

<sup>1</sup>

©Copyright 2021

<sup>2</sup>

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

52	<b>Chapter 5:</b>	<b>Reconstruction</b>	55
53	5.1	Jets	55
54	5.2	Flavor Tagging	57
55	<b>Chapter 6:</b>	<b>The Anatomy of an LHC Search</b>	67
56	6.1	Object Selection and Identification	67
57	6.2	Defining a Signal Region	67
58	6.3	Background Estimation	67
59	6.4	Uncertainty Estimation	67
60	6.5	Hypothesis Testing	67
61	<b>Chapter 7:</b>	<b>Search for pair production of Higgs bosons in the <math>b\bar{b}b\bar{b}</math> final state</b>	68
62	7.1	Data and Monte Carlo Simulation	70
63	7.2	Triggers and Object Definitions	72
64	7.3	Analysis Selection	74
65	7.4	Background Reduction and Top Veto	80
66	7.5	Kinematic Region Definition	82
67	7.6	Background Estimation	88
68	7.7	Uncertainties	149
69	7.8	Background Validation	160
70	7.9	Overview of Other $b\bar{b}b\bar{b}$ Channels	163
71	7.10	$m_{HH}$ Distributions	164
72	7.11	Statistical Analysis	171
73	7.12	Results	172
74	<b>Chapter 8:</b>	<b>Future Ideas for <math>HH \rightarrow b\bar{b}b\bar{b}</math></b>	177
75	8.1	pairAGraph: A New Method for Jet Pairing	177
76	8.2	Background Estimation with Mass Plane Interpolation	180
77	<b>Chapter 9:</b>	<b>Conclusions</b>	191

## LIST OF FIGURES

Figure Number		Page
79	1.1 Diagram of the elementary particles described by the Standard Model [1]. . . . .	3
80	2.1 Dominant contributing diagrams for non-resonant gluon-gluon fusion production of $HH$ . $\kappa_\lambda$ and $\kappa_t$ represent variations of the Higgs self-coupling and coupling to top quarks respectively, relative to that predicted by the Standard Model. . . . .	29
84	2.2 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. . . . .	30
86	3.1 Diagram of the ATLAS detector [26] . . . . .	35
87	3.2 Cross section of the ATLAS detector showing how particles interact with various detector components [28] . . . . .	37
89	3.3 2D projections of the ATLAS coordinate system . . . . .	38
90	4.1 Energy and lateral shower width variable, weta2, for 16 GeV photons with full simulation (G4) and FastCaloSimV2 (FCSV2) [43]. . . . .	48
92	4.2 Example of photon and pion average shapes in $5 \times 5$ 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. . . . .	50
97	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 $\alpha$ , 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. . . . .	51

103	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 $\alpha$ , 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. . . . .	52
111	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, $\eta$ points, and layers. . . . .	53
114	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. . . . .	54
118	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. [52] . . . . .	60
125	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. . . . .	64
130	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. [57] . . . . .	65

136	7.1 Comparison of $m_{H_1}$ vs $m_{H_2}$ 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 $m_{H_1}$ and $m_{H_2}$ is introduced by the top veto described in Section 7.4.	78
143	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. . . . .	79
148	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. . . . .	83
153	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. . . . .	85
158	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). . . . .	86
162	7.6 <b>Resonant Search:</b> Distributions of $\Delta R$ between the closest Higgs Candidate jets, $\Delta R$ between the other two, and average absolute value of HC jet $\eta$ before and after CR derived reweighting for the 2018 Control Region. . . . .	95
165	7.7 <b>Resonant Search:</b> 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. . . . .	96
168	7.8 <b>Resonant Search:</b> 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. . . . .	97
171	7.9 <b>Resonant Search:</b> Distributions of $p_T$ of the 1st and 3rd leading Higgs Candidate jets and $\Delta R$ between Higgs candidates before and after CR derived reweighting for the 2018 Control Region. . . . .	98

174	<b>7.10 Resonant Search:</b> Distributions of $\eta$ of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Control Region. . . . .	99
175		
176		
177	<b>7.11 Resonant Search:</b> Distributions of $\eta$ 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. . . . .	100
178		
179		
180	<b>7.12 Resonant Search:</b> 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. . . . .	101
181		
182		
183	<b>7.13 Resonant Search:</b> Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region. . . . .	102
184		
185		
186	<b>7.14 Resonant Search:</b> 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 . . . . .	103
187		
188		
189	<b>7.15 Resonant Search:</b> Distributions of $\Delta R$ between the closest Higgs Candidate jets, $\Delta R$ between the other two, and average absolute value of HC jet $\eta$ before and after CR derived reweighting for the 2018 Validation Region. . . . .	104
190		
191		
192	<b>7.16 Resonant Search:</b> 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. . . . .	105
193		
194		
195	<b>7.17 Resonant Search:</b> 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. . . . .	106
196		
197		
198	<b>7.18 Resonant Search:</b> Distributions of $p_T$ of the 1st and 3rd leading Higgs Candidate jets and $\Delta R$ between Higgs candidates before and after CR derived reweighting for the 2018 Validation Region. . . . .	107
199		
200		
201	<b>7.19 Resonant Search:</b> Distributions of $\eta$ of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Validation Region. . . . .	108
202		
203		
204	<b>7.20 Resonant Search:</b> Distributions of $\eta$ 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. . . . .	109
205		
206		
207	<b>7.21 Resonant Search:</b> 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. . . . .	110
208		
209		

210	<b>7.22 Resonant Search:</b> Distributions of $\Delta\phi$ between jets in the leading and 211 subleading Higgs candidates before and after CR derived reweighting for the 212 2018 Validation Region. . . . .	111
213	<b>7.23 Resonant Search:</b> Distributions of the top veto variable, $X_{Wt}$ , before and 214 after CR derived reweighting for the 2018 Validation Region. Reweighting is 215 done after the cut on this variable is applied . . . . .	112
216	<b>7.24 Non-resonant Search (4b):</b> Distributions of $\Delta R$ between the closest Higgs 217 Candidate jets, $\Delta R$ between the other two, and average absolute value of HC 218 jet $\eta$ before and after CR derived reweighting for the 2018 4b Control Region. . . . .	113
219	<b>7.25 Non-resonant Search (4b):</b> Distributions of $p_T$ of the 2nd and 4th leading 220 Higgs Candidate jets and the $p_T$ of the di-Higgs system before and after CR 221 derived reweighting for the 2018 4b Control Region. . . . .	114
222	<b>7.26 Non-resonant Search (4b):</b> Distributions of the number of jets before and 223 after CR derived reweighting for the 2018 4b Control Region. A minimum of 4 224 jets is required in each event in order to form Higgs candidates. . . . .	115
225	<b>7.27 Non-resonant Search (4b):</b> Distributions of $p_T$ of the 1st and 3rd leading 226 Higgs Candidate jets and $\Delta R$ between Higgs candidates before and after CR 227 derived reweighting for the 2018 4b Control Region. . . . .	116
228	<b>7.28 Non-resonant Search (4b):</b> Distributions of $\eta$ of the 1st, 2nd, and 3rd 229 leading Higgs Candidate jets before and after CR derived reweighting for the 230 2018 4b Control Region. . . . .	117
231	<b>7.29 Non-resonant Search (4b):</b> Distributions of $\eta$ of the 4th leading Higgs 232 Candidate jet and the $p_T$ of the leading and subleading Higgs candidates 233 before and after CR derived reweighting for the 2018 4b Control Region. . . . .	118
234	<b>7.30 Non-resonant Search (4b):</b> Distributions of mass of the leading and sub- 235 leading Higgs candidates and of the di-Higgs system before and after CR 236 derived reweighting for the 2018 4b Control Region. . . . .	119
237	<b>7.31 Non-resonant Search (4b):</b> Distributions of $\Delta\phi$ between jets in the leading 238 and subleading Higgs candidates before and after CR derived reweighting for 239 the 2018 4b Control Region. . . . .	120
240	<b>7.32 Non-resonant Search (4b):</b> Distributions of the top veto variable, $X_{Wt}$ , 241 before and after CR derived reweighting for the 2018 4b Control Region. 242 Reweighting is done after the cut on this variable is applied. . . . .	121
243	<b>7.33 Non-resonant Search (4b):</b> Distributions of $\Delta R$ between the closest Higgs 244 Candidate jets, $\Delta R$ between the other two, and average absolute value of HC 245 jet $\eta$ before and after CR derived reweighting for the 2018 4b Validation Region. . . . .	122

246	<b>7.34 Non-resonant Search (4b):</b> 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. . . . .	123
247		
248		
249	<b>7.35 Non-resonant Search (4b):</b> 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. . . . .	124
250		
251		
252	<b>7.36 Non-resonant Search (4b):</b> Distributions of $p_T$ of the 1st and 3rd leading Higgs Candidate jets and $\Delta R$ between Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region. . . . .	125
253		
254		
255	<b>7.37 Non-resonant Search (4b):</b> Distributions of $\eta$ of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Validation Region. . . . .	126
256		
257		
258	<b>7.38 Non-resonant Search (4b):</b> Distributions of $\eta$ 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. . . . .	127
259		
260		
261	<b>7.39 Non-resonant Search (4b):</b> 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. . . . .	128
262		
263		
264	<b>7.40 Non-resonant Search (4b):</b> 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. . . . .	129
265		
266		
267	<b>7.41 Non-resonant Search (4b):</b> 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. . . . .	130
268		
269		
270	<b>7.42 Non-resonant Search (3b1l):</b> Distributions of $\Delta R$ between the closest Higgs Candidate jets, $\Delta R$ between the other two, and average absolute value of HC jet $\eta$ before and after CR derived reweighting for the 2018 3b1l Control Region. . . . .	131
271		
272		
273		
274	<b>7.43 Non-resonant Search (3b1l):</b> 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. . . . .	132
275		
276		
277	<b>7.44 Non-resonant Search (3b1l):</b> 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. . . . .	133
278		
279		
280	<b>7.45 Non-resonant Search (3b1l):</b> Distributions of $p_T$ of the 1st and 3rd leading Higgs Candidate jets and $\Delta R$ between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region. . . . .	134
281		
282		

283	<b>7.46 Non-resonant Search (3b1l):</b> Distributions of $\eta$ of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Control Region. . . . .	135
286	<b>7.47 Non-resonant Search (3b1l):</b> Distributions of $\eta$ 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. . . . .	136
289	<b>7.48 Non-resonant Search (3b1l):</b> 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. . . . .	137
292	<b>7.49 Non-resonant Search (3b1l):</b> 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. . . . .	138
295	<b>7.50 Non-resonant Search (3b1l):</b> 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. . . . .	139
298	<b>7.51 Non-resonant Search (3b1l):</b> Distributions of $\Delta R$ between the closest Higgs Candidate jets, $\Delta R$ between the other two, and average absolute value of HC jet $\eta$ before and after CR derived reweighting for the 2018 3b1l Validation Region. . . . .	140
302	<b>7.52 Non-resonant Search (3b1l):</b> 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. . . . .	141
305	<b>7.53 Non-resonant Search (3b1l):</b> 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. . . . .	142
308	<b>7.54 Non-resonant Search (3b1l):</b> Distributions of $p_T$ of the 1st and 3rd leading Higgs Candidate jets and $\Delta R$ between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region. . . . .	143
311	<b>7.55 Non-resonant Search (3b1l):</b> Distributions of $\eta$ of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Validation Region. . . . .	144
314	<b>7.56 Non-resonant Search (3b1l):</b> Distributions of $\eta$ 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. . . . .	145
317	<b>7.57 Non-resonant Search (3b1l):</b> 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. . . . .	146

320	7.58 Non-resonant Search (3b1l): Distributions of $\Delta\phi$ between jets in the lead-	147
321	ing and subleading Higgs candidates before and after CR derived reweighting	
322	for the 2018 3b1l Validation Region. . . . .	
323	7.59 Non-resonant Search (3b1l): Distributions of the top veto variable, $X_{Wt}$ ,	148
324	before and after CR derived reweighting for the 2018 3b1l Validation Region.	
325	Reweighting is done after the cut on this variable is applied. . . . .	
326	7.60 Illustration of the approximate bootstrap band procedure, shown as a ratio to	153
327	the nominal estimate for the 2017 non-resonant background estimate. Each grey	
328	line is from the $m_{HH}$ prediction for a single bootstrap training. Figure 7.60(a)	
329	shows the variation histograms constructed from median weight $\pm$ the IQR of	
330	the replica weights. It can be seen that this captures the rough shape of the	
331	bootstrap envelope, but is not good estimate for the overall magnitude of the	
332	variation. Figure 7.60(b) demonstrates the applied normalization correction,	
333	and Figure 7.60(c) shows the final band (normalized Figure 7.60(a) + Figure	
334	7.60(b)). Comparing this with the IQR variation for the prediction from each	
335	bootstrap in each bin in Figure 7.60(d), the approximate envelope describes a	
336	very similar variation. . . . .	
337	7.61 Resonant Search: Example of CR vs VR variation in each $H_T$ region for	155
338	2018. The variation nicely factorizes into low and high mass components. . .	
339	7.62 Non-resonant Search (4b): Example of CR vs VR variation in each signal	156
340	region quadrant for 2018. Significantly different behavior is seen between	
341	quadrants, with the largest variation in quadrant 1 and the smallest in quadrant	
342	4. . . . .	
343	7.63 Resonant Search: Performance of the background estimation method in the	161
344	resonant analysis reversed $\Delta\eta_{HH}$ kinematic signal region. A new background	
345	estimate is trained following nominal procedures entirely within the reversed	
346	$\Delta\eta_{HH}$ region, and the resulting model, including uncertainties, is compared	
347	with 4b data in the corresponding signal region. Good agreement is shown.	
348	The the quoted $p$ -value uses the $\chi^2$ test statistic, and demonstrates no evidence	
349	that the data differs from the assessed background. . . . .	
350	7.64 Non-resonant Search: Performance of the background estimation method	162
351	in the 3b + 1 fail validation region. A new background estimate is trained	
352	following nominal procedures but with a reweighting from 2b to 3b + 1 fail	
353	events. Generally good agreement is seen, though there is some deviation at	
354	very low masses in the low $\Delta\eta_{HH}$ low $X_{HH}$ category. . . . .	

355	7.65 <b>Resonant Search:</b> Demonstration of the performance of the nominal reweighting in the control region on corrected $m_{HH}$ , with Figure 7.65(a) showing 2 <i>b</i> events normalized to the total 4 <i>b</i> yield and Figure 7.65(b) applying the reweighting procedure. Agreement is much improved with the reweighting. Note that overall reweighted 2 <i>b</i> yield agrees with 4 <i>b</i> yield in the control region by construction. . . . .	165
361	7.66 <b>Resonant Search:</b> 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. . . . .	166
366	7.67 <b>Resonant Search:</b> 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 $m_{HH}$ . . . . .	167
370	7.68 <b>Non-resonant Search (4b):</b> Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 4 <i>b</i> 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. . . . .	169
376	7.69 <b>Non-resonant Search (3b1l):</b> Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 3 <i>b</i> 1 <i>l</i> channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Conclusions are very similar to the 4 <i>b</i> 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. . . . .	170
382	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 $k/\overline{M}_{Pl} = 1$ [19] (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. . . . .	174



## GLOSSARY

426 ARGUMENT: replacement text which customizes a L<sup>A</sup>T<sub>E</sub>X macro for each particular usage.

## ACKNOWLEDGMENTS

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

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

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

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

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

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

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

457

## DEDICATION

458

To family, both given and found

459

## Chapter 1

460

# THE STANDARD MODEL OF PARTICLE PHYSICS

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

469 **1.1 Introduction: Particles and Fields**

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

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

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

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

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

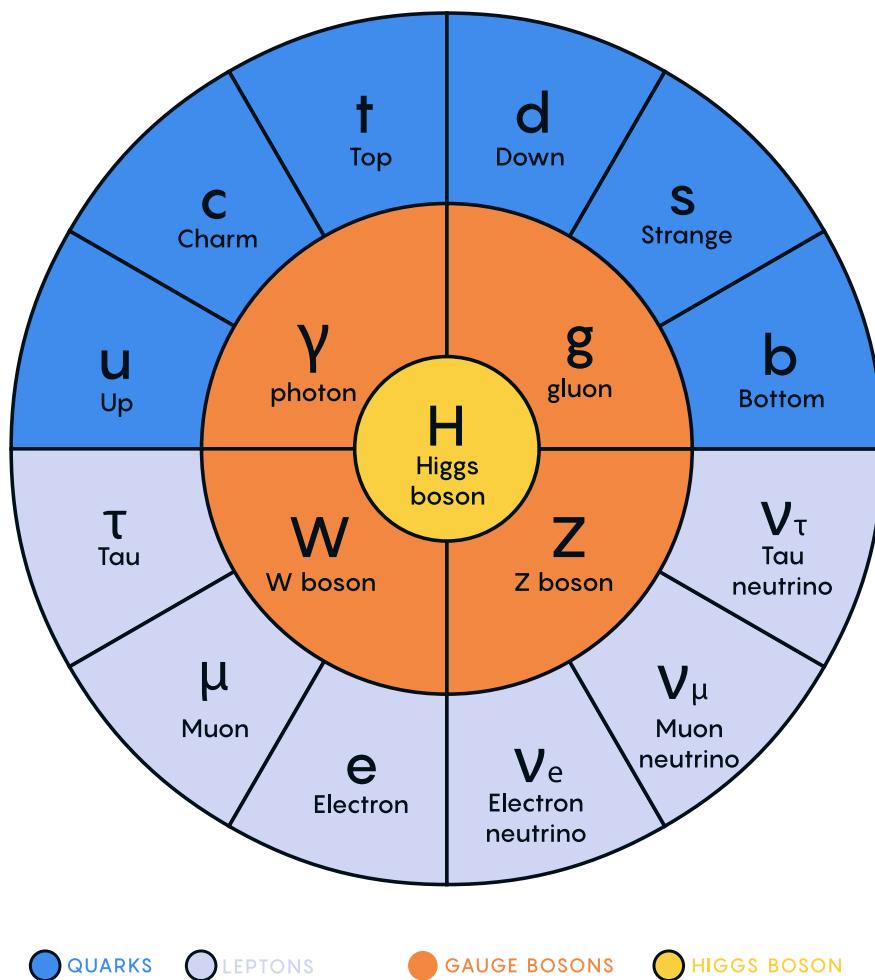


Figure 1.1: Diagram of the elementary particles described by the Standard Model [1].

<sup>498</sup> **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*:  $\phi$  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:

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

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

<sup>499</sup> 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:

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

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

505 We know that electrons experience electromagnetic interactions with other objects. Given  
 506 this, and the fact that such interactions must be transmitted *somewhat* between e.g. two  
 507 electrons, it seems natural that these interactions are facilitated by electromagnetic radiation.  
 508 More specifically, we may think of photons as *mediators* of the electromagnetic force. It  
 509 follows, then, that a description of electromagnetism on the level of particles must involve a  
 510 description of both the "source" particles (e.g. electrons), the mediators (photons), and their  
 511 interactions. Further, this description must be (1) relativistic and (2) consistent with the  
 512 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  $\partial_\mu$  is as defined above,  $\psi$  is a Dirac *spinor*, i.e. a four-component wavefunction,  $m$  is the mass of the particle, and  $\gamma^\mu$  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

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

513 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

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

<sup>514</sup> 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:

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

<sup>515</sup> where the source term may be interpreted as coming from the Dirac fields themselves, namely,

<sup>516</sup>  $-q\bar{\psi}\gamma^\mu\psi A_\mu$ .

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

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

<sup>517</sup> which is exactly the quantum electrodynamics Lagrangian.

<sup>518</sup> We have swept a few things under the rug here, however. Recall that the general form  
<sup>519</sup> of a Lagrangian is conventionally  $\mathcal{L} = T - V$ , where  $T$  is the kinetic term, and thus ought  
<sup>520</sup> to contain a derivative with respect to time (c.f. the standard  $\frac{1}{2}m\frac{\partial x}{\partial t}$  familiar from basic  
<sup>521</sup> kinematics). More particularly, given the definition of conjugate momentum as  $\partial\mathcal{L}/\partial\dot{q}$  for

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

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

528 What does this mean?  $A^\mu$  is a four component object, but it would appear that only three  
529 of the components have dynamics: we have too many degrees of freedom in the theory. This  
530 is the principle behind *gauge symmetry* – an extra constraint on  $A^\mu$  (a *gauge condition*) must  
531 be defined such that a unique  $A^\mu$  defines the theory and satisfies the condition. However,  
532 we are free to choose this extra condition – the physics content of the theory should be  
533 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

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

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

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

536 Under the same transformation of  $\vec{A}$ , the electric field  $\vec{E}$  becomes  $-\nabla\phi - \frac{\partial\vec{A}}{\partial t} - \frac{\partial\nabla\lambda}{\partial t} =$   
537  $-\nabla(\phi + \frac{\partial\lambda}{\partial t}) - \frac{\partial\vec{A}}{\partial t}$ , such that, for the  $\vec{E}$  field to be unchanged, we must additionally apply  
538 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

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

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

<sup>541</sup> 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:

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

<sup>542</sup> since by the arguments above, the physics content of the Lagrangian is invariant under the  
<sup>543</sup> transformation  $A_\mu \rightarrow A_\mu - \partial_\mu \lambda$ , we may directly make this transformation, and remove this  
<sup>544</sup> 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  
<sup>545</sup> this same transformation of  $A_\mu$ , so we may say that the QED Lagrangian is invariant under  
<sup>546</sup> local transformations of the form  $\psi \rightarrow e^{iq\lambda(x)}\psi$ .

<sup>547</sup> These arguments illuminate some important concepts which will serve us well going forward.  
<sup>548</sup> First, while we have remained grounded in the “familiar” physics of electromagnetism for the  
<sup>549</sup> above, arguments of the “top down” variety would lead us to the exact same conclusions.  
<sup>550</sup> That is, suppose we wanted to construct a theory of spin-1/2 particles that was invariant  
<sup>551</sup> under local transformations of the form  $\psi \rightarrow e^{iq\lambda(x)}\psi$ . More broadly, we could say that we  
<sup>552</sup> desire this theory to be invariant under local  $U(1)$  transformations, where  $U(1)$  is exactly  
<sup>553</sup> this group, under multiplication, of complex numbers with absolute value 1. By very similar

554 arguments as above, we would see that, to achieve invariance, this theory would necessitate  
555 an additional degree of freedom,  $A_\mu$ , with the exact properties that are familiar to us from  
556 electrodynamics. These arguments based on symmetries are extremely powerful in building  
557 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  $A_\mu$  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  $A_\mu$ ,

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

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

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

### 565 1.3 An Aside on Group Theory

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

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

574 To set up language to generalize beyond  $U(1)$ , note that we may equivalently write  $U(1)$   
575 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,  
576 with each  $\alpha^a(x)$  an associated scalar function to generator  $t^a$ , and  $g$  is some scalar strength  
577 parameter. Of course this is a bit silly for  $U(1)$ , which has a single generator, and thus  
578 reduces to the transformation we discussed above. However, this becomes much more useful  
579 for groups of higher degree, with more generators and degrees of freedom.

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

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

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

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

599 **1.4 Quantum Chromodynamics**

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

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

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

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

623 Counting up the available types of gluons, then, we come up with nine. Six of mixed  
624 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,  
625 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,  $\lambda^a$ , as  $t^a = \frac{1}{2}\lambda^a$ , and the spinor  $\psi$  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,  $G_\mu^a$ , 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  $A_\mu$  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  $G_\mu^a$  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

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

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

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

<sup>643</sup> for  $G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - f^{abc} G_\mu^b G_\nu^c$ .

<sup>644</sup> This gives the component of the field strength corresponding to a particular gauge field  $a$ ,  
<sup>645</sup> where the first two terms have the familiar form of the QED field strength, while the last  
<sup>646</sup> term is new, and explicitly related to the group structure via the  $f^{abc}$  constants. In terms  
<sup>647</sup> of the physics content of the theory, this latter term gives rise to a gluon *self-interaction*, a  
<sup>648</sup> distinguishing feature of QCD.

<sup>649</sup> Similarly as in QED, a Lorentz invariant combination of field strength tensors may be made  
<sup>650</sup> as  $G_{\mu\nu} G^{\mu\nu}$ . However, this is not manifestly gauge invariant. Under a gauge transformation

- 651  $U$ , the covariant derivative behaves as  $D^\mu \rightarrow UD^\mu U^{-1}$ , corresponding to  $G^{\mu\nu} \rightarrow UG^{\mu\nu}U^{-1}$ .  
652 The cyclic property of the trace thus ensures the gauge invariance of  $\text{tr}(G_{\mu\nu}G^{\mu\nu})$ , which we  
653 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  $\psi$  have associated color charge, which must interact appropriately with the  $G_\mu$  fields. Mathematically, the generators  $t^a$  are  $3 \times 3$  matrices, while the  $\psi$  are four component spinors. Adding a color index to the Dirac fields, i.e.,  $\psi_i$  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

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

- 654 where  $\delta_{ij}$  is the Kronecker delta, as  $\partial_\mu$  does not participate in the  $SU(3)$  structure.

The Lagrangian then becomes

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

- 655 and we have constructed QCD.

## 656 1.5 The Weak Interaction

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

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

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

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

677 This leads to three degrees of freedom for the gauge symmetry of the weak interactions, so  
678 we thus need a theory which is locally invariant under transformations of a group with three  
679 generators. The simplest such choice is  $SU(2)$ . We may follow a very similar prescription as  
680 for QED and QCD:  $SU(2)$  has three generators, which implies the existence of three gauge  
681 bosons, call them  $W_\mu^k$ . The gauge transformation may be expressed as  $\psi \rightarrow e^{ig_W \vec{\alpha}(x) \cdot \vec{T}} \psi$ , where  
682 in this case the generators are for  $SU(2)$ , which may be written in terms of the familiar Pauli  
683 matrices:  $\vec{T} = \frac{1}{2}\vec{\sigma}$ . The structure constants for  $SU(2)$  are the antisymmetric Levi-Civita  
684 tensor, so the corresponding gauge covariant derivative is  $D_\mu \equiv \partial_\mu + ig_W W_\mu^k t^k$ , and the field  
685 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

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

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

687 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  $2 \times 2$ , 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  $\pm 1$  in both weak isospin  $I_W^{(3)}$  and electric charge. We may thus line up all such doublets as:

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

711 with the top corresponding to the lower weak isospin and electric charge particles, and the  
712 lower quark entries ( $d'$ , etc) corresponding to the weak quark eigenstates (which are related  
713 to the mass eigenstates by the CKM matrix *TODO: more detail*). Similar doublets may be  
714 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

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

715 for right and left-handed chiralities respectively – acting on a Dirac field with such operators  
716 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:

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

<sup>717</sup> and right-handed particle states are placed in singlets and assigned 0 charge under  $SU(2)_L$   
<sup>718</sup> ( $I_W = I_W^{(3)} = 0$ ).

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:

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

<sup>719</sup> for left-handed doublets  $Q$  and  $L$  for quarks and electron fields respectively and right handed  
<sup>720</sup> 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

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

<sup>721</sup> where the  $f$  are understood to run over the fermion chiral doublets and singlets as above.

This then leaves our Lagrangian as

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

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

<sup>722</sup> where we have expanded the covariant derivative for clarity. You may note that we have  
<sup>723</sup> 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

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

<sup>724</sup> with *weak hypercharge*  $Y$ .

Similarly to our discussion of QED, we may write the  $U(1)_Y$  gauge field as  $B_\mu$ , 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  $B_\mu$  field and classical electrodynamics is not so obvious – however it is convenient to parametrize as

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

<sup>725</sup> where  $A_\mu$  and  $Z_\mu$  are the physical fields, and we pick  $W_\mu^3$  as the neutral weak boson.

<sup>726</sup> Note that in the  $SU(2)_L \times U(1)_Y$  theory, the Lagrangian must be invariant under all of  
<sup>727</sup> the local gauge transformations. In particular, this means that the hypercharge must be the  
<sup>728</sup> same for fermion fields in each weak doublet to preserve  $U(1)_Y$  invariance. This gives insight  
<sup>729</sup> into the relation between the charges of  $SU(2)_L \times U(1)_Y$  and electric charge. In particular  
<sup>730</sup> we know that the hypercharge,  $Y$ , of  $e^-$  ( $I_W^{(3)} = -\frac{1}{2}$ ) and  $\nu_e$  ( $I_W^{(3)} = +\frac{1}{2}$ ) 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,

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

<sup>731</sup> Some of these particular forms are best understood in the context of the Higgs mechanism  
<sup>732</sup> – we will return to this discussion below.

<sup>733</sup> **1.6 The Higgs Potential and the SM**

<sup>734</sup> In the above, we have neglected a discussion of masses. However there are several things to  
<sup>735</sup> sort out here. In the first place, we know experimentally that the weak interactions occur  
<sup>736</sup> over very short ranges at low energies (e.g., why Fermi's effective four fermion interaction was  
<sup>737</sup> such a good description). This is consistent with massive  $W^\pm$  and  $Z$  bosons (and indeed, this  
<sup>738</sup> is seen experimentally). However, requiring local gauge invariance forbids mass terms in the  
<sup>739</sup> Lagrangian. In the simple  $U(1)$  QED example, such a term would have the form  $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$ ,  
<sup>740</sup> which is not invariant under the transformation  $A_\mu \rightarrow A_\mu - \partial_\mu \lambda$ , and similar arguments hold  
<sup>741</sup> 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:

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

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

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

<sup>742</sup> 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  
<sup>743</sup> and right-handed particles transform differently under  $SU(2)_L$ , this is manifestly not gauge  
<sup>744</sup> invariant.

<sup>745</sup> The question then becomes: how do we include particle masses while preserving the  
<sup>746</sup> gauge properties of our theory? The answer, due to Robert Brout and François Englert [11],  
<sup>747</sup> Peter Higgs [12], and Gerald Guralnik, Richard Hagen, and Tom Kibble [13] comes via the  
<sup>748</sup> Higgs mechanism, which we will describe in the following. Importantly for this thesis, this  
<sup>749</sup> mechanism predicts the existence of a physical particle, the Higgs boson, and a particle  
<sup>750</sup> 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  $\phi^+$  and  $\phi^0$  to the Standard Model embedded in a weak isospin doublet  $\phi$ . We may write the doublet as

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

<sup>751</sup> 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

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

<sup>752</sup> we may notice that this has some interesting properties. Considering, as illustration, a similar  
<sup>753</sup> potential for a real scalar field,  $\mu^2 \chi^2 + \lambda \chi^4$ , taking the derivative and setting it equal to 0  
<sup>754</sup> 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  
<sup>755</sup> unique minimum at  $\chi = 0$ , and for  $\mu^2 < 0$  there are degenerate minima at  $\chi = \pm\sqrt{-\frac{\mu^2}{2\lambda}}$ .  
<sup>756</sup> 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

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

<sup>757</sup> where we have explicitly chosen a real, non-zero vacuum expectation value for the neutral  
<sup>758</sup> component of the Higgs doublet to maintain a massless photon, as we shall see. Physically,  
<sup>759</sup> 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

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

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

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

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

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

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

Expanding the kinetic term

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

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

<sup>767</sup> in a similar way, completing the story of three massless degrees of freedom (Goldstone bosons)  
<sup>768</sup> 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

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

<sup>769</sup> 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

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

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

<sup>770</sup> As identified above,  $Y = 2(Q - I_W^{(3)})$ . The Higgs has 0 electric charge, and the lower doublet  
<sup>771</sup> 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  $W_\mu^1$  and  $W_\mu^2$  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  $\theta_W$  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

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

<sup>772</sup> which is proportional to the  $W$  mass – an analysis with the form of the physical  $Z$  boson  
<sup>773</sup> 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

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

<sup>774</sup> is manifestly not gauge invariant because right and left handed particles transform differently  
<sup>775</sup> under  $SU(2)_L$ . However, because the Higgs is constructed via an  $SU(2)_L$  doublet,  $\phi$ , writing  
<sup>776</sup> 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

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

<sup>777</sup> where  $g_f$  is a corresponding Yukawa coupling,  $f_1$  and  $f_2$  have been used to denote components  
<sup>778</sup> 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

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

<sup>779</sup> Notably here, the fermion coupling to the Higgs boson scales with the mass of the fermion, a  
<sup>780</sup> 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

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

<sup>781</sup> The argument proceeds similarly to the above, with similar results for couplings and masses  
<sup>782</sup> of upper components.

### <sup>783</sup> 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

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

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

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

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

802

## Chapter 2

### 803 DI-HIGGS PHENOMENOLOGY AND PHYSICS BEYOND 804 THE STANDARD MODEL

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

812 **2.1 Intro to Di-Higgs**

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

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

821 As shown in Chapter 1, the Higgs coupling to fermions scales with particle mass. As the  
 822 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  
 823 kinematically disfavored,  $H \rightarrow b\bar{b}$  is the dominant fermionic Higgs decay mode, and, in fact,  
 824 the dominant overall decay mode, with a branching fraction of around 58 %. The dominant

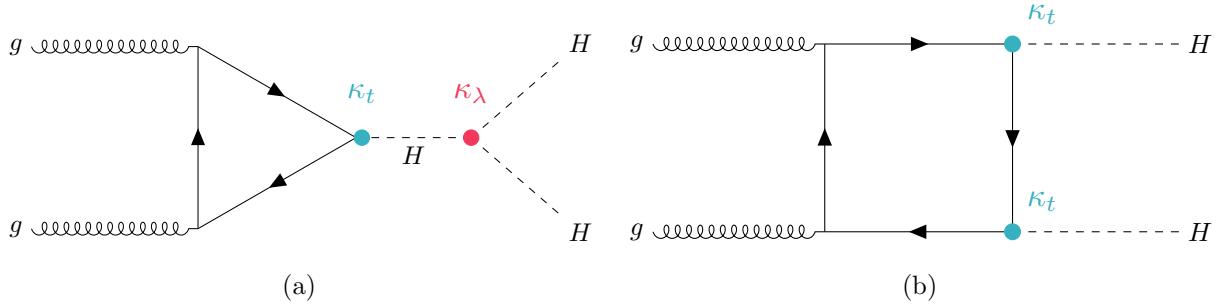


Figure 2.1: Dominant contributing diagrams for non-resonant gluon-gluon fusion production of  $HH$ .  $\kappa_\lambda$  and  $\kappa_t$  represent variations of the Higgs self-coupling and coupling to top quarks respectively, relative to that predicted by the Standard Model.

825 top quark Yukawa coupling to the  $H$  does play a role in  $H$  production, however – gluon-gluon  
826 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 2.2.

## 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.

The spin-2 signal considered is implemented within the bulk Randall-Sundrum (RS) model [18, 19], 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

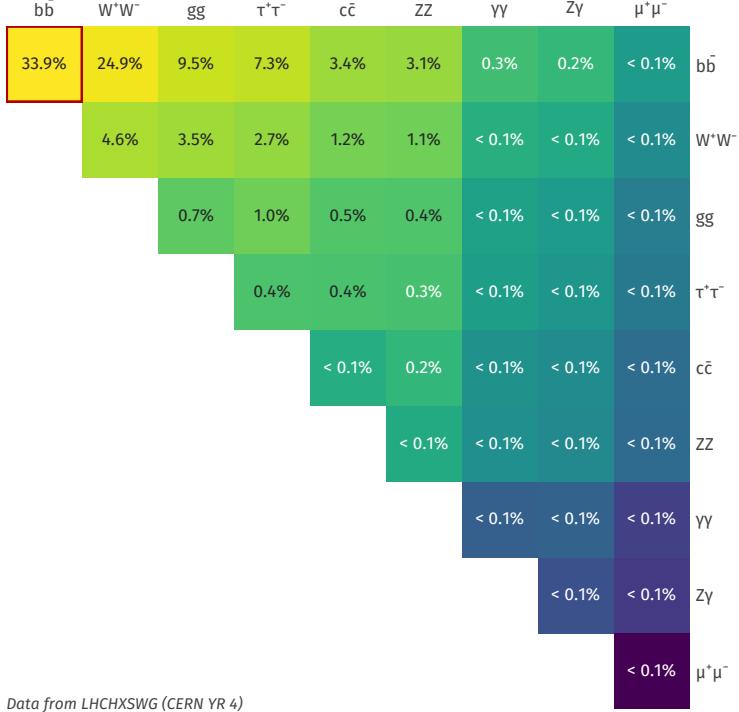


Figure 2.2: Illustration of dominant  $HH$  branching ratios.  $HH \rightarrow b\bar{b}b\bar{b}$  is the most common decay mode, representing 34 % of all  $HH$  events produced at the LHC.

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  $\overline{M}_{\text{Pl}}$  is the Planck mass. The model considered here has  $c = 1.0$ . However, this model was considered in the early Run 2  $HH$  analyses [20], 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 [21], 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

850 are “eaten” by the electroweak bosons, leaving five degrees of freedom which may correspond  
851 to physical fields.

852 **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

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

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

857 The relevant diagrams for non-resonant  $HH$  production are shown in Figure 2.1. Notably,  
858 the diagrams *interfere* with each other, which can be easily seen by counting the fermion  
859 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

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

860 where  $\lambda_{HHH}$  is a varied coupling, whereas  $\lambda_{HHH}^{SM}$  is the Standard Model prediction, given by  
861 As this variation only impacts the *triangle* diagram, significant and interesting effects are

862 observed due to the interference. Examples of the impact of this tradeoff on the di-Higgs  
863 invariant mass are shown in *TODO: include plot*. Generally speaking, the triangle diagram  
864 contributes more at low mass, while the box diagram contributes more at high mass, and for  
865 positive values of  $\kappa_\lambda$ , more events are predicted at low mass, whereas for negative values of  
866  $\kappa_\lambda$ , more events are predicted at high mass.

867

## Chapter 3

868

# EXPERIMENTAL APPARATUS

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

880     **3.1 The Large Hadron Collider**

881     The Large Hadron Collider is a particle accelerator near Geneva, Switzerland, operating  
 882     at a center of mass energy  $\sqrt{s} = 13$  TeV. In broad scope, it is a ring with a 27 kilometer  
 883     circumference. Hadrons (usually protons or heavy ions) move in two counter-circulating  
 884     beams, which are made to collide at four collision points at various points on the ring. These  
 885     four collision points correspond to the four detectors placed around the ring: two “general  
 886     purpose” experiments: ATLAS and CMS; LHCb, focused primarily on flavor physics; and  
 887     ALICE, focused primarily on heavy ions.

888     For proton-proton collisions, the focus of this thesis, the acceleration chain proceeds as  
 889     follows: first, an electric field strips hydrogen of its electrons, creating protons. A linear

890 accelerator, LINAC 2, accelerates protons to 50 MeV. The resulting beam is injected into  
 891 the Proton Synchrotron Booster (PSB), which pushes the protons to 1.4 GeV, and then the  
 892 Proton Synchrotron, which brings the beam to 25 GeV.

893 Protons are then transferred to the Super Proton Synchrotron (SPS), which ramps up  
 894 the energy to 450 GeV. Finally, the protons enter the LHC itself, bringing the beam up to  
 895 6.5 TeV. *TODO: cite: <https://home.cern/science/accelerators/accelerator-complex>*

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

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

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

917 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.

### 3.2 The ATLAS Experiment

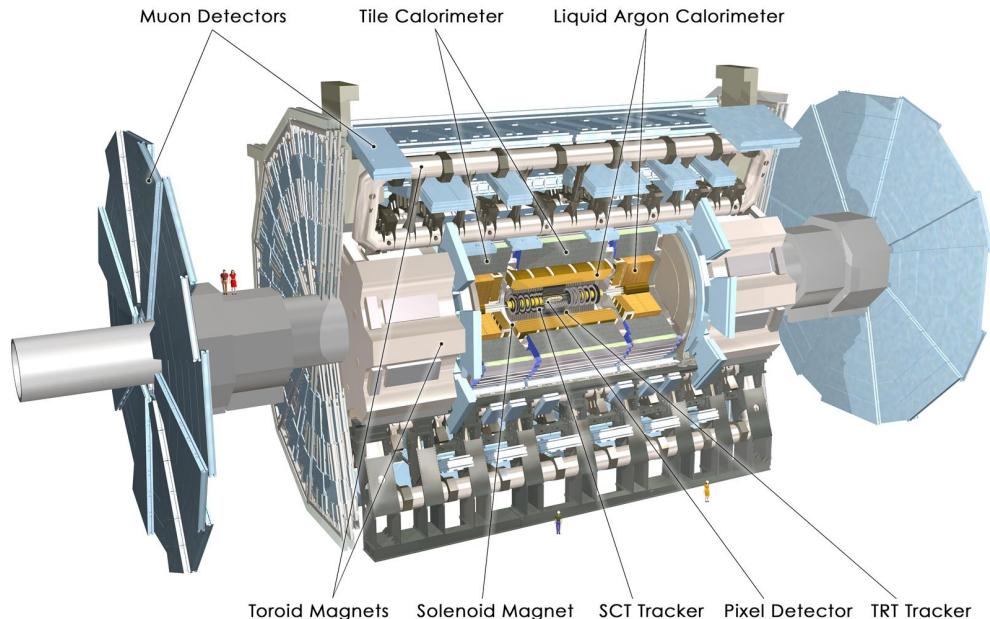


Figure 3.1: Diagram of the ATLAS detector [26]

This thesis focuses on searches done with the ATLAS experiment. As mentioned, this is one of two “general purpose” experiments at the LHC, by which we mean there is a very large and

broad variety of physics done within the experimental collaboration. This broad physics focus has a direct relation to the design of the ATLAS detector [27], pictured in Figure 3.1, which is composed of a sophisticated set of subsystems designed to fully characterize the physics of a given high energy particle collision. It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadronic calorimeters, and a muon spectrometer incorporating three large superconducting toroidal magnets. The ATLAS detector covers nearly the entire solid angle around the collision point, fully characterizing the “visible” components of a collision and allowing for indirect sensitivity to particles that do not interact with the detector (e.g. neutrinos) via “missing” energy (roughly momentum balance). We will go through the design and physics contribution of each of the detector components in the following. A schematic of how various particles interact with the detector is shown in Figure 3.2.

#### 3.2.1 ATLAS Coordinate System

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

#### 3.2.2 Inner Detector

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

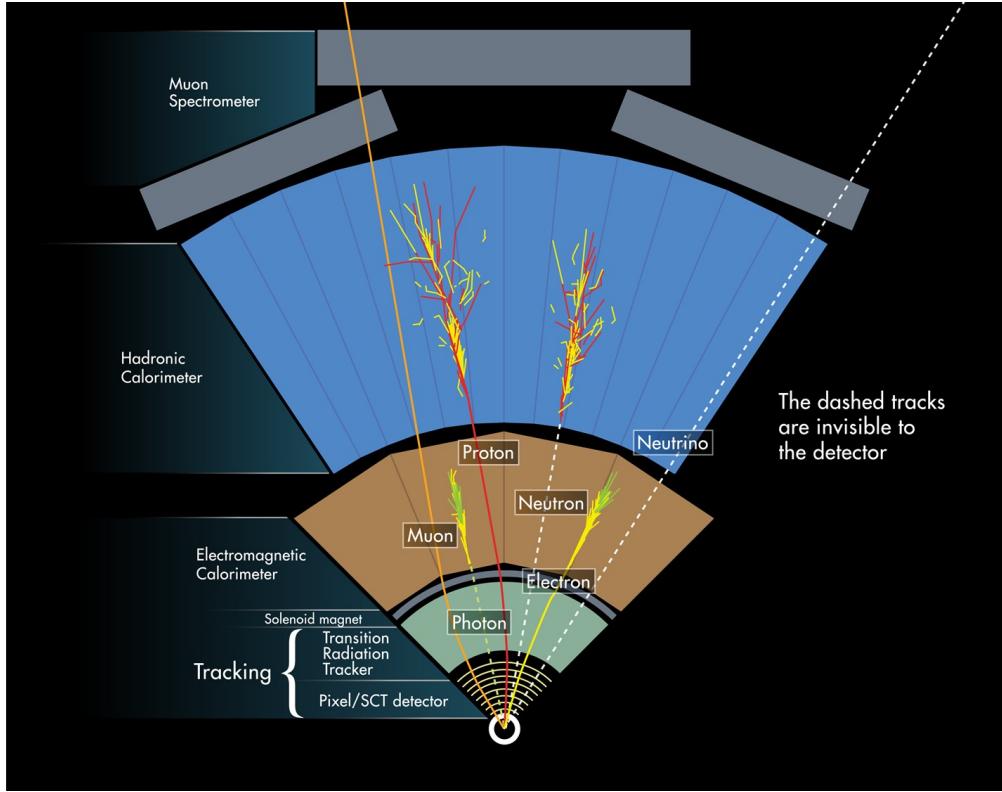


Figure 3.2: Cross section of the ATLAS detector showing how particles interact with various detector components [28]

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

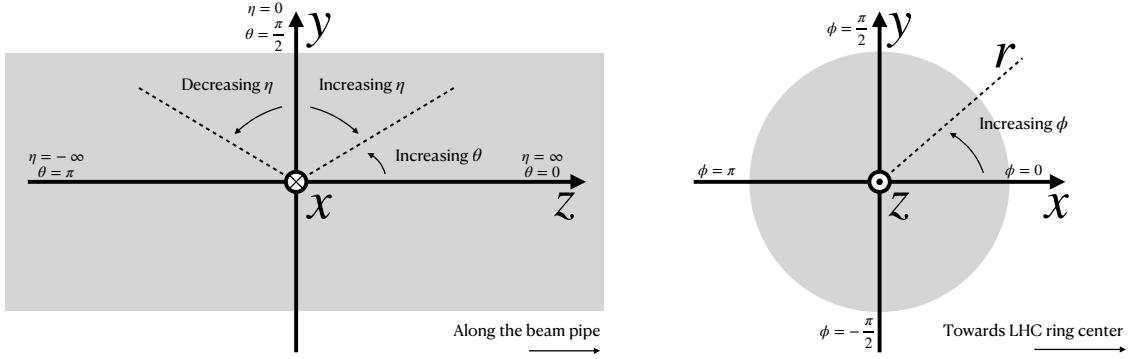


Figure 3.3: 2D projections of the ATLAS coordinate system

965 (IBL) installed before Run 2 [29, 30]. This is very close to the interaction point with a  
966 high degree of positional information, and is therefore very important for e.g.  $b$ -tagging (see  
967 Chapter 5). It is followed by the silicon microstrip tracker (SCT), which usually provides  
968 eight measurements per track. This is lower granularity, but similar in concept to the pixel  
969 detector.

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

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

981 *3.2.3 Calorimeter*

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

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

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

999 The hadronic calorimeter is composed of alternating steel and plastic scintillator tiles,  
 1000 segmented into three barrel structures within  $|\eta| < 1.7$ , in addition to two copper/liquid-argon  
 1001 endcap calorimeters.

1002 The solid angle coverage is completed with forward copper/liquid-argon and tungsten/liquid-  
 1003 argon calorimeter modules optimized for electromagnetic and hadronic energy measurements  
 1004 respectively.

1005    3.2.4 Muon Spectrometer

1006    While muons interact electromagnetically, they are around 200 times heavier than electrons  
 1007    ( $m_\mu = 106 \text{ MeV}$ , while  $m_e = 0.510 \text{ MeV}$ ). Therefore, electromagnetic interactions with ab-  
 1008    sorbers in the calorimeter are not sufficient to stop them, and, as they do not interact via the  
 1009    strong force, hard scattering with nuclei is rare. A dedicated system for muon measurements  
 1010    is therefore required.

1011    The muon spectrometer (MS) is the outermost layer of ATLAS and is designed for this  
 1012    purpose. It is composed of three parts: a set of triggering chambers, which detect if there is  
 1013    a muon and provide a coordinate measurement, in conjunction with high-precision tracking  
 1014    chambers, which measure the deflection of muons in a magnetic field to measure muon  
 1015    momentum, similar to the inner detector solenoid. The magnetic field is generated by the  
 1016    superconducting air-core toroidal magnets, with a field integral between 2.0 and 6.0 T m  
 1017    across most of the detector. The toroid magnetic field runs roughly in a circle in the  $x - y$   
 1018    plane around the beam line, leading to muon curvature along the z-axis.

1019    The precision tracking system covers the region  $|\eta| < 2.7$  via three layers of monitored  
 1020    drift tubes, and is complemented by cathode-strip chambers in the forward region, where the  
 1021    background is highest. The muon trigger system covers the range  $|\eta| < 2.4$  with resistive-plate  
 1022    chambers in the barrel, and thin-gap chambers in the endcap regions.

1023    3.2.5 Triggering

1024    During a typical run of the LHC, there are roughly 1 billion collisions in ATLAS per second  
 1025    (1 GHz), corresponding to a 40 MHz bunch crossing rate. *TODO: cite: <https://cds.cern.ch/record/1457044/file>*  
 1026    Saving the information from all of them is not only unnecessary, but infeasible. The ATLAS  
 1027    trigger system provides a sophisticated set of selections to filter the collision data and only  
 1028    keep those collision events useful for downstream analysis.

1029    These events are selected by the first-level trigger system, which is implemented in custom  
 1030    hardware, and accepts events at a rate below 100 kHz. Selections are then made by algorithms

1031 implemented in software in the high-level trigger [31], reducing this further, and, in the end,  
1032 events are recorded to disk at much more manageable rate of about 1 kHz.

1033 An extensive set of ATLAS software [32] is open source, including the software used for  
1034 real and simulated data reconstruction and analysis and that used in the trigger and data  
1035 acquisition systems of the experiment.

1036 *3.2.6 Particle Showers and the Calorimeter*

1037 The design of the ATLAS detector is directly tied to the physics it is trying to detect. Of these,  
1038 possibly the most non-trivial distinction is in the calorimeter design. It is therefore useful to  
1039 discuss in more detail the various properties of electromagnetic and hadronic interactions  
1040 with material, and how these correspond to the particle showers measured by the detector  
1041 described above.

1042 Electromagnetic showers in ATLAS predominantly occur via bremsstrahlung, or “braking  
1043 radiation”, and electron-positron pair production. This proceeds roughly as follows: an electron  
1044 entering a material is deflected by the electromagnetic field of a heavy nucleus. This results in  
1045 the radiation of a photon. That photon produces an electron-positron pair, and the process  
1046 repeats, resulting in a shower structure. At each step, characterized by *radiation length*,  $X_0$ ,  
1047 the number of particles approximately doubles and the average particle energy decreases by  
1048 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

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

1049 where  $Z$  is the nuclear charge. From a similar analysis of rates, we may see that the  
1050 bremsstrahlung rate is inversely proportional to the square of the mass of the particle. This

1051 explains why muons do not shower in a similar way, as the rate of bremsstrahlung is suppressed  
1052 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  $1.3 \times 10^3$  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

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

1053 such that for a 100 GeV shower in lead,  $x \sim 13$ . The radiation length for lead is around  
1054 0.56 cm, such that an electromagnetic shower could be expected to be captured within 10 cm  
1055 of lead.

1056 Electromagnetic showers are therefore characterized by depositing much of their energy  
1057 within a small region of space. As we show below (Chapter 4) though electromagnetic  
1058 showering is not deterministic, the large number of particles and the restricted set of processes  
1059 involved means that the shower development as a whole is very similar between individual  
1060 electromagnetic showers of the same energy.

1061 For completeness, note as well that pair production dominates for photons of energy greater  
1062 than around 10 MeV, whereas for lower energies (below around 1 MeV), the photoelectric  
1063 effect, namely atomic photon absorption and electron emission, dominates.

1064 Hadronic showers are distinguished by the fact that they interact strongly with atomic  
1065 nuclei. They are correspondingly more complex because (1) they involve a wider variety  
1066 of processes than electromagnetic showers, and (2) these processes have a wide variety of  
1067 associated length scales. Because these are heavier than electrons (e.g. protons and charged  
1068 pions) bremsstrahlung is suppressed, but ionization interactions with the electrons will cause  
1069 these particles to lose energy as they pass through the material. Hadronic showering occurs  
1070 on interaction with atomic nuclei. This may lead to production of, e.g. both charged ( $\pi^\pm$ )  
1071 and neutral ( $\pi^0$ ) pions. The  $\pi^0$  lifetime is much much shorter than that of the charged pions  
1072 (around a factor of  $10^8$ ), and immediately decays to two photons, starting an electromagnetic

1073 shower, as described above. The longer lived  $\pi^\pm$  travel further in the detector before  
1074 experiencing another strong interaction with more particles produced, also with varying  
1075 lifetimes and decay properties.

1076 It is therefore immediately apparent that hadronic showers are more complex than  
1077 electromagnetic ones (electromagnetic showers can be a subset of the hadronic!), and therefore  
1078 much more variable from shower to shower. The length scales involved are also significantly  
1079 larger due to the reliance on nuclear interactions, characterized by length  $\lambda_I$ , which is around  
1080 17 cm for iron (used in the ATLAS hadronic calorimeter). This motivates the calorimeter  
1081 design, and results in the properties demonstrated in Figure 3.2.

1082

## Chapter 4

1083

# SIMULATION

1084 Simulated physics samples are a core piece of the physics output of the Large Hadron  
 1085 Collider, providing a map from a physics theory into what is observed in our detector. This  
 1086 is crucial for searches for new physics, where simulation is necessary to describe what a given  
 1087 signal model looks like, but also extremely valuable for describing the physics of the Standard  
 1088 Model, providing detailed predictions of background processes for use in everything from  
 1089 designing simple cuts to training multivariate discriminators. Broadly, simulation can be split  
 1090 into two stages: *event generation*, in which physics theory is used to generate a description of  
 1091 particles present after a proton-proton collision, and *detector simulation*, which passes this  
 1092 particle description through a simulation of the detector material, providing a view of the  
 1093 physics event as it would be seen in ATLAS data. Such simulation is often called Monte Carlo  
 1094 in reference to the underlying mathematical framework, which relies on random sampling.

1095 **4.1 Event Generation**

1096 A variety of tools are used to simulate various aspects of event generation. MADGRAPH [33]  
 1097 is commonly used for the generation of the “hard scatter” event, i.e., two protons collide  
 1098 and some desired physics process happens. In practice, this is not quite as simple as two  
 1099 quarks or gluons interacting. Protons are composed of three “valence” quarks with various  
 1100 momenta interacting with each other via exchange of gluons, but also a sea of virtual gluons  
 1101 which may decay into other quarks. A hard scatter event is therefore characterized by  
 1102 the corresponding particle level diagrams, but additionally by a set of *parton distribution*  
 1103 *functions* (PDFs), which describe the probability to find constituent quarks or gluons at  
 1104 carrying various momenta at a given energy scale (often written  $Q^2$ ). Such PDFs are measured

1105 experimentally *TODO: cite* and the selection of a “PDF set” and a given physics process  
 1106 characterizes the hard scatter. Depending on the model being considered and the particular  
 1107 theoretical constraints, processes are often simulated at either leading (LO) or next to leading  
 1108 order (NLO), corresponding to the order of the perturbative expansion (i.e. tree level or 1  
 1109 loop diagrams). Various additional tools are developed for such NLO calculations, including  
 1110 POWHEG Box v2 [34–36], which is used for this thesis.

1111 The hard scatter is not the only component of a given collider event, however. Incoming  
 1112 and outgoing particles are themselves very energetic and may radiate particles along their  
 1113 trajectory. In particular, gluons, which have a self-interaction term as described in Chapter 1,  
 1114 may be radiated, which subsequently themselves radiate gluons or decay to quarks which can  
 1115 also radiate gluons, in a whole mess of QCD that both contributes to the particle content  
 1116 of a collider event and is not directly described by the hard scatter. This cascade, called  
 1117 a *parton shower*, has a dedicated set of simulation tools, commonly HERWIG 7 [37][38] and  
 1118 PYTHIA 8 [39], which interface with tools such as MADGRAPH for simulation.

