A First Measurement of Electroweak Production of a W Boson in Association with Two Jets with the ATLAS Detector

Robert King

Magdalen College, Oxford

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

University of Oxford, Trinity Term 2013

A First Measurement of Electroweak Production of a *W* Boson in Association with Two Jets with the ATLAS Detector

Robert King
Magdalen College, Oxford

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

July 2013

Abstract

A first measurement of Electroweak W + 2 jets production at high dijet mass is performed using $\sqrt{s} = 7$ TeV pp collision data from the ATLAS experiment corresponding to 4.6 fb^{-1} of integrated luminosity. The background only hypothesis is excluded with a significance of 4.89-σ. A cross section of $\sigma = 325 \pm 6$ (lumi) ± 32 (stat) $^{+63}_{-70}$ (syst) ± 86 (theo) fb is extracted in the fiducial region.

Acknowledgements

I would like to thank the STFC for funding me for 3.5 years and providing funding for a two year stay at CERN and travel to several useful and interesting workshops and meetings around the world. Thanks also to Magdalen College whose generous travel fund enabled me to attend to the CERN-Fermi Lab Hadron Collider Symposium.

I have been fortunate to be surrounded by knowledgeable, helpful and friendly people throughout my D.Phil. Thanks must first go to my supervisor Chris Hays. His broad knowledge and experience has been an inspiration. His enthusiasm and dedication at times daunting. Thanks for all your support.

At Oxford I would like to thank the ever helpful secretarial staff for making light work of the bureaucracy, and the computing team for ensuring efficient functioning of the local batch system while I chewed my way through 18 years worth of CPU time. I am also grateful to Richard Nickerson for teaching me what a real lab should look like - full of junk and covered in duct tape.

Within the wider Atlas collaboration too many to mention have given me their time and understanding, without which I would never have got started. Particular mention should go to Dag Gillberg and Jim Lacey from the jet intercalibration days; and Gerhard Brandt, Hugo Beauchemin, Alessandro Tricoli and Craig Sawyer from the Rjets team.

On the VBF side Gerhard (again), Sabine Lammers and Sinead Farrington provided assistance and encouragement.

The Sherpa authors have always been particularly helpful, in particular I would like to thank Frank Krauss, Frank Siegert, Jan Winter and Hendrik Hoeth for giving generously of their time. Their assistance was vital in producing the Sherpa signal Monte Carlo sample.

Most of all I would like to thank my fellow Atlas students who formed the L'Athénée Skype chat - quicker than google or twiki search 98% of the time. Will, Nick, Sarah, Ellie, Stew, JRob, Caterina, Jack, Chris and Adrian: I couldn't have done it without you!

At times I needed to get away from the physics chat, and while at CERN I was very glad to find Arno, Dan, Georg and Harri who shared my love of mountains and somehow put up with my skiing and ROOT rants. Back in Oxford the OUMC provided frequent drinking and occasional climbing distractions. Except for the rowing chat I couldn't have asked for better house mates. I'm going to miss cycling with the Manor Farm High Performance Center!

Final thanks to my family for getting me here, and to Lauren for waiting for a very long '6 month' visit to CERN and playing 'spot the typo' for more weekends than I care to remember.

Preface

The work presented in this thesis was carried out within the ATLAS collaboration; a group of over 3000 physicists and engineers who together built and ran the ATLAS detector. As a result much of the material presented in this thesis is dependent on the work of others. I have tried to make this clear in the text by citing published papers where possible, public reports where no paper yet exists, or internal notes or documentation pages where no other reference is available. Unfortunately these pages are not accessible to non-members of the collaboration, but I hope that citing something is better than citing nothing at all. In order to make my personal contribution clear I list below my major contributions by chapter:

Chapter 1 I performed necessary cross-checks and requests for the VBF signal Monte Carlo samples, described in Section 4.2.1.

Chapter 3 I produced an independent cross-check of the dijet balance jet calibration results, as described in Section 3.4.1. I evaluated p_T^{avg} shape systematic uncertainties and various technical issues such as pileup-reweighting templates for all 20 triggers.

Chapter 4 I designed and optimized the VBF aspects of the selection from Section 4.4 and onwards.

Chapter 5 I adapted the QCD background estimation method in the electron channel previously used by the 2010 $W+\text{jets}$ analysis for the 2011 $W+\text{jets}$ and W/Z ratio jets anayses. I further modified this method for the VBF phase space. In the muon channel the 2011 QCD background for $W+\text{jets}$ and W/Z ratio jet was derived by another member of the collaboration; I modified it for the VBF phase space.

Chapter 6 I designed and carried out the signal extraction, background rejection and cross section measurements presented in this chapter.

In addition to the material presented in this thesis I qualified as an Atlas author in 2010 with work on a hardware project, measuring the thermo-mechanical stability of inner detector components for the Atlas upgrade. During my long-term attachment to CERN I did shifts on the trigger desk, monitoring the event recording-rate and data-quality.

Figures

All figures are the work of the author unless cited in the caption. In particular plots with the label 'ATLAS' or 'ATLAS preliminary' are official paper or conference results from the Atlas collaboration. Further details of these results will be found in the reference cited. All Feynman diagrams were created by the author using JaxoDraw [1].

Units

As is customary in high energy physics this thesis uses natural units throughout with $c = \hbar = 1$. Thus masses and momenta are given in units of energy. Energies are generally given in multiples of electron volts (eV).

Contents

Introduction

One of the principal goals of the Large Hadron Collider (LHC) is a detailed investigation of all aspects of the electroweak model including the search for a mechanism of electroweak symmetry breaking. As part of this program it is important that all accessible elements of the model are measured.

The purpose of this thesis is to present the first measurement of the electroweak production of a W boson in association with two hadronic jets. This contains contributions from the vector-boson-fusion (VBF) production mechanism. This process is characterized by a centrally produced W boson with two hadronic jets at small angles to the beam-pipe. VBF is a significant component of the Standard Model Higgs Boson production, and so the measurement of the well-known W boson in the VBF channel provides an important demonstration of the control of both theoretical and experimental understanding of such a measurement. In addition, this channel is sensitive to anomalous WWZ and $WW\gamma$ couplings.

This thesis presents a measurement using the 2011 dataset taken with the ATLAS detector, providing a preliminary approach to extracting electroweak W production cross section and serving as a proof of principle for future studies. The thesis is structured as follows: Chapter 1 provides the theoretical background and motivation for a VBF measurement and explains the measurement strategy. Chapter 2 describes the LHC and the Atlas detector. Chapters 3, 4 and 5 explain the object and event selection, including a description of the forward jet calibration measurement to which I contributed. The signal extraction and cross section measurement is explained in Chapter 6. Conclusions and the outlook for future work is given in Chapter 7.

All unusual terms and acronymns are explained the first time they are mentioned, but as there are many a glossary is provided to assist the reader.

CHAPTER 1

Theoretical background and motivation

The Standard Model of particle physics is well established as the working theory of high energy physics describing observed phenomena up to an energy scale of ~ 1 TeV. Section 1.1 will give a brief overview of the features of the model relevant to this thesis: Section 1.2 will discuss how the predictions of the Standard Model are turned into predictions of observables at the LHC. Section 1.3 will summarize the current experimental state of electroweak physics and Section 1.3.1 will provide a description of the $W + 2$ jets phase space and the motivation for pursuing a measurement of the electroweak $W +$ jets production process. Section 1.4 will describe the characteristics of electroweak $W +$ jets production, how it is modelled and the search strategy that has been pursued in this thesis.

1.1 The Standard Model

The Standard Model of particle physics describes the fundamental particles and their interactions. It is described in detail in many textbooks, for example Reference [2].

The matter particles are formed of two groups of spin-1/2 fermions: the leptons and the quarks. Each of these groups is arranged in three generations each of which has two members, shown with their mass and charge in Figure 1.1. Each particle has an anti-particle partner, with equal mass but opposite charge. These particles are believed to be fundamental and do not have a substructure.

In the lepton sector, each generation consists of one charged lepton, with charge -1 in units of electric charge, and one neutral lepton with mass of $\mathcal{O}(\text{ eV})$ known as a neutrino. The electron is the lightest of the charged leptons and is the only one which is stable.

In the quark sector each generation consists of one up-type quark with charge $+2/3$ and one down-type quark with charge $-1/3$. In addition to electric charge the quarks carry colour charge, which is described below. Quarks are not found on their own at distances greater than ~ 1 fm. Beyond this scale they are only found in bound states.

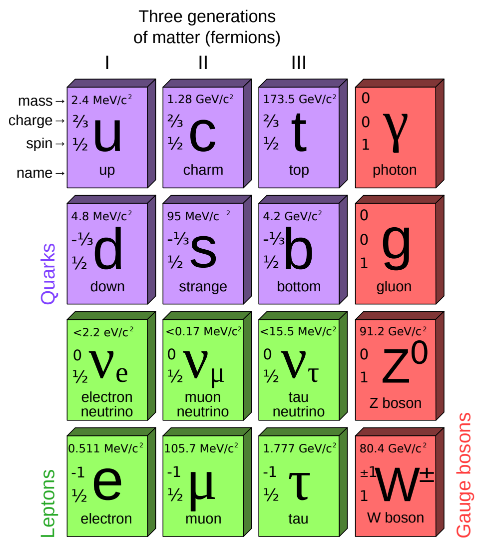


Figure 1.1: The particles of the Standard Model, including their charge, spin and mass. Figure taken from [3] with masses updated to values from [4].

The interactions of the Standard Model are described by a locally gauge-invariant quantum field theory with an $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry. The subscripts here refer to ***c**olour*, ***l**eft-handed* and *weak-**hy**percharge*. Each of these symmetries gives rise to spin-1 gauge bosons which are shown in Figure 1.1.

1.1.1 The strong force

The strong force is the interaction responsible for the binding of protons and neutrons in nuclei, and the binding of quarks within hadrons. In the Standard Model this is produced by the $SU(3)_C$ group which has 8 gluons that couple to colour charge. *Colour* is a three valued quantum number carried by quarks and gluons. The $SU(3)_C$ symmetry leads to to self interacting gluons which produces complex phenomenology, known collectively as the strong nuclear force or QCD (**Q**uantum **C**hromo**d**ynamics). An important feature of QCD is that the gluons have an anti-screening effect on the colour charge of the quark. As the quark is probed at short distances, or equivalently high energies, the apparent strong coupling α_s approaches zero. Above the energy scale, known as $\Lambda_{QCD} \sim 200$ MeV, QCD processes can be calculated with perturbation theory because $\alpha_s < 1$. This is known as 'asymptotic freedom' [5]. This is an extremely important result, since it means that at large energy scales, such as those produced by LHC collisions, quarks will interact as though they were free particles and we can make predictions of their interactions.

At the other end of the distance scale, as two quarks are separated α_s becomes stronger. The potential energy of the system increases until enough energy exists to create a new quark-antiquark pair, which occurs at the confinement limit of ~ 1 fm. The Standard Model thus explains the observation that unbound quarks do not exist at distances greater than 1 fm, and they are always found in hadrons: groups of three quarks (baryons) or quark anti-quark pairs (mesons). Perturbation theory does not work in this regime and one is forced to rely on lattice field theory calculations to predict the hadron spectrum.

1.1.2 The electroweak interaction

The electromagnetic interaction occurs only between electromagnetically charged particles and has an infinite range, mediated by photons. The weak interaction occurs between charged and uncharged leptons and quarks and has very short range. While the weak and electromagnetic are superficially different they have been combined into a single theoretical framework known as the Glashow-Salam-Weinberg (GSW) model [6] with a $SU(2)_L \otimes U(1)_Y$ symmetry.

The $SU(2)_L$ symmetry gives rise to three bosons, $\boldsymbol{W}^\mu$. These couple to weak isospin with coupling strength denoted by g. The $U(1)_Y$ symmetry gives rise to one boson, B^μ, which couples to weak hypercharge with coupling strength denoted by $g'/2$. The weak hypercharge Y is defined: $Q = \frac{1}{2}Y + I_3$, where Q is the electric charge and I_3 is the third component of weak isospin. Values of these quantum numbers for the quarks and leptons are given in Table 1.1.

Particle	I	I_3	Q	Y
		Isospin doublets		
ν	$\frac{1}{2}$	$+\frac{1}{2}$	0	-1
l_L	$\frac{1}{2}$	$-\frac{1}{2}$	-1	-1
u_L	$\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{2}{3}$	$+\frac{1}{3}$
d'_L	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$+\frac{1}{3}$
		Isospin singlets		
l_R	0	0	-1	-2
u_R	0	0	$+\frac{2}{3}$	$+\frac{4}{3}$
d'_R	0	0	$-\frac{1}{3}$	$-\frac{2}{3}$

Table 1.1: The quark and lepton assignment of weak isospin I, third component of weak isospin I_3, charge Q and weak hypercharge Y. The L (R) subscript refers to left(right)-handed particles. l is any charged lepton, ν is any neutrino, u is any up-type quark and d' is any down-type Cabibbo-rotated quark (see main text).

The weak isospin current couples to doublets of left-handed particles, transforming one member of the doublet to the other (e.g. $e_L \to \nu_e$). The weak hypercharge current couples to both the left handed doublets and right handed singlets. In the quark sector the electroweak interaction does not couple to the same eigenstates as the strong force instead coupling to "Cabibbo rotated" states. For example a W^+ boson may decay to $u\bar{d}$ or $u\bar{s}$. The mixings are parametrized by the Cabibbo-Kobayashi-Maskawa (CKM) matrix, with three mixing angles and a CP violating phase. These do not play an important role in this thesis and are not discussed further.

The physical gauge bosons of electroweak interactions are the photon (γ), the $W^\pm$ and the Z, which do not correspond to the $SU(2)_L$ and $U(1)_Y$ representations given above. The W^1 and W^2 can be identified with the physical charged vector bosons by making the following transformation:

$$W^\pm = \frac{1}{\sqrt{2}}(W^{\mu 1} \mp W^{\mu 2}) \tag{1.1}$$

The photon, denoted as A^μ, and Z boson and these are identified with a rotation in the weak-hypercharge and weak-isospin fields by the following linear combination, where θ_W is the weak mixing angle:

$$A_\mu = B_\mu \cos\theta_W + W^3_\mu \sin\theta_W,$$

$$Z_\mu = -B_\mu \sin\theta_W + W^3_\mu \cos\theta_W. \tag{1.2}$$

To preserve the regular QED interactions we must have

$$g \sin\theta_W = g' \cos\theta_W = e. \tag{1.3}$$

Thus the EM interactions and the weak interactions are unified, provided that the coupling constants satisfy Equation (1.3), known as the electroweak unification condition. By comparing the low-energy Fermi theory interactions to the charged-current interaction we have:

$$\frac{\mathrm{G_F}}{\sqrt{2}} = \frac{g^2}{8\mathrm{m_W}^2}, \tag{1.4}$$

where $\mathrm{G_F}$ is the Fermi constant. Comparing the Fermi theory to neutral current we find:

$$\frac{\mathrm{G_F}}{\sqrt{2}} = \frac{g^2}{8\cos^2\theta_W \mathrm{m_Z}^2}. \tag{1.5}$$

Therefore we find the weak mixing angle to be:

$$\frac{\mathrm{m_W}}{\mathrm{m_Z}} = \cos\theta_W. \tag{1.6}$$

While the γ is massless the Z and W bosons are known to have large masses due to the short range of the weak force. This causes a problem for electroweak gauge theory as this breaks the symmetry. The GSW model deals with this by invoking the Higgs mechanism [7]. A new scalar field with non-zero vacuum expectation is introduced which gives rise to a physical particle, the Higgs boson. This particle has spin-0 as it is a scalar, and in the Standard Model has zero electric charge. This has the effect of hiding or spontaneously breaking the $SU(2)_L \otimes U(1)_Y$ symmetry. The local gauge symmetry is preserved in weak-isospin and hypercharge space, but in low energy interactions we see the physical W and Z bosons with a mass, while the photon is massless.

Requiring local gauge invariance forces the introduction of self coupling $\boldsymbol{W^\mu}$ fields. WWZ and $WW\gamma$ are the only allowed triple gauge vertices. Neutral triple gauge vertices, such as $ZZ\gamma$, are forbidden since the W doesn't carry weak-hypercharge and the B doesn't carry weak-isospin.

There are three parameters that govern the electroweak interaction at tree-level. One possible set is g, g' and $\sin\theta_W$. These quantities do not relate directly to physical observables, since these are relations in weak-isospin space. One method to extract the parameters from physically measurable quantities uses $\mathrm{G_F}$, $\mathrm{m_Z}$ and α, the latest world-average values of which are shown in Table 1.2. The fine structure constant, $\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c}$, enters since it is the most precise method of determining e. Using Equation (1.3) one can calculate $g\sin\theta_W$ or $g'\cos\theta_W$. Using Equation (1.6) and Equation (1.4) one can calculate θ_W, g and g'.

The mass of the scalar boson (m_H) is not predicted by the tree-level theory, but for specific values of W boson mass and top quark mass the Higgs mass can be calculated from loop corrections which involve the Higgs boson. Measurements of these quantities therefore constrain the possible Higgs masses.

In addition to this minimal set of measurements a large number of complementary measurements have been made to over-constrain the electroweak model [4].

Parameter	World average value	Best experimental measurement
Parameters required to determine couplings		
G_F- Fermi constant	$1.1663787(6) \times 10^5$ GeV^2	muon lifetime [8]
m_Z- Z mass	91.1876 ± 0.0021 GeV	Z line-shape at LEP [9]
α - fine structure constant	1/137.035999074(44)	$e^{\pm}$ anomalous magnetic moment [10]
Parameters required to determine m_H from loop corrections		
m_W- W mass	80.385 ± 0.015	Tevatron [11]
m_t- top-quark mass	173.5 ± 1.0	Tevatron [12]

Table 1.2: One parametrization of the electroweak model. The first three parameters are required to calculate the electroweak couplings. The remaining two are the inputs to the Higgs mass calculation. The world average values are all taken from the Particle Data Group [4].

1.2 Modelling hadron collisions

The Standard Model describes the interactions between the fundamental quarks and leptons. However the LHC does not collide quarks but protons. Some discussion is required to explain how we can relate the large distance scale physics that is observable at the LHC to the short-distance scale theory discussed in the sections above.

The interaction of protons may be divided into “hard processes” and “soft processes” according to momentum transfer. The underlying dynamics of both are governed by QCD, but hard processes are calculable with perturbative QCD whereas soft-processes are not. As an example, the production of a W boson in association with hadronic jets is a hard process, while the total interaction cross section is largely composed of soft interactions. In this thesis we are mainly concerned with the hard processes; however we cannot ignore the soft processes altogether. As suggested by the schematic in Figure 1.2, hard-scattering events involve some soft interactions as well. I will give a brief description of how we model the interaction of protons in this section.

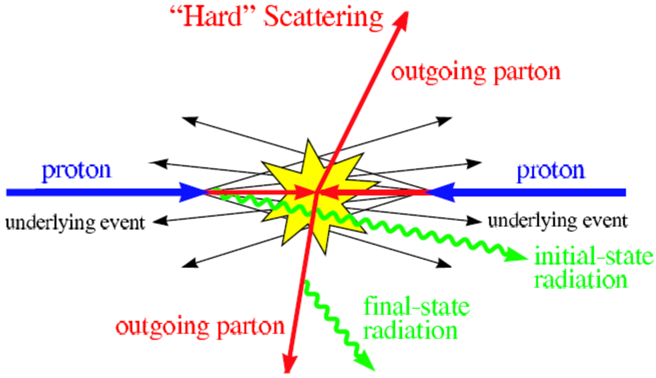


Figure 1.2: Schematic diagram of proton proton collision from [13]

Protons are composite particles made up of three valance quarks: two up quarks and one down quark, giving overall charge +1. In addition they contain an indeterminate number of "sea"-quarks — quark-antiquark pairs and gluons. The constituents of the proton are known collectively as "partons". The total momentum of the proton is distributed between the partons and the distribution is parametrized by Parton distribution functions (PDFs). In the simplest parton model, there are no interactions between the partons and the scattering probability is only a function of the Bjorken-x, which is the fraction of proton momentum carried by the struck parton. When we add QCD interactions the exact scaling with x breaks down due to the possibility of gluon emission. The PDFs are then functions of x and the momentum transfer Q^2, where the dependence on Q^2 can be calculated using the DGLAP equations [14].

The determination of PDFs is a difficult task undertaken by several groups, for example MSTW [15], CTEQ [16] and NNPDF [17]. Typically a fit is made to data from a range of experiments to extract PDFs for light quark flavours, as a function of x and includes theoretical predictions of the DGLAP evolution in Q^2. The data include fixed target electron-proton or muon-proton scattering, electron-proton collider data, proton-proton collision data from the Tevatron, and neutrino-proton scattering.

Using the PDFs one can calculate the cross section of a particular proton-proton process to produce final state X by factorizing the hard process from the PDFs using the results of the QCD factorization theorems [18]. Thus we can write the cross section of a process as

$$\sigma_{A,B\to X} = \sum_{i,j} \int dx_a dx_b f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij\to X}, \tag{1.7}$$

where i and j are parton flavours, A and B are the initial state protons, and a and b are initial state partons from these protons. $\hat{\sigma}$ is the short-distance cross section, i.e. the cross section that one calculates from the Standard Model taking the partons as free particles. The PDFs are given by $f_{i,j}$ which are functions of the fraction of momentum, x, and the factorization scale μ_F. This scale defines a separation between the short-distance and long-distance scales. If the PDFs and $\hat{\sigma}$ were calculated to all orders of perturbation theory then there would be no dependence on the choice of μ_F, but in any finite calculation it remains a parameter. The choice of scale should be appropriate to the process being considered; for example in Drell-Yan processes the scale chosen is the invariant mass of the lepton pair [13].

This formulation provides an inclusive cross section for a particular hard-process, however to compare to observations at the LHC we need predictions of differential distributions of physical observables. The final states of calculations like that in Equation (1.7) often involve unstable particles which will decay or hadronize. The initial state partons which entered the hard scatter have left behind proton-remnants which are no longer colour singlets. These must interact with one another to form a colourless final state hadrons in the phenomenon known as the 'underlying event'.

The simulation of the entire proton-proton interaction is carried out with specialist software, known as 'Monte Carlo generators'[1]. These produce event records with

[1] A distinction has to be made between 'Monte Carlo calculators', which only perform calculations of inclusive cross sections and 'Monte Carlo generators' which calculate cross sections and generate

complete lists of the final state particles and their momentum-energy four-vectors. This is invaluable as the event may be passed to a detector simulator (see Chapter 2.5) to produce a simulation of the observables that we can measure in our detector.

A number of different public codes which perform this role are available and each makes different choices in their approaches, and consequently has a different regime of accuracy. However they all follow the same general pattern[2].

First the hard interaction is calculated using perturbation theory. Effects due to initial and final state radiation (ISR and FSR) and the underlying event, i.e. all effects that are not caused by the hard process, ISR or FSR, must be included. These include multiple-parton-interactions (MPI), which are interactions between partons from the initial hadrons other than the two involved in the hard scatter[3].

Quarks or gluons in the final state are dealt with using a parton showering algorithm. This models the evolution of the partons, through sequential stochastic branching, until the hadronization scale is reached. This is the scale at which colour-connected partons dissociate into colour singlet hadrons, which may go on to decay further. This reproduces the showers of hadrons, known as 'jets', which are an observed feature of collider physics and believed to be initiated by partons in the final state.

Generators such as PYTHIA [20] or HERWIG [21] the hard-process calculation is generally limited to $2 \rightarrow 2$ processes and the calculation is performed at leading order. All additional radiation is modelled by the parton shower.

More advanced 'multileg' generators calculate matrix elements with many final state 'legs' which correspond to external legs on Feynman diagrams. An $X + 1$ parton final state may be produced either by a matrix element level calculation for $X + 1$ parton, or the matrix element level calculation for X and an additional parton produced in the parton shower. A factorization scheme is required to decide which

events.

[2]The order of the steps used here is illustrative only.

[3]ATLAS has recently released a measurement of multi-parton scattering in $W +$ jets events [19]

route to take, avoiding both doubling counting and unpopulated regions in phase space. A good scheme will also chose the best approximation given the kinematics of the event, preferring the parton shower for soft, low angle emissions, and the matrix-element for large angle, hard emissions [22]. Such a scheme has been implemented with 'MLM' jet-parton 'matching' in ALPGEN [23] and 'CKKW-L' 'merging' in SHERPA [24, 25]. They both require a 'matching scale' parameter, which sets the nominal cut-off from parton shower to matrix-element calculation. This cut-off is generally chosen to be below the scale of the jets in the final analysis, but its exact value is arbitrary and so should be varied to evaluate systematic effects. A comparison of the two algorithms is presented in Reference [22].

Next-to-leading order generators such as POWHEG [26] and MC@NLO [27] are now available to calculate some processes to next-to-leading order in QCD. SHERPA also implements the POWHEG method and can generate some processes at next-to-leading order. However, in general these calculations are limited to one real parton emission and so for mutli-parton final states multi-leg generators will give better description of the kinematics, since they rely less on the parton shower.

1.3 Testing the electroweak model

The electroweak model summarized above requires the inputs listed in Table 1.2 to predict electroweak interactions. To test that this model is a true description of nature these predictions must be compared to experimental measurements. Over the past 40 years the electroweak sector of the Standard Model has been very strongly constrained, by precision measurements of the W and Z bosons, the top quark, and diboson processes, all of which have been found to be in good agreement with theory. Current global fit to all parameters gives a p-value of 0.07 [28][4]. A review of the cur-

[4]The p-value gives the probability of at least as good a result being obtained given that the model assumed is not correct. The value for this result is slightly worse than the 2-sigma value of 0.05, and thus shows there is some tension between experiment and the standard model.

rent experimental inputs to such a fit is given in [4], and the latest summary of cross sections measured at ATLAS compared to theory is shown in Figure 1.3.

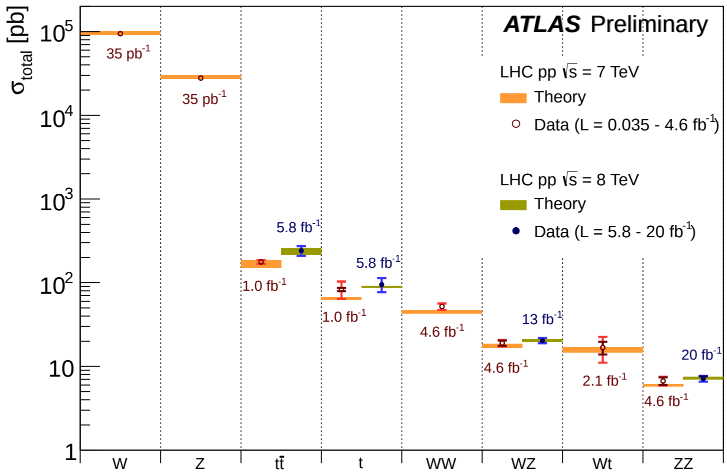
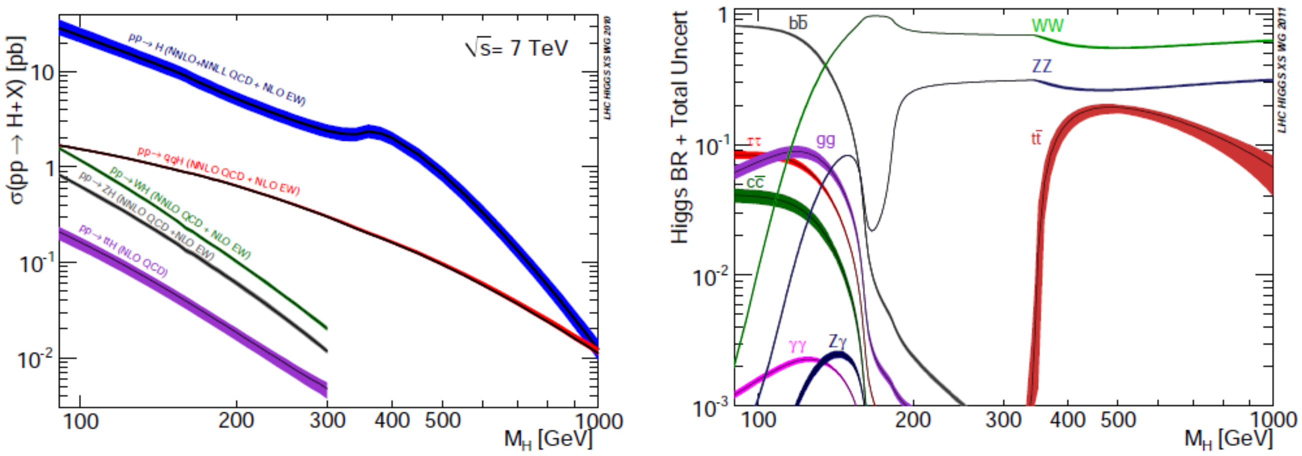


Figure 1.3: The measured cross sections for electroweak gauge boson and top-quark processes with the ATLAS detector at 7 and 8 TeV compared to theoretical calculations [29].

The search for the Higgs boson, required by the Standard Model to allow gauge-invariant masses of the electroweak bosons, has been one of the most important tests of electroweak theory. At LEP, precision measurements provided indirect upper limits on the Higgs mass, while direct searches ruled out $m_H < 114$ GeV [30]. Searches carried out at the Tevatron ruled out mass ranges around the WW resonance. In July 2012 the ATLAS and CMS collaborations announced the observation of a particle consistent with the Standard Model Higgs boson with a mass of $\sim$ 125 GeV [31] and subsequent measurements have confirmed its spin [32, 33] and couplings [34] are compatible with the Standard Model Higgs boson. Recent observations have included the $H \to \gamma\gamma$ [33, 35] and $H \to WW$ [36] processes in the vector boson fusion (VBF) channel.

1.3.1 W + jets processes

At hadron colliders W bosons are often produced in association with hadronic jets. In this thesis we are particularly interested in the $W + 2$ jets process which may be



(a) Higgs boson production cross section

(b) Higgs boson branching ratios

Figure 1.4: The production cross section (left) and branching ratio (right) of the Standard Model Higgs boson as a function of Higgs mass from [4]. $pp \to H$ is gluon-gluon fusion, $pp \to qqH$ is VBF, $pp \to WH$ and $pp \to ZH$ are associated production and $pp \to ttH$ is Higgs in association with top quarks.

produced by one of three distinct mechanisms.

The process with the largest cross section is termed '$W +$ QCD jets'. This occurs when in addition to the W boson there is some initial state QCD radiation resulting in one or more jets; some example Feynman diagrams are shown in Figure 1.5. These processes have been extensively studied at the Tevatron [37] and more recently at the LHC [38–40]. These measurements are principally tests of perturbative QCD, where the rate of additional jets may be measured and compared to theoretical predictions made by Monte Carlo generators. In addition these processes are large backgrounds to top quark and Higgs boson measurements and new physics searches and so they must be well understood. Individual measurements of W or Z bosons with jets will quickly become systematically limited, but additional precision may be achieved by comparing the ratio of W to Z production [40–42]. Since many of the systematic uncertainties in the measurement of W and Z bosons are common, they cancel in the ratio.

The second production mechanism for $W+2$ jets is diboson production with semi-leptonic decay: $WW \to \ell\nu jj$ or $WZ \to \ell\nu jj$. Some example Feynman diagrams for this process are shown in Figure 1.6. The branching ratio of W bosons to quarks

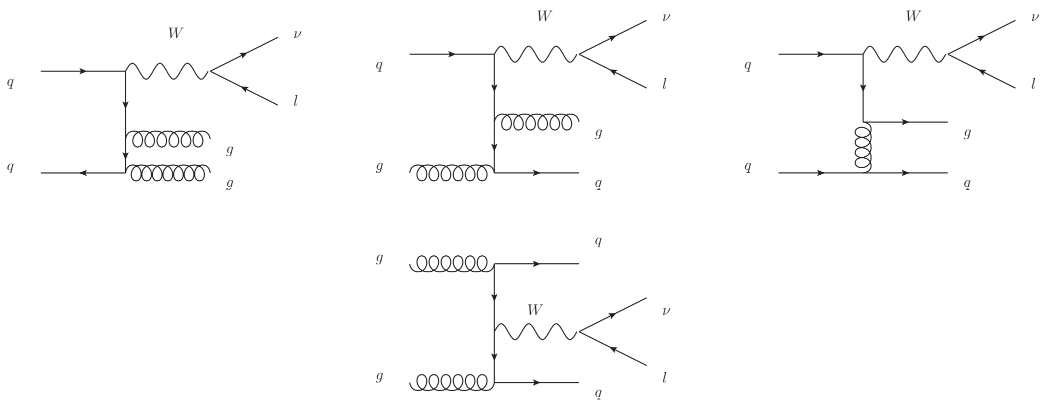
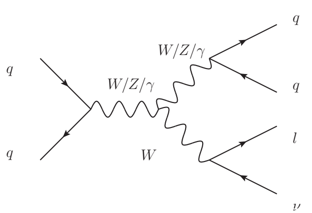


Figure 1.5: A sample of diagrams which contribute to the '$W+$ QCD jets' process.

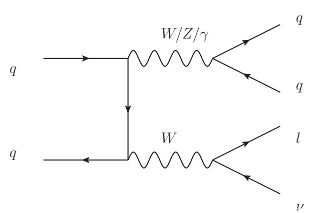
is higher than to leptons because of colour factors and so one expects much higher statistics in these channels compared to the fully leptonic diboson channel.

However, the cross section of semi-leptonic diboson decays is still much smaller than $W+$ QCD jets and so this channel remains experimentally challenging.

These processes have been studied at D0 and CDF, where limits were set on anomalous triple gauge couplings in this channel [43]. Great excitement was caused in 2011 when CDF published an analysis which observed an excess consistent with a resonance in their WW semi-leptonic M_{jj} spectrum [44]. However measurements from D0, CMS and ATLAS see no such deviation from the Standard Model [45] and it has recently been reported [46] that the CDF anomaly is due to QCD background and jet energy scale mismodelling.



(a) Diboson (annihilation or s-channel)

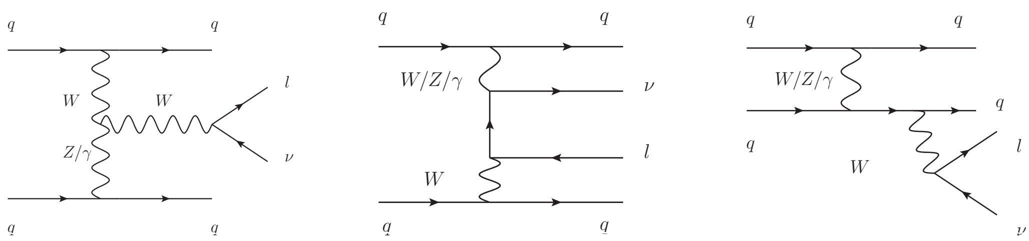


(b) Diboson (converstion or t-channel)

Figure 1.6: Example diagrams which contribute to diboson $W+$jets production.

The final production mechanism for the $W+2$ jets final state is "$W+$ EWK jets". In this process the jets are produced by the exchange of additional electroweak bosons rather than gluons. Example diagrams named following the terminology of Reference [47] are shown in Figure 1.7. It is this class of $W+$jets production that is the subject of this thesis.

A comparison of the cross-sections for the three $W+$jets processes discussed above is shown in Table 1.3.



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

Figure 1.7: Sample diagrams which contribute to $W+$ EWK jets. The full set of diagrams includes all possible permutations bremsstrahlung emission.

Process and source	Standard Model cross section (pb)	Measured cross section (pb)
$W \to e\nu + 2$ QCD jets	210 ± 20	210 ± 51 [38]
$W \to \mu\nu + 2$ QCD jets	210 ± 20	230 ± 54 [38]
$WW + WZ \to \ell\nu + 2$ jets	63 ± 3.15	71.2 ± 17.0 [48]
$W \to \ell\nu$ + 2 EWK jets	1.962 ± 0.0077	

Table 1.3: Cross sections for the three classes of $W+2$ jets production mechanisms. The theoretical and measured cross sections are taken from the ATLAS publications listed, except for the $W+$ EWK jets which was calculated with VBF@NLO [49]. The measurement of the $W+$ EWK jets cross section is covered by this thesis in Chapter 6. For $W \to \ell\nu$ + 2 EWK jets the error quoted is statistical only, while for the other cross sections the error includes PDF, renormalization, factorization, fragmentation and underlying event errors.

Motivation for a $W+$ EWK jets measurement

$W+$ EWK jets is of particular interest due to diagram 1.7(a) known as "Vector Boson Fusion" or VBF. This process has attracted great attention because of its importance as the second largest production mechanism for the Higgs boson at the LHC.

The cross section for VBF Higgs production is an order of magnitude smaller than the dominant production mechanism of gluon-gluon fusion, as shown in Figure 1.4.

Nevertheless VBF production will play an important role in determining the Higgs properties. VBF production must originate from a ZZH or WWH vertex, which is particularly important if this scalar boson is to confirm its role as the mechanism of electroweak symmetry breaking as it must couple to W and Z bosons in the correct ratio.

The identification of Higgs boson production via VBF typically involves complex phase-space cuts such as a central jet veto. These are sensitive to soft radiation which is difficult to model and for this reason it is desirable to have a reference process with similar kinematics to check calculations against [50–53]. At ATLAS this has been carried out for dijet production [54] and it is hoped that in the future a similar measurement may be made using electroweak W or Z boson production to form a cross-check for VBF Higgs production. $W+$ EWK jets production, which contains contributions from VBF diagrams, has the largest cross section of all heavy boson VBF production mechanisms and should therefore provide the most statistically precise measurement.

VBF production of a W or a Z boson also form a background to VBF Higgs and so an understanding of these is also important for precision Higgs measurements such as $H \to \tau\tau$ [52], or $H \to WW$ [55]. If massive particles from a theory beyond the Standard Model exist, such as the neutralino predicted by SUSY [56], then the Higgs boson could decay into these. These particles would not leave a signal in the ATLAS detector. A good understanding of the VBF production signature may nevertheless allow VBF $H \to$ invisible to be measured [57].

The CMS collaboration have presented a measurement of $Z+$ EWK jets production [58]. This reports the cross section in their phase space to be 154 ± 58 fb in the fiducial region $m_{ll} > 50$ GeV, $p_T^j > 10$ GeV, $|\eta^j| < 7.0$ $M_{jj} > 120$ GeV. This is in agreement with the theoretical prediction of 166 fb calculated with MadGraph. An ATLAS paper on this topic is in preparation.

As yet there has been no public result from any experiment for $W+$ EWK jets, which is the subject of this thesis. This process is predicted by the Standard Model but is one of the few remaining processes that has not yet been observed. The observation of VBF W production and a measurement of its cross section are therefore important goals in their own right.

This process also allows the probing of anomalous triple gauge vertices [59]. While these have been constrained by other diboson measurements, such measurements rely on space-like momentum transfer. The VBF process provides a unique probe into time-like momentum transfer and thus provides different sensitivities.

1.4 Modelling and measuring $W +$ EWK jets

The three $W +$ jets processes identified above all share the same final state of a W boson and two partons and thus they may interfere. If the interference is large then one cannot consider the processes separately within the Standard Model. However it has been shown [60] that the interference is small at energies well above the threshold for diboson production. The signal diagrams summarized in Figures 1.7(a)-1.7(c) form a gauge-invariant set, and thus can give a physical prediction [50, 60].

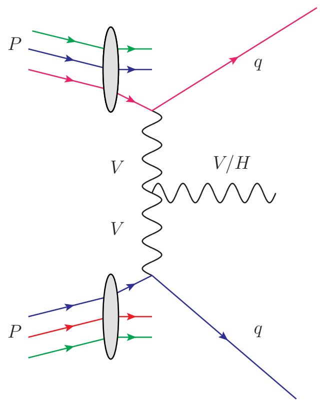


Figure 1.8: The *vector boson fusion* process. 'P' represents an incoming proton. 'V' is a W boson, Z boson or γ . 'H' is a Higgs boson.

How can one identify VBF production? The principle identifying feature comes from the two initial state quarks which each radiate a vector boson. I shall refer to these as the 'tag quarks' and the jets that they correspond to as the 'tag jets'. These quarks have large energy, but comparatively low transverse momentum since the intrinsic scale of W/Z production is $80 - 90$ GeV. In a simplified picture where we identify quarks with hadronic jets observable in our detector, this suggests that we should see two jets in opposite hemispheres of the detector at a small angle to the beam pipe.

A more detailed examination of the process requires us to think about how the colour reunification may occur. Quarks are coloured objects and yet QCD confinement means that we cannot produce coloured final states. A very important feature of VBF is the colour separation between the initial states. Consider Figure 1.8: there is no colour connection between the two quarks - the interaction is entirely mediated by colourless electroweak boson exchange. This contrasts with Figures 1.5 and 1.6 where colour charge is exchanged between the two initial state quarks. This lack of colour flow changes our expectation of how the hadronization of the two final state quarks may occur. In processes with colour exchange, hadronization may result from further colour exchange between the final states, whereas in our signal colour exchange is only likely between the tag quark and the remnants of its mother proton. The implications of this were first discussed by Dokshitzer, Khoze and Sjötrand [61], and subsequently developed by Bjorken [62]. Recently specialist Monte Carlo generators, such as VBF@NLO [49], have been developed which provide detailed differential predictions of expected kinematics of such processes.

VBF interactions result in high energy forward jets, which as they are well separated will have a high dijet invariant mass. Because of the colour structure there is expected to be few hadronic particles in the rapidity gap between the tag jets and so a 'central jet veto' may be useful in distinguishing signal from background.

1.4.1 ATLAS search strategy

In the ATLAS detector the signature for $W+$ EWK jets production will be a W boson in association with two hadronic jets that are typically produced at a small angle to the beam pipe. The W boson may be identified by its leptonic decay to a high p_{T} electron or muon in addition to missing transverse momentum due to the neutrino.

Detector level cuts are used to maximize the signal-to-background ratio and the signal significance as far as possible. These are described in detail in Chapters 3 and 4. Final evaluation of the signal component is made using a fit to the dijet invariant mass. This variable is expected to give good separation since a pair of high-energy jets produced close to the beam pipe with large rapidity separation will have a high invariant mass. The details of this fit are explained in Chapter 6.

The choice of cuts and the application of the fit require knowledge of the kinematic distributions of signal and background. The source of the modelling is discussed for the signal and each source of background in Chapter 4.

1.5 Summary

The Standard Model predicts triple-gauge boson interactions, and while space-like, interactions have been widely studied time-like vector boson fusion interactions have not. These processes are interesting for their similarity to Higgs production processes and they can act as a test region for Higgs measurements, as a background to VBF Higgs production, and in their own right as another test of Standard Model predictions. Such a test of Standard Model predictions includes further probes of anomalous triple gauge coupling.

CHAPTER 2

The LHC and the ATLAS Detector

The Large Hadron Collider (LHC), located at CERN in Geneva, is designed to produce high luminosity beams of protons and collide them at high centre-of-mass energy to allow the study of the energy frontier of particle physics. The Atlas (A Toroidal LHC ApparatuS) detector surrounds one of the four points where protons are brought into collision and it facilitates a broad range of physics measurements, from precision tests of the Standard Model to searches for new and exotic physics. The accelerator and detector are the result of over twenty years of design and construction from a multinational team of several thousand engineers and physicists. In this chapter I will provide a short overview of the elements of the LHC (Section 2.1), the Atlas detector (Section 2.2), trigger system (Section 2.3), data processing (Section 2.4) and the detector simulation (Section 2.5), relevant to the work presented in the remainder of this thesis.

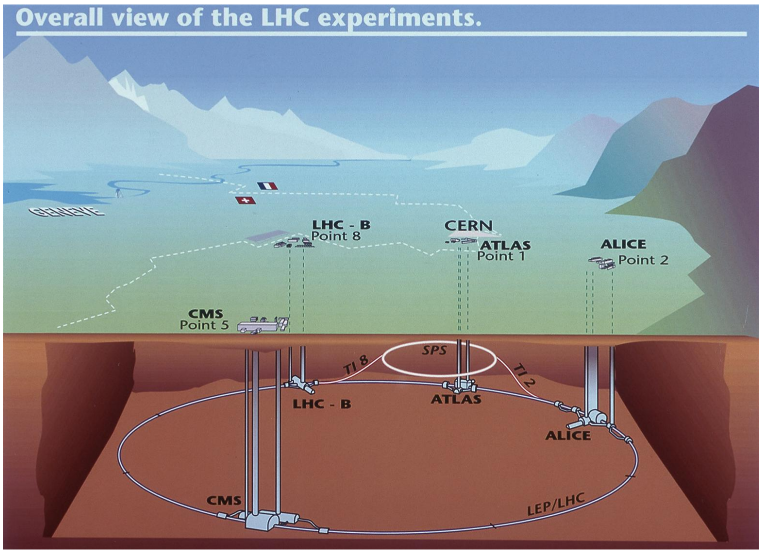


Figure 2.1: A cross section of the LHC showing the LEP tunnel and the four LHC experiments [63].

2.1 The LHC

The LHC is the largest and most energetic proton synchrotron ever built. It was constructed in the tunnel previously occupied by the Large Electron Positron Collider (LEP) with a circumference of 26.7 km and lies between 45 and 170 m below the surface. It consists of two superconducting rings which support same-sign charged-hadron beams circulating in opposite directions and brings them into collision at four points around the ring as shown in Figure 2.1. LHCb is the specialized b-physics experiment located at point-8 [64]; ALICE is the specialist heavy-ion experiment located at point 2 [65]; ATLAS [66] and CMS [67] are the so-called 'general purpose' detectors located at points 1 and 5 respectively.

Protons are supplied through the CERN accelerator chain: Linac2 50 MeV → Proton Synchrotron Booster 1.4 GeV → Proton Synchrotron, 25 GeV → Super Proton Synchrotron 450 GeV → LHC 7 TeV.

The LHC consists of superconducting dipole magnets which bend the beam around the roughly circular course, and quadrupole and higher-moment magnets which focus the beams. Eight RF cavities per beam provide power to accelerate and maintain the beam, and also provide longtitudinal beam focusing to maintain the bunch structure.

The nominal design of the LHC called for an energy of 7 TeV per beam and luminosity of 10^{34} cm^2 s^{-1} [68]. These running conditions are not expected to be achieved until at least 2015 and in the following I describe the beam conditions during the 2011 run at point 1, when the data for this thesis was taken.

In 2011 the LHC produced proton beams with 3.5 TeV per beam. These were segmented into bunch trains with a separation of 50 ns between bunches in a given train. In order to maximize the number of low cross section events which ATLAS could record, the LHC sought to increase the integrated luminosity as much as possible. This is done by increasing the number of protons per bunch, reducing the physical size of the bunch by optimizing beam optics, and/or increasing the total number of bunches in the machine.

The absolute luminosity, $\mathscr{L}$, can be calculated from beam parameters,

$$\mathscr{L} = \frac{n_b f_r n_1 n_2}{2\pi \Sigma_x \Sigma_y}, \tag{2.1}$$

where n_b is the number of bunches, f_r is the rotation frequency, n is the beam charge of each beam, and Σ measures the beam size in x and y.

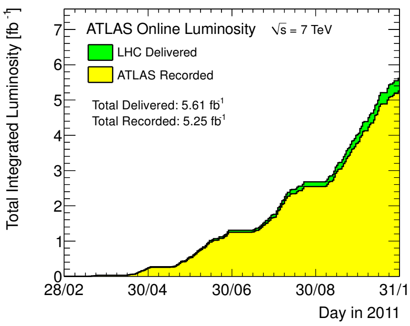
For the purposes of instantaneous luminosity determination an alternative formulation of the luminosity in terms of the interaction rate μ, detection efficiency and inelastic cross section σ is:

$$\mathscr{L} = \frac{\mu n_b f_r}{\sigma} = \frac{\mu_{vis} n_b f_r}{\epsilon \sigma} = \frac{\mu_{vis} n_b f_r}{\sigma_{vis}}. \tag{2.2}$$

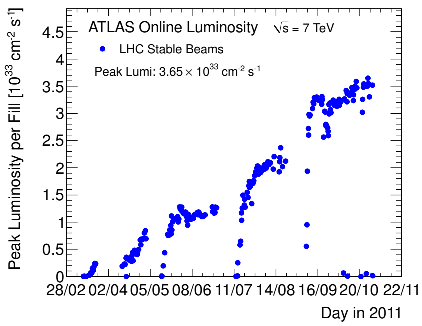
Using the absolute luminosity from Equation (2.1) and the observed interaction rate μ_{vis} the visible cross section σ_{vis} can be calculated without using the total inelastic cross section or knowledge of the detector efficiencies.

At ATLAS this is performed using Van der Meer scans [69]. In this procedure the beams are scanned across one another first in the x-direction and then in the y-direction while the event rate is monitored. This allows the beam width $\Sigma_{x,y}$ and the maximum visible interaction rate μ_{vis} to be measured. The interaction rate is measured using multiple methods including the LUCID forward Cherenkov detector, hit counts in the muon chambers, vertex counting or Z boson counting. These different measurments compliment one-another and allow for a smaller uncertainty.

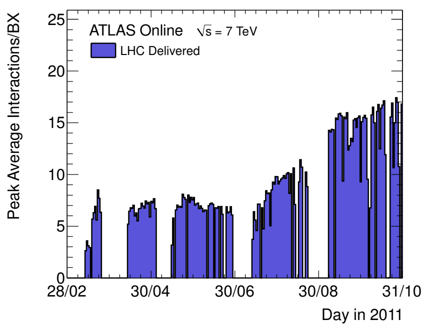
The variation in luminosity over the 2011 run can be seen in Figure 2.2 with a peak instantaneous luminosity of 3.65×10^{33} cm^{-2}s^{-1}. The luminosity measurement for 2011 has an overall uncertainty of $\pm 1.8\%$ [69, 70].



(a) Integrated luminosity by day



(b) Peak instantaneous luminosity by fill



(c) Peak μ

Figure 2.2: The evolution of integrated and peak luminosity as a function of time in the 2011 run [71]

As the number of protons per bunch increases, the average number of interactions per bunch crossing μ increases, leading to higher rates of multiple interactions recorded in the detector. This is known as *in-time pile up*. In addition, interactions from one bunch crossing can interfere with the recording of events in a subsequent bunch crossing either through detector deadtime, or due to long signal readout times. This is a particular problem in the liquid argon calorimeters which have very long pulse shapes of 400 ns. This effect is known as *out-of-time pile up.* The evolution of peak μ with time is shown in Figure 2.2(c).

2.2 The ATLAS Detector

Atlas is a general purpose hermetic detector, designed to come as close as possible to full 4π steradians coverage of the solid angle around the interaction point and to enable the reconstruction of momentum-energy 4-vectors of all visible particles emerging from interactions. It is 44 m long with a diameter of 25 m and has an overall mass of $\sim$ 7000 tonnes. An overview can be seen in Figure 2.3.

At its centre is a tracking system consisting of a silicon pixel array, a silicon strip tracker and a transition radiation tracker. These are surrounded by a solenoid which immerses the tracker in a 2 T magnetic field to allow measurement of the transverse momentum of charged particles. This is described in more detail in Section 2.2.2. At larger radii from the solenoid is an electromagnetic calorimeter, followed by hadronic calorimeters. These are described in Section 2.2.3. The final layer consists of a muon spectrometer, which is embedded in a toroidal field, and designed to measure the direction and momentum of muons. This is described in Section 2.2.4. For all detector elements the principle reference is [66].

2.2.1 Coordinate definitions

Atlas uses a right-handed coordinate system, with the z-axis parallel to the anti-clockwise beam direction, the x-axis pointing towards the centre of the ring, and the y-axis pointing upwards. $\phi = 0$ points along the x-axis and increases in the clockwise direction when looking in the positive z direction.

Since the initial state partons generally have different momenta (as described in

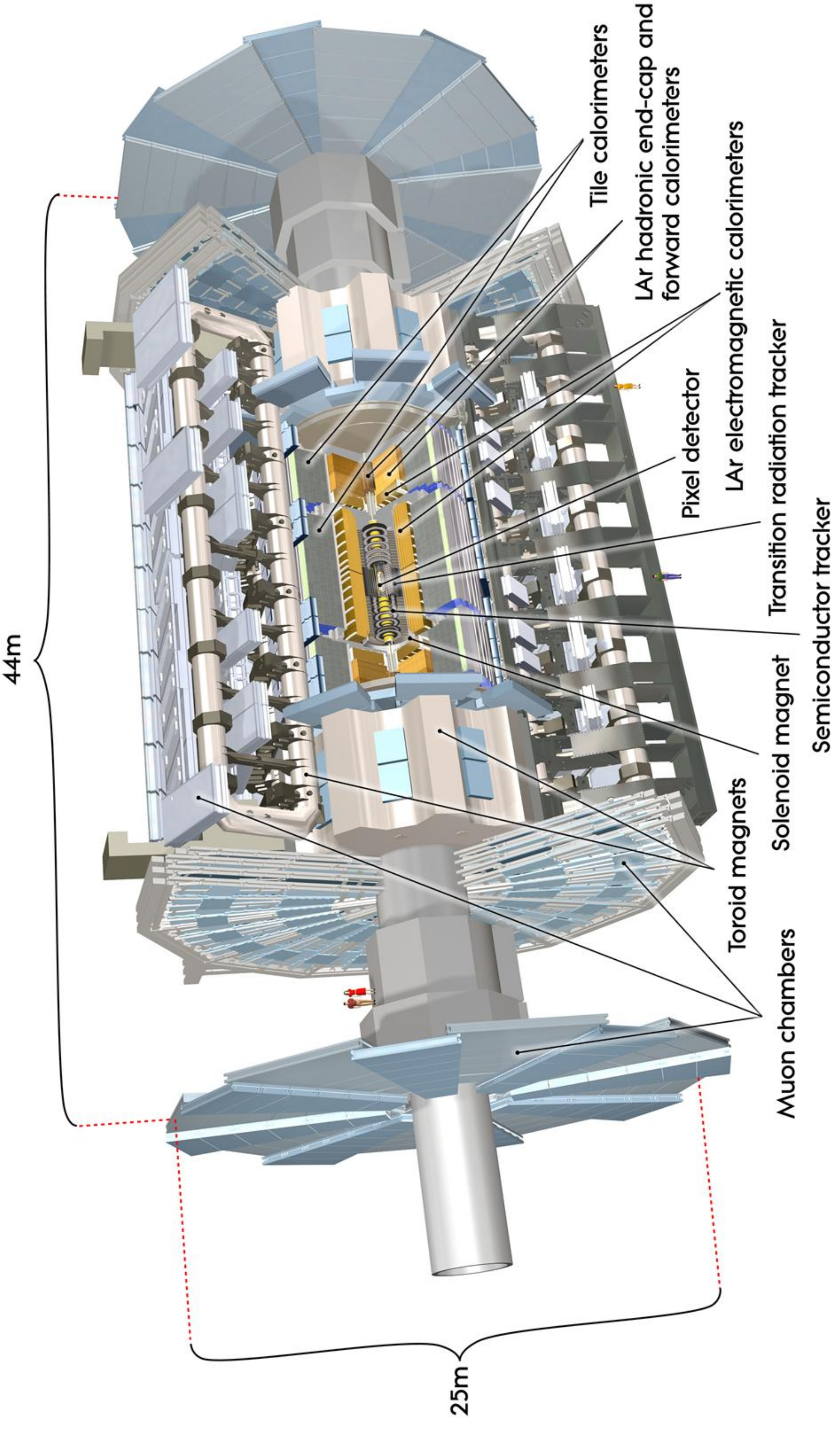


Figure 2.3: The ATLAS detector [63]

Section 1.2), the interactions produced will be given a boost in the beam direction. The rapidity, defined as:

$$y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}, \tag{2.3}$$

is a useful quantity since differences in rapidity are Lorentz-invariant to boosts in the z direction. This means distances between objects measured in units of rapidity are insensitive to the fact that our hard-collision did not occur in the lab frame. At the LHC most particles measured in the detector have $E \gg m$ and in this limit Equation (2.3) reduces to the pseudorapidity η:

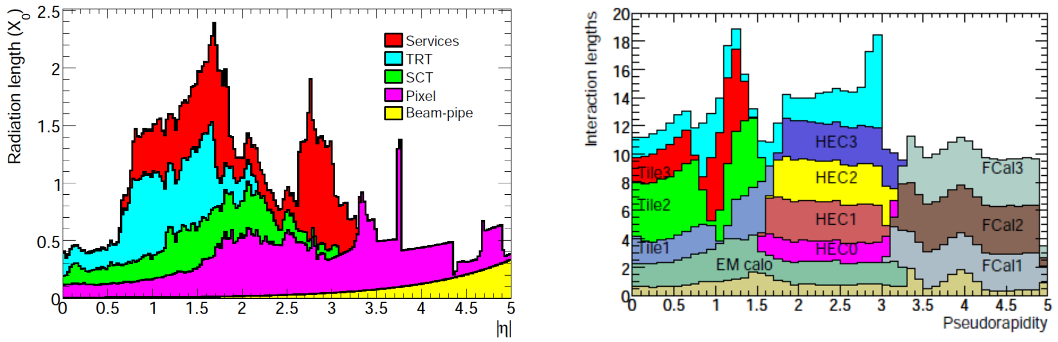
$$\begin{aligned} \eta &= \frac{1}{2}\ln\frac{|p|+p_z}{|p|-p_z} \\ &= \frac{1}{2}\ln\frac{1+\cos\theta}{1-\cos\theta} \\ &= -\ln\tan(\frac{\theta}{2}). \end{aligned} \tag{2.4}$$

This is used in many places as an alternative to θ when describing detector components.

The transverse momentum p_{T}, the transverse energy E_{T}, and the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ are defined in the x-y plane. The scalar energy E is converted as $E_{\mathrm{T}} = E\sin\theta$. This terminology is frequently used for objects measured with calorimeter information. The distance ΔR in the pseudo-rapidity-azimuthal angle space is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

2.2.2 Inner detector

The purpose of the Inner Detector (ID) is to record three-dimensional coordinates ("space-points") which form the input to pattern recognition algorithms which reconstruct the particle tracks. This allows the measurement of particle direction, bending radius and curvature, which are used to calculate the particle's charge and p_{T}. It also enables identification of the primary and secondary vertices which are useful for b and c jet tagging and for distinguishing and correcting for pile up events. The transition radiation tracker provides tracking information and also makes use of transition radia-



(a) Radiation lengths in the inner detector (b) Interaction lengths in the calorimeter

Figure 2.4: The material distribution in the inner detector and calorimeters as a function of η averaged over ϕ [66]. The inner detector is given in units of radiation lengths (X_0) and the calorimeters in units interaction length (λ).

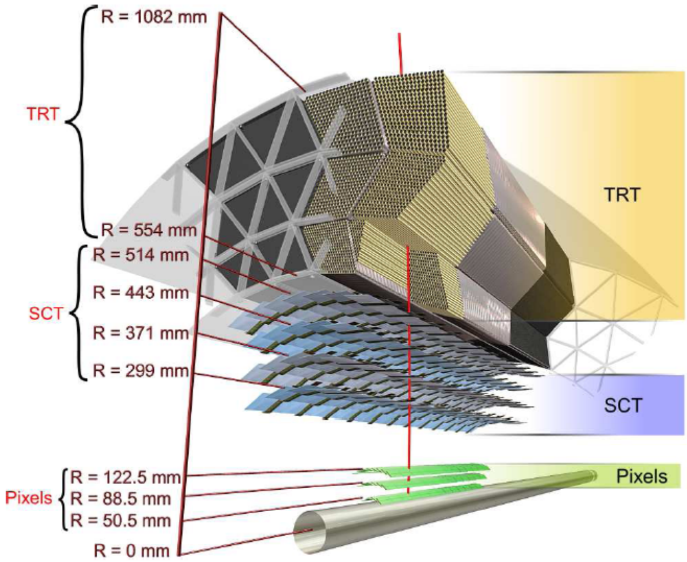
tion to improve the electron identification. The ID gives an overall coverage of $\eta < 2.5$ for particles with $p_{\mathrm{T}} > 0.5$ GeV.

The ID represents a compromise between the two competing design aims of very low mass and very high space point resolution. Low mass is required to keep bremsstrahlung and multiple scattering to a minimum. These effects degrade the momentum resolution of the tracker since they make particle track deviate from a helix. The material distribution in the inner detector in units of radiation length[1] and η is shown in Figure 2.4(a).

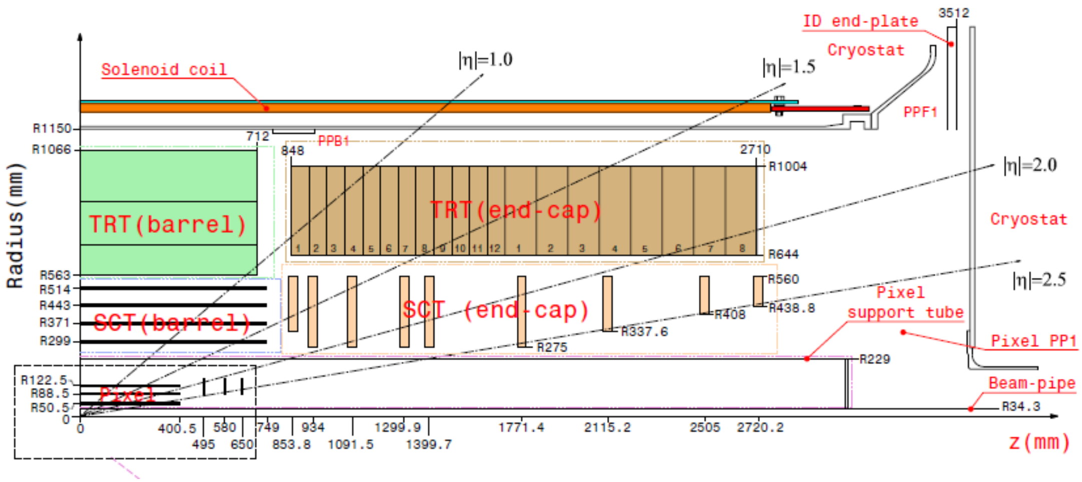
The ID consists of three detectors which are shown in Figure 2.5. In the central region ($|\eta| \lesssim 1$) the detector elements are arranged in concentric barrels. This area can be seen in Figure 2.5(a). Further forward ($1 < |\eta| < 2.5$) are the 'end-caps' — disks perpendicular to and centred on the beam pipe. This area can be seen in Figure 2.5(b)

The detector elements nearest the beam pipe have the finest granularity, since these are subjected to highest particle flux, while those at larger radii can achieve the required performance with lower granularity, thus reducing detector mass and power requirements.

[1]A 'radiation-length' (X_0) is the distance in a given material that an electron must travel to lose energy $(1 - e^{-1}) \cdot E_0$ where E_0 is its initial energy.



(a) Inner detector barrel section



(b) Inner detector forward section

Figure 2.5: The inner detector [66]

Separating the high vacuum of the LHC from the detector is a 38 m beryllium beam-pipe. It has an inner radius of 58 mm and is 0.8 mm thick. The material and thickness were chosen to maintain the LHC vacuum while minimizing interaction and radiation lengths which would disrupt particle paths before they enter the detector.

The Pixel detector [72] consists of three concentric barrels surrounding the interaction point and three end-cap disks at each end, with geometry and locations as shown in Figure 2.5. This arrangement ensures that all tracks with $|\eta| < 2.5$ should pass through at least three pixel layers. The inner-most layer, known as the b-layer, located 50.5 mm from the interaction point, is as close to the interaction point a possible to maximize vertexing precision.

