W Boson Production in Association with Hadronic Jets at ATLAS

Maria Fiascaris

Merton College, Oxford

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

Trinity Term, 2010

W Boson Production in Association with Hadronic Jets at ATLAS

Maria Fiascaris, Merton College, Oxford

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford. Trinity Term, 2010

Abstract

The Large Hadron Collider (LHC) offers unprecedented opportunities to explore unknown kinematic regions and discover new Physics. Its discovery capabilities, however, strongly depend on our understanding of the Standard Model. Leptonic decays of the W boson produced in association with jets are one of the dominant backgrounds to Physics beyond the Standard Model, particularly Supersymmetry. These processes can also serve as testing ground for the theory of strong interactions. Their study is therefore of great importance at the LHC.

The aim of this thesis is to prepare the essential tools for an early data measurement of $W \rightarrow e\nu$ + jets processes, assuming an integrated luminosity of 100 pb^{-1}. Based on Monte Carlo simulations, several aspects of the analysis are discussed, from the treatment of backgrounds, to the calculation of electron efficiencies and acceptances. The focus is on data-driven techniques, which are going to be crucial, particularly in the early-data phase, to minimise dependence on Monte Carlos. A data-driven method for the extraction of electron efficiencies using $Z \rightarrow ee$ events is thoroughly studied. The expected statistical and systematic uncertainties are evaluated and the systematic bias on the $W \rightarrow e\nu$ + jets cross-section is assessed. Finally, the major sources of systematic uncertainties are treated and the experimental prospects for measuring the $W \rightarrow e\nu + N_{jets}$ absolute cross-section and the ratios $\frac{\sigma(W+N_{jets}+1)}{\sigma(W+N_{jets})}$ and $\frac{\sigma(W+N_{jets})}{\sigma(Z+N_{jets})}$ (with N_{jets} = 1, 2) are presented. It is suggested that the ratios are to be preferred as robust measurements for the early data scenario and it is discussed what aspects of the analysis are the most critical in each case.

To Alberto and Barbara

Acknowledgements

The last four years, spent between Oxford and CERN, have been truly exciting and enriching, both for my professional and personal life. It is a pleasure to thank those who made this possible. I would like to thank the Science and Technology Facilities Council (STFC) for providing financial support giving me the opportunity to undertake this degree.

I owe my deepest gratitude to my supervisor, Prof. Amanda Cooper-Sarkar, and to Dr. Tony Weidberg, whose guidance, constant support and encouragement have been invaluable not only for the completion of this thesis but also for my future career. I would like to thank all the people of the Oxford Particle Physics sub-department, including the IT support and the secretariat, who have been extremely patient and helped me solve innumerable problems. A special thanks goes to my fellow students in the ATLAS group, who supported me in getting started when I first arrived: Ellie Dobson, Kristin Lohwasser, Guillaume Kirsch, Oleg Brandt. I also thank the numerous other students and post-docs that have offered unvaluable help, as well as entertainment.

One of the nicest aspects of scientific research, and ATLAS in particular, is the opportunity to collaborate with a large number of people, from different countries and institutions. During my thesis, I had the pleasure of working with numerous people, in Oxford and elsewhere. I am grateful to Alessandro Tricoli, for his guidance and support, which have been very important especially (but not only) at the initial stages of my DPhil. I would like to thank Hugo Beauchemin (and the rest of the ’Rjets’ team) for the fruitful collaboration and long discussions that have helped me understanding some of the complexities of a physics analysis. It has been a great pleasure to be part of the ’Tag-and-Probe’ and the ’Insitu-Performance’ teams. My thanks go to Olivier Arnaez, Daniel Froidevaux, Elzbieta Richter-Was, Matthias Schott and Mike Flowerdew. I would also like to show my gratitude to the conveners of the ATLAS subgroups for their guidance. Another important part of my DPhil has been the exciting time spent in the ATLAS control room and for this I am grateful to Sandra Ciocio and the rest of the SCT Commissioning team.

I would also like to thank all my friends, too numerous to mention, who have supported me during these years.

This thesis is dedicated to my family who have supported and encouraged me throughout the last four years and for many years before. To my parents who have always reposed confidence in me and placed great importance in my studies and work, despite it being so distant from their own world. It is thanks to them that I reached this point in the first place. To Till, who has been of enormous support and has taught me to approach work, and life in general, with passion and playfulness.

Preface

The analysis presented in this thesis was developed in the few years immediately preceding the start-up of the LHC and reached completion when the ATLAS detector was showing the first physics results from collisions with a centre of mass energy of 7 TeV. At the time of writing, however, the amount of data collected was not sufficient to perform the studies described in this thesis. Due to time constraints and the limited statistics available, the analysis presented is therefore based purely on Monte Carlo simulations.

ATLAS is a very large collaboration. Due to the complexity of the experiment, any measurement unavoidably involves contributions from a large number of people. This thesis presents my own contribution to the analysis of W + jets processes and therefore focuses on the particular aspects of the analysis I concentrated on during my studies. Contributions from other people are referenced as they arise.

The analysis presented in this document is only one part of the work I have done during my D Phil. The rest could not be included due to space limitations and is briefly described here.

As part of my technical work at ATLAS, I have been involved in the commissioning of the SemiConductor Tracker (SCT). Installed in the experimental hall from 2006 to 2008, the SCT has undergone extensive integration and commissioning with cosmic rays. During this time, I worked on the testing of the optical links used for the modules read out and monitored the individual power supplies and environmental variables. Since 2008, the SCT has participated in ATLAS combined runs for cosmic data taking and I have been part of the detector operation team. I have presented results on the commissioning of the SCT and Pixel detectors at an international conference [1].

A significant part of my thesis has been an in-situ method (so called Tag-and-Probe) to measure electron efficiencies from data using $Z \rightarrow ee$ events. This analysis was developed within the ATLAS electron/photon performance working group, with a tight connection with the W/Z boson inclusive and W/Z boson + jets working groups. The work has been presented at several ATLAS workshops, contributed to the ”Electrons and Photons” and ”Standard Model” chapters of the ATLAS CSC book [2], and has been written into an internal note [3] (another one is currently in preparation) where I am one of the main authors. I have also co-developed the official ATLAS package for the in-situ determination of detector performance, where I am responsible for the electron performance sub-package. In addition to the electron performance studies described in this document, I also investigated the feasibility of extracting tracking efficiencies from data using $Z \rightarrow ee$ and $W \rightarrow e\nu$ events. Furthermore, I performed preliminary studies of a novel statistical tool [4] for background subtraction in the context of efficiency determination.

The physics topic of my D Phil is the production of W boson in association with hadronic jets. The work, presented in this thesis, has been shown at workshops and international conferences [5] and written up in ATLAS publications [2] and internal notes [3].

Contents

Glossary

$\boldsymbol{A}$: Acceptance

$\boldsymbol{A_{Gen}}$: Acceptance at generator level

$\boldsymbol{A_{Gen \to Rec}}$: Correction factor to account for detector smearing on the acceptance

$\boldsymbol{A_{Rec}}$: Acceptance at detector level

B : Number of background events

BSM : Beyond the Standard Model

$\boldsymbol{\Delta R}$: Angular separation

$\boldsymbol{\Delta R_{ej}}$: ΔR between electron and closest jet

$\boldsymbol{\Delta R_{jj}}$: ΔR between two jets

$\boldsymbol{\Delta}$: Uncertainty

$\boldsymbol{\eta}$: Pseudo-rapidity

$\mathbf{e20\_loose}$: Specific single electron trigger

$\boldsymbol{\epsilon}$: Efficiency

$\boldsymbol{\epsilon_{\mathrm{ele}}}$: Electron efficiency ($= \epsilon_{\mathrm{reco}} \cdot \epsilon_{\mathrm{id}} \cdot \epsilon_{\mathrm{trig}}$)

$\boldsymbol{\epsilon_{\Delta \mathrm{Rej}}}$: Efficiency of ΔR_{ej} cut

$\boldsymbol{\epsilon_{\Delta \mathrm{Rjj}}}$: Efficiency of ΔR_{jj} cut

$\boldsymbol{\epsilon_{\mathrm{id}}}$: Electron identification efficiency

$\boldsymbol{\epsilon_{\mathrm{jet}}}$: Jet efficiency ($= \epsilon_{\mathrm{reco\ jet}} \cdot \epsilon_{\Delta \mathrm{Rej}} \cdot \epsilon_{\Delta \mathrm{Rjj}}$)

$\boldsymbol{\epsilon_{\mathrm{reco}}}$: Electron reconstruction efficiency

$\boldsymbol{\epsilon_{\mathrm{reco\ jet}}}$: Jet reconstruction efficiency

$\boldsymbol{\epsilon_{\mathrm{trig}}}$: Electron trigger efficiency

EF : Event Filter

EM : Electromagnetic

$\boldsymbol{E_T^{\mathrm{miss}}}$: Missing transverse energy

FSR : Final State Radiation

HLT : High Level Trigger

HT : Scalar sum of first and second jet p_T

ID : Inner Detector

ISR : Initial State Radiation

JER : Jet Energy Resolution

JES : Jet Energy Scale

L : Integrated luminosity

LHC : Large Hadron Collider

LO : Leading-Order

LVL1 : Level 1 trigger

LVL2 : Level 2 trigger

MC : Monte Carlo

ME : Matrix Element

$\boldsymbol{M_{ee}}$: Invariant mass of an electron pair

NLO : Next-to-Leading-Order

Pdf : Probability Density Function

PDF : Parton Distribution Function

PID : Particle Identification (for electrons)

PS : Parton Shower

pQCD : Perturbative QCD

$\boldsymbol{p_T}$: Transverse momentum

$\boldsymbol{p_T^W}$: W boson transverse momentum

TP : Tag-and-Probe

QCD : Quantum ChromoDynamics

QED : Quantum ElectroDynamics

RoIs : Region of Interests

$\boldsymbol{R_{jets}}$: Ratio $\frac{\sigma(W+N_{jets}+1)}{\sigma(W+N_{jets})}$

$\boldsymbol{R_W}$: Ratio $\frac{\sigma(W+N_{jets})}{\sigma(Z+N_{jets})}$

$\boldsymbol{\sigma}$: Cross-section

S : Number of signal events

SCT : SemiConductor Tracker

SM : Standard Model

TRT : Transition Radiation Tracker

Chapter 1

Introduction

The desire to understand Nature in the most fundamental way has always been the driving force behind scientific progress. Today, humanity has achieved impressive advancements in the understanding of the world around us. The great variety of physics phenomena are explained in terms of four fundamental interactions between elementary particles. All but the gravitational force are combined into the Standard Model (SM) of particle physics, currently the universally accepted theory describing the building blocks of matter. The SM has been successfully tested over the past 30 years and only one of the predicted particles, the Higgs boson, has so far eluded discovery. Yet, rather than completing our picture of nature, such advancements have raised even more questions and highlighted the shortcomings of the theory. The SM can simply not answer some of the observed features of our universe, such as the origin of dark matter and the predominance of matter over antimatter. This does not imply that the SM is wrong, but rather incomplete: its limitations can be overcome by adding new particles and new interactions.

In addition to searching for the elusive Higgs boson, the Large Hadron Collider (LHC) hopes to provide evidence for new physics. The ultimate scope of the LHC is not to further test the SM, but rather to use it as a tool in searches for unknown phenomena. Despite the valuable knowledge accumulated in previous collider experiments, further validation of the SM at the LHC conditions is going to be a fundamental pre-requisite for any new discovery.

In particular, understanding the theory of strong interactions which dominates any process at the LHC is going to be one of the challenging tasks. The study of W boson production in association with hadronic jets presented in this thesis is carried out in this perspective: it is a tool to validate the strong-interaction dynamics at the LHC energies and also an important background to both Standard Model and Beyond the Standard Model physics. This analysis could therefore directly/indirectly feed into new physics searches. Indirectly, by validating the available Monte Carlo generators and theoretical predictions, indispensable tools for any discovery. Directly, by accurately measuring the background to several potential discovery channels. For these reasons, W boson plus jets is going to be one of high-priority Standard Model processes to be measured at the LHC.

1.1 Thesis Outline

This thesis presents the work developed in preparation for the first measurements of W boson plus jet production at the LHC, with the W decaying into the electron channel. The analysis is performed on fully simulated Monte Carlo samples, generated with a centre of mass energy of $\sqrt{s}$ = 10 TeV. Throughout this thesis, an integrated luminosity of 100 pb^{-1} is assumed.

Since the work is carried out solely on Monte Carlo simulations, the focus is on the development of techniques necessary for extracting a cross-section or a ratio of cross-sections, with particular emphasis on data-driven methods which could be applied with the first $\sim$ 100 pb^{-1} of data. Rather than presenting a Monte Carlo based result, this thesis is concluded with a discussion of the experimental prospects for studying W + jets processes, with focus on the 1 and 2 jet multiplicity bins. In addition to the absolute cross-section, also ratio measurements are treated. Ratios are particularly interesting since they lead to a cancellation of several systematic uncertainties. This work should therefore be considered as a starting point for designing a robust measurement of $W \rightarrow e\nu$ + jets.

The first chapters give a brief introduction to the ATLAS experiment at the LHC (Chapter 2) and to the physics of W + jets (Chapter 3). Chapter 4 explains how $W \rightarrow e\nu$ + jets events are selected and Chapter 5 describes the main backgrounds and their estimation. An entire chapter (Chapter 6) is dedicated to the development of a data-driven method for the extraction of electron efficiencies. The methodology for applying efficiency and acceptance corrections to extract the cross-section is described in Chapter 7. Finally, Chapter 8 discusses how the techniques developed in previous chapters are tied together for a final measurement. In addition to the absolute cross-section measurement of $W \rightarrow e\nu$ + 1, 2 jets, the ratios $\frac{W+2\text{ jets}}{W+1\text{ jet}}$ and $\frac{W+1,2\text{ jets}}{Z+1,2\text{ jets}}$ are also addressed and the experimental prospects for each of these measurements are presented.

Chapter 2

The ATLAS Experiment at the LHC

The Large Hadron Collider is starting a new era in particle physics: with its unprecedented energy and luminosity, it offers a great potential for new discoveries. This chapter gives a short introduction to the physics which is going to be studied at the LHC and briefly describes the ATLAS detector. Sections 2.1-2.2 are dedicated to the Standard Model and physics Beyond the Standard Model and Section 2.3 describes the physics programme of the LHC. Finally, the components of the ATLAS detector which are most relevant for this thesis are described in Section 2.4.

2.1 The Standard Model

The Standard Model (SM) is the most widely accepted theory which incorporates all the currently known elementary particles (fermions) and all interactions other than gravity. The fundamental interactions are described in the formalism of Quantum Field Theory (QFT) and are mediated by spin-1 particles, called gauge bosons: the photon (carrier of the electromagnetic interaction, described by the theory of Quantum Electrodynamics, QED), the massive $W^{\pm}$ and Z bosons (for the weak interaction) and eight gluons (carriers of the strong interaction, described by the theory of Quantum ChromoDynamics, QCD). Masses of particles are introduced

Particle Type	Symbol	Spin
Leptons	$e^{\pm}$, $\mu^{\pm}$, $\tau^{\pm}$	1/2
	ν_e, ν_μ, ν_τ	1/2
	$\bar{\nu}_e$, $\bar{\nu}_\mu$, $\bar{\nu}_\tau$	1/2
Quarks	u, d, c, s, t, b	1/2
	$\bar{u}$, $\bar{d}$, $\bar{c}$, $\bar{s}$, $\bar{t}$, $\bar{b}$	1/2
Bosons	γ, Z^0, $W^{\pm}$, g	1
	H^0	0

Table 2.1: Particle spectrum of the Standard Model.

by their coupling to a scalar field, the Higgs boson. The particle spectrum of the SM is shown in Table 2.1.

The SM has so far been amazingly successful in explaining a wide variety of experimental results. Its predictive power was first highlighted by the discovery of weak neutral currents [6] at the European Organisation for Nuclear Research (CERN) in 1973 and subsequently confirmed by numerous other discoveries (such as the observation of the W and Z bosons [7, 8, 9, 10] and the top quark [11, 12]) and precision measurements at Large Electron-Positron Collider (LEP) and at the Tevatron proton-antiproton collider. Searches for the Higgs boson, the only missing particle, are being performed at the Tevatron and LHC.

2.2 Physics Beyond the Standard Model

Despite its successes, the SM cannot be complete. The SM is an effective theory, only valid up to a certain energy scale. At the very least, at the Plank scale (10^{19} GeV), when gravity becomes important, it is necessary to extend the SM with a quantum theory of gravity. Several arguments suggest that physics Beyond the Standard Model (BSM) could already appear well below the Plank scale. One of these is the hierarchy problem concerning radiative loop corrections to the bare Higgs mass [1]. Other theoretical arguments concern the unification of gauge couplings.

[1] In order to keep the mass of the Higgs around a few hundreds GeV, as suggested by global electroweak fits [13], and avoid a remarkable fine-tuning, new physics should appear at the TeV scale.

Furthermore, the SM does not explain the matter-antimatter asymmetry observed in the universe and does not provide a satisfactory dark matter candidate.

There are several scenarios for new physics beyond the SM at relatively low scales, invoking for example new symmetries, extra spatial dimensions or new interactions. One of the theoretically favoured candidates is Supersymmetry (SUSY) (for a comprehensive description see, for example, [14]). It postulates a theory invariant under a symmetric transformation that turns bosonic final states into fermionic final states and vice-versa. The main motivation for its proposal is that it solved the hierarchy problem by a striking cancellation between the various contributions. In addition, SUSY provides a suitable candidate for dark matter (the lightest supersymmetric particle, LSP) and even provides the possibility of unifying the gauge couplings at a scale of about 10^{16} GeV.

2.3 Physics at the LHC

The Large Hadron Collider (LHC) at CERN is the largest and most powerful particle accelerator. It has been designed to deliver proton-proton collisions[2] every 25 ns at a centre of mass energy of 14 TeV and a design luminosity of $10^{34}\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$. Since the end of March 2010, LHC has been operating at a reduced centre of mass energy of 7 TeV [3] and has been gradually increasing the delivered luminosity with the aim of achieving the goal of $10^{32}\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$ by the end of 2010. Along the 27 km long LHC ring, are situated four main experiments: ATLAS and CMS (general purpose detectors), ALICE and LHCb (specialised detectors).

The main motivation for building the LHC is the search for the Higgs boson and for evidence of physics beyond the SM. Nevertheless, the LHC physics programme covers a wide range of topics. We provide here a brief summary, focusing specifically on the physics goals of the general purpose detector ATLAS:

[2] The LHC will also collide heavy ions, in particular lead, at 5.5 TeV per nucleon pair.

[3] Although first collisions at $\sqrt{s}$ = 900 GeV occurred on November 23rd 2009.

- **SM Physics**: the high luminosity and increased cross-sections will allow precision measurements of the electroweak sector, flavour physics and tests of QCD. The LHC will be a factory of W and Z bosons, which in addition to being useful for detector performance studies, will provide constraints on QCD and the proton structure in particular. The properties of the top quark (its couplings, mass and spin) and its production cross-section will also be studied. Such measurements will not only help to test the consistency of the SM, but also provide important information for BSM physics searches and, in some cases (eg. top quark properties), might even provide a hint of new physics.

- **Higgs boson searches**: various Higgs decay channels will be searched for at the LHC, to ensure coverage over a wide Higgs mass range (about 80 GeV to 1 TeV) and ensure redundancy. The ATLAS experiment has a large discovery potential for the Higgs boson both in the SM and in the Minimal Supersymmetric Standard Model (MSSM) [4]. For a review of the ATLAS Higgs search strategies the reader can refer to [2].

- **BSM Physics**: searches for new physics cover a wide range of potential discovery channels for SUSY and more exotic scenarios (see [2]). One of the main focuses will be on SUSY models which manifest themselves through cascades of supersymmetric particle decays resulting in large missing transverse energy, high transverse momentum jets and a certain number of isolated leptons.

2.4 A Toroidal LHC ApparatuS

The design of the ATLAS (A Toroidal LHC ApparatuS) detector has been driven by the stringent requirements imposed by the complex environment of a hadron collider combined with the excellent performance necessary to accomplish the wide physics programme (see Section 2.3). The very high luminosity of the LHC on one side provides the high statistics indispensable to

[4]The MSSM is the simplest supersymmetric extension of the SM, with minimal possible particle content.

gain sensitivity on processes with low cross-sections, but on the other hand leads to very busy final states, where the event of interest is accompanied by on average 23 inelastic events per bunch-crossing. In addition, physics at hadron colliders is dominated by final states with QCD-jets which are a formidable background to numerous interesting processes. In such conditions, identifying rare interesting processes requires high performance from all sub-components, as well as an accurate and efficient trigger system.

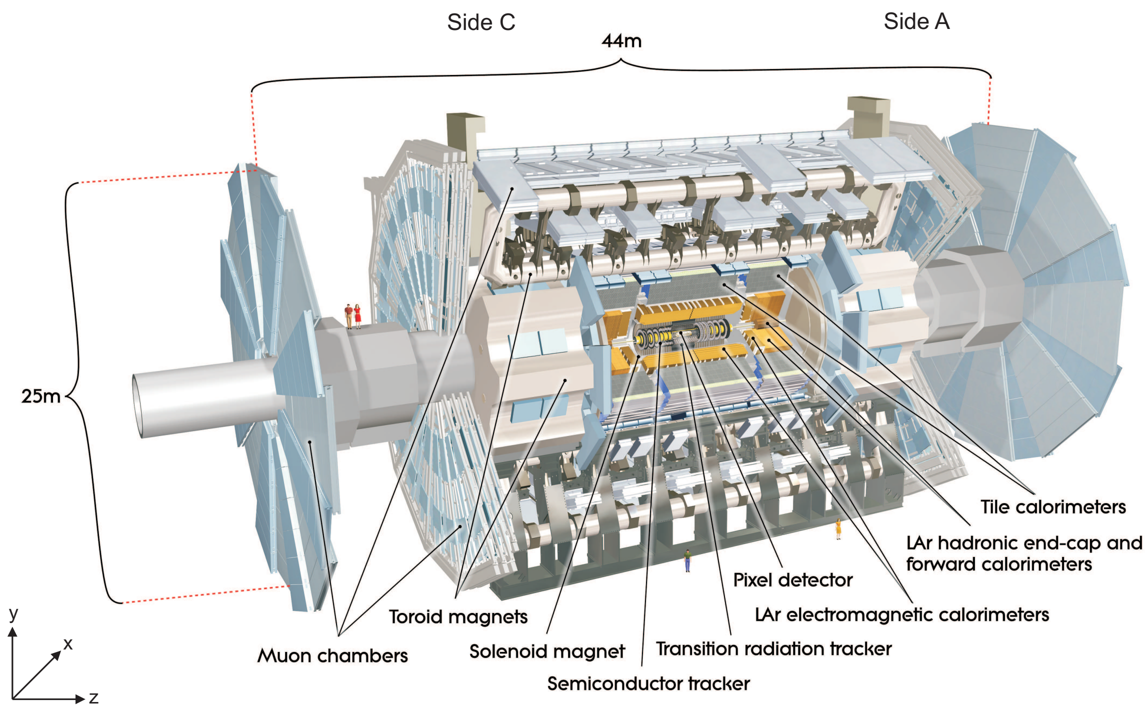

Figure 2.1: Cut-away view of the ATLAS detector and its sub-components [15]. Shown is also the ATLAS coordinate system: the z-axis is defined along the beam axis, while the $x-y$ axes define the transverse plane.

In this section the ATLAS sub-detectors and the trigger system are briefly described. Before entering into a detailed description, it is useful to define the coordinate system and some nomenclature used throughout this thesis.

2.4.1 The ATLAS Coordinate System

The ATLAS coordinate system is shown in the left bottom corner of Fig. 2.1. The z direction is along the beam axis, the positive x-axis points towards the centre of the LHC ring and the positive y-axis points upwards. In spherical coordinates (R, θ, ϕ) , the polar angle θ is the angle from the beam axis and azimuthal angle ϕ is measured around the beam axis:

$$
\begin{aligned}
R &= \sqrt{x^2 + y^2 + z^2} \\
\theta &= \tan^{-1}\left(\frac{\sqrt{\mathrm{x}^2 + \mathrm{y}^2}}{\mathrm{z}}\right) \\
\phi &= \tan^{-1}\left(\frac{\mathrm{y}}{\mathrm{x}}\right).
\end{aligned}
$$

The kinematics of a particle of energy E and momentum $\vec{p}$ are usually described in the (p_T, ϕ, η) system, where p_T is the transverse momentum and η is the pseudo-rapidity:

$$
\begin{aligned}
p_T &= \sqrt{(p_x^2 + p_y^2)} = \vec{p} \cdot \sin\theta \\
\eta &= \frac{1}{2}\ln\left(\frac{\mathrm{p} + \mathrm{p_z}}{\mathrm{p} - \mathrm{p_z}}\right) = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]
\end{aligned}
$$

where $p = |\vec{p}|$ is the magnitude of the momentum vector and p_z is the momentum in the z-direction. The angular separation between two particles, ΔR, is:

$$
\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{2.3}
$$

2.4.2 Detector Overview

A view of the ATLAS detector with its sub-components is shown in Fig. 2.1. At the innermost radius, the Inner Detector, placed inside the solenoid, provides tracking of charged particles. Outside this sits the calorimeter and finally the outer part is occupied by the muon spectrometer with large superconducting air-core toroid magnets. The sub-detectors most relevant for the analysis presented are the Inner Detector and the calorimeters.

Inner Detector

The Inner Detector (ID) is the system closest to the interaction point. Immersed in a 2T solenoidal field, it is designed to provide excellent momentum and vertex resolution for particles with $|\eta| < 2.5$ and $p_T > 0.5$ GeV [15]. It is built with high granularity, fast, radiation-hard electronics and sensor elements to be able to cope with the LHC high interaction rates and particle fluxes. The ID comprises three independent but complementary sub-detectors, shown in Fig. 2.2 (left): the Pixel Detector, the SemiConductor Tracker (SCT) and the Transition Radiation Tracker (TRT).

Pixel Detector: In the innermost part of the tracking volume, the highest granularity is achieved by the Pixel detector. With approximately 80.4 million read-out channels, the Pixel detector provides excellent vertex and impact parameter measurements, which are critical for heavy-flavour and τ-lepton tagging. A pixel module consists of a 250 μm thick silicon sensor [16], segmented in pixels of typical size $50 \times 400 \mu\text{m}^2$. As an ionising particle crosses a pixel element, it creates a number of electron-hole pairs proportional to the energy loss. An externally applied electric field separates the pairs before they recombine. The induced current is amplified and compared to a programmable threshold. If above threshold, the hit is recorded with spatial information (given by the pixel position) and a time stamp.

SCT: At higher radii, the measurement of particle momentum and spatial resolution are enhanced by the silicon strips of the SCT. An SCT module consists of two pairs of identical single-sided micro-strip sensors of typical pitch 80 μm. The two sensors are glued back-to-back at a stereo angle of 40 mrad, allowing to determine the hit position along the strip. The detection principle is identical to the Pixel detector.

TRT: Finally, at the outer radius the TRT provides particle tracking while also improving electron identification capabilities. The TRT consists of many layers of drift tubes of 4 mm in diameter and up to 1.44 m in length interleaved with transition radiation material. It exploits partly the principle of straw gas detectors for tracking (whereby a charged particle ionises the gas and generates an electric signal) and partly the principle of transition radiation (TR) [5]. Since the probability for a particle to emit TR depends solely on the Lorentz γ-factor, it can be used to separate electrons (high TR probability) from pions (low TR probability). The TRT uses therefore two thresholds for the signal read-out: a low threshold for tracking and a high threshold for electron identification by TR.

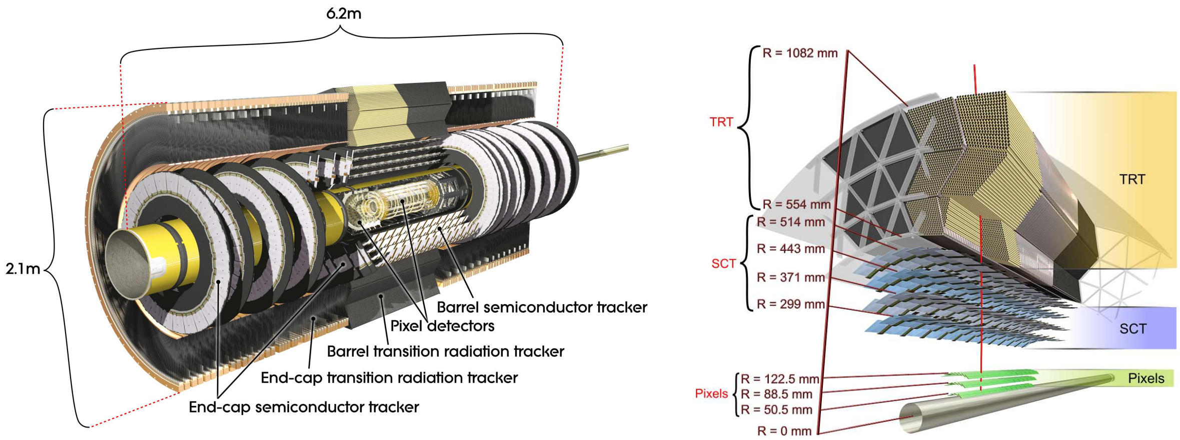

Figure 2.2: (Left) The Inner Detector envelope with its three sub-detectors: Pixels, SemiConductor Tracker and Transition Radiation Tracker. (Right) A section of the ID barrel, with the main detector components and their dimensions. A track crossing the detector is also shown.

The silicon precision trackers (Pixel Detector and SCT) are arranged in concentric cylinders around the beam axis in the central barrel region (three Pixel and four SCT layers), and in disks perpendicular to it in the end-caps (three Pixel and nine SCT disks per side). The TRT straw tubes are arranged in 73 planes parallel to the beam axis in the barrel and in 160 planes radially around it (end-caps). A section of the ATLAS ID barrel is shown in Fig. 2.2 (right): a track crossing the ID volume typically leaves three high precision hits in the Pixel Detector and four space-points in the SCT. The number of TRT hits is much higher (typically 36 per track), compensating the TRT poorer resolution.

[5] Transition radiation arises when relativistic charged particles cross a boundary between two media with different dielectric constants.

The Calorimeters

Sitting outside the solenoidal magnet, the calorimeters provide the energy measurement of electrons, photons and hadrons over a range of $|\eta| < 4.9$. As an incident particle interacts with the detector material (through electromagnetic or strong processes), it produces a shower of particles with gradually lower energy until full absorption. The energy deposited in the active medium, detected in form of light or charge, serves as an energy measurement of the incident particle. Since these are sampling calorimeters, a correct calibration is crucial for the physics performance. Issues related to the calibration of electrons and jets are treated in Chapter 4. The ATLAS calorimetry system comprises three sub-systems (see Fig. 2.3): the electromagnetic (EM), the hadronic and the forward calorimeter (FCal).

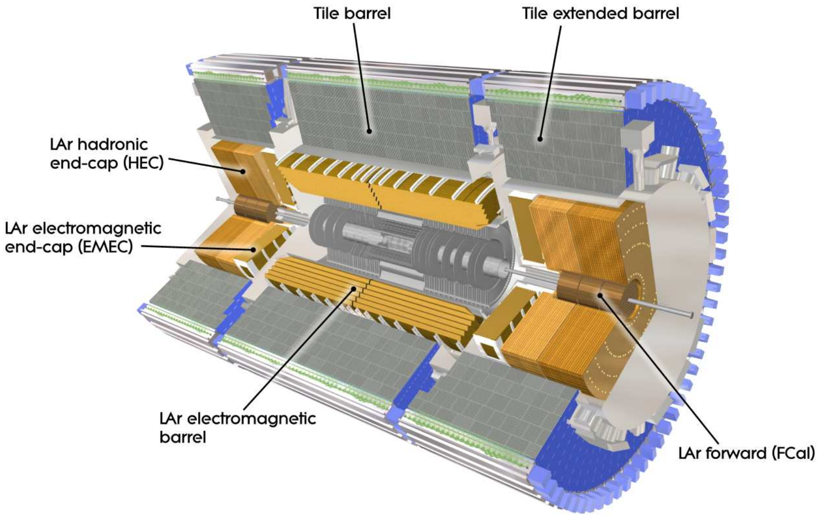

Figure 2.3: Cut-away view of the ATLAS calorimeter system [15].

EM Calorimeter: The EM calorimeter provides excellent energy and position resolution for electrons and photons in the pseudo-rapidity range $|\eta| < 3.2$. It is a liquid-argon (LAr) sampling calorimeter with lead absorber plates. The accordeon geometry is perfectly symmetrical in ϕ without azimuthal cracks. In pseudo-rapidity, the EM calorimeter is divided into a barrel part ($|\eta| < 1.475$) and two end-caps ($1.375 < |\eta| < 3.2$). The η region matched to the ID ($|\eta| < 2.5$) is segmented longitudinally into three sections with high granularity. In the region

$|\eta| < 1.8$, a pre-sampler provides corrections for energy losses in upstream material [6]. The importance of this longitudinal segmentation for precision measurements will be discussed further in Chapter 4, in the context of electron reconstruction and identification. The outer region ($|\eta| > 2.5$) has coarser granularity and is segmented into two sections in depth.

Hadronic Calorimeter and FCal: The main goal of the hadronic calorimeter is to provide good jet and missing transverse energy measurement. Crucial for this are large coverage in η and a thickness sufficient to contain the hadronic shower and limit punch-through into the muon system. Granularity is less important than for the EM calorimeter and is therefore coarser. Hadronic calorimetry in the range $|\eta|$ <1.7 is provided by the Tile calorimeter, a sampling device made out of steel as the absorber material and scintillating tiles as the active medium. In the end-caps (1.5 $<$ $|\eta|$ $<$ 3.2), matching the range covered by the EM calorimeter, LAr technology is used. The forward region, up to $|\eta|$ = 4.9, is covered by the Forward calorimeter, also built in LAr technology, which provides both EM and hadronic energy measurement.

2.4.3 The ATLAS Trigger System

The challenging LHC environment with $\sim 10^9$ p-p interactions per second, combined with the large number of electronic channels ($\sim 10^8$) of the ATLAS detector, results in a huge amount of output data. Storing all these data would not only be unfeasible, but even undesirable, since most of them are uninteresting events. The scope of the trigger is to reduce the output rate from the initial 40 MHz to about 200 Hz, while retaining interesting rare physics signatures for off-line analysis.

The ATLAS trigger system [17, 18] is organised into three levels: level-1 (LVL1), level-2 (LVL2) and event-filter (EF). LVL2 and EF together form the high-level trigger (HLT). A schematic view of the trigger system is shown in Fig. 2.4. The trigger works like a chain, with some basic selection criteria applied at each step. Each trigger level refines the decision made at

[6] Historically, the motivation for building the pre-sampler was to enhance the performance of the calorimeter in view of Higgs searches in the $H \to \gamma\gamma$ decay mode, which requires excellent energy and angular resolution.

a previous level. If an event fails all trigger criteria at any stage, it is rejected and not processed any further. This early rejection approach saves processing time, which is made available for more sophisticated selections on the remaining events.

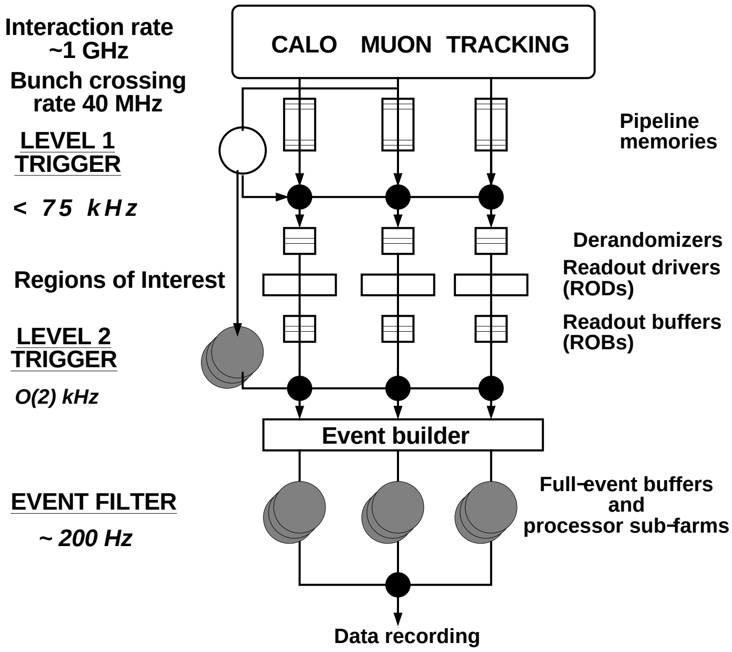

Figure 2.4: ATLAS trigger and data acquisition system.

The **LVL1** trigger is hardware-based and searches for events with high-p_T objects (leptons, jets, photons) and large missing and total transverse energy. Only coarse granularity information from the calorimeter and the muon spectrometer is used at this level. The LVL1 trigger reduces the rate from 40 MHz to about 75 kHz in a decision time of 2.5 μs.

The **HLT** trigger is software-based and runs on computing farms. At **LVL2**, the full granularity of the detector is used, including information from the ID. However, only regions around objects identified at LVL1 are considered. The rate is reduced to about 3 kHz, with a processing time of 40 ms. Finally, the **EF** adopts selection criteria similar to those used in off-line analysis (data from the full detector is available). It further reduces the rate to the desired 200 Hz, with an average processing time of the order of four seconds. Events that pass the EF are sent to permanent storage for off-line analysis.

2.4.4 The ATLAS Detector in Action

The ATLAS detector is currently operating, recording p-p collisions at a centre of mass energy of 7 TeV. W, Z bosons and, more recently, top quark candidates have been observed [19].

Figure 2.5 shows an event display of a candidate $W \rightarrow e\nu$ event. The large figure on the left is a cross-section of the ATLAS detector in the transverse ($R - \phi$) plane. The dark middle region represents the ID, while the green area is the EM calorimeter and the red is the hadronic calorimeter (the muon system is not shown). The yellow line, leaving hits in the ID and an energy deposit in the EM calorimeter, is the electron. The direction of missing transverse energy is shown by the red dotted line. The top right plot shows another view of the same event in the $R - z$ plane. The Pixel and SCT cylindrical layers and disks are clearly visible. In the bottom right, a lego plot in the η-ϕ plane shows the energy deposition in the calorimeter. The values of the reconstructed objects and of the W boson candidate transverse mass are also shown.

This is only one of the many W boson candidates that have been observed and of the many more that are expected in the coming months. The exciting LHC era has just started.

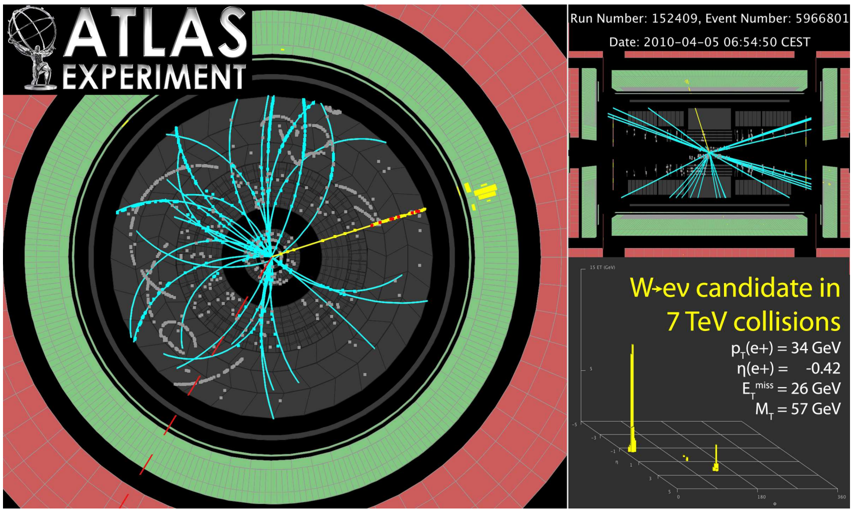

Figure 2.5: A candidate $W \rightarrow e\nu$ event recorded by the ATLAS detector at a centre of mass energy of 7 TeV [19].

Chapter 3

Production of W + Jets at the LHC

This chapter introduces the main topic of this thesis, by providing some background information on the production of W bosons in association with jets. The first part of the chapter (Section 3.1-3.2) gives a general description of QCD physics at hadronic colliders. It explains how the various processes involved in a hard scattering event are modelled in theoretical predictions and introduces the concept of jet algorithms. The second part of the chapter is dedicated more specifically to W + jets processes. Section 3.3 gives a brief review of the theoretical predictions available, while Section 3.4 explains the motivation for studying these processes. Section 3.5 gives an overview of the studies which are going to be performed at ATLAS and introduces the analysis discussed in this thesis. Finally, Section 3.6 describes the Monte Carlo samples used in the next chapters.

3.1 QCD Physics at hadronic colliders

The initial state of any scattering process at the LHC consists of a pair of protons. Its description requires knowledge of the proton structure, regardless of whether the hard process is of electroweak origin (like W production), or a strong process (such as jet production). Quantum ChromoDynamics (QCD) is the theory of the strong interaction of coloured particles. It de-

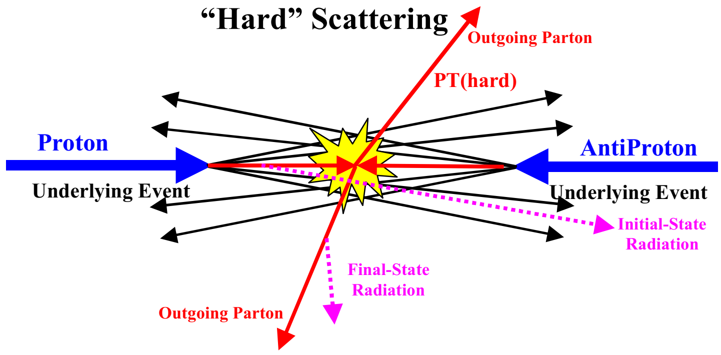

Figure 3.1: Pictorial representation of a hard scattering event in hadronic collision [20].

scribes the dynamics of partons (quarks and gluons) inside the proton and is therefore the starting point of any study of LHC physics. Our level of understanding of QCD processes, however, depends on the energy scale involved. Because of the non-Abelian nature of the $(SU)_3$ gauge symmetry group of QCD, the coupling of the strong interaction (α_S) becomes larger at decreasing energy. This implies that at high energies (short distances) quark and gluons interact very weakly and can be considered as free particles (this is the concept of asymptotic freedom). This is the perturbative regime, where calculations are accurate. On the other hand, at low energies (long distances) α_S is large and perturbative calculations are not applicable.

A collision event at the LHC evolves from the long distance protons in the beam, to the hard (short distance) scattering processes between two incoming partons and finally to the long distance outgoing states. Each of these steps involves very different scales (usually defined in terms of the momentum transfer squared, Q^2) and therefore different dynamics. Factorisation, the ability to isolate the independent phases of the collision, is an essential concept for the theoretical description of these phenomena. The short distance (large Q^2) interaction, which can be described by perturbative QCD (pQCD), is decoupled from the complexity of the proton structure and the final state hadron formation, which are non-perturbative processes. A hard scattering event at a hadron collider (see Fig. 3.1) can be summarised as follows:

- The partons involved in hard scattering processes are "picked out" from the incoming beam. The probability of finding a parton of a given type i, with momentum x in a proton probed with scale Q^2, is given by the parton distribution functions (PDFs), $f_i(x, Q^2)$. PDFs cannot be calculated in pQCD, but have to be obtained from global fits to experimental data. The evolution of the PDFs with the scale Q^2 is described by the DGLAP equations [21, 22].

- Before the scattering, partons can emit initial state radiation (ISR).
- The large momentum transfer scattering between the two incoming partons occurs. This process (referred to as partonic cross-section) can be evaluated at fixed order in pQCD.
- Gluons and quark pairs are added to the final state (final state radiation FSR). Emitted gluons are predominantly soft and collinear with the parent partons. FSR (as well as ISR) is still governed by pQCD and is described in Monte Carlo programs such as PYTHIA and HERWIG by a parton showering process. A parton shower evolves the partons produced in the hard process at a high energy scale, to partons at a lower energy scale, up to the point when non-perturbative effects arise.
- The final step of the evolution reorganises coloured final state objects into colour-singlet hadrons, which are observed in the detector as jets. This process, called hadronization, is non-perturbative and is based on phenomenological models tuned to data. Since hadronization is assumed to be independent of the hard process, it can be applied to different processes.
- Additional energy deposits from the underlying event (UE) and pile-up must also be accounted for. UE originates from the remnants of the original hadrons, which interact with each other since they are no longer colour-singlet states. Pile-up comes from multiple hadron collisions in the same data-taking period and increases with the instantaneous luminosity. Both of these give rise to additional energy deposits that tend to add energy to the jets originating from the hard scatter.

The overall cross-section of an hadronic final state is then given by the convolution of the partonic cross-section with the parton distribution functions and some transition function from the partonic final state to the hadronic observable. The partonic cross-section itself is determined by drawing all the Feynman diagrams for a given process, calculating the matrix element squared and integrating over the appropriate phase space.

3.2 Jet algorithms

To make comparisons between data and theoretical predictions it is necessary to relate the short distance physics (partons) to the long distance objects detected experimentally (charged particles and energy deposition in the calorimeter cells). This is the objective of a jet algorithm. A jet algorithm associates objects which are nearby in coordinate or momentum space. The input objects can be partons, particles or calorimeter clusters. The algorithm defines a set of rules which specify what objects have to be associated together into a jet (jet finding), and secondly defines how the 4-momenta of the objects have to be recombined to construct the kinematics of a jet (recombination scheme). Jet algorithms have been reviewed in several articles, for example [23, 24]. Here we summarise the theoretical and experimental attributes of an ideal jet:

- Infrared and collinear safety: the algorithm should be insensitive to the emission of extra soft and/or collinear radiation in the event.
- Order independence: it should yield a similar description of the hard scatter event, regardless of whether applied at parton, hadron or detector level.
- Detector independence: its performance should be independent of the signal source. Any effect due to detector segmentation and energy response or resolution should be corrected for by proper calibration.
- Environment independence: it should be stable under increase of activity in the event due to multiple interactions and pile-up.
- Fully specified and efficient: its implementation has to be clearly and completely defined. This includes the jet finding procedure, the algorithm parameters (for example the jet radius

parameter) and the recombination scheme. In addition, it should use minimum computing time.

Several algorithms exist on the market and can be broadly divided in two categories: cone and sequential recombination algorithms. The anti-k_T algorithm [25], which belongs to the second class, is the one used by default in ATLAS and adopted in this thesis. We therefore dedicate a short section to the description of this algorithm.

3.2.1 The Anti-Kt Algorithm

Sequential recombination algorithms in general (not only anti-k_T), start from a list of objects (for example pre-clusters), each with transverse momentum k_T:

- For a given pre-cluster i and pair of pre-clusters $i,\ j$ with $i \neq j$, the distances d_i and d_{ij} are computed:

$$d_i = k_{Ti}^{2p} \tag{3.1a}$$

$$d_{ij} = min(k_{Ti}^{2p}, k_{Tj}^{2p})\,\frac{\Delta R_{ij}^2}{D^2} \tag{3.1b}$$

  where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with $y_i,\ \phi_i$ being the rapidity [1] and azimuthal angle of particle i. D and p are two parameters of the jet algorithm. The former controls the size of the jet and the latter defines different jet algorithms: the value $p = -1$ corresponds to the anti-k_T algorithm.

- The minimum of all d_i and d_{ij} is found: if the smallest of these quantities is a d_i, the pre-cluster is identified as a jet and removed from the list of pre-clusters. If instead the minimum is a d_{ij}, pre-clusters $i,\ j$ are merged into a new pre-cluster (by adding their 4-momenta).

- The procedure is iterated until all pre-clusters have been removed.

[1] The rapidity y is defined as: $y = \frac{1}{2}\ln\left(\frac{\mathrm{E+p_z}}{\mathrm{E-p_z}}\right)$.

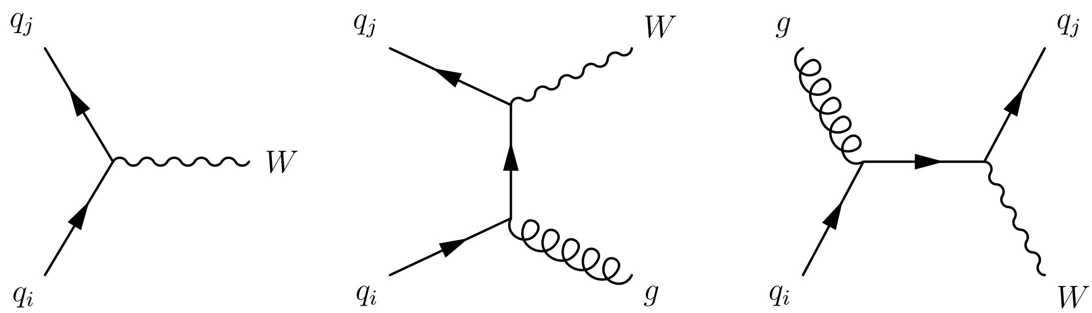

Figure 3.2: Some leading-order diagrams for W + 0 (left) and 1 (centre and right) jet production.

In practice, this means that soft particles are clustered with nearby hard particles long before any soft particles are clustered together. The results are hard jets with conical shapes which are not much affected by soft radiation. The anti-k_T algorithm has the great advantage of being fast as well as infrared and collinear safe. Furthermore, regular jet shapes facilitate the experimental jet calibration.

3.3 Predictions for W + jets: the current state of art

The previous sections introduced the basic concepts of QCD physics at the LHC. The last part of this chapter is dedicated to the production of W boson in association with jets. The hard process involves both an electroweak (EW) and strong interaction vertices. Some Feynman diagrams for the production of W + 0 and 1 jets are shown in Fig. 3.2. The leading-order (LO) diagram for W + 0 jets has only one EW vertex and requires two quarks in the initial state (for example $u\bar{d} \to W^+$). The process W + 1 jet has an additional α_S vertex and can be produced either by $qq \to Wg$ or by $qg \to Wq$ [2]. Any additional final state parton requires in an additional α_S vertex. A full treatment of the partonic cross-section calculation for W + jets events can be found in [26].

Theoretical predictions for W + jets have been calculated up to next-to-leading-order (NLO) in QCD for $N_{jets} \leq 3$. NLO predictions for processes with a vector boson + 3 jets have been a

[2]The second process, $qg \to Wq$, dominates at the LHC given the large contribution from the gluon PDF.

very recent and long-awaited achievement [27, 28, 29]. The main reason for extending theoretical calculations to NLO is the reduction of the unphysical dependence on the renormalization and factorization scales, reflecting a more reliable normalization of cross-sections and distributions. NLO calculations up to 2 jets are implemented in the MC program MCFM [30]. While being the most reliable, such predictions are parton-level cross-sections and cannot account for hadronization and underlying event.

The alternative for a direct comparison with data, are parton shower approaches which simulate final states at hadron-level. Several generators are on the market, but we will focus on ALPGEN [31], which is used by default in ATLAS for the process of interest. ALPGEN is a widely used MC generator which implements the exact matrix element (ME) LO calculation for multi-partonic final states. The hard ME calculation for a 2→n process is complemented by the parton shower (PS), implemented by interfacing ALPGEN to another generator such as HERWIG or PYTHIA. This approach, however, results in a complication: a given final state with (N+1) jets can either result from (N+1)-partons, or from a final state with N-partons and an additional jet from hard, large-angle emission during the shower evolution. A matching scheme [32] (MLM matching) is implemented to avoid double-counting (in kinematic regions where ME and PS overlap) and dead regions. Other ME-PS matching schemes (used by generators other than ALPGEN) exist, the most common alternative to the MLM matching being the so called CKKW [33] technique.

Although ME + PS predictions (such as those provided by ALPGEN), are the most suited to simulate final states similar to those observed in data, they suffer from large systematic uncertainties. The intrinsic uncertainties of these calculations can only be reduced by tuning the input parameters (such as the choice of renormaliztion scale or the parameters used for the matching scheme) to data. Therefore validation of such tools with data from the Tevatron and the LHC is essential to increase the accurancy of QCD calculations.

3.4 Why studying W + jets: history and motivations

Since its discovery in 1983 [7, 8, 10], the W boson has been extensively studied at LEP [34] and at the Tevatron [35, 36, 37, 38]. At the LHC, W boson production is considered as a precision benchmark for SM physics, for example for luminosity measurements [39] or to improve our knowledge of PDFs [40]. ATLAS has recently shown its very first measurement of the W inclusive cross-section at $\sqrt{s} = 7$ TeV [41]. As more data are recorded, the measurement will be refined and the physics programme will be extended to the measurement of the W charge asymmetry, the W/Z ratio and the production cross-section as a function of jet multiplicity.

When studying W boson in association with hadronic jets, the main interest lies in the QCD part of the interaction (rather than in the EW vertex) and therefore in the production of quarks and gluons which are observed in the detector as jets. The presence of a heavy vector boson ensures that the scale Q^2 of the interaction is large, thus making W + jets events excellent candidates for testing pQCD predictions. Furthermore, these processes form a significant background to several SM and BSM physics, such as top production, Higgs and Supersymmetry. Although the accuracy of theoretical predictions for processes which are background to new physics searches have significantly increased in the past years, precise measurement and validation of each background component separately is going to aid for any new discovery. As explained in [42], "claiming that Supersymmetry exist is far too important a conclusion to make it follow from the straight comparison against a Monte Carlo". Studying $W \rightarrow e\nu$ + jets events will therefore allow testing the current pQCD predictions and available MC tools which are going to be an essential piece of information for many searches for new physics phenomena.

Extensive studies of W boson production in association with jets have been performed at the D0 [43, 44] and CDF [45, 46] experiments. The latter, in particular, performed measurements of absolute cross-section for W plus up to 4 jets, as well as differential cross-sections in jet kinematic variables, which were compared to theoretical predictions in QCD. NLO calculation were found to reproduce the data reasonably well. However, the comparison is limited by the fact that NLO calculations do not include hadronization and underlying event and could therefore be somewhat misleading [47]. LO ME plus PS approaches were also compared to data. These were found to have some limitations in describing the differential jet p_T distributions, particularly in the low p_T region. The agreement with data also depended on what ME-PS matching scheme was used, ie. the MLM matching scheme was observed to perform less well than the alternative CKKW matching scheme.

Tevatron results have thus proven very important for validating the available MC tools. The LHC, however, is going to probe a new kinematic range and results from the Tevatron cannot simply be scaled to the LHC energy. Most processes at the LHC are going to be dominated by sea quark and gluon scattering, compared to valence quark scattering at the Tevatron. Furthermore, the LHC provides large phase space for additional hard QCD radiation, enhancing the importance of higher order terms. It is therefore necessary to test the pQCD predictions at the LHC energies by performing measurements similar to those carried out at the Tevatron. The next section describes the prospects for studying W+jets at the LHC and introduces the analysis presented in this thesis.

3.5 Prospects for studying W+jets at the LHC

With a cross-section of the order of 100 nb, W bosons are going to be copiously produced at the LHC. The production rates for several processes as a function of the centre of mass energy $\sqrt{s}$ are shown in Fig. 3.3. The W boson total cross-section (σ_W) increases by about a factor of 10 moving from the Tevatron to the LHC energies. In addition, the production in association with jets relative to the inclusive production is enhanced at the LHC. As an example, events with W plus two very hard jets ($p_T > 100$ GeV) will be produced as frequently as much softer W + 2 jets events at the Tevatron [47]. This illustrates why studying multi-jet final states and understanding pQCD is crucial at the LHC.

