

¹

©Copyright 2021

²

Sean Gasiorowski

$HH \rightarrow b\bar{b}b\bar{b}$ or How I Learned to Stop Worrying and Love the QCD Background

Sean Gasiorowski

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

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Anna Goussiou, Chair

Jason Detwiler

Shih-Chieh Hsu

David Kaplan

Henry Lubatti

Thomas Quinn

Gordon Watts

Program Authorized to Offer Degree:
Physics

22

University of Washington

23

Abstract

24

$HH \rightarrow b\bar{b}b\bar{b}$ or How I Learned to Stop Worrying and Love the QCD Background

25

Sean Gasiorowski

26

Chair of the Supervisory Committee:

27

Professor Anna Goussiou

28

Physics

29

Insert abstract here

TABLE OF CONTENTS

	Page
31 List of Figures	iii
32 Glossary	xiv
33 Chapter 1: The Standard Model of Particle Physics	1
34 1.1 Introduction: Particles and Fields	1
35 1.2 Quantum Electrodynamics	4
36 1.3 An Aside on Group Theory	9
37 1.4 Quantum Chromodynamics	11
38 1.5 The Weak Interaction	14
39 1.6 The Higgs Potential and the SM	20
40 1.7 The Standard Model: A Summary	26
41 Chapter 2: Di-Higgs Phenomenology and Physics Beyond the Standard Model	28
42 2.1 Intro to Di-Higgs	28
43 2.2 Resonant HH Searches	29
44 2.3 Non-resonant HH Searches	31
45 Chapter 3: Experimental Apparatus	35
46 3.1 The Large Hadron Collider	35
47 3.2 The ATLAS Experiment	38
48 Chapter 4: Simulation	47
49 4.1 Event Generation	47
50 4.2 Detector Simulation	48
51 4.3 Correlated Fluctuations in FastCaloSim	50

52	Chapter 5:	Reconstruction	58
53	5.1	Jets	58
54	5.2	Flavor Tagging	60
55	Chapter 6:	The Anatomy of an LHC Search	70
56	6.1	Object Selection and Identification	70
57	6.2	Defining a Signal Region	70
58	6.3	Background Estimation	70
59	6.4	Uncertainty Estimation	70
60	6.5	Hypothesis Testing	70
61	Chapter 7:	Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state	71
62	7.1	Data and Monte Carlo Simulation	73
63	7.2	Triggers and Object Definitions	75
64	7.3	Analysis Selection	77
65	7.4	Background Reduction and Top Veto	83
66	7.5	Kinematic Region Definition	85
67	7.6	Background Estimation	91
68	7.7	Uncertainties	152
69	7.8	Background Validation	163
70	7.9	Overview of Other $b\bar{b}b\bar{b}$ Channels	166
71	7.10	m_{HH} Distributions	167
72	7.11	Statistical Analysis	174
73	7.12	Results	175
74	Chapter 8:	Future Ideas for $HH \rightarrow b\bar{b}b\bar{b}$	180
75	8.1	pairAGraph: A New Method for Jet Pairing	180
76	8.2	Background Estimation with Mass Plane Interpolation	183
77	Chapter 9:	Conclusions	194

LIST OF FIGURES

	Figure Number	Page
79	1.1 Diagram of the elementary particles described by the Standard Model [1].	3
80	2.1 Representative diagrams for the gluon-gluon fusion production of spin-0 (X) and spin-2 (G_{KK}^*) resonances which decay to two Standard Model Higgs bosons. The spin-0 resonance considered for this thesis is a generic narrow width resonance which may be interpreted in the context of two Higgs doublet models [18], whereas the spin-2 resonance is considered as a Kaluza-Klein graviton within the bulk Randall-Sundrum (RS) model [19, 20].	29
81		
82		
83		
84		
85		
86	2.2 Dominant contributing diagrams for non-resonant gluon-gluon fusion production of HH . κ_λ and κ_t represent variations of the Higgs self-coupling and coupling to top quarks respectively, relative to that predicted by the Standard Model.	30
87		
88		
89		
90	2.3 Illustration of dominant HH branching ratios. $HH \rightarrow b\bar{b}b\bar{b}$ is the most common decay mode, representing 34 % of all HH events produced at the LHC.	31
91		
92		
93		
94		
95		
96		
97		
98		
99	2.4 Monte Carlo generator level m_{HH} distributions for various values of κ_λ , demonstrating the impact of the interference between the two diagrams of Figure 2.2 on the resulting m_{HH} distribution. For $\kappa_\lambda = 0$ there is no triangle diagram contribution, demonstrating the shape of the box diagram contribution, whereas for $\kappa_\lambda = 10$, the triangle diagram dominates, with a strong low mass peak. The interplay between the two is quite evident for other values, resulting in, e.g., the double peaked structure present for $\kappa_\lambda = 2$ (near maximal destructive interference) and $\kappa_\lambda = -5$. [21]	33
100		
101		
102		
103		
104	3.1 Diagram of the ATLAS detector [40]	38
105		
101	3.2 Cross section of the ATLAS detector showing how particles interact with various detector components [42]	40
102		
103	3.3 2D projections of the ATLAS coordinate system	41
104		
105	4.1 Energy and lateral shower width variable, weta2, for 16 GeV photons with full simulation (G4) and FastCaloSimV2 (FCSV2) [57].	51

106	4.2 Example of photon and pion average shapes in 5×5 calorimeter cells. The left column shows the average shape over a sample of 10000 events, while the right column shows the energy ratio, in each cell, of single GEANT4 events with respect to this average. The photon ratios are all close to 1, while the pion ratios show significant deviation from the average.	53
111	4.3 Distribution of the ratio of voxel energy in single events to the corresponding voxel energy in the average shape, with GEANT4 events in blue and Gaussian model events in orange, for 65 GeV central pions in EMB2. Moving top to bottom corresponds to increasing α , left to right corresponds to increasing R , with core voxels numbered 1, 10, 19, Agreement is quite good across all voxels. Results are similar for the VAE method.	54
117	4.4 Correlation coefficient of ratios of voxel energy in single events to the cor- responding voxel energy in the average shape, examined between the core bin from $\alpha = 0$ to $2\pi/8$ and each of the other voxels. The periodic structure represents the binning in α , and the increasing numbers in each of these periods correspond to increasing R , where the eight points with correlation coefficient 1 are the eight core bins. Both the Gaussian and VAE generated toy events are able to reproduce the major correlation structures for 65 GeV central pions in EMB2.	55
125	4.5 Comparison of the RMS fluctuations about the average shape with the Gaussian fluctuation model (red), the VAE fluctuation model (green), and without correlated fluctuations (blue) for a range of pion energies, η points, and layers.	56
128	4.6 Comparison of the Gaussian fluctuation model to the default FCSV2 version and to G4 simulation, using pions of 65 GeV energy and $0.2 < \eta < 0.25$. With the correlated fluctuations, several shape variables demonstrate improved modeling.	57
132	5.1 Illustration of an interaction producing two light jets and one b -jet in the transverse plane. While the light jets decay “promptly”, coinciding with the primary vertex of the proton-proton interaction, the longer lifetime of B hadrons leads to a secondary decay vertex, displaced from the primary vertex by length L_{xy} . This is typically a few mm, and therefore is not directly visible in the detector, but leads to a large transverse impact parameter, d_0 , for the resulting tracks. [66]	63
139	5.2 Performance of the various low and high level flavor tagging algorithms in $t\bar{t}$ simulation, demonstrating the tradeoff between b -jet efficiency and light and c -jet rejection. The high level taggers demonstrate significantly better performance than any of the individual low level taggers, with DL1 offering slight improvements over MV2 due to the inclusion of additional input variables.	67

144	5.3 Performance of the MV2, DL1, and DL1r algorithms in $t\bar{t}$ simulation, demon-	
145	strating the tradeoff between b -jet efficiency and light and c -jet rejection. f_c	
146	controls the importance of c -jet rejection in the discriminating variable, and	
147	values shown have been optimized separately for each DL1 configuration. DL1r	
148	demonstrates a significant improvement in both light and c jet rejection over	
149	MV2 and DL1. [71]	68
150	7.1 Comparison of m_{H_1} vs m_{H_2} planes for the full Run 2 2 b dataset with different	
151	pairings. As evidenced, this choice significantly impacts where events fall in	
152	this plane, and therefore which events fall into the various kinematic regions	
153	defined in this plane (see Section 7.5). Respective signal regions are shown for	
154	reference, with the min ΔR signal region shifted slightly up and to the right to	
155	match the non-resonant selection. Note that the band structure around 80 GeV	
156	in both m_{H_1} and m_{H_2} is introduced by the top veto described in Section 7.4.	81
157	7.2 Comparison of signal distributions in the respective signal regions for the	
158	min ΔR and D_{HH} pairing for various values of the Higgs trilinear coupling	
159	in the respective signal regions. The distributions are quite similar at the	
160	Standard Model point, but for other variations, min ΔR does not pick up the	
161	low mass features.	82
162	7.3 Regions used for the resonant search, shown on the 2 b data mass plane. The	
163	outermost region (the “control region”) is used for derivation of the nominal	
164	background estimate. The innermost region is the signal region, where the	
165	signal extraction fit is performed. The region in between (the “validation	
166	region”) is used for the assessment of an uncertainty.	86
167	7.4 Regions used for the non-resonant search. The “top” and “bottom” quadrants	
168	together comprise the control region, in which the nominal background estimate	
169	is derived. The “left” and “right” quadrants together comprise the validation	
170	region, which is used to assess an uncertainty. The signal region, in the center,	
171	is where the signal extraction fit is performed.	88
172	7.5 Impact of the m_{HH} correction on a range of spin-0 resonant signals. The	
173	corrected m_{HH} distributions (solid lines) are much sharper and more centered	
174	on the corresponding resonance masses than the uncorrected m_{HH} distributions	
175	(dashed).	89
176	7.6 Resonant Search: Distributions of ΔR between the closest Higgs Candidate	
177	jets, ΔR between the other two, and average absolute value of HC jet η before	
178	and after CR derived reweighting for the 2018 Control Region.	98

179	7.7 Resonant Search: Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.	99
180		
181		
182	7.8 Resonant Search: Distributions of the number of jets before and after CR derived reweighting for the 2018 Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	100
183		
184		
185	7.9 Resonant Search: Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 Control Region.	101
186		
187		
188	7.10 Resonant Search: Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Control Region.	102
189		
190		
191	7.11 Resonant Search: Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.	103
192		
193		
194	7.12 Resonant Search: Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.	104
195		
196		
197	7.13 Resonant Search: Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.	105
198		
199		
200	7.14 Resonant Search: Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 Control Region. Reweighting is done after the cut on this variable is applied	106
201		
202		
203	7.15 Resonant Search: Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 Validation Region.	107
204		
205		
206	7.16 Resonant Search: Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.	108
207		
208		
209	7.17 Resonant Search: Distributions of the number of jets before and after CR derived reweighting for the 2018 Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	109
210		
211		
212	7.18 Resonant Search: Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.	110
213		
214		

215	7.19 Resonant Search: Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Validation Region.	111
218	7.20 Resonant Search: Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.	112
221	7.21 Resonant Search: Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.	113
224	7.22 Resonant Search: Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.	114
227	7.23 Resonant Search: Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 Validation Region. Reweighting is done after the cut on this variable is applied	115
230	7.24 Non-resonant Search (4b): Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 4b Control Region.	116
233	7.25 Non-resonant Search (4b): Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.	117
236	7.26 Non-resonant Search (4b): Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	118
239	7.27 Non-resonant Search (4b): Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.	119
242	7.28 Non-resonant Search (4b): Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Control Region.	120
245	7.29 Non-resonant Search (4b): Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.	121
248	7.30 Non-resonant Search (4b): Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.	122

251	7.31 Non-resonant Search (4b): Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.	123
252		
253		
254	7.32 Non-resonant Search (4b): Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 4b Control Region. Reweighting is done after the cut on this variable is applied.	124
255		
256		
257	7.33 Non-resonant Search (4b): Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 4b Validation Region.	125
258		
259		
260	7.34 Non-resonant Search (4b): Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.	126
261		
262		
263	7.35 Non-resonant Search (4b): Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	127
264		
265		
266	7.36 Non-resonant Search (4b): Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.	128
267		
268		
269	7.37 Non-resonant Search (4b): Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Validation Region.	129
270		
271		
272	7.38 Non-resonant Search (4b): Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.	130
273		
274		
275	7.39 Non-resonant Search (4b): Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.	131
276		
277		
278	7.40 Non-resonant Search (4b): Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.	132
279		
280		
281	7.41 Non-resonant Search (4b): Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 4b Validation Region. Reweighting is done after the cut on this variable is applied.	133
282		
283		
284	7.42 Non-resonant Search (3b1l): Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 3b1l Control Region.	134
285		
286		
287		

288	7.43 Non-resonant Search (3b1l): Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Control Region.	135
291	7.44 Non-resonant Search (3b1l): Distributions of the number of jets before and after CR derived reweighting for the 2018 3b1l Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	136
294	7.45 Non-resonant Search (3b1l): Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.	137
297	7.46 Non-resonant Search (3b1l): Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Control Region.	138
300	7.47 Non-resonant Search (3b1l): Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.	139
303	7.48 Non-resonant Search (3b1l): Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Control Region.	140
306	7.49 Non-resonant Search (3b1l): Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.	141
309	7.50 Non-resonant Search (3b1l): Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 3b1l Control Region. Reweighting is done after the cut on this variable is applied.	142
312	7.51 Non-resonant Search (3b1l): Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 3b1l Validation Region.	143
316	7.52 Non-resonant Search (3b1l): Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Validation Region.	144
319	7.53 Non-resonant Search (3b1l): Distributions of the number of jets before and after CR derived reweighting for the 2018 3b1l Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	145
322	7.54 Non-resonant Search (3b1l): Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.	146

325	7.55 Non-resonant Search (3b1l): Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Validation Region.	147
328	7.56 Non-resonant Search (3b1l): Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.	148
331	7.57 Non-resonant Search (3b1l): Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Validation Region.	149
334	7.58 Non-resonant Search (3b1l): Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.	150
337	7.59 Non-resonant Search (3b1l): Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 3b1l Validation Region. Reweighting is done after the cut on this variable is applied.	151
340	7.60 Illustration of the approximate bootstrap band procedure, shown as a ratio to the nominal estimate for the 2017 non-resonant background estimate. Each grey line is from the m_{HH} prediction for a single bootstrap training. Figure 7.60(a) shows the variation histograms constructed from median weight \pm the IQR of the replica weights. It can be seen that this captures the rough shape of the bootstrap envelope, but is not good estimate for the overall magnitude of the variation. Figure 7.60(b) demonstrates the applied normalization correction, and Figure 7.60(c) shows the final band (normalized Figure 7.60(a) + Figure 7.60(b)). Comparing this with the IQR variation for the prediction from each bootstrap in each bin in Figure 7.60(d), the approximate envelope describes a very similar variation.	156
351	7.61 Resonant Search: Example of CR vs VR variation in each H_T region for 2018. The variation nicely factorizes into low and high mass components.	158
353	7.62 Non-resonant Search (4b): Example of CR vs VR variation in each signal region quadrant for 2018. Significantly different behavior is seen between quadrants, with the largest variation in quadrant 1 and the smallest in quadrant 4.	159

357	7.63 Resonant Search: Performance of the background estimation method in the resonant analysis reversed $\Delta\eta_{HH}$ kinematic signal region. A new background estimate is trained following nominal procedures entirely within the reversed $\Delta\eta_{HH}$ region, and the resulting model, including uncertainties, is compared with $4b$ data in the corresponding signal region. Good agreement is shown. The quoted p -value uses the χ^2 test statistic, and demonstrates no evidence that the data differs from the assessed background.	164
364	7.64 Non-resonant Search: Performance of the background estimation method in the $3b + 1$ fail validation region. A new background estimate is trained following nominal procedures but with a reweighting from $2b$ to $3b + 1$ fail events. Generally good agreement is seen, though there is some deviation at very low masses in the low $\Delta\eta_{HH}$ low X_{HH} category.	165
369	7.65 Resonant Search: Demonstration of the performance of the nominal reweighting in the control region on corrected m_{HH} , with Figure 7.65(a) showing $2b$ events normalized to the total $4b$ yield and Figure 7.65(b) applying the reweighting procedure. Agreement is much improved with the reweighting. Note that overall reweighted $2b$ yield agrees with $4b$ yield in the control region by construction.	168
375	7.66 Resonant Search: Demonstration of the performance of the control region derived reweighting in the validation region on corrected m_{HH} . Agreement is generally good for this extrapolated estimate. Note that the uncertainty band includes the extrapolation systematic, which is defined by a reweighting trained in the validation region.	169
380	7.67 Resonant Search: Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit. The closure is generally quite good, though there is an evident deficit in the background estimate relative to the data for higher values of corrected m_{HH}	170
384	7.68 Non-resonant Search (4b): Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the $4b$ channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Modeling is generally quite good near the Standard Model peak, but disagreements are seen at very low and high masses. A deficit is present in low $\Delta\eta_{HH}$ bins near 600 GeV.	172

426	8.1 Gaussian process sampling prediction of marginals m_{H1} and m_{H2} for $2b$ signal region events compared to real $2b$ signal region events for the 2018 dataset. Good agreement is seen. Only a small fraction (0.01) of the $2b$ dataset is used for both training and this final comparison to mimic $4b$ statistics.	189
430	8.2 Gaussian process sampling prediction for the mass plane compared to the real $2b$ dataset for 2018. Only a small fraction (0.01) of the $2b$ dataset is used for both training and this final comparison to mimic $4b$ statistics. Good agreement is seen.	191
434	8.3 Gaussian process sampling prediction of marginals m_{H1} and m_{H2} for $4b$ signal region events compared to both control and validation reweighting predictions. While there are some differences, the estimates are compatible.	192
437	8.4 Gaussian process sampling prediction for the $4b$ mass plane compared to the reweighted $2b$ estimate in the signal region. Both estimates are compatible. .	193
438		

GLOSSARY

440 ARGUMENT: replacement text which customizes a L^AT_EX macro for each particular usage.

ACKNOWLEDGMENTS

442 Five years is both a short time and a long time – many things have happened and many
443 have stayed the same. I certainly know much more physics than I did at the outset, but also
444 have learned to ski, discovered a love for hiking, eaten large amounts of cheese, and survived
445 a pandemic by making sourdough and cinnamon rolls. Of course, the most important part of
446 any journey is the friends we made along the way – the utmost gratitude and appreciation
447 goes to the Seattle friends, for the many nights both hard at work and at College Inn, to
448 the CERN friends, for adventures around Europe and days by the lake, the Chicago friends,
449 for my position as European correspondent for the Dum Dum Donut Intellectuals, and the
450 friends from home, for New Years Eve parties and visits in between. Though we’re scattered
451 across the world, I hope to see you all soon.

452 A thank you, of course, to my family for their continued and constant support, for trying
453 their best to learn physics along with me, and going along on this Ph.D. adventure every
454 step of the way.

455 I am only writing this thesis because of the incredible support of my research group, so a
456 massive thank you to Anna, for your guidance and compassion, for advising me as both a
457 physicist and a person, and for always being available, and Jana, for shaping who I am as a
458 physicist and providing feedback on almost literally every talk that I’ve written – I knew I
459 was getting ready to graduate when you started running out of comments.

460 A special set of thank yous to the SLAC group, namely Nicole, Michael, and Rafael, for
461 being an excellent set of collaborators with an excellent set of ideas to push the analysis
462 forward, many of which are in this thesis, and for being my second research family.

463 And last but not least, a thank you to the entire $HH \rightarrow 4b$ group, of course to Max

⁴⁶⁴ and Rafael for their guidance and direction, Bejan for being an excellent collaborator and
⁴⁶⁵ co-editor for the resonant search, Dale and Alex for excelling on the boosted side of things,
⁴⁶⁶ Lucas, for great work on triggers and for diving headfirst into all of my code, the ggF push
⁴⁶⁷ squad, for a huge amount of hard work on the non-resonant, and of course the $HH \rightarrow 4\text{beers}$
⁴⁶⁸ team, many of whom have already been mentioned, but who deserve an extra shout out for
⁴⁶⁹ keeping things fun even during stressful times.

⁴⁷⁰ The physics is done, the rest is paperwork. Let us begin.

471

DEDICATION

472

To family, both given and found

473

Chapter 1

474

THE STANDARD MODEL OF PARTICLE PHYSICS

475

The Standard Model of Particle Physics (SM) is a monumental historical achievement, providing a formalism with which one may describe everything from the physics of everyday experience to the physics that is studied at very high energies at the Large Hadron Collider (Chapter 3). In this chapter, we will provide a brief overview of the pieces that go into the construction of such a model. The primary focus of this thesis is searches for pair production of Higgs bosons decaying to four b -quarks. Consequently, we will pay particular attention to the relevant pieces of the Higgs Mechanism, as well as the theory behind searches at a hadronic collider.

483

1.1 *Introduction: Particles and Fields*

484

What is a particle? The Standard Model describes a set of fundamental, point-like, objects shown in Figure 1.1. These objects have distinguishing characteristics (e.g., mass and spin). These objects interact in very specific ways. The set of objects and their interactions result in a set of observable effects, and these effects are the basis of a field of experimental physics.

488

The effects of these objects and their interactions are familiar as fundamental forces: electromagnetism (photons, electrons), the strong interaction (quarks, gluons), the weak interaction (neutrinos, W and Z bosons). Gravity is not described in this model, as the weakest, with effects most relevant on much larger distance scales than the rest. However, the description of these other three is powerful – verifying and searching for cracks in this description is a large effort, and the topic of this thesis.

494

The formalism for describing these particles and their interactions is that of quantum field theory. Classical field theory is most familiar in the context of, e.g., electromagnetism – an

496 electric field exists in some region of space, and a charged point-particle experiences a force
497 characterized by the charge of the point-particle and the magnitude of the field at the location
498 of the point-particle in spacetime. The same language translates to quantum field theory.
499 Here, particles are described in terms of quantum fields in some region of spacetime. These
500 fields have associated charges which describe the forces they experience when interacting
501 with other quantum fields. Most familiar is electric charge – however this applies to e.g., the
502 strong interaction as well, where quantum fields have an associated *color charge* describing
503 behavior under the strong force.

504 Particles are observed to behave in different ways under different forces. These behaviors
505 respect certain *symmetries*, which are most naturally described in the language of group
506 theory. The respective fields, charges, and generators of these symmetry groups are the basic
507 pieces of the SM Lagrangian, which describes the full dynamics of the theory. In the following,
508 we will build up the basic components of this Lagrangian. The treatment presented here relies
509 heavily on Jackson's Classical Electrodynamics [2] for the build-up, and Thomson's Modern
510 Particle Physics [3] for the rest, with reference to Srednicki's Quantum Field Theory [4], and
511 some personal biases and interjections.

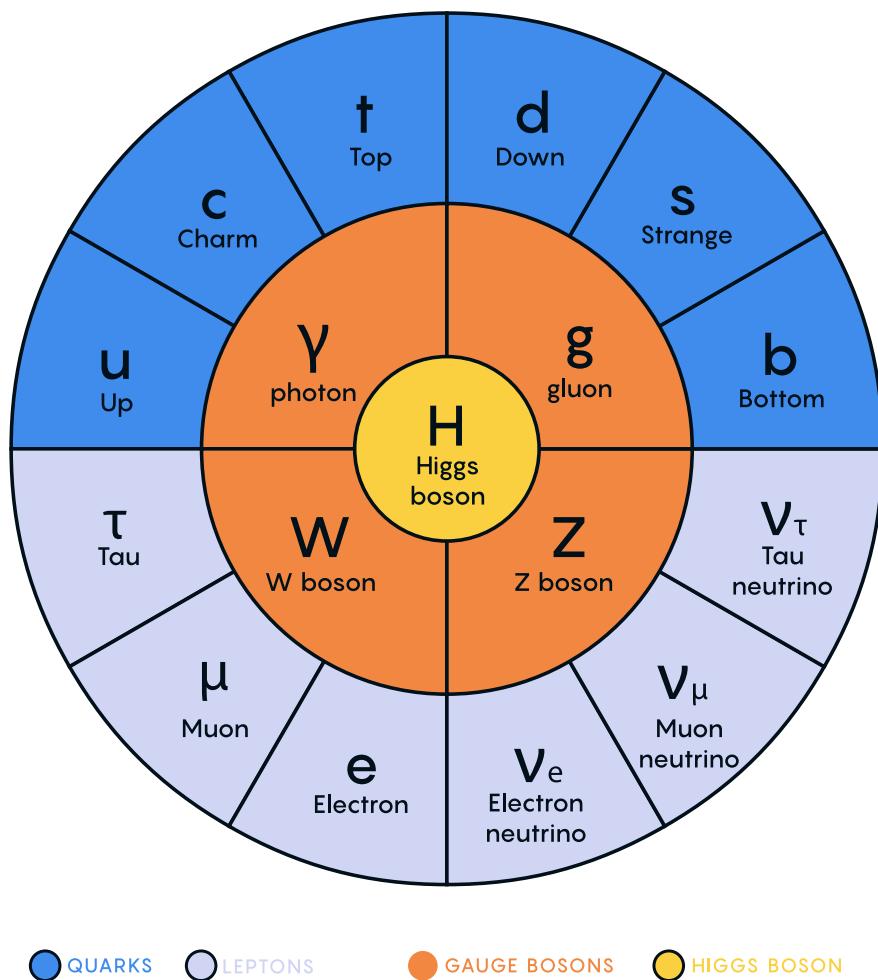


Figure 1.1: Diagram of the elementary particles described by the Standard Model [1].

512 **1.2 Quantum Electrodynamics**

Classical electrodynamics is familiar to the general physics audience: electric (\vec{E}) and magnetic (\vec{B}) fields are used to describe behavior of particles with charge q moving with velocity \vec{v} , with forces described as $\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}$. Hints at some more fundamental properties of electric and magnetic fields come via a simple thought experiment: in a frame of reference moving along with the particle at velocity \vec{v} , the particle would appear to be standing still, and therefore have no magnetic force exerted. Therefore a *relativistic* formulation of the theory is required. This is most easily accomplished with a repackaging: the fundamental objects are no longer classical fields but the electric and magnetic *potentials*: ϕ and \vec{A} respectively, with

$$\vec{E} = -\nabla\phi - \frac{\partial\vec{A}}{\partial t} \quad (1.1)$$

$$\vec{B} = \nabla \times \vec{A} \quad (1.2)$$

It is then natural to fully repackage into a relativistic *four-vector*: $A^\mu = (\phi, \vec{A})$. Considering $\partial^\mu = (\frac{\partial}{\partial t}, \nabla)$, the x components of these above two equations become:

$$E_x = -\frac{\partial\phi}{\partial x} - \frac{\partial A_x}{\partial t} = -(\partial^0 A^1 - \partial^1 A^0) \quad (1.3)$$

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = -(\partial^2 A^3 - \partial^3 A^2) \quad (1.4)$$

513 where we have used the sign convention $(+, -, -, -)$, such that $\partial^\mu = (\frac{\partial}{\partial x_0}, -\nabla)$.

This is naturally suggestive of a second rank, antisymmetric tensor to describe both the electric and magnetic fields (the *field strength tensor*), defined as:

$$F^{\alpha\beta} = \partial^\alpha A^\beta - \partial^\beta A^\alpha \quad (1.5)$$

Defining a four-current as $J_\mu = (q, \vec{J})$, with q standard electric charge, \vec{J} standard electric current, conservation of charge may be expressed via the continuity equation

$$\partial_\mu J^\mu = 0 \quad (1.6)$$

and all of classical electromagnetism may be packaged into the Lagrangian density:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - J^\mu A_\mu. \quad (1.7)$$

514 This gets us partway to our goal, but is entirely classical - the description is of classical
 515 fields and point charges, not of quantum fields and particles. To reframe this, let us go back
 516 to the zoomed out view of the particles of the Standard Model. Two of the most familiar
 517 objects associated with electromagnetism are electrons: spin-1/2 particles with charge e , mass
 518 m , and photons: massless spin-1 particles which are the "pieces" of electromagnetic radiation.

519 We know that electrons experience electromagnetic interactions with other objects. Given
 520 this, and the fact that such interactions must be transmitted *somewhat* between e.g. two
 521 electrons, it seems natural that these interactions are facilitated by electromagnetic radiation.
 522 More specifically, we may think of photons as *mediators* of the electromagnetic force. It
 523 follows, then, that a description of electromagnetism on the level of particles must involve a
 524 description of both the "source" particles (e.g. electrons), the mediators (photons), and their
 525 interactions. Further, this description must be (1) relativistic and (2) consistent with the
 526 classically derived dynamics described above.

The beginnings of a relativistic description of spin-1/2 particles is due to Paul Dirac, with the famous Dirac equation:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \quad (1.8)$$

where ∂_μ is as defined above, ψ is a Dirac *spinor*, i.e. a four-component wavefunction, m is the mass of the particle, and γ^μ are the Dirac gamma matrices, which define the algebraic structure of the theory. For the following, we also define a conjugate spinor,

$$\bar{\psi} = \psi^\dagger \gamma^0 \quad (1.9)$$

which satisfies the conjugate Dirac equation

$$\bar{\psi}(i\gamma^\mu \partial_\mu - m) = 0 \quad (1.10)$$

527 where the derivative acts to the left.

The Dirac equation is the dynamical equation for spin-1/2, but we'd like to express these dynamics via a Lagrangian density. Further, to have a relativistic description, we'd like to

have this be density be Lorentz invariant. These constraints lead to a Lagrangian of the form

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \quad (1.11)$$

⁵²⁸ where the Euler-Lagrange equation exactly recovers the Dirac equation.

The question now becomes how to marry the two Lagrangian descriptions that we have developed. Returning for a moment to classical electrodynamics, we know that the Hamiltonian for a charged particle in an electromagnetic field is described by

$$H = \frac{1}{2m}(\vec{p} - q\vec{A})^2 + q\phi. \quad (1.12)$$

Comparing this to the Hamiltonian for a free particle, we see that the modifications required are $\vec{p} \rightarrow \vec{p} - q\vec{A}$ and $E \rightarrow E - q\phi$. Using the canonical quantization trick of identifying \vec{p} with operator $-i\nabla$ and E with operator $i\frac{\partial}{\partial t}$, this identification becomes

$$i\partial_\mu \rightarrow i\partial_\mu - qA_\mu \quad (1.13)$$

Allowing for the naive substitution in the Dirac Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu(\partial_\mu + iqA_\mu) - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (1.14)$$

⁵²⁹ where the source term may be interpreted as coming from the Dirac fields themselves, namely,

⁵³⁰ $-q\bar{\psi}\gamma^\mu\psi A_\mu$.

Setting $q = e$ here (as appropriate for the case of an electron), and defining $D_\mu \equiv \partial_\mu + ieA_\mu$, this may then be written in the form

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (1.15)$$

⁵³¹ which is exactly the quantum electrodynamics Lagrangian.

⁵³² We have swept a few things under the rug here, however. Recall that the general form
⁵³³ of a Lagrangian is conventionally $\mathcal{L} = T - V$, where T is the kinetic term, and thus ought
⁵³⁴ to contain a derivative with respect to time (c.f. the standard $\frac{1}{2}m\frac{\partial x}{\partial t}$ familiar from basic
⁵³⁵ kinematics). More particularly, given the definition of conjugate momentum as $\partial\mathcal{L}/\partial\dot{q}$ for

536 $\mathcal{L}(q, \dot{q}, t)$ and $\dot{q} = \frac{\partial q}{\partial t}$, any field q which has no time derivative in the Lagrangian has 0
537 conjugate momentum, and thus no dynamics.

538 Looking at this final form, there is an easily identifiable kinetic term for the spinor fields
539 (just applying the D_μ operator). However trying to identify something similar for the A fields,
540 one comes up short – the antisymmetric nature of $F^{\mu\nu}$ term means that there is no time
541 derivative applied to A^0 .

542 What does this mean? A^μ is a four component object, but it would appear that only three
543 of the components have dynamics: we have too many degrees of freedom in the theory. This
544 is the principle behind *gauge symmetry* – an extra constraint on A^μ (a *gauge condition*) must
545 be defined such that a unique A^μ defines the theory and satisfies the condition. However,
546 we are free to choose this extra condition – the physics content of the theory should be
547 independent of this choice (that is, it should be *gauge invariant*).

To ground this a bit, let us return to basic electric and magnetic fields. These are physical quantities that can be measured, and are defined in terms of potentials as

$$\vec{E} = -\nabla\phi - \frac{\partial\vec{A}}{\partial t} \quad (1.16)$$

$$\vec{B} = \nabla \times \vec{A}. \quad (1.17)$$

548 It is easy to show, for any scalar function λ , that $\nabla \times \nabla\lambda = 0$. This implies that the physical
549 \vec{B} field is invariant under the transformation $\vec{A} \rightarrow \vec{A} + \nabla\lambda$ for any scalar function λ .

550 Under the same transformation of \vec{A} , the electric field \vec{E} becomes $-\nabla\phi - \frac{\partial\vec{A}}{\partial t} - \frac{\partial\nabla\lambda}{\partial t} =$
551 $-\nabla(\phi + \frac{\partial\lambda}{\partial t}) - \frac{\partial\vec{A}}{\partial t}$, such that, for the \vec{E} field to be unchanged, we must additionally apply
552 the transformation $\phi \rightarrow \phi - \frac{\partial\lambda}{\partial t}$.

This set of transformations to the potentials that leave the physical degrees of freedom invariant is expressed in our four vector notation naturally as

$$A_\mu \rightarrow A_\mu - \partial_\mu \lambda \quad (1.18)$$

553 where $A_\mu = (\phi, -\vec{A})$ with our sign convention. It should be noted that this function λ is an
554 arbitrary function of *local* spacetime, and thus expresses invariance of the physics content

555 under a local transformation.

Let us return to the Lagrangian for QED. In particular, focusing on the free Dirac piece

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \quad (1.19)$$

we note that if we apply a local transformation of the form $\psi \rightarrow e^{iq\lambda(x)}\psi$ (and correspondingly $\bar{\psi} \rightarrow \bar{\psi}e^{-iq\lambda(x)}$, by definition), the Lagrangian becomes

$$\bar{\psi}e^{-iq\lambda(x)}(i\gamma^\mu \partial_\mu - m)e^{iq\lambda(x)}\psi = \bar{\psi}e^{-iq\lambda(x)}(i\gamma^\mu \partial_\mu)e^{iq\lambda(x)}\psi - m\bar{\psi}\psi. \quad (1.20)$$

As $\partial_\mu(e^{iq\lambda(x)}\psi) = iq e^{iq\lambda(x)}(\partial_\mu \lambda(x))\psi + e^{iq\lambda(x)}\partial_\mu \psi$, this becomes

$$\bar{\psi}(i\gamma^\mu(\partial_\mu + iq\partial_\mu \lambda(x)) - m)\psi. \quad (1.21)$$

Thus, the free Dirac Lagrangian on its own is not invariant under this transformation. We may note, however, that on interaction with an electromagnetic field, as described above, this transformed Lagrangian may be packaged as:

$$\bar{\psi}(i\gamma^\mu(\partial_\mu + iq\partial_\mu \lambda(x) + iqA_\mu) - m)\psi = \bar{\psi}(i\gamma^\mu(\partial_\mu + iq(A_\mu + \partial_\mu \lambda(x))) - m)\psi. \quad (1.22)$$

556 since by the arguments above, the physics content of the Lagrangian is invariant under the
557 transformation $A_\mu \rightarrow A_\mu - \partial_\mu \lambda$, we may directly make this transformation, and remove this
558 extra $\partial_\mu \lambda(x)$ term. It is straightforward to verify that the $\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ term is invariant under
559 this same transformation of A_μ , so we may say that the QED Lagrangian is invariant under
560 local transformations of the form $\psi \rightarrow e^{iq\lambda(x)}\psi$.

561 These arguments illuminate some important concepts which will serve us well going forward.
562 First, while we have remained grounded in the “familiar” physics of electromagnetism for the
563 above, arguments of the “top down” variety would lead us to the exact same conclusions.
564 That is, suppose we wanted to construct a theory of spin-1/2 particles that was invariant
565 under local transformations of the form $\psi \rightarrow e^{iq\lambda(x)}\psi$. More broadly, we could say that we
566 desire this theory to be invariant under local $U(1)$ transformations, where $U(1)$ is exactly
567 this group, under multiplication, of complex numbers with absolute value 1. By very similar

arguments as above, we would see that, to achieve invariance, this theory would necessitate an additional degree of freedom, A_μ , with the exact properties that are familiar to us from electrodynamics. These arguments based on symmetries are extremely powerful in building theories with a less familiar grounding, as we will see in the following.

Second, we defined this quantity $D_\mu \equiv \partial_\mu + ieA_\mu$ above, seemingly as a matter of notational convenience. However, from the latter set of arguments, such a packaging takes on a new power: by explicitly including this gauge field A_μ which transforms in such a way as to keep invariance under a given transformation, the invariance is immediately more manifest. That is, to pose the $U(1)$ invariance in a more zoomed out way, under the transformation $\psi \rightarrow e^{iq\lambda(x)}\psi$, while

$$\bar{\psi}\partial_\mu\psi \rightarrow \bar{\psi}(\partial_\mu + iq\partial_\mu\lambda(x))\psi \quad (1.23)$$

with the extra term that gets canceled out by the gauge transformation of A_μ ,

$$\bar{\psi}D_\mu\psi \rightarrow \bar{\psi}D_\mu\psi \quad (1.24)$$

where this transformation is already folded in. This repackaging, called a *gauge covariant derivative* is much more immediately expressive of the symmetries of the theory.

Finally, to emphasize how fundamental these gauge symmetries are to the corresponding theory, let us examine the additional term needed for $U(1)$ invariance, $q\bar{\psi}\gamma^\mu A_\mu\psi$. While a first principles examination of Feynman rules is beyond the scope of this thesis, it is powerful to note that this is expressive of a QED vertex: the $U(1)$ invariance of the theory and the interaction between photons and electrons are inextricably tied together.

1.3 An Aside on Group Theory

Quantum electrodynamics is very familiar and well covered, and provides (both historically and in this thesis) a nice bridge between “standard” physics and the language of symmetries and quantum field theory. However, now that we are acquainted with the language, we may set up to dive a bit deeper. To begin, let us look again at the $U(1)$ group that is so fundamental to QED. We have expressed this via a set of transformations on our Dirac spinor

objects, ψ , of the form $e^{iq\lambda(x)}$. Note that such transformations, though they are local (i.e. a function of spacetime) are purely *phase* transformations. Relatedly, $U(1)$ is an Abelian group, meaning that group elements commute.

To set up language to generalize beyond $U(1)$, note that we may equivalently write $U(1)$ elements as $e^{ig\vec{\alpha}(x)\cdot\vec{T}}$, $\vec{\alpha}(x)$ and \vec{T} and are vectors in the space of *generators* of the group, with each $\alpha^a(x)$ an associated scalar function to generator t^a , and g is some scalar strength parameter. Of course this is a bit silly for $U(1)$, which has a single generator, and thus reduces to the transformation we discussed above. However, this becomes much more useful for groups of higher degree, with more generators and degrees of freedom.

To discuss these groups in a bit more detail, note that $U(n)$ is the unitary group of degree n , and corresponds to the group of $n \times n$ unitary matrices (that is, $U^\dagger U = UU^\dagger = 1$). Given that group elements are $n \times n$, this means that there are n^2 degrees of freedom: n^2 generators are needed to characterize the group.

For $U(1)$, this is all consistent with what we have said above – the group of 1×1 unitary matrices have a single generator, and the phases we identify above clearly satisfy unitarity. Note that these degrees of freedom for the gauge group also characterize the number of gauge bosons we need to satisfy the local symmetry: for $U(1)$, we need one gauge boson, the photon.

Of relevance for the Standard Model are also the special unitary groups $SU(n)$. These are defined similarly to the unitary groups, with the additional requirement that group elements have determinant 1. This extra constraint removes 1 degree of freedom: groups are characterized by $n^2 - 1$ generators.

In particular, we will examine the groups $SU(2)$ in the context of the weak interaction, with an associated $2^2 - 1 = 3$ gauge bosons (cf. the W^\pm and Z bosons), and $SU(3)$, with an associated $3^2 - 1 = 8$ gauge bosons (cf. gluons of different flavors). Note that these groups are non-Abelian (2×2 or 3×3 matrices do not, in general, commute), leading to a variety of complications. However, both of these theories feature interactions with spin-1/2 particles, with transformations of a very similar form: $\psi \rightarrow e^{ig\vec{\alpha}(x)\cdot\vec{T}}\psi$, and the general framing of the arguments for QED will serve us well in the following.

613 **1.4 Quantum Chromodynamics**

614 In some sense, the simplest extension the development of QED is quantum chromodynamics
615 (QCD). QCD is a theory in which, once the basic dynamics are framed (a non-trivial task!)
616 the group structure becomes apparent. The quark model, developed by Murray Gell-Mann [5]
617 and George Zweig [6], provided the fundamental particles involved in the theory, and had
618 great success in explaining the expanding zoo of experimentally observed hadronic states.

619 Some puzzles were still apparent – the Δ^{++} baryon, e.g., is composed of three up quarks,
620 u , with aligned spins. As quarks are fermions, such a state should not be allowed by the
621 Pauli exclusion principle. The existence of such a state in nature implies the existence of
622 another quantum number, and a triplet of values, called *color charge* was proposed by Oscar
623 Greenberg [7]. With these pieces in place, the structure becomes more apparent, as elucidated
624 by Han and Nambu [8].

625 Let us reason our way to the symmetries using color charge. Experimentally, we know
626 that there is this triplet of color charge values r, g, b (the “plus” values, cf. electric charge)
627 and correspondingly anti-color charge $\bar{r}, \bar{g}, \bar{b}$ (the “minus” values). Supposing that the force
628 behind QCD (the *strong force*) is, similar to QED, interactions between fermions mediated
629 by gauge bosons (quarks and gluons respectively), we can start to line up the pieces.

630 What color charge does a gluon have? Similarly to electric charge, we may associate
631 particles with color charge, anti-particles with anti-color charge. Notably, free particles
632 observed experimentally are colorless (have no color charge). Thus, in order for charge to
633 be conserved throughout such processes, this already implies that there are charged gluons.
634 Further, examining color flow diagrams such as *TODO: insert*, it is apparent first that a
635 gluon has not one but two associated color charges and second that these two must be one
636 color charge and one anti-color charge.

637 Counting up the available types of gluons, then, we come up with nine. Six of mixed
638 color type: $r\bar{b}, r\bar{g}, b\bar{r}, b\bar{g}, g\bar{b}$, and $g\bar{r}$, and three of same color type: $r\bar{r}, g\bar{g}$, and $b\bar{b}$. In practice,
639 however, these latter three are a bit redundant: all express a colorless gluon, which, if we

could observe this as a free particle, would be indistinguishable from each other. The *color singlet* state is then a mix of these, $\frac{1}{\sqrt{3}}(r\bar{r} + g\bar{g} + b\bar{b})$, leaving two unclaimed degrees of freedom, which may be satisfied by the linearly independent combinations $\frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g})$ and $\frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b})$.

We thus have an octet of color states plus a colorless singlet state. If this colorless singlet state existed, however, we would be able to observe it, not only via interactions with quarks, but as a free particle. Since do not observe this in nature, this restricts us to 8 gluons. The simplest group with a corresponding 8 generators is $SU(3)$. Under the assumption that $SU(3)$ is the local gauge symmetry of the strong interaction, we may proceed in a similar way as we did for QED. The gauge transformation is $\psi \rightarrow e^{ig_S \vec{\alpha}(x) \cdot \vec{T}} \psi$, where \vec{T} is an eight component vector of the generators of $SU(3)$, often expressed via the Gell-Mann matrices, λ^a , as $t^a = \frac{1}{2}\lambda^a$, and the spinor ψ represents the fields corresponding to quarks.

This $SU(3)$ symmetry exactly expresses the color structure elucidated above – the Gell-Mann matrices are an equivalent presentation of the color combinations described above. Proceeding by analogy to QED, gauge invariance is achieved by introducing eight new degrees of freedom, G_μ^a , which are the gauge fields corresponding to the gluons, with the gauge covariant derivative then analogously taking the form $D_\mu \equiv \partial_\mu + ig_S G_\mu^a t^a$.

Recall from the QED derivation that the field strength tensor, $F^{\mu\nu}$ is a rank two antisymmetric tensor which is manifestly gauge invariant and which describes the physical dynamics of the A_μ field. We would like to analogously define a term for the gluon fields. Repackaging this QED tensor, it is apparent that

$$[D_\mu, D_\nu] = D_\mu D_\nu - D_\nu D_\mu \quad (1.25)$$

$$= (\partial_\mu + iqA_\mu)(\partial_\nu + iqA_\nu) - (\partial_\nu + iqA_\nu)(\partial_\mu + iqA_\mu) \quad (1.26)$$

$$= \partial_\mu \partial_\nu + iq\partial_\mu A_\nu + iqA_\mu \partial_\nu + (iq)^2 A_\mu A_\nu - (\partial_\nu \partial_\mu + iq\partial_\nu A_\mu + iqA_\nu \partial_\mu + (iq)^2 A_\nu A_\mu) \quad (1.27)$$

$$= iq(\partial_\mu A_\nu - \partial_\nu A_\mu) + (iq)^2 (A_\mu A_\nu - A_\nu A_\mu) \quad (1.28)$$

$$= iq(\partial_\mu A_\nu - \partial_\nu A_\mu) + (iq)^2 [A_\mu, A_\nu]. \quad (1.29)$$

We proceed through this derivation to highlight that, in the specific case of QED, with its Abelian $U(1)$ gauge symmetry, the field commutator vanishes, leaving exactly the definition of $F_{\mu\nu}$ as described above, i.e.,

$$F_{\mu\nu} = \frac{1}{iq}[D_\mu, D_\nu]. \quad (1.30)$$

We may proceed to define an analogous field strength term for G_μ^a in a similar way:

$$G_{\mu\nu} = \frac{1}{ig_S}[D_\mu, D_\nu] \quad (1.31)$$

This has an extremely nice correspondence, but is complicated by the non-Abelian nature of $SU(3)$, with

$$G_{\mu\nu} = \partial_\mu(G_\nu^a t^a) - \partial_\nu(G_\mu^a t^a) + ig_s[G_\mu^a t^a, G_\nu^a t^a]. \quad (1.32)$$

in which the field commutator term is non-zero. In particular (since each term is summing over a , so we may relabel) as

$$[G_\mu^a t^a, G_\nu^b t^b] = [t^a, t^b]G_\mu^a G_\nu^b \quad (1.33)$$

and as $[t^a, t^b] = if^{abc}t^c$ for the Gell-Mann matrices, where f^{abc} are the structure constants of $SU(3)$, we have

$$G_{\mu\nu} = \partial_\mu(G_\nu^a t^a) - \partial_\nu(G_\mu^a t^a) - g_s f^{abc} t^c G_\mu^a G_\nu^b \quad (1.34)$$

$$= t^a(\partial_\mu G_\nu^a - \partial_\nu G_\mu^a - f^{bca}G_\mu^b G_\nu^c) \quad (1.35)$$

$$= t^a G_{\mu\nu}^a \quad (1.36)$$

⁶⁵⁷ for $G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - f^{abc}G_\mu^b G_\nu^c$.

⁶⁵⁸ This gives the component of the field strength corresponding to a particular gauge field a ,
⁶⁵⁹ where the first two terms have the familiar form of the QED field strength, while the last
⁶⁶⁰ term is new, and explicitly related to the group structure via the f^{abc} constants. In terms
⁶⁶¹ of the physics content of the theory, this latter term gives rise to a gluon *self-interaction*, a
⁶⁶² distinguishing feature of QCD.

⁶⁶³ Similarly as in QED, a Lorentz invariant combination of field strength tensors may be made
⁶⁶⁴ as $G_{\mu\nu}G^{\mu\nu}$. However, this is not manifestly gauge invariant. Under a gauge transformation

- ⁶⁶⁵ U , the covariant derivative behaves as $D^\mu \rightarrow UD^\mu U^{-1}$, corresponding to $G^{\mu\nu} \rightarrow UG^{\mu\nu}U^{-1}$.
⁶⁶⁶ The cyclic property of the trace thus ensures the gauge invariance of $\text{tr}(G_{\mu\nu}G^{\mu\nu})$, which we
⁶⁶⁷ will write as $G_{\mu\nu}^a G_a^{\mu\nu}$ with the implied sum over generators a .

Packaging up the theory, it is tempting to copy the form of the QED Lagrangian, with the identifications we have made above:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}. \quad (1.37)$$

However this is not quite correct due to the $SU(3)$ nature of the theory. In terms of the physics, the Dirac fields ψ have associated color charge, which must interact appropriately with the G_μ fields. Mathematically, the generators t^a are 3×3 matrices, while the ψ are four component spinors. Adding a color index to the Dirac fields, i.e., ψ_i where i runs over the three color charges, and similarly indexing the generators t_{ij}^a , we may then express the $SU(3)$ gauge covariant derivative component-wise as

$$(D_\mu)_{ij} = \partial_\mu \delta_{ij} + ig_S G_\mu^a t_{ij}^a \quad (1.38)$$

- ⁶⁶⁸ where δ_{ij} is the Kronecker delta, as ∂_μ does not participate in the $SU(3)$ structure.

The Lagrangian then becomes

$$\mathcal{L} = \bar{\psi}_i(i(\gamma^\mu D_\mu)_{ij} - m\delta_{ij})\psi_j - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}. \quad (1.39)$$

- ⁶⁶⁹ and we have constructed QCD.

⁶⁷⁰ 1.5 The Weak Interaction

- ⁶⁷¹ One of the first theories of the weak interaction was from Enrico Fermi [9], in an effort to
⁶⁷² explain beta decay, a process in which an electron or positron is emitted from an atomic
⁶⁷³ nucleus, resulting in the conversion of a neutron to a proton or proton to a neutron respectively.
⁶⁷⁴ Fermi's hypothesis was of a direct interaction between four fermions. However, in the advent of
⁶⁷⁵ QED, it is natural to wonder if a theory based on mediator particles and gauge symmetries
⁶⁷⁶ applies to the weak force as well. The modern formulation of such a theory is due to Sheldon

677 Glashow, Steven Weinberg, and Abdus Salam [10], and is what we will describe in the
678 following.

679 Considering emission of an electron, Fermi's theory involves an initial state neutron that
680 transitions to a proton with the emission of an electron and a neutrino. This transition
681 gives a hint that something slightly more complicated is happening than in QED: there is an
682 apparent mixing between particle types.

683 Now, with the assumption there are mediators for such an interaction, we further know
684 from beta decay and charge conservation that there must be at least two such degrees of
685 freedom: e.g. one that decays to an electron and neutrino (W^-) and one that decays to a
686 positron and neutrino (W^+). From consideration of the process $e^+e^- \rightarrow W^+W^-$, it turns
687 out that with just these two degrees of freedom, the cross section for this process increases
688 without limit as a function of center-of-mass energy, ultimately violating unitarity (more
689 W^+W^- pairs come out than e^+e^- pairs go in). This is resolved with a third, neutral degree
690 of freedom, the Z boson, whose contribution interferes negatively, regulating this process.

691 This leads to three degrees of freedom for the gauge symmetry of the weak interactions, so
692 we thus need a theory which is locally invariant under transformations of a group with three
693 generators. The simplest such choice is $SU(2)$. We may follow a very similar prescription as
694 for QED and QCD: $SU(2)$ has three generators, which implies the existence of three gauge
695 bosons, call them W_μ^k . The gauge transformation may be expressed as $\psi \rightarrow e^{ig_W \vec{\alpha}(x) \cdot \vec{T}} \psi$, where
696 in this case the generators are for $SU(2)$, which may be written in terms of the familiar Pauli
697 matrices: $\vec{T} = \frac{1}{2}\vec{\sigma}$. The structure constants for $SU(2)$ are the antisymmetric Levi-Civita
698 tensor, so the corresponding gauge covariant derivative is $D_\mu \equiv \partial_\mu + ig_W W_\mu^k t^k$, and the field
699 strength tensor is $W_{\mu\nu}^k = \partial_\mu W_\nu^k - \partial_\nu W_\mu^k - \epsilon^{ijk} W_\mu^k W_\nu^k$.

The corresponding Lagrangian would thus be

$$\mathcal{L} = \bar{\psi}_i (i(\gamma^\mu D_\mu)_{ij} - m\delta_{ij}) \psi_j - \frac{1}{4} W_{\mu\nu}^k W_k^{\mu\nu} \quad (1.40)$$

700 where indices i and j run over $SU(2)$ charges.

701 On considering some of the details, the universe unfortunately turns out to be a bit

more complicated. However, this still provides a useful starting place for elucidating the theory of weak interactions. First off, let us consider the particle content, namely, what do the Dirac fields correspond to? This is still a theory of fermionic interactions with gauge bosons. However, we might notice that the fermion content of this theory is both a) broader than QCD, as we know experimentally (cf. beta decay) that both quarks and leptons (e.g. electrons) participate in the weak interaction and b) this fermion content seemingly has a large overlap with QED. In terms of the gauge bosons, we know that at both W^+ and W^- are electrically charged – this means that we expect some interaction of the weak theory with electromagnetism.

However, before diving deeper into this apparent connection between the weak interaction and QED, let us focus on the gauge symmetry. In QCD, the $SU(3)$ content of the theory is expressed via a contraction of color indices – the theory allows for transitions between quarks of one color and quarks of another. Thinking similarly in terms of $SU(2)$ transitions, the beta decay example is already fruitful – there is a transition between an electron and its corresponding neutrino, as well as between two types of quark. In particular, for the case of neutron (with quark content udd) and proton (with quark content udu), the weak interaction provides for a transition from down to up quark.

Such $SU(2)$ dynamics are described via a quantity called *weak isospin*, denoted I_W with third component $I_W^{(3)}$, and can be thought of in a very similar way as color charge in QCD (i.e. as the charge corresponding to the weak interaction). Since $SU(2)$ is 2×2 , there are two such charge states for the fermions, denoted as $I_W^{(3)} = \pm\frac{1}{2}$. This means that the bosons must have $I_W = 1$ such that, by sign convention corresponding to electric charge, the W^+ boson has $I_W^{(3)} = +1$, the Z boson has $I_W^{(3)} = 0$, and the W^- boson has $I_W^{(3)} = -1$.

From conservation of electric charge, this means that transitions involving a W^\pm are between particles that differ by ± 1 in both weak isospin $I_W^{(3)}$ and electric charge. We may thus line up all such doublets as:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}, \begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \begin{pmatrix} t \\ b' \end{pmatrix} \quad (1.41)$$

725 with the top corresponding to the lower weak isospin and electric charge particles, and the
726 lower quark entries (d' , etc) corresponding to the weak quark eigenstates (which are related
727 to the mass eigenstates by the CKM matrix *TODO: more detail*). Similar doublets may be
728 constructed for the corresponding anti-particles.

The fundamental structuring of these transitions around both electric and weak charge is again indicative of a natural connection. However, nature is again a bit more complicated than we have described. This is because the weak interaction is a *chiral* theory. For massless particles, chirality is the same as the perhaps more intuitive *helicity*. This describes the relationship between a particle's spin and momentum: if the spin vector points in the same direction as the momentum vector, helicity is positive (the particle is “right-handed”), and if the two point in opposite directions, the helicity is negative (the particle is “left-handed”). More concretely:

$$H = \frac{\vec{s} \cdot \vec{p}}{|\vec{s} \cdot \vec{p}|}. \quad (1.42)$$

For massive particles, this generalizes a bit – in the language of Dirac fermions that we have developed, we define projection operators

$$P_R = \frac{1}{2}(1 + \gamma^5) \quad \text{and} \quad P_L = \frac{1}{2}(1 - \gamma^5) \quad (1.43)$$

729 for right and left-handed chiralities respectively – acting on a Dirac field with such operators
730 projects the field onto the corresponding chiral state.

Experimentally, this pops up via parity violation and the famous $V - A$ theory. For the scope of this thesis, it is sufficient to say that the weak interaction is only observed to take place for left-handed particles (and correspondingly, right-handed anti-particles). We therefore modify the theory stated above by projecting all fermions participating in the weak interaction onto respective chiral states – in particular, the $SU(2)$ gauge symmetry only acts on left-handed particles and right-handed anti-particles. We therefore modify the theory appropriately, denoting the chiral projected gauge symmetry as $SU(2)_L$, and similarly for the

Dirac fields. In particular, the weak isospin doublets listed above must now be left-handed:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L \quad (1.44)$$

⁷³¹ and right-handed particle states are placed in singlets and assigned 0 charge under $SU(2)_L$
⁷³² ($I_W = I_W^{(3)} = 0$).

With all of these assignments, let us revisit our guess at the form of the weak interaction Lagrangian. First, dwelling on the kinetic term $\bar{\psi}_i(i(\gamma^\mu D_\mu)_{ij}\psi_j)$, we note that the assigning of left-handed fermions to isospin doublets and right-handed fermions to isospin singlets allows us to remove explicit $SU(2)$ indices by treating these as the fundamental objects, that is, for a single *generation* of fermions, we may write:

$$\bar{Q}i\gamma^\mu D_\mu Q + \bar{u}i\gamma^\mu D_\mu u + \bar{d}i\gamma^\mu D_\mu d + \bar{L}i\gamma^\mu D_\mu L + \bar{e}i\gamma^\mu D_\mu e \quad (1.45)$$

⁷³³ for left-handed doublets Q and L for quarks and electron fields respectively and right handed
⁷³⁴ singlets u and d for up and down quark fields and e for electrons.

More concisely, and summing over the three generations of fermions, we may write

$$\sum_f \bar{f}i\gamma^\mu D_\mu f \quad (1.46)$$

⁷³⁵ where the f are understood to run over the fermion chiral doublets and singlets as above.

This then leaves our Lagrangian as

$$\mathcal{L} = \sum_f \bar{f}i\gamma^\mu D_\mu f - \frac{1}{4}W_{\mu\nu}^k W_k^{\mu\nu} \quad (1.47)$$

$$= \sum_f \bar{f}\gamma^\mu(i\partial_\mu - \frac{1}{2}g_W W_\mu^k \sigma_k)f - \frac{1}{4}W_{\mu\nu}^k W_k^{\mu\nu}, \quad (1.48)$$

⁷³⁶ where we have expanded the covariant derivative for clarity. You may note that we have
⁷³⁷ dropped the mass term in the equation above – we will discuss this in detail in just a moment.

First, however, we return to the above comment about fermion content – we neglected to include the sum over fermions in our QED derivation for simplicity. However, all of the

fermions considered in the discussion of the weak interaction have an electric charge (except for the neutrinos). It would be nice to repackage the theory into a coherent *electroweak* theory. This is fairly straightforward when considering the gauge approach – from the discussion above we should expect the electroweak gauge group to be something like $SU(2) \times U(1)$, with four corresponding gauge bosons. Consider a gauge theory with group $SU(2)_L \times U(1)_Y$ – that is, the same weak interaction as discussed previously, but a new $U(1)_Y$ gauge group for electromagnetism, with transformations defined as

$$\psi \rightarrow e^{ig' \frac{Y}{2} \lambda(x)} \psi \quad (1.49)$$

⁷³⁸ with *weak hypercharge* Y .

Similarly to our discussion of QED, we may write the $U(1)_Y$ gauge field as B_μ , and interactions with the Dirac fields take the form $g' \frac{Y}{2} \gamma^\mu B_\mu \psi$. The relationship between this hypercharge and new B_μ field and classical electrodynamics is not so obvious – however it is convenient to parametrize as

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \quad (1.50)$$

⁷³⁹ where A_μ and Z_μ are the physical fields, and we pick W_μ^3 as the neutral weak boson.

⁷⁴⁰ Note that in the $SU(2)_L \times U(1)_Y$ theory, the Lagrangian must be invariant under all of
⁷⁴¹ the local gauge transformations. In particular, this means that the hypercharge must be the
⁷⁴² same for fermion fields in each weak doublet to preserve $U(1)_Y$ invariance. This gives insight
⁷⁴³ into the relation between the charges of $SU(2)_L \times U(1)_Y$ and electric charge. In particular
⁷⁴⁴ we know that the hypercharge, Y , of e^- ($I_W^{(3)} = -\frac{1}{2}$) and ν_e ($I_W^{(3)} = +\frac{1}{2}$) is the same.

Supposing that $Y = \alpha I_W^{(3)} + \beta Q$, we must have $-\alpha \frac{1}{2} - \beta = \alpha \frac{1}{2} \implies \beta = -\alpha$. Therefore, choosing an overall scaling from convention,

$$Y = 2(Q - I_W^{(3)}). \quad (1.51)$$

⁷⁴⁵ Some of these particular forms are best understood in the context of the Higgs mechanism
⁷⁴⁶ – we will return to this discussion below.

⁷⁴⁷ **1.6 The Higgs Potential and the SM**

⁷⁴⁸ In the above, we have neglected a discussion of masses. However there are several things to
⁷⁴⁹ sort out here. In the first place, we know experimentally that the weak interactions occur
⁷⁵⁰ over very short ranges at low energies (e.g., why Fermi's effective four fermion interaction was
⁷⁵¹ such a good description). This is consistent with massive W^\pm and Z bosons (and indeed, this
⁷⁵² is seen experimentally). However, requiring local gauge invariance forbids mass terms in the
⁷⁵³ Lagrangian. In the simple $U(1)$ QED example, such a term would have the form $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$,
⁷⁵⁴ which is not invariant under the transformation $A_\mu \rightarrow A_\mu - \partial_\mu \lambda$, and similar arguments hold
⁷⁵⁵ for gauge bosons in the electroweak theory and QCD.

Similar issues are encountered with fermions – in the electroweak theory above, the gauge symmetries are separated into left and right handed chirality via doublet and singlet states. This means that a mass term would need to be separated as well. Such a term would have the form:

$$m\bar{f}f = m(\bar{f}_L + \bar{f}_R)(f_L + f_R) \quad (1.52)$$

$$= m(\bar{f}_L f_L + \bar{f}_L f_R + \bar{f}_R f_L + \bar{f}_R f_R) \quad (1.53)$$

$$= m(\bar{f}_L f_R + \bar{f}_R f_L) \quad (1.54)$$

⁷⁵⁶ where we have used that $f_{L,R} = P_{L,R}f$, $\bar{f}_{L,R} = \bar{f}P_{R,L}$, and $P_R P_L = P_L P_R = 0$. As left
⁷⁵⁷ and right-handed particles transform differently under $SU(2)_L$, this is manifestly not gauge
⁷⁵⁸ invariant.

⁷⁵⁹ The question then becomes: how do we include particle masses while preserving the
⁷⁶⁰ gauge properties of our theory? The answer, due to Robert Brout and François Englert [11],
⁷⁶¹ Peter Higgs [12], and Gerald Guralnik, Richard Hagen, and Tom Kibble [13] comes via the
⁷⁶² Higgs mechanism, which we will describe in the following. Importantly for this thesis, this
⁷⁶³ mechanism predicts the existence of a physical particle, the Higgs boson, and a particle
⁷⁶⁴ consistent with the Higgs boson was seen by both ATLAS [14] and CMS [15] in 2012.

To explain the Higgs, we focus first on generating masses for the electroweak gauge bosons.

Consider adding two complex scalar fields ϕ^+ and ϕ^0 to the Standard Model embedded in a weak isospin doublet ϕ . We may write the doublet as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (1.55)$$

765 where we explicitly note the four available degrees of freedom.

The Lagrangian for such a doublet takes the form

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \quad (1.56)$$

where V is the corresponding potential. Considering the particular form

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (1.57)$$

766 we may notice that this has some interesting properties. Considering, as illustration, a similar
767 potential for a real scalar field, $\mu^2 \chi^2 + \lambda \chi^4$, taking the derivative and setting it equal to 0
768 yields extrema when $\chi = 0$ and $(\mu^2 + 2\lambda\chi^2) = 0 \implies \chi^2 = -\frac{\mu^2}{2\lambda}$. For $\mu^2 > 0$, there is a
769 unique minimum at $\chi = 0$, and for $\mu^2 < 0$ there are degenerate minima at $\chi = \pm\sqrt{-\frac{\mu^2}{2\lambda}}$.
770 Note that we take $\lambda > 0$, otherwise the only minima in the theory are trivial.

The same simple calculus for the complex Higgs doublet above yields degenerate minima for $\mu^2 < 0$ at

$$\phi^\dagger \phi = \frac{1}{2} (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{v}{2} = -\frac{\mu^2}{2\lambda} \quad (1.58)$$

However, though there is this degenerate set of minima, there can only be a single *physical* vacuum state (we say that the symmetry is *spontaneously broken*). Without loss of generality, we may align our axes such that the physical vacuum state is at

$$\langle 0 | \phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (1.59)$$

771 where we have explicitly chosen a real, non-zero vacuum expectation value for the neutral
772 component of the Higgs doublet to maintain a massless photon, as we shall see. Physically,
773 however, this makes sense - the vacuum is not electrically charged.

The vacuum is a classical state – we want a quantum one. We may express fluctuations about this nonzero expectation value via an expansion as $v + \eta(x) + i\xi(x)$. However, renaming of fields is only meaningful for the non-zero vacuum component - we thus have:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + \eta(x) + i\phi_4 \end{pmatrix}. \quad (1.60)$$

where we may expand the Lagrangian listed above:

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2. \quad (1.61)$$

It is an exercise in algebra to plug in the expansion about v into this Lagrangian: first expanding the potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \quad (1.62)$$

$$= \mu^2 \left(\sum_i \phi_i(x)^2 + (v + \eta(x))^2 \right) + \lambda \left(\sum_i \phi_i(x)^2 + (v + \eta(x))^2 \right) \quad (1.63)$$

$$= -\frac{1}{4} \lambda v^4 + \lambda v^2 \eta^2 + \lambda v \eta^3 + \frac{1}{4} \lambda \eta^4 \quad (1.64)$$

$$+ \frac{1}{2} \lambda \sum_{i \neq j} \phi_i^2 \phi_j^2 + \lambda v \eta \sum_i \phi_i(x)^2 + \frac{1}{2} \lambda \eta^2 \sum_i \phi_i(x)^2 + \frac{1}{4} \sum_i \phi_i(x)^4 \quad (1.65)$$

⁷⁷⁴ where the sums are over the $i \in 1, 2, 4$, that is, the fields with 0 vacuum expectation, and we
⁷⁷⁵ have used the definition $\mu^2 = -\lambda v^2$.

