



**ATLAS Note**

GROUP-2017-XX



19th November 2020

Draft version 0.1

1

2      **A Deep Learning Approach to Differential**  
3      **Measurements of Higgs - Top Interactions in**  
4      **Multilepton Final States**

5

The ATLAS Collaboration

6 Several theories Beyond the Standard Model modify the momentum spectrum of the Higgs  
7 Boson, without significantly altering the overall rate of Higgs produced from top quark pairs.  
8 A differential measurement of the Higgs transverse momentum provides a way to search for  
9 these effects at the LHC. Because of the challenges inherent to reconstructing the Higgs in  
10 multilepton final states, a deep learning approach is used to predict the momentum spectrum  
11 of the Higgs for events where the Higgs Boson is produced from top quark pairs and decays  
12 to final states that include multiple leptons. The regressed Higgs  $p_T$  is fit to data for events  
13 with two or three leptons in the final state, and estimates of the sensitivity to variations in the  
14 Higgs  $p_T$  spectrum are given.

© 2020 CERN for the benefit of the ATLAS Collaboration.

15 Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

# 16 **Contents**

17	<b>I Introduction</b>	<b>4</b>
18	<b>1 Introduction</b>	<b>4</b>
19	<b>II Theoretical Motivation</b>	<b>6</b>
20	<b>2 The Standard Model and the Higgs Boson</b>	<b>6</b>
21	2.1 The Forces and Particles of the Standard Model	7
22	2.2 The Higgs Mechanism	9
23	2.2.1 The Higgs Field	10
24	2.2.2 Electroweak Symmetry Breaking	12
25	2.3 Limitations of the Standard Model	14
26	<b>3 Effective Field Theory in <math>t\bar{t}H</math> Production</b>	<b>15</b>
27	3.1 Extensions to the Higgs Sector	15
28	3.2 Six Dimensional Operators	15
29	<b>III The LHC and the ATLAS Detector</b>	<b>16</b>
30	<b>4 The LHC</b>	<b>16</b>
31	<b>5 The ATLAS Detector</b>	<b>19</b>
32	5.1 Inner Detector	20
33	5.2 Calorimeters	21
34	5.3 Muon Spectrometer	22
35	5.4 Trigger System	23
36	<b>IV Search for Dimension-Six Operators</b>	<b>24</b>
37	<b>6 Data and Monte Carlo Samples</b>	<b>24</b>
38	6.1 Data Samples	25
39	6.2 Monte Carlo Samples	26
40	<b>7 Object Reconstruction</b>	<b>26</b>
41	7.1 Trigger Requirements	27
42	7.2 Light Leptons	27
43	7.3 Jets	28
44	7.4 Missing Transverse Energy	29
45	<b>8 Higgs Momentum Reconstruction</b>	<b>29</b>

46	8.1	Decay Candidate Reconstruction	31
47	8.2	b-jet Identification	31
48	8.2.1	2lSS Channel	32
49	8.2.2	3l Channel	36
50	8.3	Higgs Reconstruction	39
51	8.3.1	2lSS Channel	40
52	8.3.2	3l Semi-leptonic Channel	44
53	8.3.3	3l Fully-leptonic Channel	48
54	8.4	$p_T$ Prediction	51
55	8.4.1	2lSS Channel	53
56	8.4.2	3l Semi-leptonic Channel	56
57	8.4.3	3l Fully-leptonic Channel	58
58	8.5	3l Decay Mode	62
59	<b>9</b>	<b>Signal Region Definitions</b>	<b>64</b>
60	9.1	Pre-MVA Event Selection	65
61	9.2	Event MVA	68
62	9.3	Signal Region Definitions	74
63	<b>10</b>	<b>Background Rejection MVA</b>	<b>78</b>
64	10.1	Pre-MVA Event Selection	79
65	10.2	Event MVA	80
66	10.3	Signal Region Definitions	86
67	10.3.1	2lSS	89
68	10.3.2	3l – Semi – leptonic	89
69	10.3.3	3l – Fully – leptonic	89
70	<b>11</b>	<b>Systematic Uncertainties</b>	<b>89</b>
71	<b>12</b>	<b>Results</b>	<b>91</b>
72	12.1	Results - $80 \text{ fb}^{-1}$	91
73	12.2	Projected Results - $140 \text{ fb}^{-1}$	94
74	<b>V</b>	<b>Conclusion</b>	<b>96</b>
75	<b>Appendices</b>		<b>98</b>
76	<b>A</b>	<b>Machine Learning Models</b>	<b>98</b>
77	A.1	Higgs Reconstruction Models	98
78	A.1.1	b-jet Identification Features - 2lSS	98
79	A.1.2	b-jet Identification Features - 3l	102
80	A.1.3	Higgs Reconstruction Features - 2lSS	107
81	A.1.4	Higgs Reconstruction Features - 3lS	112

82	A.1.5 Higgs Reconstruction Features - 3lF	117
83	A.2 Background Rejection MVAs	121
84	A.2.1 Background Rejection MVA Features - 2lSS	121
85	A.2.2 Background Rejection MVA Features - 3l	125
86	A.3 Alternate b-jet Identification Algorithm	130
87	A.4 Binary Classification of the Higgs $p_T$	130
88	A.5 Impact of Alternative Jet Selection	132

---

**89 Part I****90 Introduction****91 1 Introduction**

92 Particle physics is an attempt to describe the fundamental building blocks of the universe and  
93 their interactions. The Standard Model (SM) - our best current theory of fundamental particle  
94 physics - does a remarkable job of that. All known fundamental particles and (almost) all of the  
95 forces underlying their interactions can be explained by the SM, and the predictions from this  
96 theory agree with experiment to an incredibly precise degree. This is especially true since the  
97 Higgs Boson, the last piece of the SM predicted decades before, was finally discovered at the  
98 Large Hadron Collider (LHC) in 2012.

99 Despite the success of the SM, there remains significant work to be done. For one, the  
100 SM is incomplete: it fails to provide a description of gravity, to give an explanation for the  
101 observation of Dark Matter, or to provide a mechanism for neutrinos to gain mass. Further, a  
102 Higgs Boson with a mass of around 125 GeV, as observed at the LHC, gives rise to what is  
103 known a hierarchy problem - such a low mass Higgs requires a seemingly unnatural level of “fine  
104 tuning” that is unexplained by the SM.

105 A promising avenue for addressing these problems is to study the properties of the Higgs  
106 Boson and the way it interacts with other particles, in part simply because these interactions

107 have not been measured before. Its interactions with the Top Quark are a particularly promising  
108 place to look. Because the Higgs Field is responsible for allowing particle to acquire mass, the  
109 strength of a particle's interaction with the Higgs Boson is proportional to its mass. As the most  
110 massive of the fundamental particles, the Top Quark has the strongest coupling to the Higgs  
111 Boson, meaning any new physics in the Higgs sector is likely to present itself most prominently  
112 in its interaction with the Top Quark.

113 These interactions can be measured by directly by studying the production of a Higgs  
114 Boson in association with a pair of Top Quarks ( $t\bar{t}H$ ). While studies have been done measuring  
115 the overall rate of  $t\bar{t}H$  production, there are several theories of physics Beyond the Standard  
116 Model (BSM) that would affect the kinematics of  $t\bar{t}H$  production without altering its overall  
117 rate. This dissertation attempts to make a differential measurement of the kinematics of the  
118 Higgs Boson in  $t\bar{t}H$  events in order to search for these BSM effects.

119 An Effective Field Theory model can be used to model the low energy effects of high  
120 energy physics.

121 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-  
122 2018 provides the opportunity to make this measurement for the first time. The unprecedented  
123 energy achieved by the LHC during this period greatly increase the rate at which  $t\bar{t}H$  events are  
124 produced, and the large amount of data collected provides the necessary statistics for a differential  
125 measurement to be performed.

126 A study of  $t\bar{t}H$  events with multiple leptons in the final state is performed, using  $139 \text{ fb}^{-1}$

127 of data from proton-proton collisions at an energy  $\sqrt{s} = 13$  TeV collected by the ATLAS detector  
128 from 2015-2018. Events are separated into channels based on the number of light leptons in the  
129 final state - either two same-sign leptons, or three leptons. A deep neural network is used to  
130 reconstruct the momentum of the Higgs Boson in each event. This momentum spectrum is fit to  
131 data for each analysis channel, the result of which is used to place limits on BSM effects.

132 This dissertation begins with a brief explanation of the SM, its limitations, and the theor-  
133 etical motivation behind this work. This is followed by a description of the LHC and the ATLAS  
134 detector. The analysis strategy is then described, and the results are presented. Finally, the results  
135 of the study are summarized in the conclusion.

## 136 **Part II**

### 137 **Theoretical Motivation**

#### 138 **2 The Standard Model and the Higgs Boson**

139 The Standard Model of particle physics (SM) is a Quantum Field Theory (QFT) describing the  
140 known fundamental particles and their interactions. It accounts for three of the four known  
141 fundamental force - electromagnetism, the weak nuclear force, and the strong nuclear force, but  
142 not gravity. Further, the SM describes a mechanism for combining the weak and electromagnetic  
143 forces into a singular interaction, known as the electroweak force. It is a non-Abelian gauge

<sup>144</sup> theory, invariant under the Lie Group  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ , where C refers to color  
<sup>145</sup> charge, L, the helicity of the particle, and Y, the hypercharge.

<sup>146</sup> **2.1 The Forces and Particles of the Standard Model**

<sup>147</sup> The SM particles, summarized in figure 2.1, can be classified into two general categories based  
<sup>148</sup> on their spin: fermions, and bosons.

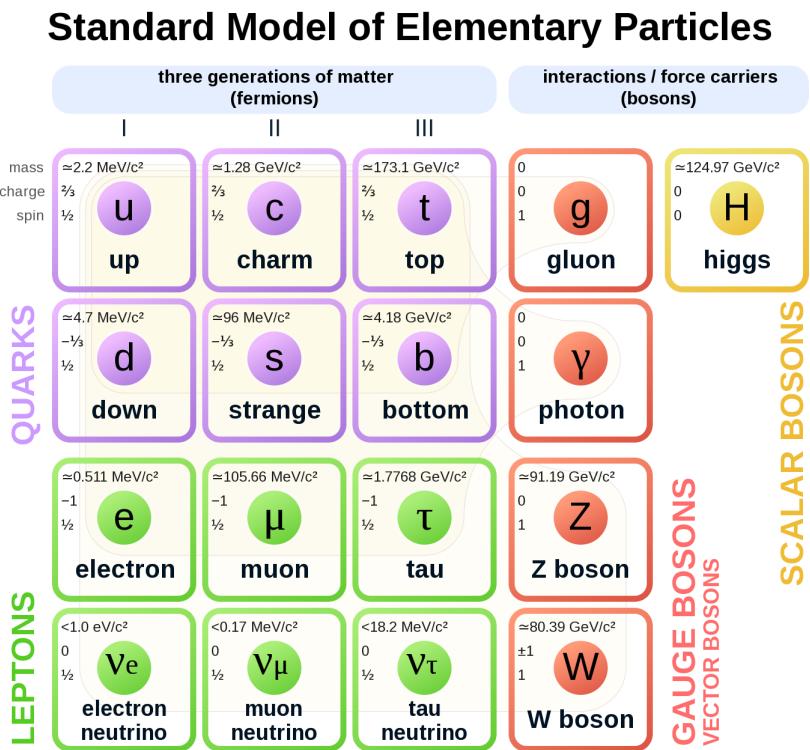


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. []

<sup>149</sup> Fermions are particles with  $\frac{1}{2}$ -integer spin, which according to the spin-statistics theorem,  
<sup>150</sup> causes them to comply with the Pauli-exclusion principle []. They can be separated into two

151 groups, leptons and quarks, each of which consist of three generations of particles with increasing  
 152 mass.

153 Leptons are fermions interact via the electroweak force, but not the strong force. The three  
 154 generation of leptons consist of the electron and electron neutrino, the muon and muon neutrino,  
 155 the tau and tau neutrino. The quarks, which do interact via the strong force - which is to say they  
 156 have color charge - in addition to the electroweak force. The three generations include the up  
 157 and down quarks, the strange and charm quarks, and the top and bottom quarks. Each of these  
 158 generations form left-handed doublets invariant under SU(2) transformations. For the leptons  
 159 these doublets are:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}_L, \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}_L, \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}_L \quad (2.1)$$

160 And for the quarks:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} s \\ c \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L \quad (2.2)$$

161 For both leptons and quarks, the heavier generations can decay into the lighter generation  
 162 of particles, while the first generation does not decay. Hence, ordinary matter generally consists  
 163 of this first generation of fermions - electrons, up quarks, and down quarks. Each of these  
 164 fermions has a corresponding anti-particle, which has an equal mass as its partner but opposite

165 charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos,

166 whose mass has been experimentally confirmed but is not accounted for in the SM.

167 Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-

168 exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1

169 particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0,

170 particle - the Higgs Boson. Of the gauge bosons, the  $W^+$ ,  $W^-$  and Z bosons - which are the

171 mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon

172 mediates the electric force, and the gluon mediates the strong force.

## 173 2.2 The Higgs Mechanism

174 A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to

175 the Lagrangian giving mass to the the gauge bosons would violate the underlying symmetry of

176 the theory. This presents a clear problem with the theory: The experimental observation that the

177 W and Z bosons have mass seems to contradict the basic structure of the SM.

178 Rather than abandoning gauge invariance, an alternative way for particles to acquire mass

179 beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout

180 in 1964 []. This procedure for introducing masses for the gauge bosons while preserving local

181 gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory

182 by Weinberg in 1967 [].

<sup>183</sup> **2.2.1 The Higgs Field**

<sup>184</sup> The Higgs mechanism introduces a complex scalar  $SU(2)$  doublet,  $\Phi$ , with the form:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}_L \quad (2.3)$$

<sup>185</sup> This field introduces a scalar potential to the Lagrangian of the form:

$$V(\Phi) = \mu^2 |\Phi^\dagger \Phi| + \lambda (|\Phi^\dagger \Phi|)^2 \quad (2.4)$$

<sup>186</sup> Where  $\mu$  and  $\lambda$  are free parameters of the new field. This represents the most general  
<sup>187</sup> potential allowed while preserving  $SU(2)_L$  invariance and renormalizability. In the case that  
<sup>188</sup>  $\mu^2 < 0$ , this potential takes the form shown in figure 2.2.

<sup>189</sup> The significant feature of this potential is that its minimum does not occur for a value of  
<sup>190</sup>  $\Phi = 0$ . Instead, it is minimized when  $|\Phi^\dagger \Phi| = -\mu^2/\lambda$ . This means that in its ground state, the  
<sup>191</sup> Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while  
<sup>192</sup> the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

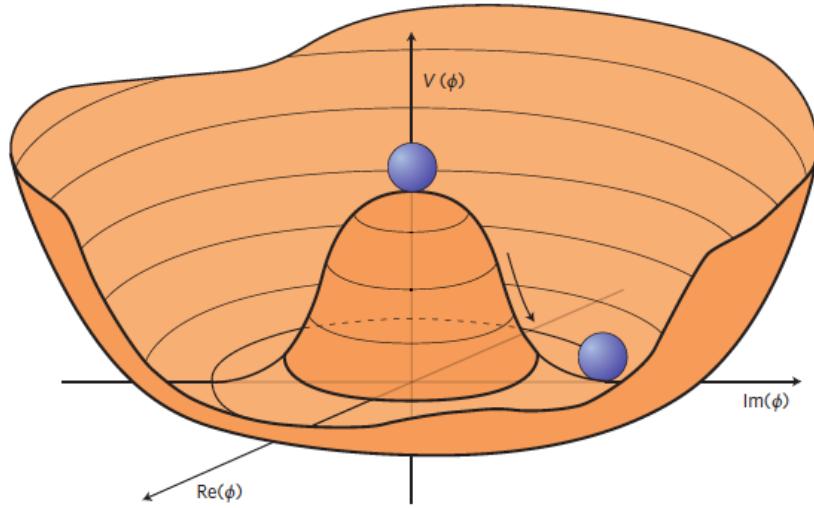


Figure 2.2: The value of the Higgs potential,  $V(\Phi)$  as a function of  $\Phi$ , for the case that  $\mu^2 < 0$  [].

<sup>193</sup> the minimum is determined only by  $\Phi^\dagger \Phi$ , there is some ambiguity in the particular definition of  
<sup>194</sup> the VEV, but it is generally represented as

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (2.5)$$

<sup>195</sup> The full value of  $\Phi$  can be written as

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H/\sqrt{2} \end{pmatrix} \quad (2.6)$$

<sup>196</sup> with  $v$  being the value of the VEV, and  $H$  being the real value of the scalar field.

197 **2.2.2 Electroweak Symmetry Breaking**

198 The Electroweak (EWK) interaction is described in the SM by a  $SU(2)_L \otimes U(1)_Y$  gauge theory.  
 199 This theory predicts three  $SU(2)_L$  gauge boson,  $W_\mu^1, W_\mu^2, W_\mu^3$ , and a single  $U(1)_Y$  gauge boson,  
 200  $B_\mu$ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar  
 201 field  $\Phi$  in the Lagrangian:

$$(D_\mu \Phi)^\dagger (D^\mu \Phi) = |(\partial_\mu - \frac{ig}{2} W_\mu^a \sigma^a - \frac{ig'}{2} B_\mu Y) \phi|^2 \quad (2.7)$$

202 Here  $D_\mu$  represents the covariant derivative required to preserve gauge invariance,  $g$  and  
 203  $g'$  represent coupling constant of the gauge bosons,  $\sigma^a$  denotes the Pauli matrices of  $SU(2)$ ,  
 204 and  $Y$  represents the hypercharge of  $U(1)$ . The terms in this interaction which contribute to the  
 205 masses of the gauge bosons can be written as:

$$\frac{1}{2}(0, v)(\frac{g}{2} W_\mu^a \sigma^a - \frac{g'}{2} B_\mu)^2 \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (2.8)$$

206 Expanding these terms into the mass eigenstates of the electroweak interaction yields four  
 207 physical gauge bosons, two charged and two neutral, which are linear combinations of the fields

<sup>208</sup>  $W_\mu^1, W_\mu^2, W_\mu^3$ , and  $B_\mu$ :

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \pm iW_\mu^2) \\ Z^\mu &= \frac{1}{\sqrt{(g^2 + g'^2)}}(-g'B_\mu + gW_\mu^3) \\ A^\mu &= \frac{1}{\sqrt{(g^2 + g'^2)}}(gB_\mu + g'W_\mu^3) \end{aligned} \tag{2.9}$$

<sup>209</sup> And the masses of these fields are given by:

$$\begin{aligned} M_W^2 &= \frac{1}{4}g^2v^2 \\ M_Z^2 &= \frac{1}{4}(g^2 + g'^2)v^2 \\ M_A^2 &= 0 \end{aligned} \tag{2.10}$$

<sup>210</sup> This produces exactly the particles we observe - three massive gauge bosons and a single  
<sup>211</sup> massless photon. The massless photon represents the portion of the gauge symmetry, a single  
<sup>212</sup>  $U(1)$  of the electromagnetic force, that remains unbroken by the VEV.

<sup>213</sup> Interactions with the Higgs field also lead to the generation of the fermion masses, which  
<sup>214</sup> in the Lagrangian take the form:

$$-\lambda_\psi(\bar{\psi}_L\phi\psi_R + \bar{\psi}_R\phi^\dagger\psi_L) \tag{2.11}$$

215 After symmetry breaking has occurred and  $\phi$  has taken on the value of the VEV as written  
 216 in equation 2.5, the mass terms for the fermions become  $\lambda_\psi v$ . Written this way, the fermion  
 217 masses are proportional to their Yukawa coupling to the VEV,  $\lambda_\psi$ .

218 Based on the equation 2.6, an additional mass term,  $\mu^2 H^2$  arises from the potential  $V(\Phi)$ .  
 219 This term can be understood as an excitation of the Higgs field, a scalar boson with mass  $M_H = \mu$ .  
 220 This is the Higgs boson, which comes about as a natural prediction of electroweak symmetry  
 221 breaking.

222 The fermion's Yukawa coupling to the VEV take the same form as the fermion's coupling  
 223 to the Higgs boson -  $\lambda_\psi$ . Therefore, the strength of a fermion's interaction with the Higgs is  
 224 directly proportional to its mass. We now have a model that predicts a Higgs boson with mass  
 225  $M_H = \mu$ , which interacts with the fermions with coupling strength  $\lambda_\psi$ . Because  $\mu$  and  $\lambda_\psi$  are  
 226 free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions  
 227 must be measured experimentally.

### 228 2.3 Limitations of the Standard Model

229 While the SM has great predictive power, there are still several experimental observations that the  
 230 SM fails to explain. For example, the SM predicts neutrinos to be massless, despite experimental  
 231 observation to the contrary.

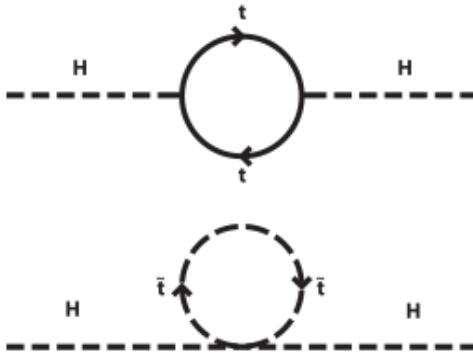


Figure 2.3: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram [6].

### <sup>232</sup> **3 Effective Field Theory in $t\bar{t}H$ Production**

<sup>233</sup> Higher dimension operators are a common way to paramaterize the effects of physics at very  
<sup>234</sup> high energies into

#### <sup>235</sup> **3.1 Extensions to the Higgs Sector**

#### <sup>236</sup> **3.2 Six Dimensional Operators**

<sup>237</sup> While the SM has been tested to great precision, particularly at the LHC, it is generally accepted  
<sup>238</sup> that it is only valid up to a certain energy scale. It is assumed that above a certain energy, at the  
<sup>239</sup> scale where something like a Grand Unified Theory (GUT) or quantum gravity become relevant,  
<sup>240</sup> the SM will not be applicable.

---

## 241 Part III

### 242 The LHC and the ATLAS Detector

#### 243 4 The LHC

244 The Large Hadron Collider (LHC) is a particle accelerator consisting of a 27 km ring, designed  
245 to collide protons at high energy. Located outside of Geneva, Switzerland and buried about 100  
246 m underground, it consists of a ring of superconducting magnets which are used to accelerate  
247 opposing beams of protons - or lead ions - which collide at the center of one of the various  
248 detectors located around the LHC ring which record the result of these collisions. These  
249 detectors include two general purpose detectors, ATLAS and CMS, which are designed to make  
250 precision measurements of a broad range of physics phenomenon, and two more specialized  
251 experiments, LHCb and ALICE, which are optimized to study b-quarks and heavy-ion physics,  
252 respectively.

253 The LHC first began running in 2009 at a proton-proton center of mass energy of  $\sqrt{s} = 8$   
254 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during  
255 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,  
256 and collected data at an increased energy of  $\sqrt{s} = 13$  TeV until 2018, a period referred to as Run  
257 2.

258 The LHC consists of a chain of accelerators, which accelerate the protons to higher and

259 higher energies until they are injected into the main ring. This process is summarized in figure  
 260 4.1. Protons extracted from a tank of ionized hydrogen are fed into a linear accelerator, LINAC2,  
 261 where they reach an energy of 50 MeV. From there, they enter a series of three separate circular  
 262 accelerators, before being injected into the main accelerator ring at an energy of 450 GeV. Within  
 263 the main ring protons are separated into two separate beams moving in opposite directions,  
 264 and their energy is increased to their full collision energy. Radiofrequency cavities are used to  
 265 accelerate these particles and sort them into bunches. From 2015-2018, these bunches consisted  
 266 of around 100 billion protons each with an energy of 6.5 TeV per proton, which collided at a rate  
 267 of 40 MHz, or every 25 ns.

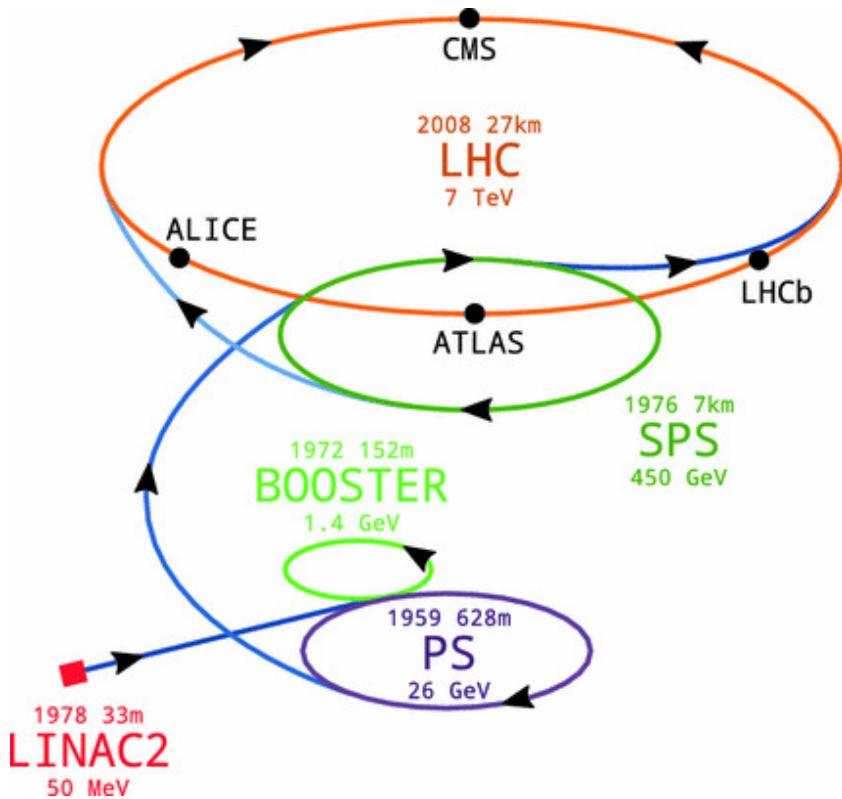


Figure 4.1: A summary of the accelerator chain used to feed protons into the LHC [].

268 Because these proton bunches consist of a large number of particles, each bunch crossing  
 269 consists of not just one, but several direct proton-proton collisions. The number of interactions  
 270 that occur per bunch crossing,  $\mu$ , is known as pileup. During Run 2, the average pileup for bunch  
 271 crossings was around  $\langle \mu \rangle = 35$ , with values typically ranging between 10 and 70.

272 The amount of data collected by the LHC is measured in terms of luminosity, which is the  
 273 ratio of the number of events detected per unit time,  $\frac{dN}{dt}$ , and the interaction cross-section,  $\sigma$ .

$$\mathcal{L} = \frac{1}{\sigma} \frac{dN}{dt} \quad (4.1)$$

274 The design luminosity of the LHC is  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , however the LHC has achieved a  
 275 luminosity of over  $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . The total luminosity is then this instantaneous luminosity  
 276 integrated over time.

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} dt \quad (4.2)$$

277 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around  
 278  $140 \text{ fb}^{-1}$ , exceeding the expected integrated luminosity of  $100 \text{ fb}^{-1}$ .

## <sup>279</sup> 5 The ATLAS Detector

<sup>280</sup> ATLAS (a not terribly natural acronym for “A Toroidal LHC Apparatus”) is a general purpose  
<sup>281</sup> detector designed to maximize the detection efficiency of all physics objects, including leptons,  
<sup>282</sup> jets, and photons. This means it is capable of measuring all SM particles, with the exception of  
<sup>283</sup> neutrinos, the presence of which can be inferred based on missing transverse momentum. The  
<sup>284</sup> detector measures 44 m long, and 25 m tall.

<sup>285</sup> The ATLAS detector consists of multiple layers, each of which serves a different purpose  
<sup>286</sup> in reconstructing collisions. At the very center of the detector is the interaction point where the  
<sup>287</sup> proton beams of the LHC collide.

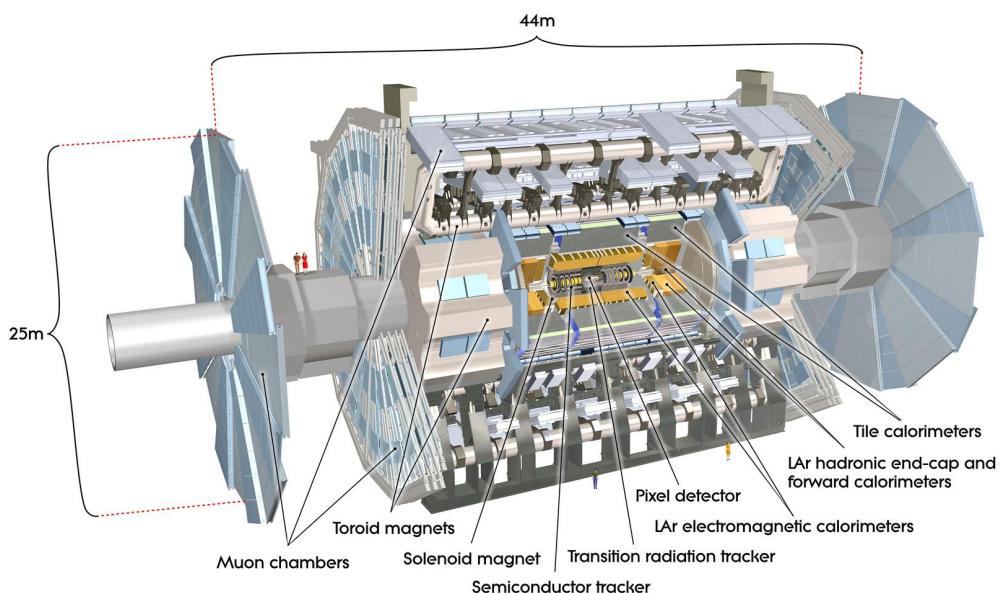


Figure 5.1: Cutaway view of the ATLAS detector, with labels of its major components [].

288 **5.1 Inner Detector**

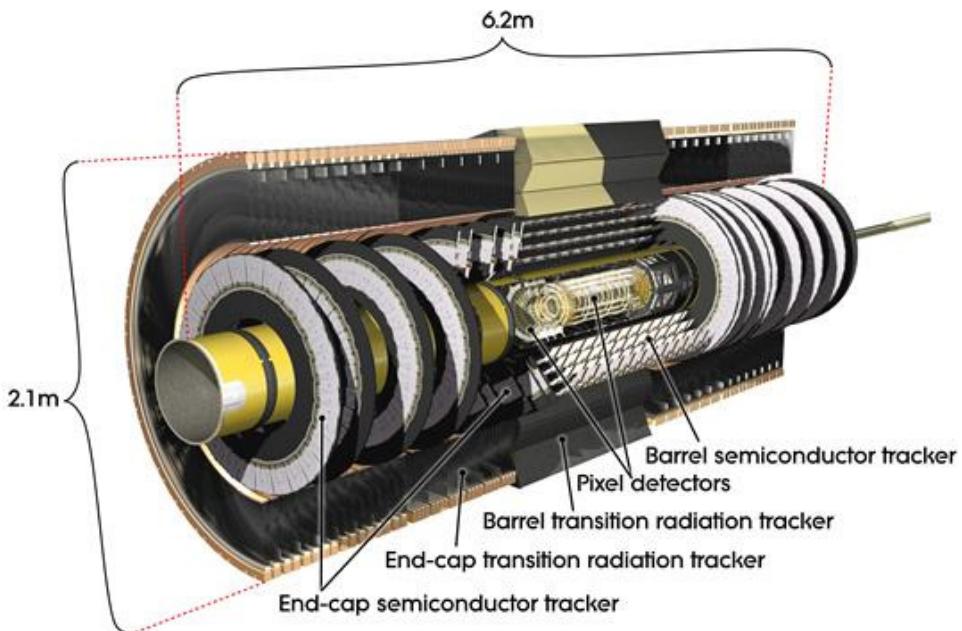


Figure 5.2: Cutaway view of the Inner Detector [].

289 Just surrounding the interaction point is the Inner Detector, designed to track the path  
290 of charged particles moving through the detector. An inner solenoid surrounding the Innder  
291 Detector is used to produces a magnetic field of 2 T. This large magnetic field causes the path  
292 of charged particles moving through the Inner Detector to bend. Because this magnetic field is  
293 uniform and well known, it can be used in conjunction with the curvature of a particles path to  
294 measure its charge and momentum.

295 The Inner Detector consists of three components - the Pixel Detector, the Semi-Conductor  
296 Tracker (SCT), and the Transition Radiation Tracker (TRT). The Pixel Detector is the innermost  
297 of these, beginning just 33.25 mm away from the beam line. It consists of three silicon layers

298 along the barrel, as well as three endcap layers, covering a range of  $|\eta| < 2.5$ .

299        The Semiconductor Tracker (SCT) is similar to the Pixel detector, but uses long strips  
 300      rather than small pixel to cover a larger spatial area.

301 **5.2 Calorimeters**

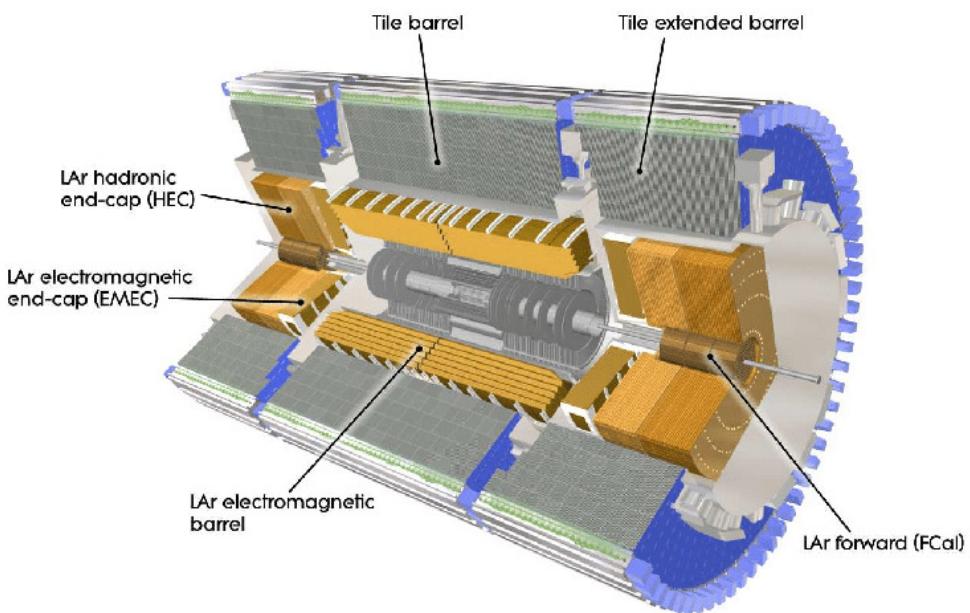


Figure 5.3: Cutaway view of the calorimeter system of the ATLAS detector [].

