

UNIVERSITY OF CALIFORNIA  
SANTA CRUZ

**SEARCH FOR WW AND WZ RESONANCES IN  $\ell\nu qq$  FINAL  
STATES IN  $pp$  COLLISIONS AT  $\sqrt{s} = 13$  TEV WITH THE ATLAS  
DETECTOR**

A dissertation submitted in partial satisfaction of the  
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

**Natasha Woods**

December 2019

The Dissertation of Natasha Woods  
is approved:

---

Abraham Seiden, Chair

---

Mike Hance

---

Bruce Schumm

---

Quentin Williams  
Vice Provost and Dean of Graduate Studies

Copyright © by

Natasha Woods

2019

# Table of Contents

List of Figures	vi
List of Tables	xvi
Abstract	xix
Dedication	xx
Acknowledgments	xxi
<b>I Introduction</b>	<b>1</b>
<b>1 Introduction</b>	<b>2</b>
<b>II Theoretical Motivation</b>	<b>6</b>
<b>2 The Standard Model of Particle Physics</b>	<b>7</b>
2.1 Introduction . . . . .	7
2.2 Quantum Field Theory . . . . .	7
2.3 $U(1)_{EM}$ Local Gauge Invariance . . . . .	8
2.4 Yang-Mills Gauge Theories . . . . .	11
2.5 Particles in the Standard Model . . . . .	12
2.6 Higgs Mechanism . . . . .	17
2.7 Electroweak Theory . . . . .	18
2.8 Quantum ChromoDynamics . . . . .	19
<b>3 Standard Model Successes and Limitations</b>	<b>25</b>
<b>4 New Physics Models with Diboson Resonances</b>	<b>28</b>
4.1 Randall Sundrum Bulk Model . . . . .	28
4.2 Simple Standard Model Extensions . . . . .	30

<b>III Experimental Setup</b>	<b>32</b>
<b>5 LHC</b>	<b>33</b>
5.1 LHC Layout and Design . . . . .	35
<b>6 The ATLAS Detector</b>	<b>39</b>
6.1 Coordinate System . . . . .	42
6.2 Inner Detector . . . . .	42
6.2.1 Pixel Detector . . . . .	44
6.2.2 Semiconductor Tracker . . . . .	45
6.2.3 Transition Radiation Tracker . . . . .	45
6.3 Calorimeters . . . . .	46
6.4 Muon Spectrometer . . . . .	51
6.5 Magnet System . . . . .	55
6.6 Trigger System . . . . .	56
<b>IV Method</b>	<b>58</b>
<b>7 Dataset and Simulated Samples</b>	<b>59</b>
7.1 Dataset . . . . .	59
7.2 Simulated Samples . . . . .	62
<b>8 Objects</b>	<b>63</b>
8.1 Electrons . . . . .	63
8.2 Muons . . . . .	64
8.3 Jets . . . . .	65
8.3.1 Small-R jets . . . . .	68
8.3.2 Large-R jets . . . . .	68
8.3.3 Variable Radius jets . . . . .	71
8.3.4 Jet Flavor Tagging . . . . .	72
8.4 MET/Neutrinos . . . . .	72
8.5 Overlap Removal . . . . .	72
8.6 Reconstructed Resonance Mass ( $m_{WV}$ ) . . . . .	73
<b>9 Event Selection and Categorization</b>	<b>75</b>
9.1 Pre-selection . . . . .	76
9.2 Trigger . . . . .	77
9.3 non-VBF/VBF RNN . . . . .	79
9.4 Signal Region Definitions . . . . .	86
9.5 Selection Acceptance and Efficiency . . . . .	88

<b>10 Background Estimate</b>	<b>91</b>
10.1 Control Regions . . . . .	92
10.2 Fake Lepton Backgrounds . . . . .	99
<b>11 Systematic Uncertainties</b>	<b>113</b>
11.1 Experimental Systematics . . . . .	113
11.2 Theory Systematics . . . . .	115
<b>12 Statistical Analysis</b>	<b>123</b>
12.1 Likelihood Function Definition . . . . .	124
12.2 Fit Configuration . . . . .	124
12.3 Best Fit $\mu$ . . . . .	127
12.4 Discovery Test . . . . .	127
12.5 Exclusion Limits . . . . .	129
<b>13 Results</b>	<b>131</b>
13.1 Expected and Measured Yields . . . . .	131
13.2 Systematic Profiling and Correlations . . . . .	145
13.3 Discovery Tests . . . . .	147
13.4 Limits . . . . .	153
<b>V Quark and Gluon Tagging</b>	<b>155</b>
<b>14 Prospects</b>	<b>156</b>
<b>15 <math>n_{trk}</math> Calibration</b>	<b>163</b>
<b>16 Application</b>	<b>169</b>
<b>VI Conclusion</b>	<b>174</b>
<b>17 Conclusions</b>	<b>175</b>
<b>Bibliography</b>	<b>176</b>

# List of Figures

2.1	The particles of the Standard Model. . . . .	14
2.2	Summary of how Standard Model particles interact with other Standard Model particles. . . . .	15
2.3	This figure shows the three dominant QCD interactions. From Ref. [18] . . . . .	21
2.4	Strength of the U(1), SU(2), and SU(3) gauge couplings as a function of the energy scale of the interaction ( $Q$ ). From Ref. [12] . .	22
2.5	A cartoon of string breaking: the QCD string spanned between quark $Q$ and antiquark $\bar{Q}$ breaks due to $q\bar{q}$ creation [4] . . . . .	24
3.1	A comparison of cross section measurements at $\sqrt{s} = 7, 8, 13$ TeV from ATLAS compared to theoretical measurements. From Ref. [6]	27
4.1	Cartoon of RS Bulk Model . . . . .	29
5.1	Scaling of various SM cross sections with $\sqrt{s}$ . . . . .	34
5.2	The layout of the LHC and the four detectors along the beam line (ATLAS, LHCb, ALICE, CMS). . . . .	36
5.3	An overview schematic of the LHC accelerator subsystems. . . . .	37
6.1	Overview schematic of the ATLAS detector. . . . .	40
6.2	A simplified schematic of how different particles interact and are detected within ATLAS. . . . .	41
6.3	Layout of ATLAS Inner Detector . . . . .	43

6.4	Layout of ATLAS ID Barrel System. . . . .	44
6.5	Overview of ATLAS electromagnetic and hadronic calorimeters. . . . .	48
6.6	Schematic of ECAL. . . . .	49
6.7	Schematic of HCAL. . . . .	50
6.8	Schematic of Muon Spectrometer [cite G35] . . . . .	53
6.9	Schematic of MDT chamber. [cite G35] . . . . .	54
6.10	Schematic of RPC chamber, which is used for triggering in the central region of the detector [cite G35]. . . . .	54
6.11	Schematic of TGC chamber, which is used for triggering in the muon end-cap region. [cite G35] . . . . .	55
6.12	Layout of ATLAS magnet systems. . . . .	56
7.1	Integrated luminosity for data collected from ATLAS from 2011 - 2018 . . . . .	60
7.2	Mean number of interactions per crossing for data collected from ATLAS from 2011 - 2018 . . . . .	61
8.1	This figure shows the breakdown of the muon reconstruction efficiency scale factor measured in $Z \rightarrow \mu\mu$ as a function of $p_T$ [5]. . . . .	65
8.2	[8] This diagram shows the calibration stages for EM jets. . . . .	67
8.3	The upper cut on $D_2$ (a) and jet mass window cut i.e. the upper and lower boundary of the mass (b) of the $W$ -tagger as a function of jet $p_T$ . Corresponding values for $Z$ -tagger are shown in (c) and (d). The optimal cut values for maximum significance are shown as solid markers and the fitted function as solid lines. Working points from $VV \rightarrow JJ$ [ATLAS-HDBS-2018-31-002] is also shown as dashed lines as a reference. Natasha reword? . . . . .	70
8.4	Natasha write caption . . . . .	71
9.1	RNN architecture. Natasha add caption . . . . .	81

9.2	This figure shows the embedded logic in LSTM cells. This image was taken from [23], where a more in depth discussion about LSTMs may be found.	82
9.3	RNN Score distribution for ggF and VBF signals and backgrounds.	83
9.4	ROC curve using k-fold validation for RNN.	84
9.5	Comparison of GGF Z' limits for different RNN score selections. The bottom panel shows the ratio of the upper limits set for different RNN cuts to the cut-based analysis. In this panel smaller numbers, indicate that the expected upper limit is smaller than the cut-based analysis, which is desired.	85
9.6	Event Categorization. Natasha write more.	88
9.7	Selection acceptance times efficiency for the $W' \rightarrow WZ \rightarrow \ell\nu qq$ events from MC simulations as a function of the $W'$ mass for Drell-Yan (left) and VBF production (right), combining the merged HP and LP signal regions of the $WW \rightarrow \ell\nu J$ selection and the resolved regions of the $WW \rightarrow \ell\nu jj$ selection. Note: the VBF selection acceptance for the DY $W'$ is approximately zero in the left plot.	89
9.8	Selection acceptance times efficiency for the $G \rightarrow WW \rightarrow \ell\nu qq$ events from MC simulations as a function of the $G$ mass for (a) Drell-Yan and (b) VBF production, combining the merged HP and LP signal regions of the $WW \rightarrow \ell\nu J$ selection and the resolved regions of the $WW \rightarrow \ell\nu jj$ selection. Note: the VBF selection acceptance for the ggF $G'_{KK}$ is approximately zero in the left plot.	90
10.1	Data MC comparison for the merged $WW$ HP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.	93

10.2 Data MC comparison for the merged $WW$ LP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	94
10.3 Data MC comparison for the merged $WZ$ HP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	95
10.4 Data MC comparison for the merged $WZ$ LP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	96
10.5 Data MC comparison for the resolved $WW$ TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	97
10.6 Data MC comparison for the resolved $WZ$ TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	98
10.7 The $E_T^{miss}$ distribution in MJCR for 2017 data in the electron channel(left), muon channel with W-boson pT < 150 GeV (center) and > 150 GeV (right). Multi-jet templates are given by the difference between the data and simulated distributions. . . . .	101
10.8 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WW$ electron channel. The MJ template is obtained from the pre-MJ-fit. . . . .	102
10.9 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WW$ muon channel. The MJ template is obtained from the pre-MJ-fit. . . . .	103

10.10 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WZ$ untag electron channel. The MJ template is obtained from the pre-MJ-fit.	104
10.11 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WZ$ untag muon channel. The MJ template is obtained from the pre-MJ-fit.	105
10.12 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WZ$ untag electron channel. The MJ template is obtained from the pre-MJ-fit.	106
10.13 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WZ$ untag muon channel. The MJ template is obtained from the pre-MJ-fit.	107
10.14 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the VBF $WW$ electron channel. The MJ template is obtained from the pre-MJ-fit.	108
10.15 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the VBF $WW$ muon channel. The MJ template is obtained from the pre-MJ-fit.	109
10.16 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the VBF $WZ$ electron channel. The MJ template is obtained from the pre-MJ-fit.	110
10.17 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the VBF $WZ$ muon channel. The MJ template is obtained from the pre-MJ-fit.	111
11.1 The $W/Z+jet$ systematics for the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The top subplot shows the nominal and variation distributions/bands, the middle shows the ratio of the two, and the final shows just the shape of the envelope (the final uncertainty).	117

11.2 The two-point generator comparison between Sherpa and MadGraph for the W/Z+jet samples in the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The normalization of the Madgraph sample is set to the Sherpa value to consider only shape effects. The bottom inset shows the ratio of the two. . . . .	118
11.3 Ratio between the variations of generator (red) and hadronization (blue) variations for the Merged regime for $t\bar{t}$ sample. . . . .	119
11.4 Ratio between the variations of generator (red) and hadronization (blue) variations for the Resolved regime for $t\bar{t}$ sample. . . . .	119
11.5 Ratio between the variations of ISR up (red) and down (blue) variations for the Merged regime for $t\bar{t}$ sample. . . . .	120
11.6 Ratio between the variations of ISR up (red) and down (blue) variations for the Resolved regime for $t\bar{t}$ sample. . . . .	121
11.7 Ratio between the variations of FSR up (red) and down (blue) variations for the Merged regime for $t\bar{t}$ sample. . . . .	122
11.8 Ratio between the variations of FSR up (red) and down (blue) variations for the Resolved regime for $t\bar{t}$ sample. . . . .	122
 12.1 The HVT signal mass resolution as a function of mass fit with a straight line in the Resolved ggF region (left) and VBF (right) region. . . . .	126
12.2 The HVT signal mass resolution as a function of mass fit with a straight line in the Merged ggF region (left) and VBF (right) region. . . . .	126
 13.1 This figure shows the distribution of $m_{\ell\nu qq}$ in the DY WW control regions. . . . .	139
13.2 This figure shows the distribution of $m_{\ell\nu qq}$ in the DY WZ control regions. . . . .	140
13.3 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WW control regions. . . . .	141

13.4 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF $WZ$ control regions. . . . .	142
13.5 This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF $WW$ signal regions. . . . .	143
13.6 This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF $WZ$ Untag signal regions. . . . .	143
13.7 This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF $WZ$ Tag signal regions. . . . .	144
13.8 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF $WZ$ Tag signal regions. . . . .	144
13.9 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF $WZ$ Tag signal regions. . . . .	145
13.10 Ranked systematics and their fitted values for $WW$ DY (right) and VBF (left) selections. . . . .	146
13.11 Ranked systematics and their fitted values for $WZ$ DY (right) and VBF (left) selections. . . . .	146
13.12 Correlations between systematics for $WW$ DY (right) and VBF (left) selections. . . . .	147
13.13 These plots show the measured $p_0$ value as a function of resonance mass for HVT $Z'$ DY production. . . . .	148
13.14 These plots show the measured $p_0$ value as a function of resonance mass for HVT $Z'$ VBF production. . . . .	149
13.15 These plots show the measured $p_0$ value as a function of resonance mass for HVT $W'$ DY production. . . . .	150
13.16 These plots show the measured $p_0$ value as a function of resonance mass for HVT $W'$ VBF production. . . . .	151
13.17 These plots show the measured $p_0$ value as a function of resonance mass for the RS Graviton DY production. . . . .	152
13.18 This figure shows theory, expected and observed limits for HVT $W'$ DY (left) and VBF (right) production. . . . .	153

13.19 This figure shows theory, expected and observed limits for HVT $Z'$ DY (left) and VBF (right) production. . . . .	154
13.20 This figure shows theory, expected and observed limits for RS Gravitons via DY production. . . . .	154
14.1 PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV $Z'$ signals and the background (all simulated backgrounds that pass SR selections). . . . .	158
14.2 The number of tracks in small-R jets in events passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets. These distributions are shown for 300, 500, and 700GeV $Z'$ signals and the background. . . . .	158
14.3 The number of tracks in background small-R jets in events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one. . . . .	159
14.4 The number of tracks in small-R jets in 300GeV $Z'$ events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$ .Note the number of total entries in these plots has been normalized to one. . . . .	159

14.5 The number of tracks in small-R jets in 500GeV Z' events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one. . . . .	160
14.6 The number of tracks in small-R jets in 700GeV Z' events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one. . . . .	160
14.7 ROC Curve for Quark and Gluon Tagging with a cut on the number of tracks that depends on the $\ln(p_T)$ . . . . .	161
14.8 The top panel shows the distribution of $m_{lvqq}$ with and without quark gluon tagging. The middle panel shows the ratio of the signals and backgrounds with and without quark gluon tagging. The bottom panel shows the change in $S/\sqrt{B}$ with quark gluon tagging. . . . .	162
16.1 The top panel shows the distribution of $m_{lvqq}$ with and without requiring jets to be true quarks. The middle panel shows the ratio of the signals and backgrounds with and without requiring jets to be true quarks. The bottom panel shows the change in $S/\sqrt{B}$ when requiring jets to be true quarks. . . . .	170
16.2 Unfolded and extracted $n_C$ qg dstbs. . . . .	171
16.3 PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV Z' signals and the background. . . . .	172

16.4 These figures show the impact of the uncertainties on the number of tracks in the leading jet in the sum of the background sample in the Resolved GGF WW SR (a) tracking efficiency (b) fake (c) PDF (d) ME (e) unfolding uncertainties. . . . . 173

# List of Tables

2.1	Representations of the SM fermions under $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry group. Rows are correspond to different weak isospin states and columns to different QCD color states. . . . .	13
9.1	The list of triggers used in the analysis. . . . .	78
9.2	Summary of selection criteria used to define the signal region (SR), $W$ +jets control region ( $W$ CR) and $t\bar{t}$ control region ( $t\bar{t}$ CR) for merged 1-lepton channel. . . . .	87
9.3	The list of selection cuts in the resolved analysis for the $WW$ and $WZ$ signal regions (SR), $W$ +jets control region (WR) and $t\bar{t}$ control region (TR). . . . .	88
10.1	Definitions of “inverted” leptons used in multijet control region. For the inverted muon selection, $ptvarcone30$ is given by sum of the $p_T$ of tracks in a cone around the muon candidate divided by the muon $p_T$ . The size of the cone, $\delta R$ used is $10\text{GeV}/p_T^\mu$ or 0.3, whichever is smaller. So, as the $p_T$ of the muon increases, the cone size used decreases. This is useful as more boosted muons are more likely to be produced in dense environments and using a smaller cone size more accurately determines the quality of the muon. . . . .	100

10.2 Fit validation result in WCRs for 2015+16 data. The fit is done in various WCRs, in order to obtain the corresponding scale factors for MJ templates: ggF resolved WCR for the $WW \rightarrow lvqq$ selection, ggF resolved untagged WCR for the $WZ \rightarrow lvqq$ selection, ggF resolved tagged WCR for the $WZ \rightarrow lvqq$ selection, VBF resolved WCR for the $WW \rightarrow lvqq$ selection, and VBF resolved WCR for the $WZ \rightarrow lvqq$ selection. Post-fit event yields for electroweak processes and MJ contributions are shown. The SF column shows the corresponding normalization scale factors for electroweak processes from the fit. R.U. stands for relative uncertainty. . . . .	112
13.1 Expected and Measured for DY $WW$ $W+jets$ , $t\bar{t}$ control regions and signal regions. . . . .	132
13.2 Expected and Measured for DY $WZ$ $W+jets$ , $t\bar{t}$ tag and untag control regions. . . . .	133
13.3 Expected and Measured for DY $WZ$ $W+jets$ , $t\bar{t}$ tag and untag signal regions. . . . .	134
13.4 Expected and Measured for VBF $WW$ $W+jets$ , $t\bar{t}$ control regions and signal regions. . . . .	135
13.5 Expected and Measured for VBF $WZ$ $W+jets$ , $t\bar{t}$ control regions and signal regions. . . . .	136
13.6 Fitted background normalizations for $t\bar{t}$ and $W+jets$ backgrounds for the DY $WW$ analysis region. . . . .	137
13.7 Fitted background normalizations for $t\bar{t}$ and $W+jets$ backgrounds for the DY $WZ$ analysis region. . . . .	137
13.8 Fitted background normalizations for $t\bar{t}$ and $W+jets$ backgrounds for the VBF $WW$ analysis region. . . . .	137
13.9 Fitted background normalizations for $t\bar{t}$ and $W+jets$ backgrounds for the VBF $WZ$ analysis region. . . . .	138

14.1 This table shows the best fit slope and intercept for the 2-d distribution of number of tracks vs. jet  $\ln(p_T)$  for the leading jet in the background and HVT  $Z'$  samples. The tagging efficiency is shown for the 90% working point in the last column. The background jets contain more gluons than the signal jets. Consequently, the best fit line for the background predicts larger values of the number of tracks in jets for the background than the considered signals. . . . 161

## Abstract

Search for  $WW$  and  $WZ$  Resonances in  $\ell\nu qq$  final states in  $pp$  collisions at

$\sqrt{s} = 13$  TeV with the ATLAS detector

by

Natasha Woods

This thesis presents a search for  $WW$  and  $WZ$  resonances using data from  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector, corresponding to an integrated luminosity of  $139 \text{ fb}^{-1}$ . Diboson resonances are predicted in a number of Standard Model (SM) extensions, such as Extended Gauge Models, and Extra dimensional models. This search looks for resonances where one  $W$  boson decays leptonically and the other  $W$  or  $Z$  boson decays hadronically. This search is sensitive to diboson resonance production via vector-boson fusion as well as quark-antiquark annihilation and gluon-gluon fusion mechanisms. No significant excess of events is observed with respect to the Standard Model backgrounds, and constraints on the masses of new  $W'$ ,  $Z'$ , and bulk-RS Gravitons are extended to up to 3.3 TeV, depending on the model. As the dominant backgrounds in this search contain gluons, classifying jets as quark-initiated or gluon-initiated would make this analysis more sensitive to new physics. Towards this end, this thesis provides a calibrated quark-gluon tagger based on the multiplicity of charged particles within a jet.

Loving Dedication

å

xx

## **Acknowledgments**

Proper acknowledgments of everyone else who helped you graduate. Write later.

# **Part I**

## **Introduction**

# <sup>3</sup> Chapter 1

## <sup>4</sup> Introduction

<sup>5</sup> In general, humanity has continually strived to understand the structure and  
<sup>6</sup> dynamics of reality for widely varying reasons. Each academic field uses a spe-  
<sup>7</sup> cific set of concepts and models to describe nature. Physics is one such field,  
<sup>8</sup> that uses mathematical objects to systematically develop testable models about  
<sup>9</sup> the universe. Currently, the most fundamental types particles are fermions and  
<sup>10</sup> bosons. Fermions are the particles that make up the "ordinary" matter of the  
<sup>11</sup> universe, while bosons are the quanta of the fundamental forces. The Standard  
<sup>12</sup> Model (SM) of particle physics describes the quantum behavior of three of the  
<sup>13</sup> four fundamental forces: electromagnetic, strong, and weak forces.

<sup>14</sup> The Standard Model has consistently described much of reality to an extreme  
<sup>15</sup> degree of accuracy. It has predicted cross sections for strong and electroweak  
<sup>16</sup> processes that span over ten orders of magnitude [see Fig. 3.1] and contains no  
<sup>17</sup> known logical inconsistencies. Despite the reality of the Standard Model, it still  
<sup>18</sup> fails to describe aspects of reality and suffers from aesthetic issues. The SM fails  
<sup>19</sup> to account for dark matter, dark energy, neutrino masses, the hierarchy of the  
<sup>20</sup> fundamental force strengths, and other issues that may have not been noticed  
<sup>21</sup> yet! This incompleteness may indicate that a more fundamental theory exists. It

22 is hoped that such a theory would address the aforementioned phenomena and  
23 the ad-hoc structure and parameter values of the SM. In particular the relative  
24 scales of the fundamental forces impose oddly fine-tuned SM parameters, unless  
25 there is additional structure at higher energies (e.g. between the electroweak and  
26 Planck scales). This and other theoretical arguments motivate the search for new  
27 physics at the TeV scale. The set of theories that hope to explain more of reality  
28 are known as Beyond the Standard Model theories (BSM). Many of these theories,  
29 if true, would revolutionize concepts of symmetry and space-time, which would  
30 be intrinsically meaningful.

31 To probe the physics at this high energy frontier, physicists often collide ener-  
32 getic particles that combine to produce massive particles, such as the Higgs boson  
33 and top quark. The more energetic the colliding particles are the more massive  
34 produced particles can be. Currently, the world's highest energy particle collider  
35 is the Large Hadron Collider (LHC) at the European Organization for Nuclear  
36 Research (CERN).

37 This thesis presents a search for  $WW$  and  $WZ$  resonances using data from  $pp$   
38 collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at CERN, corresponding  
39 to an integrated luminosity of  $139 \text{ fb}^{-1}$ . Diboson resonances are predicted in a  
40 number of BSM theories, such as Extended Gauge Models and Extra dimensional  
41 models. This search looks for resonances where one  $W$  boson decays leptonically  
42 and the other  $W$  or  $Z$  boson decays hadronically. This search is sensitive to  
43 diboson resonance production via vector-boson fusion as well as quark-antiquark  
44 annihilation and gluon-gluon fusion mechanisms (which will be collectively called  
45 non-VBF modes).

46 To search for these new resonances, Monte-Carlo simulations are used to model  
47 SM backgrounds and BSM signals. In these simulations, a series of optimized cuts

48 are used to create signal regions (SR) to identify the leptonic and hadronic decay  
49 products of the resonance, maximize signal acceptance, and minimize background  
50 contamination. In these regions, the resonance mass is calculated as the combined  
51 system mass of the leptonic and hadronic system. The expected resonance mass  
52 distribution from the simulated backgrounds and anticipated signal are compared  
53 to data to search for the existence of these BSM signals (also known as a "bump  
54 hunt"). Control regions enriched in the dominant backgrounds,  $t\bar{t}$  and  $W+\text{jets}$   
55 (TCR and WCR, respectively) are constructed to be orthogonal to SRs and used  
56 to determine the normalization of the  $t\bar{t}$  and  $W+\text{jets}$  backgrounds in SRs.

57 The VBF  $W'$  and  $Z'$  and ggF  $W'$  and  $Z'$  resonances studied have unique  
58 SR and CR selections to maximize analysis sensitivity. RS Graviton signals are  
59 probed using the same selections as the ggF  $Z'$  signal. Additionally, more mas-  
60 sive resonances are more likely to have boosted  $W/Z$  bosons. As the boost of  
61 the hadronically decaying boson increases the separation of its hadronic decay  
62 products decreases. When the hadronically decaying boson has sufficient boost,  
63 the two quarks will overlap and not be identified separately. For this reason, a  
64 set of "resolved" selections are used when the hadronic decay products are recon-  
65 structed separately, and "merged" selections when the decay products overlap and  
66 identified as a single object in the event. A  $W/Z$  tagger identifies merged jets as  
67 originating from a  $W/Z$  bosons based on jet substructure and mass cuts. However,  
68 the more boosted the jet is the less likely it is to pass the jet substructure cut, due  
69 to track merging. Consequently, the merged selection uses a high purity region  
70 (HP), which requires that the jet pass both cuts, and low purity (LP) region where  
71 the jet can fail the jet substructure cut.

72 The aforementioned SR definitions veto events with  $b$ -jets to minimize  $t\bar{t}$  con-  
73 tamination. However,  $b$ -jets are anticipated from  $W'$  resonances from the hadron-

74 ically decaying  $Z$  boson. To increase the signal acceptance of these resonances,  
75 a  $Z \rightarrow bb$  tagger is used to construct additional SR and CRs called the "tagged"  
76 regions (and "un>tagged" if the event fails the  $Z \rightarrow bb$  tagger).

77 For each signal model, the simulated and measured resonance mass distribu-  
78 tions in the relevant SR and CRs are combined to construct a likelihood. This  
79 likelihood is parameterized by the signal strength parameter,  $\mu$  and systematic  
80 uncertainties of the resonance mass distribution. This likelihood is used to quan-  
81 tify the likelihood of a certain signal model given the anticipated backgrounds and  
82 measured data.

83 No significant excess of events is observed with respect to the Standard Model  
84 backgrounds, and constraints on the masses of new  $W'$ ,  $Z'$ , and bulk-RS Gravi-  
85 tons are extended to up to 3.3 TeV, depending on the model. As the dominant  
86 backgrounds in this search contain gluons, classifying jets as quark-initiated or  
87 gluon-initiated would improve the sensitivity of this analysis to new physics. To-  
88 wards this end, this thesis provides a calibrated quark-gluon tagger based on the  
89 multiplicity of charged particles within a jet.

90 Part II reviews the SM, its successes and shortcomings, and the aforemen-  
91 tioned BSM theories that address the incompleteness of the SM. The structure  
92 and performance of the ATLAS detector used is given in Part III. Part IV summa-  
93 rizes the search for the diboson resonances using ATLAS data from  $pp$  collisions  
94 at  $\sqrt{s} = 13$  TeV. Finally, Part V examines the prospects for a quark gluon tagger  
95 based on the track multiplicity of jets and details the calibration of this tagger.

**Part II**

96

**Theoretical Motivation**

97

<sub>98</sub> **Chapter 2**

<sub>99</sub> **The Standard Model of Particle**

<sub>100</sub> **Physics**

<sub>101</sub> **2.1 Introduction**

<sub>102</sub> By determining the dynamics of the most elementary degrees of freedom, par-  
<sub>103</sub> ticle physics hopes to uncover the fundamental laws of the universe. The definition  
<sub>104</sub> of elementary has evolved through time and currently refers to matter and force  
<sub>105</sub> mediating particles: fermions and bosons, respectively. The Standard Model of  
<sub>106</sub> Particle Physics (SM) describes the quantum behavior of three of the four funda-  
<sub>107</sub> mental forces: weak, strong, and electromagnetic, via boson and fermion interac-  
<sub>108</sub> tions. Gravity is not included in the SM and still under investigation.

<sub>109</sub> **2.2 Quantum Field Theory**

<sub>110</sub> In the SM, forces (and particles) are represented as fields. In this context,  
<sub>111</sub> fields are mathematical objects that define a tensor (e.g. scalar, vector, etc) at  
<sub>112</sub> every point on a manifold, here the manifold is space-time. These fields obey laws

<sub>113</sub> dictated by Quantum Field Theory (QFT). Particles arise naturally in QFT as  
<sub>114</sub> quantized field excitations localized in spacetime.

<sub>115</sub> According to Noether's theorem, symmetries of a field give rise to conserved  
<sub>116</sub> quantities (e.g. time-translation invariance leads to energy conservation). Often  
<sub>117</sub> in the history of physics, a conserved quantity of a field is found and then the  
<sub>118</sub> underlying symmetry of the field is inferred. Gauge symmetries are symmetries  
<sub>119</sub> among the internal degrees of freedom of the field (components of the tensor),  
<sub>120</sub> which give rise to quantities associated with fields. By specifying the symmetries  
<sub>121</sub> of a system the dynamics and conserved quantities of the system may be succinctly  
<sub>122</sub> defined.

## <sub>123</sub> 2.3 $U(1)_{EM}$ Local Gauge Invariance

<sub>124</sub> The Lagrangian of Quantum Electrodynamics (QED) describes the electro-  
<sub>125</sub> magnetic force. QED may be derived by requiring local  $U(1)_{EM}$  gauge invariance  
<sub>126</sub> of the free dirac fermion Lagrangian,  $\psi$ :

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \quad (2.1)$$

<sub>127</sub> This symmetry may be represented as a complex number with unit modulus,  
<sub>128</sub>  $e^{i\theta}$ .  $U(1)$  gauge invariance requires this gauge transformation of  $\psi$  will leave the  
<sub>129</sub> Lagrangian unchanged.

$$\psi(x) \rightarrow \psi'(x) = e^{i\theta(x)}\psi(x) \quad (2.2)$$

<sub>130</sub> NB: This transformation is a local gauge transformation as  $\theta$  depends on the  
<sub>131</sub> spacetime coordinate.

<sub>132</sub> By requiring this symmetry of the free Dirac fermion Lagrangian:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \quad (2.3)$$

<sub>133</sub> The mass term is unaffected, but the kinetic term is modified due to  $\theta(x)$ .

$$\mathcal{L} \rightarrow \mathcal{L}' = i\bar{\psi}e^{-i\theta(x)}\gamma^\mu\partial_\mu\psi e^{i\theta(x)} - m\bar{\psi}e^{-i\theta(x)}\psi e^{i\theta(x)} \quad (2.4)$$

<sub>134</sub>

$$= i\psi\gamma^\mu(\partial_\mu\psi + i\psi\partial_\mu\theta) - m\bar{\psi}\psi \quad (2.5)$$

<sub>135</sub> The  $\partial_\mu\theta$  terms breaks the gauge invariance of the Lagrangian. By introducing a  
<sub>136</sub> new field,  $A_\mu$  we can recover the gauge invariance of the derivative. Now redefining  
<sub>137</sub> the derivative as the covariant derivative:

$$D_\mu\psi \equiv (\partial_\mu - iqA_\mu)\psi \quad (2.6)$$

<sub>138</sub> And letting  $A_\mu$  transform under  $U(1)$  as:

$$A_\mu \rightarrow A_\mu + \delta A_\mu \quad (2.7)$$

<sub>139</sub> The transformed covariant derivative becomes:

$$D_\mu\psi \rightarrow D_\mu\psi' = (\partial_\mu - iqA_\mu)\psi' \quad (2.8)$$

<sub>140</sub>

$$= (\partial_\mu - iq(A_\mu + \delta A_\mu))\psi e^{i\theta} \quad (2.9)$$

<sub>141</sub>

$$= e^{i\theta}D_\mu + ie^{i\theta}\psi(\partial_\mu\theta - q\delta A_\mu) \quad (2.10)$$

<sub>142</sub> The covariant derivative can be made gauage invariant by setting the last term  
<sub>143</sub> to zero.

$$\delta A_\mu = \frac{1}{q} \partial_\mu \theta \quad (2.11)$$

<sup>144</sup> So now  $A_\mu$  transforms as:

$$A_\mu \rightarrow A_\mu + \frac{1}{q} \partial_\mu \theta \quad (2.12)$$

<sup>145</sup> Finally, replacing the derivative with the covariant derivative the Dirac La-  
<sup>146</sup> grangian we have:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (2.13)$$

<sup>147</sup>

$$= \mathcal{L}_{QED} \quad (2.14)$$

<sup>148</sup> Here  $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ . This last term in the Lagrangian is the kinetic  
<sup>149</sup> energy of the gauge boson field.

<sup>150</sup> So we have derived the QED Lagrangian. By requiring the free Dirac La-  
<sup>151</sup> grangian to be invariant under local U(1) transformations we have generated a  
<sup>152</sup> new gauge boson field,  $A_\mu$ , which describes the photon. As expected the photon  
<sup>153</sup> interacts with fermions.

<sup>154</sup> Stepping back, a global U(1) gauge symmetry of the free Dirac Lagrangian  
<sup>155</sup> implies we cannot measure the absolute phase of a charged particle. A local U(1)  
<sup>156</sup> gauge symmetry changes the phase of fields differently across space time. For this  
<sup>157</sup> type of transformation to leave the Lagrangian invariant, we had to introduce an  
<sup>158</sup> additional field,  $A_\mu$ , which "communicates" these phase changes across space-time.  
<sup>159</sup> In less formal language this effectively means: if the field at one location changes,  
<sup>160</sup> this change is conferred to other particles via  $A_\mu$ .

## <sup>161</sup> 2.4 Yang-Mills Gauge Theories

<sup>162</sup> Requiring  $U(1)_{EM}$  gauge invariance of the free Dirac Lagrangian gave us QED.

<sup>163</sup> Requiring different gauge symmetries we can derive the structure of other inter-

<sup>164</sup> actions. Any gauge symmetry may be written as:

$$\psi_i \rightarrow \exp(i\theta^a T_{ij}^a) \psi_j \quad (2.15)$$

<sup>165</sup> Here  $\theta$  is a dimensionless real parameter and  $T$  is the generator of the gauge

<sup>166</sup> symmetry group. With this the covariant derivative can be written as:

$$D_\mu \psi_i \equiv \partial_\mu \psi_i + ig A_\mu^a T_{ij}^a \psi_j \quad (2.16)$$

<sup>167</sup> Then the gauge field must transform as:

$$A_\mu^a \rightarrow A_\mu^a - \frac{1}{g} \partial_\mu \theta^a - f^{abc} \theta^b A_\mu^c \quad (2.17)$$

<sup>168</sup> Here  $f$  is the structure constant of the gauge group. The field strength tensor

<sup>169</sup> is given by:

$$F_{\mu\nu}^a \equiv \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g f^{abc} A_\mu^b A_\nu^c \quad (2.18)$$

<sup>170</sup>

$$F_{\mu\nu}^a \rightarrow F_{\mu\nu}^a - f^{abc} \theta^b F_{\mu\nu}^c \quad (2.19)$$

<sup>171</sup> This gives the Yang-Mills Lagrangian:

$$\mathcal{L}_{YM} = -\frac{1}{4} F_{\mu\nu}^{a\mu\nu} F_{\mu\nu}^a + i \bar{\psi}_i \gamma^\mu D_\mu \psi_i + m \bar{\psi}_i \psi_i \quad (2.20)$$

## <sup>172</sup> 2.5 Particles in the Standard Model

<sup>173</sup> The SM consists of fermions (half-integer spin matter constituents) and bosons  
<sup>174</sup> (integer spin force mediators). Fermions are spinor representations of the Poincare  
<sup>175</sup> group and can be further separated into leptons and quarks. Bosons are the result  
<sup>176</sup> of requiring a particular symmetry among the spinor fields:

$$SU(3)_C \times SU(2)_L \times U(1)_Y \quad (2.21)$$

<sup>177</sup>  $SU(3)_C$  is the symmetry group of the strong force and generates eight gluon  
<sup>178</sup> fields,  $G_\mu$ .  $SU(2)_L$  is the symmetry group of the Electroweak force and generates  
<sup>179</sup> three electroweak boson fields. The mixing of this  $SU(2)_L$  and  $U(1)_Y$  gives rise  
<sup>180</sup> to the photon field, where Y is the weak-hypercharge:

$$Y = 2(Q - T_3) \quad (2.22)$$

<sup>181</sup> Q is the electromagnetic charge, and  $T_3$  is the z-component of the weak isospin.  
<sup>182</sup> Weak isospin is the charge associated with the  $SU(2)_L$  symmetry. The correspond-  
<sup>183</sup> ing covariant derivative is then:

$$D_\mu \phi \equiv (\partial_\mu + ig_1 B_\mu Y_{L/R} + [ig_2 W_\mu^\alpha T^\alpha]_L + [ig_3 G_\mu^\alpha \tau^\alpha]_C) \psi \quad (2.23)$$

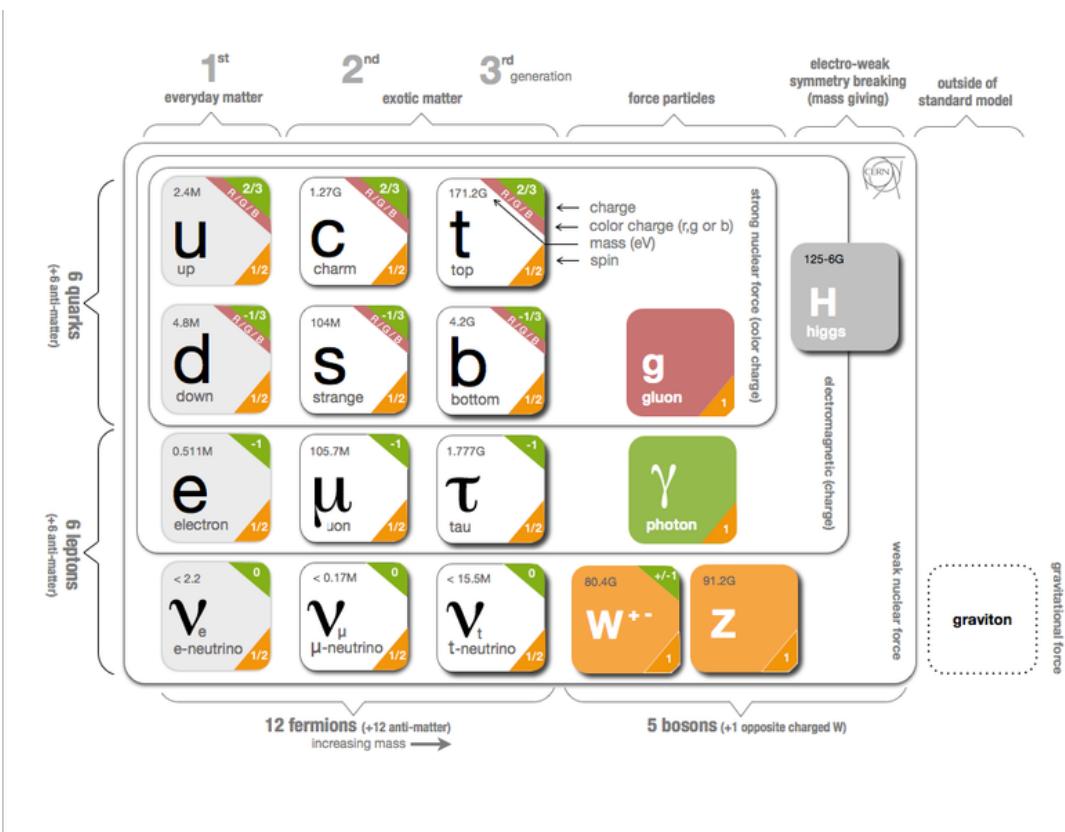
<sup>184</sup> It is important to note that the gauge symmetry of the SM yields a particular  
<sup>185</sup> structure of the fermion representations. So for a given fermion to interact with  
<sup>186</sup> a given gauge field it must have a non-zero corresponding Noether charge for  
<sup>187</sup> that gauge symmetry. If the corresponding Noether charge is zero, that fermion  
<sup>188</sup> transforms as a singlet and does not participate in that gauge interaction.

<sup>189</sup> Fermions are divided into quarks and leptons based on their transformations  
<sup>190</sup> under  $SU(3)_C$ . Quarks transform as color triplets. Leptons transform as color

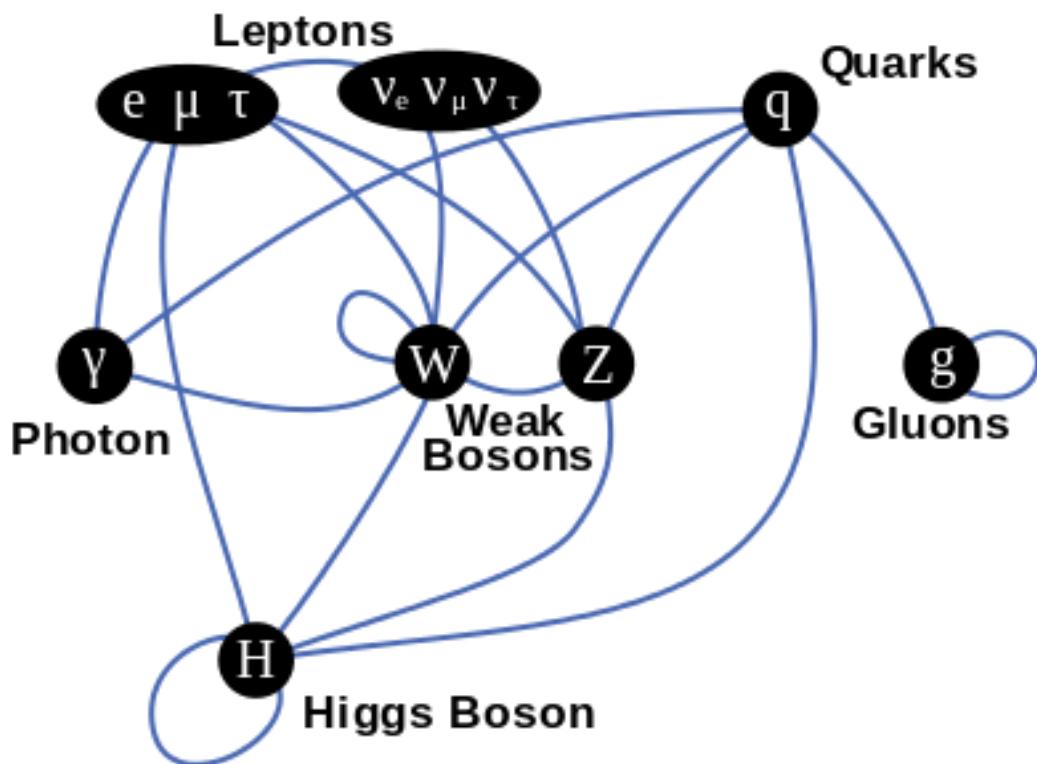
singlets and consequently do not interact with gluons. Fermions may be further  
 classified by their  $SU(2)_L$  interactions. Only the left-chiral part of fermions (denoted by L here) transform as  $SU(2)_L$  doublets, the right-chiral part forms singlets under this gauge. Lastly, all these groups of particles come in three generations, each a heavier copy of the previous, but with differing flavor quantum numbers.  
 This is summarized in Table 2.1 and shown in Figures 2.1 and 2.2.

SM Fermion Gauge Group	First Generation	Second Generation	Third Generation	$(SU(3)_C, SU(2)_L, U(1)_Y)$ Representations
Left-handed quarks	$\begin{pmatrix} u_L^r & u_L^g & u_L^b \\ d_L^r & d_L^g & d_L^b \end{pmatrix}$	$\begin{pmatrix} c_L^r & c_L^g & c_L^b \\ s_L^r & s_L^g & s_L^b \end{pmatrix}$	$\begin{pmatrix} t_L^r & t_L^g & t_L^b \\ b_L^r & b_L^g & b_L^b \end{pmatrix}$	$(3, 2, \frac{1}{6})$
Right-handed quarks	$(u_R^r, u_R^g, u_R^b)$ $(d_R^r, d_R^g, d_R^b)$	$(c_R^r, c_R^g, c_R^b)$ $(s_R^r, s_R^g, s_R^b)$	$(t_R^r, t_R^g, t_R^b)$ $(b_R^r, b_R^g, b_R^b)$	$(3, 1, \frac{2}{3})$ $(3, 1, -\frac{1}{3})$
Left-handed leptons	$\begin{pmatrix} \nu_e^L \\ e_L \end{pmatrix}$	$\begin{pmatrix} \mu_e^L \\ \mu_L \end{pmatrix}$	$\begin{pmatrix} \tau_e^L \\ \tau_L \end{pmatrix}$	$(1, 2, -\frac{1}{2})$
Right-handed leptons	$e_R$	$\mu_R$	$\tau_R$	$(1, 1, -1)$

**Table 2.1:** Representations of the SM fermions under  $SU(3)_C \times SU(2)_L \times U(1)_Y$  gauge symmetry group. Rows are correspond to different weak isospin states and columns to different QCD color states.



**Figure 2.1:** The particles of the Standard Model.



**Figure 2.2:** Summary of how Standard Model particles interact with other Standard Model particles.

197 Now we can understand the SM Lagrangian density as a Yang-Mills theory  
 198 with the gauge group:  $SU(3)_C \times SU(2)_L \times U(1)_Y$  with an additional  $SU(2)$  complex  
 199 scalar Higgs field doublet that will be discussed later.

$$\begin{aligned} \mathcal{L}_{SM} = & \underbrace{-\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4}G_{\mu\nu}^\alpha G^{\alpha\mu\nu}}_{\text{Kinetic Energies and Self-Interactions of Gauge Bosons}} \\ & + \underbrace{\bar{L}_i \gamma^\mu (i\partial_\mu - \frac{1}{2}g_1 Y_{iL} B_\mu - \frac{1}{2}g_2 \sigma^a W_\mu^a) L_i}_{\text{Kinetic Energies and EW Interactions of Left-handed Fermions}} \\ & + \underbrace{\bar{R}_i \gamma^\mu (i\partial_\mu - \frac{1}{2}g_1 Y_{iR} B_\mu) R_i}_{\text{Kinetic Energies and EW Interactions of Right-Handed Fermions}} \\ & + \underbrace{\frac{ig_3}{2} \bar{Q}_j \gamma^\mu \lambda^\alpha G_\mu^\alpha Q_j}_{\text{Strong Interactions between Quarks and Gluons}} \\ & + \underbrace{\frac{1}{2} |(i\partial_\mu - \frac{1}{2}g_1 B_\mu - \frac{1}{2}g_2 \sigma^a W_\mu^a)\Phi|^2 - V(\Phi)}_{\text{Electroweak Boson Masses and Higgs Couplings}} \\ & - \underbrace{(y_{kl}^d \bar{L}_k \Phi R_l + y_{kl}^u \bar{R}_k \tilde{\Phi} L_l + h.c.)}_{\text{Fermion Mass terms and Higgs Couplings}} \end{aligned}$$

200 Here several abstract spaces are being spanned:

- 201 – a spans the three  $SU(2)_L$  gauge fields with generators expanded in Pauli  
 202 matrices,  $T^\alpha = \frac{1}{2}\sigma^\alpha$
- 203 –  $\alpha$  spans the eight  $SU(3)_C$  gauge fields, with generators expanded in Gell-  
 204 Mann matrices,  $\tau^\alpha = \frac{1}{2}\lambda^\alpha$
- 205 – L/R represent left and right projections of Dirac fermion fields. The Strong  
 206 interaction is not chiral, so  $Q = L+R$

- 207 –  $\mu$  and  $\nu$  are four-vector indices
- 208 –  $i, j, k$  are summed over the three generations of SM particles.

## 209 2.6 Higgs Mechanism

210 The SM Lagrangian without the addition of a Higgs field does not allow for  
211 gauge boson and fermion mass terms:  $\frac{1}{2}m_A^2 A_\mu A_\mu$  and  $m(\bar{\psi}\psi)$ , as these terms are  
212 not gauge invariant. By introducing the Higgs field, mass terms for these particles  
213 may be included in a gauge invariant way. This field is a complex doublet with a  
214 potential  $V(\Phi)$ :

$$\Psi = \begin{pmatrix} \Phi^\dagger \\ \Phi^0 \end{pmatrix} \quad (2.24)$$

215

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda |\Phi^\dagger \Phi|^2 \quad (2.25)$$

216 The minima of this field occurs for  $|\Phi| = \sqrt{\frac{\mu^2}{2\lambda}} \equiv \frac{v}{2}$ . This yields degenerate  
217 minima, this symmetry is broken by choosing a specific minima (a.k.a. sponta-  
218 neous symmetry breaking). By convention  $\Phi_{min} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$  is chosen. This means  
219 the ground state of the Higgs field (Higgs vacuum) is non-zero,  $\sqrt{\frac{-\mu^2}{\lambda}}$ . The Higgs  
220 Field may now be expanded around this new ground state:

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (2.26)$$

221 This non-zero Higgs vacuum now generates mass terms for the gauge bosons  
222 from the following term in the Lagrangian:

$$|(-\frac{1}{2}g_1B_\mu - \frac{1}{2}g_2\sigma^aW_\mu^a)\Phi|^2 = \frac{1}{2}m_W^2W_\mu^+W^{-\mu} + \frac{1}{2}m_Z^2Z_\mu Z^\mu \quad (2.27)$$

223 where:

$$W_\mu^\pm \equiv \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \quad (2.28)$$

$$\begin{aligned} \text{224} \quad Z_\mu &\equiv \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2W_\mu^2 - g_1B_\mu) \end{aligned} \quad (2.29)$$

$$\begin{aligned} \text{225} \quad m_W &= \frac{vg_2}{\sqrt{2}} \end{aligned} \quad (2.30)$$

$$\begin{aligned} \text{226} \quad m_Z &= \frac{v}{\sqrt{2}}\sqrt{g_1^2 + g_2^2} \end{aligned} \quad (2.31)$$

227 The Higgs field also generates a mass term for the Higgs boson and self-  
228 interactions for the Higgs boson.

## 229 2.7 Electroweak Theory

230  $SU(2)_L$  generates  $W^\pm, W^0$  gauge bosons, which would be massless if  $SU(2)_L$   
231 was a perfect symmetry. These bosons are massive as this symmetry is broken.

