Multijet Background Estimation for Supersymmetry Searches Using the ATLAS Detector at the Large Hadron Collider

Simon Owen

Supervisor: Davide Costanzo

A Thesis Submitted for the Degree of Doctor of Philosophy in Physics

Department of Physics and Astronomy
University of Sheffield, February 2012

ABSTRACT

This thesis describes a data-driven technique for estimating the multijet background to Supersymmetry (SUSY) searches with no leptons using the ATLAS detector at the Large Hadron Collider. The technique is used to estimate multijet distributions in SUSY signal and control regions with 1 fb^{-1} of $\sqrt{s} = 7$ TeV data collected by ATLAS in 2011. The systematic uncertainty on the estimates is reduced with the development and use of novel event shape triggers. Multijet estimates provided from the technique developed in this thesis are used by the ATLAS collaboration in several different SUSY searches.

ACKNOWLEDGEMENTS

I would first like to thank my supervisor, Davide Costanzo, for his guidance and help during my PhD studies. It has been a pleasure to work with such a relaxed and supportive supervisor. Thanks also to Dan Tovey who had endless patience in helping out all of those times I knocked on his office door. A huge thank you to Rob Duxfield for all his help and encouragement in my first couple of years. I was very lucky to work with such an intelligent and (more importantly) thoroughly nice bloke. Thanks also to all of the other friendly and helpful staff of the Sheffield high-energy physics group with special mention to Matt Robinson, Paul Hodgson and Elena Korolkova for their limitless computing expertise and to Mark Hodgkinson, Mark Sutton and Kerim Suruliz for sharing their knowledge and experience. Cheers to all of the Sheffield PhD students I've worked with; despite their embarrassing lack of skill at office badminton and poor work ethics, Josh and Alan have been truly excellent company and I will miss *most* of the chat.

Thanks to all of the people who made my time at CERN a brilliant experience. On the work side, thanks for the friendly and constructive help I received from members of the SUSY and jet trigger groups and a big thank you to Teresa Fonseca Martin and Ignacio Aracena for their help with the trigger studies. Away from work, I owe John and Vik a lot of beers for almost constant use of their sofa and for starting off my professional Pro Evo career. Thanks to all of the LTA crowd and especially to Gareth, Zoe, Sarah, Dan, Graham, Jody, Tom and Giulia for being great friends and climbing partners too! Also, a special mention to the members of Chamonix chalet (2), 2010, for a ridiculously fun holiday.

Finally, thanks to Lisa for being brilliant through everything, the best company at the beginning/middle/end of every day.

Contents

List of Tables

List of Figures

Author's Contribution

The material presented in this thesis is based upon work performed by a large number of people in the ATLAS collaboration. The author's contribution is summarised here.

Service Work

The author was responsible for maintaining and developing the "Full Chain Testing" framework. This work is described in more detail in §4.6. The author also took numerous "Computing at Point 1" shifts; these involved monitoring the data flow from the CERN Tier 0 computing farm to Tier 1 and Tier 2 sites around the world.

Trigger Development

The author developed and maintained the $\Delta\phi$ triggers discussed in §7.4.3. As part of this work, the author aided the validation of the L2 jet triggers, contributing plots to a note on jet trigger performance [1]. The algorithms written for the $\Delta\phi$ triggers were also used for other event shape triggers including a trigger using the M_{CT} variable [2] for SUSY searches.

Multijet Background Estimation

The author's main contribution to the ATLAS collaboration was the development of the jet smearing technique for estimating the multijet background to SUSY searches with jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and no leptons at ATLAS. The author's work was directly used for the multijet estimation in several ATLAS papers [3, 4, 5]. The techniques developed have also been used in several other SUSY searches including searches with b-jets [6, 7, 8] and searches with high jet multiplicities [9]. The most up to date version of the jet smearing method is described in this thesis in Chapters 6 to 8 and is also documented in an internally approved ATLAS note [10].

Additional Contributions

The author has presented SUSY background studies at a number of meetings including "ATLAS UK 2009" in Oxford, where the author was awarded the poster prize; "ATLAS UK 2010" in Cambridge and "Physics at the LHC 2010" in Hamburg, where the author won the poster prize. The author has also presented results from the ATLAS collaboration on diphoton searches with $E_{\mathrm{T}}^{\mathrm{miss}}$ and heavy long-lived charged particles using the muon spectrometer at "Physics at the LHC 2011" in Perugia.

Chapter 1

Introduction

The search for SUSY is one of the most important tasks for experimental particle physicists. Searches using the particle detectors at the Large Hadron Collider (LHC) currently offer the best chance of directly discovering SUSY particles in nature. In the searches using the ATLAS detector, the measurement of jet response and its use for multijet background estimation is a vital part of this quest.

This thesis describes a data-driven technique for estimating the multijet background to fully hadronic searches for Supersymmetry (SUSY) at ATLAS. This includes the measurement of jet response, including its non-Gaussian tails, and the development and use of novel event shape triggers to increase the sensitivity of this measurement. The use of this technique on 1 fb^{-1} of data collected by the ATLAS experiment in 2011 is described.

Chapter 2 gives an overview of the Standard Model of particle physics and an introduction to SUSY. The motivation for SUSY is explained, in that it provides a solution to the hierarchy problem of the Higgs mechanism of electroweak symmetry breaking. The current status of SUSY searches is then reviewed including strategies for controlling and estimating the multijet background to fully hadronic searches at hadron colliders. Chapter 3 describes the LHC and the ATLAS detector, with particular emphasis on the calorimetry and trigger systems which are used in this thesis. Chapter 4 then describes the analysis tools used in this thesis, such as Monte Carlo simulation and how physics objects are reconstructed using ATLAS. Chapter 5 gives details on the fully hadronic SUSY search performed at ATLAS, to which the author contributed. The results from this search using 1 fb^{-1} are presented.

Chapter 6 gives an overview of the jet smearing method for multijet background estimation including the assumptions behind the technique and how low-$E_{\mathrm{T}}^{\mathrm{miss}}$ data seed events are selected. Chapter 7 then describes how jet smearing is used to measure jet response at ATLAS, including the measurement of non-Gaussian response tails. Novel event shape $\Delta\phi$ triggers are then introduced and they are shown to provide extra statistics for the measurement of non-Gaussian response. Chapter 8 describes the use of jet smearing to provide the multijet estimation for the fully hadronic SUSY search and presents the results of this analysis.

Chapter 9 gives a summary of this thesis and concludes. Future work which could improve the techniques developed in this thesis is discussed.

Chapter 2

Theoretical and Experimental Background

This chapter gives an overview of where the work of this thesis fits in the wider picture of particle physics research and particularly in searches for Supersymmetry (SUSY) at hadron colliders. §2.1 describes the Standard Model, including details on the strong interaction and the Higgs mechanism of electroweak symmetry breaking. §2.2 then introduces SUSY and explains its central ideas. The current progress in the search for SUSY is briefly reviewed in §2.3. As yet there is no experimental evidence for SUSY, the work in this thesis contributes to some of the latest and most sensitive searches performed with ATLAS. Details of jet production and reconstruction in detectors at hadron colliders are given in §2.4. The strategies for dealing with the multijet background to SUSY searches with no leptons at hadron collider experiments are discussed in §2.5, this is very relevant to the work in this thesis.

2.1 The Standard Model

The Standard Model (SM) (see e.g. [11]) is our most accurate mathematical description of the fundamental constituents of the universe and how they interact. It was formulated in the 1970's and succeeded in bringing together the electroweak and quantum chromodynamic theories into an internally consistent framework. All of the particles associated with the SM, with the exception of the Higgs boson, have been experimentally observed. The SM cannot be a complete description of the universe as it does not include gravity; furthermore, issues such as the gauge hierarchy problem (see §2.1.3) and no dark matter candidate motivate extensions to the SM.

2.1.1 Particle Content

The SM is a quantum field theory and the excitations of the fields described by the SM are the point-like fundamental particles which we observe in nature. The Lagrangian of the theory is required to be "gauge invariant"; this means that the theory does not depend on the space-time positions of the fields. To maintain this invariance, additional gauge fields are included in the theory which provide the forces between the particles. Particles in the SM possess internal angular momentum known as spin, s. Matter particles (fermions) have half-integer spin $s = 1/2$ and the forces between them are mediated by particles with integer spin (bosons). There are three forces in the SM: the strong interaction, the weak interaction and electromagnetism. Fermions which feel the strong interaction are known as quarks, otherwise they are known as leptons. Table 2.1 shows the matter particles of the SM in the three generations which have been observed, each generation containing a quark and lepton doublet. Table 2.2 shows the vector gauge bosons (excitations of the gauge fields) which mediate the forces between the particles.

2.1.2 The Strong Interaction

The theory which governs the interactions of quarks and gluons is known as the strong interaction, or Quantum Chromodynamics (QCD). The theory introduces a new type of charge, known as "colour charge", which is respected by the strong interaction in the same way that electric charge is respected by electromagnetism. The strong interaction is the strongest of the three forces and is mediated by the gluon which itself

Quarks			Leptons		
Particle	Mass (MeV)	Charge (e)	Particle	Mass (MeV)	Charge (e)
up (u)	1.7-3.1	$\frac{2}{3}$	electron (e)	0.511	-1
down (d)	4.1-5.7	$-\frac{1}{3}$	e neutrino (ν_e)	$< 2 \times 10^{-6}$	0
charm (c)	1290^{+50}_{-110}	$\frac{2}{3}$	muon (μ)	105.7	-1
strange (s)	100^{+30}_{-20}	$-\frac{1}{3}$	μ neutrino (ν_μ)	$< 2 \times 10^{-6}$	0
top (t)	1.73×10^5	$\frac{2}{3}$	tau (τ)	1777	-1
bottom (b)	4190^{+180}_{-60}	$-\frac{1}{3}$	τ neutrino (ν_τ)	$< 2 \times 10^{-6}$	0

Table 2.1: Fermions in the Standard Model. These particles have spin $s = 1/2$. Interactions between these particles are mediated by the gauge bosons shown in Table 2.2.

Force	Vector Boson	Mass (GeV)	Electric Charge (e)
Strong	gluon (g)	0	0
Weak	$W^\pm$	80.4	± 1
	Z	91.2	0
Electromagnetism	photon (γ)	0	0

Table 2.2: Vector gauge bosons in the Standard Model. These particles possess integer spin and mediate interactions between the fermions shown in Table 2.1. They are excitations of the gauge fields.

carries colour charge. The strong interaction increases in strength with increasing distance; this property, coupled with the strength of the force, leads to the phenomenon of confinement. Confinement dictates that particles carrying colour charge (quarks and gluons, collectively known as partons) can only exist in bound states (hadrons). Note that hadrons can be further subdivided into two groups: baryons, which contain three quarks and mesons, which contain quark-antiquark pairs. Partons can only exist in bound states because the energy required to separate two partons is sufficient to produce a quark-antiquark pair which can then bind to the original partons. If a parton is given sufficient energy then it can produce many new pairs of particles due to confinement, this is known as hadronisation. Hadronisation is observed in a particle detector as a group particles including baryons and mesons travelling approximately co-linearly; this is usually referred to as a jet. Jet formation is described in more detail in §2.4.

2.1.3 Electroweak Theory, the Higgs Mechanism and the Hierarchy Problem

Electromagnetism and the weak force unify at ~ 100 GeV and are described by the electroweak theory. The four vector bosons which carry the electroweak interaction are the photon, W^+, W^- and the Z. As shown in Table 2.2, the $W^\pm$ and Z bosons are observed to be massive. The mechanism for generating these masses in the SM, without violating gauge invariance, is the Higgs mechanism [12]. A self-interacting complex doublet scalar field is added to the theory which causes spontaneous symmetry breaking giving mass to the gauge bosons and fermions. The fermion masses originate due to interactions between the Higgs scalar field and the fermionic fields (interactions of this type are known as Yukawa couplings). The particle associated with the Higgs field is the undiscovered Higgs boson. Although the mass of the Higgs Boson is not predicted by the SM, other observables in the SM are sensitive to its mass and these indicate a Higgs mass of less than 161 GeV [13]. Recent results from the ATLAS experiment at the LHC indicate that Higgs mass lies between 115.5 GeV and 131 GeV [14]. The probable low mass of the Higgs Boson conflicts with the fact that radiative corrections to the Higgs mass (through such loop diagrams as shown in 2.1) diverge quadratically up to a cut-off scale, Λ, where the SM is no longer valid. If the SM is valid up to high energies then extreme amounts of fine-tuning to the bare Higgs are required to keep the observed mass at ~ 100 GeV. This unfeasible amount of fine-tuning is known as the hierarchy problem.

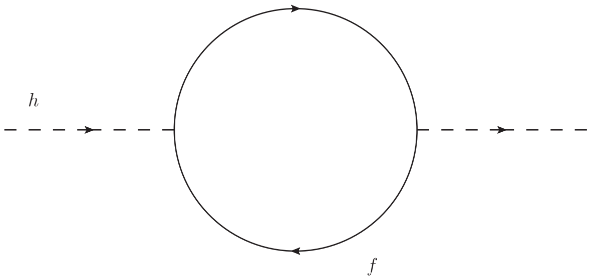


Figure 2.1: Example of a one-loop diagram which contributes to the Higgs mass squared.

2.2 Supersymmetry

Supersymmetry (SUSY) hypothesises a symmetry between bosons and fermions. For every SM particle, a partner particle ("sparticle") is introduced with a spin quantum number which differs by one half. Furthermore, no particles currently observed can be partners of other SM particles and therefore a wealth of new undiscovered particles is predicted. The motivation for the theory is that each sparticle will contribute to the radiative corrections to the Higgs mass discussed in §2.1.3. Their contributions will be opposite to the SM particles hence cancelling the divergences and removing the need for fine-tuning the bare Higgs mass. SUSY, if it exists, must be a broken symmetry as the sparticle masses must be much higher than their SM particle partners (otherwise they would have already been observed). Various mechanisms have been proposed for the breaking of the theory, more details are given later in this section.

In most SUSY models, a multiplicative conservation law is applied on a quantity known as R-parity:

$$R = (-1)^{3(B-L)+2s}, \tag{2.1}$$

where B is baryon number and L is lepton number. SM particles have even R-parity and sparticles have odd R-parity. The conservation law is imposed otherwise SUSY models would predict (amongst other things) proton decay. This naturally leads to the Lightest Supersymmetric Particle (LSP) being stable as there is no kinematically allowed decay to a sparticle and decays to SM particles would violate R-parity. At the high energies present shortly after the big bang, sparticles would have been produced copiously and then would each have decayed into a LSP. If the LSP is weakly interacting then these stable sparticles naturally provide a dark matter candidate. R-parity conservation also has important consequences for collider experiments such as the LHC:

- Sparticles must always be produced in pairs.
- Any sparticle will eventually decay into the LSP which is stable. If the LSP has no electric charge and is weakly interacting (as is required for SUSY to provide a dark matter candidate) then the LSP will leave a particle detector unobserved. Its presence can only be inferred from momentum imbalance of the observed particles.

As there is yet to be any experimental evidence for SUSY or measurements of SUSY parameters, the theory is largely unconstrained. Given the large number of additional particles postulated by the theory, there are a vast number of free parameters and therefore possible SUSY models. Assumptions need to be made to group these models to allow their predictions to be tested with experimental observations.

The simplest possible SUSY model which is consistent with the SM is called the Minimal Supersymmetric Standard Model (MSSM). It postulates the smallest number of new particles and does not specify the mechanism for SUSY being a broken symmetry. Two favoured models which are based on the MSSM, but which do specify how SUSY breaking occurs, are the constrained MSSM (cMSSM) and minimal supergravity (mSUGRA) [15, 16, 17, 18, 19]. In both of these models, gravitational interactions mediate SUSY breaking. These models are closely related and rely on only five parameters including the common boson mass, m_0 and the common fermion mass, $m_{1/2}$. Exclusion limits for these models are often shown in the $m_0, m_{1/2}$ plane. It should be noted that in the MSSM, the observable mass eigenstates are mixings of different sparticles. For example, the "neutralino" is the electromagnetically neutral mass eigenstate and the "chargino" is the charged mass eigenstate of the mixing of the sparticle partners of gauge and Higgs bosons. In many models, the LSP is the neutralino ($\tilde{\chi}_1^0$).

Recently, effort has been made to produce a set of simple models (containing few parameters) which are based on the expected phenomenology of SUSY models but dispense with much of the complexity associated with them [20]. These are known as “simplified models” and are intended to provide a framework to describe any SUSY-like excess which may be observed in LHC data.

2.3 Current Status of SUSY Searches

LEP, the Tevatron and the LHC have performed wide-reaching searches for signs that SUSY exists in nature. There is yet to be any experimental evidence for sparticle production. Searches for squarks ($\tilde{q}$) and gluinos ($\tilde{g}$), the super-partners of quarks and gluons respectively, are most relevant to this thesis and are summarised in this section.

2.3.1 SUSY Searches at Past Colliders

Before recent results from the LHC, LEP and the Tevatron provided the most stringent limits on searches for squarks and gluinos. Gluino production is forbidden at LEP as the gluino does not interact electroweakly. In squark searches, exclusion results from LEP are generally much less stringent than results from the Tevatron. However, in the stop (the SUSY partner of the top quark) sector, LEP still provides some useful information. For example, in searches for acoplanar jets with missing energy, stop masses from 96 to 99 GeV are excluded in the case of $m_{\tilde{t}} - m_{\tilde{\chi}_1^0} - m_c > 5$ GeV [21].

The Tevatron, before its deactivation, ran at a centre-of-mass energy of 1.96 TeV, substantially higher than 209 GeV achieved at LEP. This results in its mass reach for SUSY particles being much larger than at LEP. However, the search is complicated by the fact that the momenta of the colliding partons are unknown. Furthermore, the relative cross-sections between new physics processes and background processes (such as inelastic scattering and multijet production) is much higher than at LEP.

In the $p\bar{p}$ collisions at the Tevatron, coloured SUSY particles will have the highest production cross-section. Both squark and gluino production are possible depending on the SUSY model, but in both cases the final state consists of a number of jets and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). Backgrounds to these searches are of two types, those involving real $E_{\mathrm{T}}^{\mathrm{miss}}$ (e.g. $W \to l\nu + \mathrm{jets}$) and those involving fake $E_{\mathrm{T}}^{\mathrm{miss}}$ from jet mismeasurement. The exclusion results from CDF and DØ are interpreted in the mSUGRA framework in Figures 2.2 and 2.3.

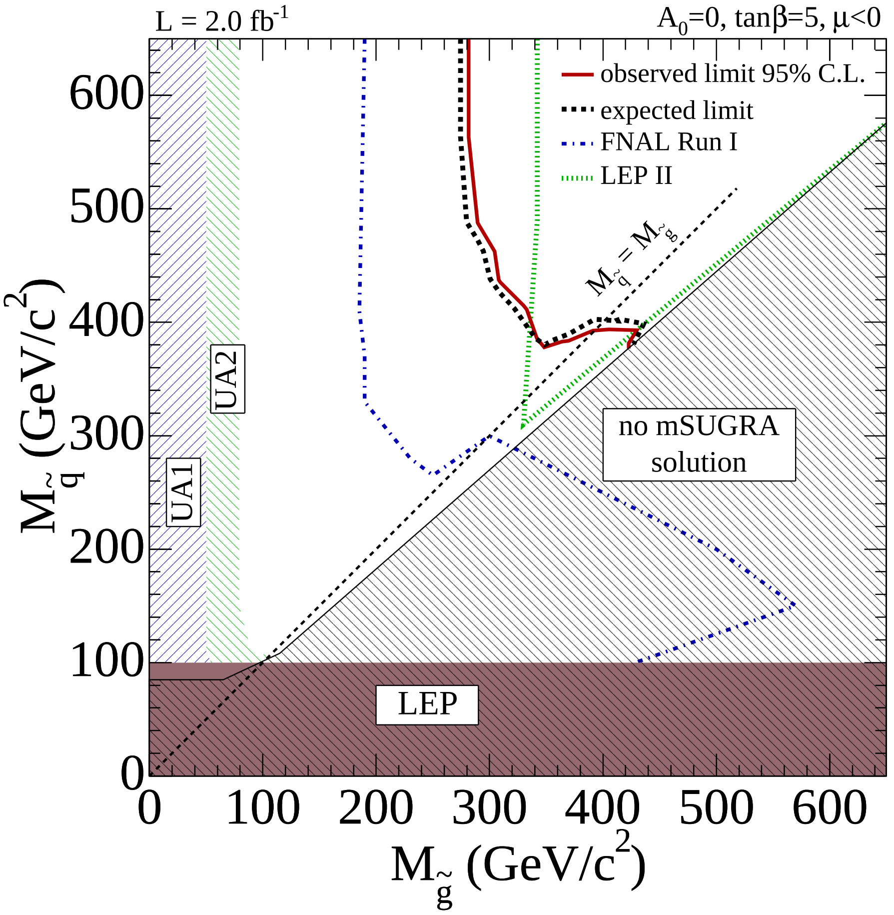


Figure 2.2: mSUGRA exclusion limits in the $m_{\tilde{g}}, m_{\tilde{q}}$ plane from the search for squarks and gluinos using CDF. From [22].

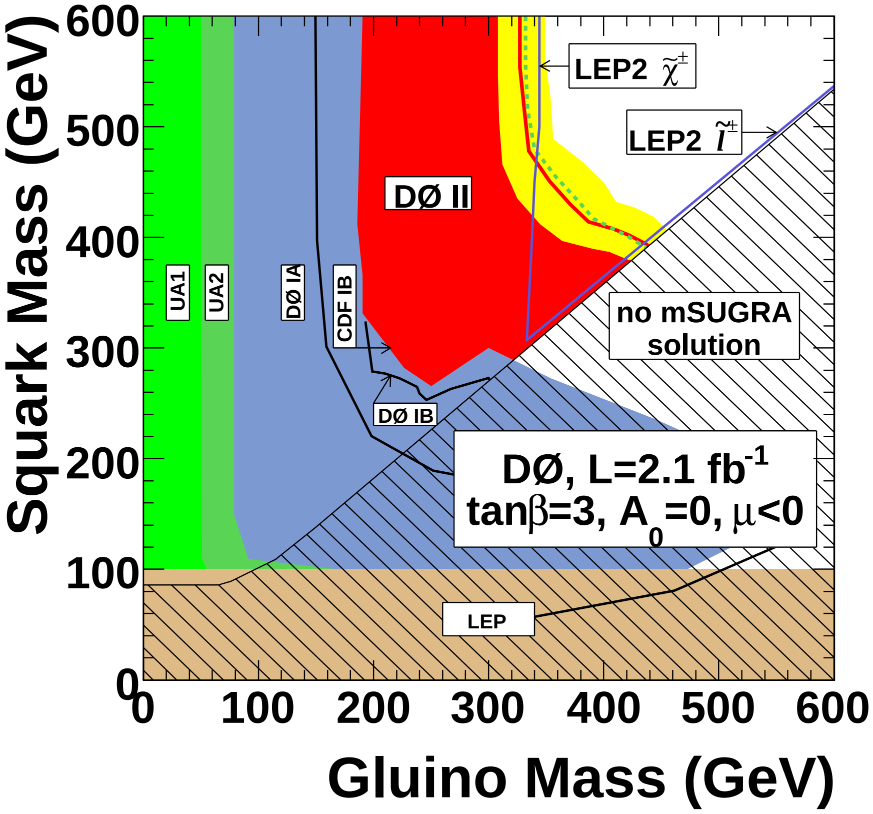


Figure 2.3: mSUGRA exclusion limits in the $m_{\tilde{g}}, m_{\tilde{q}}$ plane from the search for squarks and gluinos using DØ. From [23].

2.3.2 SUSY Searches at the LHC

Direct searches have been performed by the ATLAS and CMS collaborations. SUSY searches at ATLAS can be broadly split into two types:

- Signatures involving significant $E_{\mathrm{T}}^{\mathrm{miss}}$ and multiple jets. These are the primary discovery channels for most general SUSY models and include the zero lepton analysis which is described in more detail in Chapter 5. They are defined by the number of leptons in the final state and whether b-jets are required. In these searches, the primary production mechanisms are squark and gluino pair production since they couple to the strong interaction.

- Searches for long-lived exotic massive particles and for signatures of R-parity violating SUSY. There are a number of signatures which can be studied, including searching for particles which stop in the detector and decay at a later time, monojet analyses and searches for highly displayed vertices.

The searches with jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ are most relevant to this thesis and are summarised in this section.

The analysis involving jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and no leptons [5] is described in Chapter 5. It provides the largest discovery reach for squark and gluinos as discussed in Chapter 5. A related search is for events containing two b-jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and no leptons [7]. This search is particularly sensitive to the pair production of scalar bottom quarks (the sparticle equivalent of b-quarks) decaying to b-quarks and neutralinos. In MSSM models where this is the exclusive decay mechanism for scalar bottom quarks, the mass of these squarks is excluded beneath 390 GeV for neutralino masses below 60 GeV. ATLAS has also performed a search for high jet multiplicities (≥ 6 jets) with no leptons [9]; unlike other searches described in this section there is no direct $E_{\mathrm{T}}^{\mathrm{miss}}$ cut used. This search is sensitive to SUSY models predicting many-body decays or large cascade decays to partons. It extends the exclusion reach in mSUGRA/cMSSM models in certain regions of the $m_0, m_{1/2}$ plane above that found in the main jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and no lepton analysis.

Searches involving jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and leptons are characterised by the lepton multiplicity. Squarks and gluinos are expected to decay into SM quarks and charginos or neutralinos. The charginos ($\tilde{\chi}_i^{\pm}$) and neutralinos ($\tilde{\chi}_i^0$) can decay into SM leptons through the following main processes: $\tilde{\chi}_i^0 \rightarrow l^{\pm}\nu\tilde{\chi}_j^{\mp}$, $\tilde{\chi}_i^{\pm} \rightarrow l^{\pm}\nu\tilde{\chi}_j^0$, $\tilde{\chi}_i^0 \rightarrow l^{\pm}l^{\mp}\tilde{\chi}_j^0$ and $\tilde{\chi}_i^{\pm} \rightarrow l^{\pm}l^{\mp}\tilde{\chi}_j^{\pm}$, where $i > j$ and $i > 0$. It should be noted that l refers to any SM lepton but the searches described in this section only deal with electrons and muons due to the short lifetime of the tau lepton. The processes shown above allow for a variety of final state lepton multiplicities. The one lepton search [24] searches for an excess of events with one lepton, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ and produces exclusion limits in mSUGRA/CMSSM models which are not as stringent as those from the no lepton search. The two lepton search [25] is split into searches for same-sign leptons and opposite-sign leptons. The opposite-sign search also looks for an excess of same-flavour over different-flavour lepton pairs. Limits are set in a simplified model (see §2.2) of weak gaugino (the sparticle partner of a SM gauge boson) production. Finally, the multi-lepton search [26] uses signatures with four or more leptons to provide cross-section limits on processes creating four or more leptons (with and without a veto on Z bosons).

CMS also has a wide programme for SUSY searches including searches involving leptons, b-jets and Z bosons. The fully hadronic searches (without a b-jet requirement) are the most relevant to this thesis. These searches at CMS are split into four different search strategies:

- Using the α_{T} variable [27]. α_{T} is a kinematic variable which is designed to distinguish between real and fake $E_{\mathrm{T}}^{\mathrm{miss}}$.
- Using the M_{T2} variable [28]. M_{T2} was originally introduced to measure sparticle masses but in this search it is used as a discovery variable.
- Using razor variables [29]. This search is designed to distinguish between pair produced heavy particles (like squarks and gluinos) and pair produced SM particles.
- Using missing energy [30].

In all cases, no evidence of SUSY is found so exclusion limits are published. Figure 2.4 gives an overview of the exclusions from a variety of different searches at CMS; the fully hadronic searches mentioned above are the most sensitive.

2.4 Jets at Hadron Colliders

The theory of QCD is briefly outlined in §2.1.2 and it is noted that high energy partons form jets in a detector. This section will expand on this process as it is extremely

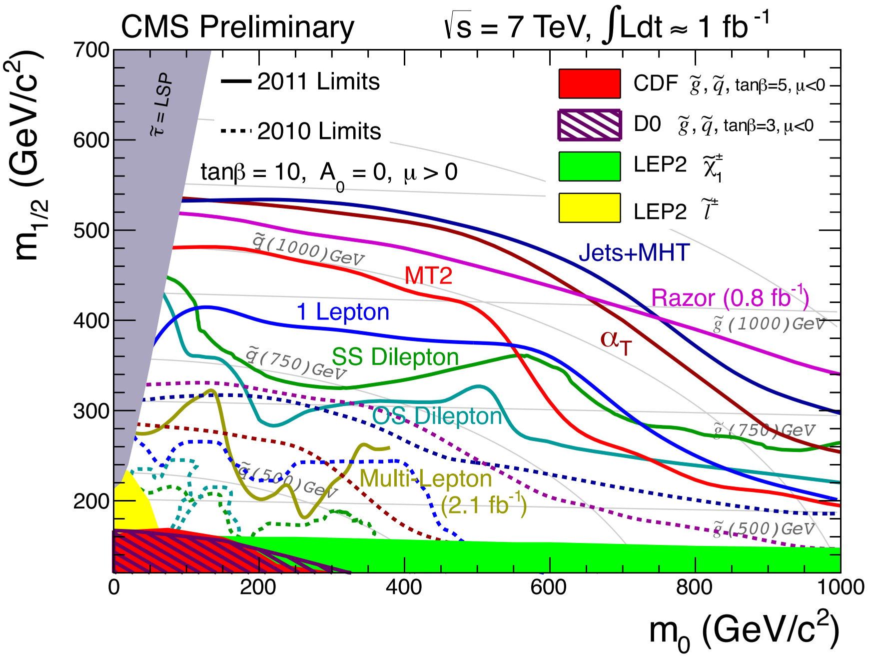


Figure 2.4: Observed limits from searches for SUSY at CMS plotted in the CMSSM plane. From [31].

relevant to this thesis. Taking the example of a 'simple' $qq \longrightarrow qq$ hard scatter at the LHC: firstly, two hard quarks are produced in the interaction. Additional partons are radiated from the incoming quarks (initial state radiation) and from the outgoing quarks (final state radiation). The other partons in the colliding protons will also interact (multi-parton interactions). This is generally, with the initial state radiation, referred to as the "underlying event". Any outgoing partons from the hard interaction or radiation will then 'shower'; this accounts for quark-antiquark pairs produced from gluons and gluon radiation from gluons and quarks. Each shower of outgoing partons will then hadronise to form a co-linear collection of colour neutral hadrons. These collections of hadrons are referred to as jets and will interact in the tracking system and deposit energy in the calorimetry of a particle detector.

It is impossible to uniquely define what constitutes a jet in a detector. For example, a jet may correspond to the group of hadrons produced from a single outgoing hard parton, but it may also be from a number of co-linear hard partons or even a boosted W decaying into a pair of quarks. In these cases, the object reconstructed in the detector can be indistinguishable. Therefore, experiments must decide their criteria for a jet and understand that this does not map directly to a parton produced in the pp interaction. Jets are then defined in experiments from the choice of jet algorithm. These use input from either theoretical objects (simulated partons or hadrons) or detector quantities like energy deposits in the calorimetry or tracks. The following are desirable properties for jet algorithms:

- Collinear safe. This means that a reconstructed jet should not be affected if a parton in the jet splits into two collinear partons. Non-collinear safe algorithms produce divergences in theoretical calculations.
- Infra-red safe. The reconstructed jet must be unaffected by the emission of a small amount of soft radiation. As with collinear safety, if the algorithm is not infra-red safe then theoretical calculations contain divergences.
- Well-defined area. To allow simple isolation and overlap removal to be applied, it is desirable that jets have well defined areas, usually in a conical shape.
- Computationally fast.

