

Measurements of  $H \rightarrow b\bar{b}$  decays and  $VH$   
production with the ATLAS detector at the  
Large Hadron Collider in proton-proton collisions  
at  $\sqrt{S} = 13$  TeV

Thomas Paul Charman

Supervised by Dr. Jonathan Hays



Queen Mary University of London

Submitted in partial fulfillment of the requirements of the Degree of  
Doctor of Philosophy 31st August 2021.

I, Thomas Paul Charman, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis. I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date: 31st August 2021

Details of collaboration and publications:

## Abstract

Proton-proton collisions with centre of mass energy of 13 TeV are analysed in order to measure the  $VH \rightarrow b\bar{b}$  process. These collisions are provided by the Large Hadron Collider and measured by the ATLAS detector both of which are based at CERN. The  $VH \rightarrow b\bar{b}$  process occurs when a Higgs boson is produced in association with a vector boson and decays to a pair of  $b$ -quarks. Measurement of the  $VH \rightarrow b\bar{b}$  process serves the purpose of furthering our understanding of the Higgs by providing more precise measurements of the  $VH$  signal strength as well as the individual  $WH$  and  $ZH$  signal strengths, testing the predictions of The Standard Model of Particle Physics. The analysis uses a number of by-hand categorisations as well as a multi-variate classifier to obtain distributions that enter into a profile-likelihood fit. Systematic uncertainties are considered in the fit as nuisance parameters. The results of the analysis performed using the full run 2 dataset ( $140 \text{ fb}^{-1}$  recorded between 2015–2018) agree with predictions for signal strength and cross-section in a number of fiducial regions.

Numerical result?

## Acknowledgements

- Thank you to Professor Jon Hays for your years of supervision and guidance.
- Thank you to Professor Adrian Bevan, for starting me on my journey in research.
- Thank you to Jack Cribben, Mattea Todd, Henry Preston and Natasha Westland for your years of love, friendship and support.
- Thank you to Stephen Jiggins for supporting me during my time at CERN.
- Thank you to Mum, Dad and James for your love and care that I can always depend on.

# Contents

<b>Abstract</b>	<b>3</b>
<b>Acknowledgements</b>	<b>4</b>
<b>Contents</b>	<b>5</b>
<b>List of Figures</b>	<b>8</b>
<b>List of Tables</b>	<b>12</b>
<b>1 Introduction</b>	<b>15</b>
<b>2 Physics Theory</b>	<b>18</b>
2.1 Historical Aside . . . . .	22
2.2 Quantum Electrodynamics . . . . .	23
2.3 Quantum Chromodynamics . . . . .	25
2.4 Electroweak theory . . . . .	26
2.5 The Brout-Englert-Higgs Mechanism . . . . .	28
2.6 Higgs Bosons at the LHC . . . . .	30
<b>3 The ATLAS Detector at the Large Hadron Collider</b>	<b>32</b>
3.1 The Large Hadron Collider . . . . .	32
3.2 The ATLAS Detector . . . . .	36
<b>4 Machine Learning</b>	<b>49</b>
4.1 Boosted Decision Trees . . . . .	50

<b>CONTENTS</b>	<b>6</b>
<b>4.2 Neural Networks</b>	<b>54</b>
<b>5 Reconstruction and Selection</b>	<b>57</b>
<b>5.1 Athena and the CxAOD Framework</b>	<b>57</b>
<b>5.2 Leptons</b>	<b>60</b>
<b>5.3 Triggers</b>	<b>62</b>
<b>5.4 Jets</b>	<b>64</b>
<b>5.5 Missing Transverse Momentum</b>	<b>72</b>
<b>5.6 Overlap Removal</b>	<b>72</b>
<b>5.7 Final Selection</b>	<b>73</b>
<b>6 Analysis Strategy</b>	<b>77</b>
<b>6.1 Categorisation Into Analysis Regions</b>	<b>78</b>
<b>6.2 Composition of Analysis Regions</b>	<b>79</b>
<b>6.3 Multi-variate Event Classification</b>	<b>81</b>
<b>6.4 Profile Likelihood Fit</b>	<b>85</b>
<b>6.5 Pre-fit Data Versus Prediction</b>	<b>87</b>
<b>6.6 Analysis Cross-checks</b>	<b>90</b>
<b>7 Systematic Uncertainties</b>	<b>91</b>
<b>7.1 Parametrising Variance Due To Shape Uncertainties</b>	<b>92</b>
<b>7.2 Experimental Systematic Uncertainties</b>	<b>96</b>
<b>7.3 Systematic Uncertainties on <math>V+jets</math> Events</b>	<b>103</b>
<b>7.4 Systematic Uncertainties on <math>t\bar{t}</math> Events</b>	<b>125</b>
<b>7.5 Systematic Uncertainties on Sub-Dominant Backgrounds</b>	<b>131</b>
<b>7.6 Systematic Uncertainties on the Signal Process</b>	<b>134</b>
<b>8 Results</b>	<b>138</b>
<b>8.1 Nuisance Parameter Pulls</b>	<b>138</b>
<b>8.2 Breakdown and Ranking of Uncertainties</b>	<b>142</b>
<b>8.3 Post-fit Data Versus Predictions</b>	<b>145</b>
<b>8.4 Signal Strength and STXS Measurements</b>	<b>152</b>

<b>9 Conclusion</b>	<b>157</b>
9.1 Future studies . . . . .	158
<b>Bibliography</b>	<b>160</b>
<b>A Supplement to Modelling of Systematic Uncertainties</b>	<b>181</b>
A.1 Extra BDT input distributions . . . . .	181
A.2 Prefit distributions . . . . .	181
<b>B Supplement to Modelling of Systematic Uncertainties</b>	<b>191</b>
B.1 Full sample tables . . . . .	191
B.2 Summary Tables . . . . .	202
B.3 Extrapolation Uncertainties Induced by Shape Systematics . . . . .	202
<b>C Truth Tagging Probability</b>	<b>204</b>
<b>D BDT Transformation</b>	<b>205</b>

# List of Figures

1.1	A roadmap of the analysis.	17
2.1	The Higgs potential.	29
2.2	Higgs boson production mechanisms.	30
2.3	Higgs boson production cross sections and branching ratios.	31
2.4	A diagram of the $VH \rightarrow b\bar{b}$ process.	31
3.1	The CERN accelerator complex	33
3.2	Magnets in a dipole configuration.	34
3.3	Magnets in a quadrupole configuration.	35
3.4	The ATLAS Detector	37
3.5	ATLAS magnetic field	39
3.6	The geometric construction of the sagitta.	40
3.7	The geometric construction of $d_0$ and $z_0$ .	40
3.8	ATLAS inner detector	41
3.9	ATLAS long strip module	43
3.10	ATLAS strip close-up	43
3.11	ATLAS Calorimeter	45
3.12	ATLAS muon subsystem	48
4.1	The structure of a decision tree.	51
4.2	An illustration of gradient descent.	52
4.3	A depiction of a neural network.	56
5.1	Data flow central to ATLAS.	58

5.2	Data flow of the analysis. . . . .	59
5.3	Shapes of the pseudo-continuous b-tagging quantiles for different jet flavours. . . . .	68
5.4	A comparison of tagging strategies in $t\bar{t}$ events. . . . .	70
5.5	A comparison of tagging strategy in $W + \text{jets}$ events. . . . .	71
6.1	A 2-dimensional histogram of signal events in the $p_T^V$ , $\Delta R(b, \bar{b})$ plane. .	78
6.2	Examples of 2-jet BDT input distributions. . . . .	84
6.3	Pre-fit distributions in the signal region of the 0-lepton channel. . . . .	87
6.4	Pre-fit distributions in the signal region of the 1-lepton channel. . . . .	88
6.5	Pre-fit distributions in the signal region of the 2-lepton channel. . . . .	89
7.1	Derivation of $p_T^V$ shape uncertainties on $W+\text{jets}$ events in the 1-lepton channel (2-jet category). . . . .	111
7.2	Derivation of $p_T^V$ shape uncertainties on $W+\text{jets}$ events in the 1-lepton channel (3-jet category). . . . .	112
7.3	Nominal, alternative and re-weighted nominal predictions of $W+\text{jets}$ events (1-lepton channel, 2-jet category, MVA input variables). . . . .	113
7.4	Nominal, alternative and re-weighted nominal predictions of $W+\text{jets}$ events (0-lepton channel, 2-jet category, MVA input variables). . . . .	114
7.5	Nominal, alternative and re-weighted nominal predictions of $W+\text{jets}$ events (1-lepton channel, MVA score). . . . .	114
7.6	Nominal, alternative and re-weighted nominal predictions of $W+\text{jets}$ events (0-lepton channel, MVA score). . . . .	115
7.7	Subtracted data versus the nominal $Z+\text{jets}$ prediction, GSC $m_{bb}$ . . . . .	122
7.8	Subtracted data versus the nominal $Z+\text{jets}$ prediction, $p_T^V$ . . . . .	123
7.9	Subtracted data versus the nominal $Z+\text{jets}$ prediction, GSC $m_{bb}$ across different analysis $p_T^V$ bins. . . . .	124
7.10	Data versus prediction comparison of $m_{bb}$ distributions used to check how well the top $e\mu$ control region data models the shape of the $t\bar{t}$ and single top processes. . . . .	129

7.10 Data versus prediction comparison of $m_{bb}$ distributions used to check how well the top $e\mu$ control region data models the shape of the $t\bar{t}$ and single top processes. . . . .	130
8.1 Nuisance parameter pulls and free parameter scale factors relating to the $W+\text{jets}$ , $Z+\text{jets}$ and top process backgrounds of the analysis, where an Asimov dataset conditional on $\mu = 1$ in red is compared with the data in black. . . . .	140
8.2 Correlations between nuisance parameter pulls. . . . .	141
8.3 The top ranked nuisance parameters are shown according to their impact on the $VH \rightarrow b\bar{b}$ signal strength as determined by the combined unconditional fit to data. . . . .	143
8.4 The top ranked nuisance parameters are shown, omitting signal uncertainties, according to their impact on the $VH \rightarrow b\bar{b}$ signal strength as determined by the combined unconditional fit to data. . . . .	144
8.5 Post-fit distributions in the 0-lepton 2-jet channel. . . . .	146
8.6 Post-fit distributions in the 0-lepton 3-jet channel. . . . .	147
8.7 Post-fit distributions in the 1-lepton 2-jet channel. . . . .	148
8.8 Post-fit distributions in the 1-lepton 3-jet channel. . . . .	149
8.9 Post-fit distributions in the 2-lepton 2-jet channel. . . . .	150
8.10 Post-fit distributions in the 2-lepton, $\geq 3$ -jet channel. . . . .	151
8.11 Best fit values for signal strength broken down by analysis channel. . .	153
8.12 Best fit values for $WH$ , $ZH$ and $VH$ signal strength. . . . .	154
8.13 Best fit values for cross-section times branching ratio in each STXS bin. .	156
8.14 Correlations between the measurements in each of the STXS bins of the analysis for the unconditional fit to data. . . . .	156
A.1 Inputs to the multi-variate analysis in the 0-lepton 2-jet region. . . . .	182
A.2 Inputs to the multi-variate analysis in the 1-lepton 2-jet region. . . . .	183
A.3 Inputs to the multi-variate analysis in the 2-lepton 2-jet region. . . . .	184
A.4 Pre-fit distributions in the 0-lepton channel in the 2-jet region. . . . .	185

A.5	Pre-fit distributions in the 0-lepton channel in the 3-jet region. . . . .	186
A.6	Pre-fit distributions in the 1-lepton channel in the 2-jet region. . . . .	187
A.7	Pre-fit distributions in the 1-lepton channel in the 3-jet region. . . . .	188
A.8	Pre-fit distributions in the 2-lepton channel in the 2-jet region. . . . .	189
A.9	Pre-fit distributions in the 2-lepton channel in the $\geq 3$ -jet region. . . . .	190

# List of Tables

2.1	Particles of The Standard Model.	19
5.1	Lepton definitions used in the analysis.	60
5.2	Triggers used in 2015–2018 data collection.	63
5.3	Jet selection requirements.	65
5.4	MV2c10 $b$ -tagging Working Points	67
5.5	The analysis event selection.	76
6.1	The samples of Monte-Carlo simulated predictions of signal processes.	79
6.2	The samples of Monte-Carlo simulated predictions of background processes.	80
6.3	Categories of events on which the multi-variate analysis is trained.	81
6.4	Variables used to train the multi-variate discriminant.	83
6.5	Hyperparameter choices used in the multi-variate analysis.	83
7.1	A summary of experimental systematic uncertainties.	97
7.2	Flavour filters used in the simulation of $V+jets$ events. † this tighter filter is only applied to $Z \rightarrow \nu\nu$ samples.	104
7.3	A summary of SHERPA internal variations.	106
7.4	$W+jets$ normalisation and acceptance uncertainties.	107
7.5	A breakdown of flavour subprocesses making up the $W +$ heavy flavour process.	109
7.6	$W+jets$ flavour composition uncertainties.	109
7.7	$W+jets$ shape uncertainties.	110
7.8	$Z+jets$ normalisation and acceptance uncertainties.	116

7.9	Extrapolation uncertainties due to the $p_T^V$ shape uncertainty. . . . .	117
7.10	$Z+$ jets flavour composition uncertainties. . . . .	118
7.11	$Z+$ jets shape uncertainties. . . . .	121
7.12	A summary of systematic uncertainties on $t\bar{t}$ events. . . . .	125
7.13	A summary of systematic uncertainties on sub-dominant backgrounds of the analysis. . . . .	132
7.14	A summary of systematic uncertainties on the signal process. . . . .	136
8.1	This table shows the impact of systematic uncertainties on the signal strength $\mu$ broken down into categories or nuisance parameter sets. Some sets are further broken down into smaller categories indicated by nested indentation. . . . .	142
8.2	Statistical significances of the background plus signal hypothesis broken down by analysis channel. . . . .	152
8.3	Statistical significances of the background plus signal hypothesis for $WH$ , $ZH$ and $VH$ measurements. . . . .	154
8.4	Statistical significances of the cross-section measurement in each STXS bin. . . . .	155
B.1	$W \rightarrow \ell\nu$ samples used in the analysis. . . . .	192
B.2	$Z \rightarrow \ell\ell$ samples used in the analysis, which are all generated using SHERPA 2.2.1. . . . .	193
B.3	Nominal $Z \rightarrow \nu\nu$ predictions used in the analysis, which are all generated using SHERPA 2.2.1. . . . .	193
B.4	Nominal $t\bar{t}$ samples used in the analysis. . . . .	194
B.5	Nominal predictions of the $VH \rightarrow b\bar{b}$ signal process used in the analysis fit and to assess the systematic uncertainties. . . . .	195
B.6	Nominal diboson predictions used in the analysis. . . . .	196
B.7	Nominal and alternative single top samples. . . . .	197
B.8	Alternative $W \rightarrow \ell\nu$ samples used in the analysis. . . . .	198
B.9	Alternative $Z+$ jets samples. . . . .	198

B.10 Alternative $t\bar{t}$ samples used in the analysis. . . . .	199
B.11 Alternative predictions of the $VH \rightarrow b\bar{b}$ signal process. . . . .	200
B.12 Alternative diboson predictions used in this analysis to study systematic uncertainties. . . . .	201
B.13 Wjets syst summary caption . . . . .	201
B.14 $Z+jets$ syssts summary . . . . .	201
B.15 Extrapolation uncertainties of the $p_T^V$ shape systematic on the $W+jets$ samples. . . . .	202
B.16 Extrapolation uncertainties of the BDTr-based shape systematic on the $W+jets$ samples. . . . .	203

# Chapter 1

## Introduction

In 2012 the Higgs boson [1–3] was discovered by the ATLAS and CMS collaborations at the Large Hadron Collider [4, 5]. It was said to form the last piece of the Standard Model of Particle Physics [6–9], a framework that describes three of the four fundamental forces of nature. Despite the theory being apparently complete, it does not describe gravity, the fourth of the known fundamental forces of nature. The theory also has other shortcomings, it cannot explain the presence of dark matter [10–20] or a number of other observed phenomena [21–25]. So far the model has stood up to all experimental tests [26, 27] concerning its own predictions but there are still parameters of the model that have not been measured. Given the theory’s understood shortcomings, it is hoped that continued scrutiny of the models predictions will yield unexpected results, perhaps hinting at a new way forward in terms of a theory that describes all matter and forces in the universe or simply exposing further gaps in our knowledge. For this reason it is more important than ever to study in detail the most recently discovered piece of the model, the Higgs boson.

This work focuses on studying Higgs boson particles that decay into two bottom quarks and are produced alongside a vector boson, denoted  $VH \rightarrow b\bar{b}$ . This decay mode is of importance as it is currently the only decay mode of the Higgs decaying to quarks that has been observed [28].

The data used to study this decay mode was collected using the ATLAS detector

I thought you were thinking of 3 generations, CKM matrix elements or Yukawa couplings here. Maybe say something like, "...recent measurements straining SM fits."

Aren't there recent fits from the EWWG?

Really? Which ones?

double decay

by members of the ATLAS collaboration. The collection of this data is only made possible by their hard work and by the hard work of everyone working on the Large Hadron Collider at CERN.

A rough blueprint of this analysis is shown in figure 1.1. The blueprint starts with the theory of the Standard Model of Particle Physics, described in chapter 2. The theory's predictions inform the design of the ATLAS detector which is detailed in chapter 3. Furthermore, the theory aids the choice of particles to collide, which events to analyse, and allows us to generate predictions of what should happen in the collisions. Chapter 4 gives an overview of two machine learning algorithms that are used throughout the analysis. Events must be reconstructed before they can be analysed, a selection process is used on the reconstructed objects to filter events not relevant to the analysis. The reconstruction and selection process is described in chapter 5. Events in data must be categorised into those that are signal-like and those that are background-like in order to extract the most signal sensitivity from the analysis. This is achieved with a number of strategies including a multi-variate algorithm which is trained on the simulations.

model ... well?

This categorisation process is described in chapter 6, along with the overall analysis strategy including the details of the profile-likelihood fit. A choice is made of simulated events that well model the data. Considering the modelling, the theory and the shortcomings of the detector<sup>or</sup>, a set of systematic uncertainties are estimated, these are detailed in 7. The results of the analysis are shown in chapter 8 and conclusions are drawn in chapter 9.

which are

responsible for

My contributions to the  $VH \rightarrow b\bar{b}$  analysis published in 2021 have been numerous and spanned a number of different areas of the analysis. At the beginning of each chapter I will aim to make clear which content represents my work and which is that of my colleagues. In short, my main contributions have been to study and estimate systematic uncertainties as well as the behaviour of the profile-likelihood fit. Additionally I was responsible with maintaining and operating the analysis software framework.

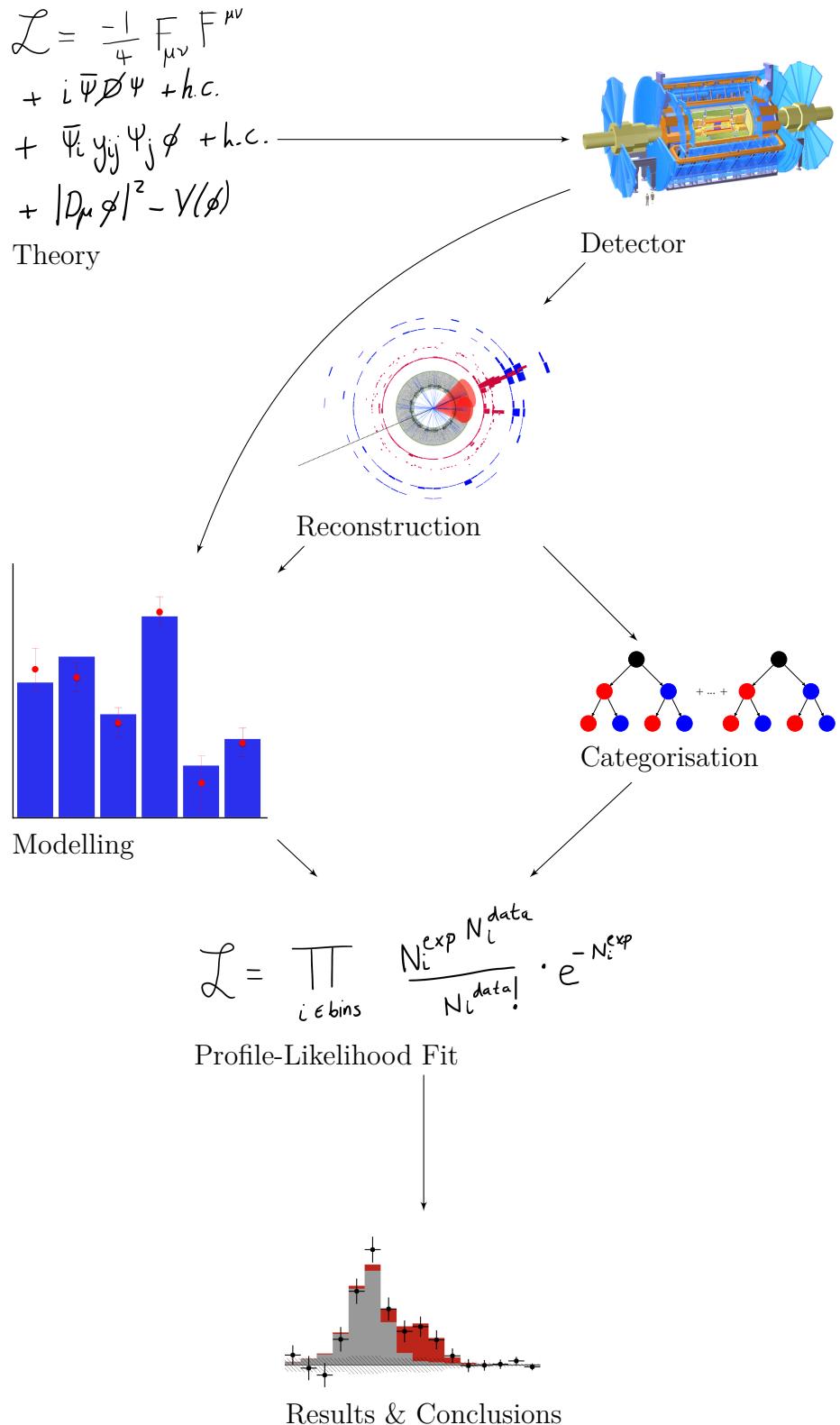


Figure 1.1: A flow chart showing the roadmap of the  $VH \rightarrow b\bar{b}$  analysis and of this thesis.

# Chapter 2

## Physics Theory

The following chapter outlines the physics theory that informs and guides the experimental process of searching for new particles or making measurements at a particle collider. Not only are the physics theories described here useful in that context but they also provide an almost complete picture of the universe at certain scales. This chapter was written with the aid of notes taken at the annual STFC High Energy Physics Summer School, and with the aid of several books [29, 30], in which a more detailed description of the theories can be found.

The Standard Model of particle physics is a theoretical framework that describes all elementary particles and three of the fundamental forces of nature. Notably the only force that is not described by the theory is gravity. Particles described by the model are listed in table 2.1 with a white gap separating the matter particles (fermions) from the force carrying particles (bosons). Fermions, which make up solid matter obey Fermi-Dirac [31, 32] statistics whereas bosons obey Bose-Einstein [33] statistics. The Higgs boson is special in that as far as we know it does not carry a force in the conventional sense, instead it is responsible for giving fundamental particles mass, discussed in more detail in section 2.5. In the table of particles quarks (blue) and leptons (red) are ordered in columns by increasing mass, apart from the neutrinos, which are massless in the theory.

The model describes forces as being mediated by certain particles, the photon ( $\gamma$ ) mediates the electromagnetic force, particles experiencing electromagnetic repulsion

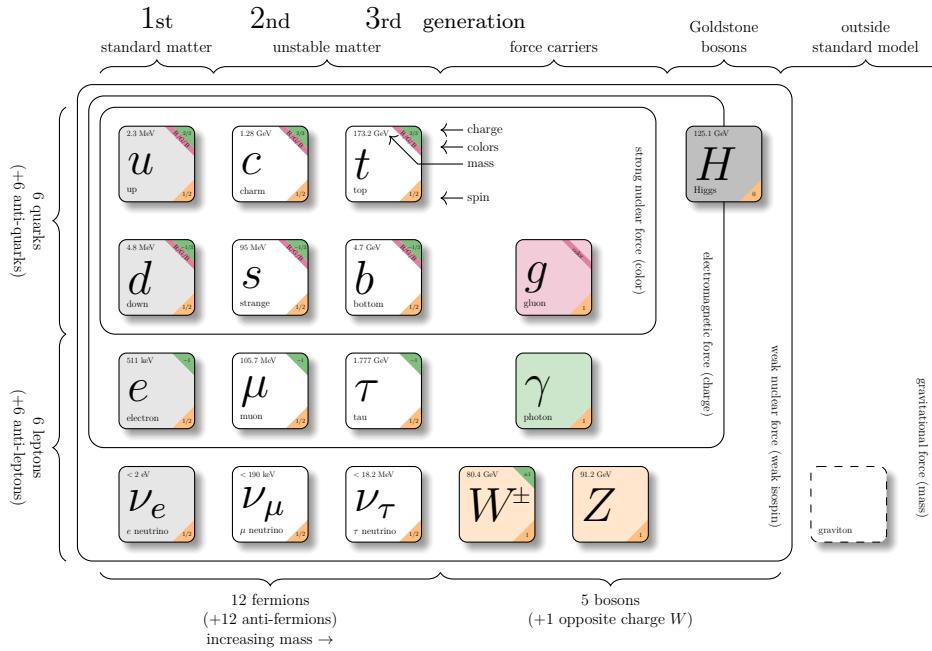


Table 2.1: A table showing the particles of The Standard Model. The particles of matter, fermions, are shown on the left and the force carrying particles, bosons, are shown on the right.

or attraction are described as exchanging photons. The strength and direction of the force experienced is proportional to the electromagnetic charge of particles involved. The Standard Model is a theory of quantum fields in which the strength of interactions between fields, or particles which are described as excitations in the fields, is parametrised by something known as a coupling constant. It is natural to assume that the strength of the interaction between photons and charged particles is related to the electromagnetic charge of the particles involved. Indeed this is the case consider the Coulomb force between two protons,

$$F = \frac{e^2}{4\pi\epsilon_0 r}, \quad (2.1)$$

where  $e$  is the elementary charge,  $\epsilon_0$  is the electric constant and  $r$  is the distance between the two protons in question and also the energy of a photon given by

$$E = \frac{hc}{\lambda}, \quad (2.2)$$

where  $h$  is Planck's constant,  $c$  is the speed of light and  $\lambda$  is the wavelength of the

photon. The value of the ratio of these two quantities

$$\alpha = \frac{e^2 \lambda}{4\pi\epsilon_0 r hc}, \quad (2.3)$$

known as the fine structure constant, is the coupling constant that describes the strength of the interactions between the photon field and fields of particles with electromagnetic charge. So as is now clear this coupling constant does indeed depend on the electromagnetic charge of the objects involved and so it is not constant.

As well as the electromagnetic force the Standard Model describes the strong nuclear force and the weak nuclear force, shortened to just the strong and weak forces respectively. Like the electromagnetic force they too are mediated by the exchange of particles, the gluons ( $g$ ) carry the strong force and the  $W^\pm$  and  $Z^0$  bosons carry the weak force.

The charge associated with the strong force is known as colour which can take values that are mapped onto colours in the visible spectrum (red, green, blue) for ease of description. For each of these colours an anti-colour is also allowed (anti-red, anti-green, anti-blue). Unlike with the electromagnetic charge, particles with colour charge are not found freely in nature. Instead we find particles known as hadrons which are bound states of quarks and anti-quarks (e.g. the proton). The phenomenon of coloured particles being bound in such a manner is known as colour confinement [34], and the bound states are described by the quantum numbers isospin ( $I$ ) and hypercharge ( $Y_c$ ). It is commonly assumed that all free particles in nature are colour singlets e.g. for a hadron the state could be written as

$$\frac{(r\bar{r} + b\bar{b} + g\bar{g})}{\sqrt{3}}, \quad (2.4)$$

where  $r$ ,  $b$  and  $g$  represent red, blue and green charges respectively. This phenomenon is known as quark confinement. Gluons carry colour and anti-colour indicating that there should be nine possible quantum mechanical states for the gluon given the available number of colour/anti-colour combinations, however when one

considers that the strong force is exclusively short range, and therefore that there should be no free gluons (disallowing colour singlet gluons) the number of possible states is reduced to eight. The state of a particle, as far as its description with respect to the strong force is concerned, is given by a vector which lives in a vector space, in which elements of the Lie group  $SU(3)_C$  act as unitary operators, where the  $C$  denotes that the group is associated with the colour charge. The  $SU(3)$  group is the group of  $3 \times 3$  unitary matrices whose determinant is one.

Describing the weak force requires introducing further quantum numbers: weak isospin  $T$  and weak hypercharge  $Y_W$ . The state of a particle with regards to the weak force is given by a vector which lives in a vector space in which elements of  $SU(2)_L \times U(1)_{Y_W}$  act as unitary operators where the  $L$  denotes that only particles in left-handed chiral states interact with the weak force<sup>1</sup>. Left-handed fermions are represented as doublets in the theory with weak isospin  $T = 1/2$  whilst right-handed fermions are singlets with weak isospin  $T = 0$ .

Along the way we have described particle states with respect to particular forces as vectors living in some vector space where the action of the element of a group has been as a unitary operator. If we are to describe a particle state taking into account the full model, the group whose elements should act as unitary operators on the particle state (the gauge group) is  $SU(3)_C \times SU(2)_L \times U(1)_{Y_W}$ . For each of the groups in the direct product we have established a (gauge) symmetry and therefore due to Noether's theorem [35] there should be an associated conserved quantity. The conserved quantities in this case are the electric charge, the weak hypercharge and isospin and the colour charge.

---

<sup>1</sup>More specifically only left-handed chiral particles participate in weak charged current interactions.

## 2.1 Historical Aside

This section provides some historical context surrounding the Dirac equation which will later be used as the starting point in the discussion of quantum electrodynamics, which is the sector of the Standard Model that describes electromagnetic interactions.

In 1905 Albert Einstein first proposed the idea of special relativity [36]. The aim of the idea was to unify the then inconsistent theories of Maxwell's electromagnetism and Newtonian mechanics. The result of Einstein's work was a theory of motion which agreed with the predictions of Newtonian mechanics at velocities much smaller than the speed of light but whose predictions were accurate also at much higher velocities (for which Newtonian predictions fail). A consequence of special relativity is that it demands that any equation of motion must be invariant under Lorentz transformations, at least in terms of the formulation of new theories is concerned.

Many physical phenomena predicted by special relativity could be considered of higher consequence in general, for example the phenomena of length contraction, time dilation, energy-mass equivalence and the universal speed limit (equal to the speed of light in vacuum), all of which are extensively scrutinised experimentally [37–43]. It is the Lorentz transformation however that should be kept in mind for the following discussion, the transformation may be written as

full stop

$$\begin{aligned} t' &= \gamma(t - vx/c^2) \\ x' &= \gamma(x - vt) \quad \text{add punctuation} \quad (2.5) \\ \text{with } \gamma &= \frac{1}{\sqrt{1 - v^2/c^2}}, \end{aligned}$$

in a single dimension of space  $x$  and one of time  $t$  where  $v$  represents the velocity of the system described by the primed coordinates relative to the unprimed coordinates and  $c$  is the speed of light in vacuum.

Twenty years after Einstein introduced the ideas of special relativity Erwin Schrödinger postulated new ideas regarding the motion of quantum mechanical systems [44].

Though he knew his new equation was not invariant under Lorentz transformations, and therefore incomplete, Schrödinger's formulation of quantum mechanics changed the way physicists thought about the universe forever. His famous equation

$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{x}, t) = \left( \frac{-\hbar^2}{2m} \nabla^2 + V(\vec{x}, t) \right) \Psi(\vec{x}, t), \quad (2.6)$$

describes the states of particles as wave-functions  $\Psi$  which can only be interpreted in a probabilistic manner and contains Planck's constant the quantum of action. This work had many consequences including the quantisation of the values of measured observables (meaning they can only take discrete values) and the descriptions of particles as waves.

his needs?

It was the aim of Paul Dirac to make the Schrödinger equation Lorentz invariant and thus provide a more complete description of quantum systems. Along the way he came to the realisation that in order for his equation to satisfy his needs the wave-function had to be replaced with a four component spinor ( $\psi$ ) and the introduction of matrices known now as the Dirac matrices (labeled  $\gamma^\mu$  with  $\mu = 0, 1, 2, 4$ ) was required. Though not the form he originally wrote down Dirac's Lagrangian density takes the form

$$\mathcal{L}_{Dirac} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \quad (2.7)$$

no ( ) needed

where the repeated up and down indices are implicitly summed over ( $\partial_\mu$  represents the partial derivative taken with respect to a spatial coordinate  $\mu = 1, 2, 3$  or time  $\mu = 0$ ).

## 2.2 Quantum Electrodynamics

In order to take Dirac's Lagrangian (eq. 2.7) and turn it into something that appropriately describes quantum electrodynamics (QED), we should consider a  $U(1)$

gauge transformation of the Dirac spinor and it's adjoint

$$\begin{aligned}\psi \rightarrow \psi' &= e^{i\alpha(x)}\psi, \\ \bar{\psi} \rightarrow \bar{\psi}' &= e^{-i\alpha(x)}\bar{\psi},\end{aligned}\tag{2.8}$$

with  $\bar{\psi} \equiv \psi^\dagger \gamma^0$  and where  $\alpha(x)$  is a local phase. Under this transformation the Lagrangian transforms as

$$\mathcal{L}_{Dirac} \rightarrow \mathcal{L}'_{Dirac} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \bar{\psi}\gamma^\mu \alpha(x)\psi\tag{2.9}$$

which is not equivalent to the original due to the factor resulting from the derivative of the transformed spinor. Instead let us change the derivative to the gauge covariant derivative

$$D_\mu = \partial_\mu + ieA_\mu \quad \text{comma}\tag{2.10}$$

where we interpret  $A_\mu$  as the photon field, with coupling constant  $e$ , parametrising the interaction strength. The field is also referred to as the electromagnetic gauge field since it arrives during the process of making the Lagrangian invariant under the  $U(1)$  group, the gauge group of electromagnetism. Note that here what we have labeled  $e$  is nothing more than the fine structure constant previously denoted  $\alpha$  in eq. 2.3. The transformation of the new field under the action of the gauge is defined as

$$A_\mu \rightarrow A'_\mu \equiv A_\mu - \frac{1}{e}\partial_\mu \alpha(x).\tag{2.11}$$

This means that the action of the gauge covariant derivative on the spinor transforms as

$$D_\mu \psi \rightarrow D'_\mu \psi' = e^{i\alpha(x)} D_\mu \psi\tag{2.12}$$

which means that the new Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi\tag{2.13}$$

is invariant under the action of the gauge as desired. What remains in order to write a description of QED is to write down a kinetic term for the photon field. An appropriately gauge and Lorentz invariant term is

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (2.14)$$

where the electromagnetic tensor is defined as

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (2.15)$$

Putting everything together we can define the Lagrangian for QED as

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (2.16)$$

## 2.3 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the theory of the strong force. Its mathematical formulation is similar to that of QED. Except the gauge group for QCD is  $SU(3)_C$  where the  $C$  denotes that the force is associated with colour charge. As previously discussed the eight generators of the group are associated with the eight gluons of the Standard Model. The generators are present in the form of the transformation of a fermion field under an element of  $SU(3)$

$$\psi \rightarrow \psi' = \exp\left(i\alpha_a(x) \cdot \frac{\lambda_a}{2}\right)\psi, \quad (2.17)$$

where the  $\lambda_a$  are the Gell-Mann matrices, generators of  $SU(3)$ . A key difference between the strong force and the other forces described by the Standard Model is

that it increases in strength with range. This property leads to a phenomena known as quark confinement which has been discussed previously. Quark confinement is the reason for many of the complications that arise when trying to detect certain particles in a particle detector such as ATLAS. Specifically, quarks and gluons that

Maybe add that gluons carry colour charge as reason

**Why does the strong force increase with distance?**

are produced in collisions undergo a process called hadronisation whereby they transition from their coloured states to colour singlets. Energy present in this process results in the creation of lots of different states, some which decay to leptons, with the overall process producing a roughly conical shower of particles known as a jet.

## 2.4 Electroweak theory

The Glashow-Salam-Weinberg model of electroweak interactions [8? ? ] describes the weak force and electromagnetism as a quantum field theory, which is gauge invariant under transformations that are elements of  $SU(2)_L \times U(1)_{Y_W}$ . As previously mentioned the  $L$  and  $Y_W$  subscripts denote that the gauge groups in the direct product that are associated with left-handed chiral particles and weak hypercharge respectively. The association with weak hypercharge distinguishes this  $U(1)$  group with the  $U(1)$  group from QED. The transformation of the fermion fields under  $SU(2)$  is given by

$$\psi \rightarrow \psi' = \exp\left(i\vec{\alpha}(x) \cdot \frac{\vec{\sigma}}{2}\right)\psi, \quad (2.18)$$

where  $\vec{\sigma}$  is a vector of the Pauli matrices  $\sigma_i$  with  $i = 1, 2, 3$ , a familiar representation of  $SU(2)$  generators. Constructing a gauge covariant derivative for the full transformation under  $SU(2) \times U(1)$  requires the addition of new fields analogous to the photon field from QED, the new derivative takes the form

$$D_\mu = \partial_\mu - i\frac{g_1}{2}Y_W B_\mu - i\frac{g_2}{2}\sigma_i W_\mu^i, \quad (2.19)$$

where coupling constants  $g_1$  and  $g_2$  parametrise the strength of interactions with each field. The index  $i$  runs over the three Pauli matrices and three new fields  $W_\mu^i$  with  $i = 1, 2, 3$  which are associated with the  $SU(2)$  gauge. The  $B_\mu$  field is associated with the  $U(1)_{Y_W}$  gauge and is obtained in the same way as the photon field in QED but is given a new symbol as it is *not* the photon field.

In fact, none of the fields added here are the physical fields that we have access to in nature associated with the electromagnetic force or the weak currents. In order

some of

usually jets are not defined with leptons

broken references

to obtain the physical fields for the weak charged current one can simply take the linear superposition

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \quad (2.20)$$

In order to recover the photon field and reveal the field for the weak neutral current the idea of weak mixing must be introduced. Weak mixing was introduced to the theory after the discovery of parity violation [45]. Parity is an intrinsic property of a particle which is similar to chirality. Chirality is not intrinsic to particles as a Lorentz boost can always appear to flip the chirality of the particles state. For particles that are massless and therefore travelling at the speed of light this cannot occur, for such particles chirality will always be equivalent to parity. A fermion field with left or right handed chirality can be obtained by multiplication with one of two corresponding projection operators defined as

$$\begin{aligned} P_L &= (1 - \gamma^5)^2 / 2, \\ P_R &= (1 + \gamma^5)^2 / 2, \end{aligned} \quad (2.21)$$

Just use a ":" with  $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ , where  $\gamma^\mu$  are the Dirac matrices, like so

$$\begin{aligned} \psi_L &= P_L\psi, \\ \psi_R &= P_R\psi. \end{aligned} \quad (2.22)$$

It is known that the weak neutral current and the electromagnetic force both interact with particles of left and right handed chirality. Spontaneous symmetry breaking, has the effect of rotating the plane defined by the  $B_\mu$  and  $W_\mu^3$  fields into the physical fields we see in nature today. The mixing of the fields due to this rotation takes the form

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \quad (2.23)$$

Glashow already used

$\theta_W$  in the earlier paper.

It's the weak mixing angle.

where  $\theta_W$  the Weinberg angle parametrises the amount of mixing. This picture shows that unification of QED, up to the inclusion of the two coupling constants,

and a description of the weak force has been achieved. Although it at first seems like the  $U(1)_{EM}$  gauge group is not present in  $SU(2)_L \times U(1)_{Y_W}$  gauge group of electroweak theory it has been shown [29] that the QED gauge symmetry is recovered by spontaneous symmetry breaking. Also the  $Y_W$  subscript in the gauge group represents weak hypercharge which is related to electric charge  $Q$  by the following relationship

$$Y_W = 2(Q - T^3), \quad (2.24)$$

where  $T^3$  is the third component of isospin, the component that must be conserved in weak interactions according to the theory.

The particles associated with the weak neutral and charged currents are observed to have masses in nature [46–49] therefore one would naively like to write mass terms of the form

$$\mathcal{L}_{mass} \propto M_B^2 B^\mu B_\mu \quad (2.25)$$

$$+ M_W^2 W_a^\mu W_\mu^a. \quad (2.26)$$

The above mass terms are however not gauge invariant therefore another solution is required, as the masses as observables which would have ambiguous predictions were they to change under gauge transformation.

## 2.5 The Brout-Englert-Higgs Mechanism

The Brout-Englert-Higgs mechanism was made complete almost simultaneously by R. Brout and F. Englert, P. Higgs and, G. Guralnik, C. R. Hagen and T. Kibble. The underlying mechanism was proposed prior to this work by P. Anderson [50], though this initial theory was not relativistic invariant. It was initially proposed as a means to give the vector bosons mass terms that were gauge invariant. The theory predicts a complex scalar field (the Higgs field) that undergoes spontaneous symmetry breaking. Interactions with this field are predicted to be mediated by a massive spin-1 scalar particle that is now known to be the Higgs boson. This particle

at first it seems like

also gives mass to the fermions via a different mechanism. In general spontaneous symmetry breaking is a process by which a symmetry breaks once conditions meet some threshold. An example of this is a hot sphere of ferromagnetic material whose spins are isotropically oriented. As the sphere cools the ferromagnetic property of the material will align the spins. In the hot scenario the sphere had symmetry in all spatial directions, by this it is meant that the changes to the sphere's orientation were indistinguishable. Once the spins have aligned however this is no longer the case, the fact that the spins point in a specific direction means that direction is special and so some of the symmetry was spontaneously broken. It can be noted though that a preserved symmetry still exists as rotations about the axis defined by the direction of the spins would leave the sphere invariant. In the Standard Model the symmetry that breaks is that of the complex scalar Higgs potential. Consider a Lagrangian involving the field  $\phi$  of the form

$$\mathcal{L} = T - V(\phi) = \partial_\mu \phi^\dagger \partial^\mu \phi - \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (2.27)$$

$$\text{with } \phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2). \quad (2.28)$$

Invariance under global phase transformations of the form  $\phi \rightarrow e^{i\theta}\phi$  depends on the parameters of the potential  $\mu$  and  $\lambda$ . Figure 2.1 shows two sketches of the potential for the scenarios where  $\mu^2 > 0$ ,  $\lambda < 0$  (left) and  $\mu^2 < 0$ ,  $\lambda < 0$  (right). To suggest

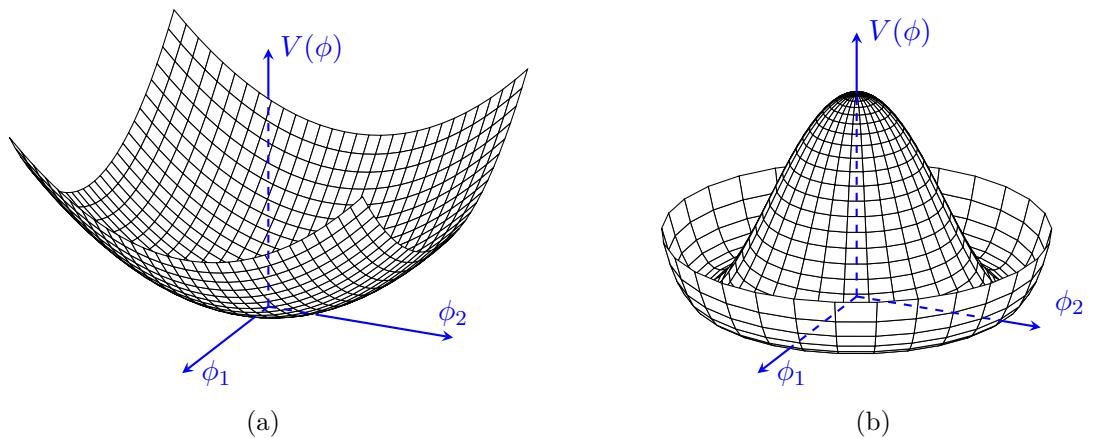


Figure 2.1: The Higgs potential in its fully and broken symmetric forms.

**Include (a) and (b) in the caption**

that in our universe this symmetry is spontaneously broken is to suggest that the

values of these parameters evolved over time from the full to the broken state. This ends up leading to masses for the vector bosons that are dependent on  $\mu^2$ .

Maybe add something theoretical about the Higgs couplings. After all you have a VH and Hbb coupling in your work.

## 2.6 Higgs Bosons at the LHC

Higgs bosons are produced at the LHC in a number of different ways, the four most common of which are shown in figure 2.2. The production cross-section

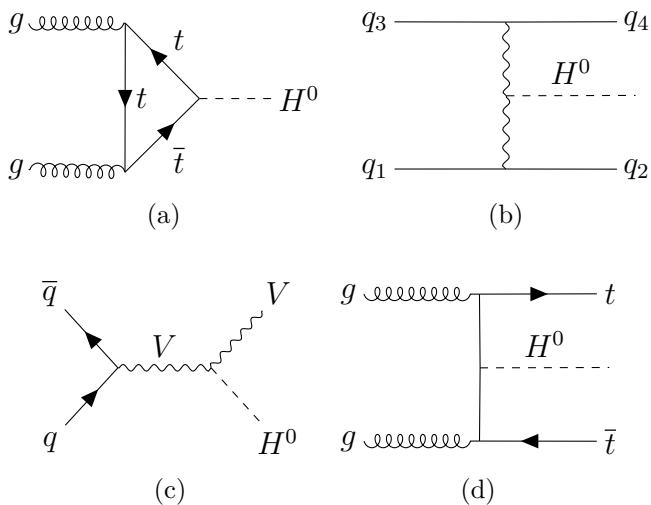


Figure 2.2: The four most common Higgs boson production methods from proton-proton collisions at the LHC.

Name (a) - (d)

of these processes with respect to the centre of mass energy of the proton-proton collision is shown in figure 2.3 (a). It can be seen gluon-gluon fusion (fig 2.2 a) is by far the dominate contributor occurring over an order of magnitude more than the next highest process which is quark associated production (fig 2.2 b). The next highest production channel with respect to cross section is vector boson associated (fig 2.2 c) which will be the focus of the rest of this report. Finally, top quark associated production (fig 2.2 d) has the smallest cross section of these processes. The Higgs boson is predicted by the Standard Model to decay in a number of different ways depending on its mass, a free parameter of the model. In figure 2.3 (b) the branching ratios of the Higgs can be seen, plotted with respect to Higgs mass. The decay that will be focused on for the rest of this report is a Higgs boson decaying to a pair of  $b$ -quarks, denoted  $H \rightarrow b\bar{b}$ . Given that the focus here is on

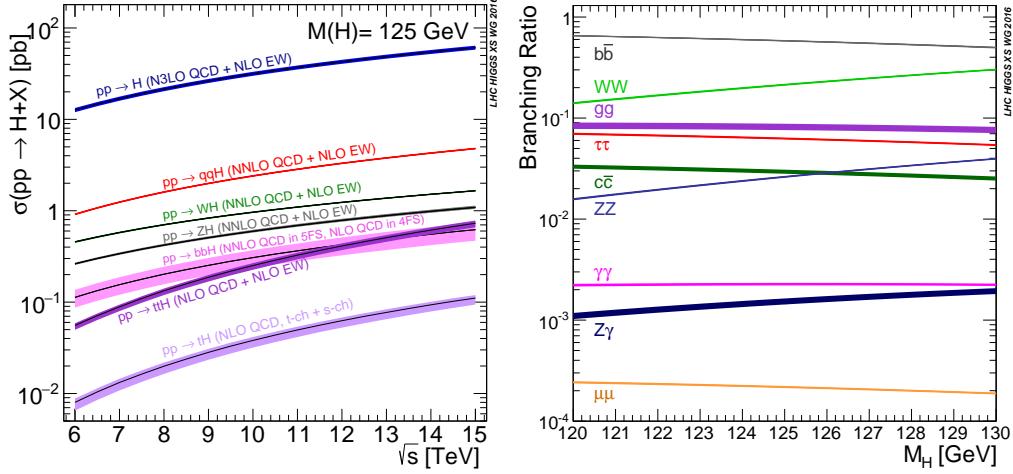


Figure 2.3: Higgs boson production cross-sections (left), and branching ratios (right) for a range of centre of meass energies and Higgs boson masses respectively [51].

0, 1, 2

vector boson associated production of a Higgs boson it is also important to consider the decay of the vector boson. Three possible scenarios are represented in figure 2.4, namely the situations where the vector boson decays to 1, 2 or 3 charged leptons and the appropriate number of neutrinos. It is these leptonic decay modes that

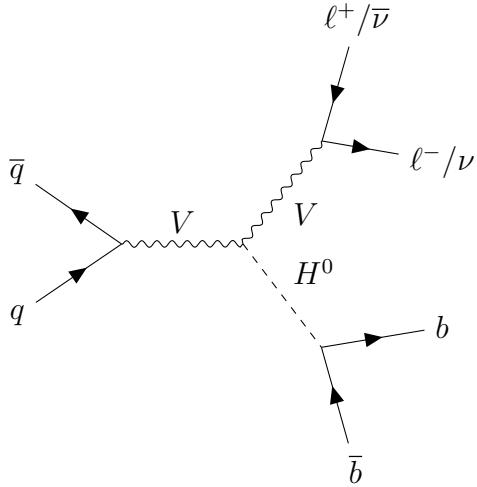


Figure 2.4: A diagram showing a Higgs boson (decaying to a pair of b quarks) produced in association with a vector boson (decaying to 0, 1, or 2 charged leptons denoted  $\ell^{+/ -}$ ).

motivate the reason for studying this production mechanism as opposed to one of the more common ones. The issue with looking at the other production modes is that very large QCD generated backgrounds are present due to initial state radiation. Whilst these backgrounds are also present when looking at the vector associated channel they can be partially suppressed by triggering on a charged lepton.

# Chapter 3

## The ATLAS Detector at the Large Hadron Collider

### 3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [52] is a large circular machine located 100 m underground straddling the Swiss-French border at the European Organisation for Nuclear Research (CERN). The LHC accelerates and collides protons and other charged particles. It has a diameter of 27 km and resides in a tunnel which was originally excavated for the Large Electron-Positron Collider [53] experiment. During its construction the tunnel was the largest civil engineering project in Europe to date. Today there are many physics experiments that take place at CERN, some of which are marked in figure 3.1. There are currently seven experiments that record data from the collisions at the LHC: ATLAS [54], CMS [55], LHCb [56], ALICE [57], MoEDAL [58], TOTEM [59] and LHCf [60].

The LHC is a synchrotron, an accelerator that uses magnets in a dipole configuration, such as in figure 3.2, to bend the path of charged particles into conformity with its circular shape. It is apparent from studying the figure that counter-rotating beams of same sign charged particles will require two sets of dipole magnets in order to rotate in opposite directions around the same ring. This is one disadvantage of a proton-proton collider with respect to a proton-anti-proton collider such as the Tev-

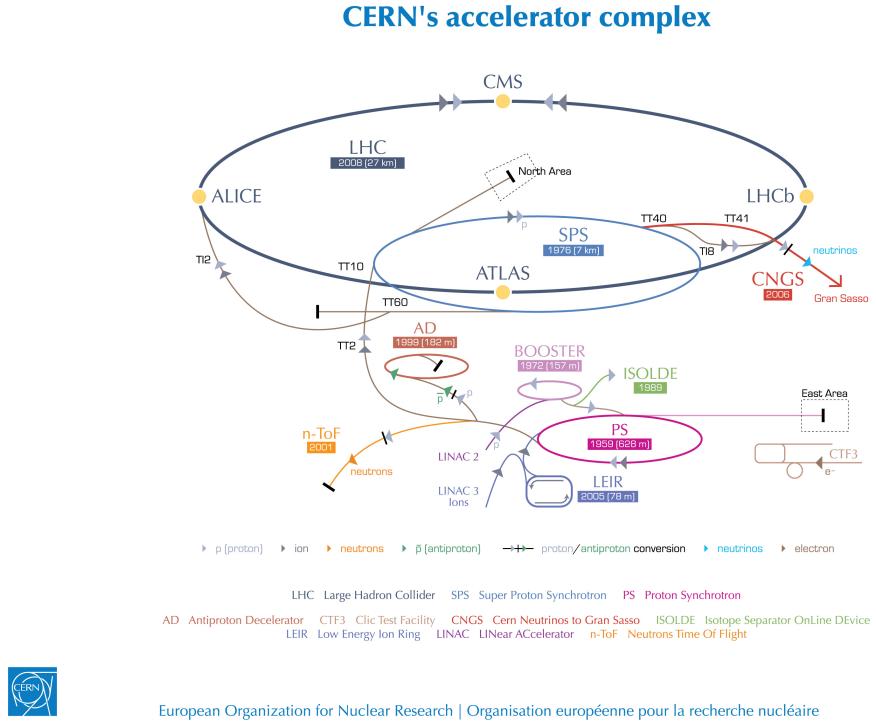


Figure 3.1: The CERN accelerator complex [61].

atron [62] which can use the same magnets for both beams. On the other hand, the proton-proton collider is able to separate the beams after they have been brought together to collide with a single dipole. The bending magnets of a synchrotron are designed to ramp up their magnetic field in synchronisation with the kinetic energy of the accelerated particles, allowing higher energies to be achieved before the beam is lost. The LHC can accelerate each beam to an energy of 6.5 TeV leading to collisions with a centre of mass energy of  $\sqrt{s} = 13$  TeV, although the design energy of the LHC is  $\sqrt{s} = 14$  TeV. The Tevatron held the previous record for centre of mass energy of collisions of  $\sqrt{s} = 2$  TeV. The LHC has 1232 dipole magnets [52] which are made of copper-clad niobium-titanium cables, a superconducting material whose electrical resistance falls to zero below 10 K. In order to maintain super-conductivity a cryogenic system using liquid helium is employed to cool the magnets. The higher the velocity of a charged particle, and the tighter the desired bending radius, the larger the magnetic field required to perform the bending. The large size of the LHC and the choice of superconducting magnet technologies are both informed by the aim to accelerate protons to the highest energy, and therefore velocity, possible.

synchronisation   comma

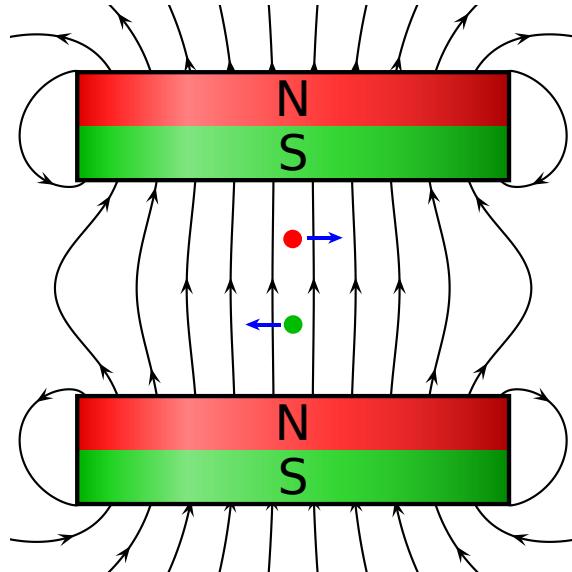


Figure 3.2: A representation of a pair of idealised cylindrical magnets in a dipole configuration. Two positively charged particles are shown as circles, the red particle is traveling out of the page, the green particle is traveling into the page. The forces experienced by each particle due to the magnetic field are shown as blue arrows.

The force which accelerates the particles is provided by radio-frequency cavities, of which the LHC has 16 [52] The electric field in the radio-frequency cavity forms a standing wave, the separation between bunches of particles to be accelerated must be matched to the frequency of this wave. Protons in the LHC are accelerated in full stop or drop “this” bunches and in vacuum, this to increase the likelihood of collisions and mitigate loss of energy and scattering effects due to interactions with air molecules. These two factors lead to the occurrence of space charge which causes an increase in the emittance of the beam, where the emittance is defined as the total area that the beam occupies in its beam-pipe. The greater the energy of the particles the more they can overcome increase in emittance due to space charge. Increased emittance is especially problematic in circular accelerators where periodic effects can quickly lead to the loss of beam. For these reasons it would be very challenging to accelerate full stop protons from rest in a synchrotron, the starting point for the protons of the LHC is therefore a linear accelerator called Linac2 which is used to overcome space charge effects before the protons move on to a series of synchrotrons as seen in figure 3.1.

Even a beam with its emittance under control would still be lost from the accelerator if only dipole magnets were used to control its path. Magnets in a quadrupole

configuration as in figure 3.3 are used to focus the beam and keep it in the beam-pipe. The quadrupoles behave such that particles feel a force that increases with the distance from the centre of the beam leading to simple harmonic motion of individual particles in a bunch. The LHC has a series of 24 quadrupole magnets each for focusing in the horizontal and vertical directions [52] as well as higher multiplicity configurations; sextupole, octupole, decapole and dodecapole which are used to correct imperfections in the fields of other magnets.

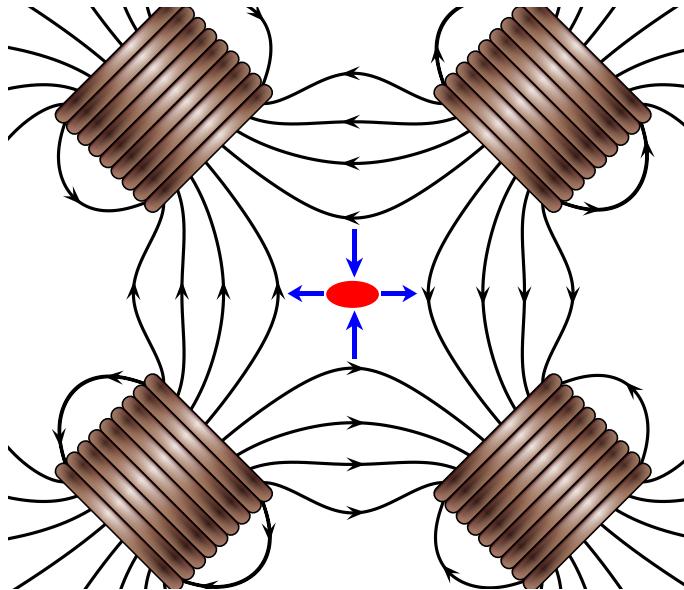


Figure 3.3: Representation of an idealised set of coil magnets in a quadrupole configuration with a proton beamspot shown as a red ellipse. The proton beam is drawn coming out of the page (*watch out!*), magnetic field lines are drawn in black and the forces acting on each bunch of protons are drawn as blue arrows.

Ignoring systematic uncertainties for the moment, the sensitivity of a measurement can be increased by having a larger dataset, in this case more recorded collisions. Constraints on the number of years the LHC is able to run mean that the best way to record more collisions is to collide more particles per second. A quantity known as the luminosity is often used to describe how much data is available for an analysis, it is written as

$$L = \frac{1}{\sigma} \frac{dN}{dt}, \quad (3.1)$$

where  $\sigma$  is the cross-section, an area within which particles must pass by one another in order to interact, and  $N$  is the number of events recorded in a period of time  $t$ .

**multiplicity → multipole**

\$N\$

For luminosity at the LHC N can be expressed as

$$\frac{dN}{dt} = n_{bp} n_1 n_2 \nu_r, \quad (3.2)$$

where  $n_{bp}$  is the number of colliding bunch pairs,  $n_1$  and  $n_2$  are the number of protons in each beam and  $\nu_r$  is the frequency with which the beams rotate around the LHC's circumference. The number of particles in the beams is limited by space charge. The number of bunches is limited by the frequency that the radio-frequency cavities can operate at. The revolution frequency is limited by the strength of the dipole magnets and the circumference of the accelerator ring. Increasing luminosity by reducing the cross-section amounts to reducing the beam widths which is limited by the emittance of the beam. The LHC has already exceeded it's design luminosity providing physicists with more data to analyse than expected and plans are well underway for the upgrade to a High-Luminosity LHC (HL-LHC) [63].

it's → its

## 3.2 The ATLAS Detector

The ATLAS detector [64] resides at a location on the LHC ring called Point 1, its full name is A Toroidal LHC ApparatuS. A diagram of the detector is shown in figure 3.4. ATLAS is considered to be a general purpose particle detector and has a wide physics program including: Higgs boson physics, top quark physics, searches for Supersymmetry and exotic states, probes of CP violation in b-quarks and light states and heavy ion physics. The detector itself is very large in size, spanning a width of 25 m and a length of 44 m and weighs 7000 tonnes which is comparable to the weight of the wrought iron content of the Eiffel tower [65].