232 The mass eigenstates,  $Z$  and  $\gamma$  given by:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \quad (2.32)$$

233 Here  $\theta_W$  is the Weinberg angle given by:

$$\cos\theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} = \frac{m_W}{m_Z} \quad (2.33)$$

## <sup>234</sup> 2.8 Quantum ChromoDynamics

<sup>235</sup> As mentioned earlier the Strong Force, which binds the proton together, is  
<sup>236</sup> mediated by gluons. Quantum Chromodynamics is the QFT which describes the  
<sup>237</sup> interactions of quarks and gluons via  $SU(3)_C$  symmetry. QCD contains features  
<sup>238</sup> not present in Electroweak Interactions due to  $SU(3)_C$  generators not commuting  
<sup>239</sup> (a.k.a.  $SU(3)_C$  is a non-abelian group) and the number of quark flavors ( $n_f$ ).  
<sup>240</sup> For example, in QCD there is color confinement and asymptotic freedom due to  
<sup>241</sup> the structure constants being non-zero. Requiring  $SU(3)_C$  local gauge invariance  
<sup>242</sup> implies:

$$\psi(x) \rightarrow \psi(x)' = \exp[i g_S \alpha(x) \cdot \hat{T}] \psi(x) \quad (2.34)$$

<sup>243</sup> where  $\alpha(x)$  is the local phase function,  $g_S$  is the strong coupling constant, and  
<sup>244</sup>  $\hat{T}$  are the eight generators of  $SU(3)$  (note  $\hat{T}^a = \frac{1}{2}\lambda^a a$ , where  $\lambda^a$  are the Gell-Mann  
<sup>245</sup> matrices). As the Gell-Mann matrices are 3x3, this means  $\psi$  has three degrees of  
<sup>246</sup> freedom under these  $SU(3)$  rotations. So we represent  $\psi$  under  $SU(3)$  rotations  
<sup>247</sup> as:

$$\psi = \begin{pmatrix} \psi_{red} \\ \psi_{green} \\ \psi_{blue} \end{pmatrix} \quad (2.35)$$

<sup>248</sup> Consequently, particle fields transforming under  $SU(3)$  rotations have three  
<sup>249</sup> components which physicists describe as color components (red, green, and blue).  
<sup>250</sup> A particle's corresponding antiparticle has the corresponding anticolor. This color  
<sup>251</sup> is the "charge" of QCD and is conserved under  $SU(3)$  rotations. Combining colors,  
<sup>252</sup> color neutral states (e.g. red and antired, or red, green and blue) may be created.  
<sup>253</sup> For the free Dirac Lagrangian to remain invariant under  $SU(3)$  transformations,

254 we must again postulate a boson field that modifies the derivative. The gluon  
255 field tensor is given by ( $\alpha = 1, \dots, 8$ ):

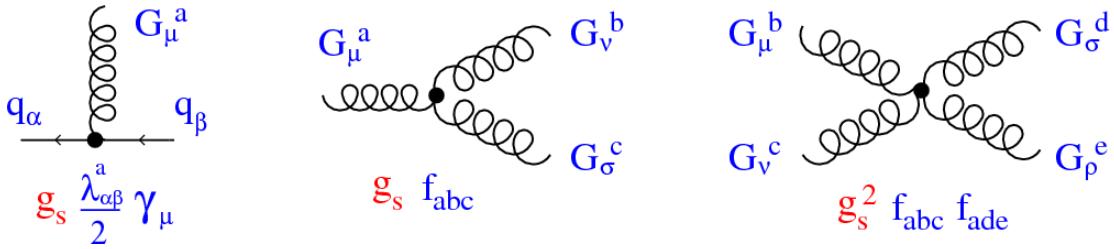
$$G_{\mu\nu}^k = \partial^\mu G_\alpha^\nu - \partial^\nu G_\alpha^\mu - g_S f^{\alpha\beta\gamma} G_\beta^\mu G_\gamma^\nu \quad (2.36)$$

256 Here  $f^{\alpha\beta\gamma}$  are the structure constants of  $SU(3)$ . Combining all this gives the  
257 QCD Lagrangian:

$$\mathcal{L}_{QCD} = \bar{\psi}_q i\gamma^\mu (D_\mu)_{ij} \psi^{qj} - m \bar{\psi}_q \psi_q - \frac{1}{4} G_{\mu\nu}^\alpha G^{\alpha\mu\nu} \quad (2.37)$$

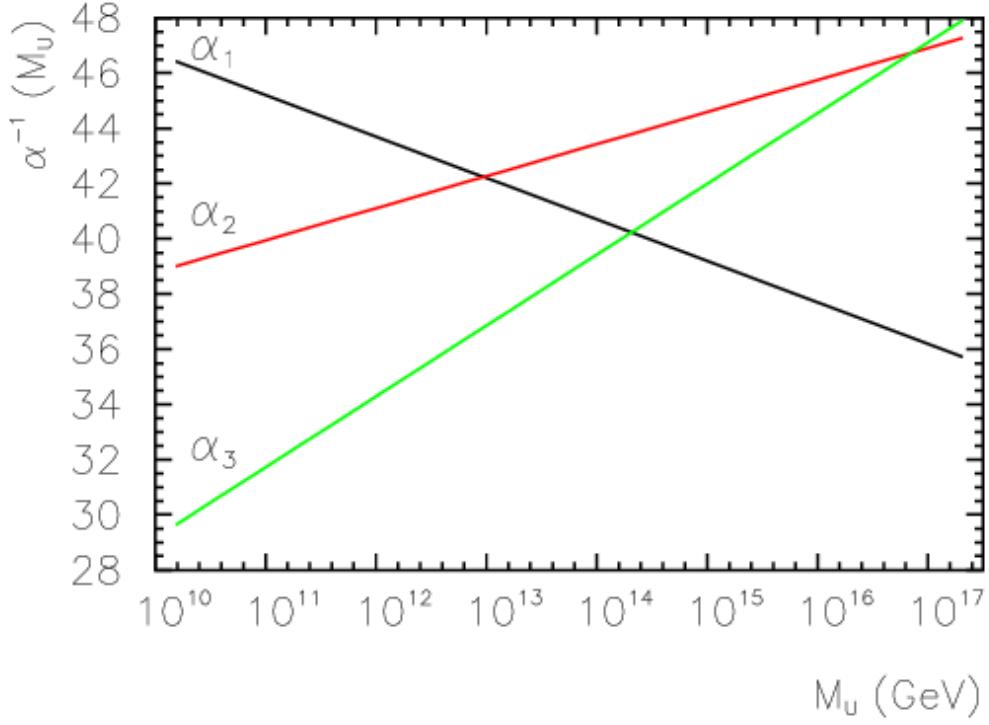
258 Here  $i$  are the color indices, and  $q$  are the quark flavors. It is important to  
259 note that quarks transform under the fundamental representation of  $SU(3)$ , while  
260 gluons transform under the adjoint representation. This means quarks carry a  
261 single color charge (red, green, blue, antired, antigreen, antiblue) and gluons carry  
262 a color and anticolor charge.

263 Figure 2.3 shows the three dominant QCD interactions. Since gluons carry  
264 color charge, they interact with one another. This does not occur in QED, as  
265 photons do not have electric charge and therefore do not interact with each other.  
266 In QED, a bare electron's effective charge is largest closest to the electron and  
267 decreases as a function of distance. This is because the QED vacuum fills with  
268 particle antiparticle pairs spontaneously, which screen the charge of the bare elec-  
269 tron. The larger the distance from the electron, the smaller the effective charge  
270 and therefore the weaker the force.



**Figure 2.3:** This figure shows the three dominant QCD interactions. From Ref. [18]

271        As the distance from a quark increases it's effective color charge increases due  
 272        to the vacuum polarization in QCD. Color charge grows as the distance from  
 273        the source increases (a.k.a. color is anti-screened in QCD). In this way, strong  
 274        interactions become stronger at large distances (low momenta interactions). At  
 275        small distances (large momenta interactions) strong interactions are significantly  
 276        weaker and considered nearly free. This effect of referred to as asymptotic freedom.  
 277        At large distances, a quark's effective charge is large and the strong force is more  
 278        significant. This force becomes so strong that quarks form colorless bound states  
 279        instead of remaining free particles. This effect is known as color confinement.  
 280        This running of all SM fields is shown in Figure 2.4.



**Figure 2.4:** Strength of the U(1), SU(2), and SU(3) gauge couplings as a function of the energy scale of the interaction ( $Q$ ). From Ref. [12]

Commonly the change in a particle's effective charge under a given force is quantified with  $\beta(r) \equiv -\frac{de(r)}{d\ln r}$ , where  $e(r)$  is the effective charge of a given particle under a force. In QED this function is positive but in QCD this function is negative leading to confinement and asymptotic freedom. Moreover, one can calculate how the coupling ( $\alpha$ ) of a force varies with energies. (More deeply this amounts to incorporating renormalization and vacuum polarization in the boson propagators).

For QCD this is:

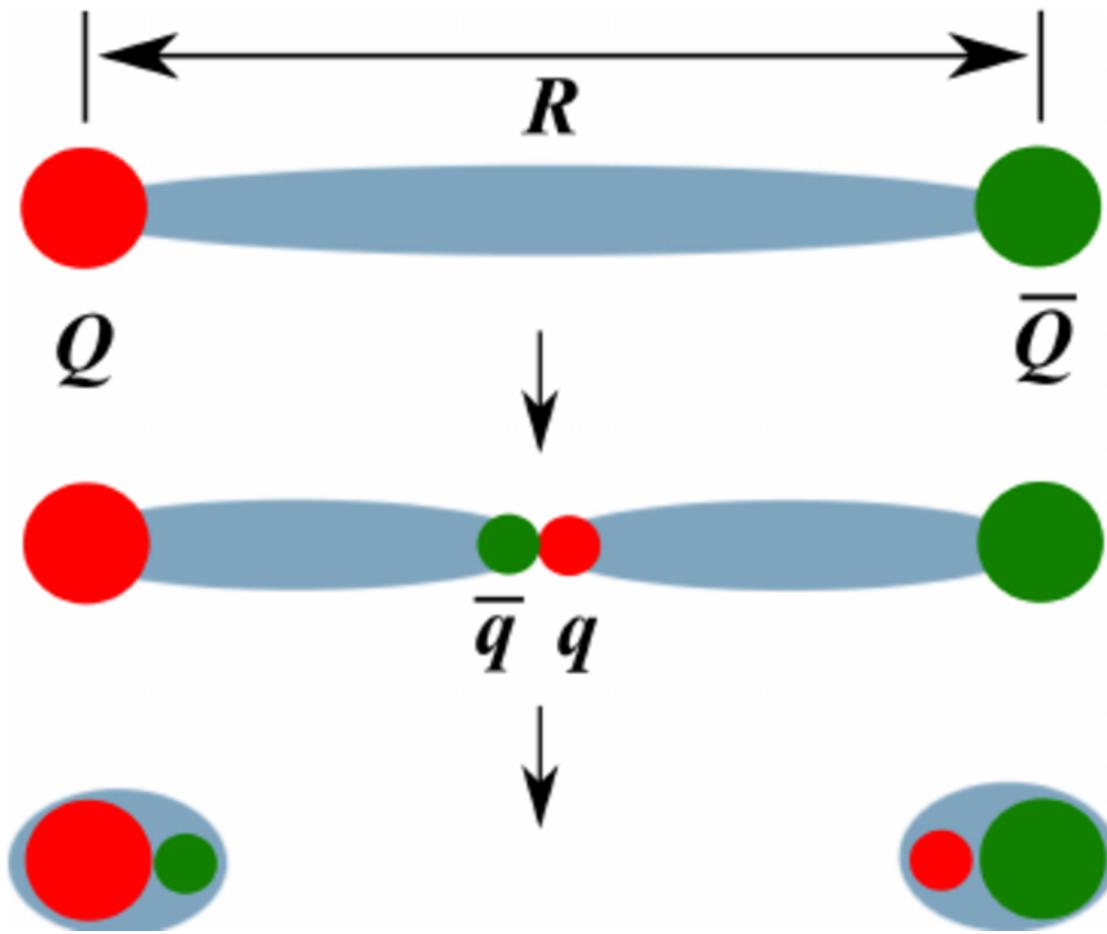
$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \frac{\alpha_s(\mu^2)}{12\pi}(33 - 2n_f)\ln(Q^2/\mu^2)} \quad (2.38)$$

where  $Q$  is the momentum of the force is probed at,  $\mu^2$  is the renormaliza-

289 tion scale,  $n_f$  is the number of quark flavors. There are six quark flavors in SM  
290 QCD, making  $33 - 2n_f > 0$ . This factor being positive and the  $\ln(Q^2/\mu^2)$  being  
291 in the denominator means that as  $Q^2$  increases  $\alpha_s$  decreases. So for large  $Q^2$ ,  $\alpha_s$   
292 is small and SM QCD is asymptotically free, while for small  $Q^2$ ,  $\alpha_s$  is large and  
293 SM QCD is confined, as mentioned earlier.

294 As stated previously, quarks and gluons have not been observed in isolation.  
295 Instead they form bound colorless states. Hadronization is the process by which  
296 quarks and gluons form hadrons. The process of hadronization is still an active  
297 area of research. One qualitative description is show in Figure 2.5. In this figure,  
298 as two quarks separate the color field between them is restricted to a tube with  
299 energy density of  $1\text{GeV}/\text{fm}$ . As they separate further, the energy in the color field  
300 increases, until there is enough energy to produce  $q\bar{q}$  pairs, which breaks the color  
301 field. This process repeats until quarks and antiquarks have low enough energy  
302 to form colorless hadrons. The resulting spray of hadrons is called a jet.

303 Since quarks and gluons carry different color charges, their respective jets have  
304 different properties. As quarks carry only a single color charge (vs. gluons which  
305 have color and anticolor charge), so their jets have less constituent particles. More  
306 precisely, the Altarelli-Parisi splitting functions [3] contain a factor  $C_A$  for gluon  
307 radiation off a gluon and  $C_F$  for gluon radiation off a quark ( $C_A/C_F = 9/4$ ). These  
308 color factors are the prefactor in the Feynman diagrams for these processes [1],  
309 which leads to gluon jets having more constituents and therefore more tracks than  
310 quark jets. Gluon jets also tend to have a larger radius with lower momentum  
311 constituents than quarks. There are many novel techniques to distinguish quarks  
312 from gluons. For this study the number of charged particles will be focused on.



**Figure 2.5:** A cartoon of string breaking: the QCD string spanned between quark  $Q$  and antiquark  $\bar{Q}$  breaks due to  $q\bar{q}$  creation [4]

<sup>313</sup> **Chapter 3**

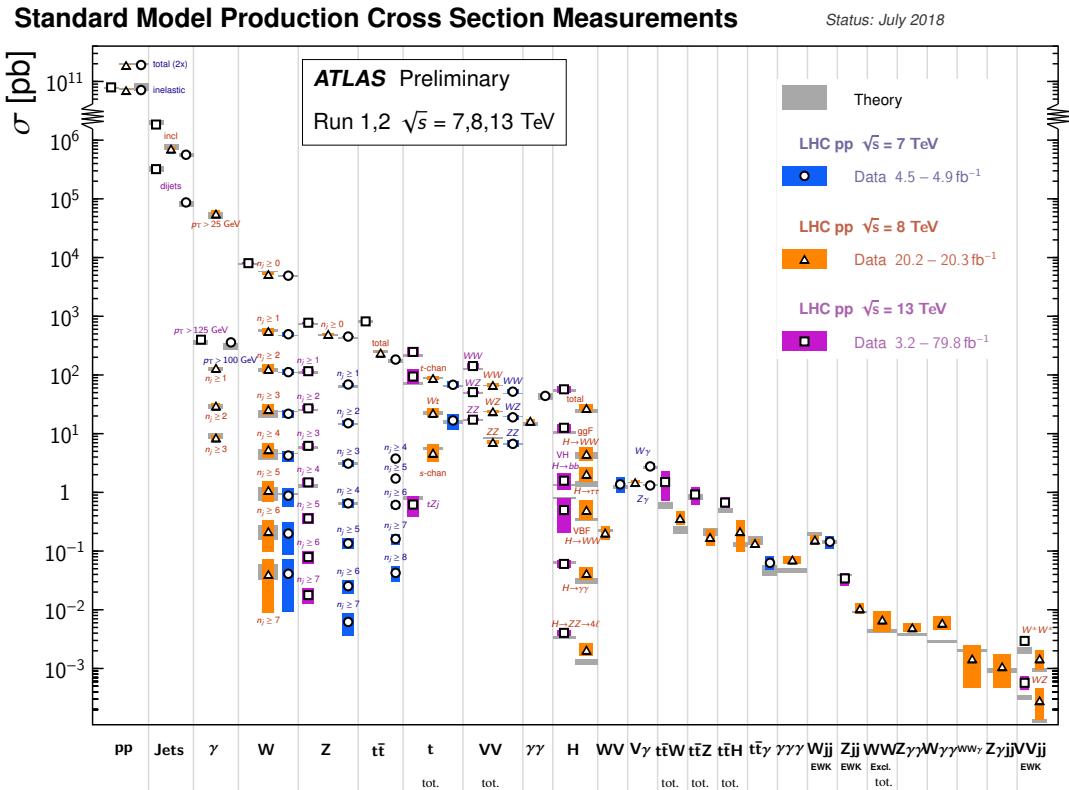
<sup>314</sup> **Standard Model Successes and  
Limitations**

<sup>316</sup> The Standard Model has accurately described most of the underlying principles  
<sup>317</sup> of nature. It has predicted cross sections for strong and electroweak processes that  
<sup>318</sup> span over ten orders of magnitude correctly [see Fig. 3.1] and contains no known  
<sup>319</sup> logical inconsistencies. Despite the strength and reality of the Standard Model, it  
<sup>320</sup> still fails to describe some important aspects of reality and suffers from aesthetic  
<sup>321</sup> issues. To date, dark matter and dark energy comprise  $\sim 95\%$  of the universe, but  
<sup>322</sup> the SM offers no explanation of their nature. Additionally, neutrinos are known  
<sup>323</sup> to have mass, but the SM offers no mass generation mechanism for left-handed  
<sup>324</sup> neutrinos without right-handed neutrinos (which do not exist). There are other  
<sup>325</sup> mechanisms for introducing massive neutrinos in the SM, but these mechanisms  
<sup>326</sup> create hierarchy problems.

<sup>327</sup> Possibly the most significant aesthetic issue is the hierarchy between the elec-  
<sup>328</sup> troweak and Planck scales. The electroweak scale is the scale of electroweak  
<sup>329</sup> symmetry breaking. The Planck scale is the scale where the gravitational force  
<sup>330</sup> is comparable in strength to the other forces. The Planck scale is where the SM

331 breaks down, as there is not an experimentally verified theory of quantum gravity,  
332 and at this scale gravity cannot be ignored (like it can at the electro-weak scale).  
333 These scales differ by  $\sim 30$  orders of magnitude. Understanding the difference  
334 in these energy scales may help explain the weakness of gravity at electroweak  
335 scales, and possibly a QFT for gravity. (NB: This hierarchy can also be framed in  
336 terms of the corrections to the Higgs mass, which depend on the UV cutoff scale -  
337 where the SM is suppose to break, which is taken at the Planck scale. This leads  
338 the quantum corrections to the Higgs mass that would force the Higgs mass to  
339  $\sim 10^{18}$  TeV.)

340 These stark contrasts in scales may indicate that a more fundamental theory  
341 exists. It is hoped that such a theory would explain and motivate some of the  
342 ad-hoc features of the SM. In particular, the values of the 19 SM parameters (6  
343 quark masses, 3 charged lepton masses, 3 gauge couplings, Higgs parameters ( $\mu^2$ ,  
344  $\lambda$ )), the structure of the fermion representations, etc.



**Figure 3.1:** A comparison of cross section measurements at  $\sqrt{s} = 7,8,13$  TeV from ATLAS compared to theoretical measurements. From Ref. [6]

<sup>345</sup> **Chapter 4**

<sup>346</sup> **New Physics Models with  
347 Diboson Resonances**

<sup>348</sup> **4.1 Randall Sundrum Bulk Model**

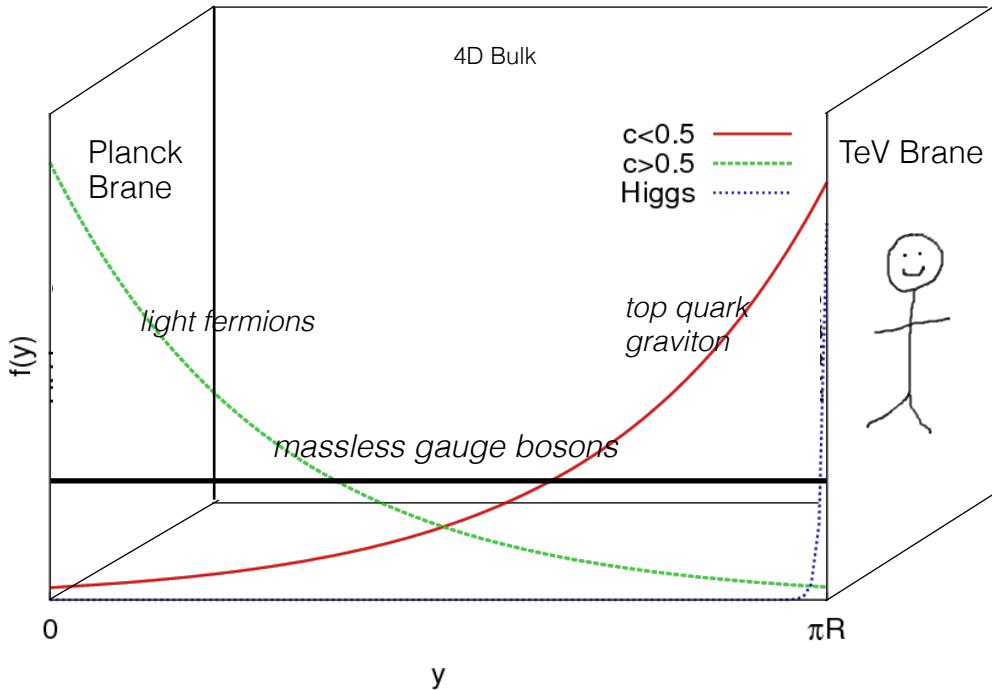
<sup>349</sup> The electroweak-planck hierarchy may be explained by the existence of extra  
<sup>350</sup> dimensions, like the 5D Randall Sundrum Bulk Model ([19], [2]). In this model,  
<sup>351</sup> there is one extra warped spatial dimension,  $y$ , with a metric:

$$ds^2 = e^{-2k|y|} \eta_{\mu\nu} dx^\mu dx^\nu + dy^2 \quad (4.1)$$

<sup>352</sup> where  $e^{-k|y|}$  is the warp factor of the extra dimension, which is compactified on  
<sup>353</sup> a  $S^1/Z_2$  orbifold (a.k.a. a circle where  $y \rightarrow -y$ ). This can be visualized as every  
<sup>354</sup> point in space time having a line extending from it a distance  $L$ , representing  
<sup>355</sup> this fifth dimension. At the end of this line is the Planck brane. This fourth  
<sup>356</sup> spatial dimension separates two 4-D branes: Planck brane and TeV brane. We  
<sup>357</sup> live on the TeV brane, as shown in Figure 4.1. The Higgs field (and to a lesser  
<sup>358</sup> degree the top quark and graviton fields) is localized near the TeV Brane, while

359 the light fermion fields are localized more near the Planck brane. Fundamental  
 360 parameters are set on the Planck brane. The warp factor may be scaled away from  
 361 all dimensionless SM terms by field redefinitions. However, the only dimensionful  
 362 parameter,  $m_H^2 = v^2$  is rescaled by  $\tilde{v} \sim e^{-kL} M_{Pl} \sim 1\text{TeV}$  for  $kL \sim 35$ , explaining  
 363 why gravity is so weak on the TeV brane. Also, by localizing the light fermion  
 364 fields near the Planck brane and top and graviton fields near the TeV brane, the  
 365 light quarks will have smaller masses.

366 The two free parameters of this theory are  $M_{Pl}$  and  $k$ . Based on this RS Bulk  
 367 model, all SM particles should have Kaluza-Klein (KK) excitations. In particular,  
 368 the graviton would have KK excitations that prefer to decay to WW or ZZ, which  
 369 is why this analysis searches for RS Gravitons.



**Figure 4.1:** Cartoon of RS Bulk Model

## 370 4.2 Simple Standard Model Extensions

371 The RS Bulk model is motivated by resolving SM hierarchies, but it does not  
372 address all of the other SM issues. There are many other interesting and well  
373 motivated new physics frameworks that address these issues, but there is a lack  
374 of completely predictive models, due to model flexibility (free parameters). It is  
375 difficult for experimentalists to know which theories to search for in data. There-  
376 fore, developing a model-independent resonance search that can be reinterperted  
377 in the context of a given BSM theory is ideal.

378 This search is sensitive to the resonance mass and its interactions, but not  
379 all of a given BSM model's parameters. Therefore, the BSM Lagrangian may be  
380 reduced to only retain this information (mass parameters and couplings) following  
381 the procedure in [17]. In this simplified approach, the new resonance searched for  
382 is represented as an additional heavy vector triplet (HVT), which is a real vector  
383 field in the adjoint representation of  $SU(2)_L$  with vanishing hypercharge. This  
384 results in one neutral and two charged bosons, defined as:

$$\small{385} \quad V^\pm = \frac{V_\mu^1 \mp iV_\mu^2}{\sqrt{2}} \quad (4.2)$$

$$V_\mu^0 = V_\mu^3 \quad (4.3)$$

386 The SM Lagrangian is then augmented with the additional terms:

$$\mathcal{L} \supset -\frac{1}{4}D_{[\mu}V_{\nu]}^a D^{[\mu}V^{\nu]}_a + \frac{m_V^2}{2}V_\mu^a V^{a\mu} + ig_V c_H V_\mu^a H^\dagger \tau^a \overset{\leftrightarrow}{D}{}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a J_F^{\mu a} \quad (4.4)$$

387 In order the terms represent: the kinetic,  $V$  mass, Higgs- $V$  interaction, and  
388  $V$ -left-handed fermion interaction terms. The  $g_V$  coupling factor determines the  
389 coupling of the new resonance to left-handed fermions and the Higgs boson.

390 As benchmark models, this search considers resonances from extended gauge  
391 symmetry (EGM) and composite Higgs models as discussed in [17] . The EGM  
392 model predicts weakly coupled resonances, where  $g_V = 1$ , referred to later as  
393 Model A. The composite Higgs Model is a strongly coupled model, where  $g_V = 3$ ,  
394 and later referred to as Model B. As shown in Eq. 4.4, the coupling of these  
395 resonances to fermions scales as  $g_f = g^2 c_F / g_V$ , where  $g$  is the SM  $SU(2)_L$  gauge  
396 coupling and  $c_F$  is a free parameter. This then means that for Model B the  
397 coupling to fermions is suppressed relative to Model A, leading to a smaller DY  
398 production rate and branching ratio (BR) to fermionic final states. The coupling  
399 of  $V$  to SM bosons scales as  $g_H = g_V c_H$ , where  $c_H$  is a free parameter on the  
400 order of one for Model A and B. Consequently Model A resonances have a smaller  
401 the BR to gauge bosons than Model B. For the  $pp$  collision data used, Model A  
402 predicts larger production cross sections decaying to leptons and fermions than  
403 Model B which decays primarily to gauge bosons.

404 Model A and B vectors are produced via quark-anti-quark annihilation and  
405 the more rare vector-boson-fusion is considered by setting  $g_H = 1$  and  $g_F = 0$ .  
406 Both production modes are probed in this resonance search.

407 In summary,  $V$  couples most strongly to left-handed fermions and  $VV$  depen-  
408 dent on  $g_V$ .

## Part III

409

## Experimental Setup

410

<sup>411</sup> **Chapter 5**

<sup>412</sup> **LHC**

<sup>413</sup> The Large Hadron Collider (LHC) is the highest-energy particle collider in the  
<sup>414</sup> world. It was designed to expand the frontier of high energy particle collisions in  
<sup>415</sup> energy and luminosity. This enables LHC experiments to test the Standard Model  
<sup>416</sup> and search for new physics at higher energies than tested with previous colliders.  
<sup>417</sup> Collisions at higher energies not only produce more massive particles but also  
<sup>418</sup> more weakly interacting particles. Fig. 5.1 shows production cross sections for  
<sup>419</sup> various processes at hadron colliders. The rate for electroweak physics processes  
<sup>420</sup> including  $W$  and  $Z$  scale with the center-of-momentum energy,  $\sqrt{s}$ .

<sup>421</sup> The LHC consists of a 26.7 km (17 miles) ring, approximately 100 m un-  
<sup>422</sup> derground, outside Geneva, Switzerland. Counter-circulating proton (and occa-  
<sup>423</sup> sionally heavy ions) beams collide inside four experiments along the beam line:  
<sup>424</sup> ATLAS, CMS, LHCb, ALICE. ATLAS and CMS are general purpose detectors de-  
<sup>425</sup> signed to explore the high energy frontier. LHCb is designed to study the physics  
<sup>426</sup> of  $b$ -quarks. ALICE specializes in studying heavy ion collisions.

## proton - (anti)proton cross sections

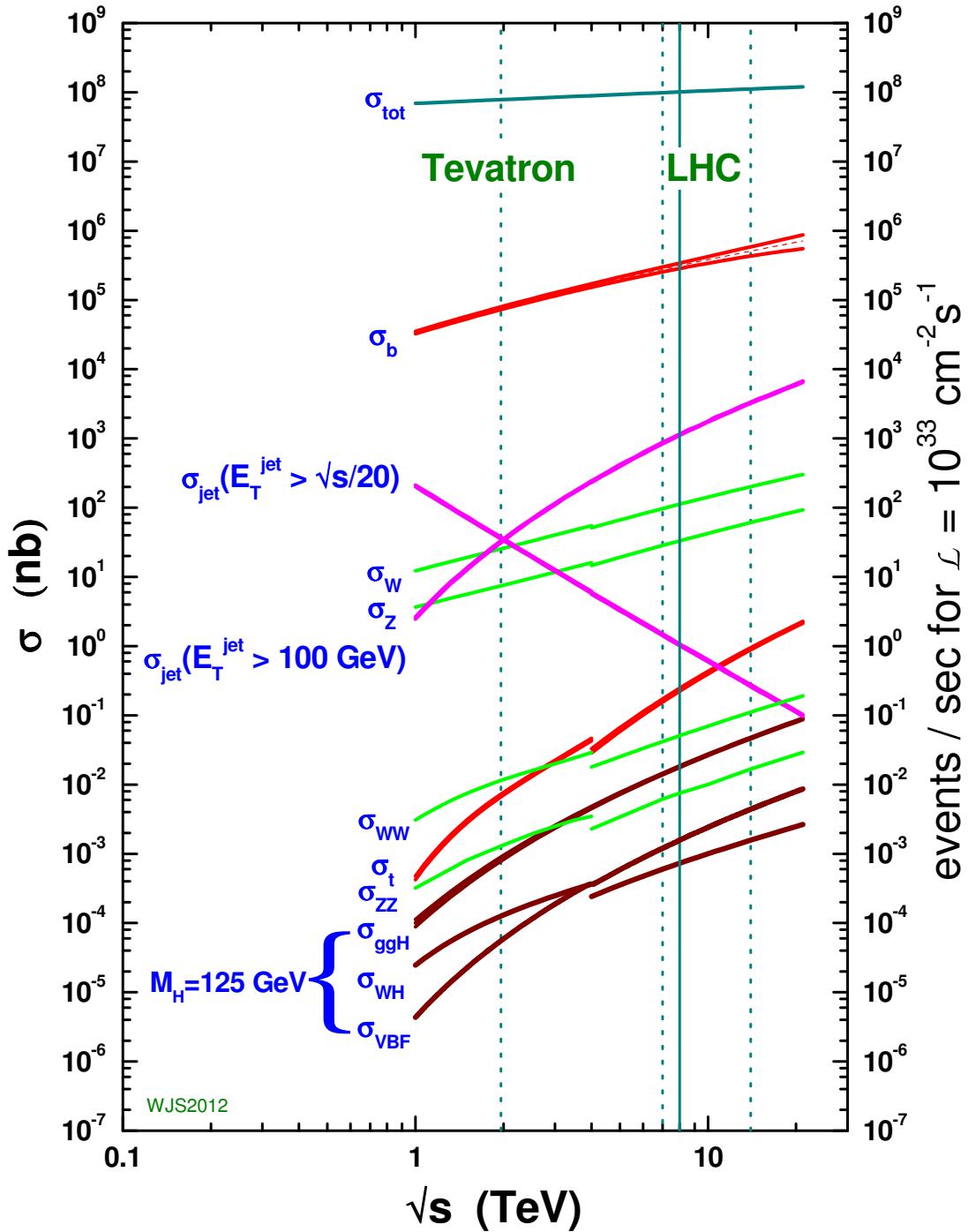
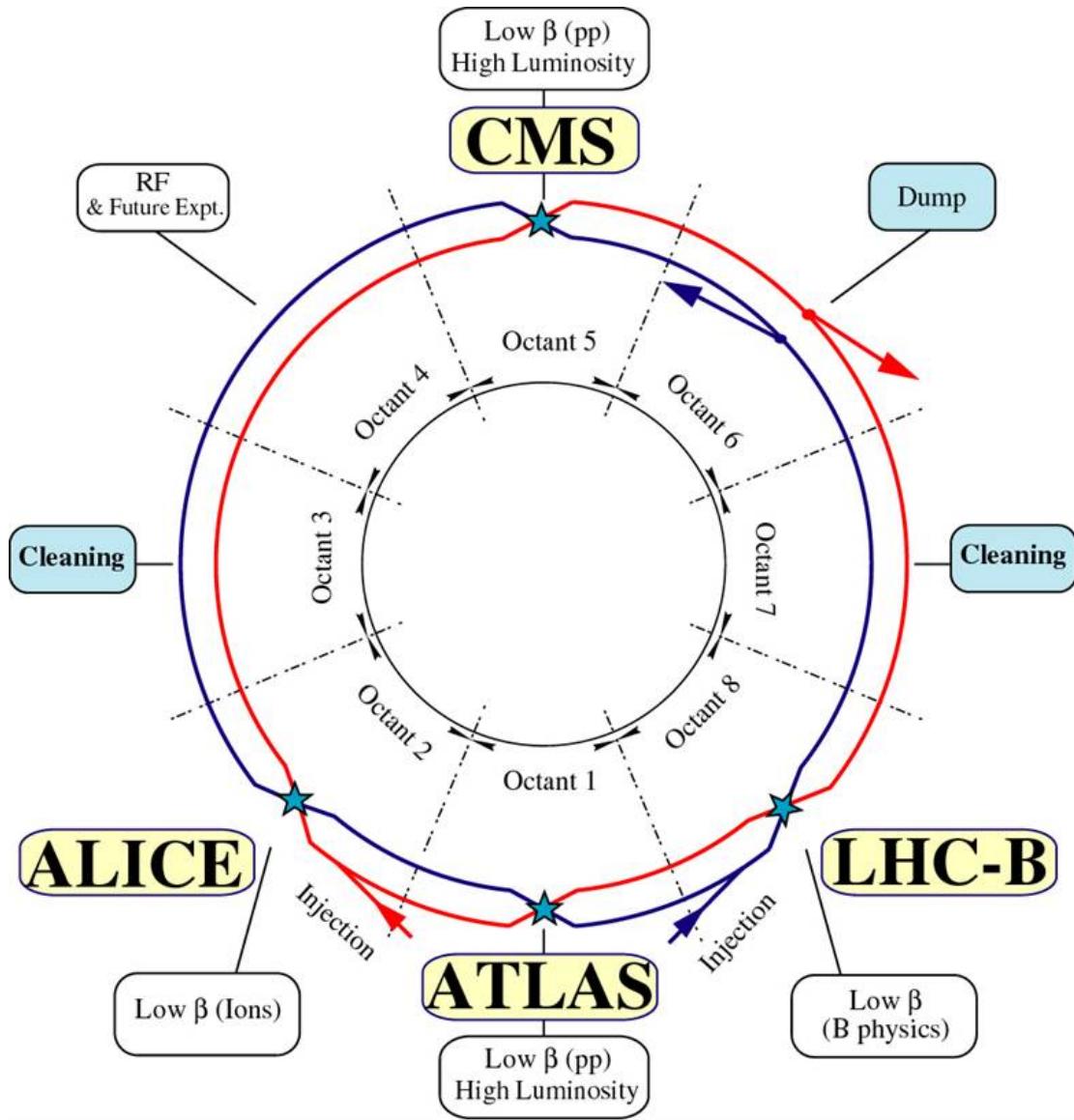


Figure 5.1: Scaling of various SM cross sections with  $\sqrt{s}$ .

427        The first proton beams circulated in September, 2008. Nine days later an elec-  
428        trical fault lead to mechanical damage and liquid helium leaks in the collider. This  
429        incident delayed further operation until November 2009, when the LHC became  
430        the world's highest energy particle collider, at 1.18TeV per beam. This first oper-  
431        ational run continued until 2013, reaching 7 and 8 TeV collision energies. During  
432        this run a particle with properties consistent with the Standard Model Higgs bo-  
433        son was discovered. The next run began after a two year shutdown after upgrades  
434        to the LHC and experiments. This run lasted from 2013 to 2018 reaching 13 TeV  
435        collision energies. This analysis uses data from the second operational run.

## 436        5.1 LHC Layout and Design

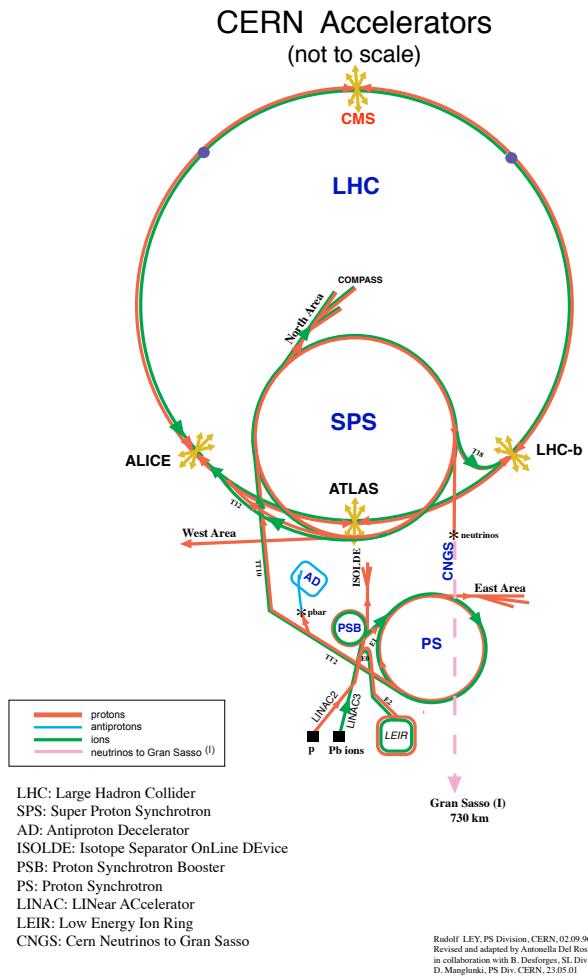
437        The layout of the LHC is shown in Figure 5.2. The red and blue lines in the  
438        figure represent the counter-circulating proton beams. The LHC is divided into  
439        eight octants. Octant 4 contains the RF cavities that accelerate the protons and  
440        octant 6 contains the beam dump system. Octants 3 and 7 house the collimation  
441        systems for beam cleaning. The beams collide inside the four aforementioned  
442        experiments. Each octant contains a curved and straight section. The LHC  
443        magnets are built with NbTi superconductors cooled with super-fluid Helium to  
444        2K, creating a 8.3T magnetic field to bend the proton beams.



**Figure 5.2:** The layout of the LHC and the four detectors along the beam line (ATLAS, LHCb, ALICE, CMS).

Four sequential particle accelerators are used to accelerate protons from rest as shown in Figure 5.3. First, Hydrogen gas is ionized to produce protons which are then accelerated to 50 MeV using Linac 2, a linear accelerator. The resulting proton beam is then passed to three circular particle accelerators: Proton Synchrotron Booster, Proton Synchrotron, and Super Proton Synchrotron (SPS),

450 accelerating protons to 1.4, 25, and 450 GeV, respectively. Once the protons exit  
 451 the SPS, they are injected into the LHC at octant 2 and 8. Each proton bunch  
 452 contains  $\sim 10^{11}$  protons. The spacing between bunches is 25 ns, which means  
 453 each beam contains 3564 bunches. However, some bunches are left empty due  
 454 to injection and safety requirements, yielding 2808 bunches per beam. Once the  
 455 proton beams are injected they are accelerated to 13 TeV.



**Figure 5.3:** An overview schematic of the LHC accelerator subsystems.

456 As many new physics models predict cross-sections below the weak scale it was  
 457 important to design the LHC to be capable of collecting enough data, by running

458 in high luminosity conditions. The machine luminosity depends only on beam  
459 parameters:

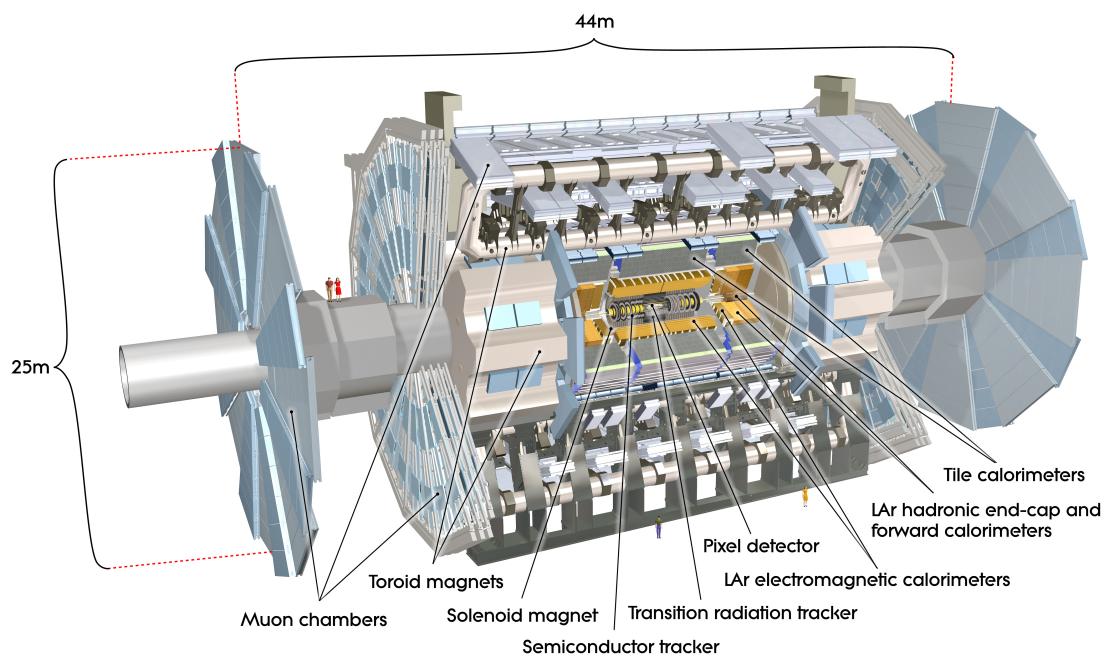
$$L = \frac{N_p^2 f}{4\epsilon\beta^*} F \quad (5.1)$$

460 where  $N_p$  is the number of protons per bunch,  $f$  is the bunch crossing frequency,  
461  $\epsilon$  is the transverse beam emittance,  $\beta^*$  is the amplitude function at the collision  
462 point, and  $F$  is the geometric luminosity reduction factor due to the beams crossing  
463 at an angle (rather than head-on).

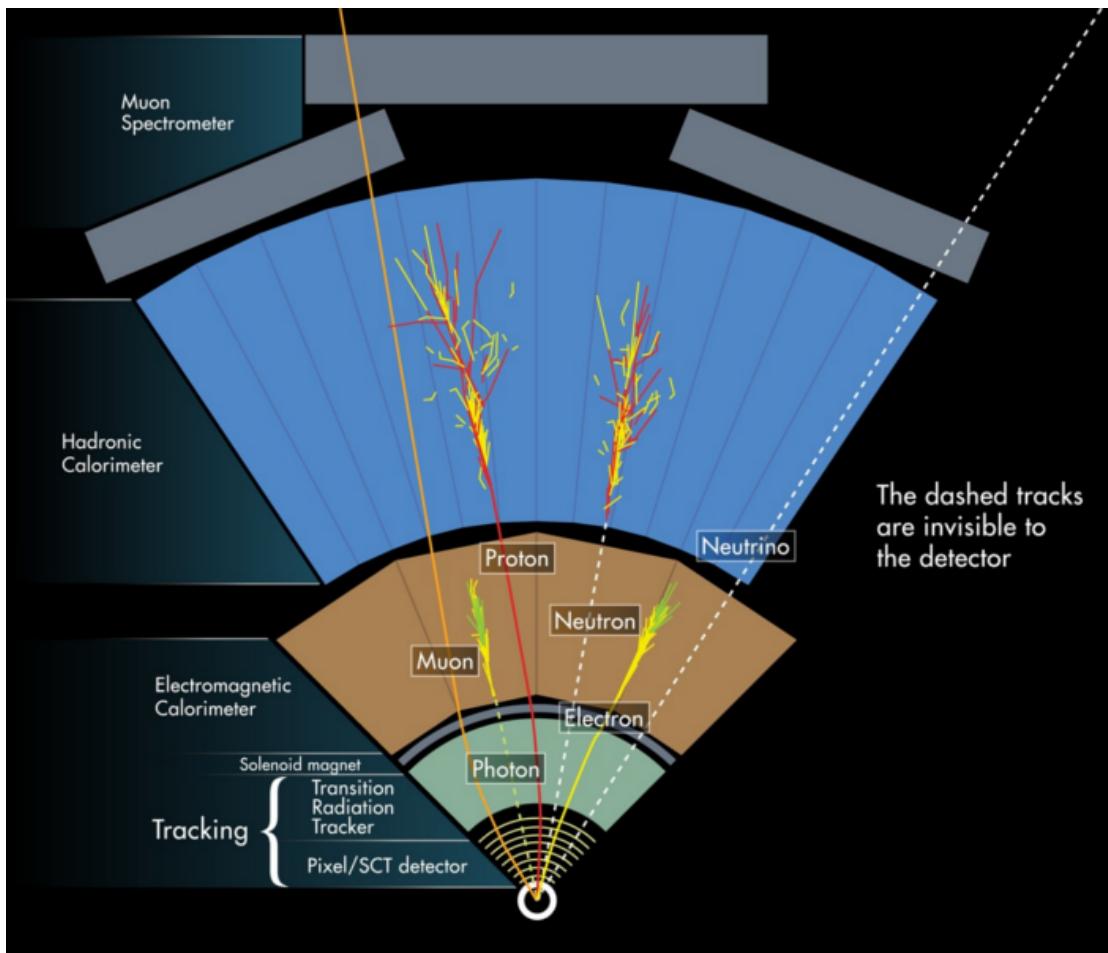
464 **Chapter 6**

465 **The ATLAS Detector**

466 The ATLAS detector measures the position, momentum and energy of parti-  
467 cles produced in the proton collisions by using magnetic fields, silicon detectors,  
468 sampling calorimeters, and gaseous wire detectors. It is located approximately  
469 100 m underground at Point-1 around the LHC beam line and weighs 7000 metric  
470 tons. The detector is 46 m long, 25 m high, 25 m wide as shown in Figure 6.1.  
471 The detector can be divided into three subsystems: the Inner Detector (ID), the  
472 Calorimeters, and the Muon Spectrometer (MS). Figure 6.2 shows an overview of  
473 how different particles interact in the detector.



**Figure 6.1:** Overview schematic of the ATLAS detector.



**Figure 6.2:** A simplified schematic of how different particles interact and are detected within ATLAS.

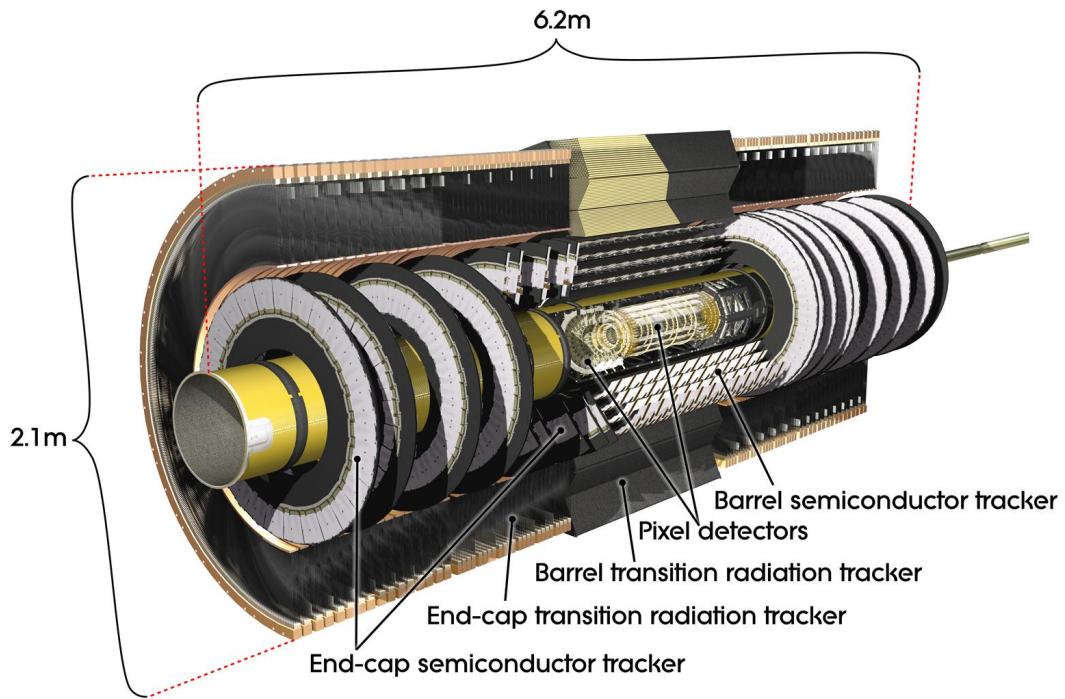
## 474 6.1 Coordinate System

475 The trajectory of particles within ATLAS is measured relative to the nominal  
476 interaction point. The  $z$ -axis points along the beam line, such that when the  
477 LHC is viewed from above, the counter-clockwise circulating beam points along  
478 the positive- $z$  direction. The  $x - y$  plane is transverse to the beam line, with the  
479 positive  $x$ -axis pointing towards the center of the LHC ring. The positive  $y$ -axis  
480 points vertically upward. The azimuthal angle,  $\phi$ , is the angular distance about  
481 the  $z$ -axis, with  $\phi = 0$  along the  $x$ -axis. The polar angle from the  $z$ -axis is denoted  
482 as  $\theta$ . However, this quantity is not Lorentz invariant, like rapidity,  $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$ ,  
483 where  $E$  is the energy of the particle considered, and  $p_z$ , is it's momentum along  
484 the  $z$ -axis. Pseudorapidity is preferred as  $\Delta\eta$  is invariant under boosts along  $z$   
485 and particle production is approximately invariant under  $\eta$ . For massless particles,  
486 rapidity and a related quantity, pseudorapidity, are the identical. The pseudora-  
487 pidity is defined as:  $\eta = -\ln \tan(\frac{\theta}{2})$ . This quantity is preferred as it is purely a  
488 geometric quantity, independent of particle energy. Angular separation between  
489 particles in ATLAS are given by  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ . The distance from the  
490 beamline is given by  $r = \sqrt{x^2 + y^2}$

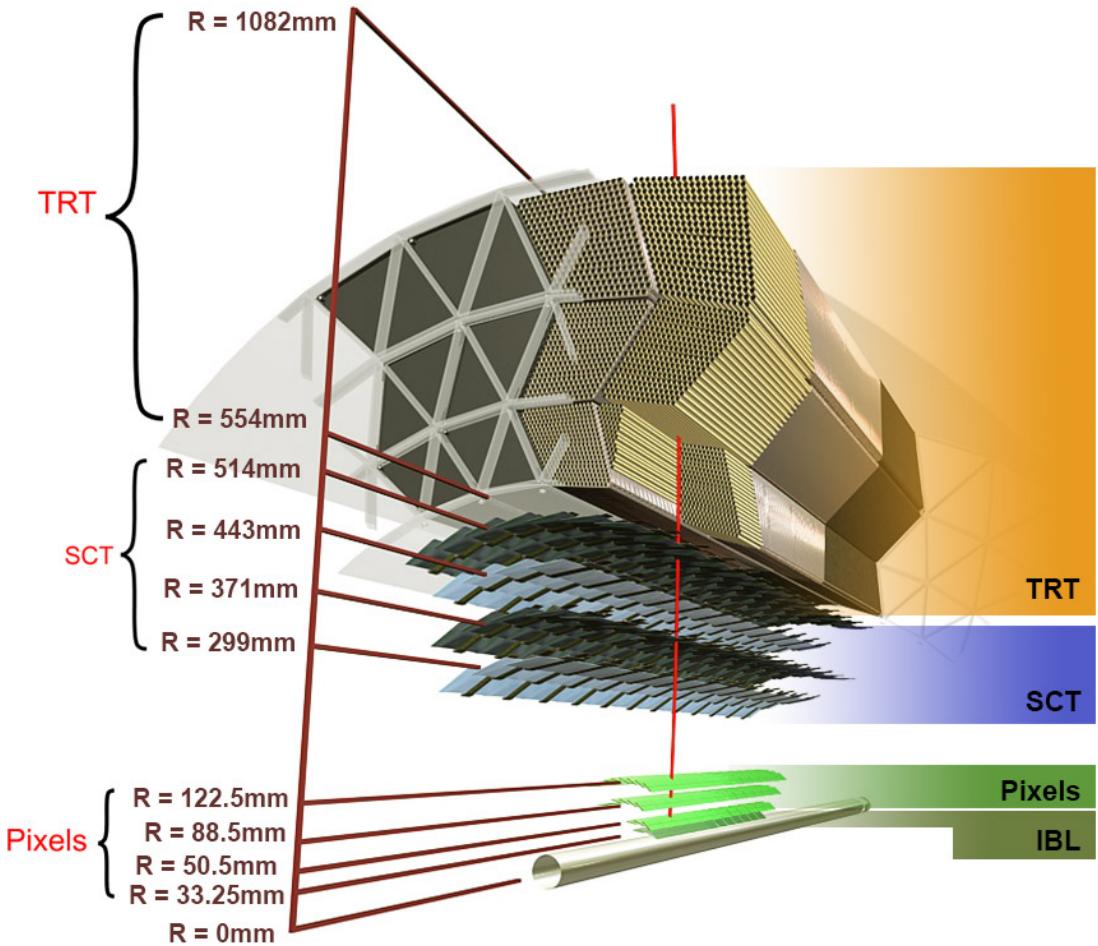
## 491 6.2 Inner Detector

492 The Inner Detector (ID) was designed to identify and reconstruct vertices,  
493 distinguish pions from electrons, and measure the momentum of charged particles.  
494 The ID uses three different technologies for particle reconstruction: the Pixel  
495 Detector, Semiconductor Tracker (SCT), and the Transition Radiation Tracker  
496 (TRT), shown in Figure 6.3 and 6.4. The entire ID is immersed in a 2T solenoidal  
497 magnetic field parallel to the  $+z$ -axis, causing charged particles to bend in the

<sup>498</sup> transverse-plane, allowing particle momentum measurements.