⁷⁷⁶ Within this potential, we note a quadratic term in $\eta(x)$ which we may identify with a
⁷⁷⁷ mass, namely $m_\eta = \sqrt{2\lambda v^2}$, whereas the ϕ_i are massless. These ϕ_i are known as *Goldstone*
⁷⁷⁸ *bosons*, and correspond to quantum fluctuations along the minimum of the potential. Of
⁷⁷⁹ particular note for this thesis are the interaction terms $\lambda v \eta^3$ and $\frac{1}{4} \lambda \eta^4$, expressing trilinear
⁷⁸⁰ and quartic self-interactions of the η field.

Expanding the kinetic term

$$(\partial_\mu \phi)^\dagger (\partial^\mu \phi) = \frac{1}{2} \sum_i (\partial_\mu \phi_i)(\partial^\mu \phi_i) + \frac{1}{2} (\partial_\mu(v + \eta(x)))(\partial^\mu(v + \eta(x))) \quad (1.66)$$

$$= \frac{1}{2} \sum_i (\partial_\mu \phi_i)(\partial^\mu \phi_i) + \frac{1}{2} (\partial_\mu \eta)(\partial^\mu \eta) \quad (1.67)$$

⁷⁸¹ in a similar way, completing the story of three massless degrees of freedom (Goldstone bosons)
⁷⁸² and one massive one.

Now, this doublet is embedded in an $SU(2)_L \times U(1)$ theory, so we would like to preserve that gauge invariance. This is achieved in the same way as for the Dirac fields, with the introduction of the electroweak gauge covariant derivative such that the Lagrangian for the Higgs doublet and the electroweak bosons is just

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 - \frac{1}{4} W_{\mu\nu}^k W_k^{\mu\nu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad (1.68)$$

⁷⁸³ with $D_\mu = \partial_\mu + ig_W W_\mu^k t^k + ig' \frac{Y}{2} B_\mu$.

We note that it is convenient to pick a gauge such that the Goldstone fields do not appear in the Lagrangian, upon which we may identify the field $\eta(x)$ with the physical Higgs field, $h(x)$. The field mass terms then very apparently come via the covariant derivative, namely, as

$$W_\mu^k \sigma^k + B_\mu = \begin{pmatrix} W_\mu^3 + B_\mu & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 + B_\mu \end{pmatrix} \quad (1.69)$$

we may then write

$$D_\mu \phi = \frac{1}{2\sqrt{2}} \begin{pmatrix} 2\partial_\mu + ig_W W_\mu^3 + ig' Y B_\mu & ig_W W_\mu^1 + \frac{1}{2} g_W W_\mu^2 \\ ig_W W_\mu^1 - g_W W_\mu^2 & 2\partial_\mu - ig_W W_\mu^3 + ig' Y B_\mu \end{pmatrix} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \quad (1.70)$$

$$= \frac{1}{2\sqrt{2}} \begin{pmatrix} ig_W (W_\mu^1 - iW_\mu^2)(v + h) \\ (2\partial_\mu - ig_W W_\mu^3 + ig' Y B_\mu)(v + h) \end{pmatrix} \quad (1.71)$$

⁷⁸⁴ As identified above, $Y = 2(Q - I_W^{(3)})$. The Higgs has 0 electric charge, and the lower doublet
⁷⁸⁵ component has $I_W^{(3)} = -\frac{1}{2}$, yielding $Y = 1$.

Computing $(D_\mu \phi)^\dagger (D^\mu \phi)$, then, yields

$$\frac{1}{8} g_W^2 (W_\mu^1 + iW_\mu^2)(W^{\mu 1} - iW^{\mu 2})(v + h)^2 + \frac{1}{8} (2\partial_\mu + ig_W W_\mu^3 - ig' B_\mu)(2\partial^\mu - ig_W W^{\mu 3} + ig' B^\mu)(v + h)^2 \quad (1.72)$$

and extracting terms quadratic in the fields gives

$$\frac{1}{8} g_W^2 v^2 (W_{\mu 1} W^{\mu 1} + W_{\mu 2} W^{\mu 2}) + \frac{1}{8} v^2 (g_W W_\mu^3 - g' B_\mu)(g_W W^{\mu 3} - g' B^\mu) \quad (1.73)$$

meaning that W_μ^1 and W_μ^2 have masses $m_W = \frac{1}{2}g_W v$. The neutral boson case is a bit more complicated. Writing the corresponding term as

$$\frac{1}{8}v^2 \begin{pmatrix} W_\mu^3 & B_\mu \end{pmatrix} \begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix} \begin{pmatrix} W^{\mu 3} \\ B^\mu \end{pmatrix} \quad (1.74)$$

we note that we must diagonalize this mass matrix to get the physical mass eigenstates. Doing so in the usual way yields eigenvalues 0 , $g'^2 + g_W^2$, thus corresponding to $m_\gamma = 0$ and $m_Z = \frac{1}{2}v\sqrt{g'^2 + g_W^2}$, with physical fields as the (normalized) eigenvectors

$$A_\mu = \frac{g' W_\mu^3 + g_W B_\mu}{\sqrt{g_W^2 + g'^2}} \quad (1.75)$$

$$Z_\mu = \frac{g_W W_\mu^3 - g' B_\mu}{\sqrt{g_W^2 + g'^2}} \quad (1.76)$$

From this form, the angular parametrization of the physical fields is very apparent, namely, defining

$$\tan \theta_W = \frac{g'}{g_W}, \quad (1.77)$$

these equations may be written in terms of the single parameter θ_W as

$$A_\mu = \cos \theta_W B_\mu + \sin \theta_W W_\mu^3 \quad (1.78)$$

$$Z_\mu = -\sin \theta_W B_\mu + \cos \theta_W W_\mu^3 \quad (1.79)$$

and, notably, from the above equations,

$$\frac{m_W}{m_Z} = \cos \theta_W. \quad (1.80)$$

To get the mass terms from Equation 1.72, we extracted those terms quadratic in fields, i.e., the v^2 terms within $(v + h)^2$. However there are also terms of the form VVh and $VVhh$ that arise, which describe the Higgs interactions with the corresponding vector bosons $V = W^\pm, Z$. Namely, identifying physical W bosons as

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2) \quad (1.81)$$

we may express the first term of Equation 1.72 as

$$\frac{1}{4}g_W^2 W_\mu^- W^{+\mu} (v + h)^2 = \frac{1}{4}g_W^2 v^2 W_\mu^- W^{+\mu} + \frac{1}{2}g_W^2 v W_\mu^- W^{+\mu} h + \frac{1}{4}g_W^2 W_\mu^- W^{+\mu} h^2 \quad (1.82)$$

with the first term corresponding to the mass term $m_W = \frac{1}{2}g_W v$, and the second two terms corresponding to hW^+W^- and hhW^+W^- vertices. Of particular note is the coupling strength

$$g_{HWW} = \frac{1}{2}g_W^2 v = g_W m_W \quad (1.83)$$

⁷⁸⁶ which is proportional to the W mass – an analysis with the form of the physical Z boson
⁷⁸⁷ finds that the coupling g_{HZZ} is also proportional to the Z mass.

The Higgs coupling to fermions (in particular to quarks) is of particular interest for this thesis. We showed above that a naive introduction of a mass term

$$m\bar{f}f = m(\bar{f}_L f_R + \bar{f}_R f_L) \quad (1.84)$$

⁷⁸⁸ is manifestly not gauge invariant because right and left handed particles transform differently
⁷⁸⁹ under $SU(2)_L$. However, because the Higgs is constructed via an $SU(2)_L$ doublet, ϕ , writing
⁷⁹⁰ a fermion doublet as L and conjugate \bar{L} , it is apparent that $\bar{L}\phi$ is invariant under $SU(2)_L$.

Combining with the right handed singlet, R , creates a term invariant under $SU(2)_L \times U(1)_Y$, $\bar{L}\phi R$ (and correspondingly $(\bar{L}\phi R)^\dagger$), such that we may include Yukawa [16] terms

$$\mathcal{L}_{Yukawa} = -g_f \left[\begin{pmatrix} \bar{f}_1 & \bar{f}_2 \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} f_R + \bar{f}_R \begin{pmatrix} \phi^{+*} & \phi^{0*} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}_L \right] \quad (1.85)$$

⁷⁹¹ where g_f is a corresponding Yukawa coupling, f_1 and f_2 have been used to denote components
⁷⁹² of the left-handed doublet and f_R the corresponding right-handed singlet.

After spontaneous symmetry breaking, with the gauge as described above to remove the Goldstone fields, the Higgs doublet becomes

$$\phi(x) = \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (1.86)$$

giving rise to terms such as

$$-\frac{1}{\sqrt{2}}g_f v(\bar{f}_{2L}\bar{f}_R + \bar{f}_R f_{2L}) - \frac{1}{\sqrt{2}}g_f h(\bar{f}_{2L}\bar{f}_R + \bar{f}_R f_{2L}) \quad (1.87)$$

where we have kept the subscript f_{2L} to emphasize that these terms *only* impact the lower component of the left-handed doublet because of the 0 in the upper component of the Higgs doublet. Leaving this aside for a second, we note that the first term has the form of the desired mass term above (identifying f_{2L} to f_L) while the second term describes the coupling of the fermion to the physical Higgs field. The corresponding Yukawa coupling may be chosen to be consistent with the observed fermion mass, namely

$$g_f = \sqrt{2} \frac{m_f}{v} \quad (1.88)$$

such that

$$\mathcal{L}_f = -m_f \bar{f}f - \frac{m_f}{v} \bar{f}fh. \quad (1.89)$$

⁷⁹³ Notably here, the fermion coupling to the Higgs boson scales with the mass of the fermion, a
⁷⁹⁴ fact that is extremely relevant for this thesis analysis.

As we said above, these terms *only* impact the lower component of the left-handed doublet. The inclusion of terms for the upper component is accomplished via the introduction of a Higgs conjugate doublet, defined as

$$\phi_c = -i\sigma_2\phi^* = \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix}. \quad (1.90)$$

⁷⁹⁵ The argument proceeds similarly to the above, with similar results for couplings and masses
⁷⁹⁶ of upper components.

⁷⁹⁷ 1.7 The Standard Model: A Summary

After all of the above, we may write the Standard Model as a theory with a local $SU(3) \times SU(2)_L \times U(1)_Y$ gauge symmetry, described by the Lagrangian

$$\mathcal{L} = \sum_f \bar{f}i\gamma^\mu D_\mu f - \frac{1}{4} \sum_{gauges} F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger(D^\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \quad (1.91)$$

where $D_\mu = \partial_\mu + ig_W W_\mu^k t^k + ig' \frac{Y}{2} B_\mu + ig_S G_\mu^a t^a$, in addition to the Yukawa terms, which we write generally as

$$\mathcal{L}_{Yukawa} = - \sum_{f,\phi=\phi,-\phi_c} y_f (\bar{f}\phi f + (\bar{f}\phi f)^\dagger) \quad (1.92)$$

798 with the sum running over running over appropriate chiral fermion and Higgs doublets.

799 The $SU(2)_L \times U(1)_Y$ subgroup is spontaneously broken to a $U(1)$ symmetry, lending mass
800 to the associated gauge bosons and fermions. Of relevance for this thesis is the resulting
801 physical Higgs field, with a predicted trilinear self-interaction and associated coupling λv ,
802 related to the experimentally observed Higgs boson mass by $m_H = \sqrt{2\lambda v^2}$, as well as the fact
803 that the strength of the Higgs coupling to fermions scales proportionally with the fermion
804 mass.

805 The Standard Model has been monumentally successful, with predictions consistent across
806 many varied experimental cross-checks. This thesis participates in one such cross check.
807 However, the Standard Model is notably not a complete theory of the universe – there is
808 no inclusion of gravity, for instance, though a consistent description may be provided with
809 the introduction of a spin-2 particle. Neutrino oscillations demonstrate that neutrinos have
810 mass, but right-handed neutrinos have not been observed, leading to questions about whether
811 there is a different mechanism to provide neutrinos with mass than that described above.
812 Cosmology tells us that dark matter exists, but there is no corresponding particle within the
813 Standard Model. This thesis therefore also participates in searches for physics beyond the
814 Standard Model. We will provide a sketch of the relevant theories in the following chapter,
815 though a detailed theoretical discussion is beyond the scope of this work.

816

Chapter 2

817

DI-HIGGS PHENOMENOLOGY AND PHYSICS BEYOND THE STANDARD MODEL

818

819 This thesis focuses on searches for di-Higgs production in the $b\bar{b}b\bar{b}$ final state. In this
 820 chapter, we will provide a brief overview of the practical theoretical information motivating
 821 such searches. Though the searches test for physics beyond the Standard Model, particularly
 822 in the search for resonances, the goal of the experimental results is to be somewhat agnostic
 823 to particular theoretical frameworks. An in depth treatment of such models is therefore
 824 beyond the scope of this thesis, though we will attempt to provide a grounding for the models
 825 that we consider.

826 **2.1 Intro to Di-Higgs**

827 Di-Higgs searches can be split into two major theoretical categories: *resonant searches*, in
 828 which a physical resonance is produced that subsequently decays into two Higgs bosons, and
 829 a *non-resonant searches* in which no physical resonance is produced, but where the HH
 830 production cross section has a contribution from an exchange of a *virtual* or *off-shell* particle.

831 The focus of this thesis is gluon initiated processes – in the case of di-Higgs this is termed
 832 gluon-gluon fusion (ggF). HH production may also occur via vector boson fusion [17]. However
 833 the cross section for such production is significantly smaller. Representative Feynman diagrams
 834 are shown for gluon-gluon fusion resonant production in Figure 2.1 and for non-resonant
 835 production in Figure 2.2.

836 As shown in Chapter 1, the Higgs coupling to fermions scales with particle mass. As the
 837 top quark has a mass of 173 GeV, whereas the H has a mass of 125 GeV, such that $H \rightarrow t\bar{t}$ is
 838 kinematically disfavored, $H \rightarrow b\bar{b}$ is the dominant fermionic Higgs decay mode, and, in fact,

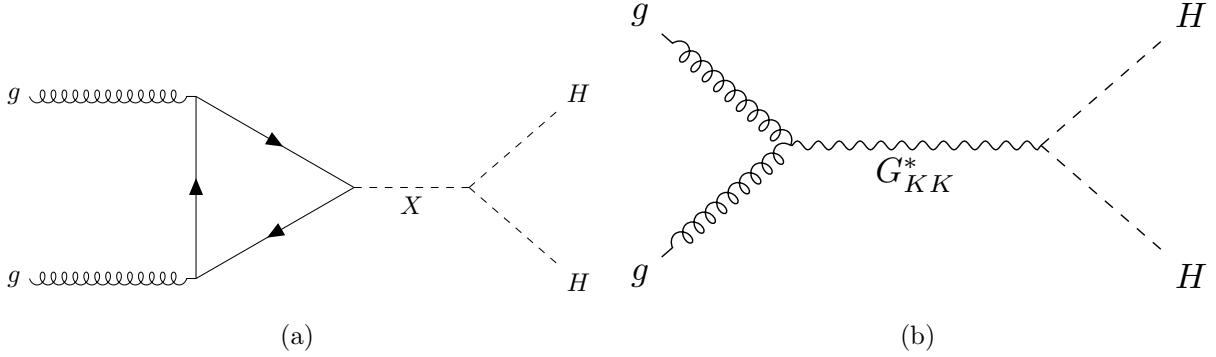


Figure 2.1: Representative diagrams for the gluon-gluon fusion production of spin-0 (X) and spin-2 (G_{KK}^*) resonances which decay to two Standard Model Higgs bosons. The spin-0 resonance considered for this thesis is a generic narrow width resonance which may be interpreted in the context of two Higgs doublet models [18], whereas the spin-2 resonance is considered as a Kaluza-Klein graviton within the bulk Randall-Sundrum (RS) model [19, 20].

the dominant overall decay mode, with a branching fraction of around 58 %. The dominant top quark Yukawa coupling to the H does play a role in H production, however – gluon-gluon fusion is dominated by processes including a top loop.

The single H properties translate to HH production, with $HH \rightarrow b\bar{b}b\bar{b}$ accounting for around 34 % of all HH decays. The H H branching fractions are shown in Figure 2.3.

2.2 Resonant HH Searches

Resonant di-Higgs production is predicted in a variety of extensions to the Standard Model. In particular, this thesis presents searches for both spin-0 and spin-2 resonances. The decay of spin-1 resonances to two identical spin-0 bosons is prohibited, as the final state must correspondingly be symmetric under particle exchange, but this process would require orbital angular momentum $\ell = 1$, and thus an anti-symmetric final state. Each model considered here is implemented in a particular theoretical context, but set up experimental results for generic searches.

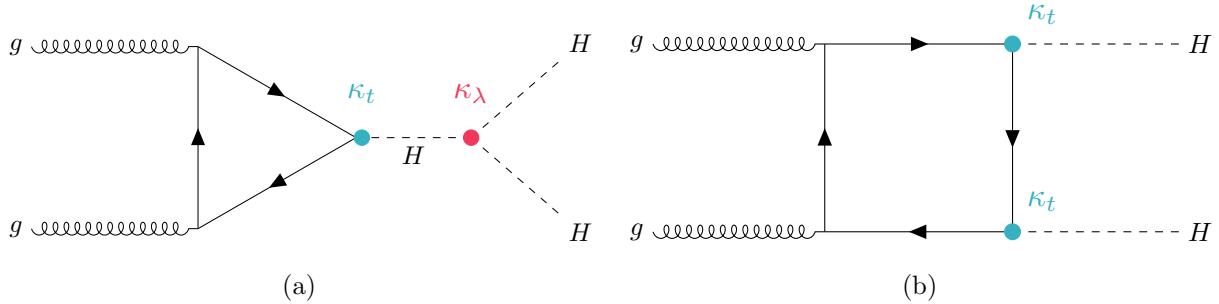


Figure 2.2: Dominant contributing diagrams for non-resonant gluon-gluon fusion production of HH . κ_λ and κ_t represent variations of the Higgs self-coupling and coupling to top quarks respectively, relative to that predicted by the Standard Model.

The spin-2 signal considered is implemented within the bulk Randall-Sundrum (RS) model [19, 20], which features spin-2 Kaluza-Klein gravitons, G_{KK}^* , that are produced via gluon-fusion and which may decay to a pair of Higgs bosons. The model predicts such gravitons as a consequence of warped extra dimensions, and is correspondingly parametrized by a value $c = k/\overline{M}_{\text{Pl}} = 1$, where k describes a curvature scale for the extra dimension and \overline{M}_{Pl} is the Planck mass. The model considered here has $c = 1.0$. However, this model was considered in the early Run 2 HH analyses [21], and was excluded across much of the relevant mass range.

The primary theoretical focus of this work is therefore the spin-0 result, which is implemented as a generic resonance with width below detector resolution. Scalar resonances are interesting, for instance, in the context of two Higgs doublet models [18], which posit the existence of a second Higgs doublet. This leads to the existence of five scalar particles in the Higgs sector – roughly, two complex doublets provide eight degrees of freedom, three of which are “eaten” by the electroweak bosons, leaving five degrees of freedom which may correspond to physical fields.

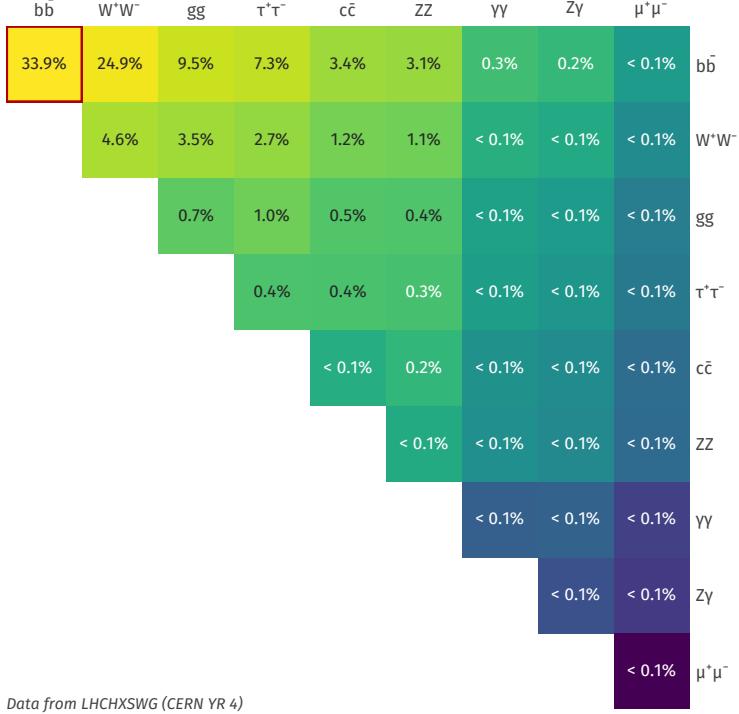


Figure 2.3: Illustration of dominant HH branching ratios. $HH \rightarrow b\bar{b}b\bar{b}$ is the most common decay mode, representing 34 % of all HH events produced at the LHC.

867 2.3 Non-resonant HH Searches

Non-resonant HH production is predicted by the Standard Model via the trilinear coupling discussed above, as well as via production in a fermion loop. More explicitly, after electroweak symmetry breaking, we have

$$\mathcal{L}_{SM} \supset -\lambda v^2 h^2 - \lambda v h^3 - \frac{1}{4} \lambda h^4 \quad (2.1)$$

$$= -\frac{1}{2} m_H^2 - \lambda_{HHH}^{SM} v h^3 - \lambda_{HHHH}^{SM} h^4 \quad (2.2)$$

where $m_H = \sqrt{2\lambda v^2}$ so that

$$\lambda_{HHH}^{SM} = \frac{m_H^2}{2v^2}. \quad (2.3)$$

868 The mass of the SM Higgs boson has been experimentally measured to be 125 GeV [22],
 869 and the vacuum expectation value $v = 246$ GeV has a precise determination from the muon
 870 lifetime [23]. This coupling is therefore precisely predicted in the Standard Model, such that
 871 an observed deviation from this prediction would be a clear sign of new physics.

872 The relevant diagrams for non-resonant HH production are shown in Figure 2.2. Notably,
 873 the diagrams *interfere* with each other, which can be easily seen by counting the fermion
 874 lines. A detailed theoretical discussion is provided by, e.g. [24].

For the searches presented here, the quark couplings to the Higgs are considered to be consistent with the Standard Model value, with measurements of the dominant top Yukawa coupling left to more sensitive direct measurements, e.g. from $t\bar{t}$ final states [25]. Variations of the trilinear coupling away from the Standard Model are considered, however. Such variations are parametrized via

$$\kappa_\lambda = \frac{\lambda_{HHH}}{\lambda_{HHH}^{SM}} \quad (2.4)$$

875 where λ_{HHH} is a varied coupling and λ_{HHH}^{SM} is the Standard Model prediction. As this
 876 variation only comes as a prefactor only with the *triangle* diagram, significant and interesting
 877 effects are observed due to the interference. Examples of the impact of this tradeoff on the
 878 di-Higgs invariant mass are shown in Figure 2.4. Generally speaking, the triangle diagram
 879 contributes more at low mass, while the box diagram contributes more at high mass.

From a quick analysis of Figure 2.2, one may see that, at leading order, the box diagram, B has amplitude proportional to κ_t^2 , defined as the ratio of the top Yukawa coupling to the value predicted by the Standard Model, whereas the triangle diagram, T has amplitude proportional to $\kappa_t \kappa_\lambda$. Therefore, the cross section is proportional to

$$\sigma(\kappa_t, \kappa_\lambda) = |A(\kappa_t, \kappa_\lambda)|^2 \quad (2.5)$$

$$\sim |\kappa_t^2 B + \kappa_t \kappa_\lambda T|^2 \quad (2.6)$$

$$= \kappa_t^4 |B|^2 + \kappa_t^3 \kappa_\lambda (BT + TB) + \kappa_t^2 \kappa_\lambda^2 |T|^2, \quad (2.7)$$

880 and thus non-resonant HH production cross section may be parametrized as a second order
 881 polynomial in κ_{lambda} .

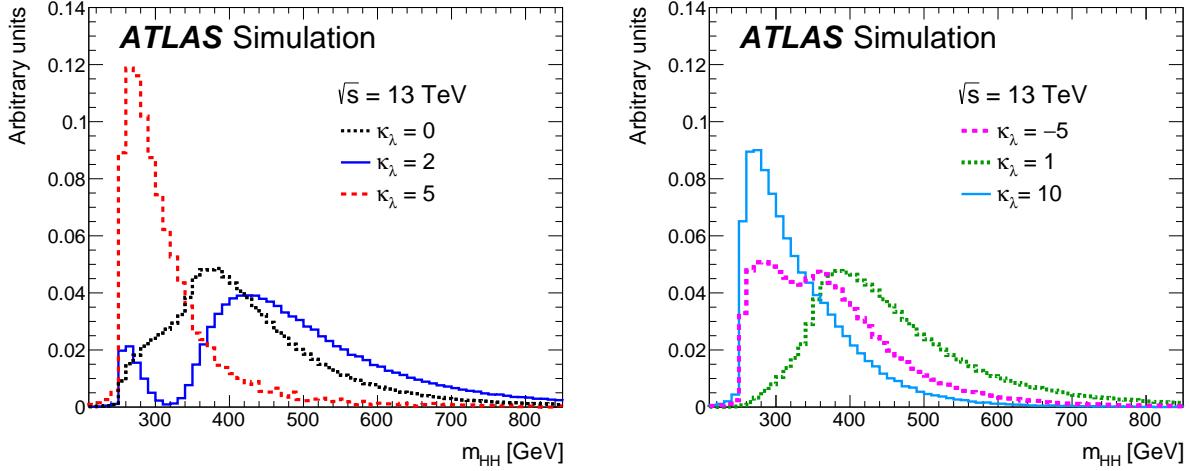


Figure 2.4: Monte Carlo generator level m_{HH} distributions for various values of κ_λ , demonstrating the impact of the interference between the two diagrams of Figure 2.2 on the resulting m_{HH} distribution. For $\kappa_\lambda = 0$ there is no triangle diagram contribution, demonstrating the shape of the box diagram contribution, whereas for $\kappa_\lambda = 10$, the triangle diagram dominates, with a strong low mass peak. The interplay between the two is quite evident for other values, resulting in, e.g., the double peaked structure present for $\kappa_\lambda = 2$ (near maximal destructive interference) and $\kappa_\lambda = -5$. [21]

For positive values of κ_λ , due to the relative minus sign between the triangle and box diagrams, the interference between the two diagrams is *destructive*, with a maximum interference near $\kappa_\lambda = 2.3$, corresponding to the minimum cross section prediction. One may note that the Standard Model value of $\kappa_\lambda = 1$ is not far away from this minimum – correspondingly the Standard Model cross section for HH production is quite small, namely 31.05 fb at $\sqrt{s} = 13 \text{ TeV}$ for production via gluon-gluon fusion [26–33] compared to, e.g. single Higgs production, with a gluon-gluon fusion production cross section of 46.86 pb at $\sqrt{s} = 13 \text{ TeV}$ [34] roughly 1500 times larger! For negative values of κ_λ , the interference is constructive.

ATLAS projections [35] of $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau^+\tau^-$ predict an expected signal strength for Standard Model HH of 3.5σ with no systematic uncertainties and 3.0σ with systematic uncertainties using the 3000 fb^{-1} of data from the HL-LHC (around $20\times$ the full Run 2 dataset considered in this thesis), constituting an *observation* of HH . As the cross section for Standard Model HHH production, corresponding to the quartic Higgs interaction, is much smaller (around 0.1 fb at $\sqrt{s} = 14\text{ TeV}$ [36]), observation of triple Higgs production is even farther in the future, and so is not considered here. However this may be interesting for future work in a variety of Beyond the Standard Model scenarios (e.g. [37–39]).

899

Chapter 3

900

EXPERIMENTAL APPARATUS

901 What machines must we build to examine the smallest pieces of the universe? The famous
 902 equation $E = m$ provides that to create massive particles, we need to provide enough energy.
 903 In order to give kinematic phase space to the types of processes that are examined in this
 904 thesis (and many others besides), a system must be created in which there is enough energy
 905 to (at bare minimum), overcome kinematic thresholds: if you want to search for HH decays,
 906 you should have at least 250 GeV ($= 2 \times m_H$) to work with. It is not enough to simply induce
 907 such processes, however. These processes need to be captured in some way, emitted energy
 908 and particles must be characterized and identified, and in the end all of this information must
 909 be put into a useful and useable form such that selections can be made, statistics can be run,
 910 and a meaningful statement can be made about the universe. In this chapter, we describe the
 911 machines behind the physics, namely the Large Hadron Collider and the ATLAS experiment.

912 **3.1 The Large Hadron Collider**

913 The Large Hadron Collider is a particle accelerator near Geneva, Switzerland. In broad scope,
 914 it is a ring with a 27 kilometer circumference. Hadrons (usually protons or heavy ions) move
 915 in two counter-circulating beams, which are made to collide at four collision points at various
 916 points on the ring. These four collision points correspond to the four detectors placed around
 917 the ring: two “general purpose” experiments: ATLAS and CMS; LHCb, focused primarily on
 918 flavor physics; and ALICE, focused primarily on heavy ions.

919 The focus of this thesis is proton-proton collisions at center of mass energy $\sqrt{s} = 13$ TeV.
 920 The process to achieve such collisions proceeds as follows: first, an electric field strips hydrogen
 921 of its electrons, creating protons. A linear accelerator, LINAC 2, accelerates protons to

922 50 MeV. The resulting beam is injected into the Proton Synchrotron Booster (PSB), which
 923 pushes the protons to 1.4 GeV, and then the Proton Synchrotron, which brings the beam to
 924 25 GeV.

925 Protons are then transferred to the Super Proton Synchrotron (SPS), which ramps up
 926 the energy to 450 GeV. Finally, the protons enter the LHC itself, bringing the beam up to
 927 6.5 TeV. *TODO: cite: <https://home.cern/science/accelerators/accelerator-complex>*

928 While there is, of course, much that goes into the Large Hadron Collider development and
 929 operation, perhaps two of the most fundamental ideas are (1) how are the beams directed
 930 and manipulated and (2) what do we mean when we say “protons are accelerated”. These
 931 questions both are directly answered by pieces of hardware, namely (1) magnets and (2)
 932 radiofrequency (RF) cavities.

933 One of fundamental components of the LHC is a large set of superconducting niobium-
 934 titanium magnets. These are cooled by liquid helium to achieve superconducting temperatures,
 935 and there are several types with very specific purposes. The obvious first question with a
 936 circular accelerator is how to keep the particle beam moving around in that circle. This job
 937 is done via a set of dipole magnets placed around the *beam pipes*: the tubes containing the
 938 beam. These are designed such that the magnetic field in the center of the beam pipe runs
 939 perpendicular to the velocity of the charged particles, providing the necessary centripetal
 940 force for the synchrotron motion.

941 A proton beam is not made of a single proton, however, but of many protons, grouped
 942 into a series of *bunches*. As all of these are positively charged, if unchecked, these bunches
 943 would become diffuse and break apart. What we want is a stable beam with tightly clustered
 944 protons to maximize the chance of a high energy collision. Such clustering is done via a series
 945 of quadropole magnets, with field distributed as in *TODO: grab image from General Exam.*
 946 Alternating sets of quadropoles provide the necessary forces for a tight, stable beam. While
 947 these are the two major components of the LHC magnet system, it is not the full story –
 948 higher order magnets are used to correct for small imperfections in the beam *TODO: expand.*

949 Magnetic fields do no work, however, so the magnet system is unable to do the job of the

actual acceleration. This is accomplished via a set of radiofrequency (RF) cavities. Within these cavities, an electric field is made to oscillate (switch direction) at a precise rate. These rates interact with the beam via in RF *buckets*, with bunches corresponding to groups of protons that fill a given bucket. The timing is such that protons will always experience an accelerating voltage, corresponding to the 25 ns bunch spacing used at the LHC.

A nice property of this bucket/bunch configuration is that there is some self-correction – there is some finite spread in the grouping of particles. If a particle arrives too early, it will experience some decelerating voltage; if too late, it will experience a higher accelerating voltage.

3.1.1 The LHC Schedule

The physics program at the Large Hadron Collider is split into a variety of data taking periods called *runs*. These runs correspond to various detector/accelerator configurations, and are interspersed with *long shutdowns* – periods used for detector/accelerator upgrades in preparation for the next run. The LHC timeline is as follows

1. Run 1 (2010–2013): First run of the LHC, operating at center of mass energy $\sqrt{s} = 7 \text{ TeV}$, increased to 8 TeV in 2012. ATLAS recorded 4.57 fb^{-1} and 20.3 fb^{-1} of data usable for physics at $\sqrt{s} = 7 \text{ TeV}$ and 8 TeV respectively.
2. Long Shutdown 1 (LS1; 2013–2015): Upgrades to accelerator complex, magnet system, to allow for increase in energy. Design energy was $\sqrt{s} = 14 \text{ TeV}$, delays in “training” of superconducting magnets led to decrease to $\sqrt{s} = 13 \text{ TeV}$.
3. Run 2 (2015–2018): Second run of the LHC, operating at center of mass energy $\sqrt{s} = 13 \text{ TeV}$. Data from this run is used in this thesis, with 139 fb^{-1} of data available for physics from the ATLAS experiment.
4. Long Shutdown 2 (LS2; 2019–2021): Upgrades to ATLAS muon spectrometer (New

974 Small Wheel), liquid argon calorimeter; upgrades in preparation for the High Luminosity
 975 LHC (HL-LHC).

976 5. Run 3 (2021–2023?): Third run of the LHC, target center of mass energy $\sqrt{s} =$
 977 $13 - 14 \text{ TeV}$, total target luminosity 300 fb^{-1} .

978 6. Long Shutdown 3 (LS3; 2024?–2026?): Further upgrades for the HL-LHC.

979 7. Run 4, 5, ...? (2026? onward): High Luminosity LHC – goal is to achieve instantaneous
 980 luminosities by a factor of five, massively enlarging available statistics for physics.
 981 Projected 3000 to 4000 fb^{-1} , > 20 times the full Run 2 ATLAS dataset.

982 3.2 The ATLAS Experiment

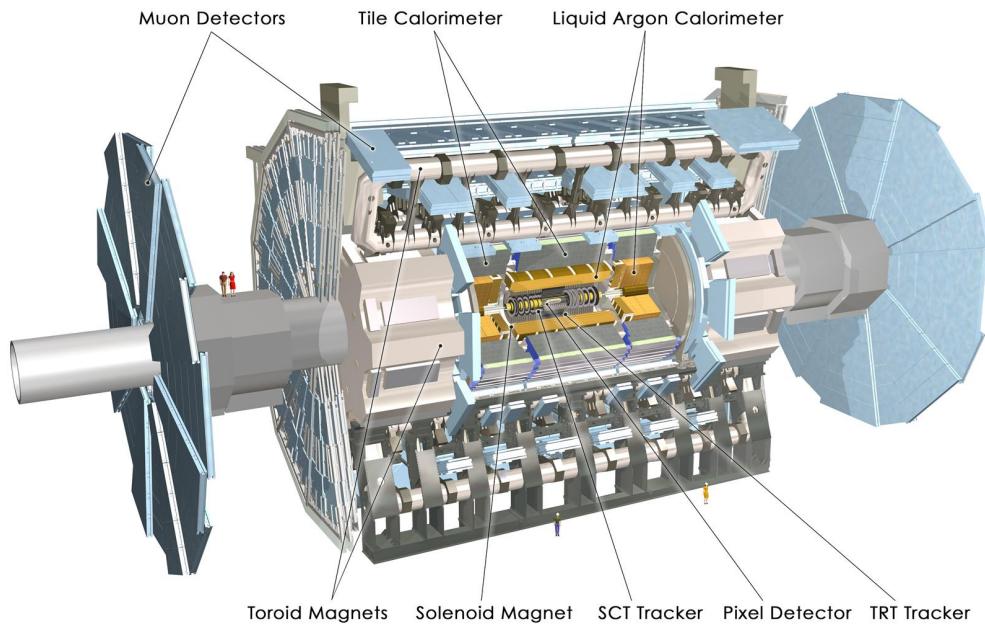


Figure 3.1: Diagram of the ATLAS detector [40]

983 This thesis focuses on searches done with the ATLAS experiment. As mentioned, this is one

984 of two “general purpose” experiments at the LHC, by which we mean there is a very large and
 985 broad variety of physics done within the experimental collaboration. This broad physics focus
 986 has a direct relation to the design of the ATLAS detector [41], pictured in Figure 3.1, which
 987 is composed of a sophisticated set of subsystems designed to fully characterize the physics of
 988 a given high energy particle collision. It consists of an inner tracking detector surrounded
 989 by a thin superconducting solenoid, electromagnetic and hadronic calorimeters, and a muon
 990 spectrometer incorporating three large superconducting toroidal magnets. The ATLAS
 991 detector covers nearly the entire solid angle around the collision point, fully characterizing
 992 the “visible” components of a collision and allowing for indirect sensitivity to particles that
 993 do not interact with the detector (e.g. neutrinos) via “missing” energy (roughly momentum
 994 balance). We will go through the design and physics contribution of each of the detector
 995 components in the following. A schematic of how various particles interact with the detector
 996 is shown in Figure 3.2.

997 3.2.1 ATLAS Coordinate System

998 Of relevance for the following discussion, as well as for the analysis presented in Chapter 7,
 999 is the ATLAS coordinate system. ATLAS uses a right-handed coordinate system with its
 1000 origin at the nominal interaction point (IP) in the center of the detector and the z -axis along
 1001 the beam pipe. The x -axis points from the IP to the centre of the LHC ring, and the y -axis
 1002 points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the
 1003 azimuthal angle around the z -axis. The pseudorapidity is defined in terms of the polar angle
 1004 θ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.
 1005 These coordinates are shown in Figure 3.3.

1006 3.2.2 Inner Detector

1007 The purpose of the inner detector is the reconstruction of the trajectory of charged particles,
 1008 called *tracking*. This is accomplished primarily through the collection of electrons displaced
 1009 when a charged particle passes through a tracking detector. By setting up multiple layers of

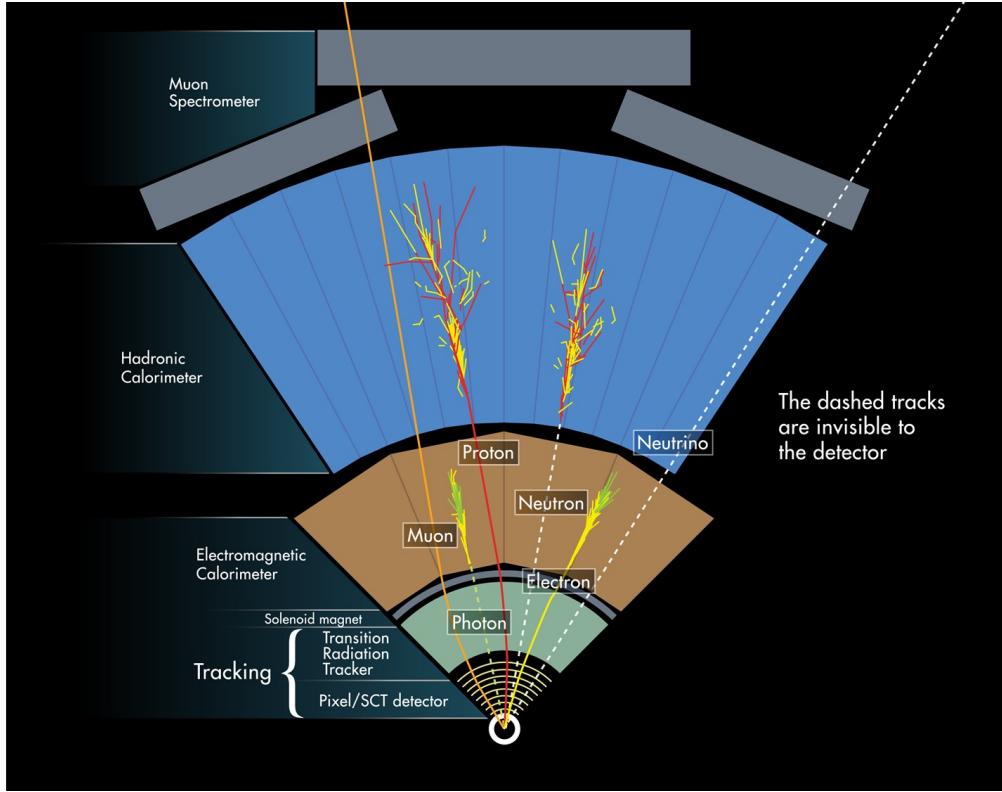


Figure 3.2: Cross section of the ATLAS detector showing how particles interact with various detector components [42]

such detectors, such that a given particle leaves a signature, known as a “hit”, in each layer, the trajectory of the particle may be inferred via “connecting the dots” between these hits.

The raw trajectory of a particle only provides positional information. However, the trajectory of a charged particle in a known magnetic field additionally provides information on particle momentum and charge via the curvature of the corresponding track (cf. $\vec{F} = q\vec{v} \times \vec{B}$). The inner detector system is therefore surrounded by a solenoid magnet, providing a 2 T magnetic field along the z -axis (yielding curvature in the transverse $x - y$ plane).

The inner detector provides charged particle tracking in the range $|\eta| < 2.5$ via a series of detector layers. The innermost of these is the high-granularity silicon pixel detector which typically provides four measurements per track, with the first hit in the insertable B-layer

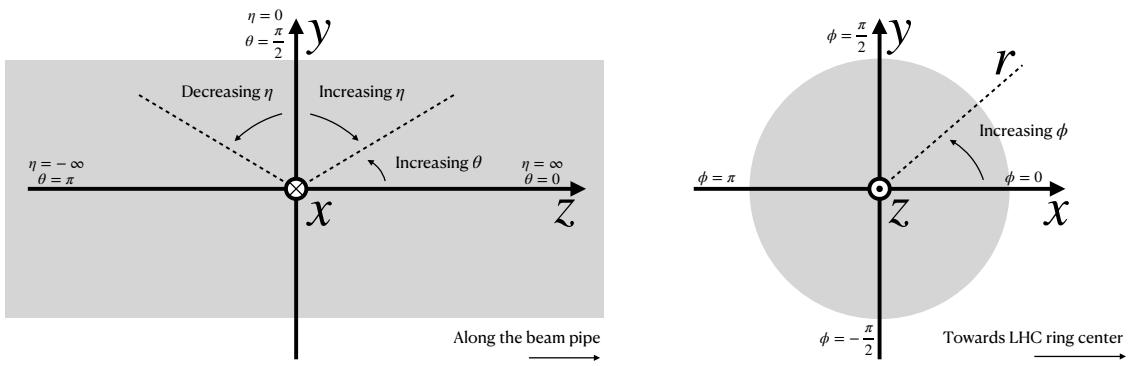


Figure 3.3: 2D projections of the ATLAS coordinate system

1020 (IBL) installed before Run 2 [43, 44]. This is very close to the interaction point with a
 1021 high degree of positional information, and is therefore very important for e.g. b -tagging (see
 1022 Chapter 5). It is followed by the silicon microstrip tracker (SCT), which usually provides
 1023 eight measurements per track. This is lower granularity, but similar in concept to the pixel
 1024 detector.

1025 Both of these silicon detectors are complemented by the transition radiation tracker
 1026 (TRT), which extends the radial track reconstruction within the range $|\eta| < 2.0$. This is
 1027 a different design, composed of *drift tubes*, i.e. straws filled with Xenon gas with a wire
 1028 in the center, but similarly collects electrons displaced by ionizing particles. In addition,
 1029 the TRT includes materials with widely varying indices of refraction, which leads to the
 1030 production of transition radiation, namely radiation produced by a charged particle passing
 1031 through an inhomogeneous medium. The energy loss on such a transition is proportional
 1032 to the Lorentz factor $\gamma = E/m$ – correspondingly, lighter particles (e.g. electrons) tend to
 1033 lose more energy and emit more photons compared to heavier particles (e.g. pions). In the
 1034 detector, this corresponds to a larger fraction of hits (typically 30 in total) above a given

1035 high energy-deposit threshold for electrons, providing particle identification information.

1036 *3.2.3 Calorimeter*

1037 Surrounding the inner detector in ATLAS is the calorimeter. The principle of the calorimeter
1038 is to completely absorb the energy of a produced particle in order to measure it. However,
1039 a pure block of absorber does not provide much information about the particle interaction
1040 with the material. The ATLAS calorimeter therefore has a *sampling calorimeter* structure,
1041 namely, layers of absorber interspersed with layers of sensitive material, giving the calorimeter
1042 “stopping power” while allowing detailed measurement of the resulting particle shower and
1043 corresponding deposited energy.

1044 The ATLAS calorimetersystem covers the pseudorapidity range $|\eta| < 4.9$, and is primarily
1045 composed of two components, an electromagnetic calorimeter, designed to measure particles
1046 which primarily interact via electromagnetism (e.g. photons and electrons), and a hadronic
1047 calorimeter, designed to measure particles which interact via the strong force (e.g. pions,
1048 other hadrons). We will return to the differences between these in a moment.

1049 In ATLAS, the electromagnetic calorimeter covers the region of $|\eta| < 3.2$, and uses
1050 lead for the absorbers and liquid-argon for the sensitive material. It is high granularity
1051 and, geometrically, has two components: the “barrel”, which covers the cylindrical body of
1052 the detector volume and the “endcap”, covering the ends. An additional thin liquid-argon
1053 presampler covers $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters.

1054 The hadronic calorimeter is composed of alternating steel and plastic scintillator tiles,
1055 segmented into three barrel structures within $|\eta| < 1.7$, in addition to two copper/liquid-argon
1056 endcap calorimeters.

1057 The solid angle coverage is completed with forward copper/liquid-argon and tungsten/liquid-
1058 argon calorimeter modules optimized for electromagnetic and hadronic energy measurements
1059 respectively.

1060 3.2.4 Muon Spectrometer

1061 While muons interact electromagnetically, they are around 200 times heavier than electrons
 1062 ($m_\mu = 106 \text{ MeV}$, while $m_e = 0.510 \text{ MeV}$). Therefore, electromagnetic interactions with ab-
 1063 sorbers in the calorimeter are not sufficient to stop them, and, as they do not interact via the
 1064 strong force, hard scattering with nuclei is rare. A dedicated system for muon measurements
 1065 is therefore required.

1066 The muon spectrometer (MS) is the outermost layer of ATLAS and is designed for this
 1067 purpose. It is composed of three parts: a set of triggering chambers, which detect if there is
 1068 a muon and provide a coordinate measurement, in conjunction with high-precision tracking
 1069 chambers, which measure the deflection of muons in a magnetic field to measure muon
 1070 momentum, similar to the inner detector solenoid. The magnetic field is generated by the
 1071 superconducting air-core toroidal magnets, with a field integral between 2.0 and 6.0 T m
 1072 across most of the detector. The toroid magnetic field runs roughly in a circle in the $x - y$
 1073 plane around the beam line, leading to muon curvature along the z-axis.

1074 The precision tracking system covers the region $|\eta| < 2.7$ via three layers of monitored
 1075 drift tubes, and is complemented by cathode-strip chambers in the forward region, where the
 1076 background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate
 1077 chambers in the barrel, and thin-gap chambers in the endcap regions.

1078 3.2.5 Triggering

1079 During a typical run of the LHC, there are roughly 1 billion collisions in ATLAS per second
 1080 (1 GHz), corresponding to a 40 MHz bunch crossing rate. *TODO: cite: <https://cds.cern.ch/record/1457044/file>*
 1081 Saving the information from all of them is not only unnecessary, but infeasible. The ATLAS
 1082 trigger system provides a sophisticated set of selections to filter the collision data and only
 1083 keep those collision events useful for downstream analysis.

1084 These events are selected by the first-level trigger system, which is implemented in custom
 1085 hardware, and accepts events at a rate below 100 kHz. Selections are then made by algorithms

1086 implemented in software in the high-level trigger [45], reducing this further, and, in the end,
1087 events are recorded to disk at much more manageable rate of about 1 kHz.

1088 An extensive set of ATLAS software [46] is open source, including the software used for
1089 real and simulated data reconstruction and analysis and that used in the trigger and data
1090 acquisition systems of the experiment.

1091 3.2.6 Particle Showers and the Calorimeter

1092 The design of the ATLAS detector is directly tied to the physics it is trying to detect. Of these,
1093 possibly the most non-trivial distinction is in the calorimeter design. It is therefore useful to
1094 discuss in more detail the various properties of electromagnetic and hadronic interactions
1095 with material, and how these correspond to the particle showers measured by the detector
1096 described above.

1097 Electromagnetic showers in ATLAS predominantly occur via bremsstrahlung, or “braking
1098 radiation”, and electron-positron pair production. This proceeds roughly as follows: an electron
1099 entering a material is deflected by the electromagnetic field of a heavy nucleus. This results in
1100 the radiation of a photon. That photon produces an electron-positron pair, and the process
1101 repeats, resulting in a shower structure. At each step, characterized by *radiation length*, X_0 ,
1102 the number of particles approximately doubles and the average particle energy decreases by
1103 approximately a factor of two. *TODO: Include nice Thomson image*

Note that bremsstrahlung and pair production only dominate in specific energy regimes, with other processes taking over depending on particle energy. For electrons, bremsstrahlung only dominates for higher energies, as low energy electrons will form ions with the atoms of the material. The point where the rates for the two processes are equal is called the *critical energy*, and is roughly

$$E_c \approx \frac{800 \text{ MeV}}{Z} \quad (3.1)$$

1104 where Z is the nuclear charge. From a similar analysis of rates, we may see that the
1105 bremsstrahlung rate is inversely proportional to the square of the mass of the particle. This

₁₁₀₆ explains why muons do not shower in a similar way, as the rate of bremsstrahlung is suppressed
₁₁₀₇ by $(m_e/m_\mu)^2$ relative to electrons.

For lead, the absorber used for the ATLAS electromagnetic calorimeter, which has $Z = 82$, this critical energy is therefore around 10 MeV. Electrons resulting from LHC collisions are of a 1.3×10^3 GeV scale. With the approximation of a reduction in particle energy by a factor of two every radiation length, the number of radiation lengths before the critical energy is reached is

$$x = \frac{\ln(E/E_c)}{\ln 2} \quad (3.2)$$

₁₁₀₈ such that for a 100 GeV shower in lead, $x \sim 13$. The radiation length for lead is around
₁₁₀₉ 0.56 cm, such that an electromagnetic shower could be expected to be captured within 10 cm
₁₁₁₀ of lead.

₁₁₁₁ Electromagnetic showers are therefore characterized by depositing much of their energy
₁₁₁₂ within a small region of space. As we show below (Chapter 4) though electromagnetic
₁₁₁₃ showering is not deterministic, the large number of particles and the restricted set of processes
₁₁₁₄ involved means that the shower development as a whole is very similar between individual
₁₁₁₅ electromagnetic showers of the same energy.

₁₁₁₆ For completeness, note as well that pair production dominates for photons of energy greater
₁₁₁₇ than around 10 MeV, whereas for lower energies (below around 1 MeV), the photoelectric
₁₁₁₈ effect, namely atomic photon absorption and electron emission, dominates.

₁₁₁₉ Hadronic showers are distinguished by the fact that they interact strongly with atomic
₁₁₂₀ nuclei. They are correspondingly more complex because (1) they involve a wider variety
₁₁₂₁ of processes than electromagnetic showers, and (2) these processes have a wide variety of
₁₁₂₂ associated length scales. Because these are heavier than electrons (e.g. protons and charged
₁₁₂₃ pions) bremsstrahlung is suppressed, but ionization interactions with the electrons will cause
₁₁₂₄ these particles to lose energy as they pass through the material. Hadronic showering occurs
₁₁₂₅ on interaction with atomic nuclei. This may lead to production of, e.g. both charged (π^\pm)
₁₁₂₆ and neutral (π^0) pions. The π^0 lifetime is much much shorter than that of the charged pions
₁₁₂₇ (around a factor of 10^8), and immediately decays to two photons, starting an electromagnetic

shower, as described above. The longer lived π^\pm travel further in the detector before experiencing another strong interaction with more particles produced, also with varying lifetimes and decay properties.

It is therefore immediately apparent that hadronic showers are more complex than electromagnetic ones (electromagnetic showers can be a subset of the hadronic!), and therefore much more variable from shower to shower. The length scales involved are also significantly larger due to the reliance on nuclear interactions, characterized by length λ_I , which is around 17 cm for iron (used in the ATLAS hadronic calorimeter). This motivates the calorimeter design, and results in the properties demonstrated in Figure 3.2.

1137

Chapter 4

1138

SIMULATION

1139 Simulated physics samples are a core piece of the physics output of the Large Hadron
 1140 Collider, providing a map from a physics theory into what is observed in our detector. This
 1141 is crucial for searches for new physics, where simulation is necessary to describe what a given
 1142 signal model looks like, but also extremely valuable for describing the physics of the Standard
 1143 Model, providing detailed predictions of background processes for use in everything from
 1144 designing simple cuts to training multivariate discriminators. Broadly, simulation can be split
 1145 into two stages: *event generation*, in which physics theory is used to generate a description of
 1146 particles present after a proton-proton collision, and *detector simulation*, which passes this
 1147 particle description through a simulation of the detector material, providing a view of the
 1148 physics event as it would be seen in ATLAS data. Such simulation is often called Monte Carlo
 1149 in reference to the underlying mathematical framework, which relies on random sampling.

1150 **4.1 Event Generation**

1151 A variety of tools are used to simulate various aspects of event generation. MADGRAPH [47]
 1152 is commonly used for the generation of the “hard scatter” event, i.e., two protons collide
 1153 and some desired physics process happens. In practice, this is not quite as simple as two
 1154 quarks or gluons interacting. Protons are composed of three “valence” quarks with various
 1155 momenta interacting with each other via exchange of gluons, but also a sea of virtual gluons
 1156 which may decay into other quarks. A hard scatter event is therefore characterized by
 1157 the corresponding particle level diagrams, but additionally by a set of *parton distribution*
 1158 *functions* (PDFs), which describe the probability to find constituent quarks or gluons at
 1159 carrying various momenta at a given energy scale (often written Q^2). Such PDFs are measured

1160 experimentally *TODO: cite* and the selection of a “PDF set” and a given physics process
 1161 characterizes the hard scatter. Depending on the model being considered and the particular
 1162 theoretical constraints, processes are often simulated at either leading (LO) or next to leading
 1163 order (NLO), corresponding to the order of the perturbative expansion (i.e. tree level or 1
 1164 loop diagrams). Various additional tools are developed for such NLO calculations, including
 1165 POWHEG Box v2 [48–50], which is used for this thesis.

1166 The hard scatter is not the only component of a given collider event, however. Incoming
 1167 and outgoing particles are themselves very energetic and may radiate particles along their
 1168 trajectory. In particular, gluons, which have a self-interaction term as described in Chapter 1,
 1169 may be radiated, which subsequently themselves radiate gluons or decay to quarks which can
 1170 also radiate gluons, in a whole mess of QCD that both contributes to the particle content
 1171 of a collider event and is not directly described by the hard scatter. This cascade, called
 1172 a *parton shower*, has a dedicated set of simulation tools, commonly HERWIG 7 [51][52] and
 1173 PYTHIA 8 [53], which interface with tools such as MADGRAPH for simulation.

1174 Due to color confinement (Chapter 1), quarks and gluons cannot be observed free particles,
 1175 but rather undergo a process called hadronization, in which they are grouped into colorless
 1176 hadrons (e.g. *mesons*, consisting of one quark and one antiquark). In simulation, this is also
 1177 handled with HERWIG 7 and PYTHIA 8.

1178 The physics of *b*-quarks is quite important for a variety of searches for new physics and
 1179 measurements of the Standard Model, including this thesis work *TODO: ref flavor tagging*
 1180 *sec?*. Correspondingly, the decay of “heavy flavor” particles (e.g. *B* and *D* mesons, containing
 1181 *b* and *c* quarks respectively) has been very well studied, and a dedicated simulation tool,
 1182 EVTGEN [54], is used for such processes.

1183 *TODO: add nice parton shower image*

1184 4.2 Detector Simulation

1185 Event generation provides a full description of the particle content of a given collider event.
 1186 In reality, however, we do not have access to such a description, and must rely on physical

detectors to collect information about said particle content. The design and components of the ATLAS detector are described in Chapter 3. Simulation of this detector quickly becomes complicated – there are a variety of different materials and subdetectors, each with particular configurations and resolutions. Interactions of particles with the detector materials can cause showering, and such showers must be simulated and characterized.

In ATLAS, the GEANT4 [55] simulation toolkit is used for detailed simulation of the ATLAS detector, often referred to as *full simulation*. The method can be thought of as proceeding step by step as a particle moves through the detector, simulating the interaction of the material at each stage, and following each branch of each resulting shower with a similarly detailed step by step simulation.

This type of simulation is very computationally intensive, especially in the calorimeter, which has a high density of material, leading to an extremely large set of material interactions to simulate. There is correspondingly a large effort within ATLAS to develop techniques to decrease the computational load – these techniques will be of increasing importance for Run 3 and the HL-LHC *TODO: include classic budget plot*.

The fast simulation used for this thesis, AtlFast-II [56], is one such technique, which uses a parametrized simulation of the calorimeter, called FastCaloSim, in conjunction with full simulation of the inner detector, to achieve an order of magnitude speed up in simulation time. This parametrized simulation uses a simplified detector geometry, in conjunction with a simulation of particle shower development based on statistical sampling of distributions from fully simulated events, to massively speed up simulation time and computational load.

Such a speed up comes at a bit of a cost in performance. In particular, the modeling of jet substructure (see Chapter 5) historically has been an issue for FastCaloSim. The ATLAS authorship qualification work supporting this thesis is an effort to improve such modeling, and is part of a suite of updates being considered for a new fast simulation targeting Run 3. We briefly describe this work in the following.

1213 **4.3 Correlated Fluctuations in FastCaloSim**

1214 A variety of developments have been made to FastCaloSim, improving on the version used for
1215 AtlFast-II. This new fast calorimeter simulation [57] is largely based on two components: one
1216 which describes the *total energy* deposited in each calorimeter layer as a shower moves from
1217 the interaction point outward, and one which describes the *shape*, i.e., the pattern of energy
1218 deposits, of a shower in each respective calorimeter layer. Both methods are parametrizations
1219 of the full simulation, and therefore are considered to be performing well if they are able
1220 to reproduce corresponding full simulation distributions. Of course, directly sampling from
1221 a library of showers would identically reproduce such distributions – however a statistical
1222 sampling of various shower *properties* provides much more generality in the simulation.

1223 For the simulation of total energy in each given layer, the primary challenge is that such en-
1224 ergy deposits are highly correlated. The new FastCaloSim thus relies on a technique called Prin-
1225 cipal Component Analysis (PCA) TODO: cite <https://root.cern.ch/doc/master/classTPrincipal.html>
1226 to de-correlate the layers, aiding parametrisation.

1227 The PCA chain transforms N energy inputs into N Gaussians and projects these Gaussians
1228 onto the eigenvectors of the corresponding covariance matrix. This results in N de-correlated
1229 components, as the eigenvectors are orthogonal. The component of the PCA decomposition
1230 with the largest corresponding eigenvalue is then used to define bins, in which showers
1231 demonstrate similar patterns of energy deposition across the calorimeter layers. To further
1232 de-correlate the inputs, the PCA chain is repeated on the showers within each such bin. This
1233 full process is reversed for the particle simulation. A full description of the method can be
1234 found in [57].

1235 Modeling of the lateral shower shape makes use of 2D histograms filled with GEANT4
1236 hit energies in each layer and PCA bin. Binned in polar $\alpha - R$ coordinates in a local plane
1237 tangential to the surface of the calorimeter system, these histograms represent the spatial
1238 distribution of energy deposits for a given particle shower. Such histograms are constructed
1239 for a number of Geant4 events, and the histograms for each event are normalized to total

1240 energy deposited in the given layer. The average of these histograms is then taken (what is
 1241 called here the “average shape”).

1242 In simulation, these average shape histograms are used as probability distributions, from
 1243 which a finite number of equal energy hits are drawn. This finite drawing of hits induces
 1244 a statistical fluctuation about the average shape which is tuned to match the expected
 1245 calorimeter sampling uncertainty.

1246 As an example, the intrinsic resolution of the ATLAS Liquid Argon calorimeter has a
 1247 sampling term of $\sigma_{\text{samp}} \approx 10\%/\sqrt{E}$ [58]. The number of hits to be drawn for each layer, $N_{\text{hits}}^{\text{layer}}$,
 1248 is thus taken from a Poisson distribution with mean $1/\sigma_{\text{samp}}^2$, where the energy assigned to
 1249 each hit is then just $E_{\text{hit}} = \frac{E_{\text{layer}}}{N_{\text{hits}}^{\text{layer}}}$. This induces a fluctuation of the order of $10\%/\sqrt{E_{\text{bin}}}$ for
 1250 each bin in the average shape.

1251 Figure 4.1 shows a comparison of energy and weta2 [59], defined as the energy weighted
 1252 lateral width of a shower in the second electromagnetic calorimeter layer, for 16 GeV photons
 1253 simulated with the new FastCaloSim and with full GEANT4 simulation. The agreement is
 1254 quite good, with FastCaloSim matching the Geant4 mean to within 0.3 and 0.03 percent
 respectively. Similar results are seen for other photon energies and η points.

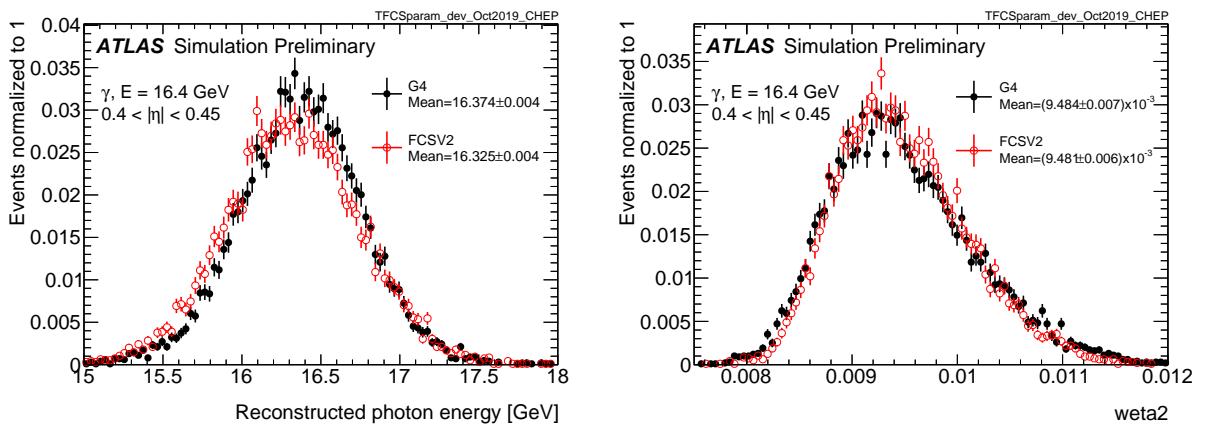


Figure 4.1: Energy and lateral shower width variable, weta2, for 16 GeV photons with full simulation (G4) and FastCaloSimV2 (FCSV2) [57].

1256 *4.3.1 Fluctuation Modeling*

1257 Figure 4.2 shows the ratio of calorimeter cell energies for single GEANT4 photon and pion
 1258 events to the corresponding cell energies in their respective average shapes. While the photon
 1259 event is quite close to the corresponding average, the pion event shows a deviation from the
 1260 average which is much larger and has a non-trivial structure, reflecting the different natures
 1261 of electromagnetic and hadronic showering.

1262 While the shape parametrization described above is thus sufficient for describing electro-
 1263 magnetic showers, we will demonstrate below that it is not sufficient for describing hadronic
 1264 showers (Figures 4.5 and 4.6). We therefore present and validate methods to improve this
 1265 hadronic shower modeling.

1266 Two methods for modeling deviations from the average shape have been studied: (1)
 1267 a neural network based approach using a Variational Autoencoder (VAE) [60] and (2) a
 1268 map through cumulative distributions to an n -dimensional Gaussian. With both methods,
 1269 the shape simulation then proceeds as described in Section ??, with the drawing of hits
 1270 according to the average shape. However, these hits no longer have equal energy, but have
 1271 weights applied to increase or decrease their energy depending on their spatial position.
 1272 This application of weights is designed to mimic a realistic shower structure and to encode
 1273 correlations between energy deposits.