1119 Due to color confinement (Chapter 1), quarks and gluons cannot be observed free particles,  
 1120 but rather undergo a process called hadronization, in which they are grouped into colorless  
 1121 hadrons (e.g. *mesons*, consisting of one quark and one antiquark). In simulation, this is also  
 1122 handled with HERWIG 7 and PYTHIA 8.

1123 The physics of  $b$ -quarks is quite important for a variety of searches for new physics and  
 1124 measurements of the Standard Model, including this thesis work *TODO: ref flavor tagging*  
 1125 *sec?*. Correspondingly, the decay of “heavy flavor” particles (e.g.  $B$  and  $D$  mesons, containing  
 1126  $b$  and  $c$  quarks respectively) has been very well studied, and a dedicated simulation tool,  
 1127 EVTGEN [40], is used for such processes.

1128 *TODO: add nice parton shower image*

## 1129 4.2 Detector Simulation

1130 Event generation provides a full description of the particle content of a given collider event.  
 1131 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 [41] 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 [42], 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.

1158    **4.3 Correlated Fluctuations in FastCaloSim**

1159    A variety of developments have been made to FastCaloSim, improving on the version used for  
 1160    AtlFast-II. This new fast calorimeter simulation [43] is largely based on two components: one  
 1161    which describes the *total energy* deposited in each calorimeter layer as a shower moves from  
 1162    the interaction point outward, and one which describes the *shape*, i.e., the pattern of energy  
 1163    deposits, of a shower in each respective calorimeter layer. Both methods are parametrizations  
 1164    of the full simulation, and therefore are considered to be performing well if they are able  
 1165    to reproduce corresponding full simulation distributions. Of course, directly sampling from  
 1166    a library of showers would identically reproduce such distributions – however a statistical  
 1167    sampling of various shower *properties* provides much more generality in the simulation.

1168    For the simulation of total energy in each given layer, the primary challenge is that such en-  
 1169    ergy deposits are highly correlated. The new FastCaloSim thus relies on a technique called Prin-  
 1170    cipal Component Analysis (PCA) *TODO: cite <https://root.cern.ch/doc/master/classTPrincipal.html>*  
 1171    to de-correlate the layers, aiding parametrisation.

1172    The PCA chain transforms  $N$  energy inputs into  $N$  Gaussians and projects these Gaussians  
 1173    onto the eigenvectors of the corresponding covariance matrix. This results in  $N$  de-correlated  
 1174    components, as the eigenvectors are orthogonal. The component of the PCA decomposition  
 1175    with the largest corresponding eigenvalue is then used to define bins, in which showers  
 1176    demonstrate similar patterns of energy deposition across the calorimeter layers. To further  
 1177    de-correlate the inputs, the PCA chain is repeated on the showers within each such bin. This  
 1178    full process is reversed for the particle simulation. A full description of the method can be  
 1179    found in [43].

1180    Modeling of the lateral shower shape makes use of 2D histograms filled with GEANT4  
 1181    hit energies in each layer and PCA bin. Binned in polar  $\alpha - R$  coordinates in a local plane  
 1182    tangential to the surface of the calorimeter system, these histograms represent the spatial  
 1183    distribution of energy deposits for a given particle shower. Such histograms are constructed  
 1184    for a number of Geant4 events, and the histograms for each event are normalized to total

1185 energy deposited in the given layer. The average of these histograms is then taken (what is  
 1186 called here the “average shape”).

1187 In simulation, these average shape histograms are used as probability distributions, from  
 1188 which a finite number of equal energy hits are drawn. This finite drawing of hits induces  
 1189 a statistical fluctuation about the average shape which is tuned to match the expected  
 1190 calorimeter sampling uncertainty.

1191 As an example, the intrinsic resolution of the ATLAS Liquid Argon calorimeter has a  
 1192 sampling term of  $\sigma_{\text{samp}} \approx 10\%/\sqrt{E}$  [44]. The number of hits to be drawn for each layer,  $N_{\text{hits}}^{\text{layer}}$ ,  
 1193 is thus taken from a Poisson distribution with mean  $1/\sigma_{\text{samp}}^2$ , where the energy assigned to  
 1194 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  
 1195 each bin in the average shape.

1196 Figure 4.1 shows a comparison of energy and weta2 [45], defined as the energy weighted  
 1197 lateral width of a shower in the second electromagnetic calorimeter layer, for 16 GeV photons  
 1198 simulated with the new FastCaloSim and with full GEANT4 simulation. The agreement is  
 1199 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  $\eta$  points.

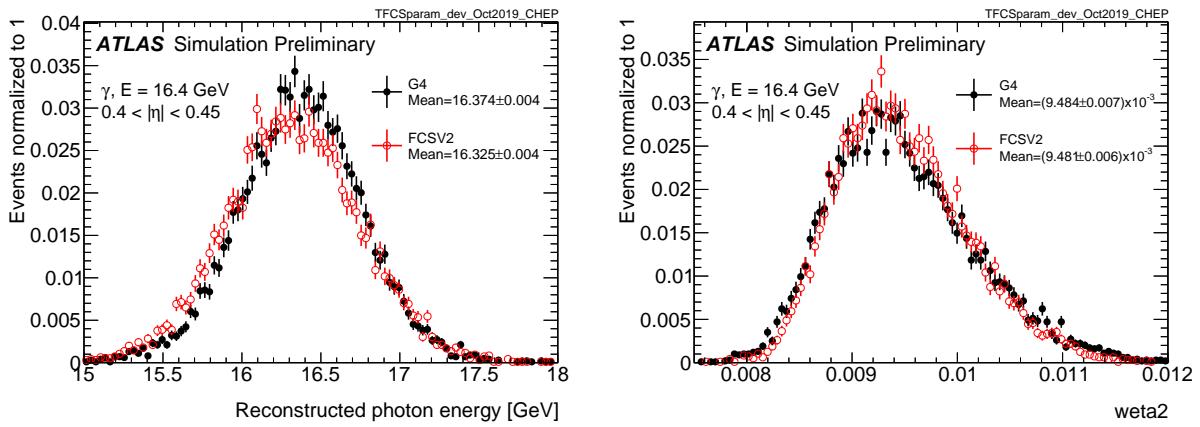


Figure 4.1: Energy and lateral shower width variable, weta2, for 16 GeV photons with full simulation (G4) and FastCaloSimV2 (FCSV2) [43].

1201 *4.3.1 Fluctuation Modeling*

1202 Figure 4.2 shows the ratio of calorimeter cell energies for single GEANT4 photon and pion  
 1203 events to the corresponding cell energies in their respective average shapes. While the photon  
 1204 event is quite close to the corresponding average, the pion event shows a deviation from the  
 1205 average which is much larger and has a non-trivial structure, reflecting the different natures  
 1206 of electromagnetic and hadronic showering.

1207 While the shape parametrization described above is thus sufficient for describing electro-  
 1208 magnetic showers, we will demonstrate below that it is not sufficient for describing hadronic  
 1209 showers (Figures 4.5 and 4.6). We therefore present and validate methods to improve this  
 1210 hadronic shower modeling.

1211 Two methods for modeling deviations from the average shape have been studied: (1)  
 1212 a neural network based approach using a Variational Autoencoder (VAE) [46] and (2) a  
 1213 map through cumulative distributions to an  $n$ -dimensional Gaussian. With both methods,  
 1214 the shape simulation then proceeds as described in Section ??, with the drawing of hits  
 1215 according to the average shape. However, these hits no longer have equal energy, but have  
 1216 weights applied to increase or decrease their energy depending on their spatial position.  
 1217 This application of weights is designed to mimic a realistic shower structure and to encode  
 1218 correlations between energy deposits.

1219 Both methods are trained on ratios of energy in binned units called voxels. This voxelization  
 1220 is performed in the same polar  $\alpha - R$  coordinates as the average shape, with a 5 mm core in  
 1221  $R$  and 20 mm binning thereafter. There are a total of 8  $\alpha$  bins from 0 to  $2\pi$  and 8 additional  
 1222  $R$  bins from 5 mm to 165 mm. The 5 mm core is filled with the average value of core voxels  
 1223 across the 8  $\alpha$  bins when creating the parametrisation. However, during simulation, each of  
 1224 these 8 core bins is treated independently. The outputs of both methods mimic these energy  
 1225 ratios and are used in the shape simulation as the weights described above. In contrast to  
 1226 an approach based on, e.g., calorimeter cells, using voxels allows for flexibility in tuning the  
 1227 binning used in creating the parametrisation. Further, due to their relatively large size, using

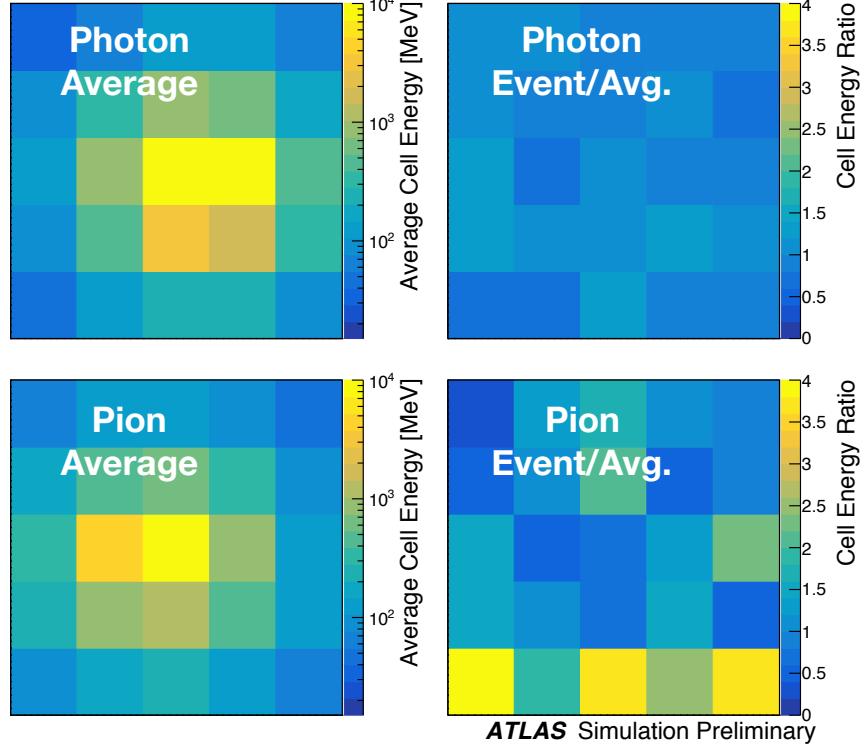


Figure 4.2: Example of photon and pion average shapes in  $5 \times 5$  calorimeter cells. The left column shows the average shape over a sample of 10000 events, while the right column shows the energy ratio, in each cell, of single GEANT4 events with respect to this average. The photon ratios are all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

1228 calorimeter cells is subject to “edge effects”, where the splitting of energy between cells has a  
 1229 non-trivial effect on the observed energy ratio. The binning used here is of the order of half  
 1230 of a cell size, mitigating this effect.

1231 The Gaussian method operates by using cumulative distributions to map GEANT4 energy  
 1232 ratios to a multidimensional Gaussian distribution. New events are generated by randomly  
 1233 sampling from this Gaussian distribution.

1234 For the VAE method, a system of two linked neural networks is trained to generate events.

1235 The first “encoder” neural network maps input GEANT4 energy ratios to a lower dimensional  
 1236 latent space. A second “decoder” neural network then samples from that latent space and  
 1237 tries to reproduce the inputs. In simulation, events are generated by taking random samples  
 1238 from the latent space and passing them through the trained decoder.

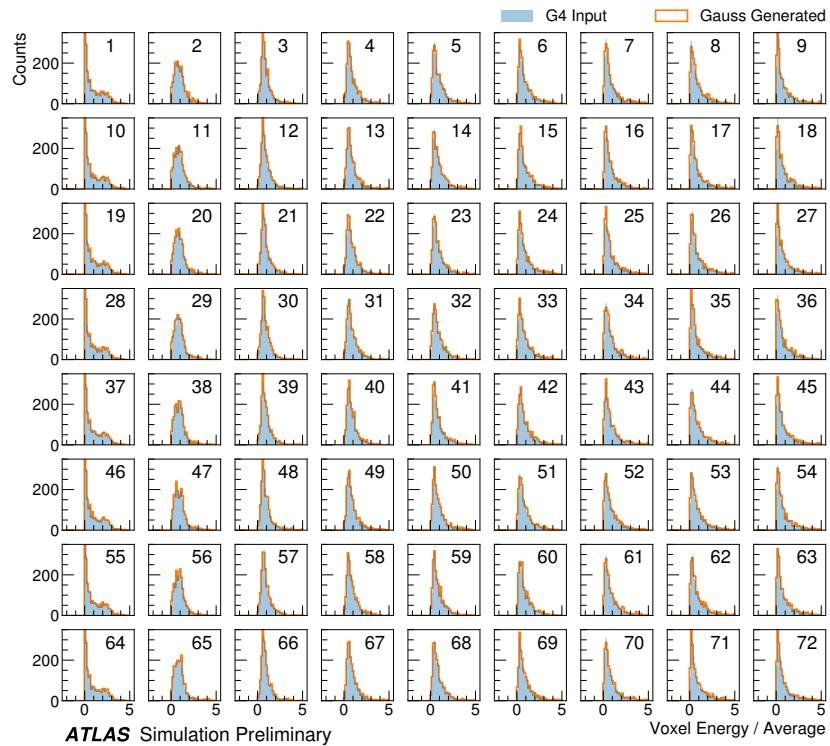


Figure 4.3: Distribution of the ratio of voxel energy in single events to the corresponding voxel energy in the average shape, with GEANT4 events in blue and Gaussian model events in orange, for 65 GeV central pions in EMB2. Moving top to bottom corresponds to increasing  $\alpha$ , left to right corresponds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

1239 Figure 4.3 shows the distributions of input GEANT4 and Gaussian method generated  
 1240 energy ratios in the grid of voxels. Figure 4.4 shows the correlation coefficient between the  
 1241 center voxel from  $\alpha = 0$  to  $2\pi/8$  for input GEANT4 and the Gaussian and VAE fluctuation

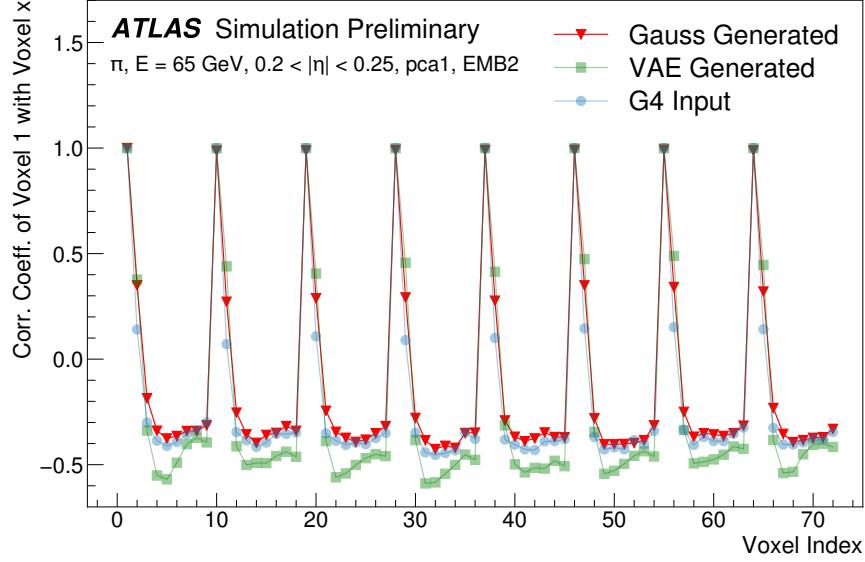


Figure 4.4: Correlation coefficient of ratios of voxel energy in single events to the corresponding voxel energy in the average shape, examined between the core bin from  $\alpha = 0$  to  $2\pi/8$  and each of the other voxels. The periodic structure represents the binning in  $\alpha$ , and the increasing 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.

1242 methods. Agreement is good throughout.

1243 Validation of the Gaussian and VAE fluctuation methods was performed within FastCaloSimV2.

1244 Figure 4.5 shows the energy ratio of cells for a given simulation to the corresponding cells in  
 1245 the average shape as a function of the distance from the shower center. The mean for all  
 1246 simulation methods is expected to be around 1, with deviation from the average (the RMS  
 1247 fluctuation) shown by the error bars. The Gaussian method RMS (red) and VAE method  
 1248 RMS (green) both match the GEANT4 RMS (yellow) better than the case without correlated  
 1249 fluctuations (blue) for a variety of energies,  $\eta$  points, and layers, often reproducing 80 – 100 %

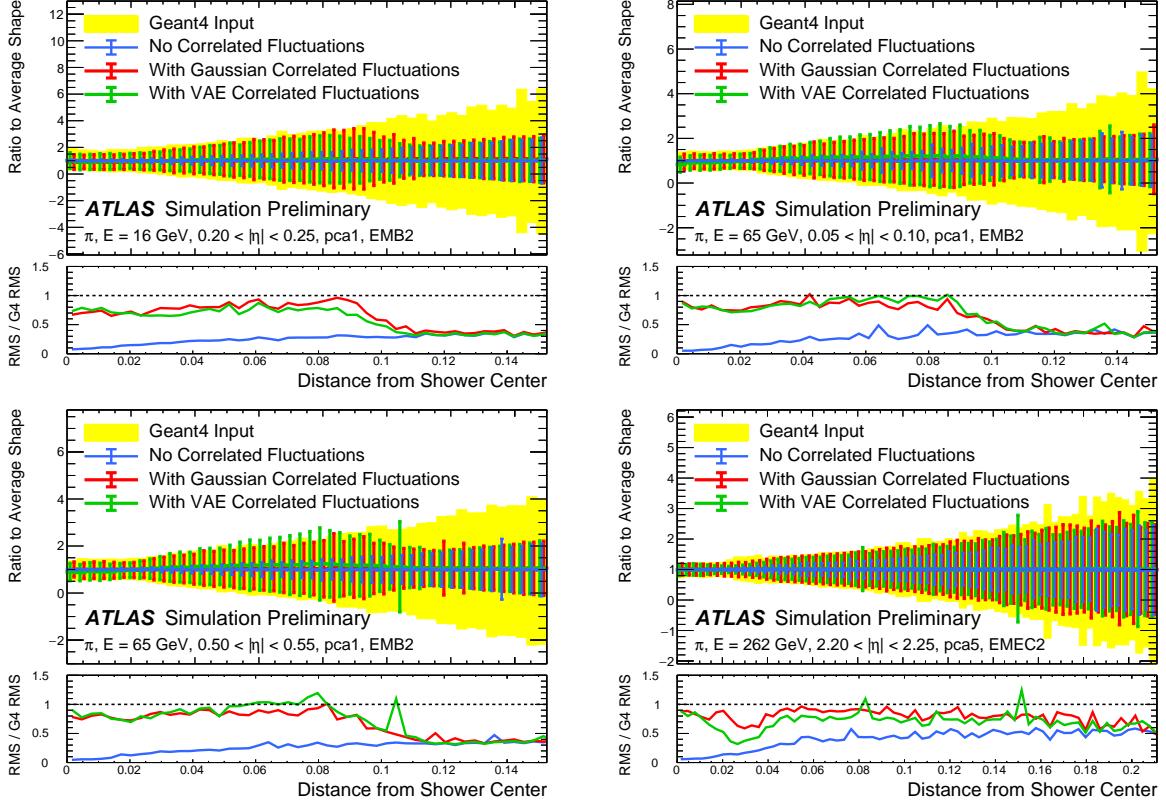


Figure 4.5: Comparison of the RMS fluctuations about the average shape with the Gaussian fluctuation model (red), the VAE fluctuation model (green), and without correlated fluctuations (blue) for a range of pion energies,  $\eta$  points, and layers.

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,

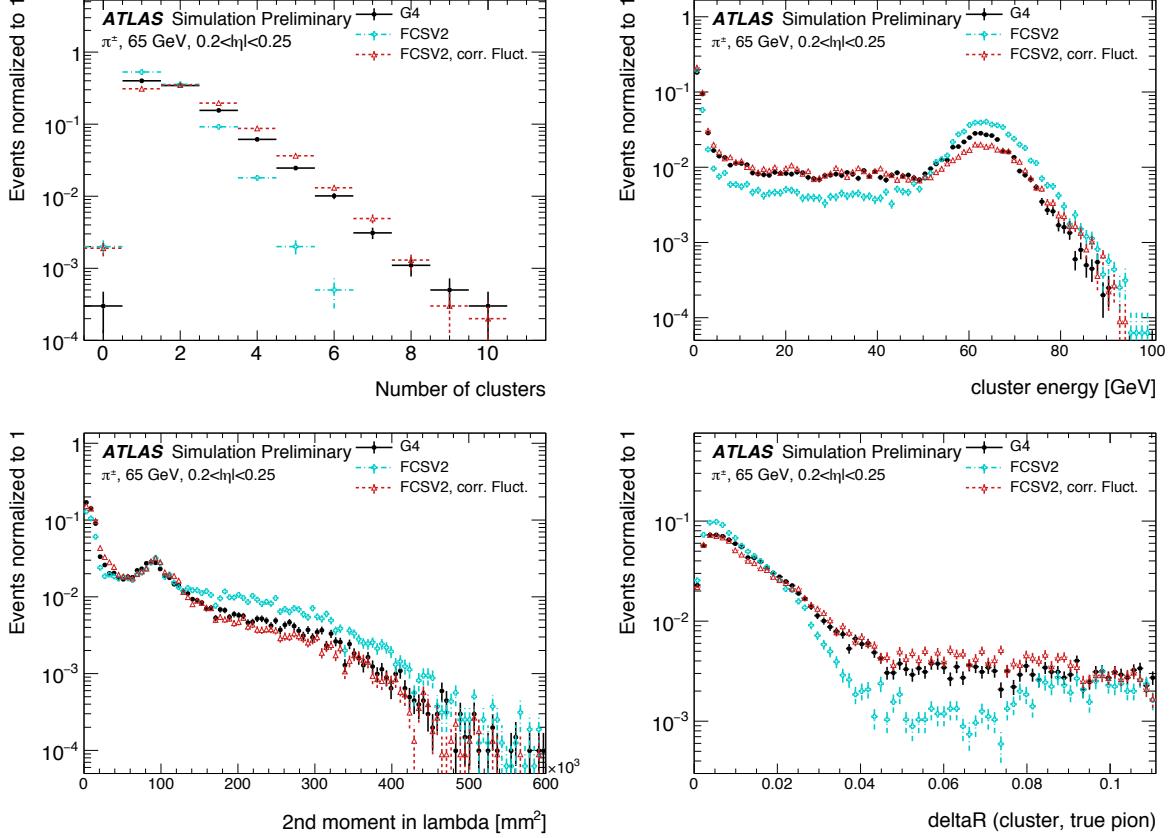


Figure 4.6: Comparison of the Gaussian fluctuation model to the default FCSV2 version and to G4 simulation, using pions of 65 GeV energy and  $0.2 < |\eta| < 0.25$ . With the correlated fluctuations, several shape variables demonstrate improved modeling.

1258 calculated over 100  $t\bar{t}$  events, is 228.9 s. Using FastCaloSim for the calorimeter simulation  
 1259 reduces this to 26.6 s, of which FastCaloSim itself is only 0.015 s. Good physics modeling is  
 1260 achieved, the correlated fluctuations method shows good proof of concept improvement for  
 1261 the modeling of hadronic showers.

1262

## Chapter 5

1263

# RECONSTRUCTION

1264 Chapter 3 discusses how a proton-proton collision may be captured by a physical detector  
 1265 and turned into data that may be stored and analyzed. Chapter 4 discusses the simulation  
 1266 of this same process. At this most basic level, however, the ATLAS detector is only a  
 1267 machine for turning particles into a set of electrical signals, albeit in a very sophisticated,  
 1268 physics motivated way. This chapter discusses the step of turning these electrical signals into  
 1269 objects which may be identified with the underlying physics processes, and therefore used to  
 1270 make statements about what occurred within a given collision event. This process is termed  
 1271 *reconstruction*, and we will focus particularly on jets and flavor tagging, as the most relevant  
 1272 pieces for this thesis work.

1273 **5.1 Jets**

1274 As discussed in Chapters 3 and 4, the production of particles with color charge from a  
 1275 proton-proton interaction is complicated both by parton showering and by confinement: a  
 1276 quark produced from a hard scatter is not seen as a quark, but rather, as a spray of particles  
 1277 with a variety of hadrons in the final state, which subsequently shower upon interaction with  
 1278 the calorimeter in a complicated way.

1279 For hard scatter electrons, photons, or muons on the other hand, the picture is much  
 1280 clearer: there is no parton showering, and each has a distinct signature in the detector:  
 1281 photons have no tracks and a very localized calorimeter shower, electrons are associated  
 1282 with tracks and are similarly localized in the calorimeter, and muons are associated with  
 1283 tracks, pass through the calorimeter due to their large mass, and leave signatures in the muon  
 1284 spectrometer.

1285 Jets are a tool to deal with the messiness of quarks and gluons. The basic concept is to  
 1286 group the multitude of particles produced by a quark or gluon decay into a single object. Such  
 1287 an object then has associated properties, including a four-vector, which may be identified  
 1288 with the corresponding initial state particle. In practice a variety of information from the  
 1289 ATLAS detector is used for such a reconstruction. The analysis considered in this thesis uses  
 1290 particle flow jets [47], which combines information from both the tracker and the calorimeter,  
 1291 where the combined objects may be identified with underlying particles. However, jets built  
 1292 from clusters of calorimeter cells [48] as well as from charged particle tracks [49] have also  
 1293 been used very effectively.

1294 A variety of algorithms are used to associate detector level objects to a given jet. The  
 1295 most commonly used in ATLAS is the anti- $k_T$  algorithm [50], which is a successor to the  
 1296  $k_T$  algorithm, among others [51], and develops as follows. Both algorithms are sequential  
 1297 recombination algorithms, which begin with the smallest distance,  $d_{ij}$  between considered  
 1298 objects (e.g. particles or intermediate groupings of particles). If  $d_{ij}$  is less than a parameter  
 1299  $d_{iB}$  (B for “beam”) object  $i$  is combined with object  $j$ , the distance  $d_{ij}$  is recomputed, and  
 1300 the process repeats. This proceeds until  $d_{ij} \geq d_{iB}$ , at which point the jet is “complete” and  
 1301 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

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

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

1302 where  $p_{Ti}$  is the transverse momentum of object  $i$ ,  $\Delta R_{ij}$  is the angular distance between  
 1303 objects  $i$  and  $j$ ,  $R$  is a radius parameter, and  $p$  controls the tradeoff between the  $p_T$  and  
 1304 angular distance terms. For the  $k_T$  algorithm  $p = 1$ ; for the anti- $k_T$  algorithm,  $p = -1$ . This  
 1305 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

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

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

1306 Therefore, a single high  $p_T$  particle ( $p_{T1}$ ) surrounded by a bunch of low  $p_T$  particles results in  
 1307 a perfectly conical jet. This shape may change with the presence of other high momentum  
 1308 particles, but the key feature of the dynamics is that the jet shape is determined by high  $p_T$   
 1309 objects due to the  $\frac{1}{p_T}$  nature of this definition. In contrast, the  $k_T$  algorithm results in jets  
 1310 influenced by low momentum particles, which results in a less regular shape. This property,  
 1311 of regular jet shapes determined by high momentum objects, as well as demonstrated good  
 1312 practical performance, makes the anti- $k_T$  algorithm the favored jet algorithm in ATLAS.

1313 Because jets are composed of multiple objects, a useful property of jets is jet *substructure*,  
 1314 that is, acknowledging that jets are composite objects, analyzing the structure of a given  
 1315 jet to infer physics information. This leads to the use of *subjets*; that is, after running jet  
 1316 clustering, often to create a “large-R”,  $R = 1.0$  anti- $k_T$  jet, a smaller radius jet clustering  
 1317 algorithm is run within the jet. Subjets are often chosen using the  $k_T$  algorithm, which again  
 1318 is sensitive to lower momentum particles, with  $R = 0.2$  or  $0.3$ . For the boosted version of this  
 1319 thesis analysis, such a strategy is used, in which the subjets are *variable radius* and depend  
 1320 on the momentum of the (proto)jet in question. Beyond this thesis work, substructure is  
 1321 used in a large variety of analyses, with a set of associated variables and tools developed for  
 1322 exploiting this structure *TODO: Cite some?*.

## 1323 5.2 Flavor Tagging

1324 For this this thesis, the physics process being considered is  $HH \rightarrow b\bar{b}b\bar{b}$ . From the previous  
 1325 section, we know that the standard practice is to identify these  $b$  quarks (or, rather, the  
 1326 resulting  $B$  hadrons, due to confinement) with jets – in our case, these  $b$ -*jets* are  $R=0.4$   
 1327 anti- $k_T$  particle flow jets (see Chapter 7). However, not all jets produced at the LHC are  
 1328 from  $B$  hadrons; rather, there are a variety of different types of jets corresponding to different

1329 flavors of quarks. These are often classified as light jets (from  $u$ ,  $d$ , or  $s$  quarks, or gluons)  
 1330 or as other *heavy flavor* jets, e.g.  $c$ -jets, involving  $c$  quarks. Distinguishing between these  
 1331 different categories is a very active area of work in ATLAS, termed *flavor tagging*, with much  
 1332 focus on *b-tagging* in particular, that is, the identification of jets from  $B$  hadron decays. We  
 1333 here briefly describe the techniques used for flavor tagging in ATLAS.

1334 What distinguishes a  $b$ -jet from any other jet? This question is fundamental to the design  
 1335 of the various  $b$ -tagging algorithms, and has two major answers: (1)  $B$  hadrons have long  
 1336 lifetimes, and (2)  $B$  hadrons have large masses. It is most illustrative to compare the  $B$  hadron  
 1337 properties to a common light meson, e.g.  $\pi^0$ , the neutral pion, with quark content  $\frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$ .  
 1338  $B$  hadrons have lifetimes around 1.5 ps, corresponding to a decay length  $c\tau \approx 0.45$  mm. In  
 1339 contrast,  $\pi^0$  has a lifetime of  $8.4 \times 10^{-5}$  ps, which is around 20,000 times shorter! Theoretically,  
 1340 this comes from CKM suppression of the  $b$  to  $c$  transition *TODO: check*, which dominates  
 1341 the  $B$  decay modes. Experimentally, this difference pops up as shown in Figure 5.1 – light  
 1342 flavor initiated jets decay almost immediately at the proton-proton interaction point, whereas  
 1343  $b$ -jets are distinguished by a displaced secondary vertex, corresponding to the 5 mm decay  
 1344 length calculated above. This displaced vertex falls short of the detector itself, but may be  
 1345 inferred from larger transverse (perpendicular to beam) and longitudinal (parallel to beam)  
 1346 impact parameters of the resulting tracks, termed  $d_0$  and  $z_0$  respectively.

1347 Coming to the mass,  $B$  mesons have masses of around 5.2 GeV, whereas the  $\pi^0$  mass  
 1348 is around 0.134 GeV, (around 40 times lighter). This higher mass gives access to a larger  
 1349 decay phase space, leading to a high multiplicity for  $b$ -jets (average of 5 charged particles per  
 1350 decay).

1351 One final distinguishing feature of  $B$  hadrons is their *fragmentation function*, a function  
 1352 describing the production of an observed final state. For  $B$  hadrons, this is particularly  
 1353 “hard”, with the  $B$  hadrons themselves contributing to an average of around 75 % of the  $b$ -jet  
 1354 energy. Thus, the identification of  $b$ -jets with  $B$  hadrons is, in some sense, descriptive.

1355 We have contrasted  $b$ -jets and light jets, demonstrating that there are several handles  
 1356 available for making this distinction.  $c$ -jets are slightly more similar to  $b$ -jets, but the same

1357 handles still apply –  $c$ -hadron lifetimes are between 0.5 and 1 ps, a factor of 2 smaller than  $B$   
1358 hadrons. Their mass is around 1.9 GeV, 2 to 3 times smaller than  $B$  hadrons, and  $c$ -hadrons  
1359 contribute to an average of around 55 % of  $c$ -jet energy. Therefore, while the gap is slightly  
1360 smaller, a distinction may still be made.

1361 The ATLAS flavor tagging framework [53] relies on developing a suite of “low-level”  
1362 taggers, which use a variety of information about tracks and vertices as inputs. The output  
1363 of these lower level taggers are then fed into a higher level tagger, which aggregates these  
1364 results into a high level discriminant. Each of these taggers is described below.

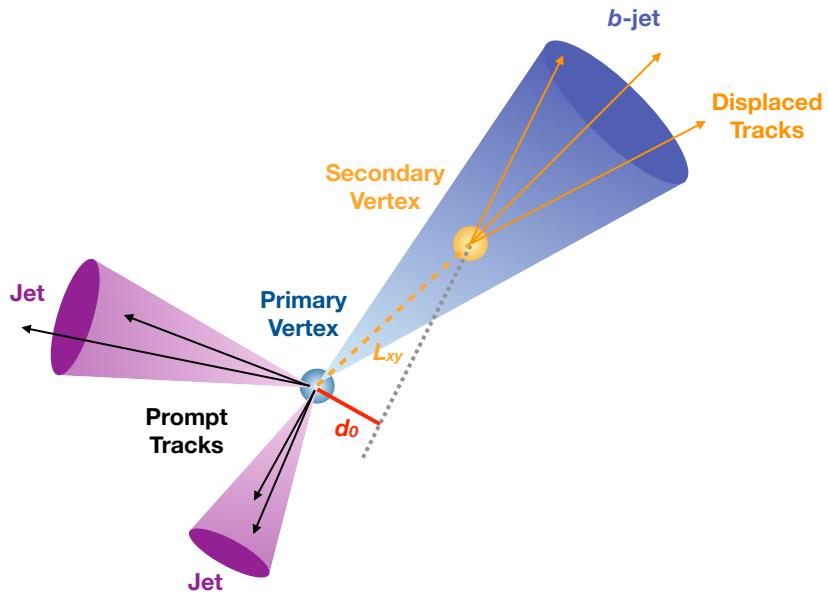


Figure 5.1: Illustration of an interaction producing two light jets and one  $b$ -jet in the transverse plane. While the light jets decay “promptly”, coinciding with the primary vertex of the proton-proton interaction, the longer lifetime of  $B$  hadrons leads to a secondary decay vertex, displaced 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. [52]

1365    5.2.1 IP2D/3D

1366    IP2D and IP3D are taggers based on the large track impact parameter (IP) nature of  $B$   
 1367    hadron decays. Both are based on histogram templates derived from Monte Carlo simulation,  
 1368    which are used as probability density functions to construct log-likelihood discriminants.  
 1369    IP2D incorporates just the transverse impact parameter information using 1D histogram  
 1370    templates, whereas IP3D uses both transverse and longitudinal impact parameters in a 2D  
 1371    template, which accounts for correlations. Importantly, these are *signed* impact parameters,  
 1372    with sign based on the angle between the impact parameter and the considered jet – positive  
 1373    impact parameters are consistent with a track extrapolation in front of the jet (angle between  
 1374    impact parameter line and jet  $< 90^\circ$ ), and therefore more consistent with tracks originating  
 1375    from a displaced decay.

1376    Rather than using the impact parameters directly, an impact parameter *significance*  
 1377    is used which incorporates an uncertainty on the impact parameter – tracks with a lower  
 1378    uncertainty but the same impact parameter will contribute more in the calculation. This  
 1379    signed significance is what is used to sample from the PDF templates, with the resulting  
 1380    discriminants the sum of probability ratios between given jet hypotheses over tracks associated  
 1381    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  
 1382     $p_{light}$  are the per-track probabilities. Similar discriminants are defined between  $b$ - and  $c$ -jets  
 1383    and  $c$  and light jets. *TODO: show distributions?*

1384    5.2.2 SV1

1385    SV1 is an algorithm which aims to find a secondary vertex (SV) in a given jet. Operating  
 1386    on all vertices associated with a considered jet, the algorithm discards tracks based on a  
 1387    variety of cleaning requirements. It then proceeds to first construct all two-track vertices,  
 1388    and then iterates over all the tracks involved in these two track vertices to try to fit a single  
 1389    secondary vertex, which would then be identified with the secondary vertex from the  $b$  or  $c$   
 1390    hadron decay. This fit proceeds by evaluating a  $\chi^2$  for the association of a track and vertex,

1391 removing the track with the largest  $\chi^2$ , and iterating until the  $\chi^2$  is acceptable and the vertex  
1392 has an invariant mass of less than 6 GeV (for consistency with  $b$  or  $c$  hadron decay).

1393 A variety of discriminating variables may then be constructed, including (1) invariant  
1394 mass of the secondary vertex, (2) number of tracks associated with the secondary vertex, (3)  
1395 number of two-track vertices, (4) energy fraction of the tracks associated to the secondary  
1396 vertex (relative to all of the tracks associated to the jet), and various metrics associated with  
1397 the secondary vertex position and decay length, including (5) transverse distance between the  
1398 primary and secondary vertex, (6) distance between the primary and secondary vertex (7)  
1399 distance between the primary and secondary vertex divided by its uncertainty, and (8)  $\Delta R$   
1400 between the jet axis and the direction of the secondary vertex relative to the primary vertex.

1401 While all eight of these variables are used as inputs to the higher level taggers, the number  
1402 of two-track vertices, the vertex mass, and the vertex energy fraction are additionally used with  
1403 3D histogram templates to evaluate flavor tagging performance by constructing log-likelihood  
1404 discriminants, similar to the procedure for IP2D/3D.

### 1405 5.2.3 *JetFitter*

1406 Rather than focusing on a particular aspect of the  $B$  hadron or  $D$  hadron decay topology  
1407 (e.g impact parameter or secondary vertex), the third low level tagger, JETFITTER [54],  
1408 tries to reconstruct the full decay chain, including all involved vertices. This is structured  
1409 around a Kalman filter formalism [55], and has the strong underlying assumption that all  
1410 tracks which stem from  $B$  and  $D$  hadron decay must intersect a common flight path. This  
1411 assumption provides significant constraints, allowing for the reconstruction of vertices from  
1412 even a single track, reducing the number of degrees of freedom in the fit, and allowing the  
1413 use of “downstream” information, e.g., compatibility of tracks with a  $B \rightarrow D$ -like decay.  
1414 The constructed topology, including primary vertex location and  $B$ -hadron flight path, is  
1415 iteratively updated over tracks associated to a given jet, and a variety of discriminating  
1416 variables related to the resulting topology and reconstructed decay are used as inputs to the  
1417 high level taggers.

1418 5.2.4 *RNNIP*

1419 The IP2D and IP3D algorithms rely on per-track probabilities, and the final discriminating  
1420 variables (and inputs to the higher level taggers) are sums (products) over these independently  
1421 considered quantities. In practice, however, the tracks are not independent – this is merely a  
1422 simplifying assumption to allow for the use of a binned likelihood, as treatment of all of the  
1423 interdependencies in such a framework quickly becomes intractable. To address this issue, a  
1424 recurrent neural network-based algorithm, RNNIP [56], is used, which takes as input a variety  
1425 of per-track variables, including the signed impact parameter significances (as in IP3D) as  
1426 well as track momentum fraction relative to the jet and  $\Delta R$  between the track and the jet.  
1427 RNNs are sequence-based, and vectors of input variables corresponding to tracks for a given  
1428 jet are ordered by magnitude of transverse impact parameter significance and then passed  
1429 to the network, which outputs class probabilities corresponding to b-jet, c-jet, light-jet, and  
1430  $\tau$ -jet hypotheses. Such a procedure allows the network to learn interdependencies between  
1431 tracks, improving performance.

1432 5.2.5 *MV2 and DL1*

1433 Outputs from the above taggers are combined into high level taggers to aggregate all of the  
1434 information and improve discriminating power relative to the respective individual taggers as,  
1435 as shown in Figure 5.2. These high level taggers are primarily in two forms: MV2, which  
1436 uses a Boosted Decision Tree (BDT) for this aggregation, and DL1, which uses a deep neural  
1437 network. For the baseline versions of these taggers, only inputs from IP2D, IP3D, SV1, and  
1438 JetFitter are used. The tagger used for this thesis analysis, DL1r, additionally incorporates  
1439 RNNIP, demonstrating improved performance over the baseline DL1, as shown in Figure 5.3.  
1440 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  $p_{\text{light}}$  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.

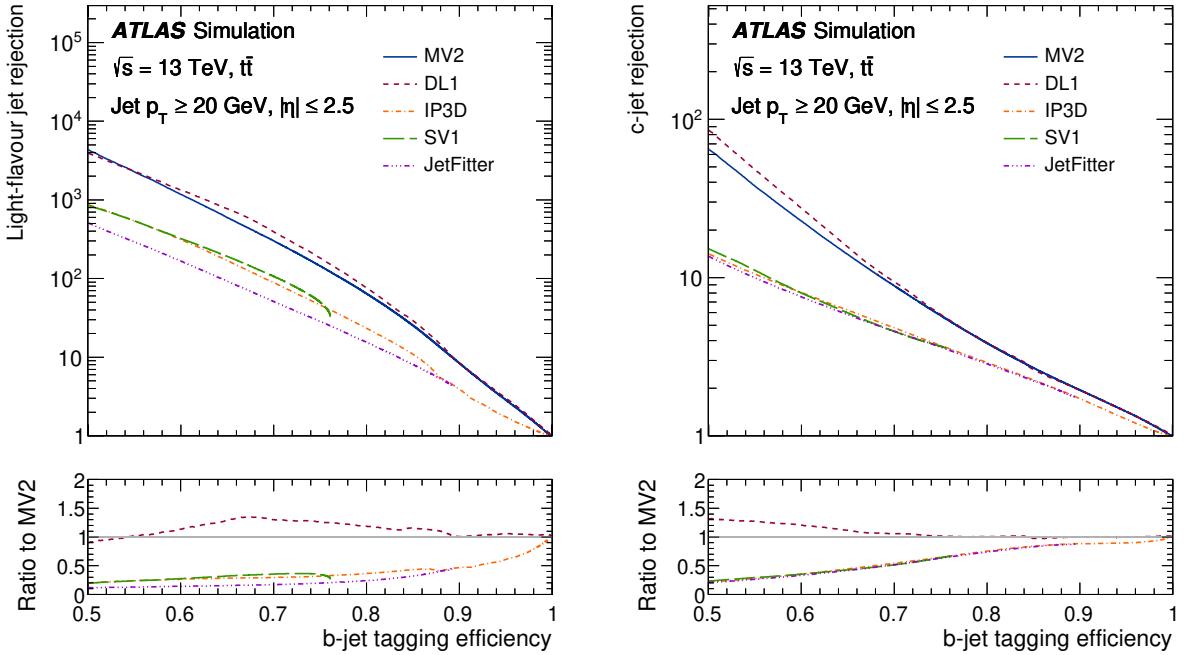


Figure 5.2: Performance of the various low and high level flavor tagging algorithms in  $t\bar{t}$  simulation, demonstrating the tradeoff between  $b$ -jet efficiency and light and  $c$ -jet rejection. The high level taggers demonstrate significantly better performance than any of the individual low 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.

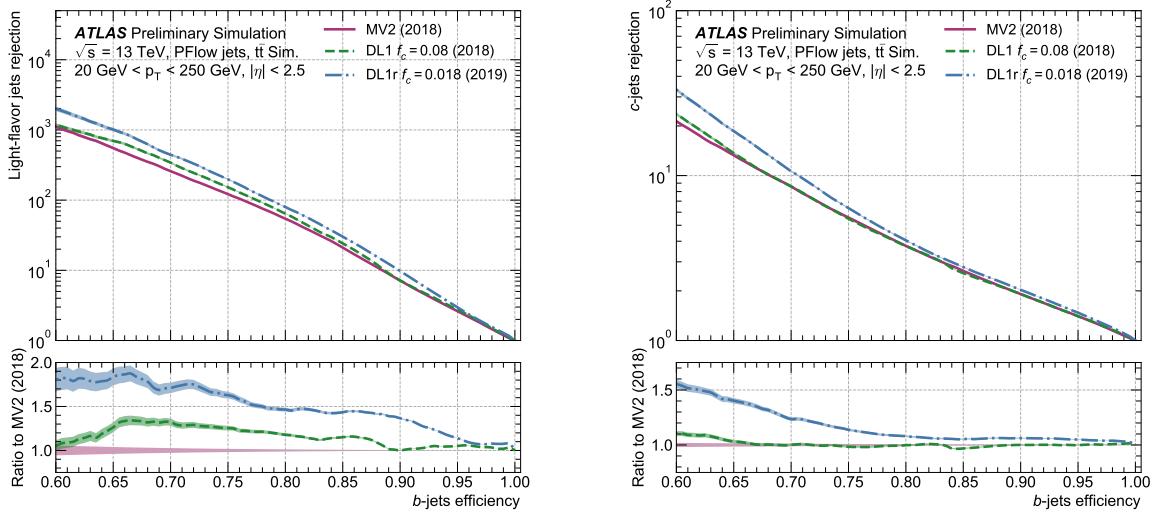


Figure 5.3: Performance of the MV2, DL1, and DL1r algorithms in  $t\bar{t}$  simulation, demonstrating the tradeoff between  $b$ -jet efficiency and light and  $c$ -jet rejection.  $f_c$  controls the importance of  $c$ -jet rejection in the discriminating variable, and values shown have been optimized separately for each DL1 configuration. DL1r demonstrates a significant improvement in both light and  $c$  jet rejection over MV2 and DL1. [57]

1455    5.2.6 *Some Practical Notes*

1456    The  $b$ -tagging metrics presented in Figures 5.2 and 5.3 correspond to evaluating a tradeoff  
1457    between  $b$ -jet efficiency and light jet and  $c$ -jet rejection. In this case,  $b$ -jet efficiency is defined  
1458    such that, e.g. for a 77 % efficiency, 77 % of the real  $b$ -jets will be tagged as such. Somewhat  
1459    counterintuitively, this means that a lower  $b$ -jet efficiency corresponds to a more aggressive  
1460    (“tighter”) selection on the discriminating variable, while a higher  $b$ -jet efficiency corresponds  
1461    to a less aggressive (“looser”) cut (100 % efficiency means no cut). Light and  $c$  jet efficiencies  
1462    are defined similarly, with rejection defined as 1/ the corresponding efficiency.

1463        In ATLAS, the respective  $b$ -tagging efficiencies are used to define various  $b$ -tagging working  
1464    points. The working point used for the nominal  $b$ -jet identification in this thesis is 77 % with  
1465    DL1r. A loosened (less aggressive) selection at the 85 % working point is additionally used.

1466        See Chapter 7 for further details.

1467

## Chapter 6

1468

### THE ANATOMY OF AN LHC SEARCH

1469 In this thesis so far, we have set the theoretical foundation for the work carried out at the  
 1470 LHC. We have described how one may translate between this theoretical foundation and what  
 1471 we are actually able to observe with the ATLAS detector. We have further stepped through  
 1472 the process of simulating production of specific physics processes and their appearance in  
 1473 our detector, allowing us to describe how a hypothetical physics model would be seen in  
 1474 our experiment. The question then becomes: all of these pieces are on the table, what do  
 1475 we do with them? This chapter attempts to answer exactly that, setting up a roadmap for  
 1476 assembling these pieces into a statement about the universe.

1477 ***6.1 Object Selection and Identification***

1478 As described in Chapter 5, there is a complicated set of steps for going from electrical signals  
 1479 in a detector to physics objects.

1480 ***6.2 Defining a Signal Region***

1481 ***6.3 Background Estimation***

1482 ***6.4 Uncertainty Estimation***

1483 ***6.5 Hypothesis Testing***

1484

## Chapter 7

1485

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

1486

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

1493

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.

1505

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

1506

1507

1508 Beyond the Standard Model signal hypotheses, which may have more events at low  $m_{HH}$ ,  
 1509 the non-resonant nature of the production allows the  $b\bar{b}b\bar{b}$  channel to retain sensitivity while  
 1510 discarding much of the challenging low mass background. Such freedom allows for decisions  
 1511 which focus on improved background modeling for the middle to upper  $HH$  mass regime,  
 1512 resulting in improved modeling and smaller uncertainties than would be obtained with a  
 1513 more generic approach.

1514 Both searches are presented in the following, with emphasis on particular motivations for,  
 1515 and consequences of, the various design decisions involved for each respective set of signal  
 1516 hypotheses.

1517 The analyses improve upon previous work ?? in several notable ways. The resonant search  
 1518 leverages a Boosted Decision Tree (BDT) based pairing algorithm, offering improved  $HH$   
 1519 pairing efficiency over a broad mass spectrum. The non-resonant adopts a different approach,  
 1520 with a simplified algorithm based on the minimum angular distance ( $\Delta R$ ) between jets in  
 1521 a Higgs candidate. Such an approach very efficiently discards low mass background events,  
 1522 resulting in an easier to estimate background with reduced systematic uncertainties.

1523 A particular contribution of this thesis is the background estimation, which uses a novel,  
 1524 neural network based approach, offering improved modeling over previous methods, as well  
 1525 as the ability to model correlations between observables. While all aspects of the analysis of  
 1526 course contribute to the final result, the author of this thesis wishes to emphasize that the  
 1527 background estimate, with the corresponding uncertainties and all other associated decisions,  
 1528 really is the core of the  $HH \rightarrow b\bar{b}b\bar{b}$  analysis – the development of this procedure, and all  
 1529 associated decisions, is similarly the core of this thesis work.

1530 ATLAS has performed a variety of searches in complementary decay channels as well, no-  
 1531 tably in the  $b\bar{b} W^+ W^-$  [58],  $b\bar{b}\tau^+\tau^-$  [59],  $W^+ W^- W^+ W^-$  [60],  $b\bar{b}\gamma\gamma$  [61], and  $W^+ W^- \gamma\gamma$  [62]  
 1532 final states, which were combined along with  $b\bar{b}b\bar{b}$  in [20].

1533 CMS has also performed searches for resonant production of Higgs boson pairs in the  
 1534  $b\bar{b}b\bar{b}$  final state (among others) at  $\sqrt{s} = 8$  TeV [63] and  $\sqrt{s} = 13$  TeV [64]. CMS have also  
 1535 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

1536 in [65].

1537 This analysis also benefits from improvements to ATLAS jet reconstruction and calibration,  
1538 and flavour tagging [53]. In particular, this analysis benefits from the introduction of particle  
1539 flow jets [47]. These make use of tracking information to supplement calorimeter energy  
1540 deposits, improving the angular and transverse momentum resolution of jets by better  
1541 measuring these quantities for charged particles in those jets.

1542 The analysis also benefits from the new DL1r ATLAS flavour tagging algorithm. Whereas  
1543 the flavour tagging algorithm used in the previous analysis (MV2) used a boosted decision  
1544 tree (BDT) to combine the output of various low level algorithms, DL1r (and the baseline  
1545 DL1 algorithm) uses a deep neural network to do this combination. In addition to the low  
1546 level algorithms used as inputs to MV2, DL1 includes a variety of additional variables used  
1547 for  $c$ -tagging. DL1r further incorporates RNNIP, a recurrent neural network designed to  
1548 identify  $b$ -jets using the impact parameters, kinematics, and quality information of the tracks  
1549 in the jets, while also taking into account the correlations between the track features.

1550 The overall analysis sensitivity further benefits from a factor of  $\sim 4.6$  increase in integrated  
1551 luminosity.

## 1552 7.1 Data and Monte Carlo Simulation

1553 Both the resonant and non-resonant searches are performed on the full ATLAS Run 2 dataset,  
1554 consisting of  $\sqrt{s} = 13\text{ TeV}$  proton-proton collision data taken from 2016 to 2018 inclusive.  
1555 Data taken in 2015 is not used due to a lack of trigger jet matching information and  $b$ -jet  
1556 trigger scale factors. The integrated luminosity collected and usable in this analysis<sup>1</sup> was:

1557 •  $24.6\text{ fb}^{-1}$  in 2016

1558 •  $43.65\text{ fb}^{-1}$  in 2017

---

<sup>1</sup>approximately  $9\text{ fb}^{-1}$  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 [66]

- 1559        •  $57.7 \text{ fb}^{-1}$  in 2018

1560        This gives a total integrated luminosity of  $126 \text{ fb}^{-1}$ . This is lower than the  $139 \text{ fb}^{-1}$  ATLAS  
 1561        collected during Run 2 [67] due to the inefficiency described in footnote 1 as well as the  
 1562         $3.2 \text{ fb}^{-1}$  of 2015 data which is unused due to the trigger scale factor issue mentioned above.

1563        In this analysis, Monte Carlo samples are used purely for modelling signal processes. The  
 1564        background is strongly dominated by events produced by QCD multijet processes, which  
 1565        are difficult to correctly model in simulation. This necessitates the use of a data-driven  
 1566        background modelling technique, which is described in Section 7.6.

1567        The scalar resonance signal model is simulated at leading order in  $\alpha_s$  using MADGRAPH  
 1568        [33]. Hadronization and parton showering are done using HERWIG 7 [37][38] with EVTGEN [40],  
 1569        and the nominal PDF is NNPDF 2.3 LO. In practice this is implemented as a two Higgs  
 1570        doublet model where the new neutral scalar is produced through gluon fusion and required  
 1571        to decay to a pair of SM Higgs bosons. The heavy scalar is assigned a width much smaller  
 1572        than detector resolution, and the other 2HDM particles do not enter the calculation.

1573        Scalar samples are produced at resonance masses between 251 and 900 GeV and the  
 1574        detector simulation is done using AtlFast-II [42]. In addition the samples at 400 GeV and  
 1575        900 GeV are also fully simulated to verify that the use of AtlFast-II is acceptable. For higher  
 1576        masses, as well as for the boosted analysis, samples are produced between 1000 and 5000 GeV,  
 1577        and the detector is fully simulated. As discussed in Chapter 4, an outstanding issue with  
 1578        AtlFast-II is the modeling of jet substructure. While such variables are not used for the  
 1579        resolved analysis, the boosted analysis begins at 900 GeV, motivating the different detector  
 1580        simulation in these two regimes.

1581        The spin-2 resonance signal model is also simulated at LO in  $\alpha_s$  using MADGRAPH.  
 1582        Hadronization and parton showering are done using PYTHIA 8 [39] with EVTGEN, and the  
 1583        nominal PDF is NNPDF 2.3 LO. In practice this is implemented as a Randall-Sundrum  
 1584        graviton with  $c = 1.0$ .

1585        Spin-2 resonance samples are produced at masses between 251 and 5000 GeV, and these

1586 samples are all produced with full detector simulation.

1587 For the non-resonant search, samples are produced at values of  $\kappa_\lambda = 1.0$  and 10.0, and are  
1588 simulated using Powheg Box v2 generator [34–36] at next-to-leading order (NLO), with full  
1589 NLO corrections with finite top mass, using the PDF4LHC [68] parton distribution function  
1590 (PDF) set. Parton showers and hadronization are simulated with PYTHIA 8.

1591 Alternative ggF samples are simulated at NLO using Powheg Box v2, but instead using  
1592 HERWIG 7 [69] for parton showering and hadronization. The comparison between these two  
1593 is used to assess an uncertainty on the parton showering.

## 1594 7.2 Triggers and Object Definitions

1595 To maximize analysis sensitivity, a combination of multi- $b$ -jet triggers is used. Due to the use  
1596 of events with two  $b$ -tagged jets in the background estimate, such triggers have a maximum  
1597 requirement of two  $b$ -tagged jets. For the resonant analysis, a combination of triggers of  
1598 various topologies is used, namely

1599 • 2b + HT, which requires two  $b$ -tagged jets and a minimum value of of  $H_T$ , defined to  
1600 be the scalar sum of  $p_T$  across all jets in the event.