Each barrel is built from identical 'staves', inclined at an angle of 20 degrees such that they overlap. Each stave has 48 pixel modules mounted on it, with 47,232 pixels per module. Due to its proximity to the interaction point the pixel detector must cope with very high particle flux, which is a challenge both in terms of space point resolution and radiation dose. To cope with the track density the nominal pixel size is 50 μm $\times$ 400 μm. Test beam measurements of of individual pixel modules found spatial resolution to be 12 μm with $\sim$ 80% of tracks giving a single pixel hit [73].

The semiconductor tracker (SCT) [74] uses stereo strips to provide space-points at larger radii. Two layers of silicon are arranged with strips at a small angle to one another. When a charged particle traverses the module a signal will be recorded on two strips and their intersection provides the space point. The use of strips rather than pixels reduces the number of channels and thus the amount of read out electronics and associated services required, thereby reducing the detector mass at some loss of position resolution. A small stereo angle reduces the probability of 'ghost hits', with the trade-off that resolution in the axis of the strips is worse than if the strips were perpendicular.

The SCT consists of four concentric barrels and nine end-cap disks at each end of the barrel as shown in Figure 2.5. This arrangement ensures that all tracks with $|\eta| < 2.5$ will pass through four Semiconductor Tracker (SCT) layers as shown in Figure 2.5(b). Test beam measurements at normal incidence give $R - \phi$ resolution of 16 µm [75].

The Transition Radiation Tracker (TRT) [76] provides both drift-circles for tracking, and transition radiation information for electron identification. The tracking element operates on the principle of a proportional straw drift chamber. It is made of plastic drift tubes or straws filled with a Xe-CO_2-O_2 gas mixture, which are 4 mm in diameter and have a drift time of 40 ns. This is sufficient to give an $R - \phi$ resolution of 130 µm.

In the barrel the straws are arranged parallel to the beam, between 73 and 22 layers depending on η, with a track in the barrel region having an average of 36 TRT hits. These are interleaved with polypropylene fibres which produce the transition radiation. The Transition Radiation Tracker (TRT) only provides an $R - \phi$ position, and does not provide a z measurement.

In the end-cap the straws are arranged radially in 96 layers and thereby give a $z - \phi$ measurement, but not an R measurement. Between the layers there are polypropylene foils which provide the transition radiation in the end-caps.

Transition radiation is produced in proportion to the Lorentz factor $\gamma = E/mc^2$ of a charged particle [77] and it is thus possible to distinguish particles based on their mass. Electrons will produce transition radiation from $\gtrsim$ 1 GeV while other particles will only radiate from $\gtrsim$ 100 GeV. Transition radiation photons are absorbed by the Xe gas and produce much higher signal amplitudes than minimum-ionizing particles. These are registered on a straw-by-straw basis by special 'high threshold' electronics.

Solenoid

A superconducting solenoid producing a nearly uniform 2 T field surrounds the tracking system. A high magnetic field is vital to enable accurate particle momentum determination up to high momentum scales. Care was taken to keep radiation length of the solenoid to a minimum with ~ 0.66 radiation lengths for particles with normal incidence.

2.2.3 Calorimeters

The calorimetry system aims to capture all particles, except muons and neutrinos, emerging from the interaction point with $|\eta| < 4.9$ and accurately measure their energy. These provide a complementary measurement to the tracking, since the momentum resolution from the tracker increases with $\sigma_{p_\mathrm{T}}/p_\mathrm{T} \propto p_\mathrm{T}$ while the calorimeter energy resolution decreases with increasing energy $\sigma_E/E \propto 1/\sqrt{E}$.

A variety of technologies are used depending on η and R. At small R, the calorimeters are designed to contain the electromagnetic showers and provide identification, direction and energy information for electrons and photons. At larger distances the technology shifts to denser media capable of containing hadronic showers. An overview of the calorimeter system is shown in Figure 2.6. A summary of the resolution of the different calorimeter components is given in Table 2.1. A summary of the material distribution as a function of interaction length[2] and η can be found in Figure 2.4(b).

The liquid-argon calorimeter provides coverage for $|\eta| < 1.475$ with a barrel system, and $1.375 < |\eta| < 3.2$ from an end-cap system. It has an overall depth of between 22 and 30 radiation lengths depending on η. It is thus the primary electromagnetic calorimeter as it will almost fully capture the showers from electrons and photons in the central region.

[2] λ is the 'interaction length', which in analogy with the radiation length is the distance a hadron must travel to lose energy $(1 - e^{-1}) \cdot E_0$ where E_0 is its initial energy.

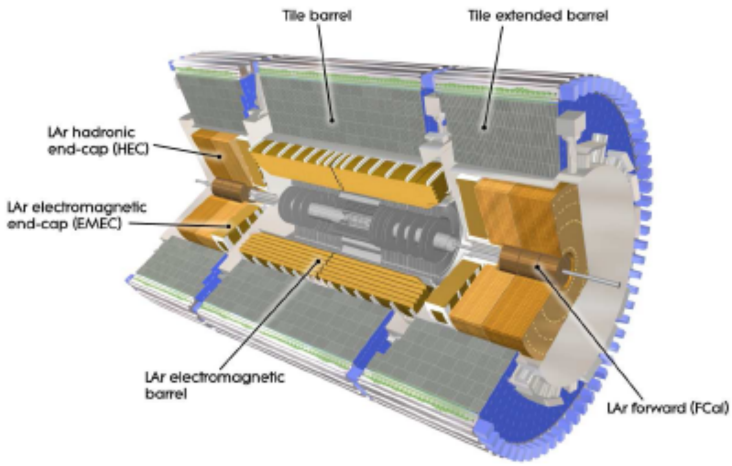


Figure 2.6: An overview of the ATLAS calorimeters [66]

Component	Expected Resolution $\frac{\Delta E}{E}$
LAr Barrel (EM)	$\frac{10\%\sqrt{\text{GeV}}}{\sqrt{E}} \oplus 0.7\%$
Tile (had)	$\frac{56.4\%\sqrt{\text{GeV}}}{\sqrt{E}} \oplus 5.5\%$
HEC (had)	$\frac{70.6\%\sqrt{\text{GeV}}}{\sqrt{E}} \oplus 5.8\%$
FCal (EM)	$\frac{28.5\%\sqrt{\text{GeV}}}{\sqrt{E}} \oplus 3.5\%$
FCal (had)	$\frac{94\%\sqrt{\text{GeV}}}{\sqrt{E}} \oplus 7.5\%$

Table 2.1: The nominal resolution of the different calorimeters taken from single particle electron (EM) or pion (had) test beams [66]. The resolution most relevant to each calorimeter is given here.

It is a sampling calorimeter, using lead as the absorber material and liquid argon as the active scintillation material. Lead is chosen as the passive absorber because of its high density, and liquid argon chosen as the scintillator due to its linear response, radiation hardness and stability over time. The lead absorber and electrodes are formed into an accordion geometry as shown in Figure 2.7. This is read out in three layers of $\eta - \phi$ cells. The first layer has very fine segmentation in η which allows good identification of $\pi^0 \to \gamma\gamma$. The depth segmentation allows good determination of the longitudinal shower shape which is important for distinguishing electromagnetic objects from hadronic ones.

In front of the main barrel calorimeter is an 11 mm layer of liquid argon which forms the 'pre-sampler'. This has no dedicated absorber and so can be used to measure the shower caused by dead material upstream of the calorimeter.

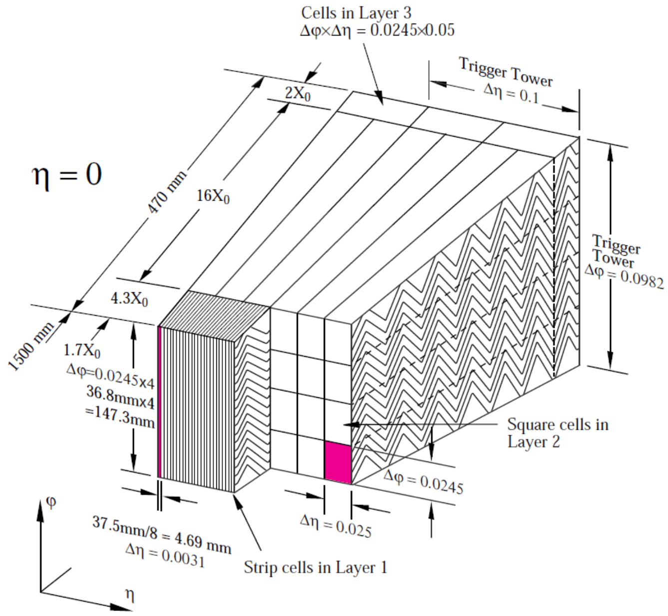


Figure 2.7: Diagram showing geometry of a barrel module. The three layers in R and and ganging in $\eta - \phi$ are visible [66].

The tile calorimeter (Tile-Cal) covers central region ($|\eta| < 1.7$). It provides a depth of 9.7 interaction lengths (λ) at $\eta = 0$. It is constructed from a barrel covering $z = \pm 2.9$ m ($|\eta| < 1.0$) and two extended barrels 2.6 m in length covering $0.8 < |\eta| < 1.7$. It is a non-compensating[3] sampling calorimeter. Steel is used as the absorber for its high density, mechanical strength and ferromagnetic properties which allow it to function as the return yoke for the solenoid. The active material is provided by plastic scintillator tiles with a absorber to active media ratio of 4.7 : 1. The readout is arranged into 'cells' of $\eta - \phi$ which are $\sim 1.5, 4.1$ and 1.8 λ thick at $\eta = 0$, with $\Delta\eta \times \Delta\phi \approx 0.1 \times 0.1$ in the first two layers and $\Delta\eta \times \Delta\phi \approx 0.2 \times 0.1$ in the final layer. Figure 2.8 shows the segmentation of modules in depth.

The LAr Hadronic Endcap (HEC) slightly overlaps with the Tile Cal, giving coverage over $1.5 < |\eta| < 3.2$. It is a liquid-argon sampling scintillation calorimeter which is desgined primarily with hadron showers in mind. Flat plates of copper, perpendicu-

[3] A large fraction of the energy of a hadronic shower is lost as non-ionizing radiation such as slow neutrons, nuclear excitations and neutrinos. A compensating calorimeter attempts to correct for this by using an absorber material such as U^{238} which will absorb slow neutrons and re-emit photons. These photons will produce visible scintillation in the active media of the calorimeter [78]. The D0 [79] and ZEUS [80] detectors made use of this technique.

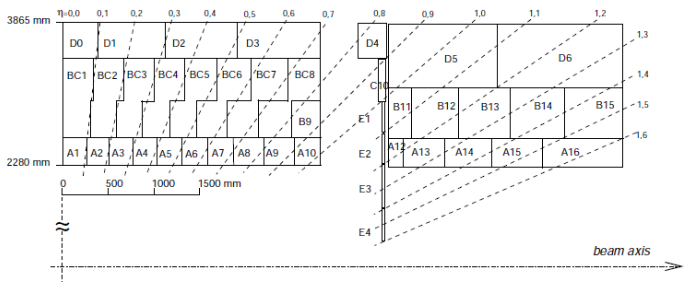


Figure 2.8: The segmenation in η and depth of the tile calorimeter [66]. The left hand section is the barrel while the right hand part is the extended barrel.

lar to the beam axis provide the absorber material. The HEC consists of two end-cap wheels HEC1 and HEC2. $|\eta| \leq 2.5$ has a readout granularity of $\Delta\eta \times \Delta\phi \approx 0.1 \times 0.1$ with further forward regions having granularity of $\Delta\eta \times \Delta\phi \approx 0.2 \times 0.2$.

The LAr Forward Calorimeter (FCal) provides measurements from $3.1 < |\eta| < 4.9$ with a depth of $\sim 10\lambda$. It is constructed from electrode rods, surrounded by a 2 mm layer of liquid argon and then a tube of absorber material. There are three modules in each FCal. The first uses an absorber of copper, which has good heat-conducting properties to deal with the very high particle flux. This layer acts mostly as an electromagnetic calorimeter with ~ 27 radiation-lengths and 2.66 interaction-lengths. The second two layers use tungsten absorbers which are extremely dense and so provide better lateral containment of hadronic showers. These have a depth of ~ 3.6 interaction-lengths each.

2.2.4 Muon Spectrometer (MS)

Downstream from the calorimeters, the only particles that should remain are muons and neutrinos (a very small number of 'punch through' hadrons that escape the calorimeters are expected). Identification of muons for triggering and offline reconstruction is important for a wide variety of physics and is used in this thesis for W boson identification and reconstruction. For physics we aim to measure p_{T} with an accuracy of 10% for 1 TeV, while for triggering fast identification is required and coarser

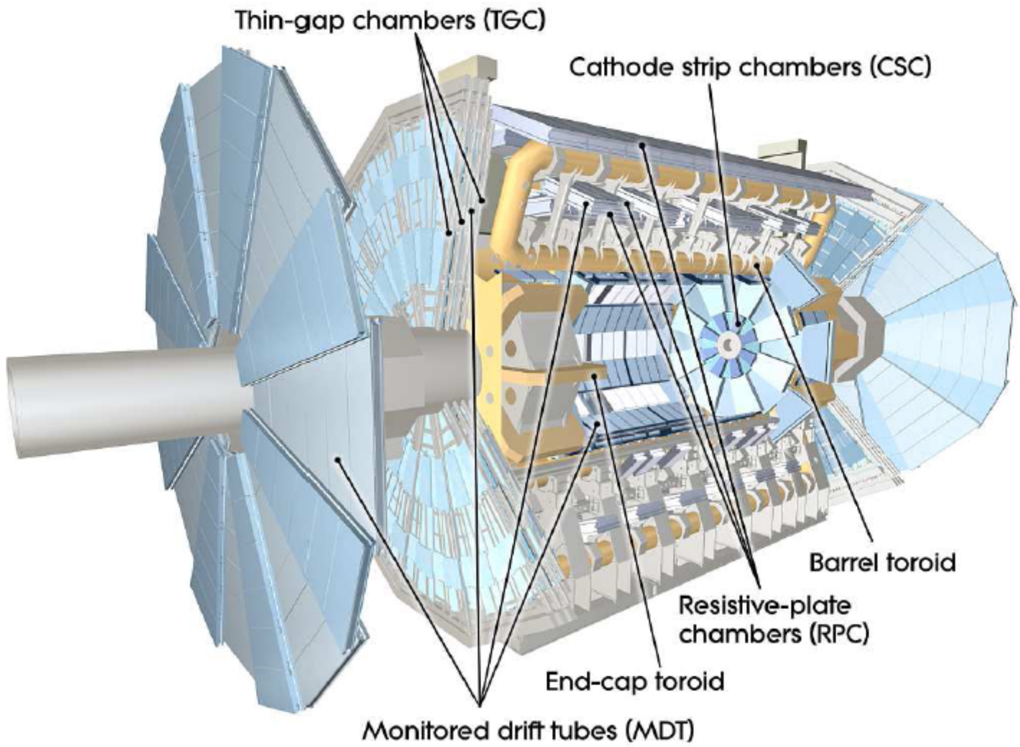


Figure 2.9: A cross section of ATLAS showing components of the muon system [66].

p_T, η and ϕ measurements are acceptable. This is achieved with a variety of tracking chambers around an air-core toroid magnet shown in Figure 2.9. For $|\eta| < 1.4$ the bending is provided by the large barrel toroid, which can be seen in Figure 2.9. The tracks in this region are recorded in detector arranged in three barrel layers. For $1.6 < |\eta| < 2.7$ bending is provided by two smaller end-cap magnets inside the main toroid. For the transition region $1.4 < |\eta| < 1.6$ bending is provided by a combination of both of the above magnet systems. For the end-cap and transition regions tracking is provided by end-cap disks perpendicular to beam-line.

Monitored Drift Tubes (MDT) cover $|\eta| < 2.7$ and provide precision tracking. This system is formed of proportional drift tubes filled with Ar-CO_2 at ~ 3 bar. The tubes are 3 cm in diameter with various lengths depending on the part of the detector. They have a maximum drift time of ~ 700 ns giving an average tube resolution of ~ 80 µm.

Cathode Strip Chambers (CSC) cover $2.0 < |\eta| < 2.7$ and provide precision tracking in the forward region, where high particle flux requires finer granularity than the MDTs can provide. They consist of multi-wire proportional chambers with cathode strips and anode wires. Each chamber has four CSC layers with wires in a radial direction and cathode strips parallel to the wire on one side and perpendicular on the other. This provides both R and ϕ coordinates for tracking. In the bending plane of the particles the CSCs are expected to provide a resolution of ~ 60 μm.

Resistive Plate Chambers (RPC) cover $|\eta| < 1.05$ and provide triggering and an additional space-point for tracking. They consist of a 2 mm gas-filled gap between two resistive plates which are oppositely charged. They are operated in 'avalanche' mode which gives very fast read-out. Pick-up strips with a width of $25 - 35$ mm are located on the outside of the resistive plates, with the strips on opposite sides arranged perpendicular to one another to provide two coordinate measurements. The RPCs are constructed in common units with the MDT and arranged as shown in Figure 2.10. Each chamber has two layers and thus tracks which traverse all three stations provide redundancy in tracking to reject background.

Thin Gap Chambers (TGC) cover $1.05 < |\eta| < 2.7$ and provide trigger and an additional coordinate in the azimuthal direction. They operate as multi-wire proportional chambers which operate in 'quasi-saturated' mode. The wires are closely spaced to provide fast read-out and allow use for triggering.

2.3 Triggers

In the 2011 run[4] the LHC provided collisions at a rate of 20 MHz, with μ up to 16. The data processing and recording infrastructure is only capable of writing out data

[4] The figures given in this section reflect the running conditions corresponding to the data used for this thesis. For design values see Reference [66].

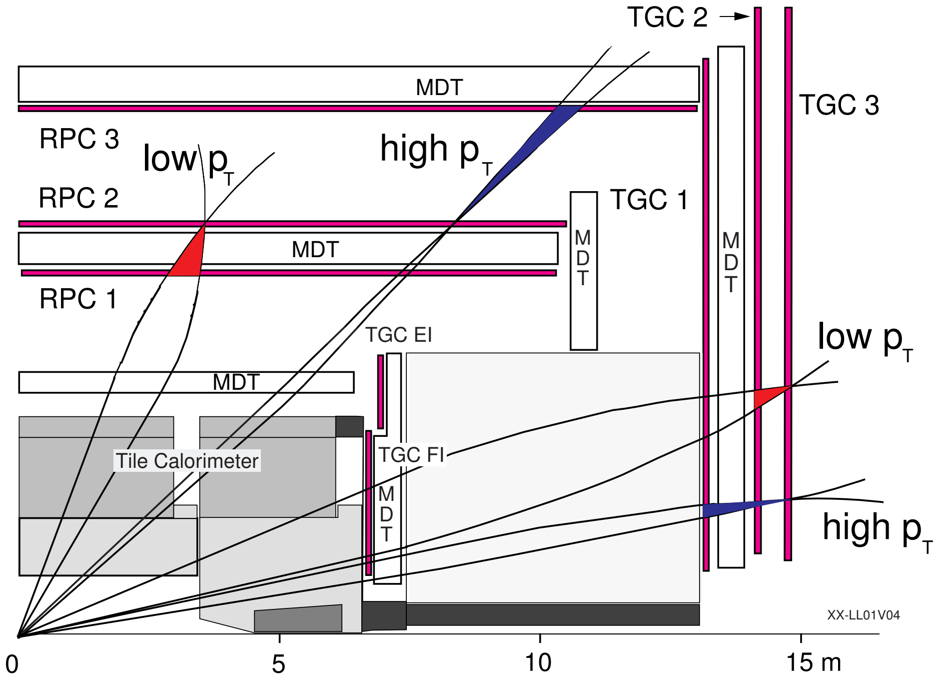


Figure 2.10: A quarter view of the muon system [81]. The muon paths show 'roads' which are used in muon triggering, see Chapter 3.3. The furthest forward MDT detectors and the CSC detectors are not shown in this diagram.

at a rate of ∼ 300 Hz, so a trigger system is vital to select events that merit further study while discarding uninteresting events. The ATLAS trigger uses information from the calorimeters and the muon spectrometer to build crude electron, jet, tau, $E_{\mathrm{T}}^{\mathrm{miss}}$ and muon objects and produces a decision based on whether these meet predefined thresholds.

It is crucial that trigger decisions are made quickly to keep the required hardware manageable. To allow this the ATLAS trigger is organized into three levels, Level 1 Trigger (L1), Level 2 Trigger (L2) and Event Filter (EF). An event is rejected if it does not pass any trigger at a particular level. Higher level triggers are seeded by triggers at lower levels forming trigger 'chains'.

The L1 trigger must reduce the data rate from the raw bunch-crossing rate to 75 kHz. It has ∼ 2 µs to reach a decision and is hardware-based, located on electronics racks close to the detector.

The High Level Trigger (HLT) runs in two stages – L2 and EF — with average

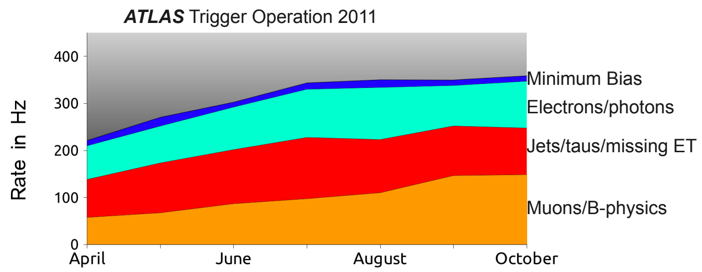


Figure 2.11: The HLT trigger output rate split by data stream [82].

processing times of 40 ms and 4 s. The HLT runs on CPU farms which allows easy maintenance of hardware and algorithms. The L2 trigger run on 'Regions of interest' (ROI) provided by the L1 trigger. It uses the full detail provided by the ATLAS sub-detectors but only using data from the ROI in question thereby using only 2% of the total event data. The EF triggers run using the full information from the complete detector.

Particular triggers may be 'pre-scaled', such that only one event in every N which passes the required conditions proceeds to the next level of the trigger[5]. For each ATLAS run a trigger menu is defined which contains the active triggers and their pre-scales. This must be carefully budgeted to allow sufficient events for all the channels of ATLAS's broad physics programme as seen in Figure 2.11. Details of the triggers used in this thesis are given in Chapter 3.

2.4 Data processing

The output from the high level trigger is divided into 'streams', with events segregated depending on which trigger(s) they fired[6]. A proportion of all events enter an 'express' stream which is processed with high priority so that data quality can be assessed within days of data taking. The data is segmented in time: the smallest unit of data taking

[5] As an LHC fill proceeds the instantaneous luminosity drops, due to collisions and beam losses, and so the trigger rate falls. To continue taking advantage of available band width trigger pre-scales are updated throughout the run. This is the job of the trigger shifter, a role I performed while at CERN.

[6] Data in this thesis is taken from the 'Egamma' and 'muon' streams.

is known as a 'lumiblock', which last $\mathcal{O}(1)$ minute; 'runs', which which typically last for one fill of the LHC; and 'periods', which are collections of runs during which the running conditions were the same. A summary of the run conditions for each period, including dates and run numbers is given in Appendix C. Data quality flags are set for each lumiblock to indicate which detector elements are in normal operating condition.

The data goes through a number of processing stages before being used for the analysis presented in this thesis. 'RAW' data from the detector, which consists of the readout from individual detector channels, is reconstructed into 'Event Summary Data' that groups the information into usable data such as tracks and clusters. This information is used by reconstruction algorithms to build the 'Analysis Object Data' (AOD), which groups the detector information into physics objects such as jets, electrons and muons. These AODs are transformed into a simple 'Ntuple' format, known within the collaboration as 'D3PDs' which can be read by the `ROOT` data analysis program [83]. These are customized by each physics group to provide the appropriate level of detail. Files from the `SM_WZ` ntuple reconstructed with ATHENA release 17 were used in the preparation of this thesis.

Data distribution is handled by the Worldwide LHC Computing Grid [84]. This is divided into tiers, with the tier-0 at CERN, a Tier-1 site for each geographical region, and many Tier-2/3 sites associated to each Tier-1. The RAW files are stored at Tier-0 and backed up to one Tier-1 site. Other data files (ESDs, AODs, D3PDs) are distributed to multiple Tier-1 and Tier-2 sites, with replication levels automatically managed to ensure popular datasets are available widely. To run analysis code users only need the name of the dataset and automated tools will find the location of datasets, and send the job to appropriate sites where the code will run.

2.5 Detector Simulation

Event records from the Monte Carlo samples are passed through a GEANT simulation of the ATLAS detector [85]. This uses a computer model of the ATLAS detector to simulate the interactions of particles as they propagate through the detector. Some of the interactions, such as nuclear recoil and hadronic excitation, are poorly understood theoretically and so empirical models are employed. Energy deposited in active regions of the detector is recorded in 'hits' files. The next stage is 'digitization' where the hits are converted to digits that are passed to the ATLAS read-out drivers. At this stage additional sets of hits are overlaid, to emulate the effect of multiple interactions (pile up) in the detector. This is done using PYTHIA 6 samples which contain low p_{T} QCD interactions known as 'min-bias'. These are overlaid on the hard interaction following a Poisson distribution whose average is the same as that found in data. The LHC bunch structure is also simulated such that additional interactions from neighbouring bunch-crossings are simulated. Effects of detector noise are also simulated. Trigger decisions are evaluated, but run in 'pass-through' mode such that no events are lost.

The simulation of Monte Carlo samples for comparison to 2011 data was carried out in four periods which correspond to the detector and accelerator conditions in periods D (0.186 fb^{-1}), E-H (1.072 fb^{-1}), I-K (1.341 fb^{-1}) and L-M (2.809 fb^{-1})[7]. This accounted for the influence of changing numbers of protons-per-bunch and defects in the liquid argon (LAr) calorimeter. The Monte Carlo events were weighted such that distribution of average interactions per bunch crossing (known as μ) equalled that found in data.

The simulated data are run through the same reconstruction algorithms as the recorded data. The end result is an event record which replicates real data event records, and thus Monte Carlo events can be passed through exactly the same analysis

[7]These luminosities correspond to the stable beam luminosity in the given period. This does not correspond to the luminosity passing data quality requirements which is the figure quoted elsewhere in this thesis.

code as data events.

One limitation of the so-called 'full sim' GEANT code is the large computational time, typically $\mathcal{O}(10\,\mathrm{s})$ per event. Because of limited computing resources a fast simulation package AtlFast-II (AFII) was developed [86]. This uses parametrized shower generation, rather than full Monte Carlo treatment of each particle's interactions. In this thesis AFII is only used in the simulation of systematic variations on the central Monte Carlo samples.

CHAPTER 3

Object reconstruction and definitions

Information from the detector must be fed into reconstruction algorithms to convert raw detector information into records of particles with momentum and energy. This chapter will describe how electrons, muons, jets and missing transverse energy are reconstructed and the subsequent quality cuts that are used to select these objects for use in this analysis. Tau leptons and photons are reconstructed in ATLAS datasets but not discussed here. A brief discussion of track reconstruction is also given.

3.1 Tracks

Charged particle tracks are reconstructed using space-points[1] from the pixel and SCT, and drift circles from the TRT [87]. Tracking begins from the inner radius, with seed

[1] In the SCT a distinction is made between 'hits' and 'space-points'. A 'hit' is any strip which records a signal, while a 'space-point' is formed in a module with a hit on either side from strips which cross. Space-points provide a more powerful constraint on a track, but hits still provide useful information.

tracks formed from the first three pixel layers and the first layer of the SCT. These are extended through the SCT. If the candidate track passes a defined set of criteria — such as goodness-of-fit, number of associated space points, number of SCT layers crossed without associated hits — then the track is extended to include information from the TRT. The track fit is then repeated with the full information from all three inner detector elements.

A complementary tracking system searches for unused track segments in the TRT and extends them back into the SCT. This improves the tracking efficiency for long-lived or secondary particles.

Finally, a vertex finding algorithm is used to associate tracks to primary vertices. This is followed by secondary and photon conversion vertex finding.

3.2 Electrons

3.2.1 Trigger

Events will only be available for study offline if they have been accepted by the trigger, as described in Section 2.3. Since high-E_{T} electrons are much rarer than jets in a hadron collider, they are excellent objects to trigger on. The operation and performance of the electron triggers is described in detail elsewhere [88] and only a brief summary of the single electron triggers is given here. The electron triggers used for the main result in this thesis are shown in Table 3.1. In each data period the lowest threshold unprescaled single electron trigger was used.

At L1 the trigger input is taken from 'trigger towers': coarse clusters of cells in the LAr calorimeter of size $\Delta\eta \times \Delta\phi \approx 0.1 \times 0.1$ including all layers of the calorimeter. Local maxima of towers are identified by using a *sliding window algorithm* with a window size of 4×4 trigger towers. The trigger is satisfied if two neighbouring cells in the cluster's central region have an energy sum greater than a threshold. This

Data period	L1 trigger	L2 trigger	HLT trigger
D-I	EM14	L2_e20_medium	EF_e20_medium
K	EM16	L2_e22_medium	EF_e22_medium
L-M	EM16VH	L2_e22vh_medium1	EF_e22vh_medium1

Table 3.1: Primary electron triggers used for each data period. The number in the name corresponds to the nominal threshold in GeV; 'medium' indicates that this trigger is close to the offline 'medium' level reconstruction; 'medium1' indicates that this trigger uses the reoptimized identification criteria. The 'vh' indicates that the hadronic leakage cut (described in the main text) has been applied.

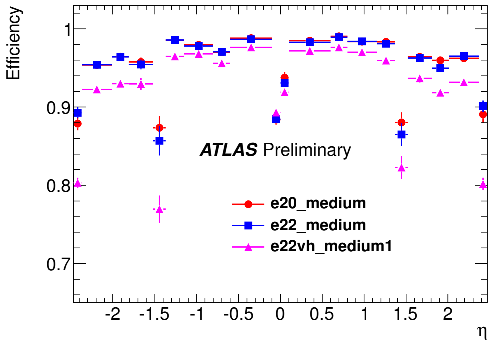
threshold can be redefined during running to control the trigger rate.

In the HLT the candidate window identified by L1 is reconstructed using a method closer to the offline electron reconstruction described in the following subsection.

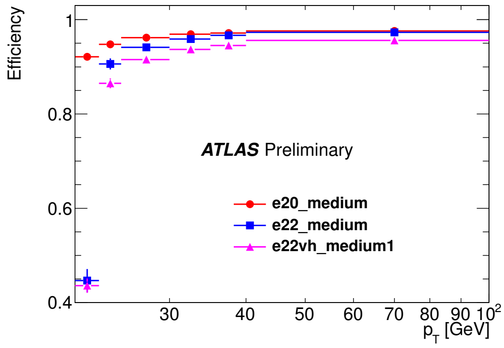
As the instantaneous luminosity increased during the 2011 run, constraints on the total trigger bandwidth required a tightening of the trigger selection at all levels in the trigger chain. This was achieved by increasing the electron energy threshold and tightening the electron identification cuts.

In period K increases in luminosity required the energy threshold to be raised in the L1 and HLT. Further increases in luminosity in periods L and M were dealt with by adding a hadronic leakage cut and allowing the E_{T} threshold to vary in η. This made it possible to raise the E_{T} threshold in regions with good electron energy response without degrading the HLT efficiency with respect to `medium++` electrons with $E_{\mathrm{T}} > 25$ GeV. The electron reconstruction cuts were also brought closer to the offline selection to avoid further increases in threshold.

The performance of the trigger was measured by the "Egamma" trigger group using a 'tag and probe' method [88]. The electrons passing the lowest unprescaled single electron trigger were identified as 'tags'. If a second reconstructed electron is found in the Z mass window of $80 < m_{ee} < 100$ GeV it is identified as the 'probe'. The probe can be used as an unbiased measure of the efficiency to pass a given electron trigger. The results of this efficiency study are shown as a function of E_{T} and η in Figure 3.1. The triggers used in this thesis are all over 90% efficient and operating close to their plateau after the offline `tight++` and $E_{\mathrm{T}} > 25$ GeV selections have been made. The mismodelling of the efficiency in Monte Carlo samples is corrected for using centrally produced scale factors.



(a)



(b)

Figure 3.1: The primary electron trigger efficiency as a function of offline `medium++` electron η and E_{T} [88]. The vertical bars represent combined statistical and systematic errors.

3.2.2 Offline reconstruction

Reconstruction begins in the LAr calorimeter. The calorimeter is divided into 'towers' of cells of size 0.025×0.025 in units of $\Delta\eta \times \Delta\phi$, which corresponds to the granularity of cells in the middle layer of the LAr, shown in Figure 2.7. A sliding window algorithm moves a window of 3×5 of these cells across the $\Delta\eta \times \Delta\phi$ grid and calculates the total transverse energy in that window. If this is greater than the threshold energy of 2.5 GeV, and it is a local maximum, then the window is recorded as a seed cluster.

A seed cluster will be reconstructed as an electron if at least one matched track is found. A track is considered to have matched a seed cluster if its impact point has $\Delta\eta < 0.05$ and $\Delta\phi < 0.1(0.05)$, where the relaxed $\Delta\phi$ requirement is allowed on one side to account for bremsstrahlung. The matched track is that with smallest ΔR to the seed cluster.

Once a matched track is established, the seed cluster is rebuilt using 3×7 towers (5×5 in the end cap), and the total energy calibrated based on the energy distribution in the different layers of the LAr. The presampler is used to estimate the energy loss upstream of the calorimeter, while the shower shape is used to estimate the lateral and longitudinal leakage, based on a comparison to Monte Carlo simulation.

The final reconstructed four-momentum of the electron is a combination of the cluster and track information, with the energy taken from the cluster, and the direction taken from the track at the primary vertex.

3.2.3 Identification

The above reconstruction still contains a large number of non-prompt electrons from hadronic jets, photon conversions and in-flight decays of heavy particles. To remove these a set of identification (ID) cuts have been derived by the Egamma group in 10 bins of cluster η and 11 bins of cluster E_{T} [89]. Three reference working points are produced for electrons with $|\eta| < 2.4$ with progressively greater background rejection and lower signal efficiency `loose++`, `medium++` and `tight++`[2]:

`loose++` makes cuts on the hadronic leakage and the shower shape in the middle layers of the electromagnetic calorimeter.

`medium++` includes all `loose++` cuts and in addition makes cuts on the shower width on the strips of the EM calorimeter, and the difference of the largest to second largest energy deposits in the cluster divided by their sum. Track-quality and track-cluster-matching requirements are added, including pixel-hit and impact-parameter cuts.

`tight++` includes all `medium++` cuts and in addition tightens the track cluster matching and impact parameter cuts. Cuts are made on the ratio of cluster energy to track momentum, a requirement is made for a b-layer hit, and a TRT particle ID is applied. Candidates matched to photon conversions are vetoed.

[2] The '++' is to differentiate these from the 2010 calibrations.

The cuts are designed based on Monte Carlo simulation and their efficiency has subsequently been validated on data using 'tag and probe' methods similar to that described in the trigger section above. A $Z \rightarrow ee$ or $J/\psi \rightarrow ee$ decay is used, with one electron reconstructed using the tightest identification requirements designated as the 'tag', while the other forms the 'probe'. The efficiency is then simply the number of probes passing a given cut over the total number of tags. The results of this study for 2010 data are available [89] with a 2011 publication in preparation [90]. The reconstruction efficiency in 2011 as a function of number of reconstructed vertices, which is a good approximation for the amount of pile up, is shown in Figure 3.2. The mismodelling of the reconstruction efficiency in Monte Carlo is corrected using centrally produced scale factors.

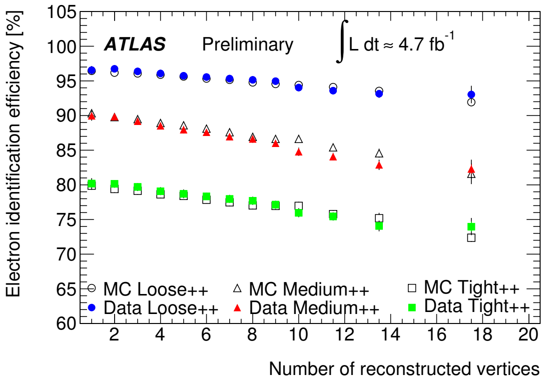


Figure 3.2: The electron reconstruction efficiency as a function of number of reconstructed vertices [91].

3.2.4 Energy scale and resolution calibration

The baseline energy scale is derived from test beam measurements [92] conducted on a single module of the LAr calorimeter. Electrons of known energy were fired into the calorimeter and the electronic signal was measured. This calibration is corrected for leakage and absorbing material using a Monte Carlo correction as described above in Section 3.2.2. This is not expected to exactly replicate the performance of the LAr calorimeter when operational in ATLAS, due to larger variations in the liquid argon temperature and other non-uniformities in the detector. A residual correction

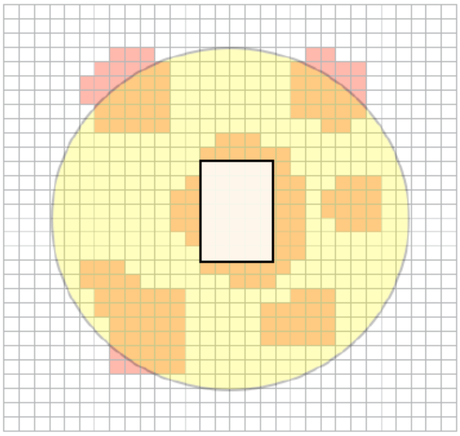


Figure 3.3: The calorimeter isolation scheme. The central 5×7 rectangle contains the majority of the candidate electron energy. The yellow circle indicates a cone of ΔR [93].

is therefore derived using $Z \rightarrow ee$ and $J/\psi \rightarrow ee$ events and fitting the observed line shape to that expected from Monte Carlo, allowing the residual energy correction parameters to float. This is performed in bins of η and the residual correction is less than 2%.

The systematic uncertainty on the energy calibration is between 0.3% and 1.6%, with the lowest uncertainty in the barrel region for electrons with $E_{\mathrm{T}} \approx 40$ GeV [89].

3.2.5 Isolation

Requiring electron candidates to be isolated in the detector can provide additional background rejection power. Two measures of isolation are used in this thesis:

Calorimeter based isolation: $E_{\mathrm{T}}^{\mathrm{coneXX}}$ is the sum of all energy deposits in the EM and hadronic calorimeters in a radius of 0.XX in $\Delta\eta \times \Delta\phi$, where XX is a cone size parameter that can be varied after subtracting the energy measured in the 5×7 cells that measure the electron energy [93]. The expected electron shower leakage into surrounding cells is taken from Monte Carlo and subtracted. A schematic of this is shown in Figure 3.3.

Track based isolation is defined as the scalar sum of the transverse momenta of inner detector tracks within a given radius of the electron track satisfying track p_{T}, quality and impact parameter cuts.

The impact parameter cuts make this definition robust against pile up as the tracks must come from the same vertex as the hard scatter. This isolation is referred to as $p_{\mathrm{T}}^{\mathrm{coneXX}}$ where 0.XX is the cone radius. More details can be found in Reference [94].

3.2.6 Analysis electron definition

Identification	`tight++`
Transverse momentum	$E_{\mathrm{T}} > 25$ GeV
η	$\lvert\eta\rvert < 2.4$
excluding crack	$1.37 < \lvert\eta\rvert < 1.52$
Isolation	`Eiso(ptcone40)` and `Eiso(etcone20)`
Cleaning cuts	No global LAr error flags
	Electron object quality bit mask passed

Table 3.2: Analysis level selection applied to electrons.

In the analysis presented in this thesis electrons must satisfy the criteria listed in Table 3.2. For the isolation the `Eiso` tool provided by the Egamma group is used [39]. This provides a variable isolation threshold as a function p_{T} and η to provide uniform efficiency of 97% and 98% for calorimeter and track based isolation respectively. Error flags are set if there is a noise burst in the liquid argon calorimeter and the electron candidate is rejected if there are problems in any of the cells which the electron reconstruction uses.

During data periods E to I, six front-end boards responsible for reading out cells in layers 2 and 3 of the LAr in the region $-0.1 < \eta < 1.5$, $-0.9 < \phi < -0.5$ failed. This resulted in a loss of efficiency in reconstructing electrons and low energy jets. From period I onwards four of the front-end boards responsible for layer two were recovered. Electrons falling in the affected region were removed using the electron quality cuts.

A central package `egammaAnalysisUtils-00-03-99` is provided by the Egamma performance group, and this is used to apply final corrections to data and Monte Carlo including a correction for mismeasurement of the energy near to the crack region. In the Monte Carlo the energy is smeared to reproduce the resolution observed in data. This tool also provides the thresholds for the isolation cut and scale factors to correct the identification and trigger efficiencies.

The systematic uncertainties on the electron energy scale, resolution, reconstruction efficiency, identification efficiency and trigger efficiencies are evaluated using the Egamma group `egammaAnalysisUtils-00-03-99` package. The uncertainty is propagated to the analysis by varying the corrections by one standard deviation and repeating the event selections and signal fits.

The Monte Carlo slightly mismodels the isolation distribution observed compared to data. This discrepancy has been measured by the inclusive W boson and Z boson cross-section analysis [95] and found to be $< 1\%$, with an uncertainty $\approx 1\%$. This is much smaller than the other uncertainties in this analysis and is neglected.

3.3 Muons

3.3.1 Muon Trigger

Muons interact very little in the detector and will typically pass through the inner tracker, calorimeter and muon spectrometer, leaving only hits in the tracking chambers and minimum ionizing deposits in the calorimeter. They therefore make very good objects on which to trigger.

At L1, signals from the Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) are used to search for coincident hits at six predefined p_{T} thresholds, using pre-calculated expected track-paths — or ‘roads’ — through the detector. Example ‘roads’ can be seen in Figure 2.10. In data periods D-I the requirement was for two (three) coincident hits in the barrel (endcap), forming a trigger known as `L1_MU10`.

Due to increased luminosity, data periods J-M required three coincident hits in both barrel and end cap, known as `L1_MU11`.

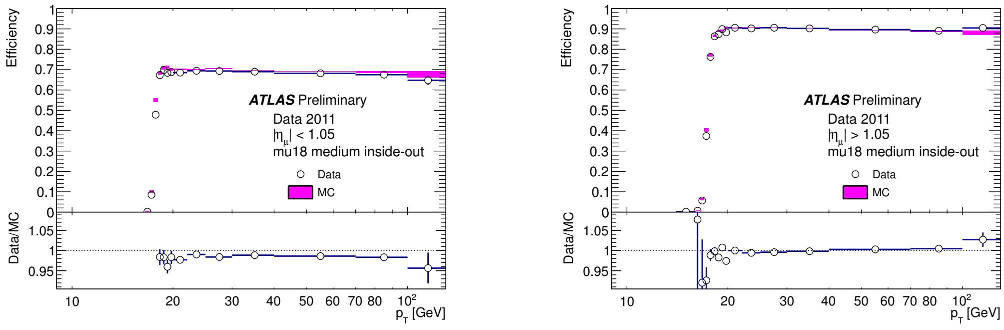
At L2, the tracks in the Region of interest (RoI) identified at L1 are reconstructed using fast reconstruction algorithms that include information from all muon detectors. Stand alone tracks using only information from the muon spectrometer are then combined with tracks from the inner detector to form so-called 'combined' candidates. If these candidates pass goodness-of-fit and p_{T} threshold cuts then the L2 trigger is passed.

At EF, information from the full detector is available. Candidates are identified either by reconstructing track segments in the muon spectrometer only and extrapolating them to the ID (outside-in) or starting with ID tracks and extrapolating them out to the muon spectrometer (inside-out). The performance of the two algorithms is very similar and for this thesis the inside-out HLT trigger is used. For all periods of 2011 data taking the EF threshold for the lowest available unprescaled single muon trigger was 18 GeV. This information is summarized in Table 3.3.

The efficiency of the trigger was measured using $Z \rightarrow \mu\mu$ tag and probe methods and the results for the inside-out trigger is shown in Figure 3.4. The plateau of 68.6% in the barrel and 89.9% in the end-cap is reached well below the offline analysis cut of 25 GeV. The low efficiency in the barrel is due to geometrical acceptance caused by the gap at $\eta = 0$ for calorimeter and inner detector services and by the detector feet. The small mismodelling of the efficiency by Monte Carlo is corrected for by centrally produced scale factors in bins of p_{T}, η and ϕ, and this is applied using the package `TrigMuonEfficiency-00-02-19` [81].

Data Period	L1 trigger	L2 trigger	HLT Trigger
D-I	L1_MU10	L2_mu18_MG	EF_mu18_MG
J-M	L1_MU11	L2_mu18_MG_medium	EF_mu18_MG_medium

Table 3.3: Primary triggers used for each data period for the results in this thesis. The 'MG' in the name signifies that this trigger used the inside-out algorithm. The 'medium' in the name indicates that this trigger is seeded by the `L1_MU11` trigger.



(a) Barrel efficiency (b) End-cap efficiency

Figure 3.4: Efficiency of the `EF_mu18_MG_medium` trigger in terms of the off-line reconstructed muon p_{T}. The error bars show the statistical uncertainties [81].

3.3.2 Offline reconstruction

The detector information can be combined in multiple ways to reconstruct muons. The `STACO`[3] algorithm [96] is used in this thesis and described briefly below.

A track from the inner detector and a track from the muon spectrometer are independently fitted, giving a p_{T} measurement from both detectors. Tracks are then paired up and these pairs are retained as combined muons if their tracks are sufficiently similar, after correcting for energy loss expected between the ID and Muon Spectrometer (MS) track. The combined four-vector is created from a statistical combination of the two track segments, giving more weight to the track segment with better resolution at the candidate muon's p_{T}. For $p_{\mathrm{T}} \lesssim 80$ GeV the ID track dominates. For 80 GeV $\lesssim p_{\mathrm{T}} \lesssim 100$ GeV the ID and MS have equal weights, and for higher p_{T} the MS dominates.

3.3.3 Identification and analysis definition

A large fraction of the objects reconstructed as muons will not be prompt muons from the interaction point, but due to decays-in-flight of hadrons. The cuts listed in Table 3.4 are recommended by the muon performance group to increase the prompt

[3] `STACO` stands for 'Statistically Combined' and is known as 'chain 1' in publications.

Kinematic acceptance	
Transverse momentum	$p_{\mathrm{T}} > 25$ GeV
Pseudorapidity	$\lvert\eta\rvert < 2.4$

Inner detector hit requirements	
ID silicon hits:	b-layer hits > 0, pixel hits ≥ 1, SCT hits ≥ 6
TRT for $\lvert\eta\rvert \leq 1.9$	hits + outliers > 5 and $\frac{outliers}{Hits+outliers} < 0.9$
TRT for $\lvert\eta\rvert > 1.9$	if(hits + outliers > 5) then $\frac{outliers}{Hits+outliers} < 0.9$

Isolation and impact parameter	
Transverse impact parameter significance	$d_0/\sigma(d_0) < 3$
Longitudinal impact parameter	$\lvert z_0\rvert < 10$mm
Isolation	$p_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}} < 0.1$

Table 3.4: Summary of the identification cuts for muons used in this thesis. The kinematic cuts are applied to the combined muon quantities, while the impact parameter cuts are with respect to the leading primary vertex of the event. The definition of track based isolation follows the same procedure as that described in the electron description above in Section 3.2.5.

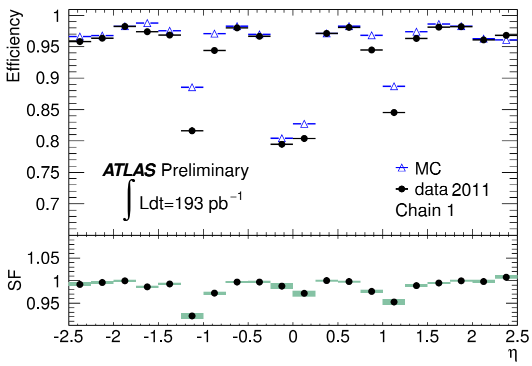


Figure 3.5: The muon reconstruction efficiency as a function of η [97]. 'SF' is the Data/Monte Carlo ratio which gives the 'scale factor' that recovers the data efficiency from the Monte Carlo.

muon purity and are applied to muons selected in this thesis.

The efficiency of muon reconstruction was measured by the muon performance group using a $Z \to \mu\mu$ tag and probe method shown in Figure 3.5 as a function of η. The small mismodelling of data by Monte Carlo is corrected for by applying scale factors derived by the muon group using the analysis package tag `MuonEfficiencyCorrections-02-01-00`, while the momentum scale and resolution is corrected for with the package `MuonMomentumCorrections-00-07-00`.

Systematic uncertainties on muon reconstruction are estimated using central packages from the muon performance group. The uncertainty on the trigger efficiency is estimated using the package `TrigMuonEfficiency-00-02-19`. The uncertainty on the reconstruction efficiency is estimated using the package `MuonEfficiencyCorrections-02-01-00`. The uncertainty on the momentum scale and resolution is estimated separately for the inner detector and muon spectrometer components using the package `MuonMomentumCorrections-00-07-00`. Scale factors to correct the isolation efficiency in simulation compared to data and associated systematic uncertainties are not included. The inclusive W and Z boson cross section analysis [95] has assessed this for a slightly different cone definition of 0.4 and find it to be $< 1\%$.

3.4 Hadronic jets

Hard parton emissions from proton-proton collisions result in showers of hadronic particles. Jet algorithms are used to group these showers into single objects — jets — which can be compared to truth level objects. Since meaningful comparisons between theoretical calculations and experimental observation are vital for discovery and measurement, the jet reconstruction should use an algorithm that can be applied at parton level, i.e. partons from matrix element calculations; at particle level, i.e. taking all particles after Monte Carlo showering; and at detector level, e.g. energy deposits in calorimeters, or tracks from the inner detector.

The jet algorithm should also be theoretically robust against soft emission and collinear emission — so called infra-red and collinear safety. If these were not satisfied then the correspondence between observed jets and matrix element calculations would be sensitive to non-perturbative details such as hadronization and fragmentation.

ATLAS uses the anti-k_t jet algorithm [98] built using FASTJET software [99]. This proceeds by making a list of all input entries[4]. A distance measure, $d_{i,j}$, is introduced

[4] The non-specific term ‘entry’ is used, since the description of the algorithm is generic whether the ‘entry’ is a parton, particle or detector quantity such as a calorimeter deposit.

which measures a distance between input objects i and j, and $d_{i,B}$ between object i and the beam. The clustering proceeds by calculating the distances between all entries — all the $d_{i,j}$ and $d_{i,B}$— and then selecting the smallest distance. If this distance is a $d_{i,j}$, then the objects i and j and combined and the process repeated. If the smallest distance is a $d_{i,B}$ then the item i is defined as a jet and removed from the list of entries. The distance measure $d_{i,j}$ is defined as:

$$d_{i,j} = \min(p_{Ti}^{-2}, p_{Tj}^{-2})\frac{\Delta R_{ij}^2}{r^2}, \tag{3.1}$$

$$d_{i,B} = p_{Ti}^{-2}, \tag{3.2}$$

where p_{Ti} is the transverse momentum of the i^{th} entry and ΔR_{ij} is the ΔR between i and j. r is the "distance parameter" and can be varied — 0.4 is used in this thesis.

For a given ΔR two soft particles will have larger $d_{i,j}$ than a hard and soft particle, since the p_{T}^{-2} term will dominate. This leads to hard items collecting up all the soft items around them. In the case of well separated hard emissions this results in conical jets with radius r. For close by hard emissions, if $p_{Ti} \approx p_{Tj}$ then the boundary between the jets will be a straight line. If $p_{Ti} > p_{Tj}$ then the i^{th} jet will be larger and more uniformly conical.

This algorithm is known to be infra-red and collinear safe [98]. It also has the favourable property of *soft resilience*, i.e. soft radiation does not affect the shape of the jet. This makes detector level corrections simpler, since jets are a more uniform shape compared to other possible algorithms such as the k_T [100] or Cambridge/Aachen [101].

Jets in ATLAS are built from topological clusters (topo-clusters) of calorimeter cells [102]. These are formed by considering the signal to noise ratio (S/N) of the cells of all of the ATLAS calorimeters. The noise is taken to be the RMS of the electronics noise added in quadrature with the noise from pile up fluctuations. Cells with $S/N > 4$ are taken as seeds, and all neighbouring cells with $S/N > 2$ are added iteratively to the cluster. Finally all neighbouring cells are added to the cluster.

If there are local maxima in cell energy with $E > 500$ MeV then these are used as new seeds for a second iteration of clustering, with the energy from neighbouring cells in the resulting topo-clusters shared by energy weights. Note that topo-clusters can have a variable number of cells in contrast to the fixed sliding window used for electron clusters. Also note that topo-clusters are based solely on calorimeter information, and so any particle that deposits energy is likely to be formed into a jet. Thus electrons and muons must be removed from the jet collection at the analysis level.

A topo-cluster's energy is taken to be the sum of all included cells. This may be calibrated to the 'EM scale', which is the scale derived from single particle electron test beams, or alternatively to a local cluster weighting (LCW) scale, which assigns a weight based on whether a cluster is identified as electromagnetic or hadronic based on energy density and shower shape. The latter improves the resolution of the cluster. The cluster's direction is taken to be a energy weighted average of the η and ϕ of the included cells relative to the centre of ATLAS. They are assumed massless (or equally $E_{\mathrm{T}} = p_{\mathrm{T}}$).

Once jets have been built they must be calibrated to reproduce the true energy of the jet. The calorimeter cells are calibrated from test beams to the electromagnetic energy scale. In an electromagnetic shower all of the energy deposited in calorimeters is visible as ionization. However the hadronic response is expected to be different since these include non-ionizing interactions, such as nuclear recoil and excitation, which produce no visible signal in the detector. The impact of energy lost due to lateral and longitudinal leakage and dead material must also be accounted for. The procedure for setting the Jet energy scale (JES) proceeds as follows:

1. **The pile-up offset correction** is applied to remove the impact of additional interactions, resulting in a jet response that is uniform as a function of number-of-primary-vertices and μ;

2. **The origin correction** is applied to correct the direction of the jet to point to the primary vertex of the interaction, rather than the centre of the detector.

3. **The energy calibration** is applied by means of a simple correction which relates the reconstructed calorimeter jet energy to the particle-level jet energy. The particle-level jet definition excludes muons and non-interacting particles, since these leave no trace in the calorimeter. The correction is derived using isolated jets from an inclusive jet Monte Carlo sample including pile up events.

4. **The residual *in-situ* calibration** is applied, which uses measurements of collision data to derive a residual correction. This is described in the following section.

If the jet is built from EM calibrated topo-clusters then it is referred to as `EM+JES` calibrated. If it is built from LCW calibrated topo-clusters then it is referred to as `LCW+JES` calibrated.

3.4.1 *In-situ* calibration by dijet balance

A variety of *in-situ* methods have been used at ATLAS to further constrain the jet energy scale. These use the expected p_{T} balance between physics objects to derive a residual correction to the calibration of jets. This requires a well measured reference object to balance the probe jet. The choice of reference object is driven by the rapidity and p_{T} of the probe jet. In the central regions of the detector, jets may be balanced with leptons (Z-jet balance [103]), photons [104] or multiple jet balance [105]. However the forward region, $|\eta_{det}| > 2.8$[5], is only instrumented with calorimeters, so lepton and photon reconstruction efficiencies are low and in addition their cross section are very small compared to dijet production. The only viable method in this region is dijet balance, where a relatively well measured central jet is used to balance a forward jet.

[5] η_{det} is the η with respect to the origin of the ATLAS detector, rather than the leading primary vertex usually used for the η of a jet.

VBF measurements require good calibration of forward jets and so I worked on this method. Here I document the areas I worked on. The final result is documented in Reference [106].

Results are produced for two cone sizes R=0.4 and R=0.6 and two calibrations: the EM calibrated clusters plus jet energy scale (`EM+JES`) calibration known as 'TopoEM', and the Local cluster weighted clusters plus jet energy scale (`LCW+JES`) calibration known as 'LCTopo'. All selections are made on jets after the pile up offset correction, origin correction and initial energy calibration defined in the previous subsection.

Methodology

From each dijet event the asymmetry A is calculated:

$$A = \frac{p_T^{\text{probe}} - p_T^{\text{ref}}}{p_T^{\text{avg}}} \text{ where } p_T^{\text{avg}} = \frac{p_T^{\text{ref}} + p_T^{\text{probe}}}{2}. \tag{3.3}$$

The reference region is defined as $|\eta_{det}| < 0.8$ which corresponds to the central barrel of the tile calorimeter (cf. Section 2.2.3).

The asymmetry is used to measure the response of the probe jet relative to the reference jet. This is known as the η-intercalibration factor c, which will restore the probe jet p_{T} to that of the reference jet:

$$\frac{p_T^{\text{probe}}}{p_T^{\text{ref}}} = \frac{2+A}{2-A} = \frac{1}{c}. \tag{3.4}$$

The average asymmetry $< A_{ik} >$ is calculated in $i = 16$ bins of η_{det}, and $k = 11$ bins of p_{T}. From this we derive the relative response c_{ik}:

$$c_{ik} = \frac{2- < A_{ik} >}{2+ < A_{ik} >}. \tag{3.5}$$

If both jets fall in the reference region then the asymmetry is filled twice, once with each jet as probe and reference. Thus the average response in the central region is one by construction.

The uncertainty on this quantity is defined to be RMS/$\sqrt{N}$, where N is the number of events for data and the square-root of the sum of the weights squared for Monte Carlo. An example distribution is shown in Figure 3.6.

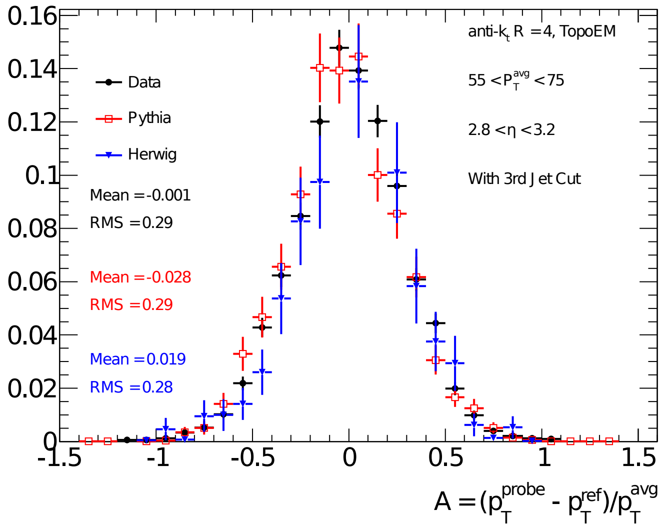


Figure 3.6: Example asymmetry distribution.

Even for a perfectly measured true dijet event the response may not be one. This is due to the underlying event and intrinsic k_t of the initial state partons which will give rise to an p_{T} imbalance in the event. An additional source of imbalance is multi-jet events, misidentified as dijet events. The dijet balance method attempts to correct for mismeasurement of jet p_{T}. As a result of these other effects the method requires a comparison to Monte Carlo simulation to understand the level of imbalance expected. The following sub-sections describe the method used to select the events of interest, and to calculate the response.

Monte Carlo samples

Data is compared to dijet Monte Carlo samples. The central sample is generated using Pythia v6.423 [20], which is the sample used for the Monte Carlo based JES. This uses p_{T}-ordered parton showers. The modified leading-order PDF set MRST LO* [107] is used with the ATLAS mc11 AUET2B MRST LO** tune.

HERWIG++ v2.5.2 [108] is used to provide a comparison sample. This uses angular ordered parton showers and a different model of multiparton interactions. The same PDFs and tune as PYTHIA are used.

These samples are produced in bins of $\hat{p}_T$, the minimum allowed parton p_{T}. This is due to the large variation in cross section as a function of parton p_{T}, which makes it impractical to generate statistics comparable to those collected by the high p_{T} triggers with a fully inclusive lower bound on parton p_{T}. The $\hat{p}_T$ bins are given in Table 3.5. The Monte Carlo events are weighted to the cross section of the sample they are generated from multiplied by luminosity of the period to which they are compared.

Pile up is modelled as described in Section 2.5, with the exception that PYTHIA 8 is used to simulate the pile up in place of PYTHIA 6[6].

Sample Name	J0	J1	J2	J3	J4	J5	J6	J7	J8
min $\hat{p}_T$ (GeV)	8	17	35	70	140	280	560	1120	2240
max $\hat{p}_T$ (GeV)	17	35	70	140	280	560	1120	2240	none

Table 3.5: $\hat{p}_T$ cuts on the parton level for PYTHIA and HERWIG++ dijet samples

Selection

Events are required the pass the following selection:

- Good run list[7]: only events taken during periods of stable running with relevant sub-detectors operating are considered;
- Monte Carlo quality cuts (see Section 3.4.1);
- At least one trigger. Single-jet event-filter triggers detailed in Table 3.6 are used;
- p_T^{avg} in efficient region for relevant trigger, see following subsection;

[6]In the ATLAS terminology these samples were from the `mc11b` production round.

[7]`data11_7TeV.periodAllYear_DetStatus-v36-pro09-03_CoolRunQuery-00-04-08_JetEtMiss.xml`

- Primary vertex with at least 5 tracks matched;
- ≥ 1 Jet $p_T^{\text{avg}} > 22$ GeV;
- LAr hole veto: If a jet points at the area of the detector with the lAr hole then the event is vetoed;
- jet cleaning cuts to suppress non-beam backgrounds;
- Third jet cuts: reject events in which a third jet is inferred, see Section 3.4.1.

Trigger

Events are selected in the $\geq 99\%$ efficiency region of the triggers used. For each p_T^{avg} bin a pair of triggers are used: an inclusive trigger which accepts events with jet p_{T} above threshold across the whole detector and a forward jet trigger which accepts events only if there is a jet above threshold in the forward region of the detector[8]. These triggers are prescaled and a weight is applied to return the number of events expected without prescale. The triggers used are listed by p_T^{avg} bin in Table 3.6. The total integrated luminosity for each trigger is given as an indication of the prescale level of each trigger.

There is an overlap in the phase space covered by these two triggers, since a jet in the forward region above threshold may pass either the inclusive trigger or the forward trigger. To account for this we check the raw trigger decision made before prescale when deciding which prescale weight to apply to the event, following the method suggested in Reference [109]. The prescales are not easily accessible on an event level and so the average prescale is used.

If the forward jet trigger passes after prescale then the event is weighted by 1/(fwd prescale). If the central jet trigger passes after prescale and the raw decision of the forward jet trigger fails (i.e. the trigger decision before prescale) then the event

[8]'Forward' is defined as the region $|\eta_{det}| > 3.2$.

p_T^{avg} (GeV)	Trigger Chain L1	L2	EF	Total Integrated Luminosity (nb^{-1})
22-30	zero-bias	zero-bias	j10	18.3
	zero-bias	zero-bias	fj10	96.3
30-40	zero-bias	zero-bias	j15	81.5
	zero-bias	zero-bias	fj15	32.2
40-55	zero-bias	zero-bias	j20	251
	zero-bias	zero-bias	fj20	400
55-75	j10	j25	j30	348
	fj10	fj25	fj30	3,626
75-100	j15	j35	j40	417
	fj10	fj25	fj30	3,626
100-130	j30	j50	j55	1,480
	fj30	fj50	fj55	58,600
130-170	j50	j70	j75	5,380
	fj50	fj70	fj75	690,000
170-220	j75	j95	j100	18,500
	fj75	fj95	fj100	4,500,000
220-300	j75	j95	j135	162,000
	fj75	fj95	fj100	4,500,000
300-400	j75	j95	j180	1,229,000
	fj75	fj95	fj100	4,500,000

Table 3.6: Triggers used for each p_T^{avg} range and the total luminosity gathered. The number in the name of the jet refers to the nominal threshold which is at EM scale. ‘Zero-bias’ triggers accept events from any filled bunch crossing after a prescale.

is weighted by 1/(central prescale). If the central jet trigger passes after prescale, and the raw decision of the forward jet trigger passes then the event is vetoed, since this phase space is already included in the forward jet trigger weights applied. Including this event would lead to double counting which cannot be correctly accounted for as event-by-event prescales are not available.