**Figure 6.3:** Layout of ATLAS Inner Detector



**Figure 6.4:** Layout of ATLAS ID Barrel System.

### 499    6.2.1 Pixel Detector

500       The pixel detector consists of four barrel layers between  $r = 32.7$  and  $122.5$   
 501   mm, extending to  $|z| = 400.5$  mm. The remaining detectors are arranged in bar-  
 502   rels and forward and backward rings. The innermost pixel barrel, the Insertable  
 503   b-Layer (IBL), only extends to  $|z| = 332$  mm. The pixel detectors closer to the  
 504   beam line (larger  $\eta$  values) consists of six parallel cylindrical rings of pixel de-  
 505   tectors transverse to the beam line. The entire pixel detector consists of 1744  
 506   identical pixel sensors each with 46080 readout channels, totaling about 80 mil-  
 507   lion individual pixels. Most of the pixel sensors are  $50 \times 400 \mu\text{m}^2$ . Each pixel has

508 a position resolution of  $14\mu\text{m}$  in  $\phi$  and  $115\mu\text{m}$  in the  $z$  direction.

### 509 **6.2.2 Semiconductor Tracker**

510 The SCT is located outside the pixel detector and has the same barrel and  
511 endcap geometry as the pixel detector. SCT sensors are  $80\mu\text{m} \times 12\text{ cm}$  with  
512 a  $80\mu\text{m}$  strip pitch. In the barrel the strips are parallel to the  $z$ -axis and are  
513 segmented in  $\phi$ . In the endcaps, the strips extend radially. Sensors are grouped in  
514 modules containing two layers of strips rotated 40 mrad with respect to each other.  
515 This offset allows for the two-dimensional position of a track to be determined by  
516 identifying the crossing point of the strips that registered a hit. SCT modules  
517 measure tracks with an accuracy of  $17\mu\text{m}$  in  $r - \phi$  and  $580\mu\text{m}$  in  $z(r)$  in the  
518 barrel (end-cap) region.

### 519 **6.2.3 Transition Radiation Tracker**

520 The transition radiation tracker (TRT), enveloping the SCT, is a gaseous  
521 straw-tube tracker mainly used for electron/pion track separation. Each straw  
522 is 4 mm in diameter and filled with a Xe- $\text{CO}_2$ - $\text{O}_2$  gas mixture. An anode wire at  
523 the center of the straw is held at ground potential, while the walls of the straw  
524 are kept at -1.4kV. When a charged particle passing through the TRT ionizes the  
525 gaseous mixture, the resulting ions form an avalanche on the anode wire with a  
526 gain of  $\sim 10^4$ . The signal from the anode wire is then digitized and discriminated.  
527 Signals passing a low threshold cutoff are used to distinguish noise from tracks.  
528 Signals passing a high threshold cutoff are sensitive to transition radiation (TR).  
529 TR photons are emitted when charged particles pass between materials with dif-  
530 ferent dielectric constants. The probability that a charged particle with energy  $E$   
531 and mass  $m$  passing between two materials emits a TR photon in the keV range

532 is proportional to  $\gamma = E/m$ . In the TRT straws these often then convert via the  
533 photoelectric effect, causing a large avalanche triggering the high-threshold. Since  
534 electrons have a smaller mass than pions, electron tracks are more likely to trig-  
535 ger the high threshold. This then provides discrimination between electrons and  
536 charged hadrons.

537 The barrel region of the TRT extends from  $r = 563\text{-}1066$  mm and  $|z| < 712$   
538 mm. Barrel Straws are 144 cm long (divided  $\sim \eta \approx 0$ ) and orientated parallel to  
539 the beam direction. End-cap straws extend radially and are 37 cm long. There  
540 are 53,544 straws in the barrel and 160,000 straws in the end-caps. Radiator mats  
541 of polypropylene/polyethylene fibers in the barrel are aligned perpendicular to the  
542 barrel straws (with holes for the straws to pass through). In the end-cap region,  
543 radiator foils are layered between the radial TRT straws.

544 The arrival time of the signal pulse is sensitive to the distance between the  
545 charged particle track and the anode wire and allows for a hit resolution of  $130\mu\text{m}$ .  
546 The TRT extends to  $|\eta| = 2.0$  and provides about 36 hits per track.

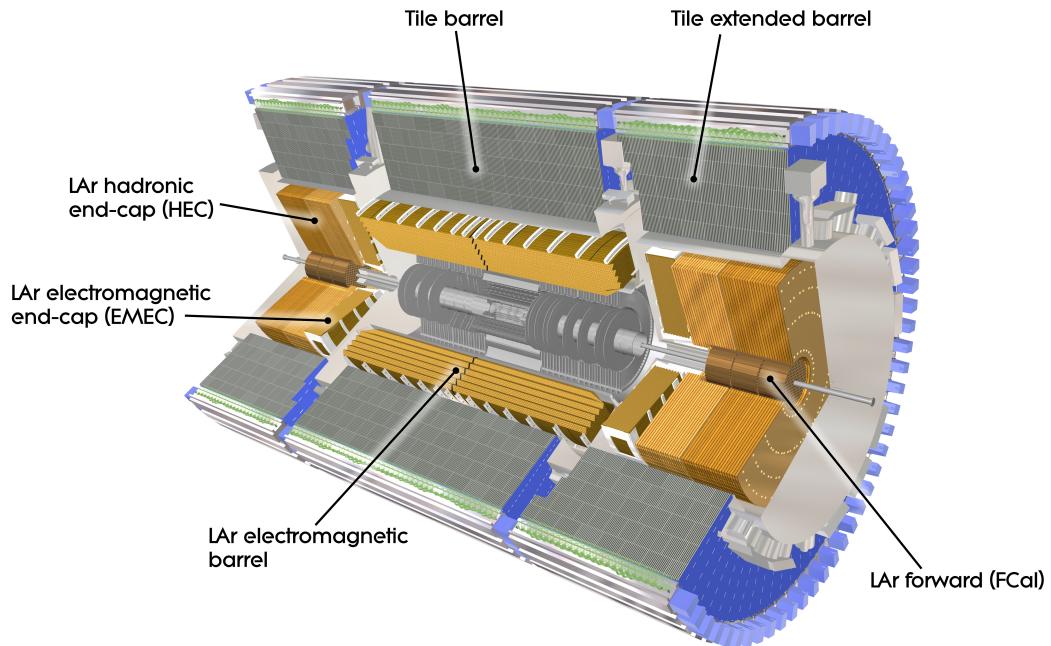
### 547 6.3 Calorimeters

548 The ATLAS electromagnetic and hadronic calorimeters (EMC and HCAL,  
549 respectively) absorb and measure the energy of high energy hadrons, photons,  
550 and electrons with  $|\eta| < 4.9$ . Both systems use sampling calorimeters which  
551 consist of alternating layers of dense absorbing and active layers. In the absorbing  
552 layer particles interact and lose energy, creating showers. These showers are then  
553 detected and measured in the active layer. The amount of charge measured in the  
554 active material scales with the energy of the incident particle, and thus provides a  
555 measurement of the particle's energy. An overview of the layout of the calorimeter  
556 system is shown in Figure 6.5.

557       The EMC measures and contains the energy of electromagnetically interacting  
558       particles. It consists of layered accordion-shaped Lead absorber plates and elec-  
559       trodes immersed in liquid Argon with 170k channels.. Using accordion-shaped  
560       electrode and absorbers ensures  $\phi$  symmetry and coverage. The EMC is com-  
561       posed of a barrel part ( $|\eta| < 1.475$ ), two end-caps ( $1.375 < |\eta| < 3.2$ ), and a  
562       presampler ( $|\eta| < 1.8$ ). The presampler, containing only liquid Argon, corrects  
563       for upstream energy losses of electrons and photons. The EMC barrel is segmented  
564       into three layers. The first layer has finest segmentation with readout cells ex-  
565       tending  $\Delta\eta \times \Delta\phi = 0.025/8 \times 0.1$ . This provides a precise shower measurements  
566       used to separate prompt photons from  $\pi^0 \rightarrow \gamma\gamma$  decays. The second layer has  
567       coarser segmentation and is approximately 16 radiation lengths long. A radiation  
568       length is the average distance an electron travels before losing all but  $1/e$  of its  
569       energy to bremsstrahlung. The last layer is the most coarse and measures the tail  
570       of the electromagnetic shower. A schematic of the ECAL is shown in Figure 6.6.

571       The hadronic calorimeter located outside the EMC and is used to contain  
572       and measure the energy of hadronically interacting particles. It consists of a tile  
573       calorimeter (TileCal), hadronic end-cap calorimeter (HEC), and liquid Argon for-  
574       ward calorimeter (FCAL). TileCal is located behind the LAr EMC and uses steel  
575       absorbers and liquid Argon as the active material. TileCal consists of three barrel  
576       layers in the central and forward regions, extending up to  $|\eta| < 1.7$ . Photons  
577       generated from hadronic interactions are collected via wavelength-shifting fibers  
578       connected to photomultiplier tubes, as shown in Figure 6.7. The HEC lies behind  
579       the EMC endcap wheels. It uses copper absorbers and liquid Argon as the active  
580       material and covers  $1.5 < |\eta| < 3.2$ . Finally, the FCAL covers  $3.1 < |\eta| < 4.9$   
581       and consists of three modules all using liquid Argon as the active material. The  
582       first module uses copper absorber and was designed for electromagnetic measure-

583 ments. The second and third modules consist of tungsten absorber and are used  
584 to measure the kinematics of hadronically interacting particles. A schematic of  
585 the HCAL is shown in Figure 6.7.



**Figure 6.5:** Overview of ATLAS electromagnetic and hadronic calorimeters.

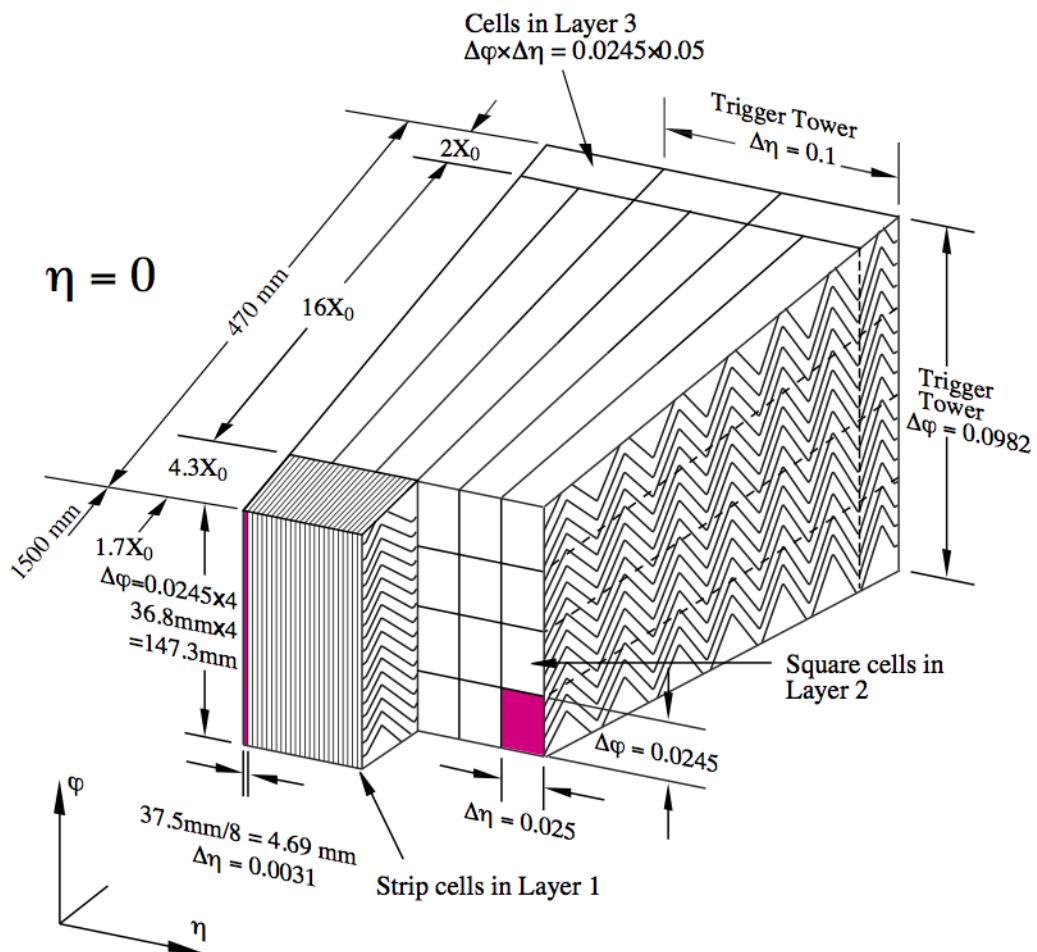


Figure 6.6: Schematic of ECAL.



Figure 6.7: Schematic of HCAL.

586       The energy resolution of the calorimeter subsystems are:

$$587 \quad \frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus \frac{0.3\%}{E} \oplus 0.4\% \text{ Electromagnetic Calorimeter}$$

$$588 \quad \frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus \frac{1.8\%}{E} \oplus 3\% \text{ Hadronic Calorimeter}$$

## 589       **6.4 Muon Spectrometer**

590       The muon spectrometer (MS) is the outermost detector system in ATLAS.

591       Muons with a  $p_T > 4$  GeV are energetic enough to reach the MS. To measure the  
592       momentum of these muons barrel and end-cap toroid magnets are used covering  
593        $|\eta| < 1.4$  and  $1.6 < |\eta| < 2.7$ . For  $1.4 < |\eta| < 1.6$ , a combination of the barrel  
594       and end-cap toroidal magnetic fields bend muon trajectories. The detector in the  
595       barrel region form three concentric rings at  $R = 5, 7.5, 10$  m and are segmented  
596       in  $\phi$  to accommodate the magnets. The end-cap region consists of three circular  
597       planes perpendicular to  $z$  and located at  $|z| = 7.4, 14, 21.5$  m from the interaction  
598       region. An additional detector at  $|z| = 10.8$  m covers the transition region between  
599       the barrel and end-cap.

600       The MS readout consists of four subsystems: Monitored Drift Tubes (MDT),  
601       Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC), and Thin Gap  
602       Chambers (TGC). The first two subsystems are used primarily for measuring  
603       muon track parameters, while the RPC and TGC subsystems are used for muon  
604       triggering. A schematic of this system is shown in Figure 6.8.

605       The MDT subsystem consists of precision tracking chambers for  $|\eta| < 2.7$ ,  
606       except for the inner most end-cap layer ( $2.0 < |\eta| < 2.7$ ), where CSCs are used.  
607       The basic unit of MDT chambers are thin walled Aluminum tubes with a diameter  
608       of 3 cm and length of 0.9-6.2 m. These tubes are filled with a mixture of Ar-CO<sub>2</sub>  
609       gas with a 50μm W-Rn wire running down the center of the tube, which is kept at  
610       3080 V. Since the maximum drift time of these chambers is ∼ 700 ns, they are not

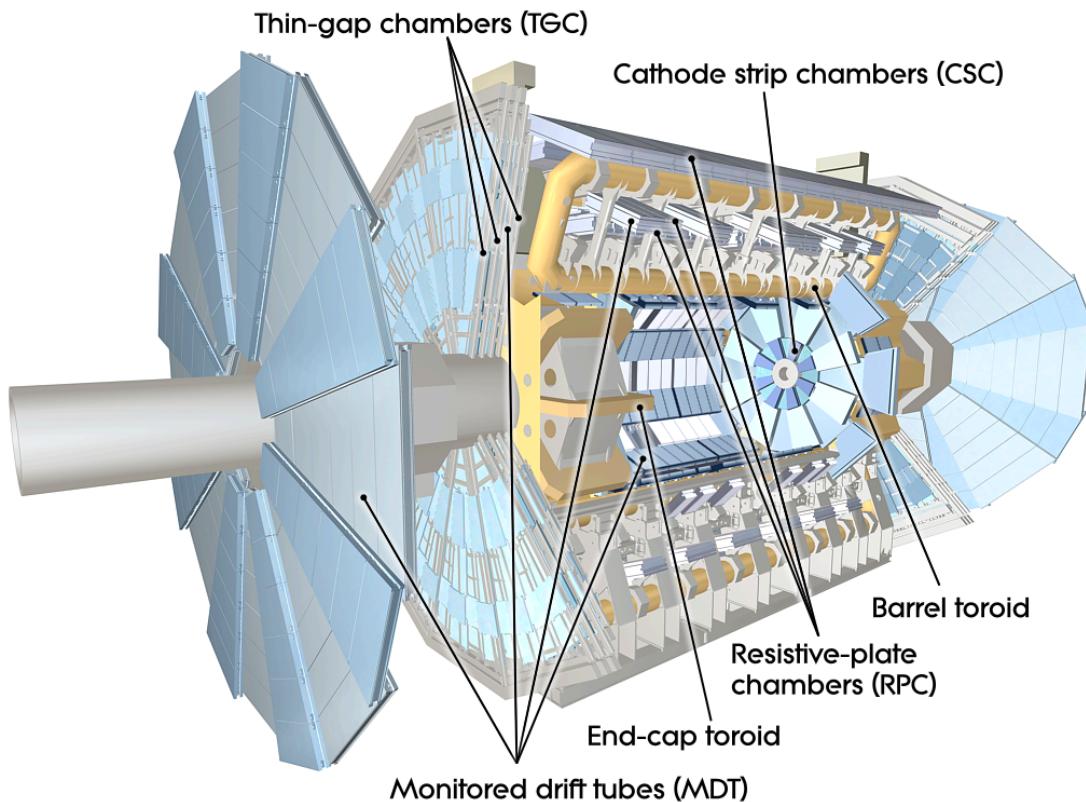
611 used for triggering. MDT chambers consist of 3-4 layers of tubes mounted on a  
612 rectangular support system, as seen in Figure 6.9, orientated along  $\phi$  to measure  
613 the coordinate in the bending plane of the magnetic field with a resolution of 35  
614  $\mu\text{m}$ .

615 The MDT subsystem can only handle hit rates below 150 Hz/cm<sup>2</sup>. For this  
616 reason, CSCs are used in the innermost end-cap layer where hit rates are larger.  
617 CSCs can handle hit rates up to 1000Hz/cm<sup>2</sup>. CSC are multiwire proportional  
618 chambers. These chambers are filled with a Ar-CO<sub>2</sub> gas mixture and evenly spaced  
619 wires kept at 1900 V. These wires are orientated in the radial direction but not  
620 read out. Instead on one side of the cathode are copper strips parallel to the wires,  
621 measuring  $\eta$ , while on the other side of the cathode are strips parallel to the wires  
622 measuring  $\phi$ . The width between strips is approximately 1.5 mm providing a  
623 resolution of 60  $\mu\text{m}$  in the bending-plane and 5 mm in the non-bending plane.

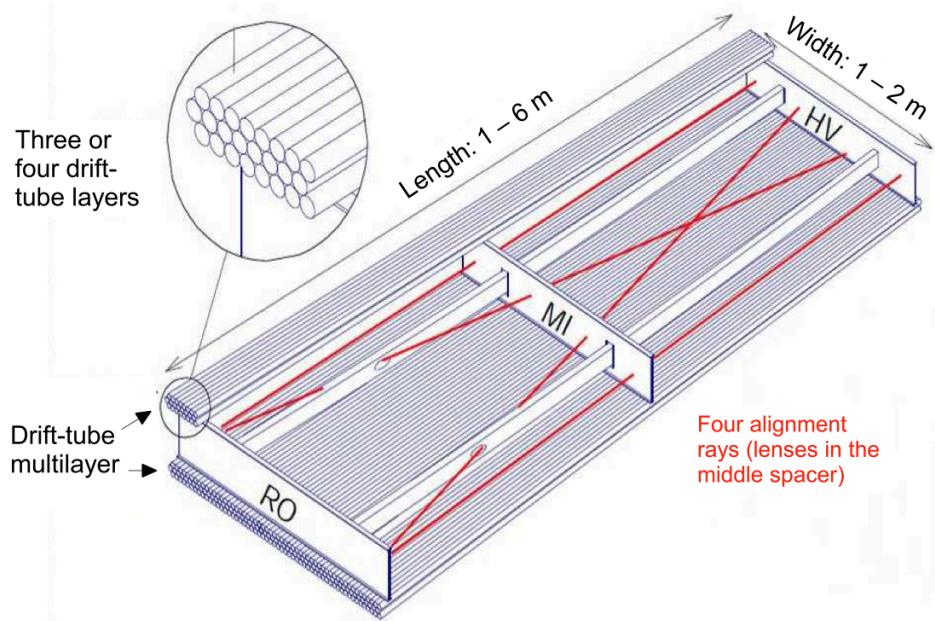
624 Since the CSC and MDT systems do not have prompt timing signals, the RPC  
625 and TGC systems are used for triggering. The RPC system is used in the barrel  
626 region ( $|\eta| < 1.05$ ). RPC consist of two parallel resistive plates separated by a  
627 2 mm insulated spacer with 100 mm spacing kept at 9.8 kV, as shown in Figure  
628 6.10. A gaseous mixture of C<sub>2</sub>H<sub>2</sub>F<sub>4</sub>, C<sub>4</sub>H<sub>10</sub>, and SF<sub>6</sub> fills the space between the  
629 two plates. Metallic strips on the outer faces of the plates are used to read out  
630 signals produced by the gas ionizing. The middle barrel layer consists of two layers  
631 of RPCs on either side of the MDT layer and one layer on the outermost MDT  
632 layer. Each layer contains two orthogonal sets of metallic strips providing  $\eta$  and  
633  $\phi$  measurements. The timing resolution of RPCs is 1.5 ns, and therefore may be  
634 used to identify bunch crossings.

635 Finally, the TGCs are used in the end-cap regions and are primarily used to  
636 provide L1 trigger decisions and  $\phi$  measurements. TGCs are multi-wire propor-

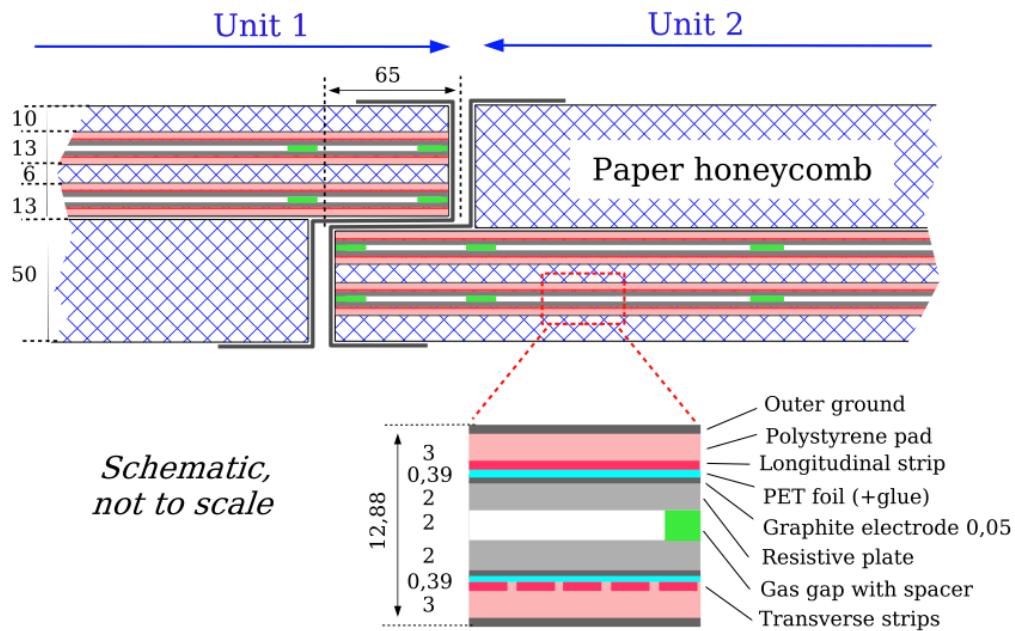
637 tional chambers consisting of arrays of gold-coated tungsten wires placed between  
 638 two cathode planes. These wires are separated by 1.8 mm and cathodes are 1.4 mm  
 639 from the wires. Orthogonal to the wires, on the opposite side of the cathode plane  
 640 are copper strips held at 2900 V. The chambers are filled with a mixture of CO<sub>2</sub>  
 641 and n-pentane gas, the latter acts as a quenching gas to prevent avalanches initi-  
 642 ated by secondary  $\gamma$ -rays from the primary avalanche. Figure 6.11 is a schematic  
 643 of a TGC. The timing resolution of TGCs is less than 25 ns and therefore they  
 644 are used for bunch crossing measurements.



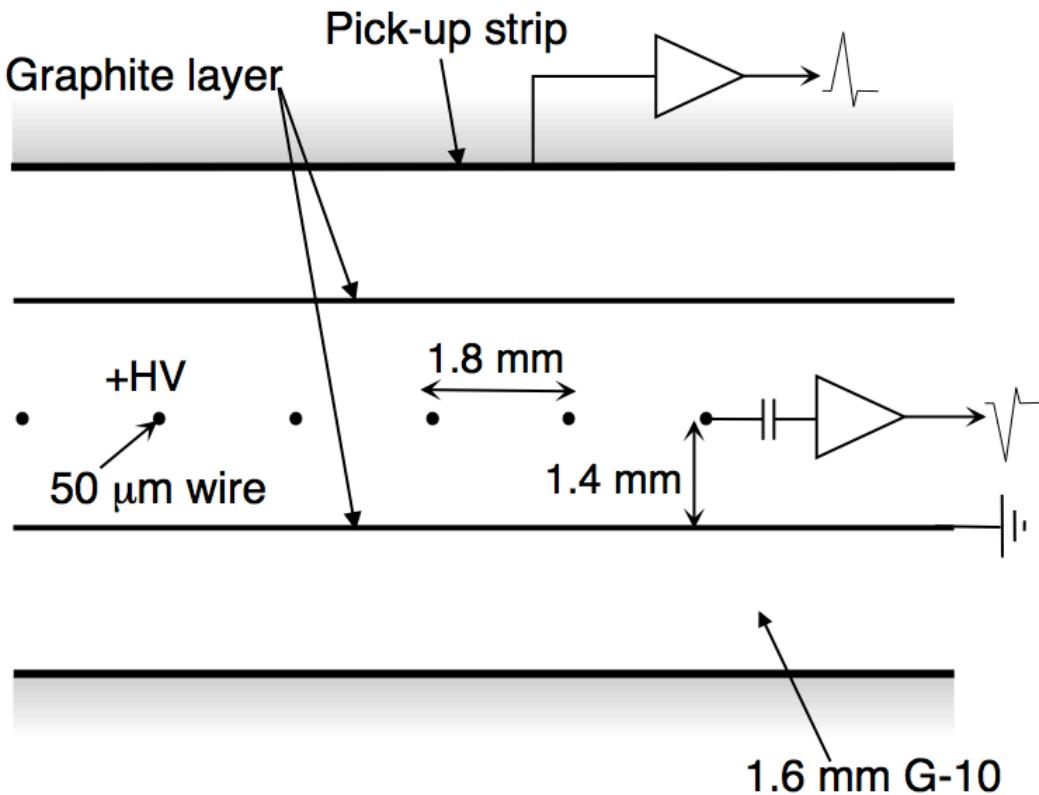
**Figure 6.8:** Schematic of Muon Spectrometer [cite G35]



**Figure 6.9:** Schematic of MDT chamber. [cite G35]



**Figure 6.10:** Schematic of RPC chamber, which is used for triggering in the central region of the detector [cite G35].



**Figure 6.11:** Schematic of TGC chamber, which is used for triggering in the muon end-cap region. [cite G35]

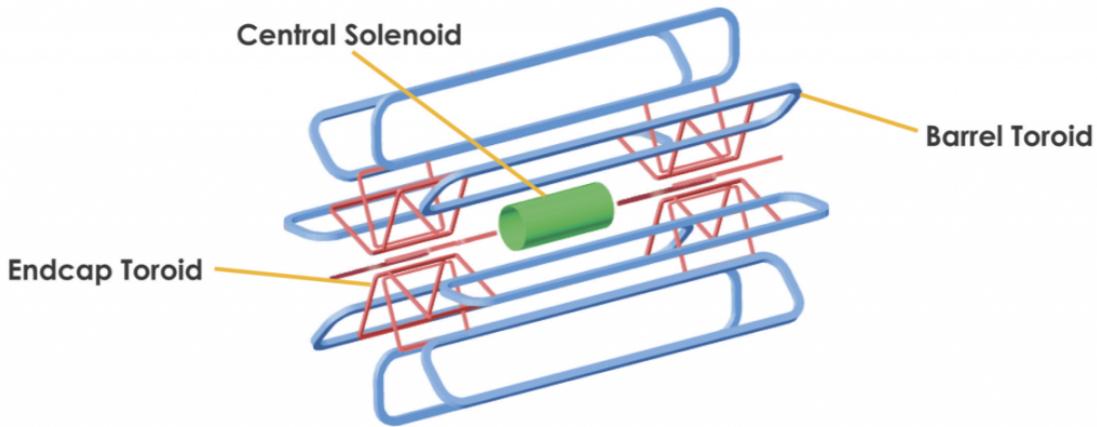
## 645 6.5 Magnet System

646 A particles with charge,  $q$ , and velocity  $v$ , moving in magnetic field,  $B$ , ex-  
 647 periences a force,  $F = qv \times B$ . This force can cause charged particles to have a  
 648 curved trajectory in magnetic fields, which the ID and MS use to determine the  
 649 particles  $p_T$ . The central solenoid provides the magnetic field for the ID and the  
 650 toroidal magnets provide the magnetic field for the MS.

651 The layout of the magnet system is shown in Figure 6.12. The central solenoid  
 652 consists of a single-layer Al-stabilized NbTi conductor coil wound inside an Al

653 support cylinder. The solenoid is 5.8 m long, 50 cm thick and has an inner radius  
654 of 1.23 m. It is cooled to 4.5 K to reach superconducting temperatures and shares  
655 the liquid argon calorimeter vacuum vessel to minimize material in the detector.  
656 A current of 7.730 kA produces a 1.998 T solenoidal magnetic field, pointing in  
657 the  $+z$  direction.

658 The toroidal magnet system consists of a barrel and two end-cap toroidal  
659 magnets used to create a magnetic field outside the calorimeters that is orientated  
660 along  $\phi$ . Each barrel toroid is 25.3 m long with an inner and outer diameter of 9.4  
661 and 20.1 m and weighs 830 tonnes. Endcap toroids are 5 m long with an inner and  
662 outer radius of 1.65 and 10.7 m. Both toroid systems use Al-stabilized Nb/Ti/Cu  
663 conductors. The magnetic field strength in the barrel and endcap regions are 0.5  
664 and 1 T, respectively.



**Figure 6.12:** Layout of ATLAS magnet systems.

## 665 **6.6 Trigger System**

666 Since collisions occur every 25 ns and reading out all detector channels and  
667 storing that information is not currently feasible (would require saving 60 million  
668 megabytes per second), the majority of events are not kept for analysis. ATLAS

669 uses a multi-stage trigger system to select approximately 1,000 of the 1.7 billion  
670 collisions that occur each second (corresponding to a rate of 1 kHz from the 40  
671 MHz proton collision rate). The first stage of the trigger system is the hardware  
672 level (L1) trigger. This trigger reduces the event rate to  $\sim$ 100 kHz by identifying  
673 Regions-of-Interest (ROIs) containing high  $p_T$  leptons, photons, jets, or  $E_T^{miss}$  by  
674 using information from RPCs, TGCs, and calorimeters to make a  $2.5 \mu\text{s}$  decision.  
675 This information is then passed to a high-level trigger (HLT) which further de-  
676 creases event rates to  $\sim$  1 kHz. The HLT uses finer granularity measurements  
677 from the MS and ID to perform simplified offline reconstruction to decide which  
678 events to keep.

## **Part IV**

679

## **Method**

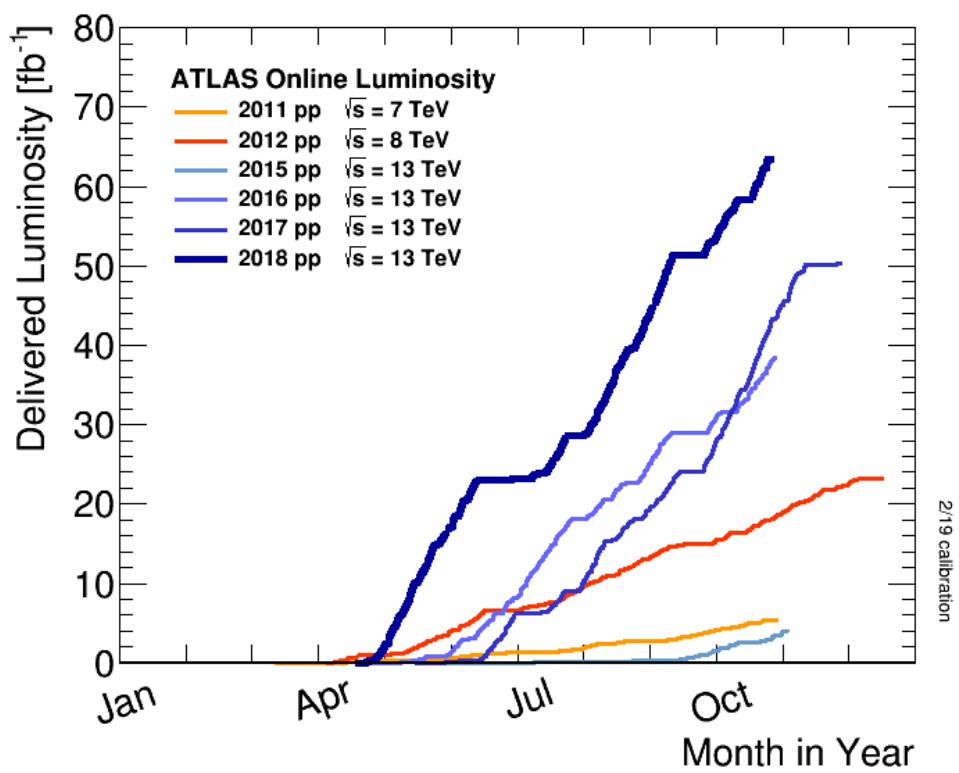
680

681 **Chapter 7**

682 **Dataset and Simulated Samples**

683 **7.1 Dataset**

684 This analysis uses  $pp$  collision data collected from 2015 to 2018 at  $\sqrt{s} = 13$   
685 TeV, corresponding to 139/fb of data as shown in Figure 7.1 and 7.2. From this  
686 dataset, only those events in which the tracker, calorimeters, and muon spectrom-  
687 eter have good data quality are used. For a given event, the solenoid and toroidal  
688 magnets must also be operating at their nominal field strengths. In addition to  
689 this, events must pass further quality checks to reject events where detector sub-  
690 systems may have failed. These selections reject events that containing LAr noise  
691 bursts, saturation in the electromagnetic calorimeter, TileCal errors, and failures  
692 in event recovery due to tracker failures. Events with information missing from  
693 subsystems (usually due to busy detector conditions) are rejected. Events must  
694 also contain a primary vertex with at least two associated tracks, where the pri-  
695 mary vertex is selected as the vertex with the largest  $\sum p_T^2$  over tracks associated  
696 with the vertex and  $p_T > 0.5$  GeV.



**Figure 7.1:** Integrated luminosity for data collected from ATLAS from 2011 - 2018



**Figure 7.2:** Mean number of interactions per crossing for data collected from ATLAS from 2011 - 2018

## 697 7.2 Simulated Samples

698 Samples are simulated in order to model backgrounds, evaluate signal ac-  
699 ceptance, optimize event selection and estimate systematic and statistical uncer-  
700 tainties. The dominant backgrounds for this analysis are  $W/Z + \text{jets}$ , diboson  
701 ( $WZ/WW$ ),  $t\bar{t}$ , single top and multijet production.

702  $W/Z+\text{jet}$  events are simulated using Sherpa 2.2.1 at NLO [cite [29]] and merged  
703 with the Sherpa parton shower using the ME+PS@NLO prescription [13]. These  
704 events are then normalized to NNLO cross sections. The  $t\bar{t}$  and single-top back-  
705 grounds are generated with Powheg-Box with NNPDF3.0NLO PDF sets in the  
706 matrix element calculation [cite[35]]. For all processes, the parton shower, frag-  
707 mentation, and underlying event are simulated using Pythia 8.320 with the A14  
708 tune set[cite[ATL-PHYS-PUB-2014-02]]. Diboson processes are generated using  
709 Sherpa 2.2.1.

710 Signal samples are simulated using MadGraph 5-2.2.2 [cite 42] and Pythia  
711 8.186 with NNPDF230LO. RS Graviton samples are generated with  $k/M_{PL}=1$ .  
712 HVT Model A (B) samples are simulated with  $g_V = 1(3)$ , as the difference in the  
713 width of the samples is smaller than detector resolution. Model C is generated by  
714 setting  $g_H = 1$  and  $g_f = 0$  to model VBF production of HVT bosons. Signals are  
715 generated for masses between 300 GeV and 5 TeV.

716 **Chapter 8**

717 **Objects**

718 **8.1 Electrons**

719 Electrons are reconstructed from electromagnetic showers in the LAr EM  
720 calorimeter. During reconstruction cells of  $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$  are grouped  
721 into  $3 \times 5$  clusters. These clusters are then scanned for local maxima that seed  
722 electron clusters. These clusters must then be matched to ID track from the PV.  
723 This requirement minimizes non-prompt electron and fake electron backgrounds.  
724 Electrons must pass identification and isolation requirements. Electron identifica-  
725 tion (loose, medium, tight) classification is based on a multivariate discriminant  
726 that identifies electrons using a likelihood based method. For this analysis, events  
727 are required to have one tight electron and no additional loose electrons. Elec-  
728 trons are also required to be isolated. The electrons are considered isolated if the  
729 quotient of the sum of the transverse momentum (of calorimeter energy deposits)  
730 in a cone around the electron of size  $\Delta R = 0.2$  and the transverse momentum  
731 of the electron to be less than  $0.015 * p_T$  or 3.5 GeV, whichever is smaller. This  
732 requirement rejects non-prompt photons and other fake leptons. Electrons in this  
733 analysis are also required to have  $p_T > 30$  GeV and  $|\eta| < 2.47$ . Electrons are also

734 required to have  $p_T > 30$  GeV.

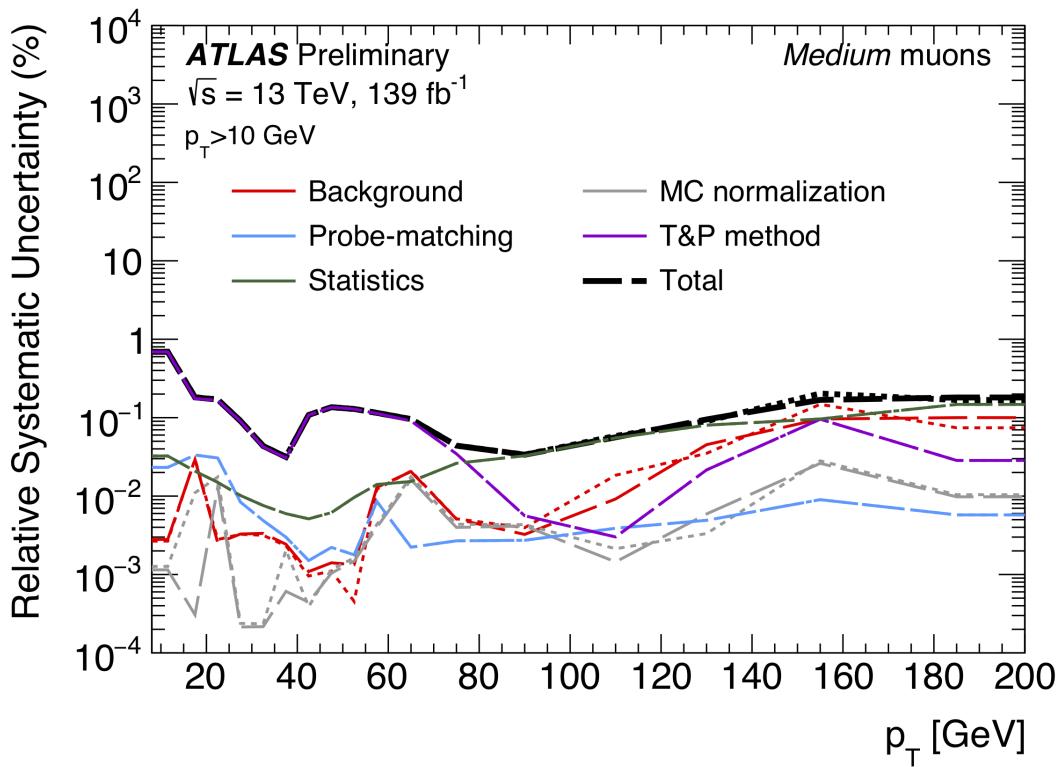
735 Electrons are calibrated to determine data-driven scale factors using  $J/\Psi \rightarrow$   
736  $ee$ ,  $Z \rightarrow ee$ ,  $Z \rightarrow \ell\ell\gamma$  processes. These corrections account for the non-uniform  
737 response of the detector which introduces modeling and reconstruction uncertain-  
738 ties.

## 739 8.2 Muons

740 As muons traverse the entire detector, they are reconstructed from ID and  
741 MS tracks. For this analysis the muon identification and isolation working points  
742 are chosen to minimize the contributions from non-prompt muons. Towards this  
743 end, each selected event must contain exactly one muon that passes the medium  
744 identification working point, and no additional muons (that pass the loose working  
745 point). For the medium working point, two types of reconstructed muons are  
746 used: combined and extrapolated muons (CB and ME, respectively). For CB  
747 muons, ID and MS tracks are reconstructed independently and a combined track  
748 fit is performed by adding or removing MS tracks to improve the fit quality.  
749 ME muons are reconstructed from only MS tracks with hits in at least two layers,  
750 which ensures the track originates from the PV. ME muons extend the acceptance  
751 for muon reconstruction outside the ID from  $2.5 < |\eta| < 2.7$ . The medium  
752 identification working point uses CB and ME tracks. CB tracks must have at  
753 least 3 hits in two MDT layers. ME tracks are required to have at least three  
754 MDT/CSC hits. To further minimize contributions from fake muons, the selected  
755 muons are required to be isolated from other tracks, as muons from  $W, Z$  decays are  
756 often isolated from other particles. To insure the selected muons are isolated, the  
757 scalar sum of the transverse momentum of tracks in a cone of  $\Delta R = 0.3$  compared  
758 to the transverse momentum of the muon must be less than 0.06. Muons are also

759 required to have  $p_T > 30$  GeV.

760 Muons are calibrated using well-studied resonances  $J/\Psi \rightarrow \mu\mu$  (low- $p_T$ ),  $Z \rightarrow$   
761  $\mu\mu$  (high- $p_T$ ). Figure 8.1 shows the combined muon  $p_T$  uncertainty from this  
762 calibration. The total systematic uncertainty is less than 1% for all  $p_T$  ranges  
763 considered in this analysis.



**Figure 8.1:** This figure shows the breakdown of the muon reconstruction efficiency scale factor measured in  $Z \rightarrow \mu\mu$  as a function of  $p_T$  [5].

### 764 8.3 Jets

765 Three types of jets are used in this analysis: variable radius, small-R and  
766 large-R jets. Variable radius jets are used to reconstruct  $Z$  bosons decaying to  
767 two  $b$ -jets in the jet catchment area of large-R jet in the Merged regime. Small-R

768 jets are used to reconstruct the hadronically decaying  $W/Z$  candidates in the  
769 resolved analysis and the forward jets from resonances produced through vector  
770 boson fusion. Large-R jets are used to reconstruct the hadronically decaying boson  
771 in the merged regime.

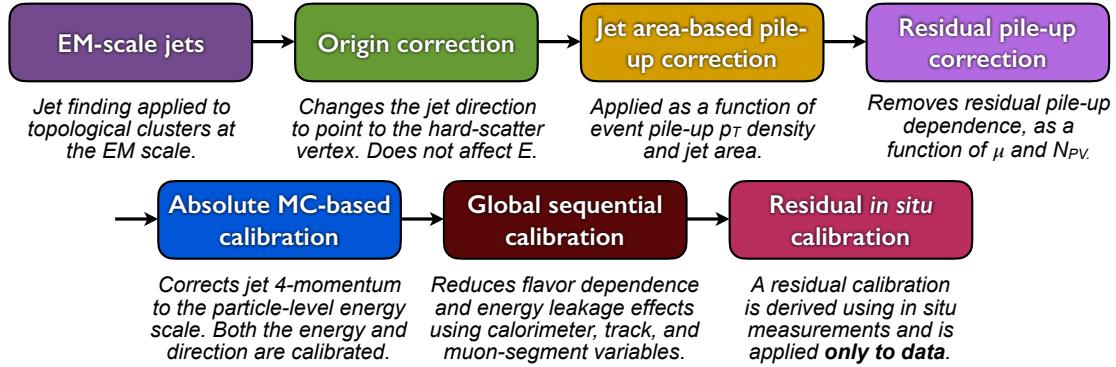
772 For all of these jet collections, the jet energy is calibrated sequentially as shown  
773 in Figure 8.2. After the jet direction is corrected to point to the PV, the energy  
774 of the jet is corrected. First, the jet energy is corrected to account for pileup  
775 contributions based on the  $p_T$  and area of the jet (these corrections are extracted  
776 from a  $pp \rightarrow jj$  sample). Following this, another pileup correction is applied that  
777 scales with  $\mu$  and  $N_{PV}$ .

778 MC-based corrections are then applied that are meant to transform the jet  
779 energy and  $\eta$  back to truth level. Therefore, these corrections account for the  
780 non-compensating nature of the ATLAS calorimeters and inhomogeneity of the  
781 detector. Following this, the Global Sequential Calibration is applied that re-  
782 duces flavor dependence of jet calibrations and accounts for energy leakage of jets  
783 outside the calorimeters. Finally, in-situ corrections are applied that account for  
784 differences in jet responses between data and simulation ( $\gamma/Z + \text{jet}$  and fake lep-  
785 ton samples are used). These differences can be due to mismodelling of the hard  
786 scatter event, pile-up, jet formation, etc.

787 To further reject fake jets, jets must pass quality requirements based on the  
788 following variables ([cite P42]):

- 789 -  $f_Q^{LAr}$ : fraction of energy of jet's LAr cells with poor signal shape
- 790 -  $f_Q^{HEC}$ : fraction of energy of jet's HEC cells with poor signal shape
- 791 -  $E_{neg}$ : sum of cells with negative energy
- 792 -  $f_{EM}$ : fraction of jet's energy deposited in EM calorimeter

- 793     -  $f_{HEC}$ : fraction of jet's energy deposited in HEC calorimeter
- 794     -  $f_{max}$ : maximum energy fraction in any single calorimeter layer
- 795     -  $f_{ch}$ : ratio of the scalar sum of the  $p_T$  of a jet's charged tracks to the jet's  $p_T$
- 796     Jets selected for the resolved analysis must pass one of the following criteria:
- 797     -  $f_{HEC} > 0.5$  and  $|f_Q^{HEC}| > 0.5$  and  $\langle Q \rangle > 0.8$
- 798     -  $|E_{neg}| > 60$  GeV
- 799     -  $f_{EM} > 0.95$  and  $f_Q^{LAr} > 0.8$  and  $\langle Q \rangle > 0.8$  and  $|\eta| < 2.8$
- 800     -  $f_{max} > 0.99$  and  $|\eta| < 2$
- 801     -  $f_{EM} < 0.05$  and  $f_{ch} < 0.05$  and  $|\eta| < 2$
- 802     -  $f_{EM} < 0.05$  and  $|\eta| > 2$



**Figure 8.2:** [8] This diagram shows the calibration stages for EM jets.

803    **8.3.1 Small-R jets**

804    Small-R jets are used to reconstruct the hadronically decaying  $W/Z$  candi-  
805    date when the two resulting jets are well-separated in  $\eta\text{-}\phi$  space. Small-R jets  
806    are also used to identify forward jets from resonances produced through vector  
807    boson fusion. Small-R jets are constructed from topologically connected clusters  
808    of calorimeter cells (topoclusters), seeded from calorimeter cells with energy de-  
809    posits significantly above the noise threshold. These cells are then used as inputs  
810    to the  $\text{anti}-k_t$  algorithm [16] with a radius parameter,  $R = 0.4$ .

811    Jets used in this analysis must have  $p_T > 30$  GeV and  $|\eta| < 2.5$ . To further  
812    reduce fake jets the jet-vertex-tagger (JVT) is used to reject pile-up jets [7]. The  
813    JVT uses two track-based variables, corrJVF and  $R_{p_T}$  to calculate the likelihood  
814    that the jet originated from the PV. The corrJVF compares the scalar sum of the  
815     $p_T$  of tracks associated with the jet and PV to the scalar sum of the  $p_T$  of tracks  
816    associated with the jet. This variable also includes a correction that reduces the  
817    dependency of corrJVF with the number of reconstructed vertices in the event.  
818    The other discriminant,  $R_{p_T}$ , is given by the ratio of the scalar sum of the  $p_T$  of  
819    tracks associated with the jet and PV to the  $p_T$  of the jet. Both of these variables  
820    peak around zero for pileup jets, as these jets are unlikely to have tracks associated  
821    with the PV. JVT cuts are applied to all jets with  $p_T > 120$  GeV. Central jets  
822    ( $|\eta| < 2.4$ ) are required to have a  $\text{JVT} > 0.59$  and forward jets ( $2.4 < |\eta| < 2.5$ )  
823    are required to have  $\text{JVT} > 0.11$ .

824    **8.3.2 Large-R jets**

825    Large-R ( $R = 1.0$ ) jets are used to reconstruct the hadronically decaying  $W/Z$   
826    candidate when the resulting jets are not well-separated in  $\eta\text{-}\phi$  space, and overlap  
827    forming one large-R jet. Track-Calorimeter Clusters (TCCs) are used to reconstruct these

jets [cite ANA 50]. These jets are constructed using a pseudo particle flow method using ID tracks matched to calorimeter clusters. To remove contamination in the jet from pileup and the underlying event, jets are trimmed using a re-clustering algorithm. This algorithm removes subjets with  $p_T^{subjett} < 0.1 p_T^{jet}$ .

The angular resolution of the calorimeter degrades sharply with jet  $p_T$ , but the jet energy resolution improves. The tracker has excellent angular resolution which improves with  $p_T$ . Therefore, by matching tracks to jets, TCCs have more precise energy and angular resolution than jets constructed from calorimeter information only. These jets are required to have  $p_T > 200$  GeV,  $|\eta| < 2.0$  and  $m_J > 50$  GeV.

TCC jets are trimmed as detailed in [cite ANA 45], which suppresses pileup and soft radiation in the jet, the jet mass is calculated as the four-vector sum of the jet's constituents (assuming massless constituents). The jet mass peaks around the  $W/Z$  boson mass for  $W/Z \rightarrow qq$  jets, and more broadly for quark and gluon induced jets.

