

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	xv
Abstract	xvii
Dedication	xviii
Acknowledgments	xix
I Introduction	1
1 Introduction	2
II Theoretical Motivation	6
2 The Standard Model of Particle Physics	7
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
3 Standard Model Successes and Limitations	24
4 New Physics Models with Diboson Resonances	27
4.1 Randall Sundrum Bulk Model	27
4.2 Extended Scalar Sector	29

4.3	Simple Standard Model Extensions	30
III	Experimental Setup	33
5	LHC	34
5.1	LHC Layout and Design	36
6	The ATLAS Detector	41
6.1	Coordinate System	43
6.2	Inner Detector	44
6.2.1	Pixel Detector	47
6.2.2	Semiconductor Tracker	47
6.2.3	Transition Radiation Tracker	47
6.3	Calorimeters	49
6.4	Muon Spectrometer	53
6.5	Magnet System	57
6.6	Trigger System	58
IV	Method	60
7	Dataset and Simulated Samples	61
7.1	Dataset	61
7.2	Simulated Samples	64
8	Objects	65
8.1	Electrons	65
8.2	Muons	66
8.3	Jets	67
8.3.1	Small-R jets	69
8.3.2	Large-R jets	70
8.3.3	Variable Radius jets	73
8.3.4	Jet Flavor Tagging	74
8.4	MET/Neutrinos	74
8.5	Overlap Removal	74
9	Event Selection and Categorization	76
9.1	Pre-selection	76
9.2	Trigger	76
9.3	non-VBF/VBF RNN	78
9.4	Topological Cuts	83
9.5	Selection Acceptance times efficiency for Signal Events	93

9.6	Background Estimate	94
9.6.1	Control Regions	94
9.6.2	Fake Lepton Backgrounds	94
10	Systematic Uncertainties	108
10.1	Experimental Systematics	108
10.2	Theory Systematics	110
11	Statistical Analysis	118
11.1	Likelihood Function Definition	118
11.2	Fit Configuration	119
11.3	Best Fit μ	121
11.4	Discovery Test	122
11.5	Exclusion Limits	123
V	Results	125
12	Statistical Interpretation	126
12.1	Discovery Tests	126
12.2	Systematic Profiling and Correlations	132
12.3	Expected and Measured Yields	133
12.4	Limits	147
VI	Quark and Gluon Tagging	149
13	Prospects	150
14	n_{trk} Calibration	157
15	Application	163
VII	Conclusion	168
16	Conclusions	169
	Bibliography	170

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. [16]	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. [11] . . .	22
3.1	A comparison of cross section measurements at $\sqrt{s} = 7, 8, 13$ TeV from ATLAS compared to theoretical measurements. From Ref. [5]	26
4.1	Cartoon of RS Bulk Model	28
5.1	Scaling of cross sections with \sqrt{s} . Natasha: write more here . . .	35
5.2	LHC Layout. Natasha write more	37
5.3	LHC Accelerator. Natasha write more	39
6.1	Big picture layout of ATLAS detector. Natasha: write more . . .	42
6.2	Big picture layout of ATLAS detector. Natasha: write more . . .	42
6.3	A simplified schematic of how different particles interact and are detected within ATLAS.	43
6.4	Layout of ATLAS Inner Detector	45
6.5	Layout of ATLAS ID Barrel System.	46
6.6	Overview of ATLAS electromagnetic and hadronic calorimeters. .	50

6.7	Schematic of ECAL	51
6.8	Schematic of HCAL	52
6.9	Schematic of Muon Spectrometer [cite G35]	55
6.10	Schematic of MDT chamber. [cite G35]	56
6.11	Schematic of RPC chamber, which is used for triggering in the central region of the detector [cite G35].	56
6.12	Schematic of TGC chamber, which is used for triggering in the muon end-cap region. [cite G35]	57
6.13	Layout of ATLAS magnet systems.	58
7.1	Integrated luminosity for data collected from ATLAS from 2011 - 2018	62
7.2	Mean number of interactions per crossing for data collected from ATLAS from 2011 - 2018	63
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 [4].	67
8.2	[7] This diagram shows the calibration stages for EM jets.	69
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?	72
8.4	Natasha write caption	73
9.1	RNN architecture. Natasha add caption	80
9.2	RNN Score distribution for ggF and VBF signals and backgrounds.	81
9.3	ROC curve using k-fold validation for RNN.	82

9.4	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.	83
9.5	Event Categorization. Natasha write more.	86
9.6	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.	87
9.7	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.	88
9.8	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.	89
9.9	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.	90
9.10	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.	91
9.11	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.	92

9.12 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 (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.	93
9.13 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.	94
9.14 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 calculated as remaining data components after excluding known MC	96
9.15 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the WW electron channel. The MJ template is obtained from the pre-MJ-fit.	97
9.16 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the WW muon channel. The MJ template is obtained from the pre-MJ-fit.	98
9.17 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the WZ untag electron channel. The MJ template is obtained from the pre-MJ-fit.	99
9.18 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the WZ untag muon channel. The MJ template is obtained from the pre-MJ-fit.	100
9.19 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the WZ untag electron channel. The MJ template is obtained from the pre-MJ-fit.	101

9.20 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the WZ untag muon channel. The MJ template is obtained from the pre-MJ-fit.	102
9.21 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the VBF WW electron channel. The MJ template is obtained from the pre-MJ-fit.	103
9.22 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the VBF WW muon channel. The MJ template is obtained from the pre-MJ-fit.	104
9.23 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the VBF WZ electron channel. The MJ template is obtained from the pre-MJ-fit.	105
9.24 Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the VBF WZ muon channel. The MJ template is obtained from the pre-MJ-fit.	106
10.1 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).	112
10.2 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 inlet shows the ratio of the two.	113
10.3 Ratio between the variations of generator (red) and hadronization (blue) variations for the Merged regime for $t\bar{t}$ sample.	114
10.4 Ratio between the variations of generator (red) and hadronization (blue) variations for the Resolved regime for $t\bar{t}$ sample.	114

10.5	Ratio between the variations of ISR up (red) and down (blue) variations for the Merged regime for $t\bar{t}$ sample.	115
10.6	Ratio between the variations of ISR up (red) and down (blue) variations for the Resolved regime for $t\bar{t}$ sample.	116
10.7	Ratio between the variations of FSR up (red) and down (blue) variations for the Merged regime for $t\bar{t}$ sample.	117
10.8	Ratio between the variations of FSR up (red) and down (blue) variations for the Resolved regime for $t\bar{t}$ sample.	117
11.1	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.	121
11.2	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.	121
12.1	These plots show the measured p_0 value as a function of resonance mass for HVT Z' DY production.	127
12.2	These plots show the measured p_0 value as a function of resonance mass for HVT Z' VBF production.	128
12.3	These plots show the measured p_0 value as a function of resonance mass for HVT W' DY production.	129
12.4	These plots show the measured p_0 value as a function of resonance mass for HVT W' VBF production.	130
12.5	These plots show the measured p_0 value as a function of resonance mass for the RS Graviton DY production.	131
12.6	Ranked systematics and their fitted values for WW DY (right) and VBF (left) selections.	132
12.7	Ranked systematics and their fitted values for WZ DY (right) and VBF (left) selections.	133
12.8	Correlations between systematics for WW DY (right) and VBF (left) selections.	133

12.9 This figure shows the distribution of $m_{\ell\nu qq}$ in the DY WW control regions.	141
12.10 This figure shows the distribution of $m_{\ell\nu qq}$ in the DY WZ control regions.	142
12.11 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WW control regions.	143
12.12 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WZ control regions.	144
12.13 This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF WW signal regions.	145
12.14 This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF WZ Untag signal regions.	145
12.15 This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF WZ Tag signal regions.	146
12.16 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WZ Tag signal regions.	146
12.17 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WZ Tag signal regions.	147
12.18 This figure shows theory, expected and observed limits for HVT W' DY (left) and VBF (right) production.	147
12.19 This figure shows theory, expected and observed limits for HVT Z' DY (left) and VBF (right) production.	148
12.20 This figure shows theory, expected and observed limits for RS Gravitons via DY production.	148
13.1 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.	152

13.2 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.	152
13.3 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.	153
13.4 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.	153
13.5 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.	154
13.6 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.	154

13.7 The number of tracks in leading small-R jets in background events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Gluons (b) Quarks 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.	155
13.8 ROC Curve for Quark and Gluon Tagging with a cut on the number of tracks that depends on the $\ln(p_T)$	155
13.9 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.	156
15.1 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.	164
15.2 Unfolded and extracted n_C qg dstbs.	165
15.3 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.	166
15.4 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.	167

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. $SU(2)_L$ gauge transformations allow one to go between rows and $SU(3)_C$ transformations allow one to go between columns in these fermion representations. [REWORD]	13
9.1	The list of triggers used in the analysis.	77
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.	85
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).	86
9.4	Definitions of “inverted” leptons used in multijet control region . .	96
9.5	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.	107

12.1	Expected and Measured for DY WW $W+\text{jets}$, $t\bar{t}$ control regions and signal regions.	134
12.2	Expected and Measured for DY WZ $W+\text{jets}$, $t\bar{t}$ tag and untag control regions.	135
12.3	Expected and Measured for DY WZ $W+\text{jets}$, $t\bar{t}$ tag and untag signal regions.	136
12.4	Expected and Measured for VBF WW $W+\text{jets}$, $t\bar{t}$ control regions and signal regions.	137
12.5	Expected and Measured for VBF WZ $W+\text{jets}$, $t\bar{t}$ control regions and signal regions.	138
12.6	Fitted background normalizations for $t\bar{t}$ and $W+\text{jets}$ backgrounds for the DY WW analysis region.	139
12.7	Fitted background normalizations for $t\bar{t}$ and $W+\text{jets}$ backgrounds for the DY WZ analysis region.	139
12.8	Fitted background normalizations for $t\bar{t}$ and $W+\text{jets}$ backgrounds for the VBF WW analysis region.	139
12.9	Fitted background normalizations for $t\bar{t}$ and $W+\text{jets}$ backgrounds for the VBF WZ analysis region.	140

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 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

xviii

Acknowledgments

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

Part I

Introduction

³ Chapter 1

⁴ Introduction

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

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

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

72 For each signal model, the simulated and measured resonance mass distribu-
73 tions in the relevant SR and CRs with their systematics are combined to construct

74 a likelihood. This likelihood parameterized by the signal strength parameter, μ
75 and considered systematics on the resonance mass distribution. This likelihood
76 is used to quantify the likelihood of a certain signal model given the anticipated
77 backgrounds and measured data.

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

Part II

85

Theoretical Motivation

86

⁸⁷ **Chapter 2**

⁸⁸ **The Standard Model of Particle**

⁸⁹ **Physics**

⁹⁰ **2.1 Introduction**

⁹¹ By determining the dynamics of the most elementary degrees of freedom, par-
⁹² ticle physics hopes to uncover the fundamental laws of the universe. The definition
⁹³ of elementary has evolved through time and currently refers to matter and force
⁹⁴ mediating particles: fermions and bosons, respectively. The Standard Model of
⁹⁵ Particle Physics (SM) describes the quantum behavior of three of the four funda-
⁹⁶ mental forces: weak, strong, and electromagnetic, via boson and fermion interac-
⁹⁷ tions. Gravity is not included in the SM and still under investigation.

⁹⁸ **2.2 Quantum Field Theory**

⁹⁹ In the SM, forces (and particles) are represented as fields. In this context,
¹⁰⁰ fields are mathematical objects that define a tensor (e.g. scalar, vector, etc) at
¹⁰¹ every point on a manifold, here the manifold is space-time. These fields obey laws

102 dictated by Quantum Field Theory (QFT). Particles arise naturally in QFT as
103 quantized field excitations localized in spacetime.

104 According to Noether's theorem, symmetries of a field give rise to conserved
105 quantities (e.g. time-translation invariance leads to energy conservation). Often
106 in the history of physics, a conserved quantity of a field is found and then the
107 underlying symmetry of the field is inferred. Gauge symmetries are symmetries
108 among the internal degrees of freedom of the field (components of the tensor),
109 which give rise to quantities associated with fields. By specifying the symmetries
110 of a system the dynamics and conserved quantities of the system may be succinctly
111 defined.

112 **2.3 $U(1)_{EM}$ Local Gauge Invariance**

113 The Lagrangian of Quantum Electrodynamics (QED) describes the electro-
114 magnetic force. QED may be derived by requiring local $U(1)_{EM}$ gauge invariance
115 of the free dirac fermion Lagrangian, ψ :

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

116 This symmetry may be represented as a complex number with unit modulus,
117 $e^{i\theta}$. $U(1)$ gauge invariance requires this gauge transformation of ψ will leave the
118 Lagrangian unchanged.

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

119 NB: This transformation is a local gauge transformation as θ depends on the
120 spacetime coordinate.

121 By requiring this symmetry of the free Dirac fermion Lagrangian:

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

₁₂₂ The mass term is unaffected, but the kinetic term is modified due to $\theta(x)$.

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

₁₂₃

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

₁₂₄ The $\partial_\mu\theta$ terms breaks the gauge invariance of the Lagrangian. By introducing a
₁₂₅ new field, A_μ we can recover the gauge invariance of the derivative. Now redefining
₁₂₆ the derivative as the covariant derivative:

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

₁₂₇ And letting A_μ transform under $U(1)$ as:

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

₁₂₈ The transformed covariant derivative becomes:

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

₁₂₉

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

₁₃₀

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

₁₃₁ The covariant derivative can be made gauage invariant by setting the last term
₁₃₂ to zero.

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

133 So now A_μ transforms as:

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

134 Finally, replacing the derivative with the covariant derivative the Dirac La-
135 grangian we have:

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

136

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

137 Here $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$. This last term in the Lagrangian is the kinetic
138 energy of the gauge boson field.

139 So we have derived the QED Lagrangian. By requiring the free Dirac La-
140 grangian to be invariant under U(1) transformations we have generated a new
141 gauge boson field, A_μ , which describes the photon. As expected the photon inter-
142 acts with fermions.

143 Stepping back, a global U(1) gauge symmetry of the free Dirac Lagrangian
144 implies we cannot measure the absolute phase of a charged particle. A local U(1)
145 gauge symmetry changes the phase of fields differently across space time. For this
146 type of transformation to leave the Lagrangian invariant, we had to introduce an
147 additional field, A_μ , which "communicates" these phase changes across space-time.
148 In less formal language this effectively means: if the field at one location changes,
149 this change is conferred to other particles via A_μ .

150 2.4 Yang-Mills Gauge Theories

151 Requiring $U(1)_{EM}$ gauge invariance of the free Dirac Lagrangian gave us QED.

152 Requiring different gauge symmetries we can derive the structure of other inter-

153 actions. Any gauge symmetry may be written as:

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

154 Here θ is a dimensionless real parameter and T is the generator of the gauge

155 symmetry group. With this the covariant derivative can be written as:

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

156 Then the gauge field must transform as:

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

157 Here f is the structure constant of the gauge group. The field strength tensor

158 is given by:

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

159

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

160 This gives the Yang-Mills Lagrangian:

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

161 2.5 Particles in the Standard Model

162 The SM consists of fermions (half-integer spin matter constituents) and bosons
163 (integer spin force mediators). Fermions are spinor representations of the Poincare
164 group and can be further separated into leptons and quarks. Bosons are the result
165 of requiring a particular symmetry among the spinor fields:

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

166 $SU(3)_C$ is the symmetry group of the strong force and generates eight gluon
167 fields, G_μ . $SU(2)_L$ is the symmetry group of the Electroweak force and generates
168 three electroweak boson fields. The mixing of this $SU(2)_L$ and $U(1)_Y$ gives rise
169 to the photon field, where Y is the weak-hypercharge:

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

170 Q is the electromagnetic charge, and T_3 is the z-component of the weak isospin.
171 Weak isospin is the charge associated with the $SU(2)_L$ symmetry. The correspond-
172 ing covariant derivative is then:

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

173 It is important to note that the gauge symmetry of the SM yields a particular
174 structure of the fermion representations. So for a given fermion to interact with
175 a given gauge field it must have a non-zero corresponding Noether charge for
176 that gauge symmetry. If the corresponding Noether charge is zero, that fermion
177 transforms as a singlet and does not participate in that gauge interaction.

178 Fermions are divided into quarks and leptons based on their transformations
179 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	μ_R	τ_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. $SU(2)_L$ gauge transformations allow one to go between rows and $SU(3)_C$ transformations allow one to go between columns in these fermion representations. [REWORD]

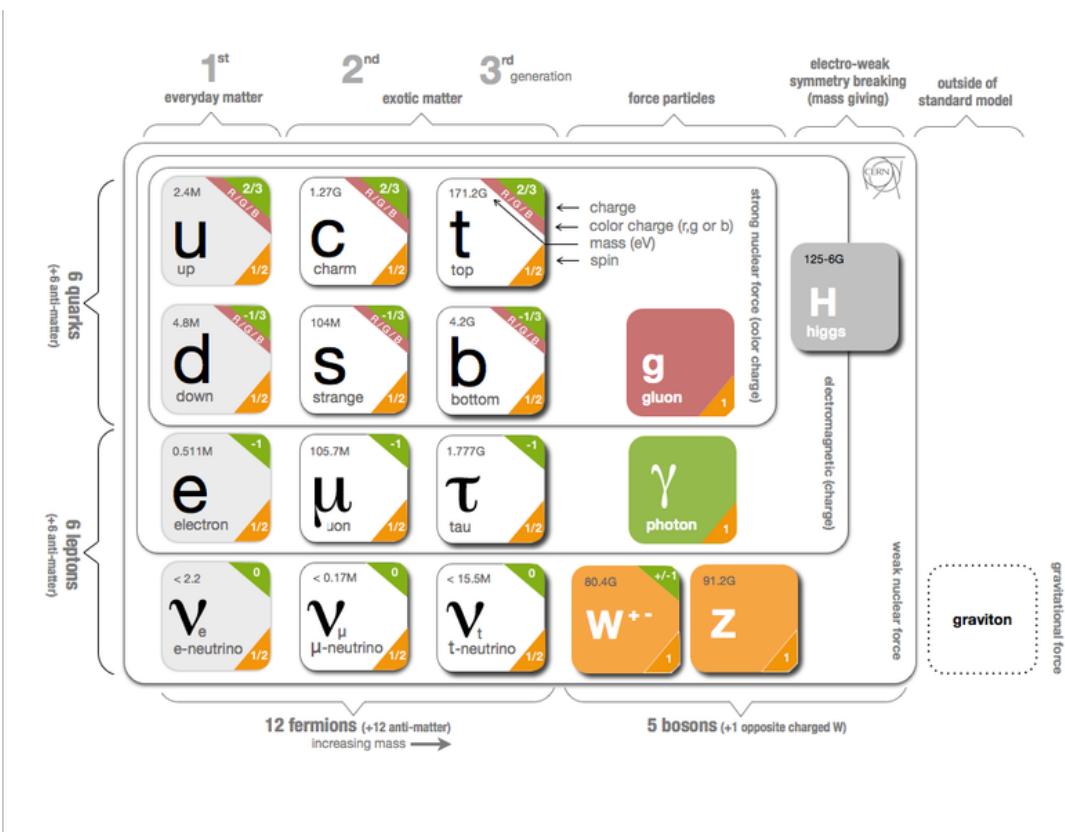


Figure 2.1: The particles of the Standard Model.

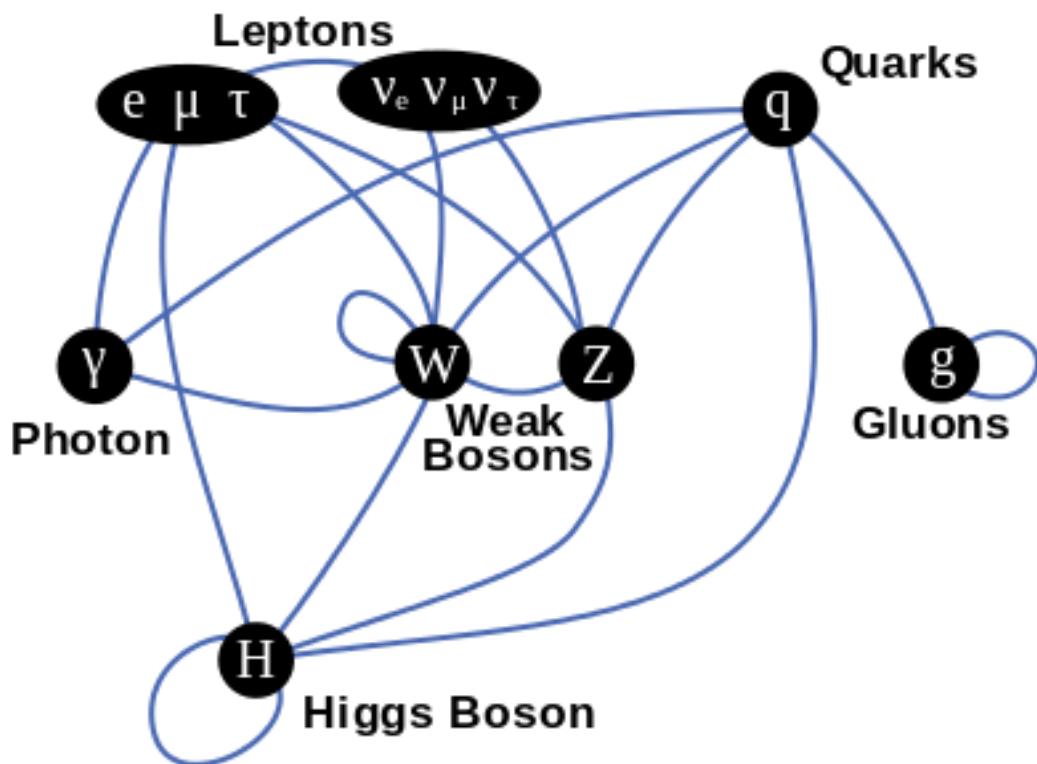


Figure 2.2: Summary of how Standard Model particles interact with other Standard Model particles.

¹⁸⁶ Now we can understand the SM Lagrangian density as a Yang-Mills theory
¹⁸⁷ with the gauge group: $SU(3)_C \times SU(2)_L \times U(1)_Y$ with an additional $SU(2)$ complex
¹⁸⁸ scalar Higgs field doublet that will be discussed later.

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

¹⁸⁹ Here several abstract spaces are being spanned:

- ¹⁹⁰ – a spans the three $SU(2)_L$ gauge fields with generators expanded in Pauli matrices, $T^\alpha = \frac{1}{2}\sigma^\alpha$
- ¹⁹¹ – α spans the eight $SU(3)_C$ gauge fields, with generators expanded in Gell-Mann matrices, $\tau^\alpha = \frac{1}{2}\lambda^\alpha$
- ¹⁹² – L/R represent left and right projections of Dirac fermion fields. The Strong interaction is not chiral, so $Q = L+R$

- 196 – μ and ν are four-vector indices
- 197 – i, j, k are summed over the three generations of SM particles.

198 **2.6 Higgs Mechanism**

199 The SM Lagrangian without the addition of a Higgs field does not allow for
200 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
201 not gauge invariant. By introducing the Higgs field, mass terms for these particles
202 may be included in a gauge invariant way. This field is a complex doublet with a
203 potential $V(\Phi)$:

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

204 $V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda |\Phi^\dagger \Phi|^2 \quad (2.25)$

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

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

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

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

212 where:

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

$$\begin{aligned} \text{213} \quad Z_\mu &\equiv \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2W_\mu^2 - g_1B_\mu) \end{aligned} \quad (2.29)$$

$$\begin{aligned} \text{214} \quad m_W &= \frac{vg_2}{\sqrt{2}} \end{aligned} \quad (2.30)$$

$$\begin{aligned} \text{215} \quad m_Z &= \frac{v}{\sqrt{2}}\sqrt{g_1^2 + g_2^2} \end{aligned} \quad (2.31)$$

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

218 2.7 Electroweak Theory

219 $SU(2)_L$ generates W^\pm, W^0 gauge bosons, which would be massless if $SU(2)_L$
220 was a perfect symmetry. These bosons are massive as this symmetry is broken.

221 The mass eigenstates, Z and γ given by:

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

222 Here θ_W is the Weinberg angle given by:

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

²²³ 2.8 Quantum ChromoDynamics

²²⁴ As mentioned earlier the Strong Force, which binds the proton together, is
²²⁵ mediated by gluons. Quantum Chromodynamics is the QFT which describes the
²²⁶ interactions of quarks and gluons via $SU(3)_C$ symmetry. QCD contains features
²²⁷ not present in Electroweak Interactions due to $SU(3)_C$ generators not commuting
²²⁸ (a.k.a. $SU(3)_C$ is a non-abelian group). For example, in QCD there is color
²²⁹ confinement and asymptotic freedom due to the structure constants being non-
²³⁰ zero. Requiring $SU(3)_C$ local gauge invariance implies:

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

²³¹ where $\alpha(x)$ is the local phase function, g_S is the strong coupling constant, and
²³² \hat{T} are the eight generators of $SU(3)$ (note $\hat{T}^a = \frac{1}{2}\lambda^a a$, where λ^a are the Gell-Mann
²³³ matrices). As the Gell-Mann matrices are 3x3, this means ψ has three degrees of
²³⁴ freedom under these $SU(3)$ rotations. So we represent ψ under $SU(3)$ rotations
²³⁵ as:

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

²³⁶ Consequently, particle fields transforming under $SU(3)$ rotations have three
²³⁷ components which physicists describe as color components (red, green, and blue).
²³⁸ A particle's corresponding antiparticle has the corresponding anticolor. This color
²³⁹ is the "charge" of QCD and is conserved under $SU(3)$ rotations. Combining colors,
²⁴⁰ color neutral states (e.g. red and antired, or red, green and blue) may be created.
²⁴¹ For the Free Dirac Lagrangian to remain invariant under $SU(3)$ transformations,
²⁴² we must again postulate a boson field that modifies the derivative. The gluon

243 field tensor is given by ($\alpha = 1, \dots, 8$):

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

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

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

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

251 Figure 2.3 shows the three dominant QCD interactions. Since gluons carry
252 color charge, they interact with one another. This does not occur in QED, as
253 photons do not have electric charge and therefore do not interact with each other.
254 In QED, a bare electron's effective charge is largest closest to the electron and
255 decreases as a function of distance. This is because the QED vacuum fills with
256 particle antiparticle pairs spontaneously, which screen the charge of the bare elec-
257 tron. The larger the distance from the electron, the smaller the effective charge
258 and therefore the weaker the force. So for a pair of electrons, as the distance
259 between them increases the repulsive force decreases and they may be observed
260 separately.

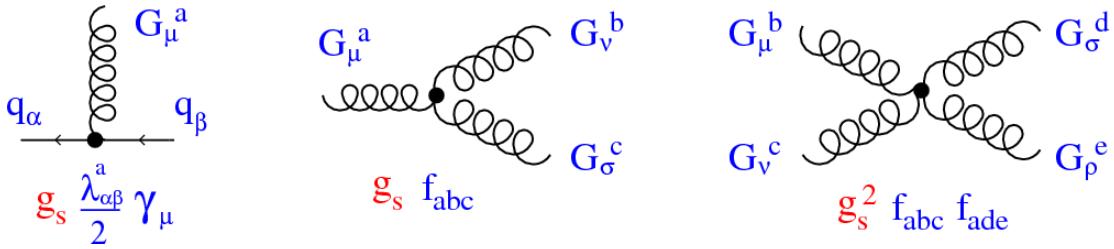


Figure 2.3: This figure shows the three dominant QCD interactions. From Ref. [16]

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

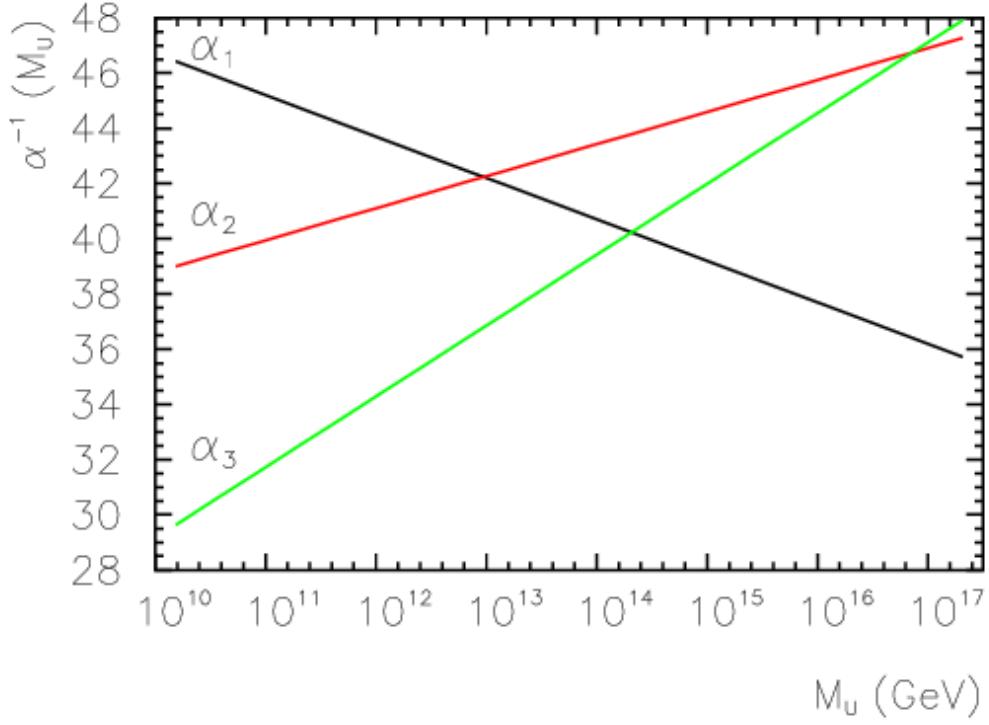


Figure 2.4: Strength of the U(1), SU(2), and SU(3) gauge couplings as a function of the energy scale of the interaction (Q). From Ref. [11]

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 (α) of a force varies with energies. (More deeply this amounts to incorporating renormalization and vacuum polarization in the boson propagators).

For QCD this is:

$$\alpha_S(x) = \frac{\alpha_S(\mu^2)}{1 + \beta_0 \alpha_S(\mu^2) \ln(Q^2/\mu^2)} \quad (2.38)$$

278

$$\beta_0 = \frac{11N_c - 2n_f}{12\pi} \quad (2.39)$$

279 where Q is the momentum of the the force is probed at, μ^2 is the renormalization scale.

281 As stated previously, quarks and gluons have not been observed in isolation.
 282 Instead they form bound colorless states. Hadronization is the process by which
 283 quarks and gluons form hadrons. The process of hadronization is still an active
 284 area of research. One qualitative description is show in Figure [Natasha add
 285 figure]. In this figure, as two quarks separate the color field between them is
 286 restricted to a tube with energy density of 1GeV/fm. As they separate further,
 287 the energy in the color field increases, until there is enough energy to produce
 288 $q\bar{q}$ pairs, which breaks the color field. This process repeats until quarks and
 289 antiquarks have low enough energy to form colorless hadrons. The resulting spray
 290 of hadrons is called a jet.

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

301 **Chapter 3**

302 **Standard Model Successes and**
303 **Limitations**

304 The Standard Model has consistently described much of reality to an extreme
305 degree of accuracy. It has predicted cross sections for strong and electroweak pro-
306 cesses that span over ten order of magnitude correctly [see Fig. 3.1] and contains
307 no known logical inconsistencies. Despite the strength and reality of the Stan-
308 dard Model, it still fails to describe aspects of reality and suffers from aesthetic
309 issues. To date, dark matter and energy comprise 95% of the universe, but are
310 not accounted for in the SM. Additionally, neutrinos are known to have mass but
311 are massless in the SM. There are mechanisms for introducing massive neutrinos
312 in the SM, but these mechanisms create hierarchy problems.

313 Possibly the most significant aesthetic issue is the hierarchy between the elec-
314 troweak and Planck scales. The electroweak scale is the scale of electroweak
315 symmetry breaking. The Planck scale is the scale where the gravitational force
316 is comparable in strength to the other forces. The Planck scale is where the SM
317 breaks down, as there is not an experimentally verified theory of quantum gravity,
318 and at this scale gravity cannot be ignored (like it can at the electro-weak scale).

319 These scales differ by ~ 30 orders of magnitude. Understanding the difference
320 in these energy scales, may help explain the weakness of gravity at electroweak
321 scales, and possibly a QFT for gravity. (NB: This hierarchy can also be framed in
322 terms of the corrections to the Higgs mass, which depend on the UV cutoff scale -
323 where the SM is suppose to break, which is taken at the Planck scale. This leads
324 the quantum corrections to the Higgs mass to force the Higgs mass to 10^{18} TeV.)

325 These stark contrasts in scales may indicate that a more fundamental theory
326 exists. It is hoped that such a theory would explain and motivate some of the ad-
327 hoc features of the SM. In particular, there currently are no experimentally verified
328 explanations of why there are three generations of fermions, the values of the 19
329 SM parameters (6 quark masses, 3 charged lepton masses, 3 gauge couplings,
330 Higgs parameters (μ^2, λ)), the structure of the fermion representations, etc.

Standard Model Production Cross Section Measurements

Status: July 2018

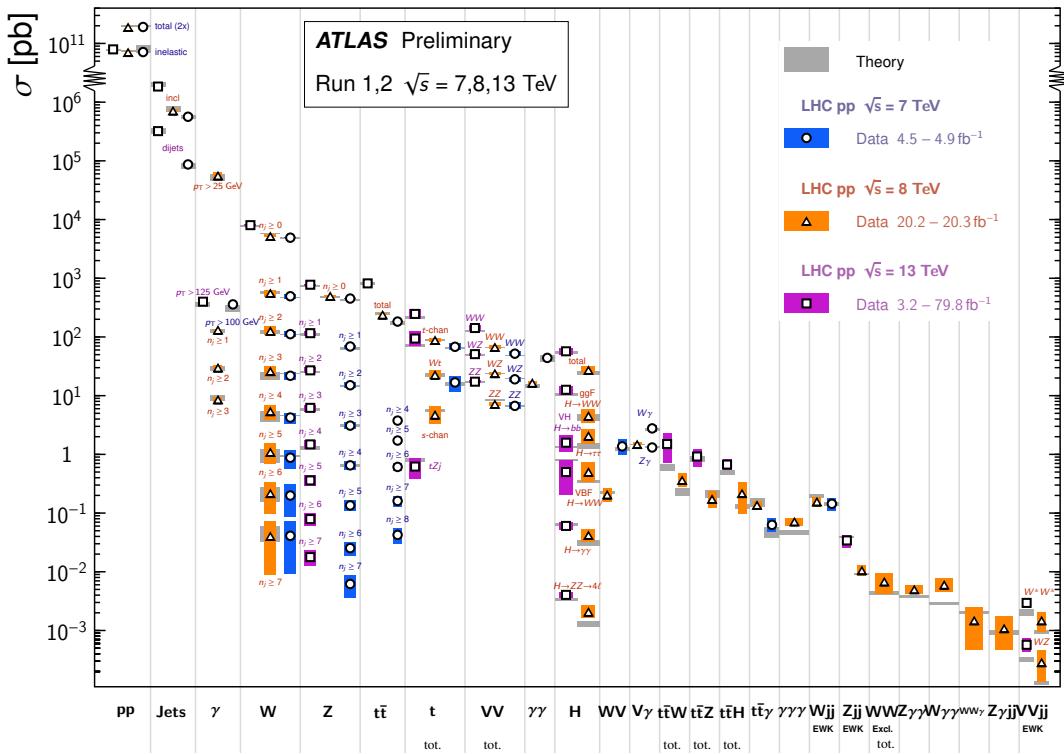


Figure 3.1: A comparison of cross section measurements at $\sqrt{s} = 7, 8, 13$ TeV from ATLAS compared to theoretical measurements. From Ref. [5]

³³¹ **Chapter 4**

³³² **New Physics Models with**

³³³ **Diboson Resonances**

³³⁴ **4.1 Randall Sundrum Bulk Model**

³³⁵ The electroweak-planck hierarchy may be explained by the existence of extra
³³⁶ dimensions, like the 5D Randall Sundrum Bulk Model ([17], [2]). In this model,
³³⁷ there is one extra warped spatial dimension, y , with a metric:

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

³³⁸ where $e^{-k|y|}$ is the warp factor of the extra dimension, which is compactified on
³³⁹ a S^1/Z_2 orbifold (a.k.a. a circle where $y \rightarrow -y$). This can be visualized as every
³⁴⁰ point in space time having a line extending from it a distance L , representing
³⁴¹ this fifth dimension. At the end of this line is the Planck brane. This fourth
³⁴² spatial dimension separates two 4-D branes: Planck brane and TeV brane. We
³⁴³ live on the TeV brane, as shown in Figure 4.1. The Higgs field (and to a lesser
³⁴⁴ degree the top quark and graviton fields) is localized near the TeV Brane, while

345 the light fermion fields are localized more near the Planck brane. Fundamental
 346 parameters are set on the Planck brane. The warp factor may be scaled away from
 347 all dimensionless SM terms by field redefinitions. However, the only dimensionful
 348 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
 349 why gravity is so weak on the TeV brane. Also, by localizing the light fermion
 350 fields near the Planck brane and top and graviton fields near the TeV brane, the
 351 light quarks will have smaller masses.

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

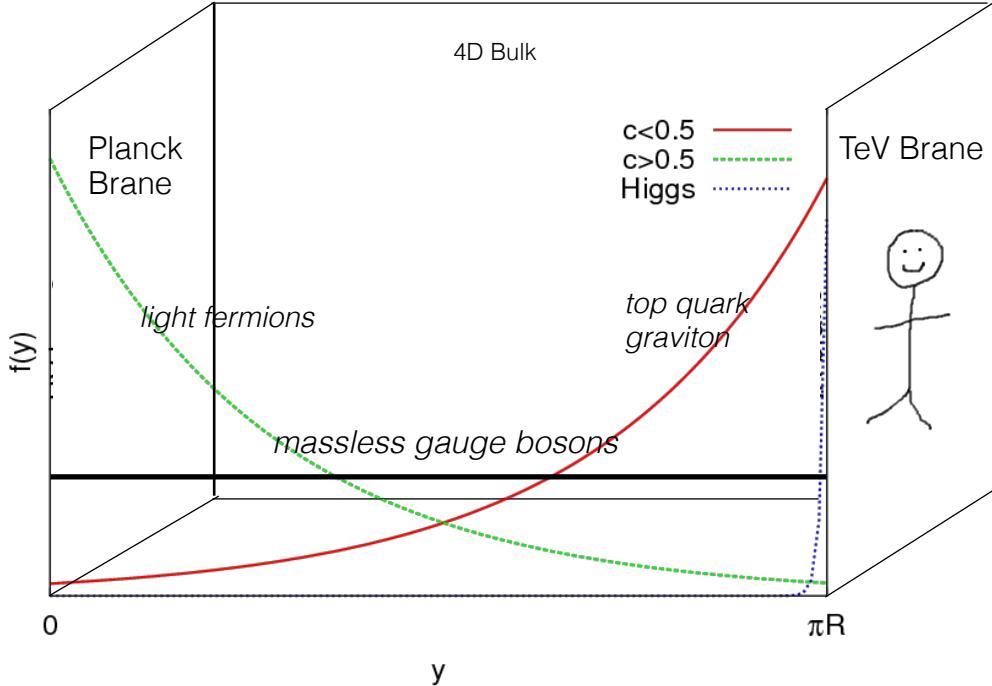


Figure 4.1: Cartoon of RS Bulk Model

³⁵⁶ 4.2 Extended Scalar Sector

³⁵⁷ A further striking asymmetry of the SM is the simplicity of the scalar sector in
³⁵⁸ comparison to the boson and fermion sectors. To date, the scalar sector has only
³⁵⁹ one member, the Higgs boson. Therefore, it is natural to posit an extension to the
³⁶⁰ scalar sector. From a theoretical standpoint this could also help generate baryon
³⁶¹ asymmetry through additional sources of CP violation. This analysis searches for
³⁶² a simple extension to the scalar sector as proposed in Ref. [18]. The extended
³⁶³ scalar sector includes a real Higgs singlet (S) and complex $SU(2)_L$ doublet (Φ)
³⁶⁴ (the SM Higgs), where mass eigenstates are mixtures of the fields. S has a vev of
³⁶⁵ v and Φ has a vev of x . This then gives a Lagrangian of:

$$\mathcal{L} \supset (D^\mu \Phi)^\dagger D_\mu \Phi + \partial^\mu S \partial_\mu S - m^2 \Phi^\dagger \Phi - \mu^2 S^2 + \lambda_1 (\Phi^\dagger \Phi)^2 + \lambda_2 S^4 + \lambda_3 \Phi^\dagger \Phi S^2 \quad (4.2)$$

³⁶⁶ The mass eigenstates of the scalar sector are then mixtures of S and Φ and
³⁶⁷ the free parameters of the theory are m_H , $\sin \alpha$, and $\tan \beta = v/x$. The fields are
³⁶⁸ then given by:

$$\Phi \equiv \begin{pmatrix} 0 \\ \frac{\tilde{h}+v}{\sqrt{2}} \end{pmatrix} \quad (4.3)$$

$$S \equiv \frac{h' + x}{\sqrt{2}} \quad (4.4)$$

³⁷⁰ Diagonalizing the mass matrix leads to the mass eigenstates h (discovered
³⁷¹ Higgs boson) and H (the physical particles):

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \quad (4.5)$$

³⁷² This suppressed h and H production and SM H couplings:

$$BR_{H \rightarrow SM} = \sin^2 \alpha \times \frac{\Gamma_{SM, H \rightarrow SM}}{\Gamma_{tot}} \quad (4.6)$$

³⁷³ Moreover, in the case that $m_H > m_h$, $H \rightarrow hh$ is possible. This further suppresses

³⁷⁴ $H \rightarrow VV/ff$. This search is most sensitive to $H \rightarrow WW$.

³⁷⁵ 4.3 Simple Standard Model Extensions

³⁷⁶ The RS Bulk model is motivated by resolving the hierarchy problem. Extending the Scalar sector is a natural space to look for new physics due to the complexity of fermion and boson groups. There are many other interesting and well motivated frameworks, but there is a lack of completely predictive models, due to model flexibility (free parameters). Therefore it is hard for experimentalists to know which theories to search for in data. However, as seen in [15], a "Simplified Model" approach may be taken. In the search for reasonably narrow width particles, as in this search, the search is not sensitive to all the details and free parameters of the theory. Generally such searches are only sensitive to the resonance mass and its interactions. Therefore, a theory's Lagrangian may be reduced to only retain this information (mass parameters and couplings). Experimental results using this framework may then be reinterpreted in a given theory.

³⁸⁸ In the simplified approach, the new resonance searched for is represented as a real vector field in the adjoint representation of $SU(2)_L$ with vanishing hypercharge. This results in one neutral and two charged bosons. Defined as:

$$V^\pm = \frac{V_\mu^1 \mp iV_\mu^2}{\sqrt{2}} \quad (4.7)$$

³⁹¹

$$V_\mu^0 = V_\mu^3 \quad (4.8)$$

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

$$\mathcal{L} \supset -\frac{1}{4}D_{[\mu}V_{\nu]}^a D^{[\mu}V^{\nu]}{}^a + \frac{m_V^2}{2}V_\mu^a V^{a\mu} + ig_V c_H V_\mu^a H^\dagger \tau^a \overset{\leftrightarrow}{D}{}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a J_F^{\mu a} \quad (4.9)$$

393 In order the terms represent: the kinetic, V mass, Higgs- V interaction, and
394 V -left-handed fermion interaction terms. Phenomenologically the three physical
395 particles this predicts are degenerate, where V couples most strongly to VV , via
396 the g_V coupling factor. The dominant production modes are DY and VBF.