302        Situated outside the Inner Detector are two concentric calorimeters. The inner calorimeter  
 303      uses liquid argon (LAr) to measure energy of particles that interact electromagnetically, which  
 304      includes photons and any charged particle. The LAr calorimeter is made of heavy metals,  
 305      primarily lead and copper, which causes electromagnetically interacting particles to shower,  
 306      depositing their energy in the detector. The showering of the high energy particles that pass

307 through calorimeter cause the liquid argon to ionize, and the ionized electrons are detected by  
308 electronic readouts. The LAr calorimeter consists of around 180,000 readout channels.

309 The outer calorimeter measures the energy from particles that pass through the EM calor-  
310 imeter, and measures the energy of particles that interact via the strong force. This is primarily  
311 hadrons. It is composed of steel plates to cause hadronic showering and scintillating tiles as the  
312 active material. The signals from the hadronic calorimeter are read out by photomultiplier tubes  
313 (PMTs).

### 314 **5.3 Muon Spectrometer**

315 Because muons are heavier than electrons and photons, and do not interact via the strong force,  
316 they generally pass through the detector without being stopped by the calorimeters. The outermost  
317 components of the detector are designed specifically to measure the energy and momentum of  
318 muons produced in the LHC. The muon spectrometer consists of tracking and triggering system.  
319 It extends from the outside of the calorimeter system, about a 4.25 m radius from the beam line,  
320 to a radius of 11 m. This large detector system is necessary to accurately measure the momentum  
321 of muons, which is essential not only for measurements involving the muons themselves, but also  
322 to accurately estimate the missing energy in each event.

323 Two large toroidal magnets within the muon system generate a large magnetic field which  
324 covers an area 26 m long with a radius of 10 m. Because the area covered by this magnet system  
325 is so large, a uniform magnetic field like the one produced in the Inner Detector is impractical.

326 Instead, the magnetic field that exists in the muon spectrometer ranges between 2 T and 8 T, and  
327 is much less uniform. The path of the muons passing through the spectrometer is bent by this  
328 field, allowing their charge to be determined.

329 1200 tracking chambers are placed in the muon system in order to precisely measure the  
330 tracks of muons with high spatial resolution.

331 **5.4 Trigger System**

332 Because of the high collision rate and large amount of data collected by the various subdetectors,  
333 ATLAS produces far more data than can actually be stored. Each event produces around 25 Mb  
334 of raw data, which multiplied by the bunch crossing rate of 40 MHz, comes out to around a  
335 petabyte of data every second. The information from every event cannot practically be stored,  
336 therefore a sophisticated trigger system is employed in real time to determine whether events are  
337 sufficiently interesting to be worth storing.

338 The trigger system in ATLAS involves multiple levels, each of which select out which  
339 events move on to the next level of scrutiny. The level-1 trigger uses hardware information from  
340 the calorimeters and muon spectrometer to select events that contain candidates for particles  
341 commonly used in analysis, such as energetic leptons and jets. The level-1 trigger reduces the  
342 rate of events from 40 MHz to around 100 kHz.

343 Events that pass the level-1 trigger move to the High-Level Trigger (HLT). The HLT takes  
344 place outside of the detector in software, and looks for properties such as a large amount of  
345 missing transverse energy, well defined leptons, and multiple high energy jets. Events that pass  
346 the HLT are stored and used for analysis. Because the specifics of the HLT are determined by  
347 software rather than hardware, the thresholds can be changed throughout the run of the detector  
348 in response to run conditions such as changes to pileup and luminosity. After the HLT is applied,  
349 the event rate is reduced to around 1000 per second, which are recorded for analysis.

## 350 **Part IV**

### 351 **Search for Dimension-Six Operators**

#### 352 **6 Data and Monte Carlo Samples**

353 For both data and Monte Carlo (MC) simulations, samples were prepared in the `xAOD` format,  
354 which was used to produce a `xAOD` based on the `HIGG8D1` derivation framework. This framework  
355 was designed for the main  $t\bar{t}H$  multi-lepton analysis. Because this analysis targets events with  
356 multiple light leptons, as well as tau hadrons, this framework skims the dataset of any events that  
357 do not meet at least one of the following requirements:

- 358 • at least two light leptons within a range  $|\eta| < 2.6$ , with leading lepton  $p_T > 15$  GeV and  
359 subleading lepton  $p_T > 5$  GeV

- 360        • at least one light lepton with  $p_T > 15$  GeV within a range  $|\eta| < 2.6$ , and at least two hadronic  
361            taus with  $p_T > 15$  GeV.

362            Samples were then generated from these HIGG8D1 derivations using a modified version of  
363            AnalysisBase version 21.2.127.

364        **6.1 Data Samples**

365        The study uses proton-proton collision data collected by the ATLAS detector from 2015 through  
366            2018, which represents an integrated luminosity of  $139 \text{ fb}^{-1}$  and an energy of  $\sqrt{s} = 13 \text{ TeV}$ . All  
367            data used in this analysis was included in one the following Good Run Lists:

- 368        • data15\_13TeV.periodAllYear\_DetStatus-v79-repro20-02\_DQDefects-00-02-02  
369            \_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- 370        • data16\_13TeV.periodAllYear\_DetStatus-v88-pro20-21\_DQDefects-00-02-04  
371            \_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- 372        • data17\_13TeV.periodAllYear\_DetStatus-v97-pro21-13\_Unknown\_PHYS\_StandardGRL  
373            \_All\_Good\_25ns\_Triggerno17e33prim.xml
- 374        • data18\_13TeV.periodAllYear\_DetStatus-v102-pro22-04\_Unknown\_PHYS\_StandardGRL  
375            \_All\_Good\_25ns\_Triggerno17e33prim.xml

<sup>376</sup> **6.2 Monte Carlo Samples**

<sup>377</sup> Several Monte Carlo (MC) generators were used to simulate both signal and background pro-  
<sup>378</sup> cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific  
<sup>379</sup> event generator used for each of these MC samples is listed in table 1.

Table 1: The configurations used for event generation of signal and background processes, including the event generator, matrix element (ME) order, parton shower algorithm, and parton distribution function (PDF).

Process	Event generator	ME order	Parton Shower	PDF
ttH	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [ <a href="#">Ball:2014uwa</a> ] (CT10 [ <a href="#">ct10</a> ])
t̄W	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
t̄t(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
t̄t	POWHEG-BOX v2 [ <a href="#">powhegtt</a> ]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄tγ	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
tZ	MG5_AMC	LO	PYTHIA 6	CTEQ6L1
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t̄t, t̄t̄t̄t	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	POWHEG-BOX v1 [ <a href="#">powhegstp</a> ]	NLO	PYTHIA 6	CT10
qqVV, VVV				
Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

<sup>380</sup> **7 Object Reconstruction**

<sup>381</sup> All analysis channels considered in this note share a common object selection for leptons and  
<sup>382</sup> jets, as well as a shared trigger selection.

383 **7.1 Trigger Requirements**

384 Events are required to be selected by dilepton triggers, as summarized in table 2.

Dilepton triggers (2015)	
$\mu\mu$ (asymm.)	HLT_mu18_mu8noL1
$ee$ (symm.)	HLT_2e12_lhloose_L12EM10VH
$e\mu, \mu e$ ( $\sim$ symm.)	HLT_e17_lhloose_mu14
Dilepton triggers (2016)	
$\mu\mu$ (asymm.)	HLT_mu22_mu8noL1
$ee$ (symm.)	HLT_2e17_lhvloose_nod0
$e\mu, \mu e$ ( $\sim$ symm.)	HLT_e17_lhloose_nod0_mu14
Dilepton triggers (2017)	
$\mu\mu$ (asymm.)	HLT_mu22_mu8noL1
$ee$ (symm.)	HLT_2e24_lhvloose_nod0
$e\mu, \mu e$ ( $\sim$ symm.)	HLT_e17_lhloose_nod0_mu14
Dilepton triggers (2018)	
$\mu\mu$ (asymm.)	HLT_mu22_mu8noL1
$ee$ (symm.)	HLT_2e24_lhvloose_nod0
$e\mu, \mu e$ ( $\sim$ symm.)	HLT_e17_lhloose_nod0_mu14

Table 2: List of lowest  $p_T$ -threshold, un-prescaled dilepton triggers used for 2015-2018 data taking.

385 **7.2 Light Leptons**

386 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that  
387 are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].  
388 Electron candidates are required to have  $p_T > 10$  GeV and  $|\eta_{\text{cluster}}| < 2.47$ . Candidates in the  
389 transition region between different electromagnetic calorimeter components,  $1.37 < |\eta_{\text{cluster}}| <$   
390 1.52, are rejected. A multivariate likelihood discriminant combining shower shape and track

391 information is used to distinguish prompt electrons from nonprompt leptons, such as those  
392 originating from hadronic showers.

393 To further reduce the non-prompt contribution, the track of each electron is required to  
394 originate from the primary vertex; requirements are imposed on the transverse impact parameter  
395 significance ( $|d_0|/\sigma_{d_0}$ ) and the longitudinal impact parameter ( $|\Delta z_0 \sin \theta_\ell|$ ).

396 Muon candidates are reconstructed by combining inner detector tracks with track segments  
397 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have  
398  $p_T > 10$  GeV and  $|\eta| < 2.5$ . All leptons are required to be isolated, and pass a non-prompt BDT  
399 selection described in detail in [**ttH\_paper**].

### 400 7.3 Jets

401 Jets are reconstructed from calibrated topological clusters built from energy deposits in the  
402 calorimeters [**ATL-PHYS-PUB-2015-015**], using the anti- $k_t$  algorithm with a radius parameter  
403  $R = 0.4$ . Jets with energy contributions likely arising from noise or detector effects are removed  
404 from consideration [**ATLAS-CONF-2015-029**], and only jets satisfying  $p_T > 25$  GeV and  
405  $|\eta| < 2.5$  are used in this analysis. For jets with  $p_T < 60$  GeV and  $|\eta| < 2.4$ , a jet-track  
406 association algorithm is used to confirm that the jet originates from the selected primary vertex,  
407 in order to reject jets arising from pileup collisions [**PERF-2014-03**].

<sup>408</sup> **7.4 Missing Transverse Energy**

<sup>409</sup> Because all  $t\bar{t}H - ML$  channels considered include multiple neutrinos, missing transverse  
<sup>410</sup> energy ( $E_T^{\text{miss}}$ ) is present in each event. The missing transverse momentum vector is defined as  
<sup>411</sup> the inverse of the sum of the transverse momenta of all reconstructed physics objects as well  
<sup>412</sup> as remaining unclustered energy, the latter of which is estimated from low- $p_T$  tracks associated  
<sup>413</sup> with the primary vertex but not assigned to a hard object [ATL-PHYS-PUB-2015-027].

<sup>414</sup> **8 Higgs Momentum Reconstruction**

<sup>415</sup> Reconstructing the momentum of the Higgs boson is a particular challenge for channels with  
<sup>416</sup> leptons in the final state: Because all channels include at least two neutrinos in the final state, the  
<sup>417</sup> Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted  
<sup>418</sup> by a neural network when provided with the four-vectors of the Higgs Boson decay products, as  
<sup>419</sup> shown in section 8.1. With this in mind, several layers of MVAs are used to reconstruction the  
<sup>420</sup> Higgs momentum.

<sup>421</sup> The first layer is a model designed to select which jets are most likely to be the b-jets  
<sup>422</sup> that came from the top decay, detailed in section 8.2. As described in section 8.3, the kinematics  
<sup>423</sup> of these jets are fed into the second layer, which is designed to identify the decay products of  
<sup>424</sup> the Higgs Boson itself. The kinematics of these particles are then fed into yet another neural-

425 network, which predicts the momentum of the Higgs (8.4). MVAs are also used in the analysis  
426 to determine the decay of the Higgs boson in the 3l channel (8.5).

427 Models are trained on Monte Carlo simulations of  $t\bar{t}H$  events generated using MG5\_AMC.  
428 Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training.

429 For all of these models, the Keras neural network framework, with Tensorflow as the  
430 backend, is used, and the number of hidden layers and nodes are determined using grid search  
431 optimization. Each neural network uses the LeakyReLU activation function, a learning rate  
432 of 0.01, and the Adam optimization algorithm, as alternatives are found to either decrease or  
433 have no impact on performance. Batch normalization is applied after each layer. For the  
434 classification algorithms (b-jet matching, Higgs reconstruction, and 3l decay identification)  
435 binary-cross entropy is used as the loss function, while the  $p_T$  reconstruction algorithm uses  
436 MSE.

437 The specific inputs features used for each model are arrived at through a process of trial  
438 and error - features considered potentially useful are tried, and those that are found to increase  
439 performance are included. While each model includes a relatively large number of features,  
440 some using upwards of 30, this inclusive approach is found to maximize the performance of each  
441 model while decreasing the variance compared to a reduced number of inputs. Each input feature  
442 is validated by comparing MC simulations to  $80 \text{ fb}^{-1}$  of data, as shown in the sections below.

---

**443 8.1 Decay Candidate Reconstruction**

444 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson  
445 using MC simulations of  $t\bar{t}H$  events. These include light leptons and jets. Reconstructed  
446 physics objects are matched to truth level particles, in order to identify the parents of these  
447 reconstructed objects. The kinematics of the decay product candidates as well as event level  
448 variables are used as inputs.

449 Leptons considered as possible Higgs and top decay candidates are required to pass the  
450 selection described in section 7.2. For jets, however, it is found that a large fraction that originate  
451 from either the top decay or the Higgs decay fall outside the selection described in section 7.3.  
452 Specifically, jets from the Higgs decay tend to be soft, with 32% having  $p_T < 25$  GeV. Therefore  
453 jets with  $p_T < 15$  GeV are considered as possible candidates in the models described below. By  
454 contrast, less than 5% of the jets originating from the Higgs fall below this  $p_T$ . The jets are found  
455 to be well modeled even down to this low  $p_T$  threshold, as shown in section 10.1. The impact of  
456 using different  $p_T$  selection for the jet candidates is considered in detail in section A.5. As they  
457 are expected to originate from the primary vertex, jets are also required to pass a JVT cut.

**458 8.2 b-jet Identification**

459 Including the kinematics of the b-jets that originate from the top decay is found to improve the  
460 identification of the Higgs decay products, and improve the accuracy with which the Higgs  
461 momentum can be reconstructed. Because these b-jets are reconstructed by the detector with

<sup>462</sup> high efficiency (just over 90% of the time), and can be identified relatively consistently, the first  
<sup>463</sup> step in reconstructing the Higgs is selecting the b-jets from the top decay.

<sup>464</sup> Exactly two b-jets are expected in the final state of  $t\bar{t}H - ML$  events. However, in both  
<sup>465</sup> the 3l and 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r b-tag  
<sup>466</sup> working point is used). Therefore, for events which have exactly one, or more than two, b-tagged  
<sup>467</sup> jets, deciding which combination of jets correspond to the top decay is non-trivial. Further,  
<sup>468</sup> events with 1 b-tagged jet represent just over half of all  $t\bar{t}H - ML$  events. Of those, both b-jets  
<sup>469</sup> are reconstructed by the detector 75% of the time. Therefore, rather than adjusting the selection  
<sup>470</sup> to require exactly 2 b-tagged jets, and losing more than half of the signal events, a neural network  
<sup>471</sup> is used to predict which pair of jets is most likely to correspond to truth b-jets.

<sup>472</sup> Once the network is trained, all possible pairings of jets are fed into the model, and the pair  
<sup>473</sup> of jets with the highest output score are taken to be b-jets in successive steps of the analysis.

### <sup>474</sup> 8.2.1 2lSS Channel

<sup>475</sup> For the 2lSS channel, the input features shown in table 3 are used for training. Here  $j_0$  and  $j_1$  are  
<sup>476</sup> the two jet candidates, while  $l_0$  and  $l_1$  are the two leptons in the event, both ordered by  $p_T$ . jet  
<sup>477</sup> DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,  
<sup>478</sup> where 5 represents the tightest working point and 1 represents the loosest. The variables nJets  
<sup>479</sup> DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and  
<sup>480</sup> 85% b-tag working points, respectively.

jet $p_T$ 0	jet $p_T$ 1	Lepton $p_T$ 0
Lepton $p_T$ 1	jet $\eta$ 0	jet $\eta$ 1
$\Delta R(j_0)(j_1)$	$M(j_0 j_1)$	$\Delta R(l_0)(j_0)$
$\Delta R(l_0)(j_1)$	$\Delta R(l_1)(j_0)$	$\Delta R(l_1)(j_1)$
$M(l_0 j_0)$	$M(l_0 j_1)$	$M(l_1 j_0)$
$M(l_1 j_1)$	jet DL1r 0	jet DL1r 1
nJets OR DL1r 85	nJets OR DL1r 60	$\Delta R(j_0 l_0)(j_1 l_1)$
$\Delta R(j_0 l_1)(j_1 l_0)$	$p_T(j_0 j_1 l_0 l_1 E_T^{\text{miss}})$	$M(j_0 j_1 l_0 l_1 E_T^{\text{miss}})$
$\Delta\phi(j_0)(E_T^{\text{miss}})$	$\Delta\phi(j_1)(E_T^{\text{miss}})$	HT jets
nJets	$E_T^{\text{miss}}$	

Table 3: Input features used in the b-jet identification algorithm for the 2lSS channel

481 As there are far more incorrect combinations than correct ones, by a factor of more than  
 482 20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random  
 483 sample of 5 million incorrect entries are used for training, along with close 1 million correct  
 484 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for  
 485 training.

486 The difference between the distributions for a few of these features for the correct (i.e.  
 487 both jets are truth b-jets), and incorrect combinations are shown in figure 8.1. The correct and  
 488 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences  
 489 in the distributions.

490 The modeling of these inputs is validated against data, with figure 8.2 showing good  
 491 general agreement between data and MC. Plots for the complete list of features can found in  
 492 section A.

493 Based on the results of grid search evaluation, the optimal architecture is found to include

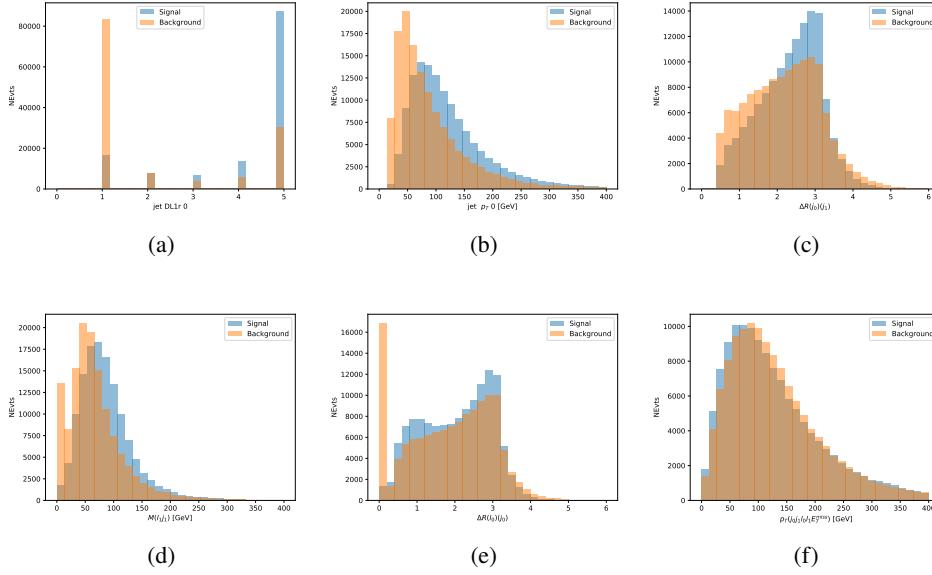


Figure 8.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the  $p_T$  of the leading jet, (c) shows the  $\Delta R$  of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the  $\Delta R$  of lepton 0 and jet 0, and (f) the  $p_T$  of both jets, both leptons, and the  $E_T^{\text{miss}}$ .

494 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as  
 495 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for  
 496 the trained model are shown in figure 8.2.1. The model is found to identify the correct pairing  
 497 of jets for 73% of 2ISS signal events on test data.

498 For point of comparison, a naïve approach to identify b-jets is used as well: The two jets  
 499 which pass the highest DL1r b-tag working point are assumed to be the b-jets from the top decay.  
 500 In the case that multiple jets meet the same b-tag working point, the jet with higher  $p_T$  is used.  
 501 This method identifies the correct jet pair 65% of the time.

502 The accuracy of the model for different values of n-bjets, compared to this naive approach,

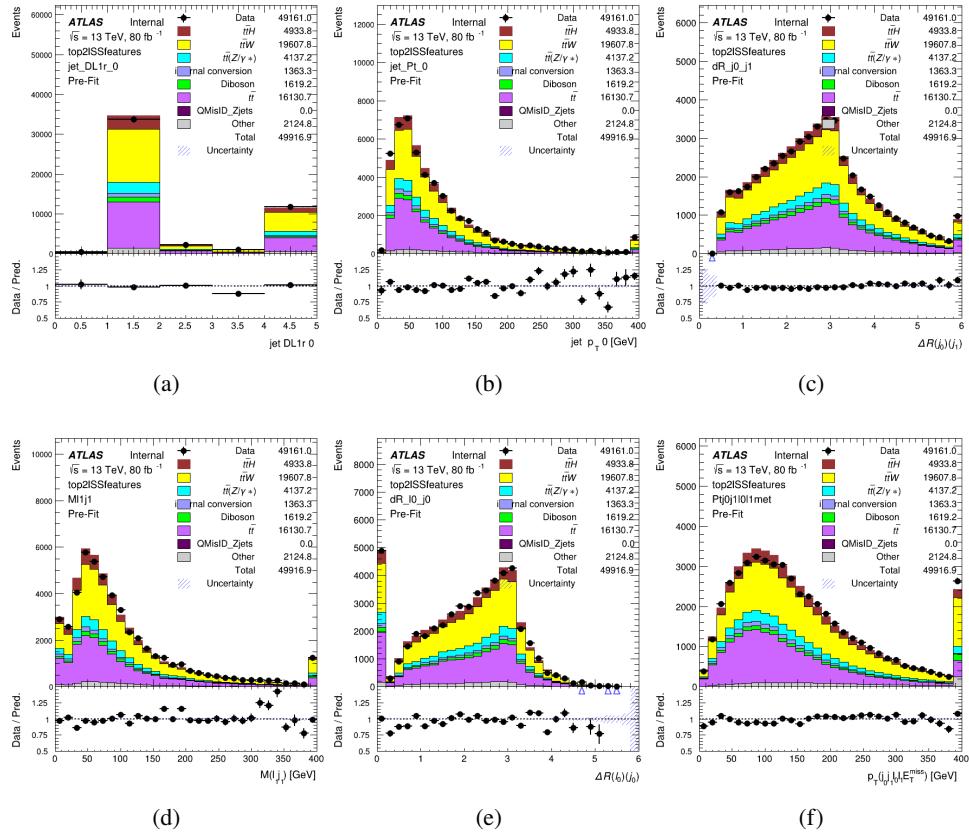


Figure 8.2: Data/MC comparisons of input features for top2ISS training for  $80 \text{ fb}^{-1}$  of data. (a) shows the DL1r working point of leading jet, (b) shows the  $p_T$  of the leading jet, (c) shows the  $\Delta R$  of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the  $\Delta R$  of lepton 0 and jet 0, and (f) the  $p_T$  of both jets, both leptons, and the  $E_T^{\text{miss}}$ .

is shown in table 4.

b-jet Selection	Neural Network	Naive
1 b-jet	58.6%	42.1%
2 b-jets	88.4%	87.1%
$\geq 3$ b-jets	61.7%	53.3%
Overall	73.9% %	67.2%

Table 4: Accuracy of the NN in identifying b-jets from tops in 2ISS events for, compared to the accuracy of taking the two highest b-tagged jets.

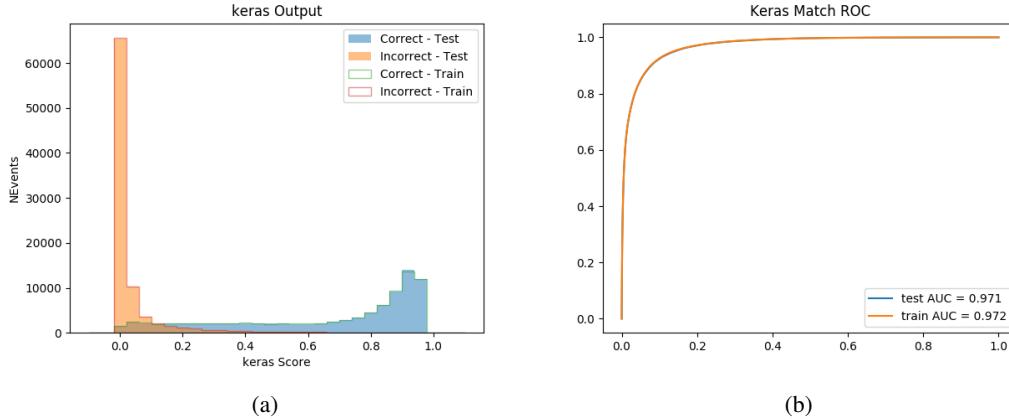


Figure 8.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

### 504 8.2.2 3l Channel

505 The input features used in the 3l channel are listed in table 5, with the same naming convention  
506 as the 2lSS channel.

jet $p_T$ 0	jet $p_T$ 1	jet $\eta$ 0
jet $\eta$ 1	Lepton $p_T$ 0	Lepton $p_T$ 1
Lepton $p_T$ 2	$\Delta R(j_0)(j_1)$	$M(j_0 j_1)$
$\Delta R(l_0)(j_0)$	$\Delta R(l_1)(j_0)$	$\Delta R(l_2)(j_0)$
$\Delta R(l_0)(j_1)$	$\Delta R(l_1)(j_1)$	$\Delta R(l_2)(j_1)$
$M(l_0 j_0)$	$M(l_1 j_0)$	$M(l_2 j_0)$
$M(l_0 j_1)$	$M(l_1 j_1)$	$M(l_2 j_1)$
$\Delta R(j_0 l_0)(j_1 l_1)$	$\Delta R(j_0 l_0)(j_1 l_2)$	$\Delta R(j_0 l_1)(j_1 l_0)$
$\Delta R(j_0 l_2)(j_1 l_0)$	jet DL1r 0	jet DL1r 1
$p_T(j_0 j_1 l_0 l_1 l_2 E_T^{\text{miss}})$	$M(t j_0 j_1 l_0 l_1 l_2 E_T^{\text{miss}})$	$\Delta\phi(j_0)(E_T^{\text{miss}})$
$\Delta\phi(j_1)(E_T^{\text{miss}})$	HT Lepton	HT jets
nJets	$E_T^{\text{miss}}$	nJets OR DL1r 85
nJets OR DL1r 60		

Table 5: Input features for the b-jet identification algorithm in the 3l channel.

507 A few of these features are shown in figure 8.4, comparing the distributions for correct and

508 incorrect combinations of jets.

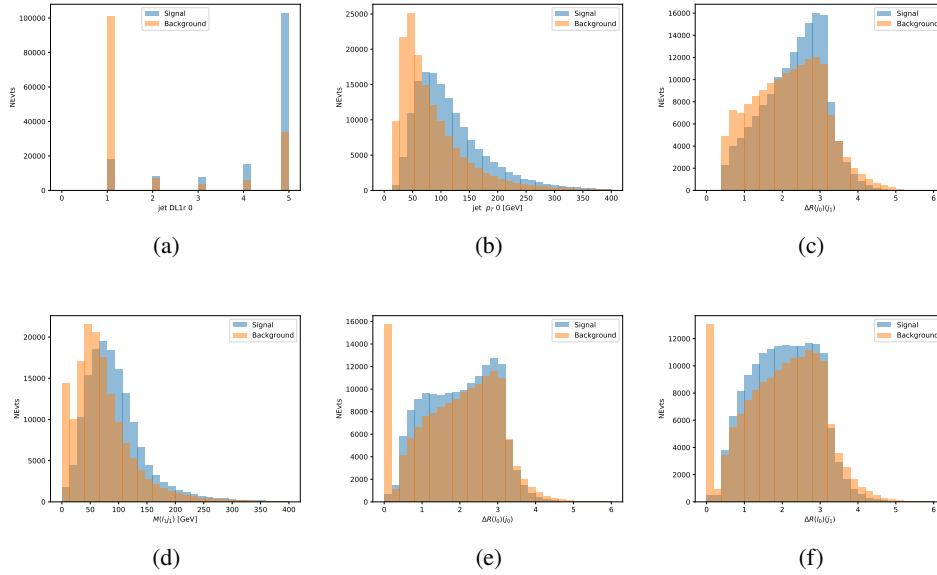


Figure 8.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the  $p_T$  of jet 0, (c)  $\Delta R$  between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e)  $\Delta R$  between lepton 0 and jet 0, and (f)  $\Delta R$  between lepton 0 and jet 1

509 The modeling of these inputs is validated against data, with figure 8.5 showing good  
510 general agreement between data and MC. Plots for the complete list of features can found in  
511 section A.

512 Again, the dataset is downsized to reduce the ratio of correct and incorrect combination  
513 from 20:1, to 5:1. Around 7 million events are used for training, with 10% set aside for testing.  
514 Based on the results of grid search evaluation, the optimal architecture is found to include 5  
515 hidden layers with 60 nodes each. The output score distribution as well as the ROC curve for the  
516 trained model are shown in figure 8.2.2.

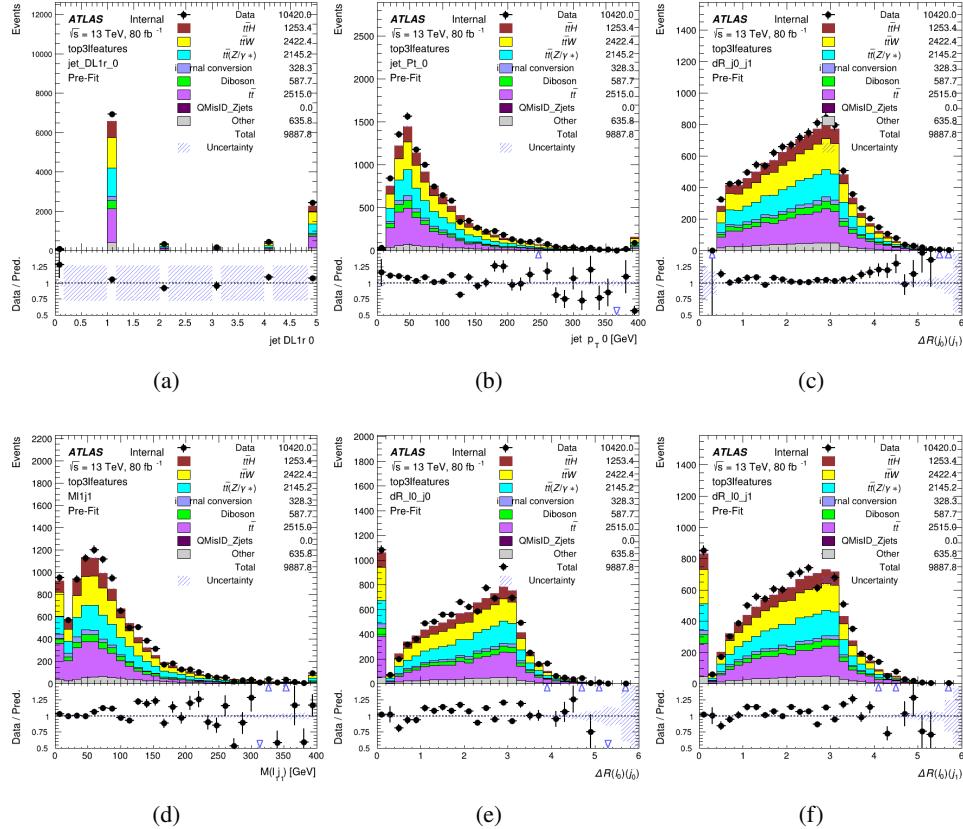


Figure 8.5: Data/MC comparisons of input features for top31 training for  $80 \text{ fb}^{-1}$  of data. (a) show the DL1r WP of jet 0, (b) the  $p_T$  of jet 0, (c)  $\Delta R$  between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e)  $\Delta R$  between lepton 0 and jet 0, and (f)  $\Delta R$  between lepton 0 and jet 1

517 This procedure is found to identify the correct pairing of jets for nearly 80% of 31 signal  
 518 events. The accuracy of the model is summarized in table 6.

Table 6: Accuracy of the NN in identifying b-jets from tops, compared to the naive method of taking the highest b-tagged jets.

	NN	Naive
1 b-jet	69.0%	48.9%
2 b-jets	89.6%	88.3%
$\geq 3$ b-jets	55.7%	52.3%
Overall	79.8%	70.2%

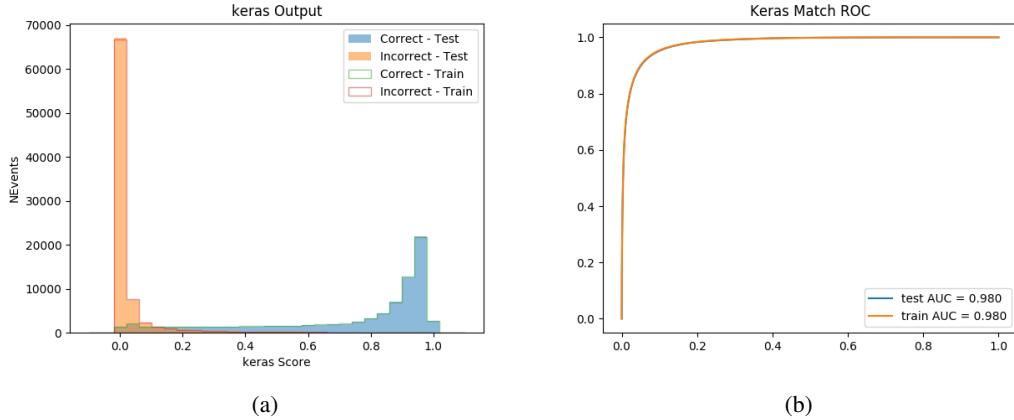


Figure 8.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

### 519 8.3 Higgs Reconstruction

520 Techniques similar to the b-jet identification algorithms are employed to select the decay products  
 521 of the Higgs: kinematics of all possible combinations of reconstructed objects are fed into a neural  
 522 network to determine which of those is most mostly to be the decay products of the Higgs.

