used in this study.

4.3 Detector Unfolding

The number of observed events passing selections defined to identify a particular physical process S cannot itself be directly compared with theory. The theoretical prediction calculated in the following chapter will first be derived at *parton level*, which means with respect to the final states of the relevant matrix elements, without inclusion of any initial or final state radiation, hadronization or the inclusion of underlying event modelling. This will then be corrected using a parton shower Monte Carlo to be defined with respect to the stable final state of the interaction, after fragmentation and hadronization but prior to any interaction with the detector. The prediction is then referred to as being at either *truth* or *particle level*. The quantities used to apply the event selections however, are those obtained after the final state particles in a collision event have interacted with the ATLAS detector. As well as containing background contamination, the resulting quantities will be subject to instrumental effects which must be removed in order that a meaningful comparison with the theory can be made. These quantities are termed *reconstruction level* quantities, and the process of correcting these quantities for instrumental effects is referred to as *unfolding*[1].

For the process S, which produces a particle level final state f, unfolding a measurement amounts to calculating the cross section $\sigma(S \to f)$, from the number of events observed passing the chosen event selections at reconstruction level. These are related via the following equation:

$$\sigma(S \to f) = \frac{N^S_{Obs} - N^S_{Bkg}}{C\epsilon_{ID}\epsilon_{Trig}\mathscr{L}} \tag{4.2}$$

In the above equation the numerator is simply the number of observed signal events, N^S_{Sig}, expressed as the difference between the number of observed events, N^S_{Obs}, and the expected number of background events, N^S_{Bkg}. The denominator consists of the following four terms:

[1] This term can, in certain contexts, have a special meaning related to the inversion of a matrix relating jet variables at truth and reconstruction level. Here it is to be understood as referring to the entire process of correcting a measurement from reconstruction level to truth level.

- C is the acceptance correction of the detector. If the reconstruction level phase space is defined by a set of geometrical and kinematic selections X_R and the corresponding truth level phase space f, in which the cross section is to be found, is defined by selections X_T, then this correction is given by the ratio of the probabilities that a given signal event passes these sets of cuts:

$$C(X) = A(X_R)/A(X_T) \tag{4.3}$$

  where $A(X)$ indicates the probability that a signal event passes the set of selections X and is referred to as an acceptance.

- ϵ_{ID} is the identification efficiency for the objects in the event. This is the probability that the relevant final state objects in an event of type S, which has passed the reconstruction level selections X_R, pass the object identification criteria. In the measurement outlined above this term would correspond to the probability that the electrons pass the quality requirements specified for each type of event given the event passes the other reconstruction level selections.

- ϵ_{Trig} is the trigger efficiency. This is the probability that a given event of type S which passes the reconstruction level selections X_R and the object identification criteria also passes the trigger requirement.

- $\mathscr{L}$ is the integrated luminosity over the period of data taking.

As the detector acceptances A are probabilities, they can be decomposed through the use of Bayes' theorem into products of separate conditional probabilities. For instance, if an event were required to pass selections 'X', 'Y' and 'Z', the acceptance for the full set of selections could be decomposed in any of the following equivalent ways [2]:

[2]Commas have been used here to denote the logical AND operation while the symbol | indicates a condition. Consequently $A(X, Y, Z)$ should be read as the probability that an event passes selections X AND Y AND Z while the statement $A(X, Y|Z)$ should be interpreted as the probability of X AND Y given Z.

$$\begin{aligned} A(X,Y,Z) &= A(Z|Y,X)A(Y|X)A(X) \\ &= A(X|Z,Y)A(Z|Y)A(Y) \\ &= A(Y|X,Z)A(X|Z)A(Z) \\ &= ... \end{aligned} \tag{4.4}$$

The separate terms in the decompositions are statistically independent by construction, however it is in principle possible that a given source of systematic uncertainty may have a correlated effect on the different terms. Such a decomposition is advantageous when terms can be factored which are subject to different sources of systematic uncertainty, as these uncertainties can then be independently studied and the uncertainties on the separate terms can still be added in quadrature. The R_{Jets} analysis is particularly concerned with the cancellation of systematic uncertainties pertaining to the jet selections and consequently the acceptance for each type of event at each level has been factored into a term giving the probability that the event passes the jet selections, and a term giving the probability that all other selections are passed given the event passes the jet requirements.

The unfolding scheme presented above corrects from the number of events observed in the fiducial volume of the detector, defined by the selections X_R, to the cross section over the phase space f defined in terms of the truth level selections X_T. This is referred to as a *fiducial measurement*. One could in principle make the set X_T empty, effectively correcting back to the full phase space of the process S. The acceptance $A(X_T)$ is then equal to unity and the acceptance correction reduces to the reconstruction level acceptance. Unfolding relies on the use of Monte Carlo simulation to calculate acceptances, correcting to the full phase space is therefore dependent on the modelling of the process S over the entire of the phase space, not just the region of phase space in the vicinity of that which is observed.

In the context of R_{Jets}, extrapolating to the full phase space of the vector bosons has the advantage that once this has been done, the available phase space for the jets in the two types

of event is close to identical and not biased by the imposed selections, potentially leading to an improved cancellation of the jet systematics. On the other hand, using Monte Carlo simulations to extrapolate outside the regions where previous experiments have validated them is likely to lead to increased generator systematics. The fiducial measurement can attempt to avoid this concern by choosing the set X_T to be as similar as possible to the set X_R[3].

In the case of the lepton selections, the fiducial measurement is reasonably successful in this regard as the effect of the detector is simply to blur the edge of the distribution around the cut value, so the correction to be obtained only requires accurate modelling of the lepton kinematic distributions in the vicinity of the cut values. In the case of the E_T^{miss} selection however, the distribution is quite heavily distorted by the detector and calculation of the acceptance correction still effectively relies on a good understanding of the behaviour of the entire distribution at both truth and reconstruction level. Comparisons of the electron P_T and missing transverse momentum distributions at truth and reconstruction level are shown in figure 4.2.

Ideally the best course of action would be to perform the measurement in such a way as to give minimal systematic uncertainties, however pragmatically there is the question of how well one can assign uncertainties to these systematics. Understanding generator systematics is more complex than understanding the uncertainties due to the jets, which will be shown in section 4.5 to have minimal effect on the ratio, and consequently it was decided initially to perform a fiducial measurement.

The truth level phase space f to which the measurement is unfolded is defined via the selections (X_T) listed in table 4.2. The selections defining this phase space closely parallel the reconstruction level selections except without the imposed quality requirements. The definition of the truth level phase space refers to the use of 'dressed' electrons. These are defined as the 4-vector sum of the 'bare' electron (the final state electron after all final state

[3]While the standard geometrical and kinematic cuts applied at reconstruction level can usually be straightforwardly carried over and applied to the corresponding truth level objects, there is nothing which corresponds to things like the quality selections applied at reconstruction level.

Jet Definition
AntiK_T4 Truth Jets, P_T >30 GeV, $\|\eta\|$ <2.8
Overlap Removal: Only count jets with $\Delta R_{\ell,j}$ >0.2
Jet-Lepton Isolation: Remove events with $\Delta R_{\ell,j}$ <0.6 after applying overlap removal
Exclusivity: Require exactly one jet passing the above selections
Electron Definition
All criteria are applied to Dressed Electrons $P_T > 20$ GeV
$\|\eta\| < 2.47$, Crack $(1.37 < \|\eta\| < 1.52)$ excluded
W Event Selections - $W \to e\nu + j$
One Electron in Phase space defined above. Neutrino $P_T > 25$ GeV
Transverse Mass $M_T > 40$ GeV
Z Event Selections - $Z \to e^+e^- + j$
Two electrons in phase space defined above.
Dilepton Invariant Mass 71 GeV $< M_{\ell\ell} <$ 111 GeV

Table 4.2: Selections defining the truth level phase space f to which the measurement is unfolded. The meaning of a dressed electron is defined in the main text.

radiation) with all radiated photons within a cone of 0.1 in $\eta - \phi$ space. This definition of an electron is more closely aligned with the reconstruction level definition where the detector automatically clusters nearby energy deposits.

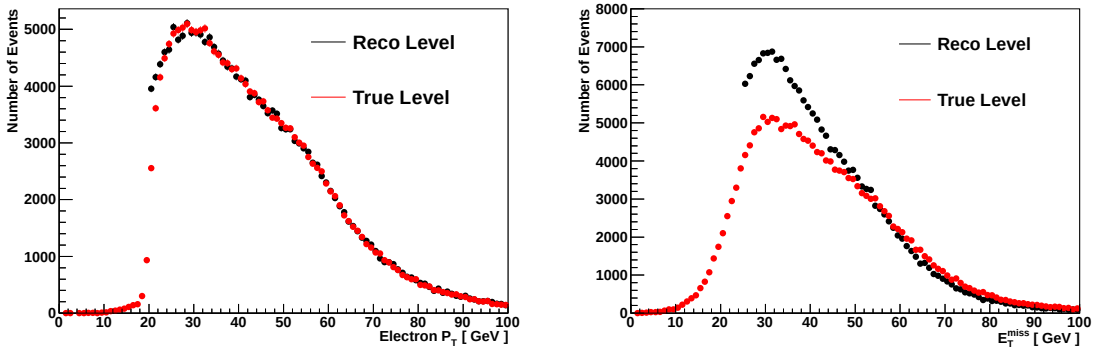


Figure 4.2: Plots comparing truth and reconstruction level distributions after the application of the reconstruction level event selections. The left hand plot shows the electron P_T distributions, which are close in shape apart from a small degree of blurring around the 20 GeV cut value. The right hand plot shows the E_T^{miss} distributions. These distributions are markedly different in shape, with the truth distribution extending below the cut value down to zero.

Using the tools developed above it is now possible to completely describe the unfolding scheme employed for correcting R_{Jets}. In the following, the symbols defined in table 4.3 are used to denote sets of selections at truth and reconstruction level. The explicit acceptance corrections for W and Z events are given in equation 4.5.

$$
\begin{aligned}
C^W &= C^W(J|\ell_T^W, E_{T,T}^{miss}, M_{T,T}) \cdot C^W(\ell^W, E_T^{miss}, M_T|J_R) \\
&= \frac{A^W(J_R|\ell_T^W, E_{T,T}^{miss}, M_{T,T})}{A^W(J_T|\ell_T^W, E_{T,T}^{miss}, M_{T,T})} \cdot \frac{A^W(\ell_R^W, E_{T,R}^{miss}, M_{T,R}|J_R)}{A^W(\ell_T^W, E_{T,T}^{miss}, M_{T,T}|J_R)} \\
C^Z &= C^Z(J|\ell_T^Z, M_{\ell\ell,T}) \cdot C^Z(\ell^Z, M_{\ell\ell}|J_R) \\
&= \frac{A^Z(J_R|\ell_T^Z, M_{\ell\ell,T})}{A^Z(J_T|\ell_T^Z, M_{\ell\ell,T})} \cdot \frac{A^Z(\ell_R^Z, M_{\ell\ell,R}|J_R)}{A^Z(\ell_T^Z, M_{\ell\ell,T}|J_R)}
\end{aligned}
\tag{4.5}
$$

This particular choice of structure for the acceptance corrections was motivated by the desire to only have the first term comparing jets at truth and reconstruction level and this is the only form which achieves this. Note that terms which refer to jets whilst not explicitly requiring a reconstruction level electron use the truth daughter electrons of the vector bosons to implement the overlap removal and isolation criteria. Where this is of particular importance a further subscript will be given indicating whether the truth or reconstruction level electrons are used. A short derivation of this form from first principles is given in appendix A. The complete expressions for the W and Z cross sections are:

$$
\begin{aligned}
\sigma(W(\to e\nu) + 1\text{Jet}) &= \frac{N_{Obs}^W - N_{Bkg}^W}{C^W(J|\ell_T^W, E_{T,T}^{miss}, M_{T,T}) C^W(\ell^W, E_T^{miss}, M_T|J_R) \epsilon_{ID}^W \epsilon_{Trig}^W \mathscr{L}} \\
\sigma(Z(\to ee) + 1\text{Jet}) &= \frac{N_{Obs}^Z - N_{Bkg}^Z}{C^Z(J|\ell_T^Z, M_{\ell\ell,T}) C^Z(\ell^Z, M_{\ell\ell}|J_R) \epsilon_{ID}^Z \epsilon_{Trig}^Z \mathscr{L}}
\end{aligned}
\tag{4.6}
$$

Taking the ratio of these two expressions one obtains the required relation between the value of R_{Jets} in the fiducial phase space at truth level and the raw numbers of W and Z events observed, N_{Obs}^W and N_{Obs}^Z respectively:

$$
R_{Jets}(x_T) = \frac{C^Z(J|\ell_T^Z, M_{\ell\ell,T})}{C^W(J|\ell_T^W, E_{T,T}^{miss}, M_{T,T})} \cdot \frac{C^Z(\ell^Z, M_{\ell\ell}|J_R)}{C^W(\ell^W, E_T^{miss}, M_T|J_R)} \cdot \frac{\epsilon_{ID}^Z}{\epsilon_{ID}^W} \cdot \frac{\epsilon_{Trig}^Z}{\epsilon_{Trig}^W} \cdot \frac{N_{Obs}^W - N_{Bkg}^W}{N_{Obs}^Z - N_{Bkg}^Z}(x_R)
\tag{4.7}
$$

Symbol	Selections
ℓ_T, ℓ_R	The set of kinematic and geometrical selections applied to the leptons in the event at truth and reconstruction level respectively. A superscript 'W' or 'Z' is used to remind the reader that two such leptons are required in Z events while only one is required in W events.
J_T, J_R	The set of kinematic, geometrical and quality selections applied to the jets in the event at truth and reconstruction level respectively.
$E^{miss}_{T,T}$, $E^{miss}_{T,R}$	The missing transverse momentum requirements applied to W events at truth and reconstruction level respectively.
$M_{T,T}$, $M_{T,R}$	The transverse mass requirements applied to W events at truth and reconstruction level respectively.
$M_{\ell\ell,T}$, $M_{\ell\ell,R}$	The invariant mass requirements applied to Z events at truth and reconstruction level respectively.

Table 4.3: A list of the shorthands used in the main text to denote particular sets of selections at truth and reconstruction level.

In the above equation, the parameters x_T and x_R are used to highlight the fact that the binning on the right hand side is in terms of the P_T threshold of the reconstructed jets, while on the left hand side the particle level jet threshold is used. Requiring that both W and Z events be subject to the same Good Run List implicitly constrains the integrated luminosity to be the same for both types of events and consequently this term cancels completely on taking the ratio. This is of great benefit as the systematic uncertainty on the luminosity was one of the largest uncertainties on the W and Z cross section measurements in 2010 [71].

The uncertainties on each of the factors in the ratio are assumed uncorrelated and consequently the overall uncertainty on R_{Jets} can be determined by independently evaluating the uncertainties on each of the individual factors in the above equation and adding them in quadrature:

$$\Delta R_{Jets} = \Delta\frac{C^Z(J|\ell^Z_T, M_{\ell\ell,T})}{C^W(J|\ell^W_T, E^{miss}_{T,T}, M_{T,T})} \oplus \Delta\frac{C^Z(\ell^Z, M_{\ell\ell}|J_R)}{C^W(\ell^W, E^{miss}_T, M_T|J_R)} \oplus \Delta\frac{\epsilon^Z_{ID}}{\epsilon^W_{ID}} \oplus \Delta\frac{\epsilon^Z_{Trig}}{\epsilon^W_{Trig}} \oplus \Delta\frac{N^W_{Obs} - N^W_{Bkg}}{N^Z_{Obs} - N^Z_{Bkg}} \tag{4.8}$$

The majority of the rest of this work will be concerned with evaluating the above correction terms and their uncertainties.

4.4 Monte Carlo Samples

Physics process ($\ell = e, \mu, \tau$)	Generator	$\sigma \cdot$ BR (nb)	
$W \to \ell\nu$ inclusive	PYTHIA 6.4.21	10.46	NNLO
$W^+ \to \ell^+\nu$		6.16	NNLO
$W^- \to \ell^-\overline{\nu}$		4.30	NNLO
$Z \to \ell\ell$	PYTHIA 6.4.21	0.99	NNLO
$W \to \ell\nu$ + jets $(0 \leq N_p \leq 5)$	ALPGEN 2.13	10.46	NNLO
$Z \to \ell\ell$ + jets $(m_{\ell\ell} > 40$ GeV, $0 \leq N_p \leq 5)$	ALPGEN 2.13	0.99	NNLO
$t\bar{t}$	POWHEG-HVQ		NLO+
	v1.01 patch 4	0.16	NNLL

Table 4.4: A list of the Monte Carlo samples used in the R_{Jets} analysis. The cross sections of the various samples are listed as well as the order of calculation to which these cross sections were calculated.

Monte Carlo simulation will be used extensively in this measurement, both directly in the calculation of the correction factors defined in the previous section, and also in the development and testing of methods and tools to be used for performing data-driven extraction of these values. Table 4.4 lists the samples which have been used in performing this measurement. The samples have all been produced with either Pythia [72], ALPGEN+HERWIG+JIMMY [73] [74] [75][4] or POWHEG [76] and have then been passed through the ATLAS GEANT4 detector simulation [77].

ALPGEN and Pythia are both leading order parton shower Monte Carlo generators. The Pythia samples only include the zero parton matrix element which is then subject to FSR and weighted such that the leading jet cross section agrees well with data. ALPGEN predictions however receive contributions from matrix elements with from 0 to 5 radiated hard partons. These hard partons are then subject to fragmentation in the usual parton shower formalism and a prescription is employed known as 'MLM' matching [78], which ensures that there is no double counting of phase space between the matrix elements and the parton shower. While both Pythia and ALPGEN should produce predictions in agreement with data for the inclusive and 1 jet cases, one would expect that for higher jet multiplicities the performance

[4]Results obtained using the ALPGEN generator with hadronisation provided by HERWIG and underlying event modelling performed by JIMMY will simply be referred to in the rest of the work as having been obtained using 'ALPGEN' for brevity.

Number of vertices	Weights ($W \to e\nu$)
1	1.628
2	1.114
3	0.858
4	0.721
5	0.656
6	0.638
7	0.630
$\geq$ 8	0.630

Table 4.5: Vertex weights applied to the pile-up MC samples to match the pile-up of the data sample.

of ALPGEN would be superior.

Some of the samples listed have been overlaid with minimum bias (minbias) events in order to simulate the effect of pile-up interactions. For each event in a sample for which this is the case, the number of minbias events to be overlaid is drawn from a Poisson distribution with a mean of two. Given that the amount of pile-up in data increases with instantaneous luminosity, a prescription for correcting the pile-up in simulation to match the amount observed in data has been developed. This procedure involves reweighting the simulated events so that the primary vertex distribution in Monte Carlo matches the distribution observed in data. The weights required for this are calculated centrally and are listed in table 4.5.

4.5 Jet Correction

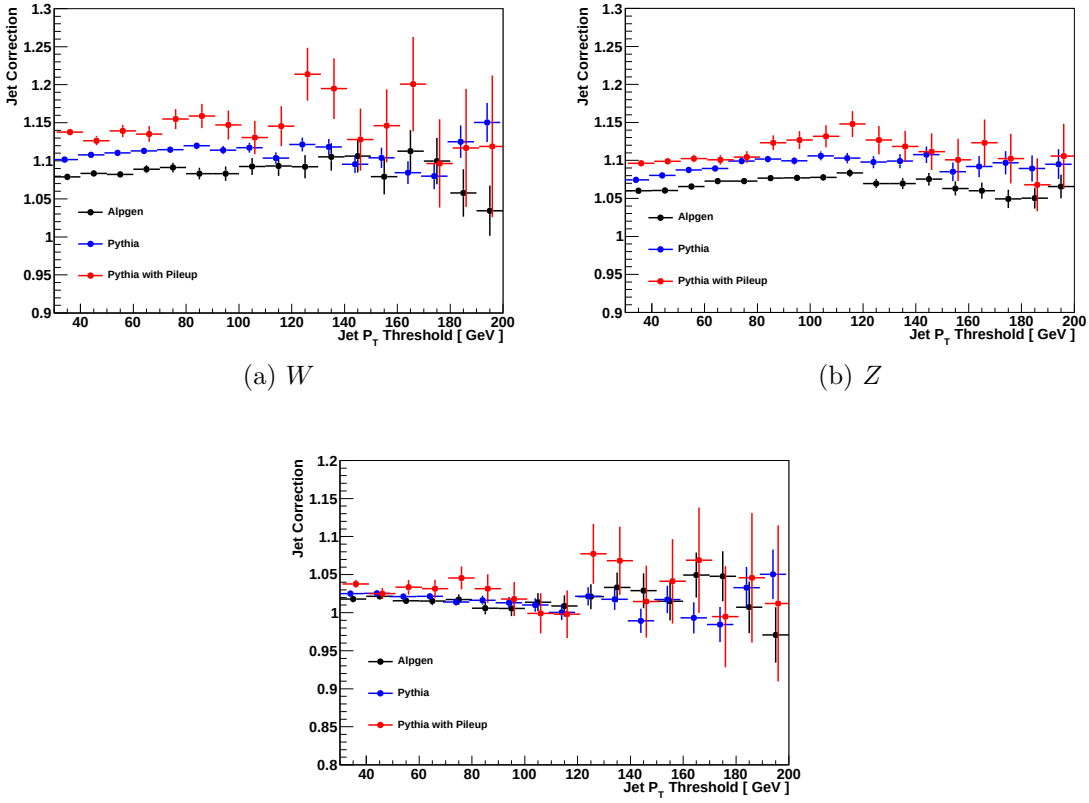
Most of the correction factors specified in equation 4.7 will be the subject of dedicated chapters, however the jet correction factor $C^Z(J|\ell_T, M_{\ell\ell,T})/C^W(J|\ell_T, E^{miss}_{T,T}, M_{T,T})$ is discussed here as its behaviour is central to the motivation for this measurement.

C^W and C^Z in the above expression give the relative numbers of events passing the jet selections at reconstruction level and truth level when subject to the truth level fiducial vector boson selections. Consequently, if the detector affects the W and Z jet spectra identically, these factors will be the same and the correction on the ratio will simply be unity. In

this instance one would expect a perfect cancellation of the jet systematics. In reality the differing masses of the vector bosons and, to a smaller extent, the differing constraints on the jet phase space imposed by the fiducial selections for the two types of event lead to the jet spectra having slightly different shapes at low jet P_T. As a result of this the ratio is not entirely flat and the detector has a slightly different effect on the spectra in the two types of event. Figure 4.3 shows the jet corrections for the W and Z events separately and for the ratio when calculated using ALPGEN and Pythia with and without pile-up simulation. The jet spectra for the W and Z events are affected significantly by the detector, showing increases of the order of 10% across all jet P_T thresholds. The exact value of the increase is also statistically incompatible between the two generators and affected markedly by pile-up. The effect on the ratio is only a 2% increase at low jet momentum and decreases to become consistent with zero at higher jet transverse momenta where the effect of the mass and selection differences become negligible. The agreement between the generators is also noticeably improved for the ratio with respect to the separate types of event.

In order to assign systematic uncertainties to the jet corrections, the jet energy scale and jet resolution uncertainties were used (see section 3.3.2). Firstly, all jets in each set of events were shifted up and down in tandem by a relative amount given by the appropriate jet energy scale uncertainty. The jet selections were then reapplied to the new shifted jet collections and the corrections were re-evaluated. The relative shifts on the jet corrections for the three cases are shown in figure 4.4. C^W and C^Z both show shifts in excess of 10% at high jet transverse momentum, while the ratio shows only a 2% deviation at low jet P_T which decreases and becomes statistically compatible with zero above 70 GeV. The ratio is clearly significantly more resilient to correlated changes in the energy scale of the jets in the events than either of the separate cross sections.

Following this, the jet energy resolution uncertainty was used in a similar way to assess the effects of incorrect simulation of the degree of smearing imposed by the detector on the jets in the events. The jets in each event were subjected to an additional smearing according to a Gaussian distribution with a mean of 1 and a width given by their relative jet energy

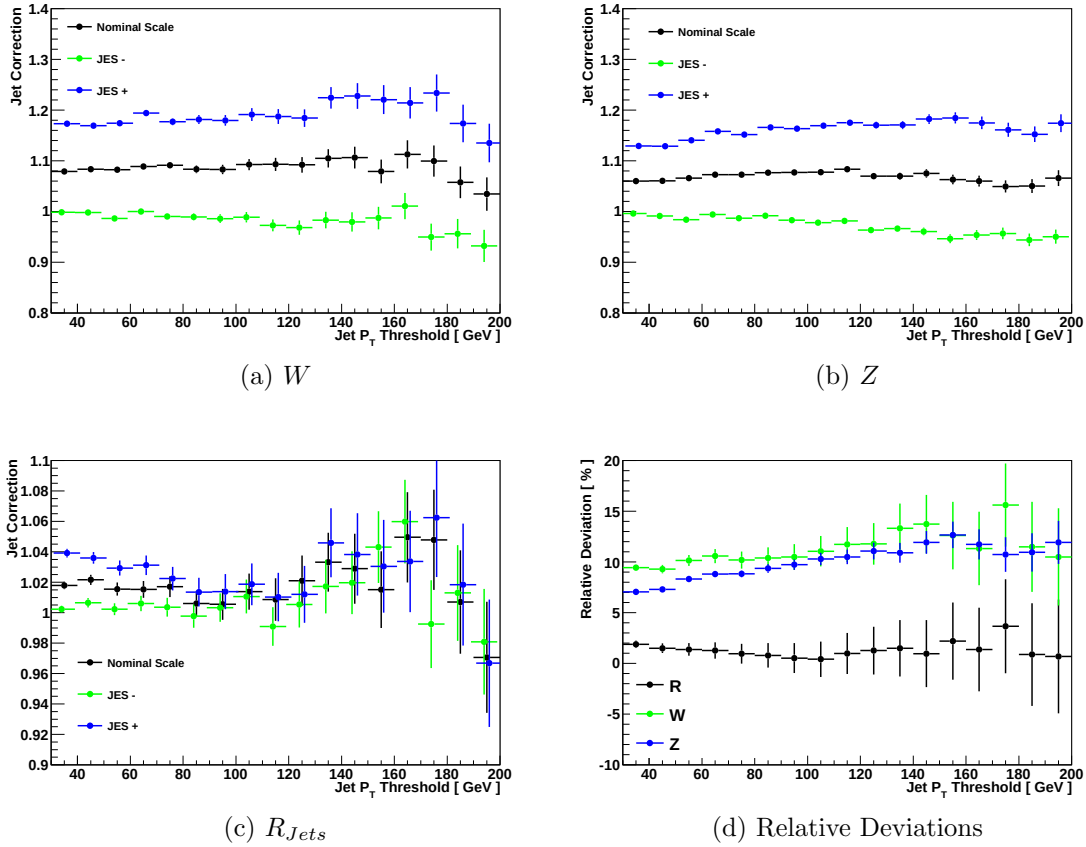


(a) W (b) Z

(c) R_{Jets}

Figure 4.3: The upper left and right plots show the jet corrections for W and Z events respectively when calculated using ALPGEN and Pythia with and without pile-up. The lower plot shows the jet correction for R_{Jets}, which is considerably closer to unity and shows improved agreement between generators.

resolution uncertainty. The jet criteria were then imposed on the smeared jets and the jet corrections were recalculated. In order to improve the statistical significance of this, 5000 pseudoexperiments were performed. The relative deviations between the nominal values of the jet corrections and the means of the distributions resulting from the pseudoexperiments are shown on the left hand side of figure 4.5. The widths of the distributions are quoted as uncertainties on these values. This process has minimal effect on any of the jet correction factors and the relative deviation on the correction to R_{Jets} is compatible with 0 over the entire range of jet transverse momenta considered. To illustrate the fact that the ratio is still more resilient to these kinds of uncorrelated effects, the process was repeated using the full jet energy resolution rather than just the uncertainty. The results of this are shown on the right hand side of the plot. In this instance C^W and C^Z are subject to upwards shifts



(a) W (b) Z

(c) R_{Jets} (d) Relative Deviations

Figure 4.4: The upper left and right plots show the effect on the jet correction of shifting the jets in the events up and down according to their jet energy scale uncertainty for W and Z events respectively. The lower left plot shows how this propagates to the ratio and the lower right hand plot shows the symmetrised relative deviations in the three cases.

of between 2% and 3%. The ratio is largely insensitive to these shifts except at low jet transverse momentum where a small residual discrepancy is observed which remains below 2%.

A total uncertainty on the jet correction to R_{Jets} was assigned as the quadrature sum of the statistical uncertainty, the mean deviation due to the jet energy scale shifts and the deviation between the nominal value and the mean of the distribution obtained from smearing the jets according to the jet energy resolution uncertainty.

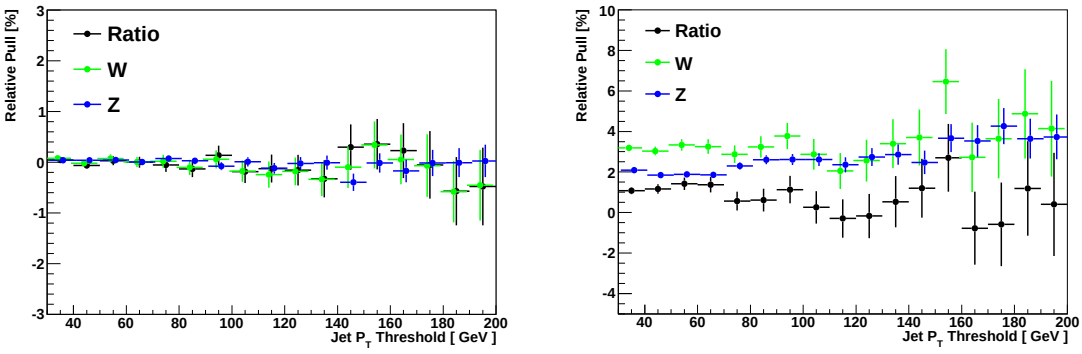


Figure 4.5: These plots show the relative deviation between the nominal jet corrections and those obtained as the mean values of distributions constructed from 5000 pseudoexperiments where the jets in the event were subjected to additional smearing. The left hand side uses the jet energy resolution uncertainty and will be used to assign an uncertainty on R_{Jets}, while the right hand side uses the full jet energy resolution and is intended as a demonstration of principle.

4.6 Conclusion

This chapter has outlined the motivation for the R_{Jets} measurement in terms of providing a robust test of the Standard Model as well as allowing model independent searches for new physics to be performed. The structure of the measurement and the procedure for implementing it have also been defined. The primary advantage of the R_{Jets} measurement, that the construction of the observational variable as a ratio allows a strong reduction of jet systematics, has been put on a solid foundation by the study of the jet correction when subject to two sources of systematic uncertainty: those arising from the jet energy scale and resolution respectively. While the component cross sections were observed to be sensitive to these variations, particularly in the case of the jet energy scale uncertainty, R_{Jets} was observed to be extremely robust against them. At high values of the jet transverse momentum threshold, where the vector boson mass difference is small compared to the scale of the interaction, the variations on R_{Jets} due to both of these uncertainties were found to be statistically consistent with zero.

Chapter 5

Theoretical Systematics

As outlined in the introduction and expanded on in the previous chapter, the primary aims of the R_{Jets} measurement are to provide a rigorous test of the Standard Model and to provide a robust means with which to search for new physics in a model independent way. Both of these objectives require a precise prediction with which to compare the results of the measurement detailed in the following chapters. This chapter describes the procedure used to obtain this prediction. Like the measurement, the theoretical prediction to which it is compared suffers from a set of uncertainties, arising both from the empirical input to the Standard Model and also the mathematical tools used to obtain numerical predictions from it. The three primary theoretical 'systematic' uncertainties and the methods used to evaluate them will therefore also be discussed. The R_{Jets} measurement will be found to benefit not only from the cancellation of experimental systematics, but also from a high degree of cancellation in the uncertainties present when calculating the theoretical prediction.

5.1 MCFM

In order to calculate a prediction for R_{Jets} and evaluate its theoretical systematic uncertainties, a program called MCFM [79] was used (Monte Carlo for FeMtobarn processes). This program allows a numerical calculation of partonic differential cross sections with up to NLO

accuracy for W and Z processes with 2 or fewer jets and LO accuracy for processes with three jets.

MCFM evaluates cross sections by approximating equation 2.23 as a sum of a set of randomly generated points over the required phase space. An event weight is calculated for each point based on the evaluated differential cross section at the sample point, the sampling probability distribution used in the generation of the sample points and the number of sample points generated. Differential distributions can then be constructed by appropriately binning these weights based on the kinematics of the sampled point in phase space. Further details of the binning procedure are given in appendix B.

At NLO this process is complicated by the need to evaluate the real emission and virtual loop contributions as well as the tree level process. The real emission contribution has a different final state phase space ($m+1$ partons) from the tree level diagram and loop contributions (m partons) and consequently requires an independent numerical integration. This presents problems however, as soft and collinear singularities present in the real emission diagrams are cancelled by singularities in the virtual loop diagrams. Separate numerical integration of these contributions therefore requires a prescription for regulating and cancelling these divergences and this is afforded by Catani-Seymour dipole subtraction [80]. This procedure firstly involves subtracting from the real emission contribution a set of additional terms, one for each pole, which have the same singularity structure as the corresponding pole, and hence render the real emission contribution finite over the entire phase space of interest. These additional terms, known as *dipoles*, are constructed from all possible partonic splittings of the m parton Born process and are formally integrable by construction. These terms can therefore be integrated over the phase space of the additional radiated parton and these integrated terms can be added to the virtual loop contribution, similarly rendering it finite over the full phase space. While this procedure changes the values of each of the two contributions, rendering each of them manifestly finite, it leaves the sum of the two integrated over the entire phase space invariant.

Implementation of this procedure into a Monte Carlo program results in some particular

features which are worth explicitly mentioning. Firstly, one is required when performing the integration over the $m+1$ parton phase space of the real emission matrix element to generate for each event a corresponding set of counter-events, one for each of the subtracted dipole terms. The kinematics of these counter-events can be different from the event to which they are associated but will tend to become the same in any limit where the real emission cross section becomes singular in the way corresponding to that counter-term. This can result in large fluctuations in adjacent bins of a histogram constructed from these events. This occurs whenever an event near a singular region of the phase space falls close to a bin edge and the counter-event with a correspondingly large opposite-sign weight falls in the adjacent bin. In order to reduce the effects of such fluctuations when constructing differential distributions in a particular kinematic variable, large statistics samples and a sufficiently coarse binning are required. The details of the samples generated for use in these studies are also described in appendix B.

The second pertinent consequence of this procedure is that it is not possible at NLO to use the weight distribution in a given bin to estimate the statistical uncertainty on that bin value. If all the event weights were independent, it would be possible due to the central limit theorem, to approximate the relative error in terms of the mean and width of the weight distribution and the number of events in a given bin as $\frac{\sigma}{\sqrt{N}\bar{w}}$, where N is the number of generated sample points, $\bar{w}$ is the mean of the weight distribution and σ is its width. At leading order and also for the virtual loop contribution this is possible, but for the real emission contribution the fact that the event has a set of statistically correlated counter-events means that this is not possible, unless the counter-events corresponding to a given event can be identified and grouped. Only the sum of these weights for each event is statistically independent. At present this has not been implemented but may be in the future [81]. Statistical errors are evaluated here by averaging over sets of distinct runs.

The theoretical predictions and uncertainties are derived for the truth level phase space defined in table 4.2 with the exception that the overlap removal and isolation criteria are not applied and the Born electron is used for the electron fiducial selections. These differences

are corrected for in the parton to particle level correction discussed in section 5.5.

5.2 Scale Uncertainties

When one renormalizes a gauge theory it becomes necessary to introduce an arbitrary scale, known as the renormalization scale μ_R, at which to evaluate the physical parameters of the theory. Cross sections calculated using the theory will then also obtain explicit logarithmic dependencies on this scale as a result of this process [53]. The renormalization group equations express the fact that any physically observable quantity should have no overall dependence on the renormalization scale. This fact allows calculation of the implicit dependence of the physical parameters of the theory on this scale, giving rise to partial differential equations such as the one shown in section 2.1.3. This is such that the explicit dependence of the observable on the renormalization scale is progressively cancelled out by the implicit dependence of the physical parameters at each successive new order. At fixed order, however, the perturbative expansion for both the hard subprocess cross section itself and the physical parameters are truncated, and this leads to a residual scale dependence of the next higher order in the coupling. Given that any residual scale dependence present at a particular fixed order should be eliminated by including successive higher orders into the calculation, it can be argued that the variation of the cross section with the renormalization scale is a measure of the effect of unknown higher order corrections. From this perspective, the scale uncertainty can be viewed as an uncertainty due to the calculation being performed at fixed order.

In an analogous way, when one factorizes the soft and collinear behaviour of the cross sections into the parton distribution functions, these also acquire a scale dependence described by the DGLAP evolution equations [82, 83, 84, 85]. This scale is known as the factorization scale μ_F.

Figure 5.1 shows the dependence of the W plus one jet cross section on the renormalization and factorization scales used when these values are held fixed throughout the calcu-

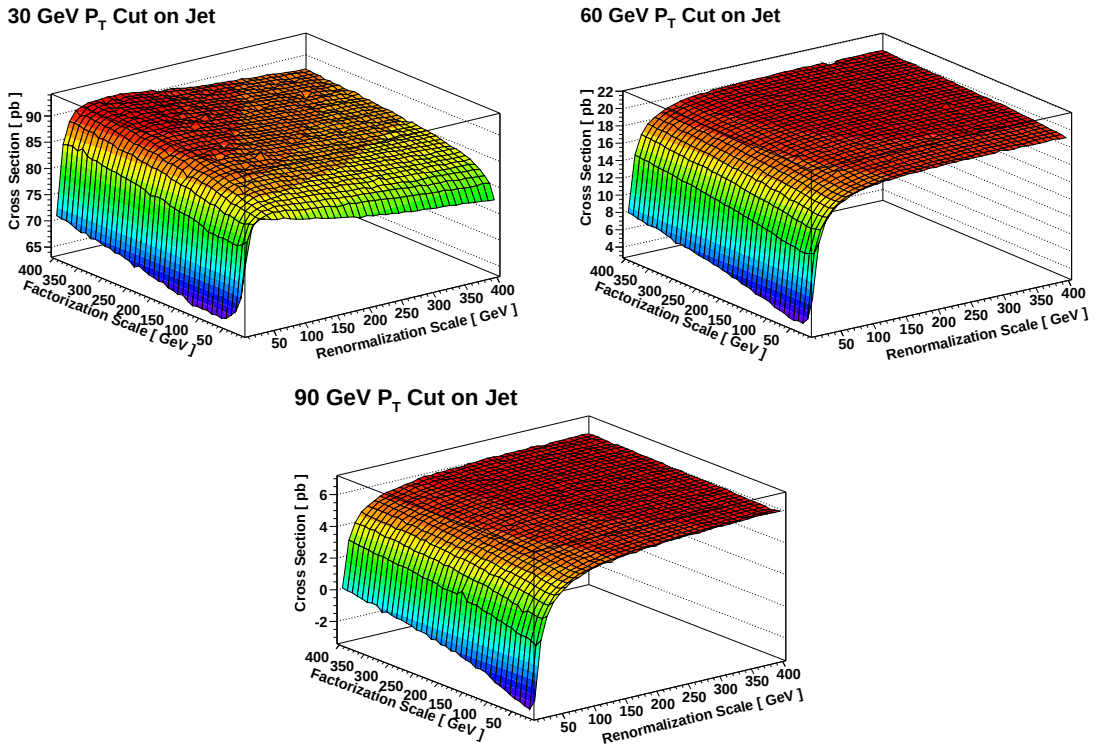


Figure 5.1: Dependences of the exclusive $W + 1$ jet cross section on the (fixed) renormalization and factorization scales used in the calculation for three different leading jet transverse momenta.

lation[1]. There are obvious regions of instability at low values of the scales where residual logarithmic terms in the expression for the cross section lead to divergent behaviour. The variation of the cross sections in terms of the scales at higher values of μ_R and μ_F is more gradual. In order to characterise the uncertainty associated with the unconstrained scales in the problem, it would be possible to choose a reasonable central scale governed by the kinematics of the problem and look at local variations around this value using plots such as the above. The disadvantage of this kind of approach is that these plots are extremely resource intensive to generate, requiring individual simulations with all different scale permutations within a reasonable range of the hard scale of the process of interest with sufficient statistics. In light of this an alternative approach has been adopted[2].

[1] The selections used in the generation of these plots are slightly different from those used in the main body of work but the general behaviour is the same.

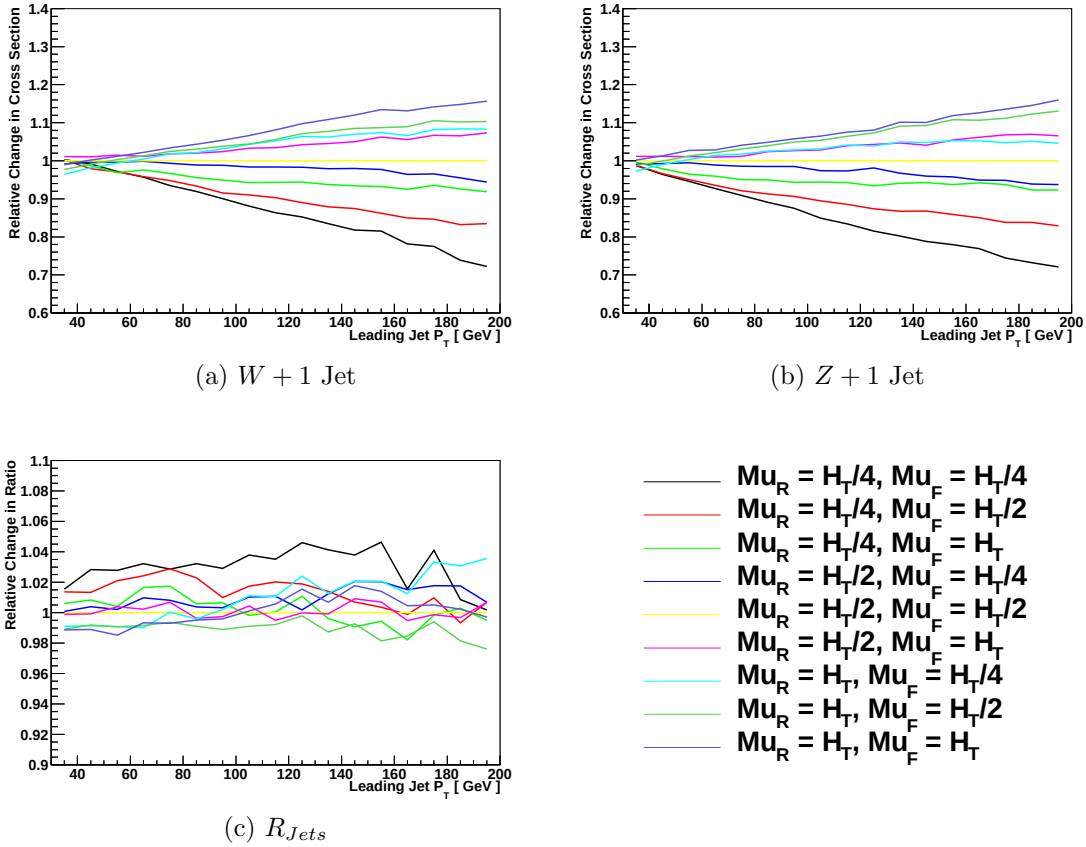
[2] More recently tools such as APPLGrid [86] have become available which allow rapid reweighting of events to different scale and PDF choices. These kind of procedures allow results such as those presented here to be obtained much more rapidly and with a much smaller overhead in terms of memory and cpu time. Progress in this area will almost certainly be made with the use of such tools.

MCFM allows dynamic allocation of the scales used for evaluating the cross section on an event by event basis depending on the kinematics of the particular event. Dynamic scale choices of this kind are deemed beneficial as they attempt to ensure that the logarithmic contributions to the NLO cross section mentioned above, which typically take ratios of either the renormalization or factorization scale with one of the real physical scales in the interaction as arguments, are always small in magnitude. This should in turn guarantee that the calculation is not susceptible to the kind of instabilities present in figure 5.1.

A scale choice of $H_T/2$ was chosen as a central value for both scales, where H_T is given by the scalar sum of the transverse momenta of all final state objects prior to the clustering of partons into jets. This choice was motivated by the fact that other collaborations have previously obtained good agreement with data [87] [88] when using scales based on H_T and also because in previous work, such scale choices were not observed to produce negative cross sections at large jet P_T.

In order to quote a systematic uncertainty, the envelope of the cross sections obtained from all possible combinations of scales $\{\mu_R, \mu_F\}\epsilon\{H_T/4, H_T/2, H_T\}$ was used with the exception of the completely anti-correlated choices, $(\mu_R, \mu_F) = (H_T/4, H_T)$ and $(\mu_R, \mu_F) = (H_T, H_T/4)$, as these were originally believed to introduce unnecessarily large logarithmic corrections and provide an overly conservative uncertainty [89]. Figure 5.2 shows the fractional deviations obtained for all of the above combinations from the nominal choice $\{H_T/2, H_T/2\}$ for the W and Z cross sections and for R_{Jets}.