ATLAS uses the anti-k_t algorithm [32] for jet reconstruction. This algorithm satisfies all of the above desirable criteria for a jet algorithm and provides a natural replacement for “iterative cone” algorithm which is collinear unsafe.

2.5 Multijet Background

At the Tevatron and LHC, the cross section for multijet production is extremely large. For example, Figure 2.5 shows that, at $\sqrt{s} = 7$ TeV, the cross section for jet production with $E_{\mathrm{T}}^{\mathrm{jet}} > 100$ GeV is higher than the cross section for W and Z boson production. Multijet production refers to any reconstructed events with multiple jets and no prompt isolated electrons or muons and is dominated by processes mediated by the strong force. Multijet events can contain significant missing energy through jet mismeasurement and due to the presence of neutrinos in heavy flavour jets. Therefore it is a potentially significant background for fully hadronic SUSY searches. This section will briefly discuss the various strategies used at the Tevatron and LHC experiments.

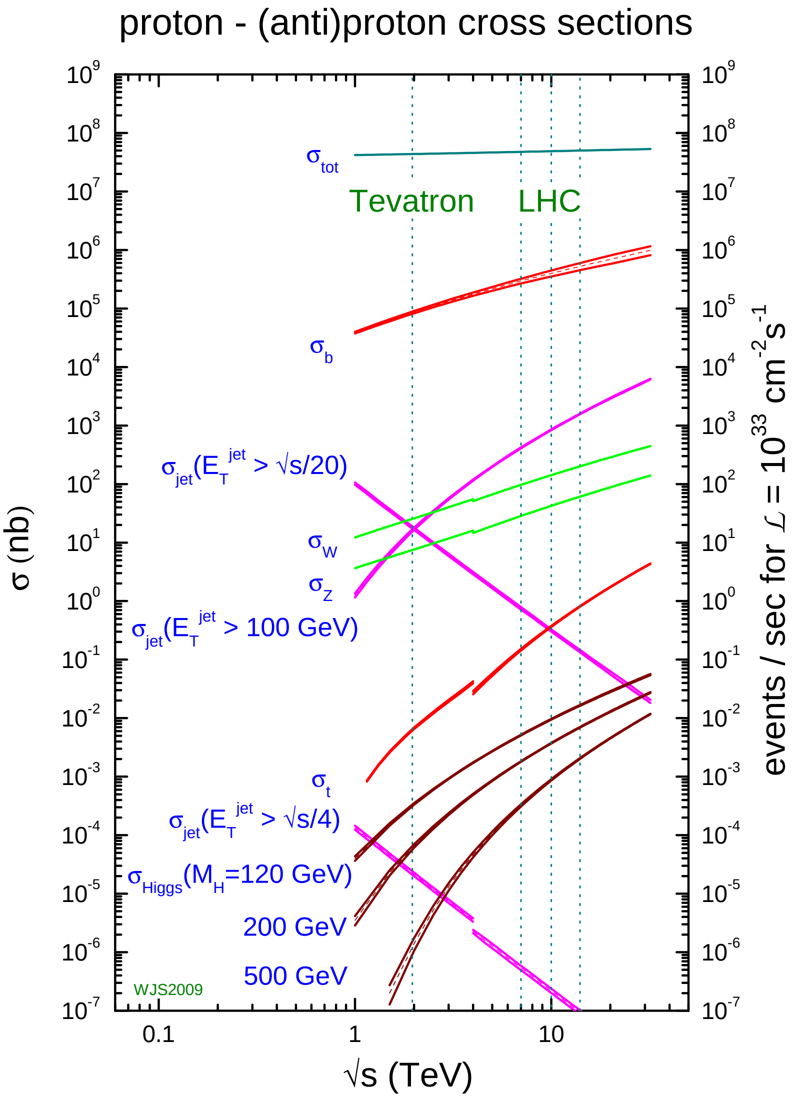


Figure 2.5: Cross sections of interesting SM and proposed processes at hadron colliders as a function of centre of mass energy. From [33].

2.5.1 Multijet Strategy at Tevatron Searches

Full details of the fully hadronic search at CDF are given in [22]. At CDF, high $E_{\mathrm{T}}^{\mathrm{miss}}$ multijet events are primarily due to significant jet mismeasurement and not from heavy flavour jets containing neutrinos. In these events the $E_{\mathrm{T}}^{\mathrm{miss}}$ will most likely align along the direction of one of the jets in the event. For this reason, CDF cut on the azimuthal angle between the jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ for selected jets in the event. The remaining multijet background is estimated from Monte Carlo (MC) and the validity of the MC estimate is checked in multijet dominated control regions. The multijet background is significant in the signal regions considered, as shown in Table 2.3.

Events in data (2 fb^{-1})			
	≥ 2 jets	≥ 3 jets	≥ 4 jets
	18	38	45
SM predictions			
QCD jets	4.4 ± 2.0	13.3 ± 4.6	15.3 ± 7.1
top	1.3 ± 1.2	7.6 ± 4.1	22.1 ± 7.0
$Z \rightarrow \nu\bar{\nu}+$ jets	3.9 ± 0.9	5.4 ± 1.4	2.7 ± 0.7
$Z/\gamma* \rightarrow l^+l^-+$ jets	0.1 ± 0.1	0.2 ± 0.1	0.1 ± 0.1
$W \rightarrow l\nu+$ jets	6.1 ± 2.2	10.7 ± 3.1	7.7 ± 2.2
WW, ZW, ZZ	0.2 ± 0.2	0.3 ± 0.2	0.5 ± 0.2
Total SM	16 ± 5	37 ± 12	48 ± 17

Table 2.3: Data and SM expectations in the fully hadronic SUSY search at CDF. Adapted from [22].

Full details of the fully hadronic search with DØ can be found in [23]. As at CDF, the multijet background at DØ is primarily due to calorimetric jet mismeasurement and cuts are applied on the azimuthal angle between the jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ to significantly reduce the background. DØ estimate the remaining multijet background using a data-driven method where the $E_{\mathrm{T}}^{\mathrm{miss}}$ spectrum is fitted with an exponential function in the low-$E_{\mathrm{T}}^{\mathrm{miss}}$ region (after the subtraction of other non-multijet sources of $E_{\mathrm{T}}^{\mathrm{miss}}$). The fit is then extrapolated to the high $E_{\mathrm{T}}^{\mathrm{miss}}$ signal regions providing an estimate of the multijet background in those regions. In all of the signal regions considered, the estimated background is sufficiently small to be ignored in setting limits.

2.5.2 Multijet Strategy at LHC Searches

The multijet estimation in fully hadronic searches at ATLAS is the subject of this thesis and will be described in detail in Chapters 6 to 8. At CMS, the four searches introduced in §2.3.2 have differing strategies to the multijet background. The searches using the α_{T} [27], M_{T2} [28] and razor [29] variables rely on reducing the background to a negligible level using a combination (depending on the search) of azimuthal cuts between the jets and $E_{\mathrm{T}}^{\mathrm{miss}}$, hard cuts on the discovery variable and b-tagging. The missing energy search [30] also uses an azimuthal cut between the jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ to reduce the multijet background and the remaining background is estimated using a "Rebalance and Smear" method. This method uses similar techniques to those described in this thesis: jets in a sample of data events with low $E_{\mathrm{T}}^{\mathrm{miss}}$ are scaled (or smeared) with a data-measured response function to produce events with potentially high $E_{\mathrm{T}}^{\mathrm{miss}}$. This sample is used as the multijet estimation in the signal regions. Table 2.4 shows the data observation and estimated backgrounds for the signal regions in this search; it can be seen that the multijet background is small compared to the other SM processes.

	Baseline ($H_{\mathrm{T}} > 350$ GeV) ($E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV)	Medium ($H_{\mathrm{T}} > 500$ GeV) ($E_{\mathrm{T}}^{\mathrm{miss}} > 350$ GeV)	High H_{T} ($H_{\mathrm{T}} > 800$ GeV) ($E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV)	High H_{T} ($H_{\mathrm{T}} > 800$ GeV) ($E_{\mathrm{T}}^{\mathrm{miss}} > 500$ GeV)
$Z \to \nu\bar{\nu}$ from γ+jets	$376 \pm 12 \pm 79$	$42.6 \pm 4.4 \pm 8.9$	$24.9 \pm 3.5 \pm 5.2$	$2.4 \pm 1.1 \pm 0.5$
$t\bar{t}/W \to e, \mu + X$	$244 \pm 20^{+30}_{-31}$	$12.7 \pm 3.3 \pm 1.5$	$22.5 \pm 6.7^{+3.0}_{-3.1}$	$0.8 \pm 0.8 \pm 0.1$
$t\bar{t}/W \to \tau_{\mathrm{h}} + X$	$263 \pm 8 \pm 7$	$17 \pm 2 \pm 0.7$	$18 \pm 2 \pm 0.5$	$0.73 \pm 0.73 \pm 0.04$
QCD	$31 \pm 35^{+17}_{-6}$	$1.3 \pm 1.3^{+0.6}_{-0.4}$	$13.5 \pm 4.1^{+7.3}_{-4.3}$	$0.09 \pm 0.31^{+0.05}_{-0.04}$
Total background	928 ± 103	73.9 ± 11.9	79.4 ± 12.2	4.6 ± 1.5
Observed in data	986	78	70	3

Table 2.4: Data and SM expectations in fully hadronic SUSY search with missing energy at CMS using 1.1 fb^{-1}. H_{T} is defined as the scalar sum of the transverse momenta of the jets in an event. Adapted from [30].

Chapter 3

The LHC and the ATLAS Detector

This chapter provides a brief introduction to the Large Hadron Collider (LHC) and the ATLAS detector. §3.1 describes the LHC, notable achievements of the collider and its future goals. §3.2 describes the ATLAS detector; emphasis is given to the systems which allow jet and missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) reconstruction and to the trigger system (see §3.2.5). ATLAS provides precision energy measurements of electromagnetic and hadronic showers and is near-hermetic, making it an ideal detector for measuring jets and $E_{\mathrm{T}}^{\mathrm{miss}}$.

3.1 The Large Hadron Collider

The LHC is the highest energy collider experiment ever built. It is a proton-proton collider running at a centre-of-mass energy of 7 TeV and in the dataset considered in this thesis, reached a peak luminosity of $\sim 10^{33}$ cm^{-2}s^{-1}. It is housed in the 27 km tunnel which was previously home to the LEP collider at CERN. It also uses many of the existing accelerator facilities of CERN. Proton beams are collided at four locations around the ring where the main LHC detector experiments are located: ALICE, ATLAS, CMS and LHCb. Some of the design parameters of the LHC are shown in Table 3.1. A schematic of the accelerator complex is shown in Figure 3.1.

Parameter	Value
Beam energy at collision	7 TeV
Beam energy at injection	0.45 TeV
Machine Circumference	26658.833 m
Dipole Field at 7 TeV	8.33 T
Luminosity	10^{34} cm$^{-1}s^{-1}$
RAMS Bunch length	7.55 cm
Number of Particles per Bunch	1.15×10^{11}
Number of bunches per beam	2808
Time between bunches at nominal luminosity	25 ns
Circulating beam current	0.1582
Dipole magnet temperature	1.9 K
Number of dipole magnets	≈ 1232
Number of quadrupole magnets	≈ 600
Number of corrector magnets	≈ 7000

Table 3.1: Machine design parameters for the LHC. Adapted from [34].

The LHC has a vast physics programme to fully utilise the extremely high energy and luminosity delivered. Primary goals include the observation of the Higgs boson (see §2.1.3), the search for physics beyond the SM, including SUSY (see §2.2), and for precision measurements of SM processes. In order to realise these goals, general purpose (ATLAS and CMS) and specialised (ALICE and LHCb) particle detectors are located at collision points on the LHC ring. The work in this thesis is performed using the ATLAS detector.

During 2010 and 2011, the LHC ran with beam energies of 3.5 TeV per beam pushing particle physics into a new energy realm. At the time of writing, the LHC has now delivered over 5 fb^{-1} of data to the ATLAS and CMS detectors. This has allowed precision SM physics to be performed and many models for physics beyond the SM have been tested to extents beyond what was possible using results from the Tevatron. Particular successes of the LHC are the achievement of the highest peak luminosity at a hadron collider [36] and limiting the available mass range of the SM Higgs to a small window between ~ 115 GeV and ~ 141 GeV [14, 37]. In 2012, the LHC will run with an energy of 4 TeV per beam [38] and aims to collect 15 fb^{-1} in order to finalise the search for the SM Higgs boson. After 2012, the LHC will undergo

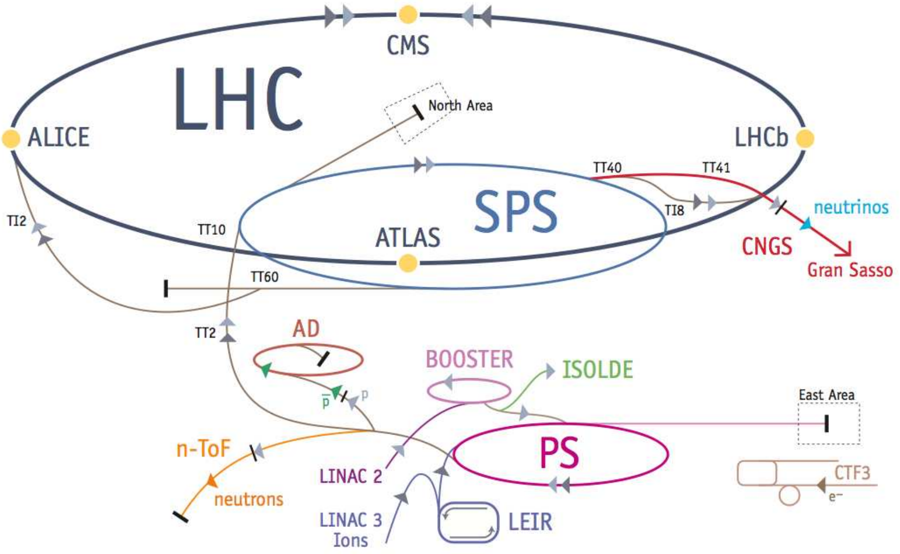


Figure 3.1: Schematic of the LHC accelerator complex. From [35].

a period of repairs and upgrades to allow it to reach its design energy in 2014-2015.

3.2 The ATLAS Detector

The ATLAS detector, “A Toroidal LHC Apparatus” is amongst the most complex detectors ever built. It is designed as a general purpose detector and is optimised to search for the Higgs boson, physics beyond the SM and to perform precision SM measurements. To allow this, it has a near hermetic design with precise particle tracking, calorimetry and muon detection systems. The sub-detector most important in the reconstruction of jets and missing energy is the calorimeter (see §3.2.3); however, the other sub-detectors are also essential. A schematic of the entire detector is shown in Figure 3.2. Full details of the detector can be found in [39].

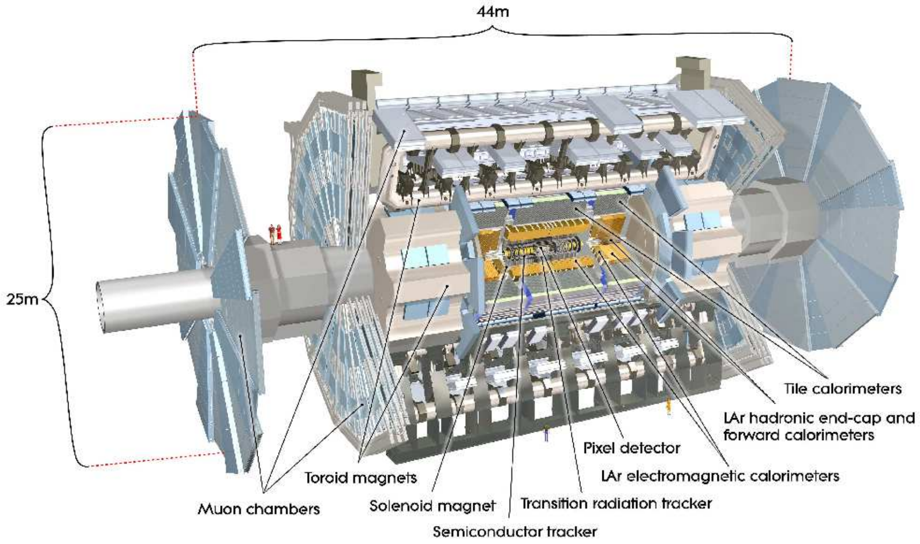


Figure 3.2: Schematic of the ATLAS detector. From [39].

3.2.1 Geometry and Transverse Quantities

The z-axis of the ATLAS detector is aligned along the beam pipe. The x-axis points towards the centre of the LHC ring and the y-axis points vertically upwards. However, as a cylindrical detector, cylindrical polar coordinates are commonly used where the azimuthal angle ϕ is measured around the z-axis and the polar angle θ is the angle from the z-axis. Pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$. This is widely used instead of θ as it is invariant under Lorentz boosts. The distance ΔR in $\eta - \phi$ space is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ and is often used to define the separation of physics objects reconstructed in the detector.

At the LHC, the hard scatter in a collision occurs between partons which carry an unknown proportion of the momenta of the protons. This means that momentum conservation can only be applied transverse to the beam axis. For this reason, transverse quantities such as p_{T} (transverse momentum) are commonly used.

3.2.2 Inner Detector

The purpose of the inner detector is to precisely reconstruct the tracks of charged particles originating from (or near to) the interaction point. To allow this, three sub-detectors are used: the pixel detector, semiconductor tracker (SCT) and transition radiation tracker (TRT). The pixel and SCT (the precision trackers) cover a pseudo-rapidity range of $|\eta| < 2.5$. The entire inner detector is immersed in a 2 T magnetic field created by the central solenoid magnet. This magnetic field results in charged particles following curved trajectories, allowing their momenta to be measured from the radius of curvature. A schematic of the inner detector is shown in Figure 3.3.

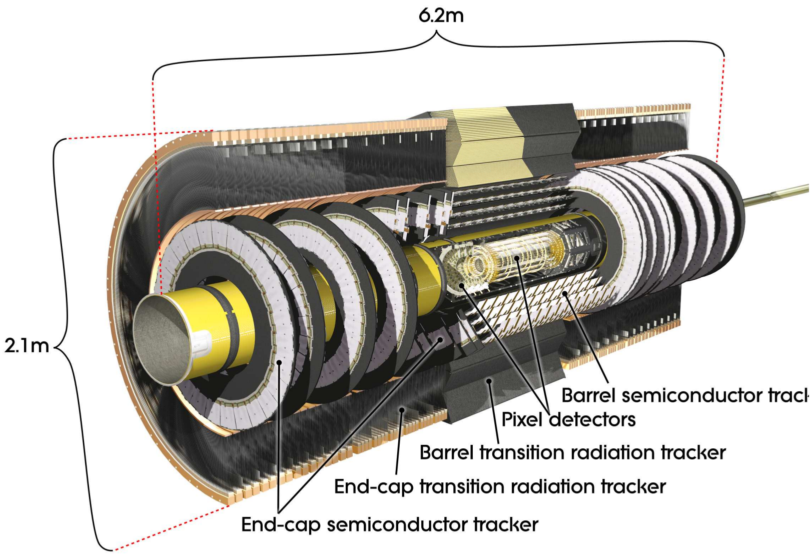


Figure 3.3: Schematic of the ATLAS inner detector. From [39].

Pixel Detector

The pixel detector is the closest detector to the interaction point and has the highest granularity of any of the sub-detectors in ATLAS. This is primarily to allow accurate locating of any secondary vertices in an event. The pixel detector contains three layers of silicon pixel arrays which have minimum pixel size in $R-\phi \times z$ space of

50×400 μm^2 (R is the perpendicular distance from the z axis). This allows an accuracy of 10 μm $\times$ 115 μm in $R - \phi \times z$ space.

SCT

The SCT provides further, coarser tracking measurements than the pixel detector. It uses layers of paired silicon microstrips, with 80 μm pitch, to give three-dimensional tracking. Particles cross eight silicon layers through the SCT providing four space-points for track reconstruction. The resolution of the SCT in the barrel is 17 μm $\times$ 580 μm in $R - \phi \times z$ space.

TRT

The TRT uses straw tubes to provide tracking measurements in $R - \phi \times z$ space. The TRT provides a large number of hits (36 per track) which makes a significant impact to momenta measurements of tracks. It also assists particle identification as the amount of transition radiation increases with particle velocity; therefore at a given energy electrons can be identified as the particles which produce the most transition radiation.

3.2.3 Calorimetry

ATLAS uses two sampling calorimeter systems (electromagnetic and hadronic) to accurately measure the energy deposited by different particles in the detector. The calorimeter is the most important component in the measurement of jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ in ATLAS. It provides precision measurements in the central region of the detector, $|\eta| < 2.5$, and coverage up to $|\eta| < 4.9$. This large range is vital in the measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$ as it ensures that the energies of all particles with significant transverse momenta can be measured. The calorimeter also has large depth (approximately ten interaction lengths) to limit punch-through of high momenta particles into the muon system and to fully contain hadronic showers. A schematic of the ATLAS calorimeter is shown in Figure 3.4.

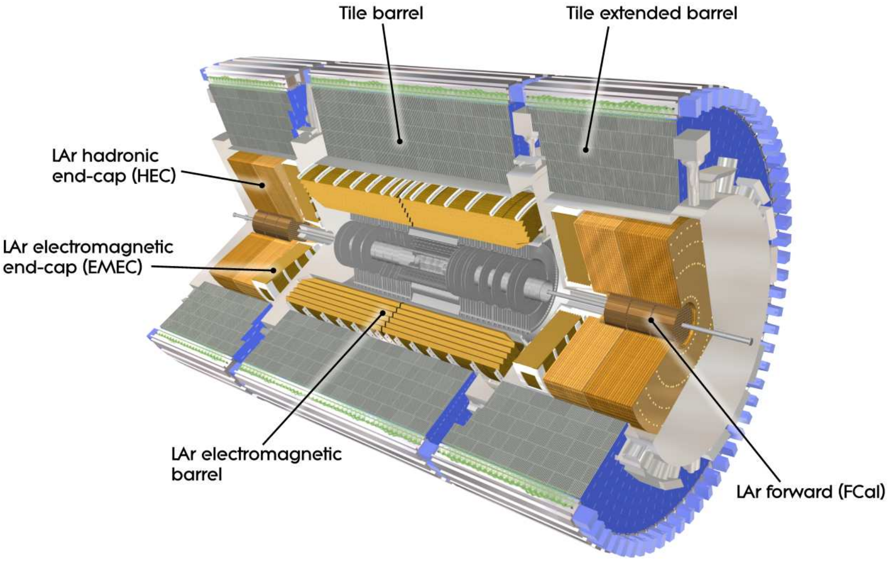


Figure 3.4: Schematic of the ATLAS calorimeter. From [39].

Both calorimeters are made up from a large number of individual cells and particles traversing the calorimetry will deposit energy in many of these cells. To provide an input to jet algorithms (see §2.4), the cells are grouped into "clusters" with a clustering algorithm. Examples of clustering algorithms are the sliding window and topological cluster algorithms.

Electromagnetic Calorimeter

The electromagnetic calorimeter is divided into a barrel region covering $|\eta| < 1.475$ and two end-caps covering $1.375 < |\eta| < 3.2$. The active material is liquid-argon and the absorber is lead. Accordion-shaped kapton electrodes provide the readout, which has complete ϕ coverage. There are three sampling layers in the $|\eta| < 2.5$ region and two in the $2.5 < |\eta| < 3.2$ region. As shown in Table 3.2, the first sampling layer in the central region has very fine η resolution to allow accurate estimation of the η coordinate of an electromagnetic shower. The second layer collects the majority of the energy of a shower and the third collects the tail of the shower with reduced granularity. There is a presampler present in the region $|\eta| < 1.8$ to allow corrections for energy lost in material upstream of the calorimeter. It should be noted that the majority of the energy in a typical hadronic shower is absorbed in the electromagnetic calorimeter so this sub-detector is vital in measuring both electromagnetic and hadronic objects. For much of the dataset used in this thesis, an electronics failure resulted in there being a dead region in two layers of the barrel region of the electromagnetic calorimeter. This is referred to as the "LAr hole". It affects the reconstruction of any physics objects which use the electromagnetic calorimeter including electrons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$.

EM calorimeter	**Barrel**	**End-cap**	
Coverage	$\|\eta\| < 1.475$	$1.375 < \|\eta\| < 3.2$	
Longitudinal segmentation	3 samplings	3 samplings	$1.5 < \|\eta\| < 2.5$
		2 samplings	$1.375 < \|\eta\| < 1.5$
			$2.5 < \|\eta\| < 3.2$
Granularity ($\Delta\eta \times \Delta\phi$)			
Sampling 1	0.003×0.1	0.025×0.1	$1.375 < \|\eta\| < 1.5$
		0.003×0.1	$1.5 < \|\eta\| < 1.8$
		0.004×0.1	$1.8 < \|\eta\| < 2.0$
		0.006×0.1	$2.0 < \|\eta\| < 2.5$
		0.1×0.1	$2.5 < \|\eta\| < 3.2$
Sampling 2	0.025×0.025	0.025×0.025	$1.375 < \|\eta\| < 2.5$
		0.1×0.1	$2.5 < \|\eta\| < 3.2$
Sampling 3	0.05×0.025	0.05×0.025	$1.5 < \|\eta\| < 2.5$

Presampler	**Barrel**	**End-cap**
Coverage	$\|\eta\| < 1.52$	$1.5 < \|\eta\| < 1.8$
Longitudinal segmentation	1 sampling	1 sampling
Granularity ($\Delta\eta \times \Delta\phi$)	0.025×0.1	0.025×0.1

Table 3.2: Parameters of the electromagnetic calorimeter. Adapted from [39].

Hadronic Calorimeter

The hadronic calorimeter consists of three subsystems: the tile calorimeter, the hadronic end-cap (HEC) and the forward calorimeter (FCal). The tile calorimeter covers the range $|\eta| < 1.7$. It uses steel absorbers with scintillating tiles as the active medium and has three layers. The HEC uses LAr as the active medium and copper as the absorber and is arranged to provide four sampling layers per end-cap. It covers the pseudorapidity range $1.5 < |\eta| < 3.2$. The FCal (covering $3.1 < |\eta| < 4.9$) also uses LAr as the active medium and uses copper (in the inner module) and tungsten (in the outer two modules) as the absorber. The overlap between the tile and FCal reduces the loss in material density. Table 3.3 shows some key parameters of the hadronic calorimeter. Coverage is provided up to $|\eta| = 4.9$ with at least 3 sampling layers providing near-hermetic coverage.

3.2.4 Muon System

The muon system of ATLAS forms the outer part of the detector and a significant fraction of the detector's volume. Superconducting air-core toroid magnets provide up to 7.4 Tm of bending power meaning that even the momenta of the highest energy muons can be determined. There are four detection sub-systems in the muon system: monitored drift tubes (MDT), cathode strip chambers (CSC), resistive plate chambers (RPC) and thin gap chambers (TPG). The MDT and CSC provide the precision tracking measurements whilst the RPC and TPG form the muon trigger system and give a second coordinate for muon track reconstruction.

3.2.5 Trigger and Data Acquisition

The design luminosity of 10^{34} cm^{-2}s^{-1} translates to an event rate of $\sim 10^9$ Hz; however, computing resources mean that only ~ 200 Hz can be permanently stored. This means the trigger system must select roughly 1 in 5×10^6 events. This is made possible through a three stage trigger system consisting of one hardware trigger

	Barrel	**Extended barrel**
Scintillator tile calorimeter		
$\lvert\eta\rvert$ coverage	$\lvert\eta\rvert < 1.0$	$0.8 < \lvert\eta\rvert < 1.7$
Number of layers	3	3
Granularity $\Delta\eta \times \Delta\phi$	0.1×0.1	0.1×0.1
Last layer	0.2×0.1	0.2×0.1

	Barrel	**End-cap**	
LAr hadronic end-cap			
$\lvert\eta\rvert$ coverage		$1.5 < \lvert\eta\rvert < 3.2$	
Number of layers		4	
Granularity $\Delta\eta \times \Delta\phi$		0.1×0.1	$1.5 < \lvert\eta\rvert < 2.5$
		0.2×0.2	$2.5 < \lvert\eta\rvert < 3.2$
LAr forward calorimeter			
$\lvert\eta\rvert$ coverage		$3.1 < \lvert\eta\rvert < 4.9$	
Number of layers		3	
Granularity $\Delta x \times \Delta y$ (cm)		FCal1: 3.1×2.6	$3.15 < \lvert\eta\rvert < 4.30$
		FCal1: $\sim$ four times finer	$3.10 < \lvert\eta\rvert < 3.15$, $4.30 < \lvert\eta\rvert < 4.83$
		FCal2: 3.3×4.2	$3.24 < \lvert\eta\rvert < 4.50$
		FCal2: $\sim$ four times finer	$3.20 < \lvert\eta\rvert < 3.24$, $4.50 < \lvert\eta\rvert < 4.81$
		FCal3: 5.4×4.7	$3.32 < \lvert\eta\rvert < 4.60$
		FCal3: $\sim$ four times finer	$3.29 < \lvert\eta\rvert < 3.32$, $4.60 < \lvert\eta\rvert < 4.75$

Table 3.3: Parameters of the hadronic calorimeter. Adapted from [39].

(level 1) and two software triggers (level 2 and event filter, collectively known as the High Level Trigger (HLT)). The trigger works at the electromagnetic energy scale (no corrections are applied for hadrons).

The level 1 (L1) trigger uses calorimeter information (with reduced granularity of 0.1×0.1 in $\Delta\eta \times \Delta\phi$ in most regions) and the muon trigger system to trigger on high momenta leptons, jets and event quantities such as $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\sum E_{\mathrm{T}}$. The L1 trigger must decide within 2.5 μs whether an event should be passed to L2 for further analysis. Given this need for extremely fast processing and the lack of full detector infromation at L1, thresholds applied to physics objects are looser than appiled in the HLT. There is a maximum L1 output rate $\sim$ 75 kHz. The L1 trigger also defines regions of interest (RoIs) in $\eta - \phi$ space for the level 2 trigger to analyse. These RoIs are located in the regions of the detector where the L1 trigger finds objects satisfying

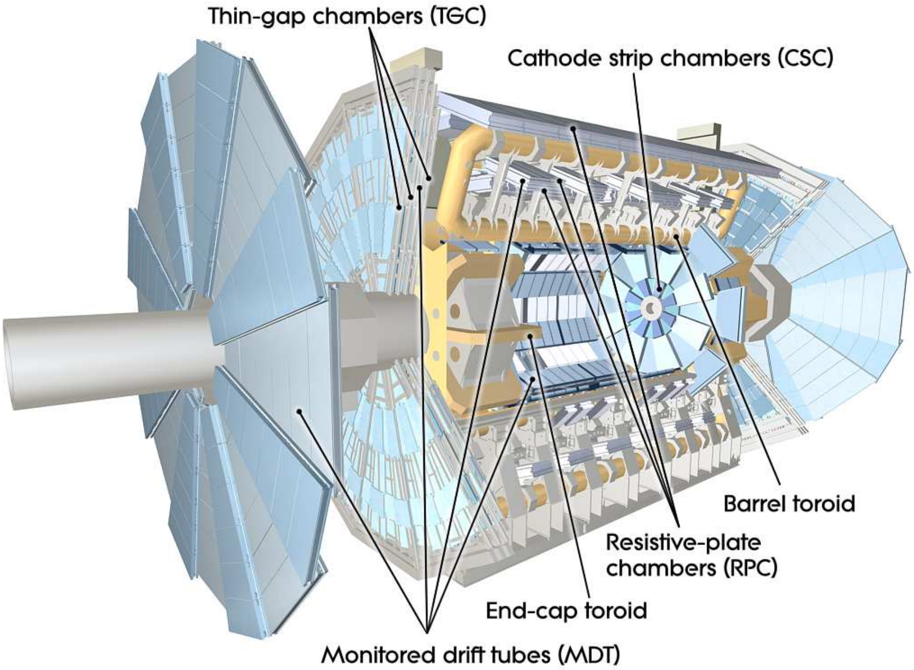


Figure 3.5: Schematic of the the ATLAS muon system. From [39].

energy thresholds defined in the L1 trigger menu (see §4.4 for information on trigger menus).