523 Again separate models are used for the 2lSS and 3l channels, while the 3l channel has now  
 524 been split into two:  $t\bar{t}H$  events with three leptons in the final state include both instances where  
 525 the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the Higgs  
 526 decays to two leptons.

527 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In  
 528 the semi-leptonic case the reconstructed decay products consist of two jets and a single leptons.  
 529 For the fully-leptonic case, the decay products include 2 of the three leptons associated with the

530 event. For training the models, events are separated into these two categories using truth level  
531 information. A separate MVA, described in section 8.5, is used to make this distinction at reco  
532 level and determine which model to use.

533 For all channels, the models described in section 8.2 are used to identify b-jet candidates,  
534 whose kinematics are used to identify the Higgs decay products. These jets are not considered  
535 as possible candidates for the Higgs decay, justified by the fact that these models are found to  
536 misidentify jets from the Higgs decay as jets from the top decay less than 1% of the time.

537 **8.3.1 2lSS Channel**

538 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural  
539 network is trained on the kinematics of different combinations of leptons and jets, as well as the  
540 b-jets identified in section 8.2, with the specific input features listed in table 7.

Lepton $p_T$ H	Lepton $p_T$ T	jet $p_T$ 0
jet $p_T$ 1	top $p_T$ 0	top $p_T$ 1
top $\eta$ 0	top $\eta$ 1	jet $\eta$ 0
jet $\eta$ 1	jet Phi 0	jet Phi 1
Lepton $\eta$ H	Lepton $\eta$ T	$\Delta R(j_0)(j_1)$
$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$	$M(j_0 j_1)$
$M(l_H j_0)$	$M(l_H j_1)$	$\Delta R(l_H)(b_0)$
$\Delta R(l_H)(b_1)$	$\Delta R(l_T)(b_0)$	$\Delta R(l_T)(b_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1)(l_T)$	$\Delta R(j_0 j_1)(b_0)$
$\Delta R(j_0 j_1)(b_1)$	$\Delta R(j_0)(b_0)$	$\Delta R(j_0)(b_1)$
$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$	$M(j_0 j_1 l_H)$
$p_T(j_0 j_1 l_H l_T b_0 b_1 E_T^{\text{miss}})$	topScore	$E_T^{\text{miss}}$
nJets	HT jets	

Table 7: Input features used to identify the Higgs decay products in 2lSS events

541 Here  $j_0$  and  $j_1$ , and  $l_H$  are the jet and lepton decay candidates, respectively. The other  
 542 lepton in the event is labeled  $l_T$ , as it is assumed to have come from the decay of one of the top  
 543 quarks.  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet identification algorithm. The b-jet  
 544 Reco Score is the output of the b-jet reconstruction algorithm.

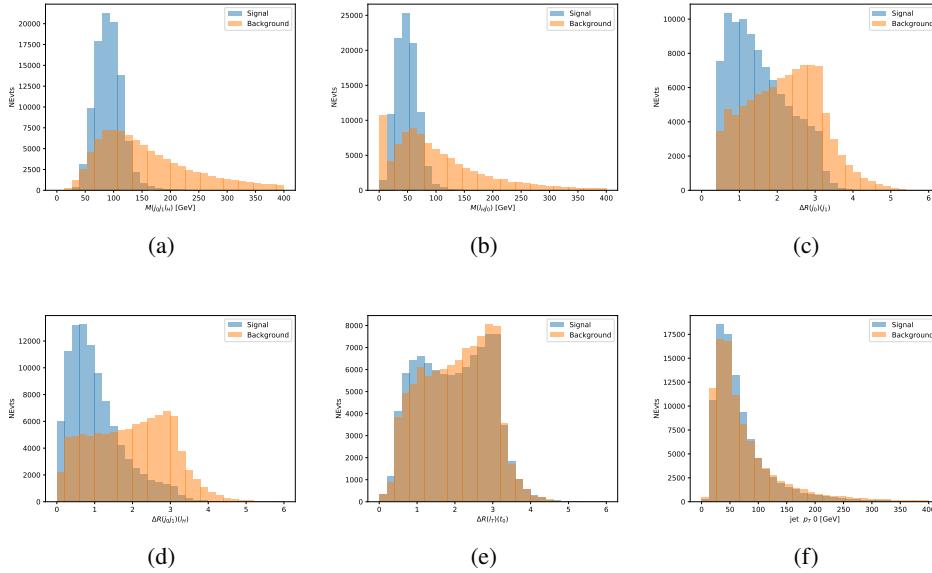


Figure 8.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c)  $\Delta R$  between the jet candidates, (d)  $\Delta R$  between jet 0 + jet 1 and the lepton candidate, (e)  $\Delta R$  between the lepton from the top and the leading b-jet, (f) the  $p_T$  of jet 0.

545 The modeling of these inputs is validated against data, with figure 8.2 showing good  
 546 general agreement between data and MC. Plots for the complete list of features can found in  
 547 section A.

548 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2  
 549 million events, with an additional 200,000 reserved for testing the model. In order to compensate

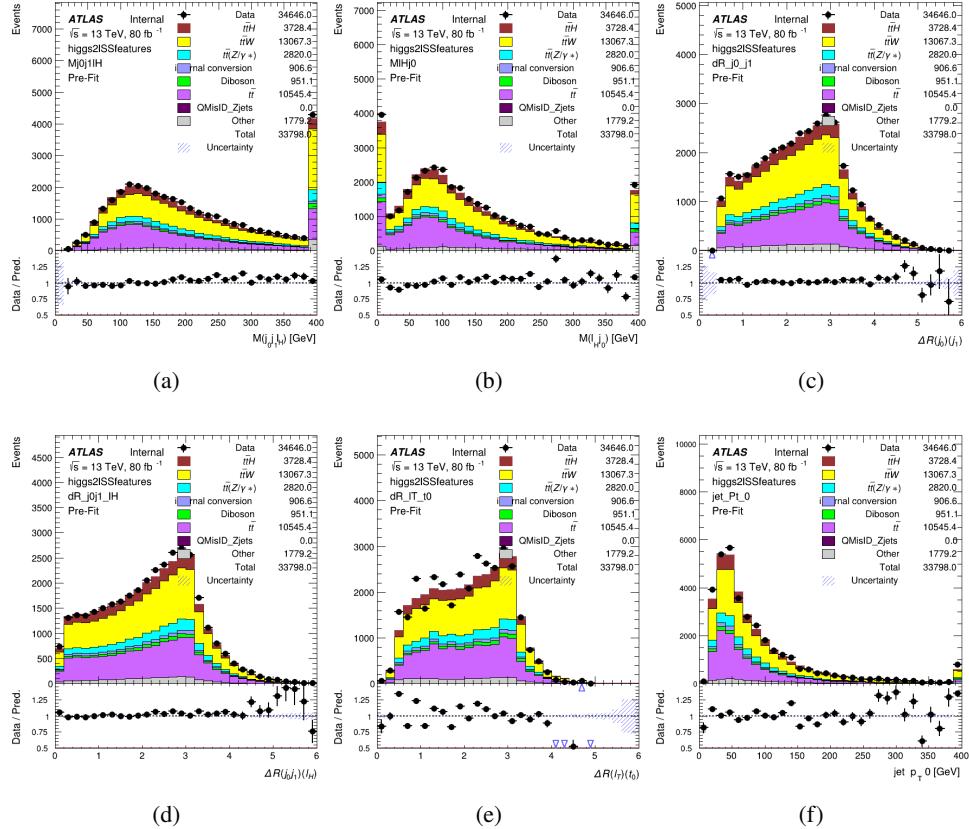


Figure 8.8: Data/MC comparisons of input features for higgs2lSS training for  $80 \text{ fb}^{-1}$  of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c)  $\Delta R$  between the jet candidates, (d)  $\Delta R$  between jet 0 + jet 1 and the lepton candidate, (e)  $\Delta R$  between the lepton from the top and the leading b-jet, (f) the  $p_T$  of jet 0.

550 for large number of incorrect combinations, these have been downsampled such that the correct  
 551 combinations represent over 10% of the training set. The output of the NN is summarized in  
 552 figure 8.3.1.

553 The neural network identifies the correct combination 55% of the time. It identifies the  
 554 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets  
 555 81% of the time.

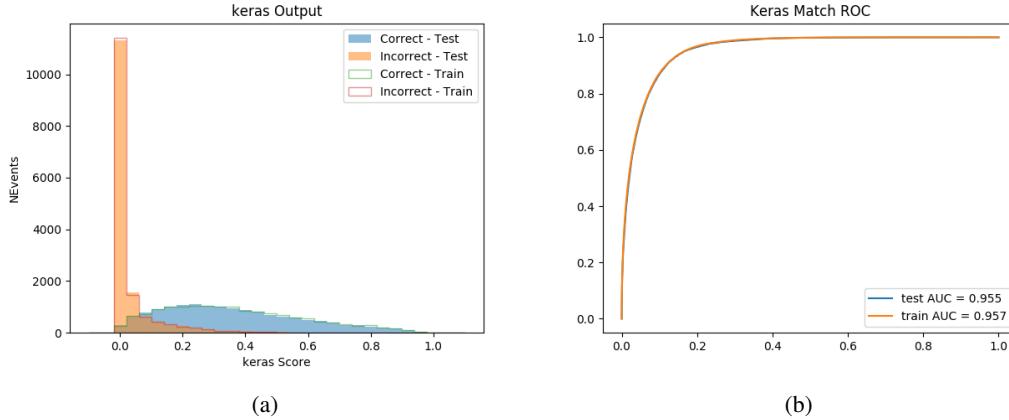


Figure 8.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

556    **8.3.2 3l Semi-leptonic Channel**

557    For 3l  $t\bar{t}H$  where the Higgs decay semi-leptonically, the decay products include one of the three  
 558    leptons and two jets. In this case, the other two leptons originated from the decay of the tops,  
 559    meaning the opposite-sign (OS) lepton cannot have come the Higgs. This leave only the two  
 560    same-sign (SS) leptons as possible Higgs decay products.

Lepton $p_T H$	Lepton $p_T T_0$	Lepton $p_T T_1$
jet $p_T 0$	jet $p_T 1$	top $p_T 0$
top $p_T 1$	jet $\eta 0$	jet $\eta 1$
jet $\phi 0$	jet $\phi 1$	$\Delta R(j_0)(j_1)$
$M(j_0 j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1)(l_{T_1})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0 j_1 l_{T_0})$	$M(j_0 j_1 l_{T_1})$
$M(j_0 j_1 l_H)$	$\Delta R(j_0 j_1 l_H)(l_{T_0})$	$\Delta R(j_0 j_1 l_H)(l_{T_1})$
$\Delta\phi(j_0 j_1 l_H)(E_T^{\text{miss}})$	$p_T(j_0 j_1 l_H l_{T_0} l_{T_1} b_0 b_1 E_T^{\text{miss}})$	$M(j_0 j_1 b_0)$
$M(j_0 j_1 b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
topScore	MET	HT jets
nJets		

Table 8: Input features used to identify the Higgs decay products in 3lS events

561 Here  $j_0$  and  $j_1$ , and  $l_H$  are the jet and lepton decay candidates, respectively. The other  
 562 two leptons in the event are labeled as  $l_{T0}$  and  $l_{T1}$ .  $b_0$  and  $b_1$  are the two b-jets identified by  
 563 the b-jet identification algorithm. The b-jet Reco Score is the output of the Higgs reconstruction  
 564 algorithm.

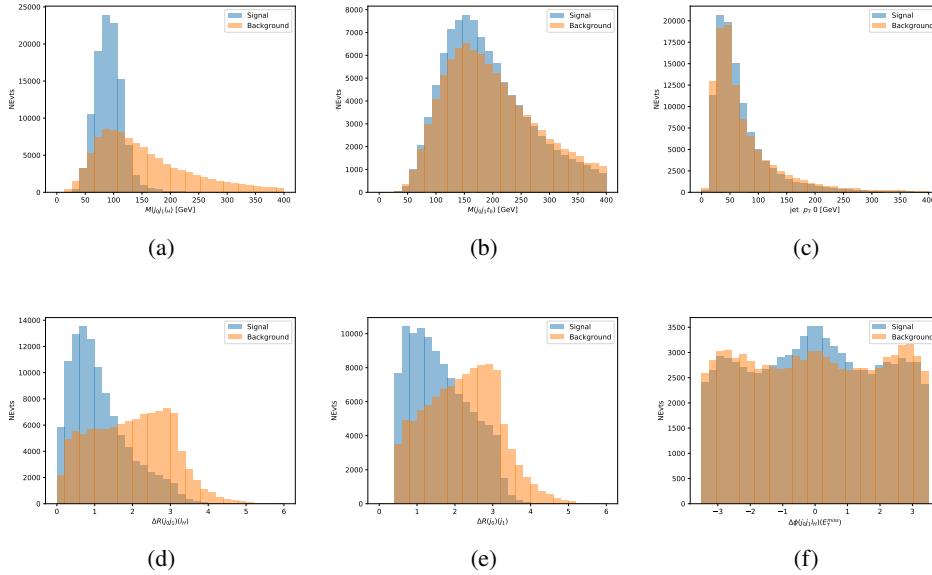


Figure 8.10: Input features for higgs3IS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

565 The modeling of these inputs is validated against data, with figure 8.11 showing good  
 566 general agreement between data and MC. Plots for the complete list of features can found in  
 567 appendix A.1.

568 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.  
 569 Once again, incorrect combinations are downsampled, such that the correct combinations are  
 570 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN

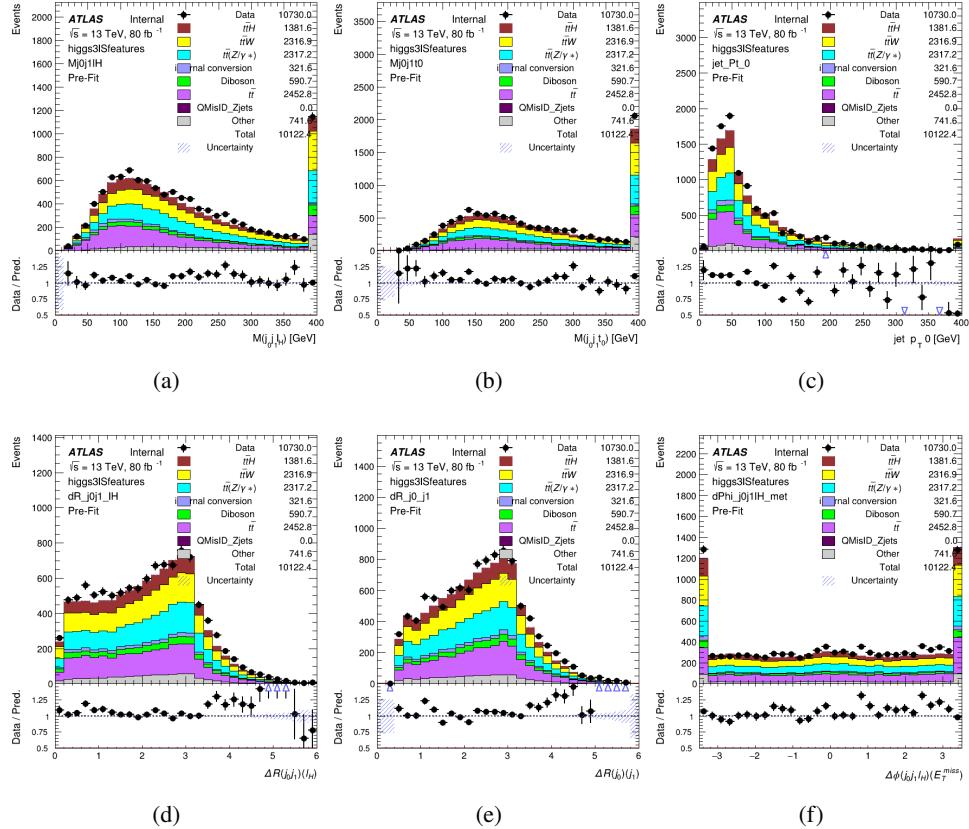


Figure 8.11: Data/MC comparisons of input features for higgs3IS training for  $80 \text{ fb}^{-1}$  of data.

571 is summarized in figure 8.3.2.

572 The neural network identifies the correct combination 64% of the time. It identifies the  
 573 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets  
 574 83% of the time.

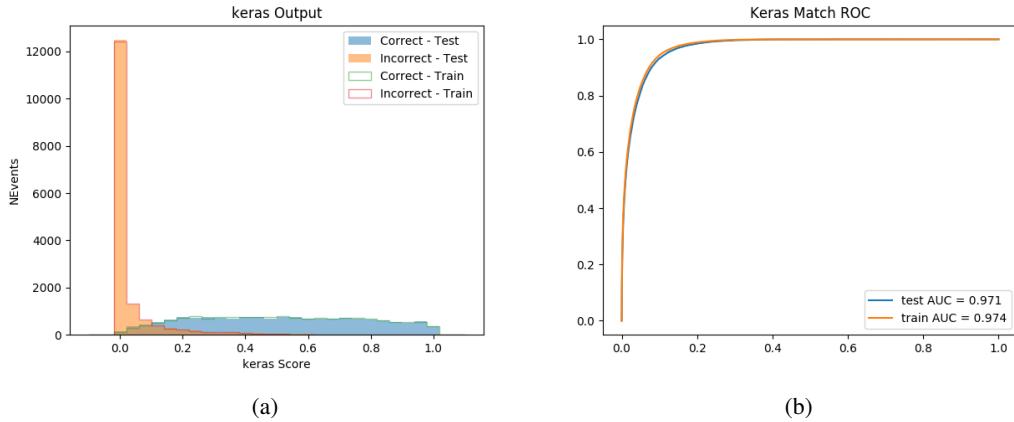


Figure 8.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

### 575 8.3.3 3l Fully-leptonic Channel

576 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from  
 577 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to  
 578 determining which of the two SS leptons originated from the Higgs. The kinematics of both  
 579 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,  
 580 and one where it is not.

Lepton $p_T H_1$	Lepton $p_T H_0$	Lepton $p_T T$
top $p_T 0$	top $p_T 1$	$\Delta\phi(l_{H_1})(E_T^{\text{miss}})$
$\Delta\phi(l_T)(E_T^{\text{miss}})$	$M(l_{H_0}l_{H_1})$	$M(l_{H_1}l_T)$
$M(l_{H_0}l_T)$	$\Delta R(l_{H_0})(l_{H_1})$	$\Delta R(l_{H_1})(l_T)$
$\Delta R(l_{H_0})(l_T)$	$\Delta R(l_{H_0}l_{H_1})(l_T)$	$\Delta R(l_{H_0}l_T)(l_{H_1})$
$\Delta R(l_{H_0}l_{H_1})(b_0)$	$\Delta R(l_{H_0}l_{H_1})(b_1)$	$\Delta R(l_{H_0}b_0)$
$M(l_{H_0}b_0)$	$\Delta R(l_{H_0}b_1)$	$M(l_{H_0}b_1)$
$\Delta R(l_{H_1}b_0)$	$M(l_{H_1}b_0)$	$\Delta R(l_{H_1}b_1)$
$M(l_{H_1}b_1)$	$E_T^{\text{miss}}$	topScore

Table 9: Input features used to identify the Higgs decay products in 3IF events

581 Here  $l_{H0}$  and  $l_{H1}$  are the Higgs decay candidates. The other lepton in the event is labeled  
 582  $l_T$ .  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco  
 583 Score is the output of the Higgs reconstruction algorithm.

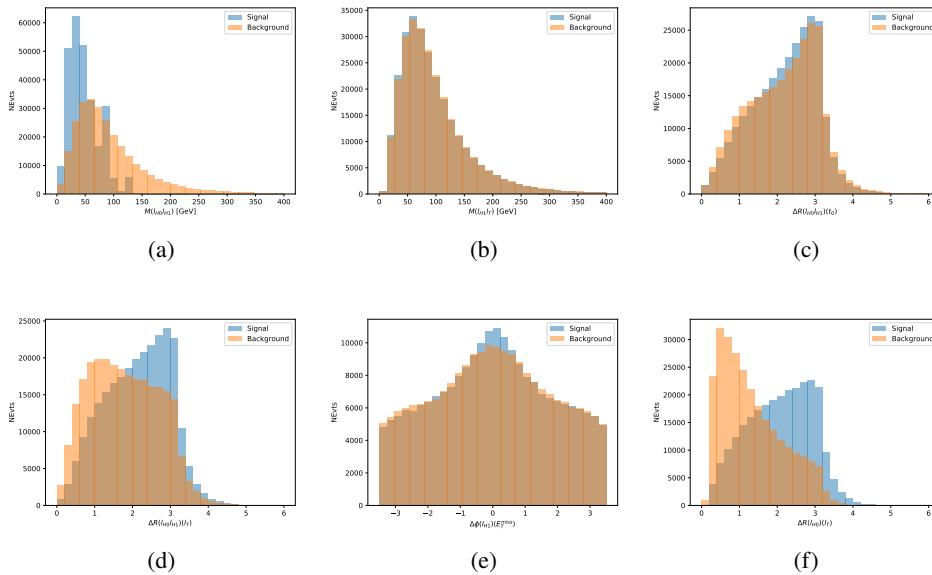


Figure 8.13: Input features for higgs3IF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

584 The modeling of these inputs is validated against data, with figure 8.14 showing good  
 585 general agreement between data and MC. Plots for the complete list of features can found in  
 586 section A.

587 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,  
 588 with 10% of the dataset reserved for testing. The output of the model is summarized in figure  
 589 8.3.3.

590 The correct lepton is identified by the model for 80% of events in the testing data set.

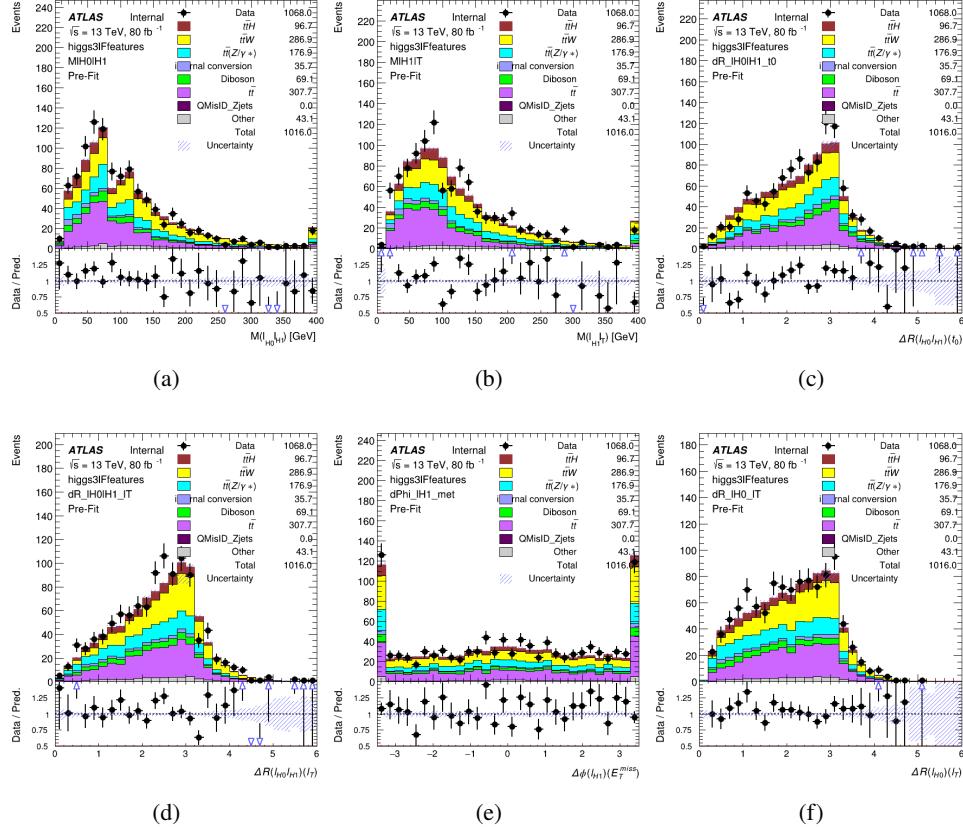


Figure 8.14: Data/MC comparisons of input features for higgs3IF training for  $80 \text{ fb}^{-1}$  of data.

## 591 8.4 $p_T$ Prediction

592 Once the most probable decay products have been identified, their kinematics are used as inputs  
 593 to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,  
 594 a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay  
 595 are included as well, as these are found to improve performance. The truth  $p_T$  of the Higgs,  
 596 as predicted by MC, are used as labels. Separate models are built for each channel - 2lSS, 3l  
 597 Semi-leptonic and 3l Fully-leptonic.

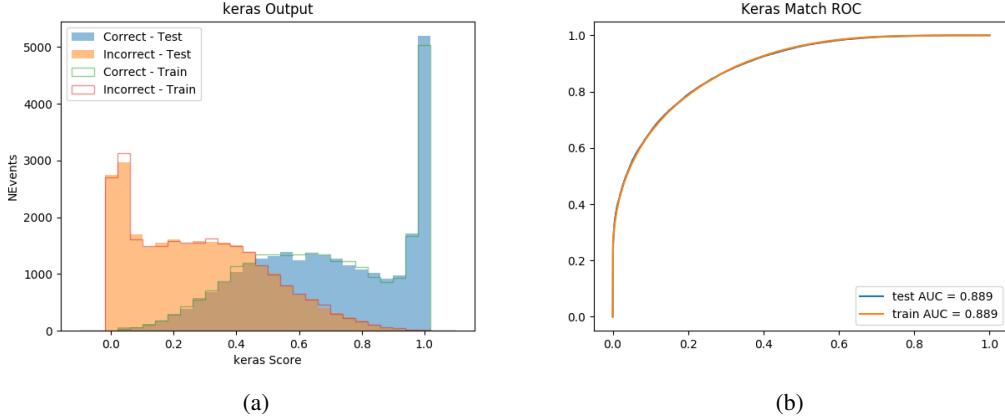


Figure 8.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

598 As a two-bin fit is targeted for the final result, some metrics evaluating the performance  
 599 of the models aim to show how well it distinguished between "high  $p_T$ " and "low  $p_T$ " events. A  
 600 cutoff point of 150 GeV is used to define these two categories.

601 Because the analysis uses a two bin fit of the Higgs  $p_T$ , the momentum reconstruction  
 602 could be treated as a binary classification problem, rather than a regression problem. This  
 603 approach is explored in detail in section A.4, and is found not to provide any significant increase  
 604 in sensitivity. The regression approach is used because it provides more flexibility for future  
 605 analyses, as it is independent of the cutoff between high and low  $p_T$ , as well as the number of  
 606 bins. Further, a regression allows the output of the neural network to be more clearly understood,  
 607 as it can be directly compared to a physics observable.

**608 8.4.1 2lSS Channel**

609 The input variables listed in table 10 are used to predict the Higgs  $p_T$  in the 2lSS channel. Here  
610  $j_0$  and  $j_1$  are the two jets identified as Higgs decay products. The lepton identified as originating  
611 from the Higgs is labeled  $l_H$ , while the other lepton is labeled  $l_T$ , as it is assumed to have come  
612 from the decay of one of the top quarks.  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet  
613 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs  
614 reconstruction algorithm, and the b-jet identification algorithm, respectively.

HT	$M(j_0 j_1)$	$M(j_0 j_1 l_H)$
$M(l_H j_0)$	$M(l_H j_1)$	$p_T(b_0 b_1)$
$p_T(j_0 j_1 l_H)$	$\Delta\phi(j_0 j_1 l_H)(E_T^{\text{miss}})$	$\Delta R(j_0)(j_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1 l_H)(l_T)$	$\Delta R(j_0 j_1 l_H)(b_0)$
$\Delta R(j_0 j_1 l_H)(b_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(b_0)$
$\Delta R(l_H)(b_1)$	$\Delta R(l_T)(b_0)$	$\Delta R(l_T)(b_1)$
$\Delta R(b_0)(b_1)$	Higgs Reco Score	jet $\eta$ 0
jet $\eta$ 1	jet Phi 0	jet Phi 1
jet $p_T$ 0	jet $p_T$ 1	Lepton $\eta$ H
Lepton $\phi$ H	Lepton $p_T$ H	Lepton $p_T$ T
$E_T^{\text{miss}}$	nJets	b-jet Reco Score
b-jet $p_T$ 0	b-jet $p_T$ 1	

Table 10: Input features for reconstructing the Higgs  $p_T$  spectrum for 2lSS events

615        The optimal neural network architecture for this channel is found to consist of 7 hidden  
 616        layers with 60 nodes each. The inputdata set includes 1.2 million events, 10% of which is used  
 617        for testing, the other 90% for training. Training is found to converge after around 150 epochs.

618        To evaluate the performance of the model, the predicted  $p_T$  spectrum is compared to the  
 619        truth Higgs  $p_T$  in figure 8.16. In order to visualize the model performance more clearly, in (a)  
 620        of that figure, the color of each point is determined by Kernal Density Estimation (KDE). The  
 621        color shown represents the logarithm of the output from KDE, to counteract the large number of  
 622        low  $p_T$  events. For that same reason, each column of the histogram shown in (b) of figure 8.16  
 623        is normalized to unity. This plot therefore demonstrates what the model predicts for each slice  
 624        of truth  $p_T$ .

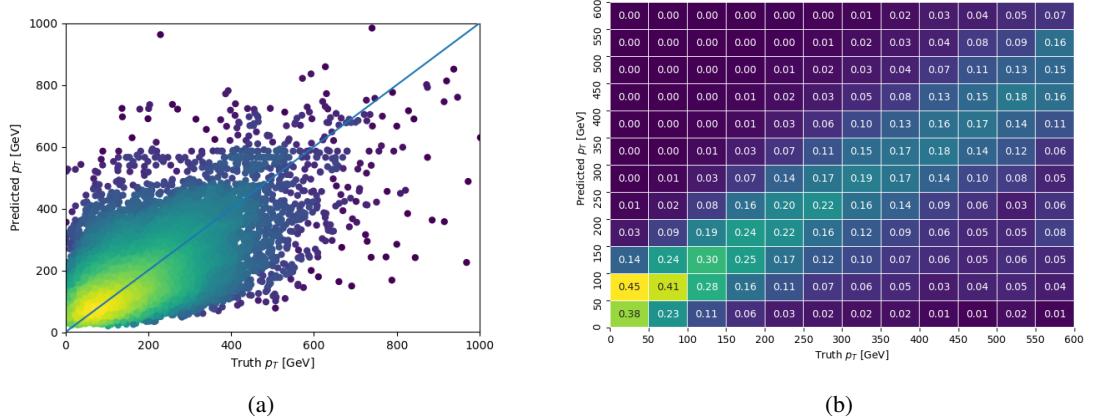


Figure 8.16: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 2ISS  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

625        We are also interested in how well the model distinguishes between events with  $p_T < 150$   
 626        GeV and  $> 150$  GeV. Figure 8.17 demonstrates the NN output for high and low  $p_T$  events based

627 on this cutoff.

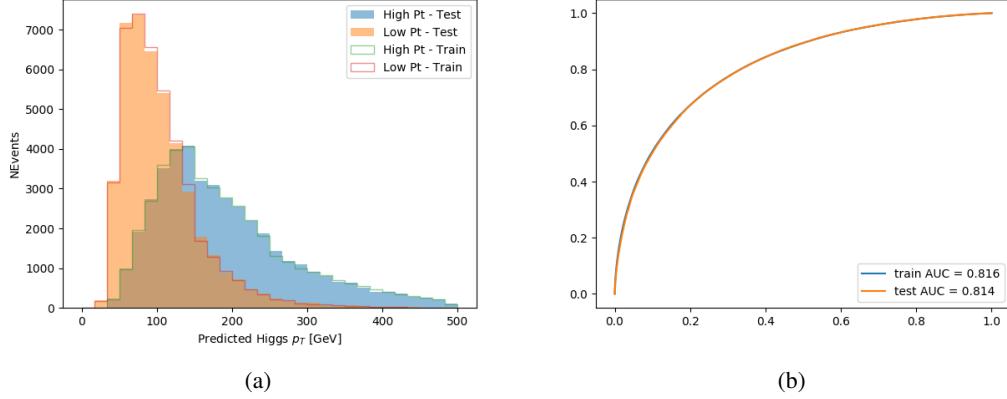


Figure 8.17: (a) shows the reconstructed Higgs  $p_T$  for 2lSS events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

#### 628 8.4.2 3l Semi-leptonic Channel

629 The following input features are used to predict the Higgs  $p_T$  for events in the 3lS channel:

HT jets	MET	$M(j_0 j_1)$
$M(j_0 j_1 l_H)$	$M(j_0 j_1 l_{T0})$	$M(j_0 j_1 l_{T1})$
$M(j_0 j_1 b_0)$	$M(j_0 j_1 b_1)$	$M(b_0 l_{T0})$
$M(b_0 l_{T1})$	$M(b_1 l_{T0})$	$M(b_1 l_{T1})$
$\Delta\phi(j_0 j_1 l_H)(E_T^{\text{miss}})$	$\Delta R(j_0)(j_1)$	$\Delta R(j_0 j_1)(l_H)$
$\Delta R(j_0 j_1)(l_{T1})$	$\Delta R(j_0 j_1)(b_0)$	$\Delta R(j_0 j_1)(b_1)$
$\Delta R(j_0 j_1 l_H)(l_{T0})$	$\Delta R(j_0 j_1 l_H)(l_{T1})$	$\Delta R(j_0 j_1 l_H)(b_0)$
$\Delta R(j_0 j_1 l_H)(b_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(l_H)(l_{T1})$	$\Delta R(l_{T0})(l_{T1})$	$\Delta R(l_{T0})(b_0)$
$\Delta R(l_{T0})(b_1)$	$\Delta R(l_{T1})(b_0)$	$\Delta R(l_{T1})(b_1)$
higgsScore	jet $\eta$ 0	jet $\eta$ 1
jet $\phi$ 0	jet $\phi$ 1	jet $p_T$ 0
jet $p_T$ 1	Lepton $\eta$ H	Lepton $\phi$ H
Lepton $p_T$ H	Lepton $p_T$ T0	Lepton $p_T$ T1
nJets	topScore	b-jet $p_T$ 0
b-jet $p_T$ 1		

Table 11: Input features for reconstructing the Higgs  $p_T$  spectrum for 3lS events

630        Again,  $j_0$  and  $j_1$  are the two jets identified as Higgs decay products, ordered by  $p_T$ . The  
 631      lepton identified as originating from the Higgs is labeled  $l_H$ , while the other two leptons are  
 632      labeled  $l_{T0}$  and  $l_{T1}$ .  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet identification algorithm.  
 633      The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs reconstruction algorithm,  
 634      and the b-jet identification algorithm, respectively.