397 Two versions of HVT are considered, Model A and B. Model A is a weakly
398 coupled model where $g_V \sim 1$, like the extended gauge symmetry discussed in Ref
399 . [16]. Model B is a strongly coupled model, where $1 < g_V < 4\pi$. The width
400 of the resonance grows with g_V so for this narrow resonance search only g_V is
401 chosen to be less than 6 (so $\Gamma/M < 10\%$). More precisely, the coupling of these
402 resonances to fermions scales as $g_f = g^2 c_F/g_V$, where g is the SM $SU(2)_L$ gauge
403 coupling and c_F is the free parameter (expected to be of order 1 for Model A and
404 B). This then means that for Model B the coupling to fermions is more suppressed
405 than for Model A, leading to a smaller DY production rate and BR to fermionic
406 final states. The coupling of V to SM bosons scales as $g_H = g_V c_H$, where c_H is
407 a free parameter on the order of 1 for Model A and B. So for small values of g_V
408 (i.e. Model A - weakly coupled theories) the BR to gauge bosons is smaller than
409 for Model B. So weakly coupled vectors have large production cross sections and
410 decay prominently to leptons or jets, while strongly coupled vectors are produced
411 less and decay predominantly to gauge bosons.

412 Vectors in Model A and B are generally produced via quark-anti-quark annihi-
413 lation. The more rare production via vector-boson-fusion is considered by setting
414 $g_H = 1$ and $g_F = 0$. In Model B diboson final states are enhanced as stated
415 previously due to g_H and moreover the BR to WZ , WH , WW , and ZH are the

⁴¹⁶ same.

⁴¹⁷ In summary, V couples most strongly to left-handed fermions and VV dependent on g_V .

419

Part III

420

Experimental Setup

421 **Chapter 5**

422 **LHC**

423 The Large Hadron Collider (LHC) is the highest-energy particle collider in the
424 world. It was designed to expand the frontier of high energy particle collisions in
425 energy and luminosity. This enables LHC experiments to test the Standard Model
426 and search for new physics at higher energies than tested with previous colliders.
427 Collisions at higher energies not only produce more massive particles but also
428 more weakly interacting particles. Fig. 5.1 shows production cross sections for
429 various processes at hadron colliders. The rate for electroweak physics processes
430 including W and Z scale with the center-of-momentum energy, \sqrt{s} .

431 The LHC consists of a 26.7 km (17 miles) ring, approximately 100 m un-
432 derground, outside Geneva, Switzerland. Counter-circulating proton (and occa-
433 sionally heavy ions) beams collide inside four experiments along the beam line:
434 ATLAS, CMS, LHCb, ALICE. ATLAS and CMS are general purpose detectors de-
435 signed to explore the high energy frontier. LHCb is designed to study the physics
436 of b -quarks. ALICE specializes in studying heavy ion collisions.

proton - (anti)proton cross sections

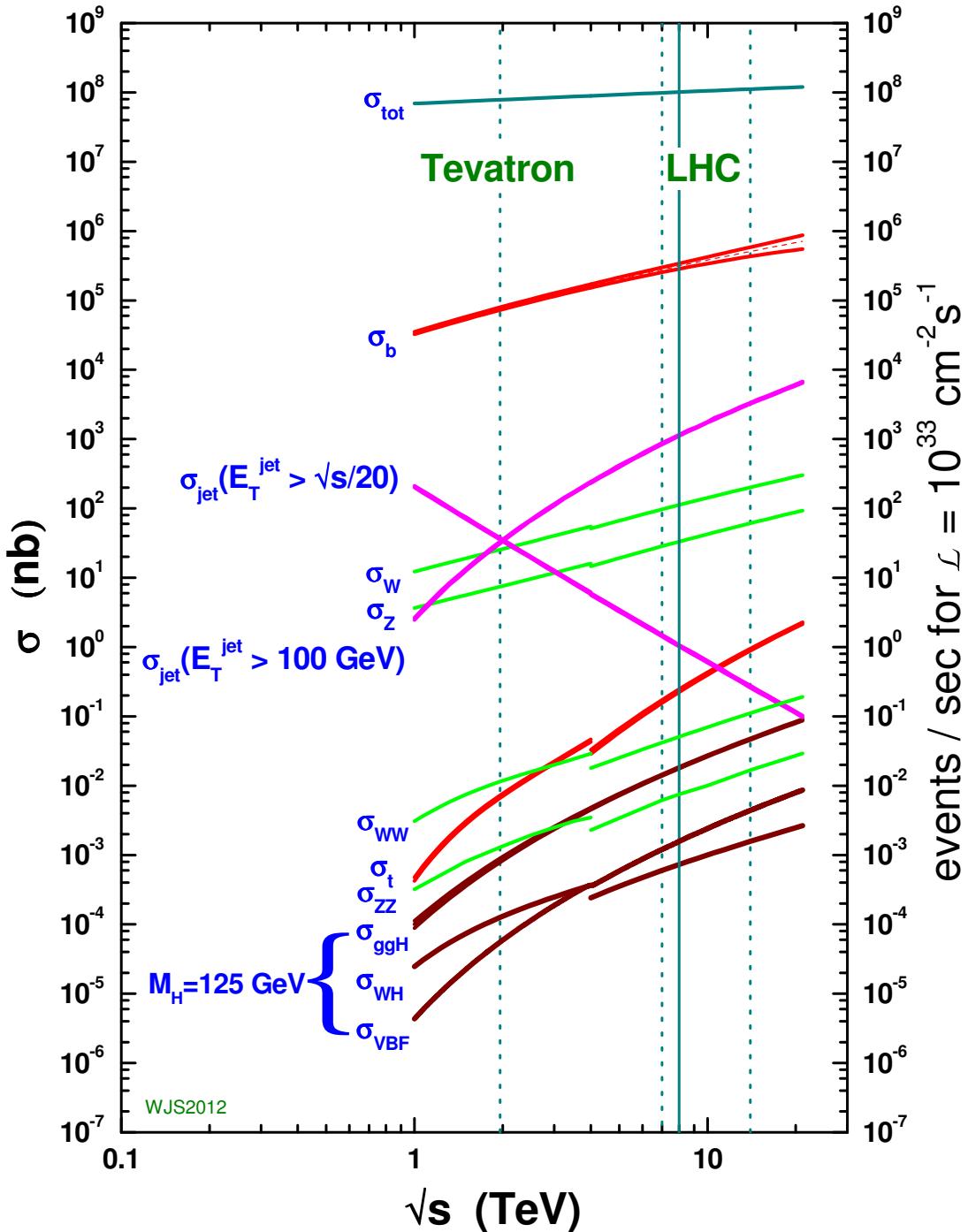


Figure 5.1: Scaling of cross sections with \sqrt{s} . Natasha: write more here

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

446 **5.1 LHC Layout and Design**

447 The layout of the LHC is shown in Figure 5.2. The red and blue lines in the
448 figure represent the counter-circulating proton beams. The LHC is divided into
449 eight octants. Octant 4 contains the RF cavities that accelerate the protons and
450 octant 6 contains the beam dump system. Octants 3 and 7 house the collimation
451 systems for beam cleaning. The beams collide inside the four aforementioned
452 experiments. Each octant contains a curved and straight section. The LHC
453 magnets are built with NbTi superconductors cooled with super-fluid Helium to
454 2K, creating a 8.3T magnetic field to bend the proton beams.

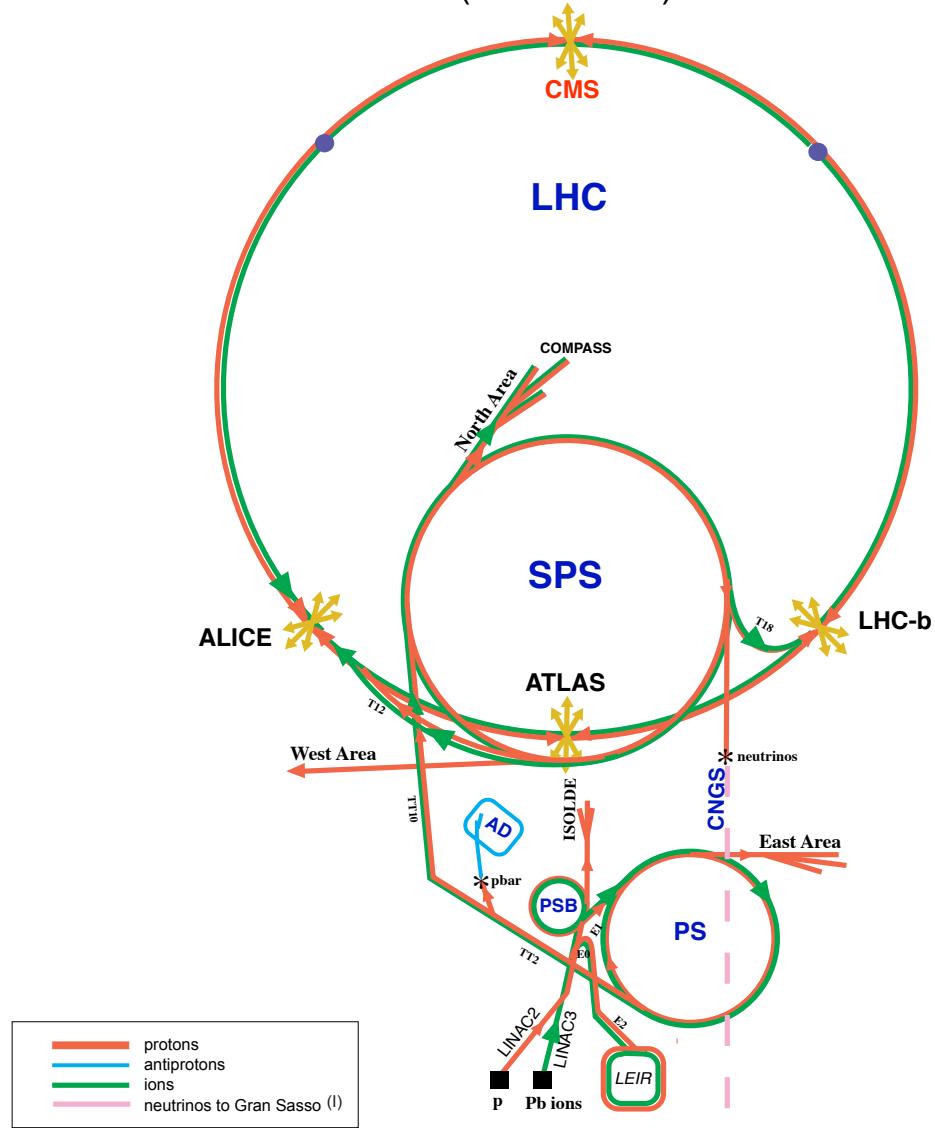


Figure 5.2: LHC Layout. Natasha write more

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

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

CERN Accelerators (not to scale)



LHC: Large Hadron Collider

SPS: Super Proton Synchrotron

AD: Antiproton Decelerator

ISOLDE: Isotope Separator OnLine DEvice

PSB: Proton Synchrotron Booster

PS: Proton Synchrotron

LINAC: LINear ACcelerator

LEIR: Low Energy Ion Ring

CNGS: Cern Neutrinos to Gran Sasso

Rudolf LEY, PS Division, CERN, 02.09.96
Revised and adapted by Antonella Del Rosso, ETT Div.,
in collaboration with B. Desforges, SL Div., and
D. Manglunki, PS Div. CERN, 23.05.01

Figure 5.3: LHC Accelerator. Natasha write more

466 As many new physics models predict cross-sections below the weak scale it was
467 important to design the LHC to be capable of collecting enough data, by running
468 in high luminosity conditions. The machine luminosity depends only on beam
469 parameters:

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

470 where N_p is the number of protons per bunch, f is the bunch crossing frequency,
471 ϵ is the transverse beam emittance, β^* is the amplitude function at the collision
472 point, and F is the geometric luminosity reduction factor due to the beams crossing
473 at an angle (rather than head-on).

474 **Chapter 6**

475 **The ATLAS Detector**

476 The ATLAS detector measures the position, momentum and energy of parti-
477 cles produced in the proton collisions by using magnetic fields, silicon detectors,
478 sampling calorimeters, and gaseous wire detectors. It is located approximately
479 100 m underground at Point-1 around the LHC beam line and weighs 7000 metric
480 tons. The detector is 46 m long, 25 m high, 25 m wide as shown in Figure 6.2.
481 The detector can be divided into three subsystems: the Inner Detector (ID), the
482 Calorimeters, and the Muon Spectrometer (MS). Figure 6.3 shows an overview of
483 how different particles interact in the detector.

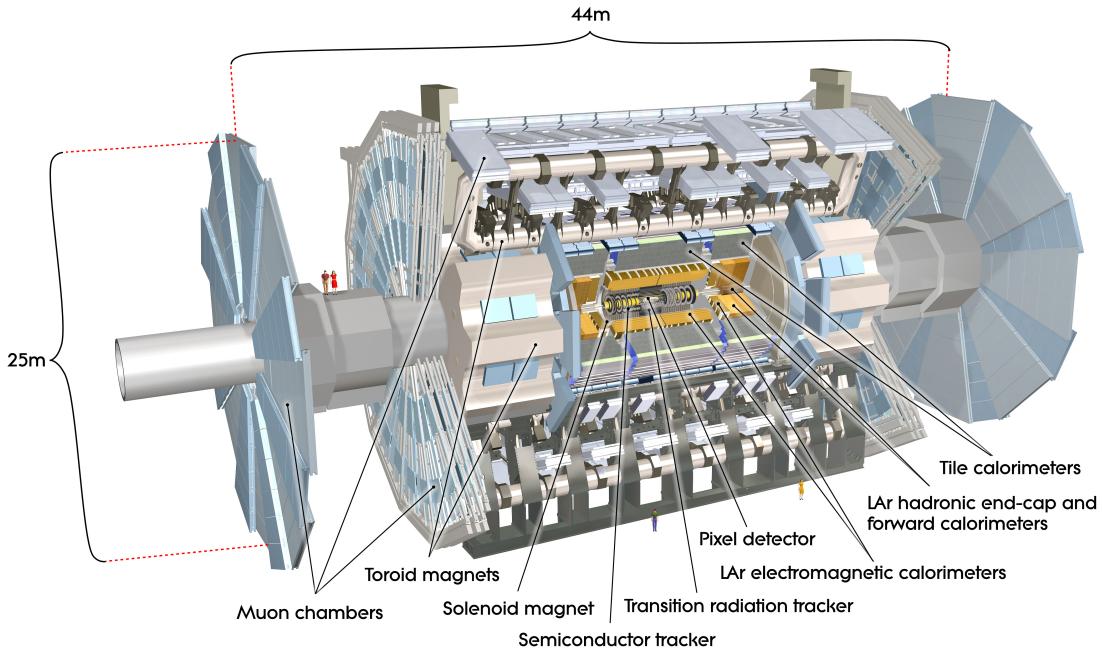


Figure 6.1: Big picture layout of ATLAS detector. Natasha: write more

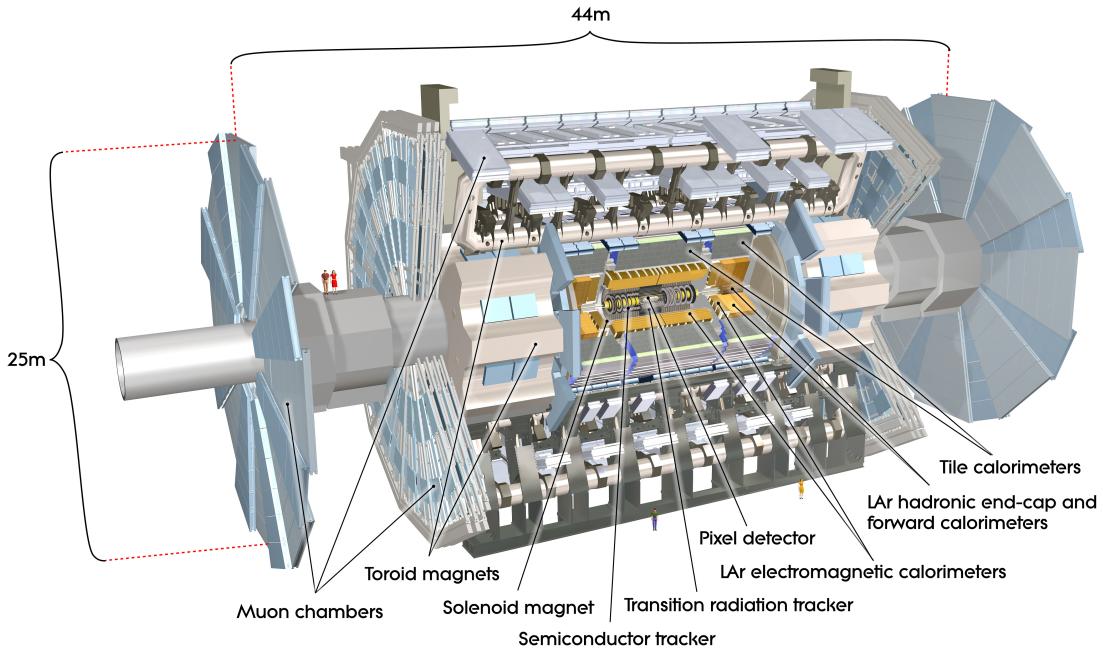


Figure 6.2: Big picture layout of ATLAS detector. Natasha: write more

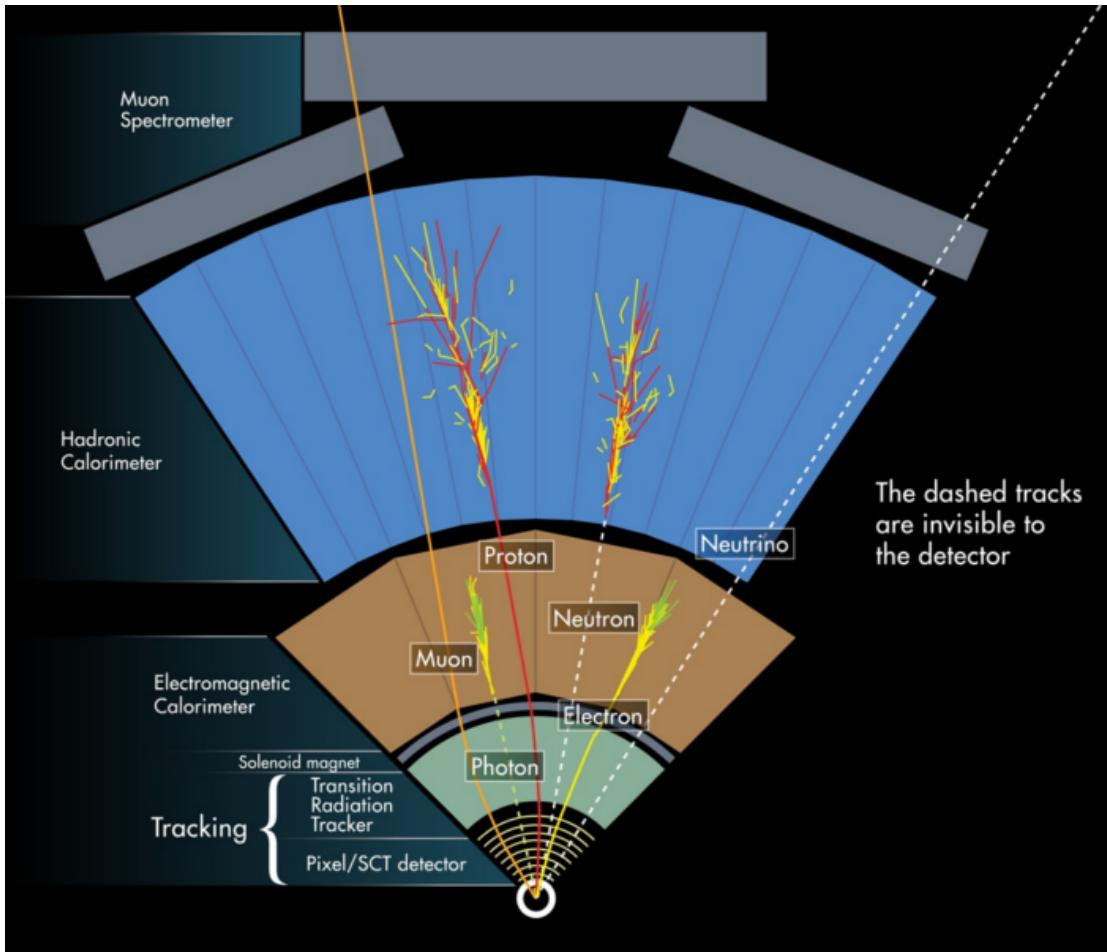


Figure 6.3: A simplified schematic of how different particles interact and are detected within ATLAS.

484 6.1 Coordinate System

485 The trajectory of particles within ATLAS is measured relative to the nominal
 486 interaction point. The z -axis points along the beam line, such that when the
 487 LHC is viewed from above, the counter-clockwise circulating beam points along
 488 the positive- z direction. The $x - y$ plane is transverse to the beam line, with the
 489 positive x -axis pointing towards the center of the LHC ring. The positive y -axis
 490 points vertically upward. The azimuthal angle, ϕ , is the angular distance about

491 the z -axis, with $\phi = 0$ along the x -axis. The polar angle from the z -axis is denoted
492 as θ . However, this quantity is not Lorentz invariant, like rapidity, $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$,
493 where E is the energy of the particle considered, and p_z , is it's momentum along
494 the z -axis. Pseudo-rapidity is preferred as $\Delta\eta$ is invariant under boosts along z
495 and particle production is approximately invariant under η . For massless particles,
496 rapidity and a related quantity, pseudorapidity, are the identical. The pseudora-
497 pidity is defined as: $\eta = -\ln \tan(\frac{\theta}{2})$. This quantity is preferred as it is purely a
498 geometric quantity, independent of particle energy. Angular separation between
499 particles in ATLAS are given by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. The distance from the
500 beamline is given by $r = \sqrt{x^2 + y^2}$

501 6.2 Inner Detector

502 The Inner Detector (ID) was designed to identify and reconstruct vertices,
503 distinguish pions from electrons, and measure the momentum of charged particles.
504 The ID uses three different technologies for particle reconstruction: the Pixel
505 Detector, Semiconductor Tracker (SCT), and the Transition Radiation Tracker
506 (TRT), shown in Figure 6.4 and 6.5. The entire ID is immersed in a 2T solenoidal
507 magnetic field parallel to the $+z$ -axis, causing charged particles to bend in the
508 transverse-plane, allowing particle momentum measurements.

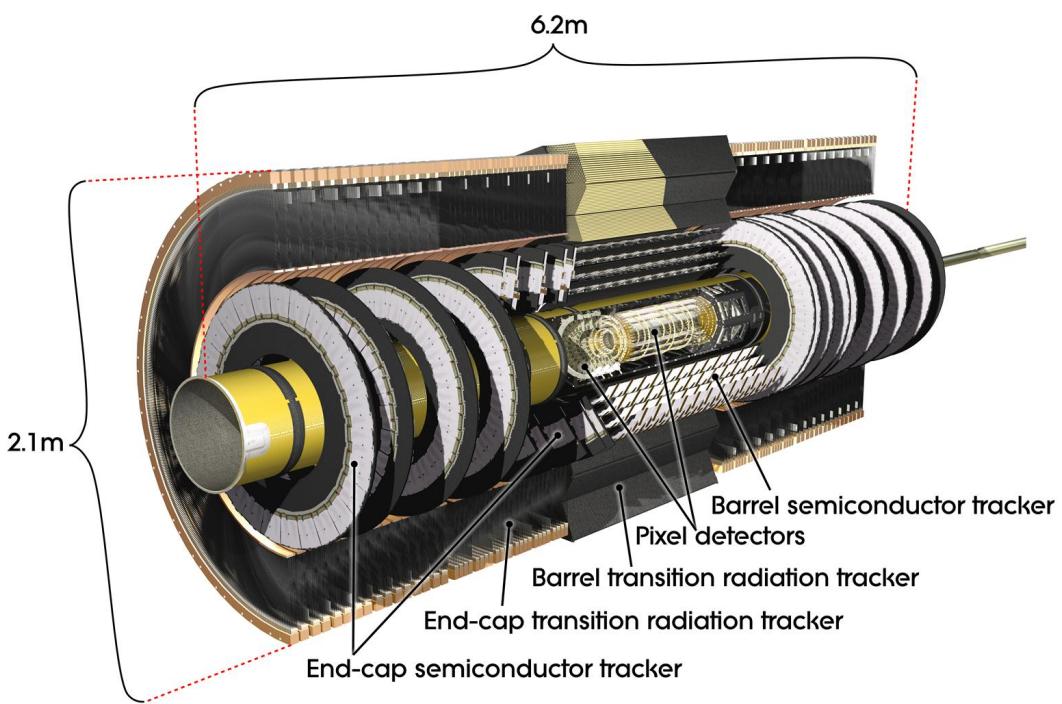


Figure 6.4: Layout of ATLAS Inner Detector

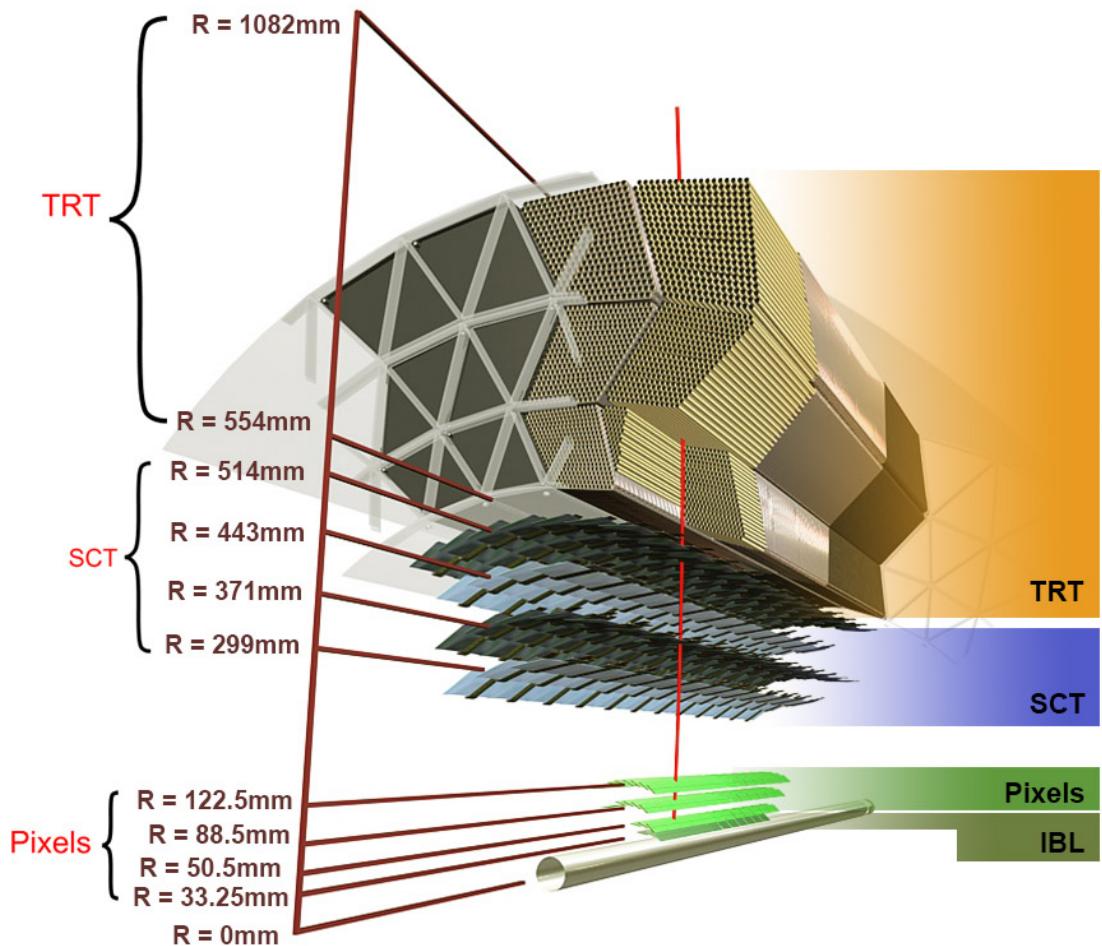


Figure 6.5: Layout of ATLAS ID Barrel System.

509 **6.2.1 Pixel Detector**

510 The pixel detector consists of four barrel layers between $r = 32.7$ and 122.5
511 mm, extending to $|z| = 400.5$ mm. The remaining detectors are arranged in bar-
512 rels and forward and backward rings. The innermost pixel barrel, the Insertable
513 b-Layer (IBL), only extends to $|z| = 332$ mm. The pixel detectors closer to the
514 beam line (larger η values) consists of six parallel cylindrical rings of pixel de-
515 tectors transverse to the beam line. The entire pixel detector consists of 1744
516 identical pixel sensors each with 46080 readout channels, totaling about 80 mil-
517 lion individual pixels. Most of the pixel sensors are $50 \times 400 \mu\text{m}^2$. Each pixel has
518 a position resolution of $14 \mu\text{m}$ in ϕ and $115 \mu\text{m}$ in the z direction.

519 **6.2.2 Semiconductor Tracker**

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

529 **6.2.3 Transition Radiation Tracker**

530 The transition radiation tracker (TRT), enveloping the SCT, is a gaseous
531 straw-tube tracker mainly used for electron/pion track separation. Each straw
532 is 4 mm in diameter and filled with a Xe- CO_2 - O_2 gas mixture. An anode wire at

533 the center of the straw is held at ground potential, while the walls of the straw
534 are kept at -1.4kV. When a charged particle passing through the TRT ionizes the
535 gaseous mixture, the resulting ions form an avalanche on the anode wire with a
536 gain of $\sim 10^4$. The signal from the anode wire is then digitized and discriminated.
537 Signals passing a low threshold cutoff are used to distinguish noise from tracks.
538 Signals passing a high threshold cutoff are sensitive to transition radiation (TR).
539 TR photons are emitted when charged particles pass between materials with dif-
540 ferent dielectric constants. The probability that a charged particle with energy E
541 and mass m passing between two materials emits a TR photon in the keV range
542 is proportional to $\gamma = E/m$. In the TRT straws these often then convert via the
543 photoelectric effect, causing a large avalanche triggering the high-threshold. Since
544 electrons have a smaller mass than pions, electron tracks are more likely to trig-
545 ger the high threshold. This then provides discrimination between electrons and
546 charged hadrons.

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

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

557 6.3 Calorimeters

558 The ATLAS electromagnetic and hadronic calorimeters (EMC and HCAL,
559 respectively) absorb and measure the energy of high energy hadrons, photons,
560 and electrons with $|\eta| < 4.9$. Both systems use sampling calorimeters which
561 consist of alternating layers of dense absorbing and active layers. In the absorbing
562 layer particles interact and lose energy, creating showers. These showers are then
563 detected and measured in the active layer. The amount of charge measured in the
564 active material scales with the energy of the incident particle, and thus provides a
565 measurement of the particle's energy. An overview of the layout of the calorimeter
566 system is shown in Figure 6.6.

567 The EMC measures and contains the energy of electromagnetically interacting
568 particles. It consists of layered accordion-shaped Lead absorber plates and elec-
569 trodes immersed in liquid Argon with 170k channels.. Using accordion-shaped
570 electrode and absorbers ensures ϕ symmetry and coverage. The EMC is com-
571 posed of a barrel part ($|\eta| < 1.475$), two end-caps ($1.375 < |\eta| < 3.2$), and a
572 presampler ($|\eta| < 1.8$). The presampler, containing only liquid Argon, corrects
573 for upstream energy losses of electrons and photons. The EMC barrel is segmented
574 into three layers. The first layer has finest segmentation with readout cells ex-
575 tending $\Delta\eta \times \Delta\phi = 0.025/8 \times 0.1$. This provides a precise shower measurements
576 used to separate prompt photons from $\pi^0 \rightarrow \gamma\gamma$ decays. The second layer has
577 coarser segmentation and is approximately 16 radiation lengths long. A radiation
578 length is the average distance an electron travels before losing all but $1/e$ of its
579 energy to bremsstrahlung. The last layer is the most coarse and measures the tail
580 of the electromagnetic shower. A schematic of the ECAL is shown in Figure 6.7.

581 The hadronic calorimeter located outside the EMC and is used to contain
582 and measure the energy of hadronically interacting particles. It consists of a tile

583 calorimeter (TileCal), hadronic end-cap calorimeter (HEC), and liquid Argon for-
 584 ward calorimeter (FCAL). TileCal is located behind the LAr EMC and uses steel
 585 absorbers and liquid Argon as the active material. TileCal consists of three barrel
 586 layers in the central and forward regions, extending up to $|\eta| < 1.7$. Photons
 587 generated from hadronic interactions are collected via wavelength-shifting fibers
 588 connected to photomultiplier tubes, as shown in Figure 6.8. The HEC lies behind
 589 the EMC endcap wheels. It uses copper absorbers and liquid Argon as the active
 590 material and covers $1.5 < |\eta| < 3.2$. Finally, the FCAL covers $3.1 < |\eta| < 4.9$
 591 and consists of three modules all using liquid Argon as the active material. The
 592 first module uses copper absorber and was designed for electromagnetic measure-
 593 ments. The second and third modules consist of tungsten absorber and are used
 594 to measure the kinematics of hadronically interacting particles. A schematic of
 595 the HCAL is shown in Figure 6.8.

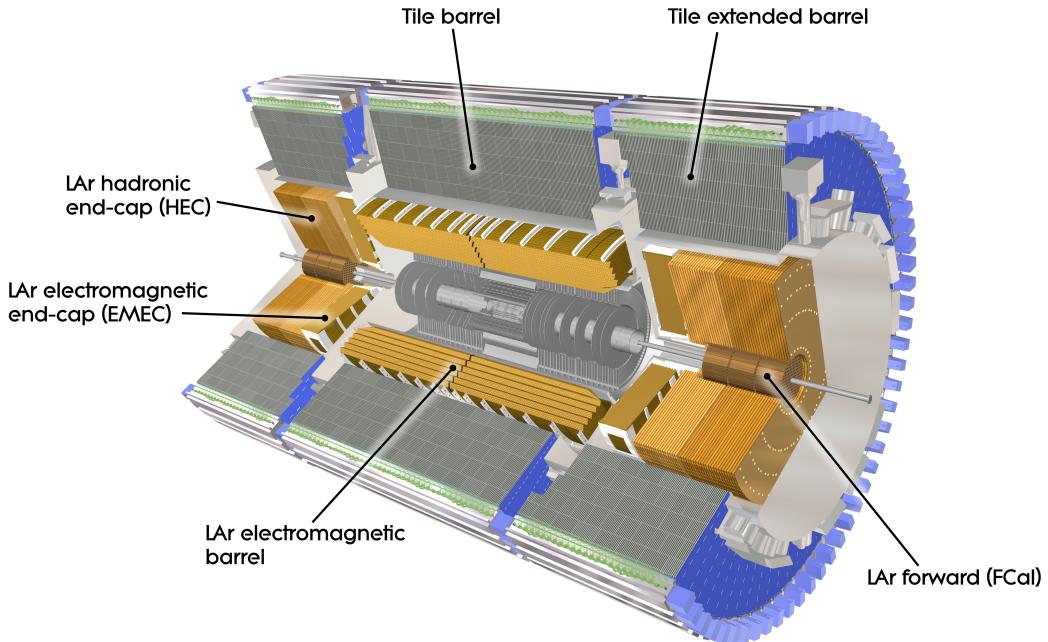


Figure 6.6: Overview of ATLAS electromagnetic and hadronic calorimeters.

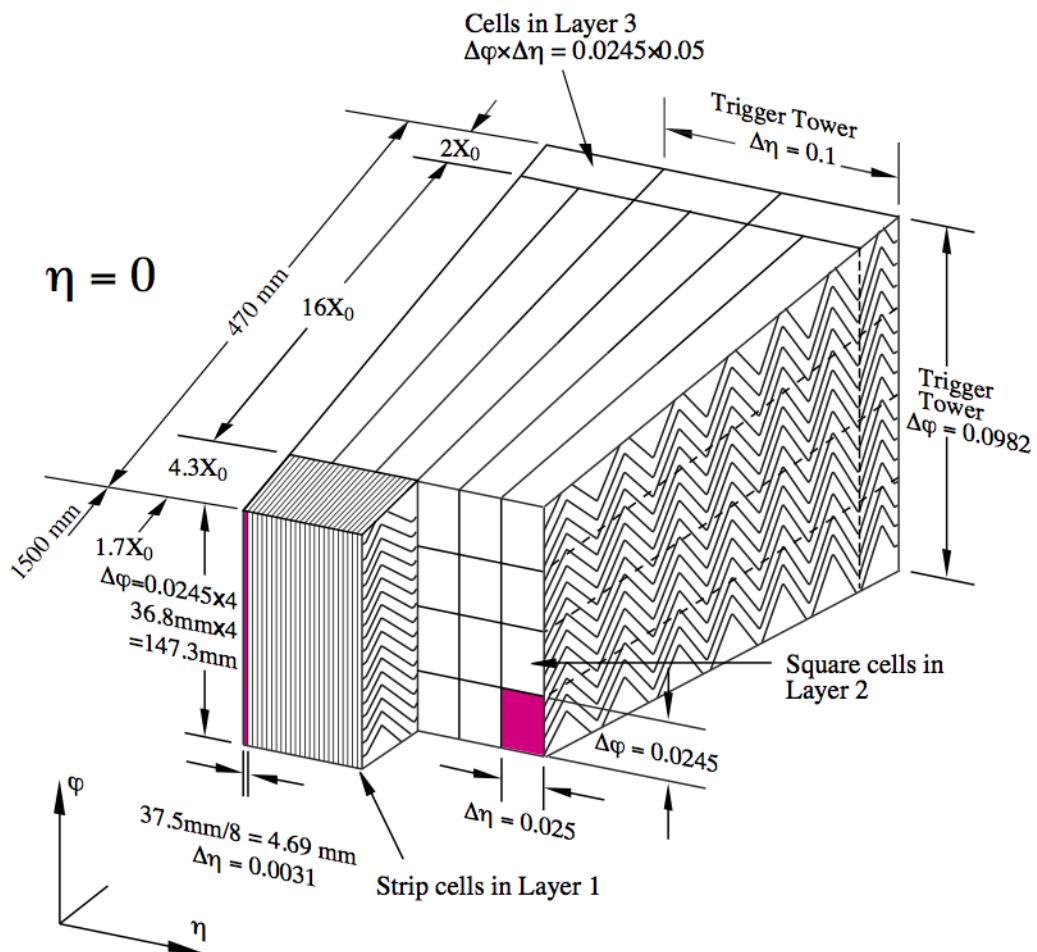


Figure 6.7: Schematic of ECAL.



Figure 6.8: Schematic of HCAL.

596 The energy resolution of the calorimeter subsystems are:

$$597 \quad \frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \text{ EMC}$$

$$598 \quad \frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \text{ hadronic barrel}$$

$$599 \quad \frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \text{ hadronic end-cap}$$

600 **6.4 Muon Spectrometer**

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

602 Muons with a $p_T > 4$ GeV are energetic enough to reach the MS. To measure the
603 momentum of these muons barrel and end-cap toroid magnets are used covering
604 $|\eta| < 1.4$ and $1.6 < |\eta| < 2.7$. For $1.4 < |\eta| < 1.6$, a combination of the barrel
605 and end-cap toroidal magnetic fields bend muon trajectories. The detector in the
606 barrel region form three concentric rings at $R = 5, 7.5, 10$ m and are segmented
607 in ϕ to accommodate the magnets. The end-cap region consists of three circular
608 planes perpendicular to z and located at $|z| = 7.4, 14, 21.5$ m from the interaction
609 region. An additional detector at $|z| = 10.8$ m covers the transition region between
610 the barrel and end-cap.

611 The MS readout consists of four subsystems: Monitored Drift Tubes (MDT),
612 Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC), and Thin Gap
613 Chambers (TGC). The first two subsystems are used primarily for measuring
614 muon track parameters, while the RPC and TGC subsystems are used for muon
615 triggering. A schematic of this system is shown in Figure 6.9.

616 The MDT subsystem consists of precision tracking chambers for $|\eta| < 2.7$,
617 except for the inner most end-cap layer ($2.0 < |\eta| < 2.7$), where CSCs are used.
618 The basic unit of MDT chambers are thin walled Aluminum tubes with a diameter
619 of 3 cm and length of 0.9-6.2 m. These tubes are filled with a mixture of Ar-CO₂
620 gas with a 50μm W-Rn wire running down the center of the tube, which is kept at

621 3080 V. Since the maximum drift time of these chambers is ~ 700 ns, they are not
622 used for triggering. MDT chambers consist of 3-4 layers of tubes mounted on a
623 rectangular support system, as seen in Figure 6.10, orientated along ϕ to measure
624 the coordinate in the bending plane of the magnetic field with a resolution of 35
625 μm .

626 The MDT subsystem can only handle hit rates below $150\text{Hz}/\text{cm}^2$. For this
627 reason, CSCs are used in the innermost end-cap layer where hit rates are larger.
628 CSCs can handle hit rates up to $1000\text{Hz}/\text{cm}^2$. CSC are multiwire proportional
629 chambers. These chambers are filled with a Ar- CO_2 gas mixture and evenly spaced
630 wires kept at 1900 V. These wires are orientated in the radial direction but not
631 read out. Instead on one side of the cathode are copper strips parallel to the wires,
632 measuring η , while on the other side of the cathode are strips parallel to the wires
633 measuring ϕ . The width between strips is approximately 1.5 mm providing a
634 resolution of 60 μm in the bending-plane and 5 mm in the non-bending plane.

635 Since the CSC and MDT systems do not have prompt timing signals, the RPC
636 and TGC systems are used for triggering. The RPC system is used in the barrel
637 region ($|\eta| < 1.05$). RPC consist of two parallel resistive plates separated by a 2
638 mm insulated spacer with 100 mm spacing kept at 9.8 kV, as shown in Figure 6.11.
639 A gaseous mixture of $\text{C}_2\text{H}_2\text{F}_4$, C_4H_{10} , and SF_6 fills the space between the two
640 plates. Metallic strips on the outer faces of the plates are used to read out signals
641 produced by the gas ionizing. The middle barrel layer consists of two layers of
642 RPCs on either side of the MDT layer and one layer on the outermost MDT
643 layer. Each layer contains two orthogonal sets of metallic strips providing η and
644 ϕ measurements. The timing resolution of RPCs is 1.5 ns, and therefore may be
645 used to identify bunch crossings.

646 Finally, the TGCs are used in the end-cap regions and are primarily used to

647 provide L1 trigger decisions and ϕ measurements. TGCs are multi-wire propor-
 648 tional chambers consisting of arrays of gold-coated tungsten wires placed between
 649 two cathode planes. These wires are separated by 1.8 mm and cathodes are 1.4 mm
 650 from the wires. Orthogonal to the wires, on the opposite side of the cathode plane
 651 are copper strips held at 2900 V. The chambers are filled with a mixture of CO_2
 652 and n-pentane gas, the latter acts as a quenching gas to prevent avalanches initi-
 653 ated by secondary γ -rays from the primary avalanche. Figure 6.12 is a schematic
 654 of a TGC. The timing resolution of TGCs is less than 25 ns and therefore they
 655 are used for bunch crossing measurements.