1274 Both methods are trained on ratios of energy in binned units called voxels. This voxelization
 1275 is performed in the same polar $\alpha - R$ coordinates as the average shape, with a 5 mm core in
 1276 R and 20 mm binning thereafter. There are a total of 8 α bins from 0 to 2π and 8 additional
 1277 R bins from 5 mm to 165 mm. The 5 mm core is filled with the average value of core voxels
 1278 across the 8 α bins when creating the parametrisation. However, during simulation, each of
 1279 these 8 core bins is treated independently. The outputs of both methods mimic these energy
 1280 ratios and are used in the shape simulation as the weights described above. In contrast to
 1281 an approach based on, e.g., calorimeter cells, using voxels allows for flexibility in tuning the
 1282 binning used in creating the parametrisation. Further, due to their relatively large size, using

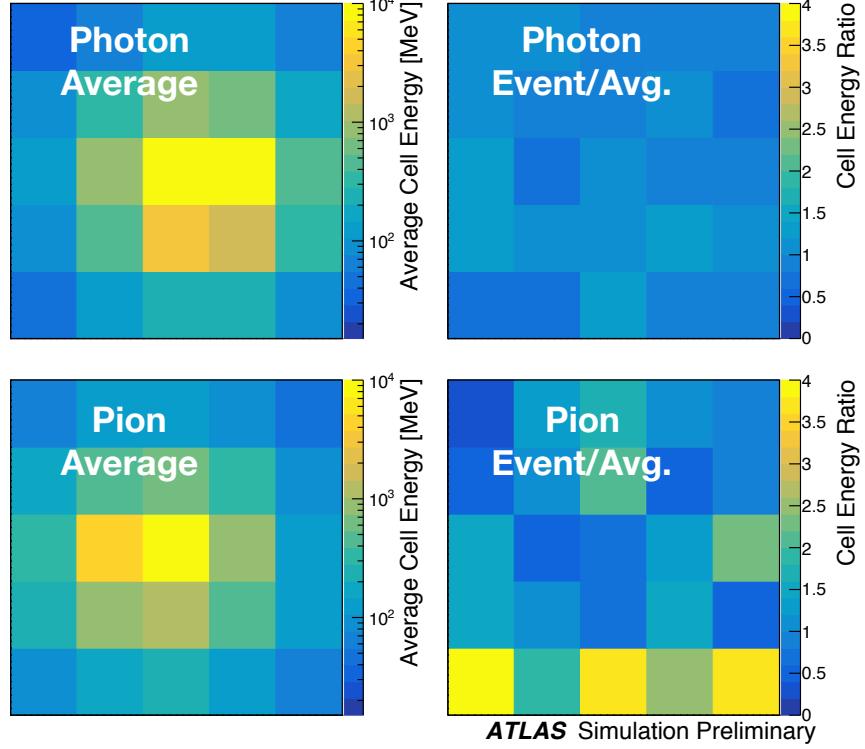


Figure 4.2: Example of photon and pion average shapes in 5×5 calorimeter cells. The left column shows the average shape over a sample of 10000 events, while the right column shows the energy ratio, in each cell, of single GEANT4 events with respect to this average. The photon ratios are all close to 1, while the pion ratios show significant deviation from the average.

1283 calorimeter cells is subject to “edge effects”, where the splitting of energy between cells has a
 1284 non-trivial effect on the observed energy ratio. The binning used here is of the order of half
 1285 of a cell size, mitigating this effect.

1286 The Gaussian method operates by using cumulative distributions to map GEANT4 energy
 1287 ratios to a multidimensional Gaussian distribution. New events are generated by randomly
 1288 sampling from this Gaussian distribution.

1289 For the VAE method, a system of two linked neural networks is trained to generate events.

1290 The first “encoder” neural network maps input GEANT4 energy ratios to a lower dimensional
 1291 latent space. A second “decoder” neural network then samples from that latent space and
 1292 tries to reproduce the inputs. In simulation, events are generated by taking random samples
 1293 from the latent space and passing them through the trained decoder.

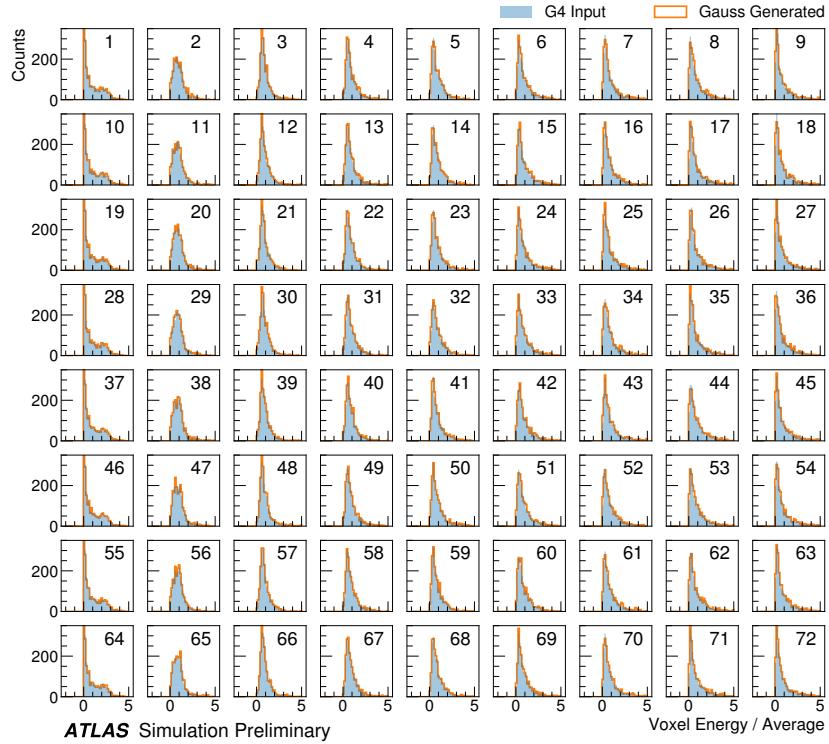


Figure 4.3: Distribution of the ratio of voxel energy in single events to the corresponding voxel energy in the average shape, with GEANT4 events in blue and Gaussian model events in orange, for 65 GeV central pions in EMB2. Moving top to bottom corresponds to increasing α , left to right corresponds to increasing R , with core voxels numbered 1, 10, 19, Agreement is quite good across all voxels. Results are similar for the VAE method.

1294 Figure 4.3 shows the distributions of input GEANT4 and Gaussian method generated
 1295 energy ratios in the grid of voxels. Figure 4.4 shows the correlation coefficient between the
 1296 center voxel from $\alpha = 0$ to $2\pi/8$ for input GEANT4 and the Gaussian and VAE fluctuation

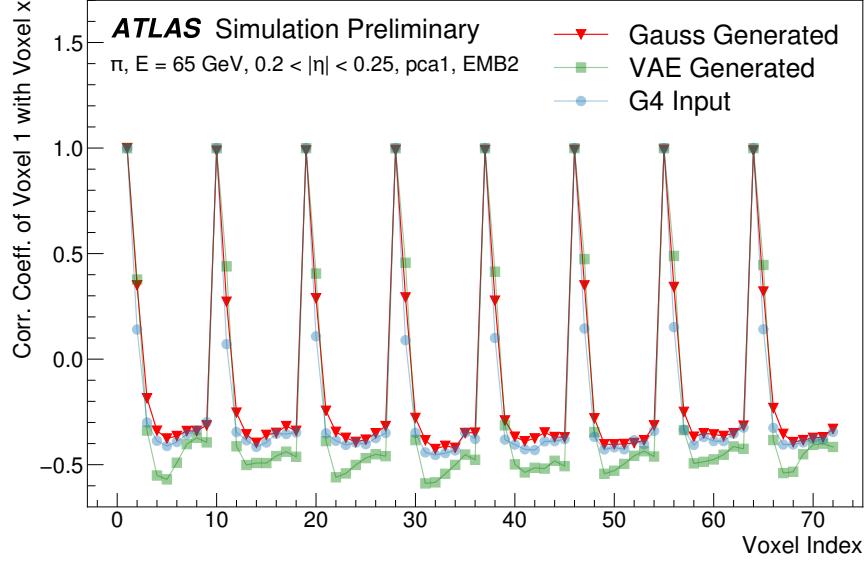


Figure 4.4: Correlation coefficient of ratios of voxel energy in single events to the corresponding voxel energy in the average shape, examined between the core bin from $\alpha = 0$ to $2\pi/8$ and each of the other voxels. The periodic structure represents the binning in α , and the increasing numbers in each of these periods correspond to increasing R , where the eight points with correlation coefficient 1 are the eight core bins. Both the Gaussian and VAE generated toy events are able to reproduce the major correlation structures for 65 GeV central pions in EMB2.

1297 methods. Agreement is good throughout.

1298 Validation of the Gaussian and VAE fluctuation methods was performed within FastCaloSimV2.

1299 Figure 4.5 shows the energy ratio of cells for a given simulation to the corresponding cells in
1300 the average shape as a function of the distance from the shower center. The mean for all
1301 simulation methods is expected to be around 1, with deviation from the average (the RMS
1302 fluctuation) shown by the error bars. The Gaussian method RMS (red) and VAE method
1303 RMS (green) both match the GEANT4 RMS (yellow) better than the case without correlated
1304 fluctuations (blue) for a variety of energies, η points, and layers, often reproducing 80 – 100 %

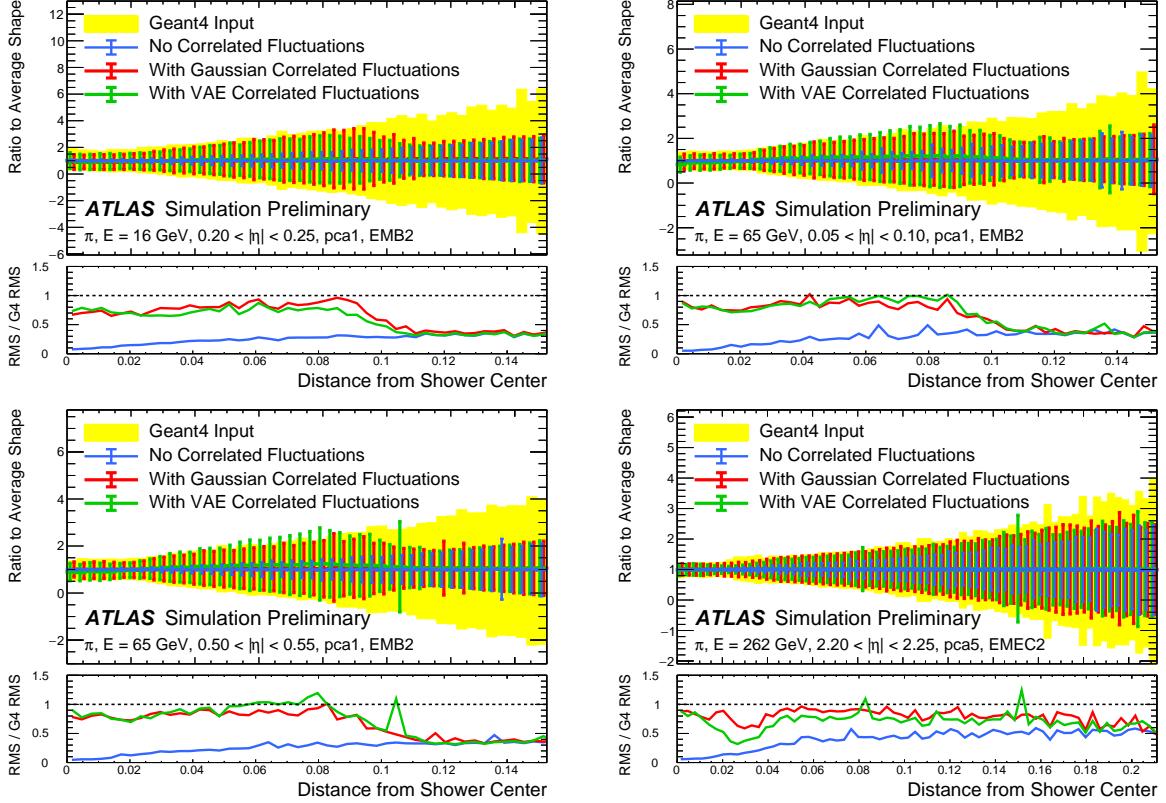


Figure 4.5: Comparison of the RMS fluctuations about the average shape with the Gaussian fluctuation model (red), the VAE fluctuation model (green), and without correlated fluctuations (blue) for a range of pion energies, η points, and layers.

of the GEANT4 RMS magnitude, compared to the 5 – 30% observed in the no correlated fluctuations case.

Figure 4.6 shows the result of a simulation with full ATLAS reconstruction for 65 GeV central pions with the Gaussian fluctuation model. The simulation with the Gaussian fluctuation model demonstrates improved modeling of several shape variables relative to baseline FastCaloSimV2, reproducing the distributions of events simulated with GEANT4.

The new fast calorimeter simulation is a crucial part of the future of simulation for the ATLAS Experiment at the LHC. The per event simulation time of the full detector with GEANT4,

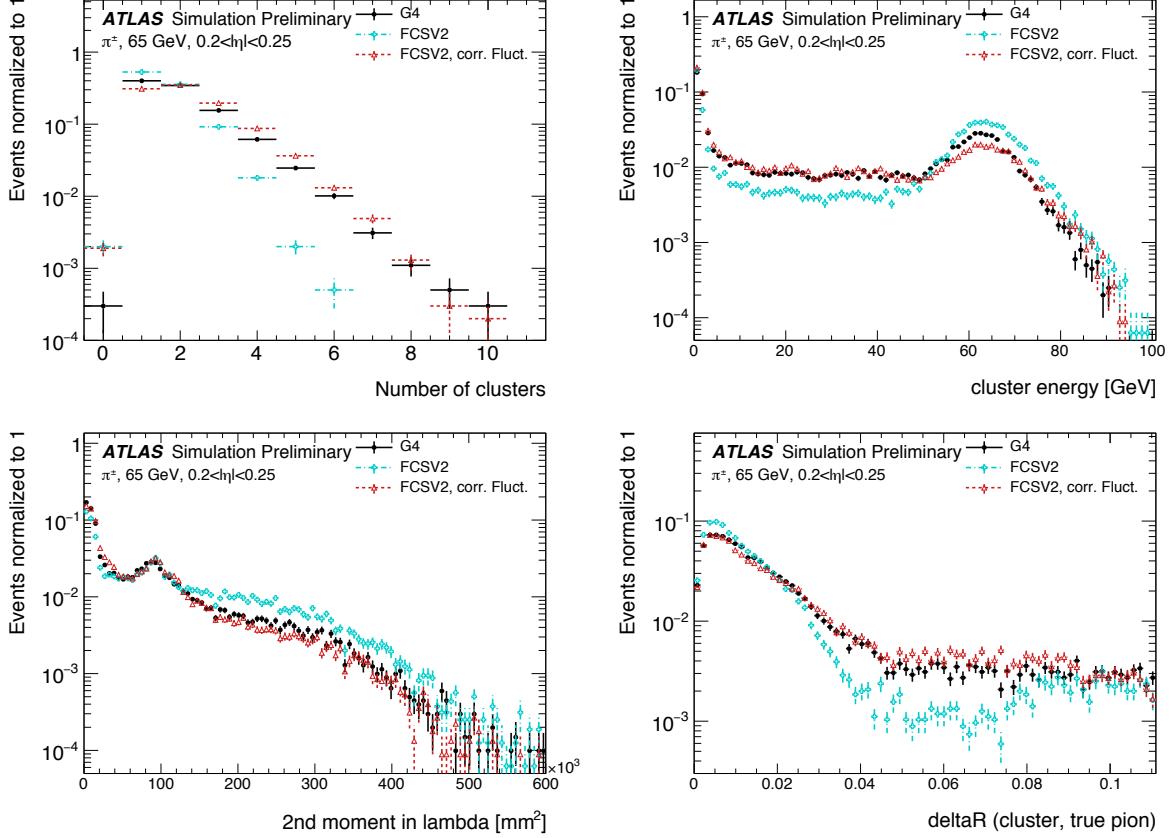


Figure 4.6: Comparison of the Gaussian fluctuation model to the default FCSV2 version and to G4 simulation, using pions of 65 GeV energy and $0.2 < |\eta| < 0.25$. With the correlated fluctuations, several shape variables demonstrate improved modeling.

calculated over 100 $t\bar{t}$ events, is 228.9 s. Using FastCaloSim for the calorimeter simulation reduces this to 26.6 s, of which FastCaloSim itself is only 0.015 s. Good physics modeling is achieved, the correlated fluctuations method shows good proof of concept improvement for the modeling of hadronic showers.

1317

Chapter 5

1318

RECONSTRUCTION

1319 Chapter 3 discusses how a proton-proton collision may be captured by a physical detector
 1320 and turned into data that may be stored and analyzed. Chapter 4 discusses the simulation
 1321 of this same process. At this most basic level, however, the ATLAS detector is only a
 1322 machine for turning particles into a set of electrical signals, albeit in a very sophisticated,
 1323 physics motivated way. This chapter discusses the step of turning these electrical signals into
 1324 objects which may be identified with the underlying physics processes, and therefore used to
 1325 make statements about what occurred within a given collision event. This process is termed
 1326 *reconstruction*, and we will focus particularly on jets and flavor tagging, as the most relevant
 1327 pieces for this thesis work.

1328 **5.1 Jets**

1329 As discussed in Chapters 3 and 4, the production of particles with color charge from a
 1330 proton-proton interaction is complicated both by parton showering and by confinement: a
 1331 quark produced from a hard scatter is not seen as a quark, but rather, as a spray of particles
 1332 with a variety of hadrons in the final state, which subsequently shower upon interaction with
 1333 the calorimeter in a complicated way.

1334 For hard scatter electrons, photons, or muons on the other hand, the picture is much
 1335 clearer: there is no parton showering, and each has a distinct signature in the detector:
 1336 photons have no tracks and a very localized calorimeter shower, electrons are associated
 1337 with tracks and are similarly localized in the calorimeter, and muons are associated with
 1338 tracks, pass through the calorimeter due to their large mass, and leave signatures in the muon
 1339 spectrometer.

1340 Jets are a tool to deal with the messiness of quarks and gluons. The basic concept is to
 1341 group the multitude of particles produced by hadronization into a single object. Such an
 1342 object then has associated properties, including a four-vector, which may be identified with
 1343 the corresponding initial state particle. In practice a variety of information from the ATLAS
 1344 detector is used for such a reconstruction. The analysis considered in this thesis uses particle
 1345 flow jets [61], which combines information from both the tracker and the calorimeter, where
 1346 the combined objects may be identified with underlying particles. However, jets built from
 1347 clusters of calorimeter cells [62] as well as from charged particle tracks [63] have also been
 1348 used very effectively.

1349 A variety of algorithms are used to associate detector level objects to a given jet. The
 1350 most commonly used in ATLAS is the anti- k_T algorithm [64], which is a successor to the
 1351 k_T algorithm, among others [65], and develops as follows. Both algorithms are sequential
 1352 recombination algorithms, which begin with the smallest distance, d_{ij} between considered
 1353 objects (e.g. particles or intermediate groupings of particles). If d_{ij} is less than a parameter
 1354 d_{iB} (B for “beam”) object i is combined with object j , the distance d_{ij} is recomputed, and
 1355 the process repeats. This proceeds until $d_{ij} \geq d_{iB}$, at which point the jet is “complete” and
 1356 removed from the list of considered objects.

The definitional difference between k_T and anti- k_T is these distance parameters. In general
 form, these are defined as

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (5.1)$$

$$d_{iB} = p_{Ti}^{2p} \quad (5.2)$$

1357 where p_{Ti} is the transverse momentum of object i , ΔR_{ij} is the angular distance between
 1358 objects i and j , R is a radius parameter, and p controls the tradeoff between the p_T and
 1359 angular distance terms. For the k_T algorithm $p = 1$; for the anti- k_T algorithm, $p = -1$. This
 1360 is a simple change, but results in significantly different behavior.

The anti- k_T algorithm can be understood as follows: for a single high p_T particle (p_{T1})
 surrounded by a bunch of low p_T particles, the low p_T particles will be clustered with the

high p_T one if

$$d_{1j} = \frac{1}{p_{T1}^2} \frac{\Delta R_{1j}^2}{R^2} < \frac{1}{p_{T1}^2} \quad (5.3)$$

$$\implies \Delta R_{1j} < R. \quad (5.4)$$

1361 Therefore, a single high p_T particle (p_{T1}) surrounded by a bunch of low p_T particles results in
 1362 a perfectly conical jet. This shape may change with the presence of other high momentum
 1363 particles, but the key feature of the dynamics is that the jet shape is determined by high p_T
 1364 objects due to the $\frac{1}{p_T}$ nature of this definition. In contrast, the k_T algorithm results in jets
 1365 influenced by low momentum particles, which results in a less regular shape. This property,
 1366 of regular jet shapes determined by high momentum objects, as well as demonstrated good
 1367 practical performance, makes the anti- k_T algorithm the favored jet algorithm in ATLAS.

1368 Because jets are composed of multiple objects, a useful property of jets is jet *substructure*,
 1369 that is, acknowledging that jets are composite objects, analyzing the structure of a given
 1370 jet to infer physics information. This leads to the use of *subjets*; that is, after running jet
 1371 clustering, often to create a “large-R”, $R = 1.0$ anti- k_T jet, a smaller radius jet clustering
 1372 algorithm is run within the jet. Subjets are often chosen using the k_T algorithm, which again
 1373 is sensitive to lower momentum particles, with $R = 0.2$ or 0.3 . For the boosted version of this
 1374 thesis analysis, such a strategy is used, in which the subjets are *variable radius* and depend
 1375 on the momentum of the (proto)jet in question. Beyond this thesis work, substructure is
 1376 used in a large variety of analyses, with a set of associated variables and tools developed for
 1377 exploiting this structure *TODO: Cite some?*.

1378 5.2 Flavor Tagging

1379 For this this thesis, the physics process being considered is $HH \rightarrow b\bar{b}b\bar{b}$. From the previous
 1380 section, we know that the standard practice is to identify these b quarks (or, rather, the
 1381 resulting B hadrons, due to confinement) with jets – in our case, these b -*jets* are $R=0.4$
 1382 anti- k_T particle flow jets (see Chapter 7). However, not all jets produced at the LHC are
 1383 from B hadrons; rather, there are a variety of different types of jets corresponding to different

1384 flavors of quarks. These are often classified as light jets (from u , d , or s quarks, or gluons)
 1385 or as other *heavy flavor* jets, e.g. c -jets, involving c quarks. Distinguishing between these
 1386 different categories is a very active area of work in ATLAS, termed *flavor tagging*, with much
 1387 focus on *b-tagging* in particular, that is, the identification of jets from B hadron decays. We
 1388 here briefly describe the techniques used for flavor tagging in ATLAS.

1389 What distinguishes a b -jet from any other jet? This question is fundamental to the design
 1390 of the various b -tagging algorithms, and has two major answers: (1) B hadrons have long
 1391 lifetimes, and (2) B hadrons have large masses. It is most illustrative to compare the B hadron
 1392 properties to a common light meson, e.g. π^0 , the neutral pion, with quark content $\frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$.
 1393 B hadrons have lifetimes around 1.5 ps, corresponding to a decay length $c\tau \approx 0.45$ mm. In
 1394 contrast, π^0 has a lifetime of 8.4×10^{-5} ps, which is around 20,000 times shorter! Theoretically,
 1395 this comes from CKM suppression of the b to c transition *TODO: check*, which dominates
 1396 the B decay modes. Experimentally, this difference pops up as shown in Figure 5.1 – light
 1397 flavor initiated jets decay almost immediately at the proton-proton interaction point, whereas
 1398 b -jets are distinguished by a displaced secondary vertex, corresponding to the 5 mm decay
 1399 length calculated above. This displaced vertex falls short of the detector itself, but may be
 1400 inferred from larger transverse (perpendicular to beam) and longitudinal (parallel to beam)
 1401 impact parameters of the resulting tracks, termed d_0 and z_0 respectively.

1402 Coming to the mass, B mesons have masses of around 5.2 GeV, whereas the π^0 mass
 1403 is around 0.134 GeV, (around 40 times lighter). This higher mass gives access to a larger
 1404 decay phase space, leading to a high multiplicity for b -jets (average of 5 charged particles per
 1405 decay).

1406 One final distinguishing feature of B hadrons is their *fragmentation function*, a function
 1407 describing the production of an observed final state. For B hadrons, this is particularly
 1408 “hard”, with the B hadrons themselves contributing to an average of around 75 % of the b -jet
 1409 energy. Thus, the identification of b -jets with B hadrons is, in some sense, descriptive.

1410 We have contrasted b -jets and light jets, demonstrating that there are several handles
 1411 available for making this distinction. c -jets are slightly more similar to b -jets, but the same

1412 handles still apply – c -hadron lifetimes are between 0.5 and 1 ps, a factor of 2 smaller than B
1413 hadrons. Their mass is around 1.9 GeV, 2 to 3 times smaller than B hadrons, and c -hadrons
1414 contribute to an average of around 55 % of c -jet energy. Therefore, while the gap is slightly
1415 smaller, a distinction may still be made.

1416 The ATLAS flavor tagging framework [67] relies on developing a suite of “low-level”
1417 taggers, which use a variety of information about tracks and vertices as inputs. The output
1418 of these lower level taggers are then fed into a higher level tagger, which aggregates these
1419 results into a high level discriminant. Each of these taggers is described below.

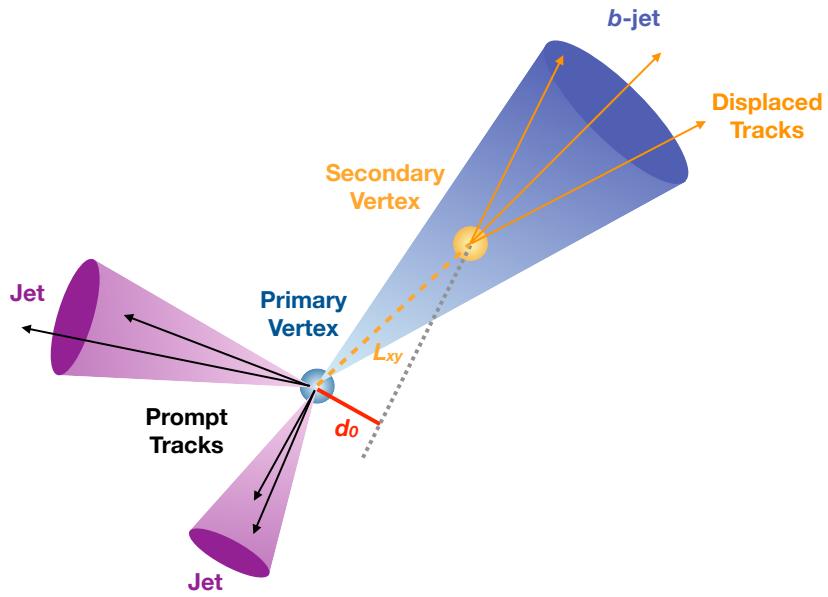


Figure 5.1: Illustration of an interaction producing two light jets and one b -jet in the transverse plane. While the light jets decay “promptly”, coinciding with the primary vertex of the proton-proton interaction, the longer lifetime of B hadrons leads to a secondary decay vertex, displaced from the primary vertex by length L_{xy} . This is typically a few mm, and therefore is not directly visible in the detector, but leads to a large transverse impact parameter, d_0 , for the resulting tracks. [66]

1420 5.2.1 IP2D/3D

1421 IP2D and IP3D are taggers based on the large track impact parameter (IP) nature of B
 1422 hadron decays. Both are based on histogram templates derived from Monte Carlo simulation,
 1423 which are used as probability density functions to construct log-likelihood discriminants.
 1424 IP2D incorporates just the transverse impact parameter information using 1D histogram
 1425 templates, whereas IP3D uses both transverse and longitudinal impact parameters in a 2D
 1426 template, which accounts for correlations. Importantly, these are *signed* impact parameters,
 1427 with sign based on the angle between the impact parameter and the considered jet – positive
 1428 impact parameters are consistent with a track extrapolation in front of the jet (angle between
 1429 impact parameter line and jet $< 90^\circ$), and therefore more consistent with tracks originating
 1430 from a displaced decay.

1431 Rather than using the impact parameters directly, an impact parameter *significance*
 1432 is used which incorporates an uncertainty on the impact parameter – tracks with a lower
 1433 uncertainty but the same impact parameter will contribute more in the calculation. This
 1434 signed significance is what is used to sample from the PDF templates, with the resulting
 1435 discriminants the sum of probability ratios between given jet hypotheses over tracks associated
 1436 to a given jet, namely $\sum_{i=1}^N \log \frac{p_b}{p_{light}}$ between b -jet and light jet hypotheses, where p_b and
 1437 p_{light} are the per-track probabilities. Similar discriminants are defined between b - and c -jets
 1438 and c and light jets. *TODO: show distributions?*

1439 5.2.2 SV1

1440 SV1 is an algorithm which aims to find a secondary vertex (SV) in a given jet. Operating
 1441 on all vertices associated with a considered jet, the algorithm discards tracks based on a
 1442 variety of cleaning requirements. It then proceeds to first construct all two-track vertices,
 1443 and then iterates over all the tracks involved in these two track vertices to try to fit a single
 1444 secondary vertex, which would then be identified with the secondary vertex from the b or c
 1445 hadron decay. This fit proceeds by evaluating a χ^2 for the association of a track and vertex,

1446 removing the track with the largest χ^2 , and iterating until the χ^2 is acceptable and the vertex
1447 has an invariant mass of less than 6 GeV (for consistency with b or c hadron decay).

1448 A variety of discriminating variables may then be constructed, including (1) invariant
1449 mass of the secondary vertex, (2) number of tracks associated with the secondary vertex, (3)
1450 number of two-track vertices, (4) energy fraction of the tracks associated to the secondary
1451 vertex (relative to all of the tracks associated to the jet), and various metrics associated with
1452 the secondary vertex position and decay length, including (5) transverse distance between the
1453 primary and secondary vertex, (6) distance between the primary and secondary vertex (7)
1454 distance between the primary and secondary vertex divided by its uncertainty, and (8) ΔR
1455 between the jet axis and the direction of the secondary vertex relative to the primary vertex.

1456 While all eight of these variables are used as inputs to the higher level taggers, the number
1457 of two-track vertices, the vertex mass, and the vertex energy fraction are additionaly used with
1458 3D histogram templates to evaluate flavor tagging performance by constructing log-likelihood
1459 discriminants, similar to the procedure for IP2D/3D.

1460 5.2.3 JetFitter

1461 Rather than focusing on a particular aspect of the B hadron or D hadron decay topology
1462 (e.g impact parameter or secondary vertex), the third low level tagger, JETFITTER [68],
1463 tries to reconstruct the full decay chain, including all involved vertices. This is structured
1464 around a Kalman filter formalism [69], and has the strong underlying assumption that all
1465 tracks which stem from B and D hadron decay must intersect a common flight path. This
1466 assumption provides significant constraints, allowing for the reconstruction of vertices from
1467 even a single track, reducing the number of degrees of freedom in the fit, and allowing the
1468 use of “downstream” information, e.g., compatibility of tracks with a $B \rightarrow D$ -like decay.
1469 The constructed topology, including primary vertex location and B -hadron flight path, is
1470 iteratively updated over tracks associated to a given jet, and a variety of discriminating
1471 variables related to the resulting topology and reconstructed decay are used as inputs to the
1472 high level taggers.

1473 5.2.4 *RNNIP*

1474 The IP2D and IP3D algorithms rely on per-track probabilities, and the final discriminating
1475 variables (and inputs to the higher level taggers) are sums (products) over these independently
1476 considered quantities. In practice, however, the tracks are not independent – this is merely a
1477 simplifying assumption to allow for the use of a binned likelihood, as treatment of all of the
1478 interdependencies in such a framework quickly becomes intractable. To address this issue, a
1479 recurrent neural network-based algorithm, RNNIP [70], is used, which takes as input a variety
1480 of per-track variables, including the signed impact parameter significances (as in IP3D) as
1481 well as track momentum fraction relative to the jet and ΔR between the track and the jet.
1482 RNNs are sequence-based, and vectors of input variables corresponding to tracks for a given
1483 jet are ordered by magnitude of transverse impact parameter significance and then passed
1484 to the network, which outputs class probabilities corresponding to b-jet, c-jet, light-jet, and
1485 τ -jet hypotheses. Such a procedure allows the network to learn interdependencies between
1486 tracks, improving performance.

1487 5.2.5 *MV2 and DL1*

1488 Outputs from the above taggers are combined into high level taggers to aggregate all of the
1489 information and improve discriminating power relative to the respective individual taggers as,
1490 as shown in Figure 5.2. These high level taggers are primarily in two forms: MV2, which
1491 uses a Boosted Decision Tree (BDT) for this aggregation, and DL1, which uses a deep neural
1492 network. For the baseline versions of these taggers, only inputs from IP2D, IP3D, SV1, and
1493 JetFitter are used. The tagger used for this thesis analysis, DL1r, additionally incorporates
1494 RNNIP, demonstrating improved performance over the baseline DL1, as shown in Figure 5.3.
1495 All high level taggers also include jet p_T and $|\eta|$.

DL1 offers a variety of improvements over MV2. Rather than a single discriminant output, as with MV2, DL1 has a multidimensional output, corresponding to probabilities for a jet to be a *b*-jet, *c*-jet, or light jet. This allows the trained network to be used for both *b*- and *c*-jet

tagging. The final discriminant for b -tagging with DL1 correspondingly takes the form

$$D_{\text{DL1}} = \ln \left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\text{light}}} \right) \quad (5.5)$$

where p_b , p_c , and p_{light} are the output b , c , and light jet probabilities, and f_c corresponds to an effective c -jet fraction, which may be tuned to optimize performance.

DL1 further includes an additional set of JETFITTER input variables relative to MV2 which correspond to c -tagging – notably properties of secondary and tertiary vertices, as would be seen in a $B \rightarrow D$ decay chain.

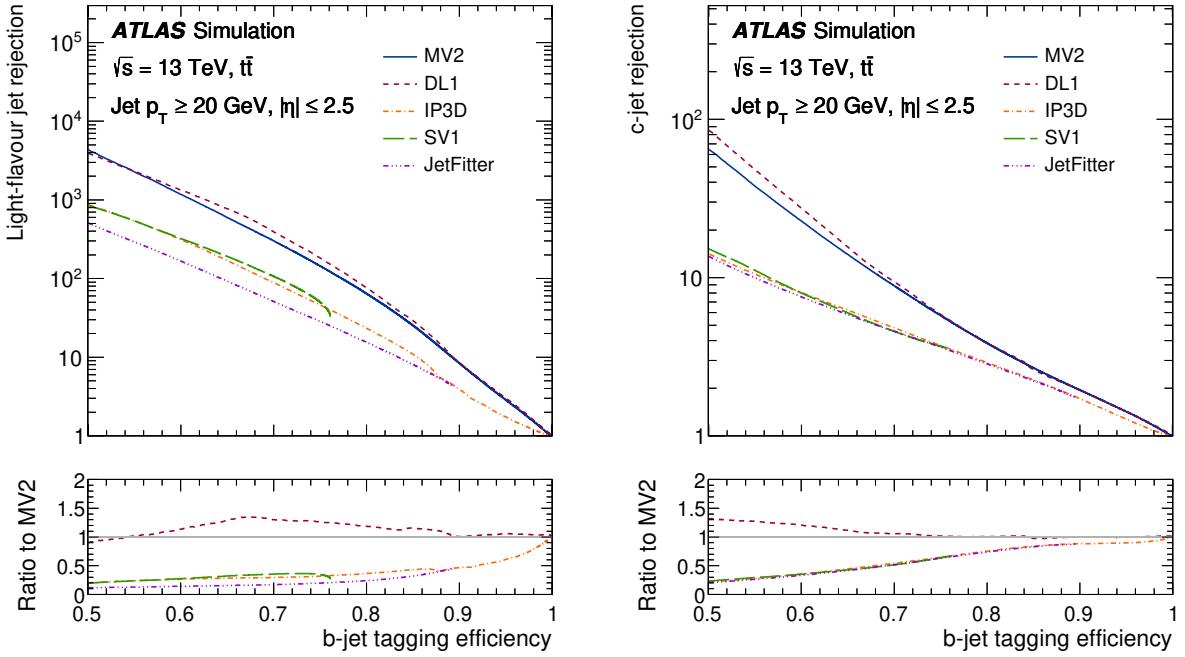


Figure 5.2: Performance of the various low and high level flavor tagging algorithms in $t\bar{t}$ simulation, demonstrating the tradeoff between b -jet efficiency and light and c -jet rejection. The high level taggers demonstrate significantly better performance than any of the individual low level taggers, with DL1 offering slight improvements over MV2 due to the inclusion of additional input variables.

Figure 5.2 shows a comparison of the performance of the various taggers. The b -tagging performance of DL1 and MV2 is found to be similar, with some improvements in light jet and c -jet rejection from the additional variables used in DL1. The performance of these high level taggers additionally is seen to be significantly better than any of the individual low level ones, even in regimes where only a single low level tagger is relevant (such as high b -tagging efficiencies, where SV1 and JETFITTER are limited by selections on tracks entering the respective algorithms).

The inclusion of RNNIP offers a significant improvement on top of baseline DL1, as shown in Figure 5.3, strongly motivating the choice of DL1r for this thesis.

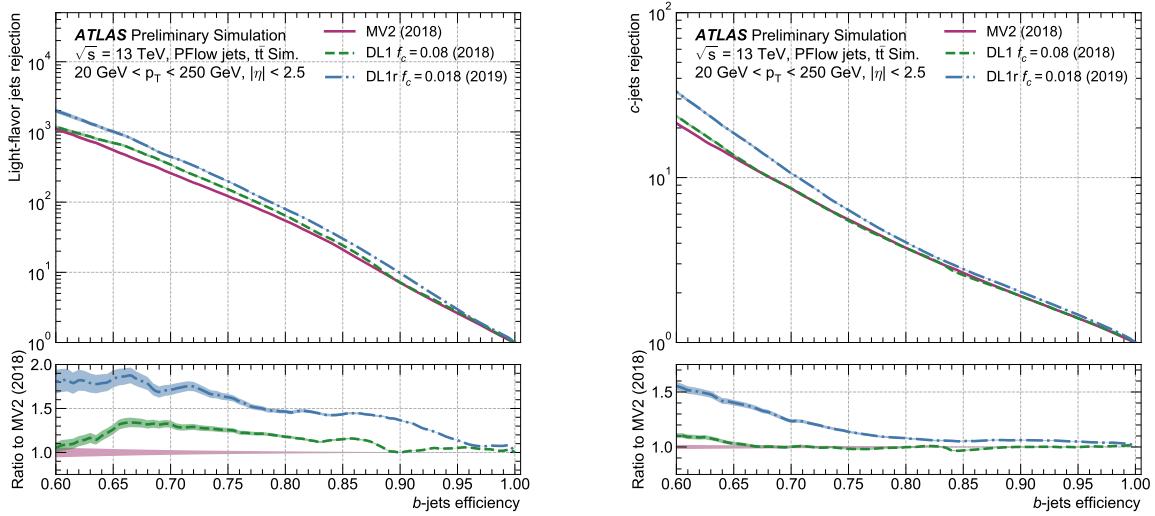


Figure 5.3: Performance of the MV2, DL1, and DL1r algorithms in $t\bar{t}$ simulation, demonstrating the tradeoff between b -jet efficiency and light and c -jet rejection. f_c controls the importance of c -jet rejection in the discriminating variable, and values shown have been optimized separately for each DL1 configuration. DL1r demonstrates a significant improvement in both light and c jet rejection over MV2 and DL1. [71]

1510 5.2.6 *Some Practical Notes*

1511 The b -tagging metrics presented in Figures 5.2 and 5.3 correspond to evaluating a tradeoff
1512 between b -jet efficiency and light jet and c -jet rejection. In this case, b -jet efficiency is defined
1513 such that, e.g. for a 77 % efficiency, 77 % of the real b -jets will be tagged as such. Somewhat
1514 counterintuitively, this means that a lower b -jet efficiency corresponds to a more aggressive
1515 (“tighter”) selection on the discriminating variable, while a higher b -jet efficiency corresponds
1516 to a less aggressive (“looser”) cut (100 % efficiency means no cut). Light and c jet efficiencies
1517 are defined similarly, with rejection defined as 1/ the corresponding efficiency.

1518 In ATLAS, the respective b -tagging efficiencies are used to define various b -tagging working
1519 points. The working point used for the nominal b -jet identification in this thesis is 77 % with
1520 DL1r. A loosened (less aggressive) selection at the 85 % working point is additionally used.
1521 See Chapter 7 for further details.

1522

Chapter 6

1523

THE ANATOMY OF AN LHC SEARCH

1524 In this thesis so far, we have set the theoretical foundation for the work carried out at the
1525 LHC. We have described how one may translate between this theoretical foundation and what
1526 we are actually able to observe with the ATLAS detector. We have further stepped through
1527 the process of simulating production of specific physics processes and their appearance in
1528 our detector, allowing us to describe how a hypothetical physics model would be seen in
1529 our experiment. The question then becomes: all of these pieces are on the table, what do
1530 we do with them? This chapter attempts to answer exactly that, setting up a roadmap for
1531 assembling these pieces into a statement about the universe.

1532 **6.1 Object Selection and Identification**

1533 As described in Chapter 5, there is a complicated set of steps for going from electrical signals
1534 in a detector to physics objects.

1535 **6.2 Defining a Signal Region**

1536 **6.3 Background Estimation**

1537 **6.4 Uncertainty Estimation**

1538 **6.5 Hypothesis Testing**

1539

Chapter 7

1540

SEARCH FOR PAIR PRODUCTION OF HIGGS BOSONS IN THE $b\bar{b}b\bar{b}$ FINAL STATE

1541

1542 This chapter presents two complementary searches for pair production of Higgs bosons
 1543 in the final state. Such searches are separated based on the signal models being considered:
 1544 resonant production, in which a new spin-0 or spin-2 particle is produced and decays to two
 1545 Standard Model Higgs bosons, and non-resonant production, which is sensitive to the value
 1546 of the Higgs self-coupling λ_{HHH} . Further information on the theory behind both channels
 1547 can be found in Chapter 2.

1548

While the searches face many similar challenges and proceed (in broad strokes) in a very
 similar manner, separate optimizations are performed to maximize the respective sensitivities
 for these two very different sets of signal hypotheses. More particularly, resonant signal
 hypotheses are (1) very peaked in values of the mass of the HH candidate system near
 the value of the resonance mass considered and (2) considered across a very broad range of
 signal mass hypotheses. The resonant searches are therefore split into resolved and boosted
 topologies based on Lorentz boost of the decay products, with the resolved channel as one of
 the primary focuses of this thesis. Further, several analysis design decisions are made to allow
 for sensitivity to a broad range of masses – in particular, though sensitivity is limited at lower
 values of m_{HH} relative to other channels *TODO: Combination, bbyy* due to the challenging
 background topology, retaining and properly reconstructing these low mass events allows the
 $b\bar{b}b\bar{b}$ channel to retain sensitivity up until the kinematic threshold at 250 GeV.

1560

In contrast, non-resonant signal hypotheses are quite broad in m_{HH} , and have a much
 more limited mass range, with Standard Model production peaking near 400 GeV, and the
 majority of the analysis sensitivity able to be captured with a resolved topology. Even for

1563 Beyond the Standard Model signal hypotheses, which may have more events at low m_{HH} ,
 1564 the non-resonant nature of the production allows the $b\bar{b}b\bar{b}$ channel to retain sensitivity while
 1565 discarding much of the challenging low mass background. Such freedom allows for decisions
 1566 which focus on improved background modeling for the middle to upper HH mass regime,
 1567 resulting in improved modeling and smaller uncertainties than would be obtained with a
 1568 more generic approach.

1569 Both searches are presented in the following, with emphasis on particular motivations for,
 1570 and consequences of, the various design decisions involved for each respective set of signal
 1571 hypotheses.

1572 The analyses improve upon previous work ?? in several notable ways. The resonant search
 1573 leverages a Boosted Decision Tree (BDT) based pairing algorithm, offering improved HH
 1574 pairing efficiency over a broad mass spectrum. The non-resonant adopts a different approach,
 1575 with a simplified algorithm based on the minimum angular distance (ΔR) between jets in
 1576 a Higgs candidate. Such an approach very efficiently discards low mass background events,
 1577 resulting in an easier to estimate background with reduced systematic uncertainties.

1578 A particular contribution of this thesis is the background estimation, which uses a novel,
 1579 neural network based approach, offering improved modeling over previous methods, as well
 1580 as the ability to model correlations between observables. While all aspects of the analysis of
 1581 course contribute to the final result, the author of this thesis wishes to emphasize that the
 1582 background estimate, with the corresponding uncertainties and all other associated decisions,
 1583 really is the core of the $HH \rightarrow b\bar{b}b\bar{b}$ analysis – the development of this procedure, and all
 1584 associated decisions, is similarly the core of this thesis work.

1585 ATLAS has performed a variety of searches in complementary decay channels as well, no-
 1586 tably in the $b\bar{b} W^+ W^-$ [72], $b\bar{b}\tau^+\tau^-$ [73], $W^+ W^- W^+ W^-$ [74], $b\bar{b}\gamma\gamma$ [75], and $W^+ W^- \gamma\gamma$ [76]
 1587 final states, which were combined along with $b\bar{b}b\bar{b}$ in [21].

1588 CMS has also performed searches for resonant production of Higgs boson pairs in the
 1589 $b\bar{b}b\bar{b}$ final state (among others) at $\sqrt{s} = 8$ TeV [77] and $\sqrt{s} = 13$ TeV [78]. CMS have also
 1590 performed a combination of their searches in the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, and $b\bar{b}VV$ channels

1591 in [79].

1592 This analysis also benefits from improvements to ATLAS jet reconstruction and calibration,
1593 and flavour tagging [67]. In particular, this analysis benefits from the introduction of particle
1594 flow jets [61]. These make use of tracking information to supplement calorimeter energy
1595 deposits, improving the angular and transverse momentum resolution of jets by better
1596 measuring these quantities for charged particles in those jets.

1597 The analysis also benefits from the new DL1r ATLAS flavour tagging algorithm. Whereas
1598 the flavour tagging algorithm used in the previous analysis (MV2) used a boosted decision
1599 tree (BDT) to combine the output of various low level algorithms, DL1r (and the baseline
1600 DL1 algorithm) uses a deep neural network to do this combination. In addition to the low
1601 level algorithms used as inputs to MV2, DL1 includes a variety of additional variables used
1602 for c -tagging. DL1r further incorporates RNNIP, a recurrent neural network designed to
1603 identify b -jets using the impact parameters, kinematics, and quality information of the tracks
1604 in the jets, while also taking into account the correlations between the track features.

1605 The overall analysis sensitivity further benefits from a factor of ~ 4.6 increase in integrated
1606 luminosity.

1607 7.1 Data and Monte Carlo Simulation

1608 Both the resonant and non-resonant searches are performed on the full ATLAS Run 2 dataset,
1609 consisting of $\sqrt{s} = 13\text{ TeV}$ proton-proton collision data taken from 2016 to 2018 inclusive.
1610 Data taken in 2015 is not used due to a lack of trigger jet matching information and b -jet
1611 trigger scale factors. The integrated luminosity collected and usable in this analysis¹ was:

1612 • 24.6 fb^{-1} in 2016

1613 • 43.65 fb^{-1} in 2017

¹approximately 9 fb^{-1} of data was collected but could not be used in this analysis due to an inefficiency in the b -jet triggers at the start of 2016 [80]

- 1614 • 57.7 fb^{-1} in 2018

1615 This gives a total integrated luminosity of 126 fb^{-1} . This is lower than the 139 fb^{-1} ATLAS
1616 collected during Run 2 [81] due to the inefficiency described in footnote 1 as well as the
1617 3.2 fb^{-1} of 2015 data which is unused due to the trigger scale factor issue mentioned above.

1618 In this analysis, Monte Carlo samples are used purely for modelling signal processes. The
1619 background is strongly dominated by events produced by QCD multijet processes, which
1620 are difficult to correctly model in simulation. This necessitates the use of a data-driven
1621 background modelling technique, which is described in Section 7.6.

1622 The scalar resonance signal model is simulated at leading order in α_s using MADGRAPH
1623 [47]. Hadronization and parton showering are done using HERWIG 7 [51][52] with EVTGEN [54],
1624 and the nominal PDF is NNPDF 2.3 LO. In practice this is implemented as a two Higgs
1625 doublet model where the new neutral scalar is produced through gluon fusion and required
1626 to decay to a pair of SM Higgs bosons. The heavy scalar is assigned a width much smaller
1627 than detector resolution, and the other 2HDM particles do not enter the calculation.

1628 Scalar samples are produced at resonance masses between 251 and 900 GeV and the
1629 detector simulation is done using AtlFast-II [56]. In addition the samples at 400 GeV and
1630 900 GeV are also fully simulated to verify that the use of AtlFast-II is acceptable. For higher
1631 masses, as well as for the boosted analysis, samples are produced between 1000 and 5000 GeV,
1632 and the detector is fully simulated. As discussed in Chapter 4, an outstanding issue with
1633 AtlFast-II is the modeling of jet substructure. While such variables are not used for the
1634 resolved analysis, the boosted analysis begins at 900 GeV, motivating the different detector
1635 simulation in these two regimes.

1636 The spin-2 resonance signal model is also simulated at LO in α_s using MADGRAPH.
1637 Hadronization and parton showering are done using PYTHIA 8 [53] with EVTGEN, and the
1638 nominal PDF is NNPDF 2.3 LO. In practice this is implemented as a Randall-Sundrum
1639 graviton with $c = 1.0$.

1640 Spin-2 resonance samples are produced at masses between 251 and 5000 GeV, and these

1641 samples are all produced with full detector simulation.

1642 For the non-resonant search, samples are produced at values of $\kappa_\lambda = 1.0$ and 10.0, and are
1643 simulated using Powheg Box v2 generator [48–50] at next-to-leading order (NLO), with full
1644 NLO corrections with finite top mass, using the PDF4LHC [82] parton distribution function
1645 (PDF) set. Parton showers and hadronization are simulated with Pythia 8.

1646 Alternative ggF samples are simulated at NLO using Powheg Box v2, but instead using
1647 Herwig 7 [83] for parton showering and hadronization. The comparison between these two
1648 is used to assess an uncertainty on the parton showering.

1649 7.2 Triggers and Object Definitions

1650 To maximize analysis sensitivity, a combination of multi- b -jet triggers is used. Due to the use
1651 of events with two b -tagged jets in the background estimate, such triggers have a maximum
1652 requirement of two b -tagged jets. For the resonant analysis, a combination of triggers of
1653 various topologies is used, namely

1654 • 2b + HT, which requires two b -tagged jets and a minimum value of of H_T , defined to
1655 be the scalar sum of p_T across all jets in the event.

1656 • 2b + 2j, which requires two b -tagged jets and two other jets matching some kinematic
1657 requirements

1658 • 2b + 1j, which requires two b -tagged jets and one other jet matching some kinematic
1659 requirements

1660 • 1b, which requires one b -tagged jet

1661 Due to minimal contributions from some of these triggers for the Standard Model non-resonant
1662 signal, a simplified strategy relying entirely on 2b + 1j and 2b + 2j triggers is used for the
1663 non-resonant search.

1664 While the use of multiple triggers is beneficial for analysis sensitivity, it comes with some
 1665 complications. Namely, a set of scale factors must be assigned to simulated events account
 1666 for trigger inefficiencies in data *TODO: check*. Because these scale factors may differ between
 1667 triggers, the use of multiple triggers becomes complicated: an event may pass more than one
 1668 trigger, while trigger scale factors are only provided for individual triggers.

1669 To simplify this calculation, a set of hierarchical offline selections is applied, closely
 1670 mimicking the trigger selection. Based on these selections, events are sorted into categories
 1671 (*trigger buckets*), after which the decision of a *single trigger* is checked.

1672 The resonant search applies such categorization in the following way, with selections
 1673 considered in order:

- 1674 1. If the leading jet is b -tagged with $p_T > 325 \text{ GeV}$, the event is in the $1b$ trigger category.
- 1675 2. Otherwise, if the leading jet is not b -tagged, but has $p_T > 168.75 \text{ GeV}$, the event is in
 1676 the $2b + 1j$ trigger category.
- 1677 3. If neither of the first two selections pass, if the scalar sum of jet p_T s, $H_T > 900 \text{ GeV}$,
 1678 the event falls into the $2b + HT$ trigger category.
- 1679 4. Events that do not pass any of the above offline selections are in the $2b + 2j$ trigger
 1680 category.

1681 Corresponding triggers are then checked in each category, and the final set of events consists
 1682 of those events that pass the trigger decision in their respective categories.

1683 For the resonant search, the $2b + 1j$ and $2b + 2j$ triggers are the dominant categories,
 1684 containing roughly 26 % and 49 % of spin-2 events, evaluated on MC16d samples with
 1685 resonance masses between 300 and 1200 GeV. Notably, the $1b$ trigger efficiency is largest at
 1686 high ($> 1 \text{ TeV}$) resonance masses.

1687 For the non-resonant search, it was noted that the $1b$ trigger has minimal contribution,
 1688 while the $2b + HT$ events are largely captured by the $2b + 2j$ trigger. Therefore, for, a

1689 simplified scheme is considered, with selections:

- 1690 1. If the 1st leading jet has $p_T > 170 \text{ GeV}$ and the 3rd leading jet has $p_T > 70 \text{ GeV}$, the
1691 event is in the $2b + 1j$ trigger category.
- 1692 2. Otherwise, the event is in the $2b + 2j$ trigger category.

1693 **7.3 Analysis Selection**

1694 After the trigger selections of Section 7.2, a variety of selections on the analysis objects are
1695 made, with the goal of (1) reconstructing a HH -like topology and (2) suppressing contributions
1696 from background processes.

1697 Both analyses begin with a common pre-selection, requiring at least four $R = 0.4$ anti- k_T
1698 jets with $|\eta| < 2.5$ and $p_T > 40 \text{ GeV}$. The $|\eta| < 2.5$ requirement is necessary for b -tagging
1699 due to the coverage of the ATLAS tracking detector (see Chapter 3) *TODO: check*, while the
1700 $p_T > 40 \text{ GeV}$ requirement is motivated by the trigger thresholds *TODO: mention low pT*. At
1701 least two of the jets passing this pre-selection are required to be b -tagged, and additional
1702 b -tagging requirements are made to define the following regions:

- 1703 • “2 b Region”: require exactly two b -tagged jets, used for background estimation
- 1704 • “4 b Region”: require at least (but possibly more) four b -tagged jets, used as a signal
1705 region for both resonant and non-resonant searches

1706 The non-resonant analysis additionally defines two 3 b regions:

- 1707 • “3 $b+1$ loose Region”: require exactly three b -tagged jets which pass the 77 % b-tagging
1708 working point (nominal) and one additional jet that fails the 77 % b-tagging working
1709 point but passes the *looser* 85 % b-tagging working point. Used as a signal region for
1710 the non-resonant search.

- “3 b +1 fail Region”: complement of 3 b +1 loose. Require exactly three b -tagged jets which pass the 77 % b-tagging working point, but require that none of the remaining jets pass the 85 % b-tagging working point. Used as a validation region for the non-resonant search.

After these requirements, four jets are chosen, ranked first by b -tagging requirement and then by p_T (e.g. for the 2 b region, the jets chosen are the two b -tagged jets and the two highest p_T non-tagged jets; for the 4 b region, the jets are the four highest p_T b -tagged jets). To match the topology of a $HH \rightarrow b\bar{b}b\bar{b}$ event, these four jets are then *paired* into *Higgs candidates*: the four jets are split into two sets of two, and each of these pairs is used to define a reconstructed object that is a proxy for a Higgs in a HH event.

For four jets there are three possible pairings. For signal events, a correct pairing can be identified (provided all necessary jets pass pre-selection) using the truth information of the Monte Carlo simulation, and such information may be used to design/select an appropriate pairing algorithm. This is only part of the story, however. The vast majority of the events in data do *not* include a real HH decay (this is a search for a reason!), either because the event originates from a background process (e.g. for 4 b events), or because the selection is not designed to maximize the signal (e.g. 2 b events). As the pairing is part of the selection, it must still be run on such events, such that various algorithms which achieve similar performance in terms of pairing efficiency may have vastly different impacts in terms of the shape of the background and the biases inherent in the background estimation procedure. The interplay between these two facets of the pairing is an important part of the choices made for this analysis.

A comparison of different shapes due to three different paring strategies is shown in Figure 7.1.

1735 7.3.1 Resonant Pairing Strategy

1736 For the resonant analysis, a Boosted Decision Tree (BDT) is used for the pairing. The boosted
 1737 decision tree is given the total separation between the two jets in each of the two pairs (ΔR_1
 1738 and ΔR_2), the pseudo-rapidity separation between the two jets in each pair ($\Delta\eta_1$ and $\Delta\eta_2$),
 1739 and the angular separation between the two jets in each pair in the $x - y$ plane ($\Delta\phi_1$ and
 1740 $\Delta\phi_2$). The total separations (ΔR_s) are provided in addition to the components in order to
 1741 avoid requiring the boosted decision tree to reconstruct these variables in order to use them.
 1742 For these variables, pair 1 is the pair with the highest scalar sum of jet p_{T} s, and pair 2 the
 1743 other pair.

1744 The boosted decision tree is also parameterized on the di-Higgs mass (m_{HH}) by providing
 1745 this as an additional feature. Since the boosted decision tree is trained on correct and
 1746 incorrect pairings in signal events, there will be exactly one correct pairing and two incorrect
 1747 pairings in the training set for each m_{HH} value present in that set. As a result, this variable
 1748 cannot, in itself, distinguish a correct pairing from an incorrect pairing, and therefore the
 1749 inclusion of this variable simply serves to parameterize the BDT on m_{HH} ².

1750 The boosted decision tree was trained on one quarter of the total AFII simulated scalar
 1751 MC statistics, using the Gradient-based One Side Sampling (GOSS) algorithm which allows
 1752 rapid training with very large datasets. A preselection was applied requiring events to have
 1753 four jets with a p_{T} of at least 35 GeV. Note that this is a looser requirement than the 40 GeV
 1754 used in the analysis selection, and is meant to increase the available statistics for events with
 1755 low m_{HH} and to ensure a better performance as a function of that variable. Events were also
 1756 required to have four distinct jets that could be geometrically matched (to within $\Delta R \leq 0.4$)
 1757 to the b -quarks. The events used to train the BDT were not included when the analysis was
 1758 run on these signal simulations. The boosted decision tree was constructed with the following
 1759 hyperparameters:

1760 `min_data_in_leaf=50,`

²That is, the conditions placed on the other variables by the BDT vary with m_{HH} .

1761 num_leaves=180,
 1762 learning_rate=0.01

1763 These hyperparameters were optimized using a Bayesian optimization procedure [84].
 1764 Three fold cross-validation was used to perform this optimization without the need for an
 1765 additional sample, while avoiding over-training on signal events.

1766 *7.3.2 Non-resonant Pairing Strategy*

1767 For the non-resonant analysis, a simpler pairing algorithm is used, which proceeds as follows:
 1768 in a given event, Higgs candidates for each possible pairing are sorted by the p_T of the vector
 1769 sum of constituent jets. The angular separation, ΔR is computed between jets in the each of
 1770 the leading Higgs candidates, and the pairing with the smallest separation (ΔR_{jj}) is selected.
 1771 This method will be referred to as $\min \Delta R$ in the following.

1772 The primary motivation for the use of this pairing in the non-resonant search is a *smooth*
 1773 *mass plane*: by efficiently discarding low mass events, $\min \Delta R$ removes the background peak
 1774 present in the resonant search while maintaining good pairing efficiency for the Standard
 1775 Model non-resonant signal. This facilitates a background estimate with small kinematic bias
 1776 – the region in which the background estimate is derived is more similar to the signal region.

1777 Along with discarding low mass background, there is a corresponding loss of low mass
 1778 signal. This predominantly impacts points away from the Standard Model (see Figure 7.2),
 1779 but, because the $4b$ channel has the strongest contribution near the Standard Model and
 1780 because of the large low mass background present with other pairing methods, the impact on
 1781 analysis sensitivity is mitigated. The $\min \Delta R$ pairing is thus adopted for the non-resonant
 1782 search.

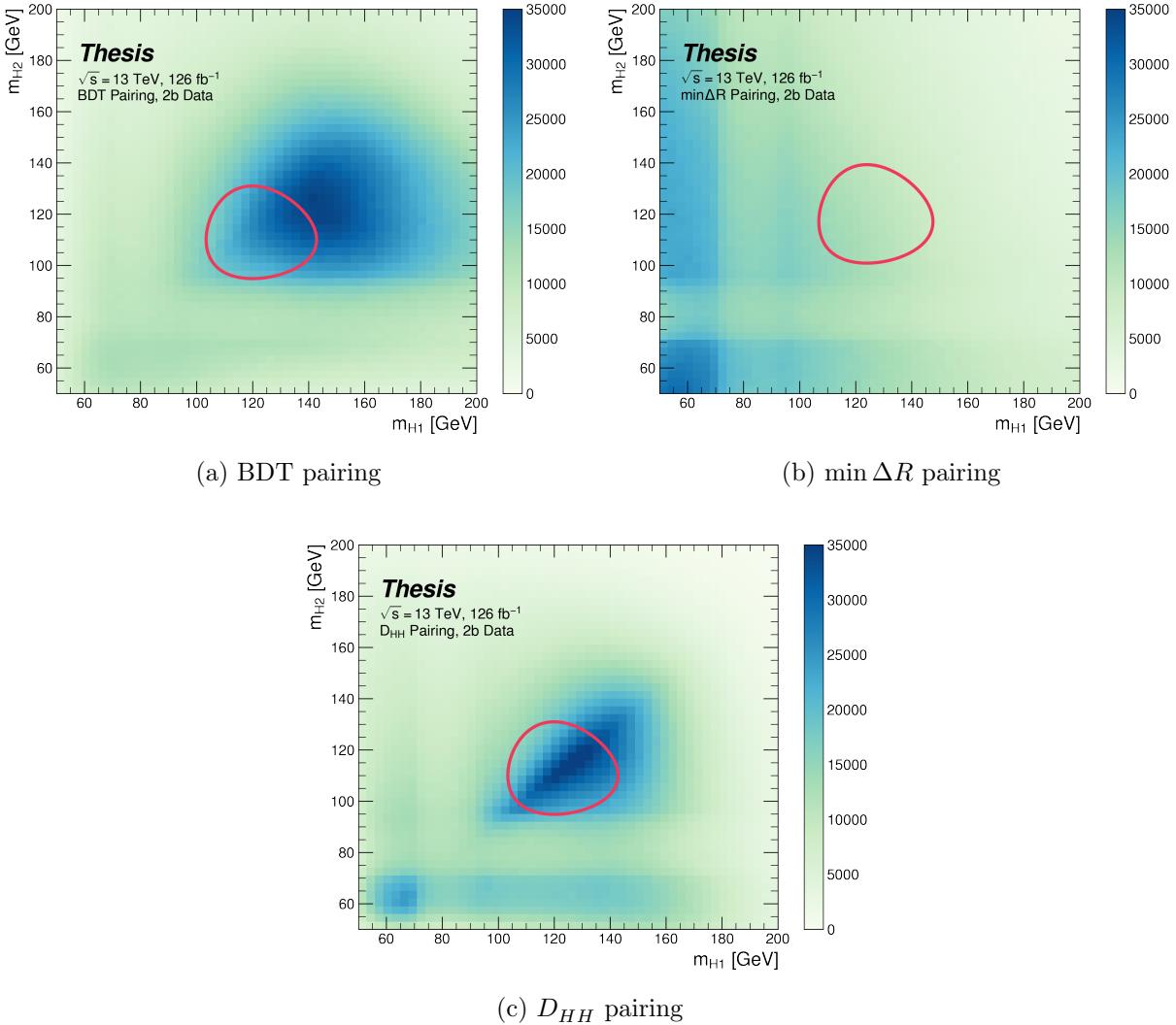


Figure 7.1: Comparison of m_{H1} vs m_{H2} planes for the full Run 2 2b dataset with different pairings. As evidenced, this choice significantly impacts where events fall in this plane, and therefore which events fall into the various kinematic regions defined in this plane (see Section 7.5). Respective signal regions are shown for reference, with the $\min \Delta R$ signal region shifted slightly up and to the right to match the non-resonant selection. Note that the band structure around 80 GeV in both m_{H1} and m_{H2} is introduced by the top veto described in Section 7.4.

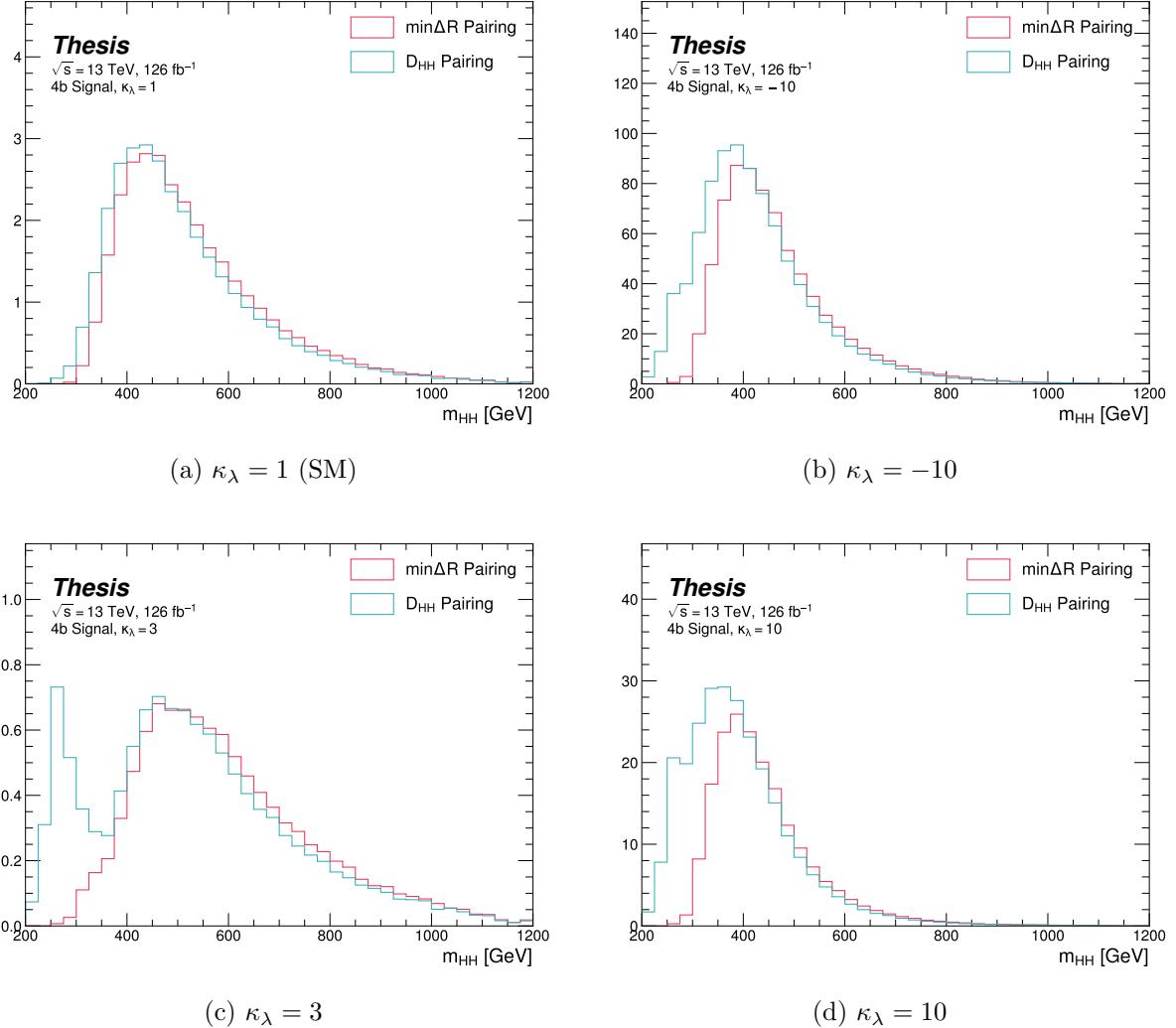


Figure 7.2: Comparison of signal distributions in the respective signal regions for the $\min \Delta R$ and D_{HH} pairing for various values of the Higgs trilinear coupling in the respective signal regions. The distributions are quite similar at the Standard Model point, but for other variations, $\min \Delta R$ does not pick up the low mass features.