635        The optimal neural network architecture for this channel is found to consist of 7 hidden  
 636      layers with 80 nodes each. The inputdata set includes one million events, 10% of which is used  
 637      for testing, the other 90% for training. Training is found to converge after around 150 epochs.

638        To evaluate the performance of the model, the predicted  $p_T$  spectrum is compared to the  
 639      truth Higgs  $p_T$  in figure 8.18. Once again, (a) of 8.18 shows a scatterplots of predicted vs truth  
 640       $p_T$ , where the color of each point corresponds to the log of the relative KDE at that point. Each  
 641      column of the the histogram in (b) is normalized to unity, to better demonstrate the output of the  
 642      NN for each slice of truth  $p_T$ .

643        Figure 8.19 shows (a) the output of the NN for events with truth  $p_T$  less than and greater  
 644      than 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low  
 645       $p_T$  events.

#### 646      **8.4.3 3l Fully-leptonic Channel**

647      The features listed in 12 are used to construct a model for predictin the Higgs  $p_T$  for 3lF events.

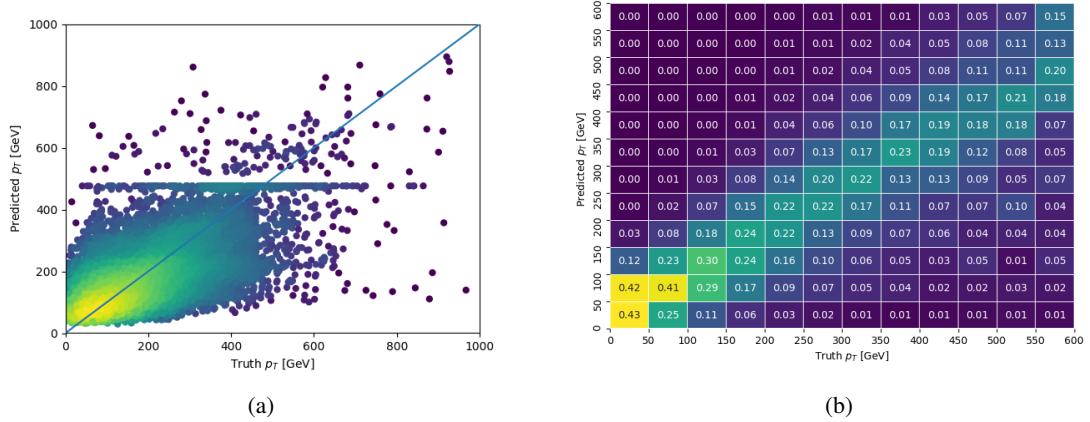


Figure 8.18: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 3lS  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

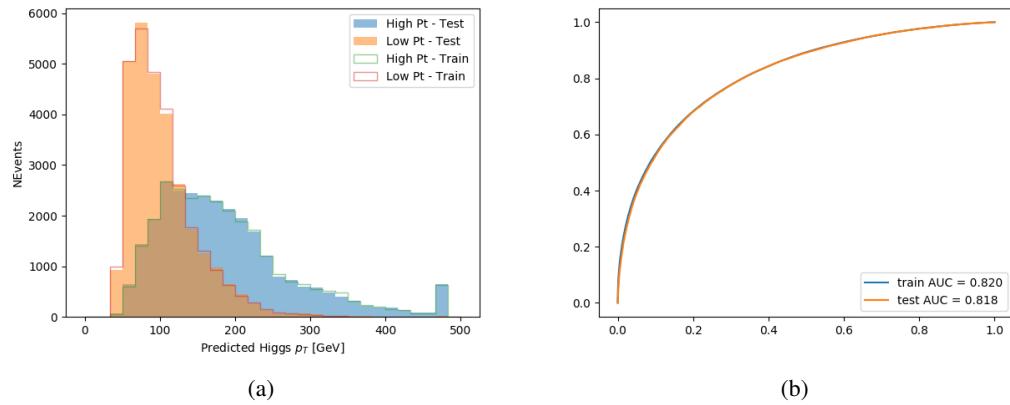


Figure 8.19: (a) shows the reconstructed Higgs  $p_T$  for 3lS events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

---

HT	$M(l_{H0}l_{H1})$	$M(l_{H0}l_T)$
$M(l_{H0}b_0)$	$M(l_{H0}b_1)$	$M(l_{H1}l_T)$
$M(l_{H1}b_0)$	$M(l_{H1}b_1)$	$\Delta R(l_{H0})(l_{H1})$
$\Delta R(l_{H0})(l_T)$	$\Delta R(l_{H0}l_{H1})(l_T)$	$\Delta R(l_{H0}l_T)(l_{H1})$
$\Delta R(l_{H1})(l_T)$	$\Delta R(l_{H0}b_0)$	$\Delta R(l_{H0}b_1)$
$\Delta R(l_{H1}b_1)$	$\Delta R(l_{H1}b_0)$	higgsScore
Lepton $\eta$ $H_0$	Lepton $\eta$ $H_1$	Lepton $\eta$ T
Lepton $p_T$ $H_0$	Lepton $p_T$ $H_1$	Lepton $p_T$ T
$E_T^{\text{miss}}$	topScore	b-jet $p_T$ 0
b-jet $p_T$ 1		

Table 12: Input features for reconstructing the Higgs  $p_T$  spectrum for 3lF events

648         $l_{H0}$  and  $l_{H1}$  represent the two leptons identified by the Higgs reconstruction model as  
 649        originating from the Higgs, while  $l_T$  is the other lepton in the event. The Higgs Reco Score and  
 650        b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification  
 651        algorithm, respectively.

652        The optimal neural network architecture for this channel is found to consist of 5 hidden  
 653        layers with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for  
 654        testing, the other 90% for training. Training is found to converge after around 150 epochs.

655        The predicted transverse momentum, as a function of the truth  $p_T$ , is shown in figure  
 656        [8.20](#).

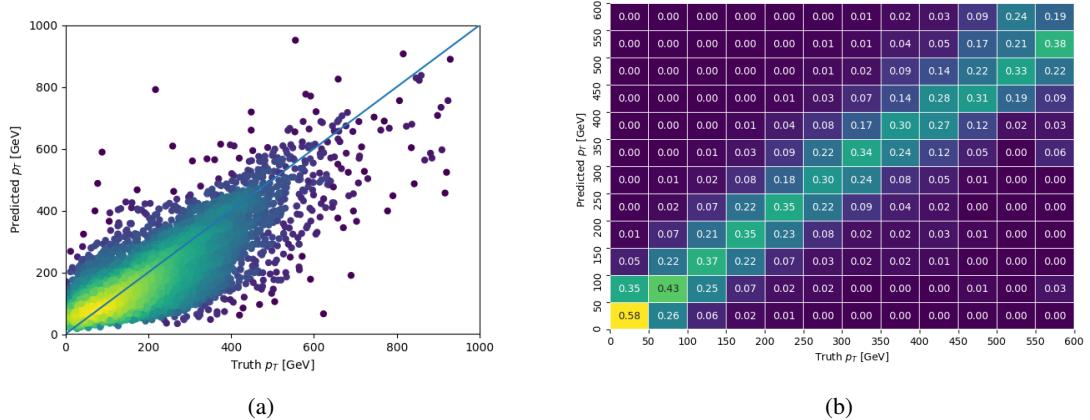


Figure 8.20: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 3lF  $t\bar{t}H$  events in  
 (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian  
 Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

657        When split into high and low  $p_T$ , based on a cutoff of 150 GeV, the

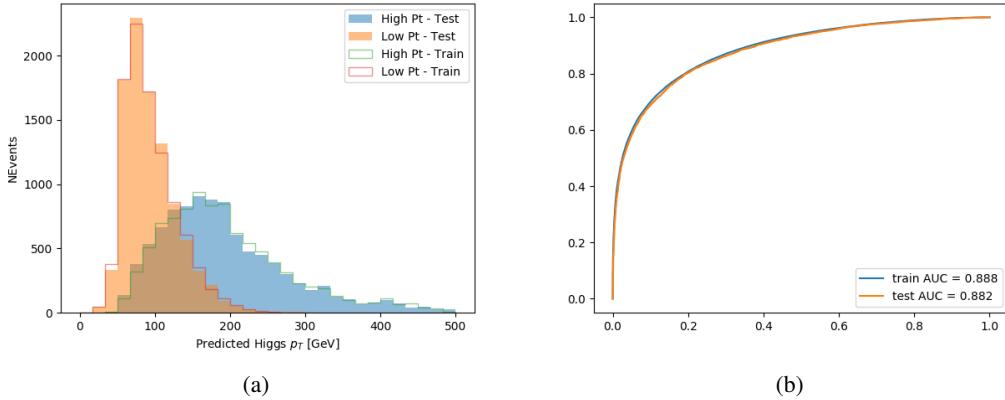


Figure 8.21: (a) shows the reconstructed Higgs  $p_T$  for 3lF events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

## 658 8.5 3l Decay Mode

659 In the 3l channel, there are two possible ways for the Higgs to decay, both involving intermediate  
 660 W boson pairs: Either both W bosons decay leptonically, in which case the reconstructed decay  
 661 consists of two leptons (referred as the fully-leptonic 3l channel), or one W decays leptonically  
 662 and the other hadronically, giving two jets and one lepton in the final state (referred to as the  
 663 semi-leptonic 3l channel). In order to accurately reconstruct the Higgs, it is necessary to identify  
 664 which of these decays took place for each 3l event.

665 The kinematics of each event, along with the output scores of the Higgs and top recon-  
 666 struction algorithms, are used to distinguish these two possible decay modes. The particular  
 667 inputs used are listed in table 13.

HT jets	$M(l_0 t_0)$	$M(l_0 t_1)$
$M(l_1 t_0)$	$M(l_1 t_1)$	$M(l_0 l_1)$
$M(l_0 l_2)$	$M(l_1 l_2)$	$\Delta R(l_0 t_0)$
$\Delta R(l_0 t_1)$	$\Delta R(l_1 t_0)$	$\Delta R(l_1 t_1)$
$\Delta R(l l_0 1)$	$\Delta R(l l_0 2)$	$\Delta R(l l_1 2)$
Lepton $\eta$ 0	Lepton $\eta$ 1	Lepton $\eta$ 2
Lepton $\phi$ 0	Lepton $\phi$ 1	Lepton $\phi$ 2
Lepton $p_T$ 0	Lepton $p_T$ 1	Lepton $p_T$ 2
$E_T^{\text{miss}}$	nJets	nJets OR DL1r 60
nJets OR DL1r 85	score3lF	score3lS
topScore	total charge	

Table 13: Input features used to distinguish semi-leptonic and fully-leptonic Higgs decays in the 3l channel.

668 Here  $l_0$  is the opposite charge lepton,  $l_1$  and  $l_2$  are the two SS leptons order by  $\Delta R$   
 669 from lepton 0. score3lF and score3lS are the outputs of the 3lS and 3lF Higgs reconstruction  
 670 algorithms, while topScore is the output of the b-jet identification algorithm.

671 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these  
 672 two decay modes. The output of the model is summarized in figure 8.22.

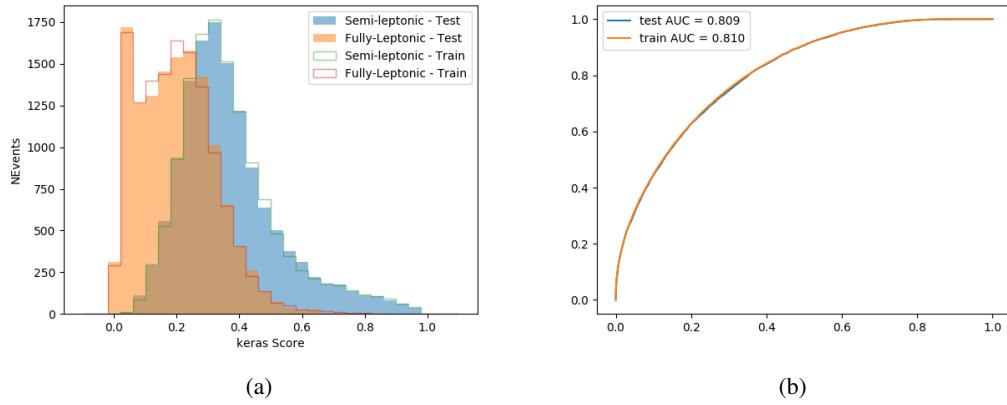


Figure 8.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

673 A cutoff of 0.23 is determined to be optimal for separating 3lS and 3lF in the fit.

## 674 9 Signal Region Definitions

675 Events are divided into two channels based on the number of leptons in the final state: one with  
 676 two same-sign leptons, the other with three leptons. The 3l channel includes events where both  
 677 leptons originated from the Higgs boson as well as events where only one of the leptons

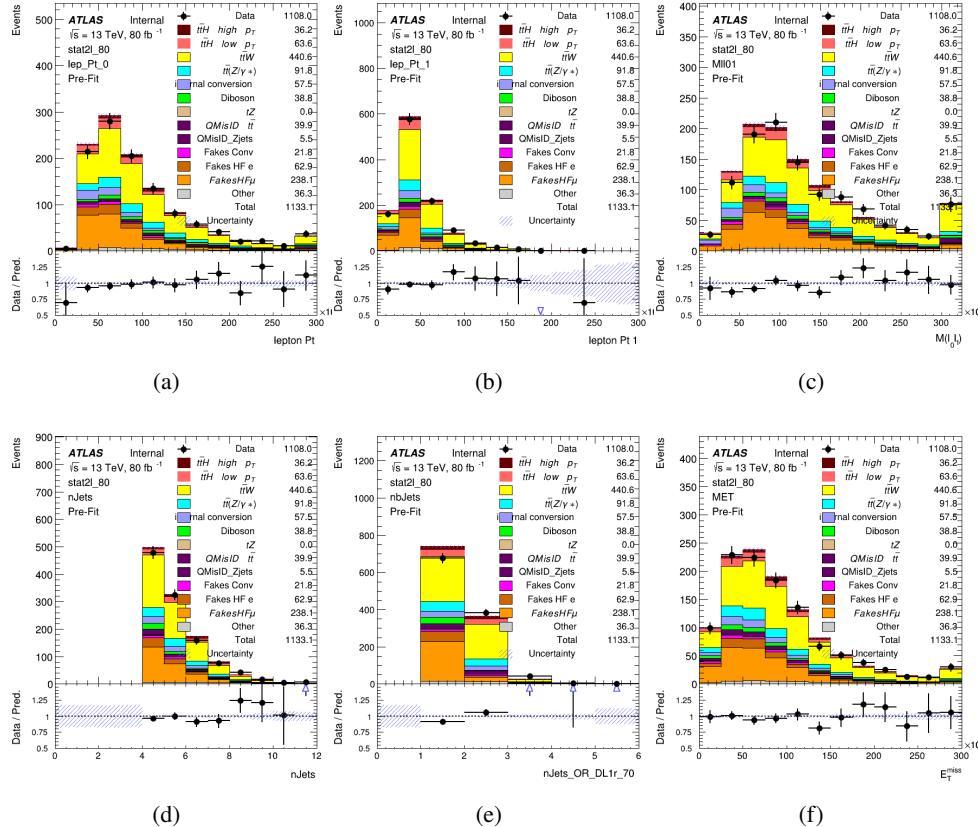
678 **9.1 Pre-MVA Event Selection**

679 A preselection is applied to define orthogonal analysis channels based on the number of leptons  
680 in each event. For the 2lSS channel, the following preselection is used:

- 681     • Two very tight, same-charge, light leptons with  $p_T > 20 \text{ GeV}$   
682     •  $\geq 4$  reconstructed jets,  $\geq 1$  b-tagged jets  
683     • No reconstructed tau candidates

684 The event yield after the 2lSS preselection has been applied, for MC and data at  $80 \text{ fb}^{-1}$ ,  
685 is shown in table ??.

figure 10.1. Good general agreement is found.

Figure 9.1: Data/MC comparisons of the 2lSS pre-selection region. (a) and (b) show the  $p_T$  of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

For the 3l channel, the following selection is applied:

- Three light leptons with total charge  $\pm 1$
- Same charge leptons are required to be very tight, with  $p_T > 20 \text{ GeV}$
- Opposite charge lepton must be loose, with  $p_T > 10 \text{ GeV}$

- 691        •  $\geq 2$  reconstructed jets,  $\geq 1$  b-tagged jets
- 692        • No reconstructed tau candidates
- 693        •  $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$  for all opposite-charge, same-flavor lepton pairs
- 694        The event yield after the 3l preselection has been applied, for MC and data at  $80 \text{ fb}^{-1}$ , is
- 695        shown in table ??.

696

Comparisons of kinematic distributions for data and MC in this region are shown in figure

697

## 10.2.

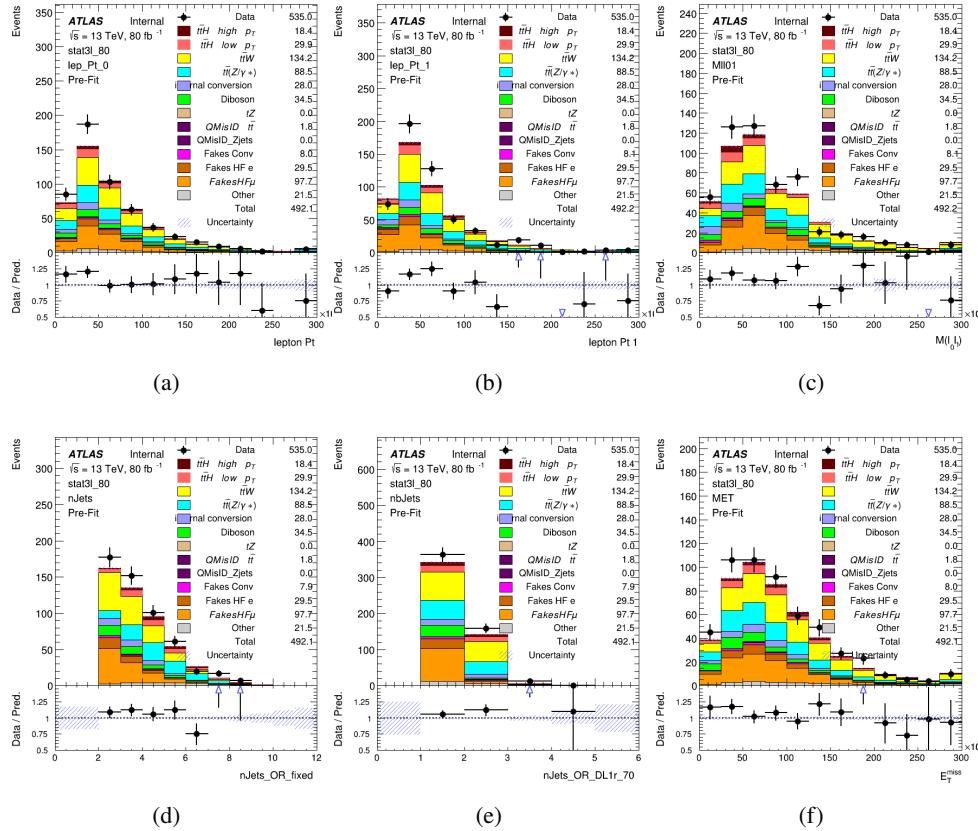


Figure 9.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the  $p_T$  of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

698

## 9.2 Event MVA

699

Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

702 [xgboost] are trained using the kinematics of signal and background events derived from Monte  
703 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo  
704 event.

705 Because the background composition differs for events with a high reconstructed Higgs  $p_T$   
706 compared to events with low reconstructed Higgs  $p_T$ , separate MVAs are produced for high and  
707 low  $p_T$  regions. This is found to provide better significance than attempting to build an inclusive  
708 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6  
709 background rejection MVAs - explicitly, 2lSS high  $p_T$ , 2lSS low  $p_T$ , 3lS high  $p_T$ , 3lS low  $p_T$ ,  
710 3lF high  $p_T$ , and 3lF low  $p_T$ .

711 The following features are used in both the high and low  $p_T$  2lSS BDTs:

HT	$M(\text{lep}, E_T^{\text{miss}})$	$M(l_0 l_1)$
binHiggs $p_T$ 2ISS	$\Delta R(l_0)(l_1)$	diLepton type
higgsRecoScore	jet $\eta$ 0	jet $\eta$ 1
jet $\phi$ 0	jet $\phi$ 1	jet $p_T$ 0
jet $p_T$ 1	Lepton $\eta$ 0	Lepton $\eta$ 1
Lepton $\phi$ 0	Lepton $\phi$ 1	Lepton $p_T$ 0
Lepton $p_T$ 1	$E_T^{\text{miss}}$	$\min \Delta R(l_0)(\text{jet})$
$\min \Delta R(l_1)(\text{jet})$	$\min \Delta R(\text{Lepton})(\text{bjet})$	mjjMax frwdJet
nJets	nJets OR DL1r 60	nJets OR DL1r 70
nJets OR DL1r 85	topRecoScore	

Table 14: Input features used to distinguish signal and background events in the 2ISS channel.

712

While for each of the 31 BDTs, the features listed below are used for training:

$M(\text{lep}, E_T^{\text{miss}})$	$M(l_0 l_1)$	$M(l_0 l_1 l_2)$
$M(l_0 l_2)$	$M(l_1 l_2)$	$\text{binHiggs } p_T \text{ 3lF}$
$\text{binHiggs } p_T \text{ 3lS}$	$\Delta R(l_0)(l_1)$	$\Delta R(l_0)(l_2)$
$\Delta R(l_1)(l_2)$	$\text{decayScore}$	$\text{higgsRecoScore3lF}$
$\text{higgsRecoScore3lS}$	$\text{jet } \eta \text{ 0}$	$\text{jet } \eta \text{ 1}$
$\text{jet } \phi \text{ 0}$	$\text{jet } \phi \text{ 1}$	$\text{jet } p_T \text{ 0}$
$\text{jet } p_T \text{ 1}$	$\text{Lepton } \eta \text{ 0}$	$\text{Lepton } \eta \text{ 1}$
$\text{Lepton } \eta \text{ 2}$	$\text{Lepton } \phi \text{ 0}$	$\text{Lepton } \phi \text{ 1}$
$\text{Lepton } \phi \text{ 2}$	$\text{Lepton } p_T \text{ 0}$	$\text{Lepton } p_T \text{ 1}$
$\text{Lepton } p_T \text{ 2}$	$E_T^{\text{miss}}$	$\min \Delta R(l_0)(\text{jet})$
$\min \Delta R(l_1)(\text{jet})$	$\min \Delta R(l_2)(\text{jet})$	$\min \Delta R(\text{Lepton})(\text{bjet})$
$mjj\text{Max frwdJet}$	$n\text{Jets}$	$n\text{Jets OR DL1r 60}$
$n\text{Jets OR DL1r 70}$	$n\text{Jets OR DL1r 85}$	$\text{topScore}$

Table 15: Input features used to distinguish signal and background events in the 3l channel.

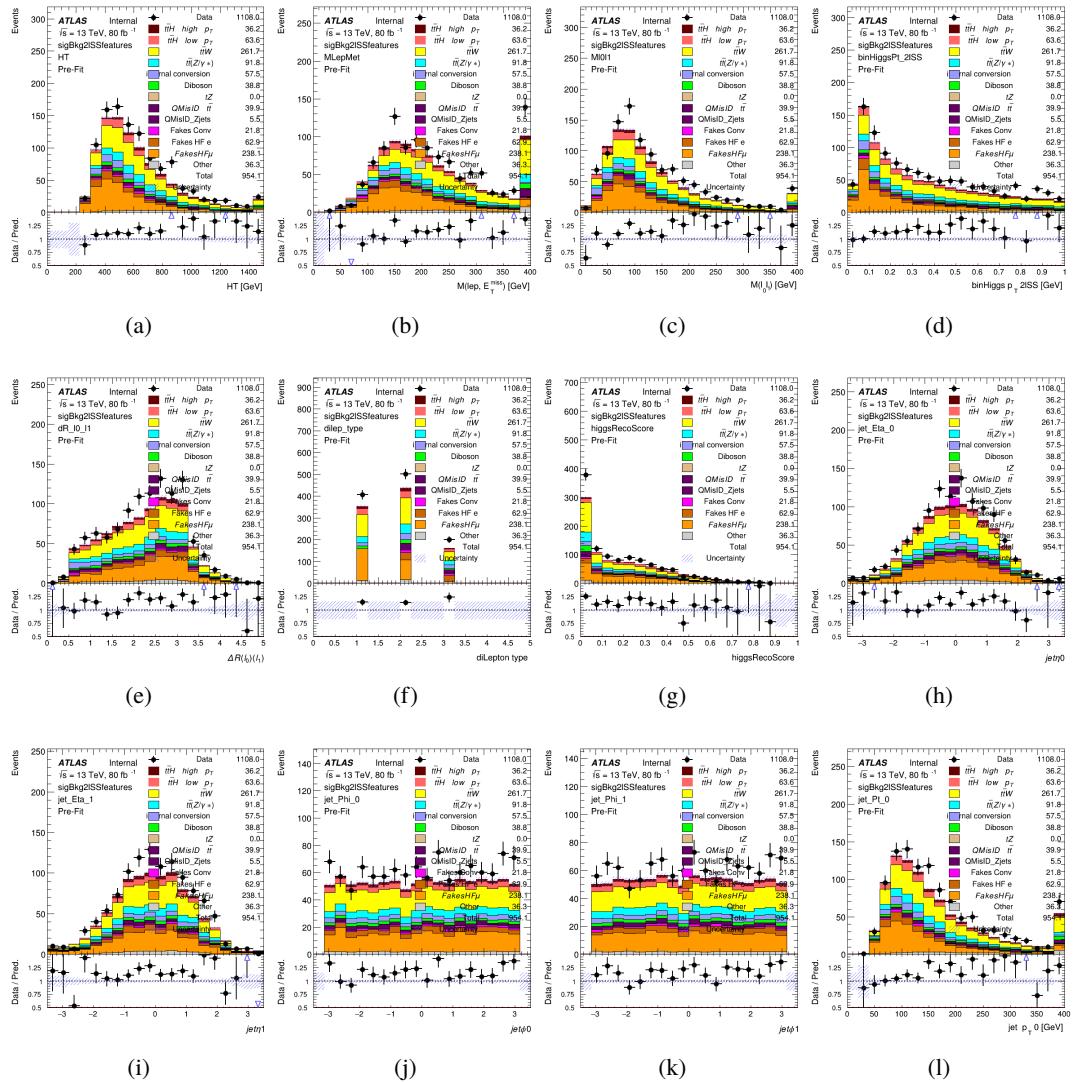


Figure 9.3:

713      The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss  
714      function.

715      Output distributions of each MVA comparing MC prediction to data at  $80 \text{ fb}^{-1}$  are shown  
716      in figures 9.7-9.2.

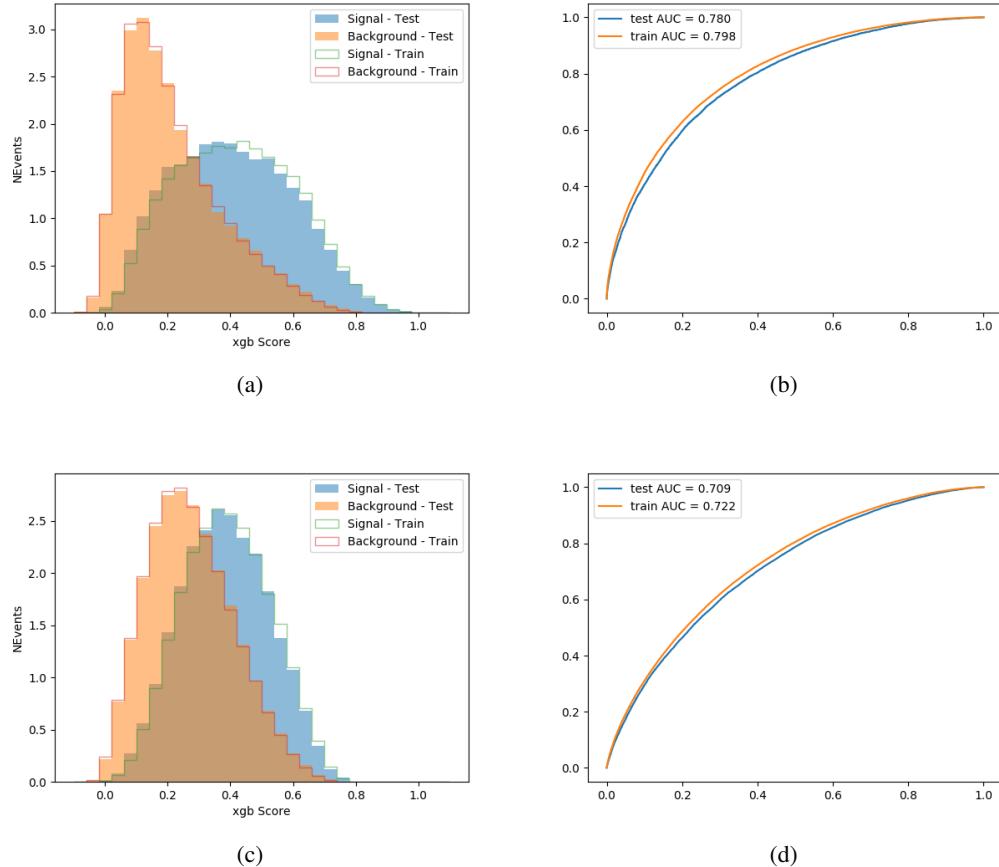


Figure 9.4:

### 717 9.3 Signal Region Definitions

718 Once pre-selection has been applied, channels are further refined based on the MVAs described  
 719 above. The output of the model described in section 8.5 is used to separate the three channel into  
 720 two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson.  
 721 This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

722 For each event, depending on the number of leptons as well as whether the  $p_T$  of the Higgs

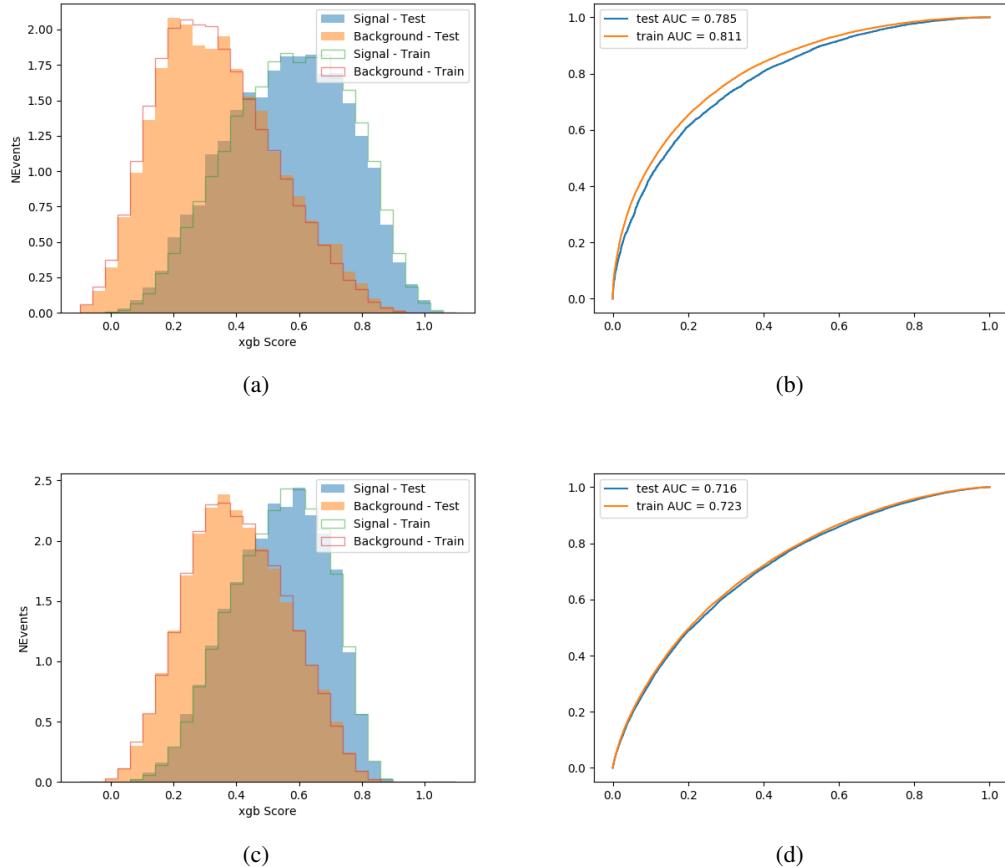


Figure 9.5:

is predicted to be high ( $> 150$  GeV) or low ( $< 150$  GeV), a cut on the appropriate background rejection MVA is applied. The particular cut values, listed in table 16, are determined by maximizing  $S/\sqrt{B}$  in each region.

The event preselection and MVA selection define the three signal regions. These signal region definitions are summarized in table 17.

