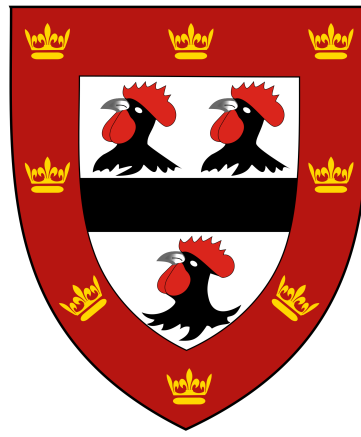
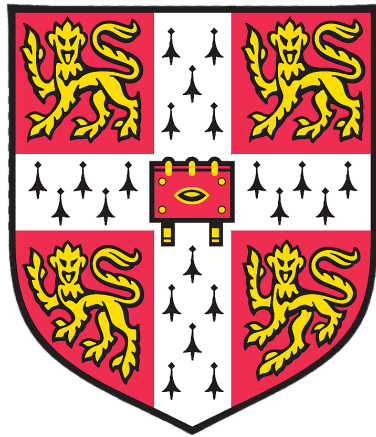


Extending the Reach of Searches for Staus, Charginos and Neutralinos with the ATLAS Experiment at the Large Hadron Collider

Dominic Matthew Jones
of Jesus College,
University of Cambridge



This thesis is submitted to the University of Cambridge
for the degree of Doctor of Philosophy
September 2023

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and stated in the text. I further state that no substantial part of this thesis has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and stated in the text. It does not exceed the prescribed word limit for the Physics and Chemistry Degree Committee.

Dominic Matthew Jones

Abstract

Extending the Reach of Searches for Staus, Charginos and Neutralinos with the ATLAS Experiment at the Large Hadron Collider

Dominic Matthew Jones

This thesis describes new approaches to extend the sensitivity reach of searches for three types of supersymmetric particles; staus, charginos and neutralinos, using the 139 fb^{-1} of data collected with the ATLAS experiment at the Large Hadron Collider between 2015 and 2018. Searches for three different production modes of these particles are considered, with new approaches for each that highlight different methods for improving sensitivity.

Firstly, a search for direct stau production, in which staus decay to tau-leptons and neutralinos, is presented. The search makes use of Boosted Decision Trees to classify the signal from the Standard Model backgrounds and in doing so significantly improves the sensitivity compared to a previous search using the same dataset. No significant differences between the observed data and the Standard Model predictions are found but new, world-leading limits on direct stau production are set. Models with mass degenerate left and right-handed staus are excluded at 95% confidence level for stau masses between 80 GeV and 500 GeV, with the low mass reach closing a previous gap in sensitivity between previous searches performed by the ATLAS experiment and searches conducted by experiments at the Large Electron Positron collider. This search also provides the first sensitivity at the

ATLAS experiment to the right-handed stau only interpretation, for which models with right-handed stau masses between 100 GeV and 350 GeV are excluded for a massless neutralino-one at 95% confidence level.

Secondly, new search channels for the production of charginos and neutralinos with highly compressed mass spectra are explored. In particular, a re-interpretation of a previous search for new physics performed at ATLAS in a final state with an energetic jet and large missing transverse momentum is compared to a sensitivity study of a vector boson fusion channel. The re-interpretation is found to have sensitivity to previously uncovered regions of parameter space; for a simplified wino/bino interpretation models with neutralino-two masses of 140 GeV are excluded at 95% confidence level, under the assumption that signal theoretical uncertainties have a negligible impact. The vector boson fusion channel is assessed to have comparatively limited sensitivity with differences in the signal modelling between this study and previous work on this channel being identified. However, it is shown that the use of machine learning in this channel can provide a significant gain in the sensitivity.

Thirdly, a statistical combination of previous ATLAS searches for chargino pair production in three different topologies is performed for the first time, leading to a notable enhancement in the overall sensitivity.

Acknowledgements

There are many people to whom I owe thanks for helping in the completion of this work. First and foremost I would like to thank my supervisor Tina Potter for her continuous support, guidance and teaching throughout this PhD. Thank you for making this experience an overall enjoyable one - I will miss our weekly chats! Thanks as well to Bart Hommels for supervising my ITk Qualification Task - I hope production goes well! Gratitude is also owed to the Science and Technology Facilities Council for funding this PhD and to Jesus College for providing conference funding.

In addition, I would like to thank those with whom I was fortunate enough to share my PhD journey and those within ATLAS I collaborated with. At Cambridge, I would like to thank fellow PhD students Ben Hodkinson, David Rouso, Fionn Bishop and Lars Henkelmann for their companionship throughout the good and the not so good times and without whom proceedings would have been an order of magnitude less fun. Further, I feel privileged to have been a part of the Cambridge High Energy Physics group over the past four years and owe thanks to all of its members for their help and for providing many interesting discussions. I would especially like to thank Will Fawcett for his invaluable technical help during the early stages of this PhD, Dan Noel and Rupert Tombs for hosting the machine learning discussion group and Jon Burr, Jesse Liu, and Sarah Williams for their support and guidance.

From my LTA experience in Switzerland, I would like to thank all those who had a hand in making it the unforgettable experience that it was, of whom there are too many to mention. It was a truly amazing time.

Within ATLAS, I would like to thank the members of the Electroweak SUSY analysis teams I was part of. From the Compressed team, I'd like to in particular thank Jeff Shahinian and Michael Holzbock for their roles as analysis contacts, Nathan Kang and Mingjie Zhai for their work on the VBF channel and Mike Hance for his perseverance in getting to the bottom of the VBF cross-sections. From the Di-Tau team, I owe thanks to Stefan Guindon and Xuai Zhuang for their work as analysis contacts, Daniela

Köck and Yuchen Cai for their technical support and assistance and again to Tina for putting the paper together. I'd also like to again thank Jeff Shahinian alongside Sara Alderweireldt, Ben Hooberman and Javier Berlingen for their feedback as convenors on the Direct Stau analysis and for helping us to make the LHCP deadlines. Lastly from the Combinations team, I'd like to once more thank Tina along with Alberto and the rest of the combinations team involved in laying the significant groundwork without which the combinations would not be possible. Thanks must also be paid to the ATLAS collaboration and CERN at large for the construction and operation of the detector and of the LHC in general.

The sanity I believe I have left after 8 years of study at Cambridge is in large part thanks to the friendship of Tudor Evans, Jamie Dougherty, Alex Barry and Ben Schreiber, so for that I also express profound acknowledgement. Apologies I didn't put a BrassEye quote in the main body, I hope this is sufficient: "TECHNOLOGY!".

Last, but by no means least, I would like to thank my family, in particular my parents, brother Adam and grandparents for their love and support over the years.

Preface

The work documented in this thesis was conducted while the author was a member of the ATLAS collaboration, the group responsible for the operation of the ATLAS experiment at the Large Hadron Collider located at CERN, Switzerland. The results presented here rely upon much previous work and effort from other members of the collaboration and of CERN in general, not least for the construction and operation of the Large Hadron Collider and the ATLAS detector. Further, the analyses in this thesis were done as part of analysis teams formed within the ATLAS Supersymmetry working group which are highlighted below. This thesis focuses on the contributions made by the author to these projects; work done by collaborators is explicitly noted. An overview of the authors contributions is provided in the following. A more detailed account can be found in Section A.1 of the appendix.

The search for direct stau production presented in Chapter 4 was done in collaboration with the ATLAS Electroweak Supersymmetry Di-Tau analysis team. The author was the primary analyzer for the direct stau search channel from March 2022-July 2023 and was responsible for many aspects of the analysis, including; designing and implementing the Boosted Decision Tree based event selection; adapting existing background estimation methods for this selection; evaluating systematic uncertainties; setting up and running the statistical likelihood fits and making the final result plots. Results of the direct stau search were presented by the author at the LHCP 2023 conference and have since been made public in a paper published by the ATLAS collaboration [1].

The studies of new search channels for charginos and neutralinos with small mass differences presented in Chapter 5 were performed by the author as part of the ATLAS Compressed Electroweak Supersymmetry analysis team. The author was one of the main analyzers of the vector boson fusion channel since its inception in April 2020 with responsibilities that included producing the ntuples used in the analysis; comparing the modelling of the dominant V +jets background by different Monte Carlo event

generators and evaluating the potential sensitivity gain from using a machine learning based event selection. Additionally, the author had an important role in identifying the differences between the studies performed within ATLAS and a CMS analysis which inspired the study of the vector boson fusion channel. The re-interpretation of the mono-jet analysis was instigated by the author and served as their main focus, alongside the search for direct stau production mentioned above, from January 2022. The work done by the author in both the vector boson fusion and mono-jet channels is intended to form part of a future ATLAS publication.

The statistical combination of searches for chargino pair production shown in Chapter 6 was performed predominantly between April 2021 - July 2022 as part of the ATLAS Electroweak Supersymmetry Combinations analysis team. The work on the chargino pair production combination was shared equally between the author and Ben Hodkinson, a fellow PhD student at Cambridge. The results of the chargino pair production combination have been published by the ATLAS collaboration alongside other combinations of searches for electroweak Supersymmetry [2].

In addition to the work presented in this thesis, the author also contributed to other projects as a member of the ATLAS collaboration during their PhD. In particular, the author completed their Qualification Task on the ATLAS Inner Tracker (ITk) upgrade, developing software for Quality Control tests of strip sensors and updating an automated wire-bonding test for barrel modules. This work provided essential automation of these tasks and has already been used extensively by the ATLAS ITk community. The work on the Quality Control tests was completed with David Rousso, another fellow PhD student at Cambridge, and was presented at the 15th Pisa Meeting on Advanced Detectors [3] and PIXEL 2022 conferences [4]. The author also performed a sensitivity study of a photon initial state radiation search channel for charginos and neutralinos with small mass differences (the subject of Chapter 5) which is not discussed in this thesis. Lastly, between August 2022 - August 2023 the author was based at CERN and performed trigger and run control shifts for the ATLAS experiment during Run 3 data taking periods.

Contents

Introduction	1
1 The Standard Model and Supersymmetry	5
1.1 The Standard Model	7
1.1.1 Particles and Interactions	7
1.1.2 Structure and Symmetries	14
1.2 Shortcomings and Open Questions	18
1.2.1 Particle-scale Observations	18
1.2.2 Astrophysical and Cosmological Observations	19
1.2.3 Suggestions of a more Fundamental Theory	21
1.3 Supersymmetry	24
1.3.1 Key Aspects	24
1.3.2 Motivations	27
1.3.3 Searching for Supersymmetric Particles	29
2 The ATLAS Experiment at the Large Hadron Collider	33
2.1 The Large Hadron Collider	34
2.1.1 Overview	34
2.1.2 The Nature of Proton-Proton Collisions	35
2.1.3 Luminosity	37
2.2 The ATLAS Experiment	39
2.2.1 Overview	39
2.2.2 Sub-Detectors	41
2.2.3 Trigger System	48
2.2.4 Object and Event Reconstruction	50
2.2.5 Simulation	57

Contents

3	Analysis Methods	59
3.1	General Setup of a Search for New Physics	60
3.1.1	Ingredients	60
3.1.2	Likelihood Construction	63
3.1.3	Fitting	64
3.1.4	Hypothesis Testing	65
3.1.5	CL _s values	67
3.1.6	Fit Types	68
3.1.7	Control Regions, Shape fits and Combinations	69
3.1.8	High Level Kinematic Variables	71
3.2	Enhancing the Sensitivity of Searches using Machine Learning	73
3.2.1	General Considerations	73
3.2.2	Boosted Decision Trees	82
4	Search for Direct Stau Production	89
4.1	Introduction	90
4.2	Overview of Analysis Strategy	94
4.3	Data and Simulated Event Samples	95
4.4	Event Reconstruction	97
4.5	Event Selection	100
4.5.1	Triggers and Baseline Selection	100
4.5.2	Machine Learning Details	102
4.5.3	Signal Region Optimization	118
4.6	Background Estimation	122
4.6.1	Multi-jet Estimation	129
4.6.2	Z+jets Estimation	137
4.6.3	W+jets Estimation	140
4.6.4	Top Estimation	144
4.6.5	Multi-boson Estimation	147
4.6.6	Inclusive Validation Region	149
4.7	Systematic Uncertainties	151
4.7.1	Experimental Uncertainties	151
4.7.2	Theoretical Uncertainties	152
4.7.3	Uncertainties on the Multi-jet background	154
4.7.4	Impact of Systematic Uncertainties	154

Contents

4.8	Results	158
4.8.1	Unblinded Signal Regions	158
4.8.2	Model Independent Upper Limits	171
4.8.3	Model Dependent Upper Limits	172
4.9	Conclusion	179
5	Searches for Ultra-compressed Charginos and Neutralinos	181
5.1	Overview of Ultra-compressed Electroweak SUSY Models and Searches	182
5.2	Re-interpretation of the ATLAS Mono-jet search	188
5.2.1	Overview of ATLAS Mono-jet search	188
5.2.2	Signal Generation and Re-interpretation setup	189
5.2.3	Results	193
5.3	Sensitivity Study of Vector Boson Fusion Channel	198
5.3.1	Overview of Analysis Strategy	198
5.3.2	Simulation and Data Samples	199
5.3.3	Event Reconstruction	203
5.3.4	Event selection and Signal Region Optimization	204
5.3.5	Background Estimation	222
5.3.6	Evaluation of Sensitivity	230
5.4	Comparison and Conclusions	240
6	Combination of Searches for Chargino Pair Production	243
6.1	Introduction	244
6.2	Combination Details	247
6.3	Systematic Uncertainties	252
6.4	Results	258
6.5	Conclusion	259
	Conclusion	261
	Appendix	263
A.1	Breakdown of Authors Contributions	263
A.1.1	Search for Direct Stau Production	263
A.1.2	Searches for Ultra-compressed Charginos and Neutralinos . . .	264
A.1.3	Combination of Searches for Chargino pair-production	265
A.2	Additional details for the Direct Stau Search	265
A.2.1	Signal and Background Shape Comparisons	265

A.2.2	Additional BDT Diagnostics	269
A.2.3	Comparison of Tight and Medium Tau-ID Working Points	273
A.2.4	Simplified Multi-jet Estimation Comparison	276
A.2.5	Cut Flow Tables	278
A.2.6	Cross-section Upper Limits	280
A.3	Additional details for the Vector Boson Fusion Sensitivity Study	281
A.3.1	Signal vs Background Shape plots	281
A.3.2	Additional BDT details	281
A.3.3	Cut Flow Tables	287
A.4	Cross-section Upper Limits for Chargino Pair-production Combination	289
Bibliography		293
List of figures		309
List of tables		315

Introduction

Particle physics is the study of the fundamental building blocks of the universe; what they are, their properties and how they interact with each other. Humanity's current best understanding of particle physics is encoded in a theoretical framework known as the *Standard Model* (SM). Since its inception, the SM has been extremely successful; it has been used to make many predictions which agree with results obtained from numerous experiments. However, there are observed phenomena, that a complete theory of particle physics perhaps ought to be able to explain, which the current formulation of the SM cannot account for. Together with theoretical arguments, they suggest the SM is incomplete and thus provide motivations for possible extensions to the SM. These extensions in general predict the existence of new particles.

Supersymmetry (SUSY) is one such extension to the SM and predicts the existence of many new particles, referred to collectively as super-particles or *sparticles*. Despite having strong theoretical motivations, SUSY does not make precise predictions on the properties of sparticles such as their masses or decay branching fractions. Dependent upon their properties, certain sparticles could be produced in detectable amounts at particle colliders. Being, at time of writing, the largest and highest energy particle collider in the world, the Large Hadron Collider (LHC) located at CERN, Geneva, provides the best prospects for producing sparticles. Consequently, searches for sparticles are performed by experiments at the LHC, with the aim of finding evidence for their existence. These searches involve looking for discrepancies, typically excesses, in the number of particle collisions recorded by the experiments that satisfy particular criteria compared to the number that would be expected from the SM alone. The criteria define *search channels* which target specific production and decay modes of sparticles. The particle collisions are referred to as *events*. In the absence of observing discrepancies in search channels, constraints can be placed on the properties of the sparticles being searched for and on parameters of SUSY in general.

Many searches for sparticles have already been conducted by experiments at the LHC, and indeed at previous particle colliders. So far no significant evidence for the existence of any sparticles has been found. However, there remains significant amounts of unexplored Supersymmetric parameter space. Given this, and the significant effort and expense that goes into the construction, design and running of particle accelerators and detectors, it is important that every effort is made to extend the sensitivity reach of sparticle searches as far as is reasonably possible. Further, so far the particle collision data used for these searches corresponds to only $\sim 5\%$ of the total dataset that aims to be recorded by the experiments at the LHC, which is planned to continue operating past 2040. Thus, now is an opportune time to devise and test strategies that can enhance the sensitivity of these searches.

In this thesis, strategies for extending the reach of searches at the ATLAS experiment, one of the experiments at the LHC, for three different sparticle production/decay scenarios are studied. All of these scenarios involve staus, charginos or neutralinos being produced via the electroweak interaction. Such production modes are amongst the most challenging to search for at a hadron collider due to their relatively small production rates. However, there are reasons to suggest that these scenarios may be amongst the first or possibly the only production/decay modes to be accessible at the LHC, including the provision of a dark matter candidate compatible with cosmological models and theoretical arguments based on the concept of naturalness.

The first scenario studied concerns the direct production of staus decaying to neutralinos. A new search for this scenario is detailed which uses machine learning to optimise the selection of events and by doing so significantly improves the sensitivity reach compared to a previous search performed with the same dataset. The second scenario is the production of charginos and neutralinos which are very close in mass to each other, with mass differences less than 1 GeV. The charginos and neutralinos in this scenario are referred to here as having an *ultra-compressed* mass spectrum. New search channels for this scenario are explored to assess their sensitivity. Specifically, a re-interpretation of a previous ATLAS search is used to determine the sensitivity of the mono-jet search channel, followed by a dedicated sensitivity study of a vector boson fusion search channel. The third scenario is the pair production of charginos decaying to W -bosons and neutralinos. Results of previous ATLAS searches for chargino pair production across three different search channels are statistically combined for the first time, leading to a significant enhancement of the total sensitivity reach.

The first three chapters of this thesis provide background information on searches for SUSY with the ATLAS experiment. Specifically, Chapter 1 provides a theoretical overview of the SM and SUSY, including a discussion of the shortcomings and open questions of the SM and how these motivate the existence of SUSY. Chapter 2 then describes the LHC and the ATLAS experiment, including details of the ATLAS detector and how the particles produced in collisions are reconstructed. How this information is used to perform searches for new physics is the subject of Chapter 3, which also includes a discussion of how machine learning techniques can be used to improve their sensitivity compared to more traditional analysis techniques. Following this background information, dedicated chapters detail each of the three studies mentioned above. The new search for direct stau production using machine learning is presented in Chapter 4, the exploration of new channels charginos and neutralinos with compressed mass spectra in Chapter 5 and the combination of searches for chargino pair production in Chapter 6, after which final conclusions are presented.

Chapter 1

The Standard Model and Supersymmetry

“The real goal of physics is to come up with an equation that could explain the universe but still be small enough to fit on a T-shirt.”

Leon M. Lederman, physicist

As mentioned in the introduction, the SM of particle physics details the current best understanding of the fundamental particles of the universe and how they interact with each other. It was constructed predominantly over the 20th century, with the Higgs boson being the most recent particle in the SM to be discovered by the ATLAS [5] and CMS [6] experiments in 2012. Over this time, and since, the SM has been tested to extremely high precision by numerous experiments, not least by those at the LHC in recent years. For example, Figure 1.1 shows measurements performed by the ATLAS experiment of the production cross-sections¹ of numerous SM processes compared to their theoretical predictions in the SM. The predictions agree across 13 orders of magnitude, including for extremely rare processes with very small cross-sections.

While the predictions of the SM agree with the vast majority of the results from experiments designed to test it², there are what will here be called “particle-level” phenomena that the SM cannot explain. Furthermore, it can be argued that theoretical

¹ Cross-sections describe the rate at which a given interaction between particles occurs.

² There are a couple of possible exceptions, which will be discussed later in this chapter.

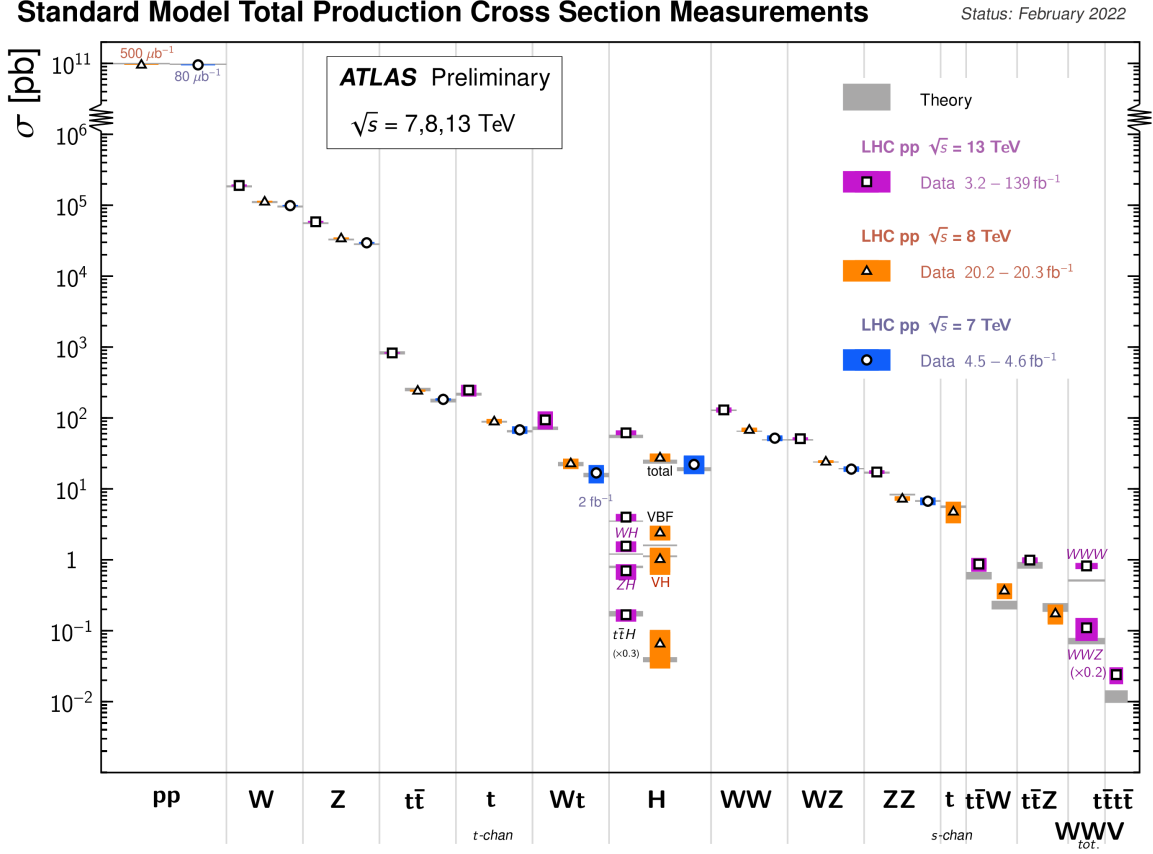


Figure 1.1: Summary of Standard Model cross-section measurements made with the ATLAS experiment, compared to their theoretical predictions. Figure taken from reference [7]. Note cross-sections are given in units of barn, b, where $1 \text{ b} = 10^{-28} \text{ m}^2$.

aspects of the SM suggest the existence of a higher level theory. Consequently, there are several motivations for extending the SM. Supersymmetry is one such proposed extension, which looks to address some of these questions. This chapter first provides a brief description of the SM in Section 1.1, which is followed by a discussion of its shortcomings and open questions in Section 1.2. Supersymmetry is then introduced in Section 1.3 as a possible solution to some of these shortcomings. In the first two sections, emphasis is placed on aspects on the SM and its shortcomings that are particularly relevant to Supersymmetry.

1.1 The Standard Model

The SM is a relativistic quantum field theory (QFT) in which particles are excitations of quantum fields that exist in four-dimensional Minkowski space-time. It describes all known particles and how they interact with each other via the electromagnetic, weak and strong forces. This section aims to review the key aspects of the SM, beginning with the constituent particles and their interactions followed by a discussion of the underlying structure of the SM, which reveals itself through symmetries.

1.1.1 Particles and Interactions

Particles

Empirically, the SM consists of 17 fundamental particles³ and at least 19 experimentally determined parameters⁴ governing their properties and interactions. The particles can be categorized by their properties in several ways. Firstly, they can be classified based on their spin, s , a property which describes a particles' intrinsic angular momentum, as either bosons or fermions depending on whether they have integer or half-integer spin respectively. Bosons are described as mediating interactions between themselves and other particles whilst fermions are described as matter particles.

Bosons can be divided into gauge-bosons, which have $s = 1$ and the Higgs boson, which has $s = 0$. The gauge-bosons consist of the photon, γ , which mediates the electromagnetic (EM) interaction; the $W^\pm$ and Z-bosons which mediate the weak interaction and gluons, g , of which there are 8 varieties, that mediate the strong interaction. The photon and gluon are massless while the $W^\pm$ and Z-bosons have masses around 100 GeV. The Higgs field, from which the Higgs boson, h , arises, meanwhile has a special role; fermions acquire their mass through interactions with the Higgs field. The Higgs field is also crucial to the process of electroweak symmetry breaking (EWSB). Prior to EWSB, which is thought to have occurred at some point in the early universe, the electromagnetic and weak interactions can be considered as different manifestations of the same interaction (the electroweak interaction) and the

3 Though this number depends exactly on how the particles are counted, for example whether left and right-handed fermions are counted separately.

4 There are exactly 19 free parameters taking neutrinos to be massless. Including neutrino masses and oscillations, discussed in Section 1.2.1, adds at least a further seven parameters.

$W^\pm$ -bosons, Z-boson and photon are replaced by four massless gauge bosons, $W^{1,2,3}$ and B with the W^1 field having positive electric charge, W^2 having negative electric charge and W^3 being electrically neutral. The importance of EWSB will be discussed in the next section on symmetries. Of the bosons, only the photon is stable; the others exist only for a short-time before decaying.

Fermions are divided into quarks, which can interact via the strong force, and leptons which do not. Six flavours of quark have been discovered; up (u), down (d), charm (c), strange (s), top (t) and bottom (b), three of which have an electric charge of $+2/3$ (u, c, t) and three with an electric charge of $-1/3$ (d, s, b). The quarks are grouped into three generations, (u, d), (c, s) and (b, t), each generation consisting of a quark with charge $+2/3$ and one with charge $-1/3$. The masses of the quarks increases across the generations with (u, d) being the lightest and (t, b) the heaviest.

Due to their interactions with the strong force, quarks carry a colour charge of either red, green or blue. Aside from the top quark, quarks do not exist freely but instead exist in bound states with zero total colour charge called hadrons (such as the proton and neutron) which are held together by the strong force. That quarks exist only in these bound states is known as confinement and is a consequence of the nature of the strong interaction, which will be discussed later in this section. The top quark decays too quickly (due to its large mass) to form these bound states. Of the known hadrons, only the proton is not known to be unstable; other hadrons only exist for a short amount of time before decaying, although fortunately for us neutrons can form stable nuclei with protons. Parton Density Functions (PDFs) are used to describe the distribution of energy and momenta of the constituent quarks and gluons, collectively known as partons, in these bound states.

Leptons can be further divided into those with an electric charge of -1 and those that are electrically neutral, called neutrinos, and again come in three generations. The three generations of charged leptons, in order of increasing mass, are the electron, muon and tau-lepton, each of which has an associated neutrino. Of the charged leptons, only the electron is stable while the muon and tau-leptons have lifetimes of $2.2 \mu\text{s}$ and 0.29 ps respectively [8]. In what will here be called the ‘vanilla’ version of the SM, the neutrinos are all massless but experimental results over the past 20 years have demonstrated that the different flavours of neutrino can oscillate into each other, which suggests they do have (very small) masses. Neutrino masses are discussed more in Section 1.2.1.

All particles have the properties of charge, spin and mass which are listed in Table 1.1. Fermions also have a property called chirality. Fundamentally, chirality is a purely quantum mechanical property related to the phase of a particles wavefunction. Fermions can have either left or right-handed chirality, with the left and right-handed variants being excitations of different fields⁵. For massless particles, chirality is the same as helicity, which describes whether a particle's spin aligns with its direction of motion. A key feature of the SM is the existence of anti-particles: for each SM particle there exists also an anti-particle which has identical mass but opposite charges. Some particles (notably the γ and the Z-boson) are their own anti-particle. Particles also have a property called parity, which describes whether their wavefunction changes sign upon spatial inversion. Fermions and the Higgs-boson have positive parity (meaning the wavefunction does not change sign) while anti-fermions and gauge-bosons have negative parity.

Interactions

Interactions can occur between fermions and bosons and between different bosons. Some bosons can also interact with themselves. These interactions can be represented by Feynman diagrams which are constructed of interaction vertices. Only certain kinds of interaction vertex are allowed. Beginning with the EM force, the only allowed interaction between fermions and photons is illustrated by Figure 1.2. These EM interactions can only occur for fermions with non-zero electric charge. As photons have no electric charge of their own, they cannot self interact, but they can interact with W-bosons (pictured later). The EM interaction is further described by Quantum Electrodynamics (QED).

Permitted interaction vertices governed by the strong force are shown in Figure 1.3. Interactions between fermions and gluons, illustrated by Figure 1.3(a), can only occur for fermions with colour charge (quarks). Unlike the photon, the gluon itself has non-zero colour charge and so interactions between gluons are possible, as shown by Figures 1.3(b) and 1.3(c). The strong interaction is further described by Quantum Chromodynamics (QCD).

⁵ As a consequence of not interacting with any of the strong, EM or weak interactions, right-handed neutrinos have not been observed and so nominally are not included in the SM. Their possible inclusion is discussed in this chapter in Section 1.2.1.

Type	Particle	Electric Charge	Spin	Mass
Quarks	up, u	$+2/3$	$1/2$	2.16 MeV
	down, d	$-1/3$		4.67 MeV
	charm, c	$+2/3$		1.27 GeV
	strange, s	$-1/3$		93.4 MeV
	top, t	$+2/3$		172.69 GeV
	bottom, b	$-1/3$		4.18 GeV
Leptons	electron, e	-1	$1/2$	0.549 MeV
	electron neutrino, ν_e	0		0^+
	muon, μ	-1		105.66 MeV
	muon neutrino, ν_μ	0		0^+
	tau-lepton, τ	-1		1.777 GeV
	tau neutrino, ν_τ	0		0^+
Gauge bosons	photon, γ	0	1	0^*
	$W^\pm$ -bosons, $W^\pm$	± 1		80.377 GeV
	Z-boson, Z	0		91.1876 GeV
	gluon(s), g	0		0^*
Scalar bosons	Higgs boson, h	0	0	125.25 GeV

Table 1.1: The particles in the SM with their electric charges, spins and experimentally determined masses, taken from [8]. *For the photon and gluon, the theoretical masses of zero are used. †Similarly, for neutrinos masses are here listed as zero but are discussed in Section 1.2.1.

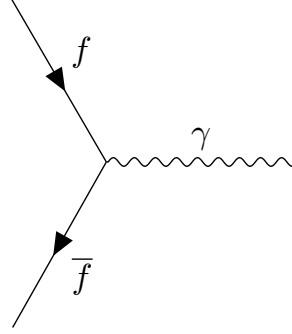
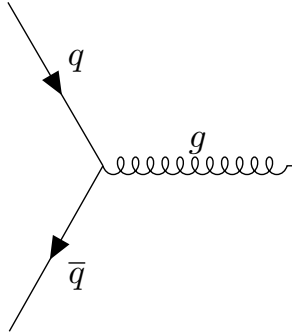
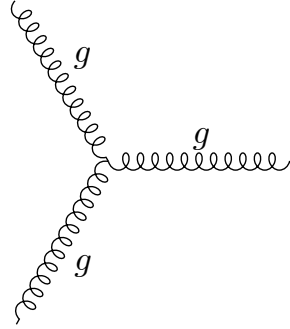


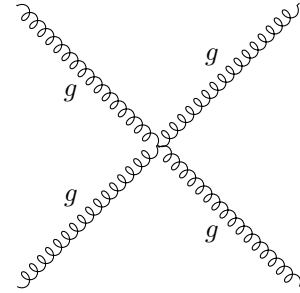
Figure 1.2: EM interaction vertex between fermions and a photon, f denotes any fermion with antiparticle $\bar{f}$. Diagrams produced using TikZ-Feynman [9].



(a) Strong interaction



(b) Triple gluon vertex



(c) Quartic gluon vertex

Figure 1.3: Strong interaction vertices, q denotes any quark. Diagrams produced using TikZ-Feynman [9].

The allowed weak interactions are shown in Figure 1.4. For fermionic weak interactions it is useful to distinguish between two types; charged current weak interactions mediated by the $W^\pm$ -bosons (Figure 1.4(a)) and neutral current weak interactions mediated by the Z -boson (Figure 1.4(b)). For fermions, only left-handed fermions (and right-handed anti-fermions) interact via the charged current weak interaction while all known fermions interact via the neutral current, though the strength of the interactions depends upon the chirality of the fermion. For interactions between leptons and $W^\pm$ -bosons, in order to conserve electric charge, charged leptons can only interact with neutrinos. Only interactions involving the same flavour of charged lepton and neutrino have been observed.

For interactions between quarks and $W^\pm$ -bosons, an up-type quark must interact with a down-type quark in order to conserve electric charge. Unlike leptons, quarks can interact across the generations, although the interactions are strongest between quarks of the same generation. The relative strengths of the charged weak interactions

between quarks are encoded in the CKM matrix. $W^\pm$ -bosons can also interact with Z-bosons, themselves and with photons as previously mentioned, as shown by Figures 1.4(c)-1.4(e). The weak interaction is the only known interaction to violate parity; weak interactions are not symmetric under a parity transformation, whereas strong and EM interactions are symmetric.

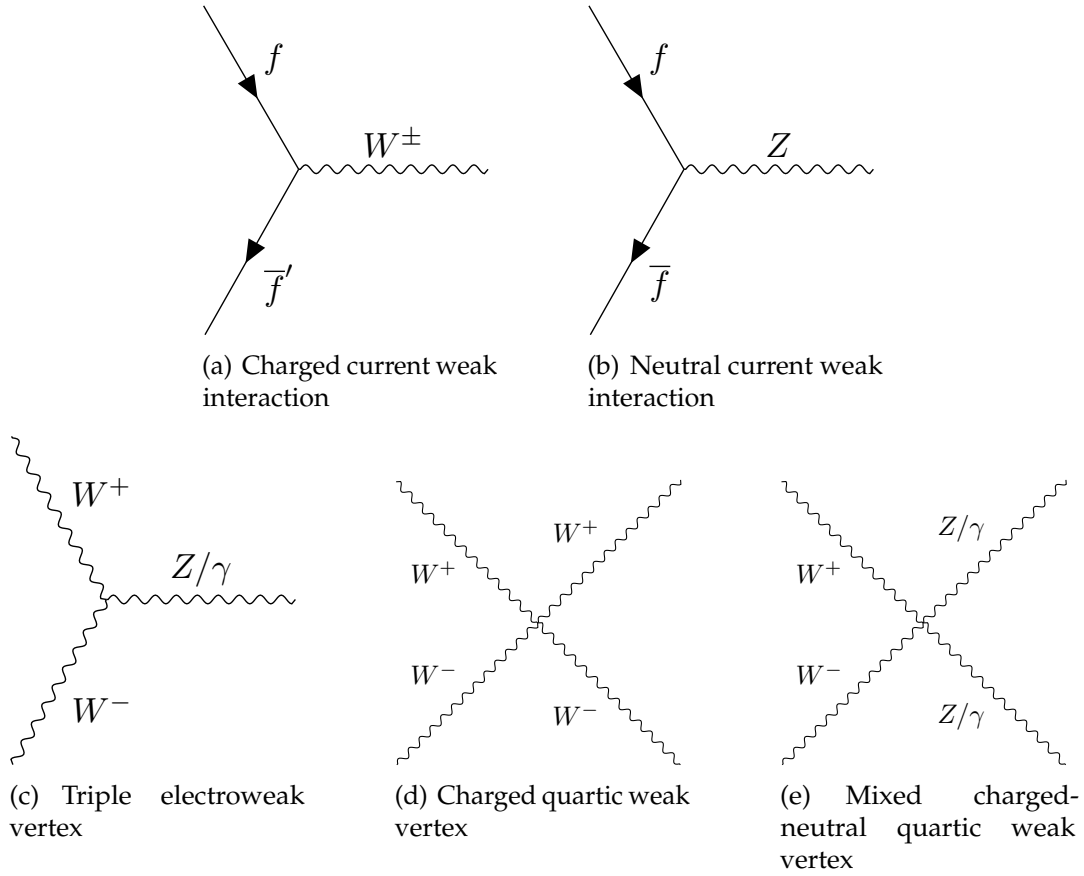


Figure 1.4: Weak interaction vertices. In these diagrams, f denotes any fermion with antiparticle $\bar{f}$, f' a different variety of fermion and Diagrams produced using TikZ-Feynman [9].

The Higgs boson interacts with all fermions and bosons that have mass, via the vertices shown in Figure 1.5. Self interactions of the Higgs boson, as shown in Figures 1.5(c) and 1.5(d), are also possible although are yet to have been experimentally verified.

An important aspect of the SM is the relative strength of the different interactions which are governed by coupling ‘constants’, here denoted α_{EM} , α_S and α_W corresponding to the electromagnetic, strong and weak interactions respectively. As a consequence of renormalization (discussed briefly in the next section), the values

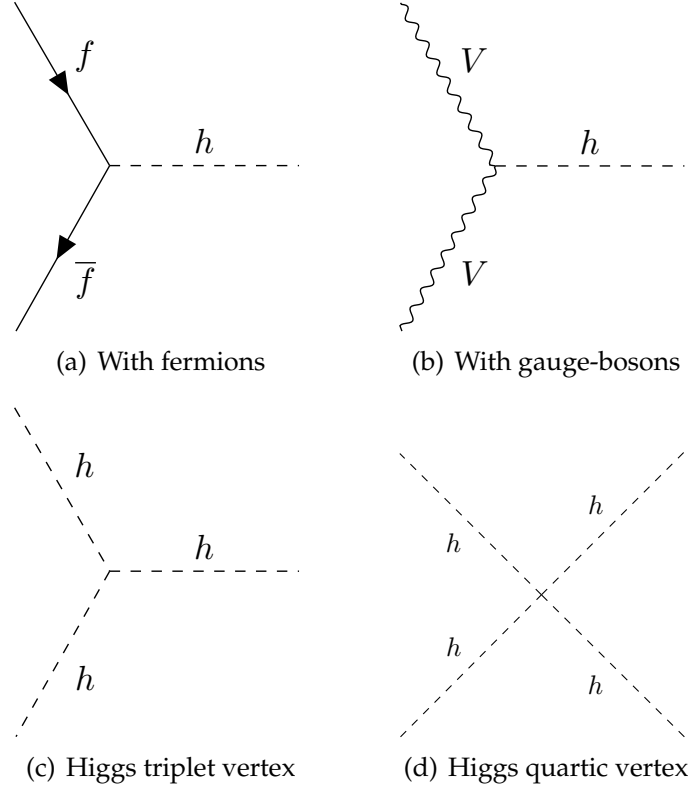


Figure 1.5: Allowed interactions between the Higgs-boson with fermions, gauge-bosons and itself. V denotes either a $W^\pm$ - or Z -boson. Diagrams produced using TikZ-Feynman [9].

taken by these coupling constants vary with the energy scale being considered, in particular the value of α_{EM} increases with energy while the value of α_S decreases. The latter is known as the asymptotic freedom of QCD. At an energy scale of 1 GeV, typical of most interactions at the LHC, these constants take values of $\alpha_{EM} \sim 0.007$, $\alpha_S \sim 1$, and $\alpha_W \sim 0.03$, hence strong interactions are dominant at the LHC. Importantly, although the value of α_W is greater than that of α_{EM} , the rate of weak interactions are suppressed with respect to strong and EM interactions due to the mediating W -bosons and Z -bosons having mass.

The confinement of quarks to bound states, mentioned earlier, is a consequence of gluons being massless and self-interacting combined with the value of α_S decreasing with energy. These effects cause the strength of the strong force to not decrease as the separation between two coloured particles increases. In addition to restricting quarks to exist in bound states, a further consequence of confinement is that when quarks or gluons are produced at particle colliders, they rapidly interact via the strong force to produce collimated collections of hadrons, called jets, a process called hadronization.

Hence, individual quarks and gluons are not able to be reconstructed at particle colliders, only the jets they form.

The strength of interactions of the gauge-bosons and fermions with the Higgs boson is also governed by α_W but for fermions the interaction is suppressed by a factor of m_f/m_W where m_f is the fermion mass and m_W is the W -boson mass. Thus, interactions with the Higgs boson are very weak for all but the heaviest of fermions.

1.1.2 Structure and Symmetries

That the particles of SM can be classified in the way described above, and that their interactions obey particular rules, suggests some high level of structure is present in the SM. This is indeed the case and the structure can be viewed more clearly through observing the symmetries of the SM. To understand these symmetries, it is necessary to know that the SM can be expressed as a Lagrangian. As in classical mechanics, the Lagrangian, L , is the kinetic energy minus the potential energy, and the action, S , is the integral of L over time. For relativistic QFTs, a more useful quantity is the Lagrange density $\mathcal{L}$ (often, and here henceforth, just referred to as the Lagrangian), which is the derivative of the Lagrangian in 3-dimensional space, and is in general a function of the particle quantum fields. The action can therefore be written as

$$S = \int L dt = \int \mathcal{L} d^4x. \quad (1.1)$$

By the principle of least action, the dynamics of a physical system can be determined from requiring the action to be extremal, i.e. $\delta S = 0$. Hence, the dynamics of the SM are determined from its Lagrangian.

The full SM Lagrangian consists of many terms which describe the dynamics of the individual particles and how they interact with each other. A condensed version of the SM Lagrangian, prior to the process of electroweak symmetry breaking (EWSB) mentioned earlier, can be written as:

$$\mathcal{L}_{\text{SM}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + y_i\bar{\psi}_{L_i}\phi\psi_{R_i} + \text{h.c.} + |D_\mu\phi|^2 - V(\phi). \quad (1.2)$$

The first term describes the dynamics and self-interactions of the gauge fields; there is one such term of each of the gluon, W and B fields. The second term encodes how the fermions interact with the gauge-bosons, with ψ and $\bar{\psi}$ being fermion and anti-fermion

fields and $\mathcal{D}$ being the covariant derivative, which includes gauge field terms. The third term details how the fermions interact with the Higgs field, ϕ , with the sum running over the three lepton and six quark flavours and the L and R denoting left and right-handed chirality respectively. The strength of these interactions are governed by the size of the couplings y_i which are known as Yukawa couplings. The “h.c.” following this term stands for the hermitian conjugate of the previous term which represents the interaction of anti-fermions with the Higgs field. The penultimate term represents the interactions of the weak gauge fields with the Higgs field and the final term represents the potential of the Higgs field. With all these terms, an important property of the SM Lagrangian is that it is renormalizable, which put simply means that divergences which arise in certain calculations (of for example cross-sections) can be dealt with in a self-consistent way.

To return to symmetries, in general, a symmetry describes a transformation that can be made to the fields under which the action, and thereby any physical observable, is invariant. By Noethers’ theorem [10], every symmetry of this kind leads to a conserved quantity, for instance invariance under translations in time yields conservation of energy. Symmetries can be described by transformations under groups with the generators of the groups controlling the transformations. Four types of symmetries are present in the SM; the fundamental symmetries of space-time, internal symmetries, accidental symmetries and approximate symmetries. The space-time symmetries are described by the Poincaré group and encode invariance under space-time translations, 3D-rotations and Lorentz transformations which in-turn ensure the SM obeys conservation of energy and momentum and complies with the postulates of special relativity.

The internal symmetries of the SM arise from redundancies in the definitions of the gauge fields; these fields can be modified in ways that do not change the action. This is known as gauge invariance. The internal symmetries differ before and after EWSB. Before EWSB, the SM had an overall internal symmetry of

$$SU(3)_C \times SU(2)_L \times U(1)_Y, \quad (1.3)$$

where $SU(3)_C$ corresponds to the eight gluon fields and the conservation of colour charge; $SU(2)_L$ corresponds to the three weak fields $W^{1,2,3}$ and the conservation of weak iso-spin; and $U(1)_Y$ the B field and conservation of weak hyper-charge. As not

Chapter 1: The Standard Model and Supersymmetry

Name	Fields	$(SU(3)_C, SU(2)_L, U(1)_Y)$
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$(3, 2, \frac{1}{3})$
	u_R	$(\bar{3}, 1, \frac{2}{3})$
	d_R	$(\bar{3}, 1, -\frac{1}{3})$
Leptons	$\begin{pmatrix} e \\ \nu \end{pmatrix}_L$	$(1, 2, -\frac{1}{2})$
	e_R	$(1, 1, 1)$
Gauge-bosons	W	$(1, 3, 0)$
	B	$(1, 1, 0)$
	g	$(8, 1, 0)$
Higgs	H	$(1, 2, \frac{1}{2})$

Table 1.2: The particles in the Standard Model, prior to electroweak symmetry breaking, and their representations in the gauge symmetry groups.

all particles experience all interactions, they have different representations under these groups, illustrated in Table 1.2.

Upon EWSB, the W^3 and B fields mix to give the Z-boson and the photon and the $SU(2)_L \times U(1)_Y$ is broken to leave only a $U(1)_{EM}$ symmetry, which corresponds to the photon and conservation of electric charge. EWSB is caused by the Higgs field acquiring a non-zero vacuum expectation value (VEV) as the temperature of the universe decreased shortly after the Big Bang. This is a result of the Higgs potential having a global minimum not situated at zero. The Higgs VEV is found to be 246 GeV (as determined from measurements of the masses of the $W^\pm$, Z and h and the strength of electroweak interactions), and is related to the energy scale at which EWSB occurs. The Higgs boson exists as a result of EWSB, thus EWSB is responsible for the origin of fermions masses. Additionally, the $W^\pm$ - and Z-bosons also acquire their mass through EWSB.

Turning to accidental symmetries, each flavour of lepton has an associated lepton number which is +1 for leptons of that flavour, -1 for anti-leptons of that flavour and 0 for all other particles. Similarly, quarks have a baryon number of +1, anti-quarks a baryon number of -1 and all other particles have a baryon number of 0. The total

lepton number(s) and baryon number are conserved in all SM interactions, but these are considered “accidental” symmetries as they are not built into the SM Lagrangian, and are not represented by a group symmetry. Parity conservation in electromagnetic and strong interactions are also accidental symmetries.

Similarly, the SM also has approximate symmetries between the different generations of quarks and leptons with each generation experiencing the same interactions. While these can be represented by groups, these symmetries are only approximate due to the masses of quarks and leptons differing across the generations. The SM is also approximately symmetric under CP transformations. CP stands for charge-conjugation and parity; a CP transformation is one in which a particle is transformed to its antiparticle and parity is inverted. The SM is only approximately CP symmetric as processes in the quark sector have been observed to occur at different rates under CP transformations, for example in the decay of neutral kaons (a meson composed of down and anti-strange quark) [11, 12]. The origin of this CP violation is from non-zero imaginary components of the CKM matrix mentioned earlier. In addition, CP violation is in principle allowed in QCD but has not been observed, a phenomenon known as the strong CP problem which is discussed briefly in the next section. Neutrino oscillations, also discussed in the next section, can provide another source of CP violation as well.

1.2 Shortcomings and Open Questions

Having reviewed key aspects of the SM, focus turns to the phenomena the SM cannot explain. Here, these are divided into purely particle scale observations and then astrophysical/cosmological scale observations which could have particle level explanations. Additionally, as mentioned earlier it can be argued that theoretical aspects of the SM suggest the existence of a more fundamental theory⁶. In the following discussion of these shortcomings, more attention is paid to those with particular relevance to Supersymmetry.

1.2.1 Particle-scale Observations

The most obvious ‘shortcoming’, already mentioned in previous section is neutrino masses. In the vanilla formulation of the SM, neutrinos do not have mass because the absence of right-handed neutrinos means there are no Yukawa terms, describing the fermionic interactions with the Higgs field, in the SM Lagrangian for neutrinos⁷. However, recent experiments have demonstrated that the different flavours of neutrinos are capable of oscillating into each other. The simplest way with which this could be understood is with the existence of neutrino mass eigenstates, at least two of which must have non-zero mass for the oscillations to be possible. Combining results of several neutrino experiments, the best fit values of the two neutrino mass differences are found to be 8.6×10^{-3} eV and 0.05 eV [13], meaning at least one of the mass eigenstates has a mass greater than 0.05 eV. An upper limit on the sum of the neutrino masses of 0.2 eV [14] has also been obtained from a combination of cosmological and astrophysical observations.

Right-handed neutrinos can be added (as singlets in all gauge groups) to the SM, giving *Dirac* mass terms in the Lagrangian analogous to those for the charged leptons. As right-handed neutrinos would not interact with any of the strong, weak or electromagnetic forces they would be considered sterile. As a consequence of this sterility, *Majorana* mass terms, representing a different type of interaction between the neutrinos and the Higgs field, could also be added to the SM Lagrangian. These terms would imply that the right-handed neutrinos were their own anti-particles. Such terms

⁶ More fundamental in the sense that the SM could be derived from it.

⁷ Generic fermion mass terms also cannot be added to the SM as such terms would not be gauge invariant.

would not be renormalizable and so the Standard Model would instead need to be treated as an Effective Field Theory (EFT), which breaks down at some high energy scale. A see-saw mechanism between the Dirac mass and Majorana mass could explain why the neutrino masses are so small. The other possibility is that there is a reason why Majorana mass terms could not be added to the SM Lagrangian and potentially also an explanation as to why the Dirac masses for neutrinos are so small. Either way, additions to the vanilla SM Lagrangian are needed to explain the existence of neutrino masses. The experimental bounds on neutrino masses put the energy scale at which the SM breaks down as being around 10^{15} GeV, although this only represents an upper bound; this treatment does not say the SM cannot break down at much lower energies.

Alongside neutrino masses, there are a couple of other recent particle-level experimental results that the SM might not be able to explain, although their status is yet to be unanimously agreed upon by the particle physics community. The first is the measurement of the anomalous magnetic moment of the muon, $g - 2$, which has recently been measured at the Fermilab National Accelerator Laboratory to be more than five standard deviations higher than was originally predicted based upon the SM [15]. However, more recent theoretical calculations have produced values of $g - 2$ that are significantly closer to the experimental value [16], so at present it is unclear whether this measurement can be taken as evidence of physics beyond the Standard Model (BSM).

The second result is a precise measurement of the W -boson mass by the CDF collaboration, which differs from the SM prediction (based upon global fits of other measured quantities, such as the Z -boson mass) by 7.2 standard deviations [17]. However, as well as being at odds with the SM prediction, this measurement is also at odds with other measurements of the W -boson mass made by other experiments, including a recent result released by the ATLAS collaboration which has comparable statistical uncertainty [18]. Hence, more work is needed to understand the sources of the discrepancies between the experimental results before this measurement can be considered as evidence for BSM physics.

1.2.2 Astrophysical and Cosmological Observations

Turning to astrophysics and cosmology, there are observed phenomena that could potentially have particle level explanations. The most apparent of these is dark matter

(DM). A variety of astrophysical observations, across a large range of length scales, suggest the existence of additional matter in the universe which has mass but does not interact via the strong or electromagnetic forces (reference [19] provides a summary). One example, at the galactic scale, is that the speed with which stars orbit spiral galaxies does not decrease with distance from the galactic centre as rapidly as would be expected based upon the distribution of visible matter (and using General Relativity), but could be explained by the existence of a “halo” of DM on the outskirts of galaxies. Another example, this time at the inter-galactic level, is that the motion of colliding galactic clusters suggests they contain a significant amount of DM. Further examples include strong and weak gravitational lensing and fits of temperature fluctuations in the Cosmic Microwave Background [20]. Under the Λ -Cold Dark Matter model of cosmology, and based upon these astrophysical and cosmological observations, DM accounts for 27% of the mass-energy of the universe, compared to around 5% for the normal matter in the SM. The only possible DM candidates in the SM, the (left-handed) neutrinos, would travel at too high velocities to be considered as cold DM, instead being a form of hot DM [21].

Several new types of particle could account for cold DM. One of the most popular candidates are Weakly Interacting Massive Particles (WIMPs); particles with masses of order 100 GeV that only interact weakly (either via the weak force or some as yet undiscovered force which is at least as weak). WIMPs are a popular DM candidate due to what is known as the ‘WIMP miracle’: if a stable WIMP exists then, based upon the current understanding of the early universe, it would naturally be produced in an abundance which matches the experimentally determined amount of DM in the Universe today, a quantity known as the DM relic density [22].

Aside from astrophysical observations, evidence for the existence of DM could be found from searches at particle colliders, such as the ones discussed in the rest of this thesis, or by direct detection experiments. While DM would pass through particle detectors without leaving a trace, its presence could be inferred from the presence of significant amounts of missing energy. This is discussed in more detail with reference to the ATLAS detector in the next chapter. Many searches for particulate DM at colliders have been performed, including by experiments at the LHC, but have so far yielded no significant evidence for its existence.

Direct detection experiments aim to observe DM from the recoil of atomic nuclei caused by interactions between the nuclei and DM particles arriving from space. Numerous experiments searching for such signatures have been constructed and are

located underground in order to reduce SM backgrounds. So far, the data collected from these experiments also does not provide any strong evidence for particle DM [23]. However, combined with the results from astrophysical observations and collider searches, limits have been placed on various DM scenarios but there remains much unexplored parameter space.

Aside from DM, there exist some arguably even greater unresolved issues when trying to marry the SM with cosmological observations. One of these is the matter-antimatter asymmetry in the universe; matter comprises almost all observed visible matter but the SM would suggest matter and anti-matter should have been created in more equal proportions during the Big-Bang. The observed amount of CP violation in the SM cannot account for the observed matter-antimatter discrepancy. Another, and arguably the biggest short-coming of the SM (if considering it as a theory of everything) is the current omission of gravity. While a spin-2 graviton, mediating gravity, could be added to the SM, there are challenges in maintaining renormalizability and accurately describing the behaviour of gravitational waves. Furthermore, one of the biggest unresolved issues in modern physics is the fact that the value of the vacuum energy determined from cosmological models is many orders of magnitude (at least 60 according to [24]) smaller than what it is calculated to be from the SM and QFT, although it is perhaps unfair to consider this disagreement as purely as a shortcoming of the SM.

1.2.3 Suggestions of a more Fundamental Theory

Gauge hierarchy problem

In addition to these unexplained phenomena, certain aspects of the SM suggest the existence of a more fundamental theory. The first of these is known as the gauge hierarchy problem. In the SM, the mass of the Higgs boson is formulated as a rest term, known as the bare mass m_0 , which is a free parameter in the SM, that receives contributions from all other particles that interact with the Higgs field:

$$m_h^2 = m_0^2 + \sum_i \delta m_{h,i}^2, \quad (1.4)$$

where the sum is over all particles (with mass) in the theory. These corrections arise from loop diagrams, examples of which are shown in Figure 1.6.

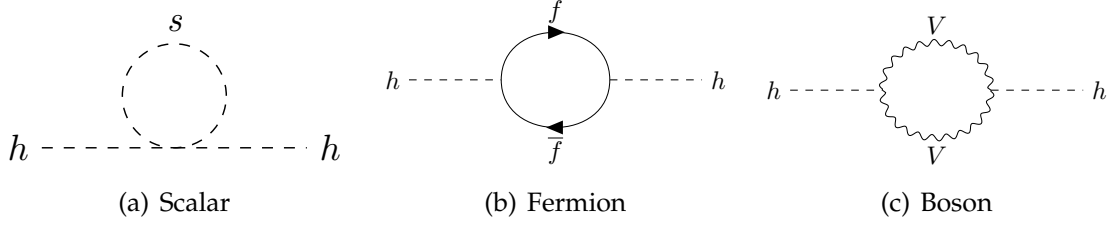


Figure 1.6: Feynman diagrams illustrating the one-loop corrections to the Higgs boson mass, produced using TikZ-Feynman [9].

Treating the SM as an effective field theory (EFT) which breaks down at some energy scale Λ , for a particle of mass m_i the one loop corrections have the form

$$\delta m_{h,i}^2 \propto \pm \lambda_i^2 \left[\Lambda^2 + C m_i^2 \ln\left(\frac{\Lambda}{m_i}\right) \right], \quad (1.5)$$

where λ_i represents the size of the coupling to the Higgs field and C is a constant, the value of which depends on whether the particle is a boson or a fermion as does the sign of the correction. The important thing to note is that part of this correction is proportional to Λ^2 , meaning the corrections to the Higgs boson mass are quadratically divergent with respect to this energy scale. Following the discussion of neutrino masses, Λ could be as large as 10^{15} GeV (or using the Planck mass (the energy scale associated with gravity), 10^{19} GeV). The central question of the gauge hierarchy problem is how the Higgs boson mass is only ~ 100 GeV if it can receive corrections many orders of magnitude larger, without the need for significant fine-tuning of m_0 (having it take a very precisely tuned value such that the total Higgs boson mass agrees with the experimentally measured value). To give strength to this problem, it is worth noting that corrections to fermion masses in the SM receive similar corrections due to their interactions with gauge bosons, but in this case the masses are protected from diverging by the gauge symmetries of the SM (which cause various diverging terms to cancel with each other).

Grand Unification

A second suggestion of a more fundamental theory is related to electroweak symmetry breaking; given that the electromagnetic and weak forces can be unified above a certain energy, then potentially the strong force can be unified with the electroweak forces at a higher energy. This idea is known as Grand Unification and in general

means that the SM has, above this energy scale, some higher symmetry than $SU(3)_C \times SU(2)_L \times U(1)_Y$. Supporting this hypothesis, the relative strengths of the forces in the SM, which as mentioned previously depend upon the energy scale being considered, are found to *almost* coincide at high energies. Potential additions to the SM could allow the couplings to exactly coincide. A larger symmetry group could also explain the quantized nature of the different charges of the SM particles.

Additional Indications

There is also the strong CP problem, introduced earlier, which put simply is that, in the current formulation of QCD, CP-violation can occur but has not been observed meaning that tight constraints can be placed on the parameter controlling the rate of CP-violating processes⁸. Again the problem is one of fine-tuning; an explanation as to why the value of this parameter is so small is desired. Finally, there are several other peculiarities in the SM for which an explanation is coveted. For instance why there are exactly three generations of quarks and leptons, an explanation for the accidental symmetries discussed in the previous section, why the charges of the quarks come in units of $1/3$ and why the masses of the SM particles occupy such a large range of masses (a more general hierarchy problem). More generally, between the parameters of the SM there are 19 degrees of freedom (not including neutrino mass oscillations, which provide at least 7 more) that at present have to be taken from experiment. A more fundamental theory could provide the origin of or place additional constraints on their values.

⁸ The strongest constraints come from the experimental upper bounds on the neutron electric dipole moment [25].

1.3 Supersymmetry

Several of the shortcomings of the SM discussed in the previous section can be addressed by introducing new particles and a new symmetry called Supersymmetry, or SUSY for short. SUSY would be a new space-time symmetry between bosons and fermions which would require the existence of a new boson and a new fermion for each existing fermion and boson in the SM respectively, each sparticle differing in spin by $1/2$ compared to its SM partner. As mentioned in the introduction, collectively these new particles are known as *superparticles* or *sparticles*. Under a supersymmetric transformation, the SM particles would transform to their superpartner and vice-versa under the super-Poincaré group (which is the Poincaré group with this additional boson-fermion symmetry).

If SUSY were an exact symmetry, then the sparticles would have the same masses as their SM partners. This is clearly not the case as sparticles with such masses would have been visible in numerous previous experiments, so, if it exists, SUSY must be a broken symmetry. The formulation of SUSY that introduces the fewest number of new particles and interactions is known as the Minimal Supersymmetric Standard Model (MSSM) [26, 27]. In the following, we will briefly review the particle content and key aspects of SUSY and the MSSM in particular before discussing the shortcomings of the SM that SUSY can address.

1.3.1 Key Aspects

Particle Content

In SUSY, the SM particles are put into supermultiplets with their superpartners. In particular, SM fermions form chiral supermultiplets with their superpartners which are scalars ($s=0$), while SM bosons form gauge supermultiplets with their superpartners which are spin- $1/2$ fermions. The superpartners of the SM fermions are given the same names as their SM partners but with an 's' prefix, so for example the superpartner of the electron is called the selectron. For the SM bosons, the superpartners, collectively called gauginos, are affixed with an 'ino' suffix, for instance the superpartner of the SM B -field is the bino field. The sparticle fields take the same label as their SM partners, but with an above tilde ($\sim$) symbol. It is important to note that although there exist

Chapter 1: The Standard Model and Supersymmetry

Chiral supermultiplets	Spin=0	Spin=1/2
Squarks, quarks	$\begin{pmatrix} \tilde{u} \\ \tilde{d} \end{pmatrix}_L$ $\tilde{u}_R$ $\tilde{d}_R$	$\begin{pmatrix} u \\ d \end{pmatrix}_L$ u_R d_R
Sleptons, leptons	$\begin{pmatrix} \tilde{e} \\ \tilde{\nu} \end{pmatrix}_L$ $\tilde{e}_R$	$\begin{pmatrix} e \\ \nu \end{pmatrix}_L$ e_R
Higgs, Higgsinos	$\begin{pmatrix} H_u^+ \\ H_u^0 \end{pmatrix}$ $\begin{pmatrix} H_d^0 \\ H_d^- \end{pmatrix}$	$\begin{pmatrix} \tilde{H}_u^+ \\ \tilde{H}_u^0 \end{pmatrix}$ $\begin{pmatrix} \tilde{H}_d^0 \\ \tilde{H}_d^- \end{pmatrix}$
Gauge supermultiplets	Spin=1/2	Spin=1
Gluino, gluon	$\tilde{g}$	g
Wino, W-bosons	$\tilde{W}^\pm, \tilde{W}^0$	$W^\pm, W^0$
Bino, B	$\tilde{B}$	B

Table 1.3: Particle content of the Minimal Supersymmetric Standard Model (MSSM). Only one generation of lepton and up and down type quarks are shown but all three are present.

separate superpartners for the left and right-handed SM fermions, which are labelled accordingly, being scalars these superpartners have no handedness of their own.

One key aspect of SUSY is that an additional Higgs doublet field, with opposite hypercharge to the one in the SM, is required in order to avoid gauge anomalies (such as violation of electric charge) as is a modification of the existing SM Higgs doublet. The new Higgs doublet fields, each of which is complex, are labelled u and d as they are responsible for giving masses to up and down type quarks respectively (with H_d also being responsible for the charged lepton masses). Upon electroweak symmetry breaking (EWSB), five Higgs scalars remain, three of which are neutral, one with a charge of +1 and one with a charge of -1. The lightest neutral scalar is typically assigned to be the SM Higgs boson. The particle content of the MSSM, prior to EWSB, is shown in Table 1.3.

After the electroweak and super symmetries are broken, the electroweak gauginos and higgsinos with the same charges mix to form mass eigenstates. The electrically charged eigenstates are known as charginos while the neutral ones are known as neutralinos. They are labelled as $\tilde{\chi}_j^\pm$ and $\tilde{\chi}_j^0$ respectively, with $j = 1, 2$ for charginos and $j = 1, 2, 3, 4$ for neutralinos in order of increasing mass. Collectively, charginos and neutralinos are referred to as electroweakinos. Depending on the level of mixing, the individual electroweakinos may be said to be more wino, bino or higgsino like. Sfermions of equal charge can also mix with each other to form mass eigenstates. The amount and nature of the mixing depends upon how SUSY is broken. For the sfermions, in all cases the mixing is expected to be most significant for the third generation (i.e the stop quarks and staus) due to their larger Yukawa couplings.

SUSY breaking and Model Building

As discussed, SUSY must be a broken symmetry which suggests there may be a mechanism behind how it is broken. The supersymmetric Lagrangian can be separated into a part that preserves SUSY and a part that does not. The part that does preserve SUSY is relatively well-defined theoretically (as so far as what terms are allowed) but there are many possibilities for the SUSY breaking part. Consequently, it is helpful to consider SUSY not as one theory with precise predictions but instead as a framework from which particular supersymmetric theories, or models, can be constructed. The nature of the terms in the SUSY breaking part of Lagrangian has important implications for the phenomenology of SUSY, in particular the sparticle mass spectrum. The MSSM reduces the freedom of the SUSY violating part due to it requiring SUSY to be only softly broken (the meaning and motivation of which is discussed in the next section (Section 1.3.2) but even then the full SUSY Lagrangian contains over 100 free parameters.

R-Parity

SUSY introduces R -parity [28] as a quantity that, if conserved, prevents baryon and lepton number violating processes (which have been strongly constrained by experiment). The R -parity of a particle is defined as

$$P_R = (-1)^{3(B-L)+2s}, \quad (1.6)$$

where B is the particles' baryon number, L its lepton number and s its spin. SM particles have an R -parity of +1 and sparticles an R -parity of -1. For a SUSY model in which R -parity is conserved, every interaction must involve an even number of sparticles and no mixing between sparticles and particles is permitted. Two further important consequences of R -parity conservation are that 1) at colliders sparticles must be produced in even numbers and 2) the lightest SUSY particle (LSP) is stable. While a good mechanism for preventing problematic processes, it is also possible that R -parity violating couplings between sparticles are sufficiently small to suppress such processes. R -parity is however conserved in the MSSM.

1.3.2 Motivations

Having established what SUSY is we can turn to its motivations of which there are three key ones. The first is that SUSY can mitigate the gauge hierarchy problem introduced in Section 1.2.3. This is because the quadratically divergent corrections to the Higgs mass have opposite signs for bosons and fermions, and the sum of all the quadratically divergent corrections would in fact perfectly cancel if SUSY were an exact symmetry [29]. While it is known that SUSY is not an exact symmetry, it is still possible for the quadratic divergences to cancel. Models for which this is the case are said to have softly broken SUSY.

However, even in softly broken SUSY models it is still possible to arrive at some kind of hierarchy problem if terms in the SUSY breaking part of the Lagrangian are themselves too large that fine-tuning of the Higgs boson mass is still required. The values of parameters controlling the size of these terms are closely related to the masses of the gluino and stop quark (which has the largest impact on the corrections to the Higgs boson mass due to the large top quark mass), such that the requirement of the soft breaking terms to not be too large implies the stop quark and gluino must be relatively light. Exactly how light they are required to be is a complex question as it depends on how the amount of fine-tuning is quantified and also on whether SUSY is treated as an effective field theory (EFT) or as part of a larger theory that works at higher energy scales, such as string theory. Historically, EFT based approaches led to requiring the stop mass to be less than ~ 400 GeV in order to maintain naturalness ([30] and references there-in). Such a scenario is now comfortably excluded by the

searches performed by the ATLAS [31] and CMS [32] experiments; stop quark masses below 1 TeV are largely⁹ excluded at 95% confidence level.

By contrast, the value of the same naturalness measures can differ greatly when SUSY is evaluated as part of a larger theory in which SUSY parameters are correlated, highlighting that EFT approaches are in this case actually quite model-dependent [33]. A model-independent approach, outlined in [30], argues that in fact the stop quark can have a mass of several TeV without breaking naturalness. What is instead required in this approach, where naturalness is evaluated from the equation for the Z-boson mass, is that the Higgsinos must have masses close to that of the weak scale, more specifically less than a few hundred GeV, a scenario *not* completely ruled out by searches by ATLAS and CMS. Hence, it is difficult to determine any clear upper limits on sparticle masses for which SUSY would present, on its own, a natural solution to the hierarchy problem and it doing so is still a possibility.

A second motivation for SUSY can be found by revisiting Grand Unification. The presence of the additional particles in SUSY changes how the strength of the couplings vary with energy scale. For certain SUSY models, the strength of the couplings can in fact converge¹⁰ at high energies as illustrated in Figure 1.7.

Thirdly, SUSY can also provide compelling DM candidates, typically in the form of the lightest neutralino or the gravitino (which would be the superpartner of the graviton) [34, 35]. More precisely, this is the case for SUSY models where R -parity is conserved and the LSP is electrically neutral. Although not all models of this kind can fulfil the restrictions on DM that have been set by astrophysical observations, direct detection experiments and collider searches, there are many models which can survive these constraints [36].

In addition to these three key motivations, the mass of the Higgs boson being less than 135 GeV is often heralded as a success of SUSY, as this is required in the MSSM (with a more general limit related to grand unification requiring the mass to be less than 200 GeV) [37]. Interestingly, although as mentioned the results are not universally accepted, SUSY could potentially also (simultaneously) explain both the muon $g - 2$ anomaly [38] and a higher than expected W -boson-mass [39]. And as a

⁹ See discussion of simplified models in the next section (1.3.3)

¹⁰ Or at least the couplings strengths are significantly closer to converging than in the SM case, with differences being more easily able to be attributed to new particles near the convergence energy.

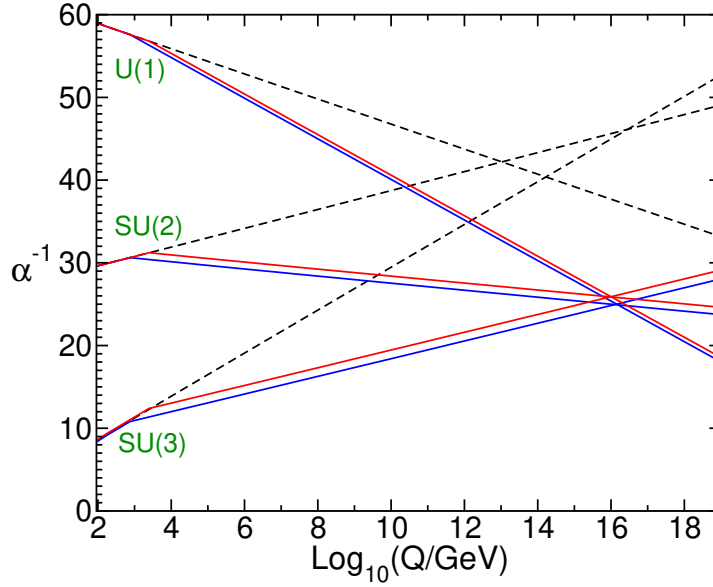


Figure 1.7: Unification of gauge couplings with and without SUSY. From top to bottom, the sets of lines plot the inverted strength of the interactions related to the B , $W^{1,2,3}$ and gluon fields (prior to EWSB) both for the SM (grey dashed lines) and for the MSSM (coloured lines) as a function of energy. For the MSSM case, the value of the strong coupling at the Z -boson mass is varied between 0.117 (red line) and 0.121 (blue line). Figure taken from [29].

final motivation, it is worth noting that SUSY is often a key component of quantum theories of gravity.

1.3.3 Searching for Supersymmetric Particles

The new particles introduced by SUSY could be produced at particle colliders such as the Large Hadron Collider (LHC) and could be detected from a wide variety of experimental signatures. Many searches for supersymmetric particles have already been performed, at the LHC with the ATLAS and CMS experiments, by the experiments at the Large Electron Positron (LEP) collider, which was housed the same ring in which the LHC now resides, and at the Tevatron collider at the Fermi National Accelerator Laboratory in the USA. A general outline of how such searches are performed can be found in Chapter 3. So far, none of these searches have yielded any concrete evidence for the existence of Supersymmetric particles, but they have been able to exclude particular Supersymmetric models.

In particular, individual searches are typically designed using simplified models, in which only a couple of the new SUSY particles have masses within experimental reach and the branching fractions of certain decay modes are fixed. Although such models do not represent ‘realistic’ models of SUSY, as argued in [40] they capture the main kinematics of the processes they target such that the searches which use them have good sensitivity to more complex models that include the target processes. Further, in general the cross-sections of the target production modes in the simplified models are in general not vastly different than they are in more complex models (although this does depend greatly on the production mode), hence the restrictions on sparticle masses that can be set using the simplified models can, in some cases, be interpreted as absolute limits on the sparticle masses. However, the assumptions of the simplified models should always be considered when interpreting the results of searches.

The phenomenological MSSM (pMSSM) [41] provides a more general framework with which to interpret the search results by imposing certain restrictions on the MSSM such that the number of free parameters is reduced from over 100 to 19. The following restrictions on top of the MSSM conditions are applied:

- No flavour changing neutral currents (as these are heavily constrained by numerous experiments).
- No new sources of CP-violation (to avoid models which violate experimental constraints on the electric dipole moments of the electron and neutron, and on the Kaon system).
- Universality between the first and second generations of squarks and fermions (to avoid models which do not satisfy experimental constraints from neutral Kaon mixing).

The results of the individual searches are re-interpreted for models satisfying the pMSSM criteria from which the fraction of models with certain characteristics that are excluded can be computed (for instance the fraction of models with a $\tilde{\chi}_1^0$ within a certain mass range). Additional constraints, such as those arising from DM models or the Higgs boson mass, can also be applied. The results from the ATLAS [36] and CMS [42] pMSSM interpretations using the data collected at the LHC between 2010-2013, show the significant impact these searches have in constraining these SUSY models. Nevertheless, many (natural) SUSY models were not, and have yet to be, excluded.

The searches discussed in this thesis are for electroweak SUSY production, in which the new SUSY particles are produced via W -bosons, Z -bosons or photons. More detail on the specific processes can be found in the corresponding chapters, but for now it is useful to note that the constraints on squark and gluino masses are much larger than the constraints on sfermions and electroweakinos (and consequently the amount of SUSY parameter space with light sfermions and electroweakinos that is excluded is much less than the fraction of excluded with comparatively light squarks or gluinos). This is primarily due to much larger production cross-sections at the LHC for squarks and gluinos. While a proton-proton collider has disadvantages for electroweak SUSY searches compared to a lepton-lepton collider, the larger centre of mass energy, large amount of data collected, and sophisticated analysis techniques that have been developed mean that searches being conducted by ATLAS and CMS at the LHC are able to probe previously unexplored regions of the electroweak parameter space, as will be seen in the latter three chapters of this thesis.

Chapter 2

The ATLAS Experiment at the Large Hadron Collider

“Theory will take you only so far.”

J. Robert Oppenheimer,
Oppenheimer (2023)

To test for possible extensions to the Standard Model (SM) like Supersymmetry, it is necessary to construct particle colliders capable of producing such particles. Located 100m underground on the Swiss-French border, the Large Hadron Collider (LHC) at CERN is currently the largest and highest energy particle collider in the world, capable of colliding protons together at centre of mass energies up to 14 TeV. Since coming online in 2010, a rich array of physics has been studied by the detector experiments at the LHC, including the discovery of the Higgs boson by the ATLAS and CMS experiments, precision measurements of rare SM processes and searches for physics beyond the Standard Model (BSM). This chapter first provides some background information on the LHC and proton-proton collisions in Section 2.1 followed by a detailed look at the ATLAS experiment in Section 2.2, with which the searches presented in this thesis have been performed.

2.1 The Large Hadron Collider

2.1.1 Overview

The LHC is a circular collider; particles lap a ring, 27 km in circumference, many times as they are accelerated. As mentioned at the end of the previous chapter, the Large Electron Positron collider (LEP) experiment was previously located in the same ring. The LHC is primarily used to collide protons at very high energies, but can also collide heavier ions at lower energies. During proton collisions, protons travel around the ring in bunches, one beam of bunches travelling clockwise the other anticlockwise. Each bunch consists of $\mathcal{O}(10^{11})$ protons. There are four points where these beams cross each other, resulting in collisions. Detector experiments are located at each of these interaction points (IPs). Of the four primary detector experiments two, ATLAS (A Toroidal LHC Apparatus) [43] and CMS (Compact Muon Solenoid) [44], are general purpose detectors, designed to study a range of phenomena while LHCb (LHC Beauty) [45] is a forward detector optimized for the study of b -quark related process and ALICE (A Large Ion Collider Experiment) [46] is focused on analyzing the aforementioned heavy-ion collisions to study quark-gluon plasma.

Protons are only collided in the LHC during ‘Runs’ with time in between Runs dedicated to upgrade and maintenance work. Run 1 of the LHC took place from 2010-2013, Run 2 from 2015-2018 and Run 3 commenced in July 2022. In Run 1, protons were collided at centre of mass energies of 7-8 TeV, while during Run 2 the energy was increased to 13 TeV. For Run 3, the collision energy has been further increased to 13.6 TeV. Beyond Run 3, significant upgrades will take place to increase the luminosity (discussed below) such that the LHC will become the High Luminosity LHC (HL-LHC). The analyses discussed in this thesis make use of the data collected by the ATLAS experiment during Run 2.

During Run 1 and 2, the protons in the beams were sourced from bottles of pressurised hydrogen. An electric field stripped hydrogen atoms of their electrons, after which they were first accelerated by a linear accelerator, Linac2, to energies of 50 MeV. For Run 3, Linac2 was replaced by another linear accelerator, Linac4, which instead accelerates negative hydrogen ions from a H⁻ ion source. The negative ions are stripped of their electrons prior to the next stage. Both Linac2 and Linac4 make use of radio frequency (RF) cavities to accelerate the protons/negative hydrogen ions. After the initial acceleration, the energy of the protons is further increased by a succession

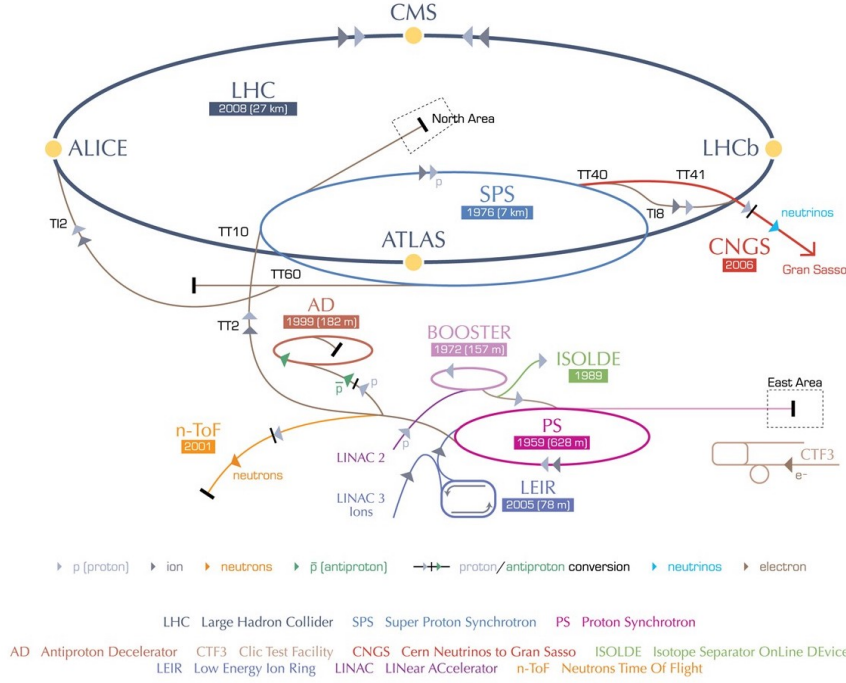


Figure 2.1: Schematic of the accelerators at CERN. Adapted from [47].

of three increasingly large synchrotrons (the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS)) until they reach an energy of 450 GeV, at which point they enter the main LHC ring. A schematic of these accelerators is shown in Figure 2.1. Each of the synchrotrons, and the main LHC ring, consist of RF cavities, which accelerate the protons, and various types of magnet to control the trajectory of the beams. The main dipoles generate magnetic fields of 8.3 T. Quadrupole, sextupole and octupole magnets are used to squeeze the beams prior to collisions, reducing the beam spot size and thus increasing the interaction rate. During Run 2, interactions occurred at intervals of roughly 25 ns. Whilst the beams of colliding, the interaction rate decreases over time as the protons in the bunches are lost (or depleted). Eventually (typically after ~ 12 hours during Run 2), the beams are dumped and a new fill of protons is later injected.

2.1.2 The Nature of Proton-Proton Collisions

Protons are one of the best candidates for exploratory colliders such as the LHC due to their relatively large mass (which keeps synchrotron radiation low), long (potentially infinite) lifetime and ease of sourcing. However, numerous complications arise from

protons not being fundamental particles (recall from Chapter 1 they are bound states of quarks and gluons). An immediate consequence of this is that the centre of mass energy of the hard scatter, the specific high energy parton-parton interaction of interest, is not known precisely as the total energy of the protons is distributed among their partons. Another consequence is that there can be additional processes on top of the hard scatter. An illustration of a proton-proton collision is shown in Figure 2.2. The complication this causes is compounded at the LHC where multiple simultaneous proton-proton interactions occur each time the bunches cross. In such events, a primary vertex has to be defined and objects from the other vertices are known as pile-up objects. During Run 2 of the LHC, there was an average of number of 34 interactions per bunch crossing at the ATLAS IP [48]. This is expected to increase to around 50 for Run 3 and to 200 at the HL-LHC.

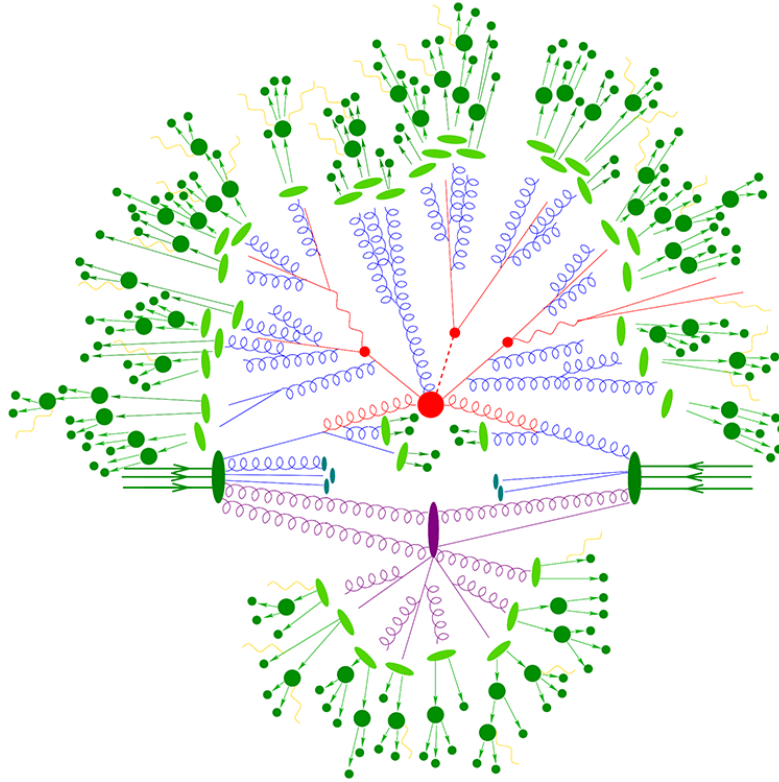


Figure 2.2: An illustration of a typical proton-proton collision. In red is highlighted the high energy parton-parton interaction of interest but other partons also interact. Image taken from [49].

What happens in each proton-proton collision is governed by the underlying Quantum Field Theory and so is inherently probabilistic (at least from a practical perspec-

tive). Many new physics and interesting SM processes are or are expected/constrained to be much rarer than the majority of SM processes that occur at the LHC, particularly generic strong interactions between partons. Consequently, most events are not interesting to analyze and, even if they were, it would not be practically possible to record all the information from every event. To avoid writing out all events to disk, ATLAS and the other experiments at CERN design algorithms that determine whether events are interesting for analysis (based on what searches or measurements are desired) and only such events are written to disk. These algorithms and selections are known as triggers. An example trigger might require an event to have two electrons with large momenta. The trigger system of the ATLAS experiment is discussed later in this chapter in Section 2.2.3.

2.1.3 Luminosity

As will be seen in the next chapter, searches for new physics and precision SM measurements benefit from having a larger total number of collisions (i.e. more data) to analyze. For a particular physics process i , the rate of that process occurring is given by

$$\frac{dN_i}{dt} = \sigma_i \cdot \mathcal{L}, \quad (2.1)$$

where σ_i is the cross-section for the process (a measure of its probability to occur) and $\mathcal{L}$ is the (instantaneous) luminosity, a property of the collider. A useful measure of the amount of data delivered to or recorded by a particle detector over a given period of time is the integrated luminosity L :

$$L = \int \mathcal{L} dt. \quad (2.2)$$

Luminosity depends upon several factors; the number of proton bunches in each beam which collide when the beams crossed, how tightly squeezed the proton bunches are, the angle at which the bunches cross, the transverse offset between the beams and the number of protons per bunch. These factors can be optimized to give a high luminosity whilst still ensuring the particle detectors have the necessary bandwidth to process the high collision rate.

Nominally, the instantaneous luminosity from a single fill (also called a run) decreases exponentially once collisions have commenced as protons in the beam are either depleted in collisions or are lost due to experimental effects. With this luminosity profile, increasing the total integrated luminosity requires increasing the maximum or peak instantaneous luminosity, which can put more severe demands on the detector components and event reconstruction algorithms, as a high peak instantaneous luminosity means a high peak number of pile-up events. As a more efficient alternative to this exponential profile, luminosity levelling can be employed whereby the initial instantaneous luminosity is reduced from its possible maximum but the beam parameters are adjusted during the run to maintain this instantaneous luminosity for a certain period. Luminosity levelling by adjusting the separation of the beams was used at the ATLAS and CMS IPs towards the end of Run 2, and has been used since Run 1 at the LHCb and ALICE IPs. For Run 3, luminosity levelling will be achieved at the ATLAS and CMS IPs by increasing the amount by which the beams are squeezed during a run, a technique known as β^* levelling [50].

Measuring the detector luminosity and total integrated luminosity accurately is of fundamental importance for physics analysis. In ATLAS, the primary luminosity measurement is from a hit-counting method using the LUCID2 sub-detector. This measurement is complemented by other measurements based on diamond sensors in the Inner Detector, track counting, and from electrical currents in the central Tile and Electromagnetic end-cap calorimeters (see Section 2.2.2 for details of each of these sub-detectors), which allow the evaluation of a long term stability uncertainty. These methods require calibration from special low-luminosity runs in order to determine the absolute luminosity (as opposed to the measured visible luminosity, which is method dependent). The analyses discussed in this thesis use the dataset collected by the ATLAS experiment during Run 2 which has a total integrated luminosity measurement of 139 fb^{-1} , which corresponds to over 10^{16} proton-proton collisions. The total uncertainty on the integrated luminosity used in the analyses in this thesis is taken as 1.7% [51], although a more recent measurement has reduced this uncertainty to 0.83% [52].

2.2 The ATLAS Experiment

Having reviewed some key aspects of the LHC, attention now turns to the ATLAS experiment. It is important to appreciate that the ATLAS experiment consists not only of the ATLAS detector, but also the software and computing infrastructure that facilitate the analysis of the data collected and the 6000 scientists, engineers and support staff from 42 countries which form the ATLAS collaboration. This section is structured as follows; an overview of the ATLAS detector including the coordinate system is given in Section 2.2.1. The different sub-detectors comprising the full ATLAS detector are described in turn in Section 2.2.2 which is followed by a discussion of the ATLAS trigger system in Section 2.2.3. Section 2.2.4 then outlines how the information recorded by the sub-detectors is used to reconstruct particles for physics analysis. Lastly, Section 2.2.5 will briefly describe how the response of the ATLAS detector is simulated as this is often used for making predictions in physics analyses. The technical information in this section is taken from the ATLAS summary paper [43] except where otherwise indicated.

2.2.1 Overview

As mentioned, the ATLAS detector is a general purpose particle physics detector designed to facilitate the study of a wide variety of phenomena. ATLAS is (approximately) cylindrically symmetric and has nearly 4π of solid angle coverage. It is located in an underground cavern at Point-1 of the LHC complex. Measuring 44 m long and 25 m tall, it is the largest detector ever constructed for a particle collider.

The primary physical components of ATLAS are its sub-detectors and its magnet systems. Closest to the interaction point is the Inner Detector (ID), which registers hits, with fine granularity, when charged particles pass through it. Beyond this is a central solenoidal magnet which produces a magnetic field inside the ID, aligned along the beam direction, which bends the path of charged particles such that the charge and momenta of the particles can be measured from the curvature of tracks reconstructed from the hits that are registered. Past the solenoid are the electromagnetic and hadronic calorimeters (ECALs and HCALs respectively), which record energy deposits of photons/electrons and hadrons respectively. The calorimeters also serve to stop the passage of these particles. The last layer of ATLAS is the muon spectrometer (MS), which consists of various measurement chambers and toroidal magnets to

precisely identify and measure the properties of muons. Figure 2.3 shows a cutaway of ATLAS for reference labelling the key components. Additionally, there are several detectors located in the forward regions of ATLAS (region close to the ends of the ATLAS ‘cylinder’) which serve special functions that will be briefly discussed in the next section.

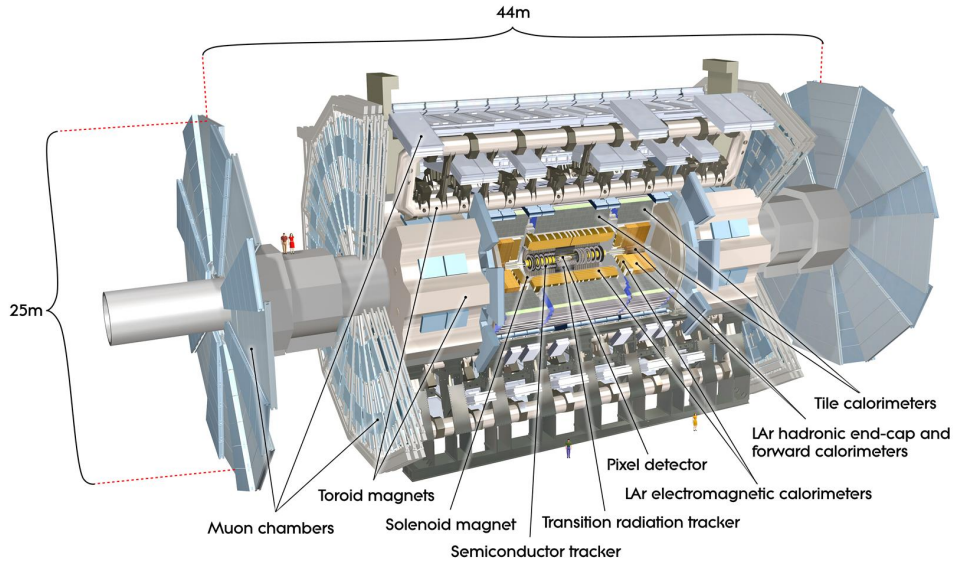


Figure 2.3: Cutaway of the ATLAS detector [43].

Coordinates

Conventionally, a right-handed set of axes is defined with the x -axis pointing towards the centre of the LHC ring, the y -axis directed vertically upwards and the z -axis running along the beam direction (the axis of the ATLAS cylinder). Cylindrical coordinates are used in the transverse (x - y) plane with R being the perpendicular distance from the interaction point and ϕ being the azimuthal angle. Transverse quantities, such as transverse momenta, p_T , refer to this plane. Pseudorapidity is defined as

$$\eta = -\ln \tan(\theta/2), \quad (2.3)$$

where θ is the polar angle from the z -axis. A pseudorapidity of zero corresponds to a particle emitted in the transverse plane while pseudorapidities with magnitude greater than 2.5 are within 10° of the z -axis. For massless particles, the pseudorapidity

is equivalent to the rapidity, y , which is defined as

$$y = \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right), \quad (2.4)$$

where E is the particles energy and p_z is the momentum in the z -direction. For highly relativistic particles, pseudorapidity is approximately equal to rapidity but has the advantage θ is more directly related to the detector than p_z and so is more often used in physics analyses. Pseudorapidity is used over θ at the LHC because *differences* in rapidity (and therefore also pseudorapidity for highly relativistic particles) are invariant under Lorentz boosts in the z -direction; which is useful as in proton-proton collisions the longitudinal momentum of the hard scatter is not known. The separation between two particles in η - ϕ space, ΔR , defined as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}, \quad (2.5)$$

is another quantity often used in physics analyses.

2.2.2 Sub-Detectors

Inner Detector

The Inner Detector (ID) is designed to record the tracks of charged particles which can be used to; identify the particles; measure their momenta; and, in events with multiple interaction vertices, determine which vertex the particles originated from. As mentioned, the ID is the closest part of ATLAS to the interaction point; coming within 33 mm of the colliding proton beams. The ID has three distinct parts; working outwards from the beam these are the Pixel Detector (PD), Semi Conductor Tracker (SCT) and the Transition Radiation Tracker (TRT). Each of these components are divided into central barrel parts and end-cap disks. In total the ID extends 1.08 m radially and 6.2 m along the z -direction. A cross-section of the ID is shown in Figure 2.4. The central solenoid magnet which lies just beyond the ID consists of over 9 km of niobium-titanium superconductor wire embedded into strengthened aluminium strips, resulting in a 2 T magnetic field inside the ID. Important to note is that the ID only has coverage for $|\eta| < 2.5$ (with full coverage in ϕ), meaning the tracking information it provides is only available in this central region.

The PD and SCT are both solid state silicon detectors. They consist of small pieces of doped silicon which are reversed bias during operation to establish an electric field. When a charged particle passes through a piece of the silicon it creates electron-hole pairs which are collected at electrodes, registering a hit. The path of the charged particle can then be constructed from the location of these hits. Additionally, the track information can be used to identify longer lived particles (both SM particles, such as tau-leptons or B-hadrons, but potentially also long-lived BSM particles) which have track origins that are displaced from the interaction point.