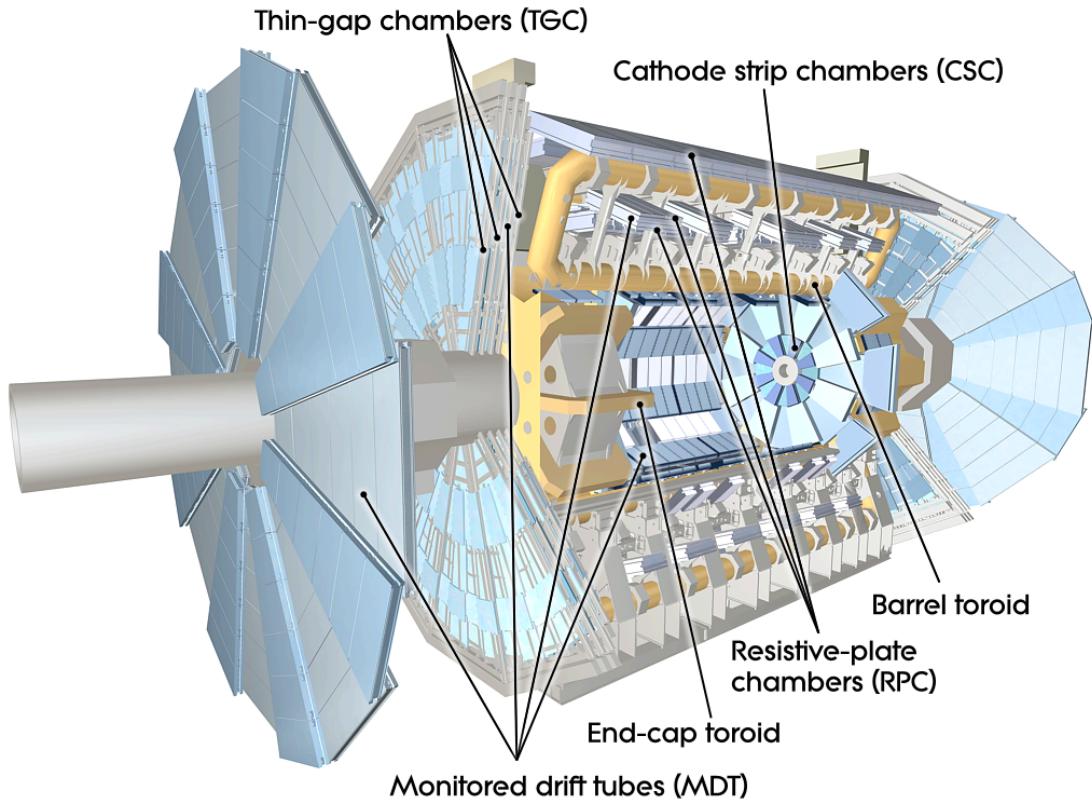


Figure 6.9: Schematic of Muon Spectrometer [cite G35]

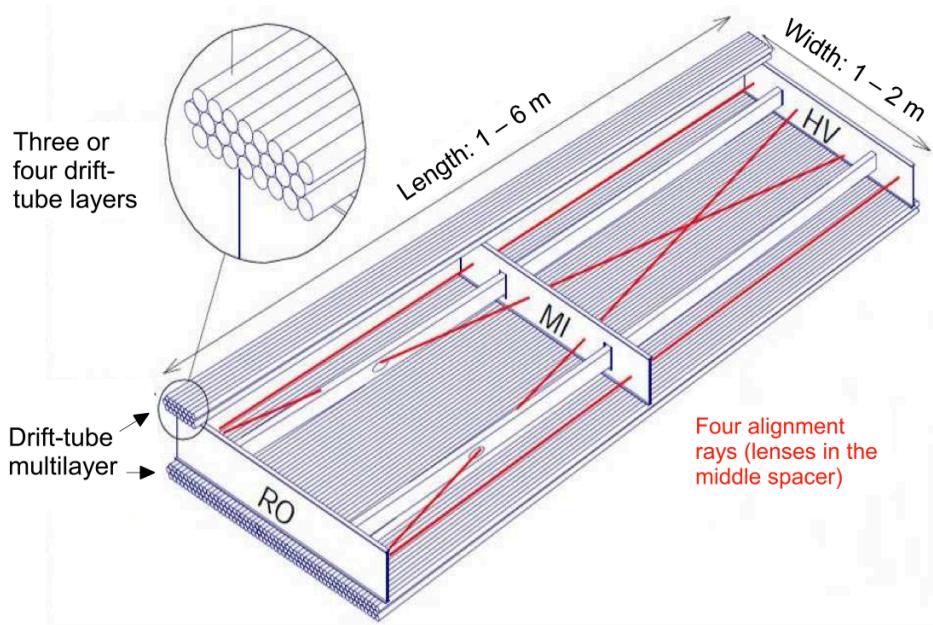


Figure 6.10: Schematic of MDT chamber. [cite G35]

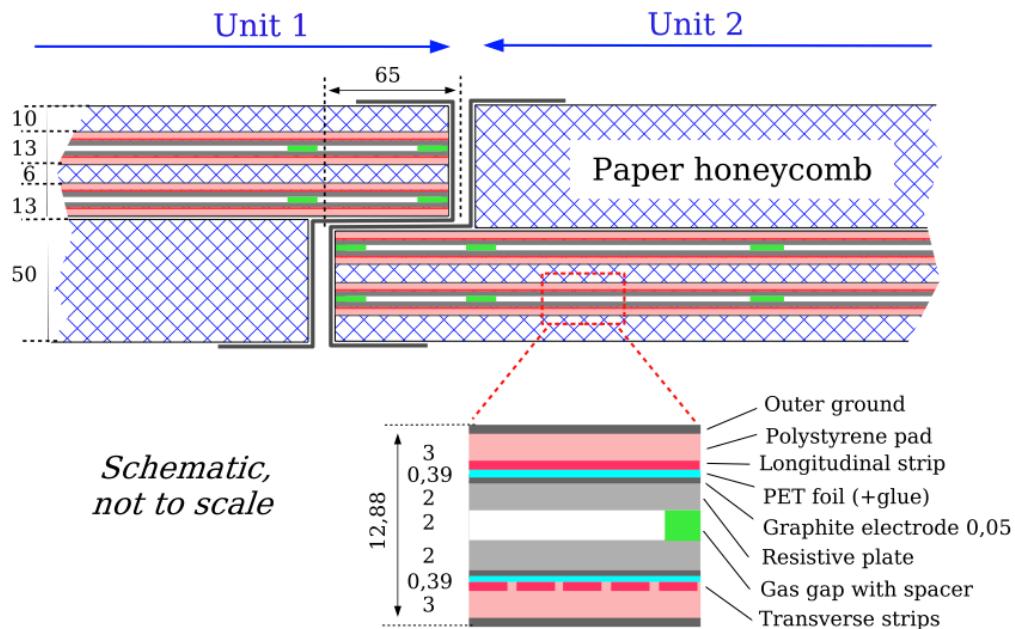


Figure 6.11: Schematic of RPC chamber, which is used for triggering in the central region of the detector [cite G35].

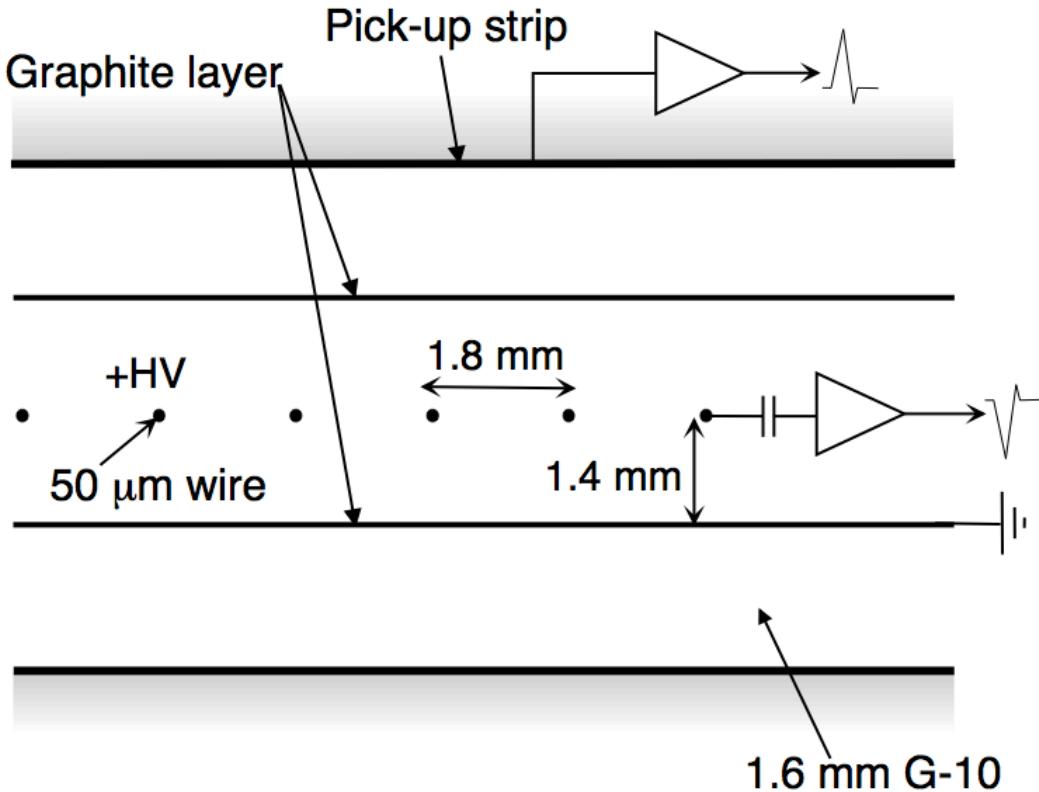


Figure 6.12: Schematic of TGC chamber, which is used for triggering in the muon end-cap region. [cite G35]

656 6.5 Magnet System

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

662 The layout of the magnet system is shown in Figure 6.13. The central solenoid
 663 consists of a single-layer Al-stabilized NbTi conductor coil wound inside an Al

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

669 The toroidal magnet system consists of a barrel and two end-cap toroidal
670 magnets used to create a magnetic field outside the calorimeters that is orientated
671 along ϕ . Each barrel toroid is 25.3 m long with an inner and outer diameter of 9.4
672 and 20.1 m and weighs 830 tonnes. Endcap toroids are 5 m long with an inner and
673 outer radius of 1.65 and 10.7 m. Both toroid systems use Al-stabilized Nb/Ti/Cu
674 conductors. The magnetic field strength in the barrel and endcap regions are 0.5
675 and 1 T, respectively.

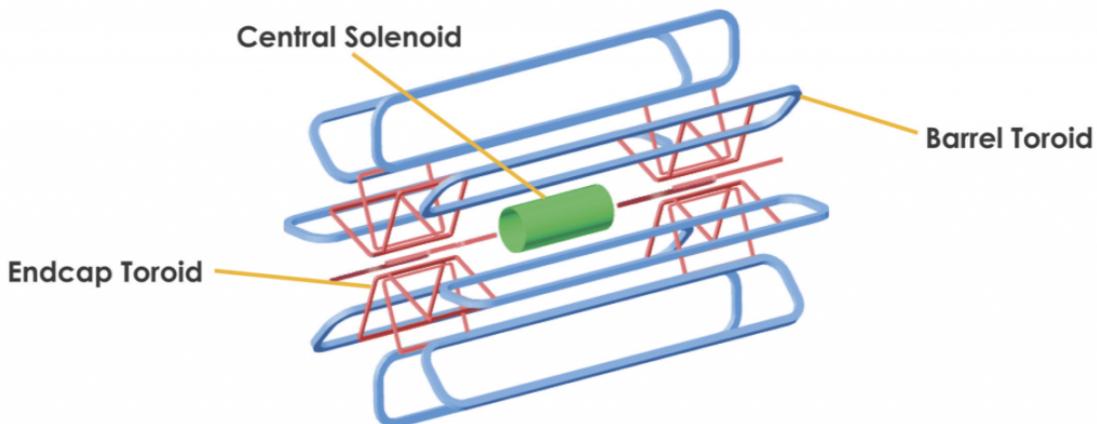


Figure 6.13: Layout of ATLAS magnet systems.

676 **6.6 Trigger System**

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

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

Part IV

690

Method

691

692 **Chapter 7**

693 **Dataset and Simulated Samples**

694 **7.1 Dataset**

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

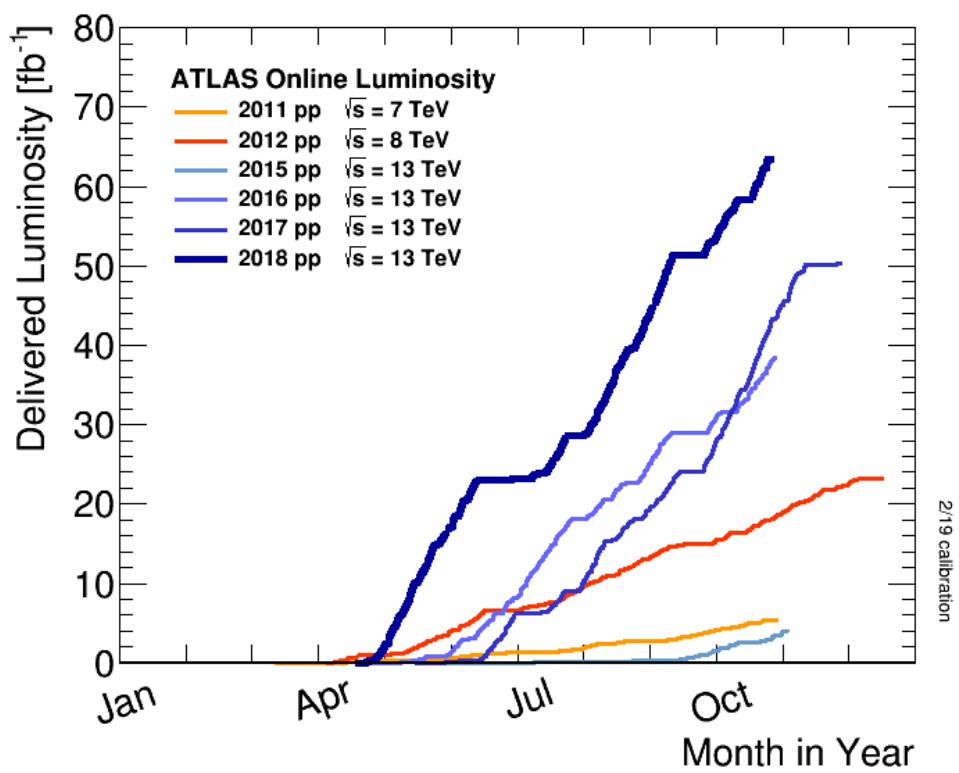


Figure 7.1: Integrated luminosity for data collected from ATLAS from 2011 - 2018



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

708 **7.2 Simulated Samples**

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

713 $W/Z+\text{jet}$ events are simulated using Sherpa 2.2.1 at NLO [cite [29]] and merged
714 with the Sherpa parton shower using the ME+PS@NLO prescription [12]. These
715 events are then normalized to NNLO cross sections. The $t\bar{t}$ and single-top back-
716 grounds are generated with Powheg-Box with NNPDF3.0NLO PDF sets in the
717 matrix element calculation [cite[35]]. For all processes, the parton shower, frag-
718 mentation, and underlying event are simulated using Pythia 8.320 with the A14
719 tune set[cite[ATL-PHYS-PUB-2014-02]]. Diboson processes are generated using
720 Sherpa 2.2.1.

721 Signal samples are simulated using MadGraph 5-2.2.2 [cite 42] and Pythia
722 8.186 with NNPDF230LO. RS Graviton samples are generated with $k/M_{PL}=1$.
723 HVT Model A (B) samples are simulated with $g_V = 1(3)$, as the difference in the
724 width of the samples is smaller than detector resolution. Model C is generated by
725 setting $g_H = 1$ and $g_f = 0$ to model VBF production of HVT bosons. Signals are
726 generated for masses between 300 GeV and 5 TeV.

727 **Chapter 8**

728 **Objects**

729 **8.1 Electrons**

730 Electrons are reconstructed from electromagnetic showers in the LAr EM
731 calorimeter. During reconstruction cells of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ are grouped
732 into 3×5 clusters. These clusters are then scanned for local maxima that seed
733 electron clusters. These clusters must then be matched to ID track from the PV.
734 This requirement minimizes non-prompt electron and fake electron backgrounds.
735 Electrons must pass identification and isolation requirements. Electron identifica-
736 tion (loose, medium, tight) classification is based on a multivariate discriminant
737 that identifies electrons using a likelihood based method. For this analysis tight
738 electrons are used. Electrons are also required to be isolated. The electrons are
739 considered isolated if the quotient of the sum of the transverse momentum (of
740 calorimeter energy deposits) in a cone around the electron of size $\Delta R = 0.2$ and
741 the transverse momentum of the electron to be less than $0.015 * p_T$ or 3.5 GeV,
742 whichever is smaller. This requirement rejects non-prompt photons and other
743 fake leptons. Electrons in this analysis are also required to have $p_T > 30$ GeV and
744 $|\eta| < 2.47$. Electrons are also required to have $p_T > 30$ GeV.

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

749 8.2 Muons

750 As muons traverse the entire detector, they are reconstructed from ID and MS
751 tracks. For this analysis the muon identification and isolation working points are
752 chosen to minimize the contributions from non-prompt muons. Towards this end,
753 the medium muon identification working point is used. For this working point,
754 two types of reconstructed muons are used: combined and extrapolated muons
755 (CB and ME, respectively). For CB muons, ID and MS tracks are reconstructed
756 independently and a combined track fit is performed by adding or removing MS
757 tracks to improve the fit quality. ME muons are reconstructed from only MS
758 tracks with hits in at least two layers, which ensures the track originates from the
759 PV. ME muons extend the acceptance for muon reconstruction outside the ID
760 from $2.5 < |\eta| < 2.7$. The medium identification working point uses CB and ME
761 tracks. CB tracks must have at least 3 hits in two MDT layers. ME tracks are
762 required to have at least three MDT/CSC hits. To further minimize contributions
763 from fake muons, the selected muons are required to be isolated from other tracks,
764 as muons from W, Z decays are often isolated from other particles. To insure the
765 selected muons are isolated, the scalar sum of the transverse momentum of tracks
766 in a cone of $\Delta R = 0.3$ compared to the transverse momentum of the muon must
767 be less than 0.06. Muons are also required to have $p_T > 30$ GeV.

768 Muons are calibrated using well-studied resonances $J/\Psi \rightarrow \mu\mu$ (low- p_T), $Z \rightarrow$
769 $\mu\mu$ (high- p_T). Figure 8.1 shows the combined muon p_T uncertainty from this

770 calibration. The total systematic uncertainty is less than 1% for all p_T ranges
 771 considered in this analysis.

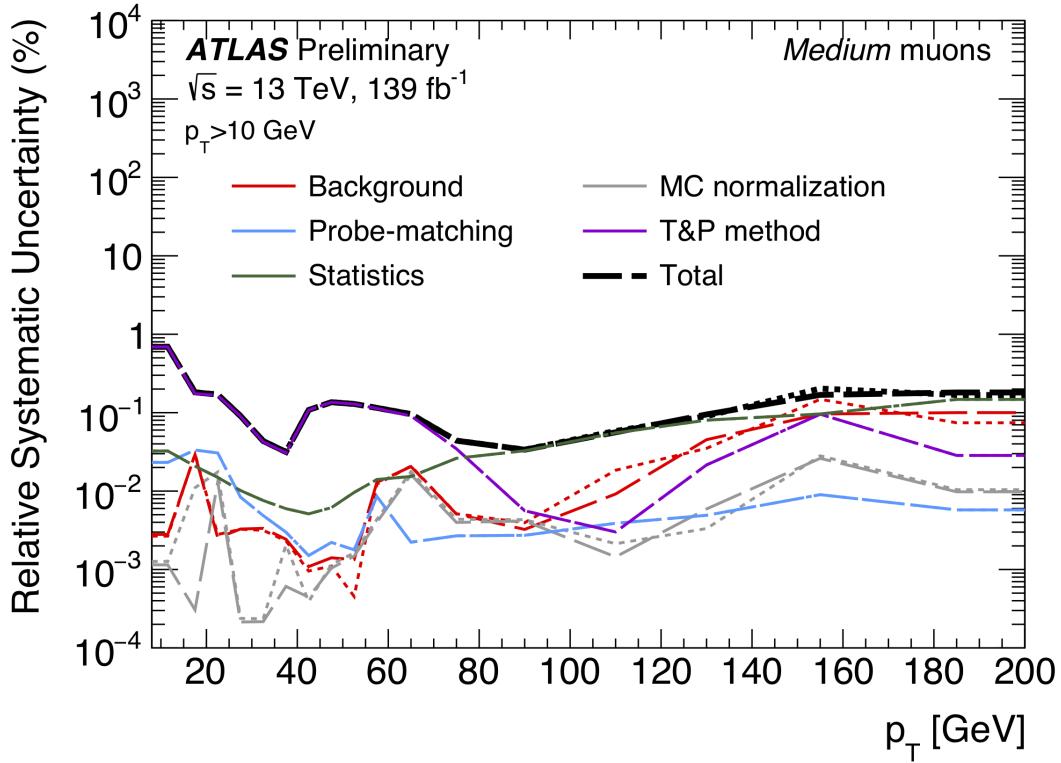


Figure 8.1: This figure shows the breakdown of the muon reconstruction efficiency scale factor measured in $Z \rightarrow \mu\mu$ as a function of p_T [4].

772 8.3 Jets

773 Three types of jets are used in this analysis: variable radius, small-R and
 774 large-R jets. Variable radius jets are used to reconstruct Z bosons decaying to
 775 two b -jets in the jet catchment area of large-R jet in the Merged regime. Small-R
 776 jets are used to reconstruct the hadronically decaying W/Z candidates in the
 777 resolved analysis and the forward jets from resonances produced through vector
 778 boson fusion. Large-R jets are used to reconstruct the hadronically decaying boson

779 in the merged regime.

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

786 MC-based corrections are then applied that are meant to transform the jet
787 energy and η back to truth level. Therefore, these corrections account for the
788 non-compensating nature of the ATLAS calorimeters and inhomogeneity of the
789 detector. Following this, the Global Sequential Calibration is applied that re-
790 duces flavor dependence of jet calibrations and accounts for energy leakage of jets
791 outside the calorimeters. Finally, in-situ corrections are applied that account for
792 differences in jet responses between data and simulation ($\gamma/Z+jet$ and fake lep-
793 ton samples are used). These differences can be due to mismodelling of the hard
794 scatter event, pile-up, jet formation, etc.

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

797 - f_Q^{LAr} : fraction of energy of jet's LAr cells with poor signal shape

798 - f_Q^{HEC} : fraction of energy of jet's HEC cells with poor signal shape

799 - E_{neg} : sum of cells with negative energy

800 - f_{EM} : fraction of jet's energy deposited in EM calorimeter

801 - f_{HEC} : fraction of jet's energy deposited in HEC calorimeter

802 - f_{max} : maximum energy fraction in any single calorimeter layer

803 - f_{ch} : ratio of the scalar sum of the p_T of a jet's charged tracks to the jet's p_T

804 Jets selected for the resolved analysis must pass one of the following criteria:

805 - $f_{HEC} > 0.5$ and $|f_Q^{HEC}| > 0.5$ and $\langle Q \rangle > 0.8$

806 - $|E_{neg}| > 60$ GeV

807 - $f_{EM} > 0.95$ and $f_Q^{LAr} > 0.8$ and $\langle Q \rangle > 0.8$ and $|\eta| < 2.8$

808 - $f_{max} > 0.99$ and $|\eta| < 2$

809 - $f_{EM} < 0.05$ and $f_{ch} < 0.05$ and $|\eta| < 2$

810 - $f_{EM} < 0.05$ and $|\eta| > 2$

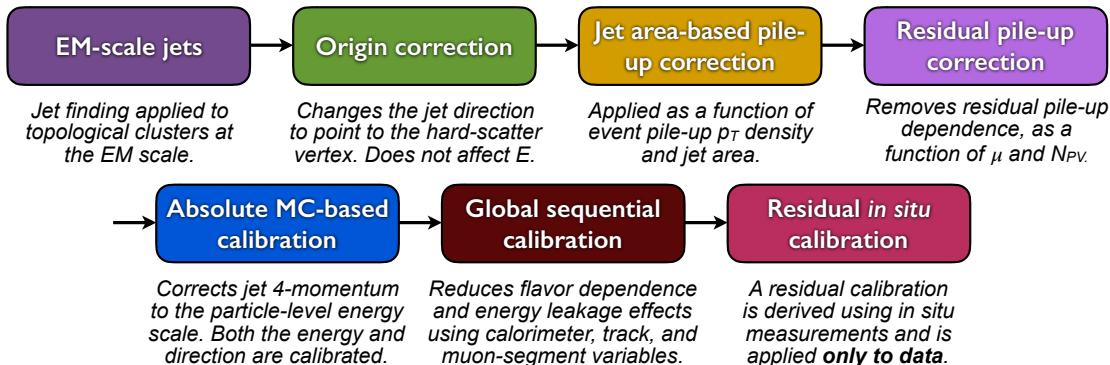


Figure 8.2: [7] This diagram shows the calibration stages for EM jets.

811 8.3.1 Small-R jets

812 Small-R jets are used to reconstruct the hadronically decaying W/Z candidate when the two resulting jets are well-separated in η - ϕ space. Small-R jets
813 are also used to identify forward jets from resonances produced through vector
814 boson fusion. Small-R jets are constructed from topologically connected clusters
815

816 of calorimeter cells (topoclusters), seeded from calorimeter cells with energy de-
817 posits significantly above the noise threshold. These cells are then used as inputs
818 to the $anti-k_t$ algorithm [14] with a radius parameter, $R = 0.4$.

819 Jets used in this analysis must have $p_T > 30$ GeV and $|\eta| < 2.5$. To further
820 reduce fake jets the jet-vertex-tagger (JVT) is used to reject pile-up jets [6]. The
821 JVT uses two track-based variables, corrJVF and R_{p_T} to calculate the likelihood
822 that the jet originated from the PV. The corrJVF compares the scalar sum of the
823 p_T of tracks associated with the jet and PV to the scalar sum of the p_T of tracks
824 associated with the jet. This variable also includes a correction that reduces the
825 dependency of corrJVF with the number of reconstructed vertices in the event.
826 The other discriminant, R_{p_T} , is given by the ratio of the scalar sum of the p_T of
827 tracks associated with the jet and PV to the p_T of the jet. Both of these variables
828 peak around zero for pileup jets, as these jets are unlikely to have tracks associated
829 with the PV. JVT cuts are applied to all jets with $p_T > 120$ GeV. Central jets
830 ($|\eta| < 2.4$) are required to have a $JVT > 0.59$ and forward jets ($2.4 < |\eta| < 2.5$)
831 are required to have $JVT > 0.11$.

832 8.3.2 Large-R jets

833 Large-R ($R = 1.0$) jets are used to reconstruct the hadronically decaying W/Z
834 candidate when the resulting jets are not well-separated in $\eta\text{-}\phi$ space, and overlap
835 forming one large-R jet. Track-Calorimeter Clusters (TCCs) are used to reconstruct these
836 jets [cite ANA 50]. These jets are constructed using a pseudo particle flow method
837 using ID tracks matched to calorimeter clusters. The angular resolution of the
838 calorimeter degrades sharply with jet p_T , but the jet energy resolution improves.
839 The tracker has excellent angular resolution which improves with p_T . Therefore,
840 by matching tracks to jets, TCCs have more precise energy and angular resolution

841 than jets constructed from calorimeter information only. These jets are required
 842 to have $p_T > 200$ GeV, $|\eta| < 2.0$ and $m_J > 50$ GeV.

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

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

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

853 Where the energy correlation functions are defined as:

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

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

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

854 A two-dimensional optimization of the jet mass and D_2 thresholds was per-
 855 formed to provide maximum sensitivity for this analysis. This optimization was
 856 done by maximizing the signal sensitivity (using HVT W' and G_{KK} samples)
 857 against the single quark and gluon jet backgrounds in bins of jet p_T . Figure 8.3
 858 shows the optimized thresholds on D_2 and jet mass as a function of p_T . Figure
 859

861 8.4 shows the efficiency of the optimized W/Z taggers as a function of jet p_T .

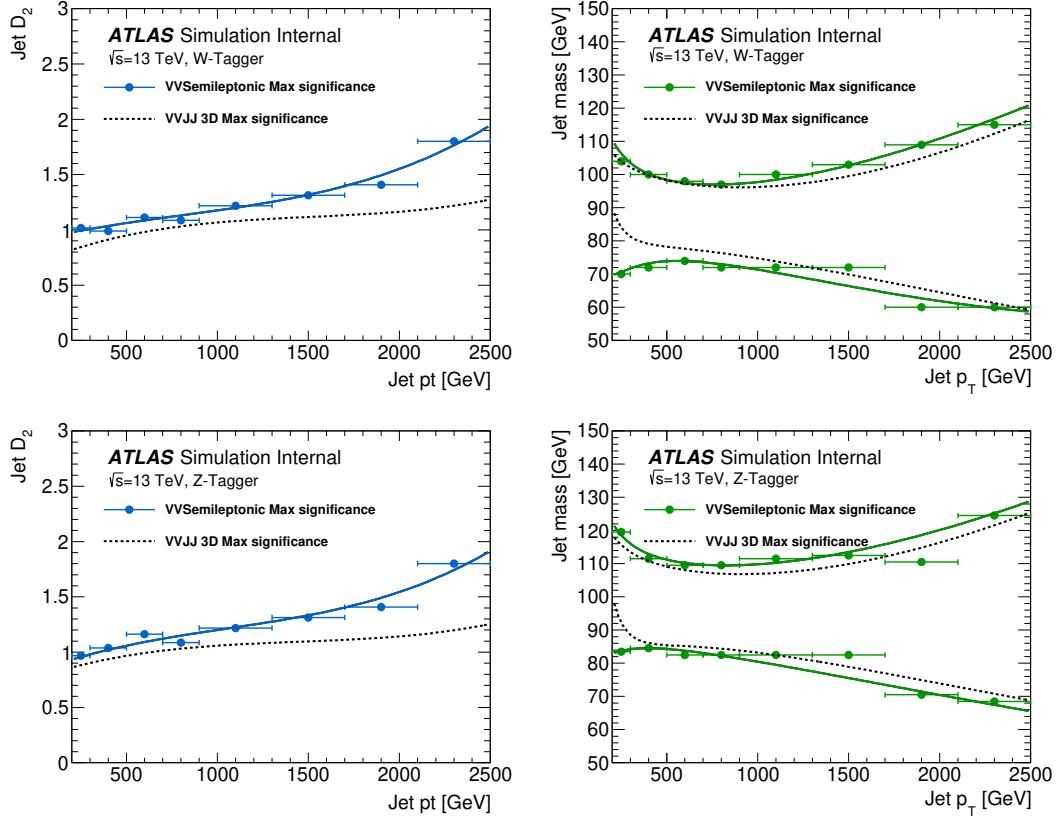


Figure 8.3: The upper cut on D_2 (a) and jet mass window cut i.e. the upper and lower boundary of the mass (b) of the W -tagger as a function of jet p_T . Corresponding values for Z -tagger are shown in (c) and (d). The optimal cut values for maximum significance are shown as solid markers and the fitted 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?

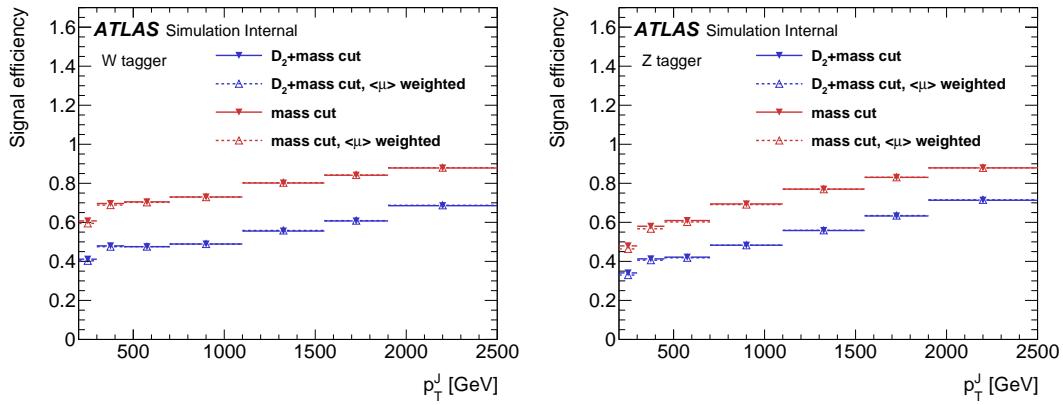


Figure 8.4: Natasha write caption

8.3.3 Variable Radius jets

To more accept more boosted Z bosons decaying to $b\bar{b}$ that would normally be rejected due to topological cuts discussed 9.4 variable radius (VR) track jets are used to identify b -jets within the catchment area of large-R jets [13]. VR jets are constructed from ID tracks using the anti- k_t algorithm with a radius parameter that depends on the p_T of the track, shown in Equation 8.5.

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

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

873 **8.3.4 Jet Flavor Tagging**

874 To further classify events, the small-R and VR jets originating from a b-quark
875 are classified using a multivariate b -tagging algorithm (BDT), MV2c10 [cite G 210
876 199]. This algorithm uses the impact parameters of the jet's ID tracks, secondary
877 vertices (if they exist), and reconstructed flight paths of b and c hadrons in the
878 jet to determine if the jet was induced by a b -quark. For this analysis the 85%
879 efficient working point of this algorithm is used giving c , τ , and light-flavor jet
880 rejection of 3, 8, and 34 respectively in simulated $t\bar{t}$ samples.

881 **8.4 MET/Neutrinos**

882 As neutrinos are uncharged and colorless they do not leave tracks or jets in the
883 detector. For this reason, neutrinos are reconstructed calculated the E_T^{miss} , the
884 negative vector sum of p_T all the physics objects and an extra "soft" term. The
885 "soft" term accounts for energy deposits not associated with any of the objects in
886 the event. For this analysis the soft term is given by the sum p_T of all ID tracks
887 not associated with objects in the event. The selected tracks must be matched to
888 the primary vertex, which decreases pile-up contamination [cite G 217 218]. The
889 tight working point is used [Natasha look up what this means].

890 **8.5 Overlap Removal**

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

896 $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ determines which object is removed from the event.

897 The overlap selections used in this analysis are:

898 - when an electron shares a track with another electron with the lower p_T
899 electron is rejected, as it is more likely to be a fake electron

900 - when a muon and electron share a track the muon is rejected if it is a
901 calo-muon, otherwise the electron is rejected

902 - when $\Delta R < 0.2$ for an electron and jet, the jet is rejected to maximize signal
903 acceptance

904 - when $\Delta R > 0.2$ for an electron and jet, the electron is rejected as likely
905 originated from decays within the jet

906 - when $\Delta R < \min(0.4, 0.04 + 10\text{GeV}/p_T^\mu)$ the muon is rejected, again maximiz-
907 ing signal acceptance, otherwise the jet is rejected

908 - when $\Delta R < 1.0$ for the a large-R jet and electron, the jet is rejected

909 **Chapter 9**

910 **Event Selection and**

911 **Categorization**

912 **9.1 Pre-selection**

913 Before applying topological cuts to suppress backgrounds and reduce data
914 size in this search, preselection cuts are applied which include trigger and event
915 requirements. Events must contain exactly one tight lepton (no additional loose
916 leptons), the $p_T^{\ell\nu} > 75$ GeV, and there must be at least two small-R jets or one
917 large-R jet.

918 **9.2 Trigger**

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

924 **9.3 non-VBF/VBF RNN**

925 To classify events as originating from non-VBF/DY or VBF production a
926 recursive neural network (RNN [19]) is used. This approach is more powerful than
927 a cut-based classification as it improves signal efficiency and analysis sensitivity
928 by exploiting correlations between variables that the RNN learns. In particular, a
929 RNN architecture is ideal as it can handle variable numbers of jets in the events.

930 The RNN uses the four-momentum of candidate VBF jets to classify events
931 as VBF or GGF topologies. As sometimes jets will be incorrectly reconstructed
932 the number of jets in the event are expected to vary across the inputs samples.
933 VBF candidate jets are identified by removing jets from the event that are likely
934 from $W/Z \rightarrow qq$. For the resolved regime this means removing the two leading
935 small-R jets from the VBF candidate jet list. For the merged regime this means
936 removing small-R jets separated by less than 1.0 in dR from the large-R jet. VBF
937 candidate jets are also required to be within $|\eta| < 4.5$. From the list of remaining
938 VBF candidate jets, the two highest- p_T jets are chosen.

939 The architecture of the RNN is show in Figure 9.1. LSTMs are a type of
940 RNN that extract meaningful information and can retain it (unlike other neural
941 networks architectures). This is useful for VBF event classification for events with
942 two jets, where using the kinematic properties of both jets (and their correlations)
943 will lead to more efficient event classification.

944 In this RNN architecture, the VBF candidates are first passed to a masking
945 layer which checks the number of jets in the event. If there is only one jet, only one
946 LSTM layer is used. The output of masking is then passed to a Long Short-Term
947 Memory (LSTM) cell, with a tanh activation function. This output is passed to
948 a dropout layer, that has a probability of 0.3 to completely forget the output of
949 the LSTM. Dropout is a regularization method, that prevents over-fitting. The

950 output of the dropout layer is then passed to the second LSTM and then through
951 another dropout layer with a probability of 0.3.

952 The weights and other parameters of the network are learned by training the
953 network with VBF and GGF signals over 200 epochs with an Adam Optimizer
954 [natasha add reference]. The training is truncated if the network parameters are
955 unchanged after ten iterations. The training, testing and validation sets are 56,
956 30, and 14 percent of the input samples, respectively. Figure [add INT figure
957 32] shows the loss function of the network as a function of training epochs. The
958 validation test set has a smaller loss function as dropout was not applied. Figure
959 9.3 shows the ROC curve for the RNN using k-fold cross validation.

960 Finally this output is passed to a dense layer [natasha ask antonio about this]
961 and then to a sigmoid activation layer, leading to an overall RNN score. Figure 9.2
962 shows the RNN discriminant for backgrounds, GGF signals, and VBF signals. The
963 RNN score is ~ 0 for GGF and background processes and ~ 1 for VBF processes.
964 Figure 9.4 shows the limits for various signal processes based on the RNN cut
965 applied. Requiring the RNN score to be > 0.8 was chosen as it provided the best
966 significance (and signal efficiency) across for this final state and the $\nu\nu qq$ and $\ell\ell qq$
967 channels this channel will be combined with for future publications.

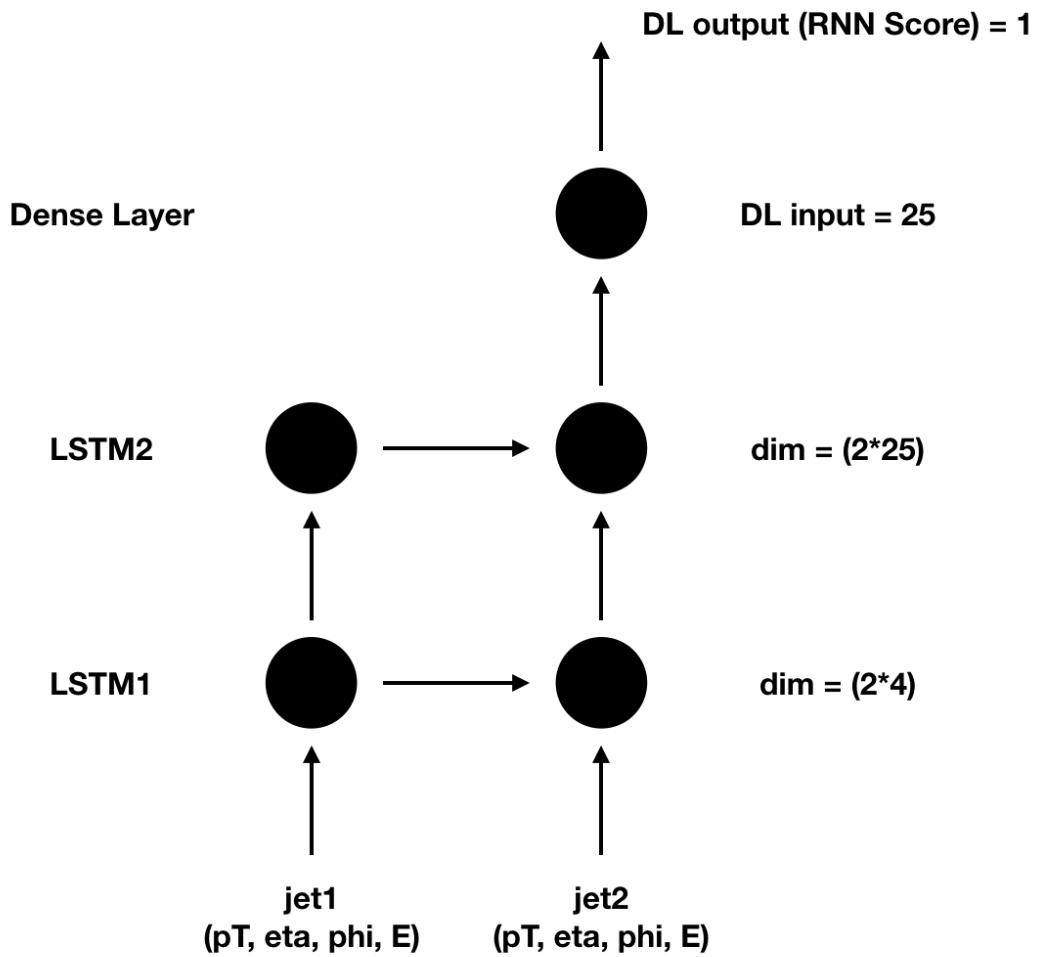


Figure 9.1: RNN architecture. Natasha add caption

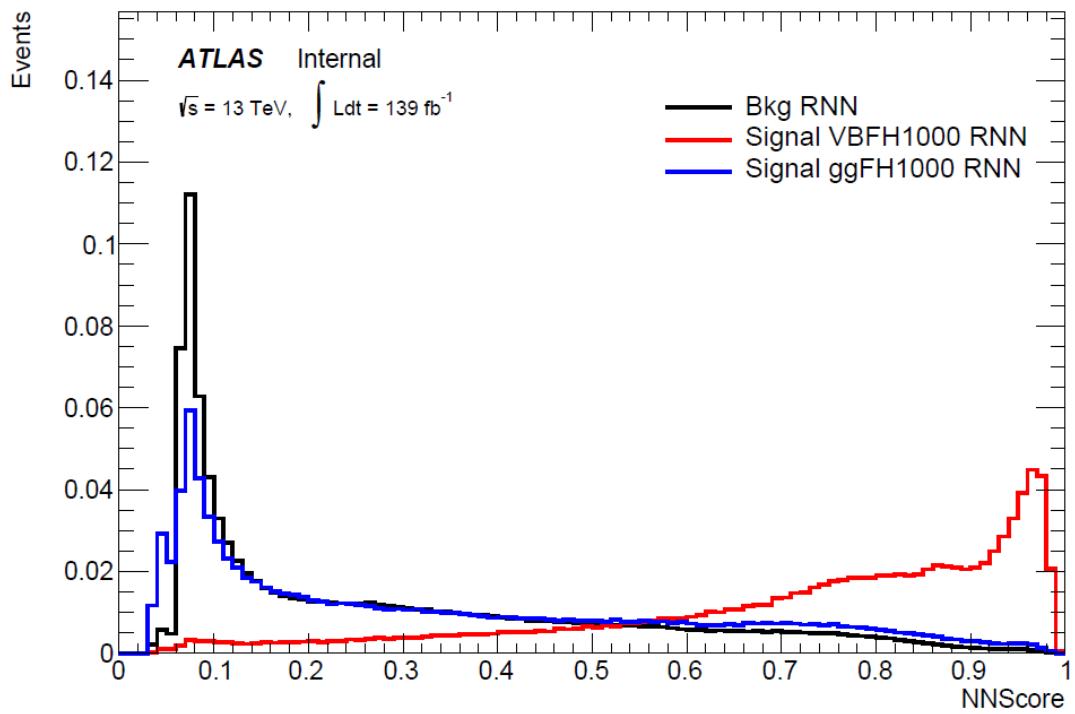


Figure 9.2: RNN Score distribution for ggF and VBF signals and backgrounds.