At low P_T the cross sections increase with decreasing renormalization scale but this behaviour rapidly inverts and at higher jet P_T the cross sections increase roughly linearly with both scales. This leads to the largest deviations from the nominal choice being due to the completely correlated variations, somewhat contrary to the concerns about logarithmic corrections mentioned in the previous paragraph. The roughly linear increase in deviation from nominal with jet transverse momentum is not a result of a strong change in the behaviour of the cross section with μ_R or μ_F at higher P_T, but instead is simply a result of the fact that at higher P_T, the absolute deviation between a given fraction of H_T and the nominal is



(a) $W + 1$ Jet

(b) $Z + 1$ Jet

(c) R_{Jets}

Figure 5.2: Relative deviations of different scale choices from the nominal set $\{H_T/2, H_T/2\}$ for the W+1-jet cross section, Z+1-jet cross section and their ratio, R_{Jets}, respectively.

larger almost in proportion in the 1-jet case. The dependence of the ratio on scale choice is weak and displays little variation with jet transverse momentum. Figure 5.3 shows the total scale uncertainty, assigned by taking the envelope of the separate scale choices as described above, for both the W and Z cross sections and the ratio. The uncertainty on the ratio remains within 4% up to 100 GeV and within 6% up to 200 GeV, although from the lower plot of figure 5.2 it is clear that this variation contains a significant statistical component.

5.3 PDF and Strong Coupling Uncertainties

The nominal predictions which are to be compared with the R_{Jets} measurement were obtained using the CTEQ66 NLO PDF set but predictions have also been obtained using the MSTW2008 NLO PDF set [51] and the Neural Net PDF set, NNPDF21 [90]. Figure 5.4

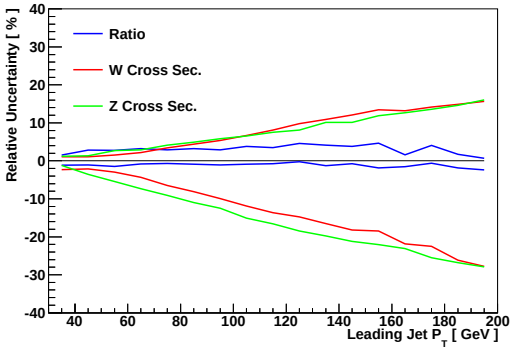


Figure 5.3: Total relative scale uncertainties on the W and Z + 1-jet cross sections and R_{Jets}, assigned using the prescription outlined in the main text.

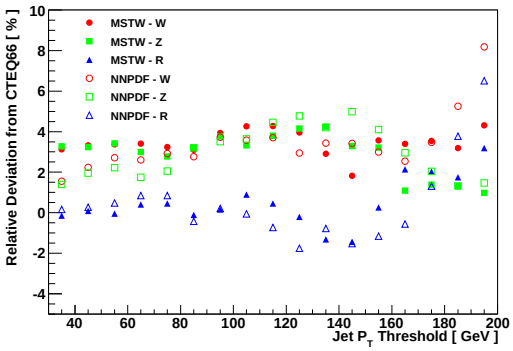


Figure 5.4: The relative deviations of the predictions for the W and Z cross sections and R_{Jets} when calculated using MSTW2008 and NNPDF21 from those obtained with CTEQ66.

shows comparisons of the predictions obtained for the W and Z cross sections and R_{Jets} when using these alternative PDF sets.

The MSTW2008 predictions show deviations on the cross sections which remain at the level of 3-4% for the range of transverse momenta considered. The NNPDF21 predictions show a deviation at low P_T ~2%, which then increases to be comparable with the MSTW prediction. The predictions for R_{Jets} however show no consistent disagreement between the three PDF sets below 160 GeV, at which point the statistics of the samples become insufficient.

PDF uncertainties have been evaluated using both the CTEQ66 and MSTW2008 NLO PDF sets. Both groups have produced a series of error PDF sets which can be used to evaluate uncertainties on the predictions of physical observables, CTEQ has 44, MSTW 40, the numbers depending on the explicit parameterisations used. These error sets are obtained

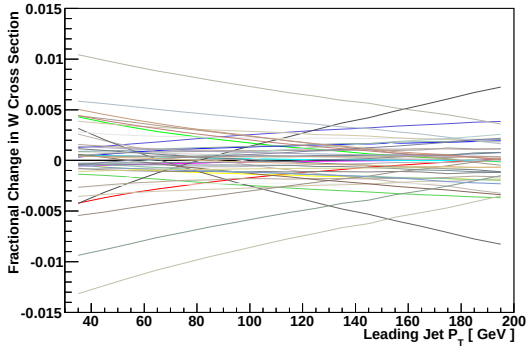
by numerically diagonalising the inverse of the Hessian matrix of second derivatives of the global χ^2 with respect to the PDF parameters [91]. This allows a determination of the allowed variation in the PDF eigenvector parameters according to some tolerance T in the deviation of the global χ^2 from its minimum value. The tolerances are obtained by finding the average deviation in the global χ^2 for which each separate dataset used in the fit is still consistent with the global fit at the P% confidence limit. For the CTEQ66 set used here P is 90%, whereas for the MSTW2008 set it is 68%. Assuming Gaussian behaviour, one can inter-convert these uncertainties by scaling with 1.645. Here the CTEQ66 errors have been scaled down to 68%. Figure 5.5 shows the relative deviations of the values obtained for the the W and $Z + 1$ jet cross sections and for R_{Jets} when using the 44 CTEQ66 error sets from those obtained using the global best fit set. Corresponding plots for the 40 MSTW error sets can be found in appendix B. The predictions correspond to reweightings of the same set of events and as a result the statistical errors are completely correlated. This leads to a much better precision on the relative deviation than was achievable for the scale variation studies where this was not the case.

The total uncertainties for each physical observable resulting from these variations in the PDF parameters were calculated using the master formula for asymmetric errors, equation 5.1 [92], which determines upper and lower uncertainties as the quadrature sum of the maximum deviation in each direction from each pair of eigenvector sets:

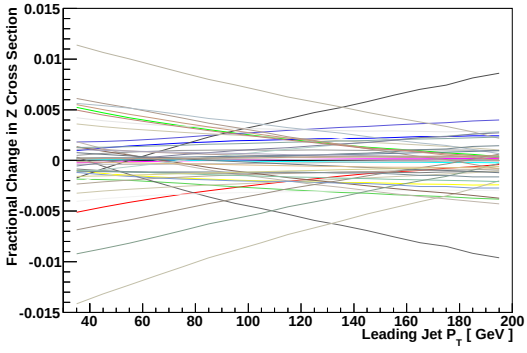
$$\Delta F(+/-) = \sqrt{\Sigma_{k=1}^{N}((max/min)[F(S_k^+) - F(S_0), F(S_k^-) - F(S_0), 0])^2} \qquad (5.1)$$

In the above equation the summation is taken over all of the eigenvectors of the inverse Hessian matrix. $F(S_0)$ indicates the value of the observable F when evaluated using the global best fit PDFs, while $F(S_k^+)$ and $F(S_k^-)$ indicate the values of F obtained when the PDF parameters are displaced in the positive and negative direction along the eigenvector k.

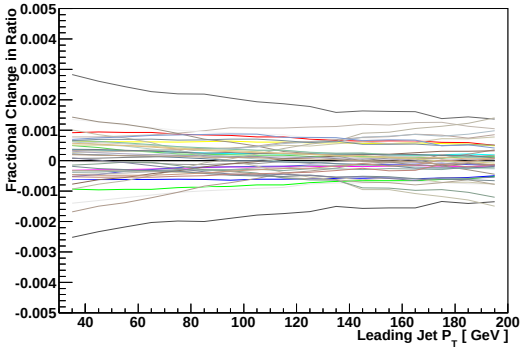
Figure 5.6 shows the total uncertainties for the cross sections and ratio assigned using this



(a) $W + 1$ Jet



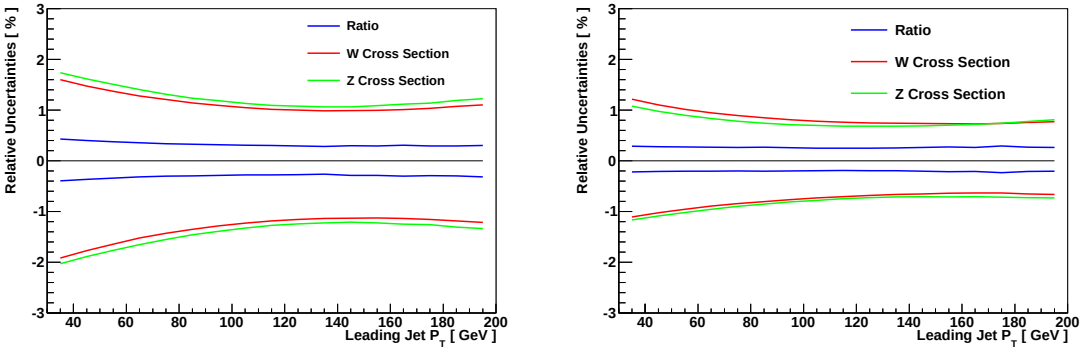
(b) $Z + 1$ Jet



(c) R_{Jets}

Figure 5.5: Relative deviations of the W and Z cross sections and ratios from those obtained using the global best fit when using different CTEQ error PDF sets. Note the difference in the scale of the ordinate.

equation for both PDF sets. For the CTEQ error sets the uncertainty on the cross sections is slightly larger in the negative direction and is within 2% up to 200 GeV in both directions. The uncertainty on the ratio is within $\sim 0.5\%$ in both directions. For MSTW the uncertainty on the cross sections is conversely slightly larger in the positive direction. Its value is just over 1% at 30 GeV which then decreases and stays within 1% up to 200 GeV. The uncertainty on the ratio is within 0.3% at all transverse momenta. The minimum in the uncertainty visible in both cases is due to a different eigenset being responsible for the dominant uncertainty at higher jet P_T than at lower jet P_T, this behaviour can be seen in Figure 5.5. Given the total PDF uncertainties as well as the relative deviations observed between the best fits shown in figure 5.4, it appears that the MSTW2008 and CTEQ66 predictions for the cross sections are not quite compatible within their respective uncertainties. The different PDF sets do however give compatible predictions for R_{Jets} despite their significantly smaller uncertainties.



(a) *CTEQ* (b) *MSTW*

Figure 5.6: Total PDF uncertainties for the cross sections and ratio evaluated using equation 5.1 for the CTEQ and MSTW error sets.

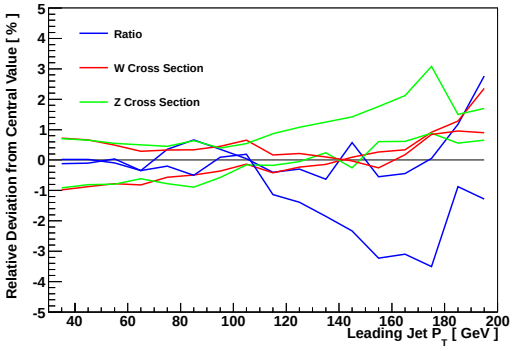


Figure 5.7: The variation in the cross sections and the ratio caused by varying the strong coupling by 0.002 around its nominal value.

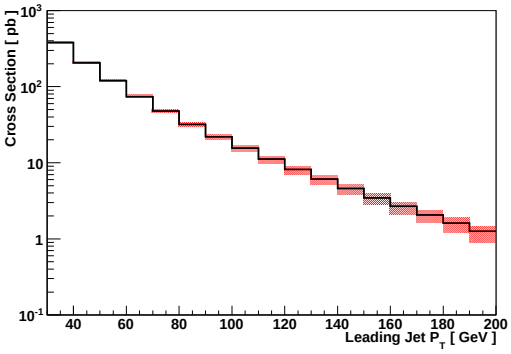
A series of additional PDF sets, CTEQ66alphas, have been created by the CTEQ group which allow for variations of the strong coupling α_S in a manner consistent with the global fitting procedure. These allow the calculation of the uncertainty on the cross sections due to uncertainties on the value of the coupling in a way which means the total uncertainty is still derivable as a quadrature sum with the other theoretical uncertainties described [93]. The samples used have variations of 0.002 around the nominal α_S value of 0.118 which corresponds to the 90% confidence limit. As above this has been scaled down to give a 68% uncertainty which is shown in figure 5.7 for the cross sections and the ratio. Below 100 GeV it is clear that there is no significant change in R_{Jets} on varying the coupling, above 100 GeV the results are dominated by statistical fluctuations due to smaller numbers of generated events in these samples.

5.4 Nominal Prediction and K-Factors

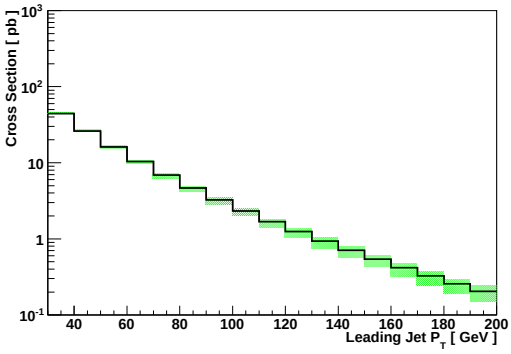
Figure 5.8 shows the W and Z cross sections and the value of R_{Jets}, calculated at next-to-leading order, with their total parton level uncertainties calculated as the quadrature sum of the scale, PDF and strong coupling uncertainties, for the relevant phase space. The combined uncertainty on the ratio in the total phase space is within 4% at 100 GeV and within 6% at 200 GeV. Figure 5.9 shows the k-factors for the cross sections and for R_{Jets}, which are defined as the ratio of the relevant quantity when evaluated at next-to-leading order and at leading order. The k-factors for the cross sections decrease markedly at higher jet P_T. This is due to the increased probability that a single high P_T parton will radiate an additional hard parton at a sufficiently large angle to form a dijet event. These events are then excluded due to the requirement that there be exactly one jet in the final state. The k-factor for the ratio is +4% at low jet P_T and gradually decreases to be statistically consistent with unity at 200 GeV.

5.5 Correction to Particle Level

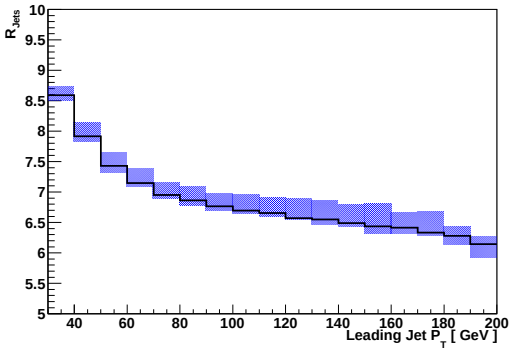
The predictions obtained from MCFM are at parton level, prior to any fragmentation, hadronisation or underlying event effects being included. In order to obtain a value for R_{Jets} which is comparable with the unfolded measurement, a correction needs to be made to the MCFM prediction in order to accommodate these effects. This correction is calculated by comparing the value of R_{Jets} obtained using a parton shower Monte Carlo when fragmentation, hadronisation and underlying event effects are included and when they are not. This correction also accommodates the difference in the definition of the truth electron described in section 4.3 and the definition used in the MCFM prediction. In the former case, when the fragmentation and hadronisation effects are included, the electron fiducial selections are applied to the dressed final state electron and overlap removal and isolation criteria are applied, while in the latter case, when these processes are not included, the Born electron is used and these criteria are not imposed. The nominal sample used to derive this correction is the



(a) $W + 1$ Jet



(b) $Z + 1$ Jet



(c) R_{Jets}

Figure 5.8: Central predictions obtained using the CTEQ6.6 global best fit PDF with strong coupling 0.118 and scale choice $\{\mu_R, \mu_F\} = \{H_T/2, H_T/2\}$.

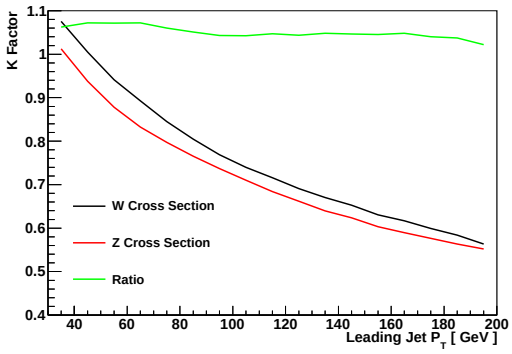


Figure 5.9: NLO / LO k-Factors for the W and Z cross sections and for R_{Jets} as a function of the jet transverse momentum threshold.

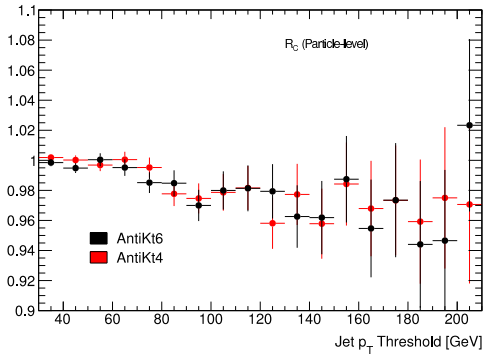


Figure 5.10: The correction factor applied to the MCFM prediction for R_{Jets} in order to accommodate the effects of initial and final state radiation as well as non-perturbative effects such as the underlying event and hadronisation. Results for two jet clustering radii are shown: 0.4 and 0.6.

Pythia 2-2 sample described in section 4.4 which utilises the ATLAS AMBT1 underlying event tune [94]. The nominal correction to R_{Jets} is shown in figure 5.10 for two different jet radii, here only ΔR =0.4 is relevant.

In order to evaluate a systematic uncertainty on this correction, samples generated using alternative underlying event tunes as well as varied levels of fragmentation were used. In the case of the underlying event, corrections were obtained using the following set of alternative tunes: AMBT1-EIGEN1M, AMBT1-EIGEN3x2, AMBT1-MARKUS1 as well as the default HERWIG+JIMMY underlying event tune. The resulting deviations are shown on the left hand side of figure 5.11. The envelope of the deviations from nominal of this set of alternative tunes was quoted as an uncertainty due to the underlying event modelling.

Similarly, uncertainties due to the modelling of fragmentation were assigned by using four tunes with increased or decreased levels of ISR and FSR. The deviations from the nominal tune observed when using these samples are shown on the right hand side of figure 5.11. In each of these two cases, the largest deviation was taken separately as an uncertainty. The total uncertainty on the correction factor was then derived as the quadrature sum of the uncertainties on the three separate contributions.

Once this correction factor has been applied, the predicted value of R_{Jets} can be directly compared with the result of the measurement. The final prediction can be seen compared

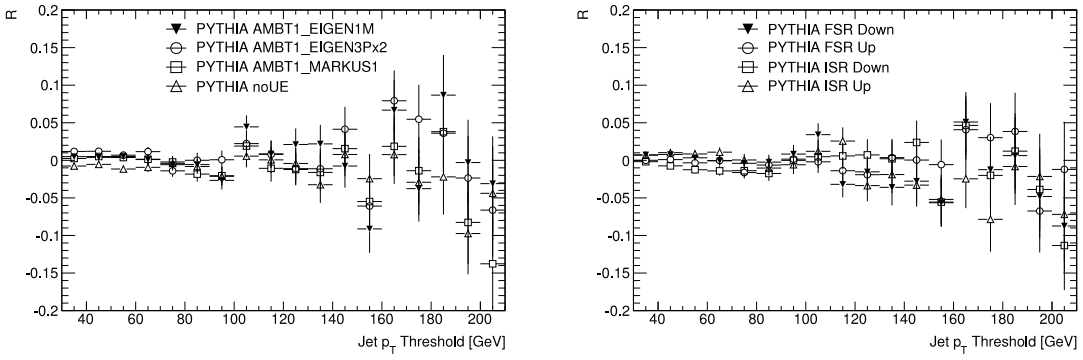


Figure 5.11: The left hand plot shows the variations of the R_{Jets} parton to particle level correction factor obtained when using the different underlying event tunes described in the text. The right hand plot shows the variations observed when the fragmentation modelling is varied.

with the result of the measurement in chapter 9.

Chapter 6

Background Subtraction

6.1 Introduction

The first step in calculating the ratio of the $W+1$ jet and $Z+1$ jet cross sections at particle level from the numbers of events of each type observed passing the selections defined in chapter 4 is to correct for the different background contaminations in the two cases. This correction proceeds in two steps, the first removes QCD background contamination from the number of observed events of each type, while the second step removes the number of electroweak background events from what remains. In order to perform the first step, QCD background fractions f_{qcd} are evaluated as functions of the leading jet transverse momentum threshold for both W and Z events using partially data-driven methods. The ratio of the number of observed events of each type is then corrected for QCD background contamination via multiplication by the factor:

$$C_{qcd}(P_T) = \frac{1 - f^W_{qcd}(P_T)}{1 - f^Z_{qcd}(P_T)} \tag{6.1}$$

Following this, electroweak background contamination is similarly corrected for via the calculation of electroweak background fractions f_{ewk} and multiplication by the ratio:

$$C_{ewk}(P_T) = \frac{1 - f^W_{ewk}(P_T)}{1 - f^Z_{ewk}(P_T)} \tag{6.2}$$

The electroweak background fractions are determined purely from Monte Carlo simulation. This is discussed in section 6.5. The majority of this chapter focuses on the evaluation of f_{qcd} for the two types of event.

The use of background fractions in the way described above has the benefit that there is no explicit dependence on the luminosity in the background subtraction procedure. Also any Monte Carlo uncertainties present in the electroweak backgrounds are expected to be dramatically reduced when dealing with fractions rather than absolute values.

6.2 Template Fitting

The approaches for calculating the QCD background fractions in W and Z events presented in this chapter rely on the construction of template distributions for the electroweak processes and the QCD background, which can then be fitted to the data in order to determine its fractional composition. The effectiveness of this procedure relies on the electroweak and QCD distributions in the template variable being sufficiently different after the application of all selection cuts that the fit can discriminate between their contributions to the data [95].

While it would in principle be possible to construct separate templates for the signal, all of the electroweak backgrounds and the QCD background and then perform a simultaneous fit to determine the composition of the data in terms of all of these contributions, in practice the shapes of the signal and electroweak backgrounds tend to be too similar for this to work. The fit has too many free parameters and does not converge, or if it does, the resulting background fractions are extremely sensitive to small changes in the input distributions. As a result of this, when attempting to determine the QCD background fractions, the signal and electroweak backgrounds are combined (weighted appropriately by cross section) to form a single electroweak template, which is then used to perform a two component fit to the data

with the QCD background template.

The template distributions could be taken from Monte Carlo, and in the case of the electroweak processes this is what has been done, but it is preferable where possible to extract the template shapes directly from data. The approaches presented here attempt to do this for the QCD background, which is not expected to be modelled sufficiently accurately in Monte Carlo. Extracting a QCD background template from data involves identifying a set of events which have minimal signal contamination, minimal bias with respect to the distribution that would be obtained if one could observe a pure set of QCD events passing all signal selections, and also have sufficient statistics to perform the template fit. There is some tension between the first two requirements and also potentially with the third. In order to obtain a pure background sample it is necessary to change (usually invert) some of the selections applied to data, however as more selections are changed, the likelihood that the template distribution taken from the sample is biased with respect to the distribution in the signal region increases. Thus the difficulty in implementing such a method lies in identifying a sensitive variable and identifying a kinematic region from which to extract a template which has sufficient statistics, sufficient purity of QCD and an adequately small bias.

The fits are performed using a ROOT class called TFractionFitter [96]. This calculates the relative contributions of a set of template distributions in a given set of data by assuming that all templates and the data itself are drawn from underlying distributions according to Poisson statistics. For any given set of assumed underlying distributions and composition fractions, the fitter calculates the log likelihood of observing the given templates and the data. It then attempts to numerically maximise this value with respect to the fractional contributions and the underlying distributions.

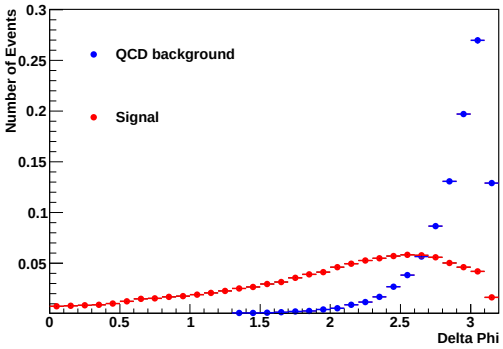
6.3 QCD Background Subtraction for W Events

6.3.1 Topological Method

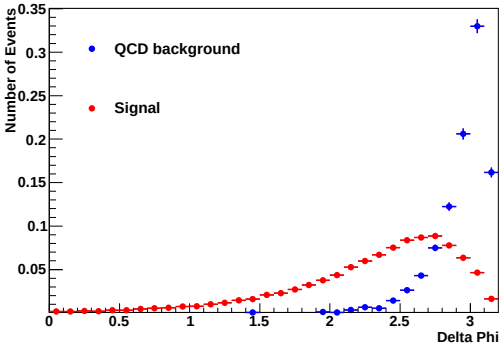
This method exploits the differing event topologies of signal and background events to distinguish between the two. The QCD background to a $W + 1$ jet process is predominantly due to QCD dijet events where a single jet fakes an electron, and dijet events containing a heavy flavour which decays to an electron. In the former case, one expects the electron to be almost back-to-back with the recoiling jet. In a genuine $W + 1$ jet event the jet is recoiling the momentum of the W boson. The W boson then decays, allowing the daughter electron to have on average smaller angular deviations from the recoiling jet. Consequently one expects the event shapes for the signal and the QCD background to be somewhat different.

As the momentum of the W boson increases, one would expect the electron in the event to be boosted along the direction of the W momentum, leading to larger angular separations between the jet and the electron and a degradation of the performance of a template fit based on such a method. This simple picture is further complicated however, by polarization effects present in the W decay. At high transverse momenta, helicity conservation paired with the left handed chiral couplings of leptons to the W bosons lead to the W^- decay preferentially boosting the electron along its direction of motion, whereas the W^+ decay preferentially decays the positron in the opposite direction. Given the composition of W events is roughly two thirds W^+ events to one third W^-, one might expect some benefit from these polarisation effects. The performance of this method at high jet momentum will be determined by the interplay of these two effects.

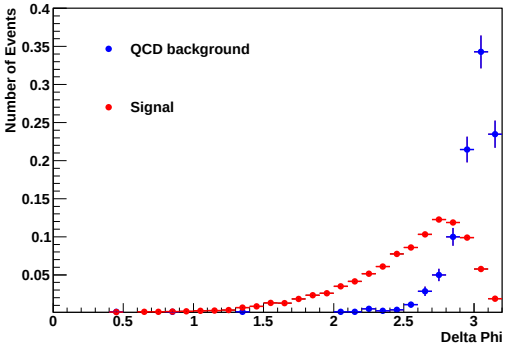
To characterise the topology of the event, the azimuthal angle between the jet and the lepton has been used. Distributions of $\Delta\phi$ for signal and QCD background events, for three different leading jet transverse momentum thresholds, are shown in figure 6.1. The signal distributions are taken from ALPGEN Monte Carlo simulations, whereas the background distributions are taken directly from data by requiring that the normal signal selections are satisfied, except that the missing transverse momentum is required to be less than 15 GeV.



(a) 30 GeV Threshold



(b) 60 GeV Threshold

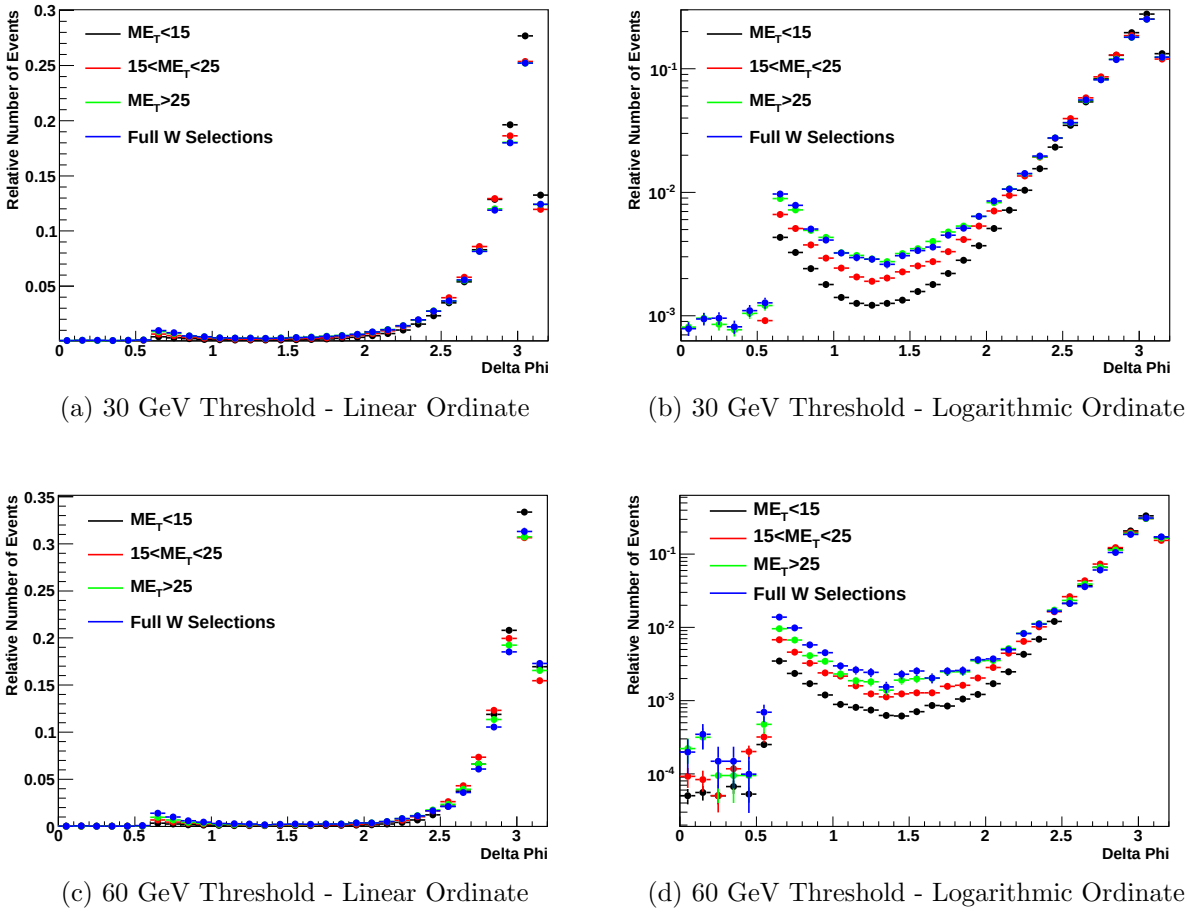


(c) 90 GeV Threshold

Figure 6.1: Comparisons of the signal and background $\Delta\phi$ distributions for different leading jet transverse momentum thresholds. The signal distributions are taken from Monte Carlo while the background distributions are obtained from data by requiring the event $E_{\mathrm{T}}^{\mathrm{miss}}$ be less than 15 GeV.

One of the primary benefits of this method is that the $\Delta\phi$ distributions of the QCD background are largely insensitive to changes in the quality requirements placed on the electron and also on the missing transverse momentum cut applied. The distributions do vary with the jet transverse momentum threshold however, becoming more strongly peaked towards π radians at higher jet transverse momenta. Consequently, any selection which changes the likelihood that a jet is systematically mismeasured will lead to a change in the shapes of the templates for the QCD background.

Figure 6.2 shows the sensitivity of the QCD $\Delta\phi$ distributions to changes in the missing transverse momentum requirement. These distributions were obtained from data by relaxing the quality requirement on the electron from 'tight' to 'loose and not tight'. Similarly figure 6.3 shows the variations in the azimuthal angle distributions when one changes the



(a) 30 GeV Threshold - Linear Ordinate

(b) 30 GeV Threshold - Logarithmic Ordinate

(c) 60 GeV Threshold - Linear Ordinate

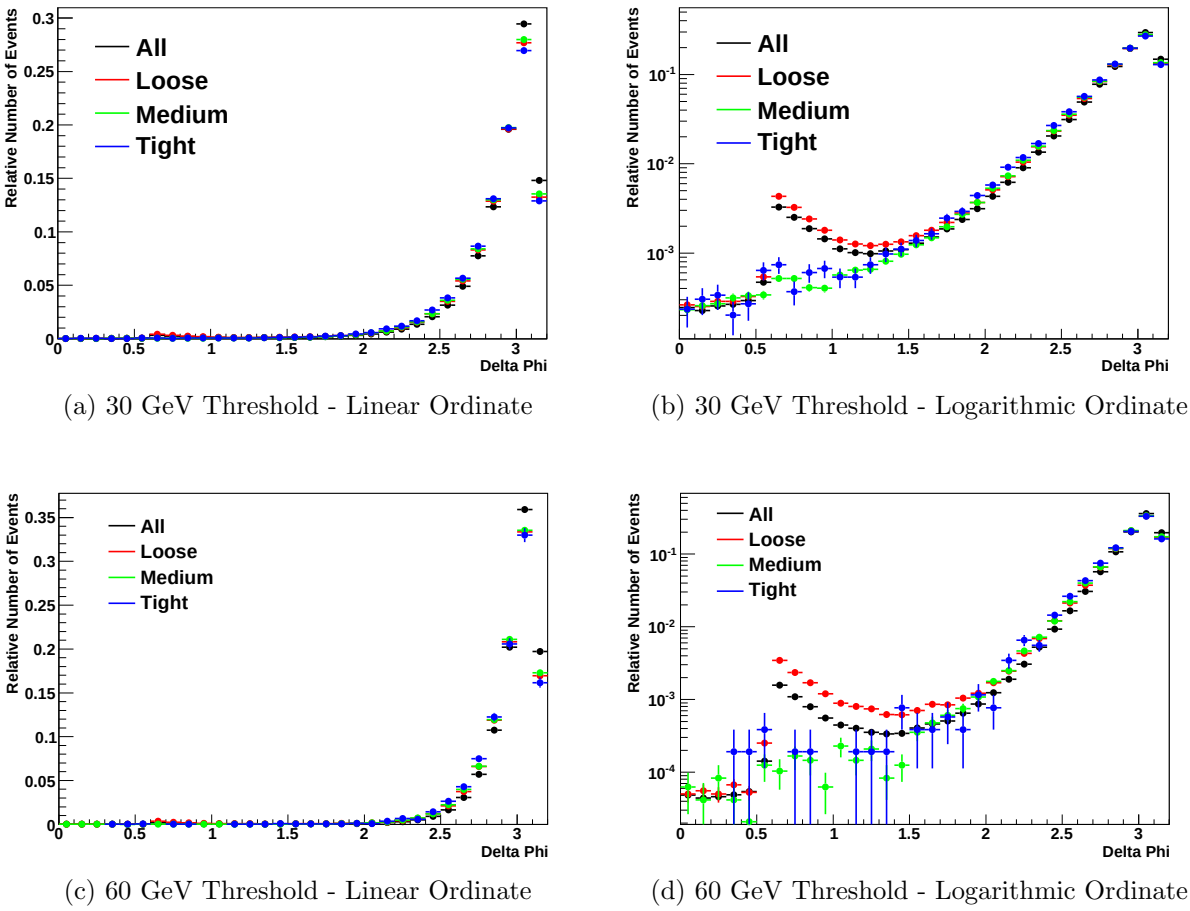
(d) 60 GeV Threshold - Logarithmic Ordinate

Figure 6.2: Plots showing the sensitivity of the azimuthal separation ($\Delta\phi$) of the jet and the fake lepton in QCD events to changes in the $E_{\mathrm{T}}^{\mathrm{miss}}$ selections. The top row have a leading jet P_T threshold of 30 GeV, the bottom have a threshold of 60 GeV. 'Full W Selection' indicates that the transverse mass selection has been applied in addition to the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut.

quality requirement on the lepton. These distributions were obtained by requiring the event missing transverse momentum be less than 15 GeV.

In order to minimise bias whilst still having sufficient statistics in the QCD distributions, the QCD background template was constructed from events which satisfy the following criteria:

- Missing transverse momentum $>$ 25 GeV
- Transverse Mass $>$ 40 GeV
- Electron candidate passes medium quality requirements



(a) 30 GeV Threshold - Linear Ordinate

(b) 30 GeV Threshold - Logarithmic Ordinate

(c) 60 GeV Threshold - Linear Ordinate

(d) 60 GeV Threshold - Logarithmic Ordinate

Figure 6.3: Plots showing the sensitivity of the azimuthal separation ($\Delta\phi$) of the jet and the fake lepton in QCD events to changes in the quality requirement imposed on the lepton. The top row have a leading jet P_T threshold of 30 GeV, the bottom row have a threshold of 60 GeV. 'All' indicates that no quality criteria were imposed and all reconstructed electrons were included.

- Electron candidate fails tight quality requirements
- $etcone20 > 4$ GeV. $etcone20$ is an isolation variable defined as the amount of transverse energy in a cone of $\Delta R < 0.2$ around the electron candidate after subtraction of the transverse energy of the electron cluster.

The fraction of the events which pass the above selections which are due to QCD processes (i.e. events with fake electrons) can be estimated from the a posteriori knowledge that the QCD background fraction in the signal region is $\sim 16\%$. The QCD fraction B_1 of events satisfying some set of selections 1 (such as the R_{Jets} signal selections) can be expressed in terms of the number of electroweak events N_1^S and the number of QCD events N_1^B as follows:

$$B_1 = \frac{N_1^B}{N_1^B + N_1^S} \Rightarrow (\frac{1}{B_1} - 1) = \frac{N_1^S}{N_1^B} \tag{6.3}$$

Now if the event selections are changed from set 1 to set 2 (i.e. the selections defining the region from which the background template is extracted) such that there is an electroweak ratio $R^S = N_2^S/N_1^S$ and a QCD background ratio $R^B = N_2^B/N_1^B$, then the QCD fraction for the new set of selections is related to the initial QCD fraction for the original set of selections via the following relation:

$$\frac{1}{B_2} - 1 = \frac{N_2^S}{N_2^B} = \frac{R^S N_1^S}{R^B N_1^B} = \frac{R^S}{R^B}(1 - \frac{1}{B_1}) \Rightarrow B_2 = \frac{B_1}{B_1 + \frac{R^S}{R^B}(1 - B_1)} \tag{6.4}$$

The electroweak ratio between the signal region and the phase space defined by the selections listed above can be calculated from Monte Carlo simulation giving a value of 0.0031. For the background ratio, it is necessary to apply some additional orthogonal selections to remove signal contamination and then assume that this additional selection does not affect the background ratio for the change $1 \rightarrow 2$. To calculate the background ratio on going from the standard tight electron selections to the configuration listed above, the missing transverse momentum was required to be less than 15 GeV. This produced a background ratio of 3.78. Using these values in equation 6.3 gives a QCD fraction of 99.6% in the region from which the QCD template is extracted. While these numbers are not expected to be extremely accurate, they give confidence that the selections chosen give a reasonable level of QCD purity for this method.

Electroweak Template

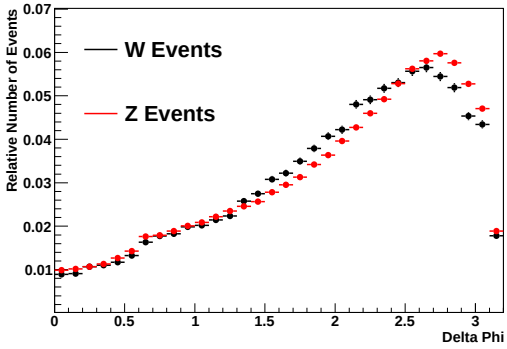
For the current study, the $We\nu$ and electroweak background templates have been taken directly from Monte Carlo simulations, including the effect of pile-up interactions as described in section 4.4. The electroweak template is constructed by forming a cross-section weighted sum of the $\Delta\phi$ distributions for events passing all signal selections from each of the pertinent electroweak background sources and the $We\nu$ signal itself. The electroweak processes

considered in the construction of the template are as follows:

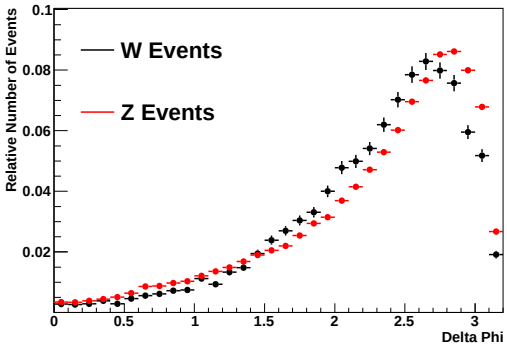
- $W \rightarrow e\nu$
- $W \rightarrow \tau\nu$
- $Z \rightarrow ee$
- $t\bar{t}$

The normalisation of the electroweak template was chosen such that the weight of the $We\nu$ ALPGEN NP1 sample was equal to one, as this sample is responsible for the majority of the signal contribution. The overall normalisation is important as it changes the cost in likelihood for the fitter to distort the shape of the electroweak template in any particular way. This can be understood by noting that the relative width of a Poisson distribution (as characterised by the ratio of its standard deviation and mean) scales as $\frac{1}{\sqrt{N}}$. Consequently, a larger normalisation results in smaller relative uncertainties on the bin contents of a histogram. This in turn means that a given fractional change of the contents of a particular bin is less likely to be the result of a Poissonian fluctuation when the normalization of the histogram is larger. A larger normalization on an input histogram therefore effectively restricts the extent to which the fitter can change the shape of that histogram in performing the fit by increasing the cost this incurs in likelihood. This will affect the results of the template fit and the estimate of the background fraction so obtained.

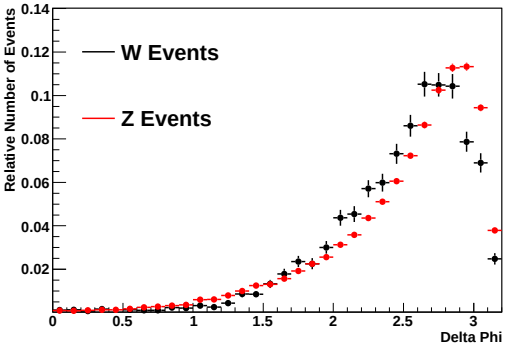
In the future it would be preferable to extract the electroweak template in a data-driven way. The most obvious way to do this would be to look at the electron - jet $\Delta\phi$ distributions in $Z+1$ jet events. These are not the same as the distributions one would expect from $W+1$ jet events due to the differing phase spaces of the observable final states and also because of the differing chiral structure of the coupling of the leptons to W and Z bosons. Due to these differences it is not possible to take the electroweak template directly from Z data, but it is possible to construct a partially data-driven electroweak template by scaling the Z angular distribution observed in data by a correction factor calculated in Monte Carlo to



(a) 30 GeV Threshold



(b) 60 GeV Threshold



(c) 90 GeV Threshold

Figure 6.4: Comparisons of the W and $Z+1$ jet $\Delta\phi$ distributions for different leading jet transverse momentum thresholds.

account for the W/Z differences (and differences due to the electroweak backgrounds). This method would be expected to have smaller systematic uncertainties, as the stability of the correction factor with respect to the various sources of systematic uncertainty is likely to be better than the stability of either of the angular distributions from which it is constructed. With the 2010 ATLAS dataset there is insufficient Z data to investigate this possibility. The $\Delta\phi$ distributions for W and $Z+1$ jet events are shown in figure 6.4 for three different jet transverse momentum thresholds. Figure 6.5 compares the $Z+1$ jet $\Delta\phi$ distributions in data and Monte Carlo.

Monte Carlo Closure Tests

Prior to the application of this method to determine the background fraction in data, a series of Monte Carlo closure tests were performed using pseudodata in order to test the

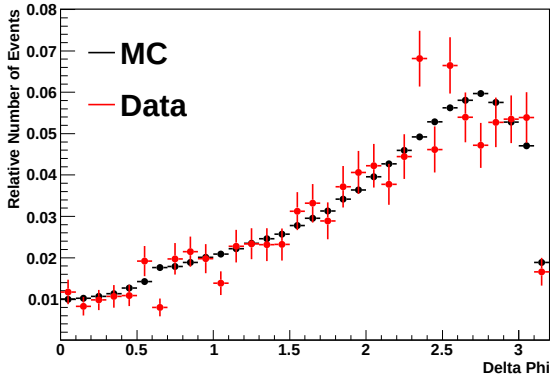
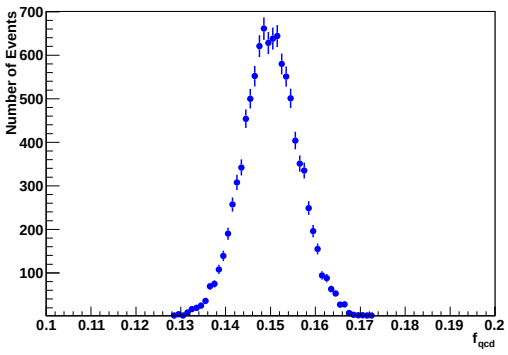
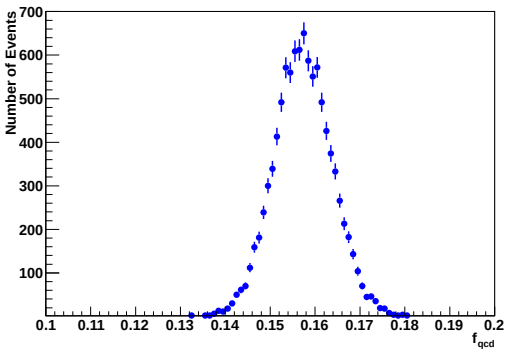


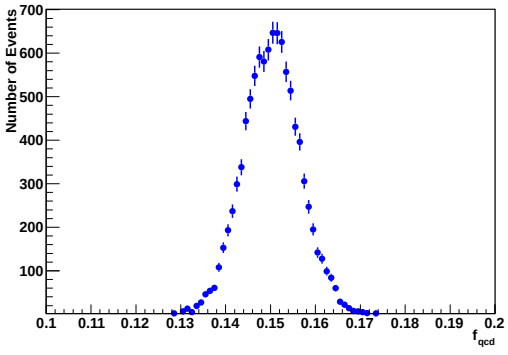
Figure 6.5: Comparison of the $Z + 1$ jet $\Delta\phi$ distribution in data and in Monte Carlo. The statistics of the data sample are insufficient to construct a partially data-driven electroweak template at present.