The PD consists of over 90 million individual silicon pixels that are arranged in four concentric layers in the barrel region and three discs at each of the end caps. The innermost layer of the PD, the Insertable B-Layer (IBL) was added between Run 1 and Run 2. This was as an alternative to replacing the innermost layer of the PD (which would have needed to be replaced due to radiation damage) while also improving the overall tracking performance. Each pixel measures $50\text{ }\mu\text{m} \times 400\text{ }\mu\text{m}$ in the $R\phi \times z$ directions in the three outer layers and $50\text{ }\mu\text{m} \times 250\text{ }\mu\text{m}$ in the IBL, resulting in a hit resolution of $12\text{ }\mu\text{m}$ and $\sim 70\text{ }\mu\text{m}$ in the $R\phi$ and z directions respectively. The larger SCT has a lower resolution of $17\text{ }\mu\text{m} \times 570\text{ }\mu\text{m}$ in $R\phi \times z$ but covers a larger area, consisting of over 4000 modules across three barrel layers and 18 end-cap disks. Each module has silicon micro-strips on each side to register hits, with over 6 million micro-strips being used in total.

The TRT is made up of 300,000 drift tubes or straws, each 4 mm in diameter, interleaved with polypropylene radiators. These straws are filled with a Xenon-based gas mixture and have a gold-plated tungsten wire at their centre. Kapton, a polyimide film, is used for the outer walls of the straws. When a charged particle passes through the straw, the gas is ionized. A 1.5 kV potential difference is maintained between the tungsten wire and the straw walls such that this ionization can be detected, and registered as a hit from which tracks can be reconstructed like in the PD and SCT. Measuring the number of hits above a given threshold across multiple straws provides a method for particle identification for particles with momentum less than 10 GeV. Additionally, when highly-relativistic charged particles pass through the radiators they emit transition radiation in the form of X-rays (due to passing through materials with different refractive indices). This transition radiation also ionises the Xenon gas mixture, but to a much greater degree than the passage of a charged particle does. The amount of transition radiation depends upon the ratio of a particle's energy and mass, such that it can be used as a method of particle identification, with typically only

electrons being produced with sufficiently large energy to mass ratio to give significant transition radiation [53].

The ID is set to be replaced by the Inner Tracker (ITk), a more advanced silicon detector, after Run 3. ITk will be a significant upgrade in resolution from the ID, which will be necessary to maintain good tracking performance with the larger pile-up conditions that will be present at the HL-LHC. The replacement is also necessary as the radiation damage the ID will have sustained after Run 3 will have severely degraded its performance.

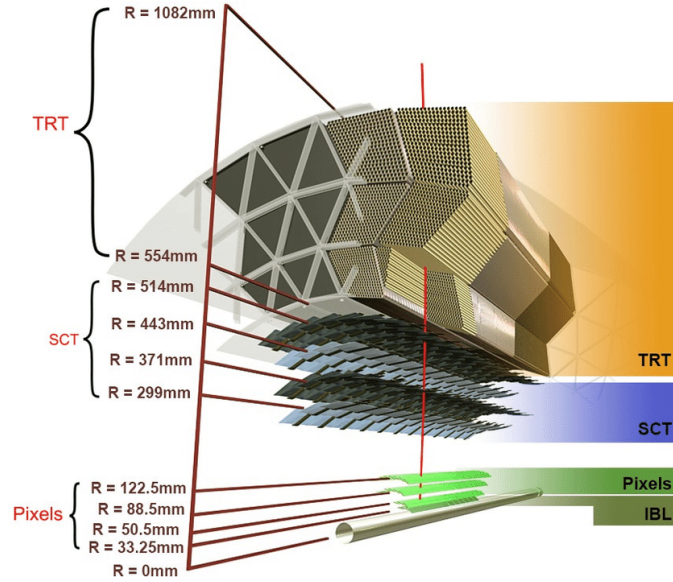


Figure 2.4: Cutaway of the Inner Detector, taken from [54]. R is the radial distance from the interaction point.

Calorimeters

Beyond the ID are several calorimeters, of both electromagnetic (EM) and hadronic variety. In particular there is a central barrel Liquid Argon (LAr) EM calorimeter, a hadronic barrel tile calorimeter, separate EM and hadronic LAr end-caps and a forward calorimeter (FCAL) with both EM and hadronic layers. A layout of these calorimeters is shown in Figure 2.5. Note there is a gap between the barrel and end-cap calorimeters covering ($1.37 < |\eta| < 1.52$).

Concerning the basic operation of calorimeters, when particles interact with matter they can produce cascades of additional particles known as a shower. Calorimeters

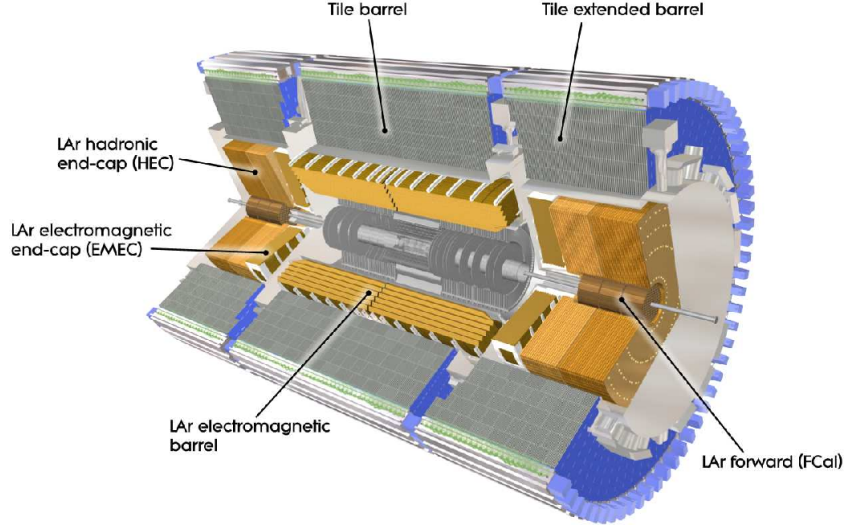


Figure 2.5: Layout of the ATLAS calorimeters, from [43].

can be used to measure the energy of incoming particles by absorbing these showers and the particles themselves. All the calorimeters in ATLAS are sampling calorimeters; they consist of interwoven layers of a passive material, which both causes the particles to shower and absorbs the particles in the shower, and an active material that facilitates measuring the energy of the showers. Electrons and photons produce electromagnetic showers which are absorbed by the electromagnetic calorimeters (ECALs). Hadrons also produce activity in the ECALs but are much more penetrating than electrons and photons due to their larger nuclear interaction lengths and so usually pass into the HCALs. Muons have even longer interaction lengths in these materials and so pass through the calorimeters without losing much energy. Having small interaction cross-sections, neutrinos pass through the calorimeters without depositing any energy.

An important property of calorimeters is their energy resolution, which is usually parameterized as

$$\frac{\Delta E}{E} = \frac{c_s}{\sqrt{E}} \oplus \frac{c_n}{E} \oplus c_c, \quad (2.6)$$

in which E is the shower energy, c_x are constants and the ' $\oplus$ ' symbol denotes that the terms are summed in quadrature. The first term is a statistical error arising from the stochastic nature of showers; the second term is a consequence of electronic noise and the final term accounts for inhomogeneities in the sampling material, radiation damage and calibration uncertainties. The energy resolution of the ATLAS calorimeters is determined from tests using electron beams (of varying energy). For all the ATLAS

calorimeters, the stochastic term is the dominant contribution to the resolution, with the later two terms often being combined in quadrature.

Examining the different calorimeters in detail, the barrel and end-caps ECALs have lead as the passive material and LAr as the active material, the latter of which requires the calorimeters to be cooled to 88K [43]. Charged particles ionise the LAr, producing measurable electric currents. The lead absorbers are arranged in an accordion pattern, with LAr filling the cavities, in order to achieve full coverage in ϕ . There is a thin (~ 11 mm) LAr pre-sampler between the ID and barrel ECAL to correct for fluctuations in the energy loss of electrons and photons from their travel through the ID and central solenoid. In total, the ECAL layers corresponds to more than 22 electron/photon interaction lengths. Following the parameterization of Equation 2.6, the ECAL has a stochastic constant $c_s = 10\% \sqrt{E}$, a noise constant of $c_n = 200$ MeV and a calibration constant $c_c = 0.2\%$.

Moving to the HCALs, the barrel Tile HCAL consists of a main segment and two extended segments, all of which use steel plates as the passive material and plastic scintillators for measurement. Photons produced by the scintillators are transmitted by wavelength shifting fibres to photomultiplier tubes to measure the signals. The barrel HCAL is significantly larger than the barrel ECAL as a consequence of hadronic showers being on average much larger than EM ones. The HCAL end-caps are subject to higher levels of radiation, predominantly from proton remnants, than the barrel component and so have LAr as the active material and radiation hardened copper as the passive material. The HCAL has poorer resolution than the ECAL, a consequence of its larger size meaning that achieving the same resolution would be significantly more expensive. The resolution parameters of the barrel HCAL are $c_s = 50\% \sqrt{E}$, $c_n = 1.6$ GeV and $c_c = 3\%$ [55].

The last calorimeter system is the FCAL, which provides coverage for $3.2 < |\eta| < 4.9$. This high η coverage is important for many physics signatures, in particular the measurement of missing transverse energy (see Section 2.2.4) and vector boson fusion signals, such as the one studied in Chapter 5. Three layers comprise the forward calorimeter; each layer using LAr as the active material, with copper and tungsten being used as the passive material in the first and outer two layers respectively. The first layer acts primarily as an EM calorimeter while the outer two layers act as hadronic calorimeters. The energy resolution terms for EM showers in the FCAL are

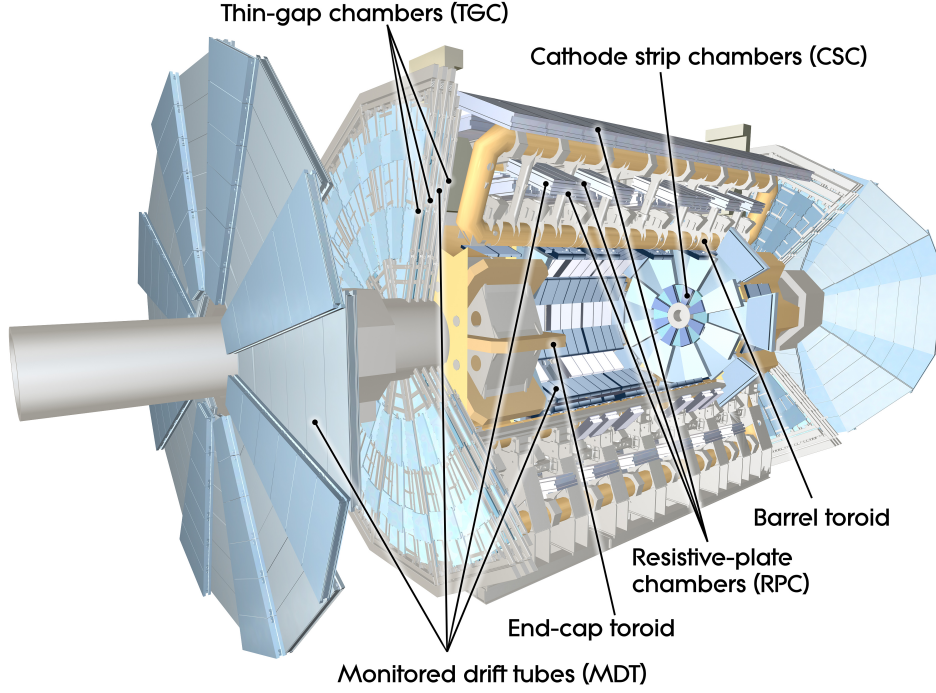


Figure 2.6: Diagram of the Muon Spectrometer, taken from [43,58].

$c_s = 25\%\sqrt{E}$, $c_n \approx 200 \text{ MeV}$, $c_c = 4\%$ while for hadronic showers they are $c_s = 100\%\sqrt{E}$, $c_n \approx 500 \text{ MeV}$, $c_c = 10\%$ [56,57]¹

Muon Spectrometer

The outermost part of the ATLAS Detector is the MS, which, as the name suggests, is dedicated to detecting muons and measuring their properties. Like the ID, the MS is subject to a magnetic field to deflect the path of muons. The magnetic field is produced by a combination of eight barrel toroidal magnets (toroids) and two end cap toroids which produce a magnetic field which is mostly orthogonal to the particle trajectories. The strength of the field varies from 0.5 T from the barrel toroids for $|\eta| < 1.6$ to 1 T from the end cap toroids for $1.6 < |\eta| < 2.7$. Four different detector technologies are used in the MS; Monitored Drift Tubes (MDTs) and Cathodes Strip Chambers (CSCs) for precise measurements of muon tracks and Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) for use in muon triggers. Figure 2.6 shows the layout of the MS, highlighting where each of these technologies is used.

¹ For the FCAL, the EM shower noise term is taken to be that of the first layer of the FCAL, while for hadronic showers it taken to be that of the second and third layers.

MDTs are used to provide the tracking information in the barrel region, where they are arranged in three concentric layers, and for all but the innermost layer of the end caps, where CSCs are used instead. MDTs operate in the same way as the drift tubes in the TRT, each tube having a 30 mm diameter aluminium casing, filled with a (93/7) Ar/CO₂ gas mixture, and a 50 μ m tungsten-rhenium wire at their centre. CSCs are multi-wire proportional chambers (MWPCs) which consist of closely spaced anode wires crossed with cathodes strips contained in a (80/20) Ar/CO₂ gas. The wires point in the radial direction. Passing muons ionize the gas, giving an electrical signal. Tracks are reconstructed by interpolating the location of these signals. The forward region, $2.0 < |\eta| < 2.7$, sees higher particle fluxes and greater muon-track densities, with both effects being greater closer to the interaction point. CDCs provide a higher readout rate and greater resolution than MDTs, hence their use in the forward region for the first plane of the end cap.

The trigger chambers cover $|\eta| < 2.4$, with RPCs being used in the barrel ($|\eta| < 1.05$) region and TGCs in the end caps ($1.05 < |\eta| < 2.4$). Although not as precise as the MDTs or CSCs, the trigger chambers have much faster readout times such that they can be used for triggering. The RPCs feature two parallel electrode plates with an electric field of 3.9 kV/mm, filled with gas (C₂H₂F₄/Iso-C₄H₁₀/SF₆ (94.7/5/0.3)) which is ionized by the passage of muons. RPCs can be used to trigger muons with transverse momenta as low as 6 GeV. Like the CSCs, the TGCs are also MWPCs but differ in that the wire-to-cathode distance is smaller than the wire-to-wire distance, which is partly why the TGCs offer better timing resolution (the other reason being a stronger electric field). TGCs have a slower response time than RPCs but are better able to handle the higher particle rates in the forward region.

Forward Detectors

In addition to the main sub-detectors, there are also three specialized detectors located in the forward regions. The first of these is LUCID (LUMinosity Cherenkov Integrating Detector), specifically LUCID-2 [59], which as previously mentioned is used as the primary way luminosity is measured in ATLAS. LUCID consists of two modules which are situated on both sides of the interaction point at $z = \pm 17$ m. Each module consists of 16 primary photomultipliers close to the beam line which have quartz windows as the Cherenkov medium. The number of hits recorded from interacting protons is used to determine the luminosity.

The second forward experiment is ALFA (Absolute Luminosity For ATLAS) and is significantly further away from the interaction point, with modules at $z = \pm 240$ m. ALFA consists of Roman pots containing scintillating fibre trackers, which get to within 1 mm of the beam. The main utility of ALFA is to measure elastic-scattering amplitudes at high scattering angles (from which it is possible to determine the luminosity). ALFA is not used in normal LHC running conditions as it is radiation sensitive but is instead used in dedicated runs. Lastly, there are also Zero-Degree Calorimeters located at $z = \pm 140$ m which are designed to measure neutrons from heavy ion collisions at $|\eta| > 8.3$, which is important for determining the number of spectator neutrons in these collisions.

2.2.3 Trigger System

As previously mentioned, the majority of proton-proton collisions are not interesting for physics analysis so detector experiments have trigger systems which determine which events to record data for. The ATLAS trigger system consists of a Level-1 hardware trigger and a software based high level trigger (HLT). The Level-1 trigger uses reduced-granularity information the calorimeters (Level-1 Calo) and from the RPCs and TGCs in the muon spectrometers (Level-1 Muon). This information is processed by a central trigger processor (CTP) where Regions of Interest (ROIs), defined by a geometrical area and object properties, are identified. The CTP makes a first selection of which events to possibly keep based on these ROIs from a trigger menu. The ROIs are then passed on to the HLT which makes use of the full detector information to make more complicated selections, from a more complicated menu. Events that are accepted by the HLT are then recorded by a Read-Out system. During Run 2, the Level-1 trigger reduced the event rate from the nominal LHC 40 MHz to 100 kHz and the HLT further reduced this to 1 kHz.

The design of the triggers in the trigger menu must balance obtaining a large data yield for analysis with keeping trigger rates low enough to stay within a manageable bandwidth. To aid with this, triggers can be pre-scaled to reduce their acceptance rate. A pre-scale indicates how many events are rejected before the physics decision is applied; a trigger with a pre-scale of 10 only computes the physics decision for 1 in every 10 events that pass its criteria. For the analyses presented in this thesis, the triggers used had the lowest (often no) pre-scale during the Run 2 data taking period

of the available triggers. Details of the specific triggers can be found in the relevant chapters.

The trigger selections are based on the ‘online’ reconstruction of physics objects, which is faster but less accurate than the ‘offline’ reconstruction discussed in the next section. To ensure the objects used for passing the triggers are reconstructed, a step known as trigger-matching is usually employed during the offline analysis that requires offline objects (of the same kind) to be reconstructed close to the online objects used to pass the trigger, otherwise the event is vetoed. As a consequence of the difference between online and offline reconstruction, triggers in general also have an efficiency, defined as the fraction of events in which the offline objects meet the trigger criteria that pass the trigger requirements. These efficiencies are measured using a mixture of simulation and data-driven techniques and in general vary with the kinematics of the objects in question. In particular, triggers tend to have lower efficiency at low transverse momenta and have an efficiency plateau above a given threshold. In general, it is recommended that events used in physics analyses are on these trigger plateaus as this tends to result in better modelling from simulation and smaller uncertainties. This recommendation is followed in the analyses presented in this thesis.

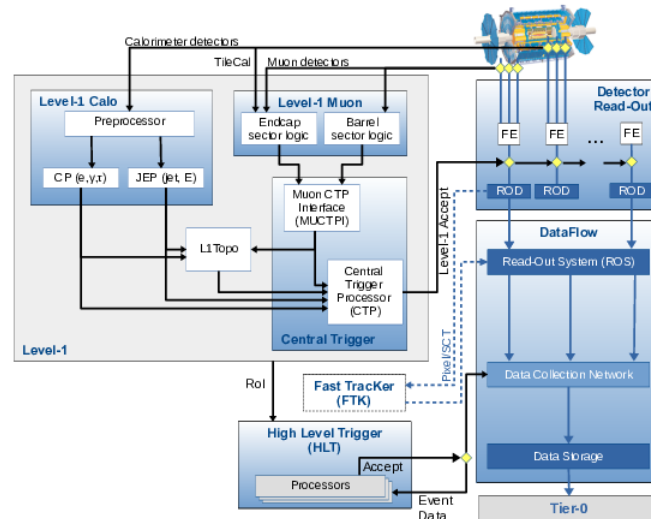


Figure 2.7: Schematic of the ATLAS trigger system during Run 2, taken from [60]

2.2.4 Object and Event Reconstruction

Once events have been selected and written to disk, to perform a physics analysis the information from the individual sub-detector components is used to reconstruct the physics objects in an event. In general all events are required to have a primary vertex; a vertex being a point in 3D space representing a single proton-proton interaction. For the analyses presented in this thesis, a primary vertex is required to have at least two associated charged-particle tracks with $p_T > 500$ MeV. The position of the vertex must also be consistent with the beam spot envelope. For events which have multiple vertices fulfilling these criteria, the primary vertex is identified as the vertex with highest Σp_T^2 of associated tracks. Reconstruction and identification of the particles originating from this primary vertex utilises the different signatures particles leave in the detector. An illustration of the general signatures, which are discussed below, is shown in Figure 2.8.

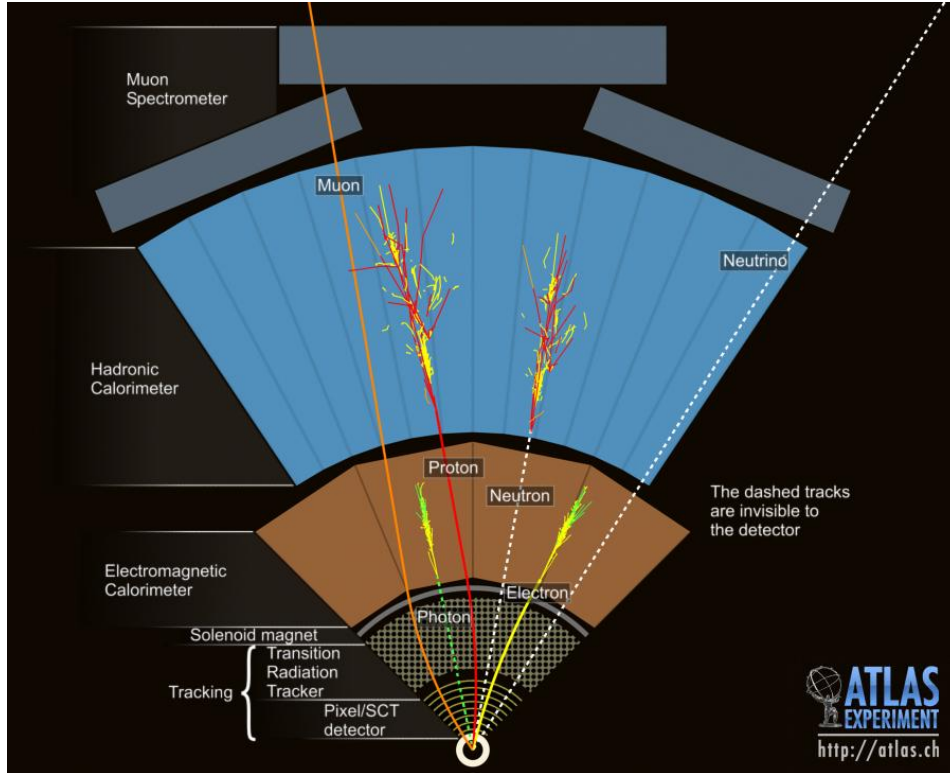


Figure 2.8: Signatures different Standard Model particles leave in the ATLAS detectors, taken from [61]. Note that protons and neutrons also deposit some energy in the electromagnetic calorimeter which is not depicted.

Along with general signature requirements, sophisticated identification (ID) algorithms have been developed to discriminate between *prompt* objects, originating from

the hard scatter, from fake and non-prompt objects that leave similar signatures. In many cases, sophisticated machine learning algorithms are used for reconstruction and identification. An overview of these algorithms is provided in the next chapter, in Section 3.2. Various working points (WPs) are defined for these algorithms which have different efficiencies, fake rates and uncertainties for the objects in question. The efficiencies and background rejection rates typically display a dependence on p_T and η .

Simulation of physics processes is very commonly used to predict possible signals and the SM background when performing physics analyses, hence it is necessary to validate that the reconstruction and ID algorithms perform similarly for simulation and for real data. Scale factors are computed to take into account where discrepancies arise, often using tag and probe based approaches which ensure that the object under consideration is not used for triggering. A general description of the reconstruction and ID algorithms used for the different physics objects is discussed below with specific details (such as which WPs are used) given in the relevant analysis chapters.

Jets

As mentioned in the previous chapter, hadronisation causes quarks and gluons produced in collisions to form collimated cones of hadrons called jets. The reconstructed jets approximate the energy and momentum of the original quark or gluon. Jets leave both tracks in the ID and topological energy deposits in the calorimeters, primarily the Tile HCALs, called topo-clusters. Forward jets, outside the region covered by the ID, can only be constructed from the calorimeter information. At the start of Run 2, jets were only reconstructed using the calorimeter energy deposits but this reconstruction method was replaced by a particle flow algorithm [62] which integrates both the tracking and calorimeter information. In both types of reconstruction, the anti- k_t algorithm [63] is used to construct the jets from the input information, typically using a radius parameter of $R=0.4$. Generally, only jets with $p_T > 20$ GeV are reconstructed and used for physics analysis. Vertex tagging algorithms [64] aim to separate jets originating from the hard scatter from those originating from pile-up, with a dedicated tagger for forward jets [65]. These algorithms exploit the fact that pile-up can be treated as noise.

It is necessary to calibrate the energy of jets as several effects would otherwise lead to significant differences between data and simulation. These effects include energy that may not be measured in the calorimeters due to leakage or inactive material and

pile-up effects which give additional energy deposits. The jet energy scale (JES) is defined as the difference between the true jet energy and the reconstructed energy divided by the true jet energy. The JES for the Run 2 data collected by ATLAS was calibrated using comparisons between data and simulation of dijet, multi-jet, Z+jets and γ +jets processes [66]. Similarly, the jet energy resolution (JER), which describes the spread of reconstructed energies about the true value, also needs to be calibrated. For the ATLAS Run 2 dataset this was done using dijet events in which the two jets are balanced against each other, allowing a precise measurement of their momenta. As jets are the most common object resulting from proton-proton collisions, a dedicated technique exists to subtract contributions to primary jets from those originating from pile-up. The method is based upon jet areas and provides event-by-event and jet-by-jet corrections [67]. Jets can also arise from detector noise or cosmic rays; events containing jets in which these origins are likely are removed [68].

The moderately long lifetime of b -hadrons ($\sim 10^{-12}$ s), means it is possible to identify jets which contain them from the displaced secondary vertex of the b -hadron decay (typical mean flight length of b -hadrons at the LHC is ~ 1 cm). Identifying these jets can be useful for physics analysis, for example in identifying jets originating from top quarks which decay primarily to a b -quark and a W -boson. This identification of these jets is the job of b -tagging algorithms; multivariate algorithms which make use of differences in the ID track parameters and possible secondary vertices for discrimination [69]. Jets passing the chosen working point of the b -tagging algorithm are known as b -tagged jets or b -jets for short. Analogous tagging algorithms have been developed for jets containing c -hadrons but do not perform as well due to the shorter lifetime of c -hadrons making such jets harder to distinguish [70]. Algorithms also exist to distinguish jets originating from quarks from jets originating from gluons [71], a task of even greater difficulty.

Electrons

Electrons are reconstructed from matching tracks in the ID to clusters of energy deposits in the ECALs. Upper bounds are placed on the ID track $|z_0 \sin \theta|$, where z_0 is the longitudinal impact parameter and θ is the polar angle, and on the significance of the transverse impact parameter, $|d_0 / \sigma(d_0)|$, to ensure tracks are close to the primary vertex and of good quality. Bremsstrahlung is taken into account in the fitting of the ID tracks. As for jets, energy calibrations of the clusters is necessary.

Electron ID algorithms are used to separate prompt electrons from hadronic jets, non-prompt electrons from hadron decays and electrons from converted photons (photons that originate from the primary vertex which pair produce electrons and positrons before reaching the ID). ECAL shower shapes, properties of the primary electron track and the matching of the track to the ECAL clusters are used for discrimination. The Electron ID algorithm used in Run 2 defined three working points, loose, medium and tight, with average acceptance efficiencies for prompt leptons of 93%, 88% and 80% respectively [72]. To further suppress fake and non-prompt backgrounds, isolation requirements are placed on electron candidates which require there to be limited ID tracks and calorimeter energy deposits surrounding them. At present, electrons can be accurately reconstructed and identified for $p_T > 4.5$ GeV.

Muons

The reconstruction of muons is based primarily upon matching ID tracks to tracks reconstructed from hits in the MS sub-systems, though calorimeter energy deposits are also used in regions for which the MS has reduced acceptance (due to services and cabling for the inner sub-detectors). As for electrons, requirements are placed on the transverse and longitudinal impact parameters. Muon ID algorithms serve to reject fake or non-prompt muons, the latter of which mainly originate from pion and kaon decays. The quality of the track fits is used a primary discriminating variable, with non-prompt tracks displaying a characteristic kink which worsen the fit quality. Loose, medium and tight WPs are defined this time with efficiencies of 96.7%, 95.5% and 89.9% respectively for muons with $4 < p_T < 20$ GeV [73].

Dedicated WPs exist for low- p_T muons, which may not reach the MS due to energy losses in the calorimeters, that allows reconstruction down to p_T as low as 3 GeV. Additionally, a high- p_T WP is defined which has lower reconstruction efficiency but greater momentum resolution for muons with $p_T > 100$ GeV. Isolation criteria are also placed on muons to further separate prompt and non-prompt muons from semi-leptonic hadron decays, with the latter often being embedded in a jet. Isolation WPs are defined by selections on track and calorimeter based isolation variables.

Tau-leptons

Tau-leptons can decay leptonically (to electrons or muons) or hadronically (to quarks) with the hadronic branching fraction being 65%. For the leptonic case, the daughter electron or muon is reconstructed as a prompt electron or muon following the descriptions given above. The hadronically decaying case can be further split into two main subsets, 1-prong and 3-prong decays depending on the number of charged mesons (predominantly pions) in the decay products, with 1-prong and 3-prong decays accounting for 72% and 22% of hadronic tau-lepton decays respectively. The signature of these decays is therefore a narrow jet cone with one or three associated tracks in the ID. Hadronic tau-lepton reconstruction is seeded by the jet reconstruction; jets with $p_T > 10$ GeV and $|\eta| < 2.5$ that satisfy these track requirements are considered as tau-lepton candidates [74].

Like jets, it is necessary to calibrate the tau-lepton energy scale (TES). The baseline method starts with the local hadronic calibrations from the JES calibration and then has additional corrections for pile up, treated separately the 1-prong and 3-prong cases. The energy resolution for this method significantly degrades at low p_T so a Tau-lepton Particle Flow (TPF) method, utilising reconstruction of individual hadrons, was developed to improve the resolution at low p_T [75]. The TPF and baseline methods are combined, together with additional tracking and calorimeter information, using a multivariate Boosted Regression Tree (BRT).

On top of the reconstruction algorithms, Tau-ID algorithms serve to reject jets that fake hadronically decaying tau-leptons. Two types of Tau-ID algorithms were developed in ATLAS during Run 2, both multivariate techniques utilising calorimeter and track based variables. In both cases, separate algorithms were trained for the 1-prong and 3-prong cases. Boosted Decision Trees (BDTs) were used up until shortly after the end of Run 2 upon which they were succeeded by Recurrent Neural Network (RNN) based algorithms which have significantly greater background rejection [76]. Both algorithms were trained using $Z/\gamma^* \rightarrow \tau\tau$ simulation as the signal and dijet events from data as the background.

In addition to jets, electrons can also fake the 1-prong hadronic tau-lepton signature. At the start of Run 2, this contamination was first reduced by considering overlapping electron and tau candidates. If the candidates were separated by $\Delta R < 0.4$ and the electron candidate passed at least the very loose electron ID WP, then the (1-prong) tau-lepton was rejected. This approach was succeeded by a dedicated BDT developed

to reduce this fake background with improved performance. For Run 2 analyses, hadronic taus can be accurately reconstructed and identified for $p_T > 20$ GeV.

Photons

The reconstruction of photons primarily uses topo-clusters in the ECALs and thus shares many similarities with electron reconstruction. Unlike electrons, unconverted photons have no associated tracks in the ID while converted photons can be matched to conversion vertices consisting of pairs of electron and positron tracks. Photon ID is designed to select prompt, isolated photons and uses a cut-based selection on shower shape variables to reject backgrounds from hadronic jets. Dedicated steps are taken to resolve ambiguities between electron and photon candidates. As for electrons, isolation criteria are also applied. Loose, Medium and Tight WPs are defined with Tight serving as the primary selection. For transverse energies above 70 GeV, the ID efficiency reaches 92% for unconverted photons and 98% for converted photons [77]. Photons are able to be reconstructed for $|\eta| < 2.37$ and transverse energies as low as 10 GeV.

Overlap Removal

The aforementioned reconstruction algorithms are run independently of each other; consequently detector signatures may be considered as part of different physics objects simultaneously. To avoid double counting, an overlap removal is performed where-by pairs of physics objects that are geometrically close to each other in the detector are identified and only one of the objects in each pair is subsequently used in the analysis. Details of the overlap removal steps taken for the analyses discussed in this thesis are given in the relevant chapters.

Missing Transverse Momentum

Although as previously discussed the total centre of mass energy of the hard scatter is not known in a hadron collider, what is known is that the total transverse momentum of the colliding partons prior to the event is zero (to a very good approximation) and so by conservation of momentum the vector sum of the transverse momenta of the physics objects after the collision should also be zero. However, not all physics objects leave

visible traces in the detector, for example neutrinos pass through without interacting. The reconstruction of physics objects is also not perfectly accurate. These effects give rise to missing transverse momentum which, without mismeasurement effects, would be equal to the sum of the transverse momenta of the invisible particles in the event. The missing transverse momentum is a three vector, $\vec{p}_T^{\text{miss}}$, with magnitude E_T^{miss} , known as the missing transverse energy, which is computed for each event as

$$\vec{p}_T^{\text{miss}} = -\sum_i \vec{p}_T^i - \vec{p}_T^{\text{soft}}, \quad (2.7)$$

where the sum in the first term, known as the hard-term, is over the reconstructed physics objects and the second term, known as the soft-term, is computed from any remaining tracks and jet candidates with $p_T < 20$ GeV that point to the primary vertex [78]. The sum is performed prior to overlap removal, but in a specific order (electrons, photons, hadronic taus, muons, jets) with candidates being rejected if they share a calorimeter signal with an object previously included in the sum. Being primarily reconstructed from ID and MS tracks, there is a special overlap treatment for non-isolated muons (muons with calorimeter signals). The x and y components of the $\vec{p}_T^{\text{miss}}$ are computed separately and combined vectorially to give the overall E_T^{miss} and ϕ angle. Similar to the hard objects, different working points are defined for $\vec{p}_T^{\text{miss}}$ which place requirements on the jets used in the calculation.

Missing transverse momentum is an important signature for many beyond the Standard Model physics scenarios; the only trace many dark matter candidates would leave in the ATLAS detector would be through contributing to the E_T^{miss} . R-parity conserving SUSY signatures in which the lightest SUSY particle is electrically neutral also often result in large E_T^{miss} , as will be seen in the analyses presented in later chapters.

In addition to the E_T^{miss} , the significance of the E_T^{miss} can be calculated to better identify events with real E_T^{miss} from those in which the E_T^{miss} is a result of mismeasurement [79]. An object based E_T^{miss} significance is computed as the log-likelihood ratio of the hypothesis that the transverse momenta of the invisible particles is zero to the hypothesis that it is different from zero. It can be written as

$$E_T^{\text{miss}} \text{significance} = \frac{E_T^{\text{miss}}}{\sqrt{\sigma_L^2(1 - \rho_{LT}^2)}}, \quad (2.8)$$

where σ_L is the resolution in the longitudinal direction (direction parallel to the E_T^{miss}) and ρ_{LT} is the correlation factor between the longitudinal and transverse measurements. Events which are likely to have real E_T^{miss} have a large E_T^{miss} significance.

Data Quality Requirements

Only data recorded during well understood detector conditions should be used for analysis. This leads to several data quality requirements needing to be imposed to ensure that the information originating from each of the detector sub-systems is complete and reliable. In addition, ‘event-cleaning’ is employed to remove events in which physics objects look to be located in problematic regions of the detector, or that are otherwise incomplete. For the data collected during Run 2, 95.6% of the data collected by ATLAS passed these criteria [80].

2.2.5 Simulation

As will be discussed in the next chapter, predictions of the rate of particular types of events are needed in order to perform physics analyses. Often these predictions use simulations of the physics processes, which aim to capture the complexity of the underlying particle interactions and the interactions of the produced particles with the ATLAS detector. Simulation first involves producing events with Monte Carlo (MC) event generators which aim to model the hard scatter of the physics process in question as well as the hadronisation of quarks and gluons. These different steps may be handled by different software packages. The interactions of other partons outside the hard scatter, referred to as the ‘underlying event’, are also modelled, which requires using the use of parton distribution functions (PDFs) mentioned in the previous chapter that model the kinematics of the quarks and gluons in the protons. These PDFs are tuned to better match results from data in dedicated studies (the original PDFs themselves are also largely determined from experimental data).

Generally, large numbers of MC events are generated (more than the relevant expected data yield) which are then weighted such that the sum of the MC event weights gives the expected yield for a given process. From the generation alone, the individual MC events in general do not have the same weight, as corrections and re-weighting (based on comparisons with data and/or theoretical calculations) may need to be applied which depend upon the event kinematics. It should be noted that

it is possible for individual MC events to have negative weights, but predictions of negative total yields are considered unphysical. Hence, when designing analyses care should be taken that such events are not separated from the positively weighted events they are designed to be (partially) cancelled with. Effects of pile-up also need to be accounted for; events are usually re-weighted such that the pile-up conditions match that of the relevant data-taking period.

An important aspect of MC generation is the precision to which the physics processes are modelled. At their simplest, simulations are of leading order (LO) and only include the dominant interactions for a physics process (typically ones with the fewest number of interacting vertices). More accurate simulations include next-to-leading order, next-to-next-leading order and so forth contributions. Dedicated software packages are often used for the computation of the cross-sections, which may be to higher levels of precision than is used to model the kinematics.

After generation, the appearance of the resulting particles in the ATLAS detector is modelled by a simulation [81] based on GEANT4 [82]. It is then necessary to apply the same reconstruction and identification algorithms, discussed at the end of the previous chapter, to the simulation that are applied to data. This requires applying additional weights multiplicatively to the MC events to account for, as an example, a reconstruction algorithm that has a different efficiency for data and simulation. Another important correction is to the distribution of the number of proton-proton interactions occurring per bunch crossing, which is known as pile-up reweighting. Other non-weight based corrections are also applied to the simulation, for instance to account for the Jet Energy Resolution mentioned in the previous section.

It is worth highlighting that fully simulating the response of the ATLAS detector is typically the most computationally intensive part of a physics analysis, due in large part to the complex geometry of the calorimeters. To help offset this, ‘fast’ simulation techniques have been developed as an alternative to the full simulation which use a parameterisation of the calorimeter responses in order to speed up the computation time [83]. Although not as accurate, fast simulation is typically useful for modelling hypothetical signals as often many slight variations of the signal need to be generated. As a side note, information from the MC generation step, prior to simulating the ATLAS detector response, referred to as ‘truth’ information, is often recorded and useful for verifying aspects of the modelling and for evaluating theoretical uncertainties.

Chapter 3

Analysis Methods

“All [statistical] models are wrong, but some are useful.”

George Box, statistician

Having seen at the end of the previous chapter how particles in proton-proton collisions are reconstructed, attention now turns to how this information can be used to search for new physics. As outlined in the introduction, searches for new physics at collider experiments involve looking for discrepancies, typically excesses, in the yields of particular types of events in data compared to the theoretical predictions of these yields based upon our understanding of the Standard Model (SM). How the significance of potential excesses is quantified, and the analysis steps necessary to make reliable predictions of these event yields, is the subject of the first part of this chapter in Section 3.1. Specific details of the HistFitter [84], pyhf [85] and underlying HistFactory [86] software packages that are used in the analyses presented in this thesis will be discussed where relevant. With this knowledge, Section 3.2 will then discuss how machine learning (ML) algorithms can be used to improve the sensitivity of these searches, with particular focus placed on Boosted Decision Trees (BDTs), the type of ML algorithm used by the analyses presented in this thesis. Subtleties related to the application of ML to the kind of analyses presented in this thesis will be discussed.

3.1 General Setup of a Search for New Physics

As discussed in the introduction, the essence of a search for new physics at collider experiments is a comparison of the number of particular types of events seen in actual data with predictions based upon the SM, with a significant discrepancy possibly indicating the presence of physics beyond the Standard Model (BSM). From here on, the most common case of searching for an excess caused by an additional physics process, referred to as the signal, will be considered.

The types of events examined, defined by the types of physics objects in them and the kinematics of the objects, are chosen based on predictions of the number of possible signal and expected SM background events, with the aim of maximizing the statistical significance of an excess the signal could cause. The event requirements define *regions* (of phase space). Observation of significant excesses in these regions that are aligned with the signal in question can be interpreted as evidence of the signal's existence, while the absence of any significant excesses can be used to place constraints on the likelihood of the signal's existence. In general, analyses are performed blinded to avoid human bias; analyzers do not look at the data in regions where significant signal is expected until they have confidence in the SM background predictions in these regions.

This section will first cover the ingredients needed for performing such an analysis before explaining how things are formulated statistically. A discussion of more advanced methods, used to enhance the sensitivity of searches (including those presented in this thesis), are then discussed at the end of this section.

3.1.1 Ingredients

Signal, background and data yields

In order to perform a search for new physics, it is necessary to have predictions of the expected number of signal and background SM events as well as the data yields in the analysis regions. This is done by simulation, discussed at the end of the previous chapter, for the signal and a mixture of simulation and fully or semi data-driven methods for the SM backgrounds. Semi-data-driven aim to account for potential shortcomings of the simulation, in the form of mis-modelling actual data, by applying

correction factors to the simulation. These correction factors usually take the form of normalisation factors for specific background processes and may be applied to regions as a whole or might vary across a kinematic range. Fully data-driven methods do not rely on MC simulation for the specific process in question and instead typically re-weight data in other regions of phase space. In general, fully data-driven methods are needed for QCD multi-jet backgrounds, which are difficult to simulate accurately.

Regions

As mentioned, the requirements placed on the events that are used for a search define regions. Regions are split into different types which serve different purposes. Regions in which a statistically significant amount of signal is expected are known as Signal Regions (SRs). Separate from SRs, Control Regions (CRs) are often defined which can serve multiple purposes, but in all cases the expected amount of signal in the CRs should be statistically insignificant, such that they can be safely unblinded before background estimates are finalised. CRs should be orthogonal to SRs; no events should be able to simultaneously enter a CR and a SR to avoid the double counting of events. CRs are often dominated by a particular type of background such that, by comparing the data to the nominal background prediction, they can be used to check the modelling of and if necessary determine correction factors for the background in question. In general, CRs also serve to constrain systematic uncertainties. This function of CRs is discussed later in this section after the statistical formalism has been explained. It is usually desirable for CRs to be kinematically similar to the SRs to give confidence that the background predictions will be accurate in the SRs, though often there is a trade-off between kinematic similarity and increasing the CR statistics.

Alongside SRs and CRs, a third type of region, Validation Regions (VRs), are also usually defined. VRs are used to check the background predictions agree well with data prior to unblinding the SRs, and so should also have low signal contamination. As with CRs, it is desirable for VRs to be kinematically close to the SRs. Unlike CRs, VRs are not used to constrain systematic uncertainties or fit normalisation factors (they are not included in the likelihood function discussed later in this section). Typically, VRs are closer to the SRs than CRs and can have slightly higher levels of signal contamination.

To conclude the discussion on regions, it is worth stating explicitly that all types of regions can be either binned in certain kinematic variables (in the form of histograms),

or unbinned. For the binned case, the yield in each bin is considered as a separate observation to compare to data.

Systematic Uncertainties

As well as the predicted yields in the analysis regions, it is necessary to estimate systematic uncertainties on both the signal and background yields. The sources of systematic uncertainty can broadly be divided into experimental uncertainties arising from detector and event reconstruction effects, theoretical uncertainties from the Monte Carlo (MC) modelling, and uncertainties arising from the finite number of MC events, here called MC statistical uncertainties. There may also be method specific uncertainties if data-driven methods are used to estimate background yields. For systematic and experimental uncertainties, what is required is an estimate of how these uncertainties effect the yields of the relevant background processes in each of the analysis regions.

For MC statistical uncertainties, as shown in [87], the sum of the MC weights squared provides an estimate of the variance, and so the square root of the sum of MC weights squared is used as the uncertainty. It is evident that increasing the number of raw MC events for a prediction (at a given integrated luminosity) should decrease the MC statistical uncertainty as the individual events will have smaller weights. It is worth noting that MC generators can produce small numbers of events with abnormally large weights which can severely increase this uncertainty even if they only occur rarely. As an example, if 10,000 raw MC events pass the selection requirements for a region, each with a weight of 0.01 (giving a total expected yield of 100 for the region), the MC statistical uncertainty on the predicted yield is 1%. However, if 10 out of the 10000 MC events (i.e. 0.1% of events) have instead a weight of 1, then the MC statistical uncertainty is over 3%, three times greater. Similarly, negatively weighted MC events contribute to increasing the MC statistical uncertainty while decreasing the total yield, thereby increasing the overall fractional MC statistical uncertainty, and so are also in general undesirable (but are often used to correct for overpredictions of the simulation or interference effects and so may be unavoidable).

3.1.2 Likelihood Construction

Having acquired the necessary ingredients, we can now review how searches are formulated statistically. Searching for new physics involves comparing the hypothesis that only the SM background is present with the hypothesis that a signal and the SM background is present. The fundamental construction for performing these tests is a likelihood function, or model, which describes the probability of getting the observed data given the theoretical predictions and uncertainties. The model consists of a number of parameters that control normalisation factors for specific physics processes and the effect of systematic uncertainties. Mathematically, the model can be written as $f(x|\phi)$ where x represents the observed data and ϕ represents the model parameters.

In general, only one of the model parameters is of interest, and is indeed referred to as the parameter of interest, μ , while the remaining parameters are known as nuisance parameters (NPs), labelled θ . For testing the existence of a new signal, the parameter of interest is the signal strength, μ_{SIG} , which for convenience will here just be labelled as μ . A μ value of one would indicate the presence of the signal with the expected cross-section, while a μ value of zero means the complete absence of the signal. Parameters may in general be constrained by a probability distribution or unconstrained. These constraints take the form of auxiliary measurements which are independent of the observed data yields. In general, μ and other normalisation factors are also required to be non-negative.

For a typical analysis, the likelihood function can be written as

$$\mathcal{L}(x|\mu, \theta) = \prod_r^{\text{regions}} \prod_b^{\text{bins}} P(x_{r,b}|\nu_{r,b}(\mu, \theta)) \prod_{\chi} C_{\chi}(a_{\chi}|\chi) \quad (3.1)$$

where the first term is a product sum over all bins in all regions being included with $P(x_{r,b}|\nu_{r,b}(\mu, \theta))$ being the probability of observing $x_{r,b}$ events in bin b of region r , given an expected number of events $\nu_{r,b}(\mu, \theta)$, which follows a Poisson distribution. The second term represents the auxiliary constraining terms with $C_{\chi}(a_{\chi}|\chi)$ being the probability of observing auxiliary data a_{χ} given the value of parameter χ . The product runs over all parameters $\phi = (\mu, \theta)$ which have constraint terms.

The expected number of events in each bin in each region is a sum over all the relevant (signal and background) physics processes. The expected yield for each process can be modified by normalisation factors or through the effect of systematic

uncertainties. The latter are controlled by parameters which have corresponding constraint terms. In HistFactory, the modifications to expected process rates from experimental and theoretical uncertainties are constructed by piecewise linear interpolation between the nominal expectation and ‘up’ and ‘down’ templates for that uncertainty provided by the user. Different methods exist for computing the ‘up’ and ‘down’ templates. Commonly, for experimental systematic uncertainties they may be determined by varying either the weights or kinematics of individual events and comparing the region yields to the nominal yield, while for theoretical uncertainties comparisons between different MC generators are often used. The constraining terms typically take the form of Gaussian distributions, with $\sigma = \pm 1$ corresponding to the ‘up’ and ‘down’ values and $\sigma = 0$ being the nominal yield. The multiplication of the NP constraining terms in Equation 3.1 assumes that the NPs are independent of each other. In HistFactory, unless otherwise specified a single NP controls the MC statistical uncertainty in each bin of each region for all the physics processes together and is also constrained by a Gaussian distribution.

3.1.3 Fitting

The most probable set of parameters values ϕ given the data is the set that maximizes the likelihood. The process of obtaining the most likely set of parameter values is known as performing a ‘fit’ or profiling. In typical analyses at ATLAS, likelihoods typically consist of over 100 NPs, hence finding the set of parameter values which maximises the likelihood cannot be done analytically. Instead, dedicated tools have been developed that perform these maximizations numerically. The fits performed in Chapters 4 and 5 use the HistFitter software package which uses the MINUIT2 software package [88] for this step, while the fits in Chapter 6 use the pyhf software package with Scipy [89] being used for the likelihood maximization. In practice, these packages minimize the negative log of likelihood functions as, given the typical form of the likelihood functions, these are easier to compute. These packages also provide measures of the quality of fits to ensure they converge to stable minima.

If, as a result of the fit, the systematic NPs take values that differ significantly from their nominal value, they are said to have been ‘pulled’ by the fit. When they occur, such pulls should be carefully examined to understand their origin. Additionally, uncertainties on the NP values can be significantly constrained after a fit and this should also be understood. Furthermore, how the systematic uncertainty NPs are

correlated with each other should be checked; in general it is desirable for the NPs to not be strongly correlated with each other as they should describe physically uncorrelated effects. However, correlations can arise in cases where the up and down templates of the uncertainties are similar, in which case it should be assessed if such a similarity is reasonable.

3.1.4 Hypothesis Testing

The fitting of the likelihood function described above is used to perform hypothesis tests from which decisions, such as whether the data indicates evidence of new physics, can be made. For search-based analyses two types of hypothesis tests are performed. The first, used when trying to discover new physics, is formulated as having the null hypothesis H_0 as only the SM background being present and the alternative hypothesis, H_1 as the signal also being present on top of the background. The second case, used for establishing limits on particular signals, is formulated the opposite way around, with H_0 being the signal plus background hypothesis and H_1 the background only hypothesis. This is because statistical tests are usually formulated in terms of rejecting the null hypothesis in favour of the alternative one.

Hypothesis testing requires the use of test statistics in order to compute p-values, which indicate the level of compatibility of a hypothesis with the data. Explicitly, the p-value for a given hypothesis test is the probability of obtaining data that is less than or equally consistent with the hypothesis as the actual data obtained. Following the Neyman-Pearson lemma [90], for making the kind of statements about discovery or exclusion that are desired, likelihood ratios are the most useful kind of test statistic. Specifically, the ratio

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta}_\mu)}{L(\hat{\mu}, \hat{\theta}_{\hat{\mu}})}, \quad (3.2)$$

where $\hat{\theta}_\mu$ are the profiled set of NPs for a given value of μ and $\hat{\mu}$ is the value of μ which maximizes the likelihood, is used to construct the test statistics for search-based analyses in the ATLAS collaboration. This ratio is known as the profile likelihood ratio. Clearly, $0 \leq \lambda(\mu) \leq 1$ as the denominator is the global maximum likelihood.

In practice, the test statistic used is given by

$$t_\mu = -2 \ln \lambda(\mu), \quad (3.3)$$

as this is easier to compute given the form of the likelihoods. For a value of μ very consistent with the data, t_μ approaches zero while for very incompatible values t_μ approaches infinity. Hence, the p-value for a particular value of t_μ is computed as

$$p_\mu = \int_{t_\mu}^{\infty} f(t_\mu|\mu) dt_\mu, \quad (3.4)$$

where $f(t_\mu|\mu)$ is the probability distribution of t_μ given a particular value of μ (this is a distribution because the value of t_μ can vary if the experiment were to be repeated for the same underlying true value of μ). When testing for the discovery of a positive signal, t_μ is taken to be zero for $\mu < 0$, as in this case it is desired that only excesses in the number of events contribute to rejecting H_0 .

The distributions $f(t_\mu|\mu)$ are in general not known and so must be approximated. Two common approaches exist for approximating them. The first is to generate pseudo-experiments using Monte Carlo, known as “toys”, from the likelihood function. Many toys ($\mathcal{O}(1000)$) are needed to model the distributions accurately, particularly the tails which are often of interest, which can be computationally expensive. An alternative approach is to use an analytical approximation [91], known as the asymptotic approximation, for t_μ given by

$$t_\mu = \left(\frac{\mu - \hat{\mu}}{\sigma} \right)^2 + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right), \quad (3.5)$$

where $\hat{\mu}$ is the estimator of μ which follows a Gaussian distribution with standard deviation σ , and N is the expected event yield in the signal region. For this to be a good approximation, N , needs to be sufficiently large (typically $N > 10$) such that the $\mathcal{O}(\frac{1}{\sqrt{N}})$ term can be neglected. The standard deviation σ is in general obtained from the covariance matrix of the estimators of all the parameters but can also be estimated as

$$\sigma^2 = \frac{(\mu - \mu')^2}{t_{\mu,A}}, \quad (3.6)$$

where μ' is the assumed signal strength and $t_{\mu,A}$ is the test statistic evaluated on the Asimov data set, which is equal to the region yields when the model parameters

take their expectation values. With this approach, the distribution of $f(t_\mu|\mu)$ is then approximated by Equation (33) in reference [91].

3.1.5 CL_s values

Instead of just using p-values, a quantity known as the CL_s value is computed using the signal plus background and background only hypothesis p-values. The explicit goal of the CL_s method [92] is to discover a true signal or exclude a non-existent signal as strongly as possible while keeping the probability of discovering a non-existent signal or excluding a true signal below a certain user defined threshold. The CL_s value is computed as

$$CL_s = \frac{P(s + b)}{P(b)}, \quad (3.7)$$

where $P(s + b)$ is the p-value for the signal plus background hypothesis given by Equation 3.4 with $\mu = 1$ and $P(b)$ is the p-value of the background only hypothesis given by Equation 3.4 with $\mu = 0$. The CL_s value is used to set limits on signals at a given confidence level; for example signals with a CL_s value less than 0.05 are said to be excluded at a 95% confidence level. The lower the CL_s value, the greater the confidence with which the signal can be said to be excluded. The motivation for using the CL_s method instead of just using $P(s + b)$ is to limit the extent to which small positive signals can be said to be excluded as a result of significant deficits in the observed data compared to the SM background predictions being seen in SRs. In such cases, the value of $P(b)$ would also be small thereby increasing the CL_s value.

As mentioned, analyses are typically performed blinded, by which the data in the SRs is only be revealed when all other aspects of the analysis are firmly established. To gauge the sensitivity of an analysis, *expected* CL_s values are often computed which use the Asimov dataset in place of the real observed values (which in turn are used after unblinding to give the final *observed* CL_s values). In searches for BSM models with different signal parameters (such as the masses of the new particles), it is common to plot 95% confidence level ‘exclusion contours’ in the signal parameter space joining points for which the CL_s value is equal to 0.05. As well as being able to exclude particular signals, CL_s values can also be used to set upper limits on the signal strength for a given signal. This involves computing the CL_s value for different values of μ from which the value of μ for which the CL_s value would fall below a given threshold (e.g.

0.05) can be determined. This value of μ is then said to be the upper limit on the signal strength at that confidence level; signals that are at least as strong as this value of μ are said to be excluded at that confidence level. Using the nominal cross-section used for the signal, these upper limits can be converted into upper limits on the cross-section for the signal process.

3.1.6 Fit Types

For future reference, it is useful to explicitly define three different types of fit that are used in the analyses presented in this thesis. The first is known as a background-only fit, the unique feature of which is that the likelihood does not include a signal model (or equivalently the signal strength is fixed to zero). Background-only fits are useful in determining the normalisation factors on specific background processes and for verifying background predictions are compatible with the observed data in validation regions (VRs). VRs are in general not included in the likelihood and as such they are not used in the determination of normalisation factors or the constraining of systematic uncertainty NPs. Background-only fits are often performed on likelihoods that only include the control and validation regions prior to unblinding the signal regions.

The second type of fits are exclusion fits, also known as model-dependent signal fits. These fits include a specific signal model and their purpose is to assess the compatibility of that signal with the observed data through the computation of CL_s values. The third type of fit are model independent or discovery fits. These are similar to the exclusion fits but instead of including a particular signal model test for the presence of a generic excess in a given (unbinned) SR. The SRs being used for discovery fits are referred to as discovery regions. In HistFitter, discovery fits are performed through the use of a dummy signal with a nominal yield of one in the SR being tested. The upper limit on the signal strength of this dummy signal can then be interpreted as a model-independent limit on the number of signal events in the SR. The upper limit can also then be converted into an upper limit on the visible cross-section, which is defined as

$$\sigma_{\text{vis}} = \epsilon \times a \times \sigma, \quad (3.8)$$

where ϵ is the reconstruction efficiency (fraction of ‘real’ physics objects that get reconstructed), a is the acceptance (fraction of events that pass the SR selection criteria,

excluding reconstruction effects) and σ the nominal production cross-section of a physics process.

3.1.7 Control Regions, Shape fits and Combinations

Before moving on it is worth discussing the operation of CRs, shape fits and combinations of individual analyses as these techniques are important in the later chapters of this thesis. To understand how they function, it is first necessary to understand how systematic uncertainties are formulated in the likelihood.

Formulation of systematic uncertainties

As previously discussed, theoretical and experimental systematic uncertainties are considered in analyses. In general, uncertainties can be correlated or uncorrelated between regions (and between bins of kinematic variables). Where uncertainties are correlated, their effect is controlled by a single NP whilst uncorrelated uncertainties are controlled by separate NPs. In the analyses considered in this thesis, theoretical uncertainties are taken to be uncorrelated between processes but correlated between analysis regions except where otherwise indicated. This means that, for each process, a single NP controls the effect of a systematic on the process yield in all regions, with each process having a separate NP (for example there would be a separate NP associated to the uncertainty on the Parton Distribution Functions for each background process effected by this uncertainty). Experimental uncertainties by contrast are correlated between processes and between regions and so each is controlled by a single NP.

Control Regions

That systematic uncertainties can be correlated across analysis regions allows their effect in certain regions to be constrained by the data yields in different regions. In particular, the size of systematic uncertainties in SRs for particular backgrounds can be constrained from dedicated CRs which have high statistics and high purity in that background. The constraining power a region has depends upon the data statistics and the size of the systematic uncertainties in the region (very large uncertainties require fewer statistics to constrain). For this reason, it is generally desirable for CRs to contain significantly more events than SRs whilst still having similar size systematic

uncertainties such that they have greater constraining power (and so that validation regions can be used to test the accuracy of the model with these constraints applied).

In addition to there constraining power, HistFitter and other fitting packages make use of a transfer factor approach for backgrounds with normalisation factors which further allows CRs to reduce the impact of systematic uncertainties. As outlined in the HistFitter paper [84], for a simple analysis with one SR and one CR, and one background process estimated from MC multiplied by a normalisation factor μ_b , the background predictions in the SR and CR are given by

$$N_b(\text{SR}, \text{est}) = \mu_b \times \text{MC}_b(\text{SR}, \text{init.}), \quad (3.9)$$

$$N_b(\text{CR}, \text{est}) = \mu_b \times \text{MC}_b(\text{CR}, \text{init.}), \quad (3.10)$$

where $\text{MC}_b(\text{SR}, \text{init.})$ and $\text{MC}_b(\text{CR}, \text{init.})$ are the initial MC background predictions from simulation. As μ_b is the same in both the SR and CR, the estimation in the SR can be re-written as

$$N_b(\text{SR}, \text{est}) = N_b(\text{CR}, \text{fit}) \times \frac{\text{MC}_b(\text{SR}, \text{init.})}{\text{MC}_b(\text{CR}, \text{init.})}. \quad (3.11)$$

where $N_b(\text{CR}, \text{fit})$ is the fitted number of background events in the CR and the ratio $\frac{\text{MC}_b(\text{SR}, \text{init.})}{\text{MC}_b(\text{CR}, \text{init.})}$ is known as the transfer factor. From this, it can be seen that if the effect of a systematic uncertainty is to simultaneously increase $\text{MC}_b(\text{SR}, \text{init.})$ and $\text{MC}_b(\text{CR}, \text{init.})$, its effect on $N_b(\text{SR}, \text{est})$ will partially cancel, thereby reducing its impact. This formalism also explains another motivation for having CRs with large statistics, as the uncertainty on $N_b(\text{CR})$ directly effects the uncertainty on $N_b(\text{SR}, \text{est})$. The TF approach is implemented for all background processes carrying normalisation factors for the analyses presented in this thesis.

Shape Fits

As previously mentioned, the regions defined in an analysis can be binned in particular kinematic variables. Fits including such regions are called shape fits. Shape fits in SRs can enhance the sensitivity of an analysis, as the presence of a signal could cause small excesses across multiple bins which combined may be significant, even if the excesses in the individual bins may on their own be insignificant. Bins with lower

signal purity, defined as the ratio of the signal yield to the background yield (S/B), may also aid in constraining systematic uncertainties, although this depends on whether the uncertainties are correlated across the bins. Unless otherwise stated, the fits in this thesis utilising shape fits correlate the uncertainties on the shape and on the normalisation in each region (i.e. their effect is controlled by a single NP.) Additionally, shape fits can be helpful when designing searches for multiple signals which may have slightly different kinematics (testing signals with different particle masses for example) which may cause excesses to occur in different bins. In these scenarios, shape fits can be helpful in providing sensitivity to multiple signals without having to define multiple regions and, in the case of an excess, discriminating between the different signal hypotheses.

Combinations

Often, BSM models, such as Supersymmetry, may lead to experimental signatures with different final states, but, for practical purposes, multiple separate analyses may be conducted each of which targets only a subset of these final states. As an example, a particular BSM model may result in final states with one lepton and a high p_T jet, or alternatively two leptons and no high p_T jets. In such cases, combining the results of the individual analyses can lead to a significant gain in the overall sensitivity, particularly for models in which the individual analyses have comparable sensitivity. Mechanically, combinations take the form of running new statistical fits that include all the signal and control regions of the individual analyses. In order to be valid, it is necessary for all regions to be orthogonal to each other, or for overlaps to be otherwise removed (or for overlaps to be sufficiently small to have negligible impact). The power of performing combinations will be demonstrated for chargino pair production in Chapter 6.

3.1.8 High Level Kinematic Variables

To aid with discriminating possible BSM signals from the SM background, high-level kinematic variables, constructed from the four vectors of reconstructed physics objects, are often used. For later reference, a couple of such variables are defined here.

The first of these variables is the transverse mass computed between a pair of physics objects. For two objects with transverse four momenta $p_T^1 = (E_T^1, \vec{p}_T^1)$ and

$p_T^2 = (E_T^2, \vec{p}_T^2)$ and masses m_1 and m_2 , the transverse mass is defined as

$$m_T^2(p_T^1, p_T^2) = m_1^2 + m_2^2 + 2(E_T^1 E_T^2 - \vec{p}_T^1 \cdot \vec{p}_T^2), \quad (3.12)$$

Hence, the transverse mass between a particle, for example a tau-lepton, and the $\vec{p}_T^{\text{miss}}$ is defined as

$$m_T^2(\tau, \vec{p}_T^{\text{miss}}) = m_\tau^2 + 2(E_T^\tau E_T^{\text{miss}} - \vec{p}_T^{\text{miss}} \cdot \vec{p}_T^\tau), \quad (3.13)$$

As the energies and momenta for the events considered at the LHC are typically much greater than the tau-lepton mass, the transverse mass can be approximated as

$$m_T^2(\tau, \vec{p}_T^{\text{miss}}) \approx 2E_T^\tau E_T^{\text{miss}}(1 - \cos(\Delta\phi(\tau, E_T^{\text{miss}}))), \quad (3.14)$$

and hence has larger values when the tau-lepton and $\vec{p}_T^{\text{miss}}$ are back-to-back.

The second variable is the “stransverse mass” [93, 94], more commonly referred to as m_{T2} , which is a useful kinematic variable for events in which a pair of particles are produced that each decay to visible and invisible daughter particles. This is the case for the direct stau signal which is the subject of the next chapter. For an event in which the two produced particles each decay to two visible daughters with transverse four momenta d_T^1 and d_T^2 and two invisible daughter particles, m_{T2} is defined as

$$m_{T2} = \min_{q_T^1 + q_T^2 = \vec{p}_T^{\text{miss}}} \left[\max \left(m_T(d_T^1, q_T^1), m_T(d_T^2, q_T^2) \right) \right] \quad (3.15)$$

where q_T^1 and q_T^2 are four vectors that are chosen to minimise the larger of the two transverse masses under the constraint that their total energy and momentum equals the total missing transverse energy and momentum. The mass of the invisible daughter(s) is (are) in general a parameter of m_{T2} (entering into Equation 3.15 in q_T^1 and q_T^2). A potentially useful property of m_{T2} for these events is that it has a kinematic endpoint at the parent particle mass. As an example, for two W -bosons decaying leptonically to visible charged leptons and invisible neutrinos, the m_{T2} distribution has an endpoint at the W -boson mass. For a BSM signal the m_{T2} distribution may extend well beyond this endpoint. Since its inception, m_{T2} has been used to great effect by many searches for supersymmetric particles performed by the ATLAS and CMS experiments and will be used in the search for direct stau production in Chapter 4.

3.2 Enhancing the Sensitivity of Searches using Machine Learning

Following the discussion in the previous section, it is clearly desirable for signal regions (SRs) to have a high signal purity (S/B). SRs are most simply defined via a ‘cut and count’ approach in which (typically one dimensional) ‘cuts’ on kinematic variables define the selection criteria, for example requiring an event to have E_T^{miss} greater than 250 GeV. Machine learning (ML) algorithms, designed to classify potential signal events from background events using kinematic variables as input features, can instead be used to define SRs. These algorithms produce models that compute a classifier score for each event. The scores can be interpreted as how likely the event is to belong to either the signal or background class and an SR can be defined by cutting on the classifier score. Such SRs should have higher signal purity than those that can be defined using a cut and count approach as ML algorithms can better exploit correlations between kinematic variables.

ML algorithms have been widely used in LHC analyses for several years in searches for BSM physics [95]. Also, as seen in the previous chapter, many of the reconstruction and identification algorithms make extensive use of ML for increased efficiency and enhanced background rejection. Only binary classifiers, which have two input classes (the signal and the background) are used in this thesis but multi-class classifiers, which classify more than two classes, have also been used in BSM searches to define CRs enriched in particular backgrounds. In this Section, general considerations of using binary classifiers for defining SRs will be discussed followed by a detailed look at the workings of Boosted Decision Trees, which are used in analyses presented in Chapters 4 and 5.

3.2.1 General Considerations

Binary Classification

The goal of classification algorithms is to produce a function or model which can, based on a vector of input features $\mathbf{x}$, predict the correct class an event belongs to, $\hat{y} = f(\mathbf{x})$ where $\hat{y}$ is the predicted class label which is desired to be 0 for the negative class (here the background) and 1 for the positive class (here the signal). Generally, these

functions consist of many internal parameters. The values of the internal parameters are determined during training in which part of a dataset, or indeed a dedicated dataset, known as the training (data)set, is used. The classifiers used in this thesis are trained using supervised learning, in which during training the class labels of the data are known. The parameters of the model are determined through minimizing a loss function. The loss function used in the models presented in this thesis is known as the binary cross entropy in which, for a single event, the contribution to the loss is given by

$$l(y_i, p_i) = -(y_i \ln p_i + (1 - y_i) \ln(1 - p_i)), \quad (3.16)$$

while the total loss is

$$L(y_i, p_i) \propto \sum_{i=1}^N -y_i \ln p_i + (1 - y_i) \ln(1 - p_i), \quad (3.17)$$

where the sum is over all events N and p_i is the predicted probability from the model that event i belongs to the positive class. This probability is typically related to a ‘raw’ predictive score from a model $\hat{y}$ by a (scaled) sigmoid transformation of the form

$$p_i = \frac{1}{1 + e^{-\hat{y}_i}}. \quad (3.18)$$

A raw score of zero means an event has a probability of 0.5 of belonging to the positive class (i.e. the same as random guessing).

Over-fitting

After training, the general performance of a model (how well it can distinguish between the different classes) should be assessed on a dataset that is statistically independent of the training dataset which in this thesis will be referred to as the evaluation (data)set. This is because it is possible for models to learn significant amounts of noise in the training dataset at which point they *might* be said to be over-fit, although this depends upon exactly how over-fitting is defined. Classically, over-fitting tends to degrade, or at least not significantly improve, a models’ performance on unseen data and/or introduce unnecessary complexity¹ and so is undesirable. Conversely, it is

¹ Though this is not necessarily always true, and recent ML research suggests optimal performance is in some cases be obtained from models which would classically be considered to be over-fit [96].

also possible for a model to not learn all the generalizable information available from the training data in which case the model may be said to be under-fit.

It should be noted though that it is possible for ML models to perform significantly better on the training set than on unseen data without them being *problematically* over-fit. This is because in some cases it might be difficult for the model to learn the general underlying characteristics of the data without learning *any* noise. Given this, simply comparing the performance of the model on the training dataset and an unseen dataset may not necessarily give a clear indication of whether a model is problematically over-fit or not.

Arguably, the unseen dataset being large enough to be able to reliably estimate a models' performance is the only thing which is important. How large this dataset needs to be depends upon the model's complexity with more complex models needing larger datasets to capture all of their behaviour. Concrete prescriptions of how much discrepancy between the train and unseen dataset performance is permissible, and/or how large an unseen dataset needs to be for a given model do not exist as such requirements depend heavily on the specific dataset.

In the absence of such prescriptions, in this thesis early stopping is used as the primary method to help prevent problematic over-fitting. In early stopping, an additional dataset, here called the validation (data)set, is used to assess the models' performance during training at regular intervals. Additional complexity ceases to be added to the model when this performance stops improving. In this way, early stopping limits the extent to which models can become over-fit. The validation set should be considered as a subset of the training set, as it is not completely unseen during training and model selection (see below), hence it is important to still have a completely unseen dataset for the final evaluation of a models' performance.

Hyperparameters and Model Selection

As well as the internal parameters of the model, ML algorithms also have hyperparameters which control aspects of the training and can set restrictions on a model's complexity. The optimal set of hyperparameters maximizes the model's general performance (the performance on unseen data). Along with the hyperparameters, there are numerous other attributes to a model and its training that can also be altered, for instance what kinematic variables are used as inputs or how much data is used

for training. To select which input/hyperparameter configuration leads to the best model, a validation dataset (which may or may not be the same as the one used for early-stopping) can be used with the optimal configuration being determined as the one which gives the best performance on this dataset. Again it is important that the final evaluation of the model uses an independent dataset to this validation set, as potentially the validation set performs slightly better with the chosen model than a completely unseen dataset.

Different strategies exist to find the optimal set of hyperparameters. A simple, brute-force approach used in the work presented in this thesis is a grid-search, whereby all possible combinations of hyperparameters are tested from a user-defined grid. Early stopping can implicitly set certain hyperparameters of a model and so can reduce the dimensionality of grid searches. There are several metrics to assess and compare the performance of models. In this thesis, the Area Under Curve (AUC) score is used which is defined as the area under a Receiver Operating Characteristic (ROC) curve, which is a plot of signal efficiency against background rejection. A perfect model, that classifies all events correctly would have an AUC score of 1.0 while an AUC score of 0.5 is equivalent to random guessing. An example ROC curve is shown in Figure 3.1.

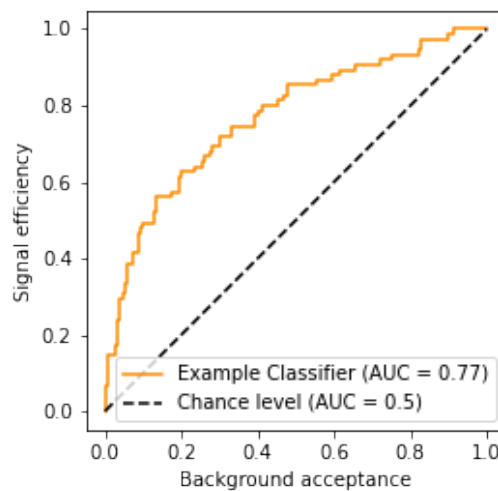


Figure 3.1: Example Receiver Operating Characteristic, with example Area Under Curve scores.

Model Interpretation

It is important to prevent ML models from being black boxes by understanding what features of the dataset they are exploiting. This can be done by evaluating the feature importance's of the input variables, which measures how often the input variables are used by the model and how much their use improves the model performance. Shapely Additive ExPlanations (SHAP) values [97,98] provide additional insight into the dependence of a model on the input features. SHAP values use concepts from game theory, treating input features as players in a game, to describe the effect of adding an input feature to a model on the output score for each event. This is computed from averaging the impact of adding the feature to the set of models defined by the possible combinations of the other input features, in order to take into account correlations between the input features. A large SHAP score for the value of a particular input feature indicates that the feature having such a value contributes significantly to the model score for that event belonging to the positive class (the class for which model scores are typically high).

The impact certain features have on a particular event for one of the models used in Chapter 4 is shown in Figure 3.2. The event (and indeed all events for this particular model) starts with a score of $E[f(X)] = 0.583$ based on the relative balance and scaling of the input classes. The impact the input features have on this baseline prediction is then shown, with the total from all features equalling the final output score $f(x) = 0.768$, indicating this event is deemed overall to be signal like. From this plot it can be seen the (relatively high) value of the variable $\Delta\phi(\tau_1, \tau_2)$ contributes most significantly to the final score being signal like, while the (relatively low) value of the m_{T2} variable lowers the overall score. More details on this specific model can be found in Chapter 4.

Cross Evaluation

Generally, it is desirable to use as much data as possible for training ML models, whilst also having enough data in the evaluation set to meaningfully assess models' performances. In the context of searches for new physics at the LHC, when using the typical setup of separate training and evaluation sets, the desire for a large training set must be further traded-off against the desire to have a large evaluation dataset to reduce the relative statistical error arising from finite (MC) statistics for the signal and

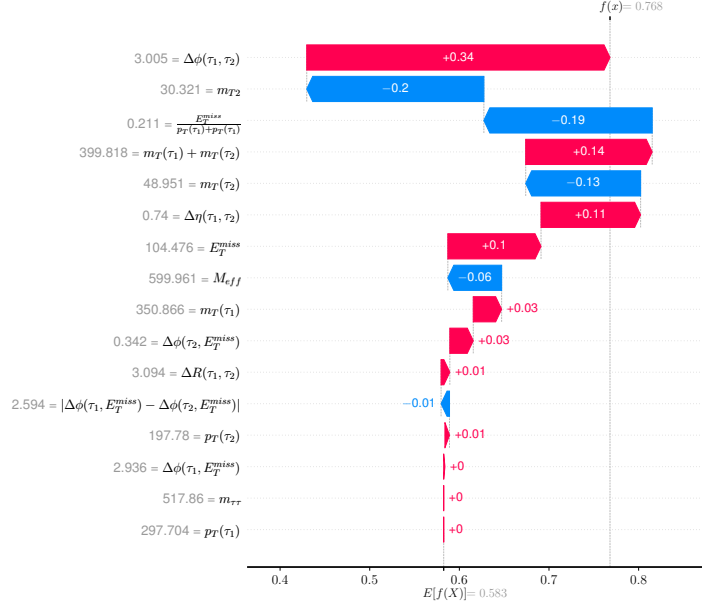


Figure 3.2: SHAP values of the input features for an example event for one of the models discussed in Chapter 4. The event initially has a score of $E[f(X)] = 0.583$ but, due to the value of the input features, has a final model score of $f(x) = 0.768$ (meaning it is more signal like).

background predictions. A setup that avoids this trade-off, is to make use of 'Cross Evaluation'² in which the total dataset is divided into N folds and N separate models are trained with each model using one of the folds as its evaluation set. An overall model score can be defined which takes, for each event, the score coming from the model for which that event was in the evaluation set (and therefore not in the training set). Such a procedure allows all the available signal and background events to be used for evaluation without incurring a bias from the models differing performance on training vs unseen data. For the evaluation of data not seen by any of the models, events would need to be randomly assigned to one of the models. The performance of each of the models should be sufficiently similar such that which model is used for an event does not greatly affect the overall output score. Validation datasets can

² 'Cross Evaluation' is made up terminology based on the more standard cross *validation* with which it shares similarities, but there are important differences. Generally, cross validation is used during model selection to verify a chosen hyperparameter configuration/setup performs similarly across different training and validation sets. Typically, a final model is then trained using all the data with the chosen hyperparameters (as using more data can in general only further improve a model). Cross validation is useful in applications where the model itself is the final product, whilst in the context of a BSM LHC search the final product is the evaluation of the signal and background predictions (and eventually the data) using the model.

also be included in cross evaluation. A schematic of 5-fold cross-evaluation, including validation datasets, is shown in Figure 3.3.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Model 1	Evaluation	Training	Training	Training	Validation
Model 2	Validation	Evaluation	Training	Training	Training
Model 3	Training	Validation	Evaluation	Training	Training
Model 4	Training	Training	Validation	Evaluation	Training
Model 5	Training	Training	Training	Validation	Evaluation

Figure 3.3: Schematic of 5-fold Cross Evaluation. Five models are trained and each fold is evaluated using the model for which it was not part of the training (or validation) dataset(s).

Importantly, when systematic uncertainties that modify the kinematics of simulated events are evaluated, for each event the systematic variations should be evaluated with the same model as was used for the nominal variation in order to not conflate any differences in the models performances' with the effect of the systematic uncertainties. Cross-evaluation has been used by many previous analyses at the ATLAS, CMS and LHCb experiments.

Imbalanced Classes and Event Weights

Often the classes in a classification problem occur with different frequencies and are said to be imbalanced, with the majority class being the one that occurs more frequently. This can complicate training a good model, as the model may be able to achieve good performance simply by classifying all events as the majority class. Imbalanced classes can often occur in the context of BSM searches, where the BSM signal tends to be much smaller than the background (at least in terms of the expected yields). A further complication arises if the individual events have their own non-uniform weights, as is also the case for BSM searches in which MC simulation is used for the signal

and/or background³. When using events weights, the loss function in Equation 3.17 is modified to include the individual event weights, w_i , as

$$L(y_i, p_i) \propto \sum_{i=1} w_i [-y_i \ln p_i + (1 - y_i) \ln(1 - p_i)]. \quad (3.19)$$

When event weights are being used, imbalance may occur in either the weighted yields or the raw number of events available or both. An imbalance in the weighted yields can be resolved by scaling the weights of one or both of the classes such that the total weight of each class is equal, and is the approach taken for the models presented in this thesis. An advantage of this approach is that the model is then agnostic towards the total cross-section of the signal compared to the background; it cares only about relative shape differences in the input variables. This can be useful as often the cross-section of hypothetical BSM signals is not known precisely. An imbalance in the raw number of events presents a more complicated issue; a model might be incentivised to isolate regions of the input phase space where the S/B might be high as a consequence of this imbalance rather than these regions actually having a high weighted S/B, analogous to binning a histogram too finely as is illustrated in Figure 3.4. The isolation of such regions by the model could be problematic as events being in such regions would contribute to these events having high classifier scores, and the background is likely under-estimated in these regions. Compounding the issue is that such an effect may not be detectable via over-fitting checks as the number of raw background events in such regions may be low in both the training and evaluation datasets.

To resolve this potential issue, it is desirable to prevent models from deliberately isolating such regions, or at least limiting the extent to which they can do so. For the models presented in this thesis, this is done through the model hyperparameters. Specific details are discussed in the relevant analysis chapters. It should be noted that simply having the same number of raw events for signal and background would not solve or prevent this issue as the raw events are unlikely to be evenly distributed across the multidimensional phase space defined by the kinematic variables being used for training, and so these ‘artificial’ regions of high S/B could still occur. Furthermore, simply restricting how tight a cut is placed on the final classifier score also does not

³ It is of course possible to not use the event weights in the training and simply set all the weights to be equal. However, it is generally desirable to use the event weights during training so that the shapes of kinematic variables are accurate.

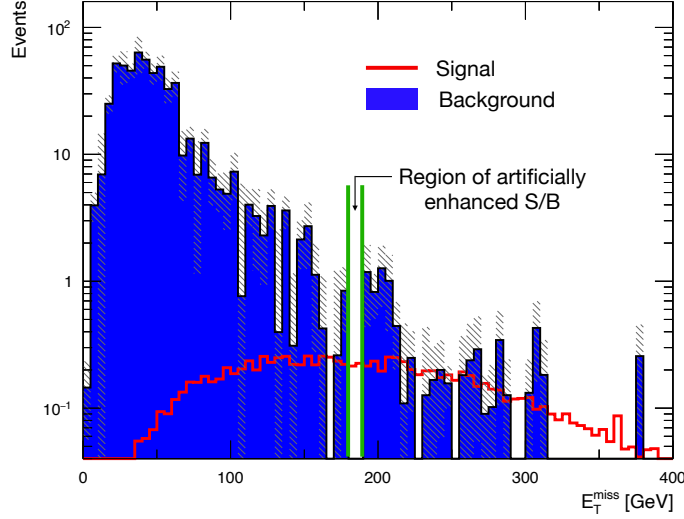


Figure 3.4: Example of a region of artificially high S/B caused by an imbalance in the number of raw signal and background statistics. The hashed bar indicates the MC statistical uncertainty on the background. The signal has a much greater number of raw MC statistics at high E_T^{miss} than the background, making it possible to define regions where the background prediction is zero. A similar problem arises from simply cutting on E_T^{miss} too tightly (above 300 GeV). Clearly the background is underestimated in such regions hence they should not be used to define SRs. It is therefore important to ensure that ML models are not making such selections.

solve this issue as different regions of the input feature phase space can end up at high scores.

Additionally, when using data with event weights special attention must be paid to the handling of negatively weighted events. From Equation 3.19 it can be seen that incorrectly classifying a negatively weighted event decreases the loss and thereby improves the performance of the model. From the model's perspective this makes sense as classifying negatively weighted background events as being very signal like will ultimately lead to a lower background yield at high scores. However, as discussed, negatively weighted yield predictions are unphysical and in general it is desirable to avoid separating these negatively weighted events from the positively weighted events they are supposed to cancel with. There are several ways negatively event weights could be dealt with. In the models presented in this thesis, a conservative approach is used whereby negatively weighted events are not used during the model training such that it is impossible for the model to be deliberately biased towards incorrectly classifying them. This comes with the disadvantage that the shapes of the input kinematic variable distributions are less accurate, particularly for physics

processes with larger fractions of negatively weighted events, but this effect is found to be small for the models presented in this thesis.

3.2.2 Boosted Decision Trees

As mentioned, the ML models used in the analyses presented in this thesis are Boosted Decision Trees (BDTs). LightGBM [99] being the software package used for their implementation. In the following, a general description of BDTs will be given including a discussion of the hyperparameters. Specific details of LightGBM will also be mentioned, including hyperparameter names and a brief discussion of some advantages LightGBM offers compared to other BDT software packages.

General Description

Boosting in general refers to combining many individual models together (which typically have weak individual classification power) into a single model; BDTs consists of a sequence of many individual decision trees. A single classification decision tree makes a series of selections based on the input features to separate the classes from each other. Initially all events start with the same score, $\hat{y}_i = \hat{y}_i^0$ which depends on the overall balance of the two classes. The selections made by the tree splits the events into nodes that have more of one class than the other. In LightGBM, for a given node the selection chosen is the one that maximizes the reduction in the loss function. BDTs and decision trees in general are therefore considered greedy algorithms, as decisions made early on in the tree are not necessarily optimal for the splits made later. Nodes cease to be split based upon various criteria controlled by the BDTs hyperparameters, discussed later in this section. The terminal nodes of a decision tree are known as leaf nodes. The new score for each event depends upon the leaf node it ends up in, with the leaf score being determined by the ratio of signal and background events that end up in it; in general nodes with a higher signal to background ratio will have a greater leaf score. These leaf scores can be used to compute the predictions after the first tree

$$\hat{y}_i = \hat{y}_i^0 + \alpha f(x_i), \quad (3.20)$$

where x_i is the vector of input features for event i , $f(x_i)$ represents the output score from the decision tree and α is the learning rate; a hyperparameter which controls how

quickly the scores are modified. A diagram of an example decision tree is shown in Figure 3.5.

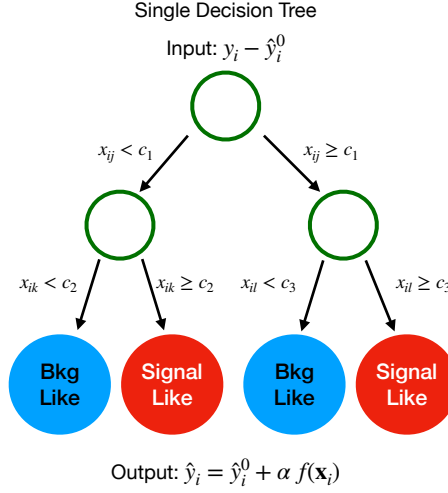


Figure 3.5: Diagram of a single decision tree. Various selection criteria are made at each node based on the input features in order to separate the classes. The end nodes are leaf nodes and can be labelled as being either signal or background like. The x_{ij} denote the values for event i of input feature j and the c_n are the ‘cut’ values on these features.

In a BDT, subsequent decision trees use either the residual after the previous tree as their input (for instance the second decision tree would use $y_i - \hat{y}_i^1$ as the input for each event), or a re-weighting based upon whether events were misclassified. In this way, the second BDT is encouraged to correctly classify events that were incorrectly classified by the first tree. When the residuals are used, the output of the second tree serves as a correction and is added to the overall prediction, which should become more accurate. When using a re-weighting approach, multiple options exist for computing an overall score, for instance taking a (weighted) majority vote from each decision tree. LightGBM uses the residuals as inputs for subsequent trees, and so only that approach will be considered in the rest of this section. A diagram showing the addition of a second decision tree is shown in Figure 3.6. This pattern is repeated, with the third decision tree using the residuals from the second tree and so on for further trees.

With this formalism, the loss function for an individual event (Equation 3.16) after m decision trees can be written (neglecting event weights) as

$$l(y_i, \hat{y}_i^m) = -y_i \ln p_i^m - (1 - y_i) \ln(1 - p_i^m), \quad (3.21)$$

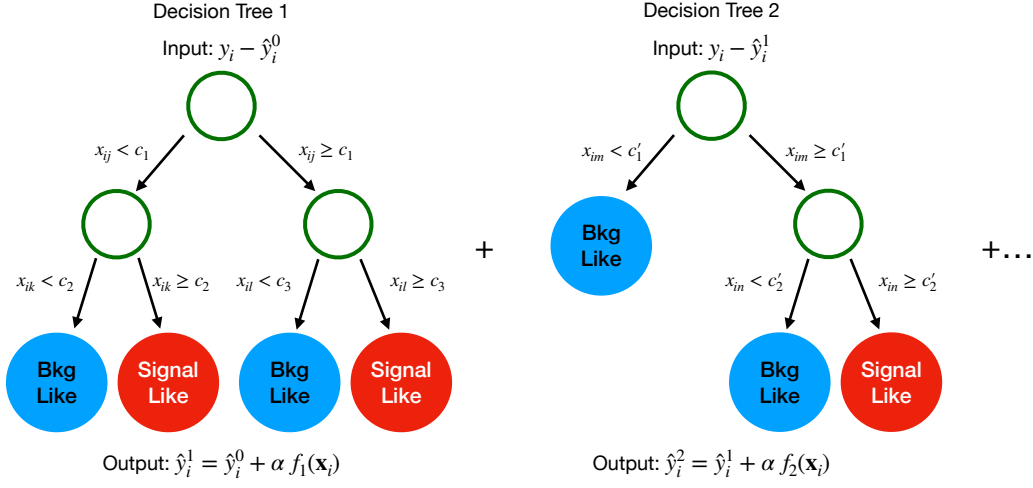


Figure 3.6: Diagram of multiple decision trees comprising a BDT in which the second decision tree uses the residuals from the first tree as its input.

where p_i^m is the sigmoid transformation of $\hat{y}_i^m$. Given that the size of the additions to the overall $\hat{y}_i$ should decrease with successive decision trees, the loss can be expanded as a Taylor series about $\hat{y}_i = \hat{y}_i^{m-1}$,

$$l(y_i, \hat{y}_i^m) = l(y_i, \hat{y}_i^{m-1}) + (\hat{y}_i^m - \hat{y}_i^{m-1}) \frac{\partial l(y_i, \hat{y}_i^m)}{\partial \hat{y}_i} \Big|_{\hat{y}_i = \hat{y}_i^{m-1}} + \dots \quad (3.22)$$

Noting that $\hat{y}_i^m - \hat{y}_i^{m-1} = \alpha f_m(\mathbf{x}_i)$ (Equation 3.20) and defining the gradient as

$$g_i^{m-1} = \frac{\partial l(y_i, \hat{y}_i^m)}{\partial \hat{y}_i} \Big|_{\hat{y}_i = \hat{y}_i^{m-1}}, \quad (3.23)$$

and the hessian as

$$h_i^{m-1} = \frac{\partial^2 l(y_i, \hat{y}_i^m)}{(\partial \hat{y}_i)^2} \Big|_{\hat{y}_i = \hat{y}_i^{m-1}}, \quad (3.24)$$

the total loss can be approximated as

$$L(y, \hat{y}^m) = \sum_i^N [l(y_i, \hat{y}_i^{m-1}) + g_i^{m-1} \alpha f_m(\mathbf{x}_i) + \frac{1}{2} h_i^{m-1} (\alpha f_m(\mathbf{x}_i))^2], \quad (3.25)$$

and it can be shown that

$$g_i^{m-1} = \left. \frac{\partial l(y_i, \hat{y}_i^m)}{\partial \hat{y}_i^m} \right|_{\hat{y}_i = \hat{y}_i^{m-1}} = y_i - p_i^{m-1}. \quad (3.26)$$

The gradient being related to the residuals in this way is known as gradient boosting or gradient descent.

Regularization terms may be added to Equation 3.25 to slow the rate of change of the loss in order to avoid over-fitting. For BDTs, the most useful kind of loss term is known as L2-regularization and adds a term of the form $\frac{1}{2}\lambda \sum_j^L w_j^2$ where λ is a hyperparameter controlling the size of this correction, w_j is the weight (raw score) of leaf node j and L is the total number of leaves. This term discourages leaves from having very large weights (very high positive or negative scores), which can be symptomatic of over-fitting. In LightGBM, the leaf weights are determined by minimizing the loss (i.e. requiring $\frac{\partial L}{\partial w_j} = 0$) from which it can be shown the leaf weights are equal to

$$w_j = -\frac{G}{H + \lambda}, \quad (3.27)$$

where G is the sum of gradients of the events in the leaf and H is the sum of Hessians.

Hyperparameters

A couple of hyperparameters controlling the construction of the BDT have already been mentioned; the learning rate α and the L2-regularization parameter λ . In addition to these, several hyperparameters exist to prevent a BDT from becoming too complex to avoid over-fitting. A general discussion of these follows, after which specific details of the names of these hyperparameters in LightGBM is given for reference.

The most straightforward way to prevent over-fitting is to place restrictions on how individual trees can grow, for instance by limiting the maximum depth and/or the maximum number of leaves a single decision tree can have. Both options exist in LightGBM (with trees ceasing to grow once either stopping condition is reached). Similarly, it is also desirable to limit the overall number of decision trees. The number of trees can be set explicitly via a hyperparameter, or alternatively early-stopping can be used in which trees stop being added once performance on a (hold-out) validation set ceases to improve (after a set number of trees). Another hyperparameter to control

over-fitting is by requiring there to be a minimum number of (raw) events in a leaf node.

Relatedly, in certain BDT packages including LightGBM, it is possible to put restrictions on the splitting conditions, specifically via a parameter known as the minimum sum hessian. As mentioned, the hessian is defined as the second derivative of the loss function with respect to the prediction. For the binary cross entropy, it can be shown that

$$h_i^m = \frac{\partial^2 l(y_i, \hat{y}_i^m)}{(\partial \hat{y}_i^m)^2} = (1 - p_i^m)p_i^m. \quad (3.28)$$

When event weights are included in the training, each term is multiplied by its respective weight. The hessian is therefore always positive for any event (provided its weight is positive) and the sum of the Hessians for the events in a given leaf j , H_j^m , is related to the number of events in the leaf and the strength of the predicted probabilities for the events. For an individual event, h_i^m is largest for $p_i^m = 0.5$ and is relatively small when p_i^m is close to zero or one. Therefore, by requiring H_j^m to have a minimum value for a given leaf, it prevents leaf nodes with very high confidence (in which most events have p_i^m close to zero or one) from existing. Such leaf nodes are more likely to occur later in the training and can result in over-fitting, hence it is useful to put a restriction on them.

For reference, the hyperparameters used in the LightGBM models presented in this thesis are:

- `learning_rate`: the α parameter described above.
- `max_depth`: controls how many levels can be added to the individual decision trees.
- `max_leaves`: limits the maximum number of leaf nodes that a single decision tree can have.
- `min_data_in_leaf`: minimum number of (raw) events required in a leaf node.
- `n_estimators`: the number of individual decision trees. When using early stopping, this represents an upper bound on the number of trees that could be added.
- `l2_regularization`: controls the λ regularization parameter discussed above.

- `min_sum_hessian_in_leaf`: minimum sum hessian in a node required for a further split to be made.
- `feature_fraction_bynode`: what fraction of the kinematic variables are available for choosing a split at each node. This can be useful in allowing sub-dominant features (i.e. features that do not necessarily offer the greatest reduction in loss for that node) to be used which may have important correlations with other input features.
- `early_stopping_rounds`: when using early stopping, how many trees can be added without performance on the validation dataset improving before training is stopped.
- `max_bin`: number of bins used for constructing histograms of the input features (see below).
- `min_data_in_bin`: minimum number of raw events in each bin during the construction of the input feature histograms (see below).

Other hyperparameters not mentioned here take their default values for the models in this thesis.

Advantages of LightGBM

In addition to the large range of hyperparameters available, LightGBM offers a couple of other key benefits compared to other BDT software packages. The first is that the input features are binned in to histograms with the selections made at each node restricted to use values at the bin boundaries. This offers a significant performance speed up compared to looking through the whole dataset to choose optimum cut values. It also provides a way to restrict the granularity with which the distributions are examined, as it is possible for the user to pre-define the binning of the input features or control the number of bins through the `max_bin` and `min_data_in_bin` hyperparameters mentioned above. The other main advantage of LightGBM is that the individual decision trees grow leaf-wise rather than depth-wise; during the growth of particular decision tree, the leaf with the largest performance gain is split first rather than splitting leaves based upon their depths. As trees are not normally grown to their full depth (due to hyperparameter restrictions), leaf-wise growth increases the likelihood

Chapter 3: Analysis Methods

that good performing nodes are added first prior to reaching the stopping conditions. The models presented in the next chapter will make use of these advantages.