1601 • 2b + 2j, which requires two  $b$ -tagged jets and two other jets matching some kinematic  
1602 requirements

1603 • 2b + 1j, which requires two  $b$ -tagged jets and one other jet matching some kinematic  
1604 requirements

1605 • 1b, which requires one  $b$ -tagged jet

1606 Due to minimal contributions from some of these triggers for the Standard Model non-resonant  
1607 signal, a simplified strategy relying entirely on 2b + 1j and 2b + 2j triggers is used for the  
1608 non-resonant search.

1609 While the use of multiple triggers is beneficial for analysis sensitivity, it comes with some  
 1610 complications. Namely, a set of scale factors must be assigned to simulated events account  
 1611 for trigger inefficiencies in data *TODO: check*. Because these scale factors may differ between  
 1612 triggers, the use of multiple triggers becomes complicated: an event may pass more than one  
 1613 trigger, while trigger scale factors are only provided for individual triggers.

1614 To simplify this calculation, a set of hierarchical offline selections is applied, closely  
 1615 mimicking the trigger selection. Based on these selections, events are sorted into categories  
 1616 (*trigger buckets*), after which the decision of a *single trigger* is checked.

1617 The resonant search applies such categorization in the following way, with selections  
 1618 considered in order:

- 1619 1. If the leading jet is  $b$ -tagged with  $p_T > 325 \text{ GeV}$ , the event is in the  $1b$  trigger category.
- 1620 2. Otherwise, if the leading jet is not  $b$ -tagged, but has  $p_T > 168.75 \text{ GeV}$ , the event is in  
 1621 the  $2b + 1j$  trigger category.
- 1622 3. If neither of the first two selections pass, if the scalar sum of jet  $p_T$ s,  $H_T > 900 \text{ GeV}$ ,  
 1623 the event falls into the  $2b + HT$  trigger category.
- 1624 4. Events that do not pass any of the above offline selections are in the  $2b + 2j$  trigger  
 1625 category.

1626 Corresponding triggers are then checked in each category, and the final set of events consists  
 1627 of those events that pass the trigger decision in their respective categories.

1628 For the resonant search, the  $2b + 1j$  and  $2b + 2j$  triggers are the dominant categories,  
 1629 containing roughly 26 % and 49 % of spin-2 events, evaluated on MC16d samples with  
 1630 resonance masses between 300 and 1200 GeV. Notably, the  $1b$  trigger efficiency is largest at  
 1631 high ( $> 1 \text{ TeV}$ ) resonance masses.

1632 For the non-resonant search, it was noted that the  $1b$  trigger has minimal contribution,  
 1633 while the  $2b + HT$  events are largely captured by the  $2b + 2j$  trigger. Therefore, for, a

1634 simplified scheme is considered, with selections:

1635 1. If the 1st leading jet has  $p_T > 170 \text{ GeV}$  and the 3rd leading jet has  $p_T > 70 \text{ GeV}$ , the  
1636 event is in the  $2b + 1j$  trigger category.

1637 2. Otherwise, the event is in the  $2b + 2j$  trigger category.

1638 **7.3 Analysis Selection**

1639 After the trigger selections of Section 7.2, a variety of selections on the analysis objects are  
1640 made, with the goal of (1) reconstructing a  $HH$ -like topology and (2) suppressing contributions  
1641 from background processes.

1642 Both analyses begin with a common pre-selection, requiring at least four  $R = 0.4$  anti- $k_T$   
1643 jets with  $|\eta| < 2.5$  and  $p_T > 40 \text{ GeV}$ . The  $|\eta| < 2.5$  requirement is necessary for  $b$ -tagging  
1644 due to the coverage of the ATLAS tracking detector (see Chapter 3) *TODO: check*, while the  
1645  $p_T > 40 \text{ GeV}$  requirement is motivated by the trigger thresholds *TODO: mention low pT*. At  
1646 least two of the jets passing this pre-selection are required to be  $b$ -tagged, and additional  
1647  $b$ -tagging requirements are made to define the following regions:

- 1648 • “2 $b$  Region”: require exactly two  $b$ -tagged jets, used for background estimation
- 1649 • “4 $b$  Region”: require at least (but possibly more) four  $b$ -tagged jets, used as a signal  
1650 region for both resonant and non-resonant searches

1651 The non-resonant analysis additionally defines two 3 $b$  regions:

- 1652 • “3 $b+1$  loose Region”: require exactly three  $b$ -tagged jets which pass the 77 % b-tagging  
1653 working point (nominal) and one additional jet that fails the 77 % b-tagging working  
1654 point but passes the *looser* 85 % b-tagging working point. Used as a signal region for  
1655 the non-resonant search.

- 1656     • “3 $b$ +1 fail Region”: complement of 3 $b$ +1 loose. Require exactly three  $b$ -tagged jets  
 1657       which pass the 77 %  $b$ -tagging working point, but require that none of the remaining jets  
 1658       pass the 85 %  $b$ -tagging working point. Used as a validation region for the non-resonant  
 1659       search.

1660   After these requirements, four jets are chosen, ranked first by  $b$ -tagging requirement and then  
 1661   by  $p_T$  (e.g. for the 2 $b$  region, the jets chosen are the two  $b$ -tagged jets and the two highest  $p_T$   
 1662   non-tagged jets; for the 4 $b$  region, the jets are the four highest  $p_T$   $b$ -tagged jets). To match  
 1663   the topology of a  $HH \rightarrow b\bar{b}b\bar{b}$  event, these four jets are then *paired* into *Higgs candidates*: the  
 1664   four jets are split into two sets of two, and each of these pairs is used to define a reconstructed  
 1665   object that is a proxy for a Higgs in a  $HH$  event.

1666   For four jets there are three possible pairings. For signal events, a correct pairing can be  
 1667   identified (provided all necessary jets pass pre-selection) using the truth information of the  
 1668   Monte Carlo simulation, and such information may be used to design/select an appropriate  
 1669   pairing algorithm. This is only part of the story, however. The vast majority of the events in  
 1670   data do *not* include a real  $HH$  decay (this is a search for a reason!), either because the event  
 1671   originates from a background process (e.g. for 4 $b$  events), or because the selection is not  
 1672   designed to maximize the signal (e.g. 2 $b$  events). As the pairing is part of the selection, it must  
 1673   still be run on such events, such that various algorithms which achieve similar performance  
 1674   in terms of pairing efficiency may have vastly different impacts in terms of the shape of the  
 1675   background and the biases inherent in the background estimation procedure. The interplay  
 1676   between these two facets of the pairing is an important part of the choices made for this  
 1677   analysis.

1678   A comparison of different shapes due to three different paring strategies is shown in Figure  
 1679   7.1.

1680    7.3.1 Resonant Pairing Strategy

1681    For the resonant analysis, a Boosted Decision Tree (BDT) is used for the pairing. The boosted  
 1682    decision tree is given the total separation between the two jets in each of the two pairs ( $\Delta R_1$   
 1683    and  $\Delta R_2$ ), the pseudo-rapidity separation between the two jets in each pair ( $\Delta\eta_1$  and  $\Delta\eta_2$ ),  
 1684    and the angular separation between the two jets in each pair in the  $x - y$  plane ( $\Delta\phi_1$  and  
 1685     $\Delta\phi_2$ ). The total separations ( $\Delta R_s$ ) are provided in addition to the components in order to  
 1686    avoid requiring the boosted decision tree to reconstruct these variables in order to use them.  
 1687    For these variables, pair 1 is the pair with the highest scalar sum of jet  $p_{\text{T}}$ s, and pair 2 the  
 1688    other pair.

1689    The boosted decision tree is also parameterized on the di-Higgs mass ( $m_{HH}$ ) by providing  
 1690    this as an additional feature. Since the boosted decision tree is trained on correct and  
 1691    incorrect pairings in signal events, there will be exactly one correct pairing and two incorrect  
 1692    pairings in the training set for each  $m_{HH}$  value present in that set. As a result, this variable  
 1693    cannot, in itself, distinguish a correct pairing from an incorrect pairing, and therefore the  
 1694    inclusion of this variable simply serves to parameterize the BDT on  $m_{HH}$ <sup>2</sup>.

1695    The boosted decision tree was trained on one quarter of the total AFII simulated scalar  
 1696    MC statistics, using the Gradient-based One Side Sampling (GOSS) algorithm which allows  
 1697    rapid training with very large datasets. A preselection was applied requiring events to have  
 1698    four jets with a  $p_{\text{T}}$  of at least 35 GeV. Note that this is a looser requirement than the 40 GeV  
 1699    used in the analysis selection, and is meant to increase the available statistics for events with  
 1700    low  $m_{HH}$  and to ensure a better performance as a function of that variable. Events were also  
 1701    required to have four distinct jets that could be geometrically matched (to within  $\Delta R \leq 0.4$ )  
 1702    to the  $b$ -quarks. The events used to train the BDT were not included when the analysis was  
 1703    run on these signal simulations. The boosted decision tree was constructed with the following  
 1704    hyperparameters:

1705    `min_data_in_leaf=50,`

---

<sup>2</sup>That is, the conditions placed on the other variables by the BDT vary with  $m_{HH}$ .

1706 num\_leaves=180,  
 1707 learning\_rate=0.01

1708 These hyperparameters were optimized using a Bayesian optimization procedure [70].  
 1709 Three fold cross-validation was used to perform this optimization without the need for an  
 1710 additional sample, while avoiding over-training on signal events.

1711 *7.3.2 Non-resonant Pairing Strategy*

1712 For the non-resonant analysis, a simpler pairing algorithm is used, which proceeds as follows:  
 1713 in a given event, Higgs candidates for each possible pairing are sorted by the  $p_T$  of the vector  
 1714 sum of constituent jets. The angular separation,  $\Delta R$  is computed between jets in the each of  
 1715 the leading Higgs candidates, and the pairing with the smallest separation ( $\Delta R_{jj}$ ) is selected.  
 1716 This method will be referred to as  $\min \Delta R$  in the following.

1717 The primary motivation for the use of this pairing in the non-resonant search is a *smooth*  
 1718 *mass plane*: by efficiently discarding low mass events,  $\min \Delta R$  removes the background peak  
 1719 present in the resonant search while maintaining good pairing efficiency for the Standard  
 1720 Model non-resonant signal. This facilitates a background estimate with small kinematic bias  
 1721 – the region in which the background estimate is derived is more similar to the signal region.

1722 Along with discarding low mass background, there is a corresponding loss of low mass  
 1723 signal. This predominantly impacts points away from the Standard Model (see Figure 7.2),  
 1724 but, because the  $4b$  channel has the strongest contribution near the Standard Model and  
 1725 because of the large low mass background present with other pairing methods, the impact on  
 1726 analysis sensitivity is mitigated. The  $\min \Delta R$  pairing is thus adopted for the non-resonant  
 1727 search.

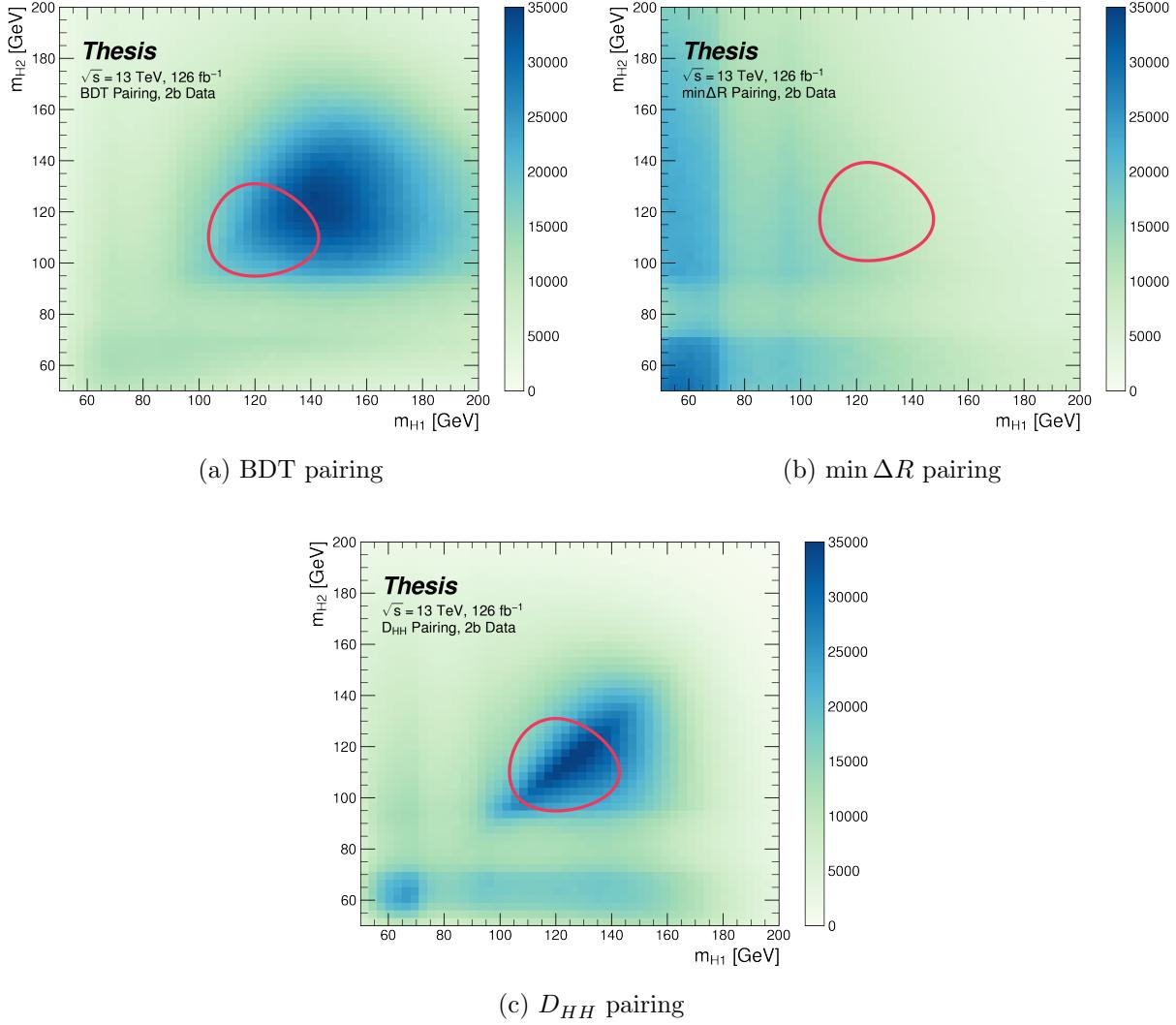


Figure 7.1: Comparison of  $m_{H1}$  vs  $m_{H2}$  planes for the full Run 2 2b dataset with different pairings. As evidenced, this choice significantly impacts where events fall in this plane, and therefore which events fall into the various kinematic regions defined in this plane (see Section 7.5). 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.

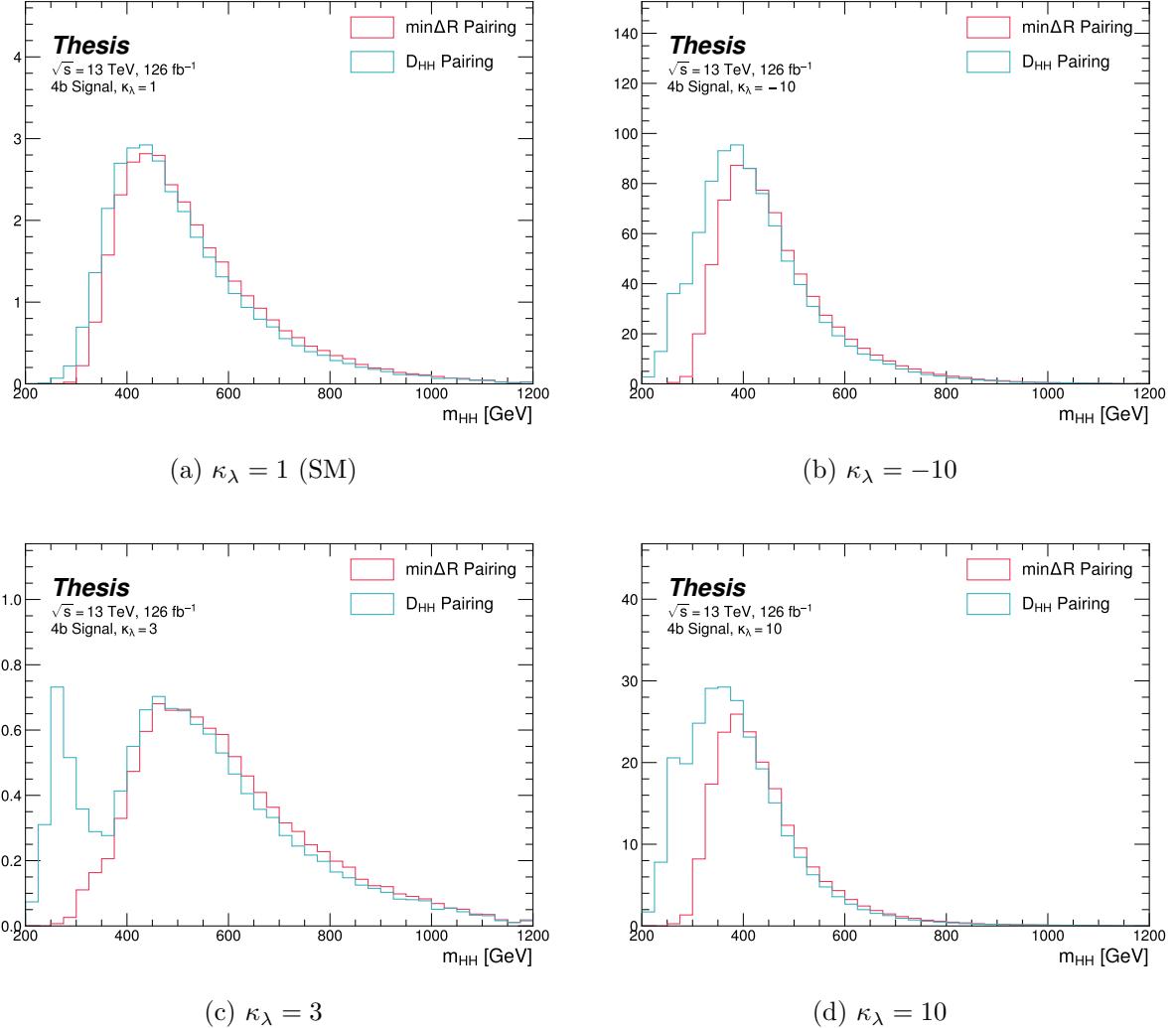


Figure 7.2: Comparison of signal distributions in the respective signal regions for the  $\min \Delta R$  and  $D_{HH}$  pairing for various values of the Higgs trilinear coupling in the respective signal regions. The distributions are quite similar at the Standard Model point, but for other variations,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

1728 **7.4 Background Reduction and Top Veto**

1729 Choosing a pairing of the four b-tagged jets fully defines the di-Higgs candidate system used  
1730 for each event in the remainder of the analysis chain. A requirement of  $|\Delta\eta_{HH}| < 1.5$  on this  
1731 di-Higgs candidate system mitigates QCD multijet background.

1732 Figure ?? illustrates this variable in the validation region (see Section ??). It demonstrates  
1733 that this selection rejects only a small fraction of signal, but a significant fraction of data  
1734 (which, in the validation region, is a good approximation of pure background).

1735 In order to mitigate the hadronic  $t\bar{t}$  background, a top veto is then applied, removing  
1736 events consistent with a  $t \rightarrow b(W \rightarrow q_1\bar{q}_2)$  decay.

1737 The jets in the event are separated into *HC jets* which are the four jets used to build the  
1738 Higgs candidates, and *non-*HC jets**, the other jets (passing the  $p_T$  and  $|\eta|$  requirements) in  
1739 the event.

1740  $W$  candidates are built by forming all possible pairs of all jets in each event. With  $n$  jets,  
1741 there are  $\binom{n}{2}$  such pairs.  $t$  candidates are then built by pairing each  $W$  candidate with each  
1742 HC jet (for  $4\binom{n}{2}$  combinations). Note that all jets in a  $t$  candidate must be distinct (i.e. a  
1743 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

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

1744 is constructed for each combination.

1745 Events are then vetoed if the minimum  $X_{Wt}$  over all combinations is less than 1.5.

1746 The same definitions and procedures are used for both the resonant and non-resonant  
1747 analyses. However, for the non-resonant search, the top candidates considered for  $X_{Wt}$  have  
1748 the additional requirement that the jet used for the  $b$  is  $b$ -tagged. While this is identical to  
1749 the resonant analysis by definition for  $4b$  events, it does change the set of events considered in  
1750 lower tag regions, in particular for the  $2b$  events considered in the derivation of the background

1751 estimate. Such a change is found to reduce the impact of background systematics by increasing

1752  $2b$  support in the high  $X_{Wt}$  kinematic region. *TODO: Insert plot*

<sup>1753</sup> **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:

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

<sup>1754</sup> with  $m(H_1)$ ,  $m(H_2)$  the leading and subleading Higgs candidate masses,  $c_1$  and  $c_2$  correspond  
<sup>1755</sup> to the center of the signal region, and the denominator provides a Higgs candidate mass  
<sup>1756</sup> dependent resolution of 10 %. For consistency with the  $HH$  decay hypothesis,  $c_1$  and  $c_2$   
<sup>1757</sup> are nominally (125 GeV, 125 GeV). However, these are allowed to vary due to energy loss,  
<sup>1758</sup> with specific values chosen described below. The selection of these values is one of several  
<sup>1759</sup> significant differences between the regions defined for the resonant and non-resonant search.  
<sup>1760</sup> We describe both below.

<sup>1761</sup> **7.5.1 Resonant Kinematic Regions**

<sup>1762</sup> For the resonant analysis, the signal region is centered at (120 GeV, 110 GeV) to account for  
<sup>1763</sup> energy loss leading to the Higgs masses being under-reconstructed. *TODO: insert signal*  
<sup>1764</sup> *location plot?* Note that leading and subleading Higgs candidates are defined according to  
<sup>1765</sup> 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

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

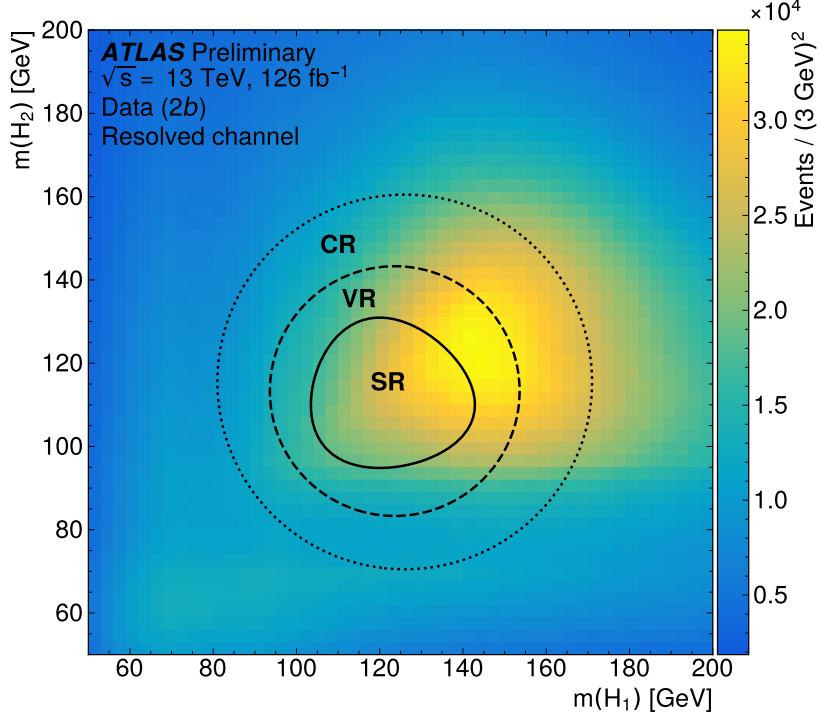


Figure 7.3: Regions used for the resonant search, shown on the  $2b$  data mass plane. The outermost region (the “control region”) is used for derivation of the nominal background estimate. The innermost region is the signal region, where the signal extraction fit is performed. The region in between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

1766 For simplicity, the SR/VR/CR definitions from the early Run 2 paper [71] were chosen  
1767 for the resonant analysis, but were found to be close to optimal. These regions are shown in  
1768 Figure 7.3.

### 1769 7.5.2 Non-resonant Kinematic Regions

1770 For the non-resonant analysis the signal region is centered at (124 GeV, 117 GeV), corre-  
1771 sponding to the means of *correctly paired* Standard Model signal events. The shape of the  
1772 signal region (other than this change of center) was found to remain optimal.

1773 For the non-resonant search, leading and subleading Higgs candidates are defined according  
 1774 to the *vector sum* of constituent jet  $p_T$ , more closely corresponding to the  $1 \rightarrow 2$  decay  
 1775 assumption behind the min  $\Delta R$  pairing algorithm.

1776 Two areas for improvement were identified in the resonant analysis, which will be dis-  
 1777 cussed in more detail below: *signal contamination* of the validation region (which impacts  
 1778 the uncertainty assessed due to extrapolation) and *large nuisance parameter pulls* on this  
 1779 uncertainty, corresponding to a rough assumption that the validation region is closer to the  
 1780 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:

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

1781 is kept, roughly corresponding to combining the regions used for the resonant analysis. In  
 1782 order to assess the variation of the background estimate, two sets of regions are desired, so  
 1783 this combined region is split into *quadrants*, that is, divided into four pieces along axes that  
 1784 intersect with the signal region center. To avoid kinematic bias, quadrants on opposite sides  
 1785 of the signal region are paired, with these pairs corresponding to the non-resonant control  
 1786 and validation regions.

1787 The particular orientation of the regions is chosen such that region centers align with the  
 1788 leading and subleading Higgs candidate masses, corresponding to a set of axes rotated at  
 1789  $45^\circ$ , with the “top” and “bottom” quadrants together comprising the control region, and the  
 1790 other set (“left” and “right”) the validation region. These regions are shown in Figure 7.4

1791 This design of regions includes a set of events closer to the signal region in the mass plane,  
 1792 leveraging the assumption that these events are more similar to signal region events, while  
 1793 also including events further away from the signal region, mitigating signal contamination.  
 1794 This region selection is found to have good performance in alternate validation regions (see  
 1795 Section 7.8).

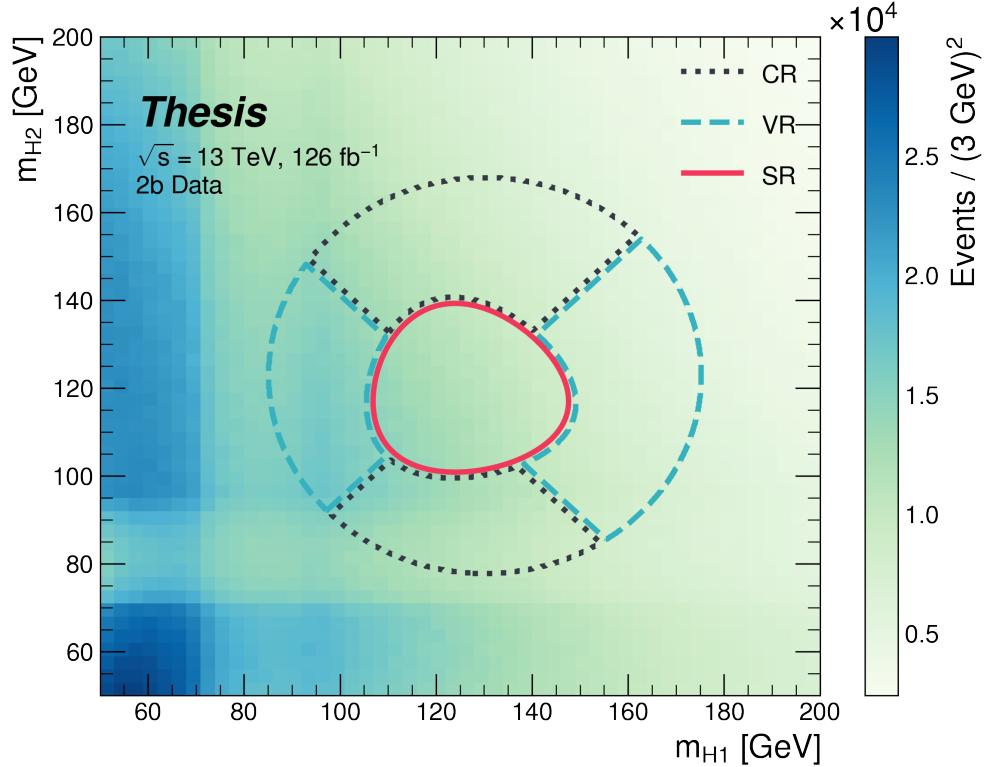


Figure 7.4: Regions used for the non-resonant search. The “top” and “bottom” quadrants together comprise the control region, in which the nominal background estimate is derived. The “left” and “right” quadrants together comprise the validation region, which is used to assess an uncertainty. The signal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

1796 7.5.3 Discriminating Variable

1797 The discriminant used for the resonant analysis is *corrected*  $m_{HH}$ . This variable is calculated  
 1798 by re-scaling the Higgs candidate four vectors such that each  $m_H = 125 \text{ GeV}$ . These re-scaled  
 1799 four-vectors are then summed, and their invariant mass is the corrected  $m_{HH}$ . These re-scaled  
 1800 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.

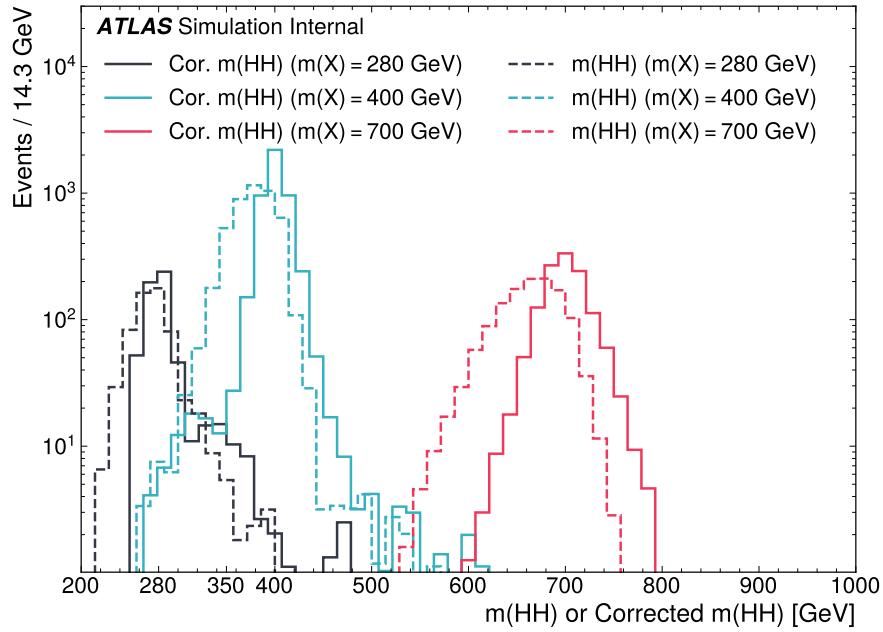


Figure 7.5: Impact of the  $m_{HH}$  correction on a range of spin-0 resonant signals. The corrected  $m_{HH}$  distributions (solid lines) are much sharper and more centered on the corresponding resonance masses than the uncorrected  $m_{HH}$  distributions (dashed).

1801

1802 For the non-resonant analysis, due to the broad nature of the signal in  $m_{HH}$ , such a  
 1803 correction is not as motivated, and, indeed, is found to have very minimal impact. The  
 1804 uncorrected  $m_{HH}$  (just referred to as  $m_{HH}$ ) is therefore used as a discriminant. To maximize

1805 sensitivity, the non-resonant analysis additionally uses two variables for categorization:  $\Delta\eta_{HH}$ ,  
1806 an angular variable which, along with  $m_{HH}$ , fully characterizes the  $HH$  system [72], and  $X_{HH}$ ,  
1807 the variable used for the signal region definition, which leverages the peaked structure of the  
1808 signal in the  $(m(H_1), m(H_2))$  plane to split the signal extraction fit into lower and higher  
1809 purity regions (highest purity near  $X_{HH} = 0$ , the center of the signal region). Distributions  
1810 of these variables are shown in *TODO: plots*. The categorization used for this thesis has been  
1811 optimized to be  $2 \times 2$  in these variables, with corresponding selections  $0 \leq \Delta\eta_{HH} \leq 0.75$  and  
1812  $0.75 \leq \Delta\eta_{HH} \leq 1.5$  for  $\Delta\eta_{HH}$ , and  $0 \leq X_{HH} \leq 0.95$  and  $0.95 \leq X_{HH} \leq 1.6$  for  $X_{HH}$ .

1813 **7.6 Background Estimation**

1814 After the event selection described above there are two major backgrounds, QCD and  $t\bar{t}$ .  
1815 A very similar approach is used for both the resonant and the non-resonant analyses, with  
1816 some small modifications. This approach is notably fully data-driven, which is warranted due  
1817 to the flexibility of the estimation method, as well as the high relative proportion of QCD  
1818 background ( $> 90\%$ ), and allows for the use of machine learning methods in the construction  
1819 of the background estimate. However, it sacrifices an explicit treatment of the  $t\bar{t}$  component.  
1820 Performance of the background estimate on the  $t\bar{t}$  component is checked explicitly *TODO:*  
1821 *add plots*, and minimal impact due to  $t\bar{t}$  mismodeling is seen.

1822 Contributions of single Higgs processes and  $ZZ$  are found to be negligible, and the  
1823 Standard Model  $HH$  background is found to have no impact on the resonant search.

1824 The foundation of the background estimate lies in the derivation of a reweighting function  
1825 which matches the kinematics of events with exactly two  $b$ -tagged jets to those of events in  
1826 the higher tagged regions (events with three or four  $b$ -tagged jets). The reweighting function  
1827 and overall normalization are derived in the control region. Systematic bias of this estimate  
1828 is assessed in the validation region.

1829 For the resonant analysis, the systematic bias is a bias due to extrapolation: the validation  
1830 region lies between the control and signal regions. For the non-resonant analysis, the bias  
1831 instead comes from different possible interpolations of the signal region kinematics – given the  
1832 choice of nominal estimate, the validation region is a conceptually equivalent, but maximally  
1833 different, signal region estimate.

1834 **7.6.1 The Two Tag Region**

1835 Events in data with exactly two  $b$ -tagged jets are used for the data driven background  
1836 estimate. The hypothesis here is that, due to the presence of multiple  $b$ -tagged jets, the  
1837 kinematics of such events are similar to the kinematics of events in higher  $b$ -tagged regions (i.e.  
1838 events with three and four  $b$ -tagged jets, respectively), and any differences can be corrected

1839 by a reweighting procedure. The region with three  $b$ -tagged jets is split into two  $b$ -tagging  
 1840 regions, with the  $3b + 1$  loose region used as an additional signal region (see Section *TODO:*  
 1841 *Add ref*). The lower tagged  $3b$  component ( $3b + 1$  fail, as described in Section ??) is reserved  
 1842 for validation of the background modelling procedure. Events with fewer than two  $b$ -tagged  
 1843 jets are not used for this analysis, as they are relatively more different from the higher tag  
 1844 regions.

1845 The nominal event selection requires at least four jets in order to form Higgs candidates.  
 1846 For the four tag region, these are the four highest  $p_T$   $b$ -tagged jets. For the three tag regions,  
 1847 these jets are the three  $b$ -tagged jets, plus the highest  $p_T$  jet satisfying a loosened  $b$ -tagging  
 1848 requirement. Similarly, and following the approach of the resonant analysis, the two tag region  
 1849 uses the two  $b$ -tagged jets and the two highest  $p_T$  non-tagged jets to form Higgs candidates.  
 1850 Combinatoric bias from selection of different numbers of  $b$ -tagged jets is corrected as a part  
 1851 of the kinematic reweighting procedure through the reweighting of the total number of jets in  
 1852 the event. In this way, the full event selection may be run on two tagged events.

### 1853 7.6.2 Kinematic Reweighting

1854 The set of two tagged data events is the fundamental piece of the data driven background  
 1855 estimate. However, kinematic differences from the four tag region exist and must be corrected  
 1856 in order for this estimate to be useful. Binned approaches based on ratios of histograms  
 1857 have been previously considered [71], [17], but are limited in their handling of correlations  
 1858 between variables and by the “curse of dimensionality”, i.e. the dataset becomes sparser and  
 1859 sparser in “reweighting space” as the number of dimensions in which to reweight increases,  
 1860 limiting the number of variables used for reweighting. This leads either to an unstable fit  
 1861 result (overfitting with finely grained bins) or a lower quality fit result (underfitting with  
 1862 coarse bins).

1863 Note that even machine learning methods such as Boosted Decision Trees (BDTs), may  
 1864 suffer from this curse of dimensionality, as the depth of each decision tree used is limited  
 1865 by the available statistics after each set of corresponding selections (cf. binning in a more

1866 sophisticated way), limiting the expressivity of the learned reweighting function.

1867 To solve these issues, a neural network based reweighting procedure is used here. This  
1868 is a truly multivariate approach, allowing for proper treatment of variable correlations. It  
1869 further overcomes the issues associated with binned approaches by learning the reweighting  
1870 function directly, allowing for greater sensitivity to local differences and helping to avoid the  
1871 curse of dimensionality.

1872 *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 [73, 74], 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

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

1873 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

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

1874 Taking the exponent then results in the desired reweighting function.

1875 Note that similar methods for density ratio estimation are available *TODO: cite*, e.g. from

1876 a more standard binary cross-entropy loss. However, these were found to perform no better  
1877 than the formulation presented here.

1878 *Variables and Results*

1879 The neural network is trained on a variety of variables sensitive to two vs. four tag differences.

1880 To help bring out these differences, the natural logarithm of some of the variables with a  
1881 large, local change is taken. The set of training variables used for the resonant analysis is

1882 1.  $\log(p_T)$  of the 4th leading Higgs candidate jet

1883 2.  $\log(p_T)$  of the 2nd leading Higgs candidate jet

1884 3.  $\log(\Delta R)$  between the closest two Higgs candidate jets

1885 4.  $\log(\Delta R)$  between the other two Higgs candidate jets

1886 5. Average absolute value of Higgs candidate jet  $\eta$

1887 6.  $\log(p_T)$  of the di-Higgs system.

1888 7.  $\Delta R$  between the two Higgs candidates

1889 8.  $\Delta\phi$  between the jets in the leading Higgs candidate

1890 9.  $\Delta\phi$  between the jets in the subleading Higgs candidate

1891 10.  $\log(X_{Wt})$ , where  $X_{Wt}$  is the variable used for the top veto

1892 11. Number of jets in the event.

1893 The non-resonant analysis uses an identical set of variables with two notable changes

1894 1. The definition of  $X_{Wt}$  differs from the resonant definition (as described in Section  
1895 *TODO: ref*)

1896 2. An integer encoding of the two trigger categories is used as an input (variable which  
1897 takes on the value 0 or 1 corresponding to each of the two categories). This was found  
1898 to improve a mismodeling near the tradeoff in  $m_{HH}$  of the two buckets.

1899 The neural network used for both resonant and non-resonant reweighting has three densely  
1900 connected hidden layers of 50 nodes each with ReLU activation functions and a single node  
1901 linear output. This configuration demonstrates good performance in the modelling of a variety  
1902 of relevant variables, including  $m_{HH}$ , when compared to a range of networks of similar size.

1903 In practice, a given training of the reweighting neural network is subject to variation  
1904 due to training statistics and initial conditions. An uncertainty is assigned to account for  
1905 this (Section 7.7), which relies on training an ensemble of reweighting networks [75]. To  
1906 increase the stability of the background estimate, the median of the predicted weight for each  
1907 event is calculated across the ensemble. This median is then used as the nominal background  
1908 estimate. This approach is indeed seen to be much more stable and to demonstrate a better  
1909 overall performance than a single arbitrary training. Each ensemble used for this analysis  
1910 consists of 100 neural networks, trained as described in Section 7.7.

1911 The training of the ensemble used for the nominal estimate is done in the kinematic  
1912 Control Region. The prediction of these networks in the Signal Region is then used for the  
1913 nominal background estimate. In addition, a separate ensemble of networks is trained in the

1914 Validation Region. The difference between the prediction of the nominal estimate and the  
 1915 estimate from the VR derived networks in the Signal Region is used to assign a systematic  
 1916 uncertainty. Further details on this systematic uncertainty are shown in Section 7.7. Note  
 1917 that although the same procedure is used for both Control and Validation Region trained  
 1918 networks, only the median estimate from the VR derived reweighting is used for assessing a  
 1919 systematic – no additional “uncertainty on the uncertainty” from VR ensemble variation is  
 1920 applied.

1921 Each reweighted estimate is normalized such that the reweighted  $2b$  yield matches the  $4b$   
 1922 yield in the corresponding training region. Note that this applies to each of the networks used  
 1923 in each ensemble, where the normalization factor is also subject to the procedure described in  
 1924 Section 7.7. As the median over these normalized weights is not guaranteed to preserve this  
 1925 normalization, a further correction is applied such that the  $2b$  yield, after the median weights  
 1926 are applied, matches the  $4b$  yield in the corresponding training region. As no preprocessing  
 1927 is applied to correct for the class imbalance between  $2b$  and  $4b$  events entering the training,  
 1928 this ratio of number of  $4b$  events ( $n(4b)$ ) over number of  $2b$  events ( $n(2b)$ ) is folded into the  
 1929 learned weights. Correspondingly, the set of normalization factors described above is near 1  
 1930 and the learned weights are centered around  $n(4b)/n(2b)$  (roughly 0.01 over the full dataset).  
 1931 This normalization procedure applies for all instances of the reweighting (e.g. those used for  
 1932 validations in Section ??), with appropriate substitutions of reweighting origin (here  $2b$ ) and  
 1933 reweighting target (here  $4b$ ).

1934 Note that, due to different trigger and pileup selections during each year, the reweighting  
 1935 is trained on each year separately. An approach of training all of the years together with a  
 1936 one-hot encoding was explored *TODO: reference study*, but was found to have minimal benefit  
 1937 over the split years approach, and in fact to increase the systematic bias of the corresponding  
 1938 background estimate. Because of this, and because trigger selections for each year significantly  
 1939 impact the kinematics of each year, such that categorizing by year is expected to reflect  
 1940 groupings of kinematically similar events and to provide a meaningful degree of freedom in  
 1941 the signal extraction fit, the split-year approach is kept.

1942      The control region closure for the 2018 dataset is shown for the resonant search in Figures  
 1943    [7.6](#) through [7.14](#) and for the non-resonant search in Figures [7.24](#) through [7.32](#) for  $4b$  and  
 1944    Figures [7.42](#) through [7.50](#) for  $3b1l$ . The impact of this control region derived reweighting  
 1945    on the validation region is shown in Figures [7.15](#) through [7.23](#) for the resonant search and  
 1946    Figures [7.33](#) through [7.41](#) for  $4b$  and Figures [7.51](#) through [7.59](#) for  $3b1l$  for the non-resonant  
 1947    search. Generally good performance is seen, with some occasional mis-modeling. For the  
 1948    resonant search, this is most notable in the case of individual jet  $p_T$ . Such mis-modeling  
 1949    may be corrected by including the variables in the input set, but this was found to not  
 1950    improve the modeling of  $m_{HH}$ , and so is not done here. This mis-modeling is notable for the  
 1951    non-resonant search in the leading Higgs candidate jet  $p_T$ , and is a direct consequence of the  
 1952    trigger category input, which improves modeling of  $m_{HH}$ . Results are similar for other years,  
 1953    but are not included here for brevity.

1954      One other salient feature of the non-resonant plots is the distributions of  $m_{H1}$  and  $m_{H2}$ ,  
 1955    which emphasize the quadrant region definitions – the control region has a peak around  
 1956    125 GeV in  $m_{H1}$ , which may be thought of as “signal region-like”, motivating this alignment,  
 1957    though consequently the distribution of  $m_{H2}$  is quite bimodal. The reverse is true for the  
 1958    validation region.

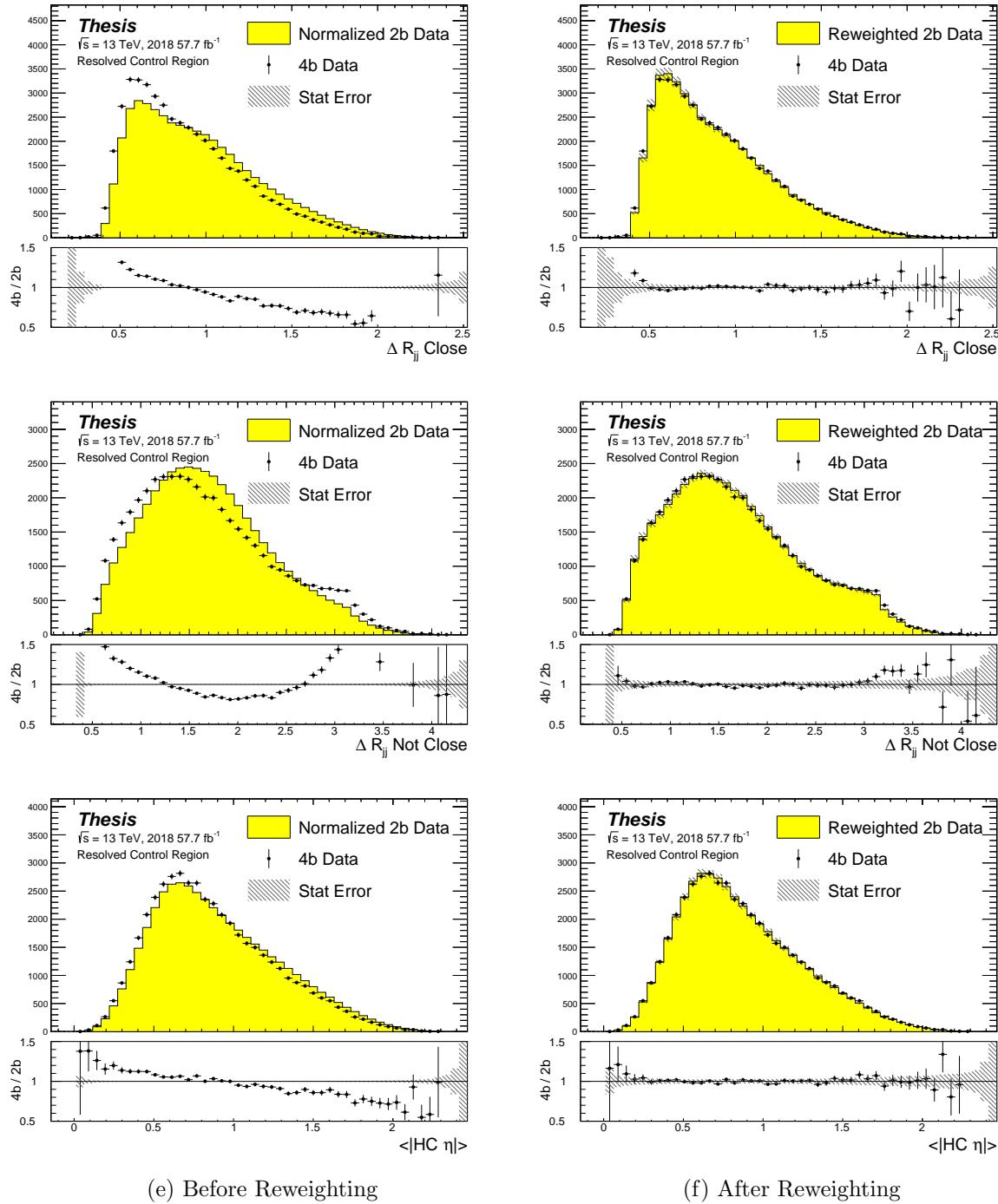


Figure 7.6: **Resonant Search:** Distributions of  $\Delta R$  between the closest Higgs Candidate jets,  $\Delta R$  between the other two, and average absolute value of HC jet  $\eta$  before and after CR derived reweighting for the 2018 Control Region.

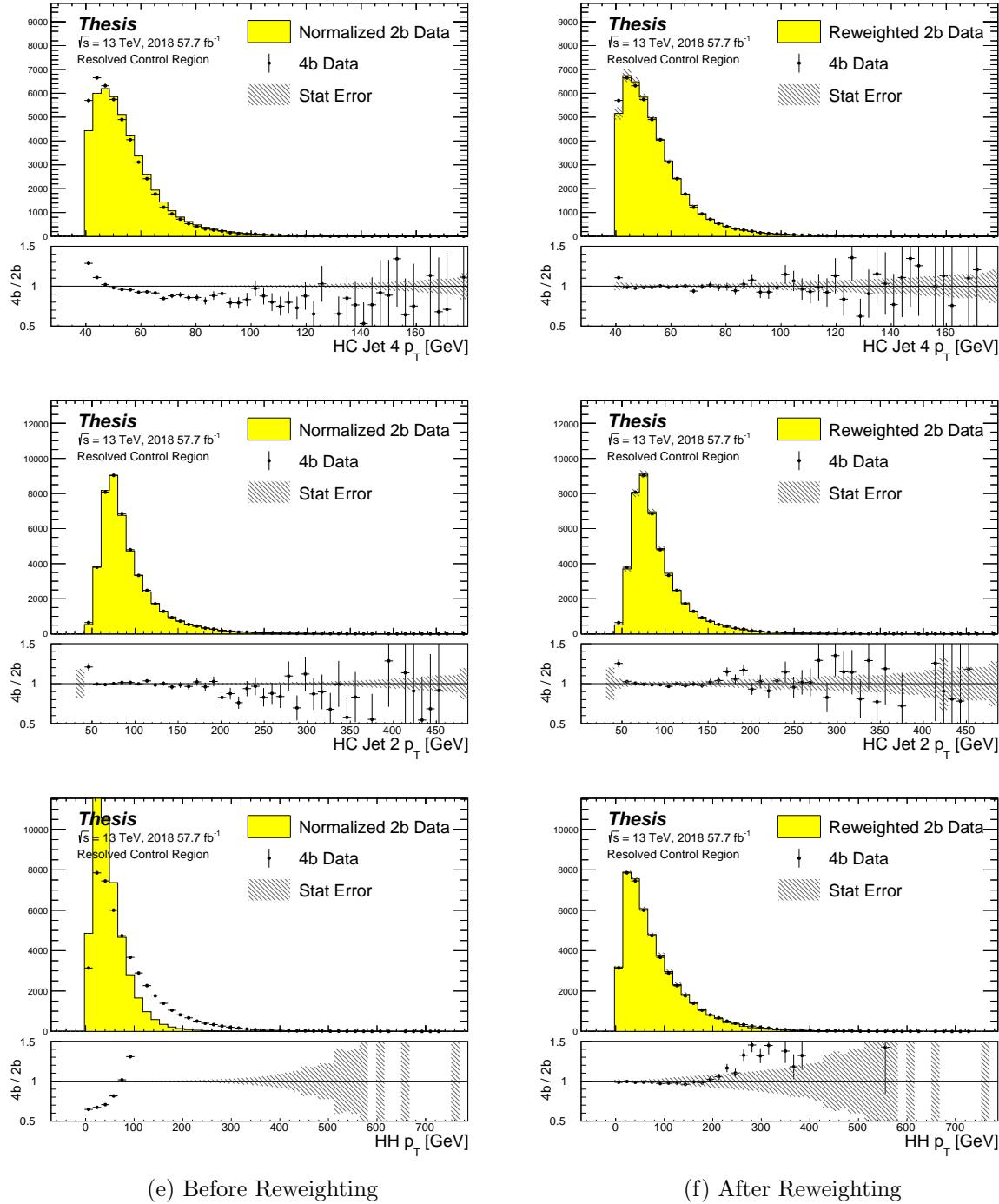
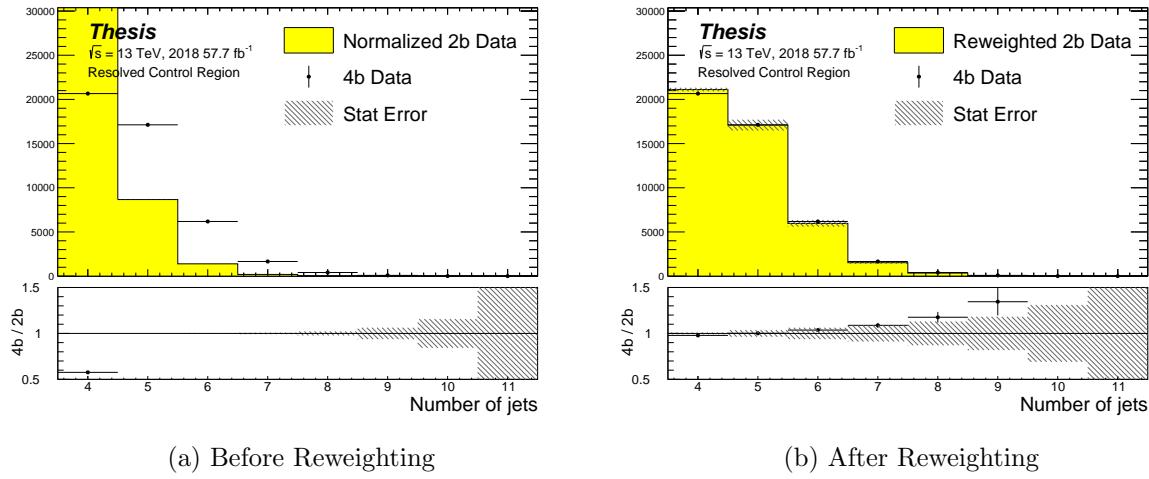


Figure 7.7: **Resonant Search:** Distributions of  $p_T$  of the 2nd and 4th leading Higgs Candidate jets and the  $p_T$  of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.