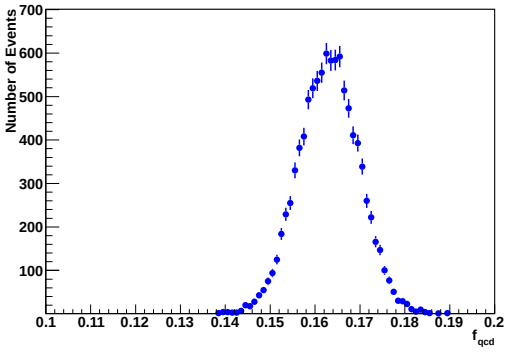
(a) Tight + $E_{\mathrm{T}}^{\mathrm{miss}} < 15$ GeV



(b) Medium + 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} < 25$ GeV



(c) Loose + $E_{\mathrm{T}}^{\mathrm{miss}} < 15$ GeV



(d) Loose + 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} < 25$ GeV

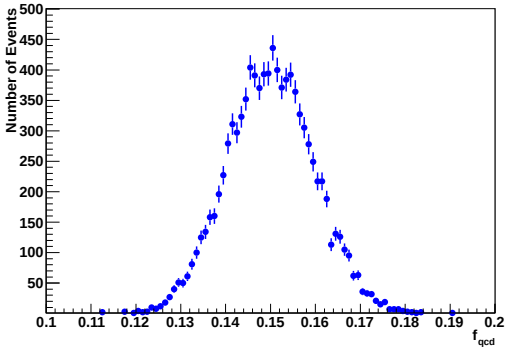
Figure 6.6: Plots showing the distributions of f_{qcd} obtained using different background templates as described in the main text for events passing a 30 GeV leading jet P_T threshold.

performance of the procedure. A set of electroweak templates were generated using Monte Carlo. These templates were then added to QCD background distributions obtained from data events with the normal tight electron quality requirement applied, but with missing transverse momentum less than 15 GeV. The composition of the templates was chosen such that the pseudodata distribution had a background fraction of 15%. The resulting distribution was then normalised to the number of events observed in data for a given jet transverse momentum threshold and fluctuated bin by bin according to a Poisson distribution with a mean given by the bin contents. This process gives a pseudodata distribution which should behave in a manner similar to the observed data.

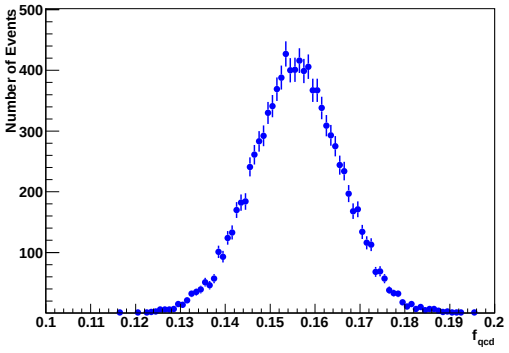
The pseudodata distribution can then be fitted with the templates used in its construction, in order to check that this returns the correct background fraction. It can also be fitted using background templates obtained from different kinematic regions and using different lepton quality criteria, in order to gauge the effect that the necessary cut reversals may have on the extracted background fraction. Figures 6.6 and 6.7 show the distributions of background fractions obtained from 10000 sets of pseudodata obtained from Poisson fluctuations as described above, for 30 GeV and 60 GeV leading jet transverse momentum thresholds. In each case, the first plot shows the distribution obtained using the same background template in the fit as was used in the construction of the pseudodata. This produced mean background fractions of 15.00% in both cases, with RMS values of 0.62% and 0.96% respectively. The following three plots show the distributions of background fractions obtained using the following alternative event selections to extract the background template:

- Medium and not Tight quality, 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 25 GeV
- Loose and not Tight quality, $E_{\mathrm{T}}^{\mathrm{miss}} <$ 15 GeV
- Loose and not Tight quality, 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 25 GeV

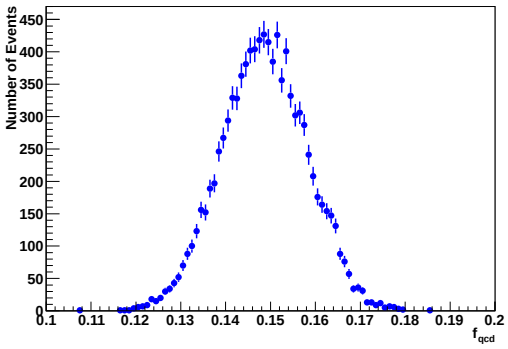
The largest deviation in both cases was produced when using the template with Loose quality and 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 25 GeV, giving values of 16.3% and 16.1% respectively for the two thresholds. These deviations would correspond to relative shifts on R_{Jets} of 1.53% and



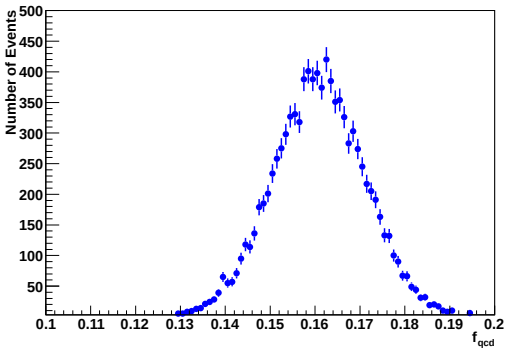
(a) Tight + $E_{\mathrm{T}}^{\mathrm{miss}} < 15$ GeV



(b) Medium + 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} < 25$ GeV



(c) Loose + $E_{\mathrm{T}}^{\mathrm{miss}} < 15$ GeV



(d) Loose + 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} < 25$ GeV

Figure 6.7: Plots showing the distributions of f_{qcd} obtained using the different background templates as described in the main text for events passing a 60 GeV leading jet P_T threshold.

	30 GeV	40 GeV	50 GeV	60 GeV	70 GeV
$f_{QCD}[\%]$	15.55 ± 0.71	15.36 ± 0.92	15.33 ± 1.13	14.36 ± 1.39	14.20 ± 1.73
	80 GeV	90 GeV	100 GeV	110 GeV	
$f_{QCD}[\%]$	14.10 ± 2.07	13.87 ± 2.36	15.03 ± 2.93	17.38 ± 3.86	

Table 6.1: The nominal values of f_{qcd} extracted from the fits along with their pure statistical uncertainties for jet P_T thresholds ranging from 30 GeV to 110 GeV.

1.29% respectively. Note that these alternative templates represent fairly dramatic changes with respect to the nominal selections used and still only produce percent level variations on the ratio.

Fit Results

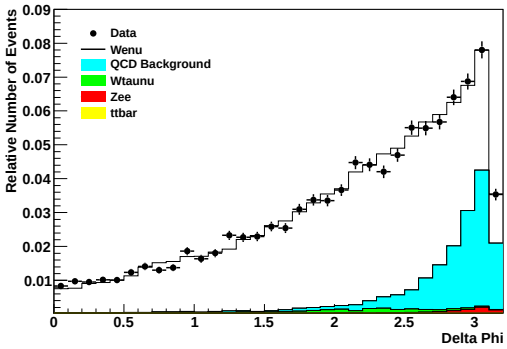
Figure 6.8 shows the results of the template fits to data for jet transverse momentum thresholds from 30 GeV to 100 GeV. The plots also show the composition of the obtained fits in

terms of the signal, the QCD background and the different processes contributing to the electroweak background. The fits were carried out using the entire range of the $\Delta\phi$ distributions between zero and π. The nominal bin width for the $\Delta\phi$ distributions was chosen to be 0.1 for reasons that will be discussed below. The values of the extracted background fractions as well as their statistical uncertainties are listed in table 6.1 for leading jet P_T thresholds up to 110 GeV, beyond which the statistics in data become inadequate to perform the fit.

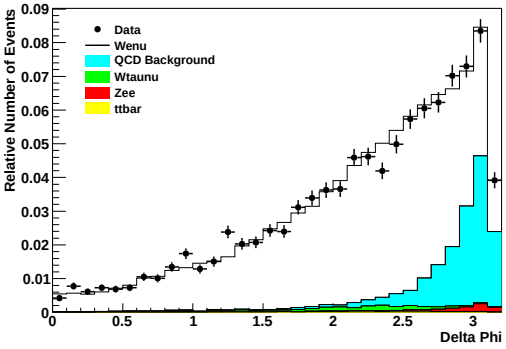
The statistical uncertainties were evaluated by fluctuating the bin contents of the data distribution and both the electroweak and QCD templates according to Poisson statistics, with a mean given by the nominal bin contents. This process was repeated 10000 times and for each repetition the data was refitted using the deformed templates to obtain a new background fraction. The width of the resulting distribution was taken as the statistical uncertainty. The statistical uncertainties for different bin widths ranging from 0.1 to 0.5 in $\Delta\phi$ are shown on the left hand side of figure 6.9. These increase roughly linearly from just over 4% at low jet P_T to 20% at 110 GeV. The right hand side of figure 6.9 shows the relative deviations between the mean values of the distributions and the nominal fit values. These should be close to 0 but if the templates used in the fit are underpopulated their values can deviate from this. From these results the nominal bin width was chosen to be 0.1 as this produces both the lowest statistical uncertainty below 100 GeV and the smallest deviation between the nominal fit value and the mean of the distribution of pseudoexperiment fits, which remained below 2% for all transverse momenta considered. The deviation between the mean and the nominal fit result was considered as an additional systematic uncertainty although it is completely dominated by the statistical uncertainty at all jet P_T.

Electroweak Template Systematics

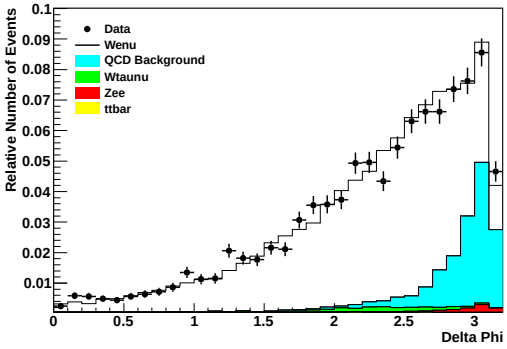
Four sources of systematic uncertainty related to the construction of the electroweak template have been considered: uncertainties on the cross sections of the samples, uncertainties related to the detector simulation, generator uncertainties present in the modelling of the event at particle level and pile-up uncertainties. With the exception of the last systematic, the



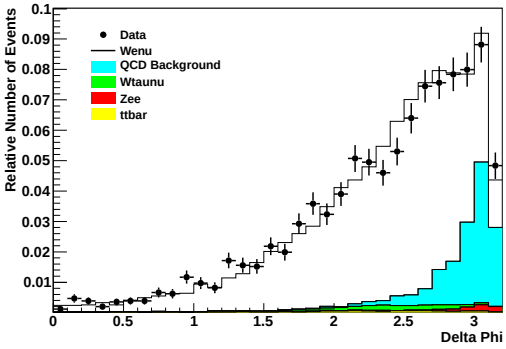
(a) 30 GeV P_T Threshold



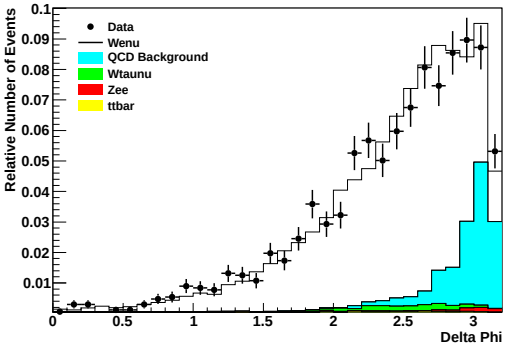
(b) 40 GeV P_T Threshold



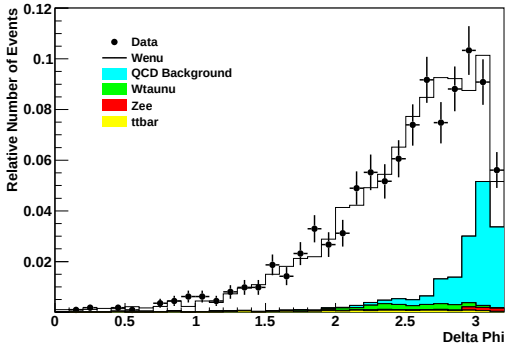
(c) 50 GeV P_T Threshold



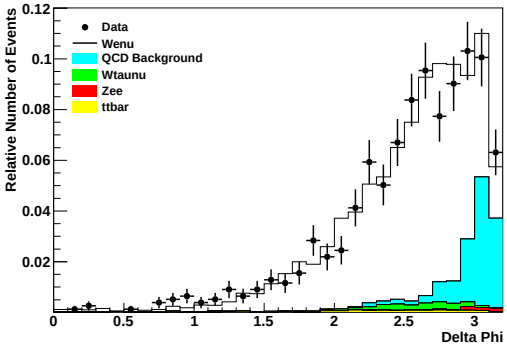
(d) 60 GeV P_T Threshold



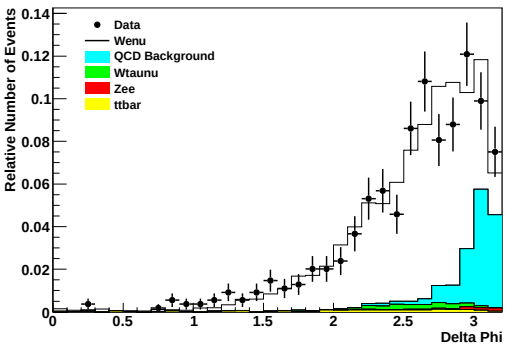
(e) 70 GeV P_T Threshold



(f) 80 GeV P_T Threshold



(g) 90 GeV P_T Threshold



(h) 100 GeV P_T Threshold

Figure 6.8: Plots showing the results of the template fits for different leading jet transverse momenta. The fit results have been decomposed into their contributions from signal (white), QCD (pale blue) and the various electroweak backgrounds.

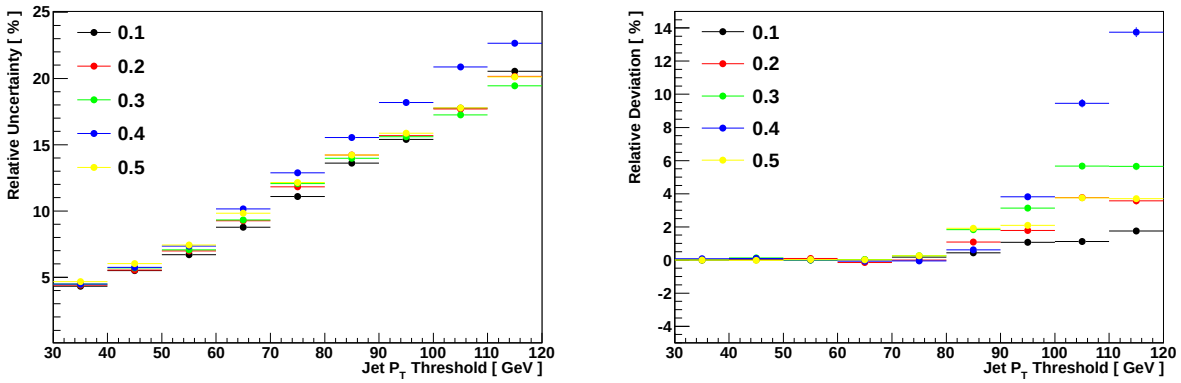


Figure 6.9: The left hand plot shows the relative statistical uncertainties on f_{qcd} as a function of the leading jet transverse momentum threshold, obtained from the width of a distribution of values constructed from 10000 pseudoexperiments performed with distorted templates and data. The right hand plot shows the relative deviation between the mean value of this distribution and the nominal fit value. Results for five different choices of bin width are shown.

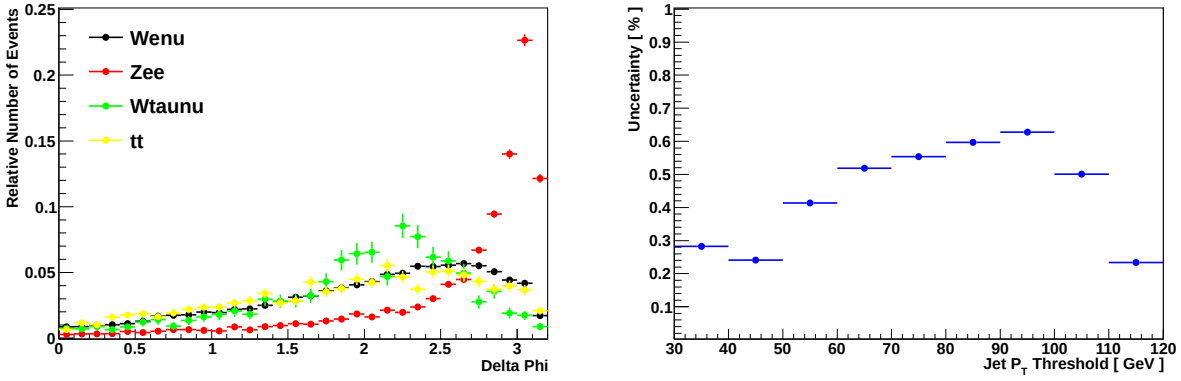


Figure 6.10: The left hand plot compares the shape of the signal $\Delta\phi$ distribution with those of the dominant electroweak backgrounds. The right hand plot shows the systematic uncertainty resulting from uncertainties on the cross sections of the Monte Carlo samples.

uncertainties were derived using Monte Carlo samples with no pile-up simulation to reduce the impact of poor statistics.

The different electroweak backgrounds and the signal produce somewhat different $\Delta\phi$ distributions and consequently uncertainties on the cross sections of the samples, which cause variations in the signal template composition, will result in variations in the shape of the template. These shape variations can result in changes in the extracted background fraction. To assess the impact of this, 10000 pseudoexperiments were performed where the cross sections of each Monte Carlo sample were drawn randomly from a Gaussian distribution

Process	Sample	Cross Section [nb]	Uncertainty [nb]
$W \to e\nu$	Np0	6.913	4.678×10^{-1}
	Np1	1.293	1.434×10^{-1}
	Np2	3.771×10^{-1}	3.665×10^{-2}
	Np3	1.009×10^{-1}	1.049×10^{-2}
	Np4	2.53×10^{-2}	1.69×10^{-3}
	Np5	6.9×10^{-3}	3.88×10^{-4}
$W \to \tau\nu$	Np0	6.836	4.257×10^{-1}
	Np1	1.277	1.386×10^{-1}
	Np2	3.766×10^{-1}	4.380×10^{-2}
	Np3	1.008×10^{-1}	8.391×10^{-3}
	Np4	2.57×10^{-2}	1.988×10^{-3}
	Np5	7×10^{-3}	3.045×10^{-4}
$Z \to ee$	Np0	6.641×10^{-1}	3.657×10^{-2}
	Np1	1.330×10^{-1}	9.436×10^{-3}
	Np2	4.023×10^{-2}	2.396×10^{-3}
	Np3	1.114×10^{-2}	6.233×10^{-4}
	Np4	2.893×10^{-3}	1.908×10^{-4}
	Np5	7.534×10^{-4}	-

Table 6.2: The above table lists the nominal cross sections and uncertainties of the samples used in the construction of the electroweak template and the evaluation of its composition uncertainty, taken from [97]. The $t\bar{t}$ uncertainty was not considered due to its negligible contribution to the background.

with mean equal to the nominal cross section and width equal to the uncertainty on this value. The fits were repeated and the widths of the resulting distributions of background fractions were taken as systematic uncertainties. Figure 6.10 shows the shapes of the different electroweak $\Delta\phi$ distributions for a 30 GeV leading jet P_T threshold as well as the uncertainty on the background fractions resulting from the cross section uncertainties. The resulting systematic uncertainty is below 1% for all jet P_T thresholds considered. The process cross sections and uncertainties are listed in table 6.2.

The dominant detector systematic affecting the electroweak template shape is due to uncertainties on the scale and resolution of the reconstructed jets. A systematically low jet energy scale (JES) for instance, would result in events being shifted to lower jet P_T bins, and as the shape of the angular distributions vary markedly with jet transverse momentum, this would result in a bias in the calculated background fractions. This uncertainty was evaluated by simply shifting the transverse momentum of all jets in the event up and down

by their JES uncertainty, reapplying the jet selections, reforming the electroweak templates and repeating the fits. The symmetrised deviation between these values and the nominal value was taken as a systematic uncertainty.

A systematic uncertainty related to the jet energy resolution (JER) uncertainty was calculated by again performing pseudoexperiments where the jets in the event were randomly smeared according to a Gaussian with a width given by the jet energy resolution uncertainty. As above, the selections were reapplied on the new jets, the template fits were repeated and the widths of the resulting distributions of f_{qcd} were taken as a systematic uncertainty.

The changes in f_{qcd} due to shifting the jets according to their JES uncertainty are shown on the left hand side of figure 6.11. The right hand plot shows the widths of the distributions of f_{qcd} obtained by smearing the jets according to their JER uncertainty. The JER uncertainty on f_{qcd} is negligible at less then 1% for all jet transverse momenta considered. The JES uncertainty however has a more pronounced effect, resulting in deviations in f_{qcd} that rise from $\sim 3\%$ to $\sim 7\%$ at high transverse momentum.

A generator uncertainty was included by calculating the difference between the background fractions obtained when taking the electroweak templates from Pythia and ALPGEN. The Pythia-ALPGEN difference is shown as a function of the jet P_T threshold on the left hand side of figure 6.12. The last systematic related to the electroweak template is the uncertainty due to modelling of the pile-up present in the event. For this, the difference between the predictions with and without pile-up was used. This method is coarse but conservative. In the future, uncertainties on the vertex weights could be calculated and used to evaluate this uncertainty. The pile-up uncertainty is shown on the right hand side of figure 6.12.

QCD Template Systematics

The systematic uncertainty on the QCD background template was treated in a manner analogous to that used for the Monte Carlo tests discussed in section 6.3.1. Fits were

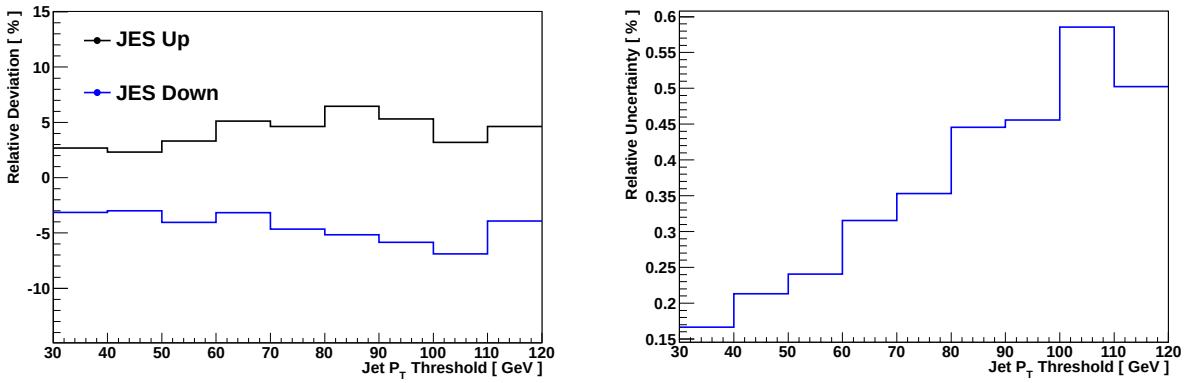


Figure 6.11: The left hand plot shows the deviations in f_{qcd} as a function of the leading jet transverse momentum threshold, obtained by shifting the jets in the events used to construct the electroweak template up and down according to their jet energy scale uncertainty. The right hand plot shows the uncertainty due to smearing the jets using their jet energy resolution uncertainty as described in the main text.

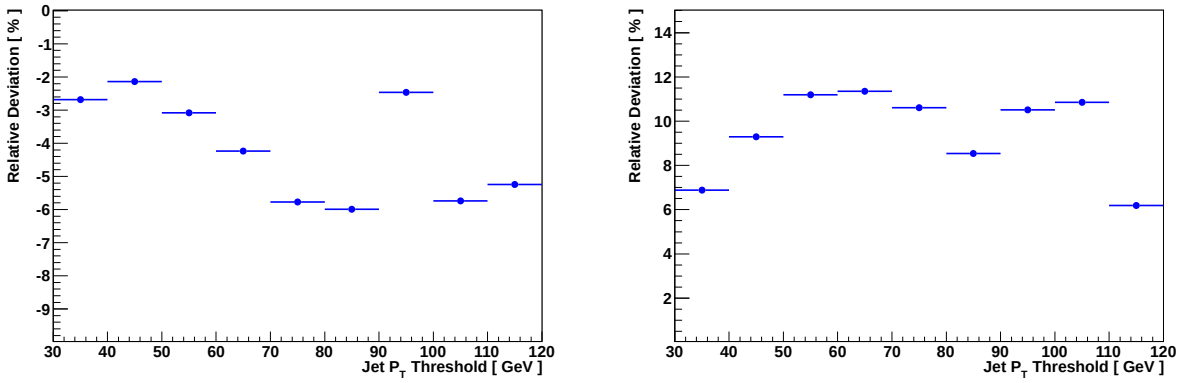


Figure 6.12: The left hand plot shows the relative deviation between background fractions calculated using ALPGEN and Pythia as a function of the jet P_T threshold. The right hand plot shows the relative deviation between background fractions obtained using events with and without pile-up simulation.

performed with background templates obtained using alternative selections, and the envelope of the deviations of the predictions obtained using these alternative templates from the nominal predictions was taken as a systematic uncertainty. The selections defining the regions of phase space from which the alternate templates were constructed are as follows:

- 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 25 GeV, No M_T cut, Loose and not Tight electron quality.
- 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 25 GeV, No M_T cut, Medium and not Tight electron quality.
- $E_{\mathrm{T}}^{\mathrm{miss}} >$ 25 GeV, $M_T >$ 40 GeV, Loose and not Tight electron quality.

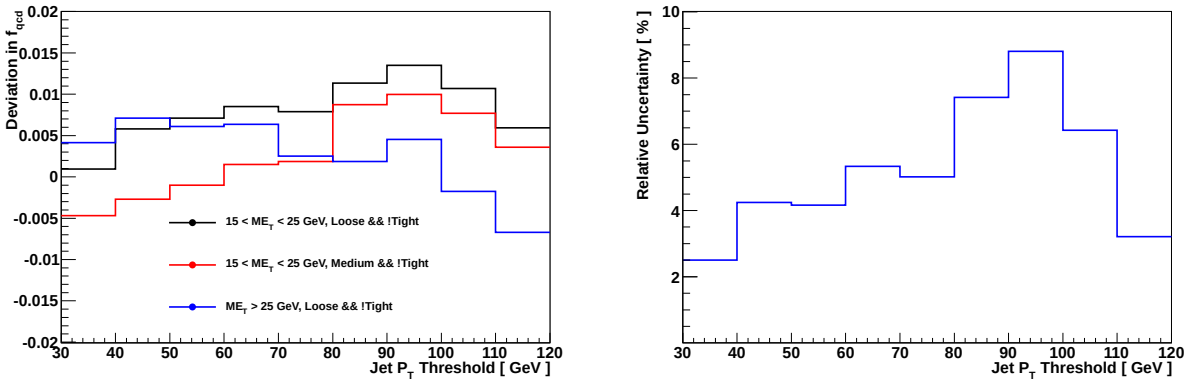


Figure 6.13: The left hand plot shows the deviations in f_{qcd} as a function of the jet P_T threshold when using the alternative background templates listed in the main text. The right hand plot shows the relative uncertainty on f_{qcd} constructed from the envelope of these deviations.

In all cases the isolation requirement listed in section 6.3.1 was applied. The deviations of the background fractions from their nominal values when using these alternative templates are shown in figure 6.13. The deviations are all within 0.01 up to a leading jet P_T of 80 GeV. The fact that the deviations are small and similar in magnitude to those seen in section 6.3.1 is an indication that a good level of signal rejection has been achieved. If this were not the case, larger variations would be seen when using the different templates due to changes in the amount of signal contamination.

Discussion

Figure 6.14 shows the QCD background fraction as a function of the leading jet P_T threshold up to 110 GeV along with its total uncertainty, which is the quadrature sum of the statistical uncertainty and all of the systematic sources discussed in the previous sections. The right hand plot shows the relative impact of these different sources of uncertainty on the total as they are successively added in quadrature. The uncertainty at present ranges from just below 10 % in the lowest jet P_T bin to just over 20 % at the highest P_T values considered.

At present, the dominant systematics are those associated with the electroweak template construction: The JES uncertainty, the pile-up uncertainty and the generator uncertainties. Using the partially data driven approach discussed in section 6.3.1 for constructing the signal

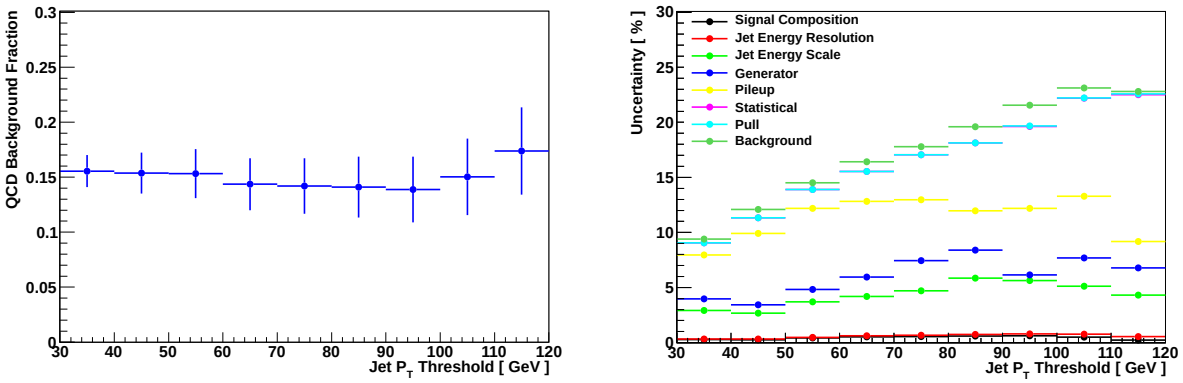


Figure 6.14: The left hand plot shows the QCD background fractions for different leading jet P_T thresholds with their total uncertainty. The right hand plot shows the composition of this uncertainty in terms of the statistical uncertainty and the different sources of systematic uncertainty discussed in prior sections. Note that in this plot the uncertainties are being successively added in quadrature.

template along with a more sophisticated treatment of the uncertainties on the vertex weights would significantly reduce the contribution of these sources to the total uncertainty.

The availability of more data may also improve the QCD template uncertainty, as fewer electron quality cuts will need to be inverted in order to have sufficient statistics and the systematic uncertainty can be evaluated by simply looking at the effect of each specific cut inverted on the template. The eventual limit on the accuracy of this method will be set by how few of the cuts comprising the Tight quality selections can be inverted whilst sill having good QCD purity in the template.

There are other options available for the construction of templates related to the topology of the event. One other example considered was the azimuthal angle between the lepton in the event and the missing transverse energy, $\Delta\phi_{lm}$. In a QCD background event, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is expected to be due to mismeasurement of one of the jets in the event and is therefore expected to be close to collinear with either the jet or the electron [1]. While the tight selections tend to favour the $E_{\mathrm{T}}^{\mathrm{miss}}$ being aligned with the jet faking an electron, these configurations are largely eliminated by the transverse mass cut, leading to QCD background distributions very similar to those seen above, with a large peak at π radians. In a genuine $W+1$ jet event, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is predominantly due to the daughter neutrino of the W and is therefore not necessarily

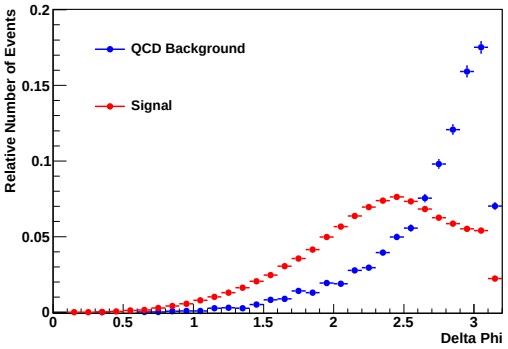
[1] Which from the above are expected to be close to collinear themselves.

collinear with the lepton and jet. In addition to this, as the transverse momentum of the W boson increases, the electron and neutrino are boosted along the W direction, leading to the signal distribution gradually moving to lower values of $\Delta\phi_{lm}$. Consequently the discriminating power of this variable improves markedly with jet P_T. This is a significant advantage of this method.

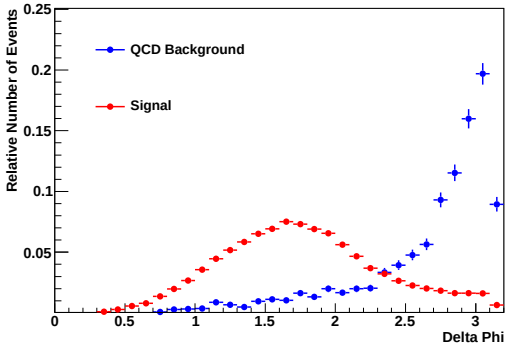
Figure 6.15 shows the $\Delta\phi_{lm}$ distributions for signal and background events for different leading jet P_T thresholds. The signal distributions are again taken from Monte Carlo, while the background distributions are constructed from data events passing medium quality criteria but failing tight, and subject to the same reverse isolation cut as listed in section 6.3.1. Figure 6.16 shows the sensitivity of the QCD $\Delta\phi_{lm}$ distributions on the $E_{\mathrm{T}}^{\mathrm{miss}}$ and quality selections for two leading jet transverse momentum thresholds. Unsurprisingly they are far less robust against variations in the event selections than the previous templates were and consequently for the same set of background selections, one would expect a larger bias in the QCD background distributions for this method than the last. This has not been investigated further here but offers an interesting complementary approach to the previous method.

Scalability

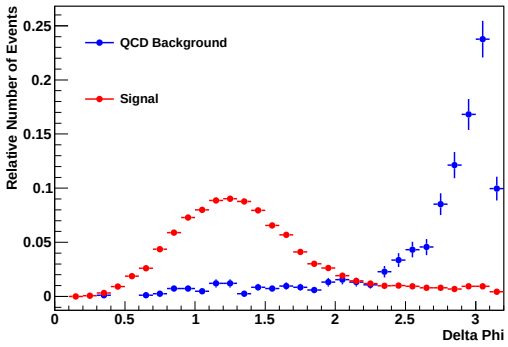
The final question related to this study concerns its scalability to higher jet multiplicities. The specific 1 jet topology, where the W boson directly recoils the jet, is not present for these cases, however, the W boson should still recoil against the total transverse momentum of the jet system and consequently it may be possible to construct a similar approach by looking at the azimuthal angle between the electron in the event and the recoil of the entire jet system. One would again expect such a distribution to be more strongly biased towards back-to-back configurations in QCD background events than in genuine signal events. As a proof of principle this has been carried out using events with two jets with transverse momentum greater than 30 GeV. The electroweak templates were again constructed from Monte Carlo, while the QCD background template was taken from data events with $E_{\mathrm{T}}^{\mathrm{miss}} < 15$ GeV and loose quality electrons. The relaxed background selections were necessitated by reduced



(a) 30 GeV Jet P_T Threshold



(b) 60 GeV Jet P_T Threshold



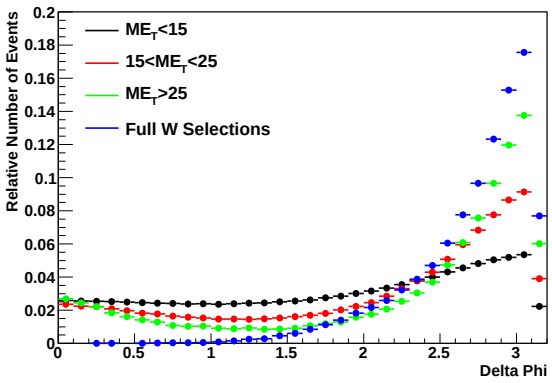
(c) 90 GeV Jet P_T Threshold

Figure 6.15: Comparisons of the signal and background $\Delta\phi_{lm}$ distributions for different leading jet transverse momentum thresholds. The signal distributions are taken from Monte Carlo, while the background distributions are obtained from data by requiring the lepton passes medium quality requirements whilst failing tight, and has etcone20 > 0.4 GeV.

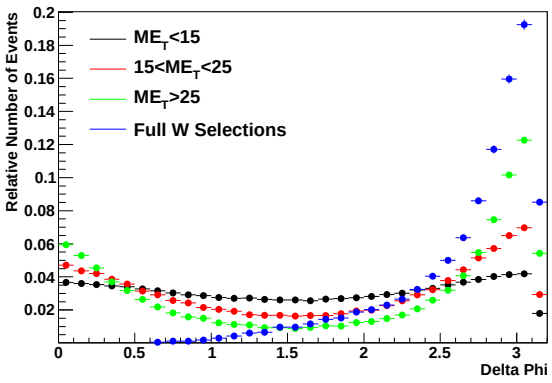
statistics in the dijet case. The left hand plot of figure 6.17 shows the signal and background angular distributions, while the right hand plot shows the template fit results. The extracted background fraction was 16.31% with a statistical uncertainty of 2.02% which is in line with what is expected [8].

6.3.2 $E_{\mathrm{T}}^{\mathrm{miss}}$ Template Method

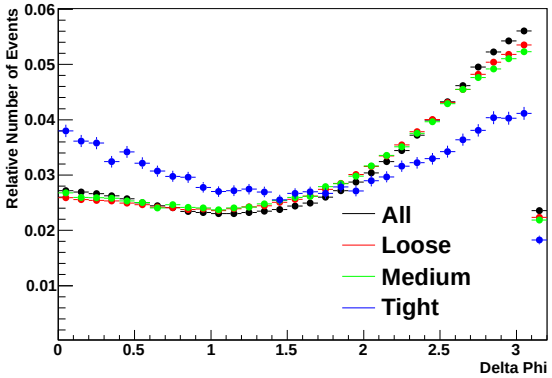
The second approach uses the missing transverse momentum as the variable with which to perform the template fit. This method is documented more thoroughly in [2]. The plots shown in this section are also taken from this source. Like the azimuthal distributions presented in the previous sections, this variable has the advantage that the signal and back-



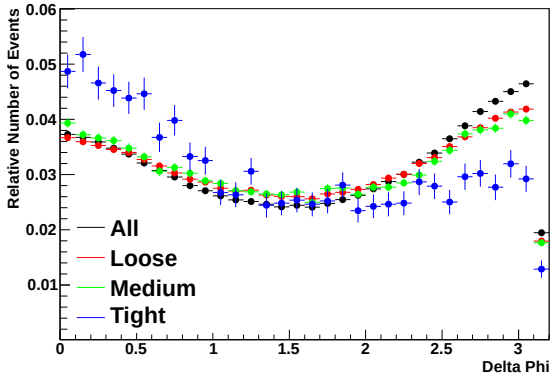
(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ sensitivity - 30 GeV P_T Threshold



(b) $E_{\mathrm{T}}^{\mathrm{miss}}$ sensitivity - 60 GeV P_T Threshold

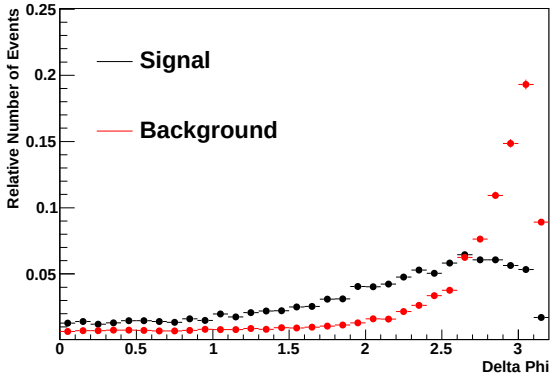


(c) Quality sensitivity - 30 GeV P_T Threshold



(d) Quality sensitivity - 60 GeV P_T Threshold

Figure 6.16: Plots showing the sensitivity of $\Delta\phi_{lm}$ in QCD events to changes in the missing transverse energy requirement and on the quality requirement imposed on the lepton.



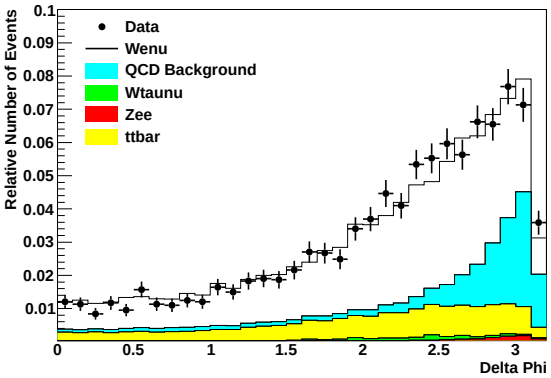


Figure 6.17: The left hand plot compares the shapes of the azimuthal angle distributions between the lepton and the jet recoil in 2 jet events for signal and background. The right hand plot shows the result of a template fit using this variable.

ground are clearly separated and differ markedly in shape, allowing the fitter to distinguish easily between the two.

The background template was constructed in a similar manner to the previous approach by requiring that the electron pass medium quality requirements but fail any two of the following tight requirements:

- **TrackBlayer** - The requirement that the electron have a hit in the inner layer of the pixel detector.
- **TrackMatchEoverP** - The requirement on the agreement between the energy of the cluster measured in the calorimeter and the momentum of the track measured in the ID.
- **TrackA0Tight** - The impact parameter requirements placed on the electron track.
- **TrackTRThits** - The requirements placed on the overall number of hits in the TRT.
- **TrackTRTratio** - The criteria placed on the ratio of high threshold TRT hits to low threshold TRT hits.

The electron is also required to be anti-isolated to ensure sufficient background purity[2]. The requirement that two of the above criteria fail was found to give optimal performance with respect to purity, compatibility and statistics as discussed in section 6.2. The electroweak template was again constructed from Monte Carlo. In future a partially data driven electroweak template could be constructed for this background subtraction method using the techniques discussed in chapter 8.

The selections presented in section 4.2 involve a cut on the missing transverse momentum at 25 GeV, however it is possible to perform the template fit over a larger region and then use the so obtained distributions and fractional composition to calculate the background

[2]The isolation requirement used in this method differs from that used in the previous section. The criterion used here involves a set of calorimeter and track based isolation cuts in a cone of 0.3 constructed to have 95% efficiency for isolated electrons.

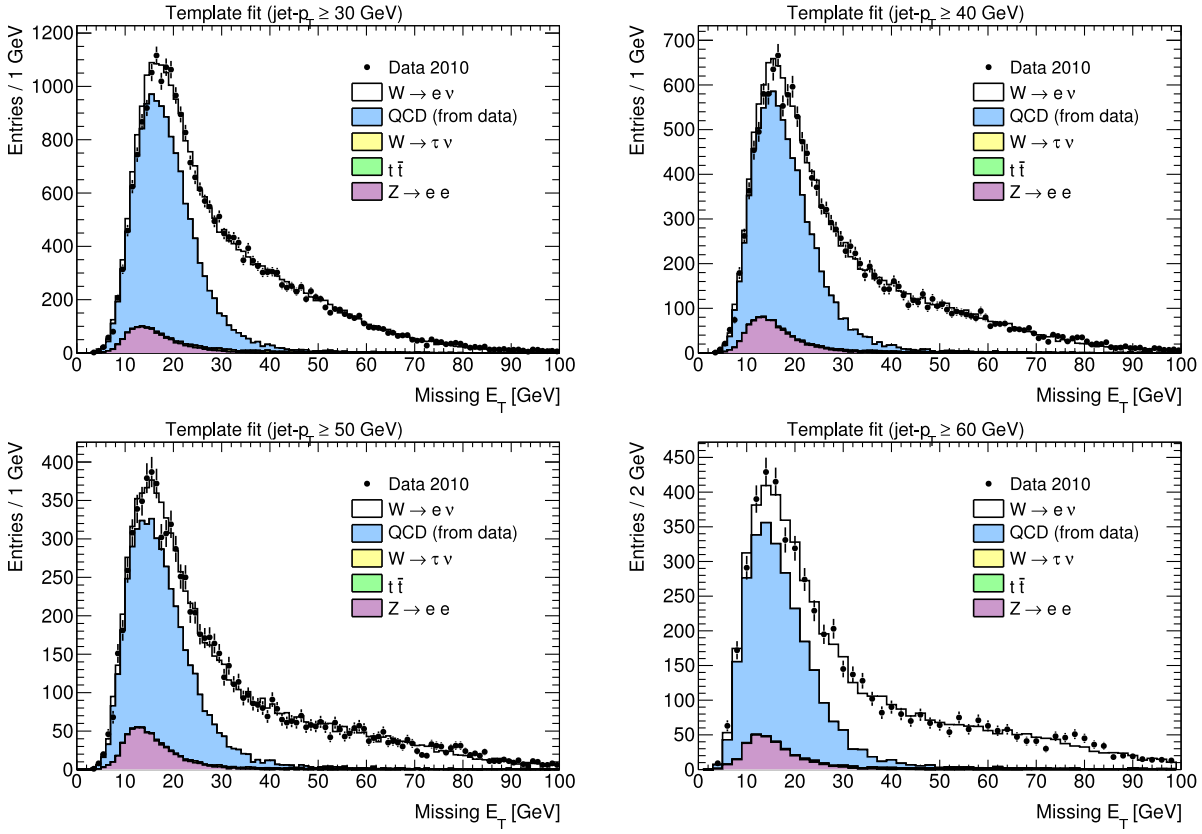


Figure 6.18: Plots showing the results of the $E_{\mathrm{T}}^{\mathrm{miss}}$ template fits for different leading jet transverse momenta. The jet P_T increases left to right and top to bottom from 30 GeV to 60 GeV inclusive.

fraction in the region $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV. Doing so has the advantage that significantly more data is available for performing the fit in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ region, leading to reduced statistical uncertainties. The fit region was consequently chosen to be 15 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 55 GeV.