Chapter 4

Search for Direct Stau Production

“Oh God! It’s the machines. They’re starting to take over.”

John Connor, Terminator 3: Rise of the Machines (2003)

The methods described in the previous chapter will now be put to use in a search for direct stau production; a possible signature of Supersymmetry with unique motivations and challenges. This search will utilise machine learning to improve the sensitivity compared to previous searches for the same signature. The search presented here uses the full 139 fb^{-1} Run 2 dataset with centre of mass energy of 13 TeV collected by the ATLAS experiment between 2015-2018 and has been published in a paper by the ATLAS collaboration [1]. This chapter is structured as follows; Section 4.1 introduces the direct stau signal and the motivation for this search after which Section 4.2 provides a high level overview of the analysis strategy. The remaining sections then describe the analysis in depth; Section 4.3 details the data and simulated event samples used in the analysis; Section 4.4 the event reconstruction, Section 4.5 the signal region optimization including information of the machine learning models used; Section 4.6 the background estimation strategy; Section 4.7 the systematic uncertainties; Section 4.8 the results and Section 4.9 the conclusions from performing this search.

4.1 Introduction

The prediction by Supersymmetry (SUSY) of the existence of additional particles leads to many possible experimental signatures to search for at the Large Hadron Collider (LHC). One such signature would be an excess of events with two oppositely charged tau-leptons and significant E_T^{miss} arising from R -parity conserving supersymmetric models in which staus ($\tilde{\tau}$) are produced via a Z-boson and decay into standard model tau-leptons and $\tilde{\chi}_1^0$, with the $\tilde{\chi}_1^0$ being the lightest SUSY particle (LSP). Being electrically neutral and stable, the only trace left by a $\tilde{\chi}_1^0$ would be through contributing to the E_T^{miss} . Such a production mode is what is referred to as direct stau production. A Feynman diagram of direct stau production is shown in Figure 4.1.

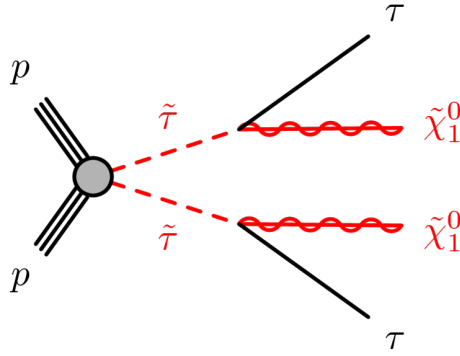


Figure 4.1: Feynman diagram illustrating direct stau production. The grey circle represents multiple diagrams in which two staus are produced from electroweak bosons produced by the proton-proton collision.

In addition to the general motivations for SUSY described in Chapter 1, supersymmetric models where the lightest neutralino is the LSP and the stau has a relatively light mass (of order 100 GeV) are particularly compelling as stau-neutralino co-annihilation occurring in the early universe could lead to a dark matter relic density consistent with cosmological observations as described in reference [100] (with the $\tilde{\chi}_1^0$ being the dark matter candidate). Furthermore, although in the minimal supersymmetric standard model (MSSM), discussed at the end of Chapter 1, the different slepton flavours (electrons, muons and staus) are assumed to be close in mass, in larger Grand Unified Theories this degeneracy does not hold and the lighter of the two stau mass eigenstates is often the next-to-lightest SUSY particle after the lightest neutralino [101]. Hence, if SUSY exists, direct stau production may be one of the first accessible signatures at particle colliders.

Given these motivations, the ATLAS and CMS collaborations have previously conducted searches for direct stau production. The latest results using the full Run 2 dataset (excluding the search presented here) can be found in references [102](ATLAS) and [103](CMS). For convenience, these searches are designed using simplified models of direct stau production in which the $\tilde{\tau}$ and $\tilde{\chi}_1^0$ are the only accessible SUSY particles at the LHC (the other SUSY particles have their masses set above 100 TeV) and the $\tilde{\tau}$ decays with 100% branching fraction to the standard model τ and a $\tilde{\chi}_1^0$. These searches are sensitive to models with $\tilde{\tau}$ masses between 100 and 400 GeV and $\tilde{\chi}_1^0$ masses between 0 and 200 GeV. The masses of the $\tilde{\tau}$ determines the predicted production cross-section while the masses of both the $\tilde{\tau}$ and $\tilde{\chi}_1^0$ effect the shapes of kinematic variables.

Both of the latest ATLAS and CMS searches targeted a final state in which both of the standard model tau-leptons decay hadronically (the “had-had” final state). Final states in which one or both of the staus decay leptonically are of course possible, and indeed CMS has conducted a search using the “lep-had” final state [104] (in which one of the taus decays hadronically and the other leptonically), but they suffer from unique challenges such that they have limited sensitivity compared to the all hadronic channel. To review first the had-had channel, the main SM backgrounds for this final state are: multi-jet QCD events where two jets are misidentified as hadronically decaying tau-leptons (referred to here-on as the multi-jet background); Z+jets (predominantly from $Z \rightarrow \tau\tau$); W+jets (where the W-boson decays to a hadronic tau-lepton and a jet fakes the other hadronic tau-lepton); backgrounds originating from a hard scatter with one or more top quarks (e.g. $t\bar{t}$ and single-top processes) in which W-bosons from the top decays themselves decay into tau-leptons, collectively referred to as the ‘top’ background and multi-boson backgrounds (primarily di-boson events with WW, WZ and ZZ production). There is also a small contribution from Higgs bosons decaying to tau-leptons. Compared to analogous searches for direct smuon and selectron production, the main challenges with stau production in the had-had final state are the large multi-jet background; the lower reconstruction efficiency of tau-leptons compared to electrons and muons; the larger systematic uncertainties associated with tau-lepton reconstruction and identification, the higher p_T thresholds of the hadronic tau-lepton triggers compared to electron and muon triggers (due to smaller rates at which jets can fake electrons and muons) and the reduction in effective cross-section from requiring both of the tau-leptons to decay hadronically.

Considering briefly the challenges of the other tau-lepton decay channels, for the “lep-had” final state, the main challenge is the large background from W+jets which is

enhanced compared to its presence in the had-had channel due to the ability of the W -boson to decay directly to electrons and muons. Although the multi-jet background is greatly reduced in the lep-had channel compared to the had-had channel, the W +jets background is more problematic as it contains real E_T^{miss} from the neutrinos and so is harder to discriminate from the signal. The state with two leptonically decaying tau-leptons is impacted by the small leptonic tau-lepton branching fraction; tau-leptons decay leptonically 35% of the time and hadronically 65% of the time, hence the “lep-lep” final state is only 12% of total branching fraction of two tau-leptons, but the fake backgrounds are greatly reduced compared to the other final states.

Though the non had-had channels have in general more limited sensitivity compared to the had-had channel due to these challenges, the lower p_T thresholds of the electron and muon triggers compared to the hadronic tau-lepton triggers mean that these channels have much better sensitivity to models which have smaller mass differences between the $\tilde{\tau}$ and $\tilde{\chi}_1^0$ as such models result in tau-leptons with lower p_T . In particular, as will be seen in the rest of this chapter, the had-had final state is not sensitive to models with $\Delta(m(\tilde{\tau}), m(\tilde{\chi}_1^0)) < \sim 70$ GeV. The non had-had channels will be necessary to obtain sensitivity to these compressed models. Further, combinations between the different channels could also enhance the overall sensitivity, particularly for moderately compressed models for which the had-had channel has some sensitivity. The rest of this chapter though concerns only the had-had final state.

The most recent ATLAS search in the had-had final state utilised two unbinned signal regions; one targeting ‘low mass’ staus the other ‘high mass’ staus; whilst the most recent CMS search used a multi-bin approach consisting of 29 signal region bins split across three kinematic variables. Neither of these searches saw any significant discrepancies with respect to the SM predictions and as such were able to set exclusion limits on the simplified SUSY models for different values of stau and neutralino-one mass ($m(\tilde{\tau})$ and $m(\tilde{\chi}_1^0)$ respectively). Limits are computed for the case where $\tilde{\tau}_L$ and $\tilde{\tau}_R$ are mass degenerate (referred to as $\tilde{\tau}_{L,R}$ production) and also for $\tilde{\tau}_L$ and $\tilde{\tau}_R$ production separately. The production of right-handed staus has a smaller cross-section than the production of left-handed staus due to the smaller couplings of the $\tilde{\tau}_R$ with Z -bosons. The ratio of $\tilde{\tau}_L$ $\tilde{\tau}_L$: $\tilde{\tau}_R$ $\tilde{\tau}_R$ production is roughly 2.5:1.

The 95% confidence level exclusion contours for $\tilde{\tau}_{L,R}$ and $\tilde{\tau}_L$ only production from the first ATLAS full Run 2 search are shown in Figure 4.2. Comparable limits were obtained by the CMS full Run 2 search, although it had slightly greater sensitivity at lower $m(\tilde{\tau})$. For $\tilde{\tau}_{L,R}$ production, models for which $90 \text{ GeV} < m(\tilde{\tau}_{L,R}) < 400 \text{ GeV}$

and $m(\tilde{\chi}_1^0) \approx 0$ are excluded at 95% CL while for $\tilde{\tau}_L$ production models for which $115 \text{ GeV} < m(\tilde{\tau}_L) < 350 \text{ GeV}$ is excluded. For $m(\tilde{\tau}_{L,R}) = 300 \text{ GeV}$, $\tilde{\chi}_1^0$ masses up to 150 GeV are excluded. As mentioned above, limits for $\tilde{\tau}_R$ production are significantly weaker. The first full Run 2 ATLAS search was not able to set any (observed) limits on $\tilde{\tau}_R$ production at 95% confidence level while the CMS search was able to obtain the first (observed) sensitivity to $\tilde{\tau}_R$ production for $150 \text{ GeV} < m(\tilde{\tau}_R) < 250 \text{ GeV}$ with neutralino masses up to 50 GeV excluded for $m(\tilde{\tau}_R) = 200 \text{ GeV}$. In addition to the ATLAS and CMS limits, models with $m(\tilde{\tau}) < 87 \text{ GeV}$ have been excluded by searches performed by the experiments at the Large Electron-Positron (LEP) collider ([105], [106], [107], [108]) for $\tilde{\chi}_1^0$ masses up to 80 GeV . Considering the latest ATLAS limits only extend as low as $m(\tilde{\tau}_{L,R}) = 110 \text{ GeV}$, this means there is a sensitivity gap between the ATLAS limits and the LEP limits.

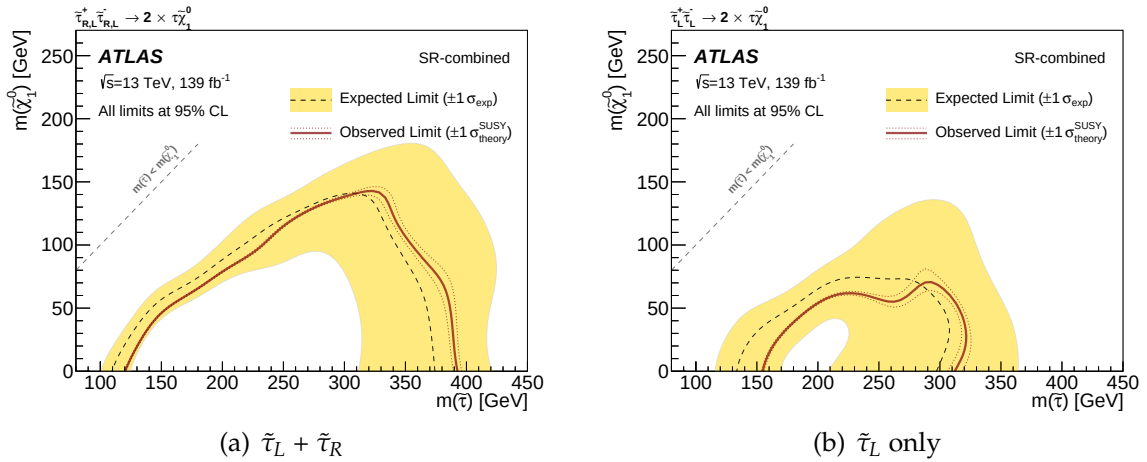


Figure 4.2: 95% CL exclusion limits set by the previous ATLAS search for direct stau production, for (a) $\tilde{\tau}_L + \tilde{\tau}_R$ and (b) $\tilde{\tau}_L$ only production. Taken from reference [102].

The aim of the new search presented here is to improve the sensitivity with respect to the previous ATLAS search (for all values of $m(\tilde{\tau})$ and $m(\tilde{\chi}_1^0)$) using the same dataset but employing machine learning algorithms, in particular Boosted Decision Trees (BDTs), trained to separate the signal from the SM backgrounds, to define the signal regions. Additionally, the new search also takes advantage of the improvements in hadronic tau-lepton identification coming from the use of a recurrent neural network algorithm developed by ATLAS [76], discussed in Chapter 2, Section 2.2.4, since the previous search which significantly reduces the fake multi-jet background.

4.2 Overview of Analysis Strategy

As mentioned in the introduction, the direct stau signal features two tau-leptons of opposite electric charge (sign), which here are required to decay hadronically, and large E_T^{miss} in the final state. These criteria therefore form the basis for the event selection in the signal regions (SRs). The asymmetric di-tau trigger, which requires two hadronically decaying tau-leptons with different p_T thresholds, is used to select these events. The invariant mass of the two tau-leptons is also required to be significantly greater than the Z-boson mass in order to suppress the background from Z+jets. Similarly, the stransverse mass (introduced in Chapter 3, Section 3.1.8) of the two tau-leptons, with the mass of invisible particles set to zero, is required to be greater than 30 GeV to suppress the multi-jet background. Events with reconstructed muons, electrons, additional tau-leptons or b -tagged jets are vetoed. On top of these baseline requirements, the primary SR selection is based upon BDT classifier scores. As the kinematics of the direct stau signal vary with both $\tilde{\tau}$ and $\tilde{\chi}_1^0$ mass, four separate BDTs are trained on groups of signals with different ranges of $m(\tilde{\tau})$ and $m(\tilde{\chi}_1^0)$ with the grouping determined via a dedicated clustering algorithm. The selections on the classifier scores are optimized to maximize the significance to the respective target mass scenarios, with two bin shape fits being employed in three out of four of the SRs to further increase the sensitivity. The SRs are not made explicitly orthogonal to each other.

All the SM backgrounds except the multi-jet background are estimated using Monte Carlo (MC) simulation with Control Regions (CRs) being defined for each of the W+jets, Z+jets and top backgrounds to determine their normalisations. The W+jets and top CRs require a muon in place of one of the hadronic tau-leptons to separate themselves from the SRs, with the top CR further requiring events to have a b -tagged jet. The Z+jets CR meanwhile is separated from the SR by inverting the selection on the invariant mass of the two tau-leptons. For the multi-jet background, a data-driven method is employed which utilises CRs separated from the SR by requiring looser tau-lepton identification criteria, dropping the opposite-sign requirement and inverting the stransverse mass requirement. The normalisations for the multi-boson and Higgs backgrounds are taken from theoretical predictions. For each of the backgrounds except the Higgs background, a Validation Region (VR), orthogonal to the SRs, is defined to check the accuracy of the prediction. Additionally, an inclusive VR, also orthogonal to the SRs, is defined to verify the compatibility of the different background

estimation strategies together. Experimental and theoretical uncertainties are evaluated for the signal and background processes. Statistical fits are performed using the HistFitter [84] software package to test the signal hypotheses. Details of all these aspects are provided in the following sections.

4.3 Data and Simulated Event Samples

As just described in the overview, simulation using Monte Carlo (MC) event generators is used to model the kinematics of all the SM backgrounds aside from the multi-jet background. MC is also used for the kinematics of the direct stau signals and to compute cross-sections for the SUSY processes; multi-boson and Higgs boson backgrounds and the different components of the top background. The event generators used for the matrix element (ME) calculations, modelling the parton showers and the sets of parton distribution functions (PDFs) and tuning used are summarised in Table 4.1 and are discussed in more detail below. The samples used in the analysis were prepared by various collaborators.

The direct stau signals are generated using MadGraph [2.2.3] [109] with ME calculation being performed at tree level. The ME calculation includes the emission of up to two additional partons. PYTHIA [8.186] [110] is used for modelling the parton shower hadronisation, and underlying event using the NNPDF2.3LO PDF set with the ATLAS A14 tune [111]. RESUMMINO v2.0.1 is used to calculate the signal cross-sections to next-to-leading order (NLO) in the strong coupling constant, with resummation of soft gluon emission added at next-to-leading-logarithm accuracy (NLL) [112, 113]. The cross-section uncertainty is calculated from an envelope of cross-section predictions using different PDF sets and factorisation and renormalization scales, following the procedure described in reference [114]. Stau masses between 80 GeV and 560 GeV are generated for neutralino-one masses between 1 GeV and 240 GeV. Cross-sections vary from 0.80 pb for $m(\tilde{\tau}_{L,R}) = 80$ GeV to 0.30 fb for a $m(\tilde{\tau}) = 560$ GeV. Uncertainties on these cross-sections vary between two and three percent. Left and right-handed staus are produced with the same mass in each sample but, for interpreting the non-degenerate mass case, whether the staus are left or right-handed is recorded for each event.

The Z+jets and W+jets backgrounds are generated with SHERPA [2.2.1] [115, 116]. The ME calculations include up to two additional partons at NLO and four addi-

tional partons at leading order (LO), using the Comix [117] and OPENLOOPS [118, 119] generators merged with the SHERPA parton shower [120] following the MENLOPS prescription [116] with the NNPDF3.0NNLO [121] PDF set. The nominal normalisation of the Z+jets and W+jets backgrounds are taken from the next-to-next-to-leading order (NNLO) cross-sections in reference [122].

The multi-boson background consists of events in which two or three vector bosons ($V = [W, Z]$) are produced and are also generated using SHERPA. Specifically, SHERPA [2.2.1] at NLO is used as the generator for tri-boson and di-boson processes where the bosons decay semi-leptonically while SHERPA [2.2.2] is used for di-boson processes where the bosons decay fully-leptonically. Multiple matrix elements are matched and merged with the SHERPA parton shower based on Catani–Seymour dipole factorisation [120] using the MENLOPS prescription [116, 123–125]. Virtual QCD corrections for matrix elements at NLO accuracy are provided by the OPENLOOPS library [119]. As for the Z+jets and W+jets processes, the NNPDF3.0NNLO set is used with the dedicated tuning developed by the SHERPA authors. The cross-sections used for the multi-boson processes have been computed to NLO by the ATLAS collaboration, using of a variety of theoretical calculations and programs [126–130].

Several processes involving top quarks contribute to the combined top background, the dominant processes being the production of top quark pairs ($t\bar{t}$) and single-top production in the Wt –, s – and electroweak t -channels. Rarer processes including the production of a top quark pair alongside a vector boson or Higgs boson are also included as are events simultaneously producing three or four top quarks. POWHEG-BOX [v2] [131–134] is used to generate the $t\bar{t}$ and single-top processes while Mad-Graph[2.2.2] [135] at NLO is used for the rarer processes. For both sets of processes PYTHIA [8.186] is again used for the parton shower, fragmentation, and the underlying event using the NNPDF2.3LO [136] PDF set and A14 tune. For all top processes, the top-quark mass is set to 172.5 GeV. Although normalised in a dedicated CR, the relative normalisations of the different top-quark containing processes are important in accurately modelling the overall top background. The cross-section prediction from TOP++ 2.0 [137–143], which is NNLO in QCD with resummation of next-to-next-to-leading logarithmic (NNLL) soft-gluons, is used to normalise the $t\bar{t}$ sample. For the single-top processes, the NLO in QCD with NNLL soft gluon corrections cross-section from [144, 145] is used for the Wt -channel while the NLO in QCD cross-section from [144, 145] is used for the t - and s -channels.

Events with Higgs bosons also provide a small contribution to the total SM background. POWHEG-BOX [v2] is used to model Higgs bosons produced via gluon–gluon fusion and vector boson fusion with PYTHIA [8.186] used for the parton shower and NNPDF3.0NNLO as the PDF set with the AZN tune. PYTHIA [8.186] is used to generate events with a Higgs boson with a vector boson while MadGraph is used for events in which a Higgs boson is produced with a $t\bar{t}$ pair. Cross-sections from Ref. [146] are used to normalise the Higgs boson processes.

For all processes, PYTHIA [8.186] [147] is used to simulate additional proton-proton interactions on top of main interaction to account for pile-up. For this modelling, the NNPDF[2.3lo] LO PDFs [136] are used with the A3 set of tuned parameters [148]. Fast simulation, described at the end of Chapter 2, is used to model the response of the ATLAS detector for the signal samples while the full simulation based on GEANT4 [82] is used for the background samples, except for a few minor EWK Z+jet processes with very small yields in the phase space considered here.

As mentioned at the beginning of this chapter, the dataset to which the SM predictions are compared to in this analysis corresponds to 139 fb^{-1} of proton-proton collision data collected between 2015 and 2018 by the ATLAS detector, at a centre-of-mass energy of 13 TeV. The data quality requirements outlined at the end of Chapter 2 are imposed to ensure that only events in which the entire ATLAS detector was functioning well are used [149].

4.4 Event Reconstruction

The event reconstruction follows the general prescriptions detailed in Chapter 2, Section 2.2.4, including the primary vertex and data quality requirements. Specific details of which reconstruction and identification (ID) algorithms are used for the objects in this analysis, and their relevant working points (WPs) are given below. These definitions follow recommendations from relevant ATLAS working groups. For the leptons, two classes of objects are given; *baseline* objects for use in the overlap removal steps also described below and *signal* objects used in the analysis region selections. Here, and in the rest of this chapter, tau-leptons refer specifically to hadronically decaying tau-leptons (as discussed in Chapter 2 leptonically decay tau-leptons are reconstructed as the electrons or muons they decay into).

Chapter 4: Search for Direct Stau Production

Process	ME Genera- tor	Parton Shower	PDF set	Tune
Direct stau sig- nals	MadGraph 2.2.3	PYTHIA 8	NNPDF2.3LO	A14
VV, VVV	SHERPA 2.2.1/2.2.2	SHERPA	NNPDF3.0NNLO	SHERPA default
$Z/W + \text{jets}$	SHERPA 2.2.1	SHERPA	NNPDF3.0NNLO	SHERPA default
$t\bar{t} + V, t\bar{t} + H, 3$ top, 4 top	MG5_AMC	PYTHIA 8	NNPDF2.3LO	A14
$t\bar{t}, s/t\text{-channel}$ and Wt single top	POWHEG- BOX	PYTHIA 8	NNPDF2.3LO	A14
$t\bar{t}H, VBFH, gg +$ H	POWHEG- BOX	PYTHIA 8	NNPDF3.0NNLO	AZN
$Z/W + H$	PYTHIA 8	PYTHIA 8	NNPDF2.3LO	A14

Table 4.1: Details of Monte Carlo generators used in the direct stau analysis. The second column denotes the generator used for modelling the hard scatter at the matrix element (ME) level; the third the generator used for modelling the parton shower, the fourth the set of parton distribution functions (PDFs) used and the fifth how the PDFs are tuned to better agree with data.

The algorithms, WPs and requirements on the physics objects used in the direct stau search are as follows:

- **Jets:** Reconstructed using the particle-flow algorithm using anti- k_t calorimeter cluster inputs with a radius parameter of 0.4. Required to have $p_T > 20$ GeV and $|\eta| < 2.8$. The medium WP of the jet vertex tagging algorithm is used to reject jets originating from pile up.
- **b-Jets:** The DL1r algorithm is used to tag jets containing b -hadrons. The WP used has a reconstruction efficiency of 77% and has a rejection factor for light (u – or d -type) jets of 140.
- **Electrons:** Baseline leptons are required to have $p_T > 4.5$ GeV, $|\eta| < 2.47$, longitudinal impact parameter $|z_0 \sin \theta| < 0.5$ mm and must pass the loose WP of the electron-ID algorithm. Signal electrons are further required to have $p_T > 25$ GeV, have transverse impact parameter significance $|d_0/\sigma(d_0)| < 5$, pass the tight electron-ID WP and satisfy loose isolation criteria.

- **Muons:** Baseline muons are required to have $p_T > 3 \text{ GeV}$ and $|\eta| < 2.7$, $|z_0 \sin \theta| < 0.5 \text{ mm}$ and pass the medium WP of the muon-ID algorithm. Signal muons must also have $p_T > 25 \text{ GeV}$, $|d_0/\sigma(d_0)| < 3$ and like electrons must pass loose isolation criteria.
- **Tau-leptons:** baseline tau-leptons are required to have $p_T > 20 \text{ GeV}$, $|\eta| < 2.5$ and must be associated with one or three Inner Detector tracks. Baseline tau-leptons are also required to pass the loose WP of the RNN tau-identification (ID) algorithm, while signal tau-leptons must pass the medium WP, which has a signal efficiency for real tau-leptons of 75% and a background rejection factor of 240¹. Tau-leptons associated to one Inner Detector track (1-prong tau-leptons) must also pass the medium WP of the BDT discriminant used to distinguish them from electrons. Additionally, tau-leptons are required to lie outside the transition region between the barrel and end-cap calorimeters ($1.37 < |\eta| < 1.52$).
- E_T^{miss} : The tight WP is used for jet component of the E_T^{miss} calculation.

Using the baseline objects, the following overlap removal steps are applied: first jets with properties suggesting they are likely to be a prompt muon (low number of tracks, sum of track p_T similar to p_T of muon candidate, little calorimeter energy around the muon candidate) are removed. Tau-lepton candidates that have $\Delta R < 0.2$, where $\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2}$ with electron or muon candidates are then removed. Electron candidates sharing an Inner Detector track with muon candidates are also removed, as are jets within $\Delta R < 0.2$ of the remaining electron candidates. Further, electron candidates within $\Delta R < r(p_T^e)$ where

$$r(p_T^e) = \max\left(0.04 + \frac{10 \text{ GeV}}{p_T^e}, 0.4\right) \quad (4.1)$$

and p_T^e is the p_T of the electron candidate, of jet candidates are also removed. Similarly, jets within $\Delta R < 0.2$ of muon candidates are removed followed by the removal of muon candidates with $\Delta R < r(p_T^\mu)$ of the remaining jet candidates. Lastly, any remaining jet candidates within $\Delta R < 0.2$ of tau-lepton candidates are removed.

¹ The use of the medium WP of the RNN tau-ID algorithm for the tau-leptons was chosen prior to the author joining the analysis based upon it offering the best balance between S/B and total statistics, the latter factor being particularly important given the use of machine learning models for the signal region selections.

4.5 Event Selection

This section describes the event selection of the Signal Regions (SRs) and is divided into three parts. The first part considers the triggers and baseline selections which are not dependent on the BDT classifiers. The second part then describes the setup of the BDT classifiers in detail. Lastly, the third part explains how selections on the resulting BDT scores are optimized to define the final SRs.

4.5.1 Triggers and Baseline Selection

Events entering the SRs are required to pass the asymmetric ditau trigger, which requires two hadronic tau-leptons with different p_T thresholds. The ‘ditau + E_T^{miss} trigger’, which in addition to two hadronic tau-leptons requires a large E_T^{miss} , was also considered but was found to add very little additional yield to the selections considered here due to the high offline E_T^{miss} threshold of 150 GeV needed to be on the efficiency plateau. For the event selection, the offline reconstructed tau-leptons are required to have transverse momenta above certain thresholds, detailed in Table 4.2, that are greater than the nominal High Level Trigger (HLT) thresholds to ensure that the events selected are on the trigger efficiency plateau. The thresholds for the second tau-lepton p_T are higher in 2018 as for this data period the tau triggers included full tracking information instead of fast tracking. The plateau requirement is also enforced in the control and validation regions that use the asymmetric ditau trigger, except for the regions used in the construction of the multi-jet estimate (discussed in Section 4.6.1). Trigger matching between the online tau-leptons used to pass the trigger and those reconstructed offline is also required.

Object	Year	HLT	Offline
leading tau p_T [GeV]	2015-2018	80	95
2nd leading tau p_T [GeV]	2015-2017	50	60
	2018	60	75

Table 4.2: HLT and offline thresholds on the tau-lepton transverse momenta required to ensure that the selected events are in the plateau region of the corresponding trigger leg.

In addition to the trigger requirements, events entering the SRs are required to pass a baseline set of selections. The final SRs definitions then include selections on the BDT

scores on top of these requirements. The baseline requirements are listed in Table 4.3. In accordance with the direct stau signal, the two highest p_T tau-leptons are required to be of opposite sign (OS), and events with an additional baseline tau-lepton, any baseline electrons or any baseline muons are vetoed. To reduce the background from top events, a b -jet veto is required. The invariant mass of the two tau-leptons $m_{\tau\tau}$, is required to be greater than 120 GeV to suppress the background from $Z \rightarrow \tau\tau + \text{jets}$ (and to a lesser extent the backgrounds from $h \rightarrow \tau\tau$ and multi-boson events). The m_{T2} variable computed between the two tau-leptons, with the mass of invisible daughter particles set to zero, is required to be above 30 GeV to reduce the multi-jet background. Following these selections, all signal events have $E_T^{\text{miss}} > 20$ GeV and separation between the two tau-leptons $\Delta R(\tau_1, \tau_2) < 4$, hence these requirements are also applied. Tables detailing the number of signal, background and data events passing each of these selections (and the final SR selections) sequentially can be found in the appendix, Section A.2.5.

Common Signal Region Selection
pass asymm. ditau trigger
asymm. ditau trigger p_T threshold requirements
two OS signal tau-leptons
N baseline taus == 2
N baseline electrons == 0
N baseline muons == 0
N b -jet == 0
$m_{\tau\tau} > 120 \text{ GeV}$
$m_{T2}(\tau_1, \tau_2) > 30 \text{ GeV}$
$E_T^{\text{miss}} > 20 \text{ GeV}$
$\Delta R(\tau_1, \tau_2) < 4$

Table 4.3: Common selection requirements for events entering the direct stau SRs. The p_T threshold requirements of the asymm. ditau trigger refer to the selections in Table 4.2 that ensure events are on the trigger’s efficiency plateau. OS denotes that the two signal tau-leptons are required to have opposite sign electric charge.

4.5.2 Machine Learning Details

As described in the overview, four BDTs are trained on separate groups of signal samples with different ranges of $m(\tilde{\tau})$ and $m(\tilde{\chi}_1^0)$. Each of the BDTs is a binary signal/background classifier, with the individual signals within each group being combined to form a single signal sample². This subsection describes the BDTs in detail and makes significant reference to the concepts described in the latter half of Chapter 3. For clarity, the signal grouping used for the final models is shown here in Figure 4.3 but how this grouping is determined is not discussed until later in this subsection. Broadly, group 1 consists of models with low $m(\tilde{\tau})$ and small mass differences between $m(\tilde{\tau})$ and $m(\tilde{\chi}_1^0)$, known as compressed models; group 2 of non-compressed models with medium $m(\tilde{\tau})$, group 3 of compressed models with medium $m(\tilde{\tau})$ and group 4 of models with high $m(\tilde{\tau})$. Some points are not used for training any of the BDTs; the reasoning for this is also explained later in this subsection.

The rest of this subsection is structured as follows; first aspects common to all the BDTs used in this analysis are discussed; including the samples used for training and evaluation, how these samples are weighted with respect to each other and the kinematic variables that are used as input features. This is then followed by the discussion of how the signal grouping was obtained. After this, attributes particular to the optimization of the final models trained using the finalized signal groups are discussed, namely restrictions that are applied during training to improve their reliability, the cross evaluation setup and how the hyperparameters are optimized.

Common Aspects of all the BDTs

LightGBM [99], described at the end of the previous chapter, is the machine learning package used for the BDTs. Binary cross-entropy (Equation 3.17) is used as the objective function. The signal and background samples detailed in Section 4.3 are used for the training and evaluation of the BDTs (and the overall yield predictions). For the signal samples, events with left and right-handed staus are used together in the BDT training as they have very similar kinematics. A simplified version of the data-driven estimate of the multi-jet background which contains no negatively weighted events (which as

² An alternative approach would have been to have trained a single machine learning model that was parameterized by $m(\tilde{\tau})$ and $m(\tilde{\chi}_1^0)$. Using a parameterized BDT was briefly studied but was found to give worse performance than the individual trained models, thought to be due to the limited number of background MC events detailed later in this section.

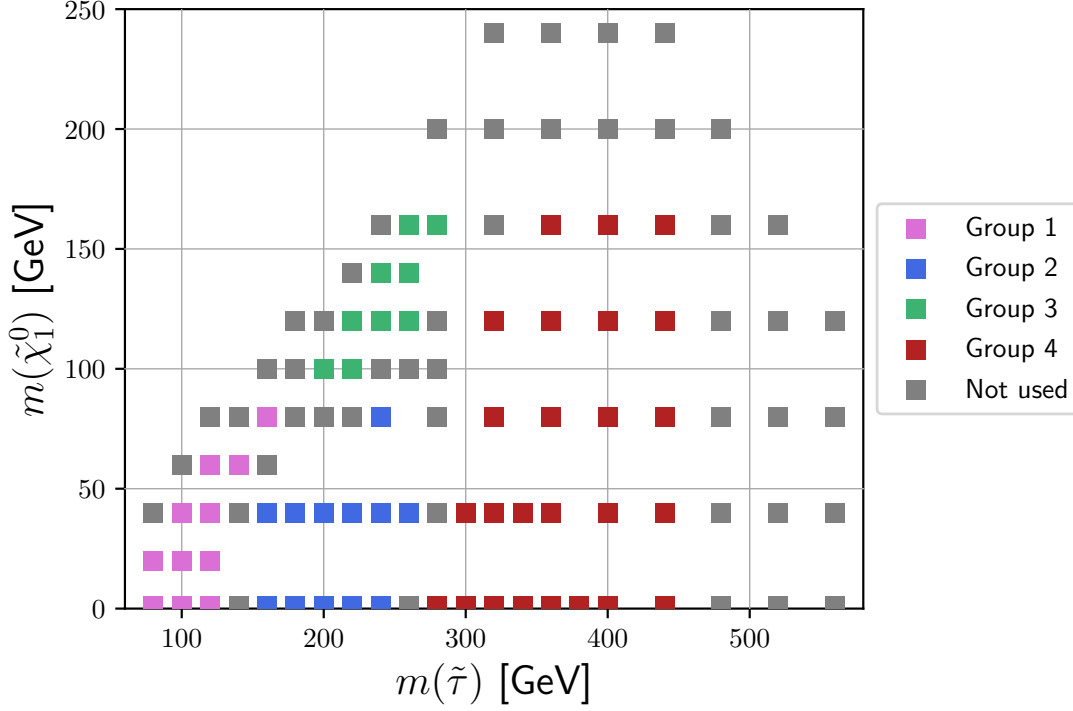


Figure 4.3: Grouping of the signal samples used to train the BDTs in the direct stau analysis. As is described in more detail later in this section, points in grey are deliberately not included in the training to increase the reliability of the interpolation of the sensitivity between the respective signal regions defined by each BDT score.

explained in Chapter 3, Section 3.2 can bias ML models) is used during training in place of the full multi-jet estimate. This simplified estimate is detailed in Section 4.6.1. Normalisation factors for the W +jets and top backgrounds, as determined from a background-only fit to the W +jets and top CRs, described in Section 4.6, are applied to ensure that the background composition used for training the BDTs is accurate (the BDTs were trained prior to the addition of a Z +jets control region, hence no normalisation factor was applied to Z +jets in the training).

The BDTs are trained with events that pass the common SR selection discussed in the previous subsection. The number of raw and weighted background events passing this preselection is shown in Table 4.4, alongside yields for each of the signal groups. For the signal groups, the weighted yield is the average of the samples within each group and is included for illustration of the total signal to background ratio for each group, while the raw number of events is the sum of all the samples in the group. From these yields a reasonably large class imbalance in the number of raw events is

evident, with signal groups 1, 2 and 4 having notably more raw events than the total SM background. Due to this imbalance, the problem outlined in Chapter 3, Section 3.2 where ML models can isolate regions of phase space with artificially enhanced S/B is of particular concern. This serves as the primary motivation for the restrictions that are applied during the training of the final models, discussed in the third part of this subsection, which limit the extent to which this isolation can occur. Looser preselections for training the BDTs were also tested to see if greater statistics could improve their performance. Significant gains in the performance of the final models were not found, which is thought to be an additional consequence of the training restrictions that are applied, as is also explained in the third part of this subsection.

Sample	Weighted Yield	No. Raw Events
Multi-jet	193.6 ± 2.6	5,564
Z+jets	142 ± 25	1,251
W+jets	105 ± 20	610
Top	28.1 ± 2.0	246
Multi-boson	29.4 ± 1.8	1,839
Higgs	2.9 ± 0.5	508
SM total	501 ± 33	10,018
Signal Group 1	65.9 ± 0.5	44,260
Signal Group 2	58.9 ± 0.6	19,943
Signal Group 3	23.2 ± 0.4	4,188
Signal Group 4	15.21 ± 0.11	45,611

Table 4.4: Weighted yields (at 139 fb^{-1}) and number of raw events passing the BDT preselection requirements for each of the background processes and signal sample groups. For the multi-jet background, the yield is from the simplified estimate discussed in Section 4.6.1. For the signal groups, the weighted yield is the average from the signals included in the group while the raw number of events is the sum of all the signals.

There are three important aspects of the relative weighting of the samples used during training. Firstly, the individual event weights are used in the training to ensure the distributions of the input variables (listed shortly) are accurate (as the kinematic shapes between the weighted and raw distributions in general differ). Negatively weighted events are not included in the training to avoid a bias towards them. To ensure the performance of the BDTs is not over-tuned to MC events with large weights, all MC events with an event weight greater than two are not included

in the training and subsequent test-train comparisons. The threshold of two was determined by first starting at a threshold of five and reducing this in integer steps until good agreement between training and validation sets (with events with weights larger than the threshold omitted) for an example BDT could be achieved. As the different background processes contain different fractions of negative and largely weighted events, the relative weights of the background samples used in the training are re-scaled such that the relative proportions of the different backgrounds with the negative and largely weight events omitted is the same as with them included.

These weight restrictions only have notably effect for the W +jets and Z +jets backgrounds (both of which were generated with SHERPA-2.2.1). For W +jets, 21/610 raw events have event weights greater than two while 23% of events are negatively weighted. The Z +jets background is less adversely effected having 7/1251 events with weight greater than two and 16% of events with a negative weight. The other backgrounds (and the signal samples) have no events with weights greater than two and have less than 8% of events with a negative weight. Although the shapes of the kinematic distributions are likely less accurate with these negatively weighted events omitted, this only effects the optimality of the BDTs, not their validity (as no restrictions on event weights are applied for the final background predictions, i.e. all events are included in the evaluation stage).

Secondly, for each signal group the weights of each of the signals in the group is scaled such the total weight of each signal is the same. This means each of the BDTs focuses on separating each of the signal points within its group from the background evenly and that the kinematics within each signal group are averaged. Thirdly, the weight of the combined signal sample is further scaled such that its total weight matches that of the background so good performance can be achieved using the binary cross-entropy as the objective function.

Sixteen kinematic variables are used as input features for the BDTs:

- $p_T(\tau_1)$: transverse momentum of the leading tau-lepton (the tau-lepton with the highest p_T).
- $p_T(\tau_2)$: transverse momentum of the sub-leading tau-lepton.
- E_T^{miss} : magnitude of missing transverse momentum.
- $\Delta\phi(\tau_1, \tau_2)$: absolute difference in ϕ between the two tau-leptons.

- $\Delta\phi(\tau_1, E_T^{\text{miss}})$: absolute difference in ϕ between the leading tau-lepton and the $\vec{p}_T^{\text{miss}}$.
- $\Delta\phi(\tau_2, E_T^{\text{miss}})$: absolute difference in ϕ between the sub-leading tau-lepton and the $\vec{p}_T^{\text{miss}}$.
- $\Delta\eta(\tau_1, \tau_2)$: absolute difference in η between the two tau-leptons.
- $\Delta R(\tau_1, \tau_2)$: separation of the two tau-leptons, equal to $\sqrt{(\Delta\phi(\tau_1, \tau_2))^2 + (\Delta\eta(\tau_1, \tau_2))^2}$.
- $m_{\tau\tau}$: invariant mass of the two tau-leptons.
- $m_T(\tau_1)$: transverse mass (defined in Chapter 3, Equation 3.12), of the leading tau-lepton with the $\vec{p}_T^{\text{miss}}$.
- $m_T(\tau_2)$: transverse mass of the sub-leading tau with the $\vec{p}_T^{\text{miss}}$.
- $m_T(\tau_1) + m_T(\tau_2)$: sum of the transverse masses of the leading and sub-leading tau-leptons with the $\vec{p}_T^{\text{miss}}$.
- $m_{T2}(\tau_1, \tau_2)$: transverse mass (defined in Chapter 3, Equation 3.15) of the two tau-leptons with the mass of invisible daughter particles set to zero.
- m_{eff} : effective mass, equal to the scalar sum of the transverse momenta of the two tau-leptons plus the E_T^{miss} .
- $\Delta\phi(\tau_1, E_T^{\text{miss}}) - \Delta\phi(\tau_2, E_T^{\text{miss}})$: difference between $\Delta\phi(\tau_1, E_T^{\text{miss}})$ and $\Delta\phi(\tau_2, E_T^{\text{miss}})$.
- $\frac{E_T^{\text{miss}}}{p_T(\tau_1) + p_T(\tau_2)}$: ratio of the E_T^{miss} to the scalar sum of the tau p_T s.

Comparisons of the shapes of each of these variables between example signals from each of the groups and the total SM background can be found in the Section A.2.1 of the appendix. Jet dependent variables are not included as the jet multiplicity is found to be poorly modelled in W +jets, and for the signal samples jets are not modelled at the matrix element level. Including additional m_{T2} variables with different masses for the invisible daughter particles of 40 GeV and 100 GeV was also tested but their inclusion was not found to offer a significant performance increase.

Signal Grouping

Having established aspects that are common to all the BDTs, how the grouping shown in Figure 4.3 was arrived at can be discussed. The first step is to create a 2-dimensional

embedding of the signal data using the UMAP [150] package. In relatively simple terms, this package creates a low dimensional graph representation of the data which preserves its topological structure. Using this embedding, the general procedure used to group the signal samples is as follows:

1. Within the reduced dimensionality space of the embedding, the euclidean distances between the centres of each signal sample are computed, the centres being defined by the average X and Y co-ordinates of the sample in this 2D space.
2. (Optional) The sample space is split into two halves; compressed models and non-compressed models.
3. N ‘seed’ signal points are chosen as the N points for which the averaged minimal distance of each of the chosen points from another of the chosen points is maximized, where N is the desired number of BDTs chosen by the user (i.e. choosing N signal points that are on average far apart from each other kinematically). These N points define the different sample groups. If splitting into compressed and non-compressed halves, this step is done separately for each half.
4. The remaining signal points are added to the group which they are closest to (as determined by the smallest average distance from the centre of the group). The signal point which is closest to any of the groups centres is the first to be added, after which the group centres are re-calculated and the process repeated.
5. Samples on the borders between the clusters are removed from the cluster they belong to. The points which are removed are those (out of pairs of points on each border) that are the furthest away from either of the group centres.

The decision whether to split into compressed and non-compressed halves is based on prioritizing the sensitivity to the compressed models or optimizing for the whole sample space evenly. The last step in the procedure is to check that there is not a significant reduction in sensitivity in samples that are kinematically similar to those used to train the BDTs, but that are not actually used in the training themselves. Additionally, points on the outside of the sample grid are also not included in the training for the same reason.

The final clustering (previously shown in Figure 4.3) arises from using the compressed/non-compressed splitting with $N = 2$ for each half. Groups 1 and 3 are from the clustering of the compressed mass points while groups 2 and 4 are from the clustering of the non-compressed points. Different groupings were tested, namely having one or three

clusters for each of compressed and non-compressed halves instead of two as well as testing $N = [3,4,5]$ without splitting the sample space into the compressed and non-compressed halves. To compare the performance of different groupings, for each grouping BDTs were trained for each cluster and the maximum significance scores obtainable by cutting on the BDT scores (i.e. requiring events to have a BDT score above a certain threshold), were computed for the signal samples in that cluster. This procedure is shown explicitly for the final models in the next subsection. For these tests, the standard ATLAS formula for estimating significance [151] was used with a flat 20% uncertainty combined in quadrature with the MC statistical uncertainty on the background. The value of 20% is based on the actual systematic error on the total background yield at BDT preselection level.

The BDTs used in testing the different groupings used a manually tuned set of hyperparameters detailed in Table 4.5 that was found to give good performance without significant over-fitting across the signal groups. These BDTs did not have the training restrictions that are applied to final models (discussed below) and cross evaluation was not used in computing the significant scores (they were computed on the training and test sets combined). Despite this, the comparison of the maximum significance scores between different models is still valid as the models were not significantly over-fit.

Hyperparameter	Value
max_depth	3
num_leaves	8
min_sum_hessian_in_leaf	100
min_data_in_bin	50
min_data_in_leaf	50
learning_rate	0.1
l2_regularization	1000
feature_fraction_bynode	0.5
Early Stopping	Yes
n_estimators	1500
early_stopping_rounds	10

Table 4.5: Set of hyperparameters for the BDTs used to compare the different signal grouping schemes (not the final models).

From testing the different grouping schemes with the compressed/non-compressed splitting, it was found that there was a significant gain in the obtainable significance scores in going from one to two clusters for each half, while the improvement in going from two to three clusters was minimal. Similarly, when not splitting into compressed/non-compressed halves, the clustering with $N = 4$ offered the best balance of sensitivity. The performance of the final grouping and the $N = 4$ grouping was similar but the final grouping had slightly better performance for the compressed models with $m(\tilde{\tau}) \geq 200$ GeV, (for which the sensitivity is close to the 95% confidence level), hence this grouping scheme was chosen for the final models.

Optimization of Final Models

As mentioned, for the final models additional restrictions are applied during training in order to reduce the extent to which regions of artificially enhanced S/B may be isolated. The first restriction is that the input variables to the BDT are pre-binned with bins defined such that (for events passing the BDT preselection) the statistical error is less than 30% in each bin. The number of bins for each variable depends upon how the event weights are distributed and ranges from 11 to 27, with an average of 16. Along with this pre-binning, the `max_depth` hyperparameter is set to one such that each tree only makes a single selection on one variable. The combination of these two restrictions prevents the BDTs from deliberately isolating regions of parameter space where the background prediction is extremely unreliable. As the pre-binning is determined from the events passing the common preselection (which is necessary as this is the phase space in which the BDTs are ultimately evaluated), this limits the potential gain of using a loosened preselection for training.

Restricting the maximum depth of the BDTs to one is certainly not ideal; a more dynamic solution would be a requirement that all leaf nodes have background yields with an MC statistical uncertainty less than 30%. However, at present such an option is not available (in LightGBM or other common BDT packages). While the isolation effect discussed above may in part be limited through use of the `min_sum_hessian_parameter` (which is the approach taken for a study in the next chapter), given the imbalance in the number of raw signal and background statistics here a more explicit restriction is deemed to be necessary.

To optimize the hyperparameters for the final models, a grid scan is performed over the `learning_rate`, `feature_fraction_by_node` and `l2_regularization` hyper-

parameters with the training restrictions described above applied. The scan ranges and intervals for these parameters are detailed in Table 4.6. Early stopping is again used with `n_estimators = 1500` and `early_stopping_rounds = 10`. For this grid scan, and for the final models, 3-fold cross evaluation is used with a validation set, hence for each signal group three BDTs are trained (for each hyperparameter configuration). Only three (equal size) folds are used due to the limited number of statistics available (recall from Chapter 3 it is important that the evaluation and validation sets have reasonable statistics so that they can accurately evaluate a model's performance). For the final models, events are split into the respective folds by their event number (which is randomised for MC events) such that events are evaluated by the same BDTs across different kinematic based systematic uncertainties.

During the grid scan, the average Area-Under-Curve (AUC) score is computed for the validation set (averaged across the 3 folds) and a 10 bin chi-squared test is performed in the BDT score between the 3 pairs of training and validation set distributions (for both the signal and background) for each hyperparameter configuration. The best hyperparameter configurations for each signal group (and the corresponding best BDTs) are selected as the configurations yielding the largest averaged validation set AUC score that also pass the requirement of the total (across all the folds) chi-squared p-value for the background and signal BDT score distributions being greater than 0.6 and 0.4 respectively. This requirement is to ensure the chosen models are not significantly over-fit. For the BDTs trained on Signal Groups 2 and 4, the best AUC score is not found to be significantly greater than the AUC score of the hyperparameter configuration with the highest background chi-squared p-value (which correspond to the models for which there is the greatest confidence that they are not significantly over-fit), so for these signal groups this hyperparameter configuration is chosen. The hyperparameter configuration used for the BDTs for each signal group is detailed in Table 4.7.

Hyperparameter	Scan Range/Intervals
<code>learning_rate</code>	0.01 - 0.1, in steps of 0.01
<code>l2_regularization</code>	1 - 1000, 10 steps geometrically spaced
<code>feature_fraction_bynode</code>	0.25 - 1.0, in steps of 0.25

Table 4.6: Hyperparameters scanned over in a grid search for the direct stau BDTs.

Signal Group	learning_rate	feature_fraction_by_node	l2_regularization
1	0.1	0.5	46
2	0.04	0.25	1000
3	0.01	1.0	21
4	0.01	0.25	1000

Table 4.7: Best BDT hyperparameter configuration for each signal group.

Figure 4.4 shows comparisons of the BDT score distributions between the training and evaluation sets for one of the three final BDTs trained using each signal group (distributions for the other two BDTs for each signal group can be found in Appendix A.2.2). In these plots, what is labelled as the training set is the actual training set plus the validation set while the test set is the evaluation set. For all the BDTs, the signal is the positive class such that a high BDT score indicates that an event is more likely to be signal. As multiple signals are used to train each BDT, the signal in these plots is composed of all the signals used in the training together, scaled to the average weight of one signal. Events with weights larger than two are not included in these distributions as these events are not included in the BDT training. The agreement is generally good and indicates the models are not significantly over-fit. The distributions for the BDTs trained on Signal Groups 2 and 4 occupy a smaller fraction of the total BDT score range (which spans from 0 to 1) due to the larger value of the `l2_regularization` hyperparameter.

Receiver operating characteristics for these BDTs are shown in Figure 4.5 and AUC scores for each of the fold splits for the final models are detailed in Table 4.8 where it can be seen the performance across the folds is similar. It is also evident that the BDTs trained on signal group 4 have the best discrimination power as they have the largest AUC scores, whilst those trained on signal group 1 have the weakest discrimination power, meaning these signals are the hardest to separate from the SM background.

As outlined in Chapter 3, the use of cross evaluation allows all the data used to train the BDTs to be used in the evaluation without incurring a bias (as in the evaluation set, the BDT score which is used from an event is from the BDT for which it was not part of the training set). For convenience, in the rest of this chapter, the scores from the BDTs used to train the three different folds on signal group X are ‘combined’ to from a single classifier score, BDT X score, where the combination involves, for each event, using the score from the BDT for which the event was in the evaluation set.

Chapter 4: Search for Direct Stau Production

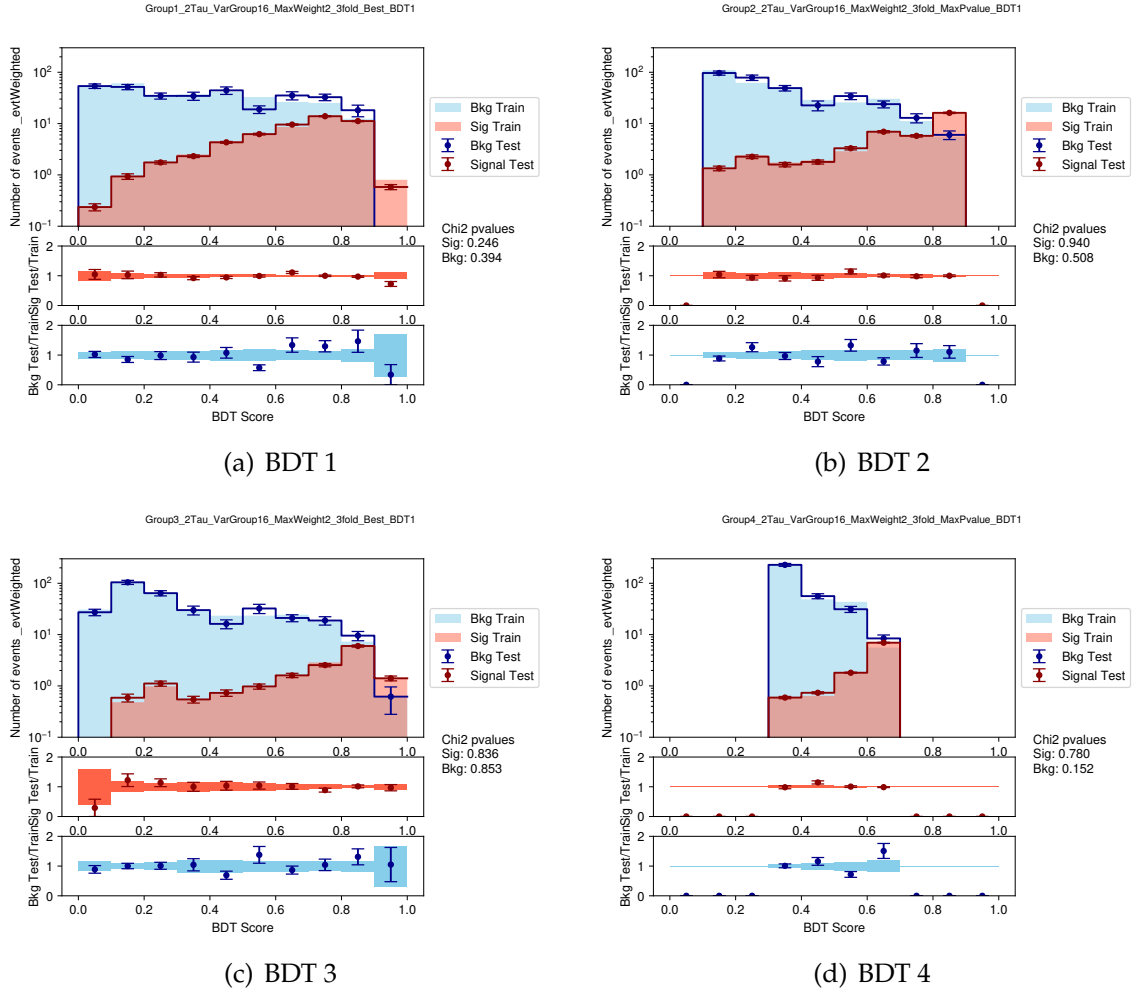


Figure 4.4: Example train-test comparisons for the BDTs trained on the different signal sample groups. All of these comparisons are using the same fold of the dataset for the evaluation set each time (here labelled as the test set). The training set in these plots includes the validation set.

To demonstrate this is valid, comparisons of the BDT distributions for the evaluation sets for each set of BDTs are shown in Figure 4.6 for the total SM background and an example signal for each set. From these it can be seen that the BDT score distributions are consistent within the statistical uncertainties and thus the scores from each fold can be safely ‘combined’. Note these plots do not include negative or largely weighted events as they were not included in the training but versions of these plots including these events can be found in the appendix, Section A.2.2 (in which agreement is also seen but the statistical uncertainties are notably larger).

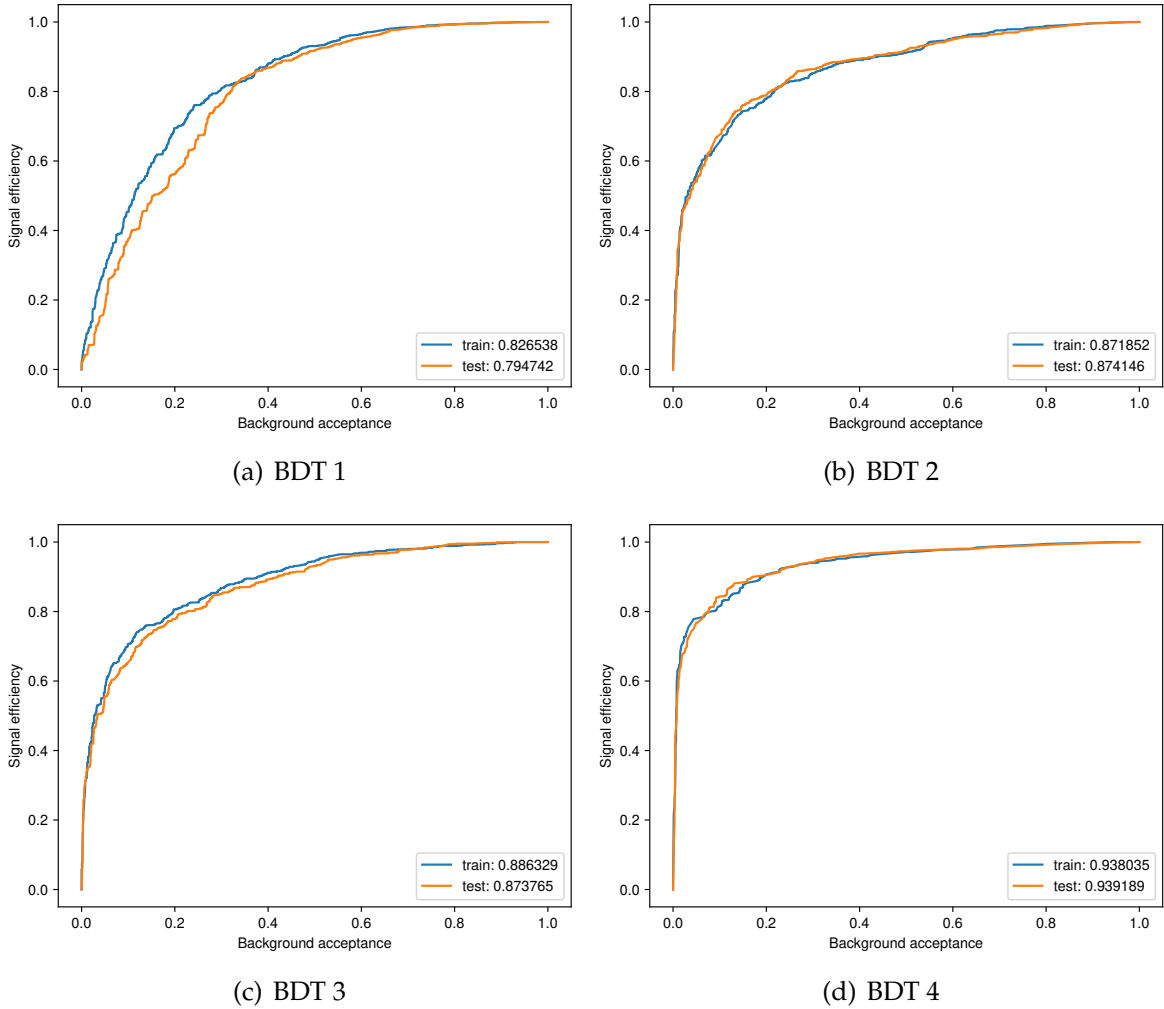


Figure 4.5: Example ROC curves for the BDTs trained on the different signal sample groups. All of these comparisons are using the same fold of the dataset for the evaluation set each time (here labelled as the test set). The training set in these plots includes the validation set.

Chapter 4: Search for Direct Stau Production

Signal Group	Split 1	Split 2	Split 3
1	0.827 (0.795)	0.834 (0.786)	0.823 (0.803)
2	0.872 (0.874)	0.882 (0.863)	0.879 (0.870)
3	0.886 (0.874)	0.900 (0.862)	0.886 (0.880)
4	0.938 (0.939)	0.944 (0.932)	0.940 (0.941)

Table 4.8: Area under curves for the BDTs trained on the different signal groups across the different train/evaluation splits for the training+validation (evaluation) datasets.

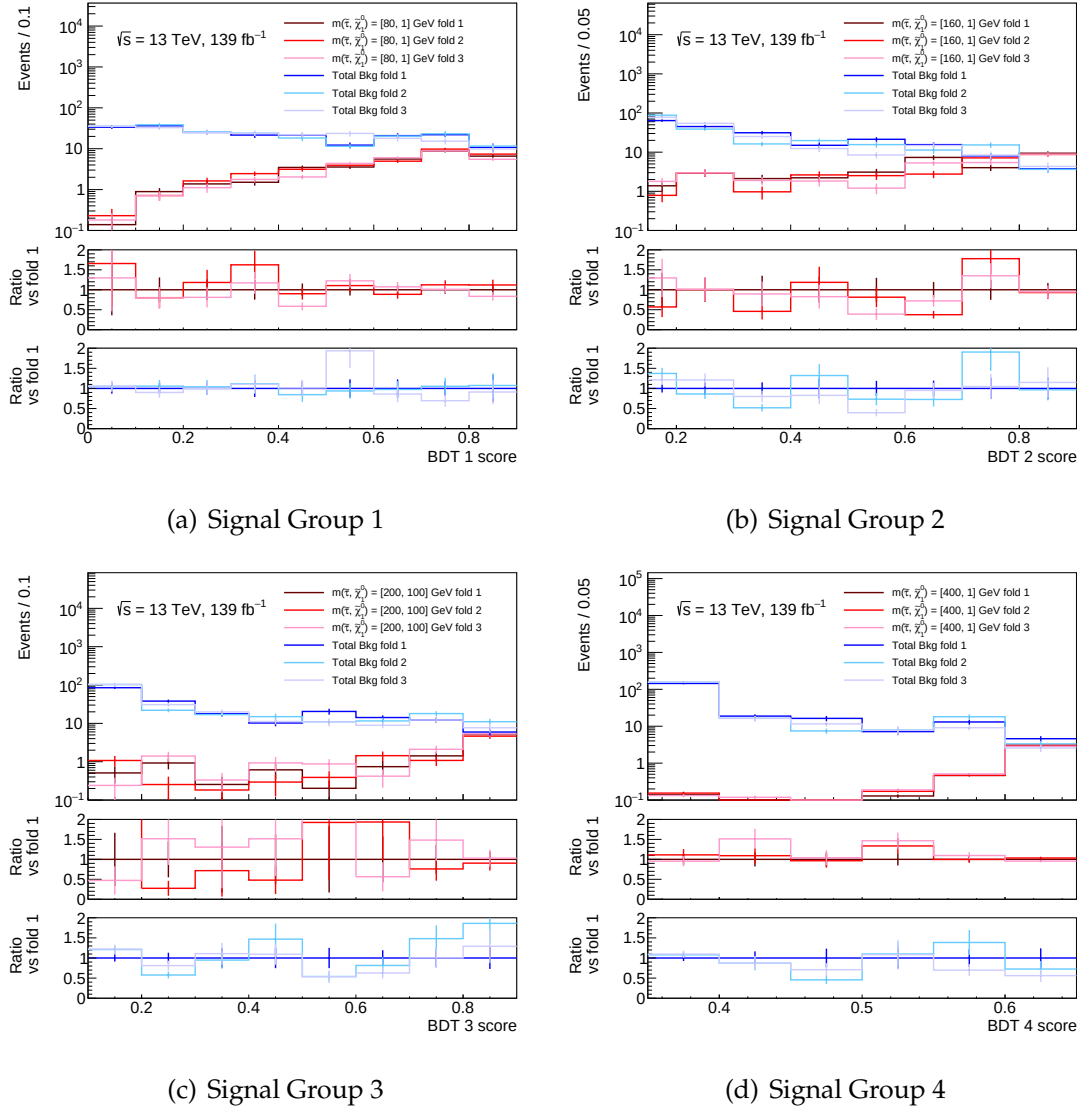


Figure 4.6: Comparison of the evaluation sets across the different train/evaluation splits, for each set of BDTs.

Chapter 4: Search for Direct Stau Production

The combined BDT score distributions of the final models for the total SM background and example signal points for each BDT are shown in Figure 4.7. The percentage purity, S/B, for the signals in each bin is also shown, demonstrating the discrimination power of the BDTs. In particular, for each signal group the BDTs are effective in reducing the large multi-jet and Z+jets backgrounds. The W+jets background is the main background at high BDT 1 score, while for the other scores the multi-boson background is dominant at high values. The bottom panel of each plot shows the ratio of signal to background yield (S/B) in each bin in for the example signals.

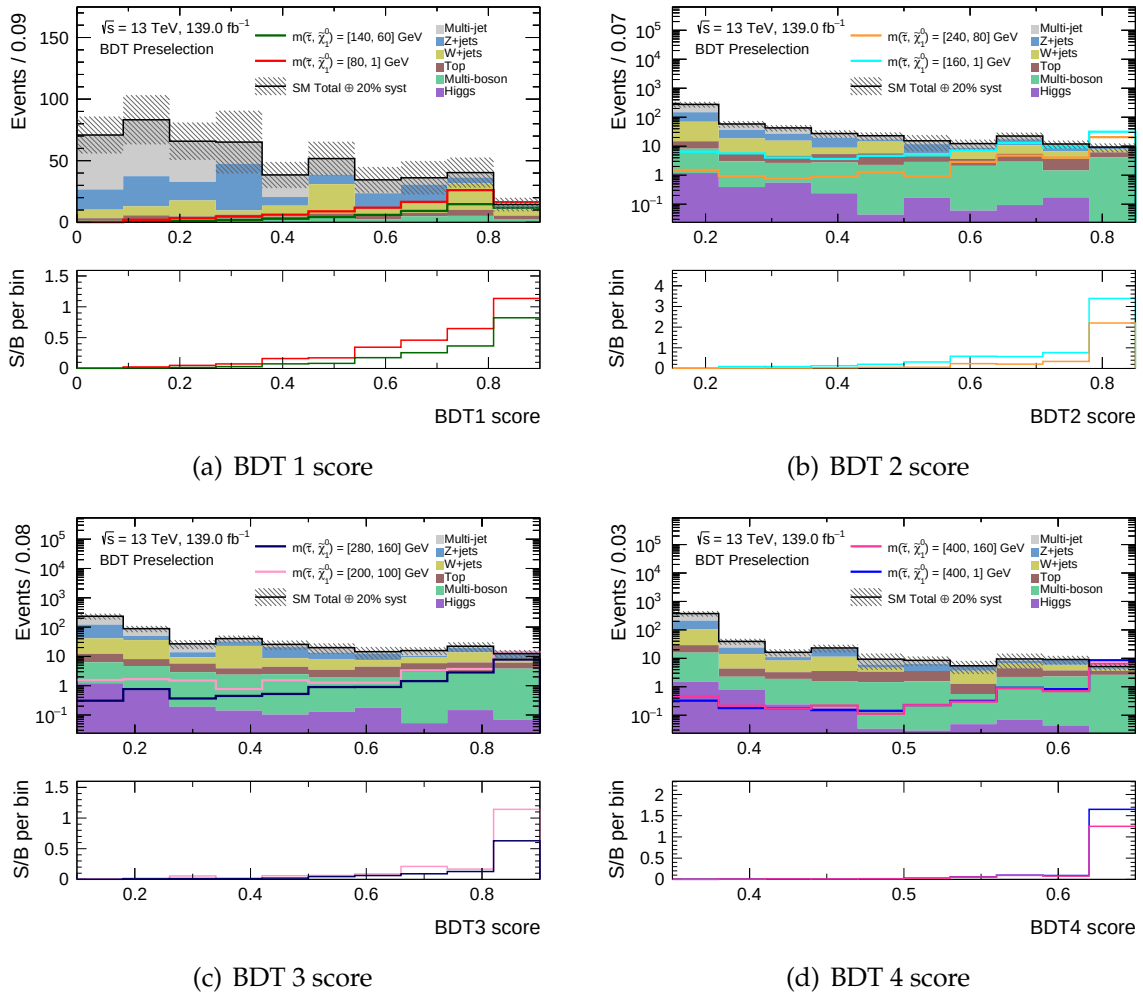


Figure 4.7: Combined BDT score distributions for the final models for the total SM background and example signal points for each BDT. In these, and subsequent similar plots in this thesis, the different background processes are stacked on top of each other while the yields from the possible SUSY signals are overlaid. A flat 20% systematic uncertainty, added in quadrature to the MC statistical uncertainty, is shown on the SM background. The bottom panel of each plot shows the ratio of signal to background yield (S/B) in each bin in for the example signals.

To help understand what the BDTs are selecting for, SHAP values, outlined in Chapter 3, Section 3.2, for an example BDT for each signal group are shown in Figure 4.8. SHAP values for the other fold splittings for each group were checked to verify the feature importances are consistent between the folds. As the signal is the positive class in all of these BDTs, a high SHAP score for the particular value an input feature takes indicates that the feature having such a value leads to the model classifying that event as being signal like. The discontinuous nature of these SHAP plots (i.e. the horizontal gaps for a given feature) is due to the restrictions applied to the BDTs during training, namely restricting the maximum depth to be 1 and that the input variables are pre-binned as together these restrictions limit the overall number of leaves in each of the models.

From Figure 4.8(a), it can be seen that for the low mass signals events with high $\Delta\phi(\tau, \tau)$, low $\Delta\eta(\tau, \tau)$, high $E_T^{\text{miss}} / (p_T^{\tau_1} + p_T^{\tau_2})$ and high $m_T(\tau_1) + m_T(\tau_2)$ are selected as being very signal like. For signal group 2 (in the bulk of the signal grid), Figure 4.8(b) shows that high $m_T(\tau_1) + m_T(\tau_2)$ is very signal like, as is just high $m_T(\tau_1)$. High m_{T2} and high E_T^{miss} are also very signal like. The BDTs trained on the high mass compressed signals in signal group 3 (Figure 4.8(c)) see high $m_T(\tau_1) + m_T(\tau_2)$ as a dominant feature, but also has significant signal like contributions from high $\Delta\phi(\tau, \tau)$, low $\Delta\eta(\tau, \tau)$, similar to the low mass compressed signals. The SHAP values for the BDTs trained on signal group 4 (Figure 4.8(d)) are similar to those trained on signal group 2 with high $m_T(\tau_1) + m_T(\tau_2)$, $m_T(\tau_1)$, m_{T2} and E_T^{miss} leading to events being classified as being very signal like.

Chapter 4: Search for Direct Stau Production

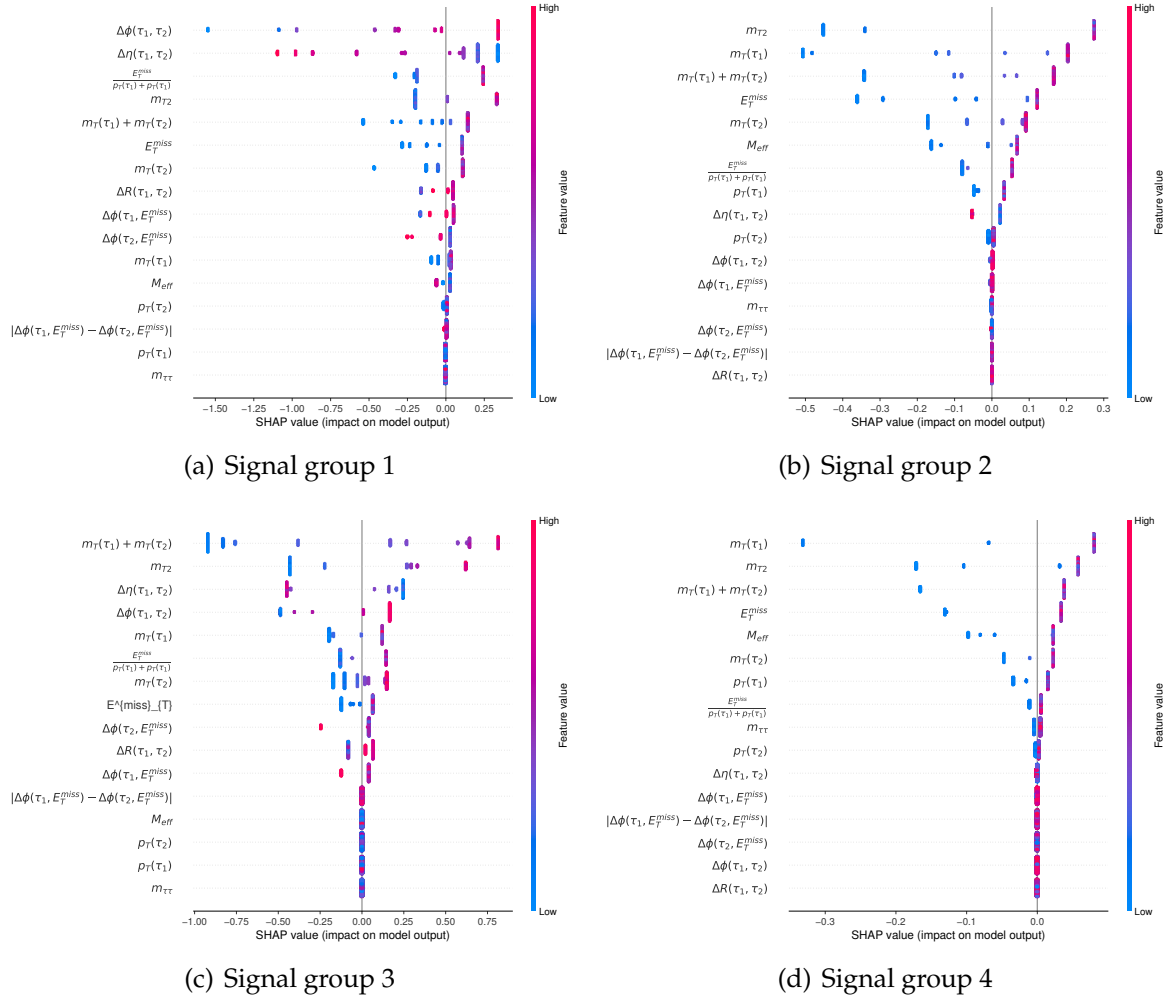


Figure 4.8: SHAP summary plots for the BDTs trained on the different signal sample groups. All of these comparisons are using BDTs with the same fold splitting.

4.5.3 Signal Region Optimization

Using these BDT classifiers, Signal Regions (SRs) are defined by cutting on each of the BDT scores and are labelled as SR-BDTX for the BDTs trained using signal group X. As mentioned in the strategy overview, the SRs defined by each BDT are not explicitly made orthogonal to each other. As was done when comparing the performance of the different signal groupings, the cut value for each BDT score is chosen to maximize the sensitivity to the signal samples used to train the respective BDTs, using the standard ATLAS significance formula [151] with the MC statistical uncertainty combined in quadrature with a flat 20% systematic uncertainty on the background. The statistical uncertainty on the total background estimation in each SR is also required to be $< 30\%$, to ensure the background prediction is reliable.

To determine the precise cut values, the BDT distributions are studied with a fine granularity as shown by the distributions in Figure 4.9. Jumps in the significance can be seen when passing events with moderately high weights. Where possible (i.e. without significantly comprising the sensitivity), the cut values are chosen to be away from such large jumps in regions of comparative stability. For SR-BDT1, 2 and 3 the signal regions are split into two bins while for SR-BDT4 only a single bin signal regions is used. The bin boundaries were determined by first cutting as tightly as possible on the significance plateaus to define the lower edge of the final bin. The lower edges of subsequent bins were then determined by requiring the background yield to approximately equal that of the final bin, in order to keep statistical and systematic uncertainties roughly constant across the bins (so that no one bin dominates the likelihood fit). The binning used, and the significance score in each bin are shown in Figure 4.10. From these plots it is evident that there is little sensitivity to be gained from adding lower bins to these signal regions. The resulting final signal regions definitions are shown in Table 4.9. As well as the binned SRs, discovery regions are defined by merging the bins for each of SR-BDT1, SR-BDT2 and SR-BDT3 (as discussed in the previous chapter, discovery regions are required to be unbinned). SR-BDT4 is also used as discovery region.

The background yields and signal yields for representative signal points in each of the SRs are shown in Table 4.10. From these yields it can be seen that W +jets is the dominant background in SR-BDT1, while the multi-boson background is dominant in SR-BDT2 and SR-BDT4. As W +jets comprises events for which one of the tau-leptons originates from a jet, there is a brief study in the appendix, Section A.2.3 assessing

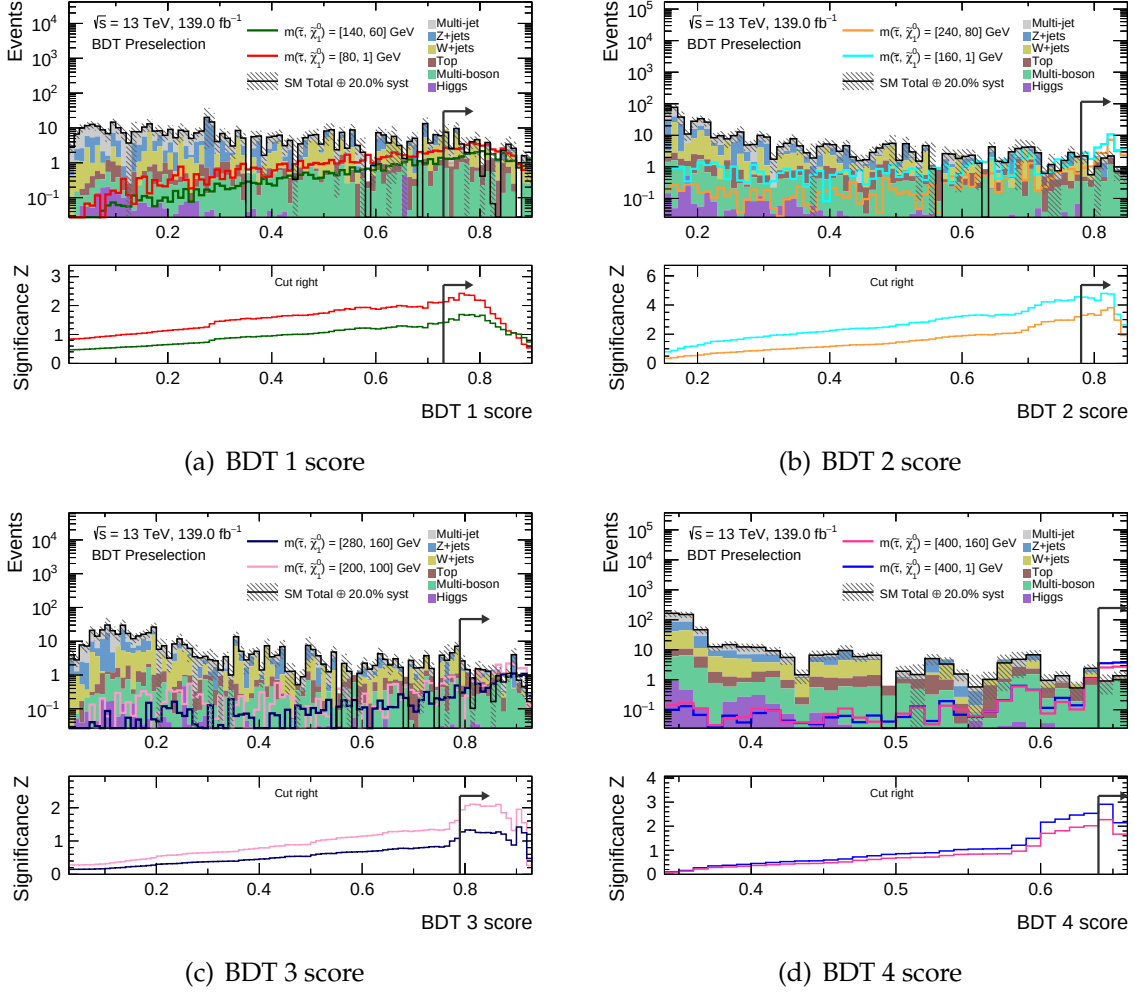


Figure 4.9: BDT score distributions with fine granularity showing the significance upon cutting to the right, using the MC statistical error plus a flat 20% systematic.

whether the use of the tight instead of the medium RNN tau-ID working point for the tau-leptons could further improve the sensitivity to signals with low $\tilde{\tau}$ -mass. Concerning the signal samples which look to fall out of the sensitivity reach of the strategy presented here (with an integrated luminosity of 139 fb^{-1}), for the models with low $m(\tilde{\tau})$ the sensitivity decreases with increasing $\tilde{\chi}_1^0$ mass as the tau-leptons become softer, and the number of signal events passing the trigger requirements decreases rapidly. For the models with high $m(\tilde{\tau})$, the limitation is mainly statistical arising from the small production cross-sections.

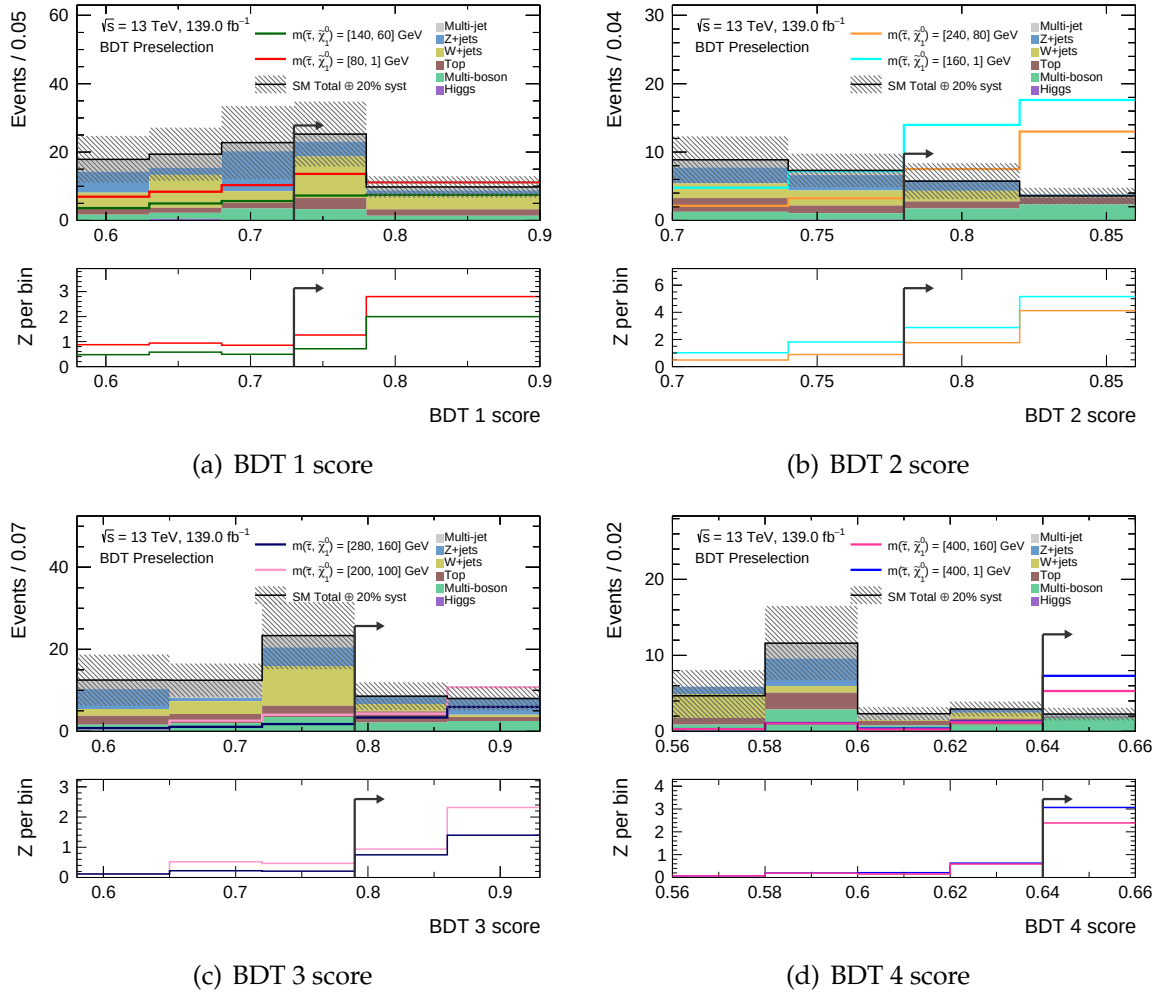


Figure 4.10: N-1 plots of the BDT score distributions for the different combined BDTs showing the binning used for the multi-bin fits and the significance score per bin (using the MC statistical error plus a flat 20% systematic). Errors bars are from the MC statistical uncertainties plus a flat 20% systematic error. The arrow indicates the lowest edge of the first bin used in the respective signal regions.

Chapter 4: Search for Direct Stau Production

SR BDT 1	SR BDT 2	SR BDT 3	SR BDT 4
asymm. ditau trigger with plateau requirements two OS signal τ 3 rd baseline τ veto light lepton veto b -veto $E_T^{\text{miss}} > 20 \text{ GeV}$ $m_{\text{inv}}(\tau_1, \tau_2) > 120 \text{ GeV}$ $m_{T2}(\tau_1, \tau_2) > 30 \text{ GeV}$			
BDT 1 score [0.73, 0.78]	> BDT 2 score [0.78, 0.82]	> BDT 3 score [0.79, 0.86]	> BDT 4 score 0.64

Table 4.9: Signal region definitions. Multiple values for the BDT1 score, BDT2 score and BDT3 score selections indicate the lower edges of the bins used for the two bin shape fits in these variables. OS denotes the tau-leptons must have opposite sign electric charge.

SM process	SR BDT 1	SR BDT 2	SR BDT 3	SR BDT 4
Multi-jet	4.6 ± 2.7	$0.0^{+2.0}_{-0.0}$	$0.0^{+2.1}_{-0.0}$	$0.2^{+0.4}_{-0.2}$
W+jets	23 ± 9	1.6 ± 1.0	2.3 ± 1.8	$0.0^{+0.4}_{-0.0}$
Z+jets	7.1 ± 3.0	2.0 ± 1.0	6.1 ± 2.3	0.25 ± 0.11
Top	7.8 ± 1.1	2.1 ± 0.6	3.7 ± 0.8	0.16 ± 0.14
Multi-boson	6.6 ± 0.9	4.1 ± 0.6	4.5 ± 0.7	1.8 ± 0.4
Higgs	—	0.021 ± 0.007	0.12 ± 0.05	—
SM total	49 ± 10	9.3 ± 2.5	17 ± 4	2.3 ± 0.7
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 1) \text{ GeV}$	40.3 ± 1.4	4.2 ± 0.4	10.8 ± 0.7	0.13 ± 0.07
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (240, 80) \text{ GeV}$	17.2 ± 0.9	20.5 ± 1.0	23.6 ± 1.1	10.4 ± 0.7
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (200, 100) \text{ GeV}$	15.4 ± 1.3	11.6 ± 1.1	15.1 ± 1.3	3.2 ± 0.6
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (400, 1) \text{ GeV}$	4.59 ± 0.08	9.46 ± 0.12	9.39 ± 0.12	7.31 ± 0.10

Table 4.10: Yields table for the direct stau signal regions. Normalisation factors, as determined from the background only fit discussed in Section 4.6 are applied to the W+jets, Z+jets and Top backgrounds. Uncertainties here are just MC statistical uncertainties.

4.6 Background Estimation

As was explained in the introduction to this chapter, the main SM backgrounds in this analysis are the multi-jet background, Z+jets, W+jets, the top background and multi-boson backgrounds with an additional small background from events with Higgs bosons decaying to two tau-leptons. Figure 4.11 shows the composition of the total SM background passing the BDT training preselection requirements (detailed in Table 4.3), from which it can be seen that the multi-jet, Z+jets and W+jets backgrounds comprise the largest portions. However, as was just shown in Table 4.10, the top and multi-boson backgrounds also have significant contributions in the SRs and hence are equally important to estimate accurately. As a reminder of the overall background estimation strategy, the Z+jets, W+jets and top backgrounds are normalised in dedicated control regions (labelled ZCR, WCR and TCR respectively) while the normalisation's for the multi-boson and Higgs backgrounds are taken from theoretical predictions, as detailed in Section 4.3. Dedicated validation regions (VRs) enriched in each of the backgrounds except the Higgs background are defined to verify the accuracy of each of the background predictions; MJVR for the multi-jet estimate, ZVR for Z+jets, WVR for W+jets, TVR for the top background and MBVR for the multi-boson background. An inclusive validation region, InclVR, is also defined to check the compatibility of the different estimation strategies together.

Details of each part of the background estimation strategy are discussed in turn in this section. For convenience, the definition of all the CRs and VRs used in the background estimate (excluding the control regions used for the multi-jet estimate) are summarised in Table 4.11. Similarly, the predicted and observed yields in each of these regions is listed in Tables 4.12- 4.15 and are summarised in Figure 4.14, where it can be seen that the predictions all agree within one standard deviation. These predictions come from a likelihood fit containing only the backgrounds and no signal (referred to as a background-only fit), in which the normalisation factors on Z+jets, W+jets and Top backgrounds are fit simultaneously from the Z+jets, W+jets and top CRs. As this fit contains three normalisation factors on backgrounds which have non-zero yields in each of the three CRs, no pulls or constraints on the systematic uncertainty nuisance parameters (NPs) are seen in this fit, as shown in Figure 4.12.

Correlations between the NPs in this fit are also checked. The largest correlations are shown in Figure 4.13. Correlations between the MC statistical NPs in the CRs and the normalisation factors of the dominant background in those CRs are expected and

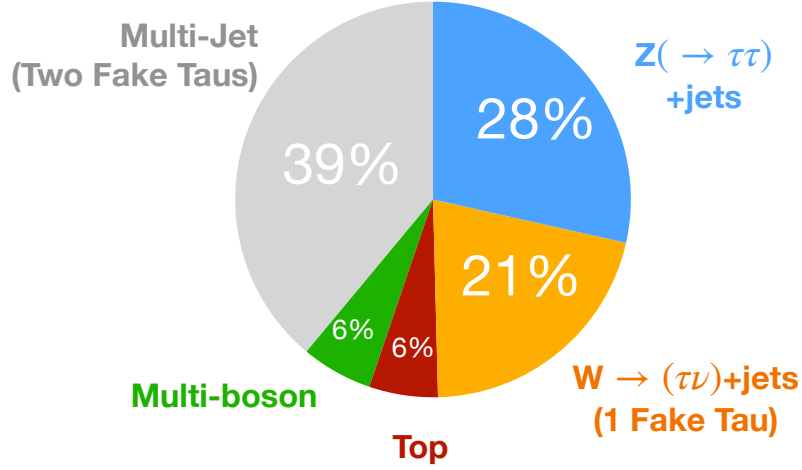


Figure 4.11: Pie-chart showing the relative sizes of the main SM backgrounds for the direct stau analysis. The Higgs background is not shown as it contributes less than 1% of the total.

are not of concern. The large correlation between the α -TES-DETECTOR NP, which describes an uncertainty in the tau-lepton energy scale, and the Z+jets CR is explained by this uncertainty being particularly large ($\sim 30\%$) on the Z+jets background in this region. The remaining NP correlations are not considered large enough to be of concern.

In the following subsections, post-fit kinematic distributions, resulting from the background-only fit to the CRs are shown in each of the CRs and VRs. These distributions include the full systematic uncertainties that are discussed in the next section. For brevity, only a selection of variables is shown in each region but all important variables, including all the variables that used as input features for the BDTs, were checked in every region prior to unblinding the SRs as part of the approval of this analysis by the ATLAS collaboration.

Chapter 4: Search for Direct Stau Production

Process	W+jets		Top	
Region	WCR	WVR	TCR	TVR
Charge combination	OS			
N signal τ	$= 1$	$= 1$	$= 1$	$= 2$
N “tight” τ	$= 0$	-	-	-
N signal e/μ	$= 1 \mu$	$= 1 \mu$	$= 1 \mu$	$= 0$
N b -jets	$= 0$	≥ 1	≥ 1	≥ 1
Trigger	single μ	single μ	single μ	asymm. di- τ
$p_T(\tau_1)$ [GeV]	-	-	-	> 95
$p_T(\tau_2)$ [GeV]	-	-	-	> 65
$\max[p_T(\tau), p_T(\mu)]$ [GeV]	> 95	> 95	> 95	-
$\min[p_T(\tau), p_T(\mu)]$ [GeV]	> 60	> 60	> 60	-
E_T^{miss} [GeV]	> 20	> 20	> 20	> 20
$m_T(\mu, E_T^{\text{miss}})$ [GeV]	$\in (50, 150]$	$\in (50, 150]$	$\in (50, 150]$	-
$m_{T2}(\tau, \mu)$ [GeV]	> 30	> 30	> 30	> 30
$m_{\tau\mu}$ [GeV]	> 120	> 120	> 120	-
$m_{\tau\tau}$ [GeV]	-	-	-	> 120
BDT scores	All < 0.5	Any > 0.5	-	-

Process	Z+jets		Multi-boson	Multi-jet	Inclusive
Region	ZCR	ZVR	MBVR	MJVR	InclVR
Charge combination	OS				
N signal τ	$= 2$	$= 2$	$= 2$	$= 2$	$= 2$
N e/μ	$= 0$	$= 0$	$= 0$	$= 0$	$= 0$
N b -jets	$= 0$	$= 0$	$= 0$	$= 0$	$= 0$
Trigger	asymm. di- τ	asymm. di- τ	asymm. di- τ	asymm. di- τ	asymm. di- τ
E_T^{miss} [GeV]	-	> 20	> 20	< 50	> 20
$m_{T2}(\tau_1, \tau_2)$ [GeV]	> 30	> 30	> 30	> 30	> 30
$m(\tau\tau)$ [GeV]	< 120	< 120	< 120	> 120	> 120
$\Delta R(\tau_1, \tau_2)$	< 4	< 4	< 4	> 3	< 4
BDT 1 score	≤ 0.10	> 0.10	-	-	≤ 0.60
BDT 2 score	-	-	-	-	≤ 0.70
BDT 3 score	-	-	-	-	≤ 0.70
BDT 4 score	≤ 0.60	≤ 0.60	> 0.61	-	≤ 0.60

Table 4.11: The definition of the control and validation regions for the direct stau channel. For the triggers, ‘asymm. di- τ ’ denotes the asymmetric di-tau trigger discussed in Section 4.5.1 including the offline p_T requirement to ensure events are on the efficiency plateau.