Results were produced separately in p_T^{avg} bins such that any particular bin is only filled from one pair of jet triggers. This method is adopted to minimize any residual trigger biases which could be introduced by using triggers with different prescales.

Third Jet Cuts

If an event is not a true dijet topology, i.e. if there is significant radiation originating from the primary vertex that is not captured in either of the two leading jets, then one would not expect the response calculated from the leading jets to be equal to unity. It is therefore important that these events are removed from the calculation of A.

In true dijet events, the two jets will be produced back-to-back in ϕ, and so a cut of $\Delta\phi(j_1, j_2) > 2.5$ is used. The cut value was changed by ± 0.4 to evaluate the systematic uncertainty on this choice.

Additional jets in the event may be due to a non-dijet primary scatter, or may be due to pile up interactions. The Jet vertex fraction (JVF) allows a distinction to be made [110]. Each jet has a set of tracks associated with it and each track is associated with a vertex. The JVF for a given jet is defined as the scalar sum p_{T} of its tracks associated to the leading primary vertex divided by the sum p_{T} of all tracks associated to the jet. This is 1 for a jet with all tracks associated to the leading primary vertex and 0 for a jet with none of its tracks associated to the leading primary vertex. Jets with no tracks are assigned a JVF of -1. The JVF is only effective for $|\eta_{det}| < 2.5$ since it requires tracking information.

p_{T}-based cuts are then applied on any third jet in the event, which vary depending on η_{det}. In the central region the JVF is used to distinguish between third jets originating from the primary interaction and third jets originating from pile up. These are summarized in Table 3.7.

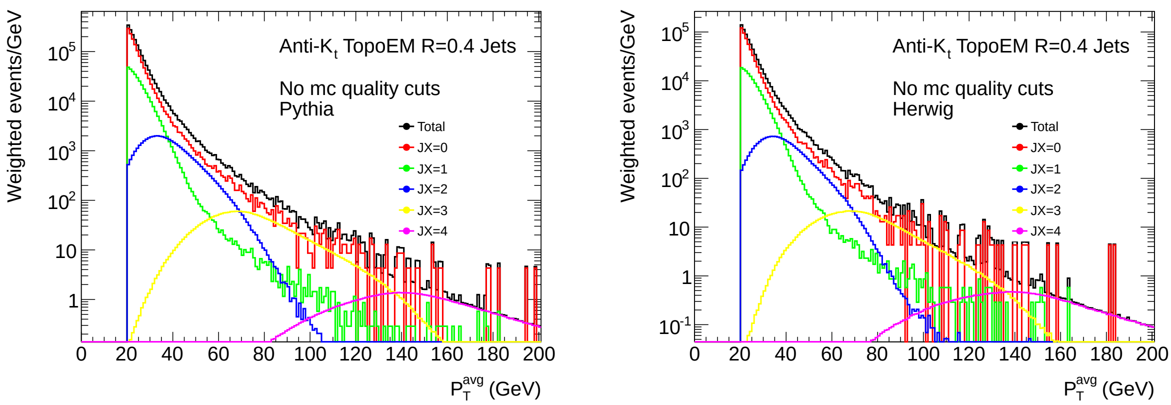
Variable	$\|\eta_{det}\| < 2.5$ Selection	$\|\eta_{det}\| > 2.5$ Selection
$\Delta\phi(j_1, j_2)$	> 2.5	> 2.5
No third jet which satisfies	$\|\mathrm{JVF}(j_3)\| > 0.6$ $p_{\mathrm{T}}(j_3) > \max(0.25 p_T^{\mathrm{avg}}, 12\ \mathrm{GeV})$	$p_{\mathrm{T}}(j_3) > \max(0.20 p_T^{\mathrm{avg}}, 10\ \mathrm{GeV})$

Table 3.7: Third jet cuts in the different η_{det} regions.

Monte Carlo quality cuts

In Monte Carlo samples, pile up events are simulated by overlaying minimum bias events simulated by Pythia 8 onto the hard event. The minimum bias events do not have an upper bound on $\hat{p}_T$, so those that overlap above the $\hat{p}_T$ in a given sample must be removed. This can be seen in Figure 3.7 by examining the red curve for the J0 sample, which is seen to extend far beyond its nominal cut of $\hat{p}_T < 17$ GeV given in Table 3.5. All events in a sample are weighted with the cross section of the nominal hard scatter of that sample, and so events with a pile up event which is harder in p_{T} than the 'hard scatter' will be weighted with the wrong cross section. This results in large discontinuities in the otherwise smoothly falling p_T^{avg} spectrum shown in Figure 3.7.

This problem is most important in the HERWIG++ samples which are included in order to provide an alternative showering model. The PYTHIA minimum-bias overlay is polluting the HERWIG++ result up to $p_T^{\mathrm{avg}} \sim 80$ GeV and so a false degree of agreement will be found in these regions.

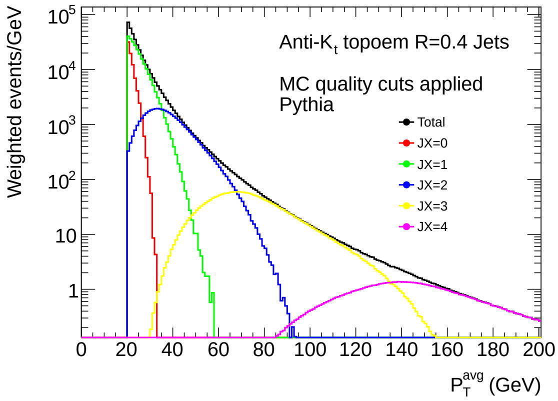


(a) PYTHIA (b) HERWIG++

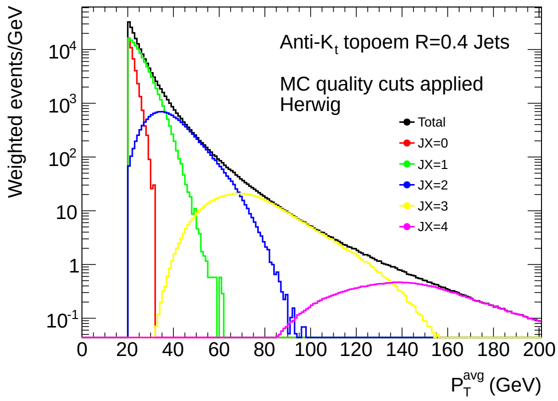
Figure 3.7: p_T^{avg} for PYTHIA and HERWIG++ sample. The individual JX samples are shown in different colours and the sum distribution is shown in black.

The most satisfactory method to address the overlap would be to include the pile up contribution in a manner that took into account the p_{T} scale of the sample[9]. However as a short term solution to this problem I investigated various cuts that could be made using the Monte Carlo truth information to remove the overlap events, while preserving Monte Carlo statistics. The following quantities were investigated:

[9]This has now been done for a 2012 set of ATLAS dijet samples. See internal ATLAS Reference [111].



(a) Pythia



(b) Herwig++

Figure 3.8: p_T^{avg} for Pythia and Herwig++ sample after Monte Carlo quality cuts have been applied.

- $\hat{p}_T$, the p_{T} of the hard scatter at Monte Carlo truth level;
- $p_{\mathrm{T}}^{\text{max truth}}$, the highest p_{T} jet in the truth collection not including the jets from pile up.
- p_T^{avg}, calculated from reconstructed quantities as defined in Equation (3.3).

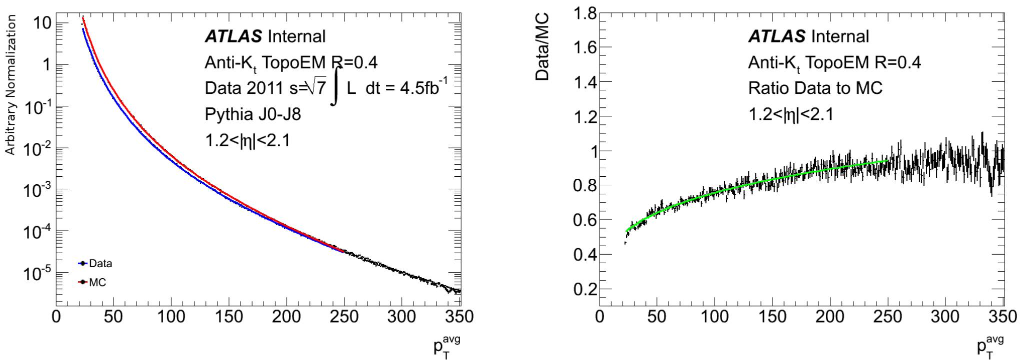
In events spuriously dominated by min-bias overlay the $p_{\mathrm{T}}^{\text{max truth}}$ will be much lower than p_T^{avg}, so ratios of these will identify problematic events. In addition it was found a few events had $p_{\mathrm{T}}^{\text{max truth}} \gg \hat{p}_T$ which gives rise to some of the spikes seen in Figure 3.7. The following cuts were applied:

- $0.6 < \frac{p_T^{avg}}{p_{\mathrm{T}}^{\text{max truth}}} < 1.4$,
- $0.6 < \frac{p_{\mathrm{T}}^{\text{max truth}}}{\hat{p}_T} < 1.4$.

The result of these cuts can be seen in Figure 3.8.

p_T^{avg} Systematics

Applying the Monte Carlo quality cuts detailed in the above section changes the p_T^{avg} shape of the Monte Carlo and this could potentially bias our Monte Carlo description of the response. Even without the Monte Carlo quality cuts one would wish to check whether the p_T^{avg} shape was well modelled and how it impacted the asymmetry. In order to assess the impact of such effects I compared fits of the data and Monte Carlo p_T^{avg} spectra. The data p_T^{avg} spectrum was taken from the same dataset as the main result with the normalization of data from each trigger-p_T^{avg} bin scaled so that adjacent trigger-p_T^{avg} bins matched and a continuous spectrum was achieved. The p_T^{avg} shapes were fitted with a function of form $x^{-a-bx-cx^2}$ in five $|\eta_{det}|$ bins with boundaries: 0, 1.2, 2.1, 2.8, 3.6 and 4.5. The functional form was empirically chosen. The ratio of the fitted data function to the fitted Monte Carlo function was taken as a weighting function and applied as event weights to the main analysis. The response, $1/c_i$, of this weighted analysis was compared to the unweighted analysis and found to be $< 0.3\%$ in almost all of the bins. Examples of the data and Pythia fits and ratio are shown for one η_{det} bin in Figure 3.9 and the goodness-of-fit for each η_{det} bin is shown in Table 3.8. An summary for all η_{det} is shown in Figure 3.10. Overall the contribution of this systematic uncertainty was considered negligible and is not further accounted for.



(a) Comparison of data and Pythia p_T^{avg} spectra.

(b) Ratio of data to Pythia.

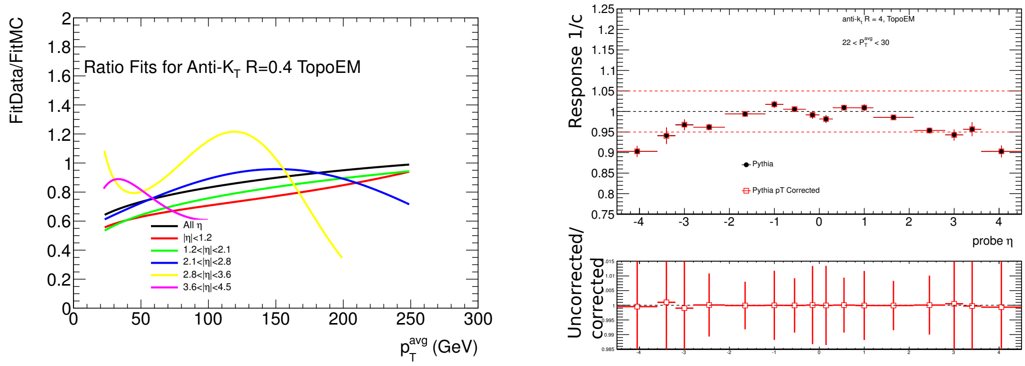
Figure 3.9: Example fits for p_T^{avg} reweighting.

Results

A sample of results for the response of anti-k_t $R = 0.4$ using the 'TopoEM' calibration is shown here. This is the jet algorithm used for results in the remainder of this thesis.

$\lvert\eta_{det}\rvert$ range	χ^2/Number of degrees of freedom
0 – 1.2	1.81
1.2 – 2.1	3.61
2.1 – 2.8	3.24
2.8 – 3.6	3.92
3.6 – 4.5	3.10
all	12.1

Table 3.8: Distribution of χ^2/Number of degrees of freedom found in the p_T^{avg} reweighting functions.



(a) Reweighting function (b) Example comparison

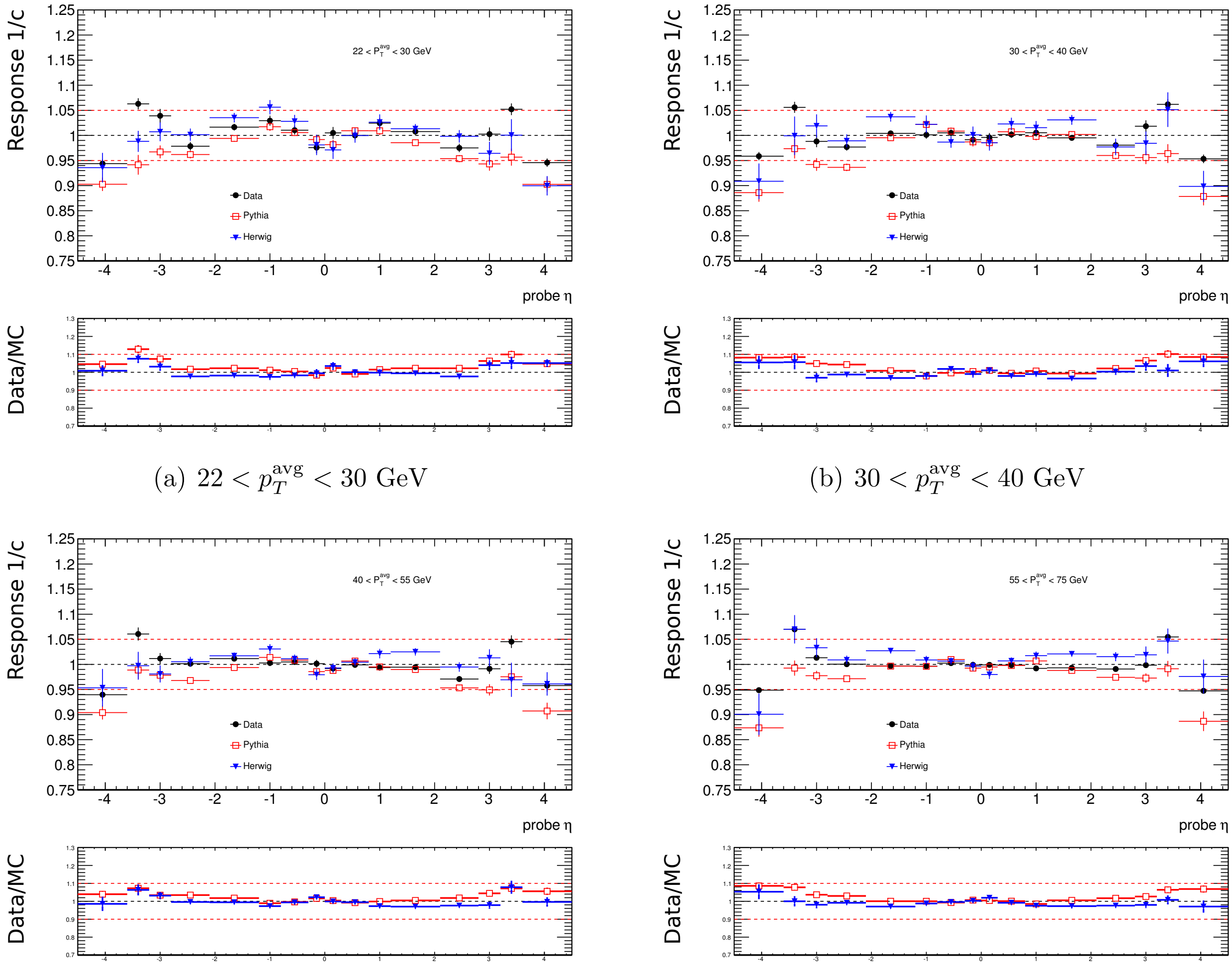
Figure 3.10: The reweighting functions used for PYTHIA anti-k_t 4 TopoEM jet and an example of the effect this has on the response. Note the scale on the ratio plot.

Results for the other jet algorithms were also produced and results from the 'LCTopo' algorithm are shown for comparison in Appendix A.

Figure 3.11 shows the relative response as a function of η_{det} and for the lowest four p_T^{avg} bins. The FCal-HEC transition region is clearly observed in the region $2.8 < \eta_{det} < 3.6$.

Figure 3.12 shows the relative response as a function of p_T^{avg} in bins of η_{det} for one half of the detector. The central η_{det} bins show a flat response up to high p_T^{avg} scales, while the forward bins show more fluctuation. As expected the very forward bins have low statistics.

The Monte Carlo prediction reproduces the overall features of the response well. However as in previous studies on 2010 data [112], the HERWIG++ response tends to be higher than the PYTHIA response. For $\lvert\eta_{det}\rvert < 2.1$ the data tends to agree with PYTHIA best; in the region $2.1 < \lvert\eta_{det}\rvert < 3.2$ the data tends to fall in between the two



(a) $22 < p_T^{\mathrm{avg}} < 30$ GeV (b) $30 < p_T^{\mathrm{avg}} < 40$ GeV

(c) $40 < p_T^{\mathrm{avg}} < 55$ GeV (d) $55 < p_T^{\mathrm{avg}} < 75$ GeV

Figure 3.11: The response, $1/c_i$, as a function of η_{det} for the four lowest p_T^{avg} bins. The dotted lines indicate response of 1 ± 0.05 in the main plots and 1 ± 0.1 in the ratio. The error bars are statistical only.

generators. For $|\eta_{det}| > 3.2$ the differences between data and Monte Carlo prediction are much greater, with around 5% typical and up to 11% observed in some bins. The data here is sometimes no longer between the generators, especially at low p_T^{avg}. This difference is seen at truth and parton level and so is thought to be due to a difference in the parton showering model used in the two generators.

Application of Results

These results were used as a cross-check to the dijet calibration performed by other members of the collaboration. The main result was produced with the *matrix method* [106] which did not require a central reference jet. Since the dijet cross section falls rapidly as the rapidity gap between the dijets increases this allows better statistical precision. The *central reference method* described above is extended by labelling the jets in terms of their η_{det} values, $\eta_{det}^{\rm left} < \eta_{det}^{\rm right}$, and then rewriting Equation (3.3) and Equation (3.5) as:

$$A = \frac{p_{\rm T}^{\rm left} - p_{\rm T}^{\rm right}}{p_T^{\rm avg}} \text{ , and } R = \frac{p_{\rm T}^{\rm left}}{p_{\rm T}^{\rm right}} = \frac{c^{\rm right}}{c^{\rm left}}, \tag{3.6}$$

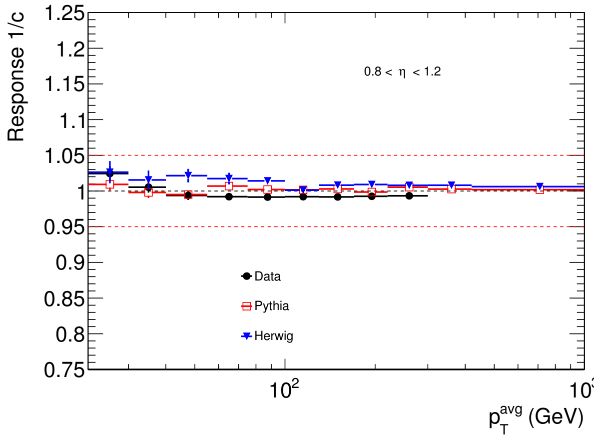
where R gives the ratio the responses and $c^{\rm right}$ and $c^{\rm left}$ are the intercalibration factors for the right and left jet. The response then forms a third rank tensor R_{ijk}, which is evaluated for each $\eta_{det}^{\rm left}$-bin i, $\eta_{det}^{\rm right}$-bin j, and $p_T^{\rm avg}$-bin k. It is assumed that a given η_{det}-bin is unaffected by where its balancing jet happens to be, and so the system of linear equations formed by R_{ij} can be solved for each $p_T^{\rm avg}$ bin. If this assumption is correct then the results should be equivalent to the *central reference method* within the statistical uncertainty.

Example response distributions from the *mathrix method* are shown in Figure 3.13. These results show the same trend as that observed in the *central reference method*. The largest variations are seen in the forward η_{det} region where variations at the 3% level are observed between the Monte Carlo and data from each method.

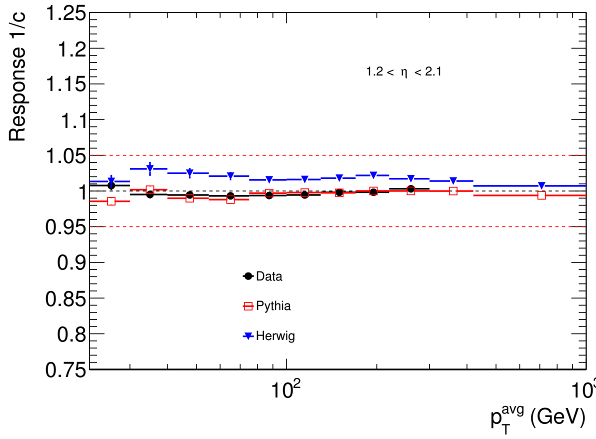
The response was corrected in the region $|\eta_{det}| < 2.4$ to agree with PYTHIA Monte Carlo, as PYTHIA is the reference Monte Carlo used to produce the main ATLAS JES. For $|\eta_{det}| > 2.4$ correcting to PYTHIA would introduce a large correction factor and cause the data response to differ significantly from the HERWIG++ response. To avoid this the correction factor was frozen at its value at $|\eta_{det}| = 2.8$.

The effects of systematic uncertainties were assessed, and the largest was found to be due to the Monte Carlo modelling. In the reference region this was taken to be zero, since the response is one by construction. For $0.8 < |\eta_{det}| < 2.4$ where data are corrected to PYTHIA the full difference between PYTHIA and HERWIG++ is taken as the uncertainty. For $|\eta_{det}| > 2.4$ the uncertainty is taken as the difference of the calibrated data to either PYTHIA or HERWIG++, whichever is larger. The overall systematic uncertainties for the dijet balance *in-situ* calibration are shown in Figure 3.14.

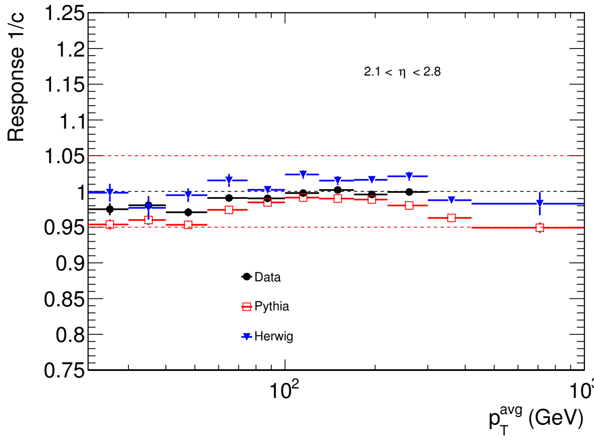
The uncertainty from dijet balance is the largest experimental systematic uncertainty in the main analysis in this thesis.



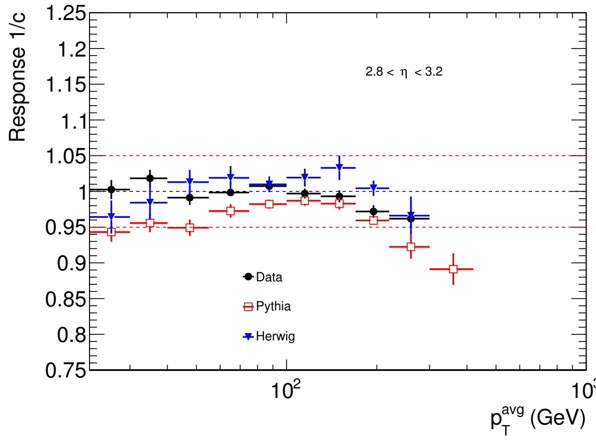
(a) $0.8 < \eta_{det} < 1.2$



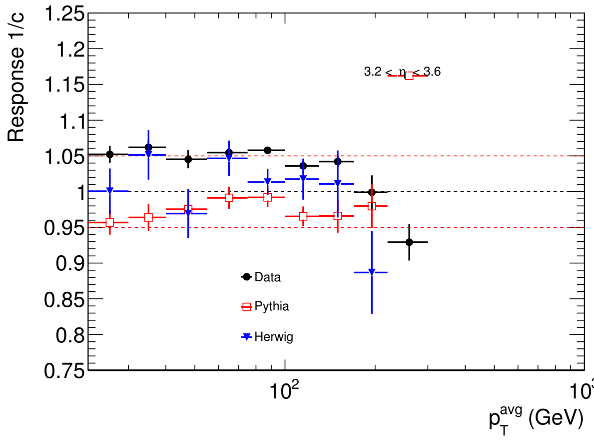
(b) $1.2 < \eta_{det} < 2.1$



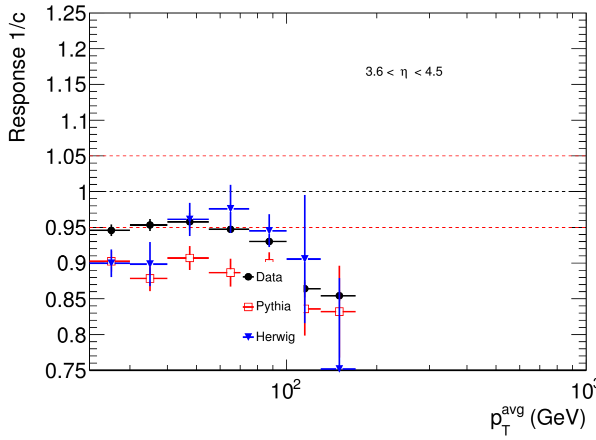
(c) $2.1 < \eta_{det} < 2.8$



(d) $2.8 < \eta_{det} < 3.2$

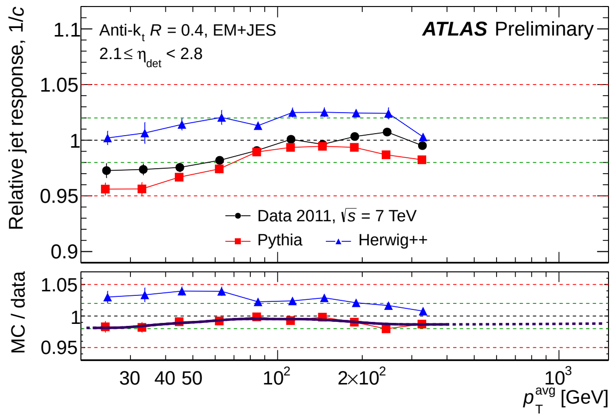


(e) $3.2 < \eta_{det} < 3.6$

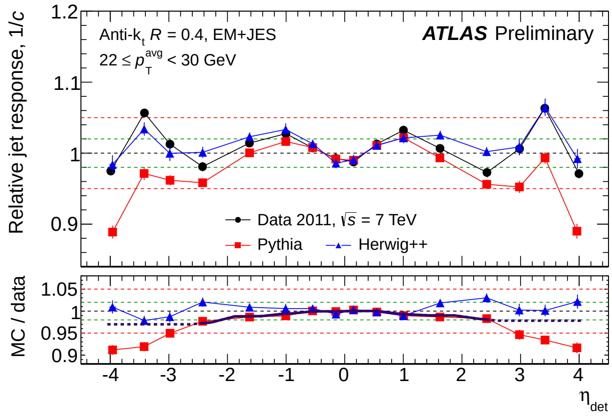


(f) $3.6 < \eta_{det} < 4.5$

Figure 3.12: Response correction as a function of p_T^{avg}. Each plot shows a different η_{det} region. The dotted lines indicate response of 1 ± 0.05 in the main plots.

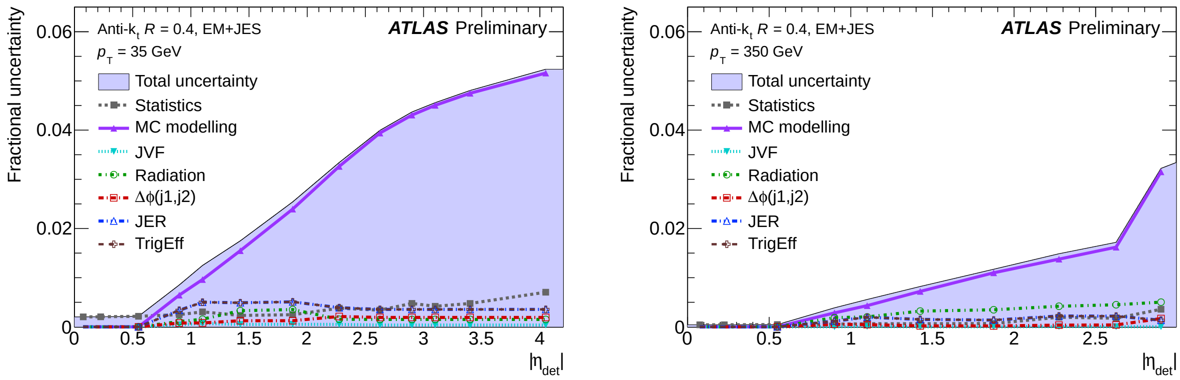


(a) Response correction as a function of p_T^{avg} for $2.1 < \eta_{det} < 2.8$



(b) Response correction as a function of η_{det} for $22 < p_T^{avg} < 30$ GeV

Figure 3.13: Example results taken from the *matrix method* used to derive the correction which is applied to data and Monte Carlo [106]. The left plot shows the response as a function of p_T^{avg} for a given η_{det} bin, and may be compared to my result from the central reference method in Figure 3.12(c). The right plot shows the response as a function of η_{det} and may be compared to my result in Figure 3.11(a).



(a) $p_{\mathrm{T}} = 35$ GeV

(b) $p_{\mathrm{T}} = 350$ GeV

Figure 3.14: The systematic uncertainty on the response calculated from dijet balance for jets with $p_{\mathrm{T}} = 35$ GeV and $p_{\mathrm{T}} = 350$ GeV. The Monte Carlo modelling is the dominant uncertainty [106].

3.4.2 Total jet energy scale and uncertainty

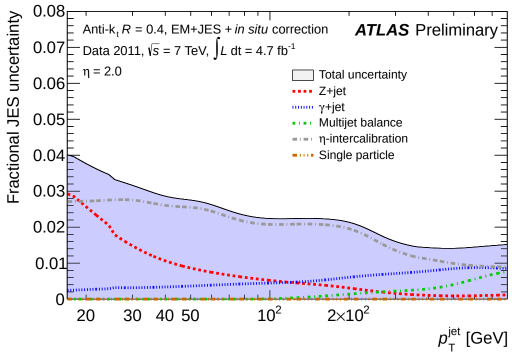
The dijet-balance response correction calculated using the *matrix method* is applied to data after the initial Monte Carlo based correction. The other methods of *in-situ* calibration, $Z/\gamma^* +$ jets balance, γ +jet balance and multijet balance are combined, taking into account their systematic uncertainties, and the resulting correction is applied to the data [113, 114]. This is performed with the jet calibration package `ApplyJetCalibration-00-02-07`.

The systematic uncertainty on the jet energy scale has over 60 sources coming from the *in-situ* calibration methods, Monte Carlo modelling and single-particle test beams. Combining all these sources into a single jet energy scale uncertainty would destroy important correlation information, while maintaining each individual source is impractical. Instead the covariance matrices of the sources are added and diagonalized, as described in Reference [114]. This results in five eigenvectors which are largely uncorrelated with a sixth component to cover the residual uncertainty.

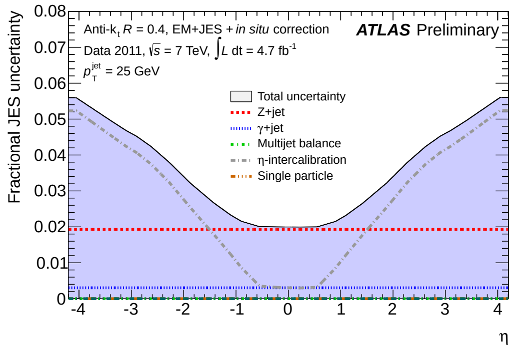
Additional sources due to the flavour composition of the calibration samples, the topology of jets with small separation in R ("close-by" uncertainty), and the pile up dependence are considered separately.

The overall results for the jet energy scale uncertainty are shown in Figure 3.15. These show that the jet energy scale is determined most precisely at high p_{T} and at central rapidities. At $|\eta_{det}| \gtrsim 1.5$ the η-intercalibration uncertainty dominates the uncertainty.

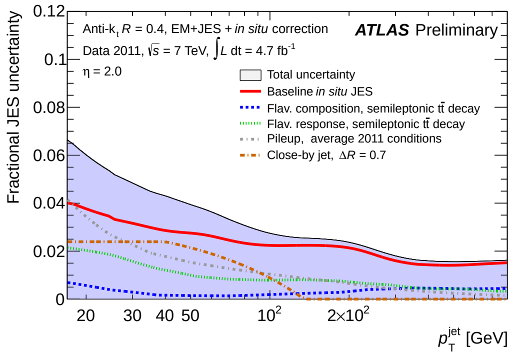
In the analysis presented in this thesis the jet energy scale uncertainty is evaluated using the `JetUncertainties-00-08-05` package [114, 115] using the `InsituJES2011_11NP.config` configuration file. This produces uncertainties as a function of the jet kinematics and pile up conditions using a reduction of the 63 nuisance parameters (60 from the nominal JES and 3 from close-by and flavour composition and response uncertainties) to the following:



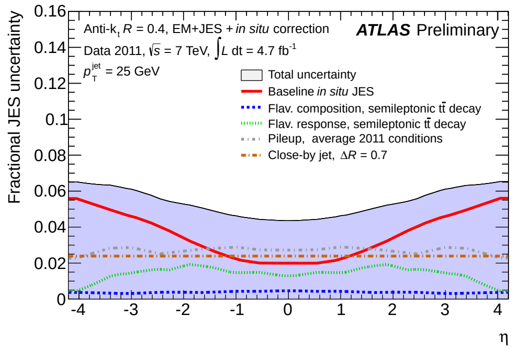
(a) Individual *in-situ* components for $\eta = 2.0$



(b) Individual *in-situ* components for $p_{\mathrm{T}} =$ 25 GeV



(c) Total JES for $\eta = 2.0$



(d) Total JES for $p_{\mathrm{T}} = 25$ GeV

Figure 3.15: Top: the contribution of the *in-situ* determined components to the total jet energy scale uncertainty [114]. Bottom: additional uncertainties on the JES [114].

- 1 component derived from the uncertainty on the η-intercalibration of the jets, referred to as 'JES η' in tables and plots;
- 2 components derived from the uncertainty on the terms comprising the pile up offset correction, referred to as 'JES (μ)' and 'JES (NPV)' in tables and plots;
- 1 component derived from the uncertainty of having other jets close to the jet of interest, referred to as 'JES (closeby)' in tables and plots;
- 2 components derived from the uncertainty on the flavour of the jets.These separate into the uncertainty on the flavour composition of the jets and the uncertainty on the response of different flavours of jet[10]. These are referred to as 'JES (Flavour response)' and 'JES (Flavour composition)' in tables and plots; and
- 1 component which corresponds to a combination of the 6 *in-situ* nuisance parameters and the high p_{T} and non-closure terms. 'Non-closure' refers to the residual difference found between calibrated reconstructed jets in Monte Carlo and the corresponding particle-level jet. If the calibration were perfect there would be no difference, and therefore perfect closure. These errors are referred to as 'JES (other)' in tables and plots.

This reduction results in a total of 7 parameters characterising the JES uncertainty which are fully propagated through the analysis and then combined in quadrature in the final result.

Jet energy resolution

The Jet energy resolution (JER) has been measured in data [116] with resolution of $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$ decreasing from $\sim 20\%$ for low $p_{\mathrm{T}} \sim 25$ GeV to $\sim 10\%$ at $p_{\mathrm{T}} > 100$ GeV. The Monte Carlo simulates a better resolution than that observed in the data and so it is corrected using the `JetResolution-02-00-02` package.

[10] This was calculated using the default parameter file supplied with the JES tool which is derived from $t\bar{t}$ events.

The uncertainty on this procedure is calculated using a different smearing which is based on a different measurement technique. For each event the jet smearing is performed three times and the average variation taken, to reduce the impact of large fluctuations. The resulting uncertainty is symmetrized by taking the average absolute difference from the nominal value and using this as the variation in both directions.

3.4.3 Analysis jet selection

In this thesis jets are selected from the anti-k_t R=0.4 algorithm calibrated with `EM+JES` scheme in common with all 2011 ATLAS Standard Model analyses. These are required to satisfy these criteria:

Kinematic cuts	
Transverse momentum	$p_{\mathrm{T}} > 30$ GeV
Rapidity	$\lvert y\rvert < 4.4$
Lepton overlap removal	$\Delta R(\ell, j) > 0.5$
Cleaning cuts	
Pileup	$\lvert JVF\rvert > 0.75$ if $\lvert\eta\rvert < 2.4$
LAr hole	veto event if jet in $-0.1 < \lvert\eta\rvert < 1.5$ and $-0.9 < \phi < -0.5$ and $p_{\mathrm{T}} > 20 \times \frac{1-\mathtt{BCH\_CORR\_JET}}{1-\mathtt{BCH\_CORR\_CELL}}$ GeV
Non-collision background	remove jet if `LooserBad`

Table 3.9: Requirements to enter the jet collection for the main results presented in this thesis. Explanation of new symbols is given in the main text.

- Lepton-jet overlap removal is applied since all energy deposits in the calorimeter that form topo-clusters above threshold will be built into jets. This includes electrons and muons so they must be explicitly removed from the jet collection.
- The ‘JVF’ is the ‘jet vertex fraction’ to reject pileup using tracks, as defined in Section 3.4.1.
- The `LooserBad` cleaning criteria are a set of cuts defined by the Jet group designed to remove jets caused by non-collision background or detector noise.

- $\frac{1-\texttt{BCH\_CORR\_JET}}{1-\texttt{BCH\_CORR\_CELL}}$ is the fractional difference between two methods of calculating energy in dead calorimeter cells. If this is greater than $\frac{1}{20}$ of the energy of the jet and the jet is in the LAr hole region then the event is rejected.

3.5 Missing Transverse Energy

Leptonic W boson decays produce neutrinos which leave the detector without depositing any energy. Conservation of momentum may be invoked to infer their presence. Due to the distribution of momenta of the partons we only know the initial momentum in the direction perpendicular to the beam, which should be zero. A vector sum of the transverse momenta of all visible final state particles is therefore made and the difference from zero designated as “missing transverse momentum”. At ATLAS this quantity is mainly measured using calorimeter information and so it is known as “missing transverse energy” or “$E_{\mathrm{T}}^{\mathrm{miss}}$ ”.

In this thesis the object based `MET_RefFinal` $E_{\mathrm{T}}^{\mathrm{miss}}$ definition is used [117]. This associates each topo-cluster with $|\eta| < 4.5$ in the calorimeters to a physics object and calibrates the energy of the cluster according to the calibration scheme of that object. Clusters are associated with objects following a heirarchical list; once a cluster has been associated to one object it is removed from the list of clusters and not considered again. This proceeds in the following order:

1. Electrons which satisfy `medium` ID and $p_{\mathrm{T}} > 10$ GeV
2. Photons which satisfy `tight` ID and $p_{\mathrm{T}} > 10$ GeV
3. Hadronic tau leptons
4. Hadronic jets calibrated with LCW. For $p_{\mathrm{T}} > 20$ GeV the JES is applied, otherwise the LCW clusters are used.
5. Muons, taking the p_{T} from their combined ID and MS measurement.

6. Remaining topo-clusters are calibrated to LCW and included. If a low p_{T} track is unambiguously associated with a cluster its p_{T} is used, since it is more precisely measured.

The reconstruction of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is handled by the package `MissingETUtility-01-01-03`.

3.5.1 Systematic uncertainties

Changes in the energy scales of the input objects to `MetRefFinal` are propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$ and included in the uncertainties associated to those objects. Further uncertainties due to the soft energy deposits included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ are included by using the `MissingETUtility-01-01-03` version of the $E_{\mathrm{T}}^{\mathrm{miss}}$ utility tool.

3.6 Truth Selection

Monte Carlo events are in addition subjected to a 'truth' selection. This uses the Monte Carlo event record without detector simulation effects to allow estimation of the acceptance. The signal selection is identical to that applied to reconstructed objects, however the physics objects of leptons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ are defined slightly differently. This standard ATLAS procedure (see e.g. [118][11]) is designed to ensure that truth cuts are not sensitive to the implementation details of different generators. In what follows 'stable' particles are final state particle (with Monte Carlo status code '1') and decay length $c\tau > 10$ mm.

Truth leptons We apply the selection to the so-called 'dressed' lepton. This is calculated from the momentum four-vector of the stable lepton after QED final state radiation (known as the 'bare' lepton) added to the momentum four-vectors of all

[11] These papers are examples of recent ATLAS results which use the procedure.

final state photons coming from the W boson decay in a cone of $\Delta R < 0.1$ around the lepton.

Truth $E_{\mathrm{T}}^{\mathrm{miss}}$ The truth $E_{\mathrm{T}}^{\mathrm{miss}}$ is taken as p_{T}^{ν}, the transverse momentum of the neutrino from the W boson decay.

Truth jets Truth jets are formed from all stable particles except those emerging from the W boson. This *includes* muons and non-interacting particles. This is a different definition from that used for the derivation of the JES described in Section 3.4 where muons and non-interacting particles are excluded.

3.7 Summary

This chapter has given a brief summary of the methods used at ATLAS to turn the raw detector signals into calibrated physics objects. A pictorial representation of a calibrated ATLAS event which passes the $W +$ EWK jets (detailed in the following chapter) is shown in Figure 3.16. Three projections of the detector information are shown: the $x - y$ projection (top left), which shows a slice perpendicular to the beam direction; the $z - \rho$ projection (bottom) which shows a slice taken parallel to the beam direction; and the 'lego plot' (centre right) which shows the calorimeter cells projected onto an $\eta - \phi$ plane. The ATLAS subdetectors can be identified as follows (cf. Chapter 2.2). The Inner Detector is shown in dark grey at the centre of the detector. Tracks identified with $p_{\mathrm{T}} > 1.0$ GeV and associated to the primary vertex are shown. The LAr calorimeter is shown in green, with energy deposits satistfying $E_{\mathrm{T}} > 200$ MeV indicated with yellow boxes proportional to E_{T}. The hadronic calorimeters are shown in red, with energy deposits satisfying $E_{\mathrm{T}} > 250$ MeV (Tile), 2.5 GeV (HEC) and 500 MeV (FCAL) shown in yellow. Beyond these are the muon chambers. In this event they record no muon candidates which can be combined with an Inner Detector track and so muon hits have not been displayed.

The electron candidate is highlighted with a green band, and can be identified by the isolated track leading to a deposit in the LAr calorimeter with no deposit in the hadronic calorimeter. The two jets reconstructed with anti-k_t $R = 0.4$ `TopoEM` algorithm are highlighted with white/grey cones. They make deposits in the LAr end caps and the Tile extended barrel.

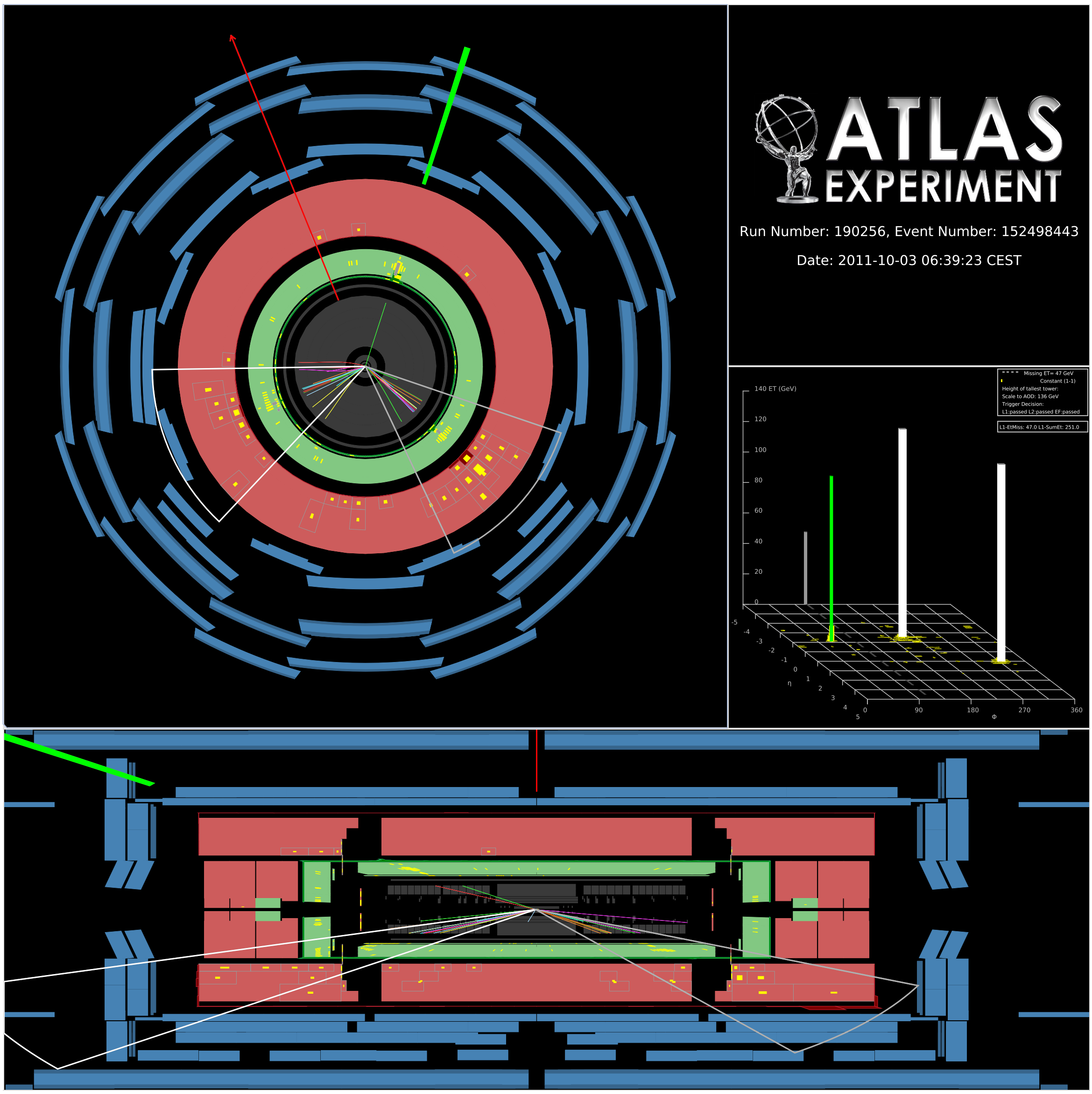


Figure 3.16: A candidate VBF W event from data taken in Period L (run 190256, event number 152498443). The electron candidate is highlighted in green, the $E_{\mathrm{T}}^{\mathrm{miss}}$ with a red arrow and the two jets indicated with white and grey cones. Further discussion is found in the main text. This event display was created using the Atlantis event display program [119].

CHAPTER 4

Event Selection

In this chapter I describe the selection used to identify events of interest. There are 4 regions of interest selected:

1. "Inclusive" W boson selection — this selection is identical to the ATLAS W plus jets analysis [38, 39].

2. "Signal" selection — a selection optimized for the observation of $W+$ EWK jets.

3. Electroweak W boson "validation" region — designed to be orthogonal to the "signal" selection to test the background modelling.

4. QCD control region — used to estimate the QCD background in the signal region. This will be discussed in detail in Chapter 5.

4.1 Data samples

The analysis is based on a sample of $\sqrt{s} = 7$ TeV proton-proton collisions collected in 2011. We use data collected during periods of stable beam operation with 50 ns bunch spacing (periods D to M). The data is only included in the analysis if it has passed all data quality requirements, including all detector components operating with correct settings and without large noise bursts. This is done by use of centrally produced Good Run List (GRL)[1] which identify the data quality lumiblock by lumiblock. All selections and channels use data corresponding to the same integrated luminosity of 4.64 fb^{-1}.

4.2 Monte Carlo samples

The signal and all background except the QCD multi-jet are simulated using Monte Carlo simulation. The QCD multi-jet background is estimated using a data-driven method which is described in Chapter 5.

4.2.1 Signal simulation

The $W+$ EWK jets process has been implemented in VBF@NLO [49], a dedicated VBF Monte Carlo calculator, capable of calculating differential cross sections to next-to-leading order in QCD. However this program is not interfaced to a parton shower and so cannot be used for direct comparisons to data[2]. Instead we used the SHERPA Monte Carlo generator [24]. This combines an automatic matrix-element calculator with the CKKW matching scheme [25], which allows multiple hard emissions to be matched to parton showers.

The VBF signal is generated with SHERPA version 1.3.1 requiring at least four

[1] `data11_7TeV.periodAllYear_DetStatus-v36-pro10_CoolRunQuery-00-04-08_WZjets_allchannels_DtoM.xml`

[2] $W+$ EWK jets has recently been included in the POWHEG box and so it is hoped that future measurements may make use of the NLO accuracy calculation [120].

electroweak vertices in the matrix-element, a lepton and a neutrino with invariant mass greater than 10 GeV and two or three partons in the matrix-element calculation. These requirements also allow for single top production, see Figure 1.7(c), where one of the intermediate quarks could be a t-quark. Single top production is simulated separately, so this contribution is removed from the signal sample by explicitly excluding the top quark.

This signal sample contains four classes of diagram. In addition to the three signal diagrams shown in Figure 1.7 — multiperipheral, bremsstrahlung and VBF— the sample contains diboson diagrams such as those shown in Figure 1.6. The diboson diagrams are a background to our signal process, however at the time of Monte Carlo production it was not technically possible to remove it from the SHERPA sample[3]. It is effectively removed with a detector level cut of $M_{jj} > 500$ GeV, as shown in Section 4.4.

A cross-check of the cross section calculated by SHERPA version 1.3.1 was performed using SHERPA version 1.4.1, where diboson diagrams can be removed, and using VBF@NLO version 2.5.3 [49]. Since VBF@NLO does not include the diboson diagrams a dijet mass cut must be used to remove the diboson contributions from the SHERPA cross section. Full details of the VBF@NLO calculation are given in Reference [60]. The results of the cross section comparison for $M_{jj} > 200$ GeV are shown in Table 4.1. The cut value of 200 GeV is used here as at generator level this is sufficient to remove the diboson contributions. A higher value is used in the main analysis since the limited jet resolution will lead to higher values of dijet mass compared to the generator level.

The difference caused by removing the diboson samples is negligible compared to the difference between SHERPA versions, which gives a change of 3%. VBF@NLO provides both a leading and next-to-leading order cross section and suggests a 'k-

[3] In SHERPA versions 1.4.1 and greater a new option `min_N_tchannels` is available which allows the generation of $W +$ EWK jets production without the diboson contribution. This method was included by the SHERPA authors in response to a request, and will be used for the 8 TeV result, beyond the time-scale of this thesis.

factor'[4] for $W+$ EWK jets production from this calcualation is 0.97. The k-factor for this process is not applied anywhere in this thesis. Sherpa 1.3.1 and vbf@nlo agree within 5%, while Sherpa 1.4.1 and vbf@nlo agree within 9%. Given the different choices of PDF, α_s, and different factorization and renormalization schemes between the versions of the generators this level of disagreement is to be expected. In addition it should be noted that vbf@nlo is a parton level Monte Carlo calculator while Sherpa includes parton shower effects. This comparison serves to show that two independant calculations of the VBF process give consisitent results, within the uncertainties described above.

	Cross section in fb		
Generator	W^+	W^-	$W^\pm$
Sherpa 1.3.1	1314 ± 13	791.3 ± 7.8	2069 ± 21
Sherpa 1.4.1	1353 ± 13	804 ± 8	2138 ± 21
Sherpa 1.4.1 no diboson	1332 ± 13	798 ± 7.9	2140 ± 21
vbf@nlo (LO)	1265 ± 2.19	742.2 ± 4.0	2008.0 ± 2.5
vbf@nlo (NLO)	1220 ± 6.5	742.9 ± 1.2	1962 ± 7.7

Table 4.1: The cross section for the $W+$ EWK jets process in the electron channel measured in fb. Calculated in Sherpa and vbf@nlo with a generator level dijet mass cut of $M_{jj} > 200$ GeV. The results for Sherpa were calculated separately for W^+, W^- and $W^\pm$; for vbf@nlo the W^+ and W^- were simply summed to give the $W^\pm$ results. The errors shown are statistical only.

The signal samples used in data analysis were produced using the Atlas production system using a central installation of Sherpa and a default set of parameters to ensure compatibility between the various generators used by Atlas. Example $W+$ EWK jets and $W+$ QCD jets run cards are given in Appendix B.

4.2.2 Background simulation

***W* boson production in association with QCD jets** This is the largest background to the $W+$ EWK jets signature with a cross section about 100 times larger

[4] 'k-factor' normalizes the generator cross section to the best available calculation. Here it is the ratio $\sigma_{NNLO}/\sigma_{(N)LO}$ which normalizes a (next-to-) leading-order sample to the next-to-next-to-leading-order cross section.

than that of $W+$ EWK jets production. This is an 'irreducible background', meaning that the final state objects are indistinguishable from the signal.

$W+$ QCD jets with the W decaying to all lepton flavours is simulated with ALPGEN version 2.13 [23] interfaced with HERWIG version 6.520 [21] for parton shower and hadronization and to JIMMY version 4.31 [21] to model underlying event and MPI contributions, using the AUET2-CTEQ6L1 tune [121]. CTEQ6L1 [122] parton density functions are employed. The program PHOTOS [123] is used to simulate final state QED radiation.

ALPGEN is a leading order 'multileg' generator. Matrix-elements with between zero and five partons are created with a separate sample for each parton multiplicity and the 'MLM' matching procedure [22] is used to ensure contributions from the matrix-element and parton shower are not double counted. For each parton multiplicity except the highest one, each parton must be matched to a particle-level jet within $\Delta R < 0.7$, where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, and with particle-jet $p_{\mathrm{T}} > 20$ GeV and the total number of partons must equal the total number of jets. For the highest parton multiplicity the generation is inclusive, allowing a higher number of jets than partons in the final state.

The standard ALPGEN $W+$jets samples consider only massless u, d, s and c quarks in the matrix-element. Massive b and c quarks can influence the kinematics of $W+$jets production and so additional $W+$ *heavy-flavour* samples were prepared. These include $b\bar{b}$, $c\bar{c}$ and c quarks in the matrix-element. Since heavy-flavour quarks may also be produced in the parton shower there is the possibility for the same final state events to be counted in both the light-flavour and heavy-flavour samples. An overlap removal process was defined by the ATLAS Top group [124] that exploits the strengths of the matrix element for hard well separated emissions and the parton shower for collinear splittings. Heavy-flavour quark pairs from the matrix element are required to be matched to different jets at generator level, while heavy-flavour quark pairs from the parton shower are required to be inside one reconstructed jet. The jet-parton matching is performed with a simple geometric cone, where partons are matched to jets if $\Delta R(\text{jet}, \text{parton}) < 0.7$.

Additional $W+$ jets samples were simulated with SHERPA version 1.4.1 [24] for decays of a W boson to $\mu\nu$ and $e\nu$ to provide a cross-check of the ALPGEN samples. The CT10 PDF set [125] and SHERPA's internal parton shower and underlying event algorithms were employed. This sample uses the MEnloPS generation procedure [126], which merges a next-to-leading order calculation of inclusive W production using the PowHeg method [26] with a leading order calculation of $W + 1 - 5$ parton and matches these to the SHERPA parton shower algorithm using the CKKW procedure [25]. This special technique is required since $W + 1$ jet events may be generated from W production plus one NLO emission, $W + 1$ jet LO matrix element, or from parton shower effects in the underlying event.

This results in a total cross section that is correct to Next-to-leading order (NLO) while including higher jet multiplicities and is therefore expected to describe the total rate accurately as well as the kinematics. The different parton multiplicities are merged internally by SHERPA. High multiplicity events are retained in the sample at higher rate than their cross section allows by the use of *enhancement factors* This ensures that sufficient numbers of high multiplicity events remain in the sample. $W + 1$ jet events are given an enhancement factor of 2, while $W + 2 - 5$ jet events are given an enhancement factor of 5. This is corrected for by the application of automatically produced Monte Carlo event weights in the analysis.

These samples are normalized globally to Next-to-next-to-leading order (NNLO) inclusive predictions as determined using the FEWZ [127] program by the ATLAS cross section task force [128, 129]. This leads to k-factors of 1.18 for the ALPGEN samples and 1.008 for SHERPA[5]. The MSTW2008 NNLO PDFs [15] are used to compute the central values and compared with other NNLO PDFs to determine the systematic uncertainty which is taken to be 5%. Further uncertainties on these cross sections

[5] This is much smaller as the SHERPA cross section is already correct to next-to-leading-order.

arise from factorization and renormalization scale dependence and the choice of α_s lead to $< 1\%$ variations. The total uncertainty is about 5%.

Top-quark pair production This results in two top quarks which decay to b-quarks and W bosons. This can pass the signal selection if one W boson decays hadronically and the other leptonically, or both leptonically with one lepton misidentified or outside the acceptance. An upper cut on number of jets will substantially reduce this contribution.

Top-quark pair production samples are generated using MC@NLO version 3.41 [27] interfaced with HERWIG for the parton shower and underlying event and using the CTEQ10 [125] PDF set. This sample was generated with a filter to require at least one leptonic top decay. The cross section is calculated at approximate NNLO in QCD with HATHOR 1.2 [130] using MSTW2008 PDF sets which incorporate PDF+α_s uncertainties [15, 131] cross checked with NLO+next-to-next-to-leading-log (NNLL) calculation of Reference [132] implemented in `Top++` [133]. This leads to a k-factor of 1.14 and a cross section uncertainty from PDF and scale variations of 10%.

Single-top production Top-quarks may be produced in isolation through electroweak processes, as observed by Tevatron [134] and LHC experiments [135] in three channels: t-channel, s-channel and Wt. This can lead to a $W +$jets final state as the decay of the t-quark produces a W and jets.

Single-top events are generated with ACERMC [136] version 3.8, interfaced with PYTHIA for parton shower, ISR, FSR and underlying event and using a modified leading-order MRST PDF [137]. These samples are normalized to cross sections calculated at next-to-next-to-leading-order [138] and theoretical uncertainties in this cross section are calculated from 90%CL variations in PDF and variations of factors of 1/2 and 2 of the renormalization and factorization scales.

Diboson processes The WW and WZ processes each contain a W which may decay leptonically and another boson that may decay to quarks leading to jets which can pass our signal selection. ZZ may pass the signal selection if one of the Z bosons decays to leptons and one of the leptons falls outside the acceptance or is otherwise not reconstructed.

The WW and WZ processes are included in the signal Monte Carlo sample, but as a cross check they are produced, along with ZZ, with Herwig using the AUET2-LO* tune [121]. Following the standard Atlas procedure [128] they are normalized to the NLO cross section from MCFM [139] using the MSTW2008NLO PDF set [15] with a cross section uncertainty of 5% for WW and ZZ production and 7% for WZ production. This leads to k-factors of 1.52, 1.58 and 1.41 for WW, WZ and ZZ respectively.

***Z* boson production in association with jets** $Z/\gamma \rightarrow \ell^+\ell^- +$ jets may pass the signal selection if one of the leptons is outside the acceptance or is misidentified. All leptonic decay modes are simulated with Alpgen in the same way as $W+$ jets as described above.

QCD Hadronic multi-jet events originating from QCD processes may pass the signal selection provided that one jet passes the reconstruction requirements of a lepton. The cross section and its uncertainty are large so a data-driven method is used to estimate this background. This is described in Chapter 5.

Monte Carlo summary A summary table of Monte Carlo samples used in this thesis are given in Table 4.2. All samples except for the signal sample are standard Atlas Monte Carlo samples produced by physics groups.

In plots, background estimations are grouped together using the key shown in Table 4.3. In each plot the backgrounds are shown 'stacked'. A ratio plot is shown for each distribution showing Data/Standard Model (SM). In addition, selected plots are shown using SHERPA for the $W+$ QCD jets. The errors are indicated in the main plot by a black hatched area. In the ratio plot the errors are shown for experimental systematics (as detailed in Chapter 3) by a yellow band, and for experimental and statistical errors added in quadrature in a green band.

4.3 Inclusive W + jets Selection

To select a W boson we follow a standard W selection procedure, in common with the 2011 $W+$ jets analysis [39] and the W/Z+jets ratio analysis [42][6].

Events which pass the lowest threshold unprescaled electron or muon triggers, as detailed in Sections 3.2.1 and 3.3.1, are considered. Events are required to have at least one primary vertex with two or more tracks and to pass the LAr-error and Tile-error flags.

In addition we require:

- exactly one 'tight' lepton with $p_{\mathrm{T}}^{\ell} > 25$ GeV, satisfying the selection criteria given in Table 3.2 (electron) or Table 3.4 (muons),
- No second lepton:
  - Electron with $p_{\mathrm{T}} > 20$ GeV and satisfying `medium++` identification,
  - Muon with $p_{\mathrm{T}} > 20$ GeV and not requiring the impact parameter and isolation requirements (given in Table 3.4),
- $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV and,

[6] These are similar in approach to the 2010 papers [38, 41] but use the larger dataset to produce a wider selection of distributions including higher jet multiplicity. Changes in selection were required due to higher trigger thresholds and increased pile up.