Due to the composite nature of the proton, the decay products of collisions are extremely numerous. Additionally when two bunches of protons cross there is the chance that more than one hard scattering event occurs and softer glancing collisions are also a possibility. The number of hard scattering events in a given bunch crossing is known as the pile-up of the collision and is often denoted with the symbol  $\mu$ . As

B.E: programme

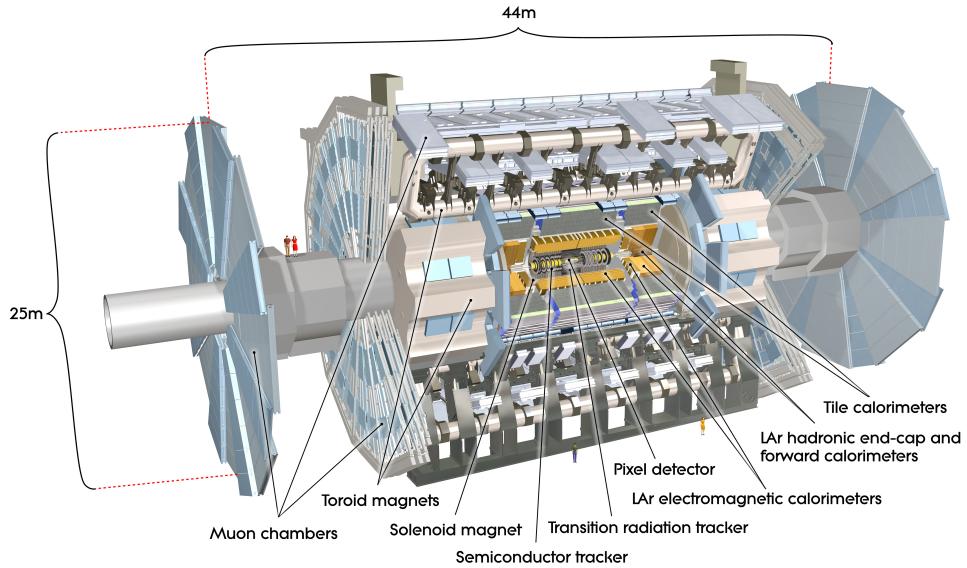


Figure 3.4: Computer generated image of the whole ATLAS detector with the major sub detectors labeled [66].

can be seen by inspecting equations 3.2 and 3.1 increasing the luminosity will often cause a higher pile-up environment in the detector. The variety of decay products of each collision necessitate the use of specialised sub-detectors in order to accurately measure the outcome of collisions. Ultimately in all cases a digital electrical signal is the desired output of a sub-detector which usually originates as an analogue signal. It is interesting to note that despite the many charges that are associated with the forces of nature discussed in chapter 2 the only one that we can directly measure is electric charge.

The ATLAS sub-systems are located in either the barrel or one of the end-caps. These two areas have a different geometry, and due to their relative positions with respect to the beam-pipe are exposed to different amounts of radiation, so the design of a sub-system in the barrel will differ from that of a sub-system measuring the same quantities in the end-cap. The details that follow are based on the ATLAS technical design report volumes [67, 68] unless another citation is present. Before detailing individual components it is important to detail certain properties of the detector relevant to all sub-systems. The coordinate system used to describe the ATLAS detector is known as right-handed. Three orthogonal axes

\$x\$ \$(x, y, z)\$ are used to describe the 3D space of the detector. The x-axis points towards  
\$y\$ the centre of the LHC ring, the y-axis points upwards and the z-axis points along  
\$y\$ the LHC beam pipe y-axis. The three axes meet at the interaction point which  
is the nominal position where bunches cross, located in the centre of the detector.  
Cylindrical coordinates \$(r, \phi)\$ are also often used to describe the physical features  
of the detector and phenomena caused by interactions in the detector that shall be  
referred to as analysis objects. Their definitions are that \$\phi\$ is the azimuthal angle  
\$x-y\$ in the x-y plane (transverse) around the beam pipe and \$r\$ is the distance from the  
interaction point. An analogy to the zenith angle of an object measured from the  
\$z\$ z-axis is pseudo-rapidity \$\eta = -\ln(\tan(\theta/2))\$ where \$\theta\$ is the ordinary zenith angle  
full stop measured from the same axis, pseudo-rapidity is relativistic invariant for massless  
particles. With these definitions an entirely angular separation between two objects  
in the detector can be written as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad \text{full stop} \quad (3.3)$$

Magnetic fields alter the trajectory of charged particles whilst preserving their energy. A large portion of the ATLAS detector is immersed in magnetic fields created by the magnet systems, and so this deflection phenomenon is present and exploited when making measurements of charged particles. There are four magnet systems

in the ATLAS solenoid in ATLAS the solenoid, the barrel toroid, and two end-cap toroids. The solenoid surrounds the inner detector whilst the toroid systems surround the muon chambers. Figure 3.5 shows a heat map of the magnetic field strengths within ATLAS, the image is from an article detailing the superconducting magnet system [69]. The magnet systems store a total energy of 1.6 GJ and produce fields of a combined volume of approximately  $12 \times 10^4 \text{ m}^3$ .

First time I have seen this,  
very interesting!

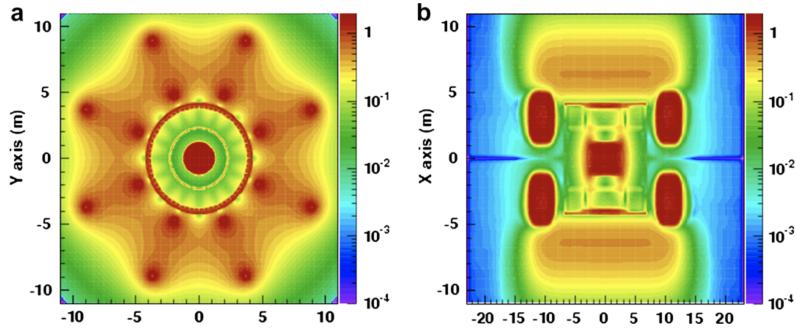


Figure 3.5: ATLAS magnetic field profile, showing a transverse cross-section in the centre of the detector (a), and a longitudinal section (b) [69]. The heat map is displayed in arbitrary units.

### 3.2.1 Tracking: The Inner Detector

The Inner Detector (ID) is comprised of a number of different tracking detector sub-systems; the pixel detectors, the semiconductor tracker (SCT) and a transition radiation tracker (TRT) as seen in figure 3.8. It covers a volume corresponding with the total  $\phi$  angle. The pixel detectors and SCT cover the range  $|\eta| < 2.5$  and the TRT covers  $|\eta| < 2.0$ . The job of the ID is to track the propagation of charged particles through the detector. This is achieved by measuring a sequence of hits for each charged particle that propagates through its volume. The sequence of hits describing the trajectory of a given charged particle must be disentangled from other hits in the detector, track-finding algorithms are applied to this end. Using these tracks we can measure; the direction of the particle, the sign of the electric charge of the particle, the rate of energy loss of the particle with respect to its distance traveled and by constructing the sagitta the transverse momentum of the particle (denoted  $p_T$ ). An example construction of the sagitta is shown in figure 3.6 and its relation to  $p_T$  is

$$S = \frac{qL^2B}{8p_T}, \quad (3.4)$$

where  $L$  is the distance between the first and last hits in the track,  $q$  is the charge of the particle, and  $B$  is the strength of the magnetic field, in this case the field produced by the solenoid magnet system. The system is made of a single layer coil with an inner diameter of 2.46 m and produces 2 T field in the axial direction

What does this mean?

What about e/π separation through transition radiation?

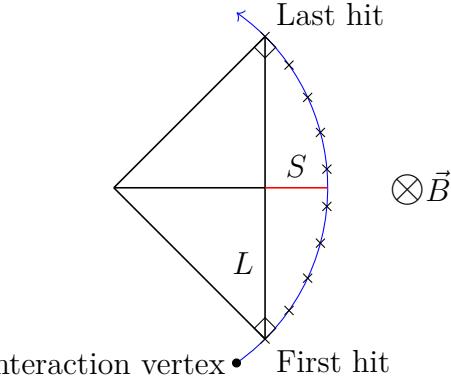


Figure 3.6: A diagram showing the geometric construction of the sagitta ( $S$ ) of a track. The track is comprised of a sequence of hits marked by crosses and the “lever-arm” distance between the first and last hit in the track is marked  $L$ . Charged tracks in a magnetic field pointing into the page, as shown, form arcs of circles. The two lines marked as normal to the track are radii of the circle to which this track’s arc belongs.

with respect to the beam-pipe. Using multiple tracks interaction vertices can also  
full stop  
 be reconstructed, the vertex which comes from the highest energy collision in a  
full stop or add “with” and drop “are”  
 given event is known as the primary vertex, the details and performance of primary  
broken reference  
 vertex finding are given in ???. Vertices coming from pile-up and glancing collisions  
 of partons also need to be reconstructed in order to separate detector response due  
 to these collisions from that of the primary vertex. So-called impact parameters

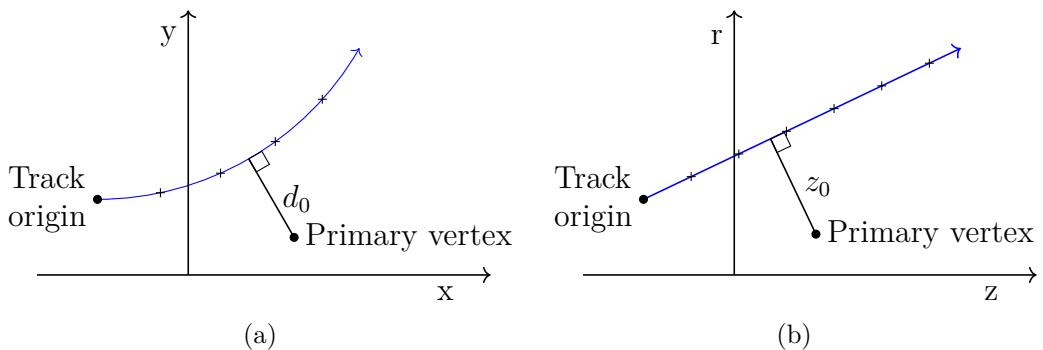


Figure 3.7: A diagram showing the geometric construction of the impact parameters  $d_0$  and  $z_0$ .

$d_0$  and  $z_0$  are used to identify secondary and pile-up vertices, their construction is shown in figure 3.7. These parameters are also used to perform identification of particles within the detector such as  $b$ -quarks and  $\tau$ s. Tracks are also used to

calculate the calorimeter impact point and in general match activity in outer regions of the detector to an interaction vertex.

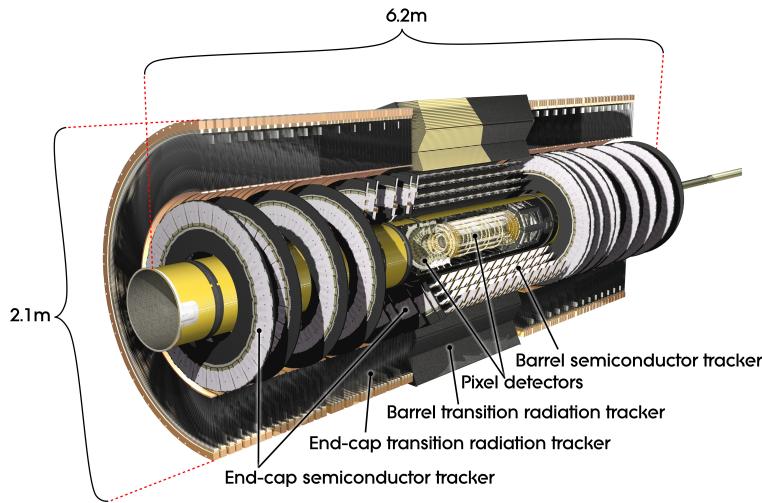


Figure 3.8: Computer generated image of the ATLAS inner detector [70].

With the exception of the TRT the tracking sub-systems are silicon based detectors. The silicon detection medium acts as a reverse bias diode. Charged particles incident on the silicon cause ionisation in the depletion layer. The products of this ionisation are electrons and holes (excess pockets of positive charge in the silicon) which produce a signal that must be handled by a read-out system. The signal is referred to as the charge collected. Application-specific integrated circuits (ASICs) are used to readout the signal performing the analogue to digital conversion. The combination of the detection medium, readout system and the printed circuit board (PCB) on which they are joined is referred to as a module.

**collected charge?**

## Pixel Detectors

There are four layers of pixel detectors that are the closest components of the ID to the beam-pipe. Pixel detectors are silicon detectors where the diodes are approximately square in shape, giving the benefit of being able to resolve hits in two directions. The design originally had three layers, each  $250\ \mu\text{m}$  thick with  $50\ \mu\text{m}$  by  $250\ \mu\text{m}$  pixels, of oxygen doped n-type silicon crystals. During LS1 a fourth layer, closest to the beam-pipe (which was also replaced for a smaller radius version) was

**far from square**

added. This layer is known as the insertable B-layer (IBL) [71], the motivation for its addition was to counteract degradation of original performance of the ID due to irreversible damage by radiation. As well as the inclusion of the IBL, performance degradation is mitigated by increasing the bias voltage across the pixels from 100 V (their starting voltage) to up to 600 V. Additionally the IBL being closer to the beam-pipe allows for interaction vertices to be measured more precisely. The need for better reconstruction of vertices is motivated by their role in the performance of algorithms that classify jets of activity in the detector that are initiated by

full stop, or “giving the IBL its name”

B-hadrons, this is where the IBL gets its name. There are no pixel detectors in the end-caps. Each pixel is small in size which mean many can fit on one module, all requiring their own conductor for readout. The solution to this challenge is to use a complex process known as bump bonding, which is both expensive and time consuming.

### Semiconductor Tracker

Next closest to the beam-pipe is the SCT whose modules have long thin strip shaped diodes. The strips provide high resolution in only a single direction. In contrast to the n-type silicon of the pixels, the strips are made from p-in-n type silicon. Each

not really wafers, as they are round.

SCT module is comprised of two back to back wafers such that the orientation of the strips are offset by a small angle in order to improve coverage. Each strip is covered

full stop

in a metalised layer, the strips are separated by a distance of 80  $\mu\text{m}$ . A rather old and dirty SCT module left over from the quality control testing stage of production that took place at Queen Mary University of London can be seen in figure 3.9. As can be seen in the figure the SCT modules are wire bonded to their ASICs which is cheaper and faster than the bump bonding used on the pixel modules. In order to calibrate the response of the strips a 100  $\text{M}\Omega$  poly-silicon resistor is located at the end of each strip. Figure 3.10 shows an image of the snake-like structure of a poly-silicon resistor from the end of an SCT module. The modules come in two different designs, short strips and long strips with the short strips forming the layer closest to the pixel detectors and the long strips on the outside. The original operating

layer implies a plane,  
better: metalisation on  
top

I don't think it is faster  
for a large  
production....

**100 $\text{M}\Omega$**   
**resistor:**  
**calibration?**

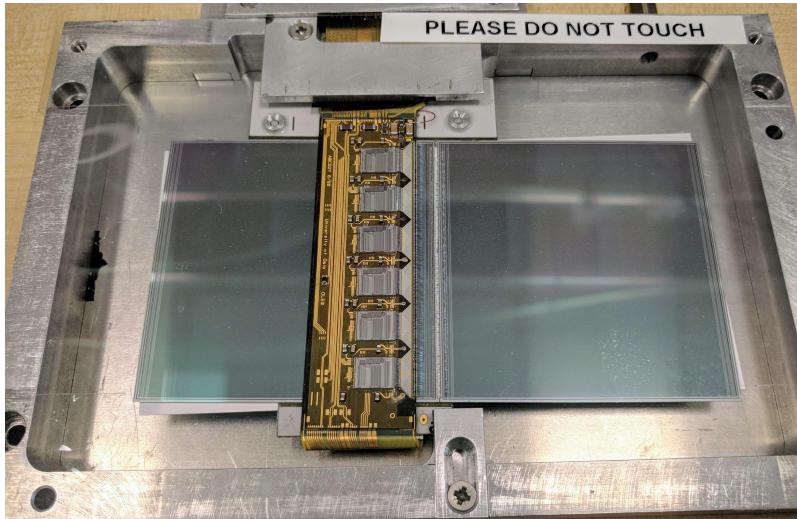


Figure 3.9: An image of an SCT long strip module mounted in a rig for testing at Queen Mary University of London.

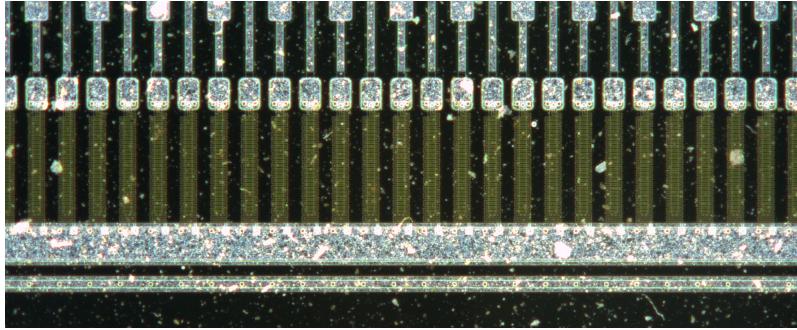


Figure 3.10: A close up image of the end of an SCT sensor in which the snake-like poly-silicon resistors are visible as a yellowish coloured structure at the end of each strip. This image was taken with a high resolution automatic area scanner commissioned by the author [72] in order to take full scans of strip sensors during the production of the ATLAS Inner Detector upgrade known as the Inner Tracker (ITk) [73, 74].

bias voltage was 150 V but again due to radiation exposure this will raise to up to 350 V over time as necessary. There are four layers of semiconductor trackers in the

**Explain the tilt.** barrel arranged so that sensors have a tilt with respect to a perfect coaxial cylinders  
**Just the stereo** of approximately 11°. This tilt increases the amount of material that particles will angle?! travel through and is optimized to the geometry of the detector. Similarly the end-cap modules are arranged in petal like structures, with a number of different geometric designed based on the position within the end-cap.

These are upgrade papers. Use SCT I papers

## Transition Radiation Tracker

I wouldn't call it layer,  
it is much more.  
Electron ID is not the  
primary role, that is just  
a bonus.

The final layer of the ID is the TRT, the primary role of the TRT is to aid electron identification by measurement of transition radiation. This is useful in distinguishing between electrons and charged hadrons which can leave similar signatures in the calorimeters and whose tracks are otherwise hard to distinguish. The TRT is mostly made up of polyimide drift tubes with a diameter of 4 mm. The drift tubes are filled with a gas mixture whose majority constituent is xenon. These tubes operate with a voltage of -1530 V and are contained within a carbon fibre support structure. Scintillating fibres and foils lie between the drift tubes which create the transition radiation. The geometric layout of the tubes is optimised individually for the barrel and end-caps.

### 3.2.2 Calorimeters

The purpose of the calorimeters is to measure the total energy of particles that pass through their volume, this is achievable only if the calorimeter stops the particle completely. A desirable side effect is that they also act as a barrier to stop particles passing through to the muon spectrometers, of course this means necessarily that muons pass through the calorimeters. There are two calorimeter systems in ATLAS the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The calorimeters are not immersed in a significant magnetic field compared to the rest of the ATLAS as seen in the heat map of figure 3.5. The geometric layout of the calorimeter systems, as well as the location of specific components can be seen in figure 3.11, in which the ID can also be seen (greyed out). Information from the two calorimeters is used in conjunction for any particles whose decay products propagate through both volumes. Both calorimeters are split up into cells of material that are used to determine the position of decay products in the detector.

High energy physics calorimeters function by measuring the shower of particles that are produced as a result of the incident particle losing its energy. Both calorimeters in ATLAS are sampling calorimeters, meaning that they are comprised of

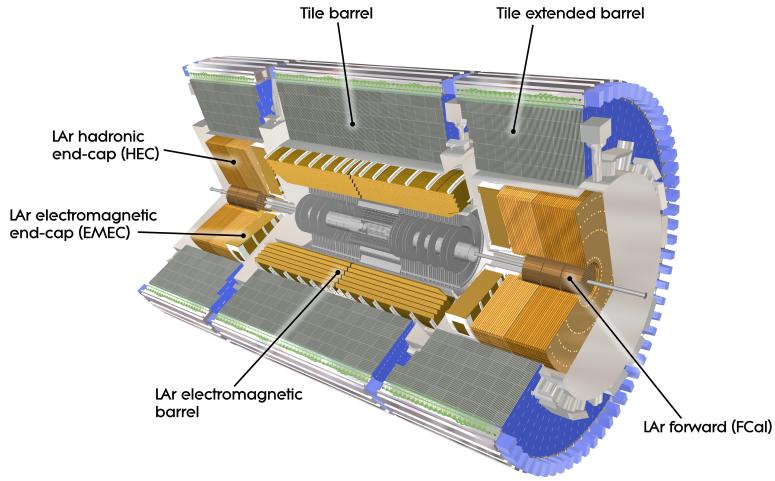


Figure 3.11: Computer Generated image of the ATLAS calorimeter [75].

alternating layers of absorber and detection medium. The purpose of the absorber medium is to provide a material in which particle showers evolve rapidly over a short distance, though they are in general not sensitive to measuring those showers. The detection medium or active material is used to then measure those showers.

In general a good calorimeter will have the following properties: it should be hermetic, meaning that there are no gaps for particles to escape through without being measured. It should have a fast response time, to keep up with the rate of collisions (actually the rate of the trigger see the following section). It should be radiation tolerant in order to perform over a long period of time. It should be able to measure energy to a high resolution in order to resolve resonances over backgrounds and have a high granularity of cells in order to accurately determine the position of energy deposits (and aid particle identification). High energy resolution is greatly aided by not having anything front of the calorimeter which may change an incident particles energy before it strikes the calorimeter itself. This requirement is therefore somewhat at odds with having a detector that has good  $p_T$  resolution for electrons, positrons and photons (muons go all the way through the calorimeter anyway) due to the process by which  $p_T$  is measured becoming more precise with a longer “lever-arm” distance  $L$  as seen in figure 3.6 and equation 3.4. A large tracker makes the calorimeter much more expensive in a barrel shaped detector as the amount of

add “in” to front

particles → particle’s

material required increases with the square of the radius of the tracker.

### Electromagnetic Calorimeter

The ECAL is primarily concerned with measuring the energy of electrons, positrons and photons. These particles primarily interact with the electromagnetic force and so produce electromagnetic showers when they lose their energy. The ECAL resides closer to the interaction point than the HCAL and has liquid argon (LAr) as its active material. LAr is a good choice for the calorimeter which is closer to the interaction point as it is naturally resistant to damage by radiation. The absorber of the ECAL is lead, it is suitable as it has a high number of nucleons ( $Z$ ) and the radiation length<sup>1</sup> of a given shower is inversely proportional to  $Z^2$ . An applied electric field causes ions produced in the EM shower to drift in such a way that the signal induced is proportional to the energy deposited by the incident particle.

### Hadronic Calorimeter

The HCAL end-caps use a LAr active material like the ECAL. The absorber is copper instead of lead and the dimensions are more suited to hadronic particle showers as opposed to electromagnetic. As noted before, LAr is naturally radiation hard which is why it is used in the end-caps which see more radiation than the barrel regions due to almost all of the momentum of the colliding bunches of particles being in the axis of the beam-pipe. The barrel section of the hadronic calorimeter is made from scintillating tiles of active material interspersed with steel as an absorber. Scintillation is the process by which a scintillation medium produces light when particles travel through it. The amount of light produced is proportional to the energy of the incident particle that initiated the shower. In order to convert the light into a digital electrical signal photo-multiplier tubes (PMTs) are used. PMTs are able to measure very small amounts of light, even the incidence of a single photon. This is achieved by initially exploiting the photo-electric effect whereby the

---

<sup>1</sup>The radiation length  $X_0$  is the thickness of material that reduces the energy of an electron by a factor of  $e$  the natural number, but is also used for other particles.

incident photon knocks an electron from a metallic part of the PMT. The current produced by this electron is then multiplied by a very large factor (up to  $10^8$  for some PMTs) by a series of dynodes (electrodes in vacuum that produce secondary emission).

### 3.2.3 Muon Spectrometers

Surrounding the calorimeters are the muon spectrometers, which form the outermost layer of the detector. Though muons are charged leptons just like

formatting

electrons, their specific properties mean that dedicated muon spectrometers are required to detect them. Muons deposit far less energy per distance traveled than other particles meaning that they punch through most materials with ease. As can be seen in figure 3.12 the components of the muon spectrometers are the thin-gap chambers, cathode strip chambers, resistive plate chambers and monitor drift tubes. The barrel and end-cap toroid magnets immerse the muon spectrometers in a magnetic field which at its peak (visible in figure 3.5) has a strength of 4 T. Despite a stronger peaking magnetic field than in the solenoid, observed muon tracks are often far less curved than that of their lighter cousins, the electrons. This is due to the larger mass of the muon. Muons do leave tracks in the ID and also deposit small amounts of energy in the calorimeters. Tracks in the muon spectrometers are matched up to tracks in the ID with the aid of the location of energy deposits in the calorimeters if possible. Tracking information for muons is used in algorithms such as overlap removal, which removes muons from jets that they have been erroneously associated with by matching the muon with its ID track.

### 3.2.4 Trigger Systems

The trigger systems in ATLAS allows data to be recorded only when an event meets certain criteria. It would be impossible to readout every interaction that occurs in the detector. The reason for this is that the geometric constraints of the detector mean that there is only a small space available for readout wires, as detection

The curvature should depend only on  $p_T$ , not the mass!

read out

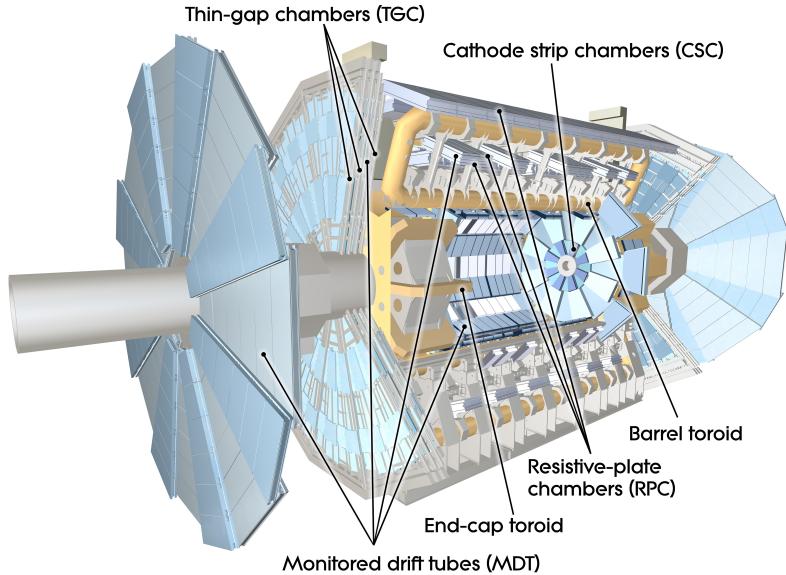


Figure 3.12: Computer generated image of the ATLAS Muons subsystem [76].

medium needs to be prioritized for sensitivity and technology limits the data rate that one can achieve through a cable of fixed area. The trigger system comes in two parts, a hardware component referred to as level one (L1) and software component referred to as the high level trigger (HLT). The L1 system is comprised of the L1 calorimeter (L1Calo) trigger which operates by searching for clusters of energy in the calorimeters and the L1 muon (L1Muon) system which coincidences in the muon systems. A third system L1 topological (L1Topo) uses regions of interest built from the L1Calo and L1Muon data which are passed to central trigger processors for selection. The various limitations of the hardware mean that these selections must be passed up to the next level of triggering, the HLT in a time window of  $2.5 \mu\text{s}$ . The HLT takes information from the L1 systems and uses faster versions of an offline style analysis in order to select or reject events for readout. The rate of the HLT was about 1.2 kHz on average over the period of data taking relevant to this measurement. In order for a trigger to fire an event must pass fully all of the requirements of one of the algorithms defined by an extensive trigger menu. More information about the triggers used in the  $VH \rightarrow b\bar{b}$  analysis will be given in a later chapter.

# Chapter 4

## Machine Learning

This used “ä and plural here...” The next chapter will focus on techniques in machine learning which are referred to interchangeably as a Multi-variate Analyses (MVA). These techniques have become very widespread in the field, they are used in several of the reconstruction algorithms that are used to obtain the events on which the analysis is performed. Furthermore an MVA is used to obtain the distribution that acts as the final discriminant for the analysis, and machine learning techniques are also used to quantify systematic uncertainties on background modelling. The algorithms at the core of these techniques are described here before diving into specific applications.

The two main algorithms used are Boosted Decision Trees (BDTs) and Neural Networks (NNs), which will be described in sections 4.1 and 4.2 respectively. These algorithms are used in many places outside of physics and so rather than referring to individual pieces of data that enter into the algorithm as an event, in this chapter they will be referred to as an example. This terminology comes from the fact that in general these algorithms must be shown a large number of examples before they are suitably “trained” for their purpose, and that in general those examples could be data that represent anything. Both of these algorithms can be operated in classification or regression modes. The main difference between these modes is that classification mode provides a score for each of a given number of classes, which can be interpreted as a probability that a given example belongs to the given class, whereas regression outputs a single number per example whose interpretation

depends on the problem.

Both algorithms can be written as a function of some inputs  $\vec{x}$ , some weights to be found during training  $\vec{w}$ <sup>comma</sup>, and a set of hyper-parameters  $\vec{\theta}$  as follows

$$F(\vec{x}, \vec{w}, \vec{\theta}) = \vec{y}, \quad (4.1)$$

outputs → elements

where  $\vec{y}$  is a vector whose outputs correspond to each class in a classification problem. If the algorithm is set up for a regression problem then the output will just be one number. The hyper-parameters control the behaviour of the algorithm in question and are set by hand in advance of training. Training either algorithm involves an iterative process where at each iteration the function is evaluated for the current set of weights, the outputs are compared against some truth labels  $\vec{t}$  via the computation of a loss function. Based on the value of the loss function the weights are then updated according to the given algorithm. It is therefore vital that firstly these truth labels are available for the data on which one wants to train (e.g. train on simulated predictions rather than real data) and secondly that the examples are split into a training and testing set so that the set which is used to iteratively update weights is not the same set on which performance is evaluated. This avoids over-fitting to noise in a given set, though this problem is not circumvented entirely.

## 4.1 Boosted Decision Trees

Decision trees have a structure as in figure 4.1, which shows a tree dividing examples into two classes, red and blue. Each circular node in the tree represents a cut on one of a number of variables provided as input to the algorithm. The tree is read top to bottom with each node being followed by two edges branching left and right that represent the path taken by examples which pass or fail the cut respectively. Square nodes represent that the termination criteria have been reached and that events in these nodes have been classified according to the colour of the node. The variable chosen at each node is optimised in order to maximise a criteria related to

cut is colloquial. use selection criteria or similar

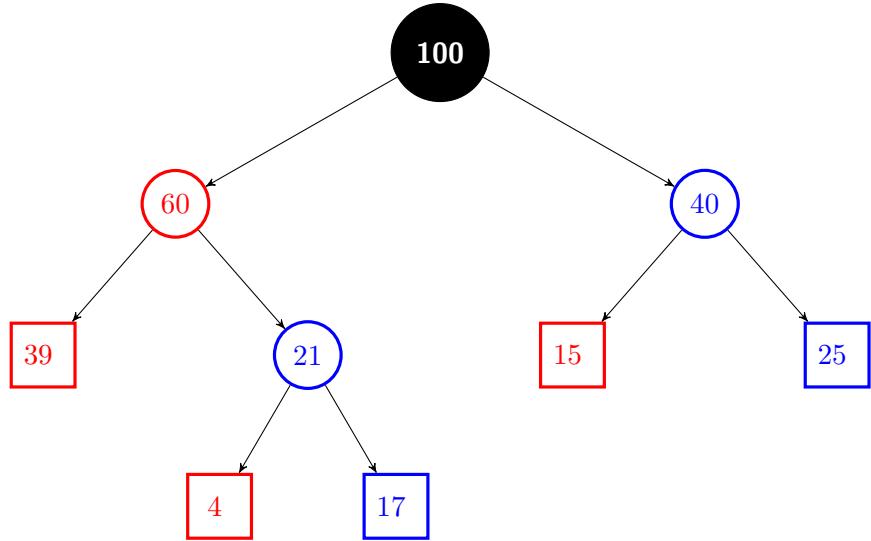


Figure 4.1: The structure of a decision tree set up for a classification problem. This diagram serves as a summary of where 100 examples end up after being passed through the tree. The number in each node corresponds to the number of examples that pass through that node.

the separation of classes. For a problem containing two classes a common separation criteria is the Gini index,

$$G = p(1 - p), \quad (4.2)$$

where  $p$  is the purity of a chosen class that one wants to maximise.

comma

A decision tree by itself is able to separate examples into a two classes, however a single tree is prone to over-fitting. The reason that decision trees are susceptible to this is that if two variables yield a similar separation criteria then a fluctuation in the training data may lead to the choice of one variable over another for a particular node, this choice will lead to a very different tree structure than if the fluctuation were not present.

In order to mitigate the over-fitting tendencies of decision trees they are often used in an ensemble algorithm such as bagging [77] or boosting [78]. Here only boosting will be discussed. Ensembles of decision trees are often referred to as random forests. Boosting works by training a sequence of trees and then weighting misclassified events from a tree so that they have more influence over the structure of the next tree in the sequence. The final classification of any given example is a weighted average over all trees, this can be weighted by the overall accuracy of each

tree, but in general can take any weighted average.

### 4.1.1 Gradient Boosting

Gradient boosting is the name of an optimisation algorithm that takes the concept of boosting and combines it with the gradient descent algorithm. A pictographic representation of gradient descent for a regression problem can be seen in figure 4.2.

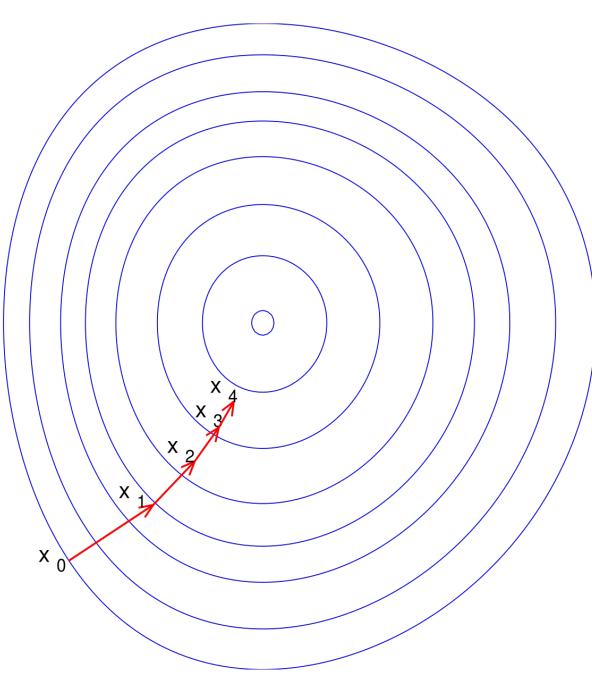


Figure 4.2: An illustration of the gradient descent algorithm. The blue lines represent level sets of a function, sets of points that have the same gradient. The  $x$ 's represent where in the range of the function the output calculated with the weights at the current step is. comma Five steps are shown numbered 0–4.

The basic principle of gradient descent is to find the minimum of some multi-variable function by taking the derivative of the function around some starting point in the space and moving in the direction of the negative gradient. This process is repeated iteratively until some termination criteria is reached. In machine learning gradient descent is most often applied to the loss function which describes the discrepancy between the predictions of a model  $\vec{y}$  and the truth labels  $\vec{t}$ . For a

loss written as

$$\mathcal{L}(\vec{y}, \vec{t}), \quad (4.3)$$

the predictions and the labels define a point in the space around which the function  
full stop  
must be locally differentiable, it is hard to determine where these points will be before training an algorithm on a given dataset and so in general a loss is chosen that is differentiable everywhere.

Considering now a boosted decision tree that has a number of iterations for each of which a decision tree is constructed using the Gini index. The model can be written as

$$F_i(\vec{x}), \quad \text{no comma} \quad (4.4)$$

at a given iteration and is evolved by the addition of a single decision tree as

$$F_i(\vec{x}) + d(\vec{x}) = F_{i+1}(\vec{x}), \quad \text{full stop} \quad (4.5)$$

the decision tree which is added is found by taking the negative gradient of the loss computed at the previous iteration    comma

$$d(\vec{x}) = -\frac{\partial \mathcal{L}(F_i, \vec{t})}{\partial F}. \quad (4.6)$$

no is

This process is known as gradient boosting. Each decision tree is known as a weak learner, by evolving the overall model by stepping in the direction of the negative gradient of the loss at the current step the algorithm aims to correct for mistakes of each weak leaner.

The aforementioned hyper-parameters that control the properties of the model come in a few forms for the BDT. They include but are not limited to, the maximum tree depth and the number of weak learners to train. The choice of boosting algorithm can also be considered a hyper-parameter, gradient boosting has risen in popularity recently but AdaBoost [79] remains popular and is still used in some of the reconstruction algorithms.

## 4.2 Neural Networks

Like boosted decision trees, neural networks can be described as a function of input data, weights and hyper-parameters  $F(\vec{x}, \vec{w}, \vec{\theta})$ . Unlike BDTs neural networks can vary a lot more in their structure, the hyper-parameters of these algorithms allow for much finer control over the behaviour of the function as an estimator. The basic structure of a neural network is shown in figure 4.3. The building blocks of the NN resemble Fisher discriminants [80] and take the form

$$a_j = \sum_{i=1}^d w_{ji}x_i + w_{j0}, \quad (4.7)$$

full stop

where the  $w_{ji}$  terms are known as weights and the  $w_{j0}$  as biases, these constructions are called activations. It would be remiss of me to cite the works of Fisher without condemning his participation in the field of eugenics, but I leave the citation here as a reminder that as scientists we have a responsibility to society to pursue causes for good. Neural network models are inspired by neurons in the brain, that fire when some threshold of neuro-transmitting chemical is reached [81] and are linked up in intricate ways to manifest complex behaviours. In order to take activations and give them a behaviour similar to neurons they must be passed through an activation function, denoted  $\mathcal{H}$ , which gives the threshold effect

Define  $h_j$  here. Threshold effect seems an unlikely definition.

$$h_j = \mathcal{H}(a_j). \quad (4.8)$$

These are similar to Rosenblatt's original perceptron [82] with the difference that  $\mathcal{H}$  must be differentiable whereas the perceptron used a step function.

All neural networks have an input layer which is a vector of input data and an output layer, a vector whose size relates to the predictions being made. Layers inbetween these are called hidden layers and are made of units  $h_j$ , known as hidden units. The most simple neural network has a single hidden layer. This model can

be written as

$$F(\vec{x}, \vec{w}, \vec{\theta}) = \mathcal{O} \left( \sum_{j=1}^m w_{kj} \mathcal{H} \left( \sum_{i=1}^d w_{ji} x_i + w_{j0} \right) + w_{k0} \right). \quad (4.9)$$

The hyper-parameters here are the choice of activation function  $\mathcal{H}$ , output function  $\mathcal{O}$ , number of hidden units  $m$  and the number of hidden layers.

The output function  $\mathcal{O}$  must have properties that allow us to interpret the outputs of the network as probabilistic i.e. comma the sum of elements in the output must equal 1. A common choice of output function for classification problems is the softmax function

Use \exp in latex

$$\mathcal{O}(z)_c = p(c|\vec{x}) = \frac{\exp(z_c)}{\sum_{i=1}^k \exp(z_i)} \quad \text{comma} \quad (4.10)$$

where  $z$  merely denotes the argument of the output function. This function gives the probability that an example represented by data  $\vec{x}$  belongs to class  $c$  where there are  $k$  possible classes. full stop

By controlling the hyper-parameters a network of any size and shape can be built, the network function can be written in general as the cumbersome

$$F(\vec{x}, \vec{w}, \vec{\theta}) = \mathcal{O} \left( \sum_{j_n=0}^{m_n} w_{kj_n} \mathcal{H}_n \left( \dots \mathcal{H}_2 \left( \sum_{j_1=0}^{m_1} w_{j_2 j_1} \mathcal{H}_1 \left( \sum_{i=0}^d w_{j_1 i} x_i \right) \right) \dots \right) \right). \quad (4.11)$$

This behaves in a similar way to the network with one hidden layer but now each layer's width and activation must be chosen individually, though often they are set to the same values. A diagram of an arbitrarily sized neural network is shown in figure 4.3.

w here .... or encapsulate w in commas

It is clear that the number of parameters that need to be learned  $\vec{w}$  and the number that need to be set by hand to tune the model to a given problem  $\vec{\theta}$  are much larger than in the BDT algorithm. For this reason adoption of neural networks is sometimes hampered by lack of data or sufficient time to tune the hyper-parameters to achieve satisfactory performance. loss functions

The loss has not been discussed in detail here but two common choices of function

same for \theta

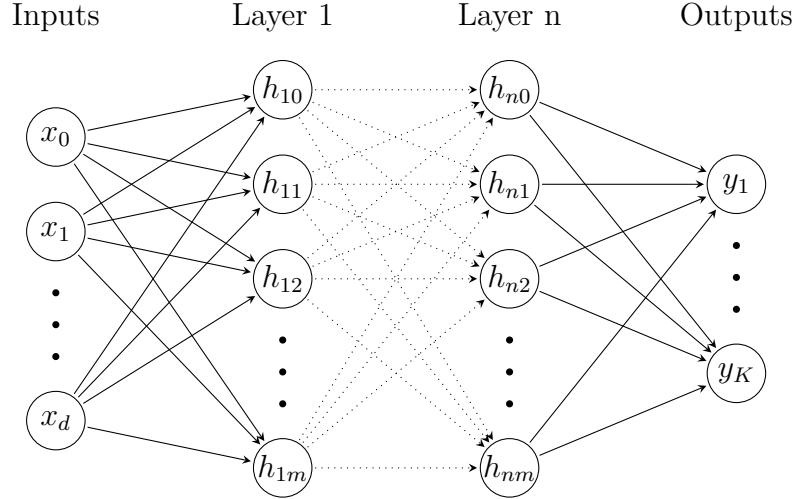


Figure 4.3: A more complex neural network containing an input layer of  $d$  nodes corresponding to data of dimensionality  $d$ ,  $n$  hidden layers of  $m$  hidden units each  $h_{ij}$  (where  $i$  indexes hidden layer and  $j$  indexes a particular unit) and an output layer of  $K$  predictive units  $y_k$ .

are the mean squared error

$$E(\vec{w}) = \frac{1}{2} \sum_{n=1}^N (y(\vec{x}_n, \vec{w}) - t_n)^2, \quad (4.12)$$

where  $t_n$  are the labels for the given data entries  $\vec{x}_n$  and the generally superior cross-entropy [83]

$$E(\vec{w}) = - \sum_{n=1}^N \left( t_n \ln(y_n) + (1 - t_n) \ln(1 - y_n) \right). \quad (4.13)$$

Likewise the choice of optimiser has not been discussed, a large number are available though in general the choice of optimiser depends on what is available in the software that one is using. For this reason minuit [84] is often used in this work. In general it would be preferred to use something more modern such as adaptive moment estimation or ADAM [85]. ADAM is a variant of the gradient descent algorithm, which is widely used and has spawned many other variants [86].

# Chapter 5

## Reconstruction and Selection

At the top level events are required to have one Higgs boson candidate and one vector boson candidate. In all cases a Higgs boson candidate is comprised of two  $b$ -jets. More on the jet collection and  $b$ -tagging strategy ~~used~~ in section 5.4. Vector boson candidates are characterised by a number of different decay products defined ~~full stop~~ in section 5.2, these decay products trigger the recording of events, specific triggers used are discussed in section 5.3. Reconstruction of basic quantities as well as higher level algorithms such as overlap removal are handled by Athena and the CxAOD Framework, more on these in section 5.1. Finally with all quantities reconstructed ~~full stop or drop “these are”~~ events are categorised in different analysis regions, these are described in section 6.1.

My contributions to the reconstruction and selection include maintenance of the CxAOD Framework, running production campaigns to produce new datasets when impactful changes have been made further up the analysis chain, and implementing new systematic uncertainties in the CxAOD Framework (not discussed until chapter 7).

### 5.1 Athena and the CxAOD Framework

Data recorded by the ATLAS detector is passed through the central collaboration software framework Athena before entering the analysis level data processing. Athena is responsible for the steps shown in figure 5.1. As can be seen in the figure ~~comma~~

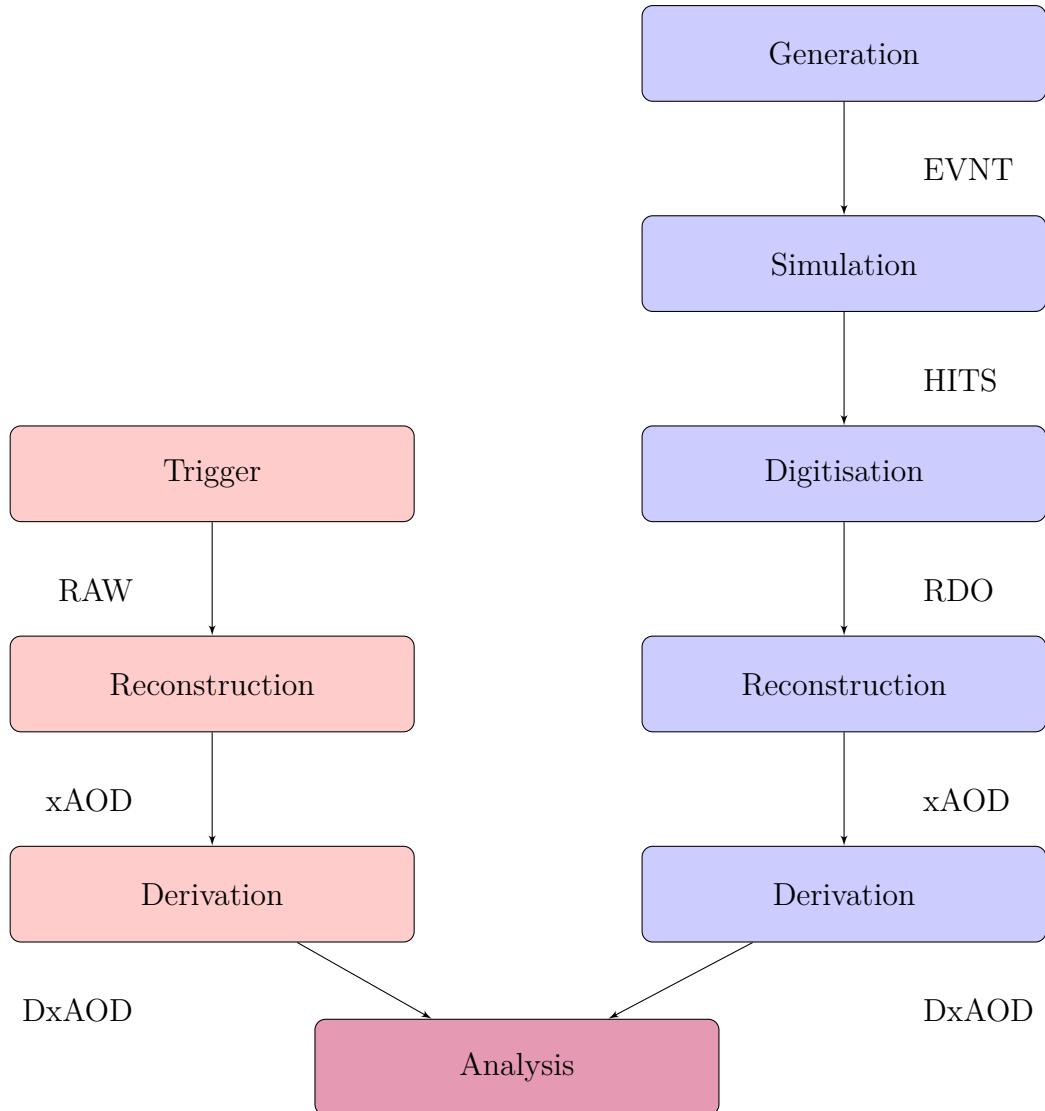


Figure 5.1: A flow chart showing the Athena data processing chain. Red nodes indicate the presence of data recorded from collisions and blue nodes indicate the presence of Monte-Carlo simulated events.

Athena processes both data recorded from collisions and Monte-Carlo simulated predictions. Steps up to and including reconstruction are required to transform the raw or simulated read-out of the detector into what are known as physics objects. These physics objects correspond to ultra-violet and infra-red safe descriptions of particles and hadron showers e.g. leptons and jets. Given the initial transverse energy of the collisions (zero) any missing transverse energy ( $E_T^{\text{miss}}$ ) is also reconstructed based on the sum of the transverse energy of all objects in an event, this missing energy can indicate the presence of particles in the event that cannot be detected by any of the ATLAS subsystems, or mis-measurement. The only particles in the Standard

Model for which this is expected are neutrinos. The files containing are referred to as Analysis Object Data (xAOD). After reconstruction a part of Athena called the derivation framework is used to produce skimmed and slimmed xAODs known as Derived xAODs (DxAODs). The reduction of these files is carried out based on a loose selection criteria.

DxAODs are the usual starting point for analysis level software, in the case of this analysis the CxAOD Framework. As in figure 5.2 The CxAOD Framework

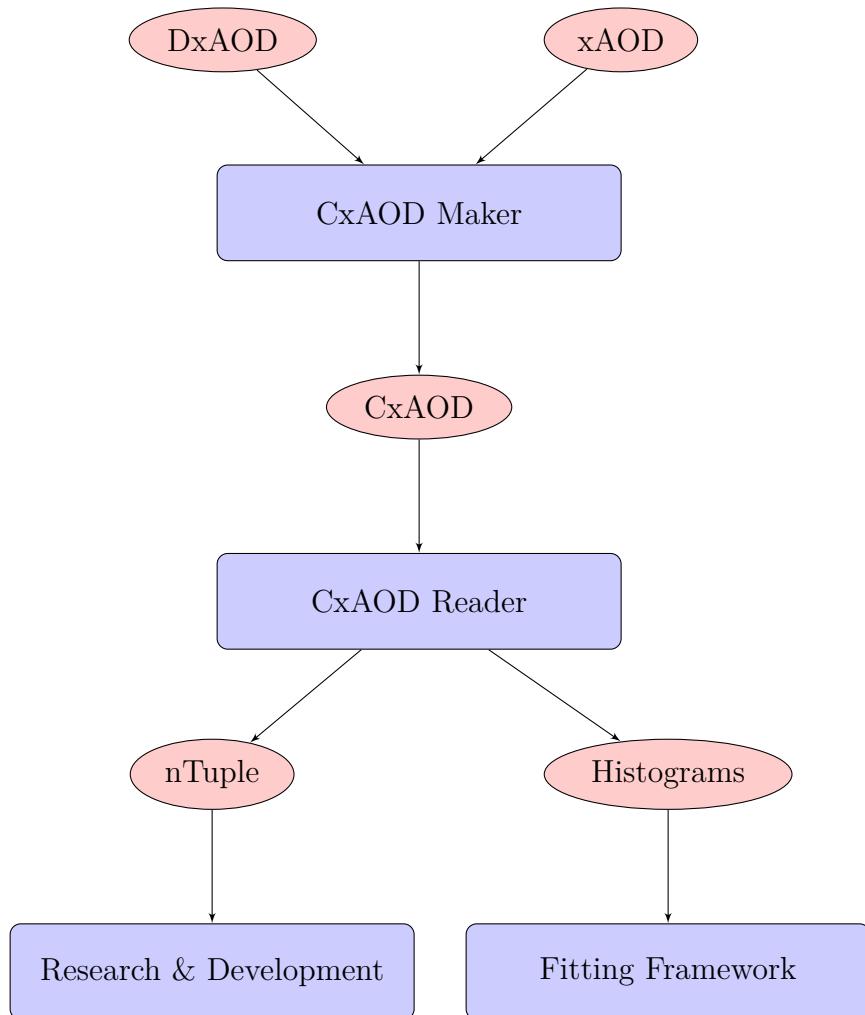


Figure 5.2: A flow chart showing the CxAOD Framework data processing chain. The red elliptical nodes indicate data formats and the blue rectangular nodes indicate software modules.

has two major modules the Maker and the Reader. The job of the Maker is to further slim the data by performing pre-selection cuts and also to apply calibrations which will be detailed below, the output of the Maker is called a Calibrated xAOD (CxAOD). The Reader takes a CxAOD as input and performs the analysis event

semi colon  
selection, it can output histograms or  $N$ -tuples.

The Maker applies the selections of each of the three analysis channels, defined in section 5.7. CxAODs are produced separately for each of the three channels as the different background compositions and signal signatures require different optimisation. Pre-selection is performed on jets based on requirements of transverse momentum and pseudo-rapidity. A tool known as the Jet Vertex Tagger (JVT) is used to remove jets resulting from pileup.

How does JVT work? p65

Reduction of pileup in Run III?

## 5.2 Leptons

The channels of the  $VH \rightarrow b\bar{b}$  analysis are categorised by the number of observed charged leptons ( $e$  or  $\mu$ ) in the decay of the vector boson. There is one channel for the study of  $WH \rightarrow b\bar{b}$  decays where the leptonic decay  $W \rightarrow \ell\nu$  yields a single charged lepton, the 1-lepton channel. There are two channels for the study of  $ZH \rightarrow b\bar{b}$  decays, the 0-lepton channel where  $Z \rightarrow \nu\nu$ , and the 2-lepton channel where  $Z \rightarrow \ell\ell$ .

Multiple classifications of lepton are defined in order to categorise events into the individual channels of the analysis, these are called  $VH$ -loose and  $VH$ -signal leptons (which are broken down into  $WH$  and  $ZH$ ), channels are defined as disjoint sets by requiring different numbers of both lepton categories. These classifications are defined in table 5.1. The characteristics of the fake lepton background from QCD

Name	$p_T$	$ \eta $	ID	$d_0^{sig}$	$ \Delta z_0 \sin $	Isolation
<b>electrons</b>						
$VH$ -loose	$> 7$ GeV	$< 2.47$	LH Loose	$< 5$	$< 0.5$ mm	FCLoose
$ZH$ -signal	$> 27$ GeV	$< 2.47$	LH Loose	$< 5$	$< 0.5$ mm	FCLoose
$WH$ -signal	$> 27$ GeV	$< 2.47$	LH Tight	$< 5$	$< 0.5$ mm	FixedCutHighPtCaloOnly
<b>muons</b>						
$VH$ -loose	$> 7$ GeV	$< 2.7$	Loose quality	$< 3$	$< 0.5$ mm	FixedCutLoose
$ZH$ -signal	$> 27$ GeV	$< 2.5$	Loose quality	$< 3$	$< 0.5$ mm	FixedCutLoose
$WH$ -signal	$> 27$ GeV	$< 2.5$	Medium quality	$< 3$	$< 0.5$ mm	FixedCutHighPtCaloOnly

Table 5.1: Definitions of the  $VH$  leptons used to define and select events for the three analysis channels where  $d_0^{sig}$  is the measured with respect to the beam line. The requirements in ID and Isolation are listed by their names in Athena which provides the tools necessary to apply them.

multi-jet processes differs between the 1– and 2–lepton channels hence the reason for two different categorisations. In general to suppress this kind of background leptons are required to be isolated from other detector activity.

### 5.2.1 Electrons

As mentioned in chapter 3 electrons leave tracks in the ID and energy deposits in the ECAL. Reconstructing electrons requires clustering the energy deposits in the ECAL, this is achieved with a sliding window algorithm [87]. Clusters must then be associated to tracks in the ID, a Gaussian Sum Filter [88] is used to account for energy loses due to bremsstrahlung radiation. The energy for electron candidates must be calibrated before it can be used in order to account for things such as non-uniformity in the detector response. Calibration is achieved by using simulated cluster activity from single particles to train a BDT regression model designed to regress the measured energy in the ECAL to the simulated energy. An in-situ data driven correction is applied to normalise the response between data and simulation [89].

Reconstruction alone is not enough to find electrons, other particles may leave similar signatures in the ATLAS sub-detectors and therefore electron identification must also be performed. Identification is performed using a likelihood-based method. Variables which have power to discriminate between electrons and other particles are used in the likelihood such as shower profiles, track quality, how closely track and cluster positions match in  $\eta$  and  $\phi$ , and the presence of a high-threshold TRT hit. This is one of the main benefits of the TRT. Performance of this method is well studied [90, 91].

### 5.2.2 Muons

Finding muons in the detector requires consideration of the coverage of the different ATLAS sub-detectors, especially as in general muons are not stopped in the detector. Muons leave charged tracks in the ID and the muon spectrometers which have

coverages of  $|\eta| < 2.7$  and  $|\eta| < 2.5$  respectively. For the region  $|\eta| > 2.5$  a stand-alone algorithm which doesn't use ID tracks can be used. All muons within the coverage of the ID require good quality ID tracks [92, 93]. A combined algorithm is used in the majority of cases. For the region  $|\eta| < 0.1$  two specialised algorithms SegmentTagged and CaloTagged are used, which require only muon segment and calorimeter deposits respectively. All aforementioned algorithms are used together in what is known as a unified chain [94–96] to reconstruct and identify muons.

### 5.2.3 Taus

The only charged leptons that are considered in the analysis are electrons and muons, leptonically decaying taus will include these as the only visible decay products, hadronically decaying taus must be considered separately. Decays are considered as one or three pronged based on the number of charged decay products, pions, with neutrinos and neutral pions also present. These decays are reconstructed in the calorimeters like jets with the anti- $k_t$  algorithm with  $\Delta R = 0.4$  [97] but the  $p_T$  of the tau is set to the total energy of the cluster of calorimeter cells within  $\Delta R < 0.2$ . Tau candidates must have  $p_T > 20\text{GeV}$ ,  $|\eta| < 2.5$  excluding  $1.37 < |\eta| < 1.52$ , and either exactly 1 or 3 tracks. A BDT based method for tau identification is used to reject fakes. The number of medium quality taus is included in each event [98, 99].

## 5.3 Triggers

The decay products of the vector boson candidate are used to trigger the recording of events for this analysis. Important triggers are the  $E_T^{\text{miss}}$ , single electron and single muon triggers. Note that it is not necessary to trigger on both charged leptons coming from the Z boson in the 2-lepton channel, the presence of one lepton allows the triggering to occur and the requirement of 2 leptons can be imposed at a later stage. Some events will be missed by not using a di-lepton trigger, however these amount to approximately only 5% of the total [100]. The list of triggers used as they

appear in the ATLAS trigger menu are shown in table 5.2, for the  $E_T^{\text{miss}}$ , electron and muon triggers respectively.

Trigger Name	Period	Threshold	Description
<b><math>E_T^{\text{miss}}</math></b>			
HLT_xe70_L1XE50	2015	70 GeV	Seeded using the level L1_XE50
HLT_xe90_mht_L1XE50	2016 (A-D3)	90 GeV	(L1_XE55) LAr and Tile calorimeter triggers, calibrated at the EM scale, with a threshold of 50(55) GeV.
HLT_xe110_mht_L1XE50	2016 ( $\geq D4$ )	110 GeV	
HLT_xe110_pufit_L1XE55	2017	110 GeV	
HLT_xe110_pufit_xe70_L1XE50	2018	110 GeV	
<b>electrons</b>			
HLT_e24_lhmedium_L1EM20VH	2015	24 GeV	Seeded using L1EM20VH level 1 trigger calibrated at the EM scale with a threshold of 20 GeV, and require medium ID quality.
HLT_e60_lhmedium	2015	60 GeV	Seeded using L1EM20VH level 1 trigger calibrated at the EM scale with a threshold of 20 GeV, and require medium ID quality.
HLT_e120_lhloose	2015	120 GeV	Seeded using L1EM20VH level 1 trigger calibrated at the EM scale with a threshold of 20 GeV, and require loose ID quality.
HLT_e26_lhtight_nod0_ivarloose	2016 – 2018	26 GeV	Tight likelihood ID required, and variable loose isolation required
HLT_e60_lhmedium(_nod0)	2016 – 2018	60 GeV	Medium ID likelihood required
HLT_e140_lhloose(_nod0)	2016 – 2018	140 GeV	Loose ID likelihood required
HLT_e300_etcut	2018	300 GeV	No ID requirements.
<b>muons</b>			
HLT_mu20_iloose_L1MU15	2015	20 GeV	Seeded using L1MU15 level 1 trigger with a threshold of 15 GeV, and requiring loose isolation requirements.
HLT_mu50	2015 – 2018	60 GeV	No isolation requirements.
HLT_mu26_ivarmedium	2016 – 2018	26 GeV	Variable cone medium isolation requirements

Table 5.2: Triggers used during the 2015, 2016, 2017 and 2018 data collection periods, notation like A or D3 denote periods during the year.