One of the ATLAS goals will be to measure the W cross-section as a function of the number

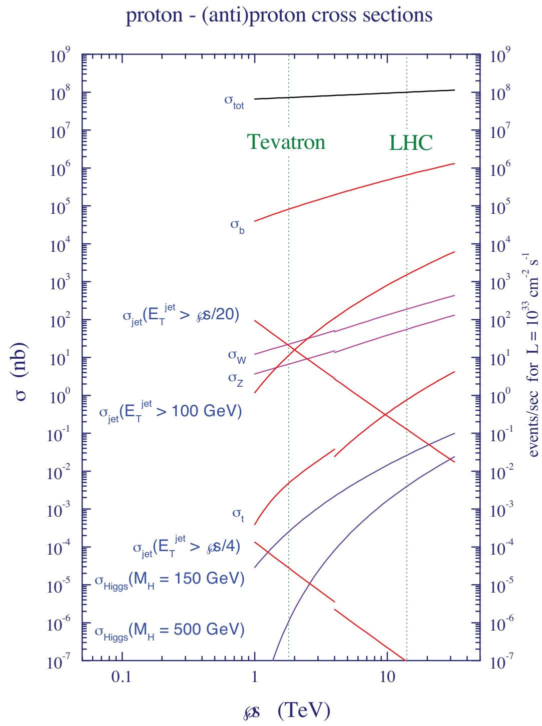

Figure 3.3: Cross-sections for a range of processes as a function of the centre of mass energy $\sqrt{s}$ [26].

of jets: $\sigma(W + N_{jets})$. The primary end-results would be hadron-level cross-sections, which can be compared to fixed order calculations at LO and NLO in QCD (such as those provided by MCFM) and to parton shower MC models (such as ALPGEN). Ratio measurements will also be interesting, in particular in the early data phase, since they lead to cancellation of several uncertainties. In addition to providing a test of QCD predictions, the following measurements could also be sensitive to new physics:

- $R_W = \dfrac{\sigma(W + N_{jets} + 1)}{\sigma(W + N_{jets})}$: this ratio is expected to be proportional to α_S. Signatures for BSM physics typically have a high number of jets and could therefore alter R_W;
- $R_{jets} = \dfrac{\sigma(W + N_{jets})}{\sigma(Z + N_{jets})}$: could provide sensitivity to new physics signals which show up differently in the single lepton (numerator) and di-lepton (denominator) channels.

This thesis will discuss techniques necessary for extracting the W + jets cross-section, as well as the above ratios, for the electron decay channel. The cross-section for any process is measured as:

$$\sigma = \frac{N - B}{\epsilon \cdot A \cdot L} \tag{3.2}$$

where N is the number of candidate events, B is the number of background events, A is the kinematic acceptance, ϵ is the event selection efficiency and L is the integrated luminosity. The relative uncertainty on the cross-section is given by the following formula:

$$\frac{\Delta\sigma}{\sigma} = \sqrt{\left(\frac{\Delta(N-B)}{N-B}\right)^2 + \left(\frac{\Delta\epsilon}{\epsilon}\right)^2 + \left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta L}{L}\right)^2}. \tag{3.3}$$

In the term $\frac{\Delta(N-B)}{N-B}$, N only contributes to the statistical error and B to the systematic uncertainties. The terms $\Delta\epsilon$, ΔA and ΔL are propagated as systematic uncertainties on the cross-section. The overall systematic uncertainty can then be rewritten as:

$$\frac{\Delta\sigma_{syst}}{\sigma} = \sqrt{\left(\frac{\Delta B}{N-B}\right)^2 + \left(\frac{\Delta\epsilon}{\epsilon}\right)^2 + \left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta L}{L}\right)^2}. \tag{3.4}$$

The individual terms of equation (3.2)-(3.4) and their uncertainties will be discussed in the next chapters.

3.6 MC Datasets

This thesis uses MC samples produced with different generators, as listed in Table 3.1. All samples were produced with a centre of mass energy $\sqrt{s}$ = 10 TeV [3]. Only in one exceptional case (indicated where appropriate in the respective chapter), some studies were performed with datasets generated with a centre of mass energy of 14 TeV. The GEANT4 [48] programme was used to provide a full simulation of the ATLAS detector. Pile-up is not included in these samples. As shown in Table 3.1, for some samples, a generator-level filter was applied to reject

[3] Although the LHC is currently operating at a centre of mass energy of 7 TeV, the full set of MC samples simulated with $\sqrt{s}$ = 7 TeV only became available when this analysis was near completion. Since it was not realistic to repeat all the studies in this thesis with the newest datasets, it was decided to have a fully consistent analysis with 10 TeV samples.

Process	Generator	$\sigma \cdot \epsilon_{EF}$(pb)	L_{Gen}(pb^{-1})	Comments
$W \to e\nu$ + 0-5 p	Alpgen + Jimmy	13249	$\geq$ 124	
W + bb + 0-3 p	Alpgen + Jimmy	14.6	$\geq$ 998	$p_T(b) > 20$ GeV $\Delta R(bb) > 0.7$
$W \to \tau\nu$ + 0-5 p	Alpgen + Jimmy	13233	$\geq$ 126	
$Z \to ee$ + 0-5 p	Alpgen + Jimmy	1206.1	$\geq$ 276	60 GeV $< M_{ee} <$ 200 GeV
$Z \to \tau\tau$ + 0-5 p	Alpgen + Jimmy	1211	$\geq$ 296	60 GeV $< M_{\tau\tau} <$ 200 GeV
$t\bar{t}$ (leptonic)	MC@NLO	205.5	7923.6	$N_l \geq 1$
$t\bar{t}$ (hadronic)	MC@NLO	168.1	4373.6	$N_l = 0$
top (Wt)	AcerMC	14.41	1385.3	
top (t chan.)	MC@NLO	14.46	1932.5	e channel
top (s chan.)	MC@NLO	0.76	10539.47	e channel
WW	Herwig	15.6	1585.45	$p_T^l > 10$ GeV, $\lvert\eta^l\rvert < 2.8$
ZZ	Herwig	1.36	7371.9	$p_T^l > 10$ GeV, $\lvert\eta^l\rvert < 2.8$
QCD-jets	Pythia	$1.03 \cdot 10^8$	0.095	$p_T^{jet} > 17$ GeV, $\lvert\eta^{jet}\rvert < 2.7$
$Z \to ee$	Pythia	1098.2	4805.70	$\lvert\eta_e\rvert < 2.8$, $M_{ee} > 60$ GeV
$W \to e\nu$	Pythia	10352.8	200.31	$\lvert\eta_e\rvert < 2.8$

Table 3.1: MC samples used in this thesis, with the effective cross-section $\sigma \cdot \epsilon_{EF}$ (given by the generator level cross-section times the event filter efficiency, if applicable) and the luminosity equivalent to the number of generated events (L_{Gen}). The ALPGEN samples are the combination of a number of sub-samples, eg. $W \to e\nu$ + 0-5 p (partons) is obtained by adding 6 sub-samples together. In this case, L_{Gen} is only quoted for the sub-sample with the smallest statistics. If a generator level filter is applied, this is indicated in the last column.

events likely to fail the analysis cuts, in order to optimize the use of computing resources. This is particularly important for QCD-jets, which have an extremely high cross-section. Despite an effective filter which retains about 7% of the generated events, the fully simulated sample corresponds to less than 0.1 pb^{-1} of integrated luminosity. For all other samples, the statistics available corresponds to $>$ 100 pb^{-1} of data.

The signal process and most of the backgrounds ($W \to \ell\nu$, $Z \to \ell\ell$) are generated with ALPGEN (see Section 3.1), interfaced with HERWIG/JIMMY for the shower evolution. While HERWIG models the hadronization, the JIMMY program [49] is used for the underlying event simulation. For all ALPGEN samples, the MLM matching parameters were set to ETCLUS = 30 GeV and RCLUS = 0.7. These parameters define the boundaries between the region of phase space generated by the ALPGEN ME and by the parton shower: the phase space described by the ME is defined by $\Delta R_{jj} >$ RCLUS and $e_{Tj} >$ ETCLUS, where ΔR_{jj} is the angular distance between two jets and e_{Tj} is the transverse energy of a jet.

The ALPGEN W/Z + n parton samples include all processes with u,d,s and c quarks, where all quarks are treated as massless. Although this is sufficient for most of the studies performed in this thesis, an additional process has to be considered for specific analyses involving tagging of jets containing b-hadrons (as in Chapter 5, Section 5.3.2). For such studies, the process $W + b\bar{b}$ + n partons is added to default signal samples $W \to e\nu$ + 0-5 partons.

Top pair production and single top production are generated with MC@NLO [50] interfaced with HERWIG/JIMMY and AcerMC [51] interfaced with PYTHIA. Finally, PYTHIA (version 6.417) is used as generator for di-boson, QCD-jets production, as well as for inclusive $W \to e\nu$ and $Z \to ee$ [4] events (used specifically for electron performance studies, see Chapter 6).

Additional packages are interfaced to the generators for specific processes. The decay of the τ-lepton is treated by the TAUOLA package [52]. The PHOTOS QED [53] package is used for the radiation of photons from charged leptons. PDFs are provided by the LHAPDF, the Les Houches accord PDF interface library [54]. LO MC generators (such as ALPGEN) are set to use the PDF set CTEQ6L1, whereas CTEQ6M is used for NLO generators (such as MC@NLO).

[4]More precisely, the process generated in the $Z \to ee$ PYTHIA and ALPGEN samples is the Drell-Yan $pp \to \gamma^*/Z \to ee$.

Chapter 4

W Boson Identification and Jet Selection at ATLAS

About 30% of W bosons produced at the LHC decay leptonically leaving a clear signature in the detector: a high transverse momentum isolated lepton and large missing transverse energy from the neutrino. Although the hadronic decay modes are enhanced relative to the leptonic ones, the significant background from QCD multi-jet production at hadronic colliders make their study unfeasible. The presence of an isolated electron or muon, instead, is an excellent discriminator against the multitude of pp interactions.

This thesis focuses exclusively on the electron decay channel. The following chapter describes the selection cuts applied to identify $W \to e\nu$ events and to subsequently divide the inclusive W sample into sub-samples according to their jet multiplicity. Sections 4.1 and 4.2 describe the ATLAS electron trigger selection and the offline reconstruction algorithms for electrons and missing transverse energy. The inclusive W sample selection is outlined in Section 4.3. Section 4.4 is dedicated to jet selection cuts. Backgrounds to W +jets are not covered in this chapter since they will be treated in detail in Chapter 5.

4.1 Electron Triggering and Offline Reconstruction

This analysis exploits the capabilities of the ATLAS detector in identifying high-p_T electrons to select a clean sample of W boson events, both online at trigger level and offline at reconstruction level. Before going into the details of the individual algorithms, which will be given in Sections 4.1.1 and 4.1.2, the features common to trigger and offline reconstruction will be described.

The Electromagnetic (EM) calorimeter and the Inner Detector (ID) are crucial for the reconstruction and identification of electrons. Segmented into three sections in depths for $|\eta| < 2.5$, the EM calorimeter provides precision measurement in the range matched by the ID (see Section 2.4.2). The first layer of the calorimeter with fine granularity provides a precise measurements of the η position and efficient γ-π^0 separation (see Section 4.1.2), while the second sampling contains most of the EM shower. A sketch of a barrel module with its three layers is shown in Fig. 4.1 left. The performance of electron reconstruction and identification is deteriorated in the transition region between barrel and end-cap calorimeters ($1.37 < |\eta| < 1.52$), because of the large amount of material in front of the active calorimeter layers (see Fig. 4.1 right). Whereas the energy of high-p_T electrons is best measured in the calorimeter, the η and ϕ direction are determined with higher accuracy in the ID. Furthermore, in the range $|\eta| < 2.0$ the TRT enhances electron identification capabilities providing better electron/pion separation.

Both ATLAS trigger and offline reconstruction algorithms for high-p_T electrons are seeded by an EM cluster in the calorimeter [1]. Cells are clustered together by a sliding window algorithm [55]. The algorithm sums together cells within a fixed size rectangular window and adjusts the window position such that its contained transverse energy is a local maximum. The window size is defined in terms of $N_\eta \times N_\phi$ elements of dimensions $\Delta\eta \times \Delta\phi$ and is different for trigger and offline (see Sections 4.1.1 - 4.1.2).

[1] A second offline reconstruction algorithm, optimised for low energy electrons and therefore not treated here, is seeded by Inner Detector tracks and subsequently finds a matching isolated energy deposit in the calorimeter.

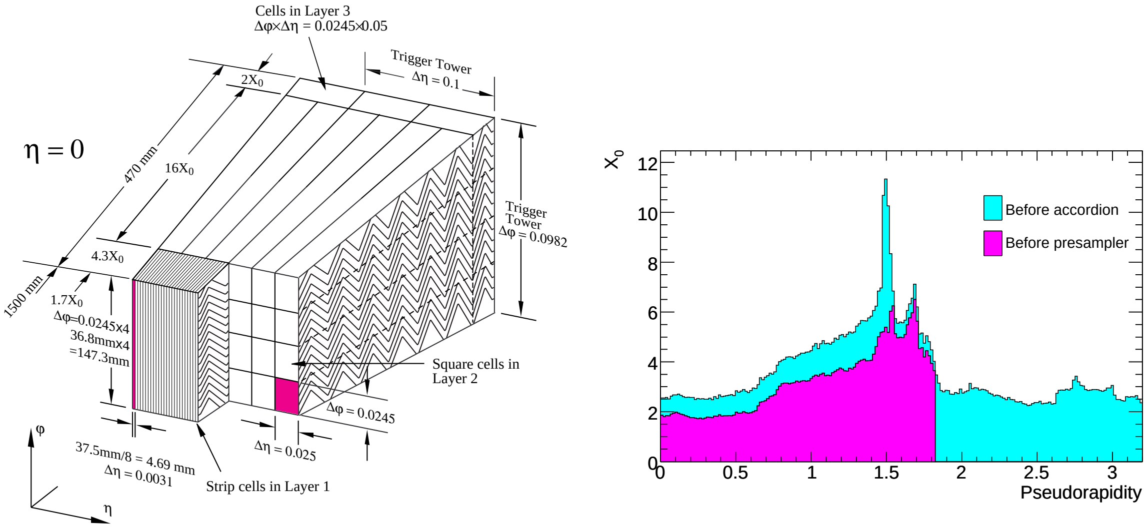

Figure 4.1: (Left) Sketch of a barrel module of the LAr calorimeter. The three layers in depth and their granularity in η and ϕ are clearly visible. Trigger towers are also shown. (Right) Amount of material in units of radiation length X_0 as a function of $|\eta|$ in front of the presampler and in front of the accordion itself. The spike around η = 1.5 corresponds to the crack region between the barrel and end-cap cryostats [15].

4.1.1 Electron Trigger

W boson events are selected using a single high-p_T electron trigger, optimised to deliver high signal efficiency and background rejection while keeping the rate to the allowed bandwidth. The ATLAS Trigger menu for the initial luminosity of $\mathrm{L} = 10^{31}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ foresees an unprescaled trigger (called $\mathrm{e20\_loose}$) requiring a single electron with p_T above a threshold of 20 GeV and satisfying some loose identification criteria. An overview of the ATLAS Trigger and its three levels (LVL1, LVL2 and EF) was given in Chapter 2. Here the $\mathrm{e20\_loose}$ trigger is described in more detail.

The first level of trigger (LVL1) has access only to calorimeter information. EM and hadronic trigger towers of dimensions $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ and depth equal to the full depth of the EM or hadronic calorimeter, are used by a sliding window algorithm to identify Regions of Interests (RoIs) corresponding to a candidate electron or photon. The LVL1 $\mathrm{e20\_loose}$ trigger is fired if the most energetic combination of two neighbouring EM towers in a $4\mathrm{x}4$ window passes the threshold of 18 GeV.

The High Level Trigger (LVL2 and EF) uses the full granularity of the calorimeter and also information from the ID. At LVL2, only regions of the detector around the LVL1 η and ϕ co-ordinates are considered. Thanks to an improved calibration, a tighter cut of 20 GeV is applied on the transverse energy. Tracking information from the ID allows to distinguish between electron and photons. Furthermore, variables related to the shower shape in the EM calorimeter and leakage into the hadronic calorimeter are used to reject background from jets (see Section 4.1.2). At the EF, information from the complete event and more precise calibration and alignment constants are available. The reconstruction algorithm run is very similar to the offline one.

Cuts applied at trigger level have been harmonised to the offline and are generally looser in order to keep high efficiency with respect to the offline selection (see Section 4.1.2). The variables used for the identification cuts are detailed in Table 4.1. No isolation cut is foreseen for the initial data-taking period, although it will be available at a later stage, once trigger variables are validated.

4.1.2 Electron Offline Reconstruction and Identification

Powerful electron identification is of utmost importance at the LHC for many physics analyses, from Higgs searches to SM physics studies such as the one described in this thesis. The expected ratio between isolated electrons and QCD di-jets in the range of 20-50 GeV is about 10^{-5}, two orders of magnitude smaller than at the Tevatron. Therefore, in order to achieve a reasonably clean electron sample in the range 20-50 GeV, ATLAS needs excellent electron identification capabilities, with a jet-rejection factor of the order of 10^{5}.

Reconstruction Algorithm

As explained earlier, the ATLAS electron reconstruction algorithm for high-p_T electrons is seeded by an EM cluster in the calorimeter identified by a sliding window algorithm. The cluster size depends on the detector region and on the particle type: electrons need a larger cluster size

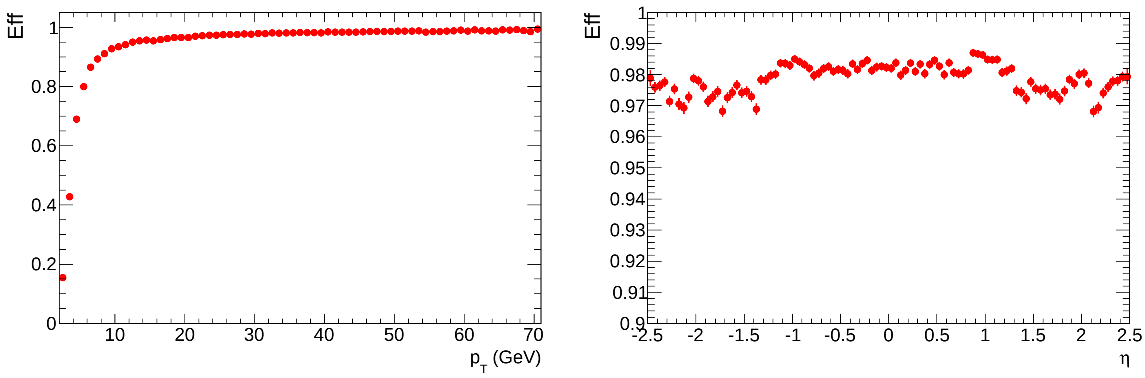

Figure 4.2: Offline reconstruction efficiency for electrons from W boson decays. (Left) Efficiency as a function of the generated p_T for electrons within the detector acceptance ($|\eta| < 2.5$). (Right) Efficiency as a function of the generated η for electrons with $p_T > 25$ GeV.

window than photons, because they are more likely to interact with upstream material and also bend in the magnetic field in the ϕ direction. In the barrel region, a window of size N_η x N_ϕ = 3x7 (3x5) cells is used for electrons (photons), whereas in the end-caps the size is 5x5 for both particle types. The reconstruction software builds different collections of clusters which are then used to identify electron and photon candidates. Once a reconstructed cluster has been identified, a matching track is searched for in the ID. A match is found if $|\Delta\eta_{\rm Cl-Tr}| < 0.05$ and $|\Delta\phi_{\rm Cl-Tr}| < 0.1$ [2], where $\Delta\eta_{\rm Cl-Tr}$ and $\Delta\phi_{\rm Cl-Tr}$ are the angular separation in η and ϕ between the cluster and the track extrapolated to the calorimeter. Any object with a track matched to a cluster is considered an electron candidate, therefore converted photons could also be reconstructed as electrons. This results in a very high efficiency for high-p_T electrons to be reconstructed, but also low purity due to photon contamination. Figure 4.2 shows the reconstruction efficiency versus η and p_T for electrons from a W boson decay lying within the detector acceptance, obtained from the ALPGEN W + jets samples. For transverse momenta above 25 GeV this efficiency is larger than 97%.

The cluster energy is obtained from the sum of the reconstructed energy in each of the three calorimeter layers and in the pre-sampler (where available) and is corrected for upstream energy losses, for lateral and longitudinal leakage. The electron/photon energy calibration is based on fully simulated MC information of energy losses inside the detector. Several test beam stud-

[2] Larger $\Delta\phi_{\rm Cl-Tr}$ values are allowed because of possible bremsstrahlung.

ies [56, 57] have allowed to improve significantly the understanding of the ATLAS detector geometry and its simulation. There are also several strategies for an in-situ measurement of the material budget with early data, using for example photon conversions [58, 59] or K^0_S [60]. The knowledge of the amount and the position of upstream material is of crucial importance since the presence of more material implies larger energy losses for electrons and higher numbers of photons converting, which in turns affects the energy measurement and the particle identification.

Electron Identification

A sample of reconstructed electron candidates [3] at the LHC is expected to contain a large number of fakes, with a signal-to-background ratio of 1/80 for electrons $E_T > 17$ GeV [58]. Isolated electrons, such as those from W and Z boson decay, can be distinguished from the large background of pions and jets by a number of discriminating variables that use calorimeter cluster or precise tracking information. A simple cut based identification criteria on these discriminating variables is foreseen for early data. Cuts have been optimised for high p_T isolated electrons in bins of η and p_T. Each identification cut can be applied independently, however a reference set of cuts have been defined to facilitate the analysis. We will refer to these as loose, medium and tight particle identification (PID) cuts. The variables used for electron identification are summarised in Table 4.1 and detailed here:

Loose:

This set of cuts is based exclusively on calorimeter information. It includes a selection based on the geometrical acceptance of the detector (limited by the precision range of the EM calorimeter and the tracker) and on shower shape variables from the second layer of the EM calorimeter, where EM showers deposit most of their energy. One of the shower shape variables is the shower width in the η direction (w_2, measured in a $\Delta\eta \times \Delta\phi = 3 \times 5$ window): electron/photon

[3] A reconstructed electron at ATLAS is an object with a calorimeter cluster loosely match to a track, as explained in the previous section.

showers are narrower than hadronic ones. It is also useful to define a lateral shower shape as the ratio in η of energy deposited in 3×7 cells to that in 7×7 cells $\left(R_\eta = \frac{E_{3\times7}}{E_{7\times7}}\right)$. Electrons and photons deposit most of their energy in the 3×7 core and will therefore typically have a ratio $> 80\%$. Finally, high energy jets deposit considerable energy in the hadronic calorimeter, contrary to electrons and photons which are contained in the EM calorimeter. Hadronic leakage, defined as the ratio of transverse energy in the first section of the hadronic calorimeter to that in the EM calorimeter, rejects such high energetic jets.

Medium:

In addition to the loose cuts, information from the tracker and the first compartment of the calorimeter is used. The first calorimeter layer with its fine granularity in rapidity is used to detect substructure within a shower and therefore helps discriminating against π^0s and ηs. Jets with $\pi^0 \to \gamma\gamma$ decays, for example, often have two local maxima. The asymmetry between the first and second largest energy deposit in the chosen window $\left(R_{Strips} = \frac{E_{1stMax} - E_{2ndMax}}{E_{1stMax} + E_{2ndMax}}\right)$ is a powerful discriminator, being close to one for electrons and photons, and tending to zero for photons from π^0 decays. After the calorimeter cuts, the contamination of charged hadrons is greatly reduced. The residual background is dominated by photon conversions and low multiplicity jets containing high-p_T π^0 mesons. Requiring a good quality track pointing to the calorimeter cluster helps reducing these backgrounds. Medium cuts include requirements on the minimum number of hits in the silicon detectors (Pixel and SCT), on the transverse impact parameter [4] and on track-cluster matching in η.

Tight:

The track quality is further refined by requiring the presence of hits in the vertexing layer (b-layer) of the Pixel detector, which has rejection power against electrons from photon conversions. Exploiting the powerful electron/pion discrimination power of the TRT (see Section 2.4.2), cuts on the number of hits in the TRT and on the ratio of high-threshold hits to the

[4] The transverse impact parameter is defined as the distance of closest approach of the track to the primary vertex point in the $r - \phi$ projection.

number of TRT hits rejects the dominant background from charged hadrons. Furthermore, the track-cluster matching is improved with additional cuts on the difference between the cluster and the extrapolated track position in ϕ. The ratio between cluster energy and track momentum ($\frac{E}{p}$) is also used. Relativistic electrons should have an energy-momentum ratio close to unity. However, this quantity is affected by bremsstrahlung which produces a tail towards high $\frac{E}{p}$ values. This happens because photons emitted collinearly will be clustered together with the parent electrons, whereas the momentum measured in the tracker is reduced. Furthermore, the track momentum resolution becomes poorer for higher energies, making $\frac{E}{p}$ inaccurate. In order to keep bremsstrahlung electrons and minimise the effect of reduced accuracy at high energy, the maximum cut on the $\frac{E}{p}$ is relaxed at p_T higher than 50 GeV.

Isolation:

In addition to the baseline selection cuts outlined above, an additional calorimetric isolation cut is available. An electron produces a localised energy deposit which is well separated from other energy deposits in the calorimeter. A fractional isolation variable is defined:

$$E_T^{\mathrm{cone20}} = \frac{E_T(\mathrm{R} < 0.2)}{E_T(\mathrm{Cluster})} \tag{4.1}$$

being the ratio of the transverse energy in a cone of radius 0.2 around the electron candidate, to the total transverse energy of the electron cluster. Calorimetric isolation, however, will not be included in the default electron identification cuts in the early phase, since it will first require validation with data. It should also be noted that the cuts applied at medium PID level (hadronic leakage and lateral shower shape) are correlated with isolation and therefore an effective isolation cut is already present.

Figure 4.3 shows some of the variables used for electron identification at trigger and offline level and their separation power between isolated electrons from Z boson decay and di-jet background.

Type	Description	PID	Trigger
Acceptance	$\|\eta\| < 2.47$	Loose	LVL1
Second Layer of EM Calo	Lateral shower shape (R_η) Lateral width (w_2)	Loose Loose	LVL2 EF
Hadronic Leakage	Ratio of E_T in first scintillator of hadronic calorimeter to E_T of EM cluster	Loose	LVL2
First Layer of EM Calo	Total shower width Difference between 1st and 2nd largest energy deposit (R_{strips})	Medium Medium	EF EF
Tracking	Number of pixel hits ≥ 1 Number of pixel and SCT hits ≥ 7 Transverse impact parameter < 5.0 mm Number of pixel hits in b-layer ≥ 1	Medium Medium Medium Tight	EF EF EF
Track Match	$\|\Delta\eta\| = \eta_{Calo}^{strips} - \eta_{Track} < 0.01$ $\|\Delta\phi\| = \phi_{Calo}^{middle} - \phi_{Track} < 0.02$ Energy-momentum match $\left(\frac{E}{p}\right)$	Medium Tight Tight	LVL2 LVL2 n/a
TRT	Number of TRT hits Ratio of high threshold to all TRT hits	Tight Tight	n/a n/a
Calorimeter Isolation	Ratio of E_T in a cone $\Delta R < 0.2$ to the total transverse cluster energy	Extra	n/a

Table 4.1: List of variables used for loose, medium and tight electron identification criteria and at different trigger levels (LVL1, LVL2 and EF). Only cut values not optimised in bins of electron η and p_T are shown.

4.2 Missing Transverse Energy Reconstruction

In the decay of a $W \to e\nu$, the neutrino escapes detection, since it interacts only weakly with matter. The presence of a significant amount of missing transverse energy (E_T^{miss}) in the event is therefore evidence for a neutrino. However, fake missing energy can arise from different sources: for example, limited coverage of the detector, dead regions, finite detector resolution, noise and large fluctuation in the jet energy measurements.

The key sub-detector for the measurement of E_T^{miss} is the calorimeter. Its hermeticity, ensuring a coverage in the full azimuthal angle ϕ and in the forward region in rapidity up to $|\eta| < 4.9$, and its thickness (11 interaction lengths at η = 0.0), are crucial ingredients for a good

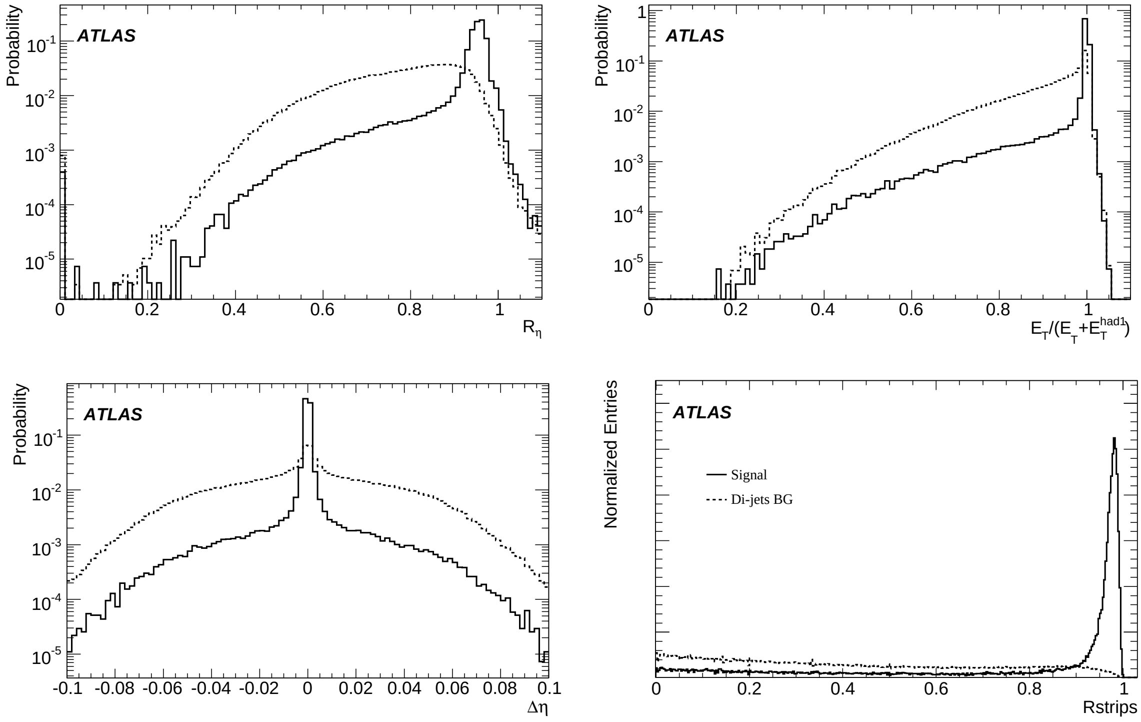

Figure 4.3: Variables used for electron identification: lateral shower shape R_η (top left), ratio between the transverse energy of the electron candidate and the sum of this transverse energy and that contained in the first layer of the hadronic calorimeter (top right), difference in η between cluster and extrapolated track position (bottom left) and asymmetry between first and second largest energy deposit R_{strips} (bottom right). Distributions are shown for electrons from $Z \rightarrow ee$ (solid line) and di-jets (dotted line) [58].

$E_T^{\rm miss}$ measurement.

$E_T^{\rm miss}$ is calculated from the x and y components of three terms, the calorimeter term (${E_{x,y}^{\rm miss}}^{Calo}$), the cryostat term (${E_{x,y}^{\rm miss}}^{Cryo}$) and the muon term (${E_{x,y}^{\rm miss}}^{Muon}$):

$$ {E_{x,y}^{\rm miss}}^{Final} = {E_{x,y}^{\rm miss}}^{Calo} + {E_{x,y}^{\rm miss}}^{Cryo} + {E_{x,y}^{\rm miss}}^{Muon}. \tag{4.2} $$

The calorimeter term is the sum of the total transverse energy deposited in all calorimeter cells. To limit the amount of fake $E_T^{\rm miss}$, it is crucial to suppress noise in the calorimeter. This is done when clustering cells together into 3-dimensional topological cluster, called TopoClusters [61],which include noise suppression (see Section 4.4.1). The energy is then calibrated through a “refined calibration” procedure where each cell is associated with a parent reconstructed high-p_T object and calibrated accordingly. So, for example, cells corresponding to an electron or photon are calibrated at the EM scale, whereas cells corresponding to jets are calibrated at the hadronic scale. TopoClusters not associated with any high-p_T objects are also used and calibrated with the global calibration scheme.

The cryostat term represents the corrections for the energy lost in the cryostat which houses the LAr calorimeters. The lost energy is inferred from the energy deposited by all reconstructed jets in the event in the last layer of the LAr calorimeter and the first layer of the hadronic calorimeter, weighted by an appropriate calibration factor.

The muon term is calculated from the momentum as measured in the muon spectrometer. This avoids any double counting, since energy lost in the calorimeter is already included in the calorimeter term [5].

The performance of the $E_T^{\rm miss}$ measurement can be evaluated in terms of linearity and resolution. The linearity, defined as the difference between true and reconstructed $E_T^{\rm miss}$, normalised to the true value, is expected to be within 1% for $W \to e\nu$ events using the refined calibration described above. The resolution of the two components of $E_T^{\rm miss}$ is shown in Fig. 4.4 [15] as a function of the total transverse energy deposited in the calorimeter $\sum E_T$ (the resolution is the same for the $E_{x,y}^{\rm miss}$ components and is hence combined into a single plot). The resolution is estimated in each bin from the width of the $E_{x,y}^{\rm miss} - E_{x,y}^{\rm miss}{}^{True}$ distribution, as obtained from a Gaussian fit. It can be seen that the $E_T^{\rm miss}$ resolution is affected by hadronic activity in the event (measured as $\sum E_T$). The best fit to the curve is $\sigma = 0.53\sqrt{\sum E_T}$. More details about the performance of $E_T^{\rm miss}$ reconstruction are given in [61] and [15].

4.3 $W \to e\nu$ Event Selection

In this analysis, W bosons decaying into the electron channel are selected according to the criteria suggested in [62]. In view of an early data analysis, priority is given to robustness

[5] In the refined calibration, the measurement of the energy lost in the calorimeter already accounts for the electron/muon signal ratio being different from one.

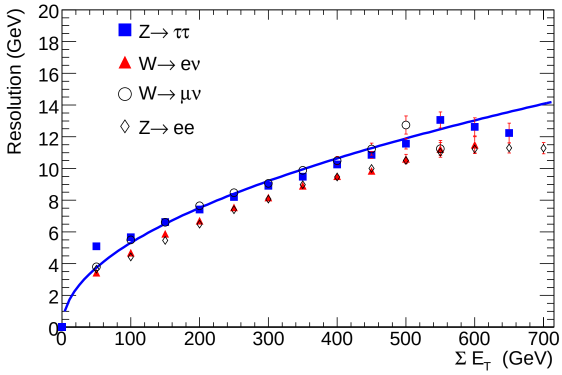

Figure 4.4: Resolution of the two components on E_T^{miss} as a function of the total transverse energy $\sum E_T$ [15].

and simplicity, over optimisation of the signal to background ratio. The selection cuts can be summarised as follows:

- One electron in the event satisfying:
  - Acceptance:

    $|\eta| < 2.46$, excluding the crack region $1.37 < |\eta| < 1.52$

    $p_T > 25$ GeV
  - Identification: Medium PID cuts
- Missing Transverse Energy: $E_T^{\text{miss}} > 25$ GeV
- Trigger e20_loose to be fired.

The acceptance cuts are based on the kinematics of the $W \rightarrow e\nu$ decay and the geometrical acceptance of the detector. The so called "cracks" ($1.37 < |\eta| < 1.52$) correspond to the transition region between barrel and end-cap calorimeters (see Fig. 4.1 right). Because of the large amount of passive material, electron efficiency and background rejection are significantly degraded and it was therefore chosen to exclude this region from the analysis. It should also be noted that no isolation requirement is applied on the electron, neither at trigger nor at offline level. Its impact on the signal efficiency and background rejection will be nevertheless shown in this and in the following chapter and arguments for and against an isolation cut will be presented.

	Ele Acceptance	Ele PID	Trigger	E_T^{miss}	Ele Iso
$W \to e\nu$	575579 ± 759	518052 ± 720	515848 ± 718	451638 ± 672	447448 ± 669
$\epsilon(\%)$	n/a	90.01 ± 0.04	99.57 ± 0.01	87.55 ± 0.05	99.07 ± 0.01

Table 4.2: Number of signal events after selection cuts and efficiency of each cut with respect to the previous one. Events are normalised to 100 pb^{-1} of integrated luminosity. Errors correspond to the statistical errors expected for the given luminosity.

The effect of the selection cuts on the signal is investigated, by applying each cut individually in the following order: electron acceptance cuts, electron identification cuts, trigger and missing transverse energy requirement. Additionally, also an electron calorimetric isolation cut is applied. Table 4.2 shows the number of $W \to e\nu$ events after each individual cut, normalised to an integrated luminosity of 100 pb^{-1}. The efficiency of each single cut relative to the previous one is also shown. The electron identification criteria have a global efficiency of 90% for electrons in the kinematic range of interest and the trigger $\mathrm{e20\_loose}$ is almost 100% efficient relative to the offline cuts. More detailed studies about the trigger and offline identification efficiencies versus kinematic variables will be shown in Chapters 6 and 7. The isolation requirement also has a negligible effect on the signal statistics, being 99% efficient.

In this section the $W \to e\nu$ inclusive sample selection has been discussed. The next section is dedicated to jet counting in order to partition the inclusive sample into jet multiplicity bins.

4.4 Jet Selection

Hard scattered partons, whose properties can be predicted by perturbative QCD, manifest themselves as jets at the experimental level. In order to study the properties of the particle interactions in the hard process it is therefore necessary to relate the kinematics of the jet with those of the partons in the hadronic final state. The reconstruction of the initial parton momentum is the ultimate goal of the jet energy measurement. This is a complex procedure and is often a source of large systematic uncertainties in physics measurements like the present analysis (see

Chapter 8). Hadronic calibration is essential for precise energy measurement, especially in a non-compensating calorimeter (as in ATLAS), whose response to hadrons is non-linear with energy.

In the following section, the full jet reconstruction sequence and analysis selection are described. Before going into the details, which are given in Sections 4.4.1 and 4.4.2, a brief overview is given. The jet reconstruction flow starts by combining together calorimeter cells to form objects with meaningful 4-momenta which are then input to the jet finding algorithm. The raw energy has to be corrected for detector effects back to the particle level, ie. to the energy that would be obtained by running the jet algorithm on all true momenta of the final state particles in the event. This is done through an hadronic calibration procedure which is based on MC corrections. Once sufficient data will be available, additional in-situ jet energy scale corrections will allow refinement of the MC-based calibration [61]. At this point we enter the analysis domain, where the jet definition is further refined by specifying the phase-space of interest. In order to recover the parton-level energy, additional analysis dependent corrections are needed: these account for physics effects such as fragmentation, underlying event and pile-up. In this analysis, however, we do not attempt to correct back to parton level and we solely rely on the MC-based corrections.

4.4.1 Calorimeter jet reconstruction and calibration

Topological Cell Clusters:

The reconstruction of jets starts in the calorimeter by identifying 4-momentum objects which are then the input to the jet finder. The input objects used in this analysis are topological clusters [55], 3-dimensional energy deposits representing the shower developed by particles entering the calorimeter. The cluster algorithm starts from seed cells with a signal-to-noise ratio (Γ) above a threshold $|\Gamma| > 4$ and adds neighbouring cells which satisfy a secondary threshold $|\Gamma| > 2$. Finally, a last ring of neighbouring cells is added if the signal-to-noise ratio is $|\Gamma| > 0$. In a second step, local maxima are searched for and clusters are re-arranged (split) around these maxima. The advantage of topological clustering is that it suppresses noise from electronics and pile-up, thanks to the signal-to-noise 4/2/0 configuration.

Local hadronic calibration:

The raw signal from the ATLAS calorimeter is at the basic electromagnetic scale. The goal of hadronic calibration is to correct the energy of the jet back to the hadronic scale, accounting for non-linearities due to the non-compensating calorimeter, dead material, cracks and particle deflection in the magnetic field. The calibration can occur at jet level (global calibration) or at cluster level (local calibration) before the jets are formed. The latter will be briefly described, since it is adopted in this analysis (see [63] for a detailed description):

- The calibration procedure starts classifying clusters as electromagnetic-like or hadronic-like depending on shower shape variables. Hadronic clusters are subject to a cell weighting procedure to correct back to the true energy deposit in each cell. The cell weights account for the invisible energy (energy released by non-ionising processes, such as nuclear binding) and the escape energy (energy leaving the volume in form of non-interacting or escaping particles such as neutrinos and muons). The weights depend on the cluster location, energy and the cell signal density.
- Out-of-cluster corrections are applied at cluster level. These corrections account for energy deposited in the calorimeter but outside the cluster due to the noise thresholds applied in topological clustering.
- Both electromagnetic and hadronic-like clusters are corrected for energy deposited outside the active zones of the calorimeter, eg. in the cryostat or in gap regions.

Although such corrections restore the cluster energy to the hadronic scale, additional inherent jet corrections should be applied to account for the magnetic field bending particles out of the jet and for particles absorbed in the ID material. At the time of writing, these corrections were still under development. All calibration and corrections are based on simulations of single neutral and charged pions and do not use the jet context. However, they rely on the correct description of the detector and of the hadronic shower development and jet fragmentation which will have to be validated with data.

Jet finding algorithm:

The sequential recombination algorithm anti-k_T, outlined in Chapter 3, Section 3.2.1, is used to combine the cluster 4-momenta (calibrated at the hadronic scale) together into jets. The choice of jet size should be dictated by the experimental considerations. Generally speaking, a large jet size is preferred when dealing with low jet multiplicities and low p_T jets, which are therefore also broader, to minimise out of cone corrections. When studying high jet multiplicities, instead, a smaller cone size is needed so not to limit the phase-space. The analysis of W + 2 jets events, where the two jets recoil against the boson and can be rather collimated, prefers small jet size. In these studies, the parameter D (see eqn. (3.1) in Chapter 3) was chosen to be $D = 0.4$, also in view of measurement at gradually higher energies and higher multiplicities.

4.4.2 Jet Counting

The reconstruction and calibration procedure described in Section 4.4.1 is performed within the official ATLAS jet reconstruction software. We now enter the analysis domain. This section is dedicated to the description of jet selection cuts used in this thesis.

Both theoretical and experimental considerations have to be taken into account when choosing a procedure for counting hadronic jets: the jet definition must be well modelled in the theory for a meaningful data-to-theory comparison and the measurement must be limited to an experimentally well understood phase-space for systematic uncertainties to be minimised.

The selection of hadronic jets used in this analysis can be summarised as follows:

- Choice of jet algorithm anti-k_T with parameter D=0.4, as described in Section 4.4.1
- Overlap removal with electrons from W boson decay

- Counting jets at detector level with:
  - Transverse momentum $p_T > 30$ GeV
  - $|\eta| < 3.1$.

  In addition, event selection cuts based on jet-jet and electron-jet separation are applied:

- Reject events where the minimum separation between the candidate electron from W boson decay and any selected jet is $\Delta R_{ej} < 0.6$, where $\Delta R = \sqrt{\Delta\eta^2_{ele-jet} + \Delta\phi^2_{ele-jet}}$.
- Reject events where the minimum separation between two jets is $\Delta R_{jj} < 0.6$.

Each of these selection cuts will be discussed below, giving the theoretical and experimental motivations that justified the choices made. When discussing the experimental motivations we will present jet performance studies on the following quantities:

- Jet Reconstruction Efficiency, defined as

$$\varepsilon = \frac{\#\ \text{of truth particle jets matched to a reco jet}}{\#\ \text{of truth particle jets}} \tag{4.3}$$

  where a MC truth jet [6] is matched to a reconstructed jet if the radial distance between them $\Delta R_{min} < 0.2$, following the prescription used in [2].

- Energy Response, defined as the relative energy difference between truth and reconstructed jets

$$\frac{\Delta E}{E} = \frac{E_{\text{Reco}} - E_{\text{Truth}}}{E_{\text{Truth}}}. \tag{4.4}$$

  The ΔE distribution is fitted to a Gaussian function, whose mean (μ) and width (σ) are the energy scale and resolution respectively.

[6] A MC truth jet is the particle-level jet, ie. as obtained by running the anti-k_T algorithm on the final state particle true momenta.

Electron-Jet Overlap Removal

The clustering procedure described in Section 4.4.1 identifies electrons as jets. Therefore it is essential to remove the candidate electrons from W boson decay when performing jet counting. There are several ways to treat the overlap between jets and electrons, intervening at different stages in the reconstruction flow (at cell, cluster or reconstruction object level):

- Cell Removal: After having removed the electron cells, re-run topological clustering and re-build the jets;
- Cluster Removal: Remove the electron topocluster and re-build the jets;
- Jet Removal: Remove closest jet to the electron as being the electron-jet.

In this study the third approach of jet removal is used. This approach has the advantage of being simpler and easier to model in the theoretical predictions. The disadvantage is that if a true jet lies very close to the electron, both the true jet and electron energy reconstruction might be deteriorated and the jet might even be lost. The additional requirement of a minimum separation between electron and jet (see later in this section) ensures that events with poorly reconstructed jets or electrons are excluded from the analysis. The overlap removal is performed before applying any jet selection based on η or p_T. The closest jet to the electron is removed only if the distance between this and the electron is within a cone of radius $\Delta R < 0.2$. Figure 4.5 shows the ΔR between electrons and jets versus the ratio of their transverse energy $\frac{E_{T\text{ele}}}{E_{T\text{jet}}}$ for any electron-jet combination (left) and for the closest jet to the electron only (right). The overlap between electrons and jets is clearly visible around $\frac{E_{T\text{ele}}}{E_{T\text{jet}}} = 1$ for $\Delta R < 0.2$, justifying the choice of $\Delta R = 0.2$ for the overlap removal. The data points at small ΔR and $\frac{E_{T\text{ele}}}{E_{T\text{jet}}} < 0.6$ are the result of some hadronic activity nearby the electron, which shifts the reconstructed jet further away from the electron and to higher energy.

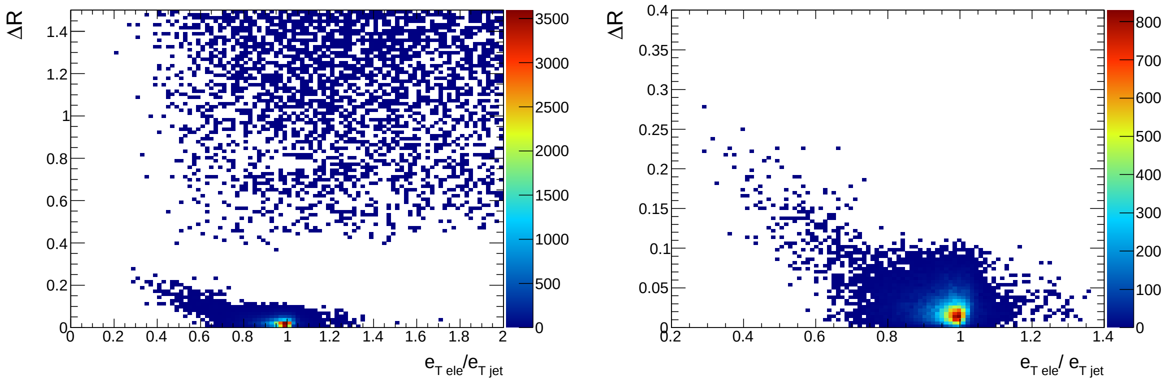

Figure 4.5: Distance in ΔR between electrons and jets versus ratio of electron E_T to jet E_T: for any jet in the event (left) and for closest jet to the electron (right). Electrons are selected according to the cuts described in Section 4.3. Jets are clustered with the anti-k_T algorithm with parameter D=0.4. No p_T or η cuts are applied on the jets.

Jet p_T Threshold

The choice of the jet p_T threshold has to be a compromise between theoretical and experimental considerations and the statistics available. Considering the theoretical side, NLO calculations are affected by logarithmic divergencies at low energy thresholds: $ln(E/\mu)$, where μ is of the order of 100 GeV ($\sim W$ boson mass). Non-perturbative effects (fragmentation and underlying events) are also more important at low energy. Whereas fragmentation tends to decrease the jet energy by out-of-cluster losses, the underlying event has the opposite effect since it increases the hadronic activity in the event. Monte Carlo studies in [64], using the PYTHIA 6.403 tune (see Fig. 4.6) showed that effects from fragmentation and underlying event are expected to cancel for jets with $p_T > 40$ GeV up to a residual correction at the percent level [7]. Between 30 and 40 GeV, underlying event corrections dominate, increasing the jet energy by less than 10%. Although the underlying event model will need validation with data, these studies give an idea of its impact on the measurement. Pile-up will also affect the measurement, particularly at low p_T. First of all, jets from additional p-p collisions can pass the jet selection, inflating the jet multiplicity. Furthermore, particles from pile-up interactions can add energy to the jets or affect the $E_T^{\rm miss}$ calculation. Since the MC samples used in this thesis do not include pile-up,

[7] The studies in [2] were performed on $Z \to \mu\mu$ Monte Carlo samples generated with PYTHIA 6.403 and with a Cone jet algorithm of radius R=0.4.

its effect on the W + jets cross-section measurement is not treated here. However, systematic uncertainties from pile-up will need to be carefully studied in the future as they will become gradually more important, as the LHC instantenous luminosity is increased.

From these arguments, the preference would be to choose a high p_T threshold of $\sim$ 40 GeV. However the statistics decrease dramatically with increasing threshold on the jet transverse momentum. Figure 4.7 (top) shows the number of expected events for the bin $N_{jets} = 2$ as a function of the p_T threshold: more than 60% in statistics is lost cutting at 40 GeV rather than at 20 GeV.

Another important consideration is the MC sample available. In the generation of the ALPGEN W + jets samples, the MLM matching parameters RCLUS and ETCLUS (see Chapter 3, Section 3.6 for the definition), which effectively define the boundary between the parts of the phase space generated by the matrix element and by the parton shower, were set to 0.7 and 20 GeV respectively. This results in an effective generator level threshold on the parton p_T of 20 GeV. In order to evaluate the threshold effect, a simple test was performed on the W + jets ALPGEN sample. An artificial cut of 30 GeV was applied on the parton transverse momentum and the p_T distribution of the associated jets at reconstruction level was plotted. Figure 4.7 (bottom) shows a comparison between the p_T distribution of jets matched to partons with transverse momenta larger than 20 GeV and 30 GeV. The ratio between the two distributions approaches unity only for jets with p_T $\sim$40 GeV, indicating that a parton level cut with a given threshold ($p_T^{Gen} > X$ GeV) has effects at detector level up to energies 10 GeV above the threshold ($p_T^{Rec} \sim X + 10$ GeV). It is therefore expected that the available ALPGEN samples are unbiased from about jet $p_T > 30$ GeV.

Considering all of the above, a minimum jet p_T threshold of 30 GeV was chosen.

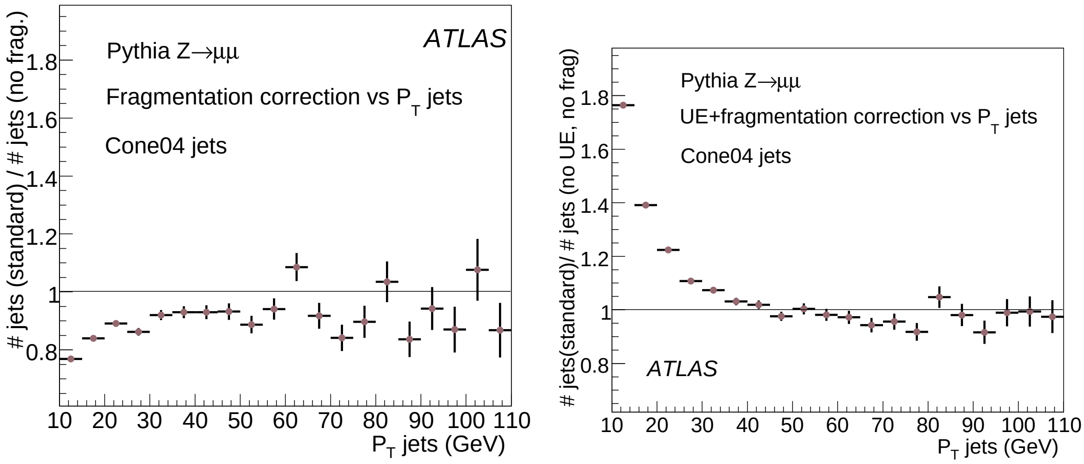

Figure 4.6: Effect of fragmentation (left) and fragmentation plus underlying event (right) as a function of the jet p_T in $Z \to \mu\mu$ + jets events, extracted using PYTHIA 6.403 and the jet Cone algorithm with radius R=0.4. [65]

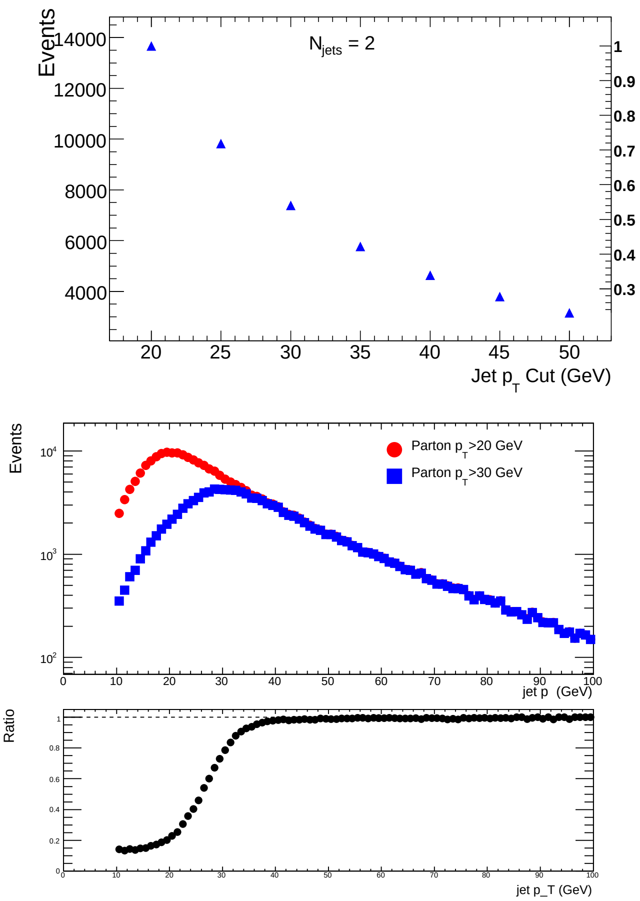

Figure 4.7: (Top) Statistics expected in W + 2 jets events from 100 pb^{-1} of data, as a function of the threshold on the jet p_T. The left axis shows the number of events, whereas the right axis shows the percentage of statistics, taking as reference (100%) the number of events expected for a jet p_T threshold of 20 GeV. (Bottom) Bias introduced on reconstructed jets by a parton level p_T cut. The top plot shows the reconstructed p_T spectra of jets matched to a parton above 20 and 30 GeV and the bottom plot is the ratio between the two distributions.

Jet η Cut

The jet reconstruction performance strongly depends on the detector geometry and therefore the jet pseudo-rapidity. If a jet falls in a poorly instrumented region of the detector, the reconstruction efficiency and energy resolution will be degraded. The presence of upstream material in front of the calorimeter also affects the energy reconstruction: low energy particles that interact with detector material or that are deflected by the ID magnetic field never reach the calorimeter resulting in a net energy loss. Figure 4.8 shows the jet performance as a function of η for jets with p_T greater than 30 GeV. The top left plot shows the jet reconstruction efficiency and the top right plot a Gaussian fit to $\frac{\Delta E}{E}$ (as defined in eqn. (4.4)) in two different η regions. The bottom plots show the energy scale and resolution obtained from a fit to $\frac{\Delta E}{E}$ as a function of η.

Both the efficiency and the energy response are deteriorated around $\eta = 3.2$. This region corresponds to the transition between the EM and hadronic calorimeter and the forward calorimeter. The performance deterioration starting at $\eta = 3.1$ is an edge effect due to the size of a jet: when a jet is close to the crack region it loses part of its energy into it. It was therefore decided to count only jets with $|\eta| < 3.1$. The fact that the jet energy is systematically underestimated ($\frac{\Delta E}{E} < 0$) is not worrisome: the local calibration corrects single clusters back to the hadronic scale, but additional corrections are needed to recover the true jet energy (see Section 4.4.1). Such scale factors were not available at the time of this thesis writing, but will be applied with data.

Minimum Separation between Electrons and Jets

When a real jet (not an electron jet) is very close to an electron, both the jet and the electron reconstruction are affected. In order to minimise the energy overlap, it is preferable to require a minimum separation between the electron from the W boson decay and high energetic jets. The choice of separation radius depends on the jet size. Figure 4.9 (left) shows that the jet reconstruction efficiency (computed after electron-jet overlap removal and considering only jets

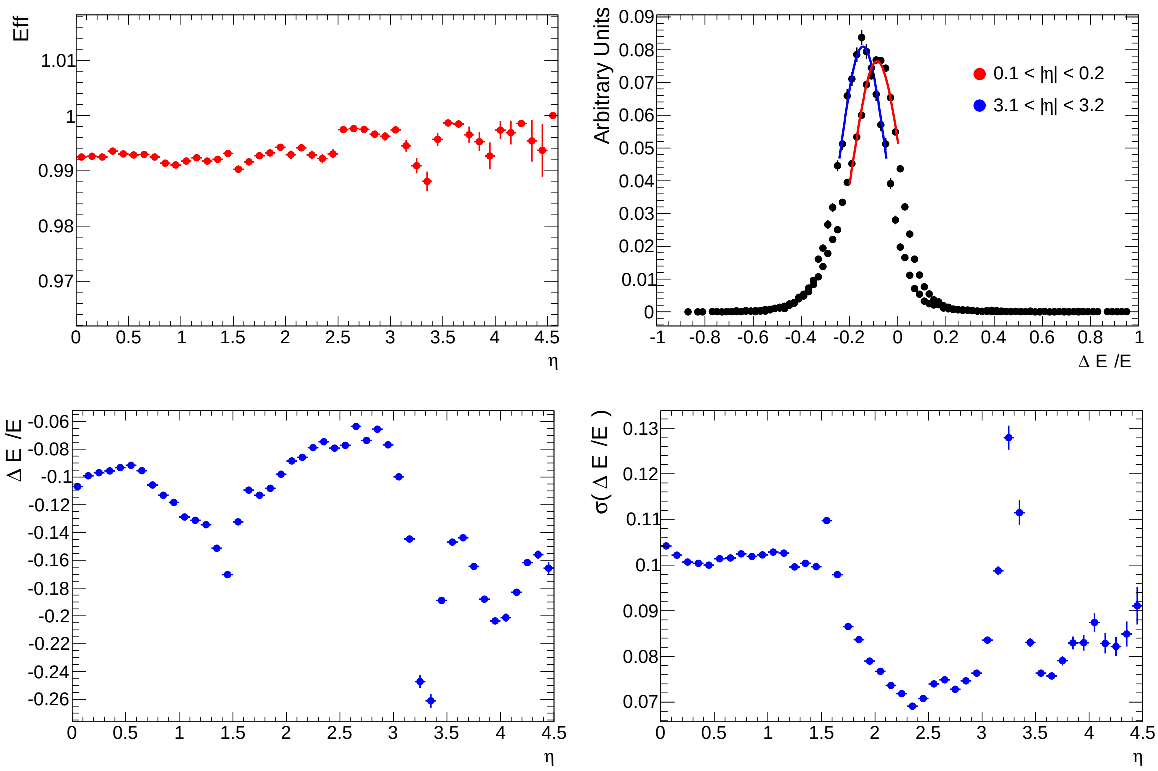

Figure 4.8: (Top Left) Jet reconstruction efficiency as a function of jet η for jets with $p_T > 30$ GeV. (Top Right) Fit with a Gaussian function to $\Delta E/E$, as defined in eqn. (4.4), for two different η bins. The bottom plots show the energy scale (left) and resolution (right), obtained as the mean and width of the Gaussian fit.

with p_T above 30 GeV) is affected when $\Delta R_{ej} < 0.6$. The energy response is also deteriorated: the right plot of Fig. 4.9 shows that for jets close in ΔR to an electron, $\frac{\Delta E}{E}$ has a low tail which is clearly non-Gaussian. It will be shown in Chapter 7 that the electron reconstruction is also deteriorated for $\Delta R_{ej} < 0.6$, suggesting that $\Delta R_{ej} = 0.6$ is a suitable value for the electron-jet minimum separation cut. Events where a high energetic jet ($p_T > 30$ GeV) and the selected electron are closer than this distance are rejected and not further considered for the analysis.

Minimum Separation between Jets

Similarly to the electron-jet case, jet performance is also affected when two jets are close to each other. The jet reconstruction efficiency and energy response as a function of ΔR_{jj} are shown in Fig. 4.10. The efficiency drops in the range $0.4 < \Delta R_{jj} < 0.6$. Again, for $\Delta R_{jj} < 0.6$ the jet energy resolution is significantly affected.

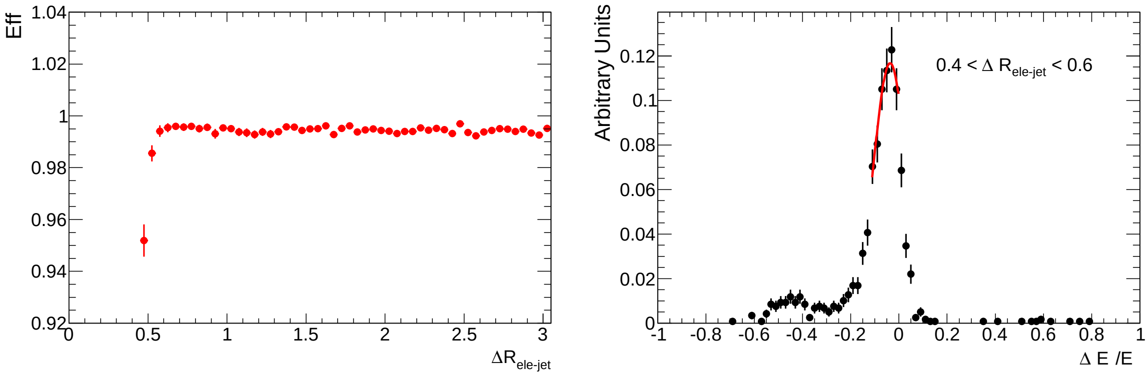

Figure 4.9: (Left) Jet reconstruction efficiency as a function of the distance ΔR to the electron from the W boson, for truth jets with $p_T > 30$ GeV. (Right) Energy response for jets which are closer than $\Delta R = 0.6$ to an electron.

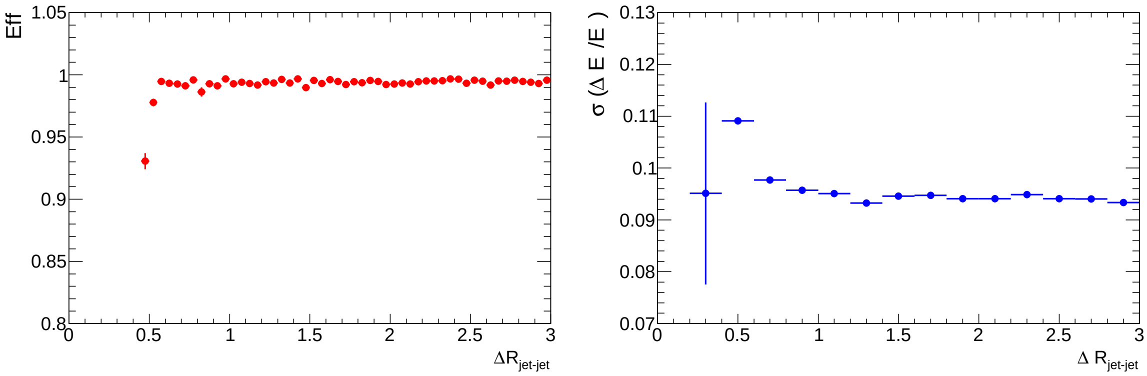

Figure 4.10: Jet reconstruction efficiency (left) and energy resolution (right) as a function of the distance ΔR to the closest jet, for truth jets with $p_T > 30$ GeV.

Jet performance was also investigated for different multiplicity bins, see Fig. 4.11. A small dependence on jet multiplicity can be seen: the jet reconstruction efficiency as a function of ΔR_{jj} approaches 100% faster in high multiplicity than in low multiplicity events. This is the result of the jet size, with jets being on average more energetic and therefore narrower in high multiplicity events.

4.4.3 Determining the $W + \mathrm{N_{jets}}$ subsamples

The selection cuts described in Section 4.4.2 are used to count the number of jets in the W inclusive sample. Table 4.3 shows the number of signal events per multiplicity bin before and after the electron-jet and jet-jet minimum separation cuts expected for 100 pb^{-1} of data at 10 TeV centre of mass energy. Studies of multiplicity up to $N_{jets} = 2$ are possible with a statistical

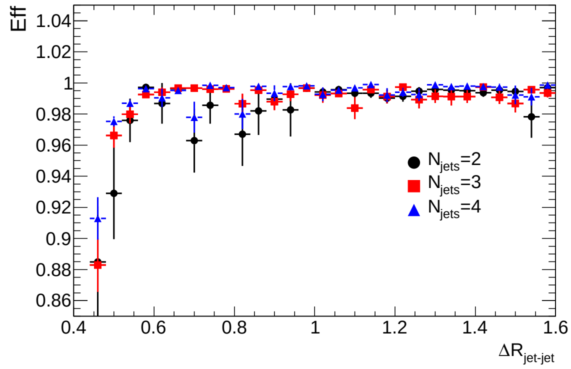

Figure 4.11: Jet reconstruction efficiency as a function of the distance ΔR to the closest jet, for truth jets with $p_T > 30$ GeV.

error of $\leq$ 1%. However, the statistics expected at the LHC in the first year of data-taking will be lower, given that the centre of mass energy reached will be 7 TeV rather than 10 TeV. The W inclusive cross-section at $\sqrt{s} = 7$ TeV is approximately 30% lower [8]. The production in association with jets is expected to be further suppressed at lower energies: the ratio of ALPGEN cross-sections at 7 TeV to 10 TeV is 0.56 for W + 2 parton and 0.49 for W + 3 partons. At this lower energy, the statistical error on the total W + 2 jets cross-section would be at the 1.4% level.

The minimum separation requirements on ΔR_{ej} and ΔR_{jj}, are effectively a reduction in the phase-space. The penalty in statistics due to these cuts increases with jet multiplicity, since the available phase-space for additional jets is gradually reduced. Whereas for $N_{jets} = 1$ this is a very small effect, it becomes more important in the $N_{jets} = 2$, where almost 5% of the events are rejected. The effect is significant for $N_{jets} \geq 3$. For this thesis, however, we will only consider jet multiplicity bins $N_{jets} = 1, 2$.

It is useful to look at the topology of W + jets events to better understand the performance studies discussed in this and following chapters. To study the kinematics of $W + 1/2$ jet events we choose the following variables:

- $W + 1$ jet: p_T of the leading jet

[8] The ratio $\frac{\sigma_W(NLO)(\sqrt{s}=7 \text{ TeV})}{\sigma_W(NLO)(\sqrt{s}=10 \text{ TeV})}$ is 0.68, calculated with the FEWZ program using MSTW08 PDF sets.

N_{jets}	Jet Count	ΔR_{ej}	$\epsilon_{Rej}(\%)$	ΔR_{jj}	$\epsilon_{Rjj}(\%)$
0	404025 ± 635	404025 ± 635	n/a	404025 ± 635	n/a
1	36802 ± 191	36592 ± 191	99.43 ± 0.04	36592 ± 191	n/a
2	8612 ± 92	8368 ± 91	97.17 ± 0.18	8213 ± 90	98.14 ± 0.15
≥ 3	2199 ± 47	2076 ± 45	94.41 ± 0.49	1836 ± 42	88.44 ± 0.70

Table 4.3: Expected number of $W \rightarrow e\nu$ + jets events for an integrated luminosity of 100 pb^{-1}. Shown separately are yields after basic jet counting (p_T and η cuts as described in the text) and after the additional cuts on electron-jet and jet-jet separation. Quoted are also efficiencies of the ΔR_{ej} and the ΔR_{jj} with respect to the previous cut.

- $W + 2\ jets$: HT, defined as the sum of the first and second leading jet p_T:

$$HT = p_T^{\text{jet1}} + p_T^{\text{jet2}}. \tag{4.5}$$

When a W boson is produced in association with hadronic jets, it is boosted to higher energies and, as a consequence, also the decay lepton will have higher p_T. As the number of jets in the event is increased, the boost is larger on average, resulting in higher energetic leptons, as shown in Fig. 4.12. The average energy per jet also increases with jet multiplicity. However, larger hadronic activity does not necessarily translate to a more energetic W boson and decay electron. The degree of correlation between the W and the jets p_T depends on the kinematic configuration of the event. We will take as an example the cases of $N_{jets} = 1, 2$ and consider the W boson p_T, p_T^W, calculated from the sum of missing transverse energy and electron p_T. For events with only one energetic jet, the correlation is clear: as the energy of the leading jet increases, so does the p_T^W (see Fig. 4.13). With $N_{jets} = 2$, the boson p_T will depend on whether jets are emitted at a small or large angle. If the two jets are emitted at a reasonably small angle, the W boson recoils against them, balancing their energy. If they are emitted back-to-to back in the transverse plane, the leading and second-leading jet balance each other, whereas the W remains rather soft. This is illustrated in Fig. 4.14, showing the p_T^W versus HT for the case where the angle in ϕ between the jets is $\Delta\phi_{jj} < \pi \cdot 3/4$ (left) and $\Delta\phi_{jj} > \pi \cdot 3/4$ (right). The configuration where the two leading jets are back-to-back dominates as the jet p_T cut is increased [66], as demonstrated in Fig. 4.15.

These features have some implications on the event reconstruction:

- If the kinematic configuration for $N_{jets} = 2$ is correlated with the jet energy, then the ΔR_{ej} and ΔR_{jj} will affect differently various parts of the phase space. In particular, the ΔR_{ej} will reject events in the jet back-to-back configuration and hence the high energy regime. The ΔR_{jj} cut, instead, will mainly affect the low energy regime. The efficiency of the two cuts was investigated as a function of HT for $N_{jets} = 2$ and found to be in agreement with the expectations, as illustrated in Fig. 4.15 (right): ΔR_{ej} efficiency drops with large HT, contrary to the ΔR_{jj} efficiency which rises slightly with HT. These effects should be taken into account when applying jet efficiency corrections, especially when measuring the cross-section as a function of jet kinematic variables.

- The boost of the W affects the kinematics of the decaying electron. It is therefore expected that electron acceptance and efficiencies exhibit some dependence on jet variables. Acceptance and efficiency corrections as a function of hadronic activity will be discussed in Chapter 7.

4.5 Conclusions

In this chapter, the criteria for selecting $W \rightarrow e\nu$ events and counting jets were described and motivated. It was explained how this analysis exploits the excellent electron identification capabilities of the ATLAS detector to trigger and identify offline leptonic decays of the W boson. Another key ingredient for this study is the measurement of $E_T^{\rm miss}$, provided by full-coverage calorimeters. It was shown that using the standard ATLAS $W \rightarrow e\nu$ selection criteria, the signal could be selected with very high efficiency.

A significant part of the chapter was dedicated to jet counting. The aim was to design a robust jet definition that would ease the comparison between data and theoretical predictions and limit the analysis to a well understood region of the phase-space. The definition of a jet involves a choice of jet algorithm, size and kinematics (such as the η range, p_T cut and minimum separation ΔR_{jj}, ΔR_{ej}). Furthermore, in a measurement involving electrons, a procedure for removing the overlap between electrons and jets has to be specified. All these aspects were treated in this chapter. Experimentally, it was shown how the jet performance is deteriorated in certain regions of the detector and is affected by the vicinity of another object (such as an electron or another jet). Based on such observations and other considerations (like the impact of non-perturbative effects and the statistics available), a baseline jet selection was proposed and this will be adopted throughout this thesis. Some jet kinematic variables were also introduced and it was briefly discussed how jet activity affects the kinematics of the W boson.

The $W \rightarrow e\nu$ + jets selection proposed here is not meant to be a finalised version, but represents a baseline which should be by all means revised with data. Comparison between data and MC, for example, will be crucial for the validation of variables used for electron identification and thus for the optimisation of the electron selection. In the context of jets, important factors for the choice of kinematic cuts will be the status of jet calibration and the availability of jet energy scale correction factors and their uncertainty. Measurements of the UE will also help identifying regions where non-perturbative effects are important, which could in turn affect the choice of the jet p_T cut.

Each cut applied on the event sample is source of some inefficiencies that have to be accounted for when measuring a cross-section and comparing data to theoretical predictions. Each selection is therefore also inevitably a source of some systematic uncertainty on the final measurement. These will be discussed in detail in Chapter 8 in the context of early data measurements of $W \rightarrow e\nu$ + jets.

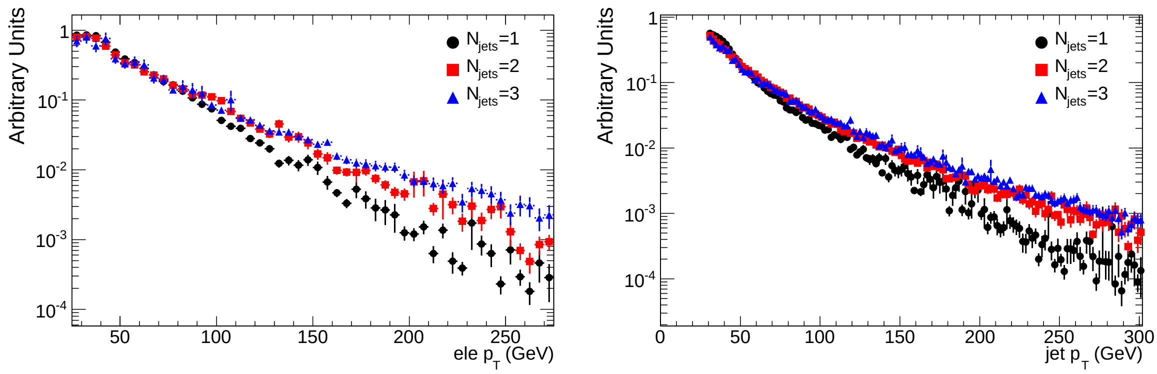

Figure 4.12: Electron (left) and jet (right) p_T distributions for different jet multiplicity bins.

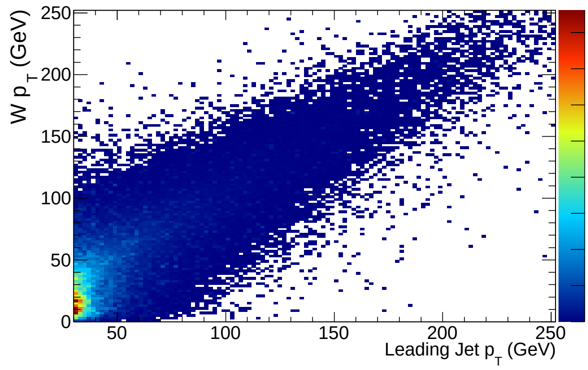

Figure 4.13: W boson p_T versus leading jet p_T for events with one jet.

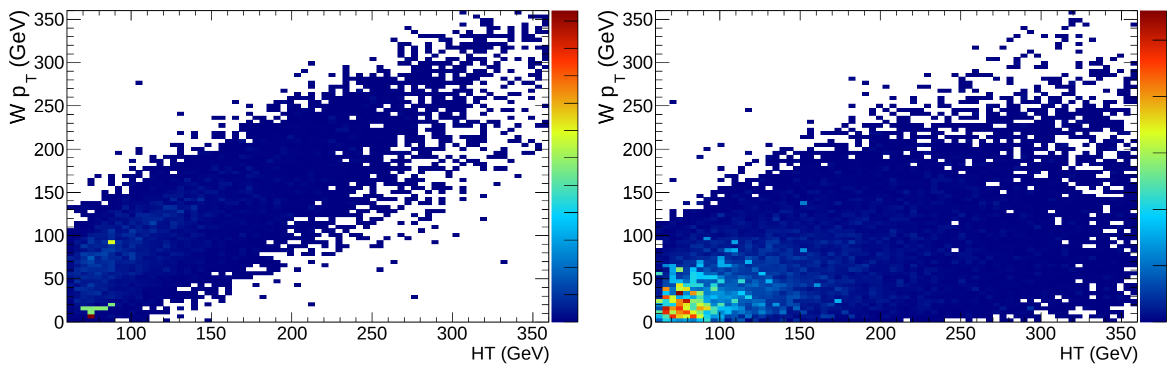

Figure 4.14: W boson p_T versus HT (defined as the sum of leading and second leading jet p_T) for events with two jets, in the case where the angle in the transverse plane between the two jets is $\Delta\phi_{jj} < \pi \cdot 3/4$ (left) and $\Delta\phi_{jj} > \pi \cdot 3/4$ (right).

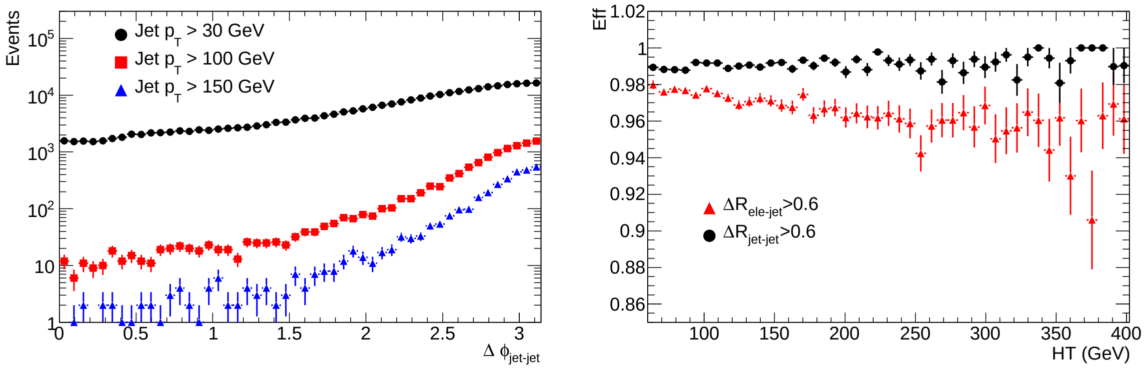

Figure 4.15: (Left) Angle $\Delta\phi_{jet-jet}$ between jets in the $N_{jets} = 2$ bin, for jets above different p_T thresholds. (Right) Efficiency of ΔR_{ej} and ΔR_{jj} relative to the baseline event selection (first column of Table 4.3) as a function of the HT variable for $N_{jets} = 2$.

Chapter 5

Backgrounds

This chapter is dedicated to the study of backgrounds to $W \rightarrow e\nu$ + jets events. Section 5.1 describes the main backgrounds and explains how they mimic signal events. Section 5.2 is dedicated to the investigation of lepton vetos for reducing leptonic backgrounds. Finally, methods for estimating backgrounds from MC and using data-driven techniques are described in Section 5.3. Particular emphasis is given to an in-situ technique for background estimation from top pair production, since it was fully developed by the author.

5.1 Backgrounds to $W \rightarrow e\nu$ + jets

Backgrounds to $W \rightarrow e\nu$ + jets can be divided into two categories: multi-jet and leptonic backgrounds. Multi-jets are produced in QCD interactions and are sources of fake electrons (ie. jets mis-reconstructed as electrons) and fake missing transverse energy (ie. due to detector inefficiencies and resolution). Leptonic backgrounds are produced in electroweak interactions and contain real leptons and, in some cases, neutrinos. One particular case of leptonic background is top quark production. Events with a top quark contain a genuine W boson and could arguably be considered either as part of the signal or background. In this analysis, they are considered as background since the aim is to measure direct single W production in association with jets.

The main backgrounds to $W \to e\nu$ + jets are listed in Table 5.1. Shown are the expected number of signal events for 100 pb^{-1} of integrated luminosity and the background fractions, expressed as background (B) to signal (S) ratio $\frac{B}{S}(\%)$ [1] in different jet multiplicity bins. The quoted errors are due to the MC statistics. The QCD-jet sample, in particular, suffers from severe lack of statistics. To overcome this problem, relaxed cuts have been applied to it for some of the plots shown in this chapter (as specified in the figure captions): the trigger requirement was removed and the E_T^{miss} threshold was lowered from 25 to 15 GeV [2]. The kinematic distributions were then normalised to the expected background fraction, as obtained applying the full selection.

The number of signal and background events as a function of jet multiplicity are plotted in Fig. 5.1. The leptonic backgrounds are combined together into backgrounds from W/Z processes and top-quark production. For the jet multiplicity bins N_{jets} = 1 and 2, the leading jet p_T and HT distributions respectively (see definition in Section 4.4.3) are also shown in Fig. 5.2. The overall background contamination is considerable, being 17% of the signal inclusively, and it increases significantly with jet multiplicity. The largest background source is QCD-jets, but top pair production becomes almost as important for three or more jets. We will concentrate however on the one and two-jet bins. The individual background sources are discussed in detail in the following:

QCD multi-jets:

This is the dominant background in all multiplicity bins, although it is difficult to draw firm conclusions for high number of jets, given the poor MC statistics available and the large uncertainties that affect theoretical predictions. QCD multi-jet events are a source of fake electrons in combination with large imbalance of transverse energy. After medium PID cuts, the largest

[1] For the cross-section measurement the quantity of interest is the overall background fraction $\frac{B}{S+B}$. However in this section it is preferred to use $\frac{B}{S}$ in oder to factorise out different background contributions and discuss each of them independently. This is desirable, in particular, given the large uncertainties on the QCD-jet background which translate in large uncertainties on the overall background fraction.

[2] Relaxed cuts are only applied to jet kinematic distributions within a given multiplicity bin, such that the correlation between E_T^{miss} and N_{jets} is taken into account. No significant bias is expected removing the trigger requirement, since the electron trigger selection is looser than the offline electron selection.

Selection : Trigger + Electron Selection + $E_T^{\text{miss}} > 25$				
Process	$N_{jets} \geq 0$	$N_{jets} = 1$	$N_{jets} = 2$	$N_{jets} \geq 3$
$W \rightarrow e\nu$	$(450.7 \pm 0.6) \cdot 10^3$	$(36.6 \pm 0.1) \cdot 10^3$	8210 ± 30	1840 ± 14
QCD-Jets(%)	14 ± 2	57 ± 13	115 ± 39	228 ± 116
$Z \rightarrow ee$ (%)	0.16 ± 0.01	0.72 ± 0.03	1.5 ± 0.1	3.4 ± 0.1
$W \rightarrow \tau\nu$ (%)	1.85 ± 0.02	4.50 ± 0.10	5.5 ± 0.1	7.3 ± 0.3
$Z \rightarrow \tau\tau$ (%)	0.19 ± 0.01	0.77 ± 0.02	1.5 ± 0.1	2.4 ± 0.1
$t\bar{t}$ (%)	0.83 ± 0.01	1.18 ± 0.01	13.4 ± 0.1	117.4 ± 1.2
single top (%)	0.20 ± 0.01	0.93 ± 0.02	4.1 ± 0.1	7.7 ± 0.2
di-bosons (%)	0.06 ± 0.001	0.32 ± 0.01	0.66 ± 0.02	0.72 ± 0.05
All (%)	17 ± 2	65 ± 13	141 ± 39	367 ± 116
All (no QCD) (%)	3.29 ± 0.03	8.4 ± 0.1	26.6 ± 0.2	138.9 ± 1.2

Table 5.1: Expected number of signal events and background fraction (eg. $N_{\text{QCD-jets}}/N_{W\rightarrow e\nu}$ in %) for the inclusive sample and different jet multiplicities. Events are normalised to 100 pb^{-1} of integrated luminosity. Errors are due to the MC statistics.

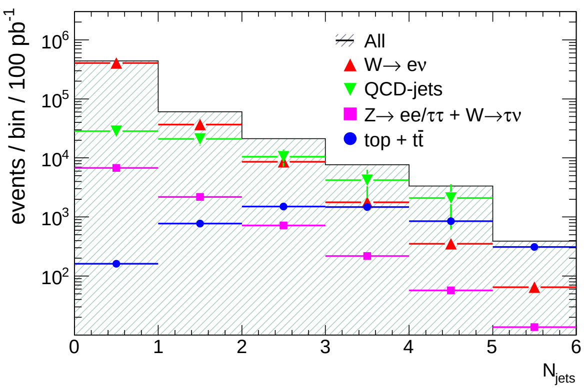

Figure 5.1: Signal and background distribution versus jet multiplicity, after the $W \rightarrow e\nu$ + jets selection cuts.

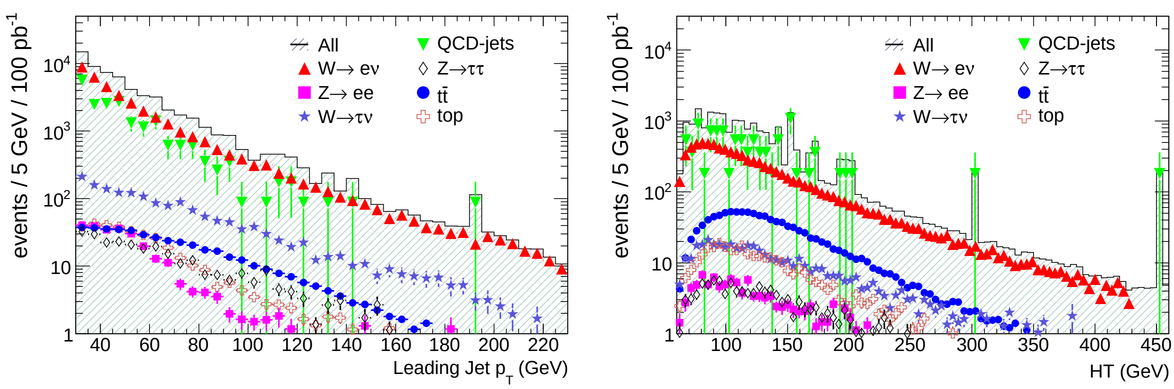

Figure 5.2: (Left) p_T of the leading jet for events with $N_{jets} = 1$. (Right) HT distribution for events with $N_{jets} = 2$. Because of poor MC statistics, relaxed cuts have been applied to the QCD-jets sample: the E_T^{miss} threshold was lowered to 15 GeV and the trigger requirement was removed. The total number of QCD-jets events was then scaled to the expected number of events with full selection (as shown in Table 5.1).

fraction of background electrons originates from photon conversions ($>$ 50 %), or are fake electrons from u/d/s hadrons (almost 20%). Another significant fraction (25%) are electrons from heavy flavour jets, these are real, but non-isolated electrons. Fake E_T^{miss} typically arises from jet mis-measurement, for example when a jet falls in the calorimeter gaps (transition regions) which have poorer resolution. Although the probability that a jet fakes an electron and that there is large fake E_T^{miss} in the same event is very small, QCD multi-jet events are still a very important source of background because of their large production rate compared to $W \to e\nu$. Figure 5.3 (left) shows the E_T^{miss} spectrum of signal and backgrounds for $N_{jets} \geq 0$, after full electron selection. The right plot shows the transverse mass distribution after requiring $E_T^{\text{miss}} > 25$ GeV. The low E_T^{miss} region is clearly dominated by QCD-jet events. A E_T^{miss} cut removes most of this background, but the residual number of QCD-jet events is non-negligible. An additional lepton isolation or transverse mass cut can help keep this background under control (a transverse mass cut of 40 GeV, for example, should reduce it by $\sim$ 50% and isolation to even lower levels). However it will be necessary to wait for data to gain accurate knowledge of the background rejection factors and optimise the selection cuts.

Even if sufficiently high statistics in the QCD multi-jet sample was available, MC simulations could not be trusted for estimating this background and data-driven estimates must be used. In fact, the jet production cross-section suffers from large theoretical uncertainties and contributions from underlying event and multiple interactions are largely unknown at the LHC. Beside an excellent understanding of the detector, an accurate prediction of electron fake rates would also require good knowledge of the event topology and the internal structure of jets (ie. how they fragment and hadronize), which are very sensitive to non-perturbative contributions. Furthermore, fake E_T^{miss} is expected to be poorly modelled in GEANT4 simulations. Several data-driven methods for calculating QCD background are under investigation at ATLAS and will be briefly described in Section 5.3.2.

$\mathbf{Z \to ee}$:

Electrons from Z boson decays are kinematically very similar to those from W bosons and

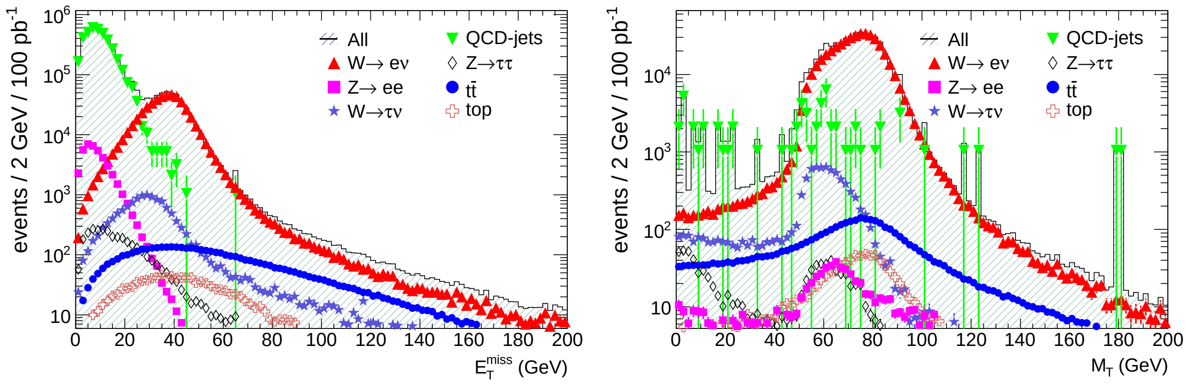

Figure 5.3: E_T^{miss} distribution in signal and background events (for $N_{jets} \geq 0$), after full electron selection (left) and transverse mass distribution after a cut on $E_T^{\text{miss}} > 25$ GeV (right).

therefore satisfy the same selection criteria. For a $Z \to ee$ event to mimic a $W \to e\nu$ decay, two conditions have to be satisfied: one of the two electrons fails the electron selection criteria (because of kinematic acceptance or reconstruction/identification inefficiencies) and there is a large amount of fake missing energy. Fake E_T^{miss} can arise either from the second electron escaping the detector through a poorly instrumented region, or from jet energy mis-measurement, as for the case of QCD-jets. These two scenarios are illustrated in Fig. 5.4. The left plot shows η of the missed electron in the $Z \to ee$ sample, after the W selection with/without a E_T^{miss} cut of 25 GeV. The jump around $|\eta| = 2.5$ corresponds to the kinematic acceptance of the detector for electron reconstruction: an electron that falls outside this region is missed and always reconstructed as a jet. Clearly, the E_T^{miss} cut greatly reduces the background contamination. Most of the residual background is in correspondence with the poorly instrumented regions of the detector (the cracks in the calorimeter around $|\eta| = 1.5$ and 3.1).

In Fig. 5.4 (right), the $Z \to ee$ background contamination is plotted as a function of jet multiplicity. Although the $Z \to ee$ production rate is about 10 times smaller than the $W \to e\nu$ one, this is still a non-negligible background because of bin migrations. In fact, if the missed electron fails the electron selection criteria but is still within the jet acceptance, a $Z \to ee + N_{jets}$ events will be reconstructed as a $W \to e\nu + N_{jets} + 1$. Given that the cross-section for each additional jet drops by a factor of ∼10, migrations of $Z \to ee$ events to upper multiplicity bins can result in a non-negligible background contamination. This explains the sudden raise by a

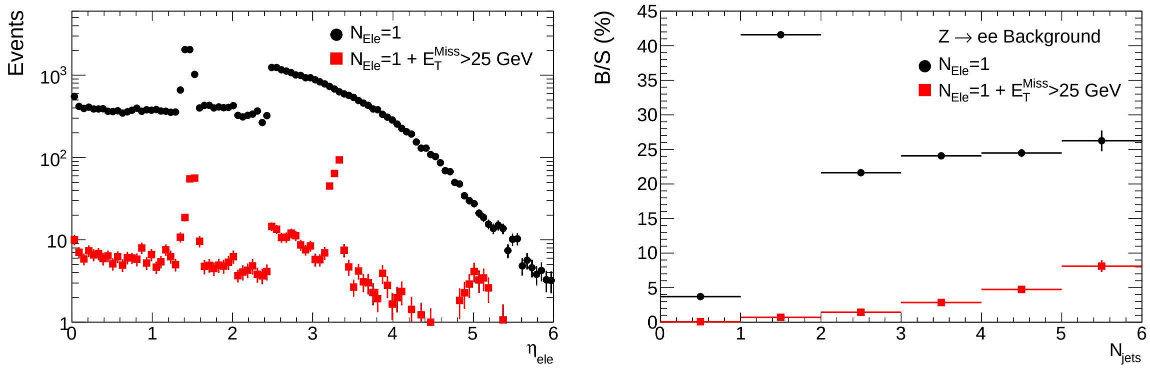

Figure 5.4: (Left) Pseudo-rapidity distribution of the missed electron in $Z \rightarrow ee$ events, after applying the $W \rightarrow e\nu$ electron selection and with/without a cut on E_T^{miss}. (Right) $Z \rightarrow ee$ background contamination expressed as $\frac{N_{Zee}}{N_{Sig}}(\%)$ as a function of jet multiplicity.

factor of 10 in $\frac{B}{S}$ between the zero and one-jet bins in Fig. 5.4 (right) before a E_T^{miss} cut: the single jet is, in most cases, an electron from a $Z \rightarrow ee + 0$ jet event. Since the electron energy is not lost, there is no fake E_T^{miss} in the event and the E_T^{miss} cut brings down the background to the percent level. As the number of jets increases, the probability of large imbalance of transverse energy due to jet mis-measurement increases, explaining the rise in the background level with jet multiplicity even after the E_T^{miss} cut. $Z \rightarrow ee$ events can be rejected by a lepton veto cut, as will be described in Section 5.2.1. The residual background can be estimated from MC.

W $\rightarrow \tau\nu$ and Z $\rightarrow \tau\tau$:

The production rate for $W \rightarrow \tau\nu$ is the same as for $W \rightarrow e\nu$ and in 18% of the cases the τ will decay into a final state electron: $\tau \rightarrow e\,\nu_e\,\nu_\tau$. These events contain therefore a real electron and real E_T^{miss}, but this electron has on average a smaller p_T than the electron from $W \rightarrow e\nu$, since the τ energy is shared among three final state particles. This is illustrated in Fig. 5.5 (left). The p_T threshold on the electron rejects most of the $W \rightarrow \tau\nu$ background, but its contamination is still at the level of $4 - 6\%$ for events with one or two jets.

Similarly, $Z \rightarrow \tau\tau$ can mimic a $W \rightarrow e\nu$ when one τ decays leptonically into an electron. The $Z \rightarrow \tau\tau$ background is at a level of a few percent, comparable with $Z \rightarrow ee$. In fact, the production rate of $Z \rightarrow \tau\tau$ events with one final state electron is smaller than that of $Z \rightarrow ee$. However, the amount of true E_T^{miss} in a $Z \rightarrow \tau\tau$ event is larger, due to the presence of neutrinos;

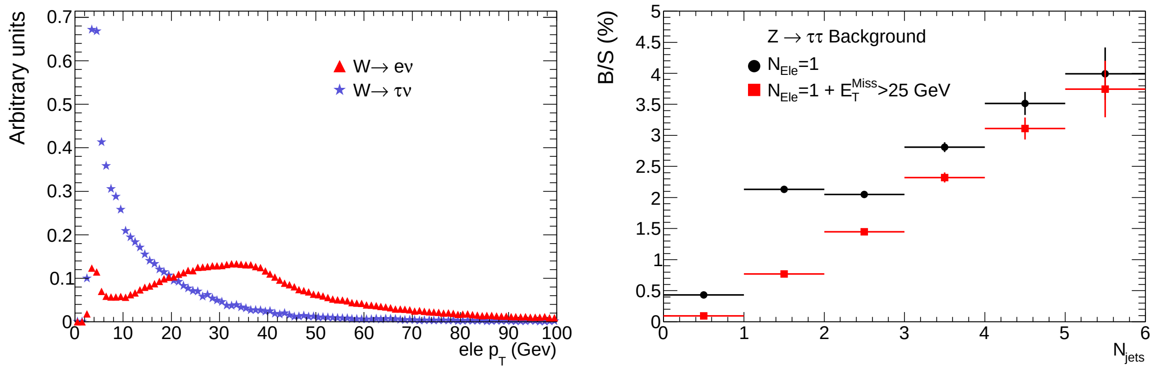

Figure 5.5: (Left) p_T distribution of the candidate electron in $W \rightarrow e\nu$ +jets and $W \rightarrow \tau\nu$ +jets events. (Right) Ratio of $Z \rightarrow \tau\tau$ +jets to $W \rightarrow e\nu$ +jets as a function of jet multiplicity before and after a cut on E_T^{miss}.

hence the E_T^{miss} cut does not have a strong rejection power against this background (see Fig. 5.5 right). Distinguishing between primary electrons and electrons from leptonic τ decays is very difficult. A generic lepton veto is not helpful against $W \rightarrow \tau\nu$ but can partly help reducing $Z \rightarrow \tau\tau$ background when also the second τ decays leptonically. Both $W \rightarrow \tau\nu$ and $Z \rightarrow \tau\tau$ background are estimated with MC, as described in Section 5.3.1.

Single Top and Top pair production:

The top quark decays almost exclusively into the $t \rightarrow Wb$ channel. Events with a top quark in the final state therefore contain a genuine W boson which can decay leptonically or hadronically and is accompanied by jets. The leading-order single top production diagrams are shown in Fig. 5.6. The dominant contribution at the LHC is the t-channel, followed by associated production of top and W boson and finally by the s-channel. An event with a single top quark production that is background to $W \rightarrow e\nu$ + jets will contain an electron and a b-quark jet from the top decay ($t \rightarrow W(e\nu)b$) plus typically an additional jet. When top quarks are produced in pairs (will refer to this process as $t\bar{t}$), there are at least two b-quark jets in the final state and a number of leptons or additional jets which depend on the decay mode of the W boson. There are three different signatures for $t\bar{t}$ production:

- fully hadronic (about 4/9 of the times), where both W bosons from $t \rightarrow Wb$ decay hadronically, resulting in six jets, of which two are b-quark jets;

- fully leptonic (about 1/9 of the times), where both W boson decays leptonically, resulting in two leptons, two neutrinos and two b-quark jets;

- semi-leptonic (about 4/9), where one W boson decays leptonically and one hadronically, resulting in one lepton, one neutrino and four jets. Fig. 5.7 is an example of a semi-leptonic $t\bar{t}$ decay, with one final state electron.

Both single top and $t\bar{t}$ production are kinematically very similar to $W \to e\nu$ +jets and, due to the presence of jets, contaminate the high jet multiplicity bins. $t\bar{t}$, in particular, competes with QCD-jets as the dominant background at high jet multiplicity and, for $N_{jets} = 2$, its contamination is already 13% (in terms of $N_{t\bar{t}}/N_{W\to e\nu}$). Table 5.2 shows the composition of the $t\bar{t}$ background to W +jets for the one and two-jet bins. For these low multiplicities the dominant component is the fully leptonic signature, since it counts the least number of jets. At least one of the top quarks has to decay into the electron channel to satisfy the lepton event selection. In fact, the $e\mu$ and $e\tau$ decay modes are favoured over the ee, since the $W \to e\nu$ selection (with $N_{ele} = 1$) already excludes events with more than one well reconstructed electron, but has no similar veto on muons or taus. Furthermore, a second electron that falls outside the acceptance will be reconstructed as a jet, pushing the event to higher jet multiplicities, whereas muons and taus decaying leptonically [3] do not contribute to jet counting. The semi-leptonic $t\bar{t}$ decay into the electron channel becomes more important for higher multiplicities, as expected since it contains four-jets. Given that more than 50% of the $t\bar{t}$ background to $W \to e\nu$ +jet events in the multiplicity bins of interest are fully-leptonic decays, lepton vetos can be a powerful tool to control this background. In addition, a new methodology to estimate this background from data is investigated using b-quark jets. This will be described in Section 5.3.2.

Di-bosons:

These are processes with WW, WZ or ZZ in the final state. Due to the small cross-section, they are expected to be a very small background (<1%), as shown in Table 5.1, and therefore can be

[3] Taus decaying into a final state electron will be reconstructed as a jet, but this will typically have a rather small p_T (cf. Fig. 5.5 left) and therefore will rarely contribute to jet counting.

Composition of $t\bar{t}$ background (%)			
Decay Mode	Channel	$N_{jets} = 1$	$N_{jets} = 2$
Hadronic	-	< 0.01	0.08 ± 0.03
Semi-leptonic	e	15.84 ± 0.77	33.09 ± 0.42
	μ	0.61 ± 0.14	0.80 ± 0.05
	τ	1.16 ± 0.18	2.58 ± 0.10
Fully-leptonic	ee	10.55 ± 0.59	8.86 ± 0.19
	$e\mu$	40.85 ± 1.30	29.08 ± 0.38
	$e\tau$	25.97 ± 0.99	22.04 ± 0.33
	$\mu\tau$	3.02 ± 0.30	2.06 ± 0.09
	$\tau\tau$	1.75 ± 0.23	1.30 ± 0.07
	$\mu\mu$	0.24 ± 0.11	0.12 ± 0.02

Table 5.2: Composition of $t\bar{t}$ background to $W \to e\nu$ + jets for the one and two jet bins. Each column sums up to 100%.The overall background fraction is $\frac{N_{t\bar{t}}}{N_{Sig}} = 1.2\%$ (1 jet), 13.4% (2 jets), as shown in Table 5.1.

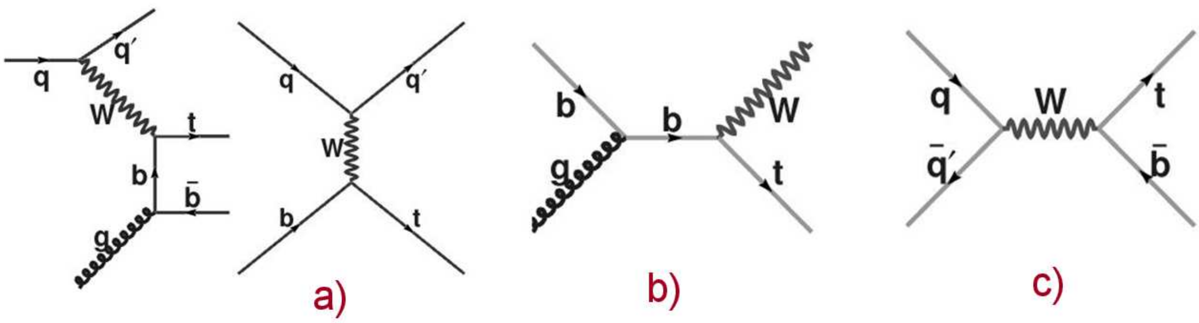

Figure 5.6: Single top quark production diagrams: (a) t-channel, (b) Wt associated production, (c) s-channel. The t-channel is the dominant mode at the LHC with a cross-section of 124.5 pb, followed by the Wt channel (32.7 pb) and the s-channel (6.6 pb) at 10 TeV centre of mass energy. [67]

safely calculated using MC.

5.2 Vetos for Rejecting Leptonic Backgrounds

In Section 5.1 it was shown that the background contamination in the signal sample is significant. Even excluding QCD-jets, leptonic backgrounds sum up to $\frac{B}{S} = 8.4\%$ for one jet and $\frac{B}{S} = 26.6\%$ for two jets (see Table 5.1). Assuming, as an example, a conservative uncertainty

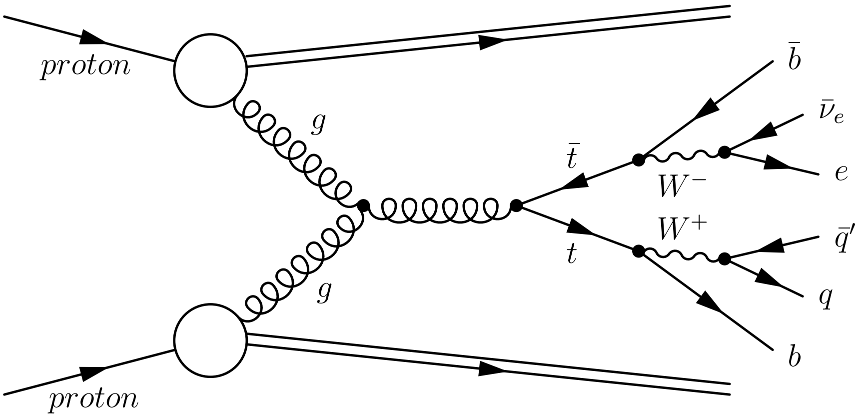

Figure 5.7: Semi-leptonic decay of a top pair.

on the background estimation of 20% [4], this would result in an uncertainty on the cross-section of 1.7% (1 jet) and 5.2% (2 jets), which is much larger than the expected statistical uncertainty of 0.5% (1 jets) and 1.1% (2 jets) for 100 pb^{-1} of ATLAS data. To decrease the systematic uncertainty, it is therefore worth to attempt to reduce these backgrounds to a lower level. This can be achieved through lepton vetos.

5.2.1 Lepton Veto Cuts

The W selection criteria described in section 4.3 requires exactly one good quality electron, but poses no limitations to the presence of a second poorer quality electron or of a muon or a tau lepton. Electroweak backgrounds that contain more than one lepton can be reduced by lepton veto cuts. Such backgrounds are: $Z \rightarrow ee$, $Z \rightarrow \tau\tau$, $t\bar{t}$ and di-bosons (WW, WZ, ZZ). Two vetos are investigated: one on isolated tracks and one on jets with high electromagnetic energy fraction.

- **Veto on an isolated track:**

  a second loose electron or muon within the inner detector acceptance ($|\eta| < 2.5$) will leave hits in the silicon detectors. Tracks from leptonic decays of Ws and Zs are mostly

[4] This value is only intended as a reference baseline. A discussion of the factors that contribute to systematic uncertainties on the background estimation is given in Section 5.3.1.

isolated, contrary to tracks left by other objects such as leptons from heavy quark decays. Figure 5.8 shows the isolation fraction of tracks in $W \to e\nu$, $W \to \mu\nu$ and $W \to \tau\nu$ events, where isolation fraction is defined as the sum in p_T of all tracks within a cone of radius $\Delta R = 0.4$ around the candidate lepton track (p_T^l), normalised by the lepton track:

$$\text{IsoFraction} = \frac{\sum_i p_T^i - p_T^l}{p_T^l}, \text{ for i} = \text{tracks} \in \ \Delta\text{R} < 0.4. \tag{5.1}$$

The peak around IsoFraction = 0 originates from poor quality tracks with very low p_T and no silicon hits (only hits in the TRT). It is in fact mostly due to fake tracks. In order to remove this contribution, only tracks satisfying all of the following quality requirements are considered:

– At least one hit in the Pixel detector

– At least one hit in the SCT detector

– Track $p_T > 5$ GeV.

A good quality track is then flagged as isolated and therefore as coming from a lepton if:

– $\text{IsoFraction} < 0.15.$

Any event with one or more tracks satisfying the above criteria, after having excluded the track from the candidate electron from the W boson decay, is rejected.

- **Veto on jets with high electromagnetic energy fraction**:

  The track veto cut can only be used to reject events with a second central lepton, ie. with $|\eta| < 2.5$, according to the acceptance of the Inner Detector. To extend the veto to the forward region, calorimeter information can be used. A good discriminator between hadronic jets and jets originating from electrons [5] is the electromagnetic energy fraction

[5] As it was explained in section 4.4.2, electrons are also reconstructed as jets by the ATLAS jet clustering algorithm.

in a jet, defined as:

$$\text{EMFraction} = \frac{E_{EM}^{jet}}{E_{EM}^{jet} + E_{HAD}^{jet}} \tag{5.2}$$

where E_{EM}^{jet} is the energy of a jet measured in the EM component of the calorimeter and E_{HAD}^{jet} is the energy measured in the hadronic part of the calorimeter, before hadronic calibration. Since electron showers are generally contained in the EM calorimeter, this variable can be used to identify extra electrons in the full range covered by calorimeter (up to $|\eta| < 4.9$), but is not effective on muons which typically do not deposit energy in the calorimeter. A veto on jets with high electromagnetic energy fraction specifically targets $Z \to ee$ events and fully-leptonic $t\bar{t}$ decays with two final state electrons. Figure 5.9 (left) shows the EM fraction of all jets in $W \to e\nu$ and $Z \to ee$ events, after the $W \to e\nu$ selection cuts and the electron overlap removal. The peak around $\text{EMFraction} = 1$ shows the presence of electron-jets in the $Z \to ee$ sample. From this plot, it was chosen to reject events for which at least one of the jets satisfies:

– $\text{EMFraction} > 0.96$

where all jets with $p_T > 30$ GeV and in the full pseudo-rapidity range are considered. In Fig. 5.9 (right), the invariant mass of the candidate electron and the jet with the highest EM fraction is plotted, for $W \to e\nu$ and $Z \to ee$ vetoed events only. The resonance around the Z mass value shows that the veto is effective in identifying $Z \to ee$ events.

5.2.2 Results and Discussion

The effect of the two veto cuts on the signal efficiency and background rejection was investigated in order to evaluate the gain of applying such cuts in the final cross-section measurement. The signal efficiency is shown in Fig. 5.10, for one and two jets. The EM fraction has a higher efficiency than the track veto, but is also dependent on the jet kinematics, dropping at low leading jet p_T and HT. The track veto, instead, is uncorrelated with jet kinematics and the efficiency

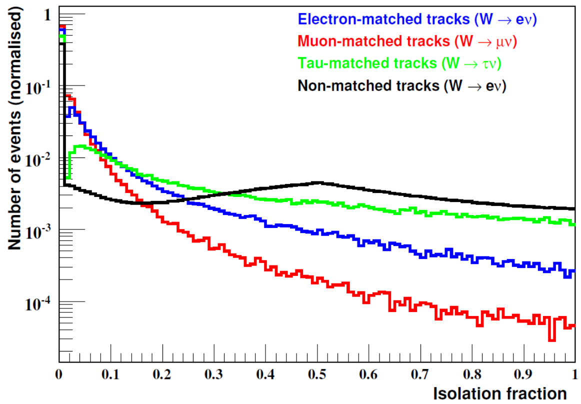

Figure 5.8: Isolation fraction (defined as in eqn. (5.1)) of tracks in $W \rightarrow l\nu$ events. Tracks matched to leptons from the W boson decay have small isolation fraction. The peak at Isolation Fraction = 0 is due to fake tracks and can be removed by track-quality cuts.

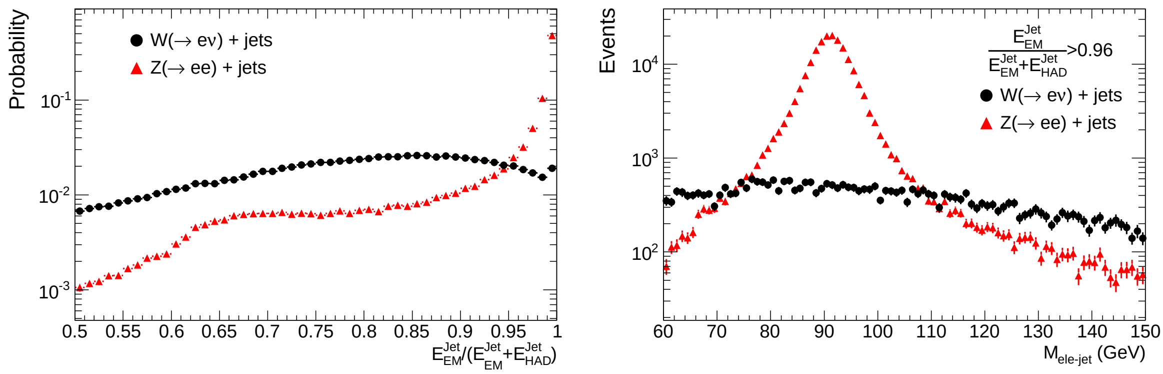

Figure 5.9: (Left) EM energy fraction of jet in $W \rightarrow e\nu$ and $Z \rightarrow ee$ events, after the W +jets selection cuts. (Right) Invariant mass of the candidate electron and the jet with EM fraction above 0.96.

is constant at 85%.