These jets are then tagged as  $W$  jet if it passes optimized jet mass and substructure ( $D_2$ ) cuts for  $W$  bosons, and a  $Z$  jet if it passes the cuts for the  $Z$  boson. The jet substructure variable  $D_2$  is given by the ratio of energy correlation functions. These functions are derived from the energies and pair-wise angles of a jet's constituents [cite ANA 46, 47]:

$$D_2^{\beta=1} = E_{CF3} \left( \frac{E_{CF1}}{E_{CF2}} \right)^3 \quad (8.1)$$

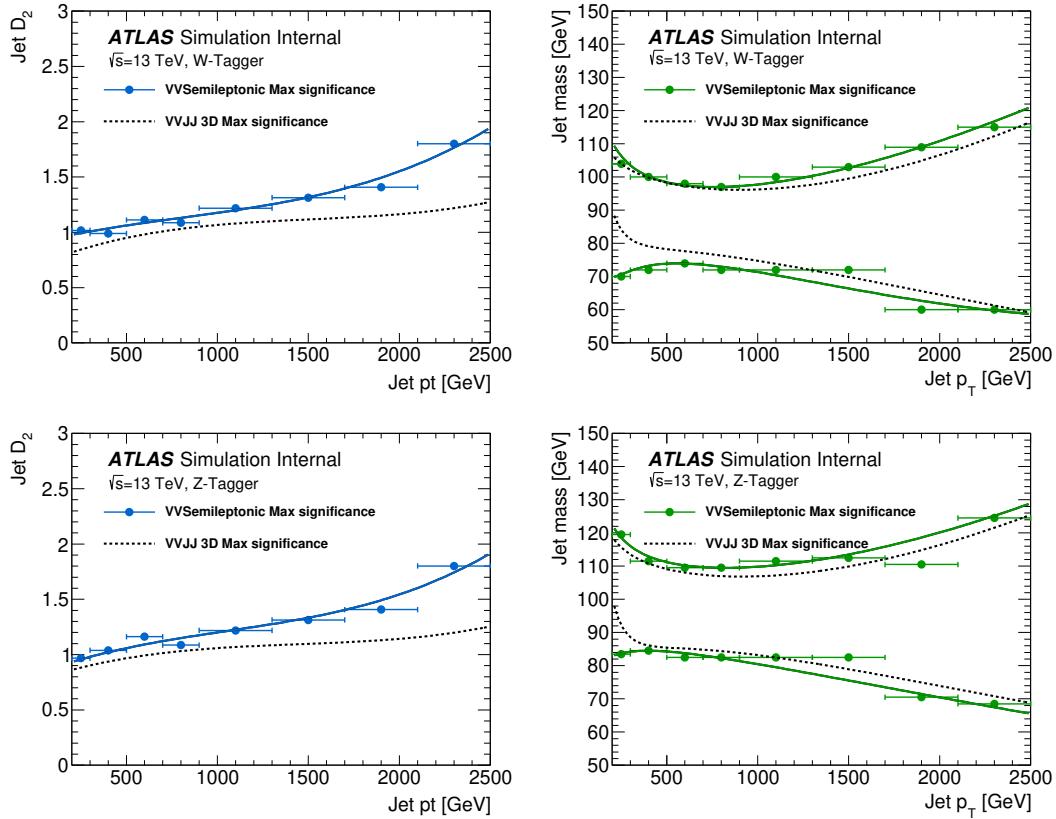
where the energy correlation functions are defined as:

$$E_{CF1} = \sum_i p_{T,i} \quad (8.2)$$

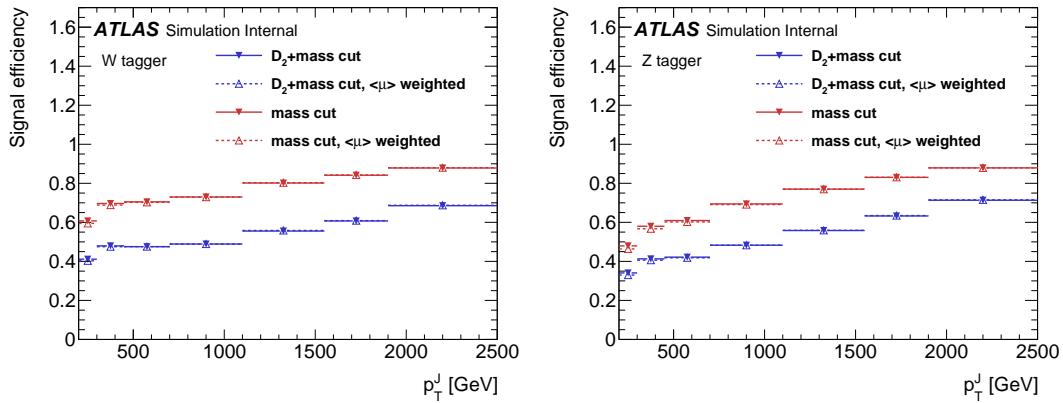
$$E_{CF2} = \sum_{ij} p_{T,i} p_{T,j} \Delta R_{ij} \quad (8.3)$$

$$E_{CF3} = \sum_{ijk} p_{T,i} p_{T,j} p_{T,k} \Delta R_{ij} \Delta R_{jk} \Delta R_{ki} \quad (8.4)$$

849 A two-dimensional optimization of the jet mass and  $D_2$  thresholds was per-  
 850 formed to provide maximum sensitivity for this analysis. This optimization was  
 851 done by maximizing the signal sensitivity (using HVT  $W'$  and  $G_{KK}$  samples)  
 852 against the single quark and gluon jet backgrounds in jet  $p_T$  bins. Figure 8.3  
 853 shows the optimized thresholds on  $D_2$  and jet mass as a function of jet  $p_T$ . Figure  
 854 8.4 shows the efficiency of the optimized  $W/Z$  taggers as a function of jet  $p_T$ .



**Figure 8.3:** The upper cut on  $D_2$  (a) and jet mass window cut i.e. the upper and lower boundary of the mass (b) of the  $W$ -tagger as a function of jet  $p_T$ . Corresponding values for  $Z$ -tagger are shown in (c) and (d). The optimal cut values for maximum significance are shown as solid markers and the fitted function as solid lines. Working points from  $VV \rightarrow JJ$ [ATLAS-HDBS-2018-31-002] is also shown as dashed lines as a reference. Natasha reword?



**Figure 8.4:** Natasha write caption

### 855 8.3.3 Variable Radius jets

856 To more accept more boosted  $Z$  bosons decaying to  $b\bar{b}$  that would normally be  
 857 rejected due to topological cuts discussed 9.4 variable radius (VR) track jets are  
 858 used to identify  $b$ -jets (discussed in 8.3.4) within the catchment area of large- $R$   
 859 jets [15]. VR jets are constructed from ID tracks using the anti- $k_t$  algorithm with  
 860 a radius parameter that depends on the  $p_T$  of the track, shown in Equation 8.5.

$$R_{eff}(p_{T,i}) = \frac{\rho}{p_{T,i}} \quad (8.5)$$

861 For this search  $\rho = 30$  GeV and an lower and upper limit on cone size are set  
 862 to 0.02 and 0.4, respectively, to prevent unphysical asymptotic behavior of  $\rho$ .  
 863 Collinear VR jets are possible, so track jets that are not separated by the the  
 864 smaller jet's cone size are not used. Additionally, VR jets are required to have  
 865  $p_T > 10$  GeV and  $|\eta| < 2.5$ .

866 **8.3.4 Jet Flavor Tagging**

867 To further classify events, the small-R and VR jets originating from a b-quark  
868 are classified using a multivariate  $b$ -tagging algorithm (BDT), MV2c10 [cite G 210  
869 199]. This algorithm uses the impact parameters of the jet's ID tracks, secondary  
870 vertices (if they exist), and reconstructed flight paths of  $b$  and  $c$  hadrons in the  
871 jet to determine if the jet was induced by a  $b$ -quark. For this analysis the 85%  
872 efficient working point of this algorithm is used giving  $c$ ,  $\tau$ , and light-flavor jet  
873 rejection of 3, 8, and 34 respectively in simulated  $t\bar{t}$  samples.

874 **8.4 MET/Neutrinos**

875 As neutrinos are uncharged and colorless they do not leave tracks or jets in  
876 the detector. For this reason, neutrinos are reconstructed as the missing energy  
877 in the event,  $E_T^{miss}$ . Mathematically,  $E_T^{miss}$  is the negative vector sum of  $p_T$  all  
878 the physics objects and an extra "soft" term. The "soft" term accounts for energy  
879 deposits not associated with any of the objects in the event. For this analysis  
880 the soft term is given by the sum  $p_T$  of all ID tracks not associated with objects  
881 in the event. The selected tracks must be matched to the primary vertex, which  
882 decreases pile-up contamination [cite G 217 218].

883 **8.5 Overlap Removal**

884 Reconstructed jets and leptons in this analysis can arise from the same energy  
885 deposits. For instance, a cluster of energy from an electron can also be a valid  
886 calorimeter seed for a jet. To mitigate this confusion of multiple objects originating  
887 from a single jet or lepton overlapping objects are removed via a procedure referred  
888 to as overlap removal. In this procedure the separation of the two objects,  $\Delta(R)$ ,

889 determines which object is removed from the event.

890 The overlap selections used in this analysis are:

- 891 - when an electron shares a track with another electron with the lower  $p_T$   
892       electron is rejected, as it is more likely to be a fake electron
- 893 - when a muon and electron share a track the muon is rejected if it is a  
894       calo-muon, otherwise the electron is rejected
- 895 - when  $\Delta R < 0.2$  for an electron and jet, the jet is rejected to maximize signal  
896       acceptance
- 897 - when  $\Delta R > 0.2$  for an electron and jet, the electron is rejected as likely  
898       originated from decays within the jet
- 899 - when  $\Delta R < \min(0.4, 0.04 + 10\text{GeV}/p_T^\mu)$  the muon is rejected, again maximiz-  
900       ing signal acceptance, otherwise the jet is rejected
- 901 - when  $\Delta R < 1.0$  for the a large-R jet and electron, the jet is rejected

## 902 8.6 Reconstructed Resonance Mass ( $m_{WV}$ )

903 The  $WV$  system mass,  $m_{WV}$  is reconstructed from the lepton, neutrino, and  
904 hadronically-decaying boson candidate. The momentum of the neutrino along the  
905  $z$ -direction is obtained by constraining the  $W$  boson mass of the lepton neutrino  
906 system to be  $80.3 \text{ GeV}/c^2$ . For complex solutions to this constraint,  $p_Z$  is taken  
907 as the real component of the solution. For real solutions, the one with the smaller  
908 absolute value is used. For the resolved analysis,  $m_{WV}$  is reconstructed by con-  
909 straining the  $W(Z)$  dijet system in the SRs, which improves the mass resolution:

$$p_{T,jj}^{corr} = p_{T,jj} \times \frac{m_{W/Z}}{m_{jj}} \quad (8.6)$$

$$m_{jj}^{corr} = m_{W/Z} \quad (8.7)$$

910 where  $m_{jj}$  and  $m_{W/Z}$  are the reconstructed invariant mass of the hadronically-  
911 decaying W/Z boson and the PDG values of the  $W/Z$  boson masses, respectively.  
912 The reconstructed resonance mass is the final discriminating variable in this anal-  
913 ysis. The distribution of this variable in the CR and SRs are used in the final  
914 likelihood fit to search for evidence of an excess of events due to BSM resonances.  
915 The distribution of  $m_{WV}$  are shown in Figures 13.1-13.9.

916 **Chapter 9**

917 **Event Selection and**

918 **Categorization**

919 To search for these new resonances, the simulated background and signal sam-  
920 ples are analyzed to determine a series of optimized cuts are used create signal  
921 regions (SR) to identify the leptonic and hadronic decay products of the reso-  
922 nance. In these regions, the resonance mass is calculated as the combined system  
923 mass of the leptonic and hadronic systems as described in 8.6. The expected res-  
924 onance mass distribution from the backgrounds and signal samples are compared  
925 to data to search for the existence of these BSM signals (also known as a "bump  
926 hunt"). Control regions enriched in the dominant backgrounds,  $t\bar{t}$  and  $W+\text{jets}$   
927 (TCR and WCR, respectively) are constructed to be orthogonal to SRs and used  
928 to determine the normalization of the  $t\bar{t}$  and  $W+\text{jets}$  backgrounds in SRs.

929 Events are classified as produced via non-VBF or VBF modes using a Re-  
930 curisve Neural Network described in 9.3. VBF  $W'$  and  $Z'$  and ggF  $W'$  and  $Z'$   
931 resonances studied have unique SR and CR selections to maximize analysis sen-  
932 sitivity. RS Graviton signals are probed using the same selections as the ggF  $Z'$   
933 signal. Additionally, more massive resonances are more likely to have boosted

934  $W/Z$  bosons. As the boost of the hadronically decaying boson increases the separation  
935 of its hadronic decay products decreases. When the hadronically decaying  
936 boson has sufficient boost, the two quarks will overlap and not be identified sep-  
937 arately. For this reason, a set of "resolved" selections are used when the hadronic  
938 decay products are reconstructed separately, and "merged" selections when the  
939 decay products overlap and identified as a single object in the event. A  $W/Z$   
940 tagger identifies merged jets as originating from a  $W/Z$  bosons based on jet sub-  
941 structure and mass cuts. However, the more boosted the jet is the less likely it is  
942 to pass the jet substructure cut, due to track merging. Consequently, the merged  
943 selection uses a high purity region (HP), which requires that the jet pass both  
944 cuts, and low purity (LP) region where the jet can fail the jet substructure cut.  
945 These selections are summarized in 9.4.

946 The aforementioned SR definitions veto events with  $b$ -jets to minimize  $t\bar{t}$  con-  
947 tamination. However,  $b$ -jets are anticipated from  $W'$  resonances from the hadron-  
948 ically decaying  $Z$  boson. To increase the signal acceptance of these resonances,  
949 a  $Z \rightarrow bb$  tagger is used to construct additional SR and CRs called the "tagged"  
950 regions (and "untagged" if the event fails the  $Z \rightarrow bb$  tagger).

## 951 9.1 Pre-selection

952 Before applying topological cuts, preselection cuts are applied which include  
953 trigger and event requirements to reduce background contamination and the dataset  
954 size. Events must contain exactly one tight lepton (no additional loose leptons),  
955 the  $p_T^{\ell\nu} > 75$  GeV, and there must be at least two small-R jets or one large-R jet,  
956 so the event is able to pass the resolved or merged selections.

957    **9.2 Trigger**

958    The data were collected using the lowest unprescaled single-lepton or  $E_T^{miss}$   
959    triggers, as summarized in Table 9.1. Since the muon term is not considered in the  
960    trigger  $E_T^{miss}$  calculation, the  $E_T^{miss}$  trigger is fully efficient to events with high- $p_T$   
961    muons. For this reason, the  $E_T^{miss}$  trigger is used for events where  $p_T^\mu > 150 \text{ GeV}$ , to  
962    compensate for the poor efficiency of the single muon trigger above  $p_T^\mu > 150 \text{ GeV}$ .

≈

**Table 9.1:** The list of triggers used in the analysis.

Data-taking period	$e\nu qq$ channel	$\mu\nu qq$ ( $p_T(\mu\nu) < 150$ GeV) channel	$\mu\nu qq$ ( $p_T(\mu\nu) > 150$ GeV) channel
2015	HLT_e24_lhmedium_L1EM20 OR HLT_e60_lhmedium OR HLT_e120_lhloose	HLT_mu20_iloose_L1MU15 OR HLT_mu50	HLT_xe70
2016a (run < 302919) $(L < 1.0 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1})$	HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e140_lhloose_nod0 HLT_e300_etcut	HLT_mu26_ivarmedium OR HLT_mu50	HLT_xe90_mht_L1XE50
2016b (run $\geq 302919$ ) $(L < 1.7 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1})$	same as above	same as above	HLT_xe110_mht_L1XE50
2017	same as above	same as above	HLT_xe110_pufit_L1XE55
2018	same as above	same as above	HLT_xe110_pufit_xe70_L1XE50

963 **9.3 non-VBF/VBF RNN**

964 To classify events as originating from non-VBF or VBF production a recursive  
965 neural network (RNN [22]) is used. This approach is more powerful than a cut-  
966 based classification as it improves signal efficiency and analysis sensitivity by  
967 exploiting correlations between variables that the RNN learns. In particular, a  
968 RNN architecture is ideal as it can handle variable numbers of jets in the events.

969 The RNN uses the four-momentum of candidate VBF jets to classify events  
970 as VBF or non-VBF topologies. Sometimes jets are incorrectly reconstructed,  
971 so the number of jets in the event is expected to vary across the input samples.  
972 VBF candidate jets are identified by removing jets from the event that are likely  
973 from  $W/Z \rightarrow qq$ . For the resolved regime this means removing the two leading  
974 small-R jets from the VBF candidate jet list. For the merged regime this means  
975 removing small-R jets separated by less than 1.0 in  $dR$  from the large-R jet. VBF  
976 candidate jets are also required to be within  $|\eta| < 4.5$ . From the list of remaining  
977 VBF candidate jets, the two highest- $p_T$  jets are chosen.

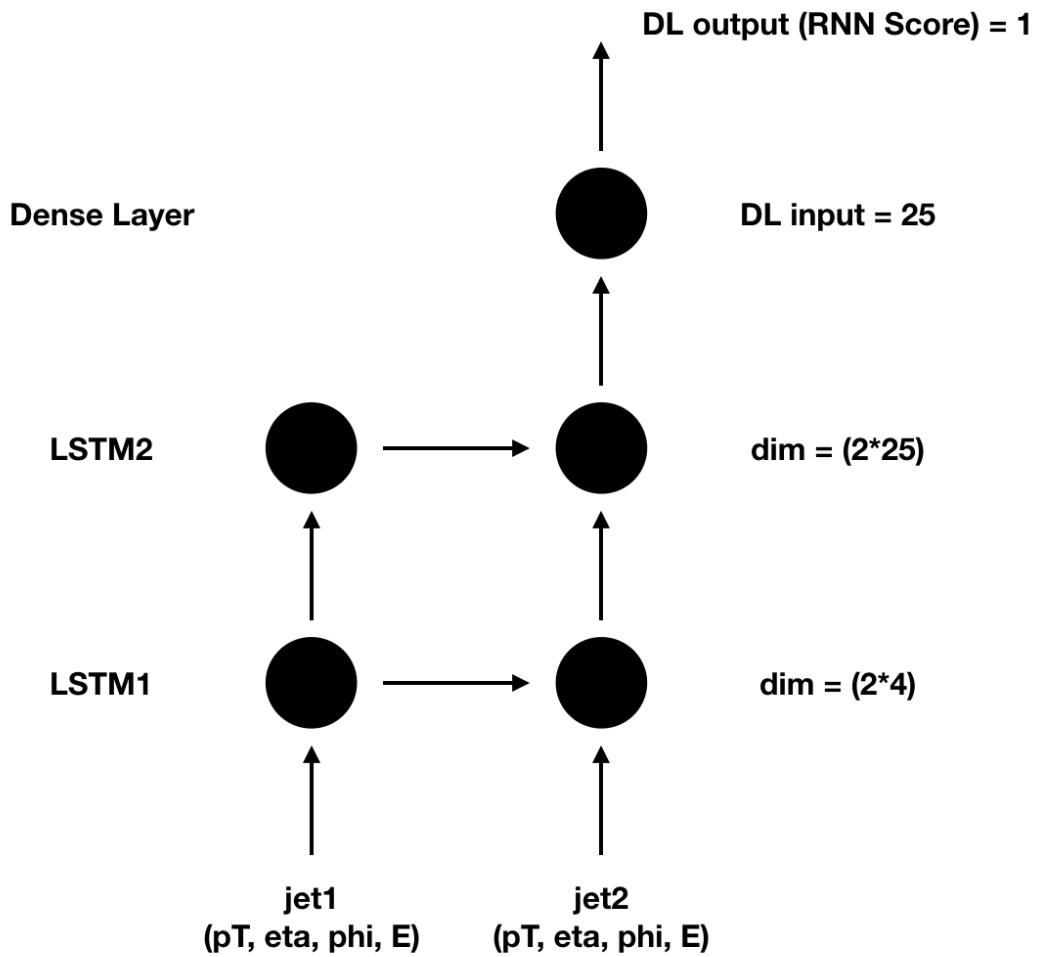
978 The architecture of the RNN is shown in Figure 9.2. The RNN is composed  
979 of Long Short Term Memory Cells (LSTM) that extract meaningful information  
980 and retain it. The logic embedded in the LSTM is shown in Figure ???. LSTMs  
981 are useful for VBF event classification for events with two jets, where using the  
982 kinematic properties of both jets (and their correlations) will lead to more efficient  
983 event classification.

984 In this RNN architecture, the VBF candidates are first passed to a masking  
985 layer which checks the number of jets in the event. If there is only one jet, only one  
986 vertical LSTM layer is used. The output of masking is then passed to a LSTM,  
987 with a tanh activation function. The output of the LSTM is then passed to a  
988 second horizontal LSTM layer (and vertical LSTM layer if there are two jets in

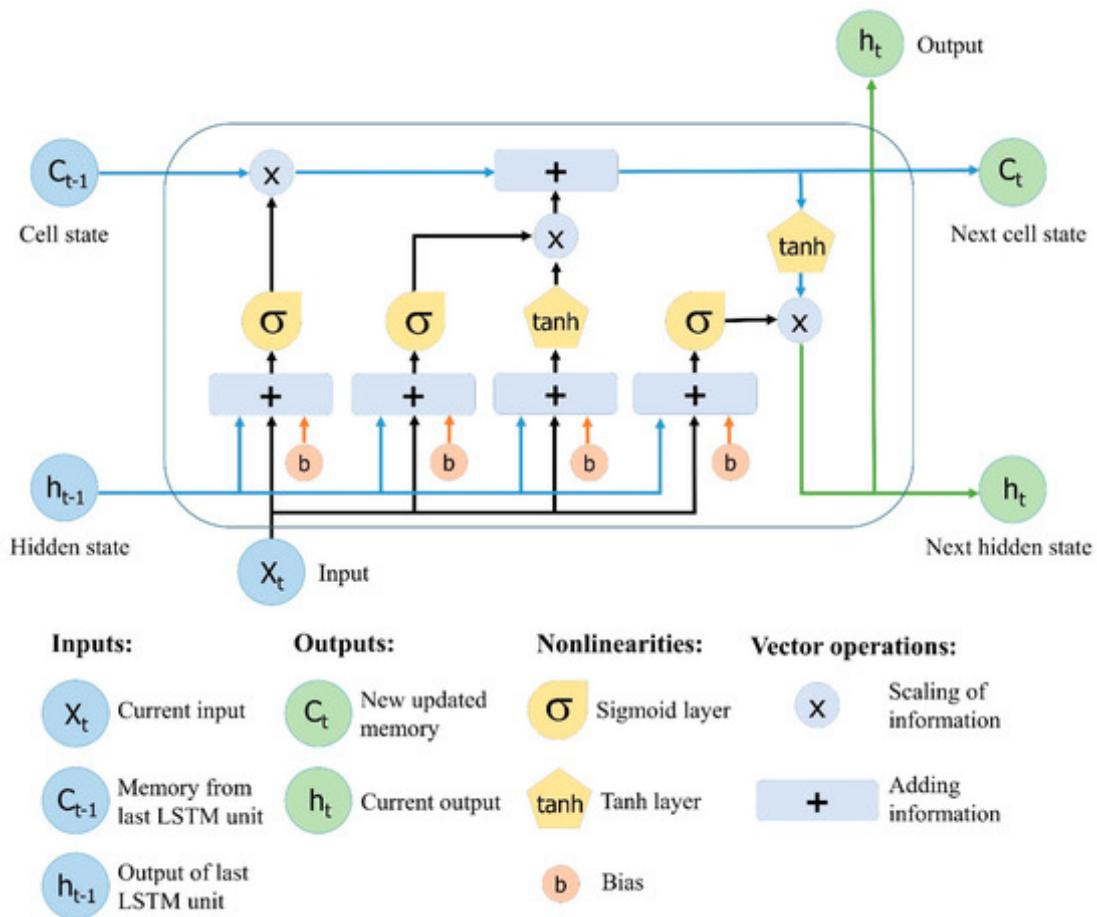
989 the event). Finally the output of the last LSTM cell is passed to a dense layer  
990 and then to a sigmoid activation layer, leading to an overall RNN score.

991 The weights and other parameters of the network are learned by training the  
992 network with HVT VBF and non-VBF signals and all simulated backgrounds over  
993 200 epochs with an Adam Optimizer [14]. To prevent overfitting during training,  
994 dropout is applied to RNN weights and training is truncated if the network pa-  
995 rameters are unchanged after ten iterations [24]. Figure 9.4 shows the ROC curve  
996 for the RNN using k-fold cross validation [20].

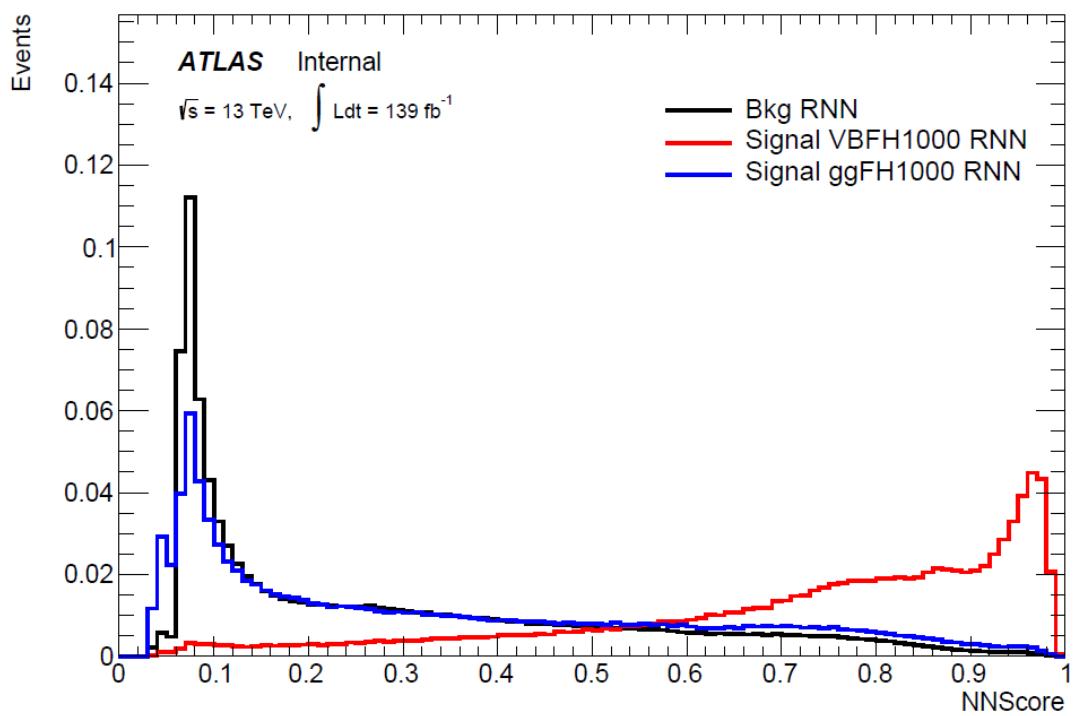
997 Figure 9.3 shows the RNN discriminant for backgrounds, non-VBF signals,  
998 and VBF signals. The RNN score is  $\sim 0$  for non-VBF signals and background  
999 processes and  $\sim 1$  for VBF processes. Figure 9.5 shows the limits for various signal  
1000 processes based on the RNN cut applied. Requiring the RNN score to be  $> 0.8$   
1001 was chosen as it provided the best analysis significance for this final state and  
1002 the  $\nu\nu qq$  and  $\ell\ell qq$  channels, which this channel will be combined with for future  
1003 publications.



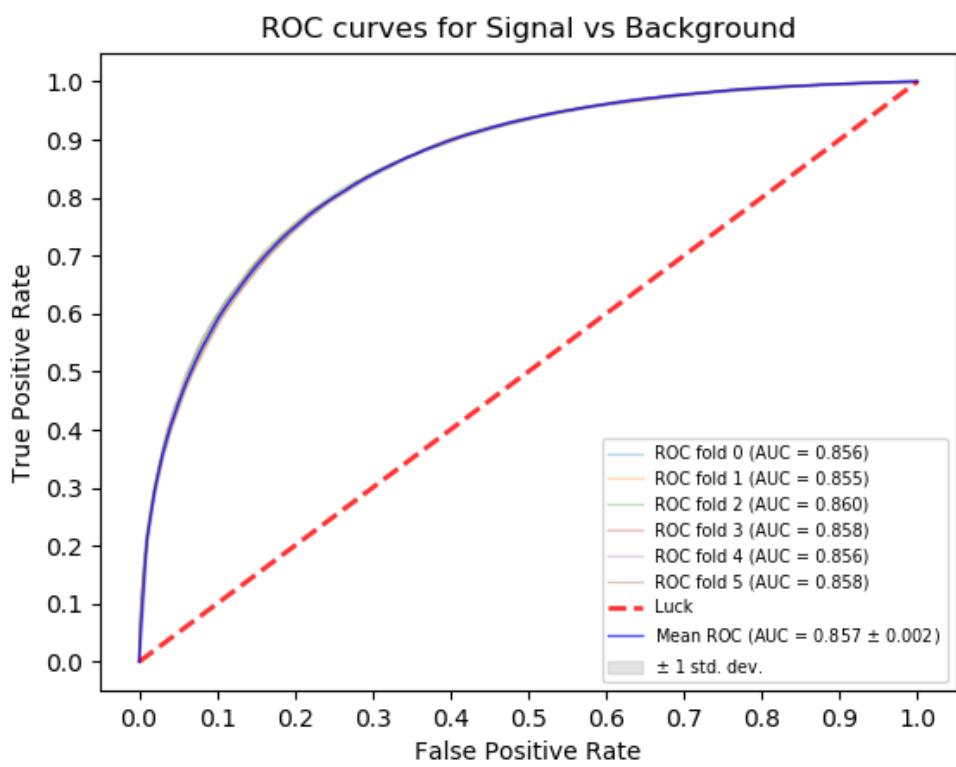
**Figure 9.1:** RNN architecture. Natasha add caption



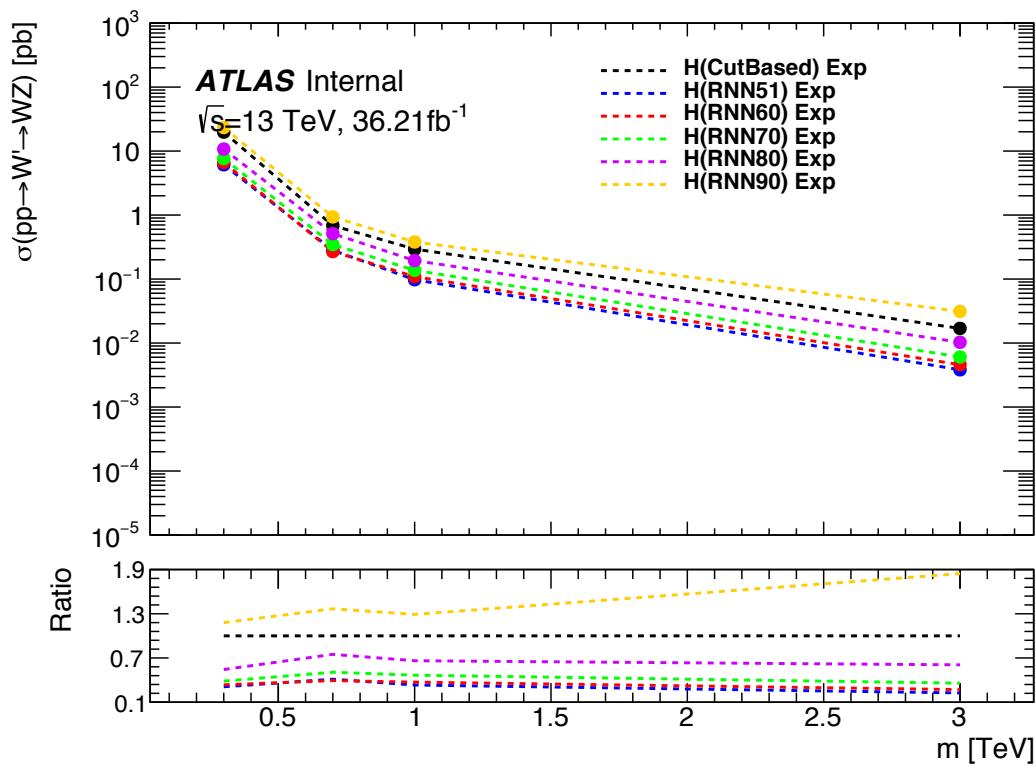
**Figure 9.2:** This figure shows the embedded logic in LSTM cells. This image was taken from [23], where a more in depth discussion about LSTMs may be found.



**Figure 9.3:** RNN Score distribution for ggF and VBF signals and backgrounds.



**Figure 9.4:** ROC curve using k-fold validation for RNN.



**Figure 9.5:** Comparison of GGF Z' limits for different RNN score selections. The bottom panel shows the ratio of the upper limits set for different RNN cuts to the cut-based analysis. In this panel smaller numbers, indicate that the expected upper limit is smaller than the cut-based analysis, which is desired.

## 1004 9.4 Signal Region Definitions

1005 Signal regions are constructed to be dominated by signal and used in the final  
1006 likelihood fit to look for a bump in the reconstructed resonance mass distribution.  
1007 Once an event is classified by the RNN, it must pass topological cuts that maximize  
1008  $S/\sqrt{B}$ . To efficiently select events with a  $W \rightarrow \ell\nu$  candidate exactly one tight  
1009 lepton is required and  $E_T^{miss} > 100(60)$  GeV and  $p_{T,\ell\nu} > 200(75)$  GeV in the  
1010 merged (resolved) analysis to suppress the fake lepton backgrounds.

1011 The resonances this search probes are expected to be produced approximately  
1012 at rest with the two resulting bosons produced back-to-back. For this reason,  
1013 it is required that the minimum value of  $(p_{T,\ell\nu}, p_{T,J})/m_{WV} > 0.35(0.25)$  for the  
1014 non-VBF (VBF) category.

1015 To reduce  $t\bar{t}$  contamination in the merged HVT  $Z'$  and  $G_{KK}$  analyses, events  
1016 with at least one  $b$ -jet with  $\Delta R > 1.0$  from the large-R jet are excluded. High  
1017 purity signal regions require the  $D_2$  and  $W/Z$  mass window cut to be passed,  
1018 whereas the low purity region only requires the  $W/Z$  mass window cut to be  
1019 passed. More boosted jets, are more likely to fail the  $D_2$  cut due to track merging.  
1020 Therefore, by using high and low purity regions, the signal acceptance is increased.

1021 The HVT  $W'$  resonance search uses tagged and untagged regions to minimize  
1022 backgrounds and increase signal acceptance. For events to be classified as tagged  
1023 the large-R jet must contain exactly two  $b$ -tagged VR jets. Untagged events must  
1024 have no more than one  $b$ -tagged jet matched to the large-R jet. These selections  
1025 are shown in Table 9.2.

1026 Events failing the merged selection are then re-analyzed in the resolved cat-  
1027 egory. To enhance resolved signals, the event should contain two high- $p_T$  boson  
1028 candidates that are back-to-back in the  $\phi$  as shown by the selections in Table 9.3.  
1029 Again, to suppress the  $t\bar{t}$  backgrounds, events are required to have no additional

**Table 9.2:** Summary of selection criteria used to define the signal region (SR),  $W$ +jets control region ( $W$  CR) and  $t\bar{t}$  control region ( $t\bar{t}$  CR) for merged 1-lepton channel.

Selection		SR		W CR (WR)		$t\bar{t}$ CR (TR1)	
		HP	LP	HP	LP	HP	LP
$W \rightarrow \ell\nu$	Num of Tight leptons	1					
	Num of Loose leptons	0					
	$E_T^{\text{miss}}$	$> 100 \text{ GeV}$					
	$p_T(\ell\nu)$	$> 200 \text{ GeV}$					
$W/Z \rightarrow J$	Num of large- $R$ jets	$\geq 1$					
	$D_2$ cut	pass	fail	pass	fail	pass	fail
	$W/Z$ mass window cut	pass	pass	fail	fail	pass	pass
	Numb. of associated VR track jets $b$ -tagged	For $Z \rightarrow J$ : $\leq 1$ ( $= 2$ ) for untagged (tagged) category					
	$\min(p_{T,\ell\nu}, p_{T,J}) / m_{WV}$	$> 0.35(0.25)$ for DY/ggF (VBF) category					
	Top-quark veto	Num of $b$ -tagged jets outside of large- $R$ jet	0	$\geq 1$			
Pass VBF selection			no (yes) for DY/ggF (VBF) category				

1030  $b$ -jets for the HVT  $Z'$  and  $G_{KK}$  analyses. A summary of the resolved selections  
 1031 is shown in Table 9.3.

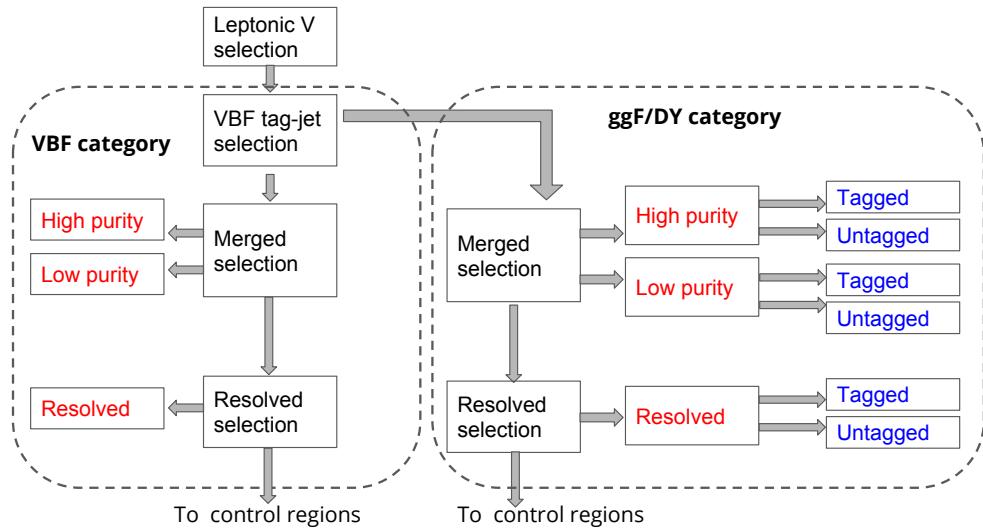
1032 The analysis cutflow in Figure 9.6 shows how the different categories are pri-  
 1033 oritized. Events classified as VBF events are classified as merged high purity, low  
 1034 purity or resolved signal region selections sequentially. If the event does not pass  
 1035 any of these selections but passes a VBF control region selection it is classified as  
 1036 a VBF CR event.

1037 If the event fails all VBF categories, it is then checked if it passes the merged  
 1038 high purity, low purity or resolved signal region selections (NB: for the  $WZ$  decay  
 1039 modes all the regions have tagged and untagged categories). If the event fails all  
 1040 the non-VBF signal region selections, it is then kept for non-VBF control region  
 1041 selections, if it passes those selections. Control region selection are discussed more  
 1042 in 10.1.

1043 Overall, for the Drell-Yan HVT  $Z'$  and gluon-gluon fusion  $G_{KK}$  signals there  
 1044 are 3 signal regions. For the Drell-Yan HVT  $W'$  signal there are 6 signal regions.  
 1045 For VBF HVT  $W'$  and  $Z'$  signals there are 3 signal regions.

**Table 9.3:** The list of selection cuts in the resolved analysis for the  $WW$  and  $WZ$  signal regions (SR),  $W+jets$  control region (WR) and  $t\bar{t}$  control region (TR).

cuts	SR	$W$ CR (WR)	$t\bar{t}$ CR (TR1)		
$W \rightarrow \ell\nu$	Number of Tight leptons	1			
	Number of Loose leptons	0			
	$E_T^{\text{miss}}$	$> 60$ GeV			
	$\cancel{p}_T(\ell\nu)$	$> 75$ GeV			
$W/Z \rightarrow jj$	Number of small-R jets	$\geq 2$			
	Leading jet $p_T$	$> 60$ GeV			
	Subleading jet $p_T$	$> 45$ GeV			
	$Z \rightarrow q\bar{q}$	$78 < m_{jj} < 105$ GeV	$50 < m_{jj} < 68$ GeV or $68 < m_{jj} < 98$ GeV	$105 < m_{jj} < 150$ GeV	$50 < m_{jj} < 150$ GeV
	Num. of $b$ -tagged jets	For $Z \rightarrow jj$ : $\leq 1$ ( $= 2$ ) for untagged (tagged) category			
Topology cuts	$\Delta\phi(j, \ell)$	$> 1.0$			
	$\Delta\phi(j, E_T^{\text{miss}})$	$> 1.0$			
	$\Delta\phi(j, j)$	$< 1.5$			
	$\Delta\phi(\ell, E_T^{\text{miss}})$	$< 1.5$			
	$\min(p_{T,\ell\nu}, p_{T,jj}) / m_{WW}$	$> 0.35(0.25)$ for DY/ggF (VBF) category			
Top veto	Number of additional $b$ -tagged jets	0	$\geq 1$		
	Pass VBF selection	no (yes) for DY/ggF (VBF) category			



**Figure 9.6:** Event Categorization. Natasha write more.

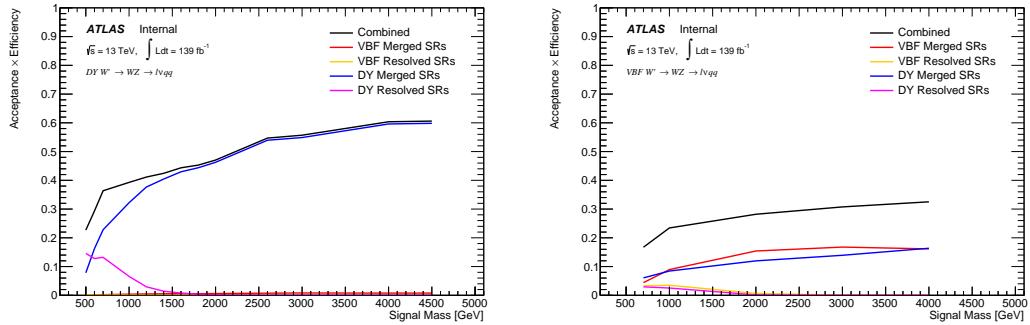
## 1046 9.5 Selection Acceptance and Efficiency

1047 The signal acceptance is the ratio of the number of signal events selected to  
 1048 the number of signal events generated at truth level, which does not account for

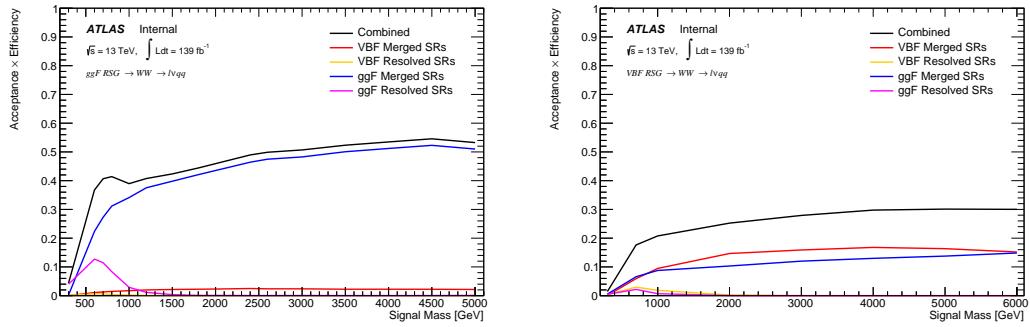
1049 detector effects. The signal efficiency is the ratio of the number of reconstructed  
 1050 events selected and the number of truth events selected, which accounts for de-  
 1051 tector effects. The expected number of signal events is given by the product of  
 1052 these two quantities:

$$A \cdot \epsilon = \frac{N_{\text{events selected}}^{\text{truth}}}{N_{\text{events generated}}^{\text{truth}}} \cdot \frac{N_{\text{events selected}}^{\text{reco}}}{N_{\text{events selected}}^{\text{truth}}} = \frac{N_{\text{events selected}}^{\text{reco}}}{N_{\text{events generated}}^{\text{truth}}} \quad (9.1)$$

1053 The distributions of  $A \cdot \epsilon$  as a function of the resonance mass for the different spin  
 1054 models are shown in Figures 9.7 - 9.8.



**Figure 9.7:** Selection acceptance times efficiency for the  $W' \rightarrow WZ \rightarrow \ell\nu qq$  events from MC simulations as a function of the  $W'$  mass for Drell-Yan (left) and VBF production (right), combining the merged HP and LP signal regions of the  $WV \rightarrow \ell\nu J$  selection and the resolved regions of the  $WV \rightarrow \ell\nu jj$  selection. Note: the VBF selection acceptance for the DY  $W'$  is approximately zero in the left plot.



**Figure 9.8:** Selection acceptance times efficiency for the  $G \rightarrow WW \rightarrow \ell\nu qq$  events from MC simulations as a function of the  $G$  mass for (a) Drell-Yan and (b) VBF production, combining the merged HP and LP signal regions of the  $WW \rightarrow \ell\nu J$  selection and the resolved regions of the  $WW \rightarrow \ell\nu jj$  selection. Note: the VBF selection acceptance for the ggF  $G''_{KK}$  is approximately zero in the left plot.

1055 **Chapter 10**

1056 **Background Estimate**

1057 Backgrounds from  $VV$ ,  $t\bar{t}$ , single-top,  $W+\text{jets}$ ,  $Z+\text{jets}$  are simulated as de-  
1058 scribed in 7.2. The dominant backgrounds for this search are from  $W+\text{jet}$  and  
1059  $t\bar{t}$  processes. To more accurately model the  $m_{VV}$  distribution from these back-  
1060 grounds in the SRs, control regions are constructed for each as described in 10.1.  
1061 The  $t\bar{t}$  and  $W+\text{jets}$  control regions are called TCR and WCR, respectively. There  
1062 are separate control regions for VBF and non-VBF regions as well as for each  
1063 region (merged HP, merged LP, resolved). For the HVT  $W'$  search there are also  
1064 tagged and untagged control regions (where tagged refers to events with two  $b$ -jets  
1065 inside the large-R jet).

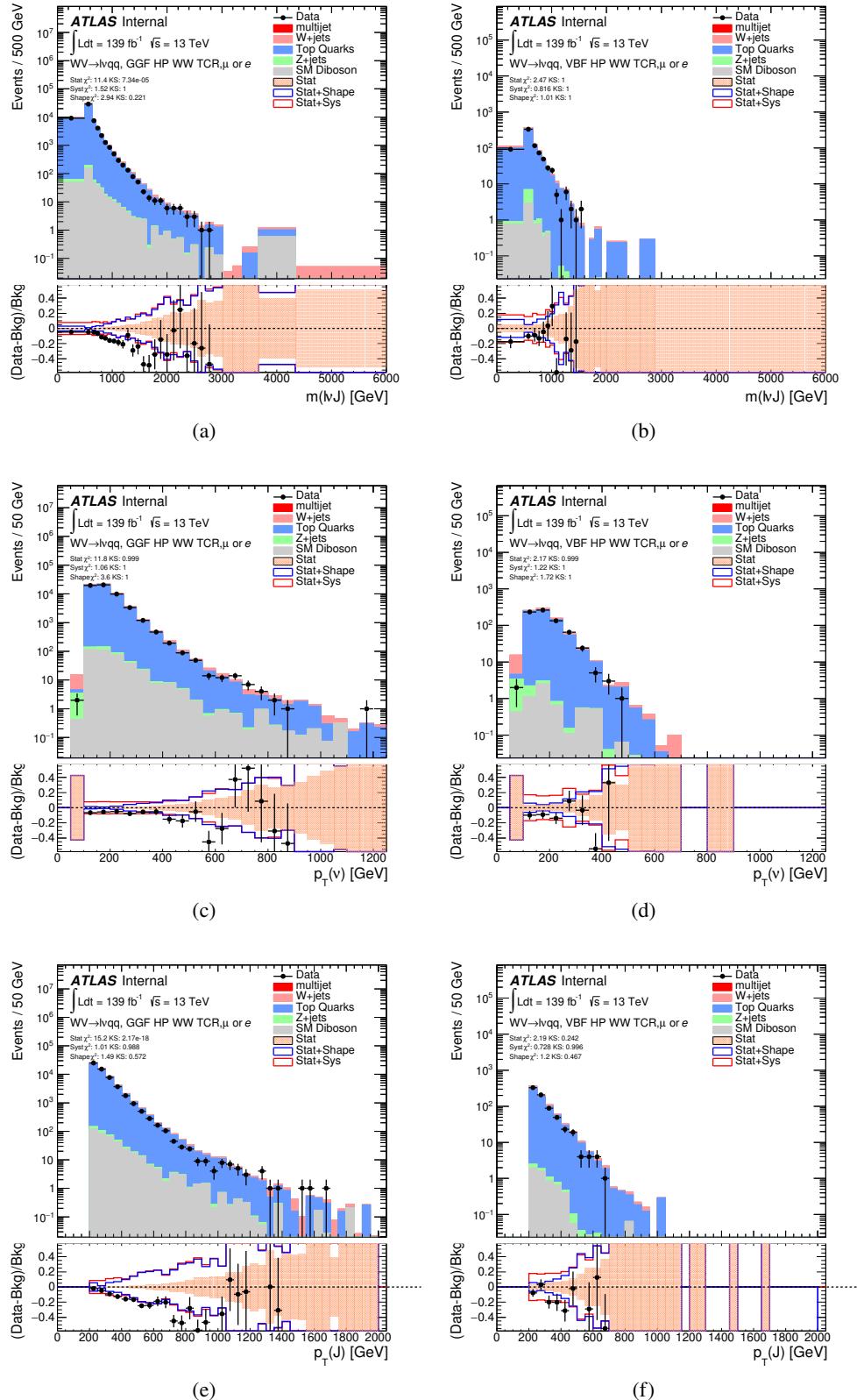
1066 The aforementioned backgrounds containing real leptons and are well-modeled  
1067 with simulated samples. Backgrounds with fake leptons (also referred to as the  
1068 multijet background) are not well-modeled with simulation. For this reason, the  
1069 multijet background is extracted from data as described in 10.2.

## 1070 10.1 Control Regions

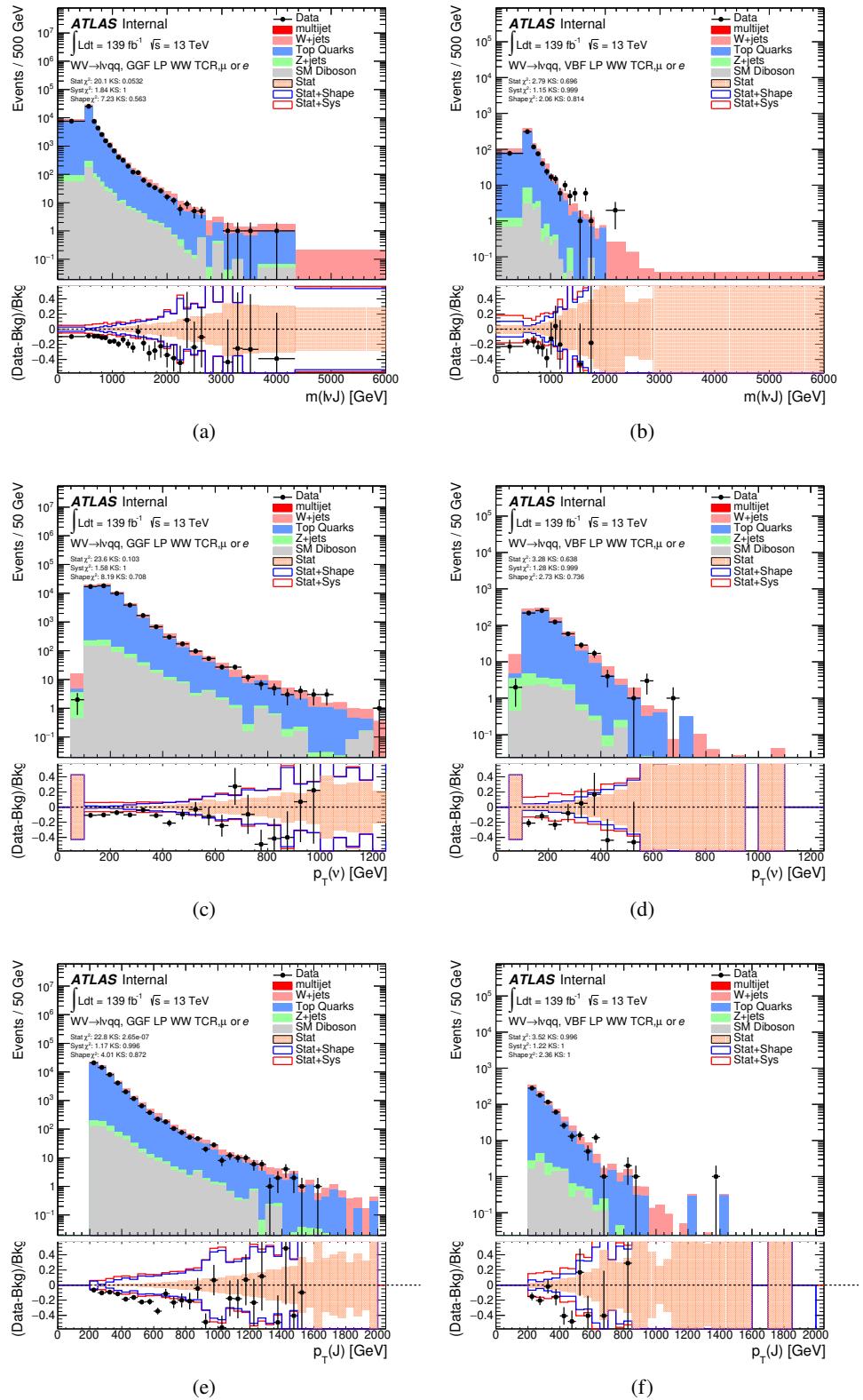
1071 The TCRs have the same selections as the SRs, but at least one  $b$ -jet in the  
1072 event (that is not within the catchement area of the large-R jet for the merged  
1073 analysis). The WCR shares the SR selections, but has different jet mass require-  
1074 ments. For the merged analyses, the large-R jet must fail the  $W/Z$  tagger jet mass  
1075 cut. In the resolved analyses,  $m_{jj}$  must be  $50 < m_{jj} < 68$  GeV or  $105 < m_{jj} < 150$   
1076 GeV.

1077 The distributions for some the variables used in merged analysis (e.g. reso-  
1078 nance mass,  $p_T(\nu)$ ,  $p_T(J)$ ) for top control regions (non-VBF and VBF HP and LP  
1079 regions) are shown in Figure 10.1-10.4. The distributions for the variables used  
1080 in the resolved analysis (e.g. resonance mass,  $p_T(\nu)$ ,  $p_T(j_1/j_2)$ ) in the TCR are  
1081 shown in Figure 10.5, 10.6. In general, in these plots the simulated distributions  
1082 match the data well, which is necessary to have confidence in the prediction yields  
1083 in the signal regions.