### 5.3.1 0–Lepton Channel Triggers

The events in the 0–lepton channel should have a  $q\bar{q}\nu\nu$  final state. We use the  $E_T^{\text{miss}}$  triggers listed in table 5.2 as the final state will manifest in the detector as  $E_T^{\text{miss}}$  with the presence of jets. At the stage of triggering  $E_T^{\text{miss}}$  is only calculated from energy measured in the calorimeters. As muons do not deposit much energy in the calorimeters the  $W \rightarrow \mu\nu + \text{jets}$  process is used to study the trigger efficiency and derive an appropriate scale factor (the energy of the muon provides constraints on the energy of the neutrino).

### 5.3.2 1– and 2–Lepton Channel Triggers

The 1– and 2–lepton channel’s final states are  $qql\nu$  and  $qq\ell\ell$  respectively. As the 2–lepton channel events are not expected to contain significant  $E_T^{\text{miss}}$  and inefficiencies in the muon trigger are mitigated by the requirement of two muons, the single lepton triggers listed in table 5.2 are used without any  $E_T^{\text{miss}}$  triggers. In the 1–lepton channel there is significant  $E_T^{\text{miss}}$  expected due to the presence of the neutrino. Single lepton triggers are used for events with  $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$  where the  $E_T^{\text{miss}}$  triggers have yet to turn on fully, for events with  $p_T^V > 150 \text{ GeV}$  in order to mitigate inefficiencies in the single muon triggers similar  $E_T^{\text{miss}}$  triggers to the 0–lepton triggers are used in conjunction with the single lepton triggers.

## 5.4 Jets

Jets are the roughly conical structure of detector activity resulting from the hadronisation of a QCD parton. Two categories of jets are considered, signal jets and forward jets, when the number of total jets is referred to it is equal to the sum of signal and forward jets. As the Higgs candidate in every channel of the analysis consists of two  $b$ -jets, both the reconstruction of jets and the  $b$ -tagging strategy have huge impacts on the final measurements. In this section the way jets are found and reconstructed will be introduced,  $b$ -tagging will be explained in general and then the specific tagging strategy of the analysis will be detailed.

### 5.4.1 Topological Calorimeter Cluster Anti- $k_t$ Jets

The jets that are found with a given algorithm are referred to as a jet collection. The jet collection relevant to this analysis uses topological calorimeter cell clusters to reconstruct jets [101]. Clusters are formed by grouping neighbouring calorimeter cells that have significant energies over the expected noise. These clusters are then passed to the anti- $k_t$  jet finding algorithm [102]. This algorithm takes a radius parameter which governs the size of jets, a radius parameter of  $R = 0.4$  is chosen.

Pileup can cause issues with reconstruction. In general there is a desire to suppress any jets which arise from pileup. The Jet Vertex Tagger (JVT) is a likelihood-based discriminant which is used to achieve this. The primary vertex location, jet  $p_T$  and the  $p_T$  of tracks associated to a given jet, serve as inputs to the JVT which outputs a 2-D likelihood that the jet arises from pileup. The likelihood is resilient to bias arising from the jet flavour. The tool is applied only to jets in region  $|\eta| < 2.5$  and  $p_T > 120$  GeV. A cut of  $\text{JVT} = 0.59$ , is applied to all jets in the collection, this cut has an average efficiency of 92 %. The definitions of signal and forward jets can be found in table 5.3.

Jet Category	Selection Requirements
Forward Jets	jet cleaning $p_T > 30$ GeV $2.5 \leq  \eta  < 4.5$
Signal Jets	jet cleaning $p_T > 20$ GeV $ \eta  < 2.5$ JVT medium for $p_T < 120$ GeV

Table 5.3: Details of jet selection requirements where jet cleaning refers to the quality criteria specified in the JetCleaningTool, which is included in Athena [103, 104].

### 5.4.2 $b$ -tagging

It is important to distinguish jets originating from  $b$ -quarks, which form our Higgs candidate, from  $c$ -jets and  $\tau$ -jets, as well as jets originating from quarks lighter than  $c$ -quark which are categorised together as light-jets. The calculation of a discriminant which ought to separate  $b$ -jets from other jets is known as  $b$ -tagging. In order to develop such a discriminant one must use simulated predictions in which the flavour of the parton which initiated a given jet is known so that the performance of the discriminant can be validated. In simulation a jet and the parton that initiated that jet are distinctly separate objects and so a set of rules must be defined in order

to decide which jet is a  $b$ -jet and likewise for other types of jets. Those rules are as follows:

1. If a weakly decaying  $b$ -hadron is found within  $\Delta R < R_{\max}$  of the jet axis, the jet is labeled a  $b$ -jet.
2. If a  $b$ -hadron isn't found, but a weakly decaying  $c$ -hadron is found within  $\Delta R < R_{\max}$  of the jet axis, then the jet is labeled as a  $c$ -jet.
3. Otherwise, if a  $\tau$ -lepton is found within  $\Delta R < R_{\max}$  of the jet axis, the jet is labeled a  $\tau$ -jet.
4. If any one hadron or  $\tau$ -lepton matches more than one jet, the closest jet is chosen as its parent.
5. All unlabeled jets after steps 1 through 4 are labeled as light-jets.

The algorithm used to tag  $b$ -jets is the MV2c10 algorithm [105], this is a BDT which is trained on kinematic and structural information about each jet. It is setup to categorise between  $b$ -jets (signal) and a mixture of light-jets and  $c$ -jets (background). The events in the training sample are simulated  $t\bar{t}$  events that have at least one lepton coming from a leptonically decaying  $W$  boson, and hadronically decaying  $Z'$  events. The training sample has 5 million  $t\bar{t}$  events and 3 million  $Z'$  events.

The kinematic training variables that enter into the MV2c10 algorithm are the jet  $p_T$  and  $\eta$ . The structural information is more complicated, IP2D and IP3D are two algorithms based on a log-likelihood ratio discriminant of impact parameters

(see chapter 3). IP3D is defined as

$$\text{IP3D} = \sum_{i=1}^N \log \frac{p_b}{p_u} \quad (5.1)$$

where

$$p_b = P\left(\text{is } b\text{-jet} \mid \frac{d_0}{\sigma_{d_0}}, \frac{z_0 \sin \theta}{\sigma_{z_0 \sin \theta}}\right), \quad (5.2)$$

$$p_u = P\left(\text{is light-jet} \mid \frac{d_0}{\sigma_{d_0}}, \frac{z_0 \sin \theta}{\sigma_{z_0 \sin \theta}}\right), \quad (5.3)$$

and

$N$  = the number of tracks for a given jet.

IP2D has the same definition but the probabilities  $p_b$  and  $p_u$  are conditional only on the transverse impact parameter  $d_0$ . The output of two algorithms designed to find secondary vertices, SV1 and JetFitter, also enter into the MV2c10 training.

A jet is defined as  $b$ -tagged if its MV2c10 score exceeds a certain threshold. This threshold is defined as the cut that gives a pre-determined efficiency value for  $b$ -jets when applied to a  $t\bar{t}$  sample. Calibrations are available for a number of these so-called working points, these working points are shown in table 5.4.

Name	MV2c10 cut	$b$ -tagging efficiency (%)	$c$ -jet rejection	light-jet rejection
FixedCutBEff_60	0.94	61.14	22	1204
FixedCutBEff_70	0.83	70.84	8	313
FixedCutBEff_77	0.64	77.52	4	113
FixedCutBEff_85	0.11	85.23	2	28

Table 5.4:  $b$ -tagging working points available in this analysis, rejection is in the inverse of efficiency.

### 5.4.3 Pseudo-continuous $b$ -tagging

The working points defined in table 5.4 are used in a so-called pseudo-continuous mode. In this mode the MV2c10 distribution is binned with bin edges corresponding to the working points listed in the table, and as can be seen in figure 5.3. This figure shows the discrimination between events with different jet flavours. Events that

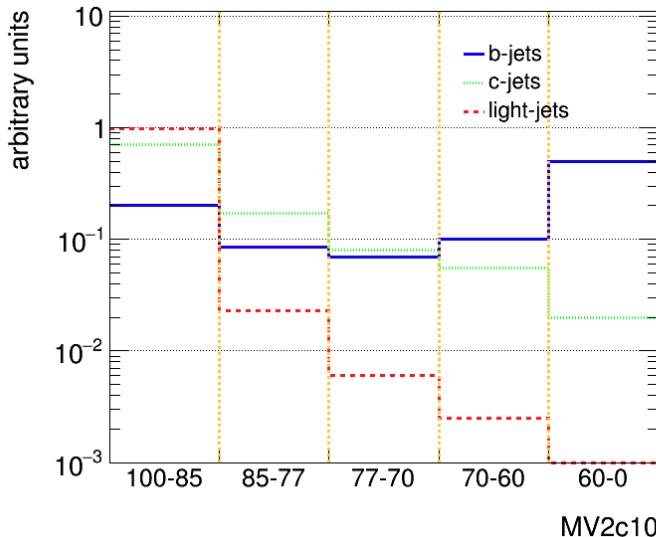


Figure 5.3: Shapes of the PCBT quantiles for  $b$ -,  $c$ -, and light-jets.

fall in the bins with ranges  $70 - 60\%$  and  $60 - 0\%$  have their MV2c10 score stored for use in the analysis. Events in the other bins are not tagged as  $b$ -jets.

#### 5.4.4 Truth Tagging

Removing all events that fail the cut associated with the  $70\%$  working point results in the loss of many events. This leaves a low number of events for analysis, which is not ideal as many of the techniques used for categorising events have a performance that is conditional on the available number of events. Instead of throwing these events away it is possible to keep all events that were simulated to have a  $b$ -hadron decay and instead weight distributions by a factor which represents the probability that an event would in fact be  $b$ -tagged and enter into the final distribution. This procedure is known as *truth tagging*, since the information as to whether a  $b$ -hadron is truly present in a simulated event is used. Tagging directly with the MV2c10 algorithm is referred to as *direct tagging*.

Given that our events contain more than one jet in all cases it is required to develop a way of generating truth tag weights for events based on all the jets in the event. This weight is calculated as the product of the  $b$ -tagging efficiency for each  $b$ -jet, multiplied by the complement of the  $b$ -tagging efficiency for each non  $b$ -jet. In case the number of jets in a given event ( $m$ ) exceeds the number of required

tagged jets in the analysis ( $n = 2$ ) all possible combinations of jets which satisfy the analysis selection are considered.

The probability to choose specific combination of truth tagged jets can be calculated as in appendix C. A single truth tagged combination can be selected based on this probability, and the event is then scaled by the factor  $w_{TT}$ . In practice due to differences between the cumulative and pseudo-continuous  $b$ -tagging efficiency distributions a modified version of truth tagging must be applied in the analysis [100].

### 5.4.5 Hybrid Tagging

There is disagreement between direct-tagged and truth-tagged events which means that truth tagging cannot directly be applied to events in the analysis. This is because direct tagging is the only strategy which can be used on data and so we can only assume that direct-tagged distributions in simulation will describe the shape and normalisation of distributions in the data. This problem is solved by implementing so-called hybrid tagging. The hybrid tagging strategy involves the following steps:

1. Divide the jets of each event in two groups, depending on the truth tag flavour:  
a group with only true  $b$ -jets and the other with non  $b$ -jets.
2. All  $b$ -jets in the first group are direct tagged.
3. The remaining group of c- and light-jets is truth tagged imposing a number of tagged jets proportional to the difference between the number of required jets in the signal region and the number of true  $b$ -jets in the first group passing the b-tag requirement.

Distributions of  $t\bar{t}$  and  $W + \text{jets}$  are shown in figures 5.4 and 5.5. Which show direct, truth and hybrid-tagged distributions. It can be seen that the agreement with respect to direct-tagged events of hybrid-tagged events is much better than that of truth-tagged events. Hybrid tagging is therefore chosen as the analysis strategy.

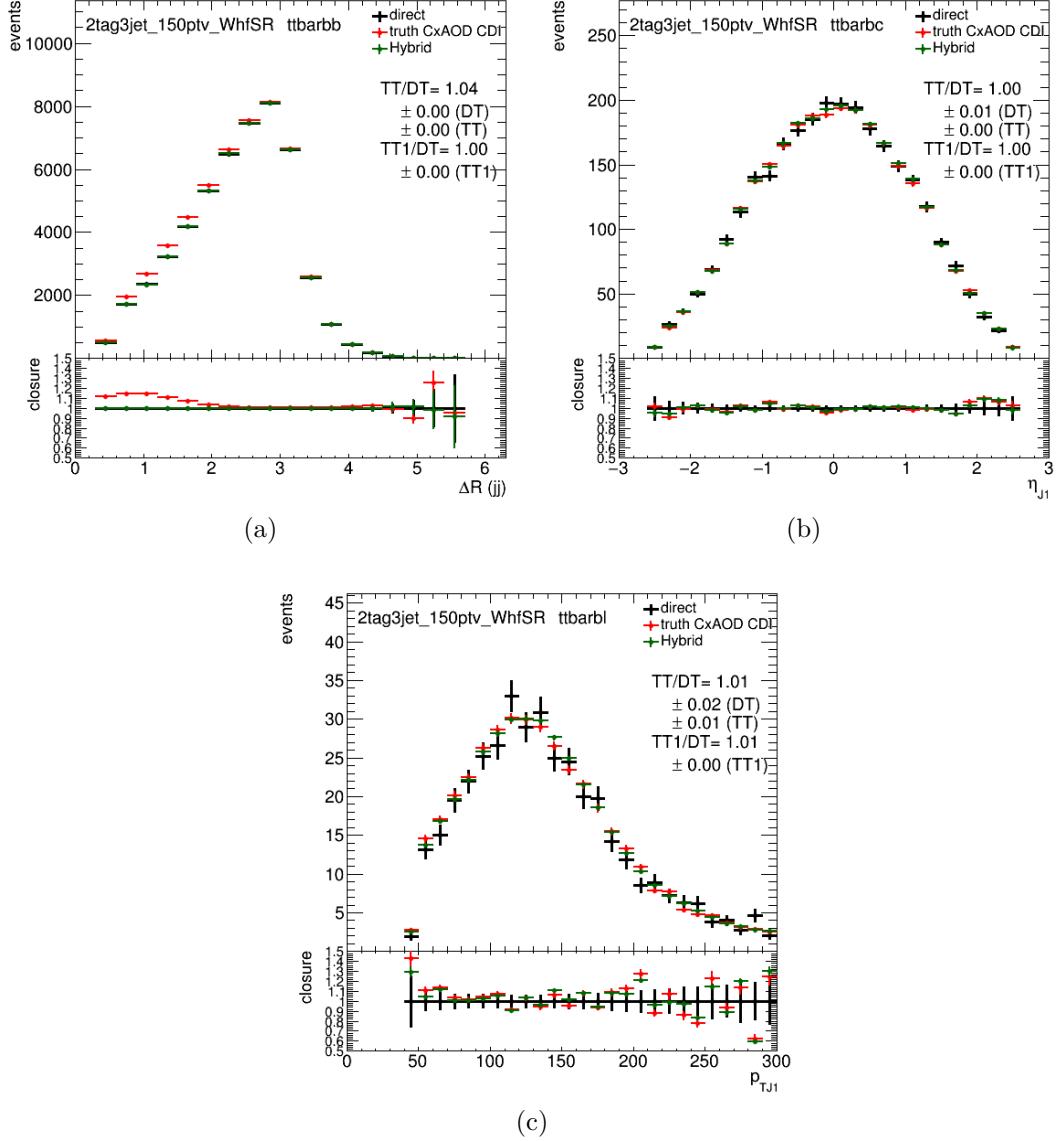


Figure 5.4: Histograms of events using the direct tagging, truth tagging and hybrid tagging strategies are shown in black, red and green respectively. (a)  $\Delta R(b, b)$  distribution of  $t\bar{t}$  events only including events with two  $b$ -jets. (b)  $\eta$  distribution of  $t\bar{t}$  events only including events with one  $b$ -jet and one  $c$ -jet. (c)  $p_T$  distribution of the leading jet in  $t\bar{t}$  events with one  $b$ -jet and one light-jet.

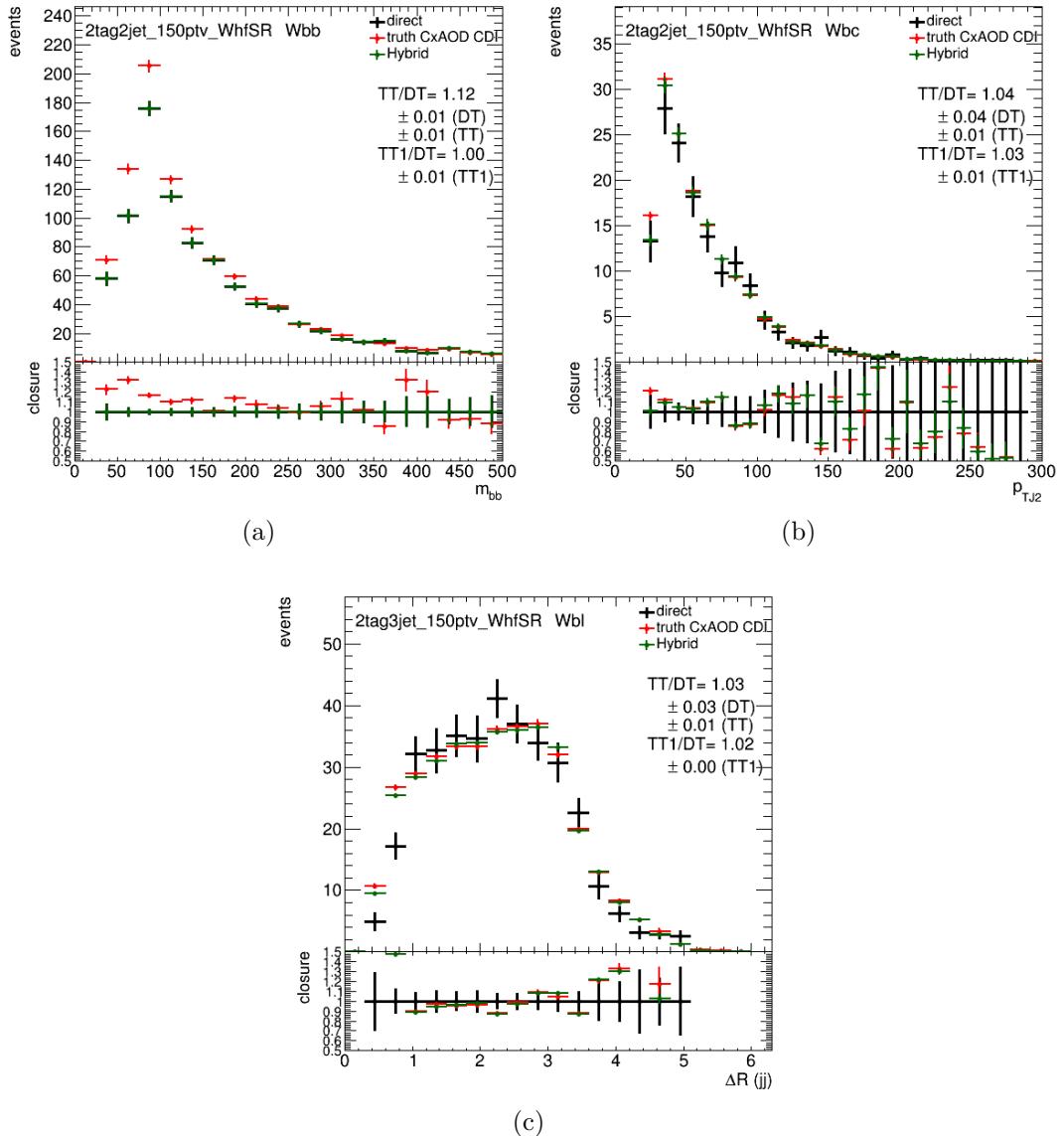


Figure 5.5: Histograms of events using the direct tagging, truth tagging and hybrid tagging strategies are shown in black, red and green respectively. (a)  $m_{bb}$  distribution in  $W + bb$  events. (b)  $p_T$  of the sub-leading jet for  $W + bc$  events. (c)  $\Delta R(jet, jet)$  distribution for  $W + bl$ .

## 5.5 Missing Transverse Momentum

This section describes how  $E_T^{\text{miss}}$  is calculated. The calculation is based on the assumption that partons entering into the hard scatter interaction have zero transverse momentum when they collide. This is not exactly true, due them having Fermi energy [106], but the transverse momentum induced by this phenomenon is minimal and therefore is ignored. Given this assumption,  $E_T^{\text{miss}}$  is calculated as the negative vector sum of the  $p_T$  of photons, electrons, muons, taus, and jets. There is also an additional soft term that enters into the sum which is made up of all good quality tracks that aren't associated with any of the aforementioned objects. These tracks must be associated with the primary vertex and therefore are robust against pileup. A second formulation of  $E_T^{\text{miss}}$  called  $E_{T,\text{Trk}}^{\text{miss}}$  is formulated just using inner detector tracks, it is even more robust to pileup but cannot account for neutral particles which do no leave tracks in the inner detector. One more variable known as  $E_T^{\text{miss}}$ -significance is also calculated as

$$E_T^{\text{miss}}\text{-significance} = \frac{E_T^{\text{miss}}}{\sqrt{\sum p_T^e + \sum p_T^\mu + \sum p_T^{jet}}}, \quad (5.4)$$

where the denominator is clearly proportional to the  $E_T^{\text{miss}}$  resolution and therefore making cuts on this variable simulates making harder cuts on  $E_T^{\text{miss}}$  with better resolution and looser cuts on  $E_T^{\text{miss}}$  with poorer resolution.

## 5.6 Overlap Removal

Overlap removal addresses what happens when overlapping detector activity could be reconstructed into more than one physics object. The treatment of overlap depends on which physics objects are involved. Below is a summary of how different pairs of overlapping objects which

are relevant to the analysis are handled, in the summary a combined muon is one which has been reconstructed with the combined muon reconstruction algorithm,

that is to say it has tracks in the ID and energy deposits in the ECAL.

- **tau-electron:** If  $\Delta R(\tau, e) < 0.2$ , the  $\tau$  lepton is removed.
- **tau-muon:** If  $\Delta R(\tau, \mu) < 0.2$ , the  $\tau$  lepton is removed, unless the  $\tau$  lepton has  $p_T > 50$  GeV and the muon is not a combined muon, then the  $\tau$  lepton is not removed.
- **electron-muon:** If a combined muon shares an ID track with an electron, the electron is removed. If a calo-tagged muon shares an ID track with an electron, the muon is removed.
- **electron-jet:** If  $\Delta R(\text{jet}, e) < 0.2$  the jet is removed. For any remaining jets, if  $\Delta R(\text{jet}, e) < 0.4$ , the electron is removed.
- **muon-jet** If  $\Delta R(\text{jet}, \mu) < 0.2$  or the muon ID track is ghost associated [107–109] to the jet, then the jet is removed if the following also holds. The jet has less than three associated tracks with  $p_T > 500$  MeV or the  $p_T$  ratio of the muon and jet is larger than 0.5 and the ratio of the muon  $p_T$  to the sum of  $p_T$  of tracks with  $p_T > 500$  MeV associated to the jet is larger than 0.7. For any remaining jets, if  $\Delta R(\text{jet}, \mu) < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^\mu)$ , the muon is removed from the jet.
- **tau-jet:** If  $\Delta R(\tau, \text{jet}) < 0.2$ , the jet is removed.

## 5.7 Final Selection

The final analysis selection varies between channels. There are however some common selections to all three channels as the Higgs candidate does not differ between them. All events must have at least two signal jets. These signal jets must be  $b$ -tagged, events with between 0 or 1  $b$ -tags are considered for study and are not used in the final analysis, events with  $\geq 3$   $b$ -tags are rejected entirely. The leading  $b$ -jet must have  $p_T > 45$  GeV. A key variable in the analysis, the di-jet mass,  $m_{jj}$  is reconstructed from the two leading jets, in the case where both jets are  $b$ -tagged  $m_{bb} \equiv m_{jj}$  and represent the mass of the Higgs candidate.

The reconstructed momentum of  $b$ -jets can be enhanced with better resolution with respect to other jets. This is achieved with so-called muon-in-jet and  $p_T$ -

reco corrections [100]. Corrections are applied after events pass the full analysis selection, but before other stages of the analysis. These corrections are not used in the calculation of any  $E_T^{\text{miss}}$  related variables.

### 5.7.1 0–Lepton Channel Selection

The vector boson candidate in the 0–lepton channel is a  $Z$  boson decaying to two neutrinos. In order to select events for this candidate a large amount of  $E_T^{\text{miss}}$  is required ( $> 150$  GeV) and there must be exactly 0  $VH$ -loose leptons in the event. Whilst events with lower  $E_T^{\text{miss}}$  may come from the physical process we desire to measure the  $E_T^{\text{miss}}$  trigger thresholds require setting the cut at it's value, the efficiency is 90% for events with  $E_T^{\text{miss}} = 150$  GeV and efficiency plateaus at  $E_T^{\text{miss}} \approx 180$  GeV.

There is a non-trivial dependence of the trigger efficiency on activity of jets in detector. This arises due to the fact that the calculation of  $E_T^{\text{miss}}$  is the total transverse energy of the event minus the transverse energy of all objects in the event. These effects are hard to model and so a requirement is put on

$$S_T = \sum_i p_T^i, \text{for } i \text{ jets in the event,} \quad (5.5)$$

such that it must be greater than 120 GeV for events with two total jets, and greater than 150 GeV for events with three total jets.

Due to the lack of charged leptons in the  $Z \rightarrow \nu\nu$  decay there is nothing to trigger on to suppress multi-jet background processes arising from QCD. This background is also enhanced due to performance limitations of the calorimeter. A set of so-called anti-QCD cuts are applied to deal with this multi-jet background, they are as follows:

- $|\Delta\phi(E_T^{\text{miss}}, E_{t,\text{trk}}^{\text{miss}})| < 90^\circ$
- $|\Delta\phi(j_1, j_2)| < 140^\circ$ , where  $j_1$  and  $j_2$  are the leading and sub-leading jets respectively,

- $|\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, h)| > 120^\circ$
- $\min(|\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \text{pre-sel. jets})|) > 20^\circ$  for 2 jets,  $> 30^\circ$  for 3 jets.

Where  $E_{t, \mathrm{trk}}^{\mathrm{miss}}$  is defined as the missing transverse momentum calculated from the negative vector sum of the transverse momenta of tracks reconstructed in the inner detector and identified as originating from the primary vertex. These cuts reduce the multi-jet background to approximately 1% of the total background in this channel.

### 5.7.2 1–Lepton Channel Selection

The vector boson candidate in the 1–lepton channel is a  $W$  boson decaying to one charged lepton and one neutrino. In order to select for events of this signature exactly one  $WH$ -signal lepton is required. At low  $p_{\mathrm{T}}$  there is increased contribution from multi-jet processes, therefore the extra requirements of  $p_{\mathrm{T}}^W > 150$  GeV and  $E_{\mathrm{T}}^{\mathrm{miss}} > 30$  GeV are applied to mitigate this. The latter cut is only applied in the electron channel.

### 5.7.3 2–Lepton Channel Selection

The vector boson candidate in the 2–lepton channel is a  $Z$  boson decaying to two charged leptons. Exactly two  $VH$ -loose leptons of the same lepton flavour are required, additionally one of the leptons must pass the  $ZH$ -signal requirements. For events with two muons, it is required that the muons are of opposite charge. Electron reconstruction suffers from a higher rate of charge misidentification (of order  $10^{-2}$  in data) and so this requirement is not applied to events with two electrons. The di-lepton invariant mass is confined to be around the  $Z$  boson mass peak, events require  $81 < m_{\ell\ell} < 101$  GeV. These cuts reduce multi-jet backgrounds to negligible levels.

The presence of two visible leptons in this channel allows for the enhancement to  $b$ -jets momentum resolution to be further improved. By inspecting the  $ZH \rightarrow \ell\ell b\bar{b}$  decay (see figure 2.4) it is clear that the momentum of the Higgs and vector

boson candidates ought to be balanced. The momentum resolution of the two charged leptons forming the  $Z$  boson candidate is better than that of the  $b$ -jets (even after corrections). The  $b$ -tagged jets momentum can therefore be corrected with a kinematic fit. The kinematic fit improves resolution by 20–30% compared with the muon-in-jet corrected quantities. This correction is only applied to events with 2 or 3 total jets as the presence of more jets results in a smearing of the effect over those additional jets.

---

### Common Selections

Jets	$\geq 2$ signal jets
$b$ -jets	2 $b$ -tagged signal jets
Leading $b$ -tagged-jet $p_T$	$> 45$ GeV

### 0 Lepton

Trigger	lowest un-prescaled $E_T^{miss}$ triggers
Leptons	0 $VH$ -loose lepton
$E_T^{miss}$	$> 150$ GeV
$S_T$	$> 120$ (2 jets), $> 150$ GeV(3 jets)
$ \min \Delta\phi(E_T^{miss}, \text{jet}) $	$> 20^\circ$ (2 jets) , $> 30^\circ$ (3 jets)
$ \Delta\phi(E_T^{miss}, h) $	$> 120^\circ$
$ \Delta\phi(j_1, j_2) $	$< 140^\circ$
$ \Delta\phi(E_T^{miss}, E_{T,\text{trk}}^{miss}) $	$< 90^\circ$
$p_T^V$ regions	[150, 250] GeV, [250, $\infty$ ] GeV

### 1 Lepton

Trigger	$e$ channel: un-prescaled single electron Tables 6 and 7 of Ref. [110]
Leptons	$\mu$ channel: see 0-lepton triggers
$E_T^{miss}$	1 $WH$ -signal lepton
$p_T^V$ regions	$> 1$ $VH$ -loose lepton veto $> 30$ GeV( $e$ channel) [150, 250] GeV, [250, $\infty$ ] GeV

### 2 Lepton

Trigger	un-prescaled single lepton Tables 6 and 7 of Ref. [110]
Leptons	2 $VH$ -loose leptons ( $\geq 1$ $ZH$ -signal lepton) Same flavor, opposite-charge for $\mu\mu$
$m_{\ell\ell}$	$81 < m_{\ell\ell} < 101$ GeV
$p_T^V$ regions	[75,150], [150, 250], [250, $\infty$ ] GeV

Table 5.5: Summary of the signal event selection in the 0–, 1– and 2–lepton analyses.

# Chapter 6

## Analysis Strategy

With the objects required for this analysis reconstructed and an event selection in place an analysis strategy is formed in order to maximise the signal strength of the  $VH \rightarrow b\bar{b}$  process and yield a robust result that is well understood in terms of modelling and systematic errors. The data the analysis strategy is executed on is  $140 \text{ fb}^{-1}$  of proton-proton collision data recorded by the ATLAS detector at a centre of mass energy of  $\sqrt{s} = 13 \text{ TeV}$  during Run 2 of the LHC (2015–2018). Data and Monte Carlo predictions are grouped into datasets called mc16a, mc16d and mc16e that refer to data taken in the periods 2015–2016, 2017 and 2018 respectively.

First the categorisation into analysis regions will be defined, then the composition of those regions will be explained detailing the simulated samples used, next the multi-variate algorithm that is used to generate some of the distributions entering into the profile-likelihood fit will be explained. Then the fit itself will be explained, including a definition of the likelihood function and systematic uncertainty model. Plots of the data versus Monte-Carlo prediction will then be shown in order to build a picture of the pre-fit status of the agreement. Finally a series of cross-checks which are used to validate the methods of the analysis will be explained.

The author’s contributions include studying the behaviour of the profile-likelihood fit with the inclusion of these regions, which are new in this round of the analysis and making comparisons between the  $80 \text{ fb}^{-1}$  and  $140 \text{ fb}^{-1}$  datasets). Contributions also include the training of the multi-variate classification algorithm with the inclusion

of new input variables with respect to the previous version of the algorithm.

Didn't you also say systematic studies?!

## 6.1 Categorisation Into Analysis Regions

Events which pass the selection detailed in table 5.5 are categorised into several regions for analysis. Firstly they are split into what are known as medium (75–150 GeV), high (150–250 GeV) and extreme ( $>250$  GeV)  $p_T^V$ . Events are further categorised by jet multiplicity, all jets are required by the selection to have at least two jets, categories are defined for events with exactly two, or exactly three jets. In the 2-lepton channel there is a requirement of three or more jets but when referring to all three channels at once categorisation is referred to as simply 3-jet or 2-jet.

Events are categorised into a signal region which is straddled either side by two control regions. This categorisation is achieved via cuts defined in the  $\Delta R(b, \bar{b}) - p_T^V$  plane, they are chosen to maximise the signal purity in the signal region and are shown for the 1-lepton channel in figure 6.1. The regions either side of the signal

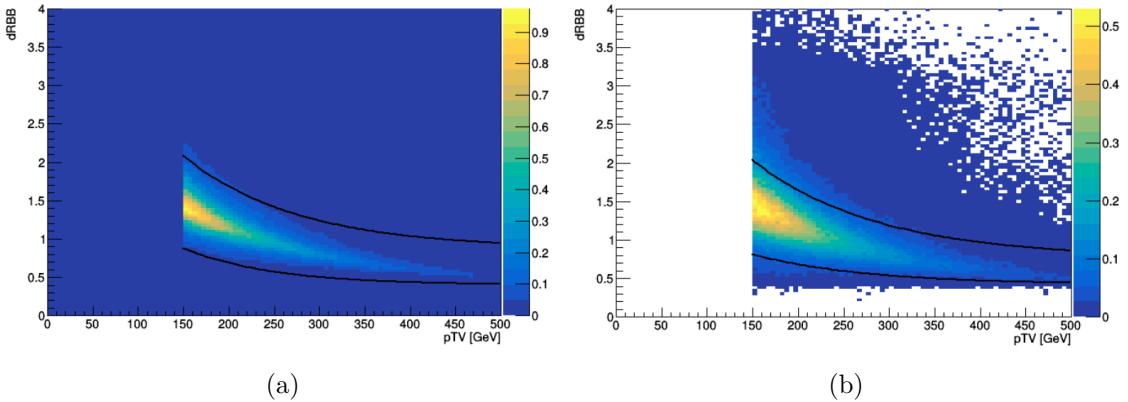


Figure 6.1: Signal distribution of  $\Delta R$  between the two selected jets as function of  $p_T^V$  in the 1-lepton channel are shown in the 2-tag 2-jet (a) and 2-tag 3-jet (b) categories. The black lines demonstrate the upper and lower continuous cuts used to categorise the events into the signal and control regions.

### How has this been optimised?

region are known as the high- $\Delta R(b, \bar{b})$  and low- $\Delta R(b, \bar{b})$  control regions, shortened to CR<sub>high</sub> and CR<sub>low</sub> respectively.

### 6.1.1 Top $e\mu$ Control Region

One more region exists only in the 2-lepton channel, it is obtained by requiring two opposite flavour leptons instead of two same flavour leptons, and keeping all other selection criteria the same. This region consists almost entirely of  $t\bar{t}$  and single top processes (whose Feynman diagrams belong to the same sum) matching very closely the number of expected events of each of these backgrounds in the 2-lepton channel. This region is therefore called the top  $e\mu$  control region. The data from this control region can be used as a prediction for the number of top process events in the 2-lepton channel once multiplied by a scale factor which accounts for differences in normalisation. Given that this region exists purely to model systematic uncertainties it will be explored in more detail in the subsequent chapter in section 7.4.1.

**Make a requirement of the flavour explicit!**

## 6.2 Composition of Analysis Regions

This section will detail the composition of each analysis region in terms of background and signal processes. For all regions the signal process is  $VH \rightarrow b\bar{b}$ , the prediction for which comes from events generated using POWHEG MiNLO [111] + PYTHIA 8 [112] for quark initiated processes and POWHEG + PYTHIA 8 for gluon initiated processes as can be seen in table 6.1.

Process	Generator	$\sigma \times BR$ [pb]	N <sub>events</sub> in millions		
			mc16a	mc16d	mc16e
$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG MiNLO + PYTHIA 8 (NNPDF3) [113]	$153.05 \times 0.582$	2	2	3.3
$qq \rightarrow WH \rightarrow l^+l^- b\bar{b}$	POWHEG MiNLO + PYTHIA 8 (NNPDF3)	$282.78 \times 0.582$	4	4	6.6
$qq \rightarrow WH \rightarrow l^-\nu b\bar{b}$	POWHEG MiNLO + PYTHIA 8 (NNPDF3)	$179.49 \times 0.582$	2	2	3.3
$qq \rightarrow ZH \rightarrow ll b\bar{b}$	POWHEG MiNLO + PYTHIA 8 (NNPDF3)	$77.04 \times 0.582$	3	3	5
$gg \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + PYTHIA 8 (NNPDF3)	$24.57 \times 0.582$	0.5	0.5	0.5
$gg \rightarrow ZH \rightarrow l^-l^+ b\bar{b}$	POWHEG + PYTHIA 8 (NNPDF3)	$12.42 \times 0.582$	0.75	0.75	0.75

Table 6.1: Details of Monte Carlo simulated predictions of the signal processes of the analysis. Here  $\ell$  is inclusive of  $e$ ,  $\mu$  and  $\tau$  leptons.

The 0-lepton channel contains the  $Z$ +jets,  $W$ +jets, top quark and diboson backgrounds. The  $Z$ +jets background dominates the mixture in the 2-jet category across signal and control regions. In the 3-jet category the top-quark processes dominate apart from in  $CR_{low}$ . There is very little signal contamination in the control regions.

As can be seen in table 6.2,  $V$ +jets events are generated using SHERPA 2.2.1, top quark events are generated using POWHEG + PYTHIA 8 and diboson events are generated also using SHERPA 2.2.1, this is true for these backgrounds across all channels where a Monte-Carlo prediction is used.

Process	Generator	$\sigma \times BR$ [pb]	$N_{\text{events}}$ in millions			
			mc16a	mc16d	mc16e	
<b>Vector boson + jets</b>						
$Z \rightarrow \nu\nu$	SHERPA 2.2.1	$56280 \times 0.200$	150	160	140	
$W \rightarrow \ell\nu$	SHERPA 2.2.1	$183600 \times 0.325$	340	400	540	
$Z/\gamma^* \rightarrow \ell\ell$	SHERPA 2.2.1	$61940 \times 0.101$	120	160	210	
<b>Top-quark</b>						
$t\bar{t}$	non-full-had (plus ET/pTW extensions)	POWHEG + PYTHIA 8	$831.76 \times 0.543$	120(55)	150(55)	200(61)
	di-leptonic	POWHEG + PYTHIA 8	$831.76 \times 0.105$	N/A	45	100
Single-top	$s$ - channel (leptonic-top)	POWHEG + PYTHIA 8	$10.32 \times 0.325$	4	5	7
	$t$ - channel (leptonic-top)	POWHEG + PYTHIA 8	$216.96 \times 0.325$	10	12	17
	$Wt$ - channel (plus di-lepton extension)	POWHEG + PYTHIA 8	$71.7 \times 1$	20	24(24)	33(33)
<b>Diboson</b>						
$qq \rightarrow WW$	$\rightarrow qqlv$	SHERPA 2.2.1	$112.6 \times 0.439$	14	50	24
$qq \rightarrow WZ$	$\rightarrow lvqq$ (with $Z \rightarrow b\bar{b}$ extension)	SHERPA 2.2.1	$50.3 \times 0.227$	7(6)	36(7)	12(10)
	$\rightarrow qqvv$	SHERPA 2.2.1	$50.3 \times 0.135$	6	6	10
	$\rightarrow qql\ell$	SHERPA 2.2.1	$50.3 \times 0.0683$	5	27	9
$qq \rightarrow ZZ$	$\rightarrow qql\ell$ (with $Z \rightarrow b\bar{b}$ extension)	SHERPA 2.2.1	$15.57 \times 0.140$	5(5)	5(6)	9(4)
	$\rightarrow qqvv$ (with $Z \rightarrow b\bar{b}$ extension)	SHERPA 2.2.1	$15.57 \times 0.280$	5(5)	5(6)	9(8)
$gg \rightarrow WW$	$\rightarrow qqlv$	SHERPA 2.2.2	$4.8 \times 0.439$	0.8	0.9	1.1
$gg \rightarrow ZZ$	$\rightarrow qql\ell$ or $qqvv$	SHERPA 2.2.2	$1.57 \times 0.420$	4	6	8

Table 6.2: Details of Monte Carlo simulated predictions of the background processes of the analysis. Here  $\ell$  is inclusive of  $e$ ,  $\mu$  and  $\tau$  leptons. For  $Z/\gamma^* \rightarrow \ell\ell$  events the requirement of  $m_{\ell\ell} > 40$  GeV was imposed.

The 1-lepton channel contains the  $W$ +jets,  $Z$ +jets, top quark, diboson and multijet backgrounds where multijet is the name given to backgrounds with jets present that do not fit into any other categorisation. The channel is dominated by a mixture of  $W$ +jets and top quark processes, CR<sub>high</sub> has a higher purity of top quark processes whereas CR<sub>low</sub> has a high purity of  $W$ +jets background. Contribution from multijet and  $Z$ +jets is small across all regions.

The 2-lepton channel contains  $Z$ +jets, top quark and diboson backgrounds. The  $Z$ +jets background dominates across all regions particularly in both  $\Delta R(b, \bar{b})$  control regions. Predictions for the top quark processes are taken from the top  $e\mu$  control region described in section 6.1.1. In all channels the diboson background is contained almost entirely within the signal region.

### 6.3 Multi-variate Event Classification

The signal regions in all channels enter into a profile-likelihood fit as distributions generated by a multi-variate analysis. The multi-variate algorithm used to generate this distribution is a BDT trained to separate  $VH \rightarrow b\bar{b}$  from all other events. Only the signal region enters into the fit as a distribution of BDT scores, despite this, the BDT itself is trained on the combination of all  $\Delta R(b, \bar{b})$  regions and takes place before the categorisation into  $p_T^V$  bins. All other analysis selection criteria defined in section 5.7 are applied before training. Training is carried out separately for each lepton channel and jet multiplicity, a single split at 150 GeV is used in the 2-lepton channel resulting in eight separate regions for training as in table 6.3. The impact of the difference between the fit regions and those that go into training the BDT on

**150GeV of what?**  
**why?**  
**as shown**

**The mitigation outweighs any change in performance????**

its performance have been studied [114], and in general the mitigation of over-fitting resulting from having more events to train on outweighs any change in performance.

In tests it has been shown that there is significant evidence of over-training when

Channel	75 GeV < $p_T^V$ < 150 GeV		$p_T^V > 250$ GeV	
	2-jets	3-jets	2-jets	3-jets
0-lepton	—	—	✓	✓
1-lepton	—	—	✓	✓
2-lepton	✓	✓	✓	✓

Table 6.3: Event categories on which the analysis BDT is trained, each box with a ✓ in represents a separate training.

training is performed on the signal region alone.

The BDT is trained on all of the samples listed in tables 6.1 and 6.2, this includes the  $t\bar{t}$  and single top Monte-Carlo predictions that are not used in the final fit in the 2-lepton channel as previously discussed. The inputs are split into two datasets based on whether or not the event number is even or odd. The model trained on the odd numbered events is evaluated on the even numbered events and vice versa, this ensures that training and evaluation take place on statistically independent datasets but also that as many events are trained on as possible. The final discriminant is

constructed by summing the results from both trainings.

The BDT inputs differ from the final analysis distributions further in that the  $b$ -tagging which is applied is not the hybrid tagging discussed in section 5.4.5 but rather just the truth tagging described in section 5.4.4. This gives the highest number of events possible given all of the tagging strategies available and thus gives the most statistically robust training. The differences between hybrid-tagged and truth-tagged distributions is small when considering only the sum of all backgrounds and thus the impact of this decision is small.

It is necessary to deal with distributions that have very long tails so that cuts are not placed by the BDT in those tails. This is achieved by placing an artificial limit on the maximum value of each input distribution that corresponds to leaving 99 % of signal events in the remaining distribution. This increases the reproducibility of the training as fewer cuts are wasted by being placed in long tails.

## What is the effect of this?

## How are data with tails classified?

It seems tails are moved into the last bin.

(Some) variables need explanation

and that was used?

Table 6.4 shows the choices of variables used as inputs to the algorithm in each analysis channel. Inputs are carefully chosen in order to maximise the performance of the algorithm. Table 6.5 shows the choice of hyper-parameters for the algorithm as described in terms of TMVA [115] the toolkit for multi-variate analysis which is built into ROOT [116], and was used for training this algorithm.

Distributions of the  $m(b, \bar{b})$  and  $\Delta R(b, \bar{b})$  observables in the 2-jet, high  $p_T^V$  region are shown in figure 6.2. The other variables are included in figures, A.1, A.2 and A.3 in the appendix for the 0–, 1– and 2–lepton channels respectively. In all channels it can be

no line break

seen that a high level of separation power in one dimension can be obtained from the  $m(b, \bar{b})$  and  $\Delta R(b, \bar{b})$  distributions which are correlated. In the 0–lepton channel a moderate level of separation can be obtained from the  $\Delta\eta(b, \bar{b})$ ,  $E_T^{\text{miss}}$  and  $p_T(b_2)$  also. In the 1–lepton channel  $\min(\Delta\phi(\ell, \text{jet}))$ ,  $m_{\text{top}}$  and  $p_T(b_2)$  provide the next best one dimensional separation. In the 2–lepton channel it is the  $\cos\theta(\ell^-, Z)$ ,  $E_T^{\text{miss}}$ –significance,  $\Delta\eta(V, H)$  and  $p_T(b_2)$  that provide next best separation.

The BDT is designed to maximise the separation between signal and background Monte-Carlo predictions. This does not take into account the statistical error on

How is the 2-jet  
150 GeV split  
treated here?  
Why the split?

Variable	Name	0-lepton	1-lepton	2-lepton
$m_{jj}$	mBB	✓	✓	✓
$\Delta R(jet_1, jet_2)$	dRBB	✓	✓	✓
$p_T^{jet1}$	pTB1	✓	✓	✓
$p_T^{jet2}$	pTB2	✓	✓	✓
$p_T^V$	pTV	✓	✓	✓
$\Delta\phi(V, H)$	dPhiVBB	✓	✓	✓
binned MV2c10(jet <sub>1</sub> )	bin_MV2c10B1	✓	✓	
binned MV2c10(jet <sub>2</sub> )	bin_MV2c10B2	✓	✓	
$ \Delta\eta(jet_1, jet_2) $	dEtaBB	✓		
$M_{eff}$	MEff	✓		
track based soft MET term	softMET	✓		
MET	MET	$\equiv p_T^V$	✓	
$\min(\Delta\phi(\ell, jet))$	dPhiLBmin		✓	
mTW	mTW		✓	
$\Delta Y(W, H)$	dYWH		✓	
$m_{top}$	mTop		✓	
MET significance	METSig			✓
$\Delta\eta(V, H)$	dEtaVBB			✓
$m_{\ell\ell}$	mLL			✓
$\cos\theta(\ell^-, Z)$	cosThetaLep			✓
<b>Only in 3 Jet Events</b>				
$p_T^{jet3}$	pTJ3	✓	✓	✓
$m_{jjj}$	mBBJ	✓	✓	✓

Table 6.4: Variables used to train the multi-variate discriminant.

tmva Setting	Value	Definition
BoostType	gradient boosting	Boost procedure
Shrinkage	0.5	Learning rate
SeparationType	Gini index	Node separation gain
PruneMethod	No Pruning	Pruning method
NTrees	200 (600 for 1-lepton VH)	Number of trees
MaxDepth	4 (2 for 1-lepton diboson)	Maximum tree depth
nCuts	100	Number of equally spaced cuts tested per variable per node
nEventsMin	5%	Minimum number of events in a node (% of total events)

Table 6.5: Hyperparameters used for the BDT trainings. Exceptions for the 1-lepton VH and diboson trainings are given in brackets.

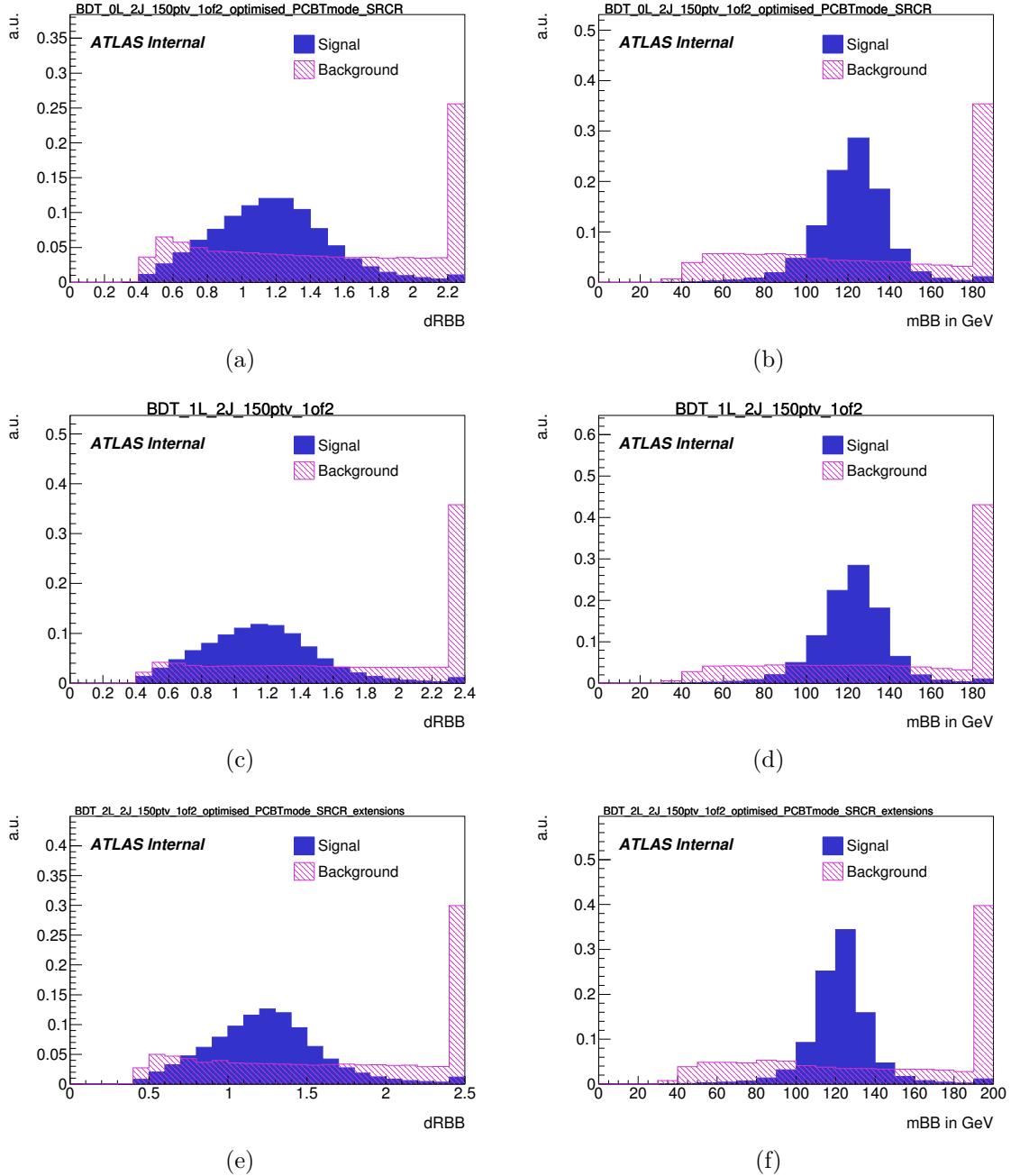


Figure 6.2: BDT input distributions are shown for the  $m(b, \bar{b})$  and  $\Delta R(b, \bar{b})$  observables in the 0–, 1– and 2–lepton channels for (a) & (b), (c) & (d) and (e) & (f) sub-figures respectively.

add a sentence that this  
is a rebinning to ensure  
that the stat. unc. is  
below 20% per bin.

the quantities of each bin in the discriminant distribution. To mitigate this issue a transformation is applied, details of which are found in appendix D. The performance of the algorithm in terms of over-fitting is found to be robust [114] and the separation between categories can be seen in section 6.5.

## 6.4 Profile Likelihood Fit

Ultimately the results of the analysis come from a profile-likelihood fit. This fit uses a binned likelihood function whose maximum corresponds to the best agreement between data and the prediction. The likelihood is defined as the product over all bins of the Poisson probability to observe  $N_i^{\text{obs}}$  data events given a prediction of  $N_i^{\text{pred}}(\mu, \mathbf{k}, \boldsymbol{\theta})$  events in a given bin  $i$ ,

$$\mathcal{L}(\mu, \mathbf{k}, \boldsymbol{\theta}) = \prod_{i \in \text{bins}} \frac{\left(N_i^{\text{pred}}(\mu, \mathbf{k}, \boldsymbol{\theta})\right)^{N_i^{\text{data}}}}{N_i^{\text{data}}!} \cdot e^{-N_i^{\text{pred}}(\mu, \mathbf{k}, \boldsymbol{\theta})}. \quad (6.1)$$

The number of predicted events  $N_i^{\text{pred}}$  is dependent on three sets of parameters, the signal strength  $\mu$ , the scale factors  $\mathbf{k} = \{k_1, \dots, k_j\}$ , and the nuisance parameters  $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_l\}$ , as follows

$$N_i^{\text{pred}}(\mu, \mathbf{k}, \boldsymbol{\theta}) = \mu \cdot N_{i,\text{sig}}^{\text{pred}}(\boldsymbol{\theta}) + \sum_{b \in \text{bkg}} k_b \cdot N_{i,b}^{\text{pred}}(\boldsymbol{\theta}). \quad (6.2)$$

either?

The ratio between the measured and the expected signal cross-sections  $\mu = \sigma/\sigma_{\text{SM}}$ , known as the signal strength, is the parameter of interest in the nominal fit. It is common to all analysis regions that enter into the fit. Each of the scale factors  $k_j$  and the signal strength scale either their associated quantity linearly without any prior constraint or penalty in the likelihood function, in further sections these parameters will be referred to as floating. Each nuisance parameter  $\theta_i$  encodes the dependence of the prediction on systematic uncertainties into continuous parameters in the likelihood. Prior knowledge of the uncertainty that these parameters encode for is expressed as a Gaussian penalty term  $\mathcal{G}(0 | \theta_i, 1)$  added to the likelihood for

each uncertainty. The parameters  $\theta_i$  are therefore expressed as a number of standard deviations of a unit Gaussian in the subsequent sections. These penalties results in a log-normal, or normal dependence of the predicted yields or shapes on the displayed parameter values.

The nominal fit result in terms of  $\mu$  and  $\sigma_\mu$  is obtained by maximizing the likelihood function with respect to all parameters. This is referred to as the maximized log-likelihood value, MLL. The profile likelihood ratio test statistic,  $q_\mu$ , is then constructed as follows:

$$q_\mu = -2 \ln \left[ \mathcal{L}(\mu, \hat{\mathbf{k}}, \hat{\boldsymbol{\theta}}_\mu) / \mathcal{L}(\hat{\mu}, \hat{\mathbf{k}}, \hat{\boldsymbol{\theta}}) \right] \quad (6.3)$$

where  $\hat{\mu}$  and  $\hat{\boldsymbol{\theta}}$  are the parameters that maximise the likelihood (with the constraint  $0 \leq \hat{\mu} \leq \mu$ ), and  $\hat{\boldsymbol{\theta}}_\mu$  are the nuisance parameter values that maximise the likelihood for a given  $\mu$ . This test statistic is used to measure the compatibility of the background only model with the observed data.

#### 6.4.1 STXS Measurements

As well as the single parameter of interest mode described above the fit can also be run in a multiple parameter of interest paradigm. The purpose of this fit is to measure the cross-section of the  $WH$  and  $ZH$  processes in a number of well defined regions of phase space. In order to achieve this a simplified template cross-section (STXS) scheme is used. As in the previous measurement [117] the STXS bins are split into many categories before being merged down to include  $p_T^V$  bins of 75–150 GeV, 150–250 GeV and  $>250$  GeV. The 75–150 GeV bin does not include a measurement of  $WH$  as there is no 75–150 GeV bin available in the 1-lepton channel and so there ought not to be any signal events present in this region of phase space.

## 6.5 Pre-fit Data Versus Prediction

This section shows the pre-fit distributions of the Monte-Carlo prediction versus the data in every analysis region that enters into the profile-likelihood fit. Figures 6.3, 6.4 and 6.5 show the distributions in the signal regions of the 0–, 1– and 2-lepton channels respectively. Plots of all distributions in all regions can be found in appendix A.2.

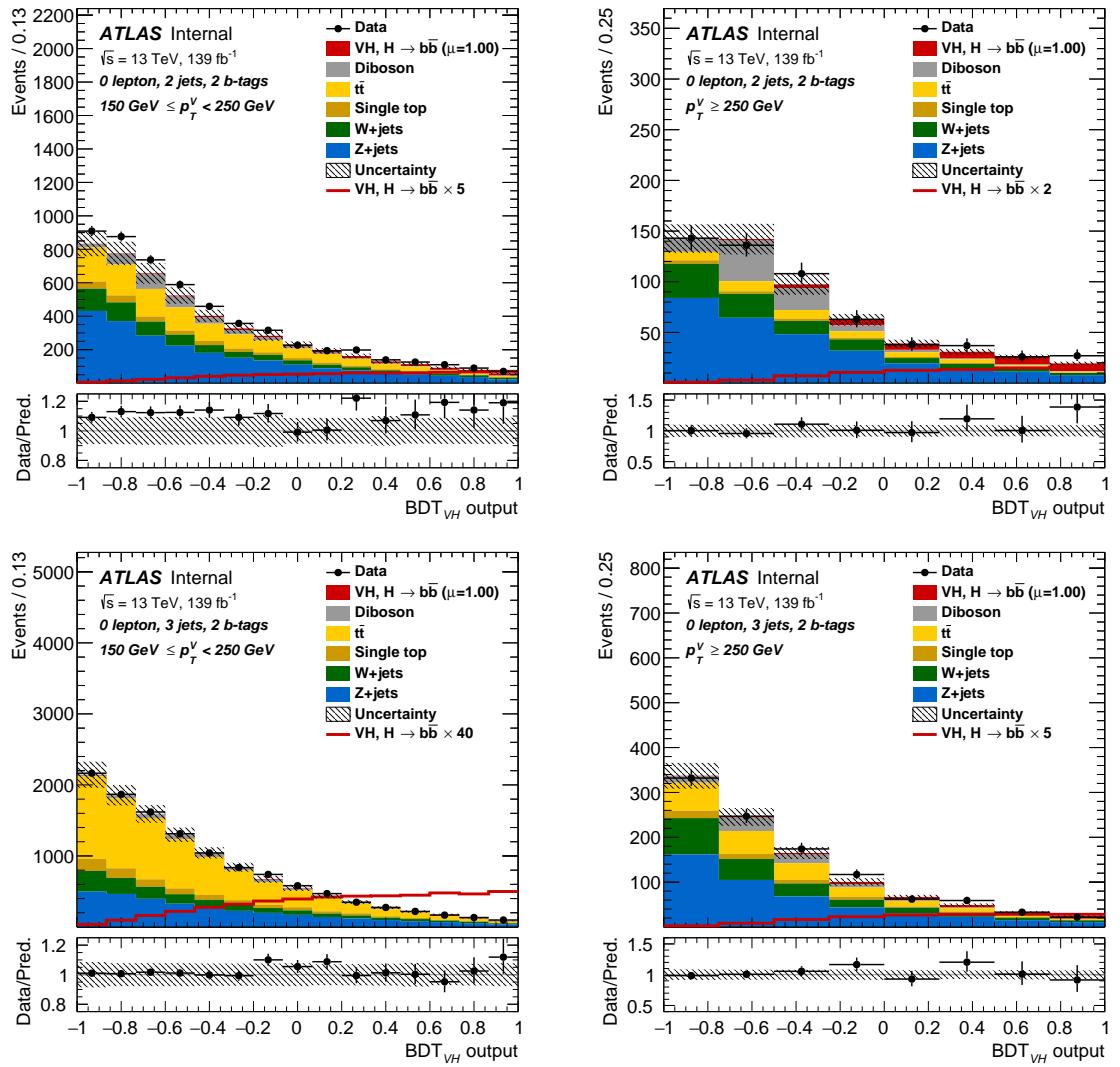


Figure 6.3: Pre-fit distributions in the signal region of the 0–lepton channel.