Chapter 4: Search for Direct Stau Production



Figure 4.12: Nuisance parameter pulls for the direct stau background-only fit to the CRs. NPs associated with normalisation factors and statistical uncertainties are shown in blue and have a γ prefix while NPs associated with systematic uncertainties are shown in black and have an α prefix.

Chapter 4: Search for Direct Stau Production

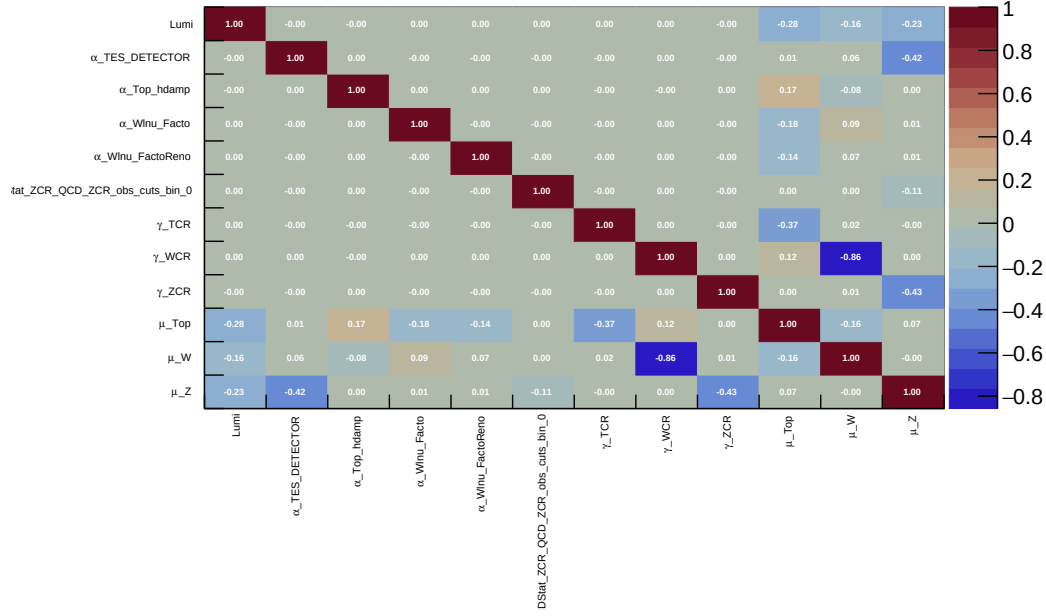


Figure 4.13: Reduced correlation matrix for the direct stau background-only fit to the CRs showing the fit parameters with the largest correlations. NPs associated with normalisation factors and statistical uncertainties have a γ prefix while NPs associated with systematic uncertainties have an α prefix.

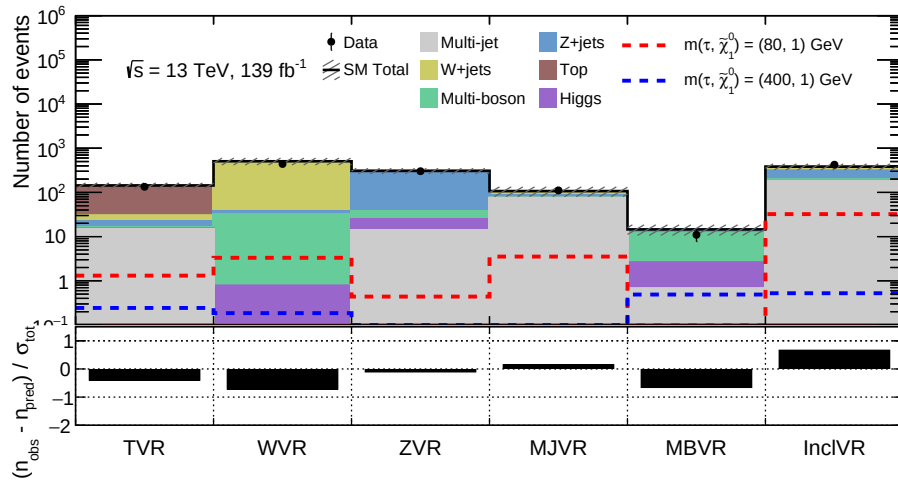


Figure 4.14: Pull plot for the direct stau VRs for the background only fit to the CRs, demonstrating the good agreement between the observed data and SM predictions.

Chapter 4: Search for Direct Stau Production

Table 4.12: Yields table (at 139 fb^{-1}) for ZCR, ZVR with full systematic uncertainties.

SM process	ZCR	ZVR
Multi-jet	16 ± 4	13.9 ± 3.5
W+jets	1.3 ± 1.0	3.6 ± 1.2
Z+jets	321 ± 20	260 ± 40
Top	$0.1^{+0.4}_{-0.1}$	1.3 ± 0.6
Multi-boson	6.5 ± 3.3	13 ± 4
Higgs	12 ± 4	10.6 ± 3.0
SM total	357 ± 19	310 ± 40
Observed	357	302
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 1) \text{ GeV}$	$0.06^{+0.12}_{-0.06}$	0.44 ± 0.17
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (240, 80) \text{ GeV}$	—	0.208 ± 0.032
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (200, 100) \text{ GeV}$	—	0.22 ± 0.11
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (400, 1) \text{ GeV}$	—	0.043 ± 0.007

Table 4.13: Yields table (at 139 fb^{-1}) for WCR, WVR with full systematic uncertainties.

SM process	WCR	WVR
Multi-jet	—	—
W+jets	509 ± 29	420 ± 90
Z+jets	15 ± 8	5.4 ± 3.5
Top	50 ± 10	46 ± 10
Multi-boson	35 ± 7	31 ± 10
Higgs	0.031 ± 0.013	0.8 ± 0.4
SM total	610 ± 25	510 ± 90
Observed	610	440
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 1) \text{ GeV}$	1.01 ± 0.21	3.3 ± 0.5
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (240, 80) \text{ GeV}$	0.09 ± 0.07	1.10 ± 0.17
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (200, 100) \text{ GeV}$	$0.10^{+0.18}_{-0.10}$	0.81 ± 0.21
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (400, 1) \text{ GeV}$	0.018 ± 0.006	0.186 ± 0.020

Chapter 4: Search for Direct Stau Production

Table 4.14: Yields table (at 139 fb^{-1}) for TCR, TVR with full systematic uncertainties.

SM process	TCR	TVR
Multi-jet	—	14.7 ± 3.4
W+jets	34 ± 8	8_{-8}^{+8}
Z+jets	$1.6_{-1.6}^{+2.4}$	5.7 ± 2.3
Top	466 ± 24	114 ± 19
Multi-boson	3.0 ± 2.0	$1.0_{-1.0}^{+1.1}$
Higgs	0.64 ± 0.09	0.6 ± 0.5
SM total	505 ± 22	143 ± 21
Observed	505	134
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 1) \text{ GeV}$	—	1.3 ± 0.5
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (400, 1) \text{ GeV}$	0.0021 ± 0.0010	0.243 ± 0.030

Table 4.15: Yields table (at 139 fb^{-1}) for MJVR, MBVR, InclVR with full systematic uncertainties.

SM process	MJVR	MBVR	InclVR
Multi-jet	74 ± 13	0.7 ± 0.5	177 ± 30
W+jets	20 ± 13	$0.00_{-0.00}^{+0.11}$	63 ± 26
Z+jets	9_{-9}^{+15}	2.5 ± 1.1	110 ± 50
Top	1.8 ± 1.6	0.28 ± 0.14	14 ± 5
Multi-boson	2.6 ± 1.6	9.3 ± 3.4	14 ± 5
Higgs	—	1.89 ± 0.25	$2.3_{-2.3}^{+2.7}$
SM total	107 ± 27	15 ± 4	380 ± 60
Observed	111	11	424
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 1) \text{ GeV}$	3.5 ± 1.1	—	32 ± 6
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (240, 80) \text{ GeV}$	0.40 ± 0.19	0.31 ± 0.11	4.5 ± 0.7
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (200, 100) \text{ GeV}$	0.61 ± 0.29	—	5.8 ± 1.2
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (400, 1) \text{ GeV}$	0.063 ± 0.021	0.49 ± 0.06	0.52 ± 0.09

4.6.1 Multi-jet Estimation

A semi-data-driven ABCD method is used to estimate the multi-jet contribution. In an ABCD method, four orthogonal regions (labelled here as CR-A, CR-B, CR-C and pre SR-D) are defined, one of which (pre SR-D) contains within it the signal region(s). The regions are separated by two variables, which should be uncorrelated with each other. Estimates for the non multi-jet background are computed for all of these regions and the data is unblinded in regions CR-A, CR-B and CR-C (which have a much lower S/B than pre SR-D). The multi-jet background is taken to be the difference between the data and the non-multi-jet background prediction in each of these regions. A transfer factor, here denoted T , is computed as the ratio of the multi-jet background between regions CR-C and CR-B. The multi-jet prediction in region pre-SR-D is then computed as the multi-jet prediction in region CR-A multiplied by T . As the variables separating the regions are uncorrelated, the real value of T is not assumed to change between regions CR-C/CR-B and regions pre-SR-D/CR-A.

In this analysis, the ABCD regions are separated by requirements on the tau-lepton-ID, the relative sign of the two tau-leptons and $m_{T2}(\tau_1, \tau_2)$. A schematic of the ABCD method for this analysis is shown in Figure 4.15 and the full region definitions are given in Table 4.16. In addition to the ABCD regions, two validation regions, VR-E and VR-F, are also defined and are used to determine a systematic uncertainty, discussed below. As in the signal region (SR) selections, the asymmetric di-tau trigger is used but unlike the SR selections the p_T threshold requirements to ensure events are on the trigger efficiency plateau are not required to improve the available statistics. How the potential bias this could introduce is checked is discussed in this section.

CR-A and pre-SR-D have the same definition except that in pre-SR-D the two leading tau-leptons are required to be of opposite sign and to pass the medium tau-lepton RNN-ID requirement ('medium' tau-leptons) while in CR-A events are required to have at least two taus passing the loose tau-lepton RNN-ID requirement ('loose' tau-leptons) and in cases where the leading two tau-leptons do pass the medium RNN-ID requirement they must be of the same sign. The same requirement on the tau-lepton-ID and the charge of the two tau-leptons as in CR-A (SR-D) is applied in CR-B (CR-C). The correlation of $m_{T2}(\tau_1, \tau_2)$ and the ID and sign requirements is checked in Figure 4.16. From this it can be seen the shape of $m_{T2}(\tau_1, \tau_2)$ is not affected by the tau-lepton-ID and sign requirements and thus is uncorrelated with them.

The non-multi-jet background in each of these regions is estimated using MC. As normalisation factors are used in the final fit for the Z+jets, W+jets and top backgrounds, it is important these are taken into account in the multi-jet estimate. To do this, the normalisation factors on each of these background and the multi-jet transfer factor T are fit simultaneously using the constraints from each of Z+jets, W+jets and top control regions plus the (data-MC) ratio between CR-B and CR-C (as this gives four equations with four unknowns). Note that the W+jets and top CRs are expected to have a negligible multi-jet contribution as they require one muon and one tau-lepton (and the rate at which the multi-jet background can fake a muon is low compared to the W+jets and top cross-sections) hence in these regions the multi-jet prediction is set to zero. For technical convenience, this simultaneous fit is done separately from the main background-only fit but the normalisation factors obtained from both fits are the same. In further fits discussed in this chapter, the multi-jet background does not have a floating normalisation factor and the multi-jet CRs are not included.

Data and MC yields in each of the ABCD regions and in VR-E and VR-F are tabulated in Table 4.17. These yields include the fit normalisation factors on the Z+jets, W+jets and top backgrounds. From these yields it can be seen the overall SM prediction in pre-SR-D and VR-F agrees with the data within the statistical uncertainties. Yields of example signal models are included to show the total signal contamination is low in all the control and validation regions.

CR-A	VR-E	CR-B	pre-SR-D	VR-F	CR-C
asymmetric di-tau trigger without p_T threshold requirements					
at least 2 loose tau-leptons					
light lepton veto					
2 loose (not medium) taus OR 2 same-sign medium tau-leptons			at least two opposite-sign medium tau-leptons		
$m_{T2} > 30 \text{ GeV}$	$20 < m_{T2} < 30 \text{ GeV}$	$10 < m_{T2} < 20 \text{ GeV}$	$m_{T2} > 30 \text{ GeV}$	$20 < m_{T2} < 30 \text{ GeV}$	$10 < m_{T2} < 20 \text{ GeV}$

Table 4.16: Region definitions for the Multi-jet background estimation (m_{T2} is equivalent to $m_{T2}(\tau_1, \tau_2)$).

The transfer factor is computed to be 0.186 with an uncertainty of 0.021 resulting from the finite data and MC statistics in regions CR-B and CR-C. To evaluate a systematic uncertainty on the transfer factor, the fit described above is repeated using VR-E and VR-F in place of CR-B and CR-C respectively. The difference between the transfer factor computed using VR-E and VR-F to the one using CR-B and CR-C is taken as a systematic uncertainty on the multi-jet predictions. Using VR-E and VR-F, the transfer factor is found to be 0.164 ± 0.022 (stat), meaning the size of the associated

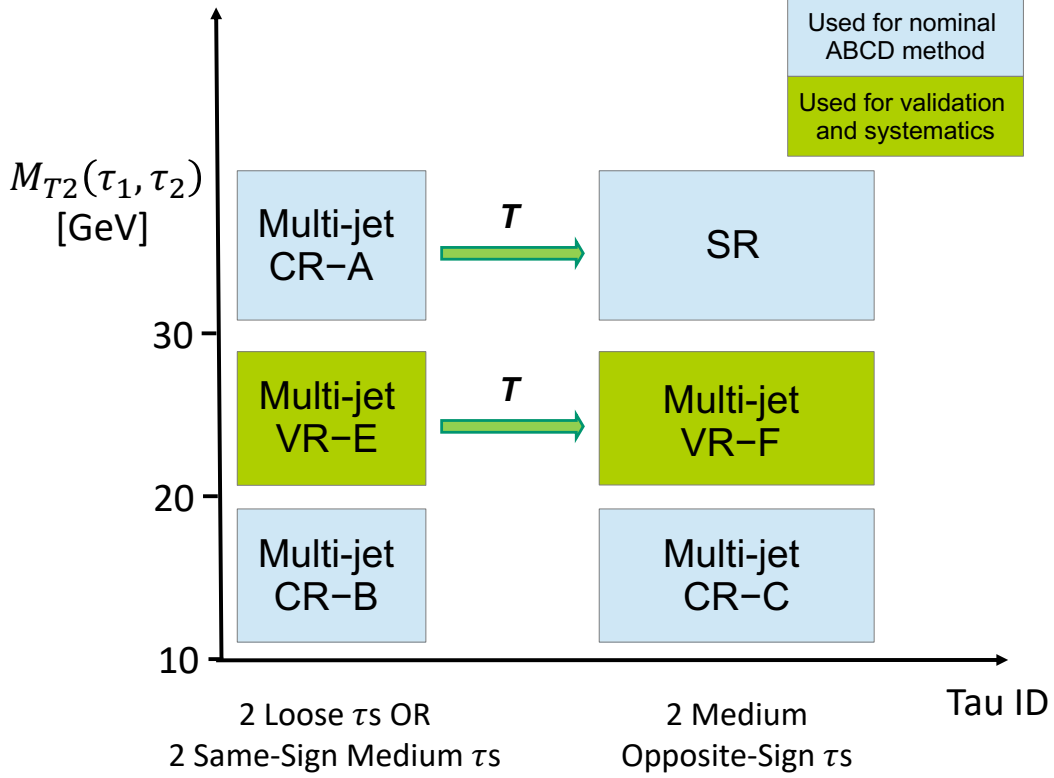


Figure 4.15: Illustration of the ABCD method for the multi-jet background determination in the signal regions. The control regions CR-A, CR-B, CR-C, and pre-SR-D are used for the ABCD method described in the text. Validation regions VR-E and VR-F, are used to validate the ABCD method and to estimate the systematic uncertainties. Figure adapted from [1].

systematic error is 12% and the final value of T is $0.186 \pm 0.022(\text{syst}) \pm 0.021(\text{stat})$. This uncertainty is used as a systematic uncertainty on the multi-jet predictions, as is the statistical uncertainty on T originating from the finite data and MC statistics in regions CR-B and CR-C (of size 11%). Note that the statistical uncertainty on T is separate from the statistical uncertainty on the final multi-jet estimates that originate from the finite data and MC statistics in CR-A, which are also included in the final fits.

A potential bias in the multi-jet estimate could be introduced due to the ABCD regions not requiring the asymmetric ditau trigger plateau selections, which are required in the other analysis regions which require two tau-leptons. To verify this is not the case, T is re-computed using the same selections for the ABCD regions but requiring the plateau cuts. With the plateau cuts applied, T is computed to be $0.22 \pm 0.07(\text{syst}) \pm 0.05(\text{stat})$, with the errors computed using the same methods

Chapter 4: Search for Direct Stau Production

	CR-A	VR-E	CR-B	pre-SR D	VR-F	CR-C
Z+jets	505 ± 31	332 ± 52	457 ± 67	1400 ± 43	939 ± 63	1651 ± 137
W+jets	390 ± 37	240 ± 53	315 ± 49	257 ± 33	188 ± 48	240 ± 57
Top	187 ± 5	79 ± 3	74 ± 3	309 ± 6	133 ± 4	148 ± 4
Multi-boson	48 ± 2	24 ± 2	25 ± 2	113 ± 3	47 ± 2	57 ± 2
Higgs	22 ± 1	8 ± 1	12 ± 1	59 ± 2	26 ± 1	37 ± 2
Total MC BG	1152 ± 48	683 ± 74	883 ± 83	2138 ± 55	1333 ± 79	2133 ± 148
Multi-jet prediction	-	-	-	821 ± 138	776 ± 131	-
Total MC + Multi-jet	-	-	-	2959 ± 149	2109 ± 153	-
Data	5564 ± 75	4855 ± 70	9402 ± 97	3127 ± 56	2018 ± 45	3719 ± 61
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (120, 1)$ GeV	41.51 ± 1.31	11.77 ± 0.70	12.05 ± 0.72	163.16 ± 2.59	45.42 ± 1.38	46.36 ± 1.39
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (200, 1)$ GeV	22.03 ± 0.50	2.37 ± 0.16	2.10 ± 0.15	88.65 ± 0.99	11.07 ± 0.36	10.43 ± 0.34
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (360, 1)$ GeV	4.36 ± 0.25	0.31 ± 0.07	0.17 ± 0.05	19.12 ± 0.51	1.09 ± 0.12	0.96 ± 0.12

Table 4.17: Event yields in regions for used for the multi-jet estimation. No multi-jet prediction is shown or CR-A, CR-B, CR-C and VR-E as the difference in data-MC in these regions is used to construct the multi-jet estimate in the other regions. Errors are from MC statistics.

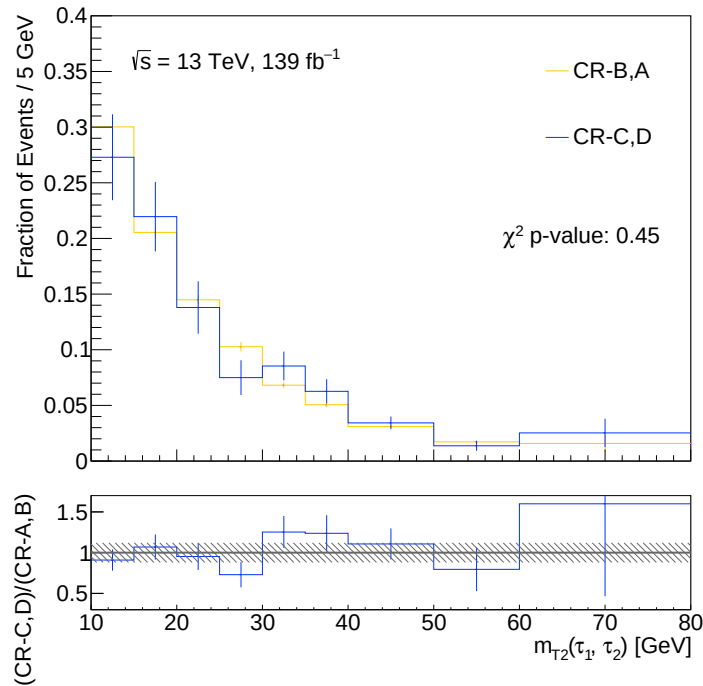


Figure 4.16: Shapes of (data - MC) for the CR-B/CR-A selection (without the $m_{T2}(\tau_1, \tau_2)$ cut) in orange and for the CR-C/pre-SR D selection in blue.

described above. This value of T agrees within the uncertainties with the nominal value of T , hence no additional uncertainty is associated with the fact that the trigger plateau cuts are not applied in the ABCD regions.

Example kinematic plots in VR-F and pre-SR D including the multi-jet prediction are shown in Figures 4.17 and 4.18 respectively. A flat 20% systematic is included in

these plots as an estimate of the total uncertainty (with the value of 20% being based upon the computed uncertainty in the inclusive validation region - see Table 4.15). Good agreement between the total background prediction and data is seen in both VR-F and pre-SR D.

For training the BDTs, it is desirable to have a multi-jet estimate without negatively weighted events (which are omitted from the BDT training). This is not possible with the ABCD method, due to subtraction of MC from data resulting in a high fraction of negatively weighted events, so an alternative multi-jet estimate is devised. In this alternative estimate, a weight factor is computed such that the data yield in CR-A multiplied by this weight factor is equal to the data-MC difference in CR-A. The alternative multi-jet estimate in pre SR-D is then given by the data in CR-A multiplied by this weight factor multiplied by T . Thus, by construction this alternative estimate has no negatively weighted events, as both the weight factor and T are positive, and the total yield prediction in pre SR-D is equivalent to that of the full multi-jet estimate. However, it should be noted that the shapes of the kinematic distributions for this alternative estimate are not as accurate as with the ABCD method, hence why this alternative estimate is only used for training the BDTs. A comparison of the shapes of key kinematic variables between the simplified and full multi-jet estimation methods for events passing the BDT preselection can be found in the appendix, Section A.2.4, where it can be seen the overall shape differences are small.

An additional validation region, MJVR, with higher purity in the multi-jet background is also checked. Unlike the regions used in the ABCD method, the plateau selections on the asymmetric di-tau trigger, detailed in Table 4.2, are required in this region. The definition of MJVR is given in Table 4.11. It is orthogonal to the signal regions due to the inverted BDT score requirements. The $E_T^{\text{miss}} < 50$ GeV and $\Delta R(\tau_1, \tau_2) > 3$ cuts are to improve the purity in the multi-jet background which is found to be 67%. This region is statistically independent of VR-F but is within pre SR-D. Predicted and observed event yields for MJVR region can be found in Table 4.15; good agreement between them is seen. A selection of post-fit kinematic distributions in MJVR are shown in Figure 4.19 in which good agreement between the SM predictions and the observed data is also seen.

Chapter 4: Search for Direct Stau Production

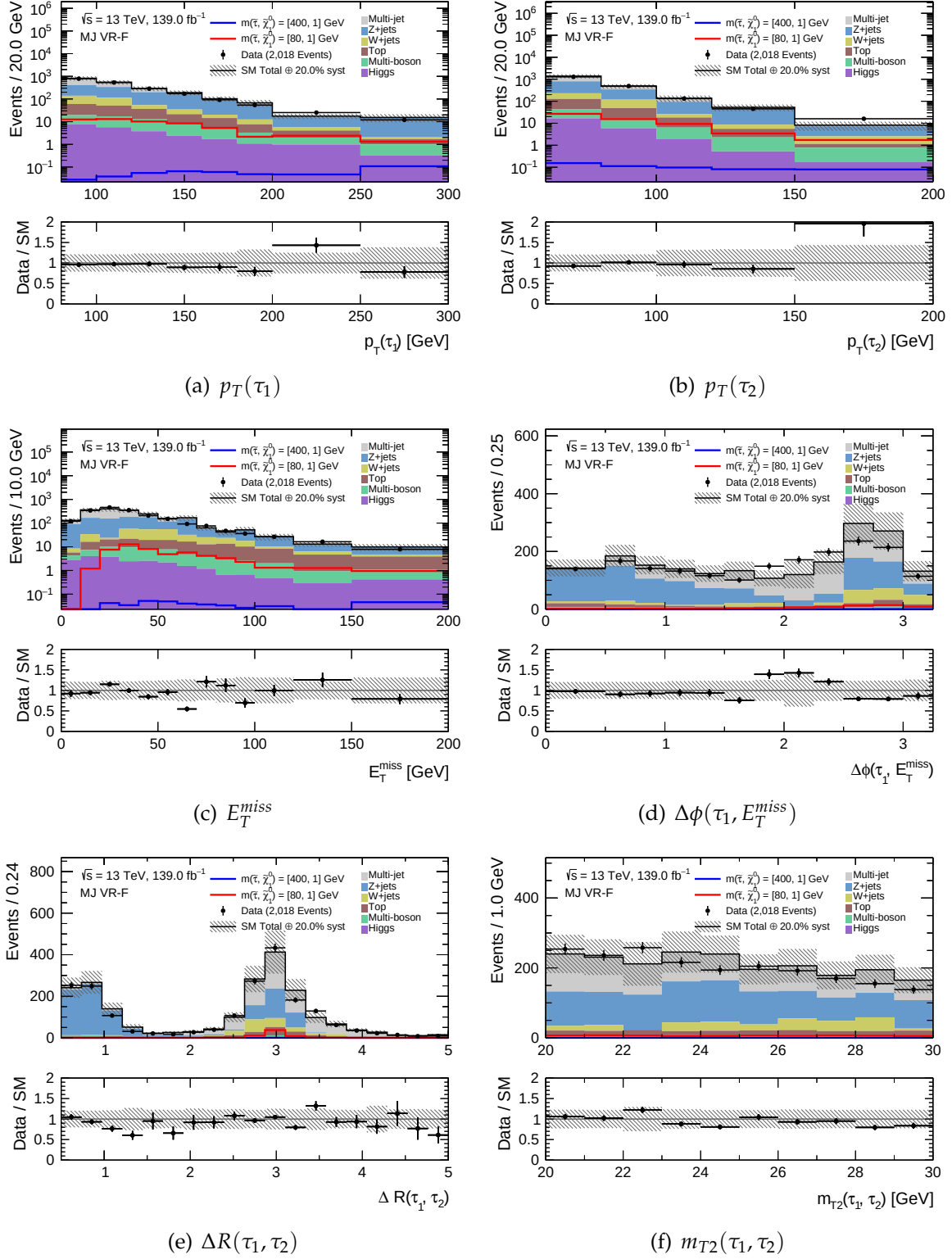


Figure 4.17: Kinematic distributions for Direct Stau in MJ VR-F. Normalisation factors are applied to the Z+jets, W+jets and Top backgrounds. A flat 20% systematic is applied on top of the statistical uncertainties.

Chapter 4: Search for Direct Stau Production

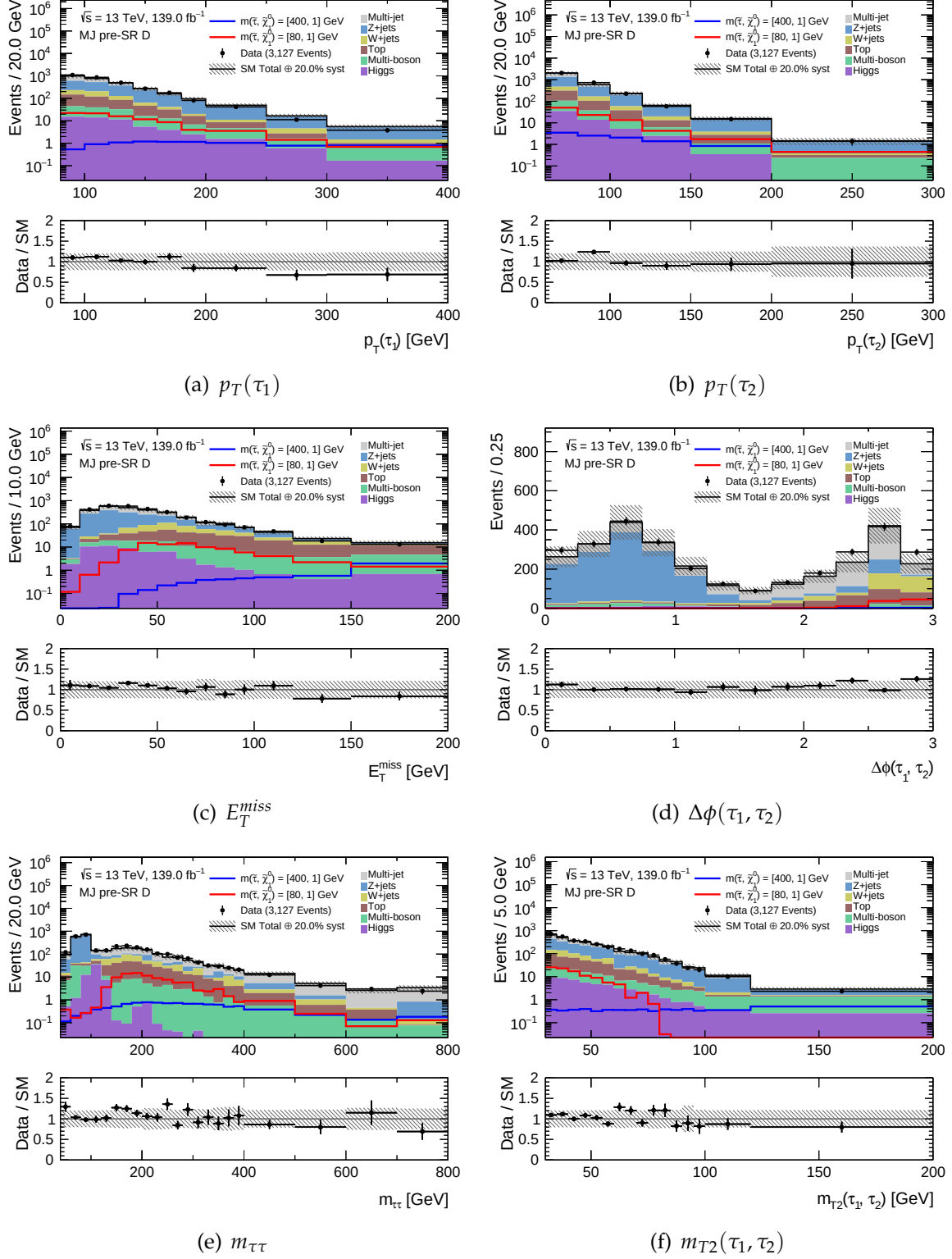


Figure 4.18: Kinematic distributions for Direct Stau in MJ pre-SR D. Normalisation factors are applied to the Z+jets, W+jets and Top backgrounds. A flat 20% systematic is applied on top of the statistical uncertainties.

Chapter 4: Search for Direct Stau Production

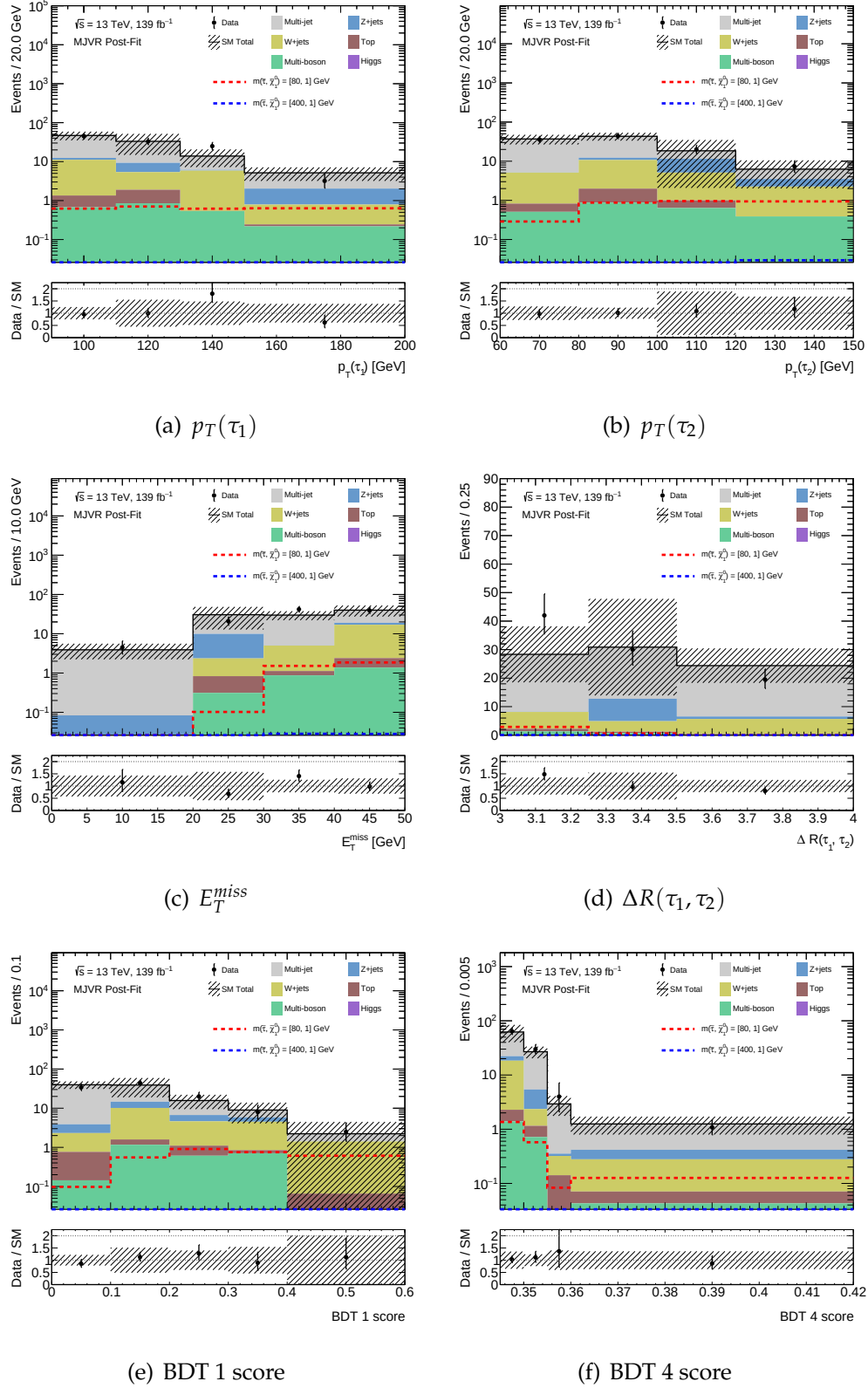


Figure 4.19: Post-fit kinematic distributions for MJVR. First and last bins include the under-flow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

4.6.2 Z+jets Estimation

The production of a Z-boson in association with jets poses an irreducible background to the final state with two hadronically decaying tau-leptons and E_T^{miss} , which can arise from the tau-lepton neutrinos. As previously mentioned, the shapes of kinematic variables for the Z+jets background are taken from MC while the normalisation factor is determined in a control region, ZCR with its use being verified in a validation region, ZVR. Both regions are defined by the selections detailed in Table 4.11. Compared to the signal regions, the cut on $m_{\tau\tau}$ is inverted which results in a greater than 90% purity in Z+jets in both regions. A cut on BDT 1 score is used to separate ZCR from ZVR, with ZVR being at higher values. The upper cut on BDT 4 score is used to make this region orthogonal with the multi-boson validation region (discussed in Section 4.6.5).

The normalisation factor on the Z+jets background, μ_Z , is determined from the background-only fit to the CRs to be 0.91 ± 0.07 . Event yields in both the ZCR and ZVR are shown in Table 4.12. In the ZVR, the total SM prediction is found to match well with the number of observed data events. A selection of post-fit kinematic distributions in the ZCR and ZVR can be found in Figures 4.20 and 4.21 respectively where good agreement between the predictions and observed data are seen.

Chapter 4: Search for Direct Stau Production

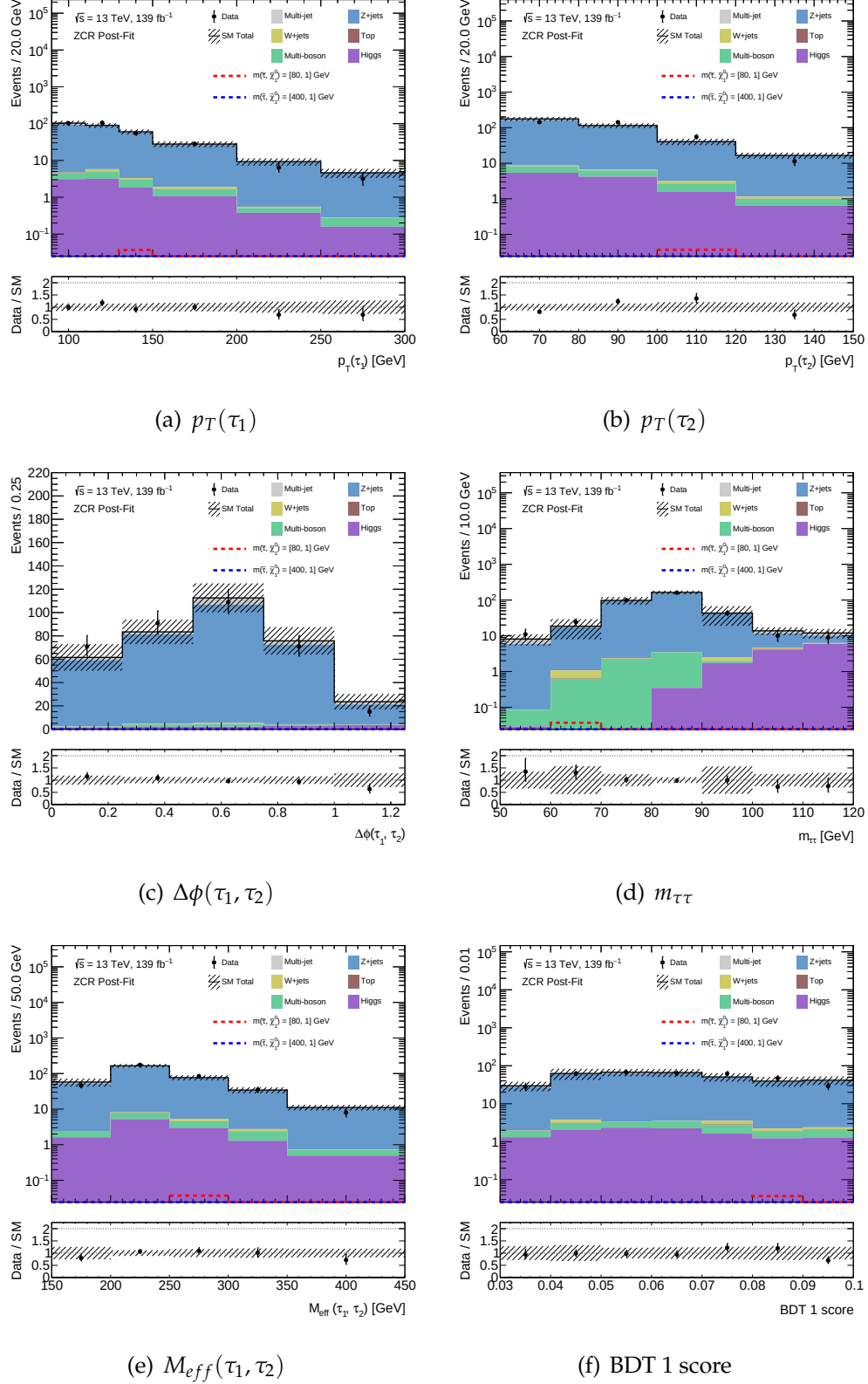


Figure 4.20: Post-fit kinematic distributions for ZCR. First and last bins include the under-flow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

Chapter 4: Search for Direct Stau Production

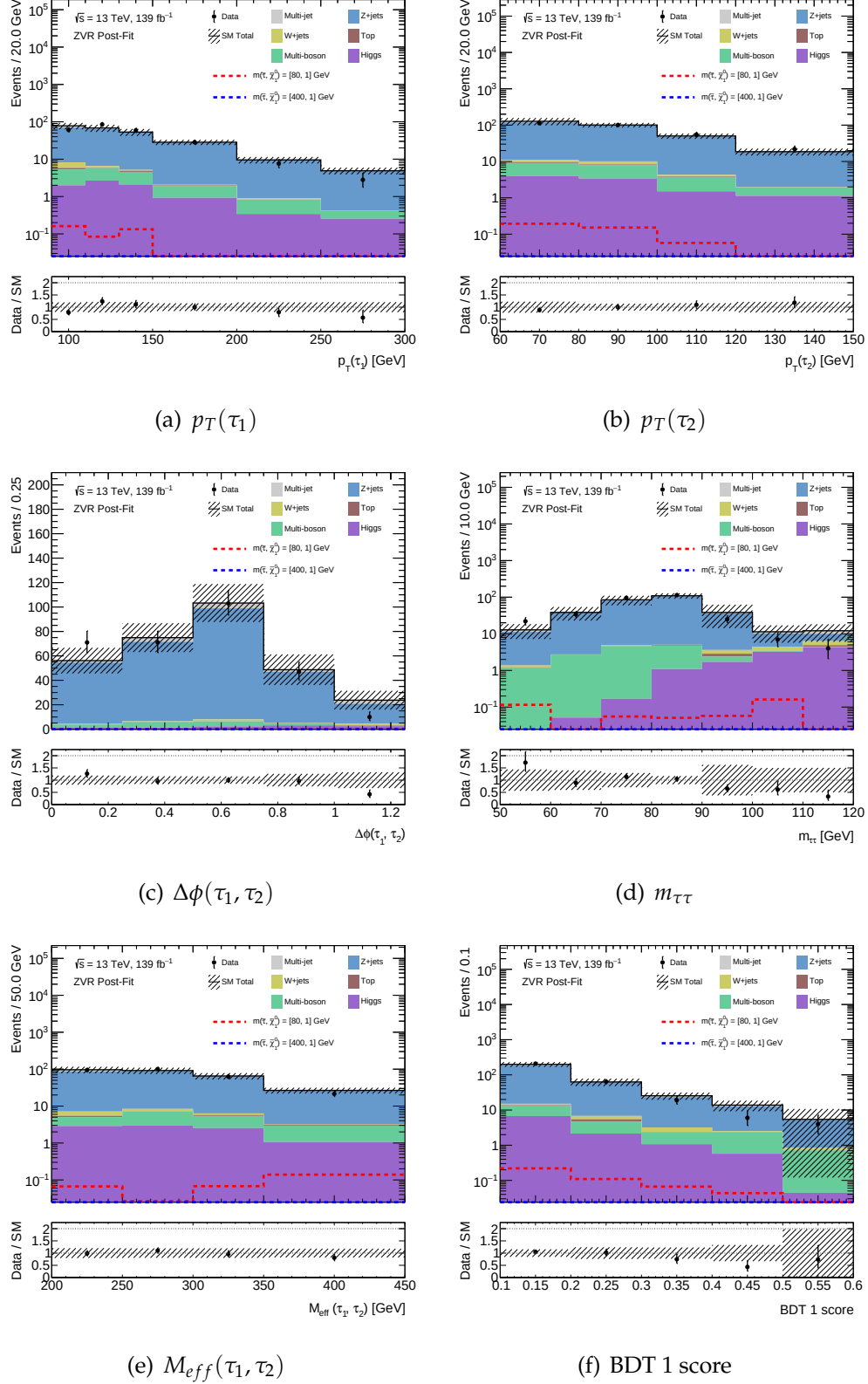


Figure 4.21: Post-fit kinematic distributions for ZVR. First and last bins include the under-flow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

4.6.3 W+jets Estimation

Events in which a W -boson is produced in association with jets give rise to a significant background, particularly in SR-BDT1 which targets the signals with low stau mass. This background consists of events where the W -boson decays to a hadronically decaying tau-lepton and one of the jets fakes the other hadronic tau-lepton. As for the Z +jets background, MC is used to estimate the kinematic shapes while the normalisation is determined in a control region, WCR, and the accuracy of the prediction is verified in a validation region, WVR.

Both WCR and WVR require events to have one hadronic tau-lepton and one isolated signal muon, with the muon being used as a proxy for the other hadronic tau-lepton required by the SR selections. The full selection requirements for the WCR and WVR are detailed in Table 4.11. These selections are designed to emulate those in the BDT training preselection (Table 4.3). The single-muon trigger is used in place of the asymmetric di-tau trigger for these regions, but selections on the maximum and minimum of the tau-lepton and muon p_T are required to emulate its effect. The requirement on the transverse mass of the muon with the E_T^{miss} , $m_T(\mu, E_T^{\text{miss}})$, is to improve the purity in W +jets while the veto on b -jets is to make this region orthogonal with the top CR, detailed in the next subsection. The tau-lepton is required to pass the “medium” ID requirement, but fail the “tight” ID requirement of the RNN tau-ID algorithm. This veto on tight tau-leptons is used to allow the selection requiring a tight tau-lepton to be used for a future search for direct stau production in the final state with one leptonically decaying tau-lepton and one hadronically decaying tau-lepton. By not including this phase space in the WCR, it allows the results of such a search to be statistically combined with this one.

BDT scores computed for events in WCR and WVR use input variables which use the muon four-vector in place of the second tau-lepton four-vector (for instance the invariant mass of tau and muon, $m_{\tau\mu}$, is used in place of $m_{\tau\tau}$) with the object with the highest p_T being used as the leading tau-lepton. The WCR and WVR are split via a requirement on the BDT scores, specifically for the WCR all four BDT scores are required to be less than 0.5 while for the WVR the inverse is required. This results in the WCR being at low BDT scores and WVR being at high BDT scores and hence kinematically more signal like. Both regions are more than 80% pure in W +jets. The fake background, originating from multi-jet events with a fake muon and a fake tau, is assumed to be negligible in these regions.

The normalisation factor on the W +jets background, μ_W , determined in the background only fit to the CRs, is found to be 0.93 ± 0.11 . Event yields in WCR and WVR are detailed in Table 4.13, from which it can be seen that the total predicted yield in the WVR agrees within one standard deviation with the observed data. A selection of post-fit kinematic distributions in the WCR and WVR can be found in Figures 4.22 and 4.23 respectively; good modelling is observed in both regions. Taken together, the WCR and WVR allow the full BDT score distribution to be inspected, to verify good data/MC agreement at both low and high scores, hence these distributions are included in both of these figures.

Chapter 4: Search for Direct Stau Production

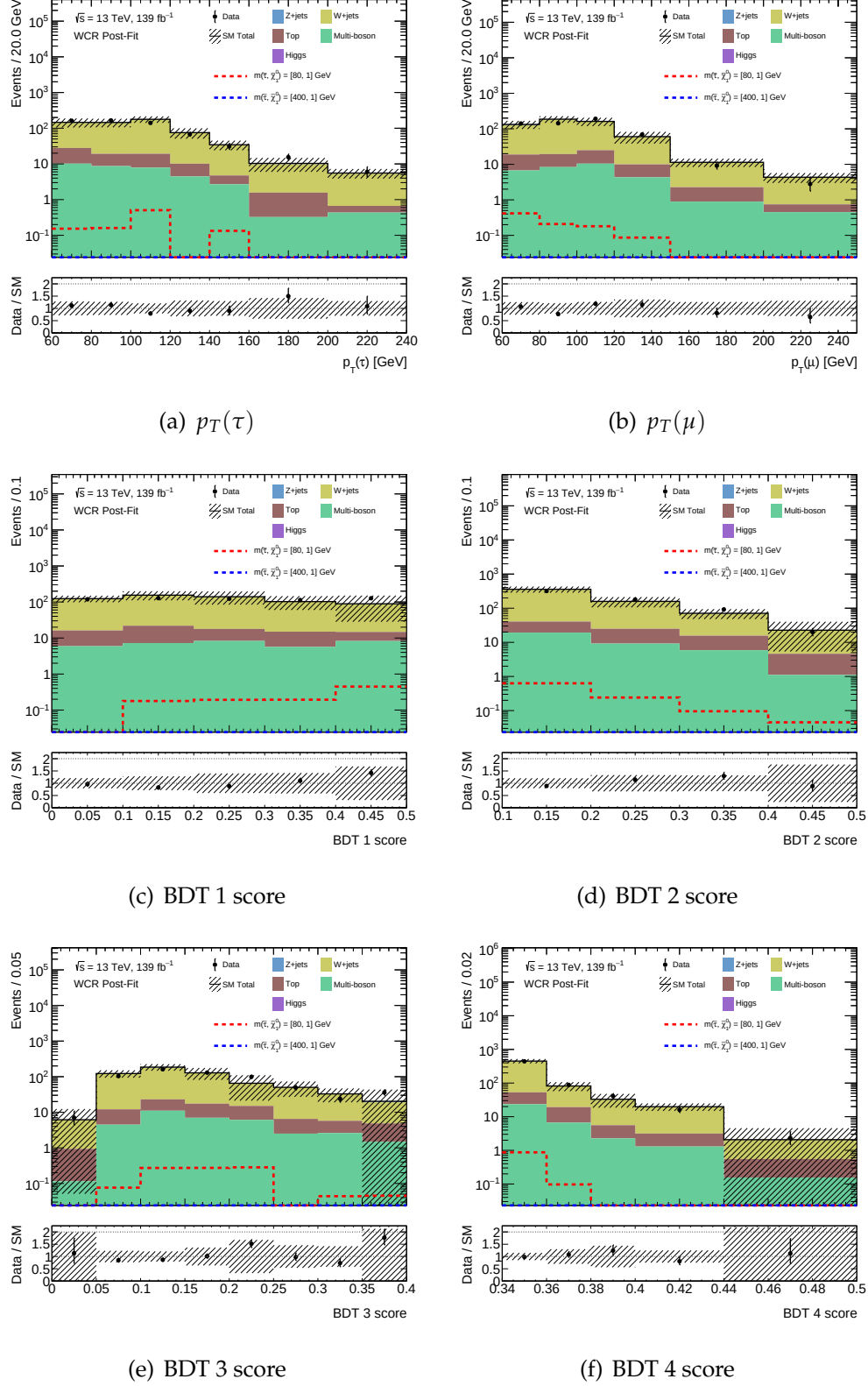


Figure 4.22: Post-fit kinematic distributions for WCR. First and last bins include the under-flow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

Chapter 4: Search for Direct Stau Production

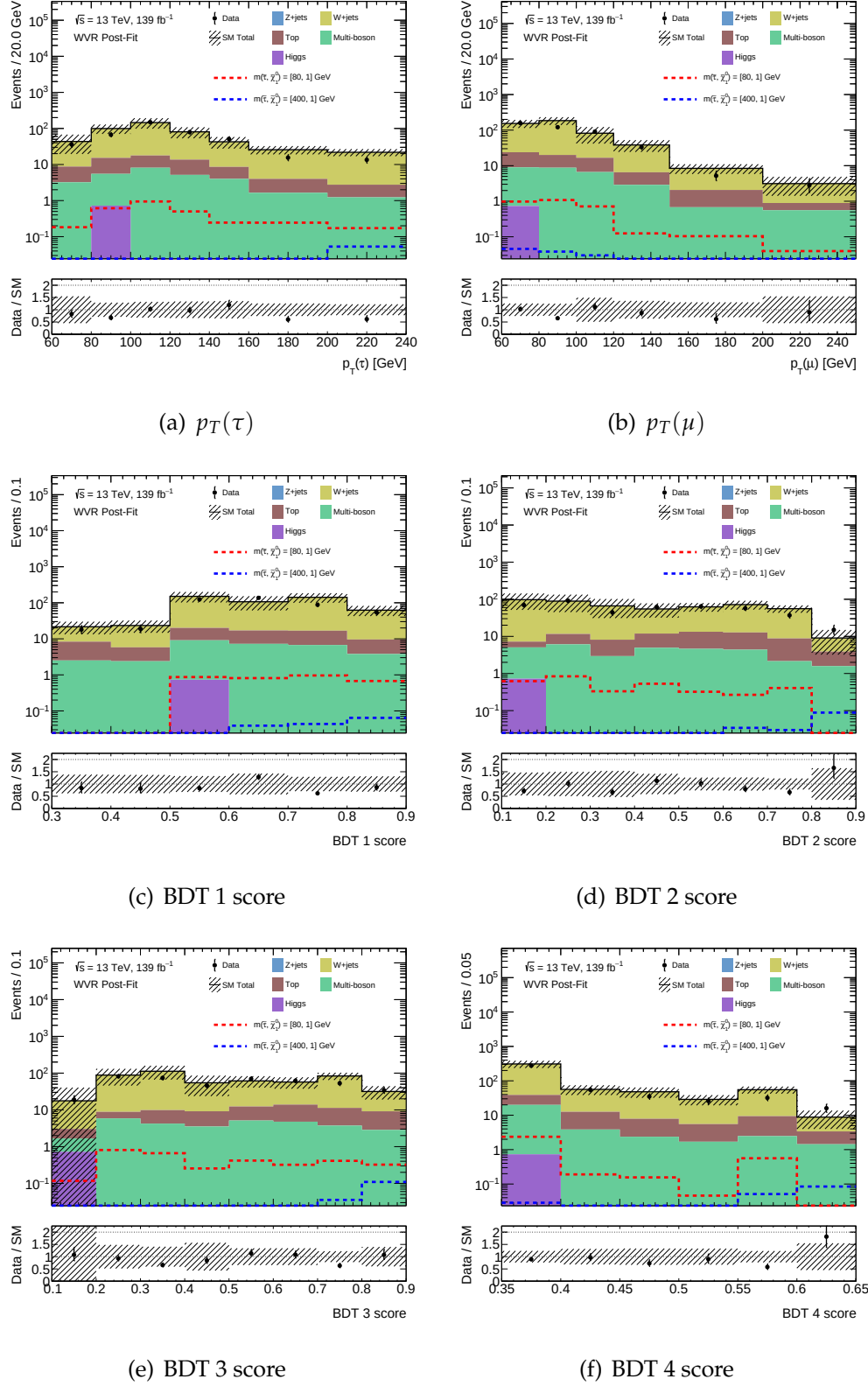


Figure 4.23: Post-fit kinematic distributions for WVR. First and last bins include the under-flow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

4.6.4 Top Estimation

Events containing one or more top quarks also contribute significantly to the total SM background. As already outlined in the analysis strategy overview, this background is also estimated using MC normalised in a control region, TCR, which, like the WCR, requires one hadronic tau-lepton and one signal muon, but additionally requires events to have at least one b -jet to give a high purity in the top quark containing backgrounds. As in the WCR, variables in the SRs computed using the second tau-lepton four-vector here use the muon four-vector and the BDT scores use these variables as inputs in place of their two-tau-lepton analogues. The top validation region (TVR) has the same selections as the BDT preselection except for replacing the b -jet veto for a requirement of at least one b -jet to enhance the purity in the top background.

The normalisation factor of the top background, μ_{Top} is determined from the background only fit to the CRs to be 0.84 ± 0.05 , which from the event yields in Table 4.14 can be seen to lead to good agreement in the TVR. As neither the TCR nor TVR are dependent on the BDT score cuts, the BDT score distributions can be plotted in their entirety to check for mis-modelling, which from Figures 4.24 and 4.25 does not appear to be present.

Chapter 4: Search for Direct Stau Production

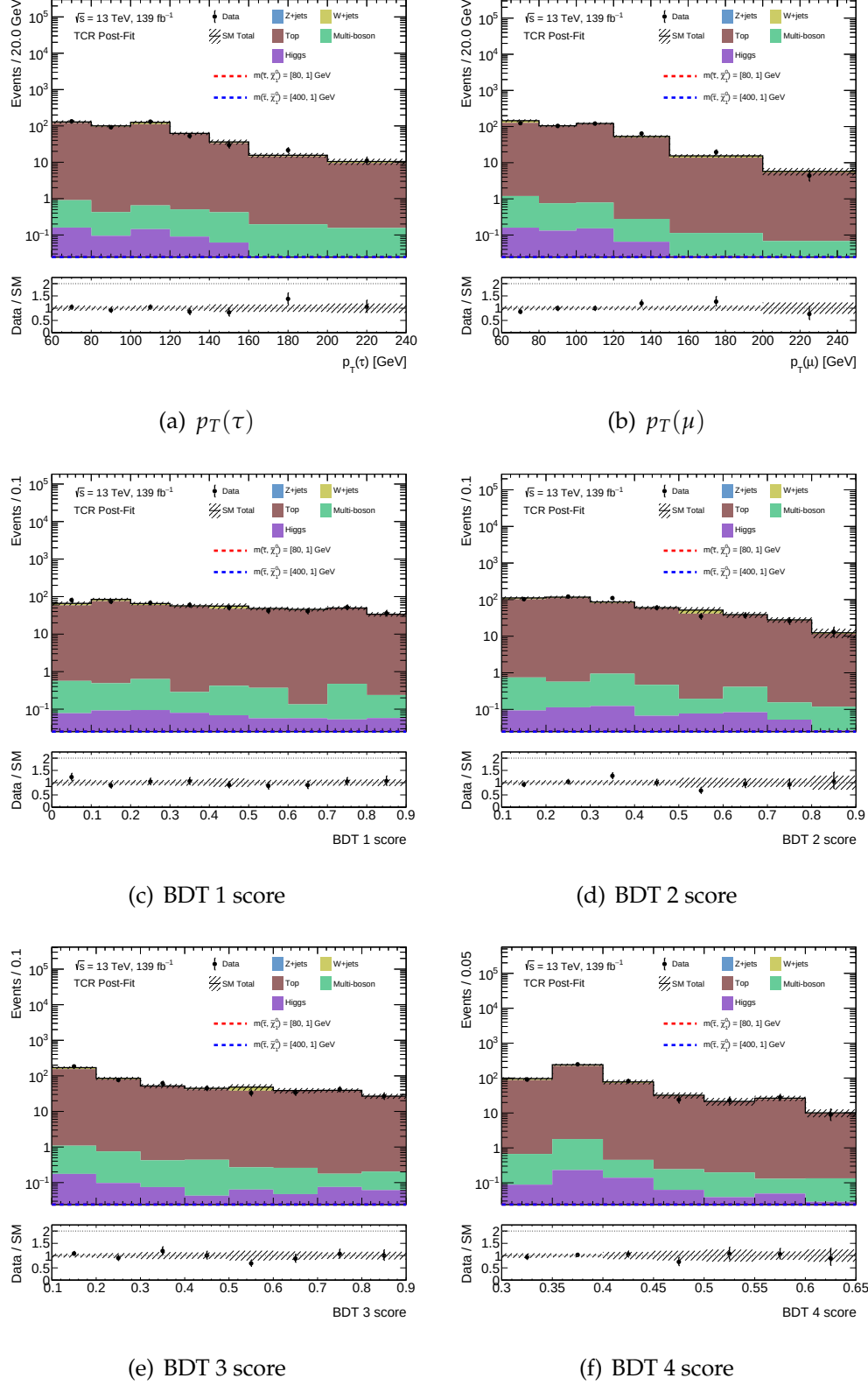


Figure 4.24: Post-fit kinematic distributions for TCR. First and last bins include the under-flow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

Chapter 4: Search for Direct Stau Production

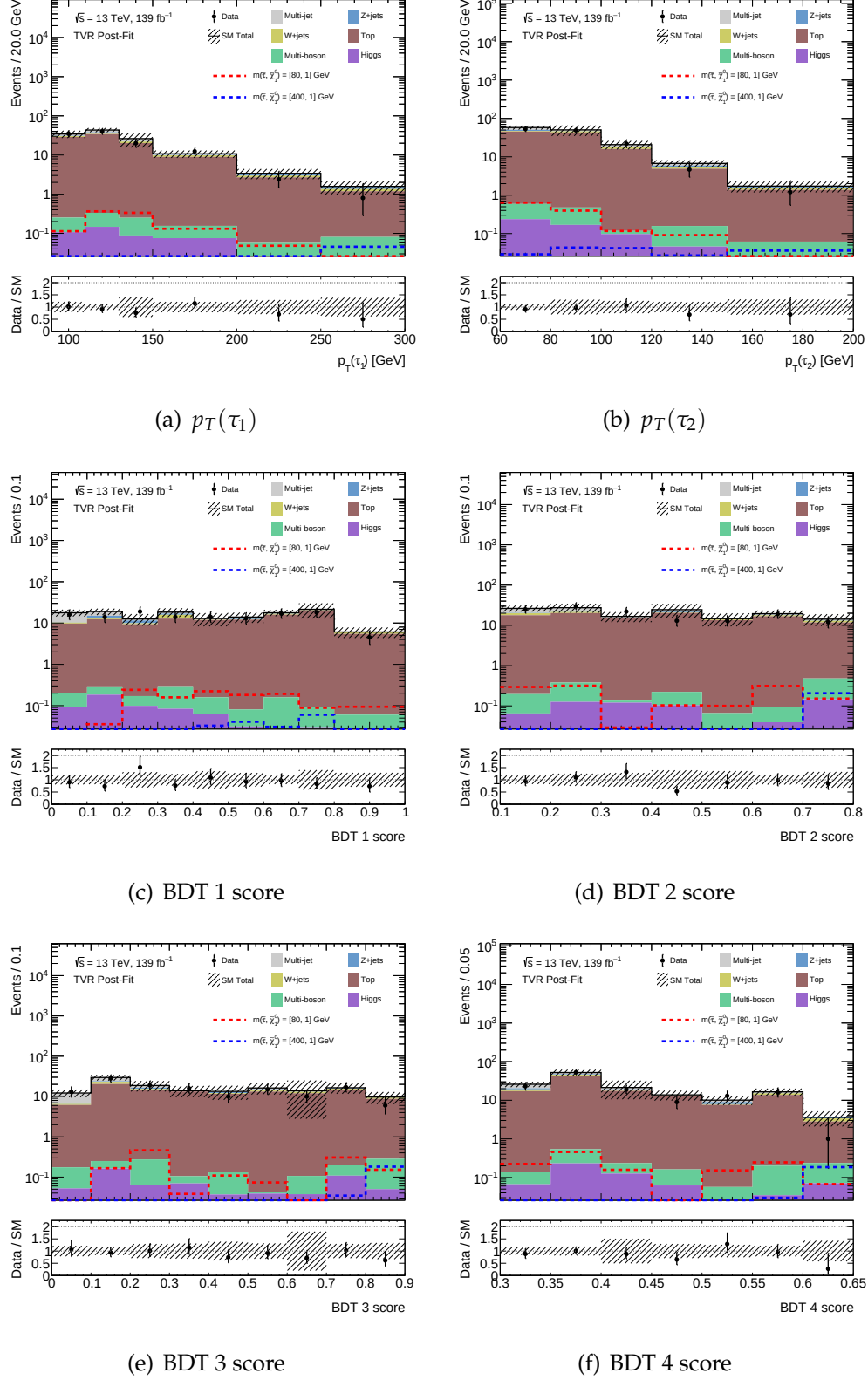


Figure 4.25: Post-fit kinematic distributions for TVR. First and last bins include the under-flow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

4.6.5 Multi-boson Estimation

Events with more than one vector boson are the dominant background in SR-BDT2 and SR-BDT4. Due to the relatively small yield of this background, and its similarities with the direct stau signal, it is difficult to isolate. Nevertheless, a dedicated validation region, MBVR, with these backgrounds as the dominant contribution is established by reversing the selection on $m_{\tau\tau}$ in the SRs and requiring a high BDT 4 score. Although there are limited statistics in these regions, as detailed in Table 4.15, reasonable data/MC agreement is seen and the region is 62% pure in multi-boson processes. Compared to the signal region, the multi-boson contribution in MBVR consists primarily of ZZ and WZ events while the SR is rich in WW events. Despite this, MBVR is still useful as a VR for this background as the same MC generator is used for all of these processes. A selection of post-fit kinematic distributions in this region can be found in Figure 4.26, where again reasonable agreement is seen, though the distributions have to be binned coarsely due to the limited statistics.

Chapter 4: Search for Direct Stau Production

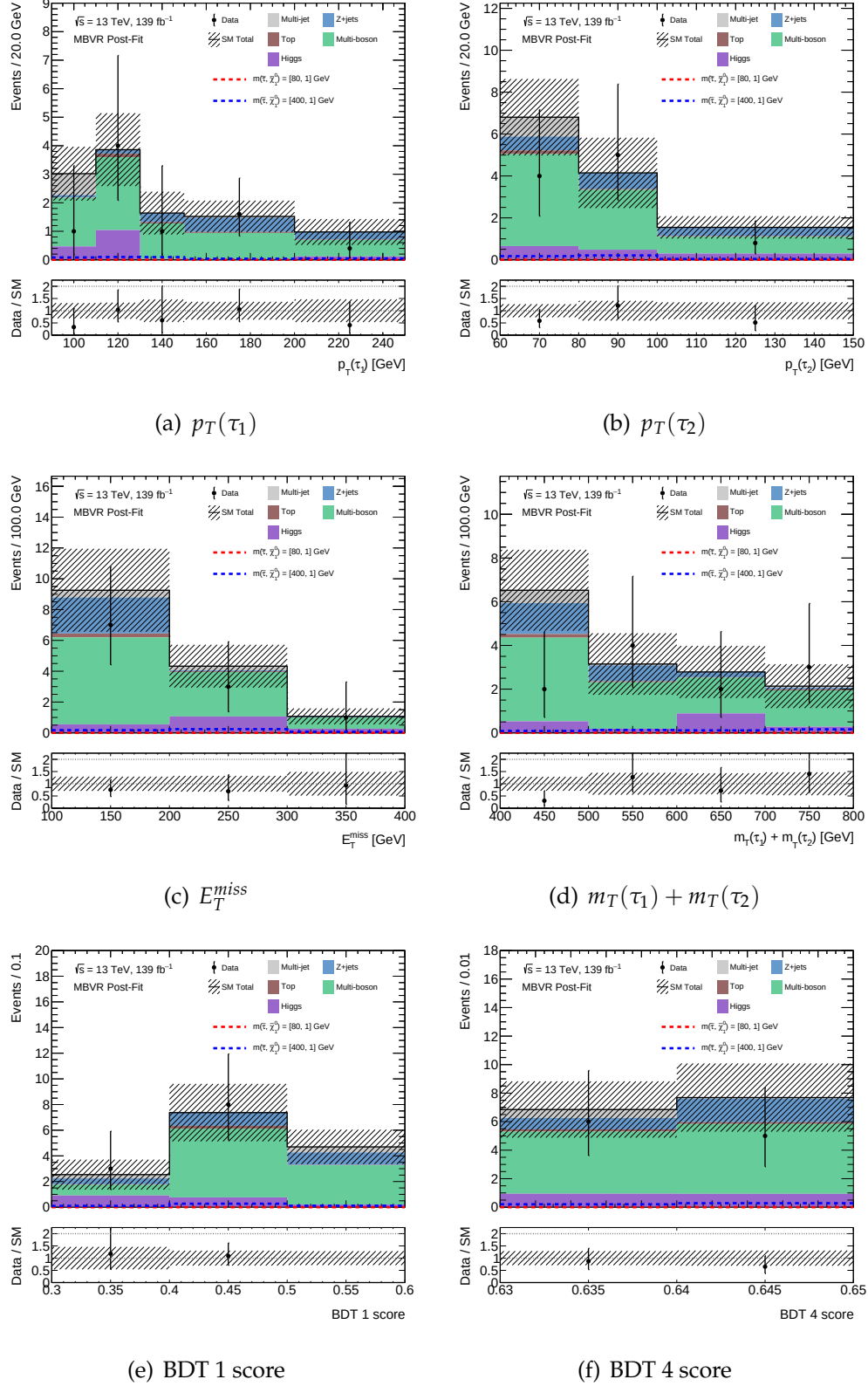


Figure 4.26: Post-fit kinematic distributions for MBVR. First and last bins include the under-flow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

4.6.6 Inclusive Validation Region

As previously discussed, in addition to the dedicated validation regions targeting specific background processes, an inclusive validation region, InclVR, is also defined to check the overall background estimation strategy and the modelling of the BDT score distributions. This region has the same selection requirements as the signal regions aside from the selections on the BDT scores. As detailed in Table 4.11, upper bounds each on each of the BDT scores are set, which keep this region orthogonal with the SRs and ensure the signal contamination is low enough that this region could be safely checked prior to unblinding the SRs. It should be noted that this region overlaps with MJVR but contains a significantly greater number of events. As well as checking the background estimates work with each other, it also allows for testing the modelling of Z+jets at $m_{\tau\tau} > 120$ GeV, the multi-jet background at $E_T^{\text{miss}} > 50$ GeV and of W+jets with two hadronic tau-leptons (albeit simultaneously).

From the overall yields in Table 4.15, it can be seen the total SM prediction agrees well with the observed yield in this region. Figure 4.27 shows some kinematic distributions in this region, including all the BDT scores, which look to be well modelled.

Chapter 4: Search for Direct Stau Production

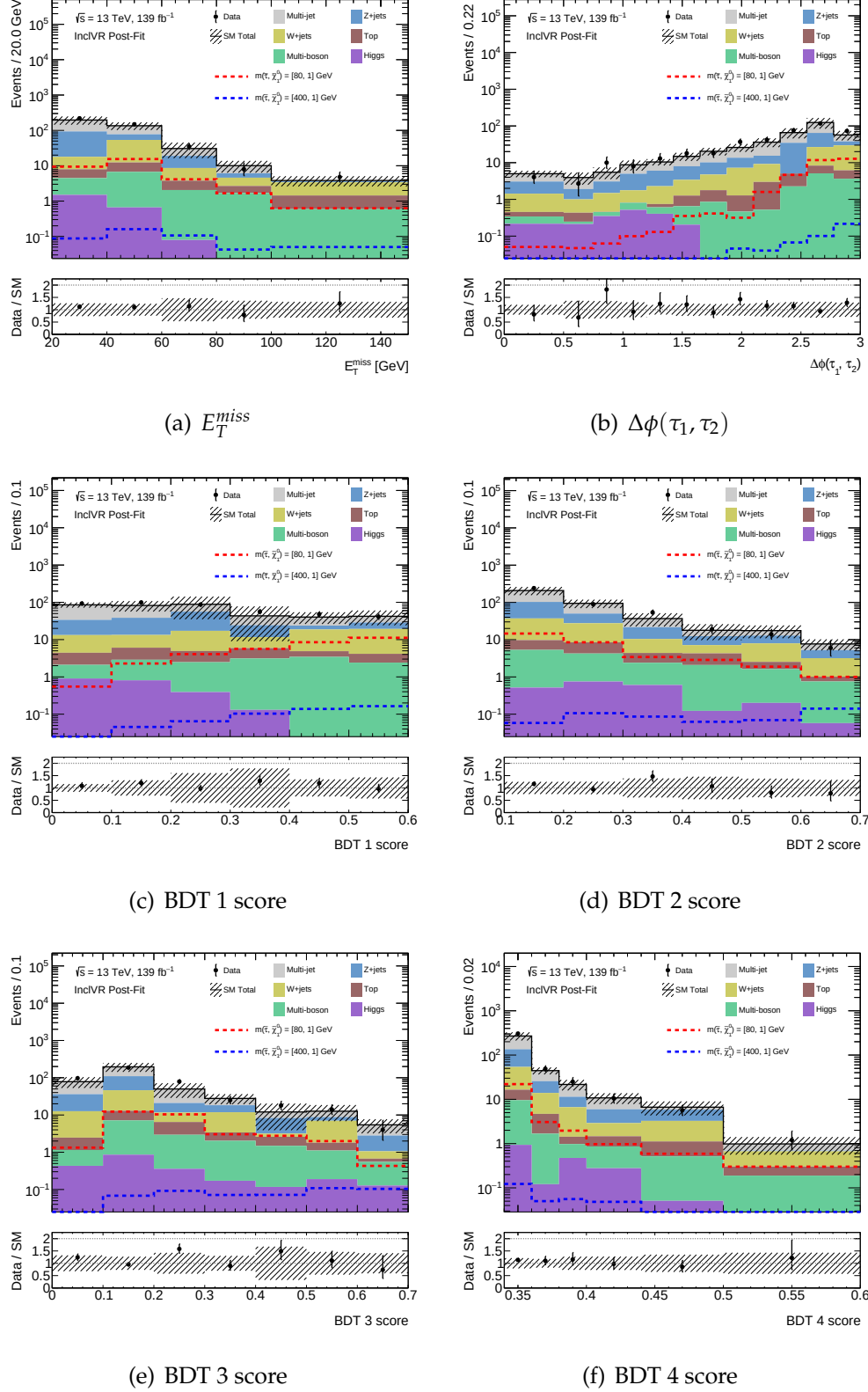


Figure 4.27: Post-fit kinematic distributions for InclVR. First and last bins include the under-flow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

4.7 Systematic Uncertainties

Several sources of systematic uncertainty are considered in the analysis. The uncertainties can be grouped into those which are experimental, resulting from uncertainties on the event/object reconstruction and on the total integrated luminosity, and those that are theoretical, which arise due to uncertainties on the modelling of the SM background processes and SUSY signals. Additionally, there are uncertainties due to the finite number of MC statistics used to model each of the physics processes and uncertainties specific to the data-driven estimate of the multi-jet background. Uncertainties are incorporated into the statistical fits using nuisance parameters (NPs) as described in Chapter 3, Section 3.1. The origin of the systematic uncertainties is discussed briefly below, followed by an assessment of their impact on the analysis.

4.7.1 Experimental Uncertainties

The estimation of the experimental uncertainties follows the recommendations from relevant working groups within the ATLAS collaboration. All experimental uncertainties are evaluated on the SM background processes estimated using MC and on the signal samples and are correlated across all the samples and across all the analysis regions. The uncertainties associated with the physics objects fall into two types, those which effect the weights of the individual MC events and those which effect the distributions of kinematic variables, both of which impact the efficiencies of the analysis selections.

Uncertainties on the reconstruction and identification of hadronically decaying tau-leptons are of obvious importance to the analysis and are handled by 15 NPs in total. In particular, five NPs describe uncertainties pertaining to the tau-lepton energy scale (TES) introduced in Chapter 2, Section 2.2.4. Uncertainties are also associated with the RNN ID algorithm distinguishing jets from tau-leptons and the BDT algorithm distinguishing electrons from 1-prong tau-leptons. The uncertainty of the asymmetric di-tau trigger efficiency is also evaluated. The E_T^{miss} is also of crucial importance to the analysis. Uncertainties on the missing transverse momentum are primarily evaluated by propagation of the uncertainties of the other objects with additional uncertainties associated with the scale and resolution of E_T^{miss} soft-term, which is handled by 3 NPs.

Uncertainties on jets are also important, primarily due to the impact they have on the E_T^{miss} . For describing the jet uncertainties, the Category Reduction Full Jet Energy Resolution (JER) scheme is used which consists of 13 NPs associated with the JER and over 30 NPs associated with the Jet Energy Scale (JES). Uncertainties on flavour and vertex taggers are also included for a total of 59 NPs related to jets. Muons are important objects in the top and W +jets CRs. Uncertainties are evaluated on the muon energy scale, reconstruction, identification and isolation requirements. In the TCR, WCR and WVR, uncertainties on the single muon trigger are also evaluated. Similarly, for electrons uncertainties on energy scale, energy resolution, reconstruction, identification and isolation requirements are included, though have only a small impact restricted to the application of an electron veto in all the analysis regions.

An uncertainty on the luminosity of 1.7% from the measurement described in Chapter 2, Section 2.1.3 is included. Lastly, an uncertainty on the distribution of pile-up interactions is evaluated by varying the parameter used for the pile-up reweighting, mentioned briefly at the end of Chapter 2, Section 2.2.4.

4.7.2 Theoretical Uncertainties

Numerous theoretical uncertainties are ascribed to the different SM background and SUSY processes. As for the experimental uncertainties, the prescription described here is in accordance with the recommendations of relevant working groups within the ATLAS collaboration. Unless otherwise indicated, these uncertainties are computed by comparing nominal MC samples to samples with particular input parameters varied. These comparisons are performed at truth-level (i.e not simulating the ATLAS detector response) as reconstruction effects are assumed to be independent. Separate NPs describing the same theoretical effect are used for different samples, for instance separate NPs describe the impact of varying the factorisation scale for W +jets and Z +jets, but only one NP per process is used to describe the effect across the different analysis regions. Aside from the cross-section uncertainties, the following uncertainties are all considered as uncertainties on the acceptance of the events in the various analysis regions; the global normalisations of the variation samples is scaled to that of the nominal.

For all the background processes, the impact of the choice of renormalization and factorisation scale is assessed by varying these scales by a factor of 0.5 and 2, both

individually and together, leading to seven total variations (including the nominal). The variation from each up and down pair is described by a single nuisance parameter, hence in total there are three NPs associated with the normalisation and factorisation scale for each background process. These variations are retrieved from internal weights of the samples. For the SUSY processes, separate samples are generated by varying the `scalefact` parameter in MadGraph, which varies the factorisation and renormalization scales in a correlated way, again by 0.5 and 2. Hence, only a single NP is used to describe the factorisation and renormalization uncertainty for the signals.

Uncertainties on the parton distribution functions (PDFs) are also considered for all background processes except for the Higgs background where they are expected to yield an overall negligible contribution to the total uncertainty. There are several origins of PDF uncertainties, including experimental uncertainties from the datasets used to extract them, theoretical uncertainties on their functional form and from missing higher order corrections. These uncertainties are evaluated by taking the envelope from a set of different PDFs with varied parameters. For the V +jets and multi-boson backgrounds, the effect of the uncertainty on the strong coupling constant is also considered in the computation of the PDF uncertainties. PDF uncertainties are assumed to have a negligible effect on the signal acceptances and so are not evaluated for the signal samples.

For the V +jets and multi-boson backgrounds, uncertainties on the matching scale (CKKW scale) and the resummation scale (QSF scale) are evaluated. The matching scale describes the transition between the perturbative QCD regime (used for the matrix element calculations of the hard scatter) and the regime where parton showers are used while the resummation scale is relevant to perturbative QCD calculations (which involve summing over many processes with different amounts of initial or final state parton emission). These uncertainties are assessed via comparisons with dedicated MC samples that vary the resummation scale by a factor of two and the resummation scale from 15 GeV to 30 GeV (with its nominal value being 20 GeV). A matching scale uncertainty is also evaluated for the signal samples by varying the `xqcut` parameter in MadGraph between 0.5 and 2.

For the top and Higgs backgrounds dedicated NPs account for uncertainties in the amount of initial and final state radiation in place of these scale variations. Radiation uncertainties are also evaluated for the signal samples. Several additional theoretical uncertainties are evaluated uniquely for the top background; in particular for $t\bar{t}$ and single-top events there are uncertainties associated with the hard scatter, parton

showering and MC generator. An uncertainty on the interference between Wt and $t\bar{t}$ is also included.

As the normalisations of the multi-boson and Higgs backgrounds are not determined from dedicated control regions, both of these samples carry uncertainties on their total cross-section. These uncertainties account for the effects of PDF uncertainties and next to leading order QCD scale variations. For the multi-boson background, the uncertainty on the cross-section is 6% (using the same tools as used for computing the cross-section [126–130]) while for the Higgs background it is 5% [146]. The signal samples also have uncertainties on their cross-sections that include PDF and scale variations, and like the cross-sections themselves are computed using RESUMMINO v2.0.1 [112,113]. These uncertainties vary between 2% and 5%.

4.7.3 Uncertainties on the Multi-jet background

The uncertainties specific to the multi-jet background were described in detail in Section 4.6.1. To summarise, two systematic uncertainties are included which originate from the systematic and statistical errors on the transfer factor respectively. Additionally, the statistical uncertainty from the finite number of events in CR-A, which is used to construct the multi-jet prediction, is also included and is controlled by a separate NP to those controlling the overall MC statistical error in each region.

4.7.4 Impact of Systematic Uncertainties

Concerning first the impact of these uncertainties on the background estimates, the sizes of the leading sources of systematic uncertainty on the total predicted background yield in each of the SRs is shown in Table 4.18. These uncertainties are post the background only fit to the control regions. From this it can be seen that the MC statistical uncertainty is a large source of uncertainty in each of the SRs. This is partly driven by the impact of W +jets and Z +jets events with relatively large events weights ($1 < \text{weight} < 5$) but is also a consequence of the relatively small number of MC events for these backgrounds.

Furthermore, the impact of these relatively largely weighted events is not confined to the MC statistical uncertainty; some of these events fluctuate in and out of the SRs with the different kinematic-based experimental uncertainties and hence artificially

enhance the size of these uncertainties. This is particularly true of the jet energy resolution uncertainties. Including these artificially large uncertainties is not necessarily a conservative approach as they can lead to artificially large constraints on the relevant NPs, which can in turn lower the uncertainty in other regions and/or on other processes. In SR-BDT1, which was most adversely affected by these events, the effect was mitigated by smoothing the problematic systematic uncertainties between the two bins. This effect also resulted in fit convergence issues for fits including SR-BDT4. To resolve this, for the Z+jets and W+jets backgrounds single placeholder NPs with conservative ranges are used for each of these processes in SR-BDT4 to represent the total experimental uncertainty. The ranges are computed such that the total uncertainty on each of these backgrounds is the same as when using the individual experimental NPs.

Region	SR-BDT1	SR-BDT2	SR-BDT3	SR-BDT4
Total background expectation	49	9.8	17	2.4
MC + multi-jet statistical	19%	17%	19%	26%
Z+jets modelling	4.7%	3.1%	12%	1.3%
W+jets modelling	5.4%	2.8%	3.2%	<1%
Top modelling	2.3%	10%	5.6%	5.4%
Multi-boson modelling	4.6%	16%	10%	30%
Multi-jet systematics	5.8%	10%	6.3%	12%
Normalisation factors	5.6%	2.8%	3.6%	<1%
Tau-lepton exp.	7.7%	18%	14%	11%
Jet exp.	22%	20%	29%	37%
E_T^{miss} soft track exp.	8.0%	6.2%	5.1%	6.9%
Other exp.	<1%	<1%	<1%	<1%
Total systematic uncertainty	31%	37%	38%	58%

Table 4.18: Breakdown of the dominant systematic uncertainties on the total background estimate in each of the signal regions, with ‘exp.’ standing for experimental. For each group of uncertainties listed, the impact of all relevant nuisance parameters are added in quadrature. Note that in general as uncertainties can be correlated the quadrature sum of each group of uncertainties does not equal the total uncertainty.

That the MC statistical uncertainties are significant suggests the sensitivity of this analysis, following the same strategy, could have notable improvement simply

from having larger background MC samples. Further, increasing the MC background samples would likely also lead to an improvement in the performance of the BDTs as some of the training restrictions imposed on the models here could be relaxed. It should however be noted that part of the origin in the limited MC statistics, in particular for W +jets, is due to one of the tau-leptons originating from a jet that fakes the tau-lepton signature. Generating MC samples with an enhanced amount of fake tau-leptons is technically difficult as the fake rate is dependent on the tau-lepton reconstruction and identification algorithms. Hence, for future searches, data-driven methods for estimating this background are likely worth (re-)investigating.

The multi-boson modelling uncertainties also represent a large contribution in each of the SRs, especially for SR-BDT2 and SR-BDT4, for which the multi-boson background is the dominant contribution. In particular, the uncertainties on the matching and resummation scale are dominant being around 25% each in these SRs. Theoretical uncertainties on other background processes are also not negligible.

For uncertainties on the signal samples, a breakdown of the (pre-fit) size of systematic uncertainties on the signal yield for example signal points in each region is shown in Table 4.19. Here, the tau-lepton experimental uncertainties are the dominant contribution to the uncertainty, although modelling uncertainties are also reasonably large for the low mass signals in SR-BDT1.

Region	SR-BDT1	SR-BDT2	SR-BDT3	SR-BDT4
Example signal $m(\tilde{\tau}, \tilde{\chi}_1^0)$	(80, 1)	(240, 80)	(200, 100)	(400, 1)
Example signal yield	40.3	20.5	23.6	7.31
MC statistical	3.5%	4.9%	8.6%	1.4%
Signal modelling	11%	4.4%	3.6%	2.9%
Tau-lepton exp.	17%	14%	13%	8.3%
Jet exp.	2.9%	4.4%	6.6%	2.1%
E_T^{miss} soft track exp.	1.1%	<1%	<1%	<1%
Other exp.	3.8%	3.8%	4.5%	4.7%
Total systematic uncertainty	21%	17%	18%	10%

Table 4.19: Breakdown of the dominant systematic uncertainties for example signal points for each signal region. For each group of uncertainties listed, the impact of all relevant nuisance parameters are added in quadrature. Note that in general as uncertainties can be correlated the quadrature sum of each group of uncertainties does not equal the total uncertainty.

4.8 Results

4.8.1 Unblinded Signal Regions

Having established the validity of the background estimation, and computed the relevant systematic uncertainties, the observed data yields in the SRs can be examined. To first assess the magnitude of any possible discrepancies in the SRs, a background-only fit is performed without the SRs being used as constraining regions (i.e. they are not included in the likelihood function). The predicted SM yields are then compared to data in each of the SRs. A comparison between the predictions and observed data is shown in Figure 4.28 and tabulated in Tables 4.20-4.23. The data yields in the SRs all show a $0.7 - 1.3 \sigma$ deficit compared to the SM prediction and hence provide no evidence for physics beyond the SM. It should be noted that these deficits are strongly correlated with each other due to the significant overlap between the SRs, in particular between SR-BDT2 and SR-BDT3, which are illustrated in Figure 4.29.

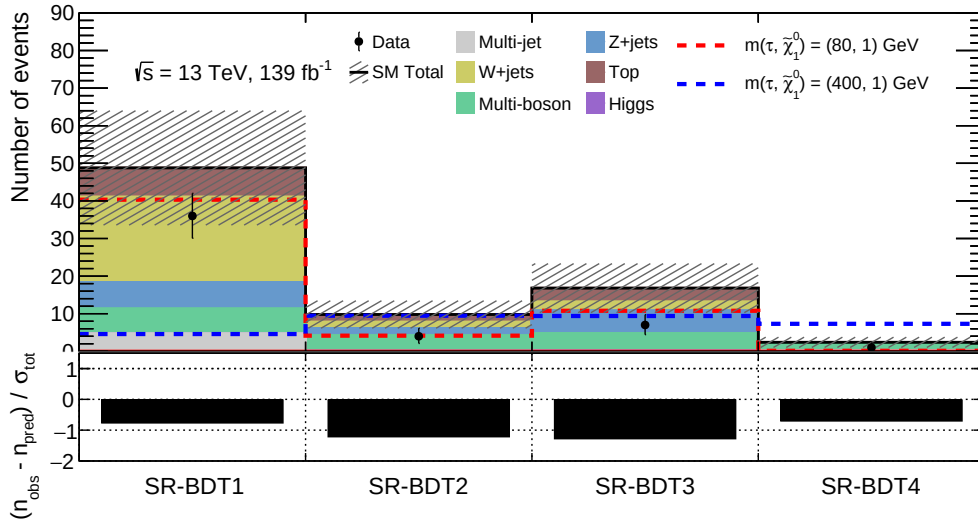


Figure 4.28: Pull plot of the SRs for the direct stau background only fit (in which the signal regions are treated as validation regions).

Chapter 4: Search for Direct Stau Production

Table 4.20: Yields table (at 139 fb^{-1}) for SR-BDT1 with full systematic uncertainties.

SM process	SR-BDT1	SR-BDT1 Bin 1	SR-BDT1 Bin 2
Multi-jet	4.6 ± 2.8	2.2 ± 1.6	2.4 ± 2.3
W+jets	23 ± 10	12 ± 7	11 ± 4
Z+jets	7 ± 7	4 ± 4	2.8 ± 2.6
Top	7.8 ± 2.6	3.3 ± 1.7	4.5 ± 1.9
Multi-boson	6.6 ± 3.1	3.3 ± 1.9	3.3 ± 1.7
Higgs	—	$0.004^{+0.006}_{-0.004}$	—
SM total	49 ± 15	25 ± 10	24 ± 8
Observed	36	22	14
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 1) \text{ GeV}$	40 ± 9	14 ± 4	27 ± 6
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (240, 80) \text{ GeV}$	17.2 ± 2.9	4.8 ± 1.5	12.4 ± 2.4
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (200, 100) \text{ GeV}$	15.4 ± 3.0	3.8 ± 1.6	11.7 ± 2.7
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (400, 1) \text{ GeV}$	4.6 ± 0.9	1.2 ± 0.4	3.4 ± 0.8

Table 4.21: Yields table (at 139 fb^{-1}) for SR-BDT2 with full systematic uncertainties.

SM process	SR-BDT2	SR-BDT2 Bin 1	SR-BDT2 Bin 2
Multi-jet	$0.0^{+1.0}_{-0.0}$	$0.0^{+1.0}_{-0.0}$	$0.00^{+0.16}_{-0.00}$
W+jets	$1.6^{+1.7}_{-1.6}$	$1.6^{+1.6}_{-1.6}$	$0.09^{+0.09}_{-0.09}$
Z+jets	2.0 ± 0.9	1.6 ± 0.8	0.40 ± 0.20
Top	2.1 ± 1.2	1.0 ± 0.9	1.0 ± 0.7
Multi-boson	4.1 ± 2.0	1.7 ± 0.9	2.3 ± 1.2
Higgs	$0.02^{+0.14}_{-0.02}$	$0.02^{+0.13}_{-0.02}$	$0.004^{+0.018}_{-0.004}$
SM total	10 ± 4	5.9 ± 2.8	3.9 ± 1.6
Observed	4	3	1
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 1) \text{ GeV}$	4.2 ± 2.0	3.4 ± 1.4	0.7 ± 0.7
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (240, 80) \text{ GeV}$	21 ± 4	7.5 ± 1.6	13.0 ± 2.4
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (200, 100) \text{ GeV}$	11.6 ± 2.4	5.4 ± 1.0	6.2 ± 1.9
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (400, 1) \text{ GeV}$	9.5 ± 1.1	2.1 ± 0.4	7.4 ± 0.9

Chapter 4: Search for Direct Stau Production

Table 4.22: Yields table (at 139 fb^{-1}) for SR-BDT3 with full systematic uncertainties.

SM process	SR-BDT3	SR-BDT3 Bin 1	SR-BDT3 Bin 2
Multi-jet	$0.0^{+1.1}_{-0.0}$	$0.0^{+1.0}_{-0.0}$	$0.00^{+0.26}_{-0.00}$
W+jets	2^{+4}_{-2}	2^{+4}_{-2}	0.5 ± 0.5
Z+jets	6 ± 4	$1.9^{+2.7}_{-1.9}$	4.1 ± 2.6
Top	3.7 ± 1.3	2.7 ± 1.1	1.0 ± 0.9
Multi-boson	4.6 ± 2.2	2.0 ± 1.2	2.5 ± 1.3
Higgs	$0.13^{+0.15}_{-0.13}$	$0.13^{+0.14}_{-0.13}$	$0.001^{+0.012}_{-0.001}$
SM total	17 ± 7	9 ± 6	8.2 ± 3.3
Observed	7	6	1
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 1) \text{ GeV}$	11 ± 4	8.6 ± 3.0	2.2 ± 1.4
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (240, 80) \text{ GeV}$	24 ± 4	8.5 ± 1.6	15.2 ± 2.6
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (200, 100) \text{ GeV}$	15.1 ± 2.9	4.3 ± 1.1	10.8 ± 2.4
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (400, 1) \text{ GeV}$	9.4 ± 1.4	3.7 ± 0.7	5.7 ± 1.1

Table 4.23: Yields table (at 139 fb^{-1}) for SR-BDT4 with full systematic uncertainties.

SM process	SR-BDT4
Multi-jet	$0.20^{+0.28}_{-0.20}$
W+jets	$0.0^{+0.7}_{-0.0}$
Z+jets	$0.2^{+0.9}_{-0.2}$
Top	$0.16^{+0.19}_{-0.16}$
Multi-boson	1.8 ± 1.0
Higgs	$0.000^{+0.015}_{-0.000}$
SM total	2.4 ± 1.4
Observed	1
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 1) \text{ GeV}$	$0.13^{+0.22}_{-0.13}$
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (240, 80) \text{ GeV}$	10.4 ± 1.8
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (200, 100) \text{ GeV}$	3.2 ± 1.1
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (400, 1) \text{ GeV}$	7.3 ± 0.9

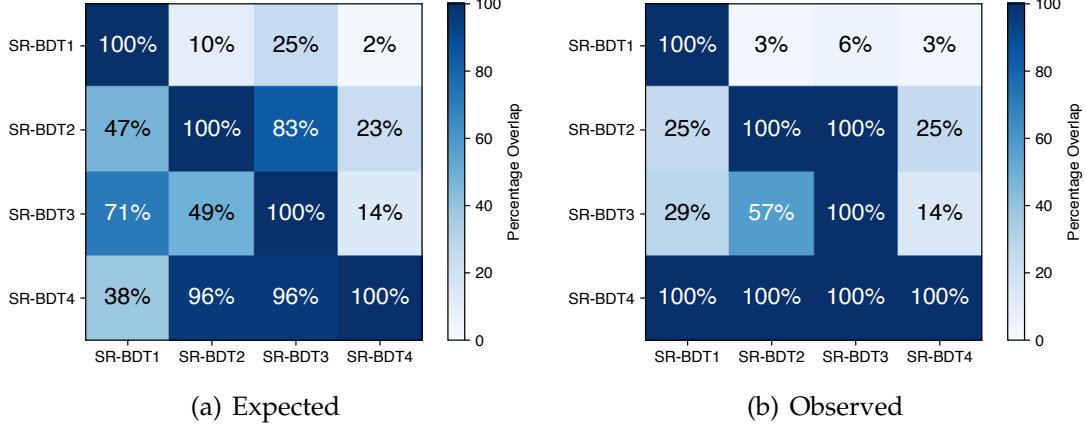


Figure 4.29: Expected and observed overlaps between the direct stau SRs. Each row represents the percentage of events shared between the labelled SR with each of the other SRs.