**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.

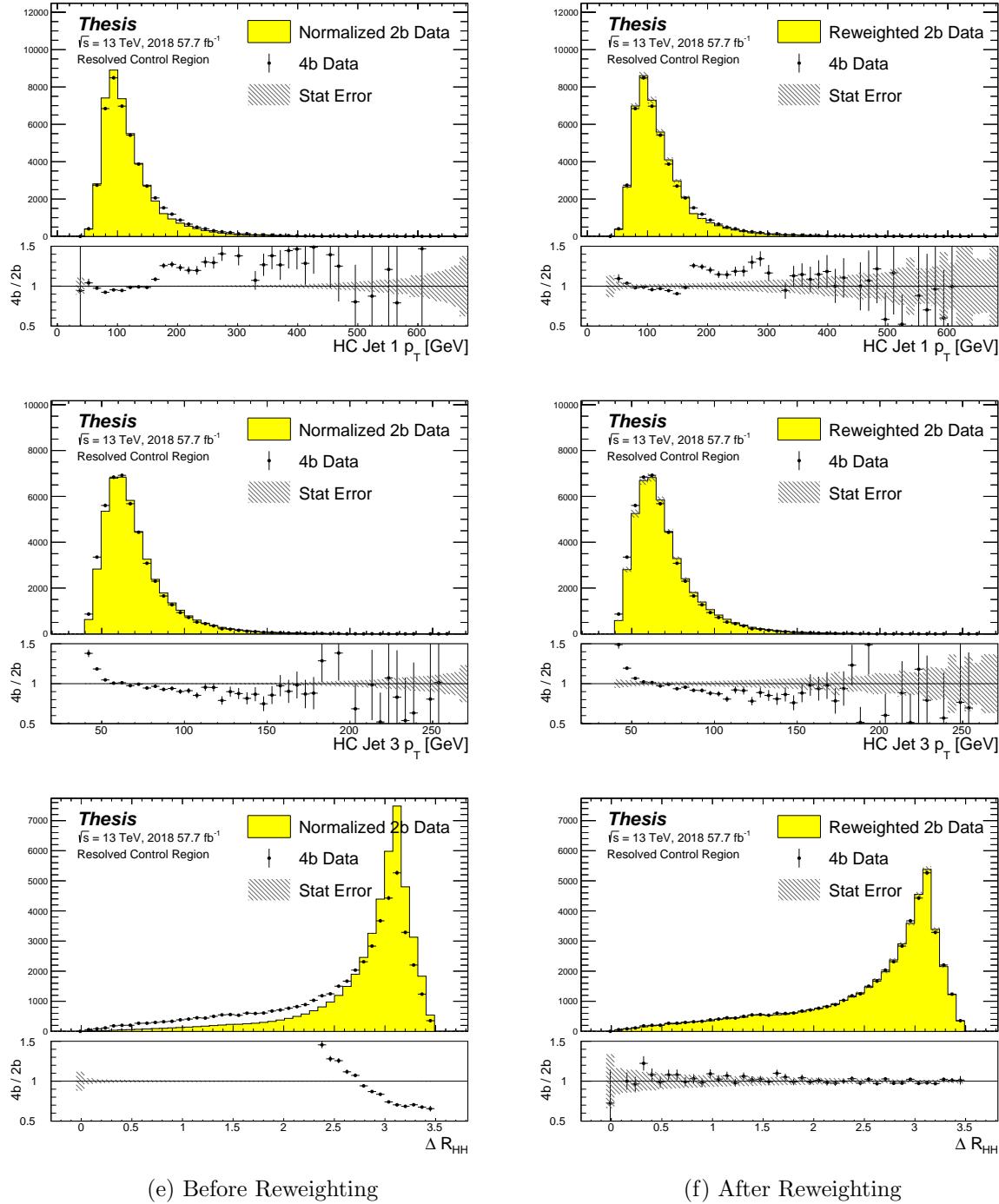


Figure 7.9: **Resonant Search:** Distributions of  $p_T$  of the 1st and 3rd leading Higgs Candidate jets and  $\Delta R$  between Higgs candidates before and after CR derived reweighting for the 2018 Control Region.

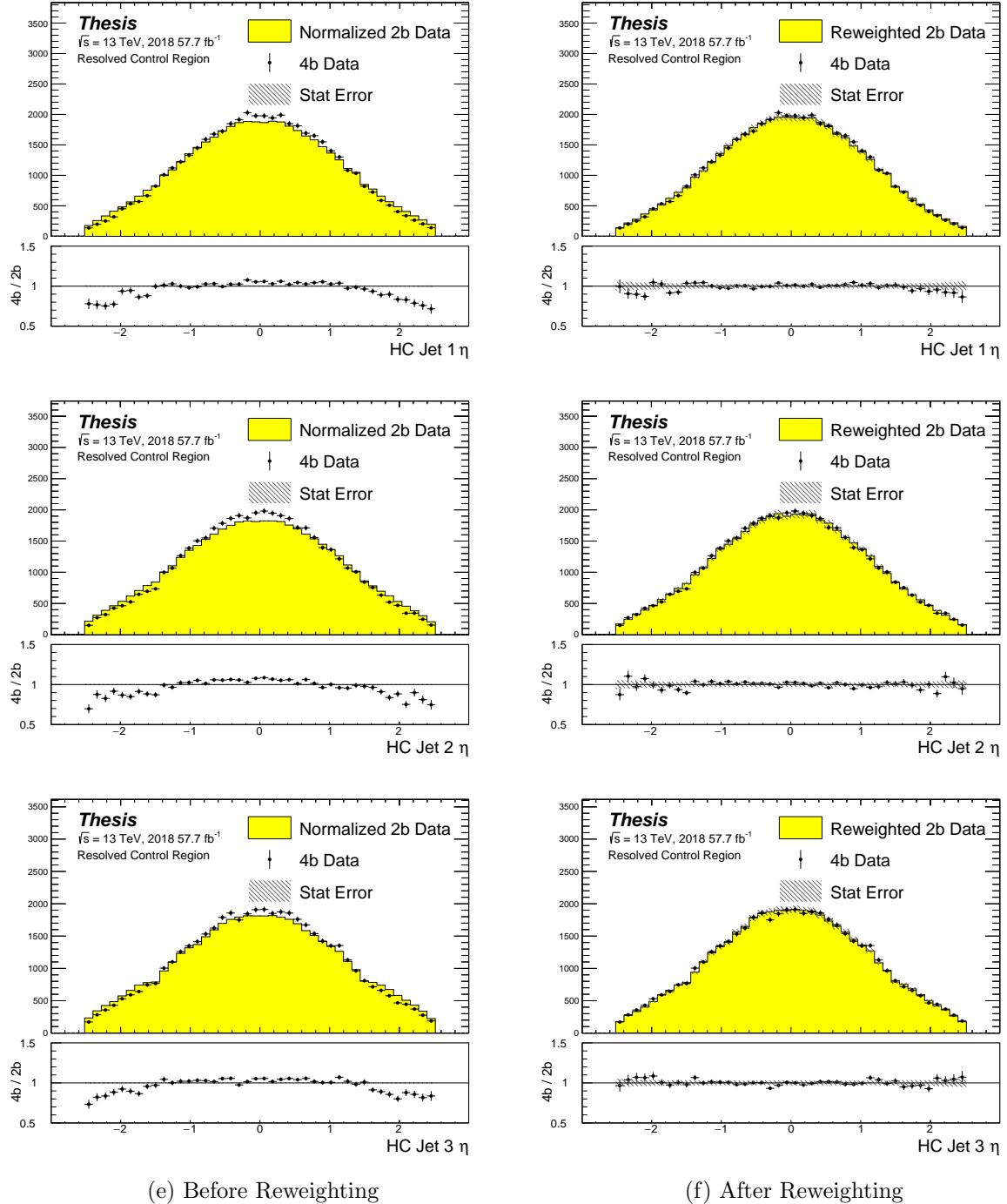


Figure 7.10: **Resonant Search:** Distributions of  $\eta$  of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Control Region.

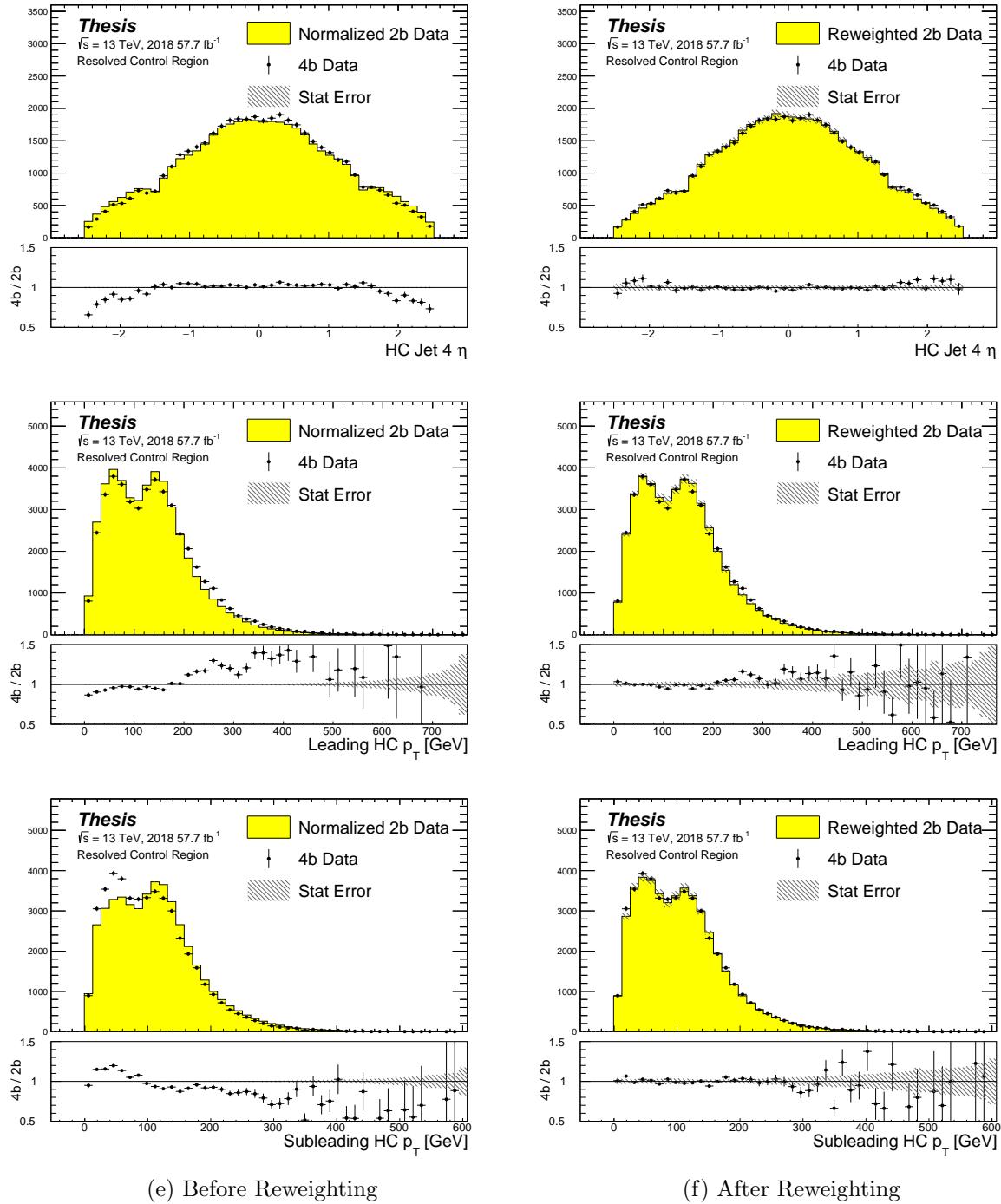


Figure 7.11: **Resonant Search:** Distributions of  $\eta$  of the 4th leading Higgs Candidate jet and the  $p_T$  of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.

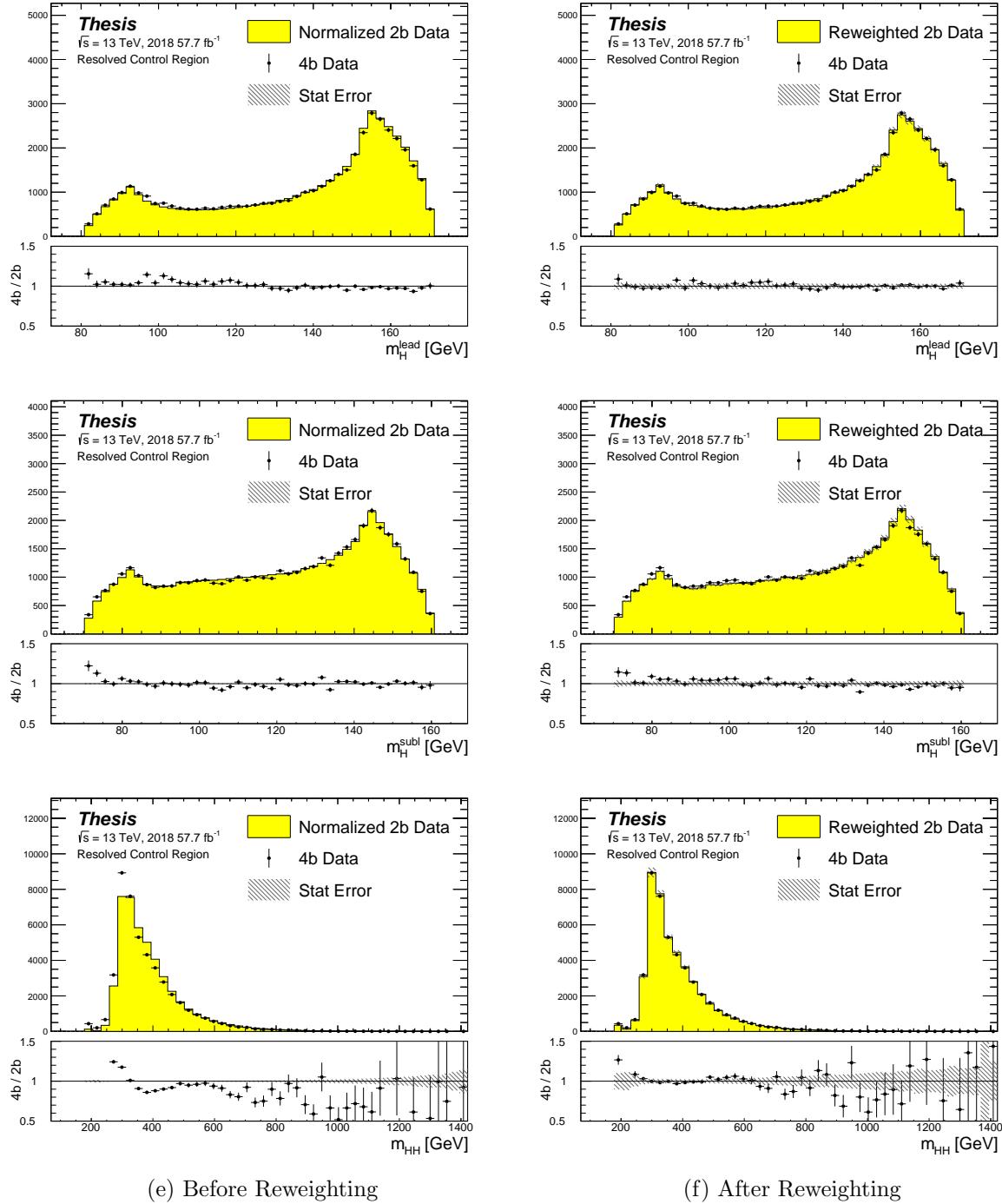


Figure 7.12: **Resonant Search:** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.

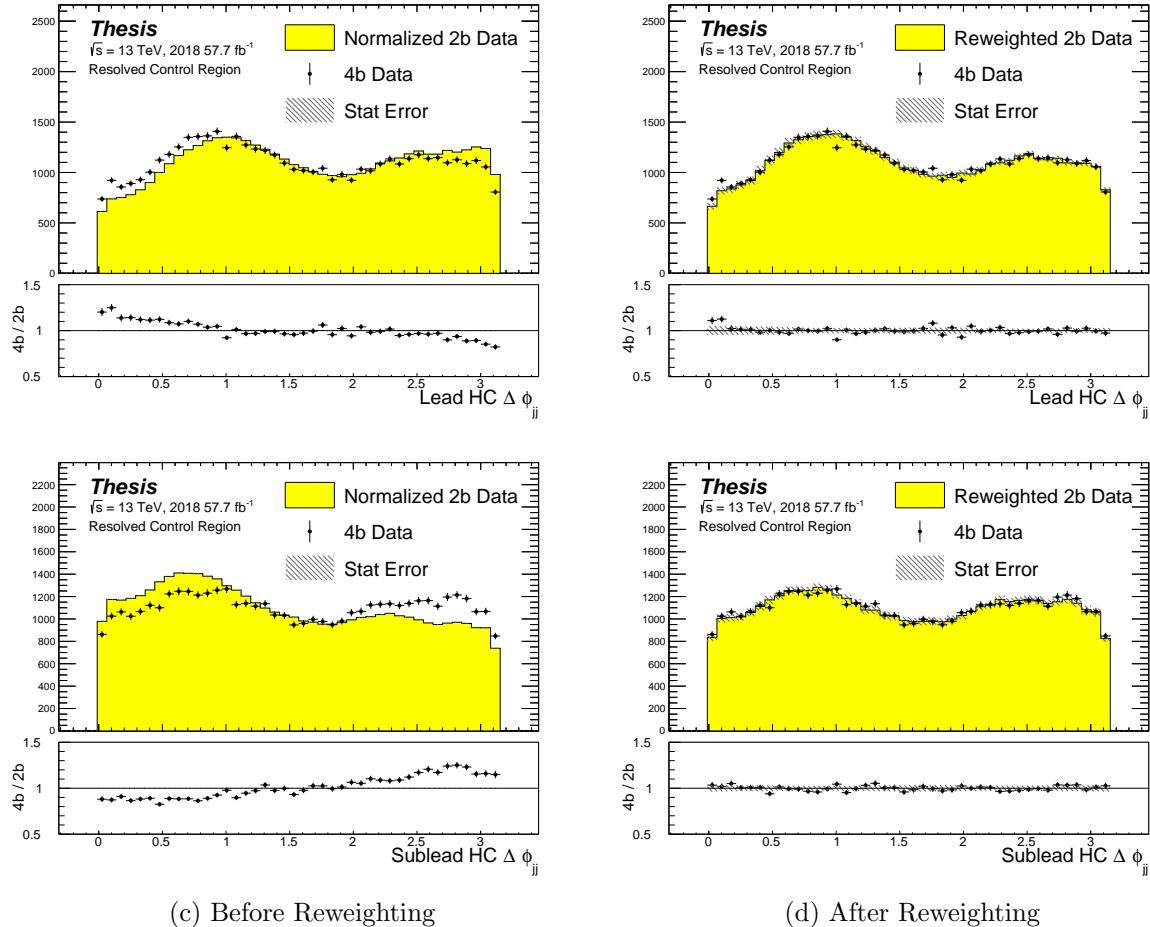


Figure 7.13: **Resonant Search:** Distributions of  $\Delta\phi$  between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.

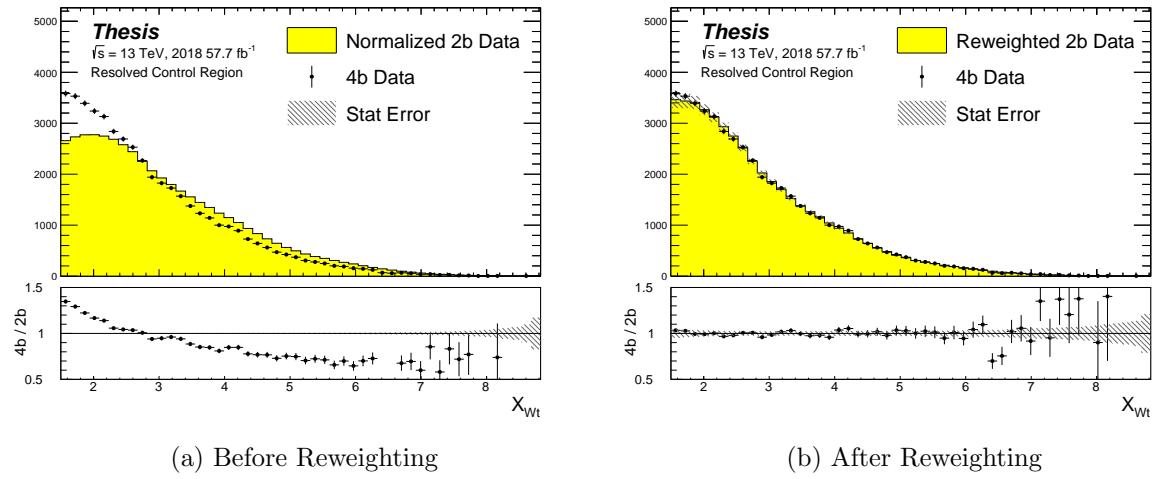


Figure 7.14: **Resonant Search:** Distributions of the top veto variable,  $X_{Wt}$ , before and after CR derived reweighting for the 2018 Control Region. Reweighting is done after the cut on this variable is applied

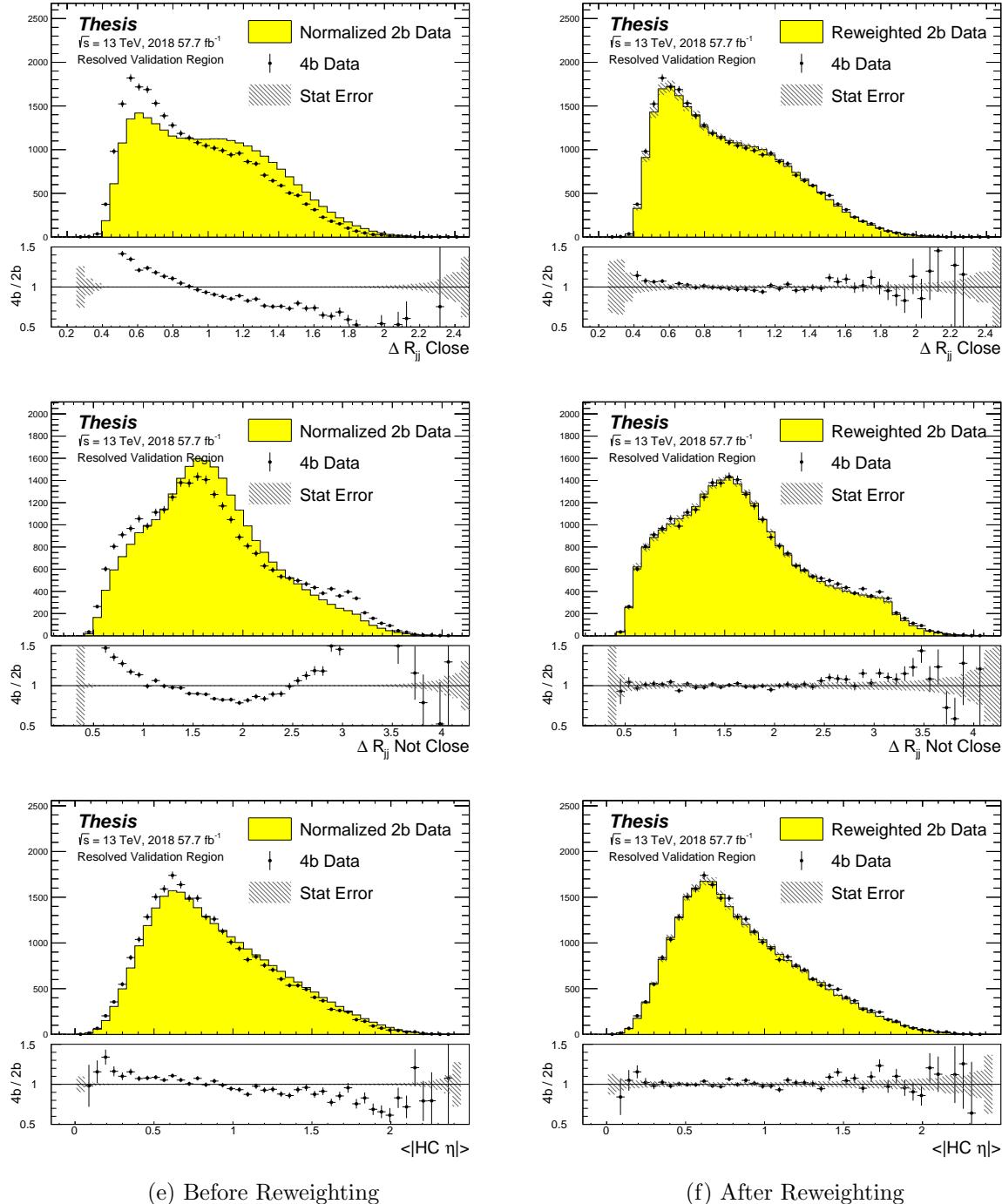


Figure 7.15: **Resonant Search:** Distributions of  $\Delta R$  between the closest Higgs Candidate jets,  $\Delta R$  between the other two, and average absolute value of HC jet  $\eta$  before and after CR derived reweighting for the 2018 Validation Region.

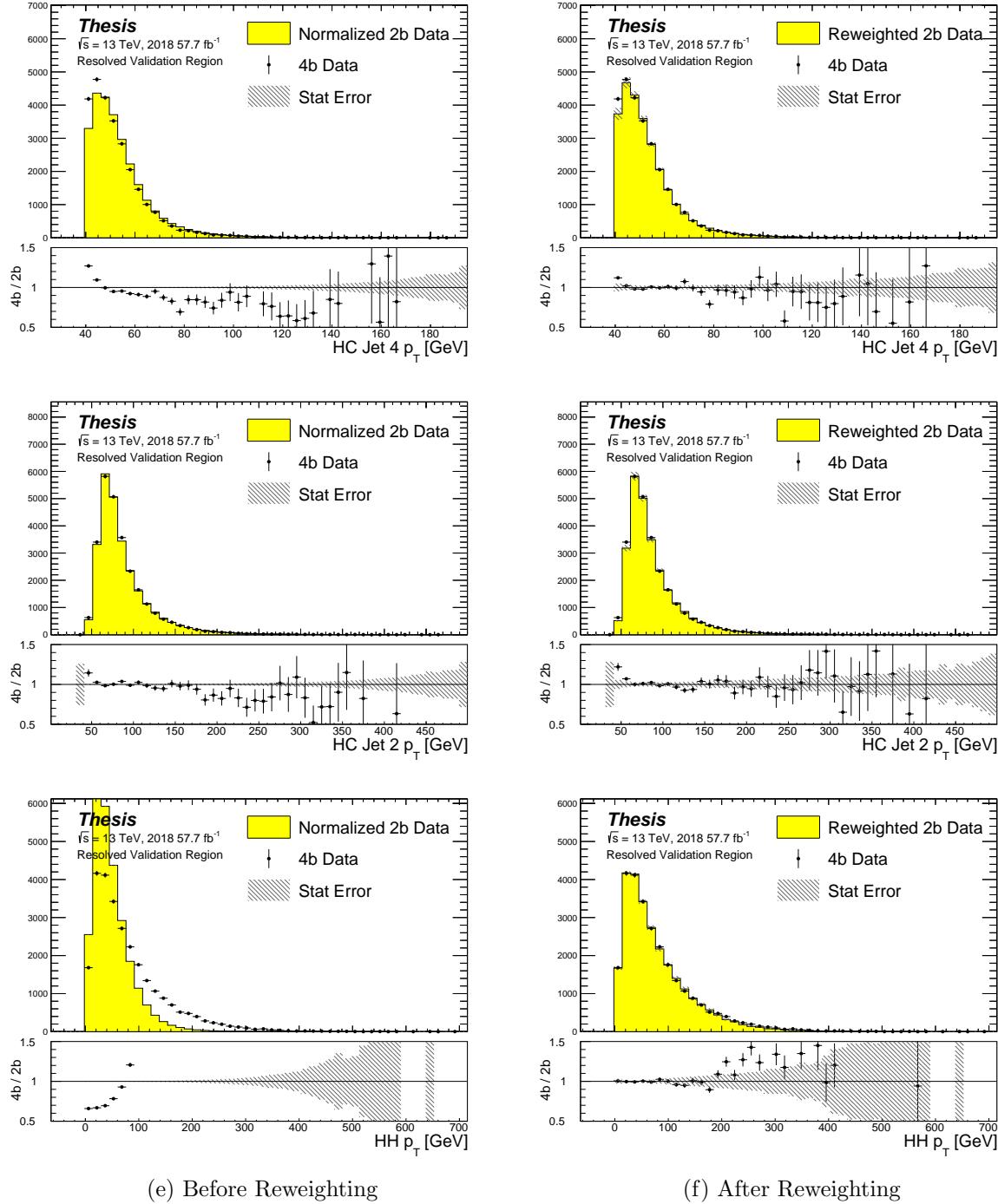


Figure 7.16: **Resonant Search:** Distributions of  $p_T$  of the 2nd and 4th leading Higgs Candidate jets and the  $p_T$  of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.

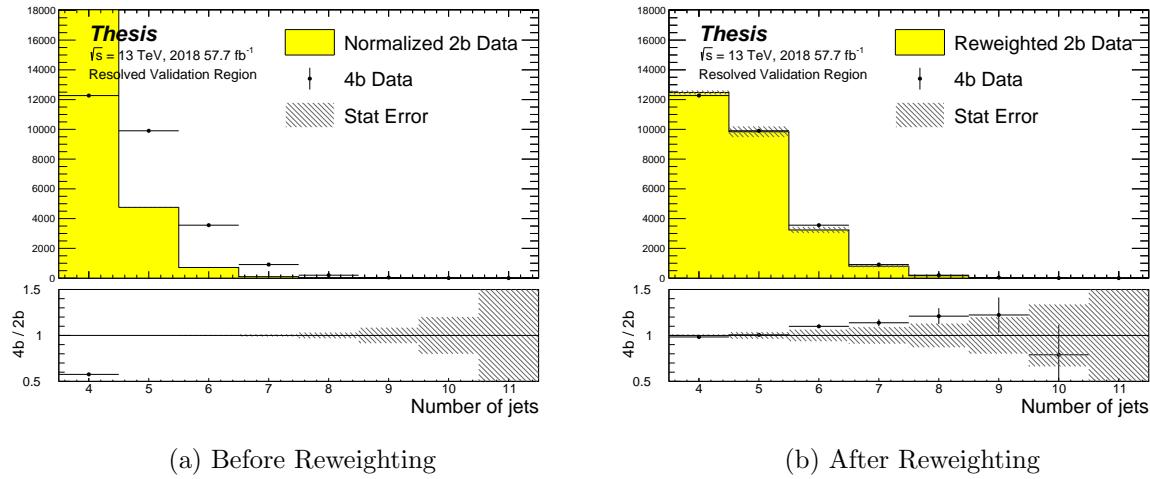


Figure 7.17: **Resonant Search:** Distributions of the number of jets before and after CR derived reweighting for the 2018 Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

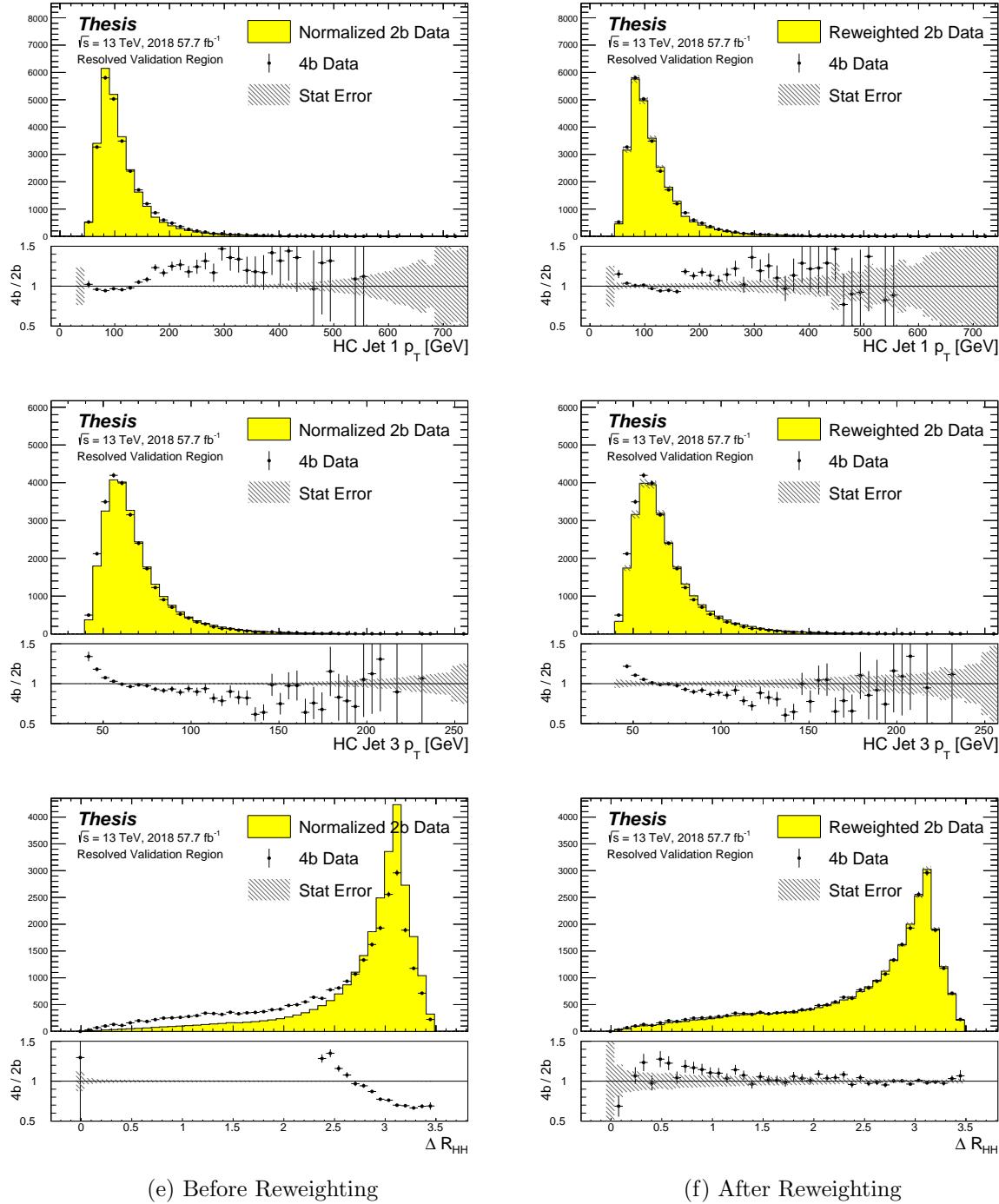


Figure 7.18: **Resonant Search:** Distributions of  $p_T$  of the 1st and 3rd leading Higgs Candidate jets and  $\Delta R$  between Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.

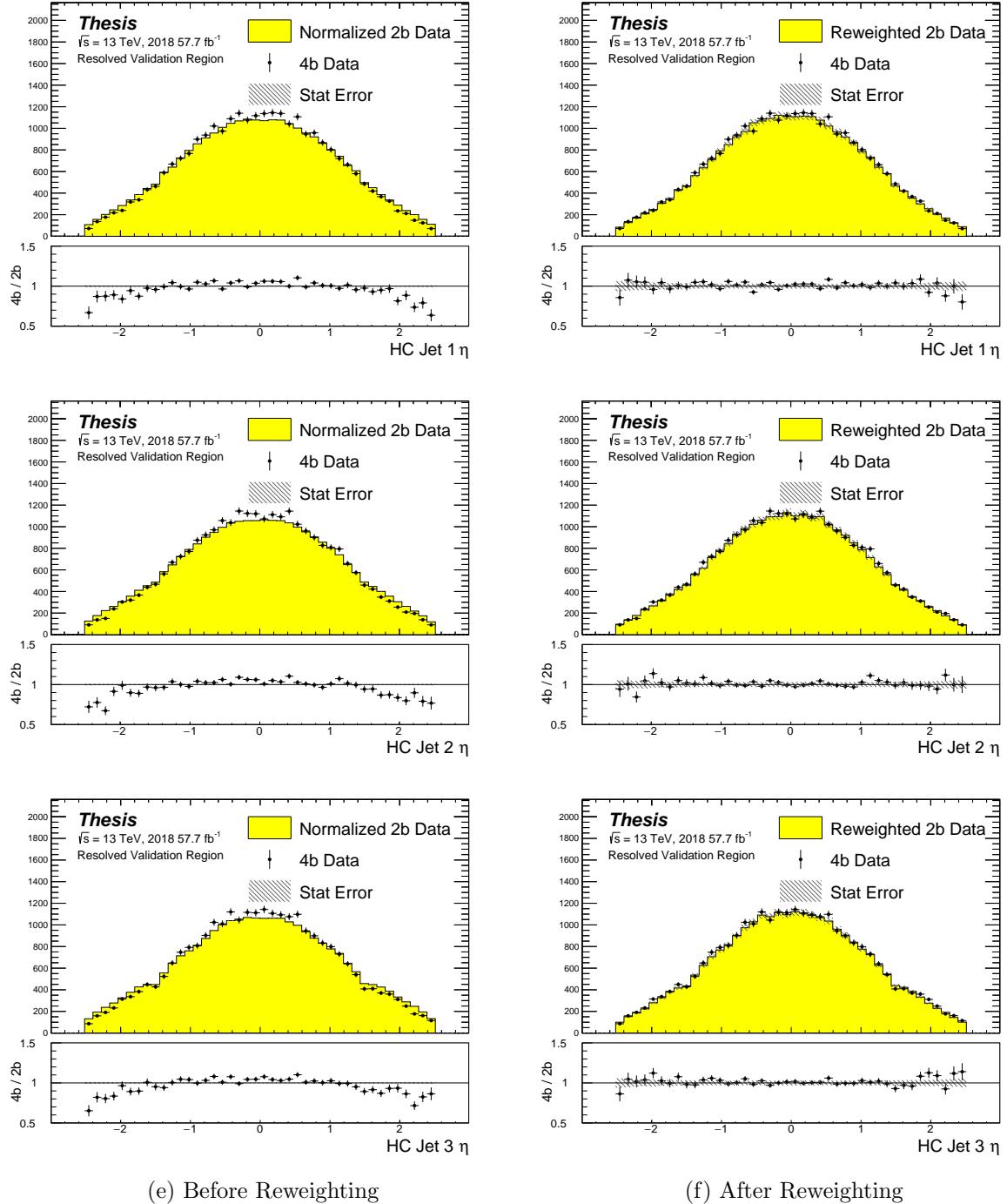


Figure 7.19: **Resonant Search:** Distributions of  $\eta$  of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Validation Region.

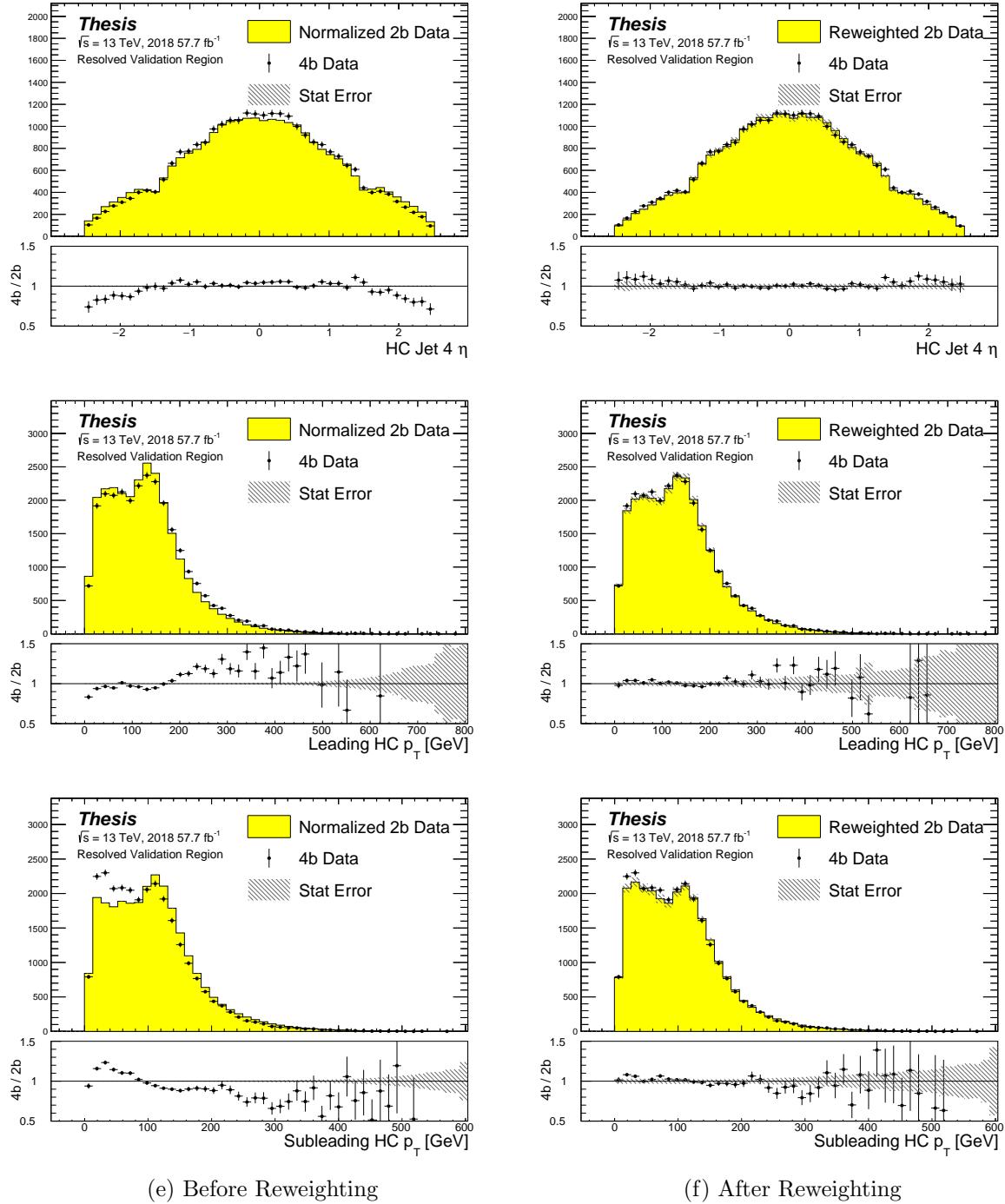


Figure 7.20: **Resonant Search:** Distributions of  $\eta$  of the 4th leading Higgs Candidate jet and the  $p_T$  of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.

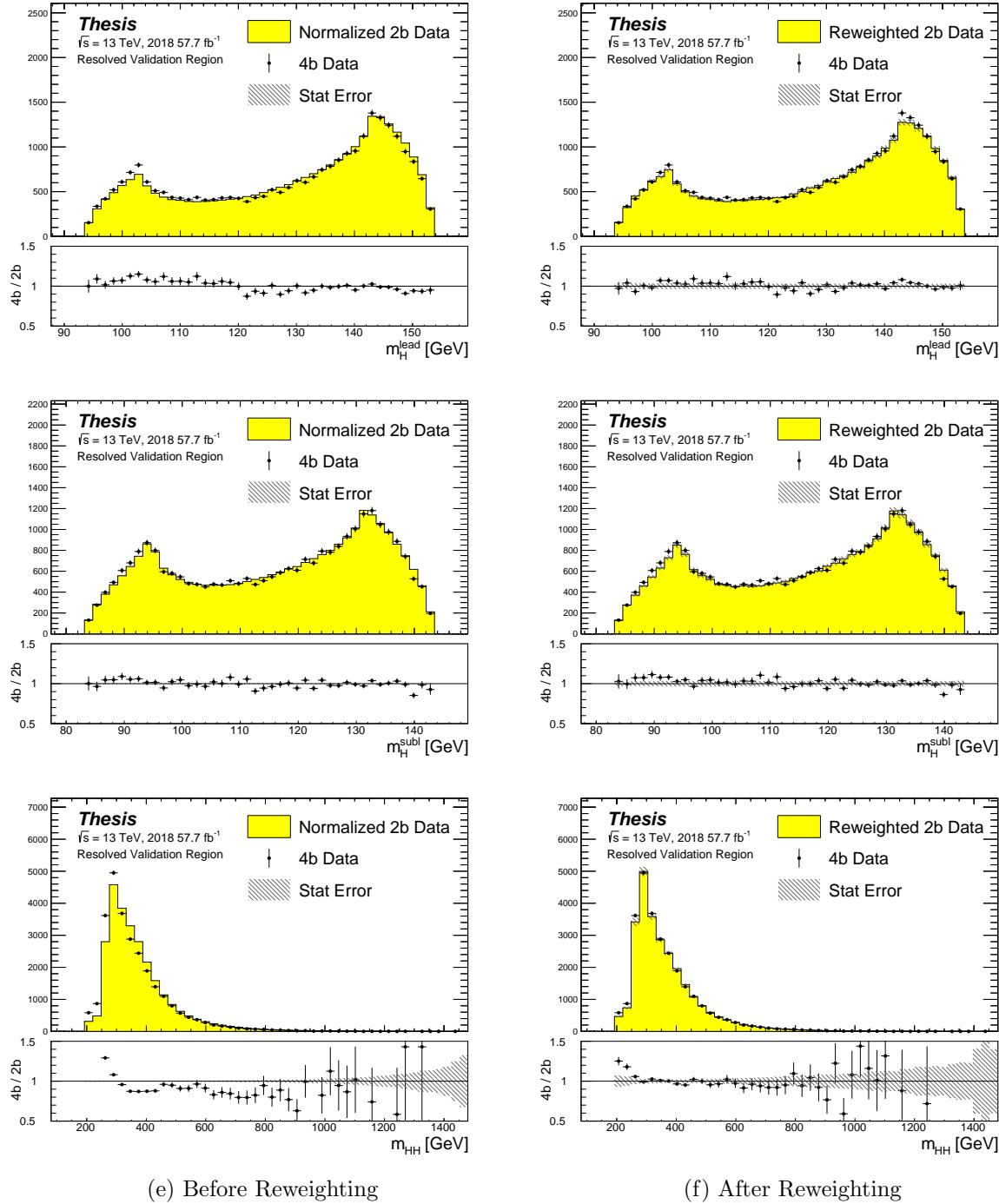


Figure 7.21: **Resonant Search:** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.

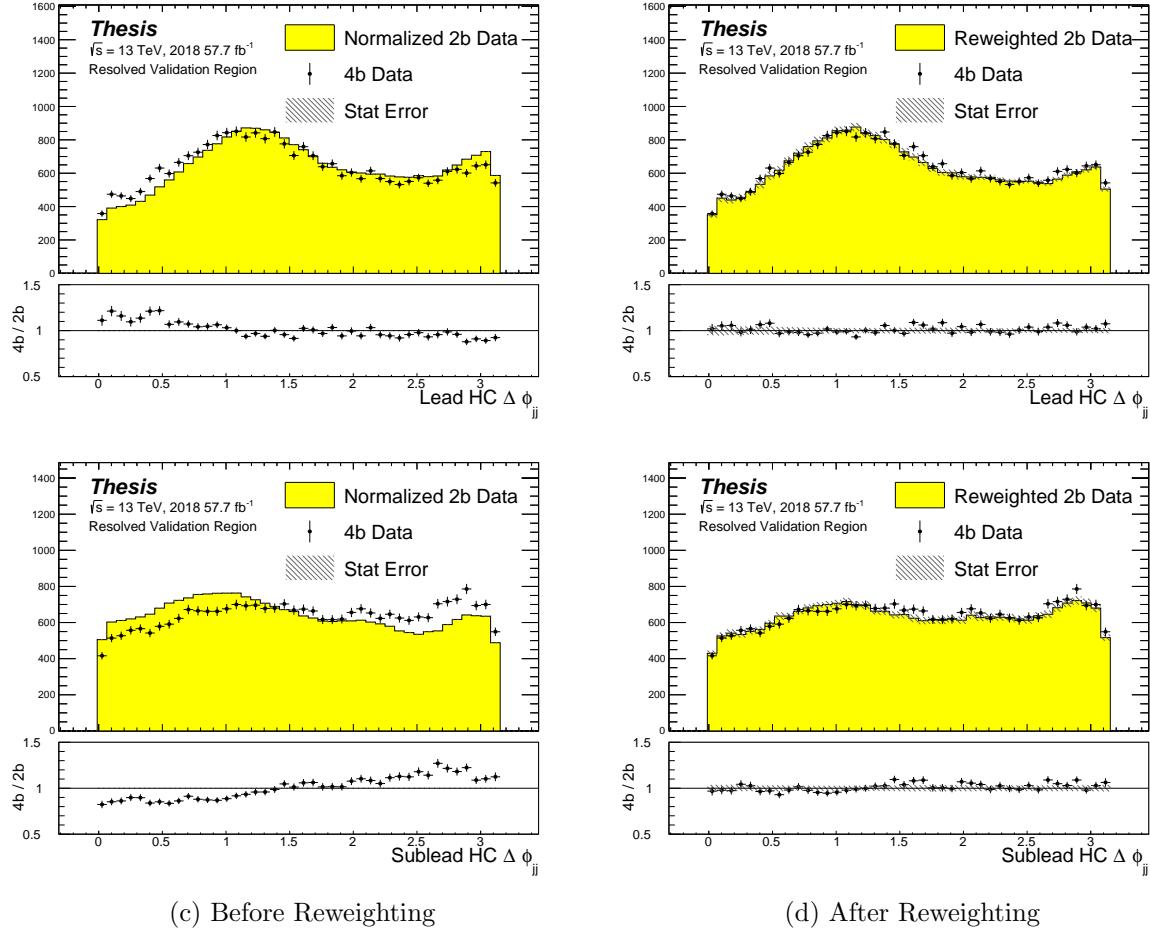


Figure 7.22: **Resonant Search:** Distributions of  $\Delta\phi$  between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.

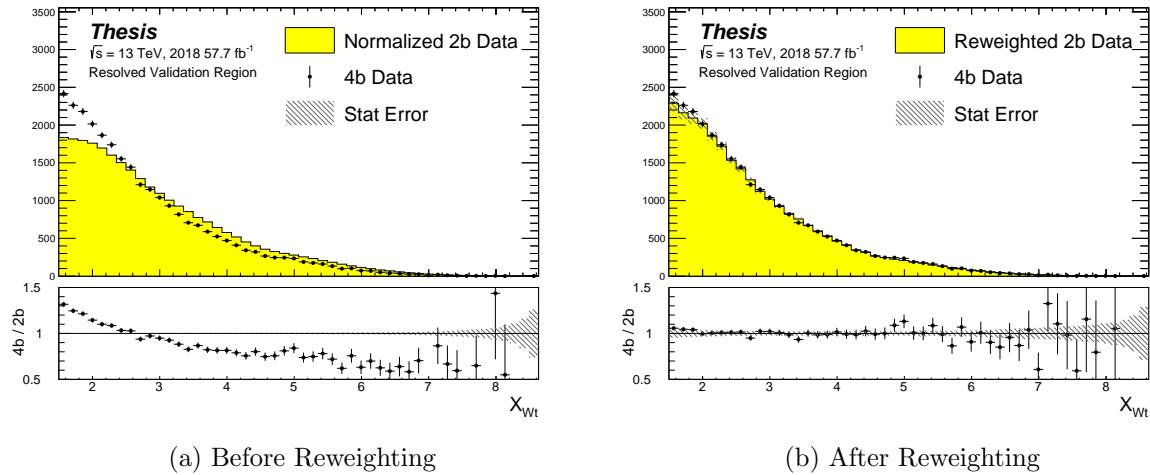


Figure 7.23: **Resonant Search:** Distributions of the top veto variable,  $X_{Wt}$ , before and after CR derived reweighting for the 2018 Validation Region. Reweighting is done after the cut on this variable is applied

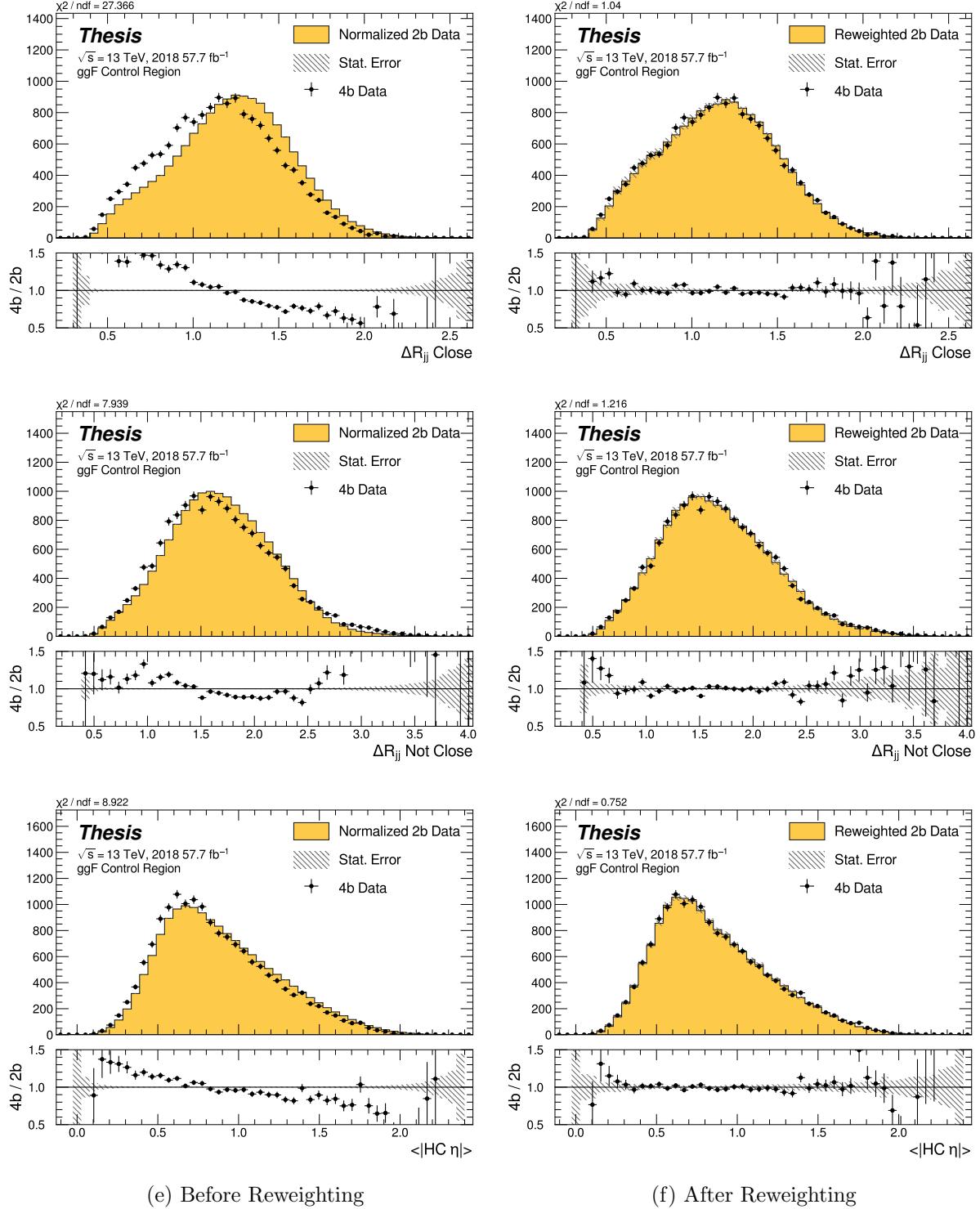


Figure 7.24: **Non-resonant Search (4b):** Distributions of  $\Delta R$  between the closest Higgs Candidate jets,  $\Delta R$  between the other two, and average absolute value of HC jet  $\eta$  before and after CR derived reweighting for the 2018 4b Control Region.