Figure 6.18 shows the results of the template fits in four bins of jet P_T. Again fits were only performed up to 110 GeV, although in this case, the kinematic reach is limited by the statistics of the background template rather than by the number of data events.

Statistical uncertainties were assigned to the fit results using the same method as for the azimuthal fits. Because the accuracy of the extracted background fraction is determined primarily by the statistics of the QCD template when using this method, only this template was fluctuated. Deviations between the nominal fit values and the mean values of the distributions obtained by fluctuating the QCD templates were also calculated and the nominal bin widths were chosen in order to minimise these. The nominal bin widths were chosen to

be 1, 2 and 5 GeV for 30-50 GeV, 60-80 GeV and 90-110 GeV respectively.

An uncertainty on the chosen fit window was evaluated by varying the fit boundaries by ±10% both individually and in tandem. This results in a further eight values of the background fraction corresponding to the following variations of the fit window:

$$\{t, b\}\epsilon\{\uparrow 0, \downarrow 0, 0 \uparrow, 0 \downarrow, \uparrow\uparrow, \uparrow\downarrow, \downarrow\uparrow, \downarrow\downarrow\} \tag{6.5}$$

where t and b refer to the shifts applied to the upper and lower boundaries of the fit window respectively, 0 indicates that no shift was applied while $\uparrow$ and $\downarrow$ indicate that shifts of 10% were applied in the positive and negative directions respectively. The deviations between the background fractions obtained using the different fit windows and the nominal were used to form an error weighted average deviation and this was added in quadrature as an additional uncertainty.

In order to assign a systematic uncertainty on the choice of background template, three alternative templates composed of restricted sets of doublet inversions were used to calculate f_{qcd}. The deviations in f_{qcd} when using these control templates were calculated in the lowest jet P_T bin relative to the statistical uncertainty and added in quadrature giving a value of $\sigma(sys)/\sigma(stat) = 1.024$. This relative uncertainty was then applied to higher jet momentum thresholds. The three control templates were those constructed by considering only the two most well populated pairs of failed cuts and all pairs of failed cuts except each of these two. These were:

- All pairs of inversions except **TrackBlayer-TrackMatchEoverP**
- All pairs of inversions except **TrackTRTratio-TrackMatchEoverP**
- Only the above two pairs.

The resulting values of f_{qcd} along with their total uncertainties are shown in figure 6.19 and listed in table 6.3.

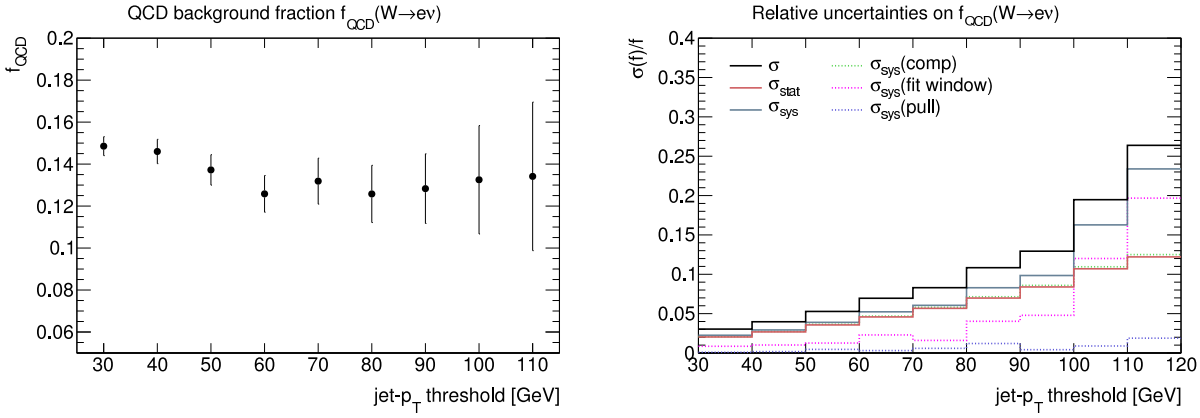


Figure 6.19: The left hand plot shows the value of f_{qcd} determined using the $E_{\mathrm{T}}^{\mathrm{miss}}$ template method along with its total uncertainty. The right hand plot shows the breakdown of these uncertainties into the components described in the main text. Note that the components are not being successively added in quadrature here although their quadrature sum σ is also shown.

	30 GeV	40 GeV	50 GeV	60 GeV	70 GeV
$f_{QCD}[\%]$	14.9 ± 0.5	14.6 ± 0.6	13.7 ± 0.7	12.6 ± 0.9	13.2 ± 1.1
	80 GeV	90 GeV	100 GeV	110 GeV	
$f_{QCD}[\%]$	12.6 ± 1.4	12.8 ± 1.7	13.3 ± 2.6	13.4 ± 3.5	

Table 6.3: f_{qcd} values along with their total uncertainties extracted using the $E_{\mathrm{T}}^{\mathrm{miss}}$ template method.

6.3.3 Method Comparison

The most pertinent uncertainties to consider in comparing the two presented methods are the statistical and QCD template uncertainties. The ability of the $E_{\mathrm{T}}^{\mathrm{miss}}$ template method to utilise the entirety of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in performing the fit, not just the signal region above 25 GeV, is a considerable advantage and as a result the statistical uncertainties for this method are roughly a factor of two smaller than for the $\Delta\phi$ method. The QCD template uncertainties however are of comparable magnitude for the two methods, rising roughly linearly from 2% in both cases, despite the fact that the alternative templates used in the evaluation of this uncertainty in the topological case represent significantly larger deviations from the nominal selections than in the $E_{\mathrm{T}}^{\mathrm{miss}}$ method.

In future, the exponentially increasing luminosity of the LHC will mean these background subtraction procedures rapidly become systematically dominated in the kinematic regions

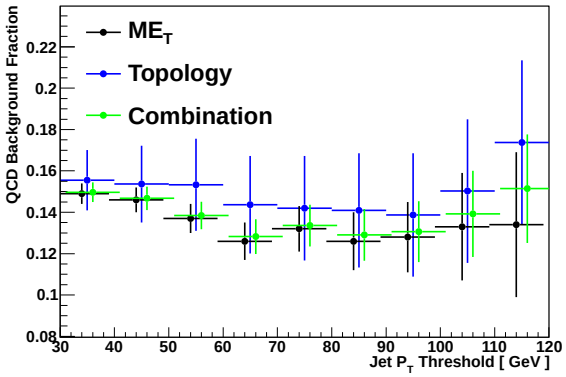


Figure 6.20: This plot shows the values of f_{qcd} determined using the two different methods presented as well as their combination as described in the main text. The values are compatible within their total uncertainties for all jet P_T thresholds considered.

considered here, and given the similar size of the background uncertainties, the performance of the two methods is likely to become more similar. This allows an improved estimation of the background fraction via an error weighted combination of their predictions as follows:

$$f_{qcd,Total} = \frac{\Sigma_i \frac{f_{qcd,i}}{\sigma_i^2}}{\Sigma_i \frac{1}{\sigma_i^2}} \qquad \frac{1}{\sigma_{Total}^2} = \Sigma_i \frac{1}{\sigma_i^2} \tag{6.6}$$

This combination has been performed and the results are show in figure 6.20. At present, the smaller statistical uncertainties and the fact that electroweak systematics were not considered, result in the combined value being strongly weighted towards the value obtained using the $E_{\mathrm{T}}^{\mathrm{miss}}$ template method.

An error weighted average of the values obtained using the $E_{\mathrm{T}}^{\mathrm{miss}}$ method in the region 60 - 110 GeV was used as an estimation of f_{qcd} in the region from 110 GeV to 200 GeV. The value obtained was $f_{qcd} = 12.8 \pm 0.6\%$. The fit uncertainty was added in quadrature with the uncertainty in the 110 GeV bin for this range.

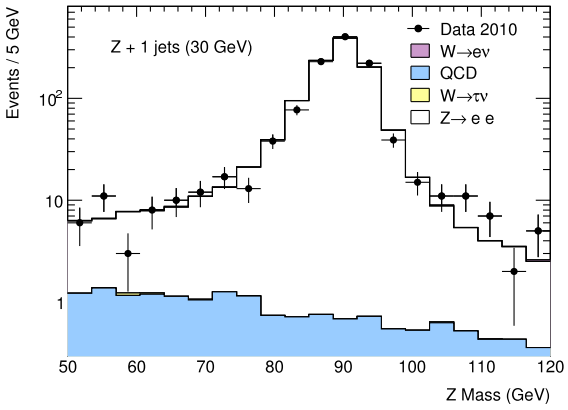


Figure 6.21: The results of the template fit to the Z invariant mass distribution constructed from events passing medium-medium quality selections [2].

6.4 QCD Background Subtraction for Z events

The QCD background fraction in Z events is evaluated using the same techniques as in the W case. The template variable chosen for performing the fit is the invariant mass of the dilepton pair in the event. This is complicated by the small background in Z events and consequently, in order to perform the fit, the signal selection requirements are relaxed from medium-tight to medium-medium. The required background fraction is then calculated from the result of the fit using the electron medium-tight rejection factor obtained from tag and probe studies [59], and the tight efficiency relative to medium for fake electrons using the procedure outlined at the end of section 6.3.1. The electroweak template is obtained from Monte Carlo as in the case of the W events and the background template is obtained by further relaxing the electron quality requirements to loose-loose with a medium veto. The result of the fit is shown in figure 6.21. Templates obtained from slightly stronger quality criteria (although still with a medium veto) were used to evaluate systematic uncertainties. Due to the weak statistics in Z data, the extracted background fraction was only evaluated in the lowest jet P_T bin and assumed constant with increasing jet P_T threshold. The resulting value for f_{qcd}^{Z} is 0.4 ± 0.3 %.

6.5 Electroweak Background Subtraction

The effect of the electroweak backgrounds can also be expressed via a background fraction. In this case, this is defined as the fraction of background events N_i, originating from a given electroweak process i, divided by the sum of all background and signal contributions as follows:

$$f^i_{ewk,V} = \frac{N_i}{N_V + \sum_j N_j} \tag{6.7}$$

The electroweak background in data is then subtracted from the signal after QCD background subtraction as follows:

$$N_{V,sig} = (1 - f_{ewk,V}) N_{V,sig+ewk} \tag{6.8}$$

$$f_{ewk,V} = \sum_i f^i_{ewk,V} \tag{6.9}$$

where the sum runs over all electroweak background processes.

Each of the contributions N_i is obtained from ALPGEN Monte Carlo simulations by taking a sample of a given background contribution and applying the appropriate selections. This method has the obvious advantage that there is no reliance on knowledge of the absolute luminosity in data. Further to this, one expects that variations due to both the generator modelling and detector simulation will contribute similarly in each process if the same generator and detector simulation are used, leading to a reduction of the associated uncertainties on the background fraction. The contributions to $f_{ewk,V}$ from various processes are listed in table 6.4. The total electroweak background is small, contributing approximately 4% of the total number of W events after QCD background subtraction and less than a percent of the Z events.

Systematic uncertainties on the background fraction are estimated by varying the dif-

ferent sources of uncertainty within a reasonable range, repeating the full background estimation and evaluating the difference with the nominal value. Sources of uncertainty which affect both W and Z events are varied in tandem when propagating their effect to R_{Jets}. The sources of uncertainty considered were:

- **Electron identification:** To estimate the impact of the lepton quality criteria on W events, the ID requirement on the electron was relaxed from tight to medium. In the case of the Z events, the ID criteria were relaxed from tight-medium to medium-loose. Since the uncertainty from fake electrons is already covered by the QCD background, the reconstructed electrons are required to be matched within a cone of 0.3 to a generated electron to prevent double counting of that contribution.

- **Electron $\mathbf{P_T}$ resolution:** The reconstructed electron P_T was replaced by the P_T of the generated electron.

- **Angular resolution:** To estimate the impact of the angular resolution of the electron track, the reconstructed track η was replaced by the generated electron η.

- $E_{\mathrm{T}}^{\mathrm{miss}}$ **resolution (W only):** To estimate the uncertainty due to the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ was replaced with the four-vector sum of the generated momenta of the neutrinos in the event.

- **Jet energy scale** To estimate the impact of the jet energy scale uncertainty on the jet acceptance and $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, the jet energy scale was varied by $\pm 5\%$.

- **Model uncertainty:** The uncertainty due to signal and background modelling was evaluated by comparing the background prediction determined using ALPGEN with that obtained using Pythia.

- **Effect of Pile-Up simulation:** Predictions calculated using Monte Carlo samples with pile-up simulation were compared with ones obtained from Monte Carlo samples without pile-up simulation.

Process	$f_{ewk,W}$ [%]	$f_{ewk,Z}$ [%]
Weν	-	0.30
Zee	1.09	-
W$\mu\nu$	0.03	0.00
W$\tau\nu$	1.96	0.02
Z$\mu\mu$	0.00	0.00
Z$\tau\tau$	0.20	0.03
$t\bar{t}$	0.34	0.21
single top	0.46	0.03
diboson	0.10	0.09
Total	4.19	0.68

Table 6.4: The calculated electroweak background fractions for a leading jet P_T threshold of 30 GeV.

Systematic	$\Delta f_{ewk,W}$ [%]	$\Delta f_{ewk,Z}$ [%]	ΔR_{Jets} [%]
JES+5%	-3.66	-7.07	-0.11
JES-5%	5.10	8.46	0.17
electron ID	-12.22	-52.41	-0.17
p_T Resolution	-1.02	-33.35	0.18
η Resolution	-0.34	-4.96	0.02
E_T^{miss} Resolution	-26.24	0.00	-1.15
Model	11.37	-24.88	0.67
Pile Up	-14.27	-6.24	-0.58

Table 6.5: This table lists the impact of the various sources of systematic uncertainty described in the main text on the yields of W and Z events and how this effect propagates through to the ratio.

Some of these variations are very conservative, for instance replacing reconstruction level quantities by particle level quantities and completely removing the pile-up simulation, however the small value of the electroweak background fraction leads to extremely small uncertainties on R_{Jets} and consequently the conservatively calculated errors on this quantity are still much smaller than other uncertainties present in the analysis. Table 6.5 shows the resulting systematic uncertainties on the yields and on the ratio. The systematic uncertainty on R_{Jets} is dominated by the uncertainty from the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution.

Chapter 7

Efficiency Corrections

7.1 Introduction

Two sets of efficiencies must be calculated in order to correct the observed value of R_{Jets} back to particle level. The first set, the determination of which will form the focus of the majority of the work in this chapter, give the probabilities that an event in a given jet P_T bin which passes the kinematic and geometrical event selections, also passes the lepton quality criteria for the corresponding type of event. These are referred to as identification efficiencies. The second set, which give the probability that an event which passes all offline selections also passes the appropriate trigger requirement, are known as trigger efficiencies. The calculation of the trigger efficiencies is discussed more briefly at the end of the chapter.

While both of these efficiencies are primarily dependent on the kinematics of the leptons in the events, specifically their transverse momenta and pseudorapidities, the R_{Jets} measurement is performed as a function of the threshold on the leading jet transverse momentum, and consequently in order to perform the correction of R_{Jets} back to particle level, average identification and trigger efficiencies must be determined in these bins. While it is not expected that the dependence of the efficiencies on the electron kinematics would change based on the jet activity given the events considered have a $\Delta R_{lj} > 0.6$ jet-lepton isolation criterion imposed, the changing jet kinematics do affect the distribution of events over the electron

$\eta - E_T$ space, and consequently the average efficiency changes. One would expect that the increasing jet transverse momentum would similarly result in the electrons from the decay of the vector boson being boosted to higher transverse energies, and as higher transverse energy electrons have higher identification and trigger efficiencies, (see the following sections) the average efficiency would be expected to steadily increase with the jet P_T threshold.

Figure 7.1 explores the claim that the dependence of the electron identification efficiency on the electron kinematics is unaffected by the jet activity, by showing the ratio of the tight identification efficiencies in W events with and without jet requirements imposed for a variety of different electron transverse energy and pseudorapidity bins. Two cases are considered, the requirement of a single jet with above 30 and 50 GeV of transverse momentum respectively. In both cases the ratios are close to one for the three different electron kinematic bins studied, suggesting that any dependence of the electron efficiencies on the jet kinematics is weak. Another demonstration of the independence of the electron efficiencies on the jet activity when binned in terms of the electron kinematic variables can be found in [98], where the efficiencies in various regions of the electron phase space are calculated as explicit functions of the jet multiplicity. Again no significant dependence is observed.

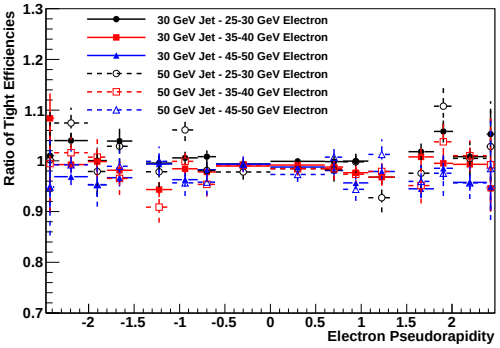


Figure 7.1: This plot shows the ratios of tight identification efficiencies calculated in W events with jets passing 30 and 50 GeV transverse momentum cuts to those calculated with no jet requirements imposed for a variety of electron η and E_T bins.

From the above discussion it is clear that the problem of determining the average identification efficiency in terms of the leading jet transverse momentum threshold reduces to knowing the dependence of the electron efficiency on the electron transverse momentum and pseudorapidity, and also the distribution of signal events over the same space for each jet P_T

bin. The former will be referred to from here onwards as an *efficiency map* and the latter will be referred to as a *population distribution*. Once these have been obtained, the efficiency for a given jet P_T bin can be determined from them by calculating a population weighted average of the efficiency map.

While in the case of the W events there is only a single efficiency which needs to be calculated, the tight electron identification efficiency, the Z selection criteria require that at least one electron passes tight identification requirements while the second passes at least medium, thus two average efficiencies must be calculated. The total efficiency for the identification of a Z event can then be calculated as follows:

$$P(Z) = P(M,T) + P(T,M) - P(T,T) \tag{7.1}$$

$$= \epsilon^M \epsilon^T + \epsilon^T \epsilon^M - \epsilon^T \epsilon^T \tag{7.2}$$

$$= \epsilon^T (2\epsilon^M - \epsilon^T) \tag{7.3}$$

where $P(i,j)$ is the probability that the first electron passes quality selection i and the second passes quality selection j, ϵ^i is the efficiency for quality criterion i and the index i takes the values M for medium quality and T for tight quality.

The following section will discuss the calculation of the efficiency map using Monte Carlo in conjunction with data-driven scale factors. Section 7.3 will then discuss two methods for the evaluation of the signal population distributions. Firstly, in section 7.3.1, the population distributions will be taken directly from Monte Carlo, then in section 7.3.2, a method will be described for calculating the population distributions using data. In both cases the final efficiencies and the correction to R_{Jets} will be determined along with a complete set of uncertainties.

7.2 Efficiency Maps

7.2.1 Data - MC Scale Factors

It has been observed that Monte Carlo simulations do not adequately reproduce the distributions of all of the cluster variables on which the electron identification selections are applied. Consequently, while the efficiency maps for the different quality criteria and the different event types could be constructed purely from Monte Carlo simulation, it is preferable once again to obtain data-driven estimates of thee values. Conversely, if estimates of the identification efficiency are taken directly from data, there is the possibility of bias if the final state, or the selections used to identify it, is different from that to which the efficiencies will be applied. For these reasons *scale factors* have been derived which allow the correction of Monte Carlo based estimates of the efficiencies in a particular final state of interest for the observed differences between data and Monte Carlo. For example, the efficiency map in some final state configuration denoted *2*, such as the final state $W \rightarrow e\nu + 1$ jet, can be recovered from the efficiencies calculated in data from the final state *1*, which could for instance be the final state $Z \rightarrow ee$, thus:

$$\epsilon^2(\eta, E_T) = \frac{\epsilon^2_{MC}(\eta, E_T)}{\epsilon^1_{MC}(\eta, E_T)} \epsilon^1_{Data}(\eta, E_T) = \epsilon^2_{MC}(\eta, E_T) \frac{\epsilon^1_{Data}(\eta, E_T)}{\epsilon^1_{MC}(\eta, E_T)} \tag{7.4}$$

$$= \epsilon^2_{MC}(\eta, E_T) \cdot SF(\eta, E_T). \tag{7.5}$$

Due to the limited amount of data collected in 2010 it was not possible to construct a fully 2-dimensional map of the scale factors and they were instead assumed to factorize in terms of the electron η and E_T. Consequently a set of η dependent scale factors was obtained from events integrated over the electron E_T as well as a set of E_T dependent corrections to be applied to these scale factors, obtained from events integrated over η.

7.2.2 Tag and Probe Efficiency Measurements

The data driven estimates of the electron identification efficiencies required for the construction of the scale factors described above were obtained through the use of tag and probe techniques. The tag and probe method involves identifying an unbiased sample of electrons, known as *probe* electrons, by applying selection cuts, known as *tag* requirements, to other objects or uncorrelated quantities in the event. Tag and probe techniques can be applied to both W and Z events although the tag requirements are different between the two types of event.

In the case of Z events, the tag selections typically involve requiring the existence of an electron within a specified kinematic region, which passes medium or tight identification criteria. A probe electron is then searched for, which again passes the required kinematic selections, and is of opposite sign to the tag electron. The pair are also required to have an invariant mass which falls within an appropriately chosen Z mass window. In the case of the W events, the tag requirements are usually imposed on the missing transverse momentum. They may require, for instance, that the $E_{\mathrm{T}}^{\mathrm{miss}}$ be greater than 25 GeV and that the transverse mass constructed using the probe electron be greater than 40 GeV. Additional background rejection in this case can be achieved by requiring $E_{\mathrm{T}}^{\mathrm{miss}}$ isolation, i.e. that the azimuthal angle between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and any jets in the event be greater than some threshold.

Once samples of probe electrons have been identified, it is necessary to perform independent background subtraction both before and after the application of the criteria one wishes to calculate the efficiency of. Once this has been performed, the efficiency of the selection is calculated straightforwardly as the fraction of probe electrons which pass the selection. Tag and probe studies have been performed centrally by the ATLAS collaboration for both W and Z events, and a set of scale factors and E_T corrections have been calculated giving the ratio of the data-driven efficiencies to those from Monte Carlo, obtained as a weighted average of the W and Z results [59]. In the case of the W events, the background subtraction was performed using sideband fits to the electron isolation distribution, while for the Z events, same-sign sideband fits and fits to the invariant mass distributions were used to

calculate the fraction of background events present at each level. The central values for each separate event type were obtained as the mean of a distribution constructed from around 100 variations of the tag and probe procedure in each case. Variations performed included altering the tag requirements in order to change the level of background, varying the form of the fit functions used in performing the background subtractions and varying the integration and sideband regions used when performing the fits, among others. The systematic uncertainty on the scale factors was taken as the root mean-squared width of the obtained distribution.

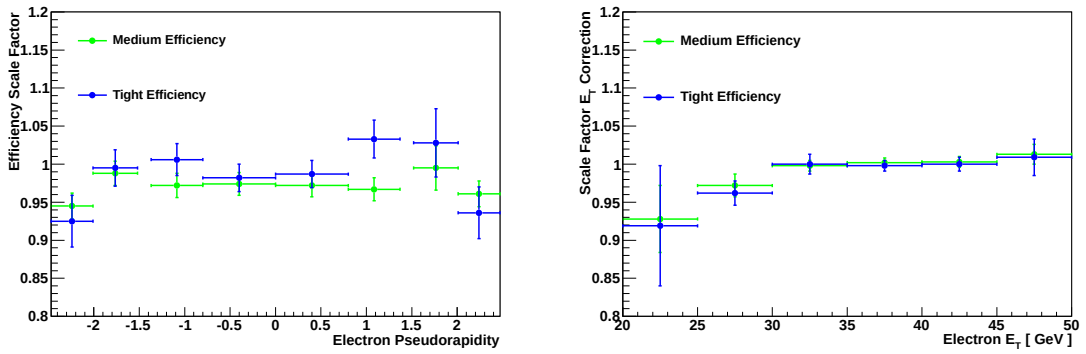


Figure 7.2: The data-MC efficiency scale factors as a function of the electron pseudorapidity (left) and their E_T corrections (right) shown for both medium and tight quality criteria.

Another benefit of the use of scale factors over raw efficiencies is that it allows an easier combination of the results of the W and Z studies. The raw efficiencies, when integrated over η and E_T separately, are different between these two processes due to different population distributions, however the scale factors themselves show closer agreement. The scale factors and the E_T corrections are shown in figure 7.2. Due to the lack of available statistics the binning is quite coarse.

7.2.3 Application to R_{Jets}

In the case of the R_{Jets} measurement, it has been demonstrated above that there is minimal dependence of the efficiency map on the amount of jet activity in the event, and consequently inclusive efficiency maps will be constructed from MC and scaled as described above. If the same binning were used as was used in the construction of the scale factors this would simply

recover the data-driven efficiencies, however the larger statistics present in Monte Carlo allow for the use of a more finely binned efficiency map.

To construct the efficiency maps from Monte Carlo, a ΔR matching of 0.3 is used between the truth daughters of the vector bosons and any reconstructed clusters. The quality criteria are then applied to any matched objects. The resulting efficiency maps for the tight quality requirements in W and Z events, as well as the medium quality requirements in Z events are shown in figure 7.3 and their projections along the electron η and E_T axes are shown in figure 7.4. The efficiency maps are calculated using Pythia samples which include pile-up modelling as described in section 4.4 as these were the samples the scale factors were calculated with respect to.

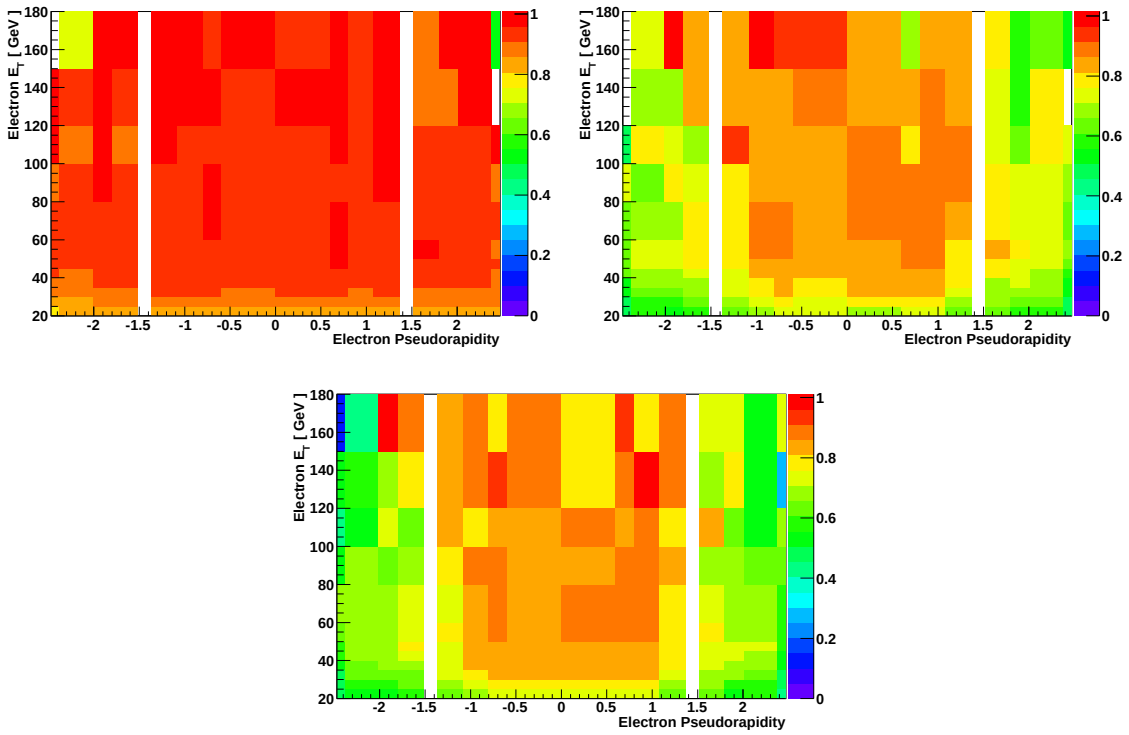


Figure 7.3: Efficiency maps calculated from Monte Carlo with the data-MC scale factors and E_T dependent corrections applied. The top left and right plots show the efficiency maps for medium and tight quality respectively in Z events while the lower plot shows the efficiency map for tight quality in W events.

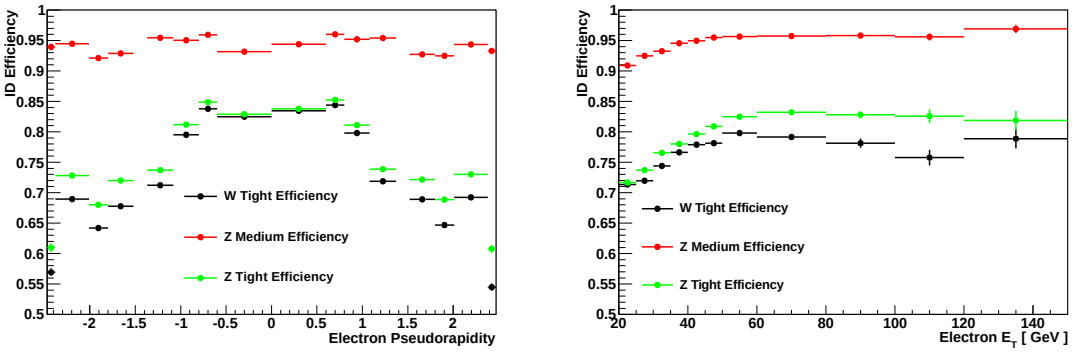


Figure 7.4: Projections of the efficiency maps along the electron η and E_T directions respectively. The slight differences between the tight efficiencies for W and Z events are the result of different population distributions between the two types of event.

7.3 Population Distributions

7.3.1 Population Distributions from Monte Carlo

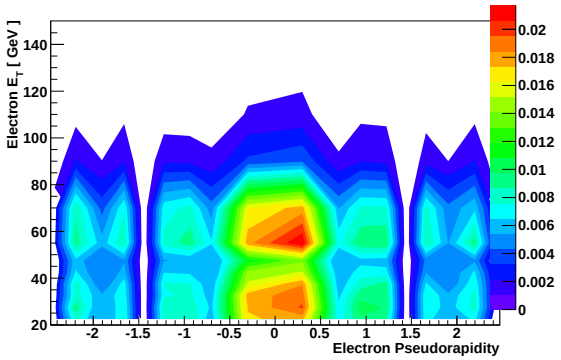
The simplest way to calculate the final efficiencies is to take the signal population distributions prior to the application of any quality selections directly from Monte Carlo. The efficiency for a given selection is then obtained as a weighted average of the efficiencies in the different electron $\eta - E_T$ bins:

$$\epsilon = \frac{\Sigma_{i,j}\epsilon_{i,j}N^{sig}_{i,j}}{\Sigma_{i,j}N^{sig}_{i,j}} = \Sigma_{i,j}\epsilon_{i,j}P^{sig}_{i,j} \tag{7.6}$$

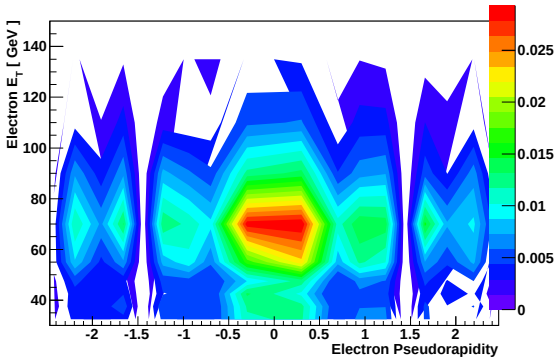
where the sum is over all bins in the efficiency map, $\epsilon_{i.j}$ is the efficiency in bin $\{i,j\}$, $N^{sig}_{i,j}$ is the number of signal events in bin $\{i.j\}$ prior to the application of the identification criteria and $P^{sig}_{i.j}$ is the population distribution obtained by normalizing $N^{sig}_{i,j}$.

In order to calculate the distributions $N^{sig}_{i,j}$ for each type of event, the same matching procedure was used as described above to identify reconstructed electrons which are daughters of the vector bosons. Subsequently all event selections were applied, except for the electron identification criteria, before forming 2-dimensional distributions of the electron pseudorapidity and transverse energy, which were then normalized to obtain the population probability distributions. In the case of the Z events, this included the opposite sign and

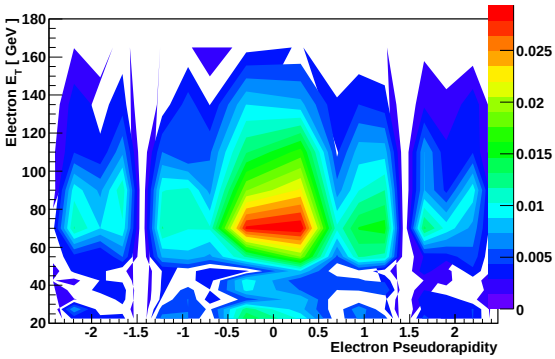
invariant mass requirements, while in the case of the W events, the transverse momentum requirement was imposed and the event was also required to satisfy the transverse mass criterion using the matched electron. Jets in the event were similarly required to pass the isolation requirement with the matched electrons after overlap removal. The resulting population distributions for three different jet P_T thresholds in W events are shown in figure 7.5. There is a clear migration of the distributions to higher values of the transverse energy as the jet threshold is increased. The efficiencies calculated using these population distributions are shown in figure 7.6.



(a) 30 GeV



(b) 60 GeV



(c) 90 GeV

Figure 7.5: Population distributions for W events with jets above different P_T thresholds.

A variety of systematic uncertainties were calculated on the final efficiencies and on the efficiency correction to R_{Jets}, the first of which stems from the uncertainties on the scale factors and the E_T dependent corrections. These uncertainties are themselves a combination of both statistical and systematic components, but given the degree of correlation between the uncertainties in different $\eta - E_T$ bins is unknown, the most conservative case is considered

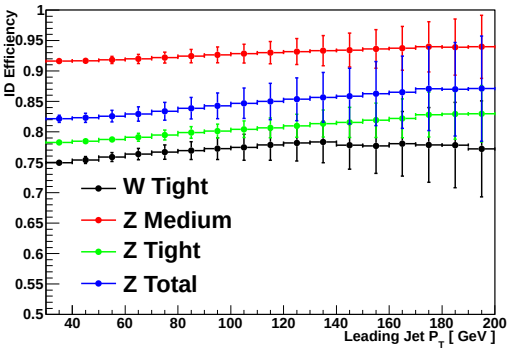
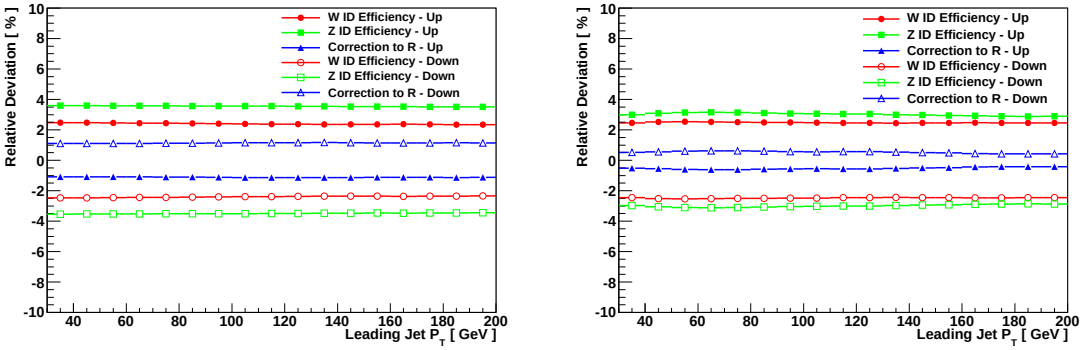


Figure 7.6: The different identification efficiencies calculated by convoluting the Monte Carlo derived population distributions with the scaled efficiency maps as per equation 7.6.

where the uncertainties on the scale factors and E_T corrections are assumed fully correlated between bins. The uncertainties on the efficiencies were therefore calculated by shifting the scale factors and E_T corrections up and down according to their uncertainty, with each bin being shifted in tandem. The efficiencies and the efficiency correction to R_{Jets} were then recalculated. Figure 7.7 shows the relative deviations on the total W and Z efficiencies and on the R_{Jets} correction resulting from this procedure.



(a) Scale Factors (b) E_T Corrections

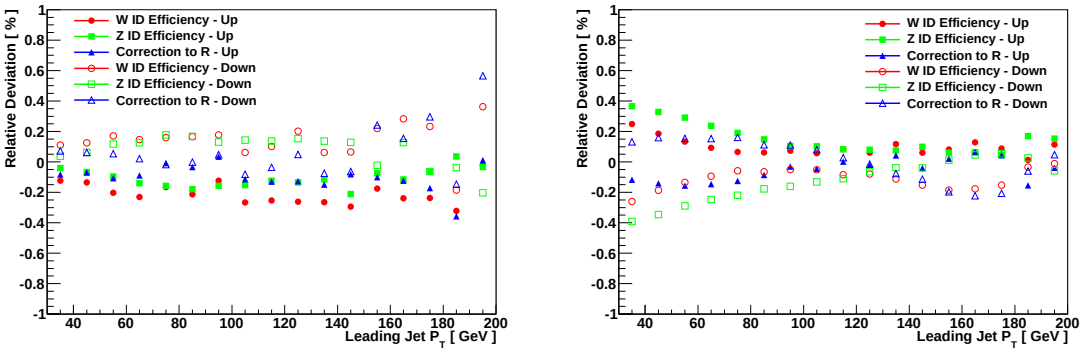
Figure 7.7: The effect on the efficiencies caused by shifting the scale factors and E_T corrections up and down according to their total uncertainties.

The effect of this variation shows very little dependence on the jet transverse momentum threshold. The W and Z efficiencies show variations between 2% and 4% but there is some partial cancellation of the variation on the correction to R_{Jets} due to the requirement of the presence of a tight electron in both types of event. The cancellation is incomplete due to the requirement of the additional electron in Z events which must pass medium quality criteria.

As a result, the variation of the Z efficiency is stronger in both cases. The variation of the correction to R_{Jets} is consequently in the opposite direction to the changes in the separate efficiencies, and is of the order of 1% for all jet transverse momenta considered.

Further to this, uncertainties on the efficiencies due to the effects of the energy scale uncertainties of the leptons and the jets in the events on the population distributions were considered. While the lepton energy scale uncertainty directly affects the population distributions of the events, the correlation between the electron transverse energies and the jet transverse momenta means that shifts in the transverse momenta of the jets in the event, and the subsequent bin migration caused, may also result in changes in the electron population distributions for events in a particular jet P_T bin.

In a similar fashion to how the scale factor uncertainties were evaluated, the leptons and jets in the events were shifted up and down according to their energy scale uncertainties and the population distributions in each jet P_T bin, and subsequently the efficiencies, were recalculated. For the electrons a fixed energy scale uncertainty of 3% was used [59]. The relative changes of the efficiencies and the correction to R_{Jets} are shown in figure 7.8 for the two cases. In both cases the resulting uncertainties are below 0.5% for all jet transverse momenta considered.

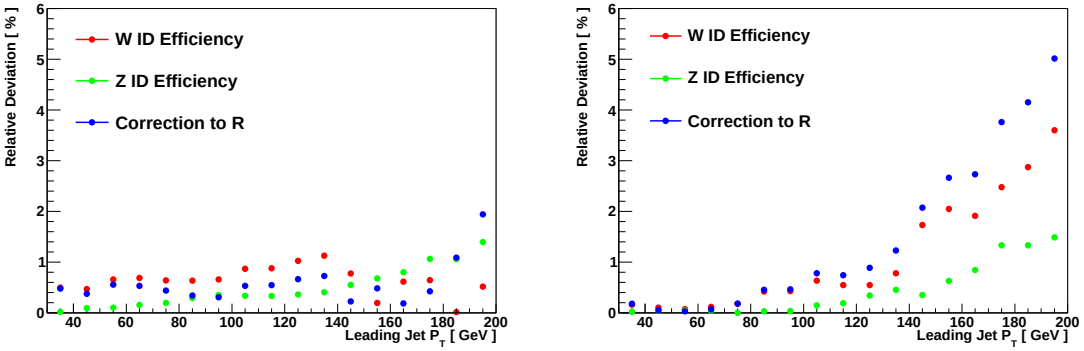


(a) Jet Energy Scale (b) Electron Energy Scale

Figure 7.8: The effect on the efficiencies caused by shifting the jet and lepton energy scales up and down according to their total uncertainties.

The last uncertainties considered on the construction of the population distributions using MC are related to the generator and pile-up modelling. Comparisons were made between

predictions for the efficiencies obtained using Pythia and ALPGEN, as well as between Pythia samples with and without pile-up modelling. Note that in doing this, only the population distributions were changed. The efficiency map used was taken from the Pythia samples with pile-up in all cases, as this was the sample used in the derivation of the scale factors. If a different generator or pile-up configuration had been used, the scale factors would change accordingly to accommodate this difference. The resulting deviations between the efficiencies are shown in figure 7.9.



(a) ALPGEN vs. Pythia.

(b) Pile-up vs. No Pile-up.

Figure 7.9: The uncertainties on the efficiencies resulting from the choice of generator and the inclusion of pile-up modelling.

Figure 7.10 shows the final efficiencies for the identification of W and Z events as well as the efficiency correction for R_{Jets}, complete with all uncertainties, evaluated using population distributions obtained from Monte Carlo. The figure also shows the breakdown of uncertainties on the correction to R_{Jets}. The dominant systematic on the efficiencies can be seen from the above to be due to the uncertainties on the scale factors, although as noted, these were treated in a particularly conservative way. The correction to R_{Jets} benefits from significant cancellation of this uncertainty although it still remains the largest contribution to the total uncertainty below 80 GeV, after which the statistical uncertainty becomes dominant. The pile-up uncertainty also increases significantly at higher jet transverse momenta, although this is likely to also be a statistical effect on the MC samples used.

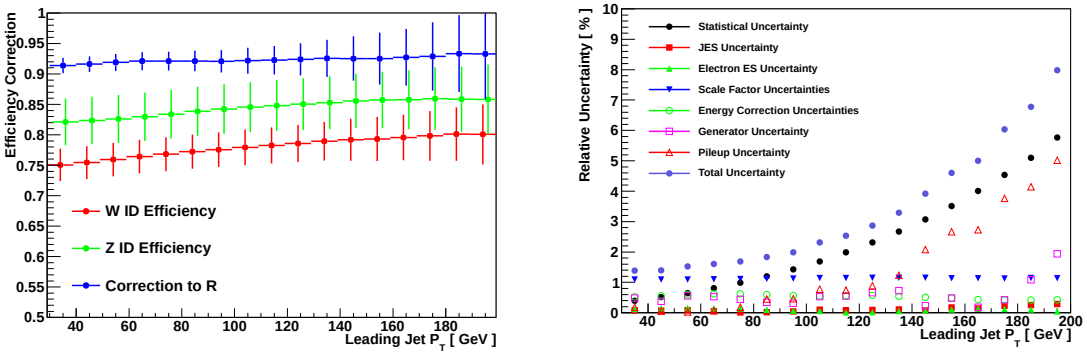


Figure 7.10: The final efficiencies and the correction to R_{Jets} including all uncertainties as well as the breakdown of the total uncertainty on R_{Jets} into its separate contributions.

7.3.2 Population Distributions from Data

The signal population distributions for the two types of event can also be obtained directly from data. This is achieved by using the efficiency maps to correct the numbers of observed signal events which pass the quality selections in each electron $\eta - E_T$ bin back to the total number that were produced in that region of phase space. The result of this is to change the form of the equation relating the total efficiency to the population distribution and the efficiency map:

$$\epsilon = \frac{\Sigma_{i,j}\epsilon_{i,j}N^{sig}_{i,j}}{\Sigma_{i,j}N^{sig}_{i,j}} = \frac{\Sigma_{i,j}\epsilon_{i,j}\frac{N^{obs}_{i,j}}{\epsilon_{i,j}}}{\Sigma_{i,j}\frac{N^{obs}_{i,j}}{\epsilon_{i,j}}} = \frac{\Sigma_{i,j}N^{obs}_{i,j}}{\Sigma_{i,j}\frac{N^{obs}_{i,j}}{\epsilon_{i,j}}} \tag{7.7}$$

$$\Rightarrow \frac{1}{\epsilon} = \frac{\Sigma_{i,j}\frac{N^{obs}_{i,j}}{\epsilon_{i,j}}}{\Sigma_{i,j}N^{obs}_{i,j}} = \Sigma_{i,j}\frac{P^{obs}_{i,j}}{\epsilon_{i,j}} \tag{7.8}$$

where $N^{sig}_{i,j}$ and $N^{obs}_{i,j}$ refer to the distributions of signal electrons prior to and after the application of the identification cuts respectively and all other symbols have the same meanings as in equation 7.6.