1783 **7.4 Background Reduction and Top Veto**

1784 Choosing a pairing of the four b-tagged jets fully defines the di-Higgs candidate system used
1785 for each event in the remainder of the analysis chain. A requirement of $|\Delta\eta_{HH}| < 1.5$ on this
1786 di-Higgs candidate system mitigates QCD multijet background.

1787 Figure ?? illustrates this variable in the validation region (see Section ??). It demonstrates
1788 that this selection rejects only a small fraction of signal, but a significant fraction of data
1789 (which, in the validation region, is a good approximation of pure background).

1790 In order to mitigate the hadronic $t\bar{t}$ background, a top veto is then applied, removing
1791 events consistent with a $t \rightarrow b(W \rightarrow q_1\bar{q}_2)$ decay.

1792 The jets in the event are separated into *HC jets* which are the four jets used to build the
1793 Higgs candidates, and *non-*HC jets**, the other jets (passing the p_T and $|\eta|$ requirements) in
1794 the event.

1795 W candidates are built by forming all possible pairs of all jets in each event. With n jets,
1796 there are $\binom{n}{2}$ such pairs. t candidates are then built by pairing each W candidate with each
1797 HC jet (for $4\binom{n}{2}$ combinations). Note that all jets in a t candidate must be distinct (i.e. a
1798 HC jet may not be used both on its own and in a W candidate).

With m_t denoting the invariant mass of the t candidate, and m_W the invariant mass of the W candidate, the quantity

$$X_{Wt} = \sqrt{\left(\frac{m_W - 80.4 \text{ GeV}}{0.1 \cdot m_W}\right)^2 + \left(\frac{m_t - 172.5 \text{ GeV}}{0.1 \cdot m_t}\right)^2} \quad (7.1)$$

1799 is constructed for each combination.

1800 Events are then vetoed if the minimum X_{Wt} over all combinations is less than 1.5.

1801 The same definitions and procedures are used for both the resonant and non-resonant
1802 analyses. However, for the non-resonant search, the top candidates considered for X_{Wt} have
1803 the additional requirement that the jet used for the b is b -tagged. While this is identical to
1804 the resonant analysis by definition for $4b$ events, it does change the set of events considered in
1805 lower tag regions, in particular for the $2b$ events considered in the derivation of the background

1806 estimate. Such a change is found to reduce the impact of background systematics by increasing

1807 $2b$ support in the high X_{Wt} kinematic region. *TODO: Insert plot*

1808 **7.5 Kinematic Region Definition**

As has been mentioned, an important piece of the analysis is the plane defined by the two Higgs candidate masses (the *Higgs candidate mass plane*). After the selection described above, a signal region is defined by requiring $X_{HH} < 1.6$, where:

$$X_{HH} = \sqrt{\left(\frac{m(H_1) - c_1}{0.1 \cdot m(H_1)}\right)^2 + \left(\frac{m(H_2) - c_2}{0.1 \cdot m(H_2)}\right)^2} \quad (7.2)$$

1809 with $m(H_1)$, $m(H_2)$ the leading and subleading Higgs candidate masses, c_1 and c_2 correspond
1810 to the center of the signal region, and the denominator provides a Higgs candidate mass
1811 dependent resolution of 10 %. For consistency with the HH decay hypothesis, c_1 and c_2
1812 are nominally (125 GeV, 125 GeV). However, these are allowed to vary due to energy loss,
1813 with specific values chosen described below. The selection of these values is one of several
1814 significant differences between the regions defined for the resonant and non-resonant search.
1815 We describe both below.

1816 **7.5.1 Resonant Kinematic Regions**

1817 For the resonant analysis, the signal region is centered at (120 GeV, 110 GeV) to account for
1818 energy loss leading to the Higgs masses being under-reconstructed. *TODO: insert signal*
1819 *location plot?* Note that leading and subleading Higgs candidates are defined according to
1820 the *scalar sum* of constituent jet p_T .

For the background estimation, two regions are defined which are roughly concentric around the signal region: a *validation region* which consists of those events not in the signal region, but which do pass

$$\sqrt{(m(H_1) - 1.03 \times 120 \text{ GeV})^2 + (m(H_2) - 1.03 \times 110 \text{ GeV})^2} < 30 \text{ GeV} \quad (7.3)$$

and a *control region* whcih consists of those events not in the signal or validation regions, but which do pass

$$\sqrt{(m(H_1) - 1.05 \times 120 \text{ GeV})^2 + (m(H_2) - 1.05 \times 110 \text{ GeV})^2} < 45 \text{ GeV} \quad (7.4)$$

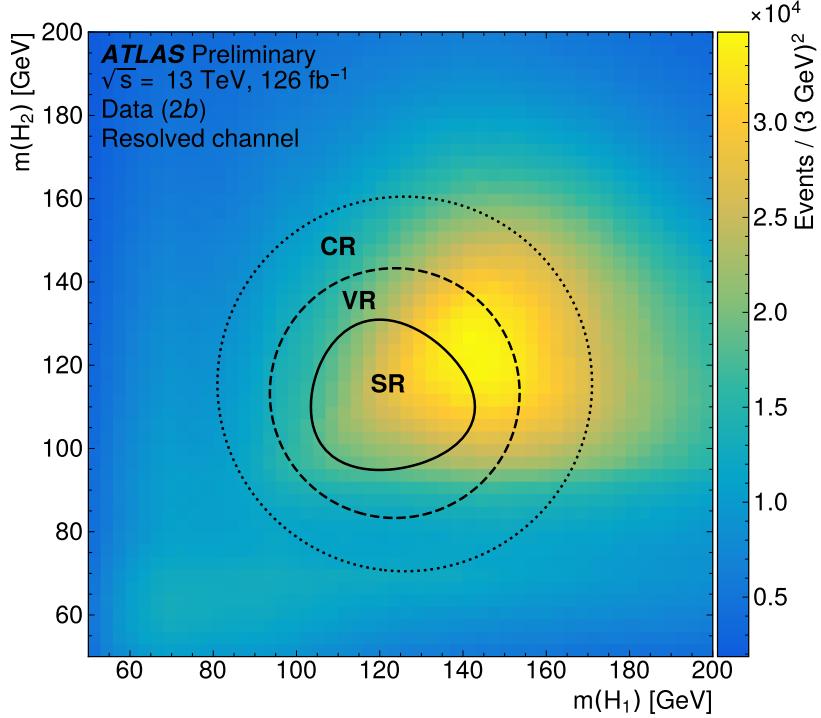


Figure 7.3: Regions used for the resonant search, shown on the $2b$ data mass plane. The outermost region (the “control region”) is used for derivation of the nominal background estimate. The innermost region is the signal region, where the signal extraction fit is performed. The region in between (the “validation region”) is used for the assessment of an uncertainty.

1821 For simplicity, the SR/VR/CR definitions from the early Run 2 paper [85] were chosen
 1822 for the resonant analysis, but were found to be close to optimal. These regions are shown in
 1823 Figure 7.3.

1824 7.5.2 Non-resonant Kinematic Regions

1825 For the non-resonant analysis the signal region is centered at $(124 \text{ GeV}, 117 \text{ GeV})$, corre-
 1826 sponding to the means of *correctly paired* Standard Model signal events. The shape of the
 1827 signal region (other than this change of center) was found to remain optimal.

1828 For the non-resonant search, leading and subleading Higgs candidates are defined according
 1829 to the *vector sum* of constituent jet p_T , more closely corresponding to the $1 \rightarrow 2$ decay
 1830 assumption behind the min ΔR pairing algorithm.

1831 Two areas for improvement were identified in the resonant analysis, which will be dis-
 1832 cussed in more detail below: *signal contamination* of the validation region (which impacts
 1833 the uncertainty assessed due to extrapolation) and *large nuisance parameter pulls* on this
 1834 uncertainty, corresponding to a rough assumption that the validation region is closer to the
 1835 signal region in the mass plane, and so offers a better estimate of the signal region.

To mitigate these two issues, a redesign of the control and validation regions was performed for the non-resonant analysis. The outer boundary defined by the shifted resonant control region:

$$\sqrt{(m(H_1) - 1.05 \times 124 \text{ GeV})^2 + (m(H_2) - 1.05 \times 117 \text{ GeV})^2} < 45 \text{ GeV} \quad (7.5)$$

1836 is kept, roughly corresponding to combining the regions used for the resonant analysis. In
 1837 order to assess the variation of the background estimate, two sets of regions are desired, so
 1838 this combined region is split into *quadrants*, that is, divided into four pieces along axes that
 1839 intersect with the signal region center. To avoid kinematic bias, quadrants on opposite sides
 1840 of the signal region are paired, with these pairs corresponding to the non-resonant control
 1841 and validation regions.

1842 The particular orientation of the regions is chosen such that region centers align with the
 1843 leading and subleading Higgs candidate masses, corresponding to a set of axes rotated at
 1844 45° , with the “top” and “bottom” quadrants together comprising the control region, and the
 1845 other set (“left” and “right”) the validation region. These regions are shown in Figure 7.4

1846 This design of regions includes a set of events closer to the signal region in the mass plane,
 1847 leveraging the assumption that these events are more similar to signal region events, while
 1848 also including events further away from the signal region, mitigating signal contamination.
 1849 This region selection is found to have good performance in alternate validation regions (see
 1850 Section 7.8).

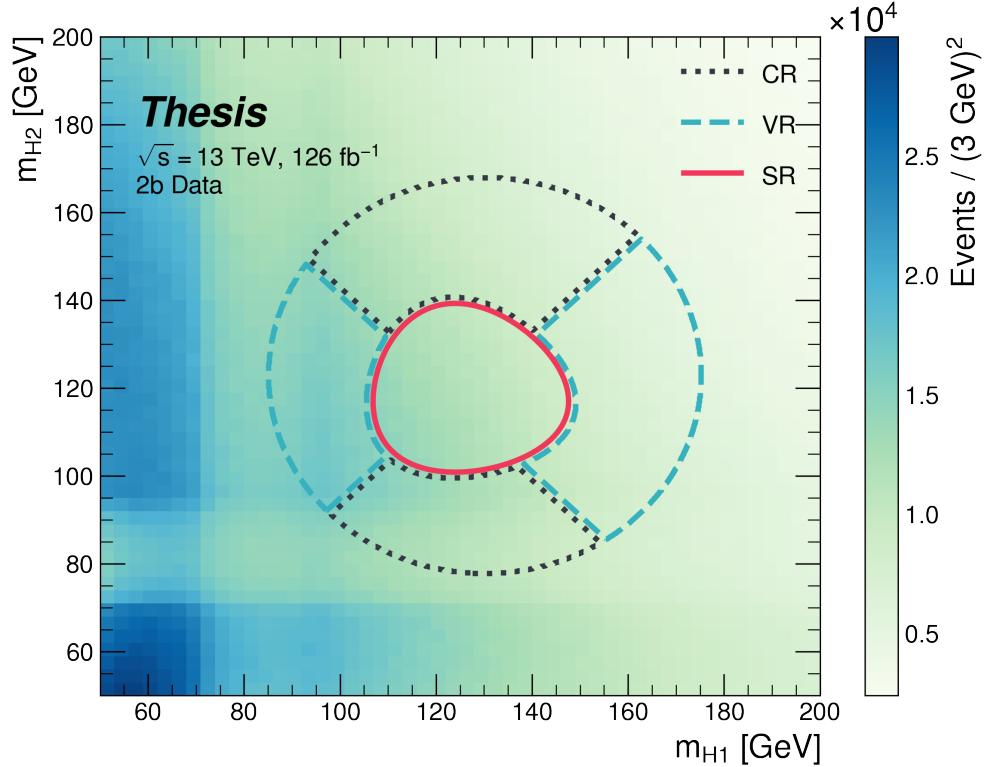


Figure 7.4: Regions used for the non-resonant search. The “top” and “bottom” quadrants together comprise the control region, in which the nominal background estimate is derived. The “left” and “right” quadrants together comprise the validation region, which is used to assess an uncertainty. The signal region, in the center, is where the signal extraction fit is performed.

1851 7.5.3 *Discriminating Variable*

1852 The discriminant used for the resonant analysis is *corrected* m_{HH} . This variable is calculated
 1853 by re-scaling the Higgs candidate four vectors such that each $m_H = 125 \text{ GeV}$. These re-scaled
 1854 four-vectors are then summed, and their invariant mass is the corrected m_{HH} . These re-scaled
 1855 four-vectors are not used for any other purpose. The effect of this correction, which sharpens
 the m_{HH} peak and correctly centres it, is shown in Figure 7.5.

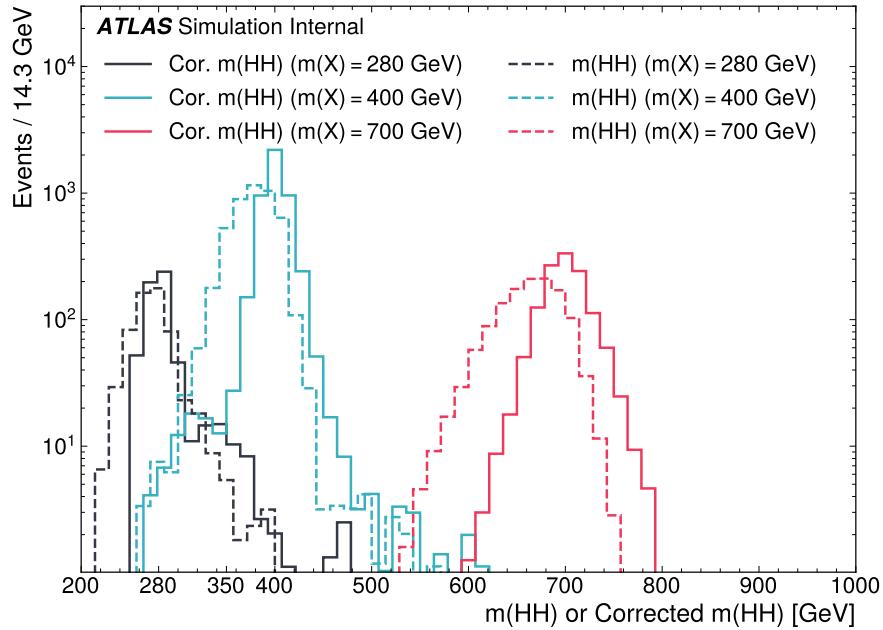


Figure 7.5: Impact of the m_{HH} correction on a range of spin-0 resonant signals. The corrected m_{HH} distributions (solid lines) are much sharper and more centered on the corresponding resonance masses than the uncorrected m_{HH} distributions (dashed).

1856

1857 For the non-resonant analysis, due to the broad nature of the signal in m_{HH} , such a
 1858 correction is not as motivated, and, indeed, is found to have very minimal impact. The
 1859 uncorrected m_{HH} (just referred to as m_{HH}) is therefore used as a discriminant. To maximize

1860 sensitivity, the non-resonant analysis additionally uses two variables for categorization: $\Delta\eta_{HH}$,
1861 an angular variable which, along with m_{HH} , fully characterizes the HH system [86], and X_{HH} ,
1862 the variable used for the signal region definition, which leverages the peaked structure of the
1863 signal in the $(m(H_1), m(H_2))$ plane to split the signal extraction fit into lower and higher
1864 purity regions (highest purity near $X_{HH} = 0$, the center of the signal region). Distributions
1865 of these variables are shown in *TODO: plots*. The categorization used for this thesis has been
1866 optimized to be 2×2 in these variables, with corresponding selections $0 \leq \Delta\eta_{HH} \leq 0.75$ and
1867 $0.75 \leq \Delta\eta_{HH} \leq 1.5$ for $\Delta\eta_{HH}$, and $0 \leq X_{HH} \leq 0.95$ and $0.95 \leq X_{HH} \leq 1.6$ for X_{HH} .

1868 **7.6 Background Estimation**

1869 After the event selection described above there are two major backgrounds, QCD and $t\bar{t}$.
1870 A very similar approach is used for both the resonant and the non-resonant analyses, with
1871 some small modifications. This approach is notably fully data-driven, which is warranted due
1872 to the flexibility of the estimation method, as well as the high relative proportion of QCD
1873 background ($> 90\%$), and allows for the use of machine learning methods in the construction
1874 of the background estimate. However, it sacrifices an explicit treatment of the $t\bar{t}$ component.
1875 Performance of the background estimate on the $t\bar{t}$ component is checked explicitly *TODO:*
1876 *add plots*, and minimal impact due to $t\bar{t}$ mismodeling is seen.

1877 Contributions of single Higgs processes and ZZ are found to be negligible, and the
1878 Standard Model HH background is found to have no impact on the resonant search.

1879 The foundation of the background estimate lies in the derivation of a reweighting function
1880 which matches the kinematics of events with exactly two b -tagged jets to those of events in
1881 the higher tagged regions (events with three or four b -tagged jets). The reweighting function
1882 and overall normalization are derived in the control region. Systematic bias of this estimate
1883 is assessed in the validation region.

1884 For the resonant analysis, the systematic bias is a bias due to extrapolation: the validation
1885 region lies between the control and signal regions. For the non-resonant analysis, the bias
1886 instead comes from different possible interpolations of the signal region kinematics – given the
1887 choice of nominal estimate, the validation region is a conceptually equivalent, but maximally
1888 different, signal region estimate.

1889 **7.6.1 The Two Tag Region**

1890 Events in data with exactly two b -tagged jets are used for the data driven background
1891 estimate. The hypothesis here is that, due to the presence of multiple b -tagged jets, the
1892 kinematics of such events are similar to the kinematics of events in higher b -tagged regions (i.e.
1893 events with three and four b -tagged jets, respectively), and any differences can be corrected

1894 by a reweighting procedure. The region with three b -tagged jets is split into two b -tagging
 1895 regions, with the $3b + 1$ loose region used as an additional signal region (see Section *TODO:*
 1896 *Add ref*). The lower tagged $3b$ component ($3b + 1$ fail, as described in Section ??) is reserved
 1897 for validation of the background modelling procedure. Events with fewer than two b -tagged
 1898 jets are not used for this analysis, as they are relatively more different from the higher tag
 1899 regions.

1900 The nominal event selection requires at least four jets in order to form Higgs candidates.
 1901 For the four tag region, these are the four highest p_T b -tagged jets. For the three tag regions,
 1902 these jets are the three b -tagged jets, plus the highest p_T jet satisfying a loosened b -tagging
 1903 requirement. Similarly, and following the approach of the resonant analysis, the two tag region
 1904 uses the two b -tagged jets and the two highest p_T non-tagged jets to form Higgs candidates.
 1905 Combinatoric bias from selection of different numbers of b -tagged jets is corrected as a part
 1906 of the kinematic reweighting procedure through the reweighting of the total number of jets in
 1907 the event. In this way, the full event selection may be run on two tagged events.

1908 7.6.2 Kinematic Reweighting

1909 The set of two tagged data events is the fundamental piece of the data driven background
 1910 estimate. However, kinematic differences from the four tag region exist and must be corrected
 1911 in order for this estimate to be useful. Binned approaches based on ratios of histograms
 1912 have been previously considered [85], [17], but are limited in their handling of correlations
 1913 between variables and by the “curse of dimensionality”, i.e. the dataset becomes sparser and
 1914 sparser in “reweighting space” as the number of dimensions in which to reweight increases,
 1915 limiting the number of variables used for reweighting. This leads either to an unstable fit
 1916 result (overfitting with finely grained bins) or a lower quality fit result (underfitting with
 1917 coarse bins).

1918 Note that even machine learning methods such as Boosted Decision Trees (BDTs), may
 1919 suffer from this curse of dimensionality, as the depth of each decision tree used is limited
 1920 by the available statistics after each set of corresponding selections (cf. binning in a more

1921 sophisticated way), limiting the expressivity of the learned reweighting function.

1922 To solve these issues, a neural network based reweighting procedure is used here. This
1923 is a truly multivariate approach, allowing for proper treatment of variable correlations. It
1924 further overcomes the issues associated with binned approaches by learning the reweighting
1925 function directly, allowing for greater sensitivity to local differences and helping to avoid the
1926 curse of dimensionality.

1927 *Neural Network Reweighting*

Let $p_{4b}(x)$ and $p_{2b}(x)$ be the probability density functions for four and two tag data respectively across some input variables x . The problem of learning the reweighting function between two and four tag data is then the problem of learning a function $w(x)$ such that

$$p_{2b}(x) \cdot w(x) = p_{4b}(x) \quad (7.6)$$

from which it follows that

$$w(x) = \frac{p_{4b}(x)}{p_{2b}(x)}. \quad (7.7)$$

This falls into the domain of density ratio estimation, for which there are a variety of approaches. The method considered here is modified from [87, 88], and depends on a loss function of the form

$$\mathcal{L}(R(x)) = \mathbb{E}_{x \sim p_{2b}}[\sqrt{R(x)}] + \mathbb{E}_{x \sim p_{4b}}\left[\frac{1}{\sqrt{R(x)}}\right]. \quad (7.8)$$

where $R(x)$ is some estimator dependent on x and $\mathbb{E}_{x \sim p_{2b}}$ and $\mathbb{E}_{x \sim p_{4b}}$ are the expectation values with respect to the 2b and 4b probability densities. A neural network trained with such a loss function has the objective of finding the estimator, $R(x)$, that minimizes this loss. It is straightforward to show (Appendix ??) that

$$\arg \min_R \mathcal{L}(R(x)) = \frac{p_{4b}(x)}{p_{2b}(x)} \quad (7.9)$$

1928 which is exactly the form of the desired reweighting function.

In practice, to avoid imposing explicit positivity constraints, the substitution $Q(x) \equiv \log R(x)$ is made. The loss function then takes the equivalent form

$$\mathcal{L}(Q(x)) = \mathbb{E}_{x \sim p_{2b}}[\sqrt{e^{Q(x)}}] + \mathbb{E}_{x \sim p_{4b}}\left[\frac{1}{\sqrt{e^{Q(x)}}}\right], \quad (7.10)$$

with solution

$$\arg \min_Q \mathcal{L}(Q(x)) = \log \frac{p_{4b}(x)}{p_{2b}(x)}. \quad (7.11)$$

1929 Taking the exponent then results in the desired reweighting function.

1930 Note that similar methods for density ratio estimation are available *TODO: cite*, e.g. from

1931 a more standard binary cross-entropy loss. However, these were found to perform no better
1932 than the formulation presented here.

1933 *Variables and Results*

1934 The neural network is trained on a variety of variables sensitive to two vs. four tag differences.

1935 To help bring out these differences, the natural logarithm of some of the variables with a
1936 large, local change is taken. The set of training variables used for the resonant analysis is

1937 1. $\log(p_T)$ of the 4th leading Higgs candidate jet

1938 2. $\log(p_T)$ of the 2nd leading Higgs candidate jet

1939 3. $\log(\Delta R)$ between the closest two Higgs candidate jets

1940 4. $\log(\Delta R)$ between the other two Higgs candidate jets

1941 5. Average absolute value of Higgs candidate jet η

1942 6. $\log(p_T)$ of the di-Higgs system.

1943 7. ΔR between the two Higgs candidates

1944 8. $\Delta\phi$ between the jets in the leading Higgs candidate

1945 9. $\Delta\phi$ between the jets in the subleading Higgs candidate

1946 10. $\log(X_{Wt})$, where X_{Wt} is the variable used for the top veto

1947 11. Number of jets in the event.

1948 The non-resonant analysis uses an identical set of variables with two notable changes

1949 1. The definition of X_{Wt} differs from the resonant definition (as described in Section
1950 *TODO: ref*)

1951 2. An integer encoding of the two trigger categories is used as an input (variable which
1952 takes on the value 0 or 1 corresponding to each of the two categories). This was found
1953 to improve a mismodeling near the tradeoff in m_{HH} of the two buckets.

1954 The neural network used for both resonant and non-resonant reweighting has three densely
1955 connected hidden layers of 50 nodes each with ReLU activation functions and a single node
1956 linear output. This configuration demonstrates good performance in the modelling of a variety
1957 of relevant variables, including m_{HH} , when compared to a range of networks of similar size.

1958 In practice, a given training of the reweighting neural network is subject to variation
1959 due to training statistics and initial conditions. An uncertainty is assigned to account for
1960 this (Section 7.7), which relies on training an ensemble of reweighting networks [89]. To
1961 increase the stability of the background estimate, the median of the predicted weight for each
1962 event is calculated across the ensemble. This median is then used as the nominal background
1963 estimate. This approach is indeed seen to be much more stable and to demonstrate a better
1964 overall performance than a single arbitrary training. Each ensemble used for this analysis
1965 consists of 100 neural networks, trained as described in Section 7.7.

1966 The training of the ensemble used for the nominal estimate is done in the kinematic
1967 Control Region. The prediction of these networks in the Signal Region is then used for the
1968 nominal background estimate. In addition, a separate ensemble of networks is trained in the

1969 Validation Region. The difference between the prediction of the nominal estimate and the
 1970 estimate from the VR derived networks in the Signal Region is used to assign a systematic
 1971 uncertainty. Further details on this systematic uncertainty are shown in Section 7.7. Note
 1972 that although the same procedure is used for both Control and Validation Region trained
 1973 networks, only the median estimate from the VR derived reweighting is used for assessing a
 1974 systematic – no additional “uncertainty on the uncertainty” from VR ensemble variation is
 1975 applied.

1976 Each reweighted estimate is normalized such that the reweighted $2b$ yield matches the $4b$
 1977 yield in the corresponding training region. Note that this applies to each of the networks used
 1978 in each ensemble, where the normalization factor is also subject to the procedure described in
 1979 Section 7.7. As the median over these normalized weights is not guaranteed to preserve this
 1980 normalization, a further correction is applied such that the $2b$ yield, after the median weights
 1981 are applied, matches the $4b$ yield in the corresponding training region. As no preprocessing
 1982 is applied to correct for the class imbalance between $2b$ and $4b$ events entering the training,
 1983 this ratio of number of $4b$ events ($n(4b)$) over number of $2b$ events ($n(2b)$) is folded into the
 1984 learned weights. Correspondingly, the set of normalization factors described above is near 1
 1985 and the learned weights are centered around $n(4b)/n(2b)$ (roughly 0.01 over the full dataset).
 1986 This normalization procedure applies for all instances of the reweighting (e.g. those used for
 1987 validations in Section ??), with appropriate substitutions of reweighting origin (here $2b$) and
 1988 reweighting target (here $4b$).

1989 Note that, due to different trigger and pileup selections during each year, the reweighting
 1990 is trained on each year separately. An approach of training all of the years together with a
 1991 one-hot encoding was explored *TODO: reference study*, but was found to have minimal benefit
 1992 over the split years approach, and in fact to increase the systematic bias of the corresponding
 1993 background estimate. Because of this, and because trigger selections for each year significantly
 1994 impact the kinematics of each year, such that categorizing by year is expected to reflect
 1995 groupings of kinematically similar events and to provide a meaningful degree of freedom in
 1996 the signal extraction fit, the split-year approach is kept.

1997 The control region closure for the 2018 dataset is shown for the resonant search in Figures
 1998 [7.6](#) through [7.14](#) and for the non-resonant search in Figures [7.24](#) through [7.32](#) for $4b$ and
 1999 Figures [7.42](#) through [7.50](#) for $3b1l$. The impact of this control region derived reweighting
 2000 on the validation region is shown in Figures [7.15](#) through [7.23](#) for the resonant search and
 2001 Figures [7.33](#) through [7.41](#) for $4b$ and Figures [7.51](#) through [7.59](#) for $3b1l$ for the non-resonant
 2002 search. Generally good performance is seen, with some occasional mis-modeling. For the
 2003 resonant search, this is most notable in the case of individual jet p_T . Such mis-modeling
 2004 may be corrected by including the variables in the input set, but this was found to not
 2005 improve the modeling of m_{HH} , and so is not done here. This mis-modeling is notable for the
 2006 non-resonant search in the leading Higgs candidate jet p_T , and is a direct consequence of the
 2007 trigger category input, which improves modeling of m_{HH} . Results are similar for other years,
 2008 but are not included here for brevity.

2009 One other salient feature of the non-resonant plots is the distributions of m_{H1} and m_{H2} ,
 2010 which emphasize the quadrant region definitions – the control region has a peak around
 2011 125 GeV in m_{H1} , which may be thought of as “signal region-like”, motivating this alignment,
 2012 though consequently the distribution of m_{H2} is quite bimodal. The reverse is true for the
 2013 validation region.

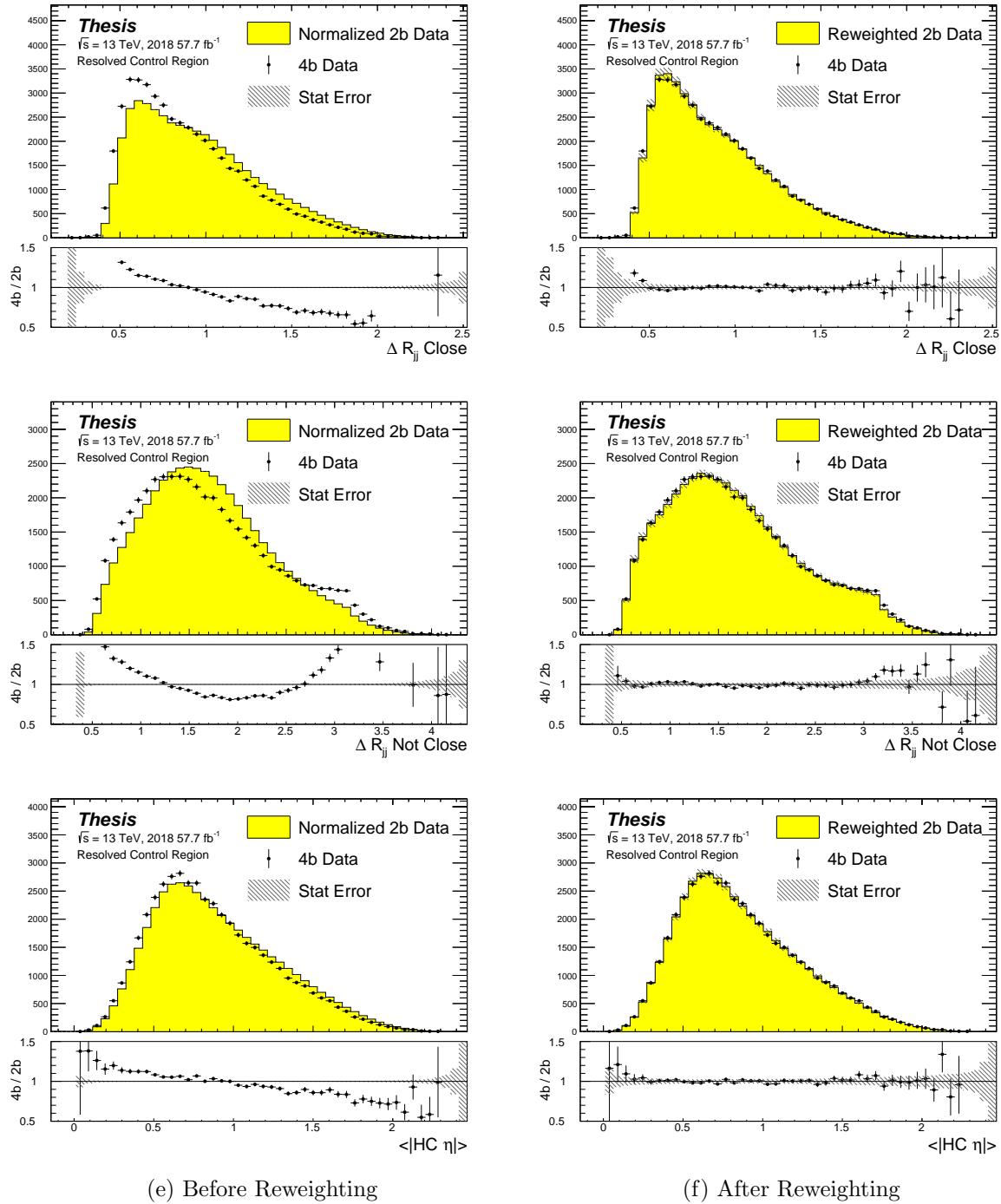


Figure 7.6: **Resonant Search:** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 Control Region.

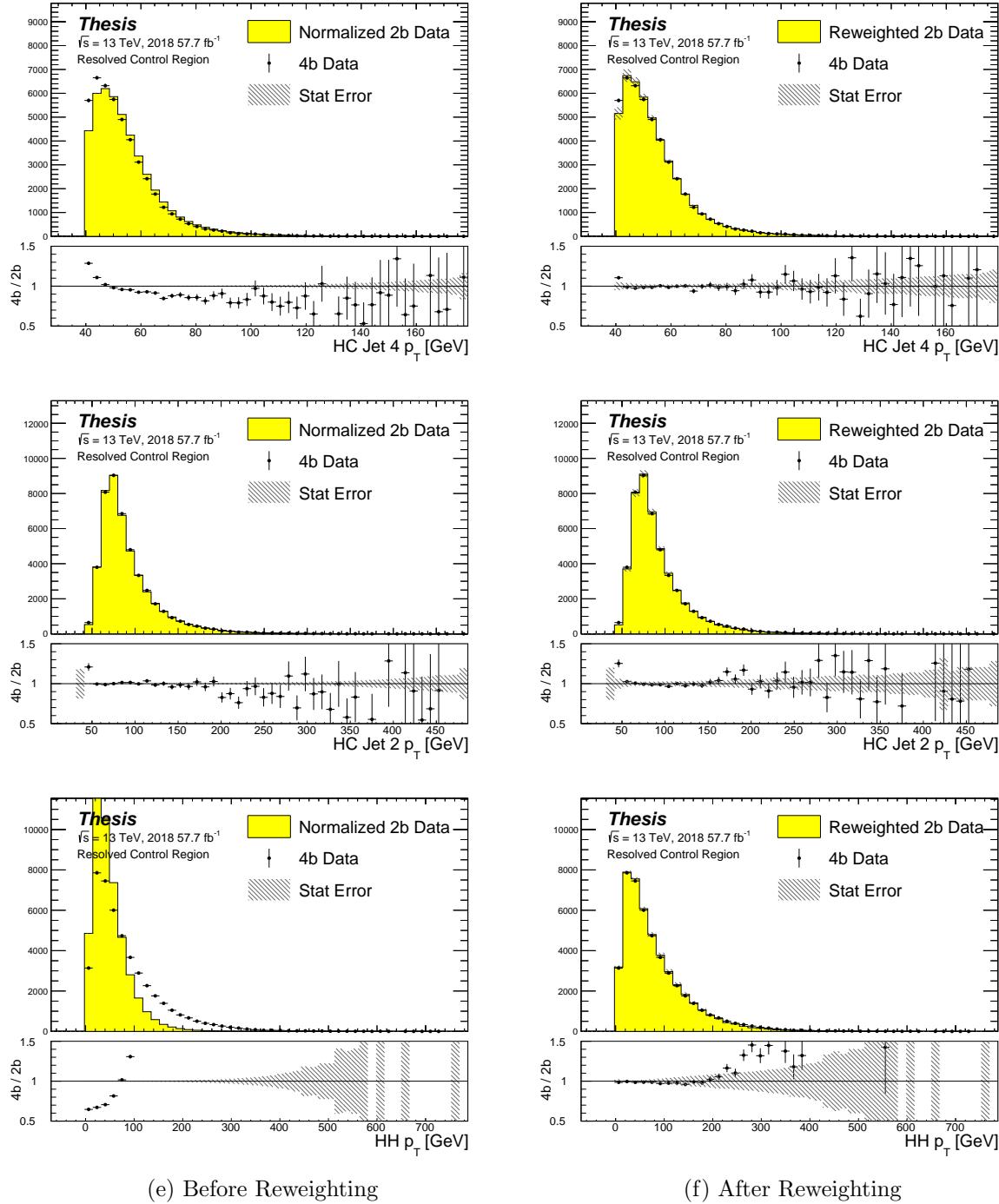


Figure 7.7: **Resonant Search:** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.

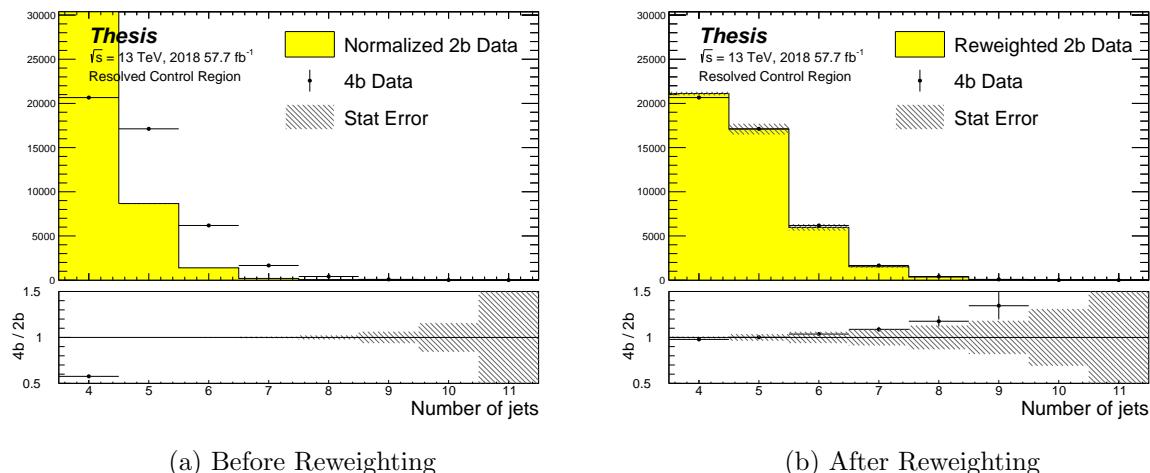


Figure 7.8: **Resonant Search:** Distributions of the number of jets before and after CR derived reweighting for the 2018 Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

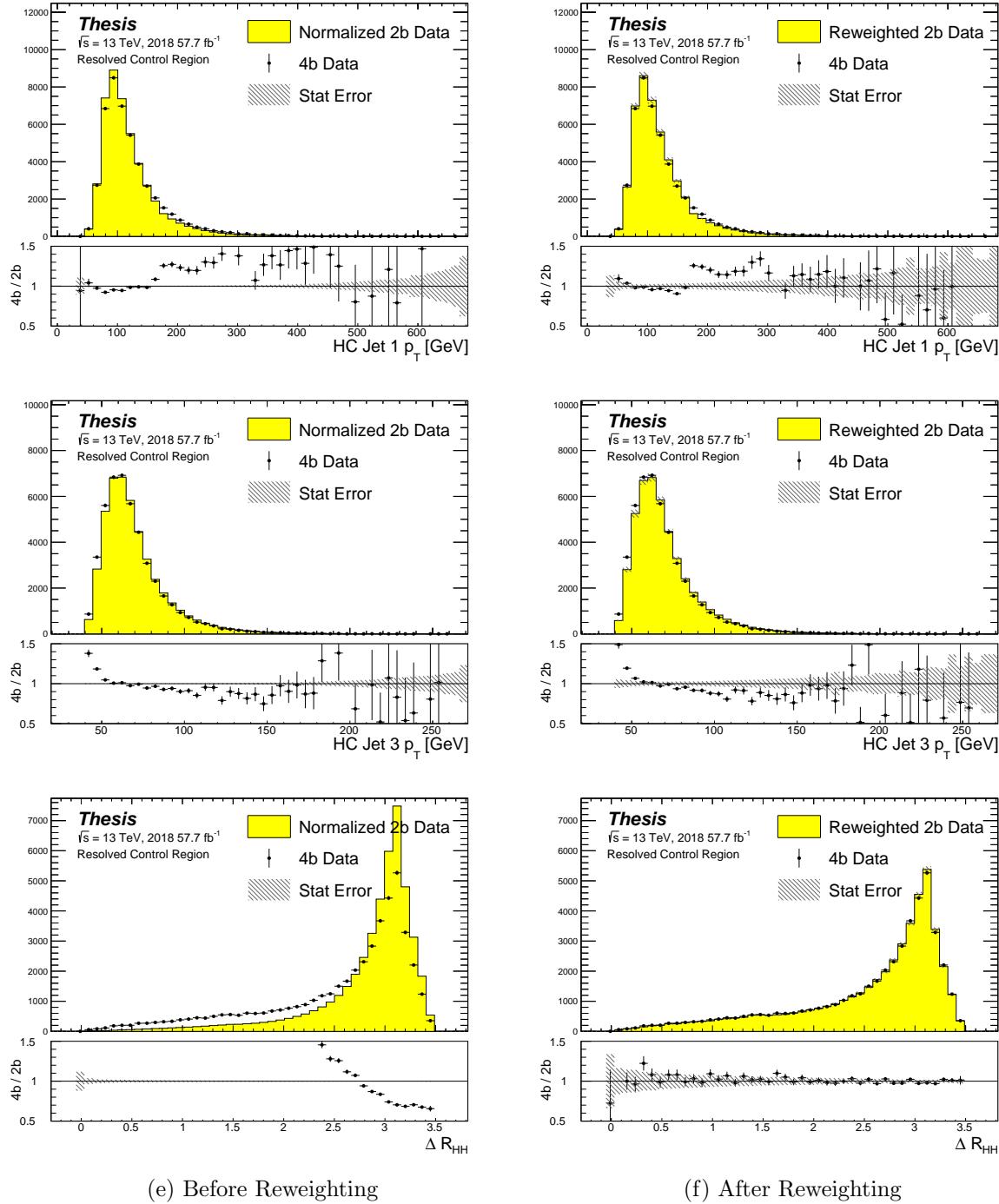


Figure 7.9: **Resonant Search:** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 Control Region.

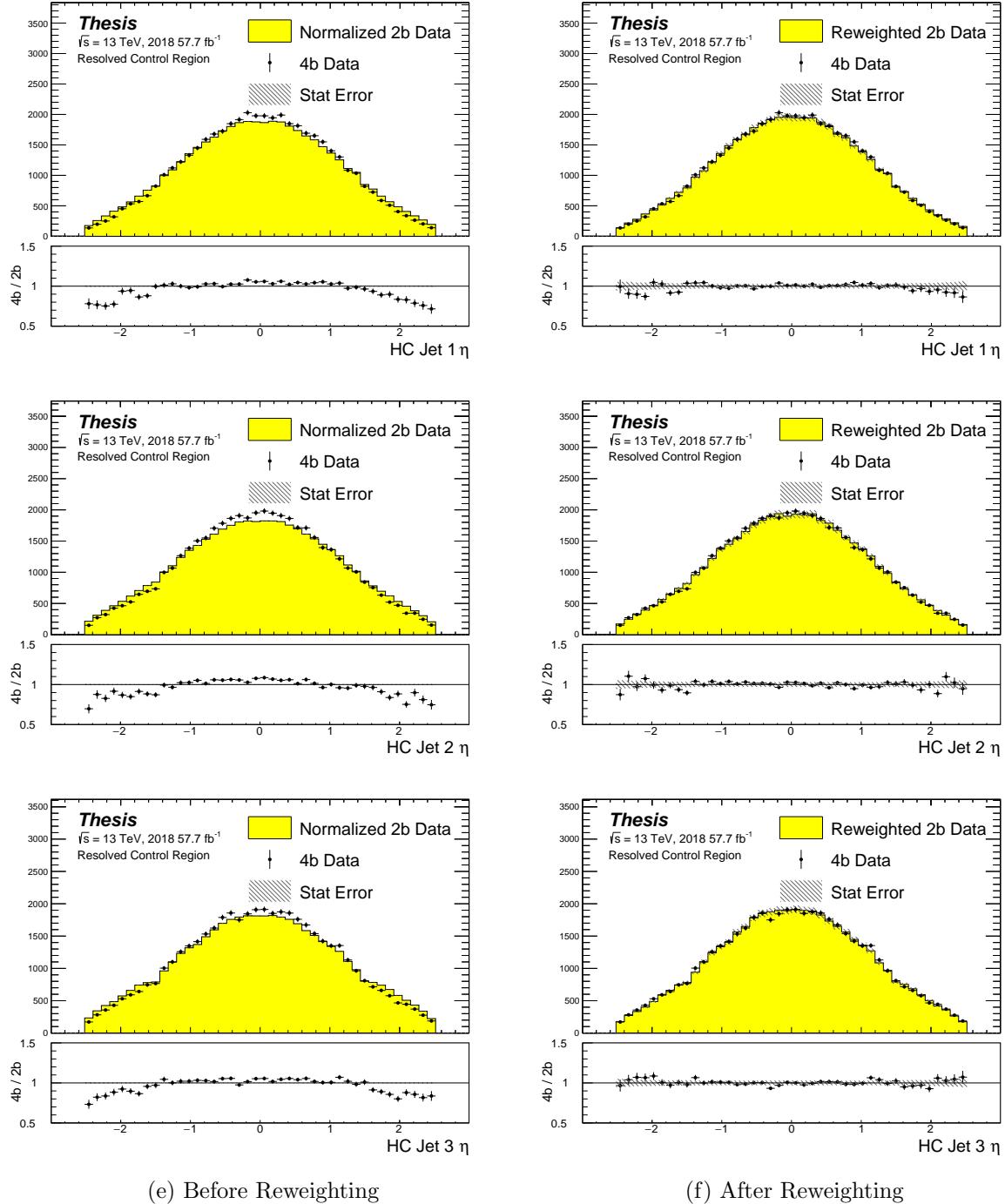


Figure 7.10: **Resonant Search:** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Control Region.

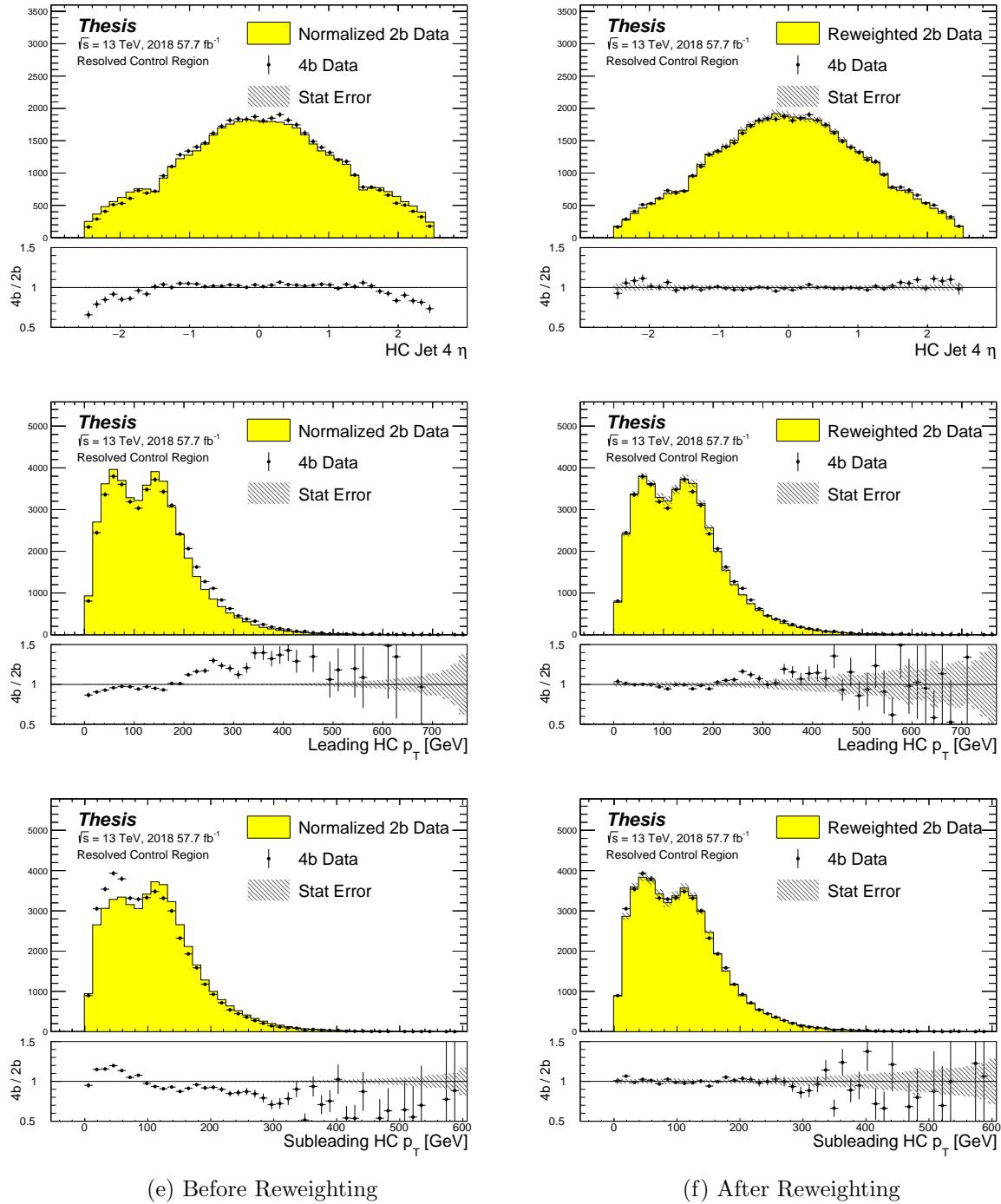


Figure 7.11: **Resonant Search:** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.

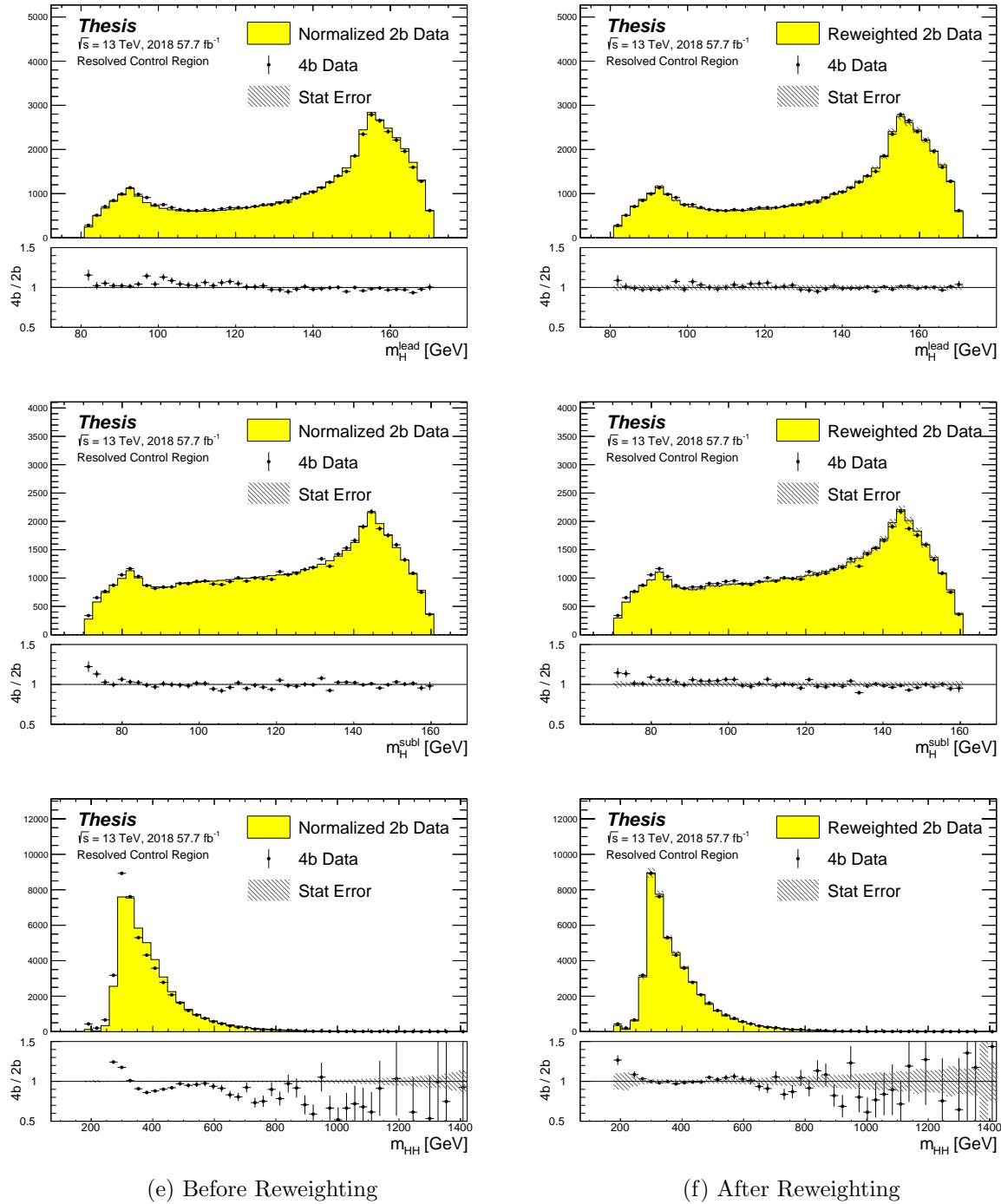


Figure 7.12: **Resonant Search:** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.

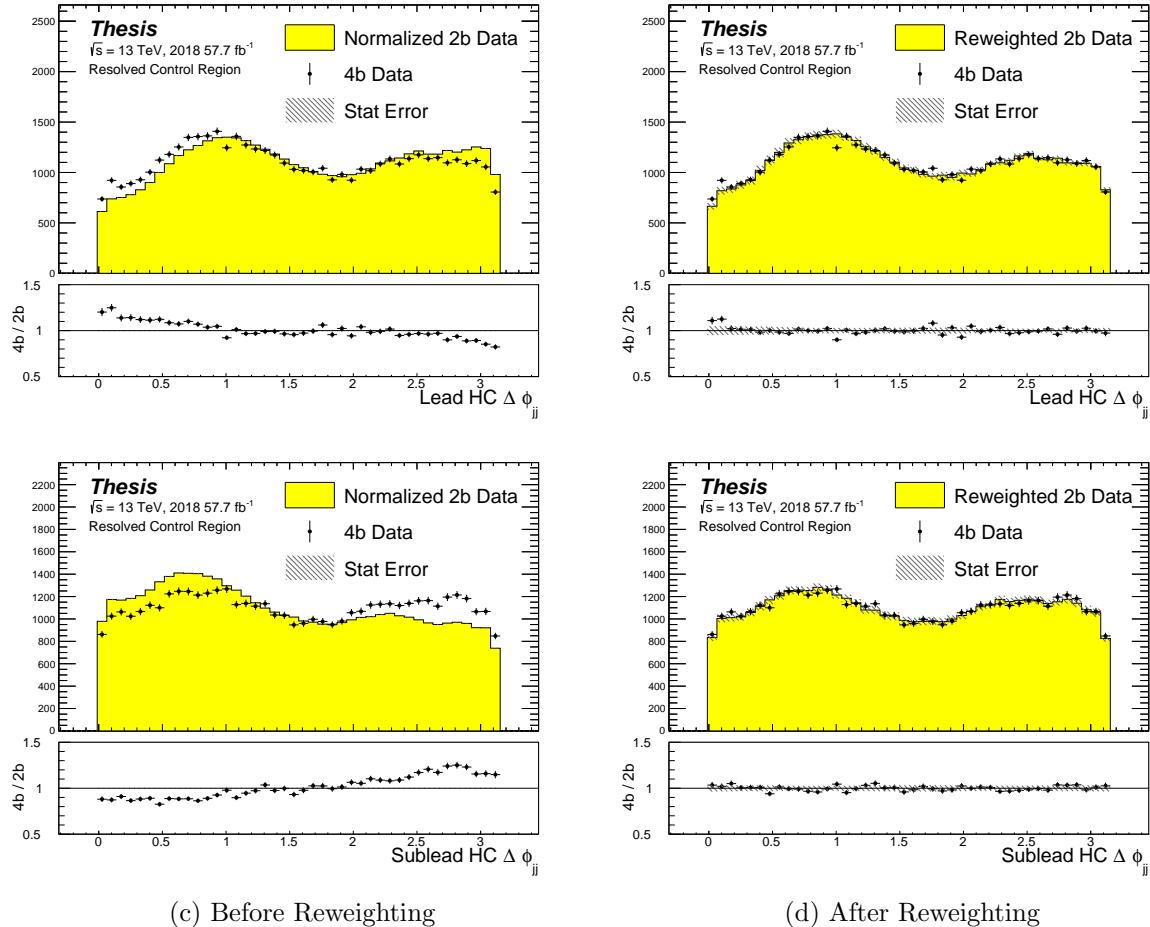


Figure 7.13: **Resonant Search:** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.

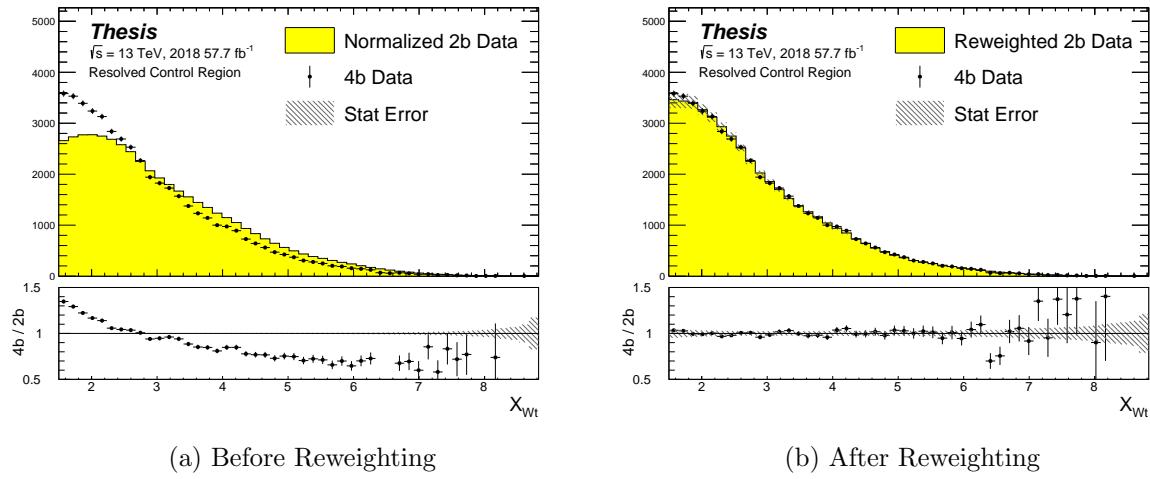


Figure 7.14: **Resonant Search:** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 Control Region. Reweighting is done after the cut on this variable is applied

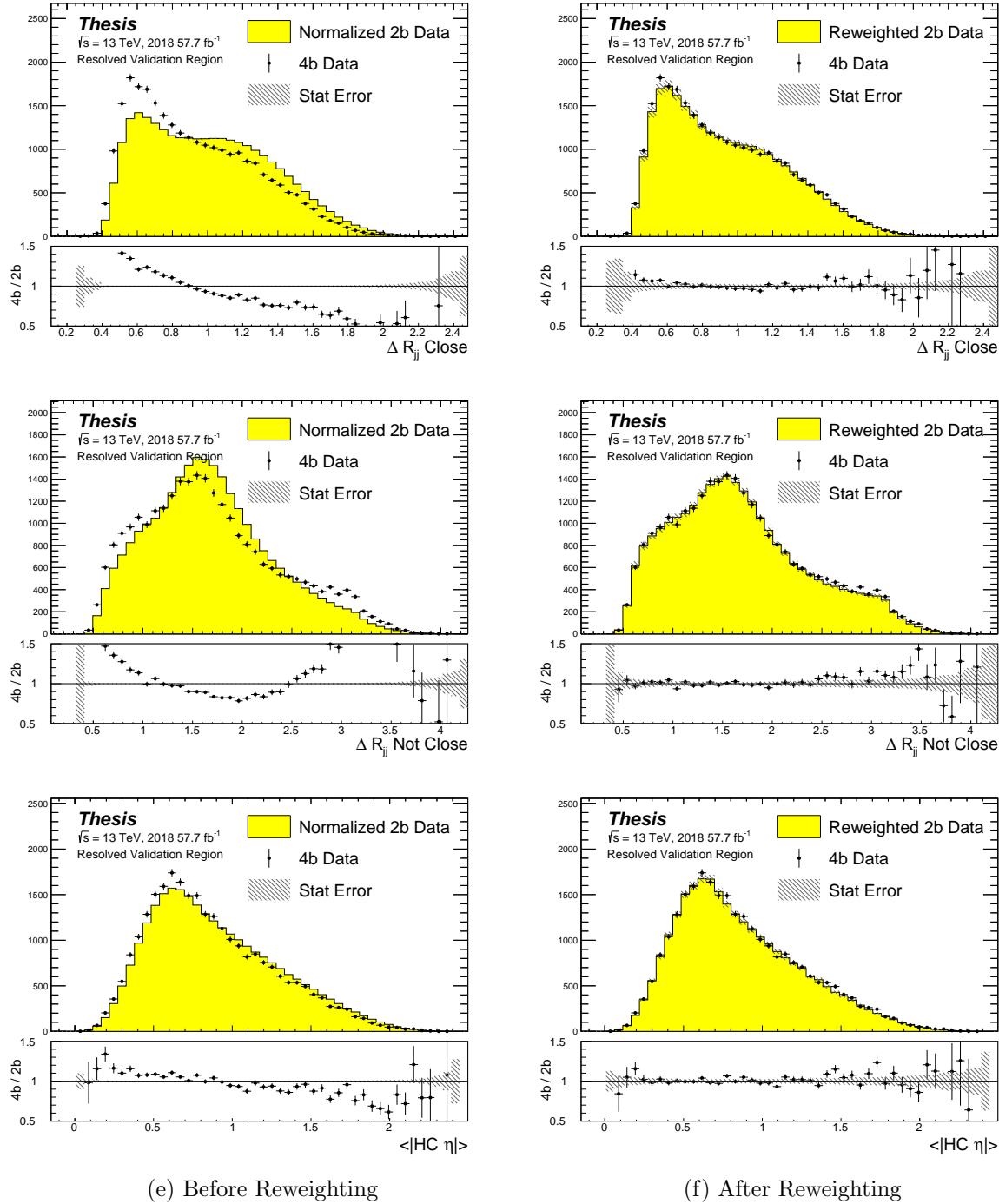


Figure 7.15: **Resonant Search:** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 Validation Region.

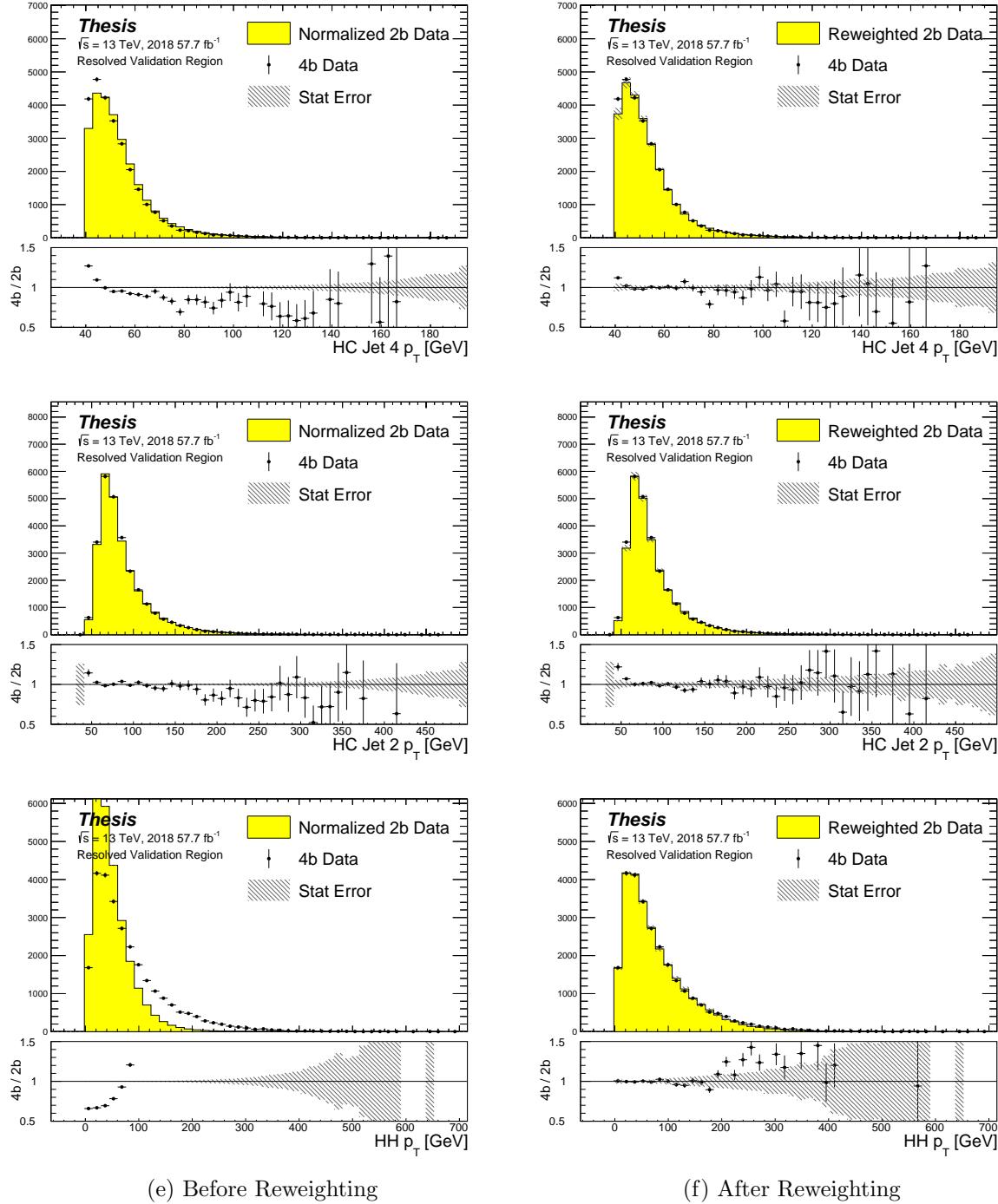


Figure 7.16: **Resonant Search:** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.

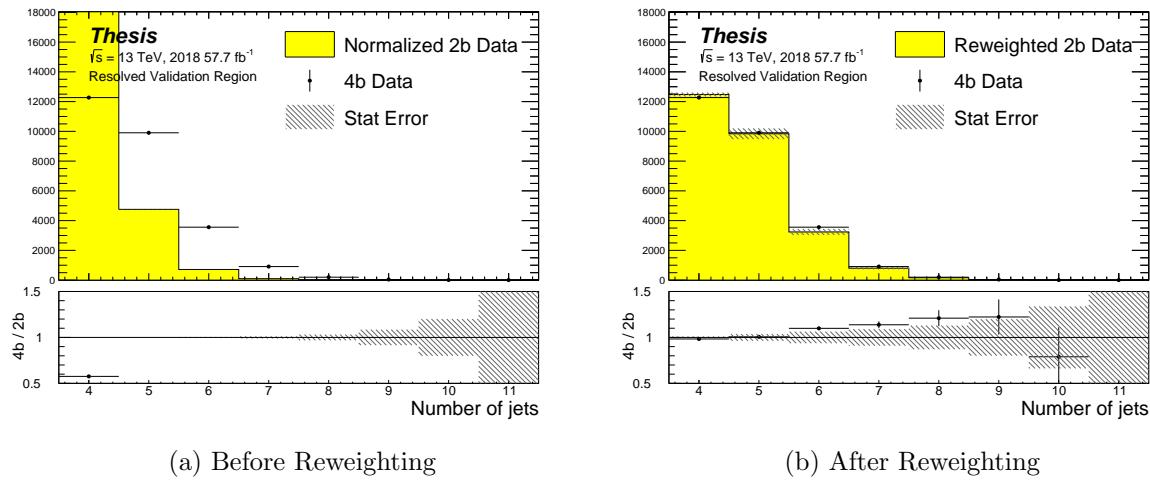


Figure 7.17: **Resonant Search:** Distributions of the number of jets before and after CR derived reweighting for the 2018 Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

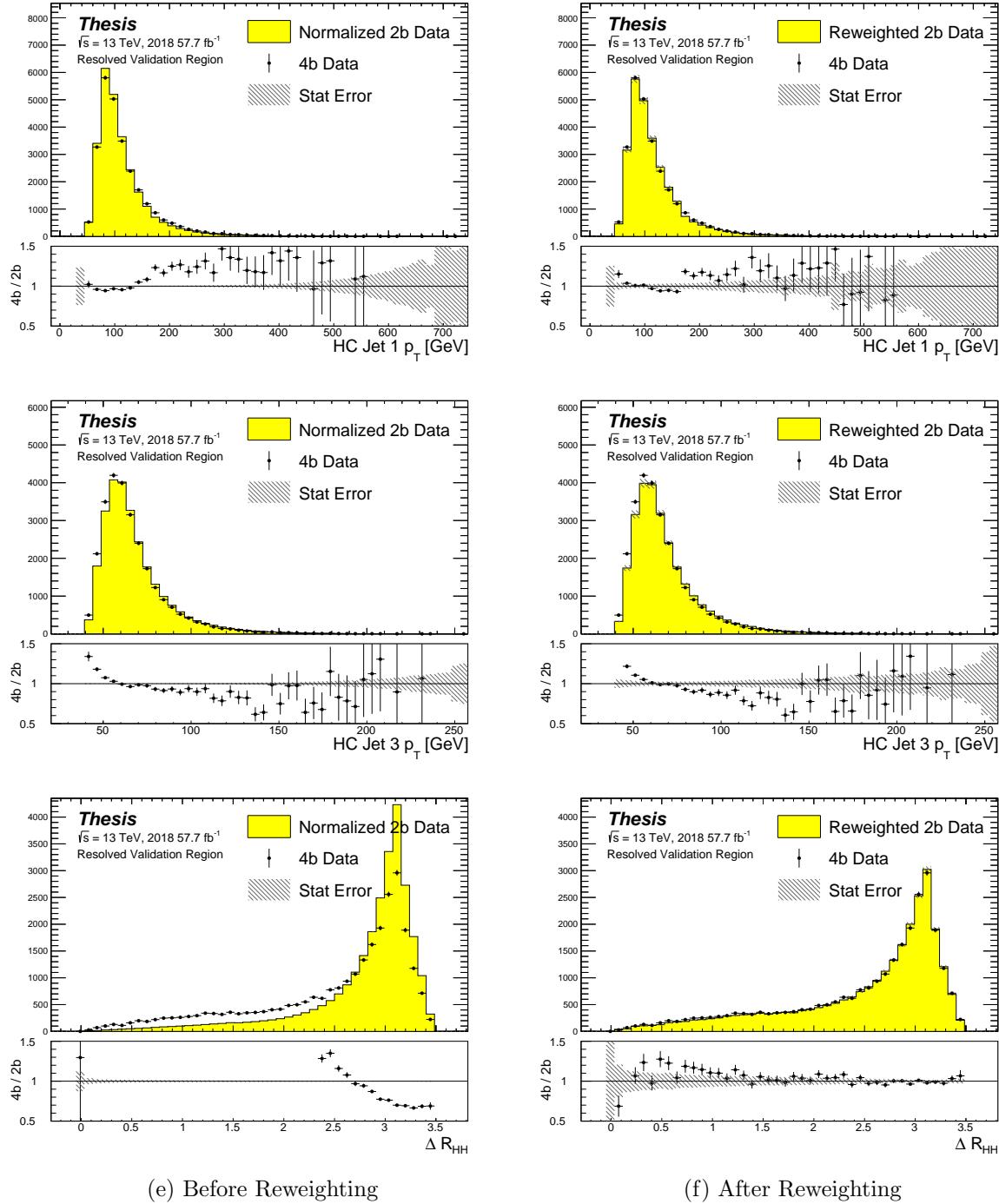


Figure 7.18: **Resonant Search:** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.

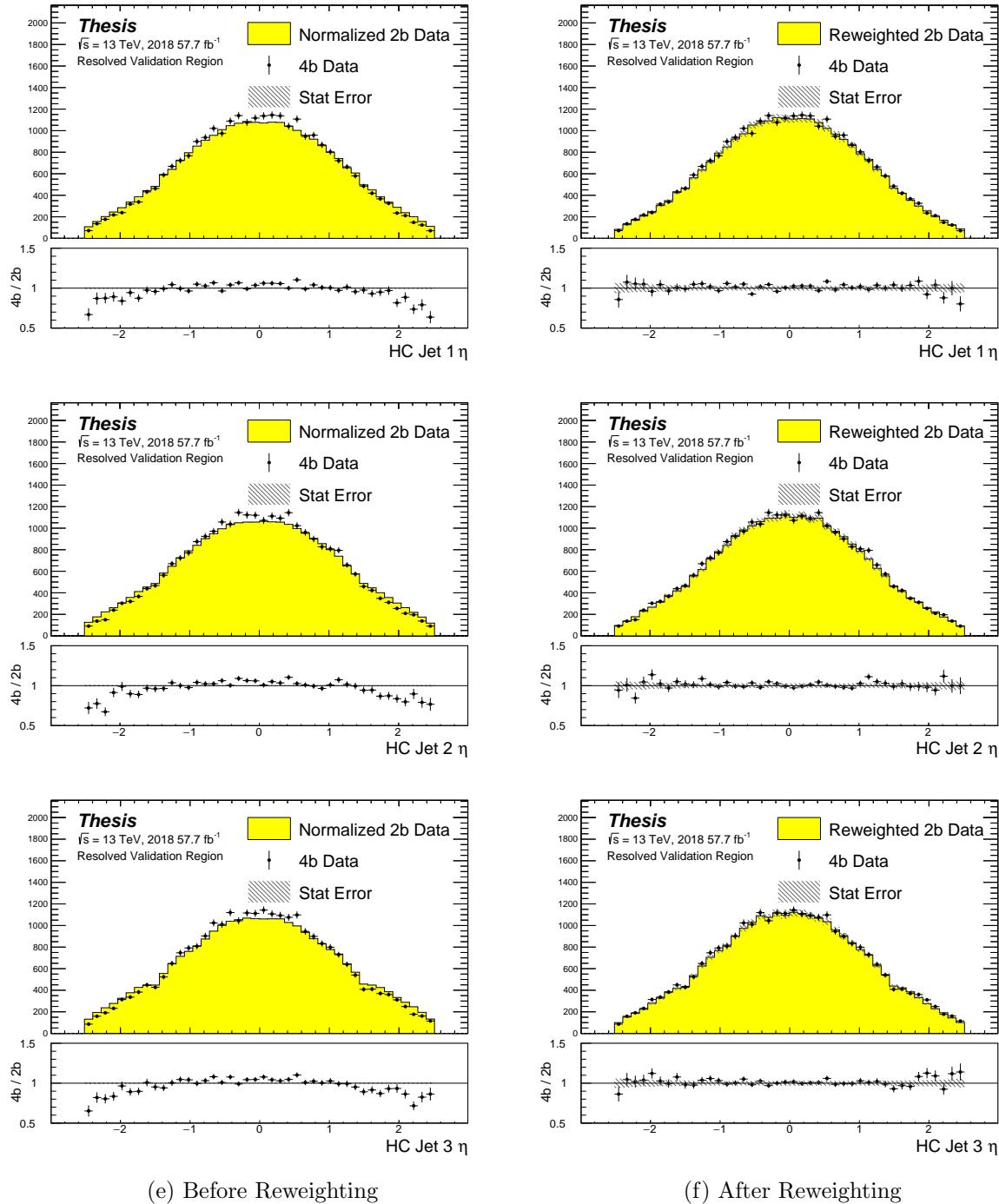


Figure 7.19: **Resonant Search:** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Validation Region.

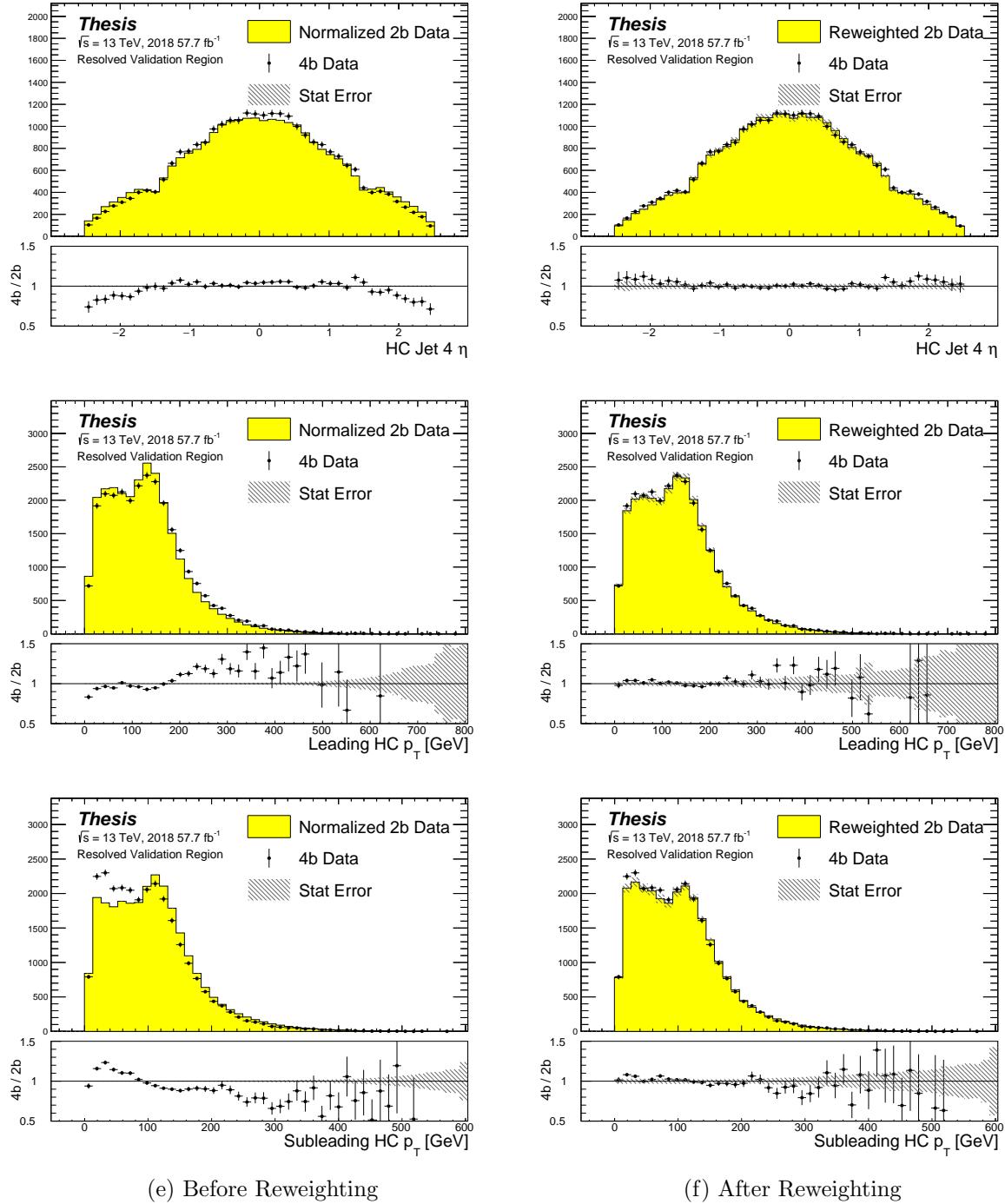


Figure 7.20: **Resonant Search:** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.

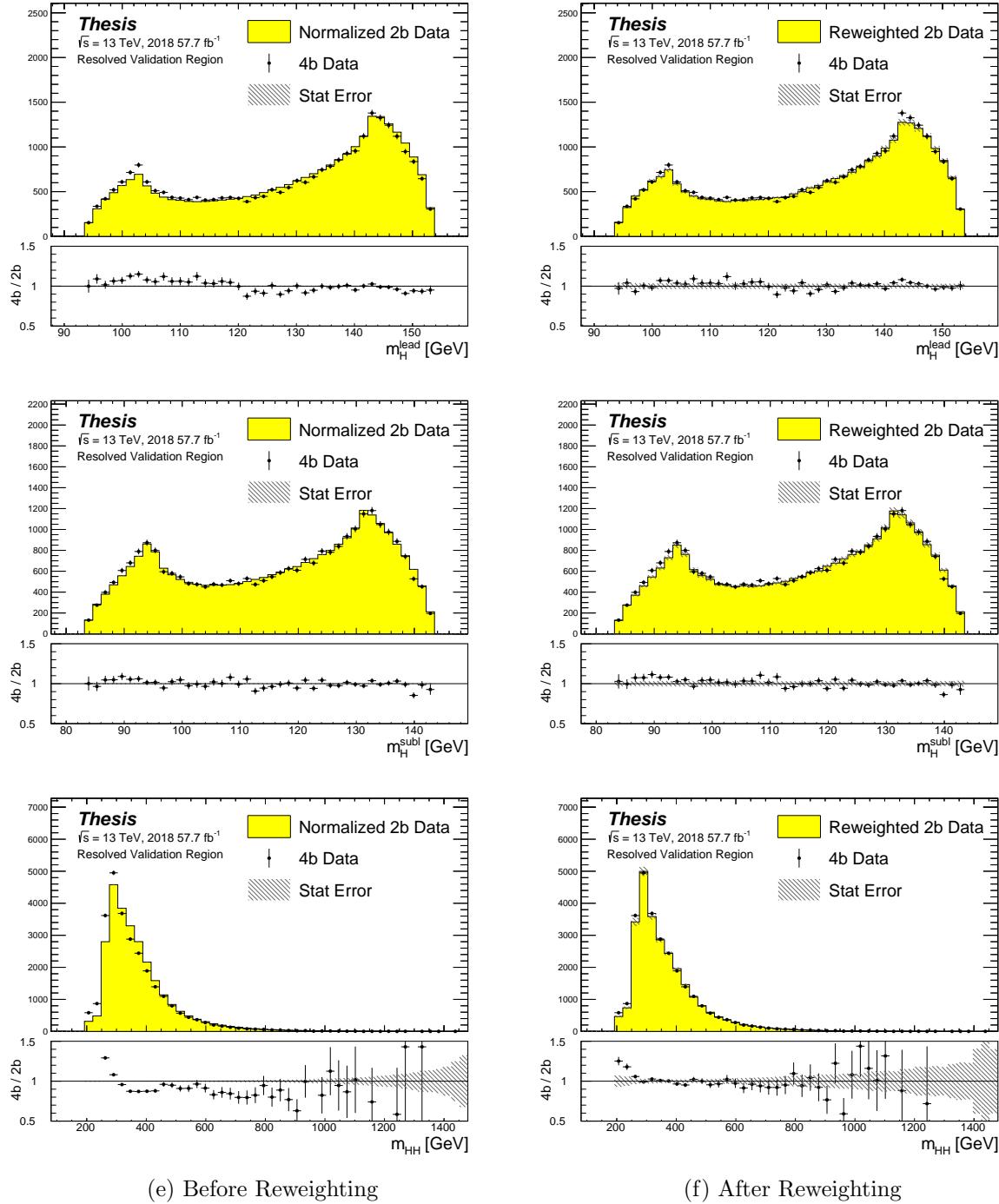


Figure 7.21: **Resonant Search:** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.

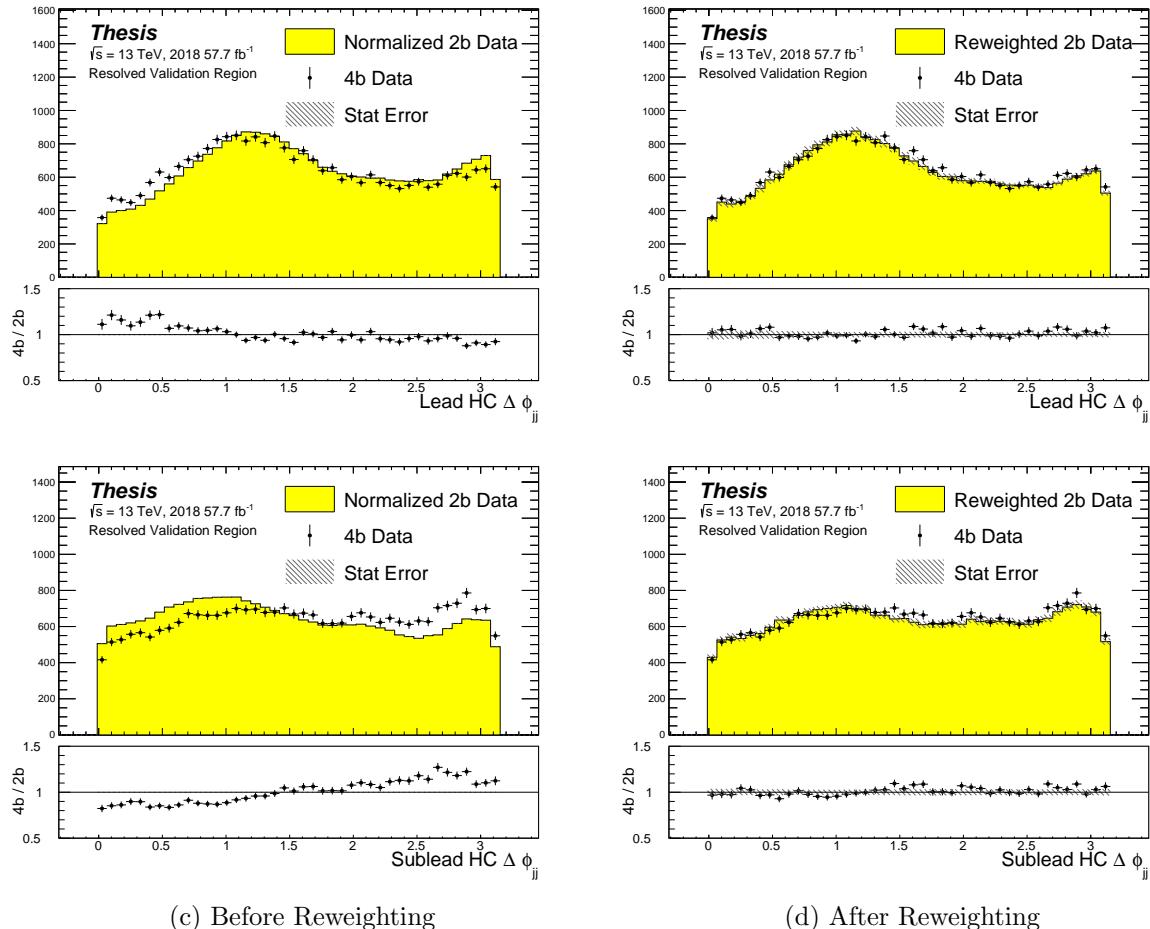


Figure 7.22: **Resonant Search:** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.

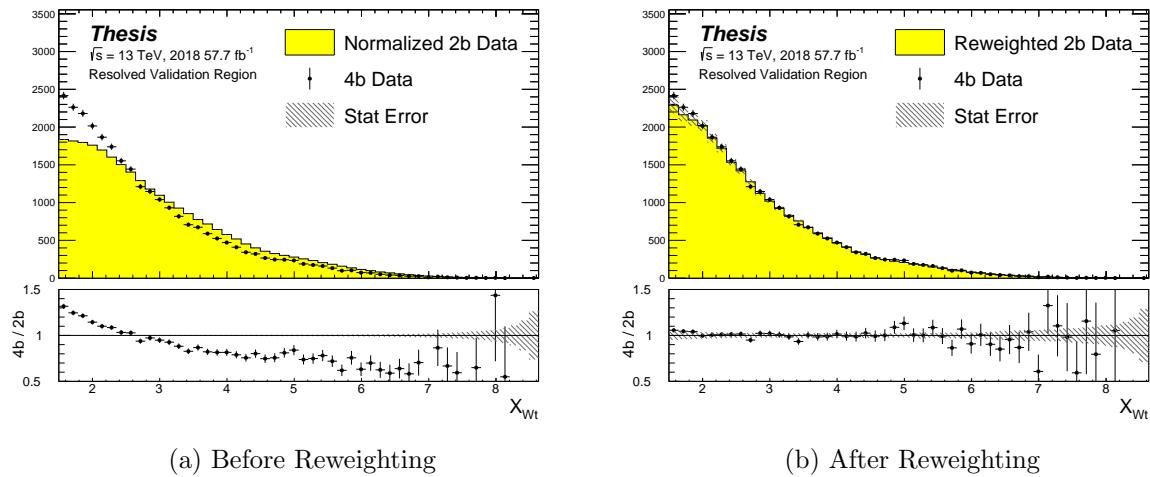


Figure 7.23: **Resonant Search:** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 Validation Region. Reweighting is done after the cut on this variable is applied

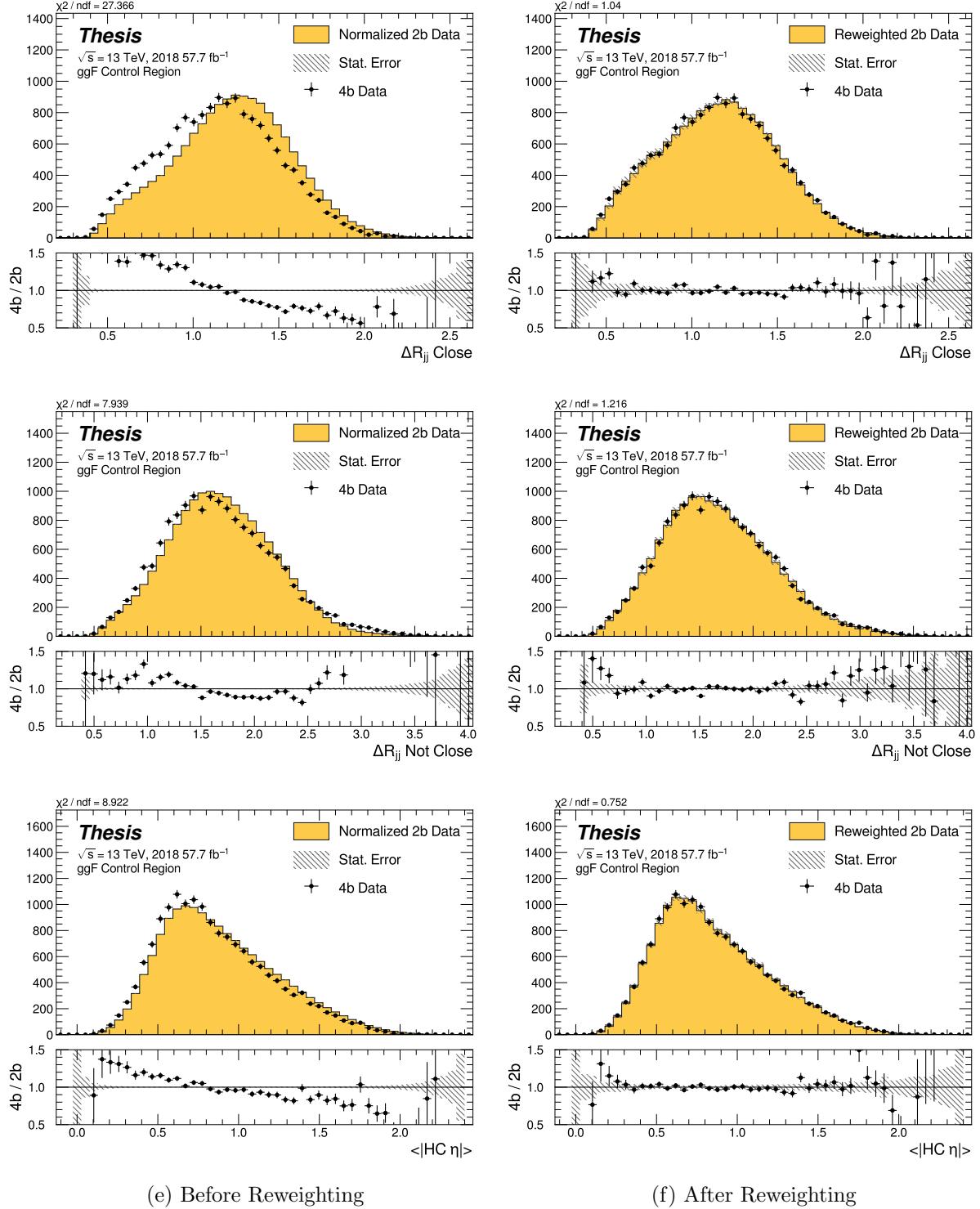


Figure 7.24: **Non-resonant Search (4b):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 4b Control Region.

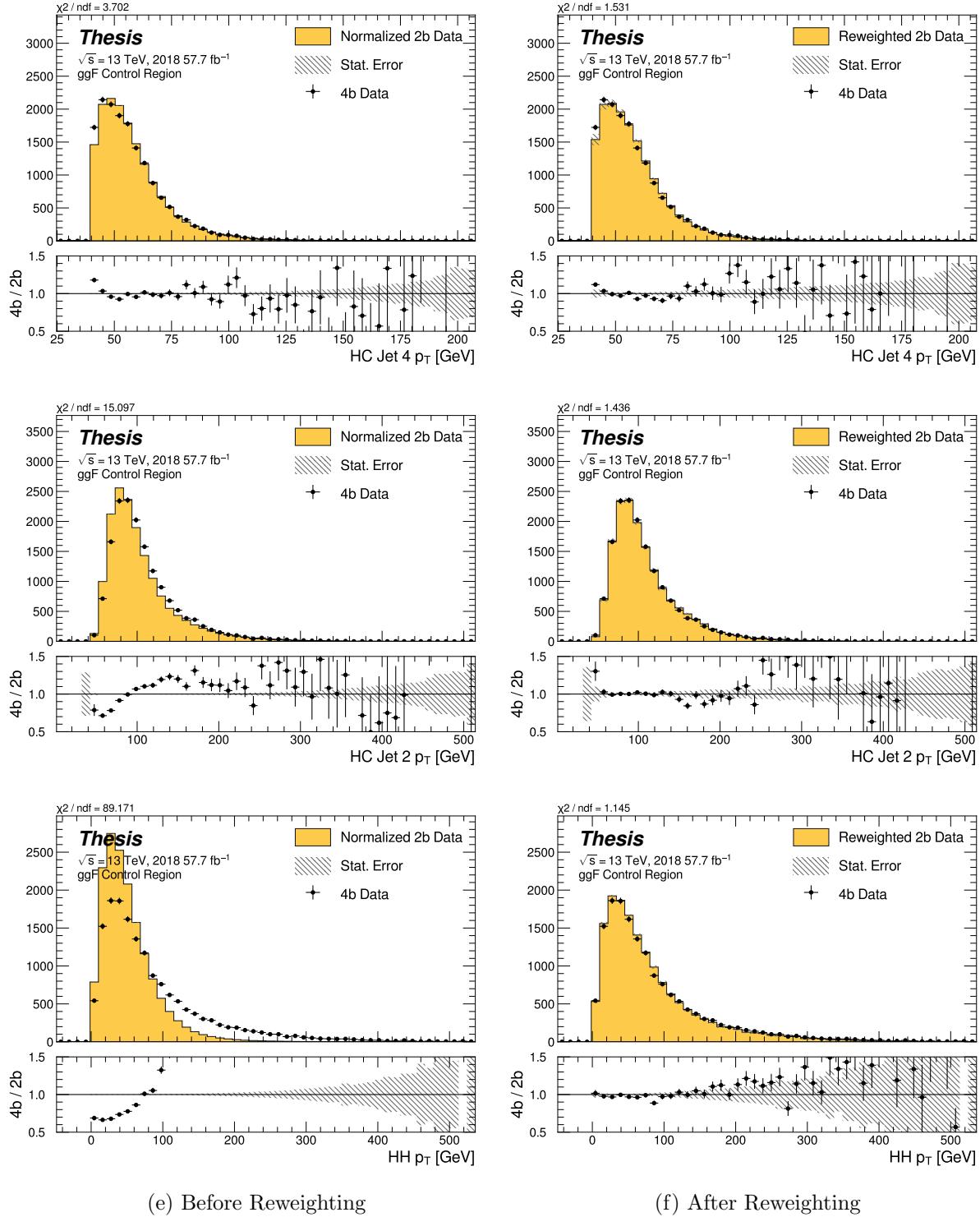


Figure 7.25: **Non-resonant Search (4b):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.

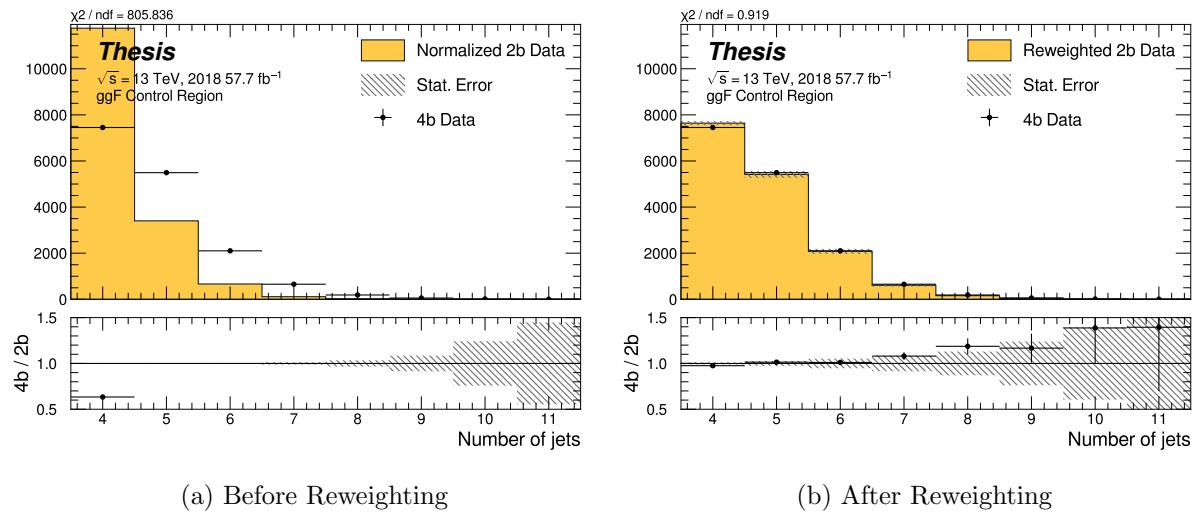


Figure 7.26: **Non-resonant Search (4b):** Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

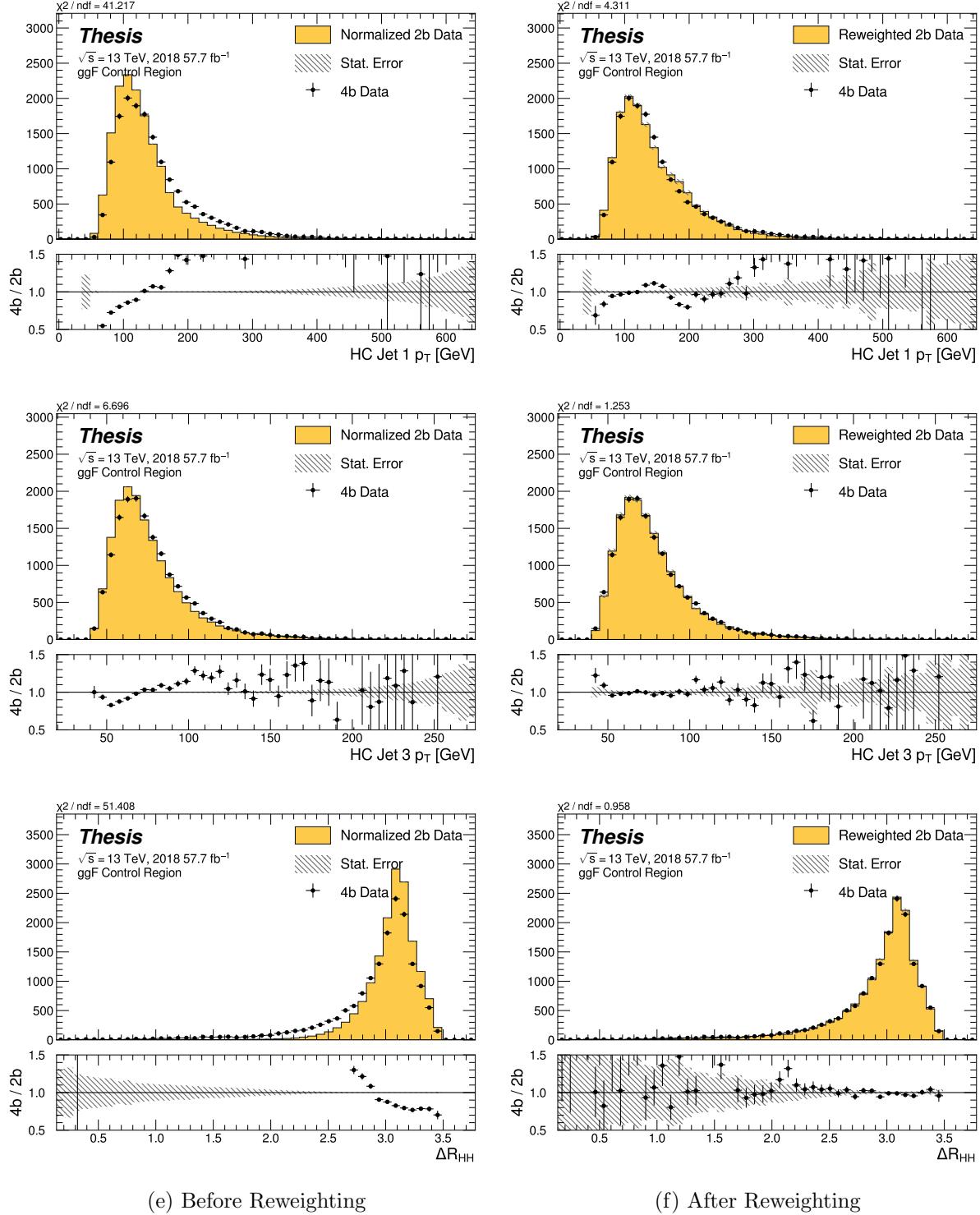


Figure 7.27: **Non-resonant Search (4b):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.

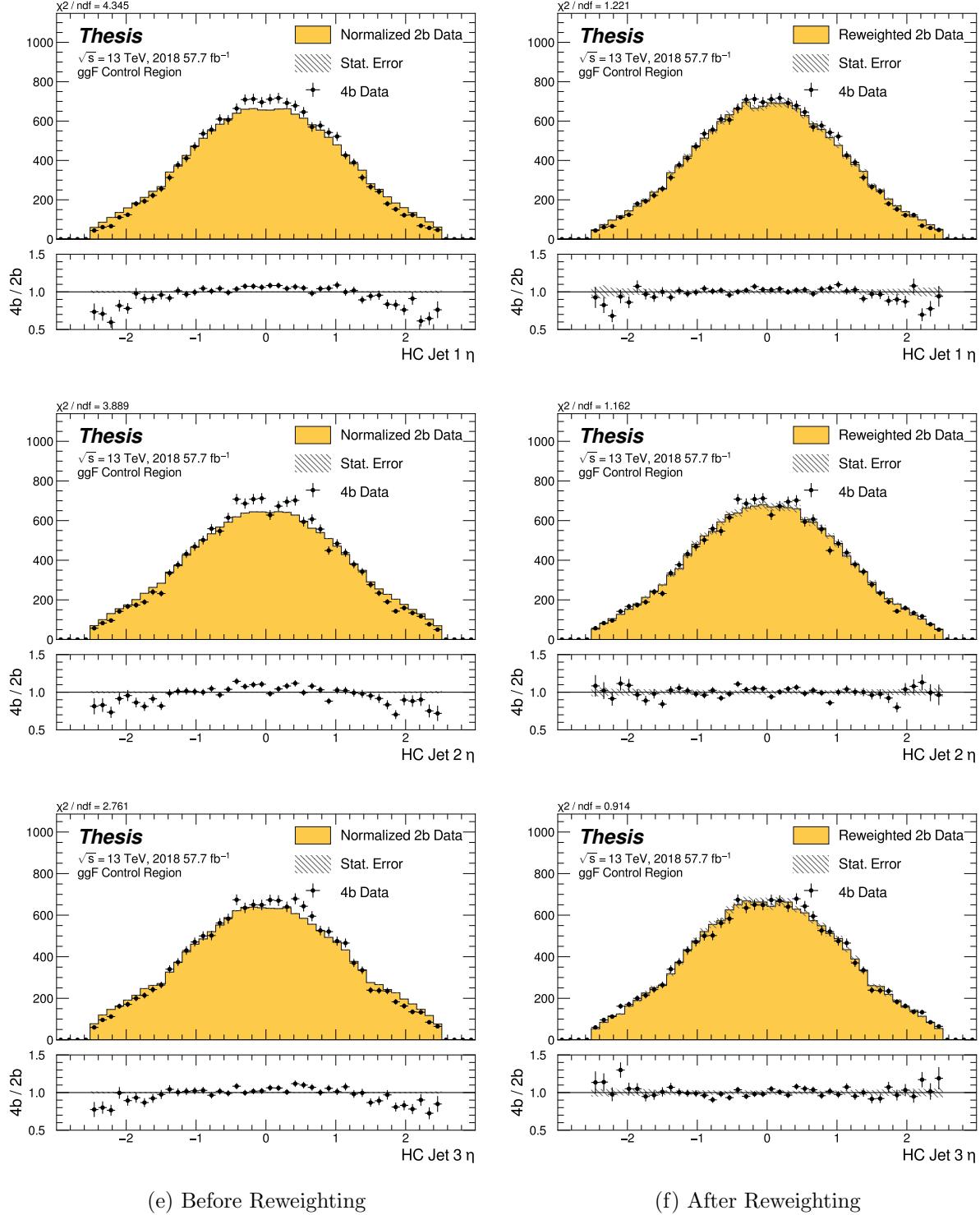


Figure 7.28: **Non-resonant Search (4b):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Control Region.

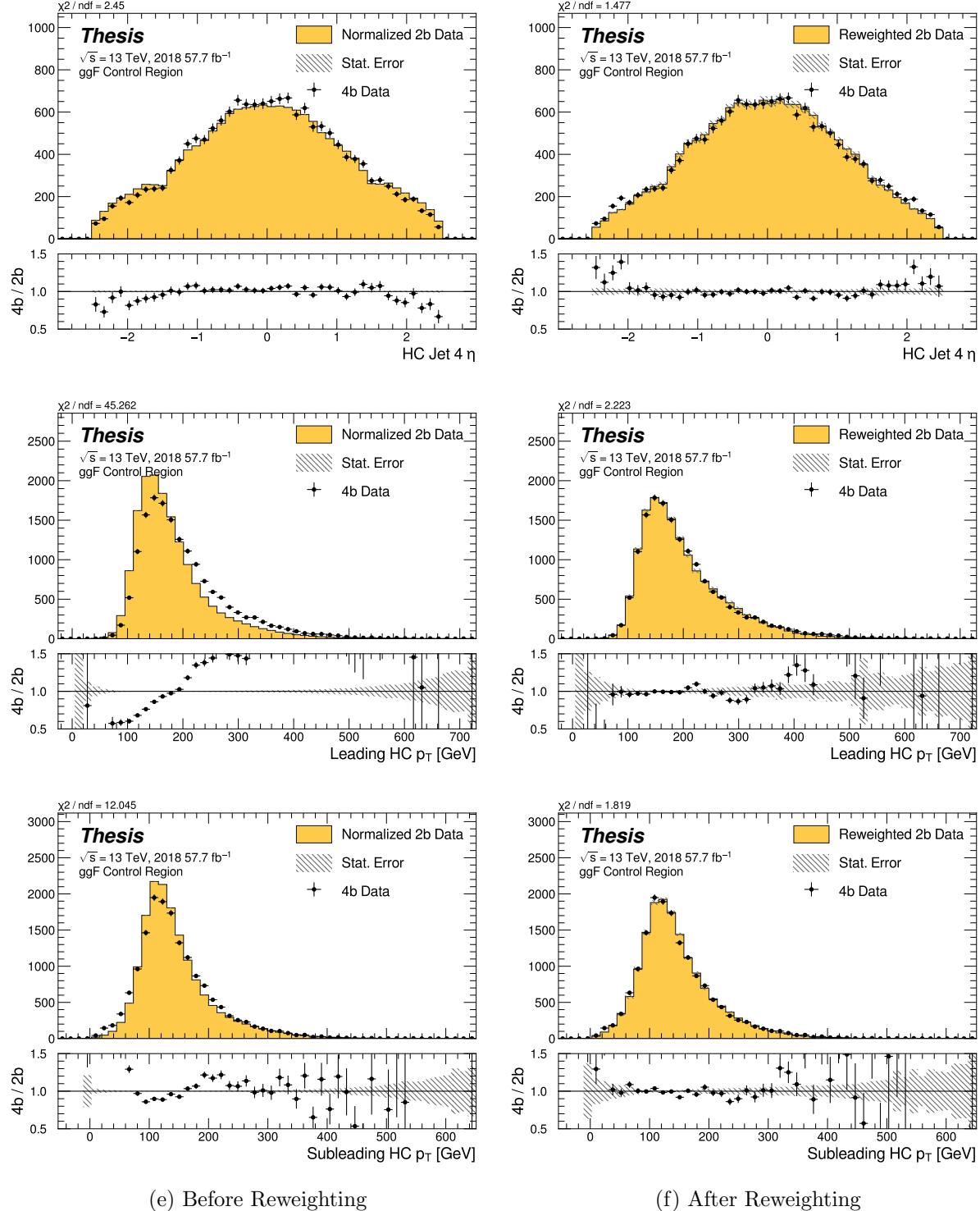


Figure 7.29: **Non-resonant Search (4b):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.

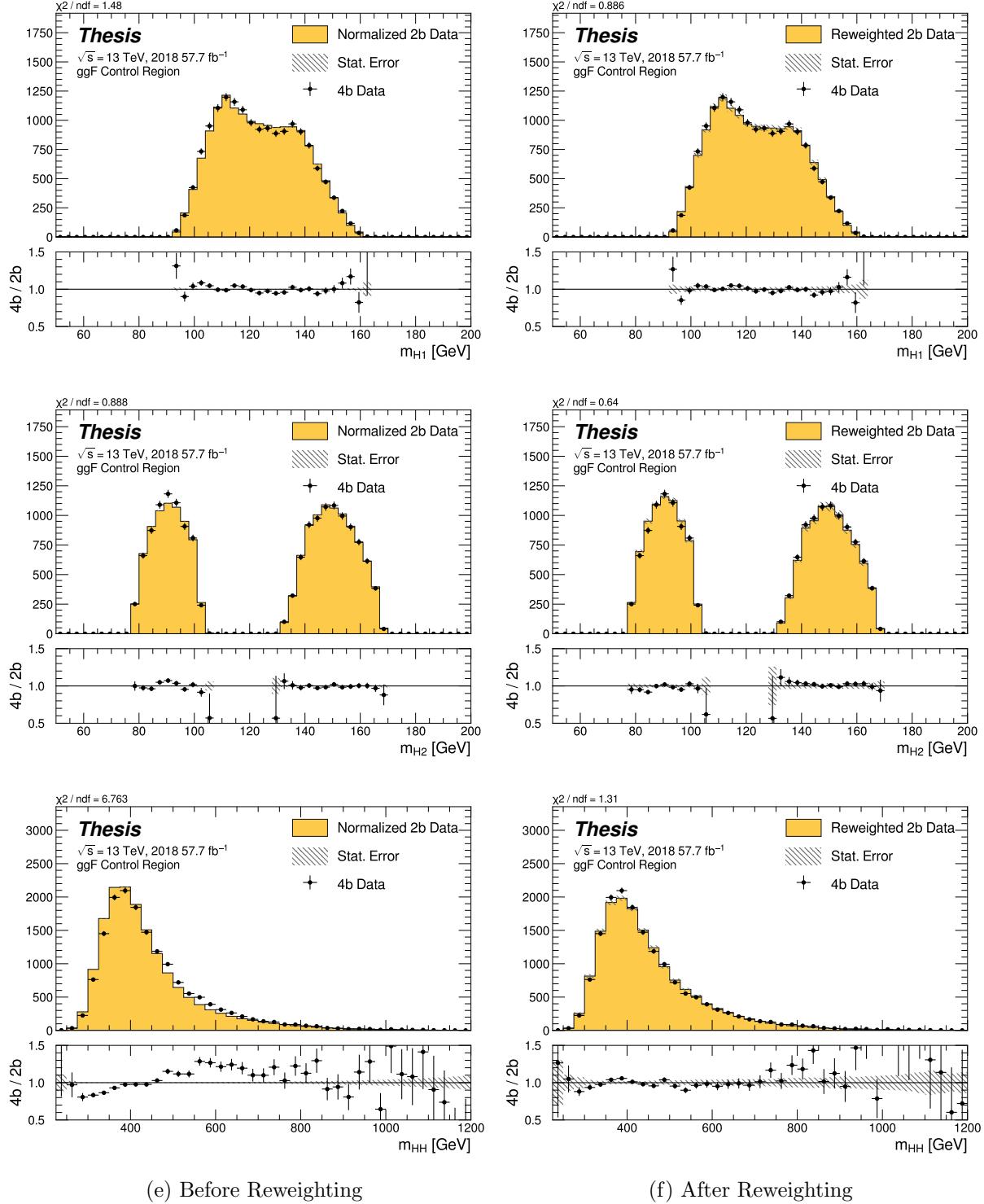


Figure 7.30: **Non-resonant Search (4b):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.

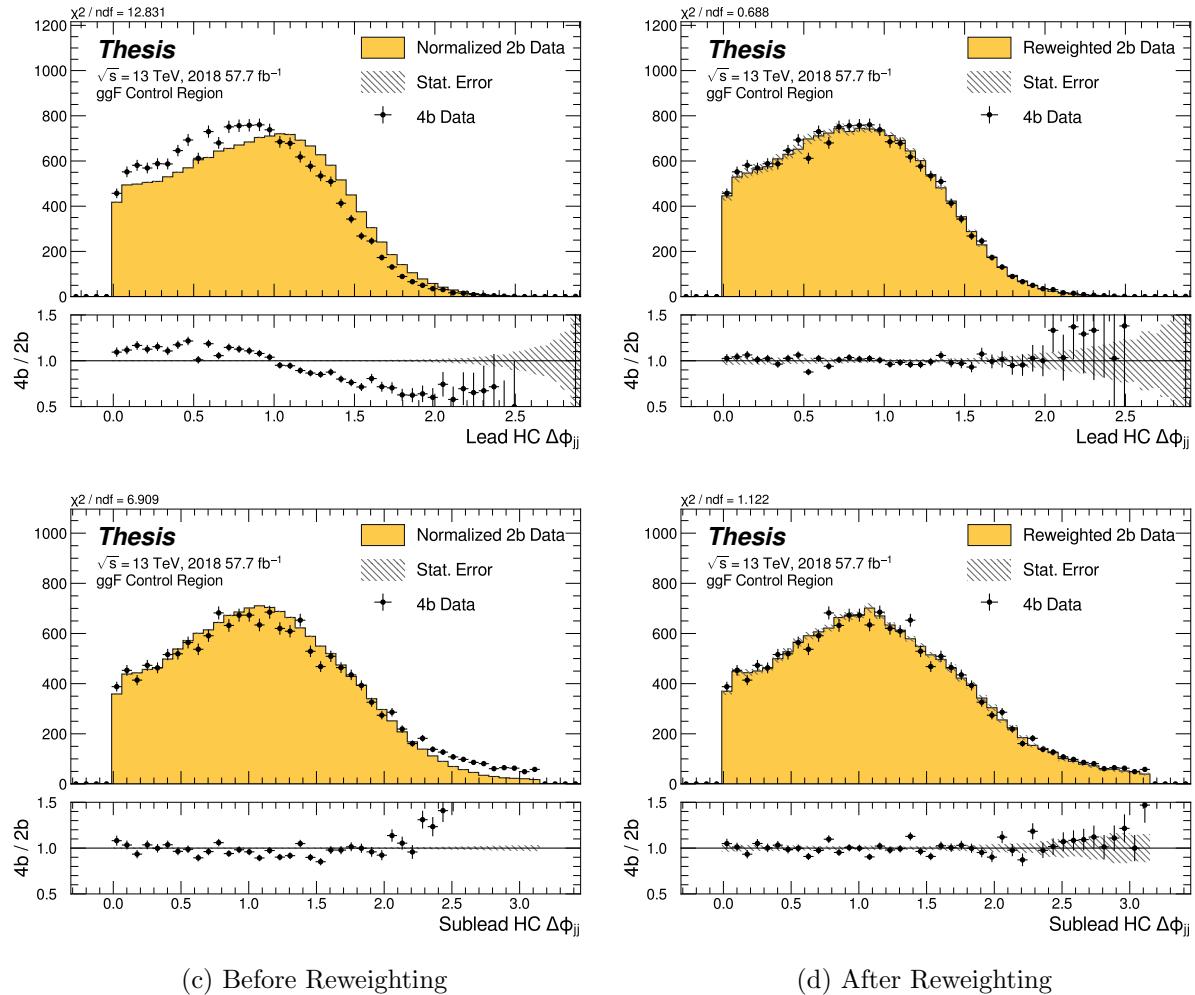


Figure 7.31: **Non-resonant Search (4b):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.

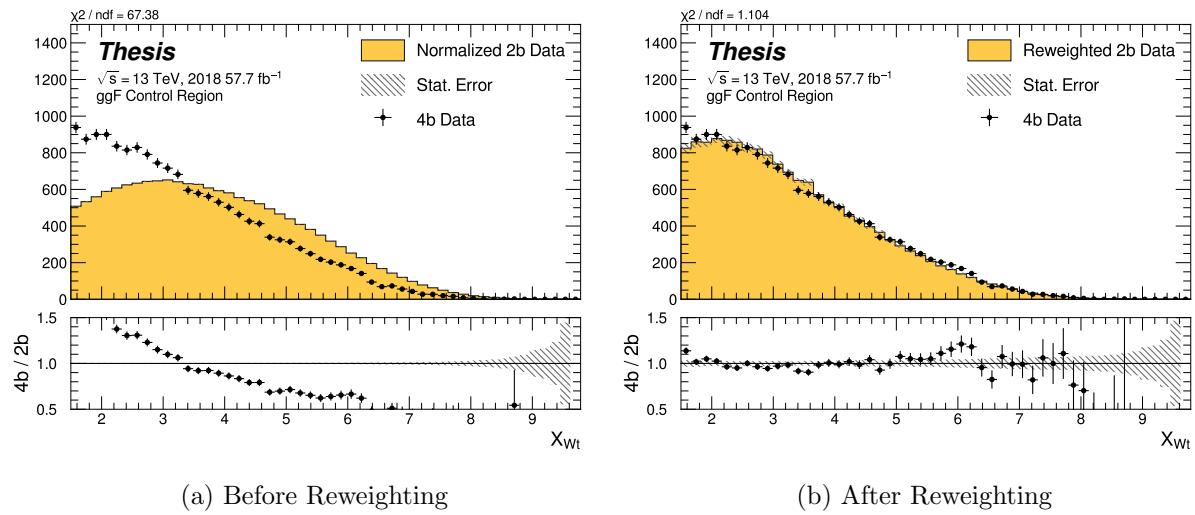


Figure 7.32: **Non-resonant Search (4b)**: Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 4b Control Region. Reweighting is done after the cut on this variable is applied.

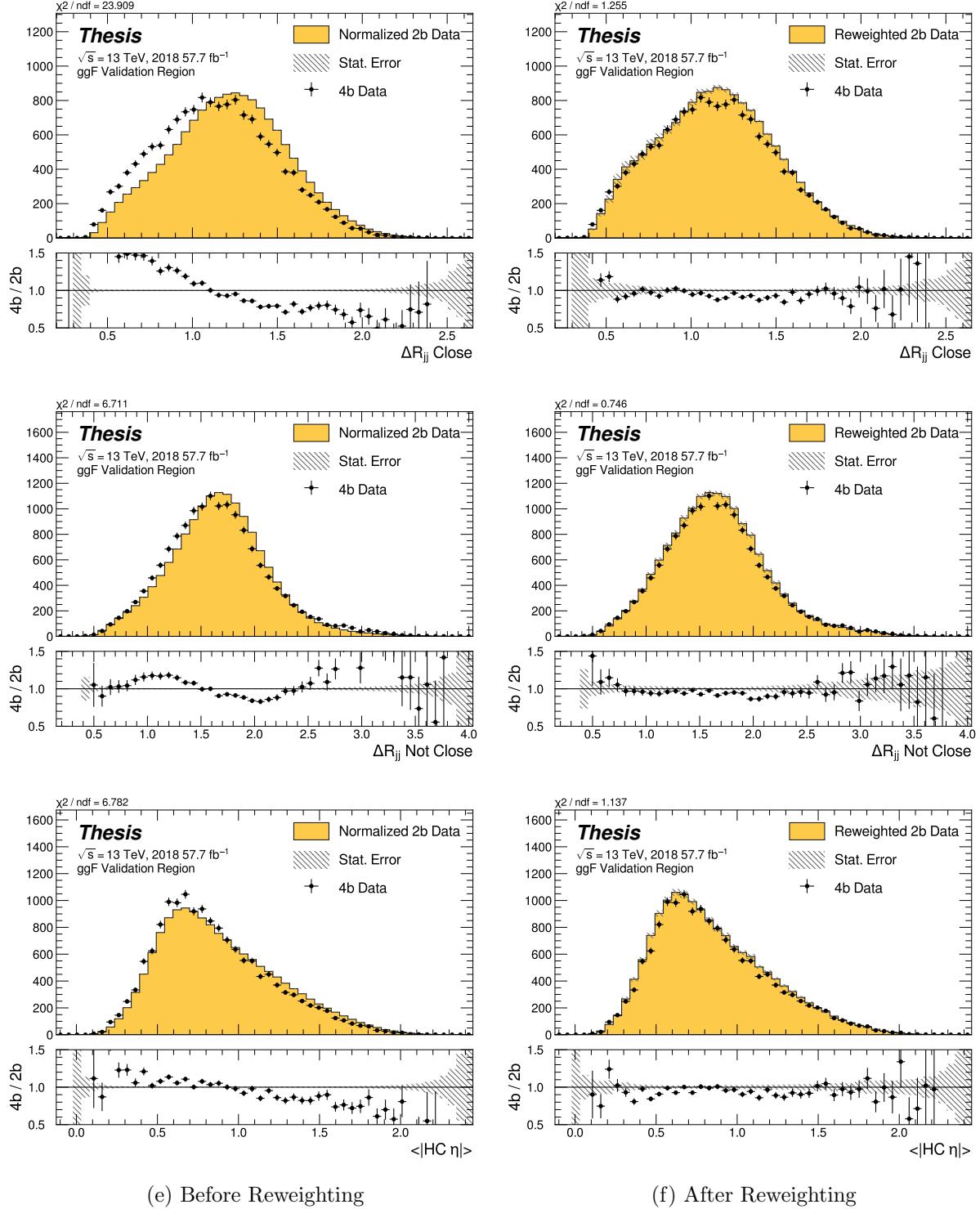


Figure 7.33: **Non-resonant Search (4b):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 4b Validation Region.

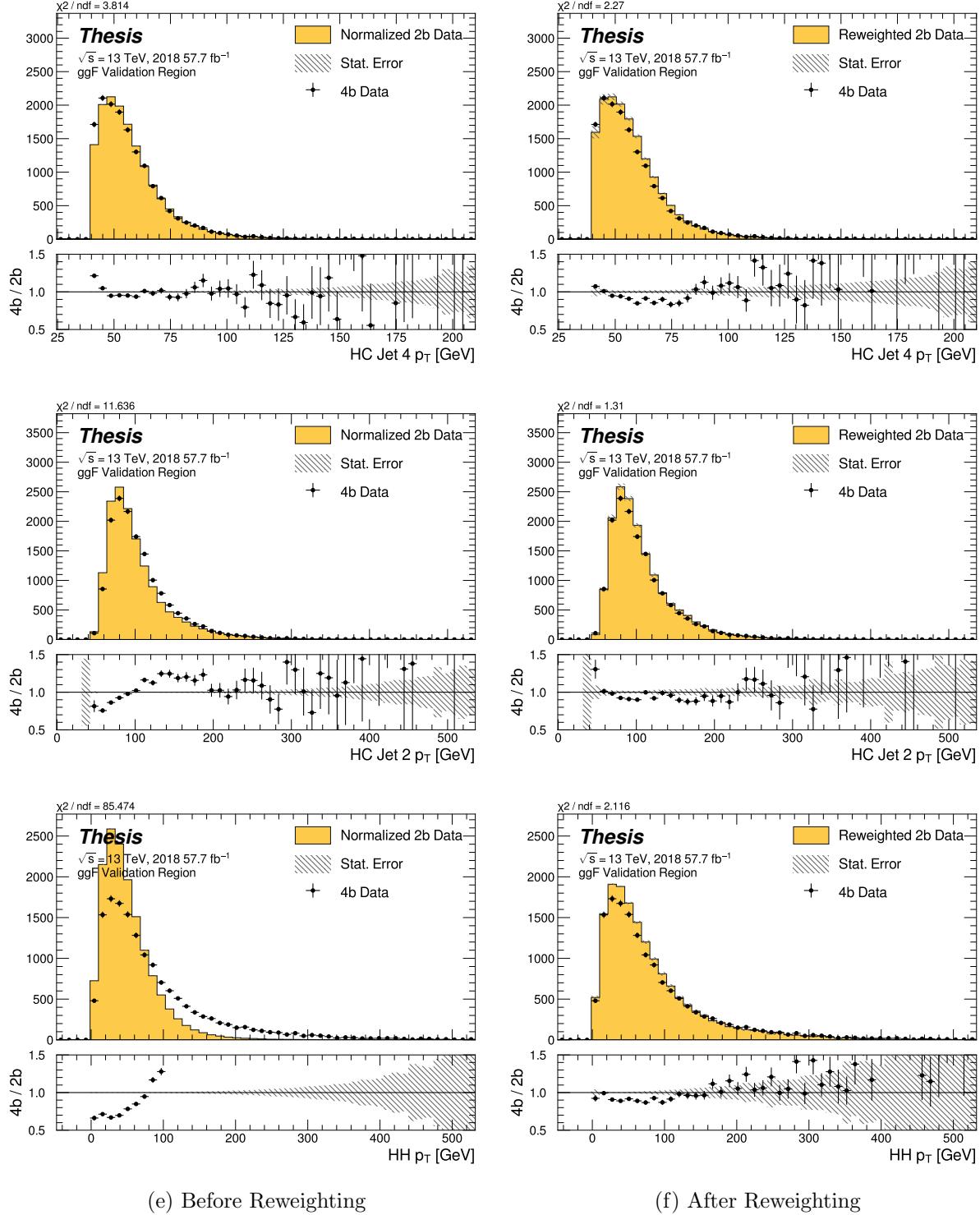


Figure 7.34: **Non-resonant Search (4b):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.

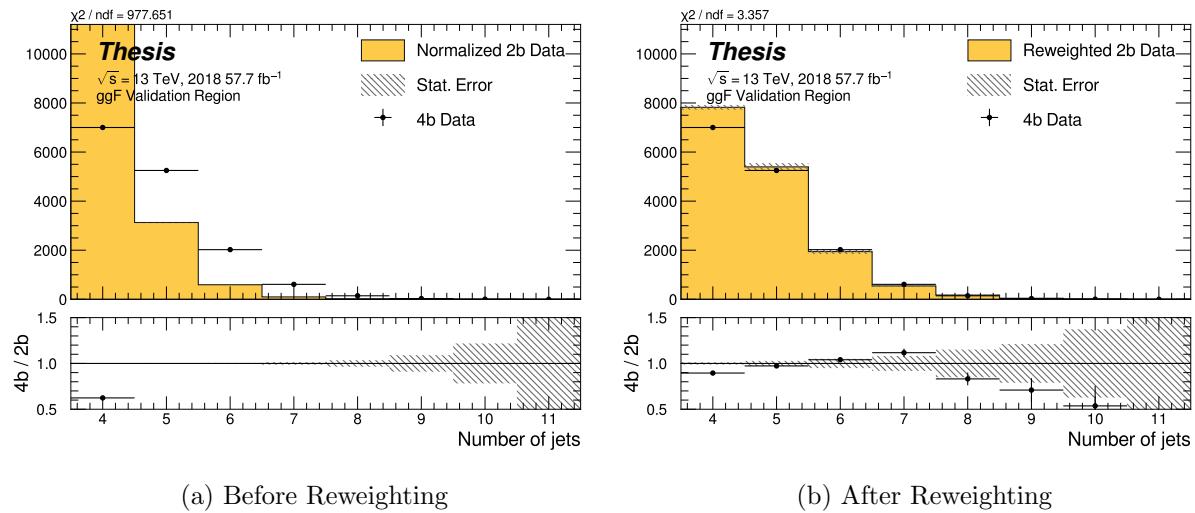


Figure 7.35: **Non-resonant Search (4b):** Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

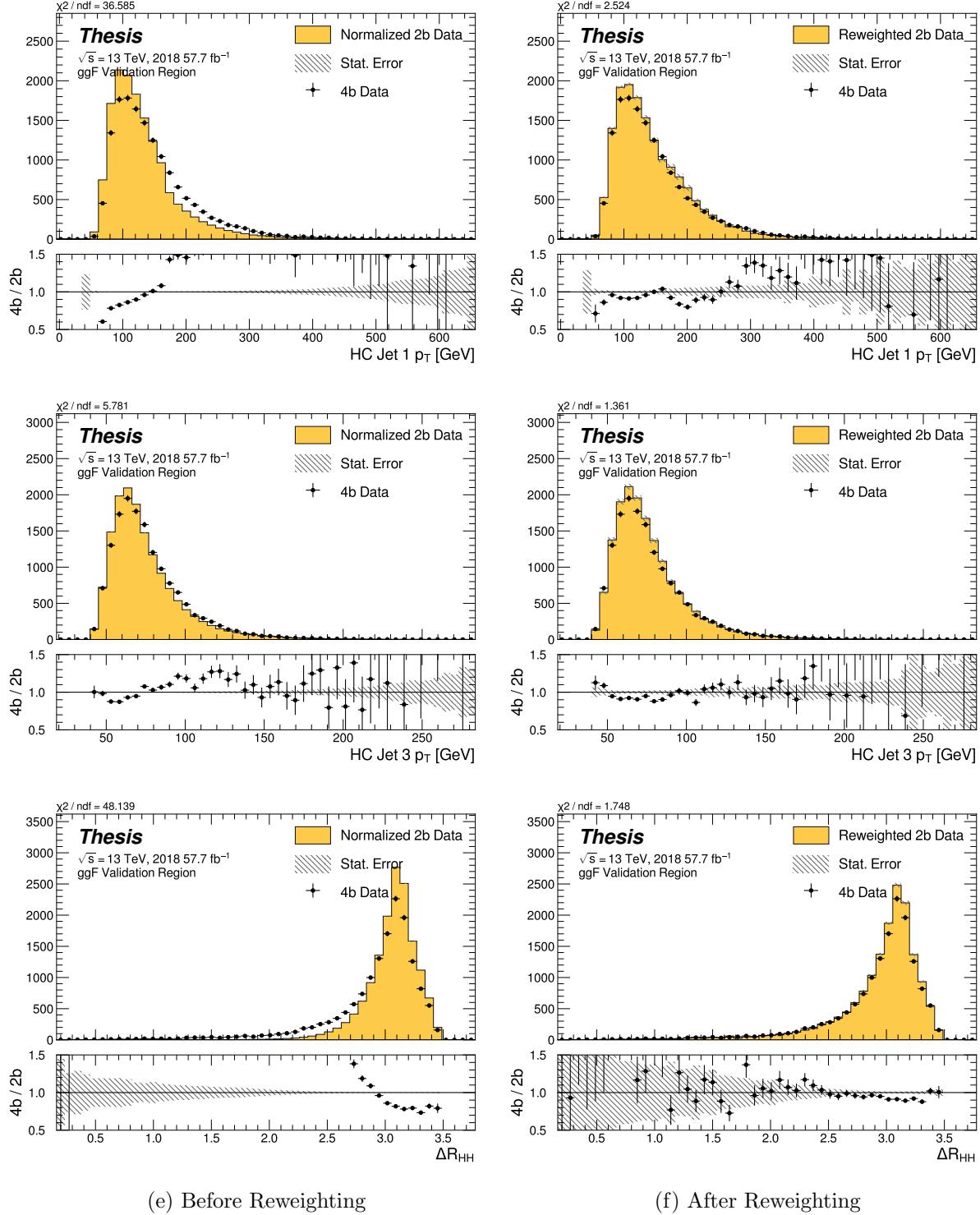


Figure 7.36: **Non-resonant Search (4b):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.

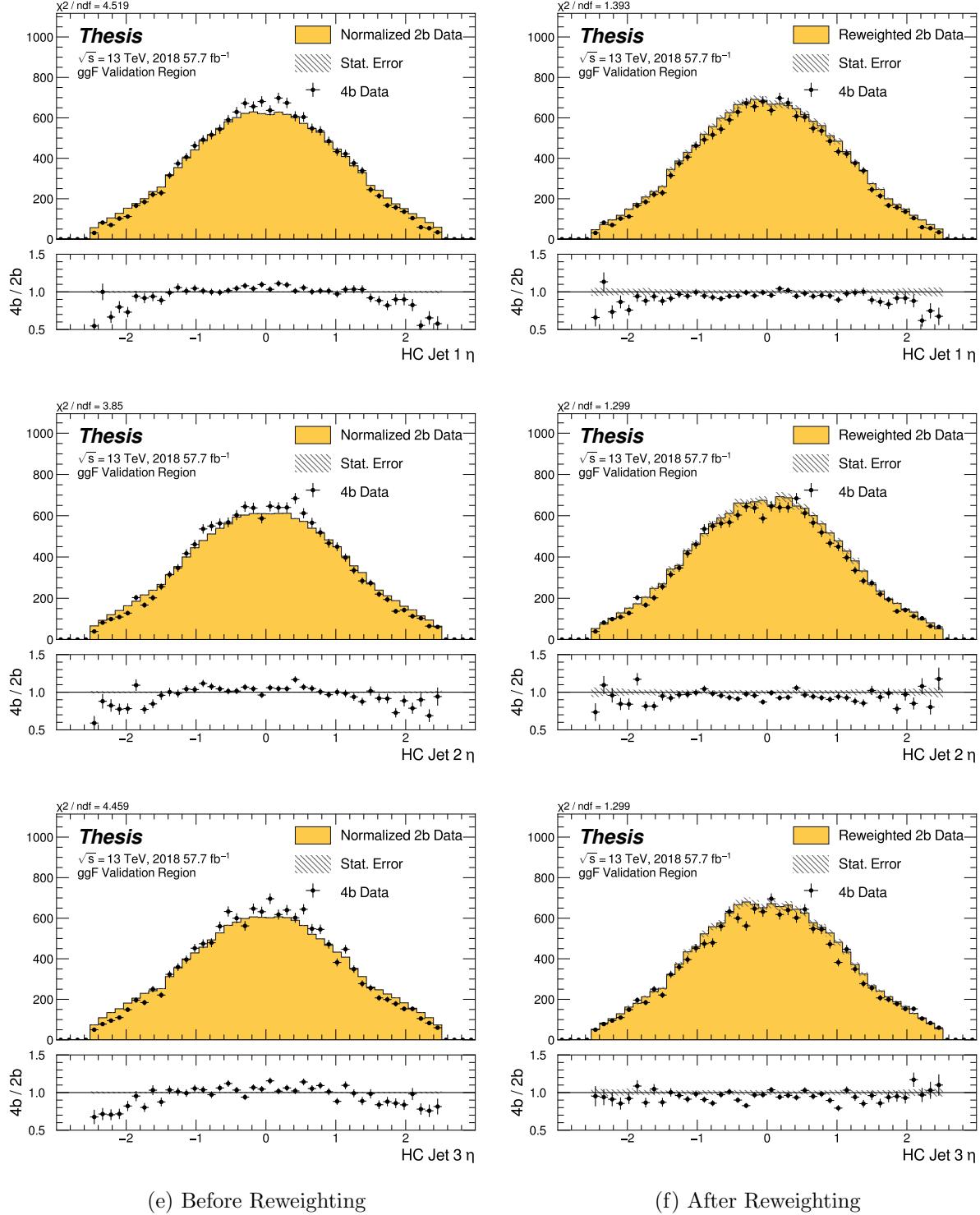


Figure 7.37: **Non-resonant Search (4b):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Validation Region.

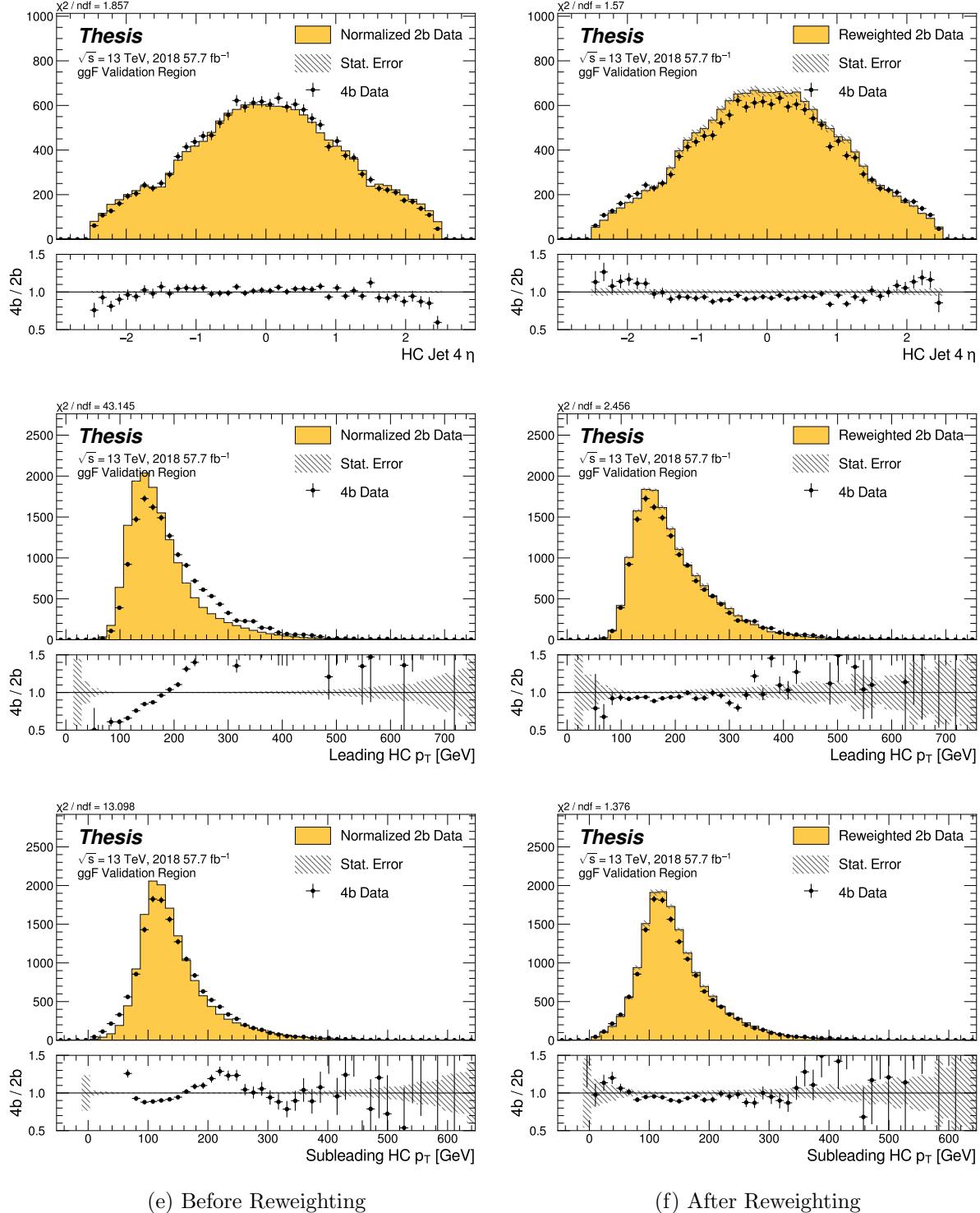


Figure 7.38: **Non-resonant Search (4b):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.

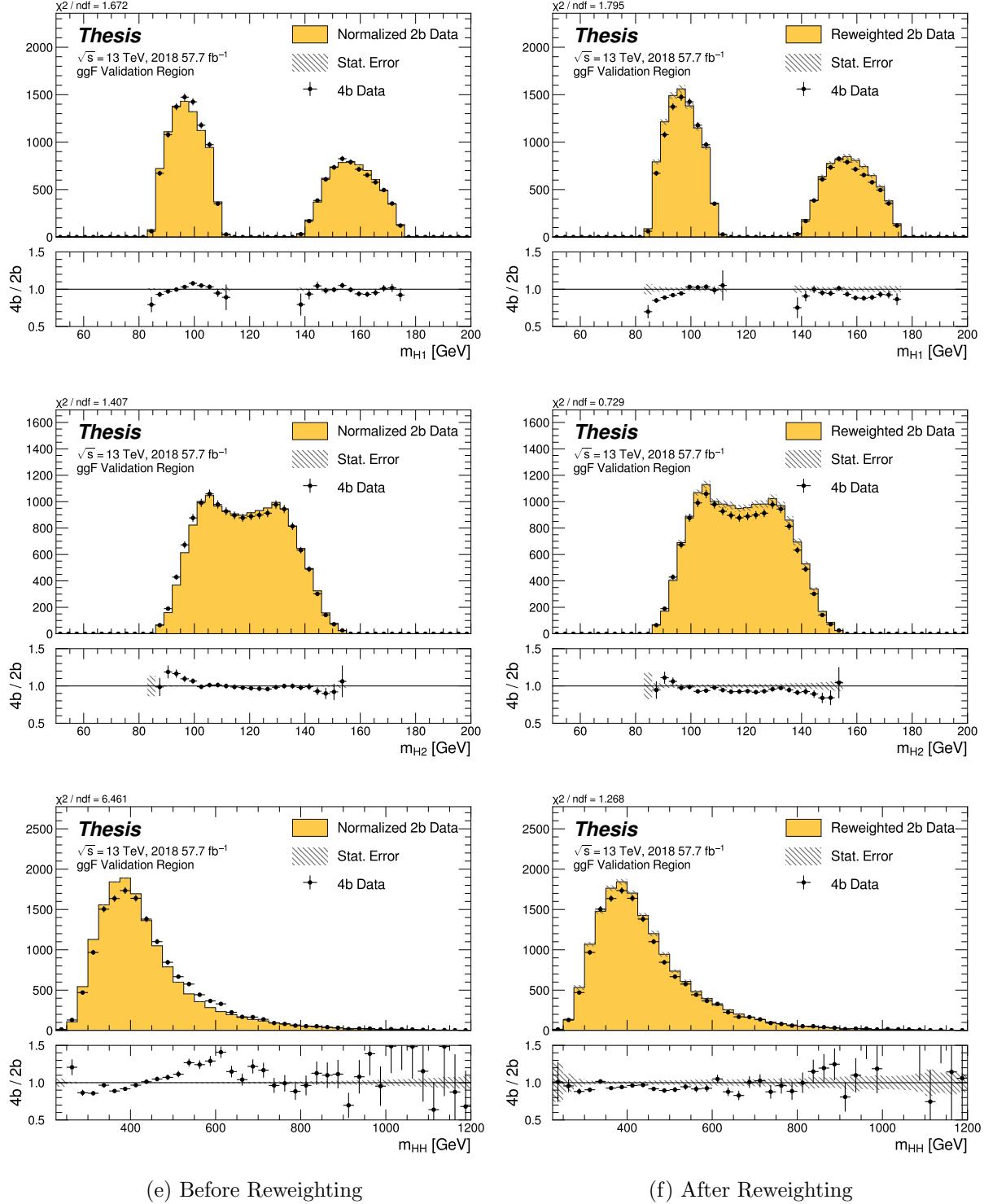


Figure 7.39: **Non-resonant Search (4b):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.

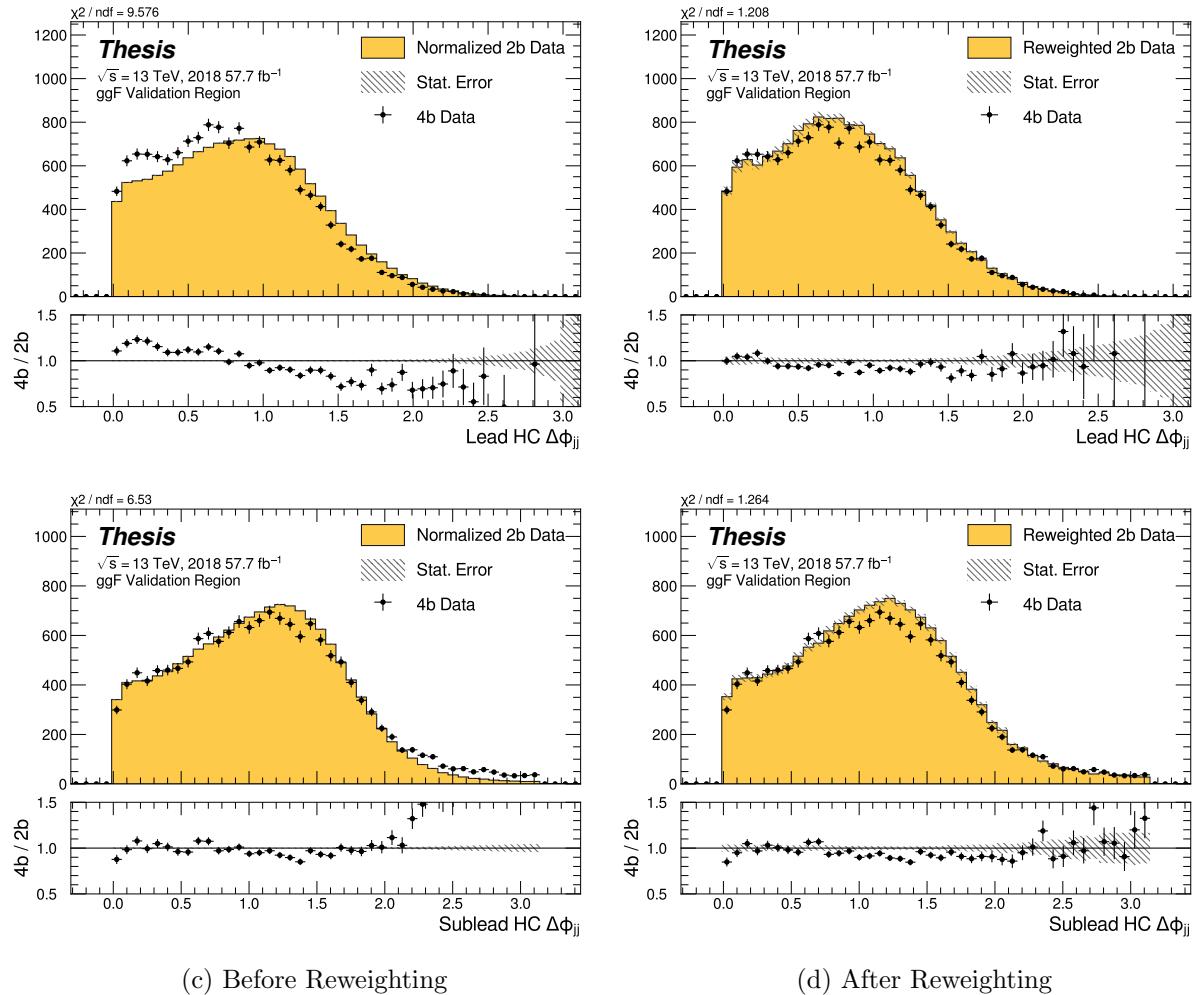


Figure 7.40: **Non-resonant Search (4b):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.

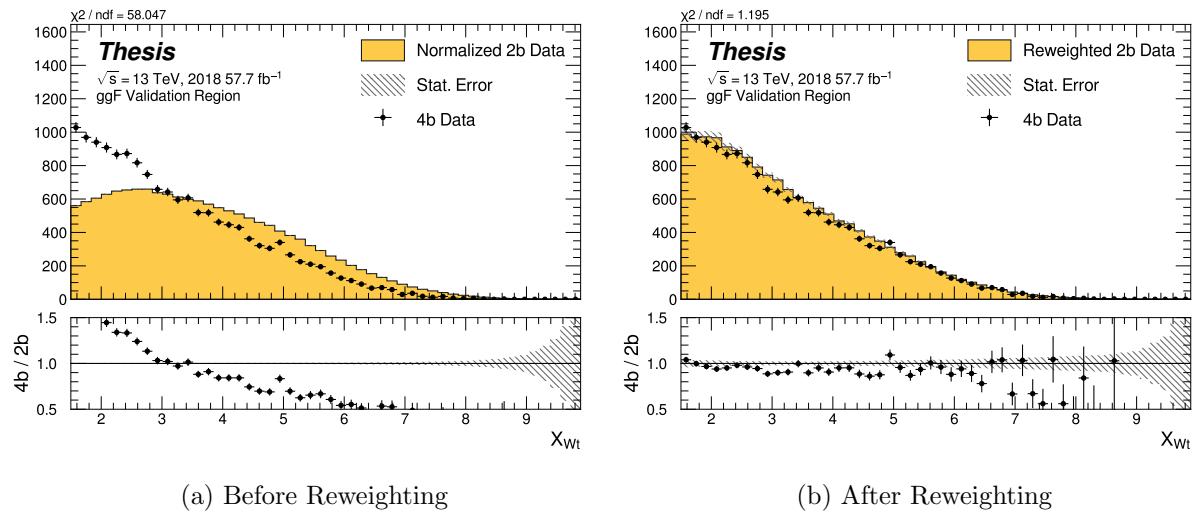


Figure 7.41: **Non-resonant Search (4b):** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 4b Validation Region. Reweighting is done after the cut on this variable is applied.

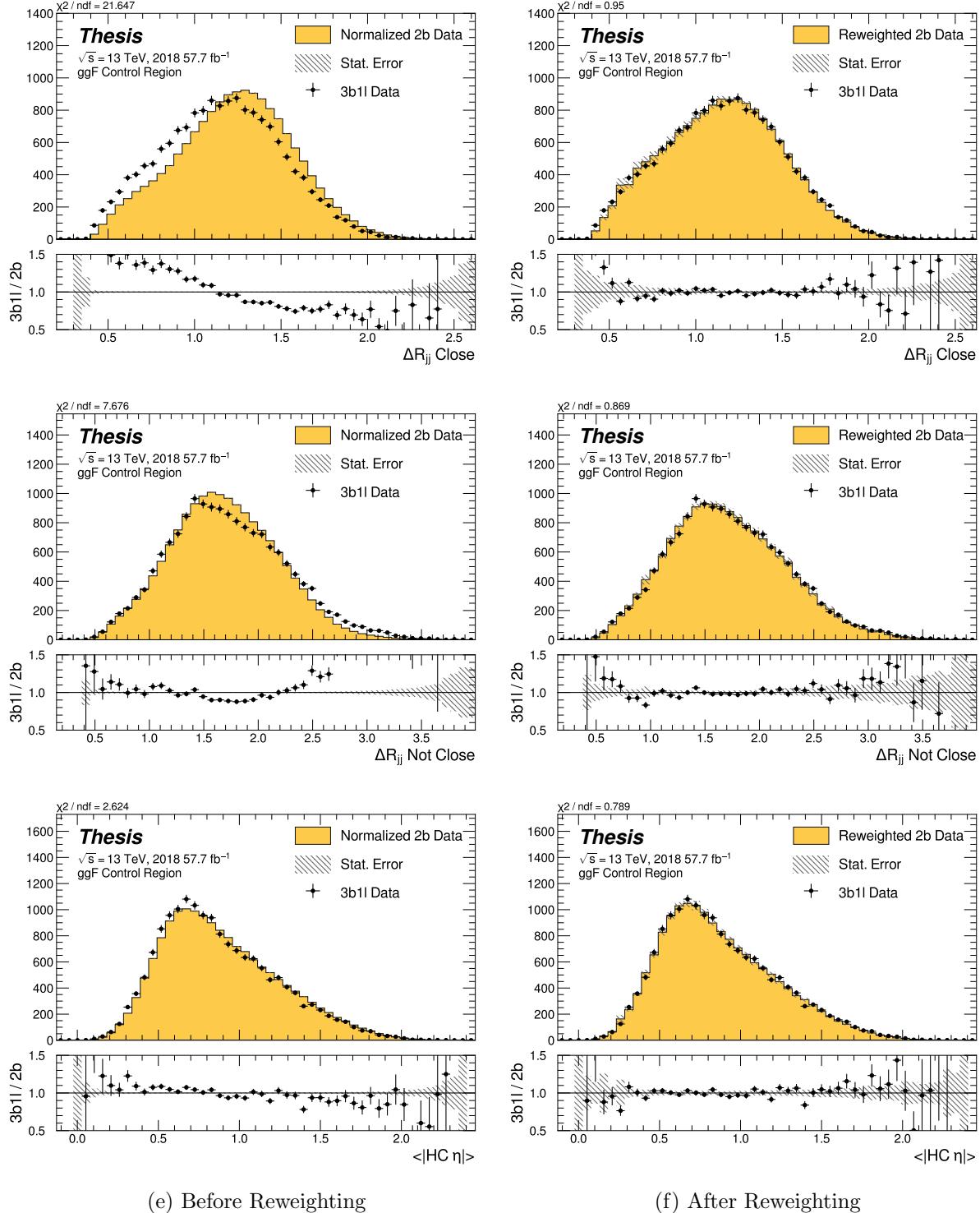


Figure 7.42: **Non-resonant Search (3b1l):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 3b1l Control Region.

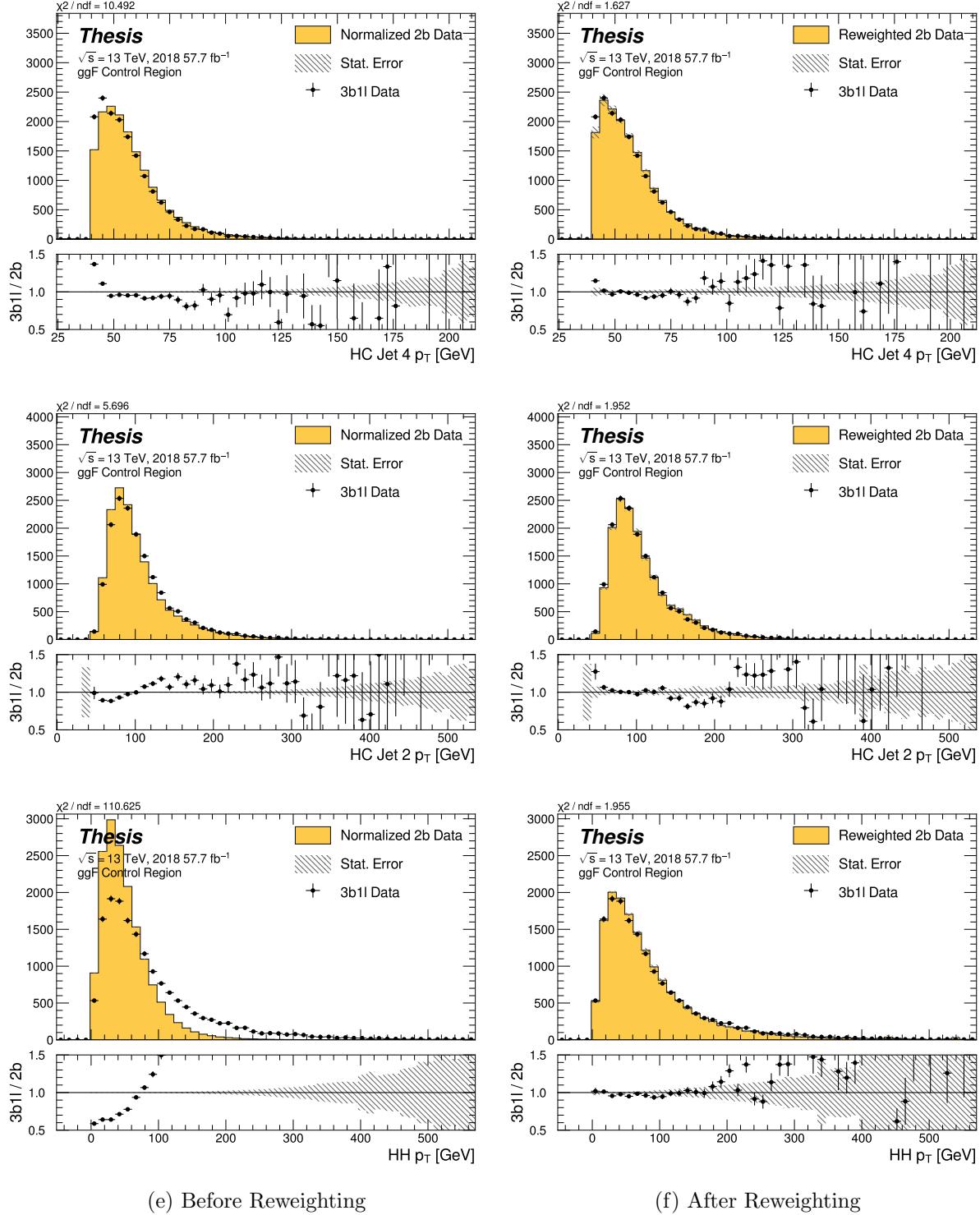


Figure 7.43: **Non-resonant Search (3b1l):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Control Region.

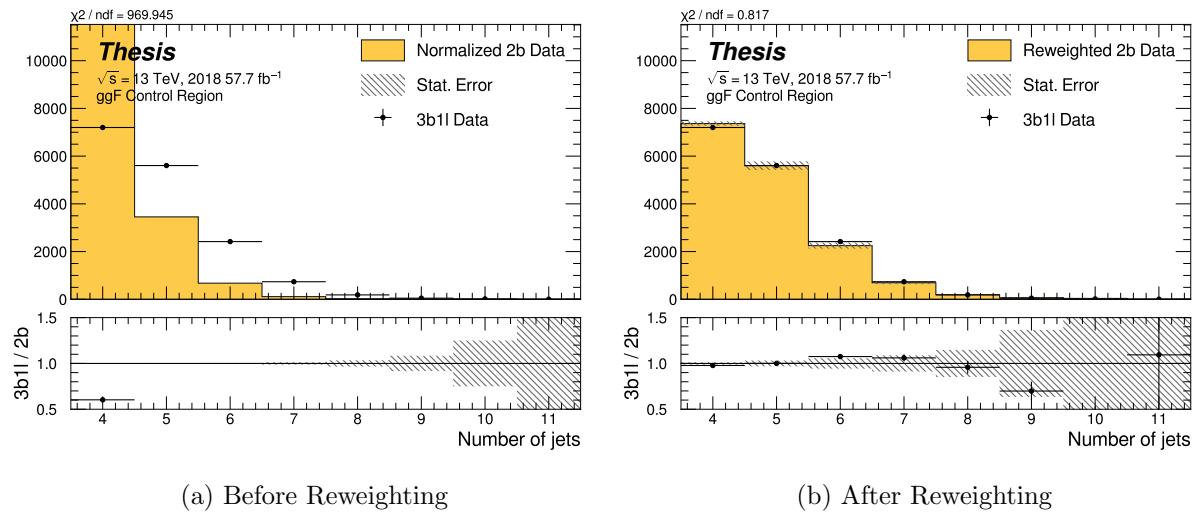


Figure 7.44: **Non-resonant Search (3b1l):** Distributions of the number of jets before and after CR derived reweighting for the 2018 3b1l Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

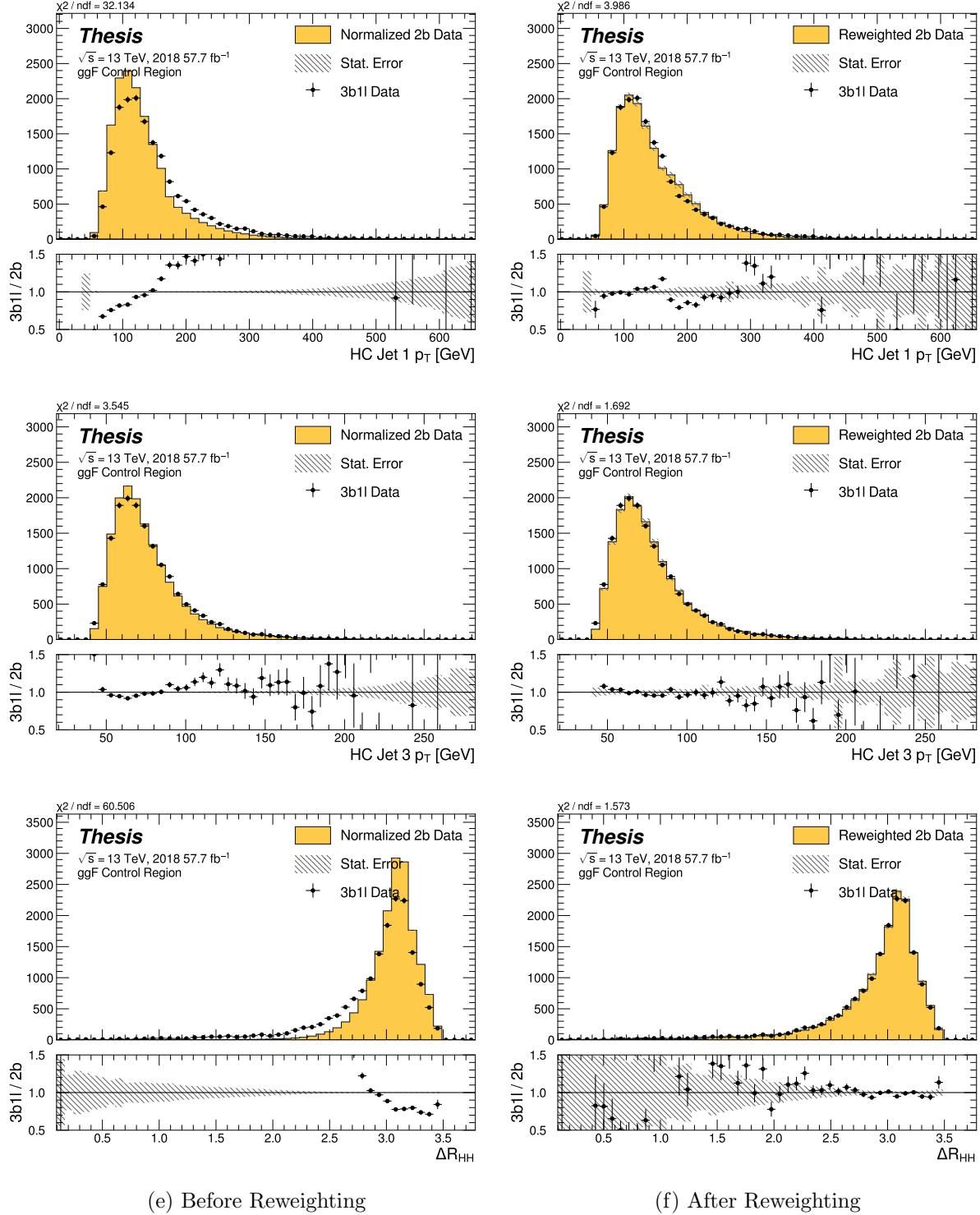


Figure 7.45: **Non-resonant Search (3b1l):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.

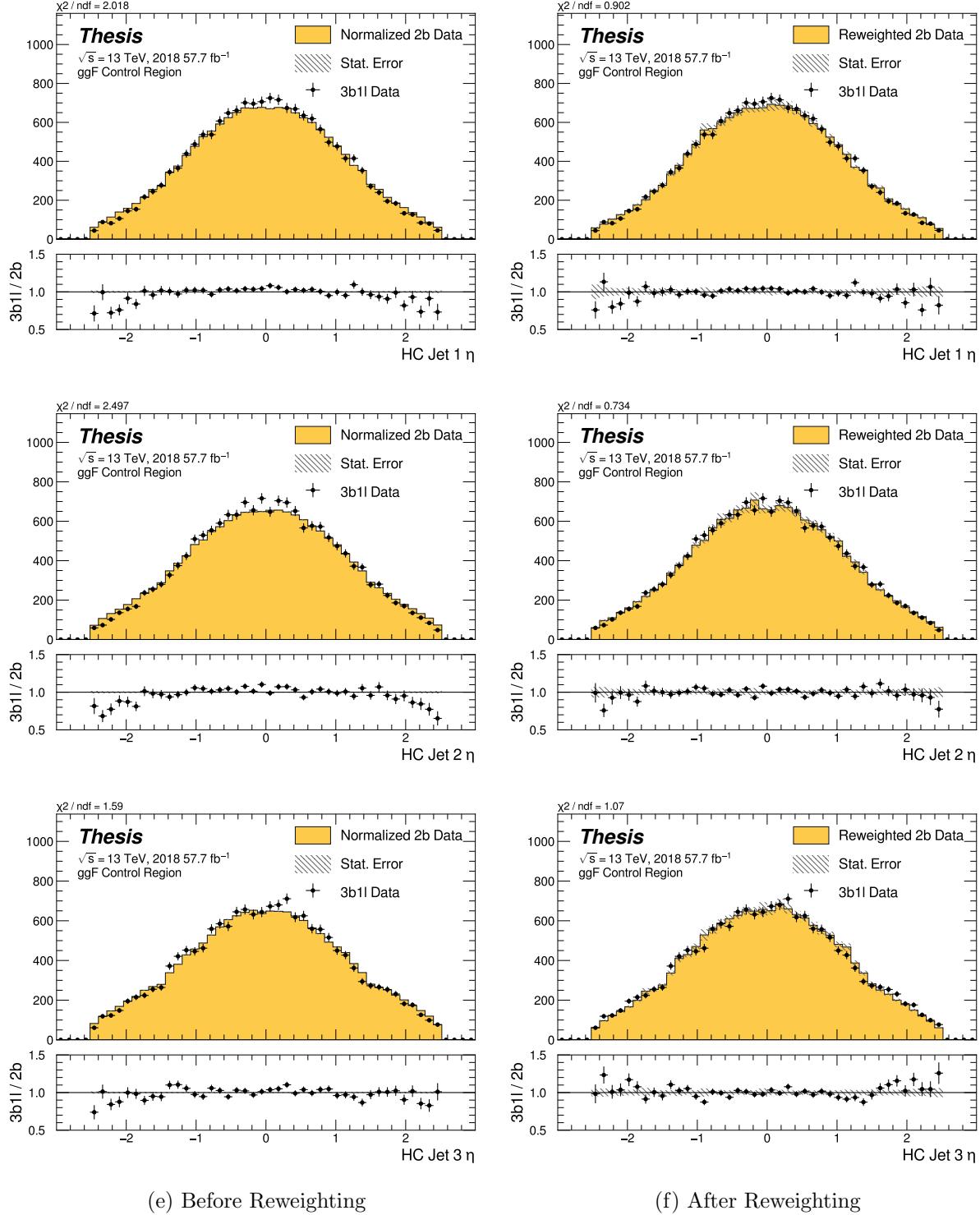


Figure 7.46: **Non-resonant Search (3b1l):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Control Region.

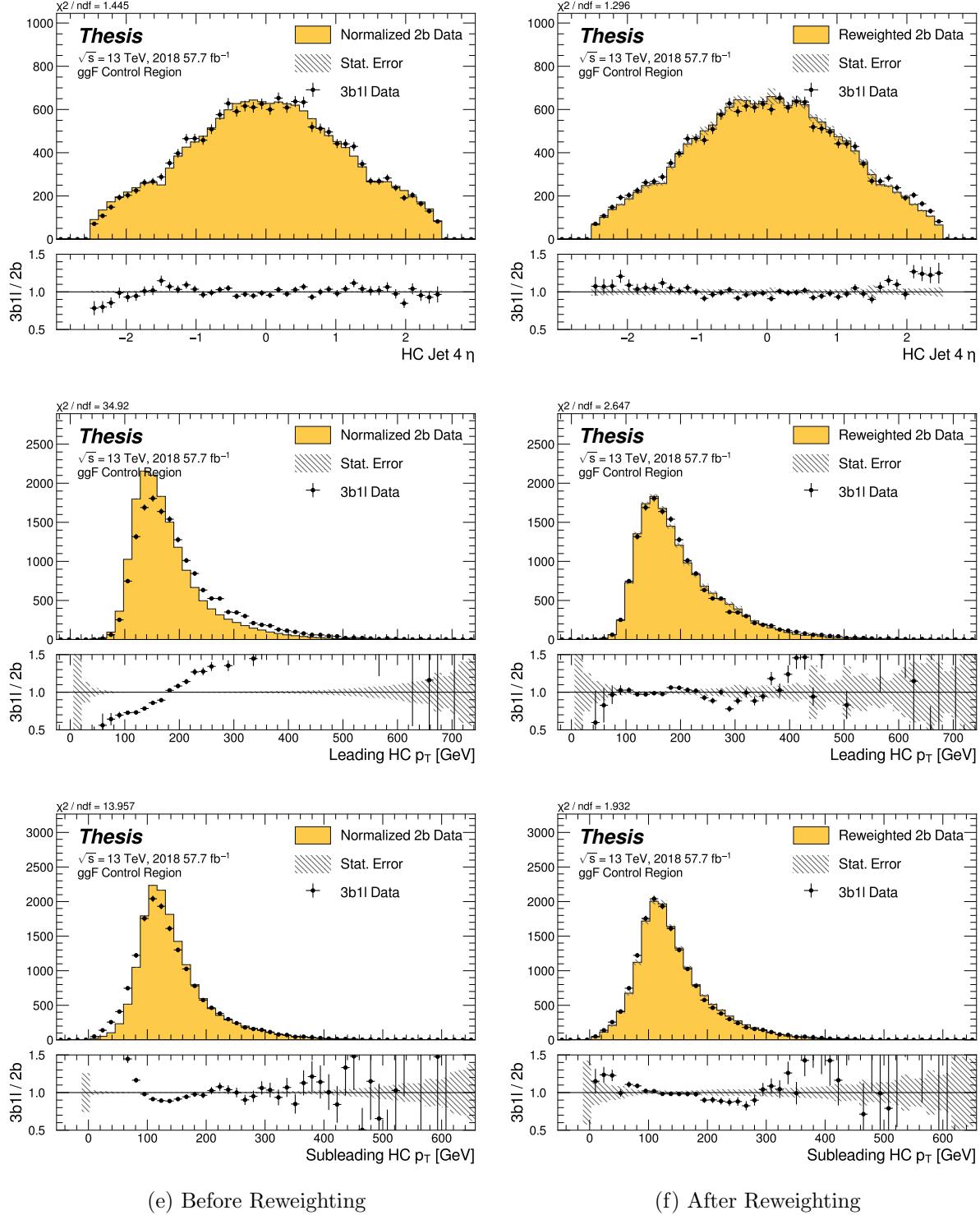


Figure 7.47: **Non-resonant Search (3b1l):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.

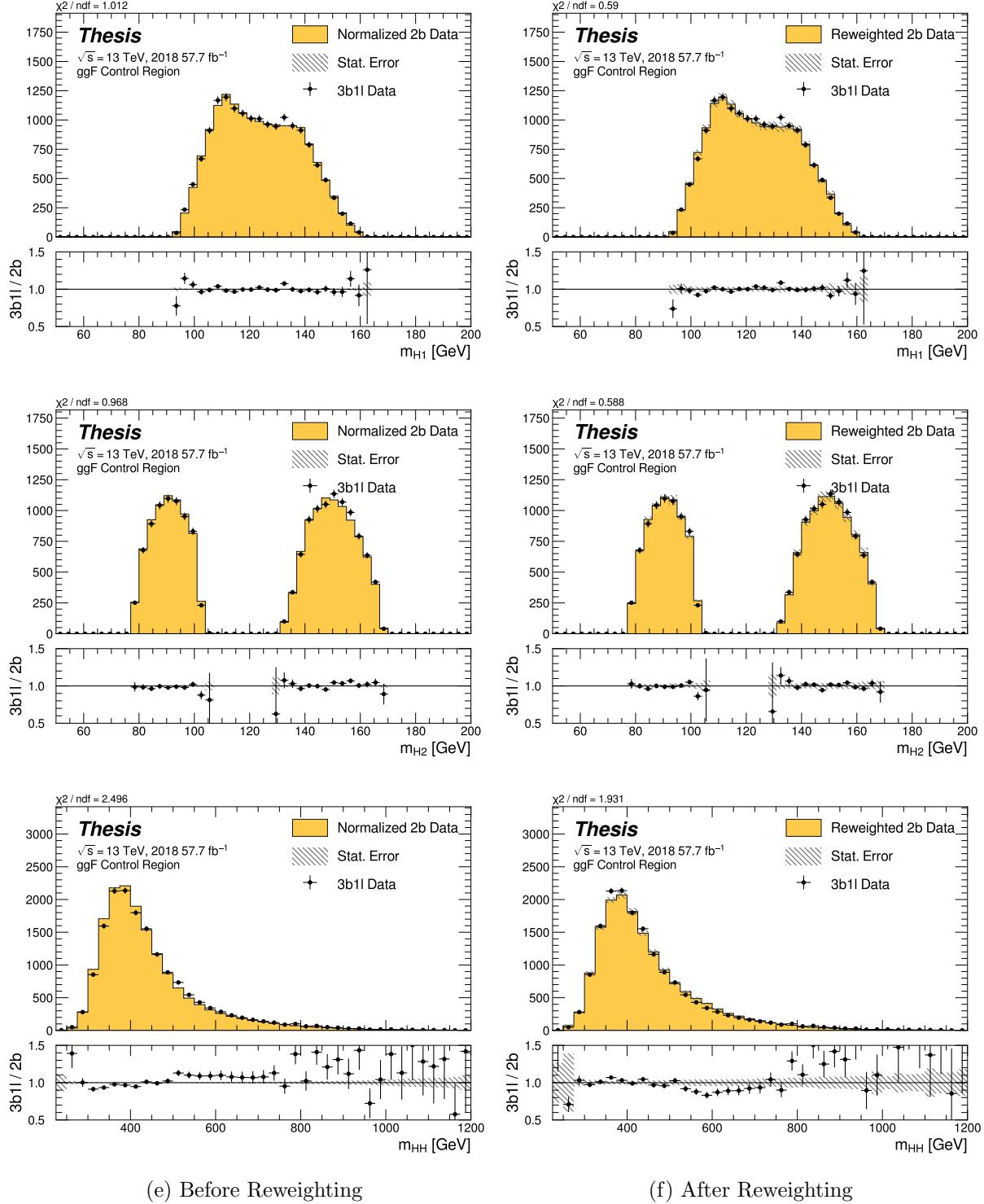


Figure 7.48: **Non-resonant Search (3b1l):** Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Control Region.

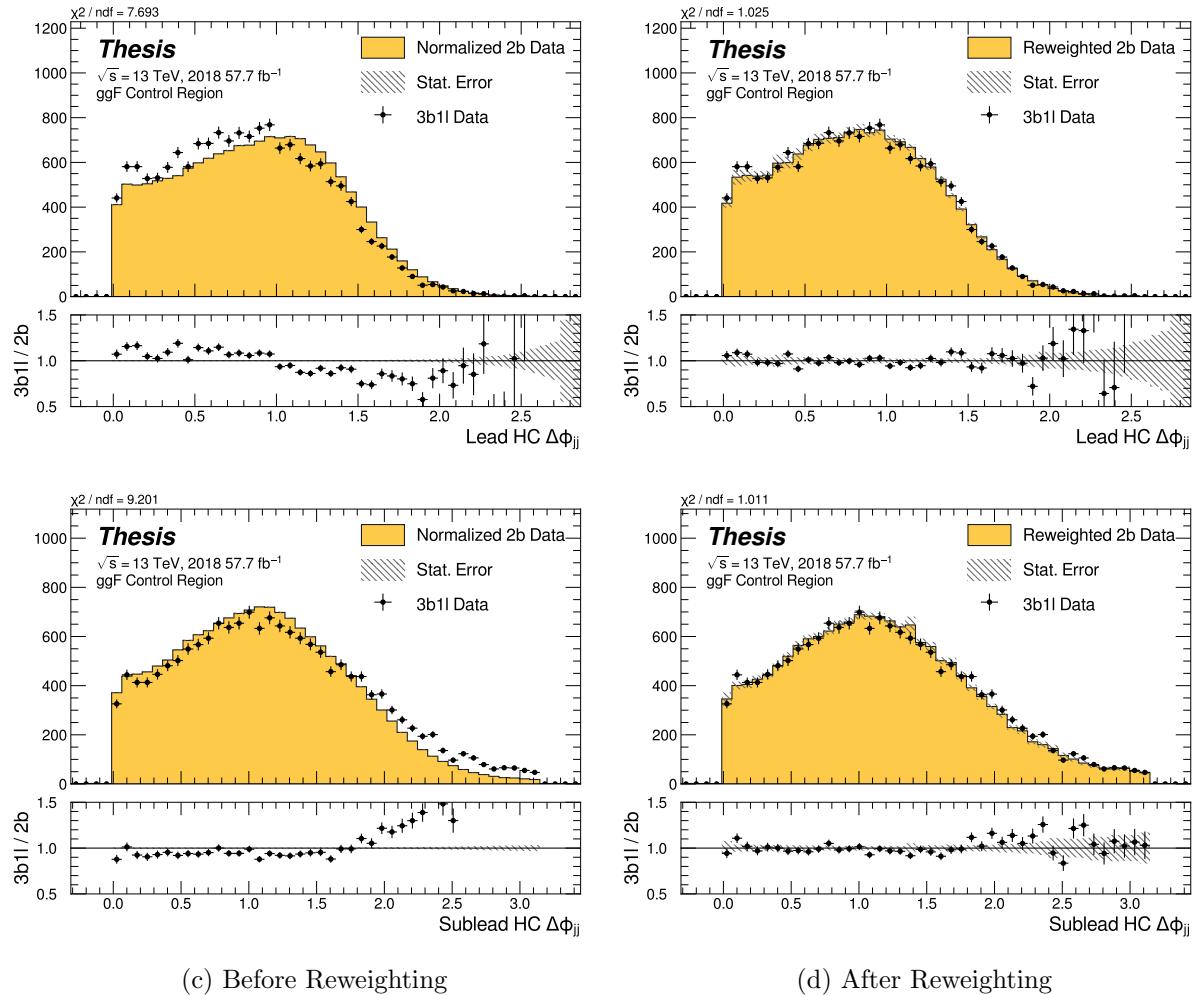


Figure 7.49: **Non-resonant Search (3b1l):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.

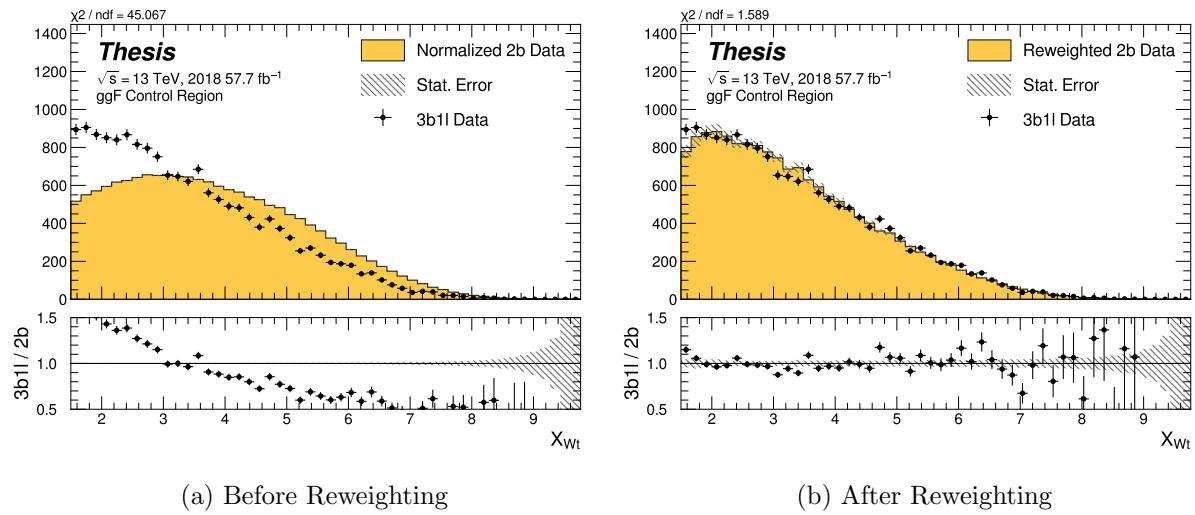


Figure 7.50: **Non-resonant Search (3b1l):** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 3b1l Control Region. Reweighting is done after the cut on this variable is applied.

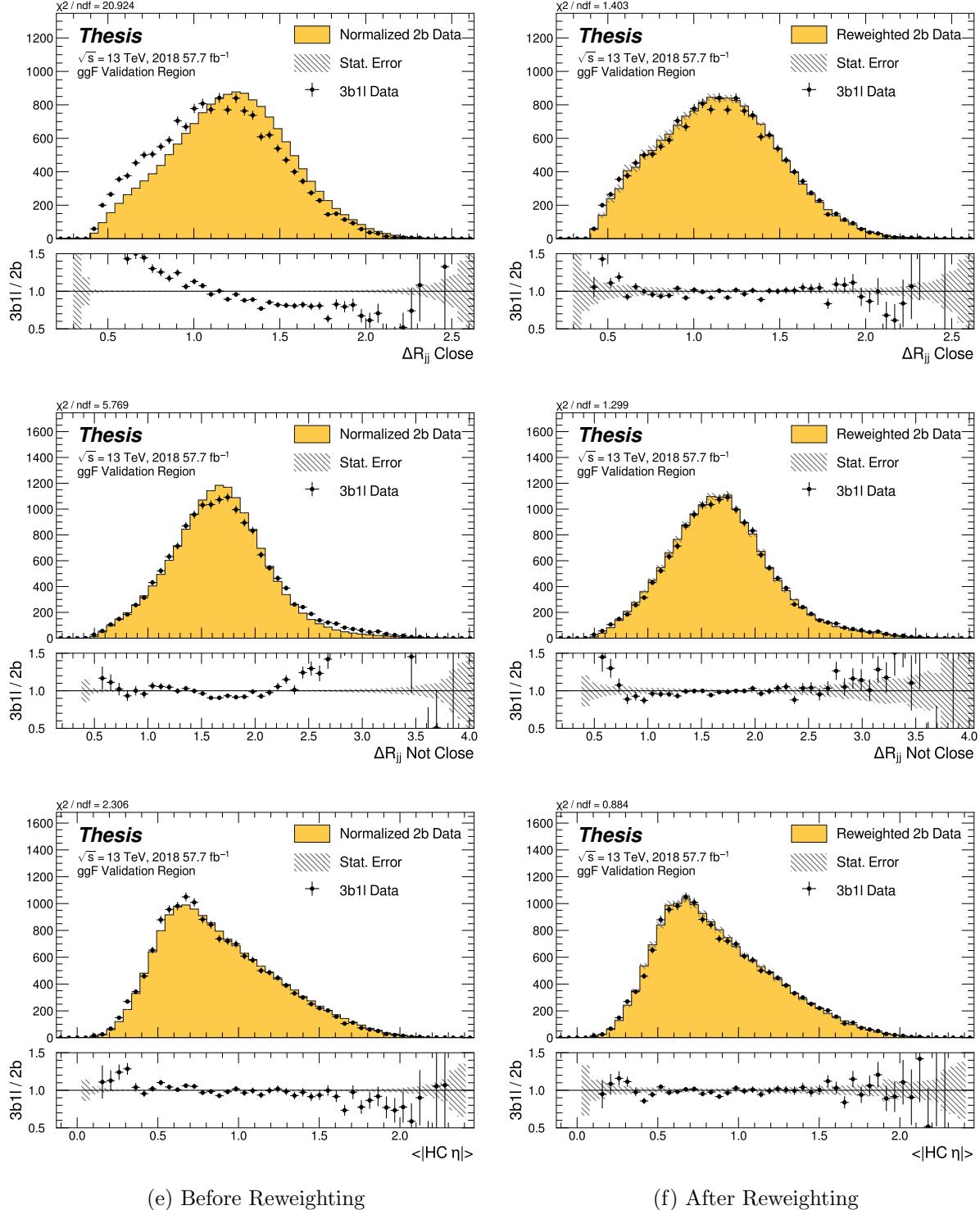


Figure 7.51: **Non-resonant Search (3b1l):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 3b1l Validation Region.

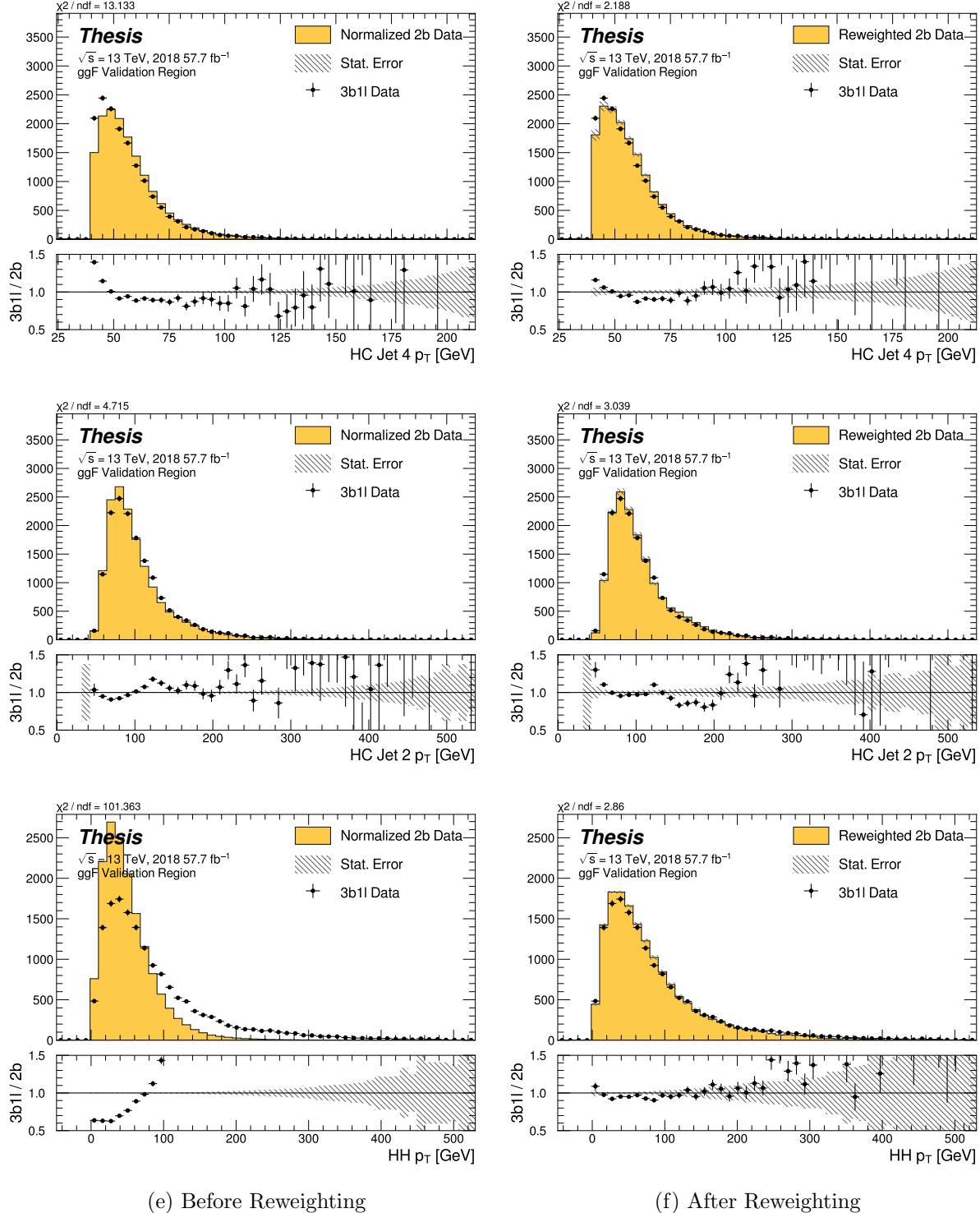


Figure 7.52: **Non-resonant Search (3b1l):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Validation Region.

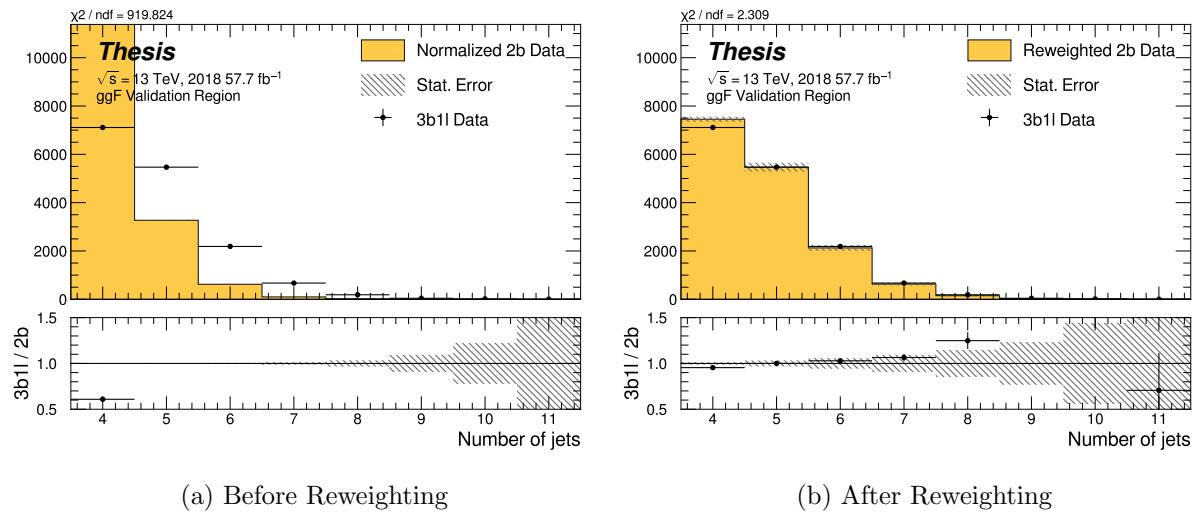


Figure 7.53: **Non-resonant Search (3b1l):** Distributions of the number of jets before and after CR derived reweighting for the 2018 3b1l Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

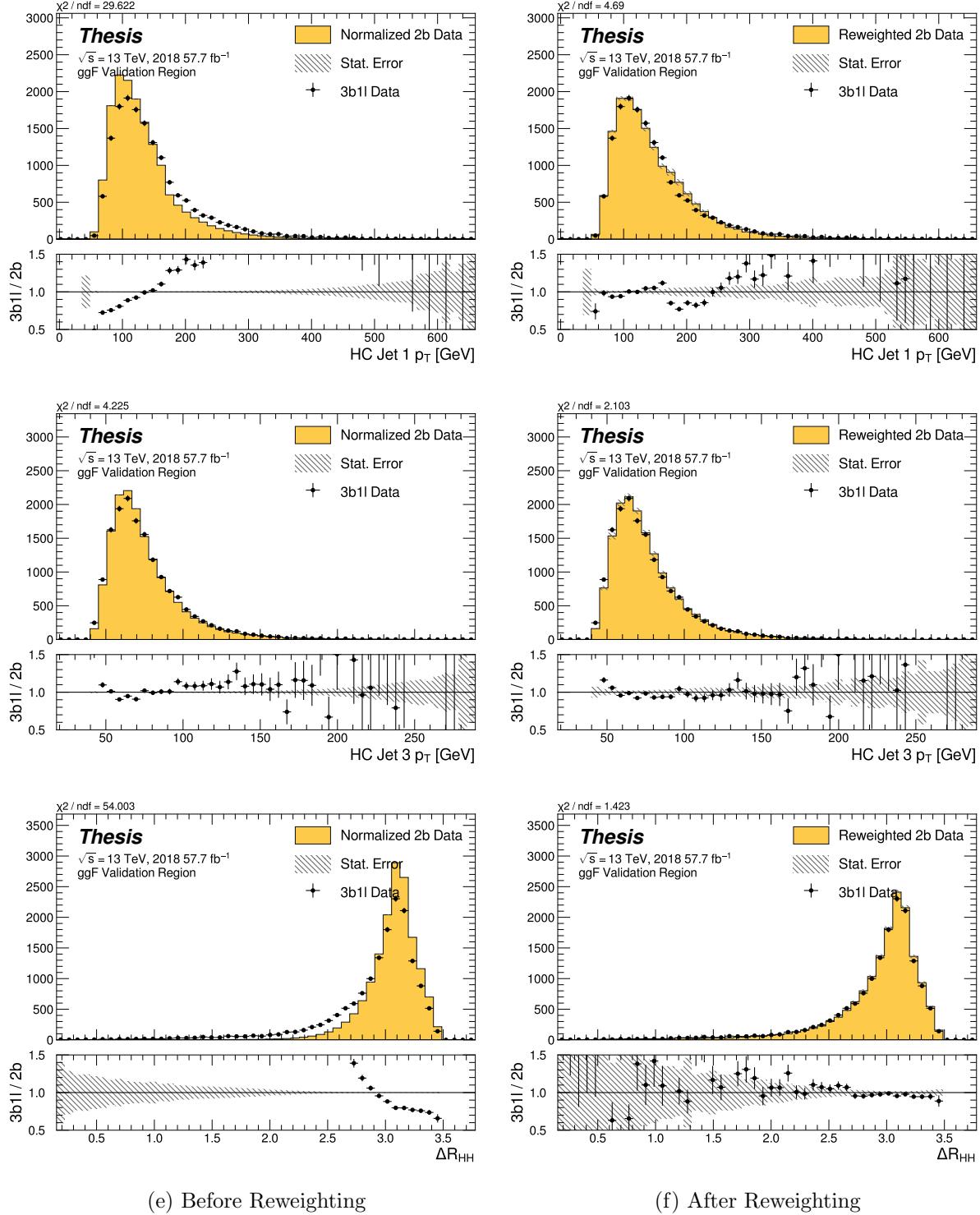


Figure 7.54: **Non-resonant Search (3b1l):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.

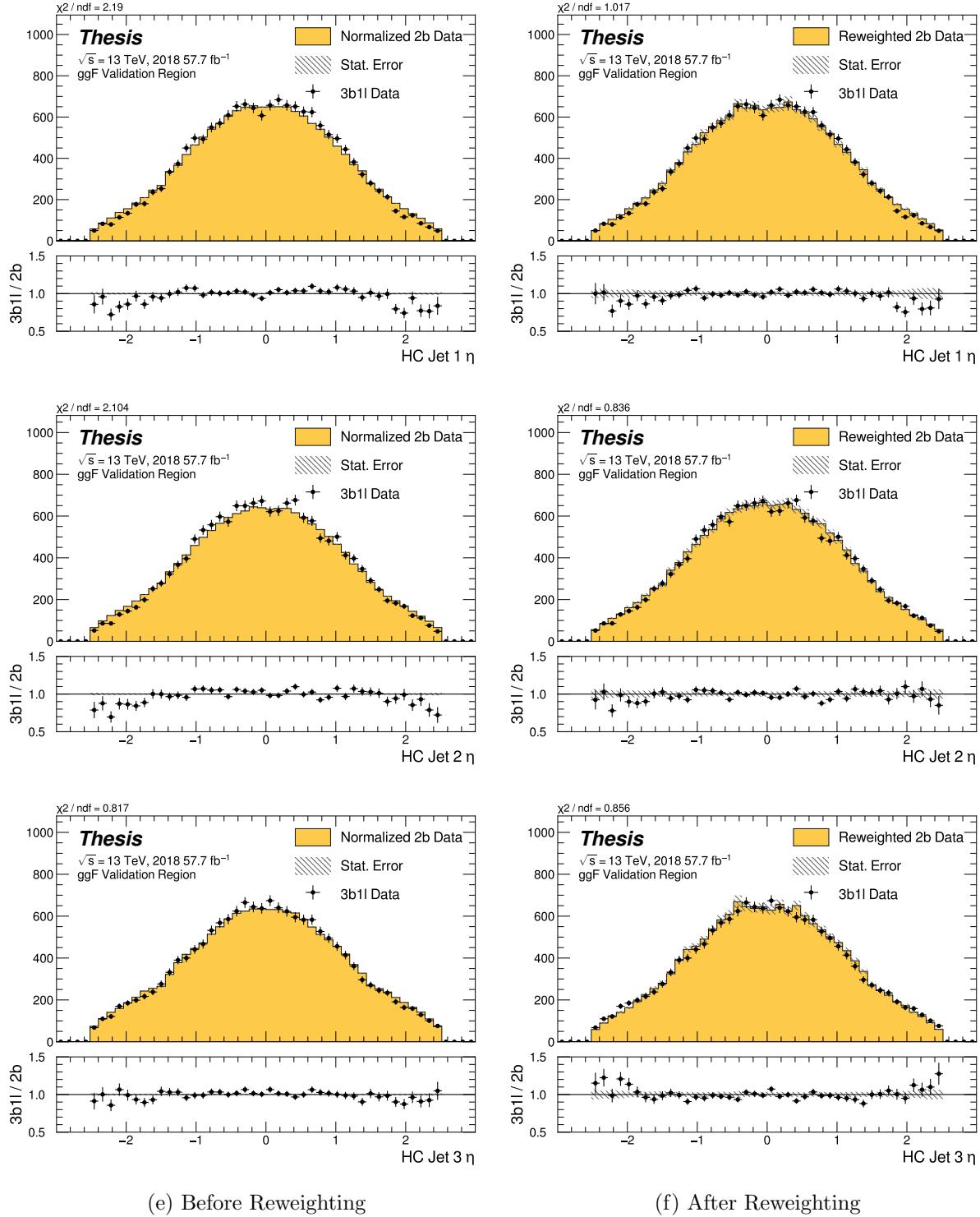


Figure 7.55: **Non-resonant Search (3b1l):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Validation Region.

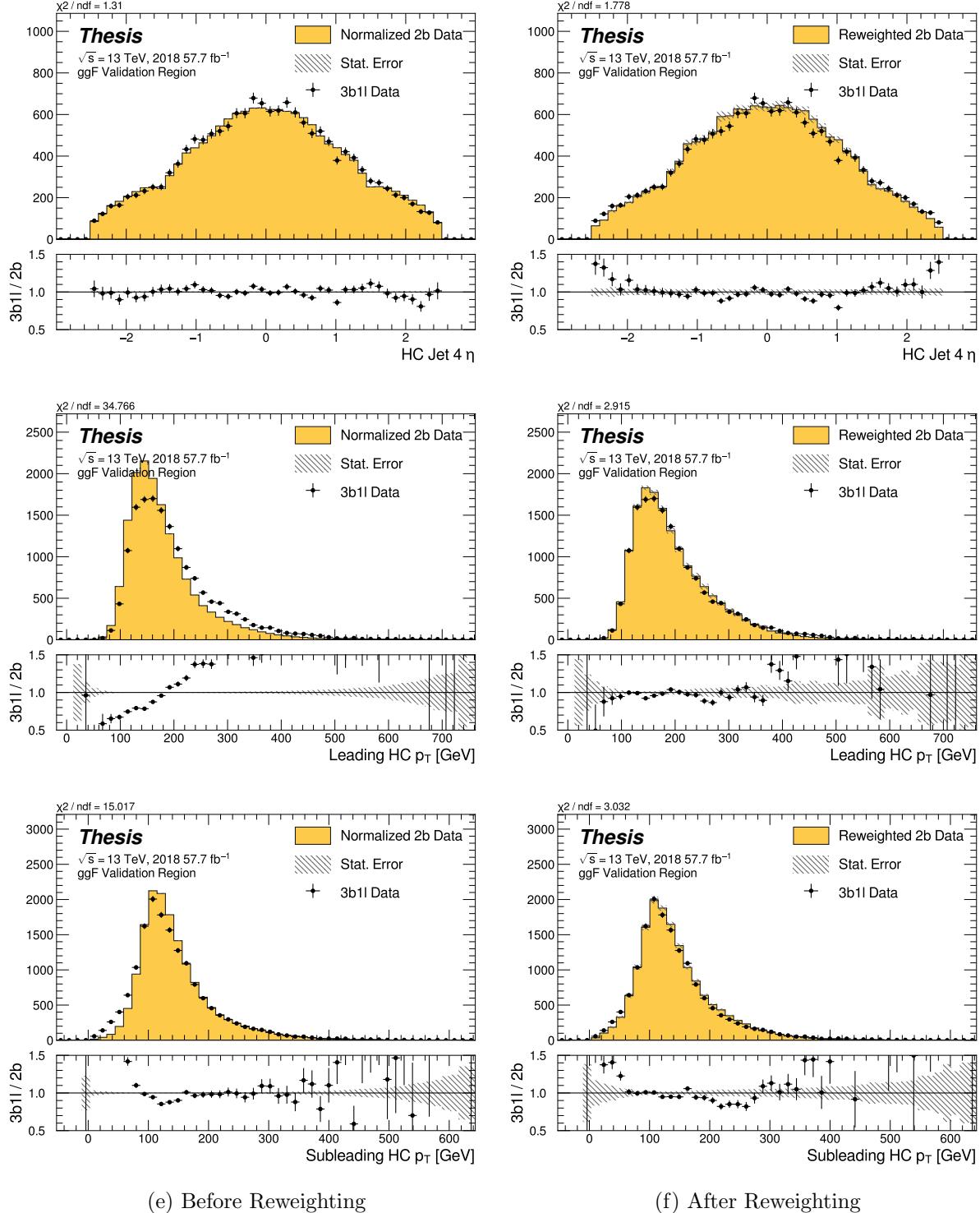


Figure 7.56: **Non-resonant Search (3b1l):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.

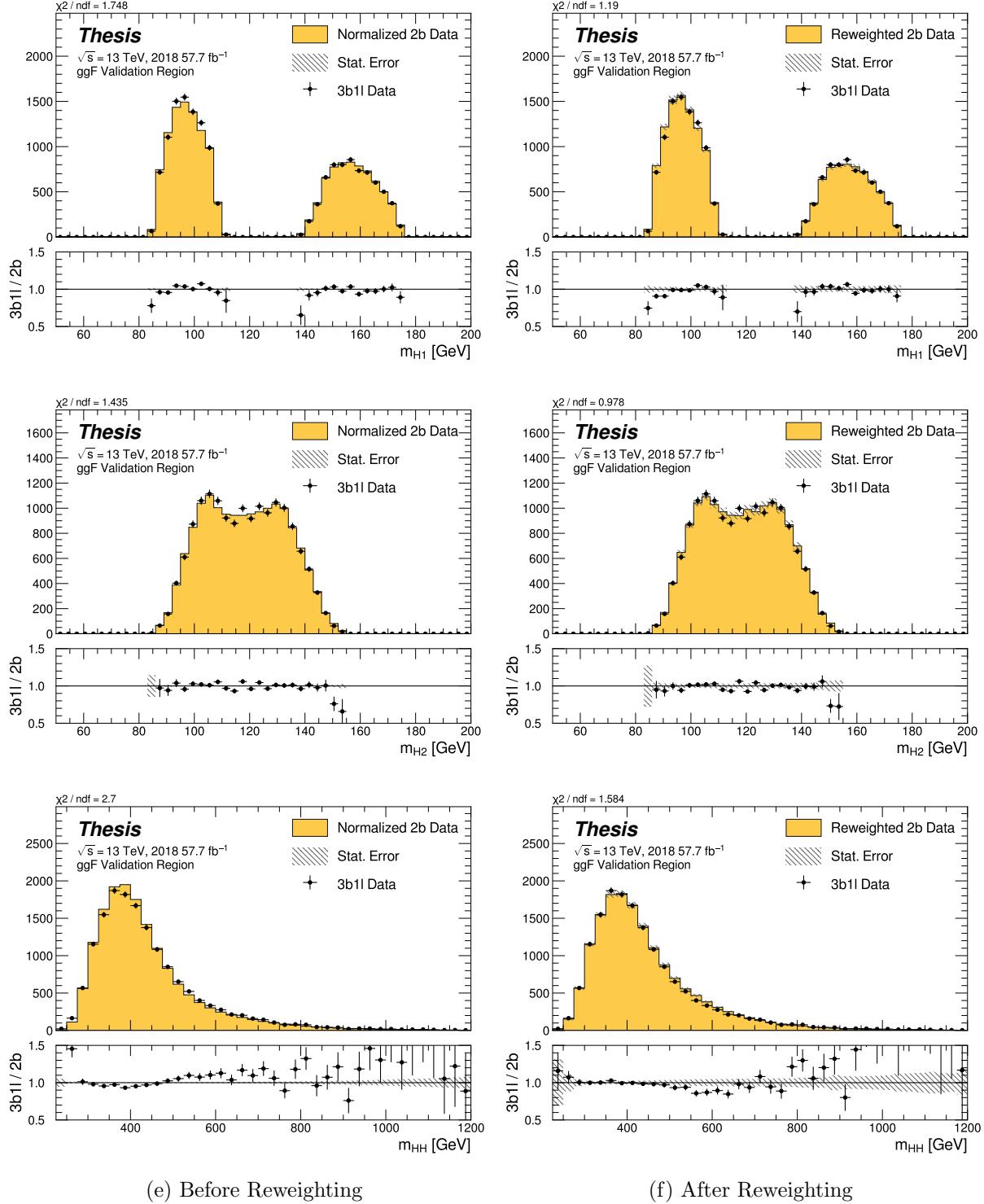


Figure 7.57: **Non-resonant Search (3b1l):** Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Validation Region.

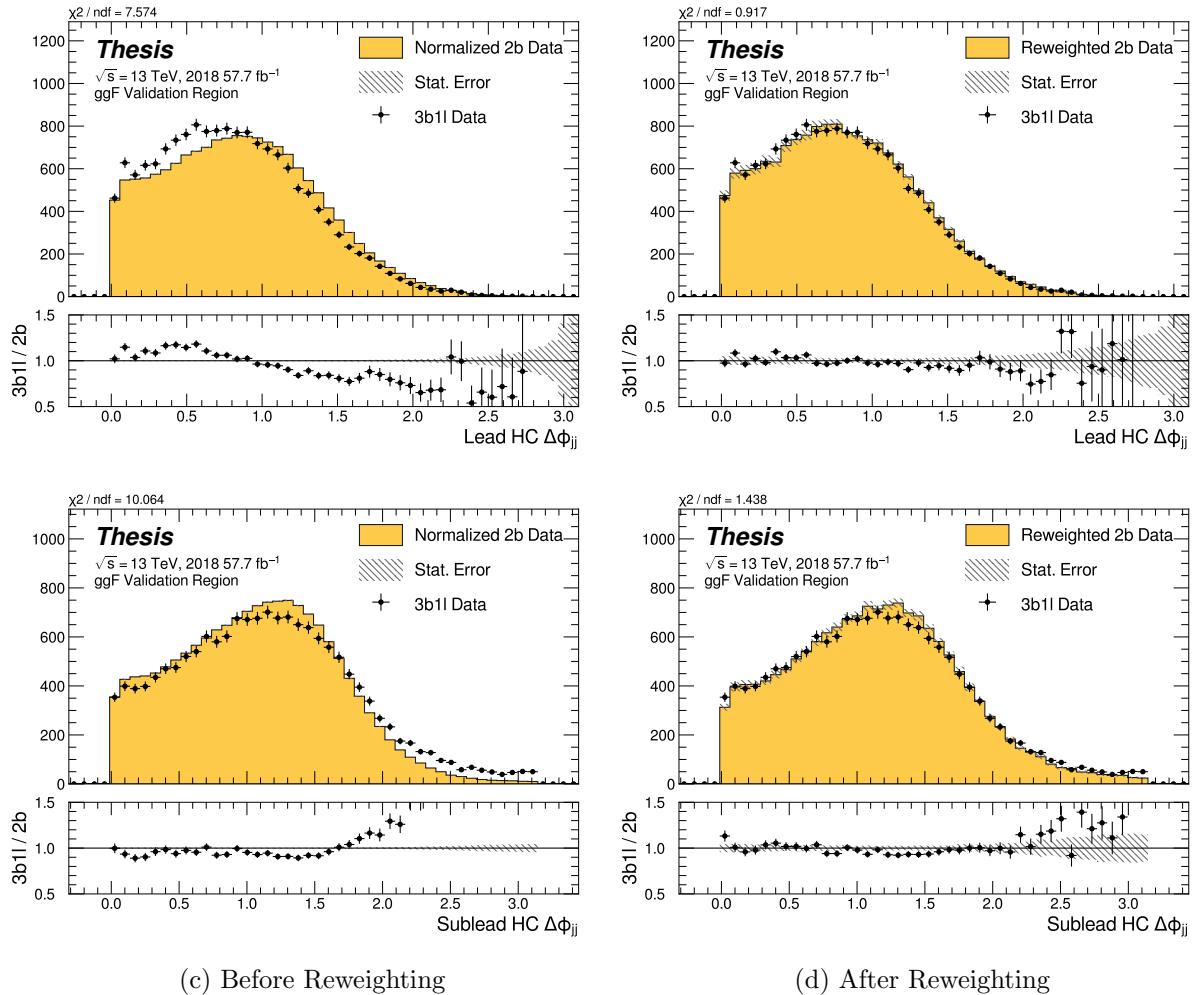


Figure 7.58: **Non-resonant Search (3b1l):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.

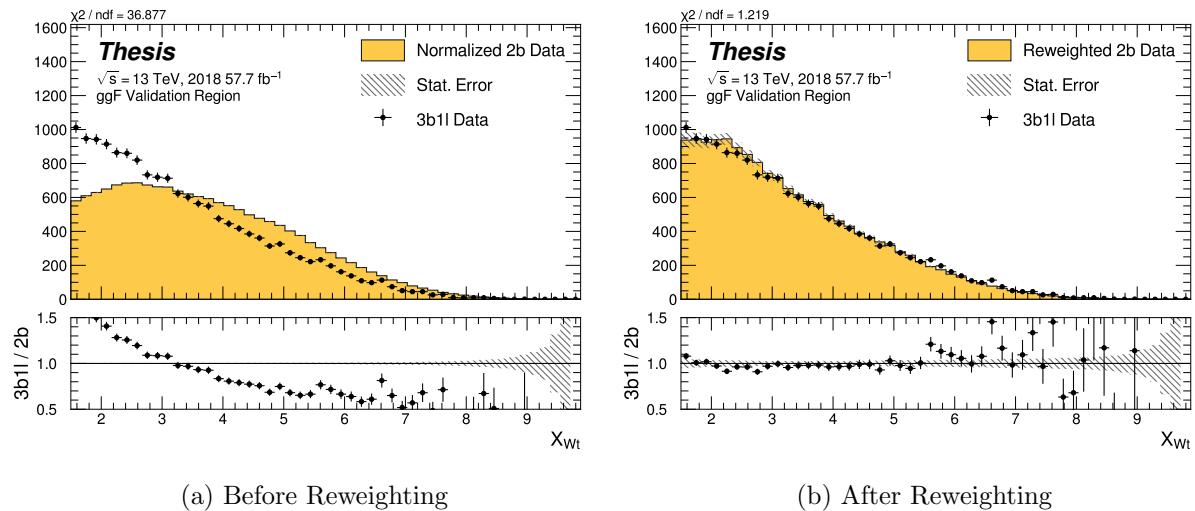


Figure 7.59: **Non-resonant Search (3b1l):** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 3b1l Validation Region. Reweighting is done after the cut on this variable is applied.

2014 **7.7 Uncertainties**

2015 A variety of uncertainties are assigned to account for known biases in the underlying methods,
2016 calibrations, and objects used for this analysis. The largest such uncertainty is associated with
2017 the kinematic bias inherent in deriving the background estimate away from the signal region.
2018 However, a statistical biasing of this same estimate has an effect of a similar magnitude.
2019 Additionally, due to the use of Monte Carlo for signal modelling and b -tagging calibration,
2020 uncertainties related to mismodellings in simulation must also be accounted for. These
2021 components, and their impact on this analysis, are described here in detail. Note that, while
2022 the Poisson error (from 2b data statistics) is negligible relative to the bootstrap error in
2023 the bulk of the distribution, it becomes relevant in the high m_{HH} tail. The final statistical
2024 uncertainty used for the limit setting is therefore the sum (in quadrature) of these two
2025 components.

2026 **7.7.1 Statistical Uncertainties and Bootstrapping**

2027 There are two components to the statistical error for the neural network background estimate.
2028 The first is standard Poisson error, i.e., a given bin, i , in the background histogram has value
2029 $n_i = \sum_{j \in i} w_j$, where w_j is the weight for an event j which falls in bin i . Standard techniques
2030 then result in statistical error $\delta n_i = \sqrt{\sum_{j \in i} w_j^2}$, which reduces to the familiar \sqrt{N} Poisson error
2031 when all w_j are equal to 1.

2032 However, this procedure does not take into account the statistical uncertainty on the
2033 w_j due to the finite training dataset. Due to the large size difference between the two tag
2034 and four tag datasets, it is the statistical uncertainty due to the four tag training data that
2035 dominates that on the background. A standard method for estimating this uncertainty is the
2036 bootstrap resampling technique [90]. Conceptually, a set of statistically equivalent sets is
2037 constructed by sampling with replacement from the original training set. The reweighting
2038 network is then trained on each of these separately, resulting in a set of statistically equivalent
2039 background estimates. Each of these sets is below referred to as a replica.

2040 In practice, as the original training set is large, the resampling procedure is able to
 2041 be simplified through the relation $\lim_{n \rightarrow \infty} \text{Binomial}(n, 1/n) = \text{Poisson}(1)$, which dictates that
 2042 sampling with replacement is approximately equivalent to applying a randomly distributed
 2043 integer weight to each event, drawn from a Poisson distribution with a mean of 1.

2044 Though the network configuration itself is the same for each bootstrap training, the
 2045 network initialization is allowed to vary. It should therefore be noted that the bootstrap
 2046 uncertainties implicitly capture the uncertainty due to this variation in addition to the
 2047 previously mentioned training set variation.

2048 The variation from this bootstrapping procedure is used to assign a bin-by-bin uncertainty
 2049 which is treated as a statistical uncertainty in the fit. Due to practical constraints, a
 2050 procedure for approximating the full bootstrap error band is developed which demonstrates
 2051 good agreement with the full bootstrap uncertainty. This procedure is described below.

2052 *Calculating the Bootstrap Error Band*

2053 The standard procedure to calculate the bootstrap uncertainty would proceed as follows: first,
 2054 each network trained on each bootstrap replica dataset would be used to produce a histogram
 2055 in the variable of interest. This would result in a set of replica histograms (e.g. for 100
 2056 bootstrap replicas, 100 histograms would be created). The nominal estimate would then be
 2057 the mean of bin values across these replica histograms, with errors set by the corresponding
 2058 standard deviation.

2059 In practice, such an approach is inflexible and demanding both in computation and in
 2060 storage, in so far as we would like to produce histograms in many variables, with a variety
 2061 of different cuts and binnings. This motivates a derivation based on event-level quantities.
 2062 However, due to non-trivial correlations between replica weights, simple linear propagation of
 2063 event weight variation is not correct.

2064 We therefore adopt an approach which has been empirically found to produce results
 2065 (for this analysis) in line with those produced by generating all of the histograms, as in the
 2066 standard procedure. This approach is described below. Note that, for robustness to outliers

2067 and weight distribution asymmetry, the median and interquartile range (IQR) are used for
 2068 the central value and width respectively (as opposed to the mean and standard deviation).

2069 The components involved in the calculation have been mentioned in Section 7.6 and are
 2070 as follows:

- 2071 1. Replica weight (w_i): weight predicted for a given event by a network trained on replica
 2072 dataset i .
- 2073 2. Replica norm (α_i): normalization factor for replica i . This normalizes the reweighting
 2074 prediction of the network trained on replica dataset i to match the correponding target
 2075 yield.
- 2076 3. Median weight (w_{med}): median weight for a given event across replica datasets, used
 2077 for the nominal estimate. Defined (for 100 bootstrap replicas) as

$$w_{med} \equiv \text{median}(\alpha_1 w_1, \dots, \alpha_{100} w_{100}) \quad (7.12)$$

- 2076 4. Normalization correction (α_{med}): normalization factor to match the predicted yield of
 2077 the median weights (w_{med}) to the target yield in the training region.

2078 As mentioned in Section 7.6, the *nominal estimate* is constructed from the set of median
 2079 weights and the normalization correction, i.e. $\alpha_{med} \cdot w_{med}$.

2080 For the bootstrap error band, a “varied” histogram is then generated by applying, for
 2081 each event, a weight equal to the median weight (with no normalization correction) plus half
 2082 the interquartile range of the replica weights: $w_{varied} = w_{med} + \frac{1}{2} \text{IQR}(w_1, \dots, w_{100})$.

2083 This varied histogram is scaled to match the yield of the nominal estimate. To account
 2084 for variation of the nominal estimate yield, a normalization variation is calculated from the
 2085 interquartile range of the replica norms: $\frac{1}{2} \text{IQR}(\alpha_1, \dots, \alpha_{100})$. This variation, multiplied into
 2086 the nominal estimate, is used to set a baseline for the varied histogram described above.

Denoting $H(\text{weights})$ as a histogram constructed from a given set of weights, $Y(\text{weights})$ as the predicted yield for a given set of weights, the final varied histogram is thus:

$$H(w_{med} + \frac{1}{2} \text{IQR}(w_1, \dots, w_{100})) \cdot \frac{Y(\alpha_{med} w_{med})}{Y(w_{med} + \frac{1}{2} \text{IQR}(w_1, \dots, w_{100}))} + \frac{1}{2} \text{IQR}(\alpha_1, \dots, \alpha_{100}) \cdot H(\alpha_{med} w_{med}) \quad (7.13)$$

where the first term roughly describes the behaviour of the bootstrap variation across

the distribution of the variable of interest while the second term describes the normalization

variation of the bootstrap replicas.

The difference between the varied histogram and the nominal histogram is then taken to

be the bootstrap statistical uncertainty on the nominal histogram.

Figure 7.60 demonstrates how each of the components described above contribute to the

uncertainty envelope for the non-resonant 2017 Control Region and compares this approximate

band to the variation of histograms from individual bootstrap estimates. The error band

constructed from the above procedure is seen to provide a good description of the bootstrap

variation.

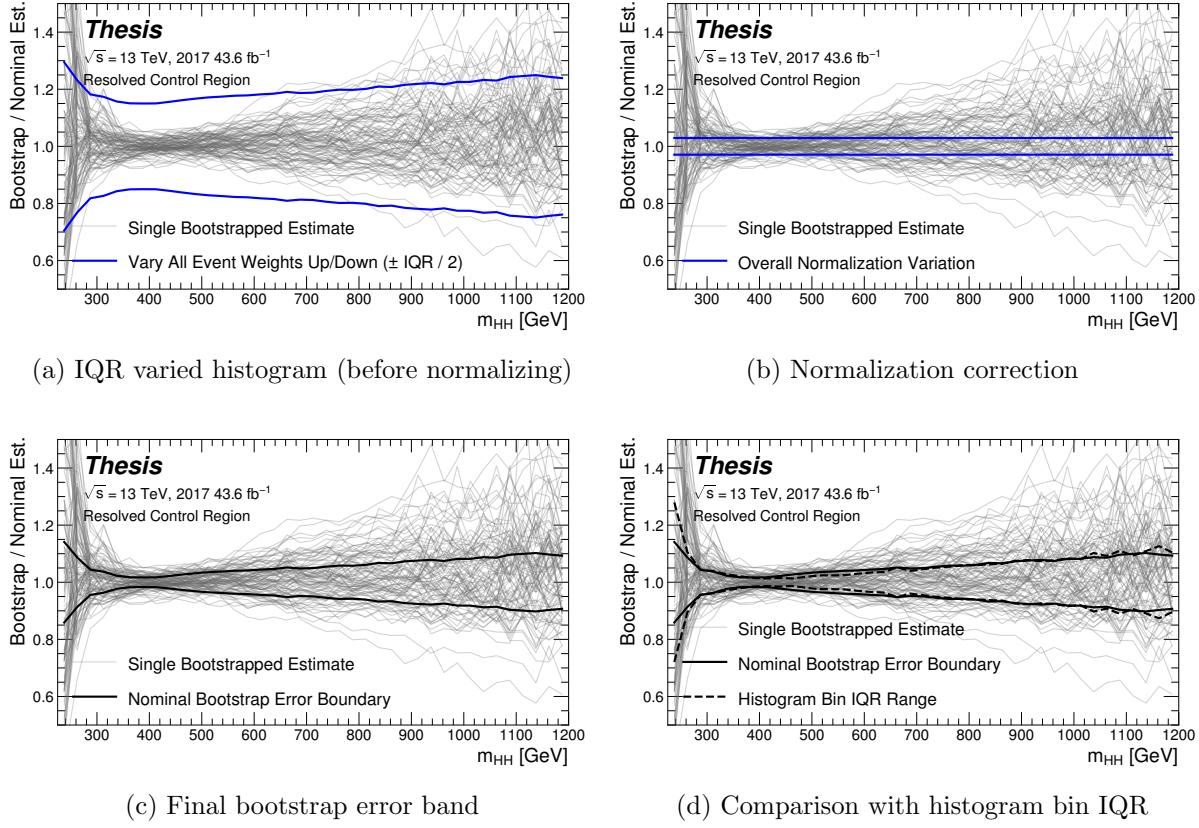


Figure 7.60: Illustration of the approximate bootstrap band procedure, shown as a ratio to the nominal estimate for the 2017 non-resonant background estimate. Each grey line is from the m_{HH} prediction for a single bootstrap training. Figure 7.60(a) shows the variation histograms constructed from median weight \pm the IQR of the replica weights. It can be seen that this captures the rough shape of the bootstrap envelope, but is not good estimate for the overall magnitude of the variation. Figure 7.60(b) demonstrates the applied normalization correction, and Figure 7.60(c) shows the final band (normalized Figure 7.60(a) + Figure 7.60(b)). Comparing this with the IQR variation for the prediction from each bootstrap in each bin in Figure 7.60(d), the approximate envelope describes a very similar variation.

2097 7.7.2 *Background Shape Uncertainties*

2098 To account for the systematic bias associated with deriving the reweighting function in the
2099 control region and extrapolating to the signal region, an alternative background model is
2100 derived in the validation region. Because of the fully data-driven nature of the background
2101 model, this is an uncertainty assessed on the full background. The alternative model and
2102 the baseline are consistent with the observed data in their training regions, and differences
2103 between the alternative and baseline models are used to define a shape uncertainty on the
2104 m_{HH} spectrum, with a two-sided uncertainty defined by symmetrizing the difference about
2105 the baseline.

2106 For the resonant analysis, this uncertainty is split into two components to allow for two
2107 independent variations of the m_{HH} spectrum: : a low- H_T and a high- H_T component, where
2108 H_T is the scalar sum of the p_T of the four jets constituting the Higgs boson candidates, and
2109 serves as a proxy for m_{HH} , while avoiding introducing a sharp discontinuity. The boundary
2110 value is 300 GeV. The low- H_T shape uncertainty primarily affects the m_{HH} spectrum below
2111 400 GeV (close to the kinematic threshold) by up to around 5%, and the high- H_T uncertainty
2112 mainly m_{HH} above this by up to around 20% relative to nominal. These separate m_{HH}
2113 regimes are by design – the H_T split is introduced to prevent low mass bins from constraining
2114 the high mass uncertainty and vice-versa.

2115 This was the *status quo* shape uncertainty decomposition from the Early Run 2 analysis.
2116 A decomposition in terms of orthogonal polynomials, which would provide increased flexibility,
2117 was also evaluated. This study revealed that both decompositions are able to account for the
2118 systematic deviations between four tag data and the background estimate (evaluated in the
2119 kinematic validation region), and produce almost identical limits. The simpler *status quo*
2120 decomposition is therefore kept.

2121 For the non-resonant analysis, the quadrant nature of the background estimation leads to
2122 a natural breakdown of the nuisance parameters: quadrants are defined in the signal region
2123 along the same axes as those used for the control and validation region definitions. Variations

2124 are then assessed in each of these signal region quadrants, corresponding to regions that
 2125 are “closer to” and “further away from” the nominal and alternate estimate regions, fully
 2126 leveraging the power of the two equivalent but systematically different estimates.

2127 Figure 7.61 shows an example of the variation in each H_T region for the 2018 resonant
 2128 analysis. Figure 7.62 shows the example quadrant variation for the 2018 4 b non-resonant
 analysis.

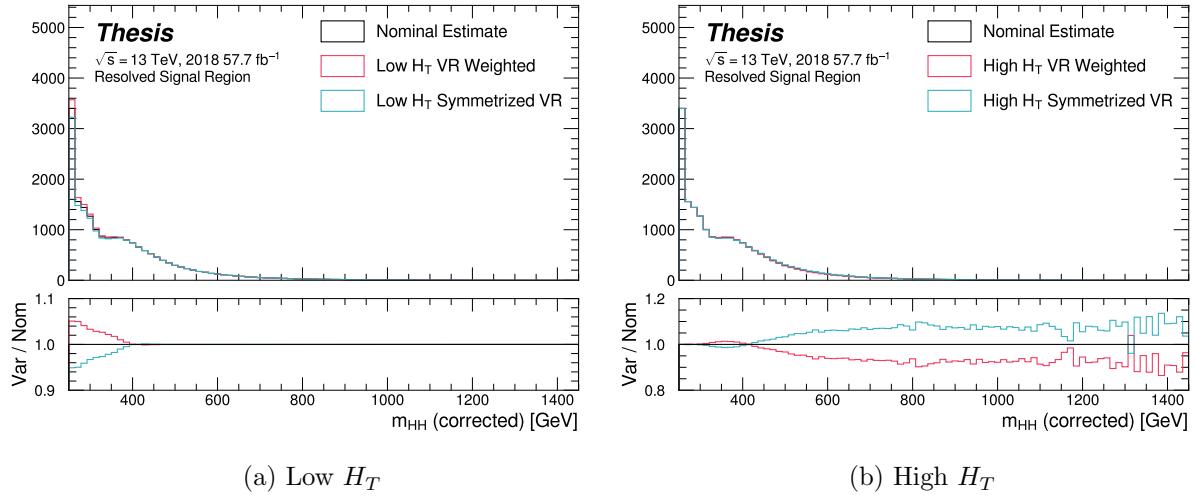
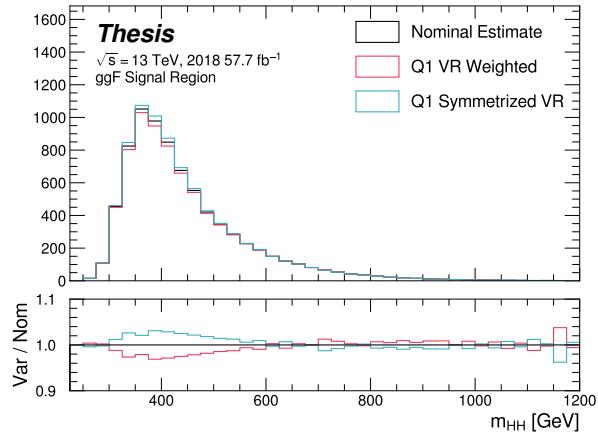
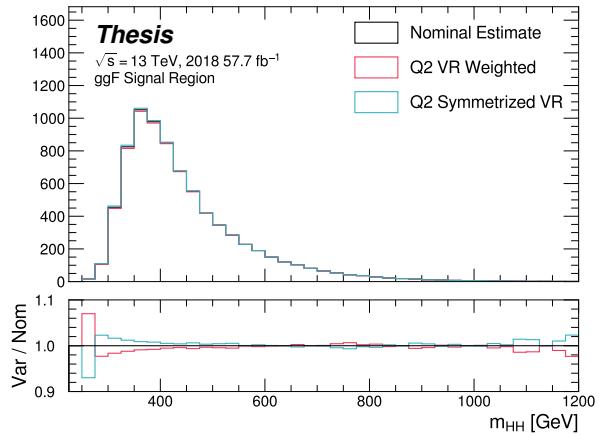


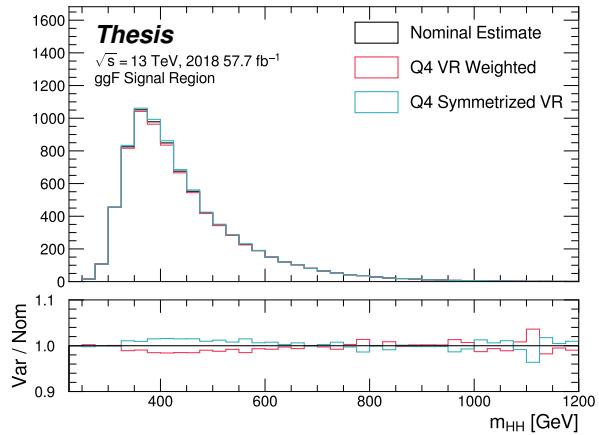
Figure 7.61: **Resonant Search:** Example of CR vs VR variation in each H_T region for 2018.
 The variation nicely factorizes into low and high mass components.



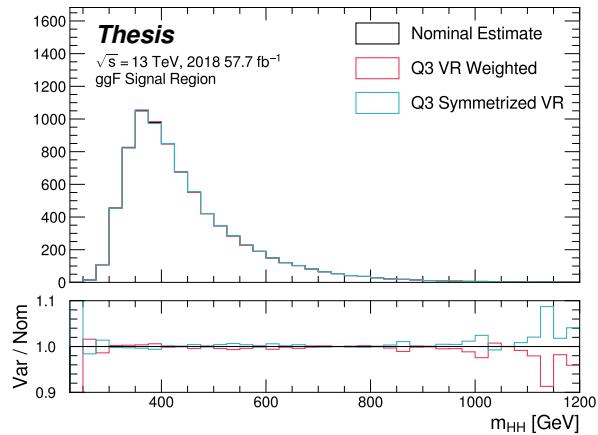
(a) Q1 (top)



(b) Q2 (left)



(c) Q4 (right)



(d) Q3 (bottom)

Figure 7.62: **Non-resonant Search (4b):** Example of CR vs VR variation in each signal region quadrant for 2018. Significantly different behavior is seen between quadrants, with the largest variation in quadrant 1 and the smallest in quadrant 4.

2130 7.7.3 *Detector Modelling and Reconstruction Uncertainties*

2131 Detector modelling and reconstruction uncertainties account for Monte Carlo simulation not
2132 being a faithful representation of real data as a result of mismodelling of the detector and
2133 differential performance of algorithms on simulation compared to data. In this analysis they
2134 consist of uncertainties related to jet properties, and uncertainties stemming from the flavour
2135 tagging procedure. The background modelling in this analysis is fully data-driven. As a
2136 result, these uncertainties are applied only to the signal simulation.

2137 The jet uncertainties are implemented as variations of the jet properties themselves. The
2138 category reduction (with ~ 30 nuisance parameters) is used for jet energy scale uncertainties
2139 and the FullJER configuration is used for jet energy resolution uncertainties (14 nuisance
2140 parameters). This is to preserve the ability to meaningfully statistically combine the results
2141 of this analysis with other di-Higgs analyses. The flavour tagging uncertainties meanwhile
2142 are implemented as scale factors applied to the Monte Carlo event weights.

2143 A systematic related to the PtReco b -jet energy correction has been studied in the
2144 $HH \rightarrow \gamma\gamma b\bar{b}$ analysis [91] and found to be negligible compared to JER. Following this
2145 example, such a systematic is therefore neglected here.

2146 7.7.4 *Trigger Uncertainties*

2147 Trigger uncertainties stem from imperfect knowledge of the ratio between the efficiency of a
2148 given trigger in data to its efficiency in Monte Carlo simulation. This ratio is applied as a
2149 scale factor to all simulated events (as described in Section ??), with the systematic variations
2150 produced by varying the scale factor up or down by one sigma.

2151 7.7.5 *Theoretical Uncertainties*

2152 The theoretical uncertainties on the acceptance times efficiency ($A \times \varepsilon$) are evaluated by
2153 analysis of specially-generated, particle-level signal samples. The generation of these samples
2154 follows the configuration of the baseline samples, but with modifications to probe the following

2155 theoretical uncertainties: uncertainties in the parton density functions (PDFs); uncertainties
 2156 due to missing higher order terms in the matrix elements; and uncertainties in the modelling
 2157 of the underlying event, which includes multi-parton interactions, of hadronic showers and of
 2158 initial and final state radiation.

2159 Uncertainties due to modelling of the parton shower and the underlying event (including
 2160 multi-parton interactions) are evaluated by switching the MC generator used. For the scalar
 2161 samples, this means switching from Herwig7.7.1.3 to Pythia 8.235. Figure ?? shows the
 2162 impact of these variations on the signal acceptance for two resonance masses: 500 GeV and
 2163 1 TeV, covering the range of the resolved analysis. No significant dependence on the variable
 2164 of interest, m_{HH} , is observed. The disagreement observed in the tails of $p_T(hh)$ and the
 2165 number of jets multiplicities is negligible with respect to the final signal acceptance. A 5%
 2166 flat systematic uncertainty is assigned to all signal samples, extracted from the acceptance
 2167 comparison for the full 4-tag selection, as seen in Figs. ?? and ??.

2168 To evaluate the potential effect of missing higher order terms in the matrix element, the
 2169 renormalization and factorization scales used in the signal generation were varied coherently
 2170 by factors of $0.5 \times$ and $2 \times$ for the signals. The alternative weights were generated as described
 2171 on the TWiki [here](#), applying on-the-fly variations using the ATLAS MadGraphControl
 2172 framework. These weights correspond to variations of the scales either together or separately
 2173 up and down by a factor of two. Seven-point scale variations are considered: $(\mu_R, \mu_F) = (0.5,$
 2174 $0.5), (1, 0.5), (0.5, 1), (1, 1), (2, 1), (1, 2), (2, 2)$. The scale uncertainties are combined by
 2175 taking an envelope of all of the uncertainties. These uncertainties are evaluated to be less
 2176 than $\pm 1\%$, thus neglected.

2177 PDF uncertainties are evaluated using the PDF4LHC15_nlo_mc set, which combined
 2178 CT14, MMHT14 and NNPDF3.0 PDF sets. The uncertainty is evaluated by calculating
 2179 the acceptance for each PDF replica. The standard deviation of these acceptance values
 2180 divided by the baseline acceptance is taken as the PDF uncertainty. For each mass point the
 2181 distribution of their corresponding ration is compatible with a Gaussian centered one. The
 2182 measured uncertainty in acceptance due to PDF uncertainties is less than $\pm 1\%$ across the

2183 full mass range considered for the analysis. For this reason, it is neglected in the statistical
2184 analysis described in Section ??.

2185 These uncertainties are implemented in the final statistical analysis as normalization
2186 uncertainties on the signals, with the value taken from the polynomial fit. This smooths out
2187 statistical fluctuations and allows interpolation between the generated mass points, if needed.

2188 The results for the non-resonant analysis presented here are preliminary and only include
2189 background systematics. However, these are expected to be by far the dominant uncertainties,
2190 and should therefore be reflective of the final results.

2191 **7.8 Background Validation**

2192 In addition to checking the performance of the background estimate in the control and
2193 validation regions, a variety of alternative selections are defined to allow for a full “dress
2194 rehearsal” of the background estimation procedure.

2195 Both the resonant and non-resonant analyses make use of a *reversed* $\Delta\eta$ region, in which
2196 the kinematic cut on $\Delta\eta_{HH}$ is reversed, so that events are required to have $\Delta\eta_{HH} > 1.5$.
2197 This is orthogonal to the nominal signal region and has minimal sensitivity, allowing for the
2198 comparison of the background estimate $4b$ data in the corresponding “signal region”. For
2199 this validation, a new reweighting is trained following nominal procedures, but entirely in the
2200 $\Delta\eta_{HH} > 1.5$ region.

2201 The non-resonant analysis additionally makes use of the $3b + 1$ fail region mentioned
2202 above, which again is orthogonal to the nominal signal regions and has minimal sensitivity.
2203 The reweighting in this case is between $2b$ and $3b + 1$ fail events rather than between $2b$
2204 and $3b + 1$ loose or $2b$ and $4b$. However, the kinematic selections of signal region events are
2205 otherwise identical, allowing for a complementary test of the background estimate.

2206 *TODO: Add shifted regions if they’re ready*

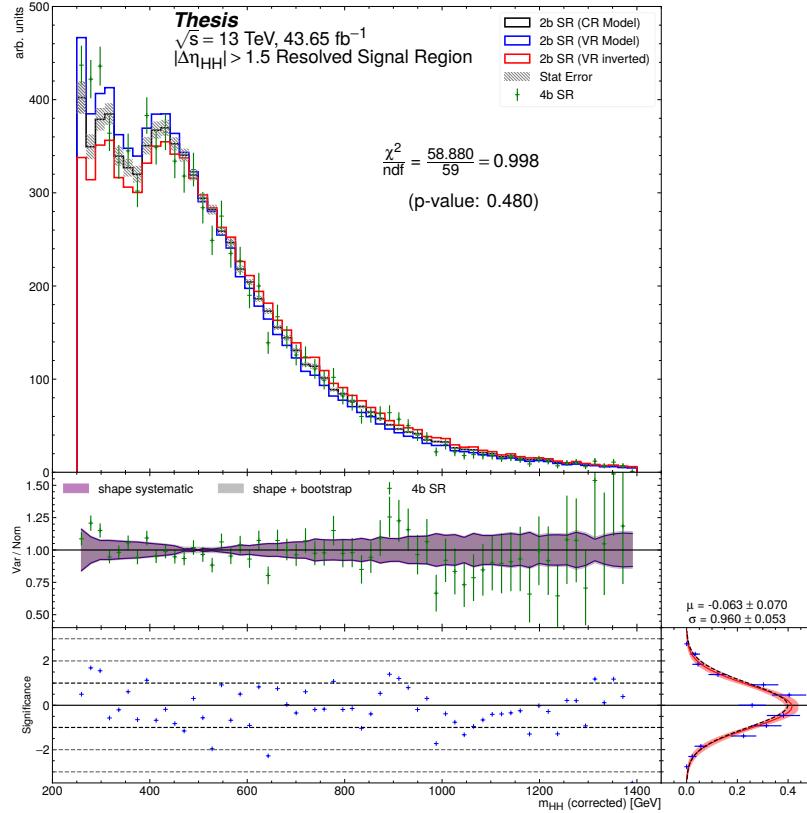


Figure 7.63: **Resonant Search:** Performance of the background estimation method in the resonant analysis reversed $\Delta\eta_{HH}$ kinematic signal region. A new background estimate is trained following nominal procedures entirely within the reversed $\Delta\eta_{HH}$ region, and the resulting model, including uncertainties, is compared with $4b$ data in the corresponding signal region. Good agreement is shown. The quoted p -value uses the χ^2 test statistic, and demonstrates no evidence that the data differs from the assessed background.

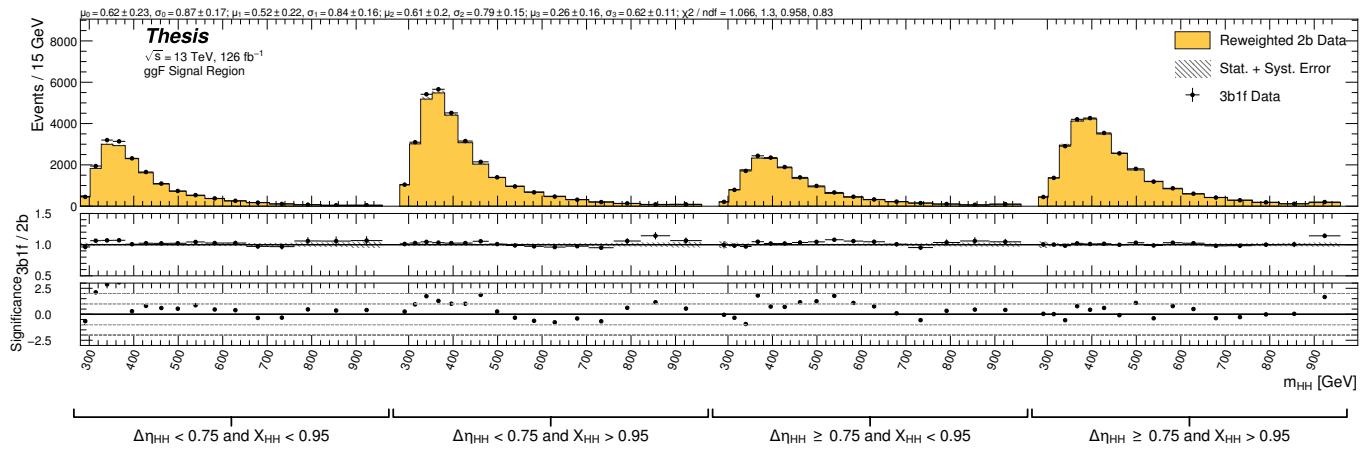


Figure 7.64: **Non-resonant Search:** Performance of the background estimation method in the $3b + 1$ fail validation region. A new background estimate is trained following nominal procedures but with a reweighting from $2b$ to $3b + 1$ fail events. Generally good agreement is seen, though there is some deviation at very low masses in the low $\Delta\eta_{HH}$ low X_{HH} category.

2207 **7.9 Overview of Other $b\bar{b}b\bar{b}$ Channels**

2208 The results discussed above have been developed in conjunction with (1) a boosted channel for
2209 the resonant search and (2) a vector boson fusion (VBF) channel for the non-resonant search.
2210 Detailed discussions of these two channels are beyond the scope of this thesis. However,
2211 a combined set of results is presented below (*TODO: or will be combined for VBF?*). We
2212 therefore briefly summarize the analyses here.

2213 **7.9.1 Resonant: Boosted Channel**

2214 The boosted analysis selection targets resonance masses from 900 GeV to 5 TeV. In such
2215 events, H decays have a high Lorentz boost, such that the $b\bar{b}$ decays are very collimated. The
2216 resolved analysis fails to reconstruct such HH events, as the $R = 0.4$ jets start to overlap.

2217 The boosted analysis instead reconstructs H decays as large radius, $R = 1.0$ jets, with
2218 corresponding b -quarks identified with variable radius subjets, that is jets with a radius that
2219 scales as ρ/p_T , the p_T is that of the jet in question, and ρ is a fixed parameter, here chosen
2220 to be 30 GeV, which is optimized to maintain truth-level double b-labelling efficiency across
2221 the full range of Higgs jet p_T *TODO: cite: <https://cds.cern.ch/record/2268678>.*

2222 Due to limited boosted b-tagging efficiency *TODO: cite* and to maintain sensitivity even
2223 when b -jets are highly collimated, the boosted analysis is divided into three categories based
2224 on the number of b -tagged jets associated to each large radius jet:

- 2225 • 4 b category: two b -tagged jets in each
- 2226 • 2 $b - 1$ category: two b -tagged jets in one, one in the other
- 2227 • 1 $b - 1$ category: one b -tagged jet in each

2228 The analysis then proceeds in each of these categories. *TODO: what other boosted details?*
2229 The resolved and boosted channels are combined for resonance masses from 900 GeV to
2230 1.5 TeV inclusive. To keep the channels statistically independent, the boosted channel vetos

2231 events passing the resolved analysis selection.

2232 *7.9.2 Non-resonant: VBF Channel*

2233 The vector boson fusion channel is only considered for the non-resonant search. While the
2234 sensitivity is in general much more limited than the gluon-gluon fusion analysis due to the
2235 much smaller production cross section, VBF is sensitive to a variety of Beyond the Standard
2236 Model physics, both complementary and orthogonal to the theoretical scope of gluon-gluon
2237 fusion. *TODO: I'll probably mention more details in the pheno section*

2238 The VBF channel proceeds very similarly to the ggF, with the primary differences being
2239 the kinematic selections and the categorization.

2240 *TODO: fill in kinematics*

2241 Note that the background estimation is inherited from the resonant and ggF analyses, an
2242 ancillary, but significant, contribution of this thesis work.

2243 **7.10 m_{HH} Distributions**

2244 *7.10.1 Resonant Search*

2245 The final discriminant used for the resonant search is corrected m_{HH} . Histogram binning
2246 was optimized for the resonant search to be 84 equal width bins from 250 GeV to 1450 GeV,
2247 corresponding to a bin width of 14.3 GeV, and overflow events (events above 1450 GeV) are
2248 included in the last bin. A demonstration of the performance of the reweighting on this
2249 distribution is shown in Figure 7.65 for the control region and Figure 7.66 for the validation
2250 region. A background-only profile likelihood fit is run for the distribution in the
2251 signal region, and results with spin-0 signals overlaid are shown in Figure 7.67. Note that the
2252 plots show the sum across all years, but the signal extraction fit and background estimate
2253 are run with the years separately. Agreement is generally good throughout.

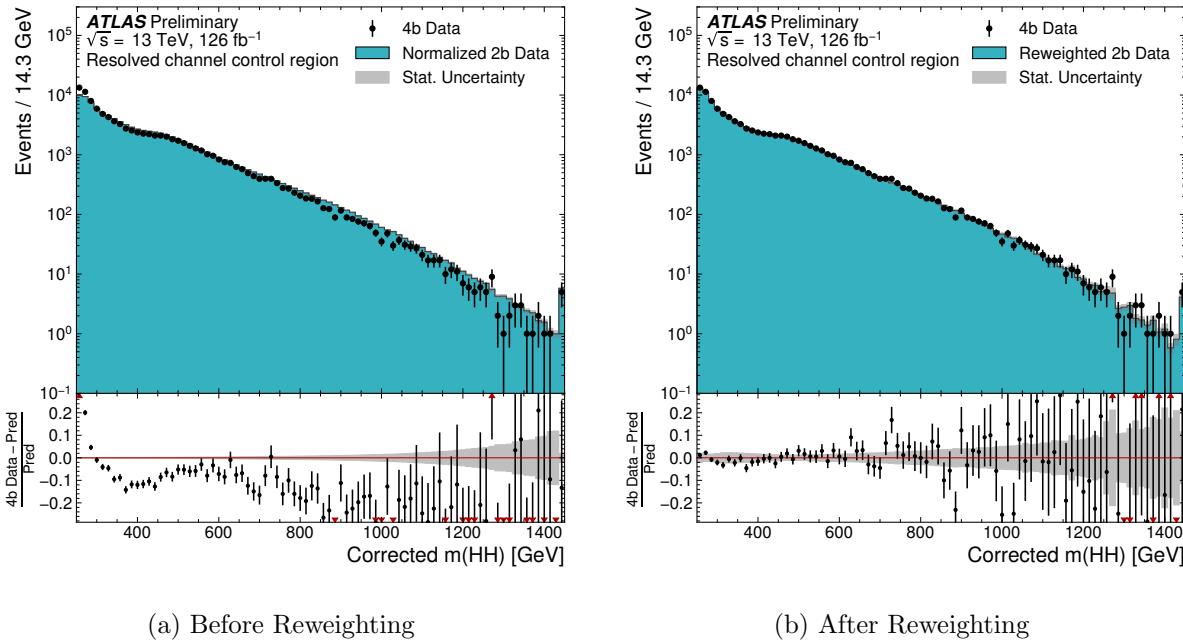


Figure 7.65: **Resonant Search:** Demonstration of the performance of the nominal reweighting in the control region on corrected m_{HH} , with Figure 7.65(a) showing $2b$ events normalized to the total $4b$ yield and Figure 7.65(b) applying the reweighting procedure. Agreement is much improved with the reweighting. Note that overall reweighted $2b$ yield agrees with $4b$ yield in the control region by construction.

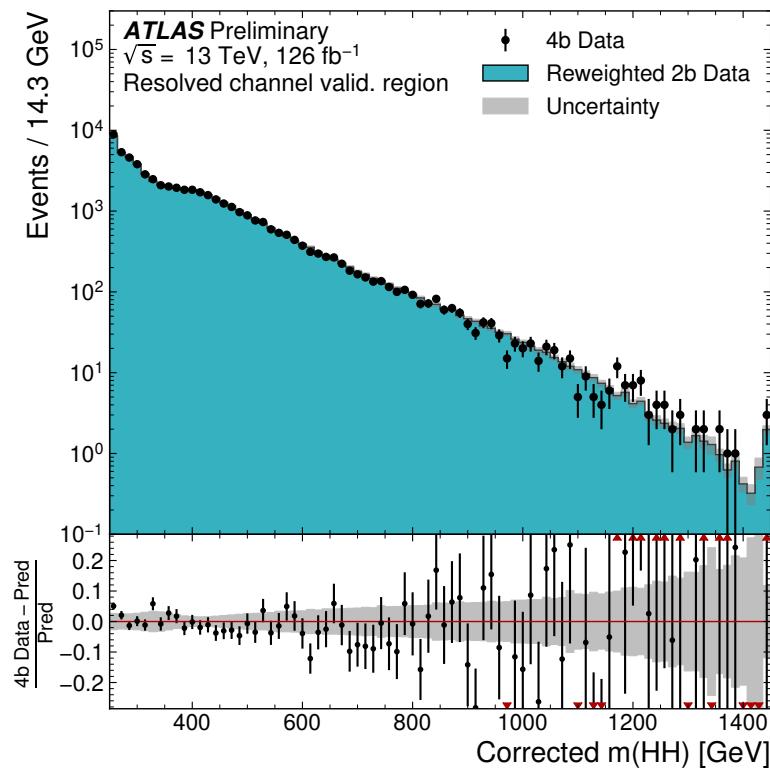


Figure 7.66: **Resonant Search:** Demonstration of the performance of the control region derived reweighting in the validation region on corrected m_{HH} . Agreement is generally good for this extrapolated estimate. Note that the uncertainty band includes the extrapolation systematic, which is defined by a reweighting trained in the validation region.

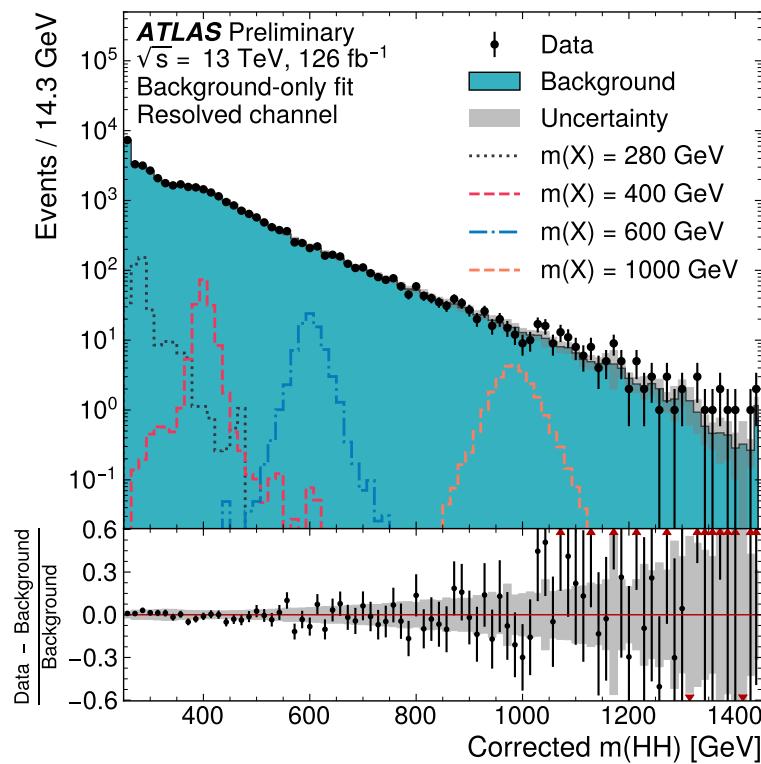


Figure 7.67: **Resonant Search:** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit. The closure is generally quite good, though there is an evident deficit in the background estimate relative to the data for higher values of corrected m_{HH} .

2254 7.10.2 Non-resonant Search

As discussed above, the non-resonant search splits the signal extraction into two categories of $\Delta\eta_{HH}$ ($0 \leq \Delta\eta_{HH} < 0.75$ and $0.75 \leq \Delta\eta_{HH} < 1.5$), and two categories of X_{HH} ($0 \leq X_{HH} < 0.95$ and $0.95 \leq X_{HH} < 1.6$). To maintain reasonable statistics in each bin entering the signal extraction fit, a variable width binning is considered defined by a resolution parameter, r , and a set range in m_{HH} , where bin edges are determined iteratively as

$$b_{low}^{i+1} = b_{low}^i + r \cdot b_{low}^i, \quad (7.14)$$

2255 where b_{low}^i is the low edge of bin i . The parameters used here are $r = 0.08$ over a range
2256 from 280 GeV to 975 GeV, and underflow and overflow are included in the intial and final
2257 bin contents respectively. m_{HH} with no correction is used as the final discriminant in each
2258 category.

2259 A demonstration of the performance of the reweighting on distributions unrolled across
2260 categories is shown in Figure *TODO: insert* for the control region and Figure *TODO:*
2261 *insert* for the validation region. A background-only profile likelihood fit is run for the
2262 distribution in the signal region, and results with the Standard Model HH signal and $\kappa_\lambda = 6$
2263 signal overlaid are shown for $4b$ in Figure 7.68 and $3b1l$ in Figure 7.69. Note that the plots
2264 show the sum across all years, but the signal extraction fit and background estimate are run
2265 with the years separately. All bins are normalized to represent a density of Events / 15 GeV.

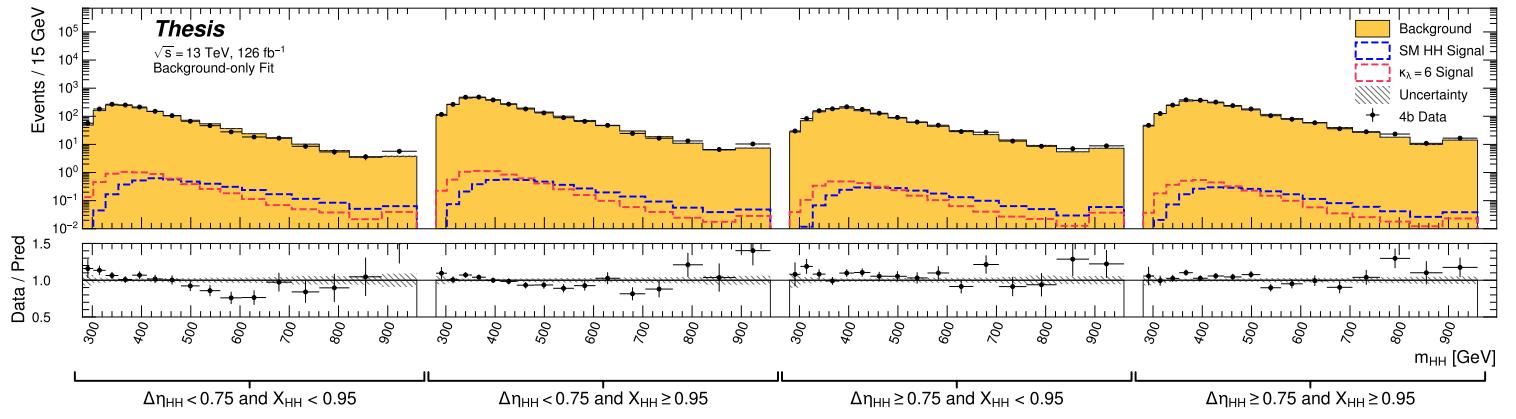


Figure 7.68: **Non-resonant Search (4b):** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 4b channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Modeling is generally quite good near the Standard Model peak, but disagreements are seen at very low and high masses. A deficit is present in low $\Delta\eta_{HH}$ bins near 600 GeV.

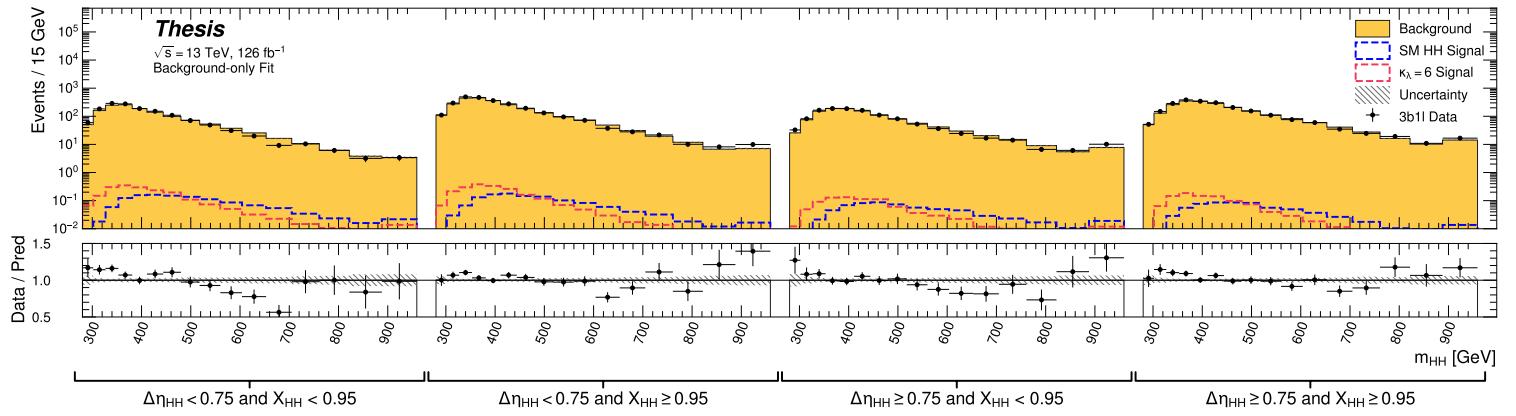


Figure 7.69: **Non-resonant Search (3b1l):** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 3b1l channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Conclusions are very similar to the 4b channels, with generally good modeling near the Standard Model peak, but disagreements at very low and high masses. A deficit is present near 600 GeV.

2266 **7.11 Statistical Analysis**

2267 The resonant analysis is used to set a 95% confidence level upper limit on the $pp \rightarrow X \rightarrow$
2268 $HH \rightarrow b\bar{b}b\bar{b}$ and $pp \rightarrow G_{KK}^* \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ cross-sections, while the non-resonant analysis
2269 is used to set a 95% confidence level upper limit on the $pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ cross sections for
2270 a variety of values of the trilinear Higgs coupling.

2271 The upper limit is extracted using the CL_s method [92]. The test statistic used is q_μ [93],
2272 where μ is the signal strength, and θ represents the nuisance parameters. Due to the use of
2273 signals normalized to 1 fb, μ is also the signal cross-section in fb. A single hat represents the
2274 maximum likelihood estimate of a parameter, while $\hat{\theta}(x)$ represents the conditional maximum
2275 likelihood estimate of the nuisance parameters if the signal cross-section is fixed at x .

$$q_\mu = \begin{cases} -2 \ln \left(\frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad (7.15)$$

2276 CL_s for some test value of μ is then defined by

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p(q_\mu \geq q_{\mu, \text{obs}} | s+b)}{p(q_\mu \geq q_{\mu, \text{obs}} | b)}, \quad (7.16)$$

2277 where the p -values are calculated in the asymptotic approximation [93], which is valid in
2278 the large sample limit.

2279 The signal cross-section μ fb is excluded at the 95% confidence level if $CL_s < 0.05$.

Observed	-2σ	-1σ	Expected	$+1\sigma$	$+2\sigma$
4.4	3.1	4.2	5.9	8.2	11.0

Table 7.1: Limits on Standard Model $HH \rightarrow b\bar{b}b\bar{b}$ production, presented in units of the predicted Standard Model cross section. Results include background systematics only.

2280 7.12 Results

2281 Figure 7.70 shows the expected limit for the spin-0 and spin-2 resonant search. The resolved
 2282 channel covers the range between 251 and 1500 GeV and is combined with the boosted channel
 2283 between 900 and 1500 GeV. The boosted channel then extends to 3 TeV. The most significant
 2284 excess is seen for a signal mass of 1100 GeV, with local significance of 2.6σ for the spin-0
 2285 signal and 2.7σ for the spin-2 signal. This is reduced to 1.0σ and 1.2σ globally.

2286 The spin-2 bulk Randall-Sundrum model with $k/\overline{M}_{\text{Pl}} = 1$ is excluded for graviton masses
 2287 between 298 and 1440 GeV.

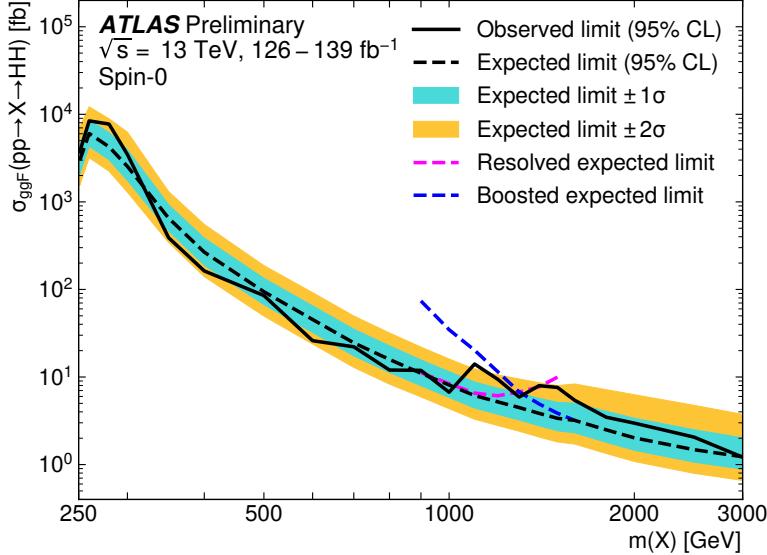
2288 Preliminary results are presented here for the gluon-gluon fusion non-resonant search,
 2289 combining results from the $4b$ and $3b + 1l$ signal regions in the 2×2 category scheme in
 2290 $\Delta\eta_{HH}$ and X_{HH} . These results will be further combined with a VBF channel as discussed,
 2291 but this is left for future work. Results shown here include background systematics only.
 2292 Limits are set for κ_λ values from -20 to 20 . The cross section limit for HH production is set
 2293 at 140 fb (180 fb) observed (expected), corresponding to an observed (expected) limit of 4.4
 2294 (5.9) times the Standard Model prediction (see Table 7.1). κ_λ is constrained to be within the
 2295 range $-4.9 \leq \kappa_\lambda \leq 14.4$ observed ($-3.9 \leq \kappa_\lambda \leq 10.9$ expected). These results are shown in
 2296 Figure 7.71.

2297 We note that this is a significant improvement over the early Run 2 result, which achieved
 2298 an observed (expected) limit of 12.9 (20.7) times the Standard Model prediction. The dataset
 2299 is 4.6 times larger, and a naive scaling of the early Run 2 result (Poisson statistics \implies a factor
 2300 of $1/\sqrt{4.6}$) would predict an observed (expected) limit of 6.0 (9.7) times the Standard Model.