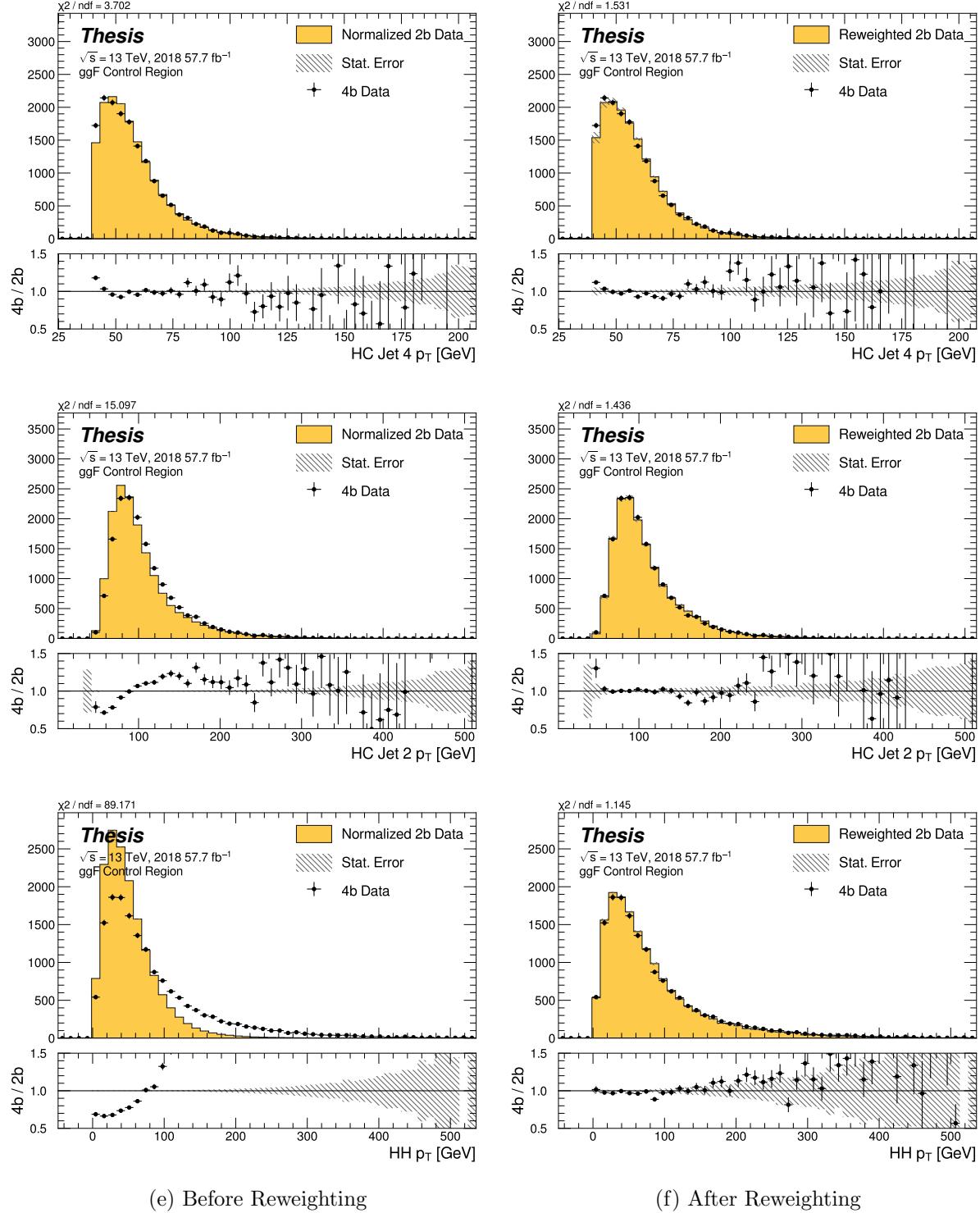


Figure 7.25: **Non-resonant Search (4b):** Distributions of  $p_T$  of the 2nd and 4th leading Higgs Candidate jets and the  $p_T$  of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.

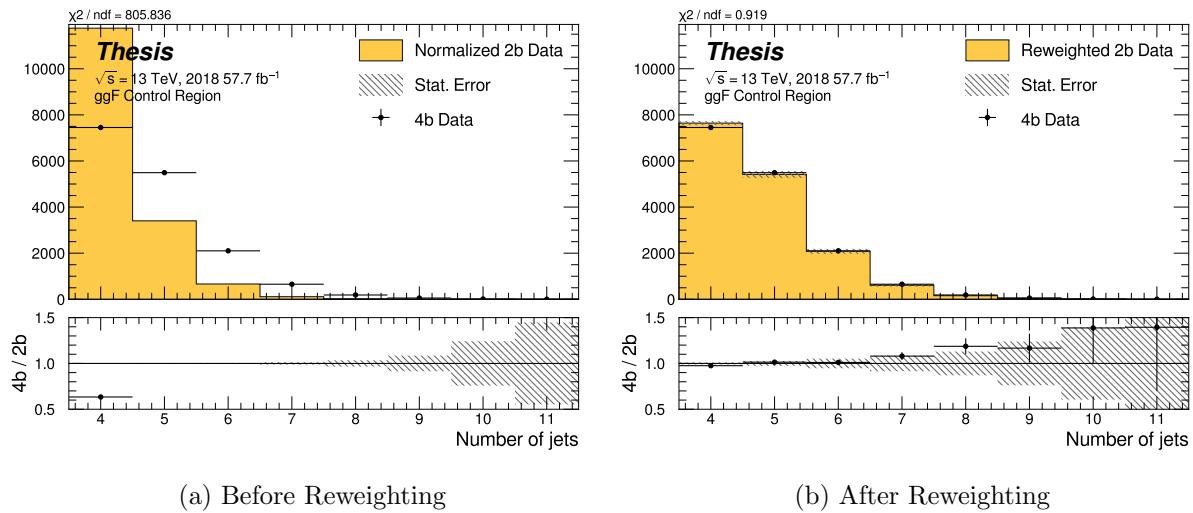


Figure 7.26: **Non-resonant Search (4b):** Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

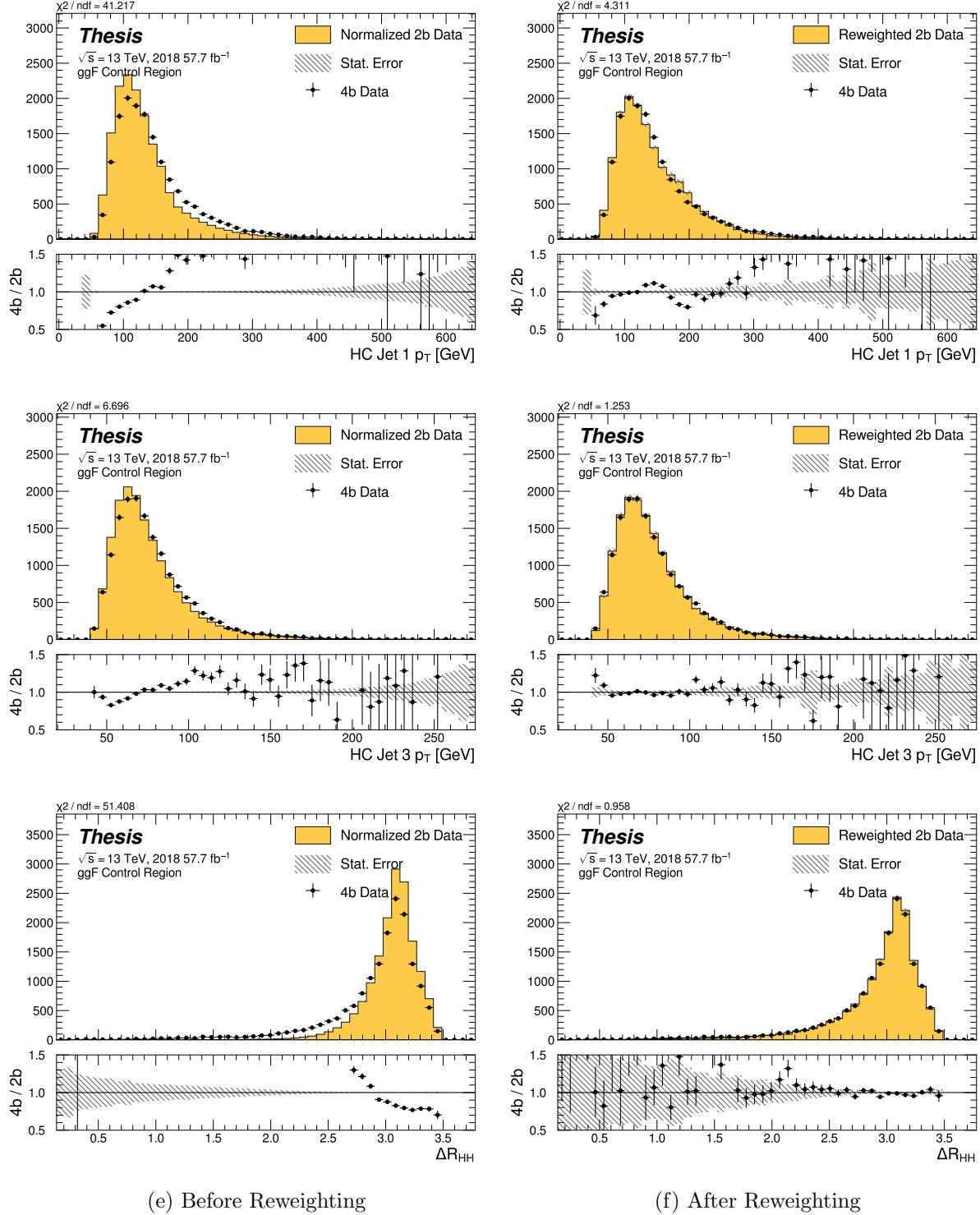


Figure 7.27: **Non-resonant Search (4b):** Distributions of  $p_T$  of the 1st and 3rd leading Higgs Candidate jets and  $\Delta R$  between Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.

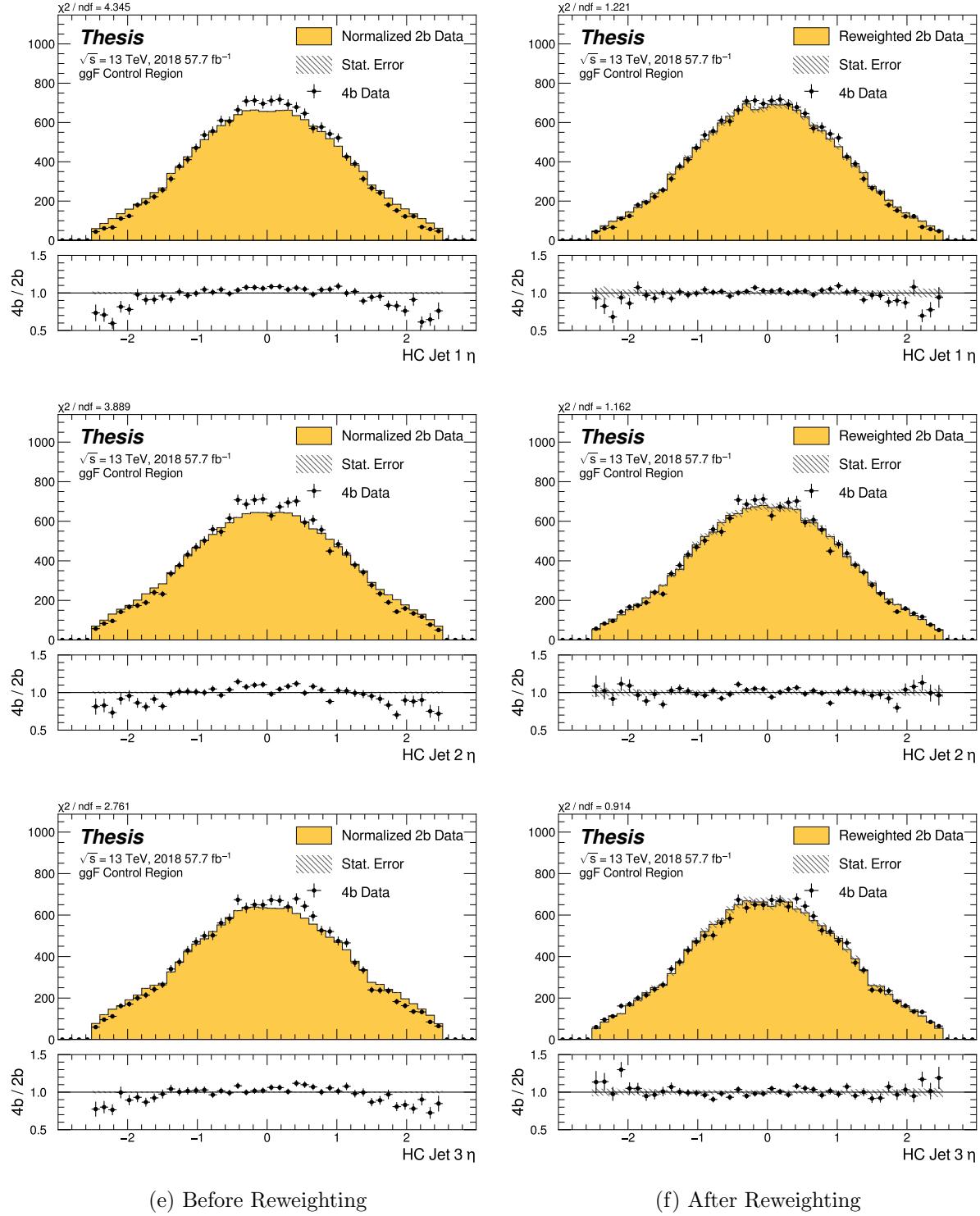


Figure 7.28: **Non-resonant Search (4b):** Distributions of  $\eta$  of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Control Region.

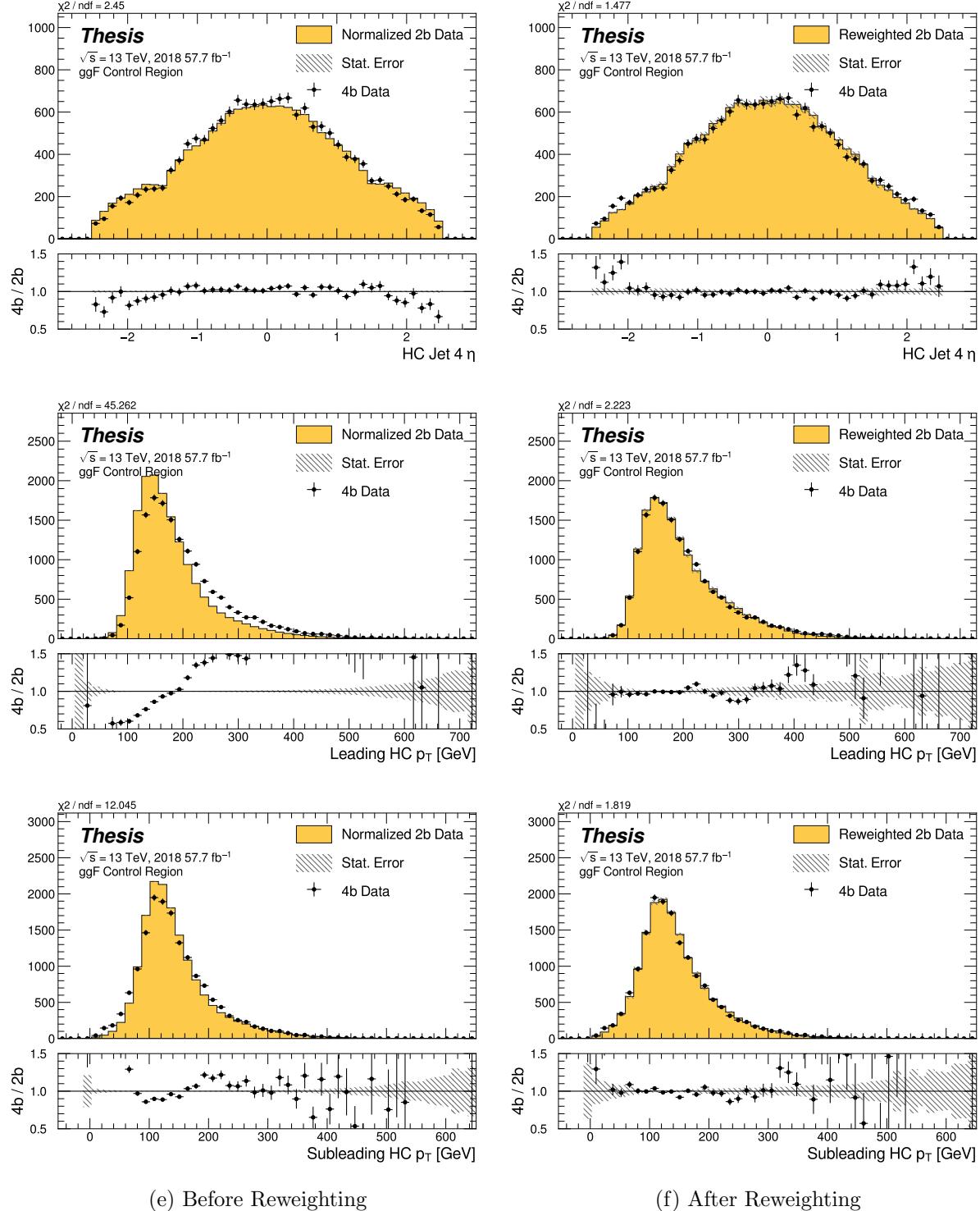


Figure 7.29: **Non-resonant Search (4b):** Distributions of  $\eta$  of the 4th leading Higgs Candidate jet and the  $p_T$  of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.

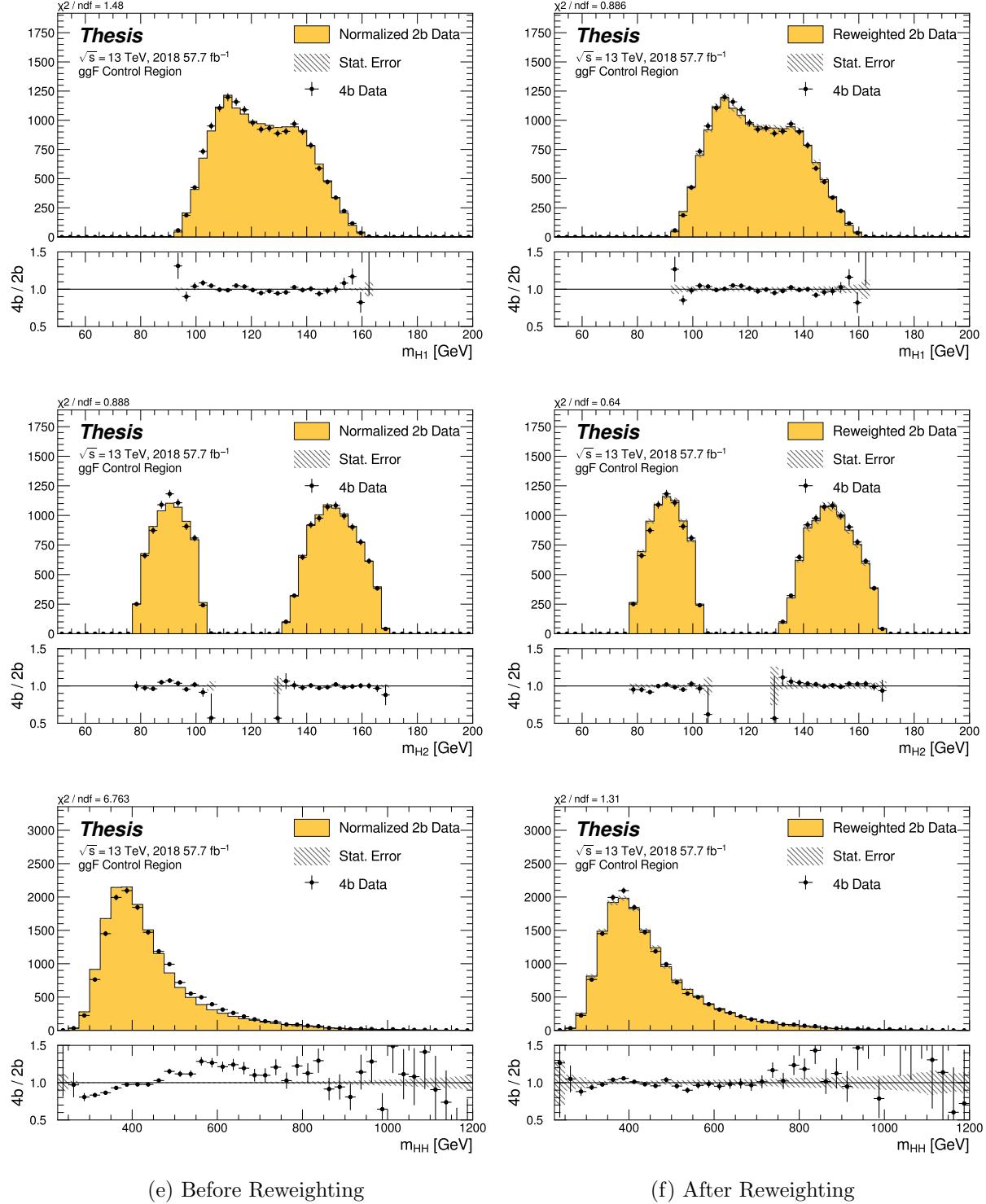


Figure 7.30: **Non-resonant Search (4b):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.

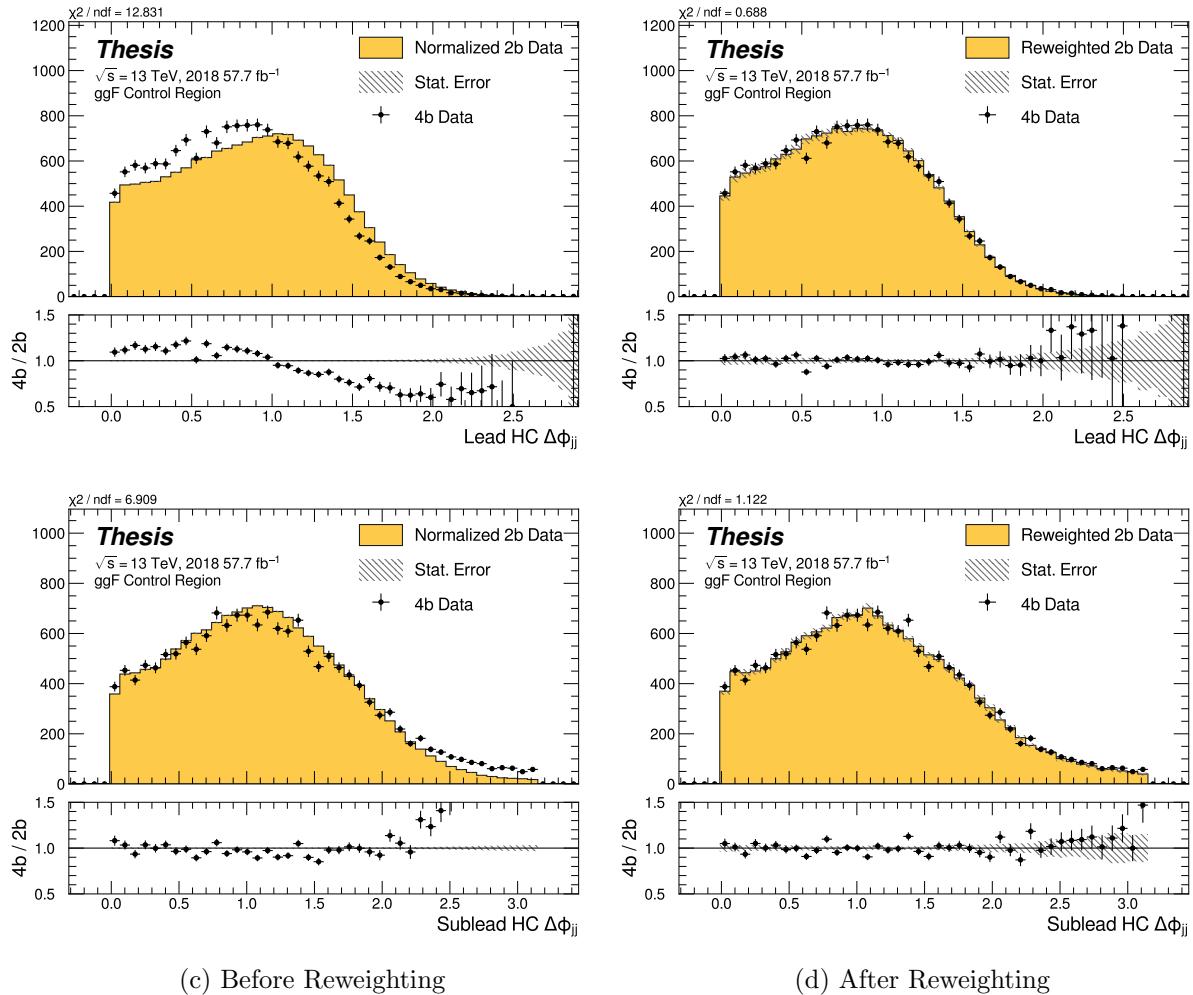


Figure 7.31: **Non-resonant Search (4b):** Distributions of  $\Delta\phi$  between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.

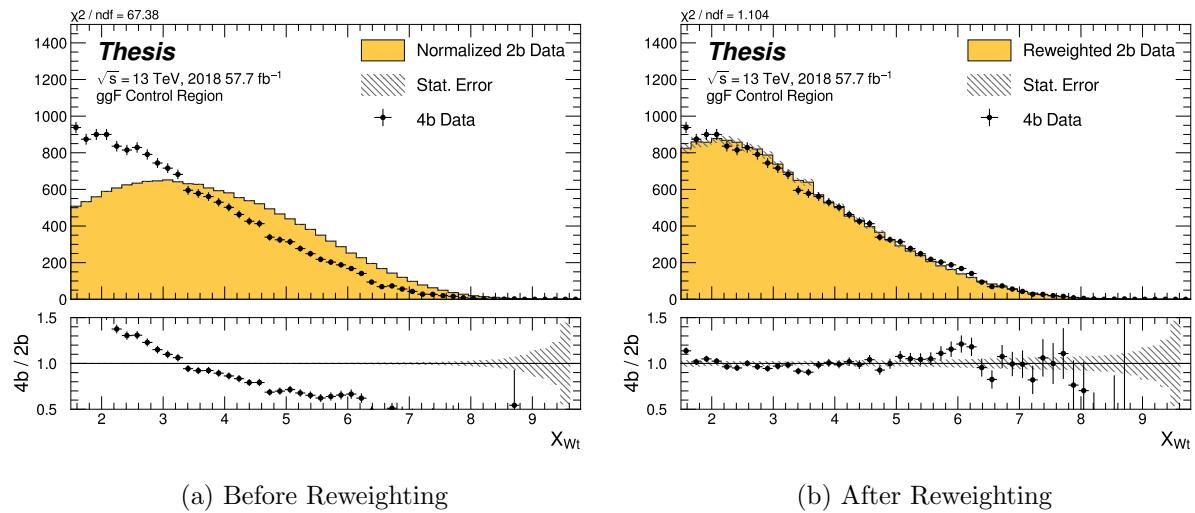


Figure 7.32: **Non-resonant Search (4b)**: Distributions of the top veto variable,  $X_{Wt}$ , before and after CR derived reweighting for the 2018 4b Control Region. Reweighting is done after the cut on this variable is applied.

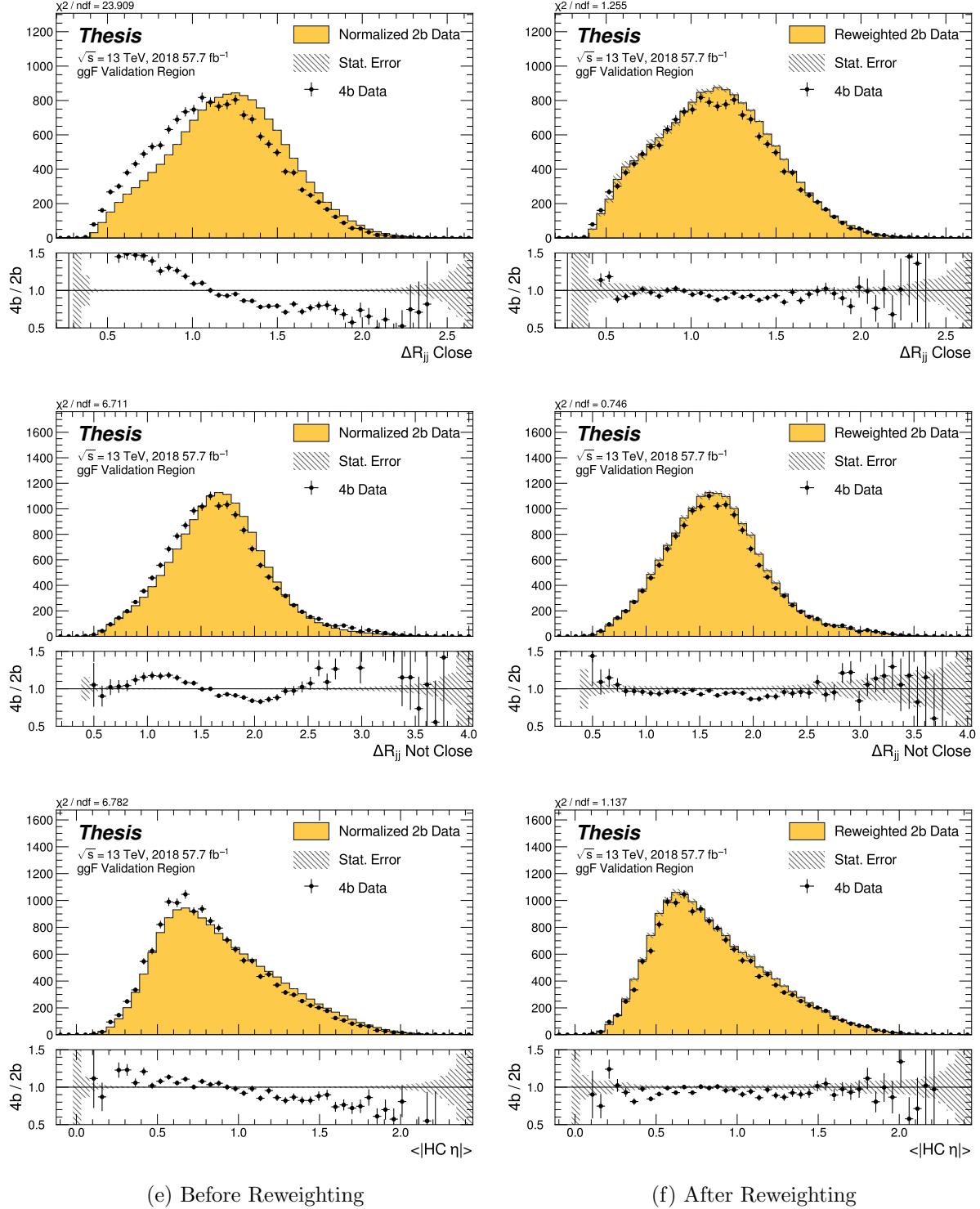


Figure 7.33: **Non-resonant Search (4b):** Distributions of  $\Delta R$  between the closest Higgs Candidate jets,  $\Delta R$  between the other two, and average absolute value of HC jet  $\eta$  before and after CR derived reweighting for the 2018 4b Validation Region.

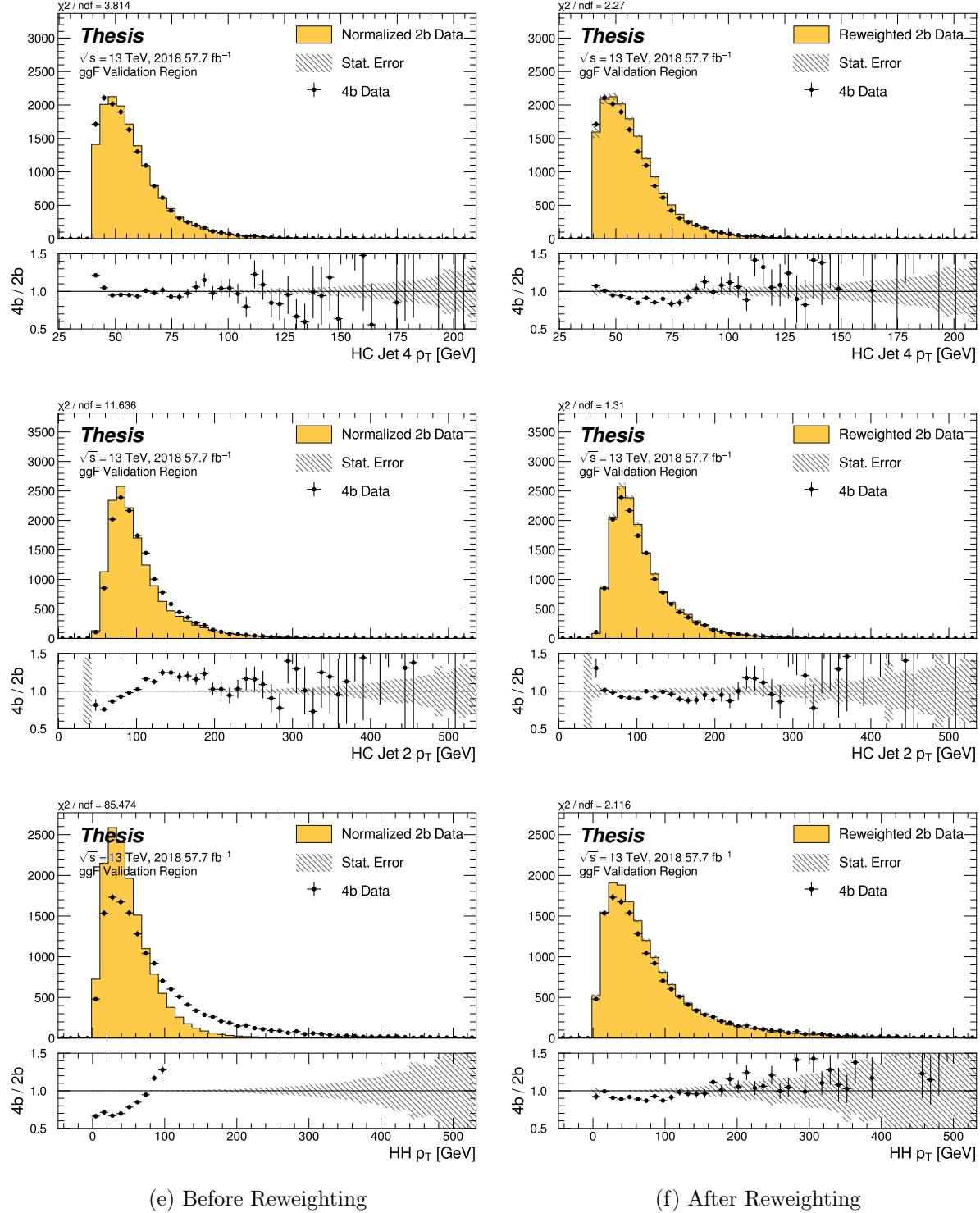


Figure 7.34: **Non-resonant Search (4b):** Distributions of  $p_T$  of the 2nd and 4th leading Higgs Candidate jets and the  $p_T$  of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.

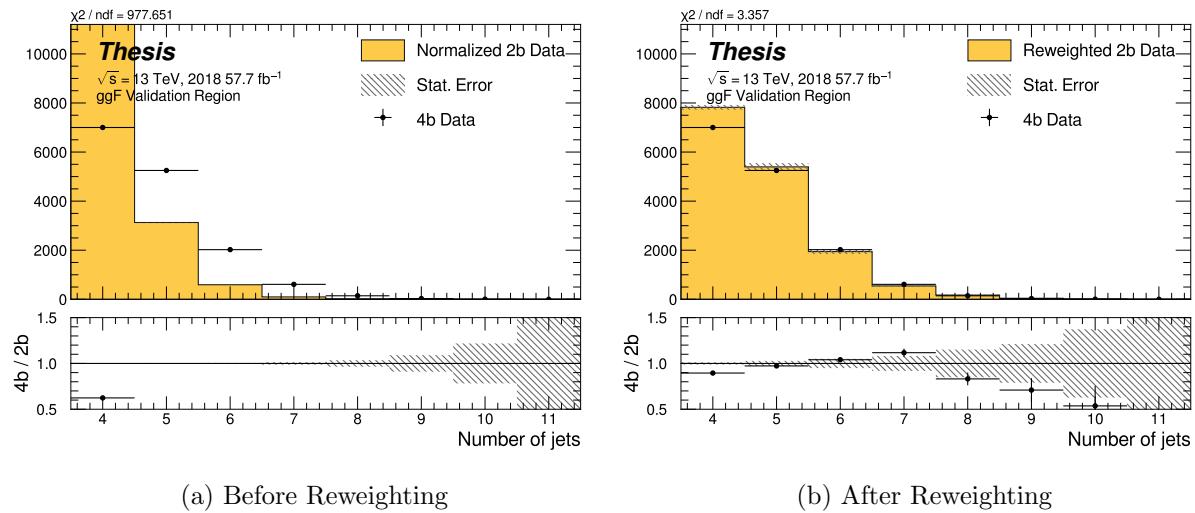


Figure 7.35: **Non-resonant Search (4b):** Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

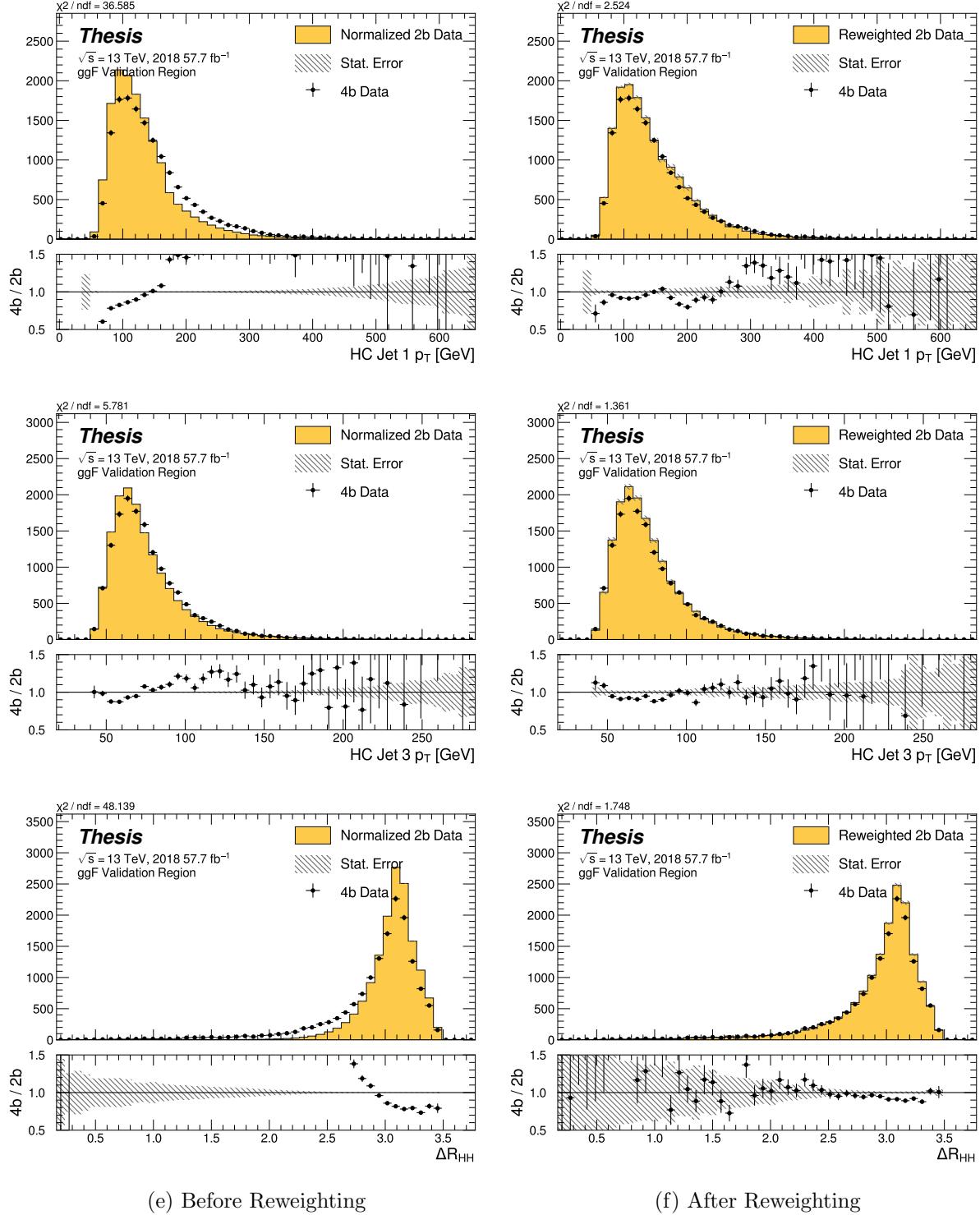


Figure 7.36: **Non-resonant Search (4b):** Distributions of  $p_T$  of the 1st and 3rd leading Higgs Candidate jets and  $\Delta R$  between Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.

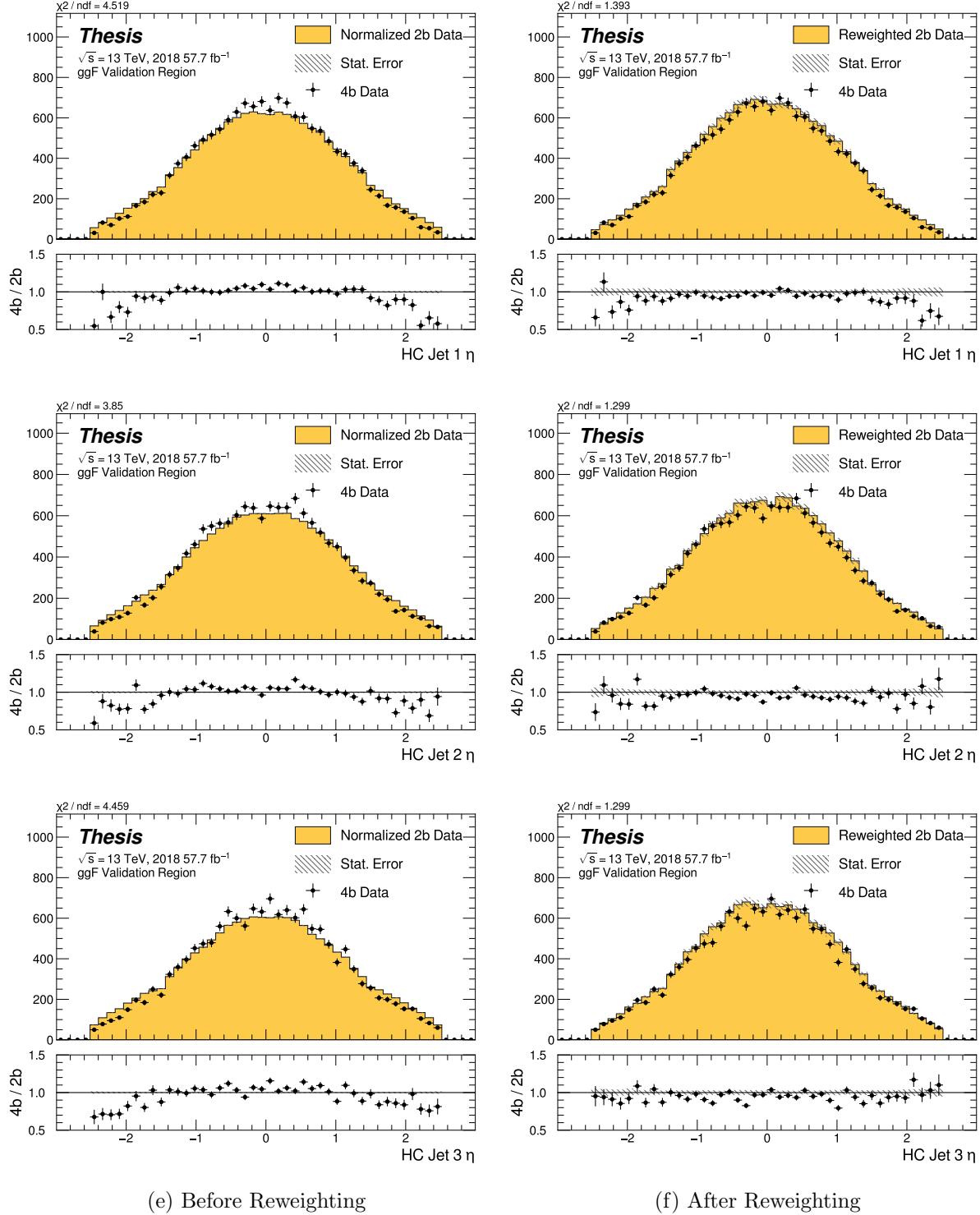


Figure 7.37: **Non-resonant Search (4b):** Distributions of  $\eta$  of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Validation Region.

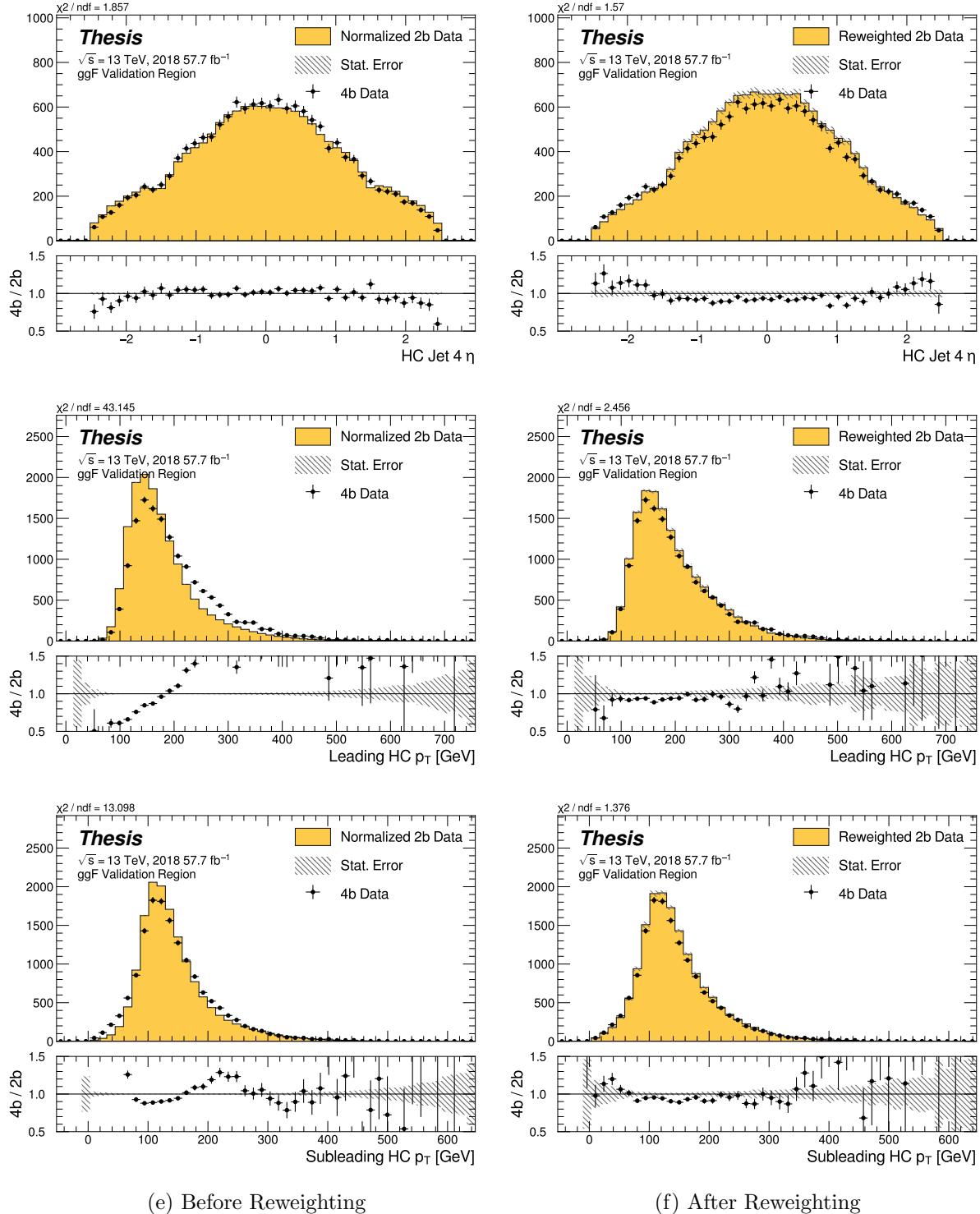


Figure 7.38: **Non-resonant Search (4b):** Distributions of  $\eta$  of the 4th leading Higgs Candidate jet and the  $p_T$  of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.

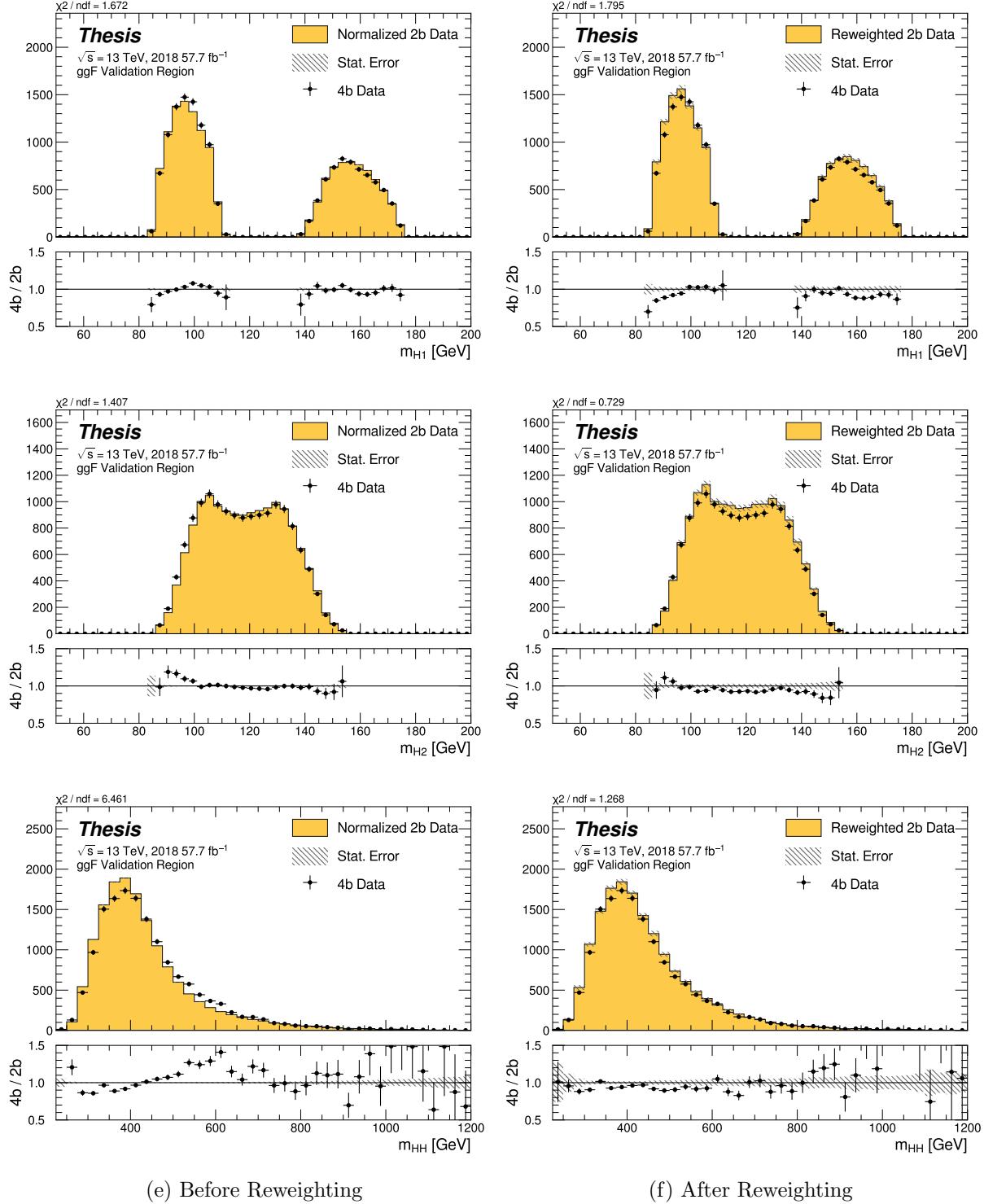


Figure 7.39: **Non-resonant Search (4b):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.

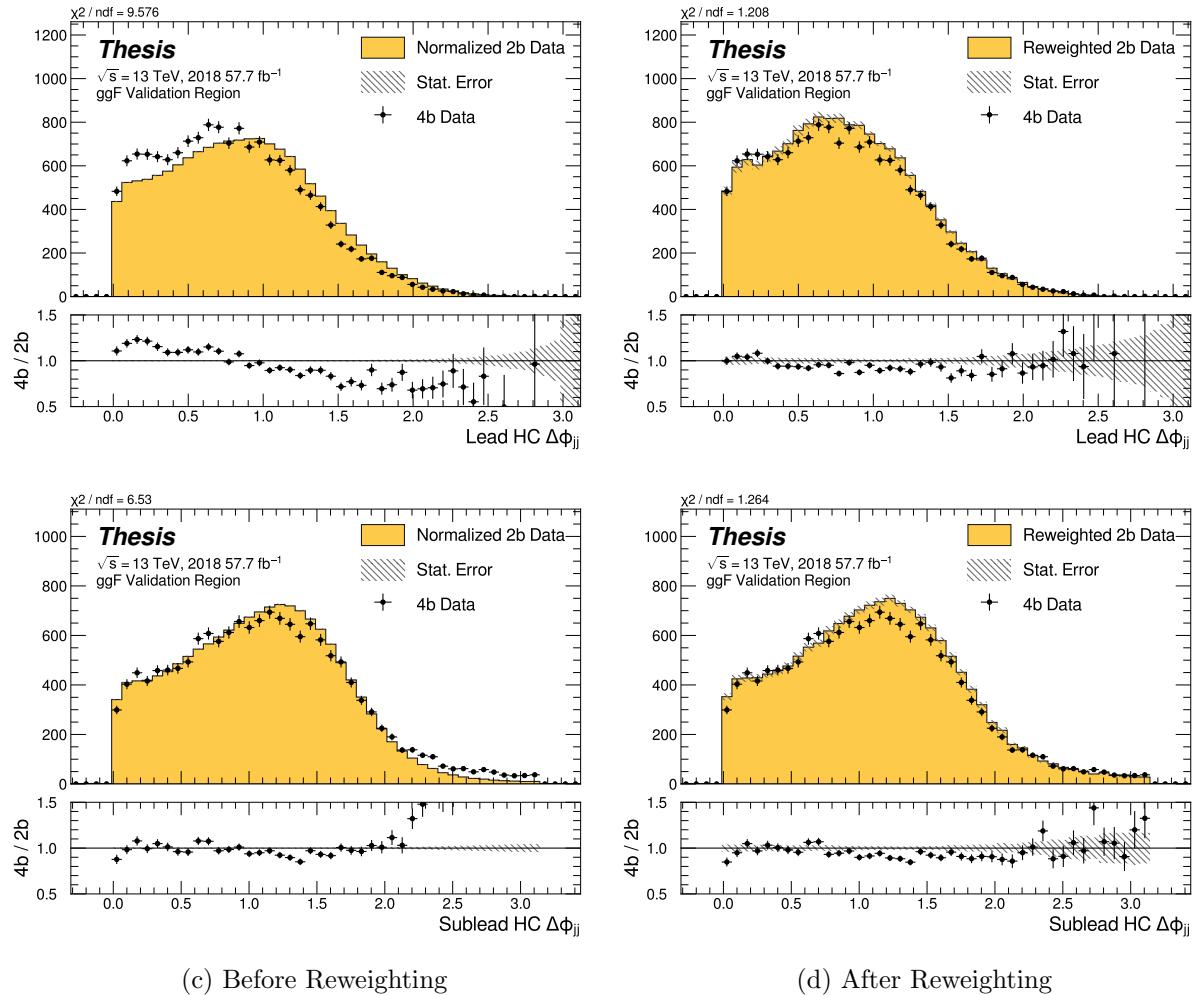


Figure 7.40: **Non-resonant Search (4b):** Distributions of  $\Delta\phi$  between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.