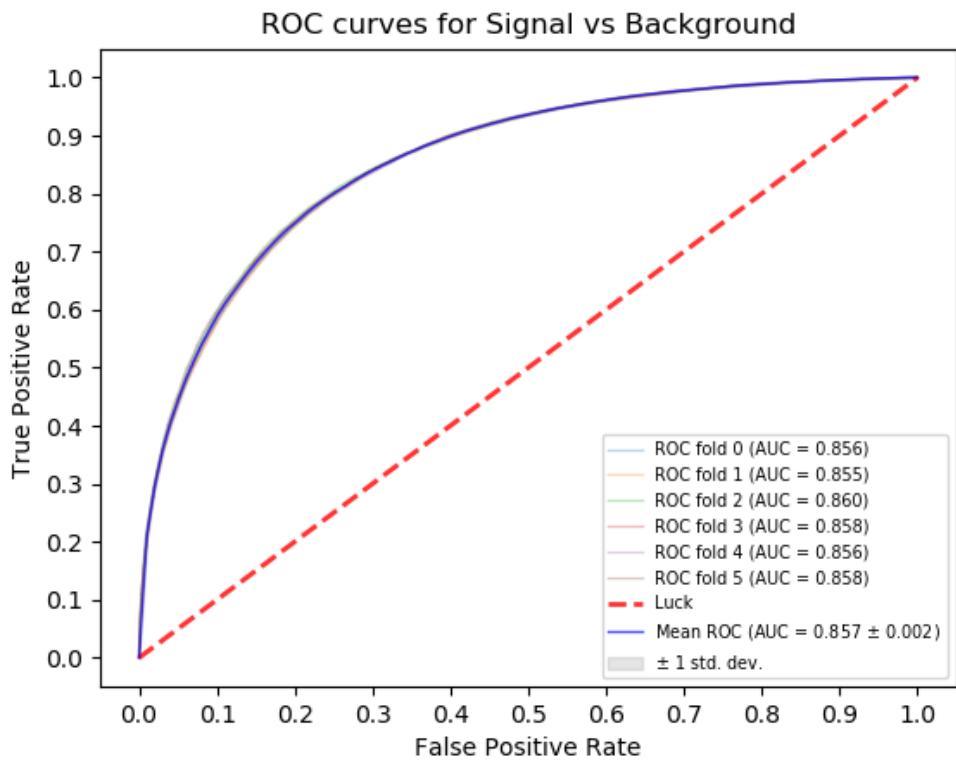


Figure 9.3: ROC curve using k-fold validation for RNN.

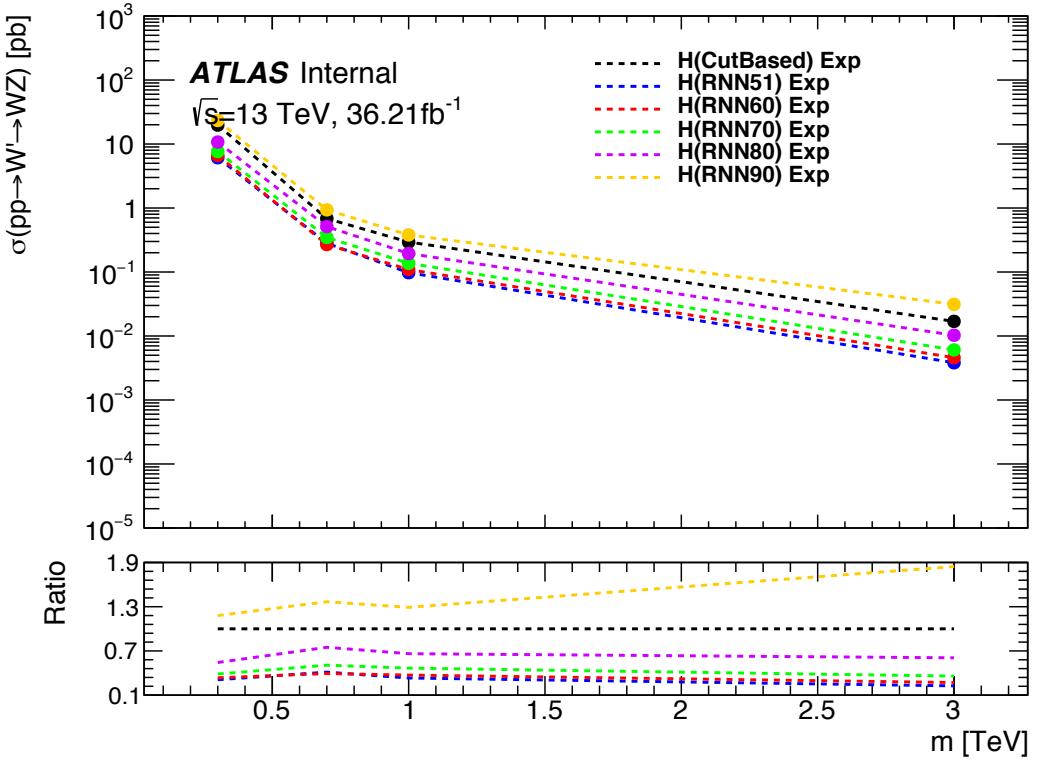


Figure 9.4: Comparison of GGF Z' limits for different RNN score selections. The bottom panel shows the ratio of the upper limits set for different RNN cuts to the cut-based analysis. In this panel smaller numbers, indicate that the expected upper limit is smaller than the cut-based analysis, which is desired.

968 9.4 Topological Cuts

969 Once an event is classified as VBF or GGF via the RNN, it must pass other
 970 topological cuts that maximize S/\sqrt{B} . First, to efficiently select events with a
 971 $W \rightarrow \ell\nu$ candidate exactly one tight lepton is required and $E_T^{miss} > 100(60)$
 972 GeV and $p_{T,\ell\nu} > 200(75)$ GeV in the merged (resolved) analysis to suppress the
 973 multi-jet background.

974 For the merged analysis, in addition to the $W \rightarrow \ell\nu$ and $W/Z \rightarrow J$ selections
 975 above, the $\min(p_{T,\ell\nu}, p_{T,J})/m_{WV} > 0.35(0.25)$ for the GGF (VBF) category. To
 976 reduce $t\bar{t}$ contamination events with the signal region and $W+jets$ control region

977 events with at least one b jet with $\Delta R > 1.0$ from the large-R jet are excluded.
 978 High purity signal regions require the D_2 and W/Z mass window cut to be passed,
 979 whereas the low purity region only requires the W/Z mass window cut to be
 980 passed. Finally for events to be classified as tagged the large-R jet must contain
 981 exactly two b-tagged jets. Untagged events must have no more than one b-tagged
 982 jet matched to the large-R jet. These selections are shown in Table 9.2. The
 983 distributions for the variables used in merged analysis for top control regions are
 984 shown in Figure 9.6-9.9.

985 Events failing the merged selection are then re-analyzed in the resolved cat-
 986 egory. To enhance resolved signals, the event should contain two high- p_T boson
 987 candidates that are back-to-back in the ϕ as shown by the cuts in Table 18. Again
 988 to suppress the $t\bar{t}$ background in the WCR and SR events are required to have no
 989 additional b-jets.

990 The WV system mass, m_{WV} is reconstructed from the lepton, neutrino, and
 991 hadronically-decaying boson candidate. The momentum of the neutrino along the
 992 z -direction is obtained by constraining the W boson mass of the lepton neutrino
 993 system to be $80.3 \text{ GeV}/c^2$. For complex solutions to this constraint, p_Z is taken
 994 as either the real component of the solution or the one with the smaller absolute
 995 value of the two real solutions. For the resolved analysis, m_{WV} is reconstructed
 996 by constraining the $W(Z)$ dijet system:

$$p_{T,jj}^{corr} = p_{T,jj} \times \frac{m_{W/Z}}{m_{jj}} \quad (9.1)$$

$$m_{jj}^{corr} = m_{W/Z} \quad (9.2)$$

998 where m_{jj} and $m_{W/Z}$ are the reconstructed invariant mass of the hadronically-
 999 decaying W/Z boson and the PDG values of the W/Z boson masses, respectively.

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^{miss}			> 100 GeV				
	$p_T(\ell\nu)$			> 200 GeV				
$W/Z \rightarrow J$	Num of large- R jets			≥ 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 $Z \rightarrow J$: ≤ 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			≥ 1		
Pass VBF selection			no (yes) for DY/ggF (VBF) category					

1000 A summary of the resolved selections is shown in Table 9.3. The distributions for
1001 the variables used in the resolved analysis in the TCR are shown in Figure 9.10,
1002 9.11.

1003 The analysis cutflow is shows in Figure 9.5. Events classified as VBF events
1004 are classified as Merged High purity, low purity or resolved signal region selections
1005 sequentially. If the event does not pass any of these selections but passes a VBF
1006 control region selection it is classified as a VBF CR event. If the event fails the
1007 VBF selection it is then checked if it passes the Merged High purity, Low purity
1008 or resolved signal region selections (NB: for the WZ decay modes all the regions
1009 have tagged and untagged categories). If the event fails all the GGF signal region
1010 selections, it is then kept for GGF control region selections, if it passes those
1011 selections.

Table 9.3: The list of selection cuts in the resolved analysis for the WW and WZ signal regions (SR), $W+\text{jets}$ 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^{miss}	$> 60 \text{ GeV}$	
	$\cancel{p}_T(\ell\nu)$	$> 75 \text{ GeV}$	
$W/Z \rightarrow jj$	Number of small-R jets	≥ 2	
	Leading jet p_T	$> 60 \text{ GeV}$	
	Subleading jet p_T	$> 45 \text{ GeV}$	
	$Z \rightarrow q\bar{q}$ $W \rightarrow q\bar{q}$	$78 < m_{jj} < 105 \text{ GeV}$ $68 < m_{jj} < 98 \text{ GeV}$	$50 < m_{jj} < 68 \text{ GeV}$ or $105 < m_{jj} < 150 \text{ GeV}$
Topology cuts	Num. of b -tagged jets	For $Z \rightarrow jj$: ≤ 1 ($= 2$) for untagged (tagged) category	
	$\Delta\phi(j, \ell)$	> 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	
Top vetos	$\min(p_{T,\ell\nu}, p_{T,jj}) / m_{WW}$	$> 0.35(0.25)$ for DY/ggF (VBF) category	
	Number of additional b -tagged jets	0	≥ 1
Pass VBF selection		no (yes) for DY/ggF (VBF) category	

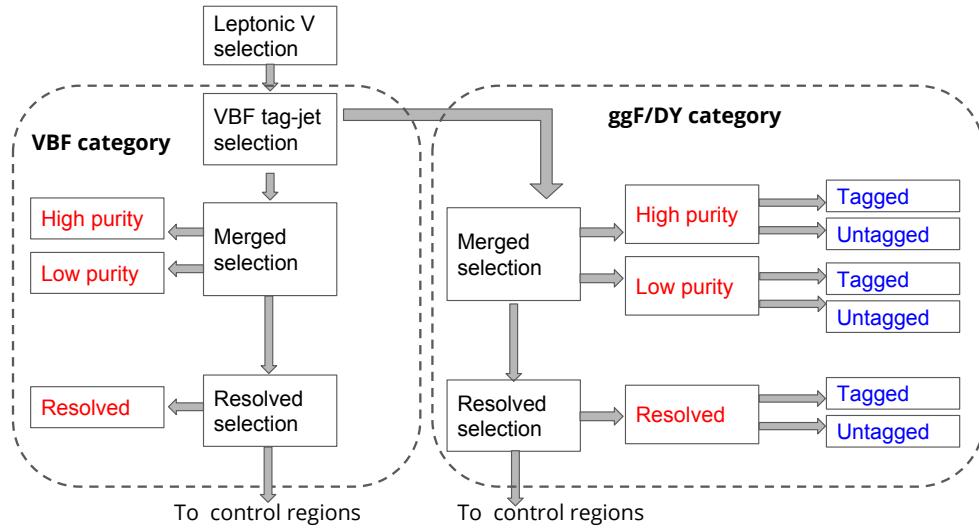


Figure 9.5: Event Categorization. Natasha write more.

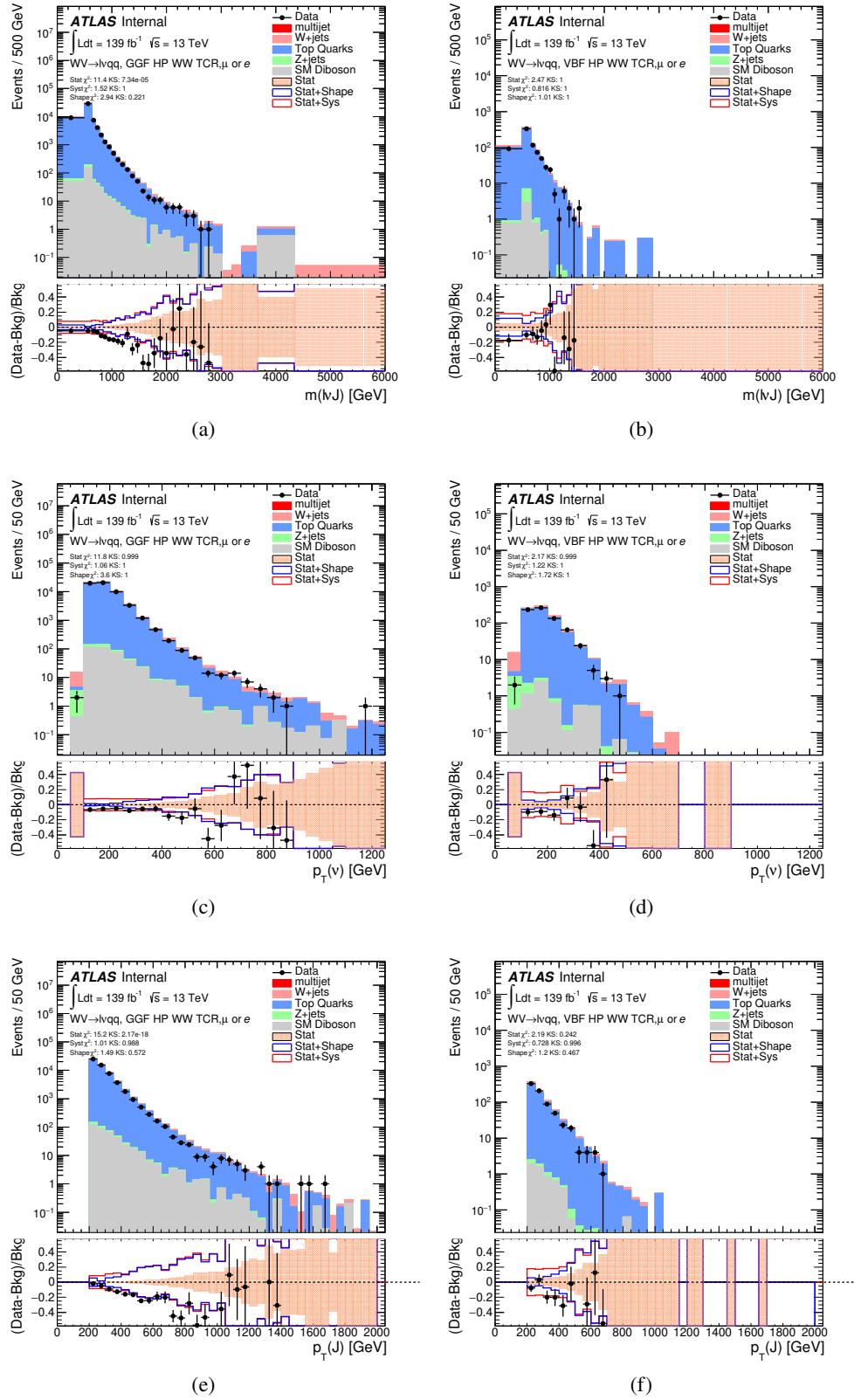


Figure 9.6: Data MC comparison for the merged WW HP TCR. The bottom panel shows the ratio of the difference⁸⁷ between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.

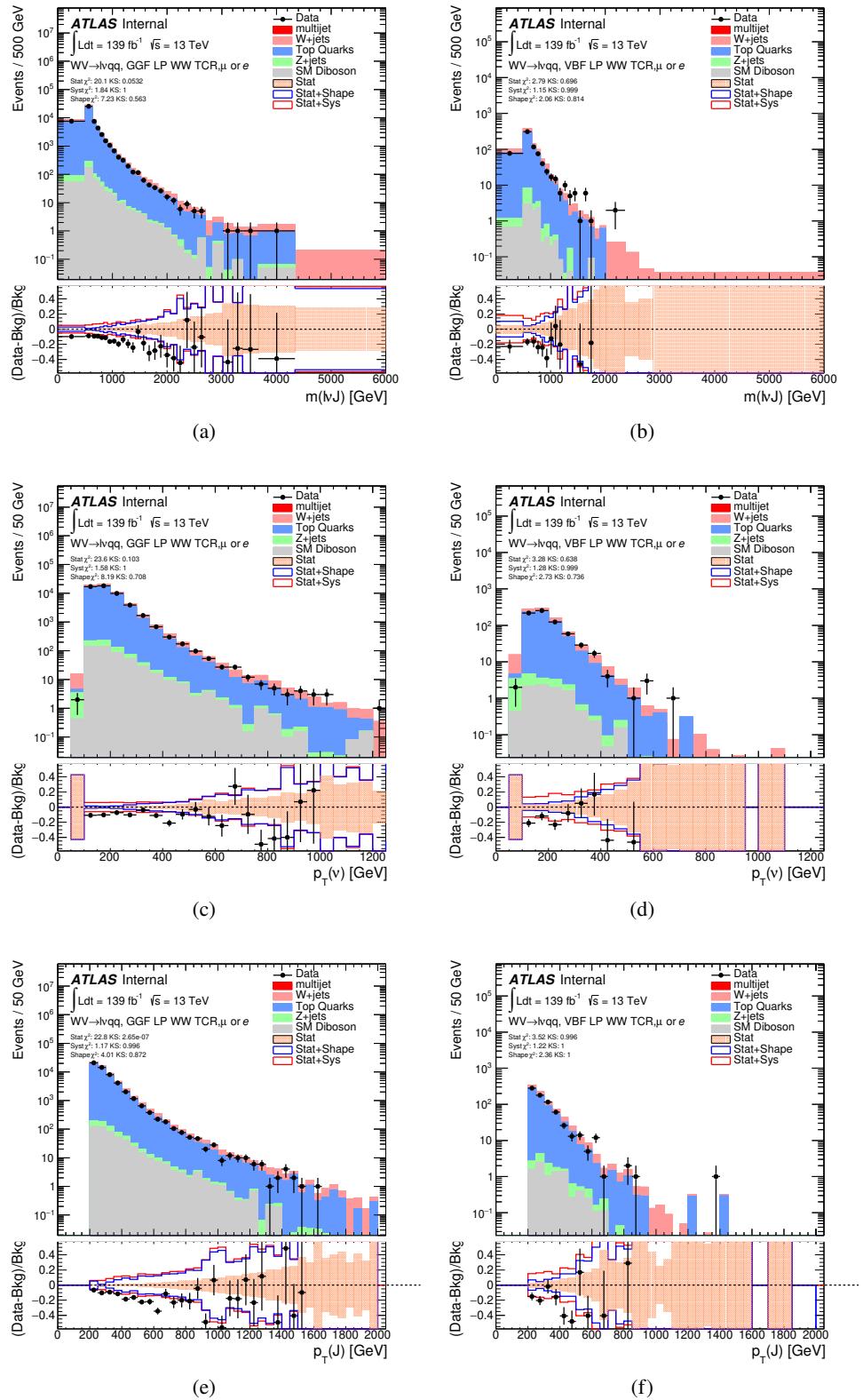


Figure 9.7: Data MC comparison for the merged WW LP TCR. The bottom panel shows the ratio of the difference⁸⁸ between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.

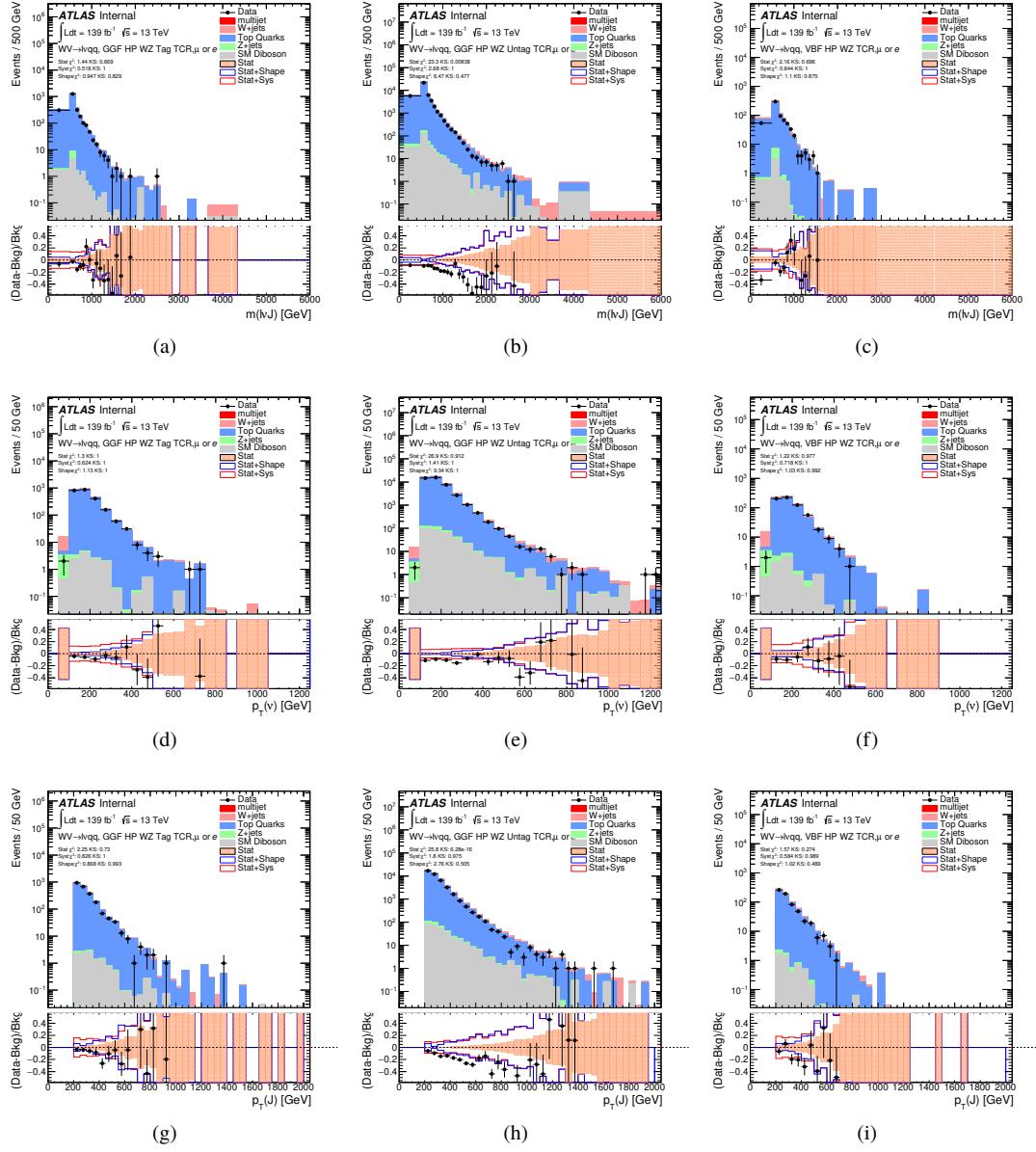


Figure 9.8: Data MC comparison for the merged WZ HP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.

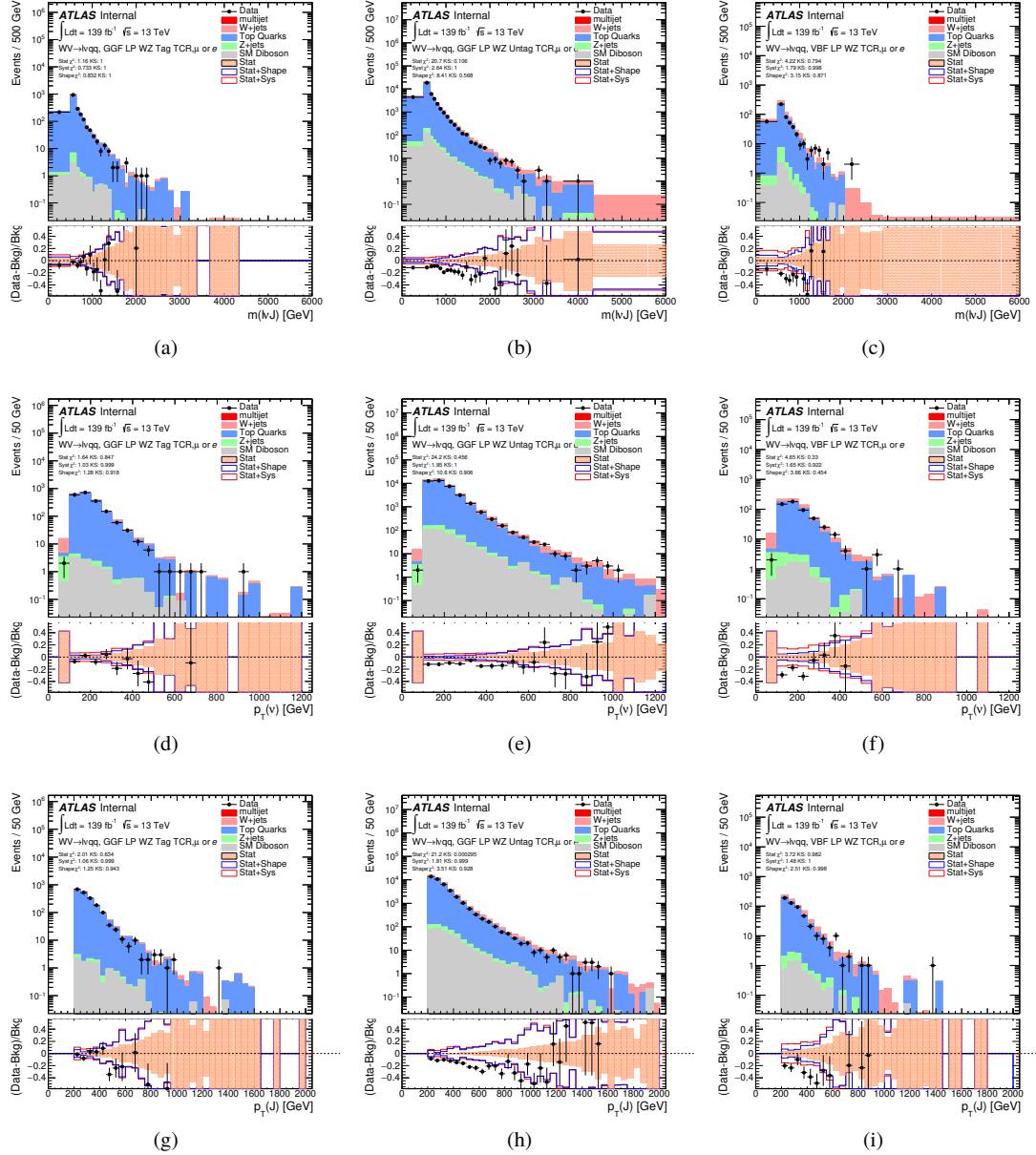


Figure 9.9: Data MC comparison for the merged WZ LP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.

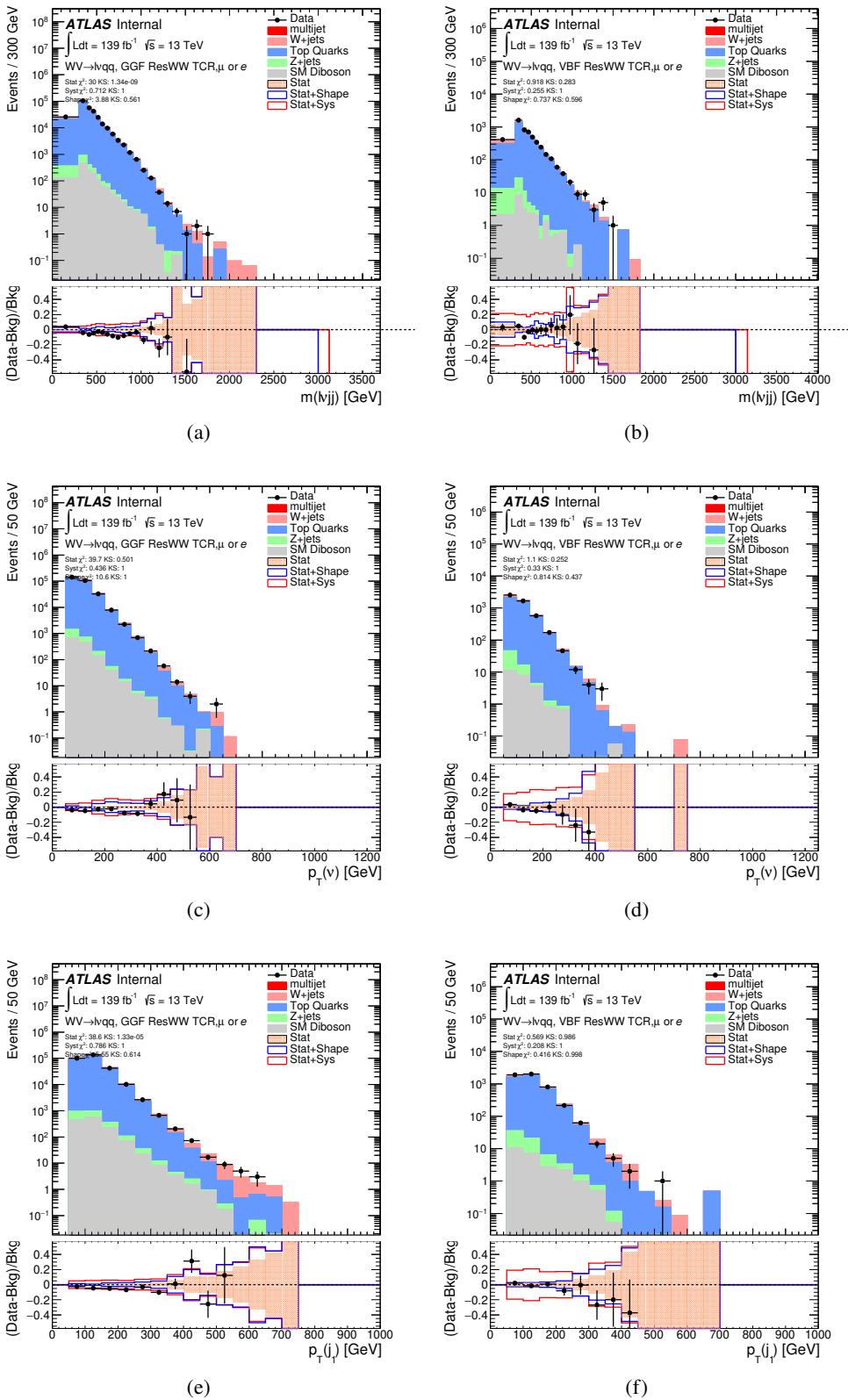


Figure 9.10: Data MC comparison for the resolved WW TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.
91

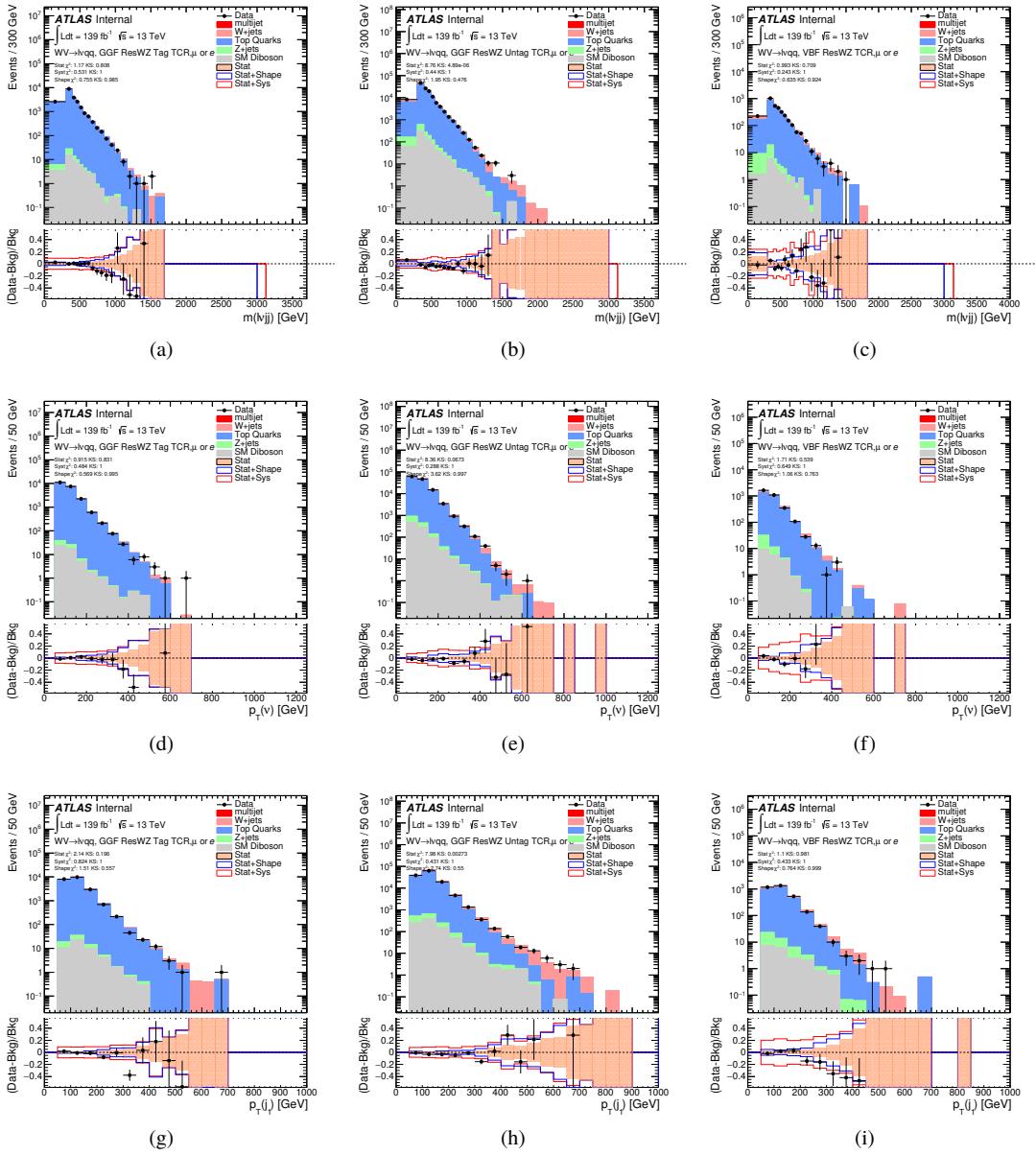


Figure 9.11: Data MC comparison for the resolved WZ TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.

1012 **9.5 Selection Acceptance times efficiency for Sig-**
 1013 **nal Events**

1014 The acceptance times efficiency for the signal region selection is defined as:

$$A \cdot \epsilon = \frac{N_{\text{events selected}}^{\text{truth}}}{N_{\text{events generated}}^{\text{truth}}} \cdot \frac{N_{\text{events selected}}^{\text{reco}}}{N_{\text{events selected}}^{\text{truth}}} = \frac{N_{\text{events selected}}^{\text{reco}}}{N_{\text{events generated}}^{\text{truth}}} \quad (9.3)$$

1015 The distributions of $A \cdot \epsilon$ as a function of the resonance mass for the different spin
 1016 models are shown in Figures 9.13 - ??.

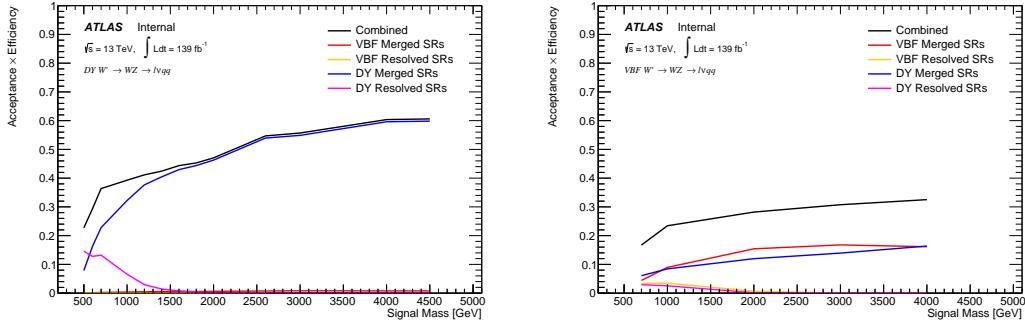


Figure 9.12: Selection acceptance times efficiency for the $W' \rightarrow WZ \rightarrow l\nu qq$ events from MC simulations as a function of the W' mass for (a) Drell-Yan and (b) VBF production, combining the merged HP and LP signal regions of the $WW \rightarrow l\nu J$ selection and the resolved regions of the $WW \rightarrow l\nu jj$ selection.

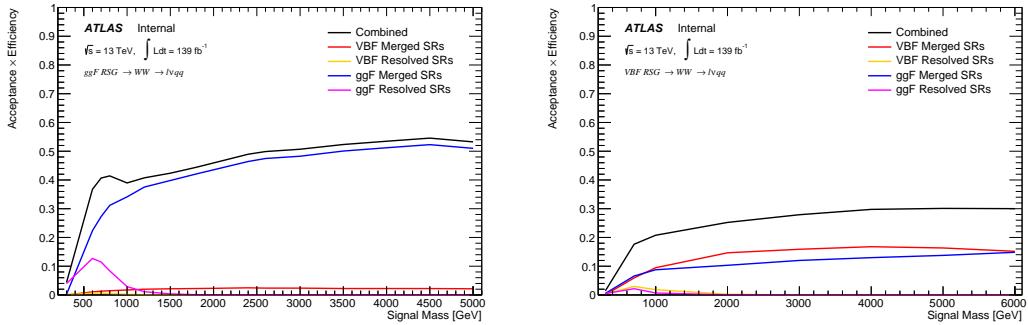


Figure 9.13: Selection acceptance times efficiency for the $G \rightarrow WW \rightarrow \ell\nu qq$ events from MC simulations as a function of the G mass for (a) Drell-Yan and (b) VBF production, combining the merged HP and LP signal regions of the $WW \rightarrow \ell\nu J$ selection and the resolved regions of the $WW \rightarrow \ell\nu jj$ selection.

1017 9.6 Background Estimate

1018 9.6.1 Control Regions

1019 To more accurately model the two dominant backgrounds in this analysis,
 1020 $W+jets$ and $t\bar{t}$, control regions are constructed for each. These control regions
 1021 are dominated by these processes and used to extract normalization factors in
 1022 the final likelihood fit that are then used in the signal region estimates. For the
 1023 $t\bar{t}$ control region the event must contain at least one such b jet. The WCR is
 1024 constructed using the $m_{jj/J}$ mass window sidebands. All other backgrounds are
 1025 estimated using simulation, except fake lepton backgrounds, which are derived
 1026 using a data-driven method.

1027 9.6.2 Fake Lepton Backgrounds

1028 Backgrounds in this analysis containing real leptons (e.g. $W/Z+jets$, diboson,
 1029 $t\bar{t}$, single- t) are well-modeled with simulated samples and constrained with data
 1030 from CRs. However, the fake lepton background (also referred to as the multijet
 1031 background) is not well-modeled with simulation. For this reason, the multijet

background is extracted from data. Heavy flavor decay products, jets, and converted photons can be mistakenly reconstructed as electrons. Fake electrons often arise from jet fakes while non-prompt muons usually arise from heavy flavor decay. For this analysis, these fake electrons generally fail the electron ID criteria and fake muons fail the muon isolation requirement. Therefore, to derive the multijet template shape the SR and CR selections and inverted lepton requirements are used as seen in Table 9.4. NB: by inverting the lepton isolation/identification criteria the SRs and CRs are orthogonal.

The template shape of the MJ background is determined by using a multijet validation region (MJVR) that requires the inverted lepton isolation/identification requirement and the two signal jets to satisfy the m_{jj} requirement used in the $W+jets$ CRs. The E_T^{miss} distribution in MJCR is shown in Figure 9.14 for 2017 data. The template is then extracted by subtracting the data in the MJVR from the electroweak background processes. The resulting template and electroweak backgrounds are then fit to data. In this fit, the E_T^{miss} distribution compared to data to extract electroweak background, multijet electron and muon background normalizations. The fitted scale factors from this MJVR template are then applied in the MJCR template. The electron and muon background normalizations in the MJCR template are parameters in the final simultaneous fit. Technically, there should be a separate template for every CR and SR, but some MJ regions have insufficient statistics to do this. Additionally, the shapes for the MJ templates for VBF and ggF regions are found to be compatible within statistical uncertainty. Therefore, the sample MJ template used for VBF and ggF CR/SRs, but with different pre-MJ-fit scale factors.

This template method was validated using WCR and full Run 2 data. The results of the fit are shown in Table 9.5. The multijet contribution in the muon

channel for $p_T^W > 150$ GeV is consistent with zero, and therefore neglected in the final fit. Applying the extracted normalization factor to MJVR in WCRs for various kinematic variables such as E_T^{miss} , W transverse mass, lepton p_T , and the invariant mass as show in Figures 9.15 -9.24. These figures show good agreement between the data and background estimate.