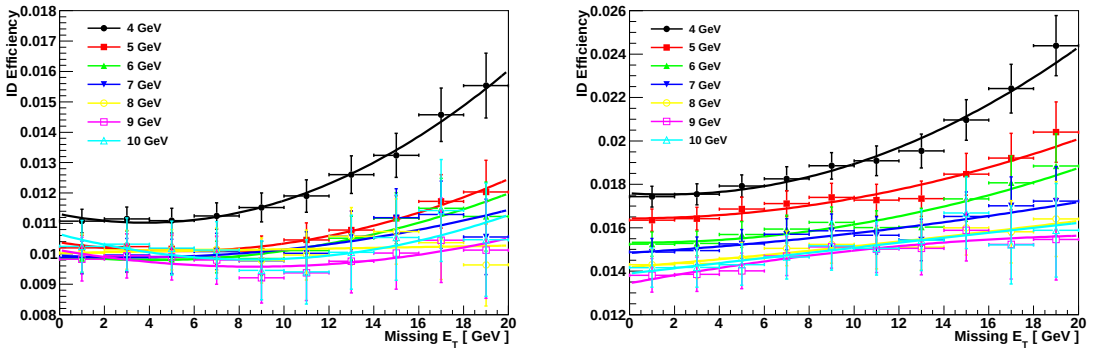
In order to perform the above summation, it is in principle necessary to first subtract background from the distribution of events observed in data to obtain the distribution of signal events, $N^{obs}_{i,j}$. Given the background fraction in Z events is a mere 0.4%, the effect

of performing the background subtraction is negligible, however for the W events where the background fractions are much higher, this is not the case and a background subtraction is required. Given the QCD background fractions f_{qcd}^{W} have already been calculated for each jet P_T bin in chapter 6, it is only necessary to know the shapes of the QCD background distributions in electron $\eta - E_T$ space for each jet P_T bin. These shapes can be determined using events which pass the W kinematic and geometrical selections but which contain only a loosely identified electron. Such a sample will be dominated by QCD background and consequently, if the identification efficiency of the medium and tight selection criteria relative to loose are known for fake electrons, the distribution of QCD background events in the signal region can be constructed by weighting each such event by the probability that it would additionally satisfy the quality criteria. The W QCD background distribution, which requires one tight electron and no second medium electron, is therefore constructed as follows:

$$N_{\{i,j\}}^{QCD}(W) \propto \Sigma_{E_W} \Sigma_{k\epsilon\{i,j\}} [\epsilon_k^T \Pi_{l\neq k}(1-\epsilon_l^M)] \tag{7.9}$$

In the above equation, the first summation is over the set E_W of all events which contain at least one loosely identified electron and pass the relevant kinematic and geometrical event selections for W events. The second summation is over all loose electrons k which are in the electron $\eta - E_T$ bin of interest, denoted here by $\{i, j\}$. Note that for a particular loose electron to be considered in this summation, the electron in question must satisfy the transverse mass requirement with the $E_{\mathrm{T}}^{\mathrm{miss}}$. ϵ_k^T is the tight identification efficiency for a *fake* electron k in bin $\{i, j\}$ which has passed the loose identification cuts. Similarly ϵ_l^M is the medium identification efficiency for a fake electron $l \neq k$ which has also passed loose identification selections. The product is taken over all loose electrons in the event other than k. Each term in the sum contributes a weight to the contents of the background population distribution in bin $\{i, j\}$ given by the probability that the fake electron k passes the tight identification criteria and all other loose electrons l in the event fail the medium identification criteria, as per the W event selections.

To calculate the efficiencies of the medium and tight identification requirements for fake electrons which pass loose identification criteria, the majority of the standard W event selections were relaxed. The events used were still required to pass the usual Good Run List and have a primary vertex, as well as having at least one reconstructed electron in the phase space of interest passing loose identification criteria. These electrons were still required to be separated by greater than 0.6 in ΔR from any jets after overlap removal, as this selection can change the resulting efficiencies. At the level of loose identification this sample is dominated by QCD background. Once one applies tight identification cuts however, the sample becomes dominated by genuine W events. Consequently, to achieve sufficient purity of background, the electrons used to determine the background identification efficiencies were also required to have an isolation energy ($etcone20$) greater than a particular value, chosen to be 8 GeV for reasons discussed later. Due to limited statistics, the background efficiencies were derived in the more coarse bins used for the derivation of the scale factors.



(a) $(-2.47 < \eta < -2.01, 25\text{GeV} < E_T < 30\text{GeV})$ (b) $(-1.37 < \eta < -0.8, 25\text{GeV} < E_T < 30\text{GeV})$

Figure 7.11: The variation of the QCD tight identification efficiency in two different electron $\eta - E_T$ bins when a variety of different lower bounds on the isolation energy are used. The fits to the data are also shown.

While one could simply use the efficiencies determined directly from this sample of fake electrons, some genuine dependence of the identification efficiencies on the missing transverse momentum is expected, and this should be accounted for. Consequently, the identification efficiencies were calculated as a function of the missing transverse momentum threshold in each electron $\eta - E_T$ bin up to 20 GeV. A quadratic fit was then made to the data and the extrapolated value at 25 GeV was taken as the efficiency for that bin. The results

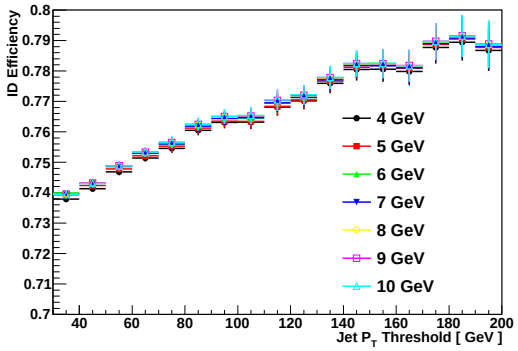
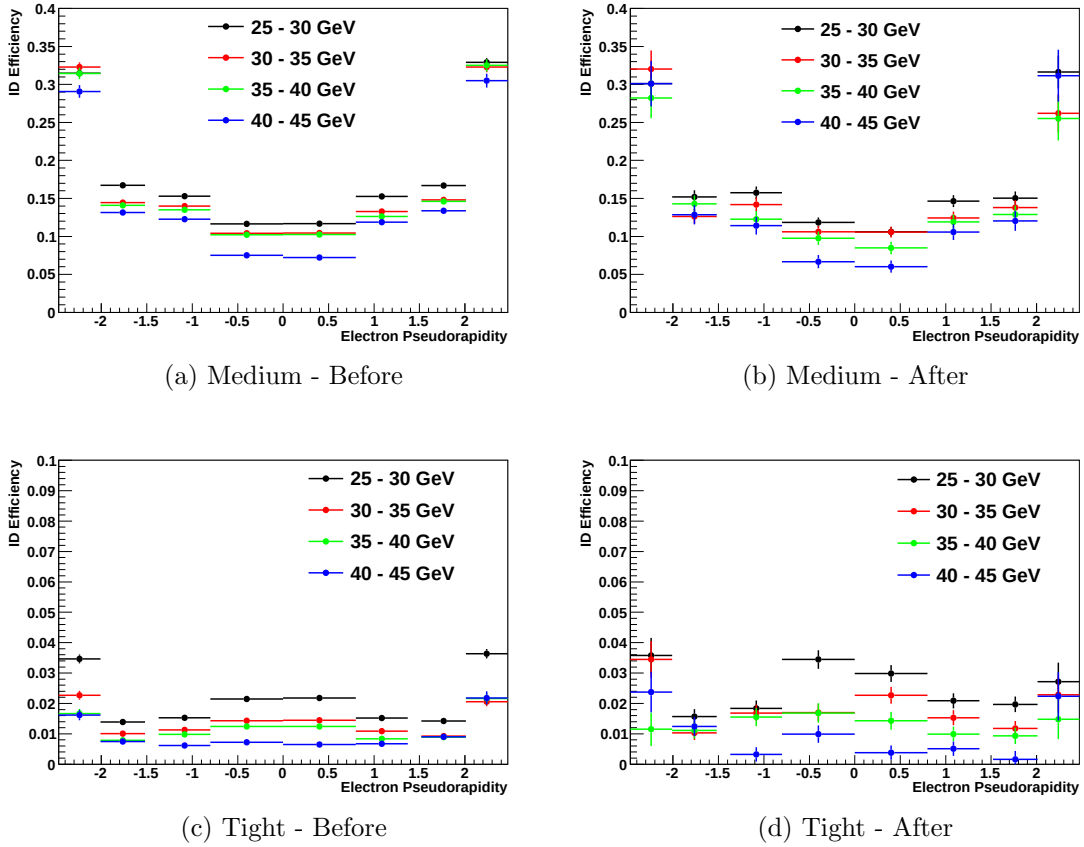


Figure 7.12: The variation of the extracted W signal efficiency due to the isolation cut chosen.

of this procedure are shown in figure 7.11 for two different bins; ($-2.47 < \eta < -2.01, 25$ GeV$< E_T <$ 30 GeV) and ($-1.37 < \eta < -0.8, 25$ GeV$< E_T <$ 30 GeV), for a variety of different choices of the minimal value of the isolation. The dependence for an isolation lower bound of 4 GeV is markedly stronger in both cases than for the other values. This was interpreted as being due to an insufficient purity of QCD background in the sample. The variation of the fit for isolation values between 6 and 10 GeV is considerably weaker and this motivated the choice of the isolation cut being placed at 8 GeV. The envelope of the efficiencies derived from the isolation values between 6 and 10 GeV inclusive were used to assign a systematic uncertainty on the use of the isolation cut. As can be seen in figure 7.12, which shows the dependence of the extracted identification efficiency on this cut, the uncertainty is extremely small. The extracted QCD identification efficiencies for medium and tight quality selections with respect to loose are shown in figure 7.13 as a function of the electron pseudorapidity both prior to and after the extrapolation procedure has been applied.

The W QCD background population distributions constructed using these efficiencies are shown in figure 7.14 for three different jet transverse momentum thresholds. These distributions were then appropriately normalized using the number of data events observed passing the signal selections and the calculated background fractions. Following this they were subtracted from the distributions of observed data events to obtain estimates of the number of electroweak events passing the identification criteria. A further Monte Carlo



(a) Medium - Before

(b) Medium - After

(c) Tight - Before

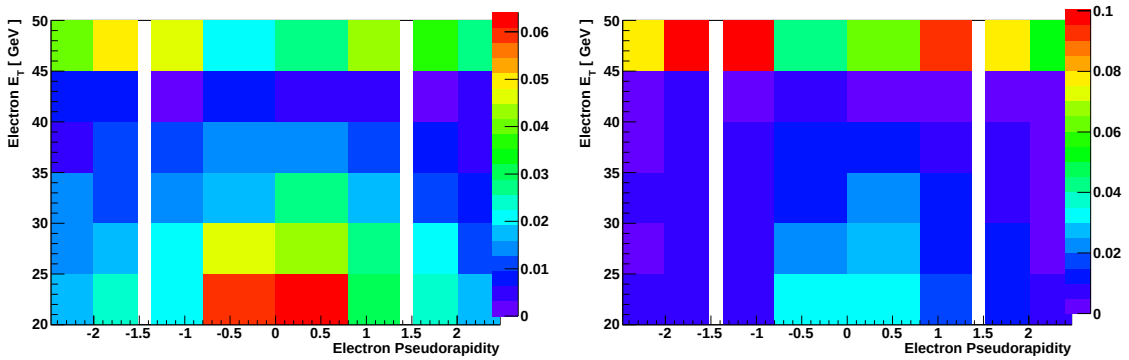
(d) Tight - After

Figure 7.13: The medium and tight identification efficiencies for the QCD background to W events as a function of the fake electron η. The left hand plots show the values extracted from the full sample of fake electrons. The right hand plots show the results of fitting the efficiencies as a function of the $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold and extrapolating to 25 GeV as described in the main text.

based correction is then applied to remove the contribution from electroweak background to the distributions. This is simply a ratio of the population distributions obtained using solely $We\nu$ MC to those obtained with the inclusion of the dominant electroweak backgrounds. This correction is extremely close to 1 in all bins. The resulting population distributions were then normalized and combined with the efficiency maps to calculate the overall event efficiencies as described in equation 7.8.

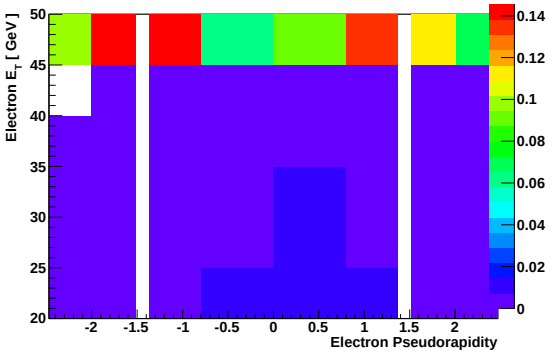
The final efficiency corrections obtained using this method including uncertainties from the fit to the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution, the isolation uncertainty and the uncertainties on the calculated background fractions are shown in figure 7.15. Uncertainties were also included due to the uncertainties on the scale factors and E_T dependent corrections in the same way as

in the previous section. The data-driven values agree extremely well with the Monte Carlo predictions for all jet P_T thresholds.



(a) 30 GeV

(b) 60 GeV



(c) 90 GeV

Figure 7.14: The background population distributions in W events for three different jet transverse momentum thresholds.

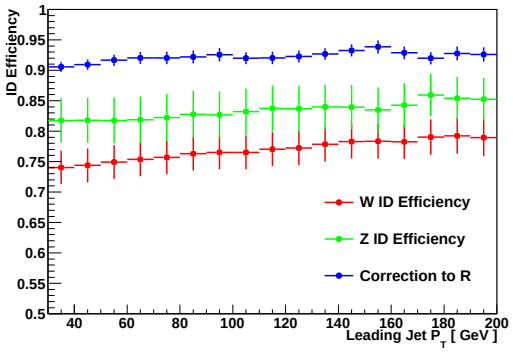


Figure 7.15: The data-driven signal efficiency corrections obtained using the method described in the main text.

A cross check of the data-driven method presented above was made for the case of the W events, by inverting some of the tight identification cuts on the electron in order to define

a QCD dominated control region from which to extract the QCD population distributions. The same five cuts were considered as in the $E_{\mathrm{T}}^{\mathrm{miss}}$ based QCD background subtraction procedure described in section 6.3.2 and again events were included in the construction of the QCD population distribution if any two of the five cuts is failed. The electron in the event was also required to have etcone20 $>$ 8GeV as in the method above to achieve the necessary signal rejection. A systematic uncertainty was also calculated in the same way as for the $E_{\mathrm{T}}^{\mathrm{miss}}$ based background subtraction procedure, by considering the variation on the extracted efficiency when subsets of the possible doublet inversions are considered. The same subsets were considered as in section 6.3.2, and the largest deviation of the efficiency calculated using one of these subsets from the nominal efficiency was considered as the uncertainty and added in quadrature with the statistical and background fraction uncertainties. The nominal efficiency calculated using this method, as well as the efficiencies calculated using the restricted sets of pairs of cut inversions, are shown on the left hand side of figure 7.16. At low jet transverse momenta, it is clear that the systematic uncertainty on the construction of the QCD population distribution is the dominant uncertainty on this method. A comparison of the the extracted efficiencies between the first method, based on calculation of the background identification efficiencies (method 1), and the method based on cut inversions (method 2), is shown on the right hand side of figure 7.16[1]. The results of the two methods are consistent over the entire range of jet transverse momentum thresholds, with the agreement improving as the jet P_T threshold increases. Method 1 will be used to provide the efficiency correction for the final R_{Jets} measurement as its overall uncertainty at low jet P_T is considerably smaller.

7.4 Trigger Efficiencies

The tag and probe method is also used on Z events to determine the trigger efficiencies for the two triggers used in this analysis: L1_EM14 and em_e15_medium, with respect to

[1]Note that the scale factor uncertainties have not been included here as they would be completely correlated between the two methods.

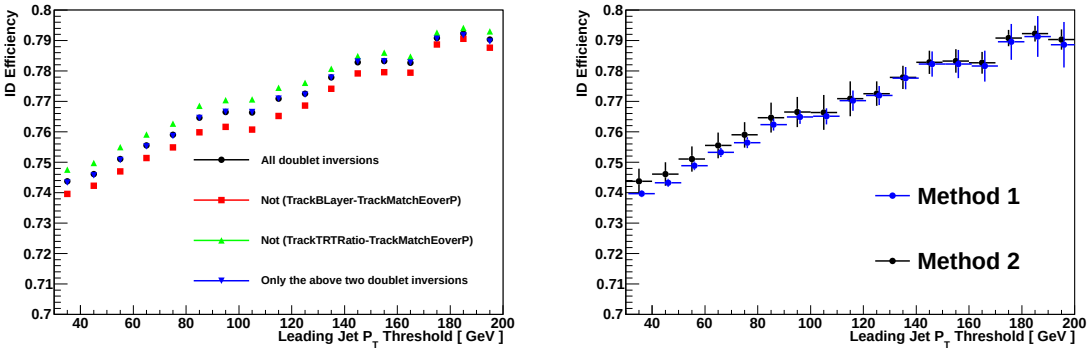


Figure 7.16: The left hand plot shows the extracted efficiencies using the cut inversion method described in the main text when the different restricted sets of doublet inversions listed in section 6.3.2 are used. The right hand plot compares the W signal efficiencies calculated using the two different data-driven methods. Method 1 relies on the calculation of background identification efficiencies, while method 2 relies on the inversion of tight cuts to evaluate the background population distributions.

medium and tight offline quality requirements. The tag requirements used to do this were the same as those described above for the determination of the identification efficiencies, with the additional requirement that the tag electron fire the relevant trigger in order that the probe sample be unbiased. The trigger efficiencies with respect to medium and tight offline quality are then simply determined by calculating what fraction of the probe electrons passing the given offline quality criteria also fire the trigger of interest. As determined in the previous chapter, the level of background after the application of offline quality criteria is less than 0.5% for Z events and consequently no background subtraction is required for the evaluation of the trigger efficiencies.

The overall trigger efficiency for a W event is simply given by the trigger efficiency with respect to the tight offline quality selections, while for a Z event, the requirement that there be at least one tight electron and a second at least medium leads to a total event trigger efficiency given by:

$$\epsilon_{Trigger}(Z) = 1 - (1 - \epsilon^{M}_{Trigger})(1 - \epsilon^{T}_{Trigger}) \tag{7.10}$$

The resulting efficiencies for the two types of event and for the two trigger requirements are given in table 7.1. The total efficiencies for the measurement are found by calculating

Trigger ϵ (%)	L1_EM14	EF_e15_medium
wrt medium	99.7 ± 0.0(stat)	99.0 ± 0.1(stat)
wrt tight	99.7 ± 0.1(stat)	99.3 ± 0.1(stat)
$R = W/Z$	99.7 ± 0.1(stat)	99.3 ± 0.1(stat)

Table 7.1: L1_EM14 and EF_e15_medium trigger efficiencies.

the luminosity weighted average of the two separate trigger efficiencies. The correction to R_{Jets} is then given by the ratio of the averaged W and Z event efficiencies and is found to be $99.3 \pm 0.5\%$ [11].

Chapter 8

Acceptance Corrections

8.1 Introduction

Having subtracted background contamination from the two processes and corrected for the identification efficiencies of the leptons in the event, the next correction which must be applied is the vector boson acceptance correction. This factor is applied to R_{Jets} to accommodate the differing acceptances of the kinematic and geometrical selections pertaining to the identification of the vector bosons at particle level and reconstruction level. The selections which are mapped from reconstruction level to particle level by this correction are described fully in section 4.2. The correction factors for the W and Z bosons separately and for the ratio are defined as follows using the notation defined in section 4.3:

$$C_A(W) = \frac{N(\ell_R^W, E_{T,R}^{miss}, M_{T,R}, J_{R,\ell_R^W})}{N(\ell_T^W, E_{T,T}^{miss}, M_{T,T}, J_{R,\ell_T^W})} = \frac{\frac{N(\ell_R^W, E_{T,R}^{miss}, M_{T,R}, J_{R,\ell_R^W})}{N(J_{R,\ell_T^W})}}{\frac{N(\ell_T^W, E_{T,T}^{miss}, M_{T,T}, J_{R,\ell_T^W})}{N(J_{R,\ell_T^W})}} = \frac{A_W(\ell_R^W, E_{T,R}^{miss}, M_{T,R}|J_R)}{A_W(\ell_T^W, E_{T,T}^{miss}, M_{T,T}|J_R)} \tag{8.1}$$

$$C_A(Z) = \frac{N(\ell_R^Z, M_{\ell\ell,R}, J_{R,\ell_R^Z})}{N(\ell_T^Z, M_{\ell\ell,T}, J_{R,\ell_T^Z})} = \frac{\frac{N(\ell_R^Z, M_{\ell\ell,R}, J_{R,\ell_R^Z})}{N(J_{R,\ell_T^Z})}}{\frac{N(\ell_T^Z, M_{\ell\ell,T}, J_{R,\ell_T^Z})}{N(J_{R,\ell_T^Z})}} = \frac{A_Z(\ell_R^Z, M_{\ell\ell,R}|J_R)}{A_Z(\ell_T^Z, M_{\ell\ell,T}|J_R)} \tag{8.2}$$

$$C_A(R) = \frac{C_A(W)}{C_A(Z)} = \frac{\frac{A_W(\ell_R^W, E_{T,R}^{miss}, M_{T,R}|J_R)}{A_W(\ell_T^W, E_{T,T}^{miss}, M_{T,T}|J_R)}}{\frac{A_Z(\ell_R^Z, M_{\ell\ell,R}|J_R)}{A_Z(\ell_T^Z, M_{\ell\ell,T}|J_R)}} \tag{8.3}$$

Note that expressing the upper fractions in equations 8.1 and 8.2 as acceptances is not strictly correct, as the numerators involve the presence of reconstructed jets where the overlap removal and isolation criteria are performed using reconstructed electrons (which are required to exist), whereas for the denominators, these criteria are applied using the dressed truth electrons. Consequently, the numerators are not formally subsets of the denominators and these are therefore not strictly probabilities. However with this caveat in mind, the notation above will be used for the rest of this chapter and these factors will be treated as acceptances for the purposes of calculating statistical uncertainties. So long as the denominator in each case is not close in value to the numerator, the errors in doing this will be small.

Calculating the acceptance correction requires knowledge of the number of events passing given selections at particle level and consequently the use of Monte Carlo simulation cannot be completely avoided here. In the following section the above corrections are calculated, and the dominant systematics evaluated, purely using Monte Carlo. In section 8.3 a partially data-driven approach is developed for calculating the $E_{\mathrm{T}}^{\mathrm{miss}}$ acceptance, which potentially allows for a significant reduction of the overall systematic uncertainty on the acceptance correction. In both cases, when calculating the acceptances for either type of event to pass the reconstruction level selections, the reconstructed electron clusters which pass the fiducial selections are also required to match a truth daughter of the vector boson within a cone of radius 0.3. This is to remove the effect of any fake clusters in a manner consistent with the evaluation of the identification efficiency.

8.2 Monte Carlo Studies

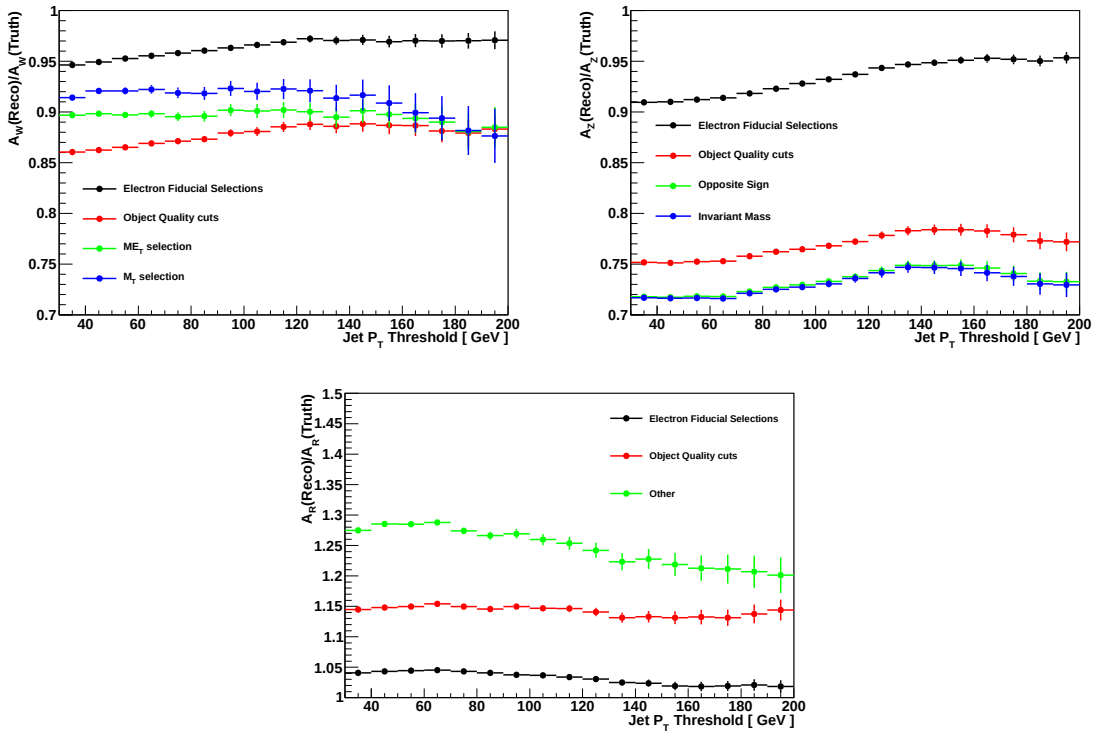


Figure 8.1: The upper left and right plots show how the W and Z acceptance corrections change as successive selections are applied. The lower plot shows the change in the correction on the ratio. For this plot 'other' indicates all other selections not currently applied and would include the $E_{\mathrm{T}}^{\mathrm{miss}}$ selection and the M_T selection in the case of the W and the invariant mass selection in the case of the Z. The uncertainties are purely statistical and are evaluated as described in the main text.

Figure 8.1 shows the acceptance corrections for the W and Z cross sections, as well as for R_{Jets}, as successive sets of selections are applied. These values were calculated using the ALPGEN samples without pile-up simulation listed in section 4.4. In both cases the electron fiducial selections, which include both the η and P_T selections, result in an acceptance correction lower than 1.0. This is primarily due to the finite resolution of the detector when reconstructing electrons. At 20 GeV, the electron P_T distribution is increasing (see the left hand plot of figure 4.2) and consequently the smearing of the distribution due to the detector resolution leads to a decrease in the number of events passing the selection with respect to particle level. The effect is higher at low P_T due to the gradient being steeper, and is larger in the case of the Z due to the presence of two electrons instead of one. The electron quality cuts also lead to a large decrease in the value of the correction as they are only applied at

reconstruction level and can only prevent events from passing. Again the effect is almost twice as large for the Z.

While the invariant mass selection has an almost negligible further effect on the Z acceptance correction, the $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_T selections tend to lead to an increase of the W acceptance correction at low jet P_T. The effect of the detector on the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution will be examined in detail in section 8.3. The net result is that the acceptance correction on the ratio is larger than one.

Statistical uncertainties on each of the separate acceptances in the above equations can be determined using the following equation, which applies to binomial efficiencies calculated using weighted events, with event i having a weight w_i [99]:

$$\delta A = \frac{\sqrt{\Sigma_+ w_i^2 (\Sigma_- w_i)^2 + \Sigma_- w_i^2 (\Sigma_+ w_i)^2}}{(\Sigma w_i)^2} \tag{8.4}$$

Where a subscript '+' indicates a sum over events which pass the given selections and a subscript '-' indicates a sum over events which failed. No subscript indicates a sum over all events.

The probabilities that an event passes a given selection at particle level and at reconstruction level will in general be quite strongly correlated, and so in order to propagate these uncertainties to the acceptance corrections, the correlations between the acceptances must be taken into account. The uncertainty on a ratio C of two quantities 'A' and 'B' which exhibit some correlation is given by the following equation[1]:

$$(\frac{\sigma_C}{C})^2 = (\frac{\sigma_A}{A})^2 + (\frac{\sigma_B}{B})^2 - 2\frac{\sigma_A \sigma_B \rho}{AB} \tag{8.5}$$

where ρ is the correlation coefficient, which is evaluated using the components of the covariance matrix. The correlation coefficients are shown in figure 8.2. The resulting statistical uncertainties are already shown on figure 8.1. The correlation coefficients take values close

[1] Derivation in appendix A

to unity when only the lepton geometrical and kinematic selections are applied due to the fact that the same η and P_T cuts are applied on the truth and reconstruction level leptons and that the detector only mildly distorts the values of these quantities. As successive cuts are applied on variables which are more markedly affected by the detector between truth and reconstruction level, the value of ρ gradually decreases.

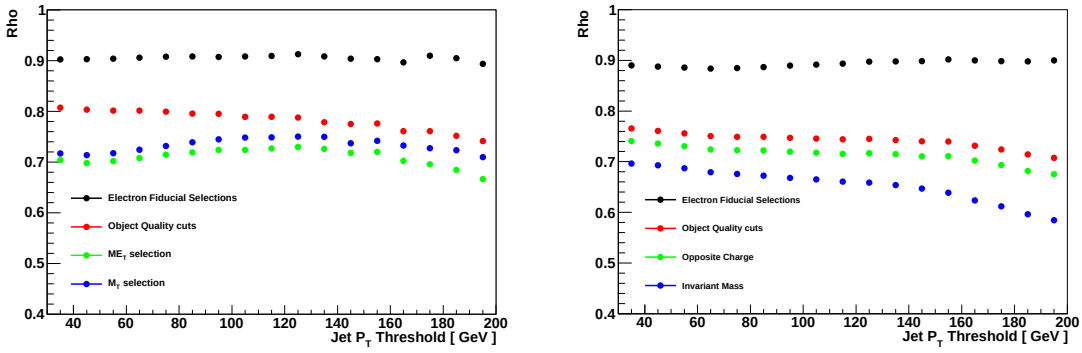
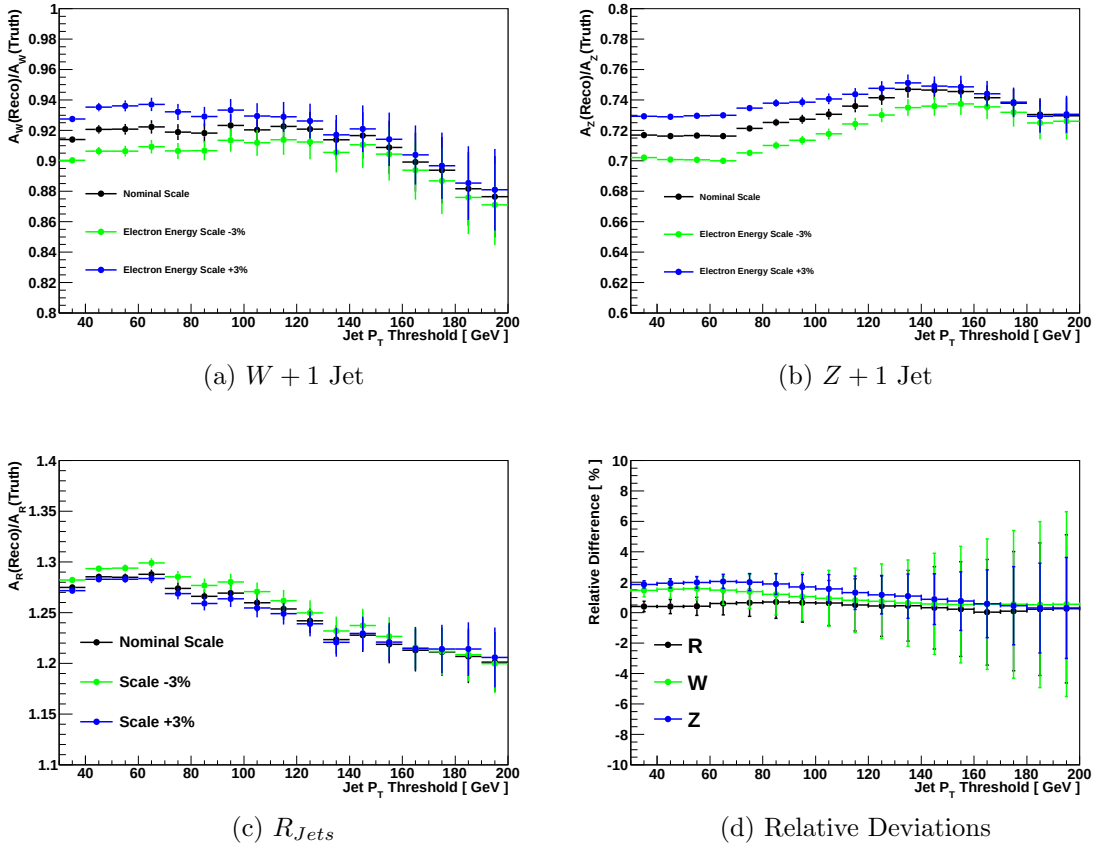


Figure 8.2: The left and right plots show the correlation coefficients obtained when different selections are applied for W and Z events respectively.

8.2.1 Systematics

The two most important systematics which must be considered with regard to the vector boson acceptance correction are the energy scale uncertainty on the electrons in the event and the uncertainty on the missing transverse momentum. To evaluate the effect of the former, all of the electrons in both sets of events were shifted in momentum up and down by 3% in tandem [59]. The effect of these shifts was then propagated to the ratio.

Figure 8.3 shows the effect of these shifts on the separate acceptance corrections for the W and Z events and the ratio, as well as the relative deviations in each case. As expected, the positive shift in the electron energy scale leads to an increase in both acceptance corrections as more events then pass the electron fiducial selections at reconstruction level. The electron energy scale uncertainty has a slightly larger effect on the Z events than the W events due to there being two electrons and consequently this uncertainty does not completely cancel on taking the ratio. The relative deviation of the Z events decreases from $\sim 2\%$ at low P_T to almost zero at high transverse momentum. The effect on the ratio is below 1% at all



(a) $W + 1$ Jet

(b) $Z + 1$ Jet

(c) R_{Jets}

(d) Relative Deviations

Figure 8.3: The deviations in the acceptance corrections caused by shifting the electron energy scale by $\pm$ 3%.

transverse momenta considered.

Uncertainties on the scale and resolution of the topoclusters used to construct the missing transverse momentum will also have an impact on the acceptance correction. In the absence of detailed knowledge of the scale and resolution of the individual topoclusters, the jet energy scale and resolution uncertainties were used to obtain reasonable variations of the missing transverse momentum on an event by event basis. The effect of these variations was then propagated to the acceptance correction on the ratio.

Firstly, the effects of uncorrelated variations of the transverse momenta of the clusters were investigated by smearing all of the jets in the event according to their jet energy resolution uncertainties. The jet energy resolutions are determined from Monte Carlo simulations by the ATLAS Jet-$E_{\mathrm{T}}^{\mathrm{miss}}$ group and the obtained values and uncertainties are validated by comparisons with data [100]. The values for the resolutions are only valid for jets within

$|\eta| < 2.8$ and with $P_T > 20$ GeV. Consequently three approaches were used to perform the smearing which are successively more conservative:

- Approach 1: Only those jets within the kinematic region of applicability were smeared.
- Approach 2: The resolution uncertainties were extrapolated outside of their region of applicability and applied to all reconstructed jets.
- Approach 3: Jets outside the kinematic region of applicability were smeared using a constant resolution of 20%.

After smearing, the vectorial difference between the smeared jets and the original jets was obtained and this difference was subtracted from the missing transverse momentum. The missing transverse momentum selection and the transverse mass selection were then reapplied and the acceptance corrections were recalculated. This procedure was repeated 1000 times for each of the above prescriptions and distributions of the deviations in the acceptance correction between the smeared and the nominal case were constructed. The widths of these distributions as well as the relative deviations between the nominal values of the acceptance corrections and the central values of the Gaussian distributions obtained from the smeared values, are shown in the upper plots of figure 8.4. Both the central shifts and the distribution widths are within 1% at all jet transverse momenta considered for all three approaches.

Correlated variations in the transverse momenta, due for example to a systematically biased calibration, were treated in a similar manner. Instead of randomly smearing each of the jets in the event, all of the jets were shifted up and down in tandem according to their energy scale uncertainty. These uncertainties are again provided by the ATLAS jet-$E_{\mathrm{T}}^{\mathrm{miss}}$ group as a function of the momentum and pseudorapidity of the jet [101]. These values are only applicable for jets with above 15 GeV of transverse momentum and $|\eta| < 4.5$. Any topoclusters not clustered into a jet within this kinematic region which were separated by $\Delta R > 0.2$ from any reconstructed electrons were also scaled in tandem with the jets according to the following function:

$$\frac{\Delta P_T}{P_T} = \pm\, a\, (1 + \frac{1.5}{P_T}) \tag{8.6}$$

where $a = 0.05$ for $|\eta| < 3.2$ and $a = 0.10$ for $|\eta| > 3.2$ and P_T is the local hadronically calibrated topocluster transverse momentum in GeV [64]. As before, the differences in momentum between the new and old objects were determined, summed vectorially and subtracted from the missing transverse momentum prior to reapplication of the selections.

The lower plot in figure 8.4 shows the percentage deviations on the acceptance correction caused from shifting the jets. This contribution to the uncertainty is dominant at low transverse momenta, causing deviations as high as 4%.

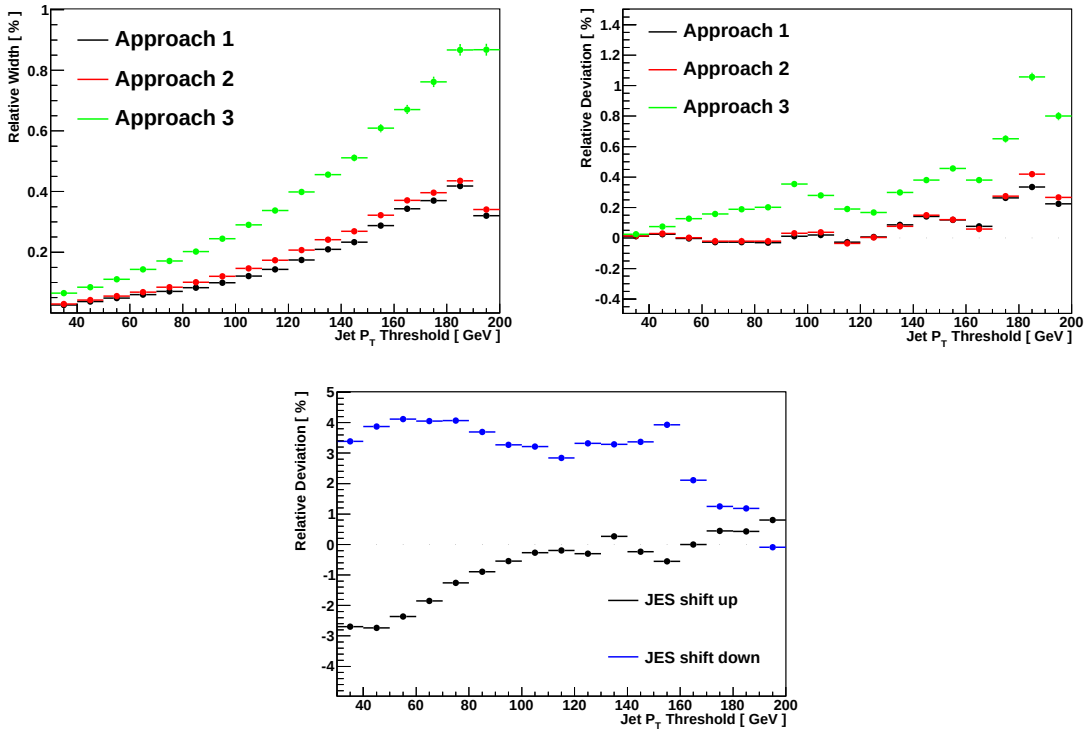
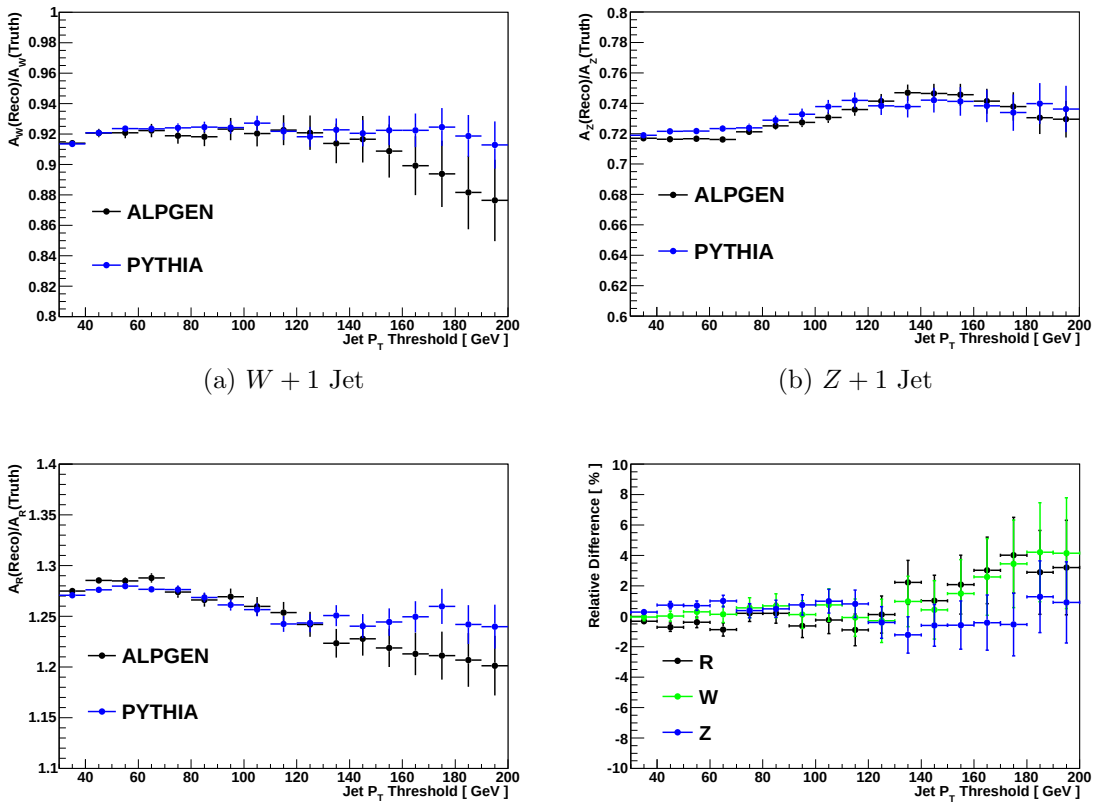


Figure 8.4: The upper left and right plots show the widths and mean values of the distributions of deviations in the acceptance correction obtained when smearing the jets according to the three prescriptions outlined in the main text. The lower plot shows the deviations obtained when shifting the jets in the event up and down according to their jet energy scale uncertainty.

Generator and pile-up uncertainties are not applied independently on this correction. These uncertainties will instead be calculated for the combined jet and acceptance correction. This will be discussed in chapter 9. Only applying this correction once avoids the unnecessary

accumulation of statistical fluctuations which would result from multiple separate inclusions of these uncertainties. For completeness, the results obtained with Pythia and ALPGEN are shown in figure 8.5 and the results from using ALPGEN samples with and without pile-up simulation are shown in figure 8.6.