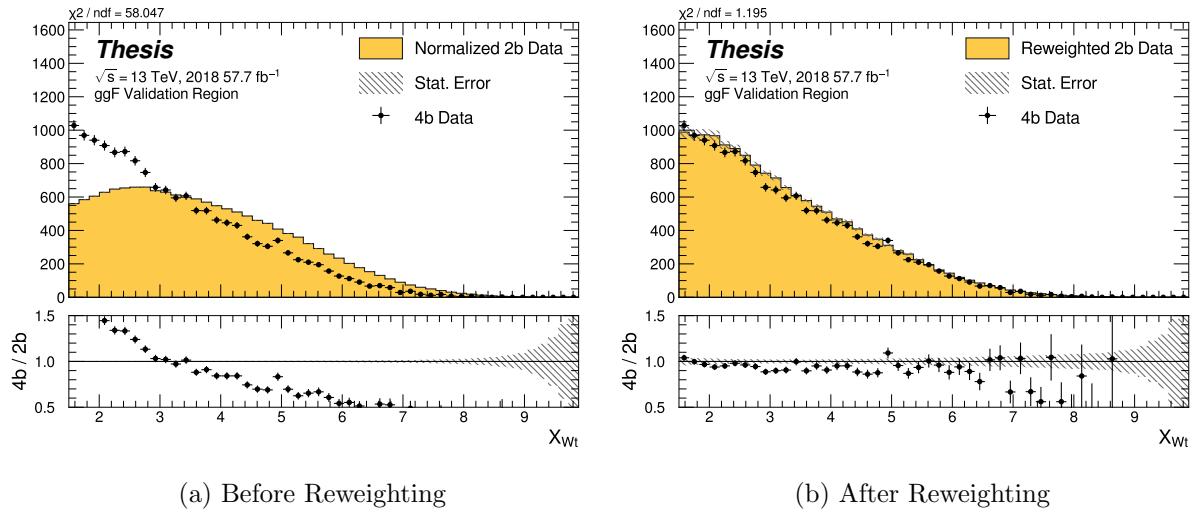


Figure 7.41: **Non-resonant Search (4b)**: Distributions of the top veto variable,  $X_{Wt}$ , before and after CR derived reweighting for the 2018 4b Validation Region. Reweighting is done after the cut on this variable is applied.

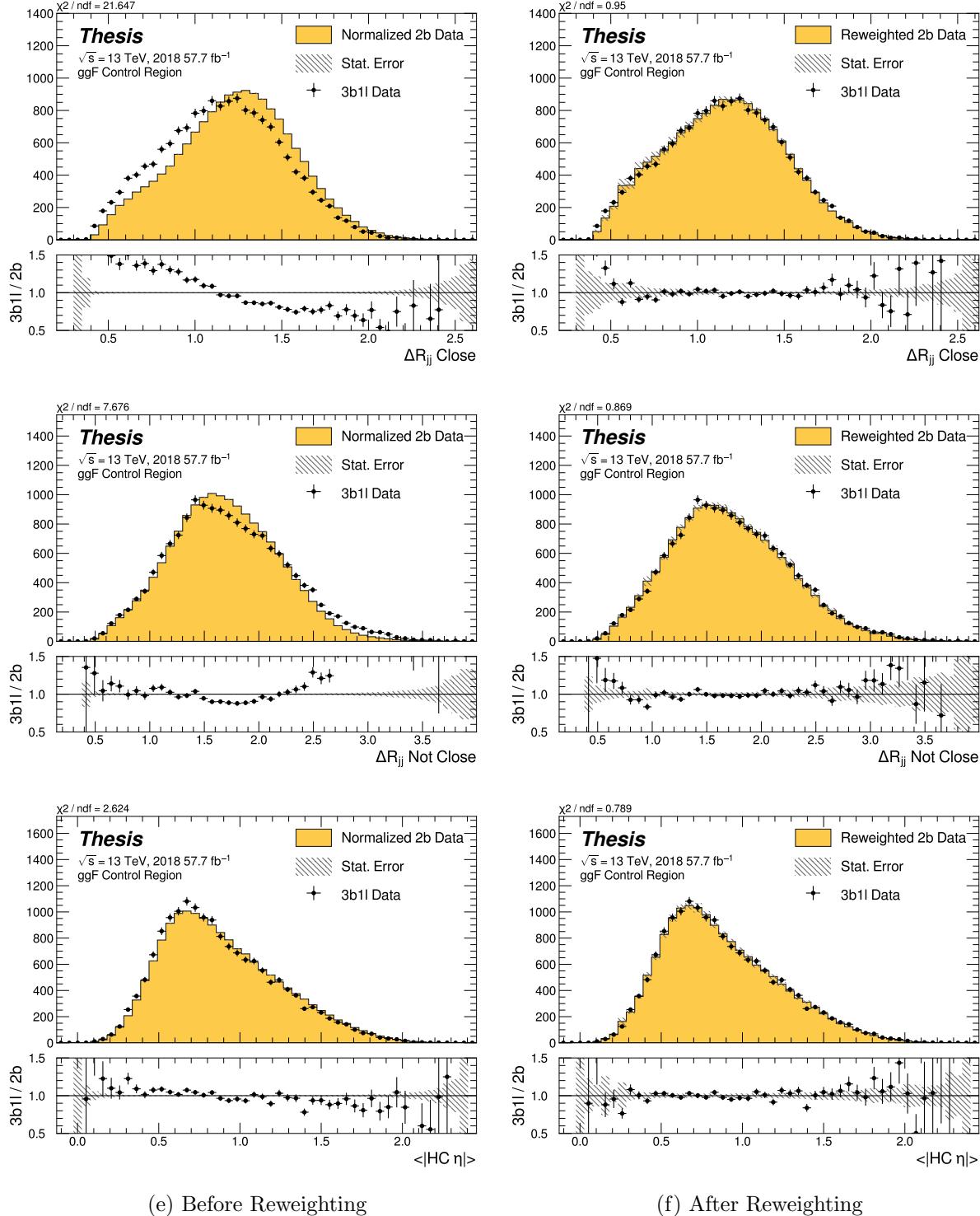


Figure 7.42: **Non-resonant Search (3b1l):** Distributions of  $\Delta R$  between the closest Higgs Candidate jets,  $\Delta R$  between the other two, and average absolute value of HC jet  $\eta$  before and after CR derived reweighting for the 2018 3b1l Control Region.

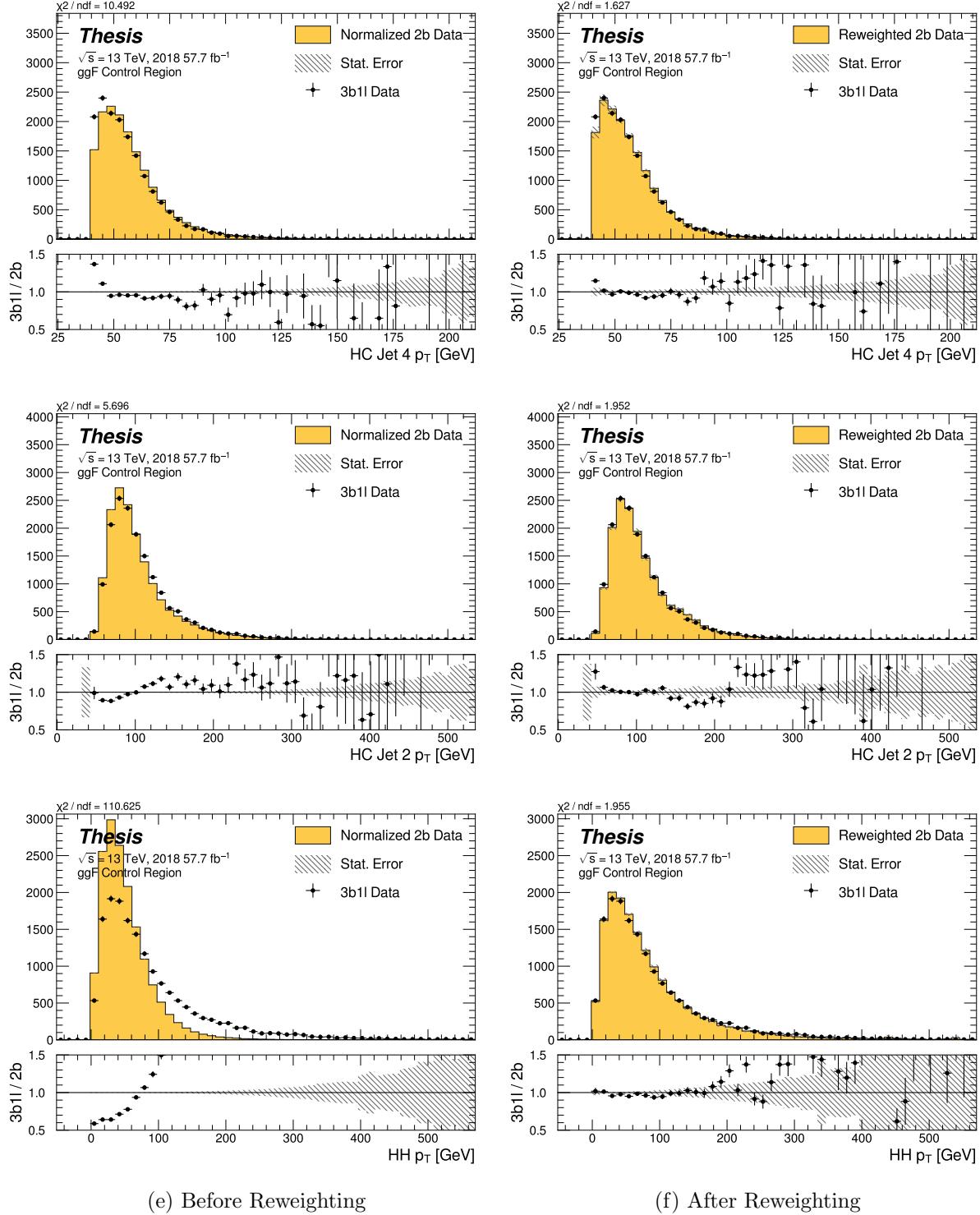


Figure 7.43: **Non-resonant Search (3b1l):** Distributions of  $p_T$  of the 2nd and 4th leading Higgs Candidate jets and the  $p_T$  of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Control Region.

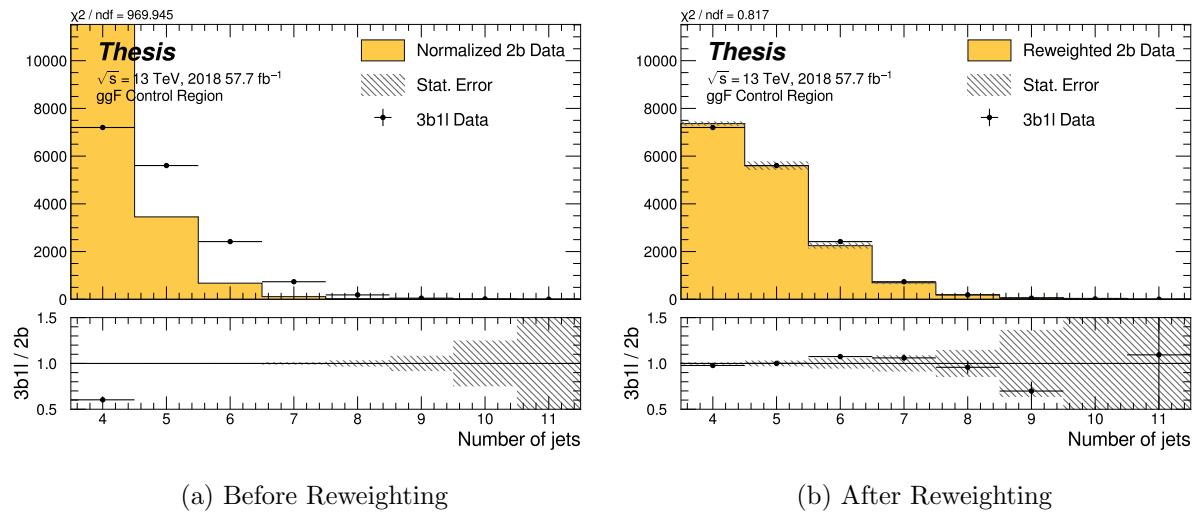


Figure 7.44: **Non-resonant Search (3b1l):** Distributions of the number of jets before and after CR derived reweighting for the 2018 3b1l Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

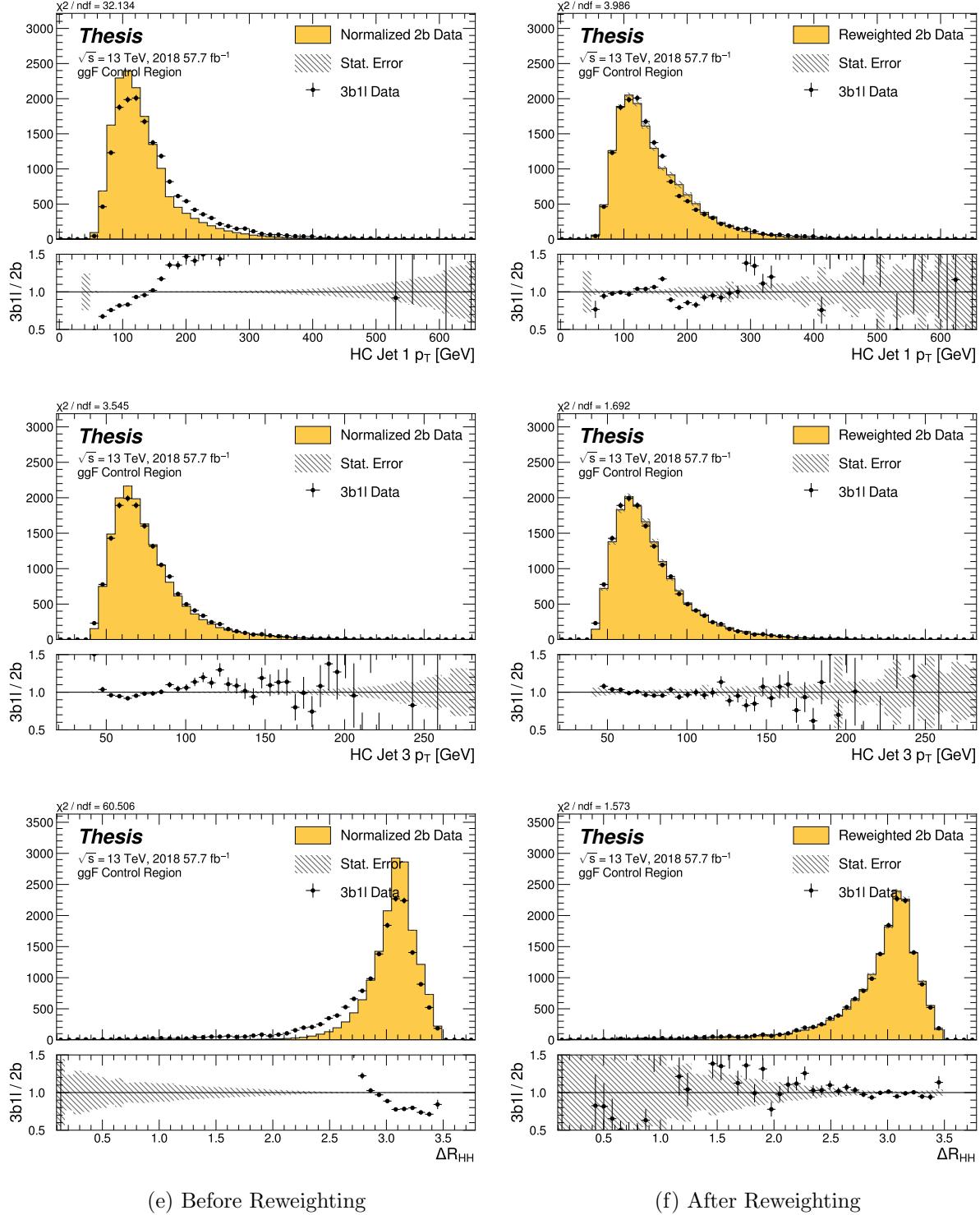


Figure 7.45: **Non-resonant Search (3b1l):** Distributions of  $p_T$  of the 1st and 3rd leading Higgs Candidate jets and  $\Delta R$  between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.

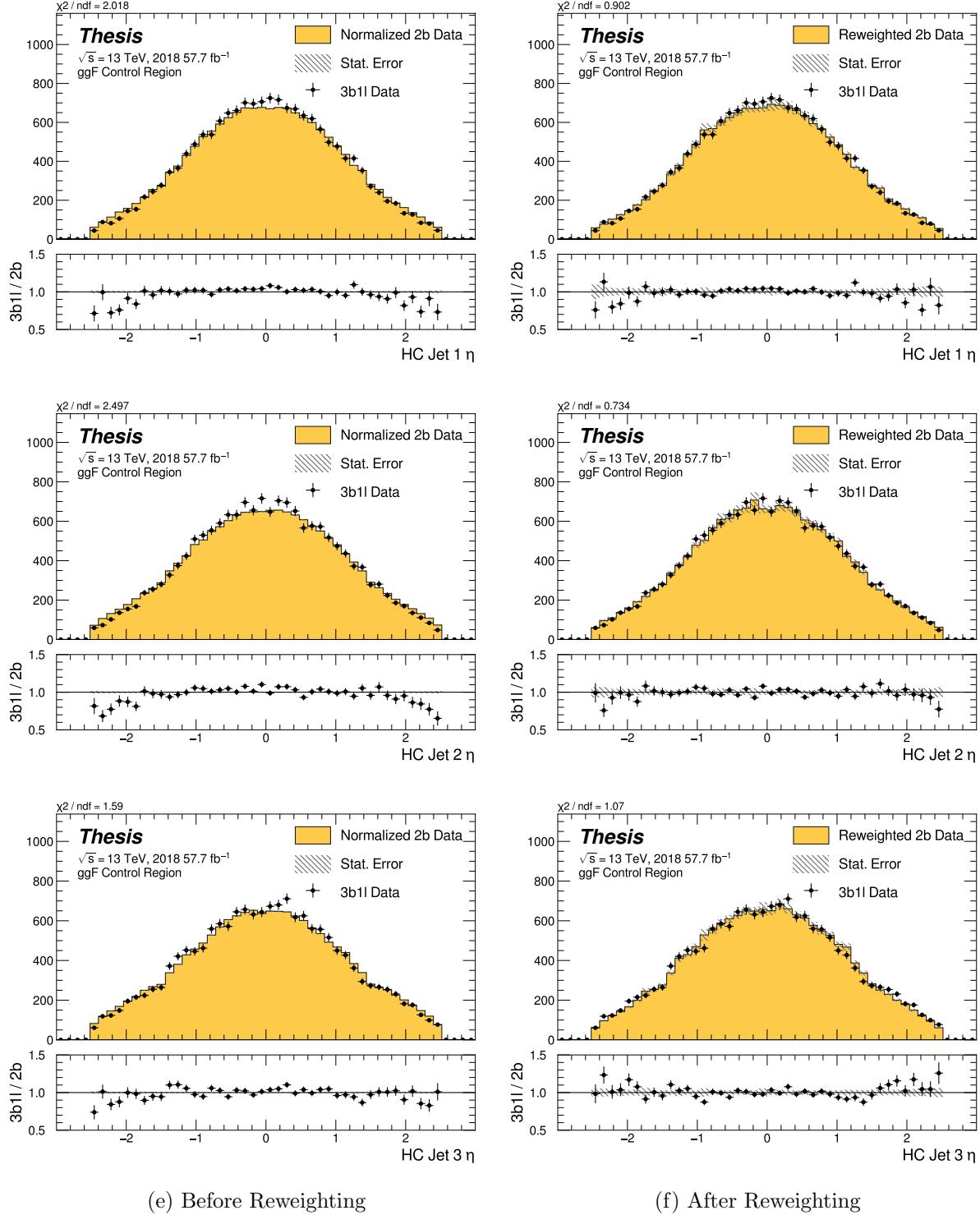


Figure 7.46: **Non-resonant Search (3b1l):** Distributions of  $\eta$  of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Control Region.

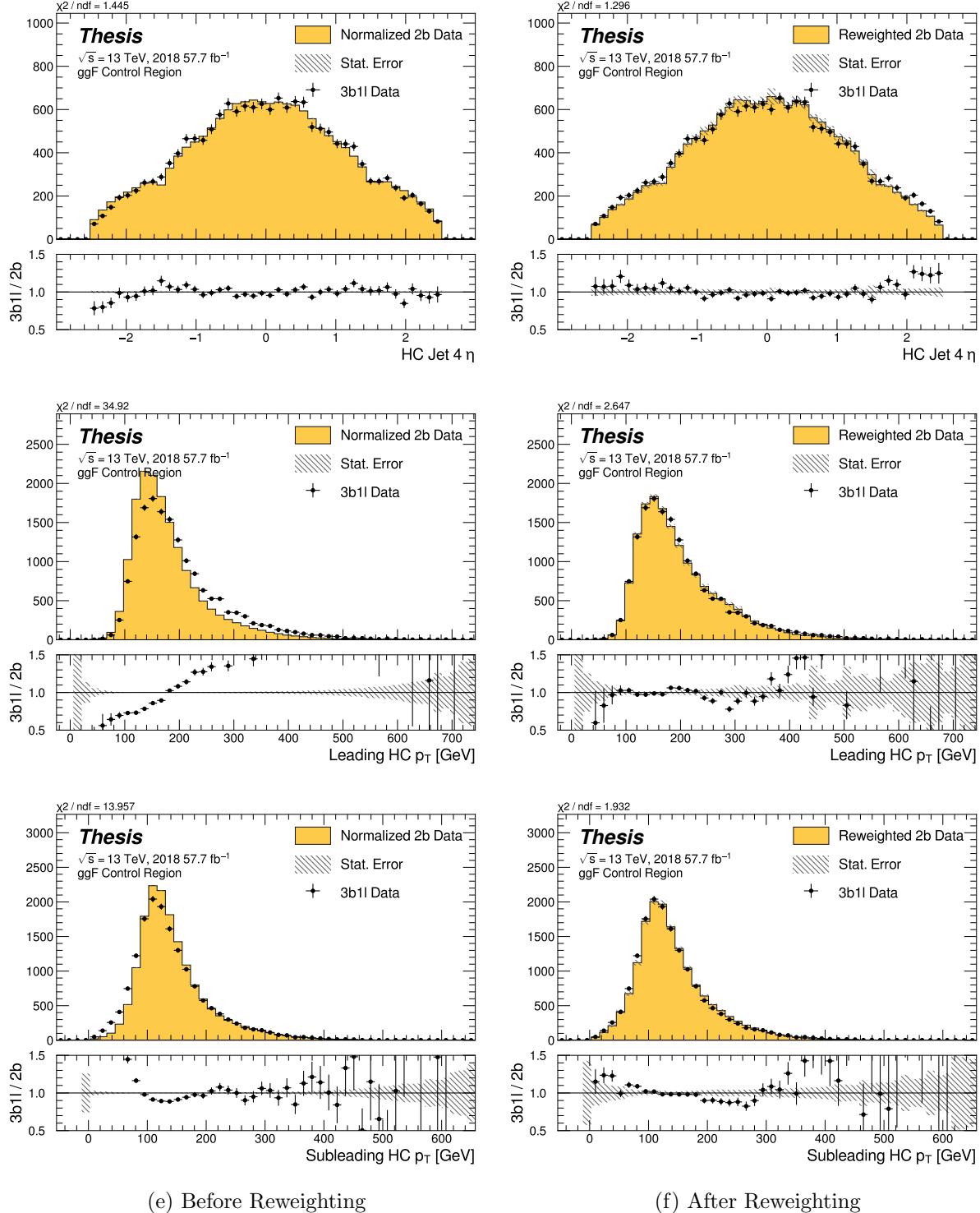


Figure 7.47: **Non-resonant Search (3b1l):** Distributions of  $\eta$  of the 4th leading Higgs Candidate jet and the  $p_T$  of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.

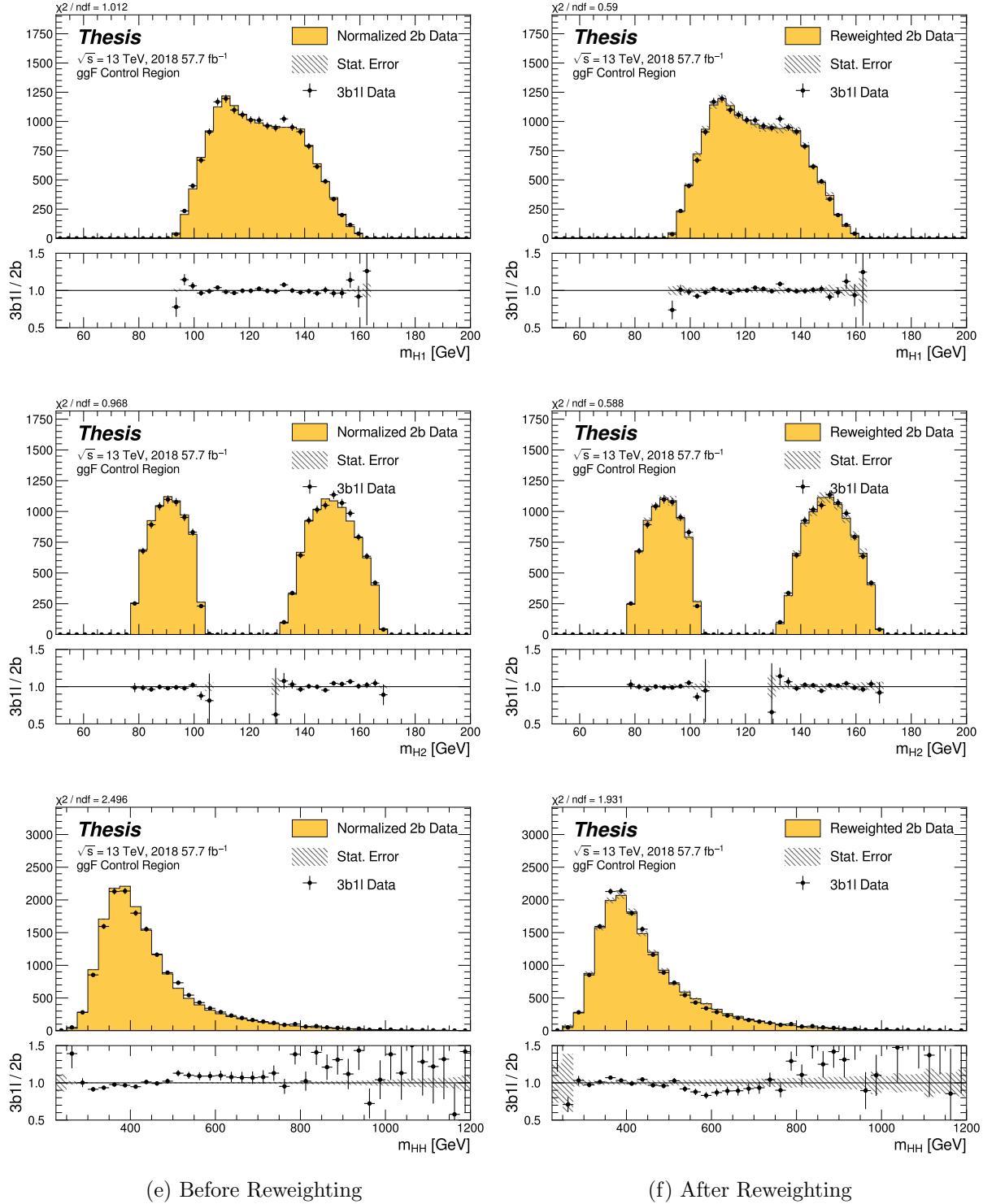


Figure 7.48: **Non-resonant Search (3b1l):** Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Control Region.

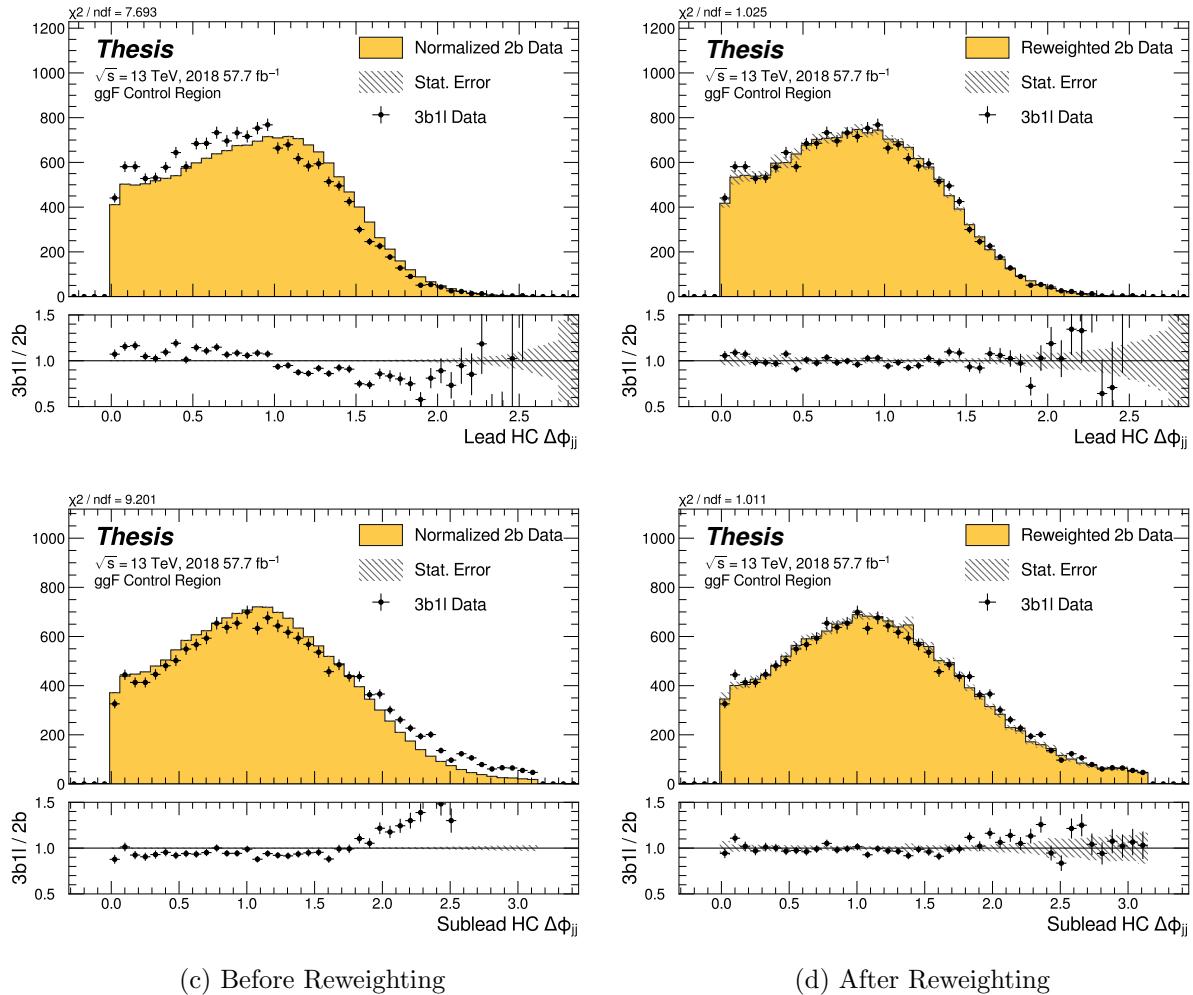


Figure 7.49: **Non-resonant Search (3b1l):** Distributions of  $\Delta\phi$  between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.

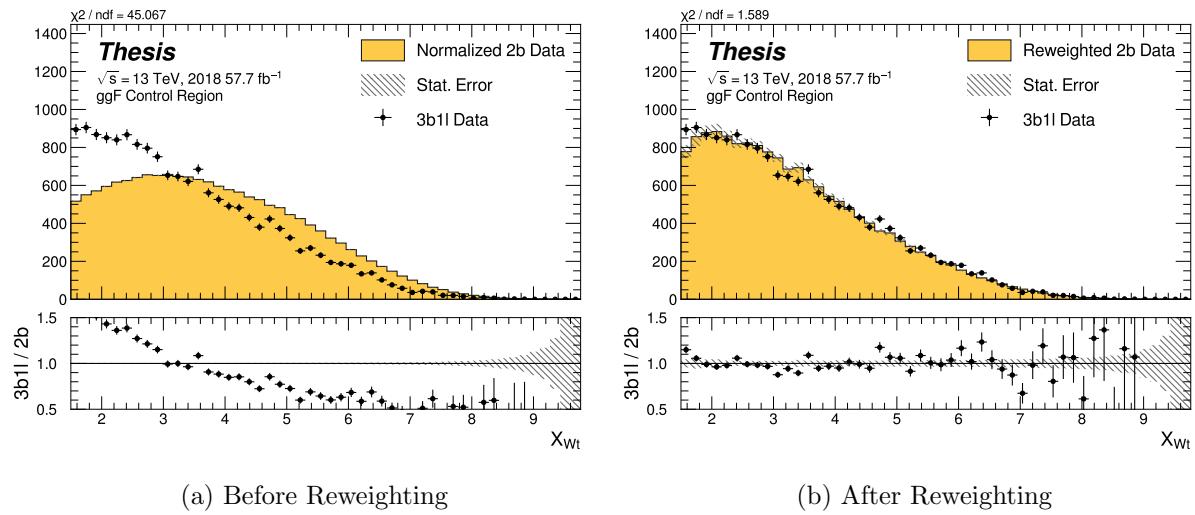


Figure 7.50: **Non-resonant Search (3b1l):** Distributions of the top veto variable,  $X_{Wt}$ , before and after CR derived reweighting for the 2018 3b1l Control Region. Reweighting is done after the cut on this variable is applied.

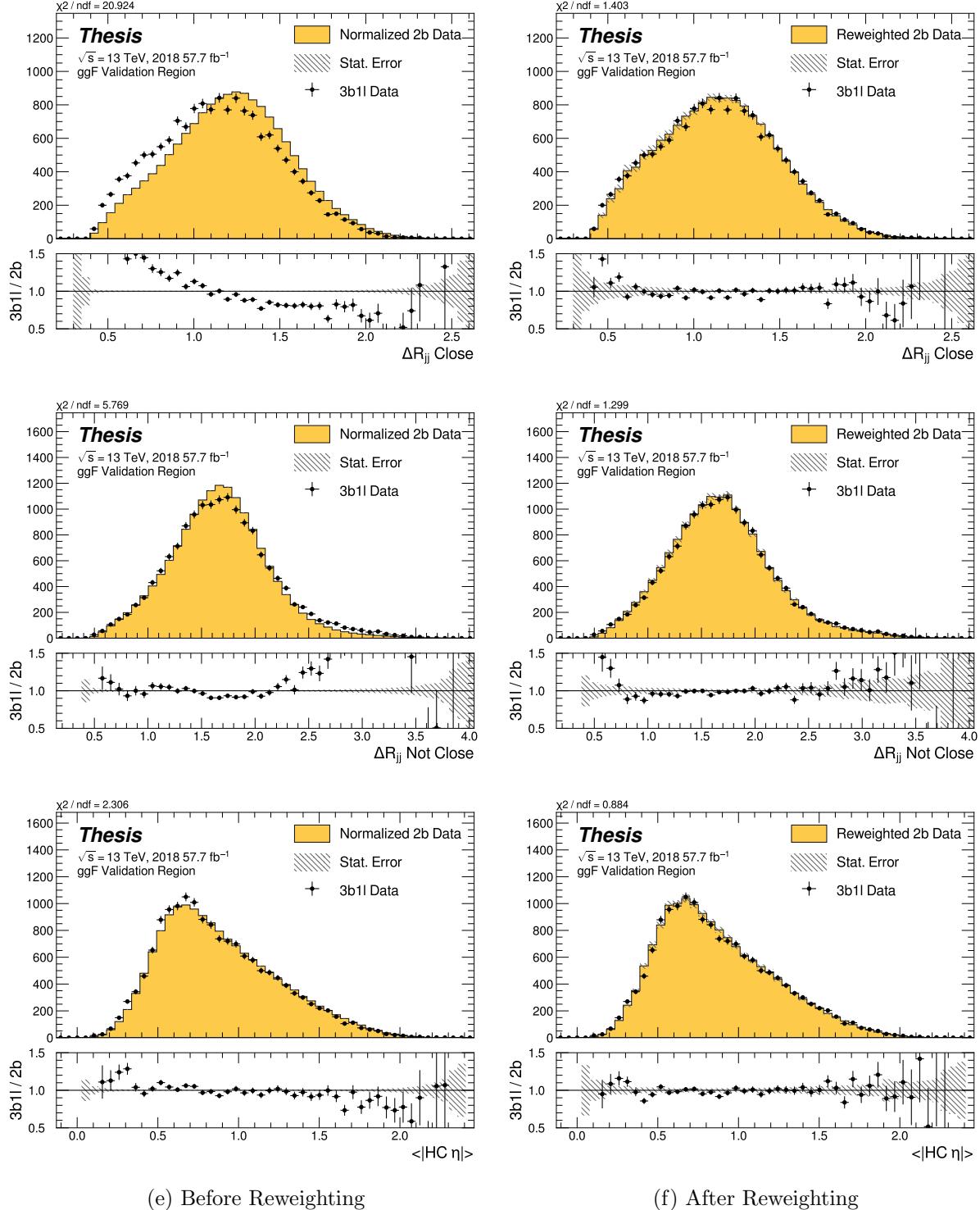


Figure 7.51: **Non-resonant Search (3b1l):** Distributions of  $\Delta R$  between the closest Higgs Candidate jets,  $\Delta R$  between the other two, and average absolute value of HC jet  $\eta$  before and after CR derived reweighting for the 2018 3b1l Validation Region.

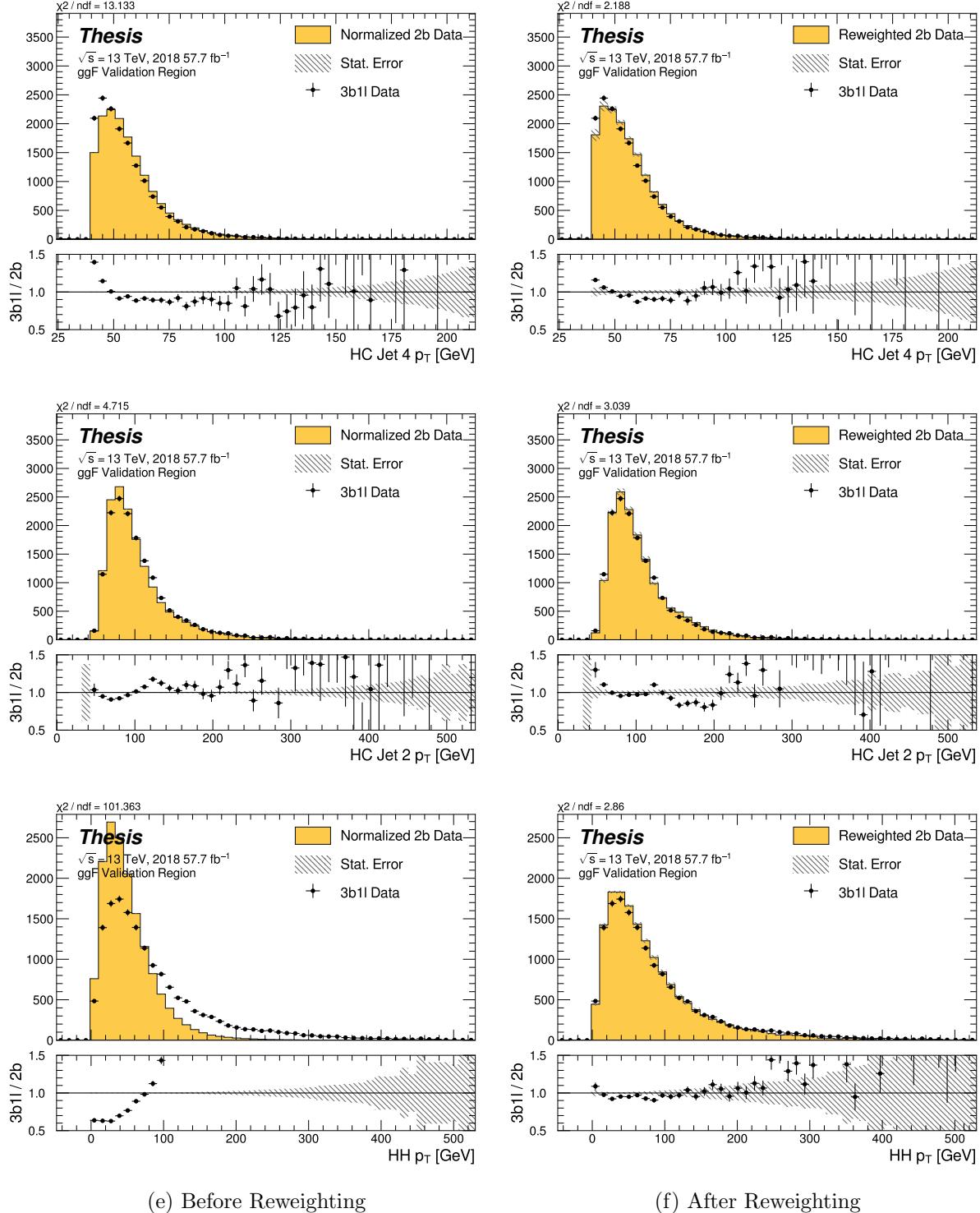


Figure 7.52: **Non-resonant Search (3b1l):** Distributions of  $p_T$  of the 2nd and 4th leading Higgs Candidate jets and the  $p_T$  of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Validation Region.

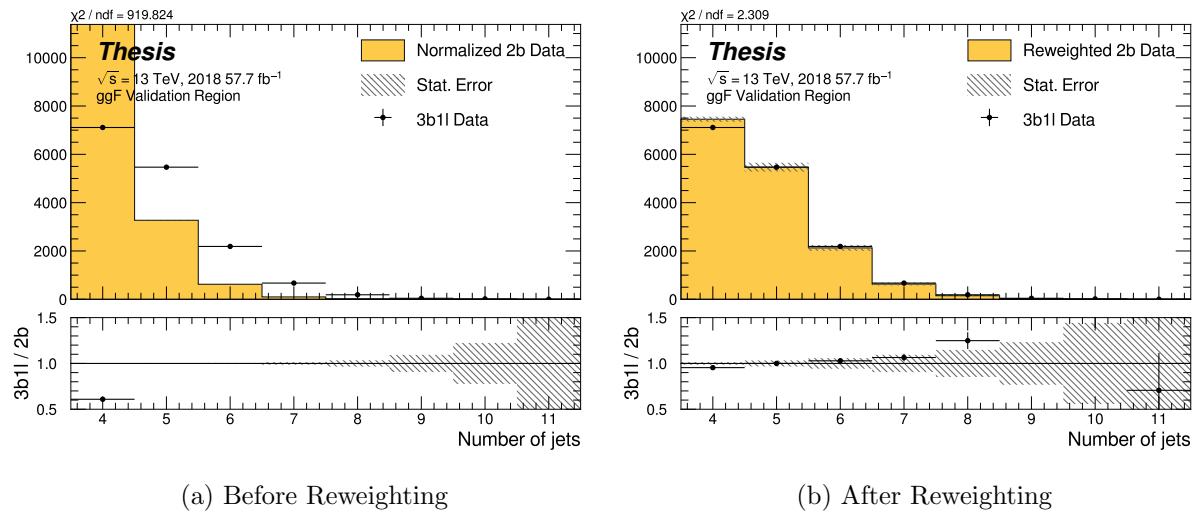


Figure 7.53: **Non-resonant Search (3b1l):** Distributions of the number of jets before and after CR derived reweighting for the 2018 3b1l Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

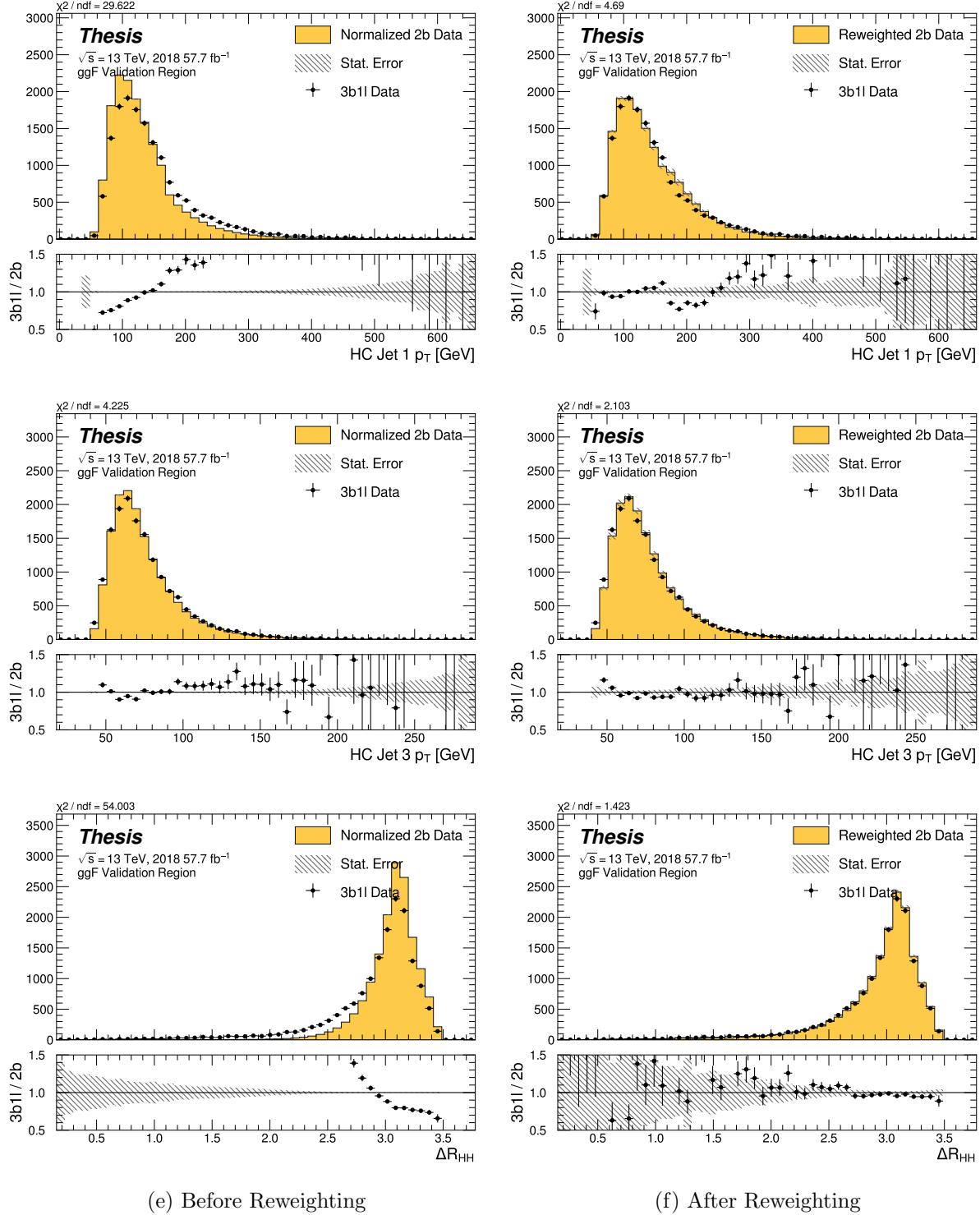


Figure 7.54: **Non-resonant Search (3b1l):** Distributions of  $p_T$  of the 1st and 3rd leading Higgs Candidate jets and  $\Delta R$  between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.

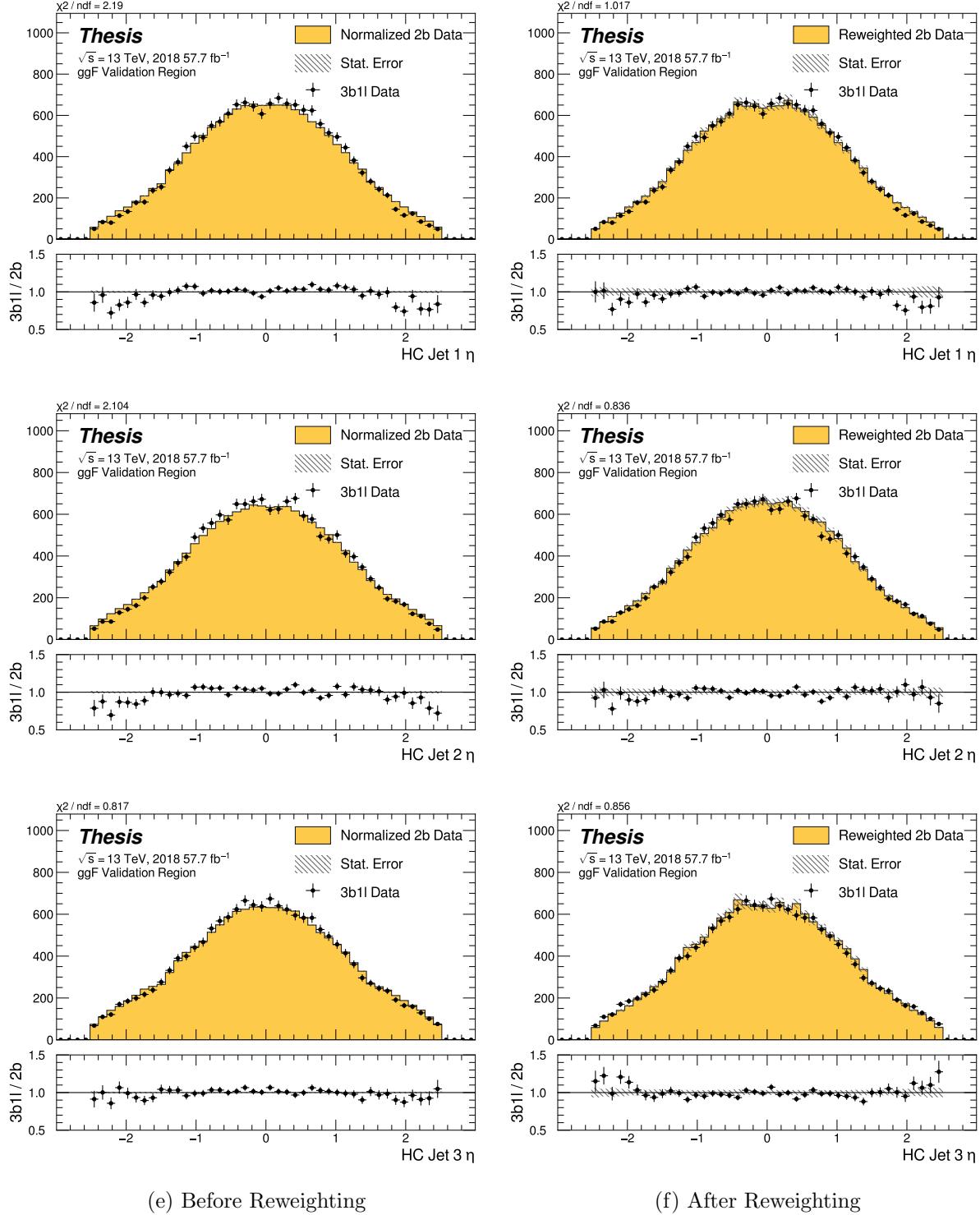


Figure 7.55: **Non-resonant Search (3b1l):** Distributions of  $\eta$  of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Validation Region.

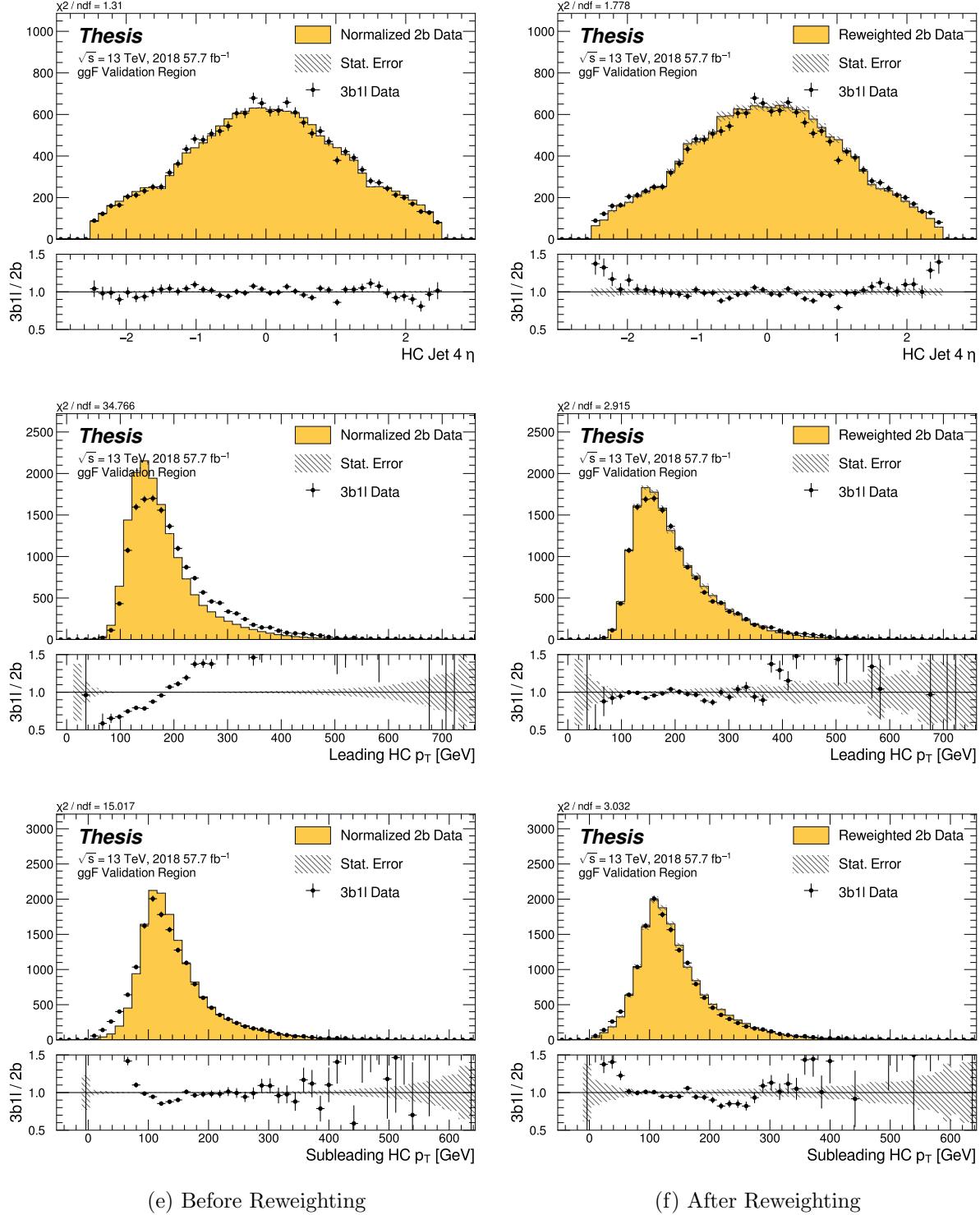


Figure 7.56: **Non-resonant Search (3b1l):** Distributions of  $\eta$  of the 4th leading Higgs Candidate jet and the  $p_T$  of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.

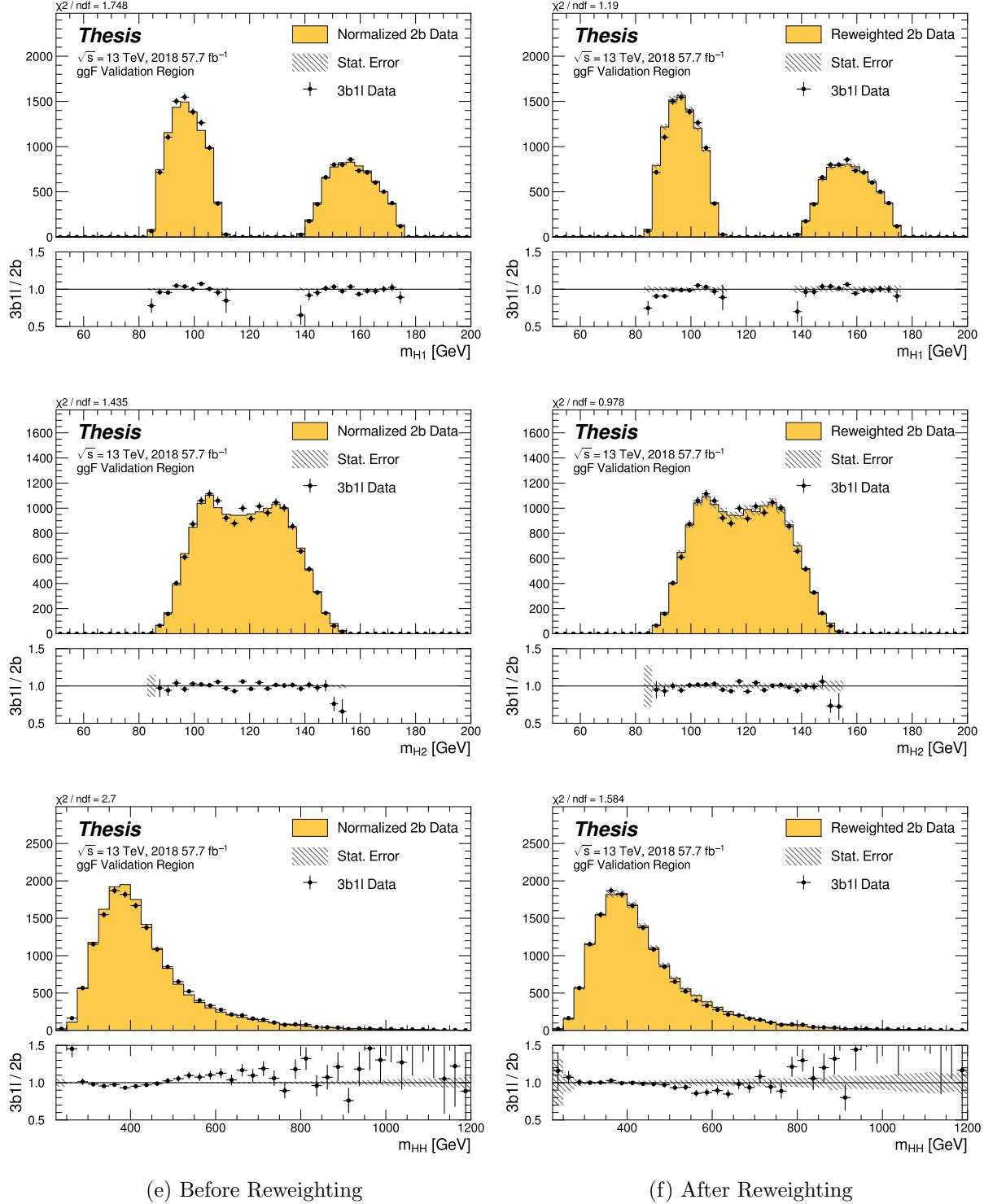


Figure 7.57: **Non-resonant Search (3b1l):** Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Validation Region.

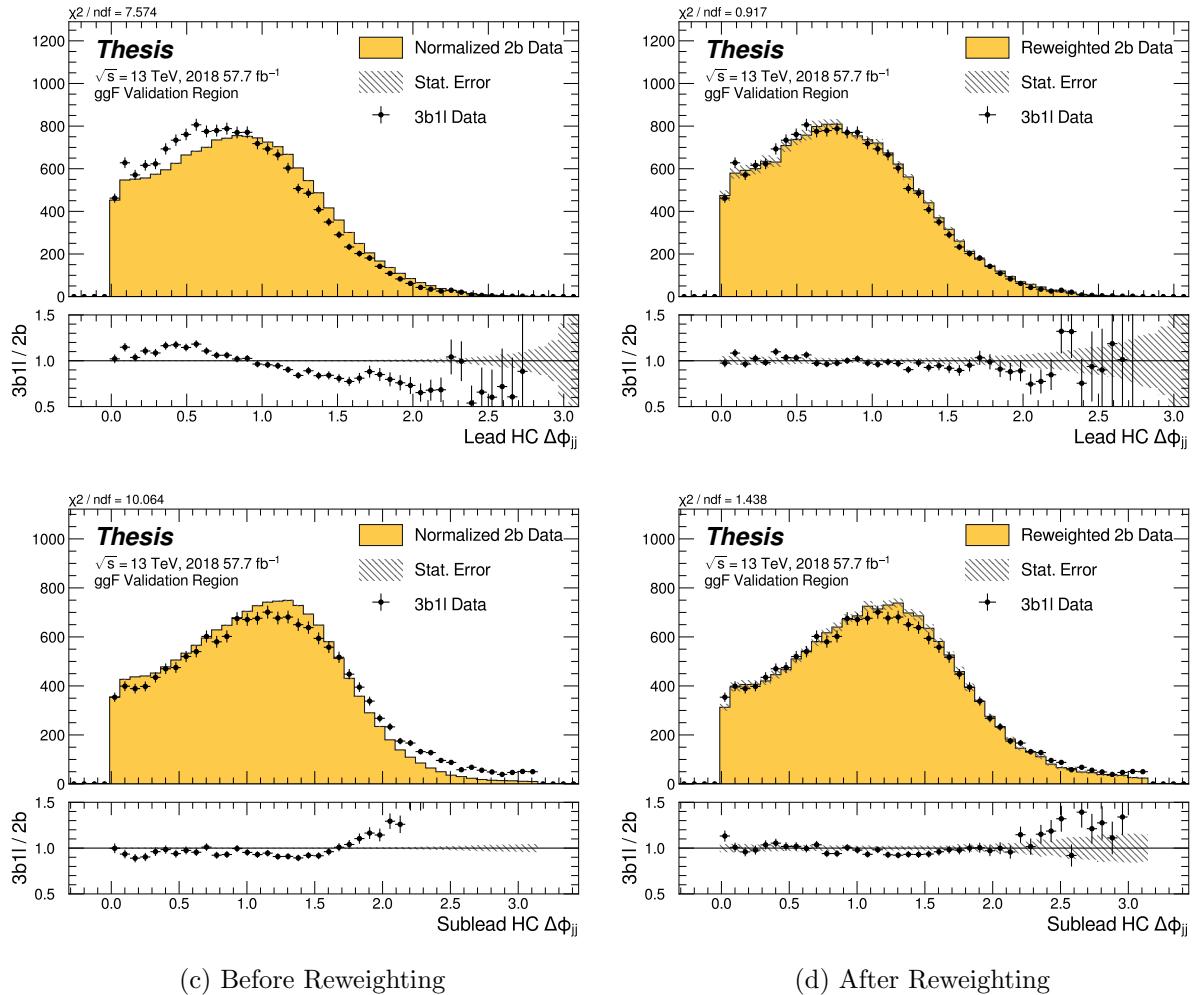


Figure 7.58: **Non-resonant Search (3b1l):** Distributions of  $\Delta\phi$  between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.

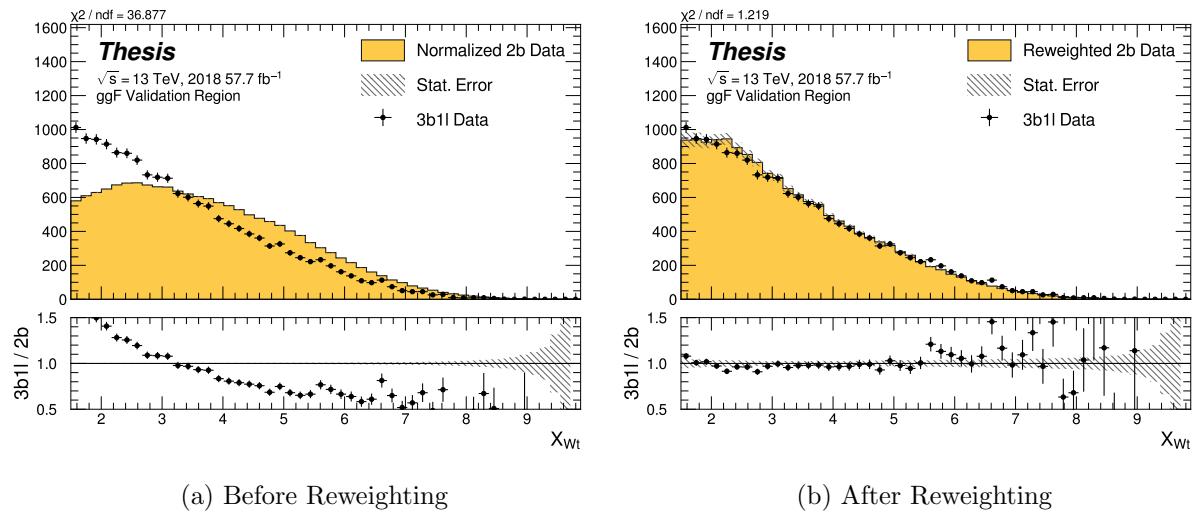


Figure 7.59: **Non-resonant Search (3b1l):** Distributions of the top veto variable,  $X_{Wt}$ , before and after CR derived reweighting for the 2018 3b1l Validation Region. Reweighting is done after the cut on this variable is applied.

<sup>1959</sup> **7.7 Uncertainties**

<sup>1960</sup> A variety of uncertainties are assigned to account for known biases in the underlying methods,  
<sup>1961</sup> calibrations, and objects used for this analysis. The largest such uncertainty is associated with  
<sup>1962</sup> the kinematic bias inherent in deriving the background estimate away from the signal region.  
<sup>1963</sup> However, a statistical biasing of this same estimate has an effect of a similar magnitude.  
<sup>1964</sup> Additionally, due to the use of Monte Carlo for signal modelling and  $b$ -tagging calibration,  
<sup>1965</sup> uncertainties related to mismodellings in simulation must also be accounted for. These  
<sup>1966</sup> components, and their impact on this analysis, are described here in detail. Note that, while  
<sup>1967</sup> the Poisson error (from 2b data statistics) is negligible relative to the bootstrap error in  
<sup>1968</sup> the bulk of the distribution, it becomes relevant in the high  $m_{HH}$  tail. The final statistical  
<sup>1969</sup> uncertainty used for the limit setting is therefore the sum (in quadrature) of these two  
<sup>1970</sup> components.

<sup>1971</sup> **7.7.1 Statistical Uncertainties and Bootstrapping**

<sup>1972</sup> There are two components to the statistical error for the neural network background estimate.  
<sup>1973</sup> The first is standard Poisson error, i.e., a given bin,  $i$ , in the background histogram has value  
<sup>1974</sup>  $n_i = \sum_{j \in i} w_j$ , where  $w_j$  is the weight for an event  $j$  which falls in bin  $i$ . Standard techniques  
<sup>1975</sup> 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  
<sup>1976</sup> when all  $w_j$  are equal to 1.

<sup>1977</sup> However, this procedure does not take into account the statistical uncertainty on the  
<sup>1978</sup>  $w_j$  due to the finite training dataset. Due to the large size difference between the two tag  
<sup>1979</sup> and four tag datasets, it is the statistical uncertainty due to the four tag training data that  
<sup>1980</sup> dominates that on the background. A standard method for estimating this uncertainty is the  
<sup>1981</sup> bootstrap resampling technique [76]. Conceptually, a set of statistically equivalent sets is  
<sup>1982</sup> constructed by sampling with replacement from the original training set. The reweighting  
<sup>1983</sup> network is then trained on each of these separately, resulting in a set of statistically equivalent  
<sup>1984</sup> background estimates. Each of these sets is below referred to as a replica.

1985 In practice, as the original training set is large, the resampling procedure is able to  
 1986 be simplified through the relation  $\lim_{n \rightarrow \infty} \text{Binomial}(n, 1/n) = \text{Poisson}(1)$ , which dictates that  
 1987 sampling with replacement is approximately equivalent to applying a randomly distributed  
 1988 integer weight to each event, drawn from a Poisson distribution with a mean of 1.

1989 Though the network configuration itself is the same for each bootstrap training, the  
 1990 network initialization is allowed to vary. It should therefore be noted that the bootstrap  
 1991 uncertainties implicitly capture the uncertainty due to this variation in addition to the  
 1992 previously mentioned training set variation.

1993 The variation from this bootstrapping procedure is used to assign a bin-by-bin uncertainty  
 1994 which is treated as a statistical uncertainty in the fit. Due to practical constraints, a  
 1995 procedure for approximating the full bootstrap error band is developed which demonstrates  
 1996 good agreement with the full bootstrap uncertainty. This procedure is described below.

1997 *Calculating the Bootstrap Error Band*

1998 The standard procedure to calculate the bootstrap uncertainty would proceed as follows: first,  
 1999 each network trained on each bootstrap replica dataset would be used to produce a histogram  
 2000 in the variable of interest. This would result in a set of replica histograms (e.g. for 100  
 2001 bootstrap replicas, 100 histograms would be created). The nominal estimate would then be  
 2002 the mean of bin values across these replica histograms, with errors set by the corresponding  
 2003 standard deviation.

2004 In practice, such an approach is inflexible and demanding both in computation and in  
 2005 storage, in so far as we would like to produce histograms in many variables, with a variety  
 2006 of different cuts and binnings. This motivates a derivation based on event-level quantities.  
 2007 However, due to non-trivial correlations between replica weights, simple linear propagation of  
 2008 event weight variation is not correct.

2009 We therefore adopt an approach which has been empirically found to produce results  
 2010 (for this analysis) in line with those produced by generating all of the histograms, as in the  
 2011 standard procedure. This approach is described below. Note that, for robustness to outliers

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

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

1. Replica weight ( $w_i$ ): weight predicted for a given event by a network trained on replica dataset  $i$ .

2. Replica norm ( $\alpha_i$ ): normalization factor for replica  $i$ . This normalizes the reweighting prediction of the network trained on replica dataset  $i$  to match the correponding target yield.

3. Median weight ( $w_{med}$ ): median weight for a given event across replica datasets, used for the nominal estimate. Defined (for 100 bootstrap replicas) as

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

4. Normalization correction ( $\alpha_{med}$ ): normalization factor to match the predicted yield of the median weights ( $w_{med}$ ) to the target yield in the training region.

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

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

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

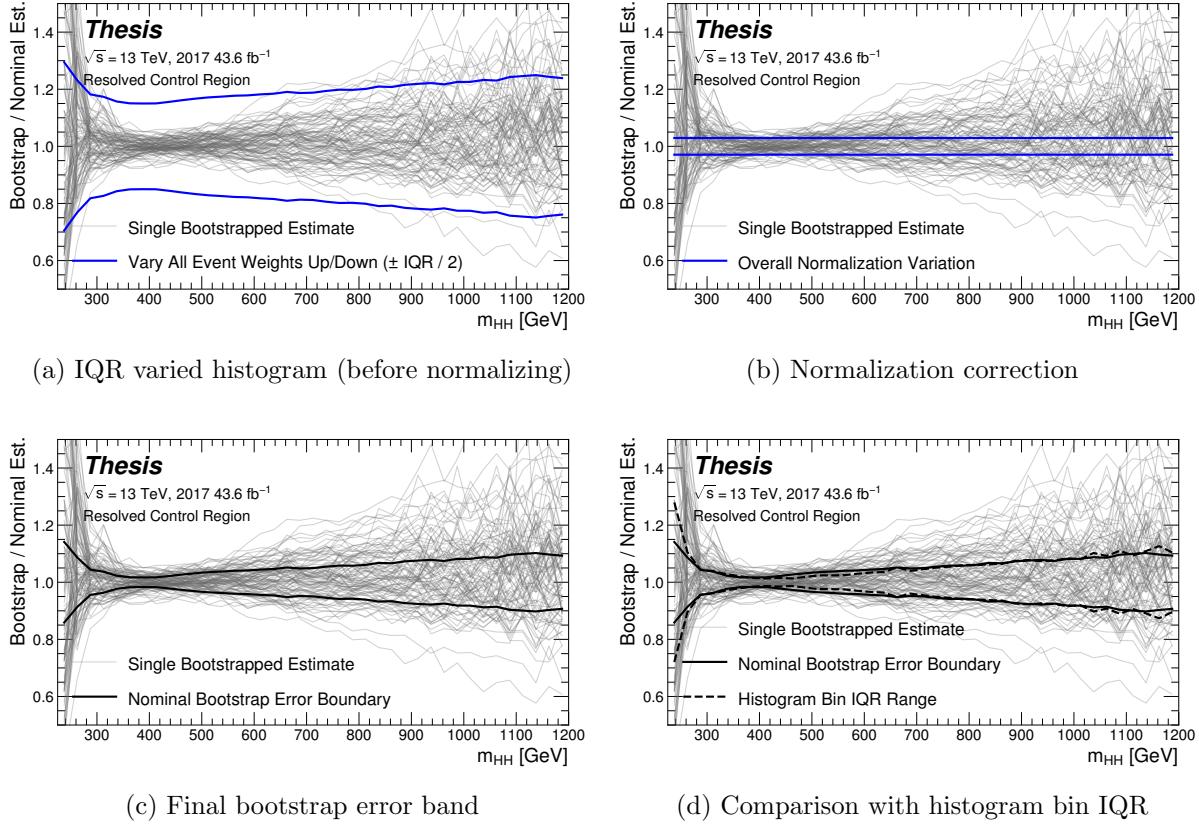


Figure 7.60: Illustration of the approximate bootstrap band procedure, shown as a ratio to the nominal estimate for the 2017 non-resonant background estimate. Each grey line is from the  $m_{HH}$  prediction for a single bootstrap training. Figure 7.60(a) shows the variation histograms constructed from 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.

2042    7.7.2 *Background Shape Uncertainties*

2043    To account for the systematic bias associated with deriving the reweighting function in the  
 2044    control region and extrapolating to the signal region, an alternative background model is  
 2045    derived in the validation region. Because of the fully data-driven nature of the background  
 2046    model, this is an uncertainty assessed on the full background. The alternative model and  
 2047    the baseline are consistent with the observed data in their training regions, and differences  
 2048    between the alternative and baseline models are used to define a shape uncertainty on the  
 2049     $m_{HH}$  spectrum, with a two-sided uncertainty defined by symmetrizing the difference about  
 2050    the baseline.

2051    For the resonant analysis, this uncertainty is split into two components to allow for two  
 2052    independent variations of the  $m_{HH}$  spectrum: : a low- $H_T$  and a high- $H_T$  component, where  
 2053     $H_T$  is the scalar sum of the  $p_T$  of the four jets constituting the Higgs boson candidates, and  
 2054    serves as a proxy for  $m_{HH}$ , while avoiding introducing a sharp discontinuity. The boundary  
 2055    value is 300 GeV. The low- $H_T$  shape uncertainty primarily affects the  $m_{HH}$  spectrum below  
 2056    400 GeV (close to the kinematic threshold) by up to around 5%, and the high- $H_T$  uncertainty  
 2057    mainly  $m_{HH}$  above this by up to around 20% relative to nominal. These separate  $m_{HH}$   
 2058    regimes are by design – the  $H_T$  split is introduced to prevent low mass bins from constraining  
 2059    the high mass uncertainty and vice-versa.

2060    This was the *status quo* shape uncertainty decomposition from the Early Run 2 analysis.  
 2061    A decomposition in terms of orthogonal polynomials, which would provide increased flexibility,  
 2062    was also evaluated. This study revealed that both decompositions are able to account for the  
 2063    systematic deviations between four tag data and the background estimate (evaluated in the  
 2064    kinematic validation region), and produce almost identical limits. The simpler *status quo*  
 2065    decomposition is therefore kept.

2066    For the non-resonant analysis, the quadrant nature of the background estimation leads to  
 2067    a natural breakdown of the nuisance parameters: quadrants are defined in the signal region  
 2068    along the same axes as those used for the control and validation region definitions. Variations

2069 are then assessed in each of these signal region quadrants, corresponding to regions that  
 2070 are “closer to” and “further away from” the nominal and alternate estimate regions, fully  
 2071 leveraging the power of the two equivalent but systematically different estimates.

2072 Figure 7.61 shows an example of the variation in each  $H_T$  region for the 2018 resonant  
 2073 analysis. Figure 7.62 shows the example quadrant variation for the 2018 4 $b$  non-resonant  
 analysis.

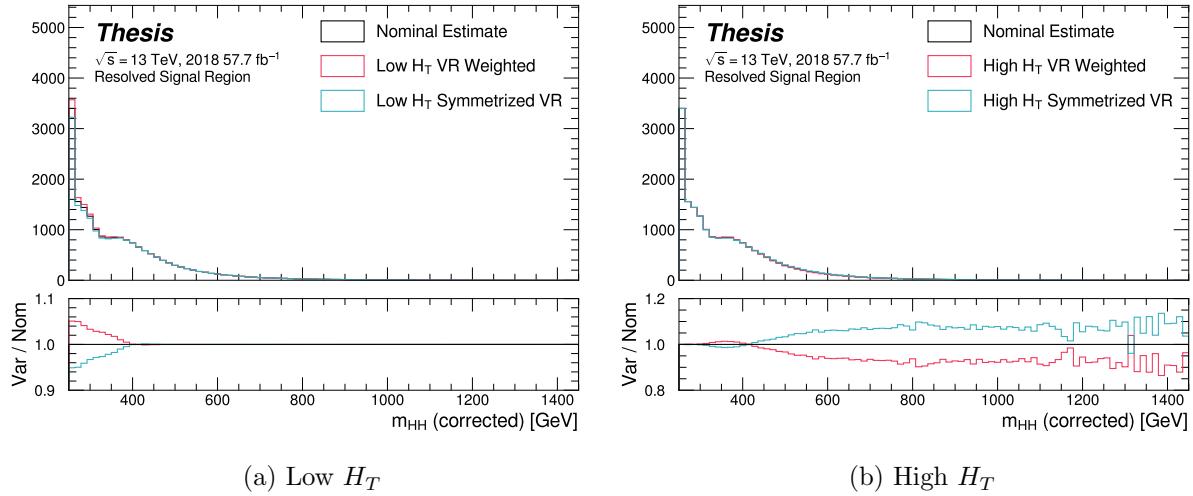
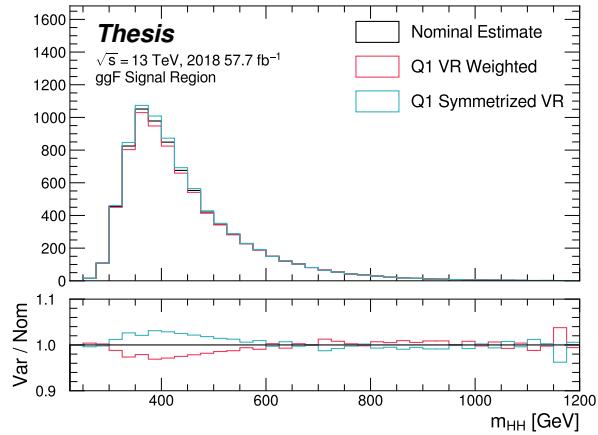
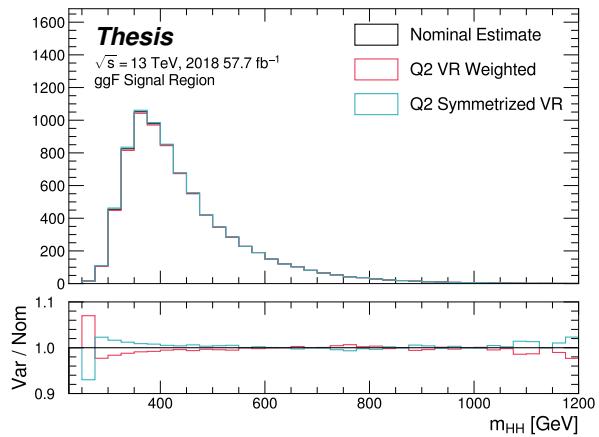


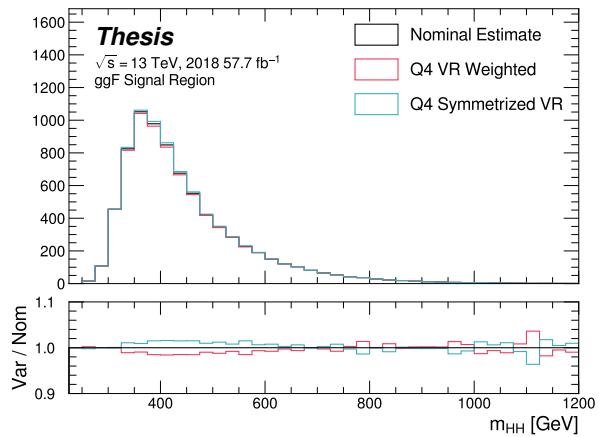
Figure 7.61: **Resonant Search:** Example of CR vs VR variation in each  $H_T$  region for 2018.  
 The variation nicely factorizes into low and high mass components.



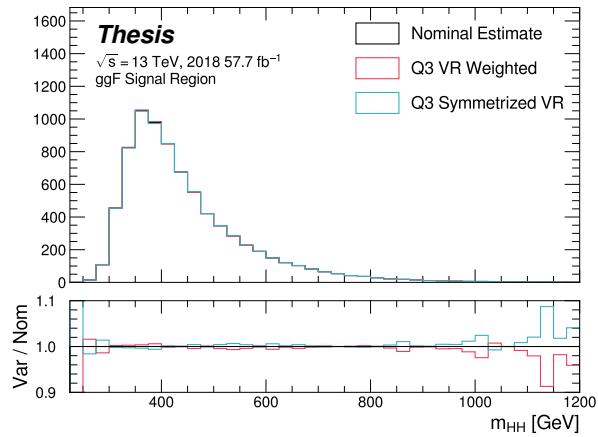
(a) Q1 (top)



(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.

2075 *7.7.3 Detector Modelling and Reconstruction Uncertainties*

2076 Detector modelling and reconstruction uncertainties account for Monte Carlo simulation not  
 2077 being a faithful representation of real data as a result of mismodelling of the detector and  
 2078 differential performance of algorithms on simulation compared to data. In this analysis they  
 2079 consist of uncertainties related to jet properties, and uncertainties stemming from the flavour  
 2080 tagging procedure. The background modelling in this analysis is fully data-driven. As a  
 2081 result, these uncertainties are applied only to the signal simulation.

2082 The jet uncertainties are implemented as variations of the jet properties themselves. The  
 2083 category reduction (with  $\sim 30$  nuisance parameters) is used for jet energy scale uncertainties  
 2084 and the FullJER configuration is used for jet energy resolution uncertainties (14 nuisance  
 2085 parameters). This is to preserve the ability to meaningfully statistically combine the results  
 2086 of this analysis with other di-Higgs analyses. The flavour tagging uncertainties meanwhile  
 2087 are implemented as scale factors applied to the Monte Carlo event weights.

2088 A systematic related to the PtReco  $b$ -jet energy correction has been studied in the  
 2089  $HH \rightarrow \gamma\gamma b\bar{b}$  analysis [77] and found to be negligible compared to JER. Following this  
 2090 example, such a systematic is therefore neglected here.

2091 *7.7.4 Trigger Uncertainties*

2092 Trigger uncertainties stem from imperfect knowledge of the ratio between the efficiency of a  
 2093 given trigger in data to its efficiency in Monte Carlo simulation. This ratio is applied as a  
 2094 scale factor to all simulated events (as described in Section ??), with the systematic variations  
 2095 produced by varying the scale factor up or down by one sigma.

2096 *7.7.5 Theoretical Uncertainties*

2097 The theoretical uncertainties on the acceptance times efficiency ( $A \times \varepsilon$ ) are evaluated by  
 2098 analysis of specially-generated, particle-level signal samples. The generation of these samples  
 2099 follows the configuration of the baseline samples, but with modifications to probe the following

2100 theoretical uncertainties: uncertainties in the parton density functions (PDFs); uncertainties  
 2101 due to missing higher order terms in the matrix elements; and uncertainties in the modelling  
 2102 of the underlying event, which includes multi-parton interactions, of hadronic showers and of  
 2103 initial and final state radiation.

2104 Uncertainties due to modelling of the parton shower and the underlying event (including  
 2105 multi-parton interactions) are evaluated by switching the MC generator used. For the scalar  
 2106 samples, this means switching from Herwig7.7.1.3 to Pythia 8.235. Figure ?? shows the  
 2107 impact of these variations on the signal acceptance for two resonance masses: 500 GeV and  
 2108 1 TeV, covering the range of the resolved analysis. No significant dependence on the variable  
 2109 of interest,  $m_{HH}$ , is observed. The disagreement observed in the tails of  $p_T(hh)$  and the  
 2110 number of jets multiplicities is negligible with respect to the final signal acceptance. A 5%  
 2111 flat systematic uncertainty is assigned to all signal samples, extracted from the acceptance  
 2112 comparison for the full 4-tag selection, as seen in Figs. ?? and ??.

2113 To evaluate the potential effect of missing higher order terms in the matrix element, the  
 2114 renormalization and factorization scales used in the signal generation were varied coherently  
 2115 by factors of  $0.5\times$  and  $2\times$  for the signals. The alternative weights were generated as described  
 2116 on the TWiki [here](#), applying on-the-fly variations using the ATLAS MadGraphControl  
 2117 framework. These weights correspond to variations of the scales either together or separately  
 2118 up and down by a factor of two. Seven-point scale variations are considered:  $(\mu_R, \mu_F) = (0.5,$   
 2119  $0.5), (1, 0.5), (0.5, 1), (1, 1), (2, 1), (1, 2), (2, 2)$ . The scale uncertainties are combined by  
 2120 taking an envelope of all of the uncertainties. These uncertainties are evaluated to be less  
 2121 than  $\pm 1\%$ , thus neglected.

2122 PDF uncertainties are evaluated using the PDF4LHC15\_nlo\_mc set, which combined  
 2123 CT14, MMHT14 and NNPDF3.0 PDF sets. The uncertainty is evaluated by calculating  
 2124 the acceptance for each PDF replica. The standard deviation of these acceptance values  
 2125 divided by the baseline acceptance is taken as the PDF uncertainty. For each mass point the  
 2126 distribution of their corresponding ration is compatible with a Gaussian centered one. The  
 2127 measured uncertainty in acceptance due to PDF uncertainties is less than  $\pm 1\%$  across the

2128 full mass range considered for the analysis. For this reason, it is neglected in the statistical  
2129 analysis described in Section ??.

2130 These uncertainties are implemented in the final statistical analysis as normalization  
2131 uncertainties on the signals, with the value taken from the polynomial fit. This smooths out  
2132 statistical fluctuations and allows interpolation between the generated mass points, if needed.

2133 The results for the non-resonant analysis presented here are preliminary and only include  
2134 background systematics. However, these are expected to be by far the dominant uncertainties,  
2135 and should therefore be reflective of the final results.

2136 **7.8 Background Validation**

2137 In addition to checking the performance of the background estimate in the control and  
2138 validation regions, a variety of alternative selections are defined to allow for a full “dress  
2139 rehearsal” of the background estimation procedure.

2140 Both the resonant and non-resonant analyses make use of a *reversed*  $\Delta\eta$  region, in which  
2141 the kinematic cut on  $\Delta\eta_{HH}$  is reversed, so that events are required to have  $\Delta\eta_{HH} > 1.5$ .  
2142 This is orthogonal to the nominal signal region and has minimal sensitivity, allowing for the  
2143 comparison of the background estimate  $4b$  data in the corresponding “signal region”. For  
2144 this validation, a new reweighting is trained following nominal procedures, but entirely in the  
2145  $\Delta\eta_{HH} > 1.5$  region.

2146 The non-resonant analysis additionally makes use of the  $3b + 1$  fail region mentioned  
2147 above, which again is orthogonal to the nominal signal regions and has minimal sensitivity.  
2148 The reweighting in this case is between  $2b$  and  $3b + 1$  fail events rather than between  $2b$   
2149 and  $3b + 1$  loose or  $2b$  and  $4b$ . However, the kinematic selections of signal region events are  
2150 otherwise identical, allowing for a complementary test of the background estimate.

2151 *TODO: Add shifted regions if they’re ready*

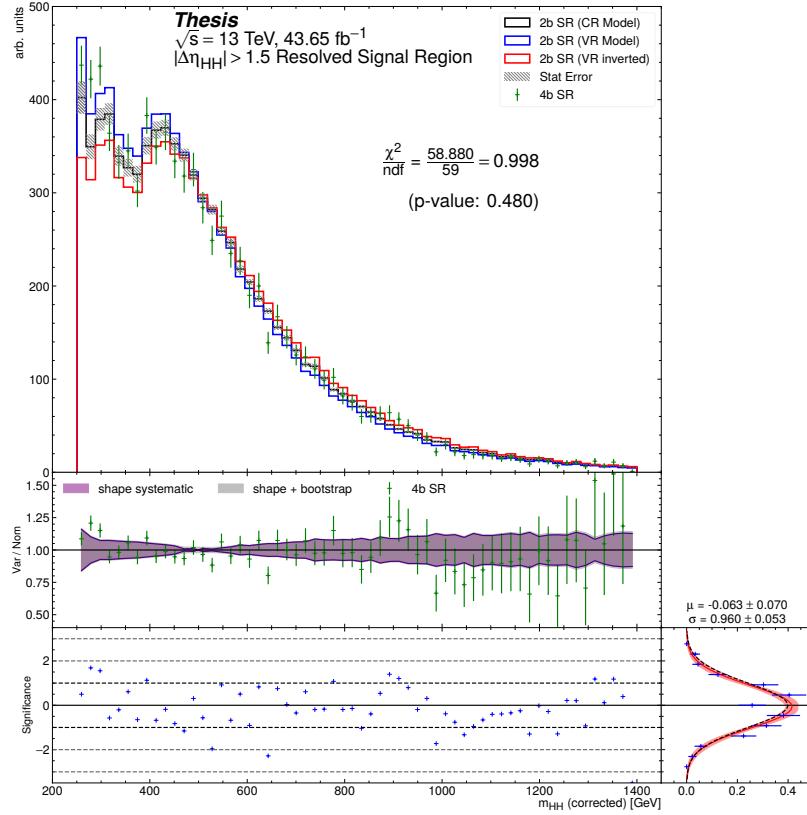


Figure 7.63: **Resonant Search:** Performance of the background estimation method in the resonant analysis reversed  $\Delta\eta_{HH}$  kinematic signal region. A new background estimate is trained following nominal procedures entirely within the reversed  $\Delta\eta_{HH}$  region, and the resulting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

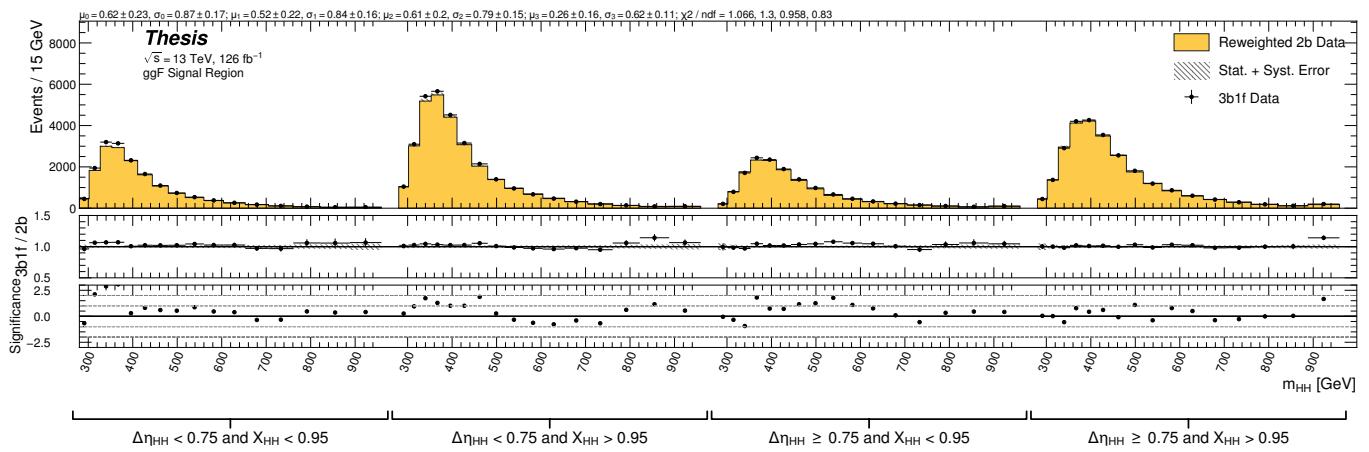


Figure 7.64: **Non-resonant Search:** Performance of the background estimation method in the  $3b + 1$  fail validation region. A new background estimate is trained following nominal procedures but with a reweighting from  $2b$  to  $3b + 1$  fail events. Generally good agreement is seen, though there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

2152 **7.9 Overview of Other  $b\bar{b}b\bar{b}$  Channels**

2153 The results discussed above have been developed in conjunction with (1) a boosted channel for  
2154 the resonant search and (2) a vector boson fusion (VBF) channel for the non-resonant search.  
2155 Detailed discussions of these two channels are beyond the scope of this thesis. However,  
2156 a combined set of results is presented below (*TODO: or will be combined for VBF?*). We  
2157 therefore briefly summarize the analyses here.

2158 **7.9.1 Resonant: Boosted Channel**

2159 The boosted analysis selection targets resonance masses from 900 GeV to 5 TeV. In such  
2160 events,  $H$  decays have a high Lorentz boost, such that the  $b\bar{b}$  decays are very collimated. The  
2161 resolved analysis fails to reconstruct such  $HH$  events, as the  $R = 0.4$  jets start to overlap.

2162 The boosted analysis instead reconstructs  $H$  decays as large radius,  $R = 1.0$  jets, with  
2163 corresponding  $b$ -quarks identified with variable radius subjets, that is jets with a radius that  
2164 scales as  $\rho/p_T$ , the  $p_T$  is that of the jet in question, and  $\rho$  is a fixed parameter, here chosen  
2165 to be 30 GeV, which is optimized to maintain truth-level double  $b$ -labelling efficiency across  
2166 the full range of Higgs jet  $p_T$  *TODO: cite: <https://cds.cern.ch/record/2268678>.*

2167 Due to limited boosted  $b$ -tagging efficiency *TODO: cite* and to maintain sensitivity even  
2168 when  $b$ -jets are highly collimated, the boosted analysis is divided into three categories based  
2169 on the number of  $b$ -tagged jets associated to each large radius jet:

- 2170 • 4 $b$  category: two  $b$ -tagged jets in each
- 2171 • 2 $b - 1$  category: two  $b$ -tagged jets in one, one in the other
- 2172 • 1 $b - 1$  category: one  $b$ -tagged jet in each

2173 The analysis then proceeds in each of these categories. *TODO: what other boosted details?*  
2174 The resolved and boosted channels are combined for resonance masses from 900 GeV to  
2175 1.5 TeV inclusive. To keep the channels statistically independent, the boosted channel vetos

2176 events passing the resolved analysis selection.

2177 *7.9.2 Non-resonant: VBF Channel*

2178 The vector boson fusion channel is only considered for the non-resonant search. While the  
2179 sensitivity is in general much more limited than the gluon-gluon fusion analysis due to the  
2180 much smaller production cross section, VBF is sensitive to a variety of Beyond the Standard  
2181 Model physics, both complementary and orthogonal to the theoretical scope of gluon-gluon  
2182 fusion. *TODO: I'll probably mention more details in the pheno section*

2183 The VBF channel proceeds very similarly to the ggF, with the primary differences being  
2184 the kinematic selections and the categorization.

2185 *TODO: fill in kinematics*

2186 Note that the background estimation is inherited from the resonant and ggF analyses, an  
2187 ancillary, but significant, contribution of this thesis work.

2188 **7.10  $m_{HH}$  Distributions**

2189 *7.10.1 Resonant Search*

2190 The final discriminant used for the resonant search is corrected  $m_{HH}$ . Histogram binning  
2191 was optimized for the resonant search to be 84 equal width bins from 250 GeV to 1450 GeV,  
2192 corresponding to a bin width of 14.3 GeV, and overflow events (events above 1450 GeV) are  
2193 included in the last bin. A demonstration of the performance of the reweighting on this  
2194 distribution is shown in Figure 7.65 for the control region and Figure 7.66 for the validation region.  
2195 A background-only profile likelihood fit is run for the distribution in the signal region, and results with spin-0 signals overlaid are shown in Figure 7.67. Note that the  
2196 plots show the sum across all years, but the signal extraction fit and background estimate  
2197 are run with the years separately. Agreement is generally good throughout.  
2198

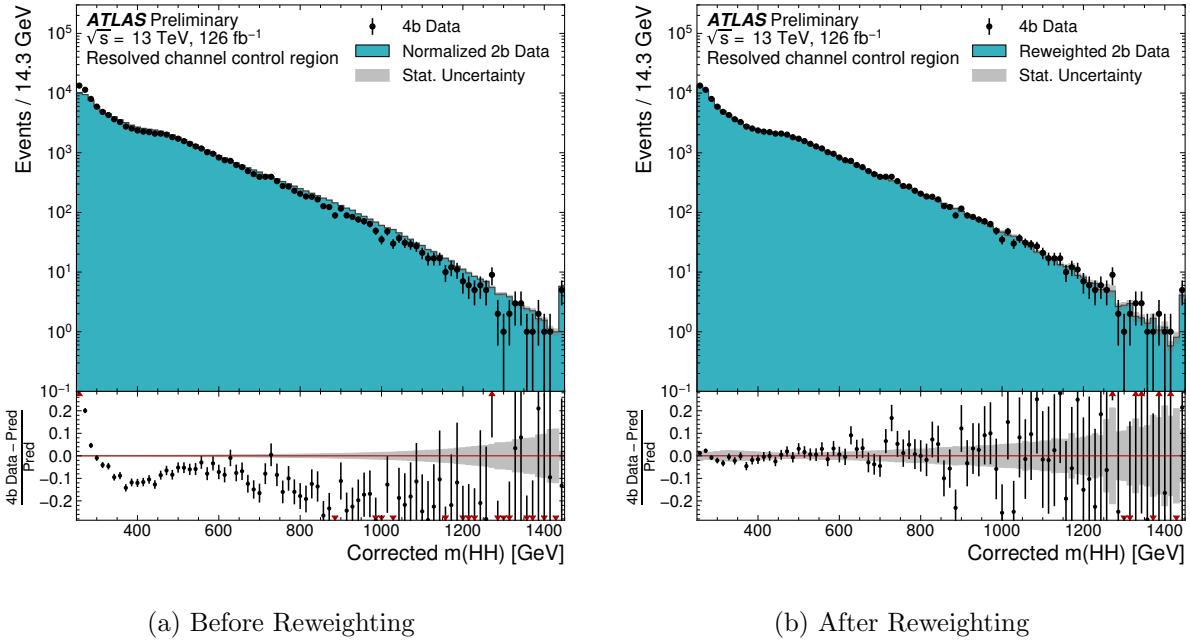


Figure 7.65: **Resonant Search:** Demonstration of the performance of the nominal reweighting in the control region on corrected  $m_{HH}$ , with Figure 7.65(a) showing  $2b$  events normalized to the total  $4b$  yield and Figure 7.65(b) applying the reweighting procedure. Agreement is much improved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

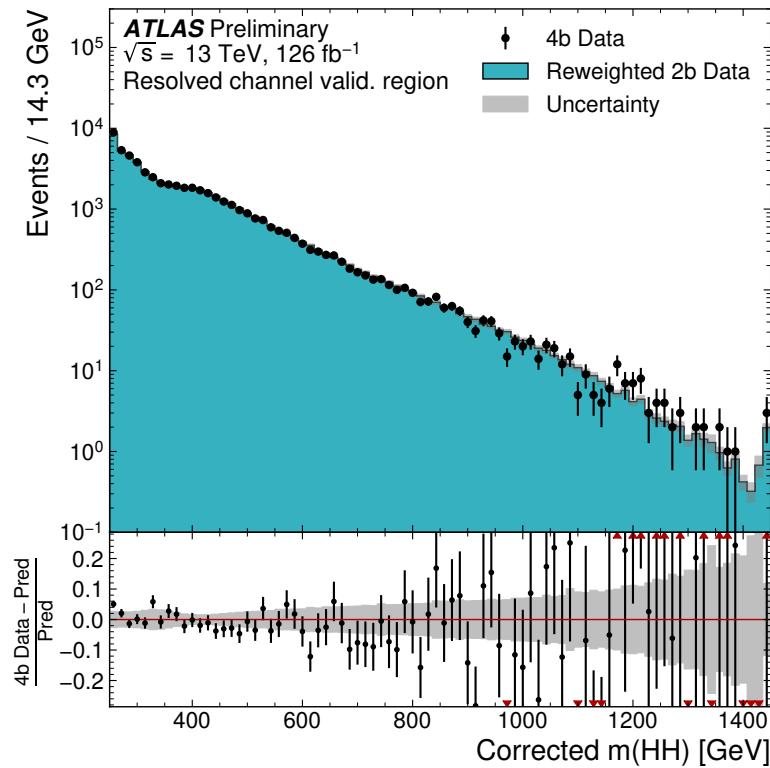


Figure 7.66: **Resonant Search:** Demonstration of the performance of the control region derived reweighting in the validation region on corrected  $m_{HH}$ . Agreement is generally good for this extrapolated estimate. Note that the uncertainty band includes the extrapolation systematic, which is defined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

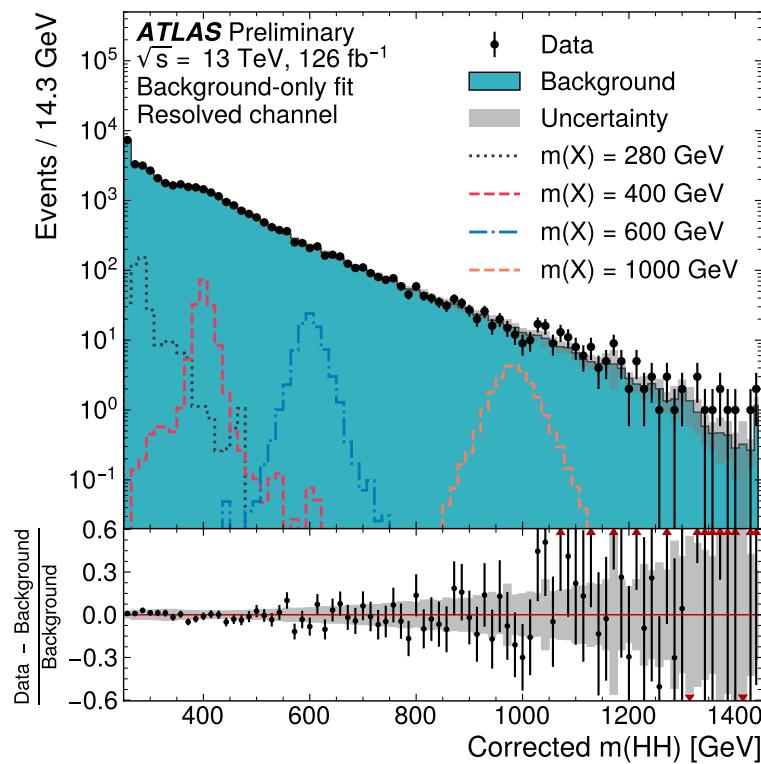


Figure 7.67: **Resonant Search:** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit. The closure is generally quite good, though there is an evident deficit in the background estimate relative to the data for higher values of corrected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

2199 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

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

2200 where  $b_{low}^i$  is the low edge of bin  $i$ . The parameters used here are  $r = 0.08$  over a range  
2201 from 280 GeV to 975 GeV, and underflow and overflow are included in the intial and final  
2202 bin contents respectively.  $m_{HH}$  with no correction is used as the final discriminant in each  
2203 category.

2204 A demonstration of the performance of the reweighting on distributions unrolled across  
2205 categories is shown in Figure *TODO: insert* for the control region and Figure *TODO:*  
2206 *insert* for the validation region. A background-only profile likelihood fit is run for the  
2207 distribution in the signal region, and results with the Standard Model  $HH$  signal and  $\kappa_\lambda = 6$   
2208 signal overlaid are shown for  $4b$  in Figure 7.68 and  $3b1l$  in Figure 7.69. Note that the plots  
2209 show the sum across all years, but the signal extraction fit and background estimate are run  
2210 with the years separately. All bins are normalized to represent a density of Events / 15 GeV.

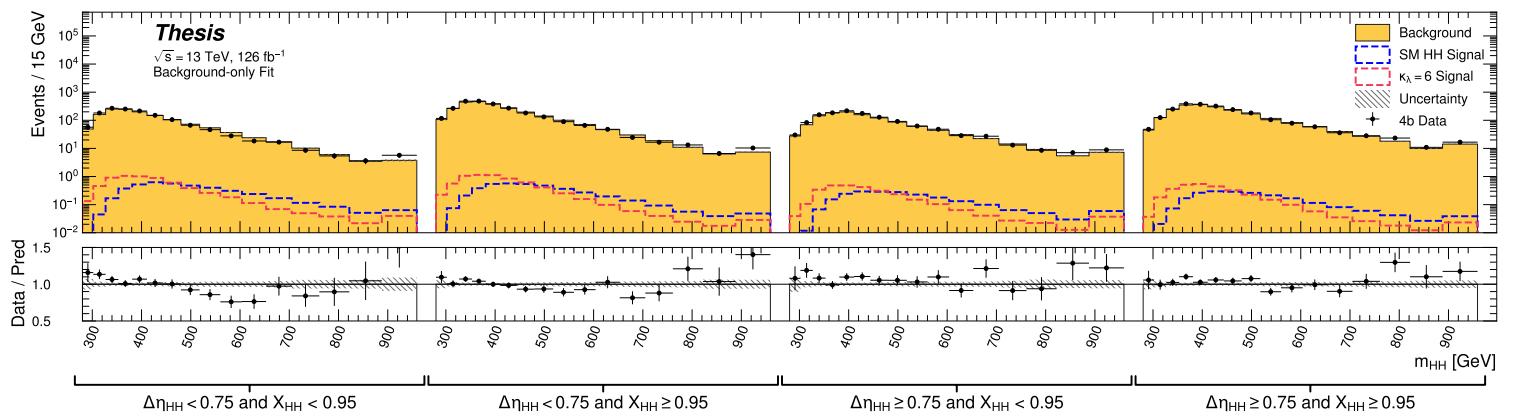


Figure 7.68: **Non-resonant Search (4b):** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 4b channels, with Standard Model and  $\kappa_\lambda = 6$  signal overlaid for reference. Modeling is generally quite good near the Standard 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.

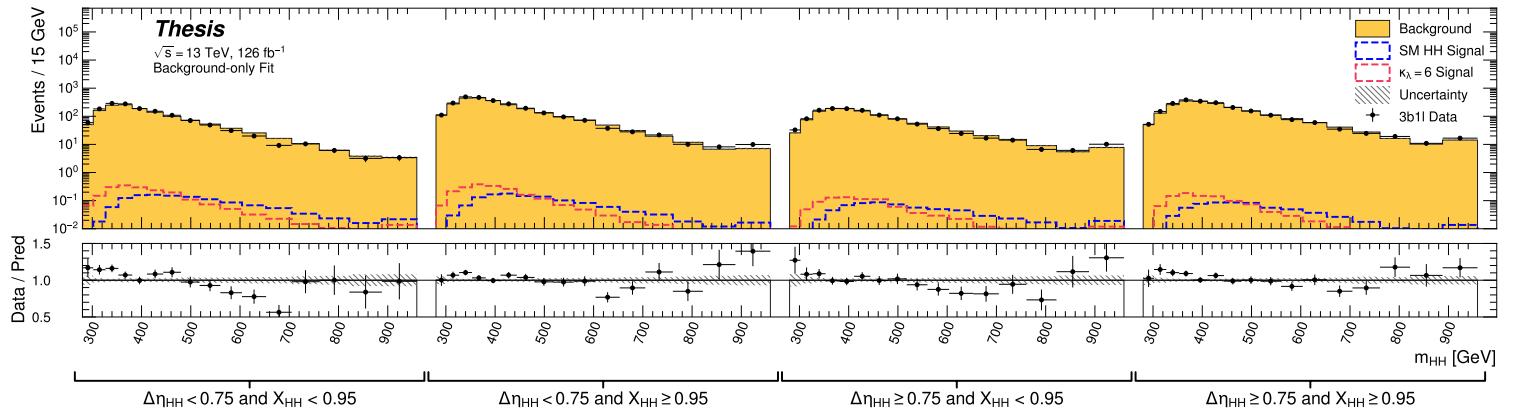


Figure 7.69: **Non-resonant Search (3b1l):** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 3b1l channels, with Standard Model and  $\kappa_\lambda = 6$  signal overlaid for reference. Conclusions are very similar to the 4b 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.

2211 **7.11 Statistical Analysis**

2212 The resonant analysis is used to set a 95% confidence level upper limit on the  $pp \rightarrow X \rightarrow$   
2213  $HH \rightarrow b\bar{b}b\bar{b}$  and  $pp \rightarrow G_{KK}^* \rightarrow HH \rightarrow b\bar{b}b\bar{b}$  cross-sections, while the non-resonant analysis  
2214 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  
2215 a variety of values of the trilinear Higgs coupling.

2216 The upper limit is extracted using the  $CL_s$  method [78]. The test statistic used is  $q_\mu$  [79],  
2217 where  $\mu$  is the signal strength, and  $\theta$  represents the nuisance parameters. Due to the use of  
2218 signals normalized to 1 fb,  $\mu$  is also the signal cross-section in fb. A single hat represents the  
2219 maximum likelihood estimate of a parameter, while  $\hat{\theta}(x)$  represents the conditional maximum  
2220 likelihood estimate of the nuisance parameters if the signal cross-section is fixed at  $x$ .

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

2221  $CL_s$  for some test value of  $\mu$  is then defined by

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

2222 where the  $p$ -values are calculated in the asymptotic approximation [79], which is valid in  
2223 the large sample limit.

2224 The signal cross-section  $\mu$  fb is excluded at the 95% confidence level if  $CL_s < 0.05$ .

Observed	$-2\sigma$	$-1\sigma$	Expected	$+1\sigma$	$+2\sigma$
<b>4.4</b>	3.1	4.2	<b>5.9</b>	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.

## 2225 7.12 Results

2226 Figure 7.70 shows the expected limit for the spin-0 and spin-2 resonant search. The resolved  
 2227 channel covers the range between 251 and 1500 GeV and is combined with the boosted channel  
 2228 between 900 and 1500 GeV. The boosted channel then extends to 3 TeV. The most significant  
 2229 excess is seen for a signal mass of 1100 GeV, with local significance of  $2.6\sigma$  for the spin-0  
 2230 signal and  $2.7\sigma$  for the spin-2 signal. This is reduced to  $1.0\sigma$  and  $1.2\sigma$  globally.

2231 The spin-2 bulk Randall-Sundrum model with  $k/\overline{M}_{\text{Pl}} = 1$  is excluded for graviton masses  
 2232 between 298 and 1440 GeV.

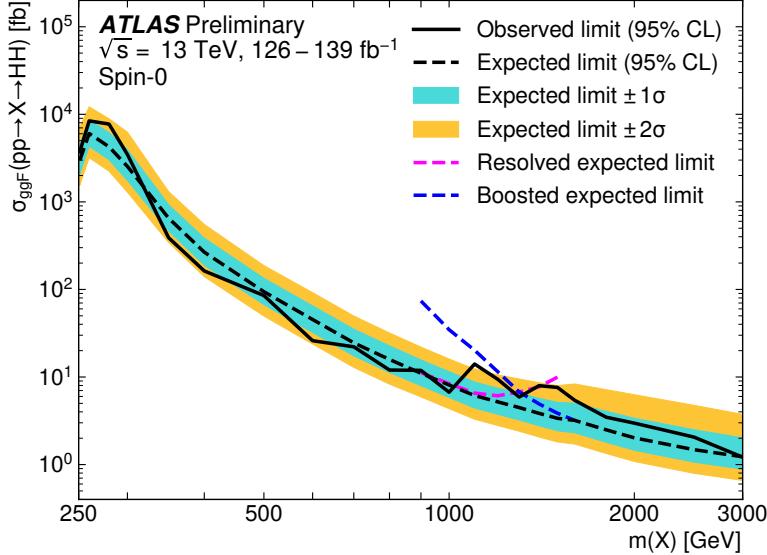
2233 Preliminary results are presented here for the gluon-gluon fusion non-resonant search,  
 2234 combining results from the  $4b$  and  $3b + 1l$  signal regions in the  $2 \times 2$  category scheme in  
 2235  $\Delta\eta_{HH}$  and  $X_{HH}$ . These results will be further combined with a VBF channel as discussed,  
 2236 but this is left for future work. Results shown here include background systematics only.  
 2237 Limits are set for  $\kappa_\lambda$  values from  $-20$  to  $20$ . The cross section limit for  $HH$  production is set  
 2238 at  $140 \text{ fb}$  ( $180 \text{ fb}$ ) observed (expected), corresponding to an observed (expected) limit of  $4.4$   
 2239 ( $5.9$ ) times the Standard Model prediction (see Table 7.1).  $\kappa_\lambda$  is constrained to be within the  
 2240 range  $-4.9 \leq \kappa_\lambda \leq 14.4$  observed ( $-3.9 \leq \kappa_\lambda \leq 10.9$  expected). These results are shown in  
 2241 Figure 7.71.

2242 We note that this is a significant improvement over the early Run 2 result, which achieved  
 2243 an observed (expected) limit of  $12.9$  ( $20.7$ ) times the Standard Model prediction. The dataset  
 2244 is 4.6 times larger, and a naive scaling of the early Run 2 result (Poisson statistics  $\implies$  a factor  
 2245 of  $1/\sqrt{4.6}$ ) would predict an observed (expected) limit of  $6.0$  ( $9.7$ ) times the Standard Model.

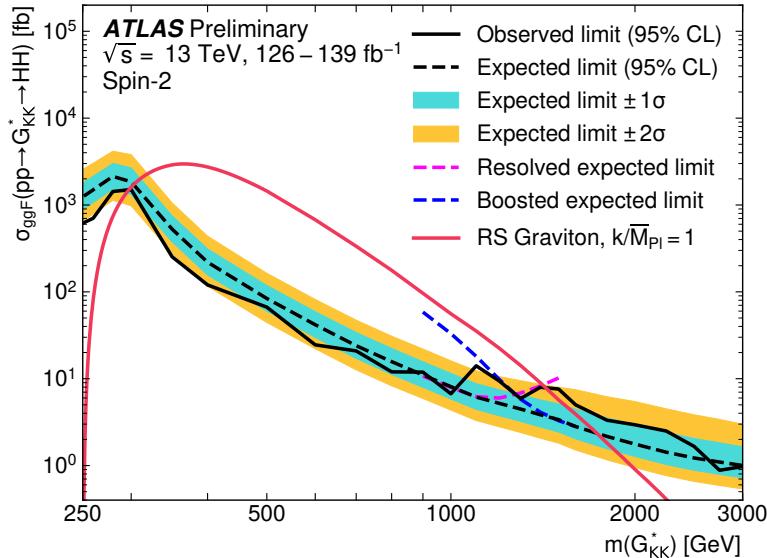
2246 The result of 4.4 (5.9) observed (expected) presented here is therefore both an improvement  
 2247 by a factor of 3 (3.5) over the previous result and also beats the statistical scaling by around  
 2248 30 (40) %, demonstrating the impact of the various analysis improvements presented here.  
 2249 We note again that these results do not include the complete set of uncertainties – however  
 2250 we expect the addition of the remaining uncertainties to have no more than a few percent  
 2251 impact.

2252 The observed limits presented in Figure 7.71 are consistently above the  $2\sigma$  band for values  
 2253 of  $\kappa_\lambda \geq 5$ , peaking at a local significance of  $3.8\sigma$  for  $\kappa_\lambda = 6$ . As this analysis is optimized for  
 2254 points near the Standard Model, and as there is no excess present in more sensitive channels  
 2255 in this same region (e.g.  $HH \rightarrow bb\gamma\gamma$  *TODO: include comparison*), we do not believe this is a  
 2256 real effect, but is rather due to a mis-modeling of the background at low mass, where the  
 2257 min  $\Delta R$  pairing has poor signal efficiency and the assumption of well behaved background in  
 2258 the mass plane breaks down. This is consistent with the location of the  $\kappa_\lambda = 6$  signal in  $m_{HH}$ ,  
 2259 as shown in Figures 7.68 and 7.69. It was considered, but not implemented, for this analysis  
 2260 to impose a cut on  $m_{HH}$  near 350 or 400 GeV to avoid such a low mass modeling issue.

2261 To check the impact of if we would have imposed such a cut, and to verify that the excess  
 2262 is due to the low mass regime, we therefore run the same set of limits without the low mass  
 2263 bins. In this case, we choose to simply drop the first few bins in  $m_{HH}$  such that everything  
 2264 else, including the higher mass bin edges, is kept the same. Due to the variable width binning,  
 2265 this corresponds to an  $m_{HH}$  cut of 381 GeV. The results of this check are shown in Figure  
 2266 7.72, overlaid with the limits of Figure 7.71 for reference. With the  $m_{HH}$  cut imposed, there  
 2267 is a slight degradation in the expected limits for larger positive and negative values of  $\kappa_\lambda$ ,  
 2268 but the points near the Standard Model are nearly identical. Further, the observed excess is  
 2269 significantly reduced, with observed limits for  $\kappa_\lambda \geq 5$  now falling entirely within the expected  
 2270  $1\sigma$  band. Due to the preliminary nature of these results, further study is left for future  
 2271 work. However, we believe, in conjunction with the  $HH \rightarrow bb\gamma\gamma$  results and our expectations  
 2272 about the difficulty of the background estimation at low mass, that this is demonstrative of a  
 2273 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  $k/\bar{M}_{Pl} = 1$  [19] (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.

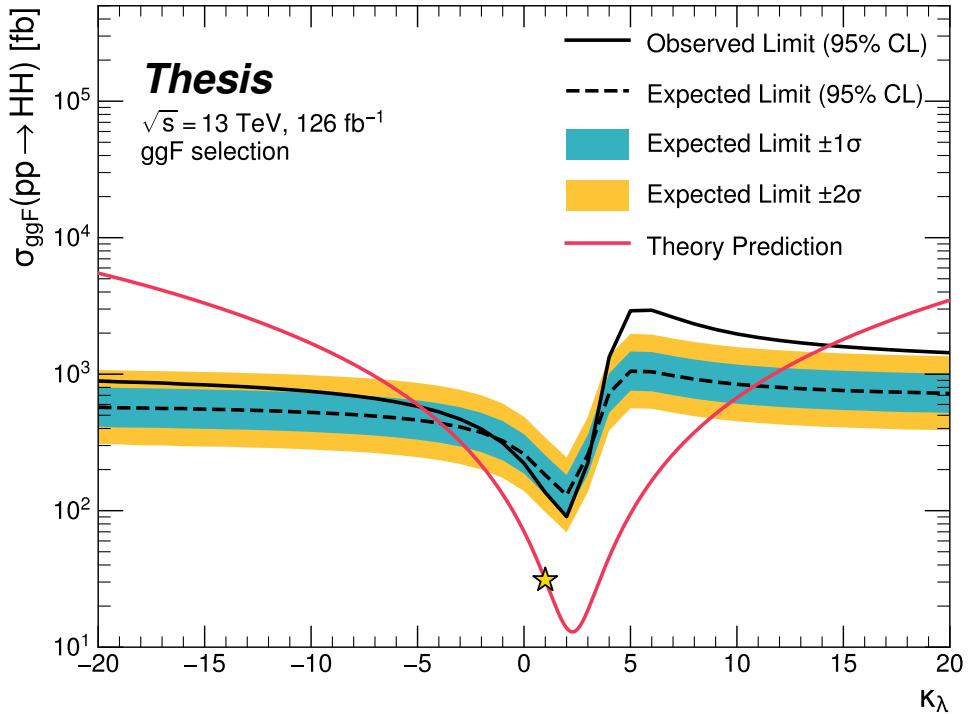


Figure 7.71: Expected (dashed black) and observed (solid black) 95% CL upper limits on the cross-section times branching ratio of non-resonant production for a range of values of the Higgs self-coupling, with the Standard Model value ( $\kappa_\lambda = 1$ ) illustrated with a star. The  $\pm 1\sigma$  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.  $\kappa_\lambda$  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.  $HH \rightarrow bb\gamma\gamma$ ).

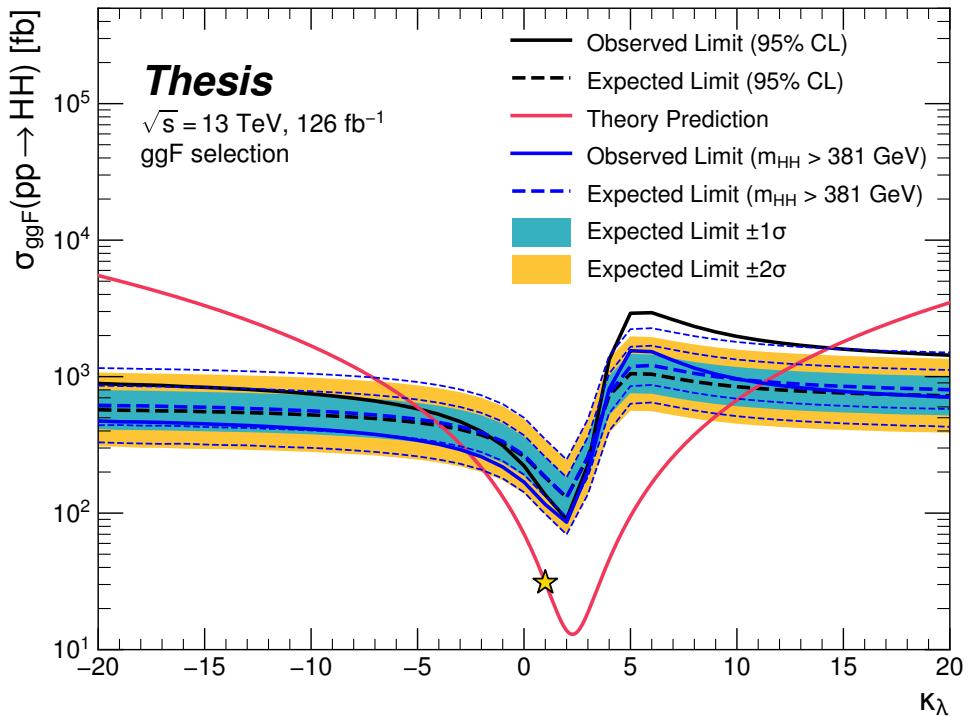


Figure 7.72: Comparison of the limits in Figure 7.71 with an equivalent set of limits that drop the  $m_{HH}$  bins below 381 GeV, with the value of 381 GeV determined by the optimized variable width binning. The expected limit band with this mass cut is shown in dashed blue, and the observed is shown 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.

2274 Chapter 8

2275 **FUTURE IDEAS FOR  $HH \rightarrow b\bar{b}b\bar{b}$**

2276 The searches presented in this thesis make use of a large suite of sophisticated techniques,  
2277 selected through careful study and validation. During this process, a variety of interesting  
2278 directions for the  $HH \rightarrow b\bar{b}b\bar{b}$  analysis were explored by this thesis author, in collaboration  
2279 with a few others<sup>1</sup>, but were not used due to a variety of constraints. We present two  
2280 such interesting directions here, with the hope of encouraging further exploration of these  
2281 techniques in future work.

2282 **8.1 pairAGraph: A New Method for Jet Pairing**

2283 As discussed in Chapter 7, one of the main problems to solve is the pairing of  $b$ -jets into  
2284 Higgs candidates. Figure 7.1 demonstrates that the choice of the pairing method, while  
2285 important for achieving good reconstruction of signal events, also significantly impacts the  
2286 structure of non- $HH$  events, leading to various biases in the background estimate. Evaluation  
2287 of the pairing method therefore must take both of these factors into account. While we have  
2288 presented some advantages in respective contexts for the pairing methods considered here,  
2289 we of course would like to explore further improvements to this important component of the  
2290 analysis.

2291 To that end, we note that all of the pairing methods considered here share a common  
2292 feature: four jets are selected, and the pairing is some discrimination between the available  
2293 three pairings of these four jets. For the methods used in this analysis, the jet selection  
2294 proceeds via a simple  $p_T$  ordering, with  $b$ -tagged jets receiving a higher priority than non-

---

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

2295 tagged jets.

2296 With the advent of a variety of machine learning methods for dealing with a variable  
 2297 number of inputs (e.g. recurrent neural networks [80], deep sets [81], graph neural networks [82],  
 2298 and transformers [83]), a natural place to improve on the pairing is to consider more than  
 2299 just four jets. The pairing and jet selection is then performed simultaneously, allowing for  
 2300 the incorporation of more event information in the pairing decision and the incorporation of  
 2301 jet correlation structure in the jet selection.

2302 In practice, the majority of  $HH \rightarrow b\bar{b}b\bar{b}$  events have either four or five jets which pass the  
 2303 kinematic preselection, and any gain from this additional freedom would come from events  
 2304 with greater than or equal to five jets. However, this five jet topology is particularly exciting  
 2305 for scenarios such as events with initial state radiation (ISR), in which the  $HH - > 4b$  jets  
 2306 are offset by a single jet with  $p_T$  similar in magnitude to that of the  $HH - > 4b$  system.  
 2307 Such events have explicit event level information which is not encoded with the inclusion  
 2308 of only the  $HH - > 4b$  jets, and are pathological if the ISR jet happens to pass  $b$ -tagging  
 2309 requirements.

2310 Additionally, with the use of lower tagged regions for background estimation and alternate  
 2311 signal regions, this extra flexibility in jet selection may provide a very useful bias – since the  
 2312 algorithm is trained on signal, the selected jets for the pairing will be the most “4b-like” jets  
 2313 available in the considered set.

2314 For the studies considered here, a transformer [83] based architecture is used. This is best  
 2315 visualized by considering the event as a graph with jets corresponding to nodes and edges  
 2316 corresponding to potential connections – for this reason, we term this algorithm “pairAGraph”.  
 2317 The approach is as follows: each jet,  $i$ , is represented by some vector of input variables,  $\vec{x}_i$ ,  
 2318 in our case the four-vector information,  $(p_T, \eta, \phi, E)$  of each jet, plus information on the  
 2319  $b$ -tagging decision. A multi-layer perceptron (MLP) is used to create a latent embedding,  
 2320  $\mathbf{h}(\vec{x}_i)$ , 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

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

where  $j$  runs over all jets in the event (including  $i = j$ ), the  $w_{ij}$  can be thought of as edge 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:

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

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

This similarity metric is calculated for all possible pairings, which are then passed through a softmax [84] activation function, which compresses these scores to between 0 and 1 with sum of 1, lending an interpretation as probability of each pairing.

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

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

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

2342 not required to use all of them in the construction of Higgs candidates, in contrast to the  
2343 other algorithms used in this thesis.

2344 **8.2 Background Estimation with Mass Plane Interpolation**

2345 The choice of a pairing algorithm that results in a smooth mass plane (such as  $\min \Delta R$ )  
2346 opens up a variety of options for the background estimation. While the method based on  
2347 reweighting of  $2b$  events used for this thesis performs well and has been extensively studied  
2348 and validated, it also relies on several assumptions. In particular, the reweighting is derived  
2349 between e.g.,  $2b$  and  $4b$  events *outside* of the signal region and then applied to  $2b$  events *inside*  
2350 the signal region, with the assumption that the  $2b$  to  $4b$  transfer function will be sufficiently  
2351 similar in both regions of the mass plane. An uncertainty is assigned to account for the bias  
2352 due to this assumption, but the extrapolation in the mass plane is never explicitly treated in  
2353 the nominal estimate. While the approach of reweighting  $2b$  events within the signal region  
2354 does have the advantage of incorporating explicit signal region information (that is, the  $2b$   
2355 signal region events), the importance of the extrapolation bias motivates consideration of  
2356 a method that operates within the  $4b$  mass plane. This additionally removes the reliance  
2357 on lower  $b$ -tagging regions, allowing for the use of, e.g.  $3b$  triggers, and future-proofing the  
2358 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:

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

2359 by the chain rule of probability. This means that the full dynamics of events in the  $HH$  mass  
2360 plane may be described by (1) the conditional probability of considered variables  $\vec{x}$ , given  
2361 values of  $m_{H1}$  and  $m_{H2}$ , and (2) the density of the mass plane itself.

2362 We present here an approach which uses normalizing flows *TODO: cite* to model the  
2363 conditional probabilities of events in the mass plane and Gaussian processes to model the  
2364 mass plane density. These models are trained in a region around, but not including, the

2365 signal region, and the trained models are then used to construct an *interpolated* estimate of  
 2366 the signal region kinematics. This approach therefore explicitly treats event behavior within  
 2367 the mass plane, avoiding the concerns associated with a reweighted estimate. Validation of  
 2368 such a method, as well as assessing of closure and biases of the method, may be done in  
 2369 alternate  $b$ -tagging or kinematic regions, notably the now unused  $2b$  region, results of which  
 2370 are shown below.

2371 *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

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

2372 where  $\left( \frac{d(f(x))}{dx} \right)$  is the Jacobian of  $f$  at  $x$ .

2373 The problem of normalizing flows then reduces to (1) choosing sets of  $f$  which are both  
 2374 tractable and sufficiently expressive to describe observed data, and (2) optimizing associated  
 2375 sets of functional parameters on observed data via maximum likelihood estimation using  
 2376 the above formula. Sampling from the learned density is done by drawing from the latent  
 2377 distribution  $z \sim p_Z$  (cf. inverse transform sampling) – the corresponding sample is then  
 2378  $x \sim p_X$  with  $x = f^{-1}(z)$ .

2379 A standard approach to the definition of these  $f$  is as a composition of affine transfor-  
 2380 mations (e.g. RealNVP *TODO: cite*), that is, transformations of the form  $\alpha z + \beta$ , with  $\alpha$  and  $\beta$   
 2381 learnable parameter vectors. This can roughly be thought of as shifting and squeezing the  
 2382 input prior density in order to match the data density. However, this has somewhat  
 2383 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  $\phi$  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  $m_{H2}$ .

2407 8.2.2 Gaussian Processes

2408 The second piece of this background estimate is the modeling of the mass plane density,  
2409  $p(m_{H1}, m_{H2})$ . This is done using Gaussian process regression – note that a similar procedure  
2410 is used to define a systematic in the boosted  $4b$  analysis. Generally, Gaussian processes  
2411 are a collection of random variables in which every finite collection of said variables is  
2412 distributed according to a multivariate normal distribution. For the context of Gaussian  
2413 process regression, what we consider is a Gaussian process over function space, that is, for a  
2414 collection of points,  $x_1, \dots, x_N$ , the space of corresponding function values,  $(f(x_1), \dots, f(x_N))$   
2415 is Gaussian process distributed, that is, described by an  $N$  dimensional normal distribution  
2416 with mean  $\mu$ , covariance matrix  $\Sigma$ .

2417 For a single point, this would correspond to a function space described entirely by a  
2418 normal distribution, with various samples from that distribution yielding various candidate  
2419 functions. For multiple points, a covariance matrix describes the relationship between each  
2420 pair of points – correspondingly, it is represented via a *kernel function*,  $K(x, x')$ . As, in  
2421 practice,  $\mu$  may always be set to 0 via a centering of the data, the kernel function fully defines  
2422 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

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

2423 where we have used that the kernel function is symmetric and assumed prior mean 0.

By standard conditioning properties of Gaussian distributions,

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

which is the sampling distribution for a Gaussian process given kernel  $K$ . In practice, the mean of this sampling distribution is used as the function estimate, with an uncertainty from 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

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

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

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

1. A binned 2d histogram of the blinded mass plane is created in a window around the “standard” analysis regions. Bins which have any overlap with the signal region are excluded.
2. A Gaussian process is trained using the bin centers, values as training points. The scikit-learn implementation [85] is used, with RBF kernel with anisotropic length scale ( $l$  is dimension 2). The length scale is initialized to  $(50, 50)$  to cover the signal region, and optimized by minimizing the negative log-marginal likelihood on the training data,  $-\log p(\vec{y}|\theta)$ . Training data is centered and scaled to mean 0, variance 1, and a statistical error is included in the fit.
3. The Gaussian process is then used to predict the density  $p(m_{H1}, m_{H2})$  in the signal region. This may then be sampled from via an inverse transform sampling to generate values  $(m_{H1}, m_{H2})$  according to the density (specifically, according to the mean of the Gaussian process posterior). Though in principle the Gaussian process sampling is not limited to bin centers, this is kept for simplicity, with a uniform smearing applied within

2445 each sampled bin to approximate the continuous estimate, namely, if a bin is sampled  
2446 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

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

2447 Note that the Monte Carlo sampling procedure is simply a set of samples of the Gaussian  
2448 process from uniformly random values of  $m_{H1}, m_{H2}$ , and is the most convenient approach  
2449 given the irregular shape of the signal region.

2450 This procedure results in a generated set of predicted  $m_{H1}, m_{H2}$  values for signal region  
2451 background events, along with an overall yield prediction.

### 2452 8.2.3 The Full Prediction

2453 Given the normalizing flow parametrization of  $p(x|m_{H1}, m_{H2})$  and the Gaussian process  
2454 generation of  $(m_{H1}, m_{H2}) \sim p(m_{H1}, m_{H2})$  and prediction of the signal region yield, all of the  
2455 pieces are in place to construct an interpolation background estimate. Namely

- 2456 1. Gaussian process sampled  $(m_{H1}, m_{H2})$  values are provided to the normalizing flow to  
2457 predict the other variables for the Higgs candidate four-vectors. These are used to  
2458 construct the  $HH$  system (notably  $m_{HH}, \cos \theta^*$ ).
- 2459 2. These final distributions are normalized according to the predicted background yield.

2460 8.2.4 Results

2461 The Gaussian process sampling procedure is trained on a small fraction (0.01) of  $2b$  data to  
 2462 mimic the available  $4b$  statistics. This fraction of  $2b$  data is blinded, and the prediction of the  
 2463 estimate trained on this blinded region may then be compared to real  $2b$  data in the signal  
 2464 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.

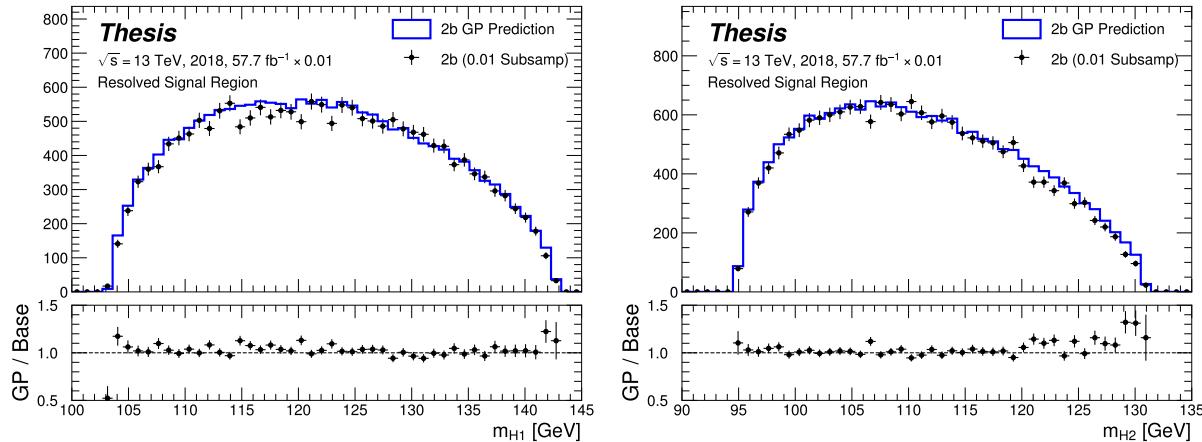


Figure 8.1: Gaussian process sampling prediction of marginals  $m_{H_1}$  and  $m_{H_2}$  for  $2b$  signal region events compared to real  $2b$  signal region events for the 2018 dataset. Good agreement is seen. Only a small fraction (0.01) of the  $2b$  dataset is used for both training and this final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

2465

2466 The  $4b$  region is kept blinded for this work, meaning that a direct comparison of the  
 2467 Gaussian process estimate in the  $4b$  signal region is not done. However, a Gaussian process is  
 2468 trained on the blinded  $4b$  region and compared to the corresponding reweighted  $2b$  estimate,  
 2469 trained per the nominal procedures from the analyses above. The predictions for signal  
 2470 region  $m_{H_1}$  and  $m_{H_2}$  individually are shown in Figure 8.3, compared to both the control and  
 2471 validation region derived reweighting estimates, and the resulting signal region mass planes  
 2472 are compared in Figure 8.4. The estimates are seen to be compatible.

2473    8.2.5 *Outstanding Points*

2474    While good performance is demonstrated from the nominal interpolated background estimate,  
2475    various uncertainties must be assigned according to the various stages of the estimate. These  
2476    notably include

2477        • Assessing a statistical uncertainty on the normalizing flow training (cf. bootstrap  
2478              uncertainty).

2479        • Propagation of the Gaussian process uncertainty through the sampling procedure.

2480        • Validation of the resulting estimate and assessment of necessary systematic uncertainties  
2481              (e.g. from validation region non-closure).

2482    These are all quite tractable, but some, especially the choice of an appropriate systematic  
2483    uncertainty, are certainly not obvious and require detailed study. In this respect, the  
2484    reweighting validation work of the non-resonant analysis is certainly quite useful as a starting  
2485    place in terms of the available regions and their correspondence to the nominal  $4b$  signal  
2486    region.

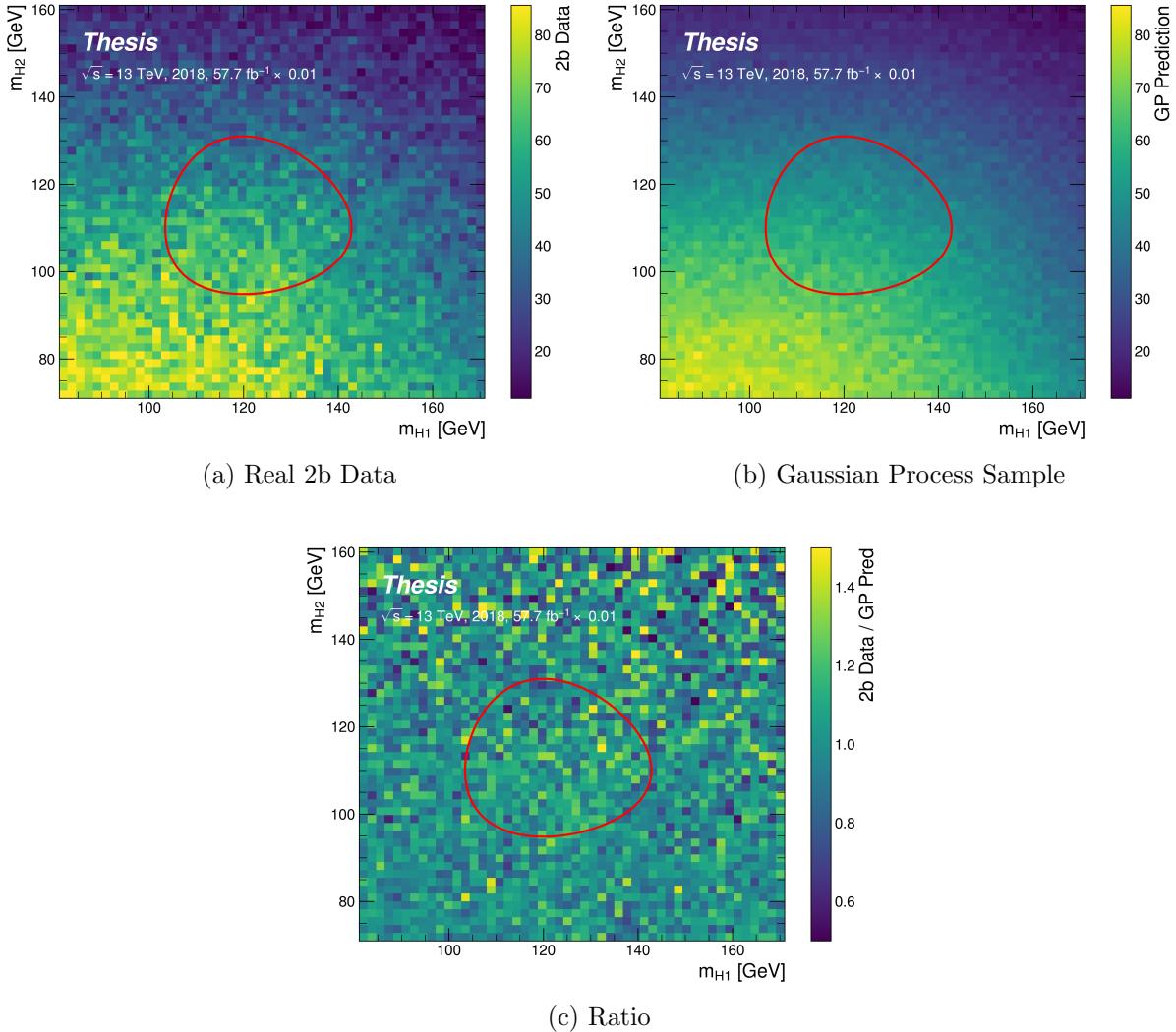


Figure 8.2: Gaussian process sampling prediction for the mass plane compared to the real 2b dataset for 2018. Only a small fraction (0.01) of the 2b dataset is used for both training and this final comparison to mimic 4b statistics. Good agreement is seen.

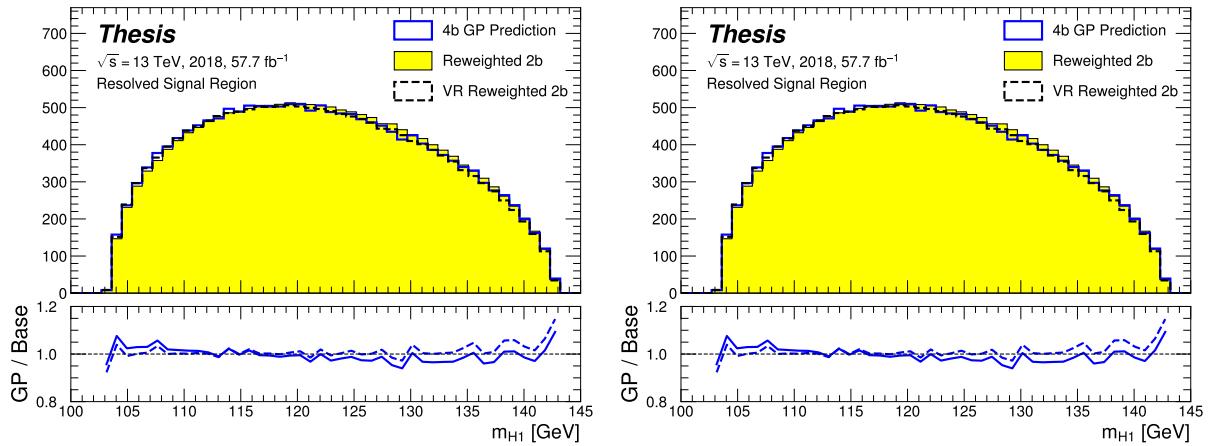


Figure 8.3: Gaussian process sampling prediction of marginals  $m_{H1}$  and  $m_{H2}$  for 4b signal region events compared to both control and validation reweighting predictions. While there are some differences, the estimates are compatible.

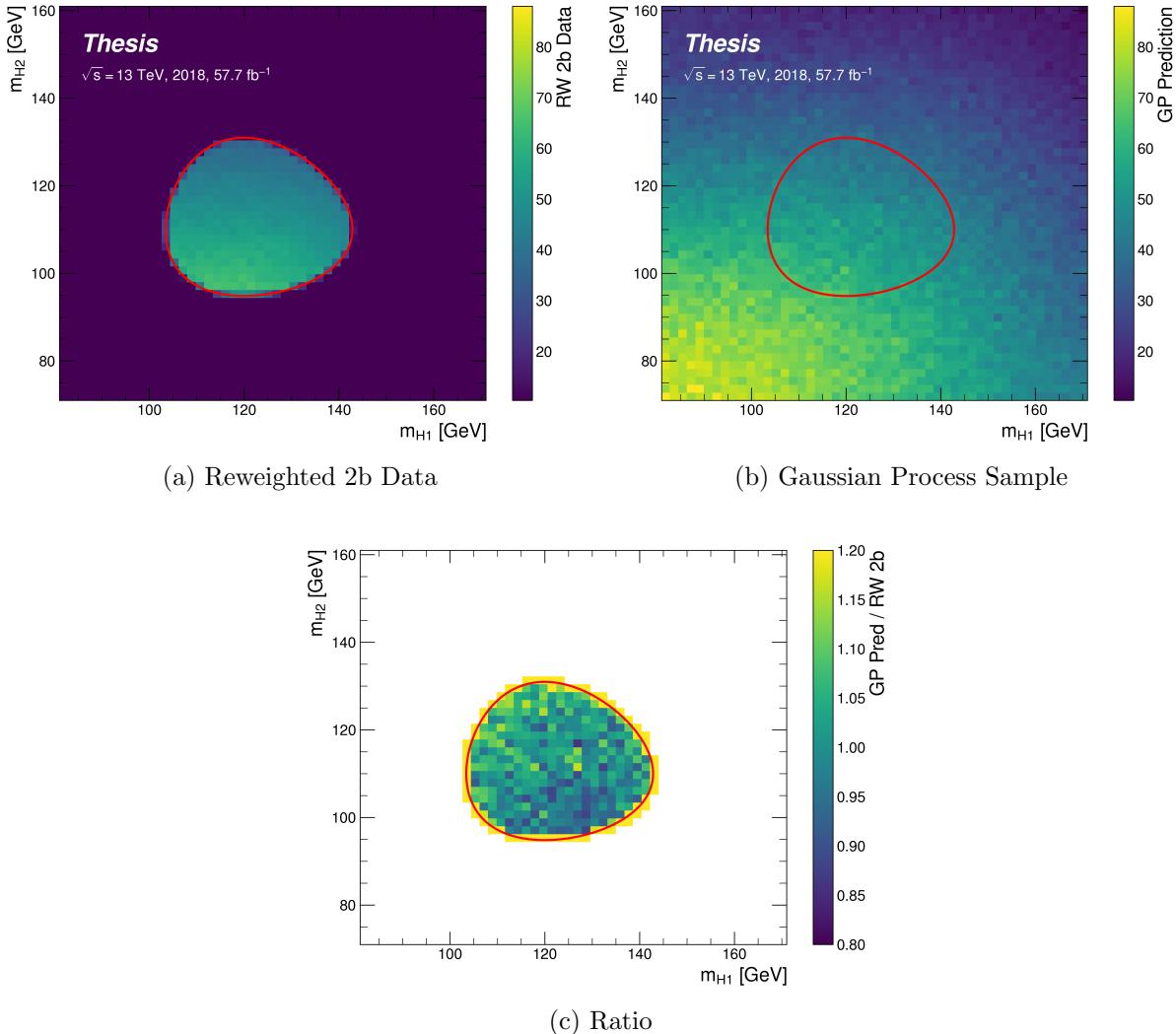


Figure 8.4: Gaussian process sampling prediction for the  $4b$  mass plane compared to the reweighted  $2b$  estimate in the signal region. Both estimates are compatible.

2487

## Chapter 9

2488

### CONCLUSIONS

2489 This thesis has provided an overview of the Standard Model, with an emphasis on pair  
2490 production of Higgs bosons and how this process may be used to both verify the Standard  
2491 Model and to search for new physics. An overview of the Large Hadron Collider and the  
2492 ATLAS detector has been provided, and the design and use of simulation infrastructure  
2493 has been explained, including work to improve hadronic shower modeling in fast detector  
2494 simulation. The translation of detector level information to analysis level information has  
2495 been explained, with an emphasis on jets and the identification of  $B$  hadron decay. Finally,  
2496 two searches for Higgs boson pair production have been presented, with a complete set of  
2497 results for resonant production included, focusing on searches beyond the Standard Model,  
2498 and a preliminary set of results for non-resonant production, targeting Standard Model  
2499 production, with variations of the Higgs self-coupling. Two advanced techniques for the  
2500 future of these analyses are further presented, along with proof-of-concept results.

2501

## BIBLIOGRAPHY

- 2502 [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:  
2503 2021-07-23 (cit. on p. 3).
- 2504  
2505 [2] J. D. Jackson, *Classical electrodynamics; 2nd ed.* Wiley, 1975 (cit. on p. 2).
- 2506 [3] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 978-1-  
2507 107-03426-6 (cit. on p. 2).
- 2508 [4] M. Srednicki, *Quantum Field Theory*, Cambridge Univ. Press, 2007 (cit. on p. 2).
- 2509 [5] M. Gell-Mann, *A schematic model of baryons and mesons*, Physics Letters **8** (1964)  
2510 214, ISSN: 0031-9163, URL: <https://www.sciencedirect.com/science/article/pii/S0031916364920013> (cit. on p. 11).
- 2511  
2512 [6] G. Zweig, *An  $SU_3$  model for strong interaction symmetry and its breaking; Version 1*,  
2513 tech. rep., CERN, 1964, URL: <https://cds.cern.ch/record/352337> (cit. on p. 11).
- 2514 [7] O. W. Greenberg, *Spin and Unitary-Spin Independence in a Paraquark Model of Baryons  
2515 and Mesons*, Phys. Rev. Lett. **13** (20 1964) 598, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.598> (cit. on p. 11).
- 2516  
2517 [8] M. Y. Han and Y. Nambu, *Three-Triplet Model with Double  $SU(3)$  Symmetry*, Phys.  
2518 Rev. **139** (4B 1965) B1006, URL: <https://link.aps.org/doi/10.1103/PhysRev.139.B1006> (cit. on p. 11).
- 2519  
2520 [9] F. L. Wilson, *Fermi's Theory of Beta Decay*, American Journal of Physics **36** (1968)  
2521 1150, eprint: <https://doi.org/10.1119/1.1974382>, URL: <https://doi.org/10.1119/1.1974382> (cit. on p. 14).
- 2522

- 2523 [10] *The Nobel Prize in Physics 1979*, <https://www.nobelprize.org/prizes/physics/1979/summary/>, Accessed: 2021-07-23 (cit. on p. 15).
- 2524
- 2525 [11] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,  
2526 Phys. Rev. Lett. **13** (9 1964) 321, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on p. 20).
- 2527
- 2528 [12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett.  
2529 **13** (16 1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>  
2530 (cit. on p. 20).
- 2531 [13] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and  
2532 Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585> (cit. on p. 20).
- 2533
- 2534 [14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard  
2535 Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,  
2536 arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214) (cit. on p. 20).
- 2537
- 2538 [15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS  
2539 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).
- 2540
- 2541 [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).
- 2542
- 2543 [17] ATLAS Collaboration, *Search for the  $HH \rightarrow b\bar{b}b\bar{b}$  process via vector-boson fusion  
2544 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, 89), Erratum: JHEP  
2545 **01** (2021) 145.
- 2546
- 2547 [18] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, *Warped gravitons at the CERN LHC  
2548 and beyond*, Phys. Rev. D **76** (3 2007) 036006, URL: <https://link.aps.org/doi/10.1103/PhysRevD.76.036006> (cit. on p. 29).

- 2549 [19] A. Carvalho, *Gravity particles from Warped Extra Dimensions, predictions for LHC*,  
 2550 2018, arXiv: 1404.0102 [hep-ph] (cit. on pp. 29, 174).
- 2551 [20] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions*  
 2552 *at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector*, Phys. Lett. B **800** (2020) 135103, arXiv:  
 2553 1906.02025 [hep-ex] (cit. on pp. 30, 69).
- 2554 [21] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Physics  
 2555 Reports **516** (2012) 1, Theory and phenomenology of two-Higgs-doublet models,  
 2556 ISSN: 0370-1573, URL: <http://www.sciencedirect.com/science/article/pii/S0370157312000695> (cit. on p. 30).
- 2558 [22] P. D. Group et al., *Review of Particle Physics*, Progress of Theoretical and Experimental  
 2559 Physics **2020** (2020), 083C01, ISSN: 2050-3911, eprint: <https://academic.oup.com/p gep/ article-pdf/2020/8/083C01/34673722/ptaa104.pdf>, URL: <https://doi.org/10.1093/p gep/ptaa104> (cit. on p. 31).
- 2562 [23] T. van Ritbergen and R. G. Stuart, *On the precise determination of the Fermi coupling*  
 2563 *constant from the muon lifetime*, Nuclear Physics B **564** (2000) 343, ISSN: 0550-3213,  
 2564 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. 31).
- 2565 [24] S. Dawson, A. Ismail, and I. Low, *What's in the loop? The anatomy of double Higgs*  
 2566 *production*, Physical Review D **91** (2015), ISSN: 1550-2368, URL: <http://dx.doi.org/10.1103/PhysRevD.91.115008> (cit. on p. 31).
- 2568 [25] A. M. Sirunyan et al., *Measurement of the top quark Yukawa coupling from  $t\bar{t}$  kinematic*  
 2569 *distributions in the dilepton final state in proton-proton collisions at  $\sqrt{s}=13 \text{ TeV}$* ,  
 2570 Physical Review D **102** (2020), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.102.092013> (cit. on p. 31).
- 2572 [26] J. Pequenao, “Computer generated image of the whole ATLAS detector”, 2008, URL:  
 2573 <https://cds.cern.ch/record/1095924> (cit. on p. 35).

- 2574 [27] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,  
 2575 JINST **3** (2008) S08003 (cit. on p. 36).
- 2576 [28] J. Pequenao and P. Schaffner, “How ATLAS detects particles: diagram of particle paths  
 2577 in the detector”, 2013, URL: <https://cds.cern.ch/record/1505342> (cit. on p. 37).
- 2578 [29] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-  
 2579 TDR-19; CERN-LHCC-2010-013, 2010, URL: <https://cds.cern.ch/record/1291633>  
 2580 (cit. on p. 38), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL:  
 2581 <https://cds.cern.ch/record/1451888>.
- 2582 [30] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST  
 2583 **13** (2018) T05008, arXiv: [1803.00844 \[physics.ins-det\]](https://arxiv.org/abs/1803.00844) (cit. on p. 38).
- 2584 [31] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys.  
 2585 J. C **77** (2017) 317, arXiv: [1611.09661 \[hep-ex\]](https://arxiv.org/abs/1611.09661) (cit. on p. 41).
- 2586 [32] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-  
 2587 PUB-2021-001, 2021, URL: <https://cds.cern.ch/record/2767187> (cit. on p. 41).
- 2588 [33] J. Alwall et al., *The automated computation of tree-level and next-to-leading order  
 2589 differential cross sections, and their matching to parton shower simulations*, Journal of  
 2590 High Energy Physics **2014** (2014), URL: [https://doi.org/10.1007/jhep07\(2014\)079](https://doi.org/10.1007/jhep07(2014)079)  
 2591 (cit. on pp. 44, 71).
- 2592 [34] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*,  
 2593 JHEP **11** (2004) 040, arXiv: [hep-ph/0409146 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0409146) (cit. on pp. 45, 72).
- 2594 [35] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton  
 2595 shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: [0709.2092](https://arxiv.org/abs/0709.2092)  
 2596 [hep-ph] (cit. on pp. 45, 72).
- 2597 [36] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO  
 2598 calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010)  
 2599 043, arXiv: [1002.2581 \[hep-ph\]](https://arxiv.org/abs/1002.2581) (cit. on pp. 45, 72).

- 2600 [37] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, The European Physical Journal  
2601 C **76** (2016), URL: <https://doi.org/10.1140/epjc/s10052-016-4018-8> (cit. on  
2602 pp. 45, 71).
- 2603 [38] M. Bähr et al., *Herwig++ physics and manual*, The European Physical Journal C **58**  
2604 (2008) 639, URL: <https://doi.org/10.1140/epjc/s10052-008-0798-9> (cit. on  
2605 pp. 45, 71).
- 2606 [39] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications  
2607 **191** (2015) 159, ISSN: 0010-4655, URL: <https://www.sciencedirect.com/science/article/pii/S0010465515000442> (cit. on pp. 45, 71).
- 2609 [40] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A  
2610 **462** (2001) 152 (cit. on pp. 45, 71).
- 2611 [41] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods  
2612 in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated  
2613 Equipment **506** (2003) 250, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688> (cit. on p. 46).
- 2615 [42] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70**  
2616 (2010) 823, arXiv: [1005.4568 \[physics.ins-det\]](https://arxiv.org/abs/1005.4568) (cit. on pp. 46, 71).
- 2617 [43] ATLAS Collaboration, *The new Fast Calorimeter Simulation in ATLAS*, ATL-SOFT-  
2618 PUB-2018-002, 2018, URL: <https://cds.cern.ch/record/2630434> (cit. on pp. 47,  
2619 48).
- 2620 [44] M. Aharrouche et al., *Energy linearity and resolution of the ATLAS electromagnetic*  
2621 *barrel calorimeter in an electron test-beam*, Nuclear Instruments and Methods in Physics  
2622 Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment  
2623 **568** (2006) 601, ISSN: 0168-9002, URL: <http://dx.doi.org/10.1016/j.nima.2006.07.053> (cit. on p. 48).

- 2625 [45] ATLAS Collaboration, *Evidence for prompt photon production in pp collisions at*  
 2626  $\sqrt{s} = 7 \text{ TeV}$  *with the ATLAS detector*, ATLAS-CONF-2010-077, 2010, URL: <https://cds.cern.ch/record/1281368> (cit. on p. 48).
- 2628 [46] D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, 2014, arXiv: [1312.6114](https://arxiv.org/abs/1312.6114)  
 2629 [[stat.ML](#)] (cit. on p. 49).
- 2630 [47] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the*  
 2631 *ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: [1703.10485](https://arxiv.org/abs/1703.10485) [[hep-ex](#)] (cit. on  
 2632 pp. 56, 70).
- 2633 [48] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its*  
 2634 *performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: [1603 . 02934](https://arxiv.org/abs/1603.02934)  
 2635 [[hep-ex](#)] (cit. on p. 56).
- 2636 [49] ATLAS Collaboration, *Variable Radius, Exclusive- $k_T$ , and Center-of-Mass Subjet Re-*  
 2637 *construction for Higgs( $\rightarrow b\bar{b}$ ) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017,  
 2638 URL: <https://cds.cern.ch/record/2268678> (cit. on p. 56).
- 2639 [50] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- $k_T$  jet clustering algorithm*, Journal of  
 2640 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)  
 2641 [1088/1126-6708/2008/04/063](#) (cit. on p. 56).
- 2642 [51] G. P. Salam, *Towards jetography*, The European Physical Journal C **67** (2010) 637, ISSN:  
 2643 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-010-1314-6> (cit. on  
 2644 p. 56).
- 2645 [52] A. Collaboration, *Configuration and performance of the ATLAS b-jet triggers in Run 2*,  
 2646 2021, arXiv: [2106.03584](https://arxiv.org/abs/2106.03584) [[hep-ex](#)] (cit. on p. 60).
- 2647 [53] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measure-*  
 2648 *ment with  $t\bar{t}$  events in pp collisions at  $\sqrt{s} = 13 \text{ TeV}$* , Eur. Phys. J. C **79** (2019) 970,  
 2649 arXiv: [1907.05120](https://arxiv.org/abs/1907.05120) [[hep-ex](#)] (cit. on pp. 59, 70).

- 2650 [54] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification of*  
 2651 *b-jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-PHYS-  
 2652 PUB-2018-025, 2018, URL: <https://cds.cern.ch/record/2645405> (cit. on p. 62).
- 2653 [55] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear  
 2654 Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,  
 2655 Detectors and Associated Equipment **262** (1987) 444, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874> (cit. on  
 2656 p. 62).
- 2657 [56] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*  
 2658 *Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, URL:  
 2659 <https://cds.cern.ch/record/2255226> (cit. on p. 63).
- 2660 [57] *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. 65).
- 2661 [58] ATLAS Collaboration, *Search for Higgs boson pair production in the  $b\bar{b}WW^*$  decay mode*  
 2662 *at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector*, JHEP **04** (2019) 092, arXiv: [1811.04671](https://arxiv.org/abs/1811.04671)  
 2663 [hep-ex] (cit. on p. 69).
- 2664 [59] ATLAS Collaboration, *A search for resonant and non-resonant Higgs boson pair pro-*  
 2665 *duction in the  $b\bar{b}\tau^+\tau^-$  decay channel in  $pp$  collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS*  
 2666 *detector*, Phys. Rev. Lett. **121** (2018) 191801, arXiv: [1808.00336](https://arxiv.org/abs/1808.00336) [hep-ex] (cit. on  
 2667 p. 69), Erratum: Phys. Rev. Lett. **122** (2019) 089901.
- 2668 [60] ATLAS Collaboration, *Search for Higgs boson pair production in the  $WW^{(*)}WW^{(*)}$*   
 2669 *decay channel using ATLAS data recorded at  $\sqrt{s} = 13 \text{ TeV}$* , JHEP **05** (2019) 124, arXiv:  
 2670 [1811.11028](https://arxiv.org/abs/1811.11028) [hep-ex] (cit. on p. 69).
- 2671 [61] ATLAS Collaboration, *Search for Higgs boson pair production in the  $\gamma\gamma b\bar{b}$  final state*  
 2672 *with  $13 \text{ TeV}$   $pp$  collision data collected by the ATLAS experiment*, JHEP **11** (2018) 040,  
 2673 arXiv: [1807.04873](https://arxiv.org/abs/1807.04873) [hep-ex] (cit. on p. 69).
- 2674
- 2675

- [62] ATLAS Collaboration, *Search for Higgs boson pair production in the  $\gamma\gamma WW^*$  channel 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 \[hep-ex\]](https://arxiv.org/abs/1807.08567) (cit. on p. 69).
- [63] CMS Collaboration, *Search for heavy resonances decaying to two Higgs bosons in final states containing four  $b$  quarks*, *Eur. Phys. J. C* **76** (2016) 371, arXiv: [1602.08762 \[hep-ex\]](https://arxiv.org/abs/1602.08762) (cit. on p. 69).
- [64] CMS Collaboration, *Search for production of Higgs boson pairs in the four  $b$  quark final state using large-area jets in proton–proton collisions at  $\sqrt{s} = 13 \text{ TeV}$* , *JHEP* **01** (2019) 040, arXiv: [1808.01473 \[hep-ex\]](https://arxiv.org/abs/1808.01473) (cit. on p. 69).
- [65] CMS Collaboration, *Combination of Searches for Higgs Boson Pair Production in Proton–Proton Collisions at  $\sqrt{s} = 13 \text{ TeV}$* , *Phys. Rev. Lett.* **122** (2019) 121803, arXiv: [1811.09689 \[hep-ex\]](https://arxiv.org/abs/1811.09689) (cit. on p. 70).
- [66] G. Bartolini et al., *Performance of the ATLAS  $b$ -jet trigger in  $p\ p$  collisions at  $\sqrt{s} = 13 \text{ TeV}$* , tech. rep. ATL-COM-DAQ-2019-150, CERN, 2019, URL: <https://cds.cern.ch/record/2688819> (cit. on p. 70).
- [67] ATLAS Collaboration, *Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 13 \text{ TeV}$  using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: <https://cds.cern.ch/record/2677054> (cit. on p. 71).
- [68] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43** (2016) 023001, arXiv: [1510.03865 \[hep-ph\]](https://arxiv.org/abs/1510.03865) (cit. on p. 72).
- [69] M. Bahr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J. C* **58** (2008) 639, arXiv: [0803.0883 \[hep-ph\]](https://arxiv.org/abs/0803.0883) (cit. on p. 72).
- [70] 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. 77).

- 2700 [71] ATLAS Collaboration, *Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state*  
 2701 *using proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, JHEP **01**  
 2702 (2019) 030, arXiv: [1804.06174 \[hep-ex\]](https://arxiv.org/abs/1804.06174) (cit. on pp. 83, 89).
- 2703 [72] A. Carvalho et al., *Higgs pair production: choosing benchmarks with cluster analysis*,  
 2704 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. 87).
- 2706 [73] G. V. Moustakides and K. Basioti, *Training Neural Networks for Likelihood/Density*  
 2707 *Ratio Estimation*, 2019, arXiv: [1911.00405 \[eess.SP\]](https://arxiv.org/abs/1911.00405) (cit. on p. 90).
- 2708 [74] T. Kanamori, S. Hido, and M. Sugiyama, *A Least-Squares Approach to Direct Importance*  
 2709 *Estimation*, J. Mach. Learn. Res. **10** (2009) 1391, ISSN: 1532-4435 (cit. on p. 90).
- 2710 [75] S. Fort, H. Hu, and B. Lakshminarayanan, *Deep Ensembles: A Loss Landscape Perspec-*  
 2711 *tive*, 2020, arXiv: [1912.02757 \[stat.ML\]](https://arxiv.org/abs/1912.02757) (cit. on p. 92).
- 2712 [76] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7** (1979) 1,  
 2713 URL: <https://doi.org/10.1214/aos/1176344552> (cit. on p. 149).
- 2714 [77] J. Adelman et al., *Search for Higgs boson pair production in the  $b\bar{b}\gamma\gamma$  final state with*  
 2715 *the full Run 2 13 TeV pp collision data collected by the ATLAS Experiment Supporting*  
 2716 *note.*, tech. rep., CERN, 2020, URL: <https://cds.cern.ch/record/2711865> (cit. on  
 2717 p. 157).
- 2718 [78] A. L. Read, *Presentation of search results: the CLs technique*, Journal of Physics G:  
 2719 Nuclear and Particle Physics **28** (2002) 2693, ISSN: 0954-3899 (cit. on p. 171).
- 2720 [79] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-*  
 2721 *based tests of new physics*, The European Physical Journal C **71** (2011), ISSN: 1434-6052  
 2722 (cit. on p. 171).

- 2723 [80] A. Sherstinsky, *Fundamentals of Recurrent Neural Network (RNN) and Long Short-*  
2724 *Term Memory (LSTM) network*, Physica D: Nonlinear Phenomena **404** (2020) 132306,  
2725 ISSN: 0167-2789, URL: <http://dx.doi.org/10.1016/j.physd.2019.132306> (cit. on  
2726 p. 178).
- 2727 [81] M. Zaheer et al., *Deep Sets*, 2018, arXiv: [1703.06114 \[cs.LG\]](https://arxiv.org/abs/1703.06114) (cit. on p. 178).
- 2728 [82] J. Zhou et al., *Graph Neural Networks: A Review of Methods and Applications*, 2021,  
2729 arXiv: [1812.08434 \[cs.LG\]](https://arxiv.org/abs/1812.08434) (cit. on p. 178).
- 2730 [83] A. Vaswani et al., *Attention Is All You Need*, 2017, arXiv: [1706.03762 \[cs.CL\]](https://arxiv.org/abs/1706.03762) (cit. on  
2731 p. 178).
- 2732 [84] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, <http://www.deeplearningbook.org>,  
2733 MIT Press, 2016 (cit. on p. 179).
- 2734 [85] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, Journal of Machine  
2735 Learning Research **12** (2011) 2825 (cit. on p. 184).