Table 9.4: Definitions of “inverted” leptons used in multijet control region

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

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

1063

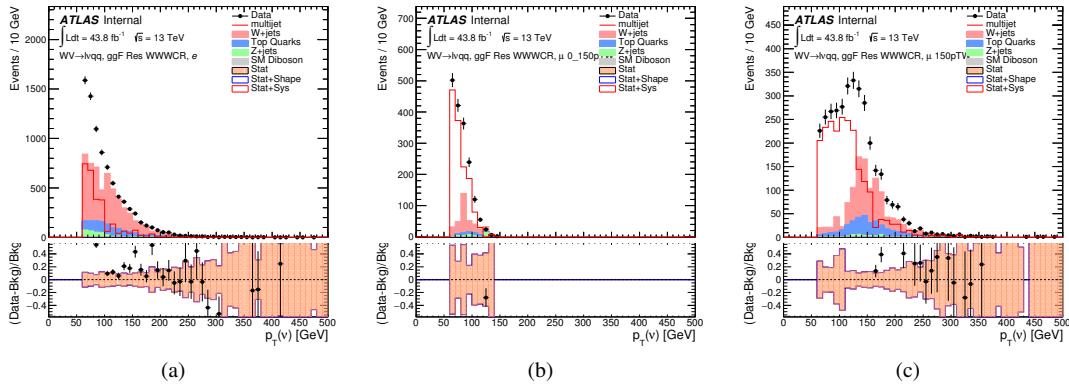


Figure 9.14: The E_T^{miss} distribution in MJCR for 2017 data in the electron channel(left), muon channel with W -boson $p_T < 150$ GeV (center) and > 150 GeV (right). Multi-jet templates are calculated as remaining data components after excluding known MC

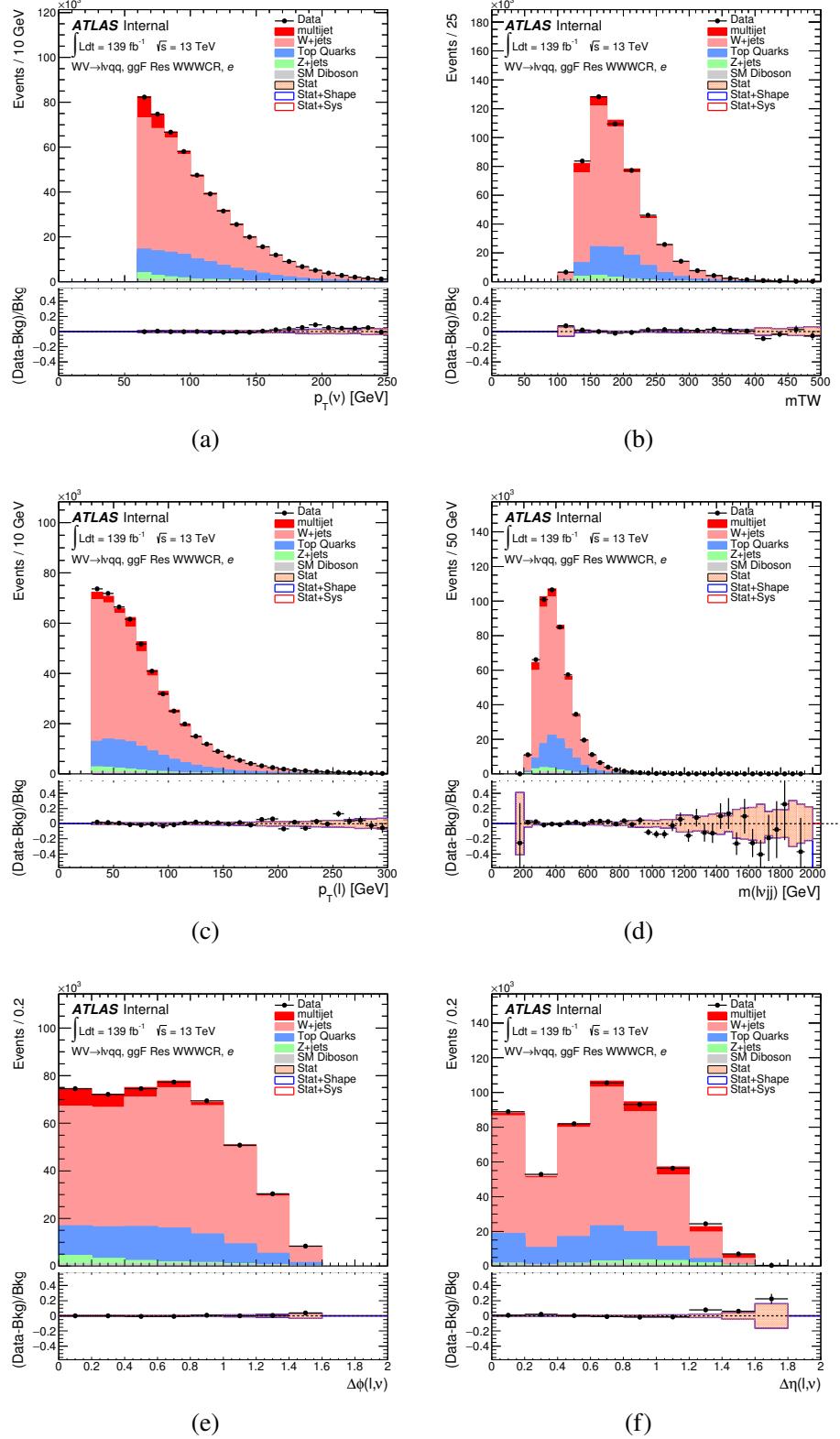


Figure 9.15: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{l\nu jj}$, lepton- ν angular distance in the WW electron channel. The MJ template is obtained from the pre-MJ-fit.

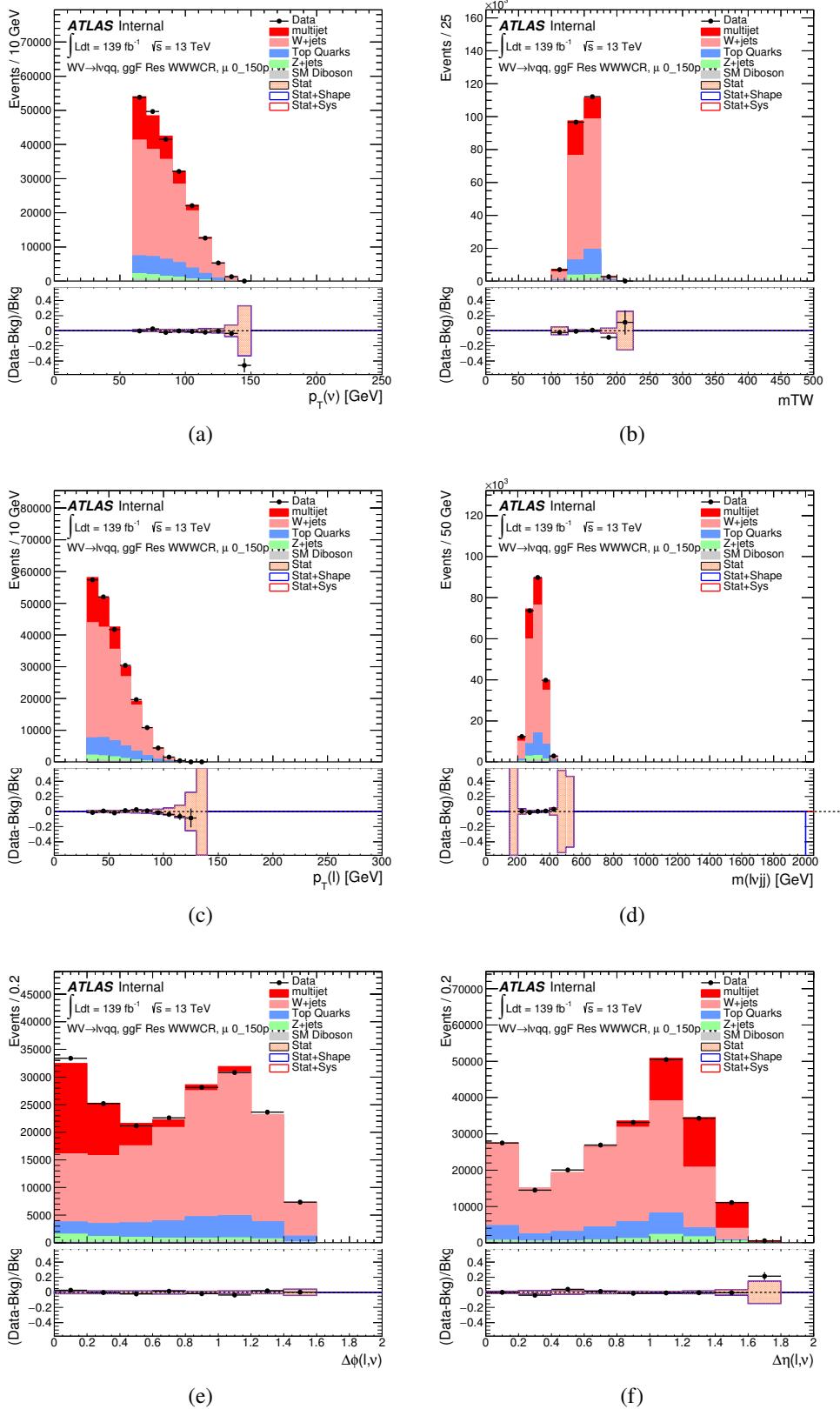


Figure 9.16: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{l\nu jj}$, lepton- ν angular distance in the WW muon channel. The MJ template is obtained from the pre-MJ-fit.

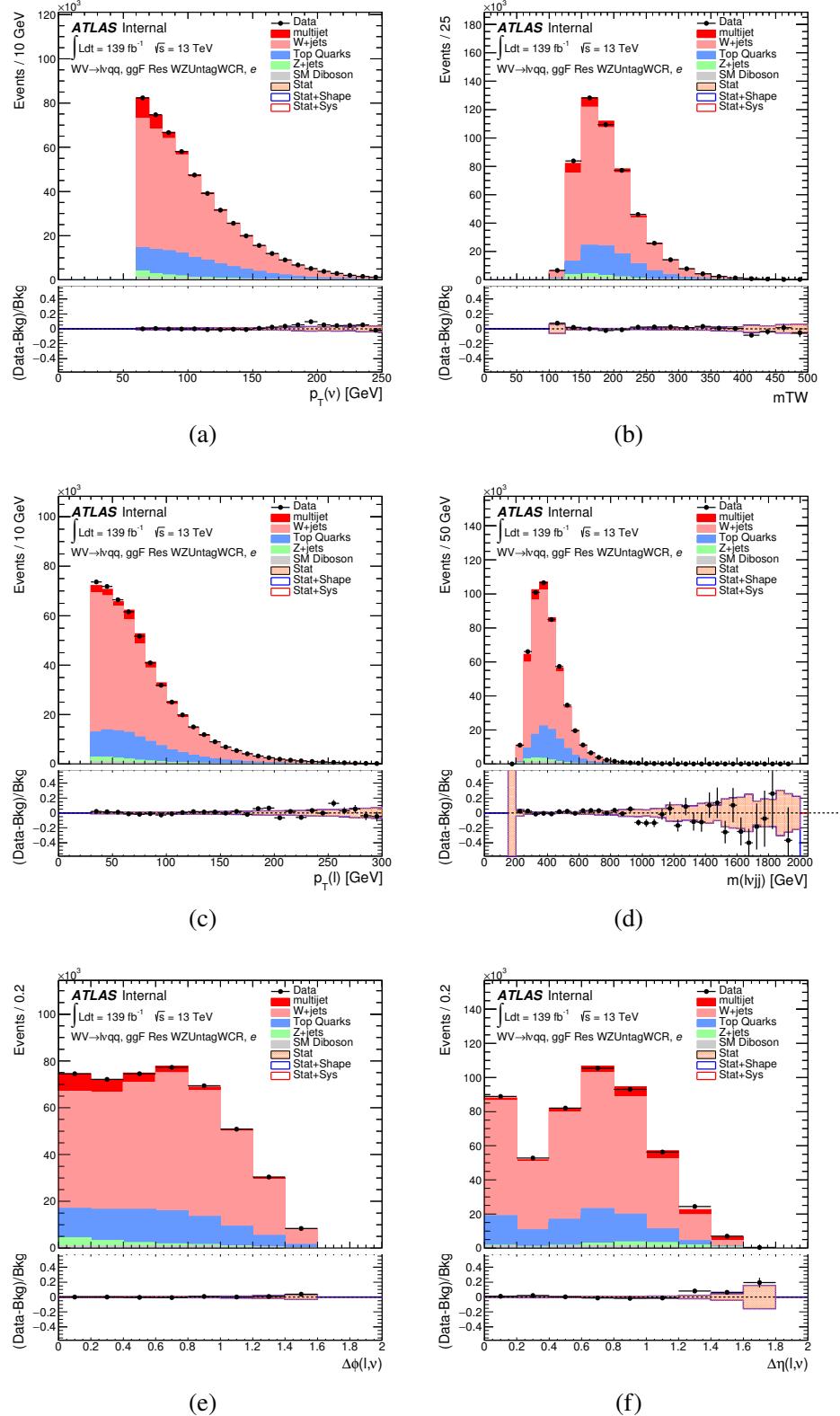


Figure 9.17: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{l\nu jj}$, lepton- ν angular distance in the WZ untag electron channel. The MJ template is obtained from the pre-MJ-fit.

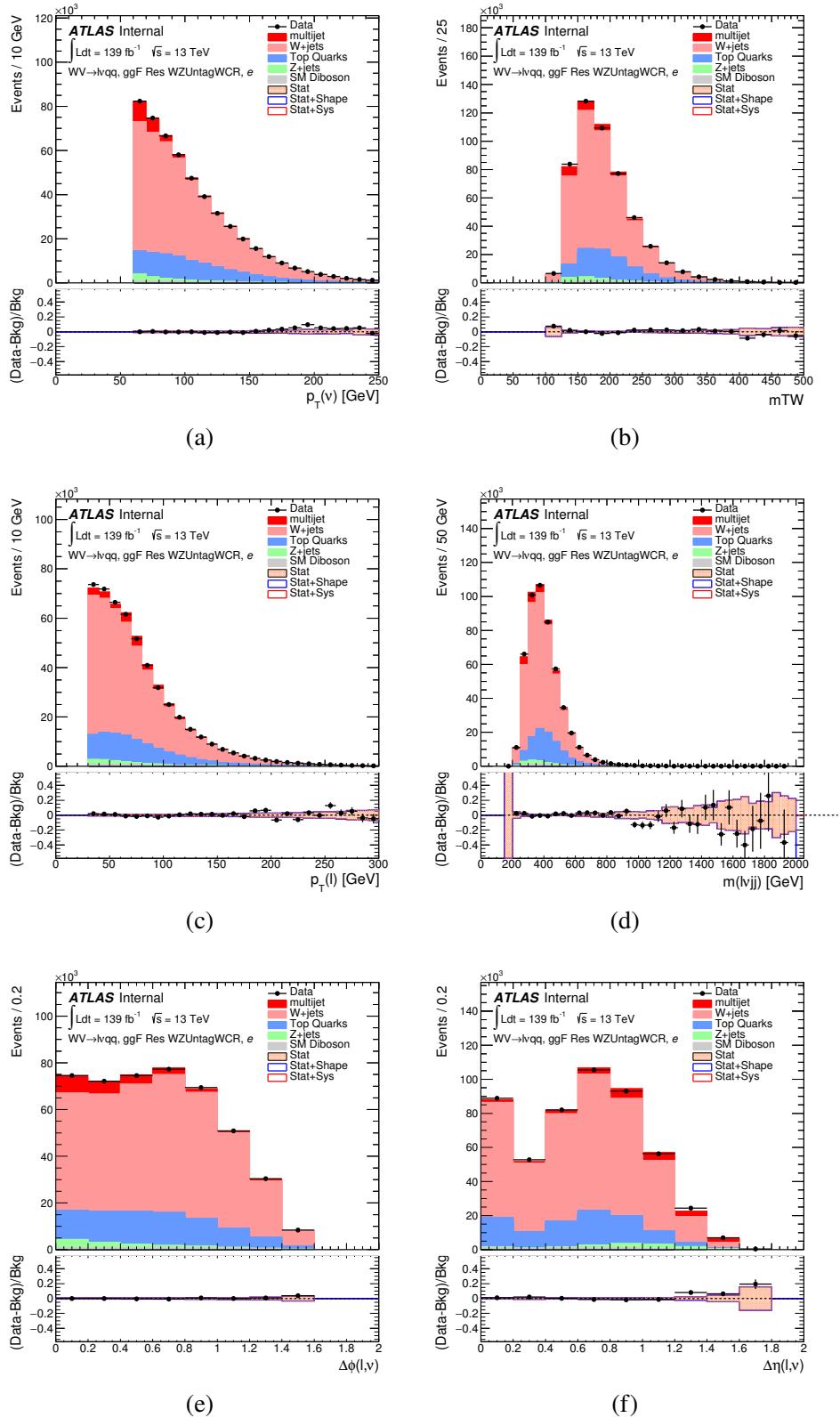


Figure 9.18: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{l\nu jj}$, lepton- ν angular distance in the WZ untag muon channel. The MJ template is obtained from the pre-MJ-fit.

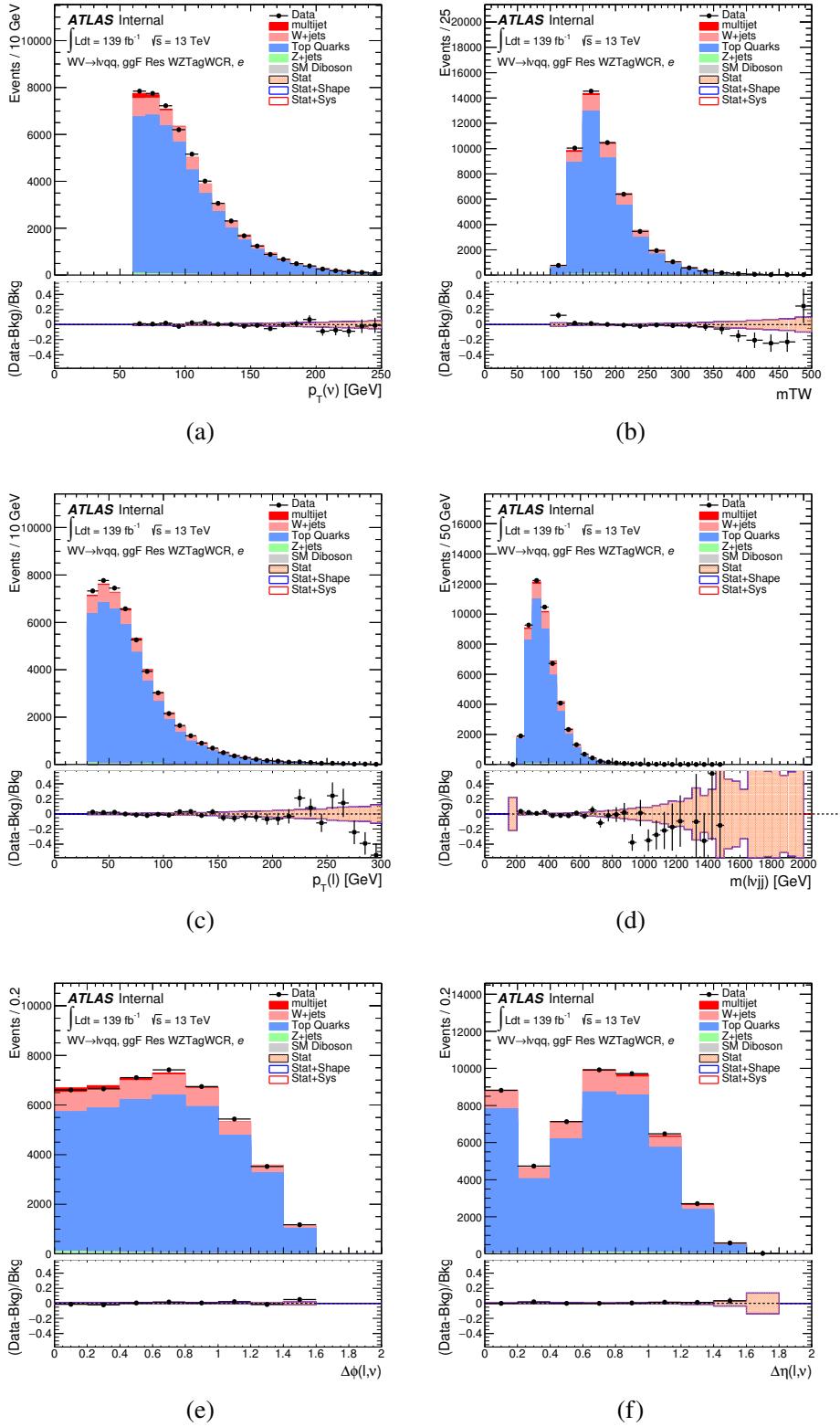


Figure 9.19: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{l\nu jj}$, lepton- ν angular distance in the WZ untag electron channel. The MJ template is obtained from the pre-MJ-fit.

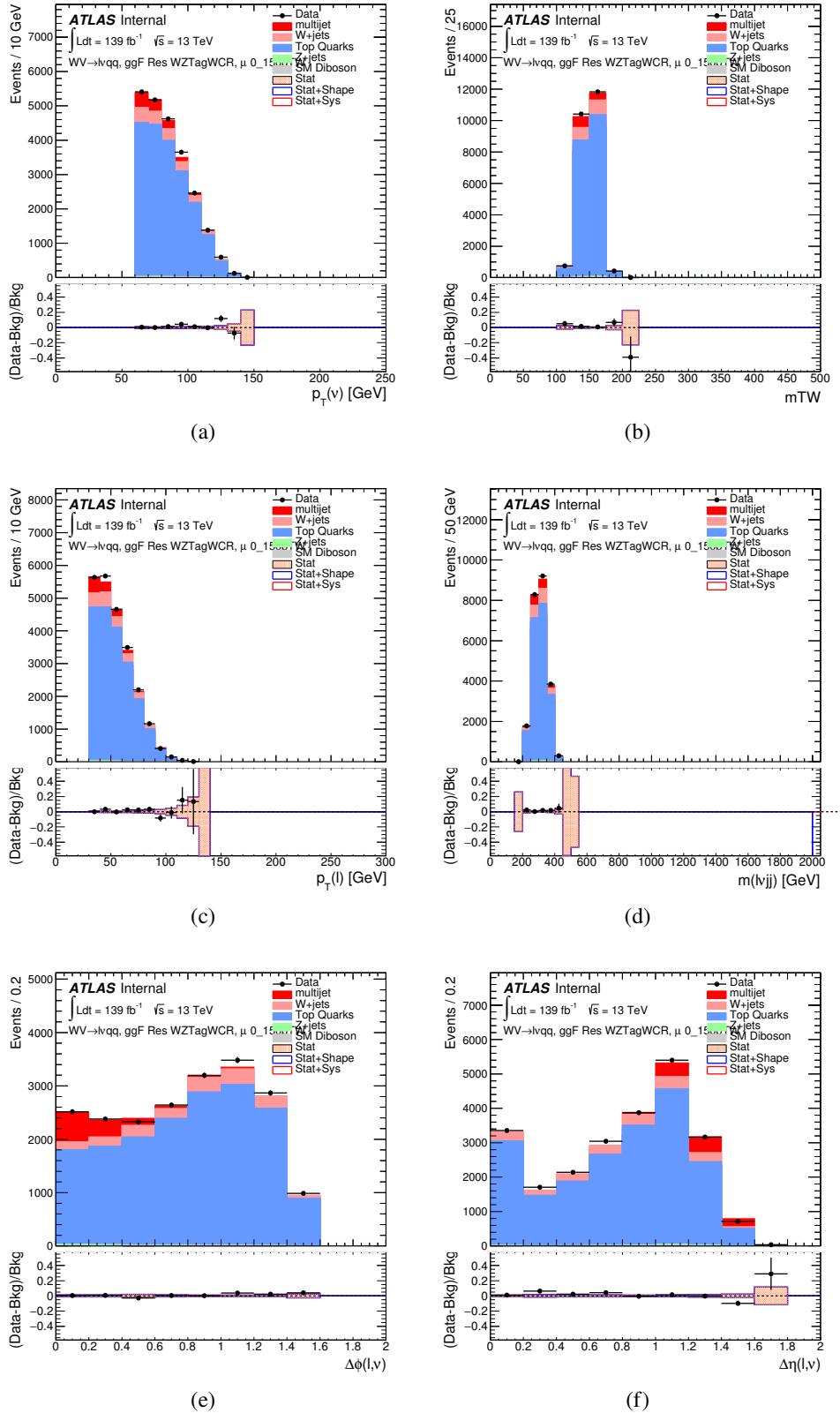


Figure 9.20: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{l\nu jj}$, lepton- ν angular distance in the WZ untag muon channel. The MJ template is obtained from the pre-MJ-fit.

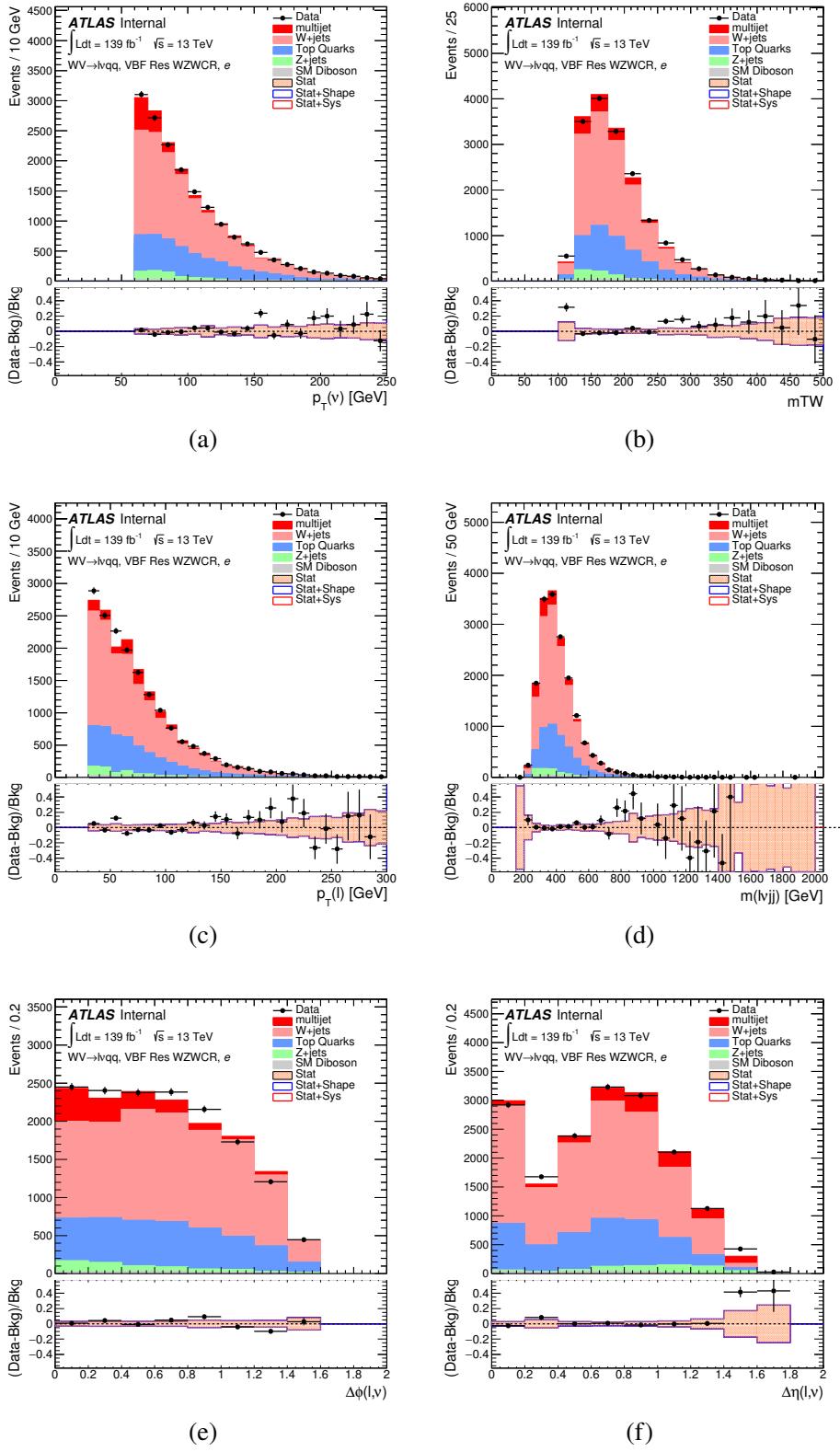


Figure 9.21: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{l\nu jj}$, lepton- ν angular distance in the VBF WW electron channel. The MJ template is obtained from the pre-MJ-fit.

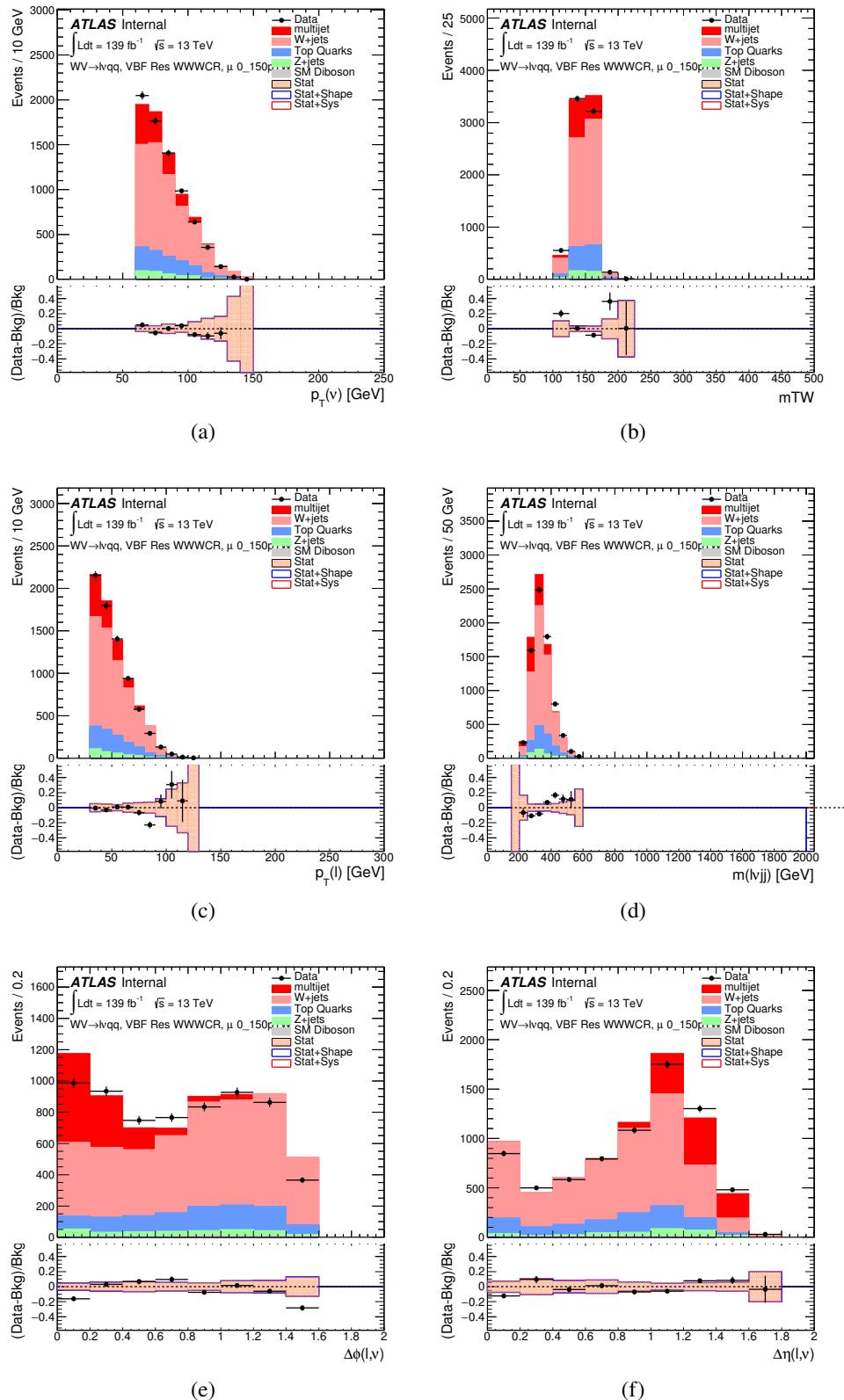


Figure 9.22: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{\ell\nu jj}$, lepton- ν angular distance in the VBF WW muon channel. The MJ template is obtained from the pre-MJ-fit. ¹⁰⁴

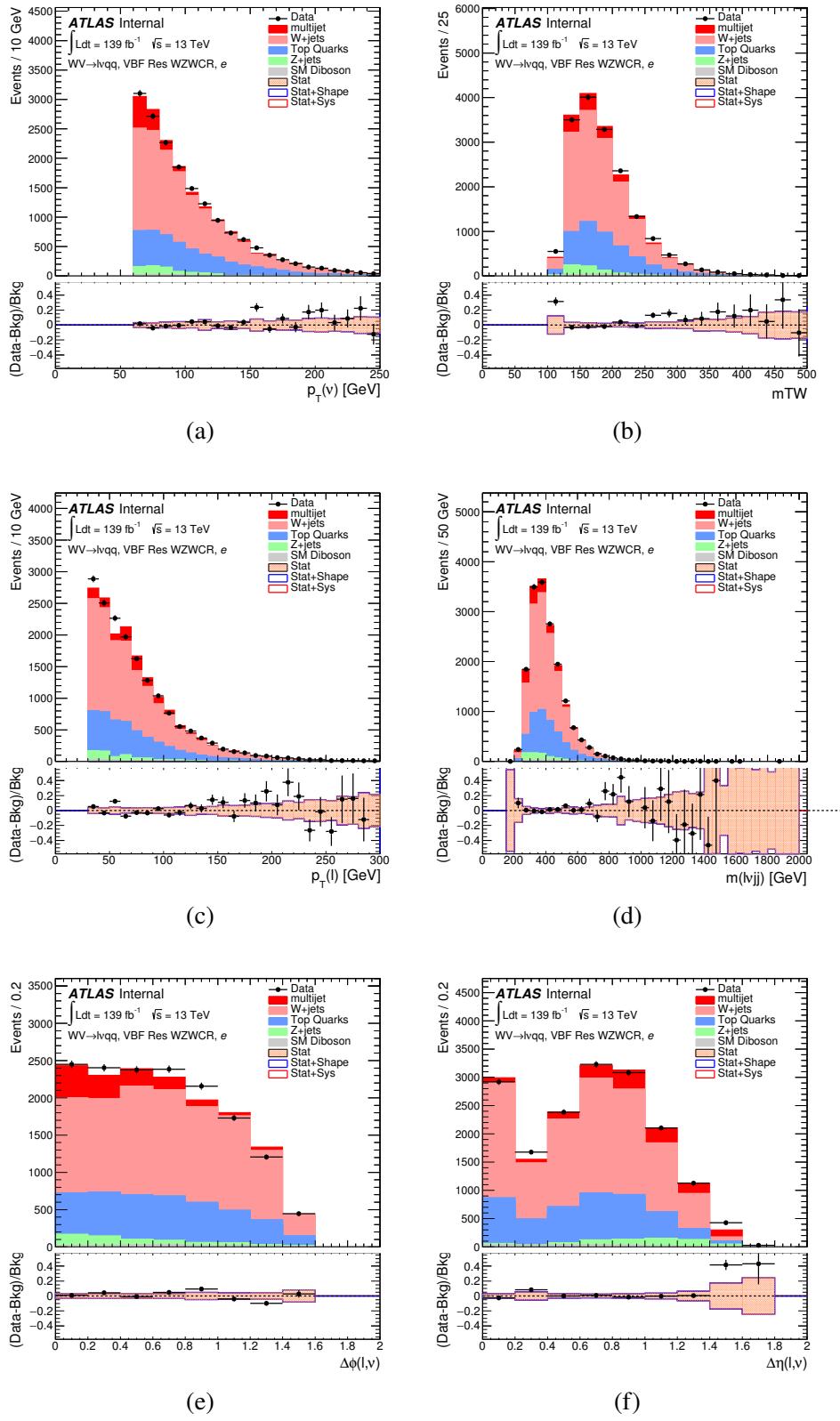


Figure 9.23: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{l\nu jj}$, lepton- ν angular distance in the VBF WZ electron channel. The MJ template is obtained from the pre-MJ-fit.

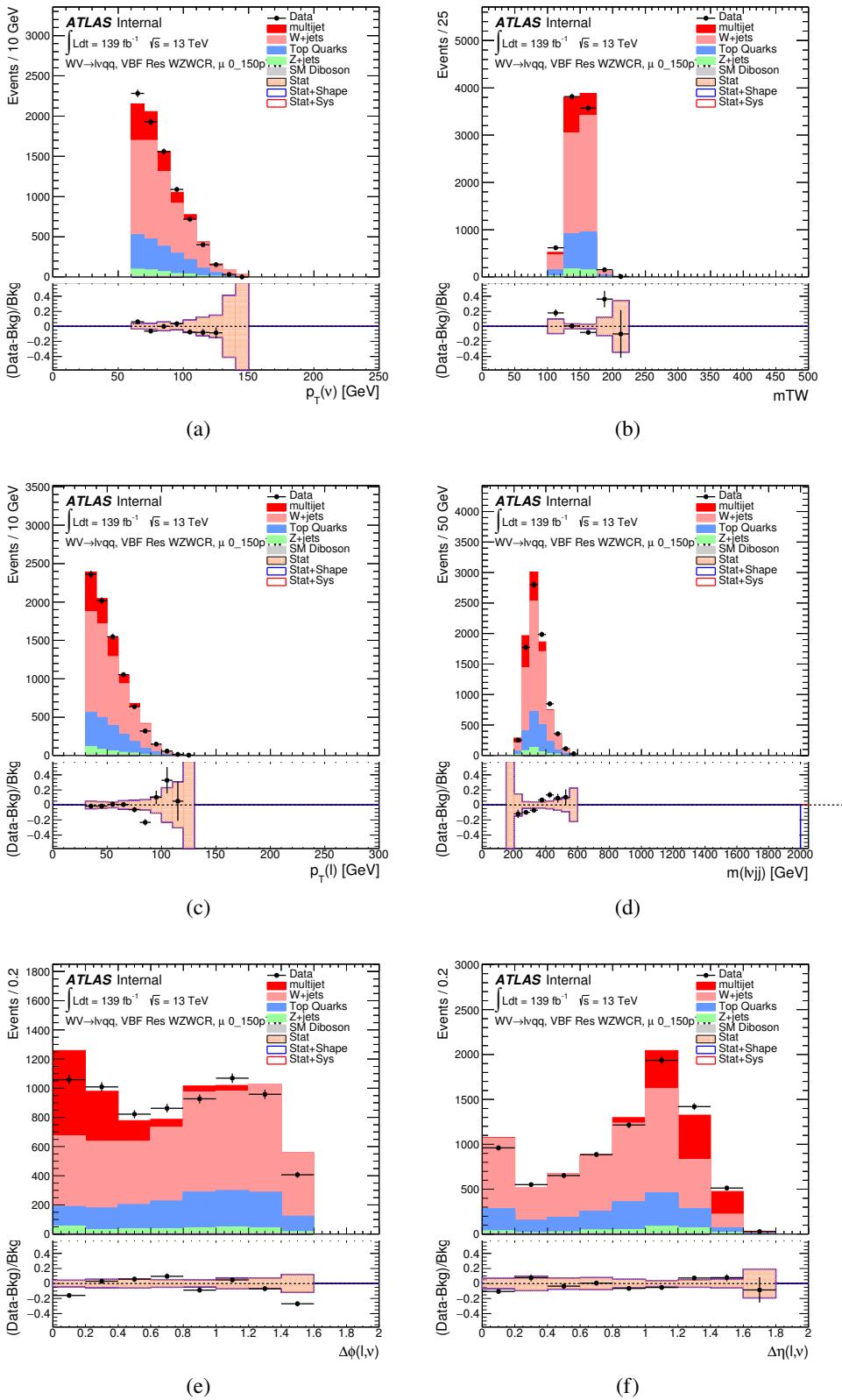


Figure 9.24: Postfit Data/MC comparison of distributions of E_T^{miss} , m_T^W , lepton and neutrino p_T , $m_{l\nu jj}$, lepton- ν angular distance in the VBF WZ muon channel. The MJ template is obtained from the pre-MJ-fit.

Full Run 2
ggF Res WWWCR

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

ggF Res WZ01bWCR

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

ggF Res WZ2bWCR

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

VBF Res WWWCR

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

VBF Res WZWCR

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

Table 9.5: 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.

1064 **Chapter 10**

1065 **Systematic Uncertainties**

1066 This section describes the sources of systematic uncertainties considered in
1067 this analysis. These uncertainties are divided into three categories: experimental
1068 uncertainties, background modeling uncertainties, and theoretical uncertainties on
1069 signal processes. In the statistical analysis each systematic uncertainty is treated
1070 as a nuisance parameter estimated on the m_{VV} distribution.

1071 **10.1 Experimental Systematics**

1072 The uncertainty on the integrated luminosity of the dataset used is 1.7% and
1073 a systematic in the final fit. This uncertainty was calculated using $x - y$ beam
1074 separation scans [ref P55].

1075 An additional source of systematic uncertainty is assigned to the pileup mod-
1076eling in MC samples. This ensures simulated detector response and particle re-
1077construction conditions are as similar as possible. The distribution of the average
1078 number of interactions per bunch crossing applied to simulation is called the μ pro-
1079file. The pileup modeling uncertainty is accounted for by re-weighting simulated
1080events so the average number of interactions per bunch crossing varies within its

1081 uncertainty due to systematics from vertex reconstruction [ref ATL-COM-SOFT-
1082 2015-119]. The associated re-weighting factors are propagated through the entire
1083 analysis chain to construct a systematic uncertainty on m_{VV} .

1084 The single-lepton and E_T^{miss} triggers used are not fully efficient, so scale factors
1085 are applied to simulation to more accurately model the data. These scale factors
1086 are given by the ratio of the distribution of offline objects before trigger selection
1087 and after trigger selection. The associated uncertainty on these scale factors are
1088 used in the final fit.

1089 Uncertainties on small-R jet energy scale and resolution are measured in-situ
1090 by calculating the response between data and simulation. This analysis uses a
1091 reduced set of JES and JER uncertainties (totaling 30 and 8 systematics, re-
1092 spectively). These reduced sets of systematics are calculated using a principal
1093 component analysis, yield largely uncorrelated independent systematics. These
1094 uncertainties on JES and JER account for the dependence on p_T , η , μ , flavor re-
1095 sponse and global sequential corrections. Systematic uncertainties associated with
1096 b -tagging are also considered. These systematics are evaluated as uncertainties on
1097 the scale factor which account for the difference in b -tagging efficiencies in data
1098 and MC, and the flavor dependence (between b, c, and light jets).

1099 The uncertainty on the p_T scale of the large-R jets is determined by comparing
1100 the jet's p_T^{calo} to p_T^{track} in di-jet simulation and data. In addition to this uncertain-
1101 ties from tracking, modeling (Pythia vs Herwig), and statistical constraints are
1102 also calculated. The large-R jet p_T resolution is given by smearing the jet p_T with
1103 a Gaussian with a 2% width.

1104 The W/Z tagging efficiency SF is estimated by comparing the tagging effi-
1105 ciency in simulation with that in data for four regions of the W/Z tagger (D_2 fail,
1106 m_J fail; D_2 pass, m_J fail; D_2 fail, m_J pass; D_2 pass, m_J pass). (Additionally,

1107 separate scale factors are determined for events with large-R jets from W bosons
 1108 and top backgrounds.) A simultaneous template fit is used to fit the signal jets
 1109 (jets initiated by W/Z bosons or top quarks) and background jets (all other jets
 1110 from the simulated backgrounds) to the data in the four regions. using the m_J
 1111 distributions. The SF for a given region is then given by:

$$SF = \frac{\epsilon_{data} = \frac{N_{fitted-signals}^{region}}{N_{all-regions}^{fitted-signals}}}{\epsilon_{MC} = \frac{N_{signal}^{region}}{N_{signal}^{all-regions}}} \quad (10.1)$$

1112 The effects of experimental and theoretical uncertainties on the efficiency scale
 1113 factor are determined by taking the ratio of efficiencies in data and simulation.
 1114 By taking this ratio the uncertainties not arising for jet mass and D_2 cancel.

1115 Lepton identification, reconstruction, isolation systematic uncertainties are de-
 1116 termined by reconstructing the Z mass peak with a tag and probe method. The
 1117 lepton energy and momentum scales are also measured with the Z mass peak.