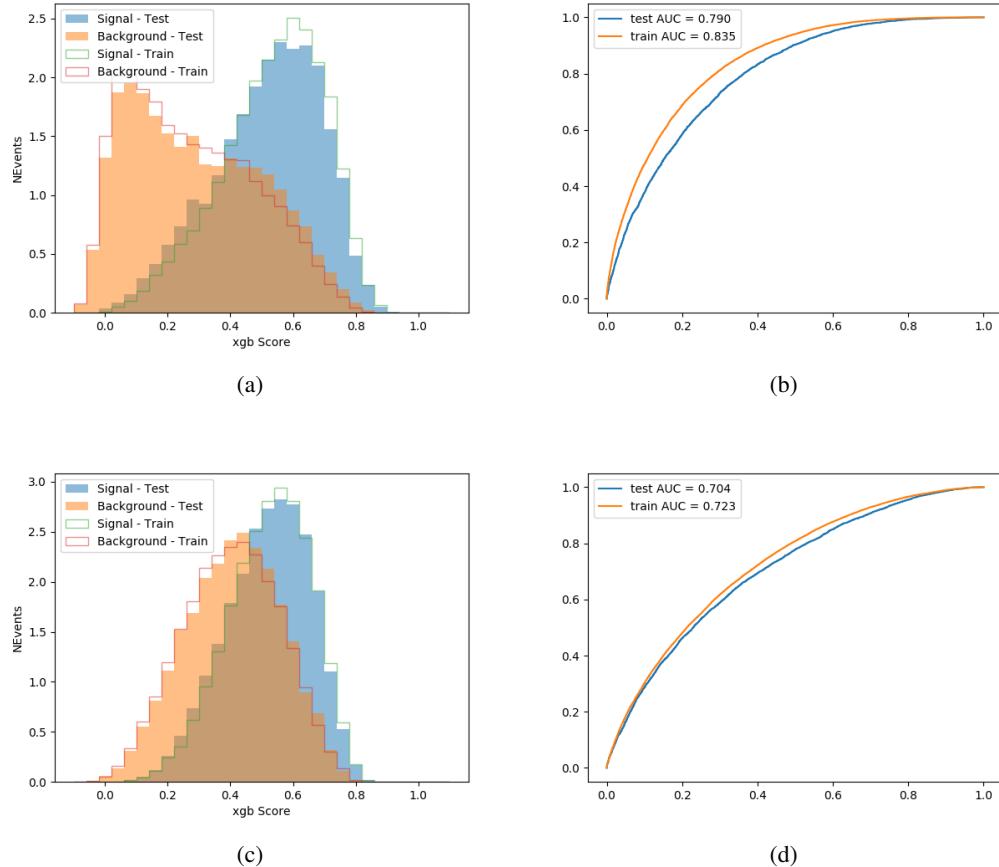
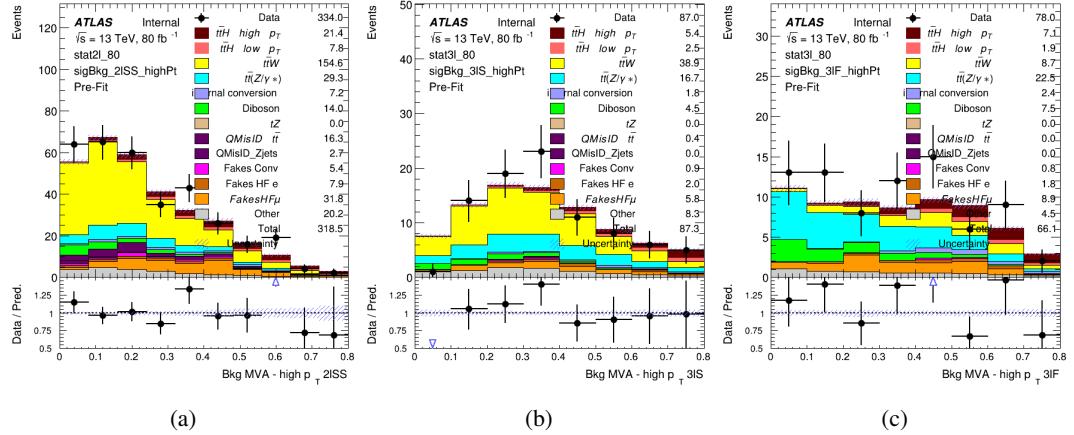
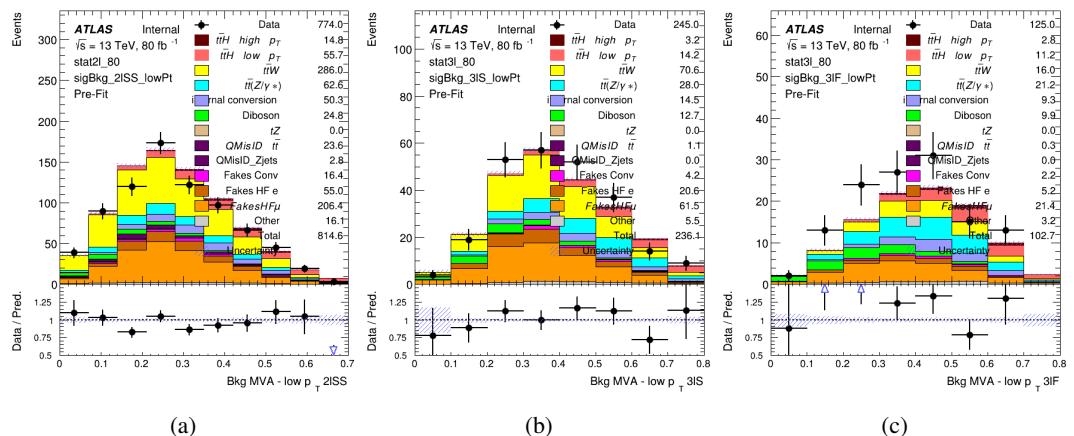


Figure 9.6:

Channel	BDT Score
2lSS high $p_T$	0.36
2lSS low $p_T$	0.34
3lS high $p_T$	0.51
3lS low $p_T$	0.43
3lF high $p_T$	0.33
3lF low $p_T$	0.41

Table 16: Cutoff values on background rejection MVA score applied to signal regions.

Figure 9.7: Output score of the high  $p_T$  BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channelsFigure 9.8: Output score of the low  $p_T$  BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channels

Region	Selection
2ISS	Two same charge tight leptons with $p_T > 20 \text{ GeV}$ $N_{\text{jets}} \geq 4, N_{\text{b-jets}} \geq 1$ b-tagged jets Zero $\tau_{\text{had}}$ $H_{p_T}^{\text{pred}} > 150 \text{ GeV}$ and BDT score $> 0.36$ <b>or</b> $H_{p_T}^{\text{pred}} < 150 \text{ GeV}$ and BDT score $> 0.34$
3IS	Three light leptons with total charge $\pm 1$ Two tight SS leptons, $p_T > 20 \text{ GeV}$ One loose OS lepton, $p_T > 10 \text{ GeV}$ $N_{\text{jets}} \geq 2, N_{\text{b-jets}} \geq 1$ b-tagged jets Zero $\tau_{\text{had}}$ $ M(l^+l^-) - 91.2 \text{ GeV}  > 10 \text{ GeV}$ for all OSSF lepton pairs Decay NN Score $< 0.23$ $H_{p_T}^{\text{pred}} > 150 \text{ GeV}$ and BDT score $> 0.51$ <b>or</b> $H_{p_T}^{\text{pred}} > 150 \text{ GeV}$ and BDT score $> 0.43$
3IF	Three light leptons with total charge $\pm 1$ Two tight SS leptons, $p_T > 20 \text{ GeV}$ One loose OS lepton, $p_T > 10 \text{ GeV}$ $N_{\text{jets}} \geq 2, N_{\text{b-jets}} \geq 1$ b-tagged jets Zero $\tau_{\text{had}}$ $ M(l^+l^-) - 91.2 \text{ GeV}  > 10 \text{ GeV}$ for all OSSF lepton pairs Decay NN Score $> 0.23$ $H_{p_T}^{\text{pred}} > 150 \text{ GeV}$ and BDT score $> 0.33$ <b>or</b> $H_{p_T}^{\text{pred}} > 150 \text{ GeV}$ and BDT score $> 0.41$

Table 17: Selection applied to define the three signal regions used in the fit.

## 10 Background Rejection MVA

728 Events are divided into two channels based on the number of leptons in the final state: one with  
 729 two same-sign leptons, the other with three leptons. The 3l channel includes events where both  
 730 leptons originated from the Higgs boson as well as events where only one of the leptons

## 732 10.1 Pre-MVA Event Selection

733 A preselection is applied to define orthogonal analysis channels based on the number of leptons  
 734 in each event. For the 2lSS channel, the following preselection is used:

- 735 • Two very tight, same-charge, light leptons with  $p_T > 20$  GeV

- 736 •  $>=4$  reconstructed jets,  $>=1$  b-tagged jets

- 737 • No reconstructed tau candidates

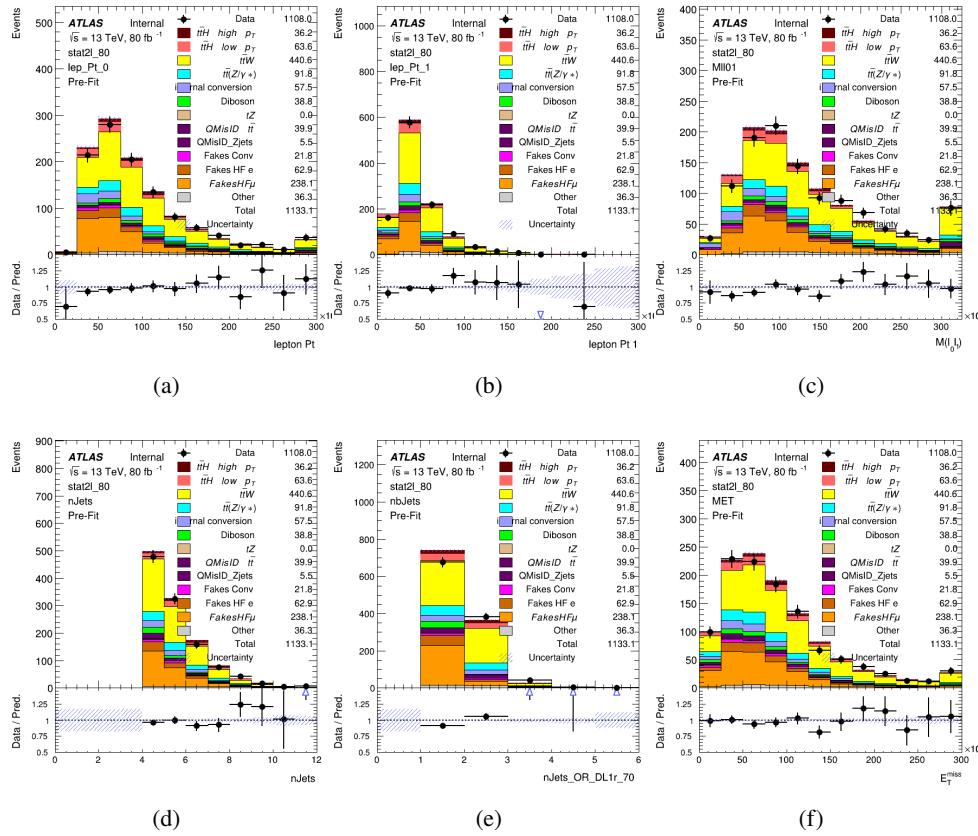


Figure 10.1:

738 For the 3l channel, the following selection is applied:

739 • Three light leptons with total charge  $\pm 1$

740 • Same charge leptons are required to be very tight, with  $p_T > 20 \text{ GeV}$

741 • Opposite charge lepton must be loose, with  $p_T > 10 \text{ GeV}$

742 •  $>= 2$  reconstructed jets,  $>= 1$  b-tagged jets

743 • No reconstructed tau candidates

744 •  $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$  for all opposite-charge, same-flavor lepton pairs

## 745 10.2 Event MVA

746 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events

747 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic

748 (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

749 [xgboost] are trained using the kinematics of signal and background events derived from Monte

750 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo

751 event.

752 Because the background composition differs for events with a high reconstructed Higgs  $p_T$

753 compared to events with low reconstructed Higgs  $p_T$ , separate MVAs are produced for high and

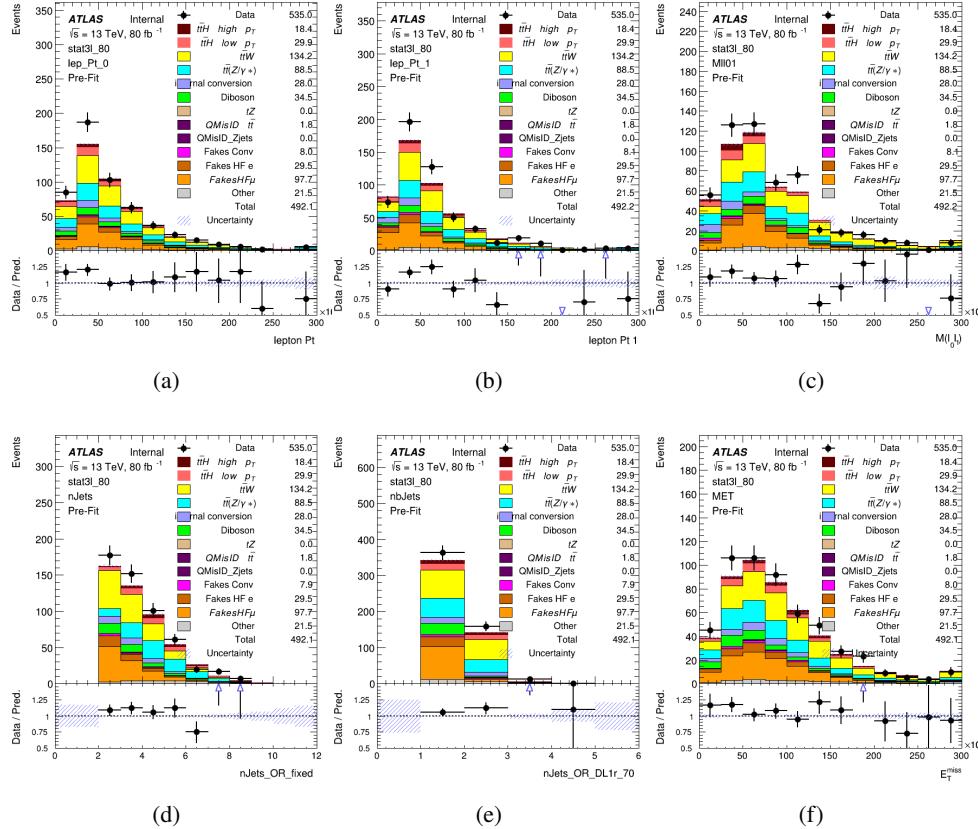


Figure 10.2:

754 low  $p_T$  regions. This is found to provide better significance than attempting to build an inclusive  
 755 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6  
 756 background rejection MVAs - explicitly, 2lSS high  $p_T$ , 2lSS low  $p_T$ , 3lS high  $p_T$ , 3lS low  $p_T$ ,  
 757 3lF high  $p_T$ , and 3lF low  $p_T$ .

758 The following features are used in both the high and low  $p_T$  2lSS BDTs:

HT	$M(\text{lep}, E_T^{\text{miss}})$	$M(l_0 l_1)$
binHiggs $p_T$ 2ISS	$\Delta R(l_0)(l_1)$	diLepton type
higgsRecoScore	jet $\eta$ 0	jet $\eta$ 1
jet $\phi$ 0	jet $\phi$ 1	jet $p_T$ 0
jet $p_T$ 1	Lepton $\eta$ 0	Lepton $\eta$ 1
Lepton $\phi$ 0	Lepton $\phi$ 1	Lepton $p_T$ 0
Lepton $p_T$ 1	$E_T^{\text{miss}}$	$\min \Delta R(l_0)(\text{jet})$
$\min \Delta R(l_1)(\text{jet})$	$\min \Delta R(\text{Lepton})(\text{bjet})$	mjjMax frwdJet
nJets	nJets OR DL1r 60	nJets OR DL1r 70
nJets OR DL1r 85	topRecoScore	

Table 18: Input features used to distinguish signal and background events in the 2ISS channel.

759

While for each of the 31 BDTs, the features listed below are used for training:

$M(\text{lep}, E_T^{\text{miss}})$	$M(l_0 l_1)$	$M(l_0 l_1 l_2)$
$M(l_0 l_2)$	$M(l_1 l_2)$	$\text{binHiggs } p_T \text{ 3lF}$
$\text{binHiggs } p_T \text{ 3lS}$	$\Delta R(l_0)(l_1)$	$\Delta R(l_0)(l_2)$
$\Delta R(l_1)(l_2)$	$\text{decayScore}$	$\text{higgsRecoScore3lF}$
$\text{higgsRecoScore3lS}$	$\text{jet } \eta \text{ 0}$	$\text{jet } \eta \text{ 1}$
$\text{jet } \phi \text{ 0}$	$\text{jet } \phi \text{ 1}$	$\text{jet } p_T \text{ 0}$
$\text{jet } p_T \text{ 1}$	$\text{Lepton } \eta \text{ 0}$	$\text{Lepton } \eta \text{ 1}$
$\text{Lepton } \eta \text{ 2}$	$\text{Lepton } \phi \text{ 0}$	$\text{Lepton } \phi \text{ 1}$
$\text{Lepton } \phi \text{ 2}$	$\text{Lepton } p_T \text{ 0}$	$\text{Lepton } p_T \text{ 1}$
$\text{Lepton } p_T \text{ 2}$	$E_T^{\text{miss}}$	$\min \Delta R(l_0)(\text{jet})$
$\min \Delta R(l_1)(\text{jet})$	$\min \Delta R(l_2)(\text{jet})$	$\min \Delta R(\text{Lepton})(\text{bjet})$
$mjj\text{Max frwdJet}$	$n\text{Jets}$	$n\text{Jets OR DL1r 60}$
$n\text{Jets OR DL1r 70}$	$n\text{Jets OR DL1r 85}$	$\text{topScore}$

Table 19: Input features used to distinguish signal and background events in the 3l channel.

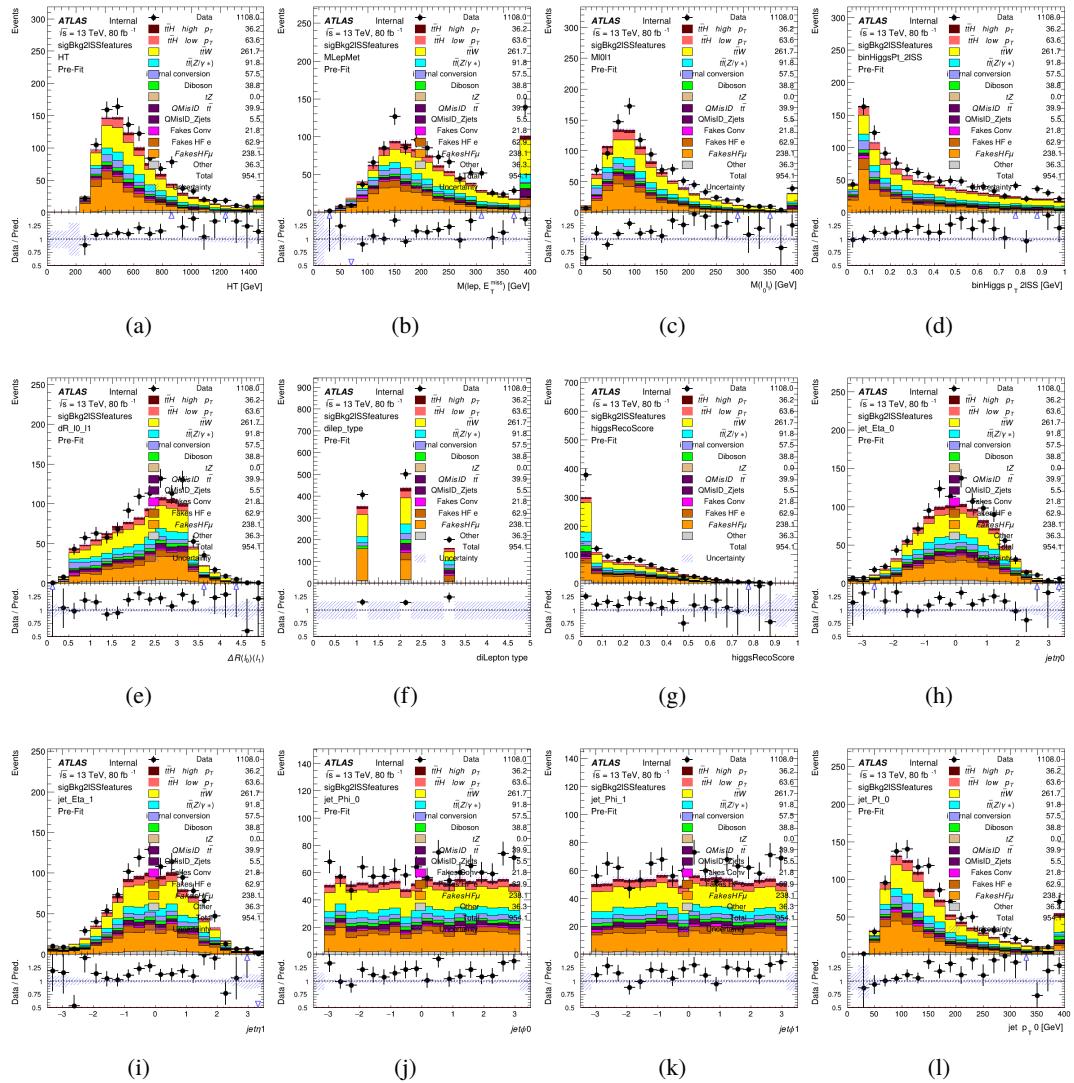


Figure 10.3:

760 The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss  
 761 function.

762 Output distributions of each MVA comparing MC prediction to data at  $80 \text{ fb}^{-1}$  are shown  
 763 in figure 10.2.

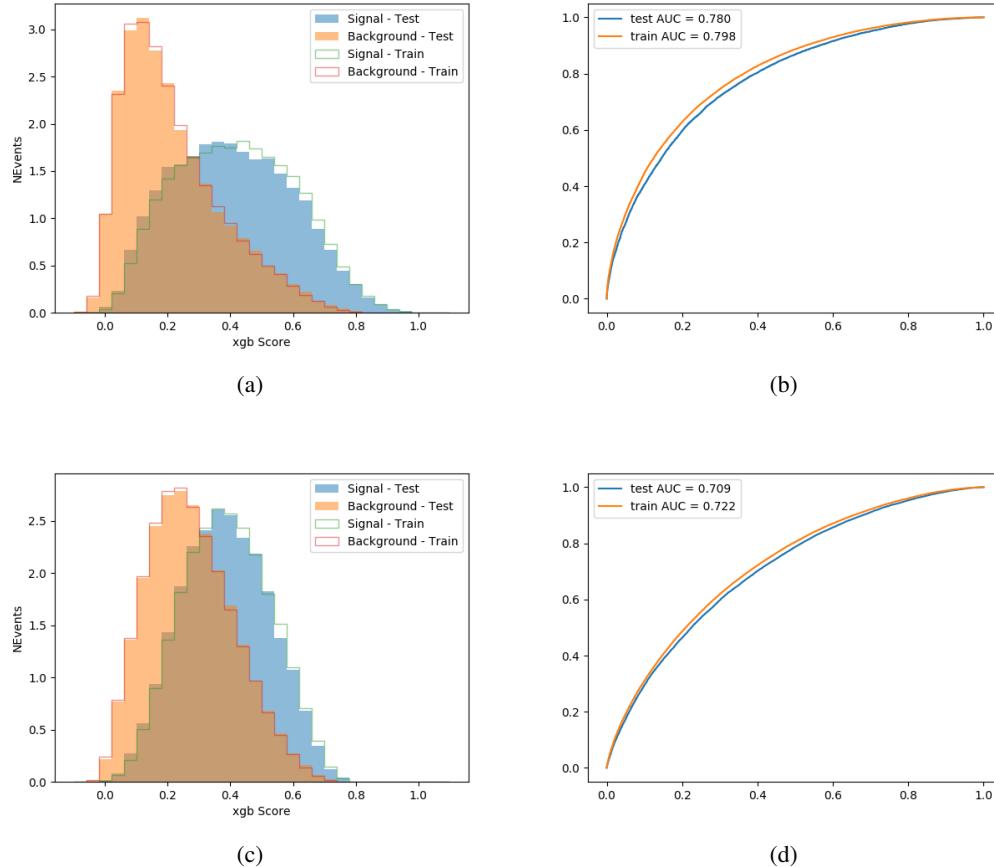


Figure 10.4:

764    **10.3 Signal Region Definitions**

765    Once pre-selection has been applied, channels are further refined based on the MVAs described  
 766    above. The output of the model described in section 8.5 is used to separate the three channel  
 767    into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs  
 768    boson.

769    For each event, depending on the channel as well as the predicted  $p_T$  of the Higgs derived

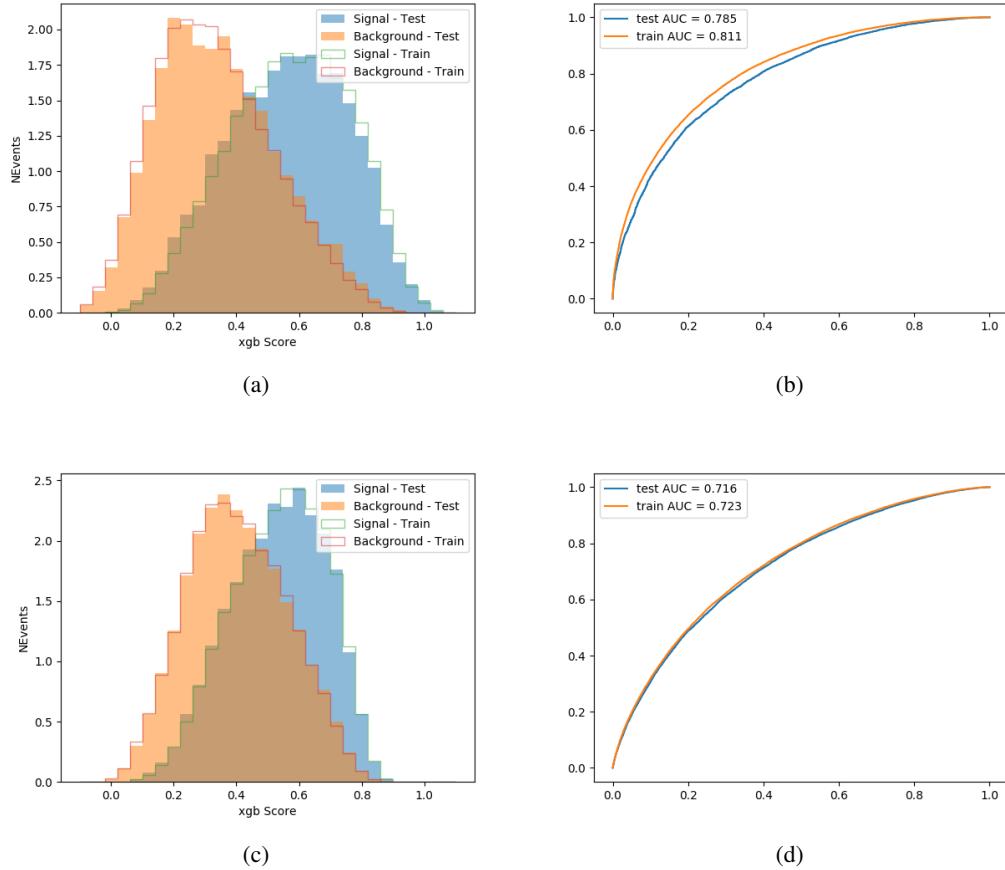


Figure 10.5:

from the algorithm described in section 8.4, a cut on the appropriate background rejection algorithm is applied. The specific selection used, and the event yield in each channel after this selection has been applied, is summarized below.

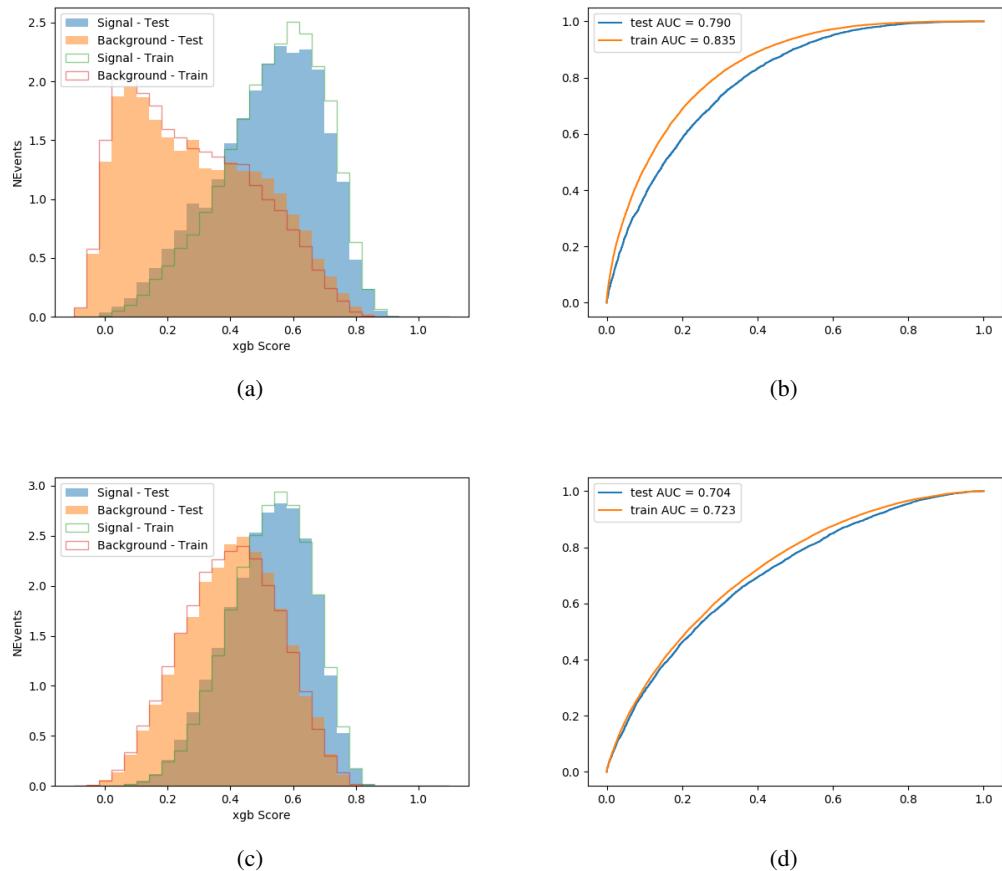


Figure 10.6:

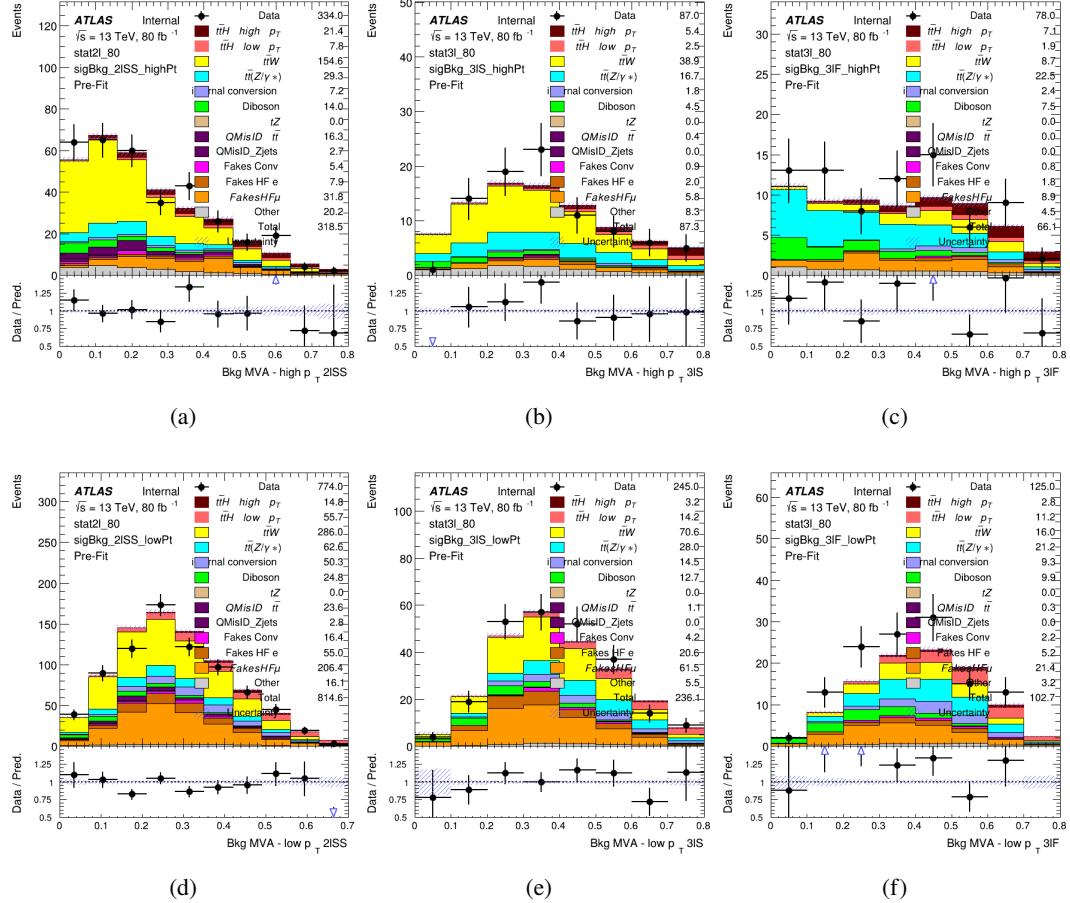


Figure 10.7: scores

773 **10.3.1 2ISS**774 **10.3.2 3l – Semi – leptonic**775 **10.3.3 3l – Fully – leptonic**776 **11 Systematic Uncertainties**

777 The systematic uncertainties that are considered are summarized in table 20. These are implemented in the fit either as a normalization factors or as a shape variation or both in the signal  
 778 19th November 2020 – 12:11

<sup>779</sup> and background estimations. The numerical impact of each of these uncertainties is outlined in  
<sup>780</sup> section 12.

Table 20: Sources of systematic uncertainty considered in the analysis. Some of the systematic uncertainties are split into several components, as indicated by the number in the rightmost column.

Systematic uncertainty	Components
Luminosity	1
Pileup reweighting	1
<b>Physics Objects</b>	
Electron	6
Muon	15
Jet energy scale and resolution	28
Jet vertex fraction	1
Jet flavor tagging	131
$E_T^{\text{miss}}$	3
Total (Experimental)	186
<b>Background Modeling</b>	
Cross section	24
Renormalization and factorization scales	10
Parton shower and hadronization model	2
Shower tune	4
Total (Signal and background modeling)	40
<b>Background Modeling</b>	
Cross section	24
Renormalization and factorization scales	10
Parton shower and hadronization model	2
Shower tune	4
Total (Signal and background modeling)	40
Total (Overall)	226

<sup>781</sup> The uncertainty in the combined 2015+2016 integrated luminosity is derived from a  
<sup>782</sup> calibration of the luminosity scale using x-y beam-separation scans performed in August 2015  
<sup>783</sup> and May 2016 [lumi].