The lower efficiency of the track veto cut is compensated by the significant larger rejection power against leptonic backgrounds, as shown in Fig. 5.11-5.13. Whereas the EM fraction cut is only effective on $Z \rightarrow ee$ events, the track veto also reduces $Z \rightarrow \tau\tau$ and $t\bar{t}$ backgrounds. Even considering $Z \rightarrow ee$ alone, the track veto rejects more events, keeping the background level below 0.5 (1.0) % for 1 (2) jets. The EM fraction veto is less powerful over the whole kinematic region and ineffective in the high energy range, where the signal efficiency is also higher. This is particularly evident in the one-jet bin: going to higher jet p_T means selecting a highly energetic electron, which is more likely to leak into the hadronic calorimeter and therefore have a low EMFraction. The fact that $t\bar{t}$ background is not reduced by this veto can also be explained

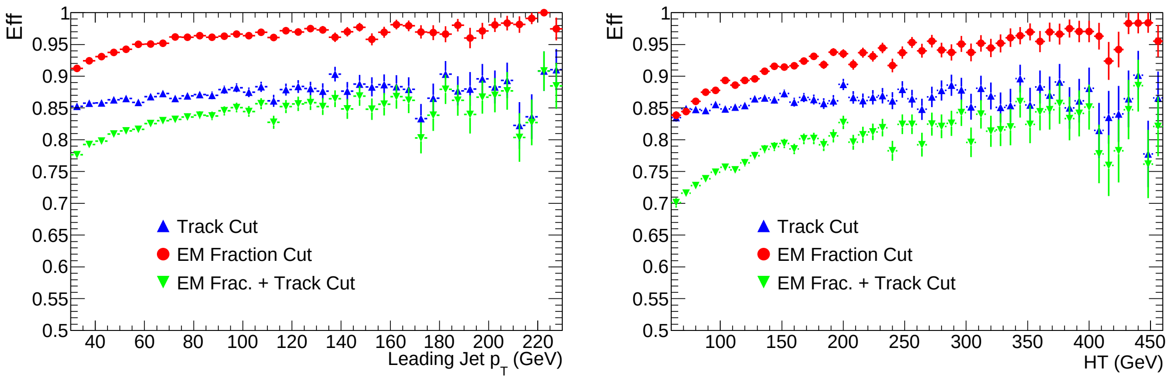

Figure 5.10: Signal efficiency of veto on isolated tracks and on jets with high EM fraction as a function of leading jet p_T for $N_{jets} = 1$ and of HT for $N_{jets} = 2$.

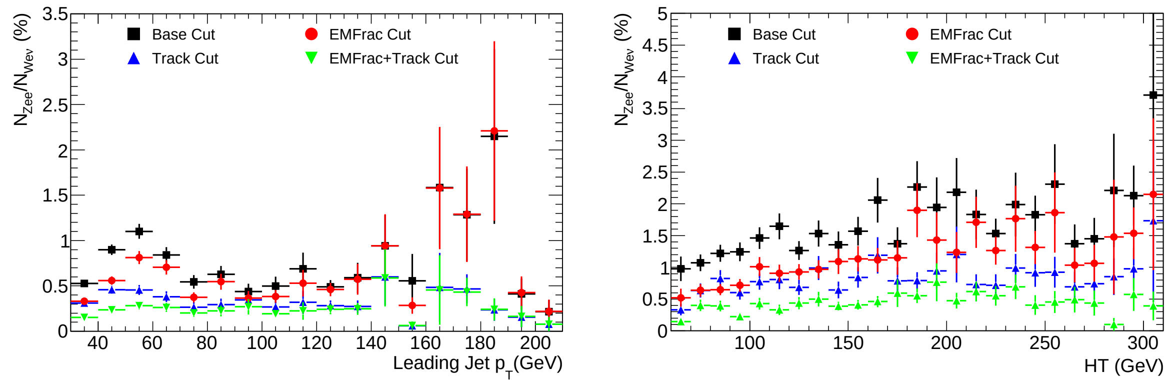

Figure 5.11: $Z \rightarrow ee$ background rejection by veto on isolated tracks and on jets with high EM fraction as a function of leading jet p_T for $N_{jets} = 1$ and of HT for $N_{jets} = 2$.

by the low rejection power against boosted electrons combined with the low fraction of fully-leptonic $t\bar{t}$ decays with two electrons in the signal sample (see Table 5.2). The track veto proves considerably more powerful, in particular since it reduces the important $t\bar{t}$ background by about $1/3$. Di-boson backgrounds are not shown, since they are a very small percentage, but they are also reduced by 25-50 % by both vetoes, similarly for $Z \rightarrow ee$. $W \rightarrow \tau\nu$ and single top are unaffected, as expected since they contain no additional lepton to the selected electron.

The overall effect of the lepton vetos on signal and background is summarised in Tables 5.3 and 5.4, where the efficiency for the selection of signal events, the increase in the statistical uncertainty due to signal rejection and the decrease in background contamination are shown.

The decision on whether or not to apply the veto cuts should be based upon the impact on cross-section measurement systematic uncertainty. An additional cut will have an effect on the

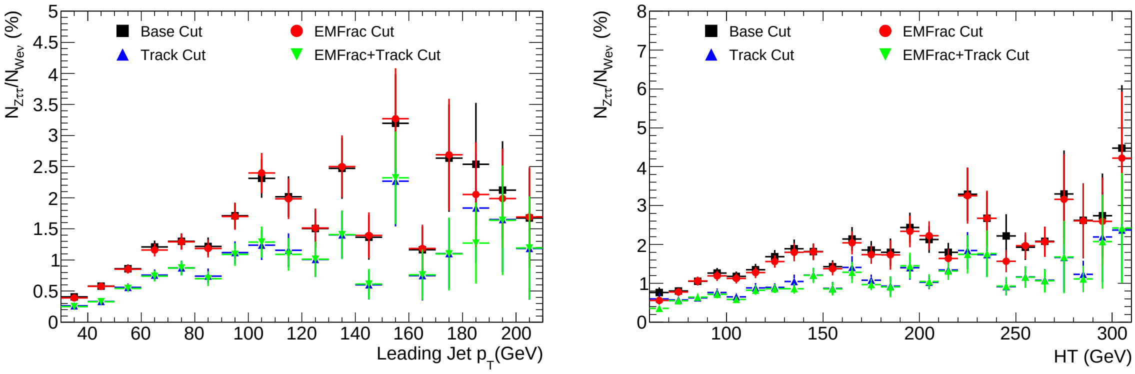

Figure 5.12: $Z \rightarrow \tau\tau$ background rejection by veto on isolated tracks and on jets with high EM fraction as a function of leading jet p_T for $N_{jets} = 1$ and of HT for $N_{jets} = 2$.

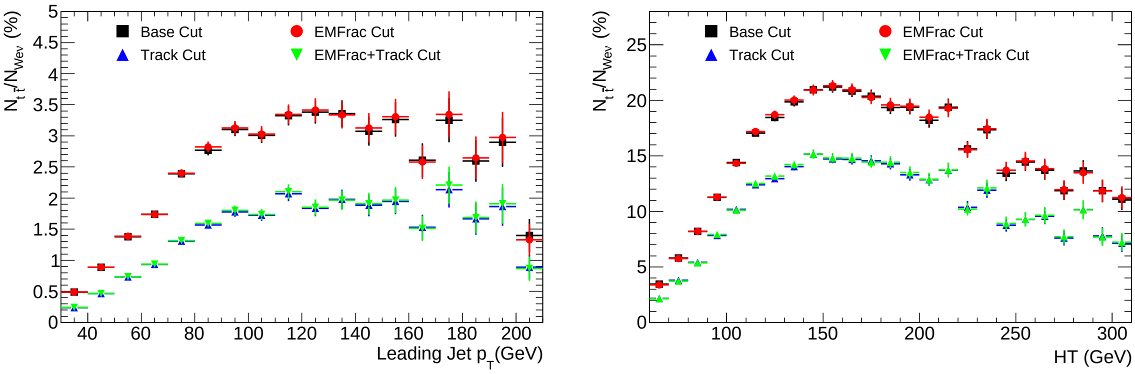

Figure 5.13: $t\bar{t}$ background rejection by veto on isolated tracks and on jets with high EM fraction as a function of leading jet p_T for $N_{jets} = 1$ and of HT for $N_{jets} = 2$.

following terms of eqns. (3.3) and (3.4) in Chapter 3:

- $\frac{\Delta(N-B)}{N-B}$: it would increase the relative statistical error on the number of estimated signal events. This term, which will be referred to as Δ_{stat}, can be rewritten as:

$$\Delta_{stat} = \frac{\Delta(N-B)}{N-B} = \frac{\sqrt{\Delta N^2 + \Delta B^2}}{N-B} \tag{5.3}$$

For simplicity we will take $\Delta N \gg \Delta B$ [6] and $N - B$ equal to the true number of signal events S, such that $\Delta_{stat} \approx \frac{\Delta N}{S}$;

- $\frac{\Delta\epsilon}{\epsilon}$: an additional cut means an additional loss of signal events. This loss has to be corrected for in the final cross-section measurement with an efficiency factor (ϵ), which

[6] This is a valid approximation, since the statistical error on the estimated background of events is expected to be negligible: leptonic backgrounds will be estimated from MC and the full statistics available can be used.

will have an associated uncertainty $\Delta\epsilon$.

- $\dfrac{\Delta B}{N-B}$: it would decrease the systematic error due to the background. This can be rewritten in terms of signal to background ratio and error on the background estimation:

$$\frac{\Delta B}{N-B} = \frac{\Delta B}{B} \cdot \frac{B}{S} \tag{5.4}$$

  where S and B are the number of signal and background events and $\frac{\Delta B}{B}$ is the relative uncertainty on the background. In the following discussion we will assume an error on the background of $\frac{\Delta B}{B} = 20\%$ as a baseline.

For the veto cuts to be advantageous, the decrease in the overall uncertainty thanks to the background reduction must be greater than the increase in the uncertainty due to the efficiency and statistical terms [7]:

$$\Delta\left(\frac{\Delta B}{B} \cdot \frac{B}{S}\right) > \sqrt{\Delta\Big(\Delta_{stat}\Big)^2 + \Delta\left(\frac{\Delta\epsilon}{\epsilon}\right)^2} \tag{5.5}$$

The expected decrease in the background contamination after lepton vetos is given in Tables 5.3-5.4 and the corresponding relative decrease in the systematic uncertainty is shown in Table 5.5. In comparison, the increase in statistical error due to veto cuts is very small (see Tables 5.3-5.4), such that the term $\Delta(\Delta_{stat})$ in eqn. (5.5) can be neglected. The efficiency term remains to be evaluated. For the absolute cross-section measurement it will have to be estimated from MC. For the measurement of $R_{jets} = \dfrac{W + N_{jets}}{Z + N_{jets}}$, however, the efficiency cancels at first order in the ratio, such that there is no need for evaluating the efficiency and a systematic uncertainty can be assigned, as will be discussed in Chapter 8.

For the EM fraction veto, the gain is very small, with $\Delta\left(\frac{\Delta B}{B} \cdot \frac{B}{S}\right)$ at the level of $0.1 - 0.2\%$.

[7] In eqn. 5.5 the Δ outside the brackets represents the change (increase/decrease) in the uncertainty. The term $\Delta\left(\frac{\Delta B}{B} \cdot \frac{B}{S}\right)$ is the change in the systematic uncertainty due to background. Similarly, $\Delta\Big(\Delta_{stat}\Big)$ is the increase in the statistical uncertainty and $\Delta\left(\frac{\Delta\epsilon}{\epsilon}\right)$ is the increase in the uncertainty due to the efficiency term.

Cut	Jet $p_T > 30$ GeV			Jet $p_T > 100$ GeV		
		$\Delta_{stat}(\%)$	$\frac{B}{S}(\%)$		$\Delta_{stat}(\%)$	$\frac{B}{S}(\%)$
Ref.		0.52	8.4 ± 0.1		2.03	19.2 ± 0.6
	Eff(%)	$\Delta_{stat}(\%)$	$\frac{B}{S}(\%)$	Eff(%)	$\Delta_{stat}(\%)$	$\frac{B}{S}(\%)$
Track	88.0 ± 0.2	0.56	6.9	89.4 ± 0.6	2.15	16.3
EMFrac	93.6 ± 0.1	0.54	7.9	97.2 ± 0.3	2.06	18.9
EM+Track	82.3 ± 0.2	0.58	6.8	86.8 ± 0.7	2.18	16.3

Table 5.3: Effect of the lepton vetos on signal and background for $N_{jets} = 1$. The first row shows the relative statistical error $\Delta_{stat}(\%)$ and the background fraction $\frac{B}{S}(\%)$ for the standard $W \to e\nu$ selection cuts. The last three rows show the efficiency of the veto cuts with respect to the standard selection and the statistical error and background contamination after lepton vetos. Only leptonic backgrounds are considered (QCD-jets are excluded). Results are shown for two different kinematic regions: leading jet $p_T > 30$ and 100 GeV.

Cut	HT > 60 GeV			HT > 200 GeV		
		$\Delta_{stat}(\%)$	$\frac{B}{S}(\%)$		$\Delta_{stat}(\%)$	$\frac{B}{S}(\%)$
Ref.		1.10	26.6 ± 0.2		3.13	32.0 ± 0.5
	Eff(%)	$\Delta_{stat}(\%)$	$\frac{B}{S}(\%)$	Eff(%)	$\Delta_{stat}(\%)$	$\frac{B}{S}(\%)$
Track	85.4 ± 0.2	1.19	20.5	86.9 ± 0.6	3.36	24.9
EMFrac	89.3 ± 0.3	1.17	25.6	94.4 ± 0.7	3.22	30.8
EM+Track	76.2 ± 0.5	1.26	20.4	81.9 ± 1.2	3.46	24.4

Table 5.4: Effect of the lepton vetos on signal and background for bin $N_{jets} = 2$. Results are shown for two different kinematic regions: HT > 60 and 200 GeV. For more details see Table 5.3.

Both in the absolute cross-section or in the ratio measurement it would be difficult to have a systematic uncertainty from the efficiency below this value, defeating the purpose of the veto. Similar conclusions can be drawn for the track veto in the one-jet case. For two jets, however, the background reduction is significant and the gain on the systematic uncertainty is $\Delta\left(\frac{\Delta B}{B} \cdot \frac{B}{S}\right) = 1.2 - 1.4\%$. If the systematic uncertainty from the efficiency term is about 0.2%, as expected for the R_{jets} ratio measurement (see Chapter 8), it is worth to apply this veto. Based on the above considerations, it is therefore suggested to apply the track veto cut for $N_{jets} = 2$.

Cut	$N_{jets} = 1$		$N_{jets} = 2$	
	Jet $p_T > 30$ GeV	Jet $p_T > 100$ GeV	HT > 60 GeV	HT > 200 GeV
Track	0.3%	0.6%	1.2%	1.4%
EMFrac	0.1%	$< 0.1\%$	0.2%	0.2%
EM+Track	0.3%	0.6%	1.2%	1.5%

Table 5.5: Relative decrease in the systematic uncertainty from the background term $\Delta\left(\frac{\Delta B}{B} \cdot \frac{B}{S}\right)$ by applying the track and EM fraction veto cuts and assuming an uncertainty on the leptonic background estimation $\frac{\Delta B}{B} = 20\%$.

5.3 Background Estimations

5.3.1 Estimation of leptonic backgrounds from MC

In this section the estimation of leptonic backgrounds from MC is discussed. Small and/or well understood backgrounds, such as $Z \to ll$, $W \to \tau\nu$ and di-bosons can be safely estimated from MC, since they will have a small impact on the total cross-section uncertainty. Theoretical predictions for single top and top pair production are also available at NLO or higher orders [68] and can therefore be estimated from MC. Nevertheless, since they represent a significant fraction of the background, especially for three or more jets, an alternative data-driven technique (described in Section 5.3.2), is also investigated. A cross-check between MC and data-driven estimates will then give a good handle on systematic uncertainties.

Di-bosons, Single Top and Top Pair Production

For di-bosons and top backgrounds, fully simulated MC samples are used and passed through the analysis selection cuts. The remaining number of events is then scaled to the required luminosity:

$$\begin{aligned} N &= N_{Gen}^{Cuts} \cdot \frac{L}{L_{Gen}} \\ &= N_{Gen}^{Cuts} \cdot L \cdot \frac{\sigma}{N_{Gen}} \end{aligned} \tag{5.6}$$

where N is the predicted number of background events in the signal sample, N_{Gen} and

N_{Gen}^{Cuts} are the number of generated events before and after the selection cuts respectively, σ is the production cross-section, L is the luminosity of the data sample (to be measured) and L_{Gen} is the luminosity equivalent to the MC dataset. The statistical error will be small, since the full available MC dataset is used. The systematic error on the estimated number of background events depends on the uncertainty on the measured luminosity ($\Delta L/L$) and on the background process cross-section ($\Delta\sigma/\sigma$):

$$\left(\frac{\Delta N}{N}\right)^2 = \left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta\sigma}{\sigma}\right)^2. \tag{5.7}$$

The luminosity uncertainty is expected to be approximately 10% with early data. Uncertainties on theoretical predictions are at the 10% [69] level for the top pair production cross-section and vary between 3-9% for single top production [70], depending on the production mechanism. In addition, these background estimates fold acceptance and efficiency factors for the background processes, which depend on the theoretical model used and on the simulation of the detector response. Giving an estimate of the uncertainty due to the kinematic acceptance and reconstruction efficiency is beyond the scope of this thesis but should be the subject of further work. Adding in quadrature the sources considered in eqn. (5.7), a lower limit of $\sim$ 15% can be assigned to the uncertainty on the top background estimates.

$\mathbf{Z} \rightarrow$ ll and $\mathbf{W} \rightarrow \tau\nu$:

For $Z \rightarrow \ell\ell$ and $W \rightarrow \tau\nu$ backgrounds, one can take advantage of the well known ratios ($W \rightarrow \ell\nu$)/($Z \rightarrow \ell\ell$) and ($W \rightarrow e\nu$)/($W \rightarrow \tau\nu$) to obtain an estimate of the number of background events without any knowledge of the integrated luminosity L. Because of lepton universality, the production diagrams for $W \rightarrow \tau\nu$ are exactly the same as for $W \rightarrow e\nu$. Therefore the ratio of $W \rightarrow \tau\nu$ to $W \rightarrow e\nu$ events after selection is:

$$f_{W\rightarrow\tau\nu} = \frac{\sigma_W \cdot \mathrm{Br(W \rightarrow \tau\nu)} \cdot \epsilon(\mathrm{W} \rightarrow \tau\nu)}{\sigma_W \cdot \mathrm{Br(W \rightarrow e\nu)} \cdot \epsilon(\mathrm{W} \rightarrow \mathrm{e}\nu)} \tag{5.8}$$

where Br are the branching ratios and $\epsilon(W \to \tau\nu)$ and $\epsilon(W \to e\nu)$ are the efficiencies for selecting a $W \to \tau\nu$ or $W \to e\nu$ event. The ratio of efficiencies is taken from MC.

Similarly, for $Z \to \ell\ell$ events, one can use the ratio R ($\sigma(W \to \ell\nu)/\sigma(Z \to \ell\ell)$):

$$f_{Z\to ee} = \frac{\sigma_Z \cdot \mathrm{Br(Z \to ee)} \cdot \epsilon(\mathrm{Z \to ee})}{\sigma_W \cdot \mathrm{Br(W \to e}\nu) \cdot \epsilon(\mathrm{W \to e}\nu)} = \frac{1}{R} \cdot \frac{\epsilon(Z \to ee)}{\epsilon(W \to e\nu)} \tag{5.9}$$

$$f_{Z\to\tau\tau} = \frac{\sigma_Z \cdot \mathrm{Br(Z \to}\,\tau\tau) \cdot \epsilon(\mathrm{Z \to}\,\tau\tau)}{\sigma_W \cdot \mathrm{Br(W \to e}\nu) \cdot \epsilon(\mathrm{W \to e}\nu)} = \frac{1}{R} \cdot \frac{\epsilon(Z \to \tau\tau)}{\epsilon(W \to e\nu)}. \tag{5.10}$$

The R ratio will be measured at ATLAS with the first 100 pb^{-1} of data, with an expected systematic error of less than 3% [71].

Extracting signal events

From the above equations, the number of signal events can be extracted for each jet multiplicity bin as:

$$N_{W\to e\nu} = \frac{N - N_{QCD-jets} - N_{t\bar{t}} - N_{top} - N_{dib}}{1 + f_{W\to\tau\nu} + f_{Z\to ee} + f_{Z\to\tau\tau}}. \tag{5.11}$$

As a consistency test, the MC sample is divided into two subsamples: one representing pseudo-data and corresponding to exactly 100 pb^{-1} of luminosity, and the second one with the remaining events representing the MC sample. QCD multi-jets are excluded since they cannot be estimated from MC. This is a purely academic exercise to test that the procedure is self-consistent. Since the detector simulation and the input cross-sections are the same for pseudo-data and MC, no systematic bias is expected in the predicted number of events. The results of the test are shown in Table 5.6: true and predicted number of events differ by less than 0.2%.

5.3.2 Data-driven background estimations using E_T^{miss} shape

Especially in the early running phase, it is preferable to rely on the MC estimations as little as possible, since these will require validation and strongly depend on the detector description, as well as on the theoretical model used as input. Furthermore, many background processes suffer

	$N_{jets} = 1$	$N_{jets} = 2$
N candidate (pseudo-data)	39651 ± 199	8396 ± 92
Di-bosons (MC)	118 ± 3	43 ± 2
Single Top (MC)	337 ± 6	287 ± 5
$t\bar{t}$ (MC)	431 ± 3	658 ± 3
$f_{W\to\tau\nu}$ (MC)	4.82 ± 0.09	5.36 ± 0.02
$f_{Z\to\tau\tau}$ (MC)	0.55 ± 0.01	0.51 ± 0.02
$f_{Z\to ee}$ (MC)	0.39 ± 0.01	0.38 ± 0.02
$N_{W\to e\nu}$ (predicted)	36653 ± 192	6972 ± 89
$N_{W\to e\nu}$ (true)	36579	6964

Table 5.6: Estimation of leptonic backgrounds from MC in the one and two jet bins. A pseudo-data sample equivalent to 100 pb^{-1} was passed through the analysis selection cuts (a track veto cut was included for the two jet bin). A statistically independent sample was used for the MC estimate, as explained in the text. The quoted errors are purely statistical.

from large theoretical uncertainties and, therefore, the real background conditions will only be known from collision data. In this section, a methodology for the data-driven estimation of QCD multi-jet and $t\bar{t}$ background is described. The QCD multi-jet background estimation was developed by the ATLAS W/Z + jets group in [3]. Emphasis will be given to the technique for estimating the $t\bar{t}$ background, since it was developed by the author. This technique was investigated in view of measurements in high multiplicity bins, where $t\bar{t}$ is a formidable background. It is however not applied to the cross-section measurement or ratios for the one and two jets bins described in this thesis: up to $N_{jets} = 2$ background from $t\bar{t}$ is relatively small and the technique is not very reliable, as will be described.

The main idea in data-driven estimates is to identify a discriminating variable x with some separation power between signal and backgrounds and a set of cuts C which allows the selection of a pure background sample. The pure background sample is called a *control sample*, in contrast to the *signal sample*, which is obtained applying the standard $W \to e\nu$ selection. The background shape in x can be extracted from the control sample, provided that the variables used for the cuts C are uncorrelated with the discriminating variable x. The overall background normalisation can then be determined by a fit to x in the signal sample, keeping the background shapes fixed.

In the present case, the discriminative variable x is E_T^{miss}. As shown in Fig. 5.3 (left), this variable has a strong discrimination power between QCD multi-jets and $W \to e\nu$ events. If the E_T^{miss} shape of QCD multi-jet is known, the low E_T^{miss} region (dominated by QCD-jets) can be used to extract the background normalisation and extrapolate the number of background events into the signal region ($E_T^{\text{miss}} > 25$ GeV). By contrast events from top pair production are kinematically very similar to the signal. E_T^{miss} does not provide a good separation power between $W \to e\nu$ + jets and $t\bar{t}$, but if templates for both signal and background are available the relative normalisations can be extracted, as will be shown in this section.

Data-driven estimation of QCD background

A pure QCD multi-jet control sample can be selected by inverting some electron identification requirements. For example, a photon, rather than an electron sample, can be selected both at trigger and offline level. This is equivalent to vetoing the presence of an inner detector track pointing to the EM cluster, while maintaining all other kinematic and calorimeter quality cuts. The sample obtained is dominated by QCD multi-jets that fail the tracking requirement but pass the EM cluster criteria. $W \to e\nu$ + jets events are largely removed from QCD-jet control region, but a small contamination at the level of $\sim$3% remains from events where no inner detector track (matching to the electron EM cluster) is found. Such inefficiencies (when $W \to e\nu$ + jets events fail the signal selection and satisfy the background control sample selection) are mainly due to large amount of material in the inner detector which affects track reconstruction and have to be accounted for using MC. This control sample, after removal of any $W \to e\nu$ contamination, can be used to extract the E_T^{miss} shape of QCD multi-jets. The normalisation of QCD multi-jet background is then extracted by fitting the photon template (for QCD multi-jets) and a MC $W \to e\nu$ + jets template to the signal sample. Figure 5.14 shows the template fit to the data, using a maximum likelihood formalism. It was shown in [3] that the estimated number of QCD multi-jet events is consistent with the truth within statistical uncertainties.

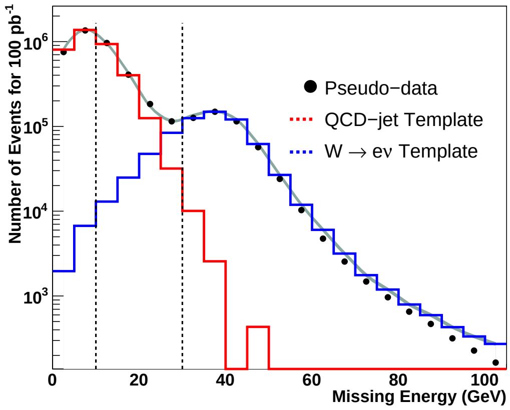

Figure 5.14: Data driven estimation of QCD multi-jet background using the photon selection: the E_T^{miss} distributions for the QCD multi-jet control sample and the $W \to e\nu$ MC are fitted to the pseudo-data sample. The results of the maximum likelihood fit is shown in the grey band. The two dashed lines indicate the range used for the fit. [3]

The assumption in this method is that the transverse missing energy distribution of QCD multi-jet events in the photon and electron samples is the same. This, however, is not obvious. Both samples are dominated by hadronic jets faking electrons or photons, but the electron sample (after medium PID) contains a non-negligible fraction of heavy flavour jets where the heavy flavour quark has decayed into an electron. Such events are more likely to pass medium PID cuts, since they contain real electrons. Additionally, leptonic heavy flavour quark decays contain additional E_T^{miss} and they could therefore modify the E_T^{miss} distribution. Given the available MC statistics, no significant difference was observed between the photon and electron QCD multi-jet samples, but further investigations are needed. This and alternative approaches for estimating the background from QCD-jets investigated by the ATLAS W/Z+ jets group are documented in [3].

Data-driven estimation of $t\bar{t}$ background

A pure $t\bar{t}$ control sample is selected using the so called b-tagging, ie. the identification of jets stemming from the hadronization of b quarks. In fact, $t\bar{t}$ events always contain two b-jets originating from the $t \to Wb$ decay. b-tagging algorithms make use of several properties of b-jets,

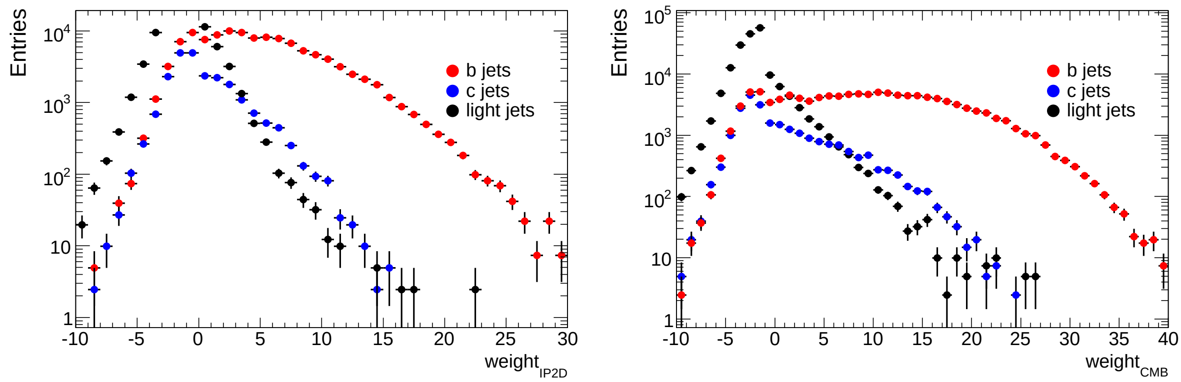

Figure 5.15: B-tagging weight of light, b and c-quark jets using two different b-tagging algorithms: IP2D (left), which uses information on the transverse impact parameter and a combined algorithm CMB (right), which uses a combination of transverse and longitudinal impact parameter and secondary vertex information.

such as for example the long lifetime [8], to distinguish them from light jets. A likelihood method is used to assign a weight to each jet, based on some discriminating variables, such as for example the impact parameter. Requiring that there are at least two jets in the event with high b-tagging weight, a relatively pure $t\bar{t}$ control sample can be obtained. Two different b-tagging algorithms are considered (see [2] for a detailed description): IP2D (which uses information on the transverse impact parameter) and a combined algorithm CMB (which uses a combination of transverse and longitudinal impact parameter and secondary vertex information). The discrimination power of these algorithms in distinguishing between jets containing a b-quark from those containing c or light quarks is shown in Fig. 5.15.

Considering, for the moment, only events from top pair production and $W \rightarrow e\nu$ + jets and ignoring all other sources of background, the steps for estimating the $t\bar{t}$ background are the following:

- Start from a pseudo-data sample where all analysis cuts have been applied with the exception of the cut on transverse missing energy. We will refer to this as the *signal sample*,

[8] A hadron containing a b-quark has a lifetime of the order of 1.5 ps. A b-jet with p_T = 50 GeV will travel on average 3mm in the transverse plane before decaying [2]. Such a displaced vertex can be identified by measuring the impact parameters of the tracks from b-hadron decay products, which tend to be larger than for tracks originating from the primary vertex.

since it is obtained using the signal selection cuts [9].

- Select a pure sample of $t\bar{t}$ events requiring at least two b-tagged jets in the event. We will refer to this as the *control sample*. It should be stressed that, contrary to the QCD background estimation method, the $t\bar{t}$ control sample is not orthogonal to the signal sample, but is a subset of it.

- Extract the E_T^{miss} shape of $t\bar{t}$ events by fitting a non-parametric probability density function (Pdf) [10] to the control sample using the RooFit package [73] (see Fig. 5.16 left). The E_T^{miss} shape for $W \to e\nu$ is taken from Monte Carlo.

- Determine the signal and background normalisation in the signal sample by a global fit in E_T^{miss}. The extended maximum likelihood method on binned dataset is used for the minimisation. The shapes of W + jets and $t\bar{t}$ have been fixed from MC or the control sample, respectively. An example of the global fit is shown in Fig. 5.16 right.

- Extract the number of $t\bar{t}$ events in the signal region for $E_T^{\text{miss}} > 25$ GeV from the predetermined normalisation and Pdf.

The only assumptions in this technique are the absence of correlation between b-tagging weight and E_T^{miss} (such that the E_T^{miss} spectrum in $t\bar{t}$ events is not biased by the b-tagging cuts) and negligible $W \to e\nu$ +jets contamination in the $t\bar{t}$ control sample. The first assumption was checked on MC by comparing the E_T^{miss} distribution in $t\bar{t}$ events in the signal region, ie. before b-tagging, and control region, ie. after b-tagging. Figure 5.17 shows the two distributions overlaid (top) and their ratio (bottom) for $N_{jets} = 2$ (left) and $N_{jets} = 3$ (right). In this case, the $t\bar{t}$ control sample has been selected requiring two jets with b-tagging weight (CMB algorithm) $w_B > 0.6$. A χ^2 and Kolmogorov-Smirnov test have also been performed (the probabilities re-

[9] This sample will contain $W \to e\nu$ + jets events as well as background events, such as $t\bar{t}$. In data other background processes, such as QCD-jets, will also be present. For the moment these are ignored but will need to be considered in future work.

[10] This is a continuous probability density function built as a superposition of Gaussians. The algorithm used is documented in [72].

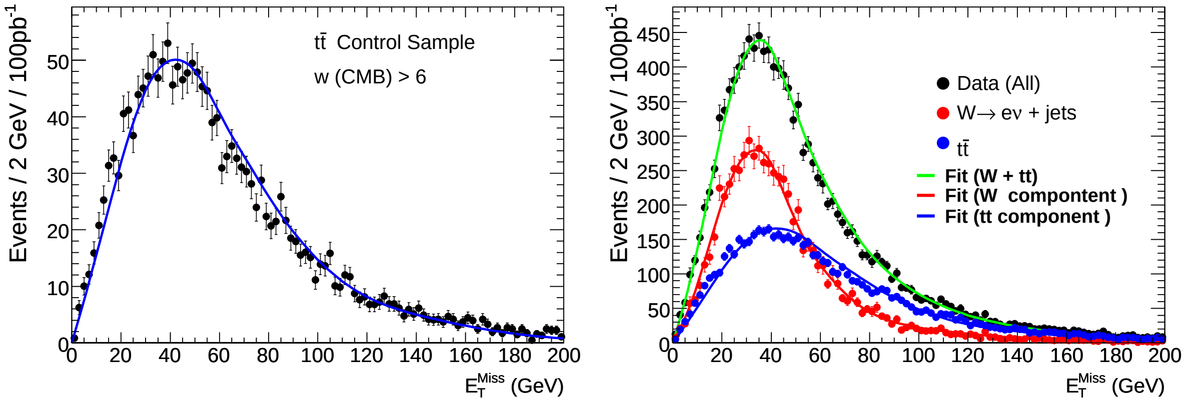

Figure 5.16: Example of $t\bar{t}$ background data-driven estimation method for $N_{jets} \geq 3$. (Left) $t\bar{t}$ background control sample fitted to a generic Pdf. (Right) global fit of signal Pdf (extracted from MC) and background Pdf (extracted from the control sample) in $E_T^{\rm miss}$ to extract the signal and background normalisations.

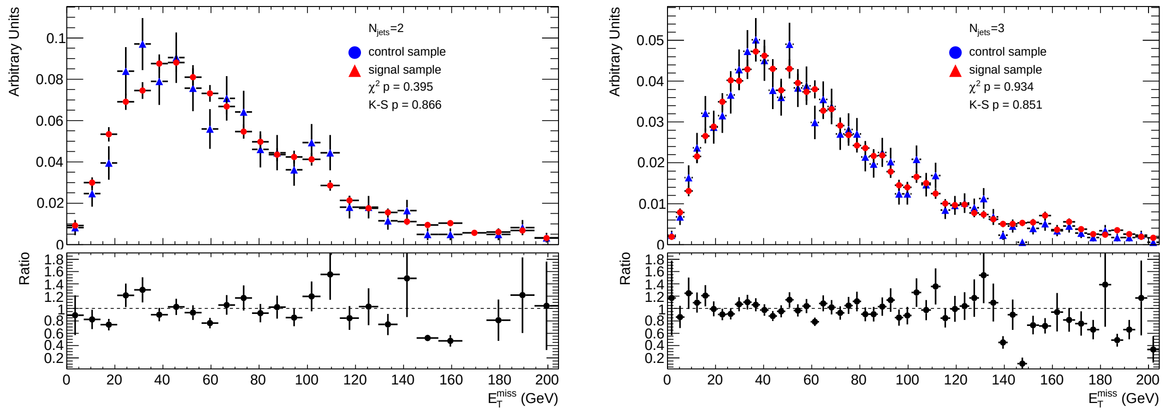

Figure 5.17: Comparison of $E_T^{\rm miss}$ shape in $t\bar{t}$ events exclusively, before (signal sample) and after (control sample) a cut on b-tagging weight (requiring at least two jets in the event with $w_b(CMB) > 6.0$). Shown is also the result of a χ^2 and a Kolmogorov-Smirnov test of compatibility between the two distributions. The bottom plot shows the ratio of the two distributions. Left: $N_{jets} = 2$, right: $N_{jets} = 3$.

turned by the tests are shown on the plot) and confirm the compatibility of the two distributions. Similar results were found for tighter/looser cuts on the weight and using the IP2D algorithm.

Signal contamination into the control sample is likely to arise from $W \to e\nu + bb$ jet events or from $W \to e\nu$ + light jets, where the light jets have been mis-tagged. $W \to e\nu + bb$ jet production, however, has a small production rate (about 1/100 of the total $W \to e\nu$ + jets cross-section) and light jet contamination can be reduced by increasing the threshold on the weight w_B, at the expenses of $t\bar{t}$ statistics. This is illustrated in Fig. 5.18, where the expected $t\bar{t}$ efficiency and control sample purity are plotted as a function of the b-tagging weight for both

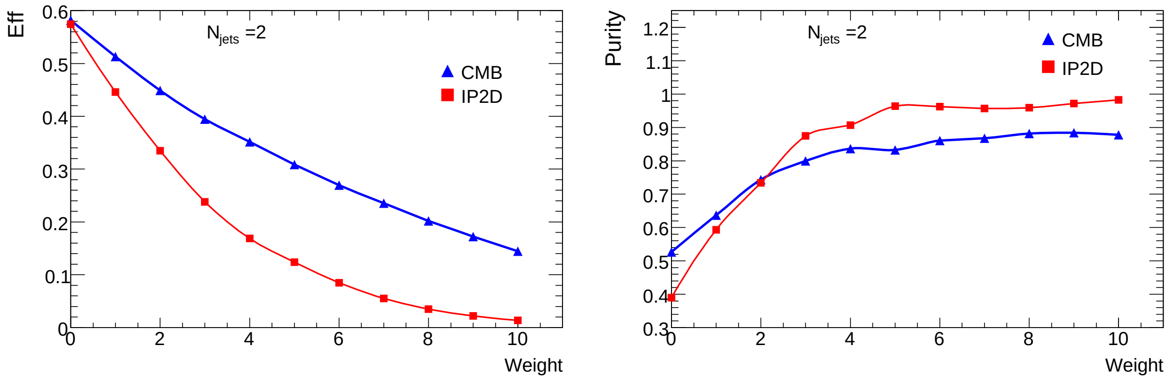

Figure 5.18: Performance of the Combined and IP2D B-Tagging algorithm in selecting a pure $t\bar{t}$ control sample. The efficiency (left) in selecting $t\bar{t}$ events and the purity (right) of the control sample ($\frac{N_{t\bar{t}}}{N_{t\bar{t}}+N_{Wjets}}$) are considered. $Wbb + jets$ events are included in the purity calculation, whereas processes other than $W \rightarrow e\nu$ + jets and $t\bar{t}$ are not.

CMB and IP2D algorithms in the two jet bin. Only contamination from $W \rightarrow e\nu$ + jets is considered [11]. The choice of the threshold on w_B for the control sample selection must be a trade off between purity and efficiency. For the two jet bin, a purity greater than 90% can be achieved with the IP2D algorithm cutting on $w_B > 5$, but the corresponding efficiency is smaller than 15%. Given that, after the full selection, approximately 1000 $t\bar{t}$ events are expected for $100pb^{-1}$ of data (see Table 5.1), the control sample would result in a few hundred events and the background template would therefore be statistically limited. This issue is the main limitation to the application of the method to low multiplicity bins. For high multiplicities, where the $t\bar{t}$ contribution is much larger, it is less problematic since the statistics are high and the $W \rightarrow e\nu$ + jets contamination is smaller. As default, it was chosen to use the CMB algorithm, since it yields higher efficiency for the same purity, with a threshold on the weight of $w_B = 6$. Results for $N_{jets} = 2$, 3 and ≥ 3 are summarised in Table 5.7. The deviation of the predicted number of events (using the data-driven method) from the MC truth is below 10%.

Systematic Uncertainties on $t\bar{t}$ background estimation

Systematic uncertainties on the background estimation are evaluated by using the full MC statistics available and varying:

[11] The signal $W \rightarrow e\nu$ + jets ALPGEN samples used throughout the analysis contain only light jets. Therefore, in order to check the contamination from $W \rightarrow e\nu$ +jets into the control sample, a sample of $W \rightarrow e\nu + bb\,jets$ was added to the $W \rightarrow e\nu$ +jets sample.

N_{jets}	$N^{True}_{W\to e\nu}$	$N^{True}_{t\bar{t}}$	$N^{Pred}_{t\bar{t}}$	$\Delta N_{t\bar{t}}/N^{True}_{t\bar{t}}(\%)$
2	15271 ± 170	963	921 ± 126	-4.4
3	4067 ± 68	1666	1517 ± 123	-8.9
≥ 3	5385 ± 73	5269	5281 ± 1134	0.2

Table 5.7: Number of signal events and MC true/predicted (using the data-driven estimate) number of $t\bar{t}$ events after full selection (including a E_T^{miss} cut of 25 GeV) for different multiplicity bins. The last column shows the deviation (%) of the predicted number of events from the MC truth. The $t\bar{t}$ control sample was selected requiring at least two b-jets with b-tagging weights $w_b(CMB) > 6$.

- B-tagging algorithm and threshold on the b-tagging weight used for the control sample selection: this effectively corresponds to changing the purity of the control sample and the available statistics for extracting the $t\bar{t}$ template. As an example, let us consider the bin $N_{jets} \geq 3$: in Table 5.8 the number of $t\bar{t}$ events and the contamination of $W \to e\nu$ + jets in the control sample are reported for different b-tagging selections. In this case, the contamination varies by a factor of three and the statistics by a factor of four, but for $N_{jets} = 2$, the variation in the $W \to e\nu$ + jets contamination is even more dramatic, as already shown in Fig. 5.18. This is one of the reasons (together with the statistical limitation discussed above) why this data-driven technique cannot be applied for such low multiplicities. The deviation of the predicted number of events from the MC truth and from the default cut value (using $w_B(CMB) > 6$) are shown in the last two columns.

- Template extraction: a histogram is used, rather than a non-parametric Pdf fit to the control sample (MC sample) for $t\bar{t}$ ($W \to e\nu$ + jets).

- Binning: the number of bins is halved/doubled.

Systematic uncertainties for each multiplicity bin are summarised in Table 5.9. For the uncertainty on the b-tagging threshold/algorithm used for the control sample selection, only the average deviation from the default value (last column of Table 5.8) is shown. Considering these sources of uncertainty as independent, the overall systematic uncertainty is given by the quadrature sum of the average values:

		$N_{jets} \geq 3$			
Algo	Cut	N(CS) $\cdot 10^3$	Cont (CS)	$\Delta N^{Truth}(\%)$	$\Delta N^{Pred}(\%)$
CMB	$w_B > 9$	4.9	1.5 ± 0.1	$+0.7$	$+0.4$
	$w_B > 8$	5.8	1.7 ± 0.2	-2.4	-2.6
	$w_B > 7$	6.8	1.7 ± 0.2	-0.1	-0.4
	$w_B > 6$	7.8	2.0 ± 0.2	$+0.2$	n/a
	$w_B > 5$	8.9	2.4 ± 0.2	$+0.9$	$+0.61$
	$w_B > 4$	10.2	2.9 ± 0.2	$+4.0$	$+3.8$
IP2D	$w_B > 6$	2.4	1.4 ± 0.2	-11.4	-11.6
	$w_B > 5$	3.6	1.7 ± 0.2	$+3.8$	$+3.6$
	$w_B > 4$	4.9	2.1 ± 0.2	-3.8	-4.0
	$w_B > 3$	6.9	3.0 ± 0.4	-10.5	-10.7
	$w_B > 2$	9.8	4.4 ± 0.4	$+0.5$	$+0.22$

Table 5.8: Systematic uncertainties on the determination of background from top pair production to $W \rightarrow e\nu + \geq 3$ jets arising from different control sample selections. The b-tagging algorithm and the threshold on the b-tagging weights are varied, resulting in different numbers of events ($N(CS)$, second column) and level of $W \rightarrow e\nu$ + jets contamination ($Cont(CS)$, third column) in the $t\bar{t}$ control sample. The last two columns show the deviation (in $\%$) from the true number of events (ΔN^{Truth}) and from the number of events predicted using default cuts (ΔN^{Pred}).

$$\Delta N_{t\bar{t}}(\%) = \sqrt{< \Delta_{B-tagging} >^2 + < \Delta_{Template} >^2 + < \Delta_{Binning} >^2}.$$

One source of systematic uncertainty which has not been considered is the input MC template for the $W \rightarrow e\nu$ + jets, which is sensitive to the theoretical model used and the simulation of E_T^{miss} scale and resolution. This should be subject of further work.

The overall uncertainty is very large ($> 35\%$) for the two jet bin. For higher multiplicities, as statistics and purity of the control sample improve, the systematic uncertainty decreases significantly and the predicted number of events is in agreement with the truth (see last column of Table 5.7) within errors.

Conclusions and Further Work

It can be concluded that this data-driven method of estimating $t\bar{t}$ background is a promising technique for high jet multiplicities (three or more jets). Since top pair production can even dominate the signal for such multiplicities, it will be of crucial importance to have alternative methods like this one, which have a reduced dependence on MC. However for early data, when

Systematics	Variation	$\lvert\Delta N^{Pred}\rvert(\%)$		
		$N_{jets} = 2$	$N_{jets} = 3$	$N_{jets} \geq 3$
B-tagging	Vary algorithm or cut on weight (default: $w_B(CMB) > 6$)	14.4	9.7	3.8
Template	Use histogram template (default: generic Pdf)	30.9	5.8	2.2
Binning	Double number of bins	14.8	2.9	1.3
	Halve number of bins	9.8	4.1	0.2
Overall	Sum in quadrature	36.2	11.8	4.4

Table 5.9: Systematic uncertainties on the $t\bar{t}$ background estimation to $W \rightarrow e\nu$ + 2, 3 and ≥ 3 jets from different sources, as explained in the text. The uncertainty from b-tagging (first line) is calculated as the average deviation obtained by varying the b-tagging algorithm and threshold cuts (see Table 5.8).

limited statistics are available and therefore only low multiplicities are accessible, this method is not accurate enough to be competitive with a MC estimate. The following points deserve further studies:

- It was observed that the main limitation to the method is the lack of statistics for low multiplicities, since the control sample is a subset of the signal sample. To overcome this limitation, alternative control sample selections could be investigated and ideally an orthogonal sample should be selected. Defining an orthogonal sample without biasing a variable which depends on event topology, such as $E_T^{\rm miss}$, is very difficult. For example, an orthogonal sample to the 2-jet bin with much higher statistics could be a sample of events with 3-jets, of which at least two from b-quarks. However, this selection biases $E_T^{\rm miss}$: the relative contamination of different $t\bar{t}$ decay modes (fully and semi-leptonic) varies with jet multiplicities (see Table 5.2) and this results in a different $E_T^{\rm miss}$ spectrum due to number of real neutrinos in the event. This deserve further investigations and the possibility of using different observables from $E_T^{\rm miss}$ should also be considered.

- The studies described in this section considered exclusively $W \rightarrow e\nu$ +jets and $t\bar{t}$ events. The contamination from other processes in the $t\bar{t}$ control sample, in particular QCD multi-jets, should also be investigated. If the contamination was found to be significant, a QCD multi-jet background template should be used to subtract it from the $t\bar{t}$ control sample, in a similar way to the QCD multi-jet background subtraction from the signal sample.

- Other sources of background have to be considered also in the signal sample, when extracting the overall normalisation. This implies that the templates for QCD multi-jets and $t\bar{t}$ (extracted from the relative control samples) should be fitted together to the signal sample. In addition, templates for $Z \to \ell\ell$ and $W \to \tau\nu$ should also be added, with shape and normalisation fixed from MC (the normalisation for $Z \to \ell\ell$ and $W \to \tau\nu$ can be determined as explained in Section 5.3.1).

5.4 Conclusions

This chapter described how different processes can mimic the decay of a $W \to e\nu$ accompanied by jets and proposed methodologies for reducing and estimating these backgrounds.

QCD multi-jets are the most significant background but also the least known. Data-driven estimations of this background are therefore mandatory. A data-driven method which uses a photon sample to extract a QCD-jet template for the E_T^{miss} spectrum was described. Although the technique seems promising, present studies are limited by the scarce MC statistics. One of the possible shortcomings is the potential bias introduced on E_T^{miss} by the photon selection with respect to the electron selection. Validation of this technique with data and identifying potential biases will therefore be necessary. A rigorous evaluation of systematic uncertainties by comparison with alternative data-driven methods is also going to be crucial for a reliable estimation of this background.

Other important sources of backgrounds are processes which contain real leptons and are going to be estimated using MC. For processes such as $W \to \tau\nu$ and $Z \to \ell\ell$, the well known ratios to the $W \to e\nu$ production cross-section can be exploited to decrease the systematic uncertainties on the background estimation. A consistency test was performed, showing that the procedure for leptonic background estimation is consistent within statistical uncertainties. Systematic uncertainties remain to be evaluated and should be subject of further work. This will require assessing the accuracy of the detector simulation, which can be achieved comparing data to MC or running the analysis on samples simulated with different experimental conditions.

Leptonic backgrounds can be partially reduced by vetoing the presence of an extra lepton. Two lepton vetos were investigated, one using tracking and the other calorimetry information. It was discussed what factor should drive the decision on whether or not to apply these vetos. Based on the available MC simulation and on the assumption that leptonic backgrounds can be known with an accuracy of 20%, it was suggested that it would be favourable applying a track veto to the 2-jet bin. This would decrease the overall background contamination from 26% to 20% and would lead to an overall reduction of the systematic uncertainty on a cross-section measurement. This discussion should serve as a baseline for future studies. When the real background conditions will be known and systematic uncertainties on lepton background estimations will be available, a similar assessment could be made.

Among leptonic backgrounds, top pair production is the most important source of signal contamination, particularly at high multiplicities. Although this background is much better understood than QCD-jets, its large impact on measurements with high numbers of jets makes it desirable to have alternative data-driven techniques for its estimation. An in-situ method that uses b-quark tagging to extract a pure $t\bar{t}$ control sample was developed. Similarly to the QCD-jet background estimation, the background E_T^{miss} distribution is extracted from the control sample and normalised to signal sample. It was shown that this technique is applicable for high multiplicity bins ($N_{jets} \geq 3$), where it predicts the number of $t\bar{t}$ background events within an accuracy of 10%. At low multiplicities (1, 2 jets) it is limited by the poor statistics of the control sample and suffers from unacceptable large systematic uncertainties. Suggestions for further work were given in the dedicated section. Given that $t\bar{t}$ becomes a dominant background for $N_{jets} \geq 3$, this method should therefore be considered as a powerful alternative to MC estimates for measurements at high multiplicities.

Chapter 6

In-situ Determination of Electron Efficiencies

Early data from the LHC is going to be of crucial importance for the commissioning of the ATLAS detector. Within 10-100 pb^{-1} it will be possible to have first data-driven estimates of electron trigger and offline reconstruction efficiencies and to compare these to MC expectations. Such studies will be important to understand the detector performance and validate the MC tools. Furthermore, an accurate knowledge of lepton efficiencies will allow to control the systematic uncertainties on electroweak boson cross-section measurements.

This chapter describes the so called Tag-and-Probe technique for the in-situ determination of electron efficiencies exploiting $Z \to ee$ events. After a brief introduction to define what is meant by the term efficiency (Section 6.1), the principles behind the Tag-and-Probe method, including the mathematical formalism, are described in Section 6.2. The second part of the chapter is dedicated to the application of the Tag-and-Probe method to $Z \to ee$ events in the context of electron identification efficiencies. After a summary of the signal selection cuts, global and differential efficiencies are discussed. Systematic uncertainties on the data-driven technique are also covered, with particular emphasis on the background subtraction. Finally, prospects for early data measurements with different assumptions on the luminosity are discussed. The next chapter will show how Tag-and-Probe efficiencies can be applied to the cross-section measurement of W + jets.

6.1 Factorization of Efficiencies

The reconstruction efficiency for an electron typically includes offline reconstruction and some identification criteria. If, in addition, the electron is used to trigger the event, a trigger efficiency is also needed. The overall efficiency can therefore be factorized as:

$$\epsilon = \epsilon_{\text{reco}} \, \epsilon_{\text{id}} \, \epsilon_{\text{trig}} \tag{6.1}$$

where ϵ_{reco} is the absolute offline reconstruction efficiency (ie. the probability that a true electron within the kinematic acceptance is reconstructed offline), ϵ_{id} is the identification efficiency relative to an offline reconstructed electron and ϵ_{trig} is the trigger efficiency relative to offline identification cuts.

This chapter describes the Tag-and-Probe method for the extraction of the identification efficiency (ϵ_{id}) exclusively, but the methodology can be easily extended to other quantities (such as ϵ_{reco} and ϵ_{trig}). ϵ_{id} is defined as the probability for an offline reconstructed electron to pass the standard cut-based identification criteria, denoted as $loose$, $medium$ and $tight$ (see Chapter 4, section 4.1.2). Since the W analysis requires a medium electron, only medium PID efficiencies are discussed in this thesis. All studies in this chapter have been performed with a high statistics PYTHIA sample (equivalent to $\sim 5fb^{-1}$) which simulates the Drell-Yan process $pp \rightarrow \gamma^*/Z \rightarrow ee$.

6.2 The Tag-and-Probe Method

The Tag-and-Probe (TP) method relies on double object final states to extract the efficiency. In this study, $Z \rightarrow ee$ events are used. One electron from the Z boson (the tag) is required to satisfy a set of tight requirements, allowing separation of Z events from the background. The efficiency of interest is then measured using the second electron from the Z (the probe).

The approach adopted here to calculate the efficiency is based on counting the number of TP pairs (rather than events) satisfying a certain selection. Let us assume that the total number of Z events in the data sample is N and that we want to extract the efficiency (ϵ) for a certain cut (C). The observable is the number of electron pairs where the tag electron satisfies the requirement C and the probe may or may not satisfy it. One can then define the number of pairs with at least one (N_1) or with exactly two electrons (N_2) satisfying the cut:

$$N_1 = \epsilon\,(1-\epsilon)\,N \;+\; (1-\epsilon)\,\epsilon\,N \;+\; 2\,\epsilon^2\,N = 2\,\epsilon\,N \tag{6.2a}$$

$$N_2 = 2\,\epsilon^2\,N. \tag{6.2b}$$

Note that events with two electrons satisfying the selection requirement result in two TP pairs, with tag and probe electrons interchanged. These pairs are considered statistically independent (since a different electron is probed in each case) and are therefore counted twice in the efficiency calculation. From equations 6.2, the efficiency ϵ can be easily extracted. It is the ratio of electron pairs where both tag and probe electrons pass the desired cut to the pairs with at least one electron passing the cut:

$$\frac{N_2}{N_1} = \frac{2\,\epsilon^2\,N}{2\,\epsilon\,(1-\epsilon)\,N \;+\; 2\,\epsilon^2\,N} = \epsilon. \tag{6.3}$$

In general, one might want to apply different selection criteria to the tag and probe electrons. This is desirable, in particular, if we are measuring the efficiency for a loose selection and therefore need to apply tighter requirements on the tag to reduce the background. The efficiency for selecting a tag electron can then be factorized in terms of the efficiency under study (ϵ) and the efficiency of an additional tight requirement ($\epsilon_{\rm t}$). Each electron can satisfy the probe selection and, additionally, the tight criteria:

Selection (ϵ)	Tight ($\epsilon_{\rm t}$)	Probability
✗	✗	$(1-\epsilon)$
✔	✗	$\epsilon\,(1-\epsilon_{\rm t})$
✔	✔	$\epsilon\,\epsilon_{\rm t}$

Redefining N_1 as the number of pairs with at least one tag and N_2 as the number of pairs with one tag and one probe passing the selection, eqn. (6.2) can be generalised as follows:

$$N_1 \;= \epsilon\,\epsilon_{\rm t}\,[\,(1-\epsilon)\;+\;\epsilon\,(1-\epsilon_{\rm t})\,]\,N\;+\;[\,(1-\epsilon)\;+\;\epsilon\,(1-\epsilon_{\rm t})\,]\,\epsilon\,\epsilon_{\rm t}\,N\;+\;2\,\epsilon^2\,\epsilon_{\rm t}^2\,N = \;2\,\epsilon\,\epsilon_{\rm t}\,N \tag{6.4a}$$

$$N_2 \;= \epsilon\,\epsilon_{\rm t}\,[\,\epsilon\,(1-\epsilon_{\rm t})\,]\,N\;+\;N\,[\,\epsilon\,(1-\epsilon_{\rm t})\,]\,\epsilon\,\epsilon_{\rm t}\,N\;+\;2\,\epsilon^2\,\epsilon_{\rm t}^2\,N\;=\;2\,\epsilon^2\,\epsilon_{\rm t}\,N. \tag{6.4b}$$

It is straightforward to show that the ratio of N_2 and N_1 in eqn. (6.4) gives the desired efficiency ϵ. Any additional event or object pre-selection cut (such as the tag pre-selection) only adds a factor $\epsilon_{\rm r}$ to both N_1 and N_2 and therefore leaves the final result unchanged.