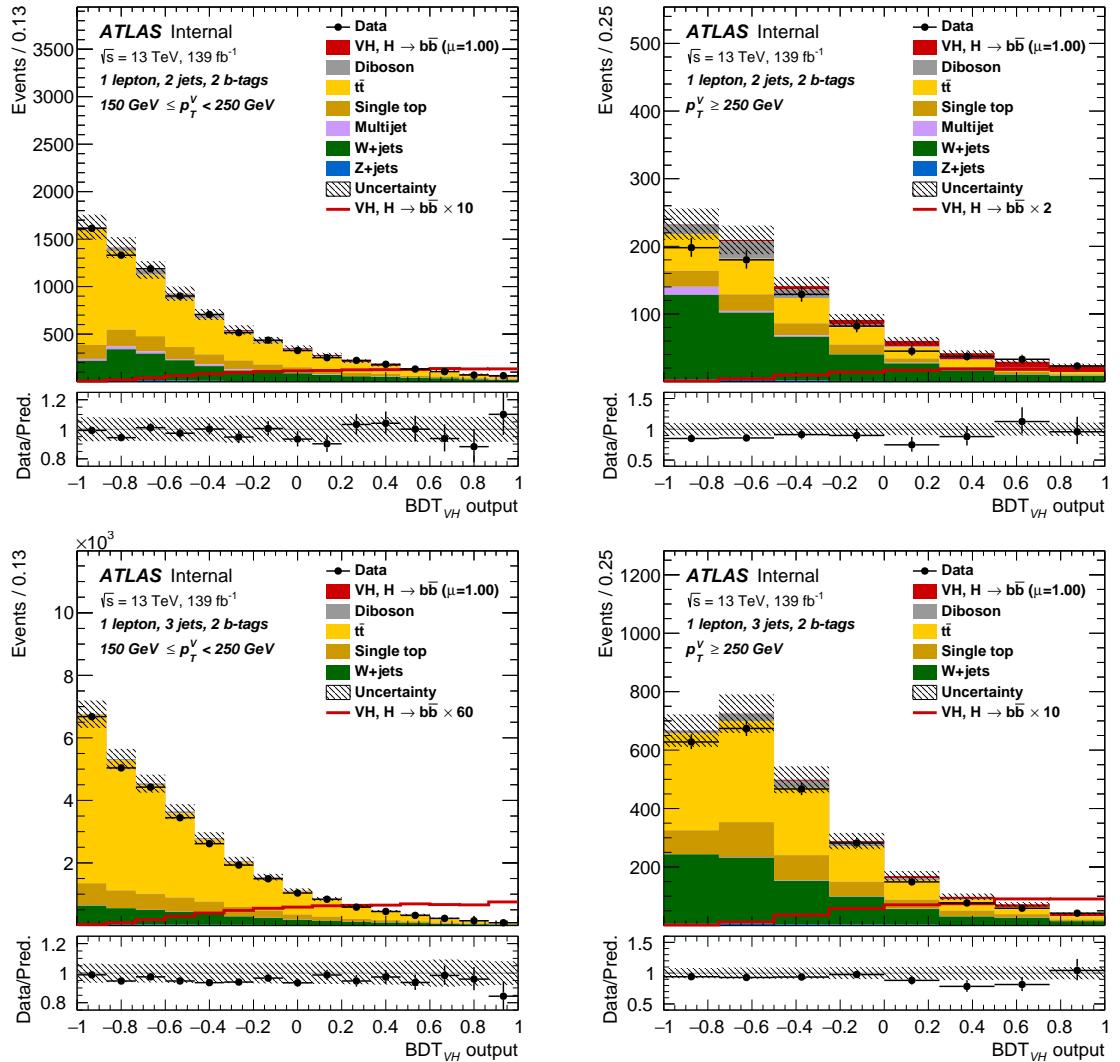


Figure 6.4: Pre-fit distributions in the signal region of the 1-lepton channel.

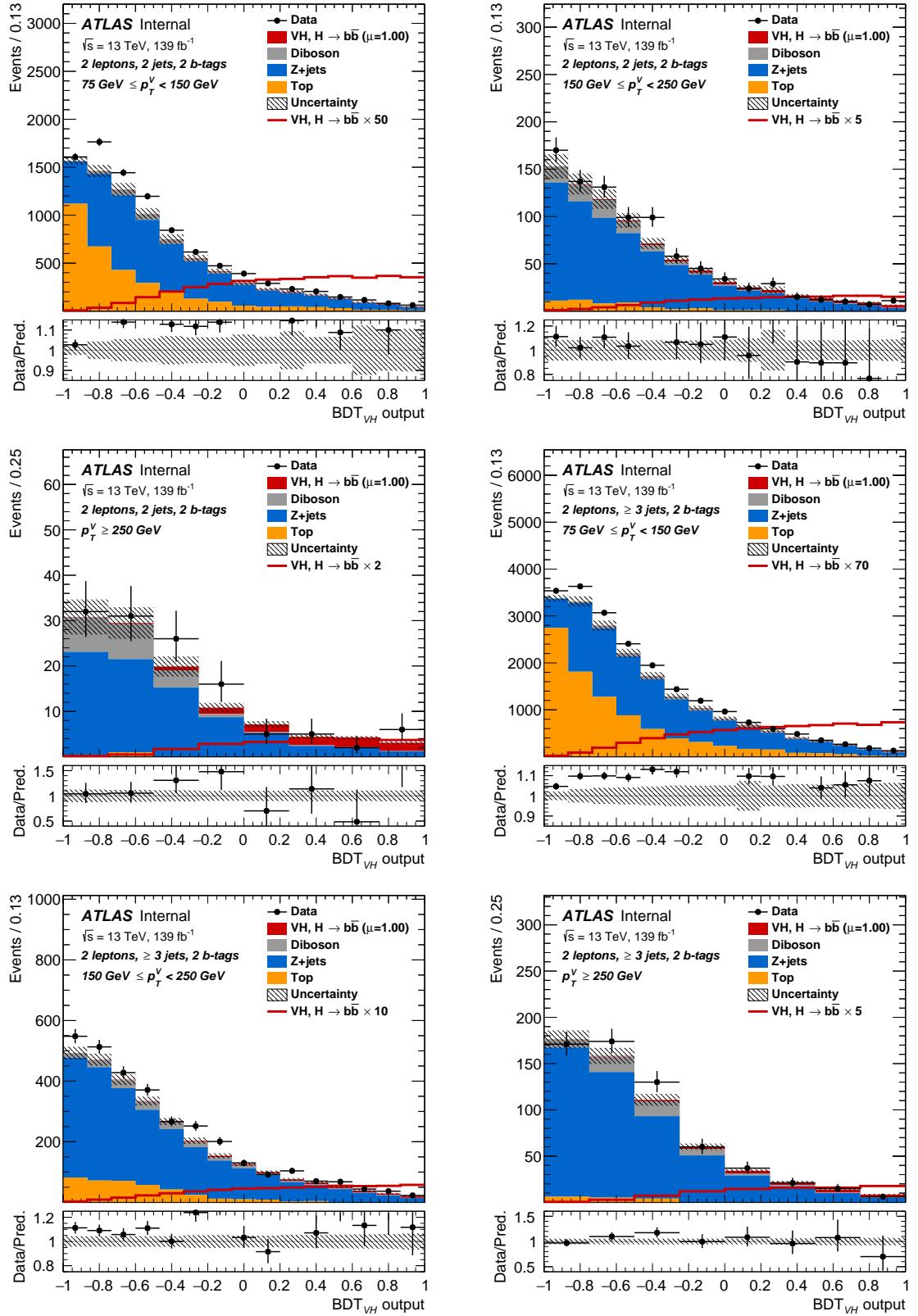


Figure 6.5: Pre-fit distributions in the signal region of the 2-lepton channel.

## 6.6 Analysis Cross-checks

The final elements of the analysis strategy are a series of cross-checks that are designed to ensure the methodology is robust. Firstly there is the di-jet mass fit, also known as the  $m_{bb}$  fit. This cross-check is designed to ensure that the multi-variate analysis has not introduced any biases that have changed the result so much that it is statistically incompatible with a version of the analysis that does not use the BDT. This cross-check is performed by simply taking the  $m_{bb}$  distribution in place of the BDT distribution in the profile-likelihood fit.

The second cross-check is a measurement of the diboson process. Diboson final states arising from proton-proton collisions are well understood and in this case are being treated as a standard candle<sup>1</sup>. The rationale here is that if the analysis methodology produces a measurement of the diboson process that is in agreement with the Standard Model prediction, and therefore existing measurements, then the methodology itself has not introduced unexpected effects on the result. Results of these cross-checks imply both of these tests have been passed [114].

---

<sup>1</sup>A standard candle is an astronomical object with a known absolute luminosity that can be used to aid astronomical measurements.

# Chapter 7

## Systematic Uncertainties

Systematic uncertainties considered in the analysis come in one of two forms, either a shape effect or a normalisation effect. Normalisation effects alter the number of events in a given sample across the entire sample simply changing the total number of events. Shape effects change where events lie in a given distribution causing events to migrate between bins of a histogram, and potentially across boundaries that are used to define analysis regions described in 6.1. Each systematic uncertainty is controlled by a nuisance parameter in the profile-likelihood fit. For all Gaussian penalty terms the a prior uncertainty is estimated in advance.

A sub-category of normalisation effect is the acceptance effect which deals with the normalisation in a particular region or set of regions. The purpose of the acceptance uncertainties is to account for any mismodelling in the theoretical prediction of the quantities used to categorise events into regions be it the leptonic channel, jet multiplicity or an analysis region. The priors of these so called acceptance uncertainties are calculated using the double ratio, unless stated otherwise,

### Example for A/B nominal / alternative

$$\frac{N_A^{\text{nominal}}}{N_B^{\text{nominal}}} \Big/ \frac{N_A^{\text{alternative}}}{N_B^{\text{alternative}}} , \quad (7.1)$$

where the formula has been written agnostic of the specific application so  $N_A^{\text{nominal}}$  is the number of events in the nominal prediction in some region or category called  $A$  and so on.

Another variant of the normalisation effect are the flavour composition uncertainties. Rather than impacting the normalisation of an entire background in terms of how they are categorised in the analysis plots, for example in section 6.5, they impact the sub-processes split by flavour of the decay products individually. A heavy flavour process is defined as one where the two leading jets have flavour ( $b, b$ ), ( $b, c$ ),  
full stop ( $b, l$ ), or ( $c, c$ ), this categorisation is often written as hf for short for example when  
sub-processes considering all heavy flavour sub-process of the  $V+jets$  background one could write  $V+hf$ . Flavour composition uncertainties affect the normalisation of one of the flavour sub-processes with respect to one of the others, specific details will follow under the sections relating to the relevant backgrounds.

The rest of this chapter will detail the sources of systematic error broken down into groups of similar origin. Before breaking down each group of systematic uncertainties, a technique called the Boosted Decision Tree Re-weighting (BDTr) method

will be explained as it used across a number of the different groups. The author's contributions include the determination of all of the  $Z+jets$  systematic uncertainties, determination of systematic uncertainties relating to the flavour composition of simulated top process events, and development and testing of the BDTr method.

## 7.1 Parametrising Variance Due To Shape Uncertainties

The predictions entering into the profile-likelihood fit of the analysis can be written as a probability density function (PDF)  $p(\vec{x}|\vec{\theta})$ , where  $\vec{x} = [x_1, \dots, x_i]$  is a vector of observable quantities, with  $N$  elements and  $\vec{\theta}$  represents the theoretical parameters of the model. The parameters may come from the Standard Model theory or phenomenological considerations that must be taken into account to turn that theory into a usable prediction.

As described above, shape effects may impact the prediction of a given Monte-Carlo generator that has a particular set of parameters. The choice of generator

will change the prediction due to different choices made by the generator creators, including but not limited to, the parton shower model, hadronisation model and non-perturbative processes. One way to get a handle on the variance of a particular generator is therefore to vary these choices by picking an alternative generator and drawing a comparison to the alternative prediction. This method is flawed in ways which will be discussed but is nonetheless fairly common and often one of the few choices available to analysers when they are trying to get an idea of what the systematic uncertainties on the modelling of complex physics processes are. Two such methods of comparison will be detailed here, the single dimension parameterisation described in section 7.1.1 and a multi-dimensional parameterisation described in section 7.1.2. Following this a hybrid approach which uses both methods will be described in section 7.1.3.

### 7.1.1 Single Dimension Parameterisation

The single dimension parameterisation uses the ratio of probability densities

$$r(\vec{x}) = \frac{p(\vec{x}|\vec{\theta}_1)}{p(\vec{x}|\vec{\theta}_0)}, \quad (7.2)$$

where the subscripts on  $\vec{\theta}$  number different choices of the parameters which govern the model, for example due to different choice of generator. Here the subscript 0 denotes the nominal prediction and 1 denotes an alternative. In order to map the nominal to the alternative it is clear that ~~the~~ one can multiply  $p(\vec{x}|\vec{\theta}_0)$  by the ratio.

The only way that this calculation is tractable is to consider a single dimension of the probability density function, specifically a single observable chosen from the vector  $\vec{x}$ . The ratio then becomes

$$r(x_i) = \frac{p(x_i|\vec{\theta}_1)}{p(x_i|\vec{\theta}_0)}, \quad (7.3)$$

which can in turn be calculated for as many or as few as the elements of  $\vec{x}$  as is desired. When the ratio is calculated using only a single variable it can still be

used to map the  $N$  dimensional nominal prediction to the alternative, however the mapping is only guaranteed to be successful in the observable  $x_i$  chosen to calculate the ratio. In general agreement between the two predictions in the other variables can not be relied upon.

In practice the ratio is used to derive a systematic shape uncertainty by weight-ing each event in the nominal prediction as a function of the variable  $x_i$  whose distribution was used to calculate the ratio. Limitations on the sample size of each prediction mean that in practice it is better to smooth the ratio via the use of a parametric fit in order to mitigate any large statistical fluctuations in a single bin of the ratio.

### 7.1.2 $N$ -dimensional Parametrisation

As already alluded to in the previous section there are issues with the parametrisa-tion in a single dimension. Not only is the mapping only guaranteed to work in a single dimension but if the technique is applied sequentially on one variable after another the mapping from the first step can be undone by the second. That is to say that if one calculates the ratio in equation 7.3 with  $i = 1$  and  $i = 2$  the re-weighting by  $r(x_2)$  will not preserve the agreement between the two predictions that would be achieved by the re-weighting with  $r(x_1)$ . It is clear that this is because a ratio for a given  $x_i$  only encode information relating to the PDFs  $p(x_i|\theta_0)$  and  $p(x_i|\theta_1)$  whereas an encoding of the multi-variate distributions  $p(\vec{x}|\theta_0)$  and  $p(\vec{x}|\theta_1)$  is desired.

A function capable of providing a univariate representation of the full multi-variate distribution has already been discussed, these are the functions that one obtains when training a machine learning classifier such as those mentioned in chapter 4. Functions of the form given in equation 4.1 can encode information from the multi-variate input  $\vec{x}$  into a lower dimensional  $\vec{y}$  that captures the correlations between the different elements of the input  $\vec{x}$ . Specifically for this parametrisation the model is trained to classify events as coming from either the Monte-Carlo gen-erator with parameters  $\theta_0$  or  $\theta_1$ , the output is therefore written as  $\vec{y} = [y_0, y_1]$  where

each term in the vector represents the probability that an event comes from  $p(\vec{x}|\theta_0)$  or  $p(\vec{x}|\theta_1)$  respectively. Note that the choice of output function in the model must ensure that probability is conserved, i.e.  $y_0 + y_1 = 1$ . This technique has been demonstrated in the literature [118] where the mathematical details are discussed in more detail.

Our setup allows us to create the following approximation

$$r(\vec{x}) = \frac{p(\vec{x}|\vec{\theta}_1)}{p(\vec{x}|\vec{\theta}_0)} \approx \frac{F(\vec{x}, \vec{w})[1]}{F(\vec{x}, \vec{w})[0]}, \quad (7.4)$$

where  $F$  is our trained model whose hyper-parameters we can consider fixed and drop from the notation (previously denoted  $\vec{\theta}$  in chapter 4). The numbers in the square brackets indicate which element of the  $\vec{y}$  is being selected. Once the above approximation is made, the steps to generate weights and perform a parameterisation of the shape discrepancy between two predictions is the same as in the one dimensional case. By placing certain restrictions on  $F$  one can turn the approximation into an equivalence, this is discussed in [119], however for what is considered here the approximation will be the focus. As mentioned this parametrisation will be referred to in the analysis as the BDTr method.

### 7.1.3 Hybrid ( $N - 1$ )-Dimensional Parametrisation

It is possible to combine the two aforementioned parametrisation strategies into a single strategy. This is achieved by first re-weighting using the 1-dimensional technique and then training the classifier on the re-weighted distributions and proceeding with the  $N$ -dimensional parametrisation as usual. In principle the 1-dimensional re-weighting can be performed a number of times sequentially before training the classifier and so therefore a general  $(N - k)$ -dimensional strategy can be formed, however in this analysis the 1-dimensional re-weighting is only applied once before training.

The hybrid approach yields two parametrisations one parametrising the difference between the two samples in the single variable used in the 1-dimensional re-

weighting and a second that parametrises the remaining variables using the density ratio formed by using the classifier trained on the re-weighted distributions. This approach will be referred to as the factorised BDTr method as the single variable that is not included in the classifier that is used to generate the multi-dimensional parametrisation is considered to be factored out.

There are a number of reasons why one might want to factor out a single variable from the multi-dimensional procedure that naively looks superior in every way to the 1-dimensional approach. Internal parameters of two predictions models need not obey a one-to-one mapping, therefore a smooth interpolation between them is not guaranteed to exist. This reveals an incongruity between the nature of the Gaussian penalty term used to control the shape uncertainty in the fit, which is varied smoothly and continuously, and the underlying mapping that it aims to characterise. A second nuisance parameter arising from the factorised variable therefore at least allows the profile-likelihood fit to control in a more well defined way the histograms relating to that variable. The inclusion of a factorised variable may also be desirable as particular variables have increased importance to the analysis, for example the histograms entering in the control regions are of  $p_T^V$  and so giving the fit a parameter to control directly the shape of this variable is considered to aid interpretability.

## Example with plots? Fig 7.3

### 7.2 Experimental Systematic Uncertainties

full stop or “which”

This section describes experimental systematic uncertainties, these arise due to limitations of the hardware discussed in chapter 3 and reconstruction algorithms described in chapter 5. All of these uncertainties are provided by different combined performance (CP) groups and made centrally available to members of the ATLAS collaboration, this ensures a consistent understanding of the performance of the detector and the reconstruction algorithms of Athena. A summary of all of the experimental systematic uncertainties used in the analysis can be found in table 7.1.

Table 7.1: A summary of the experimental systematic uncertainties considered in the analysis. They are listed by the name of the nuisance parameter entering into the profile-likelihood fit and a short description is provided of each uncertainty.

Systematic uncertainty	Short description
<b>Event</b>	
Luminosity	uncertainty on total integrated luminosity
Pileup Re-weighting	uncertainty on pileup re-weighting
<b>Triggers</b>	
EL_EFF_Trigger_Total_1NPCOR_PLUS_UNCOR	electron trigger efficiency uncertainty
MUON_EFF_TrigStatUncertainty	muon trigger efficiency uncertainty
MUON_EFF_TrigSystUncertainty	as above
METTrigStat	$E_T^{\text{miss}}$ trigger efficiency uncertainty
METTrigTop/Z	as above
METTrigSumpt	as above
<b>Electrons</b>	
EL_EFF_Reco_Total_1NPCOR_PLUS_UNCOR	reconstruction efficiency uncertainty
EL_EFF_ID_Total_1NPCOR_PLUS_UNCOR	ID efficiency uncertainty
EL_EFF_Iso_Total_1NPCOR_PLUS_UNCOR	isolation efficiency uncertainty
EG_SCALE_ALL	energy scale uncertainty
EG_RESOLUTION_ALL	energy resolution uncertainty
<b>Muons</b>	
MUON_EFF_RECO_STAT	reconstruction and ID efficiency uncertainty for muons with $p_T > 15 \text{ GeV}$
MUON_EFF_RECO_SYS	as above
MUON_EFF_RECO_STAT_LOWPT	reconstruction and ID efficiency uncertainty for muons with $p_T < 15 \text{ GeV}$
MUON_EFF_RECO_SYS_LOWPT	reconstruction and ID efficiency uncertainty for muons with $p_T < 15 \text{ GeV}$
MUON_EFF_TTVA_STAT	track-to-vertex association efficiency uncertainty
MUON_EFF_TTVA_SYS	as above
MUON_ISO_STAT	isolation efficiency uncertainty
MUON_ISO_SYS	as above
MUON_ID	momentum resolution uncertainty from inner detector
MUON_MS	momentum resolution uncertainty from muon system
MUON_SCALE	momentum scale uncertainty
MUON_SAGITTA_RHO	charge dependent momentum scale uncertainty
MUON_SAGITTA_RESBIAS	as above
<b>Jets and <math>b</math>-tagging</b>	
JET_CR_JET_EffectiveNP_Detector1,2	energy scale uncertainty from the in situ analyses (detector)
JET_CR_JET_EffectiveNP_Modelling1,...,4	energy scale uncertainty from the in situ analyses (modelling)

---

Continued

---

Systematic uncertainty	Short description
JET_CR_JET_EffectiveNP_Statistical1,...,6	energy scale uncertainty from the in situ analyses (stat)
JET_CR_JET_EffectiveNP_Mixed1,2,3	energy scale uncertainty from the in situ analyses (mixed terms)
JET_CR_JET_EtaIntercalibration_Modeling	energy scale uncertainty on eta-intercalibration (modelling)
JET_CR_JET_EtaIntercalibration_TotalStat	energy scale uncertainty on eta-intercalibrations (statistics/method)
JET_CR_JET_EtaIntercalibration_NonClosure_highE	energy scale uncertainty on eta-intercalibrations (non-closure)
JET_CR_JET_EtaIntercalibration_NonClosure_negEta	as above
JET_CR_JET_EtaIntercalibration_NonClosure_posEta	as above
JET_CR_JET_BJES_Response	
JET_CR_JET_Flavor_Composition	energy scale uncertainty on $VV$ and $VH$ sample's flavour composition → Independent NP for : <code>_Top</code> , <code>_Vjets</code> , <code>_VV</code> processes
JET_CR_JET_Flavor_Response	energy scale uncertainty on samples' flavor response
JET_CR_JET_Pileup_OffsetMu	energy scale uncertainty on pile-up (Mu dependent)
JET_CR_JET_Pileup_OffsetNPV	energy scale uncertainty on pile-up (NPV dependent)
JET_CR_JET_Pileup_PtTerm	energy scale uncertainty on pile-up (pt term)
JET_CR_JET_Pileup_RhoTopology	energy scale uncertainty on pile-up (density $\rho$ )
JET_CR_JET_PunchThrough_MC16	energy scale uncertainty for punch-through jets
JET_CR_JET_SingleParticle_HighPt	energy scale uncertainty from the behavior of high- $p_T$ jets
JET_CR_JET_JER_EffectivNP_1,...6,7restTerm	energy resolution uncertainty split into 7 components
JET_CR_JET_JER_DataVsMC	energy resolution additional uncertainty difference between data than MC resolutions
JET_JvtEfficiency	JVT efficiency uncertainty
FT_EFF_Eigen_B0,...,45	$b$ -tagging efficiency uncertainties (“BTAG_LOOSE”) in continuous mode: 45 components for $b$ jets, 20 for $c$ jets and 20 for light jets
FT_EFF_Eigen_C0,...,20	as above
FT_EFF_Eigen_L0,...,20	as above
FT_EFF_Eigen_extrapolation	$b$ -tagging efficiency uncertainty on the extrapolation to high- $p_T$ jets
FT_EFF_Eigen_extrapolation_from_charm	$b$ -tagging efficiency uncertainty on tau jets
$E_T^{\text{miss}}$	
MET_SoftTrk_ResoPara	track-based soft term related longitudinal resolution uncertainty
MET_SoftTrk_ResoPerp	track-based soft term related transverse resolution uncertainty

Continued

Systematic uncertainty	Short description
MET_SoftTrk_Scale	track-based soft term related longitudinal scale uncertainty
MET_JetTrk_Scale	track MET scale uncertainty due to tracks in jets
Taus	
TAUS_TRUEHADTAU_SME_TES_DETECTOR	energy scale uncertainty: single-particle response + threshold
TAUS_TRUEHADTAU_SME_TES_INSITU	energy scale uncertainty: total from in-situ measurement
TAUS_TRUEHADTAU_SME_TES_MODEL	energy scale uncertainty: modelling + closure

### 7.2.1 Luminosity and Pile-up

As discussed in chapter 3 luminosity is used to measure how much data is recorded by the detector in any span of time. The uncertainty is calculated for each year of running and is determined to be 2.1%, 2.6%, 2.4%, and 2.0% for the years 2015,

How does this work?  
[120] does not combine different years 2016, 2017, 2018 respectively. A combined uncertainty is calculated for the entire period 2015–2018 at 1.7%. The methodology used to calculate this figure is similar to that detailed in [120], from calibrations of the luminosity scale using x–y beam-separation scans [121].

Pile-up uncertainties are computed by changing the nominal data scale of 1.0/1.03 to 1.0/1.00 and 1.0/1.18 to get the up and down variations respectively [122].

### 7.2.2 Triggers

#### $E_T^{\text{miss}}$ Triggers

Scale factors are derived for the  $E_T^{\text{miss}}$  triggers using  $W(\mu, \nu) + \text{jets}$  events as outlined in [100]. Three uncertainties are taken into account, the statistical error on the data-set used to derived the scale factor `METTrigStat`, a parameter used to account for the choice of physics process and the effect that might have on the determination of the scale factor `METTrigTop` and `METTrigSumPt` which aims to account for dependence of the offline  $S_T$  (as defined in 5.7.1) on the trigger efficiency. The uncertainty `METTrigTop` is named as such because it is derived from a comparison of the scale

factors as calculated with a  $t\bar{t}$  sample and compared to the nominal sample. The uncertainty relating to  $S_T$  is only applied to events recorded in 2017 due to a specific trigger used in this year.

### Lepton trigger

The nuisance parameter `EL_EFF_Trigger_Total_1NPCOR_PLUS_UNCOR` is used to control the overall uncertainty on the electron trigger. For the muons, the two components are considered `MUON_EFF_TrigSystUncertainty` and `MUON_EFF_TrigStatUncertainty`, which account for the systematic error and the statistical error on scale factor respectively. An up and down variation of  $1-\sigma$  are used for all of the aforementioned nuisance parameters.

## 7.2.3 Electrons

### Electron Efficiency Uncertainties

The efficiency of the electron reconstruction and identification has a systematic uncertainty called `ElectronEfficiencyCorrection` [123] calculated using the full run 2 data. Reconstruction is 97-99% efficient across the full  $p_T$  spectrum. Identification has an efficiency scale factor available from  $p_T > 7$  GeV. An isolation efficiency scale factor is also included. The latest uncertainties that are available include scale factors for  $p_T > 150$  GeV that are unity due to a lack of enough data to measure a scale factor. An additional systematic uncertainty of  $\pm 2\%$  is assigned above 150 GeV to mitigate this shortcoming. The above uncertainties are controlled by `EL_EFF_ID_Total_1NPCOR_PLUS_UNCOR`, `EL_EFF_Reco_Total_1NPCOR_PLUS_UNCOR`, and `EL_EFF_Iso_Total_1NPCOR_PLUS_UNCOR`.

### Electron Energy Scale and Resolution Uncertainties

A large number of electron energy scale and resolution systematic uncertainties are available [124]. The analysis is not very sensitive to these quantities and therefore

only two uncertainties are considered which are called EG\_RESOLUTION\_ALL and EG\_SCALE\_ALL.

### 7.2.4 Muons

#### Muon Efficiency Systematic Uncertainties

There must be a newer ones available?! If not used, say that.

**check for  
newer  
ones,  
1201200578**

Samples of  $Z \rightarrow \mu\mu$  and  $J/\psi \rightarrow \mu\mu$  events from the full 2015 dataset (corresponding to  $3.2 \text{ fb}^{-1}$ ) are used to calculate scale factors to account for uncertainties in the reconstruction, isolation and track-to-vertex association [92]. These scale factors are valid in the full  $p_T$  spectrum with the  $J/\psi$  measurement providing more accurate determination in the  $p_T < 15 \text{ GeV}$  region and the  $Z$  measurement being more accurate in the  $p_T > 15 \text{ GeV}$  region. Four independent systematic uncertainties are considered which are called MUON\_EFF\_RECO\_STAT, MUON\_EFF\_RECO\_STAT\_LOWPT, MUON\_EFF\_RECO\_SYS, MUON\_EFF\_RECO\_SYS\_LOWPT, which are split based on the  $p_T$  measurement. Statistical and systematic uncertainties on the isolation scale factor are controlled by MUON\_EFF\_ISO\_STAT and MUON\_EFF\_ISO\_SYS respectively. They are supported in the range of  $10 < p_T < 500 \text{ GeV}$ . For muons outside of this range, a scale factor of  $1 \pm 0.05$  is used. MUON\_EFF\_TTVA\_STAT and MUON\_EFF\_TTVA\_SYS control the systematic uncertainty on the scale factor of the cuts on the impact parameter significance and the  $|z_0 \sin \theta|$  which estimate the error on track-to-vertex association. All of the above systematic uncertainties are derived using  $\pm 1\sigma$  variations in the relevant samples.

#### Muon Momentum Scale and Resolution Uncertainties

Muon momentum scale and resolution uncertainties that are considered [92] have been calibrated using a sample of  $Z \rightarrow \mu\mu$  events in the region with  $p_T > 20 \text{ GeV}$  and with a sample of  $J/\psi \rightarrow \mu\mu$  events in region with  $p_T < 20 \text{ GeV}$ . Parameters exist to control uncertainties due to the inner detector, muon system and the overall momentum scale, they are called MUONS\_ID, MUONS\_MS and MUONS\_SCALE respectively. These systematic uncertainties are derived by varying the momentum scale

and the track position in the detector by  $\pm 1\sigma$ . Two parameters which account for the charge dependence of the momentum scale uncertainty MUON\_SAGITTA\_RHO and MUON\_SAGITTA\_RESBIAS are also included.

### 7.2.5 Taus

Uncertainties on the measurement of taus do not have a large effect on the analysis.

**why do  $\tau$  uncertainties have a small effect?**

TAUS\_TRUEHADTAU\_SME\_TES\_DETECTOR, TAUS\_TRUEHADTAU\_SME\_TES\_INSITU and TAUS\_TRUEHADTAU\_SME\_TES\_MODEL, are considered, which all account for different sources of energy scale uncertainty.

### 7.2.6 Jets

A baseline set of parameters controlling uncertainties on jets is available however in the analysis a set which is reduced to 23 nuisance parameters using principal component analysis [125] is used. The baseline set accounts for effects due to eta calibration, high- $p_T$  jets, pile-up, flavour composition, flavour response,  $b$ -jets, and punch-through jets. The 23 nuisance parameters and a short description of each are displayed in table 7.1 under the category jets and  $b$ -tagging where all of the relevant nuisance parameters start with JET and JET\_CR\_Flavour\_Composition represents three independent nuisance parameters as explained in its short description. The largest source of uncertainty amongst the chosen set come from the jet energy scale and the jet energy resolution. The determination of the former is documented in [126] and the latter is determined from data versus Monte-Carlo prediction comparisons.

### 7.2.7 $E_T^{\text{miss}}$

Systematic uncertainties considered account for calorimeter and track based jets, and can be found under the heading  $E_T^{\text{miss}}$  in table 7.1.

### 7.2.8 Flavour Tagging

Uncertainties relating to flavour tagging are expected to have a large impact on the analysis. Systematic uncertainties on the tagging are implemented as scale factors. The procedure is such that for each event the scale factor is applied if it is determined to contain a  $b$ -jet by the tagger, otherwise the inefficiency scale factor is applied, which yields the nominal event weight. For each of uncertainties considered the scale factor and inefficiency scale factor are varied and applied as above, note the varied inefficiency scale factor will no longer yield the nominal event weight.

Similarly to in the determination of the systematic uncertainties on the jets, a large number of individual systematics are available (about 40 per jet flavour) and so in order to have a smaller set to work with a principle component analysis of the full set is performed. For the working point and reduction scheme chosen in this analysis there are 3 variations for  $b$ -jets, 3 variations for  $c$ -jets and 5 variations for light-jets. These are shown in table 7.1, they all start with FT and end with a descriptor of which jet flavour they apply to e.g. B0.

Two additional systematic uncertainties irrespective of reduction scheme are considered that relate to the  $p_T$  extrapolation [127] and charm-to-bottom quark extrapolation, they are called also listed in table 7.1, start with FT and contain the descriptor `extrapolation`.

## 7.3 Systematic Uncertainties on $V+jets$ Events

The  $V+jets$  processes are simulated with SHERPA 2.2.1 [128] as mentioned in section 6.2, which is interfaced with the NNPDFs [129] for both the matrix element calculation and the parton shower tuning. A feature of SHERPA 2.2.1 is used which provides a combination of matrix elements with different parton multiplicities in order to simulate events with many additional jets, which contribute significantly to the background. Up to 2 extra partons are included in the next-to-leading order matrix element, and 3 or 4 extra partons are included at leading order in QCD.

Different parton multiplicities are combined using a matching scheme based on the CKKW-L [130, 131] merging technique, with a merging scale of  $Q_{\text{cut}} = 20 \text{ GeV}^1$ . Simulation of events with more than 4 extra partons rely on the parton shower algorithm of SHERPA. The parton shower and underlying event models are included in SHERPA whose generator adopts a full 5-flavour scheme with  $b$ - and  $c$ -quarks being treated as massless in the matrix element. Massive quarks can be produced in the parton shower and heavy flavours can be produced directly in the scattering process of the underlying event.

The analysis gains a lot of sensitivity from high  $p_T^V$  regions with two  $b$ -jets. Therefore it is desirable to have a dataset of predictions of these events such that statistical fluctuations are smaller than in the data. In order to achieve this samples are simulated in specific slices of  $\max(p_T^V, H_T)$ , which is the larger of the  $p_T$  of the vector boson or the  $H_T$  of the event. The following slices are used colon

$$\max(p_T^V, H_T) = [0\text{--}70, 70\text{--}140, 140\text{--}280, 280\text{--}500, 500\text{--}1000, > 1000] \text{ GeV}.$$

Also filters are applied on the flavour of jets in the event. The filters used are shown in table 7.2, and are not applied to the highest slice in  $\max(p_T^V, H_T)$ . The filtering

Is that standard positioning,  
in the middle of a  
sentence?

Filter	Description
$b$ -filter	at least 1 $b$ -hadron with $p_T > 0 \text{ GeV}$ and $ \eta  < 4$
	at least 1 $b$ -hadron with $p_T > 5 \text{ GeV}$ and $ \eta  < 2.9^\dagger$
$c$ -filter- $b$ -veto	at least 1 $c$ -hadron with $p_T > 4 \text{ GeV}$ and $ \eta  < 3$ veto events which pass the $b$ -filter
$c$ -veto- $b$ -veto	veto events which pass the $b$ -filter or the $c$ -filter- $b$ -veto

Table 7.2: Flavour filters used in the simulation of  $V + \text{jets}$  events.  $^\dagger$  this tighter filter is only applied to  $Z \rightarrow \nu\nu$  samples.

strategy for  $Z \rightarrow \nu\nu$  samples differs in the mc16e campaign using a combination of  $p_T^Z$  and  $m_{jj}$  to better populate the region above the  $E_T^{\text{miss}}$  trigger thresholds. These samples also use a tighter  $b$ -filter compared to their mc16a/d counterpart which is also described in table 7.2. All nominal  $V + \text{jets}$  samples with the corresponding

<sup>1</sup> $Q_{\text{cut}}$  describes the parton momentum sum above which the parton showering is used and below which the matrix element calculation is used.

$\max(H_T, p_T^V)$  slices and flavour filters are listed in tables B.1 B.2, and B.3 in the appendices. A set of alternative predictions are used for study and determination of priors of systematic uncertainties. If a discrepancy is common to a nominal versus data and an alternative versus data comparison then the discrepancy may arise from experimental uncertainty rather than a shortcoming of the modelling.

The alternative samples are generated using MADGRAPH 5 [132] interfaced to PYTHIA 8 for the modelling of the parton shower and the underlying event. The MADGRAPH 5 v2 generator provides a LO (QCD) description of these processes, merging together matrix-element calculations with different parton multiplicities, up to 4 additional jets, higher jet multiplicities are modelled by the parton shower algorithm. The merging scheme is the same as for the nominal samples, but has a merging scale of  $Q_{\text{cut}} = 30$  GeV. For the LO ME calculation the NNPDF2.3 LO PDFs are used, with  $\alpha_S = 1.3$ . Similarly to SHERPA 2.2.1, also MADGRAPH adopts a full 5-flavour scheme with massless quarks in the ME calculation, while massive quarks can be produced by the parton shower. All alternative  $V + \text{jets}$  samples are listed in tables B.9 and B.8 in the appendices.

As well as using an alternative Monte-Carlo generator the nominal generator, SHERPA, includes systematic variations internally. Every SHERPA 2.2.1  $V + \text{jets}$  sample has an event weight corresponding to each of the variations detailed in table 7.3<sup>2</sup>.

### $V + \text{jets}$ Cross Section

The  $V + \text{jets}$  cross sections are known at NNLO (QCD) [133], the higher order cross sections are used to normalise the  $V + \text{jets}$  samples in the analysis. For the  $W + \text{jets}$  samples the total cross section from SHERPA or from MADGRAPH, averaged across all 3 lepton flavours taking into account the different hadron filter efficiencies, is scaled to the NNLO prediction obtaining scaling factor of  $k_{\text{NNLO}}^{\text{QCD}} = 0.9702$ .

---

<sup>2</sup>Some of the variations cannot be produced by SHERPA 2.2.1 and so SHERPA 2.1 is used instead. For these variations half of the variation in each direction is taken as the uncertainty rather than comparing to the central value of the SHERPA 2.2.1 prediction.

Variation	Values	
<b>Sherpa 2.2.1</b>		
Factorisation scale ( $\mu_F$ )	$2\mu_F$	$0.5\mu_F$
Renormalisation scale ( $\mu_R$ )	$2\mu_R$	$0.5\mu_R$
PDF Variation	MMHT2014nnlo68cl	CT14nnlo
<b>Sherpa 2.1</b>		
Re-summation scale ( $\mu_S$ )	$2\mu_S$	$0.5\mu_S$
CKKW Merging scale	15 GeV	30 GeV

Table 7.3: A summary of the SHERPA 2.2.1 and SHERPA 2.1 internal variations that are used to model  $V$ +jets processes.

cut is colloquial: requirement, selection criteria.... at before.... For  $Z$ +jets some subtleties must be considered. For  $Z \rightarrow \ell\ell$ +jets the cut at generator level,  $m_{\ell\ell} > 40$  GeV, must be taken into account. The NNLO calculation uses a cut of  $66 < m_{\ell\ell} < 116$  GeV, which can be applied at before reconstruction to the samples of this analysis in order to get the scaling correct as follows

$$k_{\text{NNLO}}^{\text{QCD}} = \frac{\sigma_{\text{NNLO}}(66 < m_{\ell\ell} < 116 \text{ GeV})}{\sigma_{\text{SHERPA, MADGRAPH}}(66 < m_{\ell\ell} < 116 \text{ GeV})}, \quad (7.5)$$

no ?? hence the scaling factor is found to be 0.9751. For  $Z \rightarrow \nu\nu$  +jets the NNLO no theoretical cross section is available. Values from the Particle Data Group (PDG) [134] are used to correct for the difference between the  $Z \rightarrow \nu\nu$  and  $Z \rightarrow \ell\ell$  branching ratios, and the NNLO cross section is used without any mass cuts and with the  $Z/\gamma^*$  interference removed. The scaling factor is therefore calculated to be 0.9728. The discrepancy in the factors for  $Z \rightarrow \ell\ell$  and  $Z \rightarrow \nu\nu$  events can be explained by the fact the two generators use different branching ratios for these processes compared with those used by the theoretical calculations.

### 7.3.1 Systematic Uncertainties on $W$ +jets Events

A number of nuisance parameters are introduced to account for modelling uncertainties on the  $W$ +jets background process. These uncertainties are considered and derived in the 0– and 1– lepton channels only, as the amount of  $W$ +jets background present in the 2–lepton channel is negligible. A summary of all of the uncertainties

for this background can be found in table B.13.

### Section 7.3

The nominal and alternative predictions described in 7.3 are used to in the determination of all of the values described in the following.

## Normalisation and Acceptance Uncertainties

A summary of the normalisation uncertainties is shown in table 7.4. A single

middle of the sentence

Name	Process	2-jet			3-jets		
		CR <sub>low</sub>	SR	CR <sub>high</sub>	CR <sub>low</sub>	SR	CR <sub>high</sub>
<b>0-Lepton</b>							
SysWbbNorm_L0	W+hf	5 %	5 %	N/A	N/A	N/A	N/A
SysWbbCRSReXtrap	W+hf	-7.7 %	N/A	14.9 %	-5.6 %	N/A	7.0 %
<b>1-Lepton</b>							
SysWclNorm	W+l				32%		
SysWlNorm	W+cl				37%		
norm_Wbb_J2	W+hf	Floating	Normalisation		N/A	N/A	N/A
norm_Wbb_J3	W+hf	N/A	N/A	N/A	Floating	Normalisation	
SysWbbCRSReXtrap	W+hf	-115 %	N/A	14.8 %	-3.6 %	N/A	5.1 %

Table 7.4: A summary of the nuisance parameters used to account for the uncertainty on the normalisation of simulated  $W$ +jets predictions. The values shown correspond to a prior representing a  $1-\sigma$  shift in the relevant Gaussian penalty term.

nuisance parameter is introduced for each of the  $W + cl$  and  $W + l$  processes which are both heavily suppressed by the analysis requirement of two b-tagged jets. There is a large mismodelling of the normalisation of the  $W+hf$  process and so floating normalisations are used separately in the 2-jet and 3-jet categories. Priors for several acceptance uncertainties are calculated using equation 7.1<sup>3</sup>, which are used to control the migration of events between the 0– and 1–lepton channels and the migration between the  $\Delta R(b, \bar{b})$  regions.

A parameter controlling migration between the two channels considered, SysWbbNorm\_L0, is applied to both the 2-jet and 3-jet regions of the 0-lepton channel. The choice to apply to the 0-lepton channel and not the 1-lepton channel is made because the 1-lepton channel provides a better constraint on the  $W+hf$

<sup>3</sup>Whilst the priors are obtained with the double ratio formula involving two regions, the effect of the uncertainty is applied to only one of the two regions.

process. Whilst all alternative predictions are considered the size of this prior is dominated by the difference between the nominal prediction and MADGRAPH.

Migration between the  $\Delta R(b, \bar{b})$  control regions and the signal region is controlled by one parameter per control region. Named `SysWbbCRSReXtrap`, it can be seen in table 7.4 that it is applied in each of the control regions, and is also correlated across jet multiplicity and leptonic channel. The shape uncertainty which uses the BDTr method discussed in section 7.1.2 induces migration effects across the analysis regions, however that systematic is taken as a shape only effect. The migrations induced by the shape uncertainty agree well with those derived from comparisons between nominal and alternative predictions, and the sign of the final priors is determined by migration due to the shape uncertainty. The magnitude of the prior is calculated as the quadrature sum of the difference between the nominal prediction and MADGRAPH as well as the scale and parton density function variations included in SHERPA. There is no dedicated nuisance parameter to control the migration between the  $p_T^V$  regions of the analysis as this migration is controlled by the  $p_T^V$  shape uncertainty.

## Flavour Composition Uncertainties

As introduced at the start of this chapter the  $W+jets$  background has a  $W+hf$  component that is comprised of the  $bb$ ,  $bc$ ,  $bl$  and  $cc$  sub-components. The fractional contribution of each of these sub-components is detailed in table 7.5. Flavour composition uncertainties are computed using equation 7.1 for each sub-component, the calculated priors are summarised in tables 7.6 separately in the 0– and 1–lepton channels<sup>4</sup>. The flavour composition uncertainties are dominated by the difference between SHERPA and MADGRAPH. Individual uncertainties for different jet multiplicities or analysis regions are not considered necessary due to a small amount of non- $bb$  events remaining after the 2  $b$ -jet requirement is imposed.

---

<sup>4</sup>Whilst these numbers were calculated in this iteration of the analysis, incomplete availability of SHERPA internal variations lead to a set of numbers from a previous iteration being used, those are displayed in this work.

Process	2 jet		SR	3 jet		SR
	CR <sub>low</sub>	CR <sub>high</sub>		CR <sub>low</sub>	CR <sub>high</sub>	
<b>0-lepton</b>						
$Wcc$	0.03	0.02	0.03	0.03	0.03	0.04
$Wbl$	0.01	0.03	0.09	0.01	0.02	0.07
$Wbc$	0.02	0.05	0.10	0.02	0.04	0.10
$Wbb$	0.94	0.90	0.78	0.94	0.91	0.79
<b>1-lepton</b>						
$Wcc$	0.03	0.05	0.03	0.03	0.04	0.03
$Wbl$	0.01	0.03	0.01	0.01	0.05	0.02
$Wbc$	0.03	0.15	0.07	0.03	0.13	0.05
$Wbb$	0.93	0.77	0.89	0.94	0.77	0.90

Table 7.5: The contribution of each  $W + \text{hf}$  subprocesses to the total  $W + \text{hf}$  yields in all 1-lepton and 0-lepton channels analysis regions expressed as a decimal.

Name	Ratio	Prior	Region
SysWbcWbbRatio	$W + bc/W + bb$	15%	0-Lepton
		30%	1-Lepton
SysWblWbbRatio	$W + bl/W + bb$	10%	0-Lepton
		30%	1-Lepton
SysWccWbbRatio	$W + cc/W + bb$	26%	0-Lepton
		23%	1-Lepton

Table 7.6: A summary of the flavour composition uncertainties applied to the components of the  $W + \text{hf}$  process. All uncertainties are only applied to events belonging to the component in the numerator of the ratio in the region specified.

### Shape Uncertainties

#### section 7.1.3

The hybrid  $(N - 1)$ -dimensional parametrisation technique described in 7.1.3 is used to generate uncertainties on the shape of the distributions of the  $W + \text{jets}$  background. The variable that has been factorised out is  $p_T^V$  and so there are nuisance parameters which control the  $p_T^V$  shape in each of the 2– and 3-jet categories as well as a nuisance parameter that controls the multi-variate shape uncertainty across all regions and jet multiplicities. These nuisance parameters are summarised in table 7.7.

Shapes are derived in regions split by leptonic channel and heavy flavour sub-component, yielding eight regions total. The  $W + l$  and  $W + cl$  components of the  $W + \text{jets}$  backgrounds are not considered as they account for less than 1 % of all

Nuisance Parameter	Description	Samples/Categories	Effect
SysWPtV_J2	$p_T^V$ shape	$W+jets$ , 2-jet region	Migration+Shape
SysWPtV_J3	$p_T^V$ shape	$W+jets$ , 3-jet region	Migration+Shape
SysBDTr_W_SHtoMG5	Multi-variate shape	$W+jets$ , all regions	Shape

Table 7.7: Systematic uncertainties aiming to account for differences in the shape between predictions and data in  $W+jets$  events.

background events in any region.

In all regions the initial 1-dimensional re-weighting is performed by comparing the ratio of the nominal SHERPA 2.2.1 ( $p_T^V, E_T^{\text{miss}}$ ) prediction with that of MADGRAPH. The requirement of  $p_T^V > 75$  GeV is applied to all distributions in the 1-lepton channel and in the 0-lepton channel the requirement of  $E_T^{\text{miss}} > 150$  GeV is imposed instead. Figures 7.1 and 7.2 show the comparison between SHERPA and MADGRAPH in the  $p_T^V$  variable in the 2- and 3-jet categories, respectively. The systematic uncertainties on the shape are shown in the ratio panel at the bottom of each plot in green.

The low number of events available in the 0-lepton channel mean that it is not statistically viable to generate a parametrisation in that channel, therefore the shapes derived in the 1-lepton channel are also used in the 0-lepton channel, a comparison of the two approaches has been studied [119].

As per the methodology of the factorised BDTr technique the nominal SHERPA 2.2.1 events are re-weighted by the  $p_T^V$  ratio with the alternative prediction. The re-weighted SHERPA 2.2.1 events are then trained against the MADGRAPH prediction in a BDT classifier that uses the input variables detailed in table 6.4. A separate training is performed in each jet category and for each sub-component of the  $W+hf$  process in congruence with the 8 regions that the  $p_T^V$  shape was derived in. Considering the output of the classifier the ratio of the scores of the nominal and alternative predictions enter into the profile-likelihood fit as a nuisance parameter called SysBDTr\_W\_SHtoMG5. All of the following plots shown contain only  $W + bb$  events as this is the dominant and most important component of the  $W+hf$  process and indeed the  $W+jets$  background in the analysis.

In order to check the validity of the method the re-weighted nominal predictions

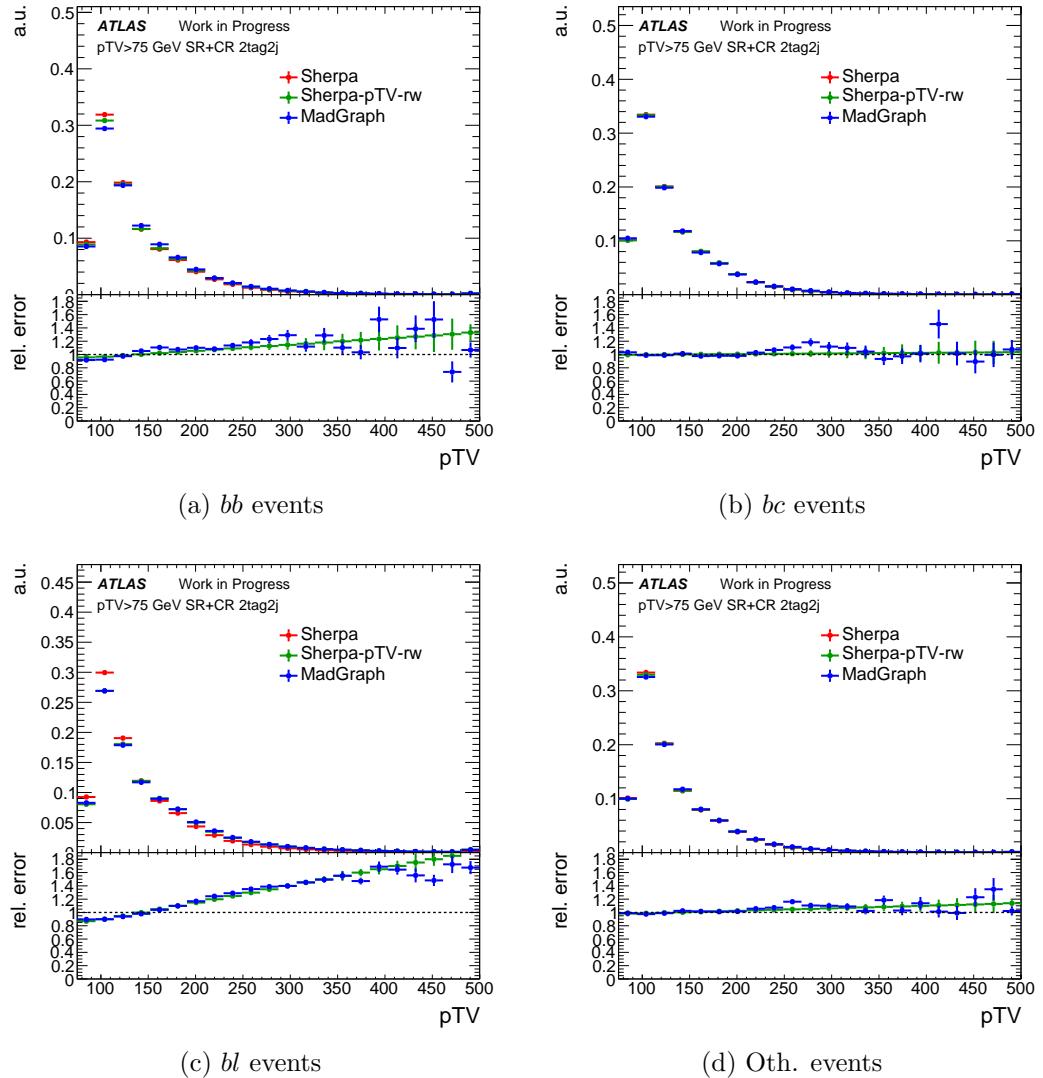


Figure 7.1:  $p_T^V$  shape systematic in the 2-jet category in the 1-lepton channel for each heavy flavour sub-component. The red and blue histograms correspond to the  $p_T^V$  prediction of SHERPA 2.2.1 and MADGRAPH respectively. The  $p_T^V$  shape systematic is plotted in green.

of inputs to the analysis MVA are plotted against the nominal and alternative predictions in figures 7.3 and 7.4 for the 1- and 0-lepton channels respectively in the 2-jet category. Should the re-weighted nominal inputs match the shapes of the alternative inputs then the method is considered to be valid. Indeed upon inspection one can see that this is largely the case across the board apart from small levels of disagreement. These small levels of disagreement are disregarded as the two most important variables in the determination of the analysis BDT score,  $m(b, \bar{b})$ ,  $p_T^V$  and  $E_T^{\text{miss}}$  have good levels of agreement.

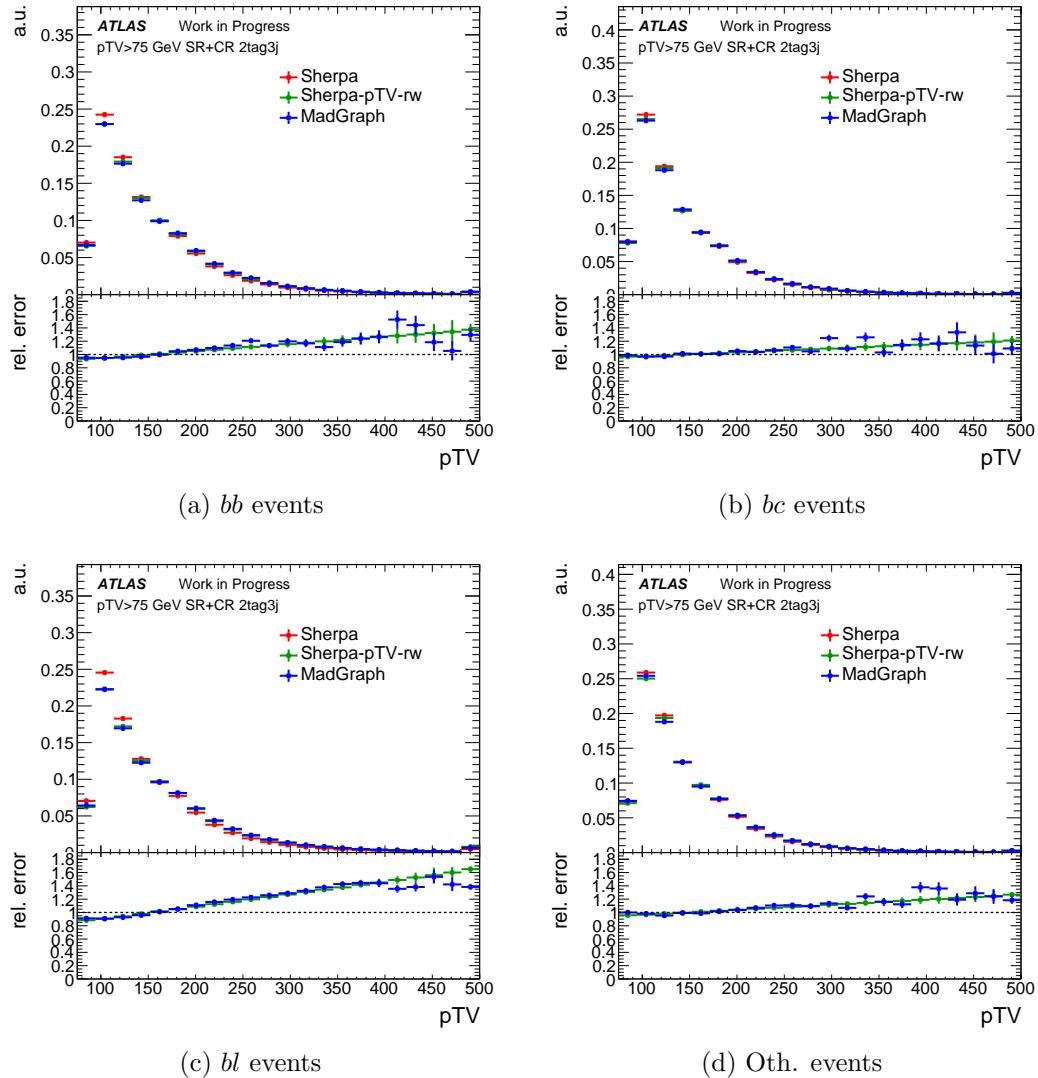


Figure 7.2:  $p_T^V$  shape systematic in the 3-jet category in the 1-lepton channel for each heavy flavour sub-component. The red and blue histograms correspond to the  $p_T^V$  prediction of SHERPA 2.2.1 and MADGRAPH respectively. The  $p_T^V$  shape systematic is plotted in green.

The actual effect of the factorised BDTr systematic uncertainty on the analysis BDT score is shown in figures 7.5 and 7.6 for the 1- and 0-lepton channels respectively. Tables B.15 and B.16 in the appendix show the extrapolation uncertainties induced by the  $p_T^V$  and factorised BDTr shape systematic uncertainties respectively.

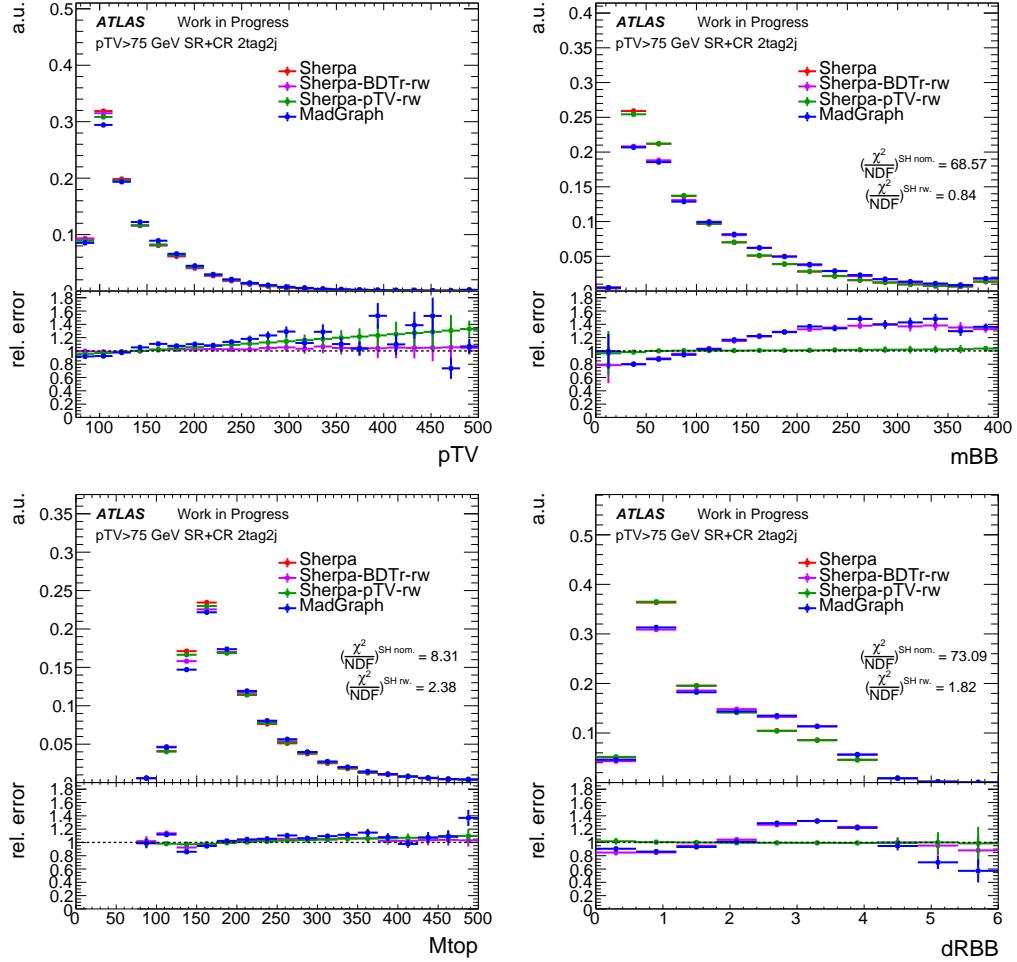


Figure 7.3: Histograms of the nominal predictions in the 1-lepton 2-jet region of variables entering into the analysis MVA are shown alongside the alternative predictions from MADGRAPH and the re-weighted nominal prediction. The  $p_{\text{T}}^V$  shape systematic which was calculated in advance of the training is shown in green.

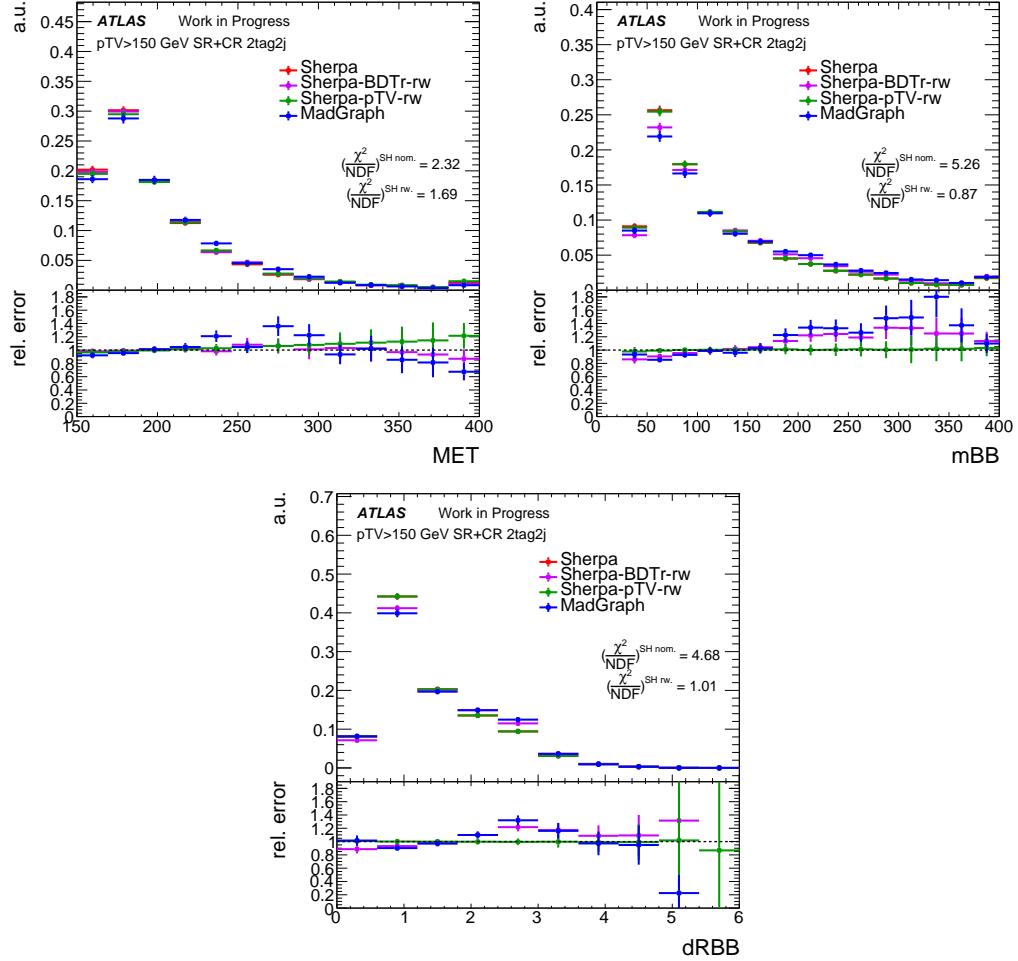


Figure 7.4: Histograms of the nominal predictions in the 0–lepton 2–jet region of variables entering into the analysis MVA are shown alongside the alternative predictions from MADGRAPH and the re-weighted nominal prediction. The  $p_T^V$  shape systematic which was calculated in advance of the training is shown in green.

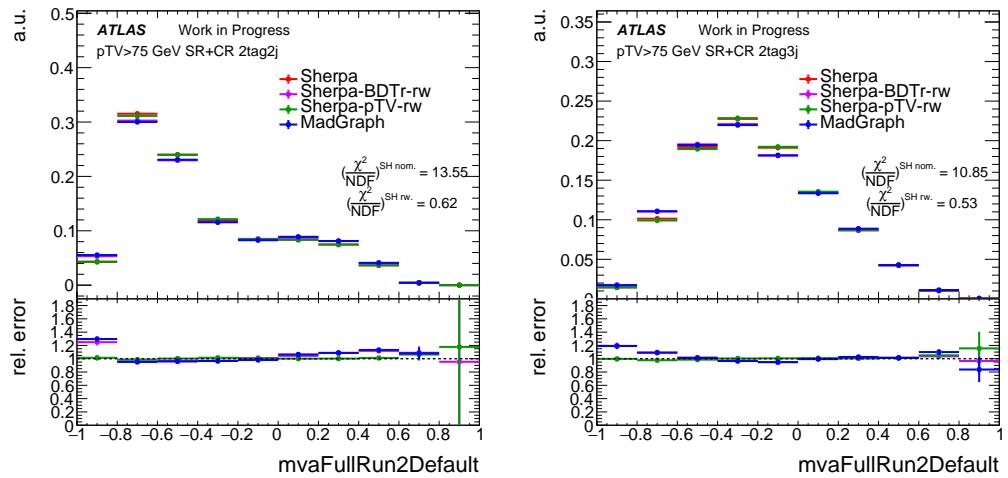


Figure 7.5: Histograms of the nominal predictions in the 1–lepton channel of the analysis MVA score are shown alongside the alternative predictions from MADGRAPH and the re-weighted nominal prediction. The  $p_T^V$  shape systematic which was calculated in advance of the training is shown in green. Only W + bb events are shown.

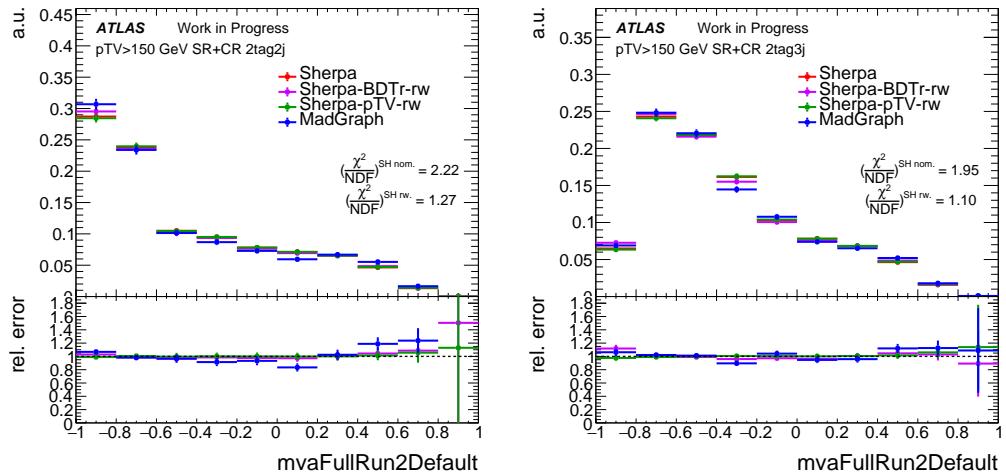


Figure 7.6: Histograms of the nominal predictions in the 0-lepton channel of the analysis MVA score are shown alongside the alternative predictions from MADGRAPH and the re-weighted nominal prediction. The  $p_T^V$  shape systematic which was calculated in advance of the training is shown in green. Only W + bb events are shown.

### 7.3.2 Systematic Uncertainties on $Z+jets$ Events

Systematic uncertainties are considered in the 0– and 2–lepton channels only as there is a negligible contribution of this background in the 1–lepton channel.

#### Normalisation and Acceptance Uncertainties

Several normalisation and acceptance uncertainties are considered for the  $Z + jets$  background, they are summarised in table 7.8. One nuisance parameter is used for

Name	Process	2-jet			$(\geq)3$ -jets		
		CR <sub>low</sub>	CR <sub>high</sub>	SR	CR <sub>low</sub>	CR <sub>high</sub>	SR
<b>0-lepton</b>							
SysZclNorm	$Z+l$				23%		
SysZ1Norm	$Z+cl$				18%		
norm_Zbb_J2	$Z+hf$		Floating Normalisation			N/A	
norm_Zbb_J3	$Z+hf$		N/A			Floating Normalisation	
SysZbbNorm_OL	$Z+hf$				7%		
SysZbbPTV	$Z+hf$		$p_T^V$ dependent migrations covered by $p_T^V$ shape systematic				
SysZbbCRSRExtrapolation_CRLow	$Z+hf$	-6.0 %	N/A	N/A	-6.6 %	N/A	N/A
SysZbbCRSRExtrapolation_CRHigh	$Z+hf$	N/A	3.8 %	N/A	N/A	3.9 %	N/A
<b>2-lepton</b>							
SysZclNorm	$Z+l$				23%		
SysZ1Norm	$Z+cl$				18%		
norm_Zbb_J2	$Z+hf$		Floating Normalisation			N/A	
norm_Zbb_J3	$Z+hf$		N/A			Floating Normalisation	
SysZbbPTV	$Z+hf$		$p_T^V$ dependent migrations covered by $p_T^V$ shape systematic				
SysZbbCRSRExtrapolation_CRLow	$Z+hf p_T^V > 150$ GeV	-9.9 %	N/A	N/A	-3.8 %	N/A	N/A
SysZbbCRSRExtrapolation_CRHigh	$Z+hf p_T^V > 150$ GeV	N/A	2.7 %	N/A	N/A	4.1 %	N/A
SysZbbCRSRExtrap_BMin75_L2_CRLow	$Z+hf 75\text{--}150$ GeV $p_T^V$	3.8 %	N/A	N/A	-9.9 %	N/A	N/A
SysZbbCRSRExtrap_BMin75_L2_CRHigh	$Z+hf 75\text{--}150$ GeV $p_T^V$	N/A	2.7 %	N/A	N/A	-4.1 %	N/A

Table 7.8: A summary of nuisance parameters which are used to control the  $Z+jets$  normalisation in the relevant regions that enter into the profile-likelihood fit. The values in the table correspond to a  $1-\sigma$  deviation of the calculated prior unless otherwise stated.

each of the  $Z+l$  and  $Z+cl$  components as they are sub-dominant accounting for less than 1 % of the total background in any region due to the requirement of two  $b$ -jets, they are called `SysZ1Norm` and `SysZclNorm` respectively. These normalisations are each correlated across all regions.

The  $Z+hf$  process has a large normalisation uncertainty and so separate nuisance parameters are used for the 2–jet and 3–jet regions, ~~these are~~ called `norm_Zbb_J2` and `norm_Zbb_J3` respectively. These parameters are heavily constrained in the fit due to a very large number of  $Z+jets$  events in the 0– and 2–lepton signal regions and due to high  $Z+jets$  purity particularly in the 2–lepton control regions.

Nuisance parameters are introduced to the model migration of events between regions. The priors for these parameters are calculated using the double ratio in equation 7.1. Firstly a parameter is introduced in order to account for the difference in the number of  $Z$ +jets in the 0– and 2–lepton channels, is it called `SysZbbNorm_0L` and is applied to  $Z$ +hf events in the 0–lepton channel as the 2–lepton channel has a higher purity of  $Z$  + jets events and therefore yields a better constraint on the normalisation. The size of the prior as calculated by comparing SHERPA 2.2.1 and MADGRAPH is similar to the size as calculated by examining the SHERPA internal weights.

The nuisance parameter that controls the  $Z$ +jets  $p_T^V$  uncertainty, `SysZPTV` (see section 7.3.2), is also allowed to control the relative number of events between analysis regions. This choice is made as the magnitude of the extrapolation uncertainties shown in table 7.9 are similar in size to the largest variations one would obtain by comparing the nominal prediction to the alternatives.

Region	$\text{CR}_{\text{low}}$	SR	$\text{CR}_{\text{high}}$	Weighted Sum	$\Delta(\text{SR}, \text{CR}_{\text{low}})$	$\Delta(\text{CR}_{\text{high}}, \text{SR})$
<b>2-lepton 2-jet</b>						
75 GeV– 150 GeV	0.68	0.78	0.55	0.685	0.1	-0.23
150 GeV– 250 GeV	-2.02	-2.08	-2.17	-2.119	0.06	-0.09
$\geq 250$ GeV	-4.31	-4.7	-4.67	-4.67	0.39	0.03
<b>Differences</b>						
$\Delta(150\text{--}250, 75\text{--}150)$	-2.7	-2.86	-2.72	-2.805	-0.04	0.14
$\Delta(\geq 250, 150\text{--}250)$	-2.29	-2.62	-2.5	-2.547	0.33	0.12
<b>2-lepton <math>\geq 3</math>-jet</b>						
75GeV– 150GeV	0.53	0.64	0	0.366	0.11	-0.64
150GeV– 250GeV	-2.16	-2.18	-2.26	-2.222	-0.01	-0.08
$\geq 250$ GeV	-4.47	-4.85	-4.91	-4.870	-0.38	-0.06
<b>Differences</b>						
$\Delta(150\text{--}250, 75\text{--}150)$	-2.69	-2.82	-2.26	-2.588	0.1	0.56
$\Delta(\geq 250, 150\text{--}250)$	-2.31	-2.67	-2.65	-2.647	-0.37	0.02
<b>0-lepton 2-jet</b>						
150 GeV– 250 GeV	-2.05	-2.10	-2.20	-2.14	-0.05	-0.10
$\geq 250$ GeV	-4.22	-4.67	-4.67	-4.64	-0.45	0.0
<b>Differences</b>						
$\Delta(\geq 250, 150\text{--}250)$	-2.17	-2.57	-2.47	-2.49	-0.40	0.10
<b>0-lepton 3-jet</b>						
3 jet 150GeV– 250GeV	-2.15	-2.16	-2.25	-2.20	-0.01	-0.09
3 jet $\geq 250$ GeV	-4.37	-4.70	-4.75	-4.71	-0.33	-0.05
<b>Differences</b>						
$\Delta(\geq 250, 150\text{--}250)$	-2.22	2.54	-2.5	-0.85	-0.32	0.04

Table 7.9: A table of values showing the extrapolation uncertainties induced by the  $p_T^V$  shape uncertainty. The weighted sum shows the induced normalisation weighted by the number of events in the relevant region.

Induced normalisation effects by the  $m_{bb}$  shape are smaller than those arising

from comparison of the different available predictions. Therefore this systematic uncertainty is applied as a shape effect only and an additional nuisance parameter is introduced to cover the migration between analysis regions. As will be explained in section 7.3.2 it is necessary to de-correlate the  $m_{bb}$  shape systematic in the lowest  $p_T^V$  bin of the analysis. The nuisance parameter controlling the migration between analysis regions in the  $m_{bb}$  variable is also de-correlated in that bin. These acceptance uncertainties are applied to the  $CR_{high}$  and  $CR_{low}$  separately, the nuisance parameter are `SysZbbCRSRExtrapolation_CRLow`, de-correlated as `SysZbbCRSRExtrap_BMin75_L2_CRLow` and `SysZbbCRSRExtrapulation_CRHigh` de-correlated as `SysZbbCRSRExtrap_BMin75_L2_CRHigh`. The determination of the prior for these nuisance parameters is dominated by the difference between SHERPA 2.2.1 and MADGRAPH.

### Flavour Composition Uncertainties

The  $Z+hf$  process breaks down in the exact same way as the  $W+hf$  process, into  $(b,b)$ ,  $(b,c)$ ,  $(b,l)$  and  $(c,c)$  sub-components. Uncertainties on the fraction that each of these makes up of the  $Z+hf$  process are found in table 7.10. They are calculated

Name	Ratio	Prior	Region
<code>SysZbcZbbRatio</code>	$\frac{Z + bc}{Z + bb}$	40%	0-Lepton
		40%	2-Lepton 2-jet
		30%	2-Lepton $\geq 3$ -jet
<code>SysZblZbbRatio</code>	$\frac{Z + bl}{Z + bb}$	25%	0-Lepton
		28%	2-Lepton 2-jet
		20%	2-Lepton $\geq 3$ -jet
<code>SysZccZbbRatio</code>	$\frac{Z + cc}{Z + bb}$	15%	0-Lepton
		16%	2-Lepton 2-jet
		13%	2-Lepton $\geq 3$ -jet

Table 7.10: A summary of the flavour composition uncertainties applied to the components of the  $Z + hf$  process. All uncertainties are only applied to events belonging to the component in the numerator of the ratio in the region specified.

using equation 7.1 and applied to the non- $(b,b)$  component of the ratio. The nuisance parameters entering into the fit are `SysZbcZbbRatio`, `SysZccZbbRatio` and

**SysZblZbbRatio.** Priors are calculated separately in the 0– and 2–lepton channels, and in the 2–lepton channel priors are also calculated separately in the 2– and 3–jet categories. The variations are nonetheless considered to be correlated across regions. The value of the priors is completely dominated by the difference between the nominal SHERPA 2.2.1 prediction and the alternative MADGRAPH prediction.

### Shape Uncertainties

The shape uncertainties for the  $Z$ +jets process are derived using the 1–dimensional parametrisation approach detailed in section 7.1.1 with one notable difference. Instead of a ratio of predictions generated with different parameters, the ratio in this case is of the nominal SHERPA 2.2.1 prediction and the data itself. In order to be valid this method has a number of requirements, the first of which is that the region of interest must have a high purity of the background for which the uncertainty is derived. In the case of the 2–lepton channel the sidebands of the region which contains both the  $VH \rightarrow b\bar{b}$  signal and the  $VZ \rightarrow b\bar{b}$  background, defined as  $80\text{GeV} < m_{bb} < 140\text{GeV}$ , is quite pure in  $Z$ +jets events. The purity can be enhanced by subtracting from the data the top quark background events which contribute second to the  $Z$ +jets events, this can be done with a high degree of certainty over the template as the data from the top  $e\mu$  CR can be used. The remaining backgrounds contribute a very small amount to the region and so even with large uncertainties on their template the method is still sufficiently accurate. Neither the signal or diboson background are subtracted from the data as they have been removed by the veto of the aforementioned range in  $m_{bb}$  and the veto is applied to the selection so that events in the range are removed from all distributions considered.

This method is used to derived shape uncertainties on the  $p_T^V$  and  $m_{bb}$  variables as they are considered to be the most important in the analysis. The

strange formatting

uncertainty is used in the 0–lepton channel which adds an additional complication. As mentioned in section 5.7.3 a kinematic fit is used in order to correct the  $b$ -jet momentum and these corrected jets are then used in the formulation of  $m_{bb}$  in the 2–lepton channel. This means that  $m_{bb}$  in the 0– and 2–lepton channels refers to a

different quantity and so a shape uncertainty derived on one may not be suitable to use on the other. In order to get around this issue the shape uncertainty for  $m_{bb}$  is derived on a variable that is calculated using globally sequentially calibrated (GSC) jets and is called GSC  $m_{bb}$ , these jets have not undergone the kinematic fit procedure and so ought to match more closely between the two leptonic channels. The aforementioned veto on events in the range  $80\text{GeV} < m_{bb} < 140\text{GeV}$  is therefore actually applied based on GSC  $m_{bb}$ .

The plots in figures 7.7 and 7.8 show the nominal SHERPA 2.2.1 prediction of the  $Z+\text{jets}$  background in the 2-lepton channel in the sum of analysis regions CR<sub>low</sub>, SR and CR<sub>high</sub> plotted against the subtracted data in the  $p_T^V$  and GSC  $m_{bb}$  variables respectively. The subtracted data is the data with a template comprised of the data from the top  $e\mu$  CR and the nominal predictions of all backgrounds other than  $Z+\text{jets}$  subtracted from it. Plots are shown in the 2-jet,  $\geq 3$ -jet and  $\geq 2$ -jet categories (the sum of the two jet multiplicity based categories in the analysis) and are plotted across all  $p_T^V$  bins of the analysis. The shape uncertainty is derived by taking a fit to the ratio in the bottom pane of each plot, the fit defines the  $1-\sigma$  up variation of the Gaussian constrained nuisance parameter, the symmetrised version of the fit (not displayed) defines the  $1-\sigma$  down variation. The  $\geq 2$ -jet region is used to calculate the prior for the systematic uncertainties SysZPtV and SysZMbb that are applied to all  $Z+\text{jets}$  events in the 0- and 2-lepton channels.

comma

The nuisance parameters are correlated across leptonic channels, however as can be seen by inspection of figure 7.7 the fitted shape does not cover the discrepancy between the subtracted data and the prediction at low values of  $m_{bb}$ . Investigating the plots in figure 7.9 shows that this discrepancy is coming from the  $75\text{ GeV} < p_T^V < 150\text{ GeV}$  analysis bin, therefore the aforementioned nuisance parameters are decorrelated in this bin. This is so that the affect of opposite sign deviations between data and the prediction do not cancel each other out. The decorrelated parameters are named SysZPtV\_BMin75\_L2 and SysMbb\_BMin75\_L2. A summary of the  $Z+\text{jets}$  shape uncertainties is found in table 7.11.

The fits to the ratios of the relevant quantities are performed on data in the

data with a template,  
reformulate

What is the fit function?

Why?????  
They are  
effect  
This is so????  
not to 80%  
correlated  
(Fig 8.2),  
only norm  
are, not  
syst.

Nuisance Parameter	Uncertainty	Region
SysZPtV	$p_T^V$ shape & normalisation	All regions with $p_T^V > 150$ GeV
SysZMbb	$m_{b\bar{b}}$ shape	All regions with $p_T^V > 150$ GeV
SysZPtV_BMin75_L2	$p_T^V$ shape & normalisation	All regions with $75 \text{ GeV} p_T^V < 150 \text{ GeV}$
SysZMbb_BMin75_L2	$m_{b\bar{b}}$ shape	All regions with $75 \text{ GeV} p_T^V < 150 \text{ GeV}$

Table 7.11: A summary of the shape uncertainties for the  $Z$ +jets background, all fits to data take place inclusive of jet multiplicity,  $p_T^V$  bin and analysis region in the 2-lepton channel.

0–1 TeV range. The functional form of the fit to the  $p_T^V$  ratio is logarithmic, whereas the GSC  $m_{b\bar{b}}$  ratio is linear and it's value at 300 GeV is levelled to the end of the range. This is to make sure that the shape uncertainty is not overly influenced by events far from the Higgs mass. Similarly the binning of the GSC  $m_{b\bar{b}}$  histograms is chosen to give the fit more data points close the Higgs mass. The binning of the  $p_T^V$  distributions is chosen as to put more points where there are more available events.

needs to  
come  
earlier

its

Explain procedure and results

Where does the template come in?

What is the fit function to the ratio?

compare to previous results

Comparison to fig 7.9

Why does this show that the nuisance par. decorrelated?

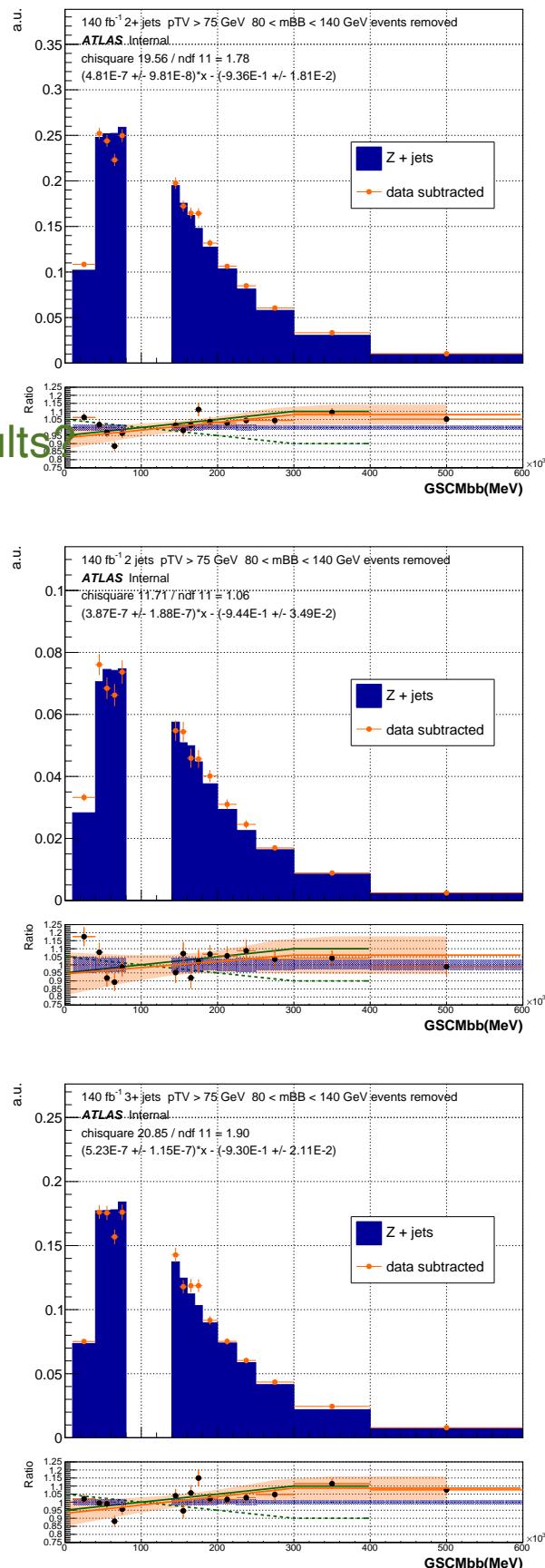


Figure 7.7: Subtracted data versus the nominal  $Z + \text{jets}$  prediction in the GSC  $m_{bb}$  distribution. The green lines in the ratio show symmetrised uncertainty used in the previous analysis. The orange line is the fit to the ratio of the subtracted data and the prediction, the shaded region represents the 95 % confidence interval.

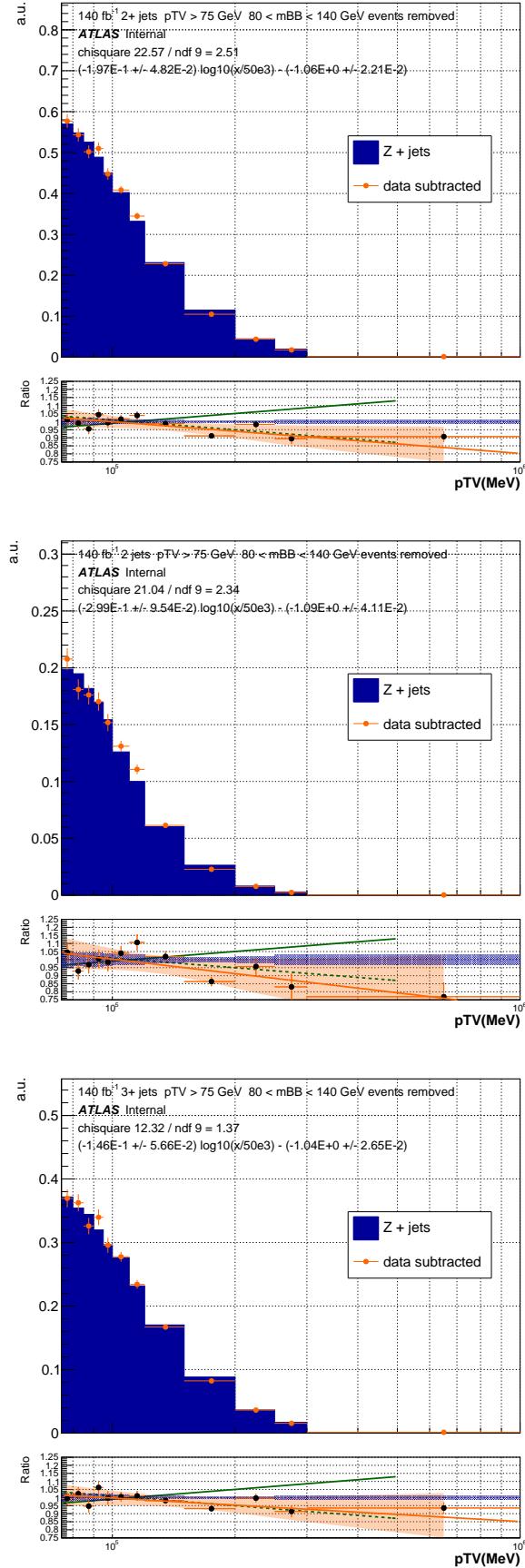


Figure 7.8: Subtracted data versus the nominal  $Z + \text{jets}$  prediction in the  $p_{\text{TV}}$  distribution. The green lines in the ratio show symmetrised uncertainty used in the previous analysis. The orange line is the fit to the ratio of the subtracted data and the prediction, the shaded region represents the 95 % confidence interval.

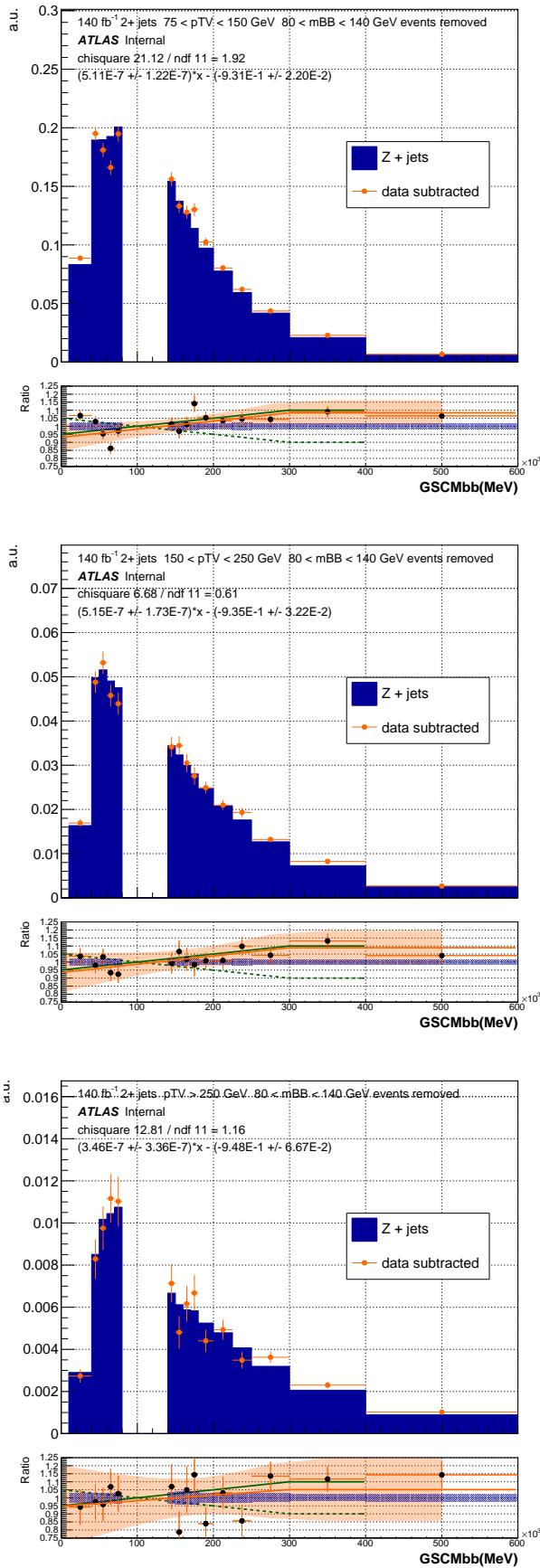


Figure 7.9: Subtracted data versus the nominal  $Z + \text{jets}$  prediction in the GSC  $m_{bb}$  distribution, shown in each of the  $p_T^V$  bins of the analysis. The green lines in the ratio show symmetrised uncertainty used in the previous analysis. The orange line is the fit to the ratio of the subtracted data and the prediction, the shaded region represents the 95 % confidence interval.

## 7.4 Systematic Uncertainties on $t\bar{t}$ Events

Modelling of  $t\bar{t}$  events differs between the leptonic channels of the analysis. Like many of the differences between channels that have been discussed already the reason for this ultimately boils down to there being two visible charged leptons in the final state of the 2-lepton channel. Using these charged leptons two regions can be defined in the 2-lepton channel, the same flavour region which breaks down into the usual  $CR_{low}$ ,  $SR$  and  $CR_{high}$  and the opposite flavour region known as the top  $e\mu$  control region. These regions can be used to get a data driven estimate of the background from top quark processes in the 2-lepton channel, described in section 7.4.1. In the 0- and 1-lepton channel Monte-Carlo predictions and theoretical uncertainties are used instead. A summary of the nuisance parameters that control the  $t\bar{t}$  systematic uncertainties can be found in table 7.12.

Nuisance Parameter	Description	Categories	Value
Normalisation			
norm_ttbar_J2	Floating $t\bar{t}$ norm	$t\bar{t}$ , 0/1-lepton, 2-jet regions	Float
norm_ttbar_J3	Floating $t\bar{t}$ norm	$t\bar{t}$ , 0/1-lepton, 3-jet regions	Float
Acceptance			
SysTTbarNorm_L0	1- to 0-lep $t\bar{t}$ norm extrapolation	0-lepton	8%
Flavour Composition			
SysTTbarbcMeACC	$bc/bb$ acceptance ratio from ME variation	0/1-lepton channels	-
SysTTbarbcPSACC	$bc/bb$ acceptance ratio from PS variation	0/1-lepton channels	-
SysTTbarOthMeACC	Oth/ $bb$ acceptance ratio from ME variation	0/1-lepton channels	-
SysTTbarOthPSACC	Oth/ $bb$ acceptance ratio from PS variation	0/1-lepton channels	-
Shape			
SysTTbarPtV_J2	0+1-lep $p_T^V$ shape	0/1-lepton channels, 2-jet regions	-
SysTTbarPtV_J3	0+1-lep $p_T^V$ shape	0/1-lepton channels, 3-jet regions	-
SysBDTr_ttbar_ME_J2	POWHEG to MADGRAPH multi-variate shape	0/1-lepton channels, 2-jet regions	-
SysBDTr_ttbar_ME_J3	POWHEG to MADGRAPH multi-variate shape	0/1-lepton channels, 3-jet regions	-
SysBDTr_ttbar_PS_LO	PYTHIA8 to HERWIG7 multi-variate shape	0-lepton channel	-
SysBDTr_ttbar_PS_Bmin150_L1	PYTHIA8 to HERWIG7 multi-variate shape	1-lepton channel, $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$	-
SysBDTr_ttbar_PS	PYTHIA8 to HERWIG7 multi-variate shape	1-lepton channel, rest	-

Table 7.12: A summary of systematic uncertainties of  $t\bar{t}$  events. Uncertainties are listed by the nuisance parameter that controls them in the profile-likelihood fit, a short description is provided for each and the categories in which the uncertainty are applied are shown. If a prior is calculated it's value is listed. All uncertainties in this table except those under normalisation are allowed to affect both the yield and shape of distributions.

### Nominal and alternative predictions

The nominal Monte-Carlo prediction for the  $t\bar{t}$  process is POWHEG + PYTHIA 8 with the POWHEG NLO matrix element (ME) generator [135, 136] interfaced to PYTHIA 8 [137] using the A14 tune [138] to model parton showering, hadronisation,

the underlying event, and multiple parton interactions. The NNPDF3.0 (NLO) and NNPDF2.3 parton distribution function sets are used in the ME calculation and parton showering respectively [139].

Details of the samples making up the nominal  $t\bar{t}$  prediction can be found in table B.4 these include the high and low radiation variations called RadHi and RadLo respectively, the samples making up the alternative prediction can be found in table B.10. Samples are simulated with POWHEG + HERWIG 7 to vary the parton shower model and MADGRAPH 5\_AMC@NLO+PYTHIA 8 to vary the matrix element calculation.

### Cross section

The  $t\bar{t}$  cross section relevant to this analysis is calculated for a top quark mass of 172.5 GeV and is  $831.76^{+40}_{-46}$  pb. The calculation is at NNLO in QCD including re-summation of NNLL soft gluon terms with TOP++2.0 [140–146]. The PDF and  $\alpha_S$  uncertainties are calculated using the PDF4LHC prescription [147] with the MSTW2008 68% CL NNLO [148, 149], CT10 NNLO [150, 151] and NNPDF2.3 5f FFN [129] PDF sets, added in quadrature to the scale uncertainty.

### Normalisation and Acceptance

In all fit regions the normalisation of the  $t\bar{t}$  background is kept as a floating normalisation in 2– and 3–jet categories independently. The nuisance parameters relevant to this normalisation are `norm_ttbar_J2` and `norm_ttbar_J3` respectively. In terms of acceptance uncertainties, most of the potential migrations such as migration between analysis regions and  $p_T^V$  bins are covered by the shape uncertainties. There is one nuisance parameter, `SysTTbarNorm_L0` which is implemented to cover migrations between the 1– and 0– lepton channels, it is implemented in the combination of the 2– and 3–jet categories, applied to events in the 0–lepton channel and has a prior calculated using the usual formula in equation 7.1.

## Flavour Composition

Flavour composition uncertainties are derived using ratios of flavour sub-components as for other backgrounds, however given that only the  $bb$  and  $bc$  sub-components are non-negligible —so— the remainder of sub-components are combined into “Oth” name

**Where is this “Oth” name coming from?** Oth only with  $bb$  on the bottom of the ratio as before. Given that neither the matrix element nor parton shower variation dominates the uncertainty heavily, nuisance parameters are implemented for each in turn, in total the nuisance parameters controlling flavour composition uncertainty are SysTTbarbcMeACC, SysTTbarbcPSACC,

SysTTbar0thMeACC, SysTTbar0thPSACC.

## Shape Uncertainties

The  $(N - 1)$ -dimensional parametrisation is used to estimate the shape uncertainties on  $t\bar{t}$  events. As in the  $W + \text{jets}$  the variable that is factorised out of the multi-dimensional parametrisation is  $p_T^V$  which is chosen for the same reasons as mentioned before. The shape systematic is not as correlated over as many of the analysis bins as in the  $W + \text{jets}$  case. The  $p_T^V$  shape uncertainty is correlated across 0– and 1-lepton channels but decorrelated in jet multiplicity as SysTTbarPtV\_J2 and SysTTbarPtV\_J3 for the 2– and 3-jet categories respectively. The  $p_T^V$  shapes used are coming from the largest variation when considering all samples which turns out to be the comparison of POWHEG + PYTHIA 8 and MADGRAPH 5 \_AMC@NLO + PYTHIA 8. The BDTr shape uncertainty is not clearly dominated by one comparison of generators and therefore multiple nuisance parameters enter to encapsulate all the potential sources of uncertainty. Differences due to the matrix element calculation are correlated in leptonic channel but decorrelated in 2– and 3-jet categories as SysBDTr\_ttbar\_ME\_J2 and SysBDTr\_ttbar\_ME\_J3 respectively.

For the parton shower variation the uncertainties are correlated in jet multiplicity but decorrelated in leptonic channel and also in the 1-lepton channel decorrelated in  $p_T^V$  bin, the nuisance parameters in question are SysBDTr\_ttbar\_PS\_L0,

Could talk about correlation and decorrelation here.

`SysBDTr_ttbar_PS_Bmin150_L1`, and `SysBDTr_ttbar_PS`.

### 7.4.1 Data Driven Estimation

As mentioned in section 6.1.1 the top  $e\mu$  control region is constructed and studied in the analysis in order to get a handle on top quark backgrounds in the 2-lepton channel. Due to the flavour symmetry of the  $t\bar{t}$  and

formatting

$Wt$

processes the shape of the top quark backgrounds in the top  $e\mu$  control region and the SR ought to be the same. In order to correct for any differences in normalisation a scale factor  $\alpha$  is derived as follows

$$N_{\text{top, data}}^{\text{SR}} = \frac{N_{\text{top, MC}}^{\text{SR}}}{N_{\text{top, MC}}^{\text{CR}}} \times N_{\text{top, data}}^{\text{CR}} = \alpha \times N_{\text{top, data}}^{\text{CR}}. \quad (7.6)$$

Once that factor is applied then the data from the top  $e\mu$  control region can be used as a template for top quark backgrounds in the SR.

Using the data driven estimate of the background eliminates the need to apply the experimental systematics detailed in section 7.2. It also means we don't have to consider theoretical modelling uncertainties associated with any Monte-Carlo generator. Note that systematic uncertainties on the Monte-Carlo prediction used to derive the scale factor  $\alpha$  cancel.

do not

part of a series of

In order to check the validity of this approach a data versus prediction check was carried out as part of series of wider checks [119]. This check includes the data from the top  $e\mu$  control region as the prediction for both  $t\bar{t}$  and single top events. The plots showing the comparison can be found in figure 7.10, which is split across two figures. In order to see how the shape is modelled the scale factors controlling normalisation of the other backgrounds have been taken from the previous iteration full stop of the analysis and applied, in this iteration there was no boundary at 250 GeV and therefore  $p_T^V$  bins 75–150 GeV and  $\geq 150$  GeV are shown. The previous iteration also

comma

full stop

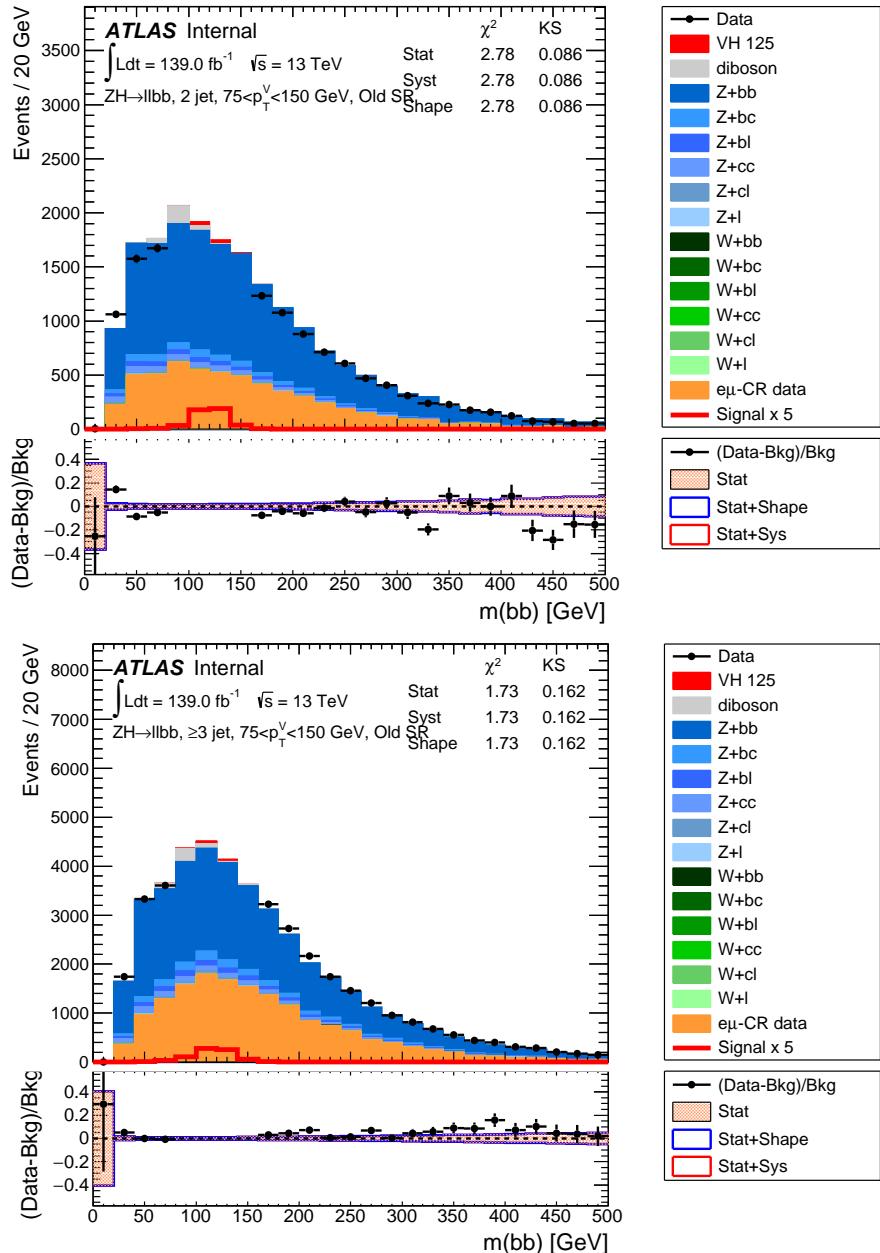


Figure 7.10: Data versus prediction comparison of  $m_{bb}$  distributions used to check how well the top  $e\mu$  control region data models the shape of the  $t\bar{t}$  and single top processes.

did not categorise events into  $\Delta R(b, b)$  based regions and therefore the sum of all the  $\Delta R(b, b)$  regions is shown. It can be seen in the ratio panels that the data and prediction agree well and therefore the prediction from the top  $e\mu$  control region is considered to be accurate.

**Why does the previous iteration inform these plots?  
Just using some previous scale factors pre-fit.**

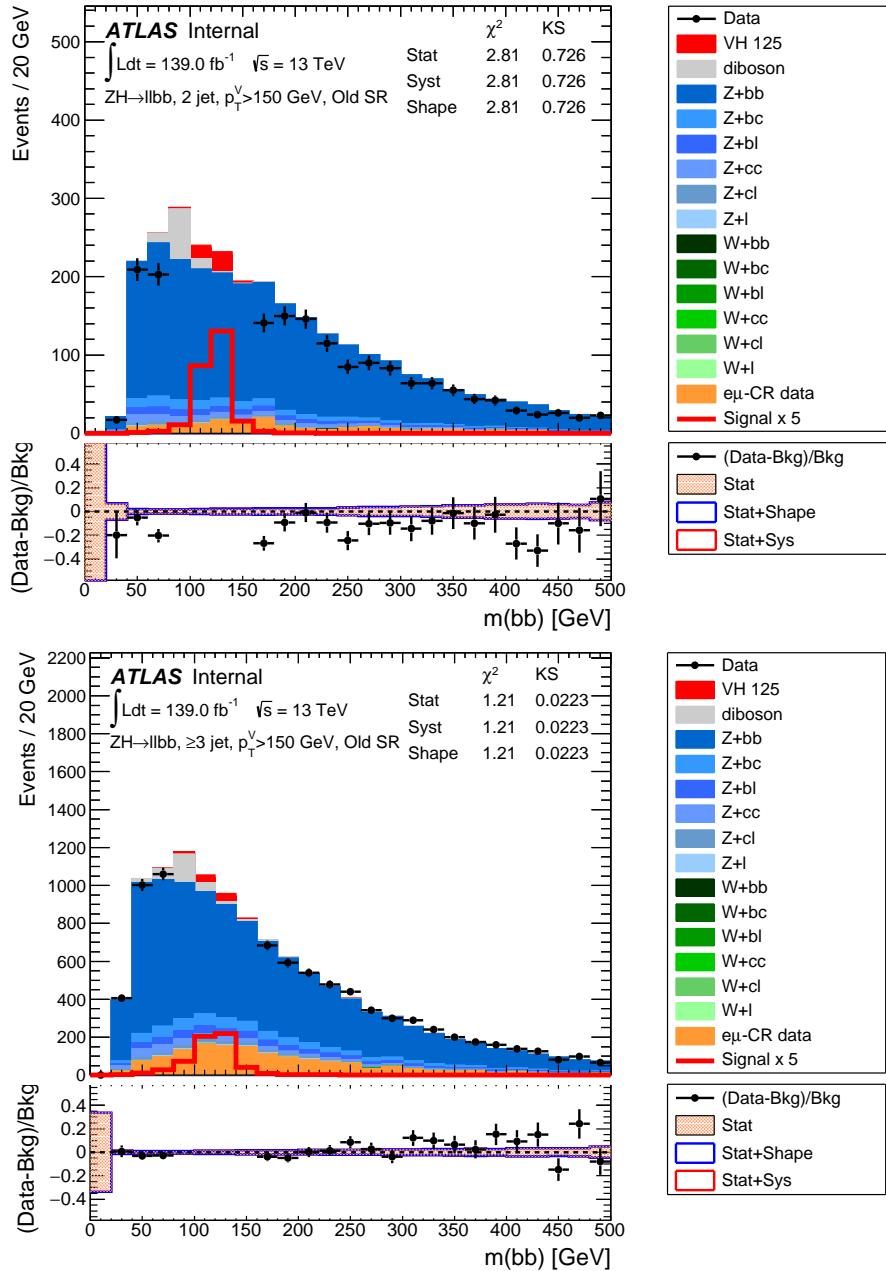


Figure 7.10: Data versus prediction comparison of  $m_{bb}$  distributions used to check how well the top  $e\mu$  control region data models the shape of the  $t\bar{t}$  and single top processes.

## 7.5 Systematic Uncertainties on Sub-Dominant Backgrounds

This section details the systematic uncertainties on sub-dominant backgrounds. None of these background processes dominate in any region of any channel of the analysis and so they are not considered to have as large of an impact on the analysis as the rest of the backgrounds. All of the uncertainties of sub-dominant backgrounds are summarised in table 7.13.

Nuisance Parameter	Description	Categories	Value	Effect
<b>Multi-jet</b>				
SysMJNorm_2J_E1	multi-jet normalization	1-lepton channel, 2-jet, $e$	+200% - 100%	Normalization
SysMJNorm_3J_E1	multi-jet normalization	1-lepton channel, 3-jet, $e$	+100% - 100%	Normalization
SysMJNorm_2J_Mu	multi-jet normalization	1-lepton channel, 2-jet, $\mu$	+47% - 29%	Normalization
SysMJNorm_3J_Mu	multi-jet normalization	1-lepton channel, 3-jet, $\mu$	+100% - 100%	Normalization
SysMJReduced	multi-jet shape change with isolation criteria	1-lepton, $e$ and $\mu$ channel	-	Shape
SysMJSFsCR	multi-jet shape change with $W+HF/Top$ scaling	1-lepton, $e$ and $\mu$ channel	-	Shape
<b>Single top</b>				
stpsNorm	single-top ( $s$ -channel) normalization	all regions	4.6%	Normalization
stptNorm	single-top ( $t$ -channel) normalization	all regions	4.4%	Normalization
stpWtNorm	single-top ( $Wt$ -channel) normalization	all regions	6.2%	Normalization
stptAcc	single-top ( $t$ -channel) acceptance	all regions	17% (2jets), 20% (3jets)	Normalization
StpwWtbbAcc	single-top ( $Wt$ channel $Wt \rightarrow b\bar{b}$ ) acceptance	all regions	54.9% (2jets), 50.7% (3jets)	Normalization
StpwWtthAcc	single-top ( $Wt$ channel $Wt \rightarrow th$ ) acceptance	all regions	24.1% (2jets), 21.2% (3jets)	Normalization
StptPTV	single-top ( $t$ -channel) $p_T^V$ shape	all regions	-	Migration+Shape
StptMBB	single-top ( $t$ -channel) $m_{bb}$ shape	all regions	-	Migration+Shape
StpwWtPTV	single-top ( $Wt$ -channel) $p_T^V$ shape	all regions	-	Migration+Shape
StpwWtMBB	single-top ( $Wt$ -channel) $m_{bb}$ shape	all regions	-	Migration+Shape
<b>Diboson</b>				
SysZZNorm	$ZZ$ normalization	all regions	20%	Normalization
SysWZNorm	$WZ$ normalization	all regions	26%	Normalization
SysWWNorm	$WW$ normalization	all regions	25%	Normalization
SysVZUEPSAcc	$VZ$ normalization	all regions	WZ $\ell\nu b\bar{b}$ : 3.9%, ZZ $\ell\nu b\bar{b}$ : 5.8%, $ZZ\nu\nu b\bar{b}$ : 5.6%	Normalization
SysVZUEPS_J3	UEPS acceptance: overall variation	3(+)-jet regions	WZ $\ell\nu b\bar{b}$ : 10.8%, ZZ $\ell\nu b\bar{b}$ : 3.1%, ZZ $\nu\nu b\bar{b}$ : 7.3%	Normalization
SysVZUEPSResid_L0	UEPS acceptance: 2- to 3(+) jet ratio	0-lepton	11.0%	Normalization
SysVZQCDscale_J2	WZ UEPS acceptance: 0-lep residual	0-lepton	6.0%	Normalization
SysVZQCDscale_J3	2-jet QCD scale acceptance variation	2-jet regions	WZ $\ell\nu b\bar{b}$ : 12.7%, ZZ $\ell\nu b\bar{b}$ : 11.9%, $ZZ\nu\nu b\bar{b}$ : 10.3%	Normalization
SysVZQCDscale_J3	2- to 3(+) jet QCD scale acceptance ratio	2-jet regions	WZ $\ell\nu b\bar{b}$ : -17.7%, ZZ $\ell\nu b\bar{b}$ : -16.4%, ZZ $\nu\nu b\bar{b}$ : -15.2%	Normalization
SysVZQCDscale_JVeto	4 <sup>th</sup> -jet veto QCD scale acceptance variation	3(+)-jet regions	WZ $\ell\nu b\bar{b}$ : 21.2%, ZZ $\ell\nu b\bar{b}$ : 10.1%, ZZ $\nu\nu b\bar{b}$ : 17.4%	Normalization
SysVggZZQCDscale	$gg \rightarrow ZZ$ QCD scale acceptance variation	3-jet, 0- and 1-lepton	WZ $\ell\nu b\bar{b}$ : 19.0%, ZZ $\nu\nu b\bar{b}$ : 18.2%	Normalization
SysVVPTVME	$gg \rightarrow ZZ$ QCD scale acceptance variation	all regions	59%	Migration+Shape
SysVVMBbME	di-boson $p_T^V$ shape	all regions	-	Migration+Shape
SysVVPTVUEPS	di-boson $m_{bb}$ shape variation	all regions	-	Migration+Shape
SysVVMBbbUEPS	di-boson $m_{bb}$ shape variation	all regions	-	Migration+Shape

Table 7.13: A summary of systematic uncertainties on sub-dominant backgrounds of the analysis.

All of these systematic uncertainties are based on the comparison of different Monte-Carlo predictions, apart from in the case of the multi-jet background which is estimated using a data-driven process, details of the nominal predictions for diboson are found in table B.6, and the alternative predictions are found in table B.12. Details of all single-top predictions are found in table B.7.

### 7.5.1 Multi-Jet

Systematic uncertainties on the multi-jet background due to QCD are considered only in the 1-lepton channel. This is due to heavy suppression of multi-jet processes in the 0– and 2–lepton channels. The uncertainties considered are on the normalisation and shape effects only. For normalisation a nuisance parameter is introduced for each category in jet multiplicity and each channel ( $e$  or  $\mu$ ) of the multi-jet process, totalling four parameters. Two parameters control the shape uncertainty which are associated to altering the isolation criteria described in section 5.2 and altering the scaling with respect to the  $W+hf$  and top quark processes.

### 7.5.2 Single Top

Systematic uncertainties on the single top process are applied to all regions of the analysis and include parameters controlling the normalisation, acceptance effects and shape effects. Normalisation is considered for the s-channel, t-channel and  $Wt$ -channel single top decays individually. Acceptances in the 2-jet and 3-jet categories are considered for the t- and  $Wt$ -channels, where the  $Wt$ -channel has a parameter for  $bb$  and Oth flavour sub-components. Shape uncertainties are considered for the  $p_T^V$  and  $m_{bb}$  variables for the t- and  $Wt$ -channels totally four parameters controlling the shapes.

### 7.5.3 Diboson

Systematic uncertainties on the diboson process are considered in all regions of the analysis and include normalisation, acceptance and shape uncertainties. There is one

parameter controlling normalisation for each of the  $WW$ ,  $ZZ$  and  $WZ$  processes. A number of parameters controlling acceptance uncertainty due to the underlying event and parton shower variations are introduced. There are two for the  $VZ$  process controlling the overall variation and the acceptance between categories in jet multiplicity. There is one parameter for each of the  $WZ$  and  $ZZ$  processes that controls the acceptance in the 0-lepton channel. Additionally there are four parameters controlling acceptance uncertainty originating from the chosen QCD strength parameter, there is one for acceptance of  $VZ$  events in the 2-jet category, one to control migration of  $VZ$  events between the 2- and 3-jet categories, one that controls acceptance changes due to the veto based on the total number of jets in any  $VZ$  event, and one that controls the acceptance of  $gg$  initiated  $ZZ$  events. Shape uncertainties are considered, for variations due to the matrix element calculation and for variations due to the underlying events and parton shower calculations, for the  $p_T^V$  and  $m_{bb}$  variables, totalling four parameters.

## 7.6 Systematic Uncertainties on the Signal Process

Predictions of the number of events expected due to the  $VH \rightarrow b\bar{b}$  signal process are generated as follows. Processes initiated by a  $q\bar{q}$  interaction are simulated with POWHEG + MiNLO + PYTHIA 8 [111, 152] at NLO whereas processes initiated by a  $gg$  interaction are simulated with POWHEG + PYTHIA 8 at LO. Both sets of samples have the AZNLO tune [153] applied with the NNPDF3.0 PDF [113] set being used. Alternative predictions are generated using variations of the internal parameters of the nominal generators and also with POWHEG + MiNLO + HERWIG 7. Details of the nominal and alternative predictions are found in tables B.5 and B.11 respectively.

The cross section that samples are normalised to is the best theoretical prediction available which is calculated at NNLO in QCD [154, 155], NLO in EW [156] and

includes higher order contributions to the gluon-induced heavy quark loop entering into the  $ZH$  calculation [157].

A summary of the systematic uncertainties considered on the signal samples can be found in table 7.14.

Nuisance Parameter	Description	Region	Value	Effect
SysTheoryBRbb	BR variation	all regions	1.7%	Normalization
QCDScaleDeltaY_qqVH	Scale uncertainty on qqVH production cross-section	all reg / $qqVH$	0.7%	Normalization
QCDScaleDelta75_qqVH	Scale uncertainty due to $p_T^V$ STXS boundary at 75 GeV	all reg / $qqVH$	-	Migration+Shape
QCDScaleDelta150_qqVH	Scale uncertainty due to $p_T^V$ STXS boundary at 150 GeV	all reg / $qqVH$	-	Migration+Shape
QCDScaleDelta250_qqVH	Scale uncertainty due to $p_T^V$ STXS boundary at 250 GeV	all reg / $qqVH$	-	Migration+Shape
QCDScaleDelta400_qqVH	Scale uncertainty due to $p_T^V$ STXS boundary at 400 GeV	all reg / $qqVH$	-	Migration+Shape
TheoryDelta1_qqVH	Scale uncertainty due to STXS boundary at $N_{jet} - N_{Hjets} == 1$	all reg / $qqVH$	-	Migration+Shape
TheoryDelta2_qqVH	Scale uncertainty due to STXS boundary at $N_{jet} - N_{Hjets} == 2$	all reg / $qqVH$	-	Migration+Shape
QCDScaleDeltaY_ggZH	Scale uncertainty on ggZH production cross-section	all reg / $ggZH$	25%	Normalization
QCDScaleDelta75_ggZH	Scale uncertainty due to $p_T^V$ STXS boundary at 75 GeV	all reg / $ggZH$	-	Migration+Shape
QCDScaleDelta150_ggZH	Scale uncertainty due to $p_T^V$ STXS boundary at 150 GeV	all reg / $ggZH$	-	Migration+Shape
QCDScaleDelta250_ggZH	Scale uncertainty due to $p_T^V$ STXS boundary at 250 GeV	all reg / $ggZH$	-	Migration+Shape
QCDScaleDelta400_ggZH	Scale uncertainty due to $p_T^V$ STXS boundary at 400 GeV	all reg / $ggZH$	-	Migration+Shape
TheoryDelta1_ggVH	Scale uncertainty due to STXS boundary at $N_{jet} - N_{Hjets} == 1$	all reg / $ggVH$	-	Migration+Shape
TheoryDelta2_ggVH	Scale uncertainty due to STXS boundary at $N_{jet} - N_{Hjets} == 2$	all reg / $ggVH$	-	Migration+Shape
TheoryPDF_ [1-30]	30 PDF4LHC uncertainties on predicted cross-section in all STXS bins	all regions	-	Normalization+Shape
TheoryPDFalphas	$\alpha_s$ variation uncertainties on predicted cross-section in all STXS bins	all regions	-	Normalization+Shape
TheoryPSUE_H7	STXS bin acceptance uncertainty comparing PYTHIA8 and HERWIG7	all regions	-	Migration+Shape
TheoryPSUE_AZNLO_Ren	STXS bin acceptance uncertainty due to AZNLO renormalisation scales	all regions	-	Migration+Shape
TheoryPSUE_AZNLO_MPI	STXS bin acceptance uncertainty due to AZNLO MPI tune	all regions	-	Migration+Shape
TheoryPSUE_AZNLO_Var1	STXS bin acceptance uncertainty after using Var1 AZNLO tune	all regions	-	Migration+Shape
TheoryPSUE_AZNLO_Var2	STXS bin acceptance uncertainty after using Var2 AZNLO tune	all regions	-	Migration+Shape
VHNLOEWK	$p_T^V$ shape variation from NLO EW correction	all regions	-	Migration+Shape
VHUEPSMbb	$m_{bb}$ shape variation from UEPS variations	all regions	-	Migration+Shape
VHQCDscaleMbb	$m_{bb}$ shape variation from scale variations	all regions	-	Migration+Shape
VHQCDscaleMbb_ggZH	→ Decorrelate: " _qqVH" and " _ggZH"	-	-	-

Table 7.14: A summary of systematic uncertainties on the  $VH \rightarrow b\bar{b}$  signal process. Most of the uncertainties account for migration between STXS bins.

The majority of these uncertainties account for migration between STXS bins (which are defined in  $p_T^V$ ) due to the QCD scale uncertainty. Parameters are decorrelated for each STXS bin boundary (75, 150, 250, 400) GeV and for the  $q\bar{q}$  and  $gg$  initiated processes. Similarly there are parameters that control for migration at the relevant  $N_{\text{jet}} - N_{H,\text{jet}}$  boundaries that are decorrelated in the same way, these boundaries come from the separations in the STXS scheme based on the number of additional jets in an event beyond the two  $b$ -jets coming from the Higgs boson, as per the previous STXS measurement [117]. Additionally there are parameters that control the uncertainty of the overall cross-section of the  $q\bar{q}$  and  $gg$  processes due to the QCD scale uncertainty. More parameters control uncertainties arising from the choice of PDF set which control both shape and normalisation. A single parameter controls for the uncertainty of the branching ratio of the  $H \rightarrow b\bar{b}$  decay. Finally several parameters control  $p_T^V$  and  $m_{bb}$  shape variations due to NLO EW corrections and underlying event and parton shower calculations, with the latter also being decorrelated based on sub-process.

# Chapter 8

## Results

In this chapter the results of the  $VH \rightarrow b\bar{b}$  multi-variate discriminant fit are shown. Other results such as those for the cross-checks of the analysis can be found in [114]. The fit is performed in accordance with the asymptotic approximation [158] for calculating the  $p_0$  values, which are converted to statistical significances expressed in terms of standard deviations. An Asimov dataset conditional on  $\mu = 1$  is generated for use in the above approximation and is used in a fit whose nuisance parameter pulls are compared with those coming from the unconditional fit to the data that is used to generate the final results. As per the prescription of the asymptotic approximation the results will be discussed in the context of a hypothesis test comparing the null hypothesis, the background only model, with the alternative hypothesis, the background plus signal model.

My contributions to the results are as already stated in other chapters plus running the combined, and 2-lepton channel fits to generate and study the results at many key milestones in the publication timeline including the final results in the paper [159].

### 8.1 Nuisance Parameter Pulls

Figure 8.1 shows the nuisance parameter pulls coming from the combined profile-likelihood fit for the  $W+\text{jets}$ ,  $Z+\text{jets}$  and top process backgrounds. The other

nuisance parameter pulls have been omitted but have been studied [114], and the `Sys` has been dropped from the names of the parameters displayed. Observation of the pull on `SysZMbb` shows that it is pulled towards the  $+1\sigma$  value which meets the expectations based on the shapes derived in section 7.3.2. The uncorrelated component of the shape is also pulled in that direction but to a lesser degree this also matches expectations. Several of the nuisance parameters relating to the BDTr shape uncertainties are pulled to either  $+1\sigma$  or  $-1\sigma$  as would be expected, whilst several have not been pulled much which goes against naive expectations. There could be many explanations for this but as the phase space is split up into many different regions limiting the statistical power of the method it is best not to over-interpret. In order to understand the behaviour of the uncertainties, the correlations between nuisance parameter pulls, shown in figure 8.2, must be studied.

this also → “, which”

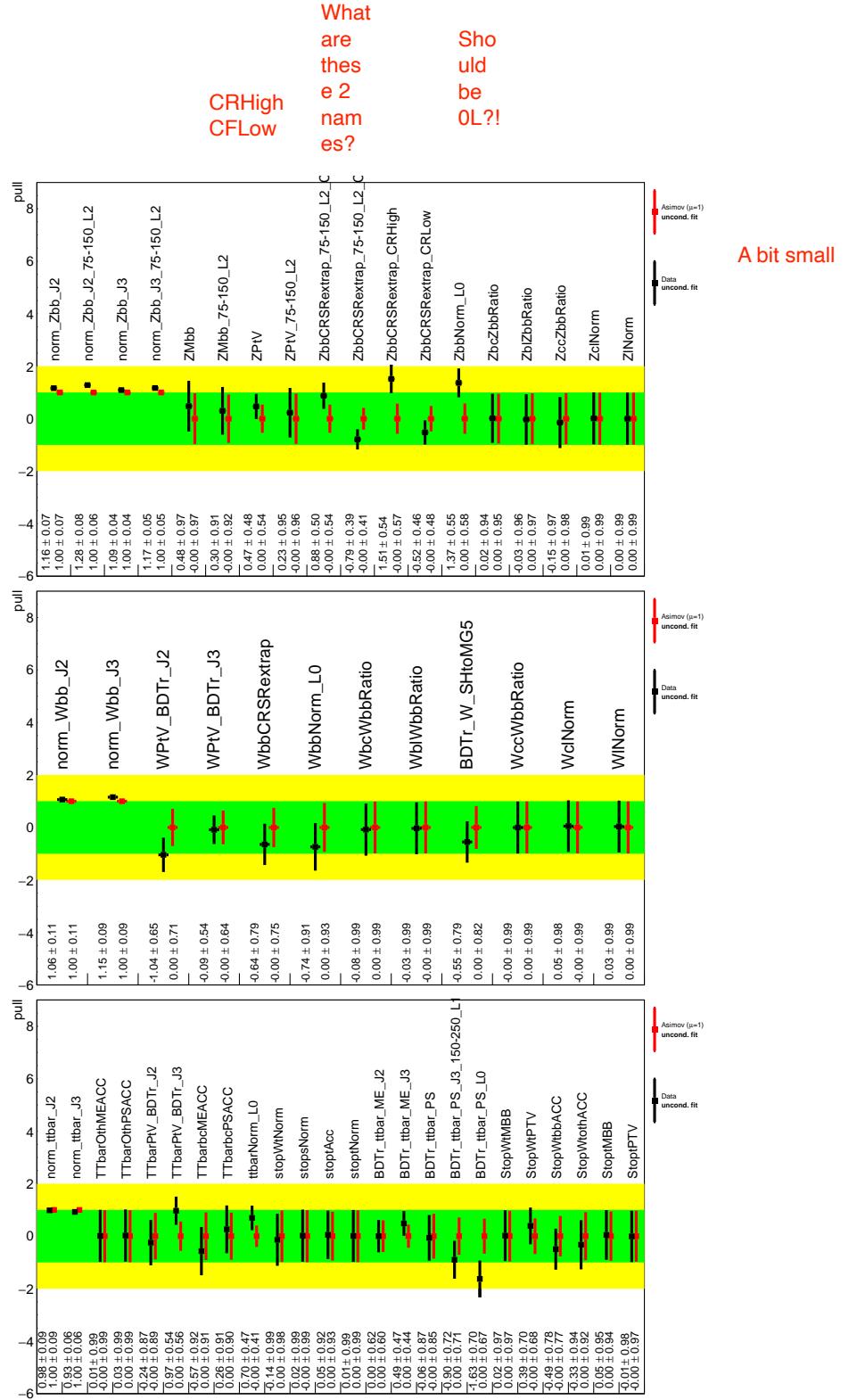


Figure 8.1: Nuisance parameter pulls and free parameter scale factors relating to the  $W+jets$ ,  $Z+jets$  and top process backgrounds of the analysis, where an Asimov dataset conditional on  $\mu = 1$  in red is compared with the data in black.

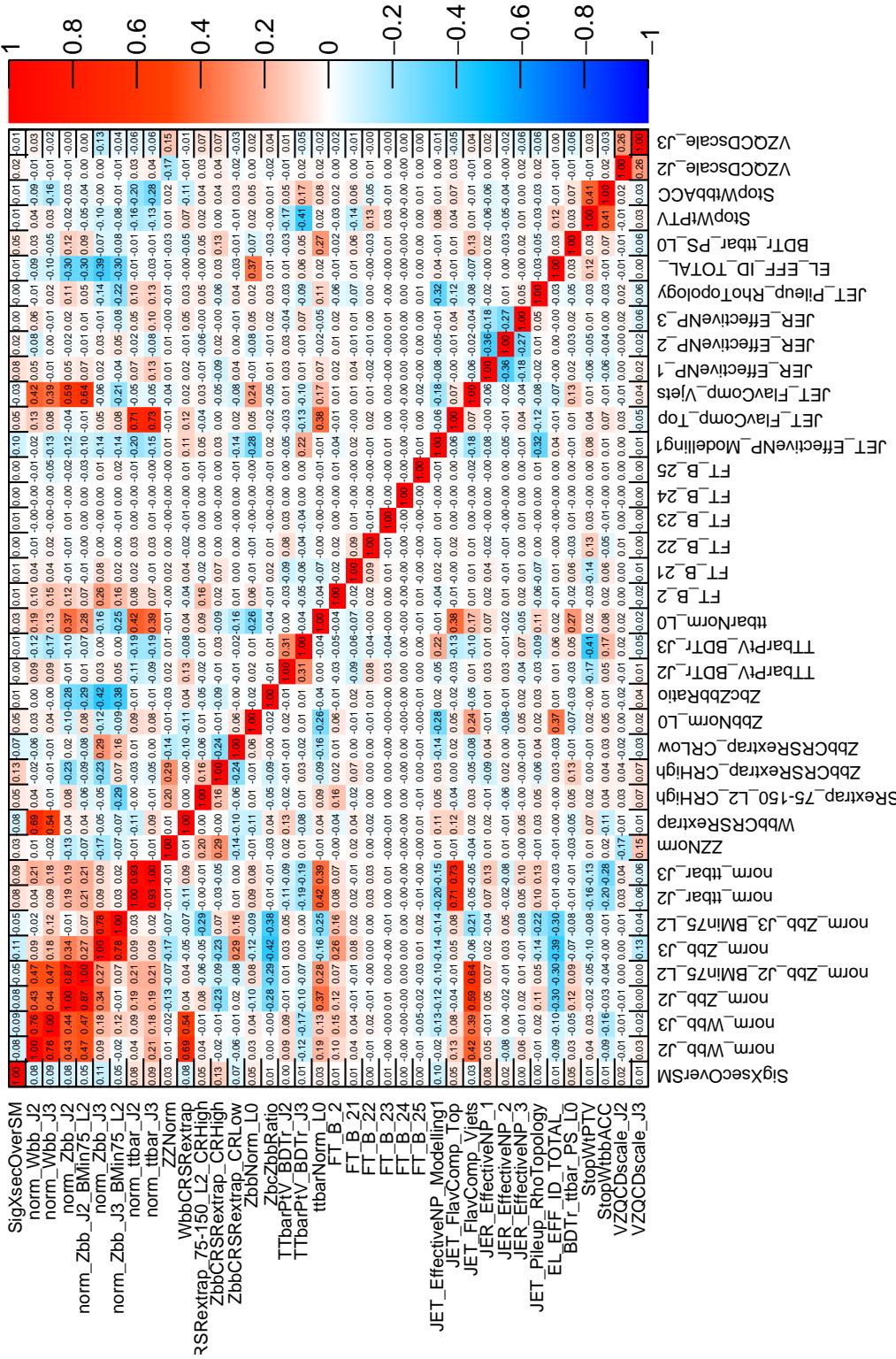


Figure 8.2: Correlations between nuisance parameter pulls where any nuisance parameter with a correlation of at least 25 % is shown and all other are omitted.

What does  
this tell us?

There are no correlations which indicate an unexpected relationship between the relevant quantities.

Is that enough study?

## 8.2 Breakdown and Ranking of Uncertainties

Table 8.1 shows a breakdown of the impact of the systematic uncertainties on the signal strength measurement. It can be seen that the total statistical uncertainty

POI: $\mu = \sigma/\sigma_{\text{SM}}$	Central Value = 1.02	
Nuisance Parameter Set	Signed impact	Un-signed impact
Total	+0.184 / -0.170	$\pm 0.177$
DataStat	+0.116 / -0.114	$\pm 0.115$
Data stat only	+0.109 / -0.107	$\pm 0.108$
Top $e\mu$ CR stat	+0.014 / -0.014	$\pm 0.014$
Floating normalizations	+0.035 / -0.033	$\pm 0.034$
FullSyst	+0.143 / -0.126	$\pm 0.134$
Modelling: $VH$	+0.083 / -0.061	$\pm 0.072$
Modelling: Background	+0.068 / -0.064	$\pm 0.066$
Multi Jet	+0.004 / -0.006	$\pm 0.005$
Modelling: single top	+0.020 / -0.019	$\pm 0.019$
Modelling: $t\bar{t}$	+0.021 / -0.020	$\pm 0.021$
Modelling: $W+jets$	+0.041 / -0.038	$\pm 0.040$
Modelling: $Z+jets$	+0.032 / -0.032	$\pm 0.032$
Modelling: Diboson	+0.034 / -0.031	$\pm 0.033$
MC stat	+0.031 / -0.030	$\pm 0.031$
Experimental Syst	+0.079 / -0.069	$\pm 0.074$
Detector: lepton	+0.004 / -0.005	$\pm 0.004$
Detector: MET	+0.015 / -0.014	$\pm 0.015$
Detector: JET	+0.048 / -0.038	$\pm 0.043$
Detector: FTAG (b-jet)	+0.047 / -0.042	$\pm 0.045$
Detector: FTAG (c-jet)	+0.037 / -0.033	$\pm 0.035$
Detector: FTAG (l-jet)	+0.009 / -0.009	$\pm 0.009$
Detector: FTAG (extrap)	+0.000 / -0.000	$\pm 0.000$
Detector: PU	+0.002 / -0.003	$\pm 0.003$
Luminosity	+0.019 / -0.013	$\pm 0.016$

Table 8.1: This table shows the impact of systematic uncertainties on the signal strength  $\mu$  broken down into categories or nuisance parameter sets. Some sets are further broken down into smaller categories indicated by nested indentation.

Impact than

(DataStat) has a smaller impact than the uncertainties due to modelling and exper-

that → than

imental sources (FullSyst). Considering the systematic uncertainties, those coming from experimental sources have a greater impact than those coming from the modelling of the signal, which in turn have a greater impact than those coming from the modelling of the background. Further scrutiny of uncertainties on the signal strength comes in the form of the ranking shown in figures 8.3 and 8.4. The rankings are

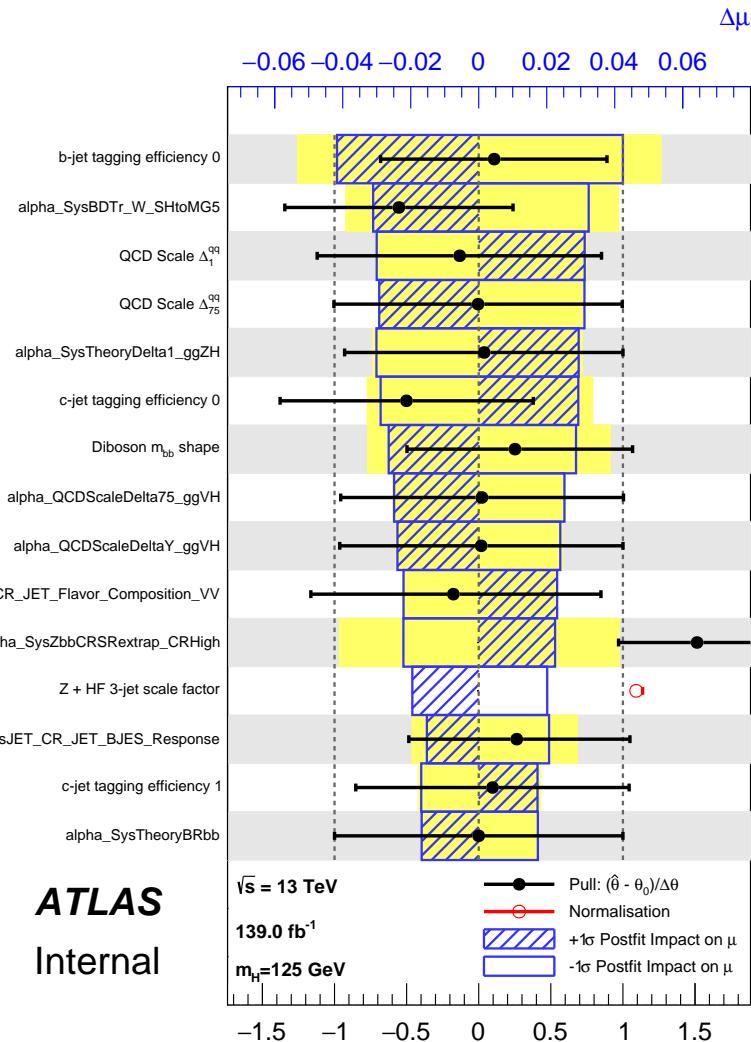


Figure 8.3: The top ranked nuisance parameters are shown according to their impact on the  $VH \rightarrow b\bar{b}$  signal strength as determined by the combined unconditional fit to data.

it's → its

determined by change in  $\mu$  resulting from shifting a given nuisance parameter. The nuisance parameter in question is moved to its up and down variation, based on post-fit uncertainty, represented by the blue boxes that are empty and hatched respectively. The variations based on the pre-fit uncertainty are represented by the

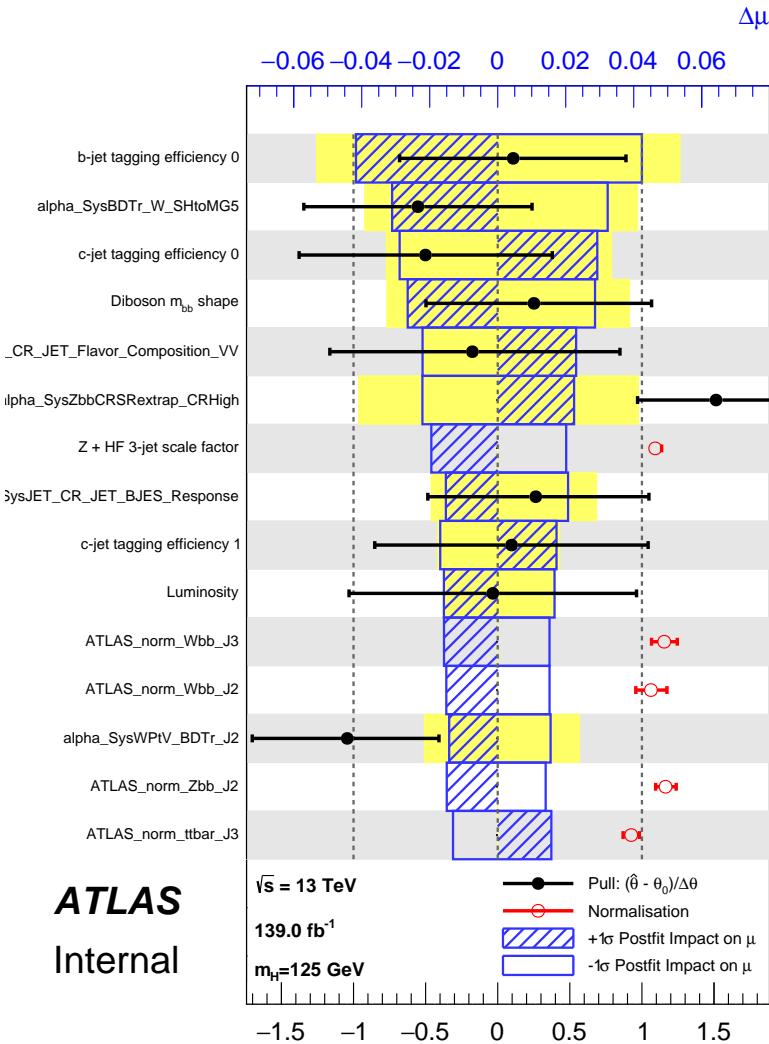


Figure 8.4: The top ranked nuisance parameters are shown, omitting signal uncertainties, according to their impact on the  $VH \rightarrow b\bar{b}$  signal strength as determined by the combined unconditional fit to data.

yellow area.

It can be seen that the ranking is dominated by uncertainties on the signal. The post-fit ranking which omits signal uncertainties is largely dominated by flavour tagging uncertainties and scale factors for normalisation and acceptance between analysis regions. The highest ranked uncertainty in both cases is the *b*-tagging efficiency, this is unsurprising as *b*-tagging plays a crucial role in the analysis selection. Notably the  $Z+jets$  shape uncertainties do not appear in the top sources of uncertainty, naively this is unexpected as the background dominates in a sensitive channel of the analysis, however other modelling uncertainties which are described

with multi-variate techniques have a greater impact.

### 8.3 Post-fit Data Versus Predictions

Figures 8.5, 8.6, 8.7, 8.8, 8.9, 8.10 show the post-fit data versus prediction comparisons for each of the analysis channel, jet multiplicity combinations. Each figure shows the individual analysis regions based on  $\Delta R(b, b)$  region and  $p_T^V$  bin. Across the board good agreement between the data and prediction can be seen, supporting the conclusion that our uncertainty model covers all of the differences between the two datasets and the ~~also~~ supporting the background plus signal hypothesis.

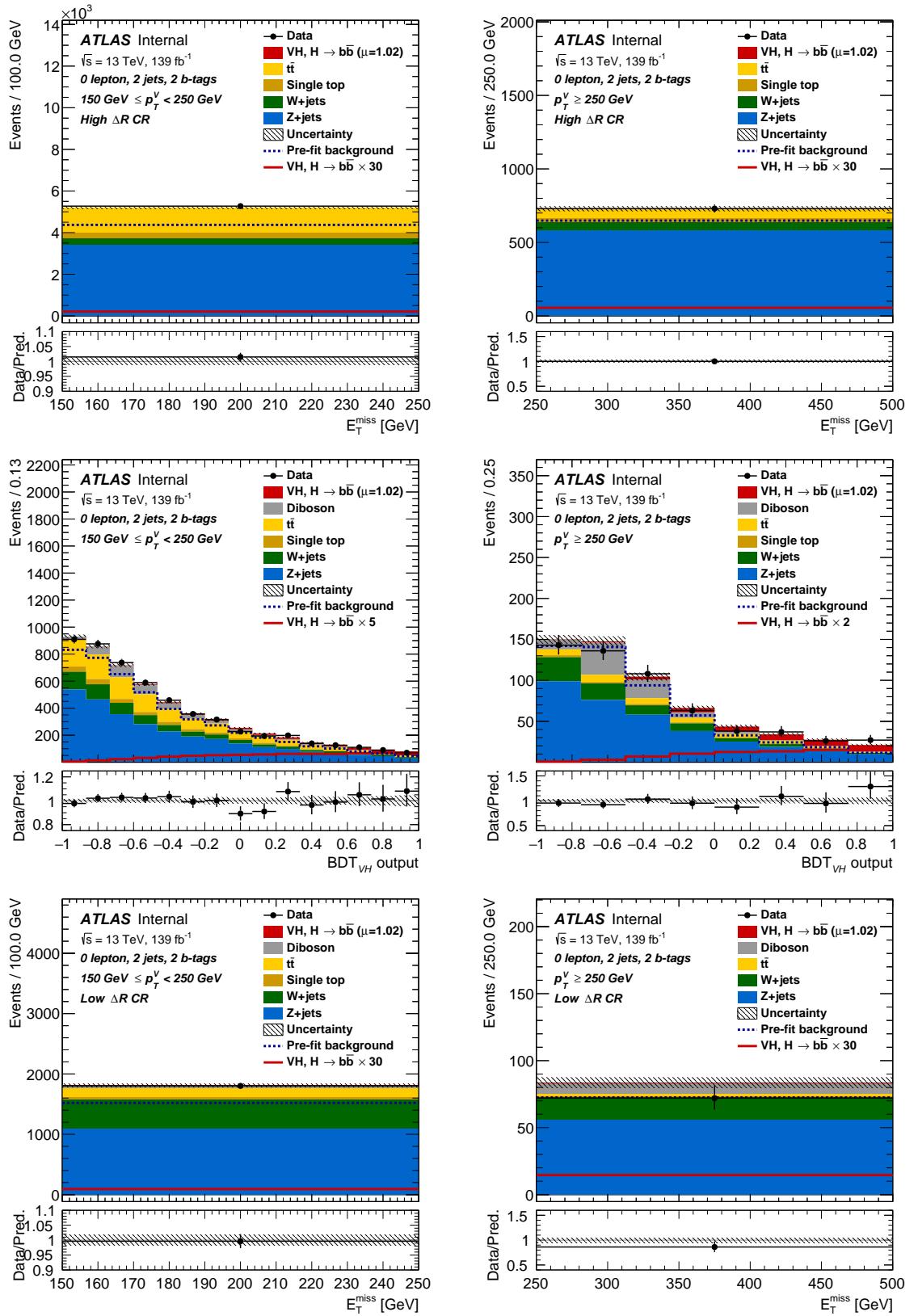


Figure 8.5: Post-fit distributions in the 0-lepton 2-jet channel.

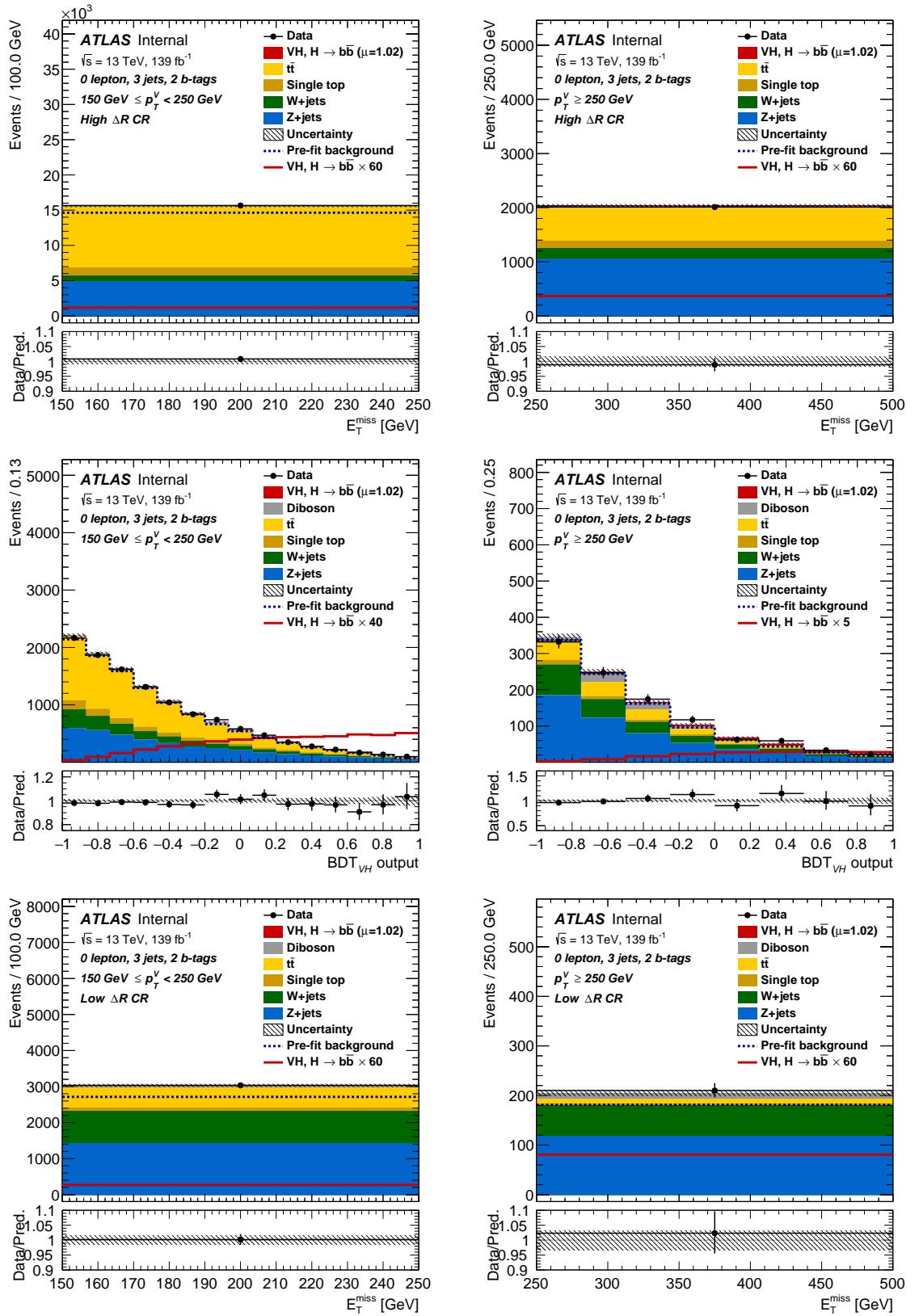


Figure 8.6: Post-fit distributions in the 0-lepton 3-jet channel.

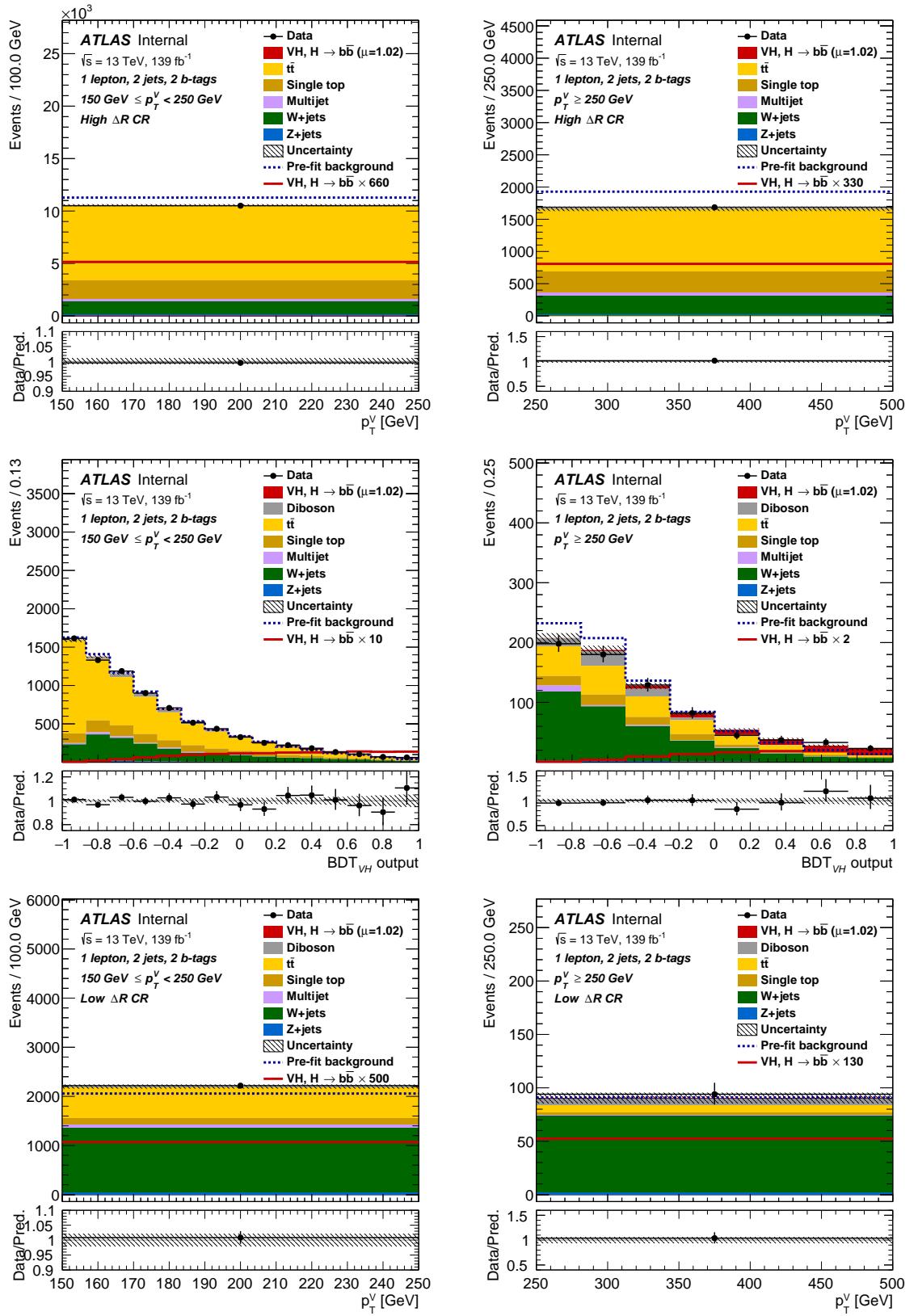


Figure 8.7: Post-fit distributions in the 1-lepton 2-jet channel.

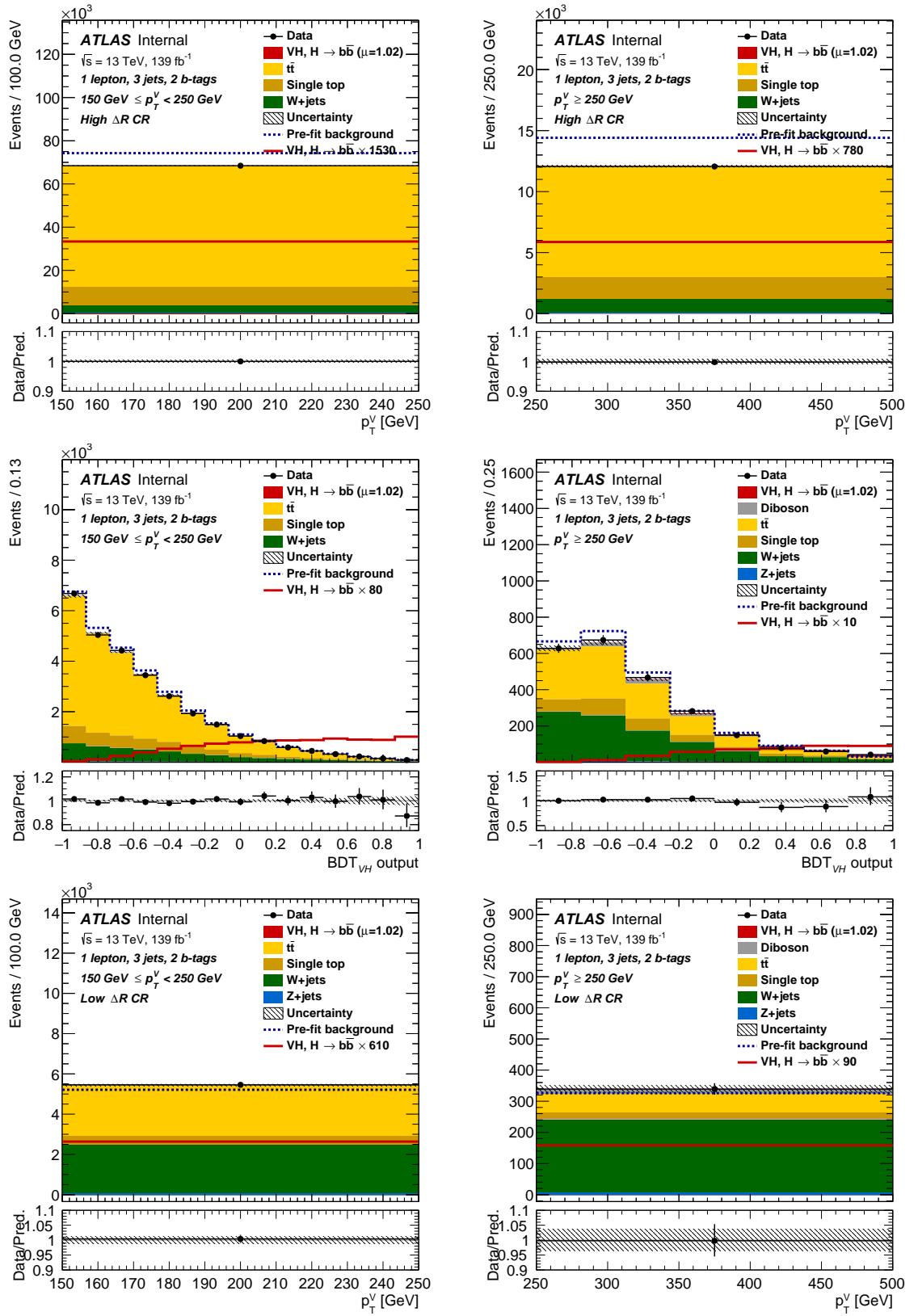


Figure 8.8: Post-fit distributions in the 1-lepton 3-jet channel.

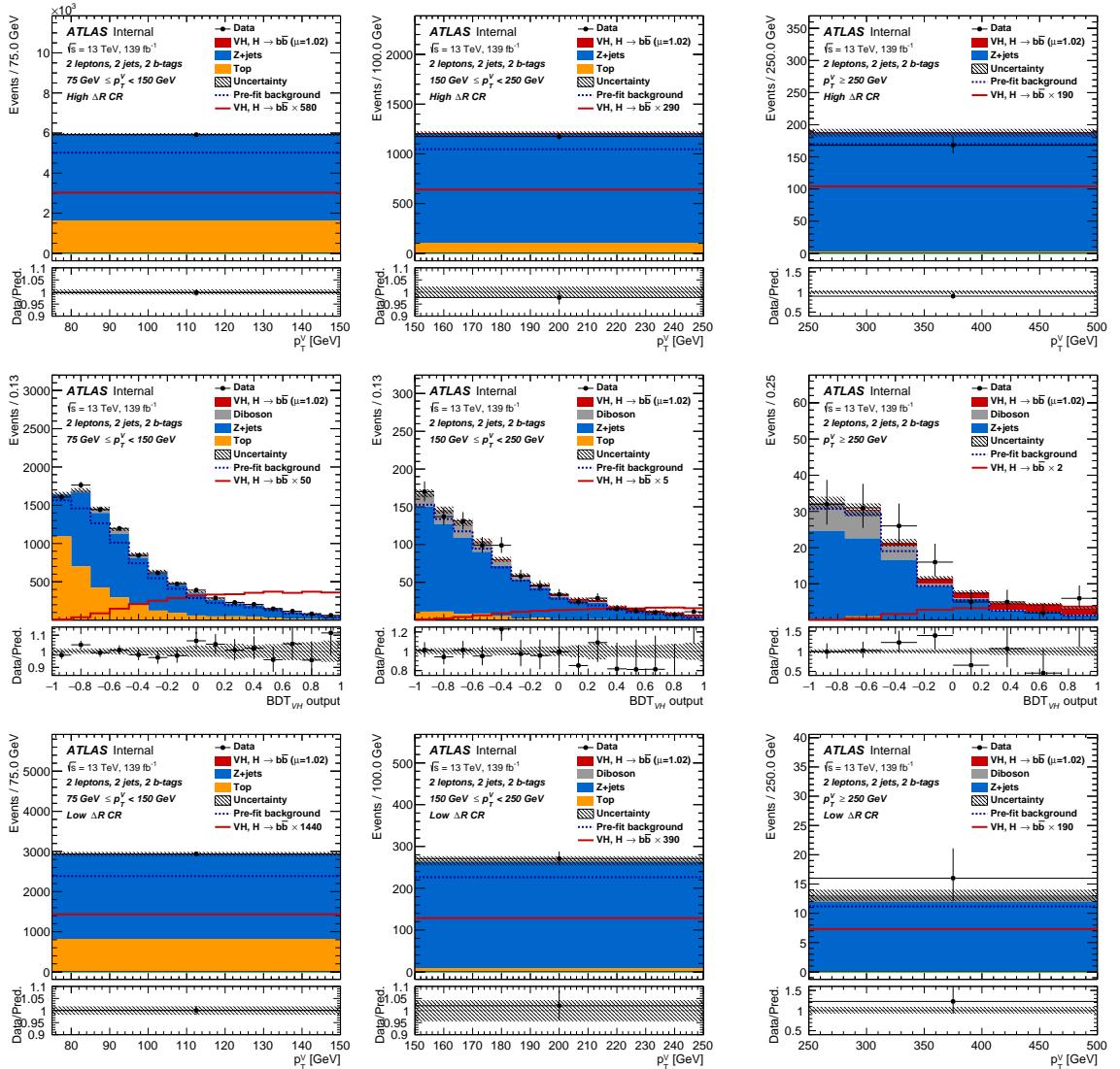
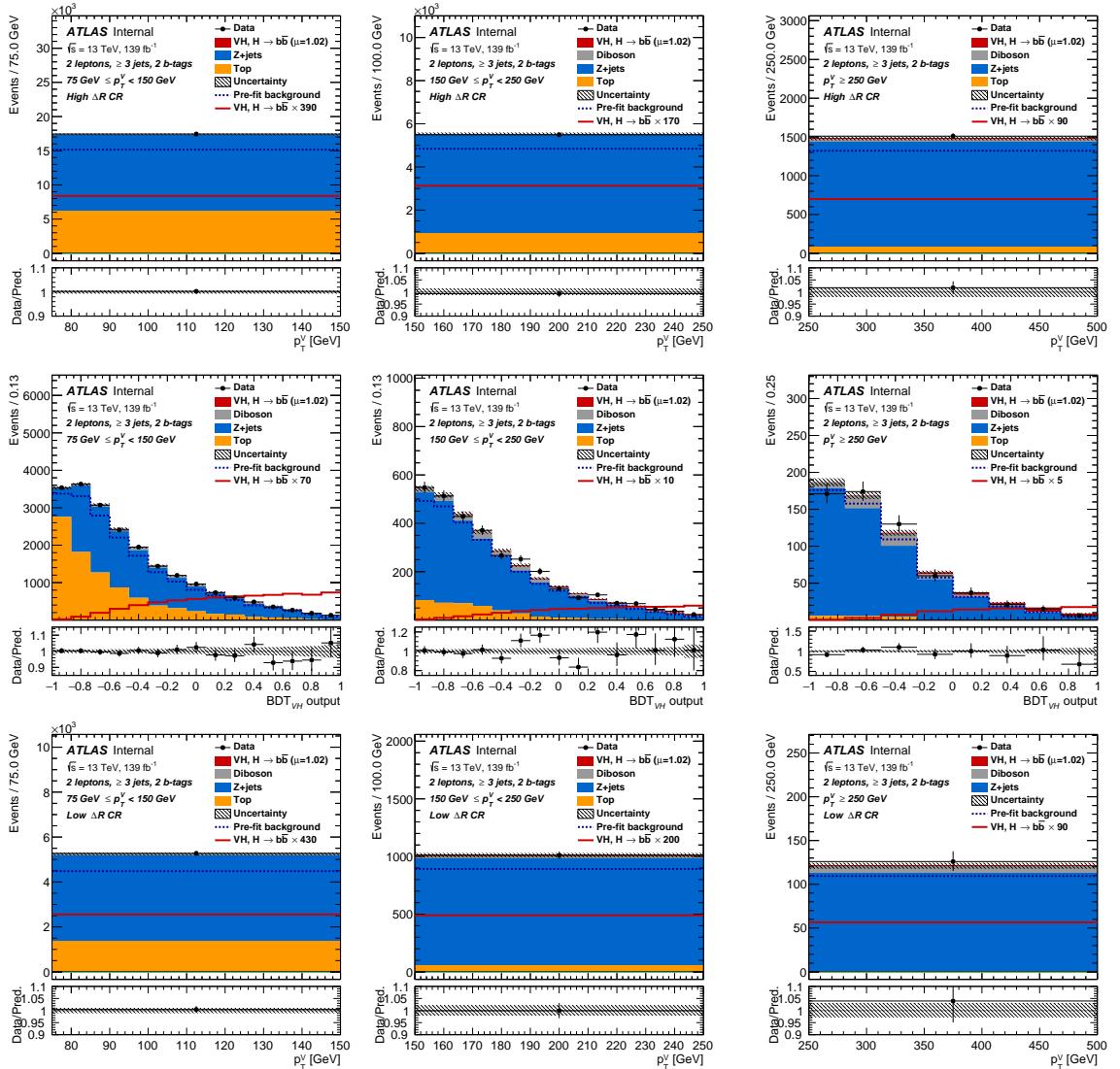


Figure 8.9: Post-fit distributions in the 2-lepton 2-jet channel.

Figure 8.10: Post-fit distributions in the 2-lepton,  $\geq 3$ -jet channel.

## 8.4 Signal Strength and STXS Measurements

In this section the signal strength and STXS measurements will be detailed.

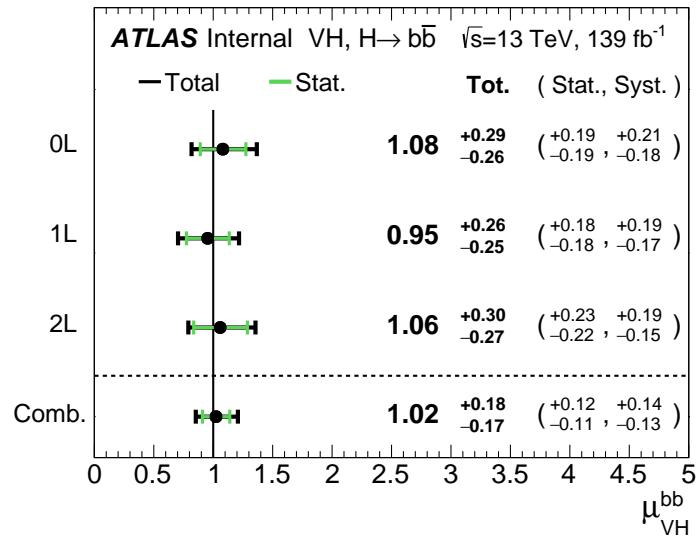
### 8.4.1 VH Signal Strength Measurement

Table 8.2 shows the expected and observed significances for the background plus signal hypothesis broken down by analysis channel. The expected significances

Channel	Expected	Observed
0L	4.1	4.4
1L	4.1	4.0
2L	4.3	4.2
Combined	6.7	6.7

Table 8.2: Statistical significances for the background plus signal hypothesis for an Asimov dataset conditional on  $\mu = 1$ , called expected and for the data, called observed. All values are given standard deviations ( $\sigma$ s). Significances are shown for a fit in each individual channel as well as the combined fit.

were determined using an Asimov dataset conditional on  $\mu = 1$ . All channels are approaching the  $5\sigma$  threshold to claim discovery and the combined fit is well above the discovery threshold, in agreement with the previous result [28]. Figure 8.11 shows the best fit values for the signal strength in each channel and in the combined paradigm. All measurements are compatible with the standard model prediction of  $\mu = 1$ , the composition and values of the uncertainties affecting each channel are similar with the 2–lepton channel being slightly more impacted by statistical rather than systematic uncertainty.



2-lepton

Figure 8.11: Best fit values of the signal strength for the 0–, 1– and 2– channels, as well as the combined (Comb.) measurement. Uncertainties are broken down into either statistical (Stat.) or systematic (Syst.), with those due to scale factors on background processes counted as statistical.

### 8.4.2 $WH$ and $ZH$ Signal Strength Measurements

Table 8.3 shows the expected and observed significances for the background plus signal hypothesis for the  $WH$   $ZH$  and  $VH$  measurements. The expected signif-

Channel	Expected	Observed
$WH$	4.1	4.0
$ZH$	5.1	5.3
$VH$	6.7	6.7

Table 8.3: Statistical significances for the background plus signal hypothesis for an Asimov dataset conditional on  $\mu = 1$ , called expected and for the data, called observed. All values are given standard deviations ( $\sigma$ s). Significances are shown the  $WH$   $ZH$  and  $VH$  measurements.

ances were determined using an Asimov dataset conditional on  $\mu = 1$ . The  $WH$  stand alone measurement is approaching the  $5\sigma$  threshold to claim discovery. The  $ZH$  stand alone measurement has surpassed the threshold and the paper associated with this measurement is first proof of discovery [159]. Figure 8.12 shows the best fit values for the signal strength for each measurement. All measurements are com-

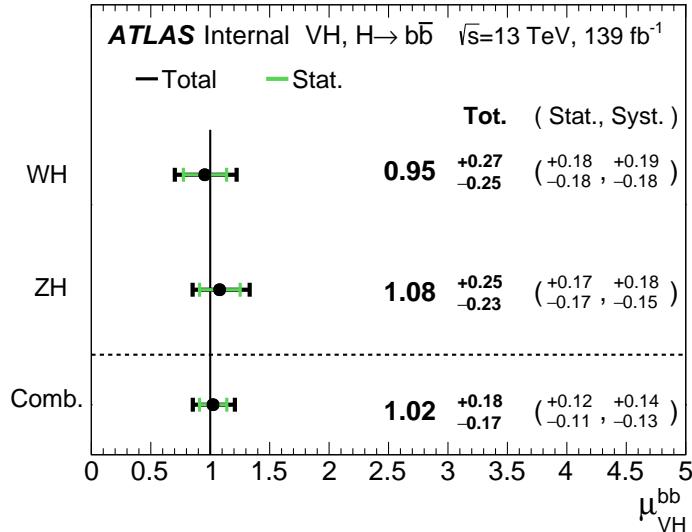


Figure 8.12: Best fit values of the signal strength for the  $WH$   $ZH$  and combined (Comb. $\equiv VH$ ) processes. Uncertainties are broken down into either statistical (Stat.) or systematic (Syst.), with those due to scale factors on background processes counted as statistical.

patible with the standard model prediction of  $\mu = 1$ , the composition and values of the uncertainties affecting each channel are very similar.

### 8.4.3 Simplified Template Cross-section Measurements

Table 8.4 shows the expected and observed significances for the  $VH \rightarrow b\bar{b}$  cross-section measurement in each STXS bin. All bins have an observed statistical significance in line with what was expected. Figure 8.13 shows the best fit values for the cross-section times branching ratio ( $\sigma \times B$ ) in each of the STXS bins.

All measurements are compatible with the standard model predictions and are

Channel	Expected	Observed
$WH, 150 \text{ GeV} < p_T^{\text{miss}} < 250 \text{ GeV}$	2.0	1.6
$WH, p_T^{\text{miss}} > 250 \text{ GeV}$	3.4	3.6
$ZH, 75 \text{ GeV} < p_T^{\text{miss}} < 150 \text{ GeV}$	1.4	1.2
$ZH, 150 \text{ GeV} < p_T^{\text{miss}} < 250 \text{ GeV}$	3.4	3.6
$ZH, p_T^{\text{miss}} > 250 \text{ GeV}$	3.5	3.6
$VH$	6.7	6.7

Table 8.4: Statistical significances on the  $VH \rightarrow b\bar{b}$  cross-section measurement in each STXS bin for an Asimov dataset conditional on  $\mu = 1$ , called expected and for the data, called observed. All values are given standard deviations ( $\sigma$ s). Significances are shown alongside the result from the single parameter combined fit, denoted  $VH$ .

weakly limited by statistical uncertainty. The theoretical uncertainty on the standard model prediction is greater for  $ZH$  bins than for  $WH$  bins. Figure 8.14 shows the correlation between each STXS bin, it can be seen that no two bins are strongly correlated with one another. The lack of correlation is largely due to the analysis categorisations which serve as independent measurements of each STXS bin.

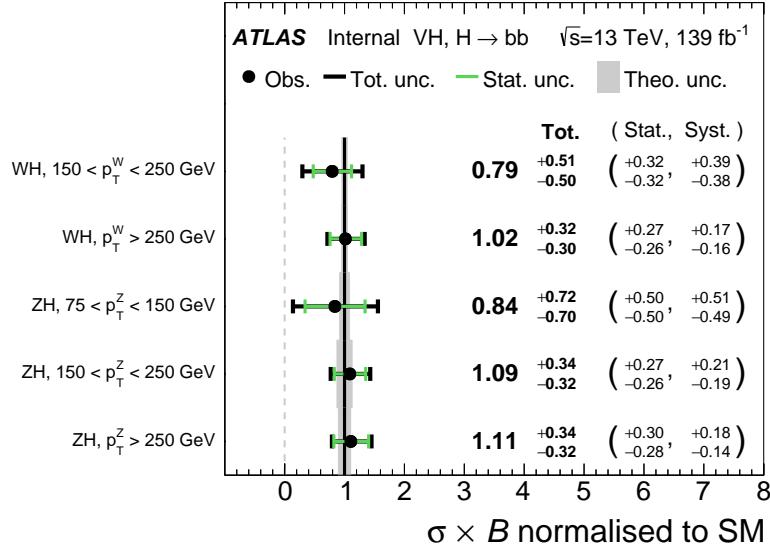


Figure 8.13: Best fit values of the cross-section times branching ratio ( $\sigma \times B$ ) in each of the STXS bins. Uncertainties quoted are broken down by source, either statistical (Stat.) or systematic (Syst.). Uncertainties are broken down into either statistical (Stat.) or systematic (Syst.), with those due to scale factors on background processes counted as statistical. The theoretical uncertainty on the SM prediction is shown in grey.

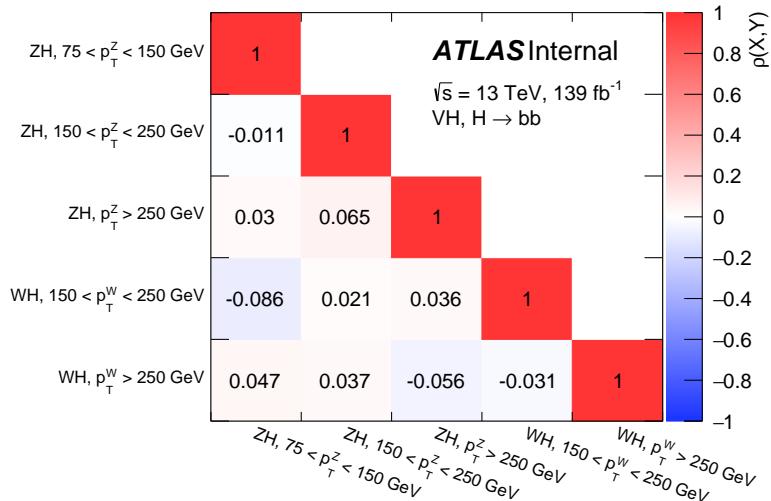


Figure 8.14: Correlations between the measurements in each of the STXS bins of the analysis for the unconditional fit to data.

# Chapter 9

## Conclusion

it's → its

The  $VH \rightarrow b\bar{b}$  analysis has been presented with its many categorisations and carefully chosen techniques to maximise the signal sensitivity and robustness of the analysis. The analysis used  $139 \text{ fb}^{-1}$  of data from proton-proton collisions at a centre of mass energy of  $\sqrt{S} = 13 \text{ TeV}$  recorded by the ATLAS detector and provided by the LHC. Measurements of the  $WH$ ,  $ZH$ , and  $VH$  signal strengths have been made along with cross-sections measurements using the so-called simplified template cross-section paradigm.

The results of the measurements of signal strengths agree with the Standard Model predictions within the uncertainties. Signal strengths are measured to be  $\mu_{WH}^{bb} = 0.95^{+0.27}_{-0.25}$ ,  $\mu_{ZH}^{bb} = 1.08^{+0.25}_{-0.23}$ , and  $\mu_{VH}^{bb} = 1.02^{+0.18}_{-0.17}$  normalised to the Standard Model prediction and for a  $m_H = 125 \text{ GeV}$ .

Why would  
the lowest  
 $p_T$  bin have  
been  
desirable for  
 $WH$ ?

Cross-sections were measured for  $WH$  and  $ZH$ . For both processes measurements were made in a number of different  $p_T^V$  bins, namely the  $75\text{--}150 \text{ GeV}$ ,  $150\text{--}250 \text{ GeV}$  and  $>250 \text{ GeV}$  bins where the first bin includes only a  $ZH$  measurement. A measurement of  $WH$  in the lowest  $p_T^V$  bin would be desirable and could have been achieved with the inclusion of a  $75\text{--}150 \text{ GeV}$  bin in the 1-lepton channel. Such an inclusion was planned but had to be dropped from the analysis due to a poor understanding of the effects of systematic uncertainties in the necessary bin. Results in all of the STXS bins agree with the Standard Model prediction for cross-section times branching ratio within the uncertainties.

significance drops....

Maybe not a  
good sign?

as well as

Modelling of background and signal processes as well of estimation of systematic uncertainties has been carried out using a wide variety of techniques. All checks indicate that the uncertainty model covers all discrepancies between data and the Monte-Carlo predictions. The breakdown of uncertainties shows that the precision of the measurement on the signal strength is limited more by the systematic than the statistical uncertainties. For this reason it is more important than ever to consider how these uncertainties can be better understood and reduced.

Efforts to model the  $V$ +jets and top processes include using a multi-variate technique, which more completely captures the difference between two datasets ~~for~~ than the traditional univariate approach. Where a univariate approach has been used, in the  $Z$ +jets modelling, comparisons have been made to data. These comparisons to data are favourable to comparing two different Monte-Carlo predictions as there is no guarantee that the true nature of the data lies anywhere within the smooth interpolation between the two predictions. Both of these improvements over the previous iteration of the analysis increase confidence in the approach to estimating systematic uncertainties which inherently relies on some prior assumptions and is therefore potentially the area of the analysis in which confidence is most needed.

It is clear from the ranking of uncertainties that an improvement to the procedure for  $b$ -tagging would enhance the precision with which the signal strength could be measured.

## 9.1 Future studies

A combination of the multi-variate and data-driven approaches to modelling would further enhance confidence in the estimation of  $Z$ +jets systematic uncertainties. This would involve using either the  $N$ -dimensional parametrisation or the hybrid  $N - 1$ -dimensional parametrisation, discussed in sections 7.1.2 and 7.1.3, with the nominal prediction being trained against the data in the classifier. In order for this to be possible the gap in the phase space due to the 80–140 GeV veto must be addressed. A BDT is an inappropriate algorithm for training on data with any gaps

~~to be possible,~~

it's → its in the input features due to it's inherently cut based nature, discussed in section 4.1. Neural networks, discussed in section 4.2, are more appropriate for this application as they can smoothly interpolate along a single dimension of the input space. It has been shown in the literature that neural networks can be trained to include a parameter that allows the network to be predictive on datasets that differ from those in the training by some choice of internal parameters [160, 161].

Improvements to the truth tagging strategy discussed in section 5.4.4 could enhance the capabilities of the analysis due to the key role of  $b$ -jets in the process being studied. It has been shown that neural networks can provide a better efficiency map than the current two dimensional maps that are used [162], allowing for better agreement between the truth and direct tagged samples.

Further measurements of the  $VH \rightarrow b\bar{b}$  process are currently underway by members of the ATLAS collaboration. These measurements aim to combine the event selection of this analysis referred to as the resolved analysis (due to the inclusion of two resolved  $b$ -jets in each event), with that of a similar analysis which studies a higher  $p_T$  phase space in which the two  $b$ -jets are merged into a large radius jet, and a measurement of the  $VH \rightarrow c\bar{c}$  process.

# Bibliography

- [1] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters*, 13:321–323, Aug 1964. doi: 10.1103/PhysRevLett.13.321. URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [2] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Physical Review Letters*, 13:508–509, Oct 1964. doi: 10.1103/PhysRevLett.13.508. URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Physical Review Letters*, 13:585–587, Nov 1964. doi: 10.1103/PhysRevLett.13.585. URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.585>.
- [4] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.
- [5] Serguei Chatrchyan et al. (CMS Collaboration). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters*, B716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.
- [6] Sheldon L. Glashow. The renormalizability of vector meson interactions. *Nuclear Physics*, 10:107–117, 1959. doi: 10.1016/0029-5582(59)90196-8.
- [7] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519: 367–377, 1968. doi: 10.1142/9789812795915\_0034.

- [8] Steven Weinberg. A Model of Leptons. *Physical Review Letters*, 19:1264–1266, 1967. doi: 10.1103/PhysRevLett.19.1264.
- [9] Murray Gell-Mann. Symmetries of baryons and mesons. *Phys. Rev.*, 125: 1067–1084, 1962. doi: 10.1103/PhysRev.125.1067.
- [10] Will J. Percival, Shaun Cole, and Eisenstein et al. Measuring the baryon acoustic oscillation scale using the sloan digital sky survey and 2df galaxy redshift survey. *Monthly Notices of the Royal Astronomical Society*, 381(3): 1053–1066, 2007. doi: 10.1111/j.1365-2966.2007.12268.x. URL <http://dx.doi.org/10.1111/j.1365-2966.2007.12268.x>.
- [11] Fabio Iocco, Gianpiero Mangano, Gennaro Miele, et al. Primordial Nucleosynthesis: from precision cosmology to fundamental physics. *Physics Reports*, 472: 1–76, 2009. doi: 10.1016/j.physrep.2009.02.002.
- [12] M. Kowalski et al. Improved Cosmological Constraints from New, Old and Combined Supernova Datasets. *The Astrophysics Journal*, 686:749–778, 2008. doi: 10.1086/589937.
- [13] Richard Massey et al. Dark matter maps reveal cosmic scaffolding. *Nature*, 445:286, 2007. doi: 10.1038/nature05497.
- [14] Douglas Clowe, Marusa Bradac, Anthony H. Gonzalez, et al. A direct empirical proof of the existence of dark matter. *The Astrophysics Journal*, 648: L109–L113, 2006. doi: 10.1086/508162.
- [15] J. Anthony Tyson, Greg P. Kochanski, and Ian P. Dell’Antonio. Detailed mass map of CL0024+1654 from strong lensing. *The Astrophysics Journal*, 498:L107, 1998. doi: 10.1086/311314.
- [16] Lars Bergstrom. Dark Matter Candidates. *New Journal of Physics*, 11:105006, 2009. doi: 10.1088/1367-2630/11/10/105006.
- [17] C. Patrignani et al. Review of Particle Physics. *Chinese Physics*, C40(10): 100001, 2016. doi: 10.1088/1674-1137/40/10/100001.

- [18] F. Zwicky. Republication of: The redshift of extragalactic nebulae. *General Relativity and Gravitation*, 41:207–224, January 2009. doi: 10.1007/s10714-008-0707-4.
- [19] Lars Bergström. Nonbaryonic dark matter: Observational evidence and detection methods. *Reports on Progress in Physics*, 63:793, 2000. doi: 10.1088/0034-4885/63/5/2r3.
- [20] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Physics Reports*, 405:279–390, 2005. doi: 10.1016/j.physrep.2004.08.031.
- [21] Roel Aaij et al. Measurement of the ratio of branching fractions  $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\mu^-\bar{\nu}_\mu)$ . *Physical Review Letters*, 115(11):111803, 2015. doi: 10.1103/PhysRevLett.115.159901, 10.1103/PhysRevLett.115.111803. [Erratum: Physical Review Letters 115, no. 15, 159901 (2015)].
- [22] J. P. Lees et al. Evidence for an excess of  $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$  decays. *Physical Review Letters*, 109:101802, 2012. doi: 10.1103/PhysRevLett.109.101802.
- [23] Thomas Blum, Achim Denig, Ivan Logashenko, et al. The Muon (g-2) Theory Value: Present and Future, 2013. URL <https://arxiv.org/abs/1311.2198>.
- [24] Carl E. Carlson. The Proton Radius Puzzle. *Progress in Particle and Nuclear Physics*, 82:59–77, 2015. doi: 10.1016/j.ppnp.2015.01.002.
- [25] Bernat Capdevila, Andreas Crivellin, and Sébastien Descotes-Genon. Patterns of New Physics in  $b \rightarrow s\ell^+\ell^-$  transitions in the light of recent data. *JHEP*, 01:093, 2018. doi: 10.1007/JHEP01(2018)093.
- [26] John D. Hobbs, Mark S. Neubauer, and Scott Willenbrock. Tests of the standard electroweak model at the energy frontier. *Reviews of Modern Physics*, 84:1477–1526, Oct 2012. doi: 10.1103/RevModPhys.84.1477. URL <https://link.aps.org/doi/10.1103/RevModPhys.84.1477>.

- [27] L. Nodulman. Experimental tests of the standard model. *NATO Science Series C*, 534:147–193, 1999. doi: 10.1007/978-94-011-4689-0\_4.
- [28] ATLAS Collaboration. Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the ATLAS detector. *Physics Letters*, B786:59–86, 2018. doi: 10.1016/j.physletb.2018.09.013.
- [29] Francis Halzen, Alan D. Martin, and Nilotpal Mitra. Quarks and leptons: An introductory course in modern particle physics. *American Journal of Physics*, 53(3):287–287, 1985. doi: 10.1119/1.14146. URL <https://doi.org/10.1119/1.14146>.
- [30] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. doi: 10.1017/CBO9781139525367.
- [31] A. Zannoni. On the Quantization of the Monoatomic Ideal Gas. *eprint arXiv:cond-mat/9912229*, December 1999.
- [32] Paul Adrien Maurice Dirac. On the theory of quantum mechanics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 112(762):661–677, 1926. ISSN 0950-1207. doi: 10.1098/rspa.1926.0133. URL <http://rspa.royalsocietypublishing.org/content/112/762/661>.
- [33] A. Borrelli. *Compendium of Quantum Physics*. 2009.
- [34] Kenneth G. Wilson. Confinement of quarks. *Phys. Rev. D*, 10:2445–2459, Oct 1974. doi: 10.1103/PhysRevD.10.2445. URL <https://link.aps.org/doi/10.1103/PhysRevD.10.2445>.
- [35] Emmy Noether. Invariant variation problems. *Transport Theory and Statistical Physics*, 1(3):186–207, Jan 1971. ISSN 1532-2424. doi: 10.1080/00411457108231446. URL <http://dx.doi.org/10.1080/00411457108231446>.

- [36] Albert Einstein. On the electrodynamics of moving bodies. *Annalen der Physik*, 17:891–921, 1905. doi: 10.1002/andp.200590006. [Annalen Phys.14,194(2005)].
- [37] V. Alan Kostelecký and Neil Russell. Data tables for lorentz and *cpt* violation. *Reviews of Modern Physics*, 83:11–31, Mar 2011. doi: 10.1103/RevModPhys.83.11. URL <https://link.aps.org/doi/10.1103/RevModPhys.83.11>.
- [38] David Mattingly. Modern tests of lorentz invariance. *Living Reviews in Relativity*, 8(1):5, Sep 2005. ISSN 1433-8351. doi: 10.12942/lrr-2005-5. URL <https://doi.org/10.12942/lrr-2005-5>.
- [39] S Liberati. Tests of lorentz invariance: a 2013 update. *Classical and Quantum Gravity*, 30(13):133001, jun 2013. doi: 10.1088/0264-9381/30/13/133001. URL <https://doi.org/10.1088%2F0264-9381%2F30%2F13%2F133001>.
- [40] Clifford M. Will Mark P. Haugan. Modern tests of special relativity. *Physics Today*, 40(5):69, 1987. doi: 10.1063/1.881074. URL <https://doi.org/10.1063/1.881074>.
- [41] Floyd W. Stecker. Limiting superluminal electron and neutrino velocities using the 2010 crab nebula flare and the icecube pev neutrino events. *Astroparticle Physics*, 56:16 – 18, 2014. ISSN 0927-6505. doi: <https://doi.org/10.1016/j.astropartphys.2014.02.007>. URL <http://www.sciencedirect.com/science/article/pii/S092765051400019X>.
- [42] R. Abbasi et al. Search for a lorentz-violating sidereal signal with atmospheric neutrinos in icecube. *Physical Review D*, 82:112003, Dec 2010. doi: 10.1103/PhysRevD.82.112003. URL <https://link.aps.org/doi/10.1103/PhysRevD.82.112003>.
- [43] Enrico Borriello, Sovan Chakraborty, Alessandro Mirizzi, et al. Stringent constraint on neutrino lorentz invariance violation from the two icecube pev neut-

- rinos. *Physical Review D*, 87:116009, Jun 2013. doi: 10.1103/PhysRevD.87.116009. URL <https://link.aps.org/doi/10.1103/PhysRevD.87.116009>.
- [44] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28:1049–1070, Dec 1926. doi: 10.1103/PhysRev.28.1049. URL <https://link.aps.org/doi/10.1103/PhysRev.28.1049>.
- [45] C. S. Wu, E. Ambler, and R. W. Hayward. Experimental test of parity conservation in beta decay. *Physical Review*, 105:1413–1415, Feb 1957. doi: 10.1103/PhysRev.105.1413. URL <https://link.aps.org/doi/10.1103/PhysRev.105.1413>.
- [46] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at  $s^{**}(1/2) = 540\text{-GeV}$ . *Physics Letters*, B122:103–116, 1983. doi: 10.1016/0370-2693(83)91177-2. [,611(1983)].
- [47] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Physics Letters*, B122:476–485, 1983. doi: 10.1016/0370-2693(83)91605-2. [,7.45(1983)].
- [48] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around  $95\text{-GeV}/c^{**2}$  at the CERN SPS Collider. *Physics Letters*, B126:398–410, 1983. doi: 10.1016/0370-2693(83)90188-0. [,7.55(1983)].
- [49] P. Bagnaia et al. Evidence for  $Z_0 \rightarrow e^+ e^-$  at the CERN anti-p p Collider. *Physics Letters*, B129:130–140, 1983. doi: 10.1016/0370-2693(83)90744-X. [,7.69(1983)].
- [50] P. W. Anderson. Plasmons, gauge invariance, and mass. *Physical Review*, 130:439–442, Apr 1963. doi: 10.1103/PhysRev.130.439. URL <https://link.aps.org/doi/10.1103/PhysRev.130.439>.

- [51] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. Technical report, CERN, 2016.
- [52] Oliver Sim Brüning, Paul Collier, and P Lebrun. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. URL <https://cds.cern.ch/record/782076>.
- [53] CERN. Geneva. LEP design report. Technical report, CERN, Geneva, 1984. URL <https://cds.cern.ch/record/102083>. Copies shelved as reports in LEP, PS and SPS libraries.
- [54] CERN. ATLAS: Letter of intent for a general purpose p p experiment at the large hadron collider at CERN. Technical report, CERN, 1992.
- [55] M. Della Negra et al. CMS: letter of intent by the CMS Collaboration for a general purpose detector at LHC. Technical Report CERN-LHCC-92-003. LHCC-I-1, CERN, Geneva, 1992. URL <https://cds.cern.ch/record/290808>. Open presentation to the LHCC 5 November 1992, M. Della Negra/CERN, CMS Spokesman.
- [56] H. Dijkstra, Hans Jürgen Hilke, Tatsuya Nakada, et al. LHCb Letter of Intent, LHCb Collaboration. Technical report, CERN, 1995.
- [57] The ALICE Collaboration. Letter of Intent for A Large Ion Collider Experiment [ALICE]. Technical Report CERN-LHCC-93-016. LHCC-I-4, CERN, Geneva, 1993. URL <https://cds.cern.ch/record/290825>.
- [58] G Giacomelli, A A Faust, and James L Pinfold. A search for highly ionizing particles and slow exotic decays at the LHC using the MOEDAL detectors: letter of intent. Technical Report CERN-LHCC-98-05. LHCC-I-19, CERN, Geneva, Feb 1998. URL <https://cds.cern.ch/record/347906>.
- [59] W Kienzle, , et al. Total cross section: elastic scattering and diffraction dissociation at the LHC. Technical report, CERN, Geneva, 1997. URL <http://cds.cern.ch/record/335255>.

- [60] O Adriani. LHCf Letter of Intent for a p-Pb run. A precise study of forward physics in  $\sqrt{s_N N} = 4.4$  TeV proton-Lead ion collisions with LHCf at the LHC. Technical Report CERN-LHCC-2011-015. LHCC-I-021, CERN, Geneva, Dec 2011. URL <https://cds.cern.ch/record/1404163>.
- [61] Christiane Lefèvre. The CERN accelerator complex. Complexe des accélérateurs du CERN. Dec 2008. URL <https://cds.cern.ch/record/1260465>.
- [62] TeVI Group. Design Report Tevatron 1 project. Technical Report FERMILAB-DESIGN-1984-01, Fermilab, 1984. URL <https://cds.cern.ch/record/1478620>.
- [63] I. Bejar Alonso and L. Rossi. HiLumi LHC Technical Design Report. Technical report, CERN, 2015.
- [64] ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, 2008. URL <http://stacks.iop.org/1748-0221/3/i=08/a=S08003>.
- [65] David Hanser. *Architecture of France*. Greenwood Press, Westport, Conn, 2006. ISBN 978-0-313-31902-0.
- [66] Joao Pequenao. Computer generated image of the whole ATLAS detector. Mar 2008. URL <https://cds.cern.ch/record/1095924>.
- [67] ATLAS Collaboration. ATLAS: Detector and physics performance technical design report. Volume 1. Technical report, CERN, 1999.
- [68] ATLAS Collaboration. ATLAS: Detector and physics performance technical design report. Volume 2. Technical report, CERN, 1999.
- [69] ATLAS Collaboration. The atlas superconducting magnet system at the large hadron collider. *Physica C: Superconductivity*, 468(15):2137 – 2142, 2008. ISSN 0921-4534. doi: <https://doi.org/10.1016/j.physc.2008.05.146>. URL <http://www.sciencedirect.com/science/article/>

- [pii/S0921453408004541](#). Proceedings of the 20th International Symposium on Superconductivity (ISS 2007).
- [70] Joao Pequenao. Computer generated image of the ATLAS inner detector. Mar 2008. URL <https://cds.cern.ch/record/1095926>.
- [71] M Capeans, G Darbo, K Einsweiller, et al. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, CERN, Sep 2010. URL <https://cds.cern.ch/record/1291633>.
- [72] Thomas Paul Charman, Bart Hommels, and Adrian John Bevan. Authorship Qualification Task: ATLAS ITk Sensor Scanner Documentation (QMUL). Technical Report ATL-COM-ITK-2019-018, CERN, Geneva, Jan 2019. URL <https://cds.cern.ch/record/2674404>.
- [73] Ben Brueers. The design and layout of the Phase-II upgrade of the Inner tracker of the ATLAS experiment. Technical report, CERN, Jul 2019. URL <https://cds.cern.ch/record/2681301>.
- [74] John Stakely Keller. The ATLAS ITk Strip Detector System for the Phase-II LHC Upgrade, Feb 2019. URL <https://cds.cern.ch/record/2663172>.
- [75] Joao Pequenao. Computer Generated image of the ATLAS calorimeter. Mar 2008. URL <https://cds.cern.ch/record/1095927>.
- [76] Joao Pequenao. Computer generated image of the ATLAS Muons subsystem. Mar 2008. URL <https://cds.cern.ch/record/1095929>.
- [77] Leo Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996. doi: 10.1023/A:1018054314350. URL <https://doi.org/10.1023/A:1018054314350>.
- [78] Robert E. Schapire. *The Boosting Approach to Machine Learning: An Overview*, pages 149–171. Springer New York, New York, NY, 2003. ISBN 978-0-387-21579-2. doi: 10.1007/978-0-387-21579-2\_9. URL [https://doi.org/10.1007/978-0-387-21579-2\\_9](https://doi.org/10.1007/978-0-387-21579-2_9).

- [79] Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. In *ICML*, pages 148–156, 1996.
- [80] R. A. FISHER. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936. ISSN 2050-1439. doi: 10.1111/j.1469-1809.1936.tb02137.x. URL <http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x>.
- [81] Zipursky SL et al. Lodish H, Berk A. chapter Section 21.1, Overview of Neuron Structure and Function. New York: W. H. Freeman, 2000. URL <https://www.ncbi.nlm.nih.gov/books/NBK21535/>.
- [82] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, pages 65–386, 1958.
- [83] John Platt Patrice Y. Simard, Dave Steinkraus. Best practices for convolutional neural networks applied to visual document analysis. Institute of Electrical and Electronics Engineers, Inc., August 2003. URL <https://www.microsoft.com/en-us/research/publication/best-practices-for-convolutional-neural-networks-applied-to-visual-document-analysis/>.
- [84] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975. doi: 10.1016/0010-4655(75)90039-9.
- [85] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL <http://arxiv.org/abs/1412.6980>.
- [86] Sebastian Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016. URL <http://arxiv.org/abs/1609.04747>.
- [87] M Delmastro et al. Photon identification efficiency measurements with the ATLAS detector using LHC Run 1 data. Technical Report ATL-COM-PHYS-

- 2014-949, CERN, Geneva, Aug 2014. URL <https://cds.cern.ch/record/1747242>.
- [88] ATLAS Collaboration. Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung. Technical Report ATLAS-CONF-2012-047, CERN, Geneva, May 2012. URL <https://cds.cern.ch/record/1449796>.
- [89] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at  $\sqrt{s} = 13$  TeV. Technical Report ATL-PHYS-PUB-2016-015, CERN, Geneva, Aug 2016. URL <https://cds.cern.ch/record/2203514>.
- [90] ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. Technical Report ATLAS-CONF-2016-024, CERN, Geneva, Jun 2016. URL <http://cds.cern.ch/record/2157687>.
- [91] ATLAS Collaboration. Electron and Photon Selection and Identification for Run2 , 2015. URL <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaIdentificationRun2>. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaIdentificationRun2>.
- [92] ATLAS Collaboration. MCPAnalysisGuidelinesMC15, 2015. URL <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesMC15>. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesMC15>.
- [93] ATLAS Collaboration. MuonSelectionTool, 2015. URL <https://twiki.cern.ch/twiki/bin/view/Atlas/MuonSelectionTool>. <https://twiki.cern.ch/twiki/bin/view/Atlas/MuonSelectionTool>.

- [94] ATLAS Collaboration. Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data. *The European Physical Journal*, C74(11):3130, 2014. doi: 10.1140/epjc/s10052-014-3130-x.
- [95] ATLAS Collaboration. Unified Muon Identification Chain, 2015. URL <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MuonsCollection>.  
<https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MuonsCollection>.
- [96] ATLAS Collaboration. Muon reconstruction performance in early  $\sqrt{s}=13$  TeV data. Technical Report ATL-PHYS-PUB-2015-037, CERN, Geneva, Aug 2015. URL <https://cds.cern.ch/record/2047831>.
- [97] ATLAS Collaboration. Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC. Technical Report ATL-PHYS-PUB-2015-045, CERN, Geneva, Nov 2015. URL <https://cds.cern.ch/record/2064383>.
- [98] ATLAS Collaboration. 2015 Tau Recommendations , 2015. URL <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TauPreRecommendations2015>. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TauPreRecommendations2015>.
- [99] ATLAS Collaboration. 2016 Tau Recommendations , 2016. URL <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TauRecommendationsSummer2016>. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TauRecommendationsSummer2016>.
- [100] Luca Ambroz et al. Technical report, CERN, Geneva, Sep 2019. URL <https://cds.cern.ch/record/2690090>.
- [101] W Lampl, S Laplace, D Lelas, et al. Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002.

- ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008. URL <https://cds.cern.ch/record/1099735>.
- [102] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.
- [103] Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector. Technical Report ATLAS-CONF-2015-029, CERN, Geneva, Jul 2015. URL <https://cds.cern.ch/record/2037702>.
- [104] Julia Gonski. Jet Cleaning in 2016 and the Event Level Cleaning Tool. Technical Report ATL-COM-PHYS-2017-982, CERN, Geneva, Jun 2017. URL <https://cds.cern.ch/record/2272136>.
- [105] Alessandro Calandri. Flavour tagging algorithms and performance at the ATLAS experiment. *PoS*, LHCP2016:180, 2016. doi: 10.22323/1.276.0180.
- [106] Jay Theodore Cremer. Chapter 3 - neutron scatter physics and differential cross sections. In Jay Theodore Cremer, editor, *Advances in Imaging and Electron Physics*, volume 173 of *Advances in Imaging and Electron Physics*, pages 167–274. Elsevier, 2012. doi: <https://doi.org/10.1016/B978-0-12-396969-9.00003-3>. URL <https://www.sciencedirect.com/science/article/pii/B9780123969699000033>.
- [107] ATLAS Collaboration. Performance of jet substructure techniques for large- $R$  jets in proton–proton collisions at  $\sqrt{s} = 7\text{ TeV}$  using the ATLAS detector. *JHEP*, 09:076, 2013. doi: 10.1007/JHEP09(2013)076.
- [108] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The catchment area of jets. *Journal of High Energy Physics*, 2008(04):005–005, apr 2008. doi: 10.1088/1126-6708/2008/04/005. URL <https://doi.org/10.1088/1126-6708/2008/04/005>.
- [109] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Physics Letters B*, 659(1):119–126, 2008. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physlettb.2007.09.051>.

- doi.org/10.1016/j.physletb.2007.09.077. URL <https://www.sciencedirect.com/science/article/pii/S0370269307011094>.
- [110] VHbb analysis group. Object selections for SM Higgs boson produced in association with a vector boson in which  $H \rightarrow b\bar{b}$  and V decays leptonically with Run-2 data: Object support note for VH(bb) 2015+2016+2017 dataset publication. Technical Report ATL-COM-PHYS-2018-517, CERN, Geneva, May 2018. URL <https://cds.cern.ch/record/2317182>. This is a support note for the VH(bb) SM publication using the 2015+2016+2017 datasets.
- [111] Gionata Luisoni, Paolo Nason, Carlo Oleari, et al.  $H_W \pm/hz + 0$  and 1 jet at nlo with the powheg box interfaced to gosam and their merging within minlo. *Journal of High Energy Physics*, 2013(10):83, 2013. ISSN 1029-8479. doi: 10.1007/JHEP10(2013)083. URL [http://dx.doi.org/10.1007/JHEP10\(2013\)083](http://dx.doi.org/10.1007/JHEP10(2013)083).
- [112] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Computer Physics Communications*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036.
- [113] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015. doi: 10.1007/JHEP04(2015)040.
- [114] Konie Al Khoury et al. Measurement of  $VH, H \rightarrow b\bar{b}$  with the ATLAS detector. Technical report, CERN, Geneva, Oct 2019. URL <https://cds.cern.ch/record/2692011>.
- [115] A. Hoecker et al. TMVA - Toolkit for Multivariate Data Analysis. *ArXiv Physics e-prints*, March 2007.
- [116] Rene Brun and Fons Rademakers. Root - an object oriented data analysis framework. In *AIHENP'96 Workshop, Lausane*, volume 389, pages 81–86, 1996.

- [117] ATLAS Collaboration. Measurement of  $v h$ ,  $H \rightarrow b\bar{b}$  production as a function of the vector-boson transverse momentum in 13 tev pp collisions with the atlas detector. *Journal of High Energy Physics*, 2019(5), May 2019. ISSN 1029-8479. doi: 10.1007/jhep05(2019)141. URL [http://dx.doi.org/10.1007/JHEP05\(2019\)141](http://dx.doi.org/10.1007/JHEP05(2019)141).
- [118] Kyle Cranmer, Juan Pavez, and Gilles Louppe. Approximating likelihood ratios with calibrated discriminative classifiers, 2016.
- [119] Konie Al Khoury et al. Data/Monte Carlo Modelling Studies for the Standard Model  $VH, H \rightarrow b\bar{b}$  Analysis. Technical report, CERN, Geneva, Sep 2019. URL <https://cds.cern.ch/record/2690042>.
- [120] ATLAS Collaboration. Luminosity Determination in pp Collisions at  $\sqrt{s} = 8$  TeV using the ATLAS Detector at the LHC. *The European Physical Journal C*, 76:653, 2016. doi: 10.1140/epjc/s10052-016-4466-1.
- [121] ATLAS Collaboration. 2012 preliminary result for luminosity uncertainty to be used for moriod 2013. <https://twiki.cern.ch/twiki/bin/viewauth/Atlas/LuminosityForPhysics>, accessed 5th April 2021.
- [122] ATLAS Collaboration. ExtendedPileupReweighting, 2015. URL <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ExtendedPileupReweighting>. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ExtendedPileupReweighting>.
- [123] ATLAS Collaboration. Electron Efficiencies for Run2 (scale factors and uncertainties) , 2015. URL <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ElectronEfficiencyRun2>. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronEfficiencyRun2>.
- [124] ATLAS Collaboration. Electron and Photon Calibration for Run2 (energy scale and resolution) , 2015. URL <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/PhotonCalibrationRun2>.

- <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaCalibrationRun2>. ht-
- [125] Ian T. Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016. doi: 10.1098/rsta.2015.0202. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2015.0202>.
- [126] ATLAS Collaboration. Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at  $\sqrt{s} = 13$  TeV. Technical Report ATL-PHYS-PUB-2015-015, CERN, Geneva, Jul 2015. URL <https://cds.cern.ch/record/2037613>.
- [127] Gordon Watts, Frank Filthaut, and Giacinto Piacquadio. Extrapolating Errors for  $b$ -tagging. Technical Report ATL-COM-PHYS-2015-711, CERN, Geneva, Jul 2015. URL <https://cds.cern.ch/record/2034234>. This is for internal information only, no approval to ever be seen outside of ATLAS.
- [128] T Gleisberg, S Höche, F Krauss, et al. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(02):007–007, feb 2009. doi: 10.1088/1126-6708/2009/02/007. URL <https://doi.org/10.1088/1126-6708/2009/02/007>.
- [129] Richard D. Ball et al. Parton distributions with LHC data. *Nuclear Physics*, B867:244–289, 2013. doi: 10.1016/j.nuclphysb.2012.10.003.
- [130] Leif Lonnblad. Correcting the color dipole cascade model with fixed order matrix elements. *JHEP*, 05:046, 2002. doi: 10.1088/1126-6708/2002/05/046.
- [131] Nils Lavesson and Leif Lonnblad. W+jets matrix elements and the dipole cascade. *JHEP*, 07:054, 2005. doi: 10.1088/1126-6708/2005/07/054.

- [132] 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. *Journal of High Energy Physics*, 2014(7):79, 2014. doi: 10.1007/JHEP07(2014)079. URL <http://dx.doi.org/10.1007/JHEP07%282014%29079>.
- [133] J. Butterworth, E. Dobson, U. Klein, et al. Single Boson and Diboson Production Cross Sections in pp Collisions at  $\text{sqrt}s=7$  TeV. Technical Report ATL-COM-PHYS-2010-695, CERN, Geneva, Aug 2010. URL <https://cds.cern.ch/record/1287902>.
- [134] C. Patrignani et al. Review of Particle Physics. *Chinese Physics*, C40(10):100001, 2016. doi: 10.1088/1674-1137/40/10/100001.
- [135] S. Frixione et al. A positive-weight next-to-leading-order monte carlo for heavy flavour hadro-production. *JHEP*, 09:126, 2007. doi: 10.1103/PhysRevD.82.074018. URL <http://link.aps.org/doi/10.1103/PhysRevD.82.074018>.
- [136] P. Nason. A new method for combining nlo qcd with shower monte carlo algorithms. *JHEP*, 11:040, 2004.
- [137] T. Sjöstrand et al. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159, 2015.
- [138] ATLAS Collaboration. ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014. URL <https://cds.cern.ch/record/1966419>.
- [139] ATLAS Collaboration. Studies on top-quark Monte Carlo Modelling for Top 2016. Technical Report ATL-PHYS-PUB-2016-020, CERN, Geneva, Sep 2016. URL <https://cds.cern.ch/record/2216168>.
- [140] M. Beneke, P. Falgari, S. Klein, and C. Schwinn. Hadronic top-quark pair production with {NNLL} threshold resummation. *Nuclear Physics B*, 855(3):695 – 741, 2012. ISSN 0550-3213. doi: <http://dx.doi.org/10.1016/>

- j.nuclphysb.2011.10.021. URL <http://www.sciencedirect.com/science/article/pii/S0550321311005803>.
- [141] Matteo Cacciari, Michał Czakon, Michelangelo Mangano, et al. Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation. *Physics Letters B*, 710(4–5):612 – 622, 2012. ISSN 0370-2693. doi: <http://dx.doi.org/10.1016/j.physletb.2012.03.013>. URL <http://www.sciencedirect.com/science/article/pii/S0370269312002766>.
- [142] Peter Bärnreuther, Michał Czakon, and Alexander Mitov. Percent-level-precision physics at the tevatron: Next-to-next-to-leading order qcd corrections to  $q\bar{q} \rightarrow t\bar{t}+x$ . *Physical Review Letters*, 109:132001, Sep 2012. doi: [10.1103/PhysRevLett.109.132001](https://doi.org/10.1103/PhysRevLett.109.132001). URL <http://link.aps.org/doi/10.1103/PhysRevLett.109.132001>.
- [143] Michal Czakon and Alexander Mitov. Nnlo corrections to top-pair production at hadron colliders: the all-fermionic scattering channels. *Journal of High Energy Physics*, 2012(12):54, 2012. doi: [10.1007/JHEP12\(2012\)054](https://doi.org/10.1007/JHEP12(2012)054). URL <http://dx.doi.org/10.1007/JHEP12%282012%29054>.
- [144] Michal Czakon and Alexander Mitov. Nnlo corrections to top pair production at hadron colliders: the quark-gluon reaction. *Journal of High Energy Physics*, 2013(1):80, 2013. doi: [10.1007/JHEP01\(2013\)080](https://doi.org/10.1007/JHEP01(2013)080). URL <http://dx.doi.org/10.1007/JHEP01%282013%29080>.
- [145] Michał Czakon, Paul Fiedler, and Alexander Mitov. Total top-quark pair-production cross section at hadron colliders through  $\mathcal{O}(\alpha_S^4)$ . *Physical Review Letters*, 110:252004, Jun 2013. doi: [10.1103/PhysRevLett.110.252004](https://doi.org/10.1103/PhysRevLett.110.252004). URL <http://link.aps.org/doi/10.1103/PhysRevLett.110.252004>.
- [146] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Computer Physics Communications*, 185:2930, 2014. doi: [10.1016/j.cpc.2014.06.021](https://doi.org/10.1016/j.cpc.2014.06.021).

- [147] Michiel Botje et al. The PDF4LHC Working Group Interim Recommendations, 2011. URL <https://arxiv.org/abs/1101.0538>.
- [148] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the lhc. *The European Physical Journal C*, 63(2):189–285, 2009. ISSN 1434-6044. doi: 10.1140/epjc/s10052-009-1072-5. URL <http://dx.doi.org/10.1140/epjc/s10052-009-1072-5>.
- [149] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Uncertainties on  $\alpha_s$  in global pdf analyses and implications for predicted hadronic cross sections. *The European Physical Journal C*, 64(4):653–680, 2009. ISSN 1434-6044. doi: 10.1140/epjc/s10052-009-1164-2. URL <http://dx.doi.org/10.1140/epjc/s10052-009-1164-2>.
- [150] Hung-Liang Lai, Marco Guzzi, Joey Huston, et al. New parton distributions for collider physics. *Physical Review D*, 82:074024, Oct 2010. doi: 10.1103/PhysRevD.82.074024. URL <http://link.aps.org/doi/10.1103/PhysRevD.82.074024>.
- [151] Jun Gao, Marco Guzzi, Joey Huston, et al. Ct10 next-to-next-to-leading order global analysis of qcd. *Physical Review D*, 89:033009, Feb 2014. doi: 10.1103/PhysRevD.89.033009. URL <http://link.aps.org/doi/10.1103/PhysRevD.89.033009>.
- [152] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. A brief introduction to {PYTHIA} 8.1. *Computer Physics Communications*, 178(11):852 – 867, 2008. ISSN 0010-4655. doi: <http://dx.doi.org/10.1016/j.cpc.2008.01.036>. URL <http://www.sciencedirect.com/science/article/pii/S0010465508000441>.
- [153] ATLAS Collaboration. Measurement of the  $Z/\gamma^*$  boson transverse momentum distribution in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector. *JHEP*, 09:145, 2014. doi: 10.1007/JHEP09(2014)145.

- [154] Oliver Brein, Abdelhak Djouadi, and Robert Harlander. NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders. *Physics Letters*, B579:149–156, 2004. doi: 10.1016/j.physletb.2003.10.112.
- [155] Oliver Brein, Robert Harlander, Marius Wiesemann, and Tom Zirke. Top-Quark Mediated Effects in Hadronic Higgs-Strahlung. *The European Physical Journal*, C72:1868, 2012. doi: 10.1140/epjc/s10052-012-1868-6.
- [156] Ansgar Denner, Stefan Dittmaier, Stefan Kallweit, and Alexander Mück. Electroweak corrections to Higgs-strahlung off  $W/Z$  bosons at the Tevatron and LHC with HAWK. *JHEP*, 03:075, 2012. doi: 10.1007/JHEP03(2012)075.
- [157] Lukas Altenkamp, Stefan Dittmaier, Robert V. Harlander, Heidi Rzehak, and Tom J. E. Zirke. Gluon-induced Higgs-strahlung at next-to-leading order QCD. *JHEP*, 02:078, 2013. doi: 10.1007/JHEP02(2013)078.
- [158] Glen Cowan et al. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal*, C71:1554, 2011. doi: 10.1140/epjc/s10052-011-1554-0.
- [159] ATLAS Collaboration. Measurements of  $WH$  and  $ZH$  production in the  $H \rightarrow b\bar{b}$  decay channel in  $pp$  collisions at 13 TeV with the ATLAS detector. *Eur. Phys. J. C*, 81:178. 41 p, Jul 2020. doi: 10.1140/epjc/s10052-020-08677-2. URL <https://cds.cern.ch/record/2723187>.
- [160] Pierre Baldi, Kyle Cranmer, Taylor Faucett, Peter Sadowski, and Daniel Whiteson. Parameterized neural networks for high-energy physics. *The European Physical Journal C*, 76(5), Apr 2016. ISSN 1434-6052. doi: 10.1140/epjc/s10052-016-4099-4. URL <http://dx.doi.org/10.1140/epjc/s10052-016-4099-4>.
- [161] Anders Andreassen, Benjamin Nachman, and David Shih. Simulation assisted likelihood-free anomaly detection. *Physical Review D*, 101(9), May 2020. ISSN

2470-0029. doi: 10.1103/physrevd.101.095004. URL <http://dx.doi.org/10.1103/PhysRevD.101.095004>.

- [162] C. Badiali, F. A. Di Bello, G. Frattari, E. Gross, V. Ippolito, M. Kado, and J. Shlomi. Efficiency parameterization with neural networks, 2020.

## **Appendix A**

# **Supplement to Modelling of Systematic Uncertainties**

### **A.1 Extra BDT input distributions**

### **A.2 Prefit distributions**

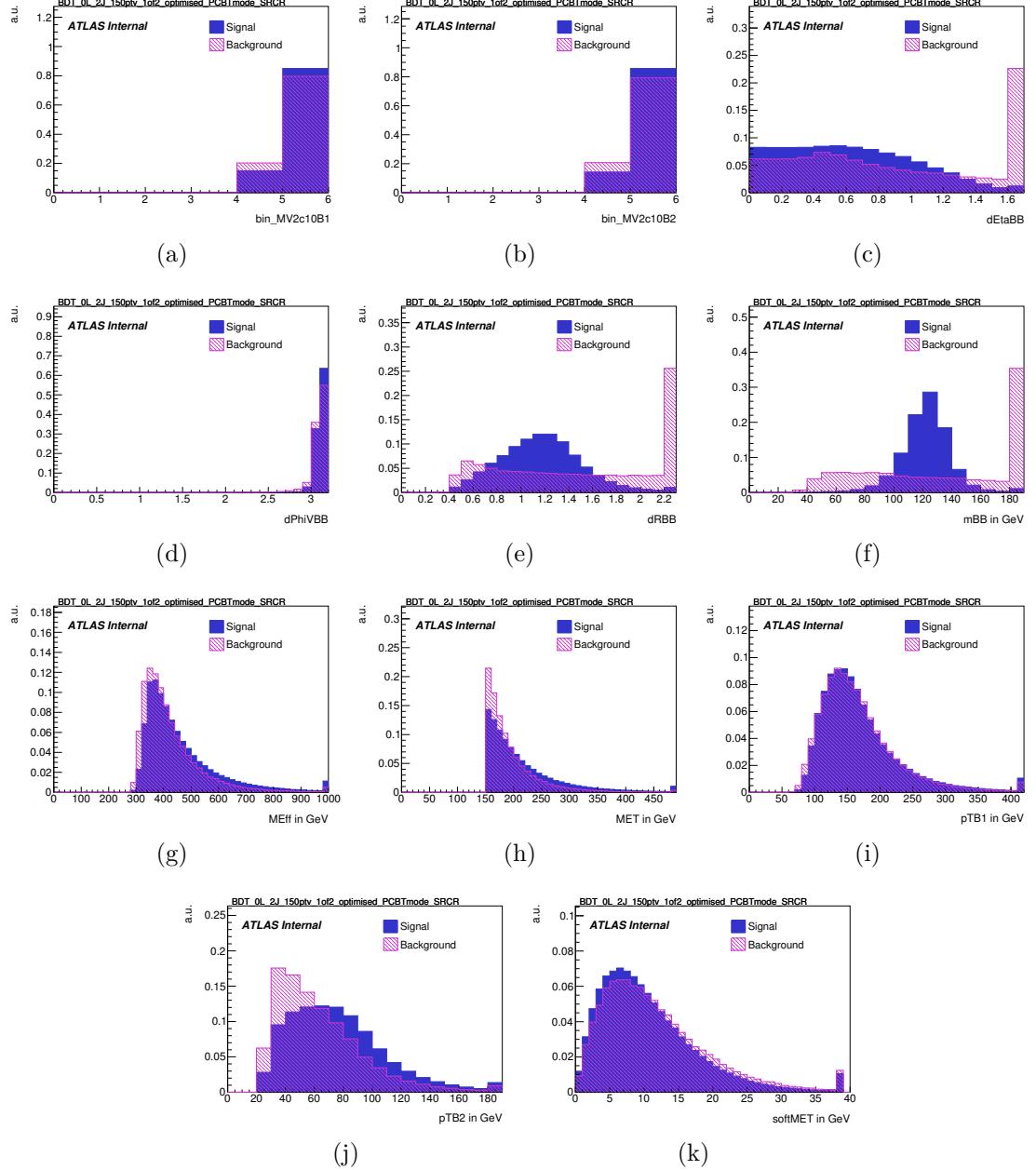


Figure A.1: Inputs to the multi-variate analysis in the 0–lepton 2–jet region. Signal events are shown in blue and background events are shown in red. The signal and background histograms have been normalised to the same area. The distributions only include events with  $E_T^{\text{miss}} > 150$  GeV.

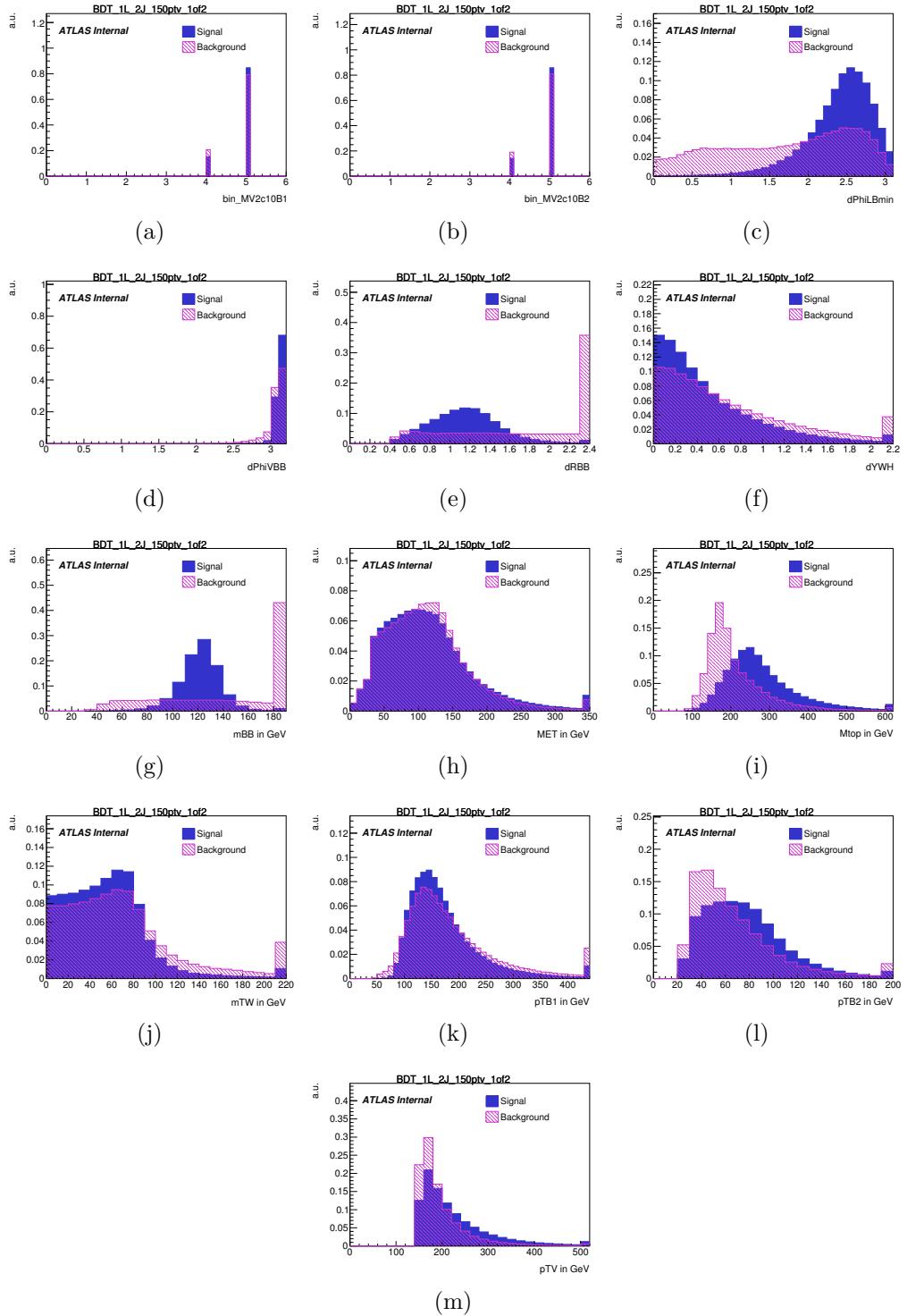


Figure A.2: Inputs to the multi-variate analysis in the 1-lepton 2-jet region. Signal events are shown in blue and background events are shown in red. The signal and background histograms have been normalised to the same area. The distributions only include events with  $p_T^W > 150$  GeV.

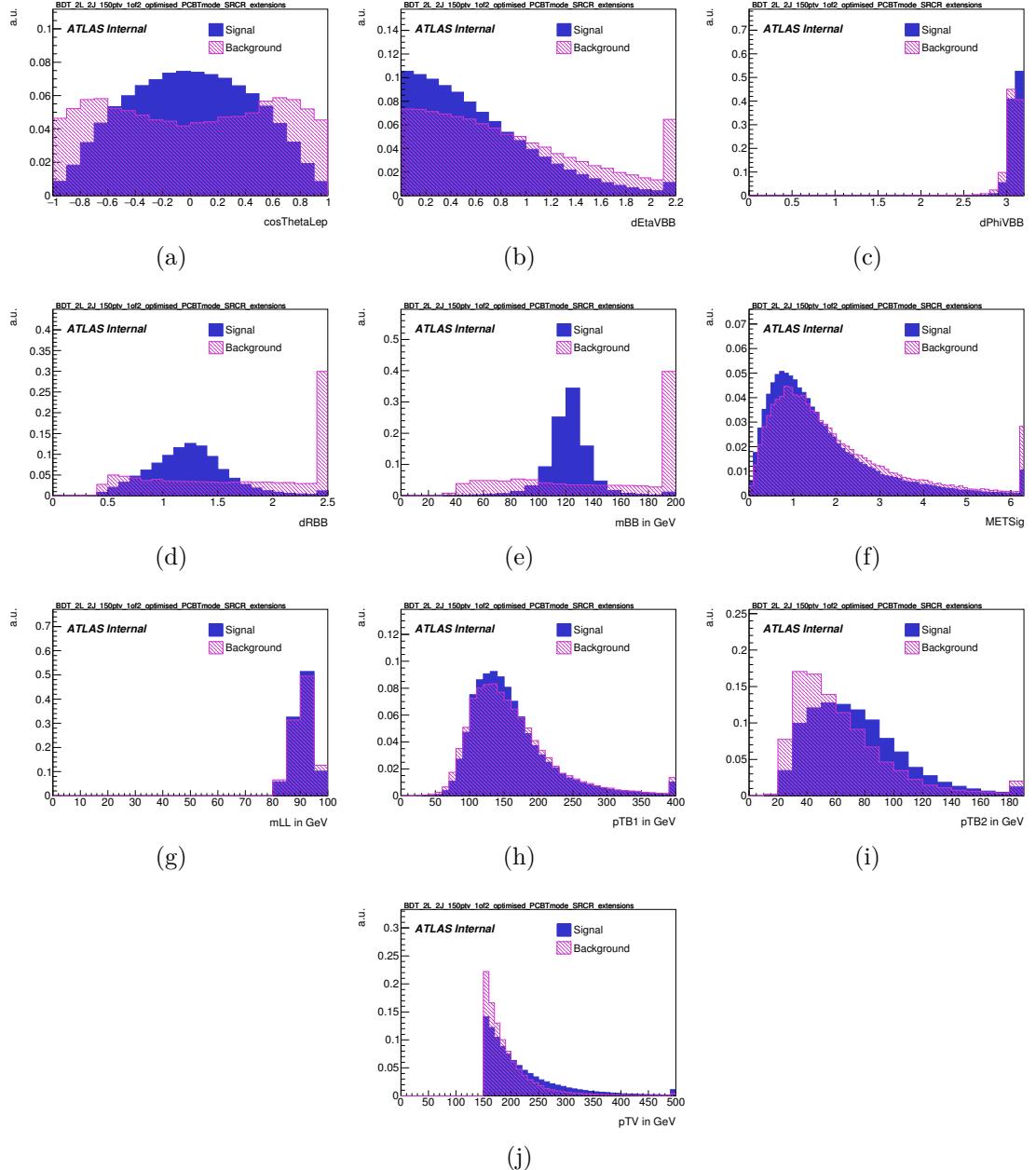


Figure A.3: Inputs to the multi-variate analysis in the 2-lepton 2-jet region. Signal events are shown in blue and background events are shown in red. The signal and background histograms have been normalised to the same area. The distributions only include events with  $p_{\text{T}}^Z > 150 \text{ GeV}$ .

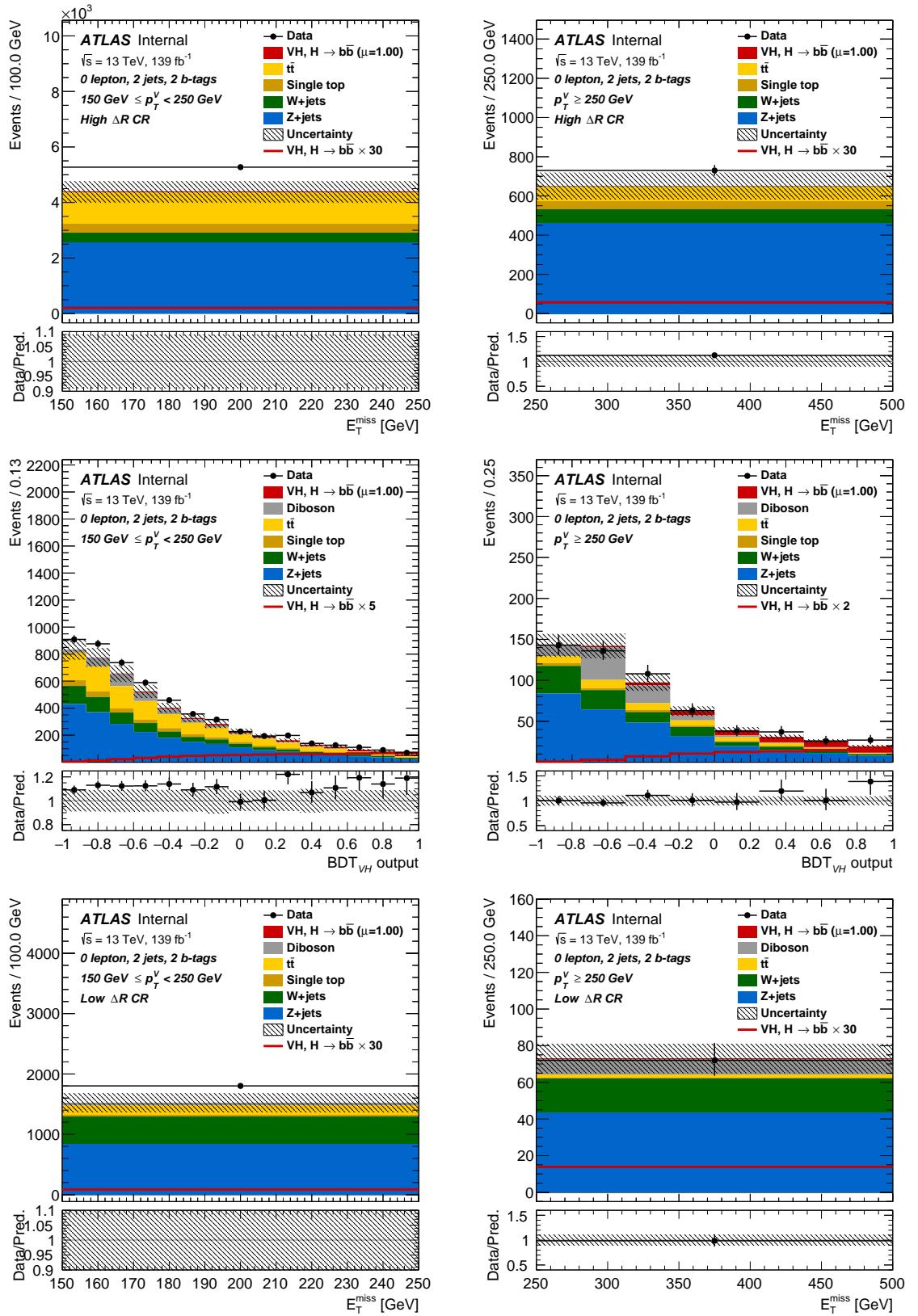


Figure A.4: Pre-fit distributions in the 0-lepton channel in the 2-jet region.

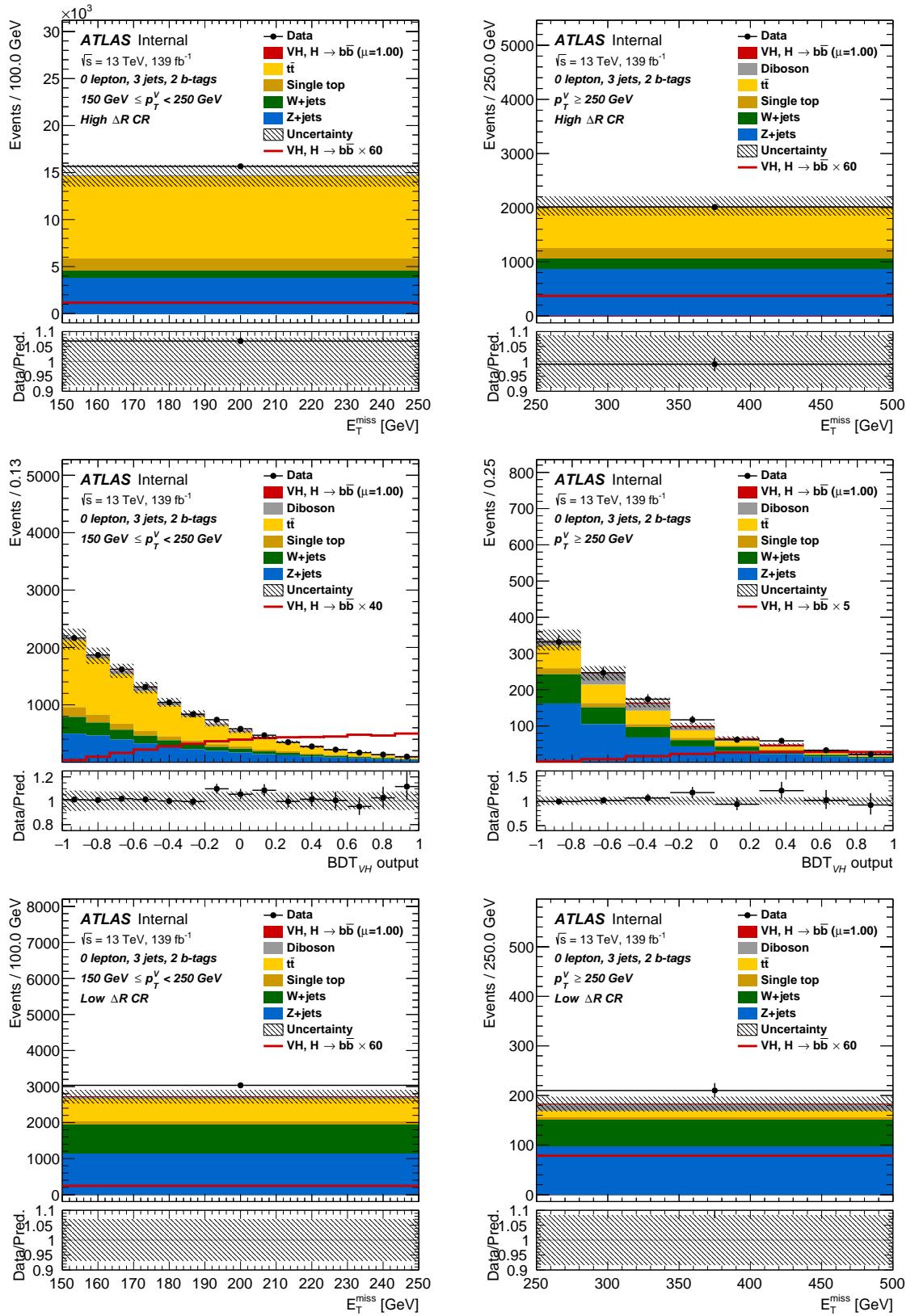


Figure A.5: Pre-fit distributions in the 0-lepton channel in the 3-jet region.

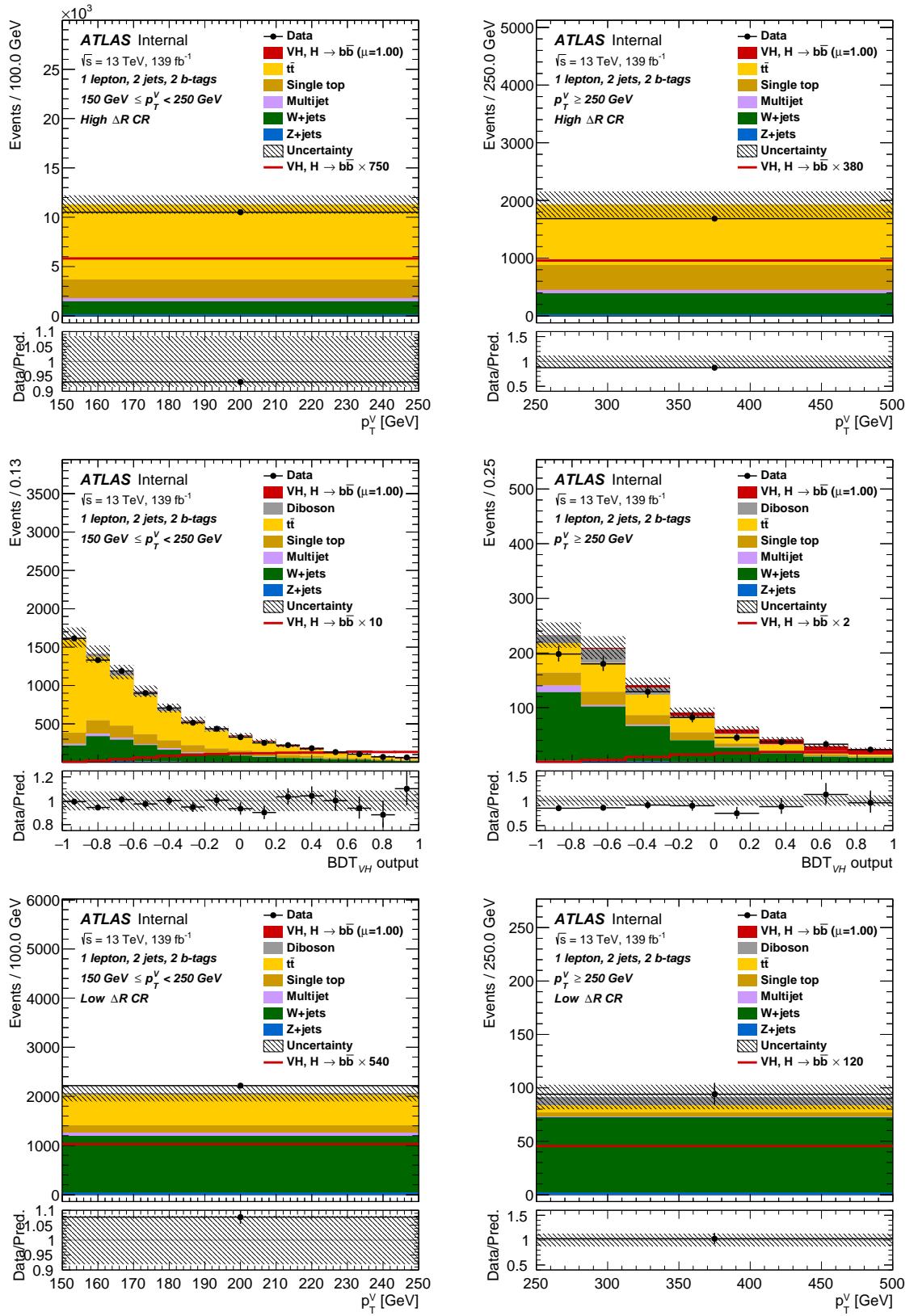


Figure A.6: Pre-fit distributions in the 1-lepton channel in the 2-jet region.

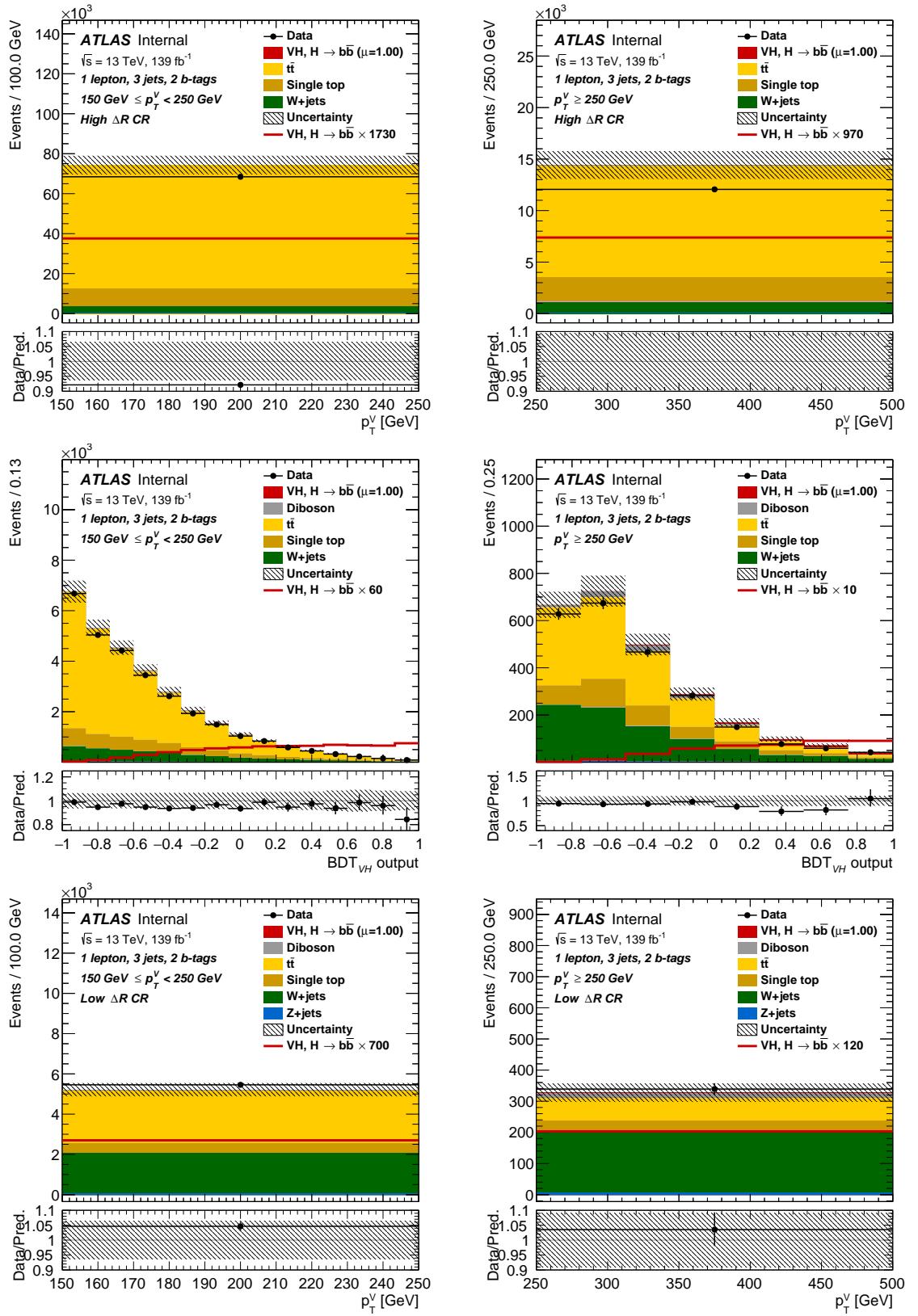


Figure A.7: Pre-fit distributions in the 1-lepton channel in the 3-jet region.