The level 2 (L2) trigger uses the full detector granularity but only investigates the RoIs identified by the L1 trigger. Software algorithms are used which are more sophisticated than the L1 algorithms but not yet at the level of the offline algorithms due to the requirement for fast processing of ~ 10 ms. The L2 output rate is ~ 1 kHz.

The event filter (EF) trigger uses the full event information and some offline algorithms to analyse the events passed by the L2 trigger to reduce the output rate to ~ 200 Hz. The trigger has ~ 1 s to make its decision and hence there is insufficient time to perform processing identical to that done offline; for example there is no Jet Energy Scale correction (see §4.3.3) applied to event filter jets.

Chapter 4

Analysis Tools

This chapter describes the analysis tools common to much of the work described in this thesis. §4.1 describes the dataset used throughout. §4.2 gives an overview of the simulation and reconstruction framework used by ATLAS and details the samples of Monte Carlo simulated data used. §4.3 describes how physics objects are reconstructed covering objects relevant to this thesis, overlap removal and event cleaning. In §4.4 and §4.5, information is given on trigger menus and the single jet triggers used extensively in the analysis detailed in this thesis. Finally, §4.6 describes the full chain testing performed at ATLAS to check for problems in the full production chain; the author performed service work for the experiment in this area.

4.1 Data Samples

The data used in this analysis were recorded by the ATLAS experiment from 22nd March 2011 to 28th June 2011 corresponding to an integrated luminosity of 1.0 fb^{-1} after the application of the official SUSY group Good Run List (GRL)[1]. ATLAS splits its collected data events into different 'streams' depending on which triggers are fired

---

[1] A GRL indicates which periods of data taking are appropriate for a particular physics analysis, with the appropriate sub-detectors active and functioning correctly.

by an event. This is convenient as events useful for lepton physics are not generally useful for fully hadronic searches. This thesis uses the ‘JetTauEtMiss’ stream, which contains events where jet, τ lepton or $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers have fired.

4.2 Simulation and Reconstruction Software

Event simulation consists of three stages: event generation, detector simulation and digitisation. Event generation produces the idealised, particle level event before detector effects are included. Detector simulation then passes these events through a model of the full detector using `GEANT4` [40]. The digitisation is designed to match the real detector so that the input to reconstruction is identical for real and simulated data.

Monte Carlo (MC) samples are produced using the available generator which is most appropriate for the required final state and the number of events needed. Generators used at ATLAS compute either the Leading-Order (LO) or Next-to-Leading-Order (NLO) contributions to a given process. Furthermore, some generators only compute the matrix elements for the hard interaction process whilst others calculate the elements for extra radiated hard partons. All generators require the input of Parton Density Functions (PDFs), which describe the probability of a collided parton having a given momentum fraction of its parent proton. At ATLAS, LO generators use the `MRST2007LO*` modified LO PDFs [41] and NLO generators use `CTEQ6.6` PDFs [42]. As well as simulating the hard process, generators also need to simulate the parton showers, hadronisation, decays and the underlying event (see §2.4). The generators used for MC samples in this thesis are described below:

- Multijet production: `PYTHIA` [43]. This generator works at LO and only calculates the matrix elements for the hard process. This means that it may not properly simulate events with more than two hard partons. However, it performs all steps of generation in one package. This makes it an good choice for producing the large number of simulated multijet events needed at a hadron collider experiment.

- Top quark pair production: `MC@NLO` [44, 45] (with a top quark mass of 172.5 GeV). `MC@NLO` is a NLO generator which only calculates the matrix elements for the hard process. It is interfaced with `HERWIG` [46, 47] for performing parton showering, hadronisation and decays. `JIMMY` [48] is used for simulating the underlying event.
- W and Z production with accompanying jets: `ALPGEN` [49]. This generator works at LO but calculates the matrix elements for both the hard process and radiated hard partons. This makes is ideal for producing electroweak bosons with accompanying jets. It is also interfaced with `HERWIG` and `JIMMY`.

All MC samples are produced using the ATLAS MC10b parameter tune [50, 51] and a `GEANT4` [40] based detector simulation. All simulation and reconstruction software is integrated into the `ATHENA` [52] framework used by ATLAS.

For the plots produced throughout this thesis, MC distributions are scaled to the data luminosity unless specified otherwise; this is often the case for the multijet MC where the normalisation requires a correction factor as the cross-section from `PYTHIA` deviates from that observed in data by $\sim 30\%$.

4.3 Object Reconstruction

The purpose of object reconstruction in ATLAS is to build analysis objects and event quantities which match as closely as possible what occurred in the detector following a collision. Firstly, standalone objects are reconstructed by the various detector subsystems; for example the ID and muon systems identify tracks and the calorimetry identifies areas of energy deposition. Combined reconstruction then pools this information together, removing overlap and constructs the final physics objects such as jets, muons and $E_{\mathrm{T}}^{\mathrm{miss}}$.

4.3.1 Electrons

Events containing electrons isolated from any jets are vetoed throughout the analysis detailed in this thesis. Electrons are identified using an algorithm which is seeded by energy deposited in clusters in the EM calorimeter. The clusters are then matched to tracks in the inner detector. The main aim of electron reconstruction is to maintain a high reconstruction efficiency whilst providing good rejection against jets faking electrons. To achieve this, selection is applied on a range of discriminating variables:

- Calorimeter shower shapes.
- Leakage into the hadronic calorimeter.
- Inner detector track quality.
- Track-cluster matching.

Electrons which pass these cuts, are not located in the LAr hole (see §3.2.3) and have $p_{\mathrm{T}} > 20$ GeV are referred to as candidate electrons and used in the overlap removal described in §4.3.4. Note that a smearing procedure is applied to the momenta of electrons in MC to match what is observed in the data.

4.3.2 Muons

Events containing muons isolated from jets are vetoed in the analysis described in this thesis. Muons are reconstructed with an algorithm (STACO) which uses the track found in the Muon Spectrometer (MS) as well as the track in the Inner Detector (ID) [39]. Two types of muons are used: “combined” muons which are made from tracks that have been independently reconstructed in both systems using the STACO algorithm and “segment-tagged” muons which use the MS to tag ID tracks as muons, without requiring a fully reconstructed MS track. Requirements are imposed on the number of hits in different layers of the ID and track quality is assessed based on information from the TRT. Finally, muons are required to have $p_{\mathrm{T}} > 10$ GeV; they are then referred to as muon candidates and are used in the overlap removal described in §4.3.4. Note that a smearing procedure is applied to the momenta of muons in MC to match what is observed in the data.

4.3.3 Jets and Missing Transverse Energy

Jet production and reconstruction are introduced in §2.4. For the analysis described in this thesis, jets are reconstructed using the anti-k_t jet algorithm [32] with distance parameter $R = 0.4$ (in $y - \phi$ space). This algorithm is both infra-red and collinear safe and also benefits from being computationally efficient and provides jets with a fixed size. Inputs to the anti-k_t algorithm are topological clusters (see §3.2.3). The topological cluster algorithm [53] works by first finding seed cells in the calorimeter above a primary noise threshold level. Neighbouring cells are then added to the cluster if they have energy greater than a secondary noise threshold. This process is continued for all cells neighbouring any primary or secondary cells until there are no more adjacent cells with energy greater than the secondary threshold. Finally, any cells neighbouring the cluster with energy greater than a third noise threshold are also included. The clustering and initial jet reconstruction is performed at the electromagnetic scale. The transverse momenta of the jets are then corrected (as a function of jet p_{T} and η) through *Jet Numerical Inversion Correction* [54] to account for the difference in response between hadrons and electrons in the detector. This is known as the Jet Energy Scale (JES) correction. The JES correction increases the energy of a typical jet by $\sim 40\%$. Finally, the position of a jet is corrected to point at the primary vertex of the interaction. Jets arising from b-quarks are identified with an algorithm which uses impact parameter and secondary vertex information [55].

In this analysis, jets must have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.8$. All jets passing this loose selection are considered when applying the overlap removal described in §4.3.4. Jet quality cuts are applied to jets after overlap removal, at which point events containing at least one jet failing these quality cuts are rejected (see §4.3.5).

Missing transverse momentum is calculated with an object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithm. Clusters are first matched to reconstructed physics objects in the following order: electrons, jets and muons. Any remaining clusters, not matched to an object, are included in a *Cellout* term. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is given by the following formula:

$$(E_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{RefFinal} = (E_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{RefEle} + (E_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{RefJet} + (E_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{RefMuo} + (E_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{CellOut}, \quad (4.1)$$

where each term is computed from the negative of the sum of calibrated cluster energies inside the corresponding objects. Contributions from electrons are included in $(E_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{RefEle}$ using electrons passing the selection described in §4.3.1 before overlap removal with the p_{T} cut lowered to $p_{\mathrm{T}} > 10$ GeV. Contributions from muons are included in $(E_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{RefMuo}$, using the muons passing the criteria described in §4.3.2 before overlap removal. Contributions from jets are included in $(E_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{RefJet}$ using jets passing the selection in §4.3.3 with no η cut applied (so that the $E_{\mathrm{T}}^{\mathrm{miss}}$ accounts for activity in the forward regions of the detector). $(E_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{CellOut}$ is computed from topological clusters at the electromagnetic scale which are not included in reconstructed objects, in particular, jets with $p_{\mathrm{T}} <20$ GeV are included in this term.

4.3.4 Resolving Overlapping Objects

When candidates passing the object selection overlap with each other, a classification is required to remove all but one of the overlapping objects. All overlap criteria are based on the simple geometric $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ variable and based on previous studies at ATLAS [56]. They are applied in the following order:

1 If an electron and a jet are found within $\Delta R < 0.2$, the object is interpreted as an electron and the overlapping 'jet' is ignored.

2 If a muon and a jet are found within $\Delta R < 0.4$, the object is treated as a jet and the muon is ignored.

3 If an electron and a jet are found within $0.2 \leq \Delta R < 0.4$, the object is interpreted as a jet and the nearby 'electron' is ignored. This is an isolation requirement.

4.3.5 Event Cleaning

Event cleaning is applied to match the standard SUSY analysis. The cleaning is designed to reject events containing jets originating from non-collision sources. Firstly, a primary vertex is required with at least five tracks associated with it to ensure the presence of a hard interaction which jets can originate from. Jets originating from non-collision sources have the following criteria where the cuts are applied with a logical OR condition:

- A large fraction of the energy of the jet (> 0.5) is deposited in the HEC with more than half of the cells flagged as having 'bad' quality. These can be dead or particularly noisy cells.
- A large fraction of energy of the jet (> 0.95) is deposited in the EM calorimeter where a large proportion of the cells (> 0.8) are flagged as having 'bad' quality.
- The jet is measured over an anomalously long time compared to the event time (> 25 ns).
- A very low proportion of its energy (< 0.05) is deposited in the EM calorimeter and the sum of the momenta of the tracks of the jet is small compared to the calorimetric p_{T} measurement ($\Sigma|p_{\mathrm{T}}^{\mathrm{trk}}|/p_{\mathrm{T}}^{\mathrm{jet}} < 0.05$).
- The fractional amount of energy measured in one layer of the calorimeter is very high (> 0.99).

Events containing a jet passing any of these criteria are vetoed. Events are also rejected if the average jet time is very low (< 5 ns) or if a noise spike is observed in the LAr calorimeter. Finally, events are rejected if one of the leading four jets in the event (after overlap removal) with $p_{\mathrm{T}} > 40$ GeV points into the LAr hole region (see §3.2.3).

4.4 Trigger Menus

The trigger system used by ATLAS, described in §3.2.5, uses a "trigger menu" to decide what properties of an event should be triggered on. The menu is made up from a set of individual trigger items. A trigger item might require an electron or jet with a certain p_{T} threshold. Trigger items also include event-wide properties such as $E_{\mathrm{T}}^{\mathrm{miss}}$ or angles between physics objects. The naming convention for trigger items is usually the type of object followed by the p_{T} threshold and then details of the algorithm. For example the j240_a4tc_EFFS trigger item requires a jet with $p_{\mathrm{T}} > 240$ GeV; it uses the anti-k_t algorithm [32] with a distance parameter of 0.4 (a4) and it uses topological clusters (see §4.3.3) with the event filter in full scan (tc_EFFS) mode. Full scan mode means that the entire event is analysed by the event filter algorithms and not just items identified by the L1 and L2 triggers. A trigger item will actually correspond to a chain of decisions with one decision made at each level of the trigger system. For example the j240_a4tc_EFFS trigger corresponds to requiring a jet at L1 with $p_{\mathrm{T}} > 75$ GeV, a jet at L2 with $p_{\mathrm{T}} > 95$ GeV and an EF jet with $p_{\mathrm{T}} > 240$ GeV. The thresholds increase with the trigger level as the rate reduction must increase as described in §3.2.5.

Some trigger items would output too high an event rate to be usable if it were not for the use of prescaling. A trigger item with a prescale of 10 will only record one in 10 of the events which would pass the normal requirements of the trigger. Prescaling allows a manageable amount of data to be collected from commonly occurring events; this is the case for single jet triggers with low p_{T} thresholds such as the j55_a4tc_EFFS trigger item.

The trigger menu and prescaling vary with the instantaneous luminosity delivered by the LHC to ATLAS. For example, in early runs in 2011, the relatively low instantaneous luminosity meant the j55_a4tc_EFFS trigger could be run unprescaled; however, in later runs the trigger was given a prescale of more than 2000. Trigger items can also be deactivated with increasing luminosity or with the availability of improved algorithms.

4.5 Single Jet Triggers

The analysis described in this thesis uses single jet triggers extensively. Background information on the ATLAS trigger system is given in §3.2.5 and the trigger menu system is described in §4.4. The single jet triggers used are shown in Table 4.1. Given that the triggers work at the EM scale, the trigger p_T threshold will not directly correspond to the offline, calibrated leading jet p_T at which the trigger is fully efficient (the trigger plateau). In order for trigger inefficiencies to not bias offline distributions, an offline p_T cut is applied at the plateau p_T value. To evaluate this value, trigger efficiencies are studied as a function of offline leading jet p_T for different trigger thresholds as shown in Figure 4.1 (prepared by the ATLAS jet trigger group [57]). Table 4.1 shows the p_T range which each trigger is used for as informed by Figure 4.1.

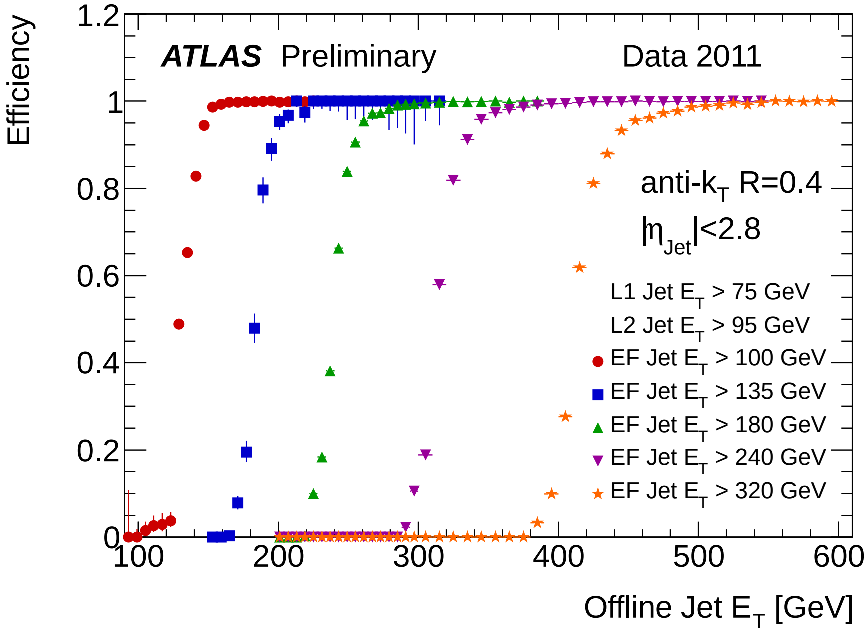


Figure 4.1: Single jet trigger efficiencies as a function of offline leading jet p_T. From [57].

The amount of data collected by prescaled triggers is less than actually occurred in the detector. This means that, to reconstruct the ‘true’ leading jet p_{T} distribution, events from prescaled triggers must be weighted appropriately. The weighting factors are found by checking the average prescale of each trigger over the dataset considered in this thesis. Table 4.1 shows these average prescales. Figure 4.2 demonstrates that this procedure gives a smooth leading jet p_{T} distribution in data.

p_{T} range [GeV]	L1 seed	L2 chain	Data period	Trigger chain	Average PS
100-130	L1_J30	L2_j45	A→B	j55_a4_EFFS	380.0
			D→H	j55_a4tc_EFFS	1480.1
130-160	L1_J55	L2_j70	A→B	j75_a4_EFFS	102.6
			D→H	j75_a4tc_EFFS	409.4
160-200	L1_J75	L2_j95	A→B	j100_a4_EFFS	28.2
			D→H	j100_a4tc_EFFS	120.2
200-260	L1_J75	L2_j95	A→B	j135_a4_EFFS	1.0
			D→H	j135_a4tc_EFFS	8.9
260-7000	L1_J75	L2_j95	A→B	j180_a4_EFFS	1.0
			D→H	j180_a4tc_EFFS	1.0

Table 4.1: Single jet triggers used throughout this thesis. Events in data are weighted by the average prescale (PS) of the appropriate trigger. The periods referred to correspond to periods of ATLAS running with similar conditions. In periods D→H, far more data were collected than in A→B.

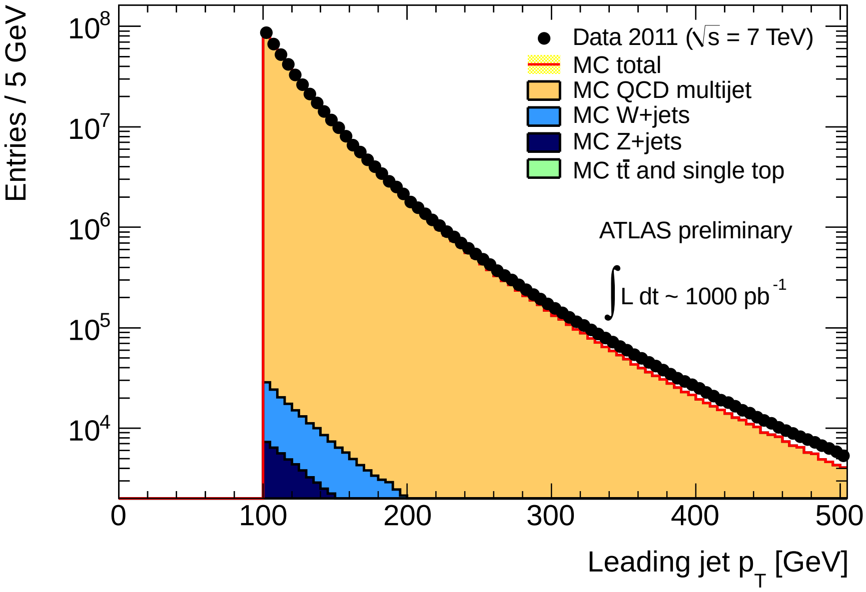


Figure 4.2: Leading jet p_{T} distribution in data and Monte Carlo simulated data after the trigger selection shown in Table 4.1. The multijet MC is normalised to the data. Statistical uncertainties are shown.

4.6 Full Chain Testing

In order to swiftly identify problems in the full production chain used by ATLAS, every day "full chain testing" is performed. This involves running a number of test jobs with several different generated samples with a small number of events compared with a full production. Any problems which are identified can then be rectified without interrupting full production. The author was involved in maintaining the full chain testing (FCT) framework as part of his service work in the ATLAS collaboration.

The FCT runs event generation, simulation, digitisation and reconstruction using batch queues at CERN. The tests are steered using an xml configuration file which specifies the details of the chains which are run. Each chain typically runs ~ 10 events to attempt to identify problems in the software. The generated samples include Higgs, Z boson, and minimum bias production. A slightly larger sample (1000 events) of top pair production is also run. This range of samples is used so that all types of different physics objects are created, hence testing as much of the software as possible. The author helped to maintain the FCT (checking for failures in the tests) and worked to improve its automation including writing scripts using Python to simplify the xml configuration file. The author also ensured that the output of the tests gave consistent results.

Chapter 5

SUSY Search with Jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and No Leptons at ATLAS

In this chapter, the ATLAS search for squark and gluino production in the jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and no lepton channel with 1 fb^{-1}, to which the author contributed, is described. Full details can be found in [5]. The analysis is introduced in §5.1, it is noted that this search has the highest potential reach for discovering SUSY. §5.2 describes the analysis method including how SM backgrounds are estimated and how systematic uncertainties are calculated. The results of the analysis are shown in §5.3. No evidence for SUSY is found and limits are set on simplified SUSY and mSUGRA/cMSSM models. These limits significantly extend on past ATLAS searches and searches performed at previous collider experiments. The multijet background estimation, which is a vital part of this analysis, is described in Chapters 6 to 8.

5.1 Introduction

In §2.3, the current status of SUSY searches using particle collider experiments is described. The discovery reach is highest at the Tevatron and LHC due to their high centre-of-mass energies. At these colliders, squarks and gluinos, which couple to the strong interaction, have the highest production cross-sections. They can decay to quarks and weakly interacting neutralinos through $\tilde{q} \rightarrow q\tilde{\chi}_1^0$ and $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_1^0$ leaving final states of $E_{\mathrm{T}}^{\mathrm{miss}}$ (from the undetected $\tilde{\chi}_1^0$) and jets. These decays have higher branching ratios than squark and gluino decays involving leptons described in §2.3.2. This means that the no lepton channel potentially has the largest reach for discovering SUSY. However, the lack of leptons in the final state makes this a challenging analysis, not least in how the multijet background is controlled and estimated.

The search presented in this chapter, also documented in [5], uses 1 fb^{-1} of data collected by the ATLAS experiment in 2011. This is the same dataset used for the multijet background estimation described in this thesis in Chapters 6 to 8. The analysis is optimised for the greatest discovery reach in simplified models (see §2.2) where particles except for the squarks, gluinos and $\tilde{\chi}_1^0$ have masses beyond the reach of the LHC.

5.2 Method

5.2.1 Trigger and Event Selection

The trigger used requires $E_{\mathrm{T}}^{\mathrm{miss}} > 45$ GeV and a jet with $p_{\mathrm{T}} > 75$ GeV, both at the electromagnetic scale. The trigger has $> 98\%$ efficiency for the offline analysis using cuts of $E_{\mathrm{T}}^{\mathrm{miss}} > 130$ GeV and leading jet $p_{\mathrm{T}} > 130$ GeV.

The object reconstruction, overlap removal and event cleaning is identical to that described in this thesis in §4.3. The event selection is split into five different searches labelled A to E, shown in Table 5.1. It is based on cuts on $E_{\mathrm{T}}^{\mathrm{miss}}$ (for inferring the presence of the $\tilde{\chi}_1^0$), the number of jets above particular p_{T} thresholds and on variables which reduce SM background contamination. The different jet requirements in A to E optimise the reach of the search for different sparticle production and decay modes. For example, pair produced squarks may decay to two quarks and two neutralinos leaving a final state of two jets and $E_{\mathrm{T}}^{\mathrm{miss}}$, while pair produced gluinos are likely to generate four quarks with two neutralinos giving a final state with four jets. Therefore searching for different jet multiplicities ensures that different production and decay modes are covered. All of the analyses use m_{eff} as a discriminating variable; this is defined as the sum of the transverse momenta of the leading n jets in the analysis and the $E_{\text{T}}^{\text{miss}}$. In signal region E (the high mass channel), all jets with $p_{\text{T}} > 40$ GeV are used in the m_{eff} calculation. The use of the m_{eff} variable is motivated by the fact that squarks and gluinos, if they exist, have higher masses than SM particles. Therefore the energy available in the transverse plane is higher than for SM events and consequently the SUSY events have higher m_{eff}. The $\Delta\phi(j_i, E_{\text{T}}^{\text{miss}})_{\text{min}}$ cut is designed to reduce the multijet background to a small level. A similar cut is used in SUSY searches at the Tevatron and at CMS (see §2.5), as mismeasured jets or jets containing neutrinos will most likely align with the $E_{\text{T}}^{\text{miss}}$. Finally, the $E_{\text{T}}^{\text{miss}}/m_{\text{eff}}$ cut further enriches the proportion of events from squark and gluino production compared to SM particle production.

Requirement	Signal Region				
	A	B	C	D	E
$E_{\text{T}}^{\text{miss}}$ [GeV]	> 130				
Leading selected jet p_{T} [GeV]	> 130				
Second selected jet p_{T} [GeV]	>40	>40	>40	>40	>80
Third selected jet p_{T} [GeV]	–	>40	>40	>40	>80
Fourth selected jet p_{T} [GeV]	–	–	>40	>40	>80
$\Delta\phi(j_i, E_{\text{T}}^{\text{miss}})$ $(i = 1, 2(, 3))$	>0.4				
$E_{\text{T}}^{\text{miss}}/m_{\text{eff}}$	> 0.3 (2j)	> 0.25 (3j)	> 0.25 (4j)	> 0.25 (4j)	> 0.2 (4j)
m_{eff} [GeV]	> 1000 (2j)	> 1000 (3j)	> 500 (4j)	> 1000 (4j)	> 1100 (incl.)

Table 5.1: Signal Regions (SRs) used in the analysis. Note that m_{eff} constructed from the leading four jets is used to calculate $E_{\text{T}}^{\text{miss}}/m_{\text{eff}}$ for SR-E, while m_{eff} constructed from all jets with $p_{\text{T}} > 40$ GeV is used for the final m_{eff} selection.

5.2.2 Standard Model Background Determination

The SM backgrounds to this SUSY search are the following:

- Multijet production. Significant $E_{\mathrm{T}}^{\mathrm{miss}}$ can be produced from heavy flavour quark decays or through severe jet mismeasurement. In order to pass the $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}}$ cut, the $E_{\mathrm{T}}^{\mathrm{miss}}$ produced must not align with the selected jets. This can occur if a jet fluctuates to such a degree that it is not reconstructed with $p_{\mathrm{T}} > 40$ GeV or if multiple jets in the event fluctuate.
- $W + \mathrm{jets}$. These events can enter the signal regions due to leptonic W decays ($W \rightarrow l\nu$) producing unmeasurable neutrinos. The charged lepton may not be reconstructed in $W \rightarrow \tau\nu$ when the τ decays hadronically; in $W \rightarrow e\nu$ when the electron is mis-identified as a jet or in $W \rightarrow \mu\nu$ when an isolated muon is not found.
- $Z + \mathrm{jets}$. This is primarily composed of $Z \rightarrow \nu\nu + \mathrm{jets}$ and is irreducible.
- Single or pair produced top quarks. Top quarks decay almost exclusively through $t \rightarrow Wb$. As for the $W + \mathrm{jets}$ background, if the W decays leptonically to a hadronic τ lepton or a misidentified/lost e or μ then the event will contain $E_{\mathrm{T}}^{\mathrm{miss}}$, jets and no leptons and can pass the signal selection.

For all of these backgrounds, Control Regions (CRs) are defined which have similar selections to the Signal Regions (SRs), except that the cuts are modified to enrich the CRs in the background of interest. Five CRs are used for each SR. Transfer Factors (TFs) are then derived for each of the 25 CRs. These are the estimated ratios of events expected between each CR and SR. The expected number of events from a particular background in a SR is then:

$$\mathrm{N(SR, estimate)} = \mathrm{N(CR, observed)} \times \mathrm{TF(estimate)} \tag{5.1}$$

The advantage of this procedure is that some systematic uncertainties cancel out in the TF estimation. The TFs are estimated using MC, except for the multijets where the data-driven technique described in this thesis is used (see Chapters 6 to 8). The CRs are defined in the following ways:

- $Z + \mathrm{jets}$: CR1. Two independent selections are used: $\gamma + \mathrm{jets}$ (CR1a) and $(Z \rightarrow ee/\mu\mu) + \mathrm{jets}$ (CR1b). The momenta of the photon or leptons are added to the $E_{\mathrm{T}}^{\mathrm{miss}}$ to simulate $(Z \to \nu\nu) + \text{jets}$ events where the neutrino is not detected.

- Multijet production: CR2. The $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ cut is changed to $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}}) < 0.2$. This selects multijet events as the $E_{\mathrm{T}}^{\mathrm{miss}}$ in these events is most likely to point in the direction of one of the jets in the event.

- $W + \text{jets}$: CR3. A lepton is required and the transverse mass of the $l + E_{\mathrm{T}}^{\mathrm{miss}}$ system is required to be consistent with the W transverse mass (30 GeV $\to$ 100 GeV). In events which populate the signal region, the lepton is misidentified as a jet and therefore to emulate these events in this CR, the lepton is treated as a jet for computing variables like m_{eff}. Top quark events are removed by requiring a veto on b-jets (identified from the impact parameter and secondary vertex, see §4.3.3).

- Top quark production: CR4. The selection is the same as CR3 except b-jets are required instead of vetoed.

5.2.3 Systematic Uncertainties

The sources of systematic uncertainty in the analysis arise from the derivation of the TFs described above and in the modelling of expected SUSY signals. These uncertainties are discussed in this section.

The Jet Energy Scale (JES) correction to account for the difference in detector response to hadrons and electrons discussed in §4.3.3 is not accurately modelled in the MC. To account for this, the uncertainty on the JES correction in MC is measured in 2010 data [54]. This uncertainty depends on jet p_{T}, η and the presence of nearby jets and is on the level of 4% on average. MC samples are then produced with recalibrated jet momenta using the upper and lower allowed values of the JES correction and TF estimates are produced using these samples. The uncertainty on the TFs is then the difference between the TF estimates in the re-calibrated samples and the nominal sample. The same procedure is applied for jet energy resolution which is also not accurately modelled in MC (with an uncertainty of up to 7%). The effects of pileup (multiple *pp* interactions occurring in one bunch crossing) are accounted for in the JES and jet resolution uncertainties. The JES and jet resolution uncertainties are also applied to the expected SUSY signal yields.

The modelling of jet radiation is not perfectly described in the MC. For example, as is mentioned in §4.2, `ALPGEN` only calculates leading order processes and `MC@NLO` only calculates the matrix elements for the hard interaction. To account for this, the backgrounds are reproduced using alternative generators or with reduced jet multiplicity and the difference in these TFs from the nominal ones is taken as the uncertainty. Differences are $\lesssim 40\%$. Parton density function uncertainties (see §4.2) are also accounted for on the samples used for TF estimation and on SUSY signal samples.

Differences between the modelling of photons and leptons (and the photon and lepton triggers) in MC and data form an additional uncertainty. Differences in the *b*-tagging efficiency are also accounted for. Finally, CR contamination with other SM processes and limited MC statistics are considered.

The uncertainties associated with the estimation of the multijet TFs are described in detail in this thesis in Chapter 8.

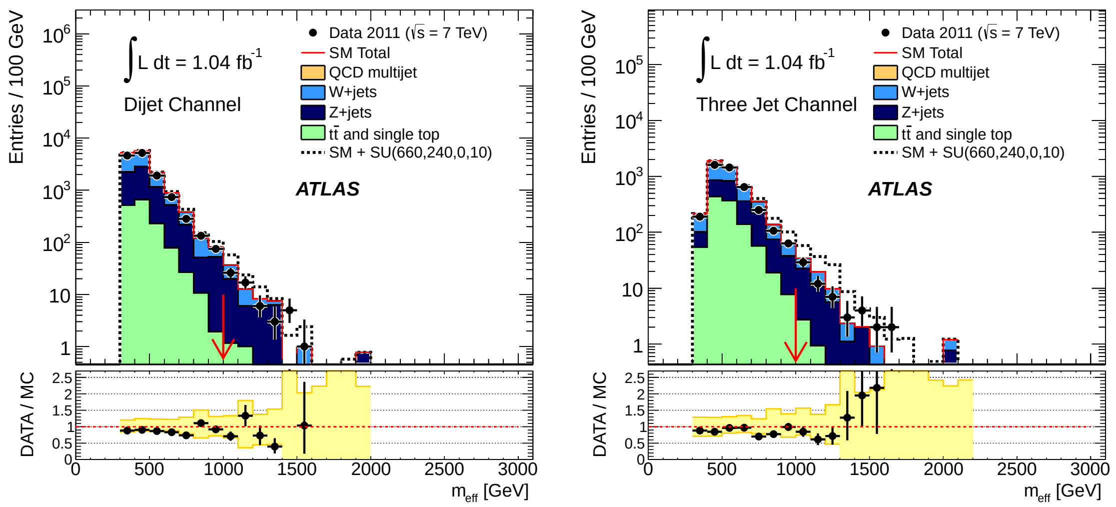
5.3 Results

Figure 5.1 shows the expected and observed m_{eff} distributions in the different SRs shown in Table 5.1. The expectation from a simulated mSUGRA model is also shown on the plots. Good agreement is observed between the SM expectations and the data. Table 5.2 shows the expected numbers of events from SM processes compared to the observations in the data. Again, good agreement is observed. As no evidence for an excess is seen, limits are set on possible SUSY parameters using the CLs technique [58]. This is a frequentist-motivated technique used for determining exclusion intervals. Exclusion contours are calculated for simplified SUSY models with

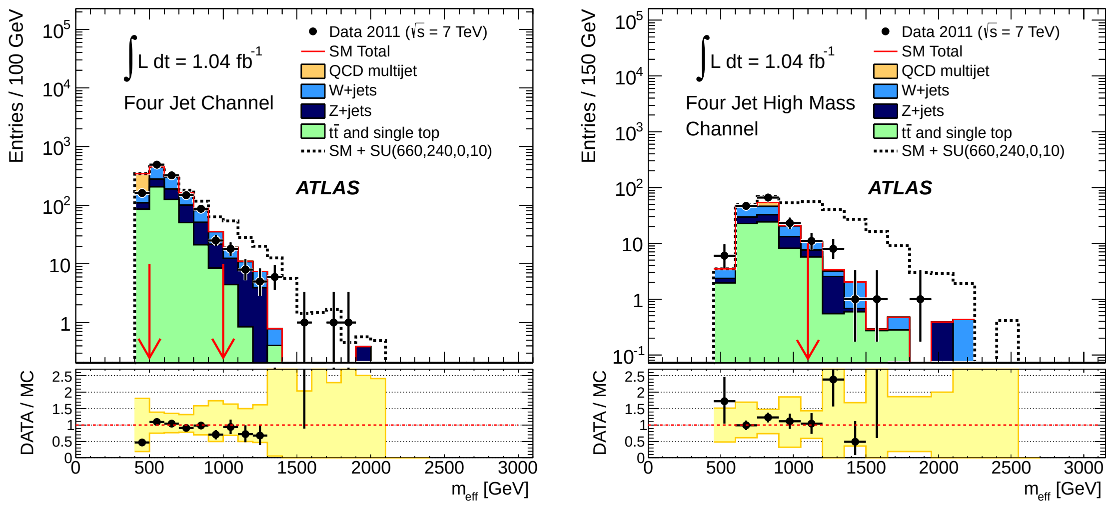
$m(\tilde{\chi}_1^0) = 0$ (Figure 5.2) and for particular mSUGRA/cMSSM models (Figure 5.3). The plots show that the limits observed are more stringent than from previous ATLAS searches and searches at LEP and the Tevatron. An example of the mass limits set is that in a simplified model with only squarks of the first two generations, a gluino octet, a massless $\tilde{\chi}_1^0$ and requiring squark and gluino masses to be less than 2 TeV, gluino and squark masses below 700 GeV and 875 GeV respectively are excluded at the 95% confidence level.

Process	Signal Region				
	$\geq$ 2-jet	$\geq$ 3-jet	$\geq$ 4-jet, $m_{\text{eff}} > 500$ GeV	$\geq$ 4-jet, $m_{\text{eff}} > 1000$ GeV	High mass
z/γ+jets	$32.3 \pm 2.6 \pm 6.9$	$25.5 \pm 2.6 \pm 4.9$	$209 \pm 9 \pm 38$	$16.2 \pm 2.2 \pm 3.7$	$3.3 \pm 1.0 \pm 1.3$
W+jets	$26.4 \pm 4.0 \pm 6.7$	$22.6 \pm 3.5 \pm 5.6$	$349 \pm 30 \pm 122$	$13.0 \pm 2.2 \pm 4.7$	$2.1 \pm 0.8 \pm 1.1$
$t\bar{t}$+ single top	$3.4 \pm 1.6 \pm 1.6$	$5.9 \pm 2.0 \pm 2.2$	$425 \pm 39 \pm 84$	$4.0 \pm 1.3 \pm 2.0$	$5.7 \pm 1.8 \pm 1.9$
QCD multi-jet	$0.22 \pm 0.06 \pm 0.24$	$0.92 \pm 0.12 \pm 0.46$	$34 \pm 2 \pm 29$	$0.73 \pm 0.14 \pm 0.50$	$2.10 \pm 0.37 \pm 0.82$
Total	$62.4 \pm 4.4 \pm 9.3$	$54.9 \pm 3.9 \pm 7.1$	$1015 \pm 41 \pm 144$	$33.9 \pm 2.9 \pm 6.2$	$13.1 \pm 1.9 \pm 2.5$
Data	58	59	1118	40	18

Table 5.2: SM background expectations and data observations in the different signal regions. The first (second) quoted uncertainty is statistical (systematic). Adapted from [5].



(a) 2 jet channel (b) 3 jet channel



(c) 4 jet channels (d) High mass channel

Figure 5.1: m_{eff} distributions in data and estimated with MC in the SRs shown in Table 5.1. The black dotted lines show the expectation from SM sources and a SUSY model identified in Figure 5.3. Statistical uncertainties are shown on the data points. Below the histograms, the ratios of the data to SM expectations are shown with the data statistical uncertainties displayed as black vertical lines and the MC systematic uncertainties shown as yellow bands. The red arrows show the m_{eff} SR cuts. From [5].

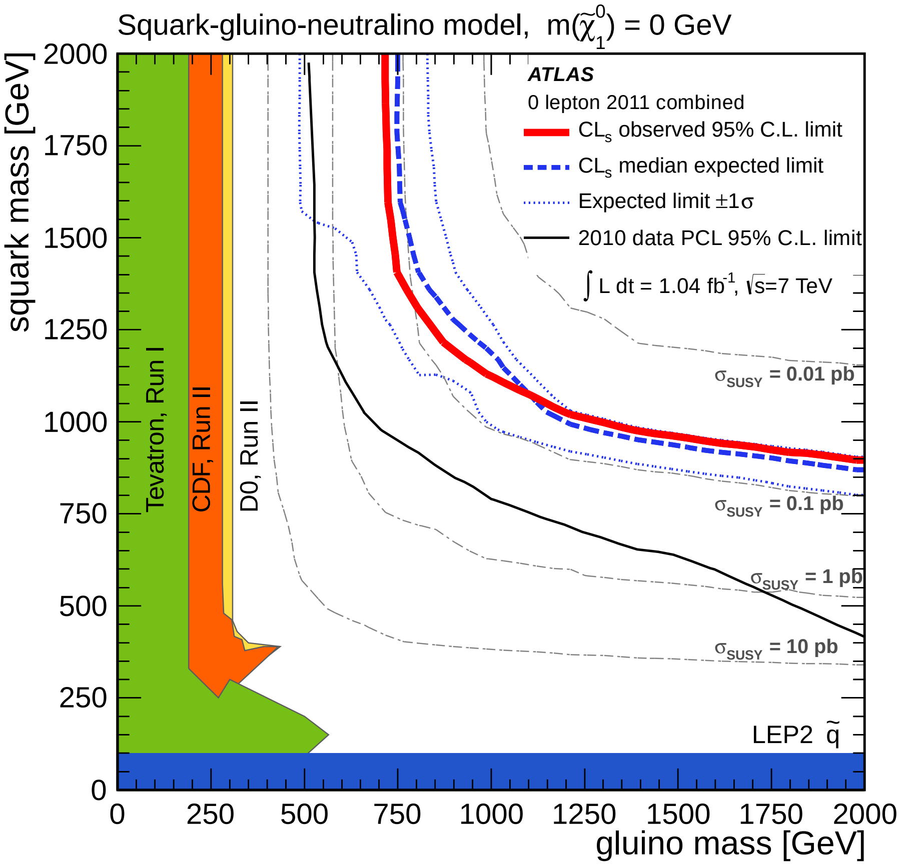


Figure 5.2: Exclusion limits for simplified SUSY models with $m(\tilde{\chi}_1^0) = 0$. From [5].

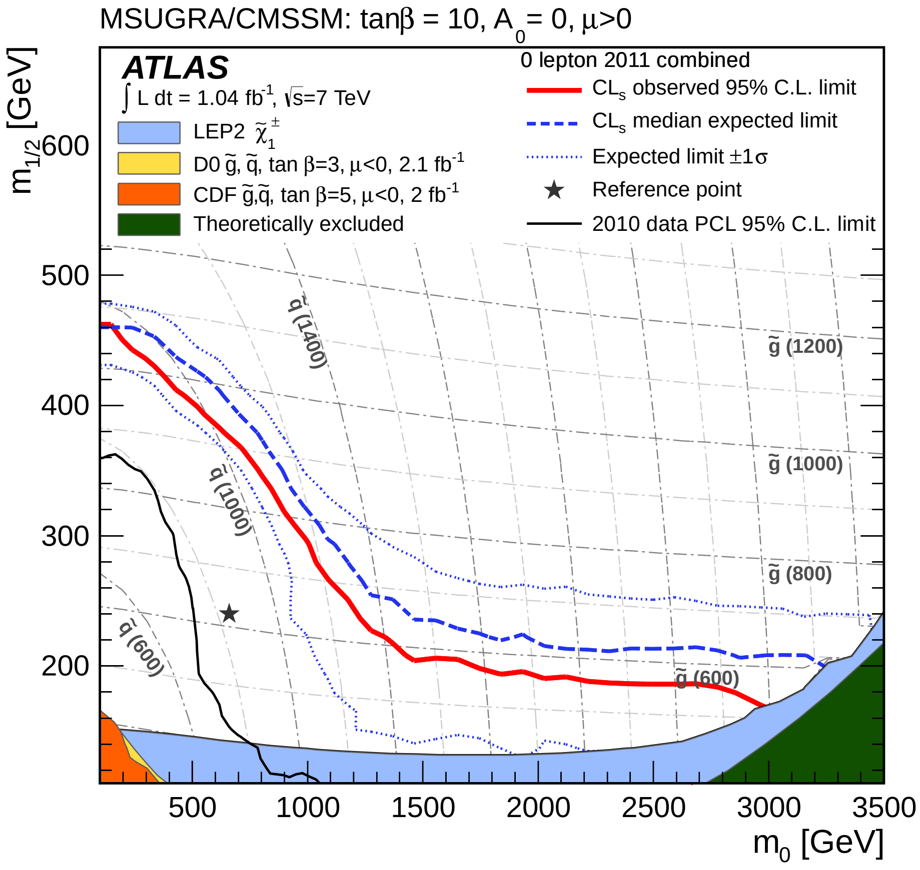


Figure 5.3: Exclusion limits for particular mSUGRA/cMSSM models. The reference point identified in this plot is used in Figure 5.1. From [5].

Chapter 6

Jet Smearing Overview

This chapter gives an overview of the jet smearing technique for estimating the multijet background to SUSY searches with no leptons. §6.1 introduces the concept of jet response. §6.2 gives the motivation for the jet smearing method and defines the process of jet smearing. §6.3 outlines the assumptions of the technique and justifications. Finally, §6.4 explains the selection of well-measured data “seed” events which are used throughout the method.

6.1 Jet Response

No particle detector can perfectly reconstruct a physics object and so every type of object will have an associated resolution or response distribution. p_{T} response is defined as $R = p_{\mathrm{T}}^{\mathrm{reco}}/p_{\mathrm{T}}^{\mathrm{true}}$ where $p_{\mathrm{T}}^{\mathrm{reco}}$ is the p_{T} of the reconstructed jet and $p_{\mathrm{T}}^{\mathrm{true}}$ is measured at the particle level. In the case of jets, the response will be relatively broad because of a number of effects:

- Fluctuations in the hadronic shower can be large.
- There is a significant amount of dead material in the detector in and leading

up to the hadronic calorimeter, which affects how precisely the energy of a jet can be measured.

- Some of the energy of the jet can fall outside of the cone of the jet.
- Jets may not be fully contained in the calorimeter, ‘punching-through’ to the muon system.
- Heavy flavour jets can contain neutrinos which escape undetected.

These effects lead to the response of a jet being broader than that of an purely electromagnetic object such as an electron. Furthermore, these effects can produce extreme deviations in the reconstructed momentum of a jet compared with the particle-level momentum. As a result of this, jet response does not follow a purely Gaussian shape but also contains non-Gaussian tails.

The response of a jet can depend on its kinematics. For example, punch-through (mentioned above) is clearly p_{T} dependent. Furthermore, as described in §3.2.3, the ATLAS calorimeter uses different technologies in different regions of η and therefore the response will depend on the direction of a jet. Jet response can also depend on the flavour of the jet as heavy flavour jets can contain neutrinos through electroweak decay. This will be revisited in §7.1. The measurement of jet response, including the non-Gaussian tails, will be discussed in detail in Chapter 7.

6.2 Motivation and Overview

A data-driven estimation of the multijet background to SUSY searches with no leptons (described in Chapter 5) is vital because the large cross-section for multijet production at the LHC means that it is impossible to simulate a comparable number of MC events. For example, at an instantaneous luminosity of 10^{33} cm^{-2}s^{-1}, the LHC produces ~ 400 events per second containing a jet with $E_{\mathrm{T}}^{\mathrm{jet}} > 100$ GeV (see Figure 2.5). A data-driven estimation is also desirable as the multijet estimation from `PYTHIA` is not expected to reproduce the data at high jet multiplicities. This is because, as mentioned in §4.2, `PYTHIA` is a leading order generator and matrix elements are only calculated for the hard interaction, and not for additional hard radiated partons.

The jet smearing method is a data-driven technique which is used to determine multijet distributions in SUSY signal and control regions. It works through scaling (smearing) the momenta of jets in well measured, low-$E_{\mathrm{T}}^{\mathrm{miss}}$ "seed events" with random numbers drawn from jet response functions measured using data. This process is repeated a large number of times for each seed event, hence producing a large number of smeared events or "pseudo-data". The pseudo-data sample then contains events with multiple jets and potentially large $E_{\mathrm{T}}^{\mathrm{miss}}$ and can be used like a MC sample providing estimated distributions of multijet events. The normalisation of pseudo-data distributions to the data can be performed in multijet dominated control regions.

Selecting the seed events from data means the kinematics of the pseudo-data match multijet data events as long as no bias is introduced though the seed selection. The smearing of the jets in the seed events acts to re-introduce $E_{\mathrm{T}}^{\mathrm{miss}}$ into the events. The precise measurement of jet response, including the non-Gaussian tails, as a function of jet angle and momentum, ensures that the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in the pseudo-data reproduces the data. The measurement of jet response is described in detail in Chapter 7. Figure 6.1 shows a cartoon of the jet smearing method, with two high-$E_{\mathrm{T}}^{\mathrm{miss}}$ events generated from a well-measured seed event.

A smeared event is generated by multiplying each jet four-vector in a seed event by a random number drawn from the appropriate jet response function. The smeared event $E_{\mathrm{T}}^{\mathrm{miss}}$ is then given by

$$\vec{\not{E}}_{\mathrm{T}}{}' = \vec{\not{E}}_{\mathrm{T}}^{\mathrm{seed}} - \sum_i \vec{p}_{\mathrm{T}}{}'(j_i) + \sum_i \vec{p}_{\mathrm{T}}(j_i), \tag{6.1}$$

primes are used to distinguish smeared from unsmeared quantities.

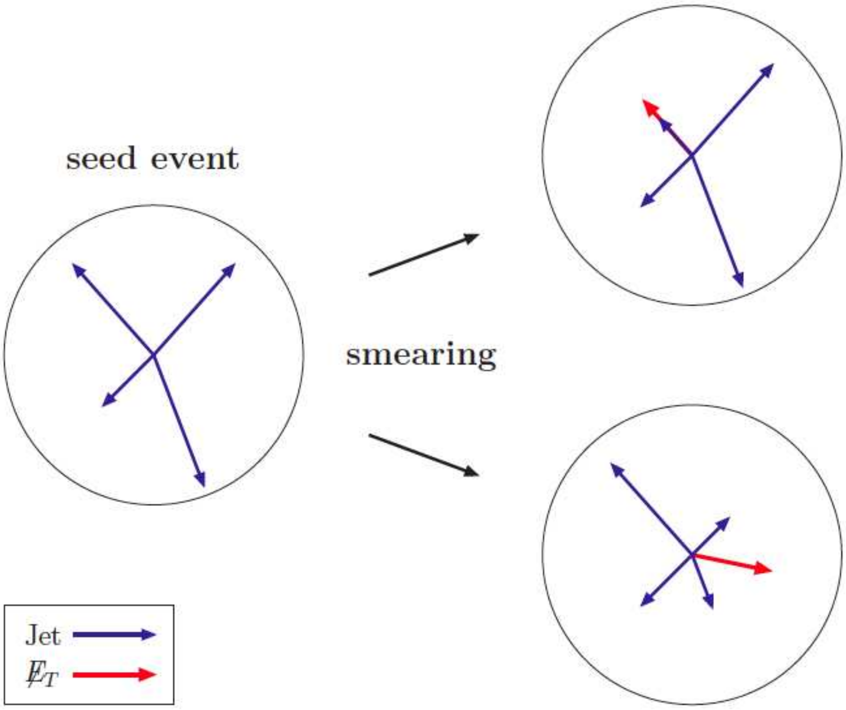


Figure 6.1: Jet smearing cartoon. From [59]

6.3 Assumptions of the Method

The jet smearing technique relies on some basic assumptions:

1 The $E_{\mathrm{T}}^{\mathrm{miss}}$ of multijet events is dominated by jet fluctuation (true or fake).

2 True and fake sources of jet fluctuation can be included in one response function and this can be applied to all jets.

3 Any dependence of jet response on event-wide properties, such as jet multiplicity or the fluctuations of other jets in the event, can be neglected. In other words, jet smearing can justifiably be applied on a jet-by-jet basis.

The main justification for these assumptions is that the technique is shown to successfully reproduce the control distributions discussed in Chapter 7 and the multijet control and validation distributions for the main SUSY analysis shown in §8.4. Assumption (2) can be checked by studying the proportion of b-jets to light-jets in the seed selection to check for potential bias. This is done in §6.4, see Figure 6.4(d).

6.4 Seed Event Selection

The seed selection is designed to collect events where the momenta of the reconstructed jets are as close as possible to the 'true', particle level jets. In other words jets in these events should have response values within the Gaussian core of jet response. Jet fluctuation can then be introduced through smearing with minimal double counting.

Events must first pass the trigger selection shown in Table 4.1. To ensure that only events in which jets have undergone small response fluctuations are selected, a cut on $E_{\mathrm{T}}^{\mathrm{miss}}$ could potentially be used. However, since $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution varies with the event scalar sum E_{T}, a cut on $E_{\mathrm{T}}^{\mathrm{miss}}$ would bias the p_{T} scale of the seed events. This is demonstrated in Figure 6.2. Instead the $E_{\mathrm{T}}^{\mathrm{miss}}$-significance, S, is used as it approximately eliminates the $\sum E_{\mathrm{T}}$ dependence.

$$S = \frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sqrt{\sum E_{\mathrm{T}}}} \tag{6.2}$$

The S distribution for events passing the trigger selection is shown in Figure 6.3. A cut on S at the peak value ensures that the selected events will typically contain jets undergoing fluctuations in the peak of the Gaussian part of the response function whilst retaining a relatively high selection efficiency. The value chosen is $S < 0.7\ \mathrm{GeV}^{1/2}$. Using a higher value would provide more seed event statistics but increase the chance of having a large fluctuation present in a jet in a seed event. Given that limited seed statistics is not the dominant uncertainty in the method,

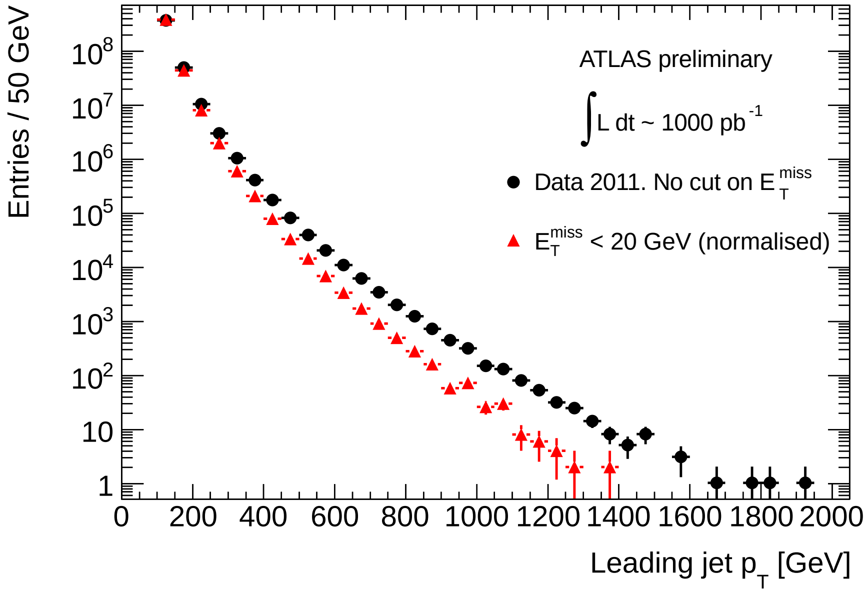


Figure 6.2: The leading jet p_{T} distribution for events passing trigger selection (see text) and after $E_{\mathrm{T}}^{\mathrm{miss}} < 20$ GeV cut. Statistical uncertainties are shown. The $E_{\mathrm{T}}^{\mathrm{miss}}$ cut clearly biases the p_{T} distribution.

higher values are not considered.

The $E_{\mathrm{T}}^{\mathrm{miss}}$, N_j and leading jet p_{T} distributions before and after the S cut are shown in Fig. 6.4. It can be seen that the seed event selection allows for the selection of low $E_{\mathrm{T}}^{\mathrm{miss}}$ events without introducing any significant bias in jet multiplicity and p_{T}. Also shown (see Figure 6.4(d)) is the proportion of b-jets before and after the S cut; no significant bias is introduced by the seed selection.

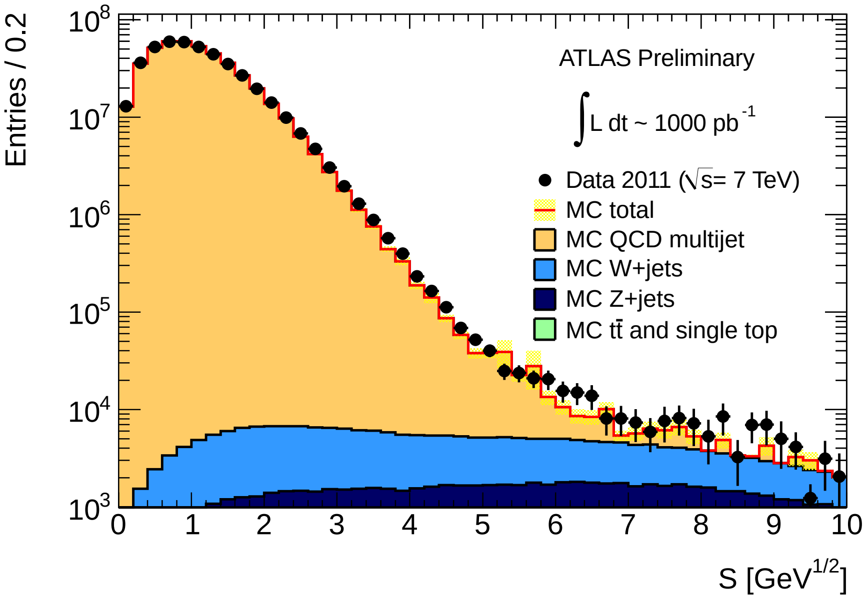
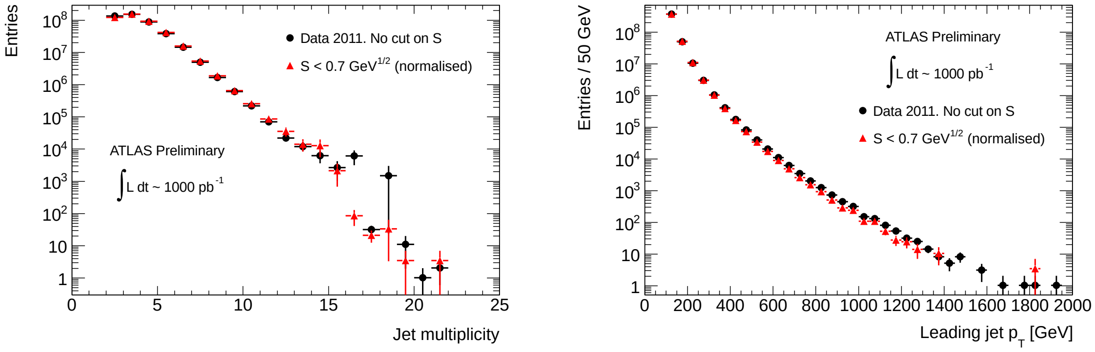
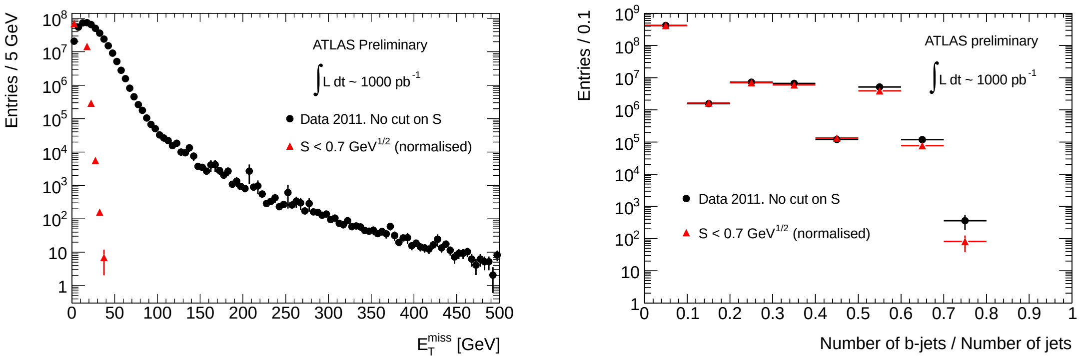


Figure 6.3: The S distribution for events passing trigger selection (see text). The cut used in the seed selection is $S < 0.7$ GeV$^{1/2}$. The multijet MC is normalised to the data. Statistical uncertainties are shown.



(a) N_j (b) Leading jet p_{T}



(c) $E_{\mathrm{T}}^{\mathrm{miss}}$ (d) b-jet proportion

Figure 6.4: The (a) N_j, (b) leading jet p_{T}, (c) $E_{\mathrm{T}}^{\mathrm{miss}}$ and (d) b-jet proportion distributions for events passing the seed selection ($S < 0.7$ GeV$^{1/2}$). The equivalent distributions for events passing the trigger selection only are shown for comparison. The histograms of the seed events are normalised to the same area as the 'data' histograms. Statistical uncertainties are shown.

Chapter 7

Measurement of Jet Response

This chapter describes the measurement of jet response used for the jet smearing multijet estimation described in Chapter 8. Firstly, jet response is measured from multijet MC samples as described in §7.1. The MC response is modified to match data using the techniques described in §7.2. The dijet analysis for constraining the Gaussian response is described in §7.3. The "Mercedes" analysis for constraining the non-Gaussian response tails is then described in §7.4. Novel event shape triggers are used to collect extra statistics for this measurement as described in §7.4.3. They allow the non-Gaussian response tail to be constrained to a greater extent than if only standard single jet triggers are used as shown in §7.4.4.

7.1 MC Response

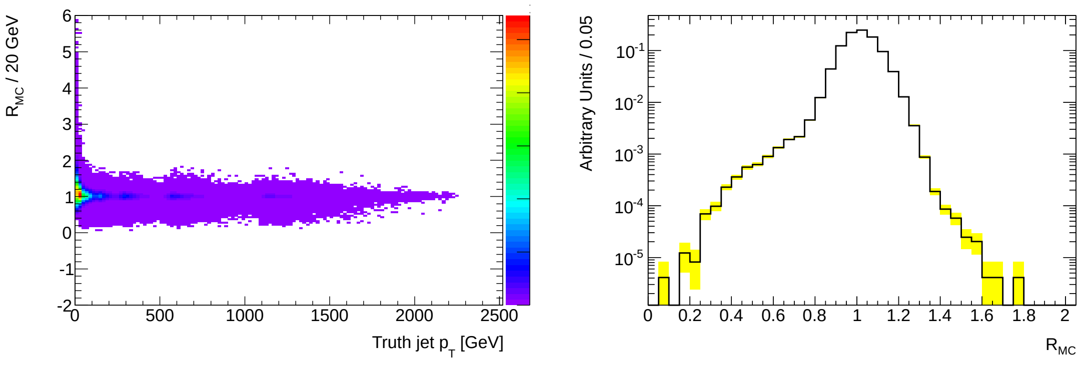
Jet response is initially taken from multijet MC samples. It is then constrained to match the data in control distributions. To measure the response from MC, the following is done:

- Multijet samples produced with `PYTHIA` (see §4.2) and standard object definitions and overlap removal are used (see §4.3). Jet cleaning is applied as in §4.3.5

so that jet response is only measured for jets which are used in the main SUSY analysis.

- Jet response is measured for any reconstructed jet in an event isolated from other reconstructed jets by $\Delta R > 1.0$ and matched to a truth jet with $\Delta R < 0.1$. The reconstructed jet is also required to be isolated from other truth jets by $\Delta R > 1.0$. Note that ΔR here refers to the geometric quantity $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$, and not jet response.
- The four-momenta of any final state neutrinos in the cone of the truth jet are added to the jet. This is vital as both true and fake sources of $E_{\mathrm{T}}^{\mathrm{miss}}$ should be encoded in the jet response function used for jet smearing.
- Finally, the jet response is measured as a function of truth jet p_{T} using $R_{MC} = p_{\mathrm{T}}^{\mathrm{reco}}/p_{T}^{\mathrm{true}}$.

Figure 7.1(a) shows R_{MC} with p_{T} bins of width 20 GeV. The response function for a jet with a particular momentum is a p_{T} slice of this distribution. The response for jets with 140 GeV$< p_{\mathrm{T}} <$ 160 GeV is shown in Figure 7.1(b).



(a) R_{MC} as a function of jet p_{T}

(b) R_{MC} for jets with 140 GeV$< p_{\mathrm{T}} <$ 160 GeV.

Figure 7.1: R_{MC} measured from multijet MC events by matching simulated truth jets to reconstructed jets. The yellow band in (b) shows the uncertainty from limited MC statistics.

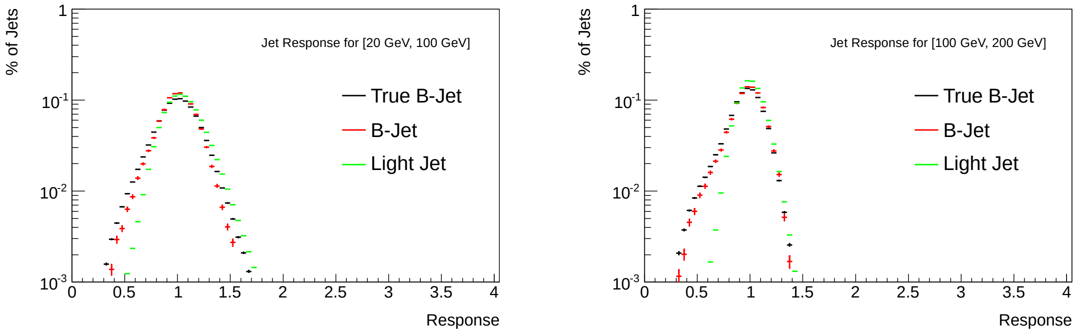
In §6.1, it is noted that the response of heavy flavour jets is expected to be different to the response of light jets. This is particularly relevant to the background estimation in SUSY searches with b-jets [6, 7, 8] where reconstructed b-jets are smeared with different response functions to those used for light jets. This technique is not used in the analysis described in this thesis since there is no requirement for b-jets in the SUSY signal or multijet control regions. In the supporting material [60] (internally approved by the ATLAS collaboration) for the analysis described in [7], the response functions for true and reconstructed b-jets and for reconstructed light jets are determined using multijet MC. These are shown in Figure 7.2. It can be seen that the non-Gaussian tails are much larger for the b-jets than for the light jets. This indicates that non-Gaussian jet response is dominated by the presence of neutrinos in heavy flavour jets.

7.2 Modification of MC Response

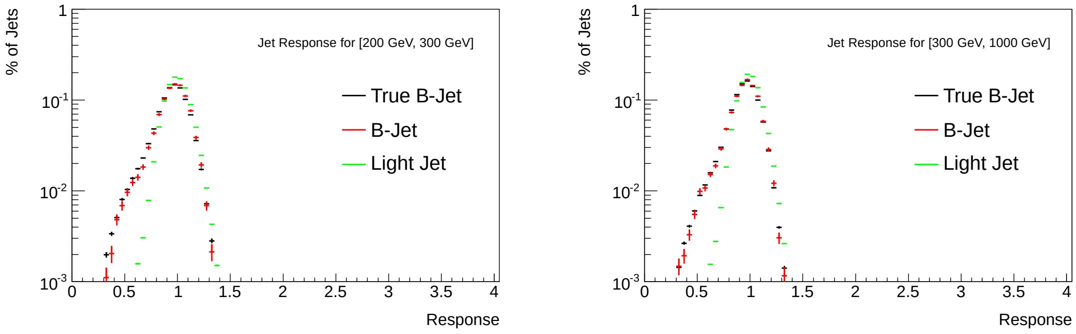
In order to constrain the response functions so that they match jet response in data, multijet dominated control distributions are used. Jet smearing on seed events is used to produce pseudo-data which can then be compared to the data. The response functions are then modified to gain agreement between the pseudo-data and the data. Two techniques are used to modify the MC response functions to improve the agreement between the pseudo-data and data in the control distributions:

(1) Additional Gaussian smearing to widen the jet response. To account for differences in the Gaussian response width in MC and data, a p_{T}-dependent Gaussian smearing is introduced. For a given jet, the smearing value from the MC response function is multiplied by a number drawn from a Gaussian with a mean of one and width $\sigma_{correction}(p_{\mathrm{T}})$. The functional form and uncertainty on $\sigma_{correction}(p_{\mathrm{T}})$ is determined using the analysis described in §7.3. This analysis uses the p_{T} asymmetry of dijet events to constrain $\sigma_{correction}(p_{\mathrm{T}})$ so that the Gaussian response in pseudo-data matches that in data.

(2) Non-Gaussian response shape modification. Understanding the shape of the



(a) 20 GeV $< p_{\mathrm{T}} <$ 100 GeV (b) 100 GeV $< p_{\mathrm{T}} <$ 200 GeV



(c) 200 GeV $< p_{\mathrm{T}} <$ 300 GeV (d) 300 GeV $< p_{\mathrm{T}} <$ 1000 GeV

Figure 7.2: Response functions for b-jets and light jets from multijet MC in different jet p_{T} ranges. From [60].

response tails is vital in estimating distributions in the SUSY signal and control regions. This is because the high $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement guarantees that at least one jet in a multijet event has undergone a very significant fluctuation. The “Mercedes”[1] analysis described in §7.4 is used to derive the correct tail shapes and uncertainties in different p_{T} bins. Events are selected with three or more jets where one of the jets is unambiguously associated with the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event. The $E_{\mathrm{T}}^{\mathrm{miss}}$ can then be assumed to be due to the fluctuation of the matched jet and the response of that jet can be measured. This analysis constrains the non-Gaussian response to match the data.

---

[1] The events passing the selection in this analysis resemble the Mercedes logo, see Fig. 7.4.

7.3 Dijet Analysis

The first control distribution is the p_{T} asymmetry of dijet events. In pure dijet events, any asymmetry is due to jet fluctuation. Events must first satisfy the trigger selection shown in Table 4.1. Seed events are selected as described in §6.4 and 1000 smeared events are produced from each seed event as described in §6.2. The following requirements are then imposed on the data (following the trigger selection from Table 4.1) and smeared data:

- Two jets with $p_{\mathrm{T}} > 130, 40$ GeV. No additional jets with $p_{\mathrm{T}} > 30$ GeV.
- $|\pi - \Delta\phi(j_1, j_2)| < 0.3$.

The p_{T} asymmetry is given by

$$A(p_{\mathrm{T},1}, p_{\mathrm{T},2}) = \frac{p_{\mathrm{T},1} - p_{\mathrm{T},2}}{p_{\mathrm{T},1} + p_{\mathrm{T},2}}, \tag{7.1}$$

where the indices correspond to the jet p_{T} ordering. The width of a Gaussian fitted to this distribution is given by:

$$\sigma_A = \frac{\sqrt{(\sigma(p_{\mathrm{T},1}))^2 + (\sigma(p_{\mathrm{T},2}))^2}}{\langle p_{\mathrm{T},1} + p_{\mathrm{T},2}\rangle}. \tag{7.2}$$

If the jets have approximately the same rapidity then $\sigma(p_{\mathrm{T},1}) = \sigma(p_{\mathrm{T},2}) = \sigma(p_{\mathrm{T}})$. Furthermore, $\langle p_{\mathrm{T},1} + p_{\mathrm{T},2}\rangle = 2\langle p_{\mathrm{T,average}}\rangle$. Therefore:

$$\sigma_A \simeq \frac{\sigma(p_{\mathrm{T}})}{\sqrt{2}p_{\mathrm{T}}}. \tag{7.3}$$

$A(p_{\mathrm{T},1}, p_{\mathrm{T},2})$ is measured in p_{T} bins (see Fig. 7.3(a)) and then the p_{T} slices are fitted with Gaussians with means fixed at zero in the range $0 < A < 0.15$. This range was found to provide a good fit for all p_{T}. See Figures 7.3(b) and 7.3(c) for examples. Fig. 7.3(d) shows the Gaussian widths measured in data ($\sigma_{A,data}$). Also shown is σ_A reproduced using jet smearing with the uncorrected response function taken from MC (referred to as $\sigma_{A,MC}$). The $\sigma_{A,data}$ and $\sigma_{A,MC}$ distributions are then fitted with the following functional form:

$$\sigma_A = \frac{A}{p_{\mathrm{T}}} + \frac{B}{\sqrt{p_{\mathrm{T}}}} + C. \tag{7.4}$$

The difference observed between $\sigma_{A,data}$ and $\sigma_{A,MC}$ is used to constrain the $\sigma_{correction}(p_{\mathrm{T}})$ parameter introduced in §7.2. The additional smearing of the MC response functions with $\sigma_{correction}(p_{\mathrm{T}})$ increases $\sigma_{MC}(p_{\mathrm{T}})$ in the same way as in the convolution of two Gaussians:

$$\left(\frac{\sigma_{total}(p_{\mathrm{T}})}{p_{\mathrm{T}}}\right)^2 = \left(\frac{\sigma_{MC}(p_{\mathrm{T}})}{p_{\mathrm{T}}}\right)^2 + \sigma_{correction}(p_{\mathrm{T}})^2. \tag{7.5}$$

$\sigma_{total}(p_{\mathrm{T}})$ is the width of the corrected response function. Requiring that this corrected response matches the data response (i.e. $\sigma_{total}(p_{\mathrm{T}}) = \sigma_{data}(p_{\mathrm{T}})$) and using Eq. 7.3 it follows that:

$$(\sqrt{2} \times \sigma_{A,data})^2 = (\sqrt{2} \times \sigma_{A,MC})^2 + \sigma_{correction}(p_{\mathrm{T}})^2. \tag{7.6}$$

Therefore:

$$\sigma_{correction}(p_{\mathrm{T}}) = \sqrt{2} \times \sqrt{\sigma^2_{A,data} - \sigma^2_{A,MC}} \tag{7.7}$$

Using this equation and the parameters in Table 7.1 from the fits to $\sigma_{A,data}$ and $\sigma_{A,MC}$ (using Eq. 7.4), $\sigma_{correction}(p_{\mathrm{T}})$ is now fully described. Table 7.1 shows the parameters calculated using different fitting ranges, this is important in deriving the systematic uncertainties on $\sigma_{correction}(p_{\mathrm{T}})$ as discussed below.

7.3.1 Uncertainty on $\sigma_{correction}(p_{\mathrm{T}})$.

The fits on $\sigma_{A,data}$ and $\sigma_{A,MC}$ are unconstrained below 150 GeV with the triggers used in this analysis so $\sigma_{correction}(p_{\mathrm{T}})$ must be extrapolated to low p_{T}. This results in the value of $\sigma_{correction}(p_{\mathrm{T}})$ at low p_{T} varying to a large extent with the fitting range used. Examples are:

Parameter	Fit range [GeV]		
	$150 < p_{\mathrm{T}} < 400$	$150 < p_{\mathrm{T}} < 600$	$150 < p_{\mathrm{T}} < 1000$
A_{data} [GeV]	6.56	7.56	9.66
B_{data} [GeV$^{1/2}$]	0.0950	-0.0299	-0.301
C_{data}	0.0418	0.0456	0.0543
A_{MC} [GeV]	6.65	6.55	4.69
B_{MC} [GeV$^{1/2}$]	0.0242	0.0364	0.274
C_{MC}	0.0421	0.0417	0.0342

Table 7.1: Parameters for fits to the $\sigma_{A,data}$ and $\sigma_{A,MC}$ distributions. These parameters are used to calculate $\sigma_{correction}(p_{\mathrm{T}})$ using Eq. 7.7 as described in the text.

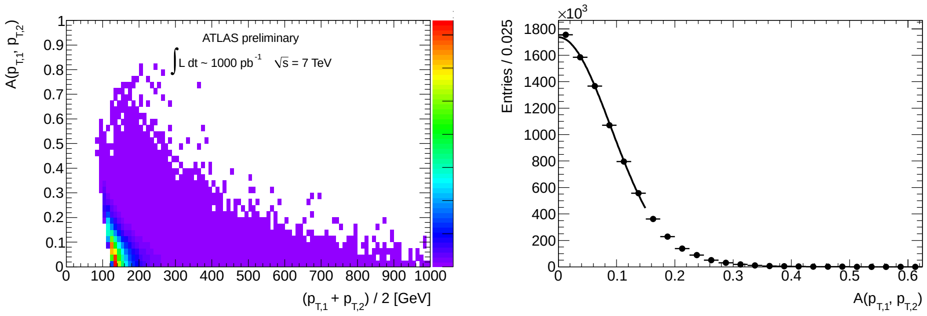
- Fitting range 150 GeV $< p_{\mathrm{T}} <$ 400 GeV $\Longrightarrow \sigma_{correction}(50\text{ GeV}) = 0.119$.
- Fitting range 150 GeV $< p_{\mathrm{T}} <$ 600 GeV $\Longrightarrow \sigma_{correction}(50\text{ GeV}) = 0.074$.
- Fitting range 150 GeV $< p_{\mathrm{T}} <$ 1000 GeV $\Longrightarrow \sigma_{correction}(50\text{ GeV}) = 0.053$.

These fits are shown in Figure 7.3(d). Using different fitting ranges therefore provides a simple method which accounts for the uncertainty on $\sigma_{correction}(p_{\mathrm{T}})$ at low p_{T}.

To account for any other differences between $\sigma_{A,total}$ and $\sigma_{A,data}$, high and low estimates are performed in the following way:

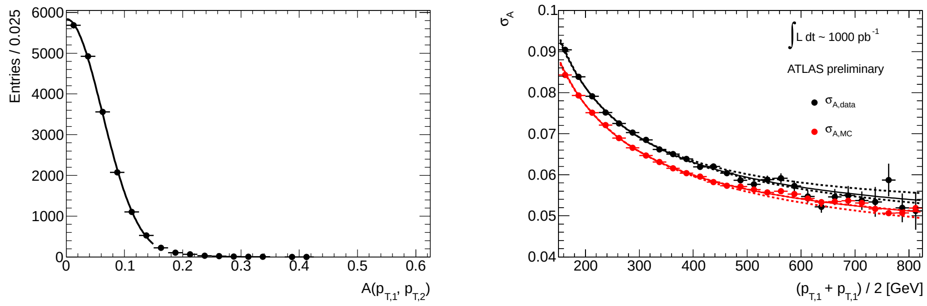
- High estimate: use parameters from fitting range 150 GeV $< p_{\mathrm{T}} <$ 400 GeV. $\sigma^{high}_{correction}(p_{\mathrm{T}}) = \sigma_{correction}(p_{\mathrm{T}}) + 0.05$.
- Optimal estimate: use parameters from fitting range 150 GeV $< p_{\mathrm{T}} <$ 600 GeV. $\sigma^{optimal}_{correction}(p_{\mathrm{T}}) = \sigma_{correction}(p_{\mathrm{T}})$.
- Low estimate: use parameters from fitting range 150 GeV $< p_{\mathrm{T}} <$ 1000 GeV. $\sigma^{low}_{correction}(p_{\mathrm{T}}) = \sigma_{correction}(p_{\mathrm{T}}) - 0.05$.

The value 0.05 is an arbitrary choice but is seen to perform well as discussed below. To obtain a given distribution or quantity, the estimate using $\sigma^{optimal}_{correction}(p_{\mathrm{T}})$ is taken as the optimal estimate. The uncertainty is then taken as half the difference between the estimates acquired using $\sigma^{high}_{correction}(p_{\mathrm{T}})$ and $\sigma^{low}_{correction}(p_{\mathrm{T}})$. This procedure is used for obtaining the jet smearing estimate of σ_A shown in Figure 7.3(e). The chi-squared value between the estimate and data is 24.1 with 34 degrees of freedom ($\mathrm{p-value} = 0.90$). This indicates that the estimate (with uncertainties accounted for) reproduces the data well.



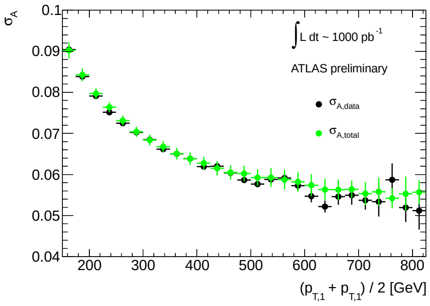
(a) p_{T} asymmetry (A) as a function of p_{T}

(b) A and Gaussian fit to A for data events with 150 GeV $< p_{\mathrm{T}}(average) <$ 175 GeV



(c) A and Gaussian fit to A for data events with 400 GeV $< p_{\mathrm{T}}(average) <$ 425 GeV

(d) $\sigma_{A,data}$ and $\sigma_{A,MC}$ (solid points) with fits in different p_{T} ranges (solid and dashed lines)



(e) $\sigma_{A,data}$ and $\sigma_{A,total}$

Figure 7.3: Dijet analysis plots. The uncertainties on the points in (d) and on the data points in (e) are uncertainties from the Gaussian fitting. The uncertainty on $\sigma_{A,total}$ in (e) is calculated as described in the text. $\sigma_{correction}(p_{\mathrm{T}})$ is determined from the fits to $\sigma_{A,data}$ and $\sigma_{A,MC}$ shown in (d) using Eq. 7.7. (e) demonstrates that the jet smearing estimate of σ_A (including uncertainties) successfully reproduces the data.

7.4 Mercedes Analysis

The measurement of the non-Gaussian jet response tails is vital in attempting to reproduce high $E_{\mathrm{T}}^{\mathrm{miss}}$ multijet events. The Mercedes analysis, developed in [59, 61], is designed to probe the response tails of a single jet. Unlike a dijet balance analysis, the Mercedes analysis can distinguish between high-side and low-side fluctuations of the candidate jet. The selection described here is based on that presented in [59].

Events are initially selected according to the trigger selection in Table 4.1. Seed events are selected as described in §6.4 and 1000 smeared events are produced from each seed event as described in §6.2. The pseudo-data accounts for the Gaussian corrections and uncertainties found in §7.3. The selection described below is then imposed on the data (following the trigger selection from Table 4.1) and smeared data.

The preselection requirements are for three or more jets with $p_{\mathrm{T}} > 260, 40, 30$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV. Three jets are required to allow the selection of three pronged Mercedes type events which allow for the unambiguous association of one jet with the $E_{\mathrm{T}}^{\mathrm{miss}}$ in these events (see Fig. 7.4 for a depiction of Mercedes events). The cut on the leading jet p_{T} is used to ensure that data events are collected using non-prescaled triggers to simplify the derivation of the tail shape. Non-multijet backgrounds are also significantly reduced with this cut.

Large $E_{\mathrm{T}}^{\mathrm{miss}}$ can be generated either by upward or downward fluctuations in jet response. Therefore it is important to select events with jets either parallel or anti-parallel to $E_{\mathrm{T}}^{\mathrm{miss}}$ in the transverse plane (see Fig. 7.5). Therefore, in each event passing the preselection cuts, all N jets with $p_{\mathrm{T}} > 30$ GeV are sorted in increasing distance in ϕ from the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the event is selected if it satisfies *either*:

$$\begin{aligned}
|\Delta\phi(j^1, E_{\mathrm{T}}^{\mathrm{miss}})| &< \pi - |\Delta\phi(j^N, E_{\mathrm{T}}^{\mathrm{miss}})| \\
|\Delta\phi(j^1, E_{\mathrm{T}}^{\mathrm{miss}})| &< \Delta\phi^{\mathrm{match}} \\
|\Delta\phi(j^2, E_{\mathrm{T}}^{\mathrm{miss}})| &> \Delta\phi_1^{\mathrm{isol}}, \qquad (7.8)
\end{aligned}$$

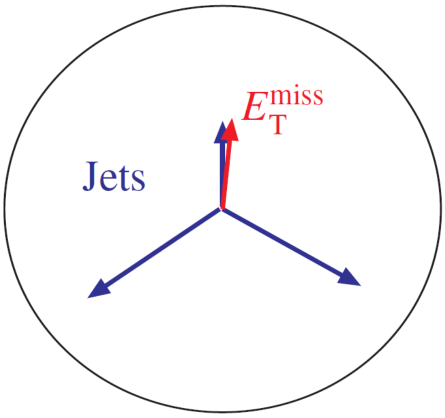


Figure 7.4: Mercedes event topology. $E_{\mathrm{T}}^{\mathrm{miss}}$ is unambiguously associated with one jet in the event. Events of this type are those used in the measurement of the non-Gaussian component of the jet response function.

or

$$\begin{aligned} \pi - |\Delta\phi(j^N, E_{\mathrm{T}}^{\mathrm{miss}})| &< |\Delta\phi(j^1, E_{\mathrm{T}}^{\mathrm{miss}})| \\ \pi - |\Delta\phi(j^N, E_{\mathrm{T}}^{\mathrm{miss}})| &< \Delta\phi^{\mathrm{match}} \\ \pi - |\Delta\phi(j^{N-1}, E_{\mathrm{T}}^{\mathrm{miss}})| &> \Delta\phi_1^{\mathrm{isol}}, \end{aligned} \tag{7.9}$$

where the indices refer to position in the $\Delta\phi$ ordering. $\Delta\phi^{\mathrm{match}}$ gives how close in ϕ the candidate jet should be to the $E_{\mathrm{T}}^{\mathrm{miss}}$. $\Delta\phi_1^{\mathrm{isol}}$ gives how isolated the next closest jet must be. Applying these cuts selects events where the jet most closely associated with the $E_{\mathrm{T}}^{\mathrm{miss}}$ is sufficiently well associated to be considered a match and is isolated from any other jets in the transverse plane. The fluctuating jet is then labelled as J and the jet closest to being back-to-back with J is labelled as K. This corresponds to $J = j^1$ and $K = j^N$ for the parallel selection (Eq. 7.8), and $J = j^N$ and $K = j^1$ for the anti-parallel selection (Eq. 7.9).

In order to avoid ambiguity, events containing jets both parallel *and* anti-parallel to the $E_{\mathrm{T}}^{\mathrm{miss}}$ are rejected. Figure 7.6 shows the $|\Delta\phi(K, E_{\mathrm{T}}^{\mathrm{miss}})|$ distribution for both parallel and anti-parallel selections with $\Delta\phi^{\mathrm{match}} = 0.1$ and $\Delta\phi_1^{\mathrm{isol}} = 1.0$. It is clear that K tends to be back-to-back with J. To resolve this ambiguity, events passing

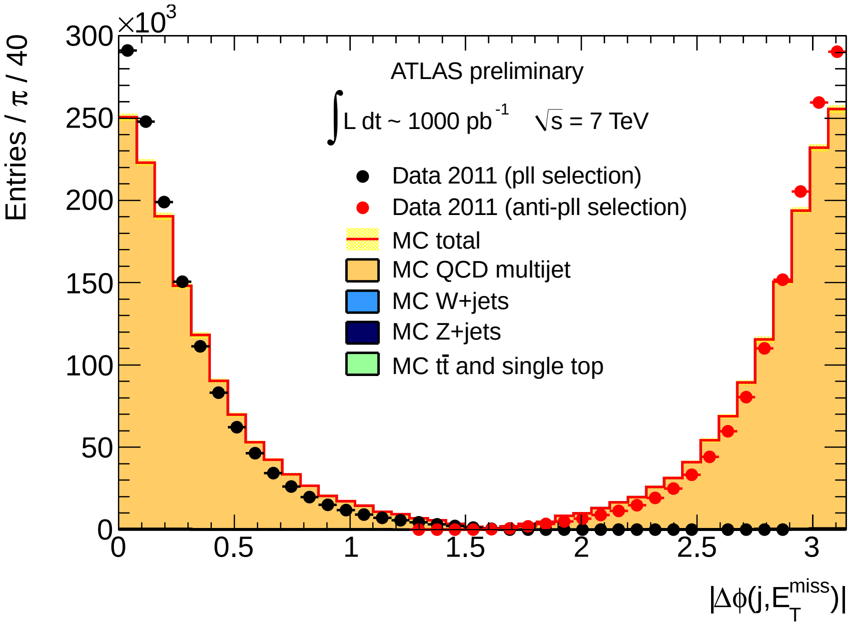


Figure 7.5: $\Delta\phi$ between $E_{\mathrm{T}}^{\mathrm{miss}}$ and nearest (black) and furthest (red) jet for events passing preselection cuts. Multijet MC is normalised to the data. Statistical uncertainties are shown.

the cuts in Eq. 7.8 and Eq. 7.9 are kept if they satisfy:

$$\pi - |\Delta\phi(K, E_{\mathrm{T}}^{\mathrm{miss}})| > \Delta\phi_2^{\mathrm{isol}} \tag{7.10}$$

or

$$|\Delta\phi(K, E_{\mathrm{T}}^{\mathrm{miss}})| > \Delta\phi_2^{\mathrm{isol}} \tag{7.11}$$

respectively.

Finally, in order to reduce the contamination from non-multijet sources (like $Z(\rightarrow \nu\bar{\nu})$ + jets), events are required to have high p_{T} jets other than the candidate fluctuating jet. In each event, jets other than J are sorted in decreasing p_{T} and the

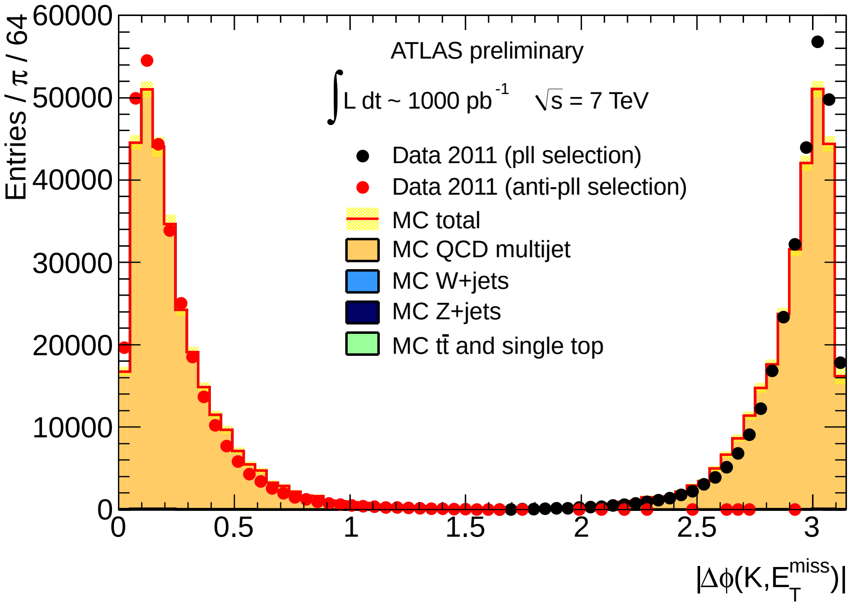


Figure 7.6: $\Delta\phi$ between $E_{\mathrm{T}}^{\mathrm{miss}}$ and K for parallel (black) and anti-parallel (red) selections. Events pass preselection cut and pass Eq. 7.8 or Eq. 7.9 with $\Delta\phi^{\mathrm{match}} = 0.1$ and $\Delta\phi_1^{\mathrm{isol}} = 1.0$. Multijet MC is normalised to the data. Statistical uncertainties are shown.

event is selected if

$$\begin{aligned} p_{\mathrm{T}}^1 &> p_{\mathrm{T}}^{\mathrm{cut1}} \\ p_{\mathrm{T}}^2 &> p_{\mathrm{T}}^{\mathrm{cut2}} \end{aligned} \tag{7.12}$$

where the indices refer to the position (excluding J) in the p_{T} ordering. The final selection is made using the cuts in Eq. 7.8 - Eq. 7.12 with the following values:

$$\begin{aligned}\Delta\phi^{\mathrm{match}} &= 0.1\\ \Delta\phi_1^{\mathrm{isol}} &= 1.0\\ \Delta\phi_2^{\mathrm{isol}} &= 0.25\\ p_{\mathrm{T}}^{\mathrm{cut1}} &= 260\ \mathrm{GeV}\\ p_{\mathrm{T}}^{\mathrm{cut2}} &= 40\ \mathrm{GeV}\end{aligned} \tag{7.13}$$

Having selected events with a jet unambiguously associated with the $E_{\mathrm{T}}^{\mathrm{miss}}$, the jet response tails can be measured. If it is assumed that the $E_{\mathrm{T}}^{\mathrm{miss}}$ in these events is dominated by the response fluctuation in J, then:

$$\vec{p}_{\mathrm{T}}^{J}(\mathrm{true}) \simeq \vec{p}_{\mathrm{T}}^{J}(\mathrm{reco}) + \vec{E}_{\mathrm{T}}^{\mathrm{miss}}. \tag{7.14}$$

With response defined as:

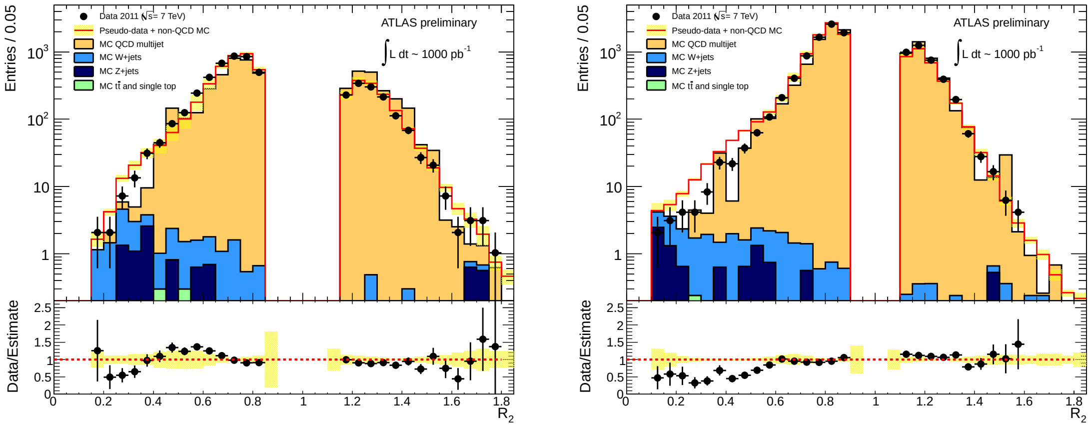
$$R = \frac{\vec{p}_{\mathrm{T}}^{J}(\mathrm{reco}) \cdot \vec{p}_{\mathrm{T}}^{J}(\mathrm{true})}{|\vec{p}_{\mathrm{T}}^{J}(\mathrm{true})|^2}, \tag{7.15}$$

then by substitution of Eq. 7.14, response measured with Mercedes events, R_2, is given by:

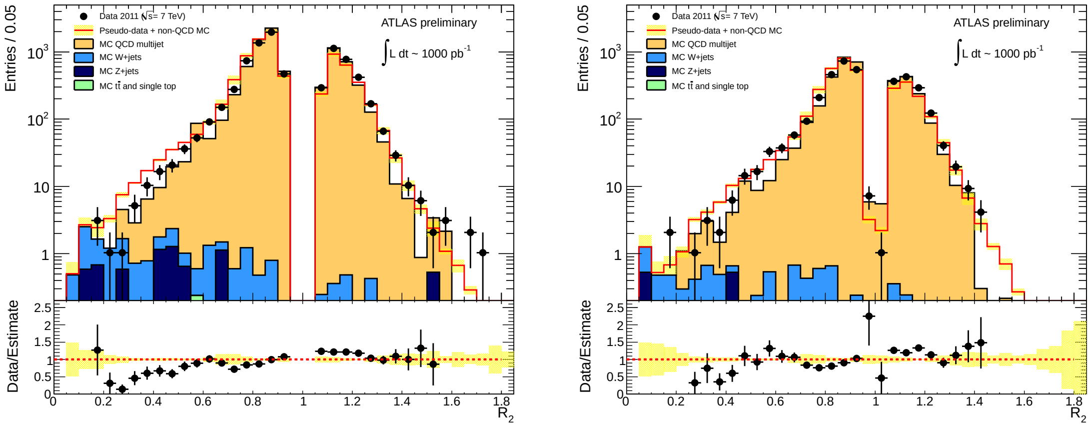
$$R_2 \simeq \frac{\vec{p}_{\mathrm{T}}^{J} \cdot (\vec{p}_{\mathrm{T}}^{J} + \vec{E}_{\mathrm{T}}^{\mathrm{miss}})}{|\vec{p}_{\mathrm{T}}^{J} + \vec{E}_{\mathrm{T}}^{\mathrm{miss}}|^2}, \tag{7.16}$$

where $\vec{p}_{\mathrm{T}}^{J}$ is the *reconstructed* p_{T} of the jet associated with the $E_{\mathrm{T}}^{\mathrm{miss}}$.

Figure 7.7 shows the R_2 distributions for different $\vec{p}_{\mathrm{T}}^{J}(\mathrm{true})$ ranges of the fluctuating jet measured with data, MC simulation and estimated using the jet smearing method with Gaussian corrections and uncertainties accounted for.



(a) R_2 for 100 GeV $< p_{\rm T}$(true) $<$ 200 GeV (b) R_2 for 200 GeV $< p_{\rm T}$(true) $<$ 300 GeV



(c) R_2 for 300 GeV $< p_{\rm T}$(true) $<$ 400 GeV (d) R_2 for 400 GeV $< p_{\rm T}$(true) $<$ 1000 GeV

Figure 7.7: R_2 in bins of fluctuating jet $\vec{p}_{\rm T}^{\,J}$(true) measured in data, MC simulation and estimated using the jet smearing method with Gaussian corrections and uncertainties accounted for. No tail modification is used in these distributions. Multijet MC and estimates are normalised to the data. Statistical uncertainties are shown on the data and Gaussian systematic uncertainties are shown on the estimates.

7.4.1 Modification of Non-Gaussian Response

Figure 7.7 shows that the jet smearing estimate without tail corrections does not perfectly reconstruct the low-side non-Gaussian tail observed in the data. This is especially clear in Figure 7.7(b) where the estimate overshoots the data in the region $0.0 < R_2 < 0.5$. In the high-side region ($R_2 \gg 1.0$), good agreement is observed between the estimate and data indicating that no special treatment is needed for the high-side non-Gaussian tail.

In order to modify the tail, firstly the MC response functions are fitted with a functional form. The tail can then be modified by changing a parameter in the functional form. The jet smearing estimates of R_2 (like those in Figure 7.7) can then be reproduced using the modified tail and the agreement between the estimate and data can be probed. By repeating this procedure (using different values for the parameter), the optimal form of non-Gaussian tail can be found with an associated uncertainty.

The functional form found to best describe the MC non-Gaussian jet response is a Gaussian with a fitting range restricted to the tail region. In order to define this region, the response is first fitted with a Crystal Ball function [62] (which has a continuous derivative) in the range $0.0 < R < 1.2$. It is important to note that this function is not expected to provide a perfect fit to the response distribution; it is only used as a tool to determine the extent of the tail region. The Crystal Ball function is defined as:

$$f(x;\alpha,n,\bar{x},\sigma,N) = N \cdot \begin{cases} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot \left(B - \frac{x-\bar{x}}{\sigma}\right)^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leq -\alpha \end{cases} \tag{7.17}$$

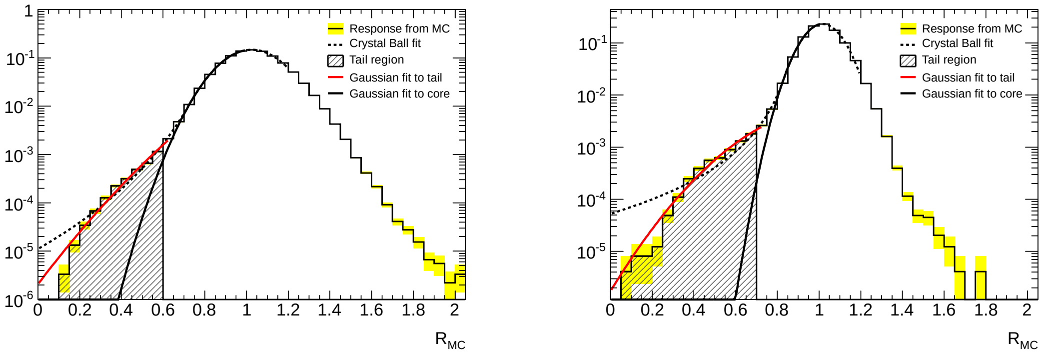
where

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right)$$

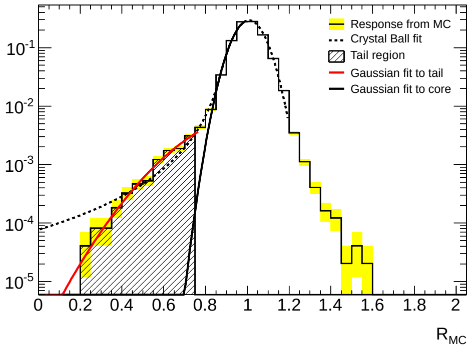
and

$$B = \frac{n}{|\alpha|} - |\alpha|.$$

The crossing point between the core and tail regions is therefore at $x = \bar{x} - \alpha\sigma$. To ensure the tail dominates, the 'tail Gaussian' is fitted in the range $0 < x < \bar{x} - \alpha\sigma - 0.1$. Examples of the Crystal Ball fits and the fitting ranges determined are shown in Figure 7.8. Also shown are Gaussians fitted to the tail region. The full jet response is then the Gaussian fit in the tail region and the original MC response histogram outside of the tail region.



(a) 40 GeV $< p_{\mathrm{T}}(\text{true}) <$ 60 GeV (b) 140 GeV $< p_{\mathrm{T}}(\text{true}) <$ 160 GeV



(c) 540 GeV $< p_{\mathrm{T}}(\text{true}) <$ 560 GeV

Figure 7.8: Fitting the MC response tail. For every p_{T} slice, a Crystal Ball function is fitted to the R distribution in the region $0.0 < R < 1.2$. The tail region is then defined as $0 < R < \bar{x} - \alpha\sigma - 0.1$ (see Eq. 7.17 and text). A Gaussian is then fitted in this region. Also shown is the Gaussian component of the Crystal Ball fit, this is used in producing a sample of pseudo-data with only Gaussian smearing as described in §7.4.2. The yellow band shows the uncertainty from limited MC statistics.

The response tail can now be modified by changing the shape of the Gaussian fit. This is achieved by multiplying the width of the Gaussian (σ_{tail}) by a factor $\Delta\sigma_{tail}$. The mean of the Gaussian is fixed and the normalisation is set such that the value of the function at $x = \bar{x} - \alpha\sigma - 0.1$ is fixed. Figure 7.9 shows examples of how the response varies with the choice of $\Delta\sigma_{tail}$ for jets with 140 GeV $< p_{\mathrm{T}}(\text{true}) <$ 160 GeV.

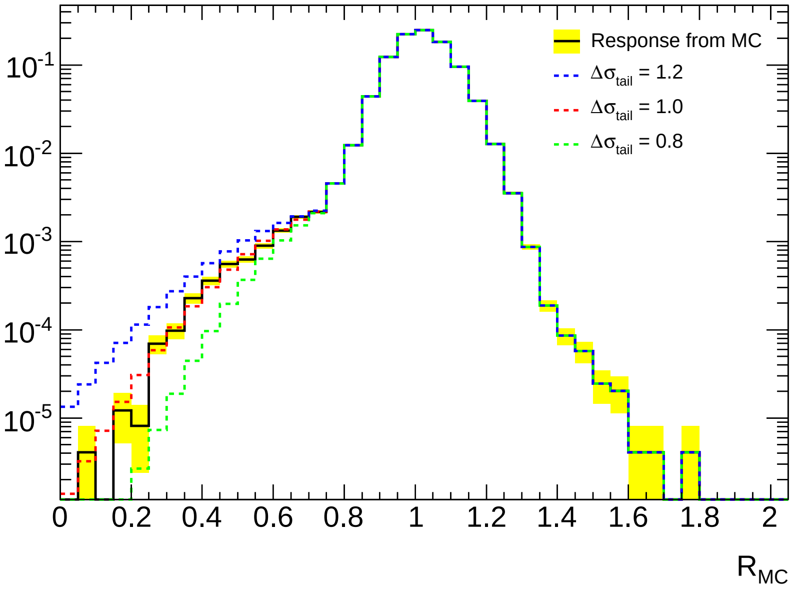
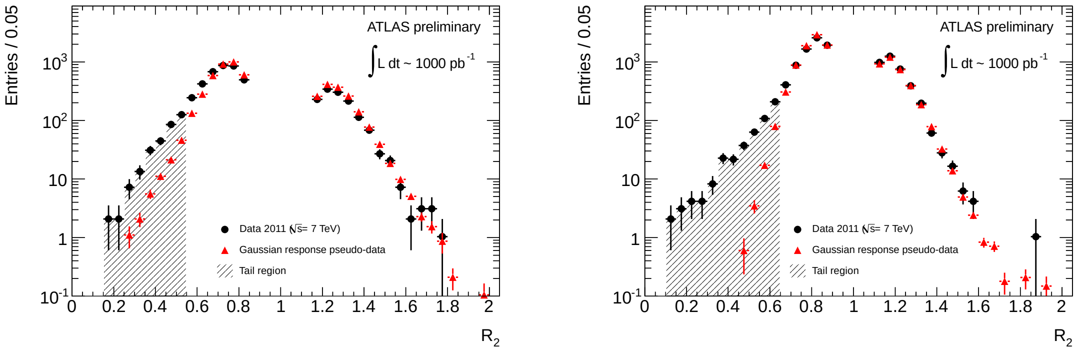


Figure 7.9: Modifying the response tail using the $\Delta\sigma_{tail}$ parameter as described in the text for jets with 140 GeV $< p_{\mathrm{T}}(\text{true}) <$ 160 GeV. The yellow band shows the uncertainty from limited MC statistics.

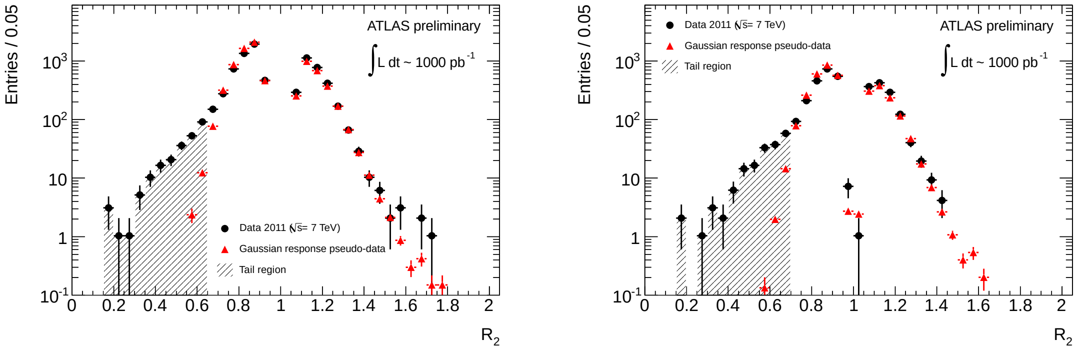
7.4.2 Determining $\Delta\sigma_{tail}$ and its Uncertainty

In order to test the compatibility of the modified tails with the observation in the data, pseudo-data samples are produced with different $\Delta\sigma_{tail}$ values. χ^2 tests are then performed between the Mercedes distributions in the pseudo-data samples and in the data.

The first task of this analysis is to define the regions in which the χ^2 tests are performed. This is important as only differences in the tail shape should be probed, and not differences in the Gaussian region of the Mercedes distributions. To achieve this, one must define in what region the Mercedes distributions are dominated by tail fluctuations. This is achieved by creating a sample of pseudo-data only using Gaussian smearing. The response functions for this sample are taken from the Crystal Ball fits described in §7.4.1. The fits are used to define the parameters for Gaussians which are a good match to the core of the response distributions as shown in Figure 7.8. For a particular Mercedes p_{T} bin, the tail region is then defined as those bins in which the bin content of the Gaussian smeared pseudo-data distribution is less than half the bin content of the data. This is demonstrated in Figure 7.10.



(a) 100 GeV $< p_{\mathrm{T}}(\mathrm{true}) <$ 200 GeV (b) 200 GeV $< p_{\mathrm{T}}(\mathrm{true}) <$ 300 GeV



(c) 300 GeV $< p_{\mathrm{T}}(\mathrm{true}) <$ 400 GeV (d) 400 GeV $< p_{\mathrm{T}}(\mathrm{true}) <$ 1000 GeV

Figure 7.10: R_2 in bins of fluctuating jet $\vec{p}_{\mathrm{T}}^{J}(\mathrm{true})$ measured in data and with a sample of pseudo-data created with Gaussian-only smearing. Comparing these distributions defines the tail region (shown in the plots) for performing χ^2 tests between data and pseudo-data (including the non-Gaussian tail) as described in the text. Statistical uncertainties are shown.

Having defined the tail regions of the Mercedes distributions, χ^2 tests can now be performed between the pseudo-data and data with different values of $\Delta\sigma_{tail}$. For these tests, the pseudo-data is compared with the data with the non-multijet MC subtracted from it. As can be seen in Figure 7.7, in the region $0.0 < R < 0.4$ there can be a significant non-multijet contamination and there can be bins with no data observed. For these reasons, the first bins are merged until the amount of data observed is greater than the non-multijet MC and furthermore there are no bins missing data to the right of this first bin. For example in Figure 7.7(b), the first bin is taken as $0.0 < R < 0.3$. Sets of pseudo-data are created with $\Delta\sigma_{tail}$ varying from 0.5 to 1.5 in steps of 0.05. The χ^2 values are then calculated. For each p_{T}(true) bin three values of $\Delta\sigma_{tail}$ are found:

- $\Delta\sigma_{tail}^{optimal}$: the value of $\Delta\sigma_{tail}$ which gives the highest p-value[2].
- $\Delta\sigma_{tail}^{high}$: the value of $\Delta\sigma_{tail}$ above $\Delta\sigma_{tail}^{optimal}$ at which the hypothesis of a good fit is rejected at a one sigma level. i.e. The first parameter value above $\Delta\sigma_{tail}^{optimal}$ at which the p-value is found to be less than 0.32.
- $\Delta\sigma_{tail}^{low}$: the value of $\Delta\sigma_{tail}$ below $\Delta\sigma_{tail}^{optimal}$ at which the hypothesis of a good fit is rejected at a one sigma level.

The parameter values found are shown in Table 7.2. These values are used to produce the pseudo-data distributions shown in Figure 7.11.

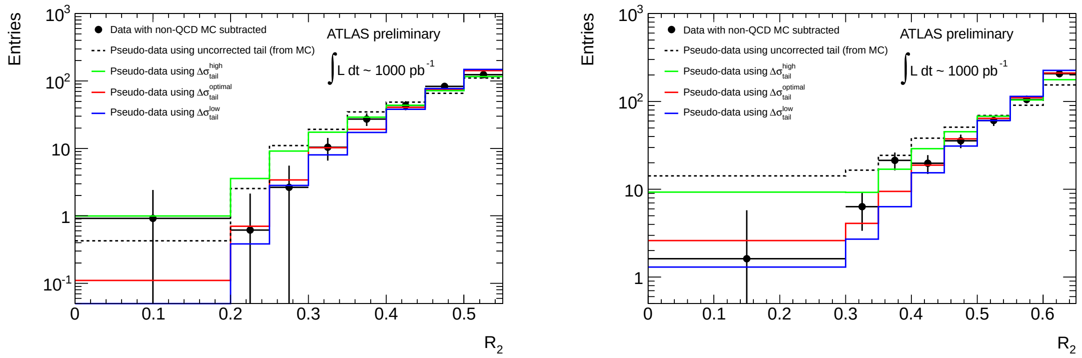
To obtain a given distribution or quantity, the estimate using $\Delta\sigma_{tail}^{optimal}$ is taken as the optimal estimate. The tail uncertainty is then taken as half the difference between the estimates acquired using $\Delta\sigma_{tail}^{high}$ and $\Delta\sigma_{tail}^{low}$. This procedure is used for obtaining the jet smearing estimate of the Mercedes distributions shown in Figure 7.12. This can be compared to Figure 7.7 where tail corrections have not been performed. For completeness, there should also be a correction and uncertainty derived for the tail normalisation (compared to the Gaussian component of jet response). However, the effect of such a correction is small compared with the effect from the shape change discussed in this section and therefore is not included in this analysis.

---

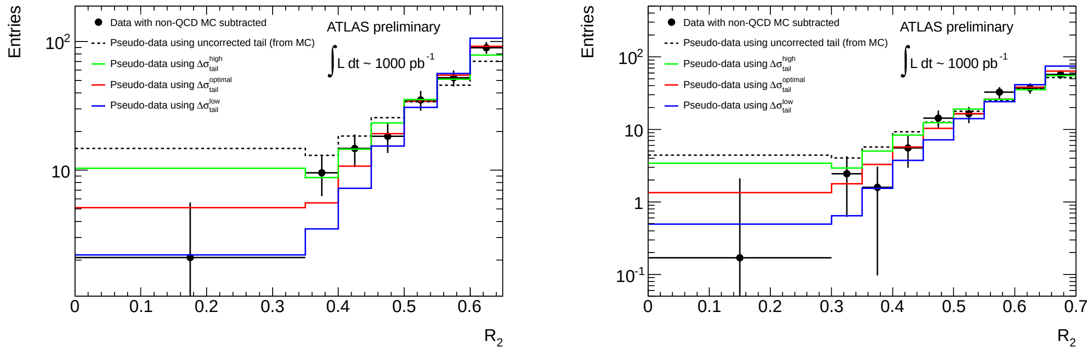
[2] The p-value gives the probability of the two distributions being consistent

p_{T} (true) range [GeV]	Parameter value, p-value		
	$\Delta\sigma_{tail}^{low}$	$\Delta\sigma_{tail}^{optimal}$	$\Delta\sigma_{tail}^{high}$
$0 < p_{\mathrm{T}} < 200$	0.75, 0.24	0.8, 0.56	1.0, 0.06
$200 < p_{\mathrm{T}} < 300$	0.75, 0.04	0.8, 0.42	0.9, 0.02
$300 < p_{\mathrm{T}} < 400$	0.75, 0.10	0.85, 0.76	0.95, 0.23
$p_{\mathrm{T}} > 400$	0.8, 0.13	0.9, 0.79	1.0, 0.18

Table 7.2: Values of $\Delta\sigma_{tail}$ calculated using χ^2 tests between the Mercedes distributions in data and pseudo-data.

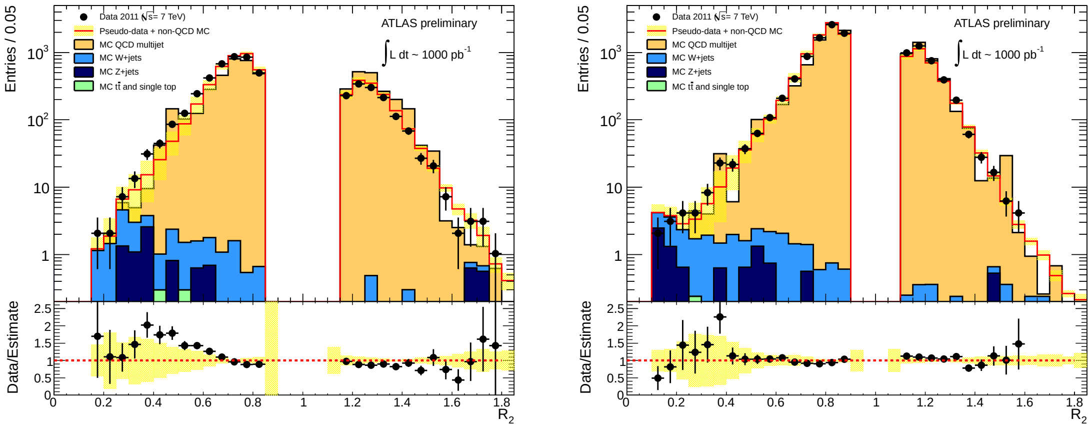


(a) 100 GeV < p_{T}(true) < 200 GeV (b) 200 GeV < p_{T}(true) < 300 GeV

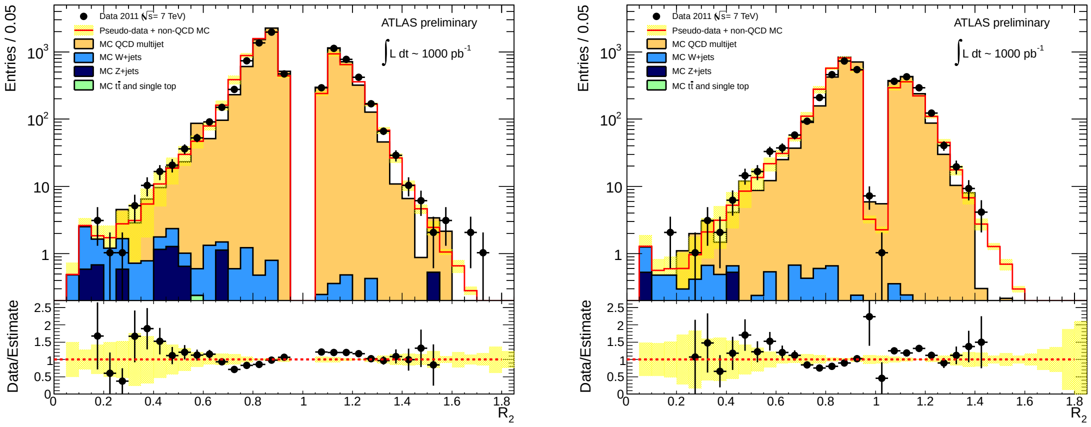


(c) 300 GeV < p_{T}(true) < 400 GeV (d) 400 GeV < p_{T}(true) < 1000 GeV

Figure 7.11: The tail regions of R_2 in bins of fluctuating jet $\vec{p}_{\mathrm{T}}^{\,J}$(true) measured in data with non-multijet MC subtracted from it and estimated using the jet smearing method using the values of $\Delta\sigma_{tail}$ shown in Table 7.2. Also shown are the estimates using the uncorrected response tails from multijet MC. Statistical uncertainties are shown on the data.



(a) R_2 for 100 GeV $< p_{\mathrm{T}}(\mathrm{true}) <$ 200 GeV (b) R_2 for 200 GeV $< p_{\mathrm{T}}(\mathrm{true}) <$ 300 GeV



(c) R_2 for 300 GeV $< p_{\mathrm{T}}(\mathrm{true}) <$ 400 GeV (d) R_2 for 400 GeV $< p_{\mathrm{T}}(\mathrm{true}) <$ 1000 GeV

Figure 7.12: R_2 in bins of fluctuating jet $\vec{p}_{\mathrm{T}}^{\,J}(\mathrm{true})$ measured in data, MC simulation and estimated using the jet smearing method with Gaussian and tail corrections and associated uncertainties accounted for. Multijet MC and estimates are normalised to the data. Statistical uncertainties are shown on the data. The yellow bands show the systematic uncertainties on the estimates.

7.4.3 Dedicated $\Delta\phi$ Trigger

The lower-p_{T} single jet triggers which could be used for collecting Mercedes events in §7.4 have high average prescaling as shown in Table 4.1; this is due to the high cross-section for multijet production at the LHC. This partly motivates the high leading jet p_{T} cut ($p_{\mathrm{T}} > 260$ GeV) used in the Mercedes analysis. To allow the use of a lower p_{T} cut, and hence have more available statistics for the measurement of non-Gaussian jet response, novel event shape triggers called "$\Delta\phi$ triggers" have been developed. These triggers are designed to be fully efficient for triggering Mercedes events whilst rejecting a significant number of other multijet events. This allows the trigger to collect a greater number of events useful for the Mercedes analysis whilst running with the same trigger bandwidth as an equivalent single jet trigger.

Figure 7.13 shows the $\min(\Delta\phi(j_1, j_2))$ distribution of all combinations of pairs of jets (referred to as $\Delta\phi^{\min}$ henceforth) passing the trigger selection described in Table 4.1. The majority of events either have jets which are back-to-back ($\Delta\phi^{\min} \sim \pi$)

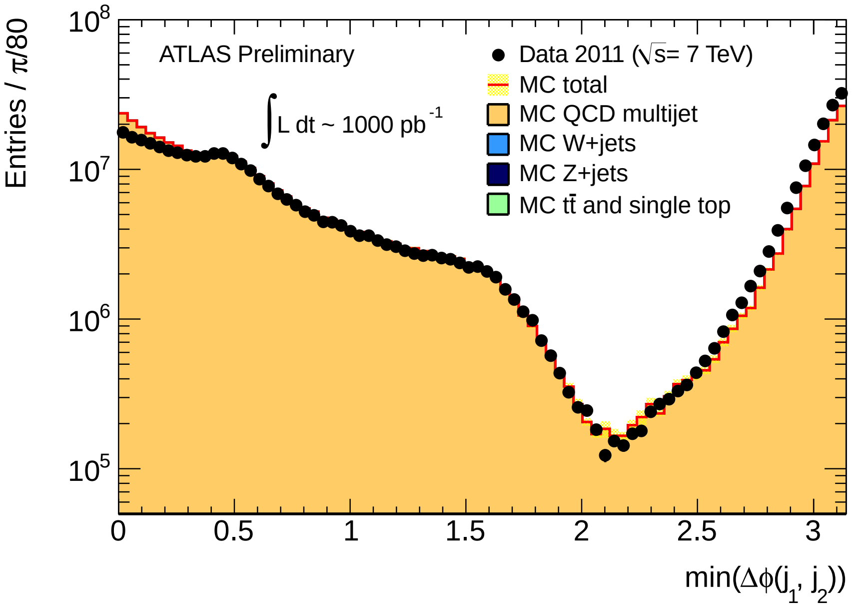


Figure 7.13: $\Delta\phi^{\min}$ between pairs of jets in data and MC. Multijet MC is normalised to the data. Statistical uncertainties are shown.

or are close to each other ($\Delta\phi^{\text{min}} \sim 0$) in ϕ. Both of these cases fail the Mercedes event selection as the candidate jet must be isolated from other jets and not back-to-back with other jets in the event. For this reason, the quantity $\Delta\phi^{\text{min}}$ is an ideal one to trigger on.

To inform the choice of the $\Delta\phi^{\text{min}}$ cut, the offline $\Delta\phi^{\text{min}}$ distribution is plotted for events passing the full Mercedes event selection described in §7.4; this is shown in Figure 7.14. It is clear that an offline cut of $\Delta\phi^{\text{min}} < 2.2$ would be fully efficient for the Mercedes selection.

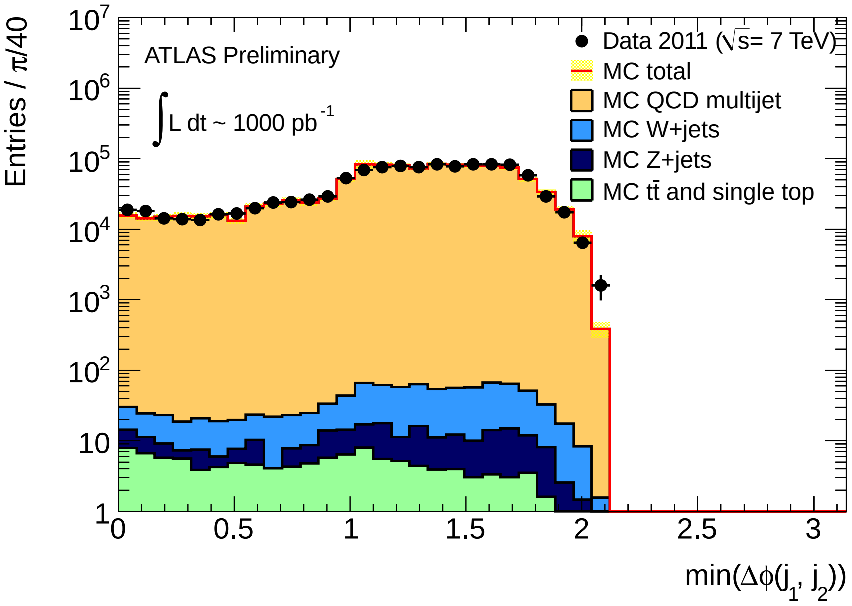
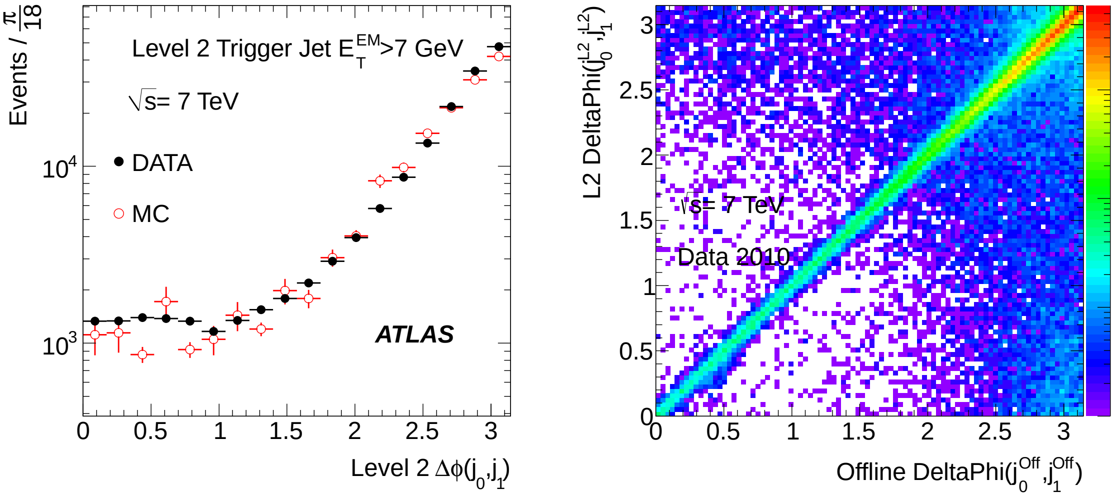
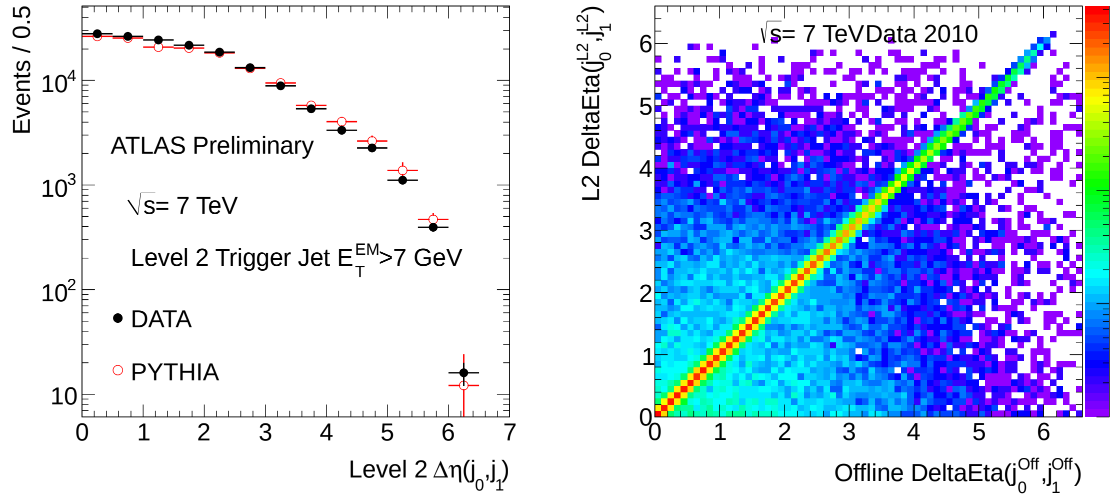


Figure 7.14: $\Delta\phi^{\text{min}}$ between pairs of jets for events passing the Mercedes selection in data and MC. Multijet MC is normalised to the data. Statistical uncertainties are shown.

The first $\Delta\phi$ triggers developed are L2 triggers. To aid their development, L2 jet kinematics are studied using ATLAS data recorded in 2010. Some key distributions are shown in Figure 7.15. Figures 7.15(a) and 7.15(c) show the $\Delta\phi$ and $\Delta\eta$ distributions between the two leading momenta jets at L2 (for jets reconstructed at L2 with $E_{\text{T}} > 7$ GeV). The MC simulation of the L2 jet kinematics is seen to model the data reasonably well. Figures 7.15(b) and 7.15(d) demonstrate that in the majority of cases, the L2 $\Delta\phi$ and $\Delta\eta$ values match the offline (events on the diagonal). There are also a large number of events lying far from the diagonal, in these events the momentum ordering of the jets is different between L2 and offline. This can occur due to the simplicity of the L2 jet algorithms compared to offline and the fact that no JES correction is applied at L2. This partly motivates the use of the $\Delta\phi^{\mathrm{min}}$ variable in the $\Delta\phi$ triggers instead of $\Delta\phi$ between the two highest momenta jets.



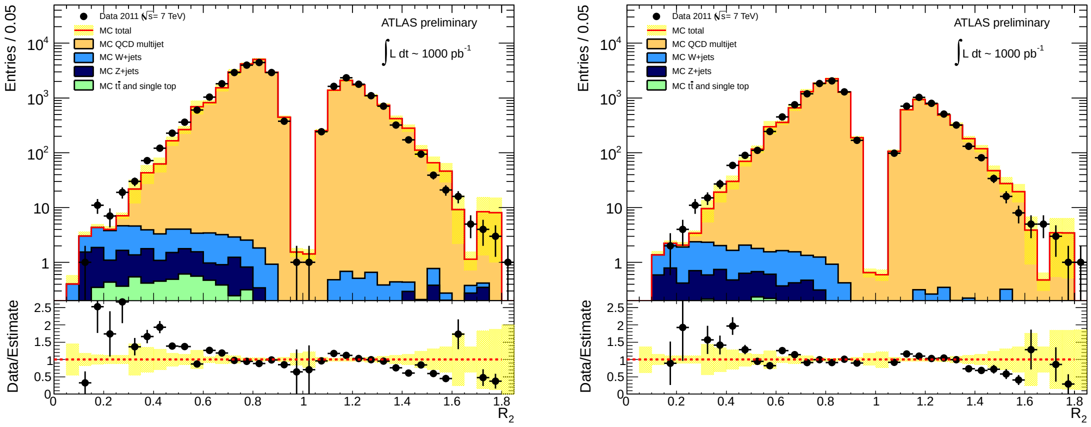
(a) $\Delta\phi$ at Level 2 (b) $\Delta\phi$ at L2 compared to offline



(c) $\Delta\eta$ at Level 2 (d) $\Delta\eta$ at L2 compared to offline

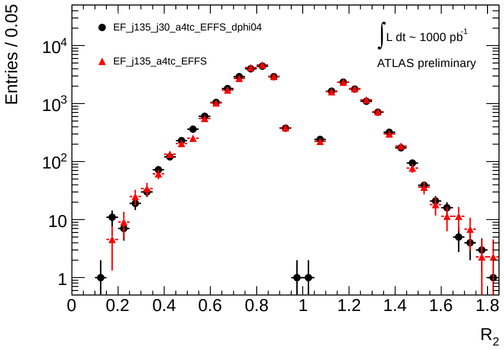
Figure 7.15: L2 jet kinematics. MC estimates are normalised to the data.

For the dataset considered in the rest of this thesis, there were two $\Delta\phi$ triggers active: EF_j135_j30_a4tc_EFFS_dphi04 and EF_j180_j30_a4tc_EFFS_dphi04. These are EF triggers giving the advantage that the jets considered are a closer match to the offline jets compared with L2 jets. The 'dphi04' notation corresponds to $\Delta\phi^{\mathrm{min}} < \pi - 0.4$. The EF_j180_j30_a4tc_EFFS_dphi04 trigger is not needed in the dataset of this thesis as the equivalent single jet trigger (EF_j180_a4tc_EFFS) was unprescaled as shown in Table 4.1. This trigger is used in later data periods beyond the scope of this thesis. The EF_j135_j30_a4tc_EFFS_dphi04 trigger is used as it had an average prescale of 3.4 compared to 8.9 for the corresponding single jet trigger: EF_j135_a4tc_EFFS. The Mercedes distribution for $p_{\mathrm{T}}(\mathrm{true}) > 100$ GeV measured using this trigger and with the leading jet p_{T} cut and $p_{\mathrm{T}}^{\mathrm{cut1}}$ set to 200 GeV is shown in Figure 7.16(a). For comparison the same distribution except using the EF_j135_a4tc_EFFS trigger is plotted in Figure 7.16(b). To easily compare the shapes, Figure 7.16(c) shows the two distributions on the same plot, with the data from EF_j135_a4tc_EFFS normalised to the $\Delta\phi$ trigger. The distribution from the $\Delta\phi$ trigger matches that from the single jet trigger (within uncertainties) indicating that the $\Delta\phi$ trigger does not bias the R_2 distribution.



(a) EF_j135_j30_a4tc_EFFS_dphi04

(b) EF_j135_a4tc_EFFS



(c) Comparison. EF_j135_a4tc_EFFS histogram normalised to EF_j135_j30_a4tc_EFFS_dphi04 histogram.

Figure 7.16: R_2 for $p_{\mathrm{T}}(\mathrm{true}) > 100$ GeV measured in data and MC simulated data with the $\Delta\phi$ trigger: EF_j135_j30_a4tc_EFFS_dphi04, and the corresponding single jet trigger: EF_j135_a4tc_EFFS. Multijet MC is normalised to the data. Statistical uncertainties are shown.

7.4.4 Using Events from $\Delta\phi$ Trigger to Constrain $\Delta\sigma_{tail}$

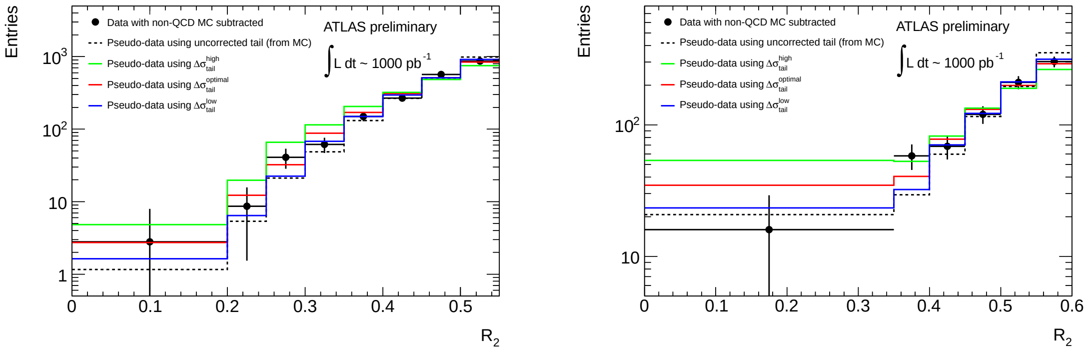
Utilising the events from the EF_j135_j30_a4tc_EFFS_dphi04 trigger may allow the $\Delta\sigma_{tail}$ parameter to be further constrained and hence reduce an uncertainty on the shape of the non-Gaussian response tail. To achieve this, the procedure described in §7.4.2 is repeated except that the leading jet p_{T} cut and $p_{\mathrm{T}}^{\mathrm{cut1}}$ are set to 200 GeV. For the 'data' distribution, the EF_j135_j30_a4tc_EFFS_dphi04 is used for events with 200 GeV $< p_{\mathrm{T}}(\mathrm{leading}) <$ 260 GeV instead of the EF_j135_a4tc_EFFS trigger. Note that pseudo-data is produced in the normal way, following the usual single jet trigger selection shown in Table 4.1.

The statistical uncertainty for bins in the data R_2 histogram is underestimated in the case of there being zero events observed from the prescaled $\Delta\phi$ trigger. For example, take a bin with N events from the unprescaled trigger and 0 events from the prescaled trigger; the statistical uncertainty would usually be $\sqrt{N}$ on this bin. However, an entry from the prescaled trigger does not just add one event, but an event with a weight: w_{PS}. Therefore the upper uncertainty on this bin should be $\sqrt{N} \oplus \sqrt{w_{PS}} = \sqrt{N^2 + w_{PS}^2}$. Therefore, for any bins whose error is less than 3.4 (this corresponds to w_{PS} for EF_j135_j30_a4tc_EFFS_dphi04), the upper error is set to $\sqrt{N^2 + 3.4^2}$. The lower error is not modified. Table 7.3 shows the values of $\Delta\sigma_{tail}$ which are obtained and Figure 7.17 shows the R_2 distributions in the tail region used for the χ^2 fits.

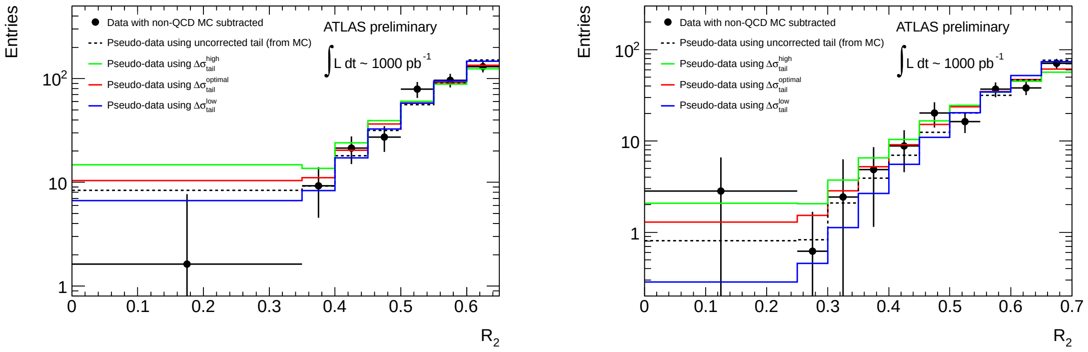
Comparing Table 7.3 with Table 7.2, it can be seen that using the $\Delta\phi$ trigger does further constrain $\Delta\sigma_{tail}$ compared to using only single jet triggers. The consequence of this will be discussed in Chapter 8.

p_{T} (true) range [GeV]	Parameter value, p-value		
	$\Delta\sigma_{tail}^{low}$	$\Delta\sigma_{tail}^{optimal}$	$\Delta\sigma_{tail}^{high}$
$0 < p_{\mathrm{T}} < 200$	0.8, 0.14	0.85, 0.68	0.95, 0.01
$200 < p_{\mathrm{T}} < 300$	0.8, 0.10	0.85, 0.39	0.9, 0.27
$300 < p_{\mathrm{T}} < 400$	0.8, 0.19	0.85, 0.36	0.9, 0.16
$p_{\mathrm{T}} > 400$	0.8, 0.14	0.95, 0.81	1.0, 0.11

Table 7.3: Values of $\Delta\sigma_{tail}$ calculated using χ^2 tests between the Mercedes distributions in data and pseudo-data using the $\Delta\phi$ trigger: EF_j135_j30_a4tc_EFFS_dphi04.



(a) 100 GeV < p_{T}(true) < 200 GeV (b) 200 GeV < p_{T}(true) < 300 GeV



(c) 300 GeV < p_{T}(true) < 400 GeV (d) 400 GeV < p_{T}(true) < 1000 GeV

Figure 7.17: The tail regions of R_2 in bins of fluctuating jet $\vec{p}_{\mathrm{T}}^{\,J}$(true) measured in data, using the $\Delta\phi$ trigger EF_j135_j30_a4tc_EFFS_dphi04 with non-multijet MC subtracted from it and estimated using the jet smearing method using the values of $\Delta\sigma_{tail}$ shown in Table 7.3. Also shown are the estimates using the uncorrected response tails from multijet MC. Statistical uncertainties are shown on the data.

Chapter 8

Jet Smearing for Multijet Estimation

This chapter describes the final multijet background estimation for the fully hadronic SUSY search (see Chapter 5) using the jet smearing method. The response functions used are those measured in Chapter 7. Firstly, in §8.1 the technique is shown to successfully reproduce the multijet $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution observed in data. In §8.2, the method used for performing the final multijet estimation is given. §8.3 gives the uncertainties associated with the jet smearing method with systematic uncertainties dominating. §8.4 presents the results of the analysis, with estimated multijet distributions compared to the data in the SUSY control, signal and validation regions. The estimated transfer factors between the control and signal regions are also shown. Finally, in §8.5, the results are discussed. The multijet distributions estimated with the jet smearing method display excellent agreement with the data in SUSY control and validation regions. In the signal regions, the benefit in the extra statistics compared with multijet MC is evident.

8.1 Validating the Method

Before obtaining the final multijet estimation for the SUSY signal and control regions, the method is tested by attempting to reproduce the multijet $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. Events must first satisfy the trigger selection shown in Table 4.1. Seed events are selected as described in §6.4 and 1000 smeared events are produced from each seed event as described in §6.2. A requirement is then made for the presence of two jets with $p_{\mathrm{T}} > 130, 40$ GeV in the data and smeared data. The $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in data, estimated with jet smearing and with MC simulated data is shown in Figure 8.1. The smeared data here is produced with response functions using the Gaussian and tail corrections and associated uncertainties described in Chapter 7.

8.2 Method for Final Multijet Estimation

Having constrained the response functions to successfully model multijet dominated distributions in the data, jet smearing can now be used to provide multijet distributions in SUSY signal and control regions. For the multijet estimation, only the signal regions (shown in Table 5.1) and CR2, the multijet control regions, are needed. For each signal region, CR2 has the same cuts except the $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ requirement changed to $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}}) < 0.2$.

Seed events are selected as shown in §6.4 and 10000 smeared events are produced from each seed as described in §6.2. The response functions used for the smearing are those constrained with multijet data as discussed in Chapter 7. The extra constraint on $\Delta\sigma_{tail}$ allowed by the $\Delta\phi$ triggers shown in §7.4.4 is not used to begin with. The smeared events are then passed through the relevant signal or control region selection. In comparison, data events are required to pass the signal trigger described in §5.2.1 and then the relevant signal or control region selection. The required inputs from this method to the full SUSY analysis are Transfer Factors (TFs) from the multijet

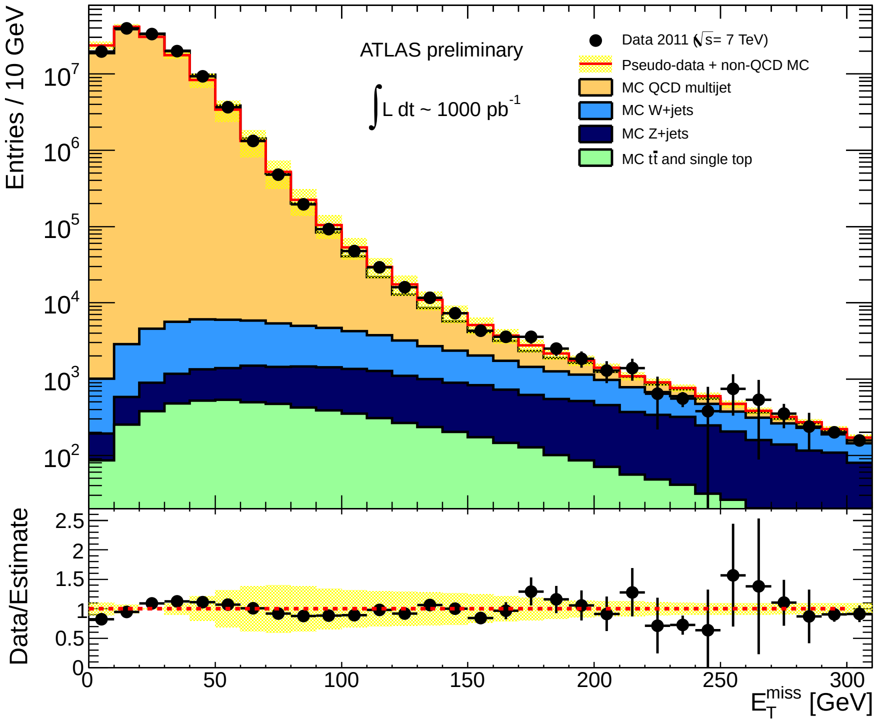


Figure 8.1: Multijet $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in data, MC simulated data and estimated using the jet smearing method with Gaussian and tail corrections and associated uncertainties accounted for. Multijet MC and estimate are normalised to the data. Statistical uncertainties are shown on the data. The yellow bands show the systematic uncertainty on the estimate.

control regions (CR2) to the signal regions (see §5.2):

$$\mathrm{TF} = \frac{N(\mathrm{SR}, \mathrm{multijet})}{N(\mathrm{CR2}, \mathrm{multijet})}. \tag{8.1}$$

This is just the predicted number of multijet events in the signal region divided by the predicted number in CR2. Therefore no normalisation of the multijet estimate is required. For the plots in this analysis, the multijet estimates from jet smearing and from Monte Carlo are normalised to the data using the CR2 cuts but without the final $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$ or m_{eff} cuts. As is explained in §5.2, the TF can be used to estimate the number of events from a particular background in a SR using the following equation:

$$\mathrm{N(SR, estimate)} = \mathrm{N(CR, observed)} \times \mathrm{TF(estimate)}. \qquad (8.2)$$

8.3 Uncertainties

The uncertainty on the final transfer factors is dominated by systematic uncertainties. The largest contribution is from the uncertainty on the non-Gaussian tail, introduced via the uncertainty on $\Delta\sigma_{tail}$.

8.3.1 Systematic Uncertainties

The systematic uncertainties accounted for in this analysis arise are from the uncertainties on the response functions used for the smearing process and from any biases introduced by the seed selection. For each uncertainty, the full estimate is repeated with the response function modified in the appropriate way; the uncertainty is then the deviation of this estimate from the original estimate.

Response function uncertainties

The uncertainties on the jet response fall into two types:

- The uncertainty on the Gaussian response width. This is introduced through the uncertainty on $\sigma_{correction}(p_{\mathrm{T}})$ described in §7.3.1 and the validity of the uncertainty can be observed in Figure 7.3(e).
- The uncertainty on the shape of the non-Gaussian tail. This is introduced via the uncertainty on the $\Delta\sigma_{tail}$ parameter as described in §7.4.2. Figure 7.12 shows the Mercedes distributions produced using the optimised values of $\Delta\sigma_{tail}$ including uncertainties.

Seed selection bias uncertainty

Figure 6.4(b) shows that a small bias is introduced to the leading jet p_{T} by the seed selection. To account for this, an estimate is performed with a constant factor of 0.1 added to jet response. This will, on average, increase smeared jet p_{T} by 10%. This is referred to as $R_{mean-shift} = +10\%$. Figure 8.2 shows leading jet p_{T} in data and estimated with jet smearing. This extra uncertainty is seen to account for differences observed between the data and original estimate.

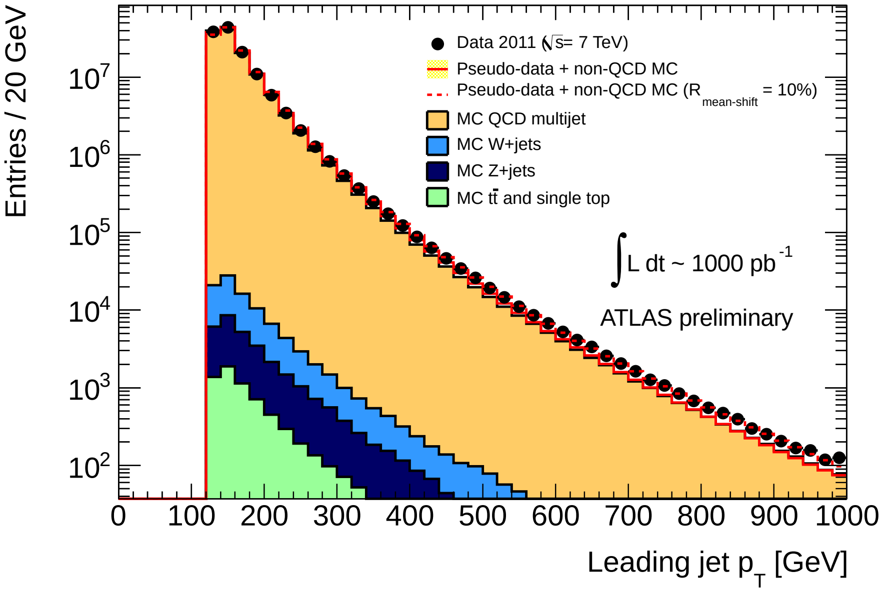


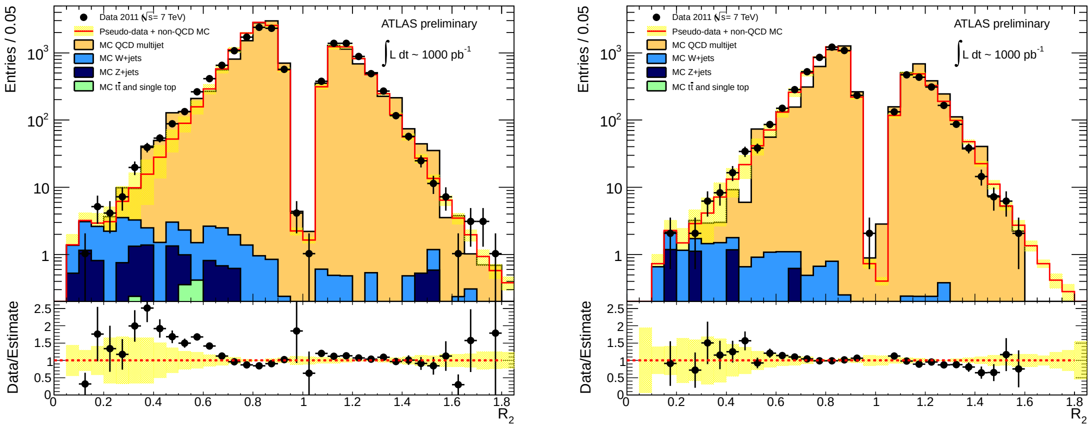
Figure 8.2: Leading jet p_{T} measured in multijet dominated data and reproduced using jet smearing. Multijet MC and estimates are normalised to the data.

Dependence of jet response on η

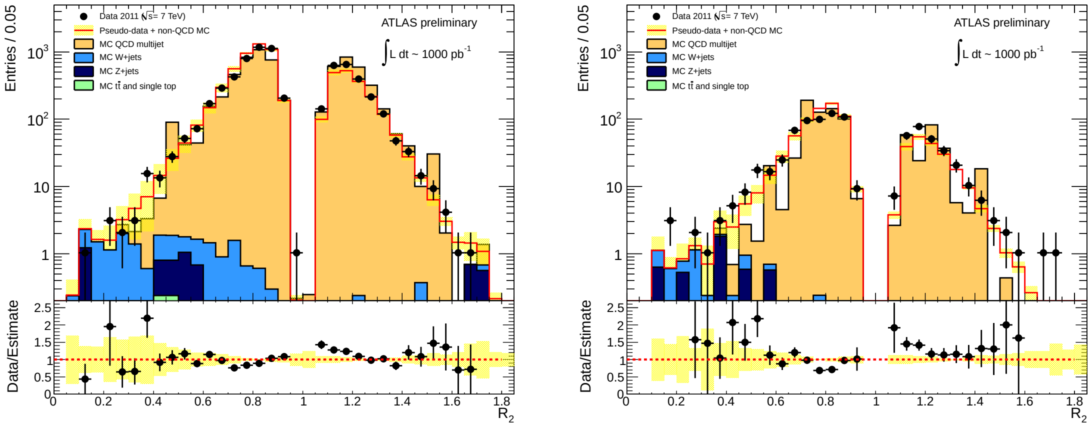
Jet response could be affected by the η coordinate of the fluctuating jet. For example, a jet in the crack region of the calorimeter would be expected to have a broader response than a jet in the central region. Binning jet response in p_{T} will partly account for these differences due to the correlation between p_{T} and η. The largest effect of η dependent response on the results of this analysis would be if the non-Gaussian tail shape were strongly dependent on η. This is checked by repeating the Mercedes analysis (see §7.4) in the following η bins:

- Central region ($|\eta| < 0.8$).
- Extended tile barrel ($0.8 < |\eta| < 1.2$).
- Crack region ($1.2 < |\eta| < 2.1$).
- End-cap region ($2.1 < |\eta| < 2.8$).

For the smearing estimates, Gaussian and tail corrections are applied to the response functions with uncertainties accounted for as described in §7.3 and §7.4. To increase statistics for this measurement, the p_{T} binning is removed so the only requirement is $p_{\mathrm{T}}(\text{true}) > 100$ GeV. The results are shown in Figure 8.3. It is clear that the estimate agrees well with the data indicating that non-Gaussian fluctuations are not strongly η dependent. The χ^2 p-values obtained between the tail regions in the data and estimates, calculated using the method described in §7.4.2 and accounting for uncertainties on $\Delta\sigma_{tail}$ and $\sigma_{correction}(p_{\mathrm{T}})$, are 0.98, 0.94, 0.66 and 0.96 for the central, extended tile barrel, crack and end-cap regions respectively. Given the good agreement observed between the data and estimates, no uncertainty is associated with the η dependence of jet response.



(a) R_2 for $p_{\mathrm{T}}(\mathrm{true}) > 100$ GeV and $|\eta| < 0.8$ (b) R_2 for $p_{\mathrm{T}}(\mathrm{true}) > 100$ GeV and $0.8 < |\eta| < 1.2$



(c) R_2 for $p_{\mathrm{T}}(\mathrm{true}) > 100$ GeV and $1.2 < |\eta| < 2.1$ (d) R_2 for $p_{\mathrm{T}}(\mathrm{true}) > 100$ GeV and $2.1 < |\eta| < 2.8$

Figure 8.3: R_2 in bins of fluctuating jet η for $p_{\mathrm{T}}(\mathrm{true}) > 100$ GeV measured in data, MC simulation and estimated using the jet smearing method with Gaussian and tail corrections and associated uncertainties accounted for. Multijet MC and estimates are normalised to the data. Statistical uncertainties are shown on the data. The yellow bands show the systematic uncertainties on the estimates.

Statistical uncertainties

The uncertainty from limited seed event statistics and pseudo-data statistics is propagated to all distributions and numbers produced using the jet smearing method. The procedure used is described in [59]. The uncertainty from limited pseudo-data tends to zero in the limit of infinite N_{smear}, the number of smeared events produced from each seed event. The number of seed events, however, relies on the number of events collected with the single jet triggers shown in Table 4.1 and the efficiency of the seed selection described in §6.4; it does not depend on N_{smear}. In a simplified case, if smeared data produced from only one seed event contributes to a signal region histogram, then the statistical uncertainty on the number of events present in the signal region should be $\sim 100\%$.

The derivation described in [59] yields a final overall statistical uncertainty on the j^{th} bin of an estimated distribution (normalised to 1) of:

$$\sigma_{\text{total}}(p_j) = \frac{1}{N_{\text{seed}} N_{\text{smear}}} \sqrt{\sum_i n_{ij}(1 + n_{ij})}. \tag{8.3}$$

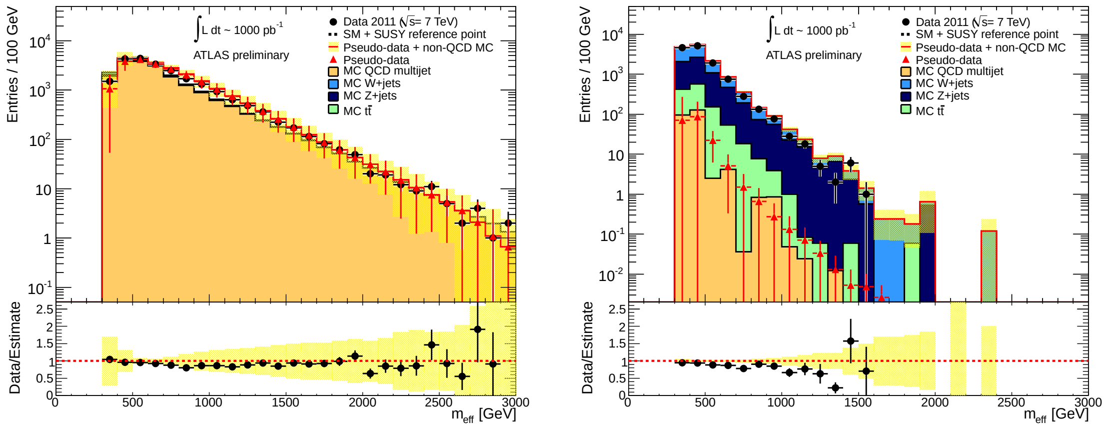
n_{ij} is the number of entries from the i^{th} seed event in the j^{th} bin. In the limit of infinite N_{smear}, Eq. 8.3 simplifies to:

$$\sigma_{\text{seed}}(p_j) = \frac{1}{N_{\text{seed}} N_{\text{smear}}} \sqrt{\sum_i n_{ij}^2}. \tag{8.4}$$

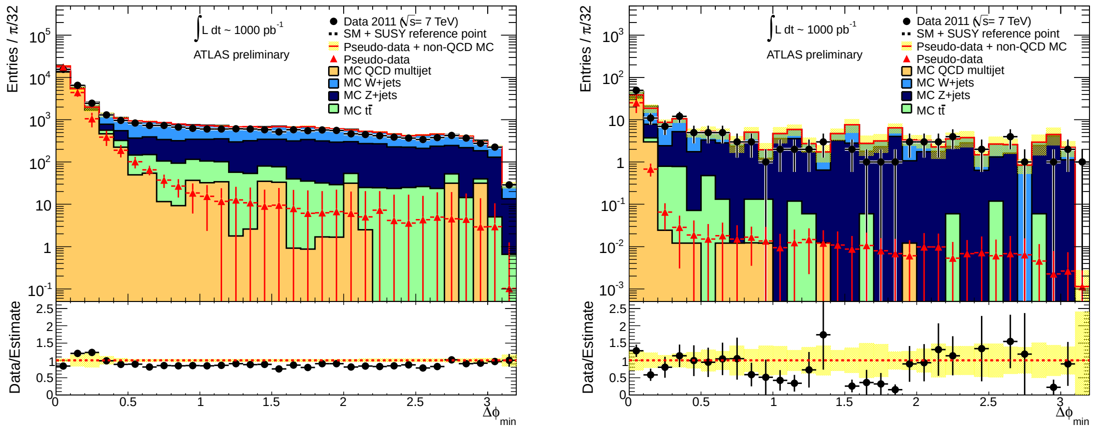
8.4 Results

Figures 8.4 to 8.8 show key distributions in SUSY signal and control regions in the data, MC and estimated with the jet smearing method. The definitions of the Signal Regions (SRs) are shown in Table 5.1. Statistical uncertainties are shown on the data and MC, and systematic and statistical uncertainties are shown on the jet smearing estimates (yellow bands).

Figure 8.9 shows the m_{eff} distributions for multijet ‘validation’ regions where the selection is the same as for CR2 (the multijet control region) except the final $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$ and m_{eff} cuts are not applied and the $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ requirement is $0.2 < \Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}}) < 0.4$.

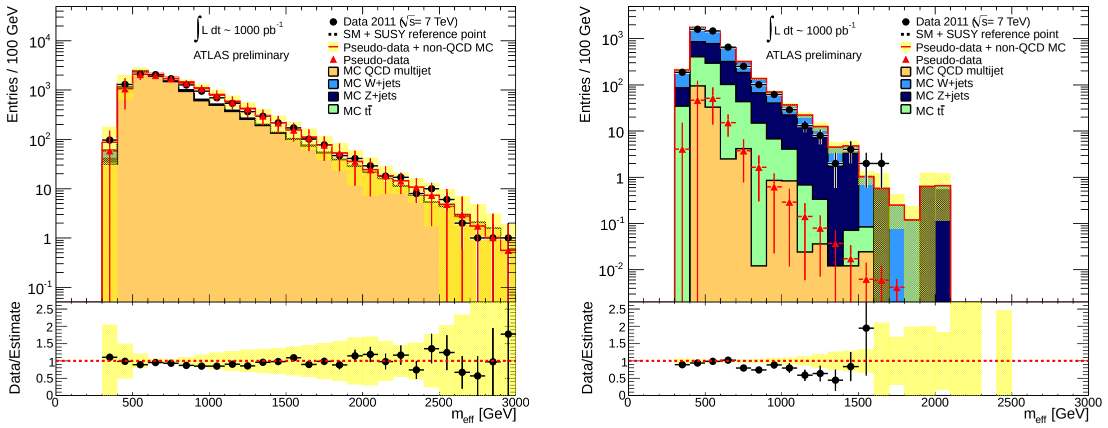


(a) $m_{\rm eff}$ distribution in CR2 without final $E_{\rm T}^{\rm miss}/m_{\rm eff}$ or $m_{\rm eff}$ cuts (b) $m_{\rm eff}$ distribution in SR without final $m_{\rm eff}$ cut

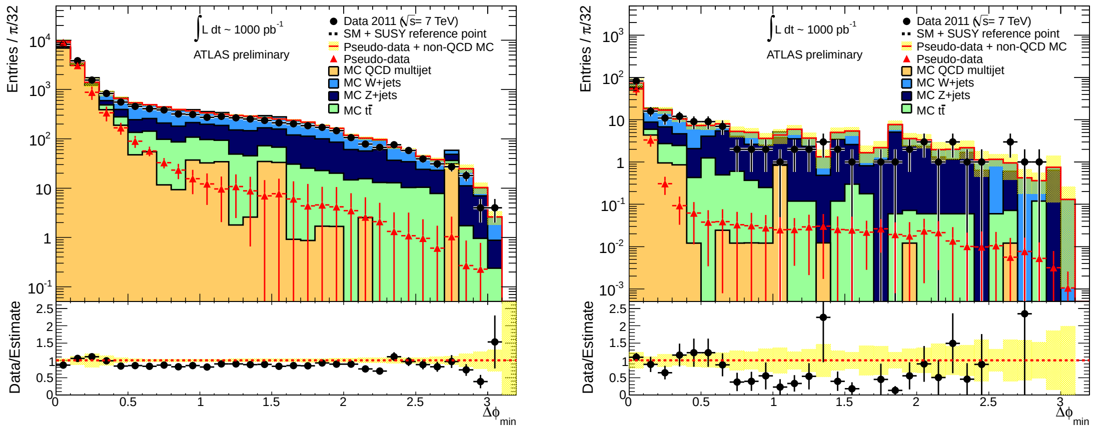


(c) $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$ distribution in CR2 without final $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$, $E_{\rm T}^{\rm miss}/m_{\rm eff}$ or $m_{\rm eff}$ cuts (d) $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$ distribution in SR without $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$ cut

Figure 8.4: $m_{\rm eff}$ and $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$ distributions in analysis A in data, MC and estimated with the jet smearing method.

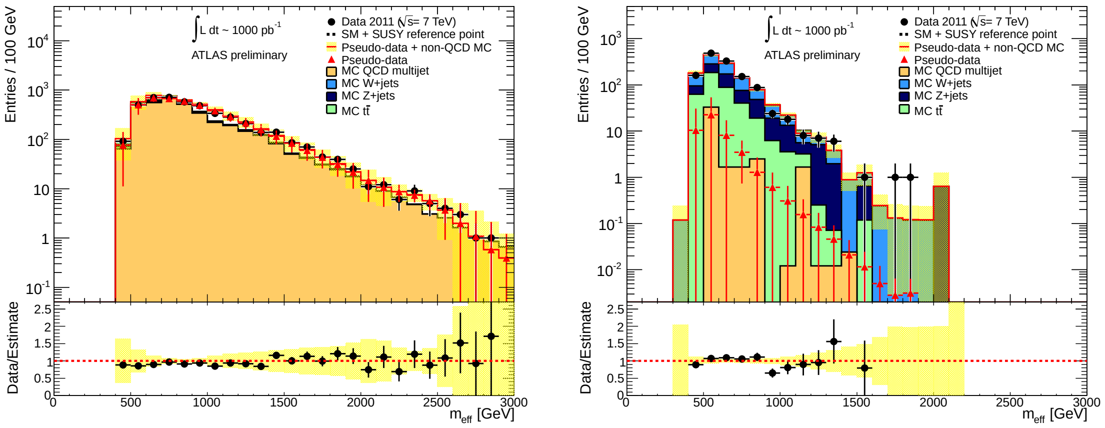


(a) m_{eff} distribution in CR2 without final $E_{\text{T}}^{\text{miss}}/m_{\text{eff}}$ or m_{eff} cuts (b) m_{eff} distribution in SR without final m_{eff} cut



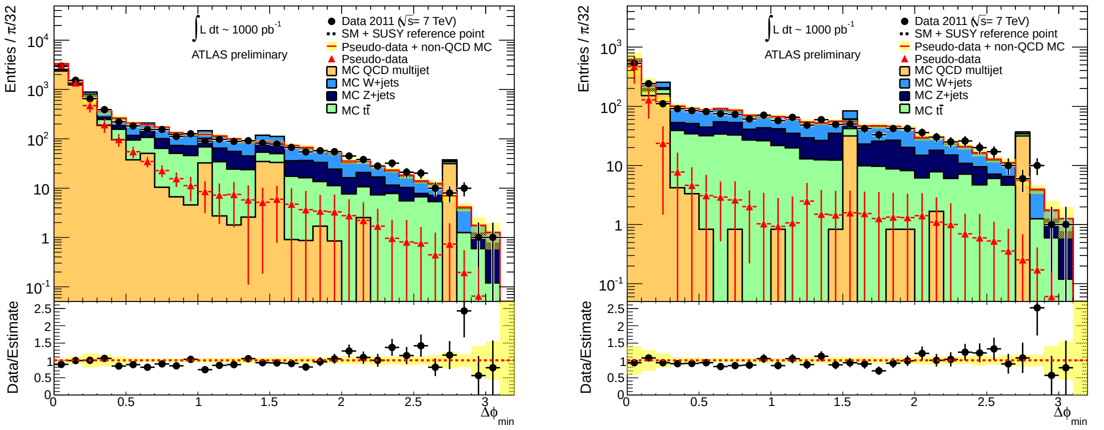
(c) $\Delta\phi(j_i, E_{\text{T}}^{\text{miss}})$ distribution in CR2 without final $\Delta\phi(j_i, E_{\text{T}}^{\text{miss}})$, $E_{\text{T}}^{\text{miss}}/m_{\text{eff}}$ or m_{eff} cuts (d) $\Delta\phi(j_i, E_{\text{T}}^{\text{miss}})$ distribution in SR without $\Delta\phi(j_i, E_{\text{T}}^{\text{miss}})$ cut

Figure 8.5: m_{eff} and $\Delta\phi(j_i, E_{\text{T}}^{\text{miss}})$ distributions in analysis B in data, MC and estimated with the jet smearing method.



(a) $m_{\rm eff}$ distribution in CR2 without final $E_{\rm T}^{\rm miss}/m_{\rm eff}$ or $m_{\rm eff}$ cuts

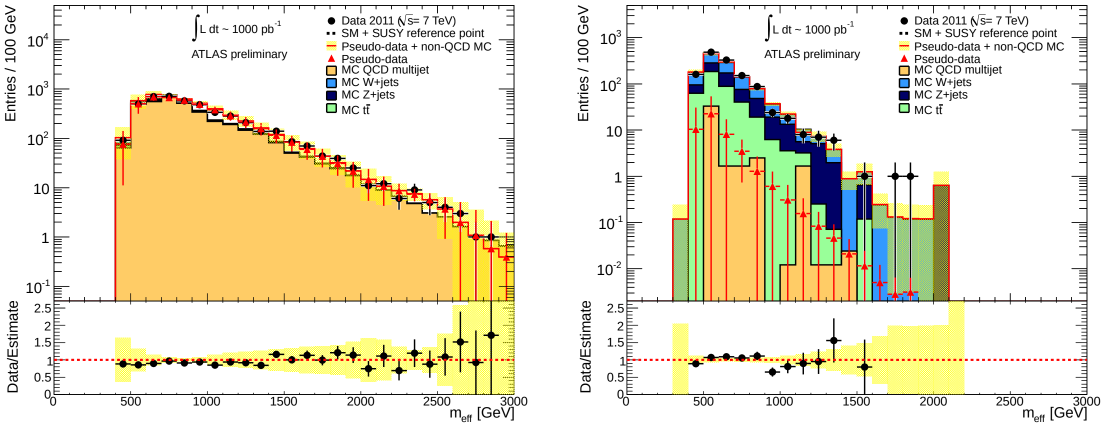
(b) $m_{\rm eff}$ distribution in SR without final $m_{\rm eff}$ cut



(c) $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$ distribution in CR2 without final $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$, $E_{\rm T}^{\rm miss}/m_{\rm eff}$ or $m_{\rm eff}$ cuts

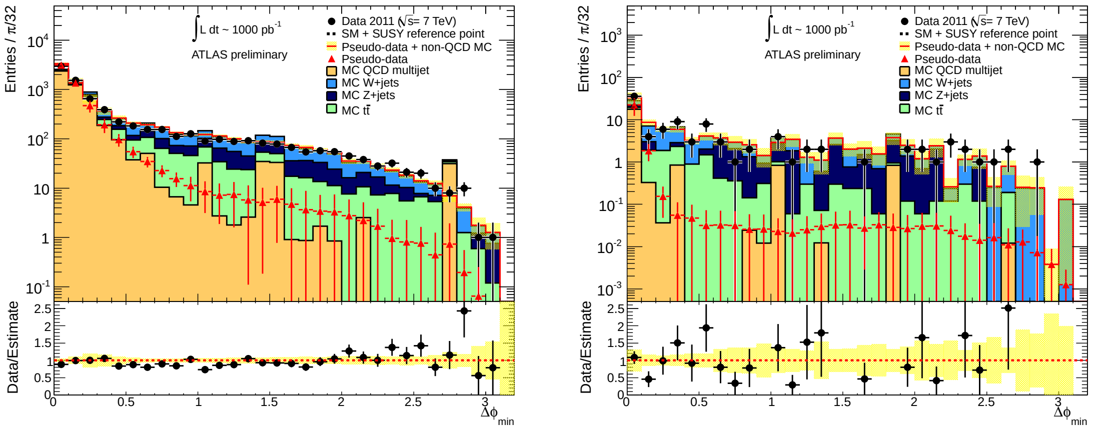
(d) $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$ distribution in SR without $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$ cut

Figure 8.6: $m_{\rm eff}$ and $\Delta\phi(j_i, E_{\rm T}^{\rm miss})$ distributions in analysis C in data, MC and estimated with the jet smearing method.



(a) m_{eff} distribution in CR2 without final $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$ or m_{eff} cuts

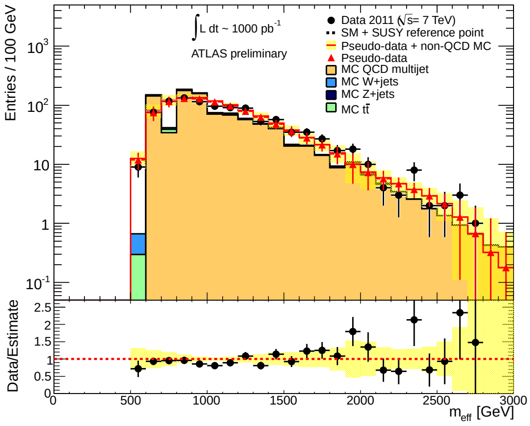
(b) m_{eff} distribution in SR without final m_{eff} cut



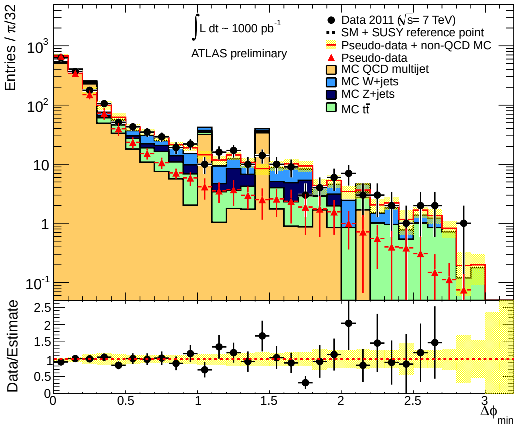
(c) $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ distribution in CR2 without final $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$, $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$ or m_{eff} cuts

(d) $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ distribution in SR without $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ cut

Figure 8.7: m_{eff} and $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ distributions in analysis D in data, MC and estimated with the jet smearing method.

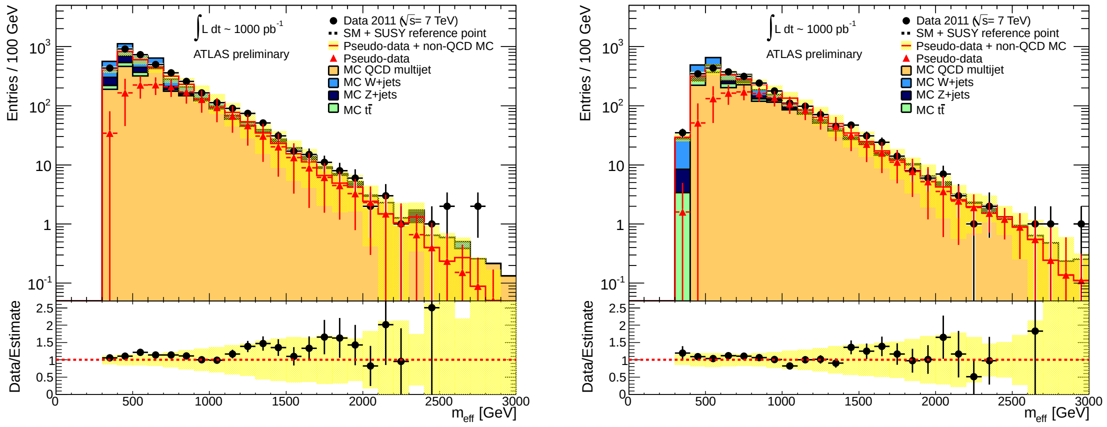


(a) m_{eff} distribution in CR2 without final $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$ or m_{eff} cuts (b) m_{eff} distribution in SR without final m_{eff} cut

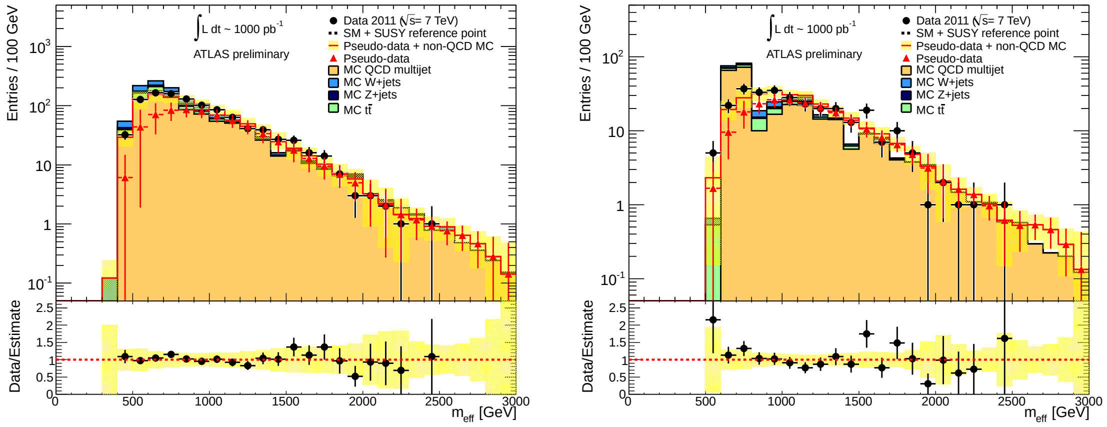


(c) $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ distribution in CR2 without final $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$, $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$ or m_{eff} cuts (d) $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ distribution in SR without $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ cut

Figure 8.8: m_{eff} and $\Delta\phi(j_i, E_{\mathrm{T}}^{\mathrm{miss}})$ distributions in analysis E in data, MC and estimated with the jet smearing method.



(a) Analysis A (b) Analysis B



(c) Analyses C and D (d) Analysis E

Figure 8.9: m_{eff} distributions in multijet validation regions in data, MC and estimated with the jet smearing method.

Table 8.1 shows the transfer factors determined for the different analyses using the jet smearing method. The transfer factors are re-evaluated using the extra constrained $\Delta\sigma_{tail}$ parameter from using the $\Delta\phi$ triggers as discussed in §7.4.4; this is shown in Table 8.2.

Signal Region	$N_{\mathrm{SR}}/N_{\mathrm{CR2}}$ $(\times 10^{-3})$
A	$10.3 \pm 0.6(\text{stat.})\ ^{+0.0}_{-1.7}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+6.7}_{-5.2}(\Delta\sigma_{tail})\ ^{+0.0}_{-5.0}(R_{mean-shift})$
B	$10.0 \pm 0.4(\text{stat.})\ ^{+0.0}_{-0.3}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+6.7}_{-4.5}(\Delta\sigma_{tail})\ ^{+0.0}_{-4.1}(R_{mean-shift})$
C	$62 \pm 3(\text{stat.})\ ^{+24}_{-0}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+63}_{-4}(\Delta\sigma_{tail})\ ^{+0}_{-11}(R_{mean-shift})$
D	$26 \pm 1(\text{stat.})\ ^{+0}_{-0}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+17}_{-13}(\Delta\sigma_{tail})\ ^{+0}_{-11}(R_{mean-shift})$
E	$42 \pm 2(\text{stat.})\ ^{+1}_{-0}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+22}_{-13}(\Delta\sigma_{tail})\ ^{+0}_{-17}(R_{mean-shift})$

Table 8.1: Predicted ratios of events between control and signal region for the five analysis channels using the jet smearing method.

Signal Region	$N_{\mathrm{SR}}/N_{\mathrm{CR2}}$ $(\times 10^{-3})$
A	$12.6 \pm 0.5(\text{stat.})\ ^{+0.0}_{-0.3}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+4.1}_{-7.3}(\Delta\sigma_{tail})\ ^{+0.0}_{-6.3}(R_{mean-shift})$
B	$11.9 \pm 0.3(\text{stat.})\ ^{+0.0}_{-0.1}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+4.0}_{-5.3}(\Delta\sigma_{tail})\ ^{+0.0}_{-5.0}(R_{mean-shift})$
C	$71 \pm 2(\text{stat.})\ ^{+15}_{-0}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+34}_{-11}(\Delta\sigma_{tail})\ ^{+0}_{-18}(R_{mean-shift})$
D	$31 \pm 1(\text{stat.})\ ^{+0}_{-1}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+9}_{-13}(\Delta\sigma_{tail})\ ^{+0}_{-13}(R_{mean-shift})$
E	$47 \pm 2(\text{stat.})\ ^{+0}_{-0}(\sigma_{correction}(p_{\mathrm{T}}))\ ^{+12}_{-14}(\Delta\sigma_{tail})\ ^{+0}_{-19}(R_{mean-shift})$

Table 8.2: Predicted ratios of events between control and signal region for the five analysis channels using the jet smearing method with the $\Delta\sigma_{tail}$ further constrained through the use of $\Delta\phi$ triggers.

8.5 Discussion

The CR2 distributions in Figures 8.4 to 8.8 and the validation plots in Figure 8.9 show good agreement between the data and the smearing estimates (with non-multijet MC accounted for). In the SR distributions in Figures 8.4 to 8.8, the benefit in statistics from using the jet smearing estimate compared to the multijet MC is clear. The transfer factors shown in Table 8.2, which are estimated using the extra-constrained $\Delta\sigma_{tail}$ parameter (from the use of the $\Delta\phi$ triggers) are seen to have smaller overall fractional systematic uncertainties associated with them than those shown in Table 8.1. Therefore the use of the $\Delta\phi$ triggers allows a more accurate estimation of the multijet background.

Chapter 9

Conclusions

A data-driven technique to estimate the multijet background to SUSY searches with no leptons at ATLAS has been described. Versions of this method were used for the multijet estimations in a number of SUSY searches [3, 4, 5], allowing a greater discovery reach than would have been possible with MC methods. As part of the method, jet response at ATLAS was measured, including the non-Gaussian tails. Novel event shape triggers were developed which provided more statistics for the response tail measurement thus increasing the sensitivity of the final multijet estimation.

Chapter 2 gave an overview of the SM and SUSY. SUSY is seen to provide a solution to the hierarchy problem. The status of SUSY searches was described with no evidence yet found for its existence. Finally, techniques for reducing and estimating the multijet background to SUSY searches with no leptons at hadron colliders were overviewed.

In Chapter 3, the LHC and ATLAS detector were described. It was seen that the calorimetry of ATLAS provides the ability to make reliable measurements of jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ which are key in fully hadronic SUSY searches.

Chapter 4 described the analysis tools used in this thesis including details of the Monte Carlo simulated multijet samples used for the initial measurement of jet response. Object reconstruction, event cleaning and details of triggers used were also given.

Chapter 5 gave an overview of SUSY searches using ATLAS. The fully hadronic search was described and it was seen to provide exclusion limits in searches for squarks and gluinos beyond those found at the Tevatron and previous searches using the ATLAS detector.

Chapter 6 introduced the jet smearing method for multijet background estimation. An introduction to jet response was given, alongside an overview of jet smearing and the assumptions which the technique relies on. The selection of well measured data seed events was also described.

In Chapter 7, the measurement of jet response using the jet smearing method was described. This included the evaluation of the non-Gaussian jet response tail. The technique was shown to adequately describe multijet dominated control distributions. Novel event shape $\Delta\phi$ triggers were then introduced and they were shown to provide extra statistics for the measurement of non-Gaussian response.

Chapter 8 described the use of jet smearing, using the data-measured response functions, to provide the multijet estimation for the fully hadronic SUSY searches. Multijet distributions were estimated and yielded excellent agreement with the data observations. Transfer functions were produced between SUSY multijet control distributions and signal regions. The use of $\Delta\phi$ triggers was seen to reduce the overall systematic uncertainty on these transfer functions.

The data-driven jet smearing technique, described in Chapters 6 to 8, is the primary method used by the ATLAS collaboration to measure the multijet background to SUSY searches with no leptons. As such it is a key part of these searches, and has resulted in the collaboration setting exclusion limits beyond that which would have been possible with MC methods. Any future SUSY discovery will need to demonstrate that the multijet background is well understood and under control, the analysis presented in this thesis will allow this.

Future work which could reduce the systematic uncertainties associated with the method could include using dijet events to further constrain the non-Gaussian response measurement and improving the seed selection to limit the bias introduced to the leading jet p_{T} distribution.

Bibliography

[1] ATLAS Collaboration, *Performance of the ATLAS Jet Trigger in the Early sqrt{s}=7 TeV Data*, ATLAS-CONF-2010-094, 2010.

[2] D. Tovey, *On measuring the masses of pair-produced semi-invisibly decaying particles at hadron colliders*, JHEP **04** (2008) 034, `arXiv:0802.2879 [hep-ph]`.

[3] ATLAS Collaboration, *Search for squarks and gluinos using final states with jets and missing transverse momentum with the ATLAS detector in sqrt(s) = 7 TeV proton-proton collisions*, Phys. Lett. **B701** (2011) 186–203.

[4] ATLAS Collaboration, *Search for squarks and gluinos using final states with jets and missing transverse momentum with the ATLAS detector in* $\sqrt{s} = 7$ *TeV proton-proton collisions*, ATLAS-CONF-2011-086, 2011.

[5] ATLAS Collaboration, *Search for squarks and gluinos using final states with jets and missing transverse momentum with the ATLAS detector in* $\sqrt{s} = 7$ *TeV proton-proton collisions*, 2011. `arXiv:1109.6572v1 [hep-ex]`. Submitted to Phys. Lett. B.

[6] ATLAS Collaboration, *Search for supersymmetry in pp collisions at* $\sqrt{s} = 7$ *TeV in final states with missing transverse momentum, b-jets and no leptons with the ATLAS detector*, ATLAS-CONF-2011-098, 2011.

[7] ATLAS Collaboration, *Search for scalar bottom pair production with the ATLAS detector in pp Collisions at sqrt{s} = 7 TeV*, 2011. `arXiv:1112.3832v1 [hep-ex]`. Submitted to Phys. Rev. Lett.

[8] ATLAS Collaboration, *Search for supersymmetry in pp collisions at* $\sqrt{s} = 7$ *TeV in final states with missing transverse momentum and b-jets with the ATLAS detector*, ATLAS-CONF-2012-003, 2012.

[9] ATLAS Collaboration, *Search for new phenomena in final states with large jet multiplicities and missing transverse momentum using sqrt(s)=7 TeV pp collisions with the ATLAS detector*, 2011. JHEP **11** (2011) 099.

[10] S. Owen, *Data-driven estimation of the QCD multijet background to SUSY searches with jets and missing transverse momentum at ATLAS using jet smearing*, ATL-PHYS-INT-2012-008, 2012. Internally approved note, only available to members of the ATLAS collaboration.

[11] F. Halzen and A. D. Martin, *Quarks and Leptons.* Wiley, 1984.

[12] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.

[13] The ALEPH, CDF, DO, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group and the SLD electroweak and heavy flavour groups, *Precision Electoweak Measurements and Constraints on the Standard Model*, 2011. `arXiv:1012.2367v2 [hep-ex]`.

[14] ATLAS Collaboration, *Combined search for the Standard Model Higgs boson using up to 4.9 fb-1 of pp collision data at sqrt(s) = 7 TeV with the ATLAS detector at the LHC*, 2012. `arXiv:1202.1408v1 [hep-ex]`.

[15] A. H. Chamseddine, R. L. Arnowitt and P. Nath, *Locally Supersymmetric Grand Unification*, Phys. Rev. Lett. **49** (1982) 970–974.

[16] R. Barbieri, S. Ferrara and C. A. Savoy, *Gauge Models with Spontaneously Broken Local Supersymmetry*, Phys. Lett. **B119** (1982) 343.

[17] L. E. Ibanez, , Phys. Lett. **B118** (1982) 73.

[18] L. J. Hall, J. D. Lykken and S. Weinberg, *Supergravity as the messenger of supersymmetry breaking*, Phys. Rev. **D27** (1983) 2359–2378.

[19] N. Ohta, *Grand Unified Theories Based on Local Supersymmetry*, Prog. Theor. Phys. **70** (1983) 542–549.

[20] Johan Alwall, Philip Schuster, Natalis Toro, *Simplified Models for a First Characterization of New Physics at the LHC*, Phys. Rev. D **79** (2009) 075020.

[21] LEP SUSY Working Group (ALEPH, DELPHI, L3, OPAL), LEPSUSYWG/04-02.1 note, http://lepsusy.web.cern.ch/lepsusy.

[22] CDF Collaboration, T. Aaltonen *et al.*, *Inclusive Search for Squark and Gluino Production in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 TeV*, Phys. Rev. Lett. **102** (2009) 121801.

[23] DØ Collaboration, *Search for squarks and gluinos in events with jets and missing transverse energy using 2.1 fb^{-1} of $p\bar{p}$ collision data at $\sqrt{s}$ = 1.96 TeV*, Phys. Lett. **B660** (2008) 449–457.

[24] ATLAS Collaboration, *Search for supersymmetry in final states with jets, missing transverse momentum and one isolated lepton in sqrt{s} = 7 TeV pp collisions using 1 fb^{-1} of ATLAS data*, Phys. Rev. **D85** (2012) 012006.

[25] ATLAS Collaboration, *Searches for supersymmetry with the ATLAS detector using final states with two leptons and missing transverse momentum in sqrt{s} = 7 TeV proton-proton collisions*, 2011. `arXiv:1110.6189v1 [hep-ex]`. Submitted to Phys. Lett. B.

[26] ATLAS Collaboration, *Search for supersymmetry in events with four or more leptons and missing transverse momentum in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, ATLAS-CONF-2012-001, 2012.

[27] CMS Collaboration, *Search for Supersymmetry at the LHC in Events with Jets and Missing Transverse Energy*, Phys. Rev. Lett. **107** (2011) 221804.

[28] CMS Collaboration, *Search for supersymmetry in hadronic final states using* M_{T2} *in 7 TeV pp collisions at the LHC*, CMS PAS SUS-11-005, August, 2011.

[29] CMS Collaboration, *Search for supersymmetry with the razor variables*, CMS PAS SUS-11-008, December, 2011.

[30] CMS Collaboration, *Search for supersymmetry in all-hadronic events with missing energy*, CMS PAS SUS-11-004, August, 2011.

[31] Figure from CMS Supersymmetry Physics Results, https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS.

[32] M. Cacciari, G.P. Salam and G. Soyez, *The anti-*k_t *jet clustering algorithm*, JHEP **04** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[33] Figure from Martin-Stirling-Thorne-Watt Parton Distribution Functions, http://projects.hepforge.org/mstwpdf/plots/plots.html.

[34] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[35] Picture from CERN public webpages, http://public.web.cern.ch/public/en/LHC/Facts-en.html.

[36] CERN Press Office, *LHC sets world record beam intensity*, http://press.web.cern.ch/press/PressReleases/Releases2011/PR02.11E.html, April, 2011.

[37] CMS Collaboration, *Combined results of searches for the standard model Higgs boson in pp collisions at sqrt(s) = 7 TeV*, 2012. `arXiv:1202.1488v1 [hep-ex]`.

[38] CERN Press Office, *LHC to run at 4 TeV per beam in 2012*, http://press.web.cern.ch/press/PressReleases/Releases2012/PR01.12E.html, February, 2012.

[39] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, 2008. JINST **3** (2008) S08003.

[40] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[41] A. Sherstnev and R.S. Thorne, *Parton Distributions for LO Generators*, Eur. Phys. J. **C55** (2008) 553–575, `arXiv:0711.2473 [hep-ph]`.

[42] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. **D78** (2008) 013004, `arXiv:0802.0007 [hep-ph]`.

[43] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026, `arXiv:hep-ph/0603175`.

[44] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06** (2002) 029, `arXiv:hep-ph/0204244`.

[45] S. Frixione, P. Nason, and B. R. Webber, *Matching NLO QCD and parton showers in heavy flavour production*, JHEP **08** (2003) 007, `arXiv:hep-ph/0305252`.

[46] G. Corcella et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, JHEP **01** (2001) 010, `arXiv:hep-ph/0011363`.

[47] G. Corcella et al., *HERWIG 6.5 release note*, 2002. `arXiv:hep-ph/0210213`.

[48] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. **C72** (1996) 637–646, `arXiv:hep-ph/9601371`.

[49] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, `arXiv:hep-ph/0206293`.

[50] ATLAS Collaboration, *First tuning of HERWIG/JIMMY to ATLAS data*, 2010. ATL-PHYS-PUB-2010-014.

[51] ATLAS Collaboration, *Charged particle multiplicities in pp interactions at* $\sqrt{s} = 0.9$ *and 7 TeV in a diffractive limited phase-space measured with the*

ATLAS detector at the LHC and new PYTHIA6 tune, 2010. ATLAS-CONF-2010-031.

[52] ATLAS Collaboration, *ATLAS Computing Technical Design Report*, 2005. CERN-LHCC-2005-022.

[53] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, Tech. Rep. ATL-LARG-PUB-2008-002, CERN, April, 2008.

[54] ATLAS Collaboration, *Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at* $\sqrt{s}$ *=7 TeV and measured with the ATLAS detector*, ATLAS-CONF-2010-056, July, 2010.

[55] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102, 2011.

[56] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, 2009. `arXiv:0901.0512 [hep-ex]`.

[57] ATLAS Experiment Public Results: Public Jet Trigger Plots for Collision Data, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetTriggerPublicResults.

[58] A. Read, *Presentation of search results: the CLs technique*, Journal of Physics G: Nucl. Part. Phys. **28** (2002) 2693–2704.

[59] R. Duxfield, *Background Control in Inclusive Supersymmetry Searches and Offline Reconstruction at ATLAS.* PhD thesis, The University of Sheffield, 2008.

[60] ATLAS Collaboration, *Search for scalar bottom pair production with the ATLAS detector in pp Collisions at sqrt{s} = 7 TeV*, ATL-PHYS-INT-2012-002, 2011. Internally approved note, only available to members of the ATLAS collaboration.

[61] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics*, CERN-OPEN-2008-020, page 1578-1582, 2008.

[62] J. Gaiser, *Appendix-F Charmonium Spectroscopy from Radiative Decays of the* J/Ψ *and* Ψ'. PhD thesis, Stanford University, 1982. SLAC-R-255.