Differential efficiencies can be considered as a further generalisation. When probing the efficiency in bin i, we consider probe electrons only in bin i, whereas the tag electron can fall in any bin. If both tag and probe electrons fall in the same bin and the probe also passes the tag selection, then both electrons are probed (resulting into two TP pairs). The algebra becomes more complicated, but it can be shown that this counting method gives again the desired efficiency for bin i.

Statistical uncertainties on the efficiencies ($\Delta\epsilon$) are calculated adopting the binomial approach. The application of a cut can be considered as a binomial process with probability of success being equal to the true efficiency ϵ_{true}[1]. Given N_1 trials, the expectation value for the number of successes is $\langle N_2 \rangle = \epsilon N_1$ and the standard deviation is:

$$\sigma_{N_2} = \sqrt{var(N_2)} = \sqrt{\epsilon_{\text{true}} \, (1 - \epsilon_{\text{true}}) \, N_1}. \tag{6.5}$$

Inserting in eqn. (6.5) the estimate ϵ, for the ϵ_{true}, yields the result:

$$\Delta\epsilon = \sqrt{\frac{\epsilon \, (1 - \epsilon)}{N_1}}. \tag{6.6}$$

The binomial error approximation formula of eqn. (6.6) has the limitation that it does not give sensible results for $\epsilon = 0$ and $\epsilon = 1$. In both cases the uncertainty on the efficiency is equal to 0 independently of the number of trials N_1. An alternative approach which solves this inconsistency uses Bayes' theorem [74]. In this analysis, however, the binomial approximation is adopted for simplicity.

6.3 $Z \to ee$ Event Selection

$Z \to ee$ events are particularly suitable for extracting electron efficiencies for two main reasons:

- High statistics: Z bosons are going to be produced at a very high rate at the LHC ($\sim$ 1000 Zs per pb per decay channel);
- Small background: Z bosons guarantee a clean sample and residual backgrounds are relatively easy to calculate thanks to the Z resonance.

It is of course desirable to apply efficiencies extracted from $Z \to ee$ to other physics processes, for example to $W \to e\nu$ events. However, the measured overall efficiency is the result of a convolution of the true efficiency with the kinematic distribution of the process under study.

[1]The number of events N2 passing the cuts is given by a Binomial distribution: $P(N_2; \epsilon, N_1) = \binom{N_1}{N_2} \epsilon^{N_2} \, (1-\epsilon)^{N_1 - N_2}$, with mean $N_2 = \epsilon N_1$ and variance $V(N_2) = \Delta^2_{N_2} = N_1 \, \epsilon \, (1 - \epsilon)$.

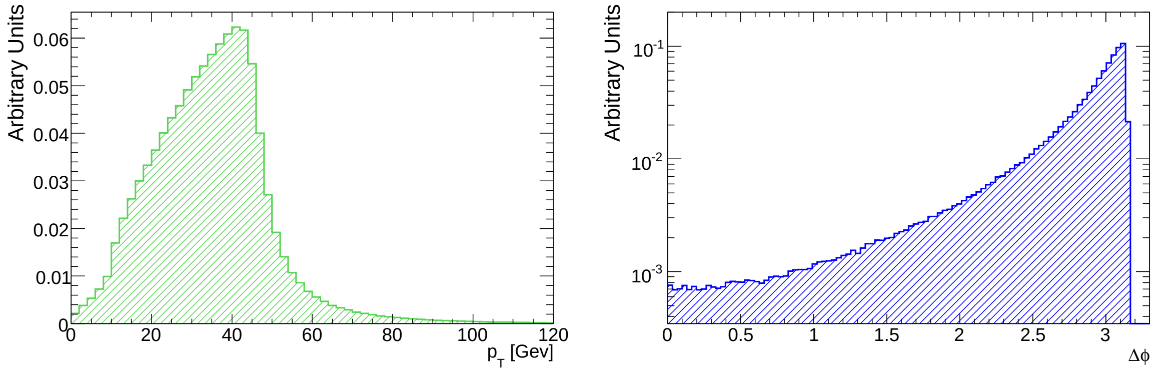

Figure 6.1: Distribution of true electrons from a Z as a function of transverse momentum p_T (left) and the angle $\Delta\phi$ between the two leptons (right). The histograms are normalised to unity.

The global efficiency from $Z \to ee$ events will therefore directly translate only to processes which are kinematically very similar. Application to kinematically different physics processes introduces the need for differential efficiencies, which allow unfolding the true efficiencies from the measured ones.

The kinematic distributions of the electrons from a Z boson are shown in Fig. 6.1. The transverse momentum spectrum peaks around 40 GeV, making Z events good candidates for the extraction of efficiencies in the transverse energy range 20-100 GeV. [2] Furthermore, thanks to the massive boson, the two leptons tend to be produced back to back. The right-most plot shows the angle $\Delta\phi$ between the two electrons: about 85% of the lepton pairs are emitted at angle $\Delta\phi$ greater than $\frac{3}{4}\pi$.

In this analysis, we focus on the measurement of electron medium PID efficiencies. The following requirements are applied to select good TP electron pairs:

- The single electron trigger $\mathrm{e20\_loose}$ is passed by the tag electron.
- One electron satisfies the tag selection cuts and the second electron satisfies the probe pre-selection cuts, as described below.
- The invariant mass (M_{ee}) of the electron pair is in the window 80-100 GeV.

[2] Figure 6.1 was obtained from a PYTHIA inclusive sample, where the Z boson is produced mainly at rest. The statistics above p_T of 80 GeV are rather low. However, if events with a high p_T jet are selected, where the Z boson is boosted, a higher kinematic range can be achieved.

- Additionally, one of these cuts can be applied to improve the sample purity: tag and probe electrons are separated in $\Delta\phi$ by at least $\frac{3}{4}\pi$ or have opposite charge. We will refer to these as the $\Delta\phi$ and the $\pm$ TP configurations. The absence of these cuts we will be referred to as the $Base\,TP$ configuration.
- The sample obtained after the application of the above criteria is the *probe collection* and represents the denominator N_1 of the efficiency calculation (see eqn. (6.3)). After requiring the probe to also satisfy the medium PID cuts, the *medium probe collection* (N_2) is obtained.

The cuts applied to tag and probe are shown in Table 6.1. Both electrons have to satisfy an acceptance cut in η and p_T. In addition, the tag is required to be outside the barrel-endcap transition region (crack region), where the detector performance is deteriorated. Finally, an identification requirement (typically tight PID, but also medium is foreseen if the background is under control) is applied to the tag to reduce the background. A geometrical matching in ΔR between offline tag electron and LVL1/LVL2/EF trigger objects is also performed. The trigger matching ensures that the e20_loose chain is fired by the tag, such that the probe is completely unbiased with respect to the trigger. The effect of the trigger was investigated using different offline-trigger matching criteria (changing the matching cone radius and/or requiring a match only between offline and EF, ignoring LVL1 and and LVL2). In $Z \to ee$ events, the trigger efficiency with respect to a tight (medium) tag was found to be 99.8% (99.7%) and independent of the matching criteria used. The signal and background yields in the probe collection are given in Tables 6.2 and 6.3 for a medium and tight tag respectively [3]. Note that the number of electron pairs rather that the number of events is considered, since this is the quantity of interest for the efficiency studies. Yields are normalised to 100 pb^{-1} of data. Statistical errors due to the size of the available MC samples are given. The QCD di-jet sample, in particular, suffers from very large statistical errors. For this reason, the full cut flow table is only shown for the case of a medium tag (Table 6.2). With a tight tag requirement only five events are left in the QCD-jet

[3] Only the probe collection is shown because this is where background contamination is more significant. After medium PID cuts on the probes, the background contamination is much lower (the jet rejection factor is increased by a factor of ~ 70).

Tag/Probe	Type	Cut
Tag	acceptance	$\|\eta\| < 2.47$
		$\|\eta\|$ outside cracks $(1.37 - 1.52)$
		$p_T > 25$ GeV
	trigger	LVL1 match: $\Delta R < 0.2$
		LVL2 match: $\Delta R < 0.1$
		EF match: $\Delta R < 0.1$
	tight	PID = Tight (Medium)
Probe	acceptance	$\|\eta\| < 2.47$
		$p_T > 25$ GeV

Table 6.1: Cuts applied on tag and probe electrons.

Medium Tag					
Cuts		$\mathrm{N_{TP}}$/Eff Signal	N_{TP} Background		S/B
		$Z \to ee$	Di-jets	$W \to e\nu$	
C0	PID + Acceptance	65540 ± 36	$(30 \pm 6) \cdot 10^3$	2131 ± 32	2.0 ± 0.4
C1	C0 + Trigger	$(99.73 \pm 0.01)\%$	$(28 \pm 5) \cdot 10^3$	2125 ± 32	2.2 ± 0.4
C2	C1 + $\Delta\phi$	$(89.43 \pm 0.02)\%$	$(20 \pm 5) \cdot 10^3$	1583 ± 28	2.7 ± 0.6
C3	C1 + Opposite Charge	$(95.82 \pm 0.01)\%$	$(10 \pm 3) \cdot 10^3$	1240 ± 24	5.4 ± 1.5

Table 6.2: Number of signal and background pairs (N_{TP}) in the invariant mass window 80-100 GeV where the tag electron is required to pass the medium PID cuts. Events are normalised to 100 pb^{-1} of integrated luminosity. The signal efficiency with respect to baseline cut (C0) is also shown. Errors are due to the MC statistics.

sample (first line of Table 6.3), making it meaningless to investigate additional cuts. For the signal, the efficiency of each cut relative to the basic selection (C0) is shown. This is calculated in terms of selected electron pairs in the signal peak. The signal to background ratio is also given in the last column. The sources of background considered are QCD jets and $W \to e\nu$. Backgrounds from $Z \to \tau\tau$ and $W \to \tau\nu$ are negligible (below 1%) and therefore not shown.

The PID requirement on the tag keeps the background under control, with a signal to background ratio $\left(\frac{S}{B}\right)$ of $2:1$ for a medium tag and larger than $7:1$ for a tight tag. The opposite charge requirement further improves the purity of the probe sample, whereas the minimum $\Delta\phi$ separation does not significantly improve the signal to background ratio. This is expected since di-jets tend to be also produced back-to-back. Although not very effective in reducing the back-

Tight Tag					
Cuts		$\mathrm{N_{TP}}$/Eff Signal $Z \to ee$	N_{TP} Background Di-jets	$W \to e\nu$	S/B
C0	PID + Acceptance	55540 ± 34	$(6 \pm 2) \cdot 10^3$ (5 events)	1828 ± 30	7.9 ± 2.6
C1	C0 + Trigger	$(99.81 \pm 0.01)\%$	-	1826 ± 30	-

Table 6.3: Number of signal and background pairs (N_{TP}) in the invariant mass window 80-100 GeV where the tag electron is required to pass the tight PID cut. Events are normalised to $100pb^{-1}$ of integrated luminosity. Errors are due to the MC statistics.

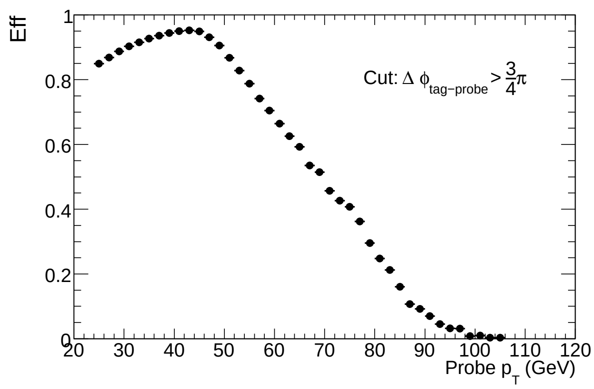

Figure 6.2: Efficiency of the $\Delta\phi$ separation cut between tag and probe electrons as a function of probe p_T.

ground, this cut ensures that tag and probe are well separated such that the efficiency is not biased by the presence of a nearby electron. However, it has another drawback: it strongly suppresses probes with high transverse momenta, as shown in Fig. 6.2, by forcing the Z boson to be produced at rest. This might not be a problem for inclusive analysis, but it is a significant bias for measurements including boosted electrons, such as W or Z bosons produced in association with jets. In such cases the $\Delta\phi$ cut should therefore be dropped.

The largest source of background is QCD jets. The W background contamination in the Z mass window (80-100 GeV) is about 3% before a cut on the opposite charge. The invariant mass distribution of signal and backgrounds for the full kinematic range is shown in Fig. 6.3: $W \to e\nu$ and QCD-jet backgrounds have different shapes, with $W \to e\nu$ peaking under the Z mass peak. Later in this chapter, it will be shown that the background shape strongly depends on the kinematic cuts applied and that also the QCD-jet invariant mass can peak below the Z

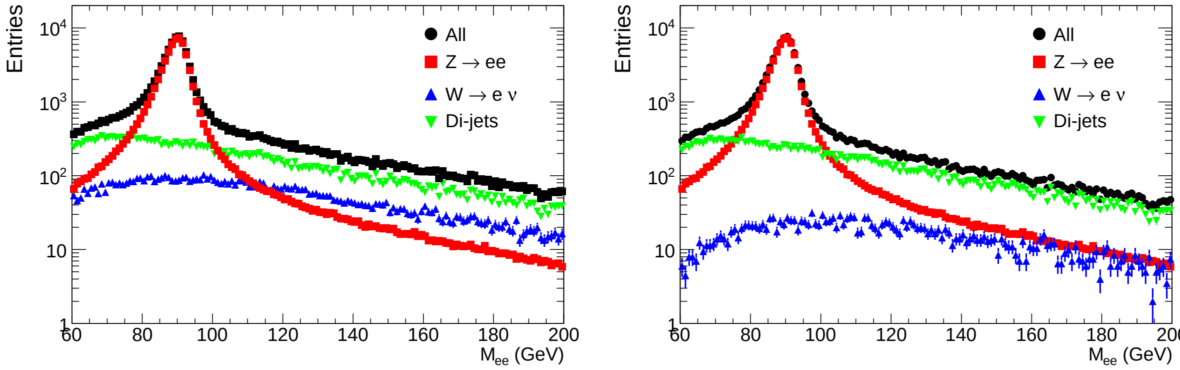

Figure 6.3: Invariant mass of signal and background before (left) and after (right) an upper cut on E_T^{miss} of 25 GeV.

mass for a specific phase space. A method for subtracting the background is the subject of Section 6.5. For the moment it is worth noticing that the $W \rightarrow e\nu$ background can be reduced to 0.8% applying an upper cut on transverse missing energy (see Fig.6.3 right) of 25 GeV, which does not introduce any bias on the efficiency.

6.4 Medium PID Efficiency in $Z \rightarrow ee$ Events

In this section, the medium PID efficiency extracted with the Tag-and-Probe method is investigated in signal events exclusively. The aim is to evaluate potential biases in the efficiency measurement introduced by the Tag-and-Probe selection and the $\Delta\phi$ or $\pm$ configurations in particular.

Probe Sample Composition in $Z \rightarrow ee$ Events

Before looking at the efficiencies, it is useful to consider the composition of the probe sample after the selection cuts described in Section 6.3. In fact, not all electrons in a $Z \rightarrow ee$ sample are genuine electrons from the Z boson decay. In MC simulations, reconstructed electrons can be classified according to the type and origin of the electron tracks [4]. We distinguish between:

[4] The truth matching procedure associates a reconstructed track to a truth track, for which the generated particle is known. Truth and reconstructed tracks are required to be close in ΔR space, such that $\Delta R < 0.2$.

Fraction of Electrons (%)					
Probe Origin	All e	All Probes		Medium Probes	
	$Z \to ee$	$\Delta\phi$ TP	$\pm$ TP	$\Delta\phi$ TP	$\pm$ TP
Non Defined	2.1 ± 0.01	0.8 ± 0.01	0.4 ± 0.01	0.1 ± 0.01	0.1 ± 0.01
Photon Conversion					
- Mother = Z	0.6 ± 0.01	0.4 ± 0.01	0.3 ± 0.01	< 0.001	< 0.001
- Mother = e from Z	4.3 ± 0.02	3.7 ± 0.03	2.1 ± 0.02	1.1 ± 0.01	0.7 ± 0.01
- Mother = other	0.4 ± 0.01	0.1 ± 0.01	0.1 ± 0.01	< 0.001	< 0.001
Z Boson	92.6 ± 0.01	94.8 ± 0.02	97.1 ± 0.02	98.9 ± 0.01	99.2 ± 0.01
Z Boson + e from Z	96.9 ± 0.01	98.5 ± 0.02	99.2 ± 0.02	99.9 ± 0.01	99.9 ± 0.01

Table 6.4: Percentage of electrons of a specified origin passing kinematic selections, in all $Z \to ee$ events (left) and those selected with the TP selection in the $\Delta\phi$ and opposite charge ($\pm$) configuration (middle). Shown are also fractions of probes that pass the medium identification (right). The last raw is the sum of two contributions: electrons originating from a Z boson decay, or from photon conversion where the photon parent particle (Mother) is an electron from Z boson decay (e from Z). Note that photons with Mother = Z are QED FSR photons, but are classified as having as parent a Z rather than an electron, to distinguish them from bremsstrahlung photons (emitted by an electron showering in the detector material).

non-defined origin (these are mainly fake electrons from hadrons), background electrons from photons converting in the detector material (where the photon's parent particle can be a Z boson [5], an electron from Z boson decay, or from other particles) and electrons from Z boson decay.

In Table 6.4 such a classification is shown for the probe electrons in inclusive $Z \to ee$ events and in two versions of the TP selection: $\Delta\phi$ separation and opposite charge ($\pm$). The purity of the probe sample improves as more cuts are applied, reaching about 97.1% for Z-boson-matched electrons (origin = Z Boson in Table 6.4) with opposite charge selection. Including also candidates from photon conversion due to interactions in the detector (origin = Z Boson + e from Z), 99.2% of the electrons in $Z \to ee$ events are accounted for. The difference between the $\Delta\phi$ and $\pm$ selections is due to the opposite charge requirement rejecting about 2.5% of pairs where the probe electron is reconstructed from photon conversion or has an undefined origin. Large differences in the probe composition are seen between barrel ($|\eta| < 1.37$) and

[5] These are actually photons from QED final state radiation (FSR). Although they are radiated by the final state electron, PHOTOS (the package which models QED FSR), records them as having a Z boson as parent.

end-cap ($|\eta| > 1.52$) regions, as a consequence of the different amount of material. In the end-caps, the fraction of electrons directly matched to the Z boson is 88.9% (with the $\Delta\phi$ configuration), almost 10% lower than in the barrel. When considering also electrons from photon conversion which have as mother electrons originating from Z bosons showering in the material of the detector, this fraction increases to 97.8% and the difference between barrel and end-caps reduces to about 1%.

After medium identification, the fraction of directly matched probes is about 99% and the difference between $\Delta\phi$ and $\pm$ configuration is below the 0.5% level: electrons that are not directly matched to a Z boson have lower quality and therefore tend to fail the PID cuts.

Global Efficiencies

Electrons with different origins also have a different probability of passing the identification cuts. Table 6.5 shows the medium identification efficiency of electrons classified according to their origin. Again both the $\Delta\phi$ and $\pm$ configuration are shown. For Z-boson matched candidates the efficiency is 94.8%, but it is significantly lower for electrons of different origins. The efficiency of each specified origin is independent of the TP selection cuts. However, since the composition of the probe sample varies with the selection cuts, the overall efficiency (considering all probes) is different in the $\Delta\phi$ and $\pm$ configuration. The difference is 1.2% globally, but strongly depends on the electron p_T and η, as will be shown in the next section. What can be measured with data strongly depends on whether or not combinatorial background caused by probes of different origins can be subtracted away with the QCD-jet background. Figure 6.4 shows the invariant mass distribution of the TP pairs for different origins of the probe electron. If the shape has a resonant structure, it will not be subtracted with the background subtraction procedure. This seems to be the case for electrons originating from photon conversions from the Z boson. Even probes of hadron origin, however, do not show a continuous falling distribution. Background subtraction will be treated in Section 6.5 and the question of combinatorial

Probe Origin	Medium Efficiency (%) $\Delta\phi$ TP	Medium Efficiency (%) $\pm$ TP
Non Defined	1.2 ± 0.1	1.3 ± 0.2
Photon Conversion		
- Mother = Z	35.8 ± 0.9	33.8 ± 1.2
- Mother = e from Z	50.7 ± 0.3	49.8 ± 0.5
- Mother = other	2.9 ± 0.6	2.8 ± 0.7
Z Boson	94.8 ± 0.03	94.8 ± 0.03
Z Boson + e from Z	93.1 ± 0.01	93.8 ± 0.01
All	92.0 ± 0.04	93.2 ± 0.01
All (Barrel)	93.1 ± 0.04	94.0 ± 0.01
All (End-caps)	89.5 ± 0.04	91.4 ± 0.01

Table 6.5: Medium identification efficiency determined from $Z \to ee$ events for specified origin of the probes. Shown separately are efficiencies from Tag-and-Probe applying the $\Delta\phi$ (left) and opposite charge (right) selection. The last rows also show the overall efficiency (considering all probes) in the barrel and end-cap regions separately.

background will be addressed. In any case, comparison between $\Delta\phi$ and $\pm$ TP measurements with data will give a handle on this combinatorial background and will provide important direct check on the absolute prediction from MC simulations.

Differential Efficiencies

The medium identification efficiency, measured with the Tag-and-Probe method on $Z \to ee$ events, is shown as a function of electron kinematic variables in Fig. 6.5. The top plots show the efficiency as a function of electron pseudo-rapidity (absolute value) for probes matched to a true electron from a Z boson (left) and all probes in the $Z \to ee$ sample (right), whereas the bottom plots show the efficiency dependence on electron ϕ and p_T. Three Tag-and-Probe configurations are considered: base cuts, $\Delta\phi$ and opposite charge ($\pm$).

Let us for the moment ignore the difference between configurations and the effect of truth matching to focus on the efficiency dependence on the detector region (η, ϕ) or electron kinematics (p_T). The variation in η and ϕ reflects the detector geometry. The efficiency is almost flat in ϕ (with a modulation due to variation in the thickness of the passive absorber), thanks to the

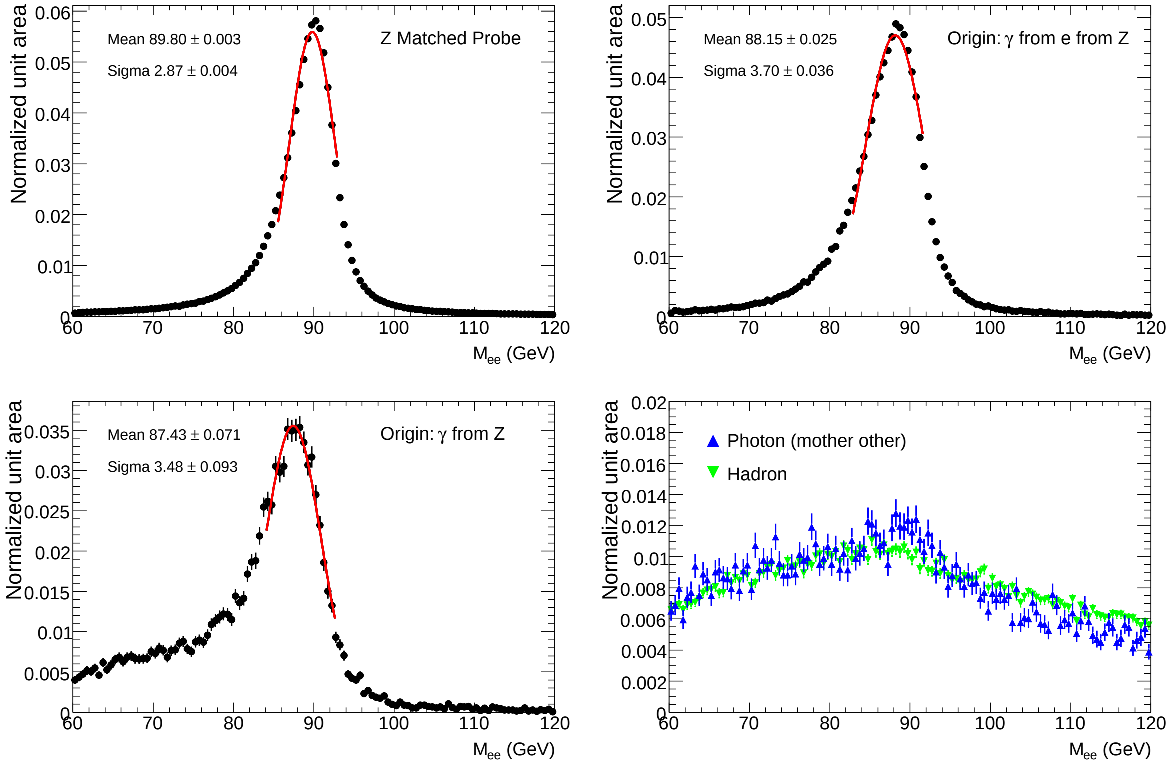

Figure 6.4: Invariant mass of tag-and-probe pairs (after $BaseTP$ selection) for probes of specified origins: directly matched to electrons from $Z \to ee$ (top left), matched to converted photons originating from electrons from $Z \to ee$ (top right) or radiated from the Z boson (bottom left) and different origin (bottom right), ie. hadrons or converted photons not coming from a Z or from an electron from Z.

the complete azimuthal symmetry without any cracks of the ATLAS EM calorimeter. The structure in η depends both on the calorimeter/inner detector geometry and on the amount of material in front of the calorimeter. In fact, PID cuts are optimised in bins of η taking into account the detector structure [6]. Low efficiencies correspond to either poorly instrumented regions of the detector or to large contributions of non-active material in the Inner Detector. The large Inner Detector material budget, which is necessary to ensure the high tracking performance requirements of ATLAS, is in fact critical for electron reconstruction. Interactions with the material mean that many electrons lose a significant fraction of their energy by bremsstrahlung before reaching the calorimeter and, similarly, most photons convert into electron-positron pairs. The correspondence between material in the Inner Detector and efficiency can be seen comparing Fig. 6.5 (top right) and Fig. 6.6 . Some interesting features of the η dependence are:

[6] The optimisation is done in the following bins: $|\eta|$ = 0.0, 0.6, 0.8, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37, 2.47.

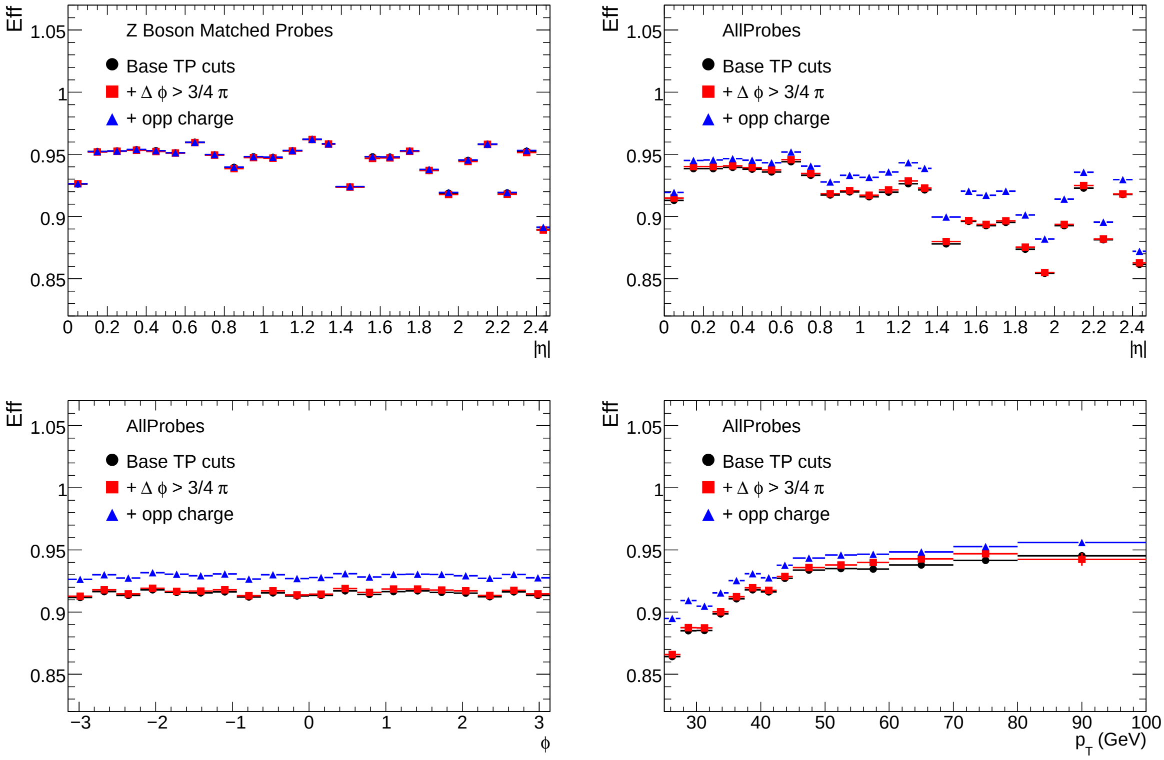

Figure 6.5: Medium electron identification efficiency determined from $Z \to ee$ events as a function of electron reconstructed η (top plots), ϕ and p_T (bottom plots). Shown is the efficiency extracted using the Tag-and-Probe method with different selection cuts. The top left plot is for probes directly matched to electrons from Z boson decay, whereas all other plots include all probes in the $Z \to ee$ sample.

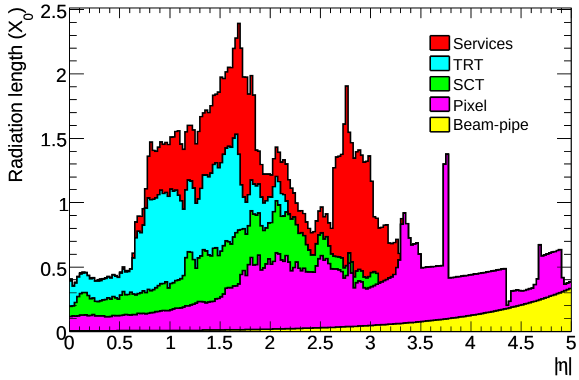

Figure 6.6: Material distribution in units of radiation length (X_0) at the exit of the ID envelope, including the services and thermal enclosure. The distribution is shown as a function of $|\eta|$ and averaged over ϕ [15].

- The drop at $|\eta| < 0.1$ corresponds to a small gap between the two identical half-barrels that make up the EM calorimeter barrel. This bin was not optimised in the reconstruction release used throughout this thesis, but is in more recent releases.

- The region between 0.6 and 0.8 corresponds to the end of the TRT barrel and the beginning of the TRT end-cap.
- The bin $|\eta|$ = 1.37-1.52 is the transition region between barrel and end-cap EM calorimeters. At this interface, there is a large fraction of service and structural materials, such as cooling and electrical connections for the Inner Detector.
- $|\eta| = 1.8$ corresponds to the end of the EM calorimeter pre-sampler.
- At $|\eta| = 2.0$ the TRT coverage ends.
- The drop in efficiency around $|\eta| = 2.2$ corresponds to another peak in the material before the calorimeter.

The identification efficiency also shows a strong dependence on the electron transverse momentum, increasing with electron p_T and reaching a plateau at about 50 GeV. The fact that the efficiency turn-on curve shows some steps at p_T = 30 and 40 GeV is due to the p_T bin optimisation of the identification cuts [7].

The difference in the efficiencies measured with different TP configurations is further evidence of what was discussed above. The effect of the selection cuts is a result of the modification in the probe sample composition, with the $\pm$ configuration rejecting many probes from photon conversion and fake electrons from hadronic jets. It is therefore not surprising that the most significant differences between configurations exist where the impact of detector material is largest: in the end-caps and at low p_T.

With data, it will be necessary to extract efficiencies in several dimensions, to be able to apply $Z \rightarrow ee$ efficiencies to other processes. Since efficiencies strongly depend on η and p_T, these will be the two first dimensions to be probed. More dimensions (such as calorimetry isolation or jet multiplicity) might be necessary for busy final states (such as $W \rightarrow e\nu$ +jets or $t\bar{t}$ cross-section measurements).

[7] For $p_T > 20$ GeV, PID cuts are optimised in 10 GeV bins in p_T.

6.5 Background Subtraction

In Section 6.3 the background contamination to the signal sample was discussed. Here the subtraction of the background from QCD-jets is treated, as it is by far the largest source of contamination into the $Z \rightarrow ee$ sample. Since efficiencies will have to be extracted differentially in bins of η and p_T, the background subtraction procedure was also tested in different bins. The signal to background ratios shown in Tables 6.3 and 6.2 are relative to the entire probe sample (ie. integrated in p_T and η). However, the background percentage strongly depends on the probe p_T and, to a lesser extent, η. Table 6.6 shows the signal to background ratio in bins of probe p_T and η for an overall $\frac{S}{B} = 7.9$, as expected for a tight tag (see first line of Table 6.3) [8]. The background contamination is much worse at low p_T and decreases with p_T up to 40-45 GeV, to then increase again above the Jacobian peak. This is a consequence of the different p_T spectra of QCD-jets (which is exponentially falling) with respect to electrons from Z boson decays. Furthermore, the kinematic cuts on the tag and probe electrons introduce a bias in the invariant mass (M_{ee}) distribution of the background. As a result, the background shape also depends on the p_T bin and cannot be described by a simple exponential function. Figure 6.7 shows the background M_{ee} for three different probe p_T bins: 25-30 GeV, 35-40 GeV and 40-45 GeV. The invariant mass peaks at some value M_{ee} which depends on the p_T and is shifted towards higher values as the cut on the probe electron p_T is increased. As a consequence, for high p_T, the background will have a maximum below the Z boson mass peak, preventing the possibility of subtracting the background using sidebands. An alternative technique, which involves a global fit of signal plus background, is proposed.

[8] Given the lack of Monte Carlo statistics for the di-jet sample, the background distribution in η and p_T was taken before applying identification cuts on the tag. The background has then been scaled to the required overall $\frac{S}{B}$. This method neglects any effect that a PID cut might have on the background distribution in p_T and η. The full selection criteria, including the PID requirement, was applied to the signal.

Signal to Background Ratio

p_T Bin	0.00 < $\|\eta\|$ < 0.80	0.80 < $\|\eta\|$ < 1.37	1.52 < $\|\eta\|$ < 2.00	2.00 < $\|\eta\|$ < 2.47
25-30 GeV	2.90 ± 0.08	3.02 ± 0.10	3.63 ± 0.15	3.90 ± 0.18
30-35 GeV	5.60 ± 0.16	5.85 ± 0.21	7.10 ± 0.35	7.28 ± 0.39
35-40 GeV	10.35 ± 0.33	10.71 ± 0.43	10.68 ± 0.54	10.97 ± 0.65
40-45 GeV	15.82 ± 0.57	16.20 ± 0.74	16.17 ± 0.94	15.68 ± 1.04
$>$ 45 GeV	8.49 ± 0.21	7.47 ± 0.22	8.08 ± 0.32	7.94 ± 0.36

Table 6.6: Signal over background ratio for the probe collection (with a tight tag) in bins of η and p_T in the invariant mass range 80 GeV $< M_{ee} < 100$ GeV. The overall ratio, integrated in η and p_T, is $\frac{S}{B} = 7.9$. The statistical error is due to the available Monte Carlo statistics.

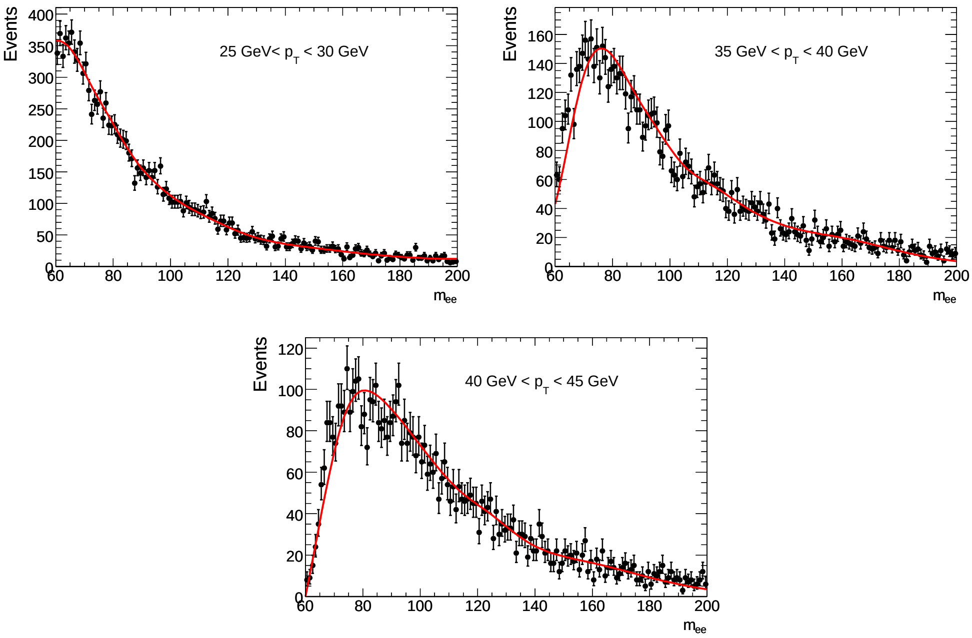

Figure 6.7: Invariant mass of TP pairs in the di-jets sample for three different p_T bins. To increase the statistics no PID cut is applied on the tag.

Global Fit of Signal and Background

Background to $Z \rightarrow ee$ can be subtracted by fitting simultaneously signal and background invariant mass distributions in the range 60-200 GeV. The global probability density function (Pdf) is the sum of two terms:

$$Pdf(M_{ee}) = N_{Signal} \cdot Pdf(M_{ee})_{Signal} + N_{Bkg} \cdot Pdf(M_{ee})_{Bkg}. \quad (6.7)$$

Here N_{Signal} and N_{Bkg} are the number of signal and background events respectively and they are constrained such that their sum is equal to the total number of events in the data sample : $N_{Signal} + N_{Bkg} = N_{Total}$.

The $Pdf s$ are:

- $Pdf(M_{ee})_{Bkg}$ is the background Pdf. Since the background M_{ee} distribution cannot be easily parametrised and its shape is largely uncertain, it will have to be extracted from data using control samples. Cut reversal of calorimetric or track cuts can be used, similar to the extraction of a QCD-jet background template for a $W \rightarrow e\nu$ analysis (see Chapter 5). For the purpose of this analysis, however, in the absence of data and with limited MC statistics, the $Pdf(M_{ee})_{Bkg}$ is taken from MC using relaxed PID cuts. Therefore it is assumed that the background shape is perfectly known. This is a purely academic exercise and, once sufficient data from the LHC is collected, plenty of QCD-jets events will be available for the extraction of the background template.

- $Pdf(M_{ee})_{Signal}$ is the signal Pdf. It is given by the convolution between a true signal Pdf and a resolution function:

$$Pdf(M_{ee})_{Signal} = \left(BW(M_Z, \Gamma_Z; M_{ee}) \cdot \left(\frac{M_Z}{M_{ee}}\right)^{\beta} \right) \otimes CB(\alpha, n, \mu, \sigma; M_{ee}) \quad (6.8)$$

  where the true signal Pdf is the product of a Breit Wigner with a parton luminosity function and the resolution function is of Crystal Ball type [9]. The only fixed parameter is the Z width (Γ_Z) and the other are free. The parton luminosity factor, which parametrises the effect of the parton density functions (PDFs) on the Z lineshape [75] affecting mainly the tails, can be switched off by setting $\beta = 0$.

Figure 6.8 shows fits to the invariant mass resolution and to the generated and reconstructed invariant mass using the functions described above. The reader should note that for the fits to

[9] A Crystal Ball function consists of a Gaussian core and a power-law tail which describes the energy loss due to bremsstrahlung radiation. The parameters μ and σ are the Gaussian mean and width, whereas α and n govern the power-law tail.

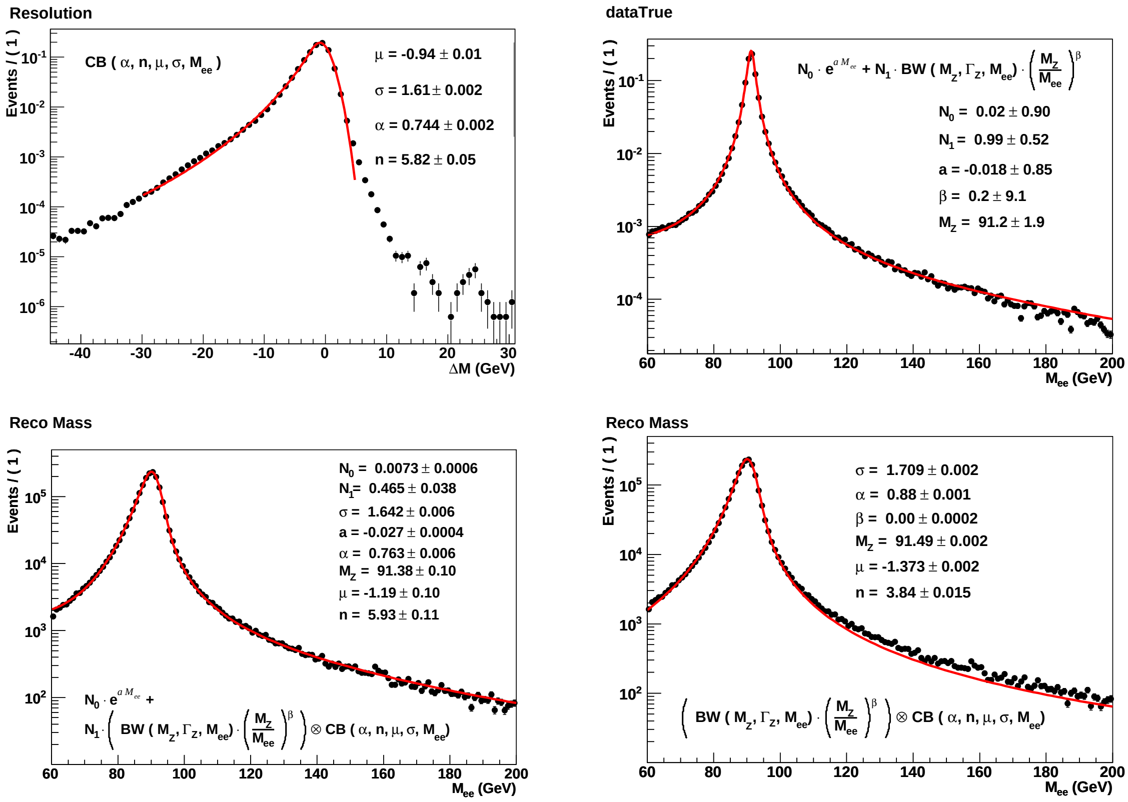

Figure 6.8: (Top Left) Fit to the M_{ee} resolution ($\Delta M = M_{Reco} - M_{Gen}$) with a Crystal Ball function. (Top Right) Fit to the generated $Z \rightarrow ee$ invariant mass using a Breit Wigner multiplied by a parton luminosity factor plus an exponential function (to account for the Drell-Yan continuum). (Bottom) Fit to the reconstructed invariant mass using a convolution of a Breit Wigner multiplied by a parton luminosity factor with Crystal Ball function and in addition with (left) or without (right) an exponential function.

the invariant mass, an exponential decay function was added to the signal Pdf. This is needed to account for the Drell-Yan $pp \rightarrow Z/\gamma* \rightarrow ee$ process which is included in the PYTHIA sample used for the signal. However it is not included in the signal Pdf, since it would introduce a component extremely similar to the background Pdf, making it very difficult to distinguish between signal and background. Therefore it is assumed that the Drell-Yan spectrum can be subtracted away together with the QCD-jet background, as will be discussed later in this section.

Toy MC to Generate QCD-jet Events

The available MC statistics for QCD-jets is too low to test the fitting procedure in p_T and η bins. To overcome this problem, a toy MC was used. The background invariant mass shape was taken from the MC sample with relaxed cuts and the rate was estimated from the probability for a jet to fake an electron. The full procedure is summarised in the following:

- The M_{ee} shape for the background was extracted by applying only the TP acceptance cuts (no PID and no trigger requirement were applied). The M_{ee} was then fitted for each p_T bin to a non-parametric Pdf using the RooFit package (as shown in Fig. 6.7). The shape dependence on η was ignored being negligible by comparison with the p_T dependence.
- The non-parametric $Pdfs$ were used to generate a high number of di-jet events in the range 60-200 GeV. The luminosity assumed was equivalent to the full statistics of the $Z \rightarrow ee$ MC PYTHIA sample (approximately 5 fb^{-1}). Each p_T-η bin was populated with the number of QCD-jet events required to obtain a given signal to background ratio. The sample obtained represents the QCD-jet background to the electron probe collection (denominator N_1 in eqn. (6.3)). The background contamination was assumed to be:
  - $\frac{S}{B} = 7.9$: this is the expected signal to background ratio for a tight tag, without any additional cut (like $\Delta\phi$ or opposite charge), see first line in Table 6.3.
  - A fraction of background which is two or three times higher was used for studies of systematic uncertainties. Such a background level could be expected using a medium rather than a tight tag.
- Using the MC QCD-jet sample, the probability for a jet to fake a medium PID electron (the so called fake rate) was extracted as a function of p_T. The number of jet probes satisfying the medium PID cuts was estimated using these fake rates, thus obtaining the QCD-jet background to the medium probe collection (numerator N_2 in eqn. (6.3)).

Table 6.7 shows the total number of signal events and generated background events in the 80-100 GeV mass window for the probe collection and the medium probe collection with the default assumption on $\frac{S}{B}$.

	N_{S+B}	N_S	S/B
Probes	$300.4 \cdot 10^4$	$266.6 \cdot 10^4$	7.9
Medium Probes	$245.4 \cdot 10^4$	$245.0 \cdot 10^4$	629.2

Table 6.7: Total number of signal and background events and signal over background ratio for the probe collection and the medium probe collection, in the invariant mass range $80\ \text{GeV} < M_{ee} < 100\ \text{GeV}$, used to test the background subtraction procedure.

Fit Procedure

For the extraction of the efficiency, two fits need to be performed: one for the probe collection (denominator) and one for the medium probe collection (numerator). An extended maximum likelihood formalisms (MLL) on a binned dataset is used. The efficiency is then computed taking the ratio of the estimated number of signal events under the Z peak in the two collections. For binned efficiencies, in one or more dimensions, this procedure has to be repeated for each bin. We take as an example two p_T bins (25-30 GeV and 35-40 GeV), integrated over η. The fits are shown in Fig. 6.9-6.10. Although the fits look rather good by eye, the pull distributions (the difference between the bin content and the expected bin content, as predicted by the fit function, divided by the statistical error in that bin) highlight some problems, which become more serious for higher p_T. The $\chi^2/ndof$ and the result of a Kolmogorov-Smirnov test (results shown on the pull plots) are also poor. The Z peak region is not well described and the fit systematically miss-estimates the data in the range 70-100 GeV. For high p_T bins, the low mass region (< 70 GeV) is also poorly described. The high mass region, however, where the background is most significant, is described rather well.

One problem identified was a number of correlated free parameters, yielding similar effects, and hence confusing the fit. The Crystal Ball resolution function, the parton luminosity factor and (in low p_T bins) the background Pdf all cause a higher number of entries in the low mass region with respect to the high mass region, resulting in the possibility that one factor absorbs the other. We may understand this by considering the bottom plots of Fig. 6.8, showing the fits to the signal $Z \to ee$ sample only: in this case there is a small amount of background due to the

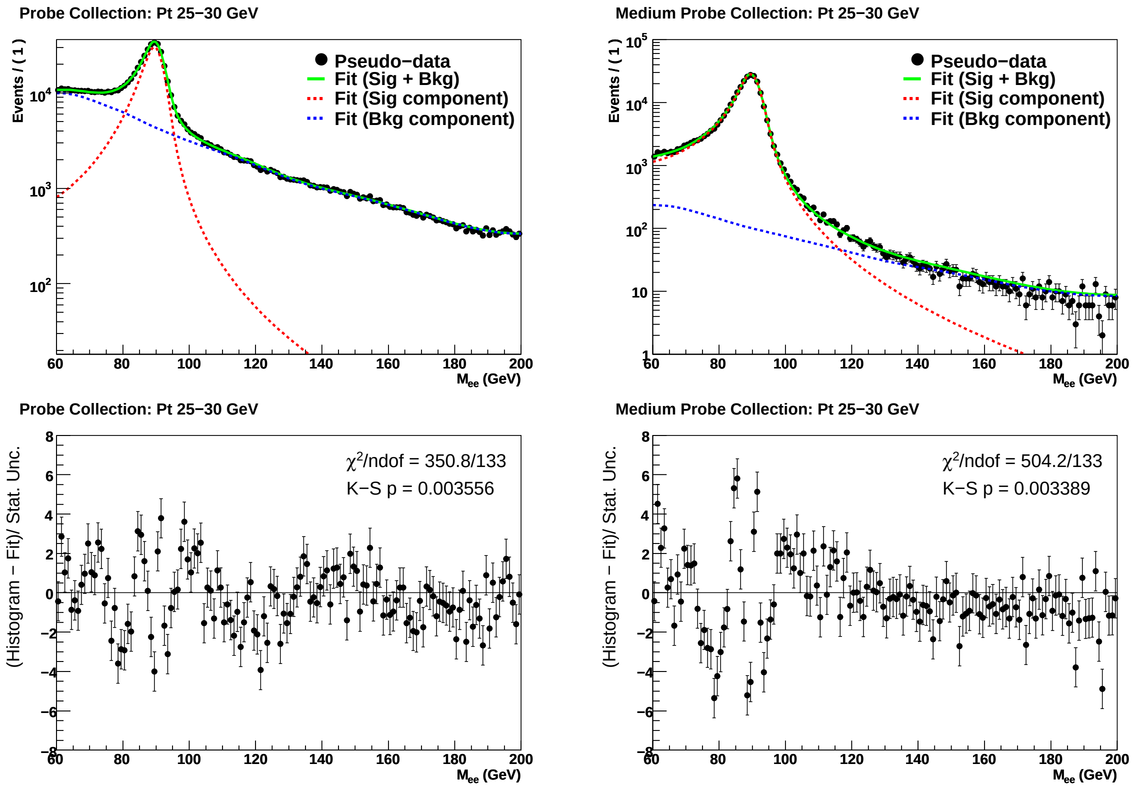

Figure 6.9: Extended MLL fit to the probe collection (top left) and medium probe collection (top right), for the probe p_T bin 25-30 GeV. Markers show the input data. The result of the fit is shown by the continuous (total $Pdf = Pdf(M_{ee})_{Signal} + Pdf(M_{ee})_{Bkg}$) and dotted lines (signal and background components). The estimated number of signal and background events and the resulting efficiency are given in Table 6.8. The bottom plots show the pull distributions of the fits to the probe (left) and medium probe collections (right). The χ^2 over number of degrees of freedom ($ndof$) and the probability obtained from a Kolmogorov-Smirnov test (K-S p) are also shown.

combinatorials and the Drell-Yan continuum. This background is correctly taken into account by an exponential function (left) [10], but when the exponential term is removed the Crystal Ball parameter n (which describes the power-law tail) absorbs part of the background contribution into the signal. This problem was already observed in [71], where it was also noted that constraining some parameters (for example setting $\beta = 0$) does not improve the goodness of the fit. It was therefore decided to keep the Crystal Ball parameter and the parton luminosity factor as free and to consider variations in the number of free parameters as systematic uncertainties. Two other competing parameters are Γ_Z and the σ of the Crystal Ball resolution function. If the Z width is let free, the fitted value is unphysically large since it also accounts for the resolution.

[10] In this case an exponential decay function describes the background because the full $Z \rightarrow ee$ sample is used and is not binned in probe p_T.

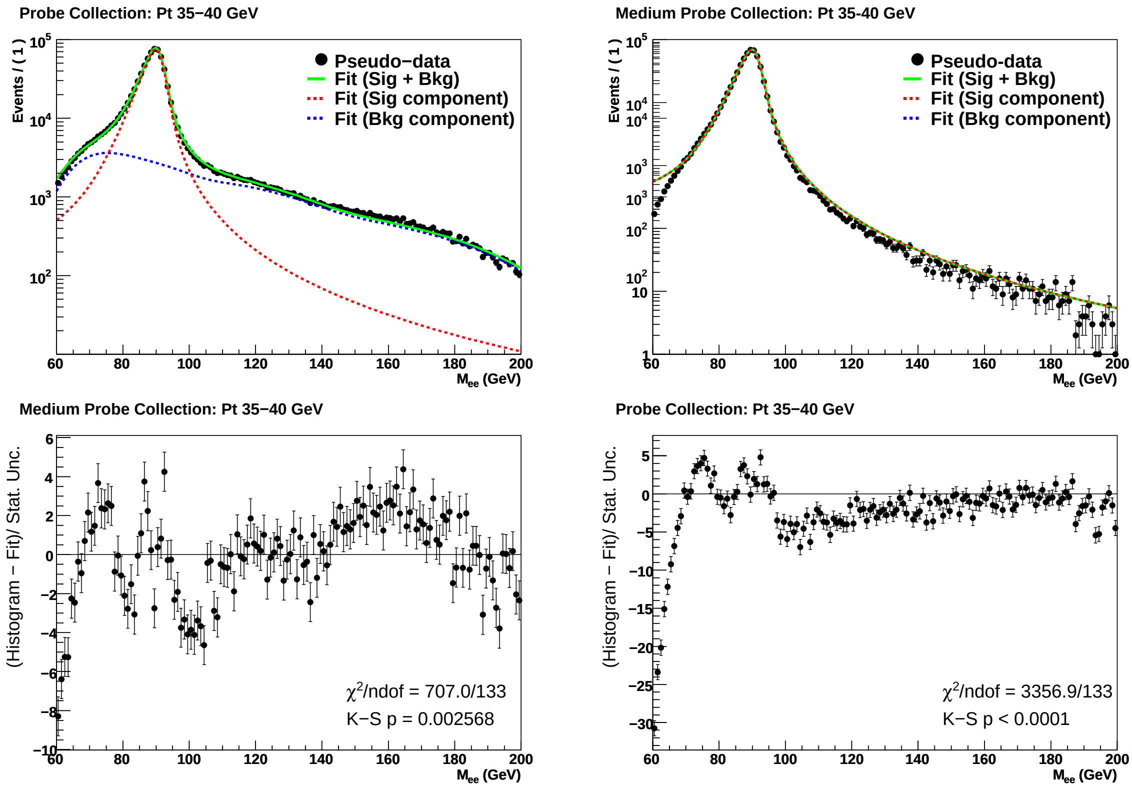

Figure 6.10: Extended MLL fit to probe collection (top left) and medium probe collection (top right), for the probe p_T bin 35-40 GeV. Markers show the input data. The result of the fit is shown by the continuous (total $Pdf = Pdf(M_{ee})_{Signal} + Pdf(M_{ee})_{Bkg}$) and dotted lines (signal and background components). Note that in the right hand plot (medium probe collection) the fit predicts essentially zero background and the background component is therefore not visible. The estimated number of signal and background events and the resulting efficiency are given in Table 6.9. The bottom plots shows the pull distributions of the fits to the probe (left) and medium probe collections (right). The χ^2 over number of degrees of freedom ($ndof$) and the probability obtained from a Kolmogorov-Smirnov test (K-S p) are also shown.