1118 As E_T^{miss} is calculated using all the physics objects in the event, all those objects
 1119 associated errors result in an uncertainty on E_T^{miss} . Additionally, the unassociated
 1120 tracks used to construct E_T^{miss} contribute to the uncertainty on E_T^{miss} .

1121 10.2 Theory Systematics

1122 Theoretical uncertainties for signal and background processes arise from un-
 1123 certainties in the parameters used in Monte Carlo simulation. In particular for
 1124 the $t\bar{t}$, $W/Z+jets$, diboson backgrounds and signal samples, the QCD scale, PDF,
 1125 generator and hadronization uncertainties were evaluated. To assess the QCD
 1126 scale uncertainty the renormalization and factorization scales were scaled up and
 1127 down by a factor of two at the event generation stage of sample production. Un-

certainties due to the choice of the parton distribution functions were evaluated by re-weighting samples from the nominal PDF to a set of error PDFs which account for the uncertainty of the fits used to produce the PDF set. In addition to this, samples are re-weighted to different PDF sets to account for the arbitrariness of the PDF choice. The difference between the m_{WV} distributions using different event generators is assessed by comparing samples generated with different generators. Similarly, the uncertainty in hadronization models is accounted for by comparing samples created using different hadronization models (e.g. $t\bar{t}$ Powheg is compared to AMC@NLO, $W + jets$ compares Sherpa and MadGraph+Pythia samples). Figures 10.2 - 10.8 show the impact of these uncertainties on the $t\bar{t}$ and $W/Z + jets$ backgrounds. Additionally, contributions to the diboson background for the VBF analysis were found to be small and were accounted for by including a 5(10)% systematic in the diboson normalization in the final fit.

The normalization of the $t\bar{t}$ and $W+jets$ processes impact the multijet template shape. The impact of these normalizations was assessed by including a shape systematic on the multijet background from varying the $t\bar{t}$ and $W+jets$ normalization factors.

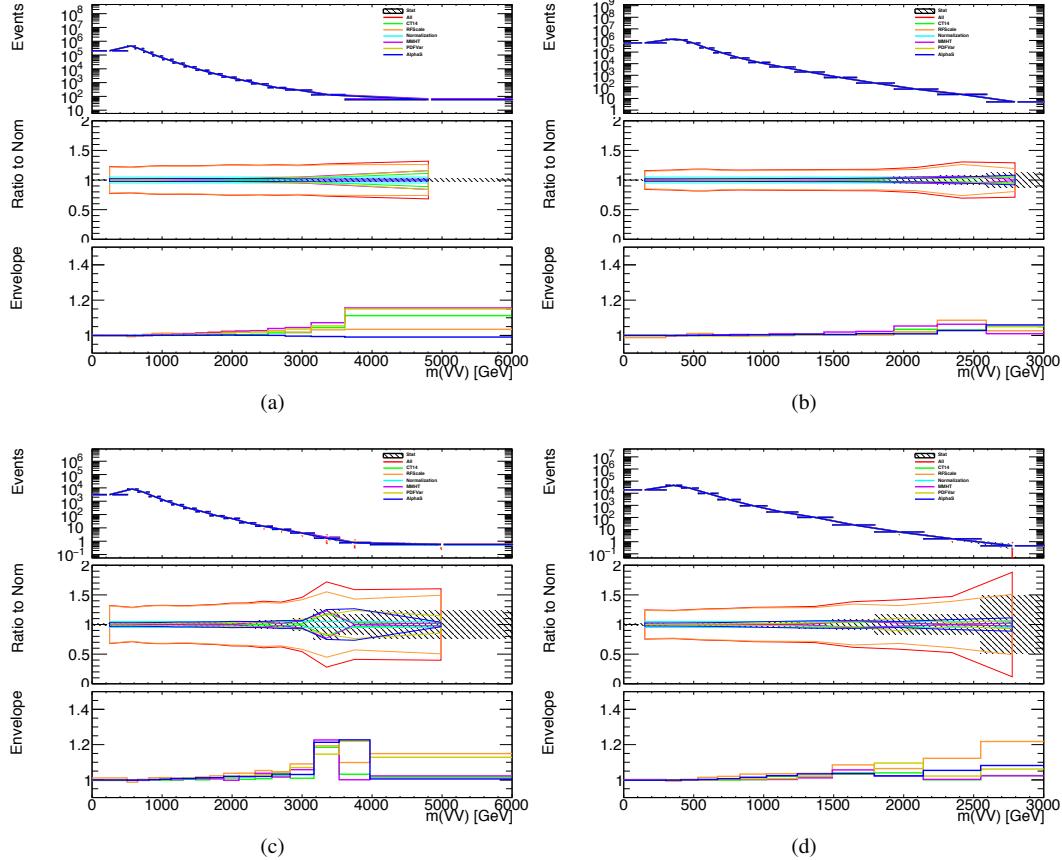


Figure 10.1: The $W/Z + \text{jet}$ systematics for the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The top subplot shows the nominal and variation distributions/bands, the middle shows the ratio of the two, and the final shows just the shape of the envelope (the final uncertainty).

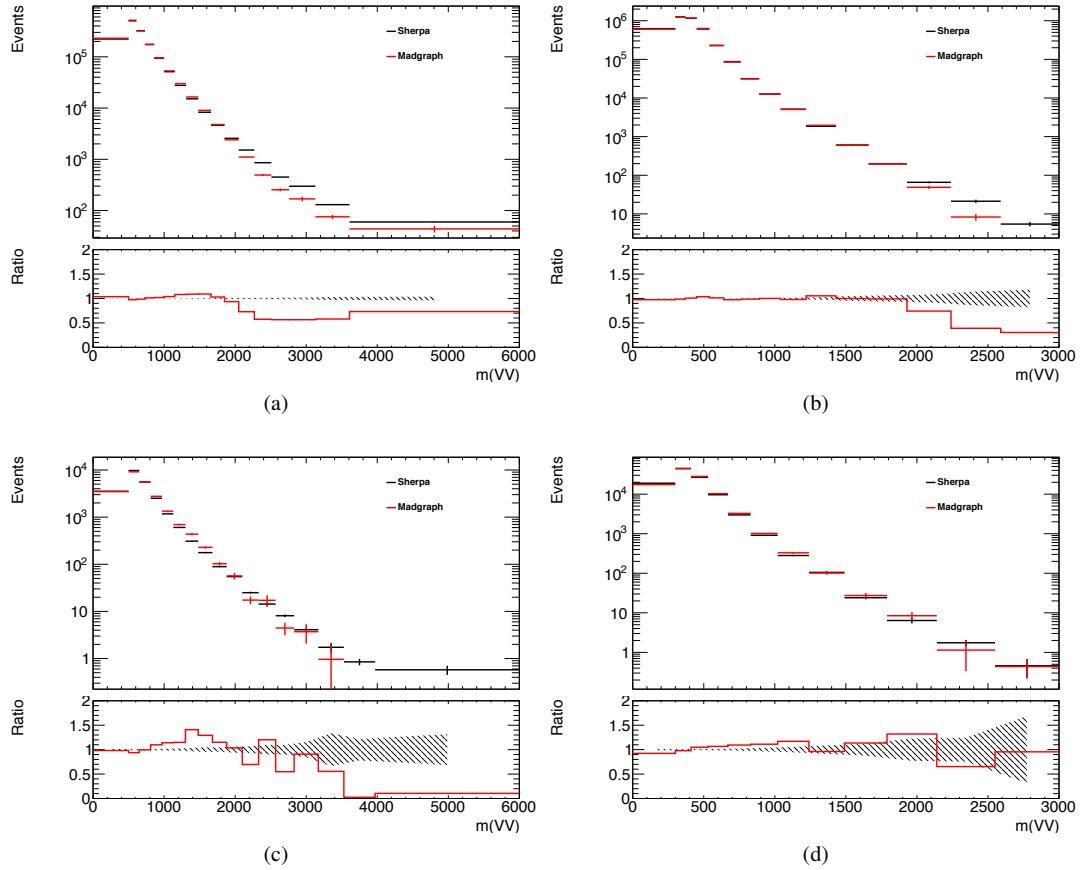


Figure 10.2: The two-point generator comparison between Sherpa and MadGraph for the $W/Z + \text{jet}$ samples in the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The normalization of the Madgraph sample is set to the Sherpa value to consider only shape effects. The bottom inset shows the ratio of the two.

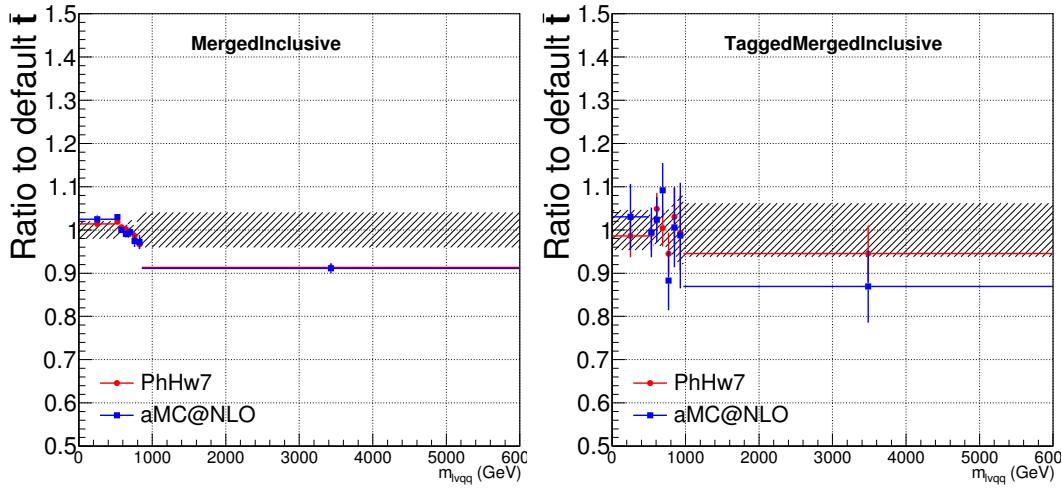


Figure 10.3: Ratio between the variations of generator (red) and hadronization (blue) variations for the Merged regime for $t\bar{t}$ sample.

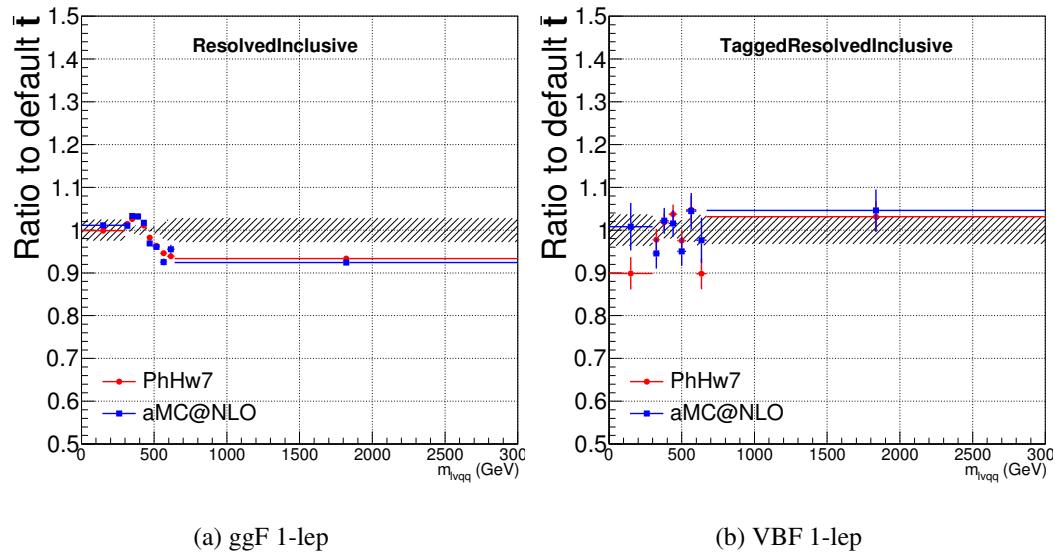
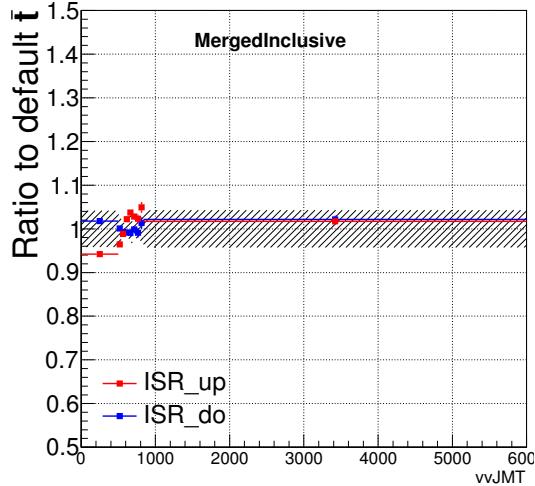
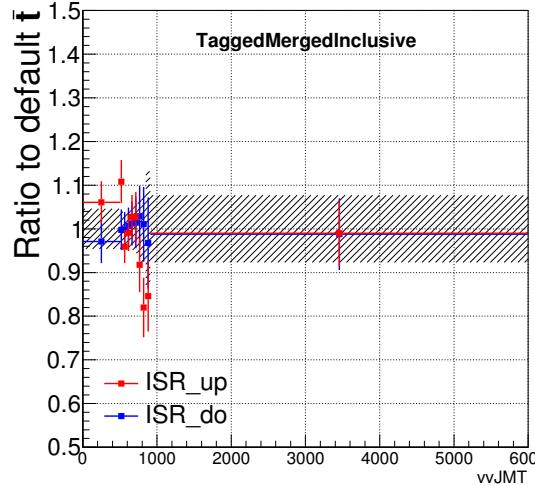


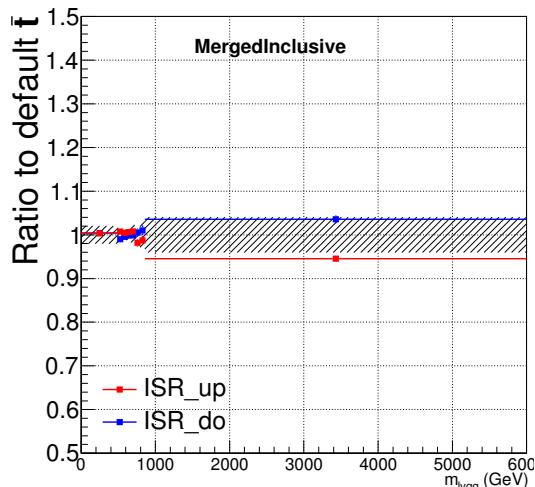
Figure 10.4: Ratio between the variations of generator (red) and hadronization (blue) variations for the Resolved regime for $t\bar{t}$ sample.



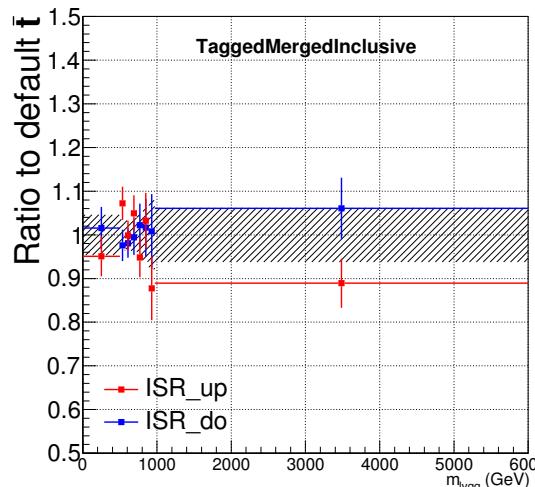
(a) ggF 0-lep



(b) VBF 0-lep



(c) ggF 1-lep



(d) VBF 1-lep

Figure 10.5: Ratio between the variations of ISR up (red) and down (blue) variations for the Merged regime for $t\bar{t}$ sample.

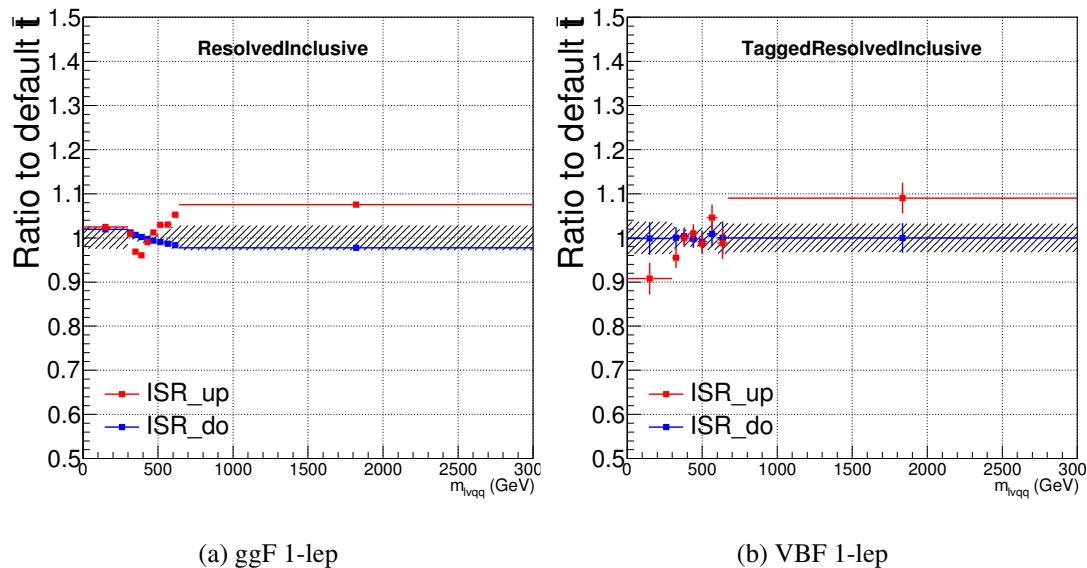


Figure 10.6: Ratio between the variations of ISR up (red) and down (blue) variations for the Resolved regime for $t\bar{t}$ sample.

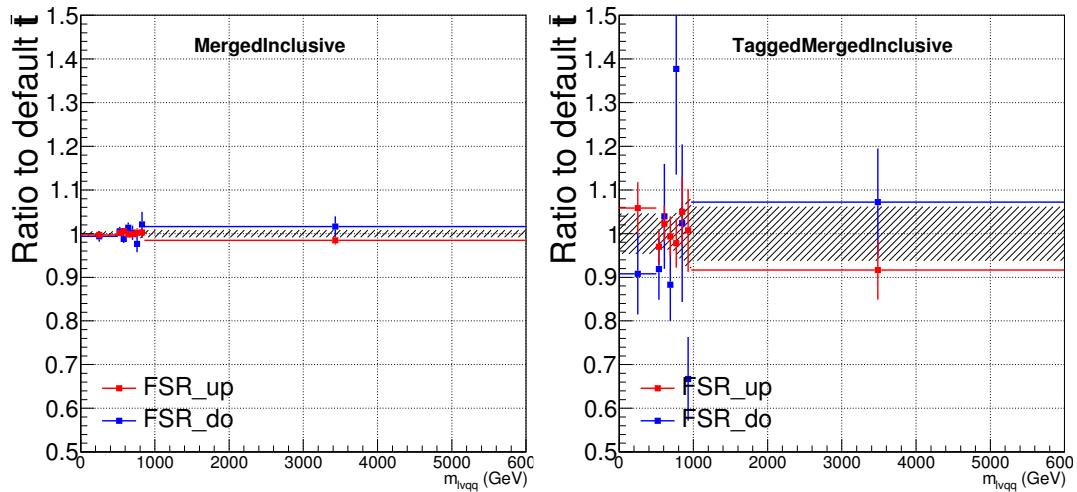


Figure 10.7: Ratio between the variations of FSR up (red) and down (blue) variations for the Merged regime for $t\bar{t}$ sample.

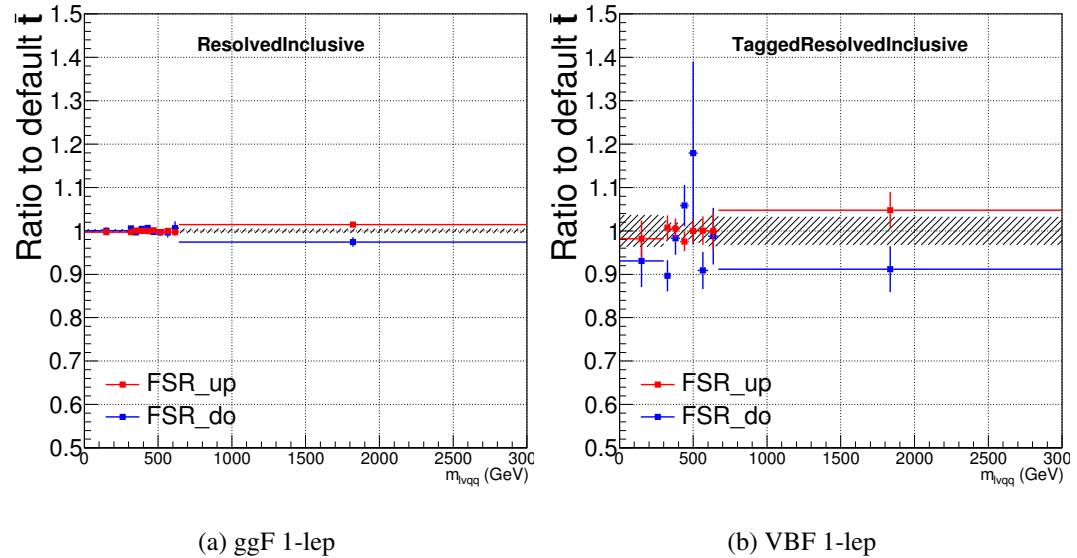


Figure 10.8: Ratio between the variations of FSR up (red) and down (blue) variations for the Resolved regime for $t\bar{t}$ sample.

₁₁₄₅ **Chapter 11**

₁₁₄₆ **Statistical Analysis**

₁₁₄₇ To determine the compatibility of the data collected with the proposed reso-
₁₁₄₈ nances a statistical procedure based on a likelihood function is used. A discovery
₁₁₄₉ test is used to measure the compatibility of the observed data with the back-
₁₁₅₀ ground only hypothesis. If the observed data is sufficiently incompatible with the
₁₁₅₁ background only hypothesis, this could indicate a discovery. In the absence of
₁₁₅₂ discovery, upper limits on the signal strength parameter, μ , are assessed using the
₁₁₅₃ CLs method.

₁₁₅₄ **11.1 Likelihood Function Definition**

₁₁₅₅ The likelihood function is product of Poisson probabilities for all analysis bins
₁₁₅₆ and systematic constraints:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_c \prod_i \frac{(\mu s_{ci}(\boldsymbol{\theta}) + b_{ci}(\boldsymbol{\theta}))^{n_{ci}}}{n_{ci}!} e^{-(\mu s_{ci}(\boldsymbol{\theta}) + b_{ci}(\boldsymbol{\theta}))} \prod_k (\theta'_k | \theta_k) \quad (11.1)$$

₁₁₅₇ Here c are the analysis channels considered and i runs over all the $m_{\ell\nu qq}$ bins

1158 used in the fit. The signal strength parameter, μ , multiplies the expected signal
1159 yield in each analysis bin, s_{ci} . The background content for channel c and bin i is
1160 given by b_{ci} . The dependence of signal and background predictions on system-
1161 atic uncertainties is described by the aforementioned set of nuisance parameters
1162 $\boldsymbol{\theta}$, which are parameterized by Gaussian or log-normal priors denoted here as
1163 θ_k . Statistical uncertainties of the simulated bin contents are also included as
1164 systematic uncertainties. Most systematics are correlated among all the analysis
1165 regions and considered to be independent from each other. The validity of this
1166 assumption is checked by evaluating the covariance of nuisance parameters.

1167 11.2 Fit Configuration

1168 The binning of $m_{\ell\nu qq}$ in signal regions for likelihood fit is determined by the
1169 statistical uncertainty of signal mass width. For each signal mass point, the signal
1170 mass resolution is given by the fitted Gaussian width of the $m_{\ell\nu qq}$. The fitted
1171 signal widths are then fit to a line to give a parameterized signal mass width, as
1172 shown in Figures 11.1 and 11.2. Bin widths are set first to this parameterized
1173 signal mass resolution. Then if the statistical uncertainty of the data or simulated
1174 background is more than 50%, bins are merged until the statistical uncertainty is
1175 less than 50%. All control regions contain only a single bin.

1176 For this analysis, each signal model is fit in the Merged and Resolved chan-
1177 nels for the relevant signal production mode simultaneously. The $W + \text{jets}$ and $t\bar{t}$
1178 normalizations are given by the best fit values in the overall fit and these fitted
1179 normalizations are then applied to those backgrounds in the SRs.

1180 Systematics may be affected by low statistics, leading to unsmooth m_{VV} dis-
1181 tributions with unphysically large fluctuations. This can lead to artificial pulls
1182 and constraints in the fit. To remove such issues a multi-step smoothing pro-

1183 cedure is applied to all systematic variation distributions in all regions. First,
1184 distributions are rebinned until the statistical error per bin is at least 5%. Next
1185 all local extrema are identified. The bins around smallest extrema are iteratively
1186 merged until only four local extrema remain. Then distributions are rebinned so
1187 that statistical uncertainties in each bin are $< 5\%$.

1188 For some systematics, up and down variations may be in the same direction
1189 with respect to the nominal distributions. This causes the variations to not cover
1190 the nominal choice, and the interpretation of the confidence interval is skewed as
1191 the nominal distribution should be bracketed by the up and down variations. This
1192 asymmetry may also lead to unconstrained systematics in the fit. To handle such
1193 asymmetric systematics, if the up and down variation for a given systematic are in
1194 the same direction for at least three m_{VV} bins the variation is averaged for those
1195 bins. The averaging procedure replaces bin-by-bin the up and down variation bins
1196 by $b_{\pm}^{new} = b_{nom} \pm \frac{|b_+ - b_-|}{2}$, where b_{nom} is the nominal bin content and b_{\pm} are the
1197 original up and down variation bin content. The same procedure is also applied to
1198 any variations where the integral of the difference between the up/down variation
1199 and the nominal distribution is twice that of the other down/up variation, further
1200 ensuring variations are symmetric around the nominal distribution.

1201 Finally, systematics that have a negligible effect on the m_{VV} distribution are
1202 not considered in the fit. Shape systematics where no bin in the variational dis-
1203 tribution deviates more than 1% from the nominal distribution (after normalizing
1204 all histograms to the nominal) are not included in the fit. Also, statistical bin
1205 uncertainties $< 1\%$ are ignored.

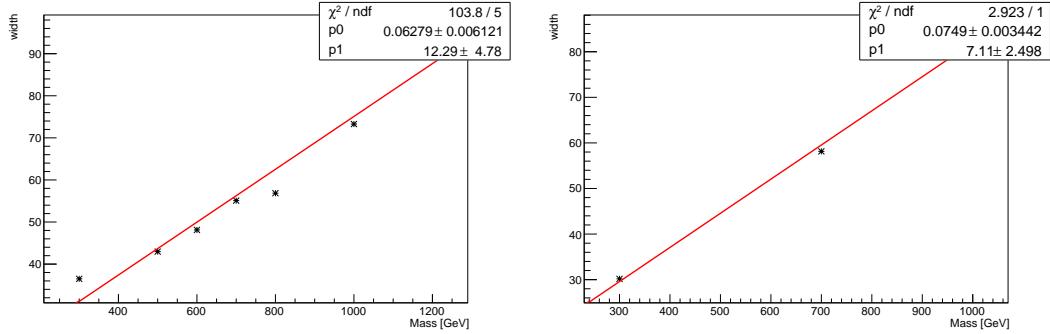


Figure 11.1: The HVT signal mass resolution as a function of mass fit with a straight line in the Resolved ggF region (left) and VBF (right) region.

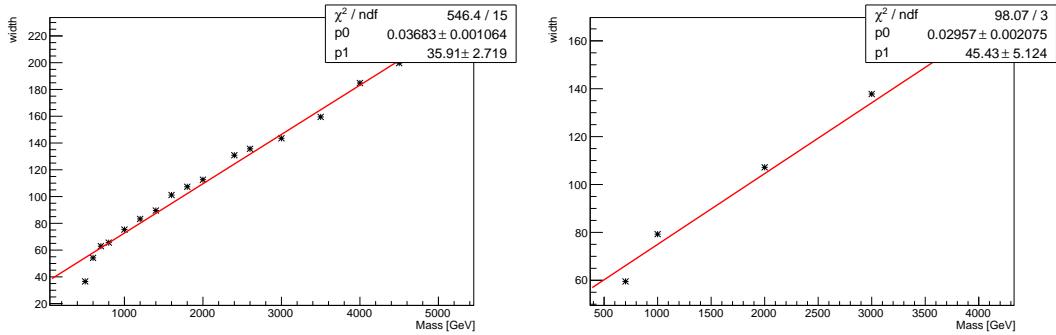


Figure 11.2: The HVT signal mass resolution as a function of mass fit with a straight line in the Merged ggF region (left) and VBF (right) region.

1206 11.3 Best Fit μ

1207 The best fit signal strength parameter is denoted by $\hat{\mu}$ and calculated by
 1208 maximizing the likelihood function with respect to all systematics and μ . The
 1209 corresponding set of systematics that maximize the likelihood are given by $\hat{\mu}$.
 1210 The first term in the likelihood is maximized when the expected number of signal
 1211 and background events is equal to the number of events in data ($n_{ci} = \mu s_{ci} +$
 1212 b_{ci}). Thus, by maximizing the likelihood, the fit determines values of μ and θ
 1213 that give the best agreement between expected and measured event yields. The

1214 second term in the likelihood is a penalty term which decreases the likelihood
1215 when systematics are shifted from their nominal values. This prevents the fit
1216 from profiling systematics in unphysical ways to maximize the likelihood. The
1217 uncertainty on μ is calculated by varying μ up and down until the natural log of
1218 the likelihood function shifts by one-half.

1219 11.4 Discovery Test

1220 To determine if the observed dataset is consistent with tested signal model a
1221 likelihood ratio is constructed:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad (11.2)$$

1222 The denominator in this equation is the maximized value of \mathcal{L} over all system-
1223 atics and μ . The numerator is the maximized likelihood over all systematics for
1224 a given μ value, where the maximized systematics are given by $\hat{\mu}_\mu$. To test for
1225 the existence of signal the observed dataset the null hypothesis (H_0) is defined as
1226 the background only hypothesis and the alternate hypothesis includes signal and
1227 background (H_1). This test quantifies the compatibility of observed data with
1228 H_0 by calculating a p-value representing the probability of observing data as dis-
1229 crepant or more than the observed data under the H_0 . The test statistic used to
1230 calculate this p-value is given by (r_0):

$$r_0 = \begin{cases} -2 \ln \lambda(0), \hat{\mu} > 0 \\ +2 \ln \lambda(0), \hat{\mu} < 0 \end{cases} \quad (11.3)$$

1231 The expected distribution of the the test statistic under H_0 ($f(r_0|0)$) is used to
1232 calculate the p-value:

$$p_0 = \int_{r_0, obs}^{\infty} f(r_0 | 0) dr_0 \quad (11.4)$$

1233 Small p-values indicate the observed data is poorly described by H_0 . This
 1234 equivalent Z-score of a given p-value is usually used to further quantify the agree-
 1235 ment between the observed data and H_0 . The Z-score is given by the number of
 1236 standard deviations away from the mean of a Gaussian distribution, the integral
 1237 of the upper tail of the distribution would equal the p-value. Mathematically:

$$Z = \Phi^{-1}(1 - p_0) \quad (11.5)$$

1238 where Φ is the Gaussian cumulative distribution function. The statistical
 1239 significance of these tests are expressed as the Z -score. In particle physics, 3σ is
 1240 considered evidence for new phenomena and 5σ is the threshold for discovery.

1241 11.5 Exclusion Limits

1242 In the absence of discovery, upper limits on the signal strength, μ are set using
 1243 the CLs method [cite P60]. The test statistic for this test, q_μ , is constructed as:

$$\tilde{\lambda}_\mu = \begin{cases} \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \hat{\mu} > 0 \\ \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(0, \hat{\theta}_0)}, \hat{\mu} < 0 \end{cases} \quad (11.6)$$

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu), \hat{\mu} < \mu \\ +2 \ln \tilde{\lambda}(\mu), \hat{\mu} > \mu \end{cases} \quad (11.7)$$

1244 As defined, larger values of q_μ correspond to increasing incompatibility between
 1245 the observed data and the background + signal hypothesis. The observed value
 1246 of the test statistic, $q_{\mu, obs}$, is then compared to its expected distribution, f , to

1247 calculate p-values to assess the likelihood of the background+signal hypothesis.

1248 Using these distributions, CL_s values are computed as:

$$1249 \quad CL_{s+b} = \int_{q_{\mu,obs}}^{\infty} f(q_{\mu}|\mu) dq_{\mu} \quad (11.8)$$

$$1250 \quad CL_b = \int_{q_0^{obs}}^{\infty} f(q_{\mu}|\mu = 0) dq_{\mu} \quad (11.9)$$

$$1251 \quad CL_s = \frac{CL_{s+b}}{CL_b} \quad (11.10)$$

1251 CL_{s+b} is the p-value for the signal + background hypothesis and CL_b is the
1252 p-value for the background only hypothesis. The CL_s value is interpreted as
1253 the probability to observe the background + signal hypothesis normalized to the
1254 probability of background-only hypothesis. Normalizing by CL_b prevents setting
1255 artificially strong exclusion limits due to downward fluctuations in data.

1256 For a given signal hypothesis, μ values are scanned simultaneously over all
1257 m_{WV} bins to find the μ value that yields $CL_s=0.05$, meaning the likelihood of
1258 finding data more incompatible with the signal+background hypothesis (relative
1259 to the background only hypothesis) is 5%. The 95% upper limit on the cross
1260 section is then calculated as the product of the μ value found, branching ratio,
1261 and theory cross section.

Part V

1262

Results

1263

1264 **Chapter 12**

1265 **Statistical Interpretation**

1266 **12.1 Discovery Tests**

1267 To test for the existence of signal in the observed dataset, the discovery tests
1268 discussed earlier are used to calculate p-values as a function of resonance mass.
1269 The results of these tests are shown in Figures 12.1 - 12.5. Across the different
1270 DY signals the largest excesses are $\sim 2.2\sigma$ at 600 GeV and 1.8σ at 2 TeV. The
1271 largest excesses for VBF signals are $< 2.5\sigma$ at for 1 TeV resonances. As these
1272 deviations do not constitute discoveries, upper limits on μ are calculated.

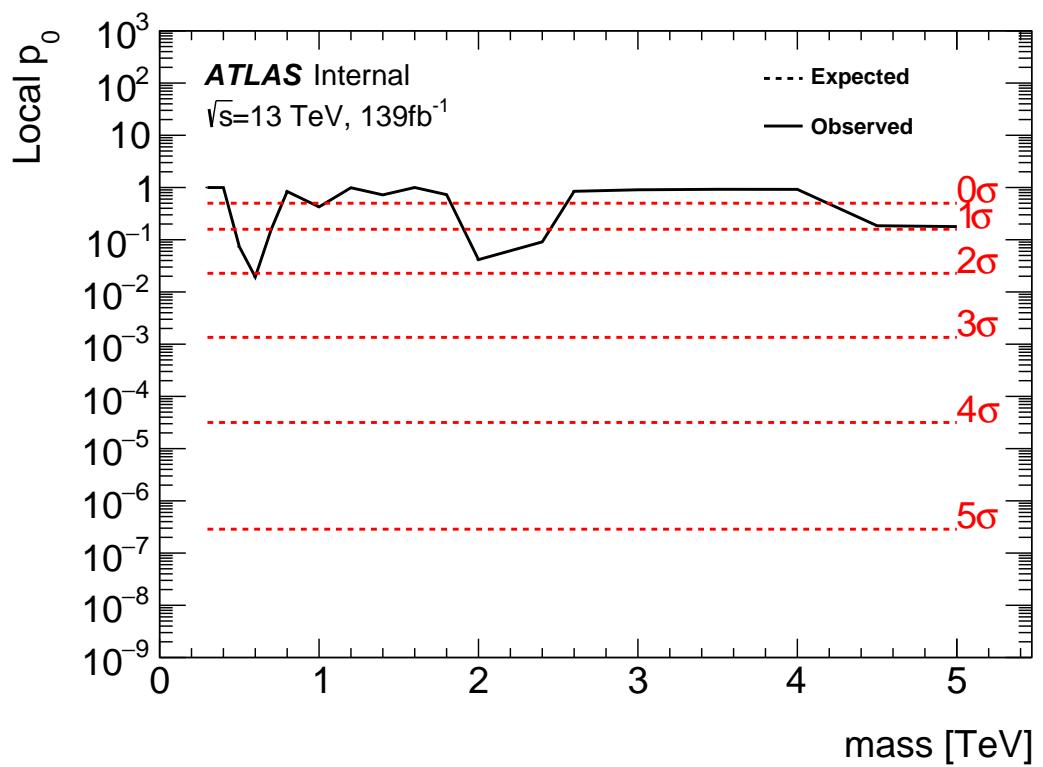


Figure 12.1: These plots show the measured p_0 value as a function of resonance mass for HVT Z' DY production.

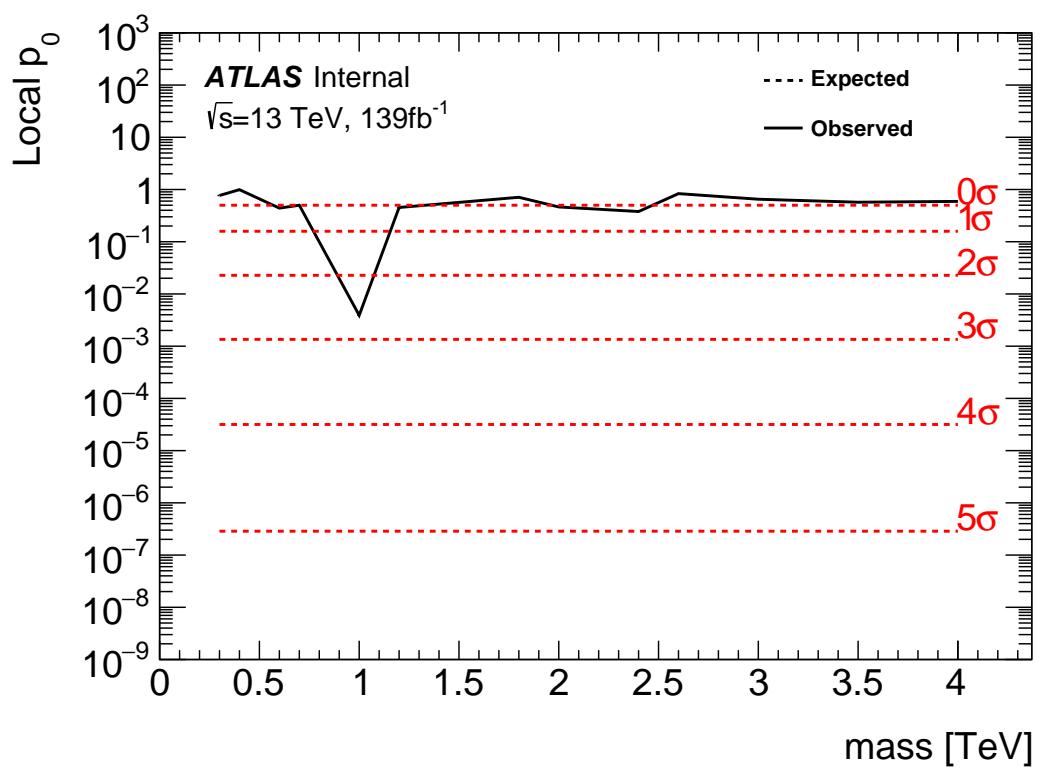


Figure 12.2: These plots show the measured p_0 value as a function of resonance mass for HVT Z' VBF production.

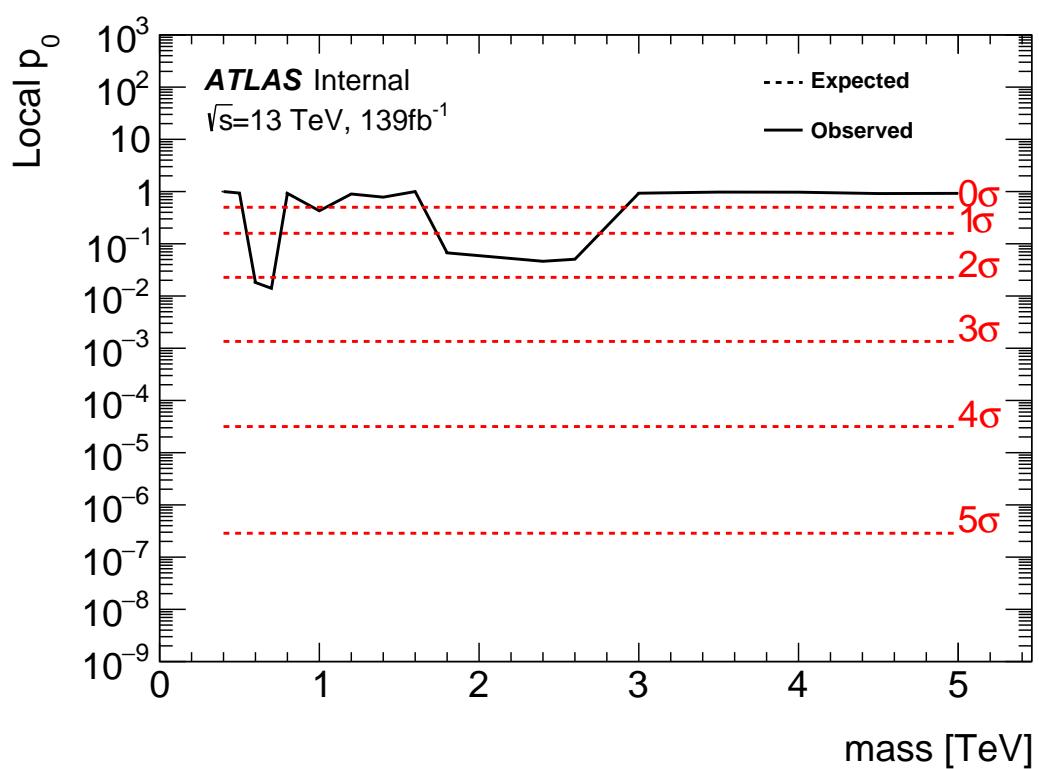


Figure 12.3: These plots show the measured p_0 value as a function of resonance mass for HVT W' DY production.