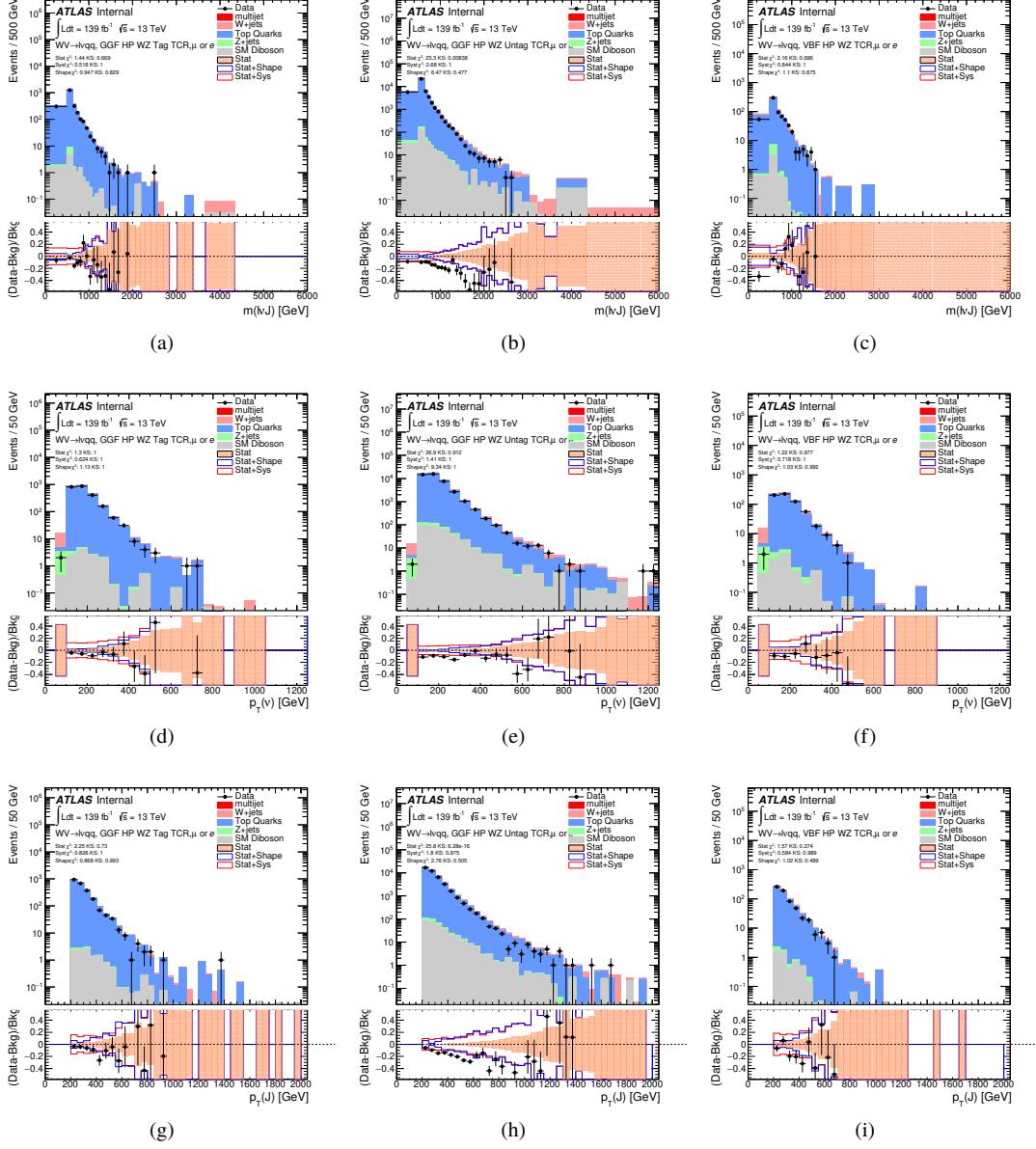
1084 To more accurately model the two dominant backgrounds in this analysis,  
1085  $W$ +jets and  $t\bar{t}$ , control regions are constructed for each. These control regions  
1086 are dominated by these processes and used to extract normalization factors in  
1087 the final likelihood fit that are then used in the signal region estimates. For the  
1088  $t\bar{t}$  control region the event must contain at least one such  $b$  jet. The WCR is  
1089 constructed using the  $m_{jj/J}$  mass window sidebands. All other backgrounds are  
1090 estimated using simulation, except fake lepton backgrounds, which are derived  
1091 using a data-driven method.



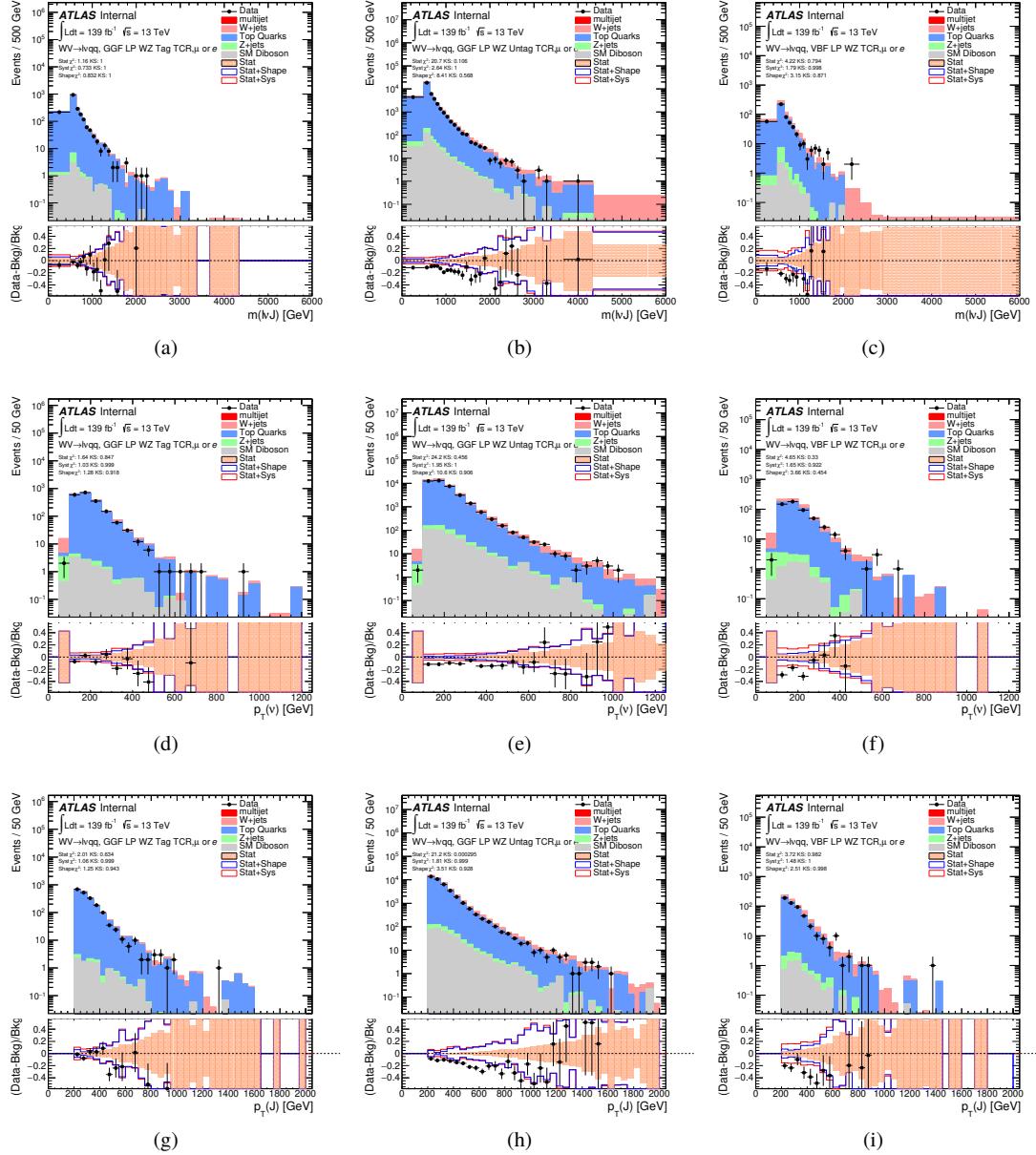
**Figure 10.1:** Data MC comparison for the merged  $WW$  HP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.<sup>93</sup>



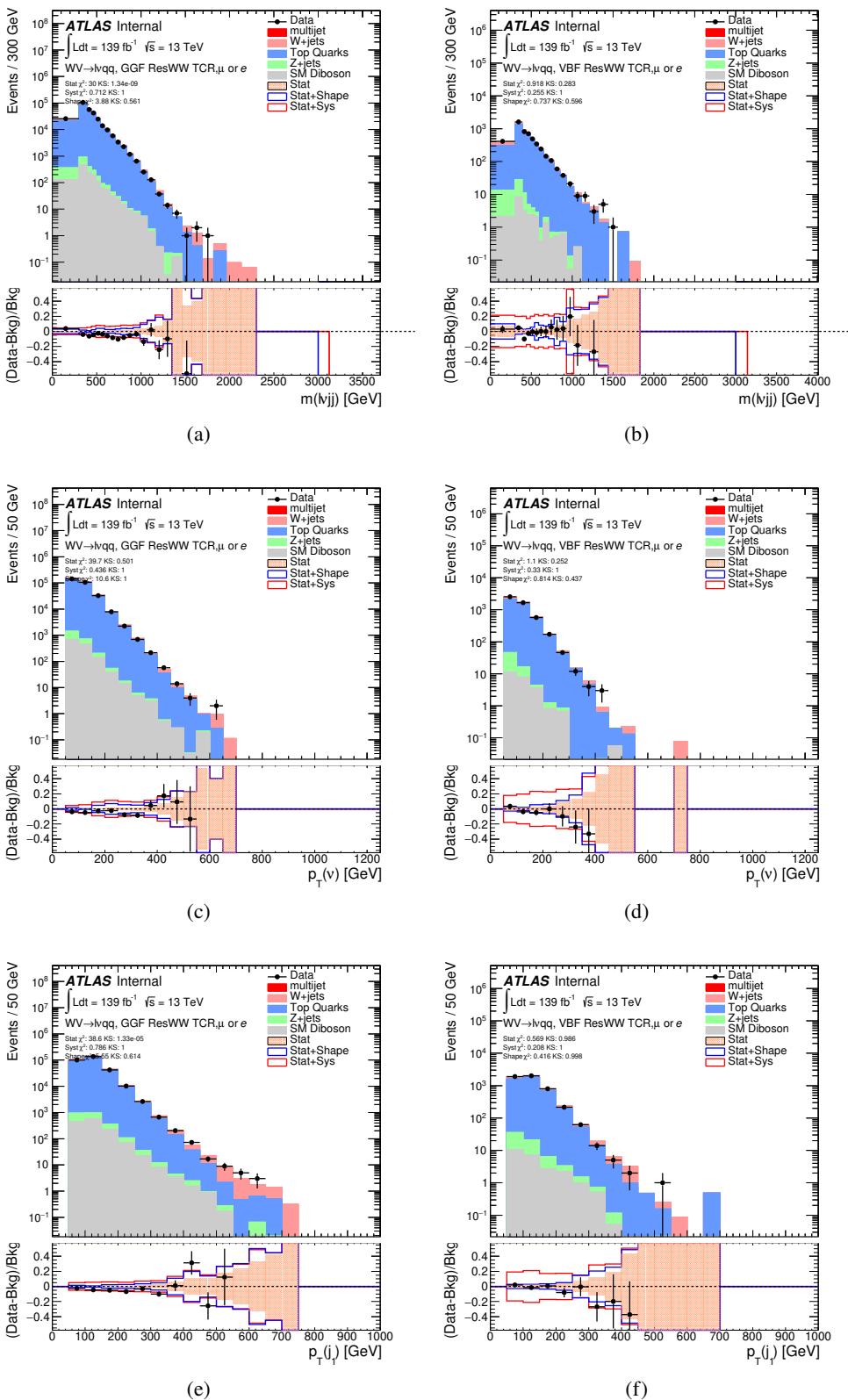
**Figure 10.2:** Data MC comparison for the merged  $WW$  LP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.



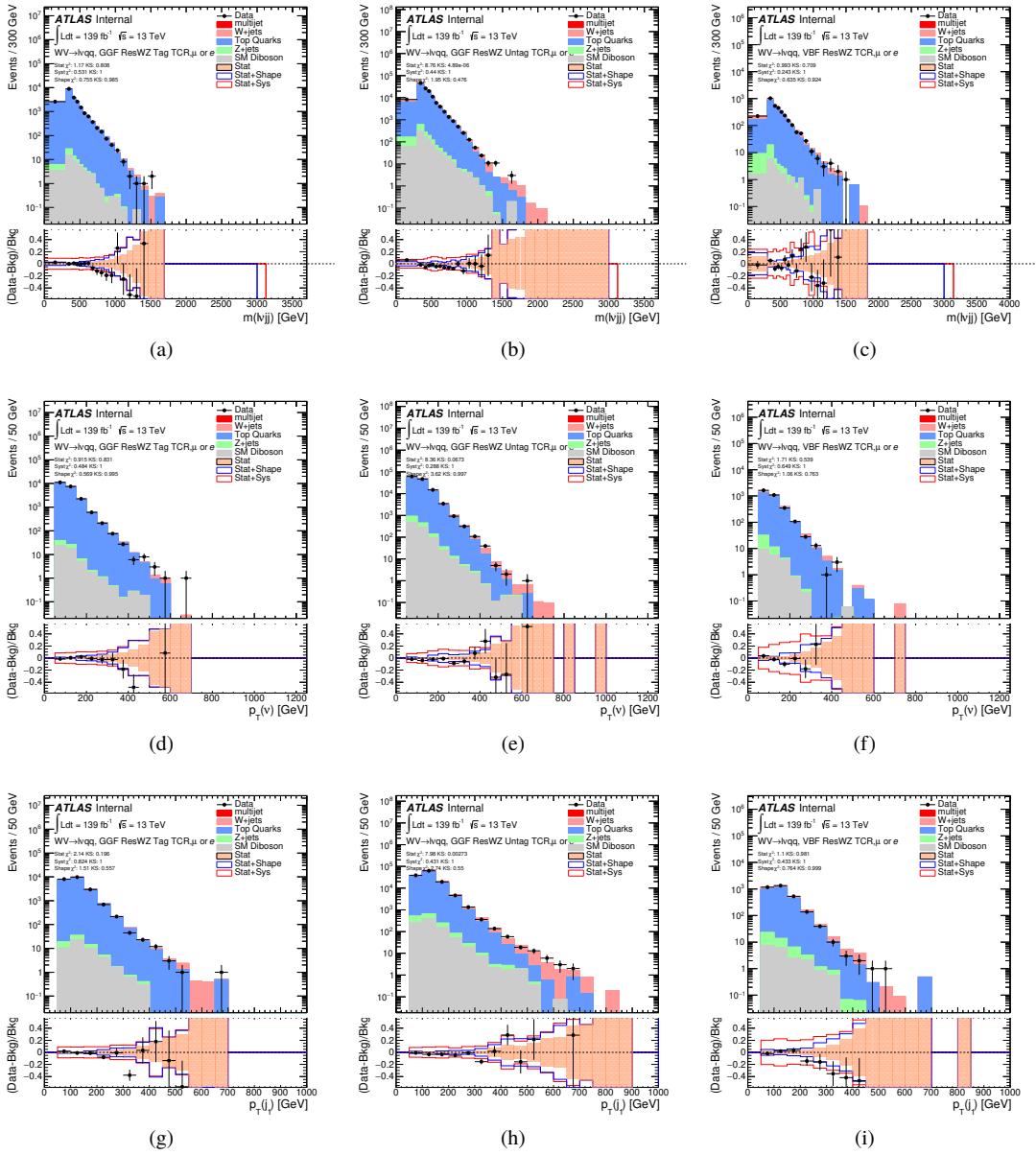
**Figure 10.3:** Data MC comparison for the merged  $WZ$  HP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.



**Figure 10.4:** Data MC comparison for the merged  $WZ$  LP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.



**Figure 10.5:** Data MC comparison for the resolved  $WW$  TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.



**Figure 10.6:** Data MC comparison for the resolved  $WZ$  TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.

## 1092 10.2 Fake Lepton Backgrounds

1093 The fake lepton backgrounds for this search are not well-modeled with simu-  
1094 lation. For this reason, this background is extracted from data. Fake electrons  
1095 often arise from fake jets and converted photons while non-prompt muons usually  
1096 arise from heavy flavor decay products. This predominately occurs at lower lepton  
1097 momentums, and therefore is only considered in the resolved analysis.

1098 Fake electrons generally fail the electron ID criteria and fake muons fail the  
1099 muon isolation requirement. Therefore, separate multijet samples are derived for  
1100 the fake electron and muon samples. For each sample the  $m_{WV}$  template shape  
1101 is derived for the SR and WCR selections using the same SR and WCR cuts but  
1102 with inverted lepton requirements as seen in Table 10.1. NB: By inverting the  
1103 lepton isolation/identification criteria the SRs and CRs are orthogonal.

1104 To derive the multijet template in a given SR, first the multijet template in  
1105 the WCR is derived, called the MJCR template. This template is calculated using  
1106 events that pass the WCR selection but with the inverted lepton criteria. The  
1107  $E_T^{miss}$  distribution for the MJCR is given by the difference between data and the  
1108 simulated samples in the MJCR. The  $E_T^{miss}$  distribution of those events is then  
1109 added to the simulated backgrounds in the WCR. The floating background and  
1110 multijet normalizations of the MJCR in this region are then fit to the data. The  
1111 fitted MJCR is then used as the multijet sample in the WCR.

1112 The fitted normalizations from the MJCR template are then used to construct  
1113 the multijet template in the SR (MJSR). The MJSR is constructed from events  
1114 that pass the SR selections but with the inverted lepton criteria. Again, the  
1115 difference between the data and simulated backgrounds in this region gives MJSR  
1116 template shape in  $m_{WV}$ . This shape is then scaled by the fitted normalizations  
1117 from the MJCR. These fitted electron and muon muon multijet templates are then

1118 used as the multijet samples in the SRs. The normalizations of the electron and  
1119 muon multijet samples are parameters in the final likelihood fit.

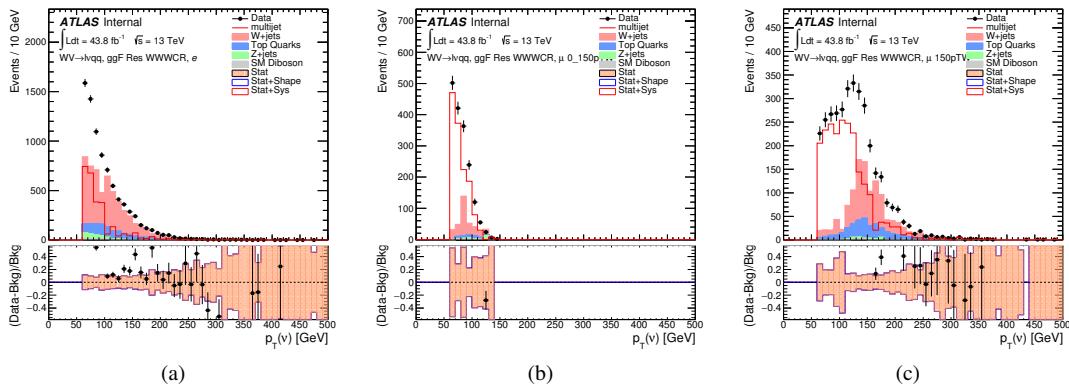
1120 This template method was validated using WCR and full Run 2 data. The  
1121 results of the fit are shown in Table 10.2. The multijet contribution in the muon  
1122 channel for  $p_T^W > 150$  GeV is consistent with zero, and therefore neglected in  
1123 the final fit. Applying the extracted normalization factor to MJCR in WCRs for  
1124 various kinematic variables such as  $E_T^{miss}$ ,  $W$  transverse mass, lepton  $p_T$ , and the  
1125 invariant mass as show in Figures 10.8 -10.17. These figures show good agreement  
1126 between the data and background estimate.

**Table 10.1:** Definitions of “inverted” leptons used in multijet control region. For the inverted muon selection,  $ptvarcone30$  is given by sum of the  $p_T$  of tracks in a cone around the muon candidate divided by the muon  $p_T$ . The size of the cone,  $\delta R$  used is  $10\text{GeV}/p_T^\mu$  or 0.3, whichever is smaller. So, as the  $p_T$  of the muon increases, the cone size used decreases. This is useful as more boosted muons are more likely to be produced in dense environments and using a smaller cone size more accurately determines the quality of the muon.

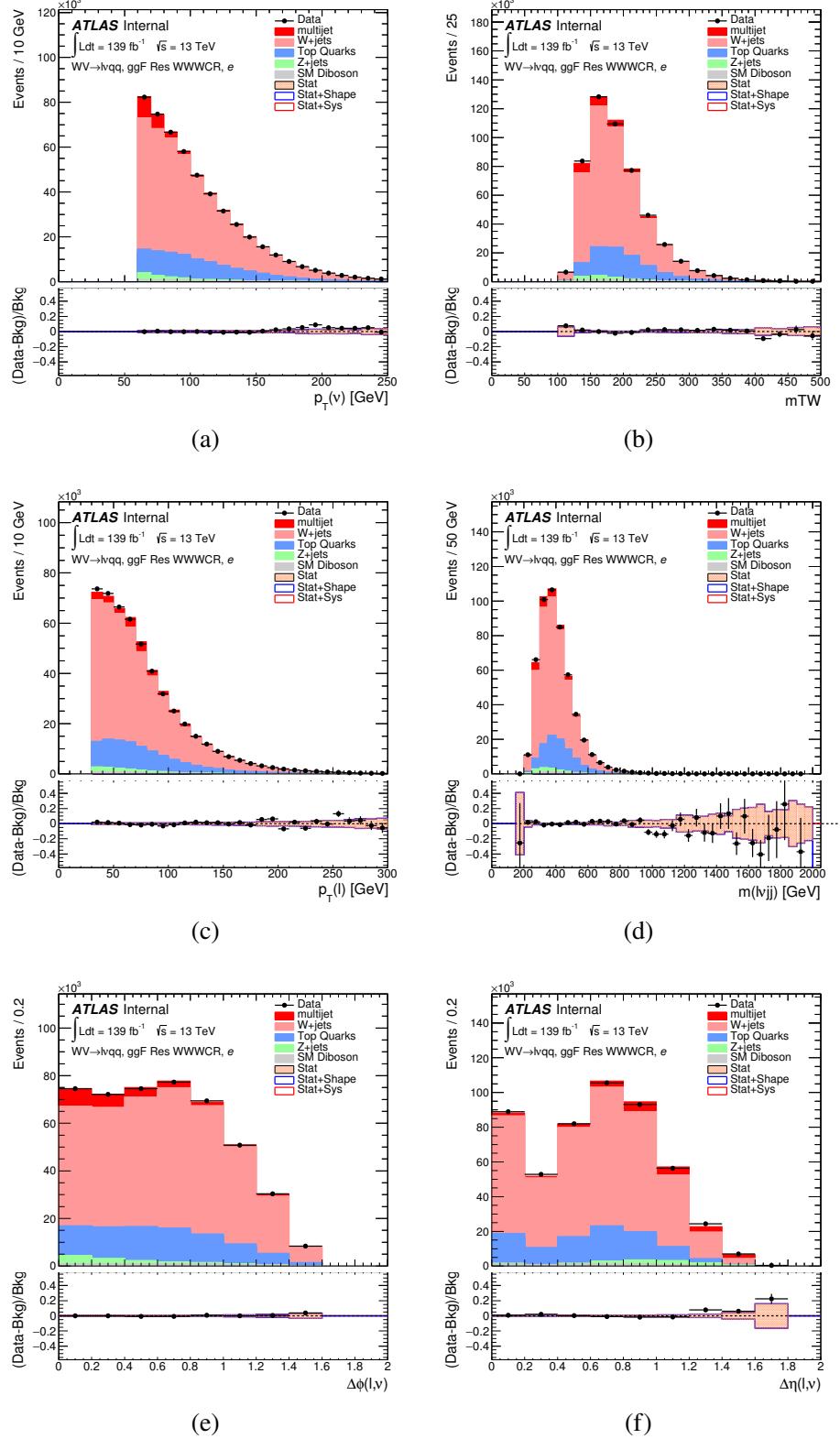
	Criterion	signal lepton	inverted lepton
Electron	ID	TightLH	MediumLH !TightLH
	Calo Isolation	FixedCutHighPtCaloOnlyIso	FixedCutHighPtCaloOnlyIso
Muon	ID	WHSignalMuon	WHSignalMuon
	Track Isolation	FixedCutTightTrackOnlyIso	!FixedCutTightTrackOnlyIso $ptvarcone30/p_T < 0.07^*$

\*Only applied to events with  $pTW < 150\text{GeV}$

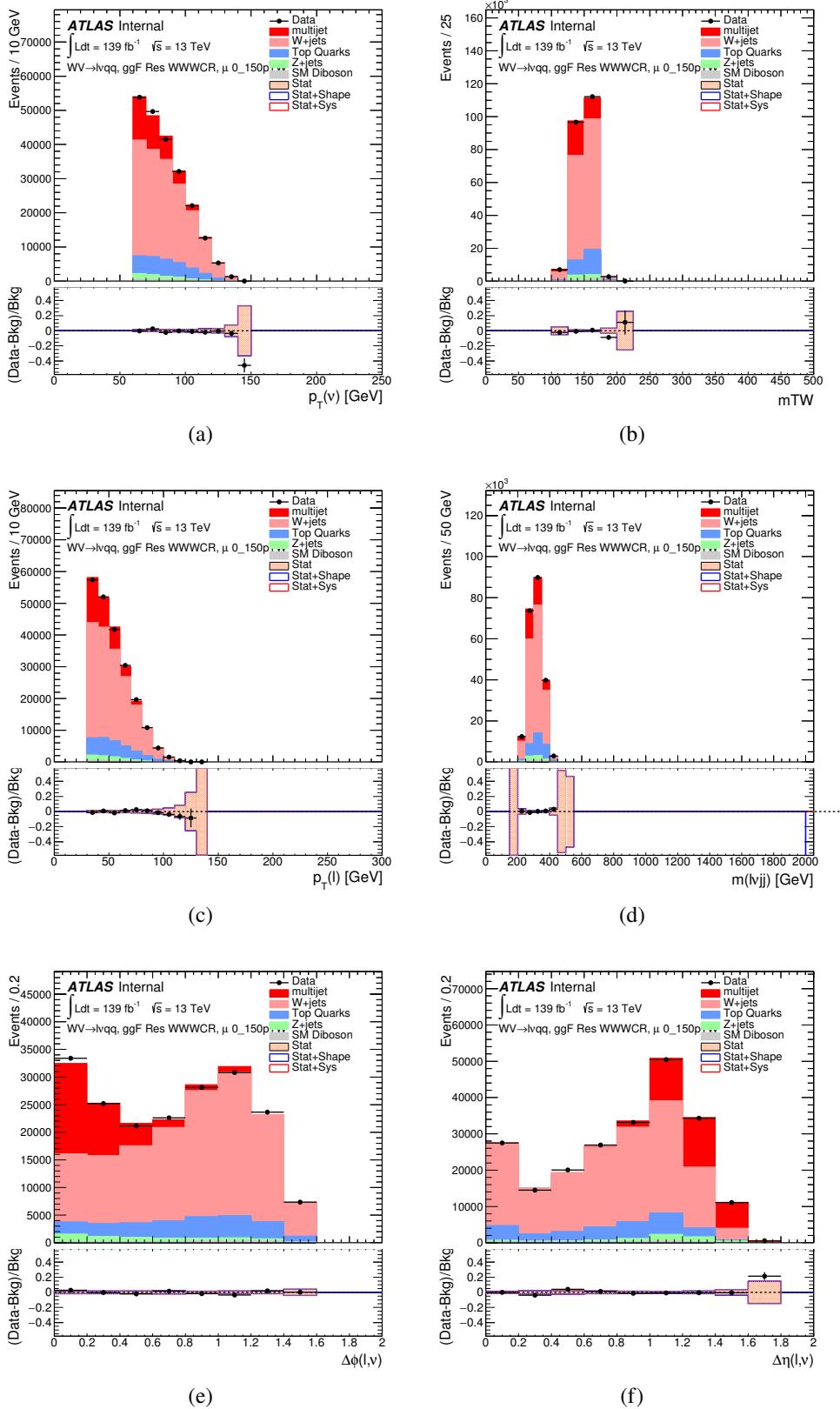
1127



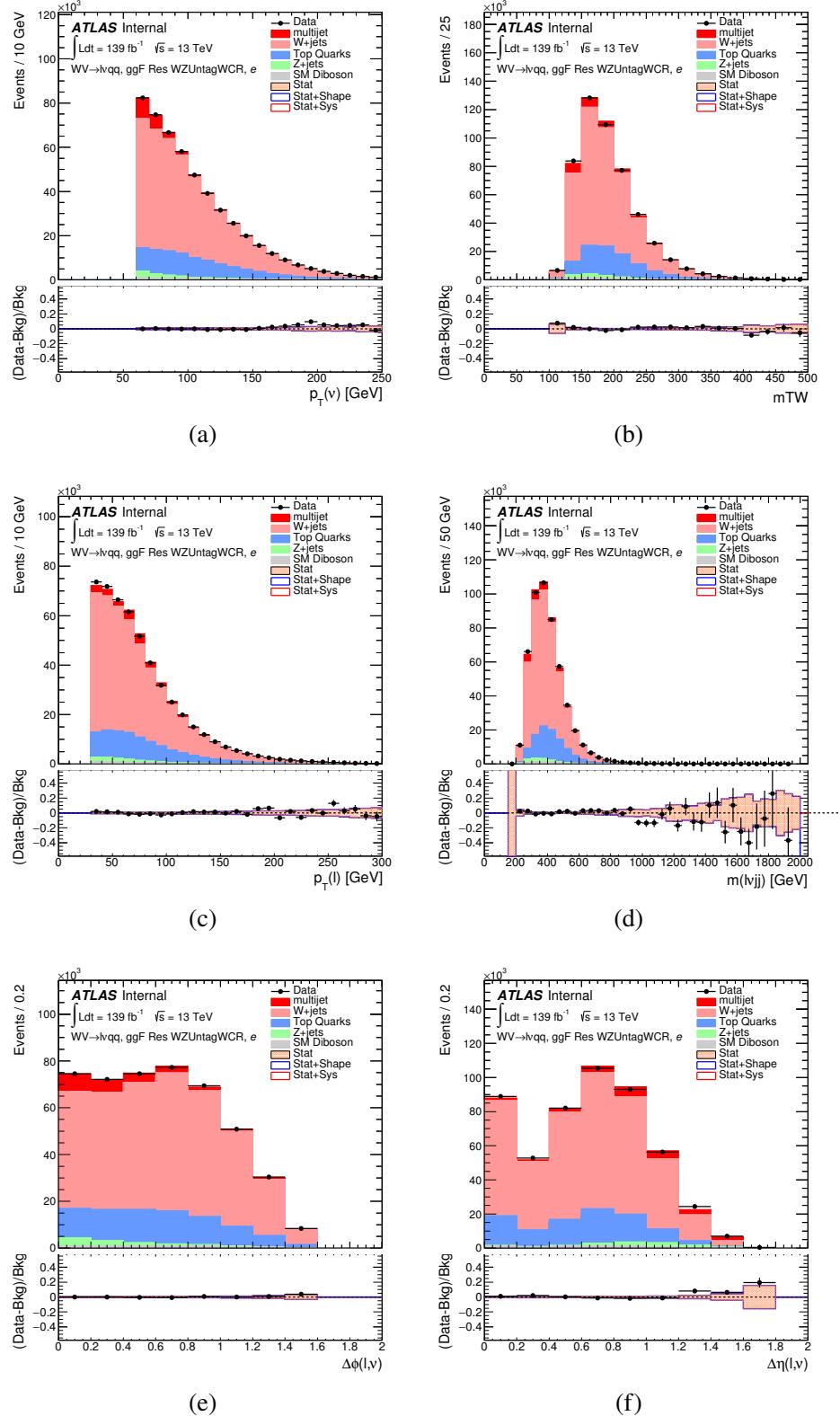
**Figure 10.7:** The  $E_T^{miss}$  distribution in MJCR for 2017 data in the electron channel(left), muon channel with  $W$ -boson  $pT < 150$  GeV (center) and  $> 150$  GeV (right). Multi-jet templates are given by the difference between the data and simulated distributions.



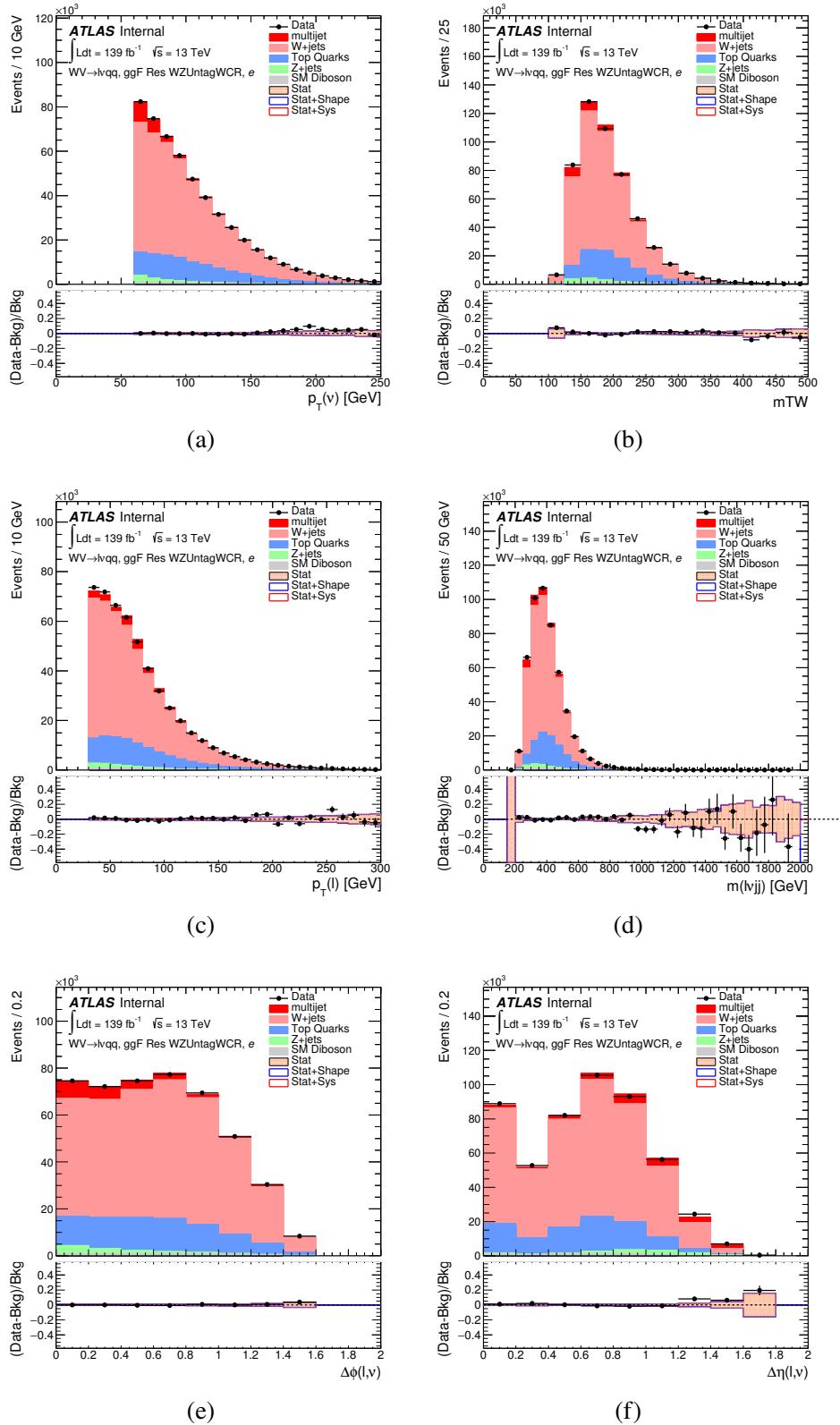
**Figure 10.8:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{\ell\nu jj}$ , lepton- $\nu$  angular distance in the  $WW$  electron channel. The MJ template is obtained from the pre-MJ-fit.



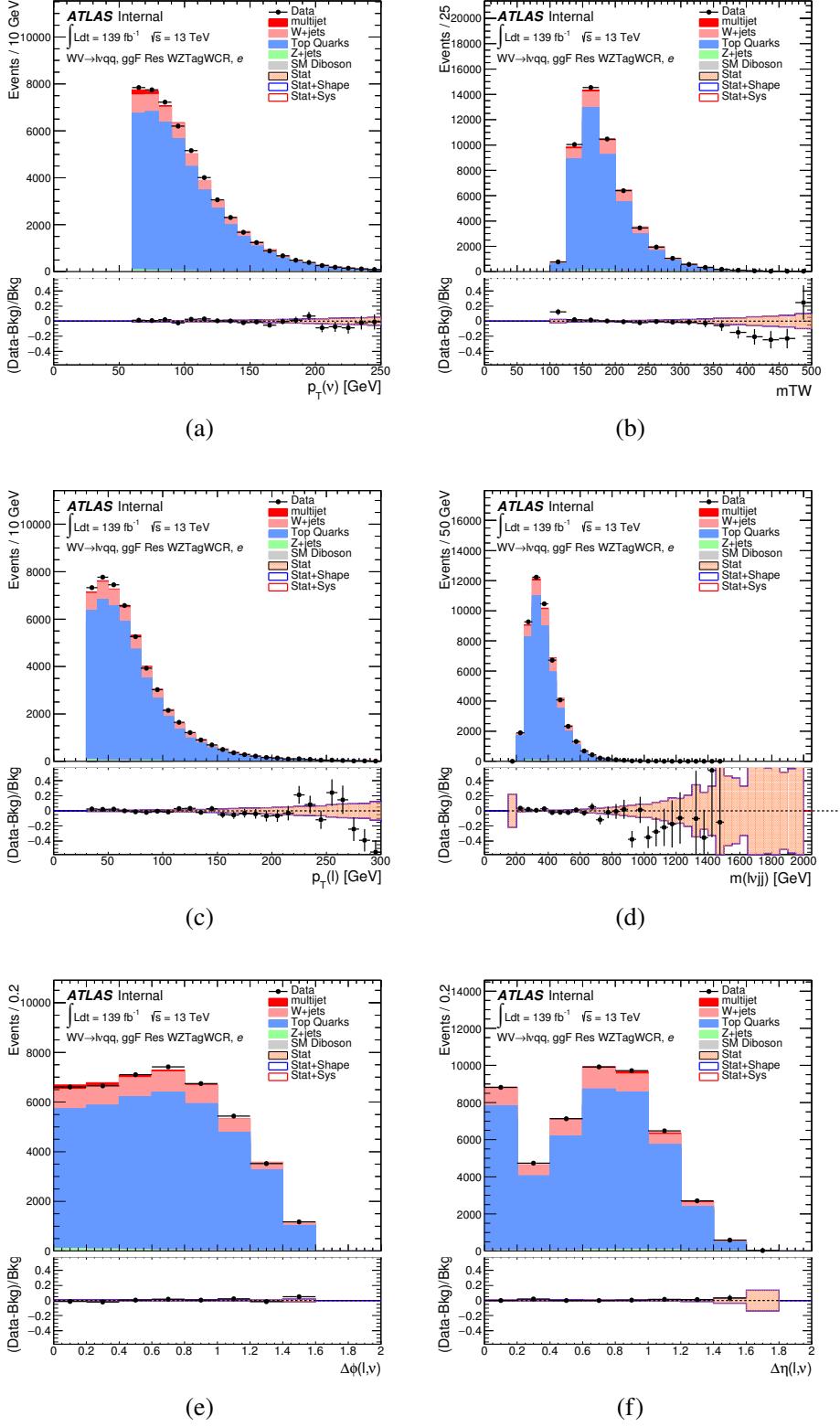
**Figure 10.9:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the  $WW$  muon channel. The MJ template is obtained from the pre-MJ-fit.



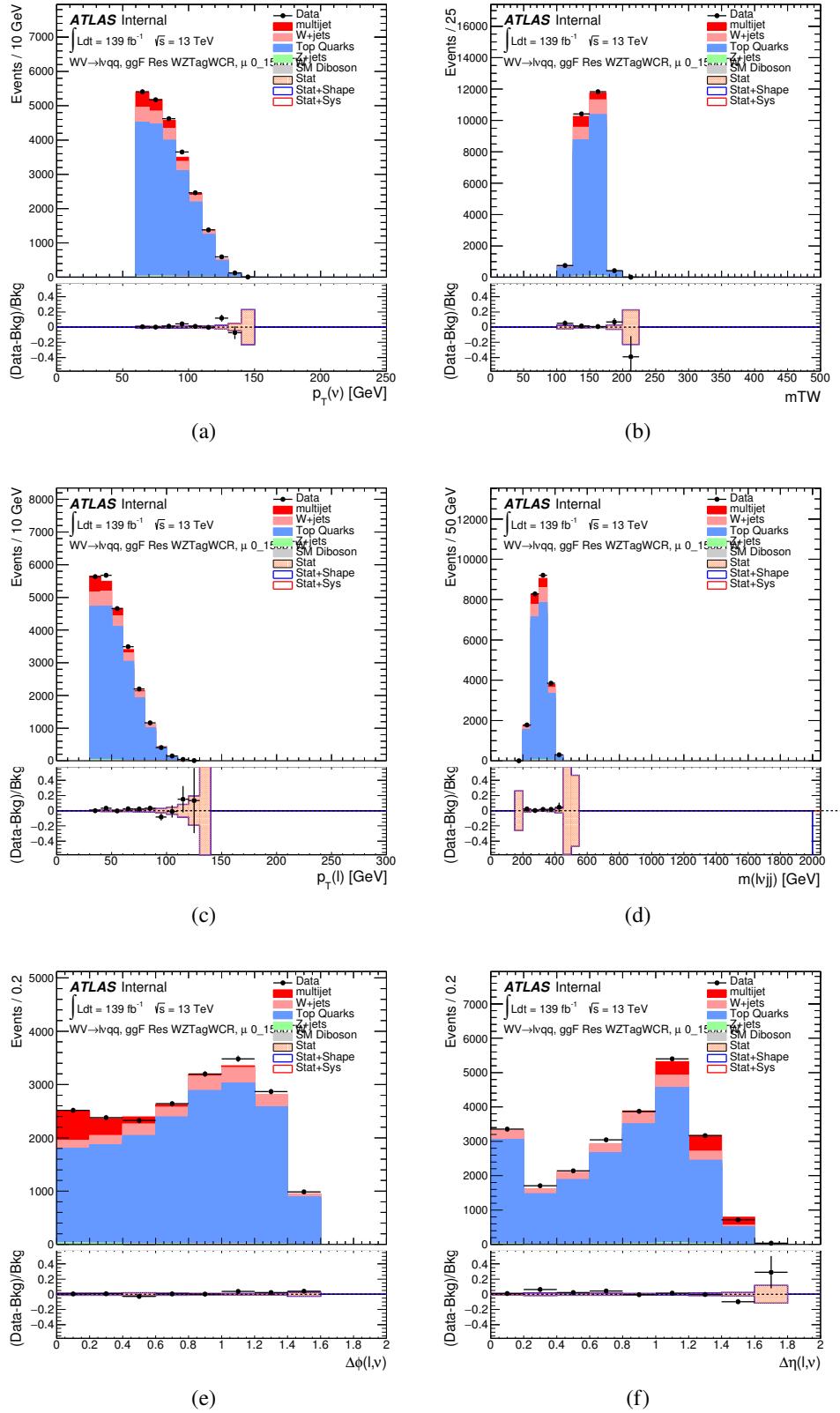
**Figure 10.10:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{\ell\nu jj}$ , lepton- $\nu$  angular distance in the  $WZ$  untag electron channel. The MJ template is obtained from the pre-MJ-fit.



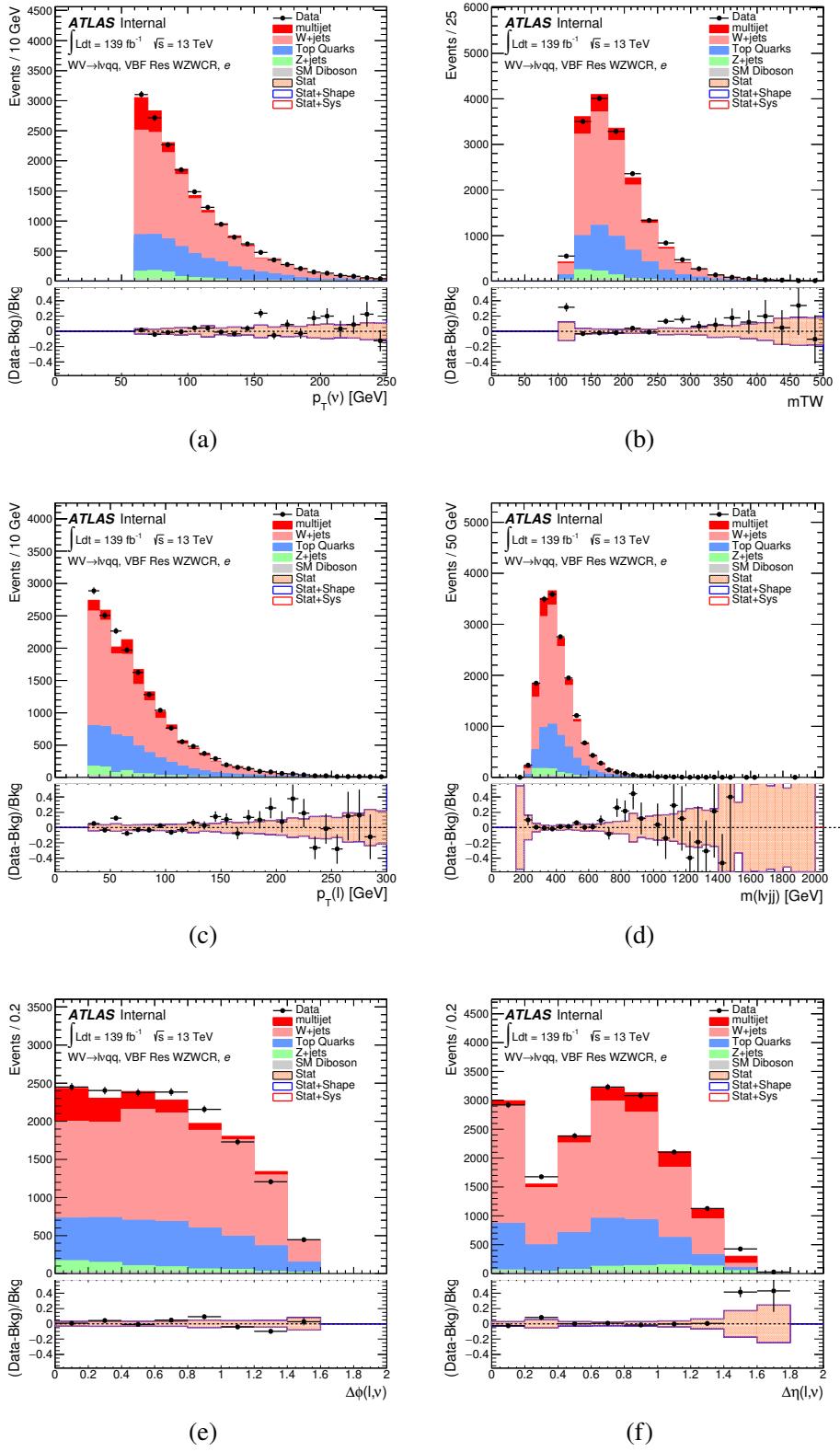
**Figure 10.11:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{\ell\nu jj}$ , lepton- $\nu$  angular distance in the  $WZ$  untag muon channel. The MJ template is obtained from the pre-MJ-fit.



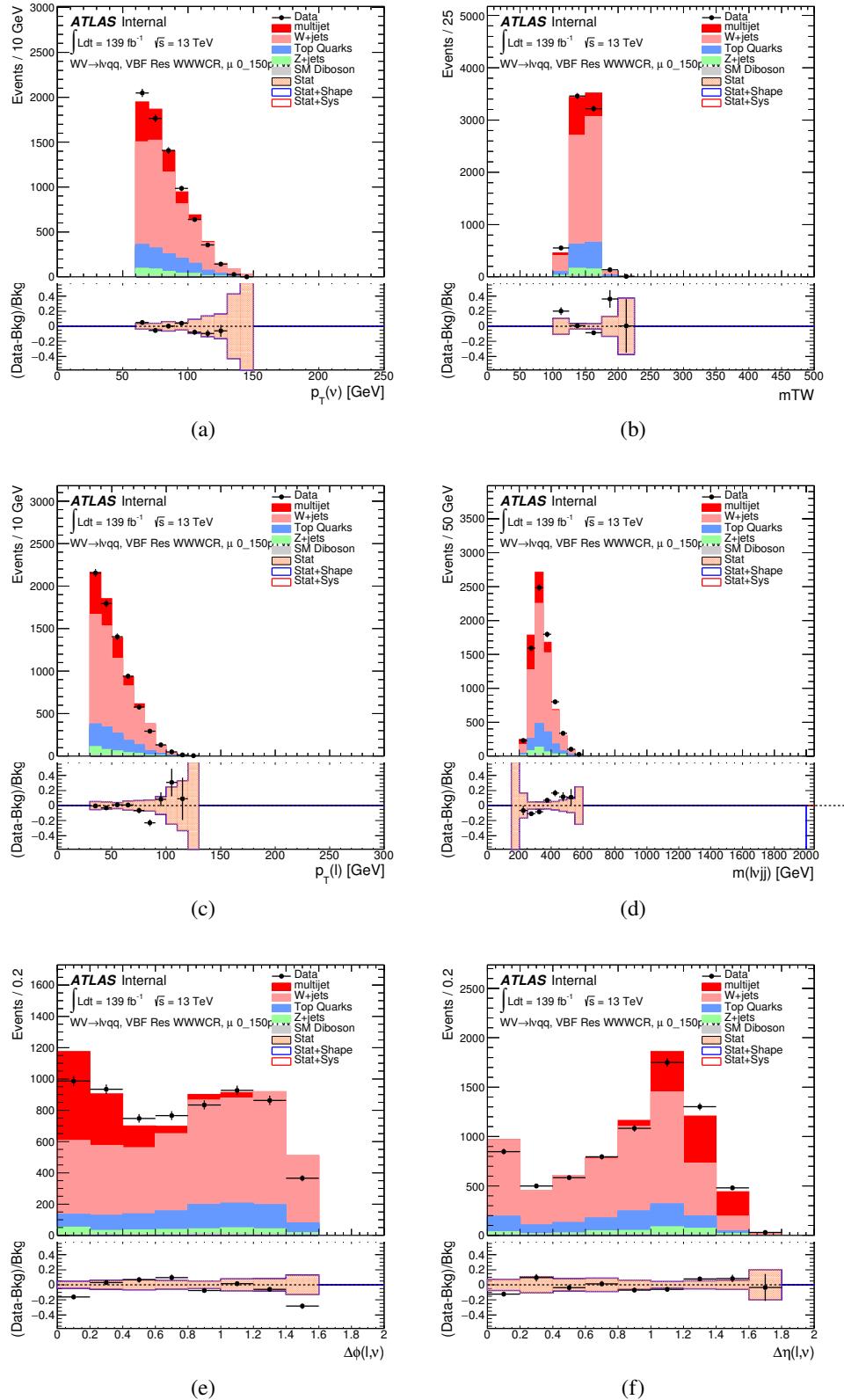
**Figure 10.12:** Postfit Data/MC comparison of distributions of  $E_T^{\text{miss}}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the  $WZ$  untag electron channel. The MJ template is obtained from the pre-MJ-fit.