<sup>784</sup> The experimental uncertainties are related to the reconstruction and identification of light  
<sup>785</sup> leptons and b-tagging of jets, and to the reconstruction of  $E_T^{\text{miss}}$ . The sources which contribute

<sup>786</sup> to the uncertainty in the jet energy scale [**jes**] are decomposed into uncorrelated components and  
<sup>787</sup> treated as independent sources in the analysis.

<sup>788</sup> The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses  
<sup>789</sup> [**btag\_cal**] are also decomposed into uncorrelated components. The large number of components  
<sup>790</sup> for b-tagging is due to the calibration of the distribution of the BDT discriminant.

<sup>791</sup> The systematic uncertainties associated with the signal and background processes are  
<sup>792</sup> accounted for by varying the cross-section of each process within its uncertainty.

## <sup>793</sup> 12 Results

<sup>794</sup> Unblinded results are shown for the  $80 \text{ fb}^{-1}$  data set, as well as MC only projections of results  
<sup>795</sup> using the full Run-2,  $140 \text{ fb}^{-1}$  dataset.

### <sup>796</sup> 12.1 Results - $80 \text{ fb}^{-1}$

<sup>797</sup> A maximum likelihood fit is performed simultaneously over the reconstructed Higgs  $p_T$  spectrum  
<sup>798</sup> in the three signal regions, 2lSS, 3lS, and 3lF, shown in figure 12.1. The  $t\bar{t}H$  MC is split into  
<sup>799</sup> high and low  $p_T$ , based on whether the truth  $p_T$  of the Higgs is above or below 150 GeV. The  
<sup>800</sup> parameters  $\mu_{t\bar{t}H\text{high}p_T}$  and  $\mu_{t\bar{t}H\text{low}p_T}$ , where  $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$ , are extracted from the fit.

801 As the data collected from 2015-2017 has been unblinded for  $t\bar{t}H$  – ML channels, rep-  
 802 resenting  $80 \text{ fb}^{-1}$ , those events are unblinded.

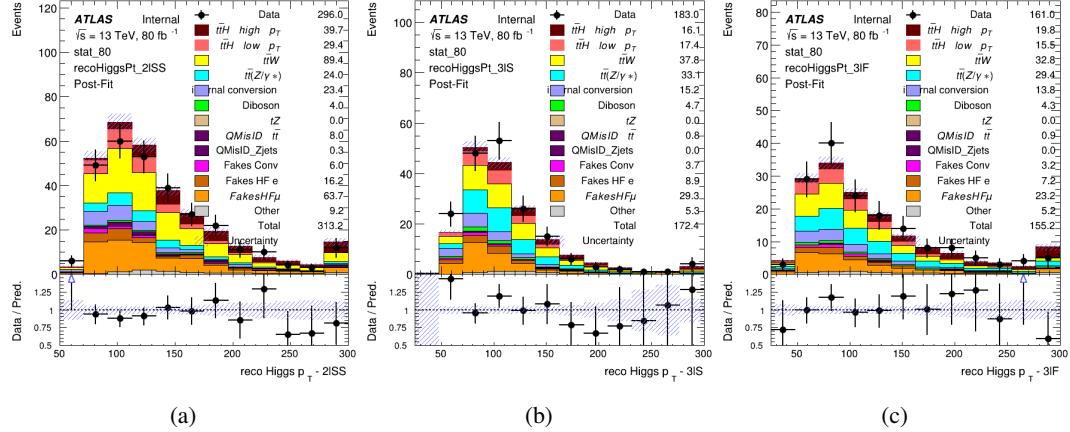


Figure 12.1: Post-fit distributions of the reconstructed Higgs  $p_T$  in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for  $80 \text{ fb}^{-1}$  of MC

803 A post-fit summary of the fitted regions is shown in figure 12.2.

The  $\mu$  values for high and low  $p_T$  Higgs are shown in 21.

$$\begin{aligned}\mu_{t\bar{t}H \text{ high } p_T} &= 1.0^{+0.0}_{-0.0} (\text{stat})^{+0.0}_{-0.0} (\text{sys}) \\ \mu_{t\bar{t}H \text{ low } p_T} &= 1.0^{+0.0}_{-0.0} (\text{stat})^{+0.0}_{-0.0} (\text{sys})\end{aligned}$$

Table 21: Best fit  $\mu$  values for  $t\bar{t}H$  high  $p_T$  and  $t\bar{t}H$  low  $p_T$ , where  $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

804 **Need to add something about systematics here**

806 The background composition of each of the fit regions is shown in figure 12.3.

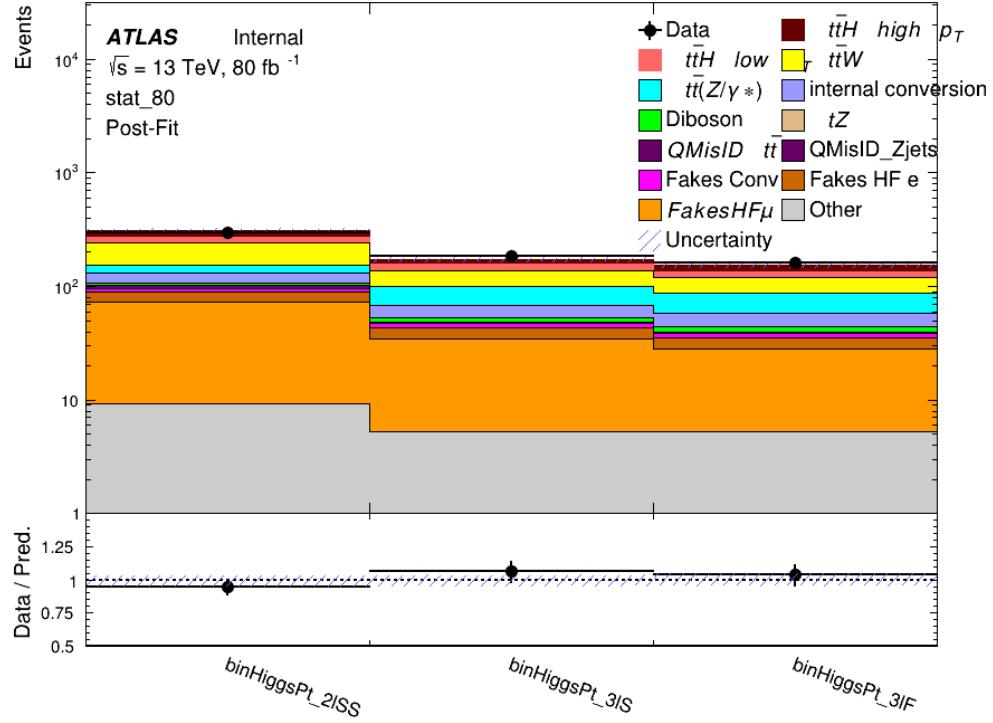


Figure 12.2: Post-fit summary of the yields in each signal region.

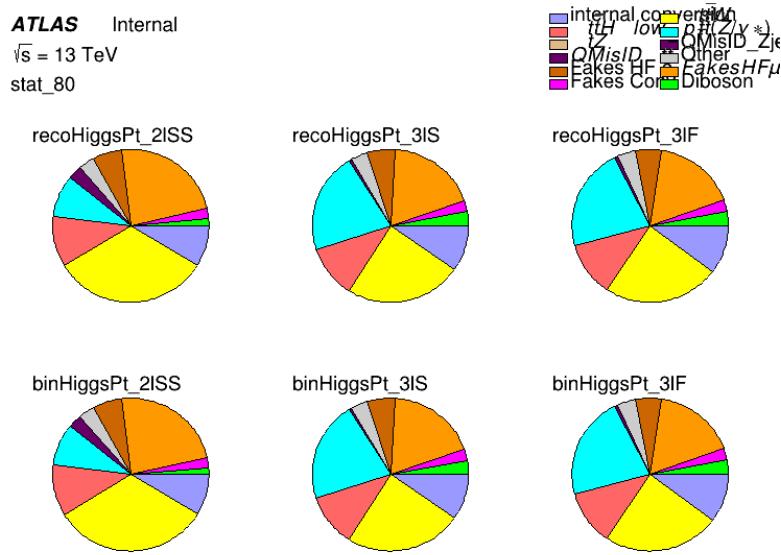


Figure 12.3: Background composition of the fit regions.

## 807 12.2 Projected Results - $140 \text{ fb}^{-1}$

808 As data collected in 2018 has not yet been unblinded for  $t\bar{t}H - \text{ML}$  at the time of this note, data  
 809 from that year remains blinded. Instead, an Asimov fit is performed - with the MC prediction  
 810 being used both as the SM prediction as well as the data in the fit - in order to give expected  
 811 results.

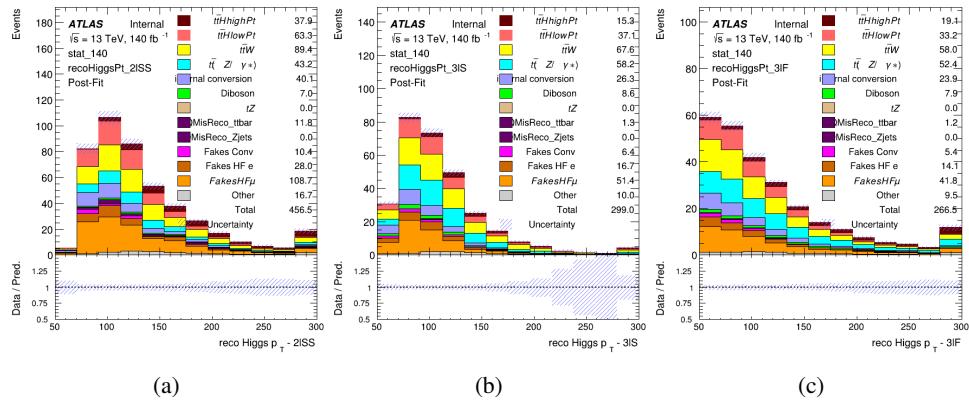


Figure 12.4: Blinded post-fit distributions of the reconstructed Higgs  $p_T$  in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for  $140 \text{ fb}^{-1}$  of data

812 The  $\mu$  values for high and low  $p_T$  Higgs are shown in 22.

$$\begin{aligned}\mu_{t\bar{t}H_{\text{high}p_T}} &= 1.0^{+0.0}_{-0.0} (\text{stat})^{+0.0}_{-0.0} (\text{sys}) \\ \mu_{t\bar{t}H_{\text{low}p_T}} &= 1.0^{+0.0}_{-0.0} (\text{stat})^{+0.0}_{-0.0} (\text{sys})\end{aligned}$$

Table 22: Best fit  $\mu$  values for  $t\bar{t}H$  high  $p_T$  and  $t\bar{t}H$  low  $p_T$ , where  $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

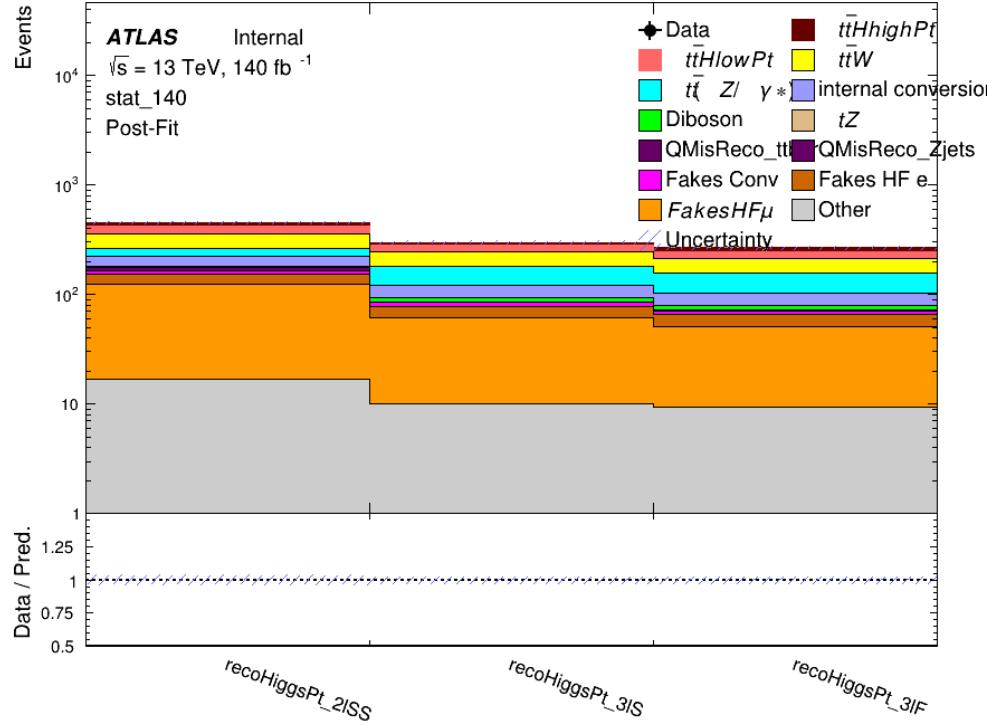


Figure 12.5: Post-fit summary of fit.

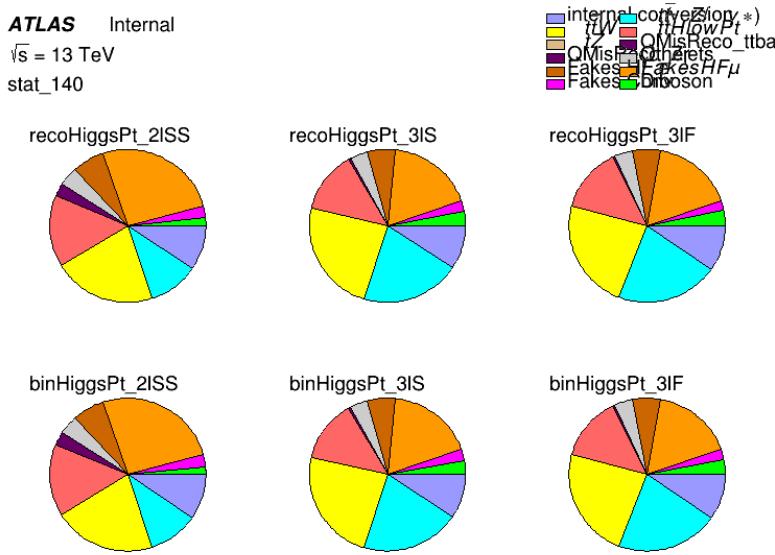


Figure 12.6: Background composition of the fit regions.

**Part V****Conclusion**

813 As search for the effects of dimension-six operators on  $t\bar{t}H$  production is performed. An effective  
814 field theory approached is used to parameterize the effects of high energy physics on the Higgs  
815 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,  
816 and the limits on dimension-six operators are limited to X.

<sup>819</sup> **List of contributions**

<sup>820</sup>

821 **Appendices**

822 **A Machine Learning Models**

823 The following section provides details of the various MVAs as well as a few studies performed  
824 in support of this analysis, exploring alternate decisions and strategies.

825 **A.1 Higgs Reconstruction Models**

826 **A.1.1 b-jet Identification Features - 2lSS**

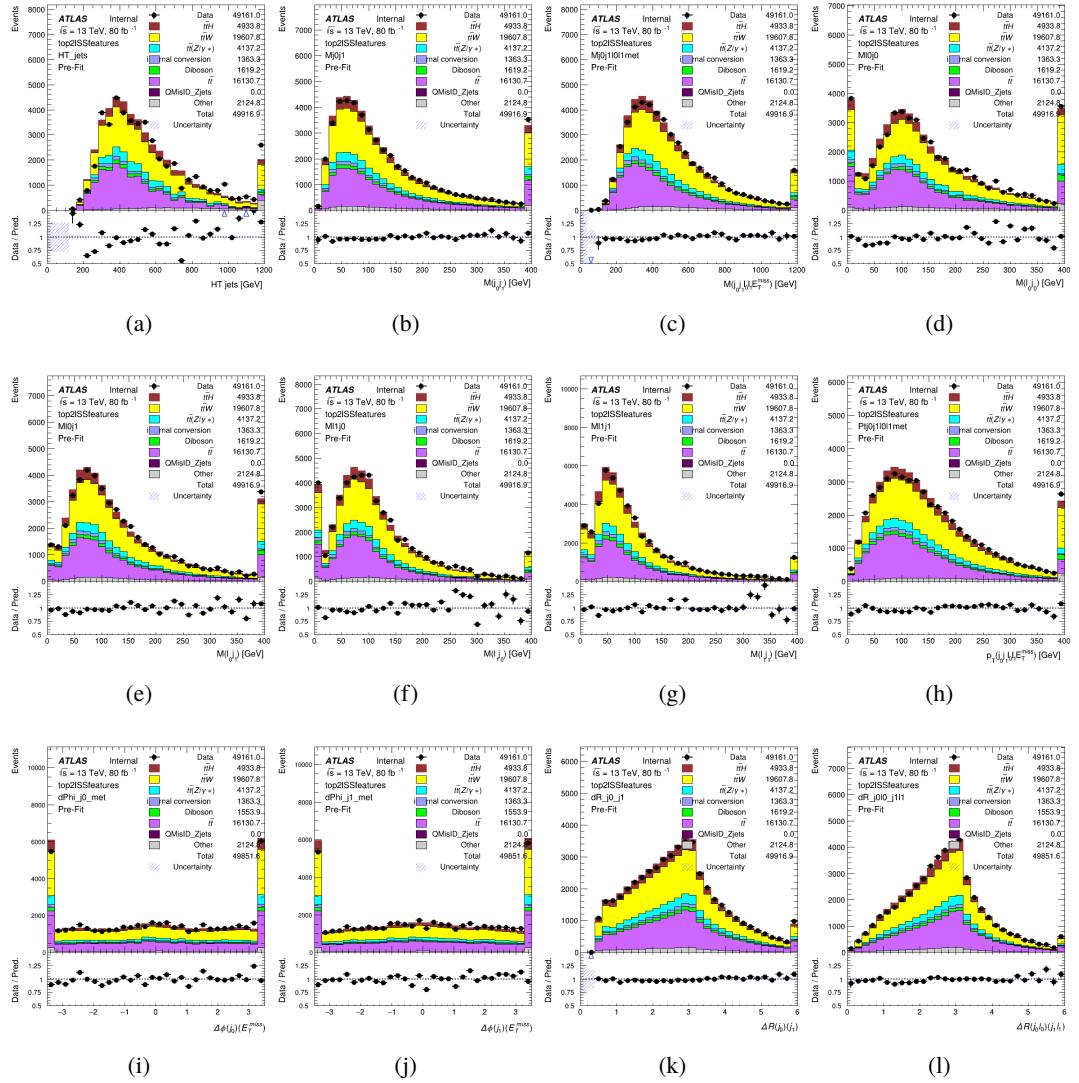


Figure A.1: Input features for top2lSS

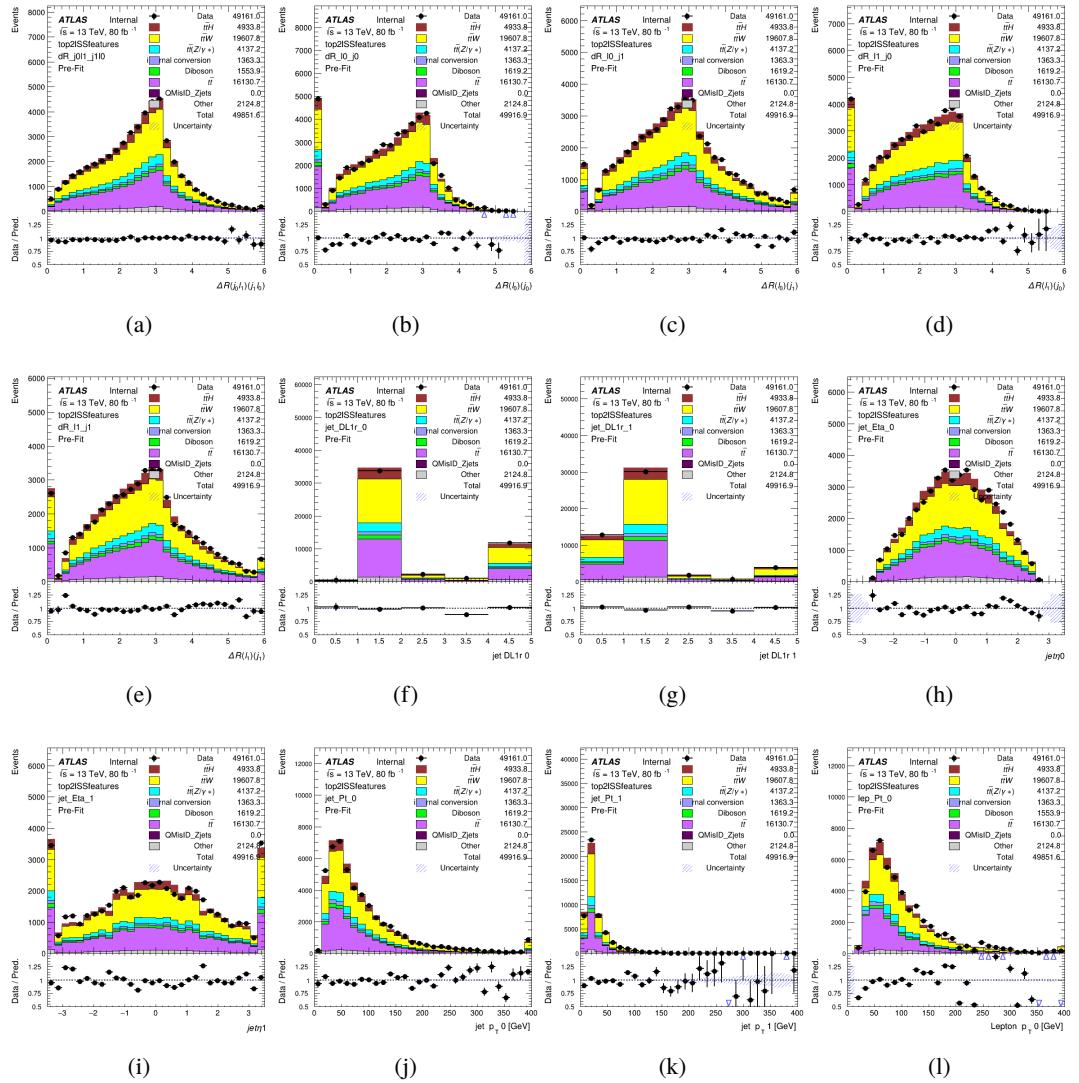


Figure A.2: Input features for top2lSS

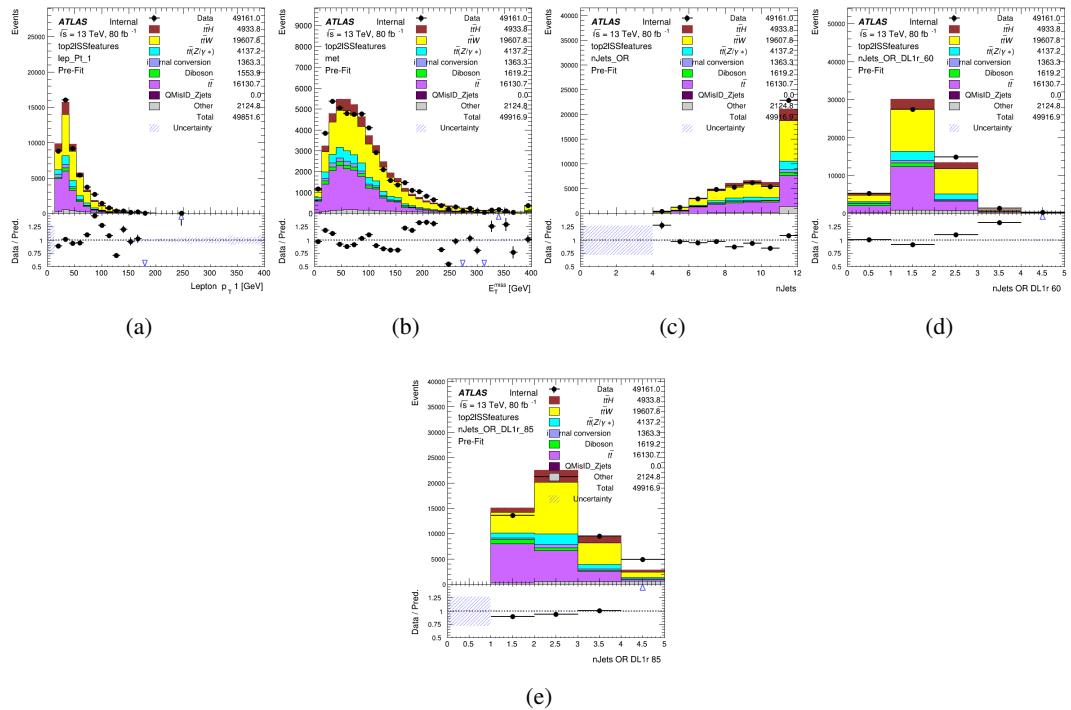


Figure A.3: Input features for top2lSS

827 **A.1.2 b-jet Identification Features - 3l**

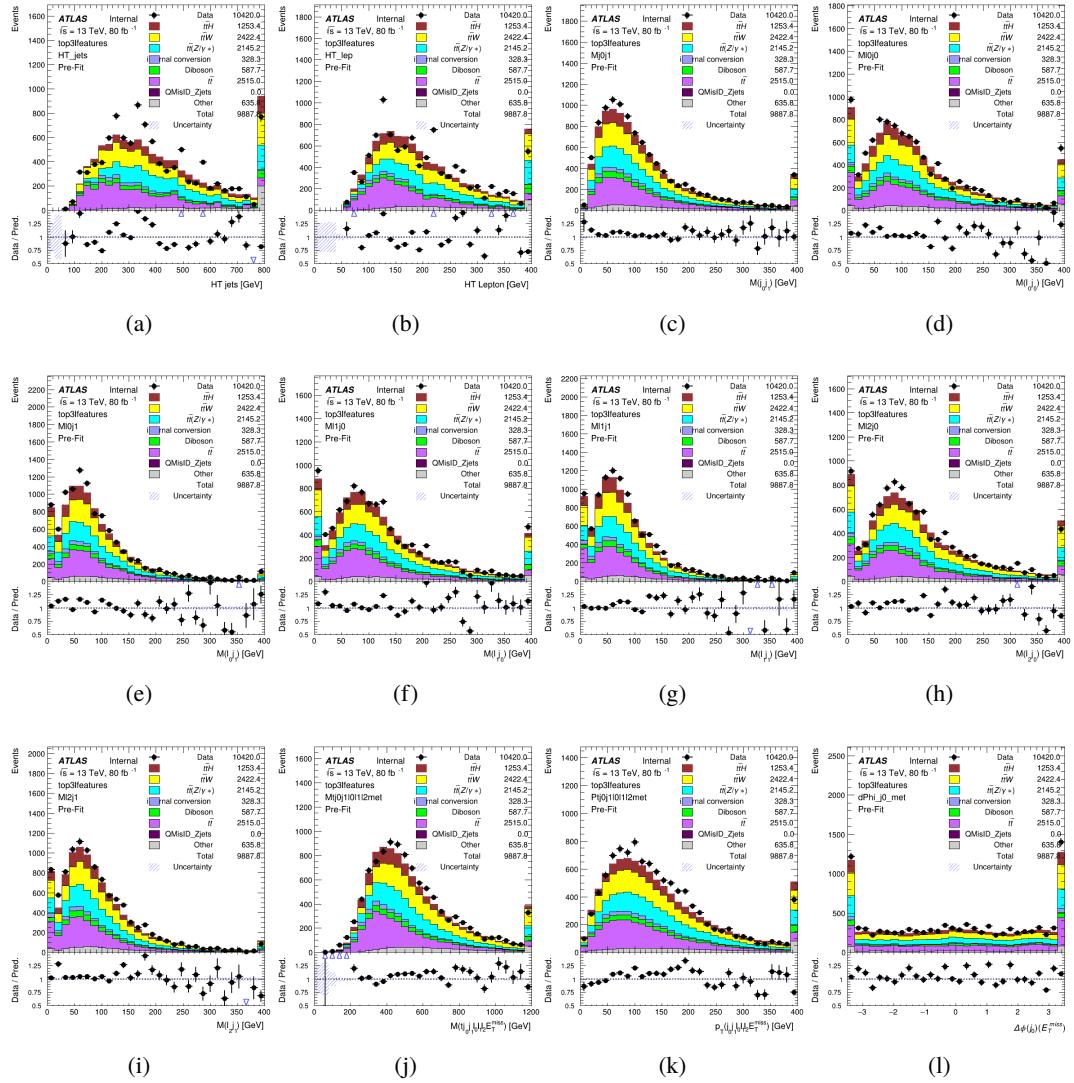


Figure A.4: Input features for top31

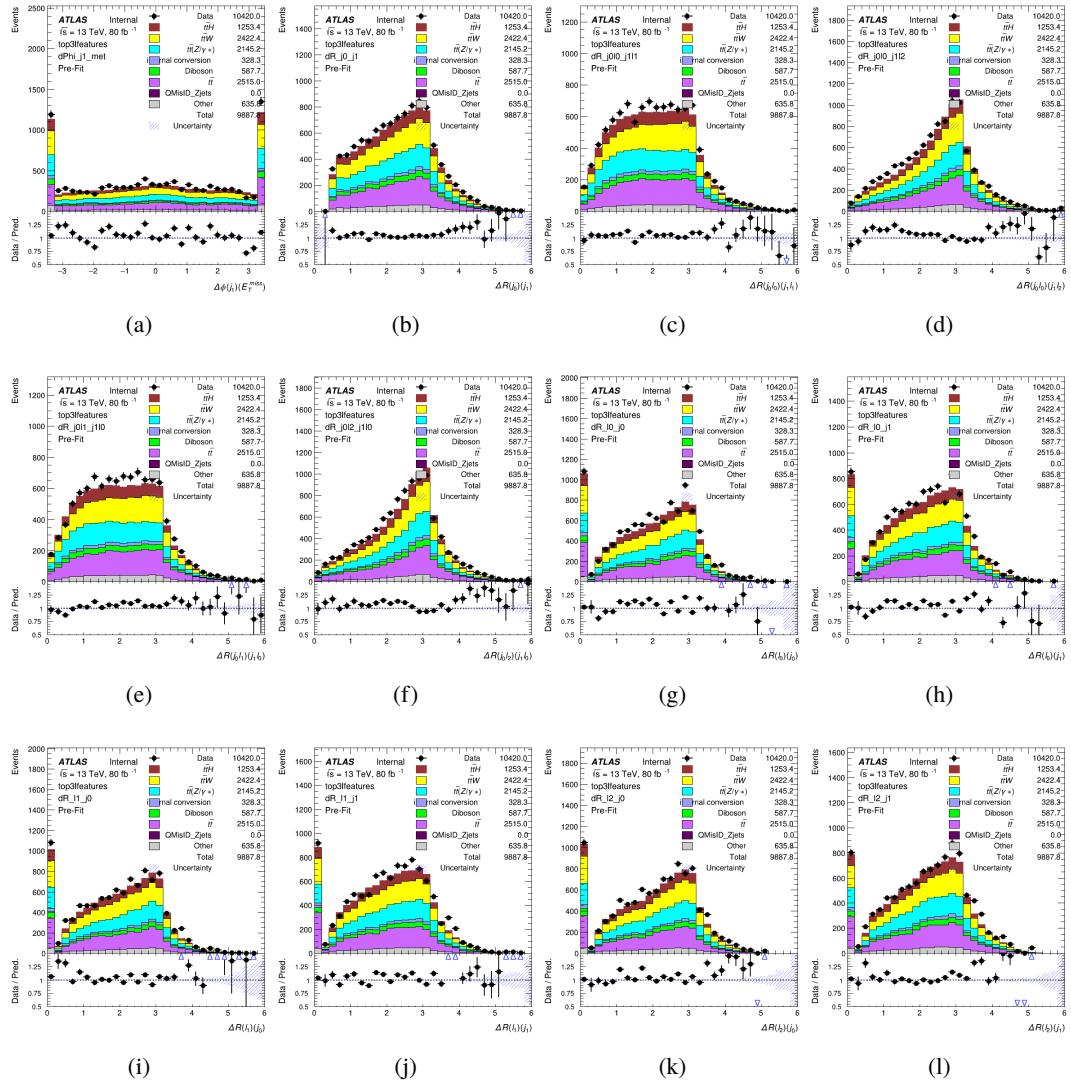


Figure A.5: Input features for top31

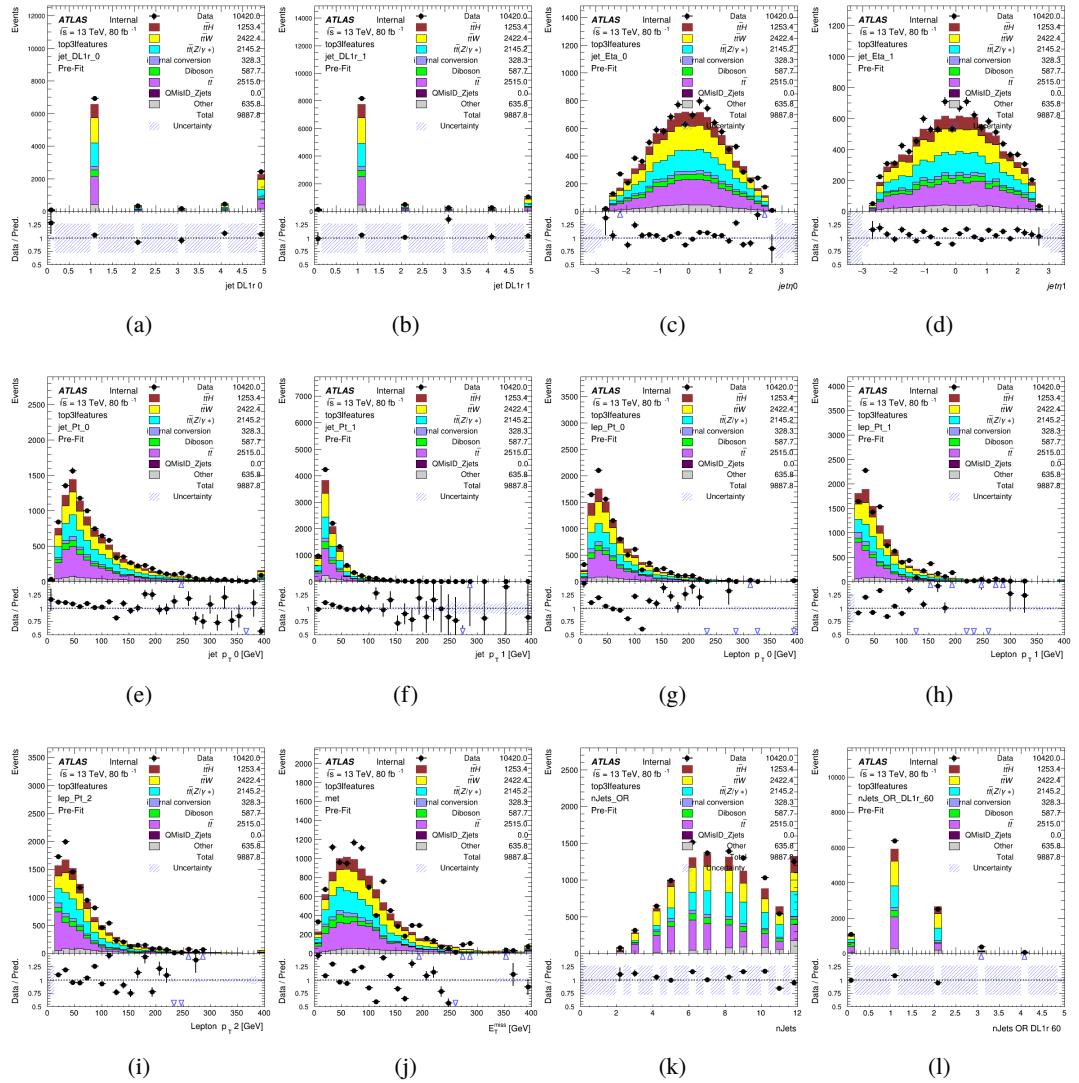
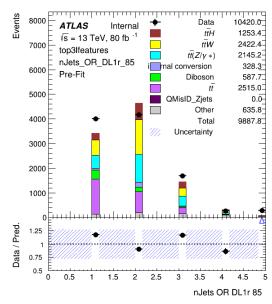