Process	Dataset ID	Generator	Cross section (pb)	Effective luminosity (fb^{-1})	Note
			V + EWK jets		
VBF $W \to e\nu$	126569	Sherpa, v1.3	8.40	23.80	EWK order 4
VBF $W \to \mu\nu$	126570	Sherpa, v1.3	8.41	23.52	EWK order 4
VBF $W \to \tau\nu$	126571	Sherpa, v1.3	8.42	23.75	EWK order 4
VBF $Z \to ee$	125950	Sherpa, v1.3	0.48	447.41	EWK order 4
VBF $Z \to \mu\mu$	125951	Sherpa, v1.3	0.48	406.41	EWK order 4
			V + QCD jets		
$W \to e\nu$	107680-107685	Alpgen +Herwig	10.46×10^3	1.00-93.77	$N_\mathrm{p} = 0 - 5$ EWK order 2
$W \to \mu\nu$	107690-107695	Alpgen +Herwig	10.46×10^3	1.00-94.33	$N_\mathrm{p} = 0 - 5$ EWK order 2
$W \to \tau\nu$	107700-107705	Alpgen +Herwig	10.46×10^3	0.49-9.32	$N_\mathrm{p} = 0 - 5$ EWK order 2
$Z/\gamma^* \to ee$	107650-107655	Alpgen +Herwig	1.07×10^3	9.88-66.83	$m_{\ell\ell} > 40$ GeV $N_\mathrm{p} = 0 - 5$ EWK order 2
$Z/\gamma^* \to \mu\mu$	107660-107665	Alpgen +Herwig	1.07×10^3	9.88-65.23	$m_{\ell\ell} > 40$ GeV $N_\mathrm{p} = 0 - 5$ EWK order 2
$Z/\gamma^* \to \tau\tau$	107670-107675	Alpgen +Herwig	1.07×10^3	15.85-59.83	$m_{\ell\ell} > 40$ GeV $N_\mathrm{p} = 0 - 5$ EWK order 2
			V + heavy flavour jets		
$W + bb$	107280-107283	Alpgen +Herwig	47.36	5.73-10.03	$N_\mathrm{p} = 0 - 3$
$W + cc$	117284-117287	Alpgen +Herwig	127.81	8.30-10.04	$N_\mathrm{p} = 0 - 3$
$W + c$	117293-117297	Alpgen +Herwig	644.15	10.09-10.19	$N_\mathrm{p} = 0 - 4$
			Alternative V + QCD jets samples for cross-checks		
$Z/\gamma^* \to ee$	147770	Sherpa, v1.4	1.07×10^3	9.4-4.6	$m_{\ell\ell} > 40$ GeV EWK order 2
$Z/\gamma^* \to \mu\mu$	147771	Sherpa, v1.4	1.07×10^3	9.4-4.6	$m_{\ell\ell} > 40$ GeV EWK order 2
$W \to e\nu$	147774	Sherpa, v1.4	10.46×10^3	0.9-4.8	EWK order 2
$W \to \mu\nu$	147775	Sherpa, v1.4	10.46×10^3	0.9-4.8	EWK order 2
			Diboson		
ZZ	105986	Herwig	6.49	54.68	$\epsilon_F = 0.211$
WZ	105987	Herwig	5.42	80.02	
WW	105985	Herwig	45.0	87.06	$\epsilon_F = 0.388$
			$t\bar{t}$ and single-top		
$t\bar{t}$	105200	mc@nlo + Herwig	166.8	102.97	$\epsilon_F = 0.556$
single top, t-channel	117360	AcerMC + Pythia	6.97	382.64	
single top, s-channel	117363	AcerMC + Pythia	0.50	1696	
W top	105500	AcerMC + Pythia	15.7	19.0	

Table 4.2: Samples of Monte Carlo simulated events used in the result. The cross sections quoted are the ones used to normalise estimates of the expected number of events. Where two generator names are given, the first is that used to generate the hard process, while the second is used for the parton shower. N_p is the number of additional partons in the sample. ϵ_F is the generator filter efficiency. "EWK order" refers to the number of electroweak vertices allowed in the matrix element. Sources for the cross sections are detailed in Section 4.2.

Sample Shorthand	Contents	
$W \to \ell\nu$	$W \to e\nu$	EWK order 2
	$W \to \mu\nu$	
	$W \to \tau\nu$	
	W + heavy flavour	
top	$t\bar{t}$	
	single top	
EWK	$W \to \tau\nu$	EWK order 4
	$Z \to \ell\ell$	
	ZZ	
QCD	Data driven QCD estimate (see following chapter)	
VBF $W \to \ell\nu$	$W \to e$ or μ	EWK order 4
	$WW + WZ$	

Table 4.3: Shorthand key for plot labels. Note that the EWK order 4 samples include WW and WZ contributions as explained in Section 4.2.1.

- $m_T > 40$ GeV where

$$m_T = \sqrt{2p_{\mathrm{T}}^{\ell} \cdot E_{\mathrm{T}}^{\mathrm{miss}}\Big(1 - \cos\left[\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})\right]\Big)}.$$

This selection is henceforth referred to as the 'inclusive' selection. The inclusive W region shows very good agreement between data and background estimates as demonstrated in Figure 4.1. The jet multiplicity plot allows a comparison to be made between the overall normalization of the data and the Monte Carlo simulation. At high jet multiplicity ($N_{jets} \geq 6$) the Monte Carlo no longer models the data well. This is understood to be due to the top modelling, since the NLO top sample will only contain up to one additional parton emission. This could be improved using the ALPGEN $t\bar{t}$ + jets samples which simulate addition parton emission in the matrix element. However these high jet bins are not a concern for the main analysis of this thesis and so this method was not pursued.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution shows good agreement between the Monte Carlo and data, with a small excess in the data compared to the Monte Carlo observed in both electron and muon channels around 80 GeV. This deviation is more than covered by the total systematic uncertainty.

The lepton p_{T} distribution displays a difference between the electron and muon channels, with the electron channel showing a slight excess from $\sim$ 50 GeV in the data compared to the Monte Carlo, while the muon channel shows a deficit above $\sim$ 150 GeV. The source of this discrepancy is unknown; the most likely source is the QCD multi-jet background since that is calculated separately for the lepton channels (see following chapter). Other possible sources are the lepton modelling or efficiency measurements. As the lepton p_{T} distribution is steeply falling it does not have a large effect on the overall normalization and so was not further considered in this analysis.

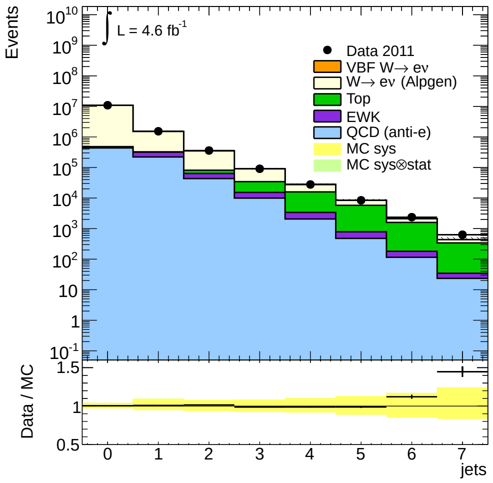
4.4 VBF selection

In the inclusive $W+2$ jets region there is very low VBF purity ($\mathcal{O}(1)\%$) so further cuts are made to enhance the signal to background ratio. These cuts were chosen in order to maximize signal purity and expected signal significance: $N_s/(\sqrt{N_b+\sigma^2})$, where N_s is the number of signal events, N_b is the number of background events and σ is the uncertainty. The 'tag jets' are defined as being the leading two jets in p_{T}. The cuts are:

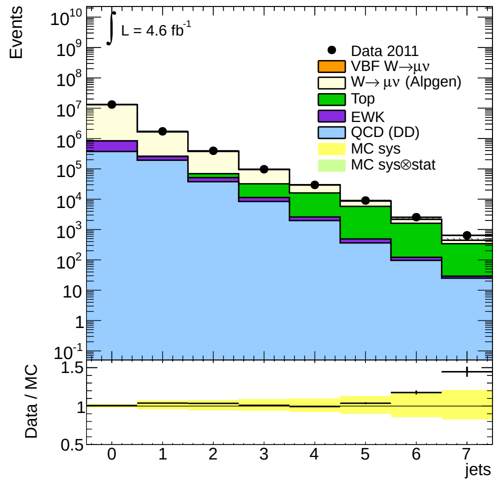
1. $N_{jets} = 2$ - no third jet with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 4.4$;
2. $M_{jj} > 500$ GeV;
3. $p_T^{j_1} > 80$ GeV, $p_T^{j_2} > 60$ GeV referred to as "high p_{T}" and;
4. $\eta_{ja} < \eta_l < \eta_{jb}$ - referred to as "outside lepton veto".

Each cut is now justified in turn:

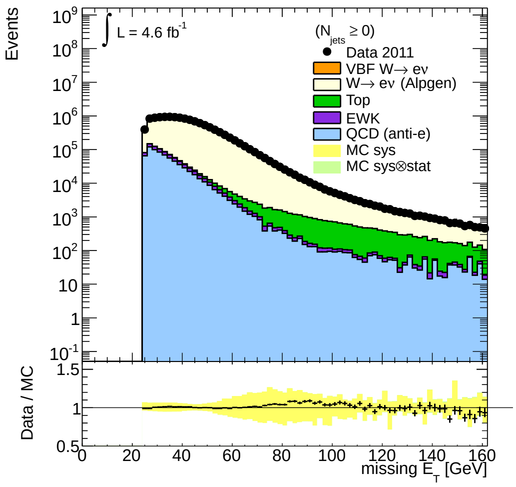
$\boldsymbol{N_{jets} = 2}$ is used to maximize signal purity. Higher jet multiplicity bins contain large contributions from top processes.



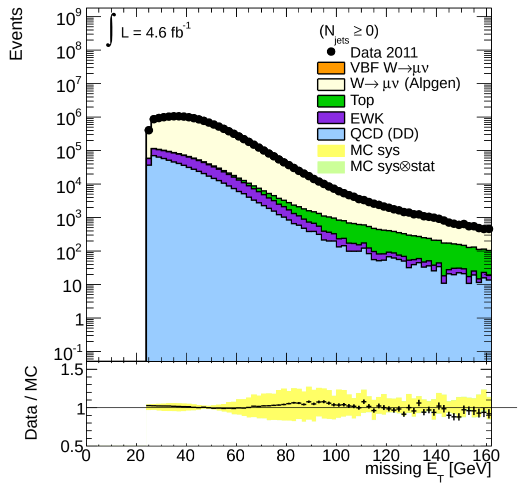
(a) Number of Jets (electron channel)



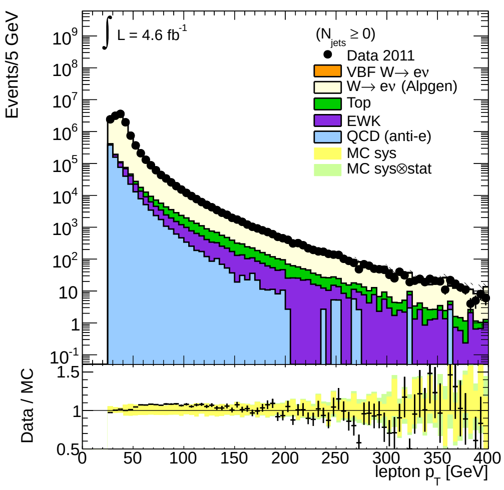
(b) Number of Jets (muon channel)



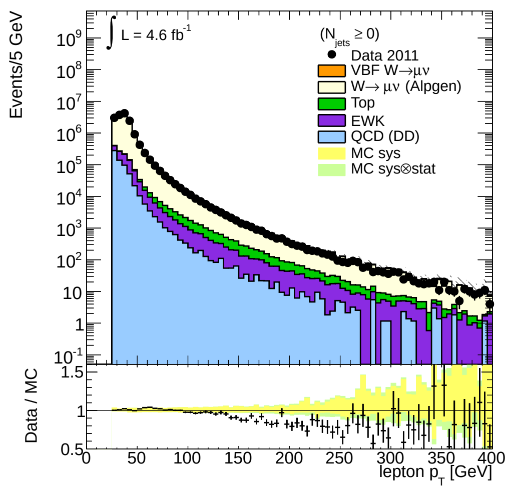
(c) $E_{\mathrm{T}}^{\mathrm{miss}}$ (electron channel)



(d) $E_{\mathrm{T}}^{\mathrm{miss}}$ (muon channel)



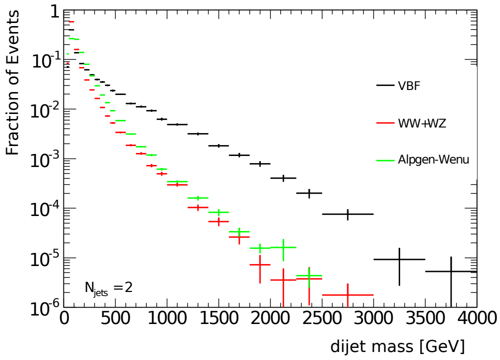
(e) Leading electron p_{T}



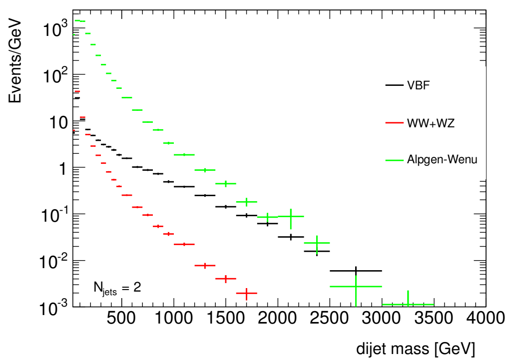
(f) Leading muon p_{T}

Figure 4.1: Inclusive distributions of electron (left) and muon (right) channels of number of jets (top), $E_{\mathrm{T}}^{\mathrm{miss}}$ (middle), and leading jet p_{T} (bottom).

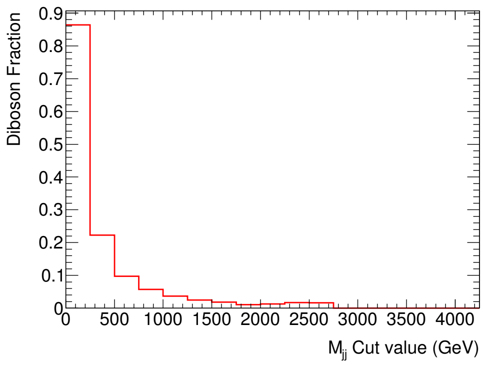
M_{jj} > 500 GeV The invariant mass of the tag jets (M_{jj}) is expected to be the primary discriminating variable for VBF. For dijet masses below $\sim$ 100 GeV the selection is dominated by diboson processes. After the cut $M_{jj} > 500$ GeV the signal is expected to have a contamination of less than 10% diboson, as shown in Figure 4.2. Figure 4.3 shows the dijet mass distributions in the electron and muon channels before the cut.



(a) Normalized to 1



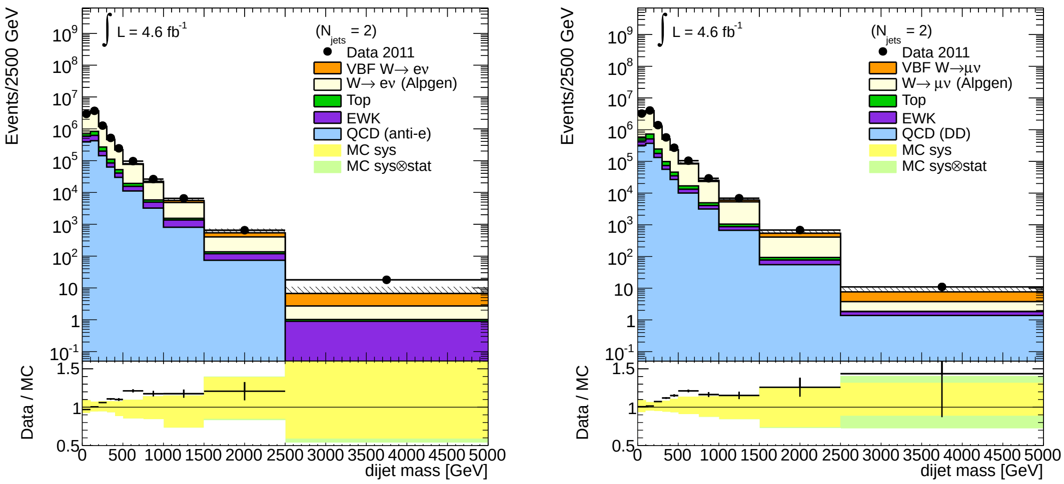
(b) Normalized to cross-section



(c) Cumulative fraction of WW+WZ processes in the VBF sample as a function of M_{jj} cut

Figure 4.2: Distributions after selection in an inclusive $W + 2$ jets phase space.

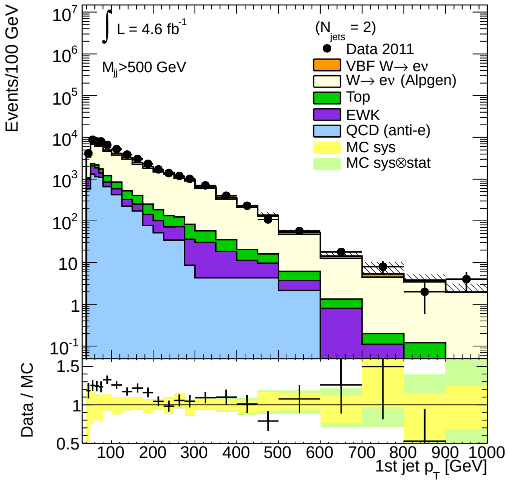
$p_T^{j_1} > 80$ GeV, $p_T^{j_2} > 60$ GeV $W+$ EWK jets production gives jets with higher p_{T} than the $W+$ QCD jets background, so this cut increases signal purity. In addition the systematic uncertainty on the jet energy scale increases rapidly at low p_{T}. Significant disagreements between data and Monte Carlo are observed without this cut due to low p_{T} jets at high absolute rapidities. These are probably due to mismodelling of pile up, which cannot be removed with the JVF cut beyond the tracker acceptance.



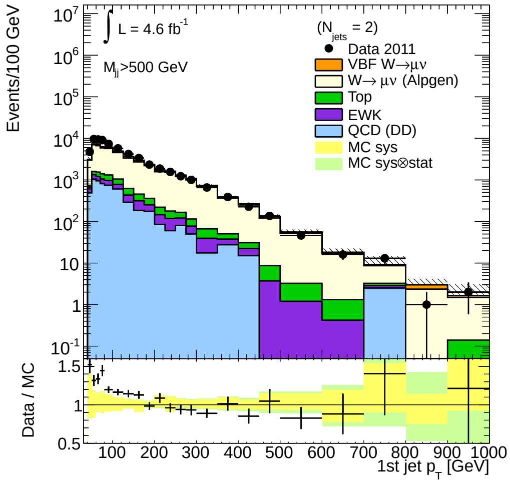
(a) M_{jj} (electron channel) (b) M_{jj} (muon channel)

Figure 4.3: M_{jj} for electron and muon channels, requiring $N_{jets} = 2$

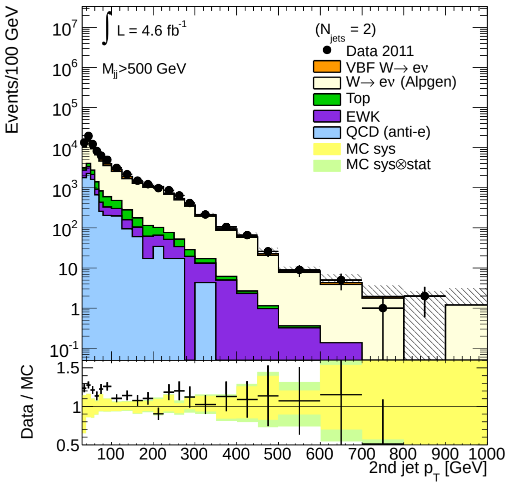
The jet p_{T} spectra for the leading and sub-leading jets before the 'high p_{T}' cuts are applied are shown in Figure 4.4, and the expected signal significance as a function of p_{T} cut is shown in Figure 4.5.



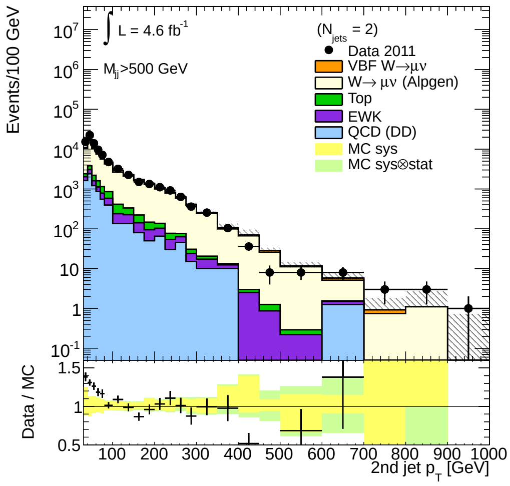
(a) Leading jet p_{T} (electron channel)



(b) Leading jet p_{T} (muon channel)

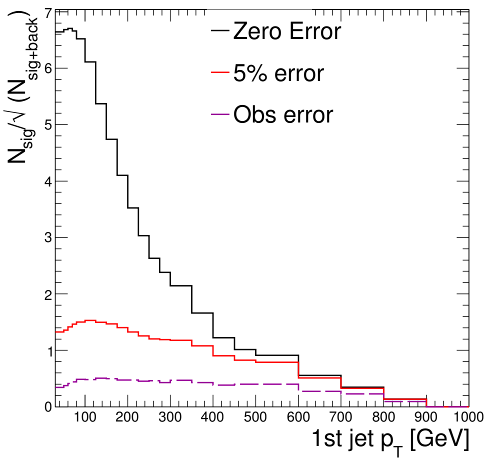


(c) Sub-leading jet p_{T} (electron channel)

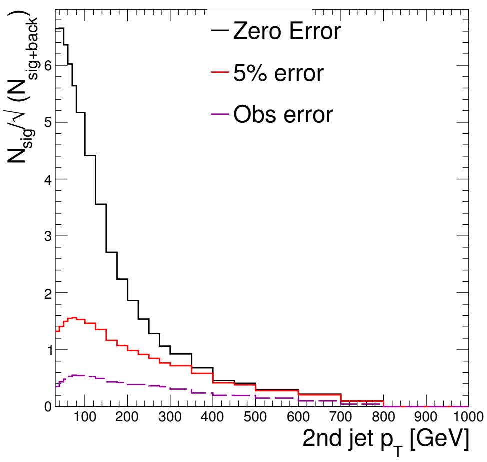


(d) Sub-leading jet p_{T} (muon channel)

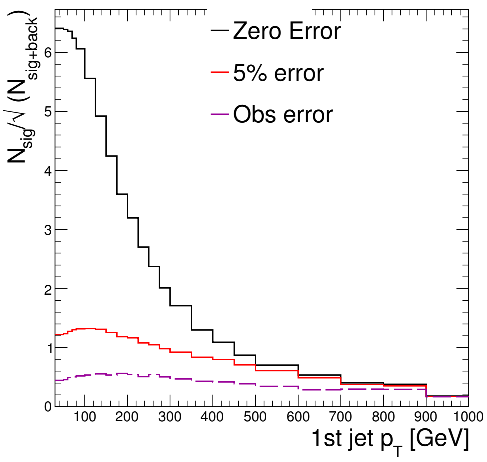
Figure 4.4: Leading and sub-leading jet p_{T} distribution for electron and muon channels after VBF cuts 1 and 2.



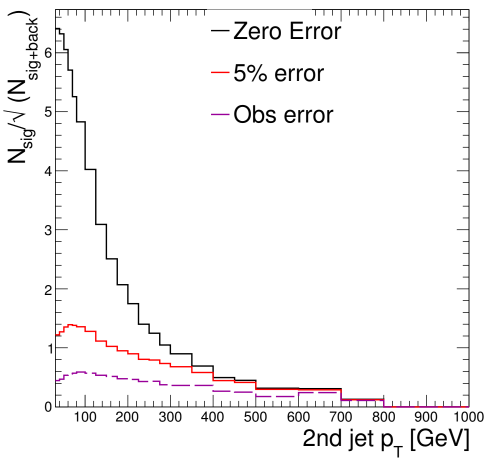
(a) Leading jet p_{T} (electron channel)



(b) Sub-leading jet p_{T} (electron channel)



(c) Leading jet p_{T} (muon channel)



(d) Sub-leading jet p_{T} (muon channel)

Figure 4.5: The expected significance as a function of minimum p_{T} cut on the leading and sub-leading jet, where significance is calculated as $s = \frac{N_{\mathrm{signal}}}{\sqrt{N_{\mathrm{signal+background}}+\sigma^2}}$ where σ is the expected experimental systematic error.

Outside lepton veto The lepton is required to be between the η of the tag jets. The W boson from $W+$ EWK jets production is expected to be found between the two tag jets in η. We cannot determine the η of the neutrino and so the η of the lepton is used as a proxy for the η of the W boson . This cut significantly reduces $W+$ QCD jets and QCD backgrounds. In addition it removes the mismodelled electrons found in the $|\eta| > 1.53$ region as shown in Figure 4.6.

The application of these cuts produces a signal region with $11-12\%$ $W+$ EWK jets. The effect of these cuts for each source of background and signal is shown in Table 4.4

Cut Name	W $N_{jets} = 2$	M_{jj}	M_{jj}, high p_{T}	M_{jj}, high p_{T}, outside lepton veto
$W \to \mu\nu$	235,293.6	5,510.1	3,082.2	1,713.8
$W \to \tau\nu$	5,904.8	150.5	81.3	42.5
$Wbb \to l\nu$	7,641.2	216.0	138.3	79.8
$Wcc \to l\nu$	18,146.3	520.7	307.1	172.1
$Wc \to l\nu$	38,675.1	660.7	386.4	185.1
QCD (data driven)	37,752.2	895.8	267.2	172.5
ZZ	59.8	0.6	0.3	0.2
$Z \to \mu\mu$	12,276.4	307.1	162.7	85.8
$Z \to \tau\tau$	1,227.2	21.4	9.4	6.1
$t\bar{t}$	11,587.2	169.3	108.3	73.8
single top	6,553.3	264.6	139.6	104.6
VBF $W \to \tau\nu$	91.1	13.4	9.8	8.1
VBF Signal Z	67.4	5.9	4.0	3.3
VBF Signal W	4,644.9	619.7	399.9	330.5
Data	393,275.0	10,844.0	5,303.0	3,024.0
SM prediction	379,920.5	9,355.7	5,096.5	2,978.2
Data/SM	1.035	1.159	1.041	1.015
Signal/SM	0.012	0.066	0.078	0.111
Signal/$\sqrt{\text{Background}}$	7.582	6.630	5.835	6.423
Signal/$\sqrt{\text{Background} + \sigma(\text{JES}\eta)^2}$	0.367	0.967	1.976	2.497

Table 4.4: The event yields for four stages of VBF cuts in the muon channel. Also shown are the expected significances for a purely statistical measurement (S/$\sqrt{\text{B}}$) and also including the dominant experimental uncertainty, the JES η component. Alpgen is used for the $W +$ QCD jets.

Cut Name	W $N_{jets} = 2$	M_{jj}	M_{jj}, high p_{T}	M_{jj}, high p_{T}, outside lepton veto
$W \rightarrow \mu\nu$	100.0	2.3	1.3	0.7
$W \rightarrow \tau\nu$	100.0	2.5	1.4	0.7
$Wbb \rightarrow l\nu$	100.0	2.8	1.8	1.0
$Wcc \rightarrow l\nu$	100.0	2.9	1.7	0.9
$Wc \rightarrow l\nu$	100.0	1.7	1.0	0.5
QCD (data driven)	100.0	2.4	0.7	0.5
ZZ	100.0	0.9	0.4	0.3
$Z \rightarrow \mu\mu$	100.0	2.5	1.3	0.7
$Z \rightarrow \tau\tau$	100.0	1.7	0.8	0.5
$t\bar{t}$	100.0	1.5	0.9	0.6
single top	100.0	4.0	2.1	1.6
VBF $W \rightarrow \tau\nu$	100.0	14.7	10.7	8.9
VBF Signal Z	100.0	8.8	6.0	4.9
VBF Signal W	100.0	13.3	8.6	7.1

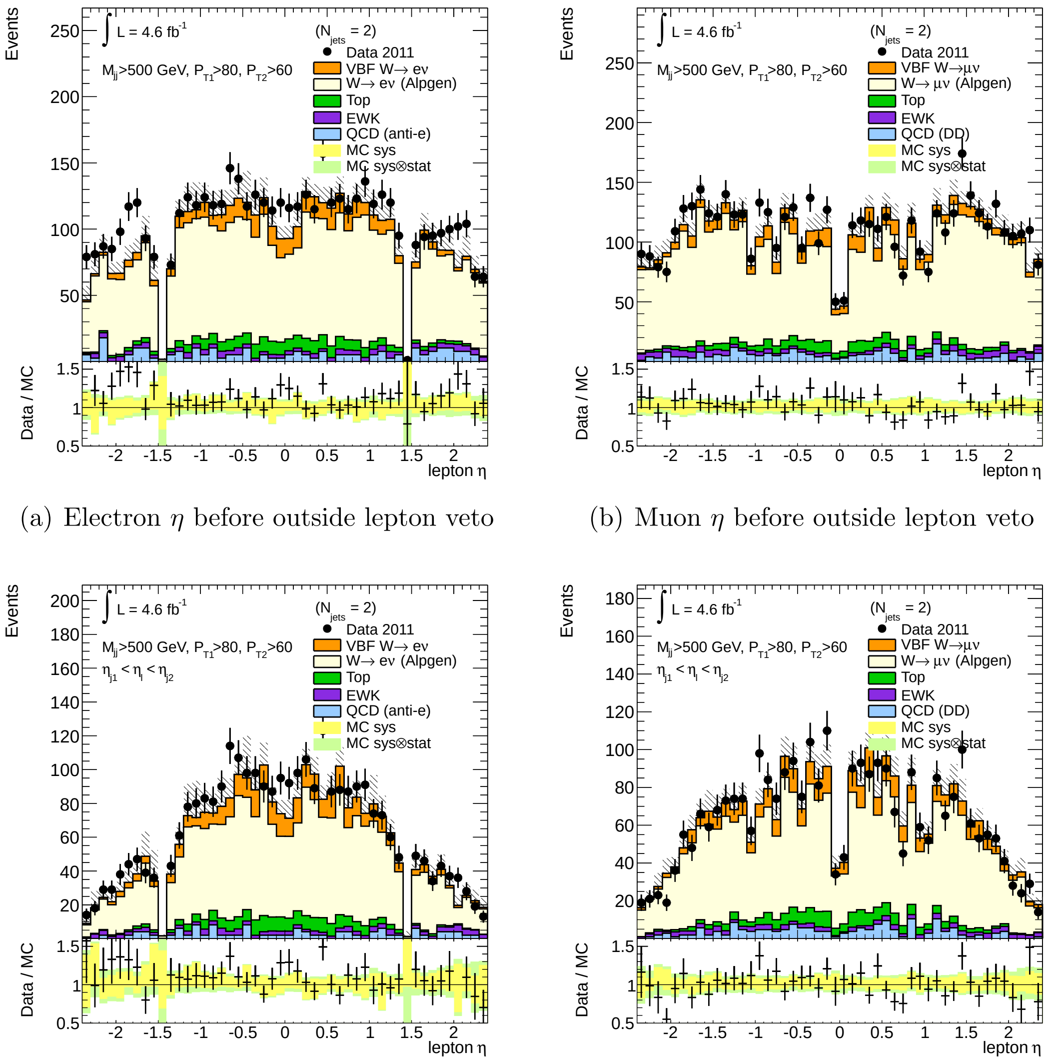
Table 4.5: Percentage events surviving after the a given cut compared to the number after the $N_{jets} = 2$ cut in the muon channel.

Cut Name	W $N_{jets} = 2$	M_{jj}	M_{jj}, high p_{T}	M_{jj}, high p_{T}, outside lepton veto
$W \rightarrow e\nu$	204,848.6	4,714.3	2,622.2	1,518.9
$W \rightarrow \tau\nu$	5,412.6	152.7	100.6	61.8
$Wbb \rightarrow l\nu$	6,508.4	169.8	104.3	60.9
$Wcc \rightarrow l\nu$	15,669.0	452.3	277.6	166.0
$Wc \rightarrow l\nu$	35,045.6	593.1	355.3	183.8
QCD (W)	43,405.3	815.7	266.5	132.3
ZZ	59.5	1.2	0.6	0.5
$Z \rightarrow ee$	18,957.0	388.8	133.8	94.3
$Z \rightarrow \tau\tau$	1,067.8	19.6	10.1	6.4
$t\bar{t}$	11,838.6	190.4	120.3	85.3
single top	6,143.3	256.7	136.6	104.2
VBF $W \rightarrow \tau\nu$	101.7	19.0	11.6	9.4
VBF Signal Z	44.8	4.0	2.9	2.3
VBF Signal W	4,166.4	608.2	397.6	338.7
Data	358,340.0	9,962.0	5,114.0	3,028.0
SM prediction	353,268.6	8,385.7	4,539.9	2,764.8
Data/SM	1.014	1.188	1.126	1.095
Signal/SM	0.012	0.073	0.088	0.123
Signal/$\sqrt{\mathrm{Background}}$	7.052	6.896	6.178	6.877
Signal/$\sqrt{\mathrm{Background} + \sigma(\mathrm{JES}\eta)^2}$	0.378	1.058	2.400	2.890

Table 4.6: The event yields for four stages of VBF cuts in the electron channel. Also shown are the expected significances for a purely statistical measurement (S/$\sqrt{\mathrm{B}}$) and also including the dominant experimental uncertainty, the JES η component. Alpgen is used for the W + QCD jets.

Cut Name	W $N_{jets} = 2$	M_{jj}	M_{jj}, high p_{T}	M_{jj}, high p_{T}, outside lepton veto
$W \to e\nu$	100.0	2.3	1.3	0.7
$W \to \tau\nu$	100.0	2.8	1.9	1.1
$Wbb \to l\nu$	100.0	2.6	1.6	0.9
$Wcc \to l\nu$	100.0	2.9	1.8	1.1
$Wc \to l\nu$	100.0	1.7	1.0	0.5
QCD (W)	100.0	1.9	0.6	0.3
ZZ	100.0	2.0	1.0	0.8
$Z \to ee$	100.0	2.1	0.7	0.5
$Z \to \tau\tau$	100.0	1.8	0.9	0.6
$t\bar{t}$	100.0	1.6	1.0	0.7
single top	100.0	4.2	2.2	1.7
VBF $W \to \tau\nu$	100.0	18.7	11.4	9.2
VBF Signal Z	100.0	9.0	6.4	5.2
VBF Signal W	100.0	14.6	9.5	8.1

Table 4.7: Percentage events surviving after the a given cut compared to the number after the $N_{jets} = 2$ cut in the elec channel.



(a) Electron η before outside lepton veto

(b) Muon η before outside lepton veto

(c) Electron η after outside lepton veto

(d) Muon η after outside lepton veto

Figure 4.6: Lepton η distributions before and after the outside lepton veto.

and Table 4.5 for the muon channel and Table 4.6 and Table 4.7 for the electron channel. The sequential application of each cut produces a decrease in the statistical significance. However when taking into account systematic errors the expected significance increases with each cut. This is shown in the final line of each table labelled "Signal/$\sqrt{\text{Background} + \sigma^2(\text{JES } \eta)}$". Only the dominant η-intercalibration component of the jet energy scale is included in this error for simplicity.

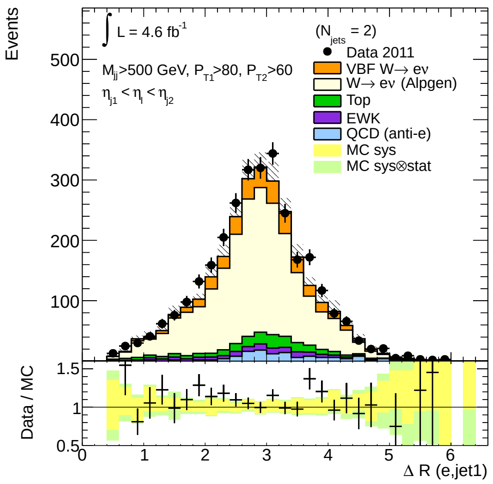
A selection of distributions after the signal selection are shown separately for electron and muon channels using ALPGEN for $W+$ QCD jets modelling in Figures 4.7-4.9. Figures 4.10- 4.12 show the same distribution using SHERPA for $W+$ QCD jets modelling. Combined lepton channels shown in Figure 4.13 with both ALPGEN and SHERPA $W+$ QCD jets Monte Carlo. Note that at this stage the tables and plots are shown with the number of events expected, normalized to the best available theoretical cross section. In Chapter 6 these plots will be shown with the normalization extracted from a fit.

In both the electron and muon channels the application of the M_{jj} cut is seen to result in an excess of data compared to the expectations from Monte Carlo simulation. This is thought to be due to poor modelling of low p_{T}, high rapidity jets, which are enhanced by the M_{jj} cut. In the muon channel this excess is removed after the application of the high p_{T} and outside lepton veto cuts. These are less effective in the electron case, where a $\sim 10\%$ discrepancy remains. Possible sources of this are the electron-jet overlap removal procedure or the QCD multi-jet background. These were both investigated but not effective remedy was obtained.

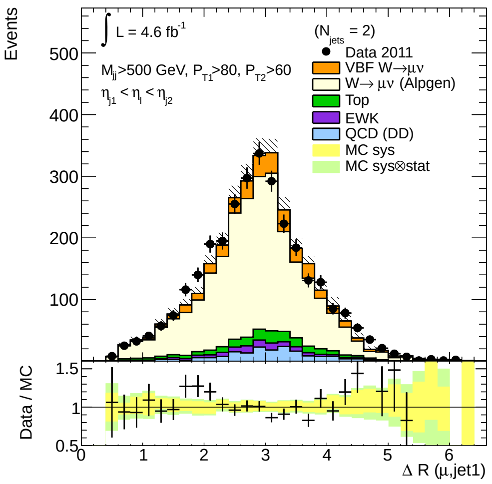
Additional cuts on quantities such as $\Delta y(j_1, j_2)$, $\eta^{j_1} \times \eta^{j_2}$ and a central jet veto were considered as these have been shown in many phenomenology papers to be useful in improving expected signal significance, since they preferentially select events with forward jets and/or reject events with central radiation (see e.g. [50, 51, 61, 62]) but not found to improve the expected significance. A cut on the azimuthal angle between the W and the vector sum p_{T} of the tag jets was found to be useful in improving the signal significance and purity, however due to poor modelling it could not be used.

4.5 Validation region selection

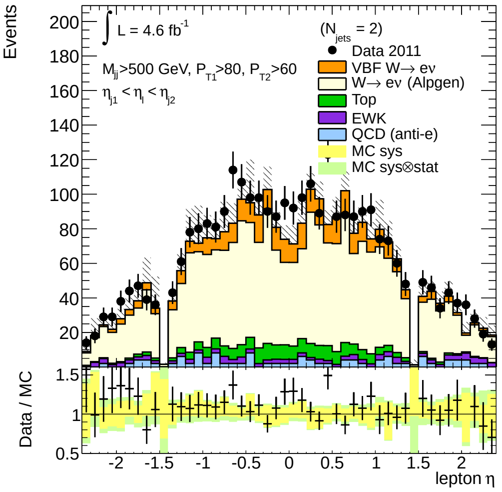
A validation region was defined to provide a cross-check for the background modelling in an orthogonal phase-space to the signal. This defined as the $N_{jets} = 3$ region, plots of kinematic variables are shown for lepton channels separately in Figure 4.14 and combined in Figure 4.15. This results in a validation region with 4-5% signal contamination; the contribution to each validation region from each source of background and



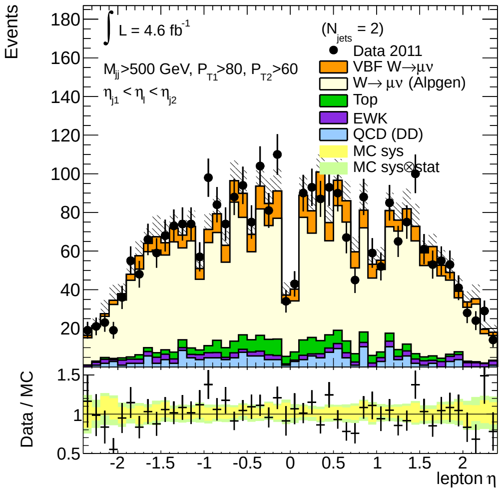
(a) $\Delta R(\ell, \text{jet})$ (electron channel)



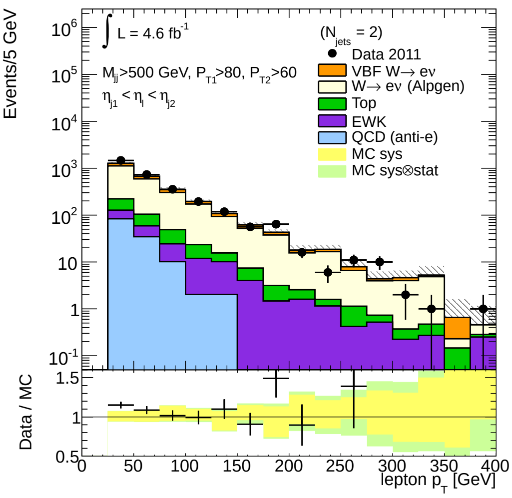
(b) $\Delta R(\ell, \text{jet})$ (muon channel)



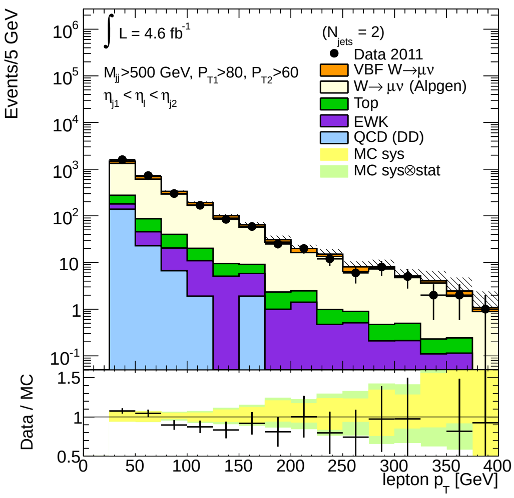
(c) Electron η



(d) Muon η

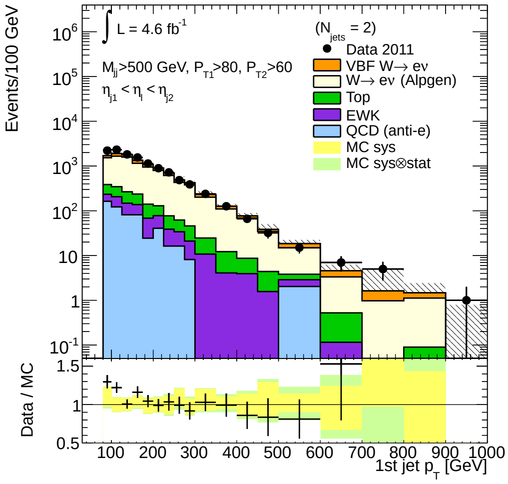


(e) Electron p_{T}

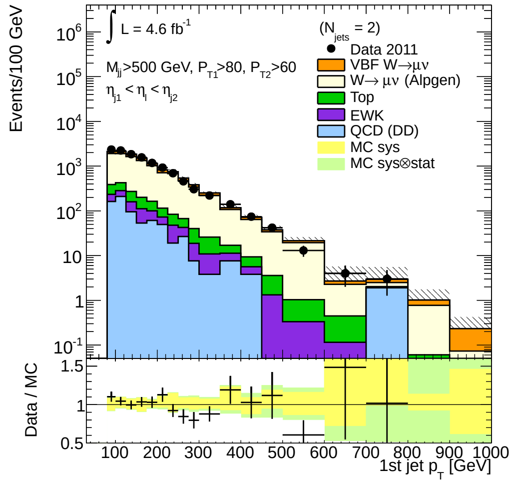


(f) Muon p_{T}

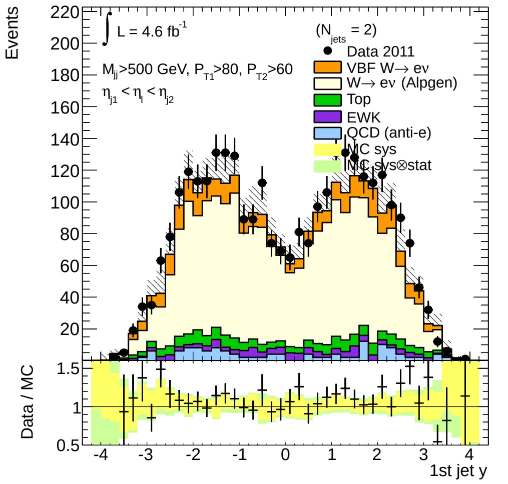
Figure 4.7: Distributions of lepton kinematic variables after all VBF selections in the electron and muon channels. ALPGEN is used for the $W+$ QCD jets simulation and plots are normalized to the luminosity of the data.



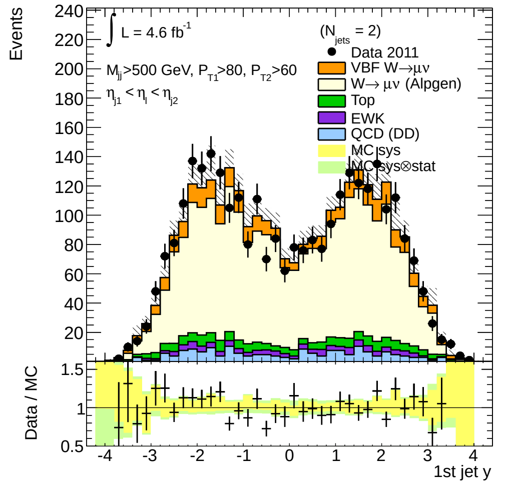
(a) Leading jet p_{T} (electron channel)



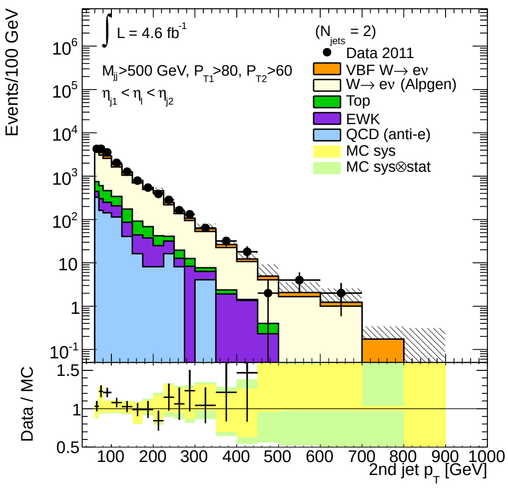
(b) Leading jet p_{T} (muon channel)



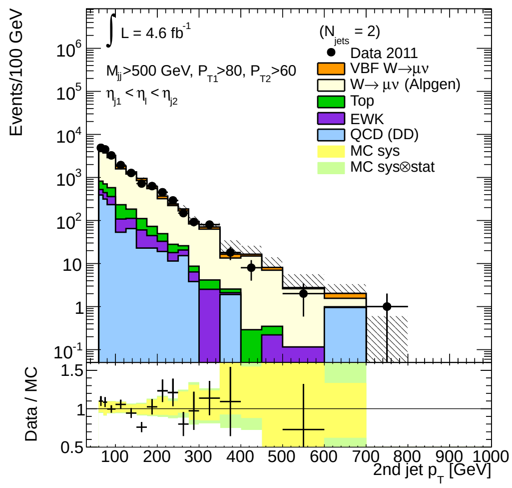
(c) Leading jet rapidity (electron channel)



(d) Leading jet rapidity (muon channel)

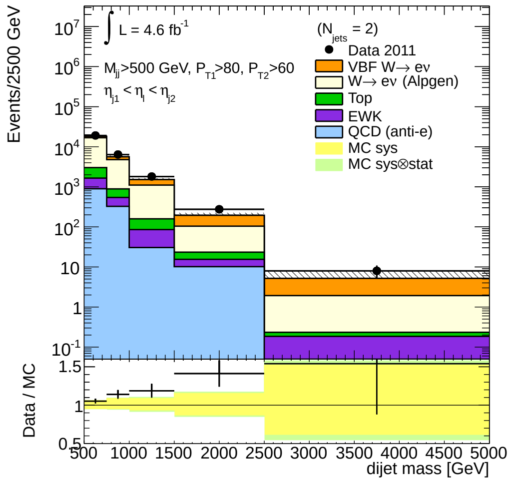


(e) Sub-leading jet p_{T} (electron channel)

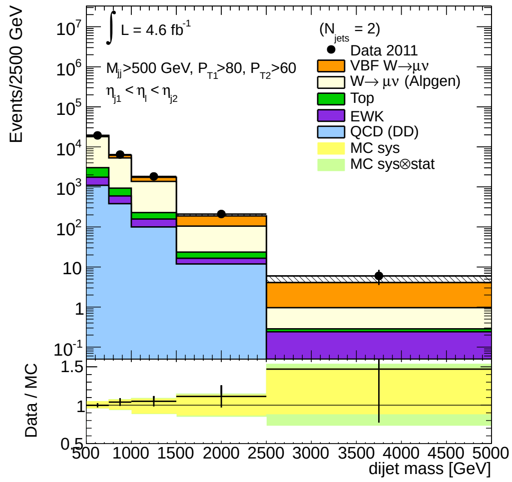


(f) Sub-leading jet p_{T} (muon channel)

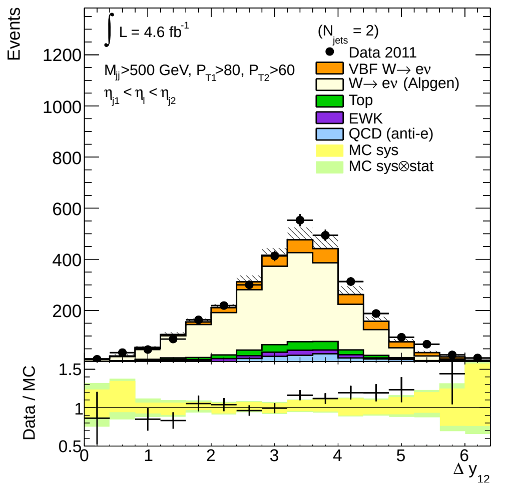
Figure 4.8: Jet distributions after all VBF selections in the electron and muon channels. ALPGEN is used for the W + QCD jets simulation and plots are normalized to the luminosity of the data.



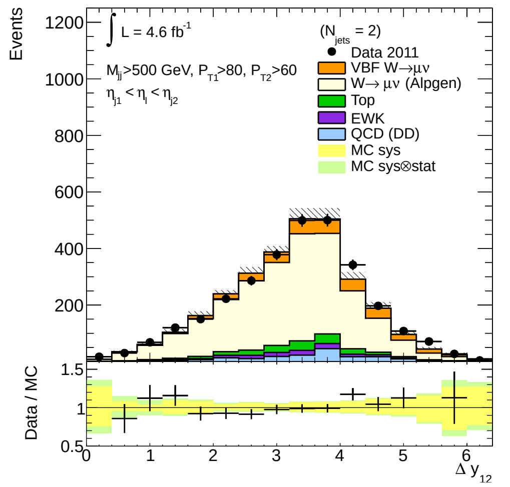
(a) M_{jj} (electron channel)



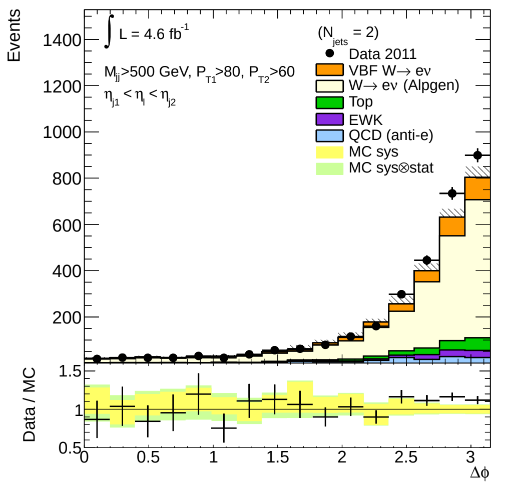
(b) M_{jj} (muon channel)



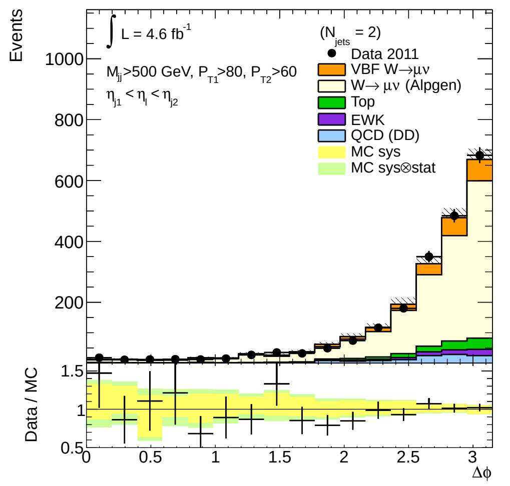
(c) $\Delta y(j_1, j_2)$ (electron channel)



(d) $\Delta y(j_1, j_2)$ (muon channel)

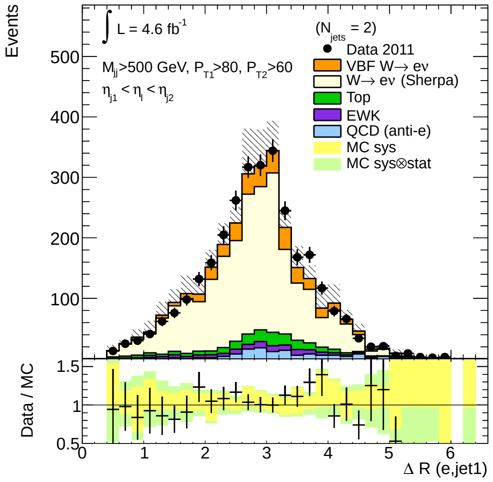


(e) $\Delta\phi(j_1, j_2)$ (electron channel)

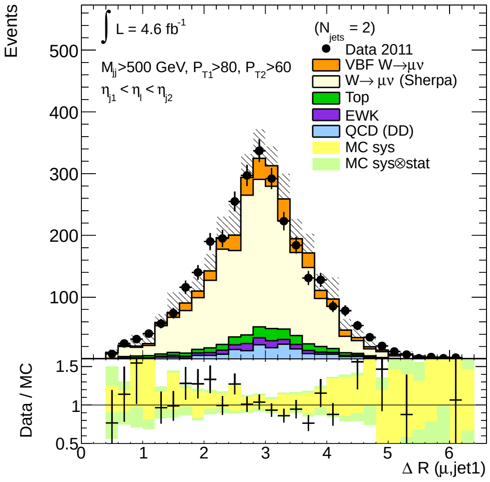


(f) $\Delta\phi(j_1, j_2)$ (muon channel)

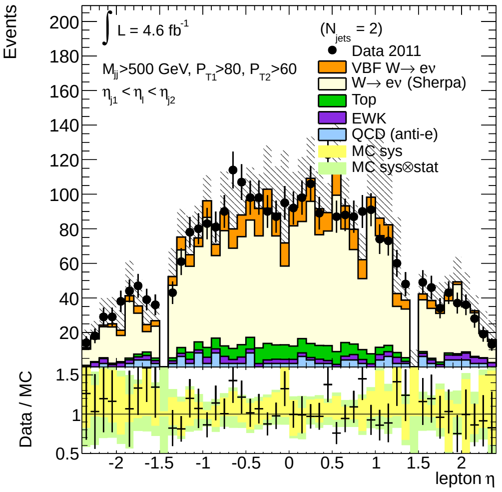
Figure 4.9: Distributions of dijet variables after all VBF selections in the electron and muon channels. ALPGEN is used for the $W+$ QCD jets simulation and plots are normalized to the luminosity of the data.



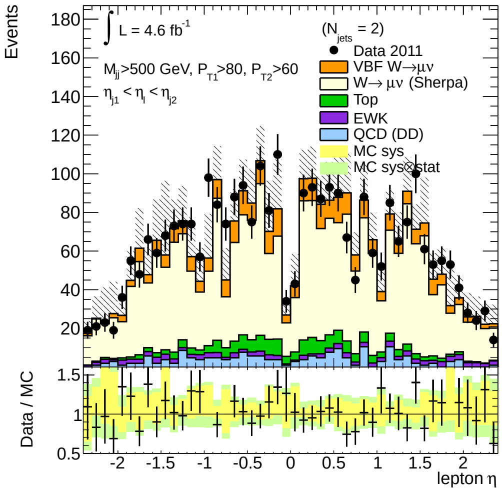
(a) $\Delta R(\ell, \text{jet})$ (electron channel)



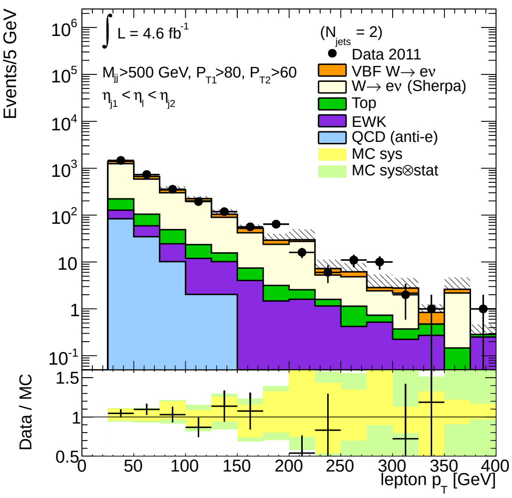
(b) $\Delta R(\ell, \text{jet})$ (muon channel)



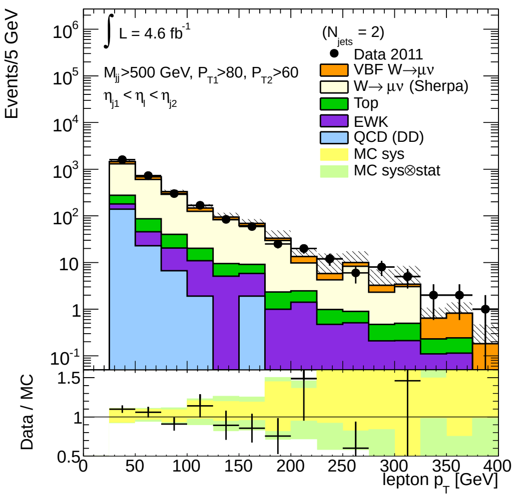
(c) Electron η



(d) Muon η

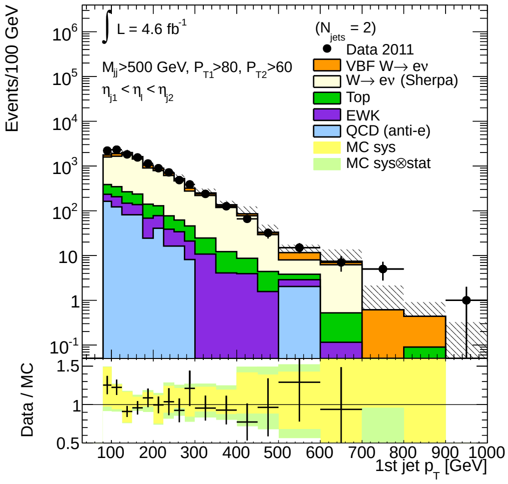


(e) Electron p_{T}

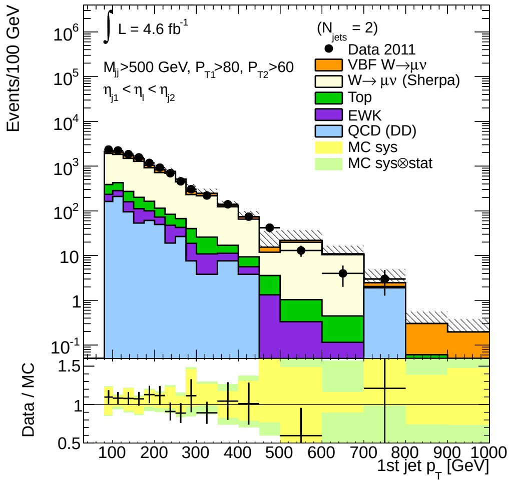


(f) Muon p_{T}

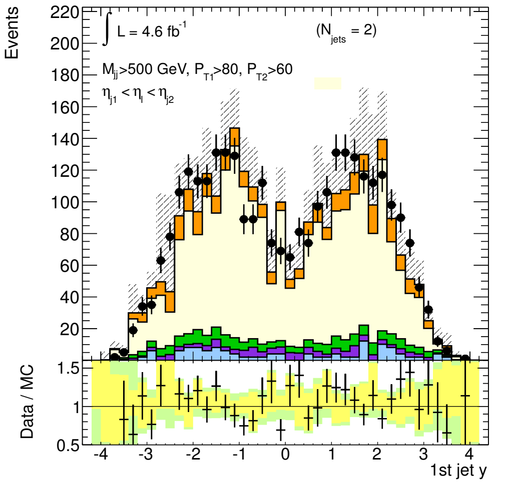
Figure 4.10: Distributions of lepton kinematic variables after all VBF selections in the electron and muon channels. SHERPA is used for the $W+$ QCD jets simulation and plots are normalized to luminosity of the data.



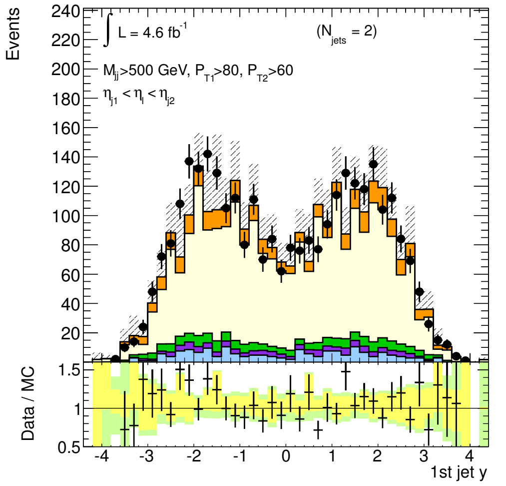
(a) Leading jet p_{T} (electron channel)



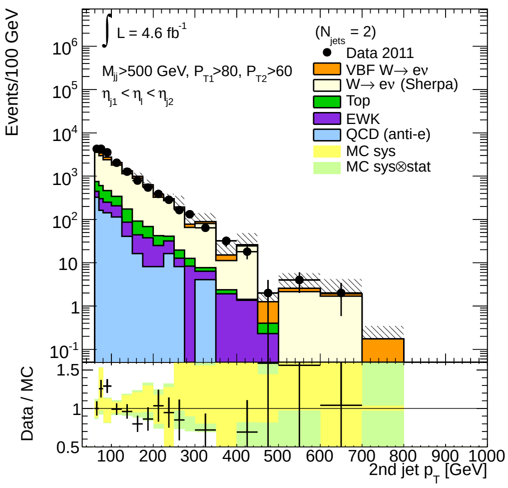
(b) Leading jet p_{T} (muon channel)



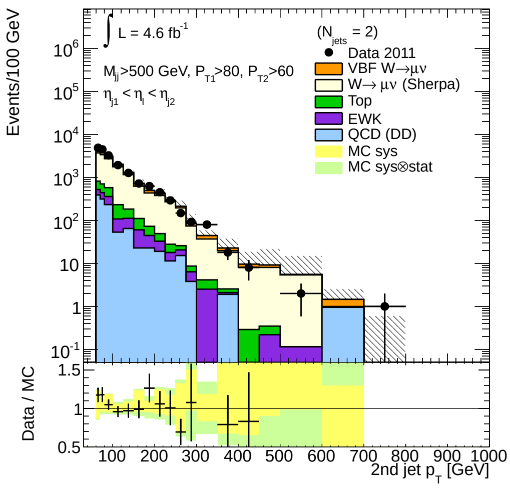
(c) Leading jet rapidity (electron channel)



(d) Leading jet rapidity (muon channel)

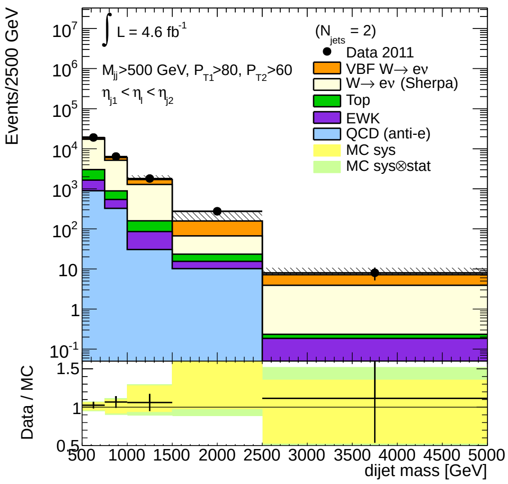


(e) Sub-leading jet p_{T} (electron channel)

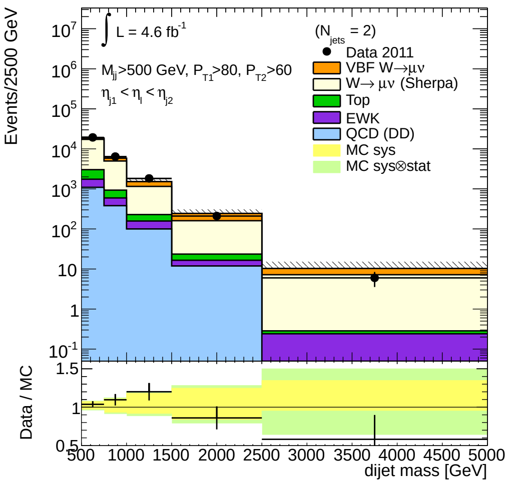


(f) Sub-leading jet p_{T} (muon channel)

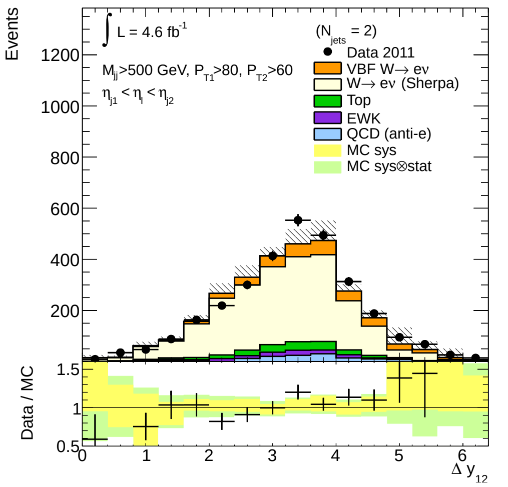
Figure 4.11: Jet distributions after all VBF selections in the electron and muon channels. Sherpa is used for the $W+$ QCD jets simulation and plots are normalized to the luminosity of the data.