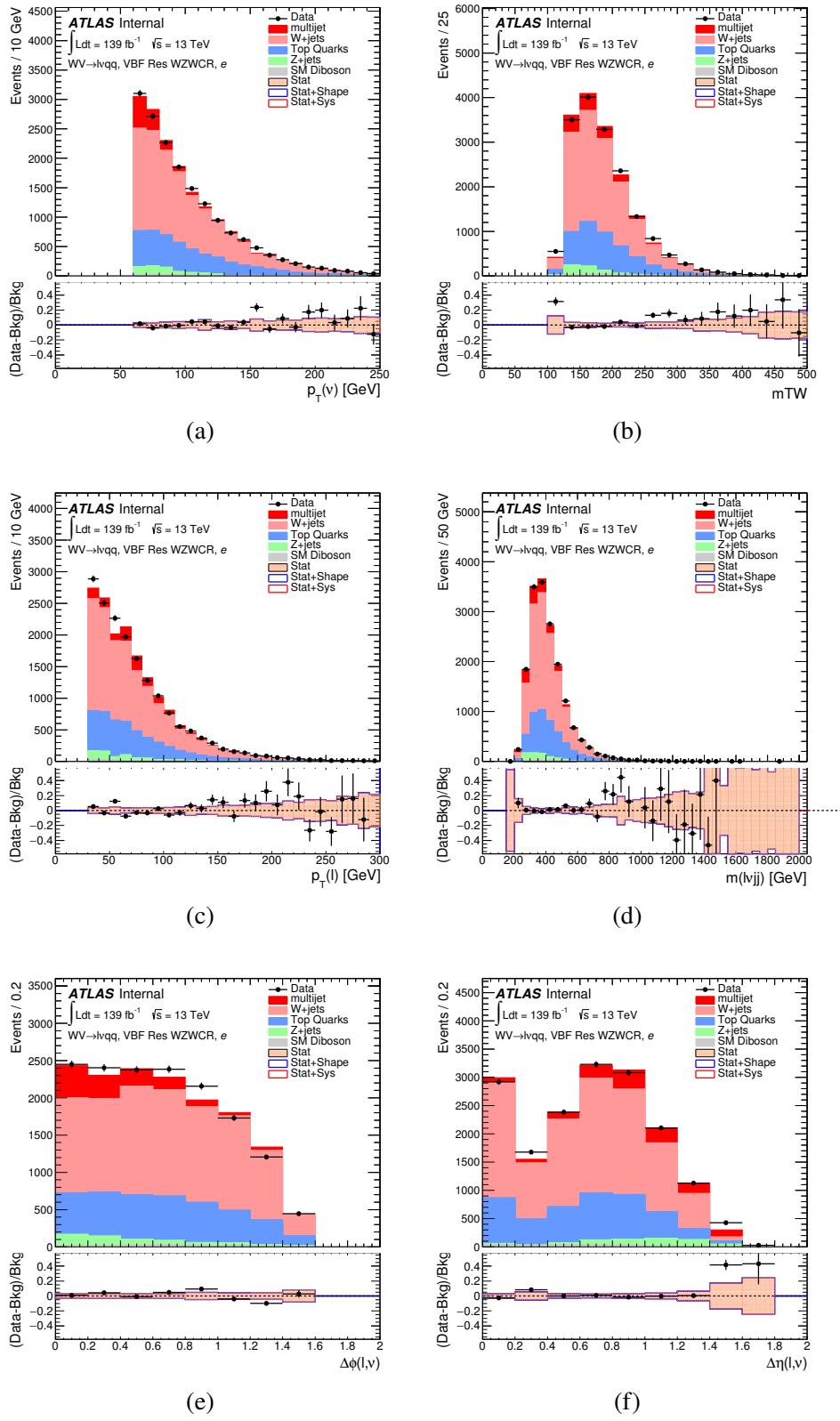
**Figure 10.13:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{\ell\nu jj}$ , lepton- $\nu$  angular distance in the  $WZ$  untag muon channel. The MJ template is obtained from the pre-MJ-fit.



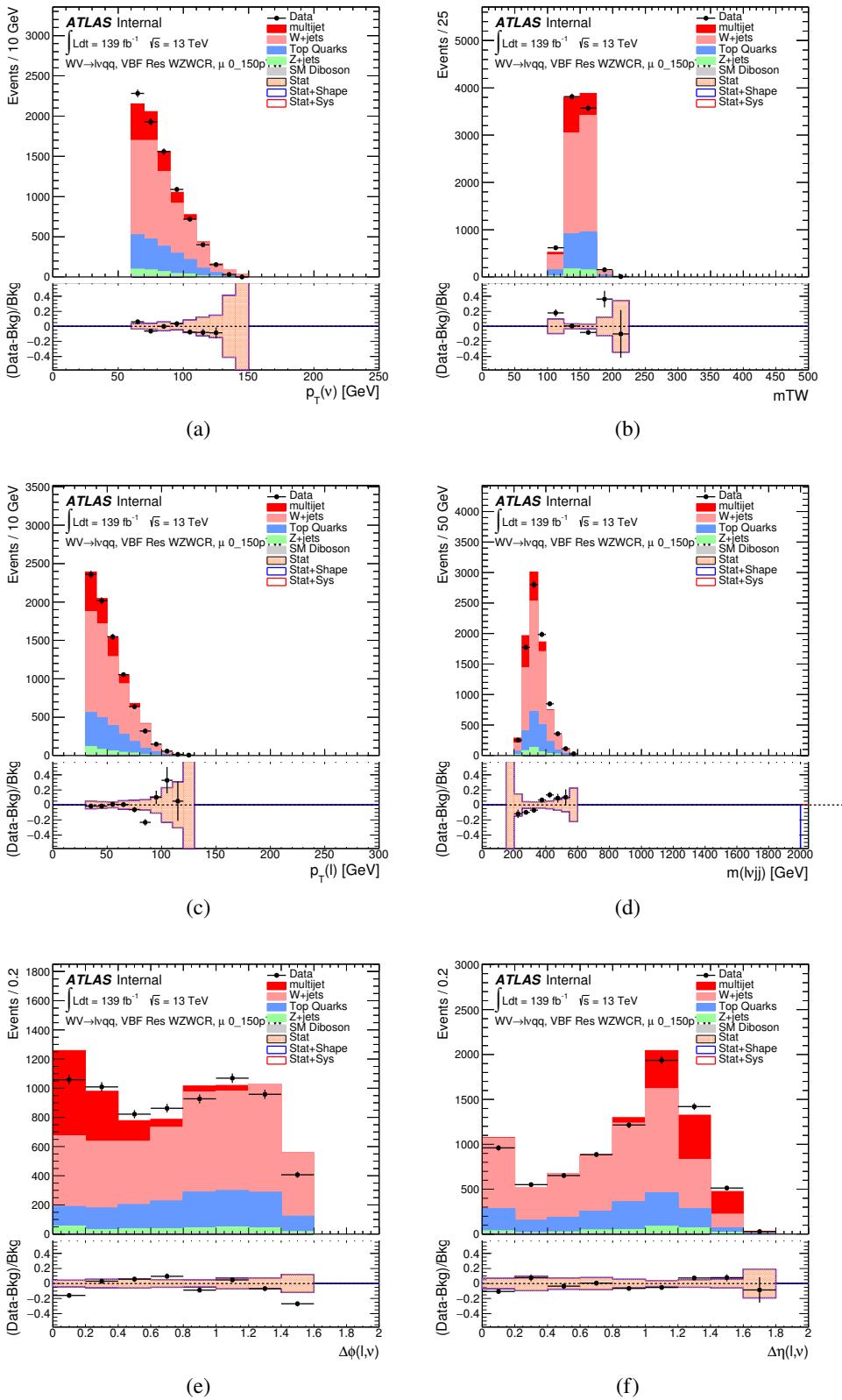
**Figure 10.14:** Postfit Data/MC comparison of distributions of  $E_T^{\text{miss}}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the VBF  $WW$  electron channel. The MJ template is obtained from the pre-MJ-fit.



**Figure 10.15:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{\ell\nu jj}$ , lepton- $\nu$  angular distance in the VBF  $WW$  muon channel. The MJ template is obtained from the pre-MJ-fit.



**Figure 10.16:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the VBF  $WZ$  electron channel. The MJ template is obtained from the pre-MJ-fit.



**Figure 10.17:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{\ell\nu jj}$ , lepton- $\nu$  angular distance in the VBF  $WZ$  muon channel. The MJ template is obtained from the pre-MJ-fit.

Full Run 2  
ggF Res WWWCR

Sample	Yield	R.U.	SF
Top&W	$645040 \pm 1971.68$	0.31%	0.998
Z&VV	24075.9		fixed
MJ_el	$24156.3 \pm 1224.62$	5.06%	3.973
MJ_mu	$35528.5 \pm 923.94$	2.60%	9.019

ggF Res WZ01bWCR

Sample	Yield	R.U.	SF
Top&W	$644690 \pm 1981.4$	0.31%	0.997
Z&VV	24075.9		fixed
MJ_el	$24366.5 \pm 1232.69$	5.05%	3.874
MJ_mu	$35528.5 \pm 921.27$	2.58%	8.746

ggF Res WZ2bWCR

Sample	Yield	R.U.	SF
Top&W	$71236.5 \pm 688.74$	0.97%	1.031
Z&VV	518.5		fixed
MJ_el	$595.63 \pm 449.34$	75.44%	0.094
MJ_mu	$1196.9 \pm 222.13$	18.56%	0.294

VBF Res WWWCR

Sample	Yield	R.U.	SF
Top&W	$19032.3 \pm 364.43$	1.91%	0.928
Z&VV	1091.63		fixed
MJ_el	$1425.73 \pm 214.42$	15.03%	0.235
MJ_mu	$1281.36 \pm 157.21$	11.83%	0.314

VBF Res WZWCR

Sample	Yield	R.U.	SF
Top&W	$21341.8 \pm 392.21$	1.84%	0.942
Z&VV	1111.75		fixed
MJ_el	$1413.76 \pm 230.36$	16.29%	0.225
MJ_mu	$1281.36 \pm 157.21$	12.27%	0.314

**Table 10.2:** Fit validation result in WCRs for 2015+16 data. The fit is done in various WCRs, in order to obtain the corresponding scale factors for MJ templates: ggF resolved WCR for the  $WW \rightarrow lvqq$  selection, ggF resolved untagged WCR for the  $WZ \rightarrow lvqq$  selection, ggF resolved tagged WCR for the  $WZ \rightarrow lvqq$  selection, VBF resolved WCR for the  $WW \rightarrow lvqq$  selection, and VBF resolved WCR for the  $WZ \rightarrow lvqq$  selection. Post-fit event yields for electroweak processes and MJ contributions are shown. The SF column shows the corresponding normalization scale factors for electroweak processes from the fit. R.U. stands for relative uncertainty.

<sub>1128</sub> **Chapter 11**

<sub>1129</sub> **Systematic Uncertainties**

<sub>1130</sub> This section describes the sources of systematic uncertainties of the  $m_{WV}$  dis-  
<sub>1131</sub> tribution. These uncertainties are divided into experimental and modeling un-  
<sub>1132</sub> certainties. Each systematic uncertainty is treated as a nuisance parameter in  
<sub>1133</sub> the final likelihood fit. The dominant systematics in this analysis arise from jet  
<sub>1134</sub> reconstruction and the generator choice for the  $V + \text{jets}$  backgrounds.

<sub>1135</sub> **11.1 Experimental Systematics**

<sub>1136</sub> The uncertainty on the integrated luminosity of the dataset used is 1.7% and  
<sub>1137</sub> a systematic in the final fit. This uncertainty was calculated using  $x - y$  beam  
<sub>1138</sub> separation scans [natasha ref P55].

<sub>1139</sub> Another source of systematic uncertainty is assigned to the pileup modeling in  
<sub>1140</sub> MC samples. This ensures simulated detector response and particle reconstruction  
<sub>1141</sub> conditions are as similar as possible. The distribution of the average number of  
<sub>1142</sub> interactions per bunch crossing applied to simulation is called the  $\mu$  profile. The  
<sub>1143</sub> pileup modeling uncertainty is accounted for by re-weighting simulated events so  
<sub>1144</sub> the average number of interactions per bunch crossing varies within its uncertainty

1145 due to systematics from vertex reconstruction [cite ATL-COM-SOFT-2015-119].

1146 The associated re-weighting factors are propagated through the entire analysis  
1147 chain to construct a systematic uncertainty on  $m_{VV}$ .

1148 The single-lepton and  $E_T^{miss}$  triggers used are not fully efficient, so scale factors  
1149 are applied to simulation to more accurately model the data. These scale factors  
1150 are given by the ratio of the distribution of offline objects before trigger selection  
1151 and after trigger selection. The associated uncertainty on these scale factors are  
1152 used in the final fit.

1153 Uncertainties on small-R jet energy scale and resolution are measured in-situ  
1154 by calculating the response between data and simulation. This analysis uses a  
1155 reduced set of JES and JER uncertainties (totaling 30 and 8 systematics, respec-  
1156 tively). This reduced set of systematics is calculated using a principal component  
1157 analysis, yielding largely uncorrelated independent systematics. These uncertain-  
1158 ties on jet energy scale and resolution (JES and JER, respectively) account for the  
1159 dependence on  $p_T$ ,  $\eta$ ,  $\mu$ , flavor response and global sequential corrections. System-  
1160 atic uncertainties associated with  $b$ -tagging are also considered. These systematics  
1161 are evaluated as uncertainties on a scale factor which accounts for the difference  
1162 in  $b$ -tagging efficiencies in data and MC, and the flavor dependence (between b,  
1163 c, and light jets).

1164 The uncertainty on the  $p_T$  scale of the large-R jets is determined by comparing  
1165 the jet's  $p_T^{calo}$  to  $p_T^{track}$  in di-jet simulation and data. In addition to this uncertain-  
1166 ties from tracking, modeling (Pythia vs Herwig), and statistical constraints are  
1167 also calculated. The large-R jet  $p_T$  resolution is given by smearing the jet  $p_T$  with  
1168 a Gaussian with a 2% width.

1169 The  $W/Z$  tagging efficiency scale factor is estimated by comparing the tagging  
1170 efficiency in simulation with that in data for four regions of the  $W/Z$  tagger ( $D_2$

fail,  $m_J$  fail;  $D_2$  pass,  $m_J$  fail;  $D_2$  fail,  $m_J$  pass;  $D_2$  pass,  $m_J$  pass). (Additionally, separate scale factors are determined for events with large-R jets from  $W$  bosons and top backgrounds.) A simultaneous template fit is used to fit the signal jets (jets initiated by  $W/Z$  bosons or top quarks) and background jets (all other jets from the simulated backgrounds) to the data in the four regions using the  $m_J$  distributions. The scale factor for a given region is then given by:

$$SF = \frac{\epsilon_{data} = \frac{N_{fitted-signal}^{region}}{N_{fitted-signal}^{all-regions}}}{\epsilon_{MC} = \frac{N_{signal}^{region}}{N_{signal}^{all-regions}}} \quad (11.1)$$

The effects of experimental and theoretical uncertainties on the efficiency scale factor are determined by taking the ratio of efficiencies in data and simulation. By taking this ratio the uncertainties not arising for jet mass and  $D_2$  cancel.

Lepton identification, reconstruction, isolation systematic uncertainties are determined by reconstructing the  $Z$  mass peak with a tag and probe method. The lepton energy and momentum scales are also measured with the  $Z$  mass peak.

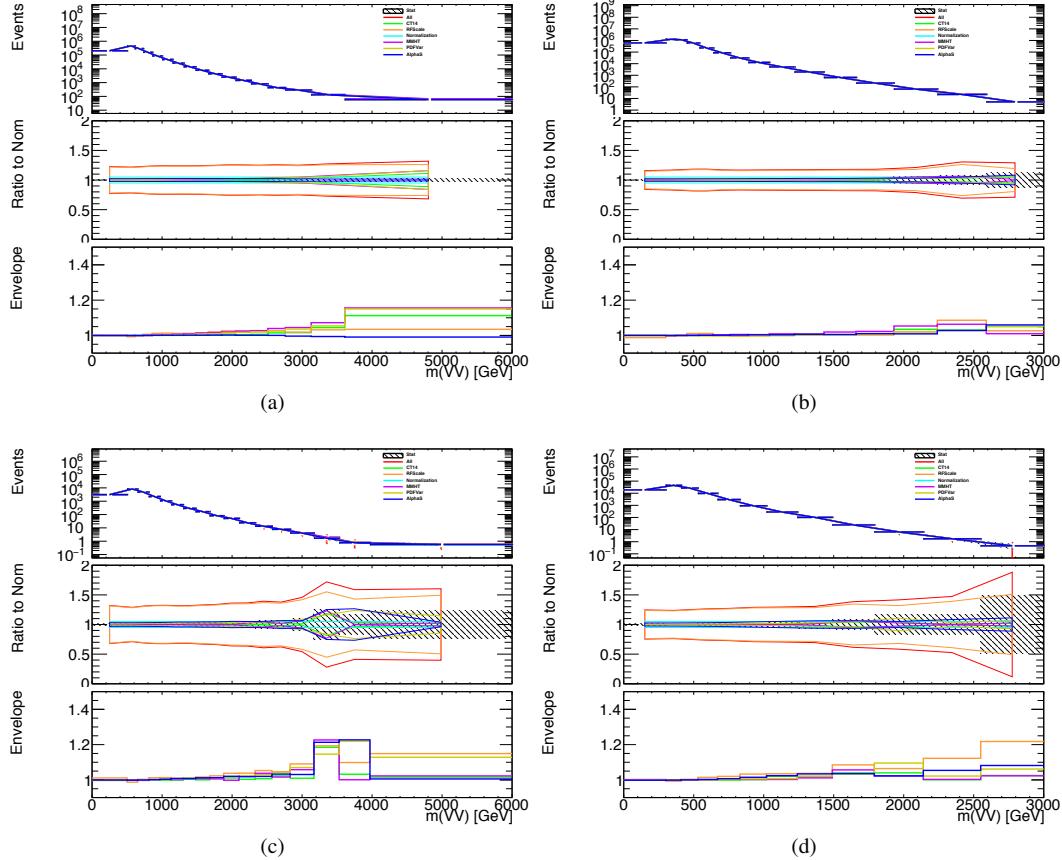
As  $E_T^{miss}$  is calculated using all the physics objects in the event, all those objects associated errors result in an uncertainty on  $E_T^{miss}$ . Additionally, the unassociated tracks used to construct  $E_T^{miss}$  contribute to the uncertainty on  $E_T^{miss}$ .

## 11.2 Theory Systematics

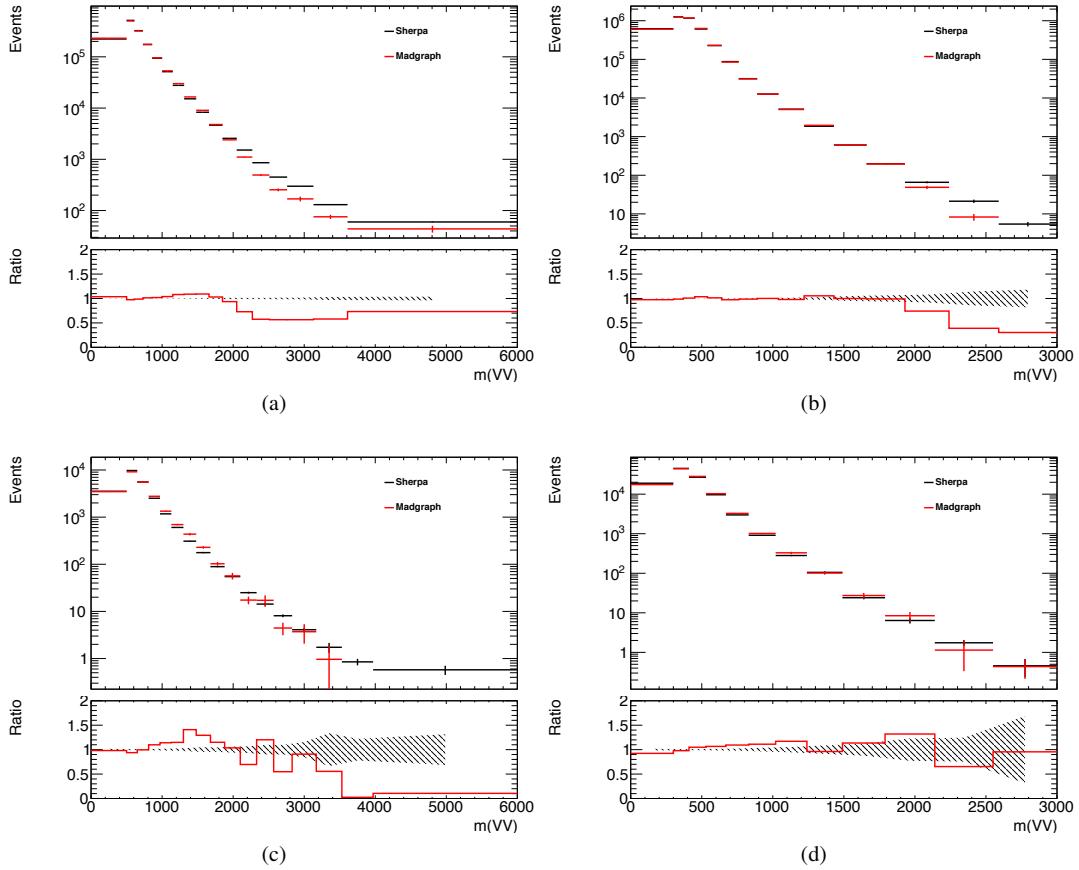
Theoretical uncertainties for signal and background processes arise from uncertainties in the parameters used in Monte Carlo simulation. In particular for the  $t\bar{t}$ ,  $W/Z+jets$ , diboson backgrounds and signal samples, the QCD scale, PDF, generator and hadronization uncertainties were evaluated. To assess the QCD scale uncertainty the renormalization and factorization scales were scaled up and

1192 down by a factor of two at the event generation stage of sample production. Un-  
1193 certainties due to the choice of the parton distribution functions were evaluated by  
1194 re-weighting samples from the nominal PDF to a set of error PDFs which account  
1195 for the uncertainty of the fits used to produce the PDF set. In addition to this,  
1196 samples are re-weighted to different PDF sets to account for the arbitrariness of  
1197 the PDF choice. The difference between the  $m_{WV}$  distributions using different  
1198 event generators is assessed by comparing samples generated with different gen-  
1199 erators. Similarly, the uncertainty in hadronization models is accounted for by  
1200 comparing samples created using different hadronization models (e.g.  $t\bar{t}$  Powheg  
1201 is compared to AMC@NLO,  $W + jets$  compares Sherpa and MadGraph+Pythia  
1202 samples). Figures 11.1 - 11.8 show the impact of these uncertainties on the  $t\bar{t}$  and  
1203  $W/Z + jets$  backgrounds. Additionally, contributions to the diboson background  
1204 for the VBF analysis were found to be small and were accounted for by including  
1205 a 5(10)% systematic in the diboson normalization in the final fit.

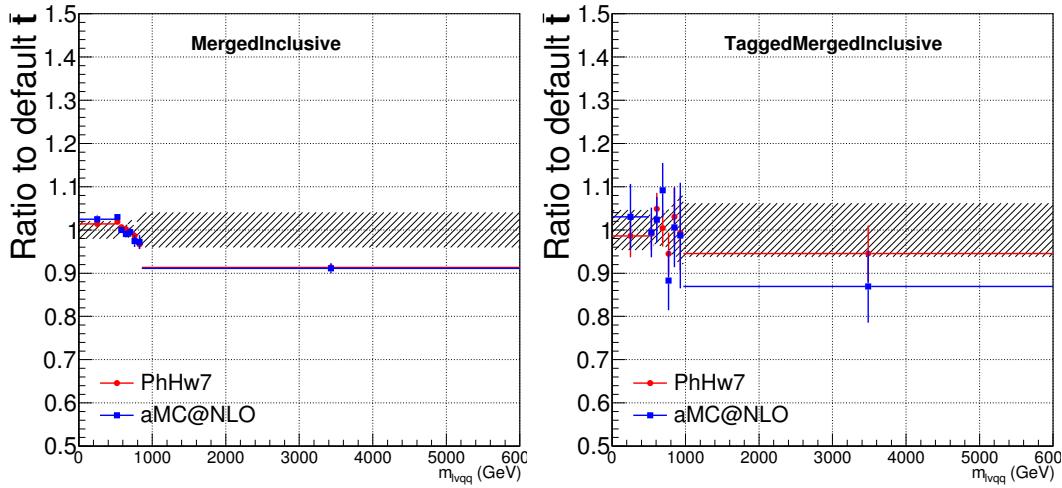
1206 The normalization of the  $t\bar{t}$  and  $W+jets$  processes impact the multijet tem-  
1207 plate shape. The impact of these normalizations was assessed by including a shape  
1208 systematic on the multijet background from varying the  $t\bar{t}$  and  $W+jets$  normal-  
1209 ization factors. The overall normalization of the template is systematic in the  
1210 final likelihood fit (account for other systematic effects on the template).



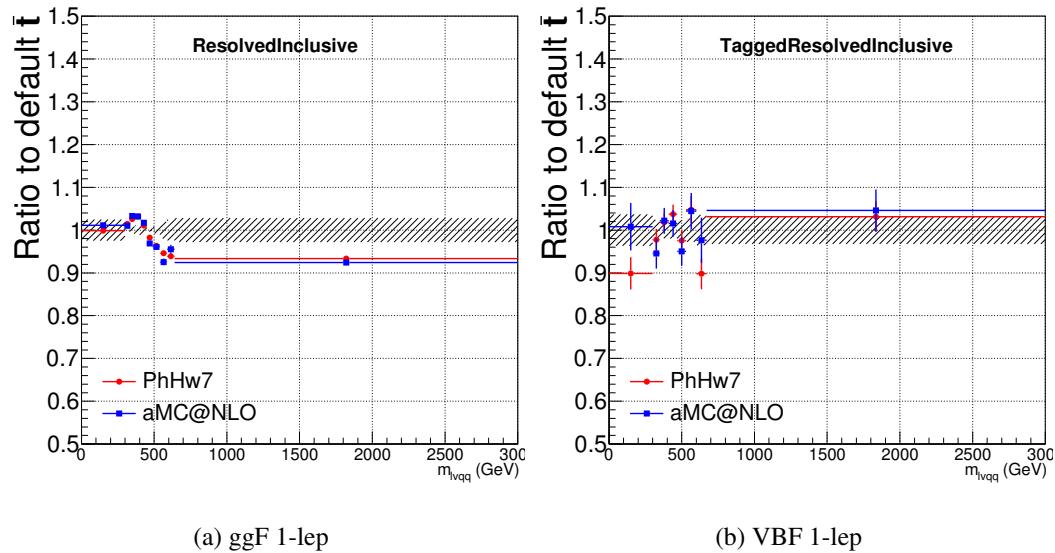
**Figure 11.1:** The  $W/Z + \text{jet}$  systematics for the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The top subplot shows the nominal and variation distributions/bands, the middle shows the ratio of the two, and the final shows just the shape of the envelope (the final uncertainty).



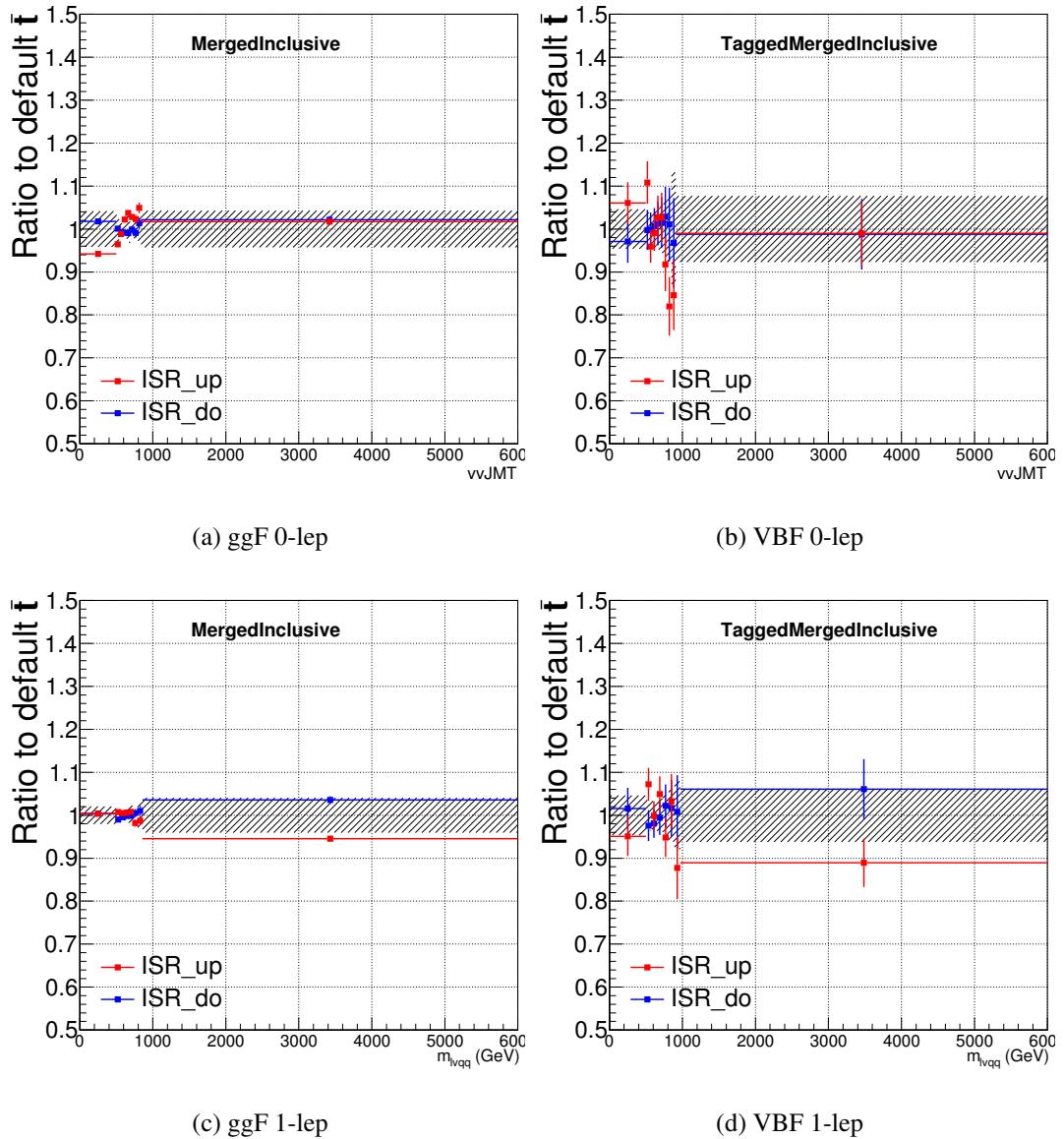
**Figure 11.2:** The two-point generator comparison between Sherpa and MadGraph for the  $W/Z + \text{jet}$  samples in the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The normalization of the Madgraph sample is set to the Sherpa value to consider only shape effects. The bottom inset shows the ratio of the two.



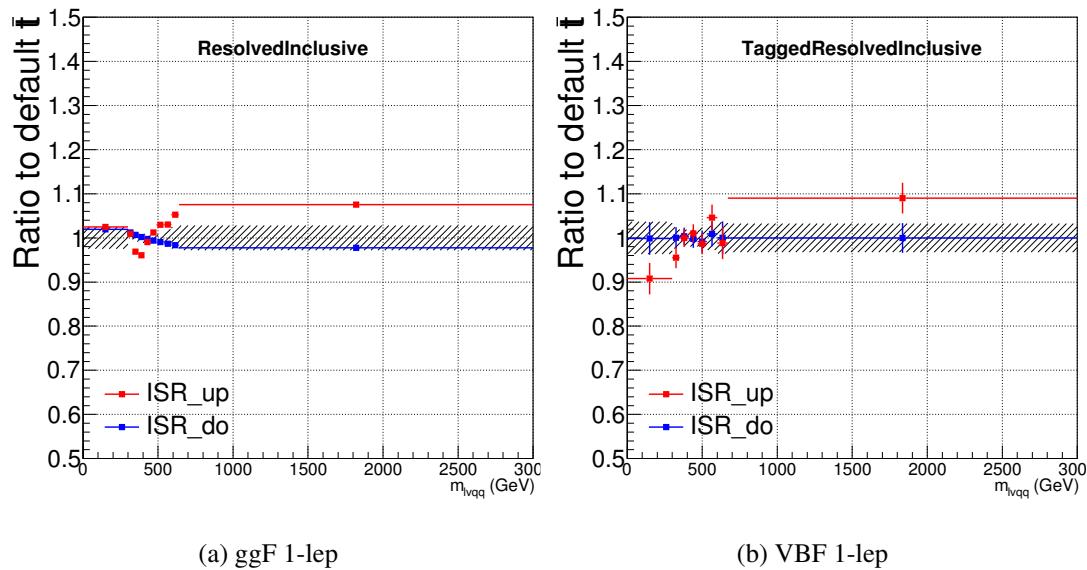
**Figure 11.3:** Ratio between the variations of generator (red) and hadronization (blue) variations for the Merged regime for  $t\bar{t}$  sample.



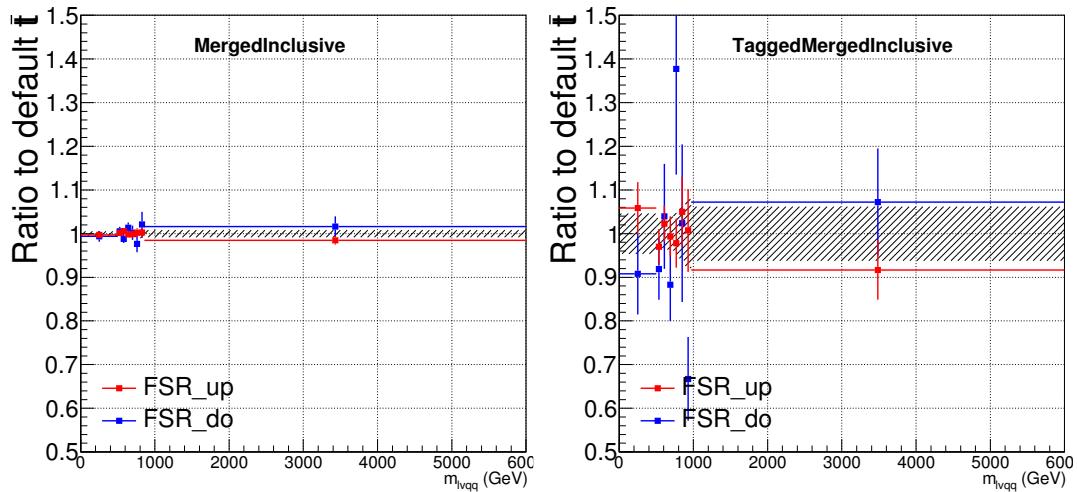
**Figure 11.4:** Ratio between the variations of generator (red) and hadronization (blue) variations for the Resolved regime for  $t\bar{t}$  sample.



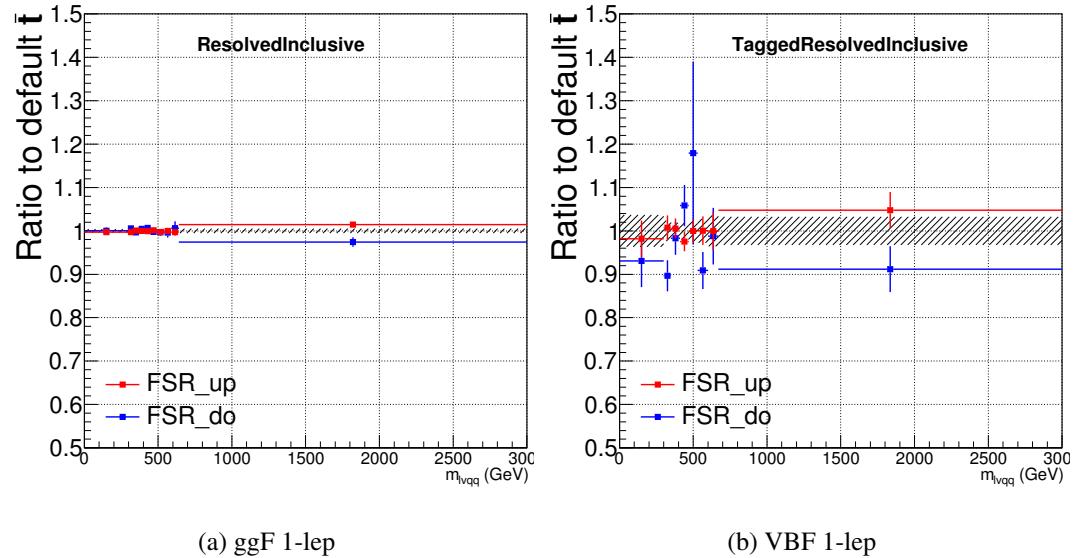
**Figure 11.5:** Ratio between the variations of ISR up (red) and down (blue) variations for the Merged regime for  $t\bar{t}$  sample.



**Figure 11.6:** Ratio between the variations of ISR up (red) and down (blue) variations for the Resolved regime for  $t\bar{t}$  sample.



**Figure 11.7:** Ratio between the variations of FSR up (red) and down (blue) variations for the Merged regime for  $t\bar{t}$  sample.



**Figure 11.8:** Ratio between the variations of FSR up (red) and down (blue) variations for the Resolved regime for  $t\bar{t}$  sample.

<sub>1211</sub> **Chapter 12**

<sub>1212</sub> **Statistical Analysis**

<sub>1213</sub> To determine the compatibility of the data collected with the proposed reso-  
<sub>1214</sub> nances a statistical procedure based on a likelihood function is used. This test  
<sub>1215</sub> compares the distribution of  $m_{WV}$  for the background only hypothesis with the  
<sub>1216</sub> background and signal hypothesis (see Figures 13.5 - 13.9 for  $m_{WV}$  SR distribu-  
<sub>1217</sub> tions). A discovery test is used to measure the compatibility of the observed data  
<sub>1218</sub> with the background only hypothesis. If the observed data are sufficiently incom-  
<sub>1219</sub> patible with the background only hypothesis, this could indicate a discovery. In  
<sub>1220</sub> the absence of discovery, upper limits on the signal strength parameter,  $\mu$ , are as-  
<sub>1221</sub> sessed using the CLs method. For signal masses below 500GeV only the resolved  
<sub>1222</sub> analysis is used, as the merged analysis is not applicable for such small resonance  
<sub>1223</sub> masses. Similarly, it is unlikely that the two jets from the hadronically decaying  
<sub>1224</sub> boson will be well separated for signal masses exceeding 1 TeV. Therefore, only  
<sub>1225</sub> the merged analysis is used above 1TeV. For signal masses between 500 - 1000  
<sub>1226</sub> GeV the merged and resolved analyses are combined for the signal production  
<sub>1227</sub> mode considered.

## 1228 12.1 Likelihood Function Definition

1229 The likelihood function is product of Poisson probabilities over all  $m_{WV}$  bins  
1230 and the associated systematics:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_c \prod_i \frac{(\mu s_{ci}(\boldsymbol{\theta}) + b_{ci}(\boldsymbol{\theta}))^{n_{ci}}}{n_{ci}!} e^{-(\mu s_{ci}(\boldsymbol{\theta}) + b_{ci}(\boldsymbol{\theta}))} \prod_k (\theta'_k | \theta_k) \quad (12.1)$$

1231 Here  $c$  are the analysis channels (e.g. merged SRs and CRs and resolved SRs  
1232 and CRs) considered and  $i$  runs over all the  $m_{WV}$  bins used in the fit. The  
1233 signal strength parameter,  $\mu$ , multiplies the expected signal yield in each anal-  
1234 ysis bin,  $s_{ci}$ . The background content for channel  $c$  and bin  $i$  is given by  $b_{ci}$ .  
1235 The dependence of signal and background predictions on systematic uncertain-  
1236 ties is described by the aforementioned set of nuisance parameters  $\boldsymbol{\theta}$ , which are  
1237 parameterized by Gaussian or log-normal priors denoted here as  $\theta_k$ . Statistical  
1238 uncertainties of the simulated bin contents are also included as systematic un-  
1239 certainties. Most systematics are correlated among all the analysis regions and  
1240 considered to be independent from each other. The validity of this assumption is  
1241 checked by evaluating the covariance of nuisance parameters.

## 1242 12.2 Fit Configuration

1243 The binning of  $m_{WV}$  in signal regions for the likelihood fit is determined by the  
1244 statistical uncertainty of signal mass width. For each signal mass point, the signal  
1245 mass resolution is given by the fitted Gaussian width in simulate of  $m_{WV}$ . The  
1246 fitted signal widths are then fit to a line to give a parameterized signal mass width,  
1247 as shown in Figures 12.1 and 12.2. Bin widths are set first to this parameterized  
1248 signal mass resolution. Then if the statistical uncertainty of the data or simulated

1249 background is more than 50%, bins are merged until the statistical uncertainty is  
1250 less than 50%. All control regions contain only a single bin.

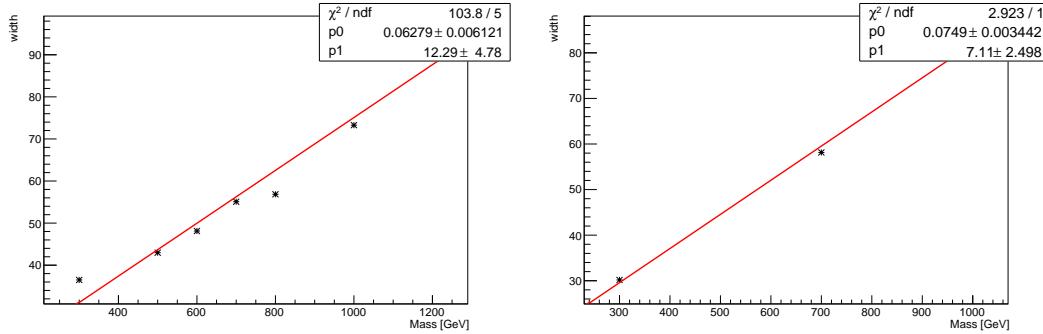
1251 For this analysis, each signal model is fit in the Merged and Resolved chan-  
1252 nels for the relevant signal production mode simultaneously. The  $W + \text{jets}$  and  $t\bar{t}$   
1253 normalizations are given by the best fit values in the overall fit and these fitted  
1254 normalizations are then applied to those backgrounds in the SRs, as mentioned  
1255 previously.

1256 The  $m_{VV}$  distributions for a given systematic may contain unphysically large  
1257 fluctuations due to  $m_{VV}$  bins with few events. This can lead to artificial pulls and  
1258 constraints in the fit. To remove such issues a multi-step smoothing procedure is  
1259 applied to all systematic variation distributions. First, distributions are rebinned  
1260 until the statistical error per bin is at least 5%. Next all local extrema are identi-  
1261 fied. The bins around smallest extrema are iteratively merged until only four local  
1262 extrema remain. Then distributions are rebinned so that statistical uncertainties  
1263 in each bin are  $< 5\%$ .

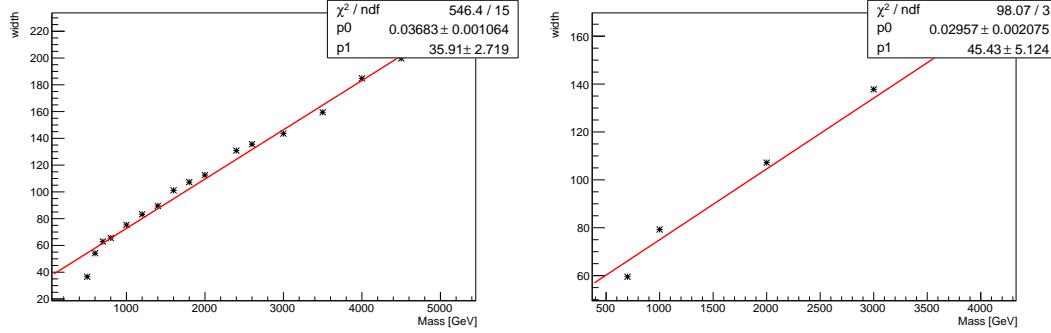
1264 For some systematics, up and down variations may be in the same direction  
1265 with respect to the nominal distributions. This causes the variations to not cover  
1266 the nominal choice, and the interpretation of the confidence interval is skewed as  
1267 the nominal distribution should be bracketed by the up and down variations. This  
1268 asymmetry may also lead to unconstrained systematics in the fit. To handle such  
1269 asymmetric systematics, if the up and down variation for a given systematic are in  
1270 the same direction for at least three  $m_{VV}$  bins the variation is averaged for those  
1271 bins. The averaging procedure replaces bin-by-bin the up and down variation bins  
1272 by  $b_{\pm}^{new} = b_{nom} \pm \frac{|b_+ - b_-|}{2}$ , where  $b_{nom}$  is the nominal bin content and  $b_{\pm}$  are the  
1273 original up and down variation bin content. The same procedure is also applied to  
1274 any variations where the integral of the difference between the up/down variation

1275 and the nominal distribution is twice that of the other down/up variation, further  
 1276 ensuring variations are symmetric around the nominal distribution.

1277 Finally, systematics that have a negligible effect on the  $m_{WW}$  distribution are  
 1278 not considered in the fit. Shape systematics where no bin in the variational dis-  
 1279 tribution deviates more than 1% from the nominal distribution (after normalizing  
 1280 all histograms to the nominal) are not included in the fit. Also, statistical bin  
 1281 uncertainties < 1% are ignored.



**Figure 12.1:** The HVT signal mass resolution as a function of mass fit with a straight line in the Resolved ggF region (left) and VBF (right) region.



**Figure 12.2:** The HVT signal mass resolution as a function of mass fit with a straight line in the Merged ggF region (left) and VBF (right) region.

1282 **12.3 Best Fit  $\mu$**

1283 The best fit signal strength parameter is denoted by  $\hat{\mu}$  and calculated by  
1284 maximizing the likelihood function over the entire  $m_{WV}$  distribution with respect  
1285 to all systematics and  $\mu$ . The corresponding set of systematics that maximize  
1286 the likelihood are given by  $\hat{\boldsymbol{\theta}}$ . The first term in the likelihood is maximized  
1287 when the expected number of signal and background events over all  $m_{WV}$  bins is  
1288 equal to the number of events in data ( $n_{ci} = \mu s_{ci} + b_{ci}$ ) . Thus, by maximizing  
1289 the likelihood, the fit determines value of  $\mu$  and  $\boldsymbol{\theta}$  that give the best agreement  
1290 between expected and measured event yields. The second term in the likelihood  
1291 is a penalty term which decreases the likelihood when systematics are shifted  
1292 from their nominal values. This prevents the fit from profiling (unexpectedly)  
1293 constraining or shifting the fitted value of a systematic far from its nominal value)  
1294 systematics in unphysical ways to maximize the likelihood. The uncertainty on  
1295  $\mu$  is calculated by varying  $\mu$  up and down until the natural log of the likelihood  
1296 function shifts by one-half.

1297 **12.4 Discovery Test**

1298 To determine if the observed dataset is consistent with tested signal model a  
1299 likelihood ratio is constructed:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \quad (12.2)$$

1300 The denominator in this equation is the maximized value of  $\mathcal{L}$  over all system-  
1301 atics and  $\mu$ . The numerator is the maximized likelihood over all systematics for  
1302 a given  $\mu$  value, where the maximized systematics are given by  $\hat{\boldsymbol{\theta}}_\mu$ . To test for  
1303 the existence of signal the observed dataset the null hypothesis ( $H_0$ ) is defined as

1304 the background only hypothesis and the alternate hypothesis includes signal and  
 1305 background ( $H_1$ ). This test quantifies the compatibility of observed data with  
 1306  $H_0$  by calculating a p-value representing the probability of observing data as dis-  
 1307 crepant or more than the observed data under the  $H_0$ . The test statistic used to  
 1308 calculate this p-value is given by ( $r_0$ ):

$$r_0 = \begin{cases} -2 \ln \lambda(0), \hat{\mu} > 0 \\ +2 \ln \lambda(0), \hat{\mu} < 0 \end{cases} \quad (12.3)$$

1309 The expected distribution of the the test statistic under  $H_0$  ( $f(r_0|0)$ ) is used  
 1310 to calculate the p-value:

$$p_0 = \int_{r_0,obs}^{\infty} f(r_0|0) dr_0 \quad (12.4)$$

1311 Small p-values indicate the observed data is poorly described by  $H_0$ . This  
 1312 equivalent Z-score of a given p-value is usually used to further quantify the agree-  
 1313 ment between the observed data and  $H_0$ . The Z-score is given by the number of  
 1314 standard deviations away from the mean of a Gaussian distribution, the integral  
 1315 of the upper tail of the distribution would equal the p-value. Mathematically:

$$Z = \Phi^{-1}(1 - p_0) \quad (12.5)$$

1316 where  $\Phi$  is the Gaussian cumulative distribution function. The statistical  
 1317 significance of these tests are expressed as the  $Z$ -score. In particle physics,  $3\sigma$  is  
 1318 considered evidence for new phenomena and  $5\sigma$  is the threshold for discovery.

## 1319 12.5 Exclusion Limits

1320 In the absence of discovery, upper limits on the signal strength,  $\mu$  are set using  
 1321 the CLs method [cite P60]. The test statistic for this test,  $q_\mu$ , is constructed as:

$$\tilde{\lambda}_\mu = \begin{cases} \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \hat{\mu} > 0 \\ \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(0, \hat{\theta}_0)}, \hat{\mu} < 0 \end{cases} \quad (12.6)$$

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu), \hat{\mu} < \mu \\ +2 \ln \tilde{\lambda}(\mu), \hat{\mu} > \mu \end{cases} \quad (12.7)$$

1322 As defined, larger values of  $q_\mu$  correspond to increasing incompatibility between  
 1323 the observed data and the background + signal hypothesis. The observed value  
 1324 of the test statistic,  $q_{\mu, obs}$ , is then compared to its expected distribution,  $f$ , to  
 1325 calculate p-values to assess the likelihood of the background+signal hypothesis.  
 1326 Using these distributions,  $CL_s$  values are computed as:

$$CL_{s+b} = \int_{q_{\mu, obs}}^{\infty} f(q_\mu | \mu) dq_\mu \quad (12.8)$$

$$CL_b = \int_{q_0^{obs}}^{\infty} f(q_\mu | \mu = 0) dq_\mu \quad (12.9)$$

$$CL_s = \frac{CL_{s+b}}{CL_b} \quad (12.10)$$

1329  $CL_{s+b}$  is the p-value for the signal + background hypothesis and  $CL_b$  is the  
 1330 p-value for the background only hypothesis. The  $CL_s$  value is interpreted as  
 1331 the probability to observe the background + signal hypothesis normalized to the  
 1332 probability of background-only hypothesis. Normalizing by  $CL_b$  prevents setting  
 1333 artificially strong exclusion limits due to downward fluctuations in data.

1334 For a given signal hypothesis,  $\mu$  values are scanned simultaneously over all

<sub>1335</sub>  $m_{WV}$  bins to find the  $\mu$  value that yields  $CL_s=0.05$ , meaning the likelihood of  
<sub>1336</sub> finding data more incompatible with the signal+background hypothesis (relative  
<sub>1337</sub> to the background only hypothesis) is 5%. The 95% upper limit on the cross  
<sub>1338</sub> section is then calculated as the product of the  $\mu$  value found, branching ratio,  
<sub>1339</sub> and theory cross section.

1340 **Chapter 13**

1341 **Results**

1342 **13.1 Expected and Measured Yields**

1343 The yield tables for the four analysis regions are shown in Tables ?? - ??.

1344 The fitted background normalizations are shown in Tables ??-??. The control

1345 region  $m_{\ell\nu qq}$  distributions are shown in Figures 13.1 - 13.4. The signal region

1346  $m_{\ell\nu qq}$  distributions are shown in Figures 13.5 - ??.