Figure A.6: Input features for top31



(a)

Figure A.7: Input features for top31

828 **A.1.3 Higgs Reconstruction Features - 2lSS**

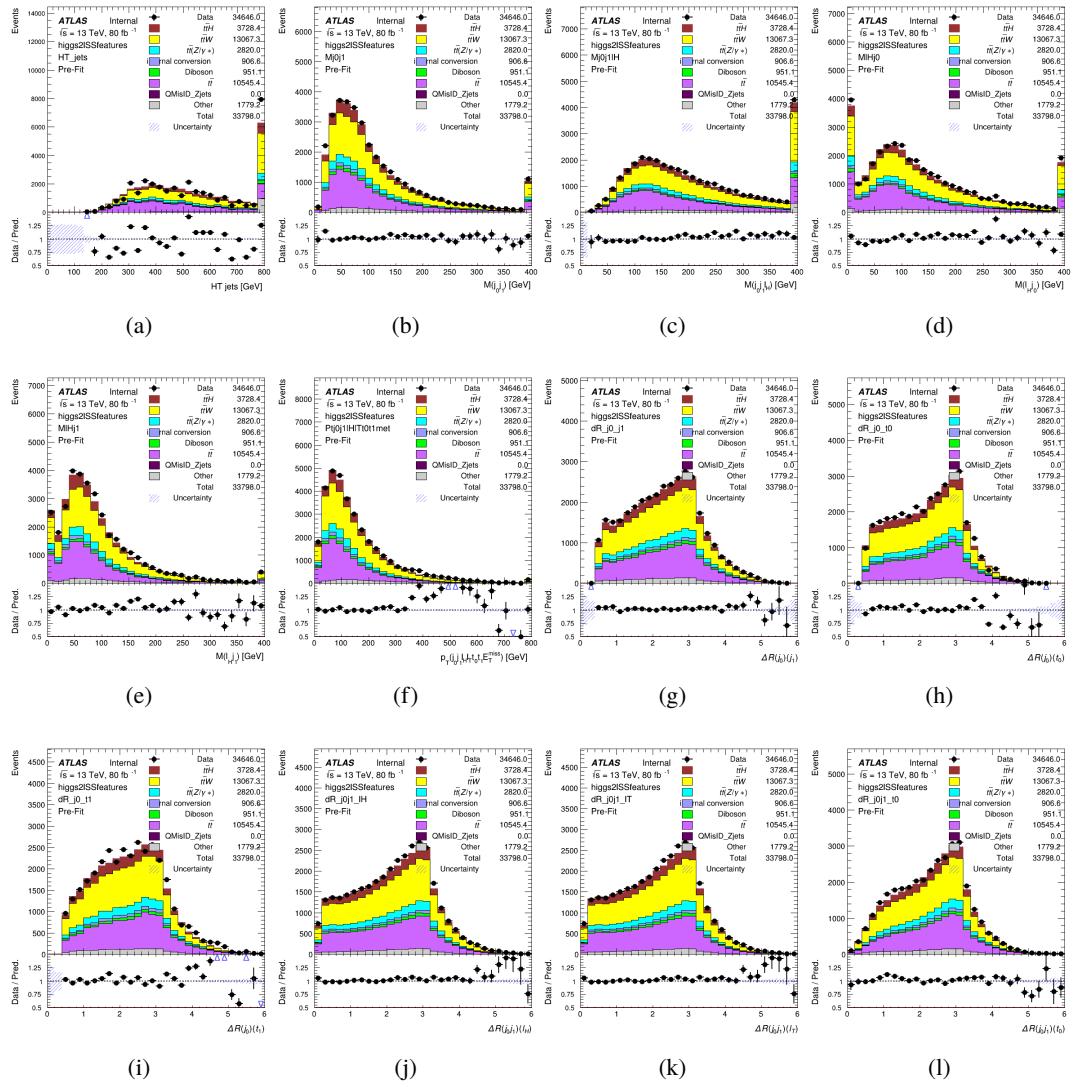


Figure A.8: Input features for higgs2lSS

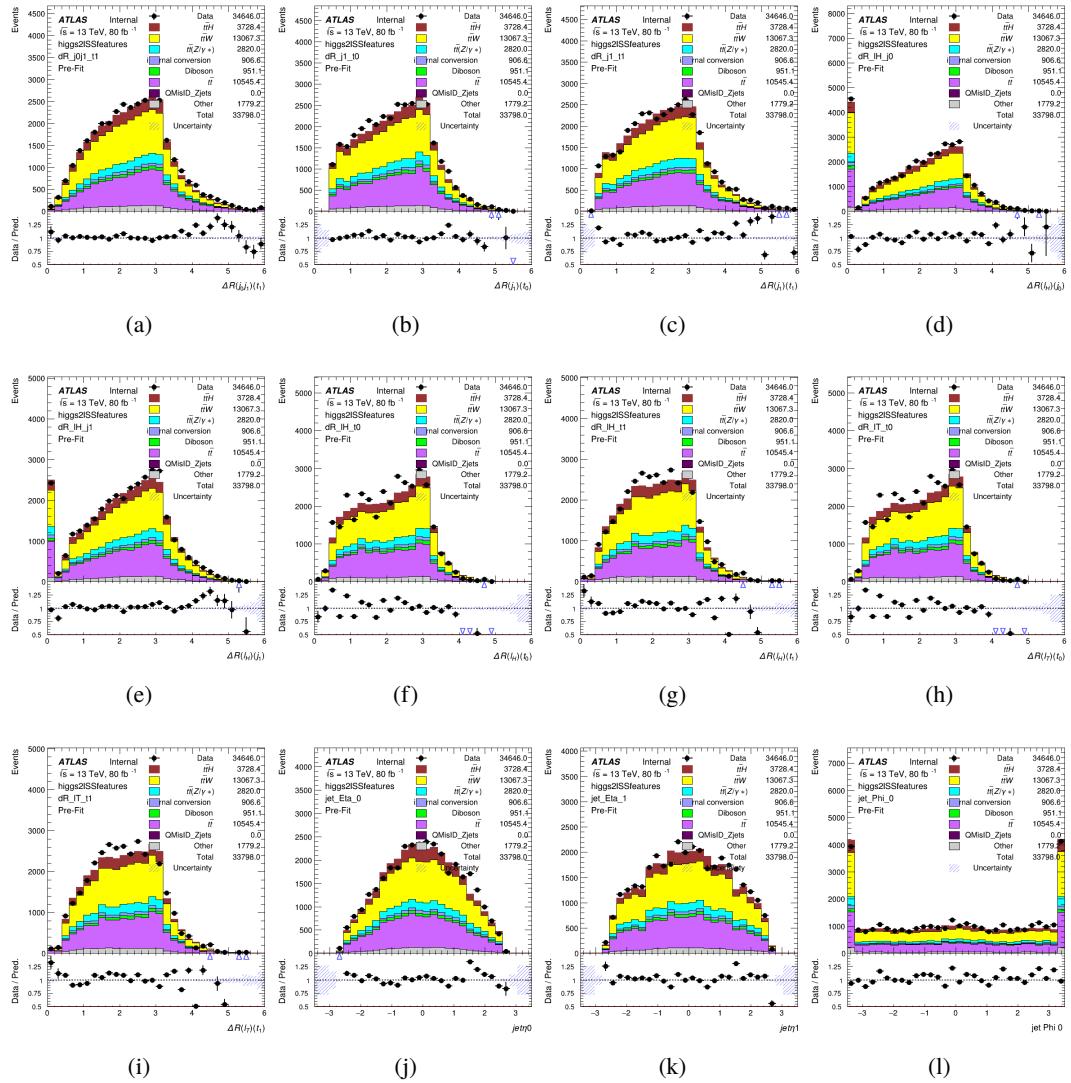


Figure A.9: Input features for higgs2lSS

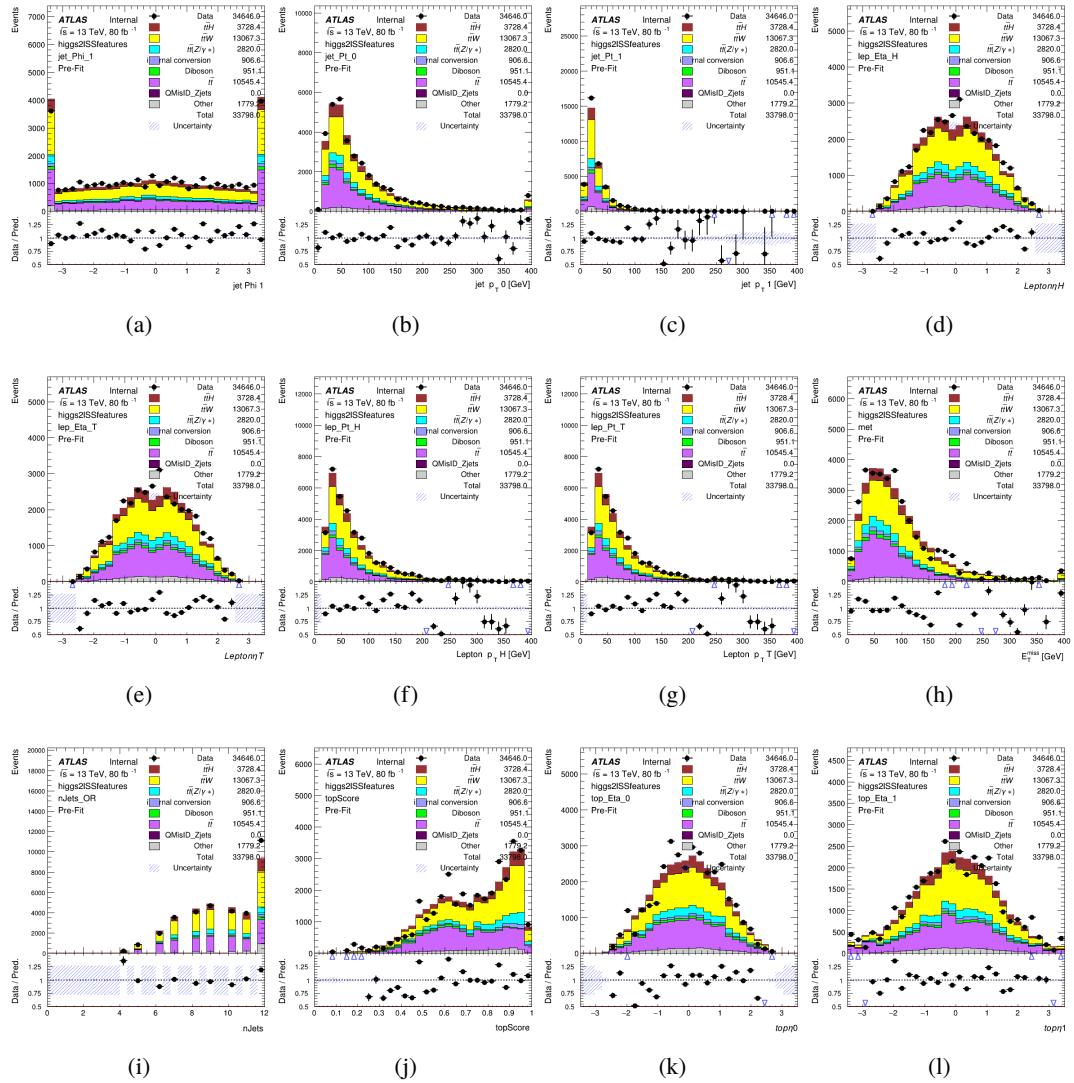


Figure A.10: Input features for higgs2ISS

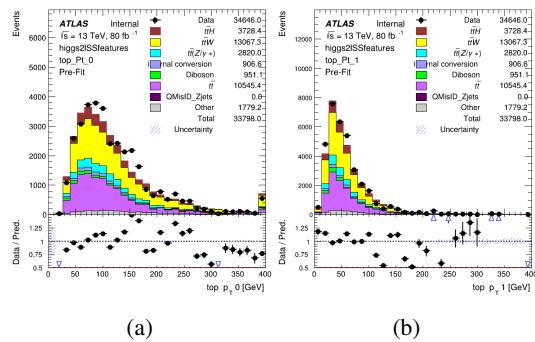


Figure A.11: Input features for higgs2ISS

829 **A.1.4 Higgs Reconstruction Features - 3lS**

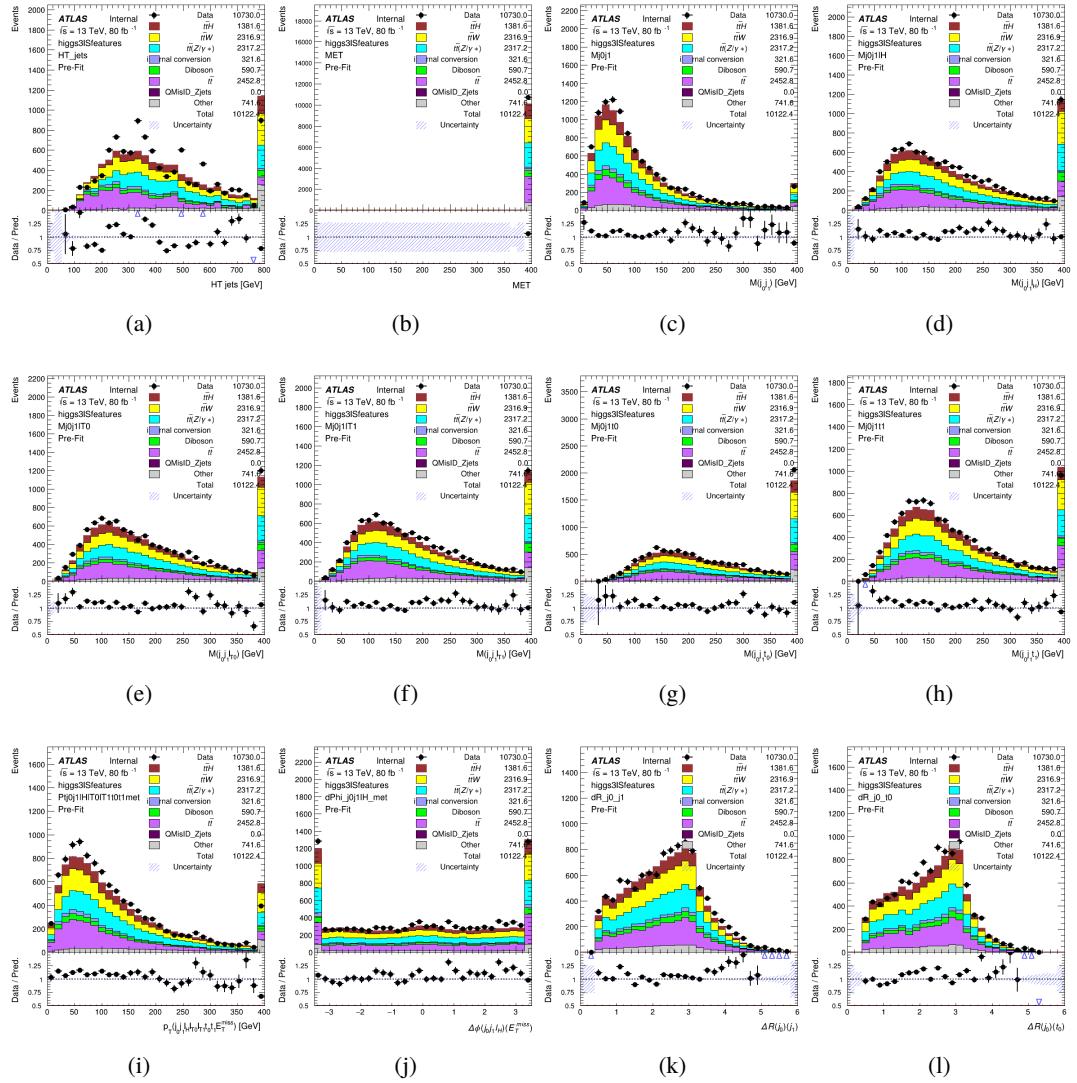


Figure A.12: Input features for higgs3IS

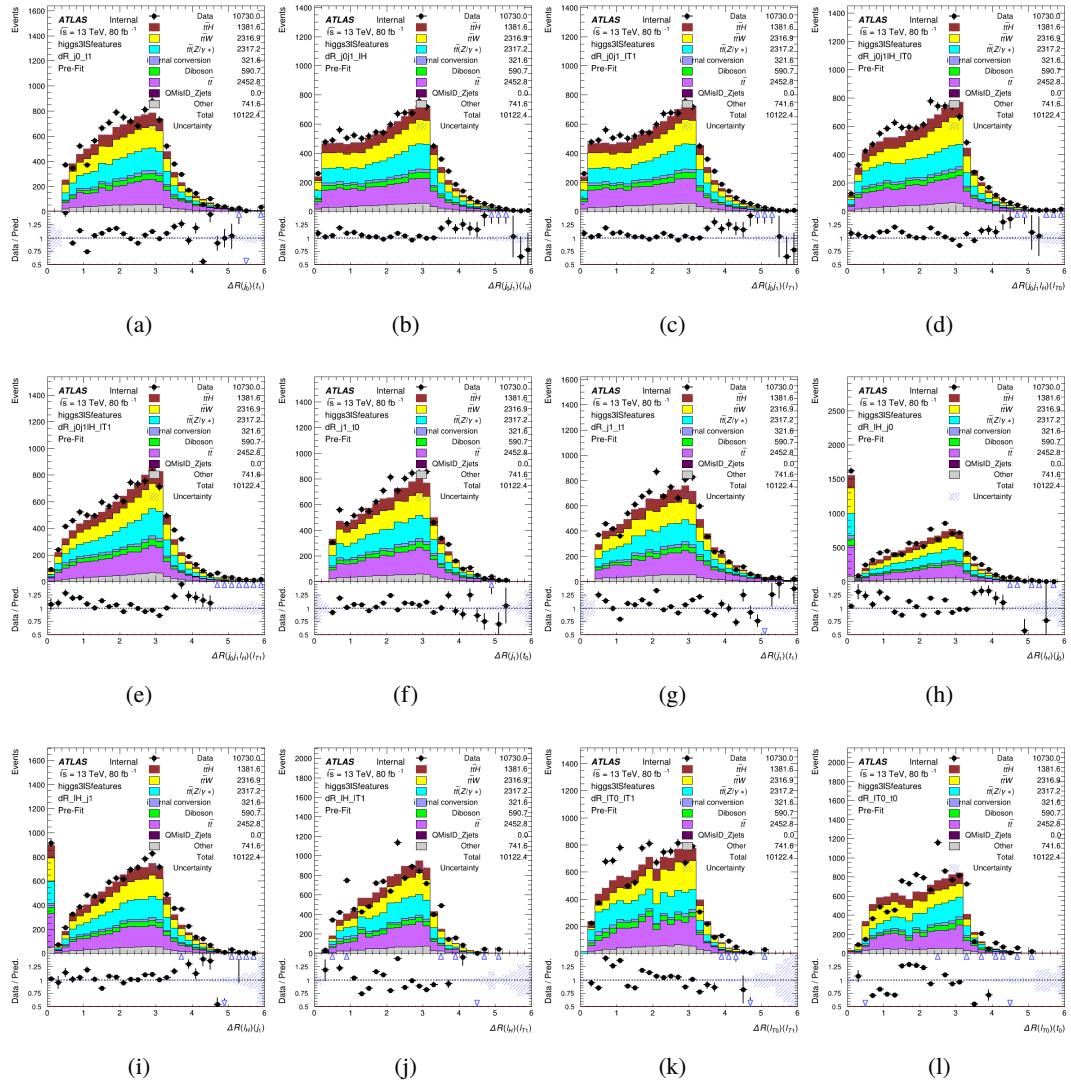


Figure A.13: Input features for higgs3lS

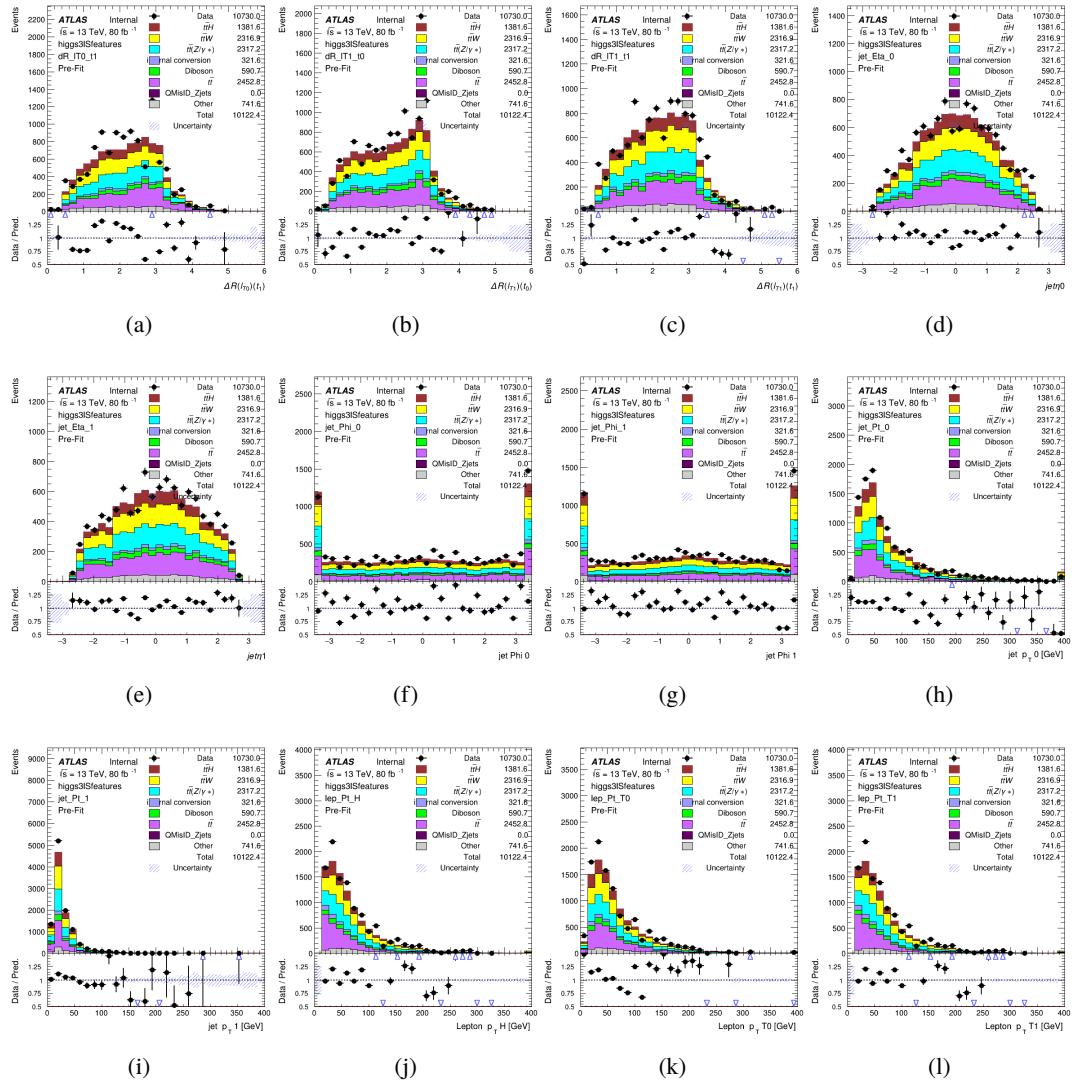


Figure A.14: Input features for higgs3lS

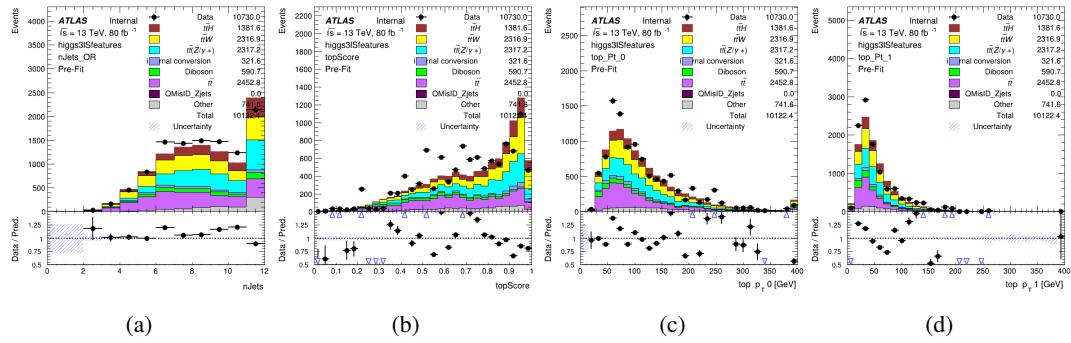


Figure A.15: Input features for higgs3lS

830 **A.1.5 Higgs Reconstruction Features - 3lF**

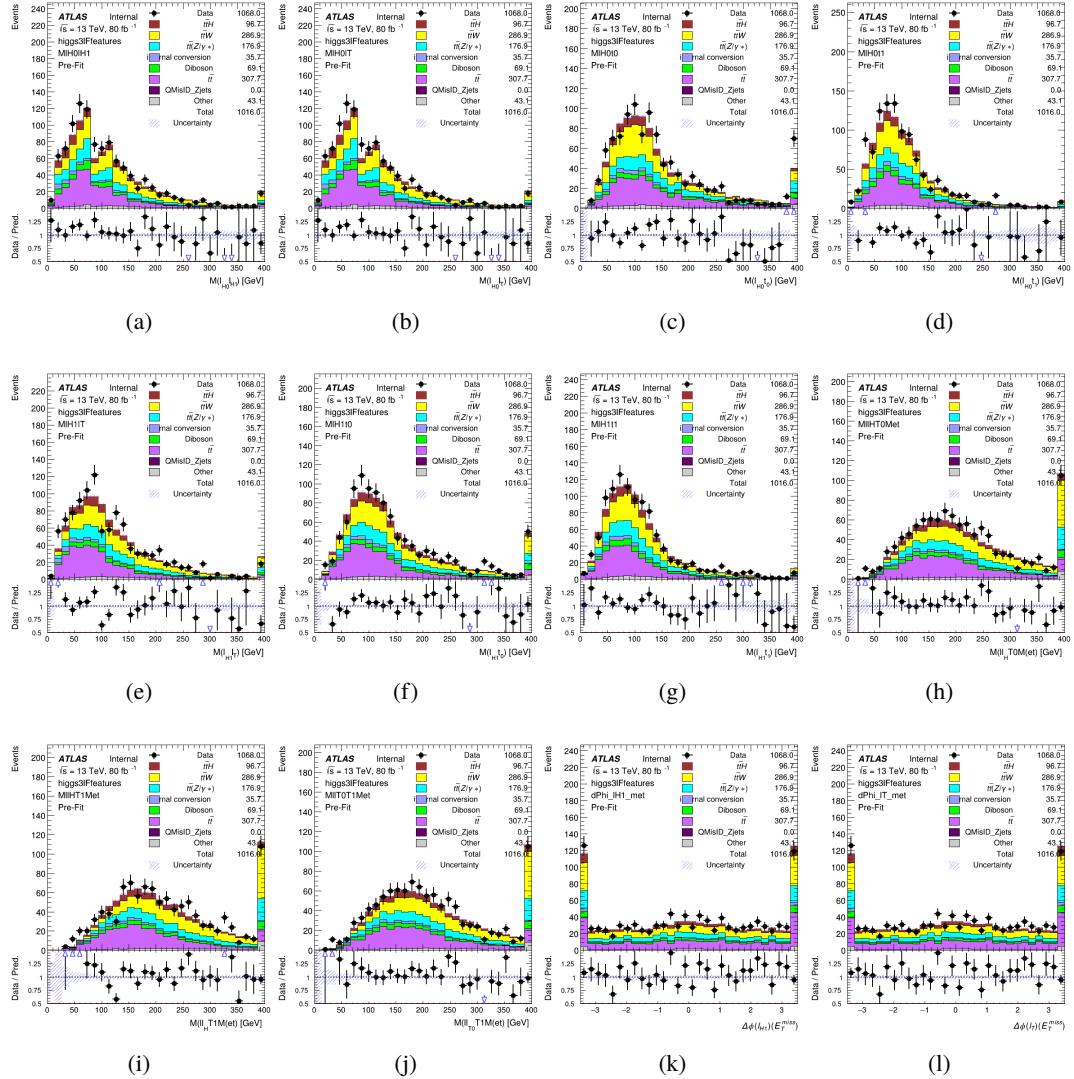


Figure A.16: Input features for higgs3IF

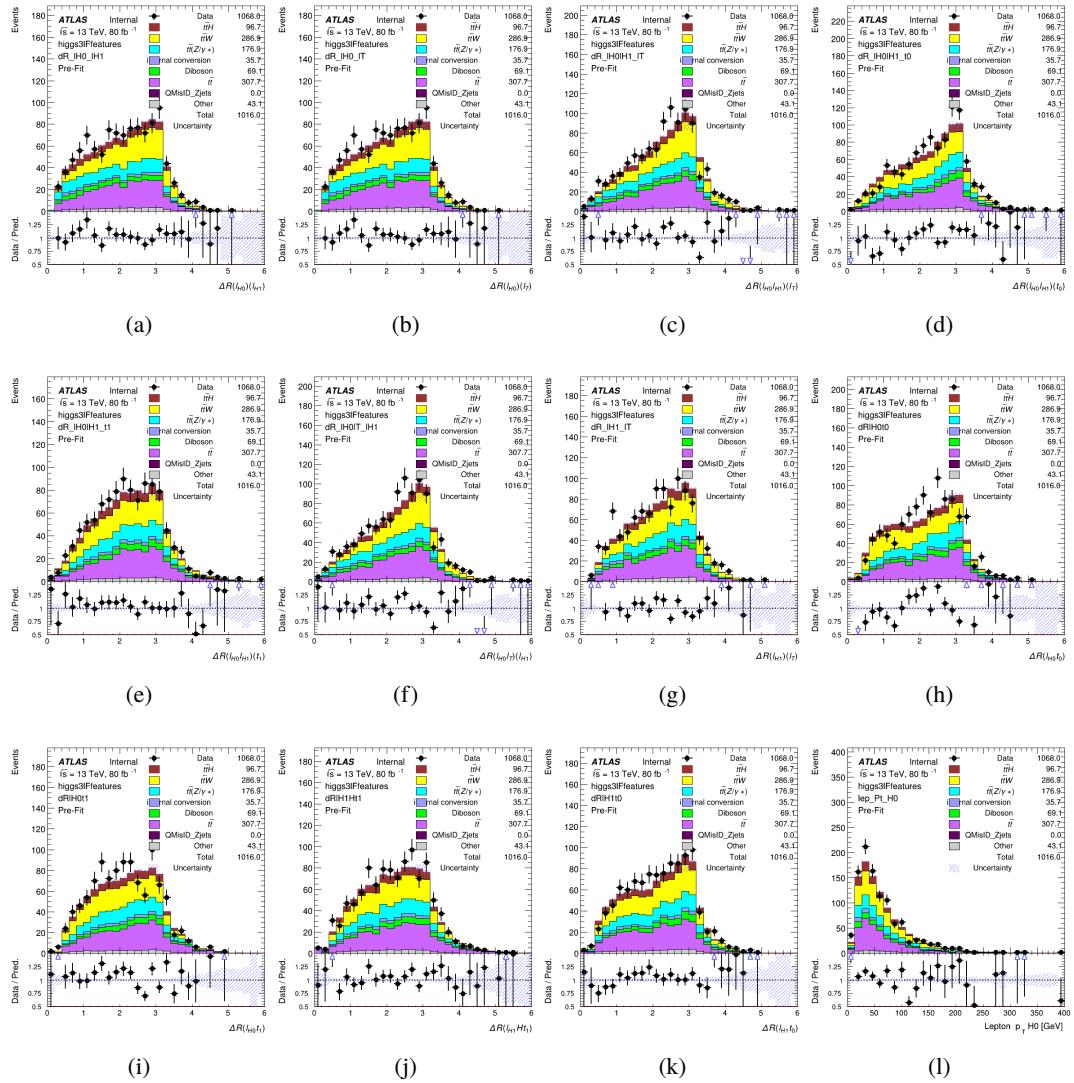


Figure A.17: Input features for higgs3lF

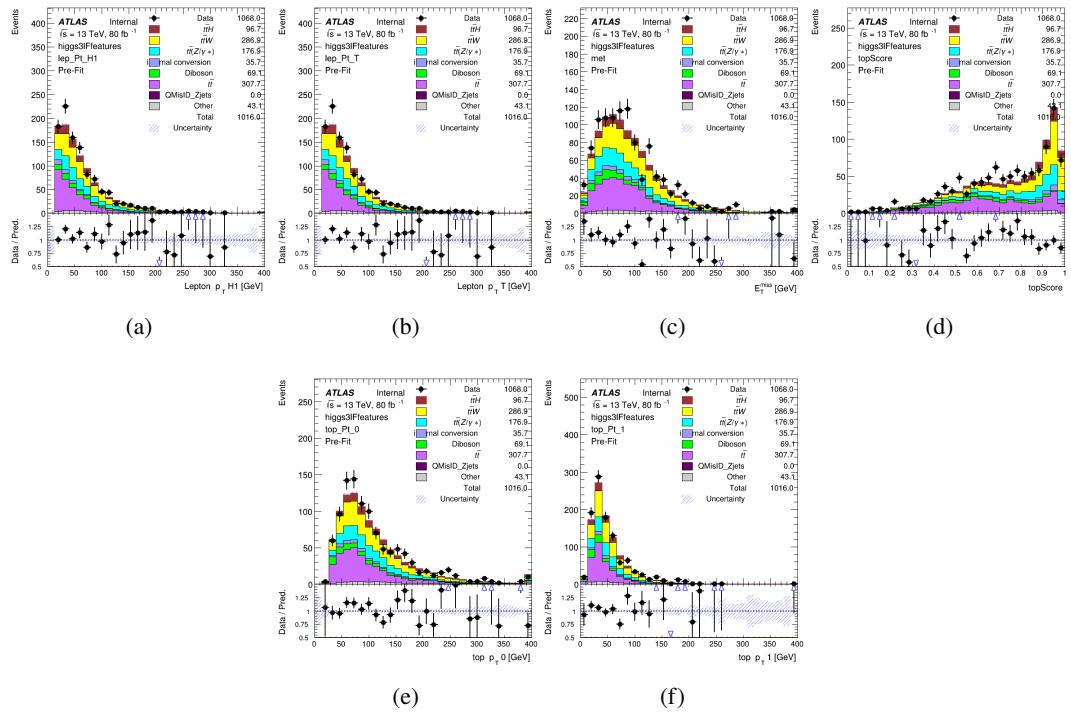
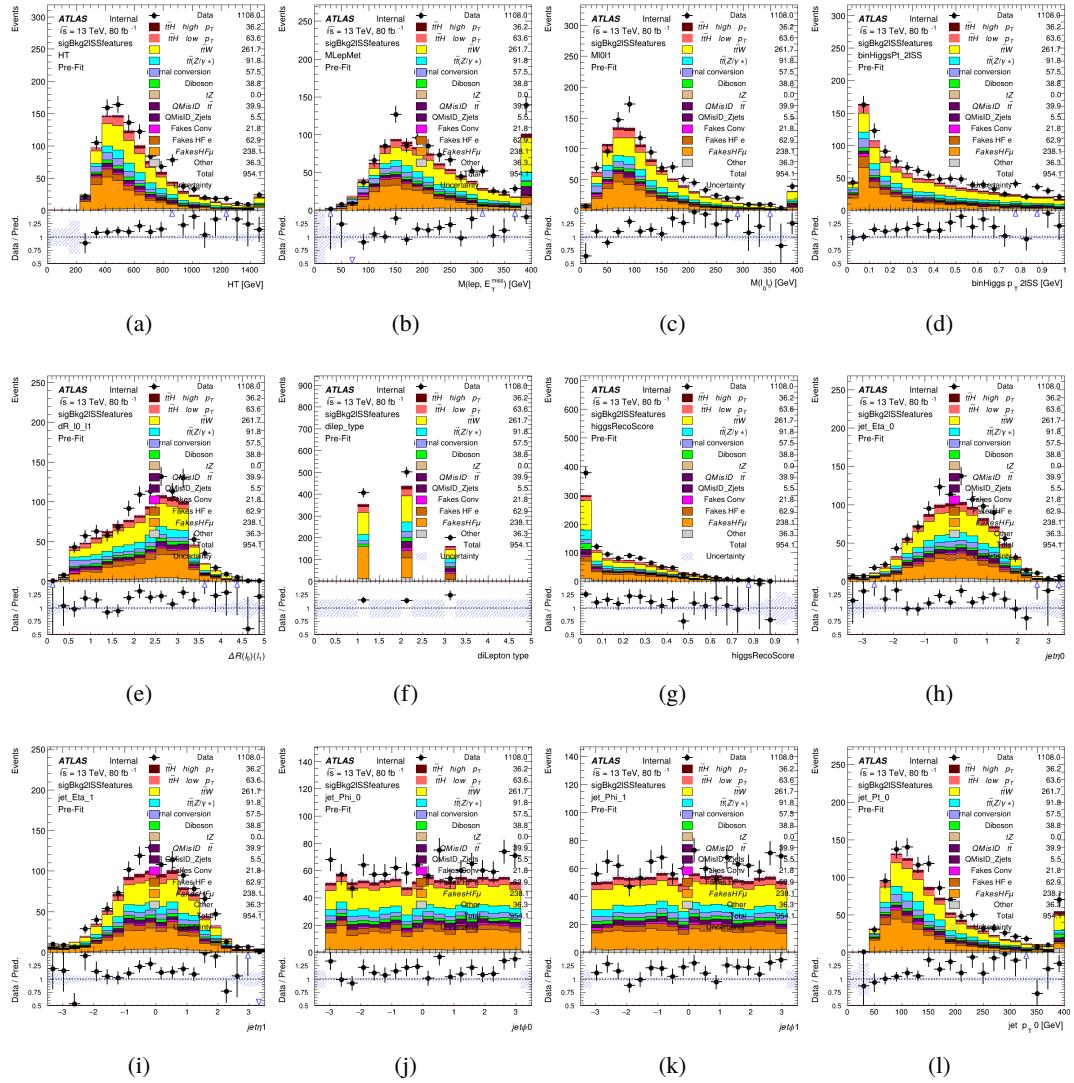


Figure A.18: Input features for higgs3IF

<sup>831</sup> **A.2 Background Rejection MVAs**

<sup>832</sup> **A.2.1 Background Rejection MVA Features - 2lSS**

Figure A.19: Input features for `sigBkg2lSS`

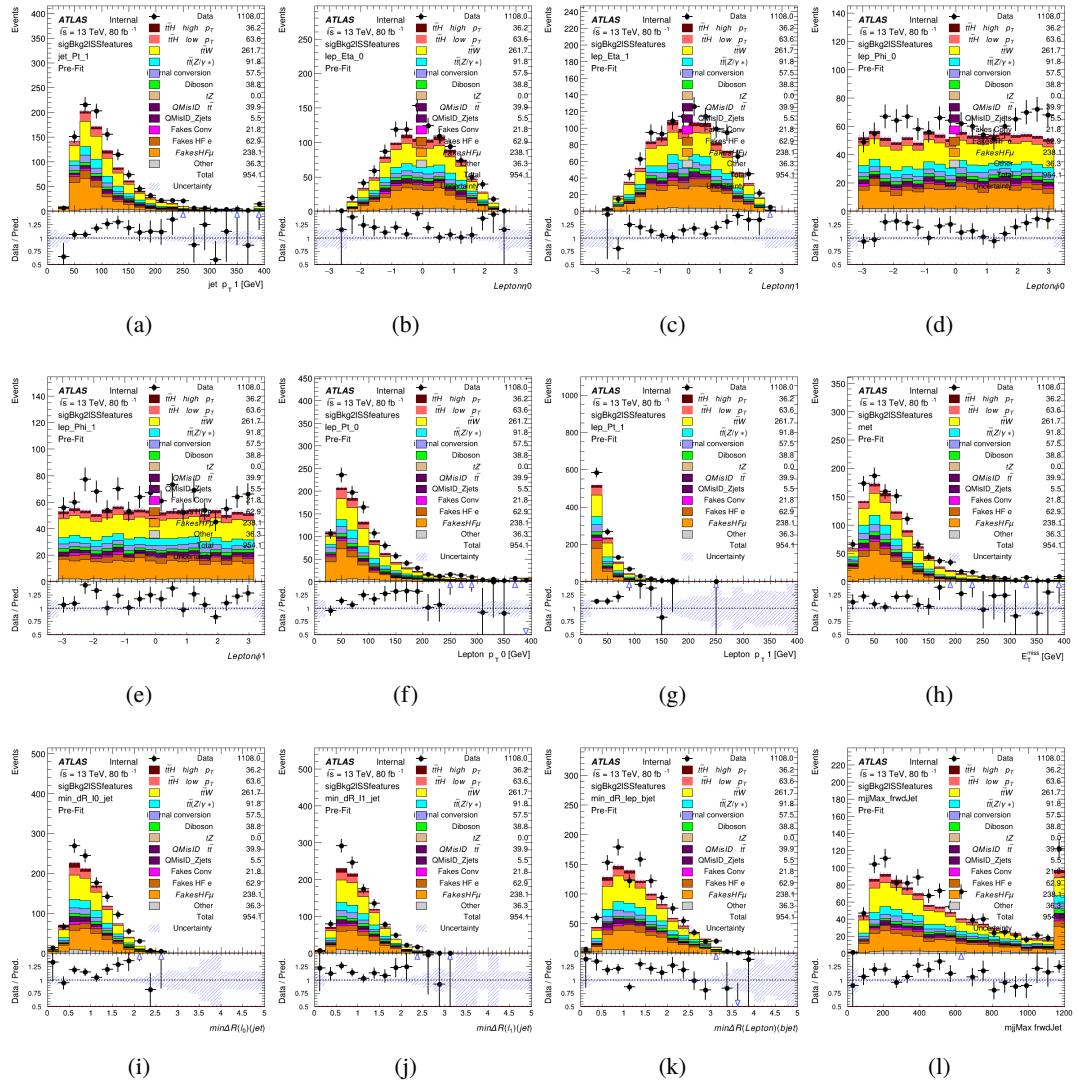


Figure A.20: Input features for sigBkg2lSS

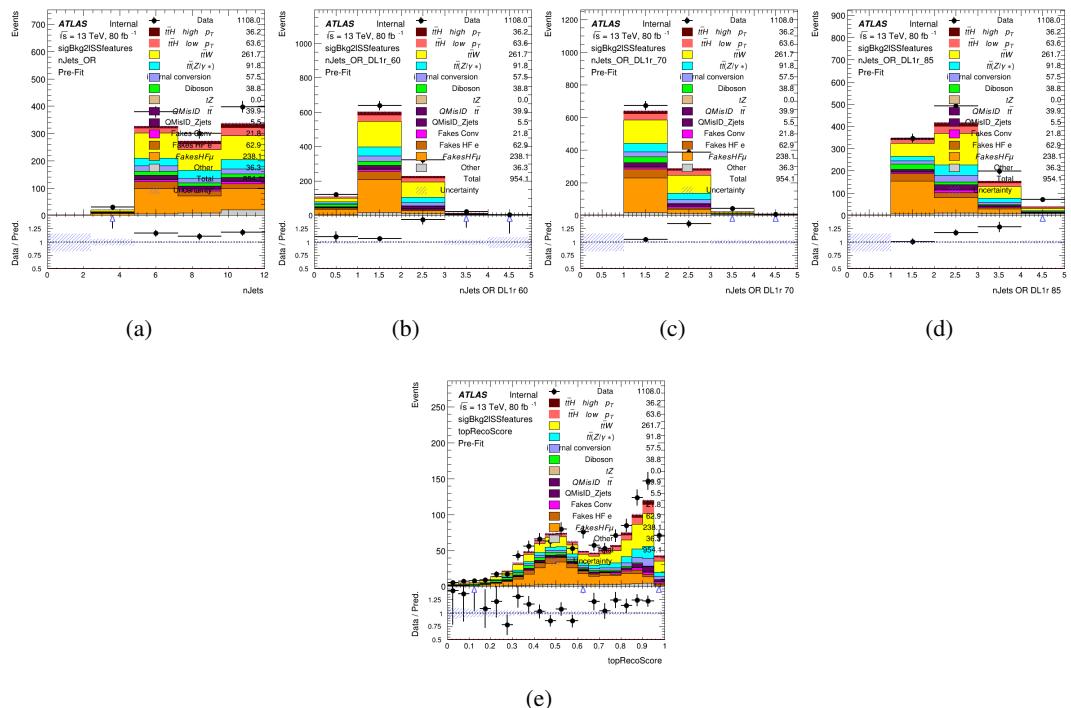


Figure A.21: Input features for sigBkg2ISS

833 **A.2.2 Background Rejection MVA Features - 3l**

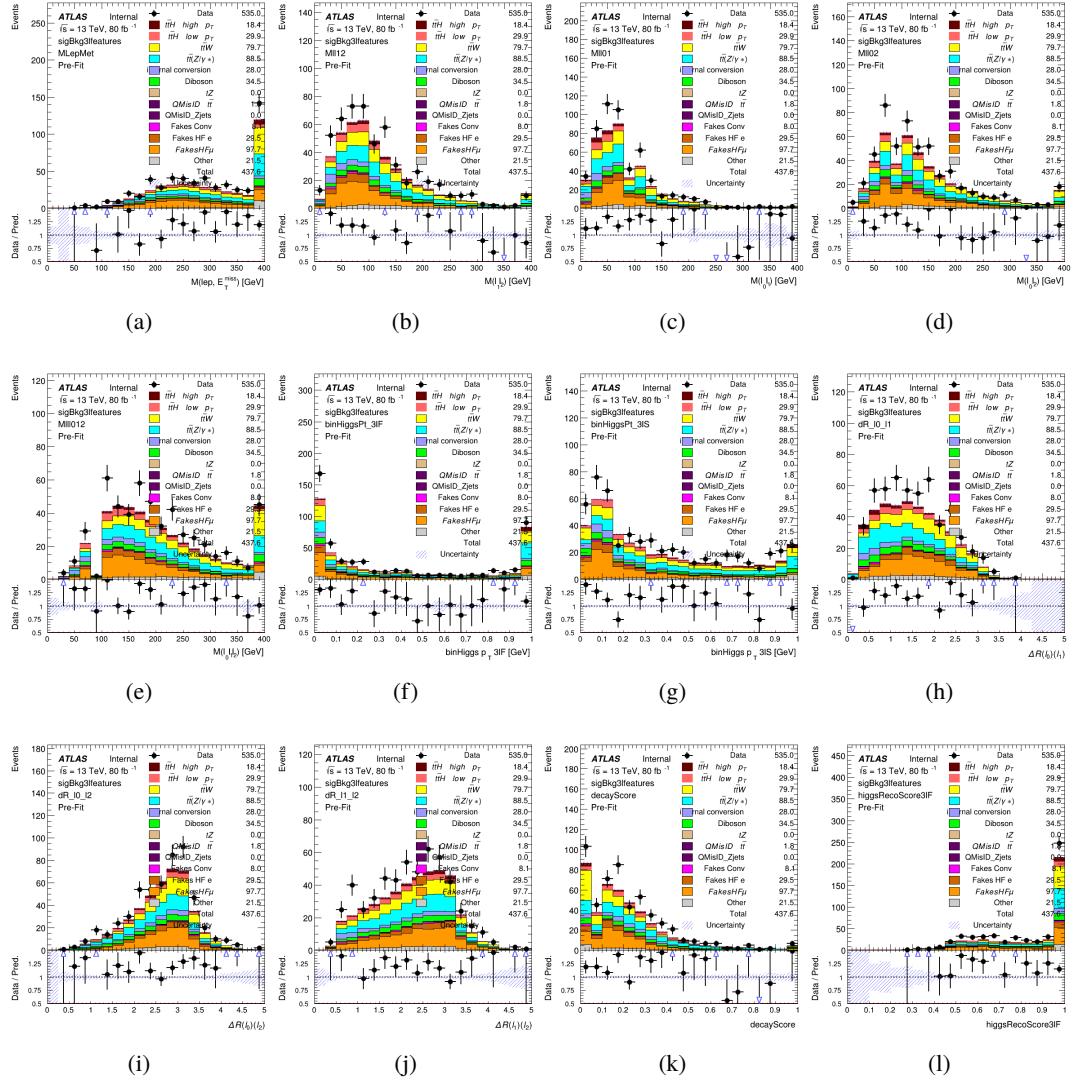


Figure A.22: Input features for sigBkg3l

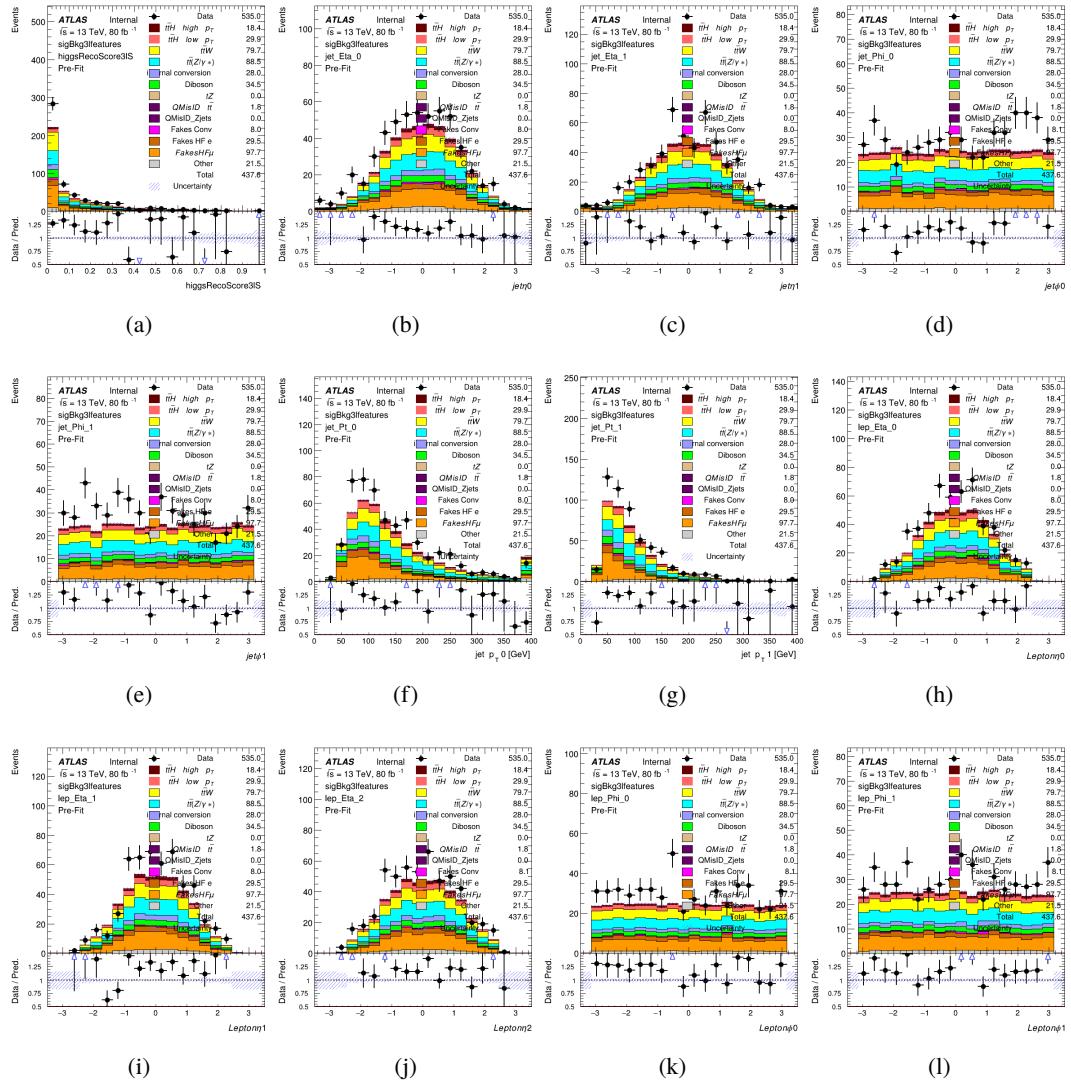


Figure A.23: Input features for sigBkg3l

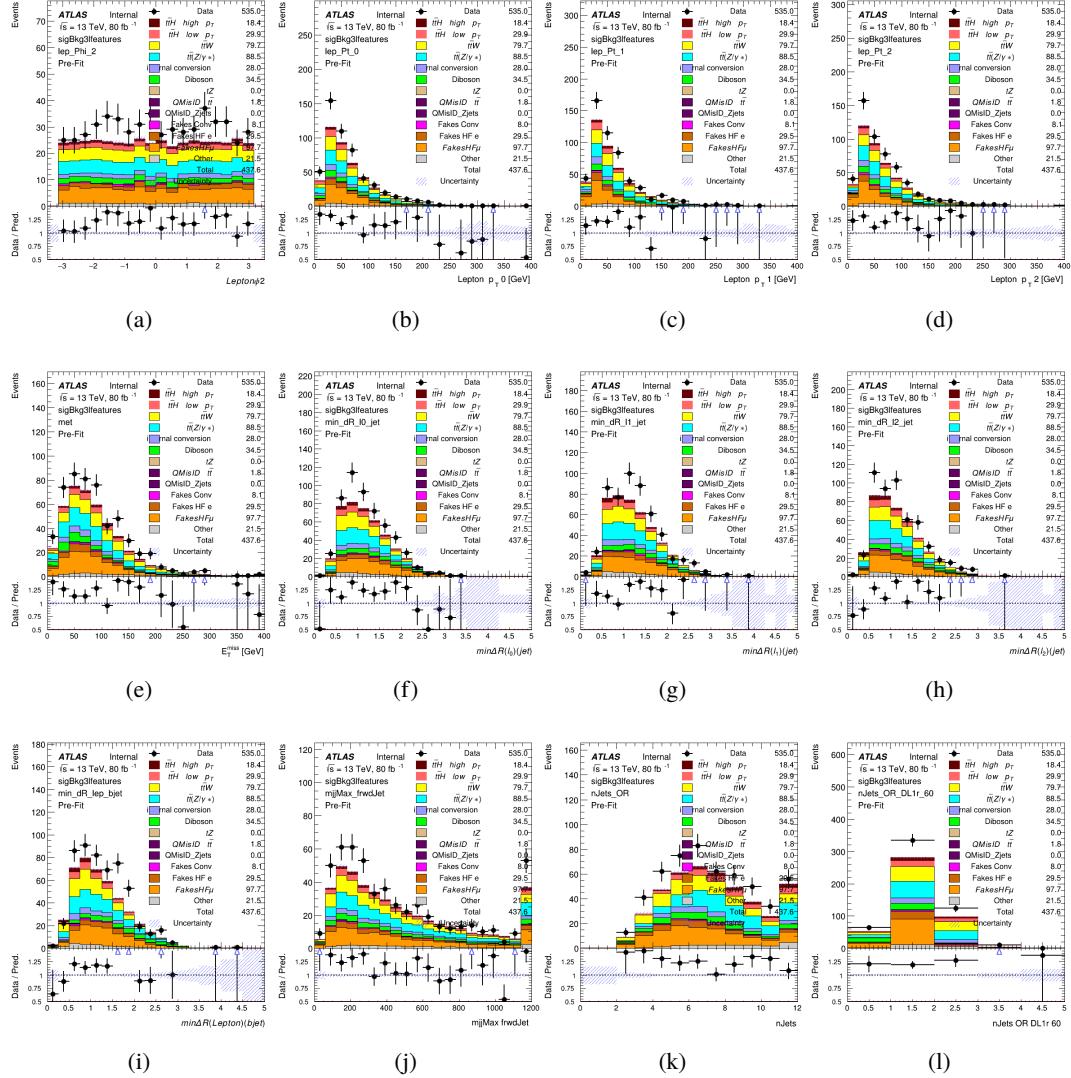


Figure A.24: Input features for sigBkg3l

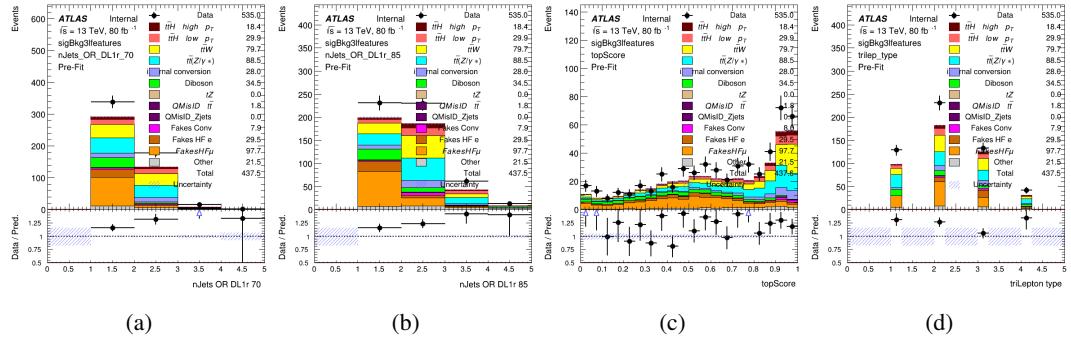


Figure A.25: Input features for sigBkg3l

834 **A.3 Alternate b-jet Identification Algorithm**

835 The nominal analysis reconstructs the b-jets by considering different combinations of jets, and  
 836 asking a neural network to determine whether each combination consists of b-jets from top quark  
 837 decays. An alternate approach would be to give the neural network about all of the jets in an event  
 838 at once, and train it to select which two are most likely to be the b-jets from top decay. It was  