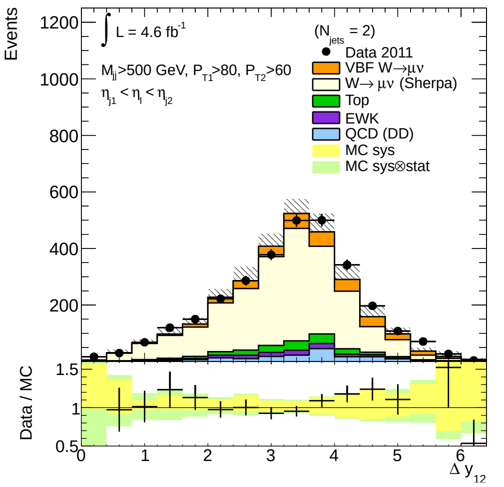
(a) M_{jj} (electron channel)



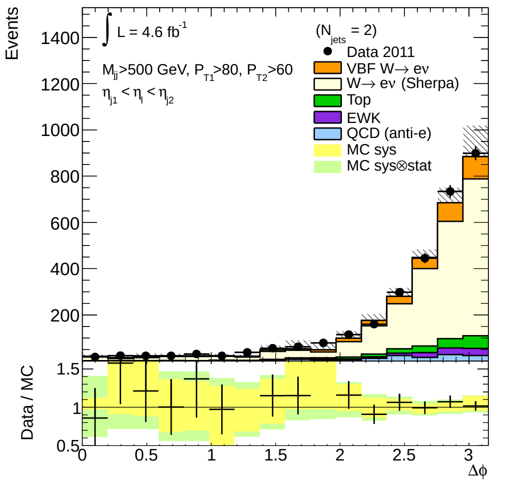
(b) M_{jj} (muon channel)



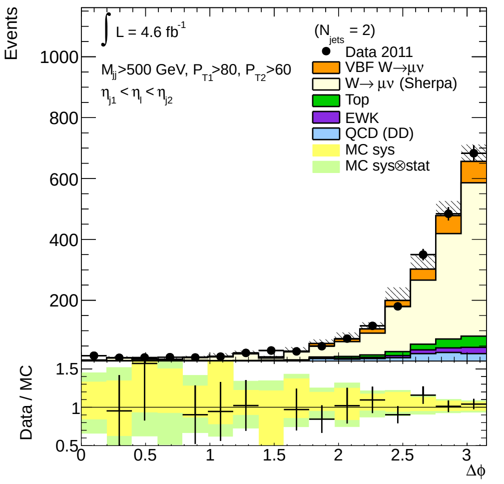
(c) $\Delta y(j_1, j_2)$ (electron channel)



(d) $\Delta y(j_1, j_2)$ (muon channel)

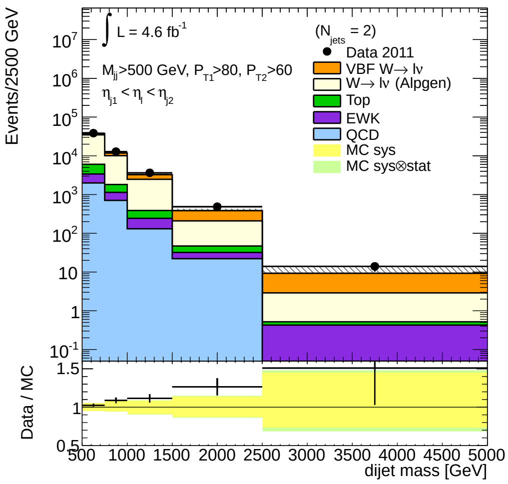


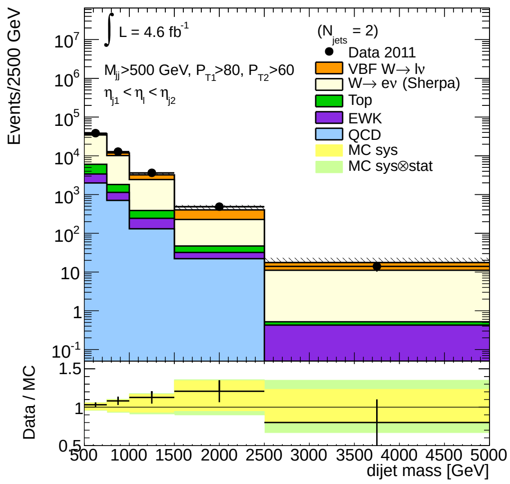
(e) $\Delta\phi(j_1, j_2)$ (electron channel)



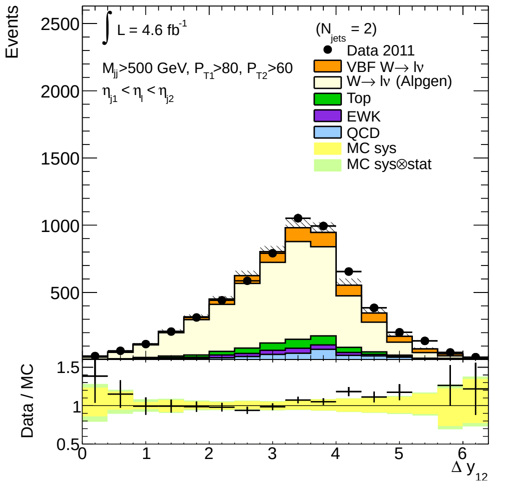
(f) $\Delta\phi(j_1, j_2)$ (muon channel)

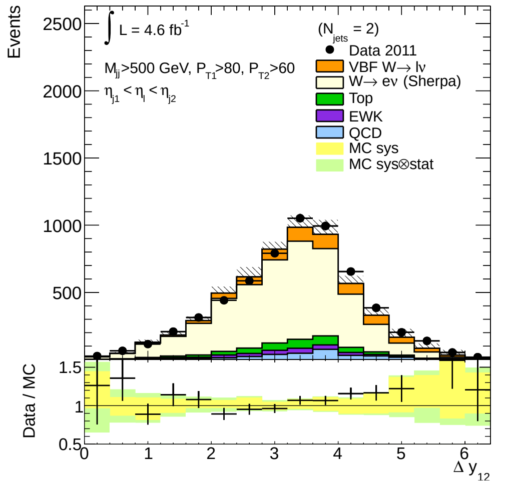
Figure 4.12: Distributions of dijet variables after all VBF selections in the electron and muon channels. SHERPA is used for the $W +$ QCD jets simulation and plots are normalized to the luminosity of the data.



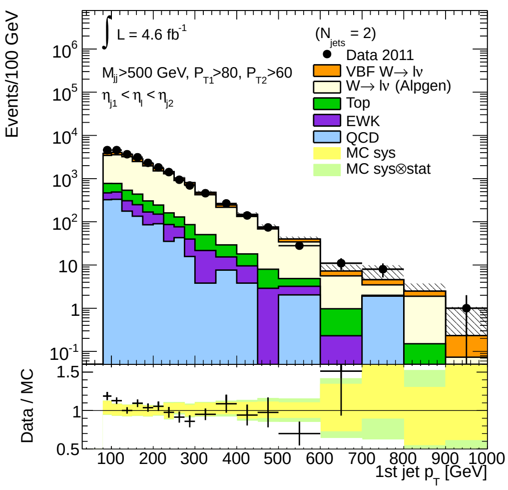


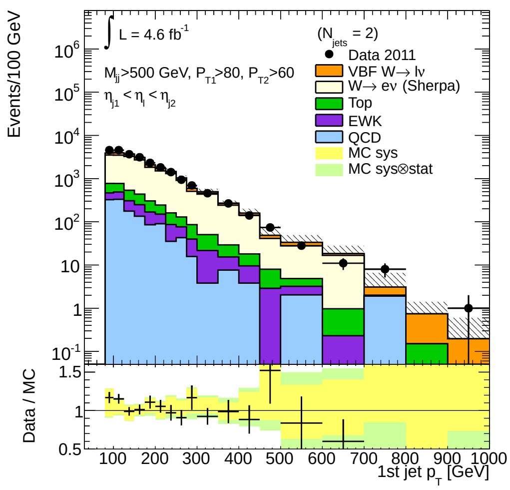
(a) Combined M_{jj} (ALPGEN $W+$ QCD jets) (b) Combined M_{jj} (SHERPA $W+$ QCD jets)





(c) Combined $\Delta y(j_1, j_2)$ (ALPGEN $W+$ QCD jets) (d) Combined $\Delta y(j_1, j_2)$ (SHERPA $W+$ QCD jets)





(e) Combined Leading jet p_{T} (ALPGEN $W+$ QCD jets) (f) Combined Leading jet p_{T} (SHERPA $W+$ QCD jets)

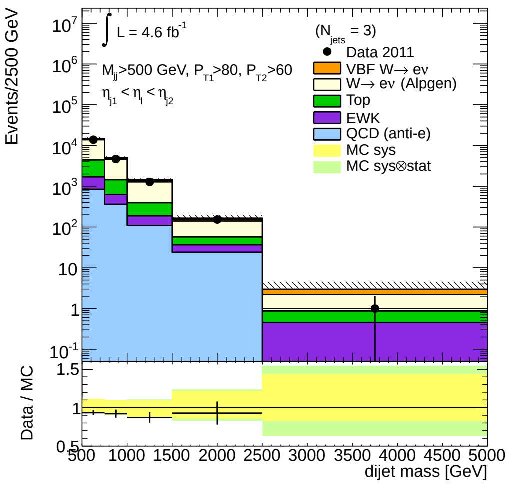
Figure 4.13: Distributions after all VBF selections in the combined electron and muon channel. Plots on the left use ALPGEN for $W+$ QCD jets and plots on the right use SHERPA. All plots are normalized to the luminosity of the data.

signal is shown in Tables 4.8 and 4.9.

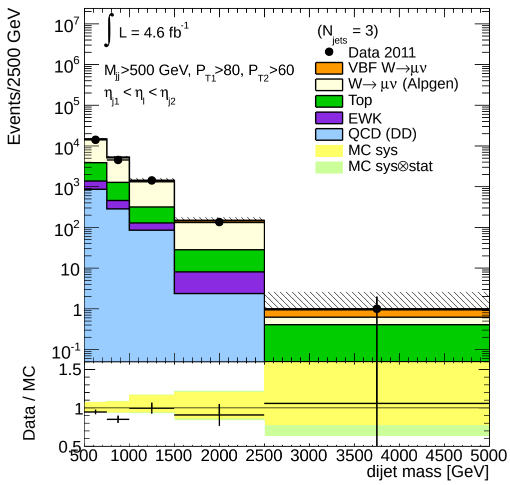
In the validation region the data agrees with the Monte Carlo at the level of 8%, which is within the systematic uncertainty in this region. These Tables also show that the data Monte Carlo agreement in the electron and muon channels agree to the 1% level. This is an interesting contrast to the signal region, where a large difference between the lepton channels is observed. The kinematic distributions also show good agreement in all distributions within the systematic errors. These results confirm that the background model is providing a good estimate in a close, but orthogonal region to the signal region.

CutName	$N_{jets} = 3$, M_{jj}, high p_{T}, outside lepton veto
$W \to e\nu$	1,215.3
$W \to \tau\nu$	40.0
$Wbb \to l\nu$	47.4
$Wcc \to l\nu$	117.3
$Wc \to l\nu$	131.5
QCD (W)	152.0
ZZ	0.4
$Z \to ee$	124.6
$Z \to \tau\tau$	5.6
$t\bar{t}$	303.4
single top	100.1
VBF $W \to \tau\nu$	3.0
VBF Signal Z	2.4
VBF Signal W	108.8
Data	2,167.0
SM prediction	2,351.8
Data/SM	0.921
Signal/SM	0.046
Signal/$\sqrt{\text{Background}}$	2.296
Signal/$\sqrt{\text{Background} + \sigma(\text{JES}(\eta))^2}$	0.785

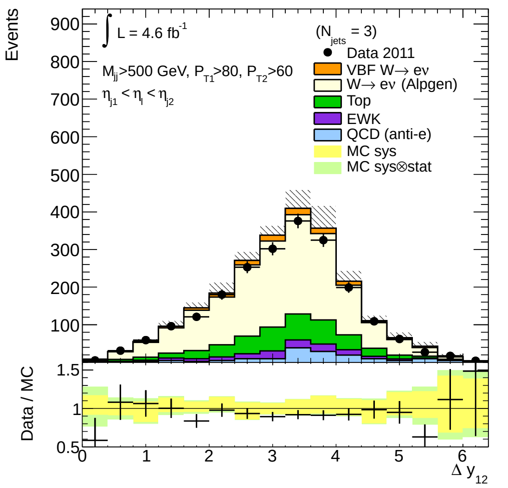
Table 4.8: Validation region in the electron channel using ALPGEN for the $W +$ QCD jets background.



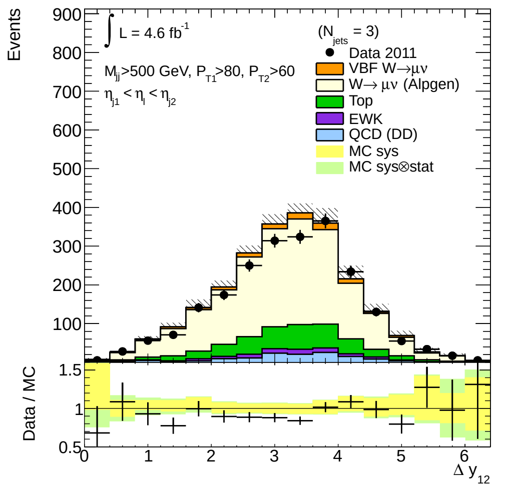
(a) M_{jj} (electron channel)



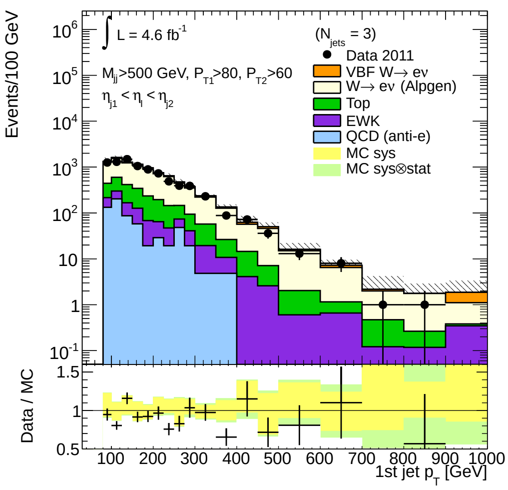
(b) M_{jj} (muon channel)



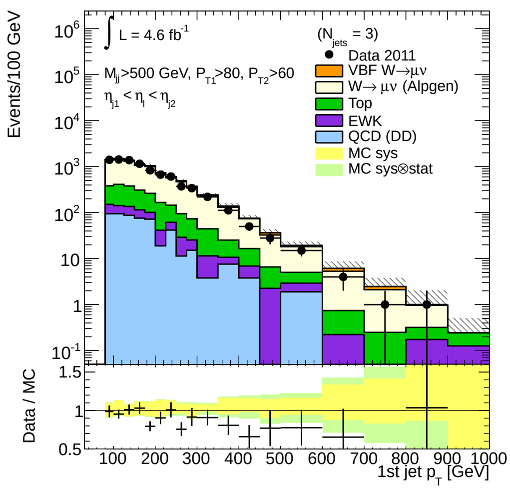
(c) $\Delta y(j_1, j_2)$ (electron channel)



(d) $\Delta y(j_1, j_2)$ (muon channel)

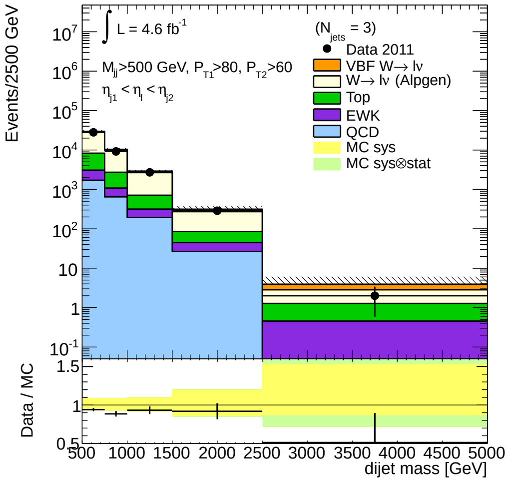


(e) Leading jet p_{T} (electron channel)

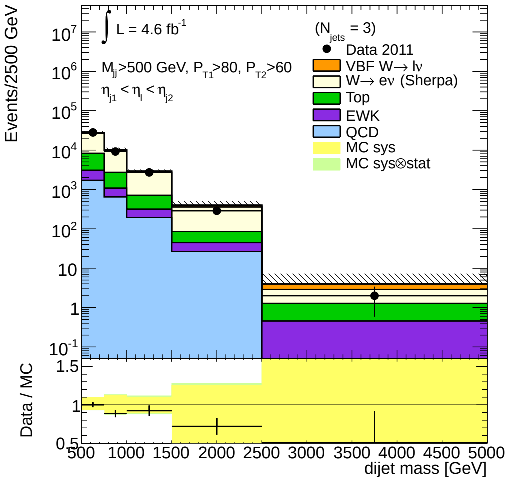


(f) Leading jet p_{T} (muon channel)

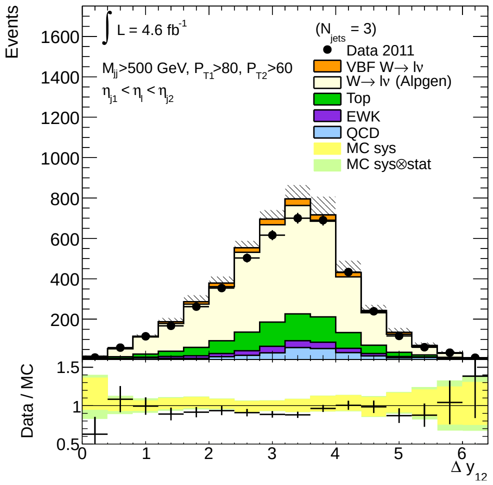
Figure 4.14: Detector-level distributions in the VBF validation region for electrons and muons normalized to the theoretical cross section.



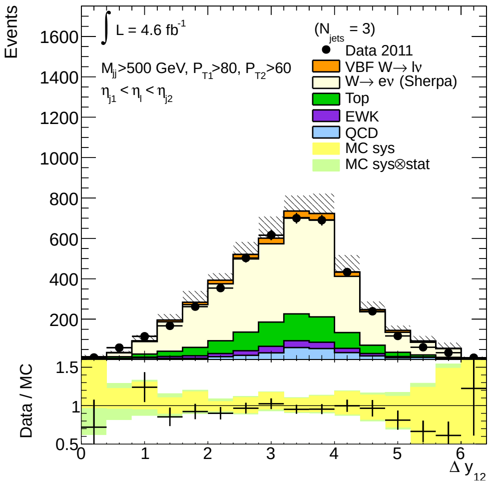
(a) Alpgen M_{jj}



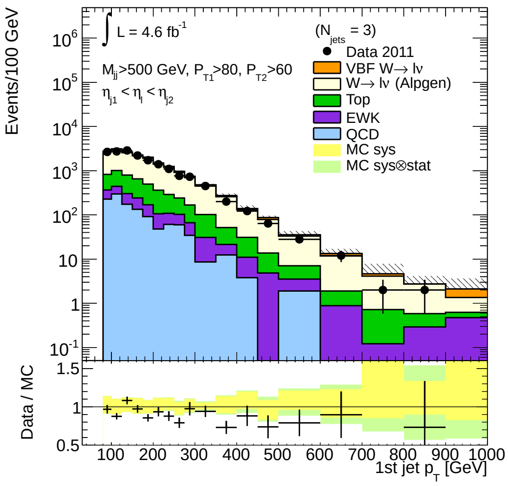
(b) Sherpa M_{jj}



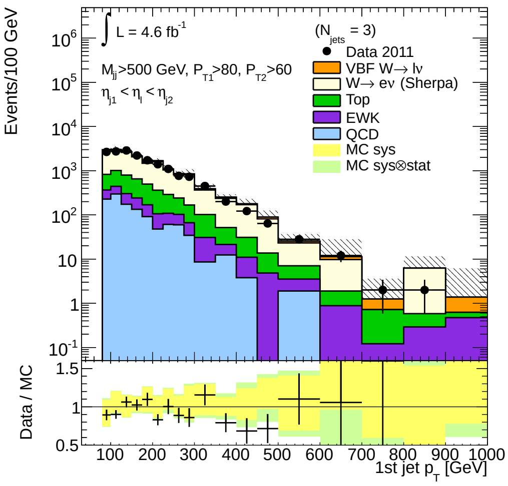
(c) Sherpa $\Delta y(j_1, j_2)$



(d) Sherpa $\Delta y(j_1, j_2)$



(e) Leading jet p_{T} Alpgen



(f) Leading jet p_{T} Sherpa

Figure 4.15: Detector-level distributions in the VBF validation region comparing the combined channels with $W+$ QCD jets modelled by Alpgen and Sherpa, normalized to the theoretical cross section.

CutName	$N_{jets} = 3$, M_{jj}, high p_{T}, outside lepton veto
$W \to \mu\nu$	1,309.7
$W \to \tau\nu$	41.2
$Wbb \to l\nu$	70.3
$Wcc \to l\nu$	124.6
$Wc \to l\nu$	134.2
QCD (data driven)	132.8
ZZ	0.3
$Z \to \mu\mu$	70.3
$Z \to \tau\tau$	6.5
$t\bar{t}$	279.8
single top	99.1
VBF $W \to \tau\nu$	2.6
VBF Signal Z	0.7
VBF Signal W	99.3
Data	2,204.0
SM prediction	2,371.3
Data/SM	0.929
Signal/SM	0.042
Signal/$\sqrt{\text{Background}}$	2.084
Signal/$\sqrt{\text{Background} + \sigma(\text{JES}\eta)^2}$	0.772

Table 4.9: Validation region in the muon channel using ALPGEN for the $W+$ QCD jets background.

4.6 Systematic uncertainties

The selection is repeated for each of the systematic uncertainties for each object, as defined in Chapter 3. The percentage change in total event yield is shown in Table 4.10[7]. The lepton identification and calibration systematics give a total uncertainty of $1.8 - 3.6\%$. The dominant contribution to the experimental uncertainty is the η-intercalibration term, labelled as "JES (η)".

4.6.1 Theoretical uncertainties

The effect of modelling, or theoretical, uncertainties were investigated. For the top, diboson, $W \to \tau\nu$, and $Z/\gamma^* +$jets backgrounds these are taken from the cross section

[7] The errors presented here do not directly correspond to the errors on the final result, since these numbers display changes in normalization, while the cross section is extracted with a fit which depends on changes in shape.

Effect	$N_{jets} = 2$	$N_{jets} = 3$
e energy scale	1.26%	1.17%
e energy smearing	0.03%	0.07%
e reco. efficiency SF	0.8%	0.79%
e id efficiency SF	0.91%	0.89%
e trigger SF	0.54%	0.53%
Electron total	1.83%	1.75%
μ pt scale turned off	0.62%	0.47%
μ MS pt smearing	0.16%	0.15%
μ ID pt smearing	0.02%	0.16%
μ trigger scale factor	1.59%	1.59%
μ reco. eff. SF	0.26%	0.26%
Muon total	1.73%	1.69%
Jet energy resolution	0.57%	1.72%
JES (other)	1.29%	2.59%
JES η	4.35%	5.66%
JES (μ)	0.48%	2.05%
JES (NPV)	0.25%	0.71%
JES (Close-by)	0.14%	0.97%
JES (Flavor response)	1.14%	2.2%
JES (Flavor composition)	0.1%	0.42%
JES	4.71%	7.03%
JES & JER	4.75%	7.24%
E_T^{miss} soft terms reso.	0.02%	0.04%
E_T^{miss} soft terms scale	0.42%	0.4%
$E_{\mathrm{T}}^{\mathrm{miss}}$ total	0.42%	0.4%
QCD Template	1.62%	1.51%
Luminosity	1.8%	1.8%

Table 4.10: Experimental systematic errors shown as percentage change in the predicted events in the VBF signal region. The electron and muon systematics are shown as percentage changes in the respective lepton channel, while the jet, $E_{\mathrm{T}}^{\mathrm{miss}}$ and QCD errors are shown for combined electron and muon channels. The derivation of the QCD errors is explained in Chapter 5, while all other errors are explained in their relevant sections in Chapter 3.

Process	Data set	Note
	Samples used for theory error estimation	
$W \to \mu\nu$	113302-113307	Alternative renormalization scale definition
$W \to \mu\nu$	113320-113325	Renormalization and factorization scales $\times 0.5$
$W \to \mu\nu$	113326-113331	Renormalization and factorization scales $\times 2$
$W \to \mu\nu$	113338-113343	p_{T} matching scale variation
$W \to \mu\nu$	113350-113355	Δr matching parameter variation

Table 4.11: $W+$ QCD jets background samples varied to evaluate theoretical uncertainties, modelled using ALPGEN +HERWIG using with AFII detector simulation (fast simulation, as described in Section 4.2.1). These variations are explained in the main text.

uncertainties as detailed in Section 4.2.2. For the $W+$ QCD jets and signal processes more detailed investigations were made. Since a fit is made to extract the signal strength, we are only interested in variations in the shape of the $W+$ QCD jets background. Changes to the shape and absolute normalization of the signal are considered since both affect the uncertainty on the signal strength.

All changes were made on the muon channel only, under the assumption that the electron channel changes identically. This is justified since lepton universality has been well established by the LEP experiments [9] and so the W couples identically to electrons and muons. The $W+$ QCD jets theoretical uncertainties were estimated using samples simulated using AFII fast simulation[8] listed in Table 4.11.

The theoretical uncertainties on the $W+$ EWK jets signal were estimated using generator level samples analysed using the Rivet framework [140]. For each variation the ratio Nominal/Systematic was calculated in four bins of M_{jj} and this was propagated to the reconstructed distributions by weighting events by Systematic/Nominal according to their generator level M_{jj}.

Jet matching uncertainties were estimated by varying criteria for matching matrix element to parton shower jets, within reasonable variation as suggested by the authors' of the respective generators. These parameters have no physical meaning and

[8] As described in Section 4.2.2.

are artefacts of the implementation of the particular Monte Carlo algorithm. These alternative parameters are equally justified to the default choices, and so any variation in the result due to these demonstrates the modelling dependence of the result.

For the $W+$ QCD jets background this was done by comparing the p_{T} matching scale of 20 GeV and 25 GeV and the ΔR criteria of 0.7 and 0.4 using the AFII samples given in Table 4.11.

For the signal, generator-level samples with different values of the CKKW matching criteria were used. The matching parameter in the run card is changed from `CKKW sqr(15/E_CMS)` nominal to `CKKW sqr(30/E_CMS)`.

Renormalization and factorization scale uncertainties were estimated by varying the renormalization (μ_R) and factorization (μ_F) scales. The dependence on the renormalization and factorization scales is an artifact of the limited number of orders of the perturbative calculation. The variation of these scales is thus commonly used to indicate the level of dependence of a given calculation on higher orders which have not been calculated. The amount of variation is arbitrary, but following convention our samples are varied up and down by factors of two; these are labelled as "qfac05" and "qfac20" respectively. In the $W+$ QCD jets samples the default scale choice is $\mu_R = \sqrt{m_W^2 + \sum p_{T\mathrm{parton}}^2}$ with the sum running over all the final state particles except for the decay products of the boson. An alternative scale choice of $\mu_R = \sqrt{m_W^2 + p_T(W)^2}$ was checked.

Hard-scatter and parton shower uncertainties were investigated by changing the background $W+$ QCD jets modelling from ALPGEN to SHERPA[9]. Ideally the modelling of the hard-scatter (i.e. the matrix element calculation) and the parton shower would be varied independently; for example by comparing ALPGEN + HERWIG to ALPGEN + PYTHIA for the parton shower variation, and to SHERPA for the hard-scatter. However in 2011 ALPGEN showered with PYTHIA were not available, and so

[9] ALPGEN samples showered with PYTHIA were not available in 2011.

SHERPA samples were used to study both effects at once.

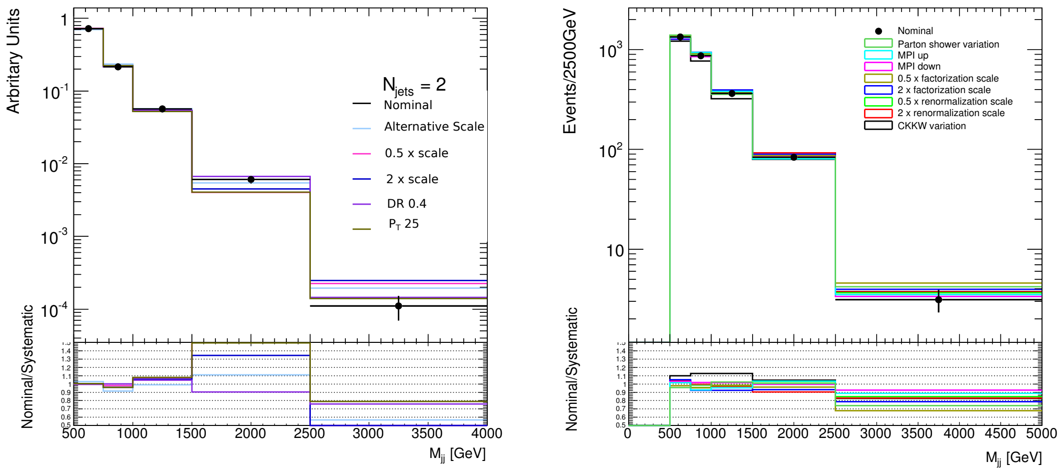
Multi-parton interaction (MPI) uncertainties were investigated only for the signal by changing the parameters that control MPI. The total non-diffractive cross section factor `SIGMA_ND_FACTOR` is changed from 0.48 (default) to 0.43 and 0.53[10]. This principally affects low p_{T} jets.

PDF Uncertainties were investigated by reweighting the signal and $W+$ QCD jets backgrounds to PDF error sets using LHAPDF version 5.8.9 [141]. Since error sets do not exist for CTEQ6L1 (the PDF set used for the generation of signal and $W+$ QCD jets backgrounds), each sample was reweighted to the central values and error sets of CTEQ61m and MSTW2008. The variation between the central value and each eigenvector was combined as recommended using the 'Master equation' [13] into a single up and down variation:

$$\begin{aligned} \Delta X^{+} &= \sqrt{\textstyle\sum_{i=1}^{N} \max(X_i^{+} - X_0, X_i^{-} - X_0, 0)^2} \\ \Delta X^{-} &= \sqrt{\textstyle\sum_{i=1}^{N} \max(X_0 - X_i^{+}, X_0 - X_i^{-}, 0)^2} \end{aligned} \tag{4.1}$$

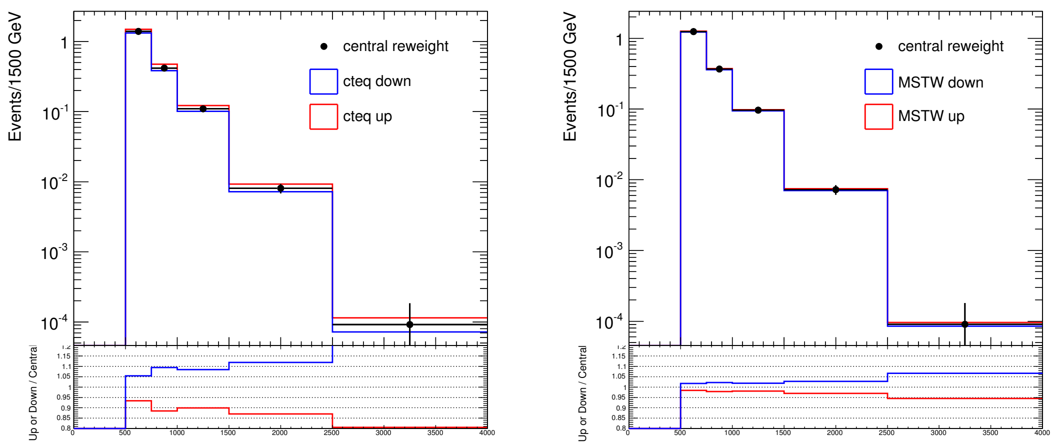
The impact of the background and signal theory errors on the M_{jj} shape is shown in Figure 4.16 for the generator variations, and Figure 4.17 for PDF variations. These plots are purely indicative of the size of the variations encountered. The propagation to the final result is done with the fit described in the Chapter 6. Nonetheless the results presented here indicate that the largest theoretical uncertainties are due to the jet matching parameters in the background simulation, with up to 40% variation in the background shape in high M_{jj} bins due to this.

[10] Recommended by SHERPA author Hendrik Hoeth for 7 TeV analysis.



(a) Background (ALPGEN) (b) Signal (SHERPA)

Figure 4.16: The variation caused by theoretical variations on the left for $W+$ QCD jets and on the right for $W+$ EWK jets. The $W+$ QCD jets variations are all normalized to the same area, since it is only changes of shape that affect the result.



(a) CTEQ variation (b) MSTW variation

Figure 4.17: The variation in M_{jj} introduced by the error eigenvectors of CTEQ and MSTW PDF sets. The CTEQ variation is larger as their error bands represent 90% C.L. compared to MSTW 68% C.L.

4.7 Summary

A set of cuts which selects a phase-space, with an expected contribution of over 10% VBF, has been defined and justified. This is expected to provide purely statistical significance of two standard deviations per lepton channel. A validation selection was defined which is orthogonal to the signal region and will provide a region to check the background modelling.

CHAPTER 5

QCD Background

In this chapter I describe the data-driven techniques used to estimate the rate and kinematic distributions of the QCD multijet background in the electron and muon channels. The methods presented here are also used in the W/Z-ratio measurement and the $W+$jets measurement, to which I also contributed[1].

5.1 Method overview

QCD background in the electron channel is principally caused by mismeasured jets which pass the identification (ID) requirements of a prompt electron (see Chapter 3). Given the jet resolution, a reasonable fraction of QCD multi-jet events have $E_{\mathrm{T}}^{\mathrm{miss}}$ above the 25 GeV threshold required to enter the signal region.

The cross section for multijet production at ATLAS is very high, and it accounts

[1] Papers in preparation [39, 42]. Results already public from 2010 data on which these methods are based [38, 41].

for $\sim 3 - 15\%$ of events in the inclusive $W +$ jets phase-space (depending on the jet bin). The rate of jets faking electrons is low, so a Monte Carlo based background estimation method would require a very high number of Monte Carlo events in order for sufficient events to pass the signal cuts. In addition, the theoretical cross section uncertainty for QCD multijet production is large, and sensitive to details of jet composition, fragmenation functions and non-Gaussian jet energy resolutions. For these reasons a data-driven technique is preferred.

In the muon channel a heavy-flavour hadron decay may result in a non-prompt muon which passes the prompt muon ID requirements given in Chapter 3. The cross section of heavy-flavour hadron production is not as large as that of inclusive multijets which are the fake candidates in the electron channel. Nonetheless a data-driven estimate is still desirable.

For both channels the general technique is the same. A QCD enhanced region — the 'control region' — with low signal contamination is selected by reversing or relaxing some of the prompt lepton identification requirements. This region is assumed to model the *shape* of the QCD multi-jet background in the signal region, and this assumption is checked using Monte Carlo. The *rate* of QCD multi-jet production is estimated by performing a fit of the data in a kinematic variable which shows good discrimination between QCD multi-jet production and other events. This fit is performed separately for each jet bin and each of the VBF regions described in Chapter 4.

5.2 Electron channel QCD control region

The primary triggers detailed in Section 3.2.1 require `medium` electron ID cuts. To select a QCD rich region we need a more relaxed identification requirement so the QCD control region is created from data passing the `loose` triggers, shown in Table 5.1. These triggers were highly prescaled and so only 113.9 pb^{-1} of integrated luminosity are available compared to 4.64 fb^{-1} available from the main trigger.

Data period(s)	Tiggers accepted
L, M	EF_e22vh_loose1, EF_e22vh_looseTrk
K	EF_e22_looseTrk, EF_e22_loose1, EF_e22_loose
J2	EF_e22_looseTrk, EF_e22_loose1, EF_e22_loose, EF_e20_looseTrk, EF_e20_loose1, EF_e20_loose
G, H, I, J1	EF_e20_looseTrk, EF_e20_loose1, EF_e20_loose
D, E, F	EF_e20_loose

Table 5.1: Triggers used for the QCD control region in the electron channel.

Using a `loose` trigger allows the electron ID requirements to be relaxed. The candidate electron was required to pass the offline `loose` ID and a subset of the `medium++` ID cuts corresponding to the track requirements, while failing a subset of the cuts for the `tight++` ID. This ensures that the QCD control selection is orthogonal to the signal selection.

Electron candidates were required to pass the following `Egamma` ID cuts from the `medium++` ID[2]:

```
egammaPID::TrackPixel_Electron
egammaPID::TrackSi_Electron
egammaPID::TrackMatchEta_Electron
```

At least one of the `tight++` identification cuts is required to have failed, with the exception of the following tracking cuts which are ignored.

```
egammaPID::ConversionMatch_Electron
egammaPID::TrackMatchPhi_Electron
egammaPID::TrackMatchEtaTight_Electron
```

This modified ID was chosen as it displayed the lowest bias in Monte Carlo studies of kinematic distributions for QCD multi-jet candidates in the signal region (i.e. passing `tight++`) and in the control region. This was checked using Monte Carlo by other members of the collaboration. This selection is referred to as the '`loose` electron

[2]These are the variable names for cuts which are listed in Section 3.2.3.

Process	Dataset ID	Generator	Cross section (pb)	Effective luminosity (fb^{-1})	Note
bbmu15X	108405	PYTHIA	73.9	0.067	$\epsilon_F = 0.9$
ccmu15X	106059	PYTHIA	28.4	0.059	$\epsilon_F = 0.9$
Min-bias	105802	PYTHIA	1.368×10^9	1.0×10^{-4}	$\epsilon_F = 6.72 \times 10^{-02}$

Table 5.2: Samples of simulated events used for QCD estimation cross-checks. The cross sections quoted are the ones used to normalise estimates of expected number of events. ϵ_F is the generator filter efficiency.

selection', in contrast to the '`tight++` electron selection', which is used in the signal region.

Prompt produced electrons are typically well isolated compared with hadronic jets, which typically consist of many tracks and many energy clusters in the calorimeter. Therefore isolation is expected to be useful in identifying genuine QCD background, as opposed to electrons which have merely been poorly reconstructed. Three isolation definitions were considered:

- $E_{\mathrm{T}}^{\mathrm{coneX}}/E_{\mathrm{T}} > 0.2$ - energy based isolation[3]
- $p_{\mathrm{T}}^{\mathrm{coneX}}/p_{\mathrm{T}} > 0.2$ - track based isolation
- Fails the signal isolation requirement: `Eiso(ptcone40)` and `Eiso(etcone20)`.

A choice between these variables was made by considering Monte Carlo closure tests, signal contamination and goodness of fit as described below.

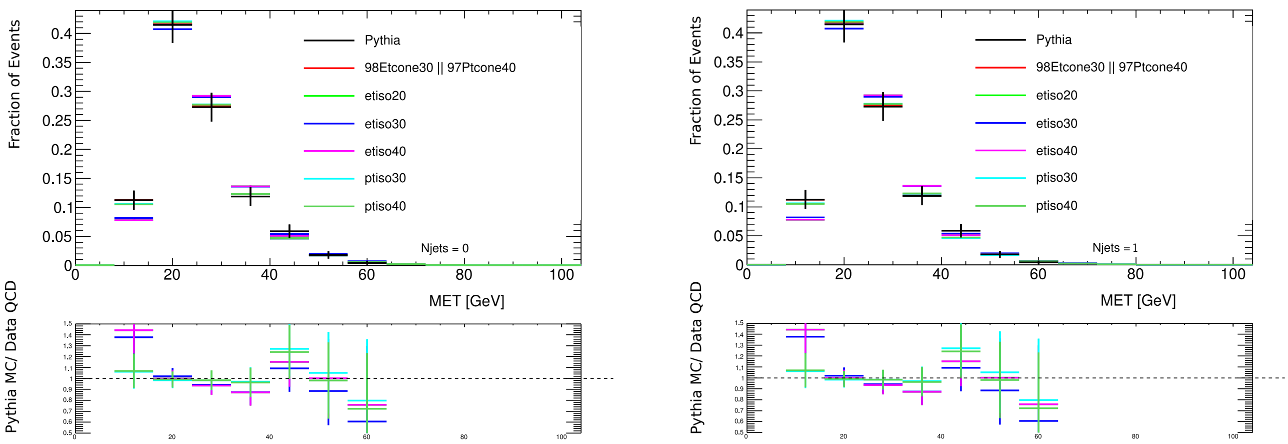
A first test in $E_{\mathrm{T}}^{\mathrm{miss}}$ is carried out using a fully-inclusive PYTHIA minimum bias Monte Carlo sample detailed in Table 5.2[4]. A filter is applied so that only events with at least one jet with $p_{\mathrm{T}} > 17$ GeV is present in each event and W and Z events are removed at generator level. The remaining events are expected to provide the best available Monte Carlo description of the QCD background.

[3] X is one of three possible cone sizes: 0.2,0.3 or 0.4.

[4] Generated including all Standard Model processes at appropriate rates.

Figure 5.1 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for events from this Monte Carlo sample which pass the `loose` electron selection and the $E_{\mathrm{T}}^{\mathrm{cone30}}$ isolation requirement labelled as 'Pythia'. This is compared to data passing the `loose` electron selection with various isolation requirements as indicated in the legend. The distributions agree within statisticical uncertainties and this confirms that the PYTHIA Monte Carlo does indeed describe the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution with non-isolated lepton candidates.

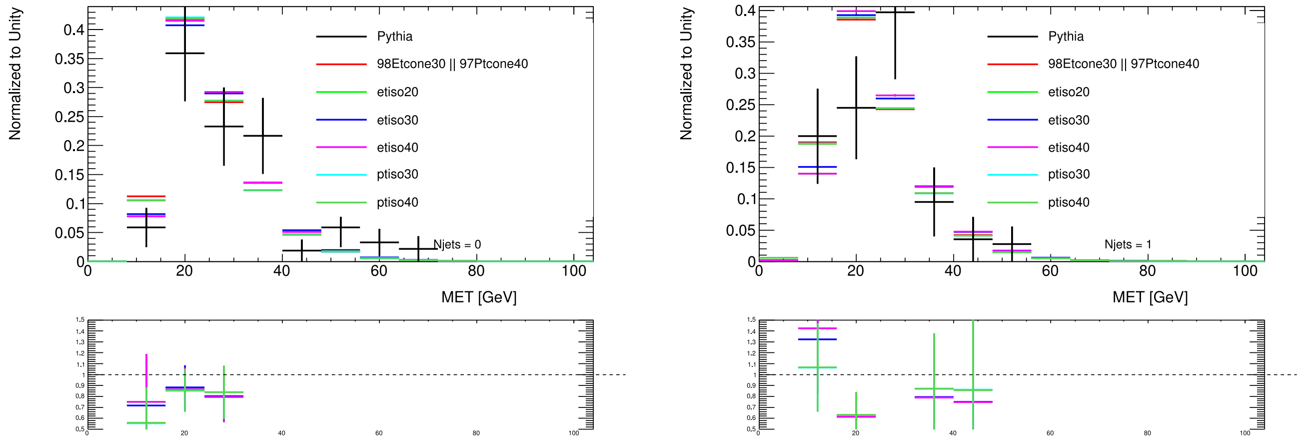
A closure test was made comparing the Monte Carlo events which passed the `tight++` selection to the data events in the QCD control region, shown in Figure 5.2. Closure in this test indicates that the PYTHIA Monte Carlo in the signal region has the same shape as the data in the QCD control region.



(a) $N_{jets} = 0$ (b) $N_{jets} = 1$

Figure 5.1: PYTHIA min-bias Monte Carlo passing the `loose` electron selection compared to data with the `loose` electron selection and various isolation requirements

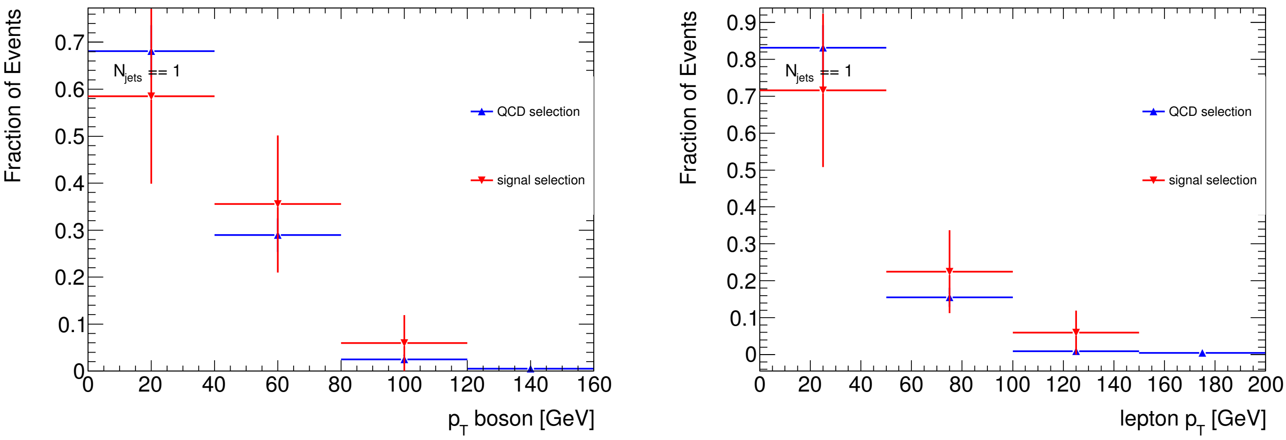
These tests taken together allow us to use the PYTHIA sample as a rudimentary test of the level of bias between the `loose` electron selection and the signal selection. Events from the PYTHIA sample which pass the signal selection are compared to events which pass the `loose` electron selection. Agreement is found within statistical errors, showing that there is not large bias between the QCD events which enter the signal region and those which enter the `loose` selection. Two examples of these tests are shown in Figure 5.3. For these plots only the reversed signal isolation is shown for clarity, as the closure tests in Figure 5.1 and 5.2 show that there is little variation



(a) $N_{jets} = 0$ (b) $N_{jets} = 1$

Figure 5.2: PYTHIA min-bias Monte Carlo passing the `tight++` electron selection compared to data passing the `loose` electron selection with various isolation requirements.

when using different isolation cone sizes.



(a) Boson p_{T} (b) Leading electron p_{T}

Figure 5.3: PYTHIA min-bias Monte Carlo passing the nominal QCD control region compared to that passing the signal selection.

To establish the level of signal contamination in the QCD selection the number of ALPGEN W+jets events which passed the QCD selection, normalized to the luminosity of the data sample passing the `loose` triggers, were compared to the number of data events passing the QCD selection. These results are shown in Table 5.3 for the inclusive W + jets region, and Table 5.4 for the VBF signal region. Contamination is largest in the zero jet bin, and uniformly below 1% in all other bins in which an isolation requirement is made. This is much less than the systematic uncertainty on the QCD estimate and so is not given further consideration.

Selection	N_{jets} =0	N_{jets} =1	N_{jets} =2	N_{jets} =3	N_{jets} =4	N_{jets} =5
$E_{\mathrm{T}}^{\mathrm{cone20}}$	1.9	0.2	0.3	0.3	0.3	0.3
$E_{\mathrm{T}}^{\mathrm{cone30}}$	3.8	0.4	0.5	0.5	0.5	0.5
$E_{\mathrm{T}}^{\mathrm{cone40}}$	1.9	0.2	0.3	0.3	0.3	0.3
$p_{\mathrm{T}}^{\mathrm{cone40}}$	7.1	0.9	1.0	1.0	0.9	1.0
no iso	30.0	4.2	4.4	3.6	3.4	2.8
Reverse signal isolation	4.9	0.6	0.7	0.6	0.7	0.8

Table 5.3: Percentage $W \to e\nu$+jets contamination for various control region isolation definitions, separated by jet bin.

Selection	N_{jets} =2	N_{jets} =3	N_{jets} =4	N_{jets} =5
$E_{\mathrm{T}}^{\mathrm{cone20}}$	0.7	1.5	0.6	0.7
$E_{\mathrm{T}}^{\mathrm{cone30}}$	1.0	1.0	0.9	0.7
$E_{\mathrm{T}}^{\mathrm{cone40}}$	0.7	1.5	0.6	0.7
$p_{\mathrm{T}}^{\mathrm{cone40}}$	2.4	1.6	1.9	1.3
no iso	7.5	5.2	3.7	2.6
Reverse signal isolation	1.6	1.7	1.2	0.9

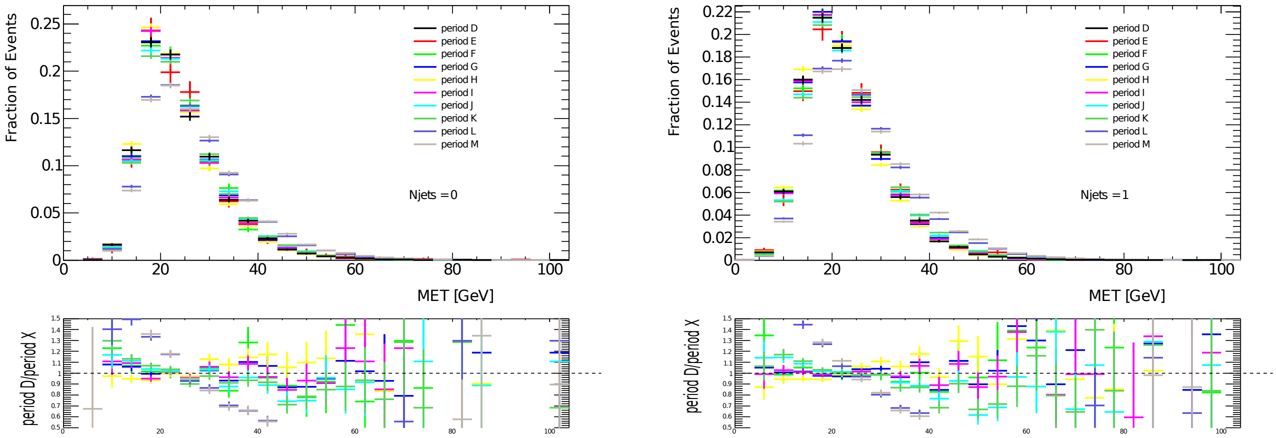
Table 5.4: Percentage $W \to e\nu$+jets contamination for various control region isolation definitions after the M_{jj}, high p_{T} and outside lepton veto cuts defined in (defined in Section 4.4).

The final choice of isolation requirements was taken by considering goodness of fit indicators described in Section 5.4. The central value was chosen to be $E_{\mathrm{T}}^{\mathrm{cone30}}$, with the others used to evaluate systematic uncertainties.

5.2.1 Period dependence

Since the prescales of the triggers changed over the course of the 2011 run, the relative contribution to the QCD control region from the various data periods is not proportional to the luminosity recorded by the primary trigger in the corresponding period. In addition, the pile up conditions changed dramatically across the 2011 run as shown in Figure 2.2. This is expected to influence the QCD background shape and rate due to additional jets from pile up interactions.

The time dependence of the QCD background was investigated and it is found



(a) $N_{jets} = 0$ (b) $N_{jets} = 1$

Figure 5.4: The $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for events passing the QCD control selection separated by period. Periods L and M are clearly separated from the others.

that the data falls into two distinct regimes. One corresponds to the early 2011 data (periods D-K) and one to the latter periods (L and M), as shown in Figure 5.4. The variations in the pile up conditions within these two sub-periods are much smaller than the variations between them. For the inclusive W+jets and W/Z ratio analyses [39,42] the QCD control region was created separately for these two periods. In the VBF region splitting the data led to statistically limited fits and so a single QCD control region was created covering all data periods.

5.3 Muon channel QCD control region

In the muon channel three identification variables are particularly useful in controlling the level of QCD background:

- $d_0/\sigma(d_0)$, the transverse impact-parameter significance. This is effective since heavy-flavour backgrounds are likely to have a displaced vertex and so the muon track will not extrapolate back to the primary vertex. In the signal region we make a cut of $d_0/\sigma(d_0) < 3$.

- z_0, the longitudinal impact-parameter, will also select heavy-flavour decays. In the signal region we make a cut of $z_0 < 10$ mm.

- $p_T^{cone20}/p_{\mathrm{T}}$, a relative isolation requirement, since heavy-flavour decays are less likely to result in isolated muons compared to prompt muons. In the signal region we make a cut of $p_T^{cone20}/p_{\mathrm{T}} < 0.1$.

Several control regions based on relaxing or reversing these cuts were investigated as summarized in Table 5.5.

Selection name	$d_0/\sigma(d_0) < 3$	$p_T^{cone20}/p_{\mathrm{T}} < 0.1$	$z_0 < 10$ mm
anti- d_0, iso, z_0	fail	pass	pass
anti- d_0, z_0	fail	—	pass
anti-iso, z_0	—	fail	pass
anti-iso	—	fail	—
anti-iso, d_0, z_0	pass	fail	pass
restricted anti-iso	—	fail restricted	—

Table 5.5: Summary of muon channel QCD control selections. The restricted anti-isolation region requires $0.1 < p_T^{cone20}/p_{\mathrm{T}} < 0.5$.

Selection	N_{jets} =0	$N_{jets} = 1$	N_{jets} =2	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} = 5$
anti- d_0, iso, z_0	42.8	5.4	4.3	3.3	3.1	1.9
anti- d_0, z_0	21.1	0.9	0.5	0.4	0.3	0.2
anti-iso, z_0	11.5	0.7	0.5	0.4	0.4	0.3
anti-iso	11.3	0.7	0.5	0.4	0.4	0.3
anti-iso, d_0, z_0	16.4	1.2	1.0	0.8	0.7	0.7
restricted anti-sio	13.4	1.0	1.0	0.8	0.8	0.7

Table 5.6: Percentage $W +$ jets contamination for various definitions of the control region in the inclusive $W \to \mu\nu +$ jets selection.

Selection	N_{jets} =2	$N_{jets} = 3$	$N_{jets} = 4$	$N_{jets} = 5$
anti- d_0, iso, z_0	6.6	16.0	4.2	8.2
anti- d_0, z_0	1.1	1.2	0.5	0.8
anti-iso, z_0	2.3	0.8	0.5	0.6
anti-iso	2.3	0.8	0.5	0.6
anti-iso , d_0, z_0	4.5	1.6	1.0	1.1
restricted anti-iso	4.2	1.4	0.9	1.4

Table 5.7: Percentage $W \to \mu\nu +$ jets contamination in the QCD control region for various definitions of the control region after the M_{jj}, high p_{T} and outside lepton veto VBF cuts (defined in Section 4.4).

Electron	Muon
loose trigger	$d_0/\sigma(d_0) > 3$
`loose` not `tight++` ID	
$E_{\mathrm{T}}^{\mathrm{cone30}}/E_{\mathrm{T}} > 0.2$	

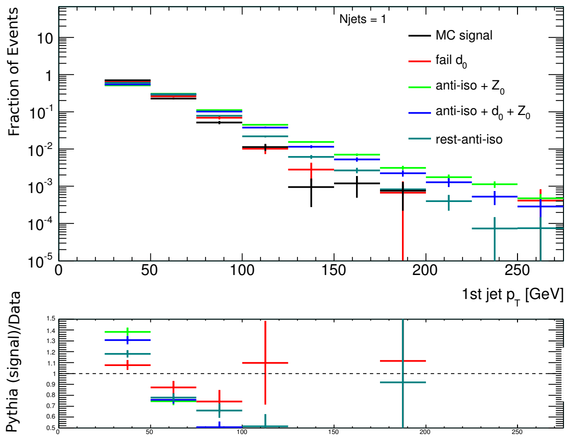
Table 5.8: Differences between signal and QCD control region selections for electron and muon channels.

The contamination of $W \to \mu\nu$ in the QCD control regions is shown in Table 5.6 and 5.7. The d_0 based control region gives higher contamination than the anti-isolation based regions, with the greatest contamination in the zero-jet bin, and $\sim 4\%$ in the two-bin region which is of most importance for this thesis.

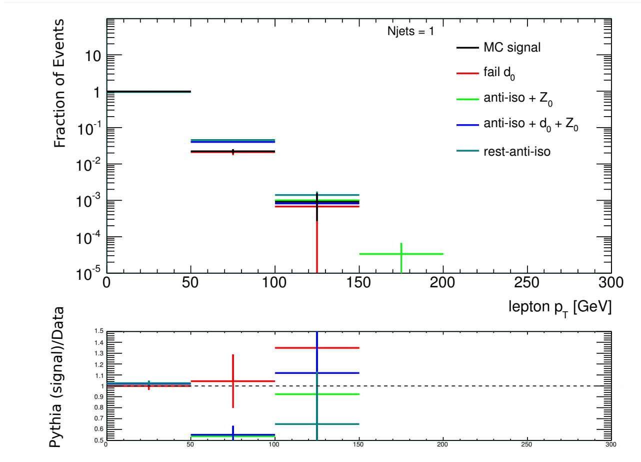
The bias in shape between the QCD control region and the inclusive signal region was tested using heavy flavour $b\bar{b}$ and $c\bar{c}$ Monte Carlo detailed in Table 4.2[5]. These samples contained b and c quarks decaying to at least one muon with $p_{\mathrm{T}} > 15$ GeV. This was passed through both the inclusive W and QCD selections and the distributions in kinematic variables of interest were compared. The leading jet p_{T} distributions are shown for four possible control region selections in Figure 5.5. All selections except the d_0 based show a very large bias to harder jet p_{T}. This may be explained as heavy-flavour decays with $p_T^{cone20}/p_{\mathrm{T}} > 0.1$ tend to be formed from higher p_{T} jets than candidates which pass the isolation cut. This means the balancing jet will have higher p_{T}. The d_0 based control region is therefore chosen for the central value of muon QCD and the other definitions are used as systematic variations.

A summary of the electron and muon QCD control selections is given in Table 5.8. For the electron channel the QCD multi-jet background is estimated from a control region consisting of events passing the `Loose` trigger, `loose` not `tight++` identification and failing the $E_{\mathrm{T}}^{\mathrm{cone30}}$ isolation requirement. The selection is otherwise identical to the main electron selection. In the muon channel the QCD multi-jet background is estimated from events which fail the d_0 significance cut. The selection is otherwise identical to the main muon selection.

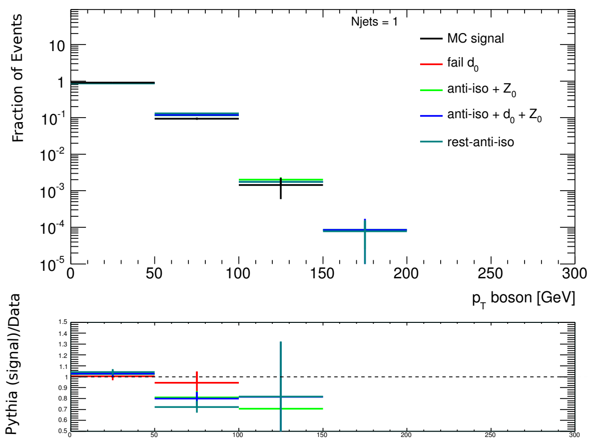
[5]Comparison with the VBF signal region was not possible due to limited Monte Carlo statistics



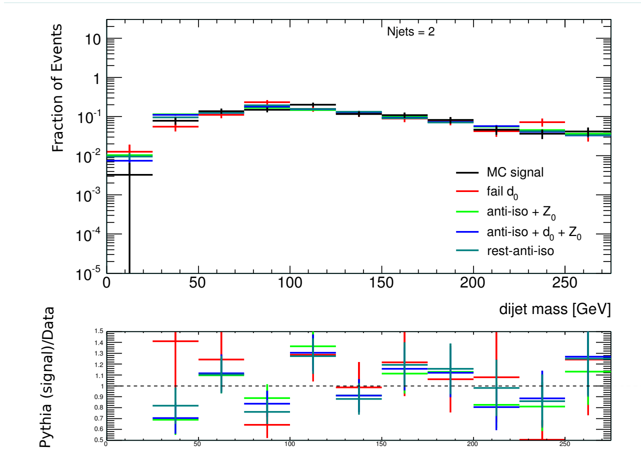
(a) Leading jet p_{T}



(b) Leading lepton p_{T}



(c) Boson p_{T}



(d) Dijet mass

Figure 5.5: Several distributions for four different QCD template selections, compared to the signal selection. The black data points show the QCD Monte Carlo passed through the signal selection. The jet p_{T} shows significant bias.

5.4 Normalization extraction

To normalize the control region a *template* variable which shows good discrimination between QCD and signal and background is chosen. The normalization is obtained using a fit performed with the `TFractionFitter` [142] class from `ROOT` [83]. This takes two input histograms — the templates — and varies their relative fractions to best match the shape of the data. In our case one template is provided by the QCD control selection described above while the other is provided by the sum of the signal and all backgrounds — known from now on as the *electroweak template* — as predicted by Monte Carlo simulation. Such a fit relies on a shape difference between the two components.