It was considered unnecessary to introduce another free parameter and thus Γ_Z was fixed to the PDG value. The effect of letting it float was considered as a source of systematics.

Another problem was observed in the bias of the $Z \to ee$ invariant mass distribution introduced by a cut on the probe p_T. When looking at different probe p_T ranges, poor description of signal in the tails becomes an issue, particularly for the high p_T bins. Figure 6.11 shows how the generated and reconstructed invariant mass shapes depend on the probe p_T bin even for probes which are matched to a Z boson. The tail in the low (high) mass region for low (high) p_T bins is due to the photon term of the Drell-Yan ($q\bar{q} \to \gamma^*/Z \to ee$) cross-section. The $\gamma^* \to ee$ process forms a continuum background which affects the invariant mass distribution in different

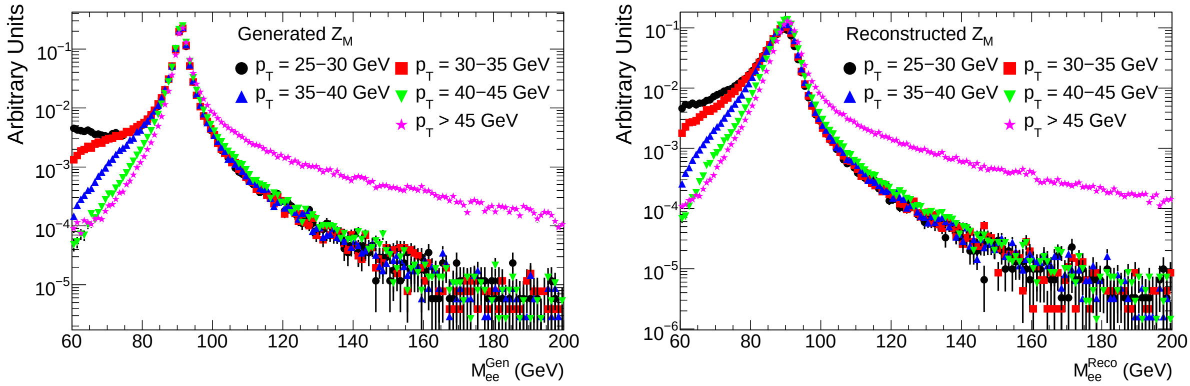

Figure 6.11: Generated (left) and reconstructed (right) invariant mass for probes matched to a Z boson in different p_T bins.

regions, depending on the p_T range. For the TP method, it does not matter whether this photon continuum is considered to be signal or background, since it will not bias the final efficiency. In fact, a tail in the high mass, like the one for $p_T > 45$ GeV, will get absorbed into the background which also has a falling exponential shape. It is however more difficult to account for the variations in low mass region which introduce further complexity to an already problematic fit. This explains, for example, the bad description of the 60-70 GeV range in Fig. 6.10. The best way to account for the variation of the signal Pdf with the p_T bin is to use a template, rather than a parametric Pdf, which can be taken from MC. This approach has the drawback of relying on simulations for the Z mass resolution and therefore is not foreseen as the default procedure, but is used to assess systematic uncertainties.

Despite such problems, this fit procedure was considered suitable for the efficiency measurement [11], since it yields stable results (see also the section on systematic uncertainties). In fact, what really matters is the correct prediction of the background and not a precise description of the signal peak. The number of $Z \to ee$ events in the range 80-100 GeV is then extracted as the number of candidate events minus the number of predicted background events [12]. The signal Pdf is used exclusively to find the correct background normalisation.

[11] For a Z cross-section measurement, however, this fit would not be satisfactory.

[12] The procedure of extracting the number of signal events by taking the signal Pdf integral in the range 80-100 GeV was also tested and found to give less stable results.

Bin p_T = 25-30 GeV

	N_B^{True}	N_B^{Fit}	N_S^{True}	N_S^{Fit}	$\frac{N_S^{Fit}-N_S^{True}}{N_S^{True}}(\%)$
Probes	81492	89198 ± 192	260582	252876 ± 615	-3.05 ± 0.32
Medium Probes	1091	2061 ± 73	230318	229348 ± 487	-0.42 ± 0.30
			$\epsilon^{True}(\%)$	$\epsilon^{Fit}(\%)$	$\Delta\epsilon(\%)$
Efficiency			88.39 ± 0.06	90.70 ± 0.16	2.31 ± 0.17

Table 6.8: Example of efficiency extraction in the p_T bin 25-30 GeV. Shown is the number of true and predicted signal (N_S) and background (N_B) events in the Z mass window 80-100 GeV, for both probe and medium probe collections, and the efficiency. Only statistical errors (including the errors from the fit) are quoted.

Results and Comparison to MC Truth

The p_T bins 25-30 GeV and 35-40 GeV are taken again as an example. The results of the fits shown in Fig. 6.9 and 6.10 are reported in Tables 6.8 and 6.9 respectively. Shown are the input number of $Z \rightarrow ee$ and QCD-jet events (in the Z mass window 80-100 GeV) and the predictions from the fit, together with the resulting efficiencies. The deviation from the MC truth is also reported. It should be noted that the reference numbers for the MC truth consider all probes in the $Z \rightarrow ee$ sample, without any distinction regarding the probe origin. The most significant point to notice is the overestimation of the background in the 25-30 GeV bin, leading to an efficiency which is more than 2% higher than in the MC truth. This is not a real problem, but rather a (desired) feature of the background subtraction procedure. If one recalls that the $Z \rightarrow ee$ sample contains poor quality probes not matched to electrons from Z bosons (such as probes from hadrons or photon conversions not coming from a Z, referred to as combinatorial background), one can deduce that the reason for the difference is that such combinatorial background is subtracted together with the QCD-jet background. A further hint to this hypothesis is the fact that after medium identification and in the higher p_T bin 35-40 GeV, the deviation from the truth is much smaller (the combinatorial background is much more significant at low p_T).

The hypothesis was tested using as input signal data the full $Z \rightarrow ee$ sample or only TP

Bin p_T = 35-40 GeV

	N_B^{True}	N_B^{Fit}	N_S^{True}	N_S^{Fit}	$\frac{N_S^{Fit}-N_S^{True}}{N_S^{True}}(\%)$
Probes	56548	54240 ± 243	581324	583632 ± 835	0.39 ± 0.19
Medium Probes	595	0.1 ± 1.7	534632	535227 ± 732	0.11 ± 0.19
			$\epsilon^{True}(\%)$	$\epsilon^{Fit}(\%)$	$\Delta\epsilon(\%)$
Efficiency			91.97 ± 0.04	91.71 ± 0.08	0.26 ± 0.09

Table 6.9: Example of efficiency extraction in the p_T bin 35-40 GeV. See Table 6.8 for more details.

pairs with a probe matched to a Z boson (origin = Z boson + e from Z, as in the last line of Table 6.4). Figure 6.12 shows the fit to the probe collection (before medium identification) in the two cases, for the p_T bin 25-30 GeV. In this bin the fraction of non-matched probes in the invariant mass range 80-100 GeV is 3.7%. When all probes are considered in the $Z \to ee$ sample, the fit underestimates the signal in the high tail, since the hadron/photon conversion probe component is absorbed into the background. The deviation from the MC truth (considering all probes) is 3.05% (see last column in Table 6.8). In contrast, the signal Pdf describes very well the Z matched probes and the deviation from the truth ΔN is smaller than 0.1%. The difference $(\Delta N_{\text{All Probes}} - \Delta N_{\text{Matched Probes}}) \approx 3\%$ has to be compared with the fraction of non-matched probes of 3.7 %. It can be concluded that combinatorial background gets subtracted with the QCD-jet background within 1% precision. This is not obvious, since the invariant mass spectrum of these non-matched probes peaks below the Z peak and is not described by the signal Pdf or by the background Pdf (see Fig. 6.13). However, the fraction of non-matched probes in the $Z \to ee$ sample dominates in the high M_{ee} region. This creates a high tail in the M_{ee} signal distribution which is absorbed into the background Pdf by the fit. It is interesting to notice that the $W \to e\nu$ background has a similar shape, as shown in Fig. 6.13 and it is therefore expected that it will also be subtracted away.

A natural question would therefore be: what should be taken as truth MC reference for the Tag-and-Probe measurement? There is no unique answer, since the definition of the truth is somewhat arbitrary. One could arguably consider as reference all probes in the $Z \to ee$

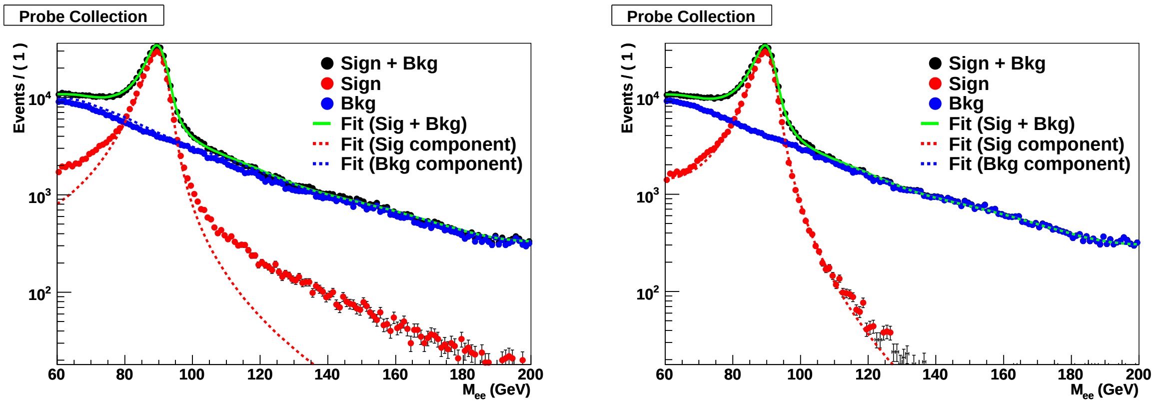

Figure 6.12: (Left) Fit of signal plus background where all probes have been included in the signal, for probes with p_T between 25 and 30 GeV. The data are represented by the markers and the predicted signal and background by the line. (Right) Same, but the input signal consists of probes matched to a Z boson only (Z boson + e from Z).

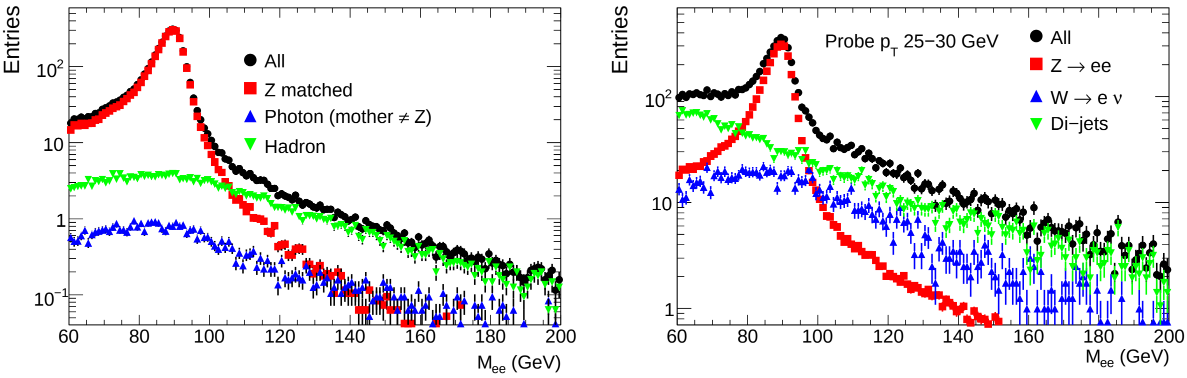

Figure 6.13: (Left) invariant mass distribution of tag-and-probe pairs in the signal, for the p_T bin 25-30 GeV. Shown separately are probes with different origins. (Right) Background contamination from di-jets and $W \rightarrow e\nu$ in the same bin.

sample, or matched probes, where truth matching can be defined in different ways. The MC truth definition is only useful as a reference to which the data measurement can be compared. In the following discussion, in addition to the truth track classification used previously in this chapter, a geometrical matching in ΔR is adopted. ΔR is defined as the geometrical distance between the reconstructed electron and the closest true electron from a Z boson decay. The reason for introducing a geometrical matching is that this criteria will be used as reference in the following chapter, in the context of the W + jets cross-section measurement. The value for the ΔR cut was set to 0.2, for reasons illustrated in Fig. 6.14. The top plot shows the normalised distribution in ΔR for probes of different origins, according to the classification used in Table 6.4. The class of hadron origin probes is a sub-set of the non-defined class.

Clearly, Z boson matched probes are close in ΔR to the true electron. Regarding electrons from γ conversion, only those reconstructed electrons which are within $\Delta R < 0.2$ from the true electron show a clear resonance around the Z mass value. In all other cases the Z peak is shifted and smeared such that it would be no longer described by the signal Pdf [13] (see bottom left plot). Similar conclusions can be drawn considering all probes (bottom right plot), independently of their origin: for $\Delta R > 0.2$ the Z resonance is no longer visible. This definition for truth matching is therefore a pragmatic approach to classify probes according to what can be clearly discriminated against the background.

The TP results (after the background subtraction procedure) were compared to the TP efficiencies as extracted from $Z \rightarrow ee$ events only, with or without truth matching. Figure 6.15 shows such a comparison for the efficiency as a function of electron p_T. Three MC truth definitions are shown: all probes in $Z \rightarrow ee$, probes geometrically matched to a true electron from Z boson decay (ΔR matched probes) and probes directly matched to an electron from a Z boson using the track classification (Z Matched Probes). The efficiency of the Z matched probes cannot be recovered, even after background subtraction. The reason is that this excludes electrons which have showered in the detector before reaching the calorimeter and which therefore have a lower efficiency. It was shown in Fig. 6.4 that such electrons exhibit a clear resonance around the Z mass value and hence cannot be subtracted away. In contrast, disagreement between pseudo-data and *all* $Z \rightarrow ee$ *probes* in the low p_T bins is evidence of the subtraction of combinatorial background. The extracted efficiency is, instead, in much better agreement with ΔR matched probes, which include a large fraction of electrons showering in the detector material, but excludes combinatorial background from hadronic jets faking an electron or from photon conversion (where the photon is not originating from the Z).

[13] Clearly, it would also not be described by the background Pdf.

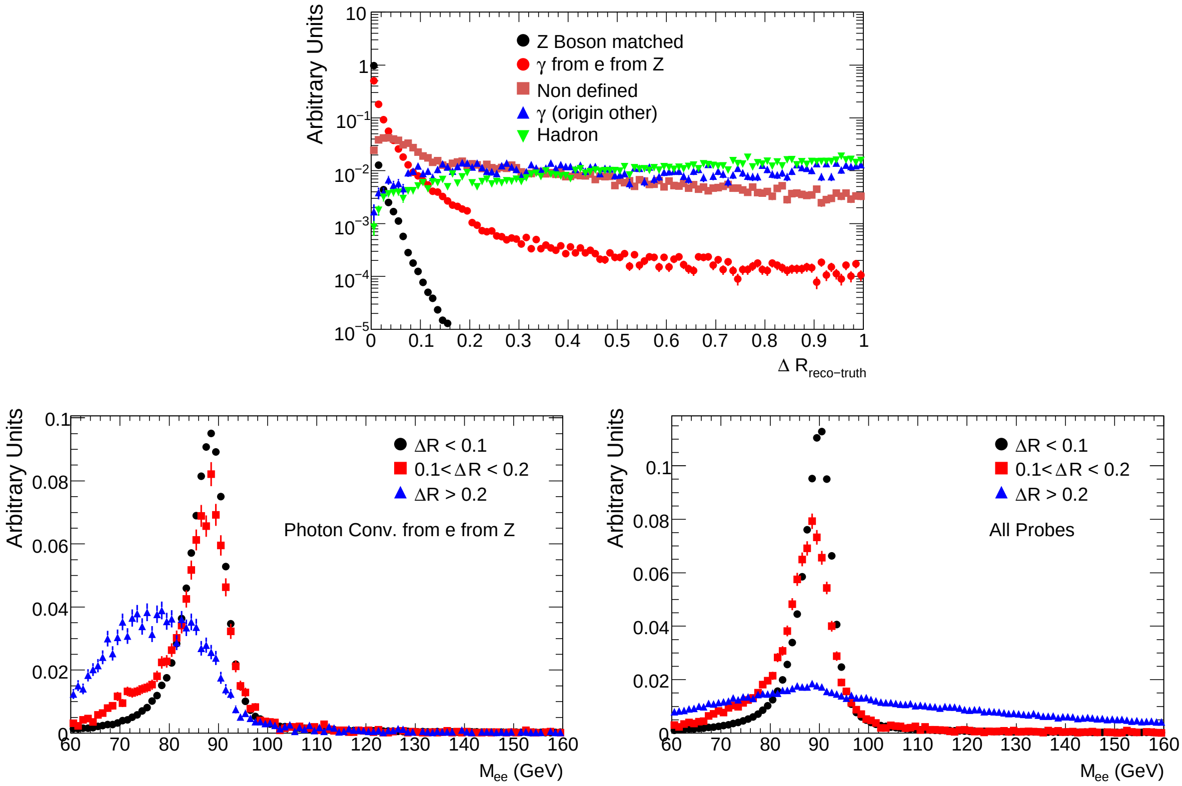

Figure 6.14: (Top) ΔR between true electron from Z boson decays and reconstructed electrons of different origins. (Bottom) invariant mass distribution of tag and probe electrons, for different values of ΔR between probe and true electron. Considered separately are probes from photon conversion, having as mother an electron from a Z boson decay (left), and all probes (right).

Systematic Uncertainties From Background Subtraction

Systematic uncertainties on the background subtraction procedure were evaluated by varying the following (see Table 6.10):

- Signal Pdf: the number of free parameters in the fit was varied. These included the Z width Γ_Z, the parton luminosity factor β and the Crystal Ball function parameter. Table 6.10 shows the default values and the variation. The Crystal Ball parameters were constrained according to the values obtained from a fit to the Z mass resolution plus and minus one sigma. The signal Pdf was also varied taking a template (from an independent MC sample) rather than using a parametrised function.
- Fit ranges: the default range used for the fit (60-200 GeV) was varied between 70-200, 60-180 and 60-160 GeV.

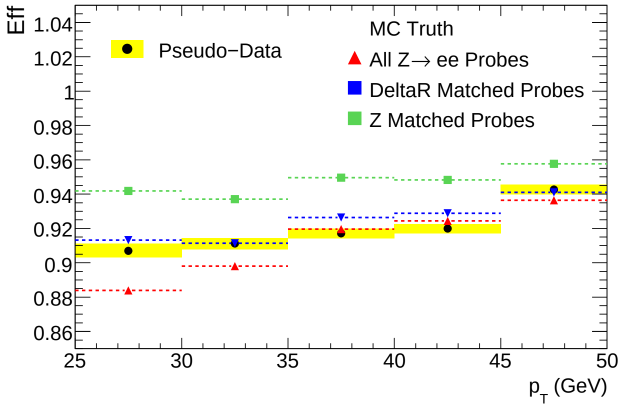

Figure 6.15: Medium identification efficiency extracted with the Tag-and-Probe method after background subtraction (including QCD-jet background). The yellow band represents the statistical error. Extracted efficiencies are compared to the MC truth with different definitions: all probes in the $Z \to ee$ sample, only probes directly matched to a Z boson (origin = Z boson) and only probes matched in ΔR to a true electron from a Z boson.

Label	Explanation	Default	Variation
S1 a	Signal Pdf	Parametrised Pdf	Template Histogram
S1 b		Γ_Z = 2.5 GeV	Γ_Z free
S1 c		β free	$\beta = 0$
S1 d		free CB parameters	Constrain CB
S2 a	Fit Range	60-200 GeV	70-200 GeV
S2 b			60-180 GeV
S2 c			60-160 GeV
S3 a	Bkg Level	S/B = 7.9 (629.2) in N_1 (N_2)	Bkg x 2
S3 b			Bkg x 3

Table 6.10: Sources of systematic uncertainties considered on the global fit of signal plus background.

- Background level: the amount of QCD-jet background was increased by a factor of 2 and 3.

The average variations in the measured efficiency from each source (signal Pdf, fit range, background level) are shown in Table 6.11 for each p_T bin. Considering these sources as independent, the overall systematics are given by their quadrature sum. The overall systematic uncertainty significantly decreases with the electron p_T (is ~0.5% for $p_T > 35$ GeV). This is to be expected since the background contamination is lower as p_T increases. Although the systematics in the low p_T bins are fairly large, this will have a limited effect on, for example, a W or Z cross-section measurement, where the bulk of the statistics are above 35 GeV.

Source	Probe p_T Bin 25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV
Signal Pdf (S1)	0.57	0.48	0.16	0.27	0.17
Fit Range (S2)	0.79	0.31	0.14	0.23	0.43
Bkg Level (S3)	0.92	0.92	0.47	0.23	0.04
Total	1.34	1.08	0.52	0.42	0.46

Table 6.11: Systematic uncertainties (in $\%$) on the global fit of signal plus background evaluated for different probe p_T bins.

6.6 Prospects for Early Data Measurements

As explained earlier in this chapter, electron efficiencies should be extracted as a function of the kinematic variables they most strongly depend upon. Most importantly, the TP method should be applied to measure efficiencies versus electron η and p_T (and possibly other variables). The number of bins obviously depends on the statistics available. Ideally, efficiencies should be extracted in a fine grid of size 18 x 6 ($\eta \times p_T$), with η boundaries corresponding to the bins used for the optimisation of the PID cuts (0.0, 0.6, 0.8, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37, 2.47) and p_T bins of a standard width of 5 GeV, up to 50 GeV when the efficiency reaches a plateau (25, 30, 35, 40, 45, 50, $>$ 50 GeV). Positive ($|\eta| > 0$) and negative ($|\eta| < 0$) regions should be considered separately, since it cannot be assumed a priori that efficiency will be perfectly symmetric in η. Differences in the amount of material between positive and negative η would in fact result in different efficiencies.

However, measuring efficiencies in such a fine grid will only be possible with $\sim$ 1 fb^{-1} of data. With less statistics, for example 100 pb^{-1} of data, it will be necessary to use a coarser grid, considering for example only end-cap and barrel η regions (4 x 6 bins), or folding together positive and negative η regions (2x 6 bins).

Table 6.12 lists the number of expected $Z \to ee$ TP pairs in an η-p_T grid of size 4 x 6, expected with 100 pb^{-1} of data at 10 TeV centre of mass energy. The corresponding statistical error on the efficiency can be determined from eqn. (6.6): to achieve 1% statistical error ap-

η/ p_T	25-30 GeV	30-35 GeV	35-40 GeV	40-45 GeV	45-50 GeV	> 50 GeV
-2.47 to -1.52	921 ± 4	1301 ± 5	1716 ± 5	1963 ± 6	1194 ± 4	931 ± 4
-1.37 to 0.00	1634 ± 5	2719 ± 7	3998 ± 9	4704 ± 9	2813 ± 7	2418 ± 7
0.00 to 1.37	1634 ± 5	2724 ± 7	3986 ± 9	4706 ± 9	2817 ± 7	2420 ± 7
1.52 to 2.47	920 ± 4	1296 ± 5	1712 ± 5	1974 ± 6	1174 ± 4	929 ± 4

Table 6.12: Number of TP pairs expected in a coarse η-p_T grid with 100 pb^{-1} of data using the $BaseTP$ selection (no $\Delta\phi$ nor opposite charge requirement) and before the medium identification requirement on the probe. The crack region between barrel and end-cap calorimeter (1.37-1.52) has been excluded.

proximately 1300 probes per bin are required [14]. Another limitation to the size of the bins is imposed by the fit for background subtraction: at least a few hundred events per bin are needed to obtain a stable fit. Furthermore, with 7 TeV centre of mass energy, the number of events per bin will be reduced by an additional $\sim$ 30%. From the numbers of Table 6.12, it seems therefore feasible to measure efficiencies in a 4x6 grid with 100 pb^{-1} of data, within a statistical error of 1% in the barrel region and 1.5% in the end-cap region, even with 7 TeV centre of mass energy. With very early data, however, only global efficiencies can be measured. For example, with 10 pb^{-1} of data at 7 TeV, approximately 3400 to 3800 pairs are expected (for $p_T > 25$ GeV and $|\eta| < 2.47$), depending on the selection cuts used.

A measurement of global efficiencies on a coarse grid will nevertheless be very valuable for testing the MC simulations and will even allow to gain sensitivity on the combinatorial background. In Section 6.5 it was shown that the TP method cannot distinguish between high quality probes directly coming from a Z boson decay and lower quality probes from the shower of Z boson electrons in the calorimeter. Since the fraction of the latter in the probe sample depends on the TP selection cuts (see Table 6.4), it is expected that efficiency extracted from data will also depend on the TP configuration. Cross-checks between MC expectations and data-driven measurements will therefore give a handle on these lower quality probes. Some interesting measurements are:

[14] This simple estimate assumes an efficiency of 85%, which is lower than the global efficiency of 92-93%. This value was chosen to account for the worst case scenario of bins in the end-caps and in the low p_T region which have an efficiency as low as 85%.

- $\Delta\epsilon = \epsilon_{\Delta\phi} - \epsilon_{\pm}$: considering only Z matched probes (Z Boson + e from Z in Table 6.5), 0.7% difference is expected between global efficiencies obtained with the two configurations. This difference increases when considering probes of other origins. Comparison between $\Delta\epsilon^{Data}$ and $\Delta\epsilon^{MC}$ would give a handle on combinatorial background subtraction as well as on the reliability of the MC predictions for low quality probes.
- Modified TP selection: the relative fraction of hadron probes in the $Z \rightarrow ee$ probe sample can also be altered by other means than those described in this chapter. For example, choosing pairs where the p_T of the lepton pair is $p_T^{ee} < 20$ GeV, the fraction of non-matched probes (category other than Z Boson matched + e from Z) in the lowest p_T bin is reduced from 2.6% to 0.8%. This selection is equivalent to requiring that the Z boson is produced at rest and the presence of jets which contribute to the combinatorial background can be effectively vetoed. A measurement of $\Delta\epsilon$ with respect to other configurations gives a handle on hadron probe contamination in particular.

6.7 Conclusions

Electron efficiencies are an important aspect of the detector performance and are crucial for many measurements, a typical example being a cross-section determination. The Tag-and-Probe method allows the extraction of electron efficiencies from $Z \rightarrow ee$ events and could be applied with the first $\sim$ 100 pb^{-1} of data to measure PID efficiencies in 4 x 6 η-p_T bins with a statistical precision of 1-1.5%. A challenging tasks will be the subtraction of QCD-jet background. Despite $Z \rightarrow ee$ being a very clean signature, background contamination is still non-negligible and can bias the efficiency. The subtraction of background is complicated by the fact that its shape strongly depends on the probe p_T bin. To achieve a measurement of differential efficiencies in electron p_T, it was suggested to use a template for the background shape and a parametrisation for the signal. Difficulties were encountered in performing the fit, which suffered from relatively large χ^2 values, despite the large number of degrees of freedom.

To achieve a higher precision, possibly a completely different Pdf should be used for the signal. Using a template from MC would be one option. Nevertheless, it was shown that with the current proposal, systematic uncertainties from background subtraction are at the 0.5% level for $p_T > 35$ GeV and increase above 1% for lower p_T bins where QCD-jet contamination is more significant.

It was also discussed how the TP selection cuts vary the composition of the probe sample, in terms of quality of the probe or its kinematics, and this in turn affects the measured efficiencies. Any data-to-MC comparison will only be meaningful if exactly the same methodology is applied on both. Measuring efficiencies with different TP configurations, on the other hand, we could gain sensitivity on combinatorial background or on bad quality probes. The bias of the TP selection also has to be considered when efficiencies are used in a cross-section measurement, as will be shown in the next chapter.

This chapter was dedicated in particular to the measurement of electron PID efficiencies. However, the TP method can also be applied to measure a number of different efficiencies: absolute reconstruction and trigger efficiencies (see Chapter 7) and even track reconstruction efficiencies. Furthermore, other signatures can be used for in-situ measurement of efficiencies. Very similar to $Z \to ee$ is the $J/\psi \to ee$, which would allow to extract efficiencies in the lower kinematic region, ie. for p_T from a few GeV to about 15-20 GeV. $W \to e\nu$ events are also good candidates for these kind of studies. Covering the same kinematic range as $Z \to ee$ events, they have the advantage of offering higher statistics (about 10 times than for $Z \to ee$), which is very important in early data. On the other hand, $W \to e\nu$ processes suffer from larger backgrounds, which are also more difficult to estimate.

Chapter 7

Acceptance and Efficiency Corrections

In a data analysis it is mandatory to apply a set of signal selection cuts. Some cuts are needed to constrain the measurement to the detector acceptance (for example, a fiducial cut on the electron pseudo-rapidity), whereas others are necessary to reduce the background contamination (for example, trigger and electron identification requirements). The application of these cuts causes a loss of signal events which has to be corrected for in the determination of the production cross-section. This chapter describes how $W \to e\nu$ + jets events are corrected for acceptance and efficiency losses. The terms acceptance and efficiency are defined in Section 7.1, which also describes the general strategy adopted in this analysis. The determination of the acceptance factor is described in Section 7.2, whereas efficiencies are covered in Section 7.3. Finally, Section 7.4 shows how the individual factors are combined together to determine the $W \to e\nu$ + jets production cross-section. A closure test is performed to verify that the correction factors return the input yields and the impact of systematic biases is evaluated. Emphasis is given to electron corrections. The Tag-and-Probe method described in Chapter 6 is used for the determination of the efficiencies. Corrections on missing transverse energy are only briefly discussed and are included in the final closure test for completeness. Jet corrections are not covered, but their impact is discussed in the next chapter.

7.1 Acceptance and Efficiencies

Losses of signal events due to selection cuts are corrected for by the terms $A{\cdot}\epsilon$ in the equation for the cross-section determination (see Chapter 3, eqn. (3.2)). In this thesis, the term acceptance (A) refers to corrections for geometrical and kinematic cuts, whereas the term efficiency (ϵ) accounts for losses due to reconstruction or quality requirements. Corrections can then be factorised into several terms, each accounting for selection cuts on different objects:

$$A \cdot \epsilon = (A \cdot \epsilon)_{jet} \cdot (A)^{jet}_{E_T^{\text{miss}}} \cdot (A \cdot \epsilon)^{jets+E_T^{\text{miss}}}_{ele}. \tag{7.1}$$

Each term in parenthesis represents a probability: the first term (jet) is the probability that in a $W \rightarrow e\nu$ event there are N jets satisfying the cuts described in Section 4.4.2; the second term (E_T^{miss}) represents the conditional probability that the event satisfies the E_T^{miss} cut, given that there are N reconstructed jets in the event; the last term (ele) represents the conditional probability that there is an electron passing the selection, given that the jet and E_T^{miss} cuts are satisfied. What really matters for the cross-section determination is, for each term, the overall product $A \cdot \epsilon$. The individual factors A and ϵ can be redefined in different ways, as long as the definition of acceptance is consistent with the definition of efficiency. It is therefore crucial to define the break point between A and ϵ in a consistent way. In this thesis, the break point is chosen at reconstruction level, as to decouple factors which have to be determined from MC from factors which can be determined from data. Each term can be rewritten as (from now on the superscripts are dropped to simplify the notation):

$$(A \cdot \epsilon)_{jet} = A_{Gen}[N_{jets}] \cdot A_{Gen \rightarrow Rec}[N_{jets}] \cdot \epsilon_{\text{jet}} \tag{7.2a}$$

$$(A)_{E_T^{\text{miss}}} = A_{Gen}[E_T^{\text{miss}}] \cdot A_{Gen \rightarrow Rec}[E_T^{\text{miss}}] \tag{7.2b}$$

$$(A \cdot \epsilon)_{ele} = A_{Gen}[ele] \cdot A_{Gen \rightarrow Rec}[ele] \cdot \epsilon_{\text{ele}} \tag{7.2c}$$

Term	Definition	Cuts
$A[N_{jets}]$	Jet Acceptance	N_{jets} in event with: $\lvert\eta_{jet}\rvert < 3.1$ $p_{T\,jet} > 30$ GeV
$\epsilon_{\text{reco jet}}$	Jet Reconstruction Eff.	jet object
$\epsilon_{\Delta\text{Rej}}$	Eff. of ΔR_{ej} cut	$\Delta R_{ej} > 0.6$
$\epsilon_{\Delta\text{Rjj}}$	Eff. of ΔR_{jj} cut	$\Delta R_{jj} > 0.6$
$A[E_T^{\text{miss}}]$	E_T^{miss} Acceptance	$E_T^{\text{miss}} > 25$ GeV
$A[ele]$	Electron Acceptance	$\lvert\eta_{ele}\rvert < 2.46$ outside cracks ($\lvert\eta\rvert = 1.37 - 1.52$) $p_{T\,ele} > 25$ GeV
ϵ_{reco}	Electron Reconstruction Eff.	electron object
ϵ_{id}	Electron Identification Eff.	medium PID
ϵ_{trig}	Electron Trigger Eff.	chain e20_loose

Table 7.1: Definition of acceptance (A) and efficiency (ϵ) correction factors.

and we define:

$$A_{Rec} = A_{Gen} \cdot A_{Gen \to Rec} \tag{7.3a}$$

$$\epsilon_{\text{jet}} = \epsilon_{\text{reco jet}} \cdot \epsilon_{\Delta\text{Rej}} \cdot \epsilon_{\Delta\text{Rjj}} \tag{7.3b}$$

$$\epsilon_{\text{ele}} = \epsilon_{\text{reco}} \cdot \epsilon_{\text{id}} \cdot \epsilon_{\text{trig}}. \tag{7.3c}$$

The individual factors are defined in Table 7.1. A_{Gen} is the acceptance at truth level (ie. the fraction of events that pass the geometrical or kinematic cuts on generator-level quantities) and must be determined from MC. $A_{Gen \to Rec}$ is a correction factor which accounts for detector smearing from generator to reconstruction level. Because of detector scale and resolution, a cut on generator level quantities will lead to a slightly different number of events than if it was applied on reconstruction level quantities. This factor can be evaluated from MC or from data, as will be discussed in Section 7.2.2. A_{Rec} is the acceptance at detector level and is simply the product of A_{Gen} and $A_{Gen \to Rec}$. Finally, the efficiency ϵ is relative to reconstruction quantities exclusively. Electron efficiencies can be determined entirely from data, using the TP method (cf. Chapter 6). Jet cut efficiencies will be determined from MC.

Since the aim of the analysis is to measure the cross-section for $W \rightarrow e\nu$ production in association with jets, the factor $A_{Gen}[N_{jets}]$ should not be included in the corrections, but rather taken to the left hand side of equation 3.2. The final measurement will then be $\sigma \cdot A_{Gen}[N_{jets}]$ which will be compared to theoretical predictions. All the other terms of eqn. (7.2) have to be determined. Jets can be corrected back to the particle level through an unfolding procedure which accounts both for efficiency and scale/resolution effects. Unfolding, however, is not treated in this thesis. The objective, instead, is to calculate the $E_T^{\rm miss}$ and electron correction terms. Throughout this chapter, jets are defined at reconstruction level, after all selections described in Chapter 4, Section 4.4.3. This is necessary to decouple any detector effect on the measurement of hadronic jets from the corrections under study. Furthermore, as will be shown later, acceptance, efficiencies and detector smearing of electron and $E_T^{\rm miss}$ objects strongly depend on the event topology, which in this case is characterised by the large hadronic activity.

7.2 Acceptance Corrections

The factors A_{Gen} and $A_{Gen \rightarrow Rec}$ are calculated using the full statistical power of the ALPGEN $W \rightarrow e\nu$ + *N partons* samples, weighted according to their relative cross-sections. The acceptance of a given cut for the jet multiplicity N_{jets} is determined as:

$$A = \frac{\text{Events with N}_{\text{jets}}\text{ and passing cut}}{\text{Events with N}_{\text{jets}}}. \qquad (7.4)$$

The fact that the number of jets enters in the denominator of A introduces a subtlety in the acceptance calculation. The jet definition includes an electron-jet overlap removal which is performed, in the measurement, at reconstruction level after a good quality electron has been identified. In the acceptance calculation, however, a good quality reconstructed electron is not yet available. The overlap removal is therefore performed between reconstructed jets and true electrons from the W boson decay, using a procedure identical to that used on pseudo-data. The

Acceptance (%)			
Definition	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} = 2$
$A_{Gen}[E_T^{\text{miss}}]$	70.25 ± 0.04	74.66 ± 0.10	76.79 ± 0.10
$A_{Gen}[E_T^{\text{miss}}] \cdot A_{Gen}[ele]$	37.04 ± 0.04	35.24 ± 0.11	39.37 ± 0.11
$A_{Rec}[E_T^{\text{miss}}]$	71.14 ± 0.04	78.00 ± 0.09	79.43 ± 0.10
$A_{Rec}[E_T^{\text{miss}}] \cdot A_{Gen}[ele]$	37.99 ± 0.04	37.69 ± 0.11	41.32 ± 0.12
$A_{Rec}[E_T^{\text{miss}}] \cdot A_{Gen}[ele + FSR\gamma]$	39.04 ± 0.04	38.54 ± 0.11	42.15 ± 0.12
$A_{Rec}[E_T^{\text{miss}}] \cdot A_{Rec}[ele + FSR\gamma]$	37.90 ± 0.04	37.54 ± 0.11	41.18 ± 0.11

Table 7.2: Acceptances for $W \to e\nu$ + 0, 1 and 2 jets estimated with ALPGEN. Only statistical errors are quoted.

effect of using true rather than reconstructed electrons was investigated on the ALPGEN sample after all selection cuts and found to be negligible with respect to the statistical errors.

7.2.1 Generator Level Acceptance

The acceptance at generator level includes the E_T^{miss} term $A_{Gen}[E_T^{\text{miss}}]$ and the electron term $A_{Gen}[\eta_{ele}, p_{T\,ele}]$ (which will be abbreviated as $A_{Gen}[Ele]$). The electron kinematic cuts are applied on truth quantities after QED final state radiation (FSR). Corrections to the acceptance to account for FSR photons clustered with the electron will be treated in Section 7.2.2.

The acceptance, shown in the first two lines of Table 7.2, varies between jet multiplicities by up to 4%, being minimal for $N_{jets} = 1$ and maximal for $N_{jets} = 2$. Such differences can be understood considering the distribution of the W boson p_T (p_T^W) for events with different number of jets, as shown in Fig. 7.1. The following features are observed:

- The acceptance strongly depends on the W boson p_T (p_T^W), reaching a minimum at $\sim$ 40-50 GeV, roughly in correspondence of half the W boson mass (M^W), and steadily increasing at higher p_T^W. This can be understood by simple kinematics: in the rest frame of the W boson, by momentum conservation, electron and neutrino are emitted back to back with equal momentum magnitudes $|\vec{p}^{\,\text{Wframe}}| = \frac{M^W}{2}$. When moving to the laboratory frame, the particles are boosted according to p_T^W. For small values of p_T^W, one particle p_T will increase while the other will decrease, causing a drop in the acceptance. At a certain point, one of the particles will stop completely (this corresponds to the minimum acceptance), after which both electron and neutrino p_T will increase. If, for example, electron and neutrino are emitted at $\eta = 0$, for one of the particles $p_T = 0$ will occur approximately when the boost $p_T^W = M^W/2$. The overall acceptance is a convolution of the p_T^W spectrum with the differential acceptance as a function of p_T^W. For $N_{jets} = 0$, most Ws decay at rest, resulting in a rather large acceptance. In the presence of jets, the W boson is boosted: for $N_{jets} = 1$ a higher number of events lie around the minimum, $p_T^W = M^W/2$, explaining why the acceptance drops with respect to the 0 and 2 jet bins (cf. Table 7.2).

- The acceptance is also affected by the requirement of a central jet, which also forces the W, and therefore the electron, to be produced centrally. The presence of jets, therefore, increases the probability that the electron will satisfy the geometrical requirement on η. This explains why the acceptance as function of p_T^W is systematically lower for the 0 jet bin with respect to higher multiplicities.

Correlated with p_T^W and therefore also affecting the acceptance are jet variables such as the leading jet p_T and HT. More energetic jets correspond in general to a more boosted W boson and thus higher acceptance (see bottom plots of Fig. 7.1). The correlation is not perfect since jets can also recoil against each other and be accompanied by a rather soft W. Therefore the variables that determine the acceptance are the presence of energetic jets and/or p_T^W.

Although the acceptance is derived from MC, it is always preferable to use binned corrections, rather than a global factor, to minimise the dependence on the theoretical model used. The choice of binning depends on the measurement. Ideally, p_T^W should be chosen since it is the driving variable for the acceptance. When measuring a cross-section as a function of leading jet p_T and HT, however, jet variables are preferable. In the closure test described in Section 7.4, it was chosen to bin the acceptance in both p_T^W (four bins: p_T^W = 0-30, 30-70, 70-140, $>$ 140 GeV) and leading jet p_T (for $N_{jets} = 1$) or HT (for $N_{jets} = 2$) (three bins: leading jet p_T = 0-80,

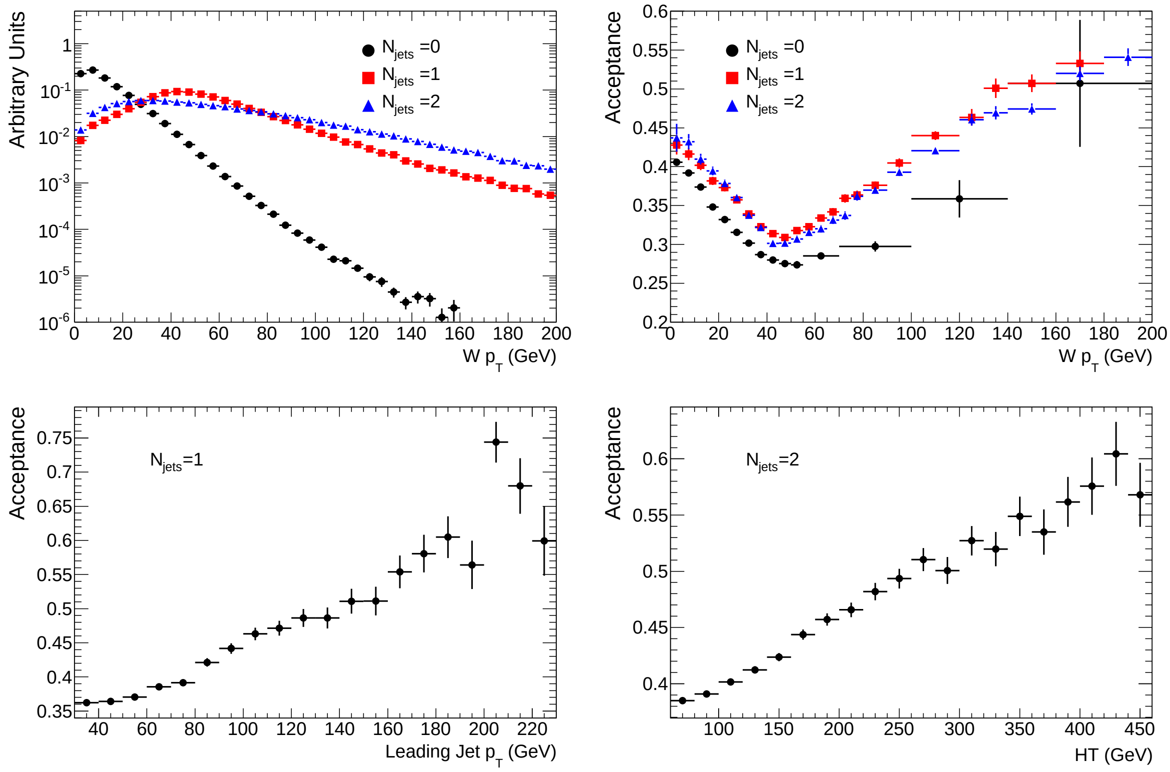

Figure 7.1: (Top) W boson p_T distribution for different jet multiplicities (left) and acceptance at generator level as a function of W boson p_T (right). Bottom: generator level acceptance as a function of leading jet p_T for the 1 jet bin (left) and HT for the 2 jet bin (right).

80-150, $>$ 150 GeV; HT = 0-120, 120-200, $>$ 200 GeV).

7.2.2 Detector Level Acceptance

Detector effects on the acceptance can be estimated using MC simulations, simply applying the kinematic requirements on reconstructed rather than truth-level quantities. However this approach relies completely on the correctness of the detector simulation. An alternative procedure is to smear by hand truth-level quantities according to scale and resolutions measured in data. This strategy was developed in [71], where both electron p_T and $E_T^{\rm miss}$ smearing procedures were studied. In this analysis detector effects on $E_T^{\rm miss}$ are taken from MC, whereas a smearing procedure is applied to electron p_T for the specific measurement of W + jets.

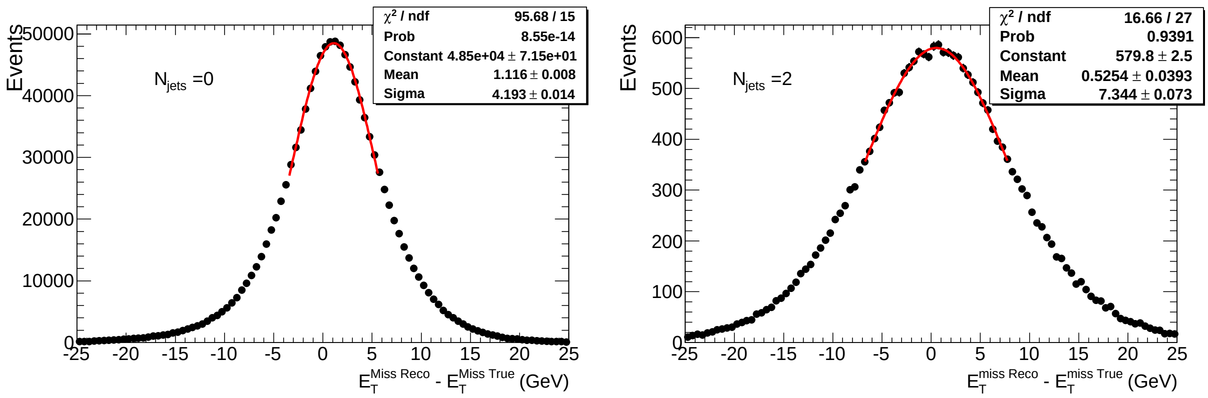

Figure 7.2: Gaussian fit to $E_T^{\rm miss\ Reco} - E_T^{\rm miss\ Truth}$ in events with 0 (left) and 2 (right) jets.

Detector effects on $E_T^{\rm miss}$

The $E_T^{\rm miss}$ measurement is affected by the calorimeter energy resolution as well as by the hadronic/EM energy calibration which can result in a $E_T^{\rm miss}$ scale bias (that is, reconstructed $E_T^{\rm miss}$ being systematically mis-estimated). The net effect of $E_T^{\rm miss}$ scale and resolution is to increase the acceptance of the $E_T^{\rm miss}$ cut (third line of Table 7.2). The magnitude of the increase varies with jet multiplicity, being less than 1% for the 0 jet bin and $\sim$ 3% for 1 and 2 jets. This is not surprising, since $E_T^{\rm miss}$ resolution depends on the hadronic activity in the event (see Section 4.2 Chapter 4) and is therefore expected to worsen with higher number of jets. This is also illustrated in Fig. 7.2 which shows the difference between reconstructed and true $E_T^{\rm miss}$ in events with 0 and 2 jets, fitted to a Gaussian function. The mean and sigma represent the scale and resolution respectively. The distributions gets broader with the number of jets (contrary to the bias, which decreases with N_{jets}), suggesting that the resolution is the dominant effect on the acceptance.

Another important factor is the position of the $E_T^{\rm miss}$ cut with respect to the peak of $E_T^{\rm miss}$ spectrum. Figure 7.3 overlays true and reconstructed $E_T^{\rm miss}$ distributions for the 0 and 2 jet bins. For $N_{jets} = 0$, the 25 GeV cut removes the low side of the spectrum, which is only weakly affected by detector smearing. In contrast, for $N_{jets} = 2$ (and similarly for $N_{jets} = 1$), the 25 GeV cut happens nearer to the peak, where the smearing effect is much larger. The net effect on the acceptance is therefore also more significant.

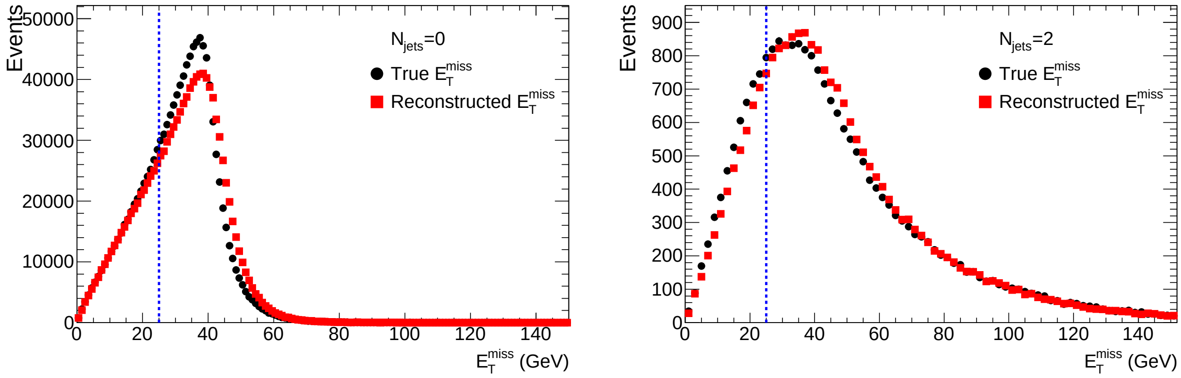

Figure 7.3: True and reconstructed $E_T^{\rm miss}$ distribution for events with 0 jets (left) and 2 jets (right). The blue dotted line shows the position of the $E_T^{\rm miss}$ cut applied in the analysis.

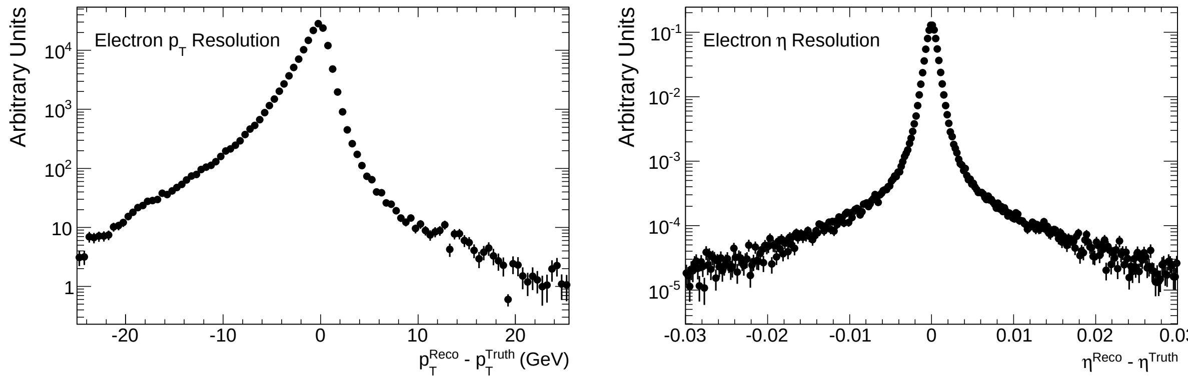

Figure 7.4: Detector response on electron p_T (left) and η (right).

Detector effects on electrons

Electron selection cuts include kinematic constraints on both electron p_T and pseudo-rapidity. The detector response on electron η and p_T is described by the resolution (or smearing) functions shown in Fig. 7.4. If fitted to a Gaussian, the mean (μ) and width (σ) of these distributions represent the scale and detector resolution in η/ p_T. From Fig. 7.4 it can be noticed that the η resolution is completely negligible compared to the energy resolution. Although the η resolution also has a smearing effect on the electron p_T, this is about ten times smaller than the smearing due to the p_T resolution itself. The η resolution can therefore be neglected. Detector effects on electron p_T, in contrast, have to be accounted for.

A procedure for smearing the electron p_T and determining the detector level acceptance is briefly described here (the individual steps will be covered in more detail later in this section):

1. Determination of the detector response: the detector response is represented by a smearing function ($f_{p_T}^{Gen\rightarrow Rec}$). When sufficient LHC data is collected, $f_{p_T}^{Gen\rightarrow Rec}$ will be derived in-situ using $Z \rightarrow ee$ events, comparing the reconstructed di-electron invariant mass with the well-known Z boson line shape, as described in [58]. In this analysis, however, the smearing function is obtained from MC.

2. p_T smearing: for a true electron in the ALPGEN $W \rightarrow e\nu$ sample, a random number is extracted from the $f_{p_T}^{Gen\rightarrow Rec}$ distribution. The smeared electron p_T is then given by:

$$p_T^{Smeared} = p_T^{Truth} + f_{p_T}^{Gen\rightarrow Rec} \rightarrow GetRandom(). \tag{7.5}$$

3. Acceptance calculation: the acceptance is then determined on the smeared electron p_T (and true η, since detector smearing on η is negligible).

1. Determination of the detector response

The idea behind this smearing technique is to determine the detector response in a data-driven approach and propagate it into the acceptance. In this way truth to detector-level corrections can be determined without relying on the MC simulation. In the absence of sufficient data, the smearing function ($f_{p_T}^{Gen\rightarrow Rec}$) is obtained from a MC sample of $Z \rightarrow ee$ + jets events, plotting the difference between reconstructed and true electron p_T ($p_T^{Rec} - p_T^{Truth}$), as shown in Fig. 7.4 (left). This procedure requires truth matching between electrons at generator and reconstruction level, which is performed requiring that the geometrical distance between the two is $\Delta R < 0.2$.

It is also important to notice from Fig. 7.4 (left) that the energy resolution function is not well described by a simple Gaussian, because of the energy losses by bremsstrahlung radiation of electrons showering in the ID (exhibited in the low tail of the curve). To avoid biases introduced by a poor functional form, the smearing function is not parametrized. Instead, the histogram obtained from ($p_T^{Rec} - p_T^{Truth}$) is directly used to represent $f_{p_T}^{Gen\rightarrow Rec}$. An example of a smearing function is given by the histogram shown in Fig. 7.4 (left).

The detector response is extracted as a function of the variables it mostly depends upon. The p_T resolution, as well as the p_T scale, is driven by the p_T of the electron itself, as shown in the top plots of Fig. 7.5. However, also the detector region crossed by the electron has an important effect. The resolution varies across η as a result of different amount of material in front of the calorimeter and is particularly deteriorated in the crack region (η = 1.37-1.52), as illustrated in the bottom plots of Fig. 7.5. The same figure also compares the detector response in $W \to e\nu$ and $Z \to ee$ events. The slightly worse resolution in $Z \to ee$ events (see plots as a function of η) is due to the harder electron p_T spectrum. The differences between $Z \to ee$ and $W \to e\nu$ events disappear when scale and resolutions are binned in p_T. Such considerations drive the choice of binning for the map of electron response. The smearing function is determined in bins of electron η and p_T ($f_{p_T}^{Gen \to Rec}(p_T, \eta)$), with 3 bins in η ($|\eta| = 0., 0.6, 1.37, 1.52, 2.46$, bin 1.37-1.52 excluded) and 9 bins in p_T (5 GeV bins in the range p_T = 15-50 GeV and 10-20 GeV bins for $p_T > 50$ GeV). Binning in jet multiplicity is not consider necessary, since no dependence on the hadronic activity in the event is observed.