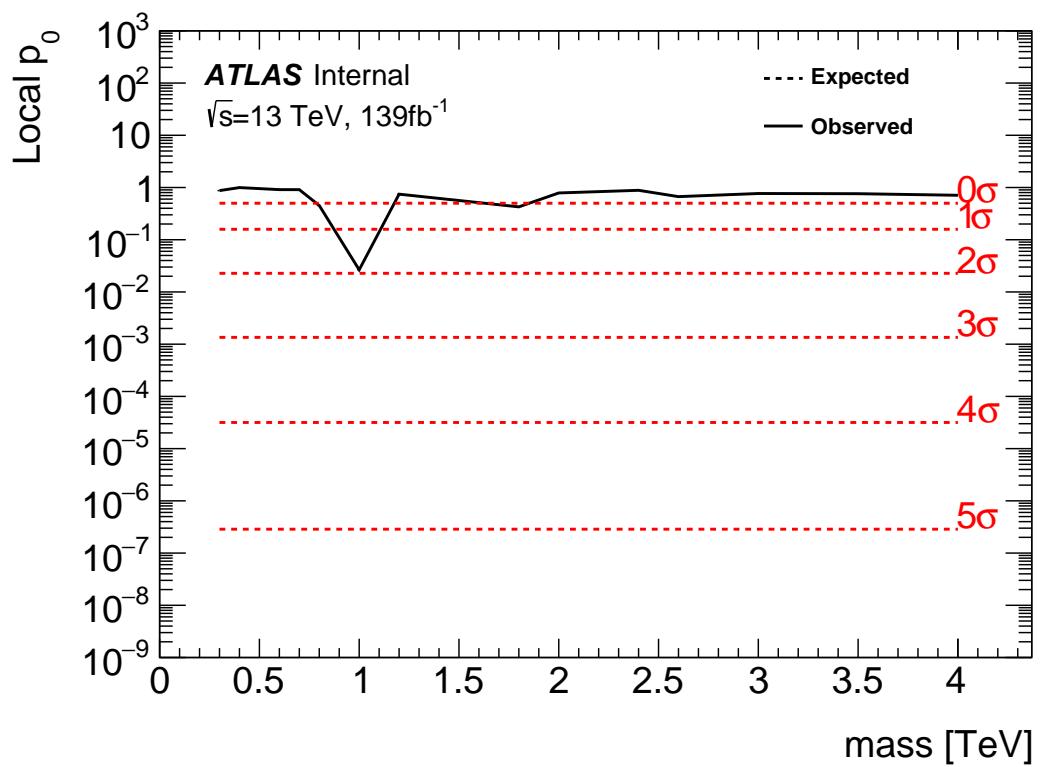


Figure 12.4: These plots show the measured p_0 value as a function of resonance mass for HVT W' VBF production.

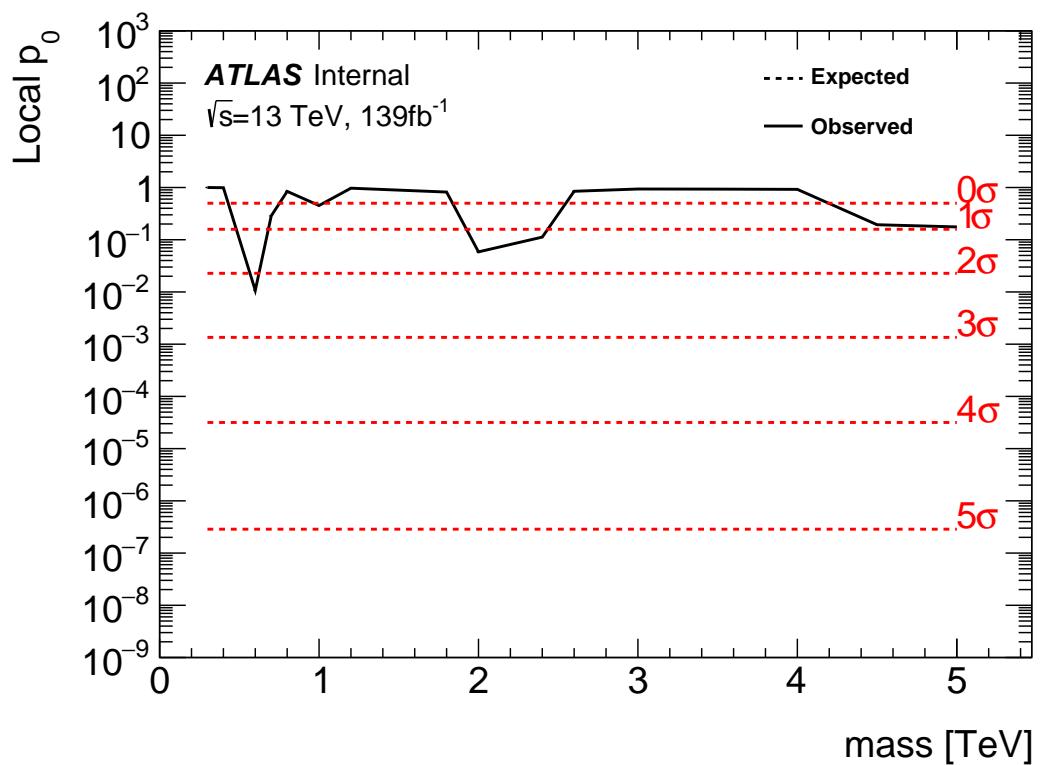


Figure 12.5: These plots show the measured p_0 value as a function of resonance mass for the RS Graviton DY production.

12.2 Systematic Profiling and Correlations

1273 The ranked systematics and their fitted values are shown for the different analysis regions in Figure 12.6 - 12.8. Note that background normalizations for $W+jets$ and $t\bar{t}$ are left free to float in the fit. This means the nominal normalization values are at one and the uncertainties are not plotted in the ranked plots. Overall, systematics are not pulled outside their uncertainties, especially for highly ranked nuisance parameters.

1280 The correlation between systematics are shown in Figures ???. Correlations 1281 between background normalization are expected. The remaining systematic correlations are not very strong or unexpected.

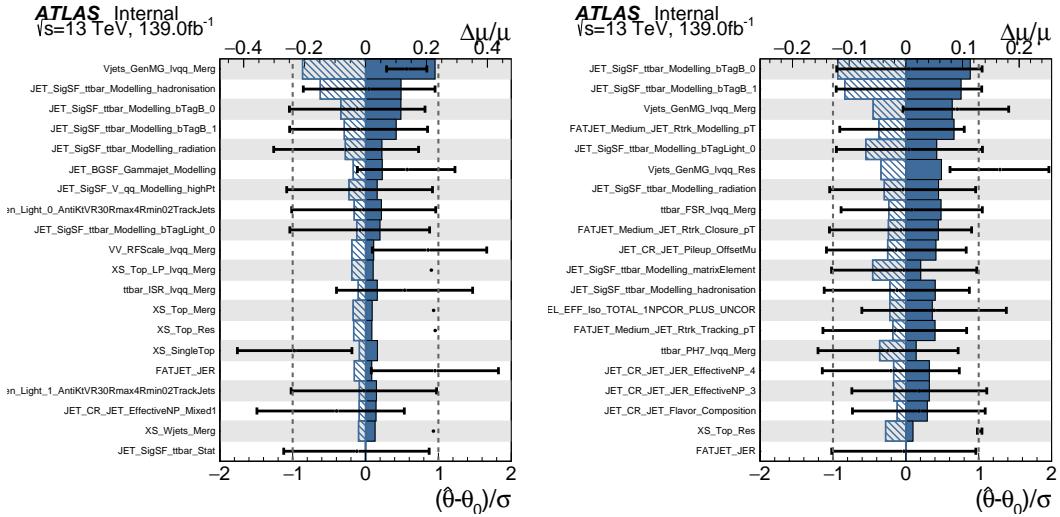


Figure 12.6: Ranked systematics and their fitted values for WW DY (right) and VBF (left) selections.

1282

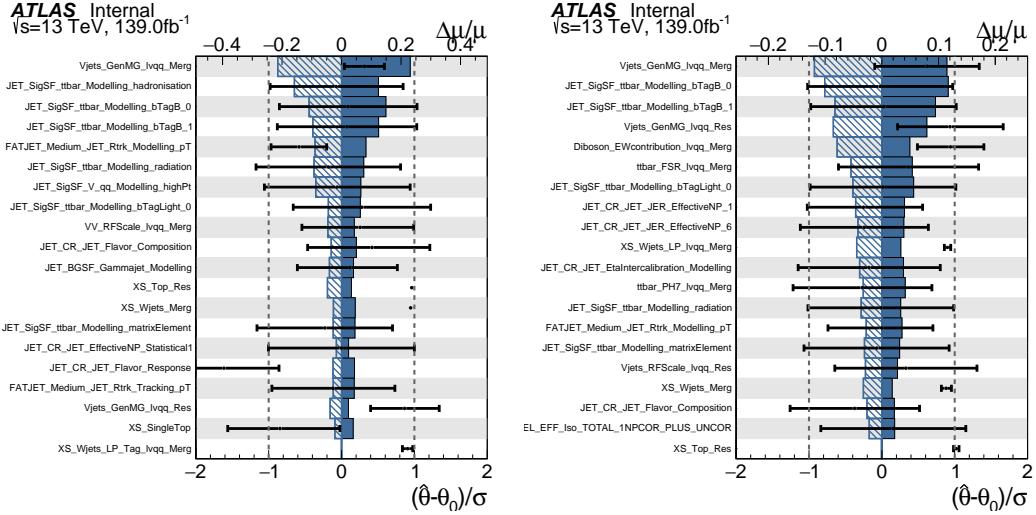


Figure 12.7: Ranked systematics and their fitted values for WZ DY (right) and VBF (left) selections.

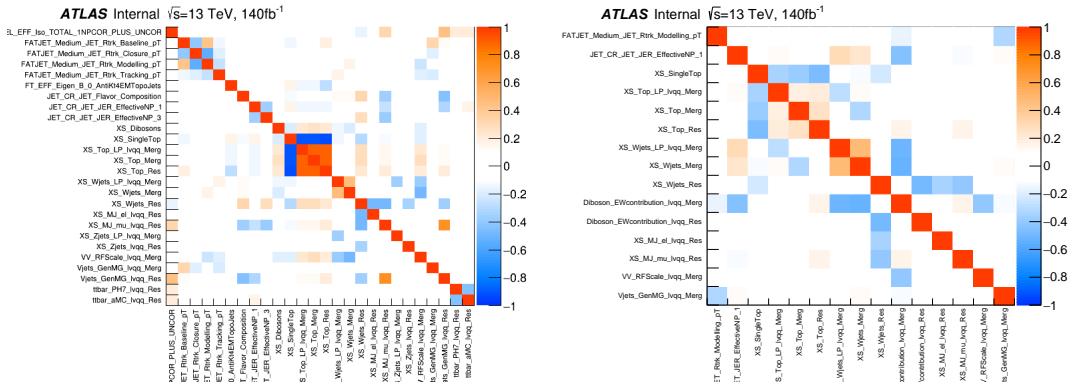


Figure 12.8: Correlations between systematics for WW DY (right) and VBF (left) selections.

12.3 Expected and Measured Yields

The yield tables for the four analysis regions are shown in Tables ?? - ?? . The fitted background normalizations are shown in Tables ??-?? . The control region $m_{\ell\nu qq}$ distributions are shown in Figures 12.9 - 12.12. The signal region $m_{\ell\nu qq}$ distributions are shown in Figures 12.13 - ?? .

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

Table 12.1: 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 ± 2277.99
Muon Multi-jet	-	-	27347.10 ± 2950.07
Diboson	1508.48 ± 154.20	2758.24 ± 284.50	9038.55 ± 728.69
Single-top	1756.59 ± 306.69	2913.18 ± 515.93	20511.74 ± 3523.47
$t\bar{t}$	13134.00 ± 238.30	21815.37 ± 334.98	140157.77 ± 2636.96
$W+jets$	40654.84 ± 333.65	87657.76 ± 501.96	665909.12 ± 4420.62
$Z+jets$	768.72 ± 77.97	1759.87 ± 178.96	16512.46 ± 1673.23
Total	57822.63 ± 540.40	116904.42 ± 862.16	894556.75 ± 7492.20
Data	57699.00	117306.00	895362.00
	HP Tagged WCR	LP Tagged WCR	Resolved Tagged WCR
Electron Multi-jet	-	-	384.58 ± 57.11
Muon Multi-jet	-	-	602.93 ± 190.12
Diboson	30.22 ± 4.69	48.95 ± 7.16	264.64 ± 28.24
Single-top	308.44 ± 56.19	371.59 ± 69.43	5752.39 ± 1029.97
$t\bar{t}$	1683.82 ± 48.73	2041.48 ± 70.00	58431.49 ± 614.30
$W+jets$	583.55 ± 75.37	1109.45 ± 85.78	11891.68 ± 903.01
$Z+jets$	13.19 ± 1.34	23.06 ± 2.34	324.74 ± 32.85
Total	2619.22 ± 106.00	3594.53 ± 130.90	77652.45 ± 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 ± 28.45	346.78 ± 35.85	650.85 ± 65.56
Single-top	3107.99 ± 538.03	2250.64 ± 385.41	9606.87 ± 1698.22
$t\bar{t}$	30992.40 ± 562.33	26954.21 ± 413.89	91893.59 ± 1728.91
$W+jets$	2236.29 ± 18.35	4874.03 ± 27.91	16122.97 ± 107.03
$Z+jets$	71.54 ± 7.26	155.50 ± 15.81	577.71 ± 58.54
Total	36697.66 ± 779.03	34581.16 ± 567.59	118851.98 ± 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 ± 1.13	8.75 ± 1.16	34.06 ± 4.98
Single-top	105.87 ± 20.65	119.66 ± 22.68	656.89 ± 132.96
$t\bar{t}$	1904.75 ± 50.61	1483.86 ± 47.05	17965.33 ± 188.87
$W+jets$	32.36 ± 4.28	85.74 ± 6.96	489.01 ± 37.13
$Z+jets$	1.27 ± 0.13	1.93 ± 0.20	19.14 ± 1.94
Total	2053.98 ± 54.84	1699.93 ± 52.70	19164.43 ± 234.01
Data	2047.00	1708.00	19143.00

Table 12.2: 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 ± 1175.56
Muon Multi-jet	-	-	17004.81 ± 1834.40
Diboson	3041.17 ± 273.77	2266.35 ± 212.79	14724.12 ± 1224.31
Single-top	2123.28 ± 373.83	1379.35 ± 240.92	18336.88 ± 3082.47
$t\bar{t}$	11678.86 ± 213.63	8906.34 ± 136.88	112669.24 ± 2122.46
$W+jets$	22741.32 ± 191.47	41726.76 ± 240.56	342934.00 ± 2280.21
$Z+jets$	442.03 ± 44.84	849.79 ± 86.42	9271.83 ± 939.52
Total	40026.65 ± 546.81	55128.59 ± 432.90	522723.03 ± 5131.71
Data	40193.00	54735.00	521813.00
	HP Tagged SR	LP Tagged SR	Resolved Tagged SR
Electron Multi-jet	-	-	199.22 ± 29.58
Muon Multi-jet	-	-	393.43 ± 124.06
Diboson	102.58 ± 11.59	65.44 ± 8.05	624.07 ± 58.10
Single-top	178.21 ± 33.62	155.53 ± 28.95	3470.39 ± 617.48
$t\bar{t}$	1017.93 ± 31.95	706.76 ± 26.20	38189.30 ± 401.91
$W+jets$	325.58 ± 41.62	575.36 ± 43.29	6161.96 ± 467.71
$Z+jets$	7.81 ± 0.80	11.62 ± 1.19	183.36 ± 18.55
Total	1632.11 ± 63.39	1514.70 ± 58.86	49221.74 ± 884.06
Data	1699.00	1559.00	48919.00

Table 12.3: 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 ± 137.82
Muon Multi-jet	-	-	601.46 ± 182.74
Diboson	107.45 ± 45.20	166.87 ± 68.11	292.10 ± 235.29
Single-top	78.19 ± 18.22	132.71 ± 31.93	879.82 ± 216.89
$t\bar{t}$	400.71 ± 28.35	569.70 ± 48.88	5067.51 ± 155.69
$W+jets$	864.49 ± 63.44	1940.80 ± 89.41	18563.70 ± 408.99
$Z+jets$	19.51 ± 2.00	46.63 ± 4.77	795.20 ± 80.89
Total	1470.35 ± 84.89	2856.71 ± 126.74	27098.28 ± 594.01
Data	1495.00	2898.00	27120.00
	HP TCR	LP TCR	Resolved TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	14.95 ± 6.61	27.57 ± 14.12	24.33 ± 20.32
Single-top	68.31 ± 16.17	58.93 ± 13.56	278.60 ± 73.04
$t\bar{t}$	496.60 ± 31.72	401.23 ± 32.13	3834.49 ± 104.60
$W+jets$	50.68 ± 4.19	144.02 ± 7.86	450.01 ± 11.87
$Z+jets$	1.32 ± 0.14	5.35 ± 0.55	29.96 ± 3.07
Total	631.87 ± 36.45	637.10 ± 38.44	4617.39 ± 129.77
Data	636.00	634.00	4615.00
	HP SR	LP SR	Resolved SR
Electron Multi-jet	-	-	596.34 ± 91.52
Muon Multi-jet	-	-	481.01 ± 144.48
Diboson	148.84 ± 48.64	181.42 ± 67.30	395.52 ± 318.06
Single-top	79.49 ± 19.80	56.82 ± 14.89	782.07 ± 190.79
$t\bar{t}$	338.42 ± 24.14	236.80 ± 20.88	4261.70 ± 138.98
$W+jets$	501.13 ± 39.36	1347.76 ± 64.50	11445.73 ± 291.49
$Z+jets$	9.25 ± 0.95	28.77 ± 2.95	567.66 ± 57.94
Total	1077.13 ± 69.93	1851.57 ± 96.73	18530.03 ± 523.88
Data	1096.00	1846.00	18530.00

Table 12.4: 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 ± 132.75
Muon Multi-jet	-	-	618.45 ± 196.90
Diboson	92.92 ± 41.77	145.90 ± 64.26	228.62 ± 114.62
Single-top	71.13 ± 16.29	118.82 ± 27.98	1209.87 ± 281.64
$t\bar{t}$	427.80 ± 29.72	509.19 ± 46.57	6860.87 ± 254.83
$W+jets$	871.68 ± 64.22	2020.67 ± 93.54	19088.50 ± 442.10
$Z+jets$	19.58 ± 2.01	47.39 ± 4.85	800.19 ± 82.02
Total	1483.11 ± 83.79	2841.97 ± 125.92	29676.50 ± 644.96
Data	1495.00	2898.00	29755.00
	HP TCR	LP TCR	Resolved TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	10.12 ± 4.51	12.73 ± 6.55	14.23 ± 7.49
Single-top	51.57 ± 12.31	35.07 ± 8.17	169.21 ± 44.54
$t\bar{t}$	470.06 ± 28.97	298.99 ± 25.28	2414.75 ± 75.42
$W+jets$	49.64 ± 4.17	109.69 ± 6.16	378.22 ± 12.05
$Z+jets$	1.28 ± 0.13	4.81 ± 0.50	17.62 ± 1.83
Total	582.67 ± 32.07	461.30 ± 28.05	2994.03 ± 88.75
Data	584.00	459.00	3001.00
	HP SR	LP SR	Resolved SR
Electron Multi-jet	-	-	444.65 ± 67.99
Muon Multi-jet	-	-	397.29 ± 125.59
Diboson	109.66 ± 44.13	112.28 ± 46.45	265.75 ± 139.43
Single-top	63.16 ± 15.20	48.02 ± 11.56	872.16 ± 205.00
$t\bar{t}$	348.95 ± 24.34	190.68 ± 17.75	5134.25 ± 193.57
$W+jets$	467.21 ± 37.12	973.73 ± 47.91	10226.83 ± 254.67
$Z+jets$	8.15 ± 0.84	23.62 ± 2.43	558.48 ± 57.25
Total	997.13 ± 64.42	1348.33 ± 70.06	17899.41 ± 432.98
Data	1018.00	1313.00	17826.00

Table 12.5: 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 12.6: 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 12.7: 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 12.8: 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 12.9: Fitted background normalizations for $t\bar{t}$ and $W+jets$ backgrounds for the VBF WZ analysis region.

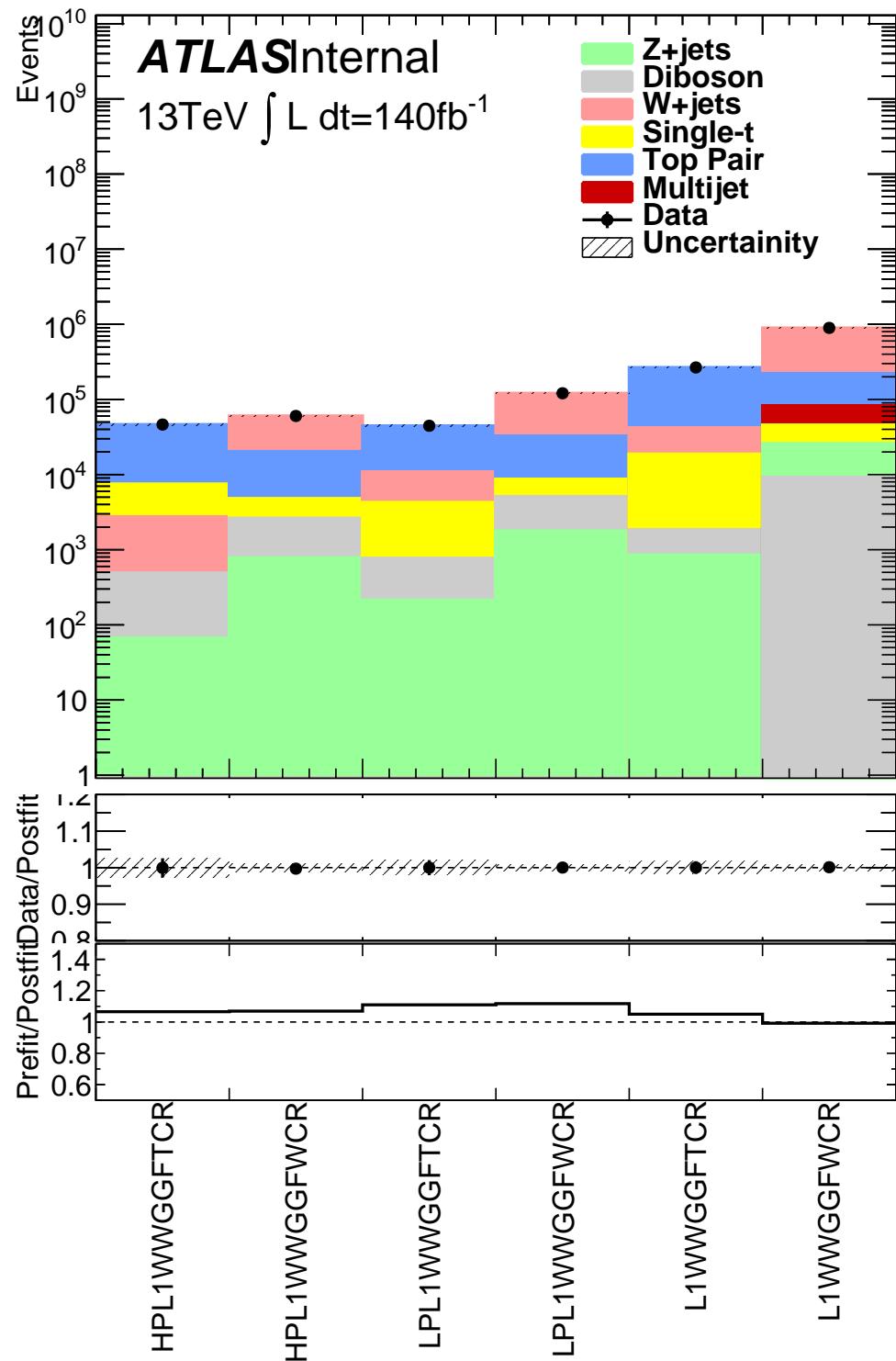


Figure 12.9: This figure shows the distribution of $m_{\ell\nu qq}$ in the DY WW control regions.

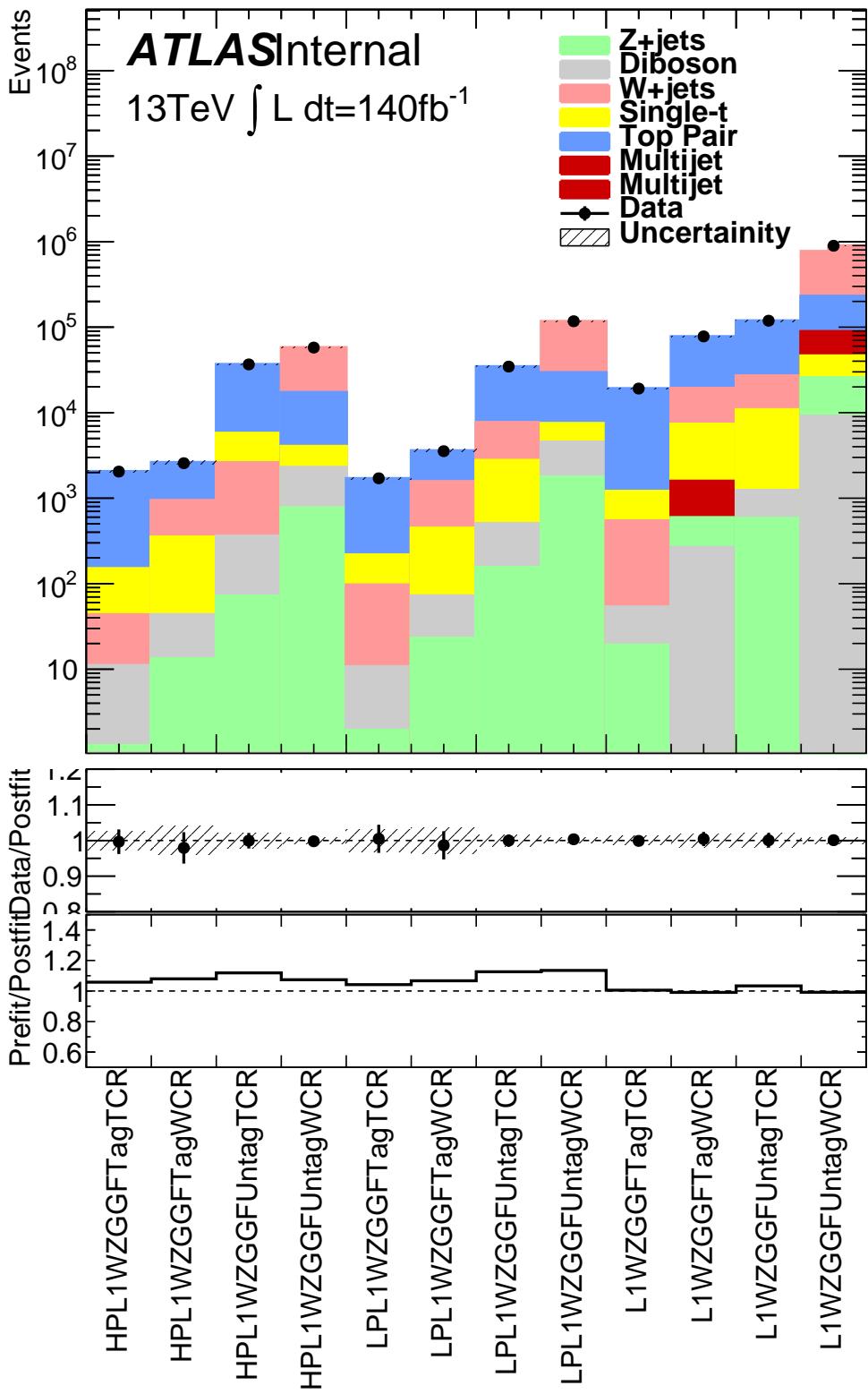


Figure 12.10: This figure shows the distribution of $m_{\ell\nu qq}$ in the DY WZ control regions.

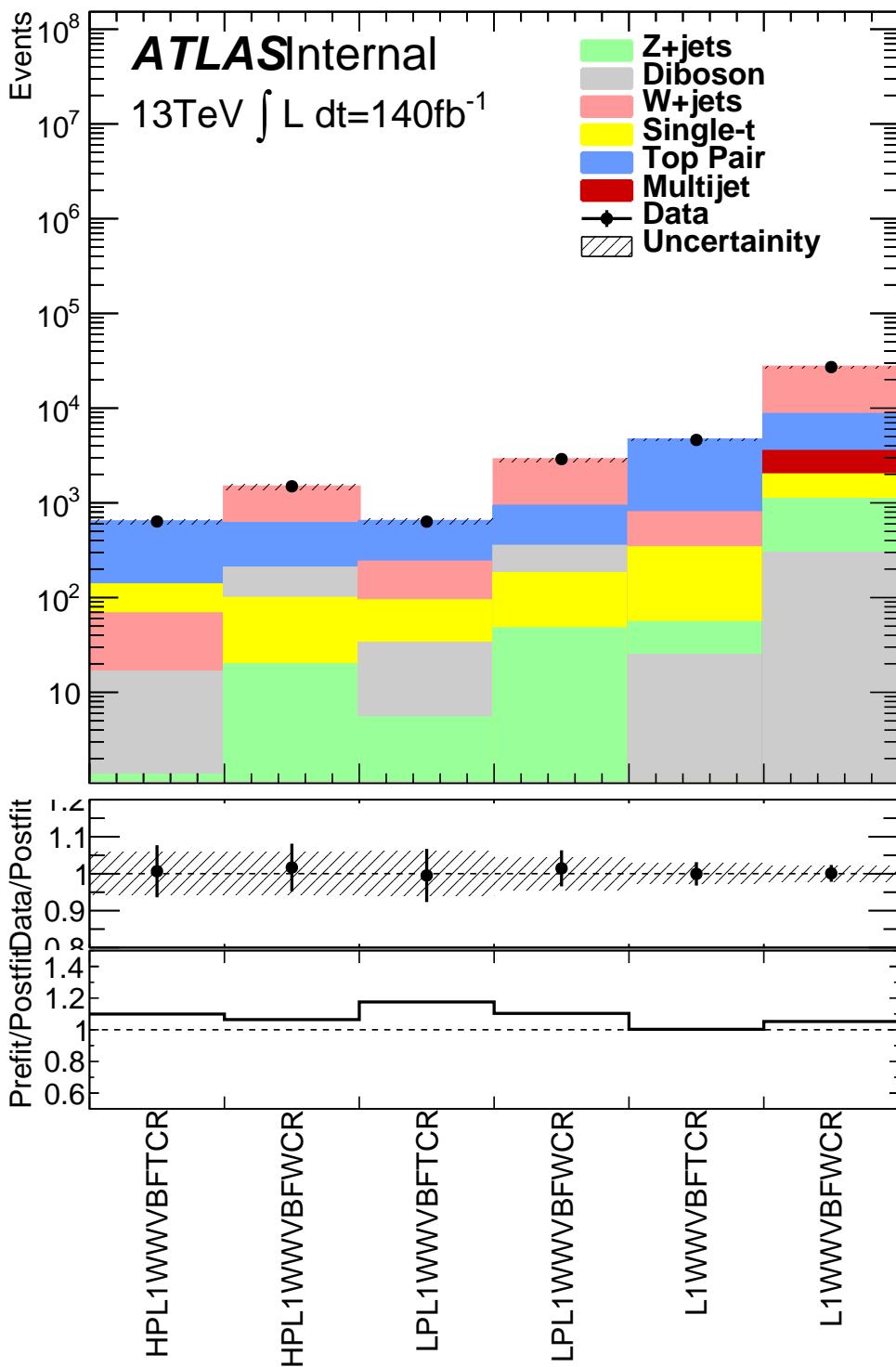


Figure 12.11: This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WW control regions.

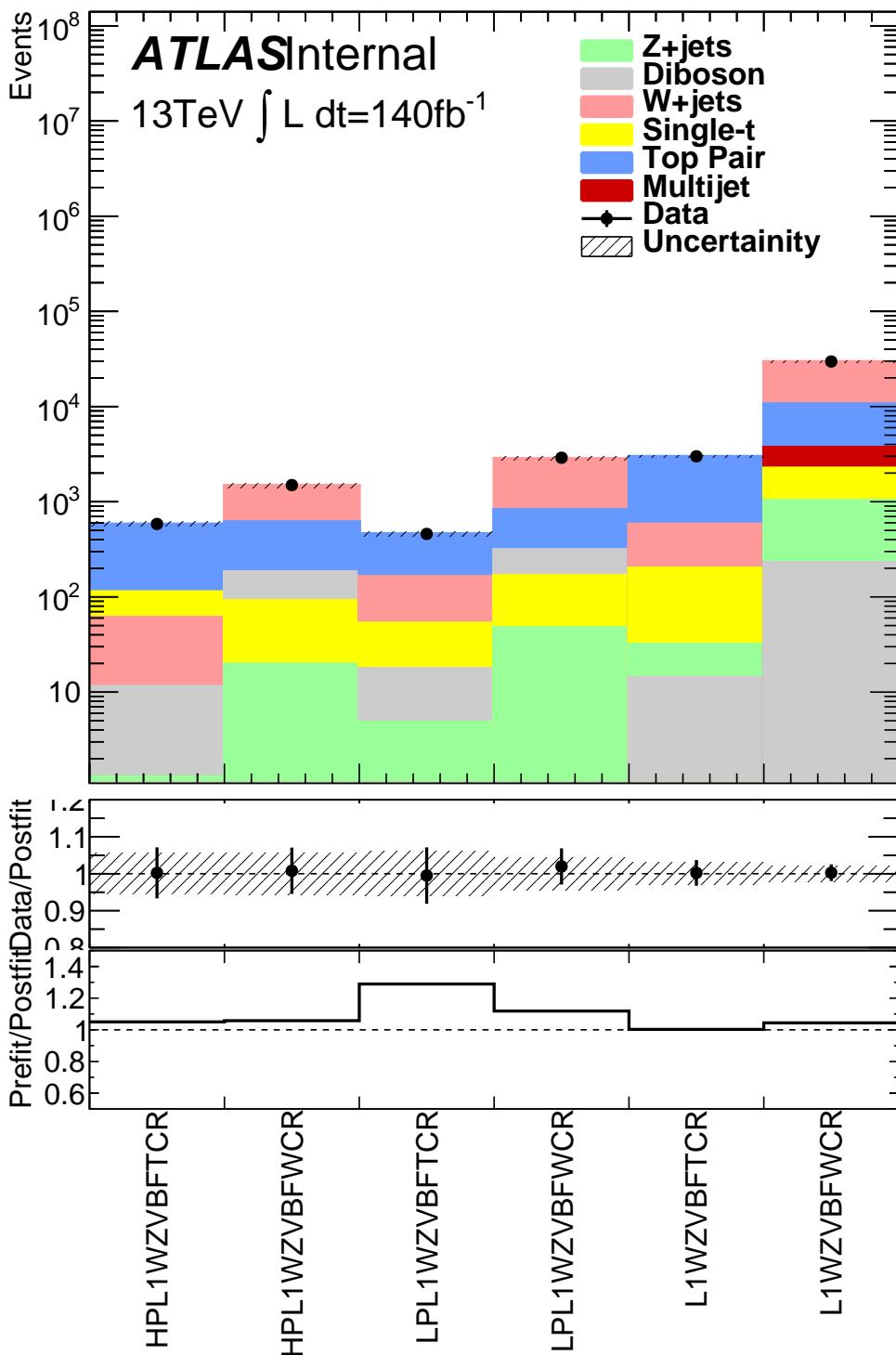


Figure 12.12: This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WZ control regions.

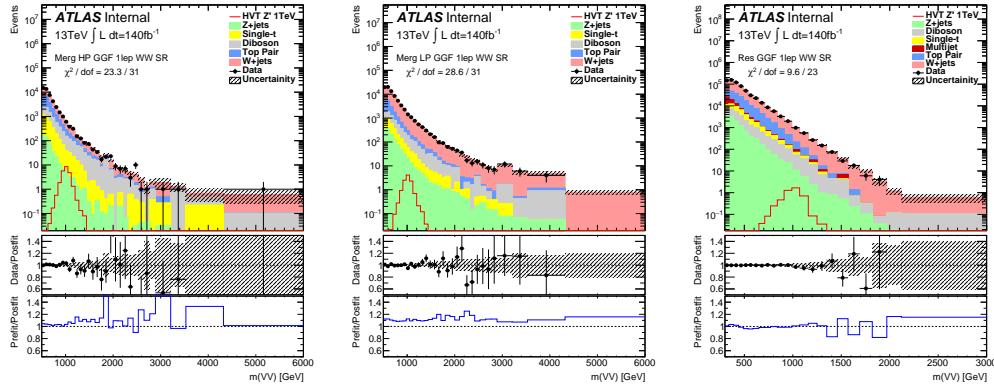


Figure 12.13: This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF WW signal regions.

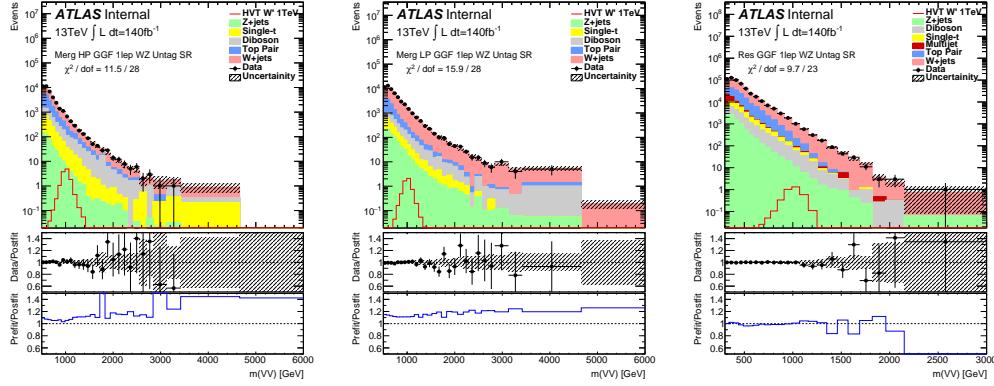


Figure 12.14: This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF WZ Untag signal regions.

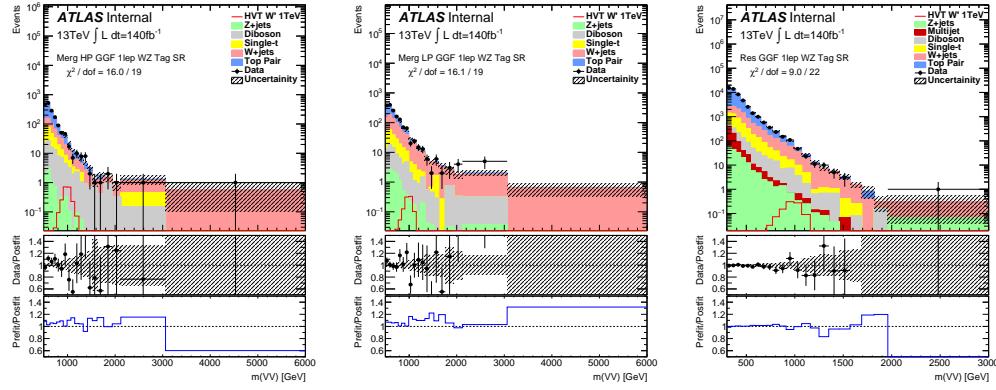


Figure 12.15: This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF WZ Tag signal regions.

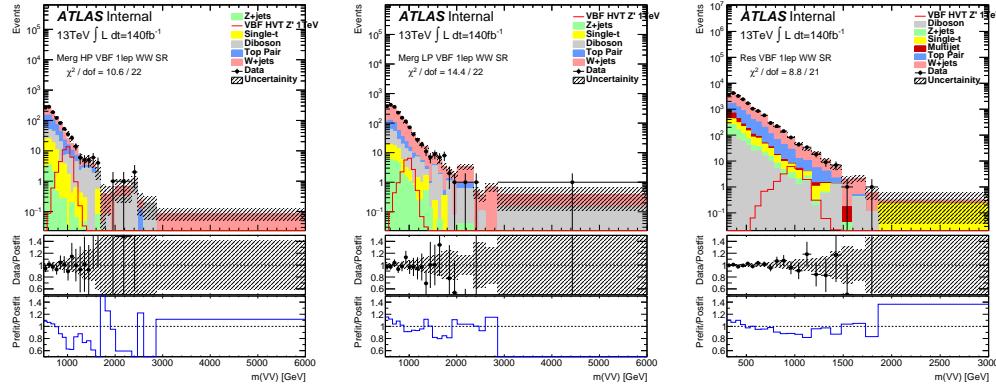


Figure 12.16: This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WZ Tag signal regions.

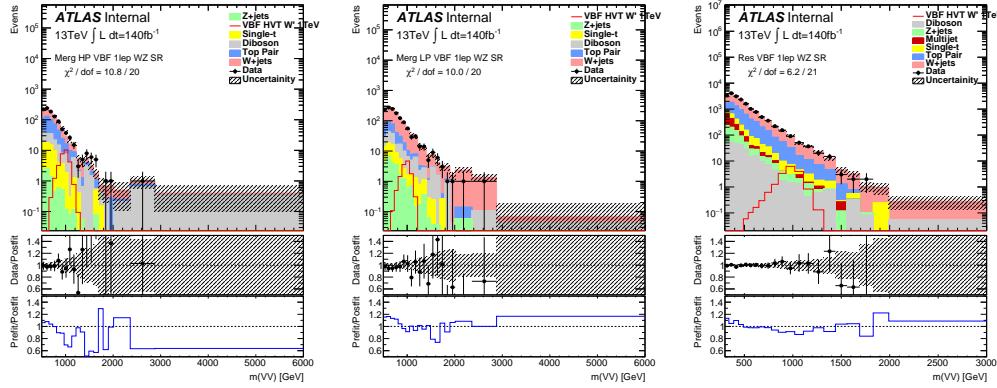


Figure 12.17: This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WZ Tag signal regions.

12.4 Limits

Using the exclusion limits tests discussed previously, exclusion limits are set on μ and consequently cross-sections for different signal models. Exclusion limits for the models considered are shown in Figure 12.18 - 12.20. These limits exclude HVT Model A $W' < 3.4$ TeV and $Z' < 3.3$ TeV and Model B $W' < 3.7$ TeV and $Z' < 3.7$ TeV. Randall Sundrum Gravitons are excluded for masses below 1.6 TeV .

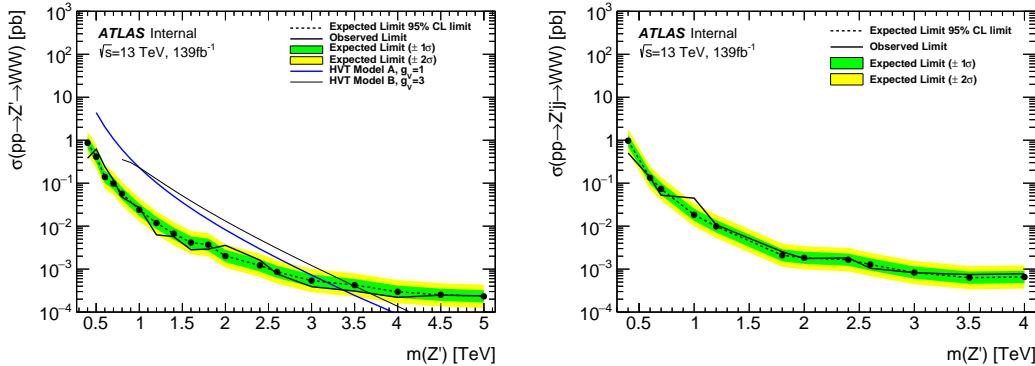


Figure 12.18: This figure shows theory, expected and observed limits for HVT W' DY (left) and VBF (right) production.