Background-only fits including each SR in the likelihood function in turn were also performed. The agreement in the respective SRs improves in these fits as the background predictions in the SRs are scaled down with respect to the CR only fit. These fits were checked to ensure no significant pulls, constraints or correlations are introduced on the systematic NPs from the inclusion of the SRs and that the fits are of good quality as assessed by the MINUIT2 [88] minimizer. Figures 4.30-4.33 show the post fit nuisance parameter (NP) pull plots for these fits. The colour and naming of the NPs follows the same convention as Figure 4.12. In each plot, the NPs labelled “QCDStat” control the uncorrelated multi-jet statistical uncertainties in each of the SR bins and have very large uncertainties on their values, both before and after the fit. Some NPs have uncertainties greater than one due to both up and down templates having yields less than the nominal in (one of) the SR bins in which a deficit in the data with respect to the SM prediction is observed. Certain NPs also have post fit constraints less than one as they are able to be constrained by the SR yield observations, despite the limited SR statistics, due to the corresponding uncertainties being large in these regions. Figures 4.34-4.37 show the NPs with the largest correlations in these fits. Comparing these plots to the correlation matrix for the background-only fit to the CRs in Figure 4.13, it can be seen no large unexpected correlations between NPs are introduced by the inclusion of the SRs.

Fully unblinded post-fit BDT score distributions for events passing all the SR selections except for the BDT score requirements are shown in Figure 4.38. These distributions also come from the background-only fit to the CRs. The lower bins

Chapter 4: Search for Direct Stau Production



Figure 4.30: Nuisance parameter pull plot for the background-only fit including the CRs and SR-BDT1.

Chapter 4: Search for Direct Stau Production

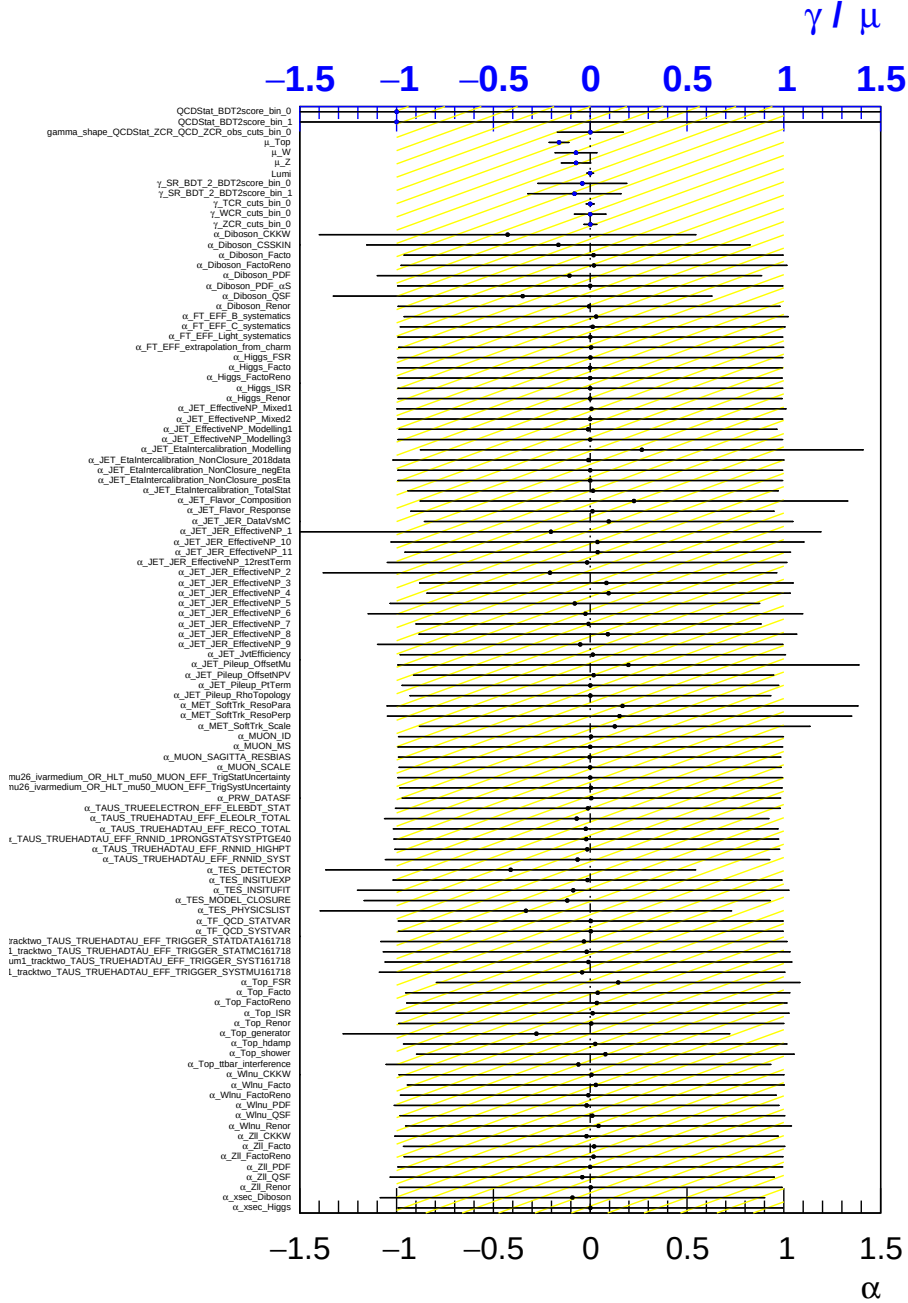
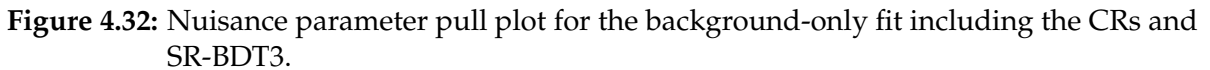
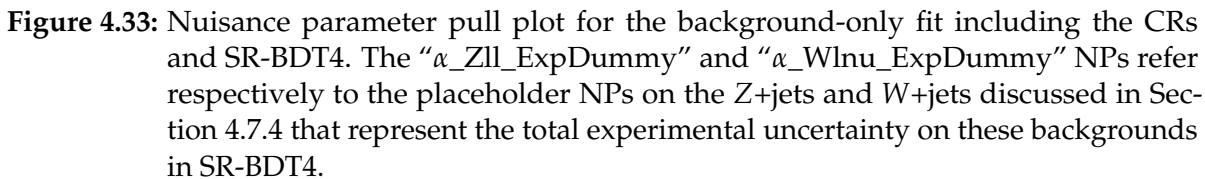


Figure 4.31: Nuisance parameter pull plot for the background-only fit including the CRs and SR-BDT2.





Chapter 4: Search for Direct Stau Production

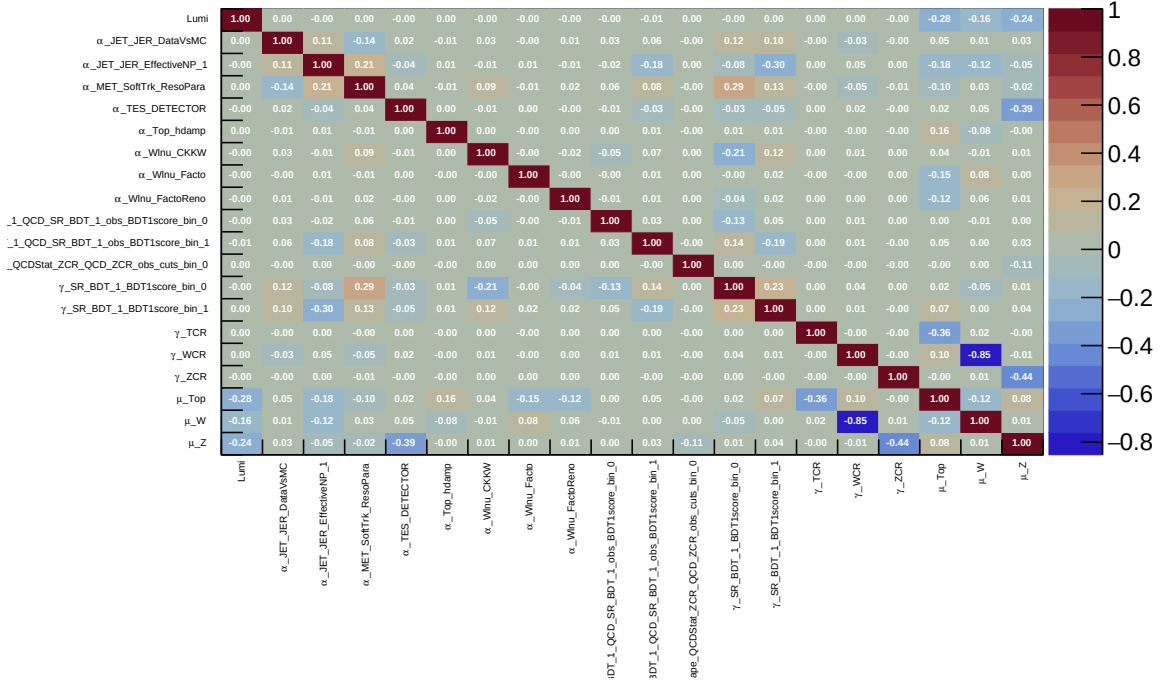


Figure 4.34: Reduced nuisance parameter correlation matrix for the background-only fit including the CRs and SR-BDT1.

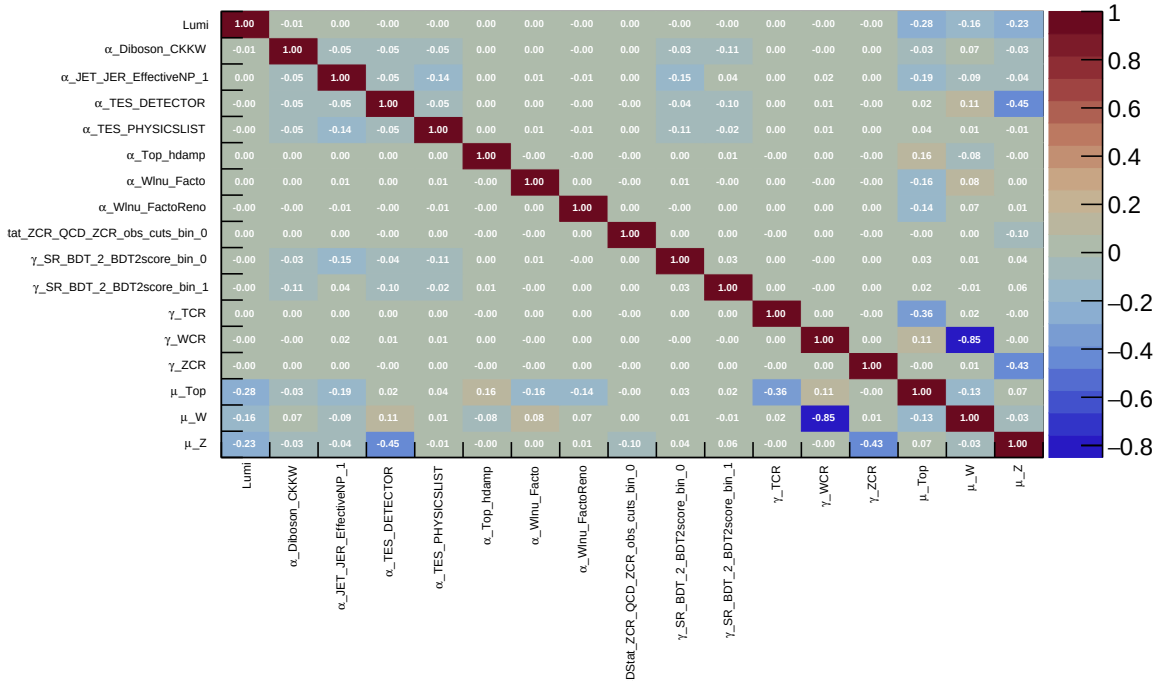


Figure 4.35: Reduced nuisance parameter correlation matrix for the background-only fit including the CRs and SR-BDT2.

Chapter 4: Search for Direct Stau Production

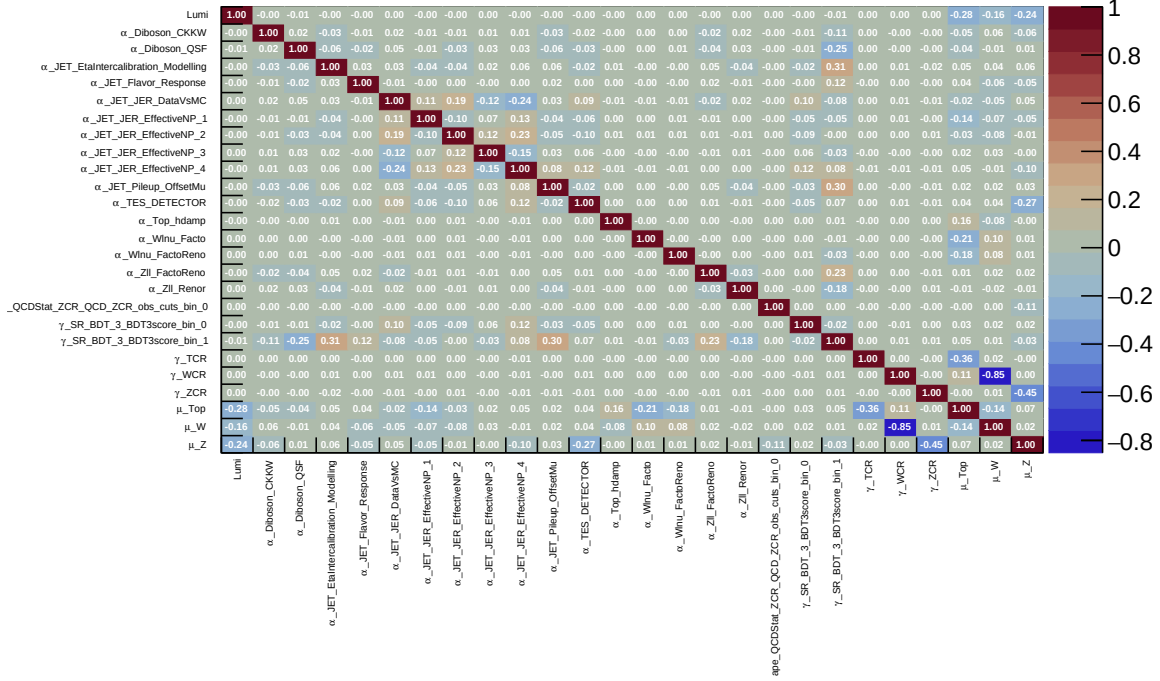


Figure 4.36: Reduced nuisance parameter correlation matrix for the background-only fit including the CRs and SR-BDT3.

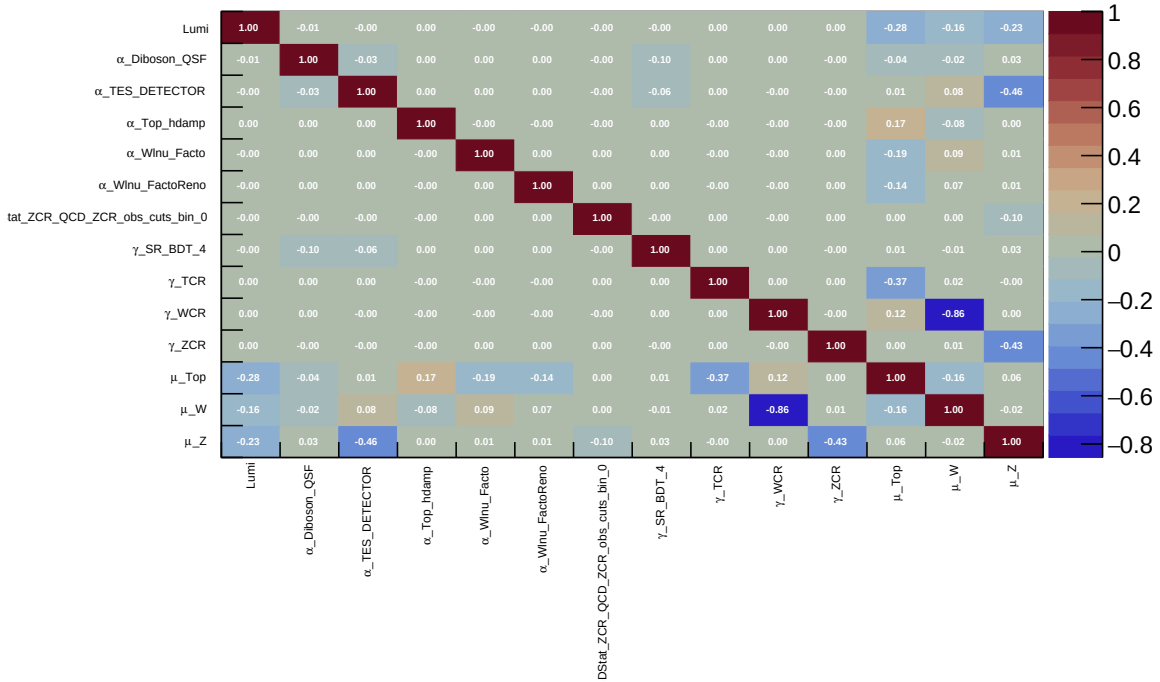


Figure 4.37: Reduced nuisance parameter correlation matrix for the background-only fit including the CRs and SR-BDT4.

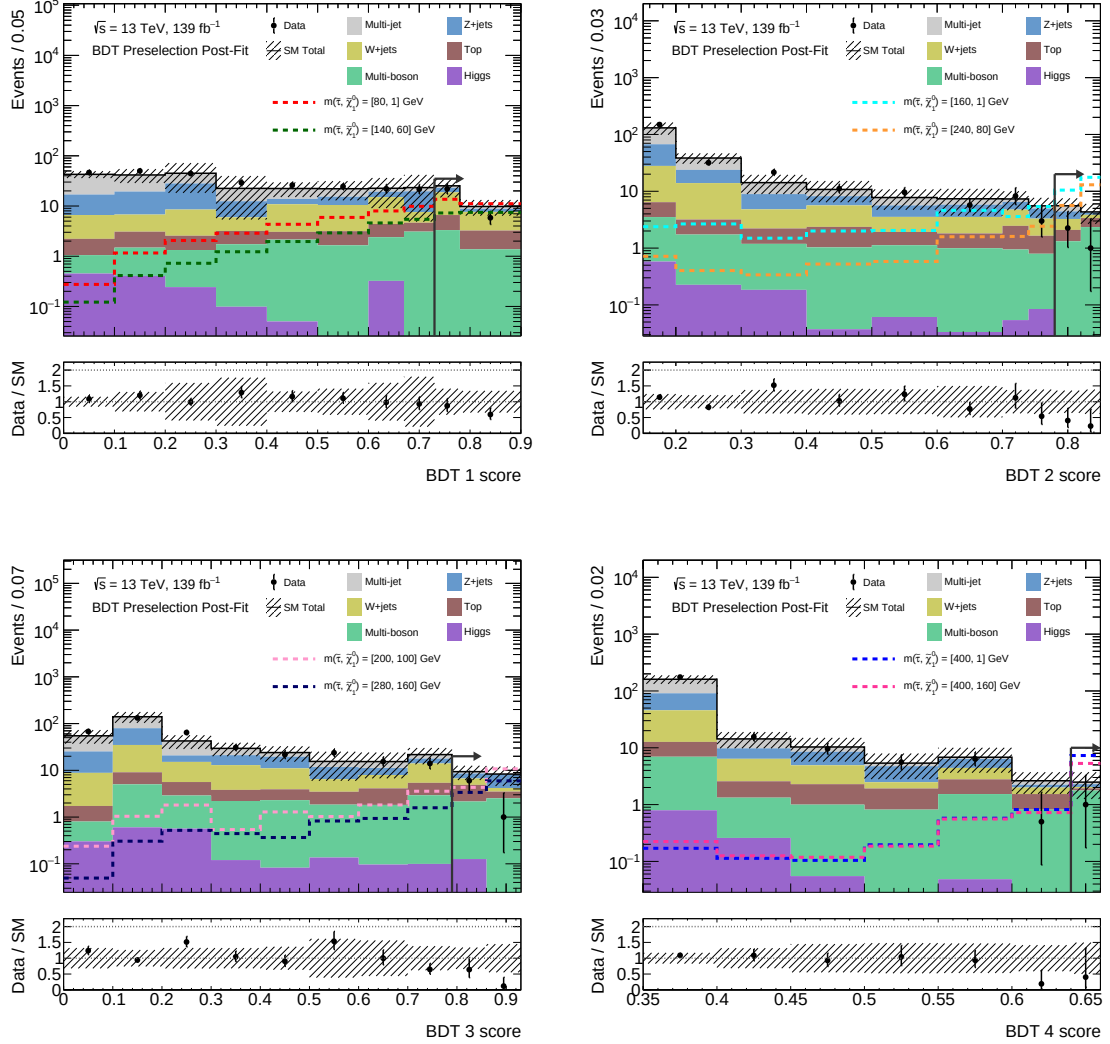


Figure 4.38: Unblinded BDT score distributions for events passing the SR selections except for the BDT score requirements, which are indicated by arrows.

in these distributions have significant overlap with the distributions plotted for the inclusive validation region InclVR (Figure 4.27) but further show there is no significant evidence of a shape trend in the comparison between data and the SM predictions when the high BDT score bins are included. As they contain a reasonable number of statistics, unblinded post-fit kinematic distributions are also shown for a selection of variables in SR-BDT1 and SR-BDT3 in Figures 4.39 and 4.40 respectively, again using only the CRs in the likelihood function. In general reasonable agreement between the data and SM predictions is seen.

Chapter 4: Search for Direct Stau Production

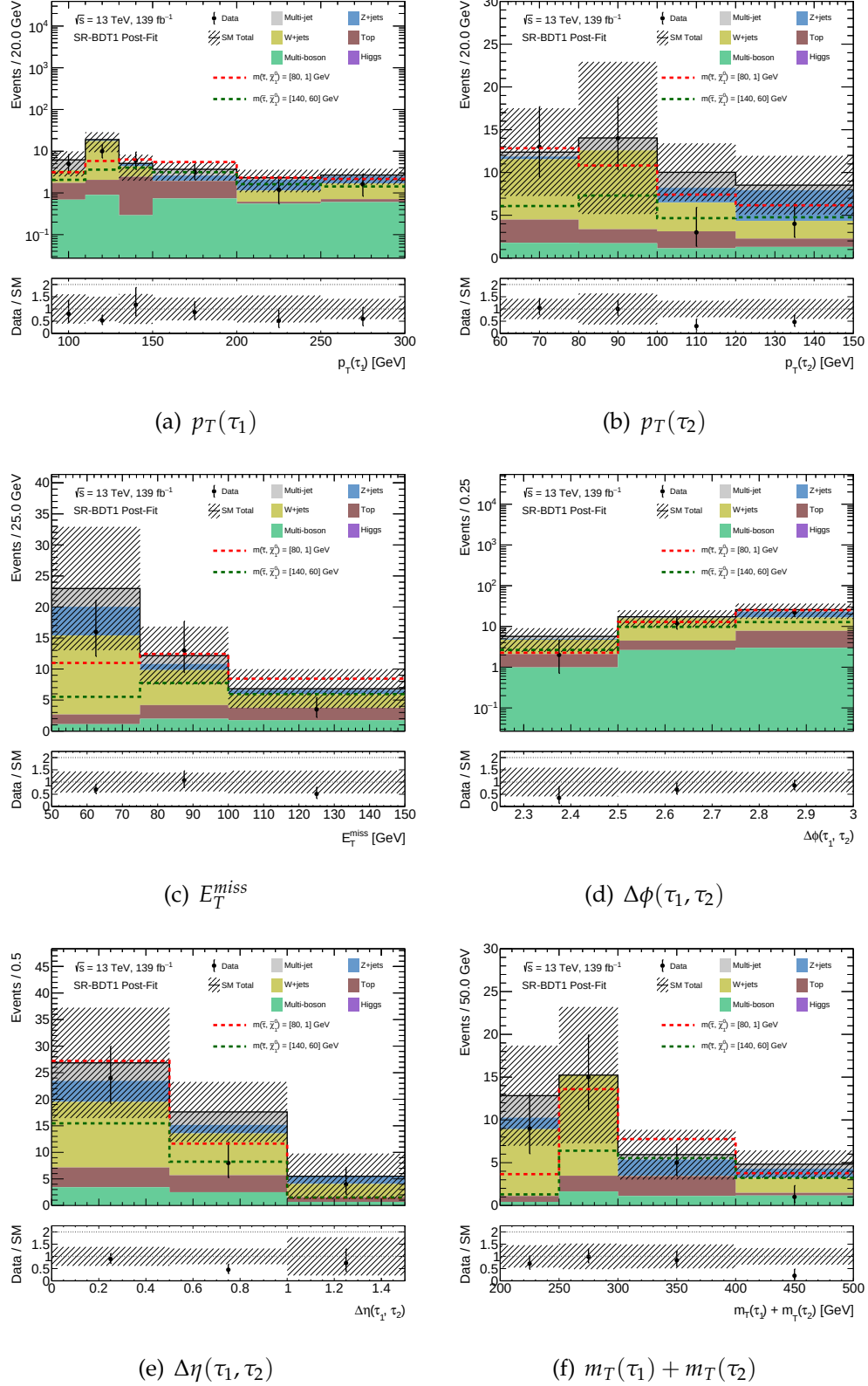


Figure 4.39: Post-fit kinematic distributions for SR-BDT-1. First and last bins include the underflow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

Chapter 4: Search for Direct Stau Production

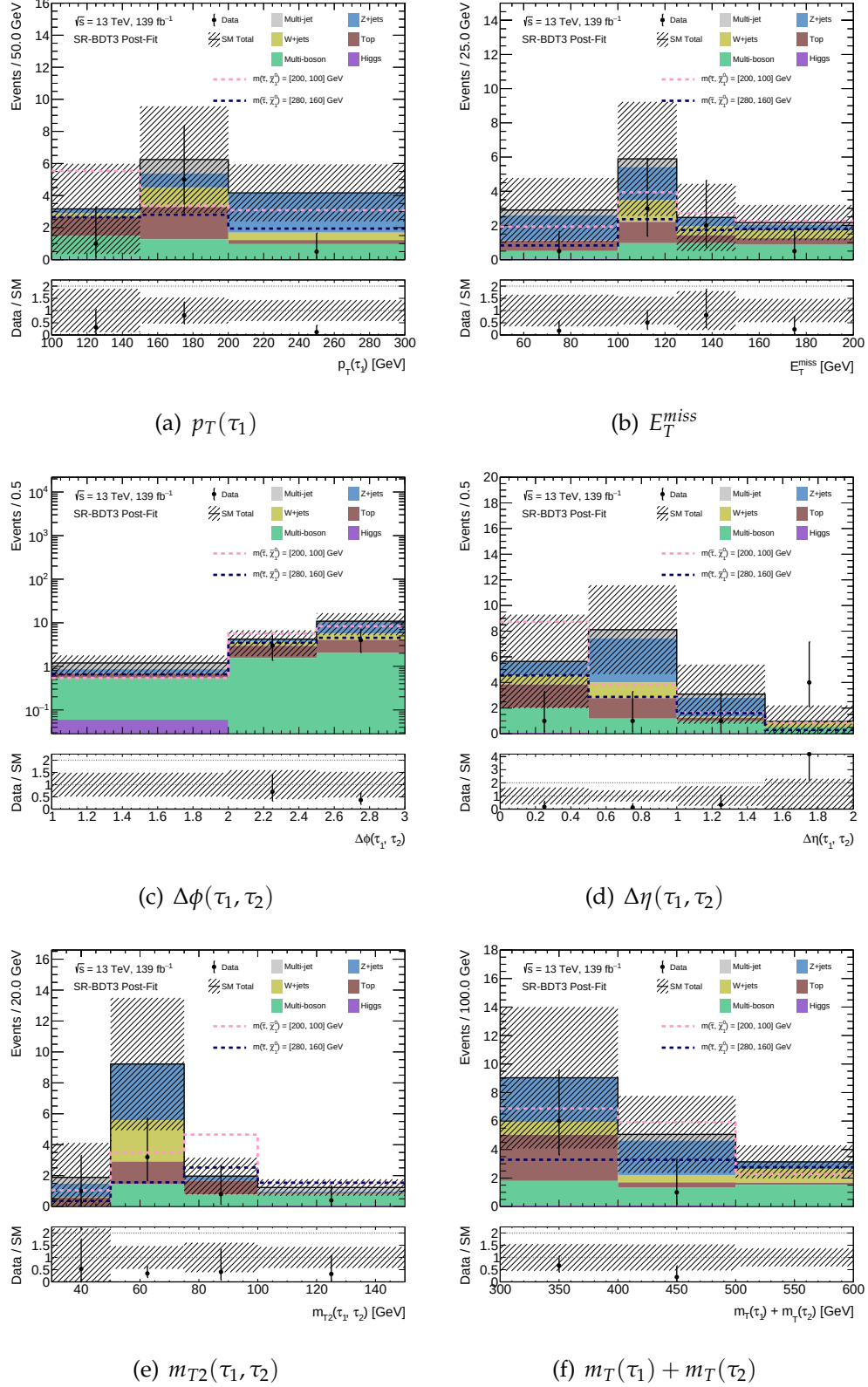


Figure 4.40: Post-fit kinematic distributions for SR-BDT-3. First and last bins include the underflow and overflow respectively. The uncertainty band represents the full post-fit systematic uncertainty on the background prediction.

4.8.2 Model Independent Upper Limits

As no significant excesses are observed, model independent limits are set on both the number of signal events and on the visible cross-section, defined in Chapter 3, Equation 3.8. To compute these, first upper limits on the number of signal events is determined by including each signal region as a constraining region in the fit in turn and fitting an upper limit to on the signal strength for a place-holder signal with a yield of one in the SR under consideration. The visible cross-section limits are calculated as the limit on the number of signal events divided by the total integrated luminosity. The expected and observed model independent 95% CL upper limits on the visible cross-section and number of signal events is shown in Table 4.24. Due to the limited number of statistics available, for SR-BDT2, SR-BDT3 and SR-BDT4 the limits have been computed using 10,000 pseudo-experiments for fits using different values of the signal strength while for SR-BDT1 the asymptotic approximation has been used. Due to the deficits compared with the SM prediction, the observed limits are stronger than the expected limits for each SR.

Signal Region	Exp. σ_{vis}^{95} [fb]	Obs. σ_{vis}^{95} [fb]	S_{exp}^{95}	S_{obs}^{95}
SR-BDT1	0.22	0.18	$30.3^{+9.7}_{-7.3}$	24.5
SR-BDT2	0.06	0.04	$8.0^{+3.7}_{-3.0}$	5.7
SR-BDT3	0.07	0.04	$9.7^{+3.5}_{-3.6}$	5.9
SR-BDT4	0.03	0.03	$4.2^{+1.8}_{-0.8}$	3.8

Table 4.24: Expected and observed 95% CL upper limits on the visible cross-section (σ_{vis}^{95}) and number of events S^{95} in each of the direct stau search signal regions.

4.8.3 Model Dependent Upper Limits

Exclusion limits are also set on the simplified direct stau models. These are computed from fits that include the signals, CRs and relevant SR. Observed and expected CLs values are computed for each signal point in each SR, for mass-degenerate left+right-handed, left-handed only and right-handed only stau interpretations. As for the model-independent limits, the fits for SR-BDT1 have been run using the asymptotic approximation while the fits using the others SRs have been run using 10,000 pseudo-experiments for each signal point. For each SR, pulls and correlations of the systematic NPs were checked for example signal points to verify the inclusion of the signals caused no large differences to the background-only fits and all fits are required to be of good quality.

Expected and observed exclusion contours drawn using the results of these fits are shown for each SR in Figures 4.41-4.43 for the different interpretations. These contours indicate where the CLs value crosses 0.05; hence models with mass points contained within these contours can be said to be excluded at 95% confidence level. The yellow bands correspond to the $\pm 1\sigma$ uncertainty on the expected exclusion limit, indicating the effect of the overall systematic and statistical uncertainties. The solid line represents the observed exclusion limit, with an error band given by varying the signal cross-sections by their uncertainties. As a deficit in data compared to the SM prediction is seen in all the SRs, the observed limit is stronger than the expected (by approximately one sigma). The limits set by previous searches at ATLAS and by the experiments at the Large Electron Positron (LEP) collider, discussed at the start of this chapter, are shown in grey.

Overall exclusion contours for the different interpretations are shown in Figure 4.44. These are constructed by, for each mass point, using the expected and observed CLs value from the fit using the SR with the best expected sensitivity (lowest expected CLs value) as determined using the asymptotic approximation for all fits (to avoid discrepancies arising from using a finite number of pseudo-experiments). Which SR is used for each mass point is illustrated in Figure 4.45 for each of the different interpretations. Compared to the expected limits from the previous ATLAS search, shown in grey, there is significantly improved coverage at low and high $\tilde{\tau}$ mass for all interpretations.

In particular, for the mass degenerate $\tilde{\tau}_L + \tilde{\tau}_R$ interpretation, the exclusion contour now extends to stau masses as low as 80 GeV, which closes the previous gap between

the ATLAS limits and the LEP limits, and stau masses as high as 500 GeV. For the left-handed only interpretation, the observed limit extends to $\tilde{\tau}_L$ masses as high as 420 GeV, while for the right-handed only interpretation, the sensitivity indicated by Figure 4.44(c) is the first sensitivity to the right-handed only scenario obtained by ATLAS and indicates how the depth of the sensitivity in this search is improved as well as the coverage. Models with right-handed staus with masses between 100 GeV and 350 GeV are excluded. Plots showing the upper limits on the cross-sections can be found in the appendix, Section A.2.6.

Comparing the best expected SRs for each mass point in Figure 4.45, to the mass points used for training each of the BDTs in Figure 4.3, it can be seen the clustering of the signal samples into low, medium and high masses has worked well as for low mass points SR-BDT1 preforms best; for medium stau masses one of SR-BDT2/SR-BDT3 performs best and SR-BDT4 performs best for the high mass points. Concerning the medium $\tilde{\tau}$ masses, the split between SR-BDT2 and SR-BDT3 differs slightly from the training groups. This difference could be due to a number of factors, including that the grouping was optimized assuming a single bin SR with a flat systematic uncertainty of 20% and using a simplified calculation of the significance. Further, consulting Figures 4.41-4.43, it can be seen that SR-BDT2 and SR-BDT3 have very similar sensitivity overall. This indicates that similar sensitivity for these points could likely have been achieved by using them to train a single BDT rather than two and that if two BDTs were to be used for these points then the grouping used here is potentially not optimal for maximising the sensitivity to all mass points equally (recall that the grouping used here was defined by prioritizing the compressed mass points over the non-compressed ones).

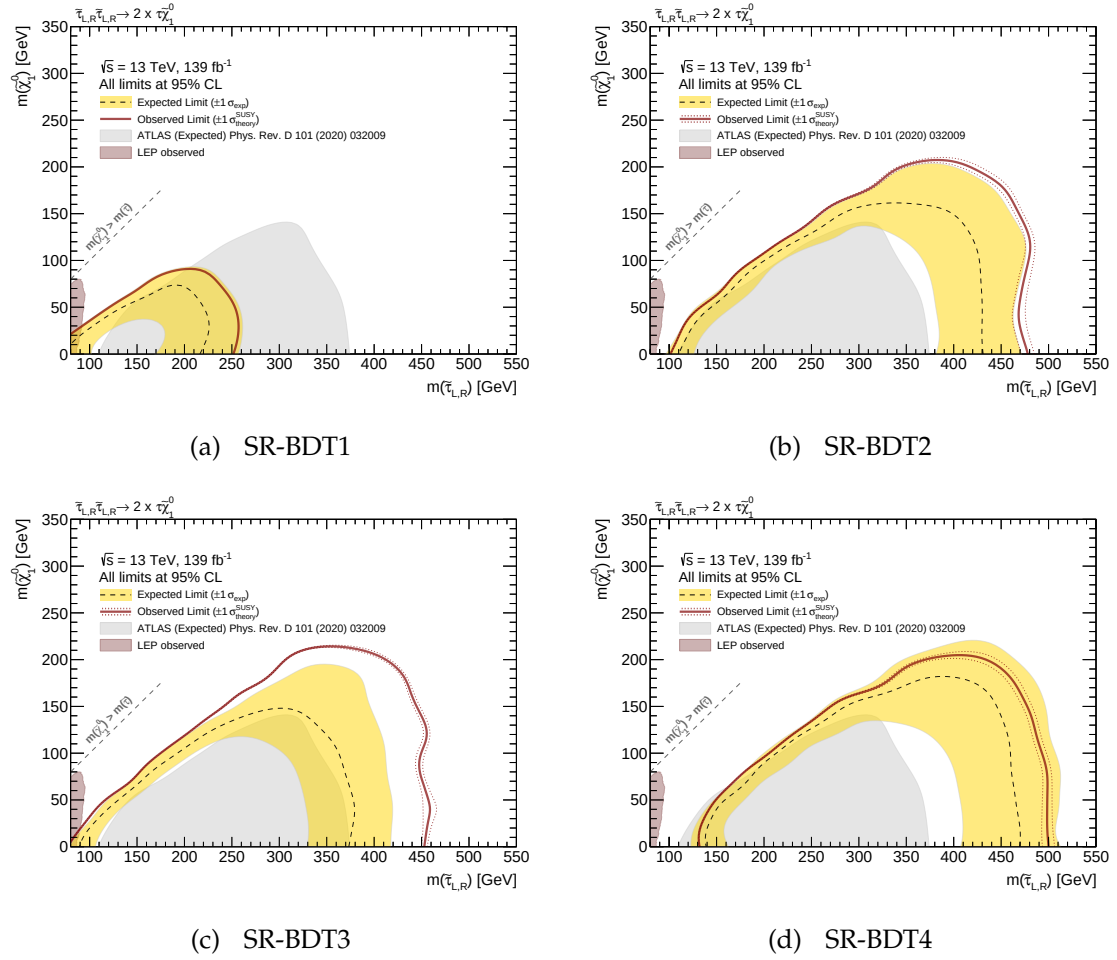


Figure 4.41: Expected and Observed 95% CL exclusion contours for each signal region for the mass degenerate $\tilde{t}_L + \tilde{t}_R$ interpretation. The yellow band corresponds to the $\pm 1\sigma$ uncertainty. The shaded grey area corresponds to the expected exclusion contour of the first ATLAS full Run 2 stau search [102].

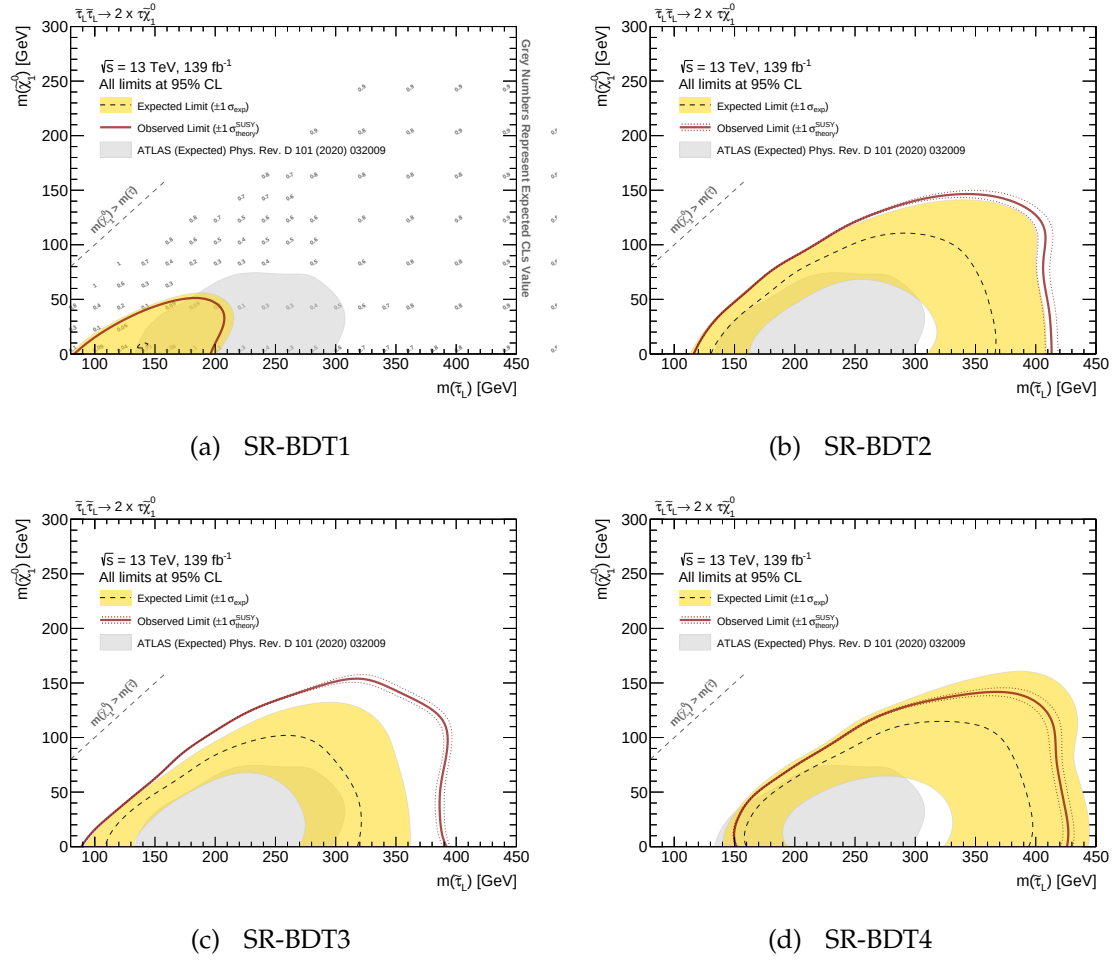
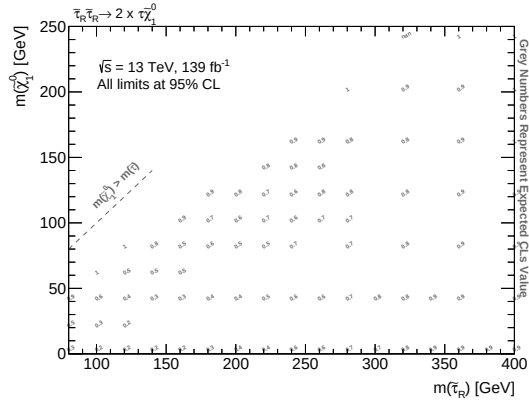
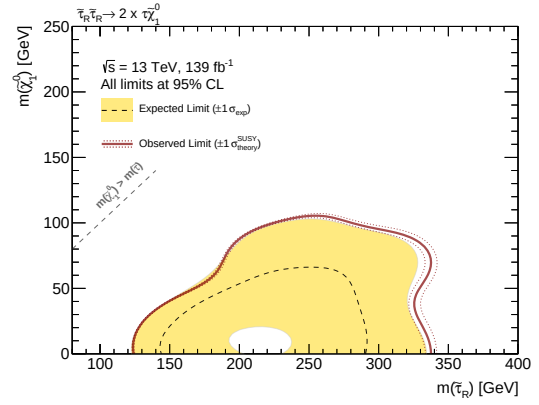


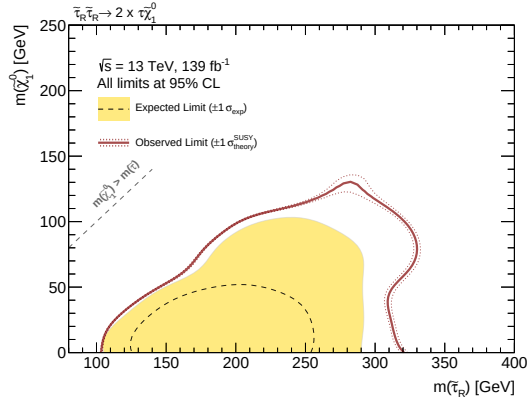
Figure 4.42: Expected and Observed 95% CL exclusion contours for each signal region for the $\tilde{\tau}_L$ only interpretation. The yellow band corresponds to the $\pm 1\sigma$ uncertainty. For SR-BDT1, only a couple of the CL_s values fall below the 0.05 threshold so a contour cannot be plotted. The shaded grey area corresponds to the expected exclusion contour of the first ATLAS full Run 2 stau search [102].



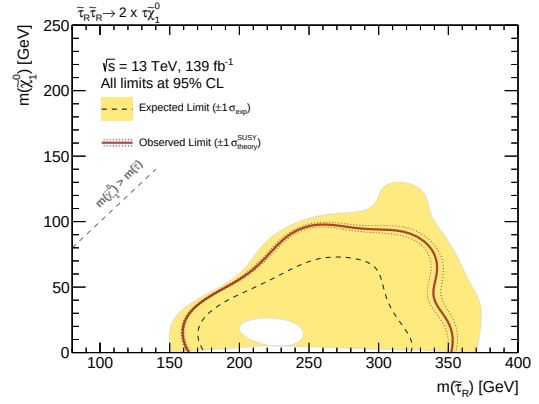
(a) SR-BDT1



(b) SR-BDT2



(c) SR-BDT3



(d) SR-BDT4

Figure 4.43: Expected and Observed 95% CL exclusion contours for each signal region for the $\tilde{\tau}_R$ only interpretation. The yellow band corresponds to the $\pm 1\sigma$ uncertainty. No limit is shown from the previous full Run 2 ATLAS search as no points were able to be excluded at 95% CL. For SR-BDT1, none of the CL_s values fall below the 0.05 threshold so a contour cannot be plotted. Instead, the expected CL_s values are shown for the different points to show how near they are to this threshold.

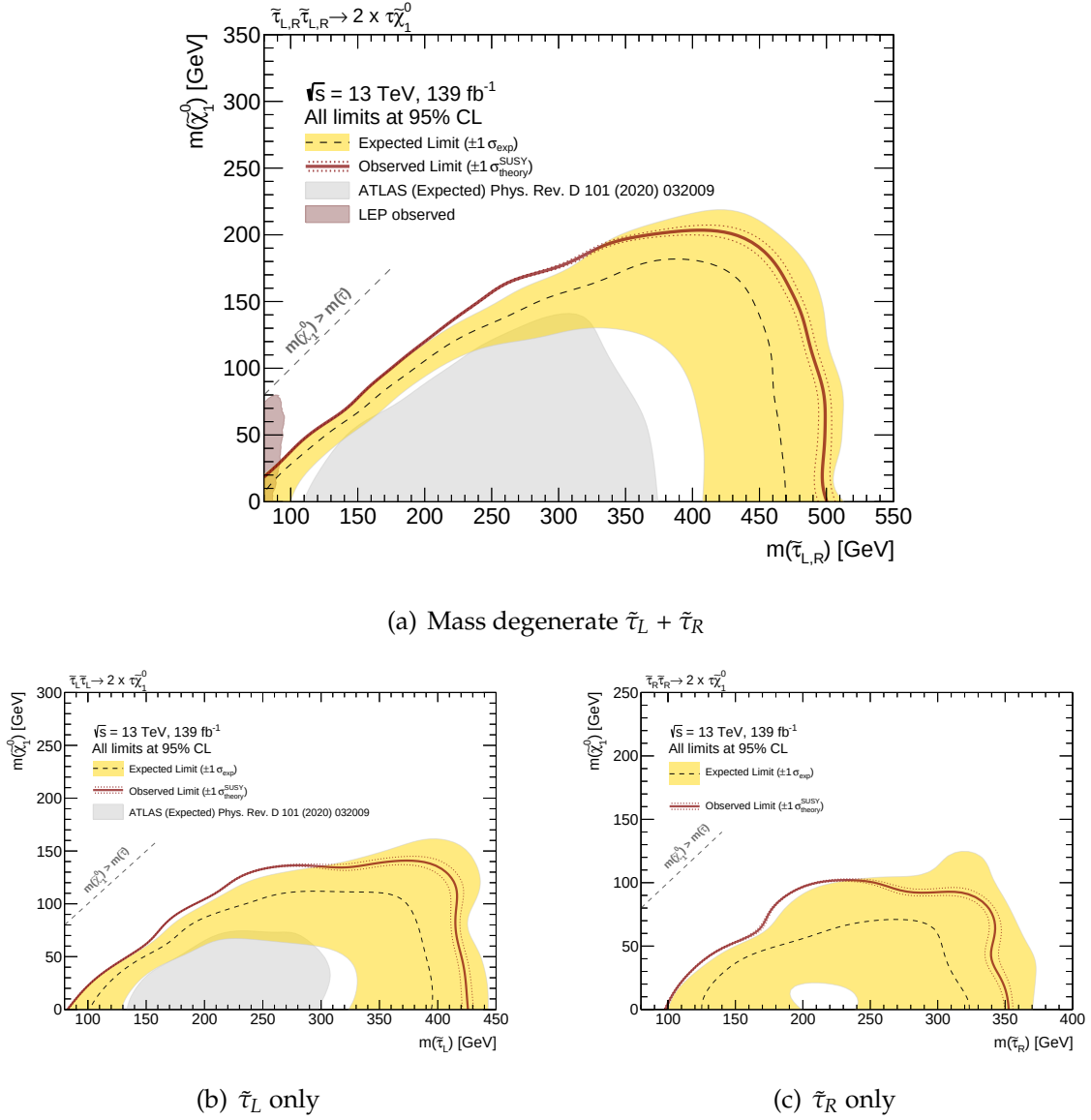


Figure 4.44: Overall exclusion contours for stau pair-production for the different interpretations. Each contour is made using, for each point, the SR with the taking the best expected CLs value, which are shown in Figure 4.45. The yellow band corresponds to the $\pm 1\sigma$ uncertainty. The shaded grey area corresponds to the expected exclusion contour of the first stau search as shown in [102]

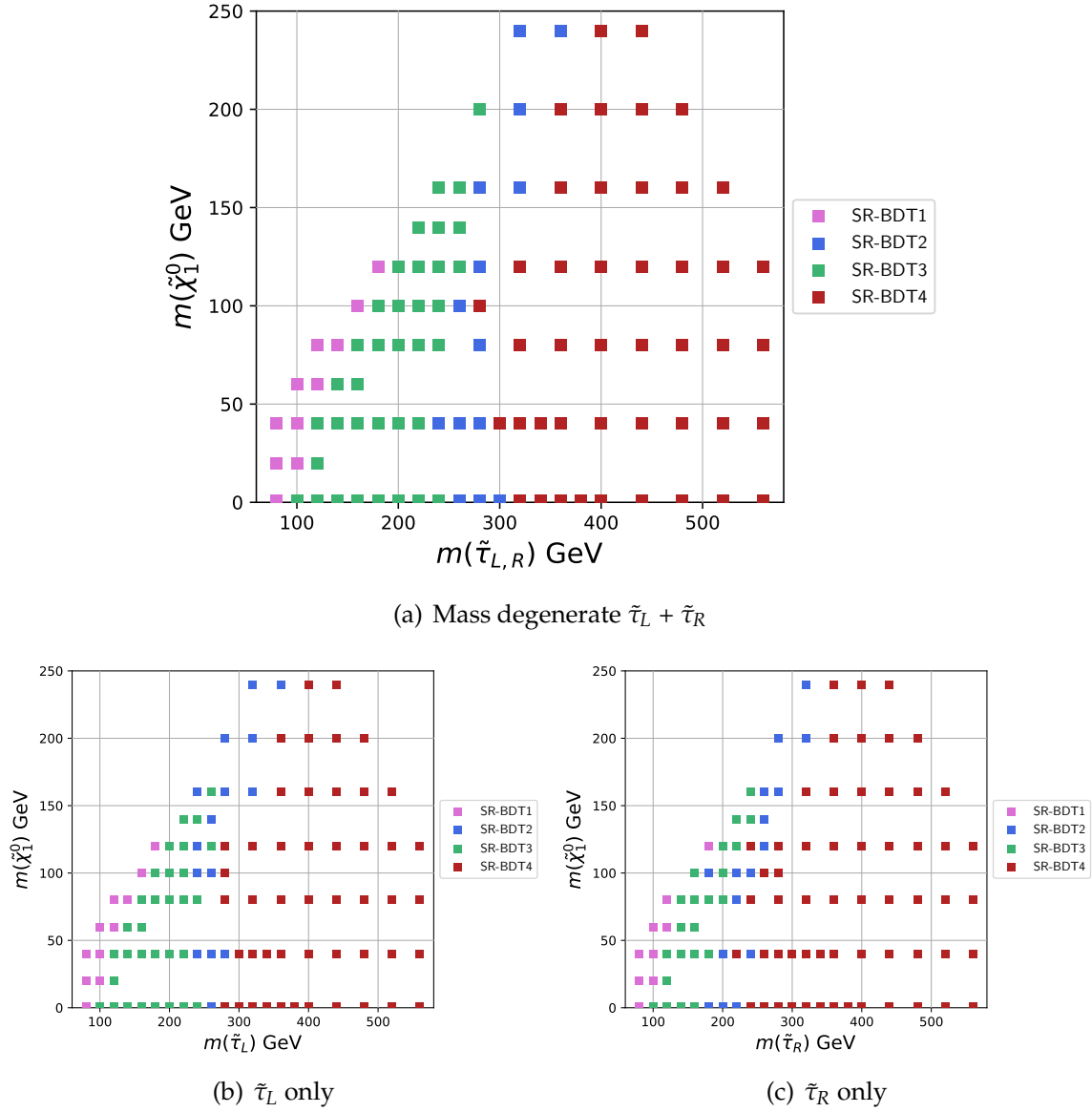


Figure 4.45: Plots showing which SR has the best exclusion limit for each signal point for each of the different interpretations.

4.9 Conclusion

A new search for direct stau production in a final state with two hadronically decaying tau-leptons and missing transverse energy, using the 139 fb^{-1} of integrated luminosity of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ collected by the ATLAS detector at the Large Hadron Collider from 2015 to 2018, has been presented. Good agreement between data and the SM predictions is observed in all signal regions, hence no evidence for the existence of staus or neutralinos is found. The results are used to set model independent limits on the visible cross-section and exclusion limits on simplified SUSY models. The use of BDTs to separate the SUSY signal from the SM backgrounds has significantly extended the reach of this search to new stau and neutralino masses for mass degenerate left+right, left-handed only and right-handed only interpretations. At 95% CL, mass-degenerate $\tilde{\tau}_{L,R}$ is excluded up to 500 GeV, $\tilde{\tau}_L$ up to 420 GeV, and $\tilde{\tau}_R$ up to 350 GeV, in all cases setting the strongest limits compared to previous searches. The sensitivity to the $\tilde{\tau}_R$ only interpretation obtained here is the first to be achieved with the ATLAS experiment.

Chapter 5

Searches for Ultra-compressed Charginos and Neutralinos

“We’ve learned from experience that the truth will out. Other experimenters will repeat your experiment and find out whether you were wrong or right.”

Richard Feynman, physicist

Having established new limits on direct stau production in the previous chapter, attention now turns to other possible signatures of Supersymmetry (SUSY). Within the landscape of SUSY models, R-parity conserving models in which charginos and neutralinos are the lightest and next to the lightest SUSY particles (LSP and NLSPs respectively) and the mass difference between the LSP and NLSPs are less than a few GeV have some of the weakest exclusion limits. Such models are said here to have ‘ultra-compressed’ electroweakino mass spectra and are referred to as Ultra-Compressed models (UC-models) in the rest of this chapter. The limits on these UC-models are relatively weak as, for reasons discussed in this chapter, the production of compressed electroweakinos is a challenging signature to search for experimentally. However, such models are well motivated due to the lightest neutralino being a compelling dark matter candidate. Hence, the ultra-compressed region is an important area of parameter space to cover.

Following an overview of these UC-models and previous searches for them in Section 5.1, this Chapter then presents results of two analyses aiming to extend the reach of the ATLAS experiment to these SUSY scenarios. The first in Section 5.2 features

a re-interpretation of a previous ATLAS search [152] in a topology consisting of a high p_T jet and large E_T^{miss} , here-on referred to as the mono-jet search/topology. This search targeted different BSM physics models yet may also have interesting sensitivity to the UC-models. As no significant excesses compared to the SM predictions were seen when this search was unblinded, the re-interpretation presented here evaluates the exclusion limits that can consequently be placed on the UC-models. The second analysis presented in Section 5.3 is a sensitivity study of the Vector Boson Fusion (VBF) production channel. This analysis is motivated by a search performed by the CMS experiment [153] which, again after seeing no significant excesses compared to the SM predictions, was able to set what, at time of writing, are nominally the strongest limits on some of these UC-models. The purpose of the sensitivity study is two-fold; 1. to assess whether a similar search in ATLAS could reproduce these limits and 2. to compare the sensitivity of a search in this channel to that of the mono-jet channel by comparing the results to the re-interpretation described above. This comparison and other conclusions are discussed at the end of this chapter in Section 5.4.

5.1 Overview of Ultra-compressed Electroweak SUSY Models and Searches

In R-parity conserving models of Electroweak SUSY, charginos and neutralinos are produced in pairs via interactions with W- and Z-bosons. Which production processes are allowed i.e. which combinations of charginos and neutralinos can be pair-produced, depends upon the mixing of the electroweakinos; whether they are wino, bino or higgsino like. The ATLAS and CMS experiments have used two particular scenarios to generate simplified models with which to design searches. The first scenario, referred to here on as the higgsino scenario (abbreviated as HH), is that the $\tilde{\chi}_2^0$, $\tilde{\chi}_1^0$ and $\tilde{\chi}_1^\pm$ are all pure higgsino states and the mass of the $\tilde{\chi}_1^\pm$ is set to be halfway between the masses of the $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$. More generally, the $\tilde{\chi}_1^\pm$ is expected to be closer in mass to the $\tilde{\chi}_2^0$, but for the UC mass spectra considered here the exact hierarchy does not affect the phenomenology. In the second scenario, referred to as the wino/bino scenario (abbreviated as WB), the $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ are mass degenerate pure wino states while the $\tilde{\chi}_1^0$ is a lighter pure bino state. Used together, these models serve as useful benchmarks to capture the phenomenology of more complex interpretations. As a result of being produced via the electroweak interaction, all electroweakino production modes have

relatively small cross-sections compared to strongly produced processes, with cross-sections typically of the order of pb. For the same process the cross-section for the wino/bino scenario is larger than in the higgsino scenario due to larger couplings between the wino-like electroweakinos and the W/Z -bosons.

Models with compressed mass spectra present additional challenges on top of these small production cross-sections. In particular, in the models considered here the $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ usually decay (with 100% branching fraction for the simplified models) to a $\tilde{\chi}_1^0$ and either a W - or Z -boson. Searches for these models generally rely on reconstructing leptons from the W - or Z -boson decays (as multi-lepton signatures have smaller SM backgrounds than multi-jet signatures, and, as discussed in Chapter 2 leptons can be reconstructed down to lower p_T). Depending on the exact production mode, it is possible for up to three leptons to be present in the final state. The magnitude of the p_T of these leptons is related to the mass splitting between the $\tilde{\chi}_2^0/\tilde{\chi}_1^\pm$ and the $\tilde{\chi}_1^0$; when this mass splitting is very small the leptons will be produced with very low p_T . Consequently, the efficiency with which these leptons can be reconstructed is low, thereby significantly hampering the sensitivity relative to models with non-compressed mass spectra. Additionally, the amount of SM background with fake leptons also tends to increase when the leptons have lower p_T and these backgrounds are generally more difficult to estimate.

To circumvent these issues, boosted topologies have been used in previous searches for these compressed models in which the produced SUSY particles recoil against initial state radiation (ISR), typically in the form of a high p_T jet emitted from one of the hard-scatter partons. Boosted topologies increase the p_T of the leptons and also result in a large E_T^{miss} (from the two $\tilde{\chi}_1^0$), which allows for discrimination against the SM background and is also used for triggering (the leptons in even these boosted scenarios are too soft to triggered on). As an alternative to fully reconstructing the leptons, searches using an ISR jet plus disappearing track signature have also been performed by the ATLAS [154] and CMS [155] experiments. The disappearing track signature results from a $\tilde{\chi}_1^\pm$ having a sufficiently long lifetime, as a result of it only being slightly heavier than the $\tilde{\chi}_1^0$, such that it decays inside the Inner Detector (ID) material, thus leaving a distinctive disappearing track which can be reconstructed.

The exclusion limits on the simplified higgsino and wino/bino models set by previous searches by ATLAS and by the experiments at the Large Electron Positron Collider (LEP) in the compressed mass parameter space are shown in Figure 5.1. Models with mass splittings greater than ~ 1.5 GeV are excluded by a search performed with an

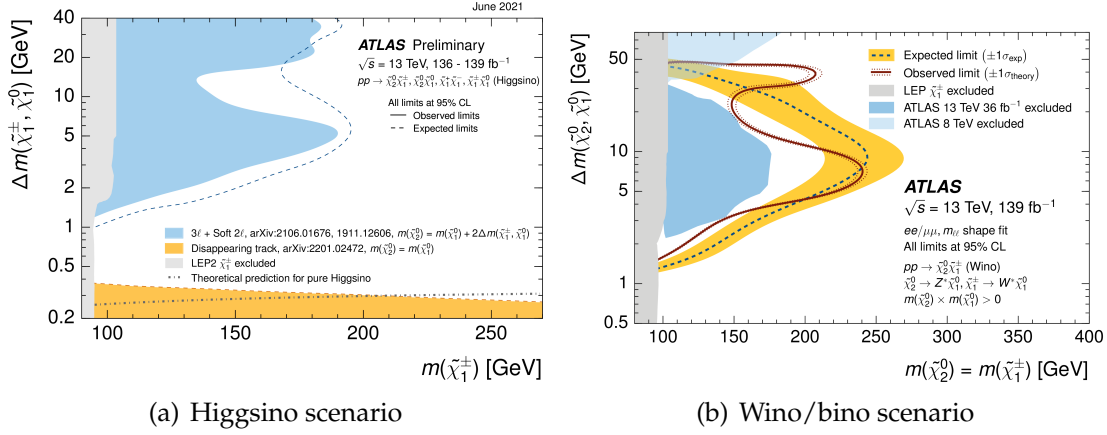


Figure 5.1: Current ATLAS limits on (a) higgsino (b) wino/bino electroweak SUSY models with compressed mass spectra. (a) taken from [31], (b) from [156]. In the left plot, solid areas indicate observed limits and dashed lines expected limits.

ISR jet and two soft leptons in the final state [156], with the sensitivity dropping as the mass splitting decreases due to the lower lepton reconstruction efficiency¹. The disappearing track search is able to exclude models with mass splittings below ~ 0.4 GeV (note that the limit set by the disappearing track search for the wino/bino scenario is not shown but should be similar to that for the higgsino scenario). The search is not sensitive to models with larger mass splittings as in such models the $\tilde{\chi}_1^\pm$ decay before reaching the ID material. For intermediate splittings in the higgsino scenario there is a notable gap where at present the best limits are from the experiments at LEP [157]. This gap is known as the “higgsino gap”. An analogous gap exists for the wino/bino scenario, although the aforementioned VBF result from CMS, discussed in more detail below, is able to exclude models with $\tilde{\chi}_1^0$ masses up to 110 GeV for a mass splitting $\Delta m(\tilde{\chi}_2^0, \tilde{\chi}_1^0)$ of 1 GeV and thus surpasses the LEP limit in the wino/bino gap region.

Broadly, there are two complimentary approaches to obtaining sensitivity to the models in these gap regions, both of which still make use of the boosted topologies. The first is to develop special reconstruction techniques for the soft SM decay products, for instance dropping the standard lepton reconstruction and ID requirements and instead developing dedicated working points or alternatively using tracks as proxies for the final state leptons. There may also be special types of tracks to reconstruct, for

1 It is worth noting that searches with a final state of one soft lepton, E_T^{miss} and ISR jet have not been performed by either ATLAS or CMS and have the potential to be sensitive to slightly smaller mass splittings than the soft two lepton searches. However, compared to the two soft lepton final state such a search would suffer from a large W +jets background. Consequently, such a search would likely only extend the sensitivity reach to slightly smaller mass splittings.

instance a soft displaced track with a characteristic kink originating from a charged pion produced in a $\tilde{\chi}_1^\pm$ decay as outlined in [158].

The second approach is to not attempt to reconstruct these soft decay products and instead just use a zero-lepton final state including a lepton veto. Such a veto has a large acceptance for these ultra-compressed models as the leptons fail the reconstruction/ID requirements, even in the boosted topologies. Compared to the first approach, the zero-lepton final state approach will have a poorer signal to background ratio, primarily due to large backgrounds from W +jets (in which the W -boson either decays hadronically or the lepton from a leptonic W -boson decay also evades reconstruction) and from Z +jets where the Z -boson decays to a pair of neutrinos. However, these backgrounds can be estimated very precisely using simulation (which may not model the special track signatures particularly well) and dedicated Control Regions (CRs) which can also be used to tightly constrain and cancel systematic uncertainties. Consequently, the overall uncertainties on the background estimates can be very low, yielding sensitivity to signals that are relatively small compared to the background.

Furthermore, for the mass splittings of interest the zero-lepton approaches should be relatively agnostic to the precise mass splitting being considered, meaning sensitivity can be obtained right down to mass splittings of zero. This is in contrast to the special track signature approach; depending on the exact life-time the nature of the signal tracks can change dramatically. This is a compelling argument when considering the global picture of searches for SUSY; these compressed models represent a small fraction of the global SUSY parameter space, hence, given available resources, it is debatable whether it is worth performing multiple dedicated searches at multiple compressed mass splittings, given that the zero-lepton final state approach requires less specialized effort.² Additionally, the zero-lepton final state approaches are often sensitive to other BSM scenarios, such as the Higgs boson decaying to a pair of invisible particles. At the very least, it is certainly important to check the sensitivity of the zero-lepton final state approach to quantify any potential gains of the special track reconstruction approaches.

The two analyses presented in this chapter are both of the zero-lepton final state variety. The first in Section 5.2 examines the most natural extension to the previous searches which is to use the boosted ISR jet topology with large E_T^{miss} , here called the

² Though a counter-argument could be made that, when viewed in lifetime parameter space, these UC-models cover a much larger fraction of phase space than non-compressed searches.

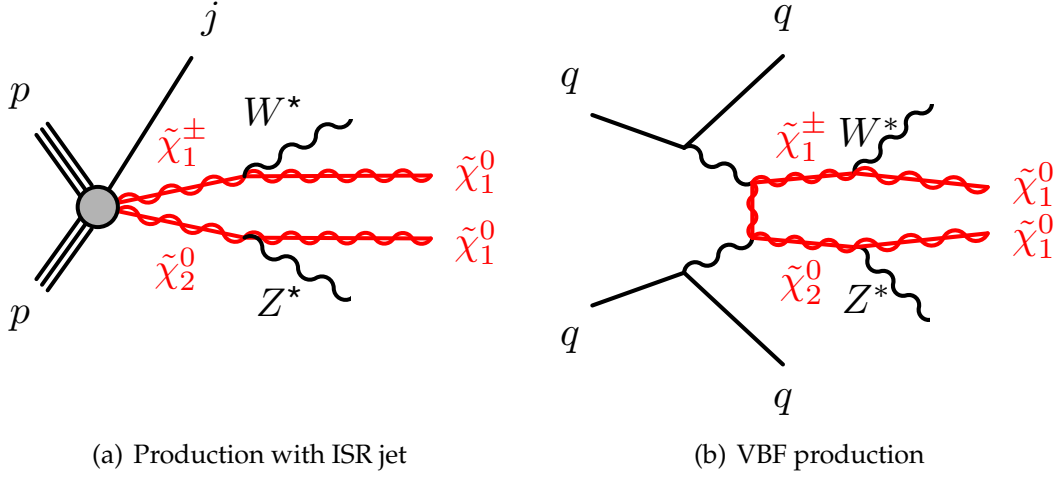


Figure 5.2: Example Feynman diagrams of $\tilde{\chi}_2^0 - \tilde{\chi}_1^\pm$ production (a) in association with an ISR jet, (b) via Vector Boson Fusion (VBF). In the left diagram, the grey circle represents multiple diagrams in which the $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ are produced from electroweak bosons produced by the proton-proton collision. Due to the small mass splittings between the $\tilde{\chi}_2^0/\tilde{\chi}_1^\pm$ and $\tilde{\chi}_1^0$ s, the W- and Z-bosons are produced off-shell and their decay products (not shown) are very soft such that they evade conventional reconstruction.

mono-jet channel. An example Feynman diagram of $\tilde{\chi}_2^0 - \tilde{\chi}_1^\pm$ production with an ISR jet is shown in Figure 5.2(a). As mentioned at the beginning of this chapter, ATLAS has already performed a search for new physics in this topology using the full Run 2 dataset [152], which targeted a range of BSM scenarios and observed no significant excesses compared to the SM predictions. As the compressed SUSY models considered here provide no unique signature compared to these other BSM scenarios, there are unlikely to be significant gains from re-optimizing the search for the SUSY signals. Hence, the results of this search are simply re-interpreted for the UC-models to see what limit can be set on them.

The second analysis presented in Section 5.3 is a sensitivity study of a search targeting the Vector Boson Fusion (VBF) production mode, in which two initial state partons radiate W-bosons or Z-bosons which ‘fuse’ to produce the electroweakinos. Figure 5.2(b) shows an example VBF Feynman diagram of $\tilde{\chi}_2^0 - \tilde{\chi}_1^\pm$ production. The experimental signature of the VBF processes is characterised by two jets with a large separation in pseudorapidity and, for the SUSY signal, large E_T^{miss} from the $\tilde{\chi}_1^0$. Compared to the mono-jet channel, the production cross-sections for the VBF channel are about an order of magnitude smaller. However, the VBF topology potentially offers

additional background discrimination power via the two VBF jets, hence it may have competitive sensitivity with the mono-jet channel. The VBF channel is motivated by a CMS result [153], which sets limits on the WB scenario that surpass those set by the LEP experiments. The sensitivity of this result is particularly striking given that it only uses about a quarter of the available Run 2 dataset. Like the Mono-jet scenario, ATLAS has already performed a search for new physics in this final state, focused on a Higgs boson decaying to invisible particles [159], here on referred to as the Higgs→invisible analysis, again with no significant excesses being found. However, unlike the mono-jet scenario, differences between the kinematics of Higgs→invisible signature and the UC-models mean that significant sensitivity can be gained from a re-optimization of the event selection to be more like that of the CMS analysis, as was demonstrated by a collaborator³. This serves as the motivation for the dedicated sensitivity study presented here.

³ In particular, it was demonstrated by members of the Higgs→invisible analysis team that a factor of two gain in the upper limit on the signal strength could be obtained by altering the signal region selections, as can be seen in reference [160] (accessible only to members of the ATLAS collaboration).

5.2 Re-interpretation of the ATLAS Mono-jet search

This section describes the re-interpretation of the ATLAS mono-jet analysis and is structured as follows; first a brief overview of important aspects of this search is given in Section 5.2.1 (more details can be found in reference [152]). This is then followed by details of the event generation of the UC-models and the re-interpretation setup. Results of the re-interpretation are then presented in Section 5.2.3.

5.2.1 Overview of ATLAS Mono-jet search

The ATLAS mono-jet search was performed using the full Run 2 dataset and, as mentioned in the introduction to this chapter, saw no significant excesses compared to the SM predictions. The lack of excesses was used to set limits on various signal models, including WIMPs produced via a generic axial vector and direct squark production. Concerning the event selection, the E_T^{miss} trigger is used to select events in the Signal Region (SR) with events being further required to have $E_T^{\text{miss}} > 200$ GeV. Events with reconstructed electrons, muons, tau-leptons or photons are vetoed. Additional requirements are placed on the p_T of the ISR jet, the number of additional jets and the azimuthal separation between the jets and the E_T^{miss} . A 13 bin shape fit in E_T^{miss} is used in the SR such that the analysis is sensitive to a range of signal models.

The dominant $Z(\rightarrow \nu\nu)$ +jets and W +jets backgrounds are estimated using simulation as are all other backgrounds except for the multi-jet background for which a data-driven method known as jet smearing, described in [161] is used. SHERPA [2.2.1] is used as the event generator for the Z +jets and W +jets processes, with a reweighting procedure used to account for higher-order QCD and electroweak corrections. The $Z(\rightarrow \nu\nu)$ +jets and $W(\rightarrow \ell\nu)$ +jets backgrounds are normalised in four dedicated Control Regions (CRs), enriched in $Z \rightarrow ee$, $Z \rightarrow \mu\mu$, $W \rightarrow e\nu$, $W \rightarrow \mu\nu$ respectively. For the electron enriched CRs, the single electron trigger was used while the E_T^{miss} trigger was used for the muon enriched regions. Backgrounds from $t\bar{t}$ and single-top events were also normalised in a dedicated CR, requiring a b -jet and also using the single electron triggers. Within these CRs, a p_T^{recoil} variable is used as a proxy for the E_T^{miss} . This variable is reconstructed in the same way as the E_T^{miss} but ignores leptons, hence the leptons are treated as being invisible. The same selection criteria on the jets in the SR is required in the CRs. Additional selection criteria are also required in for the CRs

to improve their purities in the respective target backgrounds. The CRs are binned in p_T^{recoil} using the same binning as is used for E_T^{miss} in the SR.

A full set of experimental and theoretical uncertainties are evaluated for the background predictions. HistFitter [84] is used for performing the statistical fits. In the fit, a single normalisation factor is used to control both the Z+jets and W+jets backgrounds, while separate normalisation factors are used for each of the $t\bar{t}$ and single- t backgrounds. The likelihood model was validated via a background-only fit to the CRs prior to unblinding the SR. From this fit, the Z+jets and W+jets, $t\bar{t}$ and single- t normalisation factors were determined to be 1.13 ± 0.01 , 0.9 ± 0.1 and 1.6 ± 0.4 respectively. An important aspect of the mono-jet analysis is the very tight constraining of the systematic uncertainties; the total background uncertainty in the SR varies from 1.5-4.2% between the lowest and highest E_T^{miss} bins from the CR-only fit (i.e. not including further constraints from the SR bins). While no overall significant excess is seen in the SR, a ~ 2 sigma excess is seen in one of the higher E_T^{miss} bins which tend to have greater sensitivity to the signal models. Consequently, most of the observed limits on signal models are weaker than their expected counterparts.

5.2.2 Signal Generation and Re-interpretation setup

Signal samples for the higgsino and wino/bino scenarios are generated at leading order using MadGraph[2.9.5] [109], with MadSpin [162] used to decay the electroweakinos and PYTHIA [8.306] [110] used to model the parton shower and hadronisation. In both models, the $\tilde{\chi}_2^0$ decay with 100% branching fraction to a $\tilde{\chi}_1^0$ and a Z-boson and the $\tilde{\chi}_1^\pm$ decay with 100% branching fraction to a $\tilde{\chi}_1^0$ and a W-boson. The signals are generated with an ISR requirement to have at least one jet from the hard scatter with $p_T > 100$ GeV and are required to pass a E_T^{miss} filter. Cross-sections used for the samples are computed from accurate calculations of the inclusive cross-section which is multiplied by the ratio of the cross-section with the ISR requirement to the inclusive cross-section taken from MadGraph, (multiplied by the filter efficiency). The inclusive cross-sections are calculated in the same way as for the direct stau samples; using RESUMMINO v2.0.1, which is next to leading order (NLO) in the strong coupling constant, and adds the resummation of soft gluon emission at next-to-leading-logarithm (NLL) accuracy [112, 113]. Explicitly, the cross-sections used for the

re-interpretations is given by

$$\sigma_{\text{Re-Interp}} = \frac{\sigma_{\text{ISR}}^{\text{MadGraph}}}{\sigma_{\text{Incl}}^{\text{MadGraph}}} \times \sigma_{\text{Incl}}^{\text{Resummino}} \times \epsilon_{\text{filter}}, \quad (5.1)$$

where $\sigma_{\text{ISR}}^{\text{MadGraph}}$ is the cross-section from MadGraph with the ISR requirement, $\sigma_{\text{Incl}}^{\text{MadGraph}}$ is the inclusive cross-section from MadGraph, $\sigma_{\text{Incl}}^{\text{Resummino}}$ is the inclusive cross-section from RESUMMINO and ϵ_{filter} is the MC filter efficiency. For the higgsino scenario, six production modes are considered; $\tilde{\chi}_2^0 \tilde{\chi}_1^+$, $\tilde{\chi}_2^0 \tilde{\chi}_1^-$, $\tilde{\chi}_1^0 \tilde{\chi}_1^+$, $\tilde{\chi}_1^0 \tilde{\chi}_1^-$, $\tilde{\chi}_1^+ \tilde{\chi}_1^-$ and $\tilde{\chi}_2^0 \tilde{\chi}_1^0$; while for the wino/bino scenario only three production modes are permitted; $\tilde{\chi}_2^0 \tilde{\chi}_1^+$, $\tilde{\chi}_2^0 \tilde{\chi}_1^-$ and $\tilde{\chi}_1^+ \tilde{\chi}_1^-$. The production modes are generated separately but are combined to form a single signal for a given mass spectrum. Signals are generated for multiple mass values ranging from 85 GeV to 200 GeV for mass splittings between 0.2 GeV and 10 GeV to fully assess the sensitivity of the mono-jet analysis.

Inclusive cross-sections and the ratio of the cross-section with the ISR requirement to the inclusive cross-section from MadGraph for the different production process for some benchmark mass points are shown in Table 5.1. Summed over all of the production processes, the typical inclusive cross-section is around 10 pb and the ratio of the ISR to inclusive cross-section is around 10%. For illustration of the mono-jet kinematics, distributions of the $E_{\text{T}}^{\text{miss}}$, leading jet p_{T} and the difference in ϕ between the leading jet and $E_{\text{T}}^{\text{miss}}$, $\Delta\phi(j_1, E_{\text{T}}^{\text{miss}})$, for the different higgsino and wino/bino sub-processes with $m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (100, 99)$ are shown in Figure 5.3. These distributions are constructed using an approximation of the ATLAS detector response, based on smearing the truth MC information using the SimpleAnalysis framework [163]. For the re-interpretation, the fast simulation described in Chapter 2 is used for modelling the ATLAS response for the signal samples (full simulation was used for the background processes).

The re-interpretation of the mono-jet analysis is performed using the RECAST framework [164] in which the analysis steps are repeated for the new signal samples, including statistical fits to determine the upper limit on the signal strength. The RECAST of the mono-jet search was developed by members of the original mono-jet analysis team and was adapted by a collaborator to facilitate combining multiple samples into a single signal as is done here. The fits include all the systematic uncertainties on the background predictions that were featured in the original analysis. Experimental uncertainties are also evaluated for the signal samples. At present, theoretical

Chapter 5: Searches for Ultra-compressed Charginos and Neutralinos

Process	Sub-process	Inclusive cross-section [pb]	MadGraph ISR/Inclusive ratio
WB(150, 149)	$\tilde{\chi}_2^0 \tilde{\chi}_1^+$	3.27 ± 0.13	0.11
	$\tilde{\chi}_2^0 \tilde{\chi}_1^-$	1.91 ± 0.13	0.11
	$\tilde{\chi}_1^+ \tilde{\chi}_1^-$	2.61 ± 0.14	0.12
	Total	7.79 ± 0.23	0.11
WB(100, 99)	$\tilde{\chi}_2^0 \tilde{\chi}_1^+$	13.9 ± 0.5	0.082
	$\tilde{\chi}_2^0 \tilde{\chi}_1^-$	8.8 ± 0.5	0.081
	$\tilde{\chi}_1^+ \tilde{\chi}_1^-$	11.6 ± 0.5	0.078
	Total	34.3 ± 0.9	0.080
HH(100, 99)	$\tilde{\chi}_2^0 \tilde{\chi}_1^+$	3.47 ± 0.13	0.083
	$\tilde{\chi}_2^0 \tilde{\chi}_1^-$	2.17 ± 0.13	0.083
	$\tilde{\chi}_1^0 \tilde{\chi}_1^+$	3.53 ± 0.13	0.084
	$\tilde{\chi}_1^0 \tilde{\chi}_1^-$	2.22 ± 0.13	0.083
	$\tilde{\chi}_1^+ \tilde{\chi}_1^-$	3.03 ± 0.13	0.090
	$\tilde{\chi}_2^0 \tilde{\chi}_1^0$	3.35 ± 0.13	0.079
	Total	17.76 ± 0.32	0.084

Table 5.1: Cross-sections for example signal points used for evaluating the Mono-jet sensitivity. The inclusive cross-sections are calculated with RESUMMINO. The last column indicates the ratio of the cross-section from MadGraph with the ISR requirement detailed in text to the inclusive cross-section from MadGraph.

uncertainties on the signal models have not yet been added, hence the limits presented here should be considered as optimistic, though this is discussed further in the next section.

Chapter 5: Searches for Ultra-compressed Charginos and Neutralinos

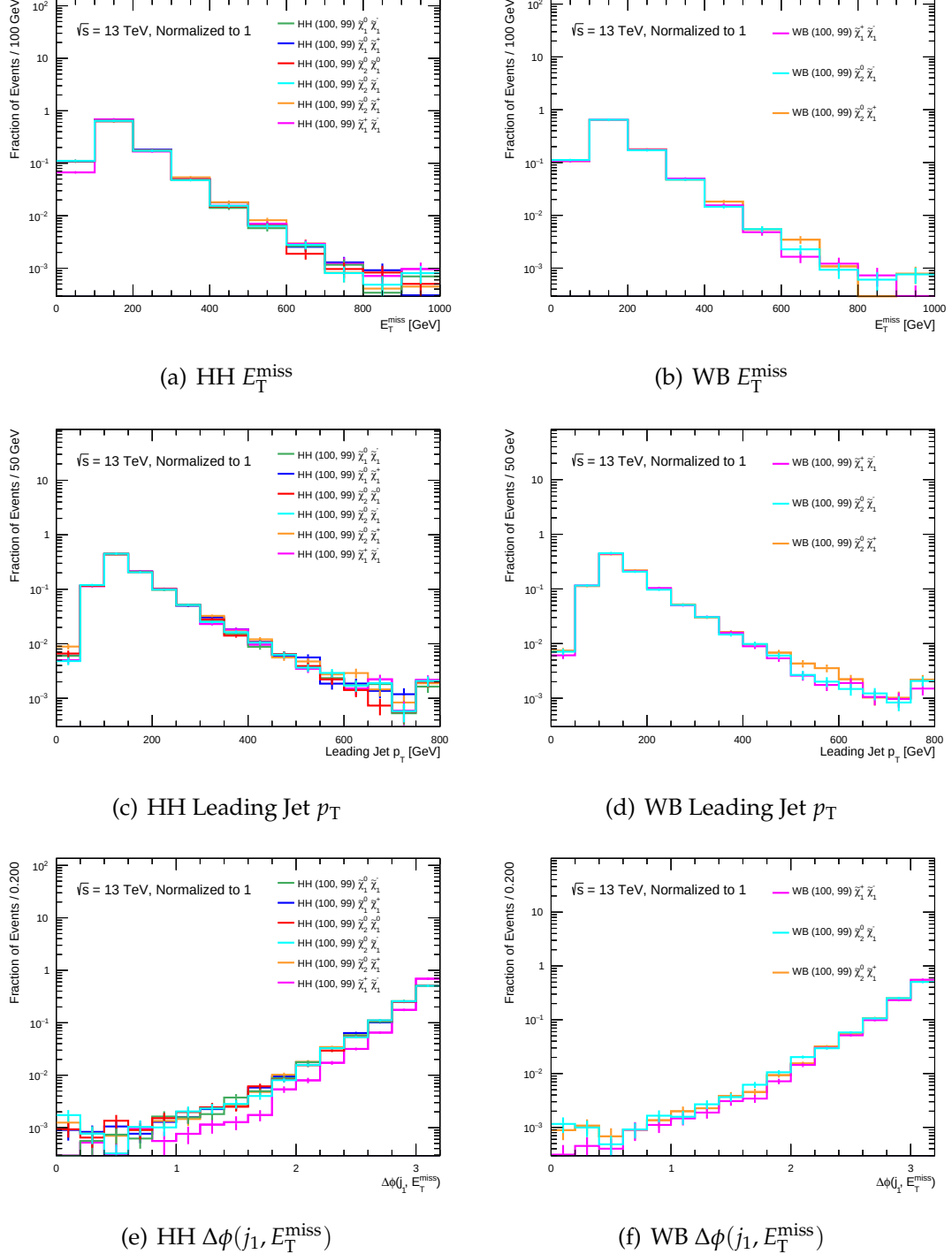


Figure 5.3: Distributions of key variables illustrating the kinematics for the Mono-jet analysis. Overflow events are included in the first and last bins. The vertical lines indicate the statistical uncertainties.

5.2.3 Results

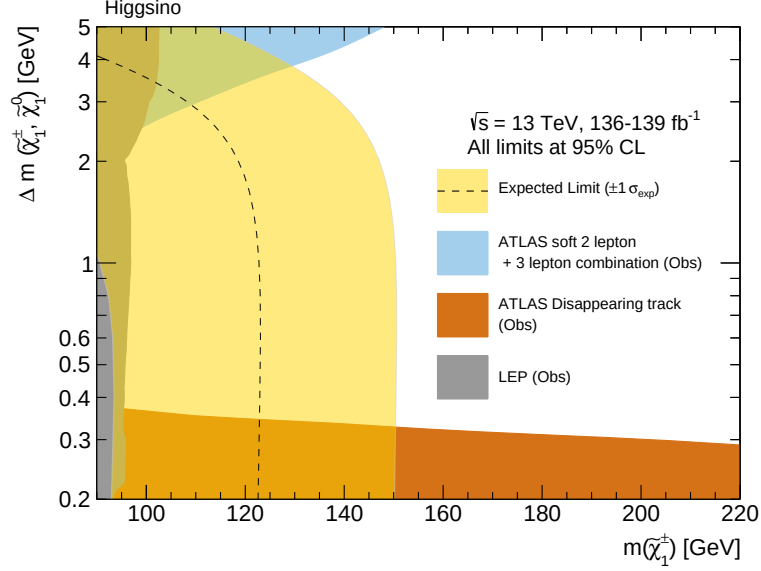
The 95% CLs exclusion contours for the higgsino and wino/bino scenarios are shown in Figure 5.4, overlaid with the results from previous searches by ATLAS and the LEP experiments discussed in the previous section. The upper limits on the signal strength for a few benchmark points are also tabulated in Table 5.2 for reference. From these figures it can be seen that the mono-jet analysis has significant interesting sensitivity to the UC-models. In particular, the expected limit on the higgsino models extends to 120 GeV, surpassing the LEP limit and hence can be considered to close the higgsino gap. However, the observed limit falls below that of the LEP experiments and hence is not shown. The expected uncertainty band also has a relatively large width of 60 GeV. Note also that for the higgsino contour a couple of points at low mass were added based on extrapolation of the existing points in order to be able to draw the lower edge of the uncertainty band.

The sensitivity to the wino/bino scenario is stronger, due to the larger production cross-sections, and extends to 150 GeV and 200 GeV for the observed and expected limits respectively. Thus, this interpretation is able to exclude, at 95% CL, wino/bino models previously not excluded by any previous analysis (with the caveat about theoretical uncertainties on the signal models). For both interpretations, the sensitivity starts to decline as the mass splitting increases as leptons are able to be reconstructed. The expected contours intersect those from the soft two lepton search at mass splittings around 3 GeV.

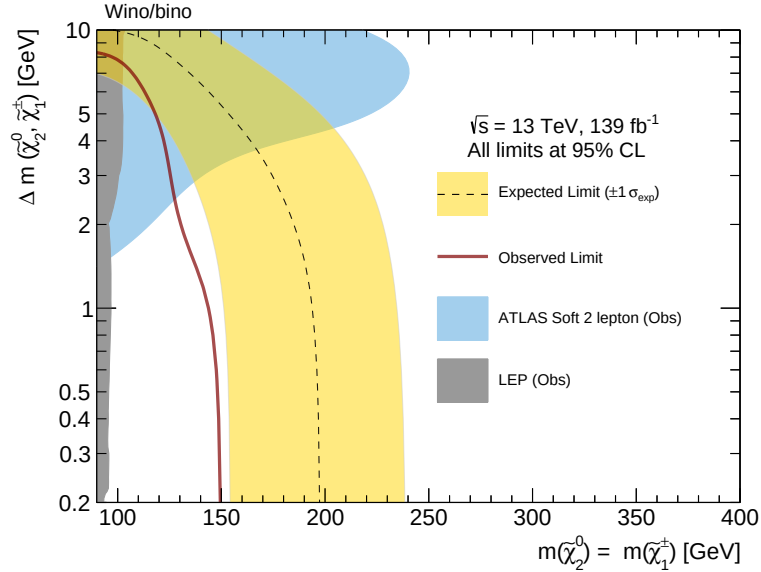
Signal	95% CL Upper Limit	
	Expected	Observed
WB (150, 149)	$0.70^{+0.27}_{-0.20}$	0.89
WB (100, 99)	$0.39^{+0.15}_{-0.11}$	0.47
HH (100, 99)	$0.80^{+0.31}_{-0.22}$	1.1

Table 5.2: 95% CL upper limits from the re-interpretation of the mono-jet analysis for example signal points.

For comparison with the VBF channel in the next section, the E_T^{miss} distribution of the total SM background and benchmark signal points is shown in Figure 5.5, together with the signal to background (S/B) ratio in each bin. For the wino/bino signal with $m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (100, 99)$, the S/B falls just short of 20% in the final bin, while for the higgsino scenario at the same mass the S/B is around 7%. Yields for the total



(a) Higgsino scenario



(b) Wino/bino scenario

Figure 5.4: Exclusion contours showing expected and observed 95% CL limits from the re-interpretation of the mono-jet analysis for the higgsino and wino/bino scenarios, overlaid with observed exclusion contours from other analyses. In particular; the limits from a combination of the ATLAS soft 2-lepton [156] and 3-lepton [165] search for the higgsino scenario, and of just the ATLAS soft 2-lepton search for the wino/bino scenario are shown in blue; the higgsino limit from the ATLAS disappearing track search [154] in orange and the observed limit from the experiments at LEP [166] in grey. The observed mono-jet limit for the higgsino scenario falls below the LEP limit and so is not shown.

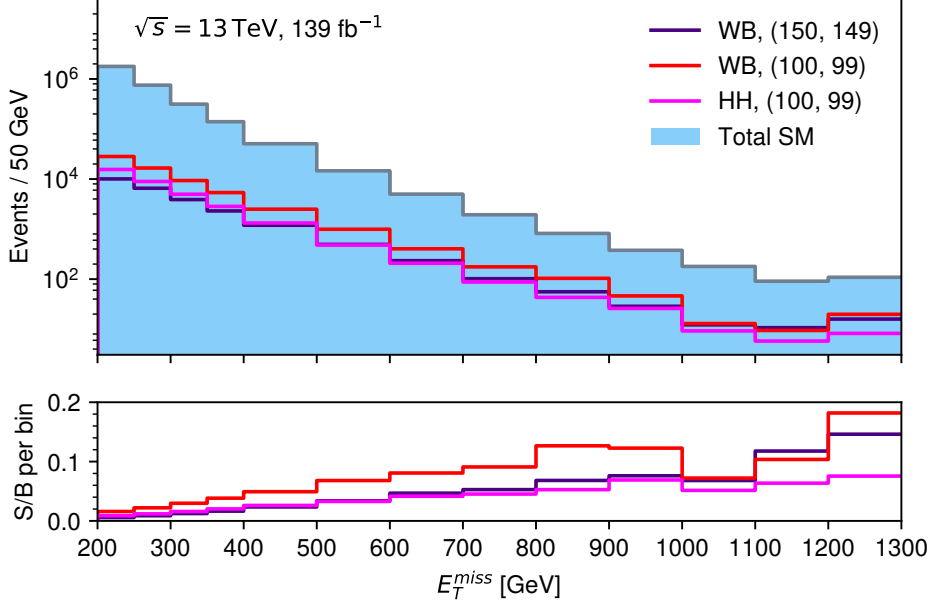


Figure 5.5: Distribution of the E_T^{miss} in the mono-jet SR for the total SM background and benchmark signal points. Overflow events are included in the last bin.

background and these signal processes in the mono-jet SR, and in the last three SR bins are detailed in Tables 5.3 and 5.4 respectively. From these yields, it can be determined that the total S/B in the mono-jet SR is less than 2%, and that $Z(\rightarrow \nu\nu)$ +jets is the dominant background in the high E_T^{miss} bins, comprising around 70% of the total background.

Additionally, from these yields it can also be seen that the uncertainty on the signal yields in the final SR bins is around 20%. This uncertainty is dominated by the MC statistical uncertainty on the signal samples. Tests performed with higher statistic samples shows a negligible improvement in the limit, illustrating that the signal MC statistical uncertainty does not have a big impact on the sensitivity. For this reason, the inclusion of signal theoretical uncertainties, which are expected to be significantly less than 20% in the high SR bins, are not expected to sufficiently degrade the limits shown here, but nonetheless should be evaluated in future work to ensure a fair comparison to the limits from other channels.

Process	SR Mono-jet
Z+jets	1,800,000
W+jets	1,100,000
VBF Z+jets	28,000
VBF W+jets	19,000
Top	120,000
Di-boson	46,000
Multi-jet + NCB	30,000
SM Total	3,100,000
WB (WB, 150)	27,000
WB (WB, 100)	68,000
HH (HH, 100)	37,000

Table 5.3: Total yields, at 139 fb^{-1} , in the mono-jet SR for the SM background and example signal processes. NCB stands for non-collisional background which is due to debris from other proton-proton collisions and non-collision components.

Process	Bin 11 E_T^{miss} [1000, 1100) GeV	Bin 12 E_T^{miss} [1100, 1200) GeV	Bin 13 $E_T^{\text{miss}} > 1200$ GeV
$Z(\rightarrow \nu\nu)+\text{jets}$	251 ± 10	126 ± 6	144 ± 8
$Z(\rightarrow ee)+\text{jets}$	-	-	-
$Z(\rightarrow \mu\mu)+\text{jets}$	0.200 ± 0.010	0.100 ± 0.010	1.19 ± 0.04
$Z(\rightarrow \tau\tau)+\text{jets}$	0.100 ± 0.010	2.10 ± 0.08	0.50 ± 0.04
VBF $Z+\text{jets}$	24 ± 6	13.7 ± 3.3	20 ± 5
$W(\rightarrow e\nu)+\text{jets}$	6.7 ± 0.5	3.01 ± 0.24	2.59 ± 0.22
$W(\rightarrow \mu\nu)+\text{jets}$	16.2 ± 0.7	7.8 ± 0.4	13.4 ± 0.7
$W(\rightarrow \tau\nu)+\text{jets}$	33.7 ± 1.5	15.5 ± 0.8	15.9 ± 1.0
VBF $W+\text{jets}$	8.4 ± 2.8	5.0 ± 1.9	8.1 ± 3.3
Di-boson	17 ± 4	8.9 ± 2.4	12 ± 4
Top	1.56 ± 0.23	0.30 ± 0.05	0.100 ± 0.020
SM Total	359 ± 10	182 ± 6	218 ± 9
Observed	413	187	207
WB (WB, 150)	24 ± 6	21 ± 6	32 ± 8
WB (WB, 100)	26 ± 10	19 ± 9	40 ± 14
HH (HH, 100)	18 ± 5	11.6 ± 3.4	16 ± 4

Table 5.4: Yields for the SM backgrounds, data and example signal points in the last three SR bins in the mono-jet analysis. Uncertainties on the background yields are from the CR-only fit.

5.3 Sensitivity Study of Vector Boson Fusion Channel

This section details a sensitivity study of the Vector Boson Fusion (VBF) channel, which as mentioned is motivated primarily by a CMS partial Run 2 search [153] in this topology. The sensitivity study is performed in view of using the full 139 fb^{-1} 13 TeV Run 2 dataset collected by ATLAS. A like-for-like comparison with the CMS search is studied, as are further optimizations of the SR, including the use of a BDT classifier. This section is structured as follows; first an overview of the general analysis strategy is provided in Section 5.3.1; Section 5.3.2 then describes the simulation and data samples used, Section 5.3.3 the event reconstruction, Section 5.3.4 the SR optimization, Section 5.3.5 the background estimation and Section 5.3.6 an evaluation of the sensitivity, including the comparison with the CMS result.

5.3.1 Overview of Analysis Strategy

The analysis strategy closely follows that of the CMS search and the ATLAS Higgs $\rightarrow$ invisible search [159] in the VBF topology, both of which also share similarities with the mono-jet search discussed in the previous section. In this study, two SRs are defined for testing the sensitivity of this channel to the UC-models; the first, SR-CMS, aims to replicate the selections used in the CMS analysis and the other, SR-BDT, relaxes some of these selections and makes use of a BDT classifier score. For both SRs, the E_T^{miss} trigger is used and events with reconstructed electrons, muons, tau-leptons or b-jets are vetoed. Shape fits are performed in both SRs using a discriminating variable. In this study, the effect of using the invariant mass of the two VBF jets (M_{jj}), the E_T^{miss} and a BDT classifier score for this variable are compared. The sensitivity to the wino/bino and higgsino scenarios are evaluated for similar benchmark mass points as used in the mono-jet re-interpretation for comparison.

As in the mono-jet analysis, the main backgrounds in the SRs are from $Z(\rightarrow \nu\nu)+\text{jets}$ and $W+\text{jets}$ events in which either the W -boson decays hadronically or in leptonic decays the lepton is not reconstructed. The modelling of these backgrounds is crucial for accurately evaluating the sensitivity of the VBF channel. These backgrounds are modelled primarily by MC simulation, specifically SHERPA [2.2.11], to which data-driven correction factors from CRs are applied. Four CRs are defined that, like those in the mono-jet and ATLAS Higgs $\rightarrow$ invisible searches, are enriched in $Z \rightarrow ee$, $Z \rightarrow \mu\mu$, $W \rightarrow e\nu$, $W \rightarrow \mu\nu$ respectively. These CRs mirror the SR selections aside from the

lepton vetoes. The modelling of the discriminating variables used for the SR shape fits are checked in each of these CRs. Although much better agreement is seen than with older simulations of these processes, some disagreement is still visible (albeit the level of disagreement cannot be accurately quantified as full systematic uncertainties are not evaluated). Other minor backgrounds from top and multi-boson events are estimated purely from simulation. A non-negligible contribution to the background from multi-jet events is expected in the SR but is neglected in this sensitivity study. Based on the ATLAS Higgs→invisible and CMS searches, the contribution of this background is expected to be around 5-7% of the total background in the SR and therefore have a non-negligible effect on the overall sensitivity. Its omission means that the study presented here should be taken as an optimistic one.

Experimental and theoretical uncertainties are not fully evaluated for this sensitivity study; instead three configurations of simplified systematic uncertainties are used which aim to emulate the dominant systematic uncertainties in each of the CMS, ATLAS Higgs→invisible and ATLAS mono-jet analyses in order to efficiently obtain a realistic assessment of the sensitivity of the VBF channel. Statistical fits are performed using HistFitter [84] to establish upper limits on the signal strengths. More details of each of these aspects is provided in the following sections.

5.3.2 Simulation and Data Samples

As for samples used for the mono-jet re-interpretation, the SUSY processes are generated using MadGraph[2.9.5] [109], interfaced with MadSpin [162] and PYTHIA [8.245] [110]. The samples generated are required to have the invariant mass of the two partons which radiate the vector bosons be greater than 500 GeV (after the radiation). The separation in pseudorapidity between these two partons is required to be greater than 3.5 to enforce the VBF topology. A filter on E_T^{miss} is also required.

Non-VBF production modes are included in the signal generation. Figure 5.6 shows an example non-VBF Feynman diagram alongside a VBF Feynman diagram for $\tilde{\chi}_2^0\text{-}\tilde{\chi}_1^\pm$ production. The non-VBF diagram shown is also labelled as QCD production due to having two QCD vertices (and two electroweak (EW) vertices), in contrast to the VBF diagram which has four EW vertices. Non-VBF diagrams with four EW vertices are also possible and are also included in the event generation. Together with the VBF diagrams, such production modes are referred to here as EW production. The

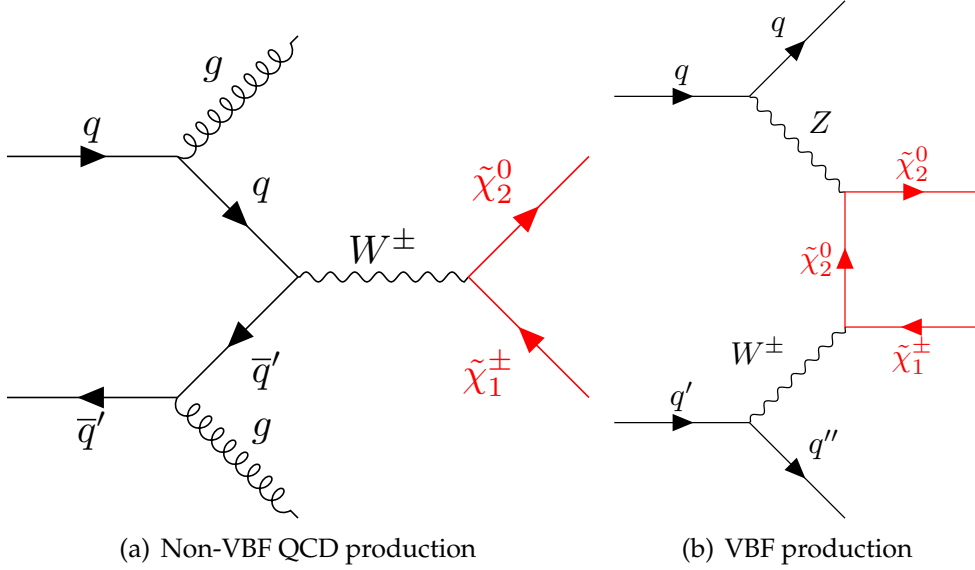


Figure 5.6: Example VBF Feynman diagrams for $\tilde{\chi}_2^0\text{-}\tilde{\chi}_1^\pm$ production via the different types of production mode. Diagrams produced using TikZ Feynman [9].

QCD production modes have larger cross-sections than the EW modes but have softer kinematics, making them less distinguishable from the majority of the SM background. These non-VBF diagrams are included in the signal generation as events resulting from such diagrams may pass the VBF selection criteria and also to ensure that interference effects between the different production modes are accounted for.

Again like in the mono-jet analysis, multiple production processes occur for the wino/bino and higgsino scenarios and are generated separately then combined into a single signal for each mass spectrum. Due to the nature of the VBF production diagrams, additional processes not considered in the mono-jet analysis are possible, for instance $\tilde{\chi}_1^+\text{-}\tilde{\chi}_1^+$ production. The cross-sections used for each of the signal processes are taken directly from MadGraph as, at present, more accurate cross-section calculations for the VBF production modes do not exist. Cross-sections for each of the production processes for a few example mass points are detailed in Table 5.5. Compared to the effective cross-sections for the mono-jet samples, the VBF cross-sections are around an order of magnitude smaller. The same branching fractions used in the mono-jet analysis are also applied here, namely forcing $\tilde{\chi}_2^0$ to decay to a $\tilde{\chi}_1^0$ and Z-boson and forcing $\tilde{\chi}_1^\pm$ to decay to a $\tilde{\chi}_1^0$ and W-boson.

As mentioned, the dominant backgrounds in the VBF channel originate from $Z(\rightarrow \nu\nu)\text{+jets}$ and $W(\rightarrow \ell\nu)\text{+jets}$, collectively known as the $V\text{+jets}$ background. Similar to

Process	Sub-process	Cross-section [fb]
WB(140, 139)	$\tilde{\chi}_2^0 \tilde{\chi}_1^+$	22
	$\tilde{\chi}_2^0 \tilde{\chi}_1^-$	11
	$\tilde{\chi}_2^0 \tilde{\chi}_2^0$	1.6
	$\tilde{\chi}_1^+ \tilde{\chi}_1^-$	19
	$\tilde{\chi}_1^- \tilde{\chi}_1^-$	0.43
	$\tilde{\chi}_1^+ \tilde{\chi}_1^+$	0.47
	Total	55
WB(100, 99)	$\tilde{\chi}_2^0 \tilde{\chi}_1^+$	81
	$\tilde{\chi}_2^0 \tilde{\chi}_1^-$	44
	$\tilde{\chi}_2^0 \tilde{\chi}_2^0$	3.4
	$\tilde{\chi}_1^+ \tilde{\chi}_1^-$	67
	$\tilde{\chi}_1^- \tilde{\chi}_1^-$	0.93
	$\tilde{\chi}_1^+ \tilde{\chi}_1^+$	2.9
	Total	200
HH(100, 99)	$\tilde{\chi}_2^0 \tilde{\chi}_1^+$	22
	$\tilde{\chi}_2^0 \tilde{\chi}_1^-$	12
	$\tilde{\chi}_1^0 \tilde{\chi}_1^+$	23
	$\tilde{\chi}_1^0 \tilde{\chi}_1^-$	12
	$\tilde{\chi}_2^0 \tilde{\chi}_2^0$	0.38
	$\tilde{\chi}_2^0 \tilde{\chi}_1^0$	19
	$\tilde{\chi}_1^+ \tilde{\chi}_1^-$	20
	Total	110

Table 5.5: Cross-sections for example signal processes used for evaluating the sensitivity of the VBF channel. These cross-sections are for the combined QCD+EW production modes. For the higgsino scenario, same-sign chargino processes are also possible but have very small cross-sections and so can be neglected. Errors on the cross-sections are not shown as they have not been accurately computed.

the SUSY signals, these processes have both EW and QCD production modes. For this study, SHERPA [115, 116] [2.2.11] is used for the event generation of all V +jets background processes. In these samples, up to two partons are modelled at next to leading order (NLO) and up to five partons at leading order (LO) using the Comix [117] and OPENLOOPS [118, 119] generators which are merged with the SHERPA Parton shower [120] using the MENLOPS prescription [116]. Contributions from all possible decays of the W - and Z -bosons are included. Electroweak virtual corrections are applied to the QCD processes as described in [167]. As in the V +jets samples used in the direct stau analysis, the NNPDF3.0NNLO [121] PDF set is used. Nominal next-to-next-to-leading order (NNLO) cross-sections for these samples are taken from

reference [168] but, as described in the overview, data driven correction factors are applied to the V +jet backgrounds.

Sub-dominant backgrounds originate from $t\bar{t}$, single- t and multi-boson processes. Collectively, the $t\bar{t}$ and single- t backgrounds are referred to as the top background in the rest of this study. The samples used for these backgrounds are the same as in the direct stau search in the previous chapter (detailed in Section 4.3), namely using SHERPA [2.2.1/2.2.2] for the multi-boson processes and POWHEG-BOX [v2] [131–134] for the top backgrounds. The parton showers, PDF sets and cross-sections used to normalise these backgrounds are also the same as in the direct stau case. Details of the event generators used for the signal and background processes are summarised in Table 5.6. Other smaller backgrounds, such as from Higgs-boson processes, are neglected in this sensitivity study, as their contribution to the total background is expected to be very small.

Process	ME Generator	Parton Shower	PDF	Tune
VBF SUSY signal	MadGraph 2.9.5	PYTHIA 8	NNPDF2.3LO	A14
Z/W + jets	SHERPA 2.2.11	SHERPA	NNPDF3.0NNLO	SHERPA default
VV, VVV	SHERPA 2.2.1/2.2.2	SHERPA	NNPDF3.0NNLO	SHERPA default
$t\bar{t}$ and single top	POWHEG-BOX	PYTHIA 8	NNPDF2.3LO	A14

Table 5.6: List of MC generators used for modelling the signal and background processes for the VBF sensitivity study. The second column denotes the generator used for modelling the hard scatter at the matrix element (ME) level; the third the generator used for modelling the parton shower, the fourth the set of parton distribution functions (PDFs) used and the fifth how the PDFs are tuned to better agree with data.

The dataset collected by ATLAS during Run 2 is used to assess and correct the modelling of the main V +jets backgrounds in the CRs. As in the Direct Stau search presented in Chapter 4, the data quality requirements discussed in Chapter 2, Section 2.2.4 are applied.

5.3.3 Event Reconstruction

The event reconstruction follows the procedures outlined at the end of Chapter 2. Fast simulation is used for the signal samples while full simulation is used for the background processes. Events are required to have at least one primary vertex, which must be consistent with the beam spot and be associated with two charged-particle tracks with $p_T > 500$ MeV. The vertex with the largest $\sum(p_T)^2$ is chosen in events with more than one vertex fulfilling these criteria. Specific details of the Identification (ID) algorithms and working points (WPs) are given below:

- **Jets:** Reconstructed using the particle-flow algorithm using anti- k_t inputs with $R=0.4$. Required to have $p_T > 20$ GeV and $|\eta| < 4.5$. Medium central JVT and loose forward-JVT WPs are used.
- **b -Jets:** DL1r algorithm is used for b -tagging, WP with 85% efficiency is used.
- **Electrons:** Two types of electrons are defined. *Baseline* electrons are used for vetoing electrons in the SR, and are required to have $p_T > 4.5$ GeV, $|\eta| < 2.47$, longitudinal impact parameter $|z_0 \sin \theta| < 0.5$ mm and must pass the loose electron-ID WP. *Signal* electrons are used for defining the V+jets CRs and have additional requirements (to reduce contamination from fakes in these regions); they must pass loose isolation criteria, pass the tight electron-ID WP and the transverse impact parameter must satisfy $|d_0/\sigma(d_0)| < 5$.
- **Muons:** As for electrons, *baseline* muons are defined for vetoing muonic events in the SR and *signal* muons for use in the V+jets CRs. Baseline muons are required to have $p_T > 3$ GeV, $|\eta| < 2.7$, $|z_0 \sin \theta| < 0.5$ mm and must pass the loose muon-ID WP. Signal muons are in addition required to pass loose isolation WP, medium muon-ID WP and have $|\eta| < 2.5$, $|d_0/\sigma(d_0)| < 3$.
- **Tau-leptons:** Hadronically decaying tau-leptons, used for a veto in the SR, are required to have $p_T > 20$ GeV, $|\eta| < 2.5$ and pass the medium WP of the RNN tau-ID algorithm.
- **E_T^{miss} :** Jets used in the E_T^{miss} reconstruction are required to pass the tight WP.

As in the direct stau search, an overlap removal is performed in order to avoid double counting objects. The steps taken are the same as those described at the end of Section 4.4, with the exception of the steps involving tau-leptons.

5.3.4 Event selection and Signal Region Optimization

The selections defining SR-CMS and SR-BDT are given in Table 5.7. The two SRs share many common selections, defining a common baseline selection. As mentioned in the overview, events in the SR are required to pass the E_T^{miss} trigger. To ensure the E_T^{miss} trigger is fully efficient, an offline $E_T^{\text{miss}} > 250$ GeV cut is also required. For compatibility with the VBF topology, events are required to have two ‘VBF tagged’ jets which have $p_T > 30$ GeV, exist in opposite hemispheres (enforced by the $\eta^{j_1} \times \eta^{j_2} < 0$ requirement), and together have an invariant mass $M_{jj} > 600$ GeV. In events in which more than one pair of jets fulfils these criteria, the pair of jets with the largest M_{jj} are selected as the VBF tagged jets. Additionally, the leading VBF jet (the one with the largest p_T) is required to have $p_T > 60$ GeV and the sub-leading jet is required to have $p_T > 40$ GeV. To suppress the multi-jet background, the jet closet in ϕ to the E_T^{miss} is required to have $\Delta\phi(j, E_T^{\text{miss}}) > 0.4$. As leptons from the SUSY signal are too soft to be reconstructed, events with a reconstructed baseline electron, baseline muon or hadronic tau-lepton are vetoed. To suppress the background from events with top quarks, a b -jet veto is also required.

Three further cuts define SR-CMS; an increased p_T threshold on the sub-leading VBF-tagged jet of 60 GeV; a requirement for the separation in η between the two jets, $\Delta\eta_{jj} > 3.8$, and increasing the threshold on the dijet invariant mass to $M_{jj} > 1000$ GeV. For SR-BDT, these selections are instead replaced by a threshold on the BDT classifier score, details of which are given below. The number of expected background events passing these selections sequentially can be found in the appendix, Section A.3.3, along with the yields for example signal points.

BDT Details

To fully explore the potential of the VBF channel, a binary BDT classifier is trained to discriminate the SUSY signal from the background, again using the LightGBM package introduced at the end of Chapter 3 and used for the direct stau models detailed in the previous chapter. Binary cross-entropy is again used as the objective function. The BDT is trained using the wino/bino signal samples with $m(\tilde{\chi}_2^0) = 100$ GeV and $\Delta m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = [0.2, 0.5, 1, 2]$ GeV and the simulated background samples. The signal samples are combined to form a single signal sample with larger statistics. The variation in the kinematics of the jets and E_T^{miss} between the different mass splittings

is negligible. For training, events are required to pass the baseline VBF SR selection shared by SR-CMS and SR-BDT. The raw and weighted MC yields for the signal and background samples passing this selection is shown in Table 5.8. Compared to yields used for training the BDTs in the direct stau search in the previous chapter (detailed in Table 4.4), it can be seen that here a much greater number of raw statistics are available. However, the weighted S/B for the VBF SUSY signal is notably smaller (0.008 compared to 0.15 - 0.03 for the different direct stau signals).

As for the direct stau models, event weights are used in the training, the total weighted signal yield is scaled to match the total weighted background yield and negatively weighted events are not used in the training. Approximately 25% of the QCD V +jets and 10% of the EW V +jets backgrounds have negatively weighted events, hence their omission could have a sizeable impact. For this reason, BDTs were tested including the negatively weighted events. These BDTs were not found to perform significantly better than those in which the negatively weighted events were omitted, hence they are not included in the training of the final models to rule out the possibility of a deliberate bias. For this sensitivity study, correction factors are not applied to the

SR-CMS	SR-BDT
Pass E_T^{miss} trigger $E_T^{\text{miss}} > 250 \text{ GeV}$ Two Non- b -tagged jets with $p_T > 30 \text{ GeV}$ $\eta^{j1} \times \eta^{j2} < 0$ $M_{jj} > 600 \text{ GeV}$ $p_T^{j1} > 60 \text{ GeV}$ $p_T^{j2} > 40 \text{ GeV}$ $\min[\Delta\phi(j, E_T^{\text{miss}})] > 0.4$ $N \text{ baseline electrons} = 0$ $N \text{ baseline muons} = 0$ $N \text{ hadronic tau-leptons} = 0$ $N \text{ } b\text{-jets} = 0$	
$p_T^{j2} > 60 \text{ GeV}$ $\Delta\eta_{jj} > 3.8$ $M_{jj} > 1000 \text{ GeV}$	BDT score > 0.65

Table 5.7: Selection requirements defining SR-CMS and SR-BDT

Sample	Weighted Yield	No. Raw Events
QCD Z+jets	$139,430 \pm 210$	4,354,972
QCD W+jets	$71,840 \pm 250$	1,425,176
EW Z+jets	$8,102 \pm 20$	175,814
EW W+jets	$4,127 \pm 14$	95,771
Multi-boson	$5,496 \pm 33$	177,691
Top	$2,413 \pm 18$	21,987
SM total	$231,410 \pm 340$	6,251,411
Training Signal	$1,939 \pm 6$	286,599

Table 5.8: Weighted yields at 139 fb^{-1} and number of raw events passing the baseline VBF SR selection shared by SR-CMS and SR-BDT. Errors shown on the weighted yields are from MC statistics. For the training signal, the weighted yield is the average from the four wino/bino signal samples with $m(\tilde{\chi}_2^0) = 100$ and $\Delta m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = [0.2, 0.5, 1, 2] \text{ GeV}$, while the raw yield is from the sum of these four samples. Note that the background yields do not include any correction factors on the V+jets backgrounds.

V+jet backgrounds during the classifier training, which presents a potential avenue for further optimization of the models.

In total 23 kinematic variables⁴ are used as input features for the BDT:

- $p_T^{j_1}$: p_T of the leading VBF-tagged jet.
- $p_T^{j_2}$: p_T of the sub-leading VBF-tagged jet.
- E_T^{miss} : missing transverse energy.
- M_{jj} : invariant mass of the two VBF-tagged jets.
- P_T^{jj} : vectorial sum of the p_T of the two VBF-tagged jets.
- H_T^{jj} : scalar sum of the p_T of the VBF-tagged jets.
- $H_T^{jj}/E_T^{\text{miss}}$: ratio of H_T^{jj} to the E_T^{miss} .

⁴ This is a relatively large number of input features, and some of the variables are strongly correlated. Although having a high number of input features should not worsen the performance of a BDT, provided of course the variables and correlations are well modelled, it can complicate the interpretation of feature importance and increase the model complexity without providing a significant performance gain. For use in a full analysis, the BDTs should be (re-)trained using a subset of the input features listed here, the variables retained being those which contribute significantly to the performance.

- $M_{\text{eff}}^{\text{rel}}$: the ‘relative effective mass’, defined as

$$M_{\text{eff}}^{\text{rel}} = \frac{M_{jj}}{\langle M_{jj} \rangle_{\text{bkg}}} + \frac{E_{\text{T}}^{\text{miss}}}{\langle E_{\text{T}}^{\text{miss}} \rangle_{\text{bkg}}}, \quad (5.2)$$

where $\langle x \rangle_{\text{bkg}}$ denotes the weight-averaged value of variable x for the total SM background passing the training preselection. This variable aims to combine the sensitivity of M_{jj} and $E_{\text{T}}^{\text{miss}}$ into a single variable.

- $\Delta\phi_{jj}$: absolute separation in ϕ between the two VBF-tagged jets.
- $\Delta\eta_{jj}$: absolute separation in pseudo-rapidity between the two VBF-tagged jets.
- ΔR_{jj} : separation between the two VBF-tagged jets, equal to $\sqrt{(\Delta\phi_{jj})^2 + (\Delta\eta_{jj})^2}$.
- $\Delta\phi(jj, E_{\text{T}}^{\text{miss}})$: absolute difference in ϕ between the vectorial combination of the two VBF-tagged jets and the $\vec{p}_{\text{T}}^{\text{miss}}$.
- $\Delta\phi(j_1, E_{\text{T}}^{\text{miss}})$: absolute difference in ϕ between the leading VBF-tagged jet and the $\vec{p}_{\text{T}}^{\text{miss}}$.
- $\Delta\phi(j_2, E_{\text{T}}^{\text{miss}})$: absolute difference in ϕ between the sub-leading VBF-tagged jet and the $\vec{p}_{\text{T}}^{\text{miss}}$.
- $\Delta(\Delta\phi(j_1, E_{\text{T}}^{\text{miss}}), \Delta\phi(j_2, E_{\text{T}}^{\text{miss}}))$: absolute difference between $\Delta\phi(j_1, E_{\text{T}}^{\text{miss}})$ and $\Delta\phi(j_2, E_{\text{T}}^{\text{miss}})$.
- $M_{\text{T}}(j_1, E_{\text{T}}^{\text{miss}})$: transverse mass (defined in Chapter 3, Equation 3.12), of the leading VBF-tagged jet and the $E_{\text{T}}^{\text{miss}}$. A definition of transverse mass can be found in Chapter 3, Equation 3.12.
- $M_{\text{T}}(j_2, E_{\text{T}}^{\text{miss}})$: transverse mass of the sub-leading VBF-tagged jet and the $E_{\text{T}}^{\text{miss}}$.
- $M_{\text{T}}(j_1, E_{\text{T}}^{\text{miss}}) + M_{\text{T}}(j_2, E_{\text{T}}^{\text{miss}})$: sum of the transverse masses of the leading and sub-leading VBF-tagged jets with the $E_{\text{T}}^{\text{miss}}$.
- $\Delta\phi(j_Y, E_{\text{T}}^{\text{miss}})$: absolute difference in ϕ between jet-Y and the $\vec{p}_{\text{T}}^{\text{miss}}$, where jet-Y is defined as the jet closet in ϕ to the $\vec{p}_{\text{T}}^{\text{miss}}$.
- $M_{\text{T}}(j_Y, E_{\text{T}}^{\text{miss}})$: transverse mass of jet-Y and the $E_{\text{T}}^{\text{miss}}$.

- $\Delta\phi(j_Z, \vec{p}_T^{\text{miss}})$: absolute difference in ϕ between jet-Z and the E_T^{miss} , where jet-Z is defined as the jet with transverse mass with the E_T^{miss} is closest to the W -boson mass.
- $M_T(j_Z, E_T^{\text{miss}})$: transverse mass of jet-Z and the E_T^{miss} .
- $\min[\Delta\phi(j, E_T^{\text{miss}})]$: minimum absolute difference in ϕ between any jet and the E_T^{miss} .

Comparisons of the shapes of these variables for the total signal and background samples used to train the BDTs can be found in Section A.3.1 of the appendix. The jet-Y and jet-Z variables are aimed at reducing backgrounds from hadronically decaying W -bosons and were suggested by a collaborator. Except for the jet-Y and jet-Z variables, kinematics of non VBF-tagged jets and the total jet multiplicity are not included as input features as only the two VBF jets are modelled at matrix-element level in the signal, whereas up to four jets are modelled at matrix-element level in the backgrounds, which could potentially cause artificial differences between these distributions for the signal and background. The inclusion of additional variables related to the kinematics of non-VBF tagged jets is another potential avenue for improving the sensitivity in the VBF channel.

Unlike the final direct stau models, the binning of the input variables is not specified explicitly and the maximum depth of the BDTs is not restricted to be one. Instead, the `min_sum_hessian_in_leaf` hyperparameter is used to prevent the isolation of regions of phase space with artificially high S/B. This is due to the better balance of signal and background statistics available here meaning that the isolation effect is of less concern. The input features are still binned into histograms but this time the LightGBM default binning construction, in which the `max_bin` hyperparameter is set to 255 and bin boundaries are chosen such that the bins contain an equal number of events (of the total training dataset), is used.

For optimizing the hyperparameters and the implementation of the final models, cross evaluation with a validation set is employed, using five-folds as opposed to three-folds due to the large number of statistics available. Early stopping is again used, with the `early_stopping_rounds` hyperparameter set to 10 and the maximum number of trees set to 1500. The `min_data_in_leaf`, `l2_regularization` and `max_leaves` hyperparameters are set to 50, 1 and 8 respectively to limit over-fitting. A grid search is then performed over the `learning_rate`, `max_depth`, `min_sum_hessian_in_leaf` and `feature_fraction_bynode` hyperparameters in order to optimize performance. The

AUC score, averaged over the five validation sets, is used to select the best hyperparameter configuration. The scan ranges and intervals are shown in Table 5.9, as well as the values these hyperparameters take in the optimal models.

Hyperparameter	Scan Range/Intervals	Value in best models
learning_rate	0.1 - 0.5, in steps of 0.1	0.1
max_depth	[3, 4, 5]	5
min_sum_hessian_in_leaf	[10, 50, 100, 100]	100
feature_fraction_bynode	[0.3, 0.5, 0.7, 1.0]	0.7

Table 5.9: Scan ranges/intervals for the hyperparameters used in the grid search for the VBF BDT, alongside the values these hyperparameters take for the optimal model.

The receiver operating characteristic for the training and evaluation sets for one of the five optimal models is shown in Figure 5.7 to show a general indicator of the performance. Compared to the direct stau models in the previous chapter, the overall AUC score is lower showing that in this case the overall discrimination power is not as strong. AUC scores for the optimum model trained on each fold are listed in Table 5.10 from which it can be seen that the performance across the folds is similar. A comparison of the training and evaluation sets distributions of the BDT score for one of these models is shown in Figure 5.8. A discrepancy between the training and test signal yields is seen at low BDT scores, where the number of raw statistics in the signal is low. Given the large number of events in the evaluation dataset, this disagreement is not thought to be problematic for this sensitivity study. Train/test comparisons for the optimum BDTs trained on the other folds can be found in the appendix, Section A.3.2.

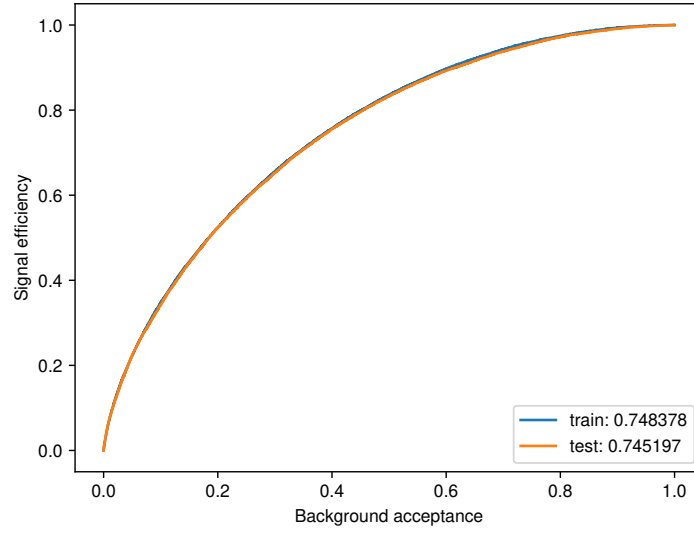


Figure 5.7: Receiver Operating Characteristic (ROC) curves for the training and evaluation (labelled 'test') sets for one of the optimum VBF BDTs.

Fold Split	AUC score	
	Training set	Evaluation set
1	0.748	0.745
2	0.750	0.742
3	0.753	0.739
4	0.751	0.740
5	0.754	0.744

Table 5.10: Area under curve scores for the VBF BDTs across the different train/evaluation splits.

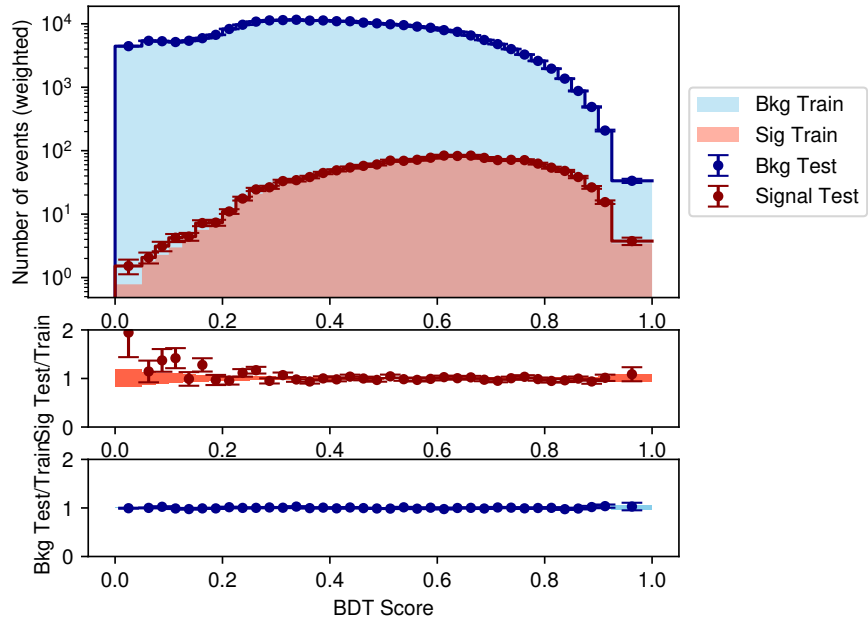


Figure 5.8: Comparison of the BDT score distributions for the training and evaluation (labelled 'test') sets for one of the optimum VBF BDTs.

A comparison of the BDT score distributions of the five different models for their respective evaluation sets are shown in Figure 5.9 for the total SM background and one of the example training signals. Consistent performance is seen for the models, except for at very high scores where the number of events represents a small fraction of the total. As in the direct stau models, the BDT scores are ‘combined’ with the final score simply being referred to as *the* BDT score in the rest of this chapter. The full distribution of this BDT score is shown in Figure 5.10 and shapes of the WB (100,99) signal and the total SM background are shown in Figure 5.11. From Figure 5.10 it can be seen that the contribution of the EW V +jet backgrounds becomes greater at higher BDT scores while Figure 5.11 better highlights the discrimination power of the BDT.

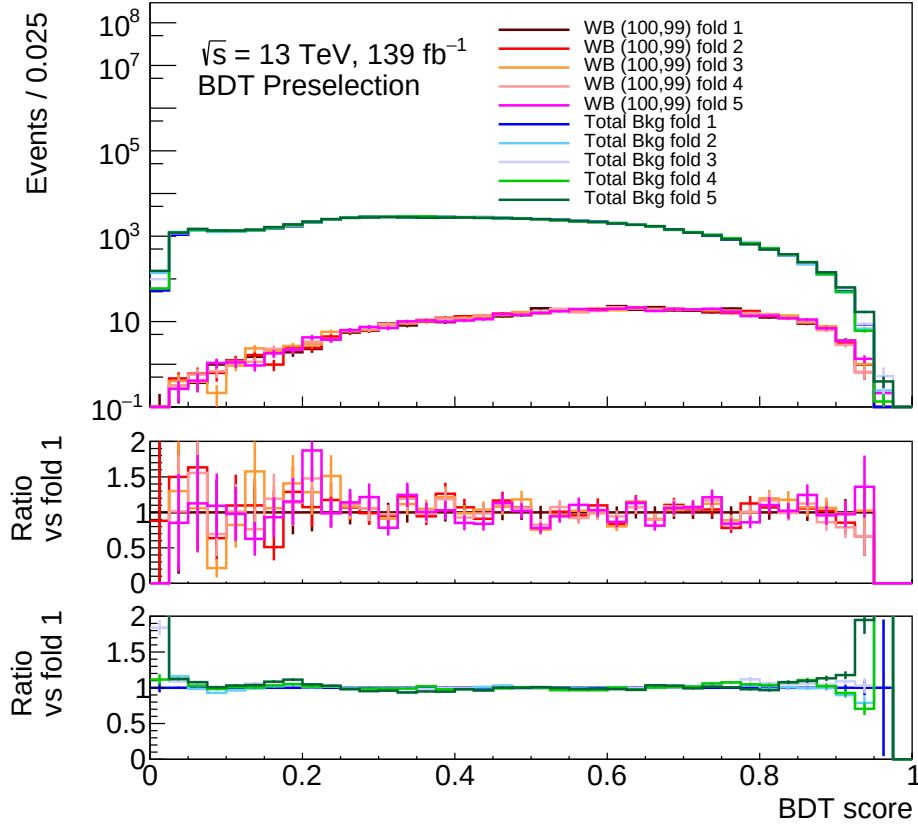


Figure 5.9: Comparison of the evaluation set BDT score distributions for the different fold splittings, for an example signal and the total SM background used in training the models.

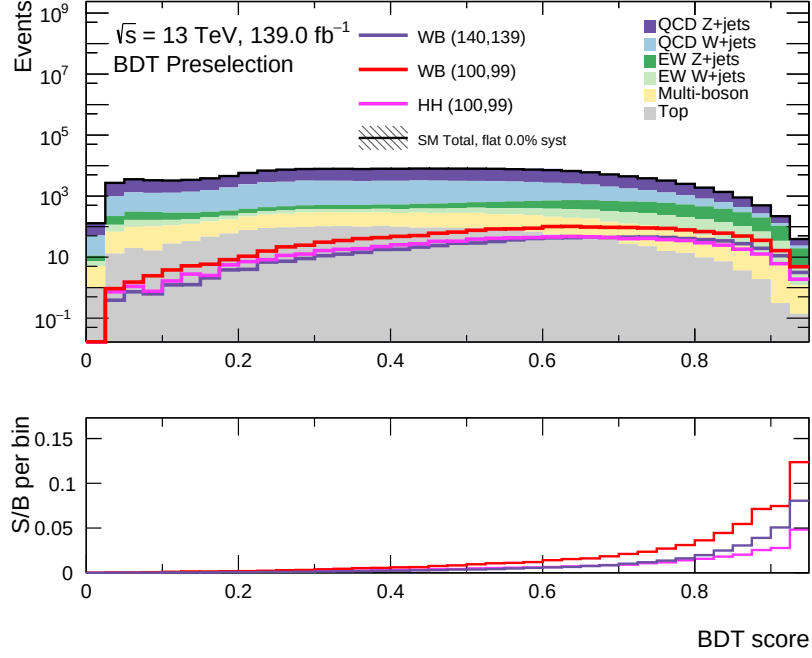


Figure 5.10: Distribution of the final BDT score combined across the different folds for the total SM background and benchmark signal points, showing the S/B per bin. No correction factors to the V +jets backgrounds are applied in this plot.

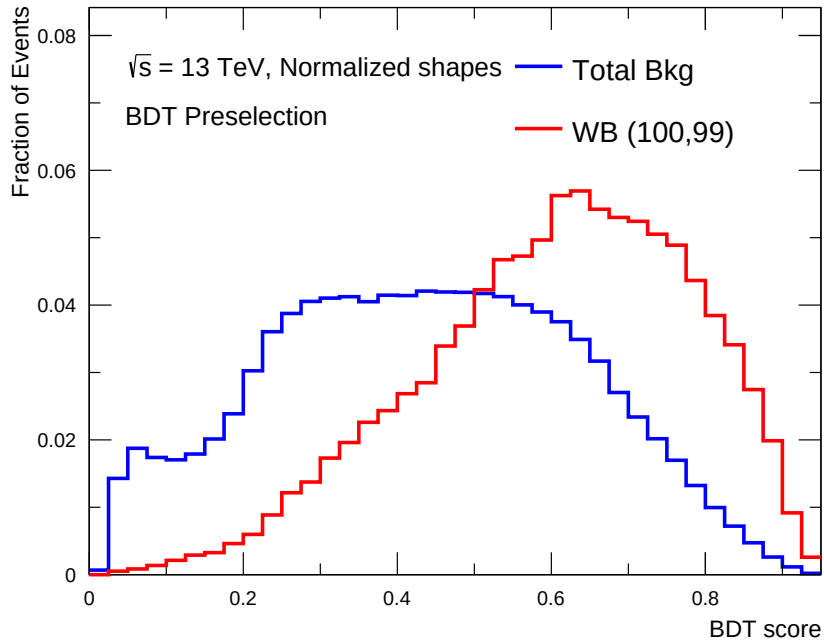


Figure 5.11: Comparison of the shape of the BDT score distribution for the WB (100,99) signal and the total SM background.

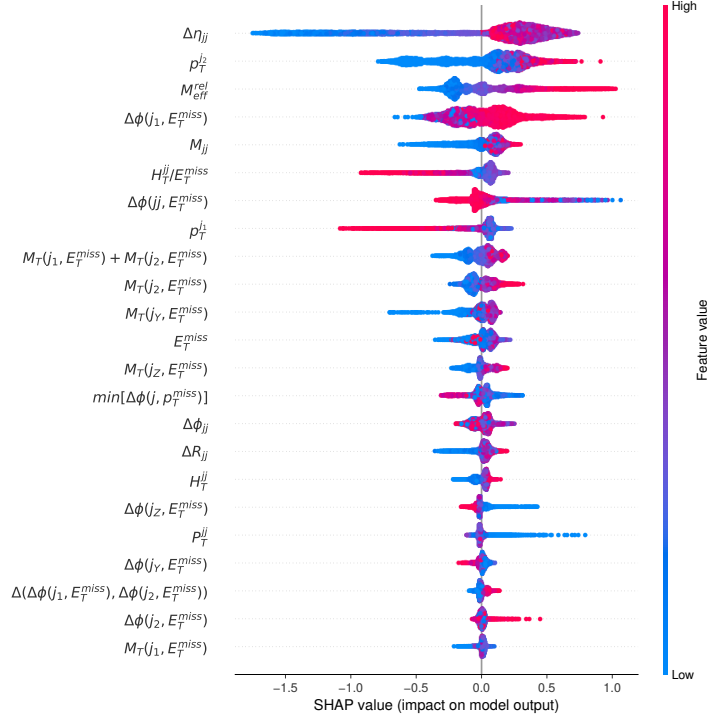


Figure 5.12: SHAP value summary plot for one of the optimum VBF BDTs.

SHAP values for the input variables for one of the models with the optimum hyperparameter configuration are shown in Figure 5.12, ranked in order of their importances. From this it can be seen that, as in the additional selections for SR-CMS, events with high $\Delta\eta_{jj}$, high p_T^{j2} and high M_{jj} contribute to high signal like scores. Additionally, events with high $M_{\text{eff}}^{\text{rel}}$ are very signal like, showing the discrimination power present in both E_T^{miss} and M_{jj} . Signal events also tend to have a high $\Delta\phi(j^1, E_T^{\text{miss}})$ while events with very high p_T^{j1} are determined to be more background like. Feature importances for the models trained on the other folds are checked and are found to be very similar.

The threshold on the BDT classifier score of 0.65 defining SR-BDT is chosen such that the total number of background events in SR-BDT is similar to that in SR-CMS. This is in order to allow a fair comparison between shape fits performed in the two SRs. Table 5.11 shows the total background and signal yields in SR-CMS and SR-BDT (including correction factors on the V +jet backgrounds). It is evident that the signal yields are $\sim 50\%$ higher in SR-BDT than in SR-CMS for approximately the same number of background as a result of the BDT selection.

Process	SR-CMS	SR-BDT
QCD Z+jets	12890 ± 70	16630 ± 60
QCD W+jets	6320 ± 70	6140 ± 60
EW Z+jets	2153 ± 10	2907 ± 12
EW W+jets	1066 ± 6	1177 ± 7
Top	320 ± 7	188 ± 5
Multi-boson	598 ± 11	671 ± 12
SM total	23350 ± 100	27710 ± 90
HH, (100, 99)	214.9 ± 3.1	341 ± 4
WB, (100, 99)	523 ± 6	782 ± 8
WB, (140, 139)	260.0 ± 3.4	399 ± 4

Table 5.11: Total yields (at 139 fb^{-1}) in SR-CMS and SR-BDT. Correction factors to the V +jets backgrounds are applied. Uncertainties are purely statistical.

Shape fit optimization

As mentioned, within each of the SRs shape fits are performed to maximize the sensitivity. For SR-CMS, three different shape fits are tested. The first replicates the binning in M_{jj} used in the CMS analysis, which features 12 bins in total. The second instead utilises a binning in E_T^{miss} , as done in the mono-jet analysis, in order to compare the discrimination power of E_T^{miss} and M_{jj} . For this fit, the binning is determined using the following procedure. First, the lower edge of the final bin is chosen to maximize the significance for the wino/bino $m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (100, 99)$ signal in the final bin, as determined by the formula provided in [151], using a flat 10% systematic uncertainty on the background predictions (combined in quadrature with the MC statistical uncertainty). The value of 10% on this final bin is motivated by the uncertainty on the final M_{jj} bin in the CMS analysis. Subsequent bin values are then selected such that the number of events in each bin is similar to the number of events in the final bins used in the mono-jet analysis SR. This procedure leads to 8 bins being defined in total. The third fit in SR-CMS is binned in M_{jj} like the first but follows the same binning procedure as used for E_T^{miss} , and so also results in 8 bins. This configuration is tested in order to allow a fairer comparison between the discrimination power of E_T^{miss} and M_{jj} . In SR-BDT, a single shape fit is performed in the BDT classifier score with the binning determined using the same procedure, thus also resulting in 8 bins.

Plots of M_{jj} , E_T^{miss} and the BDT classifier score illustrating the selection of the cut values for the final bins is shown in Figure 5.13. Significance scores are shown for the wino/bino $m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (100, 99)$, wino/bino $m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (150, 149)$, and higgsino $m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (100, 99)$ scenarios. Evident from these plots is that the significance score for cutting on the BDT score has a much faster rise than cutting on M_{jj} or E_T^{miss} . It is also evident that the optimal cut value does not differ greatly between the different signal points. Figure 5.14 shows the binning used for each of the fits and shows the ratio of signal to background yields (S/B) of each of the example signals in each bin. From these plots it can be seen that the S/B varies from $\sim 2\%$ up to a maximum of 10% in the highest BDT score bin. Yields in each of the last three bins for each shape fit are shown in Tables 5.12-5.15 for reference. These yields include the correction factors applied to the V -jet backgrounds.

Although the S/B is much smaller in the lower SR bins, they do contribute to the overall sensitivity due to having smaller statistical and systematic uncertainties.

Chapter 5: Searches for Ultra-compressed Charginos and Neutralinos

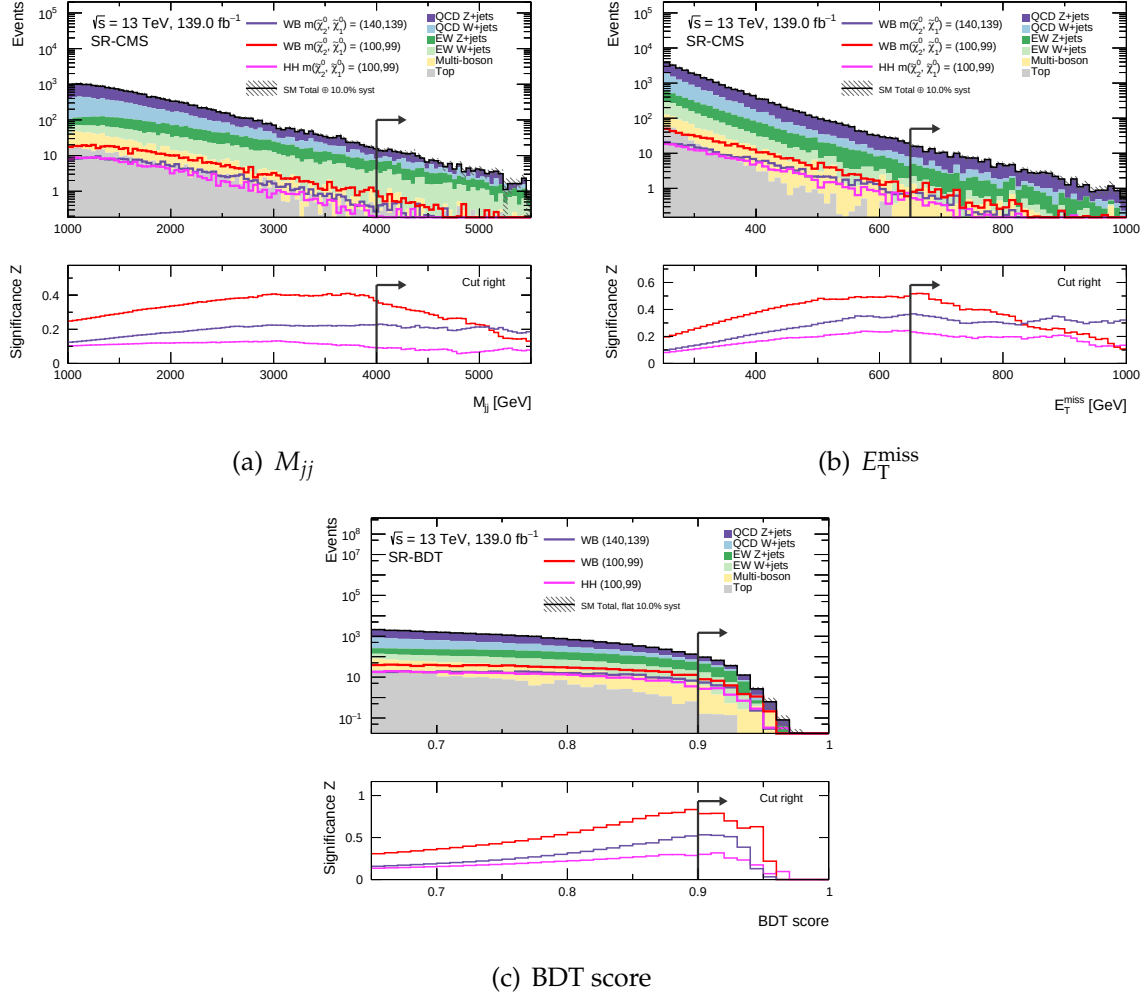


Figure 5.13: Plots showing significance score for the VBF SRs and arrows indicating the lower edge of the last bin used in each of the 8-bin shape fits.

As an indication of this, plots showing the significance score per bin for each of the shape fit configurations are shown in Figure 5.15 in which the total uncertainty on the background estimate is taken to vary linearly from 3-10% between the first and last SR bin (which is roughly consistent with the overall effect of the uncertainties that will be applied to the fits, discussed in Section 5.3.6).

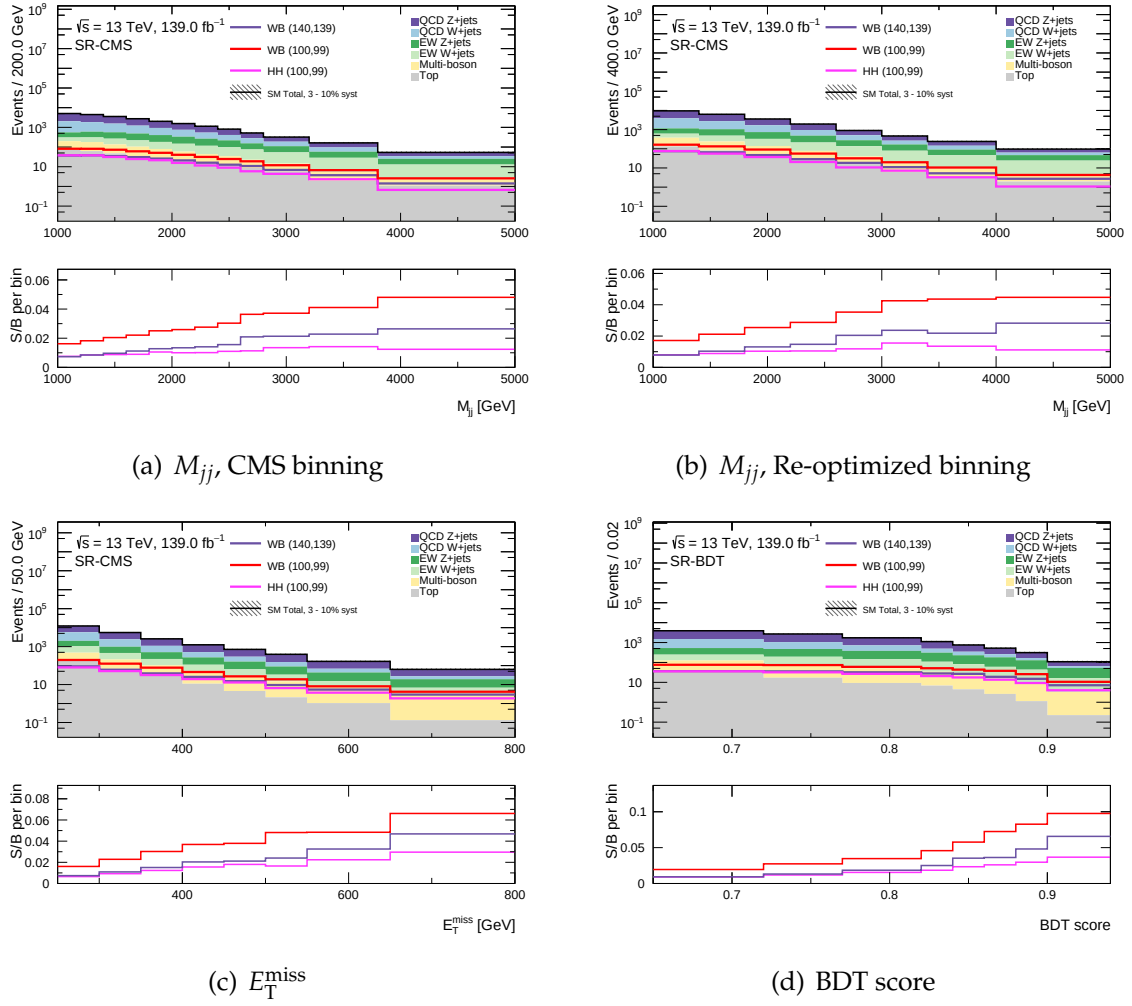


Figure 5.14: Plots showing the binning used for each of the shape fits and the S/B for each bin for example signal points. The correction factors for the V +jet backgrounds described in Section 5.3.5 are applied.

Process	Bin 10 M_{jj} [2800, 3200) GeV	Bin 11 M_{jj} [3200, 3800) GeV	Bin 12 $M_{jj} > 3800$ GeV
QCD Z+jets	267 ± 9	201 ± 9	114 ± 7
QCD W+jets	154 ± 11	114 ± 8	55 ± 5
EW Z+jets	114.0 ± 2.1	88.3 ± 1.9	74.2 ± 1.7
EW W+jets	77.8 ± 1.8	69.0 ± 1.6	63.5 ± 1.6
Top	11.3 ± 1.2	5.8 ± 0.9	6.5 ± 0.9
Multi-boson	17.8 ± 1.8	10.6 ± 1.4	7.3 ± 1.1
SM total	642 ± 14	489 ± 12	321 ± 9
HH, (100, 99)	8.7 ± 0.6	6.9 ± 0.6	4.0 ± 0.4
WB, (100, 99)	23.9 ± 1.2	20.1 ± 1.1	15.4 ± 1.0
WB, (140, 139)	13.7 ± 0.7	11.1 ± 0.7	8.5 ± 0.6

Table 5.12: Yields (at 139 fb^{-1}) for the last three bins for the fit in M_{jj} using the binning in the CMS analysis. Correction factors to the V +jets backgrounds are applied. Uncertainties are purely statistical.

Process	Bin 6 M_{jj} [3000, 3400) GeV	Bin 7 M_{jj} [3400, 4000) GeV	Bin 8 $M_{jj} > 4000$ GeV
QCD Z+jets	200 ± 8	145 ± 7	83 ± 6
QCD W+jets	97 ± 9	81 ± 6	42 ± 4
EW Z+jets	89.0 ± 1.8	67.6 ± 1.6	56.1 ± 1.5
EW W+jets	62.0 ± 1.5	54.9 ± 1.5	50.0 ± 1.4
Top	6.2 ± 0.9	5.4 ± 0.9	4.9 ± 0.8
Multi-boson	11.1 ± 1.4	8.5 ± 1.3	5.6 ± 0.9
SM total	466 ± 13	362 ± 9	241 ± 7
HH, (100, 99)	7.2 ± 0.6	4.9 ± 0.5	2.69 ± 0.34
WB, (100, 99)	19.8 ± 1.1	15.8 ± 1.0	10.8 ± 0.8
WB, (140, 139)	11.0 ± 0.7	7.9 ± 0.6	6.8 ± 0.5

Table 5.13: Yields (at 139 fb^{-1}) for the last three bins for the fit in M_{jj} using 8 bins. Correction factors to the V +jets backgrounds are applied. Uncertainties are purely statistical.

Process	Bin 6 E_T^{miss} [500, 550) GeV	Bin 7 E_T^{miss} [550, 650) GeV	Bin 8 $E_T^{miss} > 650$ GeV
QCD Z+jets	239 ± 6	197 ± 5	108.1 ± 3.3
QCD W+jets	64 ± 5	50 ± 4	24.3 ± 2.4
EW Z+jets	56.4 ± 1.7	56.3 ± 1.7	37.0 ± 1.4
EW W+jets	18.3 ± 0.9	15.7 ± 0.8	9.1 ± 0.6
Top	2.1 ± 0.5	2.1 ± 0.6	0.40 ± 0.23
Multi-boson	12.7 ± 1.4	12.5 ± 1.3	11.0 ± 1.3
SM total	393 ± 8	334 ± 7	190 ± 5
HH, (100, 99)	6.5 ± 0.5	7.5 ± 0.6	5.6 ± 0.5
WB, (100, 99)	18.9 ± 1.2	16.2 ± 1.1	12.6 ± 0.9
WB, (140, 139)	9.4 ± 0.6	10.9 ± 0.7	8.9 ± 0.6

Table 5.14: Yields (at 139 fb^{-1}) for the last three bins for the fit in E_T^{miss} . Correction factors to the V +jets backgrounds are applied. Uncertainties are purely statistical.

Process	Bin 6 BDT score [0.86, 0.88)	Bin 7 BDT score [0.88, 0.9)	Bin 8 BDT score > 0.9
QCD Z+jets	262 ± 6	145 ± 4	89 ± 4
QCD W+jets	80 ± 4	38.2 ± 3.0	23.0 ± 1.9
EW Z+jets	120.9 ± 2.3	85.1 ± 1.9	73.1 ± 1.8
EW W+jets	45.0 ± 1.3	33.4 ± 1.1	26.2 ± 1.0
Top	2.6 ± 0.6	1.1 ± 0.4	0.45 ± 0.23
Multi-boson	12.6 ± 1.4	8.9 ± 1.2	6.2 ± 1.1
SM total	523 ± 8	312 ± 6	218 ± 5
HH, (100, 99)	13.6 ± 0.8	9.3 ± 0.6	8.0 ± 0.6
WB, (100, 99)	37.9 ± 1.6	25.8 ± 1.3	21.3 ± 1.2
WB, (140, 139)	18.9 ± 0.9	15.0 ± 0.8	14.3 ± 0.8

Table 5.15: Yields (at 139 fb^{-1}) for the last three bins for the fit in the BDT score. Correction factors to the V +jets backgrounds are applied. Uncertainties are purely statistical.

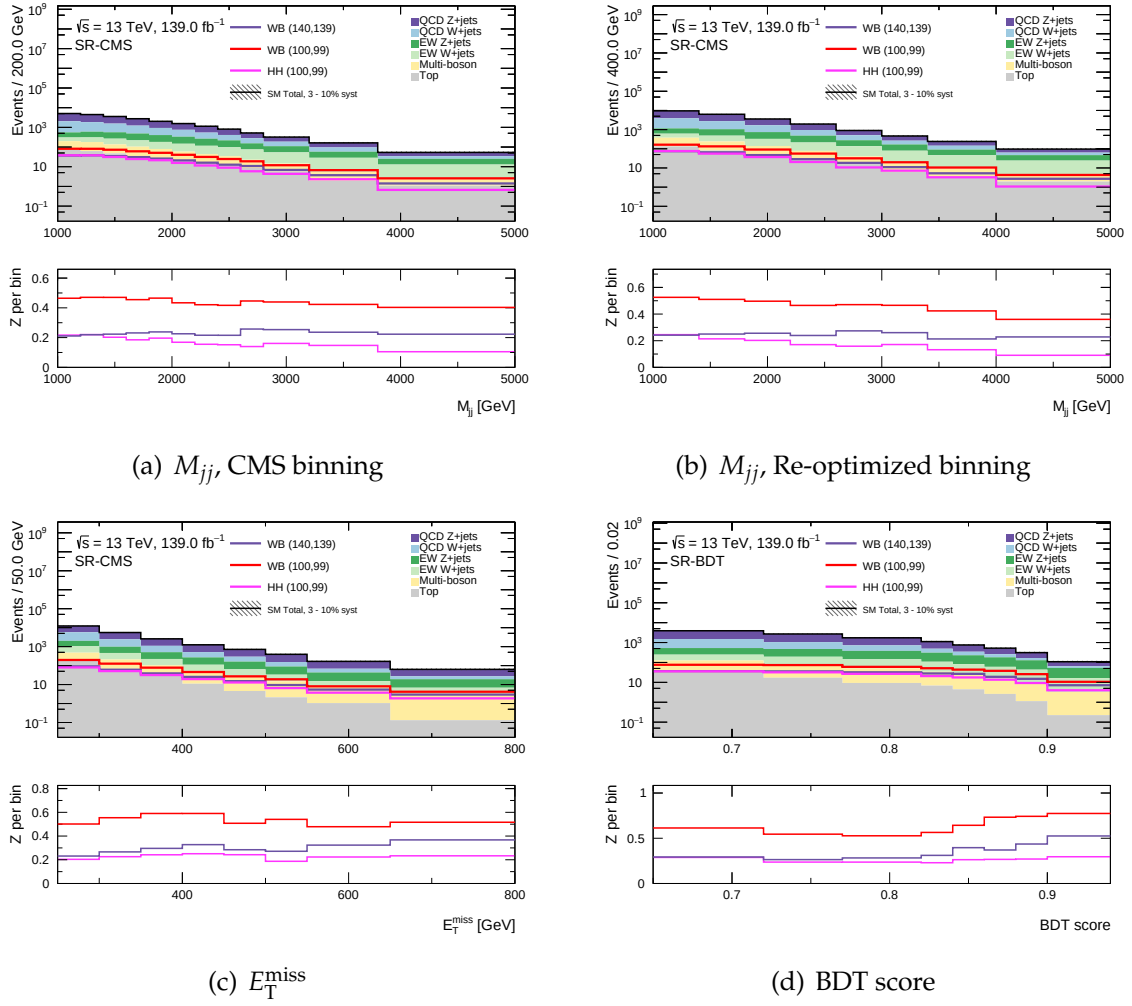


Figure 5.15: Plots showing the significance score per bin for each of the SR shape fit configurations. The correction factors for the V +jet backgrounds described in Section 5.3.5 are applied.

5.3.5 Background Estimation

As mentioned, four CRs requiring leptons are defined in order to check the modelling of the V +jets backgrounds, particularly for the discriminating variables used for the SR shape fits. These CRs are safe to unblind as they contain no signal contamination due to the lepton requirements. The selections defining these CRs are listed in Table 5.16. All the CRs have common selections on M_{jj} and the transverse momenta of the VBF-tagged jets that replicate the SR selections. To replicate the E_T^{miss} selection and distribution in the SR, a version of the E_T^{miss} is computed in which the leptons are not included and are thus treated as being invisible, as is done in the mono-jet and ATLAS Higgs $\rightarrow$ invisible analyses. Here this variable is denoted as $E_T^{\text{miss}}(\text{L.I.})$ where L.I. stands for ‘Leptons Invisible’. Versions of other variables dependent on the E_T^{miss} , including the BDT score, are computed using $E_T^{\text{miss}}(\text{L.I.})$ and have the same notation. The leading lepton is required to have $p_T > 30$ GeV in all regions, a cut taken directly from the ATLAS Higgs $\rightarrow$ invisible analysis, that suppresses fake backgrounds and ensures events are on the lepton-trigger efficiency plateaus. Events with b -jets or hadronic tau-leptons are vetoed in the CRs.

The one lepton CRs, CR-1L-El and CR-1L-Mu, are enriched in $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ events respectively and serve to validate the modelling of W +jets. Events in these regions are required to pass either the E_T^{miss} or single lepton triggers and must have exactly one signal electron or one signal muon. In addition to the common CR requirements, events entering these regions are required to have the transverse mass between the lepton and the E_T^{miss} be $30 < M_T(l_1, E_T^{\text{miss}}) < 100$ GeV in order to improve the purity in W +jets and have E_T^{miss} significance > 5 to reduce the number of events with fake leptons.

Similarly, the two lepton CRs, CR-2L-El and CR-2L-Mu, are enriched in $Z \rightarrow ee$, $Z \rightarrow \mu\mu$ respectively and serve to validate the modelling of Z +jets. For these regions, an OR of the single- and di-lepton triggers is used and events are required to have exactly two signal electrons or two signal muons of opposite sign. The invariant mass of the two leptons is further required to fulfil $61 < M_{ll} < 121$ GeV to increase the purity of $Z \rightarrow \ell\ell$ events. As these events are not expected to have real missing transverse energy, an upper cut of 70 GeV is placed on the E_T^{miss} .

Two variations of each CR are defined. The first, labelled as [CR]-LC, applies on top of the nominal selections the ‘linear cuts’ on M_{jj} , $\Delta\eta_{jj}$ and $p_T^{j_2}$ listed in Table 5.16 that are required in SR-CMS. These variations are used to validate the M_{jj} and E_T^{miss}

(L.I.) distributions and determine the V +jets correction factors used in the shape fits performed in SR-CMS. The second variety, labelled [CR]-BDT, instead applies a cut of BDT score (L.I.) > 0.65 to replicate the BDT score cut in SR-BDT and are used to verify the modelling of the BDT score (L.I.) and determine the correction factors for the shape fit in SR-BDT.

Process	W+jets		Z+jets	
Region	CR-1L-El	CR-1L-Mu	CR-2L-El	CR-2L-Mu
Triggers	$E_{\text{T}}^{\text{miss}}$ OR single-lepton		single-lepton OR di-lepton	
M_{jj}	$> 600 \text{ GeV}$			
$p_{\text{T}}^{j_1}$	$> 60 \text{ GeV}$			
$p_{\text{T}}^{j_2}$	$> 40 \text{ GeV}$			
$E_{\text{T}}^{\text{miss}}$ (L.I.)	$> 250 \text{ GeV}$			
$\min_j[\Delta\phi(j, E_{\text{T}}^{\text{miss}}(\text{L.I.}))]$	> 0.4			
$p_{\text{T}}^{l_1}$	$> 30 \text{ GeV}$			
N b -jets	$= 0$			
N hadronic tau-leptons	$= 0$			
N signal electrons	$= 1$	$= 0$	$= 2$	$= 0$
N signal muons	$= 0$	$= 1$	$= 0$	$= 2$
OS	-		Yes	
$M_T(l_1, E_{\text{T}}^{\text{miss}})$	$\epsilon[30, 100] \text{ GeV}$		-	
$E_{\text{T}}^{\text{miss}}$ Significance	> 5		-	
M_{ll}	-		$\in [61, 121] \text{ GeV}$	
$E_{\text{T}}^{\text{miss}}$	-		$< 70 \text{ GeV}$	
Linear Cuts (LC)	$M_{jj} > 1000 \text{ GeV}$			
	$\Delta\eta_{jj} > 3.8$			
	$p_{\text{T}}^{j_2} > 60 \text{ GeV}$			
BDT selection	BDT score (L.I.) > 0.65			

Table 5.16: Baseline selection requirements defining the VBF control regions. $p_T^{l_1}$ denotes the p_T of the leading lepton, $M_T(l_1, E_T^{\text{miss}})$ the transverse mass between this lepton and the E_T^{miss} and OS where two signal leptons of opposite electric charge are required.

As discussed in the analysis strategy overview, correction factors are determined from comparing the M_{jj} , E_T^{miss} (L.I.) and BDT score (L.I.) distributions to data in the CRs. Separate correction factors are determined for W+jets and Z+jets, but the fits to the relevant one and two lepton CRs are performed simultaneously. As there is

evidence of a shape trend in the data/MC comparisons for some of these distributions, for determining the correction factors each distribution is split into a low and high portion and separate correction factors are determined for each portion. The boundary between low and high is chosen to coincide with the lower edge of the fifth bin for the 8-bin shape fits, such that the first four bins comprise the low portion and the latter four the high portion (for the 12 bin CMS shape fit in M_{jj} , the lower edge of the ninth bin coincides with the lower edge of the fifth bin for the 8-bin M_{jj} shape fit).

The computed correction factors are listed in Table 5.17, for which the uncertainties arise purely from the limited MC and data statistics in each of the regions. Only the correction factors determined from the relevant distribution in the appropriate CRs are used for each SR shape fit. The correction factors are all less than one, indicating an over-prediction of the SM background in all regions. Distributions of M_{jj} , E_T^{miss} (L.I.) and BDT score (L.I.) in each of the CRs (for the relevant variations) are shown in Figures 5.16-5.19 with and without the correction factors applied (labelled as ‘Pre-fit’ and ‘Post-fit’ respectively). In these plots, an uncertainty which varies linearly from ~ 1 -10% between the lowest and highest bin is shown for illustrative purposes but this uncertainty is not included during the determination of the correction factors. The binning used for these distributions matches the binning used in the corresponding SR shape-fits. Disagreement in the ‘Pre-fit’ distributions is most evident in M_{jj} where the over-prediction worsens at higher values. Hence, the correction factors deviate most from one in the high M_{jj} regions, with the most extreme correction factor being ~ 0.75 for Z+jets. Conversely, the over-prediction is worse at lower values of E_T^{miss} (L.I.) than higher values, while the BDT score (L.I.) distributions have the flattest pre-fit data/MC ratios. With the correction factors applied, the agreement is reasonable in all regions, consistent within the illustrative uncertainty, thus the modified predictions are deemed to be useful in estimating the sensitivity of the VBF channel.

Total yields in each of the CRs are shown in Table 5.18. The background yields are from the M_{jj} shape fits, including the correction factors, and include the dummy systematic uncertainties. Within these uncertainties, the total yields agree well with the data. Evident from this table is the high purity of these CRs in their target backgrounds; the purity varies from 88% for the one lepton CRs to 96% for the two lepton CRs. Also evident is that the yields for the Z+jet CRs are about five times smaller than the W+jet CRs, owing to the smaller $Z \rightarrow \ell^+ \ell^-$ branching ratio compared to the branching ratio of $W \rightarrow \ell \nu$.

SR Shape Fit Variable	Region	Correction Factors	
		W+jets	Z+jets
M_{jj}	$M_{jj} < 2600 \text{ GeV}$	0.812 ± 0.018	0.896 ± 0.024
	$M_{jj} \geq 2600 \text{ GeV}$	0.788 ± 0.030	0.74 ± 0.05
E_T^{miss}	$E_T^{\text{miss}} < 450 \text{ GeV}$	0.796 ± 0.018	0.874 ± 0.023
	$E_T^{\text{miss}} \geq 450 \text{ GeV}$	0.940 ± 0.033	0.99 ± 0.07
BDT score	BDT score < 0.84	0.862 ± 0.012	0.939 ± 0.019
	BDT score ≥ 0.84	0.791 ± 0.025	0.85 ± 0.06

Table 5.17: Data-driven correction factors applied to the W+jets and Z+jets backgrounds. Uncertainties arise solely from the finite data and MC statistics in the respective regions.

Process	CR-1L-El-LC	CR-1L-Mu-LC	CR-2L-El-LC	CR-2L-Mu-LC
QCD Z+jets	23.2 ± 0.9	56.7 ± 2.2	1040 ± 40	1330 ± 50
QCD W+jets	4900 ± 190	4860 ± 190	0.138 ± 0.007	1.23 ± 0.04
EW Z+jets	3.63 ± 0.22	6.5 ± 0.4	209 ± 12	243 ± 14
EW W+jets	1150 ± 70	1080 ± 70	0.123 ± 0.009	0.105 ± 0.011
Top	583 ± 23	529 ± 21	2.26 ± 0.13	2.99 ± 0.11
Multi-boson	205 ± 9	194 ± 8	36.7 ± 1.5	43.4 ± 1.8
SM total	6870 ± 210	6730 ± 210	1290 ± 40	1620 ± 50
Observed	7048	6558	1344	1574

Table 5.18: Yields (at 139 fb^{-1}) for the Control Regions used in the VBF sensitivity study.

Chapter 5: Searches for Ultra-compressed Charginos and Neutralinos

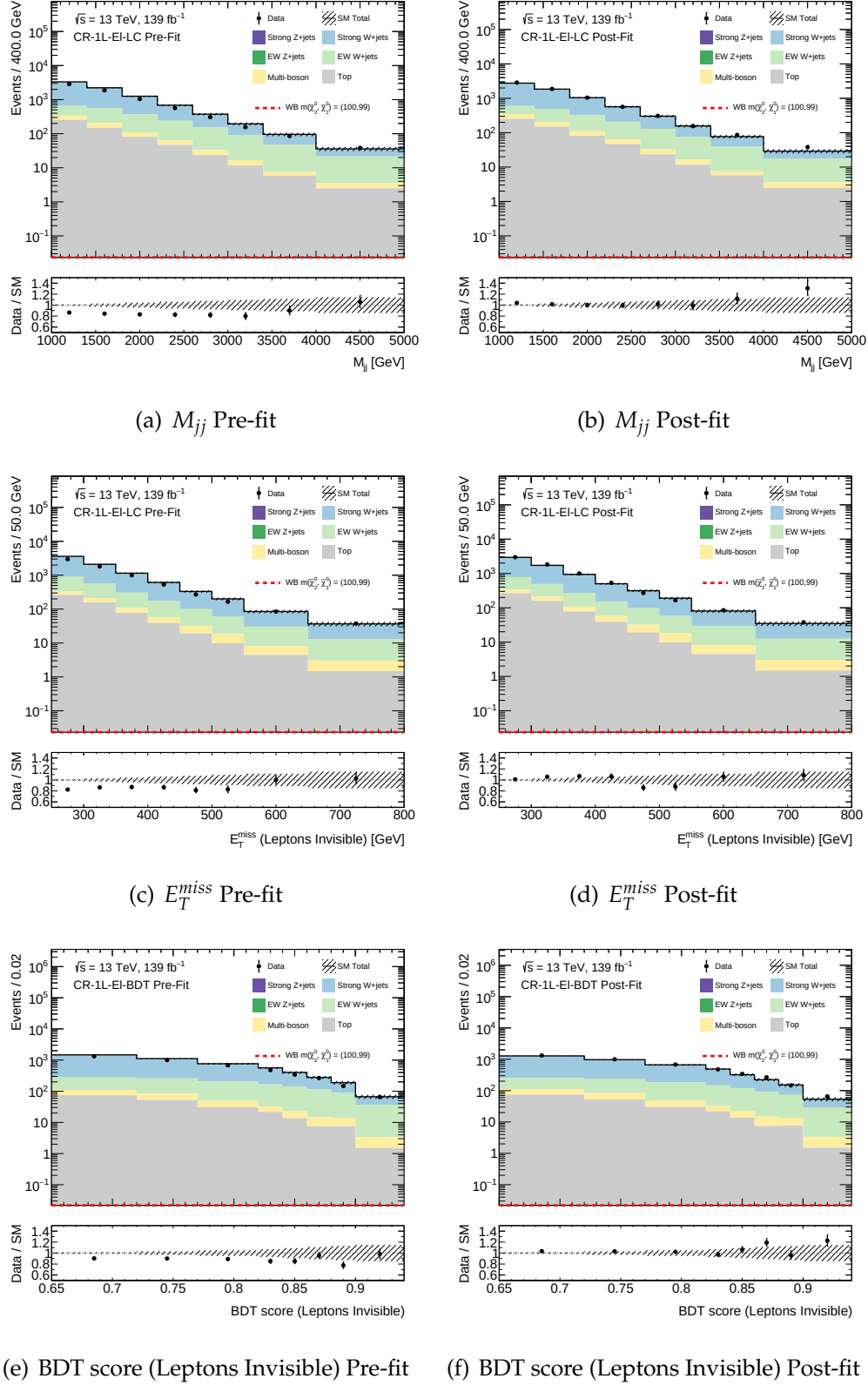


Figure 5.16: Pre and Post-fit kinematic distributions in CR-1L-El. First and last bins include the underflow and overflow respectively.

Chapter 5: Searches for Ultra-compressed Charginos and Neutralinos

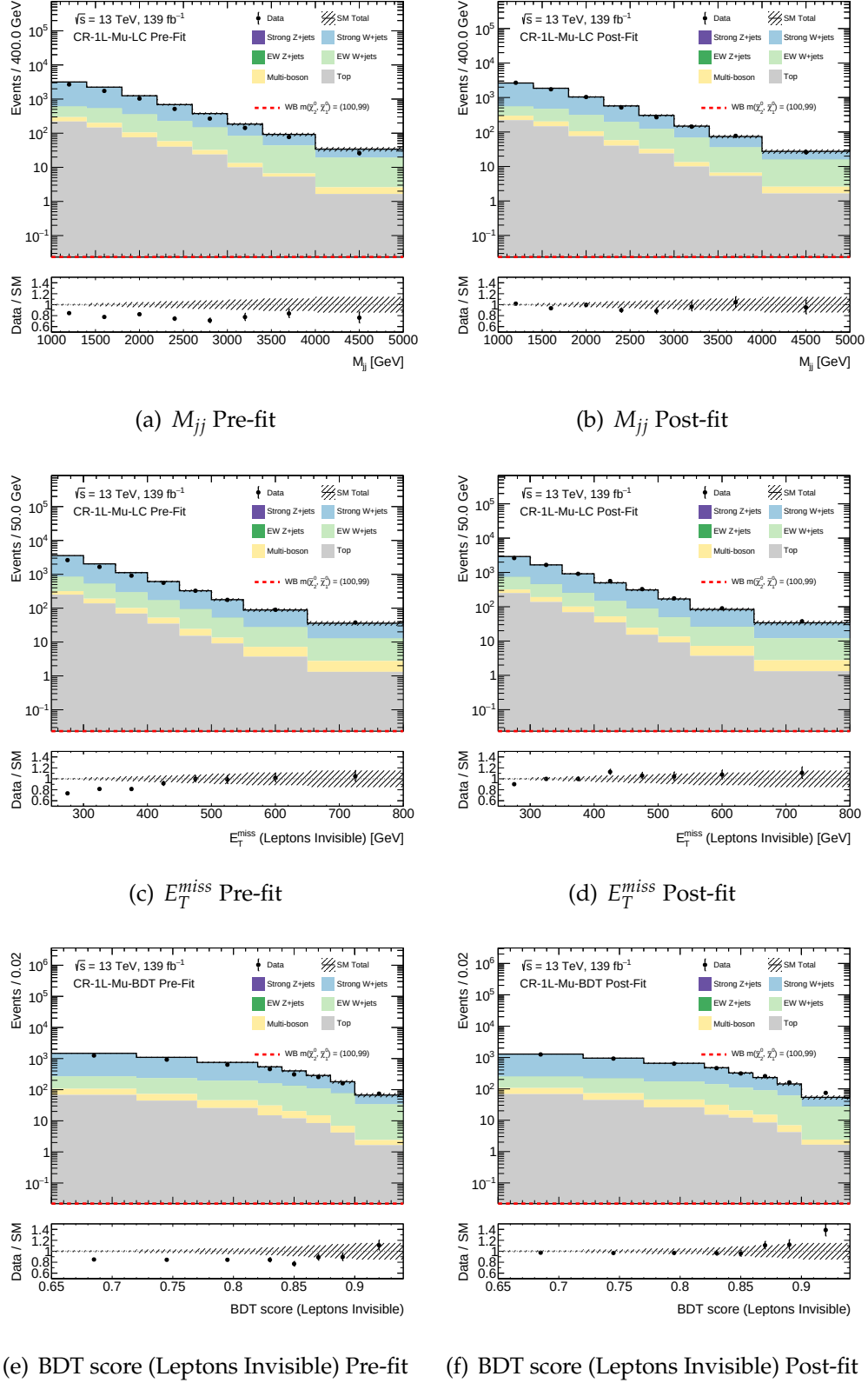


Figure 5.17: Pre and Post-fit kinematic distributions in CR-1L-Mu. First and last bins include the underflow and overflow respectively.

Chapter 5: Searches for Ultra-compressed Charginos and Neutralinos

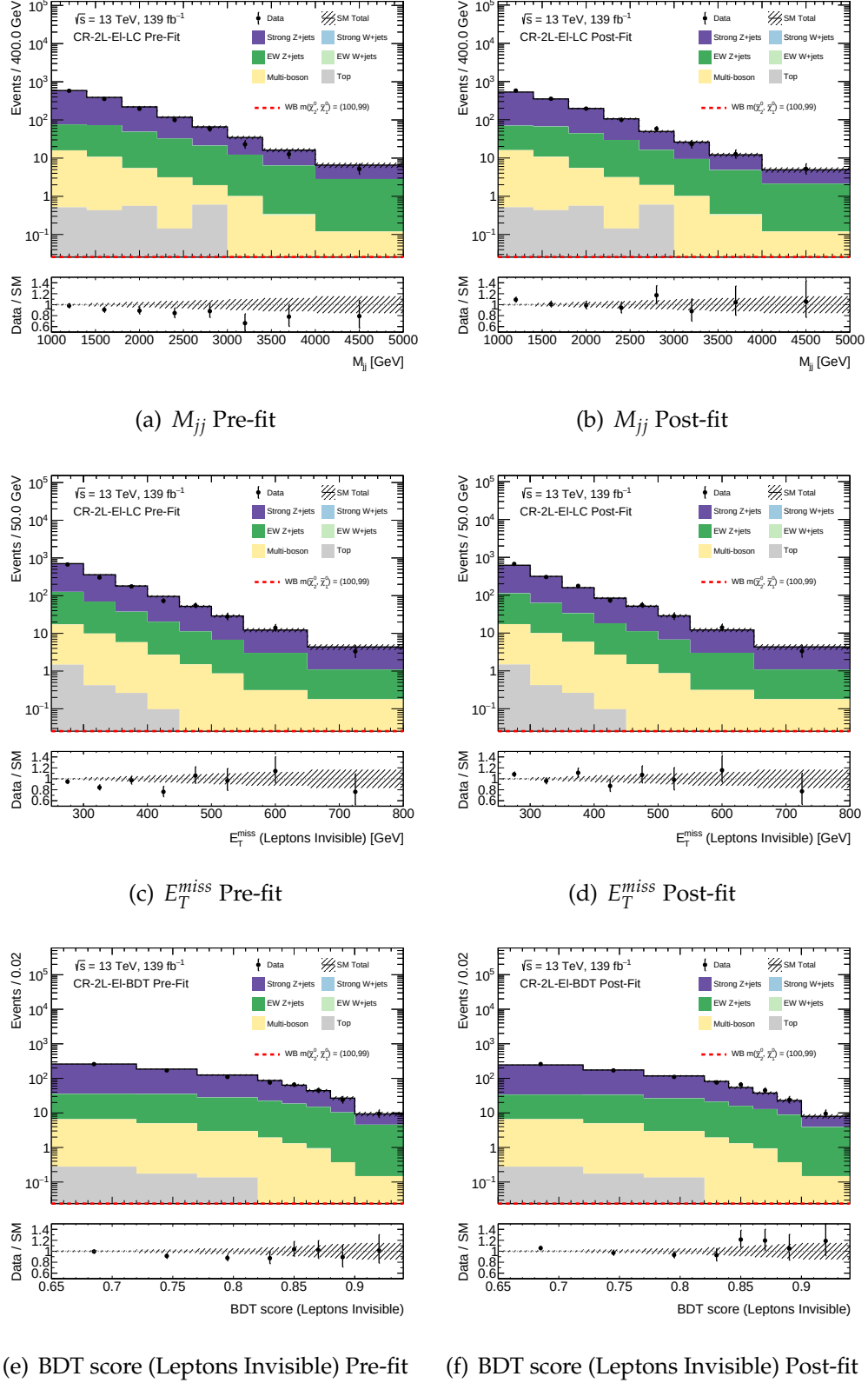


Figure 5.18: Pre and Post-fit kinematic distributions in CR-2L-El. First and last bins include the underflow and overflow respectively.

Chapter 5: Searches for Ultra-compressed Charginos and Neutralinos

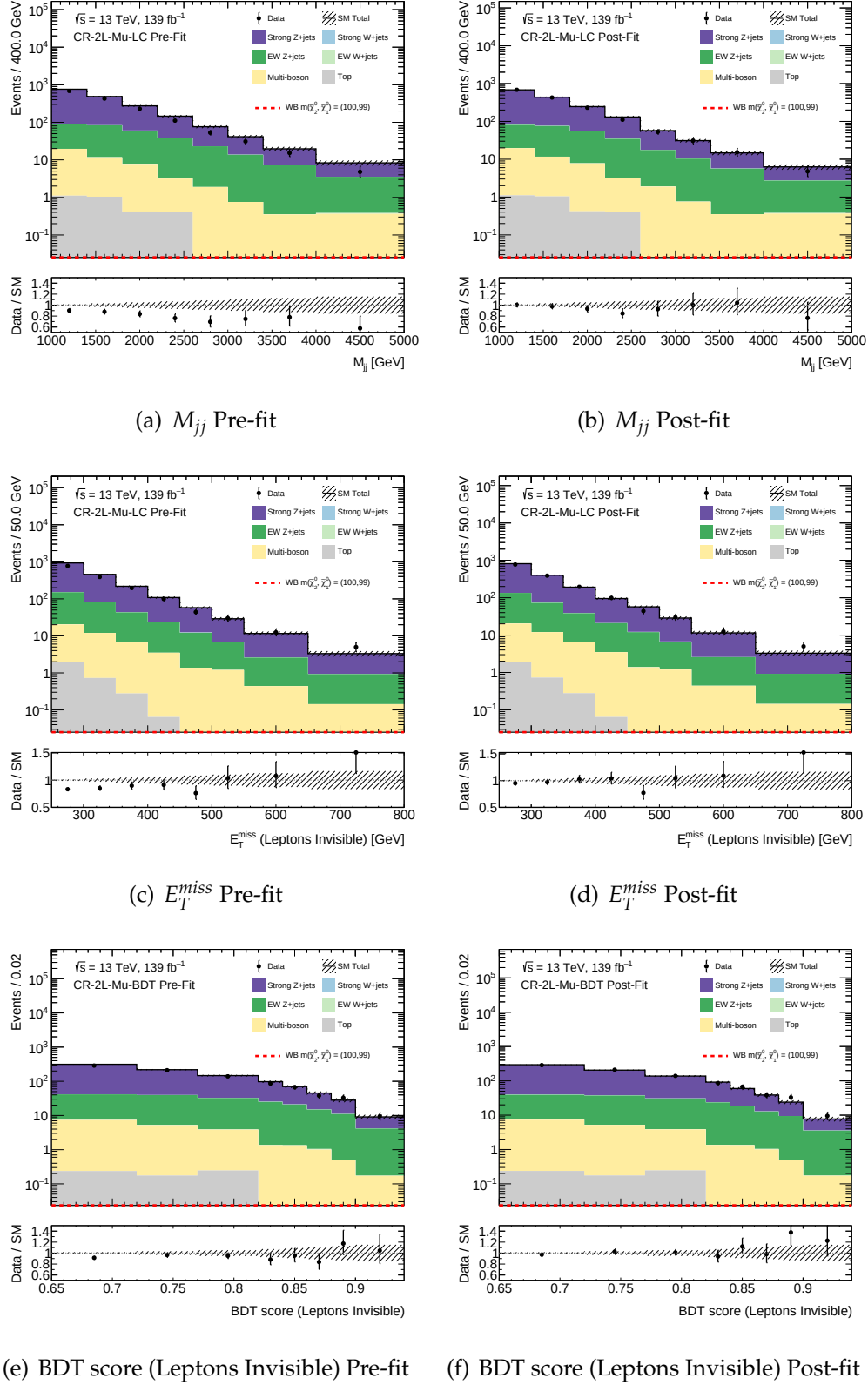


Figure 5.19: Pre and Post-fit kinematic distributions in CR-2L-Mu. First and last bins include the underflow and overflow respectively.

5.3.6 Evaluation of Sensitivity

Fit Setup and Systematic Uncertainties

As this analysis features a shape fit across many SR bins, significance scores alone cannot provide a useful estimate of the sensitivity. Instead, HistFitter [84] is used to compute upper limits on the signal strength from fits to the multi-bin SRs. To accurately assess the sensitivity, it is important to realistically model systematic uncertainties, particularly on the V +jets background estimates. As mentioned, for this sensitivity study simplified systematic uncertainties are included in the fits to represent the impact of theoretical and experimental systematic uncertainties. Three different configurations of systematic uncertainties on the V +jets backgrounds are tested, based upon the strategies taken in the CMS VBF SUSY analysis [153], the ATLAS VBF Higgs→invisible analysis [159] and the ATLAS mono-jet analysis [152] respectively, in order to build a realistic picture of the sensitivity. Each of the configurations is described in turn in the following and are summarized in Table 5.19.

In each case, the size of uncertainties is approximated based upon values taken from the respective analysis documentation. The uncertainties are taken to be the same for each of the discriminating variables used in the shape fit. The validity of this assumption is discussed more after the results are presented, but it is thought to be reasonable given that each of the SR bins contain roughly the same number of events across M_{jj} , E_T^{miss} and the BDT score. Importantly, only the SR is used in each fit; the CRs are *not* included as the uncertainties which are used for each configuration are designed to represent the uncertainties on the SR predictions taking into account the constraining and cancellation effects of the CRs. For each configuration, the total pre-fit uncertainties on the background estimates per bin are similar and vary from $\sim 3\%$ in the first SR bin to around $\sim 10\%$ in the final SR bin.

Some common uncertainties are included across all the configurations. To represent the overall impact of experimental uncertainties on the non V +jets processes, a 5% uncertainty, correlated across the SR bins, is included on the signal and multi-boson and top backgrounds. This uncertainty is correlated between all processes. It is also included on the V +jets backgrounds but takes a smaller value of 0.1%, to represent its impact after cancellations from the CRs are taken into account. In addition, uncertainties varying linearly from 10-20% across the SR bins are included on the multi-boson and top backgrounds to represent modelling uncertainties on their predictions. These

uncertainties are uncorrelated between multi-boson and top but are correlated across the SR bins for each process. A global 5% uncertainty is also applied to each of these backgrounds to represent uncertainties on their cross-sections. No uncertainties emulating modelling uncertainties are included on the signal processes to facilitate a fair comparison with the results of the mono-jet re-interpretation presented in the previous section. A 1% uncertainty on the luminosity, based upon the recent ATLAS measurement [52], is also applied in all fits and MC statistical uncertainties are included on the signal and all background processes.

Having established the uncertainties which are common to all configurations, those which are unique to each are now detailed, beginning with the CMS style configuration. In the CMS VBF SUSY analysis, separate CRs are defined for W +jets and Z +jets with global transfer factors (TFs) defined for each process. Hence, in the CMS style configuration global uncertainties are included on W +jets and Z +jets to represent the uncertainties on these TFs. Based upon the total data statistics available in the electron and muon CRs defined here, for W +jets, the uncertainty is taken to be 1%, while for Z +jets it is 2% due to two lepton CRs having fewer statistics. In the CMS analysis, shape uncertainties on M_{jj} are evaluated based upon discrepancies in its modelling in the CRs and are quoted as being as large as 10% in certain bins (though it should be noted that the statistical uncertainty in the high M_{jj} bins in the CMS Z +jets CR are $\sim 50\%$). To account for these, per bin uncertainties are included in the CMS style configuration which vary linearly from 1-10% from the first to last SR bin (it is assumed that discrepancies worsen at high values of the discriminating variable). These uncertainties are uncorrelated between bins and between W +jets and Z +jets but are correlated between the QCD and EW processes of the same bosonic variety. Theoretical modelling uncertainties are not described for the V +jets backgrounds given the ‘data-driven’ estimation method, and so are not included in the CMS style configuration.

Moving to the ATLAS Higgs $\rightarrow$ invisible style configuration, instead of global TF uncertainties on W +jets and Z +jets, per bin TF uncertainties are applied to reflect the background estimation strategy taken in the ATLAS Higgs $\rightarrow$ invisible analysis, which defines separate CRs for each SR bin. As well as being a very data-driven method, this strategy also results in significant cancellations of the modelling uncertainties which are quoted as being particularly large ($\sim 30\%$) at high M_{jj} . The analysis make use of a dedicated next-to-leading order theory calculation of the ratio of W +jets to Z +jets in a ‘VBF like’ phase space to base the Z +jets predictions off of the W +jets predictions, such

that, for each SR bin, the W +jets and Z +jets CRs can be effectively combined to reduce the size of the TF uncertainties (which would otherwise be particularly large for Z +jets due to the limited CR statistics). Given this, based upon the statistics available in the CRs using the different binning schema defined here, the TF uncertainties are taken to vary from 1.5-7%, being largest in the high SR bins which have fewer statistics. These uncertainties are uncorrelated across bins but are correlated between all the V +jets processes. To reflect the uncertainty in the W +jets to Z +jets ratio, per bin uncertainties are applied on each of the QCD and EW Z +jets predictions. These uncertainties are taken to vary from 5-10% from the low to high SR bins and are uncorrelated across the bins and between QCD Z +jets and EW Z +jets. Uncertainties varying from 0.1-3% are also applied on the V +jets processes to represent residual modelling uncertainties, as the cancellation from the CRs to the SRs is not described as being prefect. Following the description given in the analysis documentation, these uncertainties are correlated between QCD W +jets and QCD Z +jets and between EW W +jets and EW Z +jets but not between the QCD and EW processes.

The third configuration tested is based upon the ATLAS mono-jet analysis. In this configuration, a global 1% uncertainty is applied to all the V +jets backgrounds to reflect the total uncertainty on the V +jets TF that is applied to all the V +jets backgrounds in the Mono-jet analysis. The same uncertainties per bin to reflect the uncertainty on the ratio of W +jets to Z +jets as used in the Higgs→invisible configuration are also included on the Z +jets backgrounds to reflect the fact that having a global V +jets TF relies upon using the W +jets estimate for the Z +jets estimate. As a global TF is defined, the cancellation of modelling uncertainties is assumed to not be as great in the high SR bins as it would be in the TF per bin approach taken by the Higgs→invisible analysis. Hence, the residual modelling uncertainties in the high SR bins are taken to be larger; varying from 0.1-10% instead of 0.1-3% (based upon the size of the modelling uncertainties provided in the Higgs→invisible analysis documentation). As in the Higgs→invisible style configuration, these uncertainties are correlated between QCD W +jets and QCD Z +jets and between EW W +jets and EW Z +jets but not between the QCD and EW processes.