	HP WCR	LP WCR	Resolved WCR
Electron Multi-jet	-	-	$16507.83 \pm 2314.87$
Muon Multi-jet	-	-	$19977.12 \pm 2816.06$
Diboson	$1833.41 \pm 177.78$	$3323.93 \pm 320.92$	$9147.67 \pm 961.63$
Single-top	$2160.62 \pm 402.34$	$3551.09 \pm 660.00$	$20058.36 \pm 3817.26$
$t\bar{t}$	$15518.86 \pm 338.22$	$24069.54 \pm 453.15$	$138866.23 \pm 1989.71$
$W+jets$	$40141.57 \pm 357.79$	$88113.06 \pm 487.87$	$673200.38 \pm 4120.53$
$Z+jets$	$778.83 \pm 78.93$	$1765.54 \pm 179.10$	$16570.50 \pm 1672.71$
Total	$60433.29 \pm 664.92$	$120823.16 \pm 1006.99$	$894328.12 \pm 7247.12$
Data	60264.00	120852.00	895362.00
	HP TCR	LP TCR	Resolved TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$421.11 \pm 37.98$	$550.44 \pm 53.10$	$996.87 \pm 119.63$
Single-top	$4691.44 \pm 846.11$	$3466.26 \pm 631.03$	$16848.71 \pm 3258.26$
$t\bar{t}$	$38945.18 \pm 848.77$	$33836.95 \pm 637.04$	$224226.14 \pm 3212.76$
$W+jets$	$2258.34 \pm 20.13$	$6564.78 \pm 36.35$	$23466.41 \pm 143.63$
$Z+jets$	$66.35 \pm 6.72$	$213.26 \pm 21.63$	$846.66 \pm 85.47$
Total	$46382.43 \pm 1199.25$	$44631.70 \pm 899.23$	$266384.78 \pm 4580.43$
Data	46354.00	44629.00	266443.00
	WW SR	LP SR	Resolved 1-lepton SR
Electron Multi-jet	-	-	$10788.40 \pm 1512.85$
Muon Multi-jet	-	-	$15759.50 \pm 2221.53$
Diboson	$4990.30 \pm 376.50$	$3901.07 \pm 313.22$	$16971.29 \pm 1523.77$
Single-top	$3117.71 \pm 565.07$	$2176.46 \pm 400.52$	$20422.85 \pm 3731.94$
$t\bar{t}$	$13785.77 \pm 302.14$	$11005.12 \pm 207.41$	$126965.25 \pm 1819.66$
$W+jets$	$24718.56 \pm 223.72$	$60080.66 \pm 333.12$	$444133.56 \pm 2719.02$
$Z+jets$	$478.18 \pm 48.46$	$1226.69 \pm 124.44$	$11686.32 \pm 1179.69$
Total	$47090.52 \pm 777.65$	$78389.98 \pm 654.22$	$646727.19 \pm 5963.98$
Data	47330.00	78380.00	645610.00

**Table 13.1:** Expected and Measured for DY  $WW$   $W+jets$ ,  $t\bar{t}$  control regions and signal regions.

	HP Untagged WCR	LP Untagged WCR	Resolved Untagged WCR
Electron Multi-jet	-	-	$15080.03 \pm 2277.99$
Muon Multi-jet	-	-	$27347.10 \pm 2950.07$
Diboson	$1508.48 \pm 154.20$	$2758.24 \pm 284.50$	$9038.55 \pm 728.69$
Single-top	$1756.59 \pm 306.69$	$2913.18 \pm 515.93$	$20511.74 \pm 3523.47$
$t\bar{t}$	$13134.00 \pm 238.30$	$21815.37 \pm 334.98$	$140157.77 \pm 2636.96$
$W+jets$	$40654.84 \pm 333.65$	$87657.76 \pm 501.96$	$665909.12 \pm 4420.62$
$Z+jets$	$768.72 \pm 77.97$	$1759.87 \pm 178.96$	$16512.46 \pm 1673.23$
Total	$57822.63 \pm 540.40$	$116904.42 \pm 862.16$	$894556.75 \pm 7492.20$
Data	57699.00	117306.00	895362.00
	HP Tagged WCR	LP Tagged WCR	Resolved Tagged WCR
Electron Multi-jet	-	-	$384.58 \pm 57.11$
Muon Multi-jet	-	-	$602.93 \pm 190.12$
Diboson	$30.22 \pm 4.69$	$48.95 \pm 7.16$	$264.64 \pm 28.24$
Single-top	$308.44 \pm 56.19$	$371.59 \pm 69.43$	$5752.39 \pm 1029.97$
$t\bar{t}$	$1683.82 \pm 48.73$	$2041.48 \pm 70.00$	$58431.49 \pm 614.30$
$W+jets$	$583.55 \pm 75.37$	$1109.45 \pm 85.78$	$11891.68 \pm 903.01$
$Z+jets$	$13.19 \pm 1.34$	$23.06 \pm 2.34$	$324.74 \pm 32.85$
Total	$2619.22 \pm 106.00$	$3594.53 \pm 130.90$	$77652.45 \pm 1514.89$
Data	2565.00	3546.00	77973.00
	HP Untagged TCR	LP Untagged TCR	Resolved Untagged TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$289.45 \pm 28.45$	$346.78 \pm 35.85$	$650.85 \pm 65.56$
Single-top	$3107.99 \pm 538.03$	$2250.64 \pm 385.41$	$9606.87 \pm 1698.22$
$t\bar{t}$	$30992.40 \pm 562.33$	$26954.21 \pm 413.89$	$91893.59 \pm 1728.91$
$W+jets$	$2236.29 \pm 18.35$	$4874.03 \pm 27.91$	$16122.97 \pm 107.03$
$Z+jets$	$71.54 \pm 7.26$	$155.50 \pm 15.81$	$577.71 \pm 58.54$
Total	$36697.66 \pm 779.03$	$34581.16 \pm 567.59$	$118851.98 \pm 2427.40$
Data	36677.00	34573.00	118928.00
	HP Tagged TCR	LP Tagged TCR	Resolved Tagged TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$9.72 \pm 1.13$	$8.75 \pm 1.16$	$34.06 \pm 4.98$
Single-top	$105.87 \pm 20.65$	$119.66 \pm 22.68$	$656.89 \pm 132.96$
$t\bar{t}$	$1904.75 \pm 50.61$	$1483.86 \pm 47.05$	$17965.33 \pm 188.87$
$W+jets$	$32.36 \pm 4.28$	$85.74 \pm 6.96$	$489.01 \pm 37.13$
$Z+jets$	$1.27 \pm 0.13$	$1.93 \pm 0.20$	$19.14 \pm 1.94$
Total	$2053.98 \pm 54.84$	$1699.93 \pm 52.70$	$19164.43 \pm 234.01$
Data	2047.00	1708.00	19143.00

**Table 13.2:** Expected and Measured for DY  $WZ$   $W+jets$ ,  $t\bar{t}$  tag and untag control regions.

	HP Untagged SR	LP Untagged SR	Resolved Untagged SR
Electron Multi-jet	-	-	$7782.17 \pm 1175.56$
Muon Multi-jet	-	-	$17004.81 \pm 1834.40$
Diboson	$3041.17 \pm 273.77$	$2266.35 \pm 212.79$	$14724.12 \pm 1224.31$
Single-top	$2123.28 \pm 373.83$	$1379.35 \pm 240.92$	$18336.88 \pm 3082.47$
$t\bar{t}$	$11678.86 \pm 213.63$	$8906.34 \pm 136.88$	$112669.24 \pm 2122.46$
$W+jets$	$22741.32 \pm 191.47$	$41726.76 \pm 240.56$	$342934.00 \pm 2280.21$
$Z+jets$	$442.03 \pm 44.84$	$849.79 \pm 86.42$	$9271.83 \pm 939.52$
Total	$40026.65 \pm 546.81$	$55128.59 \pm 432.90$	$522723.03 \pm 5131.71$
Data	40193.00	54735.00	521813.00
	HP Tagged SR	LP Tagged SR	Resolved Tagged SR
Electron Multi-jet	-	-	$199.22 \pm 29.58$
Muon Multi-jet	-	-	$393.43 \pm 124.06$
Diboson	$102.58 \pm 11.59$	$65.44 \pm 8.05$	$624.07 \pm 58.10$
Single-top	$178.21 \pm 33.62$	$155.53 \pm 28.95$	$3470.39 \pm 617.48$
$t\bar{t}$	$1017.93 \pm 31.95$	$706.76 \pm 26.20$	$38189.30 \pm 401.91$
$W+jets$	$325.58 \pm 41.62$	$575.36 \pm 43.29$	$6161.96 \pm 467.71$
$Z+jets$	$7.81 \pm 0.80$	$11.62 \pm 1.19$	$183.36 \pm 18.55$
Total	$1632.11 \pm 63.39$	$1514.70 \pm 58.86$	$49221.74 \pm 884.06$
Data	1699.00	1559.00	48919.00

**Table 13.3:** Expected and Measured for DY  $WZ$   $W+jets$ ,  $t\bar{t}$  tag and untag signal regions.

	HP WCR	LP WCR	Resolved WCR
Electron Multi-jet	-	-	$898.48 \pm 137.82$
Muon Multi-jet	-	-	$601.46 \pm 182.74$
Diboson	$107.45 \pm 45.20$	$166.87 \pm 68.11$	$292.10 \pm 235.29$
Single-top	$78.19 \pm 18.22$	$132.71 \pm 31.93$	$879.82 \pm 216.89$
$t\bar{t}$	$400.71 \pm 28.35$	$569.70 \pm 48.88$	$5067.51 \pm 155.69$
$W+jets$	$864.49 \pm 63.44$	$1940.80 \pm 89.41$	$18563.70 \pm 408.99$
$Z+jets$	$19.51 \pm 2.00$	$46.63 \pm 4.77$	$795.20 \pm 80.89$
Total	$1470.35 \pm 84.89$	$2856.71 \pm 126.74$	$27098.28 \pm 594.01$
Data	1495.00	2898.00	27120.00
	HP TCR	LP TCR	Resolved TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$14.95 \pm 6.61$	$27.57 \pm 14.12$	$24.33 \pm 20.32$
Single-top	$68.31 \pm 16.17$	$58.93 \pm 13.56$	$278.60 \pm 73.04$
$t\bar{t}$	$496.60 \pm 31.72$	$401.23 \pm 32.13$	$3834.49 \pm 104.60$
$W+jets$	$50.68 \pm 4.19$	$144.02 \pm 7.86$	$450.01 \pm 11.87$
$Z+jets$	$1.32 \pm 0.14$	$5.35 \pm 0.55$	$29.96 \pm 3.07$
Total	$631.87 \pm 36.45$	$637.10 \pm 38.44$	$4617.39 \pm 129.77$
Data	636.00	634.00	4615.00
	HP SR	LP SR	Resolved SR
Electron Multi-jet	-	-	$596.34 \pm 91.52$
Muon Multi-jet	-	-	$481.01 \pm 144.48$
Diboson	$148.84 \pm 48.64$	$181.42 \pm 67.30$	$395.52 \pm 318.06$
Single-top	$79.49 \pm 19.80$	$56.82 \pm 14.89$	$782.07 \pm 190.79$
$t\bar{t}$	$338.42 \pm 24.14$	$236.80 \pm 20.88$	$4261.70 \pm 138.98$
$W+jets$	$501.13 \pm 39.36$	$1347.76 \pm 64.50$	$11445.73 \pm 291.49$
$Z+jets$	$9.25 \pm 0.95$	$28.77 \pm 2.95$	$567.66 \pm 57.94$
Total	$1077.13 \pm 69.93$	$1851.57 \pm 96.73$	$18530.03 \pm 523.88$
Data	1096.00	1846.00	18530.00

**Table 13.4:** Expected and Measured for VBF  $WW$   $W+jets$ ,  $t\bar{t}$  control regions and signal regions.

	HP WCR	LP WCR	Resolved WCR
Electron Multi-jet	-	-	$870.00 \pm 132.75$
Muon Multi-jet	-	-	$618.45 \pm 196.90$
Diboson	$92.92 \pm 41.77$	$145.90 \pm 64.26$	$228.62 \pm 114.62$
Single-top	$71.13 \pm 16.29$	$118.82 \pm 27.98$	$1209.87 \pm 281.64$
$t\bar{t}$	$427.80 \pm 29.72$	$509.19 \pm 46.57$	$6860.87 \pm 254.83$
$W+jets$	$871.68 \pm 64.22$	$2020.67 \pm 93.54$	$19088.50 \pm 442.10$
$Z+jets$	$19.58 \pm 2.01$	$47.39 \pm 4.85$	$800.19 \pm 82.02$
Total	$1483.11 \pm 83.79$	$2841.97 \pm 125.92$	$29676.50 \pm 644.96$
Data	1495.00	2898.00	29755.00
	HP TCR	LP TCR	Resolved TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$10.12 \pm 4.51$	$12.73 \pm 6.55$	$14.23 \pm 7.49$
Single-top	$51.57 \pm 12.31$	$35.07 \pm 8.17$	$169.21 \pm 44.54$
$t\bar{t}$	$470.06 \pm 28.97$	$298.99 \pm 25.28$	$2414.75 \pm 75.42$
$W+jets$	$49.64 \pm 4.17$	$109.69 \pm 6.16$	$378.22 \pm 12.05$
$Z+jets$	$1.28 \pm 0.13$	$4.81 \pm 0.50$	$17.62 \pm 1.83$
Total	$582.67 \pm 32.07$	$461.30 \pm 28.05$	$2994.03 \pm 88.75$
Data	584.00	459.00	3001.00
	HP SR	LP SR	Resolved SR
Electron Multi-jet	-	-	$444.65 \pm 67.99$
Muon Multi-jet	-	-	$397.29 \pm 125.59$
Diboson	$109.66 \pm 44.13$	$112.28 \pm 46.45$	$265.75 \pm 139.43$
Single-top	$63.16 \pm 15.20$	$48.02 \pm 11.56$	$872.16 \pm 205.00$
$t\bar{t}$	$348.95 \pm 24.34$	$190.68 \pm 17.75$	$5134.25 \pm 193.57$
$W+jets$	$467.21 \pm 37.12$	$973.73 \pm 47.91$	$10226.83 \pm 254.67$
$Z+jets$	$8.15 \pm 0.84$	$23.62 \pm 2.43$	$558.48 \pm 57.25$
Total	$997.13 \pm 64.42$	$1348.33 \pm 70.06$	$17899.41 \pm 432.98$
Data	1018.00	1313.00	17826.00

**Table 13.5:** Expected and Measured for VBF  $WZ$   $W+jets$ ,  $t\bar{t}$  control regions and signal regions.

Background	Fitted Normalization
XS_Top_LP_lvqq_Merg_binned	$0.905^{+0.0166}_{-0.0166}$
XS_Top_Merg	$0.936^{+0.0199}_{-0.0199}$
XS_Top_Res	$0.957^{+0.0134}_{-0.0134}$
XS_Wjets_LP_lvqq_Merg_binned	$0.884^{+0.00489}_{-0.00489}$
XS_Wjets_Merg	$0.931^{+0.00831}_{-0.00831}$
XS_Wjets_Res	$1.03^{+0.00628}_{-0.00628}$

**Table 13.6:** Fitted background normalizations for  $t\bar{t}$  and  $W+jets$  backgrounds for the DY  $WW$  analysis region.

Background	Fitted Normalization
XS_Top_LP_Tag_lvqq_Merg_binned	$0.973^{+0.0333}_{-0.0333}$
XS_Top_LP_lvqq_Merg_binned	$0.894^{+0.0135}_{-0.0135}$
XS_Top_Merg	$0.893^{+0.016}_{-0.016}$
XS_Top_Res	$0.965^{+0.0179}_{-0.0179}$
XS_Top_Tag_lvqq_Merg_binned	$0.954^{+0.0276}_{-0.0276}$
XS_Top_Tag_lvqq_Res_binned	$0.999^{+0.0105}_{-0.0105}$
XS_Wjets_LP_Tag_lvqq_Merg_binned	$0.912^{+0.0703}_{-0.0703}$
XS_Wjets_LP_lvqq_Merg_binned	$0.876^{+0.00502}_{-0.00502}$
XS_Wjets_Merg	$0.948^{+0.00779}_{-0.00779}$
XS_Wjets_Res	$1.01^{+0.00673}_{-0.00673}$
XS_Wjets_Tag_lvqq_Merg_binned	$0.906^{+0.117}_{-0.117}$
XS_Wjets_Tag_lvqq_Res_binned	$1.2^{+0.0904}_{-0.0904}$

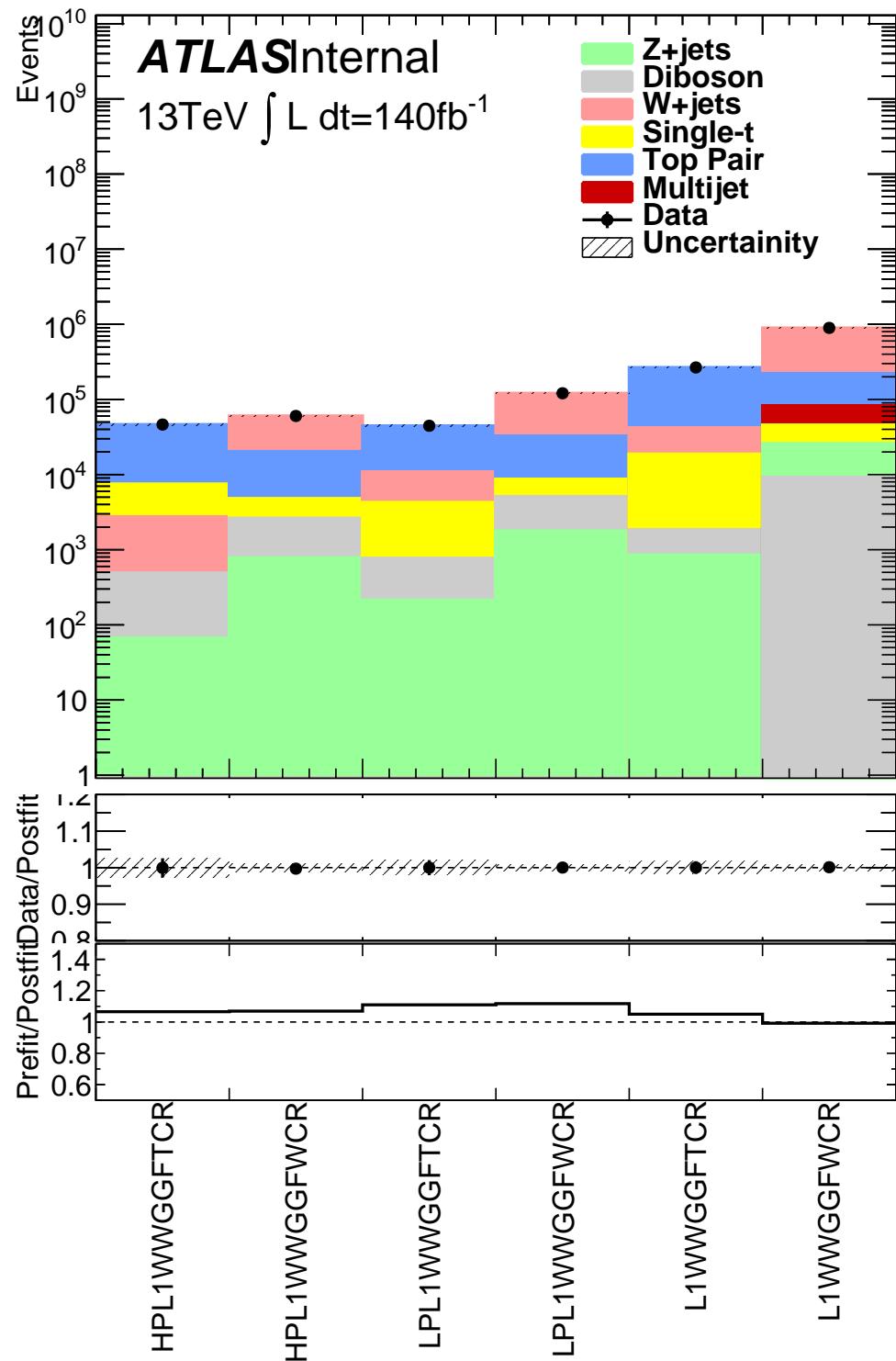
**Table 13.7:** Fitted background normalizations for  $t\bar{t}$  and  $W+jets$  backgrounds for the DY  $WZ$  analysis region.

Background	Fitted Normalization
XS_Top_LP_lvqq_Merg_binned	$0.79^{+0.0673}_{-0.0673}$
XS_Top_Merg	$0.888^{+0.061}_{-0.061}$
XS_Top_Res	$1.01^{+0.0311}_{-0.0311}$
XS_Wjets_LP_lvqq_Merg_binned	$0.88^{+0.0423}_{-0.0423}$
XS_Wjets_Merg	$0.881^{+0.0677}_{-0.0677}$
XS_Wjets_Res	$0.932^{+0.0202}_{-0.0202}$

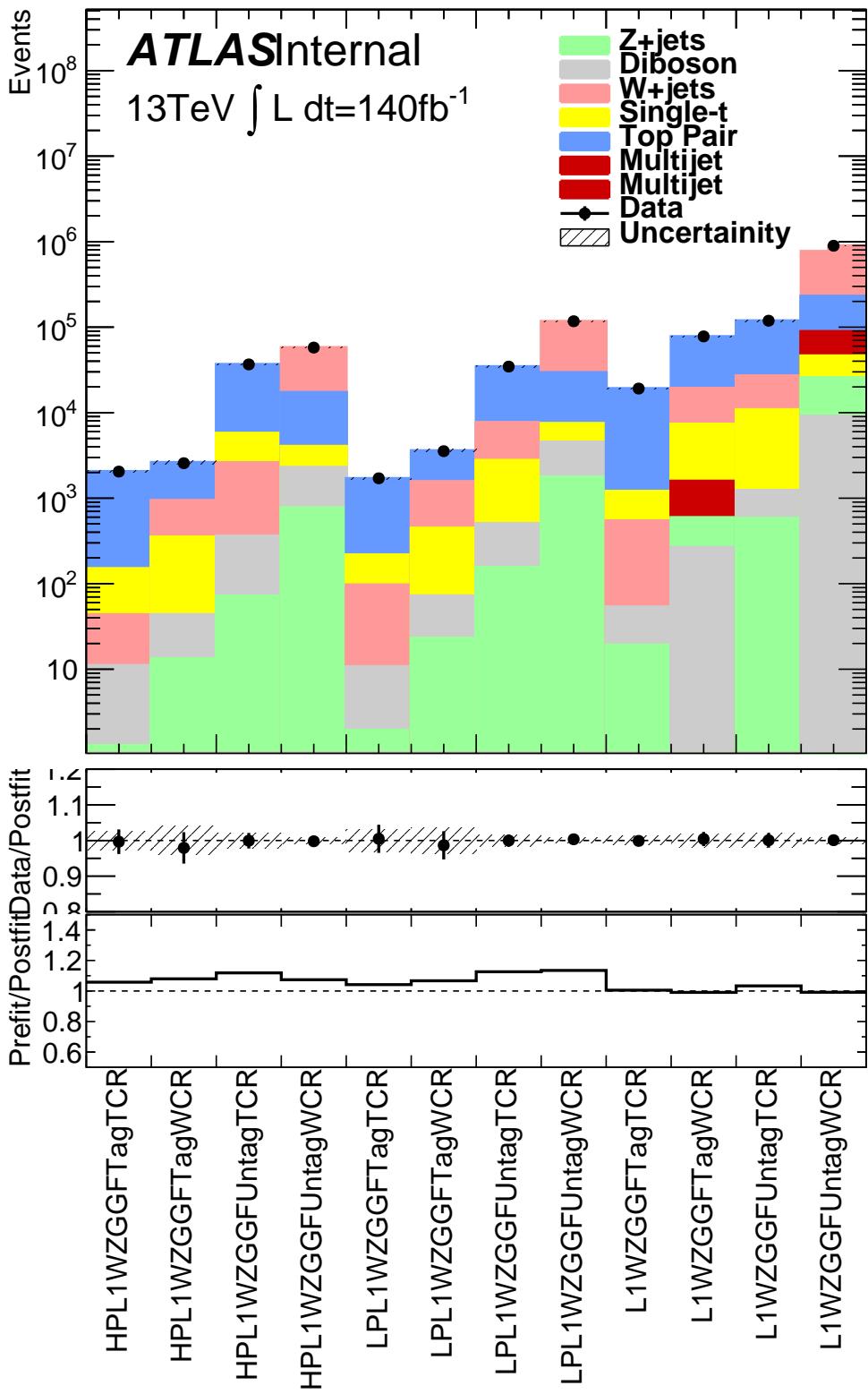
**Table 13.8:** Fitted background normalizations for  $t\bar{t}$  and  $W+jets$  backgrounds for the VBF  $WW$  analysis region.

Background	Fitted Normalization
XS_Top_LP_lvqq_Merg_binned	$0.708^{+0.064}_{-0.064}$
XS_Top_Merg	$0.958^{+0.0644}_{-0.0644}$
XS_Top_Res	$1.02^{+0.038}_{-0.038}$
XS_Wjets_LP_lvqq_Merg_binned	$0.9^{+0.0438}_{-0.0438}$
XS_Wjets_Merg	$0.883^{+0.0685}_{-0.0685}$
XS_Wjets_Res	$0.945^{+0.0219}_{-0.0219}$

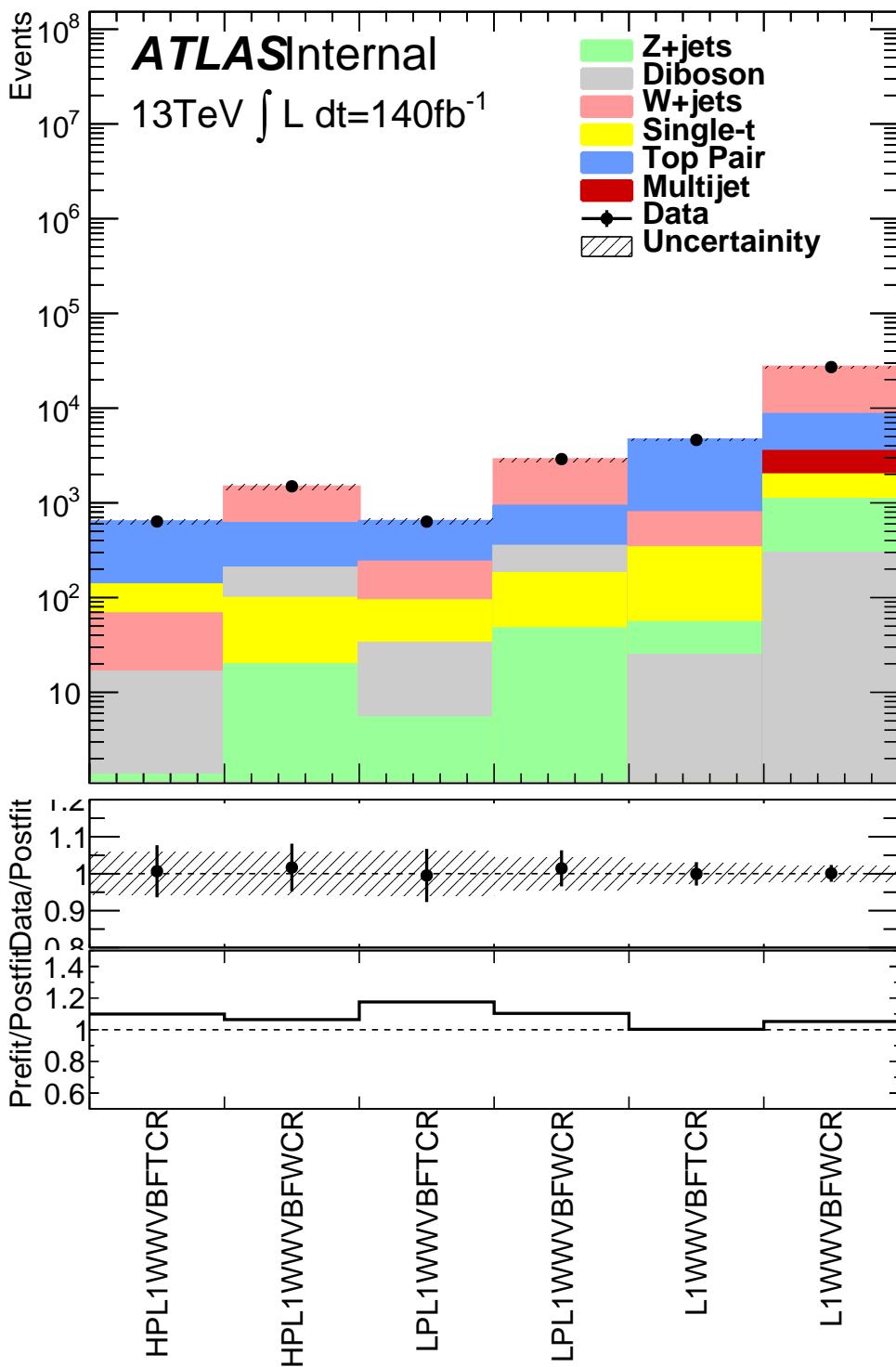
**Table 13.9:** Fitted background normalizations for  $t\bar{t}$  and  $W+jets$  backgrounds for the VBF  $WZ$  analysis region.



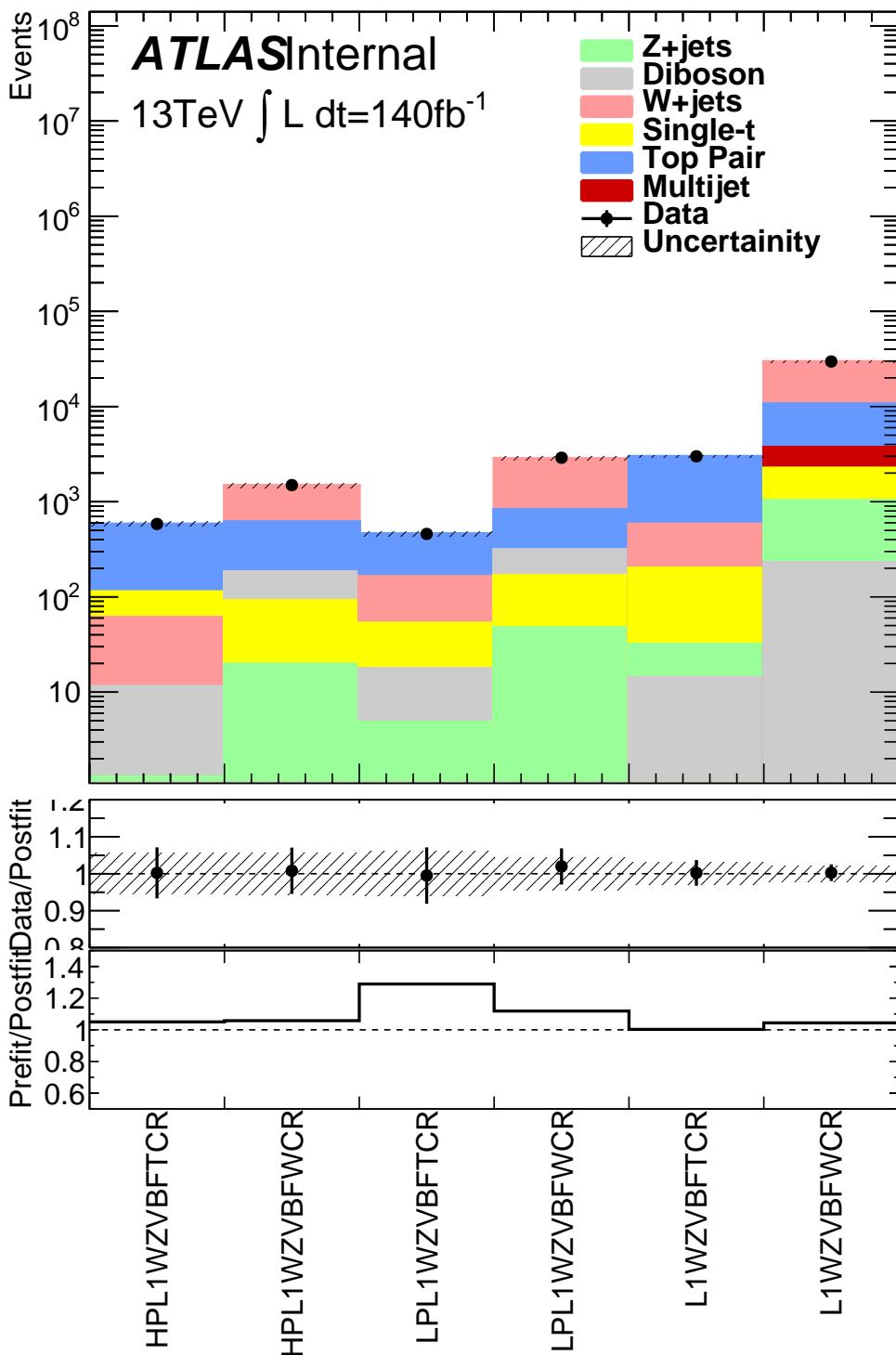
**Figure 13.1:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the DY WW control regions.



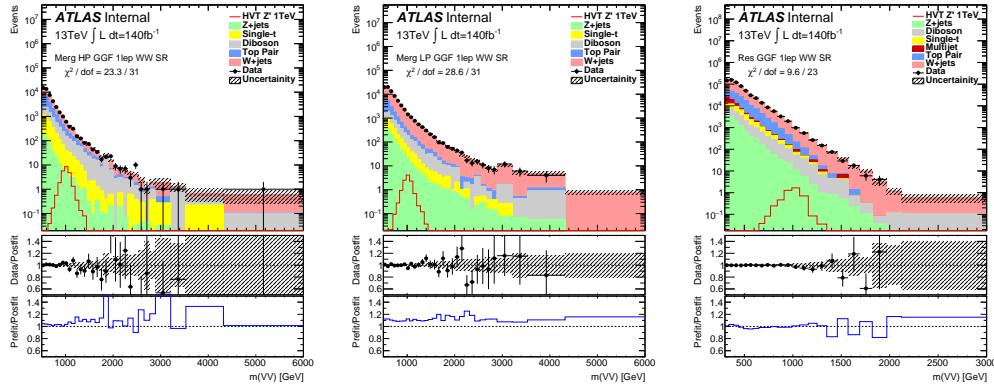
**Figure 13.2:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the DY  $WZ$  control regions.



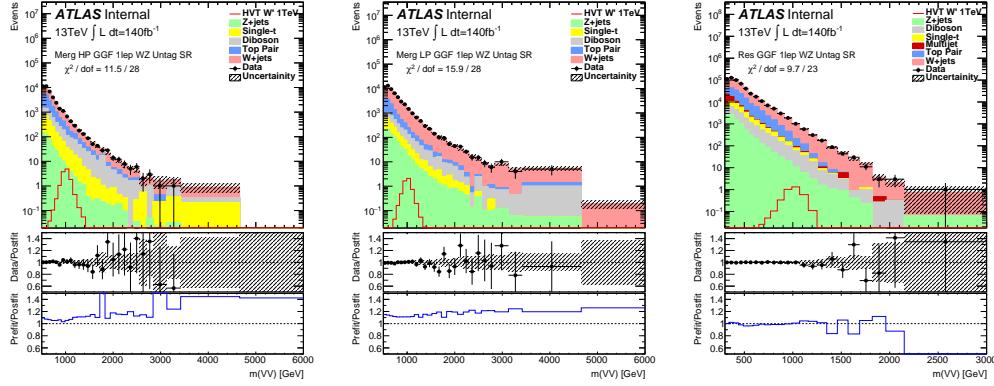
**Figure 13.3:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the VBF  $WW$  control regions.



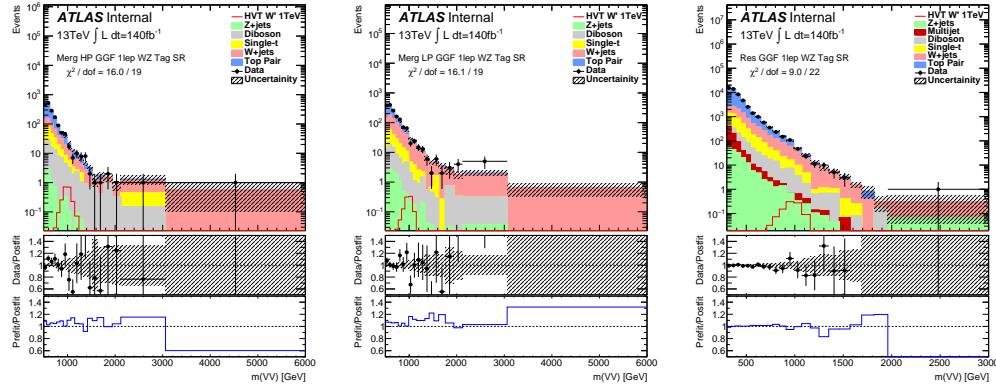
**Figure 13.4:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the VBF  $WZ$  control regions.



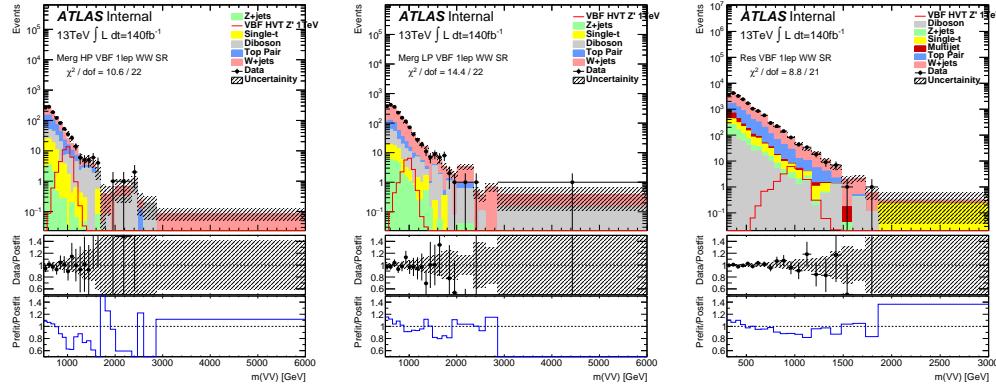
**Figure 13.5:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the GGF  $WW$  signal regions.



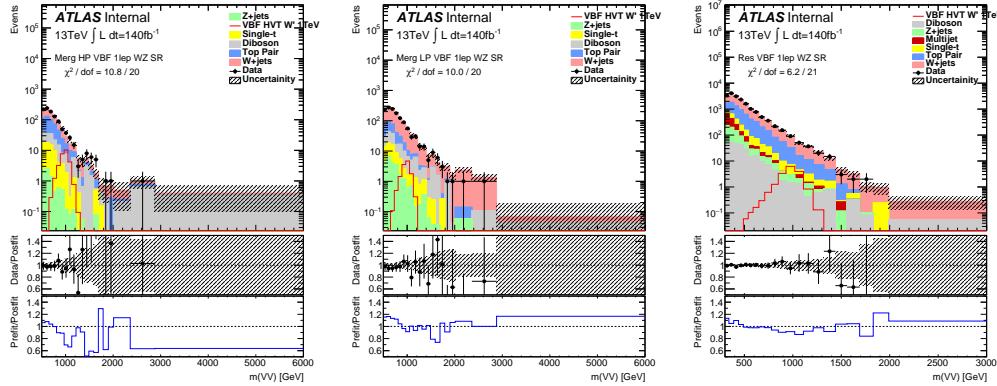
**Figure 13.6:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the GGF  $WZ$  Untag signal regions.



**Figure 13.7:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the GGF  $WZ$  Tag signal regions.



**Figure 13.8:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the VBF  $WZ$  Tag signal regions.

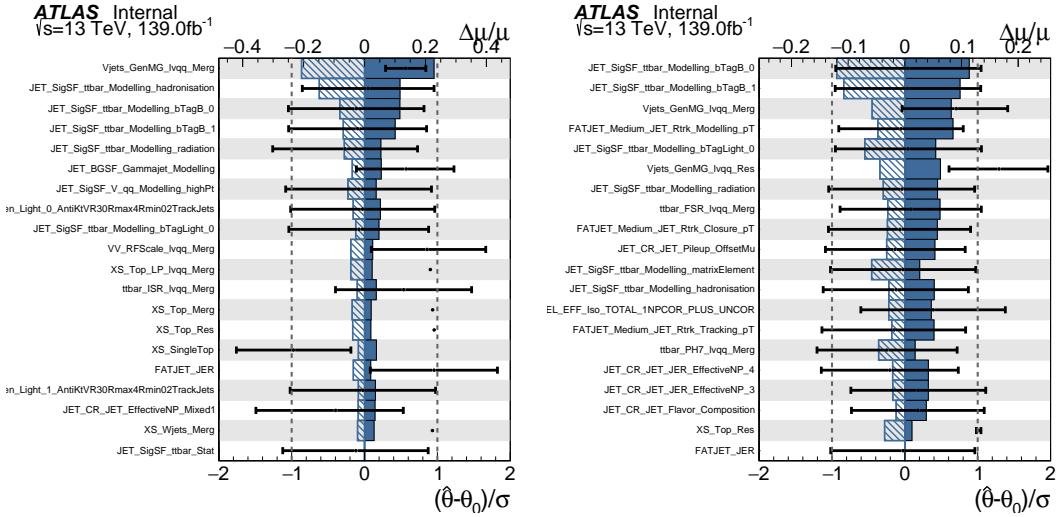


**Figure 13.9:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the VBF  $WZ$  Tag signal regions.

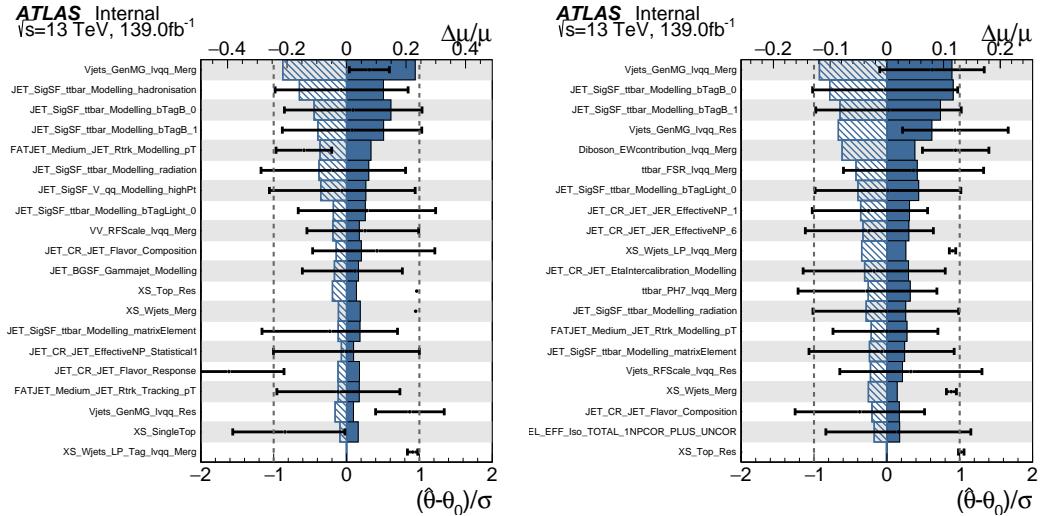
## 1347      13.2 Systematic Profiling and Correlations

1348      The ranked systematics and their fitted values are shown for the different  
 1349      analysis regions in Figure 13.10 - 13.12. Note that background normalizations for  
 1350       $W+jets$  and  $t\bar{t}$  are left free to float in the fit. This means the nominal normal-  
 1351      ization values are at one and the uncertainties are not shown in the ranked plots.  
 1352      Overall, systematics are not pulled outside their uncertainties, especially nuisance  
 1353      parameters that affect the fitted  $\mu$  value most significantly.

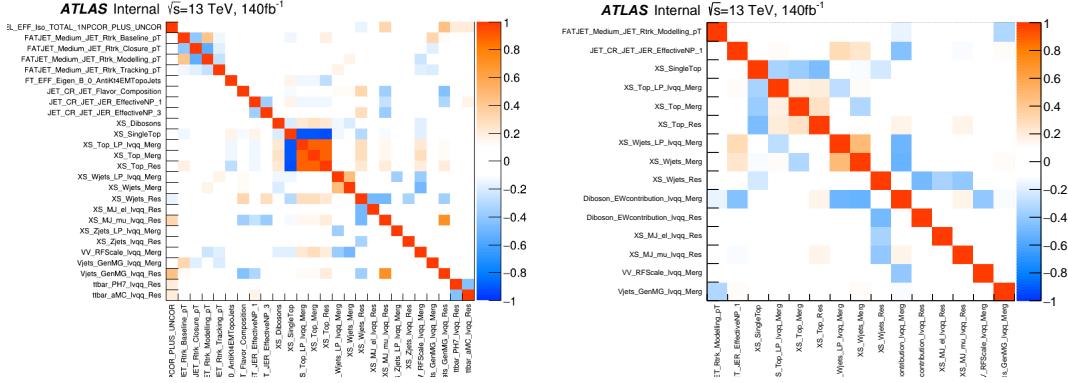
1354      The correlation between systematics are shown in Figures ???. Correlations  
 1355      between background normalization are expected. The remaining systematic cor-  
 1356      relations are not very strong or unexpected.



**Figure 13.10:** Ranked systematics and their fitted values for  $WW$  DY (right) and VBF (left) selections.



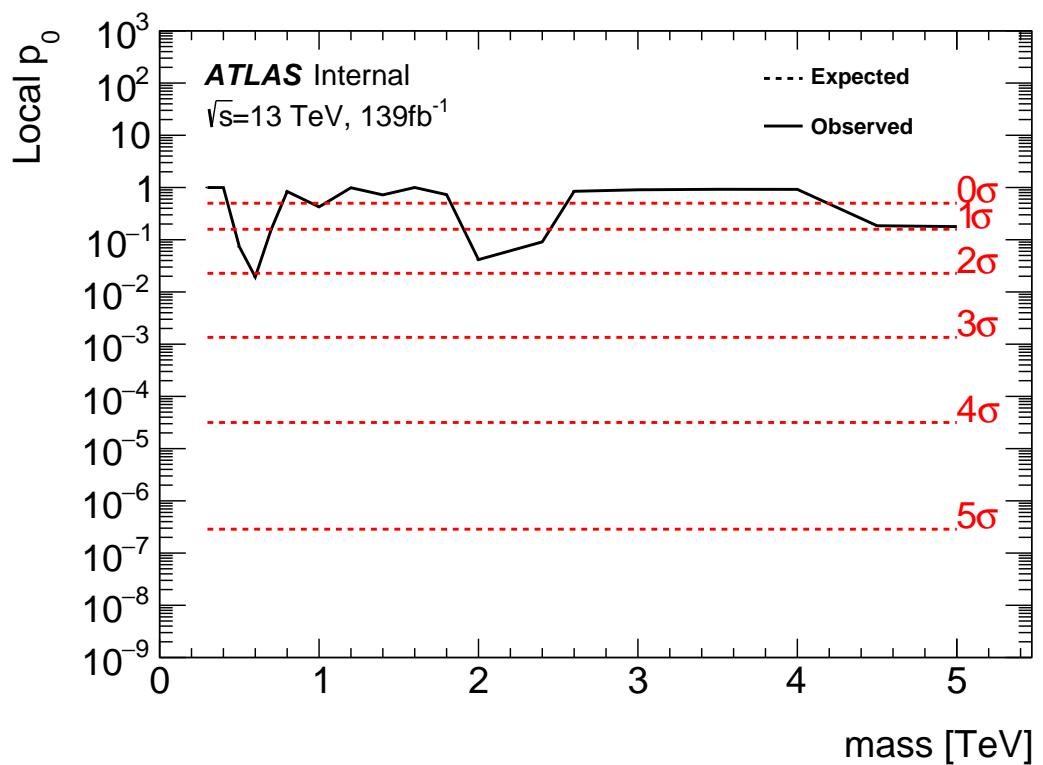
**Figure 13.11:** Ranked systematics and their fitted values for  $WZ$  DY (right) and VBF (left) selections.



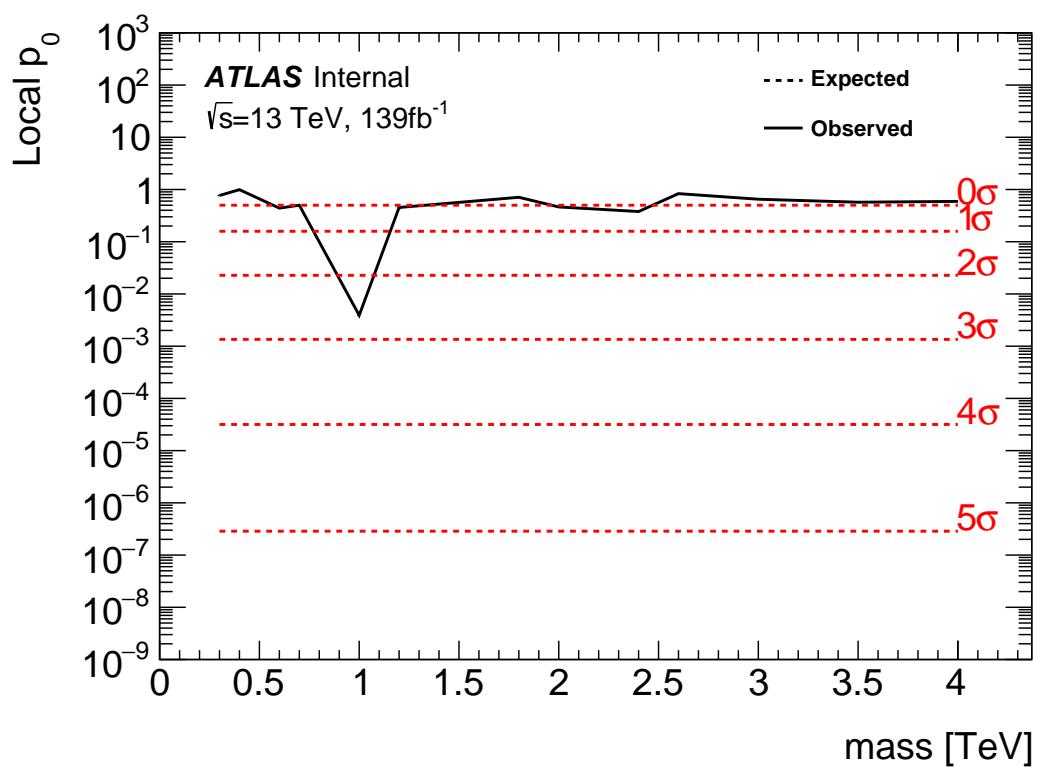
**Figure 13.12:** Correlations between systematics for  $WW$  DY (right) and VBF (left) selections.

### 13.3 Discovery Tests

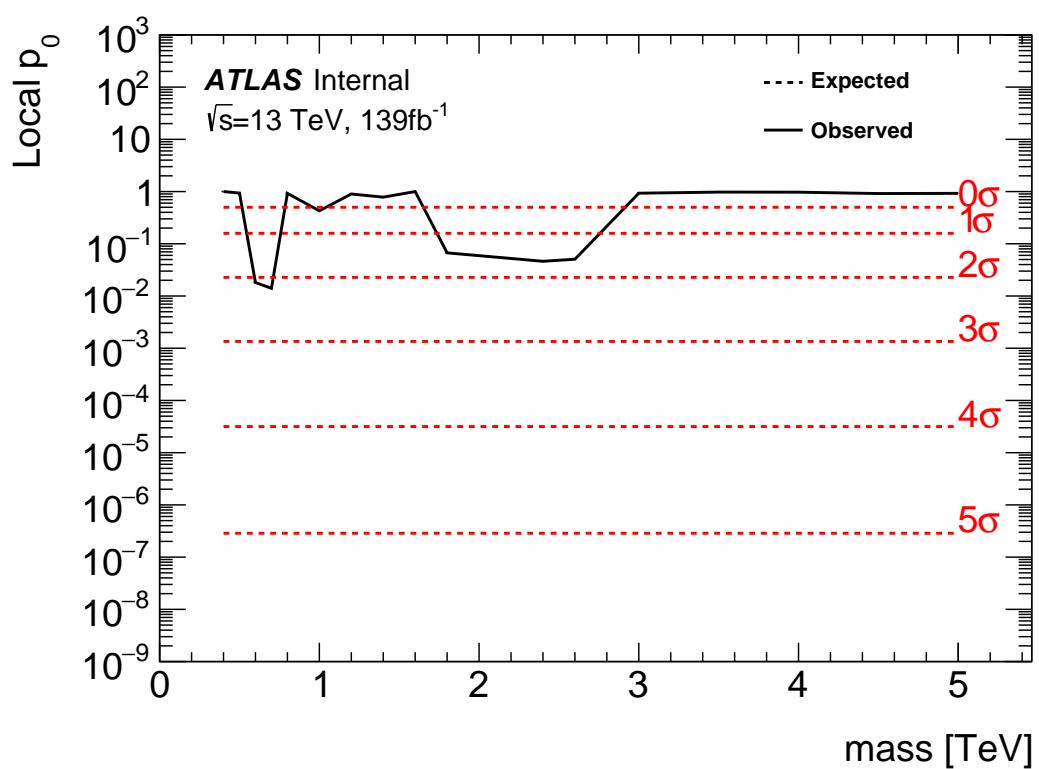
To test for the existence of signal in the observed dataset, the discovery tests discussed earlier are used to calculate p-values as a function of resonance mass. The results of these tests are shown in Figures 13.13 - 13.17. Across the different  $DY$  signals the largest excesses are  $\sim 2.2\sigma$  at 600 GeV and  $1.8\sigma$  at 2 TeV. The largest excesses for  $VBF$  signals are  $< 2.5\sigma$  at for 1 TeV resonances. As these deviations do not constitute discoveries, upper limits on  $\mu$  are calculated.



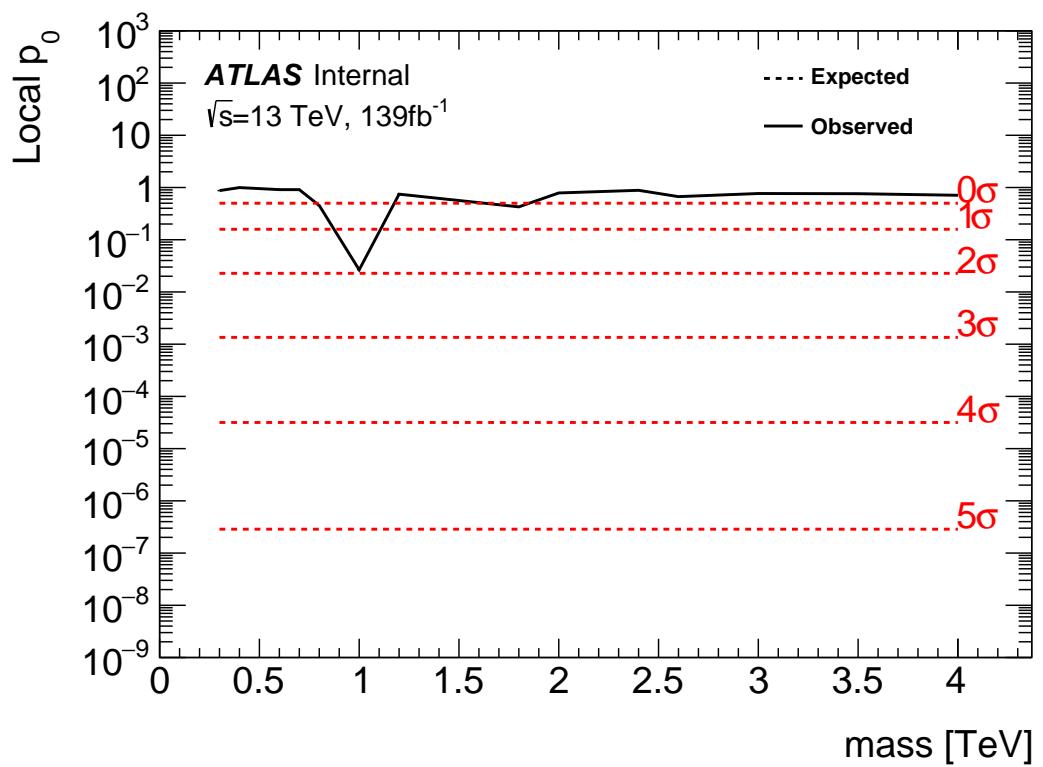
**Figure 13.13:** These plots show the measured  $p_0$  value as a function of resonance mass for HVT Z' DY production.



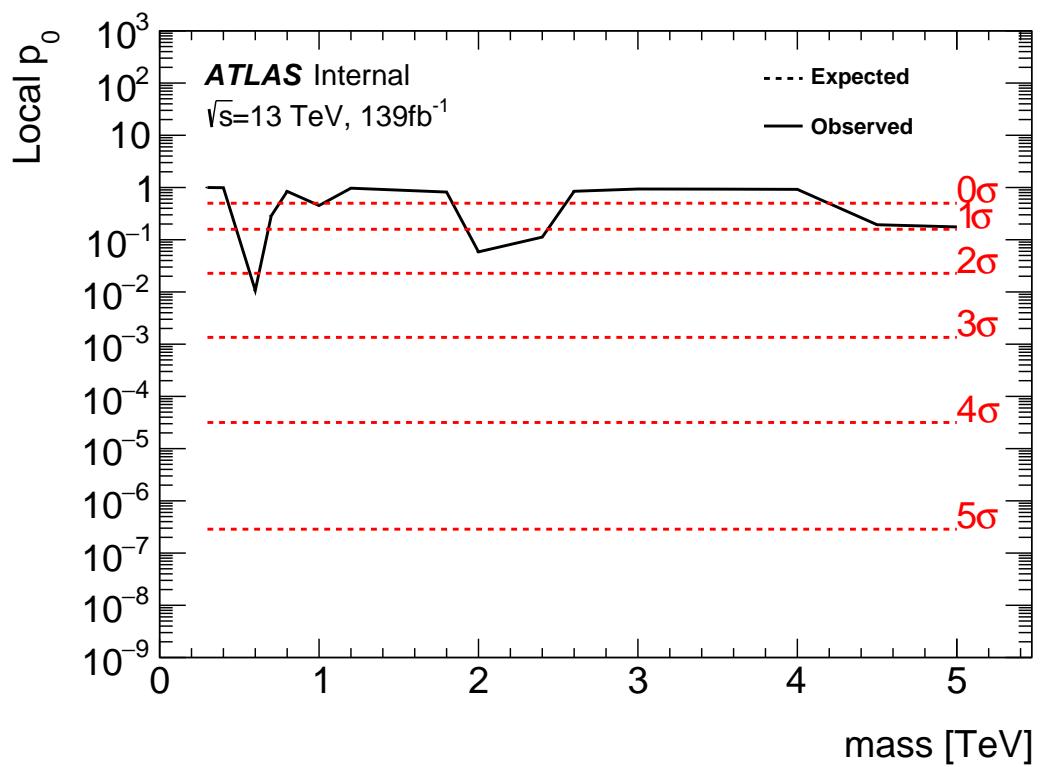
**Figure 13.14:** These plots show the measured  $p_0$  value as a function of resonance mass for HVT Z' VBF production.



**Figure 13.15:** These plots show the measured  $p_0$  value as a function of resonance mass for HVT W' DY production.



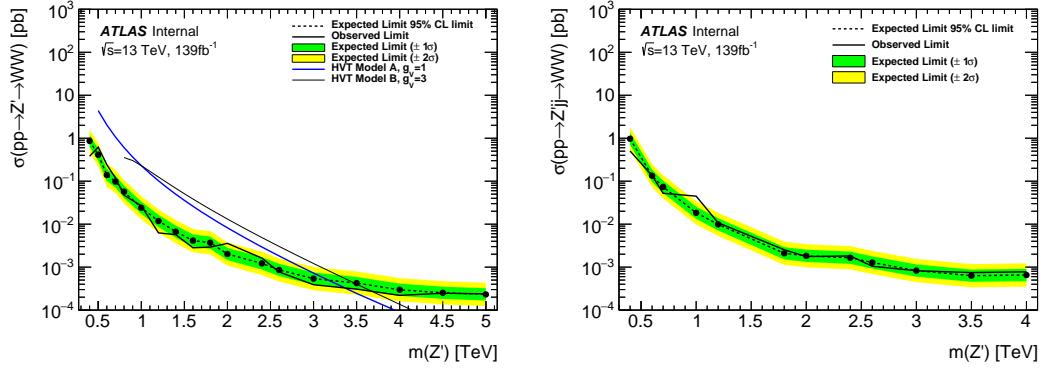
**Figure 13.16:** These plots show the measured  $p_0$  value as a function of resonance mass for HVT  $W'$  VBF production.



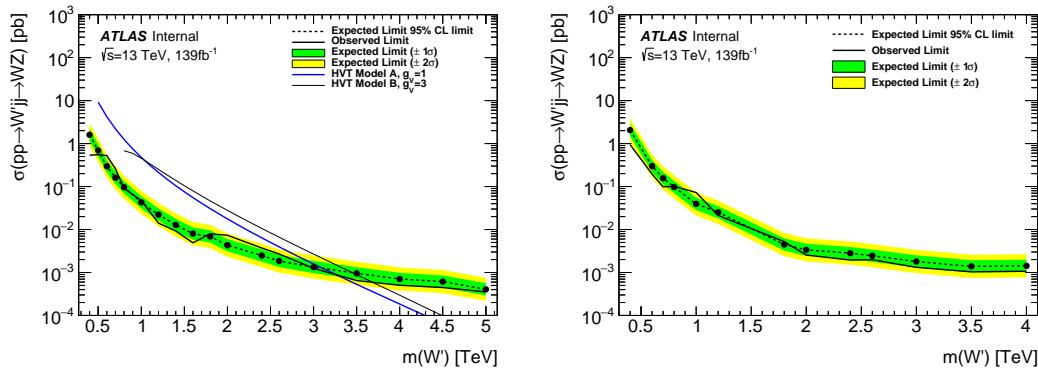
**Figure 13.17:** These plots show the measured  $p_0$  value as a function of resonance mass for the RS Graviton DY production.

## 1364 13.4 Limits

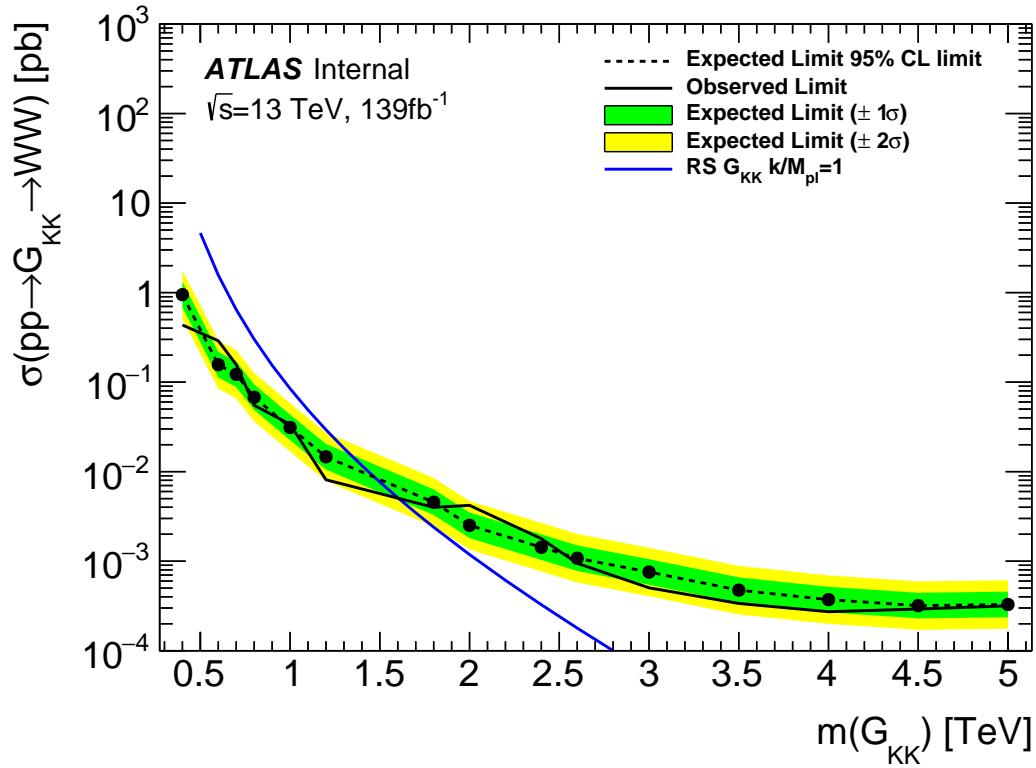
1365 Using the exclusion limits tests discussed previously, exclusion limits are set  
 1366 on  $\mu$  and consequently cross-sections for different signal models. Exclusion limits  
 1367 for the models considered are shown in Figure 13.18 - 13.20. These plots show the  
 1368 theory cross section for a given resonance to decay to  $WW/WZ$ . Also, an Asimov  
 1369 dataset is used to calculate the limits that could be set for the background only  
 1370 hypothesis with the associated errors on this predictions. Finally, the observed  
 1371 limits are shown in black. All signal mass where the theory prediction is less than  
 1372 the observed prediction are excluded at the 95% confidence level. These limits  
 1373 shown exclude HVT Model A  $W' < 3.4\text{TeV}$  and  $Z' < 3.3\text{ TeV}$  and Model B  $W'$   
 1374  $< 3.7\text{ TeV}$  and  $Z' < 3.7\text{ TeV}$ . Randall Sundrum Gravitons are excluded for masses  
 1375 below  $1.6\text{TeV}$ .



**Figure 13.18:** This figure shows theory, expected and observed limits for HVT  $W'$  DY (left) and VBF (right) production.



**Figure 13.19:** This figure shows theory, expected and observed limits for HVT  $Z'$  DY (left) and VBF (right) production.



**Figure 13.20:** This figure shows theory, expected and observed limits for RS Gravitons via DY production.

## Part V

1376

## Quark and Gluon Tagging

1377

<sub>1378</sub> **Chapter 14**

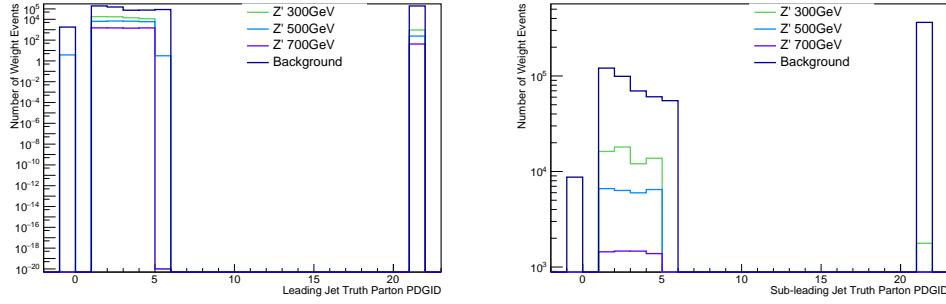
<sub>1379</sub> **Prospects**

<sub>1380</sub> For the resolved analysis, signal jets are quark enriched and background jets are  
<sub>1381</sub> gluon dominated. By classifying jets in the event as quark or gluon initiated, less  
<sub>1382</sub> background would contaminate the signal region. Figure 14.1 shows the PDGID  
<sub>1383</sub> for the truth parton matched to the jet (meaning the highest energy parton in  
<sub>1384</sub> the jet catchment area) in events passing the resolved signal region selections.  
<sub>1385</sub> PDGID = -1 corresponds to pileup jets,  $0 < \text{PDGID} < 6$  correspond to quarks  
<sub>1386</sub> and  $\text{PDGID} = 21$  corresponds to gluons. From this Figure, it is evident that a  
<sub>1387</sub> notable fraction of the background (all background events that passed the resolved  
<sub>1388</sub> SR are used) that contaminates the signal region contains gluon jets, especially  
<sub>1389</sub> for the sub-leading jet.

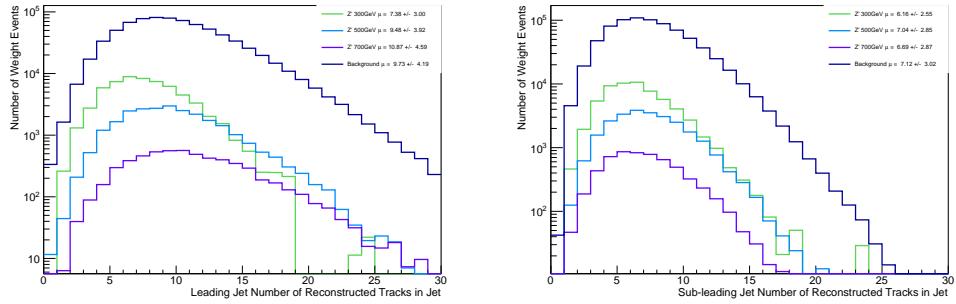
<sub>1390</sub> As gluons jets have more constituents and therefore more tracks ( $n_{trk}$ ), back-  
<sub>1391</sub> ground jets generally have more tracks than the signal jets. This is shown in  
<sub>1392</sub> Figure 14.2. Therefore, by cutting on the number of tracks in a jet, quark and  
<sub>1393</sub> gluon jets may be distinguished (i.e. jets with less than a given number of tracks  
<sub>1394</sub> are classified as a quark, otherwise the jet is classified as a gluon.) Moreover, as  
<sub>1395</sub> the momentum of the jet increases the number of tracks also increases logarith-  
<sub>1396</sub> mically [Cite nachman thesis Natasha], and improves tagging efficiency by about

1397 10% relative to a constant cut on the number of tracks. Therefore by applying a  
1398 cut on the number of tracks that scales with the  $\ln(p_T)$  is more powerful than a  
1399 threshold cut on the number of tracks. Figures 14.3-14.6 show normalized heat  
1400 maps of  $\ln(p_T)$  vs the number of reconstructed tracks for the background and  
1401 HVT  $Z'$  signals. This information is also shown in table ???. In these plots it is  
1402 evident that the number of tracks in the background jets grows more quickly with  
1403  $\ln(p_T)$  than for the signal jets. This is expected given that the signal is quark  
1404 dominated and the background is gluon dominated.

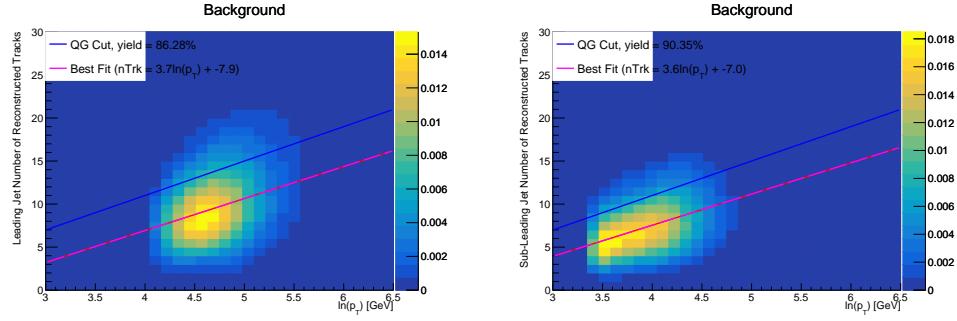
1405 In Figure 14.7 is the ROC Curve for quark gluon tagging with cut on the  
1406 number of tracks in a jet that depends on  $\ln(p_T)$ . The sum of the backgrounds  
1407 in the signal region were used for this curve. Here the quark tagging efficiency is  
1408 the ratio of quarks tagged as quarks to the total number of quarks in the signal  
1409 region. The gluon rejection is calculated as the reciprocal of the gluon tagging  
1410 efficiency. For example, choosing a 90% efficient working point with a rejection  
1411 of 1.4 corresponds to a slope of 4 and intercept of -5. Tagging both jets in this  
1412 analysis would yield an efficiency of  $90\%^{n_{jets}}$ . Focusing on the background in  
1413 Figure 14.8, this cut helps minimize gluon contamination in the signal region.  
1414 Also, from these heat maps it is obvious that the number of tracks in gluon jets  
1415 grows more quickly than those in quark jets.



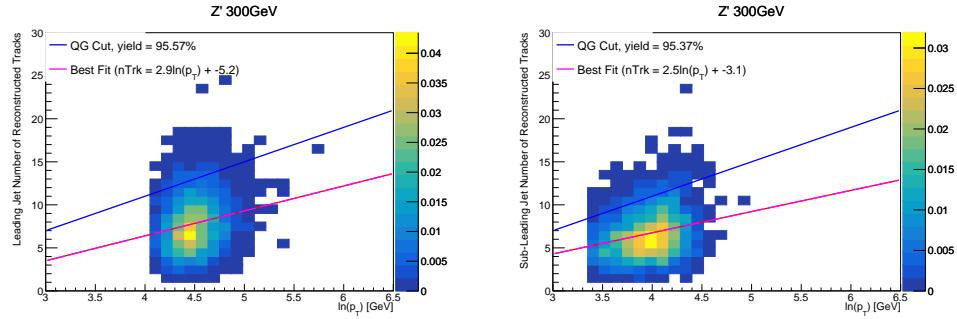
**Figure 14.1:** PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV Z' signals and the background (all simulated backgrounds that pass SR selections).



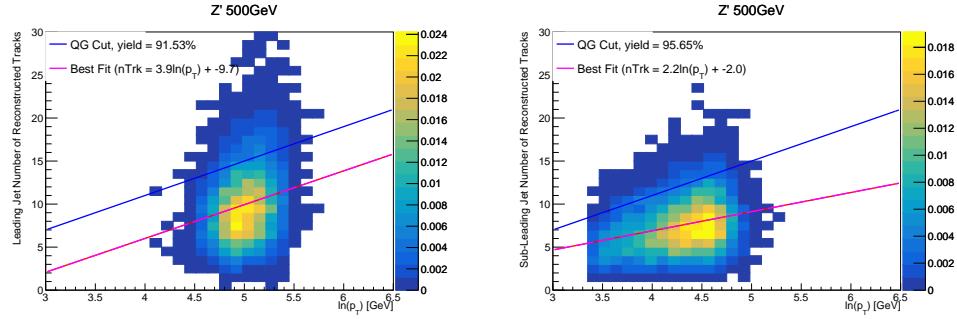
**Figure 14.2:** The number of tracks in small-R jets in events passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets. These distributions are shown for 300, 500, and 700GeV Z' signals and the background.



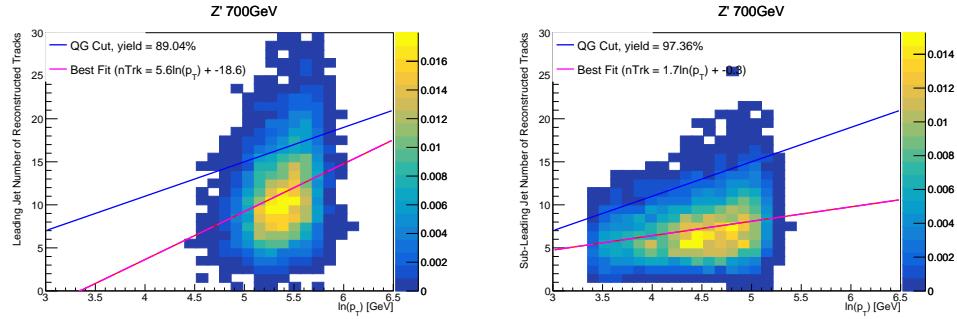
**Figure 14.3:** The number of tracks in background small-R jets in events passing the Resolved GGF WW Signal region selection vs.  $\ln(p_T)$  for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks  $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one.



**Figure 14.4:** The number of tracks in small-R jets in 300GeV  $Z'$  events passing the Resolved GGF WW Signal region selection vs.  $\ln(p_T)$  for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks  $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one.



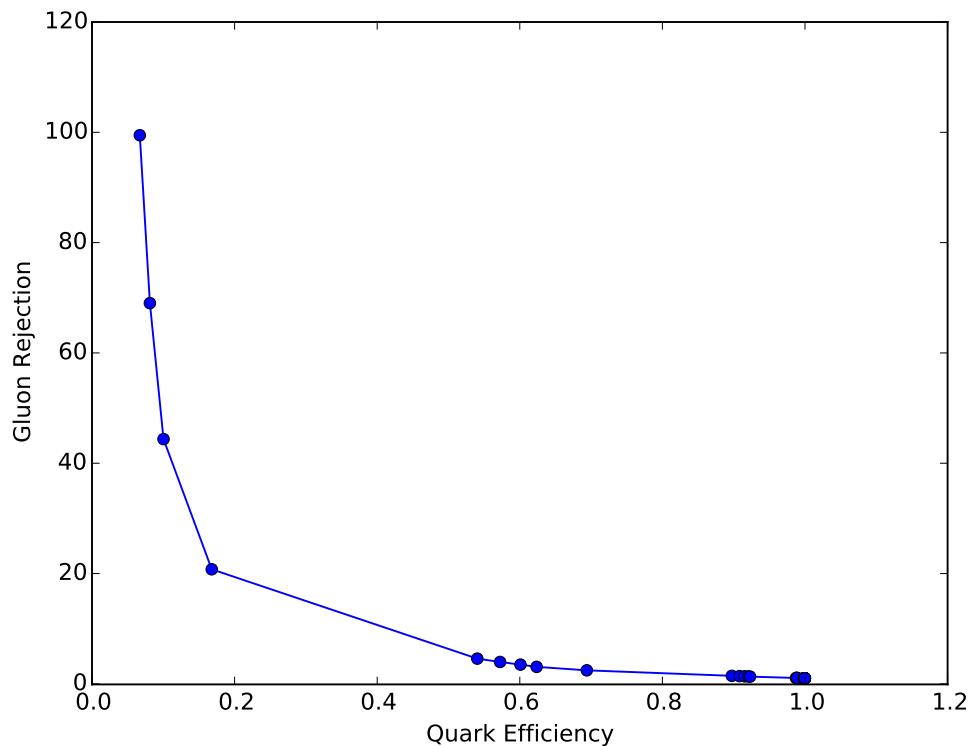
**Figure 14.5:** The number of tracks in small-R jets in 500GeV  $Z'$  events passing the Resolved GGF WW Signal region selection vs.  $\ln(p_T)$  for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks  $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one.



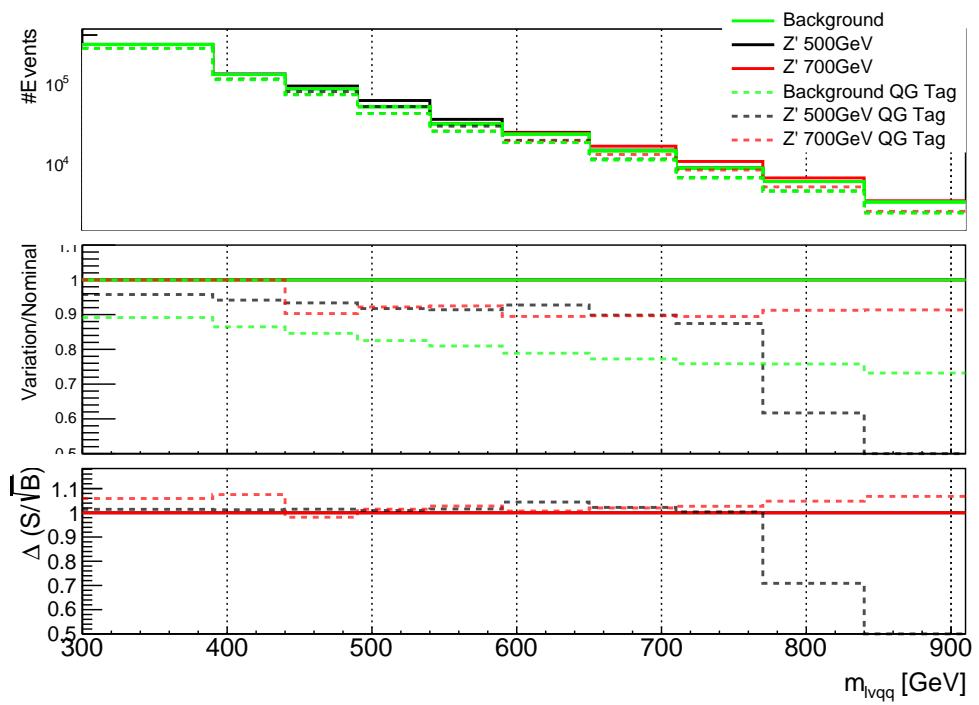
**Figure 14.6:** The number of tracks in small-R jets in 700GeV  $Z'$  events passing the Resolved GGF WW Signal region selection vs.  $\ln(p_T)$  for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks  $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one.

Sample	Best Fit Slope	Best Fit Intercept	QG Tag Yield
Backgrounds	3.7	-7.9	86%
HVT $Z'$ 300 GeV	2.9	-5.2	95%
HVT $Z'$ 500 GeV	3.9	-9.7	92%

**Table 14.1:** This table shows the best fit slope and intercept for the 2-d distribution of number of tracks vs. jet  $\ln(p_T)$  for the leading jet in the background and HVT  $Z'$  samples. The tagging efficiency is shown for the 90% working point in the last column. The background jets contain more gluons than the signal jets. Consequently, the best fit line for the background predicts larger values of the number of tracks in jets for the background than the considered signals.



**Figure 14.7:** ROC Curve for Quark and Gluon Tagging with a cut on the number of tracks that depends on the  $\ln(p_T)$ .



**Figure 14.8:** The top panel shows the distribution of  $m_{lvqq}$  with and without quark gluon tagging. The middle panel shows the ratio of the signals and backgrounds with and without quark gluon tagging. The bottom panel shows the change in  $S/\sqrt{B}$  with quark gluon tagging.

1416 **Chapter 15**

1417  $n_{trk}$  Calibration

1418 The number of tracks in jets depends on modeling and experimental systemat-  
1419 ics. Consequently, the efficiency of a quark-gluon tagger based on the number of  
1420 tracks in jets would have associated uncertainties. In the context of the resonance  
1421 search discussed, these uncertainties would be treated as systematics that impact  
1422 the  $m_{WV}$  distributions that are used for discovery tests.

1423 Modeling uncertainties are obtained by assessing PDF and ME uncertainties  
1424 on the number of charged particles in particle-level jets in dijet events. The  
1425 number of charged particles as a function of jet  $p_T$  is calculated using an Iterative  
1426 Bayesian (IB) technique [cite paper].

1427 This measurement ([9]) uses the ATLAS 2012 pp collision dataset, correspond-  
1428 ing to  $20.3/\text{fb}$  at center-of-mass energy  $\sqrt{s} = 8\text{TeV}$ . The number of charged con-  
1429 stituents depends on fragmentation modeling and matrix elements, which do not  
1430 depend on  $s$ . For this reason, it is safe to use these uncertainties for  $\text{sqrt}(s)=13\text{TeV}$ .  
1431 Monte Carlo (MC) samples are used to determine the response matrix. The MC  
1432 sample is a dijet sample generated with Pythia 8.175 using CT10 PDF and AU2  
1433 tune. The anti- $k_T$  algorithm is used to cluster jets with a radius parameter  $R$   
1434  $= 0.4$ . Jets are required to have  $|\eta| < 2.1$ . Tracks in jets are required to have

1435  $p_T > 500\text{MeV}$ ,  $|\eta| < 2.5$ , track-fit  $\chi^2 < 3.0$  and originate from the primary ver-  
1436 tex. Matching tracks to jets is accomplished using ghost-association [cite]. In this  
1437 technique, jets are re-clustered with the track collection augmented with "ghost"  
1438 versions of tracks. These "ghosts" tracks have the same direction as their parent  
1439 track, but infinitesimal track  $p_T$ . This insures meta-jet properties (e.g.  $\eta$ ,  $p_T$ , etc)  
1440 are unchanged. A track is matched to a jet if it's ghost version remains in the jet  
1441 after re-clustering. Further details of the data, object, and event selection may be  
1442 found in [cite 35].

1443 To select dijet topologies events are required to have at least two jets with  
1444  $p_T > 50\text{GeV}$  that are relatively well-balanced ( $p_T^{\text{lead}}/p_T^{\text{sub-leading}} < 1.5$ ).

1445 In the IB technique, the prior distribution and number of iterations are the  
1446 inputs [cite Bayesian paper]. The IB response matrix connects number of charged  
1447 particles to the number of tracks in jets determined using the simulated samples.  
1448 This response matrix is used to unfold data to extract the  $n_c$ . Before applying  
1449 the response matrix a fake factor is applied. This accounts for jets that pass  
1450 detector level selections, but not particle level selections. Following this, the IB  
1451 method iteratively applies the response matrix using the nominal Pythia 8.175  
1452 sample as a prior. The number of IB iterations is chosen to minimize unfolding  
1453 bias and statistical fluctuations. For this measurement four iterations was found  
1454 to be optimal by minimizing the unfolding bias from pseudodata simulated with  
1455 Herwig++ with a prior from Pythia 8 AU2. Finally, the inefficiency factor is  
1456 applied to account for events passing particle level selection but not detector  
1457 level, yielding the unfolded nCharged distribution.

1458 This process is prone to three main sources of bias: response matrix, correction  
1459 factor, and unfolding procedure uncertainties. The response matrix is sensitive to  
1460 experimental uncertainties impacting jet track reconstruction and calorimeter jet

1461     $p_T$ . Correction factors are also sensitive to experimental uncertainties (e.g. JES)  
 1462    as such uncertainties modify detector level acceptance. Sensitivity to particle  
 1463    level acceptance is calculated by comparing Pythia and Herwig. Finally, the bias  
 1464    from the IB prior choice is determined by reweighting the particle-level spectrum,  
 1465    so the simulated detector level spectrum more closely matches the uncorrected  
 1466    data. Unfolding this modified detector-level simulation and comparing it to the re-  
 1467    weighted particle-level spectrum indicates bias from the prior distribution choice.

1468    A summary of all the systematic uncertainties associated with this unfolding  
 1469    may be found in [ref paper]. Total uncertainties are < 7% for the number of  
 1470    charged particles in jets. The unfolded distribution of the nCharged in jets from  
 1471    data are further analyzed to extract the quark and gluon nCharged distributions.  
 1472    In dijet events, the jet with a larger  $\eta$  is more energetic and therefore more likely  
 1473    to be a quark. This is due to the quarks in protons generally having a larger  
 1474    fraction of the total momentum of the proton constituents. The more central jet  
 1475    is more likely to be a gluon-initiated jet. This correlation between jet  $\eta$  and flavor  
 1476    may then be used to extract nCharged in  $p_T$  bins using:

$$\langle n_c^f \rangle = f_q^f \langle n_c^q \rangle + f_g^f \langle n_c^g \rangle \quad (15.1)$$

1477

$$\langle n_c^c \rangle = f_q^c \langle n_c^q \rangle + f_g^c \langle n_c^g \rangle \quad (15.2)$$

1478    In this equation the f and c subscripts denote the more forward and central  
 1479    jets, respectively. The q and g subscripts denote quark and gluon. The fraction  
 1480    of more forward jets that are say gluons is denoted by  $f_g^f$ . The other relevant jet  
 1481    fractions are denoted with the same naming scheme. Finally,  $\langle n_c \rangle$  is the average  
 1482    number of charged particles in a jet in a given  $p_T$  bin. To show that Eq. (??) may  
 1483    be used to extract quark and gluon  $n_c$  distributions the extracted distributions  
 1484    are compared to  $n_c$  distributions determined using the jet flavor in simulation.

1485   Figure [add figure natasha] shows that the extracted and true distributions differ  
1486   by < 1% over the  $p_T$  range probed for this study. Moreover, this implies that  $n_c$   
1487   depends only on the flavor of the initiating parton and jet  $p_T$ .

1488   These extracted distributions are prone to PDF and ME biases. The bias from  
1489   the choice of the CT10 PDF for the Pythia sample is accounted for by comparing  
1490   quark/gluon fractions for the nominal CT10 sample with its eigenvector variations.  
1491   Comparing the quark/gluon fractions from Pythia 8 and Herwig++ quantify the  
1492   uncertainty from the ME calculation. These uncertainties are added in quadra-  
1493   ture with the unfolding uncertainty to give the total modelling uncertainty on  
1494   the extracted  $n_c$  distribution. This is shown in Figure 16.2.

1495   To apply these uncertainties in  $n_c$  distributions in data, per-jet event weights  
1496   are associated with each uncertainty according to:

$$w_i(n_c) = \frac{P(n_c | n_c > \pm \sigma_{n_c}^i)}{P(n_c | n_c >)} \quad (15.3)$$

1497   In Eq. (??), i denotes the uncertainty considered, P is the Poisson probability,  
1498   and  $\sigma_{n_c}^i$  represents the average impact of the uncertainty on  $n_c$ .

1499   The previous uncertainties described accounted for modeling uncertainty as-  
1500   sociated with the number of charged particles in a jet. However,  $n_c$  is not a  
1501   measurable quantity. Instead the number of tracks in a jet is measured, which is  
1502   a proxy for  $n_c$ . Therefore the uncertainties associated with the measurement of  
1503   nTracks must also be considered ([11]). These uncertainties were calculated using  
1504   a Pythia 8 dijet sample with NNPDF 23 and Run 2 data. Track reconstruction  
1505   efficiency and fake rates are the dominant sources of nTrack uncertainties.

1506   The track reconstruction efficiency is affected by the uncertainty of the de-  
1507   scription of the ID material in simulation and the modeling of charged-particle  
1508   interactions with this material. These uncertainties are accounted for by varying

1509 the ID material by 5-25% (dependent on the region of the detector considered).  
1510 The difference in the tracking efficiency between the nominal and varied simula-  
1511 tion give the uncertainty on the track reconstruction efficiency. Another important  
1512 source of track reconstruction inefficiency arises in the core of jets. The high den-  
1513 sity of tracks in the jet cores can cause ID clusters to merge. The fraction of lost  
1514 tracks due to merging is given by the fraction of tracks that have a charge of two  
1515 minimum ionizing particles. This quantity is compared between data and simu-  
1516 lation resulting in an uncertainty of 0.4% on tracks with  $\Delta R < 0.1$ . Combining  
1517 these effects gives a total uncertainty as a function of  $p_T$  and  $\eta$  that is generally  
1518  $< 2\%$  [references figure 44 from [11]].

1519      Fake tracks are the other dominant source of nTrk uncertainty. Fake tracks  
1520 are tracks that cannot be associated to a single particle. Often these tracks are a  
1521 result of random combinations of hits from charged particles that overlap in space.  
1522 In dense environments, such as the core of jets or high-pileup environments, fake  
1523 tracks are more likely. Fake tracks are estimated with a 'control region method'  
1524 which is briefly summarized here [[10]]. By applying a series of track selections  
1525 to enrich the fraction of fake tracks (e.g.  $|d_0| > 0.1$ , track  $\chi^2 > 1.4$ , etc) in  
1526 simulation, templates for fake track parameters are calculated. These templates  
1527 are then fit to data to determine the fraction of fake tracks. On average the fake  
1528 rate is found to be 30% (independent of  $p_T$  and  $\eta$ ).

1529      To assess the impact of these two detector level uncertainties, tracks are ran-  
1530 domly dropped according to the rates described above. Reconstruction and fake  
1531 uncertainties both lower the number of tracks, hence these uncertainties are one-  
1532 sided. By dropping tracks in this way a varied nTrk distribution is calculated for  
1533 both uncertainties. The associated per-jet event weights are then calculated in  
1534 the same way as the modeling weights as:

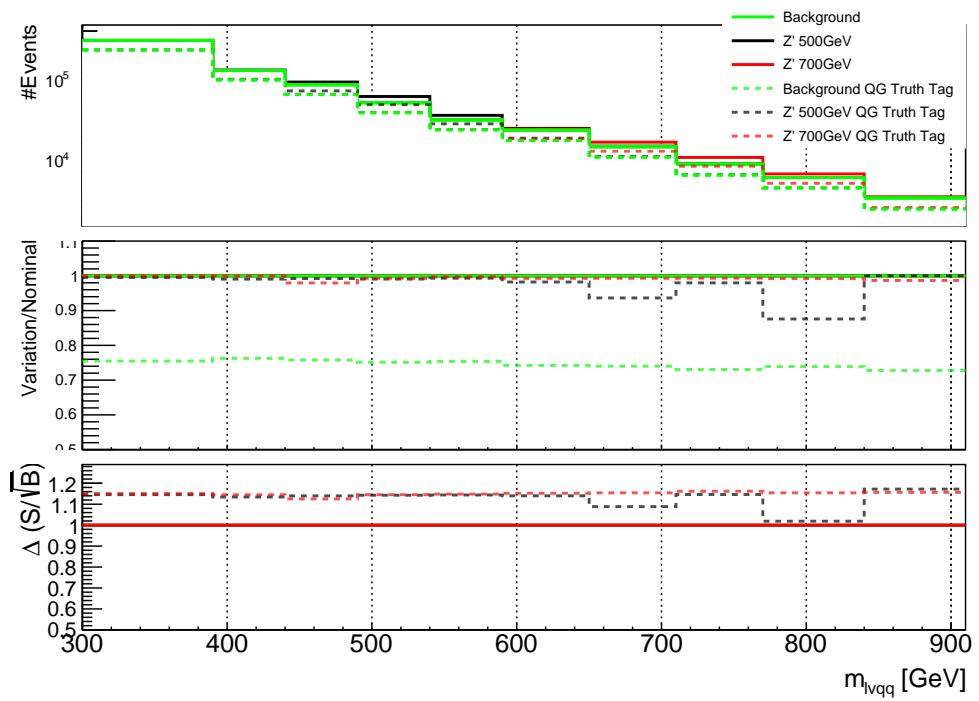
$$w_i(n_c) = \frac{P(n_{trk} | < n_{trk} > \pm \sigma_{n_{trk}}^i)}{P(n_{trk} | < n_{trk} >)} \quad (15.4)$$

1535        Adding the modeling and detector level uncertainties in quadrature gives the  
 1536        overall nTrack uncertainty. The effects of the individual uncertainties on the nTrk  
 1537        distributions can be seen in Fig 16.4. Fig 16.3 shows the  $m_{lvqq}$  and nTrk distri-  
 1538        butions for the W and Top control regions before likelihood fitting. In these plots  
 1539        the nTrk uncertainties improve agreement between data and MC. The remaining  
 1540        differences are likely covered by likelihood fitting and improving the analysis itself.

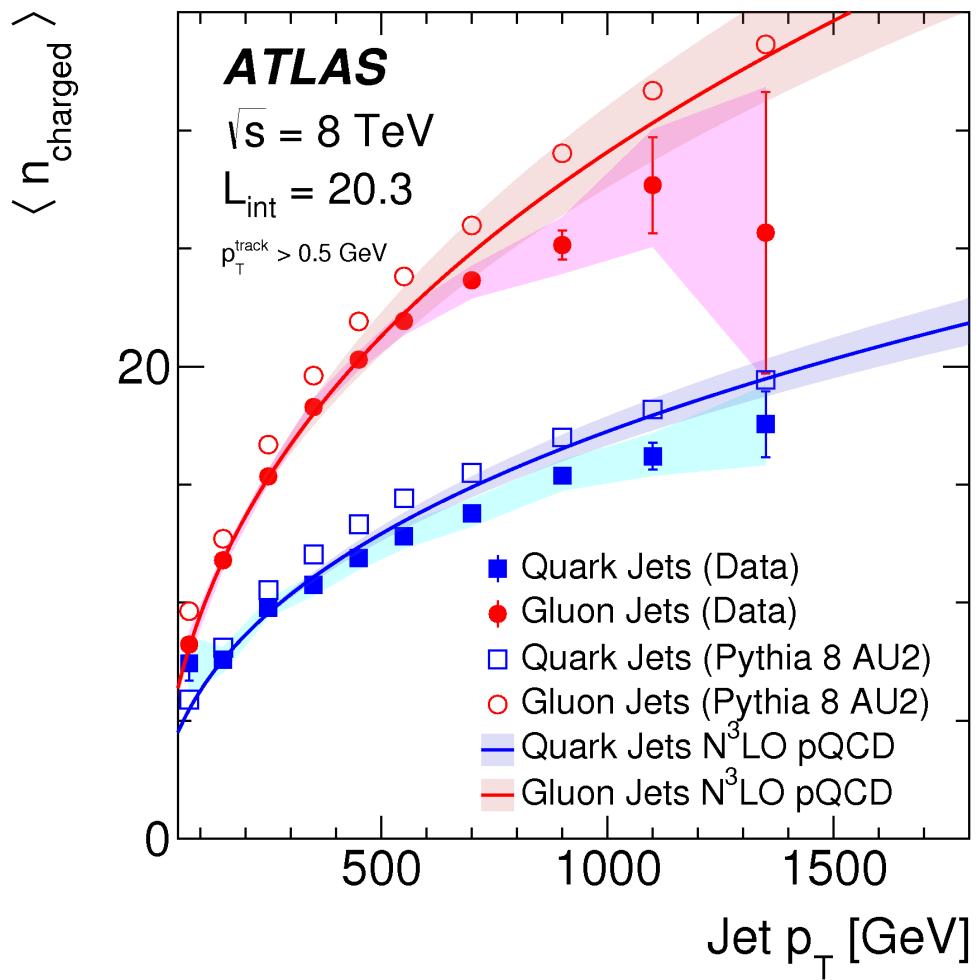
1541 **Chapter 16**

1542 **Application**

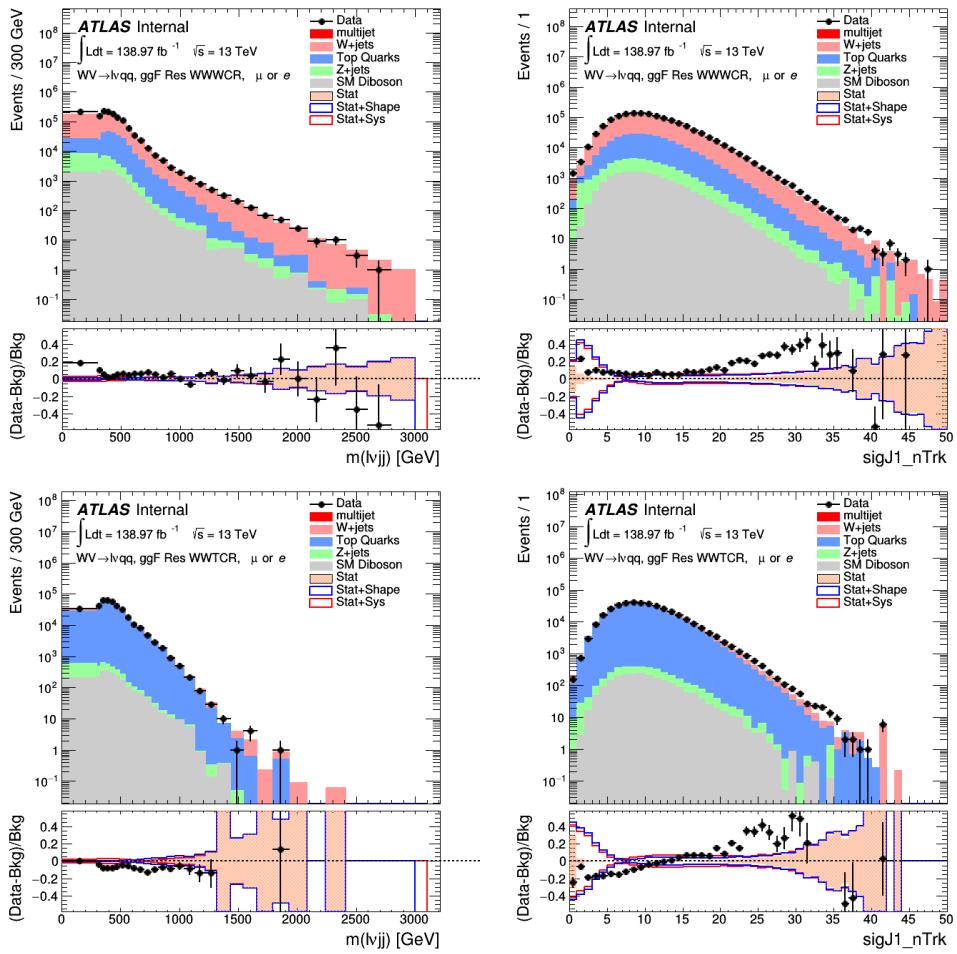
1543 Using the 90% WP of the  $n_{trk}$  tagger improves  $S/\sqrt{B}$  is  $\sim 3\%$  as shown in  
1544 Figure 14.8. Although,  $n_{trk}$  is the single most powerful discriminating variable  
1545 for quark and gluon jets, the addition of other jet variables would improve the  
1546 classification efficiency. Figure 16.1 shows the possible improvement of 10%  
1547 in jet classification using the truth label of the jets to classify jets. This type of  
1548 improvement is possible by using variables such as jet width, and energy correlata-  
1549 tors. Figure [add BDT figure/use 1612.01551.pdf] shows for a 90% quark tagging  
1550 efficiency for a 100 GeV jet, a BDT improve the gluon rejection by 0.4. Once this  
1551 tagger is calibrated it would improve the analysis sensitivity of this channel.



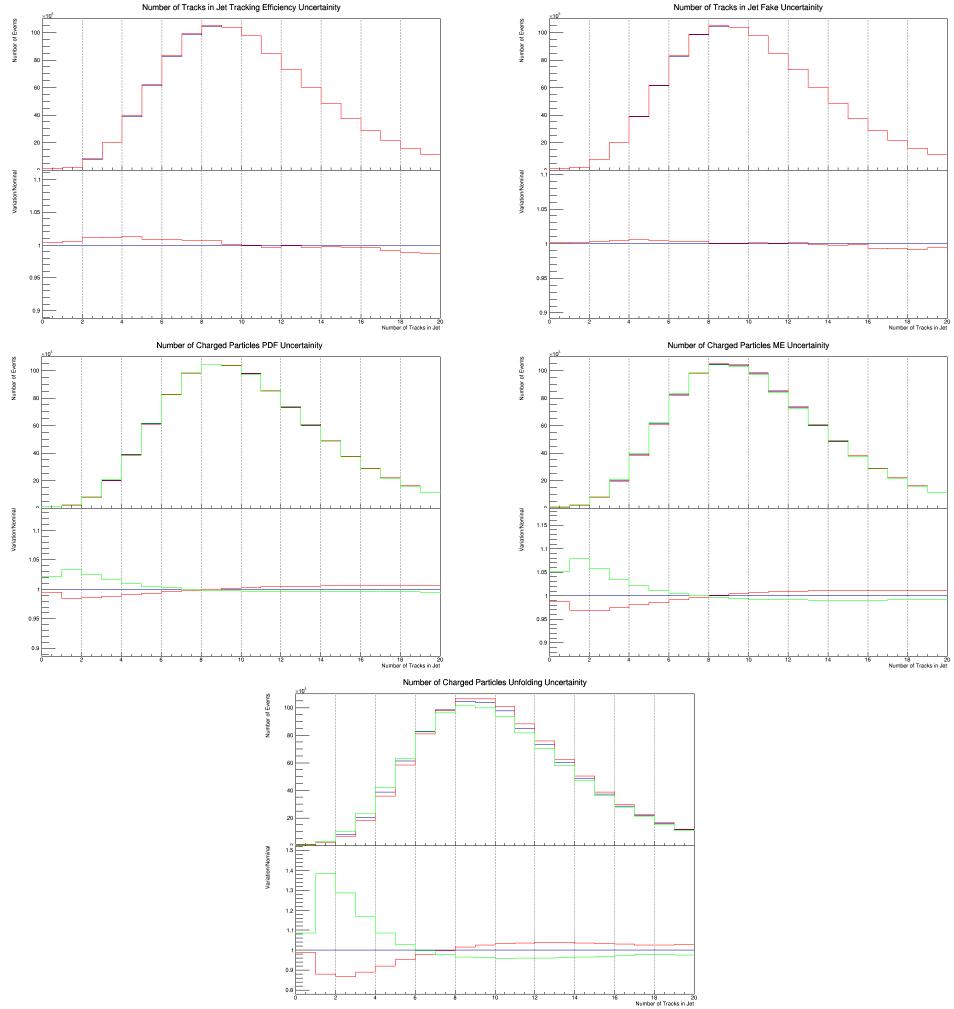
**Figure 16.1:** The top panel shows the distribution of  $m_{lvqq}$  with and without requiring jets to be true quarks. The middle panel shows the ratio of the signals and backgrounds with and without requiring jets to be true quarks. The bottom panel shows the change in  $S/\sqrt{B}$  when requiring jets to be true quarks..



**Figure 16.2:** Unfolded and extracted  $n_C$  qg dstbs..



**Figure 16.3:** PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV  $Z'$  signals and the background.



**Figure 16.4:** These figures show the impact of the uncertainties on the number of tracks in the leading jet in the sum of the background sample in the Resolved GGF WW SR (a) tracking efficiency (b) fake (c) PDF (d) ME (e) unfolding uncertainties.

## **Part VI**

1552

## **Conclusion**

1553

1554 **Chapter 17**

1555 **Conclusions**

1556 A search for  $WW$  and  $WZ$  diboson resonance production in  $\ell\nu qq$  final states  
1557 was performed using  $139\text{fb}^{-1}$  of  $pp$  collision data collected at a center-of-mass  
1558 energy of  $\sqrt{s} = 13\text{TeV}$  by that ATLAS detector at the LHC between 2015 and  
1559 2018. No excess of events above the background-only expectation was observed.  
1560 The largest local excess is approximately  $2.7\sigma$ , which is not significant. Limits  
1561 on the production cross section are obtained for the HVT  $W'$  and  $Z'$  and RS  
1562 Gravitons. Signal masses below 3.4 (3.7) TeV are excluded for HVT  $W'$  Model  
1563 A(B). Signal masses below 3.3 (3.7) TeV are excluded for HVT  $Z'$  Model A(B).  
1564 Randall Sundrum Gravitons are excluded for masses below 1.6 TeV. Going forward,  
1565 improving the classification of jets in events would improve analysis sensitivity.  
1566 To distinguish quark from gluon jets a jet tagger based on the number of tracks in  
1567 jets is studied in the context of this search. Finally, the calibration of the number  
1568 of tracks in jets is discussed.

# Bibliography

- [1] Lecture notes particle physics ii.
- [2] Kaustubh Agashe, Hooman Davoudiasl, Gilad Perez, and Amarjit Soni. Warped Gravitons at the LHC and Beyond. *Phys. Rev.*, D76:036006, 2007.
- [3] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298 – 318, 1977.
- [4] M. N. Chernodub. Background magnetic field stabilizes qcd string against breaking, 2010.
- [5] ATLAS Collaboration. Atlas muon reconstruction performance in lhc run 2.
- [6] ATLAS Collaboration. Summary plots from the atlas standard model physics group.
- [7] ATLAS Collaboration. Tagging and suppression of pileup jets with the atlas detector.
- [8] ATLAS Collaboration. Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at  $\sqrt{s} = 13$  tev with the atlas detector. arXiv: 1703.09665 [hep-ex].
- [9] ATLAS Collaboration. Measurement of the charged-particle multiplicity inside jets from  $s=\sqrt{8}$  tev pp collisions with the atlas detector. arXiv:1602.00988 [hep-ex].
- [10] ATLAS Collaboration. Performance of the atlas track reconstruction algorithms in dense environments in lhc run 2. arXiv:1704.07983 [hep-ex].
- [11] ATLAS Collaboration. Properties of jet fragmentation using charged particles measured with the atlas detector in pp collisions at  $\sqrt{s} = 13$  tev. arXiv:1906.09254 [hep-ex].
- [12] Alex Dias and V. Pleitez. Grand unification and proton stability near the peccei-quinn scale. *Physical Review D*, 70, 07 2004.

- 1595 [13] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. Qcd ma-  
1596 trix elements + parton showers. the nlo case. *Journal of High Energy Physics*,  
1597 2013(4), Apr 2013.
- 1598 [14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic opti-  
1599 mization, 2014.
- 1600 [15] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jets with variable r. *Journal*  
1601 *of High Energy Physics*, 2009(06):059–059, Jun 2009.
- 1602 [16] Gregory Soyez Matteo Cacciari, Gavin P. Salam. The anti- $k_T$  jet clustering  
1603 algorithm. arXiv:0802.1189 [hep-ph].
- 1604 [17] Duccio Pappadopulo, Andrea Thamm, Riccardo Torre, and Andrea Wulzer.  
1605 Heavy vector triplets: bridging theory and data. *Journal of High Energy*  
1606 *Physics*, 2014(9), Sep 2014.
- 1607 [18] Antonio Pich. The Standard Model of Electroweak Interactions. In *Proceed-  
1608 ings, High-energy Physics. Proceedings, 18th European School (ESHEP 2010):  
1609 Raseborg, Finland, June 20 - July 3, 2010*, pages 1–50, 2012. [,1(2012)].
- 1610 [19] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small  
1611 extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.
- 1612 [20] Sebastian Raschka. Model evaluation, model selection, and algorithm selec-  
1613 tion in machine learning, 2018.
- 1614 [21] Tania Robens and Tim Stefaniak. Lhc benchmark scenarios for the real higgs  
1615 singlet extension of the standard model. *The European Physical Journal C*,  
1616 76(5), May 2016.
- 1617 [22] Alex Sherstinsky. Fundamentals of recurrent neural network (RNN) and long  
1618 short-term memory (LSTM) network. *CoRR*, abs/1808.03314, 2018.
- 1619 [23] Muhammed Ali Sit and Ibrahim Demir. Decentralized flood forecasting using  
1620 deep neural networks. Jun 2019.
- 1621 [24] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural net-  
1622 work regularization, 2014.