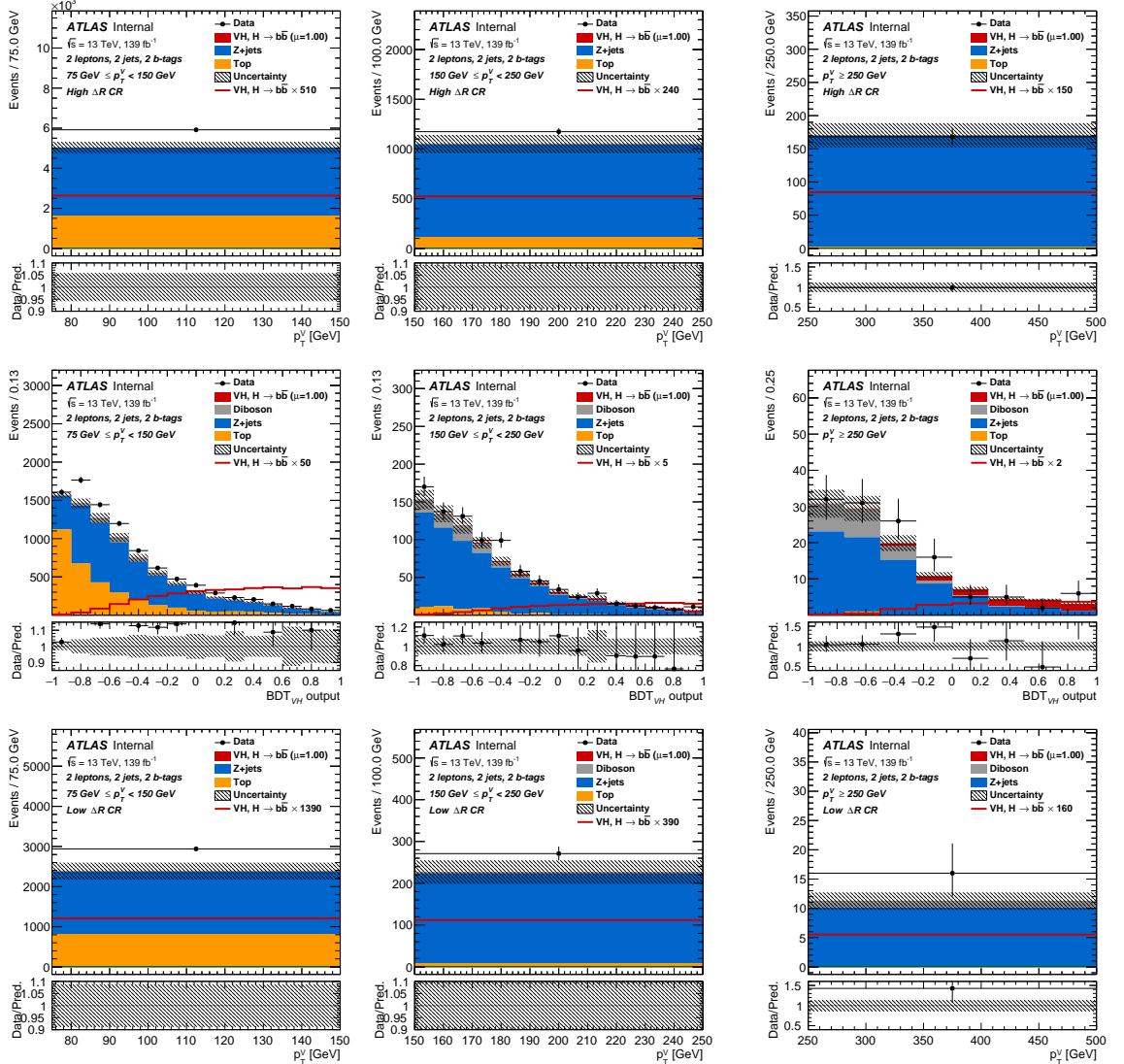
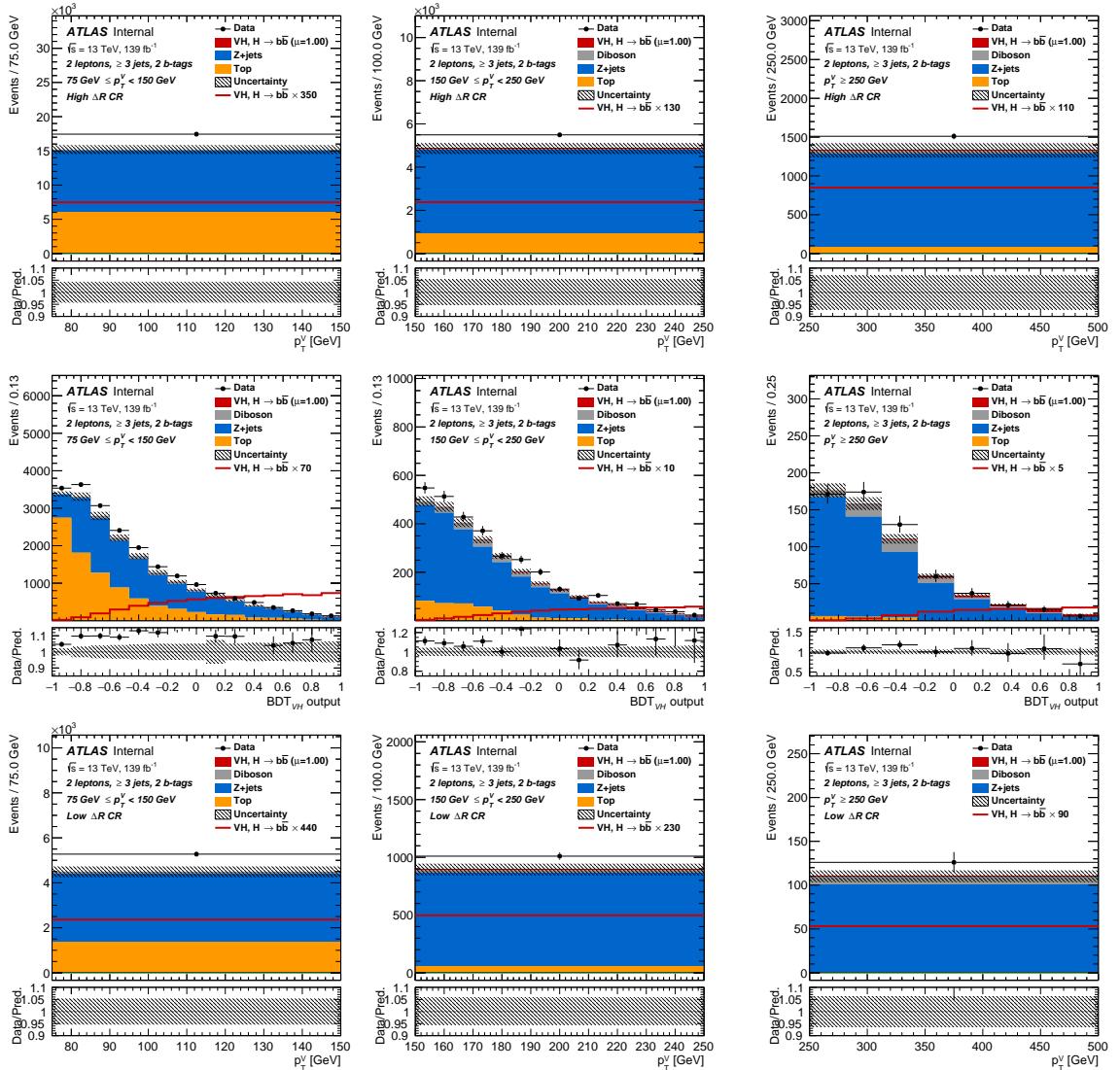


Figure A.8: Pre-fit distributions in the 2-lepton channel in the 2-jet region.


 Figure A.9: Pre-fit distributions in the 2-lepton channel in the  $\geq 3$ -jet region.

## Appendix B

# Supplement to Modelling of Systematic Uncertainties

### B.1 Full sample tables

#### B.1.1 Nominal Samples

$V + \text{jets}$

DS ID	Process	$\sigma \times \text{BR pb}$	k-factor	$\epsilon_{\text{filter}}$	Events
<b><math>W \rightarrow \tau\nu</math></b>					
364184	$W \rightarrow \tau\nu, 0 < \max(H_T, p_T^W) < 70 \text{ GeV}, \text{c-veto-b-veto}$	19152.0	0.9702	0.824 95	24 964 000
364185	$W \rightarrow \tau\nu, 0 < \max(H_T, p_T^W) < 70 \text{ GeV}, \text{c-filter-b-veto}$	19153.0	0.9702	0.129 34	9 994 600
364186	$W \rightarrow \tau\nu, 0 < \max(H_T, p_T^W) < 70 \text{ GeV}, \text{b-filter}$	19163.0	0.9702	0.044 594	17 487 200
364187	$W \rightarrow \tau\nu, 70 < \max(H_T, p_T^W) < 140 \text{ GeV}, \text{c-veto-b-veto}$	947.65	0.9702	0.673 82	14 999 500
364188	$W \rightarrow \tau\nu, 70 < \max(H_T, p_T^W) < 140 \text{ GeV}, \text{c-filter-b-veto}$	946.73	0.9702	0.222 22	10 000 000
364189	$W \rightarrow \tau\nu, 70 < \max(H_T, p_T^W) < 140 \text{ GeV}, \text{b-filter}$	943.3	0.9702	0.103 91	10 000 000
364190	$W \rightarrow \tau\nu, 140 < \max(H_T, p_T^W) < 280 \text{ GeV}, \text{c-veto-b-veto}$	339.36	0.9702	0.596 22	10 000 000
364191	$W \rightarrow \tau\nu, 140 < \max(H_T, p_T^W) < 280 \text{ GeV}, \text{c-filter-b-veto}$	339.63	0.9702	0.290 25	7 500 000
364192	$W \rightarrow \tau\nu, 140 < \max(H_T, p_T^W) < 280 \text{ GeV}, \text{b-filter}$	339.55	0.9702	0.112 29	24 999 900
364193	$W \rightarrow \tau\nu, 280 < \max(H_T, p_T^W) < 500 \text{ GeV}, \text{c-veto-b-veto}$	72.065	0.9702	0.545 69	4 999 200
364194	$W \rightarrow \tau\nu, 280 < \max(H_T, p_T^W) < 500 \text{ GeV}, \text{c-filter-b-veto}$	71.976	0.9702	0.316 48	2 998 400
364195	$W \rightarrow \tau\nu, 280 < \max(H_T, p_T^W) < 500 \text{ GeV}, \text{b-filter}$	72.026	0.9702	0.134 26	2 999 100
364196	$W \rightarrow \tau\nu, 500 < \max(H_T, p_T^W) < 1000 \text{ GeV}$	15.046	0.9702	1.0	6 000 000
364197	$W \rightarrow \tau\nu, 1000 < \max(H_T, p_T^W) < 13000 \text{ GeV}$	1.2339	0.9702	1.0	4 000 000
<b><math>W \rightarrow \mu\nu</math></b>					
364156	$W \rightarrow \mu\nu, 0 < \max(H_T, p_T^W) < 70 \text{ GeV}, \text{c-veto-b-veto}$	19143.0	0.9702	0.8238	24 986 000
364157	$W \rightarrow \mu\nu, 0 < \max(H_T, p_T^W) < 70 \text{ GeV}, \text{c-filter-b-veto}$	19146.0	0.9702	0.130 35	19 984 000
364158	$W \rightarrow \mu\nu, 0 < \max(H_T, p_T^W) < 70 \text{ GeV}, \text{b-filter}$	19147.0	0.9702	0.044 601	34 971 800
364159	$W \rightarrow \mu\nu, 70 < \max(H_T, p_T^W) < 140 \text{ GeV}, \text{c-veto-b-veto}$	945.52	0.9702	0.674 64	29 933 500
364160	$W \rightarrow \mu\nu, 70 < \max(H_T, p_T^W) < 140 \text{ GeV}, \text{c-filter-b-veto}$	945.53	0.9702	0.232 55	19 948 600
364161	$W \rightarrow \mu\nu, 70 < \max(H_T, p_T^W) < 140 \text{ GeV}, \text{b-filter}$	945.11	0.9702	0.075 648	19 915 000
364162	$W \rightarrow \mu\nu, 140 < \max(H_T, p_T^W) < 280 \text{ GeV}, \text{c-veto-b-veto}$	339.93	0.9702	0.610 58	20 000 000
364163	$W \rightarrow \mu\nu, 140 < \max(H_T, p_T^W) < 280 \text{ GeV}, \text{c-filter-b-veto}$	340.02	0.9702	0.2894	15 000 000
364164	$W \rightarrow \mu\nu, 140 < \max(H_T, p_T^W) < 280 \text{ GeV}, \text{b-filter}$	339.54	0.9702	0.108 72	24 585 000
364165	$W \rightarrow \mu\nu, 280 < \max(H_T, p_T^W) < 500 \text{ GeV}, \text{c-veto-b-veto}$	72.104	0.9702	0.546 47	4 999 000
364166	$W \rightarrow \mu\nu, 280 < \max(H_T, p_T^W) < 500 \text{ GeV}, \text{c-filter-b-veto}$	72.14	0.9702	0.317 43	2 999 000
364167	$W \rightarrow \mu\nu, 280 < \max(H_T, p_T^W) < 500 \text{ GeV}, \text{b-filter}$	72.051	0.9702	0.133 37	2 999 500
364168	$W \rightarrow \mu\nu, 500 < \max(H_T, p_T^W) < 1000 \text{ GeV}$	15.015	0.9702	1.0	5 998 500
364169	$W \rightarrow \mu\nu, 1000 < \max(H_T, p_T^W) < 13000 \text{ GeV}$	1.2348	0.9702	1.0	4 000 000
<b><math>W \rightarrow e\nu</math></b>					
364170	$W \rightarrow e\nu, 0 < \max(H_T, p_T^W) < 70 \text{ GeV}, \text{c-veto-b-veto}$	19151.0	0.9702	0.824 47	24 998 000
364171	$W \rightarrow e\nu, 0 < \max(H_T, p_T^W) < 70 \text{ GeV}, \text{c-filter-b-veto}$	19148.0	0.9702	0.130 33	19 991 000
364172	$W \rightarrow e\nu, 0 < \max(H_T, p_T^W) < 70 \text{ GeV}, \text{b-filter}$	19145.0	0.9702	0.044 141	17 492 400
364173	$W \rightarrow e\nu, 70 < \max(H_T, p_T^W) < 140 \text{ GeV}, \text{c-veto-b-veto}$	945.33	0.9702	0.671 11	29 680 000
364174	$W \rightarrow e\nu, 70 < \max(H_T, p_T^W) < 140 \text{ GeV}, \text{c-filter-b-veto}$	946.77	0.9702	0.228 23	11 580 400
364175	$W \rightarrow e\nu, 70 < \max(H_T, p_T^W) < 140 \text{ GeV}, \text{b-filter}$	946.23	0.9702	0.103 41	9 905 900
364176	$W \rightarrow e\nu, 140 < \max(H_T, p_T^W) < 280 \text{ GeV}, \text{c-veto-b-veto}$	339.88	0.9702	0.599 77	20 000 000
364177	$W \rightarrow e\nu, 140 < \max(H_T, p_T^W) < 280 \text{ GeV}, \text{c-filter-b-veto}$	340.12	0.9702	0.289 65	7 500 000
364178	$W \rightarrow e\nu, 140 < \max(H_T, p_T^W) < 280 \text{ GeV}, \text{b-filter}$	339.79	0.9702	0.108 98	24 999 800
364179	$W \rightarrow e\nu, 280 < \max(H_T, p_T^W) < 500 \text{ GeV}, \text{c-veto-b-veto}$	72.093	0.9702	0.544 41	4 998 800
364180	$W \rightarrow e\nu, 280 < \max(H_T, p_T^W) < 500 \text{ GeV}, \text{c-filter-b-veto}$	72.136	0.9702	0.316 75	2 999 400
364181	$W \rightarrow e\nu, 280 < \max(H_T, p_T^W) < 500 \text{ GeV}, \text{b-filter}$	72.111	0.9702	0.133 86	3 019 000
364182	$W \rightarrow e\nu, 500 < \max(H_T, p_T^W) < 1000 \text{ GeV}$	15.04	0.9702	1.0	5 999 600
364183	$W \rightarrow e\nu, 1000 < \max(H_T, p_T^W) < 13000 \text{ GeV}$	1.2336	0.9702	1.0	4 000 000

 Table B.1:  $W \rightarrow \ell\nu$  samples used in the analysis, which are all generated with SHERPA 2.2.1.

DS ID	Process	$\sigma \times BR \text{ pb}$	k-factor	$\epsilon_{\text{filter}}$	Events
<b><math>Z \rightarrow \tau\tau</math></b>					
364128	$Z \rightarrow \tau\tau, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, c-veto-b-veto}$	1981.6	0.9751	0.82142	7 996 000
364129	$Z \rightarrow \tau\tau, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, c-filter-b-veto}$	1978.8	0.9751	0.11314	4 999 000
364130	$Z \rightarrow \tau\tau, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, b-filter}$	1981.8	0.9751	0.064453	7 995 800
364131	$Z \rightarrow \tau\tau, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, c-veto-b-veto}$	110.37	0.9751	0.68883	5 998 500
364132	$Z \rightarrow \tau\tau, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, c-filter-b-veto}$	110.51	0.9751	0.1829	1 999 200
364133	$Z \rightarrow \tau\tau, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, b-filter}$	110.87	0.9751	0.110886	5 999 550
364134	$Z \rightarrow \tau\tau, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, c-veto-b-veto}$	40.781	0.9751	0.60821	5 000 000
364135	$Z \rightarrow \tau\tau, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, c-filter-b-veto}$	40.74	0.9751	0.22897	3 000 000
364136	$Z \rightarrow \tau\tau, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, b-filter}$	40.761	0.9751	0.13442	4 999 950
364137	$Z \rightarrow \tau\tau, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, c-veto-b-veto}$	8.5502	0.9751	0.56036	2 000 000
364138	$Z \rightarrow \tau\tau, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, c-filter-b-veto}$	8.6707	0.9751	0.26245	1 000 000
364139	$Z \rightarrow \tau\tau, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, b-filter}$	8.6804	0.9751	0.17313	1 999 950
364140	$Z \rightarrow \tau\tau, 500 < \max(H_T, p_T^Z) < 1000 \text{ GeV}$	1.8096	0.9751	1.0	2 999 800
364141	$Z \rightarrow \tau\tau, 1000 < \max(H_T, p_T^Z) < 13000 \text{ GeV}$	0.14834	0.9751	1.0	1 000 000
<b><math>Z \rightarrow \mu\mu</math></b>					
364100	$Z \rightarrow \mu\mu, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, c-veto-b-veto}$	1983.0	0.9751	0.8221	7 982 000
364101	$Z \rightarrow \mu\mu, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, c-filter-b-veto}$	1978.4	0.9751	0.11308	4 983 000
364102	$Z \rightarrow \mu\mu, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, b-filter}$	1982.2	0.9751	0.064161	7 984 000
364103	$Z \rightarrow \mu\mu, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, c-veto-b-veto}$	108.92	0.9751	0.68873	5 983 000
364104	$Z \rightarrow \mu\mu, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, c-filter-b-veto}$	109.42	0.9751	0.18596	1 996 800
364105	$Z \rightarrow \mu\mu, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, b-filter}$	108.91	0.9751	0.11375	5 981 600
364106	$Z \rightarrow \mu\mu, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, c-veto-b-veto}$	39.878	0.9751	0.60899	5 000 000
364107	$Z \rightarrow \mu\mu, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, c-filter-b-veto}$	39.795	0.9751	0.23308	3 000 000
364108	$Z \rightarrow \mu\mu, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, b-filter}$	39.908	0.9751	0.14618	12 499 900
364109	$Z \rightarrow \mu\mu, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, c-veto-b-veto}$	8.5375	0.9751	0.55906	2 000 000
364110	$Z \rightarrow \mu\mu, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, c-filter-b-veto}$	8.5403	0.9751	0.26528	999 600
364111	$Z \rightarrow \mu\mu, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, b-filter}$	8.4932	0.9751	0.17559	1 999 400
364112	$Z \rightarrow \mu\mu, 500 < \max(H_T, p_T^Z) < 1000 \text{ GeV}$	1.7881	0.9751	1.0	2 996 500
364113	$Z \rightarrow \mu\mu, 1000 < \max(H_T, p_T^Z) < 13000 \text{ GeV}$	0.14769	0.9751	1.0	1 000 000
<b><math>Z \rightarrow ee</math></b>					
364114	$Z \rightarrow ee, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, c-veto-b-veto}$	1981.8	0.9751	0.82106	8 000 000
364115	$Z \rightarrow ee, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, c-filter-b-veto}$	1980.8	0.9751	0.11295	4 999 000
364116	$Z \rightarrow ee, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, b-filter}$	1981.7	0.9751	0.063809	7 995 600
364117	$Z \rightarrow ee, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, c-veto-b-veto}$	110.5	0.9751	0.69043	5 997 000
364118	$Z \rightarrow ee, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, c-filter-b-veto}$	110.63	0.9751	0.18382	1 999 200
364119	$Z \rightarrow ee, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, b-filter}$	110.31	0.9751	0.11443	5 970 000
364120	$Z \rightarrow ee, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, c-veto-b-veto}$	40.731	0.9751	0.61452	5 000 000
364121	$Z \rightarrow ee, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, c-filter-b-veto}$	40.67	0.9751	0.23044	3 000 000
364122	$Z \rightarrow ee, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, b-filter}$	40.694	0.9751	0.14927	12 499 600
364123	$Z \rightarrow ee, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, c-veto-b-veto}$	8.6743	0.9751	0.56134	1 999 800
364124	$Z \rightarrow ee, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, c-filter-b-veto}$	8.6711	0.9751	0.26294	999 900
364125	$Z \rightarrow ee, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, b-filter}$	8.6766	0.9751	0.17223	1 999 850
364126	$Z \rightarrow ee, 500 < \max(H_T, p_T^Z) < 1000 \text{ GeV}$	1.8081	0.9751	1.0	3 000 000
364127	$Z \rightarrow ee, 1000 < \max(H_T, p_T^Z) < 13000 \text{ GeV}$	0.14857	0.9751	1.0	1 000 000

 Table B.2:  $Z \rightarrow \ell\ell$  samples used in the analysis, which are all generated using SHERPA 2.2.1.

DS ID	Process	$\sigma \times BR \text{ pb}$	k-factor	$\epsilon_{\text{filter}}$	Events
<b><math>Z \rightarrow \nu\nu</math></b>					
364142	$Z \rightarrow \nu\nu, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, c-veto-b-veto}$	10 700.0	0.9728	0.8216	10 000 000
364143	$Z \rightarrow \nu\nu, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, c-filter-b-veto}$	10 702.0	0.9728	0.11123	8 000 000
364144	$Z \rightarrow \nu\nu, 0 < \max(H_T, p_T^Z) < 70 \text{ GeV, b-filter}$	10 709.0	0.9728	0.066175	8 000 000
364145	$Z \rightarrow \nu\nu, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, c-veto-b-veto}$	603.23	0.9728	0.68924	14 974 000
364146	$Z \rightarrow \nu\nu, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, c-filter-b-veto}$	608.15	0.9728	0.18243	14 980 800
364147	$Z \rightarrow \nu\nu, 70 < \max(H_T, p_T^Z) < 140 \text{ GeV, b-filter}$	603.32	0.9728	0.11955	19 984 500
364148	$Z \rightarrow \nu\nu, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, c-veto-b-veto}$	222.28	0.9728	0.60735	14 998 800
364149	$Z \rightarrow \nu\nu, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, c-filter-b-veto}$	221.88	0.9728	0.22527	12 498 500
364150	$Z \rightarrow \nu\nu, 140 < \max(H_T, p_T^Z) < 280 \text{ GeV, b-filter}$	222.4	0.9728	0.15103	19 998 300
364151	$Z \rightarrow \nu\nu, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, c-veto-b-veto}$	47.375	0.9728	0.55887	4 996 400
364152	$Z \rightarrow \nu\nu, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, c-filter-b-veto}$	47.397	0.9728	0.26201	3 497 500
364153	$Z \rightarrow \nu\nu, 280 < \max(H_T, p_T^Z) < 500 \text{ GeV, b-filter}$	47.476	0.9728	0.17514	8 996 350
364154	$Z \rightarrow \nu\nu, 500 < \max(H_T, p_T^Z) < 1000 \text{ GeV}$	9.9099	0.9728	1.0	10 000 000
364155	$Z \rightarrow \nu\nu, 1000 < \max(H_T, p_T^Z) < 13000 \text{ GeV}$	0.81809	0.9728	1.0	5 000 000
366010	$Z \rightarrow \nu\nu, 70 < p_T^Z < 100 \text{ GeV, b-filter}$	275.19	0.9728	0.0755	26 782 000
366011	$Z \rightarrow \nu\nu, 100 < p_T^Z < 140 \text{ GeV} < M_{jj} < 500, \text{ b-filter}$	109.1	0.9728	0.0914	18 469 000
366012	$Z \rightarrow \nu\nu, 100 < p_T^Z < 140 \text{ GeV} < 500 < M_{jj} < 1000, \text{ b-filter}$	4.5528	0.9728	0.1247	1 037 000
366013	$Z \rightarrow \nu\nu, 100 < p_T^Z < 140 \text{ GeV} < M_{jj} > 1000, \text{ b-filter}$	1.2027	0.9728	0.1141	442 780
366014	$Z \rightarrow \nu\nu, 140 < p_T^Z < 280 \text{ GeV} < M_{jj} < 500, \text{ b-filter}$	51.783	0.9728	0.1020	13 246 337
366015	$Z \rightarrow \nu\nu, 140 < p_T^Z < 280 \text{ GeV} < 500 < M_{jj} < 1000, \text{ b-filter}$	4.4687	0.9728	0.1288	1 468 273
366016	$Z \rightarrow \nu\nu, 140 < p_T^Z < 280 \text{ GeV} < M_{jj} > 1000, \text{ b-filter}$	1.3760	0.9728	0.1196	473 363
366017	$Z \rightarrow \nu\nu, 280 < p_T^Z < 500 \text{ GeV, b-filter}$	4.2468	0.9728	0.1139	1 951 000

 Table B.3: Nominal  $Z \rightarrow \nu\nu$  predictions used in the analysis, which are all generated using SHERPA 2.2.1.

$t\bar{t}$ 

DS ID	Process	Generator	$\sigma \times \text{BR nb}$	k-factor	$\epsilon_{\text{filter}}$	Events
410470	non-all-had $t\bar{t}$	POWHEG + PYTHIA 8	0.831 76	1	0.543 00	709 060 000
410472	dilepton $t\bar{t}$	POWHEG + PYTHIA 8	0.831 76	1	0.105 00	554 162 000
345935	non-all-had $t\bar{t}$ , $100 < E_T^{\text{miss}} < 200$ GeV	POWHEG + PYTHIA 8	0.831 76	1	0.095 20	22 000 000
407345	non-all-had $t\bar{t}$ , $200 < E_T^{\text{miss}} < 300$ GeV	POWHEG + PYTHIA 8	0.831 76	1	0.008 06	41 936 300
407346	non-all-had $t\bar{t}$ , $300 < E_T^{\text{miss}} < 400$ GeV	POWHEG + PYTHIA 8	0.831 76	1	0.001 12	16 491 045
407347	non-all-had $t\bar{t}$ , $E_T^{\text{miss}} > 400$ GeV	POWHEG + PYTHIA 8	0.831 76	1	0.000 31	6 779 400
410480	non-all-had $t\bar{t}$ RadHi	POWHEG + PYTHIA 8	0.831 76	1	0.438 42	376 240 000
410482	dilepton $t\bar{t}$ RadHi	POWHEG + PYTHIA 8	0.831 76	1	0.105 00	312 904 000
410470	non-all-had $t\bar{t}$ RadLo	POWHEG + PYTHIA 8	0.831 76	1	0.543 00	709 060 000
410472	dilepton $t\bar{t}$ RadLo	POWHEG + PYTHIA 8	0.831 76	1	0.105 00	554 162 000
410557	single-lep $t\bar{t}$	POWHEG + HERWIG 7	0.831 76	1	0.438 42	37 826 500
410558	dilepton $t\bar{t}$	POWHEG + HERWIG 7	0.831 76	1	0.105 00	312 507 000
410464	single-lep $t\bar{t}$	aMC@NLO + PYTHIA 8.2	0.831 76	1	0.440 34	376 985 000
410465	dilepton $t\bar{t}$	aMC@NLO + PYTHIA 8.2	0.831 76	1	0.107 18	364 951 000

 Table B.4: Nominal  $t\bar{t}$  samples used in the analysis.

## Signal

DSID	Process	Generator	Events
<b><i>q</i>–<i>q̄</i>-initiated</b>			
345053	$qq \rightarrow W^+ H \rightarrow l^+ \nu b\bar{b}$	POWHEG + MiNLO + PYTHIA 8	7319850
345054	$qq \rightarrow W^- H \rightarrow l^- \nu b\bar{b}$	POWHEG + MiNLO + PYTHIA 8	14634450
345055	$qq \rightarrow ZH \rightarrow ll b\bar{b}$	POWHEG + MiNLO + PYTHIA 8	10979800
345056	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + MiNLO + PYTHIA 8	7319400
<b><i>gg</i>-initiated</b>			
345057	$gg \rightarrow ZH \rightarrow ll b\bar{b}$	POWHEG + PYTHIA 8	2740000
345058	$gg \rightarrow ZH \rightarrow \nu\nu b\bar{b}(\ast)$	POWHEG + PYTHIA 8	1835000

Table B.5: Nominal predictions of the  $VH \rightarrow b\bar{b}$  signal process used in the analysis fit and to assess the systematic uncertainties.

**Diboson**

DS ID	Process	Generator	$\sigma \times \text{BR pb}$	$k\text{-factor}$	$\epsilon_{\text{filter}}$	Events
363359	$qq \rightarrow WlvWqq$	SHERPA 2.2.1	24.71	1.00	1.0	28711000
363360	$qq \rightarrow WqqWlv$	SHERPA 2.2.1	24.725	1.00	1.0	14372000
363489	$qq \rightarrow WlvZqq$	SHERPA 2.2.1	11.413	1.0	1.0	7100000
363358	$qq \rightarrow WqqZll$	SHERPA 2.2.1	3.437	1.0	1.0	5324000
363357	$qq \rightarrow WqqZvv$	SHERPA 2.2.1	6.797	1.0	1.0	5939000
363356	$qq \rightarrow ZqqZll$	SHERPA 2.2.1	15.563	1.0	0.139 61	5317000
345045	$qq \rightarrow WlvZbb$	SHERPA 2.2.1	2.503 000	1.0	1.0	18480000
345044	$qq \rightarrow ZllZbb$	SHERPA 2.2.1	3.4739	1.0	0.139 61	15409800
345043	$qq \rightarrow ZvvZbb$	SHERPA 2.2.1	3.4278	1.0	0.279 676	15410000
363355	$qq \rightarrow ZqqZvv$	SHERPA 2.2.1	15.564	1.0	0.279 76	5324000
364304	$gg \rightarrow WlvWqq$	SHERPA 2.2.2	0.6224	1.0	1.0	440000
364305	$gg \rightarrow WlvWqq$	SHERPA 2.2.2	0.622 53	1.0	1.0	440000
364302	$gg \rightarrow ZllZqq$	SHERPA 2.2.2	0.923	1.7	0.141 77	318400
364303	$gg \rightarrow ZvvZqq$	SHERPA 2.2.2	0.924 69	1.7	0.2799	319400

Table B.6: Nominal diboson predictions used in the analysis.

**Single Top**

DS ID	Process	Generator	$\sigma \times \text{BR nb}$	k-factor	$\epsilon_{\text{filter}}$	Events
410013	Single top $Wt$ incl $t$	POWHEG+PYTHIA6	0.034009	1.054	1	4985800
410014	Single top $Wt$ incl $\bar{t}$	POWHEG+PYTHIA6	0.033989	1.054	1	4985600
410646	Single top $Wt$ incl $t$	POWHEG+PYTHIA8	0.037937	0.9450	1	11250000
410647	Single top $Wt$ incl $\bar{t}$	POWHEG+PYTHIA8	0.0379070	0.9457	1	11250000
410062	Single top DS $Wt$ incl $t$	POWHEG+PYTHIA6	0.032384	1.054	1	4948600
410063	Single top DS $Wt$ incl $\bar{t}$	POWHEG+PYTHIA6	0.03235	1.054	1	4983000
410654	Single top DS $Wt$ incl $t$	POWHEG+PYTHIA8	0.036918	0.9711	1	11250000
410655	Single top DS $Wt$ incl $\bar{t}$	POWHEG+PYTHIA8	0.037646	0.9523	1	11250000
410099	Single top $Wt$ incl $t$ radHigh	POWHEG+PYTHIA6	0.034917	1.027	1	4977000
410100	Single top $Wt$ incl $t$ radLow	POWHEG+PYTHIA6	0.033407	1.073	1	4966000
410101	Single top $Wt$ incl $\bar{t}$ radHigh	POWHEG+PYTHIA6	0.034893	1.027	1	4960000
410102	Single top $Wt$ incl $\bar{t}$ radLow	POWHEG+PYTHIA6	0.033388	1.073	1	4959000
410147	Single top $Wt$ incl $t$	POWHEG+HERWIG++	0.0340	1.054	1	4827000
410148	Single top $Wt$ incl $\bar{t}$	POWHEG+HERWIG++	0.0340	1.054	1	4842000
411036	Single top $Wt$ incl $t$	POWHEG+HERWIG7	0.0340	1.054	1	11250000
411036	Single top $Wt$ incl $\bar{t}$	POWHEG+HERWIG7	0.0340	1.054	1	6250000
410165	Single top $Wt$ incl $t + \bar{t}$	MADGRAPH5_aMC@NLO+HERWIG++	0.073588	0.9743	1	9765000
410658	Single top $t$ -chan $t$	POWHEG+PYTHIA8	0.036993	1.1935	1	11232500
410659	Single top $t$ -chan $\bar{t}$	POWHEG+PYTHIA8	0.022175	1.1849	1	11232500
410011	Single top $t$ -chan $t$	POWHEG+PYTHIA6	0.04373	1.0094	1	4986200
410012	Single top $t$ -chan $\bar{t}$	POWHEG+PYTHIA6	0.025778	1.0193	1	4989800
410017	Single top $t$ -chan $t$ radLow	POWHEG+PYTHIA6	0.040343	1.0944	1	2968000
410018	Single top $t$ -chan $t$ radHigh	POWHEG+PYTHIA6	0.04438	0.9949	1	2986000
410019	Single top $t$ -chan $\bar{t}$ radLow	POWHEG+PYTHIA6	0.023751	1.1063	1	1990000
410020	Single top $t$ -chan $\bar{t}$ radHigh	POWHEG+PYTHIA6	0.026387	0.9958	1	1985000
410047	Single top $t$ -chan $t$	POWHEG+HERWIG++	0.043614	1	1	4879000
410048	Single top $t$ -chan $\bar{t}$	POWHEG+HERWIG++	0.025744	1	1	4880000
410141	Single top $t$ -chan $t + \bar{t}$	MADGRAPH5_aMC@NLO+HERWIG++	0.068768	1	1	4872000
410644	Single top $s$ -chan $t$	POWHEG+PYTHIA8	0.0020267	1.0170	1	2498000
410026	Single top $s$ -chan $\bar{t}$	POWHEG+PYTHIA8	0.0012675	1.0167	1	2498000
410025	Single top $s$ -chan $t$	POWHEG+PYTHIA6	0.002052	1.0046	1	997800
410026	Single top $s$ -chan $\bar{t}$	POWHEG+PYTHIA6	0.001262	1.0215	1	995400
410107	Single top $s$ -chan $t$ radHigh	POWHEG+PYTHIA6	0.002076	1	1	990000
410108	Single top $s$ -chan $t$ radLow	POWHEG+PYTHIA6	0.002033	1	1	996000
410109	Single top $s$ -chan $\bar{t}$ radHigh	POWHEG+PYTHIA6	0.001275	1	1	994000
410110	Single top $s$ -chan $\bar{t}$ radLow	POWHEG+PYTHIA6	0.001251	1	1	997000

Table B.7: Nominal and alternative single top process predictions used in the analysis. Note that the POWHEG+PYTHIA8 single top predictions are used as the default configuration in the fit input, while all modelling uncertainties were derived using POWHEG+PYTHIA6 as the nominal configuration.

### B.1.2 Alternative Samples

$V+\text{jets}$

DS ID	Process	$\sigma \times \text{BR pb}$	k-factor	$\epsilon_{\text{filter}}$	Events
<b><math>W \rightarrow e\nu</math></b>					
361520	$W \rightarrow e\nu$ Np=0	13 939.0	1.2019	1.0	13 936 475
361521	$W \rightarrow e\nu$ Np=1	1894.0	1.2019	1.0	9 432 600
361522	$W \rightarrow e\nu$ Np=2	642.66	1.2019	1.0	6 490 000
361523	$W \rightarrow e\nu$ Np=3	179.18	1.2019	1.0	3 499 000
361524	$W \rightarrow e\nu$ Np=4	70.785	1.2019	1.0	4 456 600
<b><math>W \rightarrow \mu\nu</math></b>					
361525	$W \rightarrow \mu\nu$ Np=0	13 935.0	1.2019	1.0	13 922 800
361526	$W \rightarrow \mu\nu$ Np=1	1893.3	1.2019	1.0	9 456 750
361527	$W \rightarrow \mu\nu$ Np=2	642.7	1.2019	1.0	6 488 600
361528	$W \rightarrow \mu\nu$ Np=3	179.19	1.2019	1.0	3 483 000
361529	$W \rightarrow \mu\nu$ Np=4	70.761	1.2019	1.0	4 487 400
<b><math>W \rightarrow \tau\nu</math></b>					
361530	$W \rightarrow \tau\nu$ Np=0	13 920.0	1.2019	1.0	13 982 400
361531	$W \rightarrow \tau\nu$ Np=1	1891.9	1.2019	1.0	9 455 400
361532	$W \rightarrow \tau\nu$ Np=2	641.87	1.2019	1.0	6 492 400
361533	$W \rightarrow \tau\nu$ Np=3	179.21	1.2019	1.0	3 533 000
361534	$W \rightarrow \tau\nu$ Np=4	71.012	1.2019	1.0	4 473 600

 Table B.8: Alternative  $W \rightarrow \ell\nu$  samples used in the analysis, which are all generated using MADGRAPH + PYTHIA 8.

DS ID	Process	$\sigma \times \text{BR pb}$	k-factor	$\epsilon_{\text{filter}}$	Events
<b><math>Z \rightarrow ee</math></b>					
361500	$Z \rightarrow ee$ Np=0	1401.6	1.232	1.0	6 871 800
361501	$Z \rightarrow ee$ Np=1	211.99	1.232	1.0	3 597 000
361502	$Z \rightarrow ee$ Np=2	67.305	1.232	1.0	2 540 800
361503	$Z \rightarrow ee$ Np=3	18.679	1.232	0.99	634 200
361504	$Z \rightarrow ee$ Np=4	7.291	1.232	1.0	222 500
<b><math>Z \rightarrow \mu\mu</math></b>					
361505	$Z \rightarrow \mu\mu$ Np=0	1402	1.232	1.0	6 878 400
361506	$Z \rightarrow \mu\mu$ Np=1	211.95	1.232	1.0	3 599 000
361507	$Z \rightarrow \mu\mu$ Np=2	67.353	1.232	1.0	2 542 600
361508	$Z \rightarrow \mu\mu$ Np=3	18.633	1.232	1.0	633 200
361509	$Z \rightarrow \mu\mu$ Np=4	7.3013	1.232	1.0	220 500
<b><math>Z \rightarrow \tau\tau</math></b>					
361510	$Z \rightarrow \tau\tau$ Np=0	1397.8	1.232	1.0	6 840 000
361511	$Z \rightarrow \tau\tau$ Np=1	211.4	1.232	1.0	3 391 000
361512	$Z \rightarrow \tau\tau$ Np=2	67.176	1.232	1.0	2 542 000
361513	$Z \rightarrow \tau\tau$ Np=3	18.609	1.232	1.0	634 200
361514	$Z \rightarrow \tau\tau$ Np=4	7.2749	1.232	1.0	224 500
<b><math>Z \rightarrow \nu\nu</math></b>					
361515	$Z \rightarrow \nu\nu$ Np=0	7518.4	1.2283	1.0	1 645 600
361516	$Z \rightarrow \nu\nu$ Np=1	1200.1	1.2283	1.0	10 767 600
361517	$Z \rightarrow \nu\nu$ Np=2	387.16	1.2283	1.0	6 096 200
361518	$Z \rightarrow \nu\nu$ Np=3	110.08	1.2283	1.0	3 801 800
361519	$Z \rightarrow \nu\nu$ Np=4	43.389	1.2283	1.0	2 835 100

 Table B.9: Alternative  $Z+\text{jets}$  samples used in the analysis, which are all generated using MADGRAPH + PYTHIA 8.

$t\bar{t}$ 

Generator	Setup Details	Systematic Effect
POWHEG + PYTHIA 8	A14 tune NNPDF30NLO & NNPDF23LO $\text{hdamp} = 1.5 \cdot m_{top}$ nonallhad filter	nominal sample
POWHEG + PYTHIA 8	nominal setup scale variations low ( $\mu_R = \mu_F = 2$ ) $\text{hdamp} = 1.5 \cdot m_{top}$ Up variation of A14 tune (Var3c) nonallhad filter	<i>low variation</i> for additional radiation
POWHEG + PYTHIA 8	nominal setup scale variations high ( $\mu_R = \mu_F = 0.5$ ) $\text{hdamp} = 3.0 \cdot m_{top}$ Down variation of A14 tune (Var3c) nonallhad filter	<i>high variation</i> for additional radiation
POWHEG + HERWIG 7	H7UE tune CT10 & MMHT2014lo68cl $\text{hdamp}=175.2\text{GeV}$ nonallhad filter	fragmentation/hadronisation model
MADGRAPH 5 aMC@NLO + PYTHIA 8	A14 tune NNPDF30NLO & NNPDF23LO nonallhad filter	hard scatter generation and matching

Table B.10: Alternative  $t\bar{t}$  samples used in the analysis.

## Signal

DSID	Process	Generator	Events
<b><math>q\bar{q}</math>-initiated</b>			
346396	$qq \rightarrow WH \rightarrow l\nu b\bar{b}$	POWHEG + HERWIG 7	14615000
346395	$qq \rightarrow WH \rightarrow l\nu b\bar{b}$	POWHEG + HERWIG 7	7320000
346397	$qq \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + HERWIG 7	10980000
346398	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + HERWIG 7	7315000
345816	$qq \rightarrow W^+ H \rightarrow l^+\nu b\bar{b}$	POWHEG + PYTHIA 8 Var1Up	7994300
345810	$qq \rightarrow W^+ H \rightarrow l^+\nu b\bar{b}$	POWHEG + PYTHIA 8 Var1Down	7994450
345828	$qq \rightarrow W^+ H \rightarrow l^+\nu b\bar{b}$	POWHEG + PYTHIA 8 Var2Up	7994450
345822	$qq \rightarrow W^+ H \rightarrow l^+\nu b\bar{b}$	POWHEG + PYTHIA 8 Var2Down	7994400
345792	$qq \rightarrow W^+ H \rightarrow l^+\nu b\bar{b}$	POWHEG + PYTHIA 8 MPIUp	7994350
345786	$qq \rightarrow W^+ H \rightarrow l^+\nu b\bar{b}$	POWHEG + PYTHIA 8 MPIDown	7994300
345804	$qq \rightarrow W^+ H \rightarrow l^+\nu b\bar{b}$	POWHEG + PYTHIA 8 RenUp	7994250
345798	$qq \rightarrow W^+ H \rightarrow l^+\nu b\bar{b}$	POWHEG + PYTHIA 8 RenDown	7994400
345815	$qq \rightarrow W^- H \rightarrow l^-\nu b\bar{b}$	POWHEG + PYTHIA 8 Var1Up	3999850
345809	$qq \rightarrow W^- H \rightarrow l^-\nu b\bar{b}$	POWHEG + PYTHIA 8 Var1Down	3999800
345827	$qq \rightarrow W^- H \rightarrow l^-\nu b\bar{b}$	POWHEG + PYTHIA 8 Var2Up	3999800
345821	$qq \rightarrow W^- H \rightarrow l^-\nu b\bar{b}$	POWHEG + PYTHIA 8 Var2Down	3999850
345791	$qq \rightarrow W^- H \rightarrow l^-\nu b\bar{b}$	POWHEG + PYTHIA 8 MPIUp	3999750
345785	$qq \rightarrow W^- H \rightarrow l^-\nu b\bar{b}$	POWHEG + PYTHIA 8 MPIDown	3999800
345803	$qq \rightarrow W^- H \rightarrow l^-\nu b\bar{b}$	POWHEG + PYTHIA 8 RenUp	3999850
345797	$qq \rightarrow W^- H \rightarrow l^-\nu b\bar{b}$	POWHEG + PYTHIA 8 RenDown	3999850
345817	$qq \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + PYTHIA 8 Var1Up	5999900
345811	$qq \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + PYTHIA 8 Var1Down	5999900
345829	$qq \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + PYTHIA 8 Var2Up	5999950
345823	$qq \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + PYTHIA 8 Var2Down	5999950
345793	$qq \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + PYTHIA 8 MPIUp	5999900
345787	$qq \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + PYTHIA 8 MPIDown	5999900
345805	$qq \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + PYTHIA 8 RenUp	5999950
345799	$qq \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + PYTHIA 8 RenDown	5999900
345818	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + PYTHIA 8 Var1Up	3999650
345812	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + PYTHIA 8 Var1Down	3999650
345830	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + PYTHIA 8 Var2Up	3999600
345824	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + PYTHIA 8 Var2Down	3999650
345794	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + PYTHIA 8 MPIUp	3999650
345788	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + PYTHIA 8 MPIDown	3999600
345806	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + PYTHIA 8 RenUp	3999650
345800	$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}$	POWHEG + PYTHIA 8 RenDown	3999550
<b><math>gg</math>-initiated</b>			
346399	$gg \rightarrow ZH \rightarrow llb\bar{b}$	POWHEG + HERWIG 7	2240000
346400	$gg \rightarrow ZH \rightarrow \nu\nu b\bar{b}(\ast)$	POWHEG + HERWIG 7	1490000

Table B.11: Alternative predictions of the  $VH \rightarrow b\bar{b}$  signal process used in the analysis to assess the systematic uncertainties.

**Diboson**

DS ID	Process	Generator	$\sigma \times BR$ pb	k-factor	$\epsilon_{filter}$	Events
361606	$WlvWqq$	POWHEG + PYTHIA	44.18	1.0	1.0	4343000
361607	$WqqZll$	POWHEG + PYTHIA	3.2777	1.0	1.0	1469000
361608	$WqqZvv$	POWHEG + PYTHIA	5.7576	1.0	1.0	2921000
361609	$WlvZqq$	POWHEG + PYTHIA	10.086	1.0	1.0	9693000
361610	$ZqqZll$	POWHEG + PYTHIA	2.2699	1.0	1.0	3933000
361611	$ZqqZvv$	POWHEG + PYTHIA	3.9422	1.0	1.0	9591000
361592	$WlvWqq$	POWHEG + HERWIG++	44.166	1.0	1.0	4271000
361593	$WqqZll$	POWHEG + HERWIG++	3.2774	1.0	1.0	1446000
361594	$WqqZvv$	POWHEG + HERWIG++	5.7571	1.0	1.0	2888000
361595	$WlvZqq$	POWHEG + HERWIG++	10.085	1.0	1.0	9580000
361596	$ZqqZll$	POWHEG + HERWIG++	2.2699	1.0	1.0	3051000
361597	$ZqqZvv$	POWHEG + HERWIG++	3.9421	1.0	1.0	9556000

Table B.12: Alternative diboson predictions used in this analysis to study systematic uncertainties.

Nuisance Parameter	Description	Samples/Categories	Value	Effect
norm_Wbb_J2	Floating $Whf$ norm 2-jet events	$Whf$ , 2-jet categories	Float	Normalization
norm_Wbb_J3	Floating $Whf$ norm 3(+)jet events	$Whf$ , 3(+)jet categories	Float	Normalization
SysWlNorm	$Wl$ normalization	$Wl$ , all regions	32%	Normalization
SysWclNorm	$Wcl$ normalization	$Wcl$ , all regions	37%	Normalization
SysWbbNorm_L0	Uncertainty on 1- to 0-lep $Whf$ norm extrapolation	$Whf$ , 0-lepton channel	5%	Normalization
SysWccWbbRatio	$Wcc/Wbb$ ratio	$Wcc$ , all regions	26% (0-l), 23% (1-l)	Normalization
SysWbcWbbRatio	$Wbc/Wbb$ ratio	$Wbc$ , all regions	15% (0-l), 30% (1-l)	Normalization
SysWblWbbRatio	$Wbl/Wbb$ ratio	$Wbl$ , all regions	10% (0-l), 30% (1-l)	Normalization
SysWbbCRSReXtrap	Uncertainty on SR to CRs $W+jets$ extrapolation	$W+jets$ , CR <sub>low</sub> and CR <sub>high</sub>	3.6%-14.9%	Normalization
SysWPtV_J2	$p_T^V$ shape	$W+jets$ , 2-jet regions	-	Migration+Shape
SysWPtV_J3	$p_T^V$ shape	$W+jets$ , 3-jet regions	-	Migration+Shape
SysBDTr_W_SHToMG5	SHERPA to MADGRAPH mva shape from $p_T^V$ -factorized BDTr method	$W+jets$ , all regions	-	Shape

Table B.13: Wjets syst summary caption

Nuisance Parameter	Description	Samples/Categories	Value	Effect
norm_Zbb_J2_Bin75_L2	Floating $Zhf$ norm in medium $p_T^V$ ( $75 - 150$ GeV) 2-jet events	$Zhf$ , 2-jet and $p_T^V \in [75, 150]$ GeV category	Float	Normalization
norm_Zbb_J3_Bin75_L2	Floating $Zhf$ norm in medium $p_T^V$ ( $75 - 150$ GeV) 3(+)jet events	$Zhf$ , 3(+)jet, $p_T^V \in [75, 150]$ GeV category	Float	Normalization
norm_Zbb_J2	Floating $Zhf$ norm in high $p_T^V$ ( $> 150$ GeV) 2-jet events	$Zhf$ , all 2-jet and $p_T^V > 150$ GeV categories	Float	Normalization
norm_Zbb_J3	Floating $Zhf$ norm in high $p_T^V$ ( $> 150$ GeV) 3(+)jet events	$Zhf$ , all 3(+)jet and $p_T^V > 150$ GeV categories	Float	Normalization
SysZlNorm	$Zl$ normalization	$Zl$ , all regions	18%	Normalization
SysZclNorm	$Zcl$ normalization	$Zcl$ , all regions	23%	Normalization
SysZbbNorm_OL	Uncertainty on $Zhf$ norm extrapolation to 0-lep	$Zhf$ normalization in 0-lepton	7%	Normalization
SysZccZbbRatio	$Zcc/Zbb$ ratio	$Zcc$ , all regions	15% (0-l), 16% (2-l, 2-j), 13% (2-j, $\geq 3$ -j)	Normalization
SysZbcZbbRatio	$Zbc/Zbb$ ratio	$Zbc$ , all regions	40% (0-l), 40% (2-l, 2-j), 30% (2-j, $\geq 3$ -j)	Normalization
SysZblZbbRatio	$Zbl/Zbb$ ratio	$Zbl$ , all regions	25% (0-l), 28% (2-l, 2-j), 20% (2-j, $\geq 3$ -j)	Normalization
SysZbbCRSReXtrap_BMin75_L2_CRLow	Uncertainty on SR to CR <sub>low</sub> $Z+jets$ extrapolation	$Z+jets$ , CR <sub>low</sub> , $p_T^V \in [75, 150]$ GeV	3.8%-9.9%	Normalization
SysZbbCRSReXtrap_BMin75_L2_CRHigh	Uncertainty on SR to CR <sub>high</sub> $Z+jets$ extrapolation	$Z+jets$ , CR <sub>high</sub> , $p_T^V \in [75, 150]$ GeV	2.7%-4.1%	Normalization
SysZbbCRSReXtrap_CRLow	Uncertainty on SR to CR <sub>low</sub> $Z+jets$ extrapolation	$Z+jets$ , CR <sub>low</sub> , $p_T^V > 150$ GeV	3.8%-9.9%	Normalization
SysZbbCRSReXtrap_CRHigh	Uncertainty on SR to CR <sub>high</sub> $Z+jets$ extrapolation	$Z+jets$ , CR <sub>high</sub> , $p_T^V > 150$ GeV	2.7%-4.1%	Normalization
SysZPTV	$p_T^V$ shape	$Z+jets$ , all regions with $p_T^V > 150$ GeV	-	Migration+Shape
SysZPTV_BMin75_L2	$p_T^V$ shape	$Z+jets$ , all regions in $p_T^V \in [75, 150]$ GeV	-	Migration+Shape
SysZmbb	$m_{bb}$ shape	$Z+jets$ , all regions with $p_T^V > 150$ GeV	-	Shape
SysZmbb_BMin75_L2	$m_{bb}$ shape	$Z+jets$ , all regions in $p_T^V \in [75, 150]$ GeV	-	Shape

 Table B.14:  $Z+jets$  systs summary

## B.2 Summary Tables

### B.3 Extrapolation Uncertainties Induced by Shape Systematics

#### B.3.1 $W+\text{jets}$

##### $p_T^V$ Shape

$p_T^V$ region	CR <sub>low</sub>	SR	CR <sub>texthigh</sub>	weighted sum	weighted sum $\Delta(\text{CR}_{\text{low}} - \text{SR})$	weighted sum $\Delta(\text{CR}_{\text{high}} - \text{SR})$
<b>0-lepton</b>						
<b>2-jet</b>						
150-250 $p_T^V$	5.08	5.49	5.83	5.44	-1.64	-1.64
250 $p_T^V$	14.08	17.52	18.37	17.48	N/A	N/A
$\Delta(\text{high-low}) p_T^V$	9.00	12.03	12.54	12.04	N/A	N/A
<b>3-jet</b>						
150-250 $p_T^V$	4.12	4.41	4.77	4.42	-1.26	-1.41
250 $p_T^V$	14.95	16.97	18.44	17.33	N/A	N/A
$\Delta(\text{high-low}) p_T^V$	10.83	12.56	13.67	12.91	N/A	N/A
<b>1-lepton</b>						
<b>2-jet</b>						
150-250 $p_T^V$	5.14	5.41	5.29	5.30	-1.95	0.60
250 $p_T^V$	14.08	17.34	18.30	17.47	N/A	N/A
$\Delta(\text{high-low}) p_T^V$	8.94	11.93	13.01	12.17	N/A	N/A
<b>3-jet</b>						
150-250 $p_T^V$	4.08	4.38	4.61	4.39	-1.89	1.34
250 $p_T^V$	15.24	18.02	19.01	18.24	N/A	N/A
$\Delta(\text{high-low}) p_T^V$	11.16	13.64	14.40	13.85	N/A	N/A

Table B.15: Extrapolation uncertainties of the  $p_T^V$  shape systematic on the  $W+\text{jets}$  samples.

**$W+jets$  Factorised BDTr Shape**

	CR <sub>low</sub>	SR	CR <sub>high</sub>	weighted sum	weighted sum $\Delta(\text{CR}_{\text{low}} - \text{SR})$	weighted sum $\Delta(\text{CR}_{\text{high}} - \text{SR})$
<b>0-lepton</b>						
<b>2-jet</b>						
150-250 $p_{\text{T}}^V$	-1187	-4.31	12.87	-2.85	-7.32	15.70
250 $p_{\text{T}}^V$	-9.96	-5.59	3.21	-2.84		N/A
$\Delta(\text{high-low}) p_{\text{T}}^V$	1.91	-1.28	-9.66	0.01	N/A	N/A
<b>3-jet</b>						
150-250 $p_{\text{T}}^V$	-4.76	-1.73	5.84	-0.62	-2.70	6.70
250 $p_{\text{T}}^V$	0.5	-1.93	1.32	-0.34		N/A
$\Delta(\text{high-low}) p_{\text{T}}^V$	4.76	-0.20	-4.52	0.29	N/A	N/A
<b>1-lepton</b>						
<b>2-jet</b>						
150-250 $p_{\text{T}}^V$	-1164	-1.05	21.17	2.65	-1013	21.47
250 $p_{\text{T}}^V$	-7.64	-2.30	14.95	4.67		N/A
$\Delta(\text{high-low}) p_{\text{T}}^V$	4.00	-1.25	-6.76	2.03	N/A	N/A
<b>3-jet</b>						
150-250 $p_{\text{T}}^V$	-3.70	-2.15	9.56	1.55	-1.40	9.54
250 $p_{\text{T}}^V$	2.47	0.5	2.73	1.60		
$\Delta(\text{high-low}) p_{\text{T}}^V$	6.17	2.15	-6.83	0.05	N/A	N/A

 Table B.16: Extrapolation uncertainties of the BDTr-based shape systematic on the  $W+jets$  samples.

## Appendix C

### Truth Tagging Probability

Given an event with  $m$  jets and  $n$  tagged jets required, the possible combinations of tagged and non-tagged jets are  $\binom{m}{n}$ . For a given tagged configuration, referred to as the  $i^{th}$  combination and denoted as  $\binom{m}{n}_i$ , the total number of remaining configurations is  $\overline{\binom{m}{n}} = \binom{m}{n} - \binom{m}{n}_i$ . The efficiency and inefficiency products are

$$\varepsilon\left(\binom{m}{n}_i, x, f\right) = \prod_{j \in n} \epsilon_x^f(j) \quad (\text{C.1})$$

$$\varepsilon_{in}\left(\overline{\binom{m}{n}}_i, x, f\right) = \prod_{j \in m-n} (1 - \epsilon_x^f(j)) \quad (\text{C.2})$$

with  $\epsilon_x^f(j)$  the tagging efficiency of jet  $j$  of flavour  $f$  at an efficiency working point  $x$ , and where  $j \in m-n$  refers to the pool of non tagged jets. The total event weight is

$$w_{TT}(x) = \sum_i^{\binom{m}{n}} \varepsilon\left(\binom{m}{n}_i, x\right) \cdot \varepsilon_{in}\left(\overline{\binom{m}{n}}_i, x\right), \quad (\text{C.3})$$

with a probability to choose a specific combination equal to

$$P_i(x) = \frac{\varepsilon\left(\binom{m}{n}_i, x\right) \cdot \varepsilon_{in}\left(\overline{\binom{m}{n}}_i, x\right)}{w_{TT}}. \quad (\text{C.4})$$

## Appendix D

# BDT Transformation

Consider the function:

$$Z = z_s \frac{n_s}{N_s} + z_b \frac{n_b}{N_b}. \quad (\text{D.1})$$

where

- $I[k, l]$  is an interval of the histograms, containing the bins between bin  $k$  and bin  $l$ ;
- $N_s$  is the total number of signal events in the histogram;
- $N_b$  is the total number of background events in the histogram;
- $n_s(I[k, l])$  is the total number of signal events in the interval  $I[k, l]$ ;
- $n_b(I[k, l])$  is the total number of background events in the interval  $I[k, l]$ ;
- $z_s$  and  $z_b$  are parameters used to tune the algorithm.

The values for  $z_b$  and  $z_s$  are chosen such that the total number of bins in the signal regions with  $p_T^V < 250 \text{ GeV}$  is 15 ( $z_s = 10$ ,  $z_b = 5$ ) and 8 ( $z_s = 5$ ,  $z_b = 3$ ) in signal regions with  $p_T^V > 250 \text{ GeV}$ . The values are chosen such that  $z_s > Z_b$  to achieve a finer binning in the high BDT output score region which has, as expected, the best signal to background ratio with respect to the rest of the distribution.

The pre-transformation BDT score distribution has 500 equidistant bins between -1 and 1. The re-binning is then conducted using the following procedure:

1. Starting from the last bin on the right of the original histogram, increase the range of the interval  $I(k, \text{last})$  by adding, one after the other, the bins from the right to the left;
2. Calculate the value of Z at each step;
3. Once  $Z(I[k_0, \text{last}]) > 1$ , rebin all the bins in the interval  $I(k_0, \text{last})$  into a single bin;
4. Repeat steps 1-3, starting this time from the last bin on the right, not included in the previous remap (the new last is  $k_0 - 1$ ), until  $k_0$  in the first bin.

If the statistical uncertainty of the newly formed bin is larger than 20 % step 2 is extended until the statistical uncertainty is below 20 %.