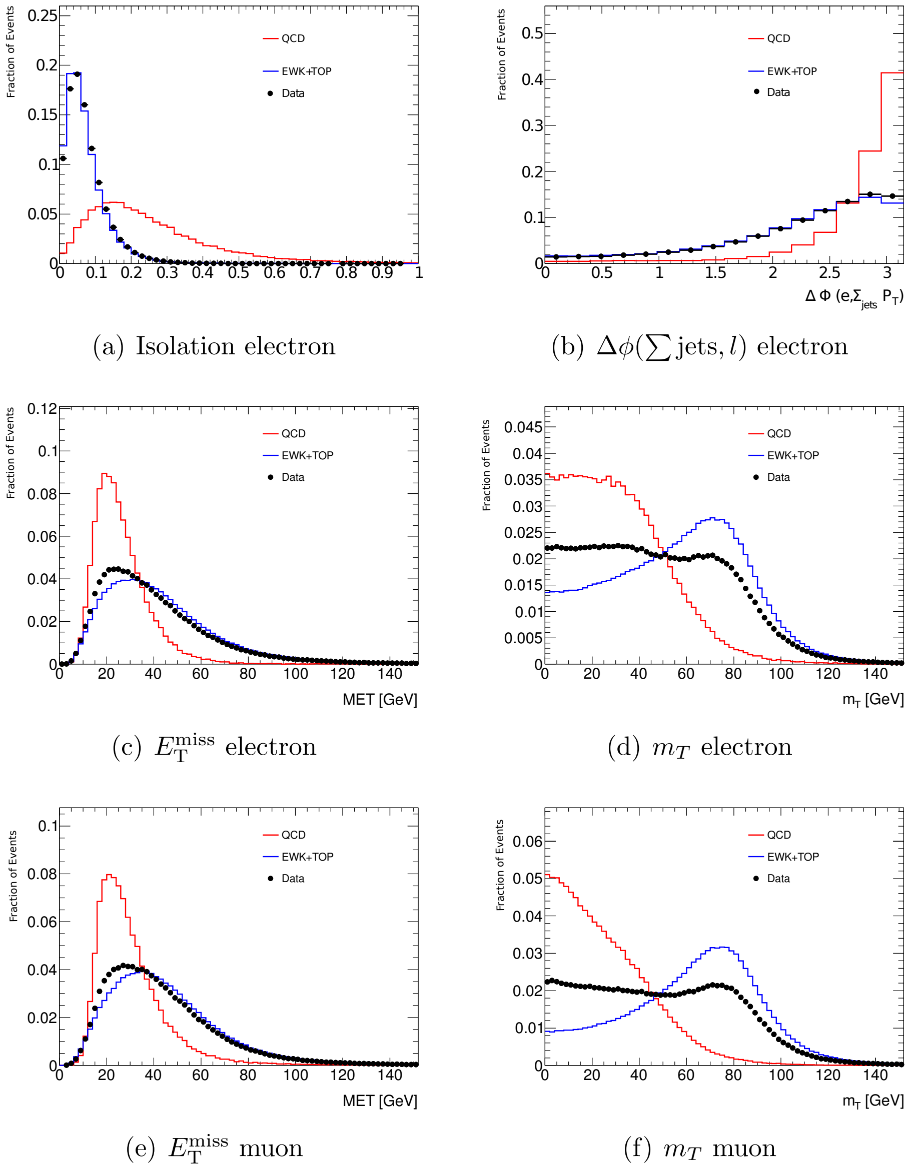
The `TFractionFitter` method accounts for the Poisson fluctuations of both data and Monte Carlo and it should therefore be noted that the plots have bin-by-bin corrections applied in addition to the overall scale factor. The implementation of the method described in Reference [142] by `TFractionFitter` does not account for weighted events. Since the electroweak template consists of the sum of all electroweak backgrounds which have been scaled by their cross section to the same effective luminosity, the error due to Monte Carlo statistical fluctuations can no longer be estimated by the square root of the bin content ($\sqrt{N_i}$). Instead an approximation of the correct treatment is made by scaling the electroweak template by N_{ewk}/σ^2, where N_{ewk} is the total content of the electroweak template, and σ is the statistical uncertainty on N_{ewk}. This ensures that the error in the scaled template follows the normal $\sigma' = \sqrt{N'}$ where N' is the scaled number of events. This scaling does not affect the overall scale factors derived and only affects the calculation of the χ^2 of the fit.

Four variables were considered as possible candidates for the template fit:

- Relative isolation cone around the lepton. QCD multi-jet background is expected to be skewed to high values compared to the electroweak template. This technique has the possibility to be fully data-driven as a signal template can be taken from $Z/\gamma^* +$jets events.

- $\Delta\phi(\sum \text{jets}, l)$— the azimuthal angle between the vector sum of the jets and the lepton. For QCD events this is expected to peak strongly at $\phi = \pi$ since the lepton candidate comes from a mismeasured jet which will balance the remaining jet system, whereas for $W +$jets events the peak is much broader, since the jets recoil against the W. This technique works well for N_{jets} =1 but is less effective for higher jet bins.

- $E_{\mathrm{T}}^{\mathrm{miss}}$ (without $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, after m_T cut) — QCD events peak at lower values of $E_{\mathrm{T}}^{\mathrm{miss}}$ since this quantity is due to mismeasurement rather than a real neutrino.



(a) Isolation electron (b) $\Delta\phi(\sum \text{jets}, l)$ electron

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

(e) $E_{\mathrm{T}}^{\mathrm{miss}}$ muon (f) m_T muon

Figure 5.6: Four candidates for use as template fit variable, each shown in the N_{jets} =2 bin.

- m_T (before m_T and $E_{\mathrm{T}}^{\mathrm{miss}}$ cut) — QCD events tend to have $E_{\mathrm{T}}^{\mathrm{miss}}$ aligned with the candidate lepton and so peak at $m_T = 0$.

In well behaved fits the normalization factor given to the electroweak template should be ~ 1 and the χ^2/degrees of freedom should be low. Deviations from this expectation suggest either biased templates and/or poor Monte Carlo modelling of the electroweak template. Only $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_T provided stable fits for all jet bins, while the statistical uncertainty on the m_T fit was substantially lower, particularly in the VBF region. In addition m_T provided more stable behaviour when jet and

Percentage QCD

	Njet = 0	Njet = 1	Njet = 2	Njet = 3	Njet = 4	Njet = 5
Electron						
Inclusive	4.3 ± 0.5	15 ± 4.9	13 ± 2.0	11 ± 1.6	7.7 ± 1.3	6.0 ± 2.6
VBF selection	–	–	5.0 ± 1.8	6.6 ± 1.8	6.1 ± 1.8	5.6 ± 1.6
Muon						
Inclusive	2.8 ± 1.2	12 ± 1	10 ± 2	9.0 ± 1.6	6.8 ± 1.8	4.3 ± 1.5
VBF selection	–	–	6.6 ± 2.0	5.9 ± 2.3	5.9 ± 2.5	5.1 ± 1.2

Table 5.9: Fraction QCD predicted in the inclusive and VBF signal regions. The uncertainty is the total systematic and statistical uncertainty which is described in Section 5.5.

$E_{\mathrm{T}}^{\mathrm{miss}}$ systematics were varied. Example distributions for $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_T are shown in Figure 5.6.

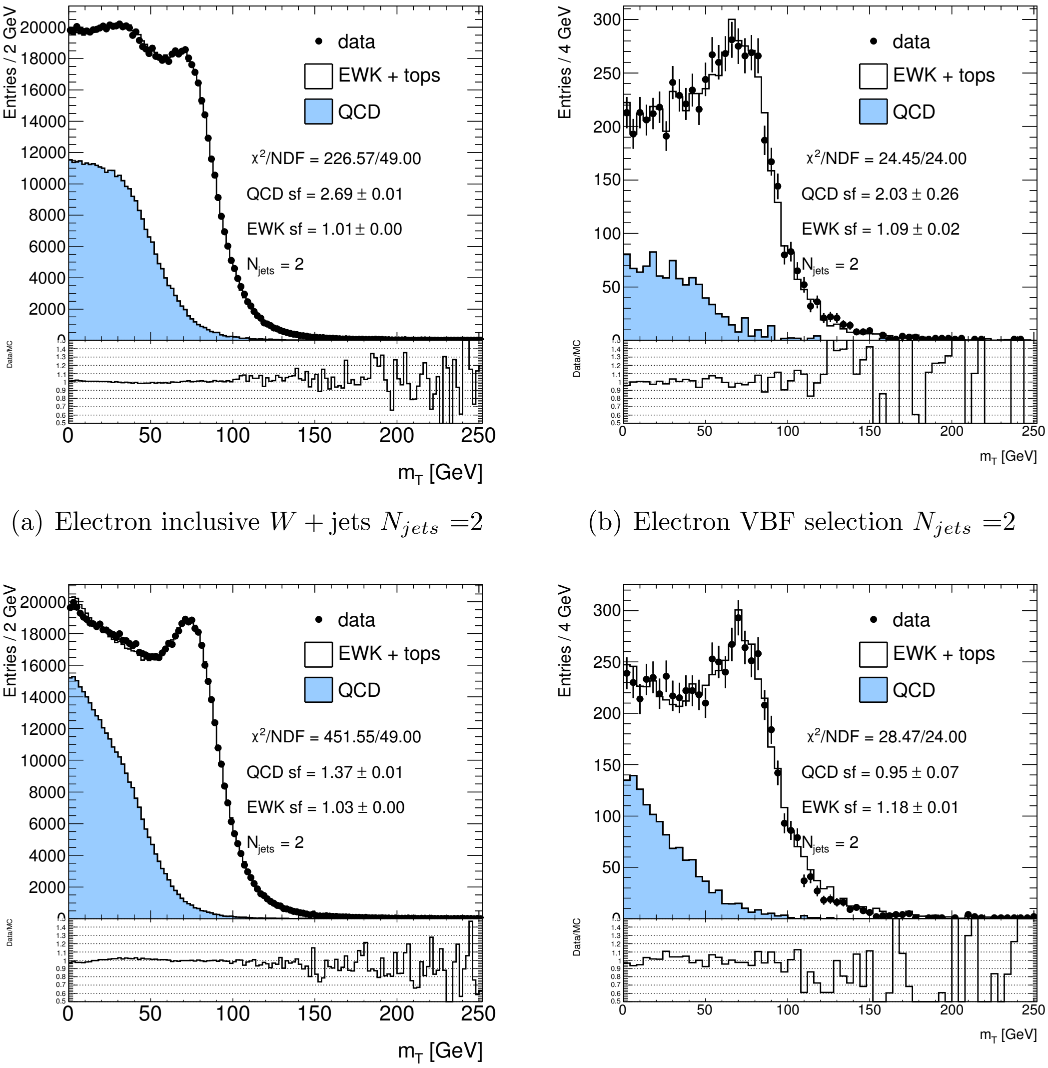
The fit is carried out in the m_T variable before the application of the cut $m_T > 40$ GeV to take advantage of the large difference in shape at low m_T values. The fit is carried out in exclusive jet bins and separately for each VBF selection and control region as described in Chapter 4. A fit range of $0 - 100$ GeV is used.

The fit provides a *scale factor* which is propagated to each kinematic distribution to normalize the QCD background contribution for each distribution. For $N_{jets} > 5$ there were insufficient statistics in the QCD templates to perform a stable fit and so the scale factor derived for $N_{jets} = 5$ was applied to all higher jet bins. The results of these fits are shown in Table 5.9 and the plots in Figure 5.7.

5.5 Systematic uncertainties

The systematic uncertainty on the contribution of QCD was estimated by varying the fit in a number of ways:

Vary fit range The nominal fit range of 0-100 GeV was varied using alternative minima of 5 or 15 GeV and alternative maxima of 80 or 120 GeV. All eight permutations



(a) Electron inclusive $W+$jets N_{jets} =2

(b) Electron VBF selection N_{jets} =2

(c) Muon inclusive $W+$jets N_{jets} =2

(d) Muon VBF selection N_{jets} =2

Figure 5.7: Examples of template fits for electron and muon channels in the inclusive $W+$jets and VBF phase spaces for the N_{jets} =2 jet bin.

of these up-down variations were fitted and then an envelope of the deviation from the nominal case was taken as the systematic uncertainty from this source. This leads to a $^{+3.6\%}_{-15\%}$ variation in QCD fraction in the VBF region for electrons, and $^{+32\%}_{-4.8\%}$ shift for the muon channel.

Change control region selection. For electrons the different isolation requirements shown in Table 5.3 were each used and fits repeated. This gave an uncertainty on QCD fraction of $^{+32\%}_{-11\%}$. For the muon channel the alternative template selections listed in Table 5.5 were used and the fits repeated. This gave an uncertainty on QCD fraction of $^{+1\%}_{-23\%}$.

Change fit variable The fit was performed using the $E_{\mathrm{T}}^{\mathrm{miss}}$ variable. This gave an uncertainty of 1.4% on the QCD fraction in the electron channel and 11% in the muon channel.

JES and $E_{\mathrm{T}}^{\mathrm{miss}}$ systematics The detector systematics due to jet energy scale and $E_{\mathrm{T}}^{\mathrm{miss}}$ were varied and the fit repeated. This gave an uncertainty of $^{11\%}_{-12\%}$ in the electron channel and $^{7.9\%}_{4.2\%}$ in the muon channel.

For each source an envelope of the largest systematic uncertainty was taken and these envelopes were added in quadrature and symmetrized separately for each signal region and jet bin. It should be noted that this systematic procedure only takes into account changes in the normalization of QCD and not any changes in shape. The results can be seen in Tables 5.10 and 5.11.

Fractional error on QCD contribution (%)		
Error source	Njet=2	Njet=3
Statistical error	13.0	14.0
Control region selection up	32.0	18.0
Control region selection down	11.0	13.0
JES and $E_{\mathrm{T}}^{\mathrm{miss}}$ up	11.0	11.0
JES and $E_{\mathrm{T}}^{\mathrm{miss}}$ down	12.0	11.0
m_T range up	3.6	15.0
m_T range down	25.0	8.2
$E_{\mathrm{T}}^{\mathrm{miss}}$ up	0.0	0.0
$E_{\mathrm{T}}^{\mathrm{miss}}$ down	1.4	12.0
Total up	37.0	29.0
Total down	33.0	26.0

Table 5.10: The systematic uncertainties for m_T fit for QCD background in the electron channel. The 'Up' and 'Down' variations correspond to the maximal Up and Down deviation.

Fractional error on QCD contribution (%)		
Error source	Njet=2	Njet=3
Statistical error	7.3	8.2
Control region selection up	1.0	1.9
Control region selection down	23.0	18.0
JES and $E_{\mathrm{T}}^{\mathrm{miss}}$ up	7.9	8.4
JES and $E_{\mathrm{T}}^{\mathrm{miss}}$ down	4.2	6.4
m_T range up	32.0	9.4
m_T range down	4.8	14.0
$E_{\mathrm{T}}^{\mathrm{miss}}$ up	11.0	0.0
$E_{\mathrm{T}}^{\mathrm{miss}}$ down	0.0	16.0
Total up	36.0	15.0
Total down	25.0	30.0

Table 5.11: The systematic uncertainties for m_T fit for QCD background in the muon channel. The 'Up' and 'Down' variations correspond to the maximal Up and Down deviation.

CHAPTER 6

Signal extraction

This chapter explains the method used to extract the number of $W+$ EWK jets events and the cross section for this process. The systematic uncertainties are assessed and several cross-checks are performed.

6.1 Fit set-up

The M_{jj} spectrum gives the best separation between background and signal, so this distribution is used to extract the observed number of signal events. Figure 6.1 shows the M_{jj} distribution with signal and background normalized to show the fraction of events in each bin. (Figures 4.9(a) and 4.9(b) showed the distributions normalized to expected cross section.) The signal extraction is performed with a maximum-log-likelihood fit[1].

The probability of observing n events, with ν expected, is given by the Possion

[1] The method of maximum-likelihood is presented in many statistics textbooks, e.g. [143].

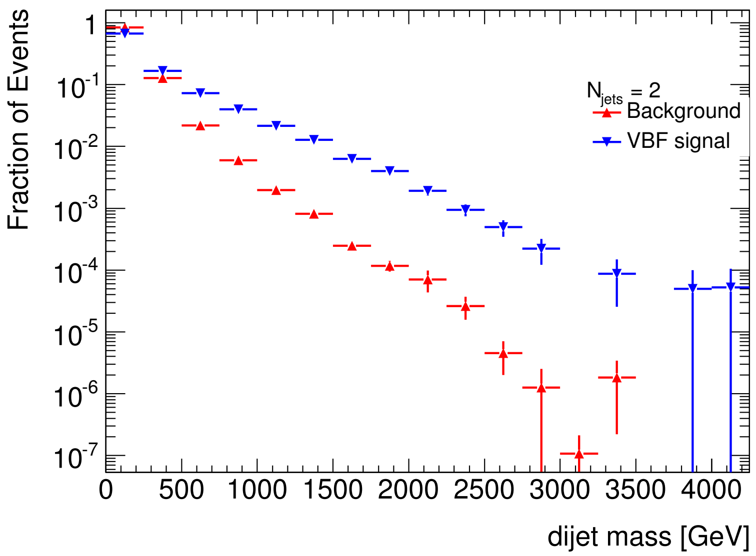


Figure 6.1: The M_{jj} distribution for signal and background normalized to one for both distributions.

distribution:

$$f(n, \nu) = \frac{\nu^n e^{-\nu}}{n!}. \tag{6.1}$$

The number of expected events for a given bin i (N_i) is divided into three sources: signal ($N_{\mathrm{S}i}$), $W+$ QCD jets background (N_{Wi}) and all other background contributions ($N_{\mathrm{B}i}$).

$$\begin{aligned} N_i &= N_{\mathrm{S}i} + N_{Wi} + N_{\mathrm{B}i}, \\ N &= \sum_{i=1}^{\mathrm{bins}} N_i. \end{aligned} \tag{6.2}$$

The sum of the bin contents, N_{S}, N_W and N_{B}, are the parameters that are extracted from the fit. The more consistent with the data our model for N is, the higher the probability as returned by the Poisson probability given in Equation (6.1). The data is divided into bins and so the 'likelihood', $\mathscr{L}$, is given by the product of the Poisson probability for each bin, with $\mathscr{D}_i$ observed in bin i and N_i expected:

$$\mathscr{L} = \prod_{i=0}^{\mathrm{bins}} \frac{N_i^{\mathscr{D}_i} e^{-N_i}}{\mathscr{D}_i!}. \tag{6.3}$$

Note that though the functional form of the likelihood is taken from the Poisson probability, it is not interpreted as the probability that the model represents that data. Rather it is the probability that this data could arise, given that the model is true. In the current work we do not use this interpretation explicitly and only use the likelihood as a test statistic to asses the best-fit values for the parameters extracted.

To turn the product into a sum we take the logarithm:

$$\ln \mathscr{L} = \sum_{i=0}^{\text{bins}} (\mathscr{D}_i \ln N_i - N_i - \ln(\mathscr{D}_i!)). \tag{6.4}$$

To find the best fit of the parameters N_{S}, N_W and N_{B} they are varied to maximize $\ln \mathscr{L}$ or equivalently minimize $-\ln \mathscr{L}$. For convienience the results of the fit are sometimes reported in terms of the "signal strength" $\mu_s = \frac{N_{\text{signal observed}}}{N_{\text{signal expected}}}$, and analogously for $W+$ QCD jets (μ_W) and other backgrounds (μ_{b2}).

The parameter N_{B} is constrained with a Gaussian whose width is determined by the cross section uncertainty on these backgrounds as described in Section 4.2.2. The uncertainty on the QCD background is described in Chapter 5. The constraint on N_{B} introduces an additional penalty term in the log-likelihood function:

$$\frac{(N_{\text{B}} - N'_{\text{B}})^2}{2\sigma_b^2}, \tag{6.5}$$

where N'_{B} is the expected value of N_{B} and σ_b is the cross section uncertainty. Thus if $N_{\text{B}} = N'_{\text{B}}$ no additional penalty is introduced, otherwise the size of the penalty is smaller for larger values of σ_b. In the minimization, terms with no dependence on N_{S}, N_W or N_{B} can be neglected as their derivative is zero, thus total negative likelihood to be minimized is:

$$-\ln \mathscr{L} = \sum_{i=0}^{\text{bins}} \Big(N_i - \mathscr{D}_i \ln N_i - \frac{(N_{\text{B}i} - N'_{\text{B}i})^2}{2\sigma_{ib}^2} \Big). \tag{6.6}$$

The likelihood is generated from "template" histograms using `RooFit` [144] and the minimization carried out using `Minuit` [145]. Examples of the likelihood are shown in

Figure 6.2. The statistical uncertainty on the fit is calcuated with the `MINOS` routine from `Minuit`. This calculates the profile likelihood ratio λ:

$$\lambda = \frac{\mathscr{L}(N_{\rm S}, \hat{\hat{N}}_W, \hat{\hat{N}}_{\rm B})}{\mathscr{L}_{max}}, \tag{6.7}$$

where $\hat{\hat{N}}_W$ and $\hat{\hat{N}}_{\rm B}$ are the values of these parameters which maximize the likelihood for the particular value of $N_{\rm S}$, and the denominator is the likelihood evaluated at at its maximum, providing a normalization. For Possion distributed data the 1σ uncertainties are found where

$$\ln(\lambda) < \tfrac{1}{2} \tag{6.8}$$

These are quoted as the 'fit stat' uncertainties. Further details of this method can be found in Reference [4]. The $\ln\lambda$ distribution for $N_{\rm S}$ from the combined fit is shown in Figure 6.3(a). The 1σ and 2σ contours in the $N_{\rm S}$-N_W plane are shown in Figure 6.3(b).

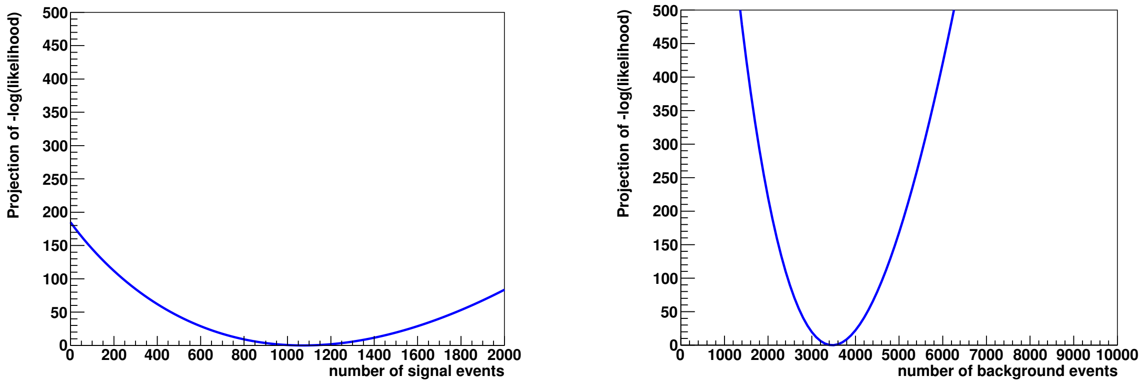
The accuracy of the fit in this result is limited by the available Monte Carlo statistics. In general more finely binned distributions should give lower 'fit stat' uncertainties, as they provide better discrimination between models and so steeper likelihood curves. If bins are made too narrow they are subject to large fluctuations in the Monte Carlo contents and so no longer sensitive to the underlying model. The binning of the M_{jj} histograms was therefore chosen such that Monte Carlo statistical errors and the fit statistical errors were equal.

The χ^2 displayed on fits and in results tables is calculated as

$$\chi^2 = \sum_{i=0}^{\rm bins} \frac{(\mathscr{D}_i - N_i)^2}{\sigma_i^2} \tag{6.9}$$

where $\mathscr{D}$ is the data, N is the fit result and σ is the statistical error from data.

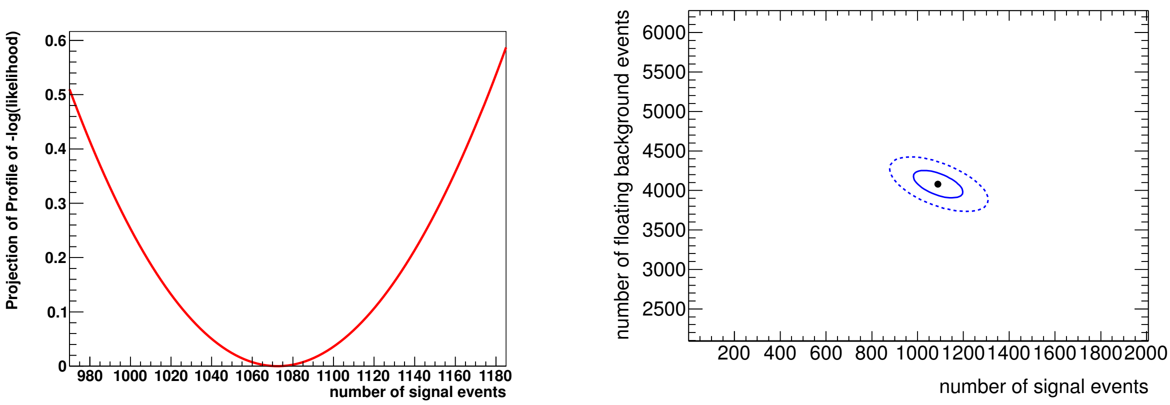
'Combined' results are obtained from the combination of the electron and muon channel by taking the sum of the input histograms from each channel. In general keeping the channels separate and doing a combined fit should lead to more accurate



(a) Likelihood in the N_S parameter

(b) Likelihood in the N_W parameter

Figure 6.2: Example $-\ln\mathscr{L}$ distributions taken from a combined electron and muon channel fit projected onto the N_S and N_W axes.



(a) $\ln(\lambda)$ in the N_S parameter

(b) Contour plot of errors in N_S and N_W

Figure 6.3: Example of the $\ln(\lambda)$ ratio distribution taken from a combined electron and muon channel fit for N_S and the contour plot showing 1σ and 2σ bands calculated as described in the text.

results, as information on individual channel background shapes and systematic uncertainties is lost by combining bins. In the current case the background shapes are similar and the dominant errors (JES) are fully correlated between the two channels. In addition we are Monte Carlo statistically limited in both, so a simple combination is preferable. All the JES and $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties are assumed to be fully correlated between the two channels and so are added linearly. The lepton uncertainties are much smaller than the Monte Carlo statistical uncertainties and so these are taken to be negligible and not included in the combination.

6.1.1 Pseudo-experiment procedure

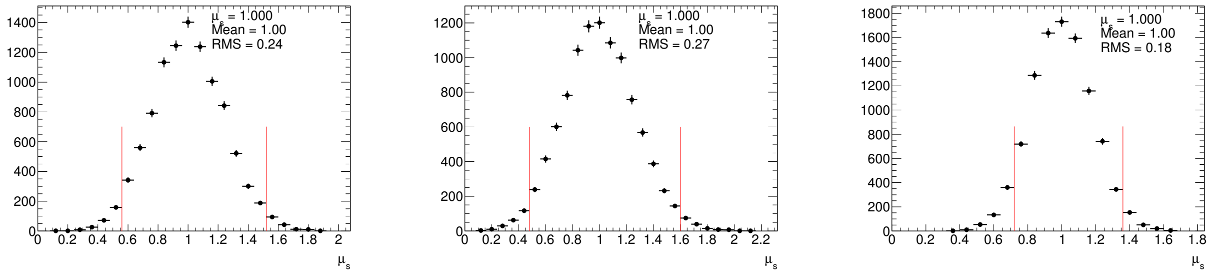
To understand the expected range of variation in μ_s, pseudo (or toy) experiments are performed. A large number of toy datasets are generated by varying some base distribution within an uncertainty and each dataset is then fitted in the same manner as the experimental data. This technique is used a number of times so a general overview is set out below.

A base histogram, known as the 'underlying distribution', is selected from which the toy datasets will be drawn. A number of toy datasets are generated, with systematic errors treated as correlated bin-to-bin and statistical errors treated as uncorrelated bin-to-bin. A single Gaussian random number p_i is drawn per systematic i per toy; and for Monte Carlo statistical uncertainties a different Gaussian random number $p_{i,j}$, is drawn for each bin j, of each toy. For bin j the content from the underlying distribution n_j^{nom} is varied to n_j^{var} according to a Poisson distribution:

$$n_j^{var} = \mathrm{Poisson}\Big(n_j^{nom} + \sum_{i=0}^{\mathrm{systematics}} p_{i(,j)} \cdot \delta_i\Big), \tag{6.10}$$

with: i = index over systematics, $\delta_i = 1\sigma$ uncertainty for systematic i and $p_{i(,j)}$ the random number for systematic i (and bin j for uncorrelated errors). Poisson(x) returns a random number from a Poisson distribution with mean x.

This technique was used to validate the fitting procedure. 10^4 toy datasets were drawn using the nominal template histograms as the underlying distribution and varied statistical uncertainty. The observed rate of fitted μ_s can be seen in Figure 6.4. The fact that each is centered around one shows that the most basic closure test is passed by our fitting procedure, i.e. it returns a signal strength distribution centered on 1 when the nominal signal is injected.



(a) Electron channel (b) Muon channel (c) Combined channel

Figure 6.4: μ_s distributions from 10^4 toy experiments for electron, muon and combined cases drawn from variations due to Poisson statistics. The red lines enclose 95% of toy experiments.

6.1.2 Inclusion of systematic uncertainties

The impact of systematic uncertainties was assessed with a method of toy experiments. For each systematic uncertainty 10^3 toy datasets were created using the signal and background template histograms, obtained from the systematic variation of the nominal selection, as the underlying distribution. This was varied within statistical uncertainties only. The mean fitted value of N_{S} from the set of toy experiments is taken as the variation in N_{S} due to that systematic uncertainty.

The JER, $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution and theoretical errors were symmetrized by taking the average absolute deviation from the nominal value. For the PDF uncertainties each eigenvector was fitted separately and the variation calculated using Equation (4.1) given in Section 4.6.1.

The results for the electron and muon channels are shown in Figure 6.5 and in Tables 6.1, 6.2, 6.3 and 6.4. The dominant uncertainties are the Monte Carlo statistical uncertainty and the jet energy scale uncertainty coming from the η-intercalibration term. The lepton uncertainties are ($\mathcal{O}$ 2%) and so we are justified in neglecting them in comparison to the JES and Monte Carlo statistical errors of ($\mathcal{O}$ 25%) in the combination of lepton channels.

The dominant systematic uncertainties display correlated positive and negative shifts between N_{S} and N_W. Some of the sub-dominant uncertainties show uncorrelated shifts between N_{S} and N_W. This is due to the fact that the fit is only sensitive to the *shape* of the templates, not their overall normalization. Thus a systematic variation that results in toy datasets with less data, may lead to more signal or more background depending on how signal-like or background-like the shape.

Systematic	% ΔN_s	% ΔN_b
Jet matching	40, −40	3, −3
MPI signal	3, −3	0, −0
MSTW2008	7, −7	3, −3
Scale	11, −11	13, −13
Shower signal	6, −6	1, −1
Total Theory Error	43, −43	14, −14

Table 6.1: Systematic uncertainties on the fraction yields of signal and background after VBF signal fit due to theory errors, derived in the muon channel as described in Section 4.6.1. These are calculated from pseudoexperiments as described in the text.

Systematic	ΔN_s	% ΔN_s	ΔN_b	% ΔN_b
MC stats	$80, -75$	$24, -22$	$23, -25$	$1, -1$
Fit stat	$74, -74$	$22, -22$	$101, -101$	$5, -5$
JER	$30, -30$	$9, -9$	$41, -41$	$2, -2$
JES eta	$80, -53$	$24, -16$	$46, -57$	$2, -3$
JES closeby	$-7, 3$	$-2, 1$	$-4, 2$	$-0, 0$
JES flavour composition	$-0, 3$	$-0, 1$	$2, -1$	$0, -0$
JES flavour response	$-7, 18$	$-2, 5$	$35, -39$	$2, -2$
JES mu	$-4, -0$	$-1, -0$	$14, -18$	$1, -1$
JES NPV	$-1, 3$	$-0, 1$	$9, -6$	$0, -0$
JES other	$7, -3$	$2, -1$	$15, -30$	$1, -2$
Met resolution	$5, -5$	$2, -2$	$12, -12$	$1, -1$
Met Scale	$-4, 1$	$-1, 0$	$30, -11$	$1, -1$
Electron ID	$1, -4$	$0, -1$	$28, -17$	$1, -1$
Electron energy scale	$5, -6$	$1, -2$	$-40, 45$	$-2, 2$
Electron energy resolution	$-1, -3$	$-0, -1$	$4, 4$	$0, 0$
Electron reconstruction	$4, -2$	$1, -1$	$23, -13$	$1, -1$
Electron trigger	$2, -4$	$0, -1$	$18, -10$	$1, -1$
Total Experimental Error	$140, -123$	$42, -37$	$144, -145$	$7, -7$
Total Theory Error	$145, -145$	$43, -43$	$278, -278$	$14, -14$
Total Error	$202, -190$	$60, -56$	$313, -313$	$16, -16$

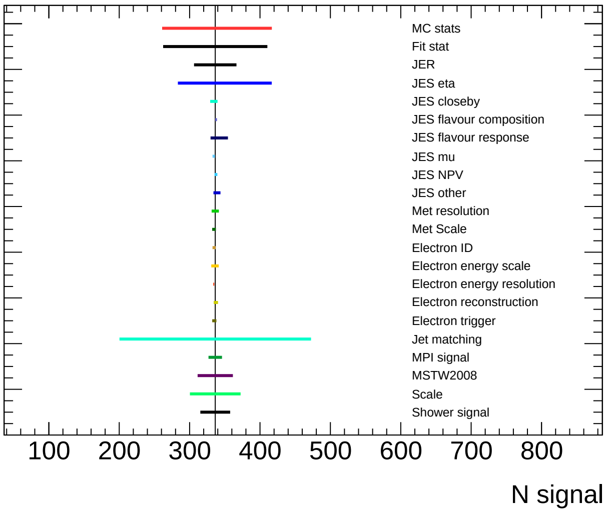
Table 6.2: Systematic uncertainties on the total and fraction yields of signal and background after VBF signal fit in the electron channel. The first number is for the 'up' variation and the second for the 'down' variation. These are calculated from pseudoexperiments as described in the text.

Systematic	ΔN_s	% ΔN_s	ΔN_b	% ΔN_b
MC stats	$85, -82$	$26, -25$	$14, -12$	$1, -1$
Fit stat	$78, -78$	$24, -24$	$112, -112$	$5, -5$
JER	$10, -10$	$3, -3$	$14, -14$	$1, -1$
JES eta	$82, -117$	$25, -35$	$45, -28$	$2, -1$
JES closeby	$-10, 2$	$-3, 1$	$17, -6$	$1, -0$
JES flavour composition	$-2, -5$	$-1, -1$	$9, 11$	$0, 1$
JES flavour response	$10, -16$	$3, -5$	$34, -22$	$2, -1$
JES mu	$6, -32$	$2, -10$	$3, 15$	$0, 1$
JES NPV	$1, -19$	$0, -6$	$8, 9$	$0, 0$
JES other	$22, -42$	$7, -13$	$27, -1$	$1, -0$
Met resolution	$13, -13$	$4, -4$	$16, -16$	$1, -1$
Met Scale	$-5, 2$	$-1, 1$	$10, -15$	$0, -1$
Muon efficiency	$1, 1$	$0, 0$	$7, -5$	$0, -0$
Muon ID resolution	$-1, 4$	$-0, 1$	$2, 1$	$0, 0$
Muon MS resolution	$4, 1$	$1, 0$	$-0, 2$	$-0, 0$
Muon trigger	$6, -5$	$2, -1$	$43, -40$	$2, -2$
Total Experimental Error	$145, -174$	$44, -53$	$141, -128$	$6, -6$
Total Theory Error	$142, -142$	$43, -43$	$306, -306$	$14, -14$
Total Error	$203, -225$	$61, -68$	$337, -331$	$15, -15$

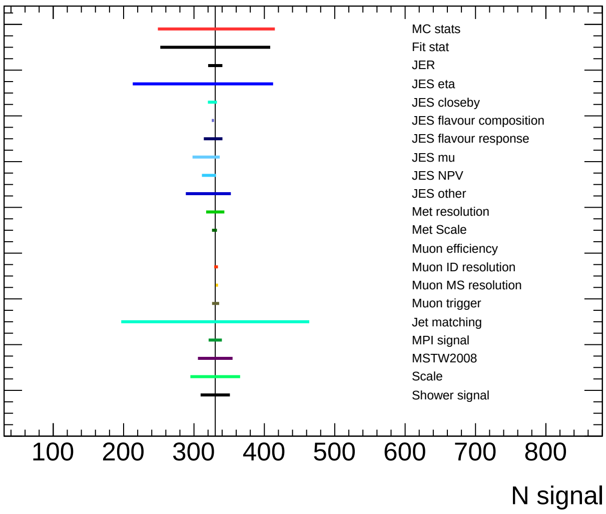
Table 6.3: Systematic uncertainties on the total and fraction yields of signal and background after VBF signal fit in the muon channel. The first number is for the ‘up’ variation and the second for the ‘down’ variation. These calculated from pseudoexperiments as described in the text.

Systematic	ΔN_s	% ΔN_s	ΔN_b	% ΔN_b
MC stats	$111, -116$	$17, -17$	$2, -13$	$0, -0$
Fit stat	$108, -108$	$16, -16$	$173, -173$	$4, -4$
JER	$27, -27$	$4, -4$	$4, -4$	$0, -0$
JES eta	$154, -171$	$23, -26$	$23, -7$	$1, -0$
JES closeby	$-5, -1$	$-1, -0$	$5, 2$	$0, 0$
JES flavour composition	$-1, -6$	$-0, -1$	$-3, 4$	$-0, 0$
JES flavour response	$9, -5$	$1, -1$	$5, -10$	$0, -0$
JES mu	$-2, -30$	$-0, -5$	$1, 5$	$0, 0$
JES NPV	$-4, -15$	$-1, -2$	$-14, 0$	$-0, 0$
JES other	$29, -32$	$4, -5$	$12, -17$	$0, -0$
Met resolution	$4, -4$	$1, -1$	$1, -1$	$0, -0$
Met Scale	$-14, 3$	$-2, 0$	$11, -12$	$0, -0$
Total Experimental Error	$222, -240$	$33, -36$	$175, -175$	$4, -4$
Total Theory Error	$286, -286$	$43, -43$	$549, -549$	$14, -14$
Total Error	$362, -373$	$54, -56$	$576, -576$	$15, -15$

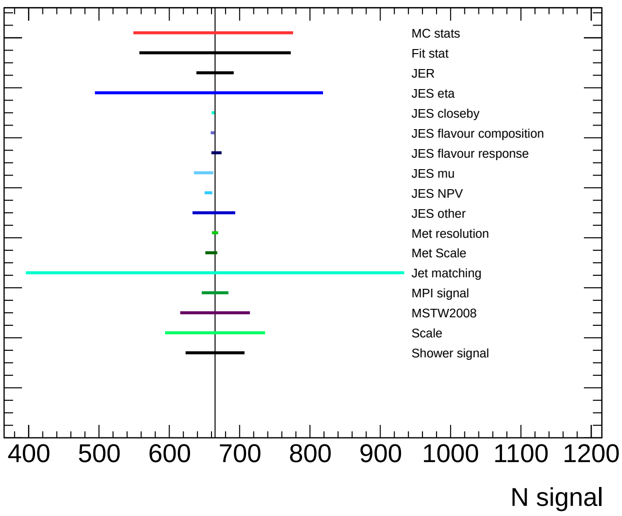
Table 6.4: Systematic uncertainties on the total and fraction yields of signal and background after VBF signal fit in the combined electron and muon channel. The first number is for the ‘up’ variation and the second for the ‘down’ variation. These calculated from pseudoexperiments as described in the text. The lepton uncertainties have been omited since they are much smaller than the statistical uncertainty.



(a) Electron



(b) Muon



(c) Combined

Figure 6.5: Variation of the expected distributions of number of signal events for each systematic variation for electron, muon and combined channels, evaluated with 10^3 toy-experiments per systematic uncertainty. Note that this is shown in terms of number of signal events, so while the combined channel has a larger uncertainty in number of signal events fitted, it has a smaller fractional error.

6.2 Results

6.2.1 Background-only exclusion

To establish that the $W+$ EWK jets process can be distinguished from a background only scenario, toy-experiments were carried out to test the consistency of the data with the background-only hypothesis. 10^7 of these were generated in the manner described in Section 6.1.1, using the background only Monte Carlo sample as the underlying distribution. The fraction of toy experiments where $\mu_s > 1$ is taken as the p-value for expected background exclusion, i.e. the probability that background-only could fluctuate up and give a larger signal than we expect. The fraction of toy experiments where $\mu_s > \hat{\mu}_s$, where $\hat{\mu}_s$ is the observed signal strength, is the observed background rejection, i.e. the probability that the background-only hypothesis could fluctuate up and give a larger signal than we observe. An example distribution obtained from toy experiments of the fitted μ_s distribution for the background-only hypothesis is shown in Figure 6.6. A summary of the expected background exclusions is given in Table 6.5. With all systematics included the background-only hypothesis is expected to be excluded at the 4.36σ level. The values for the observed case are given in the following subsection in Table 6.6.

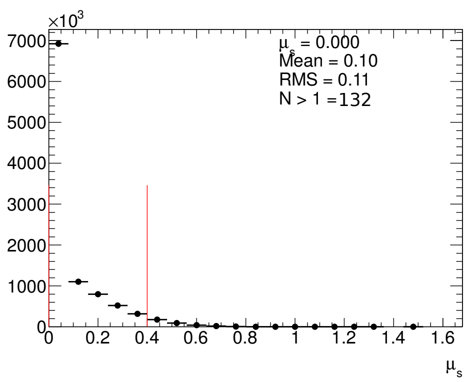


Figure 6.6: The expected distribution of fitted μ_s for the background-only hypothesis for 10^7 toy experiments. The red lines enclose 95% of toy experiments.

Scenario	N $\mu_s > 1$	p-value	Z-value ('sigma')
Electron channel	2505	2.5×10^{-4}	3.66
Muon channel	1945	1.9×10^{-4}	3.73
Combined channel	132	1.3×10^{-5}	4.36

Table 6.5: The expected exclusion of the background only hypothesis.

6.2.2 Signal strength measurement

The results of the fit to data are summarized in Table 6.7. The uncertainties calculated with toy experiments on the expected result (as described in Section 6.1.2) are propagated to this result in terms of ΔN_{events}[2]. The results are calculated using two different $W+$ QCD jets samples; the baseline ALPGEN $W+$ QCD jets sample and the alternative SHERPA $W+$ QCD jets sample. In both cases a difference is seen between the extracted value of μ_s in lepton channels. Using ALPGEN we extract $\mu_s = 1.85 \pm 0.32$ (stat.) in the electron channel and $\mu_s = 1.35 \pm 0.34$ (stat.) in the muon channel, which is consistent at the level of 1.1σ. Very similar results are found using SHERPA for the $W+$ QCD jets, with $\mu_s = 1.84 \pm 0.47$ (stat.) in the electron channel and $\mu_s = 1.27 \pm 0.49$ (stat.) in the muon. The results from electron and muon channels therefore agree at the level of 0.8σ. In both channels the χ^2/degrees of freedom is good, with better fits achieved using ALPGEN than with SHERPA.

The larger signal cross section in the electron channel is predominantly due to the presence of 50% more events in the highest mass bin as in the muon channel. Removing this bin from the fit results in measured signal strengths of 1.77 in the electron channel and 1.45 in the muon channel. Further discussion of the electron muon discrepancy is given in Section 6.3. The uncertainties in the SHERPA case are larger due to the lower statistics available in the SHERPA sample.

The fit plots are shown in Figure 6.7 and the bin contents are given in Table 6.8 for ALPGEN and Table 6.9 for SHERPA. Detector distributions after the application of the signal strengths are shown in Figures 6.8-6.11 for ALPGEN $W+$ QCD jets, and in

[2] An alternative method of assessing systematics was used as a cross check. This used the systematically varied histograms as input templates to the fitting method. Very similar results were obtained with this method, and so for clarity it is omitted here.

Figures 6.12-6.15 for Sherpa $W+$ QCD jets. The agreement between data and Monte Carlo is greatly improved by the application of μ, not just in the M_{jj} where it shows improvement by construction, but also in the $\Delta y(j_1, j_2)$ and $\Delta\phi(W, jj)$ distributions.

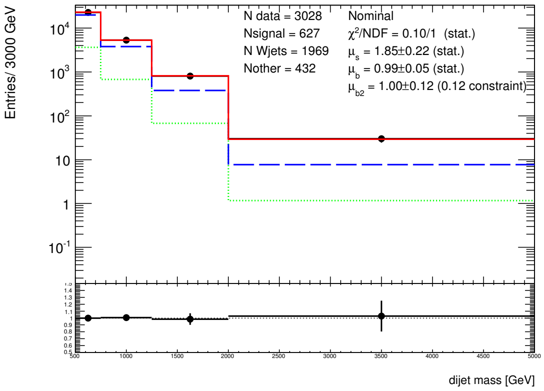
The observed background rejection is shown in Table 6.6. Since a signal strength$>$ 1 is observed the observed background rejection is greater than the expected, with the combined background-only hypothesis rejected at 4.89σ.

Scenario	N $\mu_s > \hat{\mu}_s$	p-value	Z-value ('sigma')
Electron channel	0	$< 1 \times 10^{-6}$	> 4.89
Muon channel	32	3.2×10^{-5}	4.15
Combined channel	1	1×10^{-6}	4.89

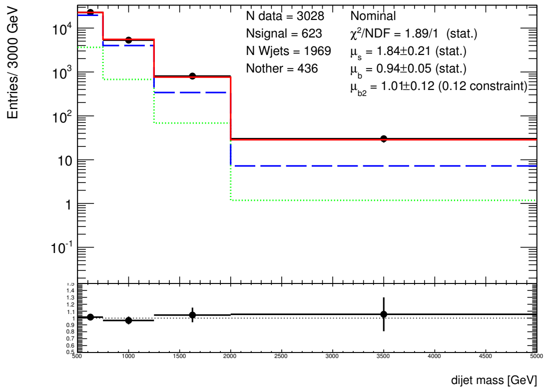
Table 6.6: The observed exclusion of the background only hypothesis including systematic and theoretical uncertainties.

Channel	N_S (exp)	N_S (obs)	N_W (exp)	N_W (obs)
ALPGEN $W+$ QCD jets				
Electron	338	627	1991	1968
	1.85 ± 0.23(fit stat.) ±0.23(MC stat.) $^{0.20}_{-0.25}$(syst.)$^{0.43}_{-0.43}$(theo.)		0.99 ± 0.05(fit stat.) ±0.01(MC stat.) $^{0.05}_{-0.05}$(syst.)$^{0.14}_{-0.14}$(theo.)	
Muon	330	447	2193	2125
	1.35 ± 0.24(fit stat.) ±0.25(MC stat.) $^{0.40}_{-0.26}$(syst.)$^{0.43}_{-0.43}$(theo.)		0.97 ± 0.05(fit stat.) ±0.01(MC stat.) $^{0.03}_{-0.04}$(syst.)$^{0.14}_{-0.14}$(theo.)	
Combined	669	1088	4184	4079
	1.63 ± 0.17(fit stat.) ±0.18(MC stat.) $^{0.28}_{-0.23}$(syst.)$^{0.43}_{-0.43}$(theo.)		0.97 ± 0.10(fit stat.) ±0.06(MC stat.) $^{0.20}_{-0.00}$(syst.)$^{0.14}_{-0.14}$(theo.)	
SHERPA $W+$ QCD jets				
Electron	338	622	2099	1968
	1.84 ± 0.21(fit stat.) ±0.33(MC stat.) $^{0.36}_{-1.06}$(syst.)		0.94 ± 0.05(fit stat.) ±0.03(MC stat.) $^{0.13}_{-0.04}$(syst.)	
Muon	330	418	2073	2143
	1.27 ± 0.26(fit stat.) ±0.42(MC stat.) $^{0.39}_{-0.52}$(syst.)		1.03 ± 0.05(fit stat.) ±0.01(MC stat.) $^{0.03}_{-0.07}$(syst.)	
Combined	669	1044	4173	4119
	1.56 ± 0.18(fit stat.) ±0.28(MC stat.) $^{0.22}_{-0.40}$(syst.)		0.99 ± 0.11(fit stat.) ±0.07(MC stat.) $^{0.24}_{-0.00}$(syst.)	

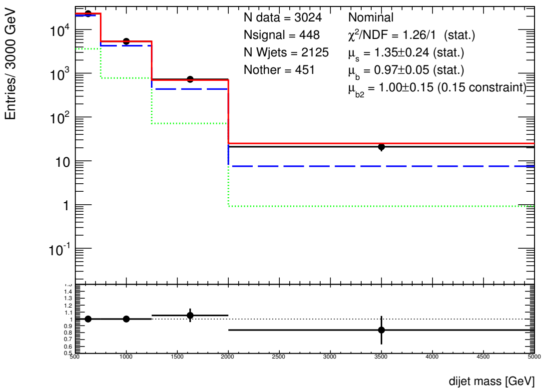
Table 6.7: Summary of fit results, with systematic errors summed in quadrature. The first line of each entry gives the number of events expected and observed and the second gives the signal strength together with the uncertainty. The uncertainty labeled 'fit stat.' is the statistical error on the fit and that labeled 'MC stat' is that from Monte Carlo statistics. The top section of the table is the result using the base-line ALPGEN $W+$ QCD jets sample, and the lower section uses the SHERPA sample.



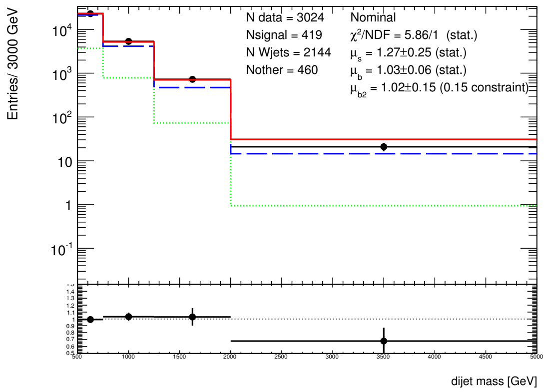
(a) Electron fit (ALPGEN)



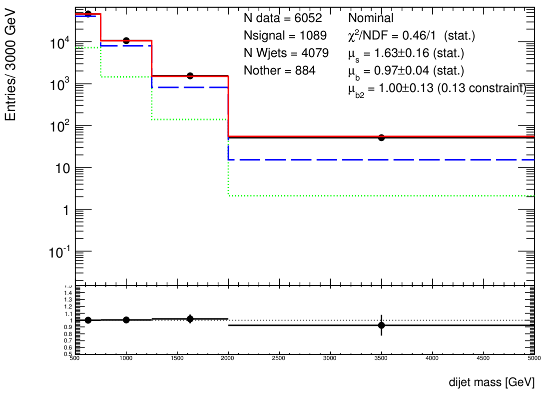
(b) Electron fit (SHERPA)



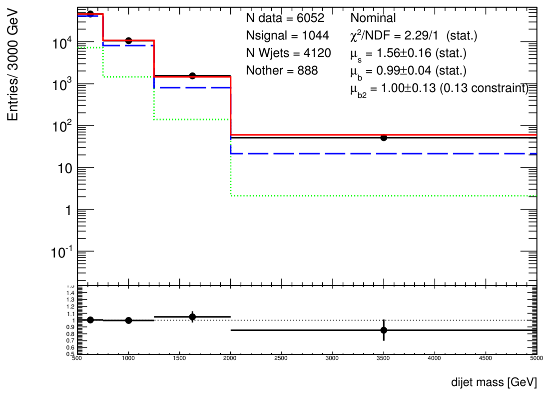
(c) Muon fit (ALPGEN)



(d) Muon fit (SHERPA)



(e) Combined fit (ALPGEN)



(f) Combined fit (SHERPA)

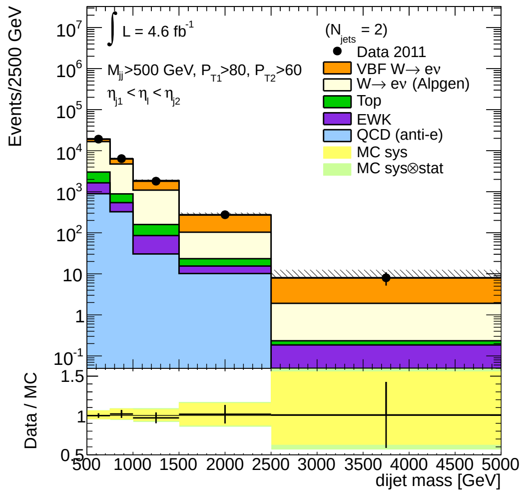
Figure 6.7: Electron, muon and combined fits. The green dotted curve is the sum of all non W+jets backgrounds. The blue dashed curve is the $W+$ QCD jets background stacked on the other backgrounds. The solid red curve is the $W+$ EWK jets signal.

Component	500-750 GeV		750-1250 GeV		1250-2000 GeV		2000-5000 GeV	
	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
Electron Channel								
Signal	133 ± 5	246 ± 114	136 ± 5	252 ± 116	58 ± 3	108 ± 50	12 ± 2	21 ± 10
$W +$ EWK jets	1379 ± 30	1363 ± 34	527 ± 21	521 ± 22	79 ± 7	78 ± 7	7 ± 2	7 ± 2
background (other)	302 ± 14	302 ± 14	112 ± 9	112 ± 9	17 ± 4	17 ± 4	1 ± 0	1 ± 0
Data	—	1909	—	889	—	200	—	30
Muon Channel								
Signal	134 ± 5	181 ± 48	136 ± 5	185 ± 49	47 ± 3	64 ± 17	13 ± 2	18 ± 5
$W +$ EWK jets	1495 ± 30	1449 ± 55	596 ± 18	578 ± 25	95 ± 8	92 ± 9	7 ± 2	7 ± 2
background (other)	303 ± 11	303 ± 11	130 ± 8	130 ± 8	18 ± 3	18 ± 3	1 ± 0	1 ± 0
Data	—	1928	—	892	—	183	—	21
Combined Channel								
Signal	267 ± 7	434 ± 168	272 ± 7	443 ± 171	106 ± 5	172 ± 67	25 ± 2	40 ± 16
$W +$ EWK jets	2874 ± 43	2802 ± 83	1123 ± 28	1095 ± 39	174 ± 11	170 ± 11	13 ± 3	13 ± 3
background (other)	605 ± 18	605 ± 18	242 ± 12	242 ± 12	35 ± 5	35 ± 5	2 ± 0	2 ± 0
Data	—	3837	—	1781	—	383	—	51

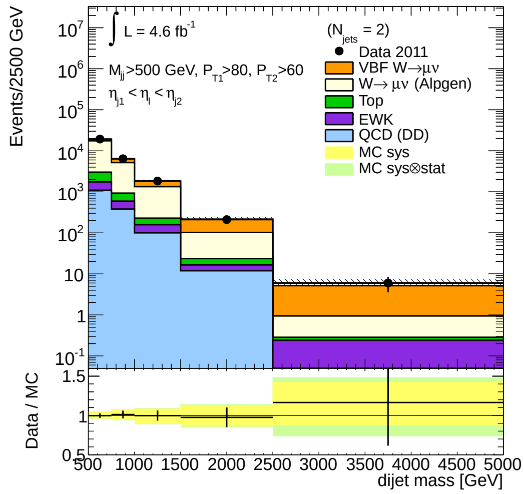
Table 6.8: Contents of the ALPGEN fit histograms displayed bin-by-bin. The errors are Monte Carlo statistical uncertainties only for expected, and Monte Carlo statistical plus fit statistical uncertainties for the observed bins. There are no events for $M_{jj} > 5000$ GeV.

Component	500-750 GeV		750-1250 GeV		1250-2000 GeV		2000-5000 GeV	
	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
Electron Channel								
Signal	133 ± 5	244 ± 112	136 ± 5	250 ± 114	58 ± 3	107 ± 49	12 ± 2	21 ± 10
$W +$ EWK jets	1426 ± 64	1337 ± 108	596 ± 45	558 ± 56	72 ± 13	68 ± 13	6 ± 3	6 ± 3
background (other)	302 ± 14	305 ± 15	112 ± 9	113 ± 9	17 ± 4	17 ± 4	1 ± 0	1 ± 0
Data	—	1909	—	889	—	200	—	30
Muon Channel								
Signal	134 ± 5	170 ± 36	136 ± 5	173 ± 37	47 ± 3	60 ± 13	13 ± 2	16 ± 4
$W +$ EWK jets	1422 ± 56	1471 ± 76	542 ± 37	560 ± 43	96 ± 17	100 ± 17	13 ± 5	14 ± 5
background (other)	303 ± 11	309 ± 13	130 ± 8	132 ± 8	18 ± 3	18 ± 3	1 ± 0	1 ± 0
Data	—	1928	—	892	—	183	—	21
Combined Channel								
Signal	267 ± 7	416 ± 150	272 ± 7	425 ± 153	106 ± 5	165 ± 60	25 ± 2	38 ± 14
$W +$ EWK jets	2848 ± 85	2812 ± 92	1138 ± 59	1123 ± 60	168 ± 21	166 ± 21	20 ± 6	19 ± 6
background (other)	605 ± 18	608 ± 18	242 ± 12	243 ± 12	35 ± 5	35 ± 5	2 ± 0	2 ± 0
Data	—	3837	—	1781	—	383	—	51

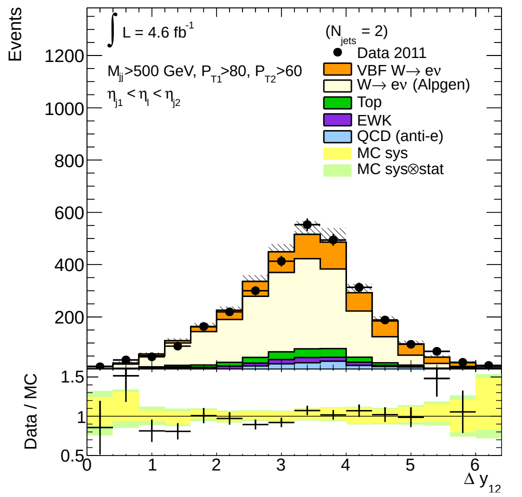
Table 6.9: Contents of the SHERPA fit histograms displayed bin-by-bin. The errors are Monte Carlo statistical uncertainties only for expected, and Monte Carlo statistical plus fit statistical uncertainties for the observed bins. There are no events for $M_{jj} > 5000$ GeV.



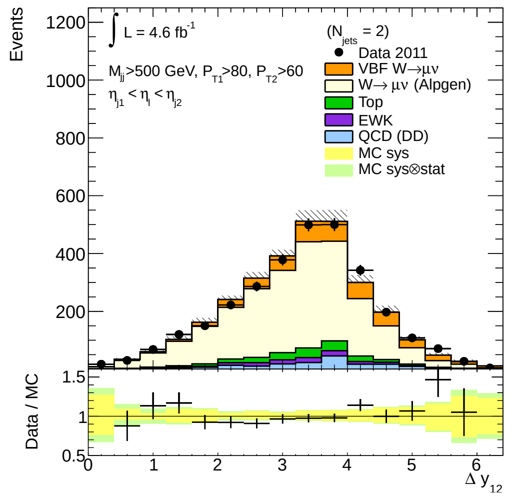
(a) Electron M_{jj}



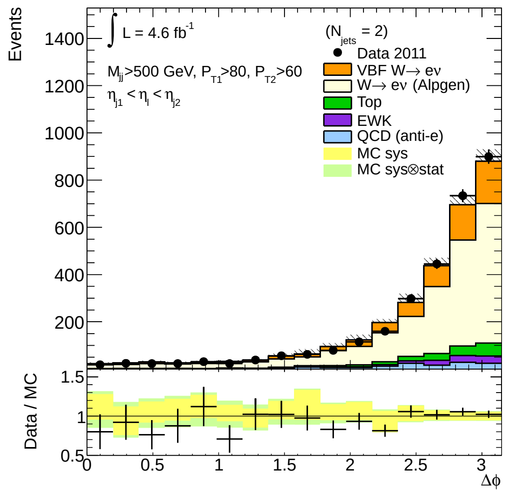
(b) Muon M_{jj}



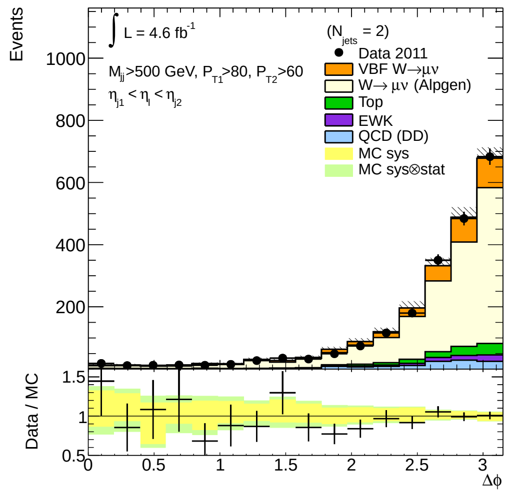
(c) Electron $\Delta y(j_1, j_2)$



(d) Muon $\Delta y(j_1, j_2)$

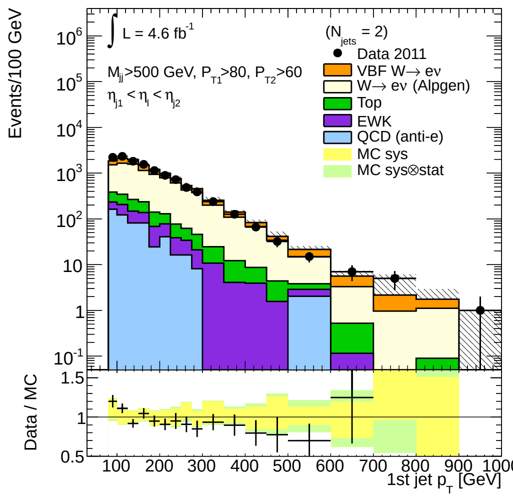


(e) Electron $\Delta\phi(j_1, j_2)$

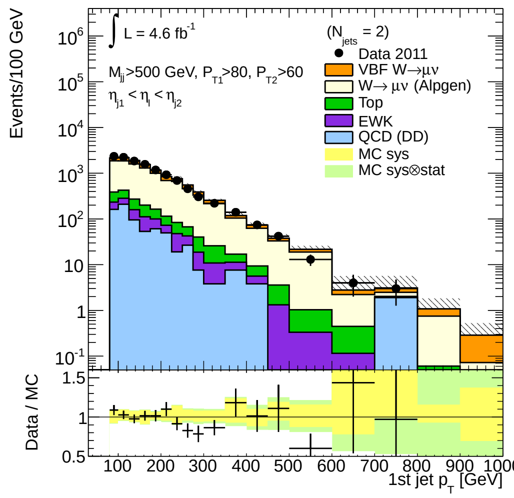


(f) Muon $\Delta\phi(j_1, j_2)$

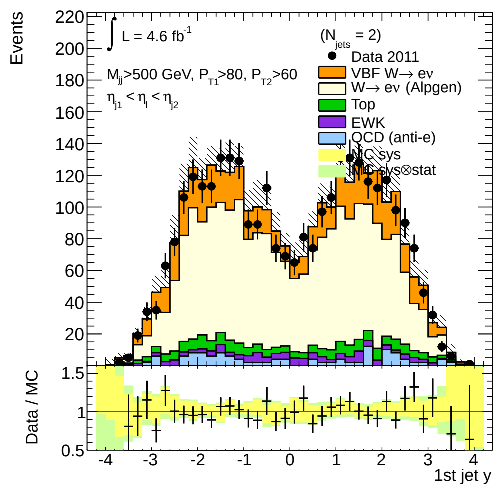
Figure 6.8: Detector level dijet distributions using ALPGEN for $W +$ QCD jets in the signal region after all VBF selections and applying the μ factors extracted from the fits.



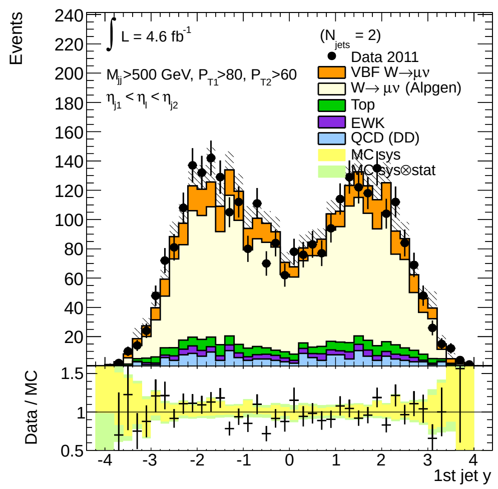
(a) Electron jet leading p_{T}^{jet}



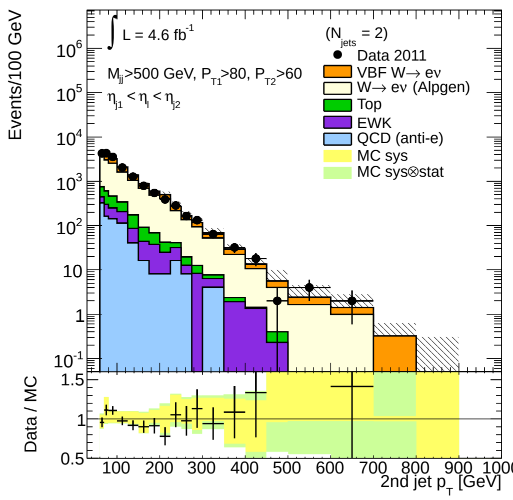
(b) Muon jet leading p_{T}^{jet}



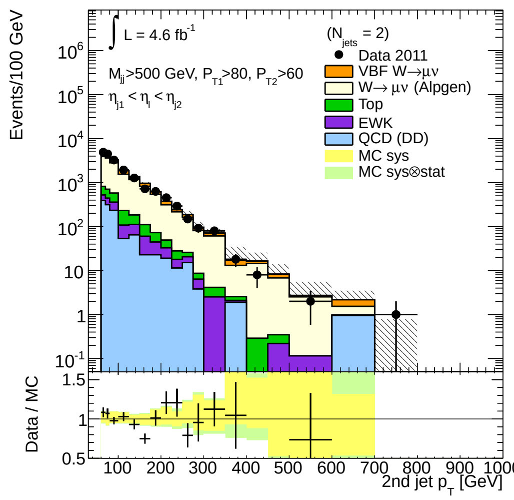
(c) Leading jet rapidity (electron channel)



(d) Leading jet rapidity (muon channel)

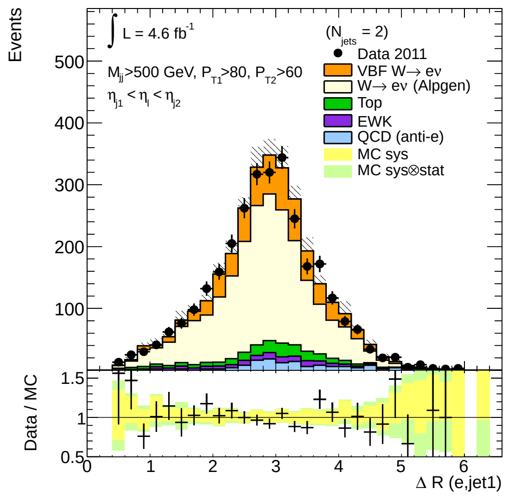


(e) Electron jet $2^{nd} p_{\mathrm{T}}^{jet}$

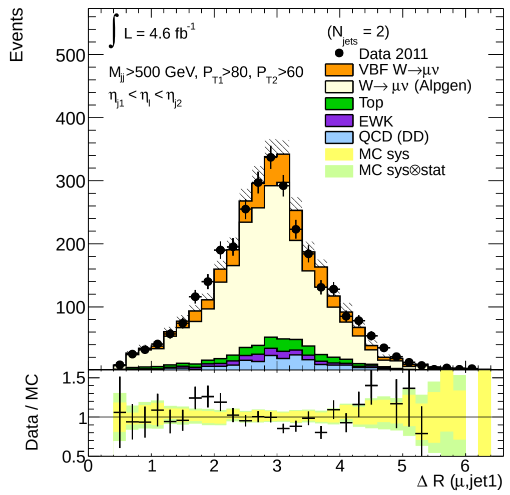


(f) Muon jet $2^{nd} p_{\mathrm{T}}^{jet}$

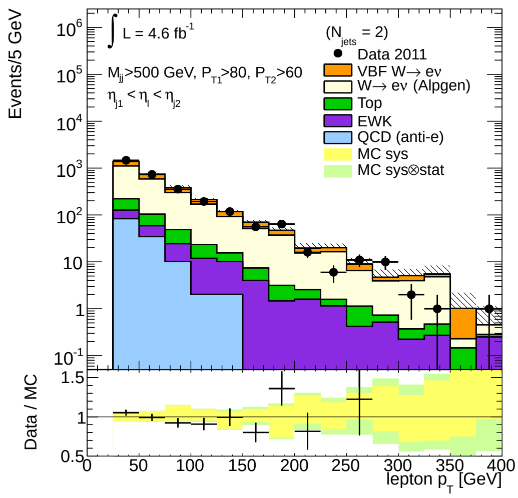
Figure 6.9: Detector level jet distributions using ALPGEN for $W+$ QCD jets in the signal region after all VBF selections and applying the μ factors extracted from the fits.



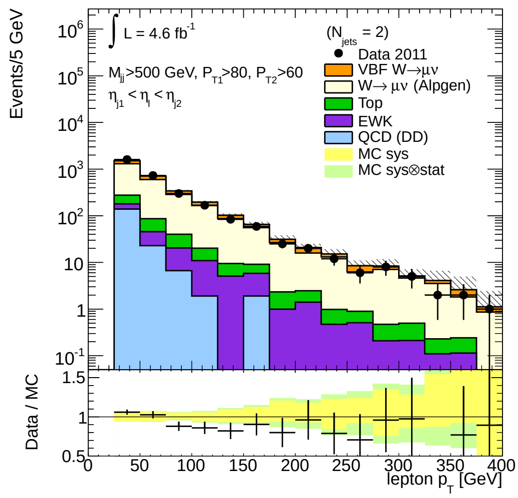
(a) Electron $\Delta R(\ell, j_1)$



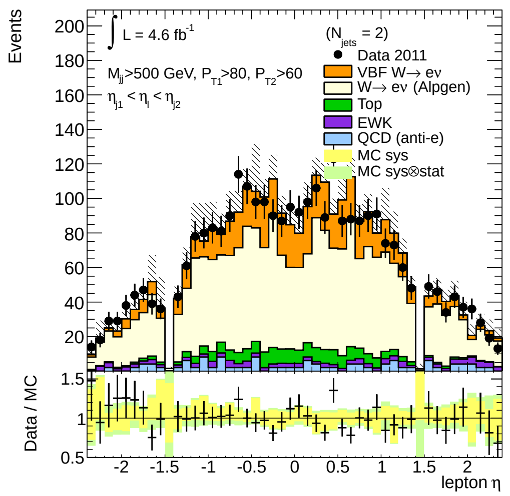
(b) Muon $\Delta R(\ell, j_1)$



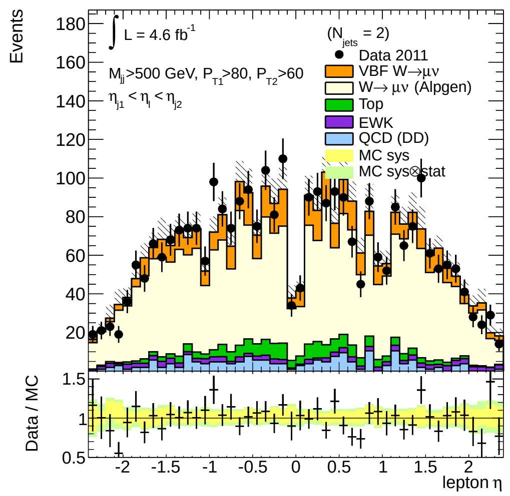
(c) Electron p_{T}



(d) Muon p_{T}

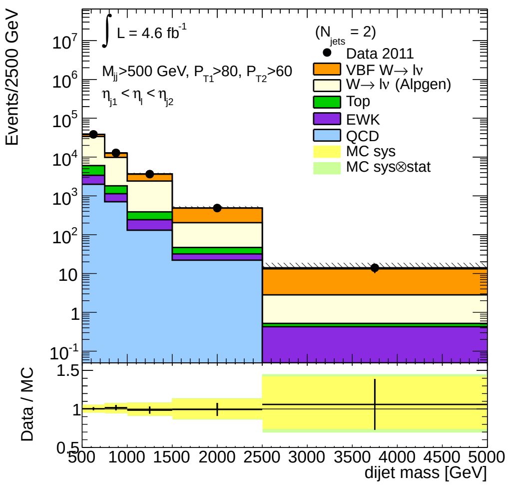


(e) Electron η

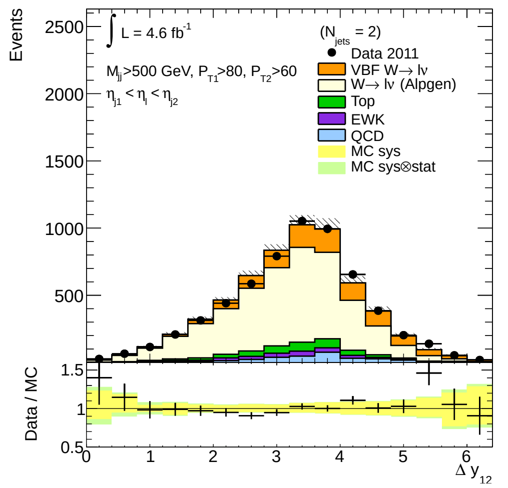


(f) Muon η

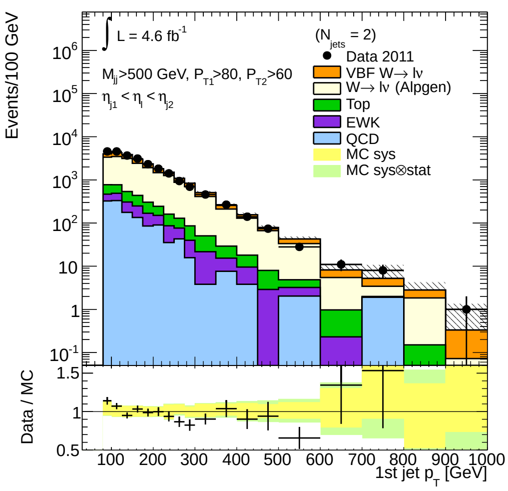
Figure 6.10: Detector level lepton distributions using ALPGEN for $W +$ QCD jets in the signal region after all VBF selections and applying the μ factors extracted from the fits.



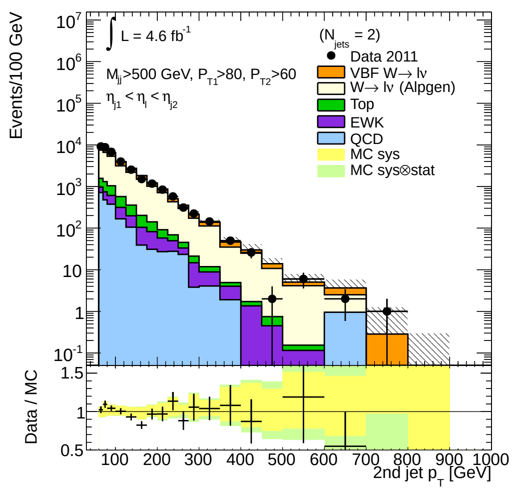
(a) Combined channel M_{jj}



(b) Combined channel $\Delta y(j_1, j_2)$

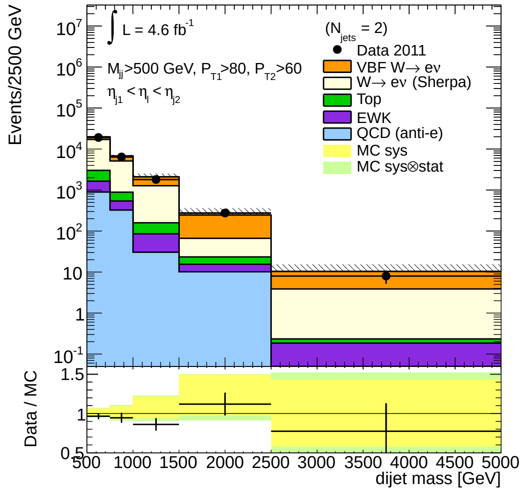


(c) Combined channel leading jet p_{T}

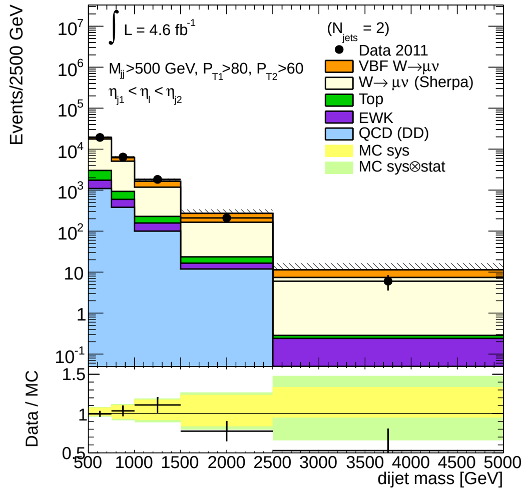


(d) Combined channel sub leading jet p_{T}

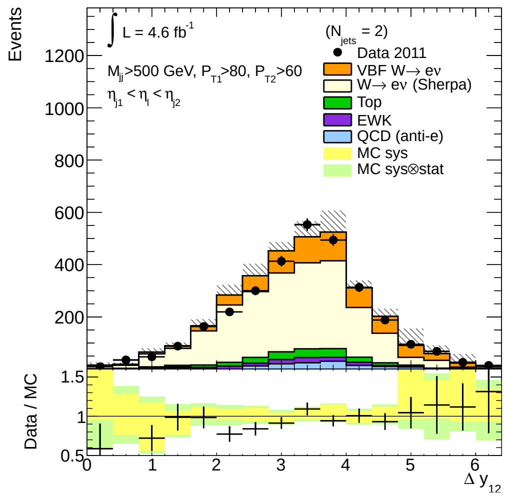
Figure 6.11: Detector level distributions using ALPGEN for W + QCD jets in the signal region for the combined electron and muon channels after all VBF selections and applying the μ factors extracted from the fits.



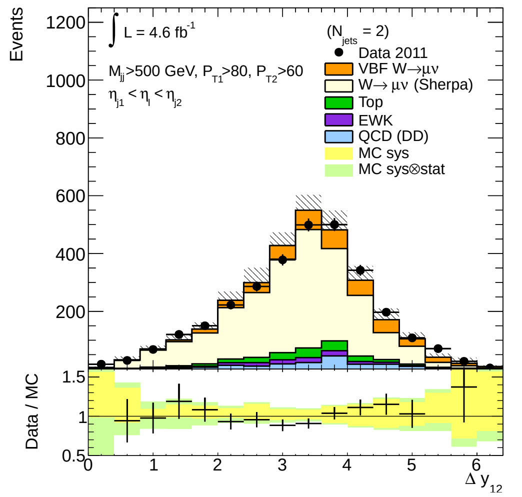
(a) Electron M_{jj}



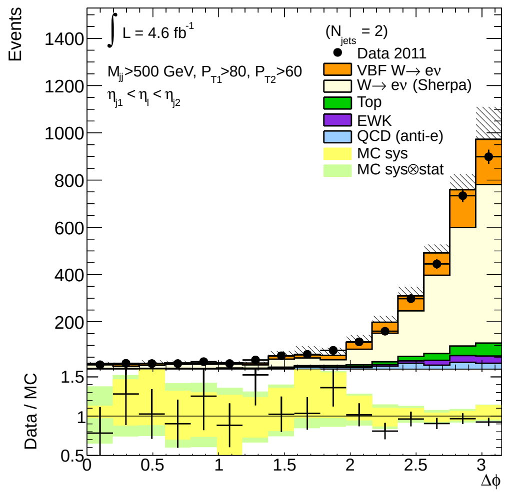
(b) Muon M_{jj}



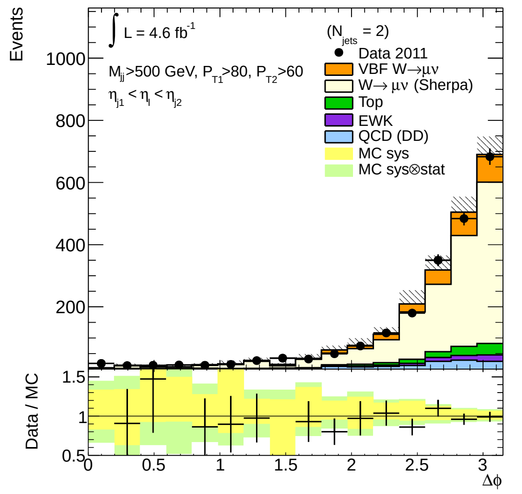
(c) Electron $\Delta y(j_1, j_2)$



(d) Muon $\Delta y(j_1, j_2)$

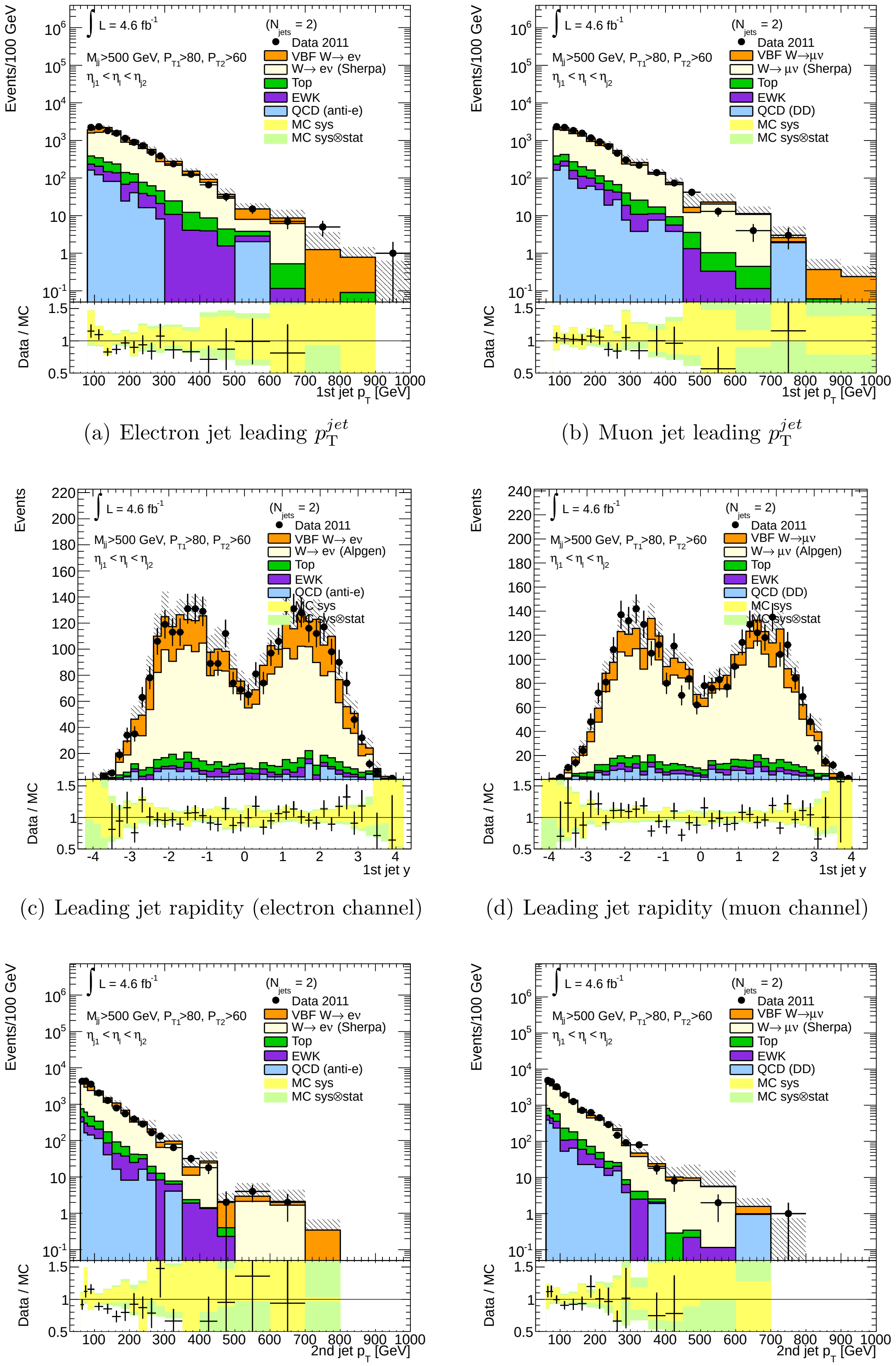


(e) Electron $\Delta\phi(j_1, j_2)$



(f) Muon $\Delta\phi(j_1, j_2)$

Figure 6.12: Detector level dijet distributions using SHERPA for $W+$ QCD jets in the signal region after all VBF selections and applying the μ factors extracted from the fits.



(a) Electron jet leading p_{T}^{jet}

(b) Muon jet leading p_{T}^{jet}

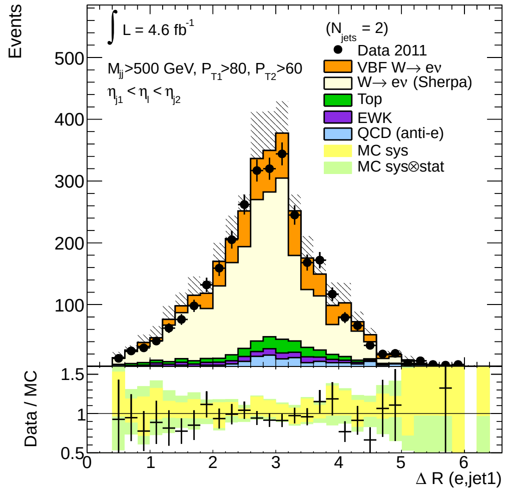
(c) Leading jet rapidity (electron channel)

(d) Leading jet rapidity (muon channel)

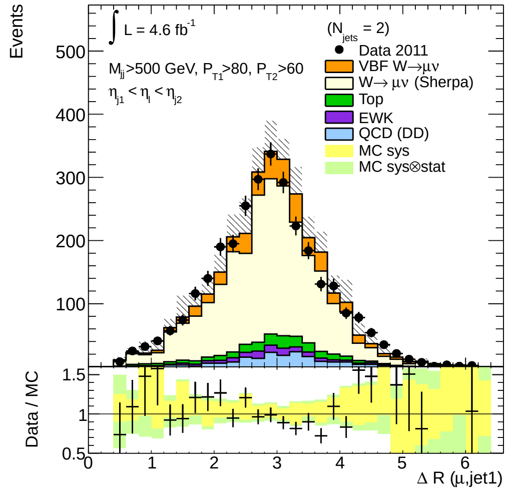
(e) Electron jet $2^{nd} p_{\mathrm{T}}^{jet}$

(f) Muon jet $2^{nd} p_{\mathrm{T}}^{jet}$

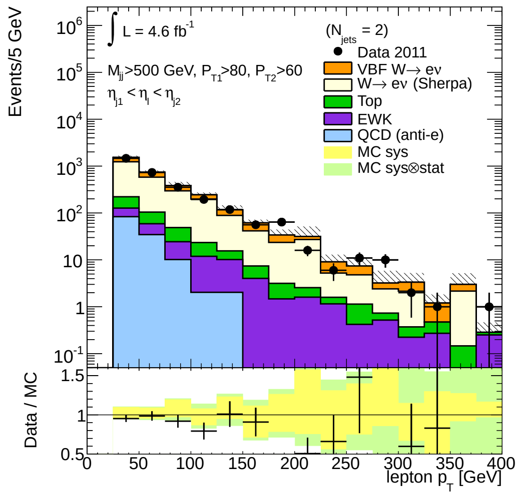
Figure 6.13: Detector level distributions using SHERPA for $W+$ QCD jets in the signal region after all VBF selections and applying the μ factors extracted from the fits.



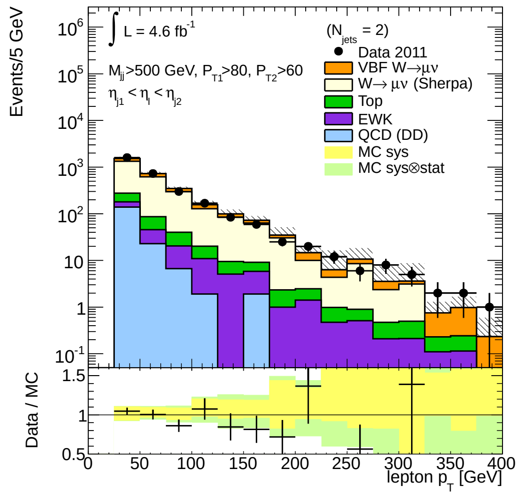
(a) Electron $\Delta R(\ell, j_1)$



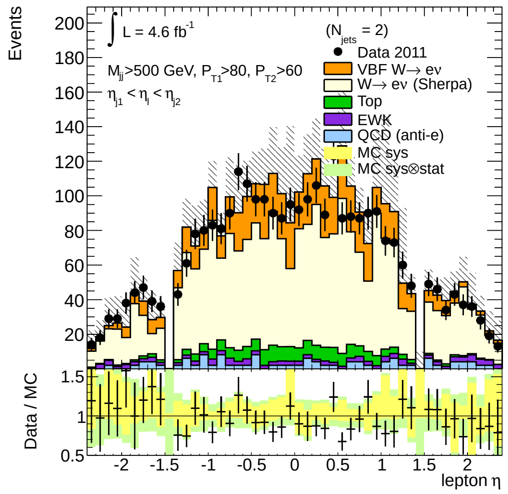
(b) Muon $\Delta R(\ell, j_1)$



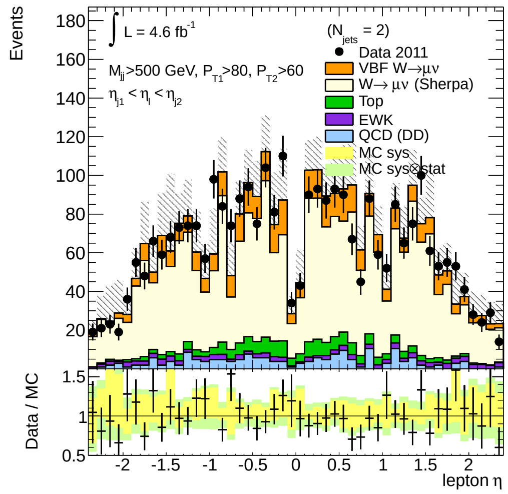
(c) Electron p_{T}



(d) Muon p_{T}

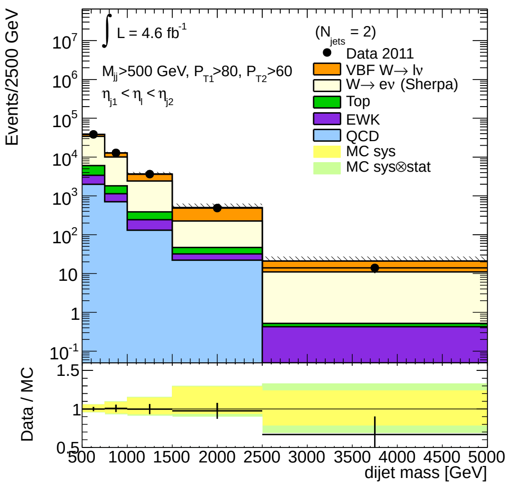


(e) Electron η

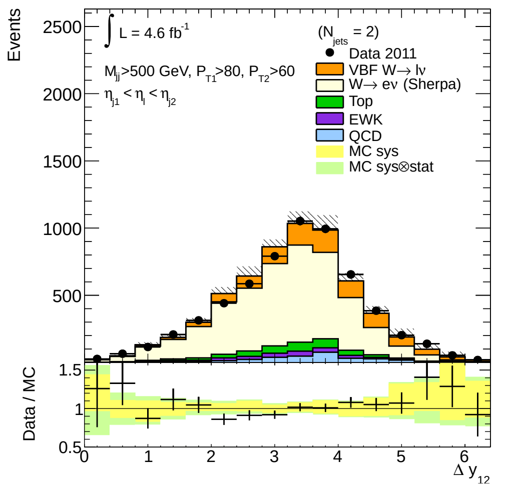


(f) Muon η

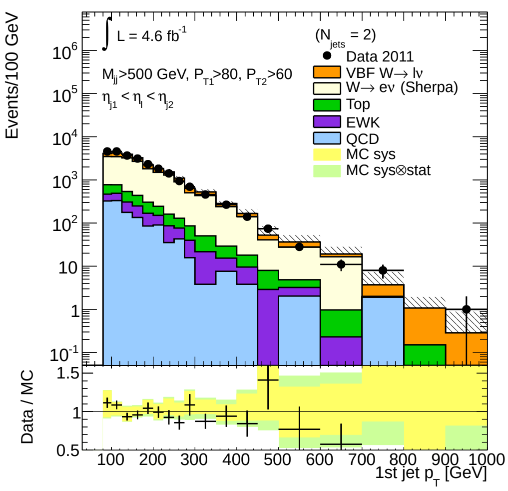
Figure 6.14: Detector level lepton distributions using SHERPA for $W +$ QCD jets in the signal region after all VBF selections and applying the μ factors extracted from the fits.



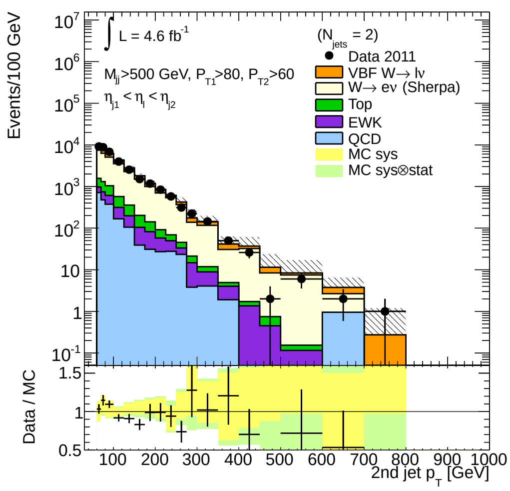
(a) Combined channel M_{jj}



(b) Combined channel $\Delta y(j_1, j_2)$



(c) Combined channel leading jet p_{T}



(d) Combined channel sub leading jet p_{T}

Figure 6.15: Detector level distributions using SHERPA for $W+$ QCD jets in the signal region for the combined electron and muon channels after all VBF selections and applying the μ factors extracted from the fits.

6.2.3 Cross-section calculation

We define two cross sections:

- Visible cross section:

$$\sigma_{vis} = \frac{N_{\mathrm{S}}}{\mathscr{L}}; \tag{6.11}$$

- Fiducial cross section:

$$\sigma_{fid} = \frac{N_{\mathrm{S}}}{\mathscr{L} \times C_{VBFW}}; \tag{6.12}$$

In these definitions:

- N_{S} is extracted from the fit,
- C_{VBFW} is the reconstruction correction; the ratio between the reconstructed and truth number of events after the VBF cuts: $C_{VBFW} = \frac{N_{\mathrm{reco}}}{N_{\mathrm{truth\ VBF}}}$. This is taken entirely from Monte Carlo without any input from the fit.
- $\mathscr{L}$ is the integrated luminosity.

The i^{th} systematic uncertainty is included on the fiducial cross section as:

$$\Delta_i\sigma^2 = \left(\frac{\partial\sigma}{\partial N_{\mathrm{S}}}\right)^2 \Delta_i N_{\mathrm{S}}^2 + \left(\frac{\partial\sigma}{\partial N_{\mathrm{reco}}}\right)^2 \Delta_i N_{\mathrm{reco}}^2 \tag{6.13}$$

where $\Delta_i N_{\mathrm{S}}$ is the difference observed in fitted signal due to this systematic and $\Delta_i N_{\mathrm{reco}}$ is the difference in the expected number of events dues to systematic i. The results are shown in Table 6.10, with combined fiducial cross section of $\sigma = 325 \pm 6$ (lumi) ± 32 (stat) ${}^{+63}_{-70}$ (syst) ± 86 (theo) fb. A calculation with VBF@NLO in the fiducial phase-space with lepton p_{T} $p_T^{\ell} > 25$ GeV, lepton pseudo-rapidity$|\eta^{\ell}| < 2.4$, missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV, transverse mass $m_T > 40$ GeV, jet rapidity $|y| < 4.4$, leading jet $p_T^{j1} > 80$ GeV, second leading jet $p_T^{j2} > 60$ GeV, no third jet satisfying $p_T^{j3} > 30$ GeV, dijet mass $M_{jj} > 500$ GeV, and outside lepton vet $\eta^{ja} < \eta^{\ell} < \eta^{jb}$ yields a cross section of 425 ± 21 fb, which is compatible with the above result at the level of 1σ if all systematic uncertainties are added in quadrature.

Cross sections for electron channel (fb)	
σ_{vis} =	135 ± 2 (lumi) ± 16 (stat) ${}^{25}_{-23}$ (syst) ± 31 (theo)
σ_{fid} =	182 ± 3 (lumi) ± 21 (stat) ${}^{34}_{-32}$ (syst) ± 42 (theo)
Efficiency factor C_{VBFW} =	0.74
Cross sections for muon channel (fb)	
σ_{vis} =	96 ± 2 (lumi) ± 17 (stat) ${}^{26}_{-34}$ (syst) ± 31 (theo)
σ_{fid} =	138 ± 2 (lumi) ± 24 (stat) ${}^{37}_{-50}$ (syst) ± 44 (theo)
Efficiency factor C_{VBFW} =	0.70
Cross sections for combined electron and muon channel (fb)	
σ_{vis} =	234 ± 4 (lumi) ± 23 (stat) ${}^{45}_{-48}$ (syst) ± 62 (theo)
σ_{fid} =	325 ± 6 (lumi) ± 32 (stat) ${}^{63}_{-70}$ (syst) ± 86 (theo)
Efficiency factor C_{VBFW} =	0.72

Table 6.10: Measured cross sections for the electron, muon and combined lepton channels extracted in femtobarns.

6.3 Cross-checks

To provide further confidence that the results in the signal region are not biased, several cross-checks were performed. In particular it is important to investigate the source of the electron and muon difference. The most informative of these cross-checks are presented here.

6.3.1 Validation region distributions

The validation region described in Section 4.5 is scaled by the cross section obtained from the fit in the signal region shown in Table 6.7. These distributions are shown in Figures 6.16-6.18 for the lepton channels separately and for the combined lepton channels in Figure 6.19. Both lepton channels and their combinations show good agreement after the application of fitted μ_s and μ_W values from the signal region.

An independent fit is carried out in the validation region and the results shown in Table 6.11. There are very large errors from all sources due to the limited statistics available in this phase space. The result is consistent with that found in the signal region for all channels.

Channel	N_S (exp)	N_S (obs)	N_W (exp)	N_W (obs)
Results extracted in the $N_{jets} = 3$ validation region				
Electron	108	8	1551	1471
	0.08±2.15(fit stat.) ±0.58(MC stat.) $^{0.88}_{-1.14}$(syst.)		0.95±0.10(fit stat.) ±0.03(MC stat.) $^{0.11}_{-0.11}$(syst.)	
Muon	99	100	1679	1497
	1.02±0.81(fit stat.) ±0.56(MC stat.) $^{0.56}_{-1.57}$(syst.)		0.89±0.09(fit stat.) ±0.01(MC stat.) $^{0.07}_{-0.07}$(syst.)	
Combined	208	104	3231	2966
	0.50±0.57(fit stat.) ±0.57(MC stat.) $^{0.81}_{-0.98}$(syst.)		0.92±0.07(fit stat.) ±0.01(MC stat.) $^{0.05}_{-0.10}$(syst.)	
Semi-inclusive: $N_{jets} = 2$ or 3				
Electron	447	664	3542	3413
	1.48±0.23(fit stat.) ±0.28(MC stat.) $^{0.42}_{-0.39}$(syst.)		0.96±0.05(fit stat.) ±0.01(MC stat.) $^{0.06}_{-0.09}$(syst.)	
Muon	429	520	3873	3668
	1.21±0.24(fit stat.) ±0.28(MC stat.) $^{0.45}_{-0.44}$(syst.)		0.95±0.06(fit stat.) ±0.00(MC stat.) $^{0.04}_{-0.05}$(syst.)	
Combined	877	1193	7415	7067
	1.36±0.17(fit stat.) ±0.19(MC stat.) $^{0.42}_{-0.38}$(syst.)		0.95±0.05(fit stat.) ±0.00(MC stat.) $^{0.04}_{-0.06}$(syst.)	
Inclusive: $N_{jets} \geq 2$				
Electron	479	600	4535	4214
	1.25±0.28(fit stat.) ±0.38(MC stat.) $^{0.67}_{-0.51}$(syst.)		0.93±0.20(fit stat.) ±0.15(MC stat.) $^{2.99}_{-0.54}$(syst.)	
Muon	459	374	4882	4564
	0.81±0.30(fit stat.) ±0.33(MC stat.) $^{0.64}_{-0.72}$(syst.)		0.93±0.07(fit stat.) ±0.00(MC stat.) $^{0.04}_{-0.09}$(syst.)	
Combined	939	986	9418	8721
	1.05±0.22(fit stat.) ±0.25(MC stat.) $^{0.64}_{-0.60}$(syst.)		0.93±0.06(fit stat.) ±0.00(MC stat.) $^{0.04}_{-0.11}$(syst.)	

Table 6.11: Fit results extracted in the validation region, with $N_{jets} = 3$, a semi-inclusive jet selection, with $N_{jets} = 2$ or 3 and a fully inclusive jet selection with $N_{jets} \geq 2$.

6.3.2 Selection cross-check

The signal selection was chosen to maximize the signal purity and significance, and to minimize the systematic errors. If this method is unbiased then results extracted before some of these cuts should be consistent with the main result, within the uncertainties.

The extracted signal strengths and uncertainties are shown in Figure 6.20. These show the errors reducing as successive cuts are applied. The fitted value of μ_s varies but is consistent within uncertainties between selections.

The result was also reproduced with a cut on the rapidity difference bewteen the tag jets $\Delta y(j_1, j_2) > 2$ cut (labelled "highdy"). The main selection requires $N_{jets} = 2$; the result was also produced with $N_{jets} \geq 2$ (labelled "inclusive") and $N_{jets} = 2$ or 3 (labelled "semi-inclusive"). These are shown in Figures 6.20 and 6.21. All these results are summarized in Table 6.12.

6.3.3 Binning and range cross-check

Additional fits are performed for binning and range variations and shown in Figures 6.22 and 6.23. These show a similar pattern, with smaller errors in our nominal selection than the alternatives, and compatible fitted values across all binnings. This shows that the fitting procedure is not heavily dependent on the binning chosen. Figure 6.23(b) shows that the discrepancy observed between electron and muon channels is partially driven by the highest M_{jj} bin. When this bin is removed from the fit the signal strengths observed are more consistent. These results are also presented in Table 6.13.

The binnings used are:

- "Binning 1" has bin edges `[0, 100, 200, 300, 400, 500, 700, 1100, 1300, 1500, 1750, 2000, 2500, 5000]`,
- "Binning 2" has bin edges `[0, 100, 200, 300, 400, 500, 800, 1100, 1500, 2000, 3000, 5000]`,
- "Binning 3" has bin edges `[0, 100, 200, 300, 400, 500, 750, 1000, 1500, 2500, 5000]`.
- The "Nominal Binning" has bin edges `[0, 100, 200, 300, 400, 500, 750, 1250, 2000, 5000]`

Channel	N_S (exp)	N_S (obs)	N_W (exp)	N_W (obs)
		$N_{jets} = 2$, $M_{jj} > 500$ GeV		
Electron	608	964	6082	7350
		1.59 ± 0.27(fit stat.) ±0.27(MC stat.) $^{0.85}_{-0.53}$(syst.)		1.21 ± 0.04(fit stat.) ±0.01(MC stat.) $^{0.11}_{-0.05}$(syst.)
Muon	618	958	7041	8224
		1.55 ± 0.22(fit stat.) ±0.31(MC stat.) $^{0.52}_{-0.42}$(syst.)		1.17 ± 0.04(fit stat.) ±0.00(MC stat.) $^{0.06}_{-0.09}$(syst.)
Combined	1226	1937	13037	15601
		1.58 ± 0.16(fit stat.) ±0.23(MC stat.) $^{0.55}_{-0.43}$(syst.)		1.20 ± 0.04(fit stat.) ±0.00(MC stat.) $^{0.06}_{-0.07}$(syst.)
		$N_{jets} = 2$, $M_{jj} > 500$ GeV, high p_T		
Elec	397	749	3459	3686
		1.88 ± 0.23(fit stat.) ±0.27(MC stat.) $^{0.22}_{-0.30}$(syst.)		1.07 ± 0.04(fit stat.) ±0.01(MC stat.) $^{0.05}_{-0.03}$(syst.)
Muon	399	530	3995	4075
		1.33 ± 0.23(fit stat.) ±0.27(MC stat.) $^{0.39}_{-0.29}$(syst.)		1.02 ± 0.04(fit stat.) ±0.00(MC stat.) $^{0.03}_{-0.04}$(syst.)
Combined	797	1293	7455	7741
		1.62 ± 0.17(fit stat.) ±0.19(MC stat.) $^{0.29}_{-0.30}$(syst.)		1.04 ± 0.03(fit stat.) ±0.00(MC stat.) $^{0.03}_{-0.02}$(syst.)
		Nominal selection: $N_{jets} = 2$, $M_{jj} > 500$ GeV, high p_T, outside lepton veto		
Electron	338	627	1991	1968
		1.85 ± 0.23(fit stat.) ±0.23(MC stat.) $^{0.20}_{-0.25}$(syst.)$^{0.43}_{-0.43}$(theo.)		0.99 ± 0.05(fit stat.) ±0.01(MC stat.) $^{0.05}_{-0.05}$(syst.)$^{0.14}_{-0.14}$(theo.)
Muon	330	447	2193	2125
		1.35 ± 0.24(fit stat.) ±0.25(MC stat.) $^{0.40}_{-0.26}$(syst.)$^{0.43}_{-0.43}$(theo.)		0.97 ± 0.05(fit stat.) ±0.01(MC stat.) $^{0.03}_{-0.04}$(syst.)$^{0.14}_{-0.14}$(theo.)
Combined	669	1088	4184	4079
		1.63 ± 0.17(fit stat.) ±0.18(MC stat.) $^{0.28}_{-0.23}$(syst.)$^{0.43}_{-0.43}$(theo.)		0.97 ± 0.10(fit stat.) ±0.06(MC stat.) $^{0.20}_{-0.00}$(syst.)$^{0.14}_{-0.14}$(theo.)
		High dy: $N_{jets} \geq 2$, $M_{jj} > 500$ GeV, high p_T, outside lepton veto, $\Delta y(j_1, j_2) > 2$		
Electron	318	562	1707	1734
		1.77 ± 0.23(fit stat.) ±0.23(MC stat.) $^{0.23}_{-0.24}$(syst.)		1.02 ± 0.05(fit stat.) ±0.01(MC stat.) $^{0.05}_{-0.07}$(syst.)
Muon	311	421	1883	1809
		1.35 ± 0.23(fit stat.) ±0.25(MC stat.) $^{0.41}_{-0.27}$(syst.)		0.96 ± 0.06(fit stat.) ±0.01(MC stat.) $^{0.04}_{-0.04}$(syst.)
Combined	630	997	3590	3529
		1.58 ± 0.16(fit stat.) ±0.17(MC stat.) $^{0.28}_{-0.23}$(syst.)		0.98 ± 0.04(fit stat.) ±0.01(MC stat.) $^{0.03}_{-0.04}$(syst.)

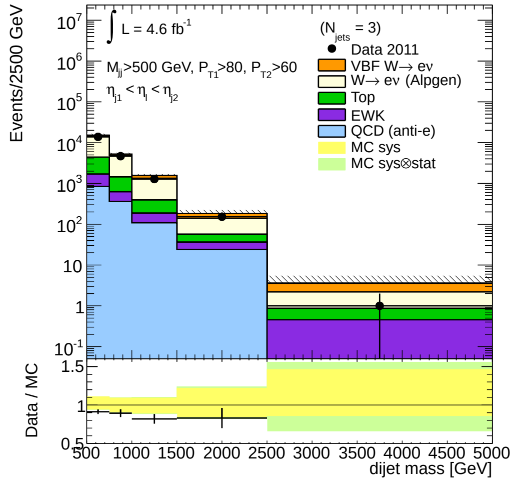
Table 6.12: Fit results for a variety of selections in the electron, muon and combined channels.

Channel	N_{S} (exp)	N_{S}	N_W (exp)	N_W (obs)
Fit min 750 GeV				
Electron	205	373	612	615
	1.81 ± 0.31(fit stat.) ± 0.27(MC stat.) $^{0.30}_{-0.29}$(syst.)		1.01 ± 0.12(fit stat.) ± 0.00(MC stat.) $^{0.10}_{-0.11}$(syst.)	
Muon	196	249	697	697
	1.27 ± 0.33(fit stat.) ± 0.31(MC stat.) $^{0.44}_{-0.27}$(syst.)		1.00 ± 0.12(fit stat.) ± 0.02(MC stat.) $^{0.04}_{-0.06}$(syst.)	
Combined	402	640	1310	1295
	1.59 ± 0.25(fit stat.) ± 0.21(MC stat.) $^{0.33}_{-0.22}$(syst.)		0.99 ± 0.08(fit stat.) ± 0.01(MC stat.) $^{0.03}_{-0.06}$(syst.)	
Fit max 2000 GeV				
Electron	338	599	1991	1966
	1.77 ± 0.28(fit stat.) ± 0.23(MC stat.) $^{0.32}_{-0.21}$(syst.)		0.99 ± 0.06(fit stat.) ± 0.02(MC stat.) $^{0.06}_{-0.04}$(syst.)	
Muon	330	478	2193	2074
	1.45 ± 0.30(fit stat.) ± 0.26(MC stat.) $^{0.50}_{-0.26}$(syst.)		0.95 ± 0.06(fit stat.) ± 0.01(MC stat.) $^{0.04}_{-0.03}$(syst.)	
Combined	669	1084	4184	4033
	1.62 ± 0.22(fit stat.) ± 0.20(MC stat.) $^{0.38}_{-0.20}$(syst.)		0.96 ± 0.11(fit stat.) ± 0.06(MC stat.) $^{0.22}_{-0.00}$(syst.)	

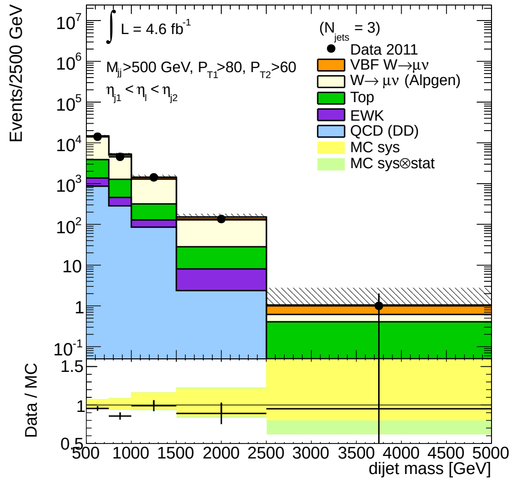
Table 6.13: The signal strength and number of events expected and fitted for each channel for two alternative fit ranges.

6.4 Summary

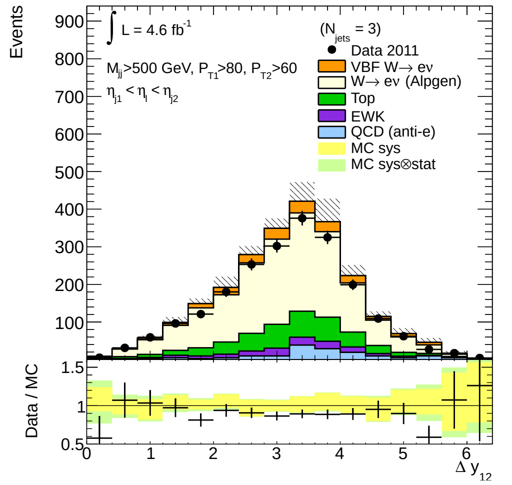
A maximum-likelihood fit was carried out to extract the number of signal events from the M_{jj} distribution in electron, muon and combined channels. The systematic uncertainties were estimated using a frequentist method and the dominant effects were forward jet energy scale uncertainty, jet matching uncertainty and Monte Carlo and fit statistics. The background only hypothesis was rejected with a significance of 4.89 (4.36 expected). The cross section was measured to be $\sigma = 325 \pm 6$ (lumi) ± 32 (stat) $^{+63}_{-70}$ (syst) ± 86 (theo) fb in the fiducial region lepton p_{T} $p_T^{\ell} > 25$ GeV, lepton pseudo-rapidity$|\eta^{\ell}| < 2.4$, missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV, transverse mass $m_T > 40$ GeV, jet rapidity $|y| < 4.4$, leading jet $p_T^{j1} > 80$ GeV, second leading jet $p_T^{j2} > 60$ GeV, no third jet satisfying $p_T^{j3} > 30$ GeV, dijet mass $M_{jj} > 500$ GeV, and outside lepton vet $\eta^{ja} < \eta^{\ell} < \eta^{jb}$. A difference of $\sim 50\%$ is observed between the electron and muon cross section, though this is within the one sigma uncertainty of the measurement.



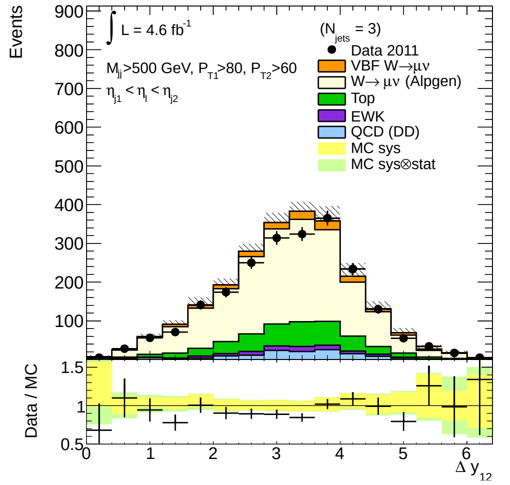
(a) Electron channel M_{jj}



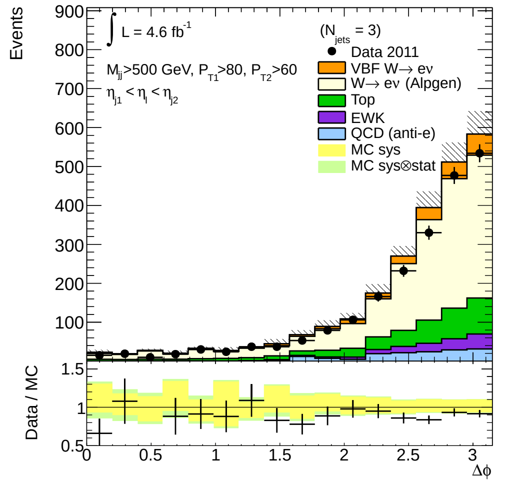
(b) Muon channel M_{jj}



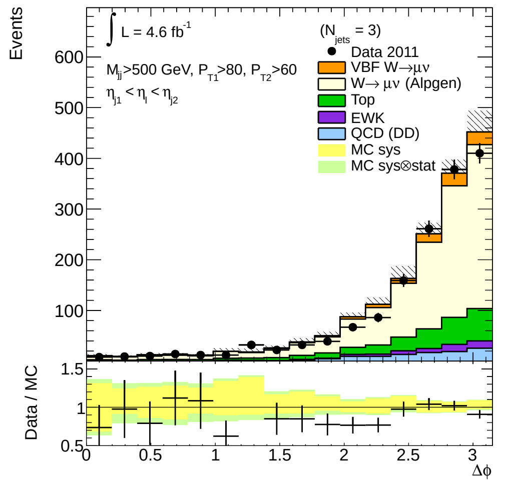
(c) Electron channel $\Delta y(j_1, j_2)$



(d) Muon channel $\Delta y(j_1, j_2)$

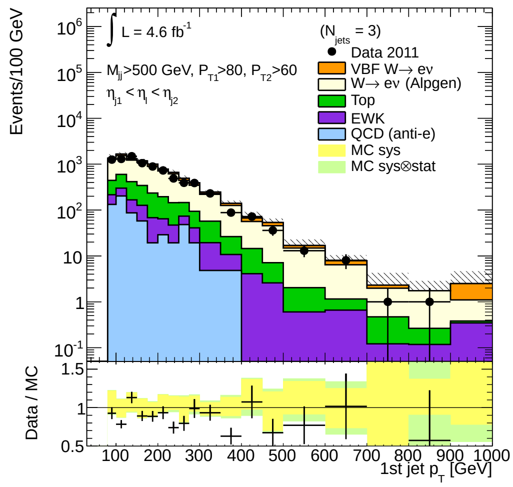


(e) Electron channel $\Delta\phi(j_1, j_2)$

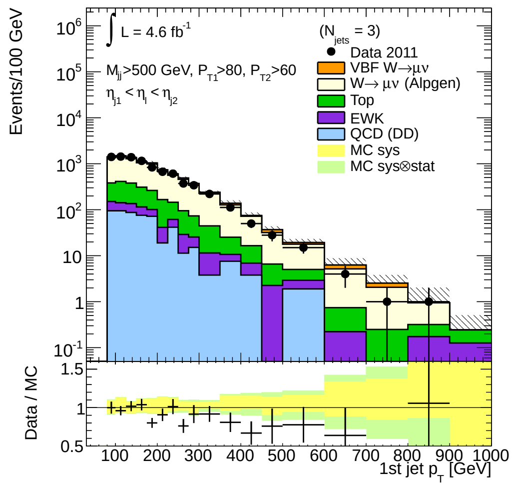


(f) Muon channel $\Delta\phi(j_1, j_2)$

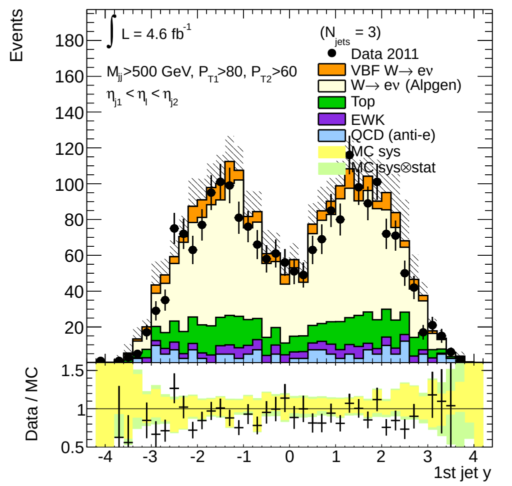
Figure 6.16: Validation region plots for dijet variables in the electron and muon channels. The $W+$ EWK jets and $W+$ QCD jets have been scaled by the signal strength values extracted in the signal region.



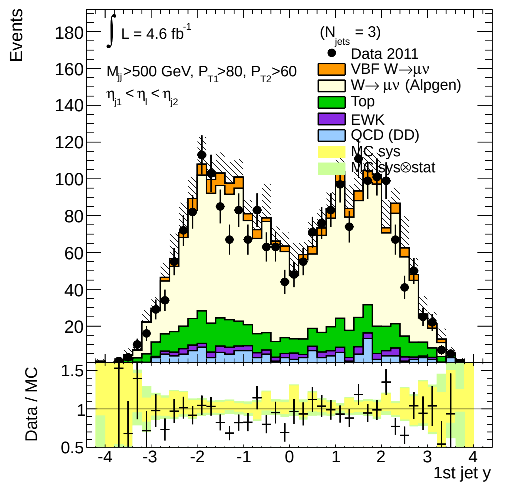
(a) Electron leading jet p_{T}



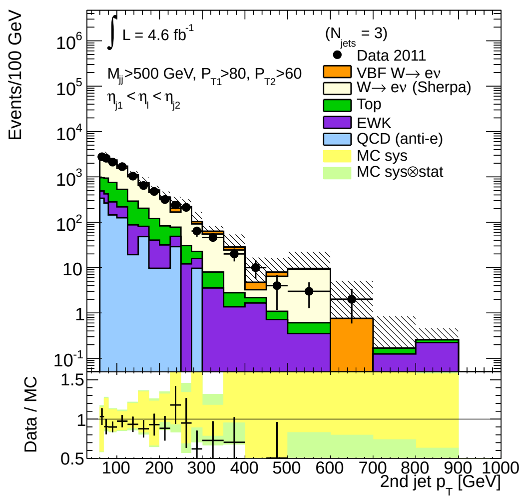
(b) Muon leading jet p_{T}



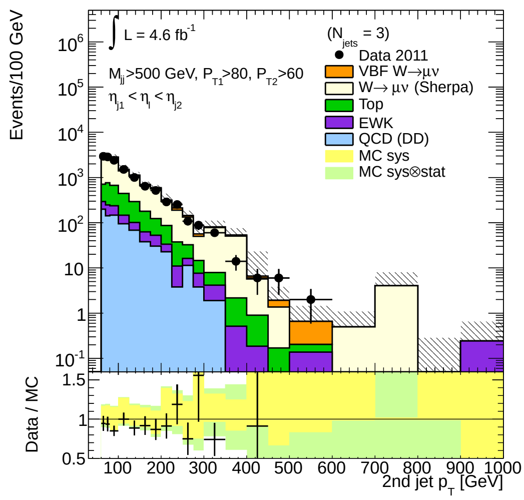
(c) Leading jet rapidity (electron channel)



(d) Leading jet rapidity (muon channel)

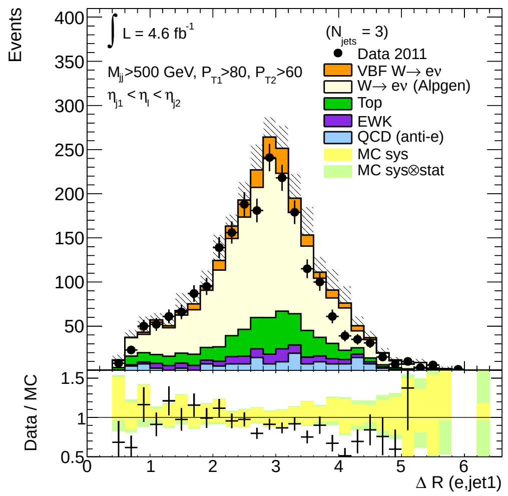


(e) Electron jet $2^{nd} p_{\mathrm{T}}^{jet}$

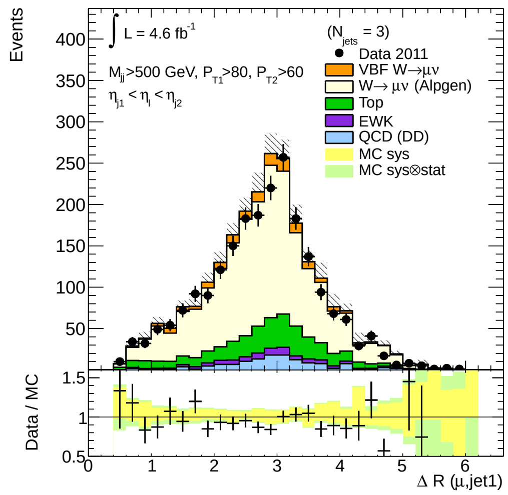


(f) Muon jet $2^{nd} p_{\mathrm{T}}^{jet}$

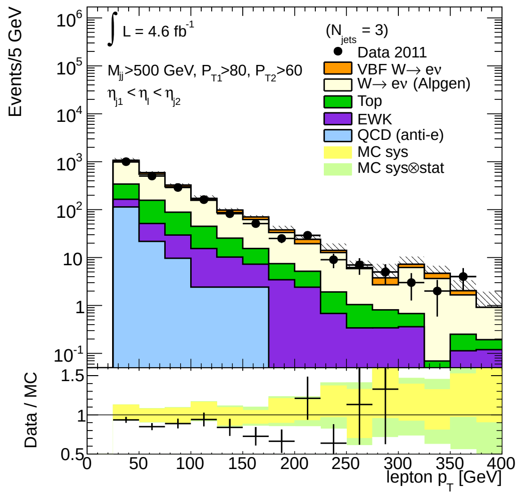
Figure 6.17: Validation region plots for jet variables in the electron and muon channels. The $W +$ EWK jets and $W +$ QCD jets have been scaled by the signal strength values extracted in the signal region.



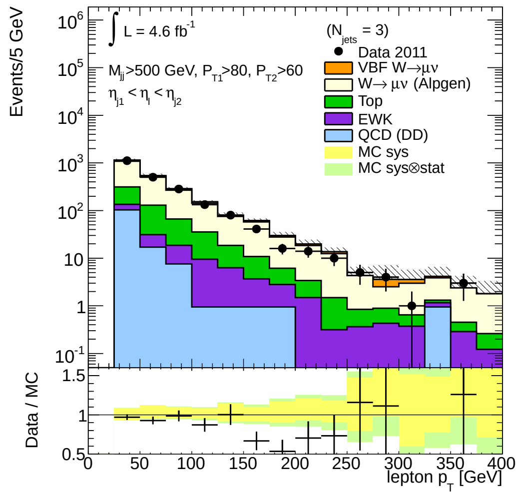
(a) Electron $\Delta R(\ell, j_1)$



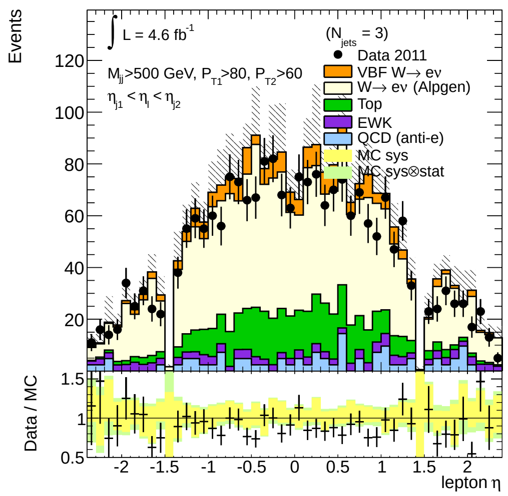
(b) Muon $\Delta R(\ell, j_1)$



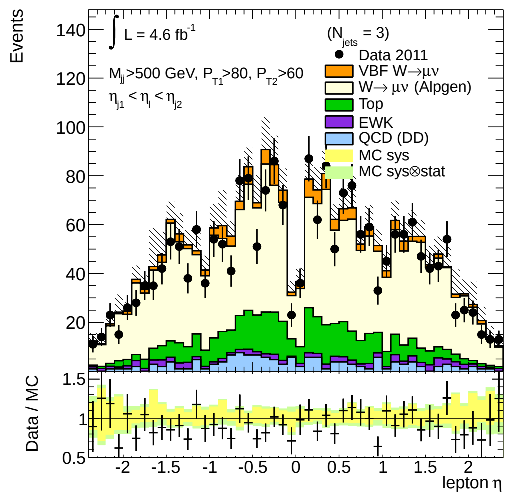
(c) Electron p_{T}



(d) Muon p_{T}

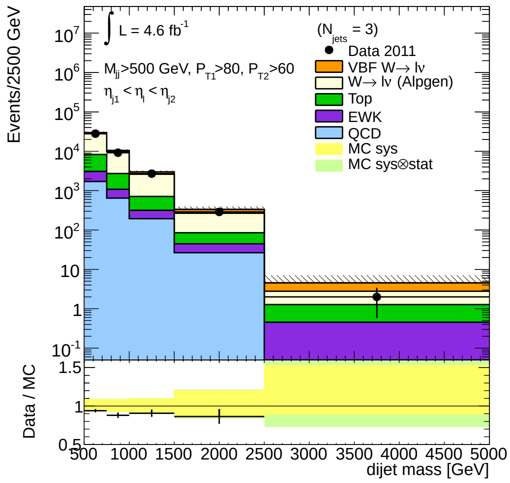


(e) Electron η

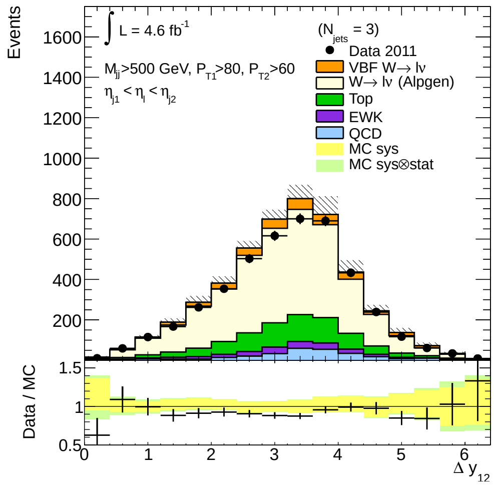


(f) Muon η

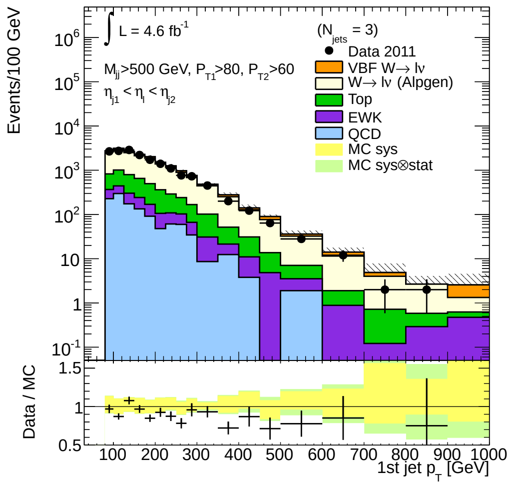
Figure 6.18: Validation region plots for lepton variables in the electron and muon channels. The $W +$ EWK jets and $W +$ QCD jets have been scaled by the signal strength values extracted in the signal region.



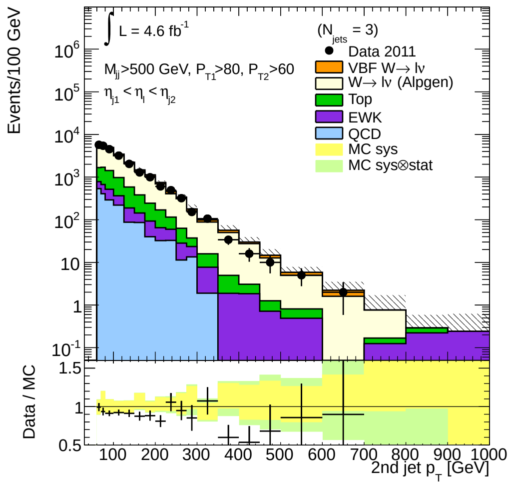
(a) Combined channel M_{jj}



(b) Combined channel $\Delta y(j_1, j_2)$

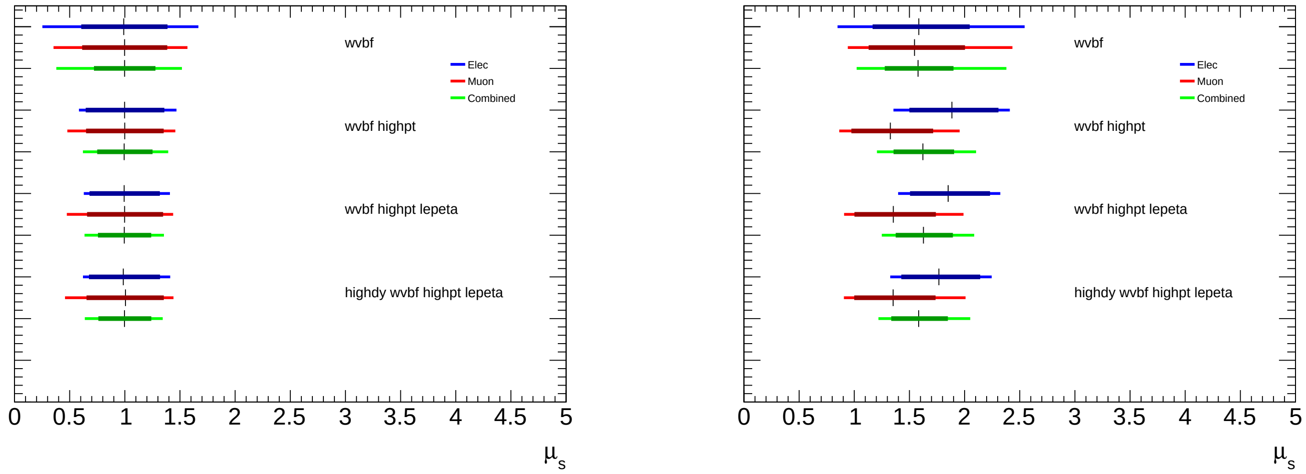


(c) Combined channel leading jet p_{T}



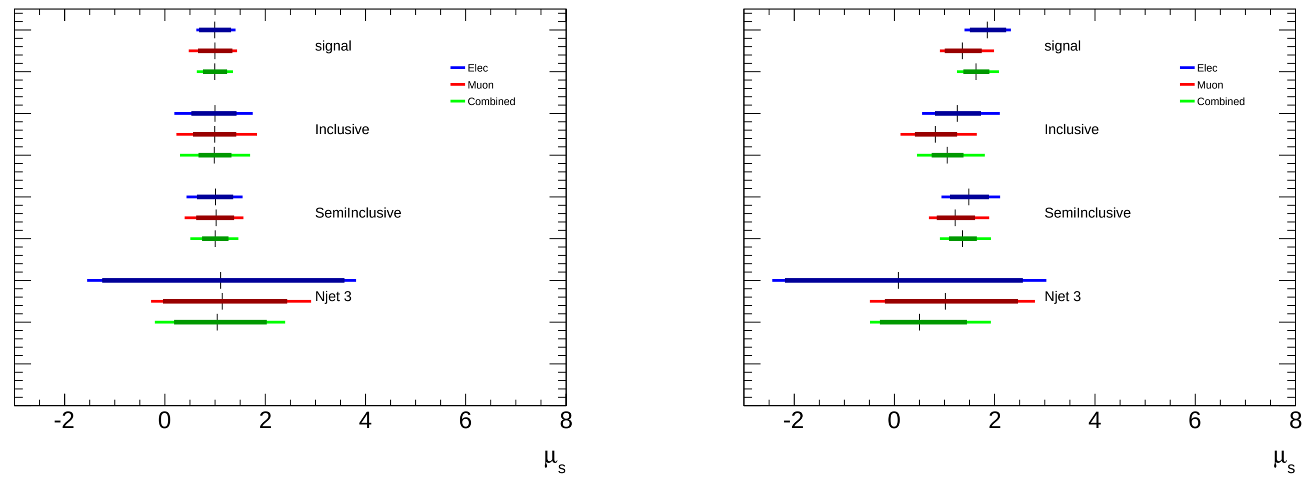
(d) Combined channel sub leading jet p_{T}

Figure 6.19: Validation region plots in the combined electron and muon channels. The $W+$ EWK jets and $W+$ QCD jets have been scaled by the signal strength values extracted in the signal region.



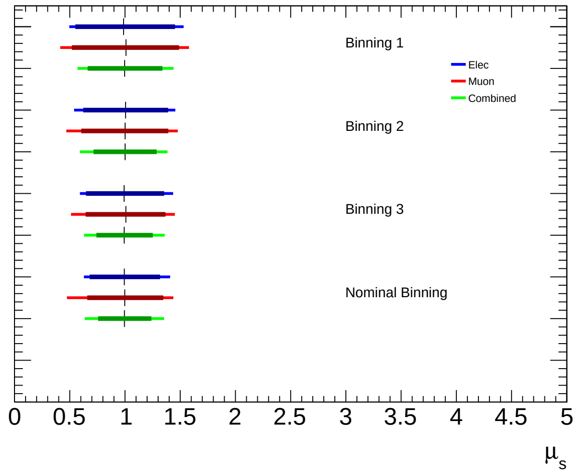
(a) Expected (b) Observed

Figure 6.20: The variation in the result with changes in cuts in expected and observed cases for the cuts applied in the signal region. The dark band is the Monte Carlo and fit statistical uncertainty while the full band is the sum of these in quadrature with other experimental systematics.

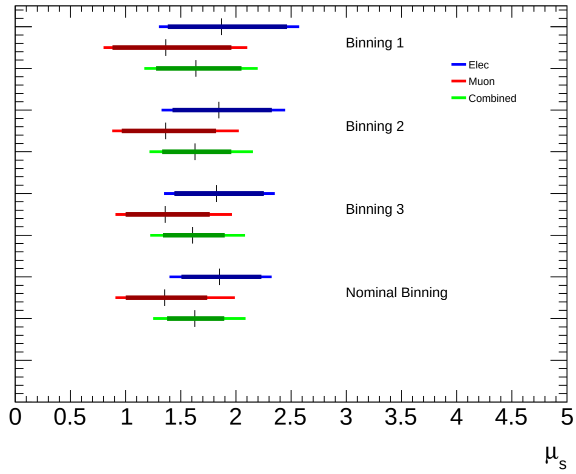


(a) Expected (b) Observed

Figure 6.21: The variation in the result with changes in N_{jets} cuts in expected and observed cases. The dark band is the Monte Carlo and fit statistical uncertainty while the full band is the sum of these in quadrature with other experimental systematics.

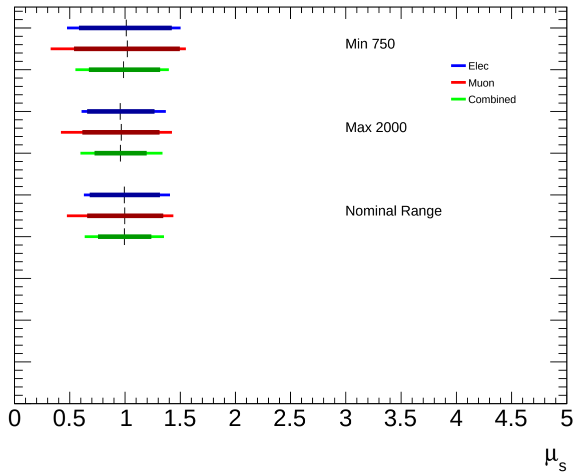


(a) Expected

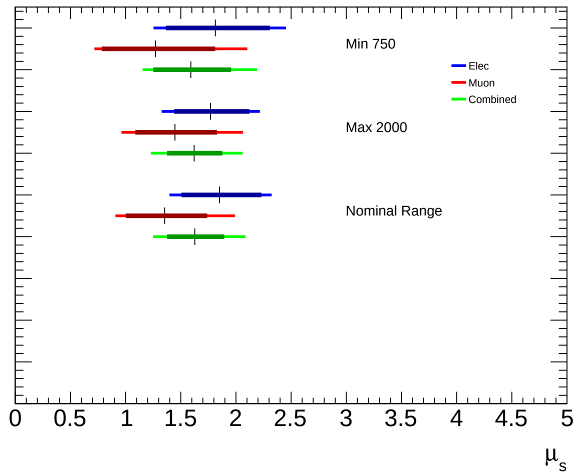


(b) Observed

Figure 6.22: Variations in the results due to changes in the binning in expected and observed cases.
The dark band is the Monte Carlo and fit statistical uncertainty while the full band is the sum of these in quadrature with other experimental systematic uncertainties.



(a) Expected



(b) Observed

Figure 6.23: Variation in the result due to changes in fit range in expected and observed cases. The dark band is the Monte Carlo and fit statistical uncertainty while the full band is the sum of these in quadrature with other experimental systematics.

CHAPTER 7

Conclusions

The work presented in this thesis has demonstrated for the first time, that it is possible to measure $W+$ EWK jets at a hadron collider. From 4.64 fb^{-1} of data taken during the 2011 LHC run I have shown it is possible to reject the background only hypothesis with a significance of 4.89σ with a value of 4.36σ expected.

A fit of the M_{jj} spectrum was used to extract the cross section of $\sigma = 325 \pm 6$ (lumi) ± 32 (stat) $^{+63}_{-70}$ (syst) ± 86 (theo) fb which is comparable to the result of 425 ± 21 fb obtained from VBF@NLO . Applying the scale factors derived from this fit improves agreement in all kinematic distributions.

This analysis has prepared the path for a more detailed analysis to be conducted with the 20 fb^{-1} of data collected at ATLAS in 2012. With the larger data set and a larger analysis team it is expected to publish unfolded distributions in addition to measuring the cross section more precisely. This should allow detailed investigations of the differential cross section and test the limits of Monte Carlo simulation of this phase space. In addition limits on anomalous triple gauge couplings will be extracted.

The current analysis is limited in several respects and it is hoped that these may be improved in the 2012 analysis.

The Monte Carlo samples used for the 2011 analysis were limiting in three important respects. The first is that the signal extraction fit is very sensitive to the lack of Monte Carlo events in the high M_{jj} tails. For the 2012 analysis a Monte Carlo filter has been developed which retains high M_{jj} events at an enhanced rate.

A second weakness of the Monte Carlo strategy is that the signal sample contained diboson production. This required an analysis cut of $M_{jj} > 500$ GeV to remove these contributions. For 2012 a new version of SHERPA is available which allows the removal of diboson production from the matrix element. This will allow the fit to be carried out over the full M_{jj} range, improving the statistical power of the fit.

The third limitation of the Monte Carlo used for this analysis is the large sensitivity to jet matching, which is the largest contribution to the theoretical uncertainty. In the future this effect should be investigated more closely using SHERPA, which uses a different matching scheme. Additionally $W + 2$ QCD jets is now available at next-to-leading-order in QCD matched to a parton shower [146], and can therefore be produced as a fully simulated Monte Carlo sample. Since this is a fixed order calculation there is no possibility of double counting emissions from the parton shower and matrix element calculation, and so no matching uncertainty.

The measurement presented in this thesis is experimentally limited by the forward component of the jet energy scale uncertainty. Measurements using the 2012 data set may benefit from improvements in the method used to determine the uncertainty, such as using the increased rate of Z production to do Z-jet balance in the forward region. The Z decays to leptons which are very well measured, and so this should allow a more precise measurement of the JES in the forward region.

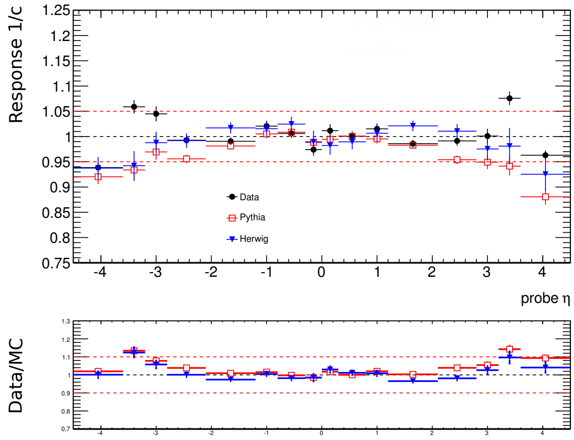
In addition it may be possible to construct a control region close to the signal region where the $W +$ QCD jets background can be estimated using a data-driven technique. This background would therefore be constrained by a data measurement and the uncertainty on the magnitude and shape of this background would be due to the extrapolation from the control region to the signal region, rather than directly due to the forward JES. This is potentially smaller and thus provides a route to an improved result.

Appendices

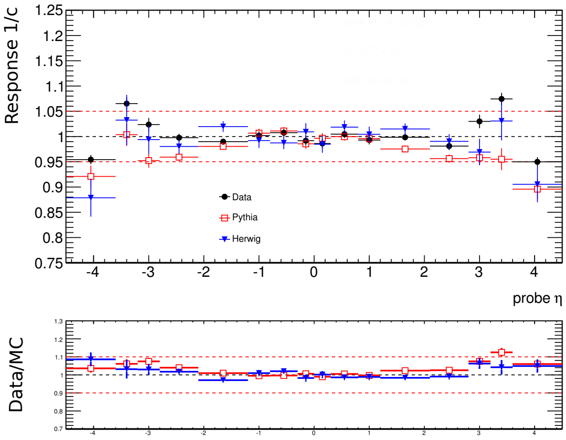
APPENDIX A

LC Topo η-intercalibration results plots

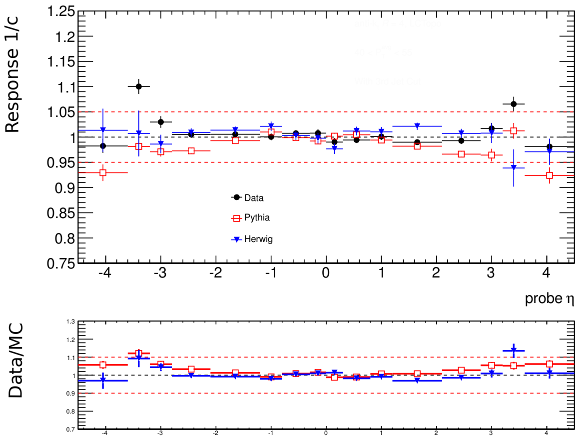
The results shown in Section 3.4.1 are repeated here using the LCTopo calibration and cone size 0.4. The response is very similar to that found for the TopoEM calibration with the largest differences seen in the $3.2 < |\eta_{det}| < 3.6$ bin.



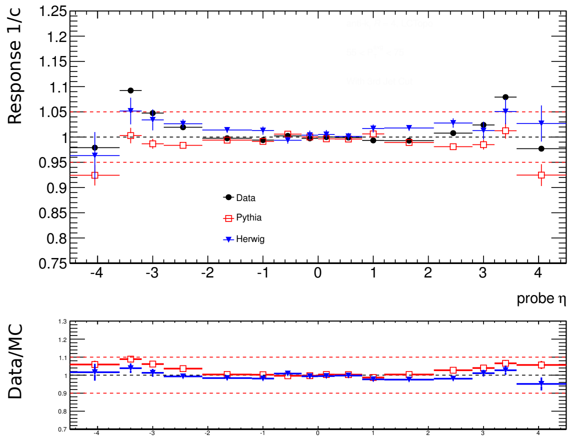
(a) $22 < p_T^{\mathrm{avg}} < 30$ GeV



(b) $30 < p_T^{\mathrm{avg}} < 40$ GeV

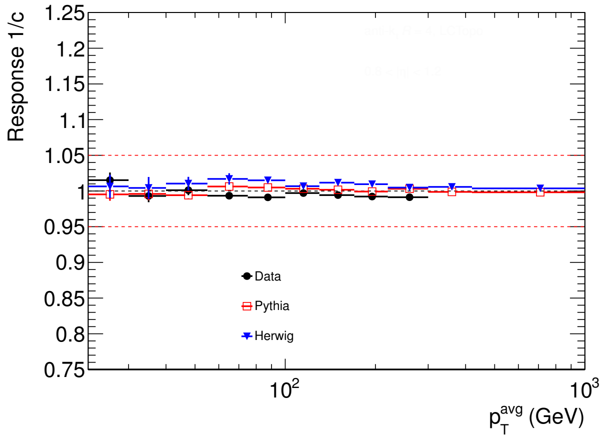


(c) $40 < p_T^{\mathrm{avg}} < 55$ GeV

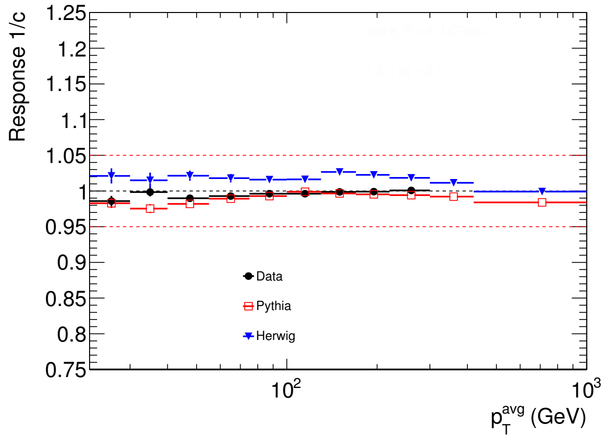


(d) $55 < p_T^{\mathrm{avg}} < 75$ GeV

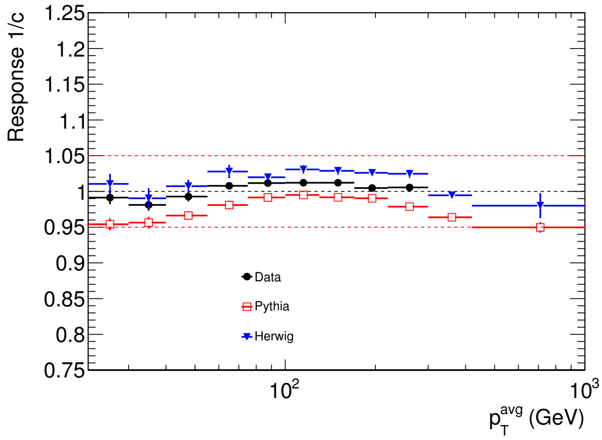
Figure A.1: The response, $1/c_i$, as a function of η_{det} for the four lowest p_T^{avg} bins. The dotted lines indicate response of 1 ± 0.05 in the main plots and 1 ± 0.1 in the ratio. The error bars are statistical only.



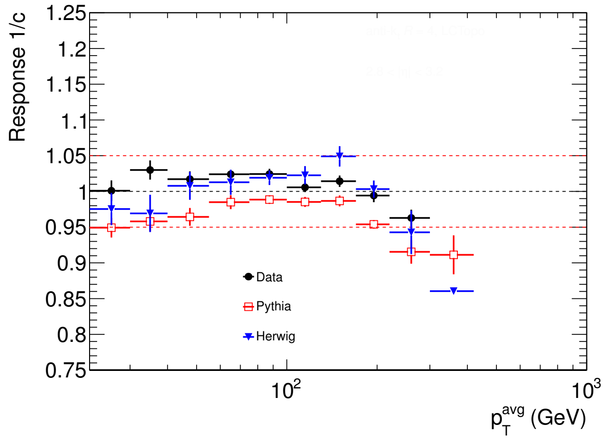
(a) $0.8 < \eta_{det} < 1.2$



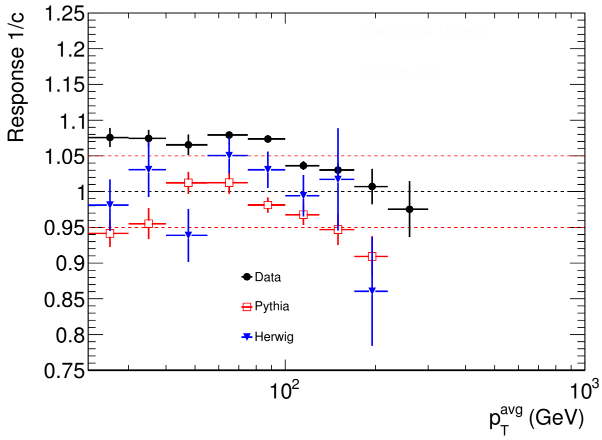
(b) $1.2 < \eta_{det} < 2.1$



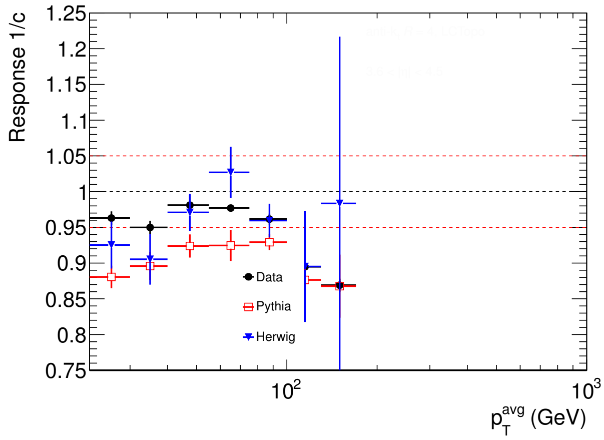
(c) $2.1 < \eta_{det} < 2.8$



(d) $2.8 < \eta_{det} < 3.2$



(e) $3.2 < \eta_{det} < 3.6$



(f) $3.6 < \eta_{det} < 4.5$

Figure A.2: Response correction as a function of p_T^{avg}. Each plot shows a different η_{det} region. The dotted lines indicate response of 1 ± 0.05 in the main plots.

APPENDIX B

Sherpa run cards

B.1 Default settings

The Atlas settings for Sherpa are given below with an explanation.

- `MASS[6]=172.5` — top quark mass
- `MASS[23]=91.1876` — Z boson mass
- `MASS[24]=80.399` — W boson mass
- `WIDTH[23]=2.4952` — Z boson width
- `WIDTH[24]=2.085` — W boson width
- `SIN2THETAW=0.23113` — weak mixing angle

B.2 Signal sample

Singal samples were generated using SHERPA v1.3.1. An example run card for the electron channel is given before. The relevant JobOptions files are given in Table B.1.

```
(processes){
!positron
Process 93 93  -> -11 12 93 93 93{1}
Order_EW 4
Integration_Error 5.e-2
CKKW  sqr(15/E_CMS)
End process;
!electron
Process 93 93  -> 11 -12 93 93 93{1}
Order_EW 4
Integration_Error 5.e-2
CKKW  sqr(15/E_CMS)
End process;
}(processes)
(selector){
Mass -11 12 10 E_CMS
Mass 11 -12 10 E_CMS
NJetFinder  2  15.0  0.0  0.4 1
}(selector)
(model){
ACTIVE[6]=0
}(model)
```

Sample	JobOption
$We\nu$	MC11.126569.SherpaVBFWenu.py
$W \to \mu\nu$	MC11.126570.SherpaVBFWmunu.py
$W \to \tau\nu$	MC11.126571.SherpaVBFWtaunu.py

Table B.1: JobOptions files for Sherpa signal files

B.3 Background samples

The QCD-W + jets Sherpa sample is a standard Atlas sample and its run card is given here for comparison to the above singal sample shown above. This is taken from MC11.147774.Sherpa_CT10_Wenu.py.

```
(processes){
  Process 93 93 -> 90 91 93{5}
  Order_EW 2
  CKKW sqr(20/E_CMS)
  NLO_QCD_Part BVIRS {2};
  ME_Generator Amegic {2};
  Integration_Error 0.05 {6,7,8}
  Scales LOOSE_METS{MU_F2}{MU_R2} {6,7,8}
  Enhance_Factor 2.0 {3}
  Enhance_Factor 5.0 {4,5,6,7,8}
  End process;
}(processes)

(selector){
  Mass 90 91 1.7 E_CMS
}(selector)

(run){
  NLO_Mode 3
  MASSIVE[13]=1
```

```
  MASSIVE[15]=1

}(run)
```

APPENDIX C

Data period summary

Table C.1 presents a summary of the period information for data used in this thesis, extracted from the COMA meta-data interace webpage [147]. Periods A-C were not included as many runs were conducted without the toroid field on, and the LHC was running with 75 ns bunch spacing.

Period name	Description of conditions	Date range (YY-Month-Date)	Run range	Number of runs	Stable lumi (pb^{-1})	Ready lumi (pb^{-1})	Max inst. lumi (10^{30} cm^2 s^{-1})	Max. μ
D	50 ns bunch train running	11-Apr-14: 11-Apr-29	179710: 180481	23	186	182 (98%)	659	7.3
E	LAr problem in EMBA: 6 FEB's not operational	11-Apr-30: 11-May-03	180614: 180776	5	53	52 (98%)	832	7.6
F	50ns data after the technical stop. For Tier0 reconstruction improved conditions were applied to correct for the EMEC sagging and for the tilt of the ID w.r.t. the solenoid field.	11-May-15: 11-May-25	182013: 182519	17	160	156 (98%)	1100	8
G	new trigger cache which uses pile up noise suppression	11-May-27: 11-Jun-14	182726: 183462	28	572	566 (99%)	1263	7.9
H	update to L1 muon firmware to fix problem with misconfiguration affecting A-side only.	11-Jun-16: 11-Jun-28	183544: 184169	13	287	283 (99%)	1264	6.8
I	data after the July technical stop: the 4 LAr FEBs in Layer2 are recovered!	11-Jul-13: 11-Jul-29	185353: 186493	27	416	406 (97%)	1887	9.1
J	LHC increasing beam currents adiabatically. Changes to primary high p_T triggers in trigger menu in reaction to this.	11-Jul-30: 11-Aug-04	186516: 186755	9	240	237 (98%)	1995	9.6
K	new trigger menu with many changes (so-called 3e33 menu)	11-Aug-04: 11-Aug-22	186873: 187815	19	685	676 (99%)	2328	11
L	new Physics_pp_v3 menu is deployed, which includes variable η cuts for electrons at L1 with hadronic veto (L1_EMxxVH), new L1_MU11 road and L1_MU0→L1_MU4 change	11-Sep-07: 11-Oct-05	188902: 190343	59	1625	1599 (98%)	3252	16
M	no known special feature	11-Oct-06: 11-Oct-30	190503: 191933	48	1184	1160 (98%)	3848	32

Table C.1: Summary of data period information.

Glossary

`EM+JES` EM calibrated clusters plus jet energy scale. 60, 192

`LCW+JES` Local cluster weighted clusters plus jet energy scale. 60, 192

VBF Vector boson fusion. 192

aTGC anaomalous triple gauge coupling. 192

b-layer The pixel layer closest to the beam pipe. 192

CSC Cathode Strip Chambers. 192

DQ Data quality. 192

EF Event Filter. 39, 40, 53, 192

$E_{\mathrm{T}}^{\mathrm{miss}}$ Missing transverse energy. 80, 192

FCal LAr Forward Calorimeter. 192

FSR Final state radiation. 192

HEC LAr Hadronic Endcap. 192

HLT High Level Trigger. 39, 40, 46, 192

ID Identification. 192

ID Inner Detector. 28, 29, 53, 54, 80, 192

interaction length (λ) is the distance a hadron must travel to lose $(1 - e^{-1}) \cdot E_0$ where E_0 is its initial energy.is the distance in a given material that an electron must travel to lose $(1 - e^{-1}) \cdot E_0$ where E_0 is its initial energy. 192

ISR Initial state radiation. 192

JER Jet energy resolution. 78, 192

JES Jet energy scale. 58, 61, 80, 192

Jet vertex fraction The JVF is defined as the sum p_T of the tracks associated to the leading primary vertex divided by the sum p_T of the tracks associated with the jet. This is 1 for a jet with all tracks associated to the leading primary vertex and 0 for a jet with none of its track associated to the leading primary vertex. Jets with no tracks are assigned a JVF of -1. 192

JVF Jet vertex fraction. 65, 97, 192, 193

L1 Level 1 Trigger. 39, 45, 46, 52, 53, 192

L2 Level 2 Trigger. 39, 40, 53, 192

LAr liquid argon. 42, 49, 51, 192

lAr hole A defect in the liquid argon calorimeter. 192

LCW local cluster weighting. 58, 80, 81, 192

leading primary vertex The primary vertex with highest $\sum p_T$ of all primary verticies in the event. 59, 65, 192, 193

LEP Large Electron Positron Collider. 23, 192

LHC Large Hadron Collider. 1, 23, 192

LO Leading order. 192

MC Monte Carlo. 192

MDT Monitored Drift Tubes. 192

MPI Multiple-parton-interaction. 192

MS Muon Spectrometer. 54, 80, 192

$\boldsymbol{\mu}$ The average number of interactions per bunch crossing. 25, 26, 38, 58, 192

NLO Next-to-leading order. 89, 192

NNLO Next-to-next-to-leading order. 89, 90, 192

$\boldsymbol{N_{PV}}$ The number of primary verticies. 192

PDFs Parton distribution functions. 10, 62, 192

primary vertex A vertex close to the beam spot, in constrast to a secondary vertex which is a result of a decay in flight. 192

QCD Quantum chromodynamics. 192

QED Quantum electrodynamics. 192

radiation length (X_0) is the distance in a given material that an electron must travel to lose $(1 - e^{-1}) \cdot E_0$ where E_0 is its initial energy. 192

RoI Region of interest. 53, 192

RPC Resistive Plate Chambers. 52, 192

SCT Semiconductor Tracker. 32, 44, 45, 192

TGC Thin Gap Chambers. 52, 192

TGCs triple gauge couplings. 192

Tile-Cal Tile Calorimeter. 192

TRT Transition Radiation Tracker. 32, 44, 45, 192

References

[1] D. Binosi and L. Theul, *JaxoDraw: A graphical user interface for drawing Feynman diagrams*, Computer Physics Communications **161** no. 12, (2004) 76 – 86. `http://www.sciencedirect.com/science/article/pii/S0010465504002115`.

[2] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory.* Westview Press, 1995.
T. Cheng and L. Li, *Gauge Theory of Elementary Particle Physics.* Oxford University Press, 1984.
F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics.* John Wiley & Sons, 1984.

[3] "Standard model of elementary particles." `http://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg:accessed20-2-2013`.

[4] Particle Data Group Collaboration, J. Beringer *et al.*, *Review of Particle Physics (RPP)*, Phys. Rev. **D86** (2012) 010001.

[5] D. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. 1*, Phys. Rev. **D8** (1973) 3633–3652.
H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346–1349.

[6] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588.
A. Salam and J. C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. **13** (1964) 168–171.
S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266. `http://link.aps.org/doi/10.1103/PhysRevLett.19.1264`.

[7] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.
P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132–133.
G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587.

T. Kibble, *Symmetry breaking in non-Abelian gauge theories*, Phys. Rev. **155** (1967) 1554–1561.

[8] MuLan Collaboration, D. Webber *et al.*, *Measurement of the Positive Muon Lifetime and Determination of the Fermi Constant to Part-per-Million Precision*, Phys. Rev. Lett. **106** (2011) 041803, `arXiv:1010.0991 [hep-ex]`.

[9] ALEPH, CDF, D0, DELPHI, L3, OPAL , SLD, LEP Electroweak Working Group, Tevatron Electroweak Working Group, SLD Electroweak and Heavy Flavour Collaboration, *Precision Electroweak Measurements and Constraints on the Standard Model*, `arXiv:1012.2367 [hep-ex]`.

[10] P. J. Mohr, B. N. Taylor, and D. B. Newell, *CODATA Recommended Values of the Fundamental Physical Constants: 2010*, Rev. Mod. Phys. **84** (2012) 1527–1605, `arXiv:1203.5425 [physics.atom-ph]`.

[11] CDF Collaboration, *Precise Measurement of the W-Boson Mass with the CDF II Detector*, Phys. Rev. Lett. **108** (2012) 151803. `http://link.aps.org/doi/10.1103/PhysRevLett.108.151803`.

[12] CDF, D0 Collaboration, *Combination of the top-quark mass measurements from the Tevatron collider*, Phys. Rev. **D86** (2012) 092003, `arXiv:1207.1069 [hep-ex]`.

[13] J. M. Campbell, J. Huston, and W. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, Rept. Prog. Phys. **70** (2007) 89, `arXiv:hep-ph/0611148 [hep-ph]`.

[14] V. Gribov and L. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438–450.
Y. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov.Phys.JETP **46** (1977) 641–653.
G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nuc. Phys. B **126** no. 2, (1977) 298 – 318. `http://www.sciencedirect.com/science/article/pii/0550321377903844`.

[15] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189–285, `arXiv:0901.0002 [hep-ph]`.

[16] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, *et al.*, *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, `arXiv:1007.2241 [hep-ph]`.

[17] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, *et al.*, *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244–289, `arXiv:1207.1303 [hep-ph]`.

[18] J. C. Collins and D. E. Soper, *The Theorems of Perturbative QCD*, Annual Review of Nuclear and Particle Science **37** no. 1, (1987) 383–409. `http://www.annualreviews.org/doi/abs/10.1146/annurev.ns.37.120187.002123`.

[19] ATLAS Collaboration, G. Aad *et al.*, *Measurement of hard double-parton interactions in $W \rightarrow \ell\nu + 2$ jet events at $\sqrt{s} = 7$ TeV with the ATLAS detector*, New J.Phys. **15** (2013) 033038, `arXiv:1301.6872 [hep-ex]`.

[20] T. Sjostrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026.

[21] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, *et al.*, *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **0101** (2001) 010,

arXiv:hep-ph/0011363 [hep-ph].
[22] J. Alwall, S. Hoche, F. Krauss, N. Lavesson, L. Lonnblad, *et al.*, *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, Eur. Phys. J. **C53** (2008) 473–500, arXiv:0706.2569 [hep-ph].
[23] M. Mangano *et al.*, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, arXiv:hep-ph/0206293 [hep-ph].
[24] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, *et al.*, *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph].
[25] S. Catani, F. Krauss, R. Kuhn, and B. Webber, *QCD matrix elements + parton showers*, JHEP **0111** (2001) 063, arXiv:hep-ph/0109231 [hep-ph].
[26] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **0411** (2004) 040, arXiv:hep-ph/0409146 [hep-ph].
[27] S. Frixione and B. R. Webber, *The MC and NLO 3.4 Event Generator*, arXiv:0812.0770 [hep-ph].
[28] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Kennedy, *et al.*, *The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC*, Eur. Phys. J. **C72** (2012) 2205, arXiv:1209.2716 [hep-ph].
[29] ATLAS Collaboration, "Atlas standard model cross-sections." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/CombinedSummaryPlots:accessed7-4-2013.
[30] LEP Working Group for Higgs boson searches, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, R. Barate *et al.*, *Search for the standard model Higgs boson at LEP*, Phys.Lett. **B565** (2003) 61–75, arXiv:hep-ex/0306033 [hep-ex].
[31] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].
CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** no. 1, (2012) 30 – 61. http://www.sciencedirect.com/science/article/pii/S0370269312008581.
[32] ATLAS Collaboration, G. Aad *et al.*, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, arXiv:1307.1432 [hep-ex]. submitted to Phys. Lett. B.
[33] *Properties of the observed Higgs-like resonance using the diphoton channel*, Tech. Rep. CMS-PAS-HIG-13-016, CERN, Geneva, 2013.
[34] ATLAS Collaboration, G. Aad *et al.*, *Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC*, arXiv:1307.1427 [hep-ex]. submitted to Phys. Lett. B.
CMS Collaboration, *Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV*, Tech. Rep. CMS-PAS-HIG-13-005, CERN, Geneva, 2013.
[35] *Differential cross sections of the Higgs boson measured in the diphoton decay channel using 8 TeV pp collisions*, Tech. Rep. ATLAS-CONF-2013-072, CERN, Geneva, Jul, 2013.
[36] CMS Collaboration, *Evidence for a particle decaying to W^+W^- in the fully leptonic final state in a standard model Higgs boson search in pp collisions at*

the LHC, Tech. Rep. CMS-PAS-HIG-12-042, CERN, Geneva, 2012.

[37] D0 Collaboration, *Studies of W boson plus jets production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, `arXiv:1302.6508 [hep-ex]`.
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, Phys. Rev. D **77** (2008) 011108. `http://link.aps.org/doi/10.1103/PhysRevD.77.011108`.

[38] 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. **B698** (2011) 325–345, `arXiv:1012.5382 [hep-ex]`.

[39] 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*, Tech. Rep. STDM-2012-24, Geneva, 2013. Paper in preparation.

[40] CMS Collaboration, *Jet Production Rates in Association with W and Z Bosons in pp Collisions at $\sqrt{s} = 7$ TeV*, JHEP **1201** (2012) 010, `arXiv:1110.3226 [hep-ex]`.

[41] ATLAS Collaboration, *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*, Phys. Lett. **B708** (2012) 221–240, `arXiv:1108.4908 [hep-ex]`.

[42] ATLAS Collaboration, *A measurement of the ratio of W to Z production cross section in association with jets with the ATLAS detector*, Tech. Rep. STDM-2012-25, 2013. Paper in preparation.

[43] D0 Collaboration, *Measurement of trilinear gauge boson couplings from $WW + WZ \to l\nu jj$ events in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$* TeV, Phys. Rev. D **80** (2009). `http://link.aps.org/doi/10.1103/PhysRevD.80.053012`.

[44] CDF Collaboration, *Invariant Mass Distribution of Jet Pairs Produced in Association with a W boson in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. Lett. **106** (2011) 171801, `arXiv:1104.0699 [hep-ex]`.

[45] D0 Collaboration, *Bounds on an anomalous dijet resonance in W+jets production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. Lett. **107** (2011) 011804, `arXiv:1106.1921 [hep-ex]`.
CMS Collaboration, *Study of the Dijet Mass Spectrum in $pp \to W$+jets Events at $\sqrt{s}$=7* TeV, Phys. Rev. Lett. **109** (2012) 251801. `http://link.aps.org/doi/10.1103/PhysRevLett.109.251801`.
ATLAS Collaboration, G. Aad *et al.*, *Search for resonant diboson production in the $\ell\nu jj$ decay channels with the ATLAS detector at 7 TeV*, Phys. Rev. **D87** (2013) 112006, `arXiv:1305.0125 [hep-ex]`.

[46] M. Trovato, *Update on dijet mass spectrum in W + 2 jets events*, Feb, 2013. `http://theory.fnal.gov/jetp/talks/Feb222013WineAndCheeseSeminar.pdf`. Fermi Lab Wine and Cheese seminar, accessed 21-4-2013.

[47] F. Boudjema, B. Mele, E. Accomando, S. Ambrosanio, A. Ballestrero, *et al.*, *Standard model processes*, `arXiv:hep-ph/9601224 [hep-ph]`.

[48] *Invariant mass distribution of jet pairs produced in association with a leptonically decaying W boson using* 1.02 fb^{-1} *of ATLAS data*, Tech. Rep. ATLAS-CONF-2011-097, CERN, Geneva, Jul, 2011.

[49] K. Arnold, J. Bellm, G. Bozzi, M. Brieg, F. Campanario, *et al.*, *VBFNLO: A Parton Level Monte Carlo for Processes with Electroweak Bosons – Manual for Version 2.5.0*, `arXiv:1107.4038 [hep-ph]`.

[50] H. Chehime and D. Zeppenfeld, *Single W and Z boson production as a probe for rapidity gaps at the SSC*, Phys. Rev. **D47** (1993) 3898–3905.

[51] D. Rainwater, R. Szalapski, and D. Zeppenfeld, *Probing color-singlet exchange in $Z + 2$-jet events at the CERN LHC*, Phys. Rev. D **54** (1996) 6680–6689. `http://link.aps.org/doi/10.1103/PhysRevD.54.6680`.

[52] D. L. Rainwater, D. Zeppenfeld, and K. Hagiwara, *Searching for $H \to \tau^+\tau^-$ in weak boson fusion at the CERN LHC*, Phys. Rev. **D59** (1998) 014037, `arXiv:hep-ph/9808468 [hep-ph]`.
T. Plehn, D. L. Rainwater, and D. Zeppenfeld, *A Method for identifying $H \to \tau^+\tau^- \to e^\pm\mu^\mp p_T$ at the CERN LHC*, Phys. Rev. **D61** (2000) 093005, `arXiv:hep-ph/9911385 [hep-ph]`.

[53] V. Khoze, M. Ryskin, W. Stirling, and P. Williams, *A Z-monitor to calibrate Higgs production via vector boson fusion with rapidity gaps at the LHC*, The European Physical Journal C - Particles and Fields **26** (2003) 429–440. `http://dx.doi.org/10.1140/epjc/s2002-01069-2`.

[54] ATLAS Collaboration, *Measurement of dijet production with a veto on additional central jet activity in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, JHEP **1109** (2011) 053, `arXiv:1107.1641 [hep-ex]`.

[55] D. L. Rainwater and D. Zeppenfeld, *Observing $H \to W^*W^* \to e^\pm\mu^\mp p_T(miss)$ in weak boson fusion with dual forward jet tagging at the CERN LHC*, Phys. Rev. **D60** (1999) 113004, `arXiv:hep-ph/9906218 [hep-ph]`.
N. Kauer, T. Plehn, D. L. Rainwater, and D. Zeppenfeld, *$H \to W^+W^-$ as the discovery mode for a light Higgs boson*, Phys. Lett. **B503** (2001) 113–120, `arXiv:hep-ph/0012351 [hep-ph]`.

[56] S. P. Martin, *A Supersymmetry primer*, `arXiv:hep-ph/9709356 [hep-ph]`.

[57] O. J. Eboli and D. Zeppenfeld, *Observing an invisible Higgs boson*, Phys. Lett. **B495** (2000) 147–154, `arXiv:hep-ph/0009158 [hep-ph]`.

[58] CMS Collaboration, *Measurement of the hadronic activity in events with a Z and two jets and extraction of the cross section for the electroweak production of a Z with two jets in pp collisions at $\sqrt{s} = 7$ TeV*, `arXiv:1305.7389 [hep-ex]`. submitted to JHEP.

[59] U. Baur and D. Zeppenfeld, *Measuring three vector boson couplings in $qq \to qqW$ at the SSC*, `arXiv:hep-ph/9309227 [hep-ph]`.

[60] C. Oleari and D. Zeppenfeld, *QCD corrections to electroweak $l\nu jj$ and l^+l^-jj production*, Phys. Rev. **D69** (2004) 093004, `arXiv:hep-ph/0310156 [hep-ph]`.

[61] Y. Dokshitzer, V. Khoze, and T. Sjötrand, *Rapidity gaps in Higgs production*, Phys. Lett. B **274** no. 1, (1992) 116 – 121. `http://www.sciencedirect.com/science/article/pii/037026939290312R`.

[62] J. D. Bjorken, *Rapidity gaps and jets as a new-physics signature in very-high-energy hadron-hadron collisions*, Phys. Rev. D **47** (1993) 101–113. `http://link.aps.org/doi/10.1103/PhysRevD.47.101`.

[63] ATLAS Collaboration, "Atlas photo gallery." `http://www.atlas.ch/photos/index.html`.

[64] The LHC*b* Collaboration, *The LHCb experiment at the CERN LHC*, JINST 3 S08005 (2008).

[65] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST 3 S08002 (2008).

[66] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large*

Hadron Collider, JINST **3** no. 08, (2008) S08003.
http://stacks.iop.org/1748-0221/3/i=08/a=S08003.

[67] The CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST 3 S08004 (2008).

[68] L. Evans and P. Bryant, *LHC Machine*, JINST **3** no. 08, (2008) S08001. http://stacks.iop.org/1748-0221/3/i=08/a=S08001.

[69] ATLAS Collaboration, *Luminosity Determination in pp Collisions at* $\sqrt{s} = 7$ *TeV Using the ATLAS Detector at the LHC*, Eur. Phys. J. **C71** (2011) 1630, arXiv:1101.2185 [hep-ex].

[70] ATLAS Collaboration, *Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC*, arXiv:1302.4393 [hep-ex]. submitted to EPJC.

[71] ATLAS Collaboration, *Luminosity working group public results*, 2011. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[72] G. Aad *et al.*, *ATLAS pixel detector electronics and sensors*, JINST **3** no. 07, (2008) P07007. http://stacks.iop.org/1748-0221/3/i=07/a=P07007.

[73] The ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** no. 07, (2008) P07007.
http://stacks.iop.org/1748-0221/3/i=07/a=P07007.
The ATLAS Collaboration, *A measurement of Lorentz angle and spatial resolution of radiation hard silicon pixel sensors*, Nucl. Intrum. M. A **481** no. 13, (2002) 204 – 221. http://www.sciencedirect.com/science/article/pii/S0168900201014139.

[74] A. Abdesselam *et al.*, *The barrel modules of the ATLAS semiconductor tracker*, Nucl. Instrum. Meth. **A568** (2006) 642–671.
A. Abdesselam *et al.*, *The ATLAS semiconductor tracker end-cap module*, Nucl. Instrum. Meth. **A575** (2007) 353–389.

[75] The ATLAS Collaboration, *Beam tests of ATLAS SCT silicon strip detector modules*, Nucl. Intrum. M. A **538** no. 13, (2005) 384 – 407. http://www.sciencedirect.com/science/article/pii/S0168900204021059.

[76] The ATLAS TRT collaboration, *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*, JINST **3** no. 02, (2008) P02013. http://stacks.iop.org/1748-0221/3/i=02/a=P02013.

[77] B. Dolgoshein, *Transition radiation detectors*, Nucl. Instrum. Meth. **A326** (1993) 434–469.

[78] C. Leroy and P.-G. Rancoita, *Physics of cascading shower generation and propagation in matter: principles of high-energy, ultrahigh-energy and compensating calorimetry*, Reports on Progress in Physics **63** no. 4, (2000) 505. http://stacks.iop.org/0034-4885/63/i=4/a=202.

[79] D0 Collaboration, S. Abachi *et al.*, *The D0 detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **338** no. 23, (1994) 185 – 253. http://www.sciencedirect.com/science/article/pii/0168900294913129.

[80] ZEUS Collaboration, *The ZEUS detector: Status report 1993*,.

[81] ATLAS Collaboration, *Performance of the ATLAS muon trigger in 2011*, Tech. Rep. ATLAS-CONF-2012-099, CERN, Geneva, Jul, 2012.

[82] The ATLAS Collaboration. https://twiki.cern.ch/twiki/bin/view/

`AtlasPublic/TriggerOperationPublicResults#2011_7_TeV`.

[83] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81–86. See also http://root.cern.ch/.

[84] C. Eck *et al.*, *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005).* Technical Design Report LCG. CERN, Geneva, 2005.

[85] The GEANT4 Collaboration, S. Agostinelli *et al.*, *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.
ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** no. CERN-PH-EP-2010-044, (2010) 823–874., `arXiv:1005.4568 [hep-ph]`.

[86] E. Richter-Was, D. Froidevaux, and L. Poggioli, *ATLFAST 2.0 a fast simulation package for ATLAS*, Tech. Rep. ATL-PHYS-98-131, CERN, Geneva, Nov, 1998.

[87] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Tech. Rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar, 2007.
ATLAS Collaboration, *The Expected Performance of the ATLAS Inner Detector*, Tech. Rep. ATL-PHYS-PUB-2009-002. ATL-COM-PHYS-2008-105, CERN, Geneva, Aug, 2008.

[88] ATLAS Collaboration, *Performance of the ATLAS Electron and Photon Trigger in p-p Collisions at* $\sqrt{s} = 7$ *TeV in 2011*, Tech. Rep. ATLAS-CONF-2012-048, CERN, Geneva, May, 2012.

[89] ATLAS Collaboration, G. Aad *et al.*, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur.Phys.J. **C72** (2012) 1909, `arXiv:1110.3174 [hep-ex]`.

[90] J. Kretzschmar and L. Iconomidou-Fayard, *Electron performances measurements using the 2011 LHC proton-proton collisions*, Tech. Rep. ATL-COM-PHYS-2012-1024, CERN, Geneva, Jul, 2012. This is a draft of the CONF Note.

[91] ATLAS Collaboration, “Atlas egamma public results page.” `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ElectronGammaPublicCollisionResults`. accessed 13-4-2013.

[92] ATLAS Electromagnetic Barrel Calorimeter Collaboration, M. Aharrouche *et al.*, *Energy linearity and resolution of the ATLAS electromagnetic barrel calorimeter in an electron test-beam*, Nucl. Instrum. Meth. **A568** (2006) 601–623, `arXiv:physics/0608012 [physics]`.

[93] S. Laplace and J. de Vivie, *Calorimeter isolation and pile-up [ATLAS Internal note]*, Tech. Rep. ATL-COM-PHYS-2012-467, CERN, Geneva, May, 2012.

[94] J. Hartert and I. Ludwig, *Electron isolation in the ATLAS experiment*, Tech. Rep. ATL-PHYS-INT-2010-052, CERN, Geneva, May, 2010.

[95] ATLAS Collaboration, *Measurement of the inclusive* $W^{\pm}$ *and Z/gamma cross sections in the electron and muon decay channels in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, Phys. Rev. **D85** (2012) 072004, `arXiv:1109.5141 [hep-ex]`.
ATLAS Collaboration, *Measurement and QCD Analysis of Differential Inclusive* $W \to l\nu$ *and* $Z/\gamma^* \to ll$ *Production and Leptonic Decay Cross Sections with ATLAS: Analysis STDM-2012-20: W and Z inclusive cross*

section with 2011 data, Tech. Rep. ATL-COM-PHYS-2013-217, CERN, Geneva, Feb, 2013. Paper in preparation.

[96] ATLAS Collaboration, *Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2011-063, CERN, Geneva, Apr, 2011.

[97] ATLAS Collaboration, "Atlas muon public results page." `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonPerformancePublicPlots`. accessed 18-4-2013.

[98] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_T jet clustering algorithm*, JHEP **2008** no. 04, (2008) 063, `arXiv:0802.1189`. `http://stacks.iop.org/1126-6708/2008/i=04/a=063`.

[99] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. **B641** (2006) 57–61, `arXiv:hep-ph/0512210 [hep-ph]`. http://fastjet.fr/.

[100] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys.Rev. **D48** (1993) 3160–3166, `arXiv:hep-ph/9305266 [hep-ph]`.

[101] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, `arXiv:hep-ph/9907280 [hep-ph]`.

[102] ATLAS Collaboration, W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, *Calorimeter Clustering Algorithms: Description and Performance*, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008.

[103] ATLAS Collaboration, *Probing the measurement of jet energies with the ATLAS detector using Z+jet events from proton-proton collisions at $\sqrt{s} = 7$ TeV*, Tech. Rep. ATLAS-CONF-2012-053, CERN, Geneva, May, 2012.

[104] ATLAS Collaboration, *Probing the measurement of jet energies with the ATLAS detector using photon+jet events in proton-proton collisions at $\sqrt{s} = 7$ TeV*, Tech. Rep. ATLAS-CONF-2012-063, CERN, Geneva, Jul, 2012.

[105] ATLAS Collaboration, *TeV-scale jet energy calibration using multijet events including close-by jet effects at the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2013-003, CERN, Geneva, Jan, 2013.

[106] ATLAS Collaboration, *In situ jet pseudorapidity intercalibration of the ATLAS detector using dijet events in $\sqrt{s} = 7$ TeV proton-proton 2011 data*, Tech. Rep. ATLAS-CONF-2012-124, CERN, Geneva, Aug, 2012.

[107] A. Sherstnev and R. S. Thorne, *Parton Distributions for LO Generators*, Eur. Phys. J. **C55** (2008) 553–575, `arXiv:0711.2473 [hep-ph]`.

[108] M. Bahr, S. Gieseke, M. Gigg, D. Grellscheid, K. Hamilton, *et al.*, *Herwig++ Physics and Manual*, Eur. Phys. J. **C58** (2008) 639–707, `arXiv:0803.0883 [hep-ph]`.

[109] V. Lendermann, J. Haller, M. Herbst, K. Kruger, H.-C. Schultz-Coulon, *et al.*, *Combining Triggers in HEP Data Analysis*, Nucl. Instrum. Meth. **A604** (2009) 707–718, `arXiv:0901.4118 [hep-ex]`.

[110] ATLAS Collaboration, D. Miller, A. Schwartzman, and D. Su, *Pile-up jet energy scale corrections using the jet-vertex fraction method*, Tech. Rep. ATL-COM-PHYS-2009-180, CERN, Geneva. Internal note.

[111] Z. Marshall, *Re-defining the Standard QCD Di-Jet Samples*, Tech. Rep.

ATL-COM-PHYS-2011-992, CERN, Geneva, Jul, 2011. Internal note.

[112] ATLAS Collaboration, *In-situ pseudorapidity intercalibration for evaluation of jet energy scale uncertainty using dijet events in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, Tech. Rep. ATLAS-CONF-2011-014, CERN, Geneva, Mar, 2011.

[113] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, Eur. Phys. J. **C73** (2013) 2304, arXiv:1112.6426 [hep-ex].

[114] ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at* $\sqrt{s} = 7$ *TeV with ATLAS 2011 data*, Tech. Rep. ATLAS-CONF-2013-004, CERN, Geneva, Jan, 2013.

[115] "Jetuncertainties2011." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/JetUncertainties2011. internal documentation page.

[116] ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at* $\sqrt{s} = 7$ *TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. **C73** (2013) 2306, arXiv:1210.6210 [hep-ex].
ATLAS Collaboration, G. Romeo, A. Schwartzman, R. Piegaia, T. Carli, and R. Teuscher, *Jet Energy Resolution from In-situ Techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy* $\sqrt{s} = 7$ *TeV*, Tech. Rep. ATL-COM-PHYS-2011-240, CERN, Geneva, Mar, 2011. Internal note.

[117] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS*, Eur. Phys. J. **C72** (2012) 1844, arXiv:1108.5602 [hep-ex].
ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS with 2011 Proton-Proton Collisions at* $\sqrt{s} = 7$ *TeV*, Tech. Rep. ATLAS-COM-CONF-2012-105, Geneva, 2012.

[118] ATLAS Collaboration, *Measurement of the production cross section of jets in association with a Z boson in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, arXiv:1304.7098 [hep-ex]. sumbitted to JHEP.
ATLAS Collaboration, *Measurement of the high-mass Drell–Yan differential cross-section in pp collisions at* $\sqrt{s}$=7 *TeV with the ATLAS detector*, arXiv:1305.4192 [hep-ex]. submitted to PLB.

[119] "Atlantis event display for atlas." http://www.hep.ucl.ac.uk/atlas/atlantis/.

[120] B. Jager, S. Schneider, and G. Zanderighi, *Next-to-leading order QCD corrections to electroweak Zjj production in the POWHEGBOX*, JHEP **1209** (2012) 083, arXiv:1207.2626 [hep-ph].
F. Schissler and D. Zeppenfeld, *Parton Shower Effects on W and Z Production via Vector Boson Fusion at NLO QCD*, arXiv:1302.2884 [hep-ph].

[121] The ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, Tech. Rep. ATL-PHYS-PUB-2011-008, CERN, Geneva, Apr, 2011.

[122] D. Stump *et al.*, *Inclusive jet production and parton distributions and the search for new physics*, JHEP **10** (2003) 046, arXiv:hep-ph/0303013.

[123] P. Golonka and Z. Was, *Next to Leading Logarithms and the PHOTOS Monte Carlo*, Eur. Phys. J. **C50** (2007) 53–62, arXiv:hep-ph/0604232 [hep-ph].

[124] ATLAS Collaboration, S. Allwood-Spires *et al.*, *Monte Carlo samples used for top physics: Top Working Group Note IX*, Tech. Rep. ATL-COM-PHYS-2010-836, CERN, Geneva, Oct, 2010. Internal note.

[125] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C.-P. Yuan, *New parton distributions for collider physics.*, Phys. Rev. **D82** (2010) 074024, `arXiv:hep-ph/0303013`.

[126] K. Hamilton and P. Nason, *Improving NLO-parton shower matched simulations with higher order matrix elements*, JHEP **1006** (2010) 039, `arXiv:1004.1764 [hep-ph]`.
S. Hoche, F. Krauss, M. Schonherr, and F. Siegert, *NLO matrix elements and truncated showers*, JHEP **1108** (2011) 123, `arXiv:1009.1127 [hep-ph]`.

[127] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, Phys. Rev. **D69** (2004) 094008.

[128] J. Butterworth, E. Dobson, U. Klein, B. Mellado Garcia, T. Nunnemann, J. Qian, D. Rebuzzi, and R. Tanaka, *Single Boson and Diboson Production Cross Sections in pp Collisions at* $\sqrt{s}$*=7 TeV*, Tech. Rep. ATL-COM-PHYS-2010-695, CERN, Geneva, Aug, 2010.

[129] ATLAS Collaboration, "Atlas standard model cross-section task force." `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/StandardModelCrossSectionTaskForce`. Internal ATLAS documentation.

[130] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, *et al.*, *HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR*, Comput. Phys. Commun. **182** (2011) 1034–1046, `arXiv:1007.1327 [hep-ph]`.

[131] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. **C64** (2009) 653–680, `arXiv:0905.3531 [hep-ph]`.

[132] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612–622, `arXiv:1111.5869 [hep-ph]`.

[133] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, `arXiv:1112.5675 [hep-ph]`.

[134] CDF Collaboration, *Observation of Single Top Quark Production and Measurement of —Vtb— with CDF*, Phys. Rev. **D82** (2010) 112005, `arXiv:1004.1181 [hep-ex]`.
D0 Collaboration, *Measurements of single top quark production cross sections and* $|V_{tb}|$ *in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, Phys. Rev. **D84** (2011) 112001, `arXiv:1108.3091 [hep-ex]`.

[135] ATLAS Collaboration, *Search for s-Channel Single Top-Quark Production in pp Collisions at* $\sqrt{s} = 7$ *TeV*, Tech. Rep. ATLAS-CONF-2011-118, CERN, Geneva, Aug, 2011.
ATLAS Collaboration, *Measurement of the t-channel single top-quark production cross section in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, Phys. Lett. **B717** (2012) 330–350, `arXiv:1205.3130 [hep-ex]`.
ATLAS Collaboration, *Evidence for the associated production of a W boson and a top quark in ATLAS at* $\sqrt{s} = 7$ *TeV*, Phys. Lett. **B716** (2012) 142–159, `arXiv:1205.5764 [hep-ex]`.
CMS Collaboration, *Measurement of the single-top-quark t-channel cross section in pp collisions at* $\sqrt{s} = 7$ *TeV*, JHEP **1212** (2012) 035, `arXiv:1209.4533 [hep-ex]`.

CMS Collaboration, S. Chatrchyan *et al.*, *Evidence for associated production of a single top quark and W boson in pp collisions at 7 TeV*, Phys. Rev. Lett. (2012), `arXiv:1209.3489 [hep-ex]`.

[136] B. P. Kersevan and E. Richter-Was, *Improved phase space treatment of massive multi-particle final states*, Eur. Phys. J. **C39** (2005) 439–450, `arXiv:hep-ph/0405248 [hep-ph]`.
B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5*, `arXiv:hep-ph/0405247 [hep-ph]`.

[137] A. Sherstnev and R. Thorne, *Parton Distributions for LO Generators*, Eur. Phys. J. **C55** (2008) 553–575, `arXiv:0711.2473 [hep-ph]`.

[138] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. **D83** (2011) 091503, `arXiv:1103.2792 [hep-ph]`.
N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. **D81** (2010) 054028, `arXiv:1001.5034 [hep-ph]`.
N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. **D82** (2010) 054018, `arXiv:1005.4451 [hep-ph]`.

[139] J. M. Campbell and R. K. Ellis, *An update on vector boson pair production at hadron colliders*, Phys. Rev. **D60** (1999) 113006, `arXiv:hep-ph/9905386 [hep-ph]`.

[140] A. Buckley, J. Butterworth, L. Lonnblad, H. Hoeth, J. Monk, *et al.*, *Rivet user manual*, `arXiv:1003.0694 [hep-ph]`.

[141] M. Whalley, D. Bourilkov, and R. Group, *The Les Houches accord PDFs (LHAPDF) and LHAGLUE*, `arXiv:hep-ph/0508110 [hep-ph]`.

[142] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput.Phys.Commun. **77** (1993) 219–228.

[143] G. Cowan, *Statistical Data Analysis.* Oxford University Press, 1998.

[144] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, `arXiv:physics/0306116 [physics]`.

[145] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Comput.Phys.Commun. **10** (1975) 343–367.

[146] J. M. Campbell, R. K. Ellis, P. Nason, and G. Zanderighi, *W and Z bosons in association with two jets using the POWHEG method*, `arXiv:1303.5447 [hep-ph]`.

[147] E. Gallas, S. Albrand, J. Fulachier, F. Lambert, K. Pachal, *et al.*, *Conditions and configuration metadata for the ATLAS experiment*, Tech. Rep. ATL-SOFT-PROC-2012-040, ATL-COM-SOFT-2012-073, Geneva, 2012.
ATLAS Collaboration, *Light-quark and Gluon Jets in ATLAS*,.