2. p_T Smearing

The true electron p_T is smeared with a random number extracted from the smearing function. For the smearing procedure to work, it is essential to define the true electron p_T correctly. As mentioned above, true electrons used in the acceptance calculation are taken after QED FSR (also called internal bremsstrahlung). However, the detector is not always able to resolve the photon from internal bremsstrahlung from the parent electron. If the photon is emitted within a small cone, it will be clustered together with the electron. In fact, in the majority of the cases, the photon is at a small distance from the electron, as shown in Fig. 7.6 left.

The radius of this so called recombination cone (ΔR_{recomb}) is determined by the spatial resolution of the calorimeter clustering algorithm, which is estimated to be approximately 0.09 [1]. To

[1] This number is estimated considering that the sliding window algorithm starts by pre-clustering towers of size 5×5 in units of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$.

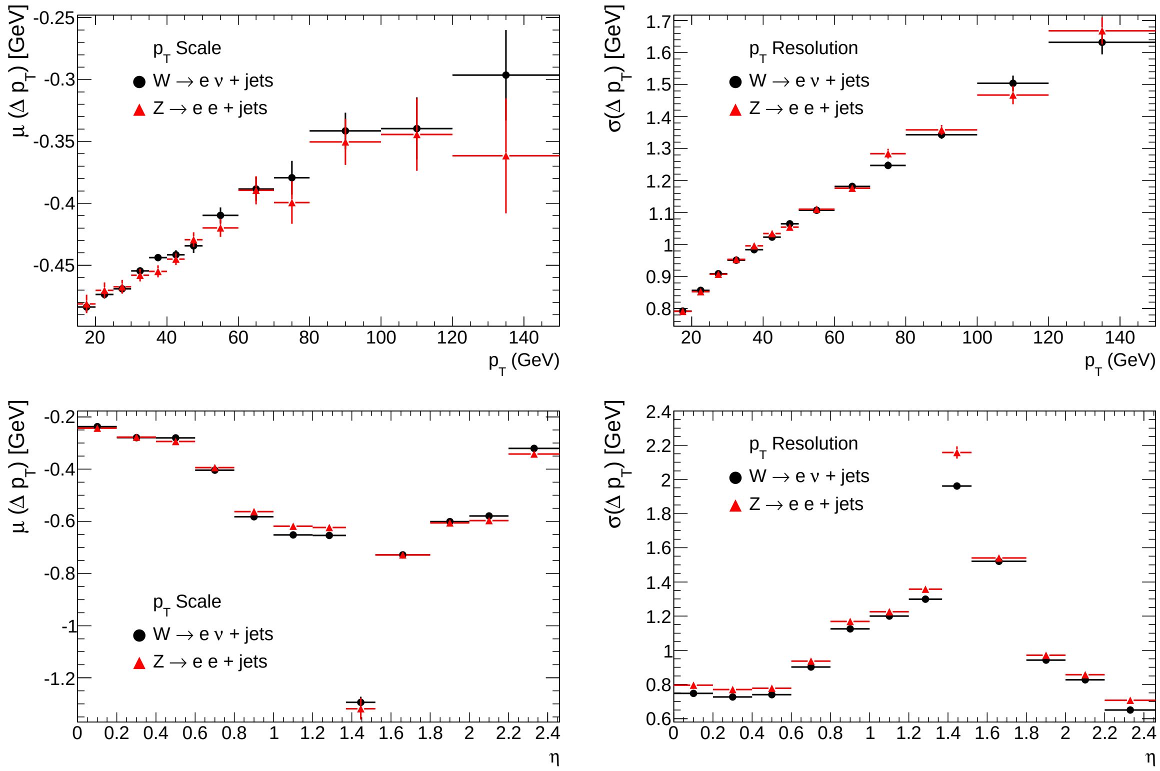

Figure 7.5: Electron p_T scale (left) and resolution (right) as a function of electron p_T (top) and η (bottom), in $W \rightarrow e\nu$ + jets and $Z \rightarrow ee$ + jets events.

check whether this is a sensible value, the distance ΔR between the W-boson electron and the closest photon at truth level ($\Delta R^{Truth}_{e-\gamma}$) is plotted against that at reconstruction level ($\Delta R^{Rec}_{e-\gamma}$). This is illustrated on the right plot of Fig 7.6: a strong correlation is visible between $\Delta R^{Truth}_{e-\gamma}$ and $\Delta R^{Rec}_{e-\gamma}$, down to approximately $\Delta R^{Truth}_{e-\gamma} = 0.09$ (indicated by the dashed red line). Below this value of ΔR, the FSR photon cannot be resolved and the reconstructed electron is matched to another photon in the event, which is not the photon from internal bremsstrahlung. However, not all photons at a distance $\Delta R > 0.09$ are resolved. In other words, the turn-on is not a sharp step but (not surprisingly) rather a slow turn-on curve from $\Delta R = 0.08$ up to $\Delta R = 0.16$. This is illustrated on the bottom plot of Fig. 7.6, showing the distance between reconstructed electrons and photons, $\Delta R^{Rec}_{e-\gamma}$.

The radius for FSR photon merging is chosen as $R_{recomb} = 0.09$, to be consistent with the clustering algorithm resolution. Given the ambiguity in the choice of merging cut, the effect of increasing R_{recomb} up to a value of 0.16 is estimated as a systematic uncertainty on the

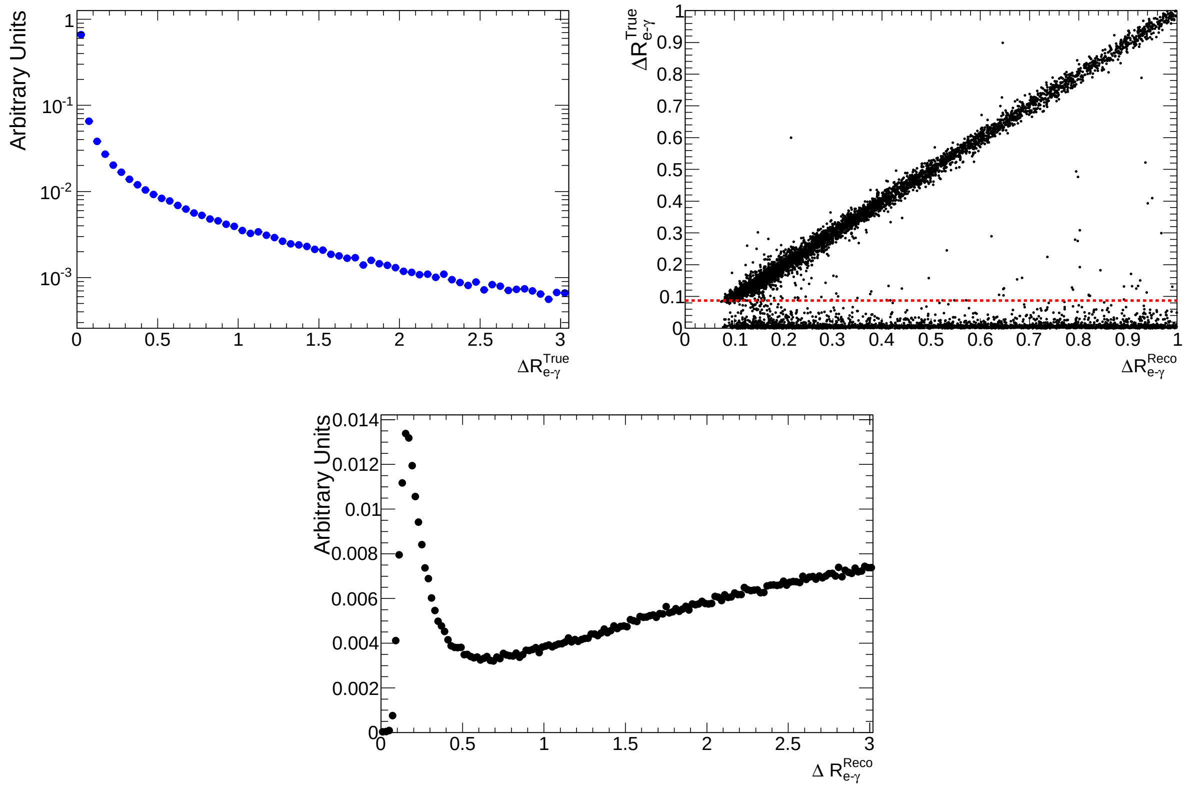

Figure 7.6: Distance between true electrons from W boson decay and FSR photons at generator-level (top left) and same variable plotted against the distance between electrons and photon objects at reconstruction level (top right). The bottom plot shows the distribution ΔR between reconstructed electrons and photons (same as x-axis of top right plot).

acceptance. In order to correctly mimic the response of the detector, any FSR photon emitted within the recombination cone $\Delta R_{recomb} = 0.09$ has to be merged with the true electron by adding their 4-momenta. As a result of this merging, the generator-level acceptance is increased by approximately 1% (see Table 7.2). The smearing procedure therefore has to be applied on the combined p_T of electron and FSR photon, since this is the quantity which will be measured. From now on, when referring to true electron p_T, this will imply that photon merging has been performed.

As a result of the smearing, electrons with $p_T < 25$ GeV at *generator* level could still contribute to the sub-sample of events with one electron having $p_T > 25$ GeV at *reconstruction* level. Vice-versa, electrons with true p_T above 25 GeV, might not satisfy the kinematic cut at *reconstruction* level. The latter case is more likely to happen, given the form of the smearing function being non-symmetric about zero (see Fig. 7.4 left). Furthermore, it was observed that

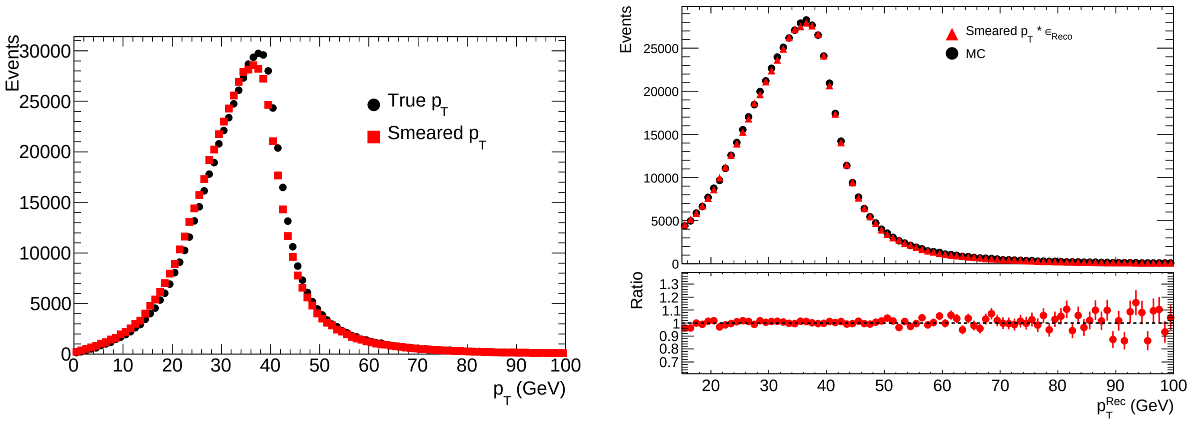

Figure 7.7: (Left) p_T distribution of the true electron from W boson decay before (black) and after (red) smearing. (Right) Smeared electron p_T distribution folded with reconstruction efficiency (red) compared with the p_T distribution of reconstructed electrons obtained from full MC simulation (black). The ratio between MC and smeared distribution is also shown.

less than 0.06% of the reconstructed electrons with $p_T > 25$ GeV originates from a true electron with $p_T < 15$ GeV. It was considered safe to apply the smearing procedure only to electrons above the threshold of 15 GeV, as was also suggested in [71], since the resolution of low p_T objects is poorly understood. It was checked that the impact on the acceptance is totally negligible with respect to the statistical error.

3. Acceptance calculation

The effect of detector smearing on the electron p_T distribution is illustrated in Fig. 7.7. It tends to flatten the peak shifting events preferably to the low p_T side. The detector-level acceptance for the electron term is calculated on this smeared distribution. The results are shown in the last line of Table 7.2: the acceptance is decreased by approximately 1%. The effect of FSR photon merging and detector smearing act, on average, in an opposite direction but the fact that they almost cancel each other is purely coincidental.

This procedure was validated by comparing the smeared p_T distribution with the reconstructed-level p_T distribution obtained from fully simulated MC. The acceptance obtained from the smearing technique is also compared to that obtained from MC. Since, in this analysis, the smearing functions are also obtained from MC, any difference in the estimated acceptance would be due to the smearing technique. This would include biases introduced by the FSR photon merging (since no merging is needed when relying entirely on fully simulated MC) or from event topology (ie. deriving $f_{p_T}^{Gen \to Rec}$ from $Z \to ee$ rather than $W \to e\nu$). There is however one additional complication in the comparison: the acceptance from MC has to be calculated on electron objects and therefore folds in the electron reconstruction efficiency. The smearing technique, in contrast, accounts for scale and resolution exclusively. In order to compare like with like, the smeared electron p_T distribution is folded with the reconstruction efficiency (extracted from MC) and the acceptance is re-computed. Note that this is a purely academic exercise, and this is not the final number used for the cross-section measurement. The comparison of the electron p_T distribution at reconstruction level between MC and smearing procedure (plus efficiency) is shown in Fig. 7.7 right. Very good agreement is observed, as also demonstrated by the ratio (lower plot) being constant around 1 over the full spectrum. The acceptance was also found to be in agreement with the one from MC, within statistical errors. The smearing technique was therefore considered suitable for estimating detector-level corrections on electrons.

7.2.3 Systematic Uncertainties on Acceptance

Systematic uncertainties on the acceptance can be classified in two different categories: theoretical uncertainties due to the input model, which affect the generator level acceptance A_{Gen} and uncertainties on the detector response, which affects the factor $A_{Gen \to Rec}$. Only the latter are discussed here.

Uncertainties on Detector Response

The systematic uncertainties on the acceptance arising from E_T^{miss} detector response were taken from [71]. Although in this thesis, the correction factor $A_{Gen \to Rec}[E_T^{\text{miss}}]$ was taken from MC, in data a similar procedure to the electron smearing technique will be applied. The expected relative systematic uncertainty is 0.8% and this includes the experimental uncertainty on the determination of $E_T^{\rm miss}$ scale and resolution from data (14.9% and 9.7% respectively) and the uncertainties from the smearing procedure.

Systematic uncertainties on the electron detector response were evaluated as follows:

1. The radius of the recombination cone R_{recomb} used for FSR photon merging were varied between 0.08 and 1.6, which are the limits of the FSR photon resolution turn-on curve (see bottom plot of Fig. 7.6). The maximum variation in the acceptance (after smearing) with respect to the default R_{recomb} value of 0.09 was taken as the systematic uncertainty. This is a very conservative estimate, since in reality the detector will not miss/or resolve all FSR photons in the range 0.08-1.6, but only a fraction of them.

2. The smearing function $f_{p_T}^{Gen \to Rec}$ was extracted from $W \to e\nu$ + jets rather than $Z \to ee$ + jets ALPGEN samples. This would account for variations in the detector response due to event topology, which should in any case be negligible if the correct binning was chosen.

3. In this analysis the detector response for electrons was perfectly known, since it was taken from MC. Ultimately, scale and resolution will be determined from data with an expected precision of 0.2% (scale) and 0.7% (resolution) [2]. It was shown in [71] that such uncertainties have a negligible effect on the acceptance and can therefore be neglected.

The systematic uncertainties on the acceptance from the detector response are reported in Table 7.3. The total (absolute) uncertainty, given by the sum in quadrature of the above, is $\sim\ 0.3\%$ and is dominated by the uncertainty on the $E_T^{\rm miss}$ corrections. The overall effect of $E_T^{\rm miss}$ and electron scale and resolution on the acceptance is much larger than this uncertainty for all multiplicity bins: $A_{Rec} - A_{Gen}$ = 0.86 (0 jets), 2.30 (1 jet), 1.81 (3 jets).

7.3 Electron Efficiency Corrections

Electron efficiencies are determined using the Tag-and-Probe (TP) method on $Z \to ee$ + jets ALPGEN samples. A detailed study of the TP method in the context of identification efficien-

Systematic uncertainties on A (%)

N_{jets}	$A \pm \Delta A_{stat}$	$\Delta A_{E_T^{\rm miss}}$	ΔA^{γ}_{ele}	$\Delta A^{Z\to ee}_{ele}$	Tot ΔA_{syst}
0	37.90 ± 0.04	0.30	0.09	0.01	0.31
1	37.54 ± 0.11	0.30	0.02	0.06	0.31
2	41.18 ± 0.11	0.33	0.06	0.05	0.34

Table 7.3: Acceptance (in %) with statistical (ΔA_{stat}) and systematic uncertainties from corrections for $E_T^{\rm miss}$ and electron smearing. Uncertainties on corrections for $E_T^{\rm miss}$ scale and resolution ($\Delta A_{E_T^{\rm miss}}$) were taken from [71]. Uncertainties on electron detector response are evaluated by varying the radius of the recombination cone for FSR photons (ΔA^{γ}_{ele}), and by using $W \to e\nu$ + jets events rather than $Z \to ee$ + jets to extract the smearing function ($\Delta A^{Z\to ee}_{ele}$).

cies determination was given in Chapter 6. The same technique is applied here with minimal variations to extract three efficiencies: $\epsilon_{\rm reco}$, $\epsilon_{\rm id}$ and $\epsilon_{\rm trig}$. Special care is given to the effect of hadronic activity on the efficiencies. Differences between efficiencies in $W \to e\nu$ and $Z \to ee$ events are also investigated to assess any systematic uncertainty arising from using TP efficiencies in the $W \to e\nu$ + jets cross-section measurement.

The TP selection on $Z \to ee$ events is adapted to this analysis as follows:

- Jet counting is performed with the same selection cuts as in the $W \to e\nu$ + jets analysis, as described in Chapter 4.
- The tag electron selection is identical to that described in Table 6.1 (with Tight PID).
- The probe definition varies depending on the efficiency under study (see Table 7.4): it is an EM cluster in the case of $\epsilon_{\rm reco}$ measurement, or an electron (medium PID electron) in the case of $\epsilon_{\rm id}$ ($\epsilon_{\rm trig}$) measurement. The kinematic cuts on the probe η and p_T are consistent with the acceptance definition.
- The invariant mass of the TP pair is required to be in the window 80-100 GeV. Two TP configurations are tested: $BaseTP$ and $\pm TP$. No additional $\Delta\phi$ separation between tag and probe is applied, to avoid introducing any bias in the probe kinematic distribution.

Efficiency	Probe (Den)	Probe (Num)
ϵ_{reco}	EM cluster	matched to electron
ϵ_{id}	electron	passes medium PID cuts
ϵ_{trig}	medium PID electron	match to trigger $\mathrm{e20\_loose}$
		LVL1 match: $\Delta R <\ 0.2$
		LVL2 match: $\Delta R <\ 0.1$
		EF match: $\Delta R <\ 0.1$

Table 7.4: Probe definition for extracting electron reconstruction (ϵ_{reco}), medium identification (ϵ_{id}) and trigger (ϵ_{trig}) efficiency. At the denominator of the efficiency (N1), the probe is a cluster or an electron object (see middle column) which satisfy the acceptance cuts: $p_T > 25$ GeV, $|\eta| < 2.46$ (cracks excluded). The last column specifies the additional requirements that have to be satisfied at the numerator (N2) of the efficiency.

7.3.1 Dependence of Efficiencies on Jet Variables

When choosing the binning for the efficiency maps it is necessary to take into account all the variables the efficiency could depend upon. In Chapter 6 it was already noted that the driving variables for the electron identification efficiencies (but the same is true for ϵ_{reco} and ϵ_{trig}) are electron kinematics, namely η and p_T. In a measurement involving jets, it is also important to check whether efficiencies exhibit any dependence on jet variables.

The identification efficiency is the quantity which is most sensitive to hadronic activity since shower shape variables used in the definition of medium PID cuts, such as hadronic leakage and lateral shower shape, are affected by the presence of nearby jets. This is illustrated in Fig. 7.8 left, which shows the medium PID efficiency as a function of distance between the true electron from W boson decay and the closest jet. For comparison, medium PID efficiency with an additional calorimeter isolation requirement is overlayed (although this is not used in the analysis). The efficiency clearly falls at small values of ΔR_{ej} and the drop is more pronounced with an explicit isolation requirement. However, for $\Delta R_{ej} > 0.6$ the efficiency is reasonably flat (with variations within 0.2%) and above 90%. Therefore, the minimum electron-jet separation imposed in the analysis (cf. Chapter 4, Section 4.4.2) ensures that the selected electrons all lie in the high efficiency region. The dependence on jet multiplicity (see right plot) is completely

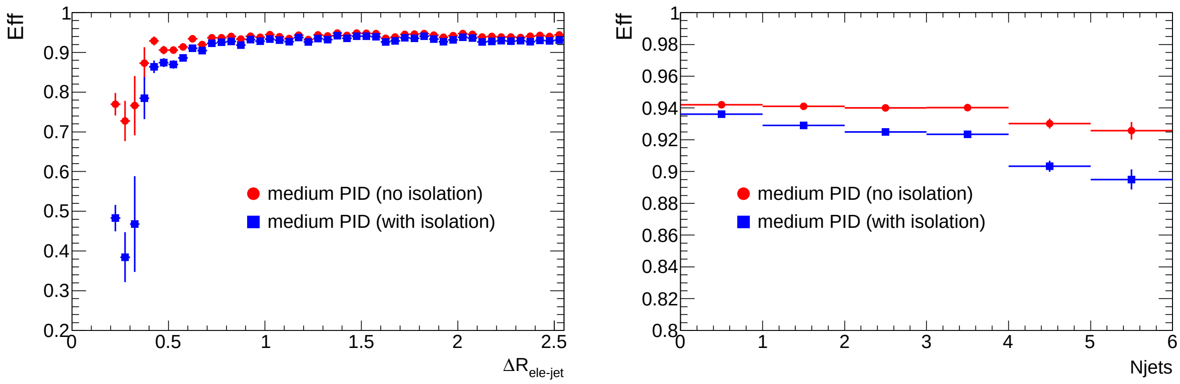

Figure 7.8: Medium PID efficiencies with and without an isolation requirement as a function of distance between electron and closest jet (left) and jet multiplicity (right).

negligible up to $N_{jets} = 3$ if no explicit isolation cut is applied. Efficiencies as a function of leading jet p_T and HT were also studied, but no significant dependence was found (ie. efficiency is flat within 0.2%), beside that due to the a harder electron p_T spectrum.

It was therefore concluded that for this analysis no explicit jet-dependent corrections are needed. It is sufficient to extract TP efficiencies as a function of electron kinematics (η and p_T). Minimising the number of variables (and hence bins) needed for the efficiency maps is advantageous especially if one is limited by statistics, as is the case for TP efficiencies from $Z \to ee$.

7.3.2 Efficiency Maps for the Cross-Section Measurement

Efficiencies extracted with the TP method are binned in a 2-dimensional grid of size 4x6 in electron η and p_T, as discussed in Chapter 6 (see Table 6.12). The statistical errors are within 1% in all bins, with the exception of the low p_T end-cap bins where they reach 1.1%. TP efficiencies for reconstruction, identification and trigger are compared as a function of η and p_T in Fig. 7.9. ϵ_{reco} and ϵ_{trig} are above 97%, with the trigger efficiency approaching 100% at high p_T. The largest source of inefficiency arises from electron identification. In fact, the electron reconstruction algorithm is designed to yield a very high efficiency by requiring only a loose match between EM cluster and ID track. Rejection of background from QCD-jets, instead, is

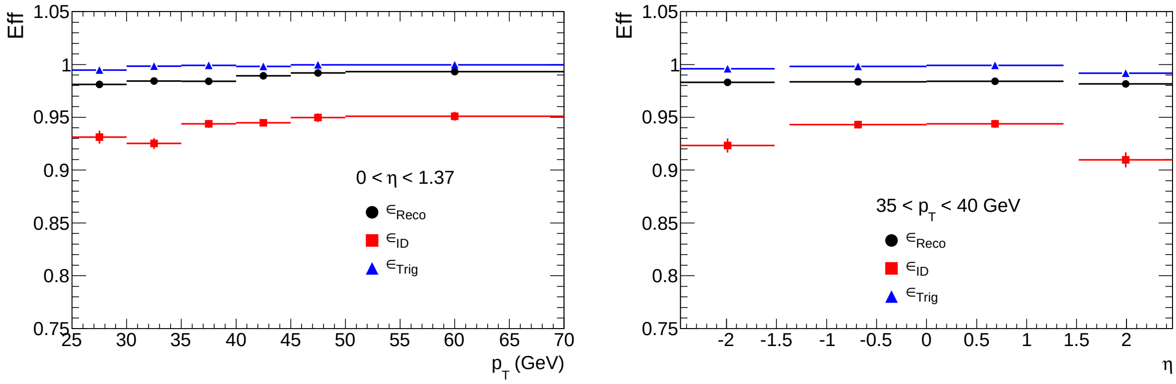

Figure 7.9: Efficiencies measured with the TP method. Left: as a function of electron p_T, for a barrel bin. Right: as a function of electron η, for $35 < p_T < 40$ GeV.

ensured by the identification cut, inevitably leading to some inefficiencies. The trigger chain $\mathrm{e20\_loose}$ is also optimised with respect to offline cuts and already reaches the plateau at 25-30 GeV.

Systematic errors on the TP method arising from background subtraction were discussed in Chapter 6 in the context of the extraction of electron identification efficiencies. The systematic error estimations shown in Table 6.11 are extended to ϵ_{reco} and ϵ_{trig} with the following assumptions:

- The same systematics are assigned to ϵ_{reco}. This assumes that the larger background level in the denominator of ϵ_{reco} (the probe is an EM cluster, rather than an electron object) does not significantly affect the invariant mass fit. This is justified by the fact that the systematic errors evaluated in Chapter 6 already included variations for a background level three times higher than expected (this was a conservative estimate for ϵ_{id}). Moreover, if the background contamination in data was too high, tighter selections on the tag electron (for example, the addition of an isolation requirement) could be applied to improve the purity.

- ϵ_{trig} is assigned systematic errors from variation of the signal Pdf and the fit range, but not of the background level. In fact, the denominator of ϵ_{id} already contains a probe satisfying the medium identification cuts, therefore the level of background is expected to be very low. The background level assumed for the Z mass fit studies shown in Chapter 6 is therefore

Tag-and-Probe Efficiencies ($Base\,TP$ Configuration) (%)

	$\epsilon(\%)$	$\Delta\epsilon_{stat}$	$\Delta\epsilon^{\text{Bkg}}_{syst}$	$\Delta\epsilon^{\text{W}}$	Tot $\Delta\epsilon$
ϵ_{reco}	98.61	0.05	0.59	+ 0.54	0.80
ϵ_{id}	92.52	0.10	0.59	+ 0.66	0.96
ϵ_{trig}	99.07	0.04	0.43	- 0.19	0.47
ϵ_{ele}	90.39	0.12	0.89	+ 0.98	1.33

Table 7.5: Global reconstruction (ϵ_{reco}), medium identification (ϵ_{id}) and trigger (ϵ_{trig}) efficiencies extracted with the Tag-and-Probe method ($Base\,TP$ Configuration) with statistical and systematic errors. The last line shows the overall electron efficiency, given by the product of ϵ_{reco}, ϵ_{id} and ϵ_{trig}. Shown separately are systematic errors arising from background subtraction ($\Delta\epsilon^{\text{Bkg}}_{syst}$) and from the difference between TP and true W efficiencies ($\Delta\epsilon^{\text{W}}$). The last column shows the overall error (Tot $\Delta\epsilon$), quadrature sum of statistical and systematic errors.

already an overestimation.

The global efficiencies (integrated over η and p_T, weighted by the number of $Z \to ee$ probes in each bin) are shown in Table 7.5 with statistical and systematic errors. The systematic errors shown in the second column ($\Delta\epsilon^{\text{Bkg}}_{syst}$) arising from background subtraction, are inherent to the TP method itself. However, when applying these efficiencies to a different process, such as $W \to e\nu$ + jets, other biases have to be accounted for (shown in the third column $\Delta\epsilon^{\text{W}}$). This is the subject of the next section.

7.3.3 Systematic biases from Tag-and-Probe

Electrons from Z and W boson decays are very similar objects, isolated high p_T electrons, and therefore there is no reason why they should have different efficiencies. However, because of the slightly different masses between Z and W boson, and because of production mechanisms involving different partons, Z and even W^+ and W^- exhibit different p_T and η spectra (see Fig. 7.10). Convoluted with efficiencies that depend on η and p_T (and potentially charge), this will result in different *global* efficiencies. This is the main reason why efficiencies have to be extracted differentially. If the binning is fine enough, no difference between Z and W true electron efficiencies should be visible. In practice, given the statistical limitations, very fine

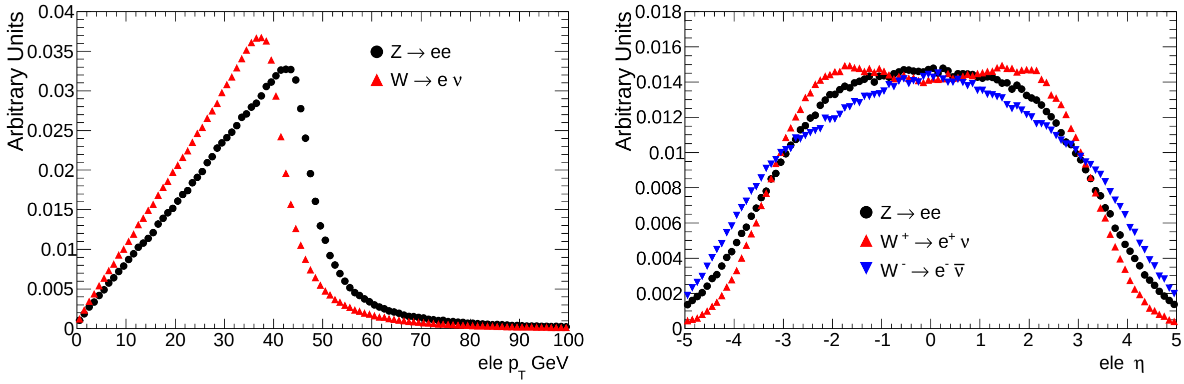

Figure 7.10: p_T (left) and η (right) spectra of W and Z bosons.

binning is not feasible and some residual differences still remain. W production asymmetry is already taken care of by extracting efficiencies as a function of electron η, provided that they are not charge sensitive. Efficiencies from $Z \to ee$ are symmetric in charge, thus no distinction is made between e^+ and e^- when extracting TP efficiencies.

Additional biases can arise from the TP method itself. It was discussed in Chapter 6, how the TP selection tends to purify the $Z \to ee$ sample and select better quality probes. Many of the rejected probes are background, but some are real Z boson electrons showering in the ID material. Any difference in the composition of the electrons in the $W \to e\nu$ signal sample with respect to the probe sample in $Z \to ee$ will therefore lead to different efficiencies. Fig. 7.11 compares W true identification efficiencies with two different configurations of the TP efficiencies (the $BaseTP$ and the $\pm TP$ configurations, see Section 6.3). TP efficiencies are systematically higher than W true efficiencies, particularly using the $\pm TP$ configurations which selects a sample with too high purity with respect to the $W \to e\nu$ sample. Adopting 2-dimensional binning in η and p_T accounts for kinematic differences between electrons from Z and W boson decays, but not for selection biases.

An additional systematic uncertainty is therefore assigned based on the difference between true W and TP efficiencies: $\Delta\epsilon^{\mathrm{W}} = \epsilon_Z^{TP} - \epsilon_W^{MC}$. True efficiencies are calculated matching the true W boson electron to a cluster/electron and then calculating ϵ_{reco}, ϵ_{id}, ϵ_{trig} relative to this object, in a similar way to the TP efficiencies. The resulting systematic uncertainties are

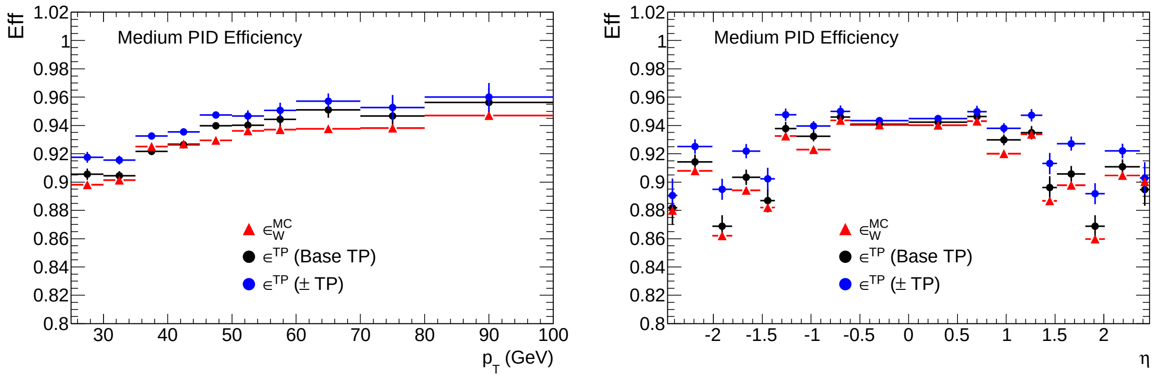

Figure 7.11: Comparison of MC efficiencies $\epsilon_{\rm id}$ extracted from $W \rightarrow e\nu$ events with Tag-and-Probe efficiencies from $Z \rightarrow ee$ events as a function of electron η and p_T.

estimated on the global efficiency and are shown, for the $Base\,TP$ configuration and inclusive number of jets [2], in the third column of Table 7.5. Medium PID and reconstruction efficiencies are the largest source of uncertainties, since they are also the most sensitive to the electron sample composition. Since the trigger efficiency is computed after medium PID cuts, most bad quality electrons are rejected and thus Z and W boson electron samples are very similar. The $\pm TP$ configuration yields 1% higher $\epsilon_{\rm id}$ efficiencies while leaving $\epsilon_{\rm trig}$ unchanged and would therefore lead to an overestimation of the overall efficiency of $\Delta\epsilon^{\rm W} = 1.89\%$, to be compared to the value of 0.98% for the $Base\,TP$ (see last line of Table 7.5). For this reason in this particular analysis, the $Base\,TP$ configuration is preferred over the $\pm TP$ configuration.

For $\epsilon_{\rm reco}$, there is one additional detail to be considered: $\epsilon_{\rm reco}$ is computed with respect to cluster level and therefore it is equivalent to the offline electron reconstruction efficiency only if an electron is always reconstructed as a cluster. The efficiency for a true electron within the acceptance to be reconstructed as a cluster is found to be $(99.998 \pm 0.001)\%$, therefore no additional correction is needed to account for cluster reconstruction inefficiencies.

[2] $\Delta\epsilon^{\rm W}$ is also estimated separately for each jet multiplicity. This jet-multiplicity-dependent uncertainty is used in Section 7.4.

7.4 Closure Test on A and ϵ Corrections

This section describes how the A and ϵ correction factors calculated in previous sections are used to extract the cross-section. We assume at first no background contamination and compare production rates after all jet selection cuts have been applied. Specifically, we calculated the following term:

$$N_{W+Njets} = \sigma_W \cdot \mathrm{BR}(\mathrm{W} \to \mathrm{e}\nu) \cdot \mathrm{L} \cdot (\mathrm{A} \cdot \epsilon)_{\mathrm{jets}} = \frac{\mathrm{N} - \mathrm{B}\,(=0)}{(\mathrm{A})^{\mathrm{jet}}_{\mathrm{E_T^{miss}}} \cdot (\mathrm{A} \cdot \epsilon)^{\mathrm{jets+E_T^{miss}}}_{\mathrm{ele}}} = \frac{\mathrm{N}}{\mathrm{A} \cdot \epsilon} \tag{7.6}$$

where $N_{W+Njets}$ are the true number of signal events and $N - B\,(=\ 0)$ are the measured number of events. A and ϵ are redefined as the acceptance and efficiency factors on E_T^{miss} and electrons with respect to reconstructed jets, as discussed throughout this chapter. This is a purely academic exercise to verify that A and ϵ have been correctly factorised and identify potential biases. Performing such a closure test on MC is essential to ensure that the procedure is self-consistent and can be safely applied to LHC data when this will be available. To extract the final cross-section, jet corrections will have to be applied and the final yield then divided by the integrated luminosity.

In the case of this measurement, with multi-dimensional efficiency and/or acceptance maps, the formula in eqn. (7.6) is generalised as:

$$N_{W+Njets} = \sum_i^{N_A} \sum_j^{N_\epsilon} \frac{N_{ij}}{A_i \cdot \epsilon_j} \tag{7.7}$$

where A and ϵ are binned in different dimensions (eg: $\epsilon\,:\,\eta^{ele},\,p_T^{ele};\;A\,:\,p_T^W,\,p_T^{leading\ jet}/HT$) and the indices $i\,,j$ run over N_A (number of A bins) and N_ϵ (number of ϵ bins) respectively.

The statistical uncertainties are:

$$\Delta^2_{stat} = \sum_i^{N_A}\sum_j^{N_\epsilon}\frac{\Delta^2 Nij}{(A_i\cdot\epsilon_j)^2} \tag{7.8}$$

$$\Delta^{A,\epsilon\;2}_{stat} = \sum_i^{N_A}\Big[\Big(\sum_j^{N_\epsilon}\frac{N_{ij}}{A_i\epsilon_j}\Big)^2\cdot\Big(\frac{\Delta A_i}{A_i}\Big)^2\Big]+\sum_j^{N_\epsilon}\Big[\Big(\sum_i^{N_A}\frac{N_{ij}}{A_i\epsilon_j}\Big)^2\cdot\Big(\frac{\Delta\epsilon_j}{\epsilon_j}\Big)^2\Big] \tag{7.9}$$

where the first term (Δ_{stat}) accounts for the statistical errors on the number of signal events and the second term ($\Delta^{A,\epsilon}_{stat}$) for the statistical errors on A and ϵ. The latter is propagated as a systematic uncertainty on the cross-section and is added in quadrature to the other sources of systematic uncertainties which are estimated globally [3] (see Tables 7.3-7.5):

$$\Delta^2_{syst} = \Big(\Delta^{A,\epsilon}_{stat}\Big)^2+\Big(\sum_i^{N_A}\sum_j^{N_\epsilon}\frac{N_{ij}-B_{ij}}{A_i\cdot\epsilon_j}\Big)^2\cdot\Big[\Big(\frac{\Delta A_{syst}}{A}\Big)^2+\Big(\frac{\Delta\epsilon^{\rm Bkg}_{syst}}{\epsilon}\Big)^2+\Big(\frac{\Delta\epsilon^{\rm W}}{\epsilon}\Big)^2\Big]. \tag{7.10}$$

The closure test is performed as follows:

- Reference: the full statistics of the $W\to e\nu$ + N partons ALPGEN sample is passed through the jet selection cuts only and is divided into sub-samples of N_{jets}. Yields are normalised to 100 pb^{-1} of integrated luminosity. This will be the reference sample the pseudo-data is compared to.
- ϵ: electron efficiencies are extracted using the TP method from a pseudo-data sample of $Z\to ee$ + jets events equivalent to 100 pb^{-1} of data. A two dimensional efficiency map in η- p_T is extracted.
- A: the full statistical power of the MC $W\to e\nu$ ALPGEN sample is used to derive the $E_T^{\rm miss}$ and electron acceptance factors binned in p_T^W- leading jet p_T/ HT (for $N_{jets}=0$ a single dimension, p_T^W, is used).
- Pseudo-experiment: a pseudo-data sample of $W\to e\nu$ + jets equivalent to 100 pb^{-1} of luminosity is passed through the full selection cuts and is divided into sub-samples according

[3] The systematic uncertainty on the acceptance from $E_T^{\rm miss}$ corrections is not considered for this test as it is not relevant, given that the $E_T^{\rm miss}$ correction factor is taken from MC.

to the jet multiplicity. For a N_{jets} sub-sample, a multidimensional histogram, with dimensions and binning determined by the ϵ and A maps, is filled. Each bin is then weighted by the corresponding $\frac{1}{A \cdot \epsilon}$ and the sum over the bins gives the overall yield for the given $N_{W+Njets}$.

Results and Discussion

The corrected yields for 100 pb^{-1} of integrated luminosity are reported in Table 7.6. Shown are the MC reference and the pseudo-data corrected for electron and E_T^{miss} acceptance and efficiencies, using two dimensional maps. The $Base\,TP$ configuration is used for extracting the efficiencies and the systematic error therefore include the term $\Delta\epsilon^{\mathrm{W}}$.

The deviation from the MC truth is less than 1% and within the statistical plus systematic errors, demonstrating that the procedure is self-consistent. Typically, corrected yields are underestimated (with the exception of the 0 jet bin), as expected by the fact that data-driven efficiencies are larger than MC efficiencies. The measurement is dominated by the systematic uncertainties, particularly for the 0-jet bin where statistical errors are below the 0.2% level. Electron efficiencies (ϵ_{id} and ϵ_{reco} in particular) are the largest sources of uncertainty.

To check the stability of the results, two variations are investigated:

- Global rather then binned acceptance and efficiency corrections are applied. Acceptances are determined separately for each jet multiplicity bin.
- The $\pm TP$ (rather than the $Base\,TP$) configuration is used to extract the efficiencies.

Table 7.7 summarises the results by comparing the deviation from the MC truth using global factors and the two TP configurations. The difference between global data-driven and true efficiencies ($\Delta\epsilon^{\mathrm{W}}$) is also shown in each case. It is observed that:

- Using binned rather than global factors improves the agreement with the MC truth. This is particularly visible in the 0-jet and 1-jet bins, where the deviations from MC are significantly reduced.

- Using the $\pm TP$ configuration, predicted yields are systematically underestimated by $1-2\%$. This is expected since the $\pm TP$ efficiencies are systematically higher than the true efficiencies, therefore leading to an under-correction of the number of events. The bias is still present even using 2-dimensional efficiency maps, demonstrating that it is not caused by differences in the W and Z bosons η and p_T spectra, but rather due to the quality of the electron sample. To account for this bias it is therefore necessary to include the systematic uncertainty $\Delta\epsilon^{\mathrm{W}}$.

- The systematic uncertainty $\Delta\epsilon^{\mathrm{W}}$ is significantly lower in the 1 and 2-jet bins with respect to the 0-jet bin for both $\pm TP$ and $Base\,TP$ configurations. In fact, as the W boson is boosted, the electron p_T spectrum is shifted to higher p_T bins and the kinematic differences between electrons from W and Z bosons are less important.

- Even for global efficiencies, the shift $\Delta_{\mathrm{data-MC}}$ is typically smaller than the value of $\Delta\epsilon^{\mathrm{W}}$. The reason is that a small correction factor in the trigger efficiency is missing and this counterbalances to a certain extend the fact that $\epsilon_Z^{TP} > \epsilon_W^{MC}$. In fact, the efficiency for an event to pass the trigger is slightly higher than the object trigger efficiency [4]. The factor ϵ_{trig} only calculates the probability that the candidate W electron fires the trigger, but other poorer quality electrons or jets in the event could fire the trigger too (although this is much less likely). The difference between object and event-based trigger efficiency is estimated to be $\sim 0.3\%$. Since this factor is much smaller than the total systematic uncertainty (> 1%), it can be neglected for the time being.

[4] The object trigger efficiency is the probability that one object, such as an electron, fires the trigger and is computed requiring matching between trigger and offline object.

N_{jets}	MC Reference	Pseudo-Data $\pm\Delta_{stat} \pm \Delta_{syst}$	Δ_{stat} (%)	Δ_{syst} (%)	$\Delta_{\rm data-MC}(\%)$
0	1186050 ± 953	$1188751 \pm 1871 \pm 18005$	0.16	1.51	0.23
1	107837 ± 245	$107735 \pm 565 \pm 1200$	0.52	1.11	-0.09
2	22066 ± 52	$21898 \pm 244 \pm 250$	1.11	1.14	-0.76

Table 7.6: Corrected yields for 100 pb^{-1} of data for $W \to e\nu$ + 0, 1, 2 jets. Shown is the MC reference, obtained using the full ALPGEN sample, and the pseudo-data after $E_T^{\rm miss}$ and electron acceptance and efficiency corrections. The errors on the MC reference are purely statistical. For pseudo-data, both statistical and systematic errors are shown in their absolute and relative (fourth and fifth columns) values. $\Delta_{\rm MC-data}$ (last column) is the deviation of the predicted value from the MC truth.

	$BaseTP$ Config.		$\pm TP$ Config.	
N_{jets}	$\Delta_{\rm data-MC}(\%)$	$\Delta\epsilon^{\rm W}(\%)$	$\Delta_{\rm data-MC}(\%)$	$\Delta\epsilon^{\rm W}(\%)$
0	-0.57	1.12	-1.57	1.81
1	0.36	0.39	-0.65	0.97
2	-0.08	0.45	-1.09	1.04

Table 7.7: Closure test using global correction factors for acceptance and efficiencies, using two different TP configurations for the extraction of data-driven efficiencies: the default $BaseTP$ (left) and $\pm TP$ (right). For each of them the relative deviation (in $\%$) between true and corrected number of events ($\Delta_{\rm data-MC}$) is shown, as well as the difference between TP and true W efficiencies ($\Delta\epsilon^{\rm W}$).

7.5 Conclusions

This chapter showed how event selection cuts are corrected for by acceptance and efficiencies when performing a measurement. The approach adopted was to use data-driven methods whenever possible to minimise the dependence on MC. Factors that can be entirely extracted from data (efficiencies) were decoupled from the corrections for kinematic cuts that must be extracted from MC (generator-level acceptance). Corrections for generator-to-detector-level acceptance were estimated using a smearing technique, relying on the fact that electron and $E_T^{\rm miss}$ scale and resolution will also be determined in-situ.

Experimental systematic uncertainties on the acceptance were evaluated. Uncertainties on smearing corrections on electrons were observed to be negligible compared to those on $E_T^{\rm miss}$. Theoretical uncertainties on the generator-level acceptance were not considered but should be studied in the future. Systematic uncertainties on the efficiencies were found to be at the 1.5% level (relative), larger than experimental uncertainties on the acceptance (0.8%, relative error). Uncertainties on the efficiency are dominated by the uncertainty on $\epsilon_{\rm reco}$ and $\epsilon_{\rm id}$. A significant contribution to this uncertainty is due to the difference between TP and true efficiencies and is therefore intrinsic to the TP method.

A closure test for acceptance and efficiency corrections was performed. It was shown that the full procedure is consistent and that the true number of corrected events can be recovered within uncertainties. Some conditions were crucial for the accuracy of the result:

- All correction factors should be binned in the variables they mostly depend upon. Applying global rather than binned correction factors leads to larger discrepancies between truth and corrected yields.
- Each step (acceptance at generator, detector level and efficiencies) should be factorised in a consistent way. Furthermore, in order to account for detector smearing on the electron p_T, a merging between final state electrons and QED photons has to be performed.
- When applying efficiencies extracted from $Z \to ee$ events to $W \to e\nu$ events, care has to be taken on the TP configuration to be used. A selection which requires tag and probe electrons to have opposite charge, for example, leads to a sample which is too pure and therefore overestimates the efficiency of $W \to e\nu$ events by 1-2%.

It has been shown that differences between TP and MC efficiencies result in an additional systematic uncertainty, which would be then propagated to the cross-section. Choosing the optimal TP configuration is therefore important to minimise the overall systematic uncertainty on the cross-section. An alternative solution would be to use TP efficiencies only to validate the MC and, if needed, extract a data-derived correction factor. For the cross-section measurement, MC efficiencies could be used, with this correction factor. This approach should be preferred when TP efficiencies are significantly different from true efficiencies.

Chapter 8

Prospects for the Measurement of the $W + N_{jets}$ Cross-Section and Ratios

The previous chapters described techniques for event selection, background estimation and extraction of efficiency and acceptance corrections in the context of the $W \rightarrow e\nu$ + jets cross-section measurement. Several topics were covered, with particular emphasis on electron related issued (from backgrounds, to efficiencies and detector smearing). This chapter brings all those topics together. The objective is to assess the feasibility of the measurement of the absolute $W \rightarrow e\nu$ + N_{jets} cross-section for the 1-2 jet bins with the first 100 pb^{-1} of ATLAS data, as well as to identify ratio measurements of vector boson V + jets which are robust against the largest sources of systematic uncertainties. The first part of the chapter (Section 8.1) is dedicated to the absolute cross-section measurement. The expected systematic uncertainties arising from the sources considered in previous chapters and other sources (such as jet scale and resolution) are discussed. In the second part (Section 8.2), two ratio measurements are presented (R_W and R_{jets}) and the advantages of measuring ratios over absolute cross-sections are described.

8.1 Measuring the $W \rightarrow e\nu + N_{jets}$ Cross-Section

The absolute cross-section for $W \rightarrow e\nu$ produced in association with N hadronic jets (σ_N) can be calculated as:

$$\sigma_N = \sigma(W \rightarrow e\nu + N_{jets}) = \frac{1}{L} \cdot \frac{N_n - B_n}{A_{Gen \rightarrow Rec}[N_{jets}] \cdot \epsilon_{\text{jet}} \cdot A_{Rec}[E_T^{\text{miss}} + ele] \cdot \epsilon_{\text{ele}} \cdot \epsilon_{veto}^{lep}} \quad (8.1)$$

where N_n and B_n are the number of candidate and background events with $n = N_{jets}$ reconstructed jets and L is the integrated luminosity. The efficiency/acceptance terms were defined in Chapter 7, eqn. (7.2), and the variable ϵ_{veto}^{lep} is the efficiency of the lepton veto cuts (track veto and/or EM fraction veto, see Section 5.2.1), *if* applied. It should be noted that the cross-section is for jets in the phase-space: $|\eta_{jet}| < 3.1$ and $p_{T\,jet} > 30$ GeV (see Section 7.1 in Chapter 7). The cross-section σ_N is extracted as follows:

- W candidate events are selected as described in Chapter 4 and divided in sub-samples according to the number of jets, thus giving N_n.

- The number of W + N_{jets} events after background subtraction is determined ($N_n - B_n$), according to eqn. (5.11) in Chapter 5: first background from QCD-jets is estimated using data-driven techniques, then contributions from single top, $t\bar{t}$ and dibosons are subtracted (based on MC estimates) and the final yield is corrected for $Z \rightarrow \ell\ell$ and $W \rightarrow \tau\nu$ contaminations.

- Yields are corrected for all inefficiencies due to event selections on electrons and E_T^{miss} (terms $A_{Rec}[E_T^{\text{miss}} + ele] \cdot \epsilon_{\text{ele}}$). Efficiencies and acceptance are extracted from data or MC as described in the Chapters 6 and 7. An example of how to apply such corrections was also given in Chapter 7. If lepton veto cuts are applied to reduce the background, these also have to be corrected for by an efficiency (ϵ_{veto}^{lep}) which can be estimated from MC.

- Jet corrections have to be applied. Jet reconstruction efficiencies, jet energy scale and

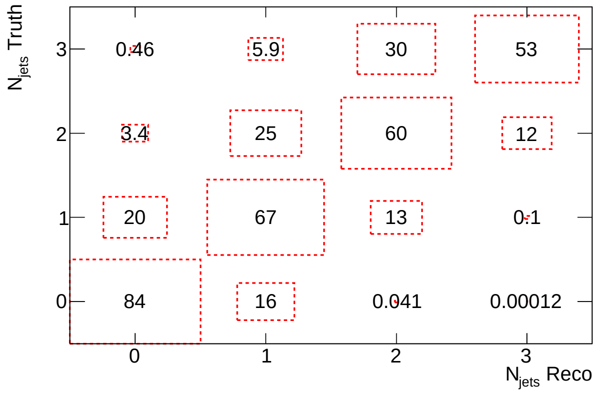

Figure 8.1: Migration matrix mapping *W + N true jets* to *W + N reconstructed jets* events. Events are normalised to the number of truth jets, ie. each row of the matrix is normalised to 100% if the bin N_{jets} $Reco \geq 4$ (overflow bin) is considered.

resolution cause migrations between jet multiplicity bins: an event with n true jets can be reconstructed as an m-jet event with $m \neq n$. Figure 8.1 shows an example of a migration matrix obtained with fully simulated ALPGEN samples: the x- and y-axis are the number of reconstructed and true (particle level) jets respectively and each row of the matrix is normalised to 100%. Although most events lie on the diagonal, less than 70% of events with $N > 0$ truth-jets have the same number of reconstructed jets. Bin migrations are more significant to lower multiplicity bins and are mainly due to jet energy scale and resolution [1]. In order to unfold reconstructed- to truth-level jets (ie. account for the terms $A_{Gen \rightarrow Rec}[N_{jets}] \cdot \epsilon_{\text{reco jet}}$ [2]), the migration matrix has to be inverted. Methods for unfolding W and Z + jets events with matrix inversion were investigated in [3] and have shown that true jet multiplicity distributions can be recovered within 2% precision. After unfolding, event-level inefficiencies due to the ΔR_{ej} and ΔR_{jj} requirements, are corrected for by event weights (eg. $1/\epsilon_{\Delta \text{Rej}}$, $1/\epsilon_{\Delta \text{Rjj}}$).

The cross-section obtained would be defined at truth-level jets (hadron level) and could therefore be compared to predictions provided by MC generators like ALPGEN (interfaced with

[1] The jet reconstruction efficiency is around 99.5 % and can therefore not explain migrations of the order of 30-40%

[2] Recall from Chapter 7 that $\epsilon_{\text{jet}} = \epsilon_{\text{reco jet}} \cdot \epsilon_{\Delta \text{Rej}} \cdot \epsilon_{\Delta \text{Rjj}}$. Unfolding accounts only for the jet reconstruction efficiency $\epsilon_{\text{reco jet}}$.

Quantity	$N_{jets} = 1$	$N_{jets} = 2$	Comment
			Relative Uncertainties on σ_N (%)
Statistical			
$N_n - B_n$	0.52	1.27	n/a
Systematics			
$N_n - B_n$	0.27	1.52	Background from QCD-jets excluded
$A_{Rec}[E_T^{\text{miss}} + ele]$	0.88	0.88	Experimental uncertainties (E_T^{miss} and ele smearing)
ϵ_{ele}	1.07	1.09	From TP efficiencies
L	10	10	Expected uncertainty for early data

Table 8.1: Statistical and systematic uncertainties on the absolute W + 1, 2 jets cross-section measurement from the sources considered in previous chapters.

HERWIG/JIMMY) and PYTHIA. For comparisons with NLO theoretical predictions at parton level provided by MCFM, additional corrections for non-perturbative effects (fragmentation and underlying event) should also be applied.

In preparation for the measurements which will take place in the coming months or years, we investigate the experimental prospects for measuring σ_N. The focus is on sources of experimental systematic uncertainties. For comparison, one of the largest sources of theoretical systematic uncertainties, parton distribution functions (PDFs), is also investigated.

8.1.1 Experimental Systematic Uncertainties

Each term in equation 8.1 has an associated uncertainty, which contributes to the overall statistical/systematic uncertainty as shown in eqn. (3.3). Table 8.1 summarises the relative uncertainties (in $\%$) from the sources considered in the previous chapters:

- $N_n - B_n$: this term contributes to the statistical, as well as to the systematic uncertainty. The systematic uncertainty is determined by the uncertainty in the background estimation, which includes here only EW backgrounds. For the 2-jet bin, a veto on isolated tracks is applied. Given the poor MC statistics, background from QCD-jets is not included. Furthermore, no uncertainty on the detector simulation is included, since this has yet to be assessed. Therefore, although the values shown in Table 8.1 are comparable to the statistical uncertainty, in data the impact of the background is expected to be much larger, depending on the contamination of QCD-jets and the reliability of the MC simulation.

- $A_{Rec}[E_T^{\text{miss}} + ele]$: the overall systematic uncertainty on the acceptance is less than 0.9%, considering only sources of experimental uncertainty. Corrections for E_T^{miss} smearing (see [71] and Section 7.2.3) have the largest impact (0.8%).

- ϵ_{ele}: uncertainties from electron efficiencies are at the level of 1% and are the largest source of uncertainties considered so far for the 1-jet bin. They are dominated by uncertainties on ϵ_{id}.

- L: the luminosity is measured at ATLAS with several forward detectors (LUCID, ALFA Roman Pots, ZDC [15]), which, however need to be calibrated. As of June 2010, the luminosity was measured with an uncertainty of 11% [76], but this is expected to improve in the future. In this thesis a 10% uncertainty is assumed on the integrated luminosity.

Uncertainties from Jet Energy Scale and Resolution

We now turn our attention to the effects of jet energy resolution and scale on the cross-section measurement. In this section we treat each source of uncertainty as an unknown bias and evaluate its impact on the cross-section assuming that no corrections (such as unfolding) are applied. Although from previous discussions (see Fig. 8.1), it is expected that the effect will be large, this is still a useful exercise to isolate potentially large sources of uncertainty. The following are considered:

- **Jet Energy Resolution (JER)**: the energy of a jet, as measured in the calorimeter, fluctuates on a jet-by-jet basis with respect to the true energy. The width of this fluctuations is the energy resolution $\sigma(E)$, which can be parametrised as $\frac{\sigma(E)}{E} = \frac{a}{\sqrt{(E)}} + b + \frac{c}{E}$ where: a, called the stochastic term, is the calorimeter intrinsic resolution due to fluctuations of the showers in different samplings; b is the constant term and reflects the effect of the calorimeter non-compensation and non-uniformities in the energy measurement; c is the noise term and parametrises electronic noise from the read-out chain. c dominates at very low energies, b at high energies and a at middle energies. The ATLAS hadronic calorimeter (neglecting c) has: $\frac{\sigma(E)}{E} = \frac{50\%}{\sqrt{(E)}} + 3\%$ (the units for E are in GeV).

- **Jet Energy Scale (JES)**: the measured energy of a jet can also be systematically shifted with respect to the true energy. The absolute jet energy scale can be restored by proper calibration and corrections as described in Chapter 4, Section 4.4.1. Such corrections, however, are affected by an uncertainty which will be reflected in an uncertainty on the cross-section. Although the ATLAS goal is to achieve 1% precision on the jet energy scale, with early data this uncertainty is likely to be around $5 - 10\%$ [3].

The impact of scale and resolution are evaluated independently by shifting or smearing the true energy of a jet. For the resolution, the jet energy is obtained from a random number generated from a Gaussian distribution with mean at the true jet energy and a width of 0.10. This is equivalent to assuming a constant term $b = 10\%$, which is a very conservative estimate and dominates the energy dependence of the resolution for jets with $p_T > 30$ GeV. For the energy scale, the energies of all jets are shifted by a given value ($\pm 5, \pm 10\%$), which corresponds to the experimental uncertainty on the JES. In each case, the jet p_T is recomputed and jet counting is performed again. The relative uncertainty on the cross-section is given by the difference in the number of signal events before ($N_N^W(Ref)$) and after the shift or smearing ($N_N^W(JER/JES)$):

$$\frac{\Delta\sigma_N}{\sigma_N} = \frac{N_N^W(JER/JES) - N_N^W(Ref)}{N_N^W(Ref)}. \tag{8.2}$$

Table 8.2 shows the impact on of scale and resolution on the absolute $W \rightarrow e\nu$ + 1 and 2 jets cross-section. In Figure 8.2, the effect of a 5% and 10% shift in the jet energy scale are compared. The resulting uncertainties are significantly larger than any source of uncertainty

[3] Recent studies [77] have estimated the jet energy scale uncertainty to be about 9 (7) % for jets with $p_T > 20(60)$ GeV and $|\eta| < 2.8$. These results have been obtained for inclusive jets using the algorithm anti-k_T with parameter D=0.4.

Systematic Uncertainties on σ_N (%)		
Sources of systematics	$N_{jets} = 1$	$N_{jets} = 2$
Jet energy resolution 10%	0.8	5.0
Jet energy scale 5%	6.0	8.0
Jet energy scale 10%	12.0	16.0

Table 8.2: Systematic uncertainties from jet energy scale and resolution on the W + 1, 2 jets cross-section measurement, obtained by smearing/shifting the jet energy.

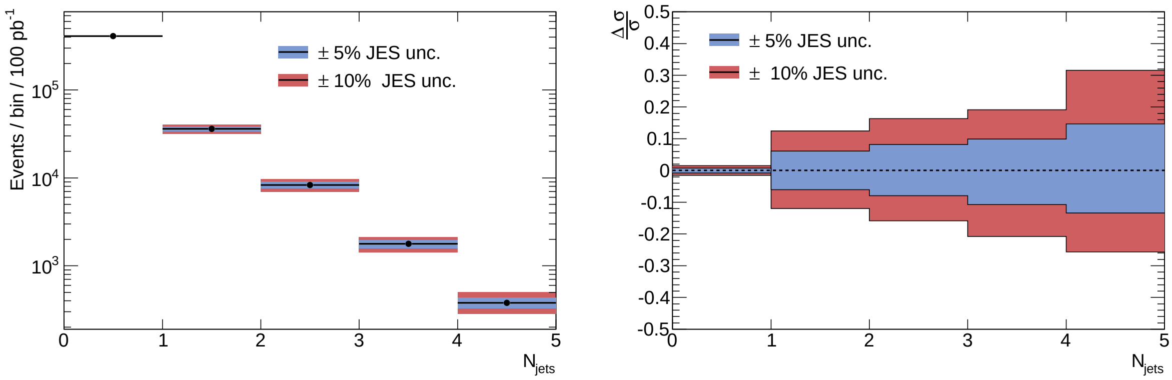

Figure 8.2: (Left) W + jets cross-section as a function of jet multiplicity with systematic uncertainties arising from 5 and 10 % uncertainty on the jet energy scale (JES). The right plot shows the relative uncertainties on the cross-section.

considered in Table 8.1 (excluding the luminosity). In particular, an uncertainty on the jet energy scale of 10% has an effect on $\sigma_N(N_{jets} = 2)$ as large as 16%. This uncertainty cannot be reduced by further corrections (such as unfolding), but only by improving calibrations which will require time. This is therefore expected to be the largest source of experimental systematic uncertainty in early data.

Smearing due to energy resolution (which, if uncorrected for, would result in the uncertainty reported in the first line of Table 8.2) can be corrected for by unfolding, provided that the detector resolution function is known. The unfolding procedure, however, will also be affected by an uncertainty and, in addition, any uncertainty on the energy resolution function itself will result in an uncertainty on the measured cross-section. Nonetheless, even a 10% uncertainty on a resolution of 10% will have a negligible impact compared with the effect of a 5-10% uncertainty on jet energy scale.

8.1.2 Theoretical Uncertainties on Parton Distribution Functions

An important source of theoretical systematic uncertainties at the LHC are the Parton Distribution Functions (PDFs). PDFs are part of predictions for any cross-section at hadronic colliders. Thus, assessing the level of PDF uncertainties is crucial for studying both processes within and beyond the SM.

The CTEQ and MRST groups provide PDF errors estimated with a Hessian technique [78], whereby a matrix of dimension N x N (where N is the number of free parameters used in the PDF fits) is diagonalized, resulting in a basis of N eigenvectors. By varying each eigenvector along the positive/negative direction a PDF error set is obtained. The CTEQ group provides 40 PDF error sets (for 20 free parameters in the fit) for the NLO PDF set CTEQ6M. A rigorous estimation of PDF uncertainties would require generating a large number of events for each PDF set, ie. 41 times for the 40 PDF errors plus the central value. This is a very expensive process in terms of CPU time and storage space. PDF reweighting is a powerful technique which overcomes this problem, allowing to estimate uncertainties using only one sample of events generated with a conventional PDF set. Predictions for any other PDF set are calculated by applying a weight to the hard process on an event-by-event basis. The event weight (EW) is a function of momentum fraction x of the two incoming partons and the scale Q^2 of the interaction:

$$EW = \frac{f_{pdf_2}(x_1, Q^2)}{f_{pdf_1}(x_1, Q^2)} \frac{f_{pdf_2}(x_2, Q^2)}{f_{pdf_1}(x_2, Q^2)} \tag{8.3}$$

where the scale Q^2 should be chosen consistently with the definition used in the generation [4].

In this thesis PDF reweighting is applied to the ALPGEN W + jets samples to investigate the effect of PDF uncertainties on physical distributions. The studies presented in this section were performed with samples generated with a centre of mass energy $\sqrt{s}$ = 14 TeV, but no significant variations are expected for a lower centre of mass energy of 10 TeV [79]. As mentioned above,

[4]For the ALPGEN W + jets samples used in this studies, the scale was calculated from the mass (M^W) and transverse momentum (p_T^W) of the W boson as $Q^2 = (M^W)^2 + (p_T^W)^2$.

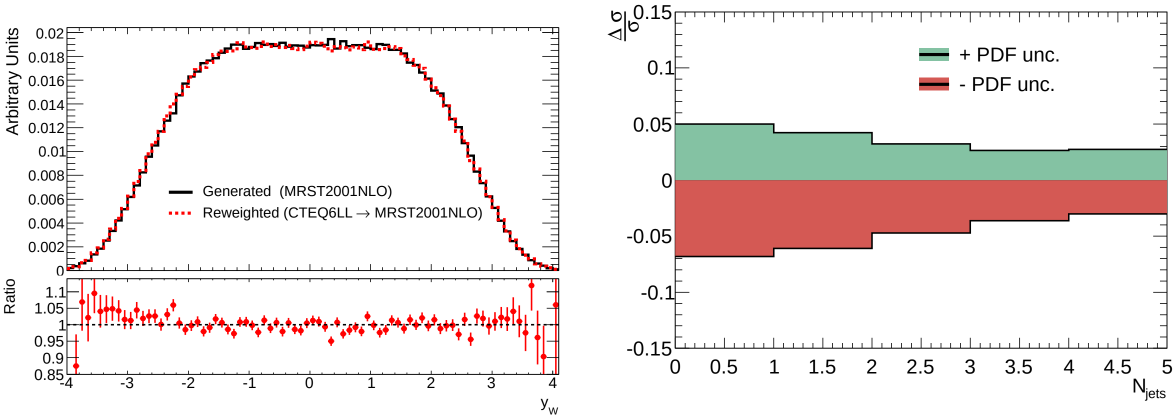

Figure 8.3: (Left) Test of PDF reweighting technique: comparison between the W boson rapidity (y) distribution obtained by directly generating events with the PDF set MRST2001NLO and by reweighting events generated with CTEQ6LL. The lower plot shows the ratio of the distributions. (Right) Relative error on the W + jets cross-section arising from PDF uncertainties given by the CTEQ6M set. The upper (+ PDF unc.) and lower (- PDF unc.) uncertainties are obtained by combining the positive/negative variations of the 20 eigenvectors, as described in [26].

the CTEQ group provides error sets for the NLO PDF CTEQ6M, whereas the ALPGEN datasets were generated with the LO PDF set CTEQ6LL. It is therefore necessary to reweight the sample from a LO to a NLO PDF.

The accuracy of PDF reweighting on kinematic distributions for events generated with HERWIG was investigated in previous studies [80] and found to be better than 1% (see below). A similar analysis was performed on ALPGEN events to test the accuracy of reweighting from a LO to a NLO PDF set. Two samples of 500,000 W +3 parton events were generated with two different PDF sets (CTEQ6LL and MRST2001NLO). Events generated with CTEQ6LL were then reweighted to MRST2001NLO and compared to those generated directly with MRST2001NLO. Figure 8.3 (left) shows such a comparison for the W boson rapidity distribution: the agreement is good, particularly in the central region. By considering different observables, it was shown that the PDF reweighting technique is accurate to better than 1% in the measurable kinematic regions, sufficient for its applicability to this study.

The ALPGEN samples were therefore reweighted from CTEQ6LL to the central value of CTEQ6M and the 40 error sets. The upper and lower uncertainties around the NLO central value are calculated using the asymmetric error formula recommended in [26].

The relative uncertainties on the jet multiplicity spectrum, shown in the right plot of Fig. 8.3, are below 7% for all bins . For the 1- and 2-jet bin, in particular, they are between $3-6\%$, lower than jet scale uncertainties arising from an energy miscalibration of 5 or 10%. In order to gain sensitivity to the PDFs and improve the PDF fits, an uncertainty on the JES of 1-2% is required. Although this is not feasible for the first year of running, it is a potentially interesting study in the longer term.

8.2 Ratio Measurements

In Section 8.1, the feasibility of measuring the absolute $W \to e\nu$ + 1, 2 jets cross-section was investigated. It was shown that the measurement is totally dominated by systematic uncertainties on jets. An uncertainty of 10% on the jet energy scale, in particular, is expected to result in an uncertainty as large as 16% on the absolute cross-section. Uncertainties from electron efficiencies and experimental uncertainties on the acceptance are an order of magnitude smaller (around the 1% level). These observations show the need for designing a measurement of W +jets which is robust against jet uncertainties. Ratio measurements are an obvious choice as they lead to cancellation of several systematic uncertainties. In this section, two ratio measurements are investigated:

- $R_W = \dfrac{\sigma(W + N_{jets} + 1)}{\sigma(W + N_{jets})}$ where $N_{jets} = 1$
- $R_{jets} = \dfrac{\sigma(W + N_{jets})}{\sigma(Z + N_{jets})}$ where $N_{jets} = 1, 2$.

One first obvious source of uncertainty that cancels in both ratios is the luminosity. In addition, one would expect some detector effects to almost cancel and others to be significantly reduced. Some examples are detector response on electrons and missing transverse energy in the R_W ratio and jet energy scale/resolution in the R_{jets} ratio. When designing a ratio measurement it is necessary to estimate to what extent detector effects cancel in the ratio and compare these to the systematic uncertainties arising from corrections for such detector effects. In this section, the following approach is adopted: detector effects, such as reconstruction efficiency or energy scale and resolution on electron/E_T^{miss}/jets, are considered as uncorrected biases. For example, the impact of electron PID efficiencies on R ($R = R_W$ or $\mathrm{R_{jets}}$) can be estimated by computing R on reconstructed quantities *before* and *after* applying a medium PID cut. The bias from the electron PID cuts can be considered as a systematic uncertainty. In general, the effect of any cut on the ratios is evaluated as:

$$\frac{\Delta R}{R} = \frac{R^{After\ Cut} - R^{Before\ Cut}}{R^{Before\ Cut}}. \tag{8.4}$$

The value of ΔR is then compared to the systematic uncertainty on R that would derive from applying a correction for such a cut. In the example of the electron medium PID, the correction is the ϵ_{id} with its associated uncertainty. If the uncertainty on the correction is larger than the bias ΔR introduced by the cut itself, it might be desirable to leave the ratio R uncorrected. Similarly, the effect of the detector response on the ratio can be assessed by computing R on truth and reconstructed (or smeared/shifted) quantities. The systematic bias introduced by the detector response, to be compared with the uncertainty on the factors $A_{Gen \rightarrow Rec}$, is:

$$\frac{\Delta R}{R} = \frac{R^{Reco} - R^{Truth}}{R^{Truth}}. \tag{8.5}$$

It should be noticed that the systematic uncertainties from acceptance and efficiency corrections have to be recomputed for each ratio. Since uncertainties on numerator/denominator correction factors are correlated, they cannot be simply added in quadrature. The R_W and R_{jets} ratios and their uncertainties are treated separately in the following sections.

8.2.1 R_W Ratio

The R_W ratio is determined as:

$$\begin{aligned} R_W &= \frac{N_2 - B_2}{N_1 - B_1} \cdot \frac{A_{Gen\to Rec}[N_{jets=1}]}{A_{Gen\to Rec}[N_{jets=2}]} \cdot \frac{(\epsilon_{\text{jet}})_1}{(\epsilon_{\text{jet}})_2} \cdot \frac{(A_{Rec}[E_T^{\text{miss}} + ele])_1}{(A_{Rec}[E_T^{\text{miss}} + ele])_2} \cdot \frac{(\epsilon_{\text{ele}})_1}{(\epsilon_{\text{ele}})_2} \\ &= \frac{N_2 - B_2}{N_1 - B_1} \cdot A^{R_W}_{Gen\to Rec}[N_{jets}] \cdot \epsilon^{R_W}_{\text{jet}} \cdot A^{R_W}_{Rec}[E_T^{\text{miss}} + ele] \cdot \epsilon^{R_W}_{\text{ele}} \end{aligned}$$

where the subscripts 1 and 2 indicate the number of jets.

Each term in the second line of the equation has an associated uncertainty. We first examine the largest source of experimental uncertainties: jet energy scale and resolution. In the R_W ratio, it is not expected that such uncertainties cancel out completely, given that the number of jets in the numerator and denominator is different, but it is expected that their impact is smaller than on σ_N. The value of R_W as a function of the number of jets N_{jets} is shown in Fig. 8.4 (left), with systematic uncertainties arising from 5% and 10% uncertainty on the jet energy scale. The right plot compares the effect of JES uncertainties on R_W with that on the absolute cross-section. The overall uncertainty on the ratio measurement is significantly reduced for $N_{jets} > 0$. In the ratio $\frac{W+1}{W+0}$, instead, no cancellation occurs and the JES still accounts for an uncertainty larger than 10%. The $\frac{W+1}{W+0}$ is therefore discarded and we focus the attention on the measurement of the $\frac{W+2}{W+1}$ ratio.

Different sources of systematic uncertainties and their impact on the $\frac{W+2}{W+1}$ ratio are summarised in Table 8.3 (left column) and 8.4. The first table shows the impact of detector effects on jets, jet selection cuts and lepton vetos if considered as uncorrected biases. The second table compares biases from electron and E_T^{miss} selections (if uncorrected) with the systematic uncertainties on the acceptance and efficiency corrections. Each source of uncertainty is discussed below:

- Uncertainties from jet energy resolution and scale are significantly reduced with respect

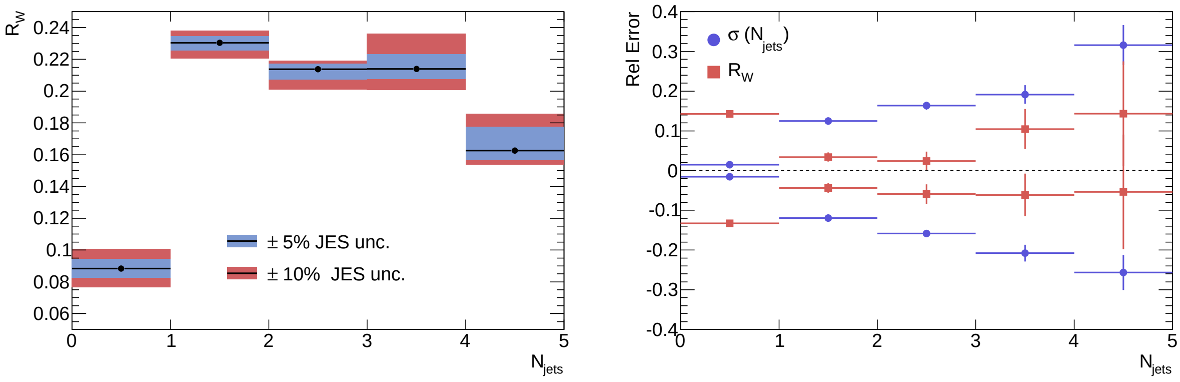

Figure 8.4: (Left) Measurement of the ratio $\frac{\sigma(W + N_{jets} + 1)}{\sigma(W + N_{jets})}$ for different jet multiplicities with systematic uncertainties from a miscalibration in the jet energy scale of 5% and 10%. (Right) Comparison of relative uncertainties from the jet energy scale (assuming a miscalibration of 10%) on the ratio and on the absolute $W \to e\nu$ + jets cross-section.

.

Systematic Uncertainties on R_W and R_{jets} (%)

Sources of systematics	$\frac{\Delta R_W}{R_W}$	$\frac{\Delta R_{jets}}{R_{jets}}$ $(N_{jets} = 1)$	$\frac{\Delta R_{jets}}{R_{jets}}$ $(N_{jets} = 2)$
Jet energy resolution 10%	0.6	0.5	0.4
Jet energy scale 10%	5.0	0.7	0.5
$\epsilon_{\Delta\mathrm{Rej}}$	1.8	0.7	1.9
$\epsilon_{\Delta\mathrm{Rjj}}$	1.2	n/a	< 0.1
$\epsilon_{\mathrm{reco\ jet}}$	0.6	0.1	0.1
ϵ_{veto}^{lep} (Track)	2.9	0.5	0.2
ϵ_{veto}^{lep} (EM Frac)	4.6	0.3	0.5
Statistical	1.4	2.0	4.2

Table 8.3: Systematic uncertainties due to jet energy scale/resolution, jet selection and background rejection cuts on the ratios $R_W = \frac{W+2}{W+1}$ (first column) and R_{jets} with $N_{jets} = 1$ (second column) or $N_{jets} = 2$ (third column). Uncertainties are computed as uncorrected bias on the measurement. Expected statistical uncertainties for an integrated luminosity of 100 pb^{-1} are shown for comparison.

to the absolute cross-section. The JES uncertainty still remains the dominant source of error, leading to a 5% uncertainty on R_W for the most pessimistic scenario of 10% JES uncertainty.

- Jet efficiencies also do not cancel in the ratio. This is hardly surprising, since the ΔR_{ej} and ΔR_{jj} selections are effectively phase-space cuts and therefore depend on the number of jets in the event. Jet reconstruction, in comparison, has a relatively small impact of

Systematic Uncertainties on R_W (%)

Sources of systematics	$\frac{\Delta R_W}{R_W}$	Unc. from correction
E_T^{miss} smearing	1.88	< 1.12
electron p_T smearing	0.29	0.31
ϵ_{ele}	0.30	0.25

Table 8.4: Systematic uncertainties on the R_W measurement from E_T^{miss} and electron p_T smearing and electron selection cuts. The left column shows the effect of each source when it is considered as an unknown and uncorrected bias. The right column shows the systematic uncertainties deriving from correcting such biases with an acceptance (A) or efficiency (ϵ) factor, as described in Chapter 7.

0.6% because jet reconstruction efficiencies are very high ($\sim 99.5\%$). Considering all jet effects (efficiencies and energy response), it is concluded that jet corrections cannot be neglected, implying that jet unfolding has to be performed on the R_W measurement.

- If lepton veto cuts are desired to reject background, they should be applied to both the 1- and 2-jet bin, to favour their cancellation. Even then, however, the bias on the ratio is of $2-4\%$, depending on the lepton cut applied. Signal inefficiencies caused by lepton vetos are in fact largely due to jets that leave hits in the Inner Detector or have large EM fraction, introducing a dependency on the jet multiplicity.

- Detector effects on E_T^{miss} do not cancel in the ratio and, if uncorrected, would lead to a bias of almost 2%. As discussed in Chapter 7, systematic uncertainties from E_T^{miss} smearing corrections were estimated in [71] to be 0.8% on the absolute W boson cross-section measurement. In the ratio R_W part of this systematic uncertainty will cancel out. In the absence of specific studies on the R_W ratio, an upper limit is set by assuming that no cancellation occurs and uncertainties on the corrections for the 1- and 2-jet bin can be added in quadrature. The overall uncertainty from E_T^{miss} smearing correction on the ratio would be 1.12%, which is still smaller than the bias introduced by detector effects on E_T^{miss}. It is therefore concluded that E_T^{miss} detector effects should be corrected for even in the ratio measurement.

- Detector response on electrons, instead, has an effect of 0.29% on the ratio. This is comparable to the systematic uncertainty introduced by correcting the acceptance for electron p_T scale/resolution and is in any case much smaller than the statistical uncertainty. It can therefore be argued that electron p_T smearing can be neglected in the acceptance calculation.

- Similarly, also electron efficiencies largely cancel out in the ratio, with a residual bias of 0.3%. The overall systematic uncertainties from applying TP efficiencies to the ratio measurement are also significantly smaller than in the absolute cross-section measurement. In fact, since TP efficiencies are extracted independently of the jet multiplicity, uncertainties from the background fits totally cancel out. The residual uncertainty is dominated by the statistical error on the TP efficiencies (0.17% for both jet bins) and has a small contribution (0.06%) from the difference between TP and true W efficiencies: $\Delta\epsilon^{\mathrm{W}} = \epsilon_{\mathrm{ele}}^{R_W} - \frac{e^{\mathrm{TP}}(2\ jet)}{e^{\mathrm{TP}}(1\ jet)}$. The overall systematic uncertainty from efficiency corrections of 0.25% is comparable to the bias on R_W and is in any case smaller than jet uncertainties. It would therefore be acceptable to neglect electron efficiency corrections in the R_W measurement.

In comparison to the absolute cross-section, the ratio R_W has some clear advantages. It has been shown that taking the ratio of $W + 2$ and $W + 1$ jets leads to a significant reduction of the largest source of experimental systematic uncertainties, namely the JES uncertainties. These still remain the largest sources of systematics, but are reduced from $12 - 16\%$ to the 5% level. Electron corrections (detector response and efficiencies) have a negligible impact on the ratio, whereas jet and lepton veto efficiencies and E_T^{miss} corrections still have an important effect and cannot be neglected. This implies that a measurement of R_W will require good understanding of jet unfolding and E_T^{miss} reconstruction. While the dependencies of electron efficiencies versus jet variables will have to be checked with data, they are expected to play a minor role in the measurement. Precise predictions of backgrounds will also be crucial, since the background contamination depends on the jet multiplicity.

8.2.2 R_{jets} Ratio

The R_{jets} ratios for N_{jets} =1,2 are determined as:

$$\begin{aligned} R_{jets} &= \frac{N_W - B_W}{N_Z - B_Z} \cdot \frac{(A_{Gen \to Rec}[N_{jets}])_Z}{(A_{Gen \to Rec}[N_{jets}])_W} \cdot \frac{(\epsilon_{\text{jet}})_Z}{(\epsilon_{\text{jet}})_W} \cdot \frac{(A_{Rec}[E_T^{\text{miss}} + ele])_Z}{(A_{Rec}[E_T^{\text{miss}} + ele])_W} \cdot \frac{(\epsilon_{\text{ele}})_Z}{(\epsilon_{\text{ele}})_W} \\ &= \frac{N_W - B_W}{N_Z - B_Z} \cdot A_{Gen \to Rec}^{R_{jets}}[N_{jets}] \cdot \epsilon_{\text{jet}}^{R_{jets}} \cdot A_{Rec}^{R_{jets}}[E_T^{\text{miss}} + ele] \cdot \epsilon_{\text{ele}}^{R_{jets}}. \end{aligned}$$

In the following discussion we assume that Z bosons are selected requiring the presence of exactly two electrons that satisfy the same electron selection criteria as for the W decay. The trigger and jet requirement are identical to the W boson analysis.

A measurement of the R_{jets} ratio has the advantage that uncertainties due to jet selection cuts and jet energy scale and resolution cancel almost completely. The second and third columns of Table 8.3 show the systematic uncertainties from jet selections and background removal cuts for the 1- and 2-jet bins respectively. The JES uncertainty, which dominates in the measurement of σ_N and R_W, is reduced below the 1% level in the R_{jets} ratio. Uncertainties due to the jet energy resolution are also small (within 0.5%) and efficiencies for jet reconstruction and quality selection criteria ($\epsilon_{\text{reco jet}}$, $\epsilon_{\Delta \text{Rjj}}$) cancel at the 0.1% level. An exception is the efficiency of the ΔR_{ej} cut, which cannot cancel in the ratio given the different number of final state electrons in W and Z events. However, this cut is expected to be well modelled in theoretical predictions. A requirement of minimum electron-jet separation (and similarly jet-jet separation) can therefore be included in the definition of ratio measurements, avoiding the need for applying any further correction to the data. The efficiency of background rejection cuts also cancels in the ratio, such that a veto on an isolated track for the 2-jet bin, for example, would have an impact on $R_{jets}(N_{jets} = 2)$ of 0.2%, while significantly reducing contamination from top pair production in the W boson sample (see Chapter 5).

Other sources of uncertainty, such as acceptances and electron efficiencies will not cancel in the R_{jets} ratio. Although specific studies should be performed to quantify their effect on this ratio, an estimate can be obtained by simple considerations:

- Acceptance: experimental systematic uncertainties on the acceptance are expected to be roughly the same as for the W absolute cross-section (see Table 8.1). In the W case, the dominant source of uncertainties are corrections for detector effects on E_T^{miss} ($\sim 0.8\%$), which would equally affect the ratio W/Z. Uncertainties due to electron corrections are expected to be much smaller, but would not totally cancel in the ratio either.

- Electron efficiencies: considering for simplicity global (rather than binned) efficiency corrections, the event-level efficiency for a $Z \to ee$ event (ϵ_Z) is:

$$\epsilon_Z = \epsilon_{\text{reco}}^2 \cdot \epsilon_{\text{id}}^2 \cdot \epsilon_{\text{trig}} \cdot (2 - \epsilon_{\text{trig}}) \tag{8.8}$$

  where the trigger efficiency term $\epsilon_{\text{trig}} \cdot (2 - \epsilon_{\text{trig}})$ is derived considering that either electrons from Z boson decay can trigger the event. The relative systematic uncertainty on R_{jets} from efficiency corrections is:

$$\left(\frac{\Delta\epsilon_{\text{ele}}^{R_{jets}}}{\epsilon_{\text{ele}}^{R_{jets}}}\right)^2 = \left(\frac{\Delta\epsilon_{\text{reco}}}{\epsilon_{\text{reco}}}\right)^2 + \left(\frac{\Delta\epsilon_{\text{id}}}{\epsilon_{\text{id}}}\right)^2 + \left(\frac{\Delta\epsilon_{\text{trig}}}{2 - \epsilon_{\text{trig}}}\right)^2 + \left(\frac{\Delta\epsilon^{\text{W}}}{\epsilon_{\text{reco}} \cdot \epsilon_{\text{id}} \cdot \epsilon_{\text{trig}}}\right)^2$$
$$\text{where} \quad (\Delta\epsilon_{\text{Reco/ ID/ Trig}})^2 = \Delta\epsilon_{\text{stat}}{}^2 + \Delta\epsilon_{\text{syst}}^{\text{Bkg}\,2}.$$

  Note that the uncertainty $\Delta\epsilon^{\text{W}}$ (defined in Chapter 7) affects only the W boson and therefore is added separately. Plugging in the numbers yields an overall uncertainty of 1.06% and 1.08% for the the 1- and 2-jet bins respectively.

Among the sources of systematic uncertainties considered here, electron efficiencies are therefore expected to have the largest impact ($\sim$ 1%) on the ratio R_{jets}. Although not discussed here, accurate predictions of background contamination in the W and Z samples will also be crucial for this measurement. Leaving out the ΔR_{ej} selection cut, the impact of detector response on jets and of jet reconstruction efficiencies is so small compared to other sources of uncertainties, that jet unfolding could even be omitted. The robustness against jet-related uncertainties makes this measurement particularly interesting, not only for early data but also in the long run. Statistical limitations, driven by the Z, will represent the largest constraint with early data: with 100 pb^{-1} of data, the statistical errors are expected to be about 2% for $N_{jets} = 1$ and 4.2% for $N_{jets} = 2$. As more data are collected, however, pinning down systematic uncertainties on electron efficiencies and background predictions will become gradually more important.

8.2.3 Theoretical Uncertainties on R_W and R_{jets} Ratios

In the previous section it was shown that the R_W and R_{jets} ratios are less sensitive to experimental systematic uncertainties than the absolute W + jets cross-section measurement. Whether or not these ratios can be used to extract QCD parameters (such as PDFs or α_{S}) depends on the level of cancellation of the underlying theoretical uncertainties.

Let us again consider uncertainties due to the choice of PDFs. The PDF contribution to a given process cross-section depends on the incoming parton types, as well as their momentum fractions x and the scale Q^2 of the interaction (see Chapter 3). The production W + 0 jets, for example, requiring an initial state $q\bar{q}$, is sensitive to different PDFs with respect to the production of W + N_{jets} (with $N_{jets} > 0$), which can proceed from the initial states $q\bar{q}$ or qg. Furthermore, as more jets are added to the final state, the scale Q^2 increases and with it the average x value probed such that PDF uncertainties will depend on the jet multiplicity, as shown in Fig. 8.3 right. Although there are correlations between different multiplicity bins, the PDF uncertainties are expected to only partially cancel in the R_{jets} ratio. The W + jets ATLAS working group has investigated PDF uncertainties on MCFM predictions for the R_{jets} ratio using the CTEQ6.6 NLO PDF set [3]. Such studies have shown that uncertainties due to PDFs on the ratio $\frac{\sigma(W+2jet)}{\sigma(W+1jets)}$ are smaller than 2.0%. As a comparison, a 5% uncertainty on the JES results in a 3% uncertainty on the ratio measurement. Therefore, although the uncertainties due to the JES are significantly reduced in the R_W ratio with respect to the absolute W cross-section measurement, they still represent the main limitation to the extraction of PDFs. Sensitivity to PDF uncertainties in the R_W ratio can be achieved only if uncertainties on the JES are reduced below the 5% level.

Another interesting property of the R_W ratio is the fact that it is proportional to the strong coupling constant $\alpha_{\rm S}$. For this reason it is often suggested to use these ratios for the extraction of $\alpha_{\rm S}$. However, the sensitivity to $\alpha_{\rm S}$ is small compared to the PDF uncertainties [81]. Therefore a measurement of $\alpha_{\rm S}$ using the R_W ratio will only be possible after constraining the PDFs.

The R_{jets} ratio is also robust against experimental uncertainties and has less sensitivity to theoretical uncertainties than the R_W ratio. Ongoing studies [3] in the ATLAS $W - Z$ + jets working group have shown that PDF uncertainties are reduced to the 1% level, from approximately 4-5% on the absolute cross-sections. Since the numerator and denominator of the R_{jets} ratio are of the same order in $\alpha_{\rm S}$, the ratio shows also little dependence on the choice of factorization and renormalization scales. Given that it is almost insensitive to theoretical uncertainties, the R_{jets} ratio cannot be used to extract QCD parameters. However, rather than a limitation, robustness against both theoretical and experimental uncertainties makes it a powerful measurement for testing SM predictions in search for new physics. In fact, the R_{jets} ratio has first been proposed [82] with the purpose of improving sensitivity to new phenomena which can contaminate the one- and two-lepton channels.

8.3 Conclusions and Outlook

This chapter presented an overview of the $W \to e\nu$ + jets cross-section measurement and the R_W and R_{jets} ratio measurements with early ATLAS data (~100 pb^{-1}). Various sources of systematic uncertainties were investigated and their cancellations in the ratio measurements were discussed.

It was observed that detector effects on jets represent the major limitation for an accurate determination of the absolute cross-section. An uncertainty on the jet energy scale of 5-10 % alone would result in an uncertainty of 6-16% on the cross-section for $W \rightarrow e\nu$ + 1, 2 jets. It was shown that this value is larger than the expected uncertainties due to parton distribution functions, thus preventing the possibility of using an early $W \rightarrow e\nu$ + jets measurement to improve our knowledge of PDF fits.

Ratios are more attractive measurements in an early data scenario, since they suffer from much smaller systematic uncertainties. Cancellation of uncertainties due to detector effects on electrons occur in the R_W ratio and on jets in the R_{jets} ratio. Considering the impact of systematic uncertainties and the way they cancel is of critical importance when designing a measurement. It helps to identify those effects that need deep understanding and, conversely, disregard those that have a negligible impact. For example, pinning down uncertainties due to jet selections will be crucial in the R_W measurement, whereas understanding electron corrections (such as efficiencies extracted with the Tag-and-Probe method) will be a primary focus in the R_{jets} measurement. Among the measurements considered, the R_{jets} ratio is expected to be the one with the smallest systematic uncertainty, but at the same time it will be statistically limited given the smaller production rate of the Z boson.

In the longer term (L $>$ 100 pb^{-1}), not only it will be possible to refine the accuracy of the ratio measurements (thanks to larger statistics and improved calibrations), but also to probe higher kinematic regions. The R_W ratio, for example, could be extended to higher jet multiplicities, where signatures for physics BSM are expected. A scenario foreseen by ATLAS for the R_{jets} measurement is to gradually increase the jet p_T threshold for jet counting, thus shrinking the phase space towards a higher scale. Figure 8.5 shows the expected cumulative distribution for the 1-jet bin. An agreement between data and theoretical predictions in the low kinematic range would be essential to validate the available tools, such as fixed order calculation at LO

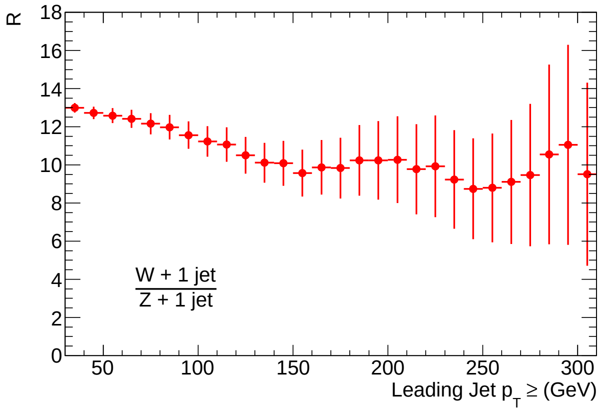

Figure 8.5: Ratio R_{jets} for N_{jets} = 1, assuming 100 pb^{-1} of integrated luminosity. The distribution is cumulative, with the x-axis indicating the threshold applied on the leading jet p_T.

and NLO in QCD and parton shower MC models. At the same time, studying the high kinematic range will be a critical ingredient to many new physics studies, such as Supersymmetry, to which W + jets is a significant background.

Chapter 9

Conclusions

The production of a W boson in association with hadronic jets is an interesting process on its own, as a tool to test QCD predictions, and is also a major background to top production and new physics searches. This thesis presented the work that was developed in preparation for the first measurements of $W \rightarrow e\nu$ + jets at ATLAS. Based on MC simulations, several techniques were developed for the reduction/estimation of backgrounds, extraction of efficiencies and acceptances. The focus was on data-driven methods, which will be very useful in particular in the early data scenario to minimise dependency on MC generators. It was discussed how such techniques can be employed to extract a cross-section or a ratio of cross-sections. The impact of several sources of systematic uncertainties were assessed and suggestions were made for robust ratio measurements which could be performed with the first 100 pb^{-1} of data. Here we summarise the major observations made in previous chapters.

Recommendations were made for the choice of jet selection to be adopted. This analysis considered anti-k_T jets with distance parameter D=0.4, with $p_T > 30$ GeV and $|\eta| < 3.1$. A minimum separation between two jets or the candidate W electron and a jet was also required. While causing a small loss of statistics, these conditions ensured that phase-space regions with poor electron or jet performance were excluded. The minimum electron-jet separation requirement, in particular, was also advantageous since it significantly reduced the dependence of electron efficiencies on jet variables.

Several sources of background were treated. Background from QCD-jets could not be studied in detail because of poor MC statistics, but a qualitative description of a data-driven method for its estimation was presented. Emphasis in this thesis was given to leptonic backgrounds. Lepton vetos were proposed to reduce background from processes with a second lepton in addition to the selected electron. It was observed that a veto exploiting the fraction of EM energy deposit to identify electrons would only lead to a small background reduction. In contrast, a veto on isolated tracks is much more effective in reducing the background and would result in a sizable decrease in the overall systematic uncertainty in a W + N_{jets} cross-section measurement, at least for the 2 jet bin. Residual leptonic backgrounds can be estimated using MC simulations, in some cases using well known ratios to decrease the systematic uncertainty on the prediction. A data-driven method for the estimation of background from top pair production was also developed in view of measurements at high jet multiplicities, where $t\bar{t}$ is a formidable background. Despite some limitations in predictions for the 1 and 2 jet bins, this technique was shown to be powerful for $N_{jets} \geq 3$. Suggestions for improvements were made (using, for example, different cuts for selecting the control sample). Furthermore, in data several backgrounds will have to be considered together and methods for QCD-jet and $t\bar{t}$ background subtraction should therefore be applied simultaneously.

The Tag-and-Probe method for extracting electron efficiencies in-situ from $Z \rightarrow ee$ events was studied in detail. A technique for subtracting the QCD-jet background using a simultaneous fit of signal and background was developed. If efficiencies are binned in some variables, a fit has to be performed for each individual bin. This introduces some complications, since the background shape varies with p_T bins. Despite the difficulties encountered to achieve a good fit, the resulting efficiencies were accurate within 0.5% for $p_T > 35$ GeV.
Prospects for measuring efficiencies with early data were also discussed. With 100 pb^{-1} of data, the Tag-and-Probe method allows measuring 2-dimensional efficiencies with a statistical precision of 1-1.5%. In addition, measurements of efficiencies using different Tag-and-Probe

configurations would also be a powerful tool to test the accuracy of the detector simulation for low quality electrons.

It was shown how acceptance and efficiency corrections enter in the measurement of a cross-section. Although the generator level acceptance has to be determined from MC, detector level corrections can be calculated applying a smearing technique, which allows using data-driven measurements of electron scale and resolution rather than MC estimates. Tag-and-Probe efficiencies can be applied to $W \to e\nu$ events, provided that care is taken on the $Z \to ee$ selection criteria used. Some Tag-and-Probe selections can bias the kinematics of the probe or the composition of the probe sample, therefore affecting the efficiency. Assessing the potential biases introduced by this method is therefore crucial to minimise systematic uncertainties on the measurement. For an accurate result, it is also important to bin acceptance and efficiencies in their driving variables.

The prospects for measuring $W \to e\nu$ + jets processes at the LHC were presented. The measurement of the W plus jets cross-section is totally dominated by systematic uncertainties on jet energy scale, which results in an uncertainty as large as 16%. Much more robust are measurements of ratios, which gain from the cancellation of several sources of uncertainties. The $R_W = \dfrac{\sigma(W + N_{jets} + 1)}{\sigma(W + N_{jets})}$ ratio benefits from a major cancellation of electron corrections and a partial cancellation of uncertainties on jets. The latter still remain the dominant uncertainties, but are reduced to the 5% level. Even more robust is the $R_{jets} = \dfrac{\sigma(W + N_{jets})}{\sigma(Z + N_{jets})}$ ratio, which is almost insensitive to jet effects and has systematic uncertainties at the few percent level. The most important aspects for this measurement are the determination of backgrounds, electron efficiencies and detector effects on transverse missing energy. The major limitation in early data, however, will be statistical.

The R_W and R_{jets} measurements will be crucial in the early LHC phase to validate theoretical predictions and MC tools. As more statistics become available, the measurements will be extended to the corners of phase-space which are most relevant for Beyond the Standard Model searches and might even provide hints for new physics.

Bibliography

[1] M. Fiascaris. ATLAS Silicon Pixel and Strip Detector: Commissioning with Cosmic Data. Technical Report ATL-INDET-PROC-2009-007, 2009.

[2] ATLAS Collaboration (G. Aad et al.). Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics. Technical Report CERN-OPEN-2008-020, 2008.

[3] J.F. Arguin et al. Prospects for W/Z + jets early data measurements with the ATLAS detector at the LHC - An analysis walkthrough. Technical Report ATL-COM-PHYS-2010-150, 2010.

[4] M. Pivk and F. R. Le Diberder. sPlot: a statistical tool to unfold data distributions. Technical Report physics/0402083. LAL-2004-07, Paris 11. Lab. Accl. Linaire, Orsay, 2004.

[5] M Fiascaris. Production of jets in association with W and Z at ATLAS. Technical Report ATL-PHYS-PROC-2008-024, 2008.

[6] F J Hasert et al. Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment. *Nucl. Phys. B*, 73(1):1–22, 1974.

[7] UA1 Collaboration. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}$ = 540 GeV. *Phys. Lett.*, B122:103–116, 1983.

[8] UA1 Collaboration. Further evidence for charged intermediate vector bosons at the SPS collider. *Phys. Lett. B*, 129:273–282, 1985.

[9] UA1 Collaboration. Experimental observation of lepton pairs of invariant mass around $95\,GeV/c^2$ at the CERN SPS collider. *Phys. Lett. B*, 126:398–410, 1985.

[10] UA2 Collaboration. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN anti-p p collider. *Phys. Lett.*, B122:476–485, 1983.

[11] F. Abe et al. Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 74(14):2626–2631, 1995.

[12] S. Abachi et al. Search for High Mass Top Quark Production in $\bar{p}p$ Collisions at $s = 1.8$ TeV. *Phys. Rev. Lett.*, 74(13):2422–2426, 1995.

[13] The LEP Electroweak Working Group (LEP EWWG). http://lepewwg.web.cern.ch/LEPEWWG/.

[14] S P Martin. A Supersymmetry Primer. Technical Report hep-ph/9709356, 1997.

[15] ATLAS Collaboration (G. Aad at al.). The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[16] ATLAS Collaboration (G. Aad et al.). ATLAS pixel detector electronics and sensors. *J. Instrum.*, 3:P07007, 2008.

[17] ATLAS Collaboration (G. Aad et al.). *ATLAS First-Level Trigger Technical Design Report*. CERN, 1998. CERN-LHCC-98-14.

[18] ATLAS Collaboration (G. Aad et al.). *ATLAS High-Level Trigger, Data Aquisition and Controls Technical Design Report*. CERN, 2003. CERN-LHCC-2003-022.

[19] ATLAS Event Displays . https://twiki.cern.ch/twiki/bin/view/Atlas/EventDisplayPublicResults.

[20] T. Affolder et al. Charged jet evolution and the underlying event in proton-antiproton collisions at 1.8 tev. *Phys. Rev. D*, 65(9):092002, Apr 2002.

[21] V. N. Gribov and L. N. Lipatov. Deep inelastic ep scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15(4):438–450, 1972.

[22] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nucl. Phys. B*, 126:298–318, 1977.

[23] G. C. Blazey et al. Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop. Technical Report hep-ex/0005012, May 2000.

[24] C. Buttar et al. Standard Model Handles and Candles Working Group: Tools and Jets Summary Report. Technical Report arXiv:0803.0678, 2008.

[25] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *J. High Energy Phys.*, 04:063, Feb 2008.

[26] J. M. Campbell, J. W. Huston, and W. J. Stirling. Hard Interactions of Quarks and Gluons: a Primer for LHC Physics. *Rep. Prog. Phys.*, 70:89, Nov 2006.

[27] R. K. Ellis, K. Melnikov, and G. Zanderighi. Generalized unitarity at work: first NLO QCD results for hadronic W + 3 jet production. *J. High Energy Phys.*, 04:077, Jan 2009.

[28] C. F. Berger et al. Next-to-Leading Order QCD Predictions for W + 3-Jet Distributions at Hadron Colliders. Technical Report arXiv:0907.1984, Jul 2009.

[29] C. F. Berger et al. Precise Predictions for W + 3 Jet Production at Hadron Colliders. Technical Report arXiv:0902.2760, Feb 2009.

[30] J. M. Campbell and R. K. Ellis. MCFM for the Tevatron and the LHC. Technical Report arXiv:1007.3492. FERMILAB-CONF-10-244-T, Jul 2010.

[31] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *J. High Energy Phys.*, 07:001, Jun 2002.

[32] S. Hoche, F. Krauss, N. Lavesson, L. Lonnblad, M. L. Mangano, A. Schlicke, and S. Schumann. Matching parton showers and matrix elements. (hep-ph/0602031), Feb 2006.

[33] S Catani, F Krauss, R Khn, and Bryan R Webber. Qcd matrix elements + parton showers. *J. High Energy Phys.*, 11:063. 21 p, Sep 2001.

[34] J Alcaraz et al. A combination of preliminary electroweak measurements and constraints on the Standard Model. Technical Report arXiv:hep-ex/0612034, Dec 2006.

[35] T. Aaltonen et al. First run II measurement of the W boson mass at the Fermilab Tevatron. *Phys. Rev. D*, 77(11):112001, Jun 2008.

[36] T. Aaltonen et al. Direct Measurement of the W Boson Width in $\bar{p}p$ Collisions at $\sqrt{s} = 1.96\ TeV$. *Phys. Rev. Lett.*, 100(7):071801, Feb 2008.

[37] V. M. Abazov et al. Measurement of the W Boson Mass. *Phys. Rev. Lett.*, 103(14):141801, Oct 2009.

[38] V. M. Abazov et al. Direct Measurement of the W Boson Width. *Phys. Rev. Lett.*, 103(23):231802, Dec 2009.

[39] V. A. Khoze, A. D. Martin, R. Orava, and M. G. Ryskin. Luminosity monitors at the LHC. *Eur. Phys. J. C*, 19:313–322, Oct 2000.

[40] C. Anastasiou, L. Dixon, K. Melnikov, and F. J. Petriello. High-precision QCD at hadron colliders: electroweak gauge boson rapidity distributions at NNLO. *Phys. Rev. D*, 69:94008–1–27. 47 p, Dec 2003.

[41] ATLAS Collaboration (G. Aad at al.). Measurement of the $W \to l\nu$ production cross-section and observation of $Z \to ll$ production in proton-proton collisions at $\sqrt{s}$=7 TeV with the ATLAS detector. Technical Report ATLAS-CONF-2010-051, Jul 2010.

[42] M. L. Mangano. Understanding the Standard Model, as a bridge to the discovery of new phenomena at the LHC. *Int. J. Mod. Phys. A*, 23:3833–3848, Feb 2008.

[43] V M Abazov et al. Measurement of the ratio of the $p\bar{p} \to W$+c-jet cross section to the inclusive $p\bar{p} \to W$+jets cross section. *Physics Letters B*, 666(1):23 – 30, 2008.

[44] A Study of the strong coupling constant using W+ jets processes. Technical Report FERMILAB-PUB-95-219-E, FERMILAB, Batavia, IL, 1995.

[45] T. Aaltonen et al. Measurement of the cross section for W-boson production in association with jets in $p\bar{p}$ collisions at $s = 1.96$ TeV. *Phys. Rev. D*, 77(1):011108, Jan 2008.

[46] T. Affolder et al. Tests of Enhanced Leading Order QCD in W Boson plus Jets Events from 1.8 TeV $p\bar{p}$ Collisions. *Phys. Rev. D*, 63:72003, 2000.

[47] J. M. Campbell. Overview of the theory of W/Z + jets and heavy flavor. Technical Report arXiv:0808.3517, Aug 2008.

[48] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Methods Phys. Res., A*, 506:250–303. 86 p, Jul 2002.

[49] J. Butterworth, J. R. Forshaw, and M. H. Seymour. Multiparton interactions in photoproduction at HERA. *Z. Phys. C*, 72:637–646. 22 p, Jan 1996.

[50] S. Frixione and B. R. Webber. Matching NLO QCD computations and parton shower simulations. *J. High Energy Phys.*, 06:029. 69 p, Apr 2002.

[51] B. P. Kersevan and E. Richter-Was. The Monte Carlo Event Generator AcerMC 2.0 with Interfaces to PYTHIA 6.2 and HERWIG 6.5. AcerMC 2.0 (thoroughly improved) program manual. Technical Report ATL-PHYS-2004-020, May 2004.

[52] S. Jadach, Z. Was, R. Decker, and J. H. Kuhn. The τ decay library TAUOLA, version 2.4. *Comput. Phys. Commun.*, 76:361–380. 22 p, Feb 1993.

[53] E. Barberio and Z. Was. PHOTOS: a universal Monte Carlo for QED radiative corrections; version 2.0. *Comput. Phys. Commun.*, 79:291–308. 23 p, Oct 1993.

[54] M. R. Whalley and D. Bourilkov. The Les Houches Accord PDFs (LHAPDF) and Lhaglue. (hep-ph/0508110), Aug 2005.

[55] W. Lampl et al. Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, 2008.

[56] B. Aubert et al. Performance of the ATLAS Electromagnetic Calorimeter End-cap Module 0. *Nucl. Instrum. Methods Phys. Res., A*, 500:178–201r, Nov 2002.

[57] B. Aubert et al. Performance of the ATLAS Electromagnetic Calorimeter Barrel module 0. *Nucl. Instrum. Methods Phys. Res., A*, 500:202–231, Nov 2003.

[58] ATLAS Collaboration (G. Aad et al.). Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics (Electrons and Photons). Technical Report CERN-OPEN-2008-020, 2008.

[59] ATLAS Collaboration (G. Aad at al.). Photon Conversions at $\sqrt{s} = 900$ GeV measured with the ATLAS Detector. Technical Report ATLAS-CONF-2010-007, Jun 2010.

[60] ATLAS Collaboration (G. Aad at al.). Study of the Material Budget in the ATLAS Inner Detector with K^0_S decays in collision data at $\sqrt{(}s) = 900$ GeV. Technical Report ATLAS-CONF-2010-019, Jul 2010.

[61] ATLAS Collaboration (G. Aad et al.). Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics (Jets and Missing Transverse Energy). Technical Report CERN-OPEN-2008-020, 2008.

[62] ATLAS Collaboration (G. Aad et al.). Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics (Standard Model: Electroweak Boson Cross-Section Measurements). Technical Report CERN-OPEN-2008-020, 2008.

[63] Barillari et al. Local Hadronic Calibration. Technical Report ATL-LARG-PUB-2009-001. ATL-COM-LARG-2008-006, 2008.

[64] ATLAS Collaboration (G. Aad et al.). Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics (Standard Model: Production of Jets in Association with Z Bosons). Technical Report CERN-OPEN-2008-020.

[65] M Fiascaris. Production of Jets in association with W and Z at ATLAS. 2008. http://www.lpta.univ-montp2.fr/users/qcd/qcd08_talks/fiascaris.pdf.

[66] C. W. Bauer and B. O Lange. Scale setting and resummation of logarithms in pp $\rightarrow$ V + jets. Technical Report arXiv:0905.4739, Jun 2009.

[67] ATLAS Collaboration (G. Aad et al.). Strategy to Search for Single-Top Events using early Data of the ATLAS Detector at the LHC. Technical Report ATL-COM-PHYS-2010-093, 2010.

[68] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi. Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC. *J. High Energy Phys.*, 09:127, Apr 2008.

[69] ATLAS Collaboration (G. Aad et al.). Prospects for measuring top pair production in the dilepton channel with early ATLAS data at $\sqrt{s}$=10 TeV. Technical Report ATL-PHYS-PUB-2009-086. ATL-COM-PHYS-2009-402, 2009.

[70] ATLAS Collaboration (G. Aad et al.). Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics (Top Quark Physics). Technical Report CERN-OPEN-2008-020.

[71] E. Dobson and A. Weidberg. *A feasibility study into the measurement of W and Z cross sections with the ATLAS detector*. PhD thesis, Oxford U., Oxford, 2009.

[72] K. Cranmer. Kernel Estimation in High-Energy Physics. *Comput. Phys. Commun.*, 136:198–207, 2001.

[73] W. Verkerke and Kirkby D. The RooFit toolkit for data modelling. http://roofit.sourceforge.net/.

[74] T. Ullrich and Z. Xu. Treatment of Errors in Efficiency Calculations. Technical Report arXiv:physics/0701199v1, Jan 2007.

[75] M. J. Flowerdew, J. H. Vossebeld, and S. J. Maxfield. *Standard Model Z/gamma to ee production in early data at ATLAS*. PhD thesis, Liverpool U., Liverpool, 2009.

[76] ATLAS Collaboration (G. Aad et al.). Luminosity Determination Using the ATLAS Detector. Technical Report ATLAS-CONF-2010-060. ATLAS-COM-CONF-2010-060, Jul 2010.

[77] ATLAS Collaboration (G. Aad et al.). Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at sqrt(s) = 7 TeV and measured with the ATLAS detector. Technical Report ATLAS-CONF-2010-056, Jul 2010.

[78] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung. New Generation of Parton Distributions with Uncertainties from Global QCD Analysis. *J. High Energy Phys.*, 07:012. 44 p, Jan 2002.

[79] A. Cooper-Sarkar. Private communication.

[80] A. Tricoli, A. M. Cooper-Sarkar, and C. Gwenlan. Uncertainties on W and Z production at the LHC. Technical Report hep-ex/0509002, Sep 2005.

[81] G Kirsch and J Tseng. *Study of the sensitivity of jet parameter distributions to the strong coupling constant α_s in $Z^0 \to e^+e^-$ events at ATLAS*. PhD thesis, Oxford U., Oxford, 2009.

[82] Abouzaid E. and H. Frisch. The Ratio of W + N jets To Z/gamma + N jets As a Precision Test of the Standard Model. Technical Report arXiv:hep-ph/0303088v1, 2003.