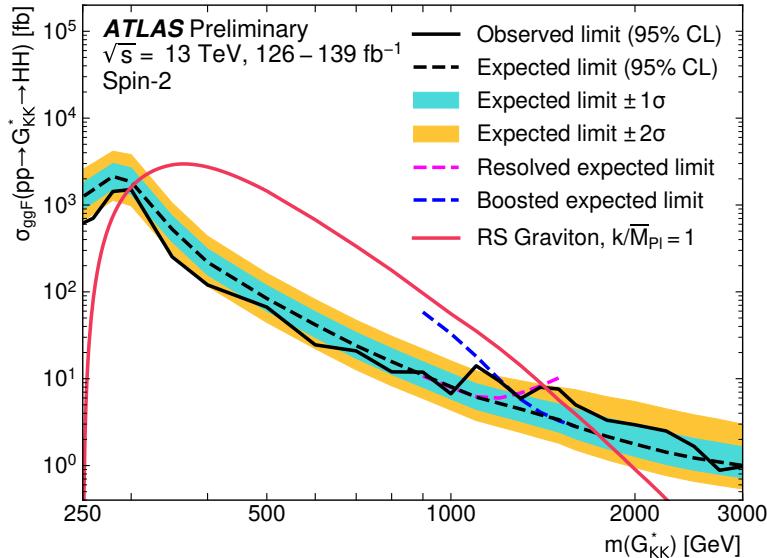
2301 The result of 4.4 (5.9) observed (expected) presented here is therefore both an improvement
 2302 by a factor of 3 (3.5) over the previous result and also beats the statistical scaling by around
 2303 30 (40) %, demonstrating the impact of the various analysis improvements presented here.
 2304 We note again that these results do not include the complete set of uncertainties – however
 2305 we expect the addition of the remaining uncertainties to have no more than a few percent
 2306 impact.

2307 The observed limits presented in Figure 7.71 are consistently above the 2σ band for values
 2308 of $\kappa_\lambda \geq 5$, peaking at a local significance of 3.8σ for $\kappa_\lambda = 6$. As this analysis is optimized for
 2309 points near the Standard Model, and as there is no excess present in more sensitive channels
 2310 in this same region (e.g. $HH \rightarrow bb\gamma\gamma$ *TODO: include comparison*), we do not believe this is a
 2311 real effect, but is rather due to a mis-modeling of the background at low mass, where the
 2312 min ΔR pairing has poor signal efficiency and the assumption of well behaved background in
 2313 the mass plane breaks down. This is consistent with the location of the $\kappa_\lambda = 6$ signal in m_{HH} ,
 2314 as shown in Figures 7.68 and 7.69. It was considered, but not implemented, for this analysis
 2315 to impose a cut on m_{HH} near 350 or 400 GeV to avoid such a low mass modeling issue.

2316 To check the impact of if we would have imposed such a cut, and to verify that the excess
 2317 is due to the low mass regime, we therefore run the same set of limits without the low mass
 2318 bins. In this case, we choose to simply drop the first few bins in m_{HH} such that everything
 2319 else, including the higher mass bin edges, is kept the same. Due to the variable width binning,
 2320 this corresponds to an m_{HH} cut of 381 GeV. The results of this check are shown in Figure
 2321 7.72, overlaid with the limits of Figure 7.71 for reference. With the m_{HH} cut imposed, there
 2322 is a slight degradation in the expected limits for larger positive and negative values of κ_λ ,
 2323 but the points near the Standard Model are nearly identical. Further, the observed excess is
 2324 significantly reduced, with observed limits for $\kappa_\lambda \geq 5$ now falling entirely within the expected
 2325 1σ band. Due to the preliminary nature of these results, further study is left for future
 2326 work. However, we believe, in conjunction with the $HH \rightarrow bb\gamma\gamma$ results and our expectations
 2327 about the difficulty of the background estimation at low mass, that this is demonstrative of a
 2328 mismodeling rather than a real excess.



(a)



(b)

Figure 7.70: Expected (dashed black) and observed (solid black) 95% CL upper limits on the cross-section times branching ratio of resonant production for spin-0 ($X \rightarrow HH$) and spin-2 $G_{KK}^* \rightarrow HH$. The $\pm 1\sigma$ and $\pm 2\sigma$ ranges for the expected limits are shown in the colored bands. The resolved channel expected limit is shown in dashed pink and covers the range from 251 and 1500 GeV. It is combined with the boosted channel (dashed blue) between 900 and 1500 GeV. The theoretical prediction for the bulk RS model with $k/\bar{M}_{Pl} = 1$ [20] (solid red line) is shown, with the decrease below 350 GeV due to a sharp reduction in the $G_{KK}^* \rightarrow HH$ branching ratio. The nominal $H \rightarrow b\bar{b}$ branching ratio is taken as 0.582.

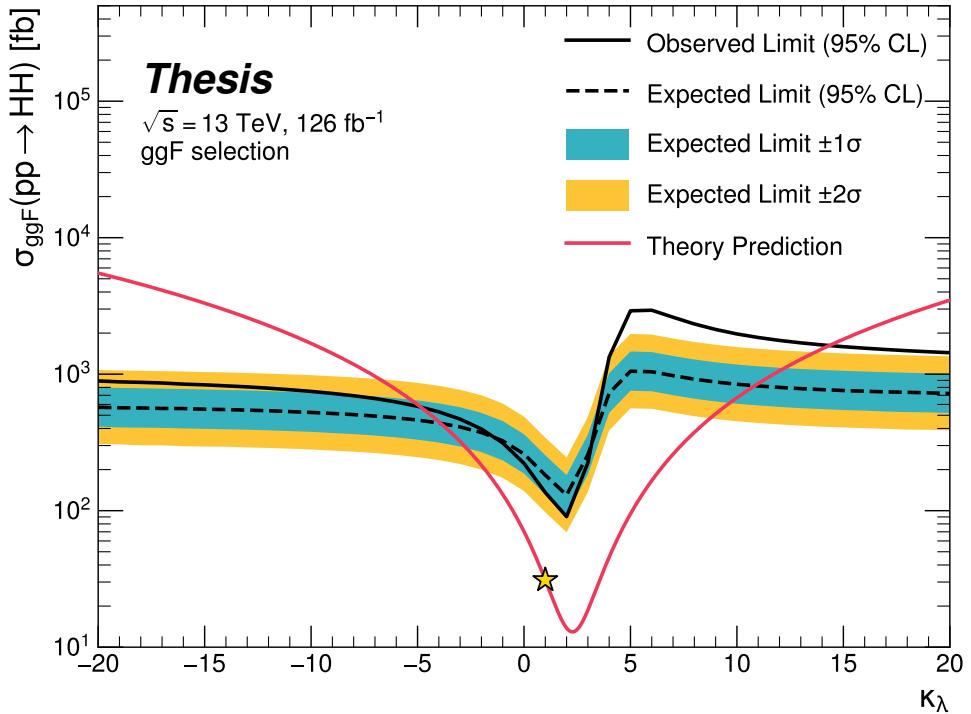


Figure 7.71: Expected (dashed black) and observed (solid black) 95% CL upper limits on the cross-section times branching ratio of non-resonant production for a range of values of the Higgs self-coupling, with the Standard Model value ($\kappa_\lambda = 1$) illustrated with a star. The $\pm 1\sigma$ and $\pm 2\sigma$ ranges for the expected limits are shown in the colored bands. The cross section limit for HH production is set at 140 fb (180 fb) observed (expected), corresponding to an observed (expected) limit of 4.4 (5.9) times the Standard Model prediction. κ_λ is constrained to be within the range $-4.9 \leq \kappa_\lambda \leq 14.4$ observed ($-3.9 \leq \kappa_\lambda \leq 10.9$ expected). The nominal $H \rightarrow b\bar{b}$ branching ratio is taken as 0.582. We note that the excess present for $\kappa_\lambda \geq 5$ is thought to be due to a low mass background mis-modeling, present due to the optimization of this analysis for the Standard Model point, and is not present in more sensitive channels in this same region (e.g. $HH \rightarrow bb\gamma\gamma$).

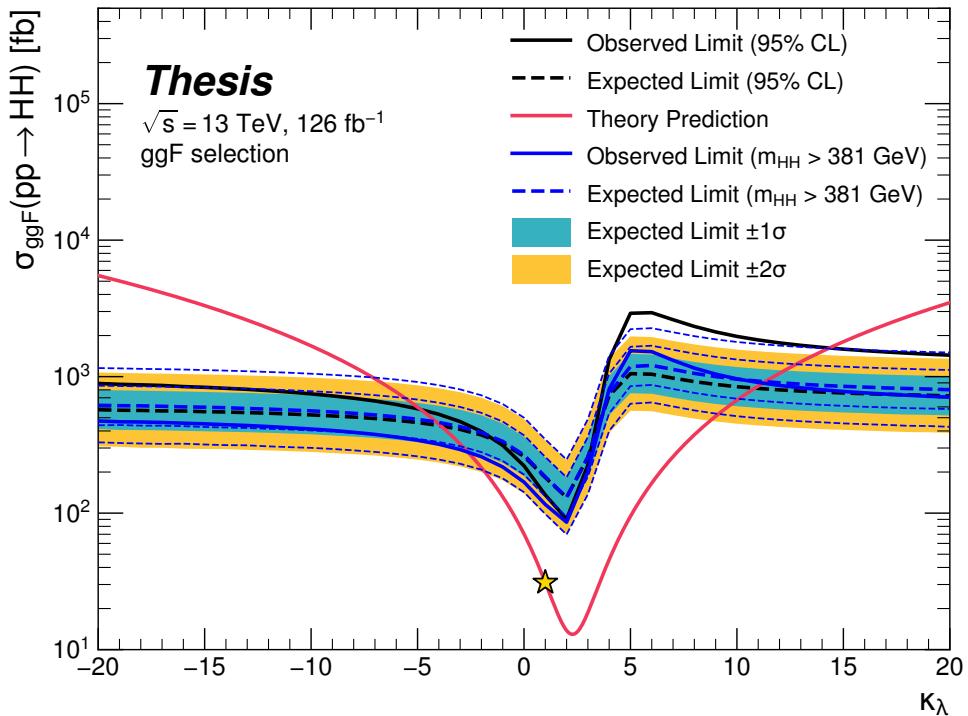


Figure 7.72: Comparison of the limits in Figure 7.71 with an equivalent set of limits that drop the m_{HH} bins below 381 GeV, with the value of 381 GeV determined by the optimized variable width binning. The expected limit band with this mass cut is shown in dashed blue, and the observed is shown in solid blue. The excess at and above $\kappa_\lambda = 5$ is significantly reduced, demonstrating that this is driven by low mass. Notably, there is minimal impact on the expected sensitivity with this m_{HH} cut.

2329

Chapter 8

2330

FUTURE IDEAS FOR $HH \rightarrow b\bar{b}b\bar{b}$

2331 The searches presented in this thesis make use of a large suite of sophisticated techniques,
 2332 selected through careful study and validation. During this process, a variety of interesting
 2333 directions for the $HH \rightarrow b\bar{b}b\bar{b}$ analysis were explored by this thesis author, in collaboration
 2334 with a few others¹, but were not used due to a variety of constraints. We present two
 2335 such interesting directions here, with the hope of encouraging further exploration of these
 2336 techniques in future work.

2337 **8.1 pairAGraph: A New Method for Jet Pairing**

2338 As discussed in Chapter 7, one of the main problems to solve is the pairing of b -jets into
 2339 Higgs candidates. Figure 7.1 demonstrates that the choice of the pairing method, while
 2340 important for achieving good reconstruction of signal events, also significantly impacts the
 2341 structure of non- HH events, leading to various biases in the background estimate. Evaluation
 2342 of the pairing method therefore must take both of these factors into account. While we have
 2343 presented some advantages in respective contexts for the pairing methods considered here,
 2344 we of course would like to explore further improvements to this important component of the
 2345 analysis.

2346 To that end, we note that all of the pairing methods considered here share a common
 2347 feature: four jets are selected, and the pairing is some discrimination between the available
 2348 three pairings of these four jets. For the methods used in this analysis, the jet selection
 2349 proceeds via a simple p_T ordering, with b -tagged jets receiving a higher priority than non-

¹Notably Nicole Hartman (SLAC), who spearheaded much of the development and proof of concept work, in collaboration with Michael Kagan and Rafael Teixeira De Lima.

2350 tagged jets.

2351 With the advent of a variety of machine learning methods for dealing with a variable
2352 number of inputs (e.g. recurrent neural networks [94], deep sets [95], graph neural networks [96],
2353 and transformers [97]), a natural place to improve on the pairing is to consider more than
2354 just four jets. The pairing and jet selection is then performed simultaneously, allowing for
2355 the incorporation of more event information in the pairing decision and the incorporation of
2356 jet correlation structure in the jet selection.

2357 In practice, the majority of $HH \rightarrow b\bar{b}b\bar{b}$ events have either four or five jets which pass the
2358 kinematic preselection, and any gain from this additional freedom would come from events
2359 with greater than or equal to five jets. However, this five jet topology is particularly exciting
2360 for scenarios such as events with initial state radiation (ISR), in which the $HH - > 4b$ jets
2361 are offset by a single jet with p_T similar in magnitude to that of the $HH - > 4b$ system.
2362 Such events have explicit event level information which is not encoded with the inclusion
2363 of only the $HH - > 4b$ jets, and are pathological if the ISR jet happens to pass b -tagging
2364 requirements.

2365 Additionally, with the use of lower tagged regions for background estimation and alternate
2366 signal regions, this extra flexibility in jet selection may provide a very useful bias – since the
2367 algorithm is trained on signal, the selected jets for the pairing will be the most “4b-like” jets
2368 available in the considered set.

2369 For the studies considered here, a transformer [97] based architecture is used. This is best
2370 visualized by considering the event as a graph with jets corresponding to nodes and edges
2371 corresponding to potential connections – for this reason, we term this algorithm “pairAGraph”.
2372 The approach is as follows: each jet, i , is represented by some vector of input variables, \vec{x}_i ,
2373 in our case the four-vector information, (p_T, η, ϕ, E) of each jet, plus information on the
2374 b -tagging decision. A multi-layer perceptron (MLP) is used to create a latent embedding,
2375 $\mathbf{h}(\vec{x}_i)$, of this input vector.

To describe the relationship between various jets in the event, we then define a vector \vec{z}_i

for each jet as

$$\vec{z}_i = \sum_j w_{ij} \mathbf{h}(\vec{x}_j) \quad (8.1)$$

2376 where j runs over all jets in the event (including $i = j$), the w_{ij} can be thought of as edge
 2377 weights, and $\mathbf{h}(\vec{x}_j)$ is the latent embedding for jet j mentioned above.

Within this formula, both \mathbf{h} and the w_{ij} are learnable. To learn an appropriate latent mapping and set of edge weights, we define a similarity metric corresponding to each possible jet pairing:

$$\vec{z}_{1a} \cdot \vec{z}_{1b} + \vec{z}_{2a} \cdot \vec{z}_{2b} \quad (8.2)$$

2378 where subscripts $1a$ and $1b$ correspond to the two jets in pair 1, $2a$ and $2b$ to the jets in pair
 2379 2 for a given pairing of four distinct jets.

2380 This similarity metric is calculated for all possible pairings, which are then passed through
 2381 a softmax [98] activation function, which compresses these scores to between 0 and 1 with
 2382 sum of 1, lending an interpretation as probability of each pairing.

2383 In training, the ground truth pairing is set by *truth matching* jets to the b -jets in the
 2384 HH signal simulation – a jet is considered to match if it is < 0.3 in ΔR away from a b -jet in
 2385 the simulation record. Given this ground truth, a cross-entropy loss *TODO: cite* is used on
 2386 the softmax outputs, and w_{ij} and \mathbf{h} are updated correspondingly. Training in such a way
 2387 corresponds to updating w_{ij} and \mathbf{h} to maximize the similarity metric for the correct pairing.

2388 In evaluation, the pairings with a higher score (and therefore higher softmax output)
 2389 given the trained h and w_{ij} therefore correspond to the pairings that are most “ HH -like”.
 2390 The maximum over these scores is therefore the pairing used as the predicted result from the
 2391 algorithm.

2392 Because the majority of $HH \rightarrow b\bar{b}b\bar{b}$ events have either four or five jets, it was found to
 2393 be sufficient to only consider a maximum of 5 jets. Consideration of more is in principle
 2394 possible, but the quickly expanding combinatorics leads to a rapidly more difficult problem.
 2395 The jets considered are the five leading jets in p_T . Notably, this set of jets may include jets
 2396 which are not b -tagged, even for the nominal $4b$ region – therefore events with 4 b -jets are

2397 not required to use all of them in the construction of Higgs candidates, in contrast to the
2398 other algorithms used in this thesis.

2399 **8.2 Background Estimation with Mass Plane Interpolation**

2400 The choice of a pairing algorithm that results in a smooth mass plane (such as $\min \Delta R$)
2401 opens up a variety of options for the background estimation. While the method based on
2402 reweighting of $2b$ events used for this thesis performs well and has been extensively studied
2403 and validated, it also relies on several assumptions. In particular, the reweighting is derived
2404 between e.g., $2b$ and $4b$ events *outside* of the signal region and then applied to $2b$ events *inside*
2405 the signal region, with the assumption that the $2b$ to $4b$ transfer function will be sufficiently
2406 similar in both regions of the mass plane. An uncertainty is assigned to account for the bias
2407 due to this assumption, but the extrapolation in the mass plane is never explicitly treated in
2408 the nominal estimate. While the approach of reweighting $2b$ events within the signal region
2409 does have the advantage of incorporating explicit signal region information (that is, the $2b$
2410 signal region events), the importance of the extrapolation bias motivates consideration of
2411 a method that operates within the $4b$ mass plane. This additionally removes the reliance
2412 on lower b -tagging regions, allowing for the use of, e.g. $3b$ triggers, and future-proofing the
2413 analysis against trigger bandwidth constraints in the low tag regions.

The method considered here relies on the following: for a given vector of input variables (event kinematics, etc), \vec{x} , the joint probability in the HH mass plane may be written as:

$$p(\vec{x}, m_{H1}, m_{H2}) = p(\vec{x}|m_{H1}, m_{H2})p(m_{H1}, m_{H2}) \quad (8.3)$$

2414 by the chain rule of probability. This means that the full dynamics of events in the HH mass
2415 plane may be described by (1) the conditional probability of considered variables \vec{x} , given
2416 values of m_{H1} and m_{H2} , and (2) the density of the mass plane itself.

2417 We present here an approach which uses normalizing flows *TODO: cite* to model the
2418 conditional probabilities of events in the mass plane and Gaussian processes to model the
2419 mass plane density. These models are trained in a region around, but not including, the

2420 signal region, and the trained models are then used to construct an *interpolated* estimate of
 2421 the signal region kinematics. This approach therefore explicitly treats event behavior within
 2422 the mass plane, avoiding the concerns associated with a reweighted estimate. Validation of
 2423 such a method, as well as assessing of closure and biases of the method, may be done in
 2424 alternate b -tagging or kinematic regions, notably the now unused $2b$ region, results of which
 2425 are shown below.

2426 *8.2.1 Normalizing Flows*

Normalizing flows model observed data $x \in X$, with $x \sim p_X$, as the output of an invertible,
 differentiable function $f : X \rightarrow Z$, with $z \in Z$ a latent variable with a simple prior probability
 distribution (often standard normal), $z \sim p_Z$. From a change of variables, given such a
 function, we may write

$$p_X(x) = p_Z(f(x)) \left| \det \left(\frac{d(f(x))}{dx} \right) \right| \quad (8.4)$$

2427 where $\left(\frac{d(f(x))}{dx} \right)$ is the Jacobian of f at x .

2428 The problem of normalizing flows then reduces to (1) choosing sets of f which are both
 2429 tractable and sufficiently expressive to describe observed data, and (2) optimizing associated
 2430 sets of functional parameters on observed data via maximum likelihood estimation using
 2431 the above formula. Sampling from the learned density is done by drawing from the latent
 2432 distribution $z \sim p_Z$ (cf. inverse transform sampling) – the corresponding sample is then
 2433 $x \sim p_X$ with $x = f^{-1}(z)$.

2434 A standard approach to the definition of these f is as a composition of affine transfor-
 2435 mations (e.g. RealNVP *TODO: cite*), that is, transformations of the form $\alpha z + \beta$, with α and β
 2436 learnable parameter vectors. This can roughly be thought of as shifting and squeezing the
 2437 input prior density in order to match the data density. However, this has somewhat
 2438 limited expressivity, for instance in the case of a multi-modal density.

This work thus instead relies on neural spline flows *TODO: cite: <https://arxiv.org/pdf/1906.04032.pdf>*
 in which the functions considered are monotonic rational-quadratic splines, which have an

analytic inverse. A rational quadratic function has the form of a quotient of two quadratic polynomials, namely,

$$f_j(x_i) = \frac{a_{ij}x_i^2 + b_{ij}x_{ij} + c_{ij}}{d_{ij}x_i^2 + e_{ij}x_i + f_{ij}} \quad (8.5)$$

with six associated parameters (a_{ij} through f_{ij}) per each piecewise bin j of the spline and each input dimension i . This is explicitly more flexible and expressive than a simple affine transformation, allowing, e.g., the treatment of multi-modality via the piecewise nature of the spline.

The rational quadratic spline is defined on an set interval. The transformation outside of this interval is set to the identity, with these linear tails allowing for unconstrained inputs. The boundaries between bins of the spline are set by coordinates scalled *knots*, with $K + 1$ knots for K bins – the two endpoints for the spline interval plus the $K - 1$ internal boundaries. The derivatives at these points are constrained to be positive for the internal knots, and boundary derivatives are set to 1 to match the linear tails.

The bin widths and heights are learnable ($2 \cdot K$ parameters) as are the internal knot derivatives ($K - 1$ parameters), and these $3K - 1$ ouputs of the neural network are sufficient to define a monotonic rational-quadratic spline which passes through each knot and has the given derivative value at each knot.

In the context of the $HH \rightarrow 4b$ analysis, a neural spline flow is used to model the four vector information of each Higgs candidate, conditional on their respective masses. The resulting flow is therefore five dimensional, with inputs $x = (p_{T,H1}, p_{T,H2}, \eta_{H1}, \eta_{H2}, \Delta\phi_{HH})$, where the ATLAS ϕ symmetry has been encdoded by assuming $\phi_{H1} = 0$. Conditional variables m_{H1} and m_{H2} are not modeled by the flow, but “come along for the ride”. A standard normal distribution in 5 dimensions is used for the underlying prior. Modeling of the four vectors was chosen in order to reduce bias from modeling m_{HH} directly.

The trained flow model then gives a model for $p(x|m_{H1}, m_{H2})$ which may be sampled from to reconstruct distributions of HH kinematics given values of m_{H1} and m_{H2} .

2462 8.2.2 Gaussian Processes

2463 The second piece of this background estimate is the modeling of the mass plane density,
2464 $p(m_{H1}, m_{H2})$. This is done using Gaussian process regression – note that a similar procedure
2465 is used to define a systematic in the boosted $4b$ analysis. Generally, Gaussian processes
2466 are a collection of random variables in which every finite collection of said variables is
2467 distributed according to a multivariate normal distribution. For the context of Gaussian
2468 process regression, what we consider is a Gaussian process over function space, that is, for a
2469 collection of points, x_1, \dots, x_N , the space of corresponding function values, $(f(x_1), \dots, f(x_N))$
2470 is Gaussian process distributed, that is, described by an N dimensional normal distribution
2471 with mean μ , covariance matrix Σ .

2472 For a single point, this would correspond to a function space described entirely by a
2473 normal distribution, with various samples from that distribution yielding various candidate
2474 functions. For multiple points, a covariance matrix describes the relationship between each
2475 pair of points – correspondingly, it is represented via a *kernel function*, $K(x, x')$. As, in
2476 practice, μ may always be set to 0 via a centering of the data, the kernel function fully defines
2477 the considered family of functions.

The considered family of functions describes a Bayesian *prior* for the data. This prior may be conditioned on a set of training data points (X_1, \vec{y}_1) . This conditional *posterior* may then be used to make predictions $\vec{y}_2 = f(X_2)$ at a set of new points X_2 . Because of the Gaussian process prior assumption, \vec{y}_1 and \vec{y}_2 are assumed to be jointly Gaussian. We may therefore write

$$\begin{pmatrix} \vec{y}_1 \\ \vec{y}_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K(X_1, X_1) & K(X_1, X_2) \\ K(X_1, X_2) & K(X_2, X_2) \end{pmatrix} \right) \quad (8.6)$$

2478 where we have used that the kernel function is symmetric and assumed prior mean 0.

By standard conditioning properties of Gaussian distributions,

$$\vec{y}_2 | \vec{y}_1 \sim \mathcal{N}(K(X_2, X_1)K(X_1, X_1)^{-1}\vec{y}_1, K(X_2, X_2) - K(X_2, X_1)K(X_1, X_1)^{-1}K(X_1, X_2)) \quad (8.7)$$

2479 which is the sampling distribution for a Gaussian process given kernel K . In practice, the
2480 mean of this sampling distribution is used as the function estimate, with an uncertainty from
2481 the predicted variance at a given point.

The choice of kernel function has a very strong impact on the fitted curve, and must therefore be chosen to express the expected dynamics of the data. A common such choice is a radial basis function (RBF) kernel, which takes the form

$$K(x, x') = \exp\left(-\frac{d(x, x')^2}{2l^2}\right) \quad (8.8)$$

2482 where $d(\cdot, \cdot)$ is the Euclidean distance and $l > 0$ is a length scale parameter. Conceptually, as
2483 distances $d(x, x')$ increase relative to the chosen length scale, the kernel smoothly dies off –
2484 further away points influence each other less.

2485 Coming back to our case of the mass plane, the procedure runs as follows:

2486 1. A binned 2d histogram of the blinded mass plane is created in a window around the
2487 “standard” analysis regions. Bins which have any overlap with the signal region are
2488 excluded.

2489 2. A Gaussian process is trained using the bin centers, values as training points. The
2490 scikit-learn implementation [99] is used, with RBF kernel with anisotropic length scale
2491 (l is dimension 2). The length scale is initialized to $(50, 50)$ to cover the signal region,
2492 and optimized by minimizing the negative log-marginal likelihood on the training data,
2493 $-\log p(\vec{y}|\theta)$. Training data is centered and scaled to mean 0, variance 1, and a statistical
2494 error is included in the fit.

2495 3. The Gaussian process is then used to predict the density $p(m_{H1}, m_{H2})$ in the signal
2496 region. This may then be sampled from via an inverse transform sampling to generate
2497 values (m_{H1}, m_{H2}) according to the density (specifically, according to the mean of the
2498 Gaussian process posterior). Though in principle the Gaussian process sampling is not
2499 limited to bin centers, this is kept for simplicity, with a uniform smearing applied within

2500 each sampled bin to approximate the continuous estimate, namely, if a bin is sampled
 2501 from, the returned value is drawn uniformly at random within the sampled bin.

4. The sampling in the previous step can be arbitrary – to set the overall normalization, a Monte Carlo sampling of the Gaussian process is done to approximate the relative fraction of events predicted both inside (f_{in}) and outside (f_{out}) of the signal region, within the training box. The number of events outside of the signal region (n_{out}) is known, therefore, the number of events inside of the signal region, n_{in} , may be estimated as

$$n_{in} = \frac{n_{out}}{f_{out}} \cdot f_{in}. \quad (8.9)$$

2502 Note that the Monte Carlo sampling procedure is simply a set of samples of the Gaussian
 2503 process from uniformly random values of m_{H1}, m_{H2} , and is the most convenient approach
 2504 given the irregular shape of the signal region.

2505 This procedure results in a generated set of predicted m_{H1}, m_{H2} values for signal region
 2506 background events, along with an overall yield prediction.

2507 8.2.3 The Full Prediction

2508 Given the normalizing flow parametrization of $p(x|m_{H1}, m_{H2})$ and the Gaussian process
 2509 generation of $(m_{H1}, m_{H2}) \sim p(m_{H1}, m_{H2})$ and prediction of the signal region yield, all of the
 2510 pieces are in place to construct an interpolation background estimate. Namely

- 2511 1. Gaussian process sampled (m_{H1}, m_{H2}) values are provided to the normalizing flow to
 2512 predict the other variables for the Higgs candidate four-vectors. These are used to
 2513 construct the HH system (notably $m_{HH}, \cos \theta^*$).
- 2514 2. These final distributions are normalized according to the predicted background yield.

2515 8.2.4 Results

2516 The Gaussian process sampling procedure is trained on a small fraction (0.01) of $2b$ data to
 2517 mimic the available $4b$ statistics. This fraction of $2b$ data is blinded, and the prediction of the
 2518 estimate trained on this blinded region may then be compared to real $2b$ data in the signal
 2519 region. The predictions for signal region m_{H_1} and m_{H_2} individually are shown in Figure 8.1,
 and the resulting mass planes are compared in Figure 8.2. Good agreement is seen.

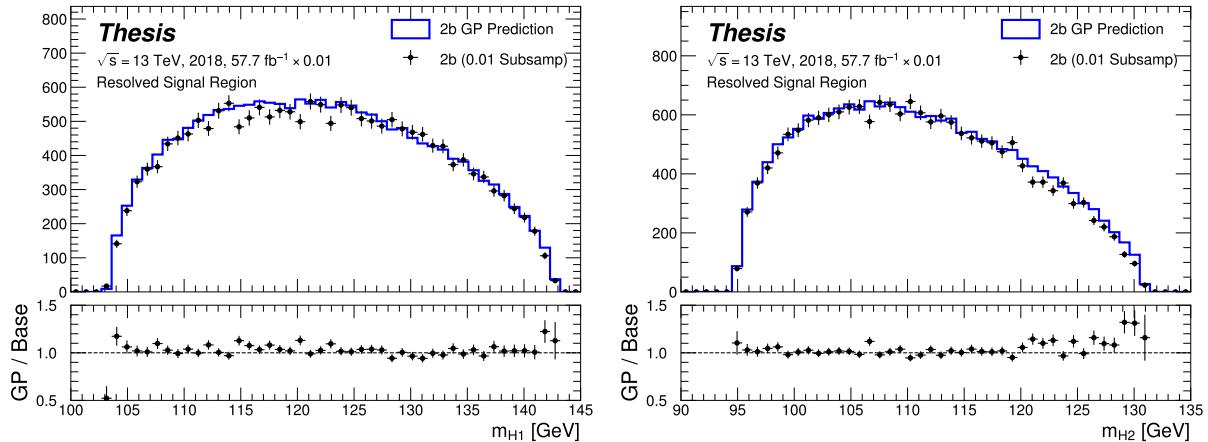


Figure 8.1: Gaussian process sampling prediction of marginals m_{H_1} and m_{H_2} for $2b$ signal region events compared to real $2b$ signal region events for the 2018 dataset. Good agreement is seen. Only a small fraction (0.01) of the $2b$ dataset is used for both training and this final comparison to mimic $4b$ statistics.

2520

2521 The $4b$ region is kept blinded for this work, meaning that a direct comparison of the
 2522 Gaussian process estimate in the $4b$ signal region is not done. However, a Gaussian process is
 2523 trained on the blinded $4b$ region and compared to the corresponding reweighted $2b$ estimate,
 2524 trained per the nominal procedures from the analyses above. The predictions for signal
 2525 region m_{H_1} and m_{H_2} individually are shown in Figure 8.3, compared to both the control and
 2526 validation region derived reweighting estimates, and the resulting signal region mass planes
 2527 are compared in Figure 8.4. The estimates are seen to be compatible.

2528 8.2.5 *Outstanding Points*

2529 While good performance is demonstrated from the nominal interpolated background estimate,
2530 various uncertainties must be assigned according to the various stages of the estimate. These
2531 notably include

2532 • Assessing a statistical uncertainty on the normalizing flow training (cf. bootstrap
2533 uncertainty).

2534 • Propagation of the Gaussian process uncertainty through the sampling procedure.

2535 • Validation of the resulting estimate and assessment of necessary systematic uncertainties
2536 (e.g. from validation region non-closure).

2537 These are all quite tractable, but some, especially the choice of an appropriate systematic
2538 uncertainty, are certainly not obvious and require detailed study. In this respect, the
2539 reweighting validation work of the non-resonant analysis is certainly quite useful as a starting
2540 place in terms of the available regions and their correspondence to the nominal $4b$ signal
2541 region.

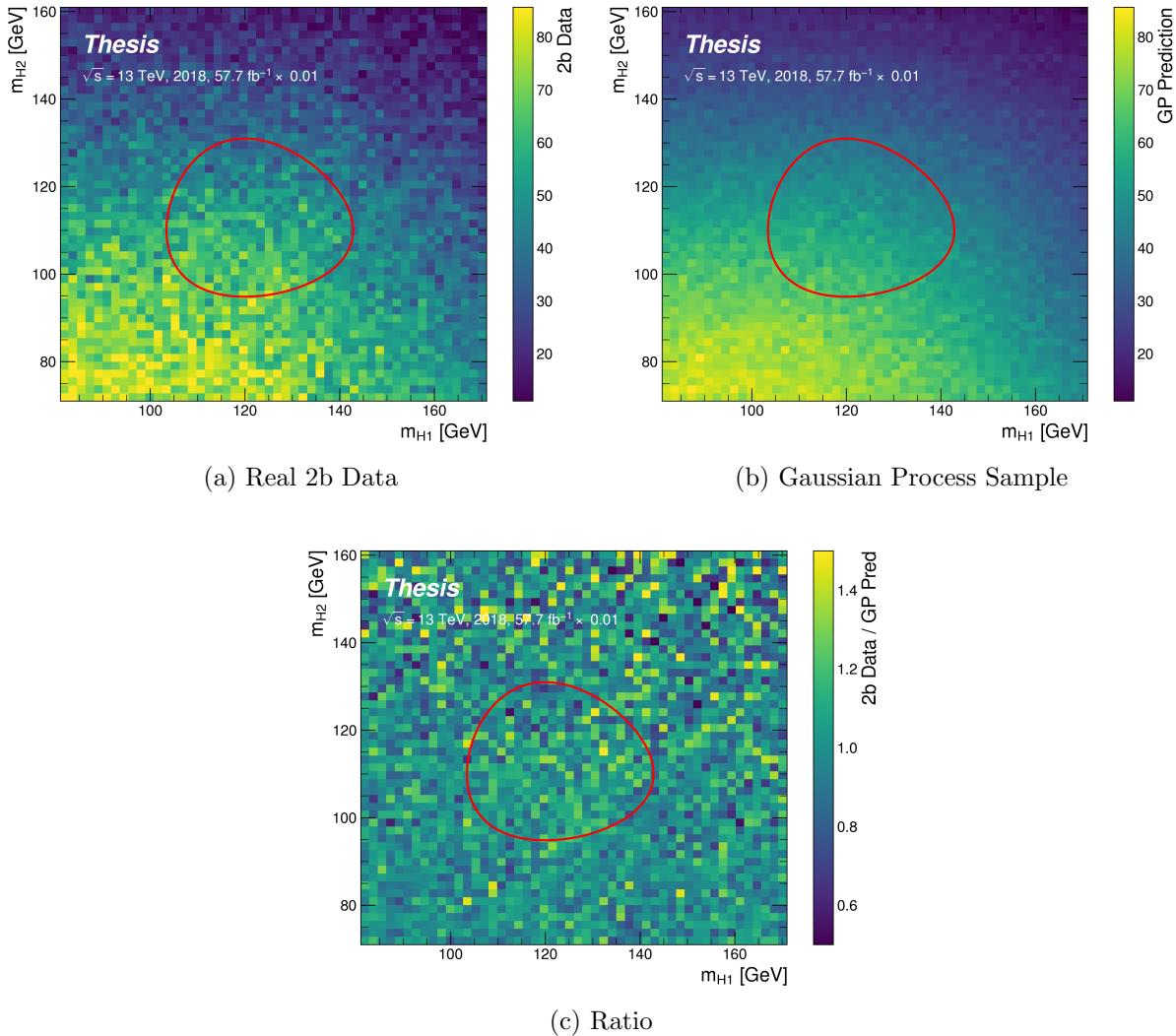


Figure 8.2: Gaussian process sampling prediction for the mass plane compared to the real 2b dataset for 2018. Only a small fraction (0.01) of the 2b dataset is used for both training and this final comparison to mimic 4b statistics. Good agreement is seen.

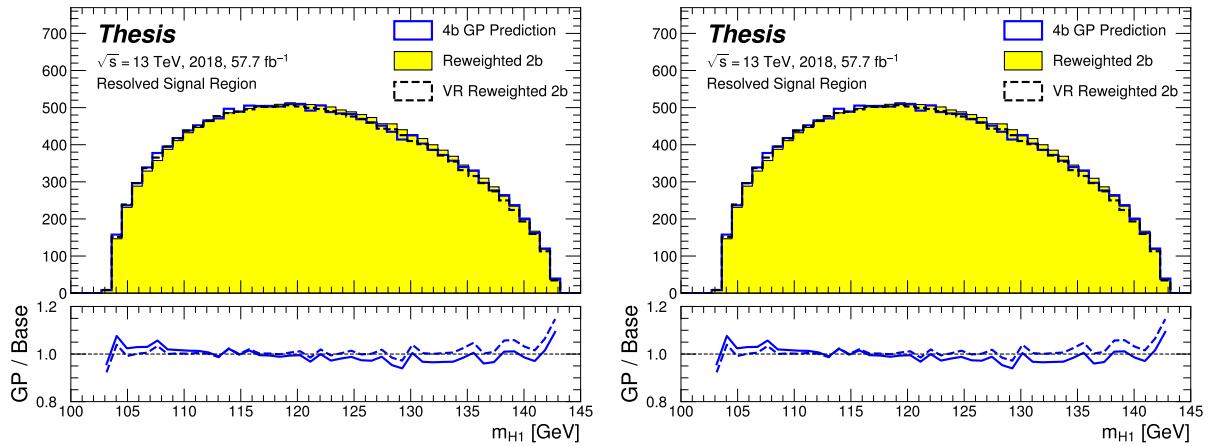


Figure 8.3: Gaussian process sampling prediction of marginals m_{H1} and m_{H2} for 4b signal region events compared to both control and validation reweighting predictions. While there are some differences, the estimates are compatible.

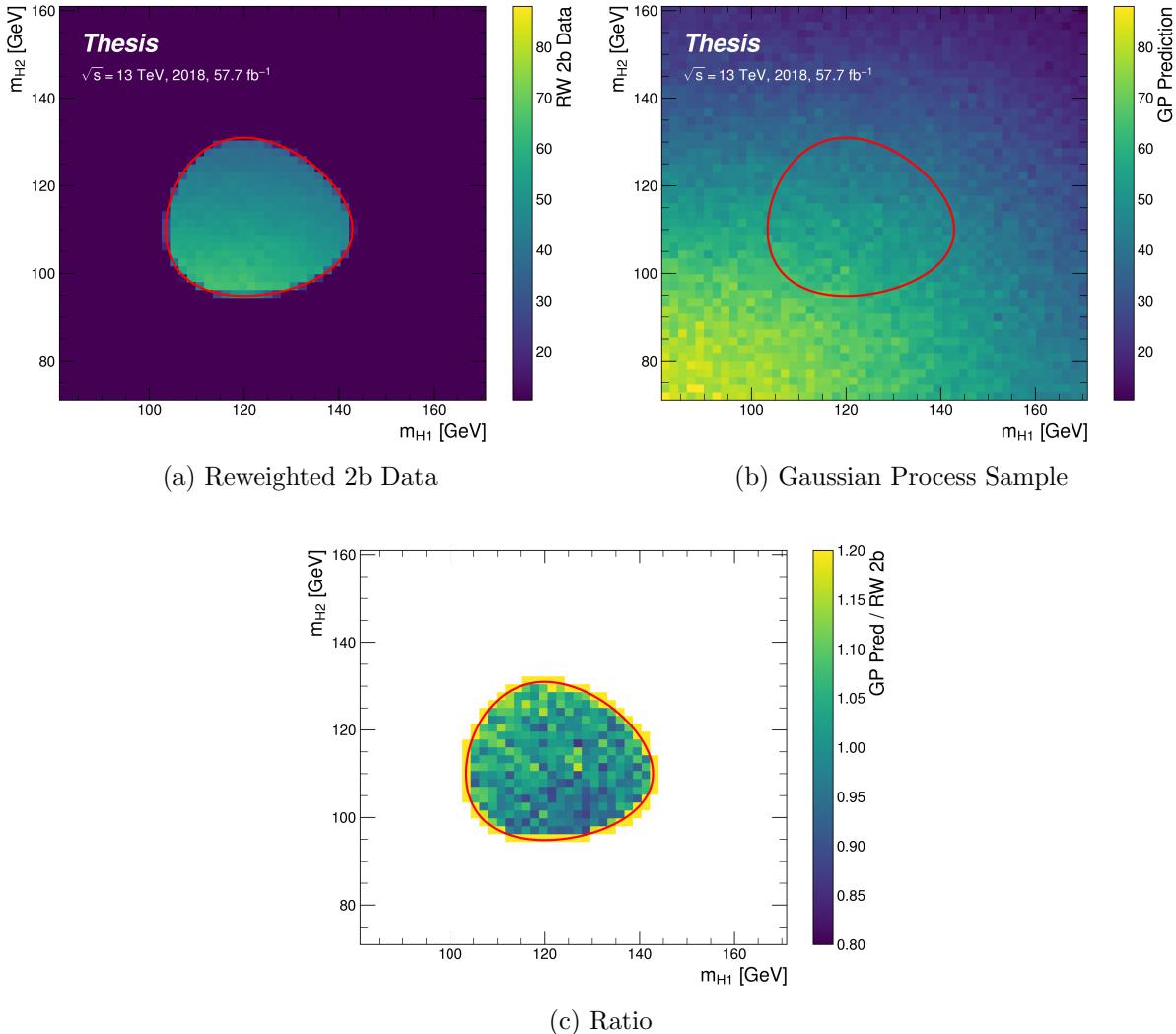


Figure 8.4: Gaussian process sampling prediction for the $4b$ mass plane compared to the reweighted $2b$ estimate in the signal region. Both estimates are compatible.

2542

Chapter 9

2543

CONCLUSIONS

2544 This thesis has provided an overview of the Standard Model, with an emphasis on pair
2545 production of Higgs bosons and how this process may be used to both verify the Standard
2546 Model and to search for new physics. An overview of the Large Hadron Collider and the
2547 ATLAS detector has been provided, and the design and use of simulation infrastructure
2548 has been explained, including work to improve hadronic shower modeling in fast detector
2549 simulation. The translation of detector level information to analysis level information has
2550 been explained, with an emphasis on jets and the identification of B hadron decay. Finally,
2551 two searches for Higgs boson pair production have been presented, with a complete set of
2552 results for resonant production included, focusing on searches beyond the Standard Model,
2553 and a preliminary set of results for non-resonant production, targeting Standard Model
2554 production, with variations of the Higgs self-coupling. Two advanced techniques for the
2555 future of these analyses are further presented, along with proof-of-concept results.

2556

BIBLIOGRAPHY

- 2557 [1] *A New Map of All the Particles and Forces*, <https://www.quantamagazine.org/a-new-map-of-the-standard-model-of-particle-physics-20201022/>, Accessed:
2558
2559 2021-07-23 (cit. on p. 3).
- 2560 [2] J. D. Jackson, *Classical electrodynamics; 2nd ed.* Wiley, 1975 (cit. on p. 2).
- 2561 [3] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 978-1-
2562 107-03426-6 (cit. on p. 2).
- 2563 [4] M. Srednicki, *Quantum Field Theory*, Cambridge Univ. Press, 2007 (cit. on p. 2).
- 2564 [5] M. Gell-Mann, *A schematic model of baryons and mesons*, Physics Letters **8** (1964)
2565 214, ISSN: 0031-9163, URL: <https://www.sciencedirect.com/science/article/pii/S0031916364920013> (cit. on p. 11).
- 2567 [6] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version 1*,
2568 tech. rep., CERN, 1964, URL: <https://cds.cern.ch/record/352337> (cit. on p. 11).
- 2569 [7] O. W. Greenberg, *Spin and Unitary-Spin Independence in a Paraquark Model of Baryons*
2570 and Mesons, Phys. Rev. Lett. **13** (20 1964) 598, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.598> (cit. on p. 11).
- 2572 [8] M. Y. Han and Y. Nambu, *Three-Triplet Model with Double $SU(3)$ Symmetry*, Phys.
2573 Rev. **139** (4B 1965) B1006, URL: <https://link.aps.org/doi/10.1103/PhysRev.139.B1006> (cit. on p. 11).
- 2575 [9] F. L. Wilson, *Fermi's Theory of Beta Decay*, American Journal of Physics **36** (1968)
2576 1150, eprint: <https://doi.org/10.1119/1.1974382>, URL: <https://doi.org/10.1119/1.1974382> (cit. on p. 14).

- 2578 [10] *The Nobel Prize in Physics 1979*, <https://www.nobelprize.org/prizes/physics/1979/summary/>, Accessed: 2021-07-23 (cit. on p. 15).
- 2579
- 2580 [11] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
2581 Phys. Rev. Lett. **13** (9 1964) 321, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on p. 20).
- 2582
- 2583 [12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett.
2584 **13** (16 1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>
2585 (cit. on p. 20).
- 2586 [13] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and
2587 Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585> (cit. on p. 20).
- 2588
- 2589 [14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard
2590 Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,
2591 arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214) (cit. on p. 20).
- 2592
- 2593 [15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS
2594 experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: [1207.7235 \[hep-ex\]](https://arxiv.org/abs/1207.7235)
(cit. on p. 20).
- 2595
- 2596 [16] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264> (cit. on p. 25).
- 2597
- 2598 [17] ATLAS Collaboration, *Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion
2599 production using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
JHEP **07** (2020) 108, arXiv: [2001.05178 \[hep-ex\]](https://arxiv.org/abs/2001.05178) (cit. on pp. 28, 92), Erratum: JHEP
2600 **01** (2021) 145.
- 2601
- 2602 [18] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Physics
Reports **516** (2012) 1, Theory and phenomenology of two-Higgs-doublet models,

- 2603 ISSN: 0370-1573, URL: <http://www.sciencedirect.com/science/article/pii/S0370157312000695> (cit. on pp. 29, 30).
- 2604
- 2605 [19] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, *Warped gravitons at the CERN LHC*
2606 *and beyond*, Phys. Rev. D **76** (3 2007) 036006, URL: <https://link.aps.org/doi/10.1103/PhysRevD.76.036006> (cit. on pp. 29, 30).
- 2607
- 2608 [20] A. Carvalho, *Gravity particles from Warped Extra Dimensions, predictions for LHC*,
2609 2018, arXiv: [1404.0102 \[hep-ph\]](https://arxiv.org/abs/1404.0102) (cit. on pp. 29, 30, 177).
- 2610 [21] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions*
2611 *at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, Phys. Lett. B **800** (2020) 135103, arXiv:
2612 [1906.02025 \[hep-ex\]](https://arxiv.org/abs/1906.02025) (cit. on pp. 30, 33, 72).
- 2613 [22] P. D. Group et al., *Review of Particle Physics*, Progress of Theoretical and Experimental
2614 Physics **2020** (2020), 083C01, ISSN: 2050-3911, eprint: <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf>, URL: <https://doi.org/10.1093/ptep/ptaa104> (cit. on p. 32).
- 2615
- 2616 [23] T. van Ritbergen and R. G. Stuart, *On the precise determination of the Fermi coupling*
2617 *constant from the muon lifetime*, Nuclear Physics B **564** (2000) 343, ISSN: 0550-3213,
2618 URL: [http://dx.doi.org/10.1016/S0550-3213\(99\)00572-6](http://dx.doi.org/10.1016/S0550-3213(99)00572-6) (cit. on p. 32).
- 2619
- 2620 [24] S. Dawson, A. Ismail, and I. Low, *What's in the loop? The anatomy of double Higgs*
2621 *production*, Physical Review D **91** (2015), ISSN: 1550-2368, URL: <http://dx.doi.org/10.1103/PhysRevD.91.115008> (cit. on p. 32).
- 2622
- 2623 [25] A. M. Sirunyan et al., *Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic*
2624 *distributions in the dilepton final state in proton-proton collisions at $\sqrt{s}=13 \text{ TeV}$* ,
2625 *Physical Review D* **102** (2020), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.102.092013> (cit. on p. 32).
- 2626

- 2627 [26] S. Dawson, S. Dittmaier, and M. Spira, *Neutral Higgs-boson pair production at hadron*
 2628 *colliders: QCD corrections*, Physical Review D **58** (1998), ISSN: 1089-4918, URL: <http://dx.doi.org/10.1103/PhysRevD.58.115012> (cit. on p. 33).
- 2630 [27] S. Borowka et al., *Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading*
 2631 *Order with Full Top-Quark Mass Dependence*, Physical Review Letters **117** (2016), ISSN:
 2632 1079-7114, URL: <http://dx.doi.org/10.1103/PhysRevLett.117.012001> (cit. on
 2633 p. 33).
- 2634 [28] J. Baglio et al., *Gluon fusion into Higgs pairs at NLO QCD and the top mass scheme*,
 2635 The European Physical Journal C **79** (2019), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6973-3> (cit. on p. 33).
- 2637 [29] D. de Florian and J. Mazzitelli, *Higgs Boson Pair Production at Next-to-Next-to-*
 2638 *Leading Order in QCD*, Physical Review Letters **111** (2013), ISSN: 1079-7114, URL:
 2639 <http://dx.doi.org/10.1103/PhysRevLett.111.201801> (cit. on p. 33).
- 2640 [30] D. Y. Shao, C. S. Li, H. T. Li, and J. Wang, *Threshold resummation effects in Higgs*
 2641 *boson pair production at the LHC*, Journal of High Energy Physics **2013** (2013), ISSN:
 2642 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)169](http://dx.doi.org/10.1007/JHEP07(2013)169) (cit. on p. 33).
- 2643 [31] D. de Florian and J. Mazzitelli, *Higgs pair production at next-to-next-to-leading loga-*
 2644 *rithmic accuracy at the LHC*, 2015, arXiv: [1505.07122 \[hep-ph\]](https://arxiv.org/abs/1505.07122) (cit. on p. 33).
- 2645 [32] M. Grazzini et al., *Higgs boson pair production at NNLO with top quark mass effects*,
 2646 Journal of High Energy Physics **2018** (2018), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP05\(2018\)059](http://dx.doi.org/10.1007/JHEP05(2018)059) (cit. on p. 33).
- 2648 [33] J. Baglio et al., *gg → HH: Combined uncertainties*, Physical Review D **103** (2021),
 2649 ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.103.056002> (cit. on
 2650 p. 33).

- 2651 [34] F. Dulat, A. Lazopoulos, and B. Mistlberger, *iHixs 2 — Inclusive Higgs cross sections*,
 2652 Computer Physics Communications **233** (2018) 243, ISSN: 0010-4655, URL: <http://dx.doi.org/10.1016/j.cpc.2018.06.025> (cit. on p. 33).
- 2654 [35] ATLAS Collaboration, *Measurement prospects of the pair production and self-coupling of*
 2655 *the Higgs boson with the ATLAS experiment at the HL-LHC*, ATL-PHYS-PUB-2018-053,
 2656 2018, URL: <https://cds.cern.ch/record/2652727> (cit. on p. 34).
- 2657 [36] D. de Florian, I. Fabre, and J. Mazzitelli, *Triple Higgs production at hadron colliders at*
 2658 *NNLO in QCD*, Journal of High Energy Physics **2020** (2020), ISSN: 1029-8479, URL:
 2659 [http://dx.doi.org/10.1007/JHEP03\(2020\)155](http://dx.doi.org/10.1007/JHEP03(2020)155) (cit. on p. 34).
- 2660 [37] A. Papaefstathiou, T. Robens, and G. Tetlalmatzi-Xolocotzi, *Triple Higgs boson pro-*
 2661 *duction at the Large Hadron Collider with Two Real Singlet scalars*, Journal of High
 2662 Energy Physics **2021** (2021), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP05\(2021\)193](http://dx.doi.org/10.1007/JHEP05(2021)193) (cit. on p. 34).
- 2664 [38] T. Robens, T. Stefaniak, and J. Wittbrodt, *Two-real-scalar-singlet extension of the SM:*
 2665 *LHC phenomenology and benchmark scenarios*, The European Physical Journal C **80**
 2666 (2020), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-020-7655-x> (cit. on p. 34).
- 2668 [39] D. Egana-Ugrinovic, S. Homiller, and P. Meade, *Multi-Higgs boson production probes*
 2669 *Higgs sector flavor*, Physical Review D **103** (2021), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.103.115005> (cit. on p. 34).
- 2671 [40] J. Pequenao, “Computer generated image of the whole ATLAS detector”, 2008, URL:
 2672 <https://cds.cern.ch/record/1095924> (cit. on p. 38).
- 2673 [41] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
 2674 JINST **3** (2008) S08003 (cit. on p. 39).
- 2675 [42] J. Pequenao and P. Schaffner, “How ATLAS detects particles: diagram of particle paths
 2676 in the detector”, 2013, URL: <https://cds.cern.ch/record/1505342> (cit. on p. 40).

- 2677 [43] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-
 2678 TDR-19; CERN-LHCC-2010-013, 2010, URL: <https://cds.cern.ch/record/1291633>
 2679 (cit. on p. 41), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL:
 2680 <https://cds.cern.ch/record/1451888>.
- 2681 [44] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST
 2682 **13** (2018) T05008, arXiv: [1803.00844 \[physics.ins-det\]](https://arxiv.org/abs/1803.00844) (cit. on p. 41).
- 2683 [45] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys.
 2684 J. C **77** (2017) 317, arXiv: [1611.09661 \[hep-ex\]](https://arxiv.org/abs/1611.09661) (cit. on p. 44).
- 2685 [46] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-
 2686 PUB-2021-001, 2021, URL: <https://cds.cern.ch/record/2767187> (cit. on p. 44).
- 2687 [47] J. Alwall et al., *The automated computation of tree-level and next-to-leading order*
 2688 *differential cross sections, and their matching to parton shower simulations*, Journal of
 2689 High Energy Physics **2014** (2014), URL: [https://doi.org/10.1007/jhep07\(2014\)079](https://doi.org/10.1007/jhep07(2014)079)
 2690 (cit. on pp. 47, 74).
- 2691 [48] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*,
 2692 JHEP **11** (2004) 040, arXiv: [hep-ph/0409146 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0409146) (cit. on pp. 48, 75).
- 2693 [49] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton*
 2694 *shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: [0709.2092](https://arxiv.org/abs/0709.2092)
 2695 [hep-ph] (cit. on pp. 48, 75).
- 2696 [50] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO*
 2697 *calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010)
 2698 043, arXiv: [1002.2581 \[hep-ph\]](https://arxiv.org/abs/1002.2581) (cit. on pp. 48, 75).
- 2699 [51] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, The European Physical Journal
 2700 C **76** (2016), URL: <https://doi.org/10.1140/epjc/s10052-016-4018-8> (cit. on
 2701 pp. 48, 74).

- 2702 [52] M. Bähr et al., *Herwig++ physics and manual*, The European Physical Journal C **58**
2703 (2008) 639, URL: <https://doi.org/10.1140/epjc/s10052-008-0798-9> (cit. on
2704 pp. 48, 74).
- 2705 [53] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications
2706 **191** (2015) 159, ISSN: 0010-4655, URL: <https://www.sciencedirect.com/science/article/pii/S0010465515000442> (cit. on pp. 48, 74).
- 2707 [54] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A
2709 **462** (2001) 152 (cit. on pp. 48, 74).
- 2710 [55] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods
2711 in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated
2712 Equipment **506** (2003) 250, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688> (cit. on p. 49).
- 2714 [56] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70**
2715 (2010) 823, arXiv: [1005.4568 \[physics.ins-det\]](https://arxiv.org/abs/1005.4568) (cit. on pp. 49, 74).
- 2716 [57] ATLAS Collaboration, *The new Fast Calorimeter Simulation in ATLAS*, ATL-SOFT-
2717 PUB-2018-002, 2018, URL: <https://cds.cern.ch/record/2630434> (cit. on pp. 50,
2718 51).
- 2719 [58] M. Aharrouche et al., *Energy linearity and resolution of the ATLAS electromagnetic
2720 barrel calorimeter in an electron test-beam*, Nuclear Instruments and Methods in Physics
2721 Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment
2722 **568** (2006) 601, ISSN: 0168-9002, URL: <http://dx.doi.org/10.1016/j.nima.2006.07.053> (cit. on p. 51).
- 2724 [59] ATLAS Collaboration, *Evidence for prompt photon production in pp collisions at
2725 $\sqrt{s} = 7 \text{ TeV}$ with the ATLAS detector*, ATLAS-CONF-2010-077, 2010, URL: <https://cds.cern.ch/record/1281368> (cit. on p. 51).

- 2727 [60] D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, 2014, arXiv: [1312.6114](https://arxiv.org/abs/1312.6114)
 2728 [[stat.ML](#)] (cit. on p. 52).
- 2729 [61] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the*
 2730 *ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) 466, arXiv: [1703.10485](https://arxiv.org/abs/1703.10485) [[hep-ex](#)] (cit. on
 2731 pp. 59, 73).
- 2732 [62] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its*
 2733 *performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) 490, arXiv: [1603 . 02934](https://arxiv.org/abs/1603.02934)
 2734 [[hep-ex](#)] (cit. on p. 59).
- 2735 [63] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet Re-*
 2736 *construction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017,
 2737 URL: <https://cds.cern.ch/record/2268678> (cit. on p. 59).
- 2738 [64] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_T jet clustering algorithm*, *Journal of*
 2739 *High Energy Physics* **2008** (2008) 063, ISSN: 1029-8479, URL: [http://dx.doi.org/10.](http://dx.doi.org/10.1088/1126-6708/2008/04/063)
 2740 [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063) (cit. on p. 59).
- 2741 [65] G. P. Salam, *Towards jetography*, *The European Physical Journal C* **67** (2010) 637, ISSN:
 2742 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-010-1314-6> (cit. on
 2743 p. 59).
- 2744 [66] A. Collaboration, *Configuration and performance of the ATLAS b-jet triggers in Run 2*,
 2745 2021, arXiv: [2106 . 03584](https://arxiv.org/abs/2106.03584) [[hep-ex](#)] (cit. on p. 63).
- 2746 [67] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measure-*
 2747 *ment with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13 \text{ TeV}$* , *Eur. Phys. J. C* **79** (2019) 970,
 2748 arXiv: [1907 . 05120](https://arxiv.org/abs/1907.05120) [[hep-ex](#)] (cit. on pp. 62, 73).
- 2749 [68] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification of*
 2750 *b-jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-PHYS-
 2751 PUB-2018-025, 2018, URL: <https://cds.cern.ch/record/2645405> (cit. on p. 65).

- 2752 [69] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear
 2753 Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,
 2754 Detectors and Associated Equipment **262** (1987) 444, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874> (cit. on
 2755 p. 65).
- 2756
- 2757 [70] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent
 2758 Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, URL:
 2759 <https://cds.cern.ch/record/2255226> (cit. on p. 66).
- 2760
- 2761 [71] *Expected performance of the 2019 ATLAS b-taggers*, <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/>, Accessed: 2021-07-14 (cit. on p. 68).
- 2762
- 2763 [72] ATLAS Collaboration, *Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode
 2764 at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, JHEP **04** (2019) 092, arXiv: [1811.04671](https://arxiv.org/abs/1811.04671)
 [hep-ex] (cit. on p. 72).
- 2765
- 2766 [73] ATLAS Collaboration, *A search for resonant and non-resonant Higgs boson pair pro-
 2767 duction in the $b\bar{b}\tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS
 2768 detector*, Phys. Rev. Lett. **121** (2018) 191801, arXiv: [1808.00336](https://arxiv.org/abs/1808.00336) [hep-ex] (cit. on
 p. 72), Erratum: Phys. Rev. Lett. **122** (2019) 089901.
- 2769
- 2770 [74] ATLAS Collaboration, *Search for Higgs boson pair production in the $WW^{(*)}WW^{(*)}$
 2771 decay channel using ATLAS data recorded at $\sqrt{s} = 13 \text{ TeV}$* , JHEP **05** (2019) 124, arXiv:
[1811.11028](https://arxiv.org/abs/1811.11028) [hep-ex] (cit. on p. 72).
- 2772
- 2773 [75] ATLAS Collaboration, *Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state
 2774 with 13 TeV pp collision data collected by the ATLAS experiment*, JHEP **11** (2018) 040,
 arXiv: [1807.04873](https://arxiv.org/abs/1807.04873) [hep-ex] (cit. on p. 72).
- 2775
- 2776 [76] ATLAS Collaboration, *Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel
 2777 using pp collision data recorded at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, Eur. Phys.
 J. C **78** (2018) 1007, arXiv: [1807.08567](https://arxiv.org/abs/1807.08567) [hep-ex] (cit. on p. 72).
- 2778

- 2778 [77] CMS Collaboration, *Search for heavy resonances decaying to two Higgs bosons in final*
 2779 *states containing four b quarks*, *Eur. Phys. J. C* **76** (2016) 371, arXiv: [1602.08762](https://arxiv.org/abs/1602.08762)
 2780 [[hep-ex](#)] (cit. on p. 72).
- 2781 [78] CMS Collaboration, *Search for production of Higgs boson pairs in the four b quark final*
 2782 *state using large-area jets in proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$* , *JHEP* **01** (2019)
 2783 040, arXiv: [1808.01473](https://arxiv.org/abs/1808.01473) [[hep-ex](#)] (cit. on p. 72).
- 2784 [79] CMS Collaboration, *Combination of Searches for Higgs Boson Pair Production in*
 2785 *Proton–Proton Collisions at $\sqrt{s} = 13 \text{ TeV}$* , *Phys. Rev. Lett.* **122** (2019) 121803, arXiv:
 2786 [1811.09689](https://arxiv.org/abs/1811.09689) [[hep-ex](#)] (cit. on p. 73).
- 2787 [80] G. Bartolini et al., *Performance of the ATLAS b-jet trigger in p p collisions at $\sqrt{s} =$*
 2788 *13 TeV* , tech. rep. ATL-COM-DAQ-2019-150, CERN, 2019, URL: <https://cds.cern.ch/record/2688819> (cit. on p. 73).
- 2790 [81] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13 \text{ TeV}$*
 2791 *using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: <https://cds.cern.ch/record/2677054> (cit. on p. 74).
- 2793 [82] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43**
 2794 (2016) 023001, arXiv: [1510.03865](https://arxiv.org/abs/1510.03865) [[hep-ph](#)] (cit. on p. 75).
- 2795 [83] M. Bahr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J. C* **58** (2008) 639, arXiv:
 2796 [0803.0883](https://arxiv.org/abs/0803.0883) [[hep-ph](#)] (cit. on p. 75).
- 2797 [84] T. Head et al., *scikit-optimize/scikit-optimize: v0.5.2*, version v0.5.2, 2018, URL: <https://doi.org/10.5281/zenodo.1207017> (cit. on p. 80).
- 2799 [85] ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state*
 2800 *using proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, *JHEP* **01**
 2801 (2019) 030, arXiv: [1804.06174](https://arxiv.org/abs/1804.06174) [[hep-ex](#)] (cit. on pp. 86, 92).

- 2802 [86] A. Carvalho et al., *Higgs pair production: choosing benchmarks with cluster analysis*,
 2803 Journal of High Energy Physics **2016** (2016) 1, ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP04\(2016\)126](http://dx.doi.org/10.1007/JHEP04(2016)126) (cit. on p. 90).
- 2805 [87] G. V. Moustakides and K. Basioti, *Training Neural Networks for Likelihood/Density*
 2806 *Ratio Estimation*, 2019, arXiv: [1911.00405 \[eess.SP\]](https://arxiv.org/abs/1911.00405) (cit. on p. 93).
- 2807 [88] T. Kanamori, S. Hido, and M. Sugiyama, *A Least-Squares Approach to Direct Importance*
 2808 *Estimation*, J. Mach. Learn. Res. **10** (2009) 1391, ISSN: 1532-4435 (cit. on p. 93).
- 2809 [89] S. Fort, H. Hu, and B. Lakshminarayanan, *Deep Ensembles: A Loss Landscape Perspec-*
 2810 *tive*, 2020, arXiv: [1912.02757 \[stat.ML\]](https://arxiv.org/abs/1912.02757) (cit. on p. 95).
- 2811 [90] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7** (1979) 1,
 2812 URL: <https://doi.org/10.1214/aos/1176344552> (cit. on p. 152).
- 2813 [91] J. Adelman et al., *Search for Higgs boson pair production in the $b\bar{b}\gamma\gamma$ final state with*
 2814 *the full Run 2 13 TeV pp collision data collected by the ATLAS Experiment Supporting*
 2815 *note.*, tech. rep., CERN, 2020, URL: <https://cds.cern.ch/record/2711865> (cit. on
 2816 p. 160).
- 2817 [92] A. L. Read, *Presentation of search results: the CLs technique*, Journal of Physics G:
 2818 Nuclear and Particle Physics **28** (2002) 2693, ISSN: 0954-3899 (cit. on p. 174).
- 2819 [93] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-*
 2820 *based tests of new physics*, The European Physical Journal C **71** (2011), ISSN: 1434-6052
 2821 (cit. on p. 174).
- 2822 [94] A. Sherstinsky, *Fundamentals of Recurrent Neural Network (RNN) and Long Short-*
 2823 *Term Memory (LSTM) network*, Physica D: Nonlinear Phenomena **404** (2020) 132306,
 2824 ISSN: 0167-2789, URL: <http://dx.doi.org/10.1016/j.physd.2019.132306> (cit. on
 2825 p. 181).
- 2826 [95] M. Zaheer et al., *Deep Sets*, 2018, arXiv: [1703.06114 \[cs.LG\]](https://arxiv.org/abs/1703.06114) (cit. on p. 181).

- 2827 [96] J. Zhou et al., *Graph Neural Networks: A Review of Methods and Applications*, 2021,
2828 arXiv: [1812.08434 \[cs.LG\]](https://arxiv.org/abs/1812.08434) (cit. on p. 181).
- 2829 [97] A. Vaswani et al., *Attention Is All You Need*, 2017, arXiv: [1706.03762 \[cs.CL\]](https://arxiv.org/abs/1706.03762) (cit. on
2830 p. 181).
- 2831 [98] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, <http://www.deeplearningbook.org>,
2832 MIT Press, 2016 (cit. on p. 182).
- 2833 [99] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, Journal of Machine
2834 Learning Research **12** (2011) 2825 (cit. on p. 187).