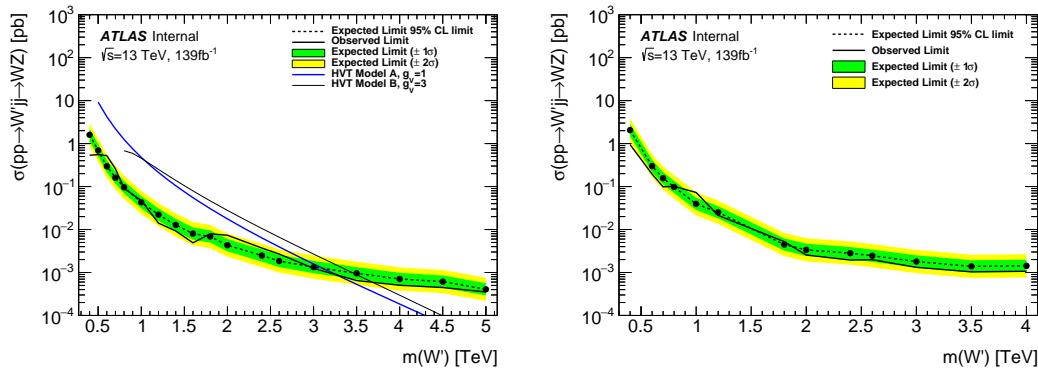


Figure 12.19: This figure shows theory, expected and observed limits for HVT Z' DY (left) and VBF (right) production.

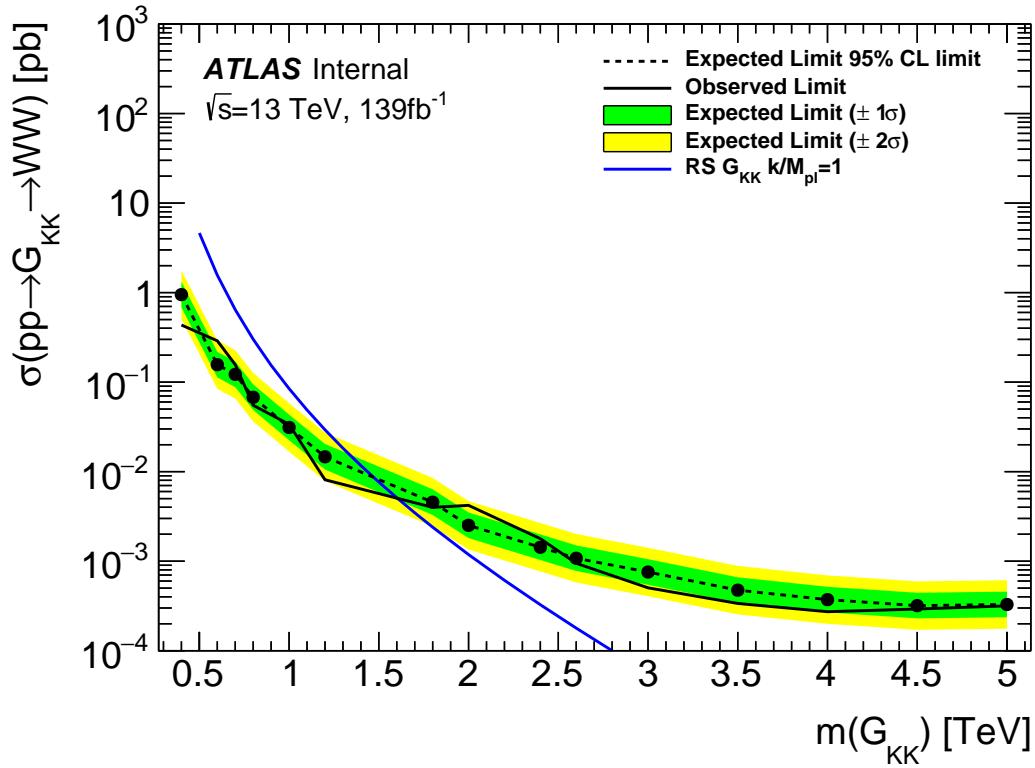


Figure 12.20: This figure shows theory, expected and observed limits for RS Gravitons via DY production.

Part VI

1294

Quark and Gluon Tagging

1295

₁₂₉₆ **Chapter 13**

₁₂₉₇ **Prospects**

₁₂₉₈ For the resolved analysis, signal jets are quark enriched and background jets are
₁₂₉₉ gluon dominated. By classifying jets in the event as quark or gluon initiated, less
₁₃₀₀ background would contaminate the signal region. Figure 13.1 shows the PDGID
₁₃₀₁ for the truth parton matched to the jet (meaning the highest energy parton in
₁₃₀₂ the jet catchment area) in events passing the resolved signal region selections.
₁₃₀₃ PDGID = -1 corresponds to pileup jets, $0 < \text{PDGID} < 6$ correspond to quarks
₁₃₀₄ and $\text{PDGID} = 21$ corresponds to gluons. From this Figure, it is evident that a
₁₃₀₅ notable fraction of the background that contaminates the signal region contains
₁₃₀₆ gluon jets, especially for the sub-leading jet.

₁₃₀₇ As gluons jets have more constituents and therefore more tracks (n_{trk}), the
₁₃₀₈ background jets have more tracks than the signal jets. This is shown in Fig-
₁₃₀₉ ure 13.2. Therefore, by cutting on the number of tracks in a jet, quark and gluon
₁₃₁₀ jets may be distinguished (i.e. jets with less than a given number of tracks are
₁₃₁₁ classified as a quark, otherwise the jet is classified as a gluon.) Moreover, as the
₁₃₁₂ momentum of the jet increases the number of tracks also increases logarithmically
₁₃₁₃ [Cite nachman thesis Natasha]. Therefore by applying a cut on the number of
₁₃₁₄ tracks that scales with the $\ln(p_T)$ is more powerful than a threshold cut on the

1315 number of tracks. Figure 13.3-Figure 13.6 show normalized heat maps of $\ln(p_T)$
1316 vs the number of reconstructed tracks for the background and a 300 GeV Z' signal.
1317 In these plots it is evident that the number of tracks in the background jets grows
1318 more quickly with $\ln(p_T)$ than for the signal jets. This is expected given that the
1319 signal is quark dominated and the background is gluon dominated.

1320 In Figure 13.8 is the ROC Curve for quark gluon tagging with cut on the
1321 number of tracks in a jet that depends on $\ln(p_T)$. The sum of the backgrounds in
1322 the signal region were used for this curve. Here the quark tagging efficiency is the
1323 ratio of quarks tagged as quarks to the total number of quarks in the signal region.
1324 The gluon rejection is calculated as the reciprocal of the gluon tagging efficiency.
1325 Choosing a 90% efficient working point with a rejection of 1.4 corresponds to a
1326 slope of 4 and intercept of -5. Tagging both jets in this analysis would yield an
1327 efficiency of $90\%^{n_{jets}}$. Focusing on the background in Figure 13.9, this cut helps
1328 minimize gluon contamination in the signal region. Also, from these heat maps it
1329 is obvious that the number of tracks in gluon jets grows more quickly than those
1330 in quark jets.

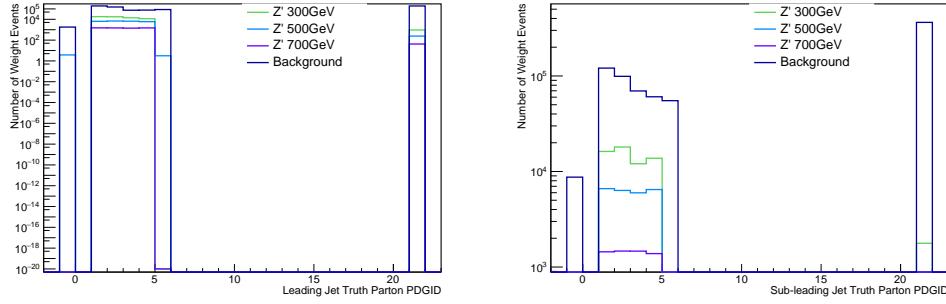


Figure 13.1: PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV Z' signals and the background.

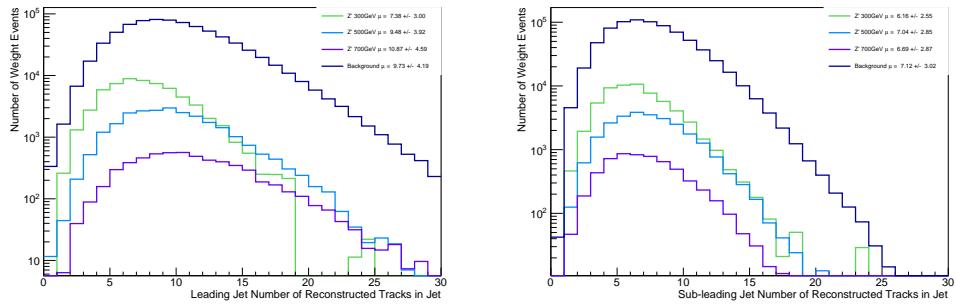


Figure 13.2: The number of tracks in small-R jets in events passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets. These distributions are shown for 300, 500, and 700GeV Z' signals and the background.

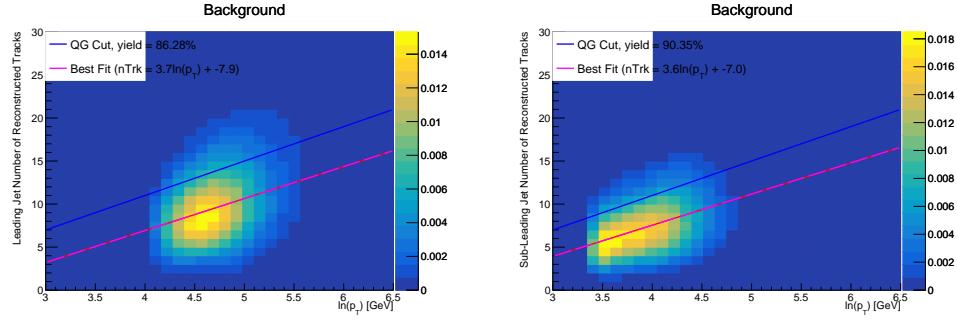


Figure 13.3: The number of tracks in background small-R jets in events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$. Note the number of total entries in these plots has been normalized to one.

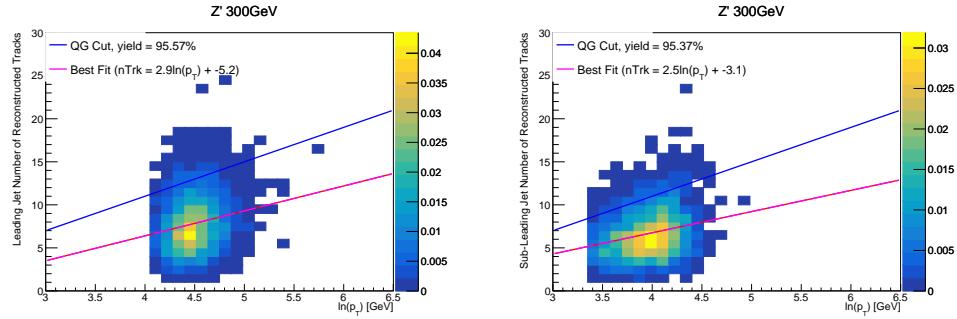


Figure 13.4: The number of tracks in small-R jets in 300GeV Z' events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$. Note the number of total entries in these plots has been normalized to one.

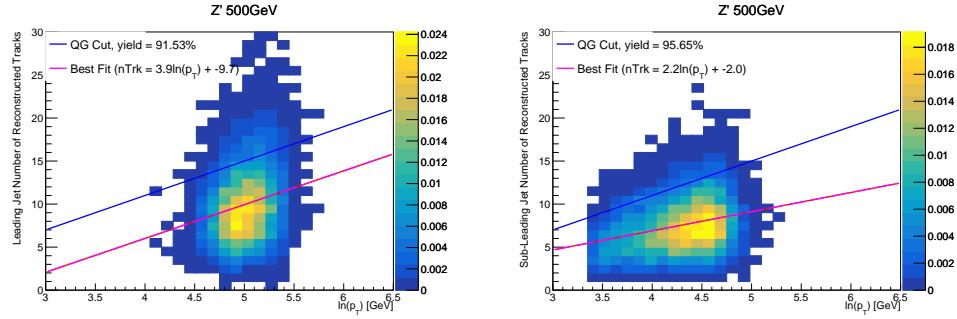


Figure 13.5: The number of tracks in small-R jets in 500GeV Z' events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$. Note the number of total entries in these plots has been normalized to one.

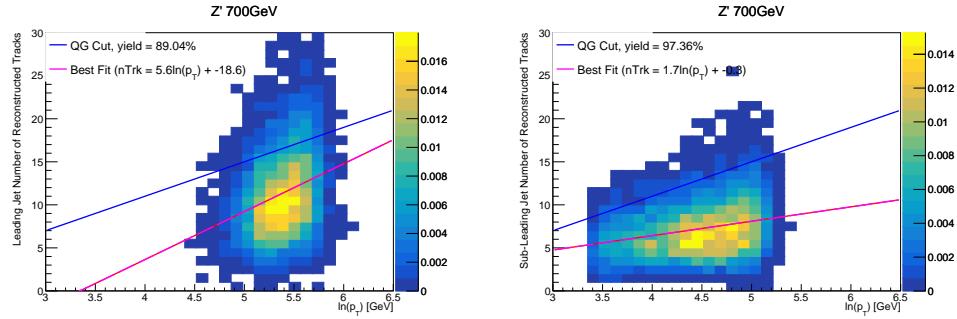


Figure 13.6: The number of tracks in small-R jets in 700GeV Z' events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$. Note the number of total entries in these plots has been normalized to one.

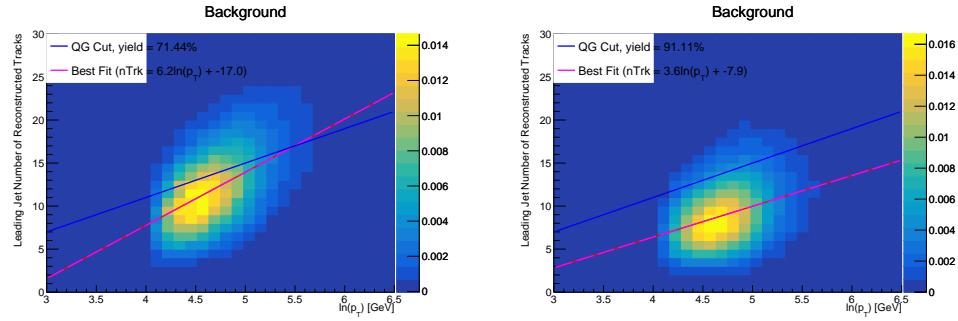


Figure 13.7: The number of tracks in leading small-R jets in background events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Gluons (b) Quarks jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$.Note the number of total entries in these plots has been normalized to one.

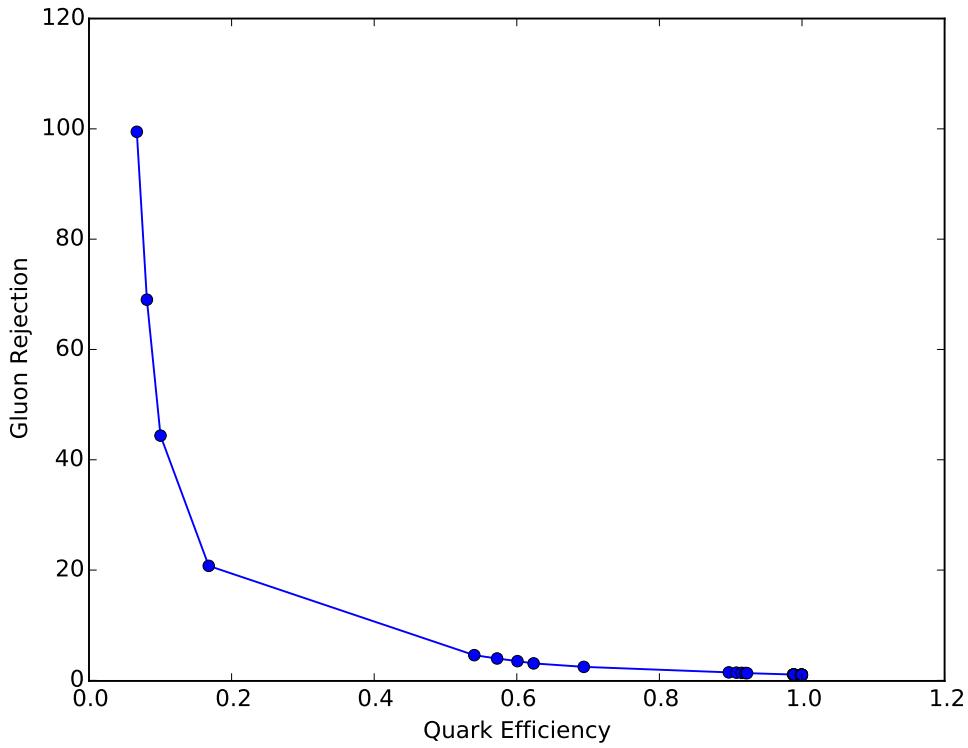


Figure 13.8: ROC Curve for Quark and Gluon Tagging with a cut on the number of tracks that depends on the $\ln(p_T)$.

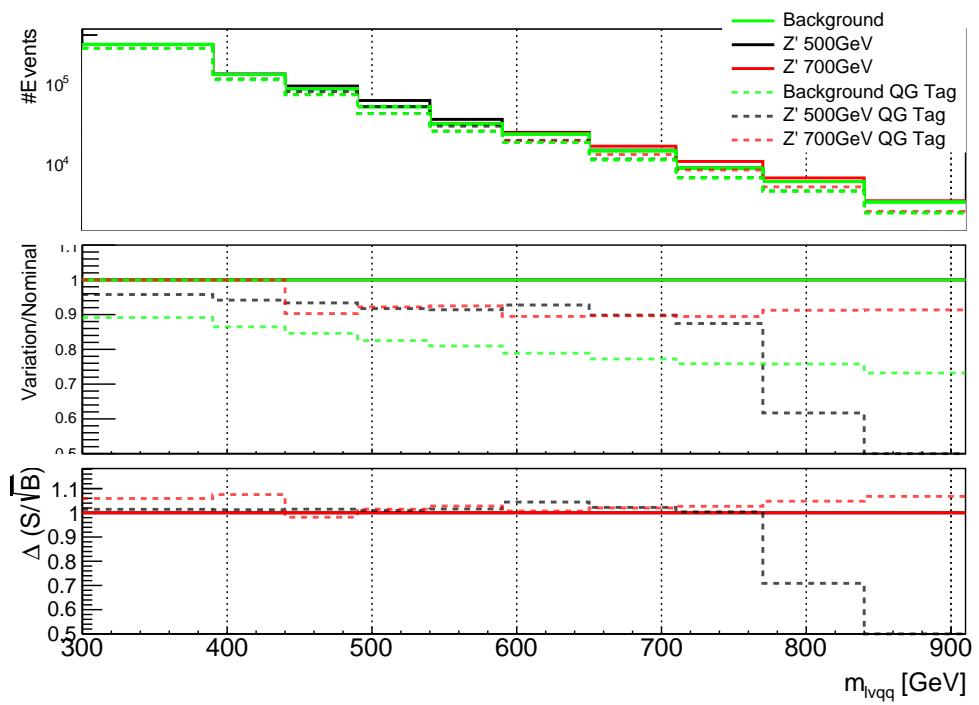


Figure 13.9: The top panel shows the distribution of m_{lvqq} with and without quark gluon tagging. The middle panel shows the ratio of the signals and backgrounds with and without quark gluon tagging. The bottom panel shows the change in S/\sqrt{B} with quark gluon tagging.

₁₃₃₁ **Chapter 14**

₁₃₃₂ **n_{trk} Calibration**

₁₃₃₃ As tagger based on nTrk cuts on the number of tracks in jets, a quantity that
₁₃₃₄ is not known with infinite precision, means that relevant systematic uncertainties
₁₃₃₅ must be evaluated. The sources of uncertainty in n_{trk} may be split into modeling
₁₃₃₆ and experimental uncertainties.

₁₃₃₇ Modeling uncertainties are obtained by assessing PDF and ME uncertainties
₁₃₃₈ on the number of charged particles in particle-level jets in dijet events. The
₁₃₃₉ number of charged particles as a function of jet p_T is calculated using an Iterative
₁₃₄₀ Bayesian (IB) technique [cite paper].

₁₃₄₁ This measurement ([8]) uses the ATLAS 2012 pp collision dataset, correspond-
₁₃₄₂ ing to $20.3/\text{fb}$ at center-of-mass energy $\sqrt{s} = 8\text{TeV}$. The number of charged con-
₁₃₄₃ stituents depends on fragmentation modeling and matrix elements, which do not
₁₃₄₄ depend on s . For this reason, it is safe to use these uncertainties for $\text{sqrt}(s)=13\text{TeV}$.
₁₃₄₅ Monte Carlo (MC) samples are used to determine the response matrix. The MC
₁₃₄₆ sample is a dijet sample generated with Pythia 8.175 using CT10 PDF and AU2
₁₃₄₇ tune. The anti- k_T algorithm is used to cluster jets with a radius parameter R
₁₃₄₈ $= 0.4$. Jets are required to have $|\eta| < 2.1$. Tracks in jets are required to have
₁₃₄₉ $p_T > 500\text{MeV}$, $|\eta| < 2.5$, track-fit $\chi^2 < 3.0$ and originate from the primary ver-

tex. Matching tracks to jets is accomplished using ghost-association [cite]. In this technique, jets are re-clustered with the track collection augmented with "ghost" versions of tracks. These "ghosts" tracks have the same direction as their parent track, but infinitesimal track p_T . This insures meta-jet properties (e.g. η , p_T , etc) are unchanged. A track is matched to a jet if it's ghost version remains in the jet after re-clustering. Further details of the data, object, and event selection may be found in [cite 35].

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

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

This process is prone to three main sources of bias: response matrix, correction factor, and unfolding procedure uncertainties. The response matrix is sensitive to experimental uncertainties impacting jet track reconstruction and calorimeter jet p_T . Correction factors are also sensitive to experimental uncertainties (e.g. JES)

1376 as such uncertainties modify detector level acceptance. Sensitivity to particle
 1377 level acceptance is calculated by comparing Pythia and Herwig. Finally, the bias
 1378 from the IB prior choice is determined by reweighting the particle-level spectrum,
 1379 so the simulated detector level spectrum more closely matches the uncorrected
 1380 data. Unfolding this modified detector-level simulation and comparing it to the re-
 1381 weighted particle-level spectrum indicates bias from the prior distribution choice.
 1382

1382 A summary of all the systematic uncertainties associated with this unfolding
 1383 may be found in [ref paper]. Total uncertainties are < 7% for the number of
 1384 charged particles in jets. The unfolded distribution of the nCharged in jets from
 1385 data are further analyzed to extract the quark and gluon nCharged distributions.
 1386 In dijet events, the jet with a larger η is more energetic and therefore more likely
 1387 to be a quark. This is due to the quarks in protons generally having a larger
 1388 fraction of the total momentum of the proton constituents. The more central jet
 1389 is more likely to be a gluon-initiated jet. This correlation between jet η and flavor
 1390 may then be used to extract nCharged in p_T bins using:

$$\langle n_c^f \rangle = f_q^f \langle n_c^q \rangle + f_g^f \langle n_c^g \rangle \quad (14.1)$$

1391

$$\langle n_c^c \rangle = f_q^c \langle n_c^q \rangle + f_g^c \langle n_c^g \rangle \quad (14.2)$$

1392 In this equation the f and c subscripts denote the more forward and central
 1393 jets, respectively. The q and g subscripts denote quark and gluon. The fraction
 1394 of more forward jets that are say gluons is denoted by f_g^f . The other relevant jet
 1395 fractions are denoted with the same naming scheme. Finally, $\langle n_c \rangle$ is the average
 1396 number of charged particles in a jet in a given p_T bin. To show that Eq. (??) may
 1397 be used to extract quark and gluon n_c distributions the extracted distributions
 1398 are compared to n_c distributions determined using the jet flavor in simulation.
 1399 Figure [add figure natasha] shows that the extracted and true distributions differ

1400 by < 1% over the p_T range probed for this study. Moreover, this implies that n_c
1401 depends only on the flavor of the initiating parton and jet p_T .

1402 These extracted distributions are prone to PDF and ME biases. The bias from
1403 the choice of the CT10 PDF for the Pythia sample is accounted for by comparing
1404 quark/gluon fractions for the nominal CT10 sample with its eigenvector variations.
1405 Comparing the quark/gluon fractions from Pythia 8 and Herwig++ quantify the
1406 uncertainty from the ME calculation. These uncertainties are added in quadra-
1407 ture with the unfolding uncertainty to give the total modelling uncertainty on
1408 the extracted n_c distribution. This is shown in Figure 15.2.

1409 To apply these uncertainties in n_c distributions in data, per-jet event weights
1410 are associated with each uncertainty according to:

$$w_i(n_c) = \frac{P(n_c | n_c > \pm \sigma_{n_c}^i)}{P(n_c | n_c >)} \quad (14.3)$$

1411 In Eq. (??), i denotes the uncertainty considered, P is the Poisson probability,
1412 and $\sigma_{n_c}^i$ represents the average impact of the uncertainty on n_c .

1413 The previous uncertainties described accounted for modeling uncertainty as-
1414 sociated with the number of charged particles in a jet. However, n_c is not a
1415 measurable quantity. Instead the number of tracks in a jet is measured, which is
1416 a proxy for n_c . Therefore the uncertainties associated with the measurement of
1417 nTracks must also be considered ([10]). These uncertainties were calculated using
1418 a Pythia 8 dijet sample with NNPDF 23 and Run 2 data. Track reconstruction
1419 efficiency and fake rates are the dominant sources of nTrack uncertainties.

1420 The track reconstruction efficiency is affected by the uncertainty of the de-
1421 scription of the ID material in simulation and the modeling of charged-particle
1422 interactions with this material. These uncertainties are accounted for by varying
1423 the ID material by 5-25% (dependent on the region of the detector considered).

1424 The difference in the tracking efficiency between the nominal and varied simula-
1425 tion give the uncertainty on the track reconstruction efficiency. Another important
1426 source of track reconstruction inefficiency arises in the core of jets. The high den-
1427 sity of tracks in the jet cores can cause ID clusters to merge. The fraction of lost
1428 tracks due to merging is given by the fraction of tracks that have a charge of two
1429 minimum ionizing particles. This quantity is compared between data and simu-
1430 lation resulting in an uncertainty of 0.4% on tracks with $\Delta R < 0.1$. Combining
1431 these effects gives a total uncertainty as a function of p_T and η that is generally
1432 $< 2\%$ [references figure 44 from [10]].

1433 Fake tracks are the other dominant source of nTrk uncertainty. Fake tracks
1434 are tracks that cannot be associated to a single particle. Often these tracks are a
1435 result of random combinations of hits from charged particles that overlap in space.
1436 In dense environments, such as the core of jets or high-pileup environments, fake
1437 tracks are more likely. Fake tracks are estimated with a 'control region method'
1438 which is briefly summarized here [[9]]. By applying a series of track selections
1439 to enrich the fraction of fake tracks (e.g. $|d_0| > 0.1$, track $\chi^2 > 1.4$, etc) in
1440 simulation, templates for fake track parameters are calculated. These templates
1441 are then fit to data to determine the fraction of fake tracks. On average the fake
1442 rate is found to be 30% (independent of p_T and η).

1443 To assess the impact of these two detector level uncertainties, tracks are ran-
1444 domly dropped according to the rates described above. Reconstruction and fake
1445 uncertainties both lower the number of tracks, hence these uncertainties are one-
1446 sided. By dropping tracks in this way a varied nTrk distribution is calculated for
1447 both uncertainties. The associated per-jet event weights are then calculated in
1448 the same way as the modeling weights as:

$$w_i(n_c) = \frac{P(n_{trk} | < n_{trk} > \pm \sigma_{n_{trk}}^i)}{P(n_{trk} | < n_{trk} >)} \quad (14.4)$$

1449 Adding the modeling and detector level uncertainties in quadrature gives the
 1450 overall nTrack uncertainty. The effects of the individual uncertainties on the nTrk
 1451 distributions can be seen in Fig 15.4. Fig 15.3 shows the m_{lvqq} and nTrk distri-
 1452 butions for the W and Top control regions before likelihood fitting. In these plots
 1453 the nTrk uncertainties improve agreement between data and MC. The remaining
 1454 differences are likely covered by likelihood fitting and improving the analysis itself.

1455 **Chapter 15**

1456 **Application**

1457 Using the 90% WP of the n_{trk} tagger improves S/\sqrt{B} is $\sim 3\%$ as shown in
1458 Figure 13.9. Although, n_{trk} is the single most powerful discriminating variable
1459 for quark and gluon jets, the addition of other jet variables would improve the
1460 classification efficiency. Figure 15.1 shows the possible improvement of 10%
1461 in jet classification using the truth label of the jets to classify jets. This type of
1462 improvement is possible by using variables such as jet width, and energy correlata-
1463 tors. Figure [add BDT figure/use 1612.01551.pdf] shows for a 90% quark tagging
1464 efficiency for a 100 GeV jet, a BDT improve the gluon rejection by 0.4. Once this
1465 tagger is calibrated it would improve the analysis sensitivity of this channel.

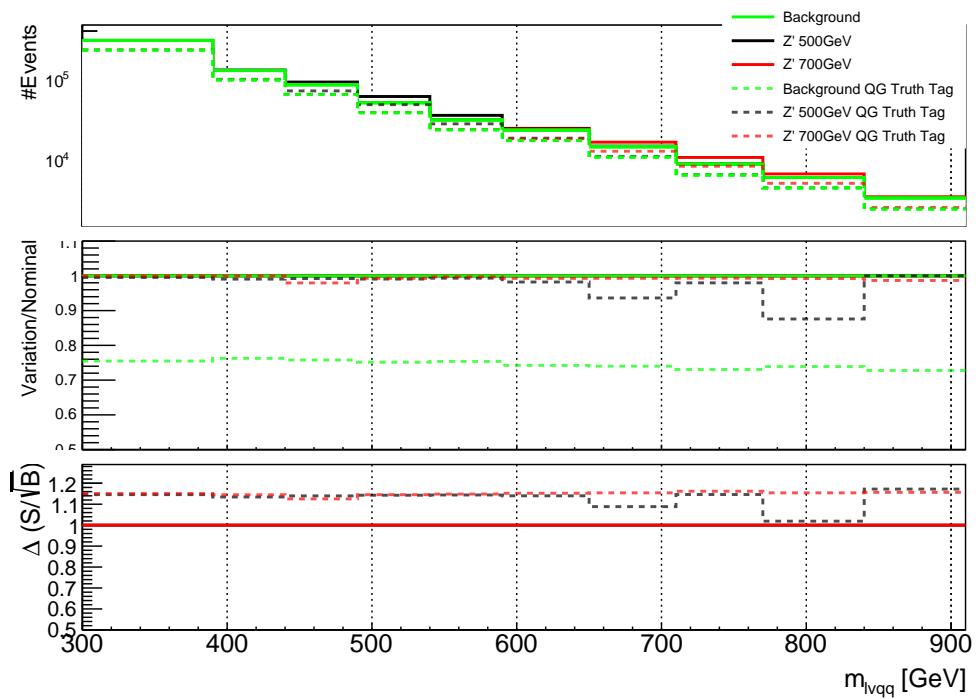


Figure 15.1: The top panel shows the distribution of m_{lvqq} with and without requiring jets to be true quarks. The middle panel shows the ratio of the signals and backgrounds with and without requiring jets to be true quarks. The bottom panel shows the change in S/\sqrt{B} when requiring jets to be true quarks..

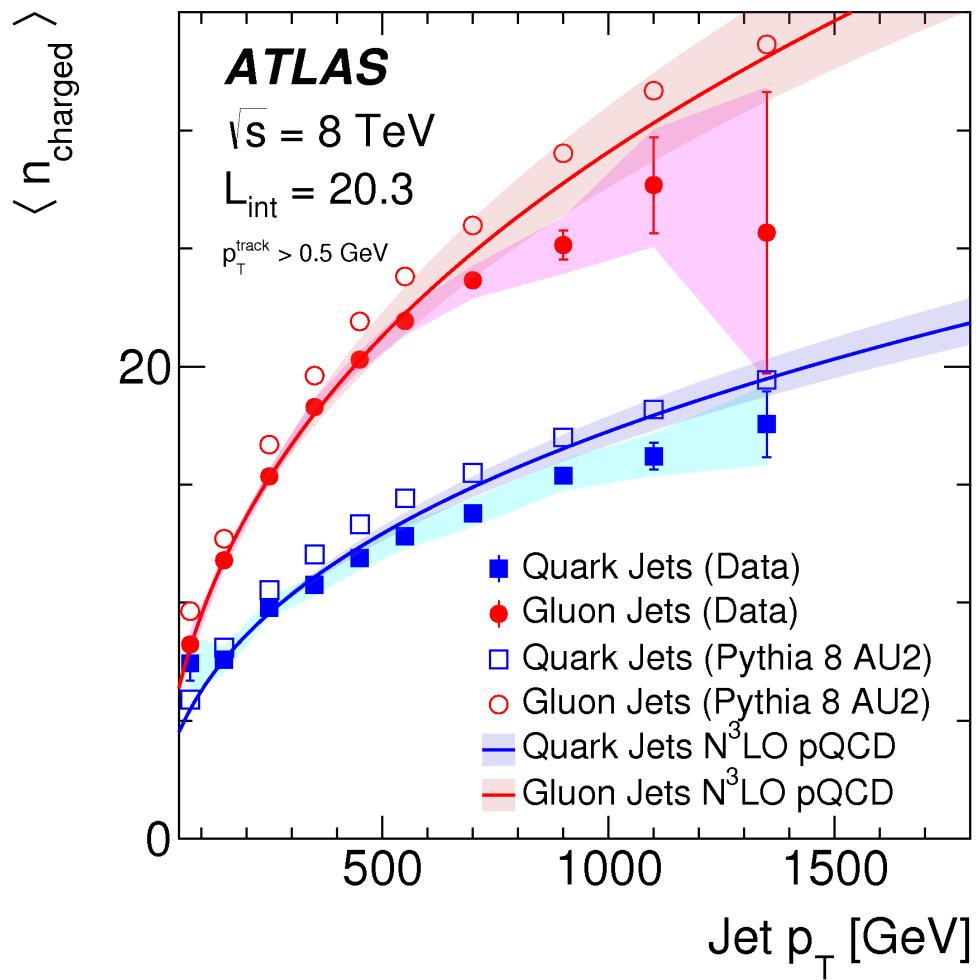


Figure 15.2: Unfolded and extracted n_C qg dstbs..

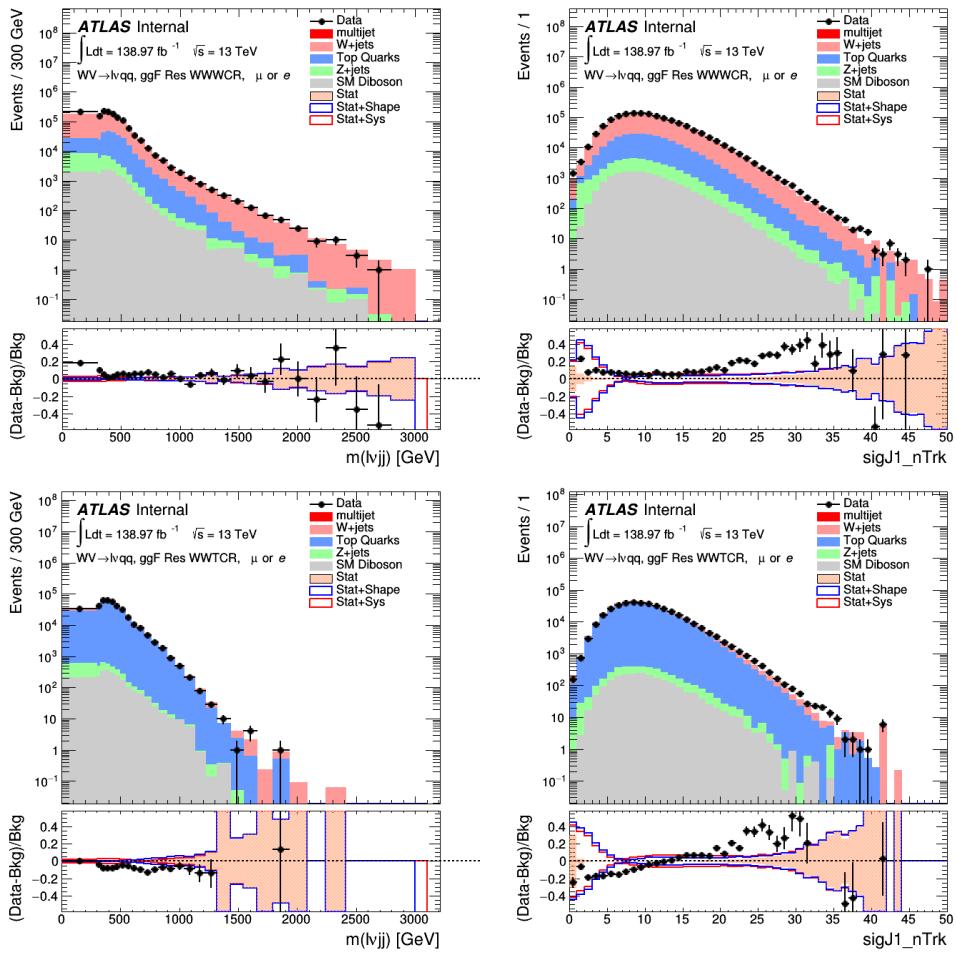


Figure 15.3: PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV Z' signals and the background.

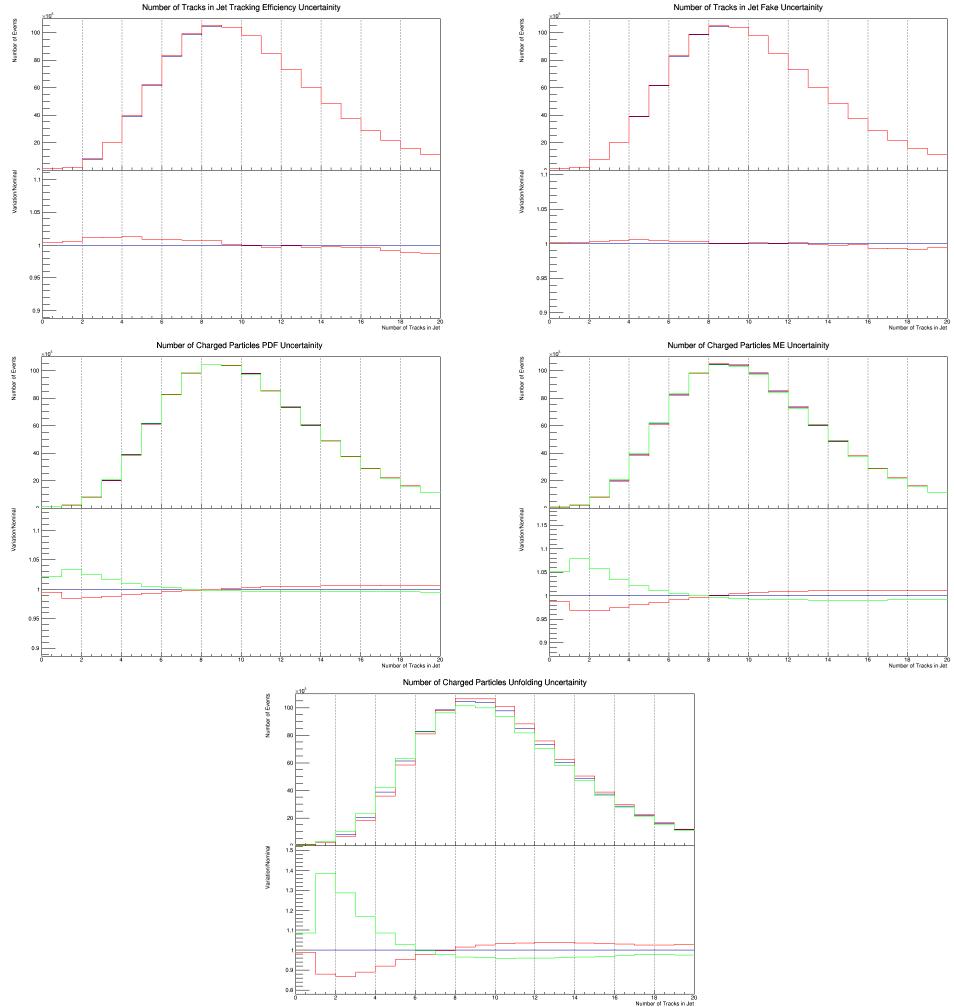


Figure 15.4: These figures show the impact of the uncertainties on the number of tracks in the leading jet in the sum of the background sample in the Resolved GGF WW SR (a) tracking efficiency (b) fake (c) PDF (d) ME (e) unfolding uncertainties.

Part VII

1466

Conclusion

1467

1468 **Chapter 16**

1469 **Conclusions**

1470 A search for WW and WZ diboson resonance production in $\ell\nu qq$ final states
1471 was performed using 139fb^{-1} of pp collision data collected at a center-of-mass
1472 energy of $\sqrt{s} = 13\text{TeV}$ by that ATLAS detector at the LHC between 2015 and
1473 2018. No excess of events above the background-only expectation was observed.
1474 The largest local excess is approximately 2.7σ , which is not significant. Limits
1475 on the production cross section are obtained for the HVT W' and Z' and RS
1476 Gravitons. Signal masses below 3.4 (3.7) TeV are excluded for HVT W' Model
1477 A(B). Signal masses below 3.3 (3.7) TeV are excluded for HVT Z' Model A(B).
1478 Randall Sundrum Gravitons are excluded for masses below 1.6 TeV. Going forward,
1479 improving the classification of jets in events would improve analysis sensitivity.
1480 To distinguish quark from gluon jets a jet tagger based on the number of tracks in
1481 jets is studied in the context of this search. Finally, the calibration of the number
1482 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] ATLAS Collaboration. Atlas muon reconstruction performance in lhc run 2.
- [5] ATLAS Collaboration. Summary plots from the atlas standard model physics group.
- [6] ATLAS Collaboration. Tagging and suppression of pileup jets with the atlas detector.
- [7] 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].
- [8] 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].
- [9] ATLAS Collaboration. Performance of the atlas track reconstruction algorithms in dense environments in lhc run 2. arXiv:1704.07983 [hep-ex].
- [10] 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].
- [11] Alex Dias and V. Pleitez. Grand unification and proton stability near the peccei-quinn scale. *Physical Review D*, 70, 07 2004.
- [12] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. Qcd matrix elements + parton showers. the nlo case. *Journal of High Energy Physics*, 2013(4), Apr 2013.

- 1510 [13] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jets with variable r. *Journal*
1511 *of High Energy Physics*, 2009(06):059–059, Jun 2009.
- 1512 [14] Gregory Soyez Matteo Cacciari, Gavin P. Salam. The anti- k_T jet clustering
1513 algorithm. arXiv:0802.1189 [hep-ph].
- 1514 [15] Duccio Pappadopulo, Andrea Thamm, Riccardo Torre, and Andrea Wulzer.
1515 Heavy vector triplets: bridging theory and data. *Journal of High Energy*
1516 *Physics*, 2014(9), Sep 2014.
- 1517 [16] Antonio Pich. The Standard Model of Electroweak Interactions. In *Proceed-
1518 ings, High-energy Physics. Proceedings, 18th European School (ESHEP 2010):
1519 Raseborg, Finland, June 20 - July 3, 2010*, pages 1–50, 2012. [,1(2012)].
- 1520 [17] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small
1521 extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.
- 1522 [18] Tania Robens and Tim Stefaniak. Lhc benchmark scenarios for the real higgs
1523 singlet extension of the standard model. *The European Physical Journal C*,
1524 76(5), May 2016.
- 1525 [19] Alex Sherstinsky. Fundamentals of recurrent neural network (RNN) and long
1526 short-term memory (LSTM) network. *CoRR*, abs/1808.03314, 2018.