Chapter 5: Searches for Ultra-compressed Charginos and Neutralinos

Configuration	Uncertainties
CMS VBF SUSY style	• Global 1% uncertainty on the W+jets backgrounds to represent overall TF uncertainty, correlated between the QCD and EW processes. • Global 2% uncertainty on the Z+jets backgrounds to represent overall TF uncertainty, correlated between the QCD and EW processes. • Per bin uncertainties varying from 1-10% for each of the V+jets backgrounds to represent CR modelling discrepancies. Uncorrelated between bins and W+jets and Z+jets but correlated between QCD and EW processes of the same bosonic variety. • Global 0.1% uncertainty correlated between all processes to represent residual experimental uncertainty.
ATLAS VBF Higgs→Invisible style	• Per bin uncertainties varying from 1.5-7% to represent the TF uncertainties per bin. Uncorrelated across bins but correlated between all V+jets processes. • 0.1-3% linearly varying uncertainty to represent residual modelling uncertainties. Correlated across bins, between EW W+jets and EW Z+jets and between QCD W+jets and QCD Z+jets, but uncorrelated between EW and QCD processes. • 5-10% uncertainties on the Z+jets processes to represent the uncertainties on the W+jets/Z+jets ratio. Uncorrelated across bins and between QCD Z+jets and EW Z+jets. • Global 0.1% uncertainty correlated between all processes to represent residual experimental uncertainty.
ATLAS Mono-jet style	• Global 1% uncertainty on all V+jets processes to represent overall TF error. Correlated between all V+jet processes. • 0.1-10% uncertainties on V+jet processes to represent residual modelling uncertainties, correlated across bins, between EW W+jets and EW Z+jets and between QCD W+jets and QCD Z+jets, but uncorrelated between EW and QCD processes. • 5-10% uncertainties on the Z+jets processes to represent the uncertainties on the W+jets/Z+jets ratio. Uncorrelated across bins and between QCD Z+jets and EW Z+jets. • Global 0.1% uncertainty correlated between all processes to represent residual experimental uncertainty.

Table 5.19: Summary of the different systematic uncertainty configurations applied on the V+jets background estimates in the VBF sensitivity study.

Results

The 95% CLs (expected) upper limits on the signal strength for the benchmark signal points for the different systematic uncertainty configurations and shape fit variables are shown in Table 5.20. A configuration in which only the (data) statistical⁵ plus luminosity uncertainties are included is also shown for reference. From these limits the first thing which is apparent is that none of the limits with systematic uncertainties included are below one, indicating that an analysis would not have sensitivity to any of these signals at 95% CLs level with a 139 fb^{-1} dataset. In particular, the limit for the WB (100,99) point using the CMS style uncertainty configuration with M_{jj} is around 1.6 which is significantly weaker than would be expected based upon the CMS result which has sensitivity to this point, particularly given that the study shown here uses a four times higher integrated luminosity⁶, and that the upper limit with only data plus luminosity uncertainties is only just below one. Reasons for this discrepancy will be discussed shortly after the results using the different systematic configurations and shape fit variables are compared.

Comparing the different uncertainty configurations, despite having rather different underlying uncertainties the CMS and ATLAS Higgs→invisible style configurations produce very similar limits which are significantly stronger than the mono-jet style setup. This is due to the effect of the residual modelling uncertainties in the mono-jet setup, which as a reminder are reasonably large in the high SR bins due to the use of a global V +jets transfer factor. These modelling uncertainties are in general much larger in the VBF channel than they are in the mono-jet analysis, hence why the binned TF approach is preferable here. This comparison suggests that adopting a configuration similar to that of the ATLAS Higgs→invisible analysis would provide the best sensitivity whilst being significantly more reliable and data-driven than the CMS style approach.

Comparing the different shape fit variables, it can be seen that the 8-bin and 12-bin M_{jj} shape fits produce very similar limits, highlighting that the uncertainty configurations tested are relatively insensitive to the precise binning being used. However, it can be seen that the shape fits in E_T^{miss} are able to provide a 10-20%

5 MC statistical uncertainties are *not* included in the statistical plus luminosity configuration.

6 Using a luminosity of 36 fb^{-1} the 95% CL_s upper limit on the signal strength using the CMS style uncertainty configuration on the WB (100,99) point is around 2.

Uncertainty Configuration	Signal	Shape fit variable			
		M_{jj} 12 bins	M_{jj} 8 bins	E_T^{miss}	BDT score
CMS VBF SUSY Style	WB (140, 139)	$3.09^{+1.24}_{-0.87}$	$3.03^{+1.21}_{-0.86}$	$2.62^{+1.05}_{-0.73}$	$1.81^{+0.73}_{-0.50}$
	WB (100, 99)	$1.64^{+0.65}_{-0.47}$	$1.62^{+0.65}_{-0.47}$	$1.48^{+0.60}_{-0.42}$	$1.04^{+0.43}_{-0.28}$
	HH (100, 99)	$4.29^{+1.74}_{-1.21}$	$4.25^{+1.74}_{-1.18}$	$3.50^{+1.41}_{-0.98}$	$2.66^{+1.06}_{-0.74}$
ATLAS Higgs $\rightarrow$ Invisible Style	WB (140, 139)	$3.10^{+1.23}_{-0.89}$	$3.07^{+1.26}_{-0.85}$	$2.69^{+1.07}_{-0.76}$	$1.86^{+0.74}_{-0.52}$
	WB (100, 99)	$1.61^{+0.65}_{-0.45}$	$1.64^{+0.65}_{-0.47}$	$1.50^{+0.61}_{-0.42}$	$1.06^{+0.43}_{-0.29}$
	HH (100, 99)	$4.17^{+1.71}_{-1.16}$	$4.35^{+1.76}_{-1.24}$	$3.56^{+1.43}_{-1.02}$	$2.66^{+1.06}_{-0.74}$
ATLAS Mono-jet Style	WB (140, 139)	$4.99^{+2.01}_{-1.39}$	$4.52^{+1.82}_{-1.27}$	$3.93^{+1.58}_{-1.10}$	$2.71^{+1.08}_{-0.76}$
	WB (100, 99)	$2.51^{+1.00}_{-0.70}$	$2.34^{+0.95}_{-0.66}$	$2.15^{+0.87}_{-0.60}$	$1.55^{+0.62}_{-0.43}$
	HH (100, 99)	$5.94^{+2.35}_{-1.70}$	$5.74^{+2.30}_{-1.61}$	$5.12^{+2.08}_{-1.42}$	$3.84^{+1.54}_{-1.07}$
Statistical + Luminosity only	WB (140, 139)	$1.77^{+0.71}_{-0.49}$	$1.77^{+0.71}_{-0.49}$	$1.53^{+0.61}_{-0.43}$	$1.06^{+0.43}_{-0.28}$
	WB (100, 99)	$0.94^{+0.38}_{-0.26}$	$0.94^{+0.38}_{-0.26}$	$0.86^{+0.35}_{-0.24}$	$0.60^{+0.24}_{-0.17}$
	HH (100, 99)	$2.49^{+0.99}_{-0.69}$	$2.48^{+1.00}_{-0.70}$	$2.04^{+0.81}_{-0.57}$	$1.52^{+0.60}_{-0.42}$

Table 5.20: 95% CL upper limits at 139 fb^{-1} on the signal strength for example signal points from shape fits performed in different kinematic variables with different systematic uncertainty configurations, as described in text.

stronger limit than the M_{jj} shape fits. Further, the fits performed using the BDT score improve upon the limits using E_T^{miss} by 25-30% for the different benchmark points, with the limit approaching one for the CMS and Higgs $\rightarrow$ invisible style uncertainty configurations. This improvement is also visible in the statistical plus luminosity only uncertainty configuration, indicating it is not a consequence of how the systematic uncertainties are set up. Hence, in a full analysis the use of ML classifier should be able to significantly improve the sensitivity, although it should be recalled that this study has assumed similar uncertainties for each of the shape fit variables; it is possible that when fully evaluated the shape uncertainties for the BDT score distribution may be higher than for M_{jj} or E_T^{miss} (for example, the uncertainty on the W +jets to Z +jets ratio if the Higgs $\rightarrow$ invisible style approach were to be adopted). Nevertheless, given the significant potential gains, the use of an ML classifier should be explored in a full analysis.

However, even in the fit using the BDT score, the limit that can be obtained for the WB (100,99) point is still above one, suggesting if this analysis were to be performed it would not have sensitivity above the limits set by the experiments at LEP. To help understand the lack of sensitivity compared to the CMS analysis, the M_{jj} distributions

of the WB (100,99) point and of the total (expected) SM background used in this study and in the CMS analysis are shown in Figure 5.20 (with the same binning as used in the CMS analysis). These plots have been made using the publicly available information from the CMS analysis for the CMS yields, and for the ATLAS yields by using the Monte Carlo production period associated with the data collected between 2015-2016 by ATLAS, corresponding to an integrated luminosity of 36 fb^{-1} , matching that of the CMS analysis. Total yields in SR-CMS are also shown in Table 5.21. Comparing the signal yields, the signal presented here, labelled as the ATLAS signal, is roughly half that of the CMS yield across all the M_{jj} bins. For the background, the total yields are more similar but at low values of M_{jj} the CMS background is around a factor of two greater than the ATLAS yield while in the last M_{jj} bin the ATLAS yield is around a factor of two larger.

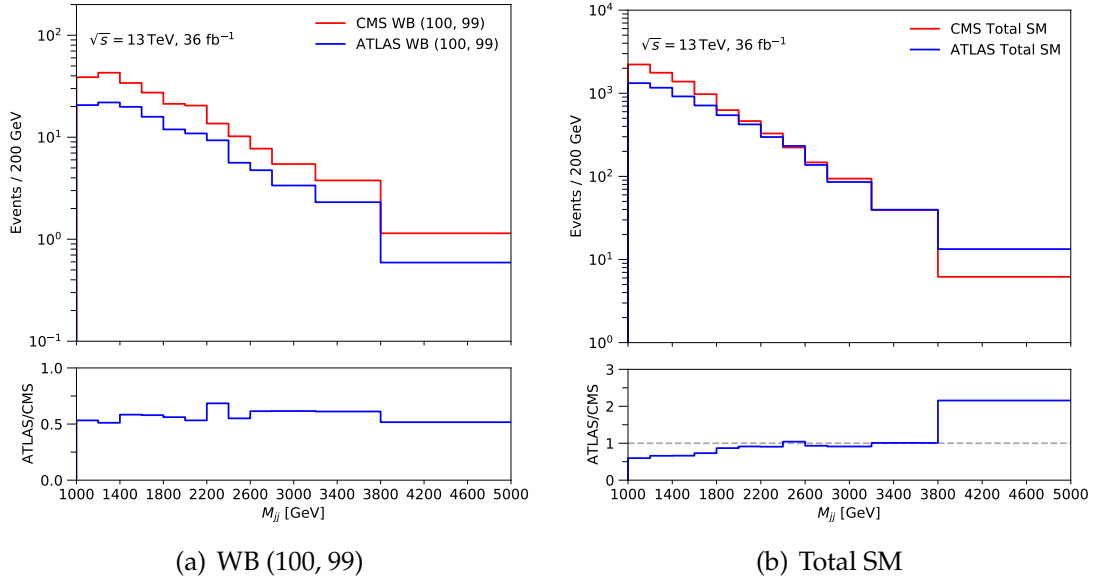


Figure 5.20: Comparison of the yields used in the VBF sensitivity study to those in the CMS analysis. Only the background estimate corresponding to the data collected between 2015-2016 is used for the ATLAS estimate, to match the integrated luminosity of the CMS analysis.

Considering first the discrepancies in the background yields, the lower background at low M_{jj} for ATLAS can, at least in part, be explained by the omission of the multi-jet background in this study. From the yields table, it can be seen that the W +jets background is also around 50% smaller for this study while the Z +jets yields are comparable; this is attributed to the improved hadronic tau-lepton veto used here rejecting a greater amount of W +jets events in which the W -boson decays to a hadronically

Process	CMS	ATLAS
Z+jets	3714 ± 760	3894.8 ± 34.1
W+jets	2999 ± 620	1890.8 ± 38.4
Multi-boson	77 ± 18	252.7 ± 7.3
Top	681 ± 128	84.7 ± 3.5
Multi-jet	546 ± 69	-
Total SM	8017 ± 992	6123.1 ± 52.1
WB (100, 99)	259	138

Table 5.21: Comparison of the yields in SR-CMS between the CMS analysis and the sensitivity study discussed here, labelled as ATLAS, at 36 fb^{-1} . Uncertainties on the CMS yields include the full systematic uncertainties evaluated in the CMS analysis, while the ATLAS uncertainties are only statistical.

decaying tau-lepton. Similarly, the smaller top background can be attributed in part to an improved b -jet veto. While the multi-boson background is around a factor of three larger in this study, its overall effect on the sensitivity is negligible as it constitutes less than 5% of the total background in even the high M_{jj} bins and so here is not of concern.

For the last M_{jj} bin, the difference in the background yields could be a result of the different MC generators used for modelling the V +jet backgrounds; the CMS analysis used leading order MadGraph for these backgrounds whereas the study presented here uses NLO SHERPA [2.2.11], with electroweak corrections. Although V +jets CRs are used in both this study and the CMS analysis to verify the modelling, the data statistical uncertainties in the final M_{jj} bins in the Z+jets CRs are approximately 25% and 50% respectively. Combined with including other systematic uncertainties, this may mean the discrepancy in the background yields in the last M_{jj} bin is in fact not that significant and instead highlights a limitation of using global transfer factors for estimating the V +jets background in this kind of analysis; that at high M_{jj} there is room for significant variation in the modelling for which the accuracy of the modelling in the CRs may be deemed acceptable due to the large data statistical uncertainties.

To understand the differences in the signal yields, discussions were held between the ATLAS and CMS analyzers. First, the signal cross-sections used were compared. Both analyses use MadGraph to generate the signals and compute the cross-sections. The cross-section used for the WB (100, 99) point in the CMS analysis is 288 fb , which is

around 50% greater than the cross-section used for the same sample in this study. This difference primarily arises due to different parton-level requirements being applied during the signal generation; the CMS analysis has looser requirements and hence the signals have an appropriately larger cross-section. Based upon studies performed by a collaborator, adjusting these parton level requirements looks to result in only a small possible increase in the SR yields of up to 10%.

Other differences between the ATLAS and CMS signal generation setups have also been identified (again by collaborators) including; slight differences in the wino/bino mixing of the electroweakinos, whether VBF diagrams with squarks are allowed during the generation, the value of the `bwcutoff` parameter (which controls what is considered to be on-shell in s-channel resonances) and the decay widths of different SUSY particles. These differences effect not only the signal cross-sections but also the shapes of kinematic variables. Combined with the adjusted parton level requirements, these differences can potentially explain why the yields from the CMS setup could be around 20% larger than with the ATLAS setup.

Some of these differences in setup between ATLAS and CMS are also present for other, non-VBF, signal topologies, but in these scenarios the differences have less of an impact than in the VBF case due to the VBF diagrams being in general more complex. For example, VBF diagrams producing charginos and neutralinos that involve intermediate squarks are possible and turn out to have non-negligible contributions to the cross-section even when the squark masses are much larger than the electroweakino masses. For these effects, there is not necessarily a preferable configuration (for usefully assessing the general sensitivity of the VBF channel to the UC-models compared to other channels). Instead, what they highlight is that the simplified model paradigm adopted for designing SUSY searches has certain complexities for VBF production modes.

The combination of the differences listed above can still not completely account for the overall factor of two difference between the ATLAS and CMS signal yields. However, an additional difference in the generation commands used has also been identified. In the CMS analysis, EW and QCD diagrams (discussed earlier in Section 5.3.2) were generated with separate commands in MadGraph and consequently, as confirmed by MadGraph developers, interference effects between the EW and QCD diagrams were not considered. These interference effects turn out to significantly decrease the signal yield in the SR like phase space. It is thought that this effect, combined with the previous differences, can potentially explain the total discrepancy

between this study and the CMS result, but more work is needed to confirm this. Due to the omission of these interference effects, the simplified model interpretations presented in the CMS analysis likely overstate the sensitivity of the VBF channel to the ultra-compressed electroweakino models.

To summarise this discussion, the signals used in this study originate from a valid generator configuration, but ambiguities in the setup mean that the SR yields could be around 20% larger by using an equally valid configuration. However, the yields cannot seemingly be as large as would be expected based upon the CMS analysis without the unphysical neglect of interference effects between QCD and EW production diagrams. Work by collaborators is ongoing to establish a common VBF signal configuration between ATLAS and CMS for use in future searches. Based upon the limits established in the sensitivity study shown in this section, a 20% increase in effective cross-section would imply that, were such a search to be conducted in ATLAS using a multivariate classifier and 139 fb^{-1} of data, it could potentially have sensitivity to wino/bino models with $\tilde{\chi}_1^0$ masses just above the LEP limit. However, without this boost in cross-section, no sensitivity to models above the LEP limit would be obtainable. It should also be recalled that the study shown here likely provides optimistic limits due to the omission of the multi-jet background and its uncertainties.

5.4 Comparison and Conclusions

To recap the results presented in this chapter, first in Section 5.2 a re-interpretation of the ATLAS mono-jet search was conducted to assess the channel's sensitivity to the UC-models described at the beginning of this chapter. The resulting expected 95% CLs upper limits on the higgsino scenario extends to models with a $\tilde{\chi}_1^0$ mass of 120 GeV, while the observed limit falls below the limits set by LEP. For the wino/bino model, the expected limit extends to $\tilde{\chi}_1^0$ masses of 200 GeV while the observed limit exceeds 140 GeV. These limits thus look to surpass those from the LEP experiments and provide the current best limits on these models, with the caveat that theoretical uncertainties on the signals were not included in the re-interpretation.

Second, in Section 5.3 a sensitivity study of the Vector Boson Fusion (VBF) channel to the UC-models was performed. Replicating the selections of a CMS analysis from which it was inspired, the sensitivity was found to be significantly worse than expected, falling below the LEP limits. The disagreement between the CMS analysis has been determined by collaborators to be primarily due to differences in how the signal samples were generated. For some of these differences, it is ambiguous as to which choice yields models that are more useful in assessing the sensitivity of the VBF channel to more general SUSY models with ultra-compressed charginos and neutralinos. Within these differences, the relevant production processes could result in signal yields 10-20% higher in the signal regions than reported in this study. However, it has also been established that interference between electroweak and QCD production diagrams was neglected in the CMS analysis, thus the limits it provides are likely optimistic for more general models. Additionally, the use of a BDT classifier was shown to have the potential to significantly improve the sensitivity of the VBF channel, although even with its use the upper limits on the models used in this study still fall (just) below the LEP limits.

Comparing the limits of the mono-jet re-interpretation and the VBF sensitivity study, it is safe to conclude that the mono-jet channel has significantly greater sensitivity than the VBF channel to the UC-models, even within the ambiguities of the VBF signal generation. Although the two studies shown here do not provide an 'apples-to-apples' comparison, the uncertainties used in the VBF study are based upon those in analyses using the same region of phase space, thus the VBF study presented here is thought to be a reasonably accurate assessment of the sensitivity. Further, comparing the limits for the common points in Table 5.2 and Table 5.20, it can be seen that the

mono-jet limits are stronger than the VBF limits that include only the statistical and luminosity uncertainties. The superior sensitivity of the mono-jet analysis is largely due to the increased statistics available; as has been shown the cross-sections are approximately an order of magnitude larger. The main benefit of these increased statistics is that, although for the signal models used here both the mono-jet and VBF SRs have a similar total signal to background ratio (S/B), the high SR bins in the mono-jet analysis have greater S/B than those in the VBF study, even when the BDT score is used as the binning variable in the VBF SR (20% versus 10% for the WB (100,99) point). While a higher S/B can be achieved in the VBF channel by cutting more tightly on the VBF BDT score, the resulting SR bin would have a much smaller number of events than the highest mono-jet bin (~ 10 versus ~ 200) and consequently have worse sensitivity due to the increased statistical uncertainties. The potential benefit of the VBF channel in providing additional discrimination power, and thus having a greater total S/B, appears to be in part hampered by the interference between the QCD and EW production diagrams reducing the yield in the VBF SR phase space, though further study would be necessary to confirm this. Additionally, the systematic uncertainties are smaller in the mono-jet channel than in the VBF channel ($\sim 4\%$ versus $\sim 10\%$ in the high SR bins). This is likely in part a further consequence of the increased statistics available in the mono-jet channel, but is also due to the VBF processes being in general more difficult to model accurately.

However, it is worth noting that potential improvements to the VBF analysis presented here are possible, on top of the potentially $\sim 20\%$ larger signal yields from changes in the signal generation. For example, correction factors on the V +jet backgrounds could be included during the classifier training and the discrimination power of the kinematics of additional, non-VBF tagged jets could also be used in the training, provided additional jets are modelled at matrix element level in the signals. While such improvements are unlikely to provide sufficient gain that the sensitivity of the VBF channel can exceed that of the mono-jet channel, it may become more competitive.

In such a scenario, it is then likely that a statistical combination of the VBF and mono-jet channels could notably improve the overall sensitivity to the UC-models. A complication of doing such a combination would be that, with the selections discussed here, the SRs of these analyses are not orthogonal to each other, with the VBF selections representing a small portion of the total mono-jet phase space. To combine the analyses this overlap would need to be removed. Such an overlap removal is done for the Higgs $\rightarrow$ invisible signature in ATLAS [169], and so could in principle also be done for

the UC-models⁷. Other search channels are also considered in the Higgs→invisible combination, such as one with a final state consisting of a Z-boson plus E_T^{miss} which targets the production of a Higgs-boson with a Z-boson. These channels should also have some sensitivity to analogous signal processes for the UC-models, both for the zero-lepton approach discussed here but also for the special track reconstruction approach discussed in the introduction. Due to having the largest statistics, it is perhaps reasonable to expect the mono-jet analysis to be able to set the strongest limit on the UC-models out of all the available channels. However, the sensitivity from other channels may be competitive, indeed a phenomenological study in [170] suggests the ‘mono-V’ channel, in which charginos and neutralinos are produced alongside a hadronically decaying W/Z-boson, could set a limit of 110 GeV on the higgsino models using the ATLAS Run 2 dataset, which is only 10 GeV below that of the mono-jet limit presented here. To maximize the overall sensitivity to these UC models, all of these channels should be re-interpreted and the results combined where possible. As will be demonstrated in next chapter, such combinations can lead to significant gains in the overall sensitivity.

⁷ Interestingly, it is worth highlighting that for the Higgs→invisible signature the VBF channel is significantly more sensitive than the mono-jet channel. This is due to the relative size of the Higgs boson couplings to the SM vector-bosons leading to the Higgs→invisible signal having an enhanced S/B in the VBF channel compared to the mono-jet channel, as well as benefiting from the enhanced discrimination power of the VBF topology.

Chapter 6

Combination of Searches for Chargino Pair Production

“Apes together strong!”

Caesar, Rise of the Planet of the Apes
(2011)

As discussed in Chapter 3, statistically combining the results of individual searches for new physics can yield a large gain in the overall sensitivity. This will be demonstrated in this chapter for chargino pair production, another possible signature of Supersymmetry, by combining the results of four previous ATLAS searches covering three different topologies, all using the Run 2 dataset. These searches were originally designed with the intention that this combination be performed, hence the results presented here are the culmination of significant previous effort by other members of the ATLAS collaboration. This is also the first time searches targeting chargino pair production in these topologies have been combined. The result has been made publicly available in a paper published by the ATLAS collaboration [2]. An overview of chargino pair-production and the searches that are combined is first given in Section 6.1, followed by technical details of the combination in Section 6.2 and a description of how systematic uncertainties are handled in Section 6.3. The results of the combination are then presented in Section 6.4 followed by a brief conclusion in Section 6.5.

6.1 Introduction

Chargino pair production is a possible signature of Supersymmetry and was one of the wino/bino subprocesses considered in the previous chapter on ultra-compressed charginos and neutralinos. The searches discussed in this chapter target this production mode for models with more general mass spectra, specifically $\tilde{\chi}_1^\pm$ masses between 100 GeV and 1000 GeV and $\tilde{\chi}_1^0$ masses between 0 GeV and 250 GeV, as charginos with masses below 100 GeV have been excluded by the experiments at the LEP collider [157]. In the simplified models used for these searches, the $\tilde{\chi}_1^\pm$ are purely wino, are produced in opposite sign pairs via a Z-boson and decay with 100% branching fraction to a W-boson and a bino $\tilde{\chi}_1^0$. A diagram illustrating chargino pair production is shown in Figure 6.1. Although in general for supersymmetric models in which there are charginos $\tilde{\chi}_2^0$ - $\tilde{\chi}_1^\pm$ pair production can also occur, there is no concrete prediction for the mass difference between the $\tilde{\chi}_1^\pm$ and the $\tilde{\chi}_2^0$. Hence, it is worthwhile to search for chargino pair production in isolation as it may be visible before $\tilde{\chi}_2^0$ - $\tilde{\chi}_1^\pm$ pair production.

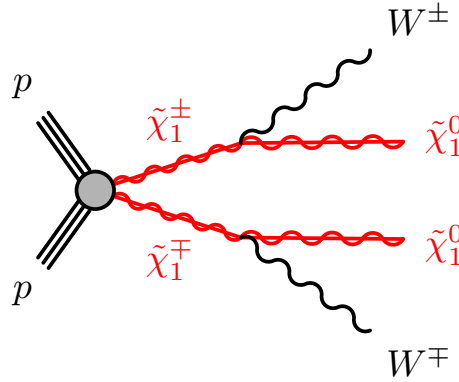


Figure 6.1: Feynman diagram of chargino pair production. The grey circle represents multiple diagrams in which a pair of opposite sign $\tilde{\chi}_1^\pm$ are produced from electroweak bosons produced by the proton-proton collision.

All final states targeting this model have significant E_T^{miss} from the $\tilde{\chi}_1^0$, but the different possible decays of the W-bosons give rise to different final states. In particular, whether one or both of the W-bosons decay leptonically gives three different final states; two leptons, one lepton plus hadronic activity or just hadronic activity. With the full Run 2 dataset, the ATLAS collaboration has performed searches targeting each of these final states¹. Specifically, two searches have been performed targeting the two

¹ For the leptonic decay modes, the searches include decays to electrons and muons as well as decays to tau-leptons which in turn decay leptonically to electrons or muons.

lepton final state with no significant hadronic activity (abbreviated as 2L0J), the first using an event selection based on E_T^{miss} and a shape fit in m_{T2} [171], while the second used BDTs to define SRs targeting mass splittings close to the W -boson mass [172]. Pertaining to the order in which these searches were done, the m_{T2} analysis is here referred to as the first wave 2L0J analysis, while the BDT analysis is referred to as the second wave 2L0J analysis. For the one-lepton (1L) final state, ATLAS has recently published a search [173] targeting chargino-neutralino and chargino pair production. Finally, a search targeting the all-hadronic final state [174] has also been performed. The 1L and all-hadronic analyses were the first searches in these topologies targeting chargino pair production to be performed at the LHC.

None of these searches saw any significant excesses compared to the SM predictions, hence each was able to set exclusion limits on the simplified model of chargino pair production for a range of $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_1^0$ masses. Overlays of the observed and expected exclusion contours for the individual analyses are shown in Figure 6.2, along with the limits from the LEP experiments. From this figure, it can be seen that the two 2L0J searches have the best sensitivity at low chargino masses, owing to the lower p_T thresholds of the lepton triggers and the cleaner experimental signature. At high masses, the all hadronic search has the best sensitivity due to the larger acceptance courtesy of the greater hadronic branching fraction of the W -boson decays. Having a mix of the above factors, the 1L search has the best sensitivity at intermediate $\tilde{\chi}_1^\pm$ mass. Notably, there is a gap in the observed limits between the 1L and all hadronic analyses for $540 \text{ GeV} < m(\tilde{\chi}_1^\pm) < 620 \text{ GeV}$ where models with massless $\tilde{\chi}_1^0$ are not excluded at 95% CL. The CMS experiment has also performed searches for chargino pair-production, but with the Run 2 dataset has only specifically targeted the all hadronic channel [175, 176]. Again no excess beyond the SM predictions was observed and the resulting 95% CL exclusion contour is shown in Figure 6.3, which covers $\tilde{\chi}_1^\pm$ masses between 300 GeV and 660 GeV.

The goal of statistically combining the Run 2 ATLAS searches is to further extend the exclusion limits beyond what can be achieved from the overlay the individual contours. The largest gain is likely to come at intermediate chargino masses where several of the analyses have comparable sensitivity. Before proceeding, it should be noted that the two 2L0J analyses overlap significantly, and so cannot be statistically combined. When performing the combination, the 2L0J analysis out of the two which has the best expected sensitivity, determined from the expected CL_s values, is used for each of the relevant mass points. As the second wave analysis only has the best

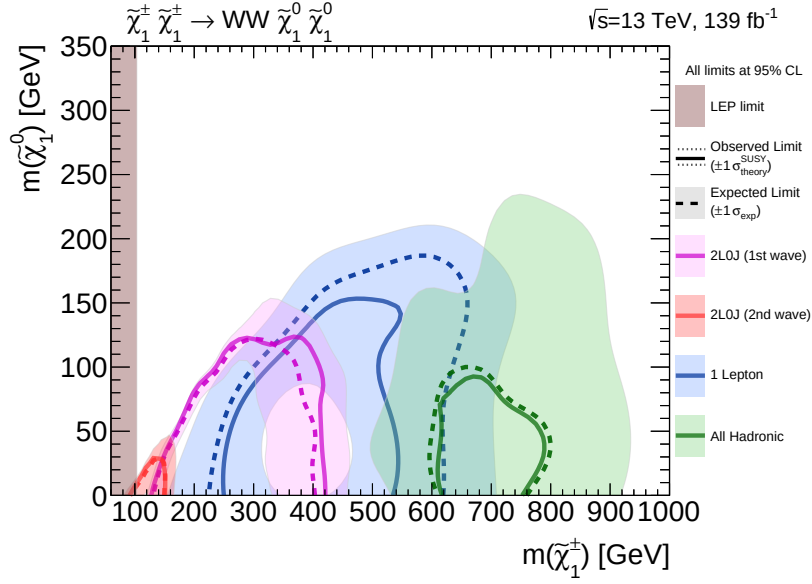


Figure 6.2: The original expected (dashed lines) and observed (full lines) exclusion contours for each of the analyses used in the combination. The limits from the LEP experiment are also shown.

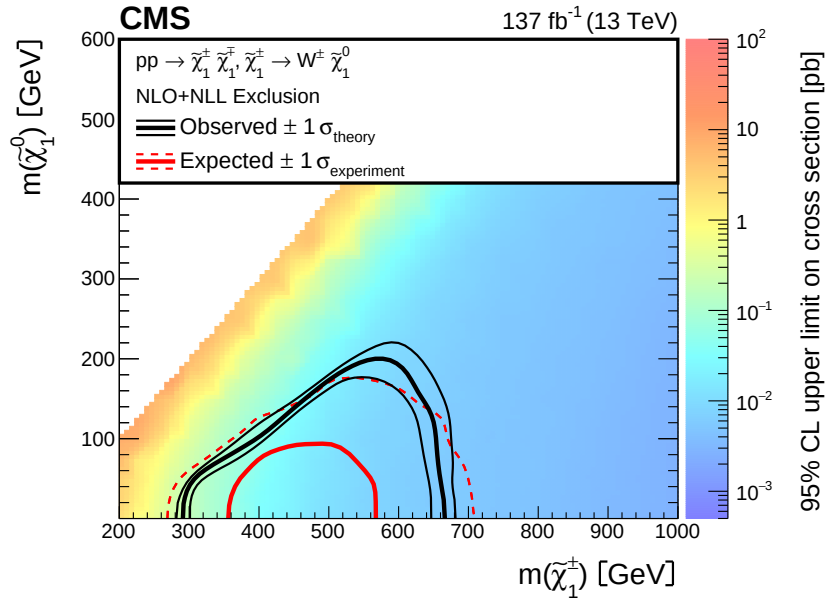


Figure 6.3: 95% CL exclusion limits from the chargino pair-production search performed by the CMS experiment using the Run 2 dataset, targeting the all hadronic final state, from [176].

sensitivity at $\tilde{\chi}_1^\pm$ masses less than 150 GeV, where neither the 1L nor all-hadronic analysis has significant sensitivity, in practice this means only the first wave 2L0J analysis is actually used in the combined fits.

6.2 Combination Details

Concerning the selections made in each analysis, the first wave 2L0J analysis features four signal regions (SRs), in each of which a shape fit in m_{T2} is performed, split by whether the leptons are of opposite or same flavour and whether there are one or two jets. This search also has separate control regions (CRs) for WW , VZ and top backgrounds. Similarly, the second wave 2L0J analysis has two SRs split by the flavour of the leptons, in each of which a shape fit in a BDT score is performed, and also has a top CR and a di-boson CR. The 1L analysis features three SRs targeting low, medium and high $\tilde{\chi}_1^\pm$ masses respectively, a combined W +jets and one-lepton di-boson CR, a 2-lepton di-boson CR and a top CR. For the all-hadronic analysis, in total 10 SRs and two CRs are defined but only the SR targeting the WW final state, and the CR used to estimate the background in this SR, are used for this combination.

To ensure these selections are orthogonal with each other, a common baseline lepton definition was adopted by all the analyses upon which to base their lepton multiplicity selections. As a result, the signal regions for the analyses are orthogonal by construction. Checks for overlaps between the control and signal regions between the different analyses were performed by collaborators, using both data and high statistic signal samples. No overlaps for the regions used in the combination were found.

The primary inputs for the combination are workspaces from the individual analyses, which detail the signal, background and data yields in each region, systematic uncertainties and which backgrounds have normalisation factors². The pyhf [85] software package was used to run the combined statistical fits, using Scipy [89] to minimize the negative log-likelihood in the fits. To have confidence in the results of the combination, it was important to first reproduce the results of the individual analyses. This is particularly important as pyhf and Scipy were not used for all the original fits, but are used for the combination fits due to being significantly faster than other software packages. All fits use the asymptotic approximation in the calculation of the CL_s values as all the analysis regions contain more than ten background events. The original expected and observed CL_s values for each of the analyses are shown in Figures 6.4 and 6.5 respectively, together with the differences from recomputing them using pyhf. Small differences are seen for several points, but these differences are in

² For the first wave 2L0J analysis, a RECAST [164] implementation of the search was produced by a collaborator in order to reproduce the workspaces.

general where CL_s values are large and so are not considered problematic³. These differences are mainly due to the difference in the software package used to minimize the negative log-likelihood; as mentioned for the recalculated values Scipy is used whereas the original fits were performed using MINUIT2 [88].

³ Note the large discrepancy at (700,1) for the 2L0J wave 1 analysis is also not problematic as this point is comfortably excluded by the all-hadronic analysis and so is not important for the combination. Note also that the recalculated CL_s values for the (300,0) and (200,50) points for the 1L analysis are missing because the fits for these points fail to converge. This is not problematic as these points are comfortably excluded by the first wave 2L0J analysis.

Chapter 6: Combination of Searches for Chargino Pair Production

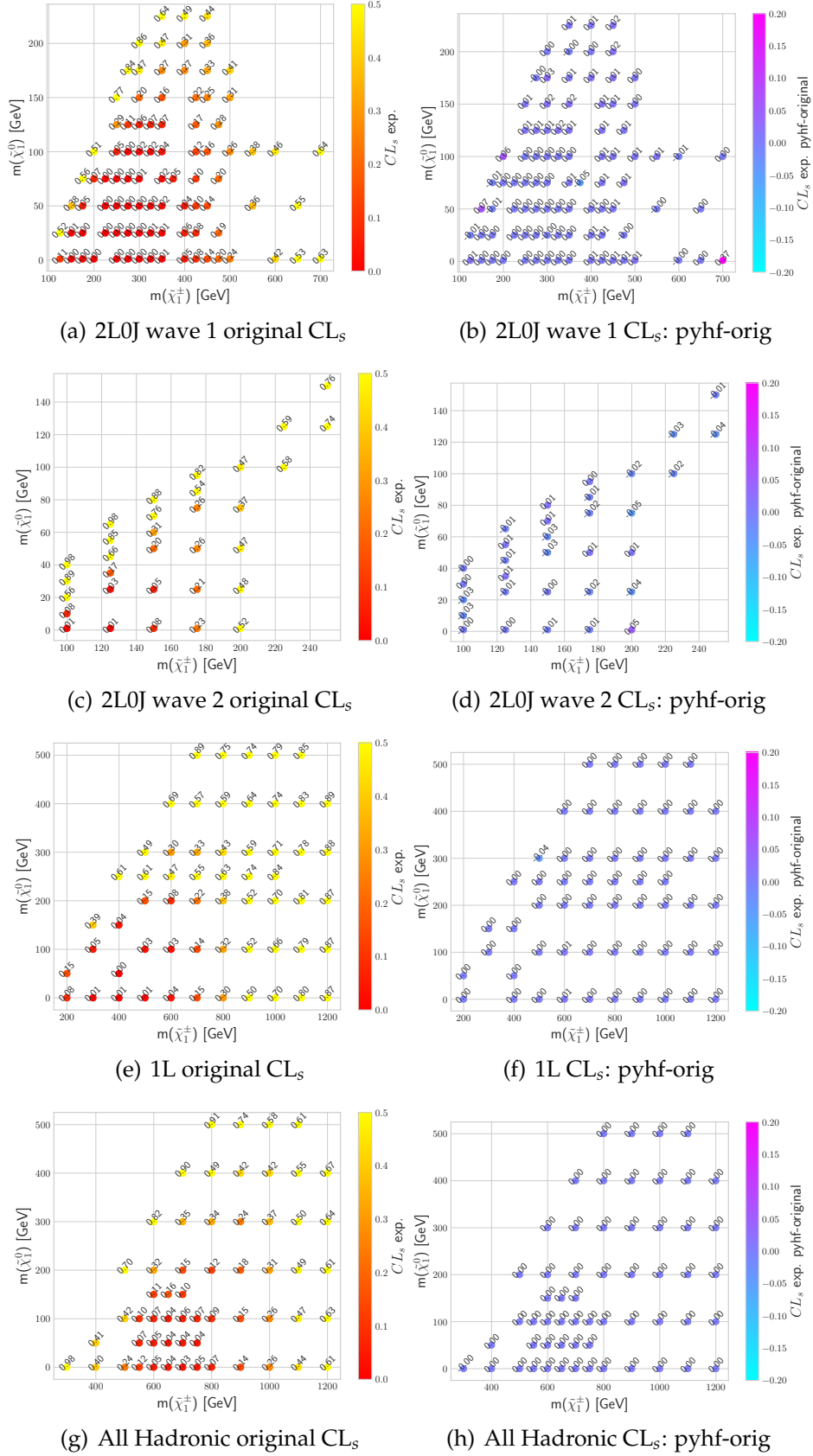


Figure 6.4: The original expected C1C1WW CL_s values for each analysis, and the difference of the recalculated values using pyhf.

Chapter 6: Combination of Searches for Chargino Pair Production

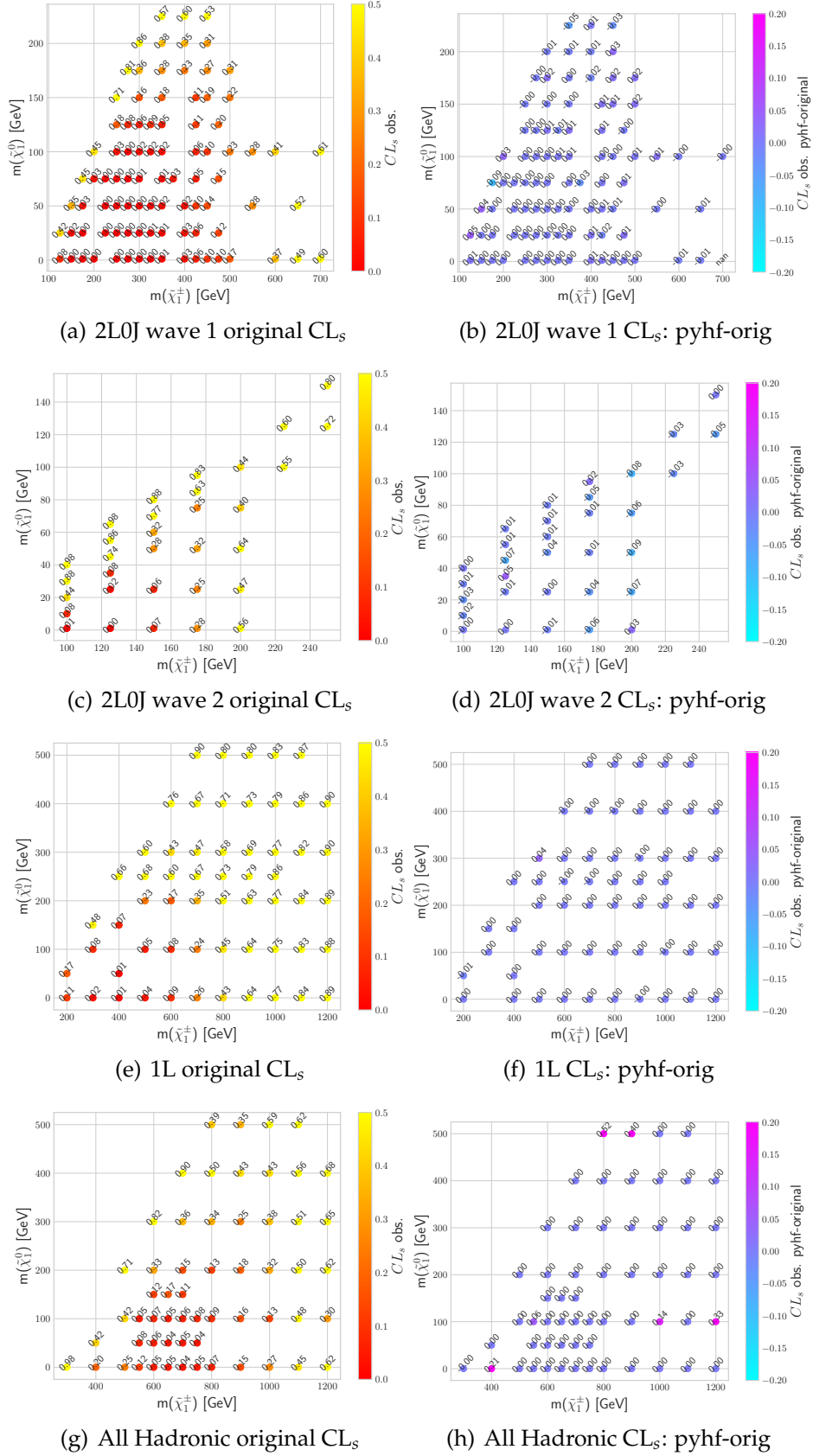


Figure 6.5: The original observed C1C1WW CL_s values for each analysis, and the difference of the recalculated values using pyhf.

As for the analyses in the previous two chapters, MadGraph was used for the event generation with NLO cross-sections taken from RESUMMINO v2.0.1 [112, 113]. As the individual analyses target different regions of the $m(\tilde{\chi}_1^\pm)-m(\tilde{\chi}_1^0)$ parameter space, workspaces for all $m(\tilde{\chi}_1^\pm)-m(\tilde{\chi}_1^0)$ points were not originally produced for all the analyses. Collaborators generated additional points for the combination that were then processed through the individual analyses⁴. The grid points available for each of the analyses, including these additional points, is shown in Figure 6.6(a). Including all points in the combination is not desirable; in particular points for which multiple analyses are expected to have sensitivity, but that have only been processed for a subset of the analyses, should not be included as such points would underestimate the sensitivity.⁵ The points which are included in the combination are shown in Figure 6.6(b). The majority of points that are not included in the combination lie on the high $\tilde{\chi}_1^\pm$ mass edge of the 2L0J wave 1 contour (around $m(\tilde{\chi}_1^\pm) = 400$ GeV) as they have only been processed for the 2L0J wave 1 analysis but the 1L analysis is expected to have sensitivity to them. Additionally, there are several other points that are not included in the combination as they have only been processed for one of the 1L or all-hadronic analyses where both are expected to contribute to the combination.

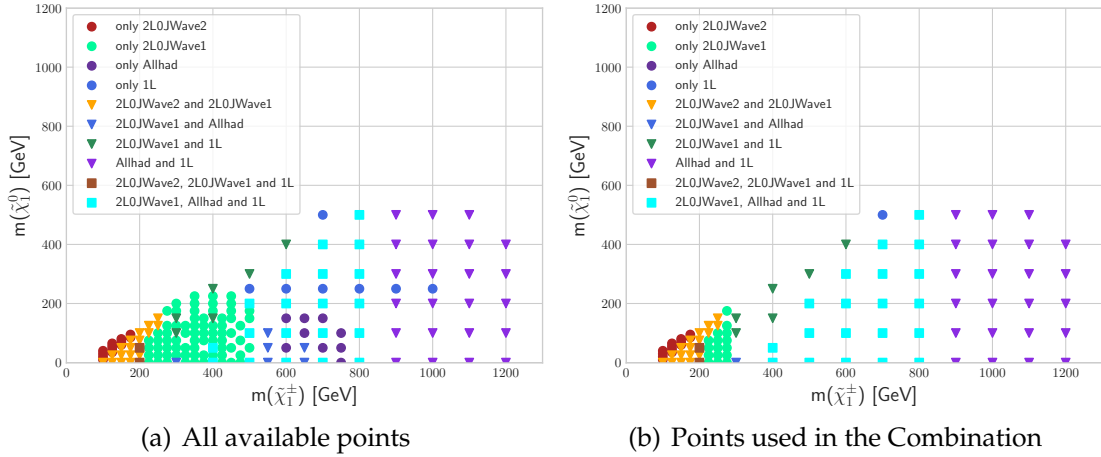


Figure 6.6: Points available for each of the analyses.

- 4 Note that different filters are applied during the MC generation to the signal samples that are evaluated by the different analyses to ensure good statistics. For example a two-lepton filter is applied to the samples used to evaluate the 2L0J analyses.
- 5 The optimal strategy of course would be to process all points for all analyses, but this would require generating a large amount of MC and is not necessary to establish the overall sensitivity of the combination.

6.3 Systematic Uncertainties

When performing combinations it is important to consider how systematic uncertainties are correlated between the different analyses. Many sources of systematic uncertainty are common between the searches considered here and their effect should in general be correlated across the different analysis regions where appropriate (not correlating the effects of uncertainties between analyses is not necessarily conservative). However, in practice correlating these systematic uncertainties can be difficult if the analyses used different methods for estimating the same underlying sources of uncertainty, resulting in a different number of nuisance parameters (NPs), or NPs with different meanings. For the searches considered here, the overall impact of systematic uncertainties on the analyses' sensitivities are in general comparable to the impact of statistical uncertainties, though for points that lie on the edge of the 95% confidence level (CL) exclusion limits the statistical uncertainties are in general dominant. Hence overall, the effect of correlating or not-correlating uncertainties across the analyses is not expected to have a big impact on the combination, but nevertheless the impact should be tested.

To assess the impact of correlating the uncertainties, the combined fits were run both with all NPs uncorrelated between analyses and then with NPs correlated where possible. In general, it is desired to use the approach outline in Chapter 3, Section 3.1.7 whereby experimental uncertainties are correlated between different SM processes and across regions while theoretical uncertainties are correlated across regions but uncorrelated between processes. In both correlation schemes, normalisation factors on the same background processes are not correlated across the different analyses as their value can differ depending on the area of phase space.

In the correlated where possible scheme, the luminosity, lepton, pileup re-weighting, E_T^{miss} , and signal cross-section uncertainties were correlated across all the analyses. Concerning uncertainties associated with small radius jets, the matter is more complicated as the analyses used different configurations of NPs for uncertainties associated with the Jet Energy Resolution (JER) and Jet Energy Scale (JES). Only where the configurations are the same can the NPs be safely correlated. The names of different JER/JES uncertainty configurations used by the analyses, are shown in Table 6.1 alongside a column which indicates whether the JES or JER uncertainties can be safely correlated. The first part of the naming of these uncertainty configurations describes the number of NPs used to describe the JES uncertainties, with the "CategoryReduction" having

30 NPs, while the “GlobalReduction” has 20 NPs and hence models the uncertainties with less precision. The second part of the naming refers to the number of NPs used to describe the JER uncertainties, “FullJER” having 13 NPs and “SimpleJER” having 8 NPs.

As both the 2nd wave 2L0J analysis and 1L analysis use the CategoryReduction_SimpleJER systematic configuration, their JES and JER uncertainties can be fully correlated. The first wave 2L0J analysis uses the CategoryReduction_FullJER so only the JES uncertainties can be correlated with 1L. The use of the R4_GlobalReduction_SimpleJER by the all-hadronic analysis means that neither the JER or JES uncertainties can be correlated with any of the other analyses.

	JES/JER scheme	Correlate	
		JES	JER
2L0JWave1	CategoryReduction_FullJER	✓	×
2L0JWave2	CategoryReduction_SimpleJER	✓	✓
1L	CategoryReduction_SimpleJER	✓	✓
All had	GlobalReduction_SimpleJER	×	×

Table 6.1: The Jet Energy uncertainty scheme used by each analysis for the C1C1WW combination and an indication of what can be correlated.

Some theoretical uncertainties are also shared between analyses but the legitimacy of correlating them is less straightforward on account of different (versions of) MC generators being used to model the same background processes across different analyses. The analyses in general also target different regions of phase space, thus in any case correlating these uncertainties may not be appropriate. To test the possible impact, the following theoretical uncertainties are correlated in the correlated where possible scheme: the factorisation, renormalization, PDF, matching (CKKW) and resummation uncertainties of the Z+jets and W+jets backgrounds (for the all-hadronic and 1L analyses) and the di-boson renormalization, factorisation and PDF uncertainties (between the 2L0J and 1L analyses).

Figure 6.7 shows the relative difference in the expected upper limit on the signal strength from the combined fits with and without the uncertainties correlated. The quality and stability of these fits was verified by a collaborator. For the majority of points, the impact of correlating systematic uncertainties is negligible, although a handful of points around $(m(\tilde{\chi}_1^\pm) = 400\text{-}500 \text{ GeV}, m(\tilde{\chi}_1^0) = 200 \text{ GeV})$, where the

impact of the combination is greatest, do see around a 5% reduction in the limit (thereby improving the sensitivity) when correlating the systematic uncertainties. This improvement originates from marginally increased constraints on several nuisance parameters (NPs) when performing the combined fit. This can be seen in Figure 6.8 which shows a comparison of NP constraints from performing the combined fit with correlated uncertainties compared to the fits from the individual analysis, for the $m(\tilde{\chi}_1^\pm) = 500$ GeV, $m(\tilde{\chi}_1^0) = 200$ GeV point. As illustration on how to interpret this figure, the first systematic, “syst_VV_PDF”, which describes the PDF uncertainty on the di-boson background, has a slightly tighter constraint from the combination (shown in red), compared to the constraint coming from the 1L analysis alone (shown in black), as it is correlated across the 1L and 2L0J wave 1 analyses. Only NPs that see a difference in their constraints greater than 1% are included in Figure 6.8, demonstrating that most NPs are not significantly impacted by being correlated.

The NP pulls on fits to observed data are also examined to check if correlating the systematic uncertainties significantly changes the pulls compared to the individual analysis fits. As an example, a comparison of the observed pulls for the same $m(\tilde{\chi}_1^\pm) = 500$ GeV, $m(\tilde{\chi}_1^0) = 200$ GeV mass point is shown in Figure 6.9. Only NPs for which the pull has changed by more than 0.05 are shown. For this mass point, a significant change in the pull of the “MUON_ID” NP, which describes an uncertainty in the muon identification algorithm, is seen. Consequently, it would not be recommended to correlate this NP between these two analyses.

Given the small differences in the CL_s values when correlating the systematic uncertainties, and for harmony with other SUSY analysis combinations, it was ultimately chosen to not correlate any systematic uncertainties aside from the luminosity and signal cross-section uncertainties for the final results. The comparison of the expected CL_s values between the correlation schemes discussed here indicates that this choice leads to slightly conservative limits for the points which stand to gain the most from the combination, but that the overall impact from correlating uncertainties is small.

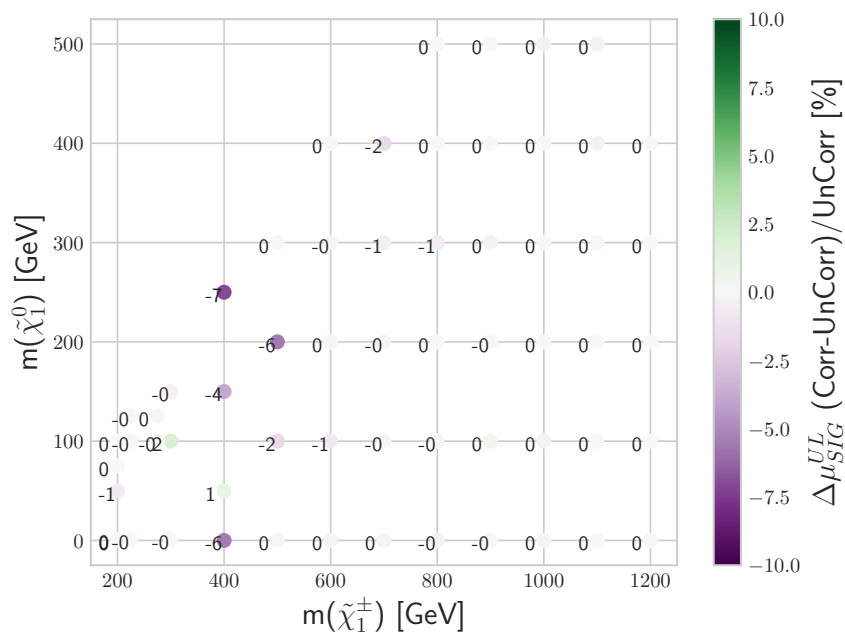


Figure 6.7: The relative difference in the expected upper limit on the signal strength when correlating (Corr) vs not-correlating (UnCorr) systematic uncertainties across the different analyses. Negative values indicate where the correlated limit is lower, and thus where there is a sensitivity gain from correlating systematic uncertainties. Points with $\tilde{\chi}_1^\pm$ masses less than 300 GeV are not shown as they are only evaluated for the 2L0J analyses, hence they are not effected by the choice to correlate uncertainties.

Chapter 6: Combination of Searches for Chargino Pair Production

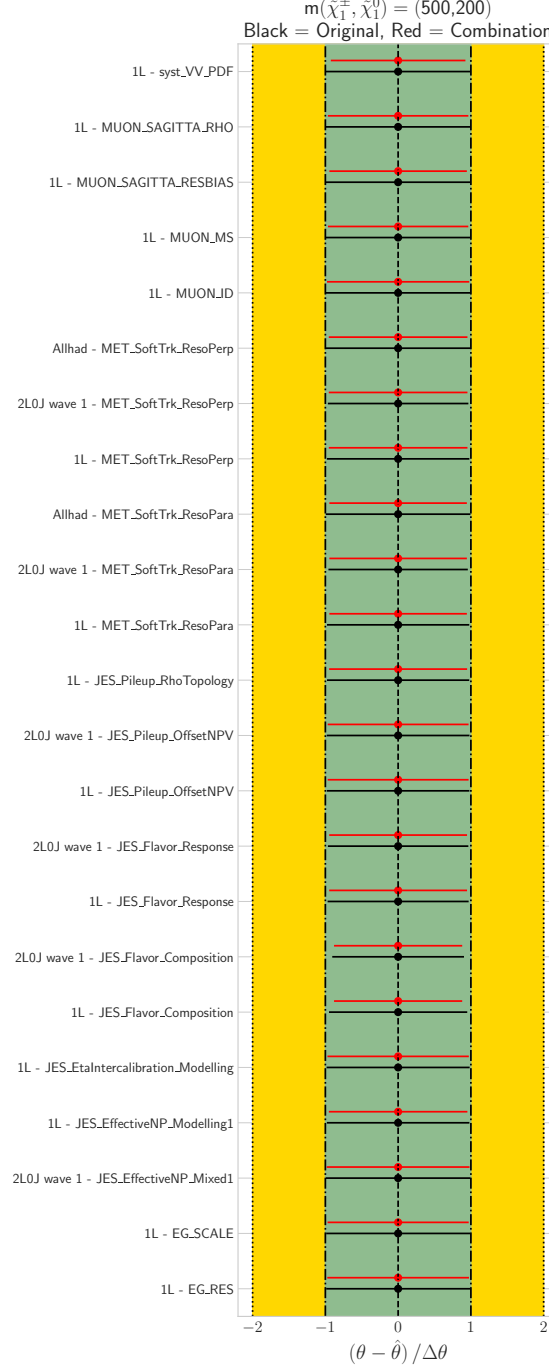


Figure 6.8: Comparison of the NP constraints from the expected combined fit for $m(\tilde{\chi}_1^\pm, \tilde{\chi}_1^0) = (500, 200)$ GeV. Points in red show the fitted value of the NPs with their errors from the combined fit, while points in black show the fitted value for the same NP from the fit of one of the analyses, which analysis being indicated by the left-hand label. Only NPs which exhibit a change greater than 1% are included. No pulls are seen due to this being a fit using expected data yields.

Chapter 6: Combination of Searches for Chargino Pair Production

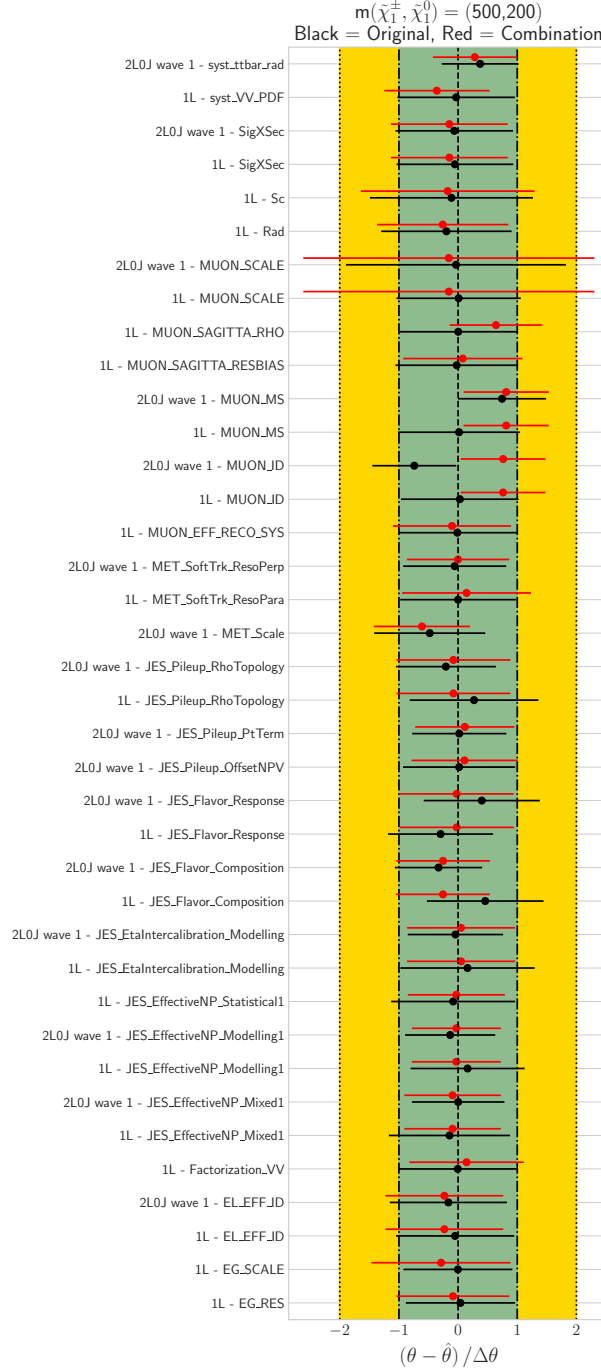


Figure 6.9: Comparison of NP constraints from the observed combined fit for $m(\tilde{\chi}_1^\pm, \tilde{\chi}_1^0) = (500, 200)$ GeV, following the same colour and labelling scheme as Figure 6.8. Only NPs which exhibit a change greater than 1% are included.

6.4 Results

The expected and observed exclusion contours from the combined fits are shown in Figure 6.10, overlaid with the contours from the individual analyses for comparison. Significant gain from the combination is evident. In particular, it can be seen that the improvement from the combination is greatest for $\tilde{\chi}_1^\pm$ masses around 700 GeV, where the expected reach in $\tilde{\chi}_1^0$ mass is extended by 100 GeV due to both the 1L and all-hadronic analysis having significant sensitivity. The gap in the observed limit between the 1L and all hadronic analyses is also closed due to the combination. A notable improvement in sensitivity can be seen at $\tilde{\chi}_1^\pm$ masses around 350 GeV as well, primarily from the combination of the 2L0J and 1L analyses. To further demonstrate the gain from the combination, Figure 6.11 shows the relative change in the expected upper limit on the signal strength between the combined fit and the otherwise best expected analysis. From this figure it can be seen that the upper limit on the signal strength is improved by up to 30%. Upper limits on the cross-section resulting from the combination are shown in Section A.4 of the appendix.

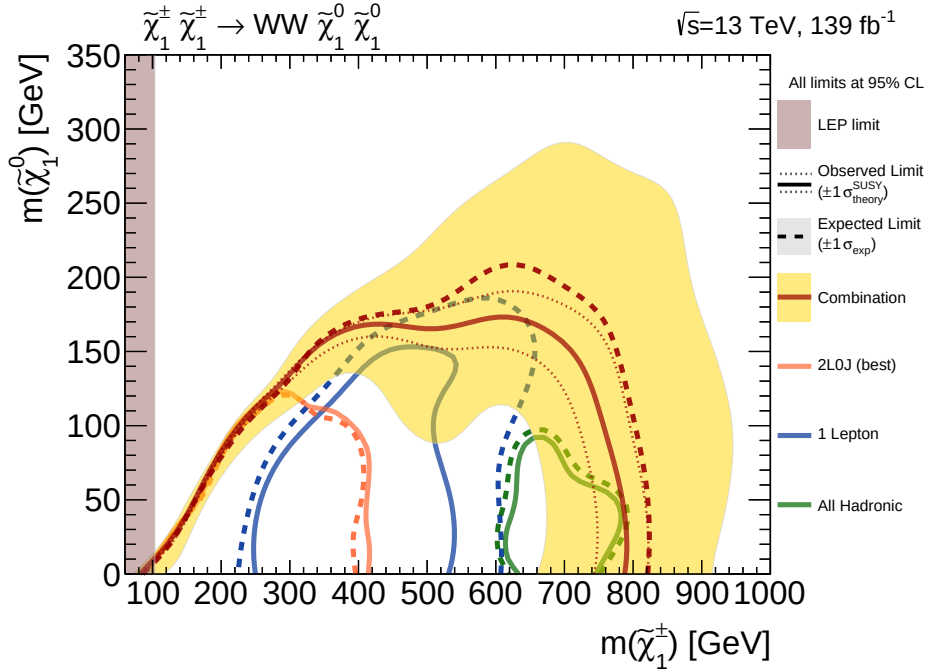


Figure 6.10: Expected and observed 95% CL exclusion contours from each individual analysis, and from the statistical combination. Uncertainty bands are shown only for the combination.

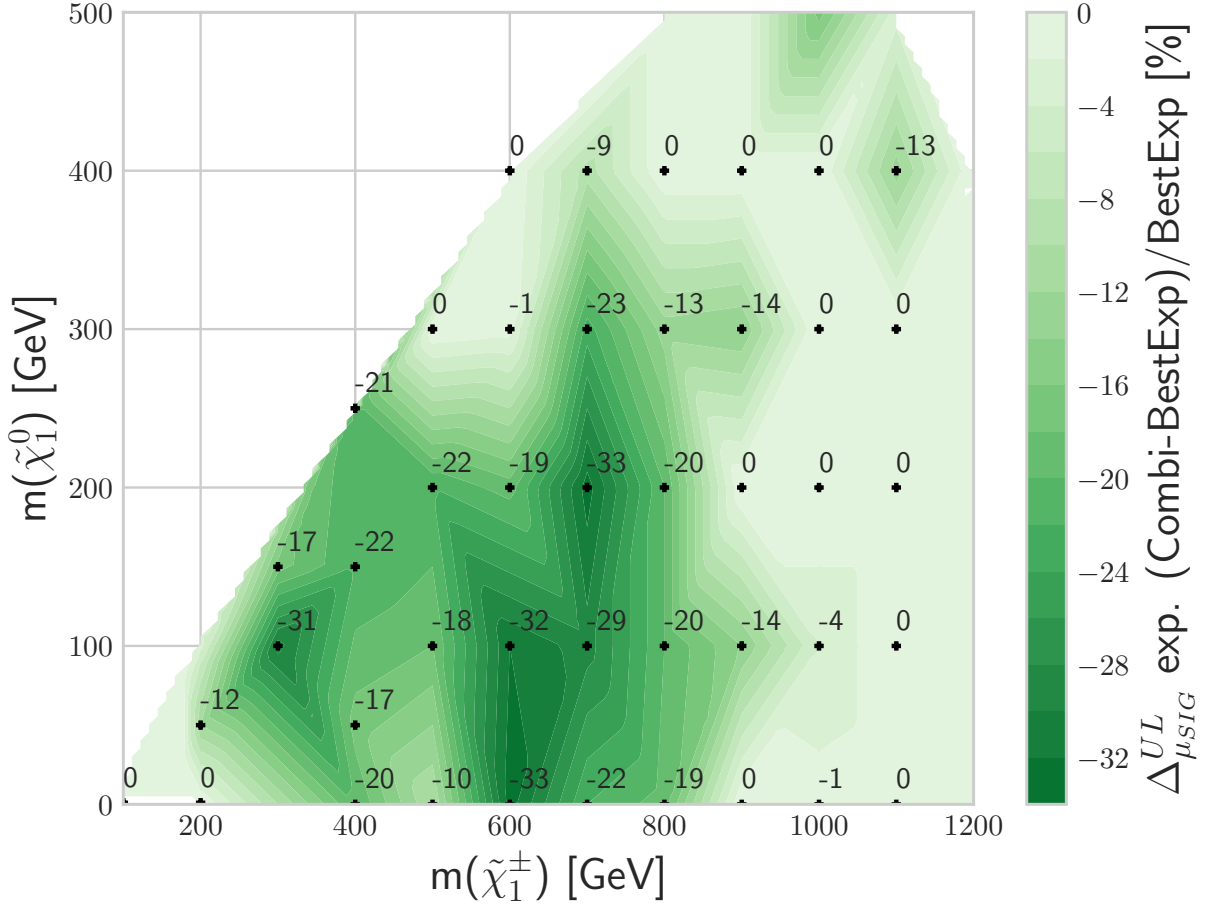


Figure 6.11: The relative difference in the expected upper limits on the signal strength between the combination and best expected analysis for each point. The shading is constructed using all the points used in the combination, of which only a subset are labelled.

6.5 Conclusion

In this chapter, the results of four previous ATLAS searches for chargino pair-production have been statistically combined for the first time. A significant gain in the overall sensitivity to this production mode results from the combination, particularly for chargino masses between 350 GeV and 700 GeV, for which models with $\tilde{\chi}_1^0$ masses of up to 160 GeV are excluded. The expected limits presented here are the strongest for all values of $\tilde{\chi}_1^\pm$ mass, while the observed limits are the strongest for all values chargino mass except for masses between 550 GeV and 650 GeV, for which the observed limits from the CMS all hadronic search (Figure 6.3) are marginally stronger.

Conclusion

In this thesis, three strategies for extending the reach of searches for staus, charginos and neutralinos with the ATLAS experiment at the Large Hadron Collider have been presented. Firstly, in Chapter 4 a new search for direct stau production utilising machine learning was detailed. No significant excesses were seen compared to the Standard Model (SM) predictions but new, world-leading limits were set on the simplified model used to design the search. In particular, the sensitivity at 95% confidence level (CL) to mass degenerate $\tilde{\tau}_L$ and $\tilde{\tau}_R$ was extended down to stau masses as low as 80 GeV, closing a sensitivity gap between previous ATLAS limits and those set by experiments at the Large Electron Positron (LEP) collider. At high $\tilde{\tau}$ masses the sensitivity extends to 500 GeV, 100 GeV greater than the previous ATLAS search using the same dataset. The depth of the sensitivity was also improved such that this new search was able to set the first 95% CL limits at the ATLAS experiment on $\tilde{\tau}_R$ production, excluding models for which $100 < m(\tilde{\tau}_R) < 350$ GeV. A large portion of the improved sensitivity is due to the machine learning based event selection. Further sensitivity gains are likely possible using this approach with increased statistics for training more complex models. Additionally, combining this search channel, in which both of the tau-leptons decay hadronically, with future searches in which one of the tau-leptons decays leptonically, will also improve the sensitivity for models with smaller mass differences between the $\tilde{\tau}$ and $\tilde{\chi}_1^0$.

Secondly, two new search channels for ultra-compressed charginos and neutralinos were explored in Chapter 5. A re-interpretation of a previous ATLAS search for physics beyond the Standard Model in a final state with a highly energetic jet and large missing transverse momentum was shown to have significant sensitivity to simplified SUSY models with ultra-compressed charginos and neutralinos. For the wino/bino scenario, the observed limits set by this search look to be able to exclude, at 95% CL, models with $\tilde{\chi}_2^0$ masses as high as 140 GeV, which have not been excluded by any previous analysis. For the higgsino scenario, the expected limit closes the gap in sensitivity between previous ATLAS searches and the LEP limit for mass splittings

Conclusion

between 0.4 and 3 GeV, though the observed limit falls below that of the experiments at LEP. Given the sensitivity demonstrated by this study, the mono-jet search should be included in reinterpretations of ATLAS searches for more general SUSY models with ultra-compressed charginos and neutralinos.

A sensitivity study of the Vector Boson Fusion (VBF) channel to the ultra-compressed models was also performed in Chapter 5. The VBF channel was found to have significantly less sensitivity than the mono-jet channel discussed above, not having sensitivity at 95% CL to models not already excluded by the LEP experiments, even when employing a BDT for enhanced background reduction. The channel was also found to be less sensitive than a similar analysis performed by the CMS experiment due to differences in the signal generation and resulting signal models used to assess the sensitivity. At present, some ambiguity remains as to what generation setup yields models that are more useful in assessing the general sensitivity of the VBF channel to the production of ultra-compressed charginos and neutralinos, with the signals used here representing a conservative case. Thus, the possibility of combining the VBF channel with the mono-jet channel and indeed other search channels was discussed but its execution is left to future work.

Lastly, in Chapter 6 the results of previous ATLAS searches for chargino pair production in two lepton, one lepton, and all hadronic final states were statistically combined resulting in a significant extension of the total sensitivity reach to this process. For $\tilde{\chi}_1^\pm$ masses around 700 GeV, the expected reach in $\tilde{\chi}_1^0$ mass was extended by 100 GeV and the upper limits on the signal strengths were increased by up to 30%.

As a closing note, it is worth recalling that the searches presented in this thesis make use of the data collected by the ATLAS experiment between 2015 and 2018 which represents only a small fraction ($\sim 5\%$) of the total data that will be collected by the experiment across the operational lifetime of the Large Hadron Collider. Accordingly, more searches for sparticles are certain to be performed in the years to come which will make use of this additional data. This likely includes future iterations of the searches for staus, charginos and neutralinos via the production modes discussed here. Having demonstrated their effectiveness for these SUSY signatures, the strategies detailed here should be reused and further developed when performing these searches in order to maximize their sensitivity reach.

Appendix

A.1 Breakdown of Authors Contributions

For clarity, a breakdown of the authors contributions to the work presented in Chapters 4 (search for direct stau production), 5 (searches for ultra-compressed charginos and neutralinos) and 6 (combination of searches for chargino pair production) are detailed below.

A.1.1 Search for Direct Stau Production

The author was responsible for the following aspects of the analysis:

- Designing the signal clustering for the BDTs which enabled the search to obtain the first sensitivity to right-handed only stau production.
- Training and optimization of all BDTs, implementing cross evaluation to ensure predictions are valid.
- Adapting the multi-jet background estimate to be done simultaneously with the normalisation of other background processes, which led to better agreement with the Standard Model predictions in relevant validation regions
- Proposed and designed the Inclusive Validation Region, which significantly enhances the reliability of the search.
- Addition of a Z+jets Control and Validation Region, on the suggestion of collaborators.
- Refinement of all previously defined Control and Validation regions.
- Re-evaluation of signal and background theoretical uncertainties.

- Re-processing of ntuples for the direct stau channel to fix bugs, adapting code originally written by a collaborator.
- Request of higher statistic signal samples for the low and compressed mass models, which resulted in improved sensitivity in this region of parameter space.
- Setting up, running and validating all statistical fits, including unblinding of the signal regions.
- Making all final post fit kinematic distributions and exclusion contours.
- SimpleAnalysis [163] implementation and computation of quantities for HEP-Data.
- Presenting a poster on the direct stau channel at the LHCP 2023 conference.

A.1.2 Searches for Ultra-compressed Charginos and Neutralinos

For the Vector Boson Fusion sensitivity channel, the authors main contributions were:

- Development of analysis framework.
- Collating background Monte Carlo samples.
- Initial ntuple productions.
- Preparing signal sample requests.
- Comparisons of MadGraph, Sherpa-2.2.1 and Sherpa-2.2.11 V+jets Monte Carlo samples.
- Fit studies comparing the sensitivity of binning in missing transverse energy, dijet invariant mass and the BDT score.
- Studies showing the potential gain of using the kinematics of additional jets for training the BDTs and dividing the signal region by jet multiplicity.
- Studies to help identify the origin of the discrepancy with the CMS analysis.
- Preliminary testing of a ‘double ratio’ method for estimating the V+jet backgrounds.

For the re-interpretation of the Mono-jet search, the authors main contributions were:

- Request and validation of generated signals.
- Computation of signal cross-sections to be used for the RECAST based upon the inclusive cross-sections.
- Preparing inputs for and running of the RECAST.
- Studies assessing the viability of performing a combination with the VBF channel.

A.1.3 Combination of Searches for Chargino pair-production

For the combination of searches for chargino pair-production, broadly, the author was responsible for recalculating the CL_s values for the individual analyses, selecting the grid points to be included in the combination and assessing which systematic uncertainties could be correlated between the analyses while a collaborator ran most of the final combinations, assessed the stability of the statistical fits, and setup the final plots, but there was substantial overlap between these responsibilities.

A.2 Additional details for the Direct Stau Search

This section of the appendix provides additional details on the search for direct stau production presented in Chapter 4.

A.2.1 Signal and Background Shape Comparisons

Figure A.1 show the shapes of the BDT input kinematic variables for example signal points and the total SM background used in the training. These plots are for events passing the nominal BDT training preselection, detailed in Table 4.3.

Appendix

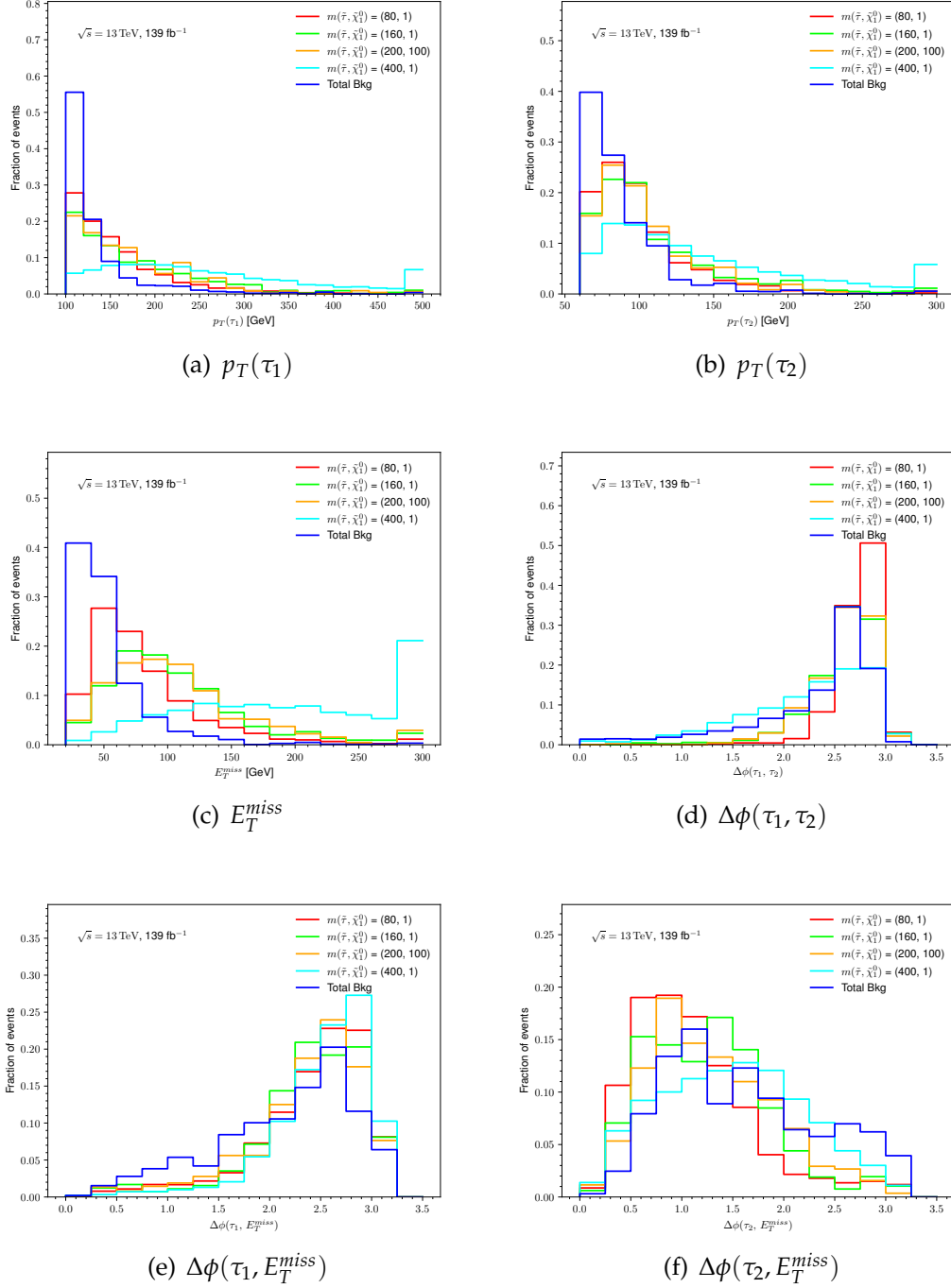


Figure A.1: Shape comparison of total background and example signal point passing the BDT preselection cuts.

Appendix

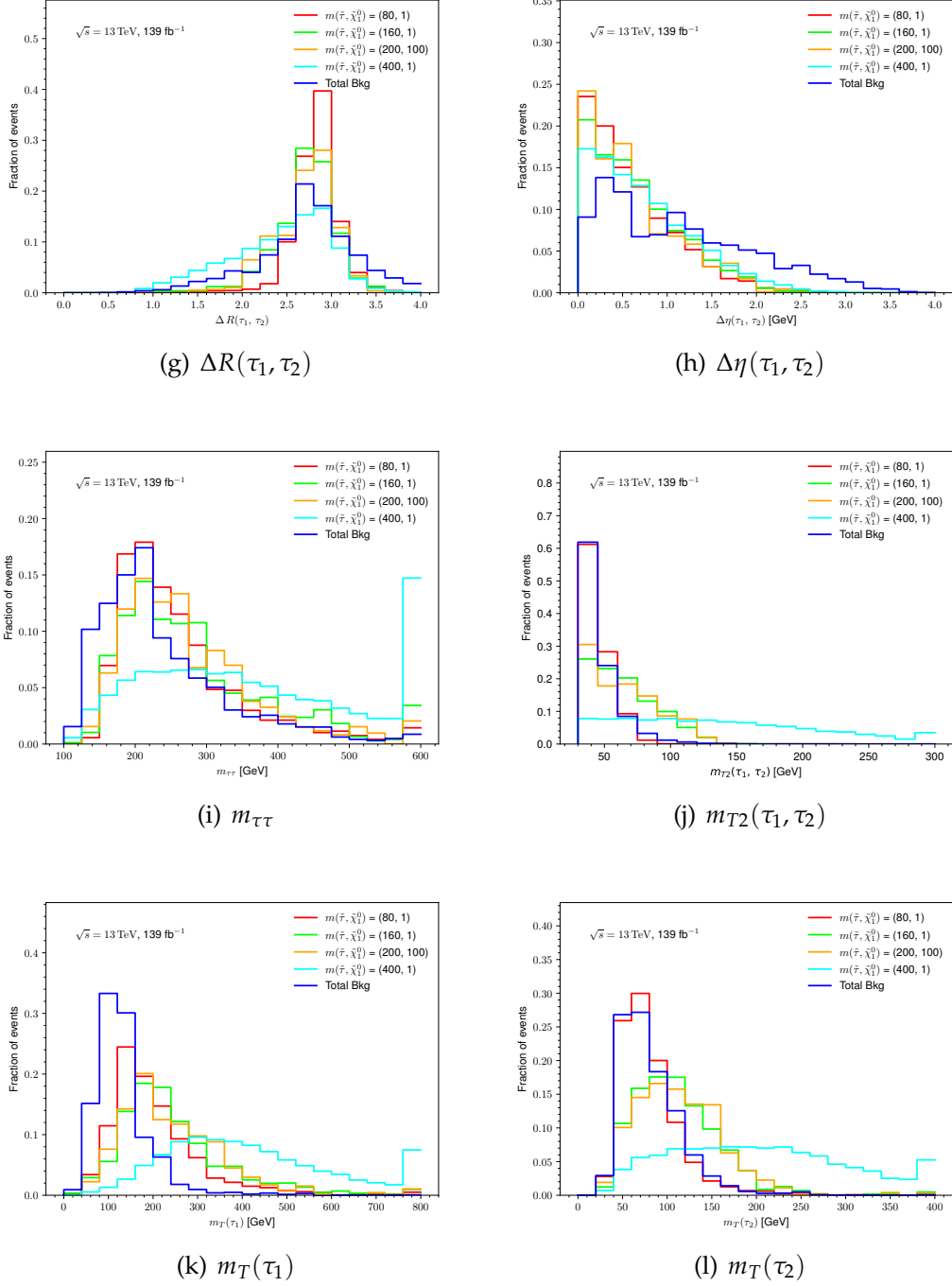


Figure A.1: Shape comparison of total background and example signal point passing the BDT preselection cuts.

Appendix

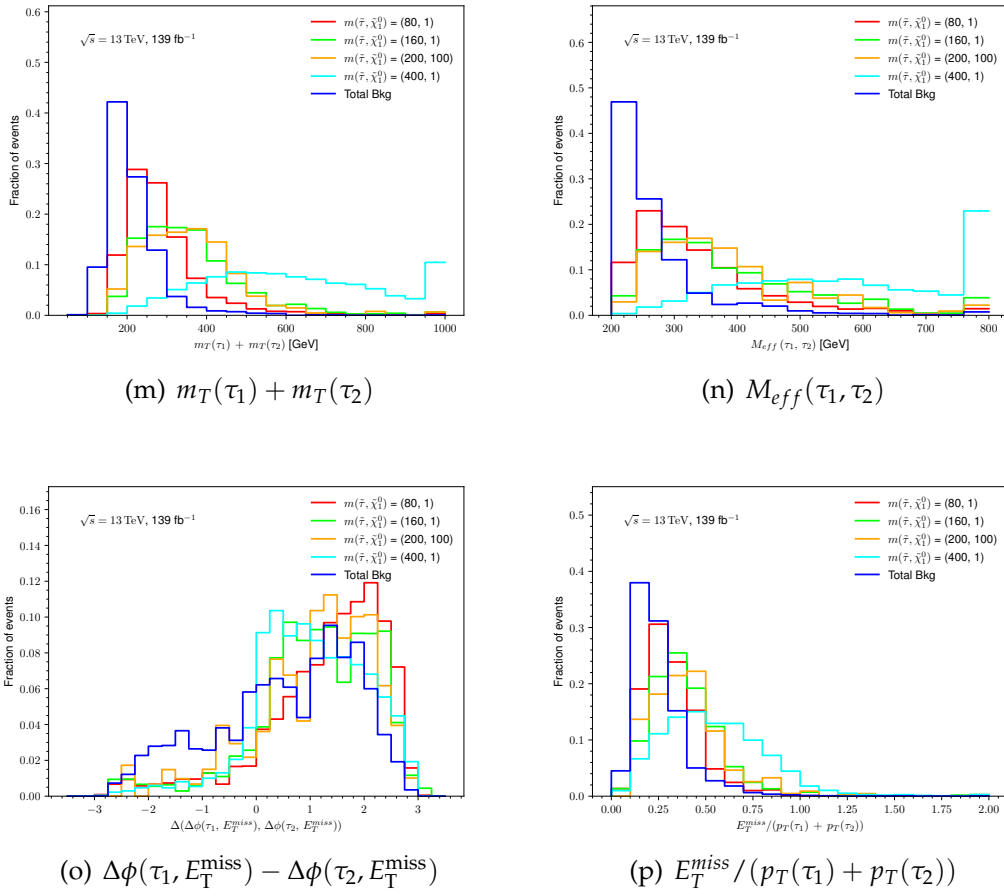


Figure A.1: Shape comparison of total background and example signal point passing the BDT preselection cuts.

A.2.2 Additional BDT Diagnostics

Comparisons of BDT score distributions for training and evaluation sets for the other fold splittings for each of the four signal groups are shown in Figures A.2 and A.3. The same notes as for the distributions for fold 1 in the main text (Figure 4.4) apply. Although for some models the test background prediction is greater than the prediction for the training set, the disagreement is not significant according to the chi-squared test. Figure A.4 shows a comparison of the BDT score distributions of the evaluation sets for each fold as in Figure 4.6 in the main text but also includes negatively weighted events and events with weights larger than two (which are not included in the BDT training, and thus could be evaluated with any of the BDTs). Agreement between the folds is seen and the effect of the largely weighted events is apparent in the size of the uncertainties which (as in Figure 4.6) are purely statistical.

Appendix

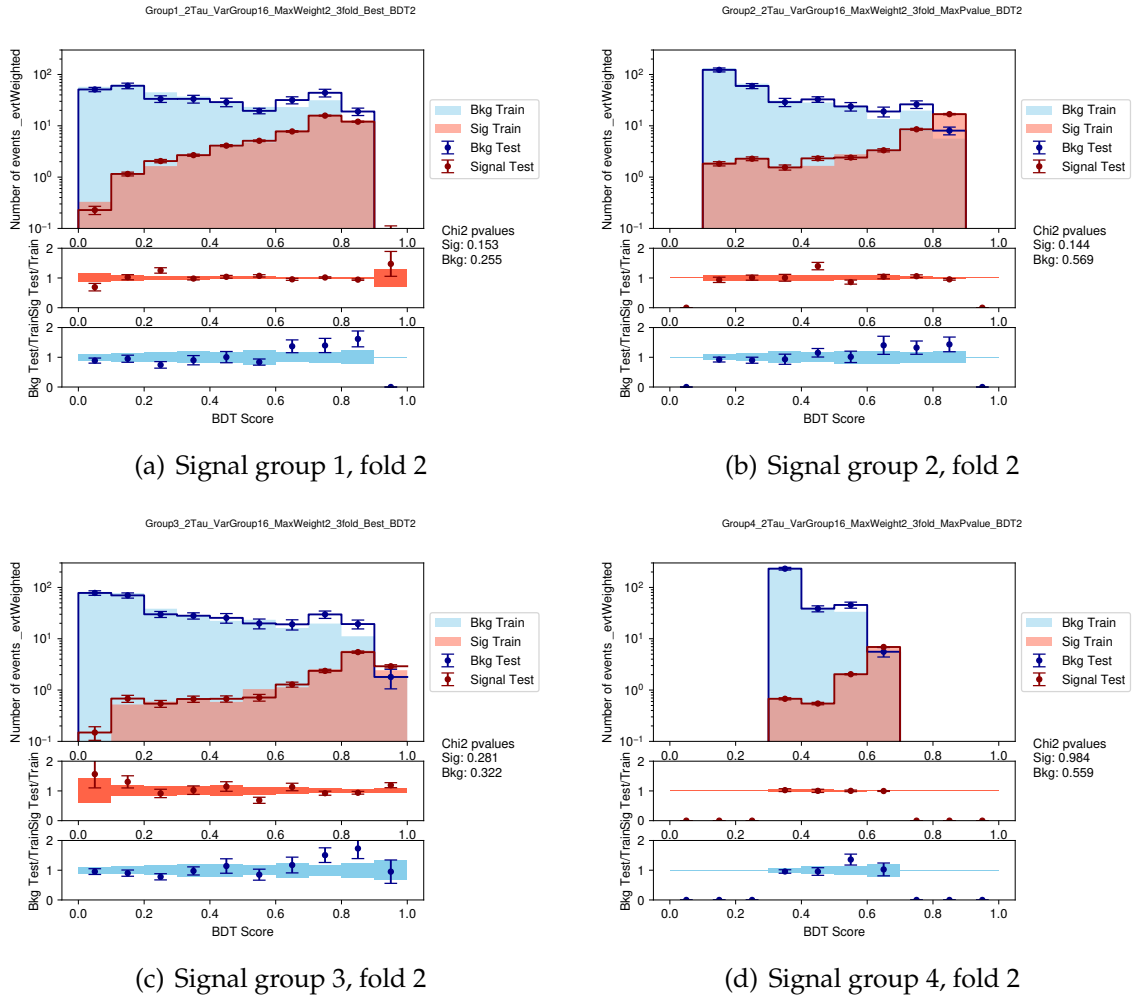


Figure A.2: Additional train/test comparisons for fold splitting 2 for BDTs trained with the different signal groups.

Appendix

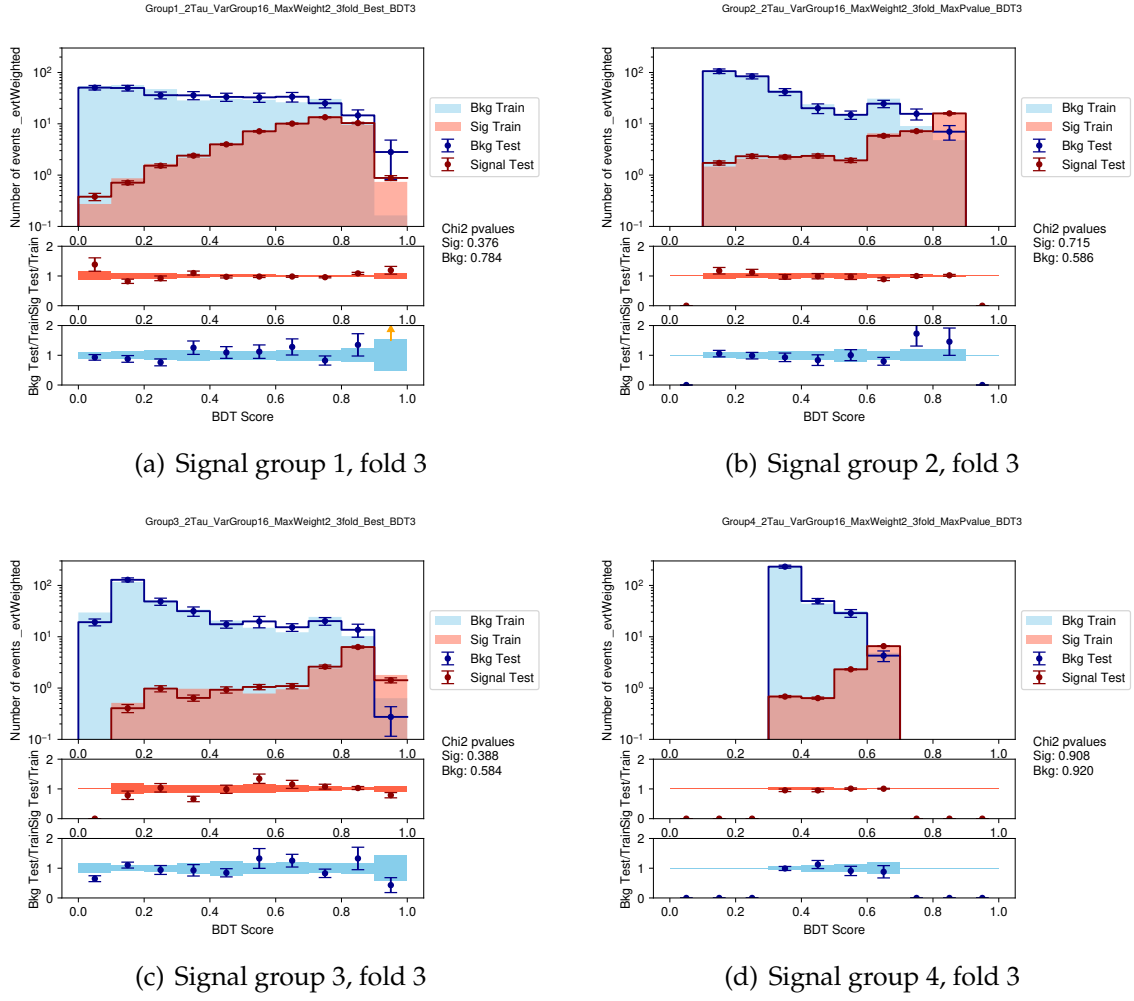


Figure A.3: Additional train/test comparisons for fold splitting 3 for BDTs trained with the different signal groups.

Appendix

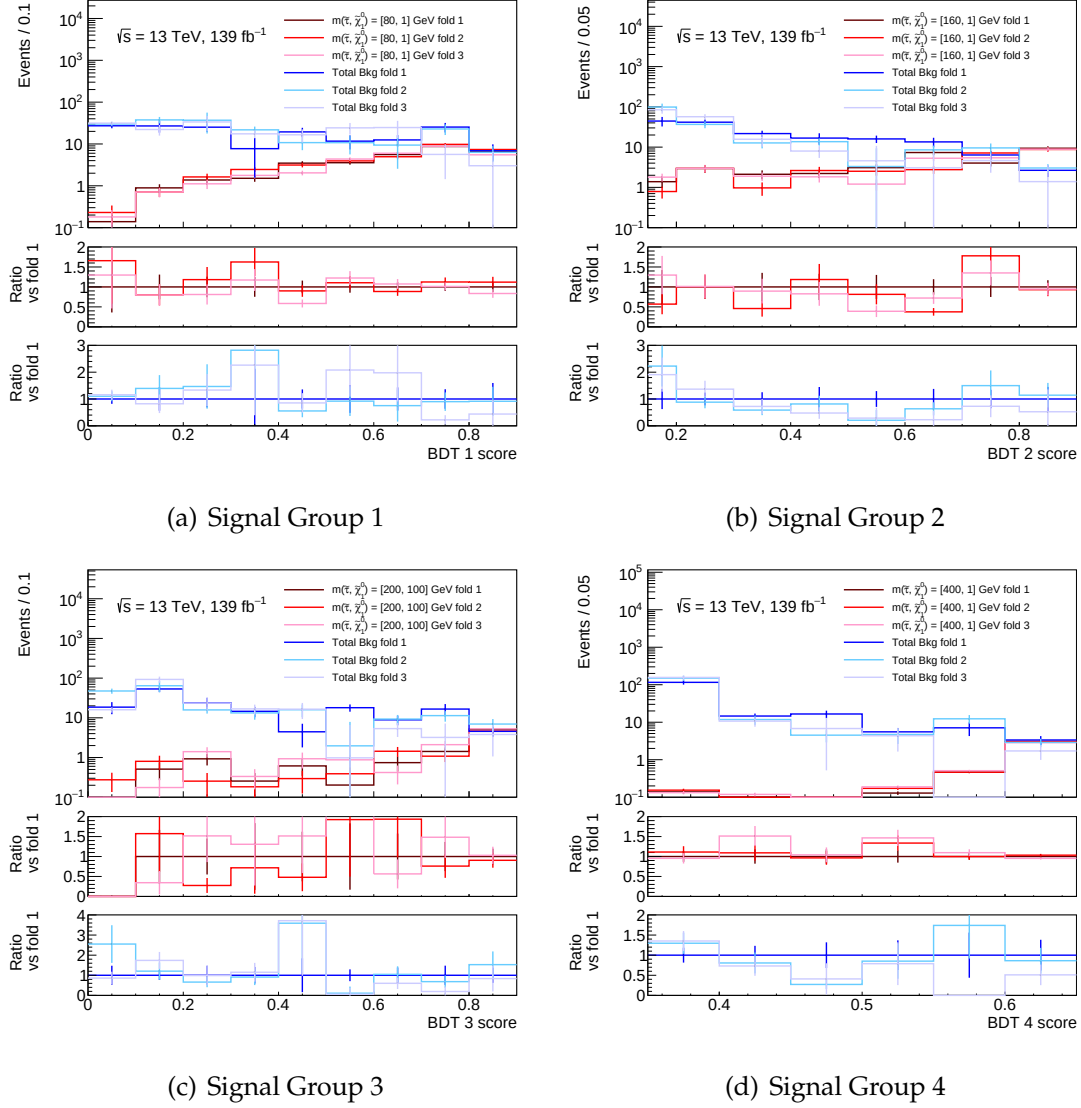


Figure A.4: Comparison of the evaluation sets across the different train/evaluation splits, for each set of BDTs, including weights with large (event weight >2) and negatively weighted events.

A.2.3 Comparison of Tight and Medium Tau-ID Working Points

Given the W +jets background, in which one of the reconstructed tau-leptons originates from a jet, is the dominant background in SR-BDT1, which targets the low mass signal points, it is worth evaluating whether using the tight working point (WP) of the RNN tau-ID algorithm instead of the medium WP could enhance the sensitivity in this signal region (SR). Whether such a change is beneficial is a trade-off between enhanced S/B and lower overall statistics, the latter factor being exacerbated in this case by the use of machine learning for the signal region selections. For reference, the average signal efficiencies and background rejection factors for the tight and medium WPs are shown in Table A.2.

Working point	Signal efficiency		Background rejection	
	1-prong	3-prong	1-prong	3-prong
Tight	60%	45%	70	700
Medium	75%	60%	35	240
Loose	85%	75%	21	90

Table A.2: Average signal efficiencies and background rejection factors for the RNN tau-ID algorithm, reproduced from [76], as determined from simulated $\gamma^* \rightarrow \tau\tau$ (signal) and dijet (background) data events.

To assess the effectiveness of the different WPs, first the yields passing the BDT training preselection requirements (Table 4.3) are compared for the cases where both tau-leptons must pass the medium WP (as in the main text) and the tight WP. These yields are shown in Table A.3. For the case where the tau-leptons are required to pass the tight WP, the multi-jet prediction, where both of the tau-leptons originate from jets, is taken as the nominal prediction for the medium WP scaled down by a factor of nine. The factor of nine is based upon the W +jets background (in which only of the tau-leptons originates from a jet) being reduced by a factor of three in moving from the medium to tight WP. In moving to the tight WP, the total background yield reduces by $\sim 60\%$, with the dominant reduction coming from the multi-jet and W +jets backgrounds, while the signal yields are reduced by $\sim 30\%$, hence there is a significant improvement in the total S/B.

Appendix

Sample	Medium	Tight
Multi-jet	193.6 ± 2.6	21.2 ± 0.9
Z+jets	142 ± 25	116 ± 24
W+jets	105 ± 20	35 ± 12
Top	28.1 ± 2.0	15.1 ± 1.5
Multi-boson	29.4 ± 1.8	17.9 ± 1.4
Higgs	2.9 ± 0.5	1.28 ± 0.13
SM total	501 ± 33	207 ± 27
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 20)$ GeV	71.9 ± 1.9	49.0 ± 1.6
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (100, 40)$ GeV	53.5 ± 1.2	35 ± 1.0
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (120, 60)$ GeV	40.1 ± 3.3	28.3 ± 2.8

Table A.3: Weighted yields at 139 fb^{-1} of events passing the BDT Preselection requirements for the SM backgrounds and low mass signal processes in the cases where both the tau-leptons are required to mass the medium or tight WP of the RNN Tau-ID algorithm discussed in the main text. Uncertainties are purely from MC statistics.

Evaluating the impact this could have on the sensitivity, high granularity distributions of BDT 1 score are plotted for events passing this selection with significance scores for example low mass signal points, using the standard ATLAS formula for estimating significance [151] with a flat 20% plus MC statistical uncertainty on the background (as was done for the SR selections in the main body), and are shown in Figure A.5. These distributions are also shown with a coarser binning and significance score per bin in Figure A.6 and the maximum single bin significance scores obtainable by cutting on BDT 1 score are tabulated in Table A.3. From these maximum scores, a gain of $\sim 25\%$ is apparent for the $m(\tilde{\tau}, \tilde{\chi}_1^0) = (120, 60)$ GeV point, while the $m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 20)$ GeV point sees a smaller 10% increase and the $m(\tilde{\tau}, \tilde{\chi}_1^0) = (100, 40)$ GeV point sees a slight decrease in moving from the medium to tight WP. Further, the distribution of the significance scores with the tight WP used are noticeably flatter than for the medium WP, suggesting an SR defined using the tight WP in the BDT 1 score is more likely to benefit from the use of a shape fit. Additionally, the kinematic distributions for events passing the medium/tight WPs are also different, hence further gain for the tight WP case may be possible from re-training the BDTs using events passing only the tight WP.

However, such a gain could potentially be offset by the reduced statistics that would be available for training the BDTs. Moreover, if the strategy used in the main

Appendix

text is followed, than the BDTs would *have* to be retrained with the restrictions applied during training to prevent regions of artificially high S/B being isolated modified to reflect the reduced statistics (in particular, the pre-binning of the input features would be significantly coarser). It is also likely that, though the yield is greatly reduced, the uncertainty on the multi-jet estimate would be larger were it to be estimated in the same way due to the lower statistics in the ABCD control regions that would require the tight rather than medium tau-leptons. Given the low MC statistics, an alternative method for estimating the W -jets background may be necessary as well. Hence, from this simple study, it is not possible to fully determine whether a significant gain would be possible for the low mass signals in moving from the medium to tight RNN tau-ID WPs, but there is certainly potential. Thus, the use of the tight WP for SRs targeting signals with low $m(\tilde{\tau})$ should be studied in more detail in future iterations of the analysis, particularly if a greater integrated luminosity is used.

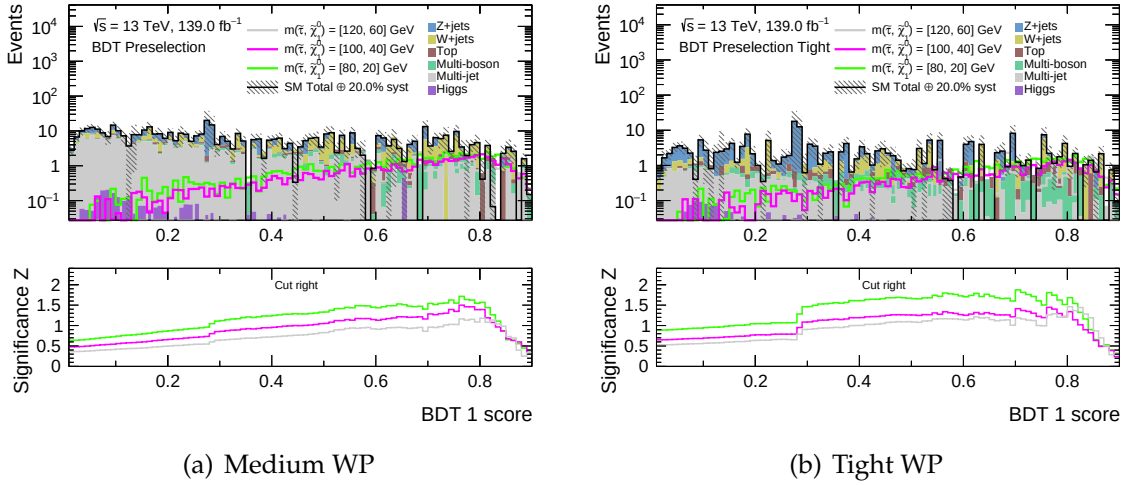


Figure A.5: Significance scores for selecting on BDT 1 score for using the medium and tight tau-ID working points. Scores are shown for example signal points with low $m(\tilde{\tau})$.

Signal	Medium	Tight
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (80, 20) \text{ GeV}$	1.71	1.88
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (100, 40) \text{ GeV}$	1.50	1.46
$m(\tilde{\tau}, \tilde{\chi}_1^0) = (120, 60) \text{ GeV}$	1.18	1.46

Table A.4: Maximum significance scores obtainable by cutting on BDT 1 score for example signal points in the cases where both tau-leptons are required to mass either the medium or tight WP of the RNN tau-ID algorithm.

Appendix

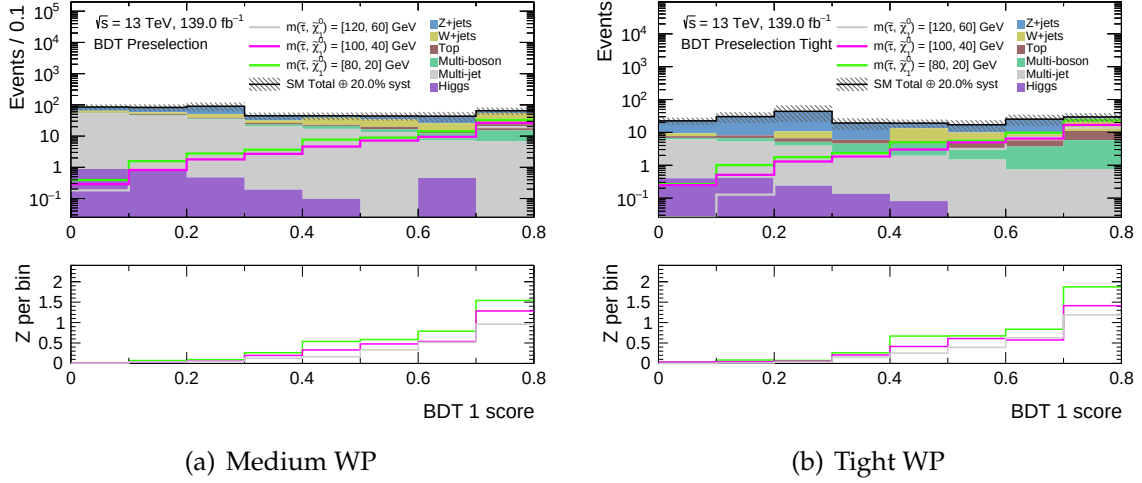


Figure A.6: Significance scores per bin of the BDT 1 score distribution for using the medium and tight tau-ID working points. Scores are shown for example signal points with low $m(\tilde{\tau})$. Note that a finer binning than the one shown here (like the one used in the main text) is more optimal when using the medium WP.

A.2.4 Simplified Multi-jet Estimation Comparison

Figure A.7 shows comparisons of a selection of kinematic variables between the simplified and full multi-jet estimates, described in Section 4.6.1, at BDT preselection level. In general, the agreement between the different methods is good, although the simplified estimate is larger at high E_T^{miss} and high p_T where the overall multi-jet contribution is small.

Appendix

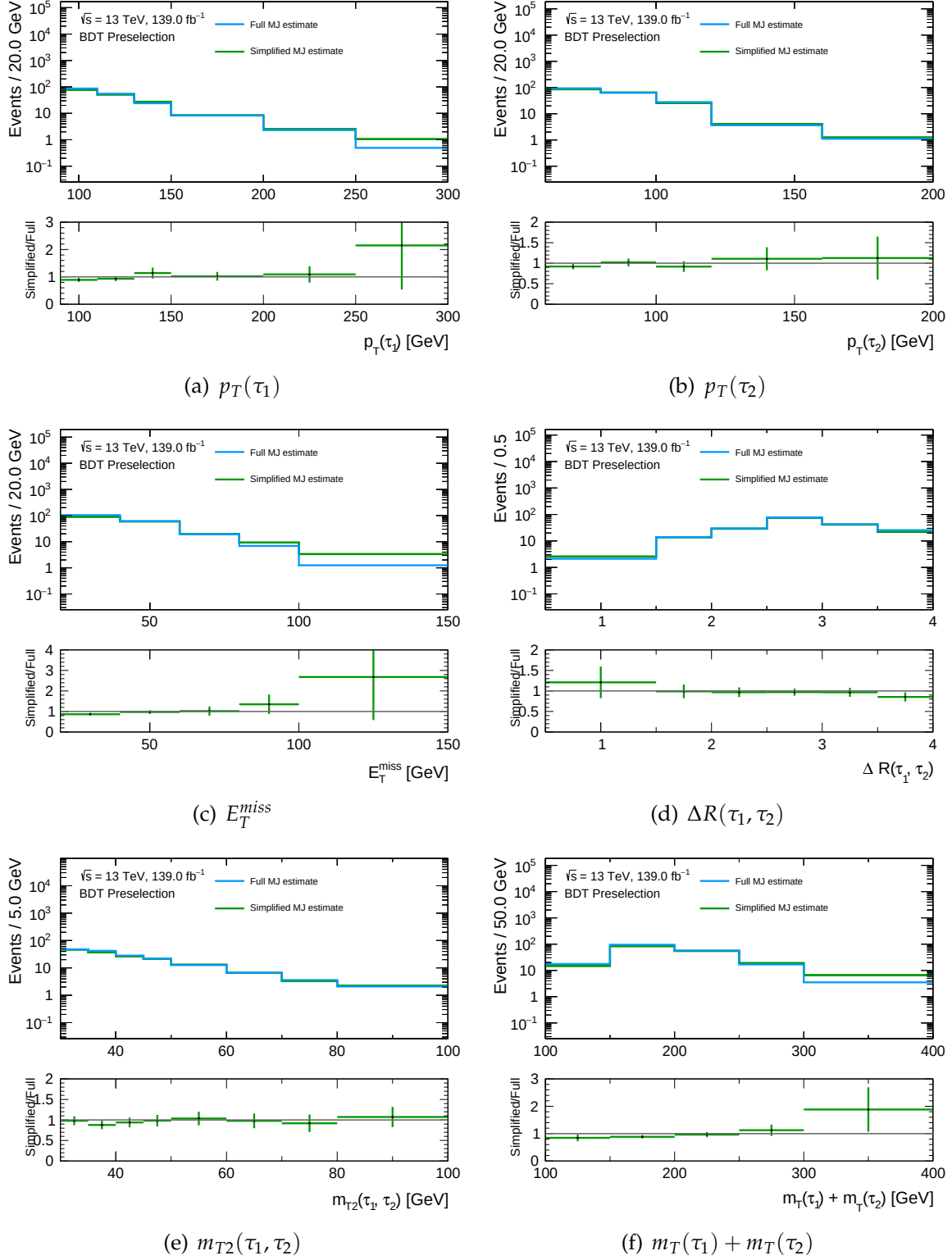


Figure A.7: Comparison of the simplified and full multi-jet estimation methods for events passing the BDT preselection. The simplified method includes no negatively weighted events and is used only to train the BDTs.

A.2.5 Cut Flow Tables

Tables A.5 and A.6 show the expected number of events at 139 fb^{-1} sequentially passing the signal region selections.

Selection	[80, 1]	[160, 1]	[200, 100]	[400, 1]
Preselection	413.0	230.0	89.7	18.5
trigger matching	410.2	227.3	89.3	18.3
Electron veto	407.1	224.9	88.9	18.2
Muon veto	404.9	224.4	88.3	18.1
OS τ pair	399.3	221.3	87.1	17.8
$m_{T2}(\tau_1, \tau_2) > 30 \text{ GeV}$	99.8	125.6	47.6	14.5
asymm. ditau τ p_T thresholds	64.5	93.8	32.0	12.7
3rd baseline Tau veto	64.5	93.6	31.9	12.7
b -jet veto	63.4	91.1	31.0	12.4
$m_{\tau\tau} > 120 \text{ GeV}$	63.1	90.6	30.8	11.9
$\Delta R(\tau_1, \tau_2) < 4$	63.1	90.6	30.8	11.8
$E_T^{\text{miss}} > 20 \text{ GeV}$	62.5	90.3	30.8	11.8
SR BDT-1 (BDT1score > 0.73)	40.3	45.3	15.4	4.6
SR BDT-2 (BDT2score > 0.78)	4.2	31.6	11.6	9.5
SR BDT-3 (BDT3score > 0.79)	10.8	41.5	15.1	9.4
SR BDT-4 (BDT4score > 0.64)	0.1	10.3	3.2	7.3

Table A.5: Number of expected signal events at 139 fb^{-1} passing the selections for the direct stau signal regions for example $[m(\tilde{\tau}), m(\tilde{\chi}_1^0)]$ signal points, with masses in GeV. Here ‘Preselection’ refers to events passing the asymmetric ditau trigger and having at least two signal tau-leptons, which was required in the ntuple production of these signals. The OS τ pair requirement is for the two highest p_T tau-leptons to be of opposite sign. Cuts except the BDT score cuts are applied consecutively, while the BDT score cuts are applied one at a time.

Appendix

Selection	Multi-jet	Z+jets	W+jets	Top
Preselection	-	18984.2	4085.7	2403.2
trigger matching	-	18615.1	4001.2	2215.5
N sig. $\tau \geq 2$	-	14412.2	2153.2	1486.9
Electron veto	-	14117.4	1919.0	1397.5
Muon veto	-	13880.7	1871.8	1316.9
OS τ pair	-	13612.6	1593.4	1259.7
$m_{T2}(\tau_1, \tau_2) > 30$ GeV	820.9	1400.4	256.6	309.2
asymm. ditau τ p_T thresholds	291.8	838.1	120.6	154.3
3rd baseline Tau veto	288.6	835.1	119.7	153.1
b -jet veto	269.9	735.8	111.4	30.4
$m_{\tau\tau} > 120$ GeV	239.3	145.2	106.3	28.9
$\Delta R(\tau_1, \tau_2) < 4$	219.8	146.2	107.5	28.5
$E_T^{\text{miss}} > 20$ GeV	193.6	142.1	104.5	28.1
SR BDT-1 (BDT1score > 0.73)	4.6	7.1	22.7	7.8
SR BDT-2 (BDT2score > 0.78)	0.0	2.0	1.6	2.1
SR BDT-3 (BDT3score > 0.79)	0.0	6.1	2.3	3.7
SR BDT-4 (BDT4score > 0.64)	0.2	0.2	0.0	0.2
Selection	Higgs	Multi-boson	SM Total	Data
Preselection	478.6	819.7	-	77530
trigger matching	462.1	735.8	-	76540
N sig. $\tau \geq 2$	365.0	560.7	-	31427
Electron veto	352.3	507.2	-	30958
Muon veto	339.1	456.0	-	30500
OS τ pair	327.5	410.9	-	24528
$m_{T2}(\tau_1, \tau_2) > 30$ GeV	58.9	112.6	2958.5	3127
asymm. ditau τ p_T thresholds	35.2	63.6	1503.5	1562
3rd baseline Tau veto	34.1	61.0	1491.6	1554
b -jet veto	28.5	55.8	1231.8	1274
$m_{\tau\tau} > 120$ GeV	4.1	30.0	554.0	605
$\Delta R(\tau_1, \tau_2) < 4$	4.1	29.8	536.0	587
$E_T^{\text{miss}} > 20$ GeV	2.9	29.4	500.6	534
SR BDT-1 (BDT1score > 0.73)	0.0	6.6	48.8	36
SR BDT-2 (BDT2score > 0.78)	0.0	4.1	9.3	4
SR BDT-3 (BDT3score > 0.79)	0.1	4.5	16.6	7
SR BDT-4 (BDT4score > 0.64)	0.0	1.8	2.3	1

Table A.6: Number of expected background and data events at 139 fb^{-1} passing the selections for the direct stau signal regions. Here, the ‘Preselection’ requirement is for events to be selected by the asymmetric ditau trigger and to have at least two baseline tau-leptons. The OS τ pair requirement is for the two highest p_T tau-leptons to be of opposite sign. Normalisation factors from the background-only fit to the control regions are applied to Z+jets, W+jets and top backgrounds. Selections except those on the BDT score cuts are applied consecutively, while the BDT score cuts are applied exclusively. No yields are shown for the multi-jet or total SM prior to the selection on $m_{T2}(\tau_1, \tau_2)$ as prior to this selection there is no multi-jet estimate.

A.2.6 Cross-section Upper Limits

Figure A.8 shows the observed 95% CL cross-section upper limits set by the direct stau analysis (using the best expected signal region for each point). Expected and observed contours showing where the upper limits are equal to the cross-sections of the simplified models are overlaid. Note all the fits used to compute these limits use the asymptotic approximation instead of pseudo-experiments as was done in the main text for the computation of the CL_s values for the fits using SR-BDT2, SR-BDT3 and SR-BDT4. This results in slightly stronger limits at high $m(\tilde{\tau})$ compared to the results shown in the main text.

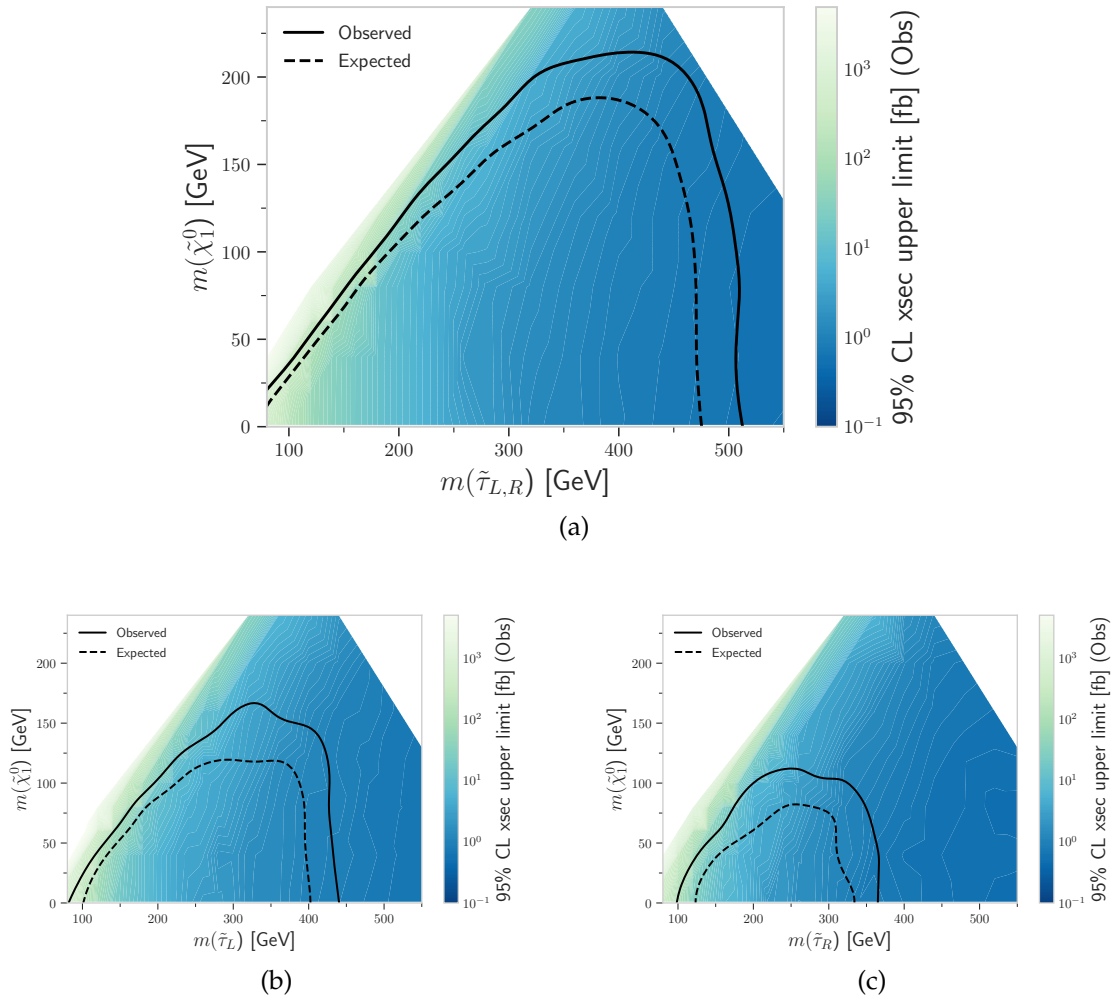


Figure A.8: Observed 95% CL upper limits on the cross-section for (a), $\tilde{\tau}_{L,R}$, (b), $\tilde{\tau}_L$ and (c), $\tilde{\tau}_R$ production. The contour lines indicate where the upper limits equal that of the theoretical predictions.

A.3 Additional details for the Vector Boson Fusion Sensitivity Study

This section provides additional information on the Vector Boson Fusion sensitivity study presented in Chapter 5, specifically the BDT used.

A.3.1 Signal vs Background Shape plots

Figure A.9 shows comparisons of the shapes of the kinematic variables that are used as input features for the BDTs between the signal and background samples used in the training.

A.3.2 Additional BDT details

Figure A.10 shows comparisons of the BDT score distributions between the training and evaluation sets of the optimal models used in the VBF sensitivity study, for the four other fold splittings. In general the models are very consistent with the first fold splitting shown in the main text in Figure 5.8.

Appendix

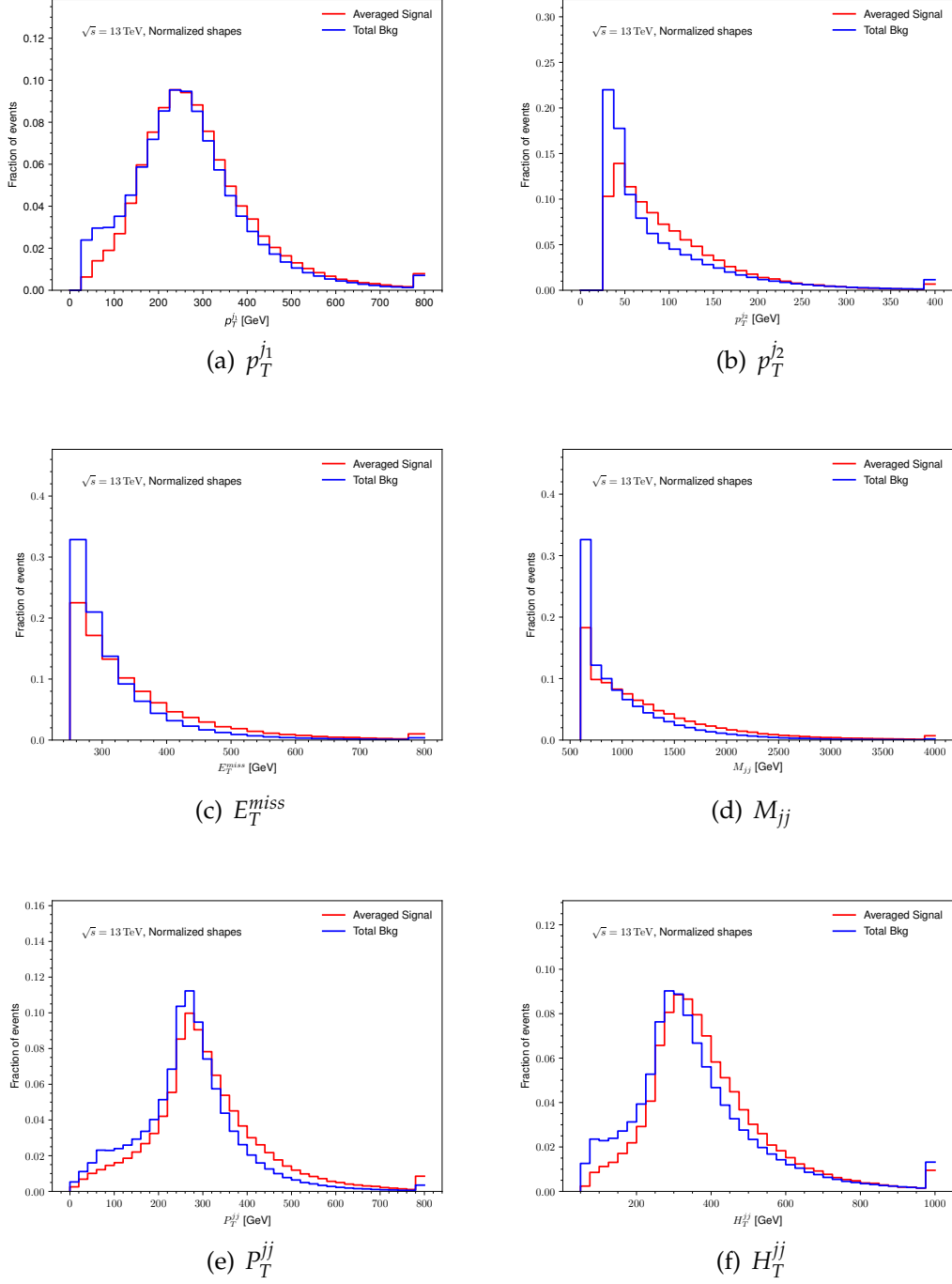


Figure A.9: Shape comparison of total background and averaged signal for at the VBF BDT preselection level.

Appendix

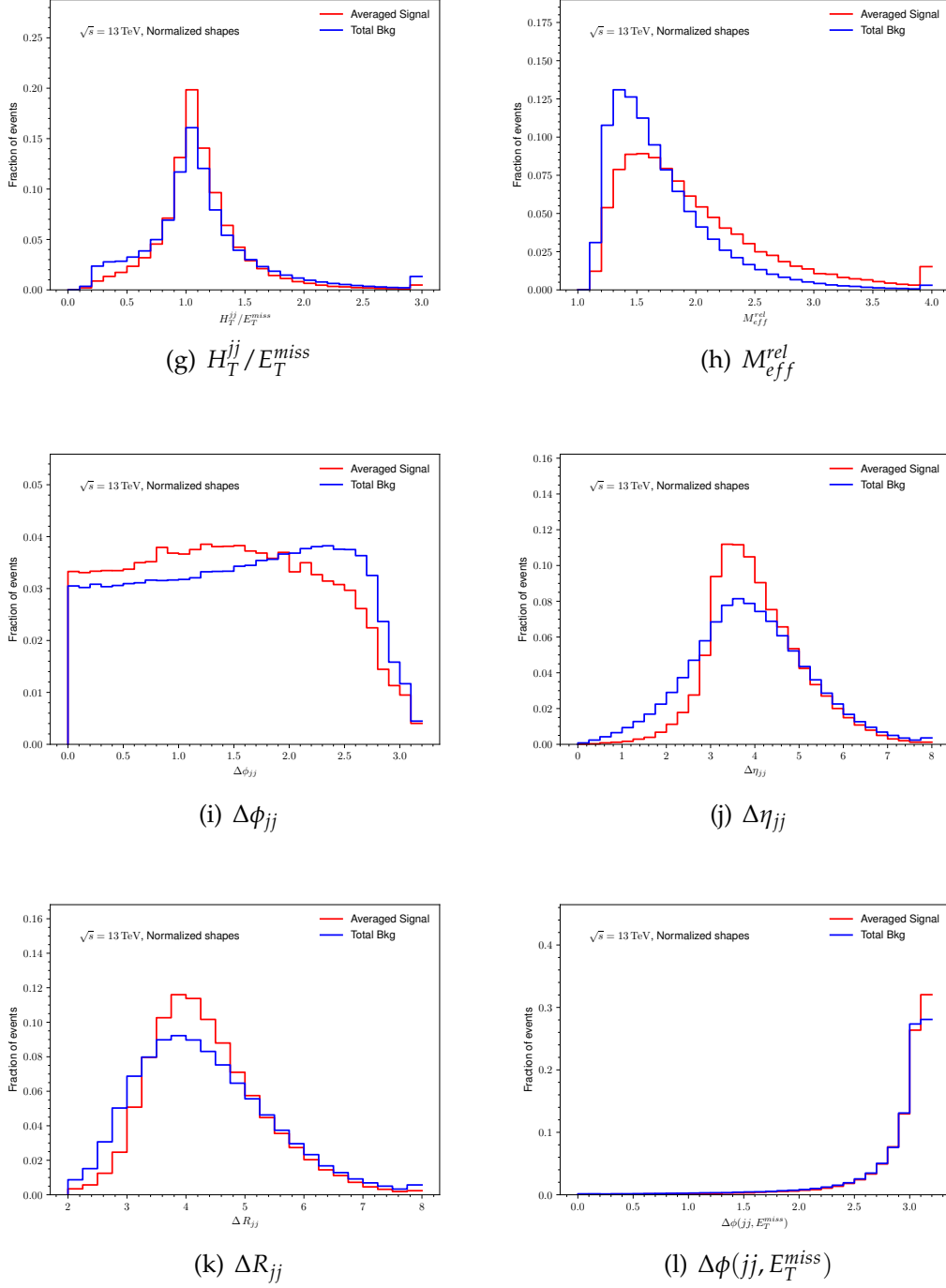


Figure A.9: Shape comparison of total background and averaged signal for at the VBF BDT preselection level.

Appendix

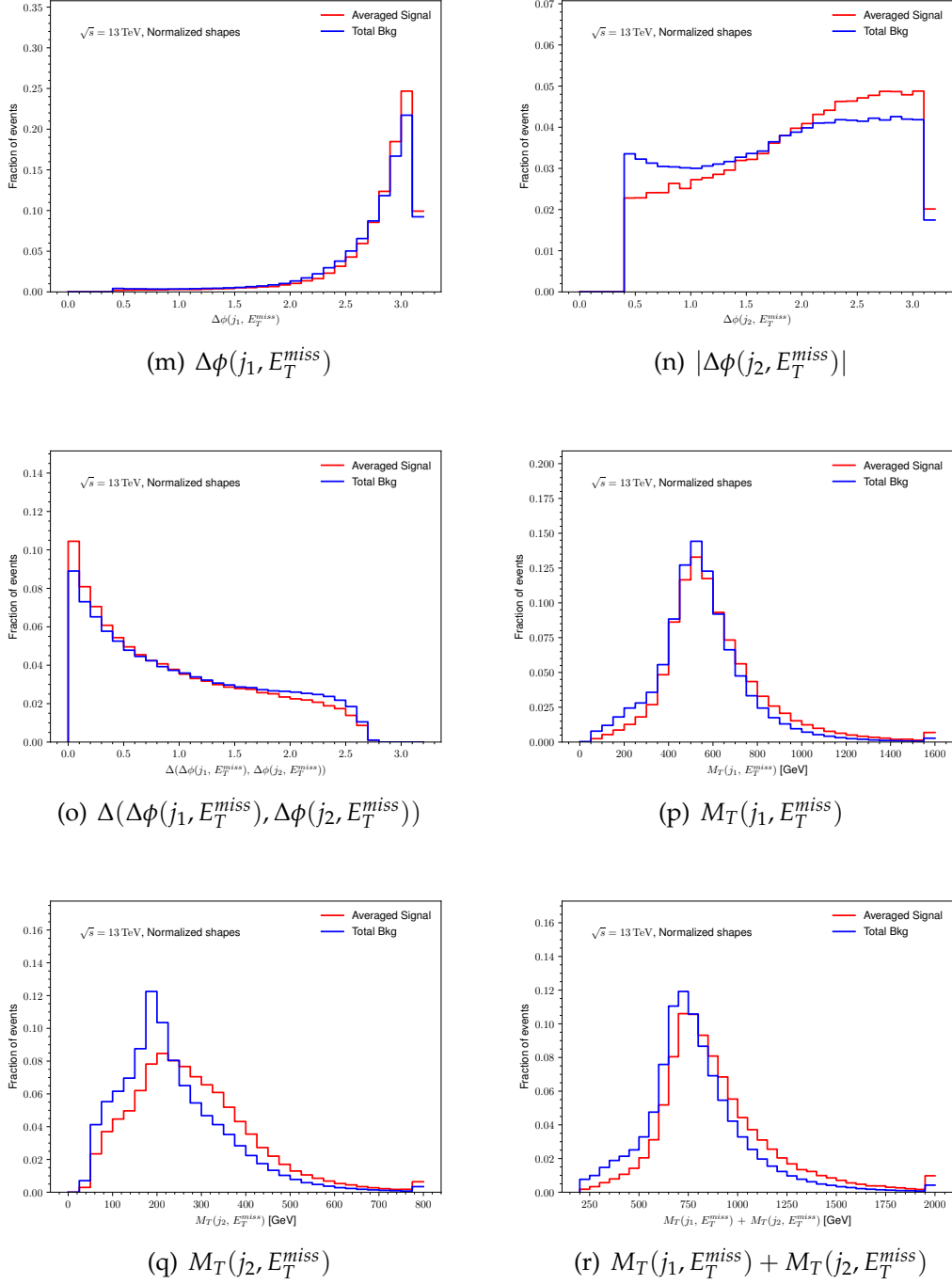


Figure A.9: Shape comparison of total background and averaged signal for at the VBF BDT preselection level.

Appendix

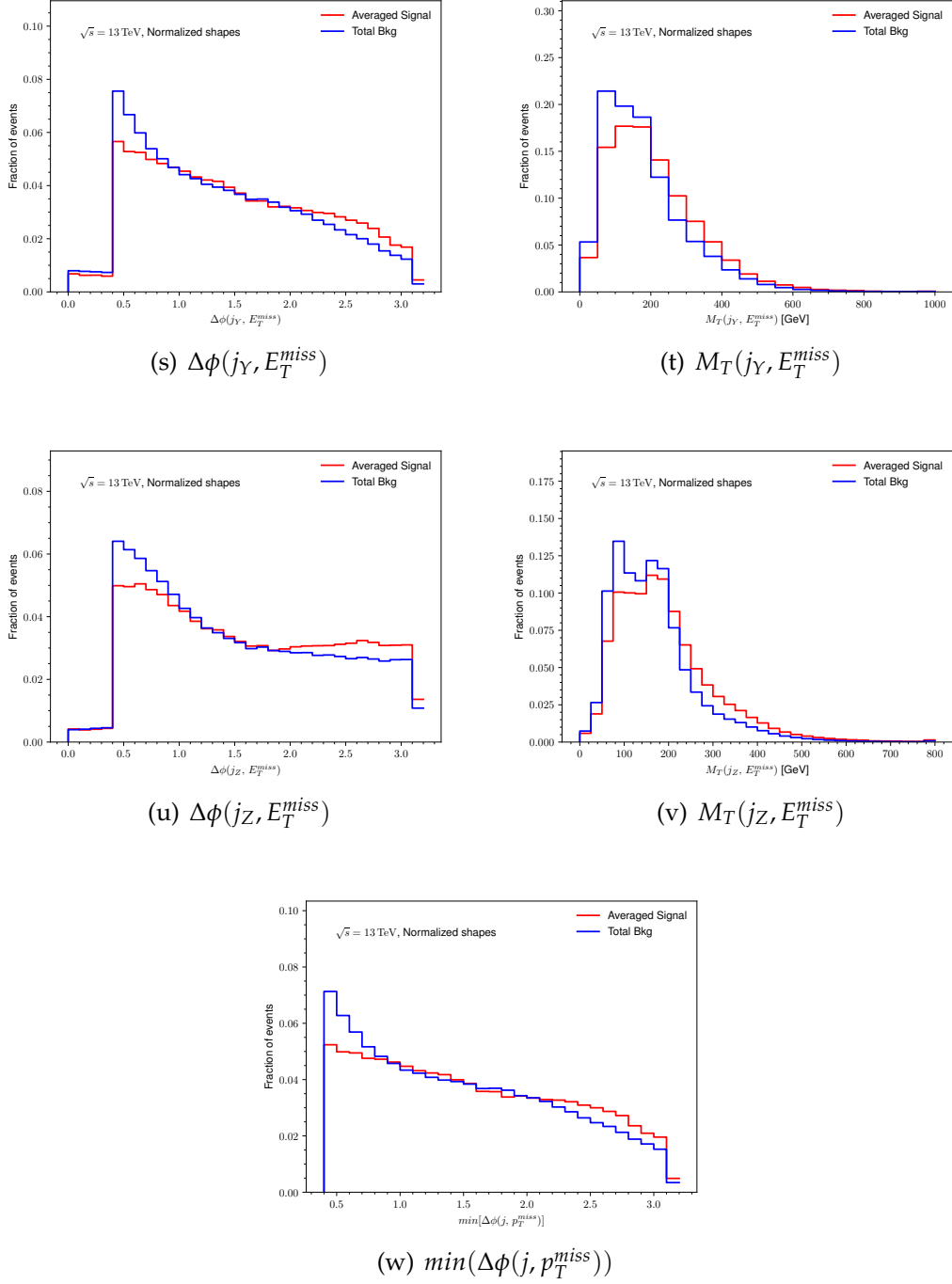


Figure A.9: Shape comparison of total background and averaged signal for at the VBF BDT preselection level.

Appendix

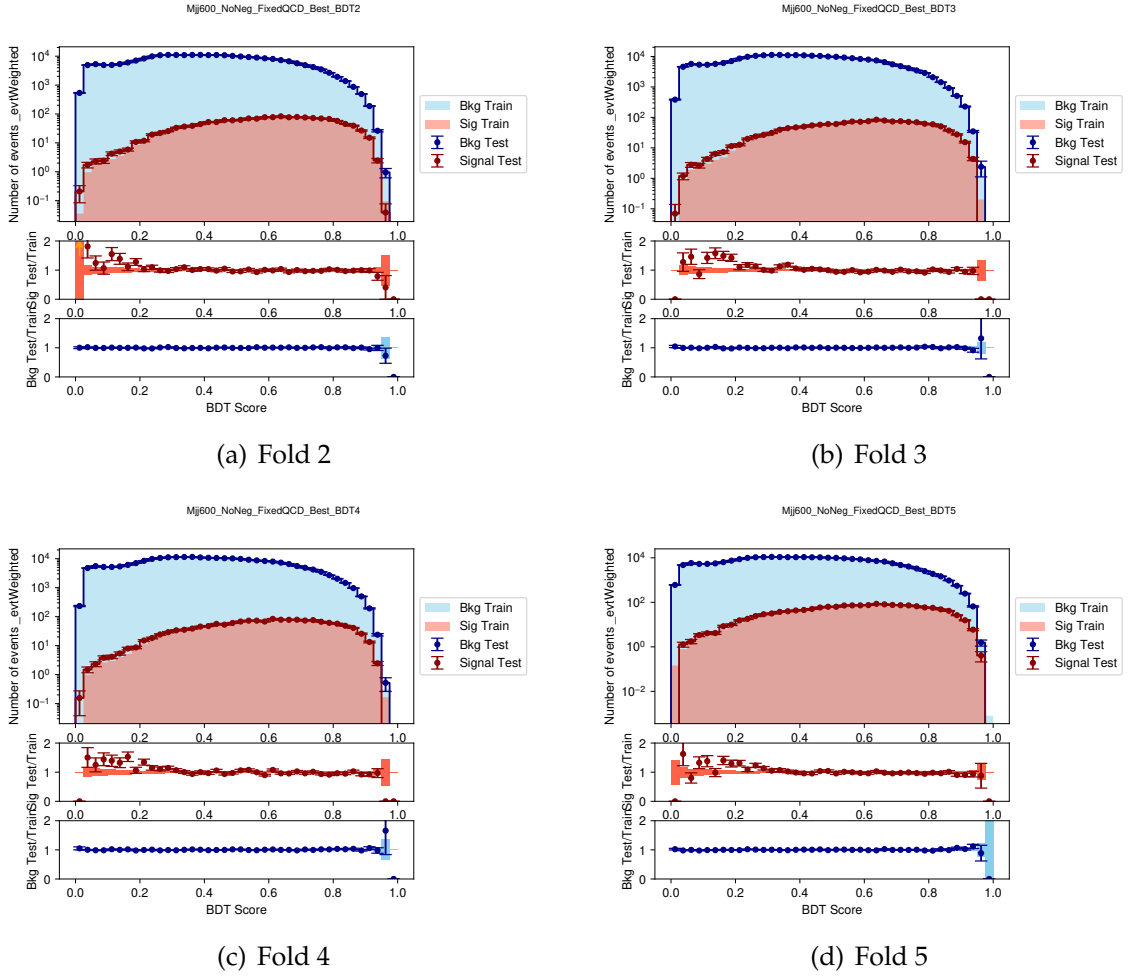


Figure A.10: Additional training/evaluation set comparisons for the BDTs used in the VBF sensitivity study.

A.3.3 Cut Flow Tables

Tables A.7 and A.8 show the number of expected events at 139fb^{-1} passing the VBF signal region selections as they are applied sequentially.

Selection	HH (100,99)	WB (100,99)	WB (140,139)
Preselection	2296.8	4728.9	1923.1
$E_T^{\text{miss}} > 250 \text{ GeV}$	1414.7	2918.0	1254.1
$\min[\Delta\phi(j, E_T^{\text{miss}})] > 0.4$	1109.1	2330.2	1024.7
Baseline lepton veto	1071.9	2230.0	985.5
b -jet veto	934.2	1953.6	873.0
$p_T^{j_1} > 60 \text{ GeV}$	923.0	1930.7	864.5
$p_T^{j_2} > 60 \text{ GeV}$	599.7	1295.7	609.1
$\Delta\eta_{jj} > 3.8$	248.4	592.4	289.9
$M_{jj} > 1000 \text{ GeV}$	214.9	522.7	260.0
BDT score > 0.65	340.8	781.7	399.1

Table A.7: Number of expected signal events at 139 fb^{-1} passing the VBF signal region selections for benchmark ($m(\tilde{\chi}_2^0) \text{ GeV}$, $m(\tilde{\chi}_1^0) \text{ GeV}$) mass points. Requirements are applied sequentially except for the additional requirements, separated by horizontal lines, that define SR-CMS and SR-BDT, which are applied separately. Preselection refers to events passing and matching the E_T^{miss} trigger, and having two VBF-tagged jets (non- b -tagged, $p_T > 30 \text{ GeV}$) in opposite hemispheres ($\eta^{j_1} \times \eta^{j_2} < 0$) with invariant mass $M_{jj} > 600 \text{ GeV}$.

Appendix

Selection	QCD Z+jets	QCD W+jets	EW Z+jets	EW W+jets
Preselection	562841.7	1328706.1	23242.6	68690.6
$E_T^{\text{miss}} > 250 \text{ GeV}$	218411.6	327225.1	10762.4	20201.4
$\min[\Delta\phi(j, E_T^{\text{miss}})] > 0.4$	165479.3	219217.9	8993.8	14805.1
Baseline lepton veto	157668.0	84080.3	8697.9	4660.1
b -jet veto	139434.8	71838.6	8102.4	4126.7
$p_T^{j_1} > 60 \text{ GeV}$	134437.4	69141.4	8030.6	4084.5
$p_T^{j_2} > 60 \text{ GeV}$	71631.2	35833.5	6232.2	3086.9
$\Delta\eta_{jj} > 3.8$	18952.3	10158.5	2666.9	1415.0
$M_{jj} > 1000 \text{ GeV}$	14568.8	7805.8	2484.0	1322.8
BDT score > 0.65	17811.6	7144.3	3143.3	1382.3

Selection	Multi- boson	Top	Total SM
Preselection	65669.1	649372.5	2698522.7
$E_T^{\text{miss}} > 250 \text{ GeV}$	23835.3	152946.4	753382.2
$\min[\Delta\phi(j, E_T^{\text{miss}})] > 0.4$	14751.7	69953.8	493201.6
Baseline lepton veto	7074.9	17804.8	279986.1
b -jet veto	5496.3	2413.7	231412.5
$p_T^{j_1} > 60 \text{ GeV}$	5299.3	2284.8	223278.1
$p_T^{j_2} > 60 \text{ GeV}$	3101.9	1187.6	121073.3
$\Delta\eta_{jj} > 3.8$	805.5	445.5	34443.8
$M_{jj} > 1000 \text{ GeV}$	598.3	320.2	27100.0
BDT score > 0.65	671.0	188.1	30340.5

Table A.8: Number of expected background events at 139 fb^{-1} passing the VBF signal region selections. Requirements are applied sequentially except for the additional requirements, separated by horizontal lines, that define SR-CMS and SR-BDT, which are applied separately. Preselection refers to events passing and matching the E_T^{miss} trigger, and having two VBF-tagged jets (non- b -tagged, $p_T > 30 \text{ GeV}$) in opposite hemispheres ($\eta^{j_1} \times \eta^{j_2} < 0$) with invariant mass $M_{jj} > 600 \text{ GeV}$. No correction factors are applied to the V +jets backgrounds.

A.4 Cross-section Upper Limits for Chargino Pair-production Combination

Figure A.11 shows the observed 95% CL cross-section upper limits resulting from the combination of searches for chargino pair-production in Chapter 6. Expected and observed contours showing where the upper limits are equal to the cross-sections of the simplified models are overlaid.

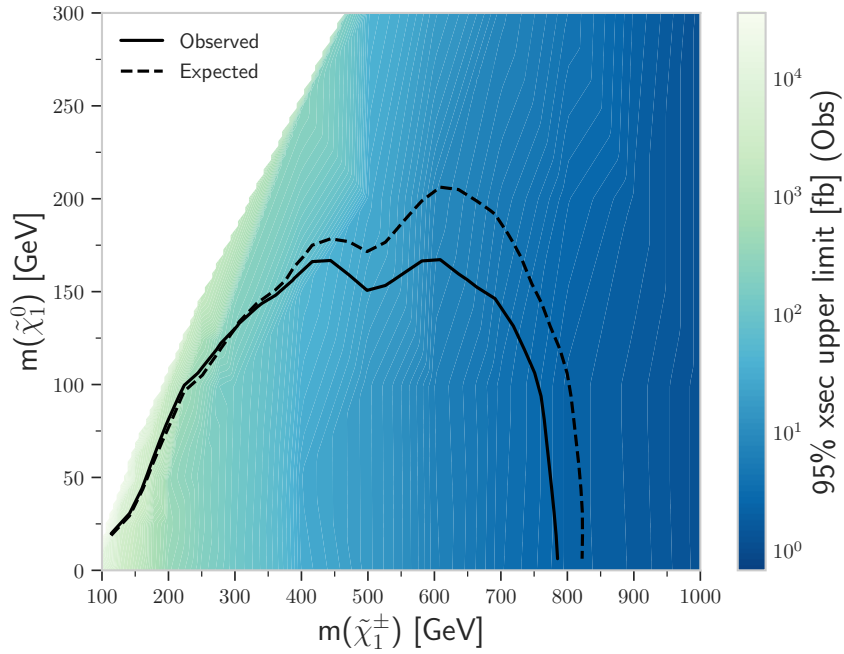


Figure A.11: Observed 95% CL cross-section upper limits.

Colophon

This thesis was made in $\text{\LaTeX}2_{\epsilon}$ using the “hepthesis” class [[177](#)].

Bibliography

- [1] ATLAS Collaboration, *Search for electroweak production of supersymmetric particles in final states with two τ -leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, [arXiv:2402.00603 \[hep-ex\]](#).
- [2] ATLAS Collaboration, *A statistical combination of ATLAS Run 2 searches for charginos and neutralinos at the LHC*, [arXiv:2402.08347 \[hep-ex\]](#).
- [3] D. Rousso, D. Jones, P. Federicova, et al., *Test and extraction methods for the QC parameters of silicon strip sensors for ATLAS upgrade tracker*, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment](#) **1045** (2023) 167608.
- [4] V. Fadeyev, M. Beranek, E. Bach Marques, et al., *Monitoring Quality of ATLAS ITk Strip Sensors through Database*, CERN, Geneva, 2023. tech. rep., ATL-ITK-PROC-2023-004, <https://cds.cern.ch/record/2847934>.
- [5] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, [Phys. Lett. B](#) **716** (2012) 1–29, [arXiv:1207.7214 \[hep-ex\]](#).
- [6] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, [Phys. Lett. B](#) **716** (2012) 30–61, [arXiv:1207.7235 \[hep-ex\]](#).
- [7] ATLAS Collaboration, *Standard Model Summary Plots February 2022*, CERN, Geneva, 2022. tech. rep., ATL-PHYS-PUB-2022-009, <http://cds.cern.ch/record/2804061>.
- [8] Particle Data Group Collaboration, *Review of Particle Physics*, [PTEP](#) **2022** (2022) 083C01.
- [9] J. Ellis, *TikZ-Feynman: Feynman diagrams with TikZ*, [Comput. Phys. Commun.](#)

BIBLIOGRAPHY

- 210 (2017) 103–123, [arXiv:1601.05437 \[hep-ph\]](#).
- [10] E. Noether, *Invariante Variationsprobleme*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918** (1918) 235–257.
- [11] NA48 Collaboration, *A New measurement of direct CP violation in two pion decays of the neutral kaon*, *Phys. Lett. B* **465** (1999) 335–348, [arXiv:hep-ex/9909022](#).
- [12] KTeV Collaboration, *Observation of direct CP violation in $K_{S,L} \rightarrow \pi\pi$ decays*, *Phys. Rev. Lett.* **83** (1999) 22–27, [arXiv:hep-ex/9905060](#).
- [13] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, T. Schwetz, and A. Zhou, *The fate of hints: updated global analysis of three-flavor neutrino oscillations*, *JHEP* **09** (2020) 178, [arXiv:2007.14792 \[hep-ph\]](#).
- [14] C. Patrignani, *Review of Particle Physics*, *Chinese Physics C* **40** no. 10, (2016) 100001.
- [15] Muon g-2 Collaboration, *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm*, [arXiv:2308.06230 \[hep-ex\]](#).
- [16] S. Borsanyi, Z. Fodor, J. N. Guenther, et al., *Leading hadronic contribution to the muon magnetic moment from lattice QCD*, *Nature* **593** no. 7857, (2021) 51–55, [arXiv:2002.12347 \[hep-lat\]](#).
- [17] CDF Collaboration, *High-precision measurement of the W boson mass with the CDF II detector*, *Science* **376** no. 6589, (2022) 170–176.
- [18] ATLAS Collaboration, *Improved W boson Mass Measurement using 7 TeV Proton-Proton Collisions with the ATLAS Detector*, CERN, Geneva, 2023. tech. rep., ATLAS-CONF-2023-004, <https://cds.cern.ch/record/2853290>.
- [19] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279–390, [arXiv:hep-ph/0404175](#).
- [20] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6, [arXiv:1807.06209 \[astro-ph.CO\]](#).
- [21] J. R. Primack and M. A. K. Gross, *Hot dark matter in cosmology*, [arXiv:astro-ph/0007165](#).
- [22] M. S. Seigar, *Dark Matter in the Universe*. 2053-2571. Morgan & Claypool

BIBLIOGRAPHY

- Publishers, 2015. <https://dx.doi.org/10.1088/978-1-6817-4118-5>.
- [23] C. Tao, *Dark Matter searches: an overview*, **JINST** **15** no. 06, (2020) C06054.
- [24] J. Martin, *Everything you always wanted to know about the cosmological constant problem (but were afraid to ask)*, **Comptes Rendus Physique** **13** no. 6-7, (2012) 566–665.
- [25] C. Abel et al., *Measurement of the Permanent Electric Dipole Moment of the Neutron*, **Phys. Rev. Lett.** **124** no. 8, (2020) 081803, [arXiv:2001.11966](https://arxiv.org/abs/2001.11966) [hep-ex].
- [26] P. Fayet, *Supersymmetry and Weak, Electromagnetic and Strong Interactions*, **Phys. Lett. B** **64** (1976) 159.
- [27] P. Fayet, *Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions*, **Phys. Lett. B** **69** (1977) 489.
- [28] G. R. Farrar and P. Fayet, *Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry*, **Phys. Lett. B** **76** (1978) 575.
- [29] S. P. Martin, *A Supersymmetry primer*, **Adv. Ser. Direct. High Energy Phys.** **18** (1998) 1–98, [arXiv:hep-ph/9709356](https://arxiv.org/abs/hep-ph/9709356).
- [30] H. Baer, V. Barger, S. Salam, D. Sengupta, and X. Tata, *The LHC higgsino discovery plane for present and future SUSY searches*, **Physics Letters B** **810** (2020) 135777.
- [31] ATLAS Collaboration, *SUSY March 2023 Summary Plot Update*, CERN, Geneva, 2023. tech. rep., ATL-PHYS-PUB-2023-005, <https://cds.cern.ch/record/2852738>.
- [32] CMS Collaboration, *Combined searches for the production of supersymmetric top quark partners in proton–proton collisions at $\sqrt{s} = 13$ TeV*, **Eur. Phys. J. C** **81** no. 11, (2021) 970, [arXiv:2107.10892](https://arxiv.org/abs/2107.10892) [hep-ex].
- [33] A. Mustafayev and X. Tata, *Supersymmetry, Naturalness, and Light Higgsinos*, **Indian J. Phys.** **88** (2014) 991–1004, [arXiv:1404.1386](https://arxiv.org/abs/1404.1386) [hep-ph].
- [34] H. Goldberg, *Constraint on the Photino Mass from Cosmology*, **Phys. Rev. Lett.** **50** (1983) 1419.
- [35] J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive, and M. Srednicki, *Supersymmetric Relics from the Big Bang*, **Nucl. Phys. B** **238** (1984) 453.

BIBLIOGRAPHY

- [36] ATLAS Collaboration, *Summary of the ATLAS experiment's sensitivity to supersymmetry after LHC Run 1 — interpreted in the phenomenological MSSM*, **JHEP** **10** (2015) 134, [arXiv:1508.06608 \[hep-ex\]](#).
- [37] M. Carena and H. E. Haber, *Higgs Boson Theory and Phenomenology*, **Prog. Part. Nucl. Phys.** **50** (2003) 63–152, [arXiv:hep-ph/0208209](#).
- [38] M. Chakraborti, S. Iwamoto, J. S. Kim, R. Masełek, and K. Sakurai, *Supersymmetric explanation of the muon $g - 2$ anomaly with and without stable neutralino*, **JHEP** **08** (2022) 124, [arXiv:2202.12928 \[hep-ph\]](#).
- [39] J. M. Yang and Y. Zhang, *Low energy SUSY confronted with new measurements of W-boson mass and muon $g-2$* , **Sci. Bull.** **67** no. 14, (2022) 1430–1436, [arXiv:2204.04202 \[hep-ph\]](#).
- [40] J. Alwall, P. C. Schuster, and N. Toro, *Simplified models for a first characterization of new physics at the LHC*, **Physical Review D** **79** no. 7, (2009).
- [41] MSSM Working Group Collaboration, A. Djouadi et al., *The Minimal supersymmetric standard model: Group summary report*, in GDR (Groupement De Recherche) - Supersymetrie. 12, 1998. [arXiv:hep-ph/9901246](#).
- [42] CMS Collaboration, *Phenomenological MSSM interpretation of CMS searches in pp collisions at $\sqrt{s} = 7$ and 8 TeV*, **JHEP** **10** (2016) 129, [arXiv:1606.03577 \[hep-ex\]](#).
- [43] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, **JINST** **3** (2008) S08003. 437 p.
- [44] CMS Collaboration, *The CMS Experiment at the CERN LHC*, **JINST** **3** (2008) S08004.
- [45] LHCb Collaboration, *The LHCb Detector at the LHC*, **JINST** **3** (2008) S08005.
- [46] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, **JINST** **3** (2008) S08002.
- [47] C. Lefèvre, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, 2008. CERN-DI-0812015, <https://cds.cern.ch/record/1260465>.
- [48] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/>

BIBLIOGRAPHY

- [LuminosityPublicResultsRun2?rev=82](#). Accessed: 12-07-23.
- [49] T. Gleisberg, S. Hoeche, F. Krauss, et al., *Event generation with SHERPA 1.1*, **JHEP** **02** (2009) 007, [arXiv:0811.4622 \[hep-ph\]](#).
- [50] R. Assmann, P. McIntosh, A. Fabris, et al., *Operational beta* levelling at the LHC in 2022 and beyond*, **JACoW IPAC2023** (2023) MOPL045.
- [51] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, CERN, Geneva, 2019. tech. rep., ATLAS-CONF-2019-021, <https://cds.cern.ch/record/2677054>.
- [52] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, [arXiv:2212.09379 \[hep-ex\]](#).
- [53] A. Bingül, *The ATLAS TRT and its Performance at LHC*, **Journal of Physics: Conference Series** **347** no. 1, (2012) 012025.
- [54] *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV*, CERN, Geneva, Jul, 2015. Tech. Rep. ATL-PHYS-PUB-2015-018, ATL-PHYS-PUB-2015-018, <https://cds.cern.ch/record/2037683>.
- [55] ATLAS Collaboration, A. Henriques, *The ATLAS tile calorimeter*, in *4th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications*, p. , 7465554. 2015.
- [56] J. del Peso, *Design and Performance of the ATLAS LAr Calorimeter*.
- [57] H. Abreu et al., *Performance of the electronic readout of the ATLAS liquid argon calorimeters*, **JINST** **5** (2010) P09003.
- [58] J. Pequeno, *Computer generated image of the ATLAS Muons subsystem*, 2008. CERN-GE-0803017, <https://cds.cern.ch/record/1095929>.
- [59] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, **JINST** **13** no. 07, (2018) P07017.
- [60] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, **JINST** **15** no. 10, (2020) P10004, [arXiv:2007.12539 \[physics.ins-det\]](#).
- [61] J. Pequeno, *Event Cross Section in a computer generated image of the ATLAS detector*, 2008. CERN-GE-0803022, <https://cds.cern.ch/record/1096081>.

BIBLIOGRAPHY

- [62] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** no. 7, (2017) 466, [arXiv:1703.10485 \[hep-ex\]](#).
- [63] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_T jet clustering algorithm*, *JHEP* **04** (2008) 063, [arXiv:0802.1189 \[hep-ph\]](#).
- [64] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** no. 11, (2016) 581, [arXiv:1510.03823 \[hep-ex\]](#).
- [65] ATLAS Collaboration, *Forward jet vertex tagging using the particle flow algorithm*, 2019. tech. rep., ATL-PHYS-PUB-2019-026.
- [66] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** no. 8, (2021) 689, [arXiv:2007.02645 \[hep-ex\]](#).
- [67] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett. B* **659** (2008) 119, [arXiv:0707.1378](#).
- [68] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector*, ATL-CONF-2015-029, 2015. <https://cds.cern.ch/record/2037702>.
- [69] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** no. 11, (2019) 970, [arXiv:1907.05120 \[hep-ex\]](#).
- [70] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, *Eur. Phys. J. C* **83** no. 7, (2023) 681, [arXiv:2211.16345 \[physics.data-an\]](#).
- [71] ATLAS Collaboration, *Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector*, 7, 2017. tech. rep., ATL-PHYS-PUB-2017-017.
- [72] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** no. 8, (2019) 639, [arXiv:1902.04655 \[physics.ins-det\]](#).
- [73] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **81** no. 7,

BIBLIOGRAPHY

- (2021) 578, [arXiv:2012.00578 \[hep-ex\]](#).
- [74] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*, CERN, Geneva, 2017. tech. rep., ATLAS-CONF-2017-029, <https://cds.cern.ch/record/2261772>.
- [75] ATLAS Collaboration, *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, *Eur. Phys. J. C* **76** no. 5, (2016) 295, [arXiv:1512.05955 \[hep-ex\]](#).
- [76] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATLAS-PHYS-PUB-2019-033, 2019. <https://cds.cern.ch/record/2688062>.
- [77] G. Aad, B. Abbott, D. Abbott, et al., *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, *Journal of Instrumentation* **14** no. 12, (2019) P12006–P12006.
- [78] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **78** no. 11, (2018) 903, [arXiv:1802.08168 \[hep-ex\]](#).
- [79] ATLAS Collaboration, *Object-based missing transverse momentum significance in the ATLAS detector*, 7, 2018. tech. rep., ATLAS-CONF-2018-038.
- [80] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, *JINST* **15** no. 04, (2020) P04003, [arXiv:1911.04632 \[physics.ins-det\]](#).
- [81] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, [arXiv:1005.4568 \[physics.ins-det\]](#).
- [82] GEANT4 Collaboration, *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.
- [83] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823–874, [arXiv:1005.4568 \[physics.ins-det\]](#).
- [84] M. Baak, G. J. Besjes, D. Côte, et al., *HistFitter software framework for statistical data analysis*, *Eur. Phys. J. C* **75** (2015) 153, [arXiv:1410.1280 \[hep-ex\]](#).

BIBLIOGRAPHY

- [85] L. Heinrich, M. Feickert, and G. Stark, *pyhf: v0.6.3*,
<https://doi.org/10.5281/zenodo.1169739>,
<https://github.com/scikit-hep/pyhf/releases/tag/v0.6.3>.
- [86] ROOT Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, New York U., New York, 2012. tech. rep., CERN-OPEN-2012-016,
<https://cds.cern.ch/record/1456844>.
- [87] G. Bohm and G. Zech, *Statistics of weighted Poisson events and its applications*,
Nucl. Instrum. Meth. A **748** (2014) 1–6, [arXiv:1309.1287 \[physics.data-an\]](#).
- [88] F. James, *MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1*, 1994. tech. rep., CERN-D-506, CERN-D506,
<https://cds.cern.ch/record/2296388?ln=en>.
- [89] P. Virtanen, R. Gommers, T. E. Oliphant, et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Methods* **17** (2020) 261–272.
- [90] J. Neyman and E. S. Pearson, *On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I*, *Biometrika* **20A** no. 1/2, (1928) 175–240.
- [91] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554,
[arXiv:1007.1727 \[physics.data-an\]](#).
- [92] A. L. Read, *Modified frequentist analysis of search results (the CL_s method)*, 2000. Tech. Rep. CERN-OPEN-2000-205, CERN-OPEN-2000-205,
<https://cds.cern.ch/record/451614>.
- [93] C. G. Lester and D. J. Summers, *Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders*, *Phys. Lett. B* **463** (1999) 99–103,
[arXiv:hep-ph/9906349](#).
- [94] A. Barr, C. Lester, and P. Stephens, *m(T2) : The Truth behind the glamour*, *J. Phys. G* **29** (2003) 2343–2363, [arXiv:hep-ph/0304226](#).
- [95] M. Feickert and B. Nachman, *A Living Review of Machine Learning for Particle Physics*, [arXiv:2102.02770 \[hep-ph\]](#).
- [96] M. Belkin, D. Hsu, S. Ma, and S. Mandal, *Reconciling modern machine-learning practice and the classical bias–variance trade-off*, *Proceedings of the National*

BIBLIOGRAPHY

- Academy of Sciences* **116** no. 32, (2019) 15849–15854.
- [97] S. M. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*, pp. , 4765–4774. Curran Associates, Inc., 2017. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
- [98] S. M. Lundberg, G. Erion, H. Chen, et al., *From local explanations to global understanding with explainable AI for trees*, *Nature Machine Intelligence* **2** no. 1, (2020) 2522–5839.
- [99] G. Ke, Q. Meng, T. Finley, et al., *Lightgbm: A highly efficient gradient boosting decision tree*, *Advances in neural information processing systems* **30** (2017) 3146–3154.
- [100] D. Alborno Vázquez, G. Bélanger, and C. Boehm, *Revisiting light neutralino scenarios in the MSSM*, *Phys. Rev. D* **84** (2011) 095015, [arXiv:1108.1338 \[hep-ph\]](#).
- [101] M. Chakraborti, S. Heinemeyer, and I. Saha, $(g - 2)_\mu$ and Stau coannihilation : *Dark Matter and Collider Analysis*, [arXiv:2308.05723 \[hep-ph\]](#).
- [102] ATLAS Collaboration, *Search for direct stau production in events with two hadronic τ -leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, *Phys. Rev. D* **101** (2020) 032009, [arXiv:1911.06660 \[hep-ex\]](#).
- [103] CMS Collaboration, *Search for direct pair production of supersymmetric partners of τ leptons in the final state with two hadronically decaying τ leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **108** no. 1, (2023) 012011, [arXiv:2207.02254 \[hep-ex\]](#).
- [104] CMS Collaboration, *Search for direct pair production of supersymmetric partners to the τ lepton in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **80** no. 3, (2020) 189, [arXiv:1907.13179 \[hep-ex\]](#).
- [105] ALEPH Collaboration, *Search for scalar leptons in $e^+ e^-$ collisions at center-of-mass energies up to 209-GeV*, *Phys. Lett. B* **526** (2002) 206–220, [arXiv:hep-ex/0112011](#).
- [106] DELPHI Collaboration, *Searches for supersymmetric particles in $e^+ e^-$ collisions up to 208-GeV and interpretation of the results within the MSSM*, *Eur. Phys. J. C* **31** (2003) 421–479, [arXiv:hep-ex/0311019](#).
- [107] L3 Collaboration, *Search for scalar leptons and scalar quarks at LEP*, *Phys. Lett. B*

BIBLIOGRAPHY

- 580 (2004) 37–49, [arXiv:hep-ex/0310007](#).
- [108] OPAL Collaboration, *Search for anomalous production of dilepton events with missing transverse momentum in e^+e^- collisions at $\sqrt{s} = 183$ GeV to 209 GeV*, [Eur. Phys. J. C **32** \(2004\) 453–473](#), [arXiv:hep-ex/0309014](#).
- [109] J. Alwall, R. Frederix, S. Frixione, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, [JHEP **07** \(2014\) 079](#), [arXiv:1405.0301 \[hep-ph\]](#).
- [110] T. Sjöstrand, S. Ask, J. R. Christiansen, et al., *An introduction to PYTHIA 8.2*, [Comput. Phys. Commun. **191** \(2015\) 159–177](#), [arXiv:1410.3012 \[hep-ph\]](#).
- [111] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014. <https://cds.cern.ch/record/1966419>.
- [112] B. Fuks, M. Klasen, D. R. Lamprea, and M. Rothering, *Gaugino production in proton-proton collisions at a center-of-mass energy of 8 TeV*, [JHEP **10** \(2012\) 081](#), [arXiv:1207.2159 \[hep-ph\]](#).
- [113] B. Fuks, M. Klasen, D. R. Lamprea, and M. Rothering, *Precision predictions for electroweak superpartner production at hadron colliders with Resummino*, [Eur. Phys. J. C **73** \(2013\) 2480](#), [arXiv:1304.0790 \[hep-ph\]](#).
- [114] C. Borschensky, M. Krämer, A. Kulesza, et al., *Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33$ and 100 TeV*, [Eur. Phys. J. C **74** \(2014\) 3174](#), [arXiv:1407.5066 \[hep-ph\]](#).
- [115] T. Gleisberg et al., *Event generation with SHERPA 1.1*, [JHEP **02** \(2009\) 007](#), [arXiv:0811.4622 \[hep-ph\]](#).
- [116] S. Höche, F. Krauss, M. Schonherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, [JHEP **04** \(2013\) 027](#), [arXiv:1207.5030 \[hep-ph\]](#).
- [117] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, [JHEP **12** \(2008\) 039](#), [arXiv:0808.3674 \[hep-ph\]](#).
- [118] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, [Phys. Rev. Lett. **108** \(2012\) 111601](#), [arXiv:1111.5206 \[hep-ph\]](#).
- [119] A. Denner, S. Dittmaier, and L. Hofer, *Collier: a fortran-based Complex One-Loop Library in Extended Regularizations*, [Comput. Phys. Commun. **212** \(2017\) 220–238](#),

BIBLIOGRAPHY

- [arXiv:1604.06792 \[hep-ph\]](#).
- [120] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, **JHEP** **03** (2008) 038, [arXiv:0709.1027 \[hep-ph\]](#).
- [121] R. D. Ball et al., *Parton distributions for the LHC Run II*, **JHEP** **04** (2015) 040, [arXiv:1410.8849 \[hep-ph\]](#).
- [122] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini, *Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order*, **Phys. Rev. Lett.** **103** (2009) 082001, [arXiv:0903.2120 \[hep-ph\]](#).
- [123] S. Höche, F. Krauss, M. Schonherr, and F. Siegert, *A critical appraisal of NLO+PS matching methods*, **JHEP** **09** (2012) 049, [arXiv:1111.1220 \[hep-ph\]](#).
- [124] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *QCD matrix elements + parton showers*, **JHEP** **11** (2001) 063, [arXiv:hep-ph/0109231](#).
- [125] S. Höche, F. Krauss, S. Schumann, and F. Siegert, *QCD matrix elements and truncated showers*, **JHEP** **05** (2009) 053, [arXiv:0903.1219 \[hep-ph\]](#).
- [126] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, **JHEP** **06** (2002) 029, [arXiv:hep-ph/0204244](#).
- [127] G. Corcella, I. G. Knowles, G. Marchesini, et al., *HERWIG 6.5 release note*, [arXiv:hep-ph/0210213](#).
- [128] J. M. Campbell and R. K. Ellis, *An Update on vector boson pair production at hadron colliders*, **Phys. Rev. D** **60** (1999) 113006, [arXiv:hep-ph/9905386](#).
- [129] T. Binoth, M. Ciccolini, N. Kauer, and M. Kramer, *Gluon-induced W-boson pair production at the LHC*, **JHEP** **12** (2006) 046, [arXiv:hep-ph/0611170](#).
- [130] T. Binoth, N. Kauer, and P. Mertsch, *Gluon-induced QCD corrections to $pp \rightarrow ZZ \rightarrow \ell\bar{\ell}\ell'\bar{\ell}'$* , p. , 142. 7, 2008. [arXiv:0807.0024 \[hep-ph\]](#).
- [131] S. Frixione, P. Nason, and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, **JHEP** **09** (2007) 126, [arXiv:0707.3088 \[hep-ph\]](#).
- [132] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, **JHEP** **11** (2004) 040, [arXiv:hep-ph/0409146](#).

BIBLIOGRAPHY

- [133] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, **JHEP** **11** (2007) 070, [arXiv:0709.2092 \[hep-ph\]](#).
- [134] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, **JHEP** **06** (2010) 043, [arXiv:1002.2581 \[hep-ph\]](#).
- [135] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, **JHEP** **07** (2014) 079, [arXiv:1405.0301 \[hep-ph\]](#).
- [136] R. D. Ball et al., *Parton distributions with LHC data*, **Nucl. Phys. B** **867** (2013) 244, [arXiv:1207.1303 \[hep-ph\]](#).
- [137] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, *Hadronic top-quark pair production with NNLL threshold resummation*, **Nucl. Phys. B** **855** (2012) 695–741, [arXiv:1109.1536 \[hep-ph\]](#).
- [138] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, **Phys. Lett. B** **710** (2012) 612–622, [arXiv:1111.5869 \[hep-ph\]](#).
- [139] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \rightarrow t\bar{t} + X$* , **Phys. Rev. Lett.** **109** (2012) 132001, [arXiv:1204.5201 \[hep-ph\]](#).
- [140] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, **JHEP** **12** (2012) 054, [arXiv:1207.0236 \[hep-ph\]](#).
- [141] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, **JHEP** **01** (2013) 080, [arXiv:1210.6832 \[hep-ph\]](#).
- [142] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_s^4)$* , **Phys. Rev. Lett.** **110** (2013) 252004, [arXiv:1303.6254 \[hep-ph\]](#).
- [143] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, **Comput. Phys. Commun.** **185** (2014) 2930,

BIBLIOGRAPHY

- [arXiv:1112.5675 \[hep-ph\]](#).
- [144] M. Aliev, H. Lacker, U. Langenfeld, et al., *HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR*, *Comput. Phys. Commun.* **182** (2011) 1034–1046, [arXiv:1007.1327 \[hep-ph\]](#).
- [145] P. Kant et al., *HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, *Comput. Phys. Commun.* **191** (2015) 74, [arXiv:1406.4403 \[hep-ph\]](#).
- [146] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 2016. [arXiv:1610.07922 \[hep-ph\]](#).
- [147] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [arXiv:0710.3820 \[hep-ph\]](#).
- [148] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016. <https://cds.cern.ch/record/2206965>.
- [149] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, *JINST* **15** (2020) P04003, [arXiv:1911.04632 \[physics.ins-det\]](#).
- [150] L. McInnes, J. Healy, and J. Melville, *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, 2018. <https://arxiv.org/abs/1802.03426>.
- [151] ATLAS Collaboration, *Formulae for Estimating Significance*, CERN, Geneva, 2020. tech. rep., ATL-PHYS-PUB-2020-025, <https://cds.cern.ch/record/2736148>.
- [152] ATLAS Collaboration, *Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector*, *Phys. Rev. D* **103** no. 11, (2021) 112006, [arXiv:2102.10874 \[hep-ex\]](#).
- [153] CMS Collaboration, *Search for supersymmetry with a compressed mass spectrum in the vector boson fusion topology with 1-lepton and 0-lepton final states in proton-proton collisions at $\sqrt{s}=13$ TeV*, *JHEP* **08** (2019) 150, [arXiv:1905.13059 \[hep-ex\]](#).
- [154] ATLAS Collaboration, *Search for long-lived charginos based on a disappearing-track signature using 136 fb^{-1} of pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector*, *Eur.*

BIBLIOGRAPHY

- Phys. J. C **82** no. 7, (2022) 606, [arXiv:2201.02472 \[hep-ex\]](#).
- [155] CMS Collaboration, *Search for disappearing tracks in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **806** (2020) 135502, [arXiv:2004.05153 \[hep-ex\]](#).
- [156] ATLAS Collaboration, *Searches for electroweak production of supersymmetric particles with compressed mass spectra in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, Phys. Rev. D **101** no. 5, (2020) 052005, [arXiv:1911.12606 \[hep-ex\]](#).
- [157] Lep2 susy working group - combined lep chargino results, up to 208 gev for large m_0 , http://lepsusy.web.cern.ch/lepsusy/www/inos_moriond01/charginos_pub.html.
- [158] H. Fukuda, N. Nagata, H. Oide, H. Otono, and S. Shirai, *Cornering Higgsinos Using Soft Displaced Tracks*, Phys. Rev. Lett. **124** no. 10, (2020) 101801, [arXiv:1910.08065 \[hep-ph\]](#).
- [159] ATLAS Collaboration, *Search for invisible Higgs-boson decays in events with vector-boson fusion signatures using 139 fb^{-1} of proton-proton data recorded by the ATLAS experiment*, JHEP **08** (2022) 104, [arXiv:2202.07953 \[hep-ex\]](#).
- [160] ATLAS Collaboration, *SUSY VBF Check*, https://indico.cern.ch/event/939498/contributions/3947455/attachments/2074314/3485665/vbfHInv_susy.pdf. Accessed: 21-03-2024. (Accessible only by members of the ATLAS collaboration).
- [161] ATLAS Collaboration, *Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum using 4.7 fb^{-1} of $\sqrt{s} = 7$ TeV proton-proton collision data*, Phys. Rev. D **87** no. 1, (2013) 012008, [arXiv:1208.0949 \[hep-ex\]](#).
- [162] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015, [arXiv:1212.3460 \[hep-ph\]](#).
- [163] ATLAS Collaboration, *SimpleAnalysis: Truth-level Analysis Framework*, CERN, Geneva, 2022. tech. rep., ATL-PHYS-PUB-2022-017, <https://cds.cern.ch/record/2805991>.
- [164] ATLAS Collaboration, *RECAST framework reinterpretation of an ATLAS Dark Matter Search constraining a model of a dark Higgs boson decaying to two b-quarks*,

BIBLIOGRAPHY

- CERN, Geneva, 2019. tech. rep., ATL-PHYS-PUB-2019-032, <https://cds.cern.ch/record/2686290>.
- [165] ATLAS Collaboration, *Search for chargino–neutralino pair production in final states with three leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, *Eur. Phys. J. C* **81** no. 12, (2021) 1118, [arXiv:2106.01676](https://arxiv.org/abs/2106.01676) [hep-ex].
- [166] LEP SUSY Working Group (ALEPH, DELPHI, L3, OPAL), Notes LEPSUSYWG/01-03.1 and 04-01.1, <http://lepsusy.web.cern.ch/lepsusy/Welcome.html>.
- [167] S. Kallweit, J. M. Lindert, P. Maierhöfer, S. Pozzorini, and M. Schönherr, *NLO electroweak automation and precise predictions for W+multijet production at the LHC*, *JHEP* **04** (2015) 012, [arXiv:1412.5157](https://arxiv.org/abs/1412.5157) [hep-ph].
- [168] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, *Phys. Rev. D* **69** (2004) 094008, [arXiv:hep-ph/0312266](https://arxiv.org/abs/hep-ph/0312266).
- [169] ATLAS Collaboration, *Combination of searches for invisible decays of the Higgs boson using 139 fb⁻¹ of proton-proton collision data at $s=13$ TeV collected with the ATLAS experiment*, *Phys. Lett. B* **842** (2023) 137963, [arXiv:2301.10731](https://arxiv.org/abs/2301.10731) [hep-ex].
- [170] L. M. Carpenter, H. Gilmer, and J. Kawamura, *Exploring nearly degenerate higgsinos using mono-Z/W signal*, *Physics Letters B* **831** (2022) 137191.
- [171] ATLAS Collaboration, *Search for electroweak production of charginos and sleptons decaying into final states with two leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 123, [arXiv:1908.08215](https://arxiv.org/abs/1908.08215) [hep-ex].
- [172] ATLAS Collaboration, *Search for direct pair production of sleptons and charginos decaying to two leptons and neutralinos with mass splittings near the W-boson mass in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, *JHEP* **06** (2023) 031, [arXiv:2209.13935](https://arxiv.org/abs/2209.13935) [hep-ex].
- [173] ATLAS Collaboration, *Search for direct production of electroweakinos in final states with one lepton, jets and missing transverse momentum and in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, CERN, Geneva, 2022. tech. rep., ATLAS-CONF-2022-059, <https://cds.cern.ch/record/2826702>.

BIBLIOGRAPHY

- [174] ATLAS Collaboration, *Search for charginos and neutralinos in final states with two boosted hadronically decaying bosons and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **104** no. 11, (2021) 112010, [arXiv:2108.07586 \[hep-ex\]](#).
- [175] CMS Collaboration, *Combined search for electroweak production of winos, binos, higgsinos, and sleptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*,.
- [176] CMS Collaboration, *Search for electroweak production of charginos and neutralinos at $\sqrt{s}=13$ TeV in final states containing hadronic decays of WW , WZ , or WH and missing transverse momentum*, *Phys. Lett. B* **842** (2023) 137460, [arXiv:2205.09597 \[hep-ex\]](#).
- [177] A. Buckley, *The hepthesis L^AT_EX class*, 2017. <https://ctan.org/tex-archive/macros/latex/contrib/hepthesis?lang=en>.

List of figures

1.1	Standard Model cross-section measurements	6
1.2	EM interaction vertex	11
1.3	Strong interaction vertices	11
1.4	Weak interaction vertices	12
1.5	Higgs boson interaction vertices	13
1.6	Feynman diagrams showing one loop corrections to the Higgs boson mass	22
1.7	Unification of gauge couplings	29
2.1	Schematic of the accelerators at CERN	35
2.2	Diagram of a proton-proton collision	36
2.3	Cutaway of the ATLAS detector	40
2.4	Cutaway of the Inner Detector	43
2.5	Layout of the ATLAS calorimeters	44
2.6	Diagram of the Muon Spectrometer	46
2.7	Schematic of the ATLAS trigger system	49
2.8	Signatures different particles leave in the ATLAS detector	50
3.1	Example Receiver Operating Characteristic curve	76
3.2	Example SHAP values for a given event	78

LIST OF FIGURES

3.3	Schematic of Cross Evaluation	79
3.4	Plot demonstrating how regions of artificially high S/B can arise . . .	81
3.5	Diagram of a single decision tree	83
3.6	Diagram of a Boosted Decision Tree	84
4.1	Direct stau: Feynman diagram	90
4.2	Direct stau: Previous ATLAS search limits	93
4.3	Direct stau BDT signal grouping	103
4.4	Direct stau BDT example train/test comparisons	112
4.5	Direct stau BDTs example ROC curves	113
4.6	Direct stau cross evaluation fold comparison	114
4.7	Direct stau BDT score distributions	115
4.9	Direct stau N-1 significance score plots of BDT score distributions . . .	119
4.10	Direct stau BDT binned significance score plots	120
4.11	Direct stau Background Composition	123
4.12	Direct stau nuisance parameter pulls for the background-only fit to the CRs	125
4.13	Direct stau nuisance parameter correlation matrix for the background- only fit to the CRs	126
4.14	Direct stau validation region summary plot	126
4.15	Direct stau multi-jet estimation schematic	131
4.16	Direct stau m_{T2} correlation check	132
4.17	Direct stau MJ VR-F kinematics	134
4.18	Direct stau MJ pre-SR D kinematics	135
4.19	Direct stau MJVR kinematics	136
4.20	Direct stau ZCR kinematics	138

LIST OF FIGURES

4.21	Direct stau ZVR kinematics	139
4.22	Direct stau WCR kinematics	142
4.23	Direct stau WVR kinematics	143
4.24	Direct stau TCR kinematics	145
4.25	Direct stau TVR kinematics	146
4.26	Direct stau MBVR kinematics	148
4.27	Direct stau InclVR kinematics	150
4.28	Direct stau pull plot of SRs	158
4.29	Direct stau SR overlap grids	161
4.30	Direct stau NP pull plot for SR-BDT1	162
4.31	Direct stau NP pull plot for SR-BDT2	163
4.32	Direct stau NP pull plot for SR-BDT3	164
4.33	Direct stau NP pull plot for SR-BDT4	165
4.34	Direct stau NP correlation matrix for SR-BDT1	166
4.35	Direct stau NP correlation matrix for SR-BDT2	166
4.36	Direct stau NP correlation matrix for SR-BDT3	167
4.37	Direct stau NP correlation matrix for SR-BDT4	167
4.38	DS Unblinded BDT plots	168
4.39	Direct stau SR-BDT-1 kinematics	169
4.40	Direct stau SR-BDT-3 kinematics	170
4.41	Direct stau $\tilde{\tau}_{L,R}$ exclusion contours per signal region	174
4.42	Direct stau $\tilde{\tau}_L$ exclusion contours per signal region	175
4.43	Direct stau $\tilde{\tau}_R$ exclusion contours per signal region	176
4.44	Direct stau exclusion contours for combined best expected signal region	177
4.45	Direct stau best expected signal region breakdown	178

LIST OF FIGURES

5.1	Current ATLAS limits on Compressed Mass Electroweak SUSY models	184
5.2	Compressed Mass models Feynman diagrams	186
5.3	Mono-Jet signal kinematics	192
5.4	Exclusion contours from mono-jet re-interpretation	194
5.5	Reconstructed E_T^{miss} distribution in the mono-jet signal region	195
5.6	Example non-VBF and VBF Feynman diagrams of $\tilde{\chi}_2^0\text{-}\tilde{\chi}_1^\pm$ production .	200
5.7	VBF BDT Example ROC curve	210
5.8	VBF BDT Example Train/Test comparison	211
5.9	VBF BDT cross evaluation fold comparison	212
5.10	VBF full BDT score distribution	213
5.11	VBF BDT score shape comparison	213
5.12	VBF BDT Example SHAP values comparison	214
5.13	VBF N-1 significance score plots of discriminating variables	217
5.14	VBF binned S/B plots for discriminating variables	218
5.15	VBF binned significance scores for discriminating variables	221
5.16	VBF CR-1L-El kinematics	226
5.17	VBF CR-1L-Mu kinematics	227
5.18	VBF CR-2L-El kinematics	228
5.19	VBF CR-2L-Mu kinematics	229
5.20	VBF CMS signal region yields comparison	236
6.1	Feynman diagram of chargino pair-production	244
6.2	C1C1WW combination overlay of exclusion contours of each analysis .	246
6.3	C1C1WW combination CMS limits	246
6.4	C1C1WW combination: comparison of recalculated and original expected CL_s values for each analysis	249

LIST OF FIGURES

6.5	C1C1WW combination: comparison of recalculated and original observed CL_s values for each analysis	250
6.6	C1C1WW combination: points available for each analysis	251
6.7	C1C1WW combination: Difference in UL from correlating systematic uncertainties	255
6.8	C1C1WW combination: expected nuisance parameter pulls comparison	256
6.9	C1C1WW combination: observed nuisance parameter pulls comparison	257
6.10	C1C1WW combination: combination exclusion contours	258
6.11	C1C1WW combination: difference in upper limit from combination . .	259
A.1	Direct stau BDT input features shape comparisons	268
A.2	Direct stau: train/test comparisons for fold splitting 2	270
A.3	Direct stau: train/test comparisons for fold splitting 3	271
A.4	Direct stau cross evaluation fold comparison including negative and largely weighted events	272
A.5	Direct stau medium vs tight tau-ID comparison fine binning	275
A.6	Direct stau medium vs tight tau-ID comparison coarse binning	276
A.7	Direct stau multi-jet comparison plots	277
A.8	Direct stau cross-section upper limits	280
A.9	VBF BDT input feature shape comparisons	285
A.10	VBF BDT additional train/test comparisons	286
A.11	C1C1WW combination: cross-section upper limits	289

List of tables

1.1	Particles in the Standard Model	10
1.2	Gauge group representations of the particles in the Standard Model . .	16
1.3	Table of particles in the MSSM	25
4.1	Direct stau simulation details	98
4.2	Direct stau trigger thresholds	100
4.3	Direct stau preliminary selections	101
4.4	Direct stau BDT preselection yields	104
4.5	Direct stau signal grouping comparison hyperparameters	108
4.6	Direct stau BDT hyperparameter scan ranges	110
4.7	Direct stau optimum hyperparameter configurations	111
4.8	Direct Stau AUC scores	114
4.9	Direct stau signal region definitions	121
4.10	Direct stau signal region yields table	121
4.11	Direct stau control and validation region definitions	124
4.12	Direct stau yields table for ZCR, ZVR	127
4.13	Direct stau yields table for WCR, WVR	127
4.14	Direct stau yields table for TCR, TVR	128
4.15	Direct stau yields table for MJVR, MBVR, InclVR	128

LIST OF TABLES

4.16	Direct stau multi-jet control region definitions	130
4.17	Direct stau multi-jet control regions yields table	132
4.18	Direct stau systematic uncertainties summary	155
4.19	Direct stau signal systematic uncertainties summary	157
4.20	Direct stau yields table for SR-BDT1	159
4.21	Direct stau yields table for SR-BDT2	159
4.22	Direct stau yields table for SR-BDT3	160
4.23	Direct stau yields table for SR-BDT4	160
4.24	Direct stau model-independent upper limits.	171
5.1	Mono-jet signal cross-sections	191
5.2	Mono-jet signal strength upper limits	193
5.3	Mono-jet signal region total yields	196
5.4	Mono-jet yields in the last three signal region bins	197
5.5	VBF signal cross-sections	201
5.6	VBF simulation details	202
5.7	VBF signal region selections	205
5.8	VBF BDT preselection yields table	206
5.9	VBF BDT hyperparameter table	209
5.10	VBF AUC scores	210
5.11	VBF signal regions total yields	215
5.12	VBF yields for the last three SR bins for the fit using M_{jj} using the binning in the CMS analysis.	219
5.13	VBF yields for the last three SR bins for the fit using M_{jj} using 8 bins. .	219
5.14	VBF yields for the last three SR bins for the fit using E_T^{miss}	220
5.15	VBF yields for the last three SR bins for the fit using the BDT score. . .	220

LIST OF TABLES

5.16	VBF CR Selections	223
5.17	VBF V+jet data-driven correction factors	225
5.18	VBF control region yields table	225
5.19	VBF V+jets systematic uncertainty configurations	233
5.20	VBF upper limits	235
5.21	VBF ATLAS and CMS background yields comparison	237
6.1	C1C1WW combination: jet energy uncertainty correlation schemes . .	253
A.2	Direct stau RNN tau-ID WP details	273
A.3	Direct stau medium vs tight yield comparison	274
A.4	Direct stau tight comparison max significance scores	275
A.5	Direct stau signal cut flows	278
A.6	Direct stau background and data cut flow	279
A.7	VBF signal cut flow table	287
A.8	VBF background cut flow table	288