(a) $W + 1$ Jet

(b) $Z + 1$ Jet

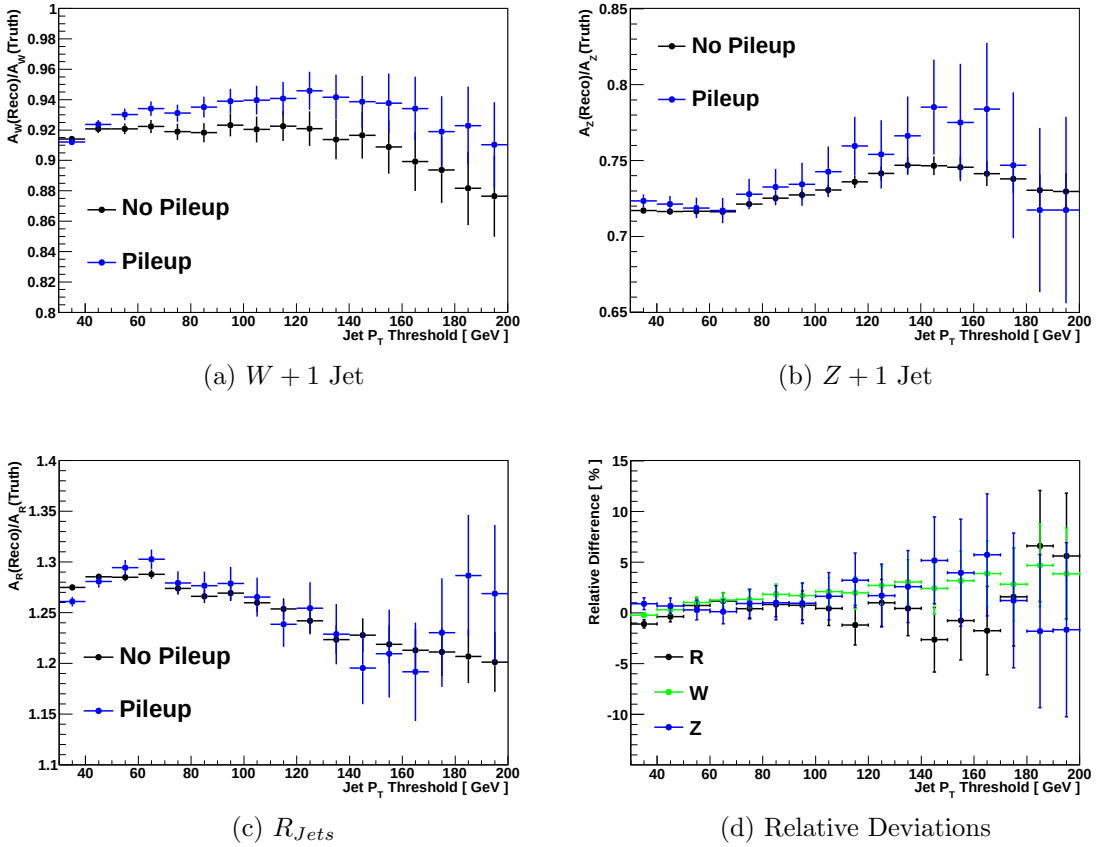
(c) R_{Jets}

(d) Relative Deviations

Figure 8.5: The upper left and right plots show the acceptance corrections separately for the W and Z cross sections respectively when calculated using ALPGEN and Pythia. The lower left plot shows the corrections on the ratio and the right hand plot shows the relative deviations between the two results.

8.3 Data-Driven Decomposition Method

The dominant source of systematic uncertainty on the acceptance correction to R_{jets} can be seen from the previous section to be due to uncertainties on the missing transverse momentum in the event. In light of this, a more sophisticated data-driven approach to calculate the $E_{\mathrm{T}}^{\mathrm{miss}}$ acceptance has been developed with the aim of enabling a more refined



(a) $W + 1$ Jet

(b) $Z + 1$ Jet

(c) R_{Jets}

(d) Relative Deviations

Figure 8.6: The upper left and right plots compare the corrections obtained for the W and Z cross sections respectively when using ALPGEN samples with and without pile-up simulation. The lower left and right plots show the ratio and the relative deviations between the corrections.

treatment of systematics leading to reduced uncertainties [102].

The data-driven method presented here is based on the assumption that while the decay of the W and Z bosons may differ due to the boson masses, the chiral structure of the couplings and the decay products, the behaviour of the hadronic recoil for a given event topology (as characterised by the boson transverse momentum and the total event ΣE_T) is similar between the two types of event. If this assumption holds, one would similarly expect the fake component of the missing transverse energy, defined here as the component of the $E_{\mathrm{T}}^{\mathrm{miss}}$ not due to the daughter neutrino of a vector boson, to also be similar between the two types of event. It is then possible to extract information regarding the fake component of the missing transverse energy from Z data, where the $E_{\mathrm{T}}^{\mathrm{miss}}$ is entirely fake, and apply this information to truth level W Monte Carlo in order to model the reconstruction level $E_{\mathrm{T}}^{\mathrm{miss}}$

with minimal explicit reference to detector simulation.

In reality the situation is complicated by the decay products of the vector bosons and to a lesser extent by flavour differences in the hadronic activity between the two types of event. Both electrons and muons will deposit clusters of energy in the calorimeters and these clusters make a contribution to the value of the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$. The presence of the leptons therefore affects the bias and resolution of the fake component of the $E_{\mathrm{T}}^{\mathrm{miss}}$, the extent of which is dependent on the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithm used.

Unlike previous chapters, results will be shown here in both the muon and electron channels and also for the inclusive as well as the one jet case. This is because these additional results aid significantly in the interpretation of the performance of this method when used with the different $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithms discussed in section 3.3.3. Also note that for the main body of this study slightly different selections were used to the ones documented in section 4.2 although this should not have any significant impact on the main conclusions. The selections used are documented in appendix C.

8.3.1 Parameterisation of Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ in W and Z Events

Conservation of momentum ensures that prior to detector effects being folded in, the transverse momentum of the boson is equal and opposite to the transverse momentum of the total hadronic recoil in the event. This axisymmetry present at truth level makes it natural to consider a coordinate system for describing the $E_{\mathrm{T}}^{\mathrm{miss}}$ which is aligned with the boson transverse momentum. The $E_{\mathrm{T}}^{\mathrm{miss}}$ in the different events is therefore considered resolved along axes parallel ($\not{E}_{\parallel}$) and perpendicular ($\not{E}_{\perp}$) to the boson P_T direction.

The detector can have two types of effect on the missing transverse momentum in a W or Z event. The first is due to some random source, for example hadrons which are lost in cracks or faked by instrumental noise, causing fake $E_{\mathrm{T}}^{\mathrm{miss}}$ which is uncorrelated with the vector boson direction. This introduces a smearing effect on the true $E_{\mathrm{T}}^{\mathrm{miss}}$. The second kind is due to some systematic bias in the measurement, for instance from an incorrect energy

scale. Due to the P_T cancellation between true hadrons in the perpendicular direction, the effect from a bias is only expected in the $\not\!\!E_{\parallel}$ distribution.

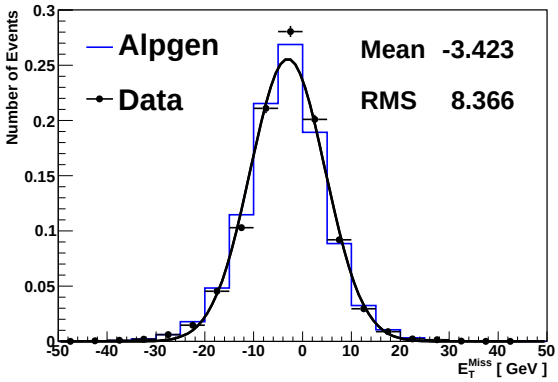
The magnitude of the bias is expected to have a strong dependence on the boson P_T. Any scale bias would be a result of some systematic mismeasurement of the hadronic momentum and the absolute deviation would therefore depend on the amount of hadronic recoil itself, which is strongly correlated with the boson transverse momentum as described above. The fake $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is expected to depend on the scalar sum E_T of the hadrons in the event since more hadronic energy in the calorimeter allows for larger absolute fluctuations. For these reasons the bias and resolution of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions are parameterised as functions of the vector boson P_T and the ΣE_T respectively, and are then used to smear the true $E_{\mathrm{T}}^{\mathrm{miss}}$ in W events as described further in section 8.3.3.

Since the Z events do not possess any real $E_{\mathrm{T}}^{\mathrm{miss}}$ the behaviour of the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ is extracted directly from Z data by estimating the mean and standard deviation of the $\not\!\!E_{\parallel}$ and $\not\!\!E_{\perp}$ distributions. The biases are evaluated first by calculating the means of the $\not\!\!E_{\parallel}$ and $\not\!\!E_{\perp}$ distributions in terms of the transverse momentum of the Z boson. One cannot in the same instance calculate the resolutions in terms of a different variable as these resolutions would then contain a contribution from the scale bias itself, which must first be removed. The events are therefore iterated over and the average bias is removed from the $E_{\mathrm{T}}^{\mathrm{miss}}$ in each event. The resolutions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions along the two coordinate axes are then determined by finding the standard deviation of the resulting distributions in terms of the scalar hadronic ΣE_T [2].

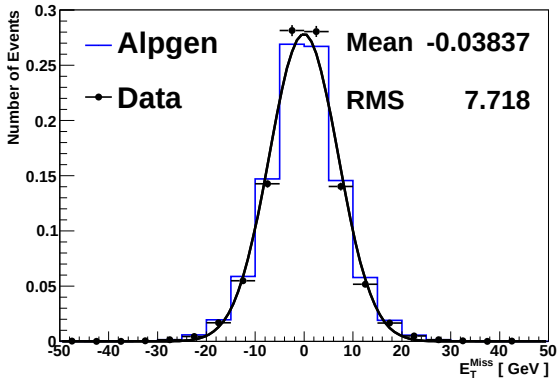
8.3.2 Measured Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ in Z Events

Figure 8.7 shows the `LocHadTopo` $\not\!\!E_{\parallel}$ and $\not\!\!E_{\perp}$ distributions from data and MC Z events without applying any jet requirements. As expected, the distribution along the perpendicular axis is centered around zero, whereas the distribution along the parallel axis is slightly

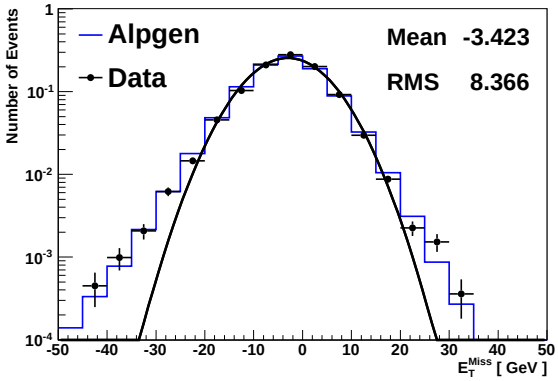
[2] In the muon channel this is simply the calorimeter ΣE_T, in the electron channel the transverse momenta of any reconstructed leptons are subtracted from the calorimeter ΣE_T.



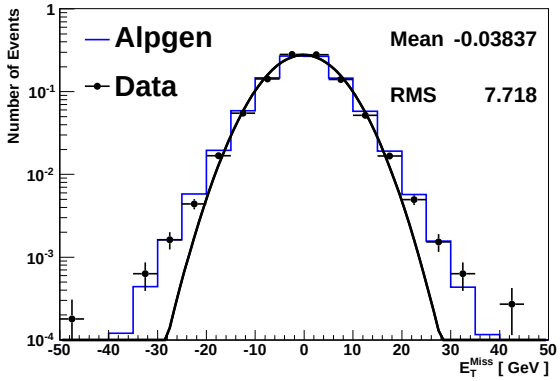
(a) $\not{E}_{\parallel}$ - Linear Ordinate



(b) $\not{E}_{\perp}$ - Linear Ordinate



(c) $\not{E}_{\parallel}$ - Logarithmic Ordinate



(d) $\not{E}_{\perp}$ - Logarithmic Ordinate

Figure 8.7: `LocHadTopo` $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions parallel and perpendicular to the P_T of the Z, for inclusive Z events in the muon channel.

shifted towards negative values, indicating a bias in the transverse energy measurement of the calorimeter. A negative value of the bias indicates an underestimation of the hadronic recoil in the event. The widths of the distributions are expected to be similar along both axes for inclusive events and this is observed to be the case. One would expect that at higher vector boson P_T the resolution in the parallel direction would become larger due to the requirement that the recoil momentum in this direction balance the vector boson P_T. The plots are shown fitted with Gaussian functions and while the core of the distributions are well described by the fits. both parallel and perpendicular distributions exhibit non-Gaussian tails and in the case of the $\not{E}_{\parallel}$ direction these are asymmetric.

Figure 8.8 shows the P_T of the reconstructed Z (P_T^Z) as well as the distribution of the scalar sum of the hadronic transverse energy measured by the calorimeter ($\sum E_T$) using

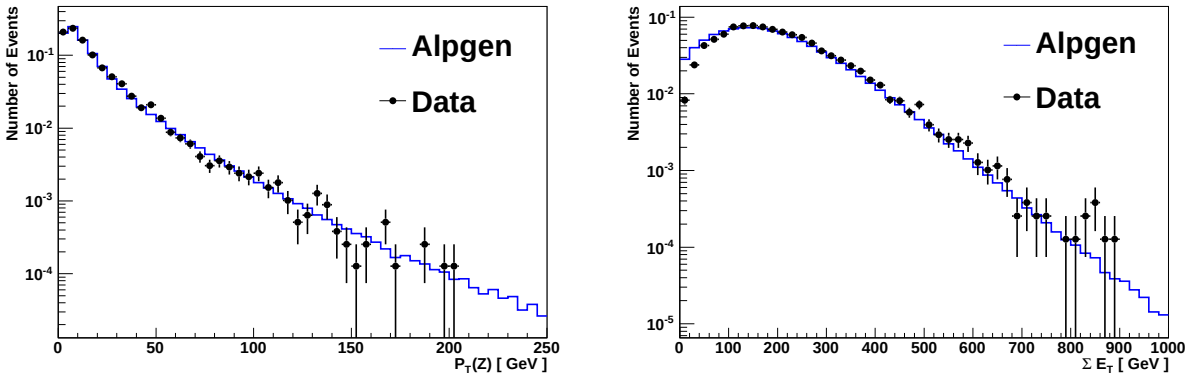


Figure 8.8: The Z boson P_T (left) and event ΣE_T (right) distributions from `RefFinal` for inclusive Z events in the electron channel.

`RefFinal` , obtained from events passing the electron channel Z selection. A good agreement is observed between data and the prediction from MC except at low values of the $\sum E_T$, where the values measured in data fall markedly below the values observed in Monte Carlo. The implications of discrepancies in the $\sum E_T$ distributions are discussed further in section 8.3.4. The distributions obtained using the other $E_{\mathrm{T}}^{\mathrm{miss}}$ definitions show similar agreement.

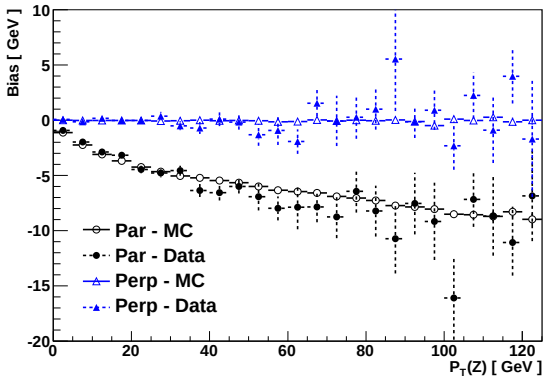
Figures 8.9 and Fig. 8.10 show the scale bias and resolutions respectively obtained using the three $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithms described in section 3.3.3 for events passing the inclusive Z selections. Both `LocHadTopo` and `HadRecoil` show a negative bias in the parallel direction strongly dependent on the boson P_T. The biases are of similar magnitude in these two cases, which is expected as they are both based on the same calibration scheme albeit with different treatment of the leptons in the event. The bias indicates that the local hadronic calibration scheme to some extent underestimates the scale of the hadronic recoil in the event. `RefFinal` on the other hand shows a more complicated behaviour but with, on average, a much smaller bias. The reduced bias is a result of the more complex object-based reconstruction algorithm implemented in `RefFinal` . It is interesting to note that the bias is opposite in sign at large vector boson P_T in the two channels although still within 3 GeV in each case at the highest boson P_T considered (120 GeV). The minimum can also be understood by looking at the behaviour of the bias in terms of jet multiplicity and noting at which boson P_T the events move from being dominated by 0 jet topologies to 1 jet topologies [69].

As described above, the axisymmetry of the event leads us to expect no bias in the direction perpendicular to the direction of the vector boson and this is observed in both data and MC for all 3 algorithms. The agreement between data and Monte Carlo in both directions is very good. The resolution plots in all cases show a slightly larger value in the parallel direction both in data and MC for the reasons discussed above. The resolutions obtained from data are however smaller by roughly 1 GeV than those obtained from MC. At low values of the $\sum E_T$ the `HadRecoil` algorithm gives a markedly smaller resolution as a result of its treatment of the leptons, while for higher values it tends towards that obtained using `LocHadTopo` as one would expect. `RefFinal` has a resolution close to that of `LocHadTopo` at low $\sum E_T$ but due to a behaviour that is closer to being linear than `LocHadTopo` , its value rises above that of `LocHadTopo` for larger values of the $\sum E_T$.

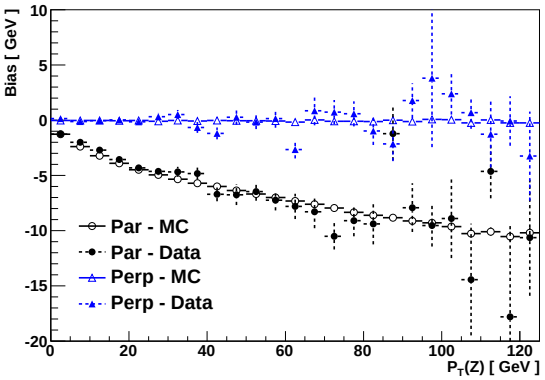
The dependence on $\sum E_T$ shows the typical $\sqrt{\sum E_T}$ behaviour from the stochastic energy resolution of the calorimeter, which implies a rather modest increase in the $E_{\mathrm{T}}^{\mathrm{miss}}$ with increasing $\sum E_T$. Furthermore, it was found that the impact from including pile-up in the MC results had no significant effect on the functional dependence between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the P_T^Z or $\sum E_T$. Hence the impact from pile-up only originates from the fact that a given event will acquire a larger $\sum E_T$ value from the increased hadronic activity.

Corresponding results for the one jet case are shown in appendix C. Due to the reduction in the amount of available data resulting from requiring a jet, the bias and resolution plots are subject to markedly larger statistical fluctuations than in the inclusive case. The behaviour of the bias and resolution plots is similar to the inclusive case with some minor exceptions. The bias in the one jet case is smaller at low WP_T when a jet is required due to the improved calibration of a single high P_T object as compared with several clusters of lower energy. The two sets of plots converge at high WP_T as this region is dominated in the inclusive case by events which contain jets. The resolution plots are again similar in structure although the one jet plots show a more pronounced deviation between the values in the parallel and perpendicular directions for the reasons outlined above.

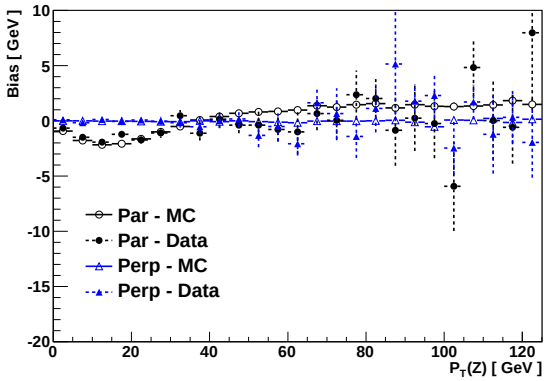
These results show that the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ from hadronic activity in the Z data is well



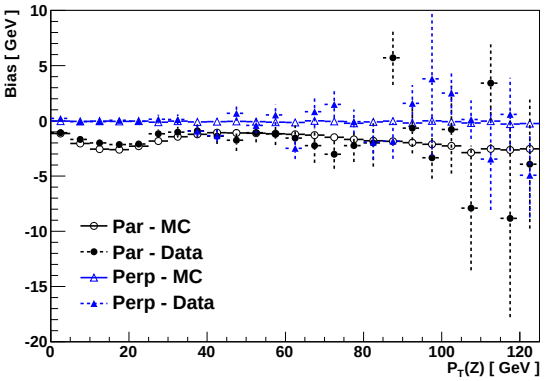
(a) Electron Channel - `LocHadTopo`



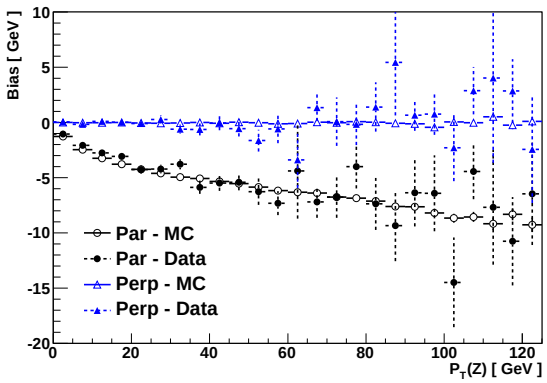
(b) Muon Channel - `LocHadTopo`



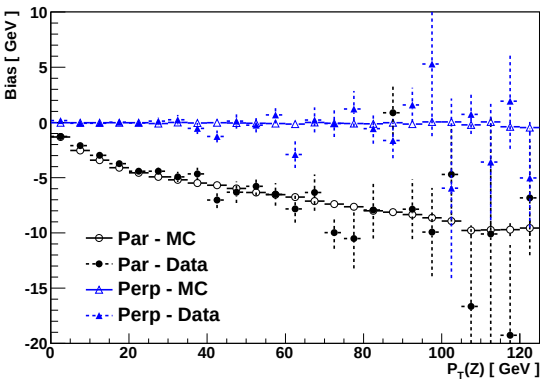
(c) Electron Channel - `RefFinal`



(d) Muon Channel - `RefFinal`

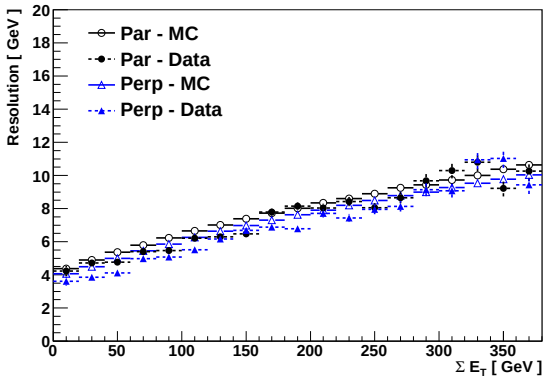


(e) Electron Channel - `HadRecoil`

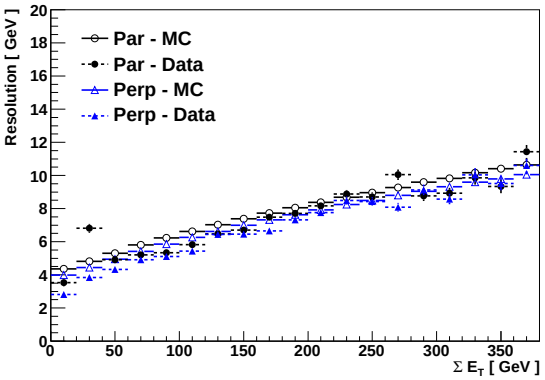


(f) Muon Channel - `HadRecoil`

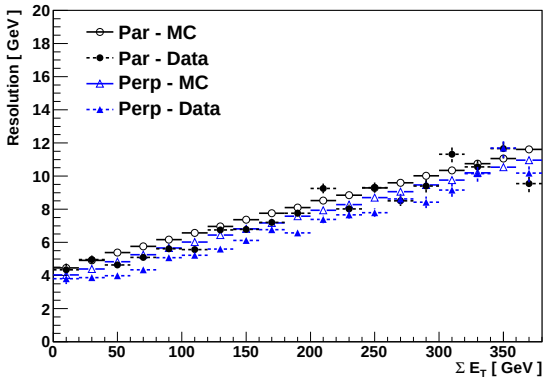
Figure 8.9: $E_{\mathrm{T}}^{\mathrm{miss}}$ scale biases (means of projected $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions) as a function of ZP_T obtained using the three $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithms in both channels.



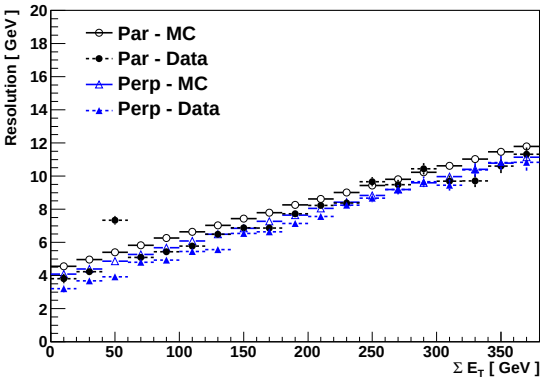
(a) Electron Channel - `LocHadTopo`



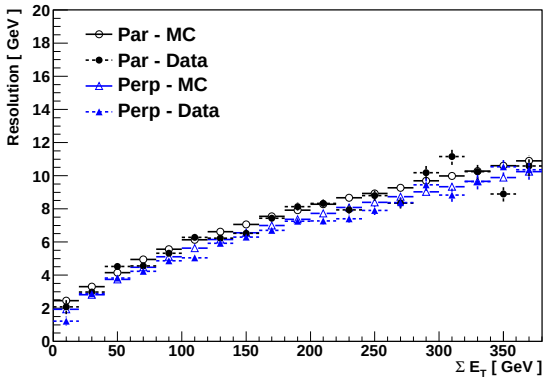
(b) Muon Channel - `LocHadTopo`



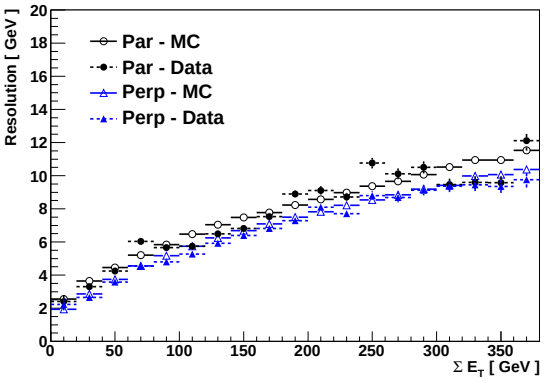
(c) Electron Channel - `RefFinal`



(d) Muon Channel - `RefFinal`



(e) Electron Channel - `HadRecoil`



(f) Muon Channel - `HadRecoil`

Figure 8.10: $E_{\mathrm{T}}^{\mathrm{miss}}$ resolutions (widths of projected $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions) as functions of the ΣE_T obtained using the three $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithms in both channels.

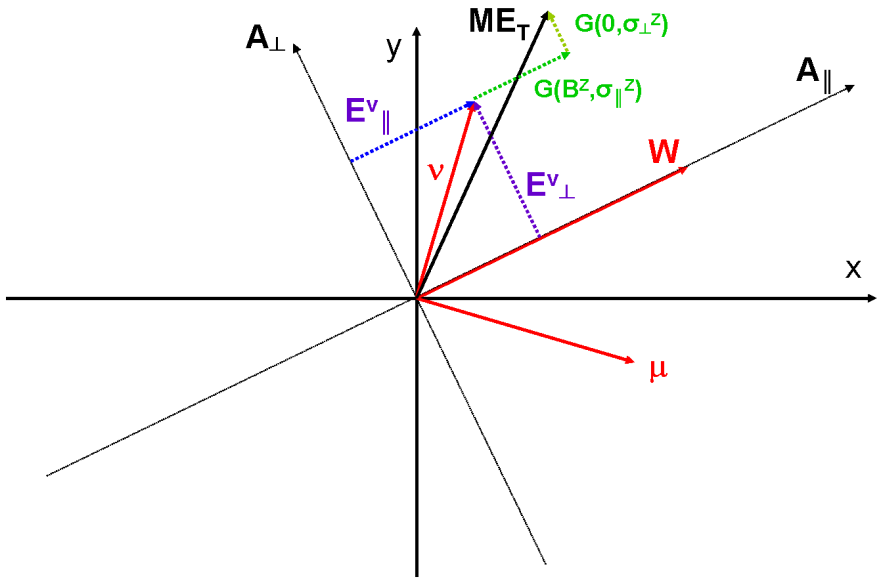


Figure 8.11: This diagram shows the transverse plane in a typical W event and illustrates how the neutrino momentum is smeared along axes parallel and perpendicular to the vector boson direction to recover the $E_{\mathrm{T}}^{\mathrm{miss}}$.

described by the MC simulations. Although this is the case, there are numerous sources of systematic uncertainty on the Monte Carlo predictions. Understanding all of these sources and enumerating their effect on the $E_{\mathrm{T}}^{\mathrm{miss}}$ acceptance in W events is difficult, and the resulting uncertainties can be large, as was shown in the previous sections. For this reason it is still valuable to develop a data-driven method which affords a simple procedure for evaluating the systematic uncertainties, and this is what is outlined in the next section.

The background contamination of the selected Z sample is generally small and therefore no significant effects from background are expected.

8.3.3 The Decomposition Method and $E_{\mathrm{T}}^{\mathrm{miss}}$ in W Events

Having discussed how to optimally decompose and parameterise the fake component of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in Z events, we now discuss how one applies this information to truth W MC. Leptonically decaying W events, unlike Z events, contain true $E_{\mathrm{T}}^{\mathrm{miss}}$ from the daughter neutrino of the W. We wish to add to this true $E_{\mathrm{T}}^{\mathrm{miss}}$ a fake component which accurately reflects the folding of detector effects into the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. In order to do this, the neutrino P_T is decomposed into components along a coordinate system aligned with the

direction of the truth W boson, analogously to how the $E_{\mathrm{T}}^{\mathrm{miss}}$ was decomposed in the Z events discussed in section 8.3.2. Each component is then independently smeared by addition of a random value drawn from a Gaussian distribution, with a mean given by the scale bias corresponding to the value of the W P_T and a width given by the corresponding resolution at the value of the reconstructed $\sum E_T$:

$$\begin{aligned} \not{E}_{\parallel} &= \not{E}^{\nu}_{\parallel} + G(B^Z(P_T^W), \sigma^Z_{\parallel}(\sum E_T^{\ Reco})) \\ \not{E}_{\perp} &= \not{E}^{\nu}_{\perp} + G(0, \sigma^Z_{\perp}(\sum E_T^{\ Reco})) \end{aligned} \tag{8.7}$$

In the above equation 'G' represents a Gaussian function with the mean and width given in parenthesis. B^Z represents the scale bias obtained from Z events as a function of the boson P_T, while $\sigma^Z_{\parallel}$ and $\sigma^Z_{\perp}$ represent the parallel and perpendicular resolution functions obtained from Z events as a function of the reconstructed $\sum E_T$. This procedure is illustrated in figure 8.11. Once this procedure is complete, one has a $E_{\mathrm{T}}^{\mathrm{miss}}$ value which behaves as a reconstruction level quantity, having made minimal explicit use of the GEANT4 simulation of the ATLAS detector.

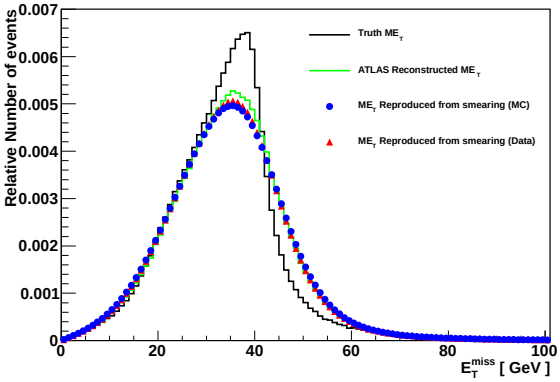
It was shown in figure 8.7 that although the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions are well modelled by a Gaussian in the core, the tails of the distributions are distinctly non-Gaussian and also asymmetric along the parallel axis. An alternative method, which would accommodate the tails, would be to smear the W events directly using the measured $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions in Z events. As above, the appropriate scale bias would be applied, then a random value would be drawn from the distribution obtained from Z events with the corresponding value of the hadronic $\sum E_T$. This was tried and found to have little effect on the results. The rest of the work presented refers to results obtained using the Gaussian method outlined above.

In order to test the effectiveness of this method, one can use Monte Carlo Z events to derive the resolution functions and then apply these to Monte Carlo W events generated in the same way. Any residual non-closure between quantities obtained using the standard ATLAS reconstruction and those obtained using the decomposition method described here

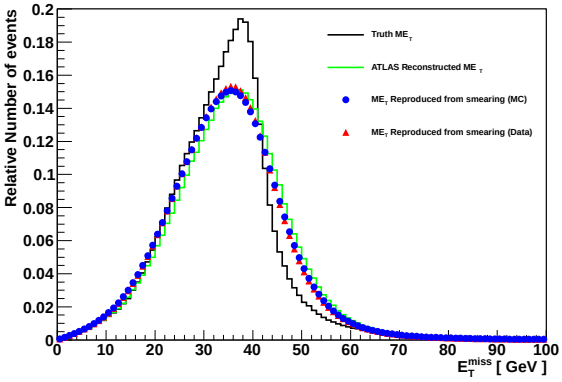
indicate an imperfection in the method and can be regarded as an additional systematic to those described in section 8.3.4.

Figure 8.12 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions in inclusive W events for the three different algorithms considered. The black solid histogram depicts the neutrino transverse momentum (truth $E_{\mathrm{T}}^{\mathrm{miss}}$), which is obviously reconstruction algorithm independent. The solid green histogram in each case shows the ATLAS reconstruction level $E_{\mathrm{T}}^{\mathrm{miss}}$ for that algorithm. The blue markers show the $E_{\mathrm{T}}^{\mathrm{miss}}$ obtained using the decomposition method described here along with their total systematic uncertainty where the resolution functions have been obtained from Z MC. Finally the red markers show the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution obtained with the decomposition method and resolution functions obtained from Z data. If the procedure was perfect, the blue markers would follow the solid green histogram exactly. Note that the red markers need not follow the green histogram and that deviations between blue and red indicate genuine data-MC differences. The agreement between the blue and red histograms is actually very good, which is consistent with the good agreement observed between data and MC for the resolution functions taken from Z events in section 8.3.2. Corresponding histograms showing the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions in the one jet case are shown in appendix C. Despite the lack of data available in the one jet case, the smeared distributions still agree very well between data and Monte Carlo and with the nominal reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$.

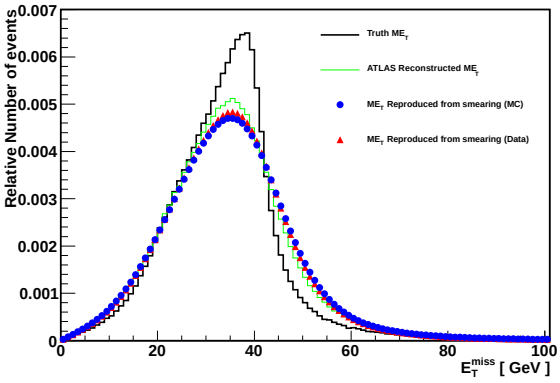
While it is clear in all cases that the decomposition method produces smeared distributions markedly closer to the reconstructed ones than to the truth, there are still residual deviations with respect to the reconstructed distributions which result in non-closures on calculating acceptances. The deviations are most noticeable for `LocHadTopo` and `RefFinal` in the muon channel, where the smeared distributions show a shift to lower $E_{\mathrm{T}}^{\mathrm{miss}}$ values when compared with the reconstructed distribution. The differences are less pronounced in the electron channel, where the reconstructed distributions are more strongly peaked than the smeared distributions. In the inclusive case the smallest deviations between the smeared and reconstructed distributions are observed when using the `HadRecoil` algorithm. In the muon channel the agreement is extremely good while in the electron channel there is a small



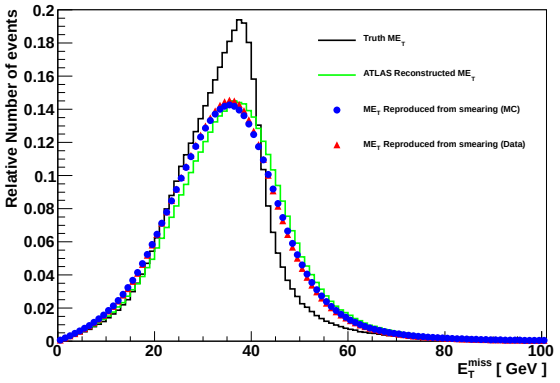
(a) Electron Channel - `LocHadTopo`



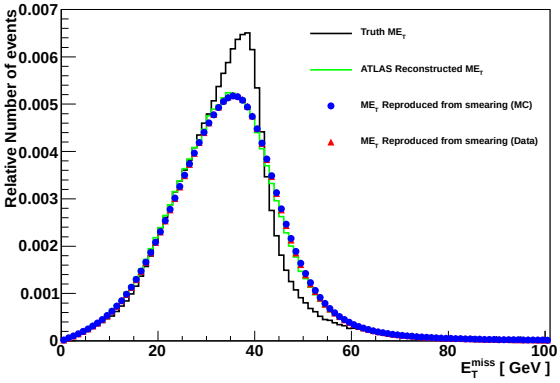
(b) Muon Channel - `LocHadTopo`



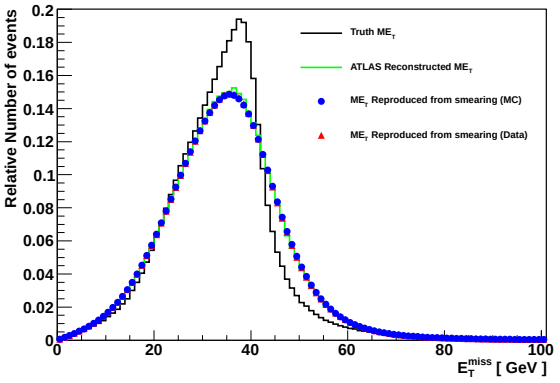
(c) Electron Channel - `RefFinal`



(d) Muon Channel - `RefFinal`



(e) Electron Channel - `HadRecoil`



(f) Muon Channel - `HadRecoil`

Figure 8.12: Comparison of the standard ATLAS reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution obtained from the GEANT4 detector simulation (green) with the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions obtained using the $E_{\mathrm{T}}^{\mathrm{miss}}$ decomposition method described in the text (Blue - Resolution functions obtained from MC, Red - Resolution functions obtained from data).

Algorithm	Type	Acceptance	Uncertainty	Non-Closure
Truth		0.8121		
`LocHadTopo`	Reconstructed	0.8253		2.80%
	MC Decomp.	0.8022	0.194%	
	Data Decomp.	0.8035	0.337%	
`RefFinal`	Reconstructed	0.8242		3.53%
	MC Decomp.	0.7951	0.192%	
	Data Decomp.	0.7964	0.285%	
`HadRecoil`	Reconstructed	0.8038		0.36%
	MC Decomp.	0.8009	0.256%	
	Data Decomp.	0.8012	0.363%	

Table 8.1: Inclusive $E_{\mathrm{T}}^{\mathrm{miss}}$ acceptances calculated in the muon channel using the different $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithms. The uncertainties and non-closures are also shown. MC and Data Decomp. refer respectively to values obtained using the $E_{\mathrm{T}}^{\mathrm{miss}}$ decomposition method presented in this note with resolution functions obtained from MC and data.

discrepancy, but the agreement is still better than that produced by the other algorithms. Given the only difference of the `HadRecoil` algorithm with respect to `LocHadTopo` is in its treatment of the leptons in the event, it can be concluded that the residual differences observable when using the other algorithms are largely a result of the contribution of the leptons to the resolution functions being different between W and Z events and no explicit consideration being given to this fact when using the decomposition method with these algorithms.

Tables 8.1 and 8.2 list the inclusive $E_{\mathrm{T}}^{\mathrm{miss}}$ acceptances calculated using the different algorithms and their associated uncertainties in the muon and electron channels respectively. The non-closures are smaller in the electron channel, consistent with the improved agreement between the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions above. The smaller non-closure of `RefFinal` as compared to `HadRecoil` in the electron channel is an artifact of the particular choice of $E_{\mathrm{T}}^{\mathrm{miss}}$ cut used in this analysis, if one moved the cut in either direction `HadRecoil` would perform similarly well, while increasing the cut would result in quite large deviations when using `RefFinal` .

As well as simply calculating the inclusive acceptance, one can calculate acceptances as a function of some variable related to the event topology. As an illustration of this, the acceptance is calculated as a function of the transverse momentum of the W boson in

Algorithm	Type	Acceptance	Uncertainty	Non-Closure
Truth		0.7987		
`LocHadTopo`	Reconstructed	0.8002		
	MC Decomp.	0.7889	0.185%	1.41%
	Data Decomp.	0.7916	0.289%	
`RefFinal`	Reconstructed	0.7812		
	MC Decomp.	0.7832	0.172%	0.26%
	Data Decomp.	0.7847	0.267%	
`HadRecoil`	Reconstructed	0.7849		
	MC Decomp.	0.7927	0.211%	0.99%
	Data Decomp.	0.7933	0.288%	

Table 8.2: Inclusive $E_{\mathrm{T}}^{\mathrm{miss}}$ acceptances calculated in the electron channel using the different $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithms. The uncertainties and non-closures are also shown. MC and Data Decomp. refer respectively to values obtained using the $E_{\mathrm{T}}^{\mathrm{miss}}$ decomposition method presented in this note with resolution functions obtained from MC and data.

the inclusive case and as a function of the leading jet transverse momentum in the one jet case. Figure 8.13 shows the acceptance as a function of W P_T calculated using the different algorithms for both channels. The colour coding of the histograms is the same as for the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions above and again one hopes that the blue markers reproduce the green histogram. Figure 8.14 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ acceptance calculated as a function of the leading jet P_T in one jet events.

Figure 8.15 summarises these results in a form which makes it easier to judge the performance of the $E_{\mathrm{T}}^{\mathrm{miss}}$ decomposition method. The plots show the relative deviations of the acceptances calculated using the decomposition method, with resolution functions taken from Z MC, from the acceptances calculated using the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ from the detector simulation for the three algorithms (the ratio of the blue histogram with the green histogram). This allows one to clearly see how the non-closure varies as a function of the W P_T and jet P_T respectively, and contrast this with the overall uncertainty on the method. In the muon channel the smeared distributions obtained using `LocHadTopo` and `RefFinal` give rise to markedly low values of the acceptance at low W P_T, whereas the acceptance calculated using `HadRecoil` is close to being consistent with one within the systematic errors associated with the method. The agreement is generally better in the electron channel, at low $W P_T$ `RefFinal` seems to perform the best although, as discussed above, this is likely to

be particular to the choice of $E_{\mathrm{T}}^{\mathrm{miss}}$ cut used. `HadRecoil` shows a slight positive deviation, this could possibly be an indication that the $\Delta R = 0.2$ cone used to remove the lepton is not always removing all clusters due to the electron. One could look at the behaviour of the non-closure as a function of the cone size used to perform the lepton removal in the `HadRecoil` algorithm to see if this is the case. If the non-closure is a result of incomplete removal of the lepton clusters, then increasing the cone size would be expected to improve the performance of the procedure, however if the non-closure is due to resolution differences between W and Z events caused by the underlying event compensation procedure, an increased cone size would make this difference more pronounced and degrade the performance. Given the non-closure is smaller in the muon channel, one might suspect that the former is true and that increasing the cone size should improve matters, in which case it may be preferable to sacrifice some degradation of the resolution in order to improve the performance of this method and reduce the systematics on the acceptance calculation[3].

The decomposition method shows similarly promising results in the one jet case although it is more difficult to speculate on the origin of any observed non-closures in this instance. The behaviour of `LocHadTopo` is similar in both channels showing a systematically low value of the acceptance at low jet P_T, again likely attributable to the lepton differences. `HadRecoil` also shows an identical behaviour between the two channels with excellent agreement at low and high jet P_T but with an upward deviation rising to 2% between 50 and 80 GeV. This is not fully understood, however the `HadRecoil` algorithm does not currently demand that the underlying event correction cone be separated in $\eta - \phi$ space from any jets, only from the recoil direction and the lepton. In the inclusive case this is perfectly acceptable but for dijet events the recoil direction may not be close to the jets in the event leaving the possibility that one uses a systematically large underlying event correction if the UE replacement cone is close to a jet. It is possible that setting requirements on the separation of the UE cone from jets in the event may improve the agreement here, but it may not be the entire cause

[3]The value of the cone size employed in the `HadRecoil` algorithm was chosen so as to minimize the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. This does not necessary imply that all of the clusters due to the electron are being removed, simply that the increase in resolution from the increased amount of underlying event compensation has more effect than the decrease from removing additional electron clusters in a small number of cases.

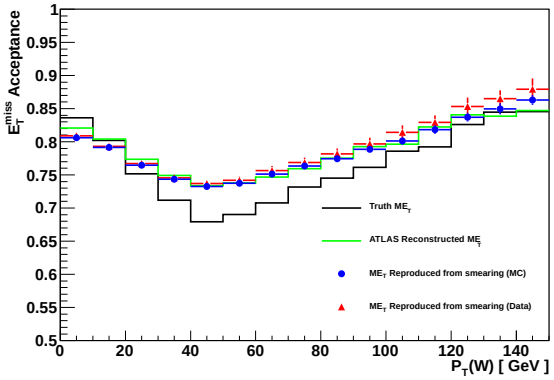
of this discrepancy. Similarly, in cases where there is a jet close to the removed lepton, the removal process will be removing a substantial amount of hadronic energy along with the energy due to the lepton itself and this energy will not be replaced. This effect is harder to control but an idea of its impact could be understood by looking at different lepton isolation criteria. `RefFinal` shows markedly different agreement at low jet P_T between the two channels, being $\sim 2\%$ deviated in both cases although in opposite directions.

8.3.4 Systematic uncertainties

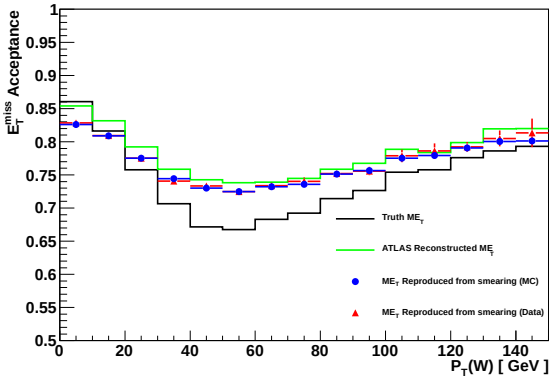
In contrast to the Monte Carlo based method, where important systematics had to be considered individually, the method presented here would in ideal circumstances (i.e. when the leptons do not influence the $E_{\mathrm{T}}^{\mathrm{miss}}$) automatically fold all pertinent effects related to the $E_{\mathrm{T}}^{\mathrm{miss}}$ into the resolution functions obtained from data. Consequently, systematic uncertainties in this method are obtained by making reasonable variations of the resolution functions, reconstructing the $E_{\mathrm{T}}^{\mathrm{miss}}$ based on these new functions and recalculating acceptances based on the new $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions. 8 combinations of different variations of the resolution functions were considered:

$$\{B_Z, \sigma_{\parallel}, \sigma_{\perp}\} \in \{000, +00, -00, 0++, 0--, +++, +--, -++, ---\} \tag{8.8}$$

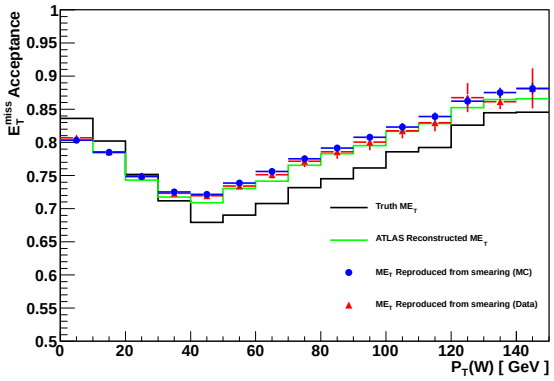
Where '0' indicates no variation, '+' indicates the value of the corresponding function has been shifted upwards by its statistical uncertainty and '-' indicates the corresponding value has been shifted down by its statistical uncertainty. For each of these cases 300 pseudoexperiments were performed. The mean value of the acceptance in the nominal configuration (000) was taken as the central value and the width of the distribution was considered as an uncertainty. Then the deviations between the nominal mean and the means of the other 8 configurations were determined. The largest deviation in each direction was found, their moduli were averaged and then this value was added in quadrature with the width of the



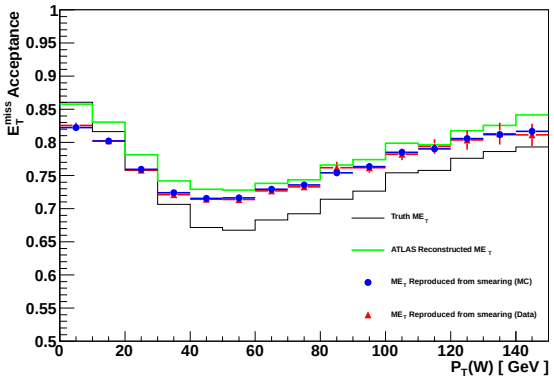
(a) Electron Channel - `LocHadTopo`



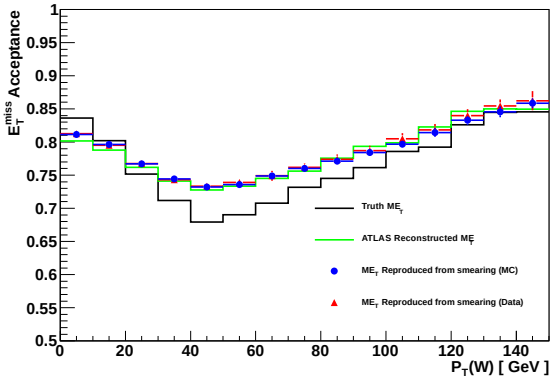
(b) Muon Channel - `LocHadTopo`



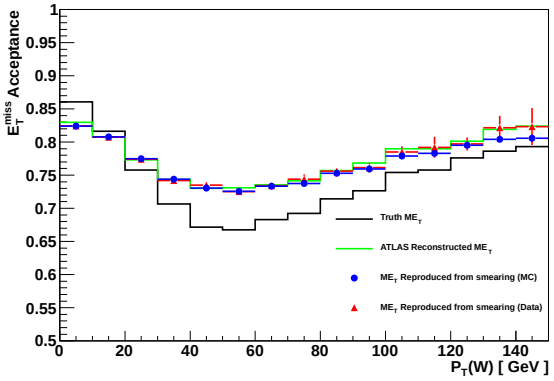
(c) Electron Channel - `RefFinal`



(d) Muon Channel - `RefFinal`

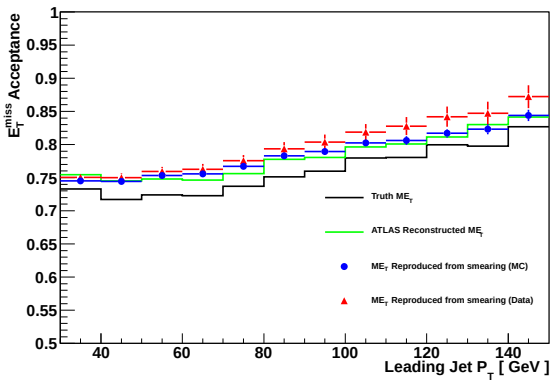


(e) Electron Channel - `HadRecoil`

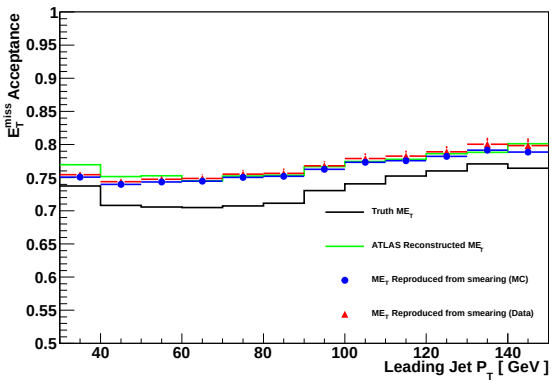


(f) Muon Channel - `HadRecoil`

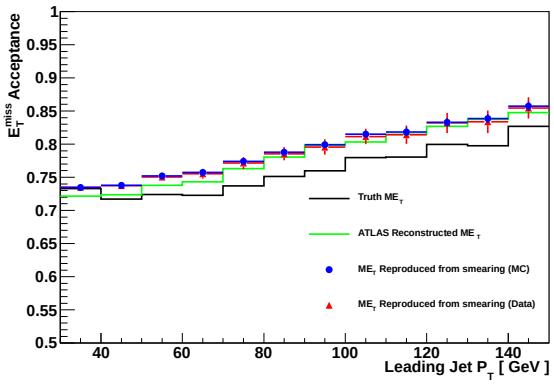
Figure 8.13: Comparison of the $E_{\mathrm{T}}^{\mathrm{miss}}$ acceptance calculated as a function of the W P_T using the standard ATLAS reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ obtained from the GEANT4 detector simulation (green) with that obtained using the $E_{\mathrm{T}}^{\mathrm{miss}}$ decomposition method described in the text (Blue - Resolution functions obtained from MC, Red - Resolution functions obtained from data).



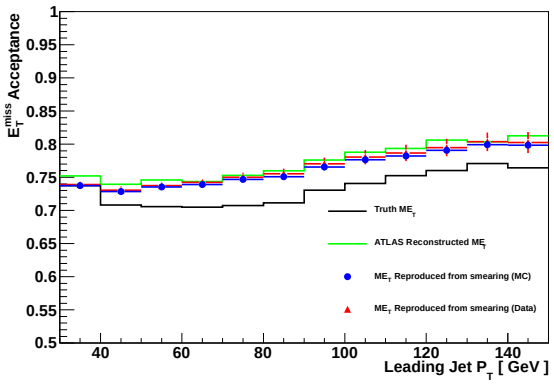
(a) Electron Channel - `LocHadTopo`



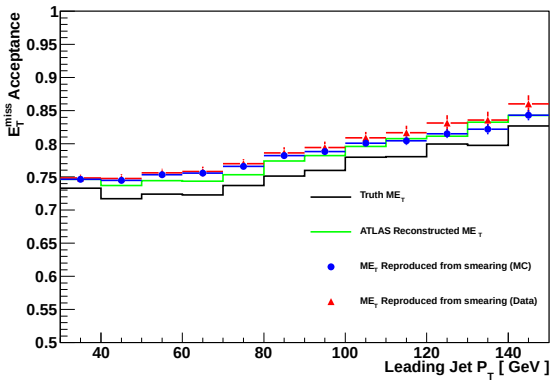
(b) Muon Channel - `LocHadTopo`



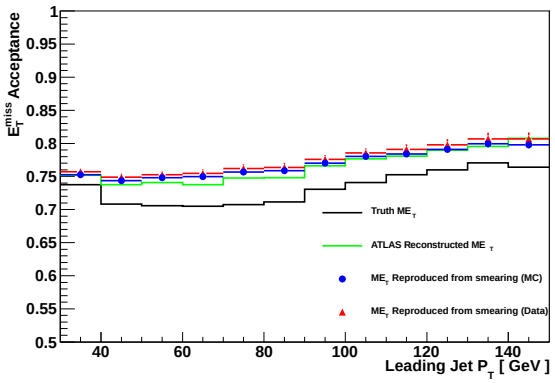
(c) Electron Channel - `RefFinal`



(d) Muon Channel - `RefFinal`

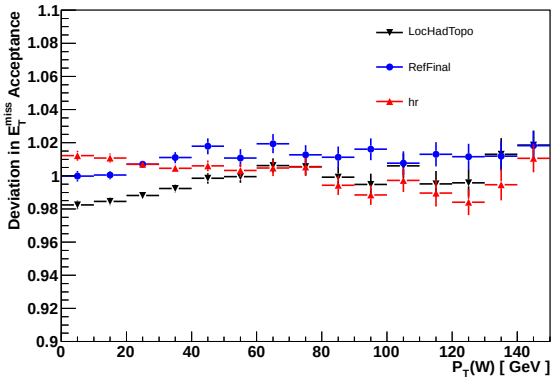


(e) Electron Channel - `HadRecoil`

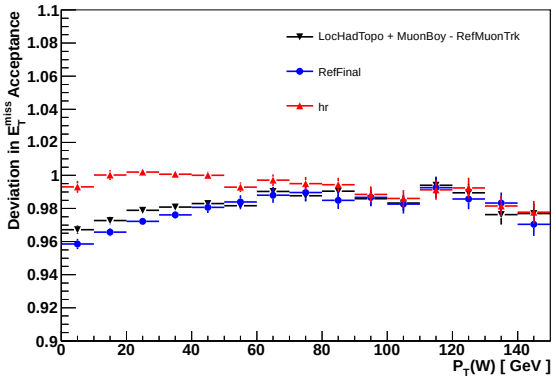


(f) Muon Channel - `HadRecoil`

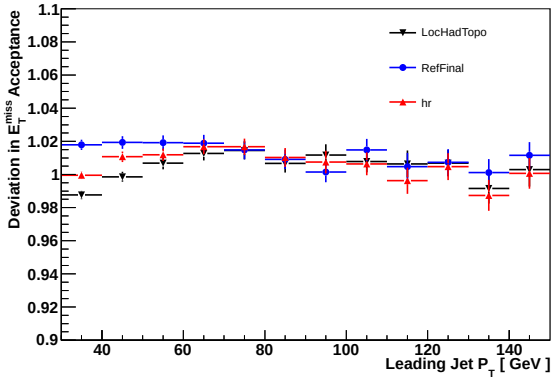
Figure 8.14: Comparison of the $E_{\mathrm{T}}^{\mathrm{miss}}$acceptance calculated as a function of the leading jet P_T using the standard ATLAS reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ obtained from the GEANT4 detector simulation (green) with that obtained using the $E_{\mathrm{T}}^{\mathrm{miss}}$ decomposition method described in the text (Blue - Resolution functions obtained from MC, Red - Resolution functions obtained from data).



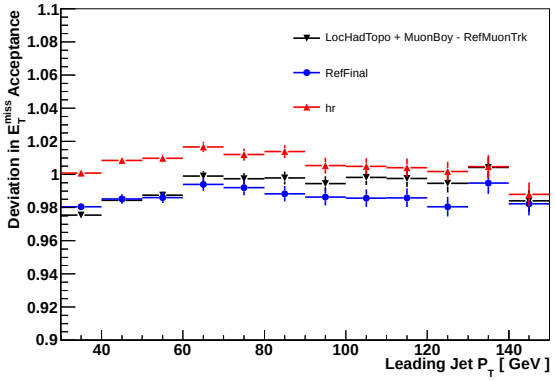
(a) Electron Channel - Inclusive



(b) Muon Channel - Inclusive

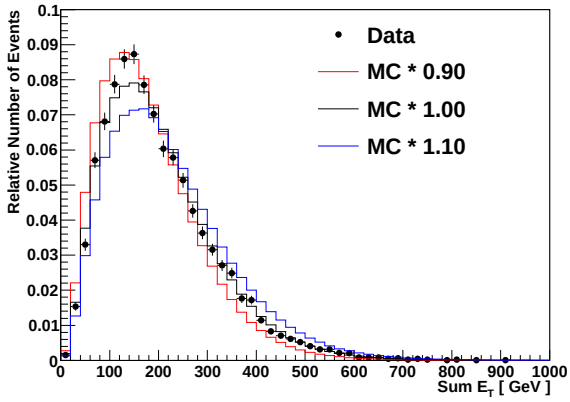


(c) Electron Channel - 1 Jet



(d) Muon Channel - 1 Jet

Figure 8.15: Plots showing the non-closure of the acceptance as a function of the W P_T in the inclusive case and the leading jet P_T in the one jet case.



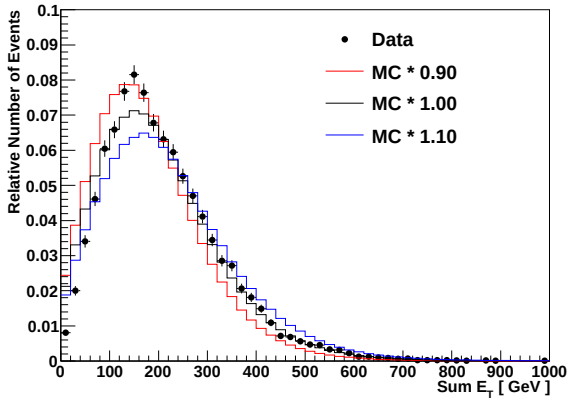


Figure 8.16: Plots comparing the $\sum E_T$ distributions observed in data with the nominal Monte Carlo prediction as well as predictions obtained by shifting each event's $\sum E_T$ up or down by 10%. The plots are taken from the muon channel, the left uses `HadRecoil` , the right uses `RefFinal` .

nominal distribution.

The last systematic considered is a consequence of the imperfect modelling of the reconstructed $\sum E_T$ in Monte Carlo. If the $\sum E_T$ is incorrectly determined for a given event, the values of the resolutions used to smear the $E_{\mathrm{T}}^{\mathrm{miss}}$ of that event will also be incorrect and this would lead to wrongly calculated acceptances. Given the $\sum E_T$ distributions agree quite well between data and MC, as shown in section 8.3.2, and also that the resolution functions are not strongly varying functions of the $\sum E_T (\sim \sqrt{\sum E_T})$, it is not expected that this source of systematic will contribute a large uncertainty to the acceptance. To determine the size of this uncertainty, the Monte Carlo $\sum E_T$ values were shifted by 10% in either direction and the acceptances were recalculated (again the mean of 300 pseudoexperiments was taken). The deviations from nominal were found, averaged as above and added in quadrature with the other uncertainties. One can see in figure 8.16 that shifting the $\sum E_T$ in this way produces distributions with lower and higher values of $\sum E_T$ than those seen in data. As such this method, while coarse, is believed to be conservative.

Figure 8.17 shows the contribution of the different systematics to the overall uncertainty for `RefFinal` in the electron channel as a function of W P_T. When the resolution functions are obtained from Monte Carlo, the uncertainty is dominated at low P_T by the $\sum E_T$ uncertainty, while at higher P_T the statistical width becomes dominant. When the resolution functions are taken from data, the statistical uncertainties are larger, which results in the different contributions to the overall uncertainty having comparable values over the entire range of the W P_T spectrum considered. The total uncertainty in MC is within 1% over the range of vector boson momenta considered. In data the uncertainty rises monotonically from below 0.5% in the first P_T bin and remains within 2% for all transverse momenta considered.

8.4 Method Comparison

The data driven method presented in the previous section was applied to the R_{Jets} measurement. The results of this are shown in figure 8.18. The reduced amount of data in the one jet

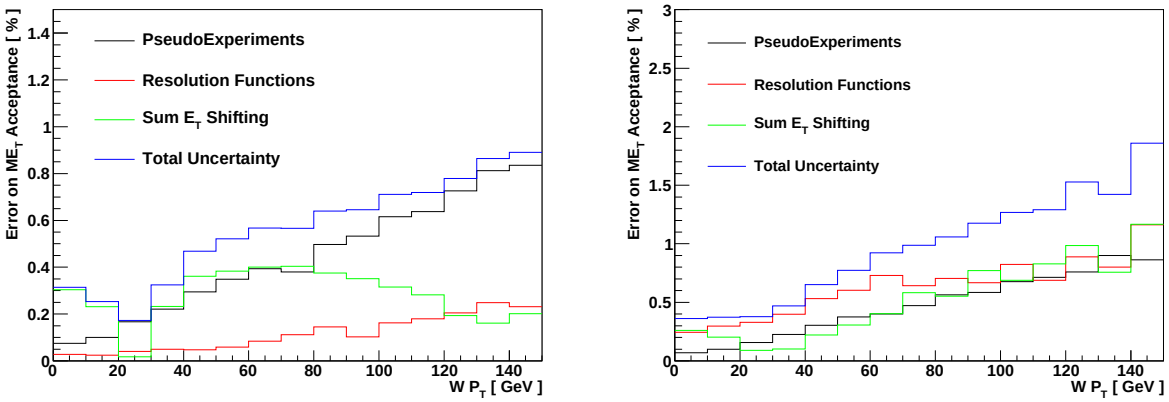


Figure 8.17: Plots showing the different contributions to the systematic uncertainty as a function of the W P_T in the electron channel. In the left hand plot the resolution functions are obtained from Z MC while in the right hand plot they are taken from data. Note that the scale of the ordinate is different between the two plots.

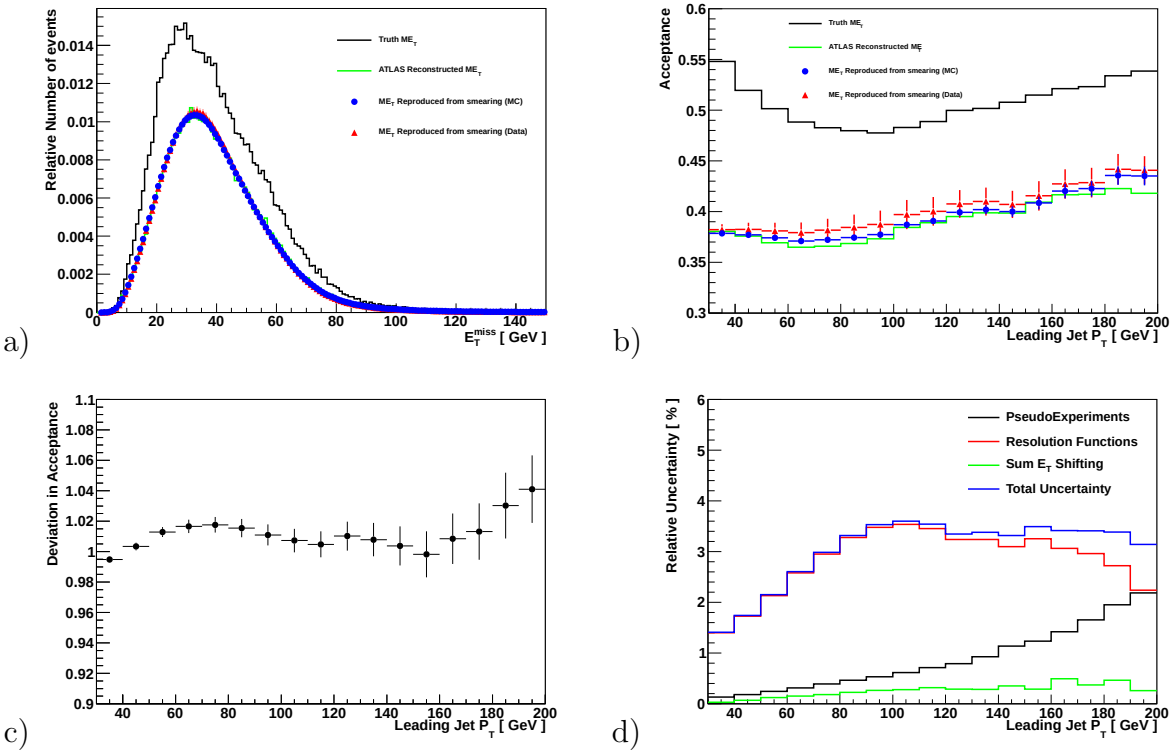


Figure 8.18: Plots showing the results of applying the data-driven decomposition method to the R_{Jets} measurement. a) Shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions after applying all selections, b) shows the W acceptance at reconstruction level $A(\ell_R, E_{T,R}^{miss}, M_{T,R}|J_R)$ as a function of the leading jet transverse momentum threshold, c) shows the non-closure and d) shows the uncertainties on the acceptance.

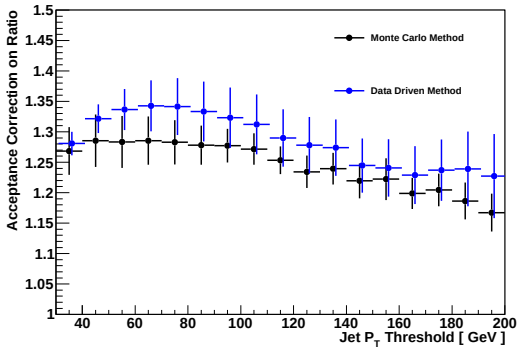


Figure 8.19: This plot shows the acceptance correction on R_{Jets} obtained using the Monte Carlo based and data-driven methods along with their respective total uncertainties.

case leads to significantly larger uncertainties due to shifting the Z resolution functions than in the inclusive case. This uncertainty is dominant at all leading jet transverse momenta, rising rapidly from $\sim 2.5\%$ at 30 GeV to a maximum $\sim 3.5\%$ at 100 GeV. This uncertainty will be reduced dramatically with the availability of more data, but at present this large uncertainty makes it more favourable to use the Monte Carlo based approach for the rest of the analysis.

The two results for the R_{Jets} acceptance correction are shown in figure 8.19. They are mostly compatible, except for the region between 40 and 90 GeV where the deviation is slightly larger than the uncertainty. This is also the region where the non-closure is the largest. This suggests that understanding the origin of the non-closure and how to remove or correct for it will be vital in the application of this method to future analyses. It is possible that whatever the physical cause of the non-closure, the effect is more pronounced in data than MC. As described above there are several avenues of research open for tuning the method to reduce the non-closure, but these will require optimisation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ definition used and consequently will only be employed in later analyses.

Chapter 9

Results and Conclusions

9.1 Modelling Uncertainties

All of the factors required to correct the value of R_{Jets} observed in data for detector effects have now been calculated along with their respective uncertainties. Although uncertainties on the jet spectrum correction and the acceptance correction due to generator and pile-up modelling were separately calculated, these were not included at the time and are instead included here as a single uncertainty evaluated on the total correction from reconstruction level (prior to the application of identification criteria) to truth level. These factors, evaluated using ALPGEN and Pythia, with and without the inclusion of pile-up modelling as described in section 4.4, are shown in figure 9.1. A total modelling uncertainty is calculated as the quadrature sum of the difference between this correction when calculated using Pythia and ALPGEN and when calculated using the Pythia samples with and without pile-up modelling. This additional uncertainty is then added in quadrature with the other systematic uncertainties discussed in the previous chapters. Comparing two generators is not a rigorous procedure for assigning model systematics, in future it is intended that such systematics be assigned by the variation of relevant parameters within a single model. The predictions of different models should then, in principle, be consistent with the nominal prediction according to the uncertainties derived in this way. It will be seen below that the modelling

uncertainty constitutes, at present, the largest systematic at high jet transverse momentum and consequently it is important that this is better understood for future analyses which will no longer be statistically dominated in this region of phase space.

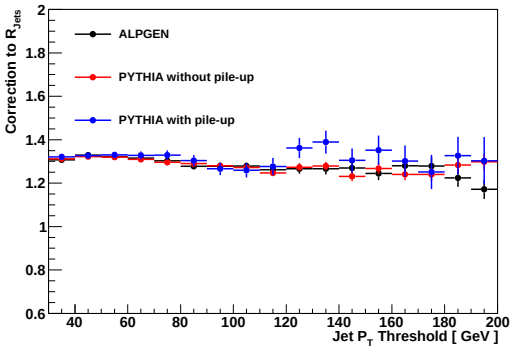


Figure 9.1: The R_{Jets} detector unfolding correction calculated using ALPGEN and Pythia with and without the inclusion of pile-up. The modelling uncertainty on the measurement is calculated as the quadrature sum of the differences between the values calculated using ALPGEN and Pythia (Black and Red) and Pythia with and without the inclusion of pile-up modelling (Red and Blue).

9.2 Results

The raw numbers of events passing the W and Z event selections, as well as their ratio, are listed as a function of the reconstruction level jet transverse momentum threshold in table 9.1. The table also lists the numerical values of the correction factors derived in the previous chapters, which are used to correct the raw value of R_{Jets} to particle level. Each of these corrections is expressed in the table as a multiplicative factor to be applied to the raw value of the ratio. As a result for some of the corrections, notably the acceptances, the values shown in the table are reciprocals of the values given in the chapters where the corrections were calculated. The final measured values of R_{Jets}, obtained using the prescription outlined in section 4.3, are shown in figure 9.2 along with the total statistical, systematic and modelling uncertainties on the measurement. The statistical uncertainty is evaluated by fluctuating the bin contents of the differential distributions of W and Z events obtained in data 100k times according to a Poisson distribution. For each trial, the W and Z distributions are cumulated and the value of R_{Jets} as a function of the leading jet P_T is

computed. The boundaries of the central 68.2% of the resulting distributions in each bin are quoted as the statistical uncertainty bounds. Due to the low statistics, the uncertainties tend to become asymmetric at high jet transverse momentum. It is important to recall here that due to the fact that the R_{Jets} measurement is made as a function of the jet transverse momentum *threshold*, the statistical uncertainties will be strongly correlated between bins. The theoretical prediction from MCFM, corrected using the parton to particle level correction discussed in section 5.5 is also shown in the figure with its associated theoretical uncertainty.

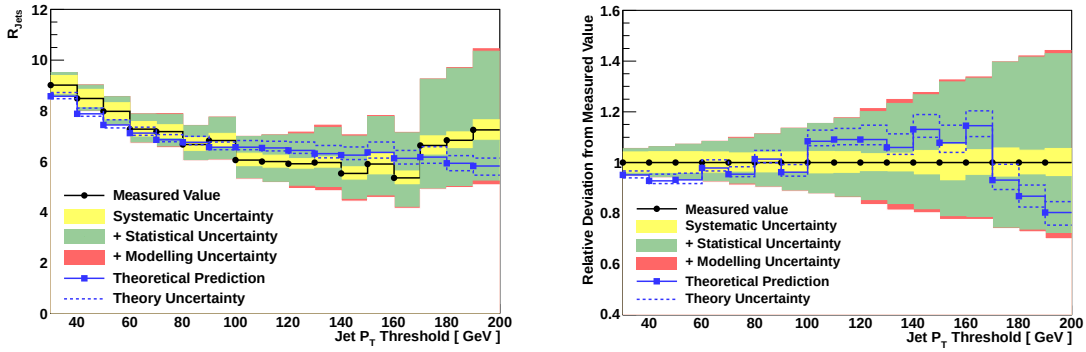


Figure 9.2: The measured value of R_{Jets}. In the above plots the measurement is shown with the contributions to its uncertainty successively added in quadrature. The points indicate the jet momentum threshold for which a particular measured value applies. The NLO theoretical prediction obtained using MCFM is also shown in the plots complete with its theoretical uncertainty. The right hand plot shows the value of the theoretical prediction relative to the measured value.

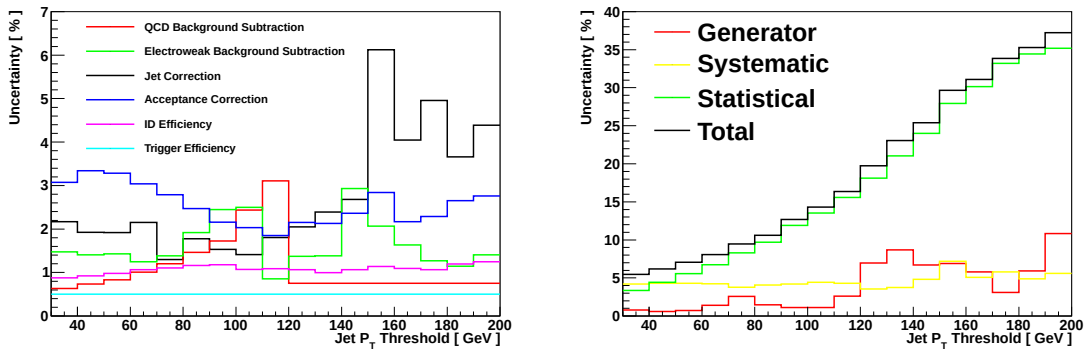


Figure 9.3: The left hand plot shows a breakdown of the systematic uncertainty on the R_{Jets} measurement into its components from the different correction factors. The right hand plot compares the combined systematic uncertainty with the statistical and generator uncertainties.

As can be seen from the figure, there is good agreement between the measured value of R_{Jets} and the Next-to-Leading Order predictions from perturbative QCD across the range of

jet transverse momenta considered. The lowest jet P_T bin is at present the only one for which the systematic uncertainty is the largest contribution to the total measurement uncertainty. While the systematic uncertainty remains fairly constant across the jet spectrum, the statistical uncertainty rapidly increases as the jet transverse momentum threshold increases and quickly becomes the dominant contribution to the uncertainty. The breakdown of the total uncertainty into its constituents is shown in figure 9.3 and enumerated in table 9.2. For low jet transverse momentum thresholds the dominant systematic component comes from the uncertainty on the acceptance correction, which is itself dominated by the $E_{\mathrm{T}}^{\mathrm{miss}}$ acceptance uncertainty. Hopefully with the availability of more data recorded in 2011, refinement of the data-driven techniques discussed in chapter 8 will allow a reduction of this uncertainty. At higher jet P_T thresholds, the uncertainties on the generator modelling and the jet correction overtake the acceptance correction as the dominant systematic components, although in the former case this is at least partly due to the lack of statistics present in the Monte Carlo samples with pile-up modelling.

For completeness, the official ATLAS results for the R_{Jets} measurement are also shown in figure 9.4 for both the muon and electron channel. The electron channel result is similar to the result presented in this work, which is unsurprising as the measurements share several common elements. The most important differences between the measurement presented in this work and the ATLAS electron channel measurement are due to the inclusion here of the topological method for obtaining the W QCD background and the efficiency calculation, which was performed differently here to in the official ATLAS result. The only noticeable difference between the two results is a slightly better agreement between theory and data for low leading jet P_T values in the ATLAS result as compared with the result presented here. The muon channel result also agrees extremely well with the NLO theoretical predictions across the full phase space of the measurement. The muon channel results are slightly lower due to different fiducial selections being applied when performing the measurement in this channel.

Jet P_T Threshold [GeV]	30	40	50	60	70	80	90	100	110
Number of W Events	12146	6826	4070	2576	1696	1155	807	575	444
Number of Z Events	952	558	355	248	166	124	84	67	52
Raw value of Ratio	12.76	12.23	11.46	10.39	10.2	9.31	9.61	8.58	8.54
QCD Background Correction	0.85	0.86	0.86	0.88	0.87	0.87	0.87	0.86	0.85
Ewk Background Correction	0.96	0.96	0.95	0.95	0.96	0.95	0.95	0.95	0.95
Jet Correction	0.98	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.99
Acceptance Correction	0.78	0.78	0.78	0.78	0.78	0.79	0.79	0.79	0.80
ID Efficiency Correction	1.11	1.10	1.09	1.09	1.09	1.09	1.08	1.09	1.09
Trigger Efficiency Correction	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
Total Correction Factor	0.71	0.69	0.70	0.70	0.70	0.72	0.71	0.71	0.70
Measured value of R_{Jets}	9.01	8.48	7.98	7.27	7.17	6.67	6.82	6.05	5.99

Jet P_T Threshold [GeV]	120	130	140	150	160	170	180	190	
Number of W Events	330	259	196	155	128	101	91	83	
Number of Z Events	40	31	25	19	17	11	10	9	
Raw value of Ratio	8.25	8.35	7.84	8.16	7.53	9.18	9.10	9.22	
QCD Background Correction	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	
Ewk Background Correction	0.95	0.95	0.94	0.95	0.95	0.95	0.96	0.96	
Jet Correction	0.98	0.97	0.97	0.99	0.95	0.95	0.99	1.03	
Acceptance Correction	0.81	0.82	0.81	0.82	0.82	0.83	0.83	0.83	
ID Efficiency Correction	1.08	1.08	1.07	1.07	1.08	1.09	1.08	1.08	
Trigger Efficiency Correction	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	
Total Correction Factor	0.72	0.71	0.70	0.72	0.71	0.72	0.75	0.79	
Measured value of R_{Jets}	5.90	5.95	5.52	5.89	5.34	6.63	6.83	7.24	

Table 9.1: This table lists the raw numbers of W and Z events in each jet P_T bin as well as the raw value of the ratio and the values of the numerical factors used to correct the raw value to particle level. All of the corrections listed are multiplicative factors to be applied to the raw value of the ratio and consequently some of the values listed in the table are reciprocals of the values derived in the preceding chapters.

Jet P_T Threshold [GeV]	30	40	50	60	70	80	90	100	110
Theoretical Prediction	$8.58^{+0.14}_{-0.11}$	$7.87^{+0.23}_{-0.10}$	$7.43^{+0.21}_{-0.12}$	$7.11^{+0.24}_{-0.08}$	$6.85^{+0.20}_{-0.07}$	$6.76^{+0.23}_{-0.10}$	$6.56^{+0.21}_{-0.11}$	$6.56^{+0.26}_{-0.11}$	$6.53^{+0.26}_{-0.12}$
Measured Value	$9.01^{+0.50}_{-0.49}$	$8.48^{+0.53}_{-0.51}$	$7.98^{+0.58}_{-0.55}$	$7.27^{+0.61}_{-0.56}$	$7.17^{+0.71}_{-0.64}$	$6.67^{+0.75}_{-0.65}$	$6.82^{+0.92}_{-0.77}$	$6.05^{+0.93}_{-0.75}$	$5.99^{+1.06}_{-0.83}$
QCD Background Uncertainty [%]	0.63	0.73	0.83	1.01	1.20	1.46	1.73	2.44	3.11
Ewk Background Uncertainty [%]	1.48	1.40	1.43	1.24	1.38	1.92	2.45	2.50	0.85
Jet Correction Uncertainty [%]	2.17	1.93	1.92	2.15	1.30	1.78	1.53	1.41	1.81
Acceptance Correction Uncertainty [%]	3.08	3.34	3.29	3.04	2.79	2.47	2.16	2.03	1.85
ID Efficiency Uncertainty [%]	0.88	0.92	0.98	1.06	1.11	1.16	1.18	1.07	1.09
Trigger Efficiency Uncertainty [%]	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Total Systematic Uncertainty [%]	4.22	4.30	4.29	4.22	3.78	4.09	4.20	4.44	4.30
Generator Uncertainty [%]	0.80	0.58	0.71	1.40	2.58	1.48	1.10	1.10	2.62
Statistical Uncertainty [%]	$^{+3.48}_{-3.25}$	$^{+4.54}_{-4.24}$	$^{+5.82}_{-5.27}$	$^{+7.13}_{-6.28}$	$^{+8.80}_{-7.59}$	$^{+10.40}_{-8.70}$	$^{+12.80}_{-10.40}$	$^{+14.71}_{-11.54}$	$^{+16.88}_{-12.92}$
Total Uncertainty [%]	$^{+5.52}_{-5.38}$	$^{+6.28}_{-6.07}$	$^{+7.26}_{-6.83}$	$^{+8.40}_{-7.70}$	$^{+9.92}_{-8.86}$	$^{+11.27}_{-9.73}$	$^{+13.52}_{-11.27}$	$^{+15.41}_{-12.41}$	$^{+17.61}_{-13.87}$

Jet P_T Threshold [GeV]	120	130	140	150	160	170	180	190	
Theoretical Prediction	$6.44^{+0.33}_{-0.12}$	$6.31^{+0.32}_{-0.16}$	$6.24^{+0.32}_{-0.17}$	$6.35^{+0.41}_{-0.22}$	$6.12^{+0.31}_{-0.23}$	$6.17^{+0.42}_{-0.24}$	$5.93^{+0.30}_{-0.30}$	$5.82^{+0.31}_{-0.36}$	
Measured Value	$5.90^{+1.25}_{-0.98}$	$5.95^{+1.48}_{-1.12}$	$5.52^{+1.52}_{-1.09}$	$5.89^{+1.92}_{-1.32}$	$5.34^{+1.80}_{-1.20}$	$6.63^{+2.63}_{-1.73}$	$6.83^{+2.87}_{-1.87}$	$7.24^{+3.20}_{-2.17}$	
QCD Background Uncertainty [%]	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	
Ewk Background Uncertainty [%]	1.37	1.39	2.93	2.07	1.64	1.27	1.15	1.41	
Jet Correction Uncertainty [%]	2.05	2.39	2.68	6.12	4.05	4.96	3.66	4.39	
Acceptance Correction Uncertainty [%]	2.15	2.13	2.36	2.84	2.17	2.29	2.65	2.76	
ID Efficiency Uncertainty [%]	1.06	1.00	1.06	1.14	1.09	1.06	1.19	1.25	
Trigger Efficiency Uncertainty [%]	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	
Total Systematic Uncertainty [%]	3.56	3.74	4.83	7.21	5.08	5.78	4.90	5.59	
Generator Uncertainty [%]	6.97	8.67	6.72	6.90	5.80	3.10	5.93	1.08	
Statistical Uncertainty [%]	$^{+19.75}_{-14.58}$	$^{+23.03}_{-16.19}$	$^{+26.33}_{-17.98}$	$^{+31.01}_{-20.07}$	$^{+32.86}_{-21.02}$	$^{+39.16}_{-25.23}$	$^{+41.29}_{-26.21}$	$^{+42.51}_{-27.40}$	
Total Uncertainty [%]	$^{+21.24}_{-16.55}$	$^{+24.89}_{-18.74}$	$^{+27.60}_{-19.79}$	$^{+32.58}_{-22.41}$	$^{+33.75}_{-22.40}$	$^{+39.71}_{-26.07}$	$^{+42.00}_{-27.32}$	$^{+44.22}_{-29.99}$	

Table 9.2: The theoretical predictions and measured values of R_{Jets} for different leading jet transverse momentum thresholds as well as a breakdown of the contributing uncertainties.

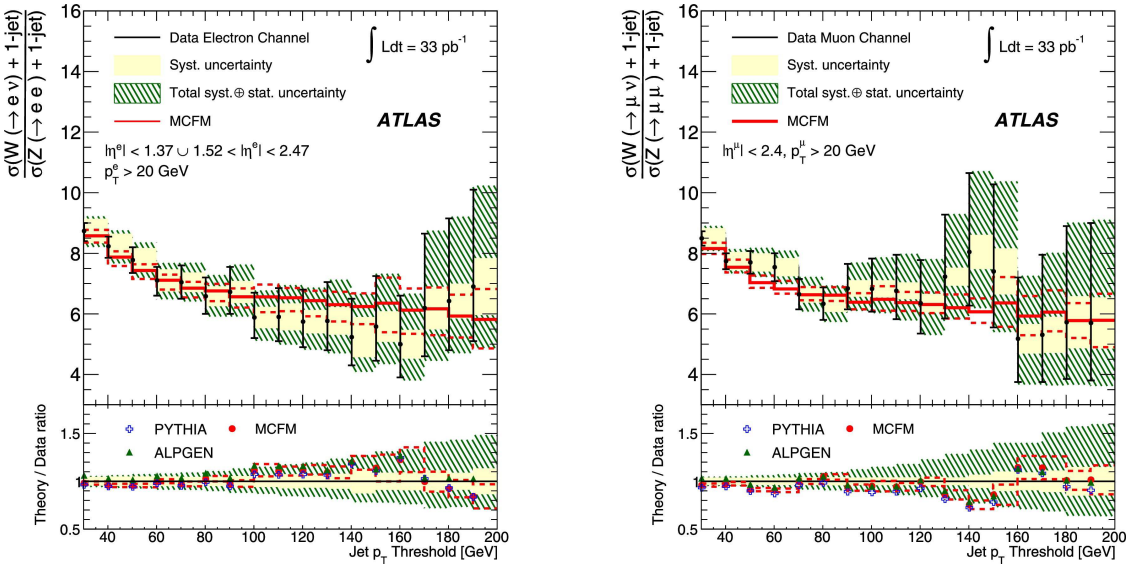


Figure 9.4: This figure shows the results of the official ATLAS measurements of R_{Jets} in the electron channel (left) and the muon channel (right) for their respective fiducial regions. The measured values are shown as black data points at the corresponding jet P_T threshold with a black error bar indicating the statistical uncertainty. The systematic uncertainty is shown as a yellow band around the measurement while the green band shows the total measurement uncertainty. The results are compared with predictions from MCFM, Pythia and ALPGEN.

9.3 Conclusions

The R_{Jets} measurement was motivated both by the desire to test the ability of QCD to make predictions at scales previously unexplored at collider experiments and also to provide a robust means for searching for new physics beyond the Standard Model in final states with leptons and jets. Other QCD measurements with similar goals have suffered from large systematic uncertainties related to the unfolding of jet quantities. Construction of the R_{Jets} observable as a ratio of cross sections for the production of W and Z vector bosons in association with jets subject to the same kinematic constraints leads to a significant suppression of such systematics.

In this work the R_{Jets} observable has been defined, a general procedure for measuring it using the ATLAS detector has been outlined and a series of computational tools have been developed for implementing this procedure, many of which are data-driven. The first such measurement has also been performed using 33pb^{-1} of integrated luminosity recorded by the ATLAS detector during the year 2010. In this measurement the vector bosons were

required to be produced in association with a single jet with above 30 GeV of transverse momentum and decay via the electron channel. The measurement was performed as a function of the threshold on the leading jet transverse momentum up to a value of 200 GeV. The results obtained were compared with the most precise theoretical predictions available, which were computed at Next-to-Leading Order in perturbative QCD with MCFM. Due to the cancellation of jet systematics described above, this measurement was the most precise performed by ATLAS during 2010 with a jet final state. The measurement is only systematically dominated for the lowest value of the jet transverse momentum threshold, 30 GeV, having a value of $9.01^{+0.50}_{-0.49}$ compared with a predicted value of $8.58^{+0.14}_{-0.11}$. The results of the measurement were observed to agree well with the predictions of fixed order pQCD over the entire range of values of the transverse momentum threshold.

The absence of a statistically significant deviation between the measured value of R_{Jets} and the predictions obtained from perturbative QCD mean this measurement shows no evidence of the existence of beyond Standard Model physics at the scales currently probed. Given the limited kinematic reach presently accessible this is unsurprising. In the context of using this measurement to aid in the understanding of QCD, due to the high degree of cancellation of the different theoretical uncertainties, the measurement uncertainties are still currently too large, even at low jet P_T, to place any significant constraints on the theory.

The availability of over 5fb^{-1} of data recorded in 2011, and the resulting factor ten reduction in the statistical uncertainties on the R_{Jets} measurement, will drastically change this situation, leading to the next incarnation of this measurement being systematically limited over almost the entire phase space considered here. This in turn allows an extension of the measurement to higher values of the jet transverse momentum threshold, where new physics may still be observable. The availability of a factor of over one hundred more data will also allow for a reduction of most of the systematic uncertainties. Improved jet and electron calibrations will result in smaller uncertainties on the unfolding corrections and the availability of improved tunes of Monte Carlo simulations to data will hopefully result in smaller modelling uncertainties. Of the tools developed in this work, the availability

of more data will allow for the improvement of both the data-driven $E_{\mathrm{T}}^{\mathrm{miss}}$ decomposition method for calculating the acceptances, which will be implemented in the next version of this analysis, and in both data-driven methods for performing the QCD background subtraction. One negative consequence of the increased luminosity in 2011 is that the amount of pile-up present in the data has drastically increased and in order to keep the systematic uncertainties related to this under control, more sophisticated tools for modelling pile-up in Monte Carlo will need to be developed [103].

The reduced uncertainties on future measurements of R_{Jets} will allow for a vastly improved scope both in terms of searching for new physics and understanding QCD. As well as repeating the particular measurement described here, R_{Jets} measurements can be performed with different jet multiplicities or with more specific criteria imposed on the jet final state, increasing the range of BSM phenomenologies to which the measurements are sensitive.

Appendix A

Theoretical Derivations

A.1 Derivation of Equation 2.13

We wish to determine the contributions to the Standard Model Lagrangian from the kinetic term in the Higgs sector which are bilinear in the gauge fields. To do this we begin by considering the action of the covariant derivative on the field ϕ:

$$D_\mu\phi = \partial_\mu\phi - ig_Y\frac{Y}{2}B_\mu\phi - ig_L T^a_{su(2)}W^a_\mu\phi \tag{A.1}$$

The Higgs doublet ϕ was introduced as an $SU(2)$ doublet with hypercharge $Y = 1$ so the generators T can be substituted for one half of the corresponding Pauli matrices:

$$D_\mu\phi = \partial_\mu\phi - i\frac{g_Y}{2}B_\mu\phi - i\frac{g_L}{2}\sigma^a W^a_\mu\phi \tag{A.2}$$

$$= \partial_\mu\begin{pmatrix}0\\ v+H\end{pmatrix} - i\begin{pmatrix}\frac{g_Y}{2}B_\mu + \frac{g_L}{2}W^3_\mu & \frac{g_L}{2}(W^1_\mu - iW^2_\mu)\\ \frac{g_L}{2}(W^1_\mu + iW^2_\mu) & \frac{g_Y}{2}B_\mu - \frac{g_L}{2}W^3_\mu\end{pmatrix}\begin{pmatrix}0\\ v+H\end{pmatrix} \tag{A.3}$$

$$\Rightarrow (D^\mu\phi)^\dagger(D_\mu\phi) = v^2(\frac{g_L}{2}(W^{1,\mu} + iW^{2,\mu}), \frac{1}{2}(g_Y B^\mu - g_L W^{3,\mu}))\begin{pmatrix}\frac{g_L}{2}(W^1_\mu - iW^2_\mu)\\ \frac{1}{2}(g_Y B_\mu - g_L W^3_\mu)\end{pmatrix} + \cdots \tag{A.4}$$

$$= \frac{v^2}{4}(g_L^2(W_\mu^1 W^{1,\mu} + W_\mu^2 W^{2,\mu}) + (g_Y B^\mu - g_L W^{3,\mu})(g_Y B_\mu - g_L W_\mu^3)) + \cdots \tag{A.5}$$

$$= \frac{v^2}{4}(g_L^2(W_\mu^1 W^{1,\mu} + W_\mu^2 W^{2,\mu}) + (B^\mu, W^{3,\mu}) \begin{pmatrix} g_Y^2 & -g_Y g_L \\ -g_Y g_L & g_L^2 \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}) + \cdots \tag{A.6}$$

which is equation 2.13. The ellipsis represent other terms not bilinear in the gauge fields.

A.2 Diagonalization of Equation 2.13

We now wish to diagonalise the real symmetric matrix of coefficients in the above equation in order to find the linear combinations of the fields B_μ and W_μ^3 which are mass eigenstates. In order to do this the characteristic equation for the matrix, denoted here by M, must be solved. It is therefore required that:

$$|\mathbf{M} - \lambda \mathbf{I}| = 0 \tag{A.7}$$

$$\Rightarrow | \begin{pmatrix} g_Y^2 - \lambda & -g_Y g_L \\ -g_Y g_L & g_L^2 - \lambda \end{pmatrix} | = 0 \tag{A.8}$$

$$\Rightarrow (g_Y^2 - \lambda)(g_L^2 - \lambda) - g_Y^2 g_L^2 = 0 \tag{A.9}$$

$$\Rightarrow \lambda(\lambda - (g_Y^2 + g_L^2)) = 0 \tag{A.10}$$

giving eigenvalues of 0 and $g_Y^2 + g_L^2$ as claimed. In order to relate the value of $tan\theta_W$ to the couplings it suffices to determine one of the eigenvectors. It is simplest to consider the zero eigenvalue for this which relates the photon field A_μ to the fields B_μ and W_μ^3.

$$\begin{pmatrix} g_Y^2 & -g_Y g_L \\ -g_Y g_L & g_L^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{A.11}$$

considering only the top row of this equation yields:

$$g_Y^2 x - g_Y g_L y = 0 \tag{A.12}$$

$$\Rightarrow \frac{y}{x} = \frac{g_Y}{g_L} = tan\theta_W \tag{A.13}$$

where the identification with $tan\theta_W$ comes from comparison of the above with equation 2.15.

A.3 Derivation of Acceptance Correction

This section outlines the construction of the form of the acceptance correction used in the R_{Jets} measurement. An alternative form is also derived which is equivalent, and the motivation for the particular choice made is discussed.

Suppose we wish to perform a fiducial measurement of a quantity S using three sets of selections, A, B and C. Implementing the measurement involves correcting the number of signal events passing the reconstruction level selections to the number of signal events passing the corresponding truth level selections. To do this we can write:

$$N_S(A_T, B_T, C_T) = [\frac{N_S(A_T, B_T, C_T)}{N_S(A_R, B_T, C_T)}][\frac{N_S(A_R, B_T, C_T)}{N_S(A_R, B_R, C_T)}][\frac{N_S(A_R, B_R, C_T)}{N_S(A_R, B_R, C_R)}]N_S(A_R, B_R, C_R) \tag{A.14}$$

$$= [\frac{A_S(A_T|B_T, C_T)N_S(B_T, C_T)}{A_S(A_R|B_T, C_T)N_S(B_T, C_T)}][\frac{A_S(B_T|A_R, C_T)N_S(A_R, C_T)}{A_S(B_R|A_R, C_T)N_S(A_R, C_T)}] \cdot [\frac{A_S(C_T|A_R, B_R)N_S(A_R, B_R)}{A_S(C_R|A_R, B_R)N_S(A_R, B_R)}]N_S(A_R, B_R, C_R) \tag{A.15}$$

$$= [\frac{A_S(A_T|B_T, C_T)}{A_S(A_R|B_T, C_T)}][\frac{A_S(B_T|A_R, C_T)}{A_S(B_R|A_R, C_T)}][\frac{A_S(C_T|A_R, B_R)}{A_S(C_R|A_R, B_R)}]N_S(A_R, B_R, C_R) \tag{A.16}$$

Equation A.16 is the form used for both the W and Z acceptance corrections in the R_{Jets} measurement. For R_{Jets} the event selections are only divided into two sets, A being the jet selections and B being all other geometrical and kinematic selections employed. An alternative correction scheme would be as follows:

$$N_S(A_T, B_T, C_T) = \frac{N_S(A_T, B_T, C_T)}{N_S(A_R, B_R, C_R)} N_S(A_R, B_R, C_R) \tag{A.17}$$

$$= \frac{A_S(A_T, B_T, C_T) N_S^{Total}}{A_S(A_R, B_R, C_R) N_S^{Total}} N_S(A_R, B_R, C_R) \tag{A.18}$$

$$= \frac{A_S(A_T, B_T, C_T)}{A_S(A_R, B_R, C_R)} N_S(A_R, B_R, C_R) \tag{A.19}$$

$$= \frac{A_S(A_T) A_S(B_T|A_T) A_S(C_T|A_T, B_T)}{A_S(A_R) A_S(B_R|A_R) A_S(C_R|A_R, B_R)} N_S(A_R, B_R, C_R) \tag{A.20}$$

$$= [\frac{A_S(A_T)}{A_S(A_R)}][\frac{A_S(B_T|A_T)}{A_S(B_R|A_R)}][\frac{A_S(C_T|A_T, B_T)}{A_S(C_R|A_R, B_R)}] N_S(A_R, B_R, C_R) \tag{A.21}$$

These forms are obviously equivalent and either could be used. The former was chosen as each term in the expansion only involves changing one set of selections from truth level to reconstruction level between numerator and denominator.

A.4 Derivation of Equation 8.5

Consider the quantity C which is given by the ratio of two quantities A and B. One can then consider statistical fluctuations about the average values of the three quantities as follows:

$$C = \frac{A}{B} \tag{A.22}$$

$$\Rightarrow dC = \frac{dA}{B} - \frac{AdB}{B^2} \tag{A.23}$$

$$\Rightarrow \frac{dC}{C} = \frac{dA}{A} - \frac{dB}{B} \tag{A.24}$$

$$\Rightarrow (\frac{dC}{C})^2 = (\frac{dA}{A})^2 + (\frac{dB}{B})^2 - \frac{dAdB}{AB} \tag{A.25}$$

$$\Rightarrow \langle(\frac{dC}{C})^2\rangle = \langle(\frac{dA}{A})^2\rangle + \langle(\frac{dB}{B})^2\rangle - \langle\frac{dAdB}{AB}\rangle \tag{A.26}$$

In the above equation the average squared fluctuation of each quantity, e.g. $\langle dC^2\rangle$, can be equated to the variance σ_C^2 of the corresponding quantity. The averages have no impact on the quantities A, B and C which are themselves average values. If A and B are independent then the quantity $\langle dAdB\rangle$ is equal to zero, however if this is not the case then one can define a non-zero correlation coefficient according to:

$$\rho = \frac{\langle dAdB\rangle}{\sigma_A \sigma_B} \tag{A.27}$$

Substituting this into equation A.26 gives equation 8.5:

$$(\frac{\sigma_C}{C})^2 = (\frac{\sigma_A}{A})^2 + (\frac{\sigma_B}{B})^2 - \frac{\sigma_A \sigma_B}{AB}\rho \tag{A.28}$$

Appendix B

Theoretical Systematics Support Material

B.1 Outline of Event Weighting Scheme

This section outlines how Monte Carlo integration software, such as MCFM, can be used to construct differential distributions of cross sections in terms of any given kinematic variable. Events are generated at random points in the phase space of interest according to a probability distribution chosen to optimise the resulting uncertainty. The generated events are assigned weights and these weights are then binned accordingly in order to obtain distributions of cross sections. In order to be free to make the choice of binning after the generation, the form of the weights must not depend on this choice in any way. The actual procedure implemented by MCFM is more complex than the one described below and employs several iterations to improve the statistical accuracy of the obtained results, but the basic premise is the same.

Suppose we generate N points within a phase space Ω according to a probability distribution $P(\omega)$ and we wish to calculate the cross section σ_{bin} in some given region Ω_{bin} of the total phase space. We then have that:

$$\sigma_{bin} = \int_{\Omega_{bin}} \sigma(\omega) d\omega \tag{B.1}$$

$$= \int_{\Omega_{bin}} P(\omega) \frac{\sigma(\omega)}{P(\omega)} d\omega \tag{B.2}$$

$$= \int_{\Omega_{bin}} P(\omega) W'(\omega) d\omega \tag{B.3}$$

where $W'(\omega) = \frac{\sigma(\omega)}{P(\omega)}$. $P(\omega)$ is the probability distribution used to sample points over the entire phase space Ω, this can be normalized by its integral over Ω_{bin} to find the probability distribution for events within the bin of interest, P_{bin}:

$$P_{bin}(\omega) = \frac{P(\omega)}{\int_{\Omega_{bin}} P(\omega') d\omega'} \tag{B.4}$$

$$\Rightarrow P(\omega) = P_{bin}(\omega) \int_{\Omega_{bin}} P(\omega') d\omega' \tag{B.5}$$

$$\Rightarrow \sigma_{bin} = \int_{\Omega_{bin}} P_{bin}(\omega) W'(\omega) d\omega \int_{\Omega_{bin}} P(\omega') d\omega' \tag{B.6}$$

$$= \langle W'(\omega) \rangle_{\Omega_{bin}}^{P_{bin}} \int_{\Omega_{bin}} P(\omega) d\omega \tag{B.7}$$

where $\langle W'(\omega) \rangle_{\Omega_{bin}}^{P_{bin}}$ is the average value of W' over the phase space Ω_{bin} according to the probability distribution $P_{bin}(\omega)$. The two terms in the above expression can now be approximated as follows:

- $\langle W'(\omega) \rangle_{\Omega_{bin}}^{P_{bin}} \approx \frac{1}{N_{bin}} \Sigma_i^{(P_{bin})} W'_i$

- $\int_{\Omega_{bin}} P(\omega) d\omega \approx \frac{N_{bin}}{N}$

where N_{bin} is the number of randomly generated events which fall within the phase space Ω_{bin}. The superscript P_{bin} on the summation is to indicate that the sample points are to be generated across the phase space Ω_{bin} according to the probability distribution P_{bin}. Inserting

these approximations into equation B.7 gives:

$$\sigma_{bin} = \frac{N_{bin}}{N} \cdot \frac{1}{N_{bin}} \Sigma_i^{P_{bin}} W_i' = \Sigma_i^{P_{bin}} \frac{W_i'}{N} \tag{B.8}$$

$$= \Sigma_i^{P} W_i \tag{B.9}$$

where $W_i = \frac{\sigma_i}{P_i N}$ and the summation is over all events i within the phase space region Ω_{bin}. In this way a set of sample points i in phase space can be chosen and weights can be evaluated for them based on the differential cross section at the point σ_i, the value of the sampling probability distribution P_i and the number of events generated. To generate a differential distribution in terms of some chosen kinematic variable one simply chooses a binning and sums the weights of all events which fall within a particular bin in order to determine the cross section for that bin.

MCFM calculates the final probability distribution to use when generating events iteratively. In each step the phase space is coarsely populated with points and cross section distributions are determined as above. For the following step, the phase space is more finely divided and the normalised modulus squared of the cross section from the previous iteration is used as the input probability distribution for the current one.

B.2 MCFM Samples

The following table lists the values of the Standard Model parameters required in the generation of the MCFM samples used in chapter 5. Electroweak scheme 1 was employed, whereby the values of the strong coupling at the Z pole and the Weinberg angle are derived from the other input parameters.

The samples were generated over a larger phase space than that defined in table 4.2 in order to accommodate changing selection criteria as well as to have a generator level

Quantity	Value	Quantity	Value
G_F	1.6637×10^{-5}	M_T (Top Mass)	172.5 GeV
M_W	80.403 GeV	M_Z	91.1876 GeV
Γ_W	2.0910 GeV	Γ_Z	2.4952 GeV
$sin(\theta_W)$	Calculated	$\alpha(M_Z)$	Calculated
M_C (Charm Mass)	1.29 GeV	M_B (Bottom Mass)	4.2 GeV
M_τ (Tau Mass)	1.777 GeV	Γ_τ (Tau Width)	2.269×10^{-12}
V_{ud} (CKM element)	0.9738	V_{us} (CKM element)	0.2272
V_{ub} (CKM element)	0.0040	V_{cd} (CKM element)	0.2271
V_{cs} (CKM element)	0.9730	V_{cb} (CKM element)	0.0422

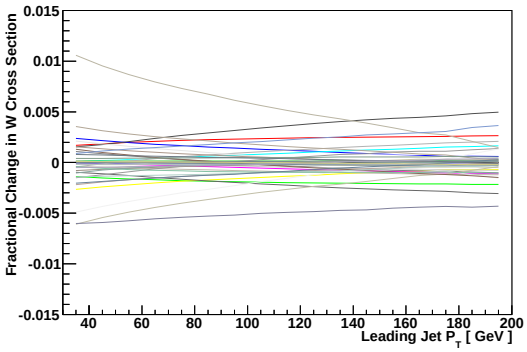
Table B.1: Values of the required Standard Model parameters used in the generation of the MCFM samples.

jet threshold somewhat lower than the analysis level jet threshold, reducing the impact of potential edge effects. Consequently, the generator jet transverse momentum threshold was chosen to be 20 GeV and the jet pseudorapidity threshold was chosen to be 6.

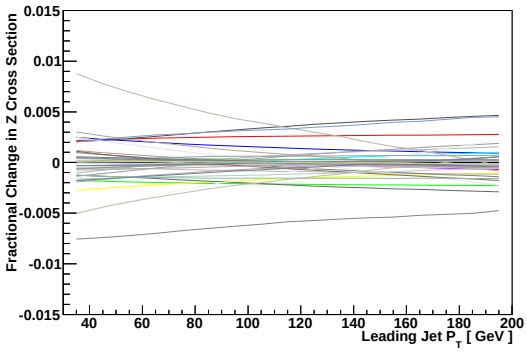
Twelve separate runs were performed for the nominal configuration while only five runs were performed for each of the scale and strong coupling variation samples. Each of these runs consisted of ten pre-conditioning iterations with 250k sample points each and a further ten final iterations using 500k sample points. For the PDF errors, a single run was performed with 10 pre-conditioning iterations each with 250k sample points, and 10 final stage iterations each with 750k sample points. MCFM version 5.8 was used throughout. Full details of the generated samples, their configuration and how to run analyses over them can be found here [104].

B.3 MSTW PDF Error Set Results

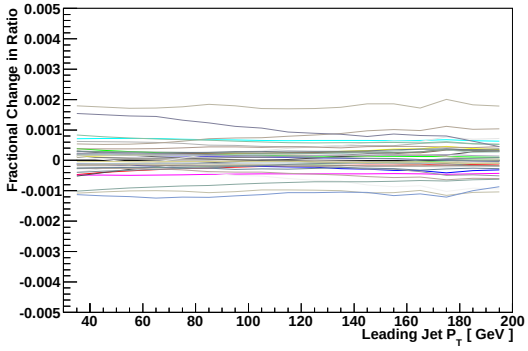
Here are the plots corresponding to figure 5.5 obtained using MSTW 2008.



(a) $W + 1$ Jet



(b) $Z + 1$ Jet



(c) R_{Jets}

Figure B.1: Relative deviations of the W and Z cross sections and their ratio from those obtained using the global best fit when using different MSTW2008 error PDF sets. Note the difference in the scale of the ordinate.

Appendix C

Data Driven $E_{\mathrm{T}}^{\mathrm{miss}}$ Acceptance Support Material

C.1 Event Selections

Table C.1 lists the event selections used for the studies related to the data-driven $E_{\mathrm{T}}^{\mathrm{miss}}$ decomposition method. These selections are different from those used in the main analyses reflecting instead the standard choices made for 2011 analyses.

C.2 1 Jet Results

The following plots show the 1 jet results corresponding to those shown in the main text for the inclusive case. Note that the plots were made by requiring there be at least one jet satisfying the requirements as opposed to requiring exactly one jet as was the case for the R_{Jets} analysis.

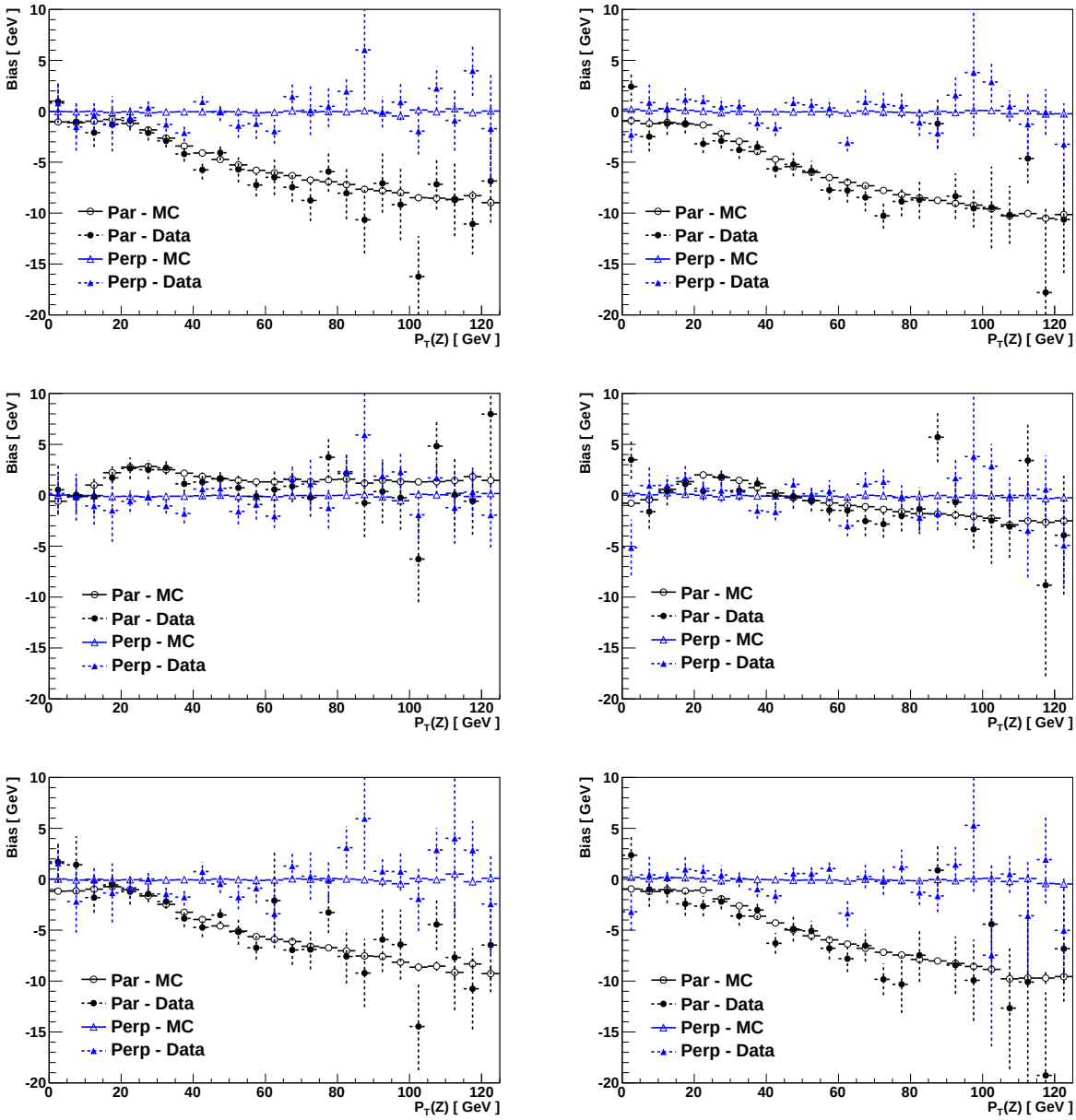


Figure C.1: $E_{\mathrm{T}}^{\mathrm{miss}}$ scale bias (mean of $\not{E}_{\parallel}$) as a function of the ZP_T in the inclusive 1 jet case obtained using `LocHadTopo` , `RefFinal` and `HadRecoil` for the three rows respectively. The left hand side shows results obtained in the electron channel, the right hand side shows results obtained in the muon channel.

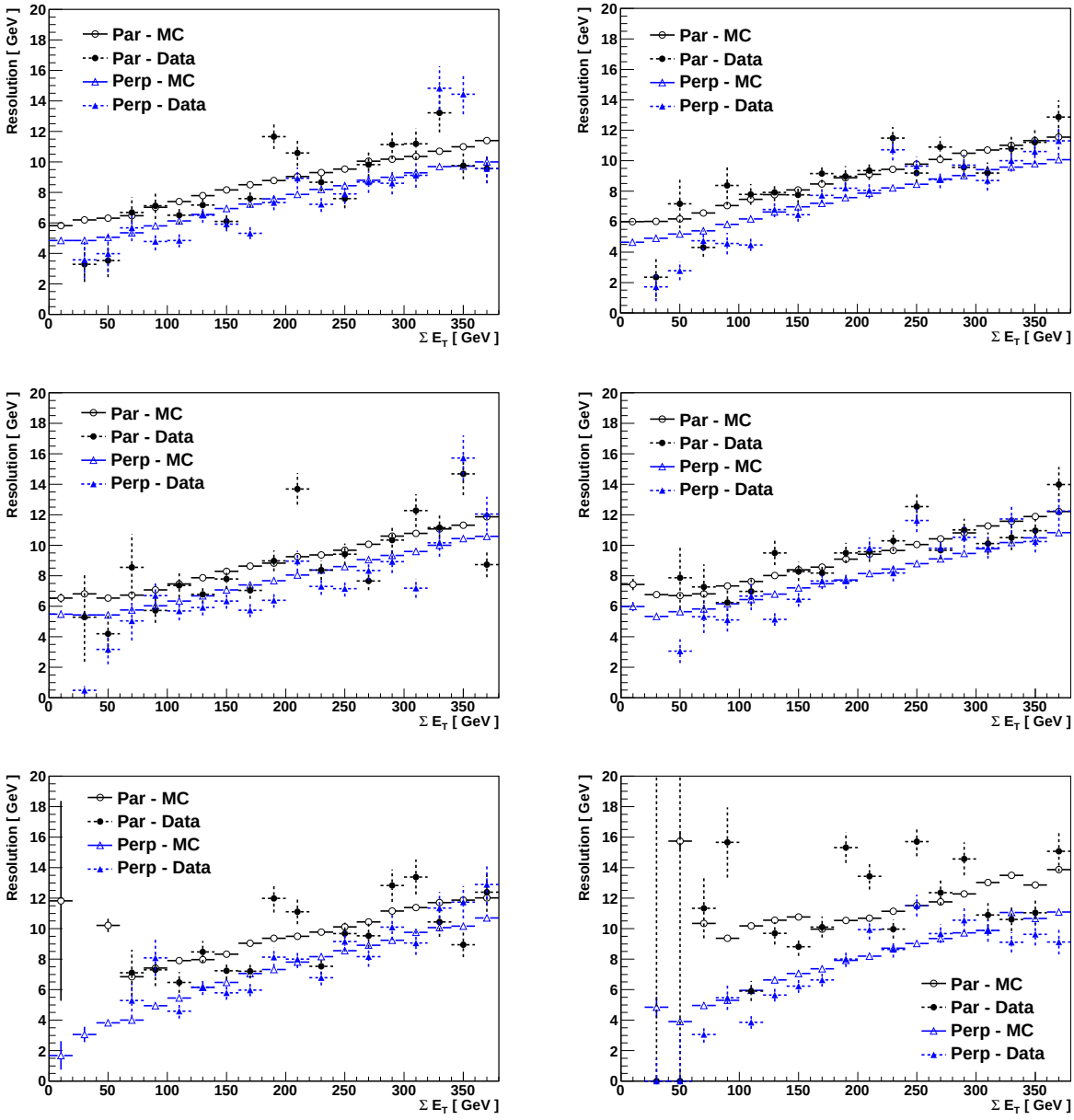


Figure C.2: $E_{\mathrm{T}}^{\mathrm{miss}}$ resolutions in the inclusive 1 jet case (widths of $\not{E}_{\parallel}$ and $\not{E}_{\perp}$) as functions of the ΣE_T obtained using `LocHadTopo` , `RefFinal` and `HadRecoil` for the three rows respectively. The left hand side shows results obtained in the electron channel, the right hand side shows results obtained in the muon channel.

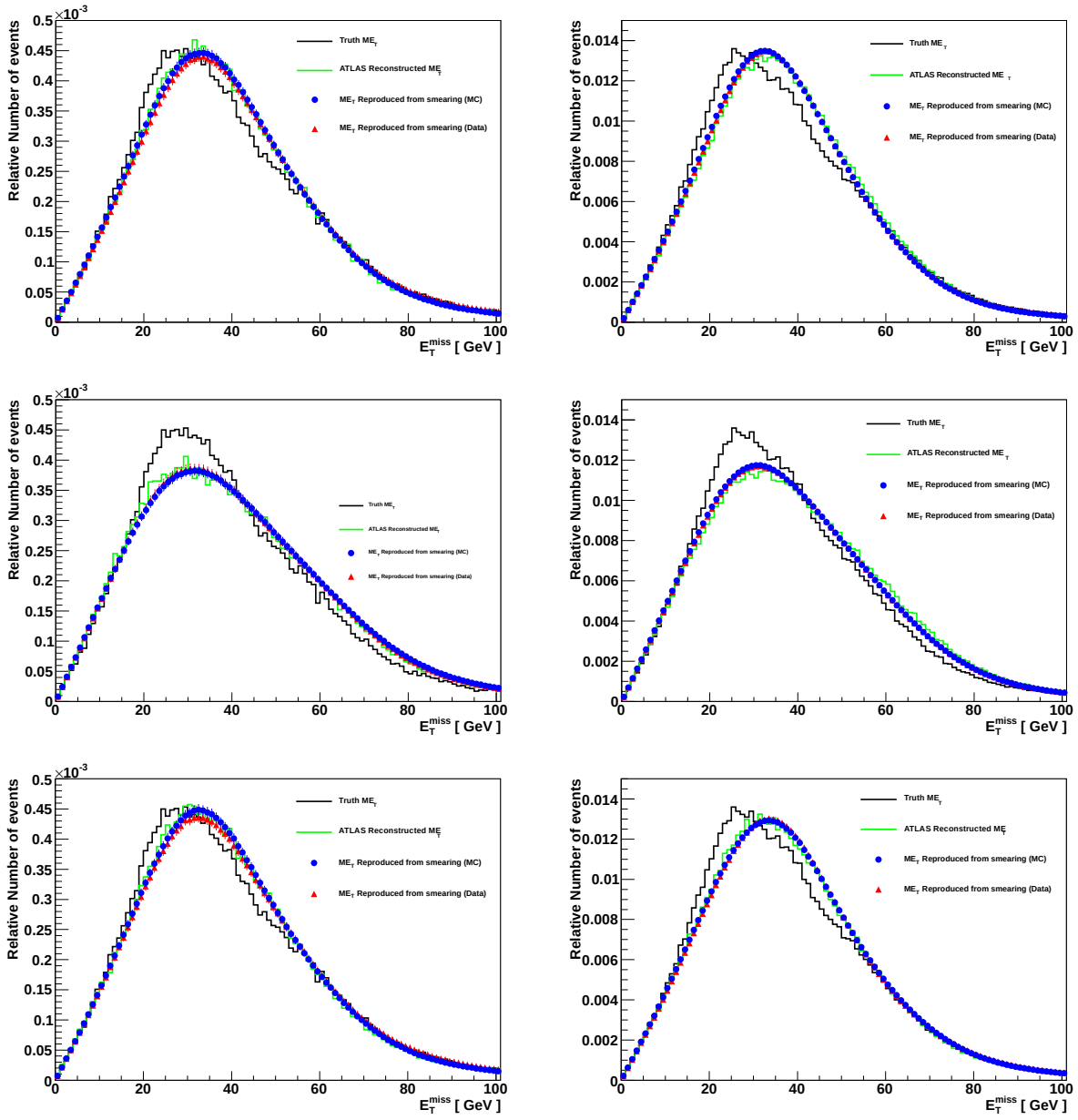


Figure C.3: Comparison of the standard ATLAS reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution obtained from the GEANT4 detector simulation (green) with the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions obtained using the $E_{\mathrm{T}}^{\mathrm{miss}}$ decomposition method described in the text (Blue - Resolution functions obtained from MC, Red - Resolution functions obtained from data) in the inclusive 1 jet case. The left hand side shows results obtained in the electron channel, the right hand side shows results obtained in the muon channel. The three rows show `LocHadTopo` , `RefFinal` and `HadRecoil` respectively.

General Selections	
Events pass Common WZ Good Run List	
Primary Vertex Selection : N_{vtx} >1 with N_{trk} >3	
Jet Cleaning: Reject events with bad or ugly jets	
Jet Def: AntiKT4 Jets, P_T^{EMJES} >20 GeV, $\|\eta\|$ <2.8, $\Delta R_{\ell,j}$ >0.5	
Z Event Selections	
Two medium electrons or two good combined muons with opposite charge	
Invariant Mass: 66< M_Z <116 GeV	
W Event Selections	
One Tight electron or good combined muon	
No second Medium electron or good combined muon	
Electron	**Muon**
Trigger Requirements	
Periods A-E3: L1_EM14 Periods E4-I2: EF_e15_Medium	Periods A-E3: L1_MU10 Periods E4-G1: EF_mu10_MG Periods G2-I1: EF_mu13_MG Period I2: EF_mu13_MG_TIGHT
$E_T = E_{cluster}/cosh(\eta_{trk}) > 20$ GeV $\|\eta\| < 2.47$, Crack ($1.37 < \|\eta\| < 1.52$) excluded Author = 1 || Author = 3 etcone20<4 GeV Passes Dead OTX Selections	P_T^μ >20 GeV $\|\eta\| < 2.4$ Author = 6 Ptcone20/$P_T^\mu < 0.1$ Z Extrapolation $\|z_{ext} - z_{PV}\|$ <10 mm Standard Track Quality Requirements

Table C.1: A summary of the selections used in the data-driven acceptance study.

Bibliography

[1] The ATLAS Collaboration. ATLAS public results - luminosity, 2011. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`.

[2] R. Apolle. *Thesis in Preparation.* PhD thesis, University of Oxford, 2012.

[3] K. Nakamura et al. The review of particle physics. *Journal of Physics G*, 37, 2010.

[4] The CDF Collaboration. Measurement of the cross section for W-boson production in association with jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physical Review D*, 77:011108, 2008.

[5] The CDF Collaboration. Measurement of inclusive jet cross sections in $Z/gamma*(\rightarrow e^+e^-)$+jets production in p anti-p collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. Lett.*, 100:102001, 2008.

[6] The D0 Collaboration. Measurements of inclusive W+jets production rates as a function of jet transverse momentum in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Lett. B.*, 705:200, 2011.

[7] The D0 Collaboration. Measurements of differential $Z/gamma*+jet+X$ cross sections with the D0 detector. *Phys. Lett. B.*, 669:278, 2008.

[8] The ATLAS Collaboration. Measurement of the production cross section for W-bosons in association with jets in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector. *Phys. Lett. B.*, 698:325–345, 2011.

[9] The ATLAS Collaboration. Measurement of the production cross section for $z/gamma*$ in association with jets in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector. *Phys. Rev. D.*, 85:032009, 2011.

[10] E. Abouzaid and H. Frisch. The ratio of $W + N$ jets to $Z^0/\gamma^* + N$ jets versus N as a precision test of the Standard Model. *Phys. Rev. D.*, 68:033014, 2003.

[11] G. Aad et al. A measurement of the ratio of the W and Z cross sections with exactly one associated jet in pp collisions at $\sqrt{S} = 7$ TeV with ATLAS. *arXiv:1108.4908v1.*

[12] MissMJ. Standard model of elementary particles, 2006. `http://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`.

[13] C.M. Becchi and G. Ridolfi. *An Introduction to Relativistic Processes and the Standard Model of Electroweak Interactions*. Springer, 2006.

[14] E. Fermi. Quantum theory of radiation. *Rev. Mod. Phys.*, 4:87–132, 1932.

[15] W. Pauli. Relativistic field theories of elementary particles. *Rev. Mod. Phys.*, 13:203–232, 1941.

[16] J. Schwinger. Quantum electrodynamics. I. A covariant formulation. *Physical Review*, 74:1439–1461, 1948.

[17] R. Feynman. Mathematical formulation of the quantum theory of electromagnetic interaction. *Physical Review*, 80:440–457, 1950.

[18] C.N. Yang and R. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Physical Review*, 96:191–195, 1954.

[19] G. t'Hooft. Renormalizable Lagrangians for massive Yang-Mills fields. *Nuclear Physics B*, 35:167–448, 1971.

[20] C.S. Wu et al. Experimental test of parity conservation in beta decay. *Physical Review*, 105:1413–1415, 1957.

[21] R.L. Garwin et al. Observations of the failure of conservation of parity and charge conjugation in meson decays: The magnetic moment of the free muon. *Physical Review*, 105:1415–1417, 1957.

[22] G. Arnison et al. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV. *Phys. Lett.*, 122:103–116, 1983.

[23] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 GeV at the CERN SPS collider. *Phys. Lett.*, 126:398–410, 1983.

[24] P.W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Letters*, 13:508–509, 1964.

[25] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Letters*, 13:321–323, 1964.

[26] C.R. Hagen et al. Global conservation laws and massless particles. *Phys. Rev. Letters*, 13:585–587, 1964.

[27] M. Kobayashi and T. Maskawa. CP-violation in the renormalizable theory of weak interactions. *Progress of Theoretical Physics*, 42:652–657, 1973.

[28] K. Nakamura et al. Review of particle physics: The CKM quark-mixing matrix. *Journal of Physics G*, 37:150, 2010.

[29] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650, 1962.

[30] T.D. Lee and M. Nauenberg. Degenerate systems and mass singularities. *Phys. Rev.*, 133:1549, 1964.

[31] F. Bloch and A. Nordsieck. Note on the radiation field of the electron. *Phys. Rev.*, 52:54–59, 1937.

[32] R. Devenish and A. Cooper-Sarkar. *Deep Inelastic Scattering*. Oxford University Press, USA, 2004.

[33] J.J. Aubert et al. Experimental observation of a heavy particle J. *Phys. Rev. Lett.*, 33:1404–1406, 1974.

[34] J. Augustin et al. Discovery of a narrow resonance in e^+ e^- annihilation. *Phys. Rev. Lett.*, 33:1406, 1974.

[35] S.L. Glashow et al. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D.*, 2:1285, 1970.

[36] M. Kobayashi and T. Maskawa. CP-violation in the renormalizable theory of the weak interaction. *Progress of Theoretical Physics*, 49:652, 1973.

[37] S.W. Herb et al. Observation of a dimuon resonance at 9.5 GeV in 400-GeV proton-nucleus collisions. *Phys. Rev. Lett.*, 39:252, 1977.

[38] F. Abe et al. Observation of top quark production in pp collisions with the collider detector at Fermilab. *Phys. Rev. Lett.*, 74:2626–2631, 1995.

[39] S. Abachi et al. Search for high mass top quark production in pp collisions at $\sqrt{s} =$ 1.8 TeV. *Phys. Rev. Lett.*, 74:2422–2426, 1995.

[40] F.J. Hasert et al. Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment. *Nucl. Phys. B.*, 73:1, 1974.

[41] M. Banner et al. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $p\bar{p}$ collider. *Phys. Lett.*, 122:476–485, 1983.

[42] P. Bagnaia et al. Evidence for $z^0 \rightarrow e^+e^-$ at the CERN $p\bar{p}$ collider. *Phys. Lett.*, 129:130–140, 1983.

[43] The ALEPH Collaboration. Measurement of the W mass in e^+e^- collisions at 183 GeV. *Phys. Lett. B.*, 453:121–137, 1999.

[44] The ALEPH Collaboration. Measurement of α_s from scaling violations in fragmentation functions in e^+e^- annihilation. *Phys. Lett. B.*, 357:487–499, 1995.

[45] The L3 Collaboration. Determination of the number of light neutrino species from single photon production at LEP. *Phys. Lett. B.*, 431:199–208, 1998.

[46] The Babar Collaboration. Observation of CP violation in the B^0 meson system. *Phys. Rev. Lett.*, 87:091801, 2001.

[47] The Belle Collaboration. Observation of large CP violation in the neutral B meson system. *Phys. Rev. Lett.*, 87:091802, 2001.

[48] S.P. Martin. A supersymmetry primer. *arXiv:hep-ph/9709356v6*, 1997.

[49] J.C. Pati and A. Salam. Unified lepton-hadron symmetry and a gauge theory of the basic interactions. *Phys.Rev.D*, 8:1240–1251, 21972.

[50] H. Georgi and S. Glashow. Unity of all elementary-particle forces. *Phys.Rev.Lett.*, 32:438, 21974.

[51] A.D. Martin et al. Parton distributions for the LHC. *The European Physical Journal C*, 63:189–285, 2009.

[52] Close F.E. *Introduction to Quarks and Partons*. Academic Press Inc. New York, 1979.

[53] R.K. Ellis et al. *QCD and Collider Physics*. Cambridge, 1996.

[54] The ATLAS Collaboration. Measurement of the $W \to \ell\nu$ and $Z/\gamma^* \to \ell\ell$ production cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 1012:060, 2010.

[55] L. Evans and P. Bryant. LHC machine. *Journal of Instrumentation*, 3:S08001, 2008.

[56] The European Organisation for Nuclear Research. CERN Accelerator Complex website. `http://public.web.cern.ch/public/en/research/AccelComplex-en.html`.

[57] The ATLAS Collaboration. *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics*, volume CERN-OPEN-2008-020. CERN, 2009.

[58] G. Aad et al. The ATLAS experiment at the Large Hadron Collider. *Journal of Instrumentation*, 3:S08003, 2008.

[59] The ATLAS Collaboration. Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data. *Eur. Phys. J. C.*, 2011.

[60] M. Cacciari et al. The anti-kt jet clustering algorithm. *JHEP*, 4, 2008.

[61] S. Ellis and D. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D.*, 48:3160–3166, 1993.

[62] G. Salam and Soyez. G. A practical seedless infrared-safe cone jet algorithm. *JHEP*, 5, 2007.

[63] S. Bentvelsen and I. Meyer. The Cambridge jet algorithm: features and applications. *Eur. Phys. J. C.*, 4:623–629, 1998.

[64] The ATLAS Collaboration. Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector. *Eur. Phys. J. C.*, 71:1512, 2011.

[65] The ATLAS Collaboration. Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ tev with atlas 2010 data. *ATLAS-CONF-2011-032*, 2011.

[66] C. Doglioni. *Measurement of the inclusive jet cross section with the ATLAS detector at the Large Hadron Collider*. PhD thesis, Merton College, University of Oxford, 2011.

[67] The ATLAS Collaboration. Measurement of the $W \to \ell\nu$ and $Z/\gamma* \to \ell\ell$ production cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 12:1–65, 2010.

[68] The ATLAS Collaboration. Measurement of the transverse momentum distribution of W bosons in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *arXiv:1108.6308v1*, page 64, 2011.

[69] The ATLAS Collaboration. Performance of missing transverse momentum reconstruction in proton-proton collisions at 7 TeV with ATLAS. *arXiv:1108.5602v1*, page 33, 2011.

[70] S. Whitehead. *A Measurement of the W Boson Charge Asymmetry with the ATLAS Detector*. PhD thesis, Somerville College, University of Oxford, 2012.

[71] The ATLAS Collaboration. Luminosity determination in pp collisions at sqrt(s)=7 TeV using the ATLAS detector at the LHC. *Eur. Phys. J. C.*, 71:1630, 2011.

[72] S. Mrenna T. Sjöstrand and P. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.

[73] M. Mangano et al. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 0307:001, 2003.

[74] J.M. Butterworth et al. Multiparton interactions in photoproduction at HERA. *Z. Phys.*, C72:637, 1996.

[75] G. Corcella et al. HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes) . *Journal of High Enery Physics*, 01, 2001.

[76] S. Frixione et al. Matching NLO QCD computations with parton shower simulations:the POWHEG method. *JHEP*, 11:070, 2007.

[77] G. Aad et al. The ATLAS simulation infrastructure. *arXiv:1005.4568*.

[78] M. Mangano et. al. Matching matrix elements and shower evolution for top pair production in hadronic collisions. *Journal of High Enery Physics*, 013, 2007.

[79] J. Campbell et al. Next-to-leading order QCD predictions for W+2 jet and Z+2 jet production at the CERN LHC. *Physical Review D*, 68(9):094021, 2003.

[80] S. Catani and M.H. Seymour. A general algorithm for calculating jet cross-sections in NLO QCD. *Nucl. Phys. B.*, 485:291–419, 1997.

[81] John Campbell. Private communication, 2010.

[82] G. Altarelli and G. Parisi. Aymptotic freedom in parton language. *Nucl. Phys. B*, 126:298, 1977.

[83] Y.L. Dokshitzer. Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.

[84] V.N. Gribov and L.N. Lipatov. Deep inelastic *ep* scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.

[85] L.N. Lipatov. The parton model and perturbation theory. *Sov. J. Nucl. Phys.*, 20:94–102, 1975.

[86] T. Carli et al. A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project. *The European Physical Journal C-Particles and Fields*, 66(3):503–524, 2010.

[87] C. Berger et al. Next-to-leading order QCD predictions for W + 3-jet distributions at hadron colliders. *Phys. Rev. D*, 80:074036, 2009.

[88] C. Berger et al. Next-to-leading order QCD predictions for Z/γ^* + 3-jet distributions at the Tevatron. *arXiv:1004.1659 [hep-ph]*.

[89] Standard Model Cross Section Task Force. Private communication, 2009.

[90] R.D. Ball et al. A first unbiased global NLO determination of parton distributions and their uncertainties. *arxiv:1002.4407*, 2010.

[91] J. Pumplin et al. Uncertainties of predictions from parton distribution functions. II. The Hessian method. *Phys. Rev. D*, 65:014013, 2001.

[92] J. Huston et al. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

[93] L. Hung-Liang et al. Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions. *arXiv:1004.4624v3 [hep-ph]*.

[94] G. Aad et al. Charged particle multiplicities in pp interactions at $\sqrt{s} = 0.9$ and 7 TeV in a diffractive limited phase space measured with the ATLAS detector at the LHC and a new PYTHIA6 tune. *ATLAS-CONF-2010-031*.

[95] R. Barlow and C. Beeston. Fitting using finite Monte Carlo samples. *Computer Physics Communications*, 77(2):219–228, 1993.

[96] R. Brun et al. ROOT website, 1995-2011. `http://root.cern.ch/drupal/`.

[97] The ATLAS Metadata Interface. `http://atlastagcollector.in2p3.fr/AMI/`.

[98] E. Berglund. *Cross Section of W and Z inclusive production and the W / Z cross section ratio in association with one jet for the boson decaying into electron(s) with ATLAS 7 TeV data*. PhD thesis, Faculte des Sciences, Universite de Geneve, 2011.

[99] List. B. Statistical error of efficiency determination from weighted events. 1999. https://www.desy.de/~blist/.

[100] G. Aad et al. Jet energy resolution and reconstruction efficiencies from in-situ techniques with the ATLAS detector in proton-proton collisions at a center-of-mass energy of 7TeV. *ATLAS-CONF-2010-054*.

[101] G. Aad et al. Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at $\sqrt{s} = 7$ TeV and measured with the ATLAS detector. *ATLAS-CONF-2010-056*.

[102] E. Dobson. *A feasibility study into the measurement of W and Z cross sections with the ATLAS detector*. PhD thesis, Linacre College, University of Oxford, 2009.

[103] D. Short. *Searching for strongly produced supersymmetry in high luminosity events containing leptons, jets and missing transverse momentum using kinematic shape constraints with the ATLAS detector and monitoring of the SCT optical links*. PhD thesis, University of Oxford, 2012.

[104] MCFM for ATLAS Twiki. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MCFMForAtlas.