 839 hypothesized that this could perform better than considering each combination independently, as  
 840 the neural network could consider the event as a whole. While this is not found to be the case,  
 841 these studies are documented here as a point of interest and comparison.

842 For these studies, the kinematics of the 10 highest  $p_T$  jets in each event are used for  
 843 training. This includes the vast majority of truth b-jets. Specifically the  $p_T$ ,  $\eta$ ,  $\phi$ ,  $E$ , and DL1r  
 844 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.  
 845 The  $p_T$ ,  $\eta$ ,  $\phi$ , and  $E$  of the leptons and  $E_T^{\text{miss}}$  are included as well. Categorical cross entropy is  
 846 used as the loss function.

Table 23: Accuracy of the NN in identifying b-jets from tops in 2lSS events for the alternate categorical method compared to the nominal method.

Channel	Categorical	Nominal
2lSS	70.6%	73.9%
3l	76.1%	79.8%

847 **A.4 Binary Classification of the Higgs  $p_T$**

848 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer  
 849 resolution. This means separating high and low  $p_T$  events is sufficient for this analysis. As

such, rather than attempting the reconstruct the full Higgs  $p_T$  spectrum, a binary classification approach is explored.

A model is built to determine whether  $t\bar{t}H$  events include a high  $p_T$  ( $>150$  GeV) or low  $p_T$  ( $<150$  GeV) Higgs Boson. While this is now a classification model, it uses the same input features described in section 8.4. Binary crossentropy is used as the loss function.

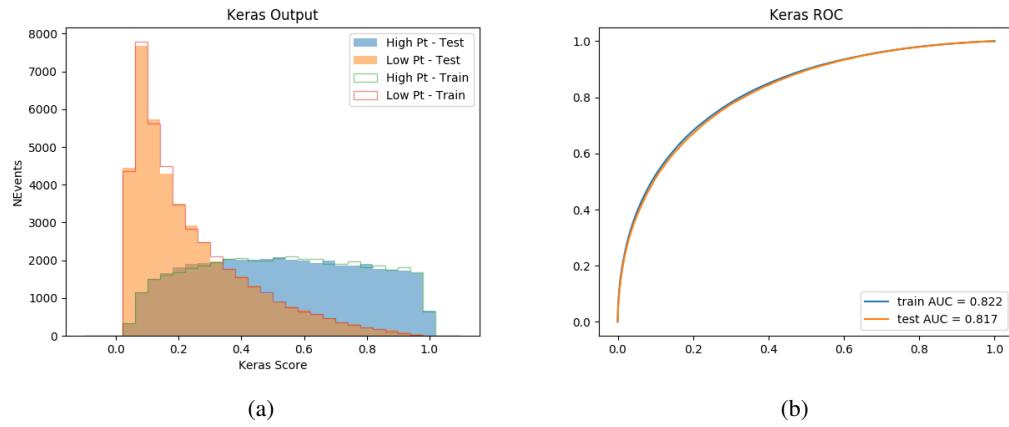


Figure A.26:

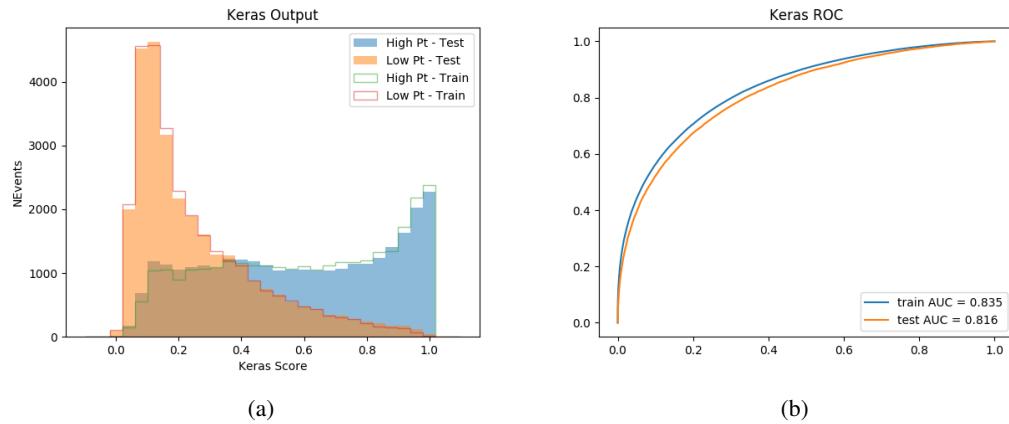


Figure A.27:

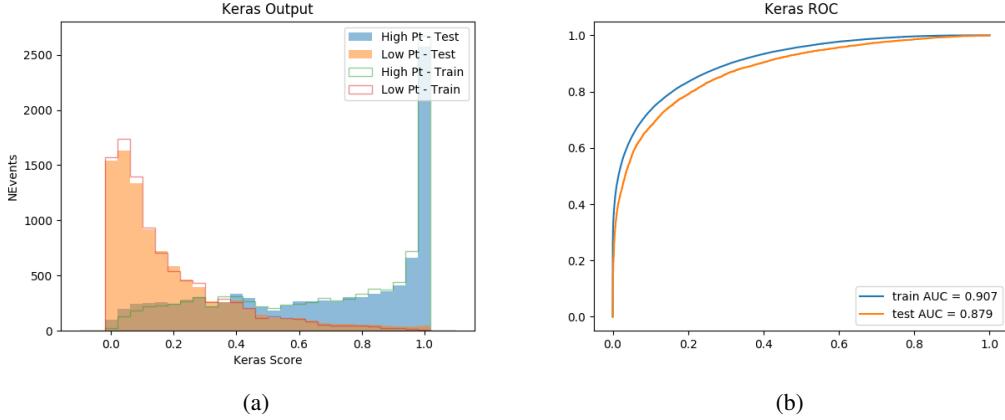


Figure A.28:

## 855    A.5 Impact of Alternative Jet Selection

856    A relatively low  $p_T$  threshold of 15 GeV is used to determine jet candidates, as the jets originating  
 857    from the Higgs decay are found to fall between 15 and 25 GeV a large fraction of the time. The  
 858    impact of different jet  $p_T$  cuts on our ability to reconstruct the Higgs  $p_T$  is explored here. The  
 859    performance of the Higgs  $p_T$  prediction models is evaluated for jet  $p_T$  cuts of 10, 15, 20, and 25  
 860    GeV.

861    **B**

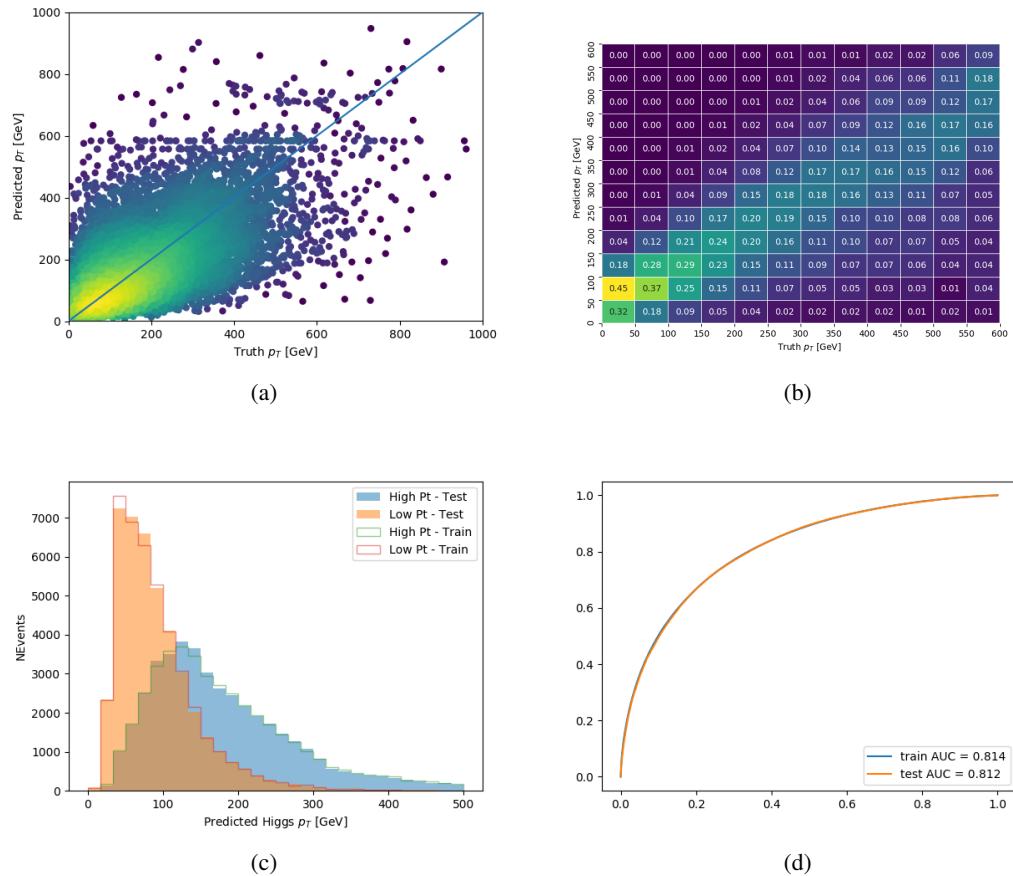


Figure A.29: