



**ATLAS Note**

GROUP-2017-XX



8th 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>5</b>
18	<b>1 Introduction</b>	<b>5</b>
19	<b>II Theoretical Motivation</b>	<b>7</b>
20	<b>2 The Standard Model and the Higgs Boson</b>	<b>7</b>
21	2.1 The Forces and Particles of the Standard Model	8
22	2.2 The Higgs Mechanism	10
23	2.2.1 The Higgs Field	11
24	2.2.2 Electroweak Symmetry Breaking	13
25	2.3 Limitations of the Standard Model	15
26	<b>3 Effective Field Theory in <math>t\bar{t}H</math> Production</b>	<b>16</b>
27	3.1 Extensions to the Higgs Sector	16
28	3.2 Six Dimensional Operators	16
29	<b>III The LHC and the ATLAS Detector</b>	<b>17</b>
30	<b>4 The LHC</b>	<b>17</b>
31	<b>5 The ATLAS Detector</b>	<b>20</b>
32	5.1 Inner Detector	21
33	5.2 Calorimeters	22
34	5.3 Muon Spectrometer	23
35	5.4 Trigger System	24
36	<b>IV Search for Dimension-Six Operators</b>	<b>25</b>
37	<b>6 Data and Monte Carlo Samples</b>	<b>25</b>
38	6.1 Data Samples	26
39	6.2 Monte Carlo Samples	27
40	<b>7 Object Reconstruction</b>	<b>27</b>
41	7.1 Trigger Requirements	28
42	7.2 Light Leptons	28
43	7.3 Jets	29
44	7.4 Missing Transverse Energy	30
45	<b>8 Higgs Momentum Reconstruction</b>	<b>30</b>

---

46	8.1	Decay Candidate Reconstruction	31
47	8.2	b-jet Identification	32
48	8.2.1	2lSS Channel	33
49	8.2.2	3l Channel	36
50	8.3	Higgs Reconstruction	40
51	8.3.1	2lSS Channel	41
52	8.3.2	3l Semi-leptonic Channel	43
53	8.3.3	3l Fully-leptonic Channel	46
54	8.4	$p_T$ Prediction	48
55	8.4.1	2lSS Channel	50
56	8.4.2	3l Semi-leptonic Channel	51
57	8.4.3	3l Fully-leptonic Channel	52
58	8.5	3l Decay Mode	54
59	<b>9</b>	<b>Signal Region Definitions</b>	<b>55</b>
60	9.1	Pre-MVA Event Selection	55
61	9.2	Event MVA	58
62	9.3	Signal Region Definitions	58
63	9.3.1	2lSS	60
64	9.3.2	3l – Semi – leptonic	60
65	9.3.3	3l – Fully – leptonic	60
66	<b>10</b>	<b>Background Rejection MVA</b>	<b>60</b>
67	10.1	Background Rejection MVAs	60
68	10.1.1	2lSS - High $p_T$	62
69	10.1.2	2lSS - Low $p_T$	62
70	10.1.3	3l Semi-Leptonic - High $p_T$	62
71	10.1.4	3l Semi-Leptonic - Low $p_T$	62
72	10.1.5	3l Fully Leptonic - High $p_T$	62
73	10.1.6	3l Fully Leptonic - Low $p_T$	62
74	<b>11</b>	<b>Systematic Uncertainties</b>	<b>62</b>
75	<b>12</b>	<b>Results</b>	<b>64</b>
76	<b>V</b>	<b>Conclusion</b>	<b>64</b>
77	<b>Appendices</b>		<b>66</b>
78	<b>A</b>	<b>Machine Learning Models</b>	<b>66</b>
79	A.1	Alternate b-jet Identification Algorithm	66
80	A.1.1	Binary Classification of the Higgs $p_T$	67
81	A.1.2	Impact of Alternative Jet Selection	67

---

**Part I****Introduction****1 Introduction**

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

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

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

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

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

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

114 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-  
115 2018 provides the opportunity to make this measurement for the first time. The unprecedented  
116 energy achieved by the LHC during this period greatly increase the rate at which  $t\bar{t}H$  events are  
117 produced, and the large amount of data collected provides the necessary statistics for a differential  
118 measurement to be performed.

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

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

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

## 129 **Part II**

### 130 **Theoretical Motivation**

131 **2 The Standard Model and the Higgs Boson**

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

<sup>137</sup> theory, invariant under the Lie Group  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ , where C refers to color  
<sup>138</sup> charge, L, the helicity of the particle, and Y, the hypercharge.

## <sup>139</sup> 2.1 The Forces and Particles of the Standard Model

<sup>140</sup> The SM particles, summarized in figure 2.1, can be classified into two general categories based  
<sup>141</sup> on their spin: fermions, and bosons.

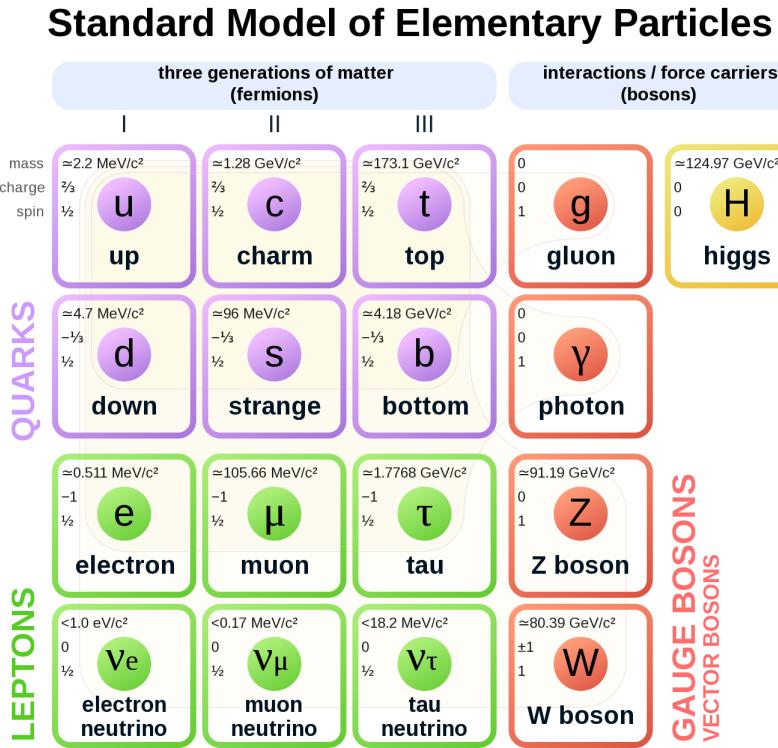


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. [ ]

<sup>142</sup> Fermions are particles with  $\frac{1}{2}$ -integer spin, which according to the spin-statistics theorem,  
<sup>143</sup> causes them to comply with the Pauli-exclusion principle [ ]. They can be separated into two

<sup>144</sup> groups, leptons and quarks, each of which consist of three generations of particles with increasing  
<sup>145</sup> mass.

<sup>146</sup> Leptons are fermions interact via the electroweak force, but not the strong force. The three  
<sup>147</sup> generation of leptons consist of the electron and electron neutrino, the muon and muon neutrino,  
<sup>148</sup> the tau and tau neutrino. The quarks, which do interact via the strong force - which is to say they  
<sup>149</sup> have color charge - in addition to the electroweak force. The three generations include the up  
<sup>150</sup> and down quarks, the strange and charm quarks, and the top and bottom quarks. Each of these  
<sup>151</sup> generations form left-handed doublets invariant under SU(2) transformations. For the leptons  
<sup>152</sup> these doublets are:

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

<sup>153</sup> And for the quarks:

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

<sup>154</sup> For both leptons and quarks, the heavier generations can decay into the lighter generation  
<sup>155</sup> of particles, while the first generation does not decay. Hence, ordinary matter generally consists  
<sup>156</sup> of this first generation of fermions - electrons, up quarks, and down quarks. Each of these  
<sup>157</sup> fermions has a corresponding anti-particle, which has an equal mass as its partner but opposite

158 charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos,  
159 whose mass has been experimentally confirmed but is not accounted for in the SM.

160 Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-  
161 exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1  
162 particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0,  
163 particle - the Higgs Boson. Of the gauge bosons, the  $W^+$ ,  $W^-$  and Z bosons - which are the  
164 mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon  
165 mediates the electric force, and the gluon mediates the strong force.

166 **2.2 The Higgs Mechanism**

167 A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to  
168 the Lagrangian giving mass to the the gauge bosons would violate the underlying symmetry of  
169 the theory. This presents a clear problem with the theory: The experimental observation that the  
170 W and Z bosons have mass seems to contradict the basic structure of the SM.

171 Rather than abandoning gauge invariance, an alternative way for particles to acquire mass  
172 beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout  
173 in 1964 []. This procedure for introducing masses for the gauge bosons while preserving local  
174 gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory  
175 by Weinberg in 1967 [].

<sup>176</sup> **2.2.1 The Higgs Field**

<sup>177</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>178</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>179</sup> Where  $\mu$  and  $\lambda$  are free parameters of the new field. This represents the most general  
<sup>180</sup> potential allowed while preserving  $SU(2)_L$  invariance and renormalizability. In the case that  
<sup>181</sup>  $\mu^2 < 0$ , this potential takes the form shown in figure 2.2.

<sup>182</sup> The significant feature of this potential is that its minimum does not occur for a value of  
<sup>183</sup>  $\Phi = 0$ . Instead, it is minimized when  $|\Phi^\dagger \Phi| = -\mu^2/\lambda$ . This means that in its ground state, the  
<sup>184</sup> Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while  
<sup>185</sup> the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

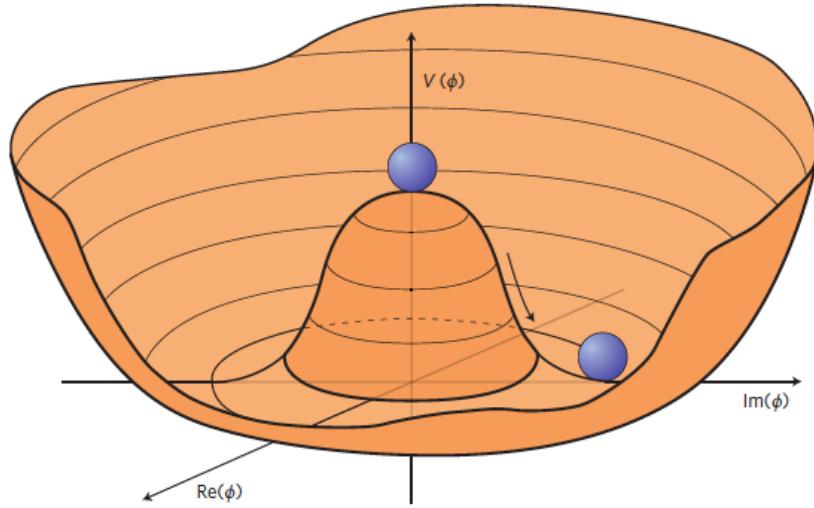


Figure 2.2: The value of the Higgs potential,  $V(\Phi)$  as a function of  $\Phi$ , for the case that  $\mu^2 < 0$  [].

<sup>186</sup> the minimum is determined only by  $\Phi^\dagger \Phi$ , there is some ambiguity in the particular definition of  
<sup>187</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>188</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>189</sup> with  $v$  being the value of the VEV, and  $H$  being the real value of the scalar field.

190 **2.2.2 Electroweak Symmetry Breaking**

191 The Electroweak (EWK) interaction is described in the SM by a  $SU(2)_L \otimes U(1)_Y$  gauge theory.  
 192 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,  
 193  $B_\mu$ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar  
 194 field  $\Phi$  in the Lagrangian:

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

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

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

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

<sup>201</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>202</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>203</sup> This produces exactly the particles we observe - three massive gauge bosons and a single  
<sup>204</sup> massless photon. The massless photon represents the portion of the gauge symmetry, a single  
<sup>205</sup>  $U(1)$  of the electromagnetic force, that remains unbroken by the VEV.

<sup>206</sup> Interactions with the Higgs field also lead to the generation of the fermion masses, which  
<sup>207</sup> in the Lagrangian take the form:

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

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

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

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

### 221        2.3 Limitations of the Standard Model

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

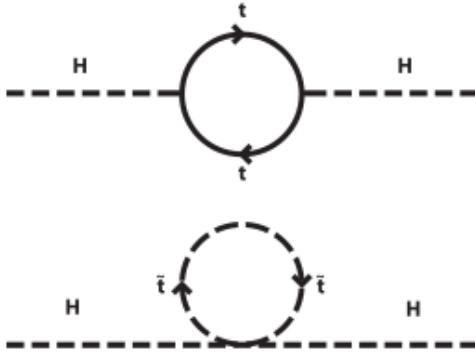


Figure 2.3: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram [6].

### <sup>225</sup> **3 Effective Field Theory in $t\bar{t}H$ Production**

<sup>226</sup> Higher dimension operators are a common way to paramaterize the effects of physics at very  
<sup>227</sup> high energies into

#### <sup>228</sup> **3.1 Extensions to the Higgs Sector**

#### <sup>229</sup> **3.2 Six Dimensional Operators**

<sup>230</sup> While the SM has been tested to great precision, particularly at the LHC, it is generally accepted  
<sup>231</sup> that it is only valid up to a certain energy scale. It is assumed that above a certain energy, at the  
<sup>232</sup> scale where something like a Grand Unified Theory (GUT) or quantum gravity become relevant,  
<sup>233</sup> the SM will not be applicable.

---

## 234 **Part III**

### 235 **The LHC and the ATLAS Detector**

#### 236 **4 The LHC**

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

246 The LHC first began running in 2009 at a proton-proton center of mass energy of  $\sqrt{s} = 8$   
247 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during  
248 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,  
249 and collected data at an increased energy of  $\sqrt{s} = 13$  TeV until 2018, a period referred to as Run  
250 2.

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

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

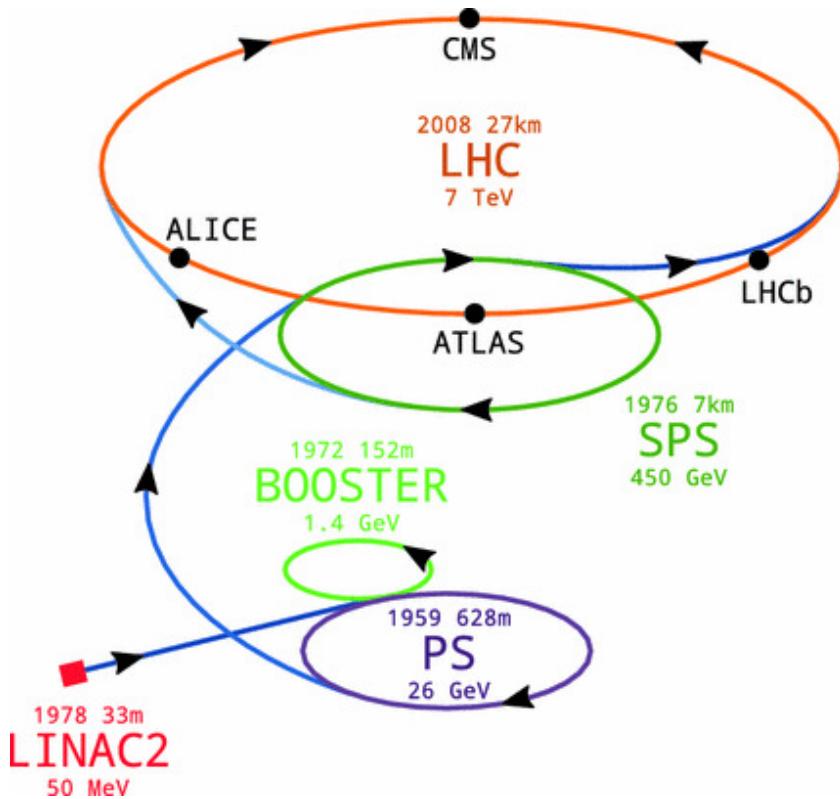


Figure 4.1: A summary of the accelerator chain used to feed protons into the LHC [1].

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

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

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

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

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

270 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around  
 271  $140 \text{ fb}^{-1}$ , exceeding the expected integrated luminosity of  $100 \text{ fb}^{-1}$ .

## <sup>272</sup> 5 The ATLAS Detector

<sup>273</sup> ATLAS (a not terribly natural acronym for “A Toroidal LHC Apparatus”) is a general purpose  
<sup>274</sup> detector designed to maximize the detection efficiency of all physics objects, including leptons,  
<sup>275</sup> jets, and photons. This means it is capable of measuring all SM particles, with the exception of  
<sup>276</sup> neutrinos, the presence of which can be inferred based on missing transverse momentum. The  
<sup>277</sup> detector measures 44 m long, and 25 m tall.

<sup>278</sup> The ATLAS detector consists of multiple layers, each of which serves a different purpose  
<sup>279</sup> in reconstructing collisions. At the very center of the detector is the interaction point where the  
<sup>280</sup> proton beams of the LHC collide.

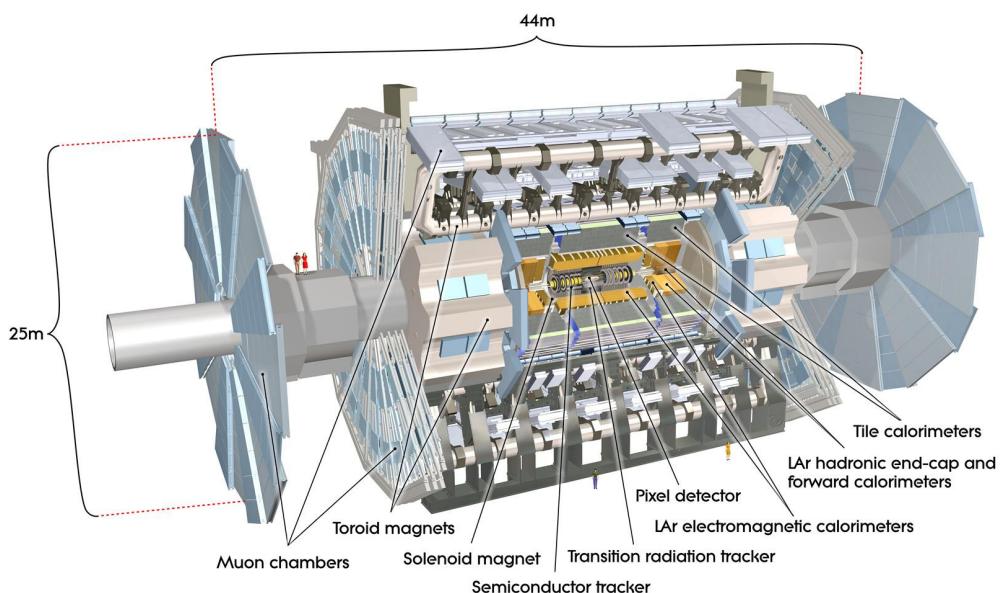


Figure 5.1: Cutaway view of the ATLAS detector, with labels of its major components [].

281 **5.1 Inner Detector**

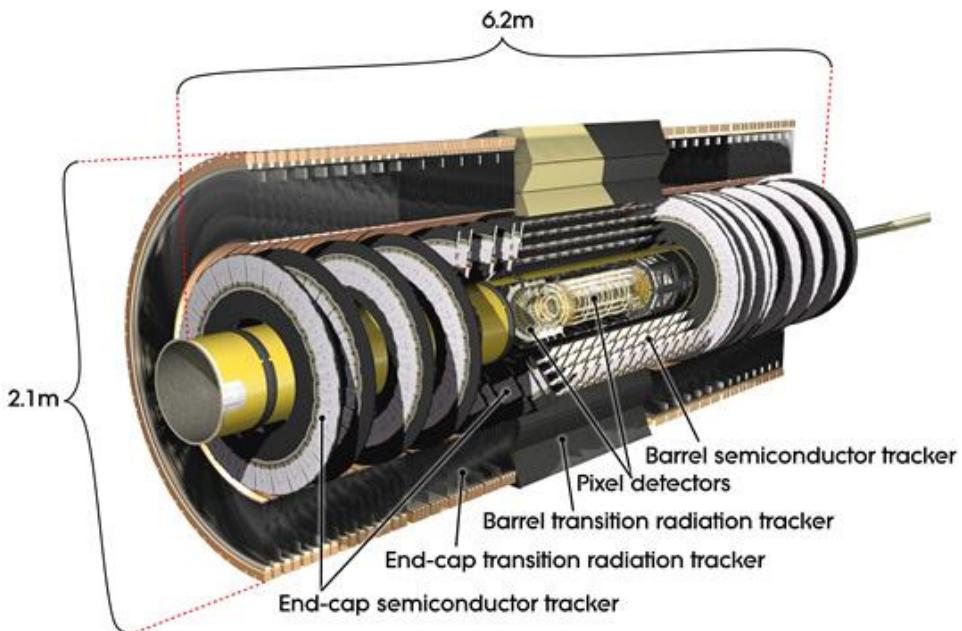


Figure 5.2: Cutaway view of the Inner Detector [].

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

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

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

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

294 **5.2 Calorimeters**

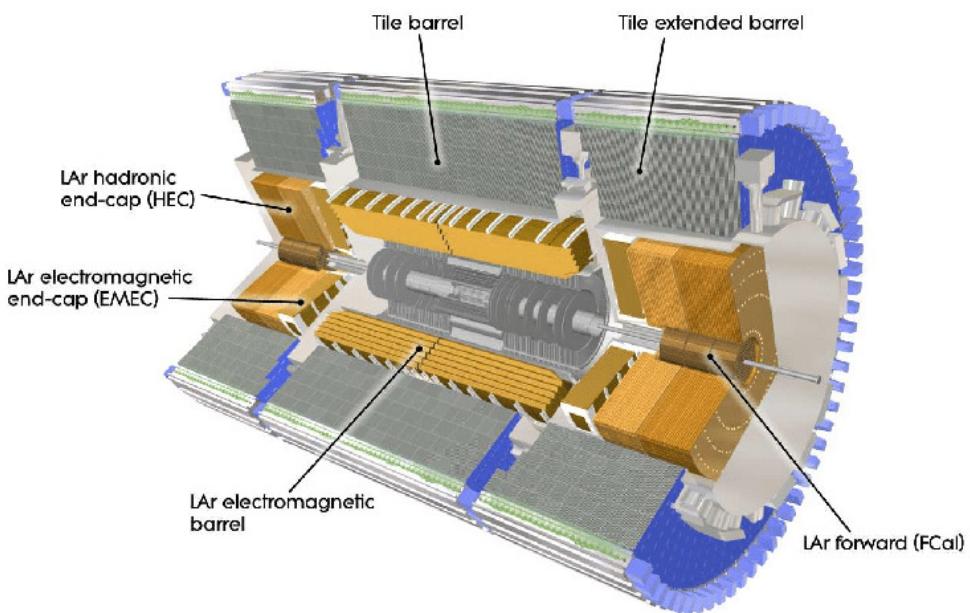


Figure 5.3: Cutaway view of the calorimeter system of the ATLAS detector [].

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

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

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

307 **5.3 Muon Spectrometer**

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

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

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

322 1200 tracking chambers are placed in the muon system in order to precisely measure the  
323 tracks of muons with high spatial resolution.

324 **5.4 Trigger System**

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

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

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

## 343 **Part IV**

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

#### 345 **6 Data and Monte Carlo Samples**

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

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

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

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

357       **6.1 Data Samples**

358       The study uses proton-proton collision data collected by the ATLAS detector from 2015 through  
359       2018, which represents an integrated luminosity of  $139 \text{ fb}^{-1}$  and an energy of  $\sqrt{s} = 13 \text{ TeV}$ . All  
360       data used in this analysis was included in one the following Good Run Lists:

- 361       • data15\_13TeV.periodAllYear\_DetStatus-v79-repro20-02\_DQDefects-00-02-02  
362            \_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- 363       • data16\_13TeV.periodAllYear\_DetStatus-v88-pro20-21\_DQDefects-00-02-04  
364            \_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- 365       • data17\_13TeV.periodAllYear\_DetStatus-v97-pro21-13\_Unknown\_PHYS\_StandardGRL  
366            \_All\_Good\_25ns\_Triggerno17e33prim.xml
- 367       • data18\_13TeV.periodAllYear\_DetStatus-v102-pro22-04\_Unknown\_PHYS\_StandardGRL  
368            \_All\_Good\_25ns\_Triggerno17e33prim.xml

<sup>369</sup> **6.2 Monte Carlo Samples**

<sup>370</sup> Several Monte Carlo (MC) generators were used to simulate both signal and background pro-  
<sup>371</sup> cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific  
<sup>372</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>373</sup> **7 Object Reconstruction**

<sup>374</sup> All analysis channels considered in this note share a common object selection for leptons and  
<sup>375</sup> jets, as well as a shared trigger selection.

<sup>376</sup> **7.1 Trigger Requirements**

<sup>377</sup> 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.

<sup>378</sup> **7.2 Light Leptons**

<sup>379</sup> Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that  
<sup>380</sup> are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].  
<sup>381</sup> Electron candidates are required to have  $p_T > 10$  GeV and  $|\eta_{\text{cluster}}| < 2.47$ . Candidates in the  
<sup>382</sup> transition region between different electromagnetic calorimeter components,  $1.37 < |\eta_{\text{cluster}}| <$   
<sup>383</sup> 1.52, are rejected. A multivariate likelihood discriminant combining shower shape and track

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

386 To further reduce the non-prompt contribution, the track of each electron is required to  
387 originate from the primary vertex; requirements are imposed on the transverse impact parameter  
388 significance ( $|d_0|/\sigma_{d_0}$ ) and the longitudinal impact parameter ( $|\Delta z_0 \sin \theta_\ell|$ ), as shown in table  
389 ??.

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

### 394 7.3 Jets

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

---

<sup>402</sup> **7.4 Missing Transverse Energy**

<sup>403</sup> Because all  $t\bar{t}H - ML$  channels considered include multiple neutrinos, missing transverse  
<sup>404</sup> energy ( $E_T^{\text{miss}}$ ) is present in each event. The missing transverse momentum vector is defined as  
<sup>405</sup> the inverse of the sum of the transverse momenta of all reconstructed physics objects as well  
<sup>406</sup> as remaining unclustered energy, the latter of which is estimated from low- $p_T$  tracks associated  
<sup>407</sup> with the primary vertex but not assigned to a hard object [ATL-PHYS-PUB-2015-027].

<sup>408</sup> **8 Higgs Momentum Reconstruction**

<sup>409</sup> Reconstructing the momentum of the Higgs boson is a particular challenge for channels with  
<sup>410</sup> leptons in the final state: Because all channels include at least two neutrinos in the final state, the  
<sup>411</sup> Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted  
<sup>412</sup> by a neural network when provided with the four-vectors of the Higgs Boson decay products, as  
<sup>413</sup> shown in section 8.1. With this in mind, several layers of MVAs are used to reconstruction the  
<sup>414</sup> Higgs momentum.

<sup>415</sup> The first layer is a model designed to select which jets are most likely to be the b-jets  
<sup>416</sup> that came from the top decay, detailed in section 8.2. As described in section 8.3, the kinematics  
<sup>417</sup> of these jets are fed into the second layer, which is designed to identify the decay products of  
<sup>418</sup> the Higgs Boson itself. The kinematics of these particles are then fed into yet another neural-

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

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

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

## 435 8.1 Decay Candidate Reconstruction

436 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson  
437 using MC simulations of  $t\bar{t}H$  events. These include light leptons and jets. Reconstructed

438 physics objects are matched to truth level particles, in order to identify the parents of these  
439 reconstructed objects. The kinematics of the decay product candidates as well as event level  
440 variables are used as inputs.

441 Leptons considered as possible Higgs and top decay candidates are required to pass the  
442 selection described in section 7.2. For jets, however, it is found that a large fraction that originate  
443 from either the top decay or the Higgs decay fall outside the selection described in section 7.3.  
444 Specifically, jets from the Higgs decay tend to be soft, with 32% having  $p_T < 25$  GeV. Therefore  
445 jets with  $p_T < 15$  GeV are considered as possible candidates in the models described below. By  
446 contrast, less than 5% of the jets originating from the Higgs fall below this  $p_T$ . The jets are found  
447 to be well modeled even down to this low  $p_T$  threshold, as shown in section 9.1. The impact of  
448 using different  $p_T$  selection for the jet candidates is considered in detail in section ???. As they  
449 are expected to originate from the primary vertex, jets are also required to pass a JVT cut.

## 450 8.2 b-jet Identification

451 Including the kinematics of the b-jets that originate from the top decay is found to improve the  
452 identification of the Higgs decay products, and improve the accuracy with which the Higgs  
453 momentum can be reconstructed. Because these b-jets are reconstructed by the detector with  
454 high efficiency (just over 90% of the time), and can be identified relatively consistently, the first  
455 step in reconstructing the Higgs is selecting the b-jets from the top decay.

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

464        Once the network is trained, all possible pairings of jets are fed into the model, and the pair  
 465        of jets with the highest output score are taken to be b-jets in successive steps of the analysis.

### 466        8.2.1 2lSS Channel

467        For the 2lSS channel, the input features shown in table 3 are used for training. Here  $j_0$  and  $j_1$   
 468        are the two jet candidates, while  $l_0$  and  $l_1$  are the two leptons in the event, ordered by  $p_T$ . jet  
 469        DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,  
 470        where 5 represents the tightest working point and 1 represents the loosest. The variables nJets  
 471        DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and  
 472        85% b-tag working points, respectively.

473        As there are far more incorrect combinations than correct ones, by a factor of more than  
 474        20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random

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 2ISS b-jet identification algorithm

475 sample of 5 million incorrect entries are used for training, along with close 1 million correct  
 476 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for  
 477 training.

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

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

485 Based on the results of grid search evaluation, the optimal architecture is found to include  
 486 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as  
 487 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for

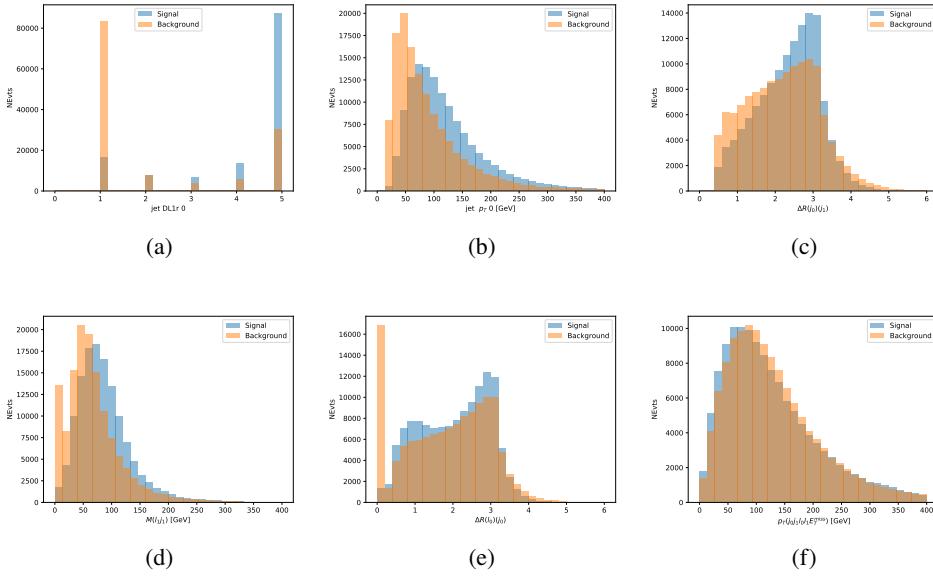


Figure 8.1: Input features for top2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

488 the trained model are shown in figure 8.2.1. The model is found to identify the correct pairing  
489 of jets for 73% of 2lSS signal events on test data.

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

494 The accuracy of the model for different values of n-bjets, compared to this naive approach,  
495 is shown in table 4.

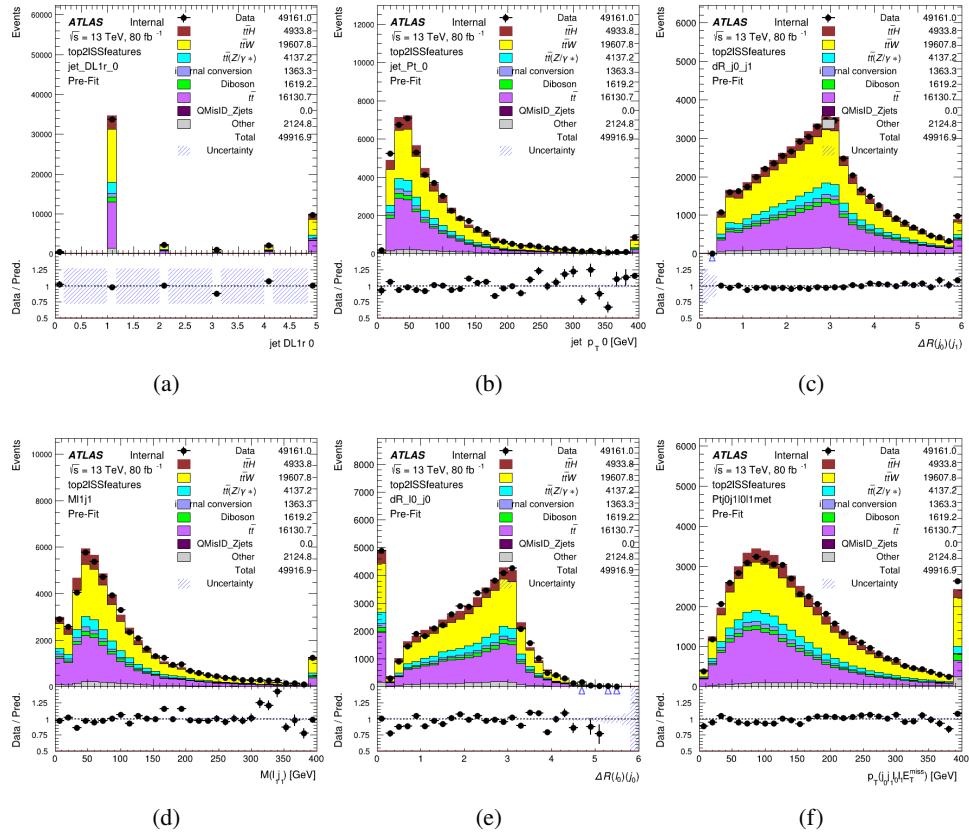


Figure 8.2: Data/MC comparisons of input features for top2ISS training for  $80 \text{ fb}^{-1}$  of data.

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.

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%

### 496 8.2.2 3l Channel

497 The input features used in the 3l channel are listed in table 5, with the same naming convention  
498 as the 2ISS channel.

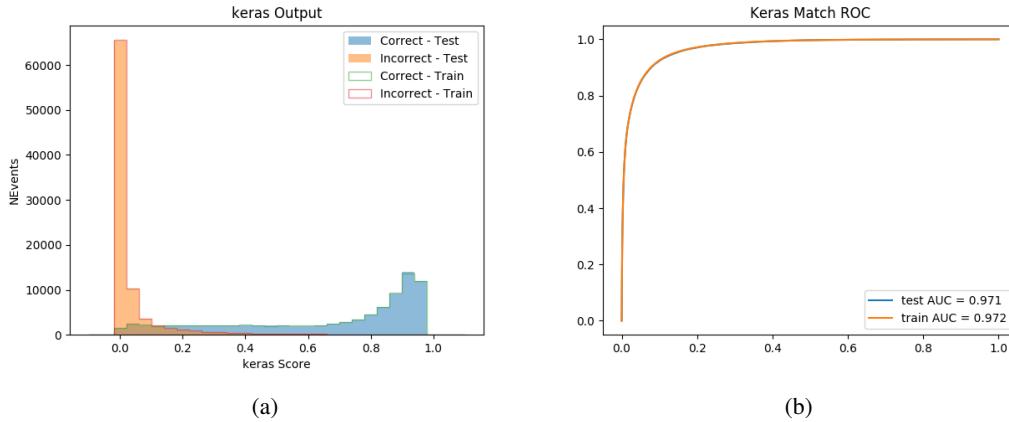


Figure 8.3: (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

499        A few of these features are shown in figure 8.4, comparing the distributions for correct and  
 500        incorrect combinations of jets.

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

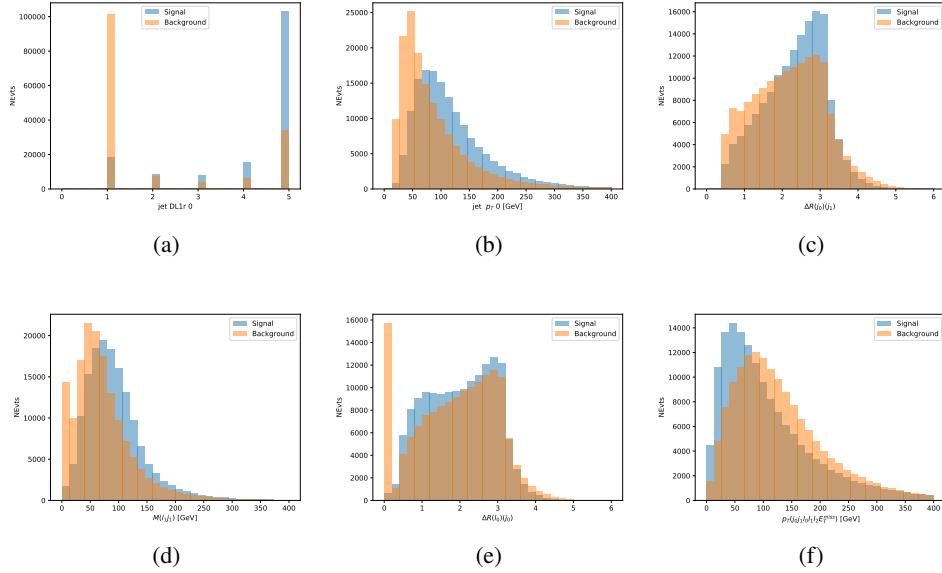


Figure 8.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

509 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal  
 510 events. The accuracy of the model is summarized in table 6.

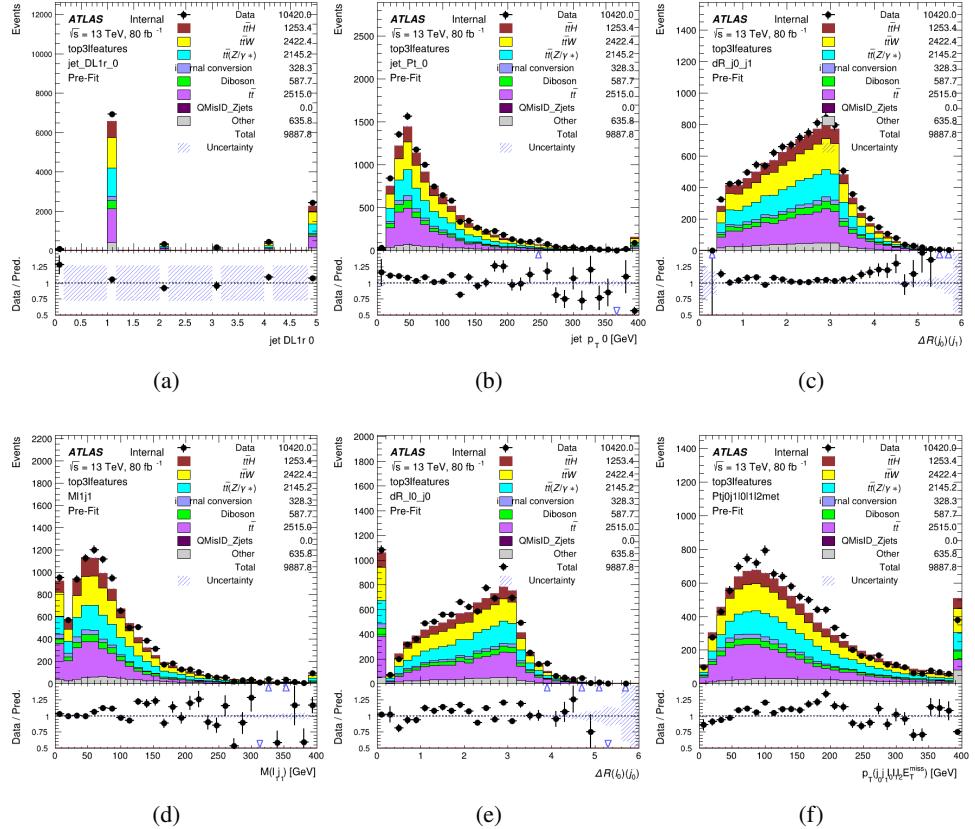
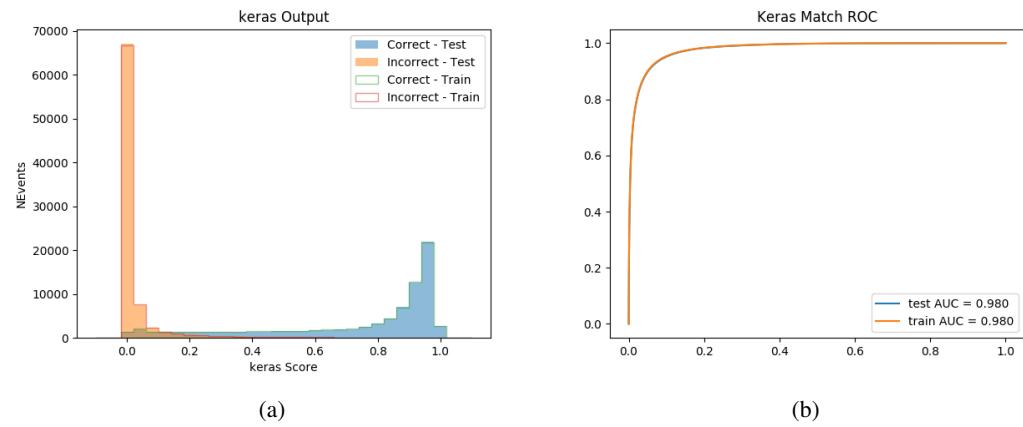
Figure 8.5: Data/MC comparisons of input features for top3l training for  $80 \text{ fb}^{-1}$  of data.

Figure 8.6: tmp

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%

### 511 8.3 Higgs Reconstruction

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

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

519 3l events are therefore categorized as either semi-leptonic or fully-leptonic. In the semi-  
 520 leptonic case the reconstructed decay products consist of two jets and a single leptons. For  
 521 the fully-leptonic case, the decay products include 2 of the three leptons associated with the  
 522 event. For training the models, events are separated into these two categories using truth level  
 523 information. A separate MVA, described in section 8.5, is used to make this distinction at reco  
 524 level and determine which model to use.

525 For all channels, the models described in section 8.2 are used to identify b-jet candidates,

526 whose kinematics are used to identify the Higgs decay products. These jets are not considered  
527 as possible candidates for the Higgs decay, justified by the fact that these models are found to  
528 misidentify jets from the Higgs decay as jets from the top decay less than 1% of the time.

529 **8.3.1 2ISS Channel**

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

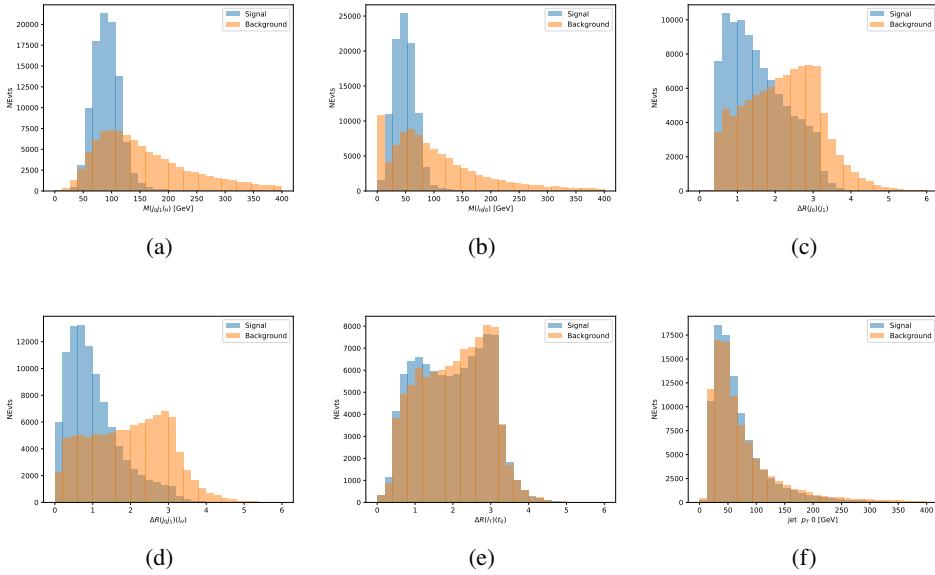


Figure 8.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

536        The neural network identifies the correct combination 55% of the time. It identifies the  
 537        correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets  
 538        81% of the time.

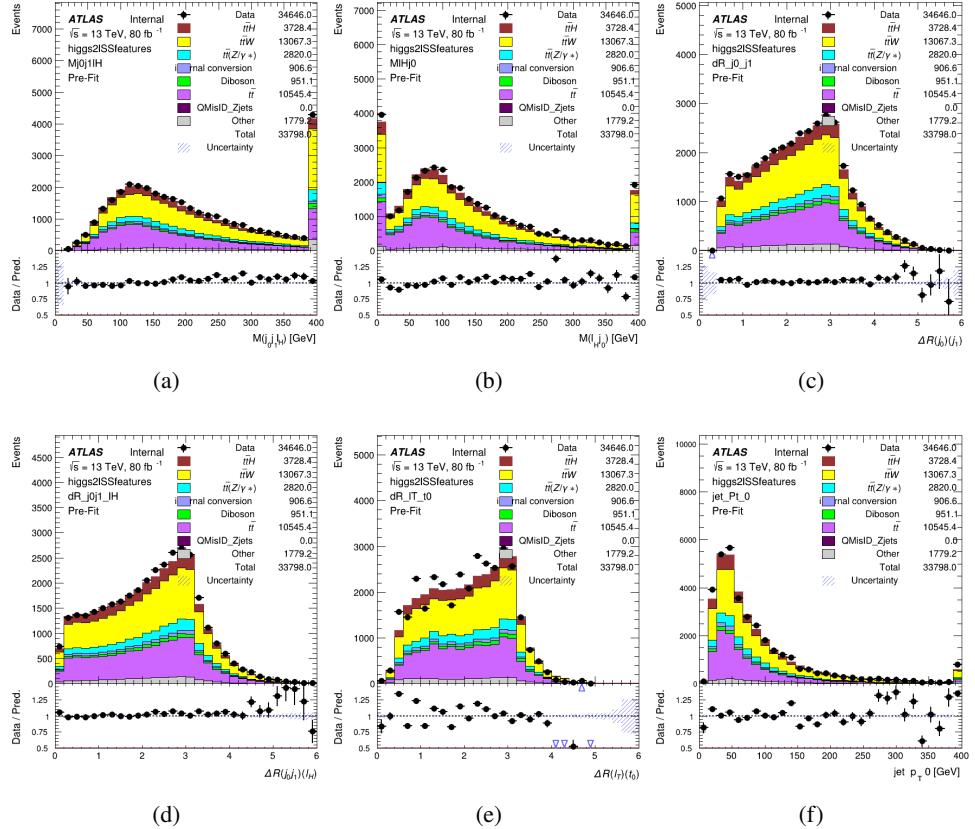


Figure 8.8: Data/MC comparisons of input features for higgs2lSS training for  $80 \text{ fb}^{-1}$  of data.

### 539 8.3.2 3l Semi-leptonic Channel

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

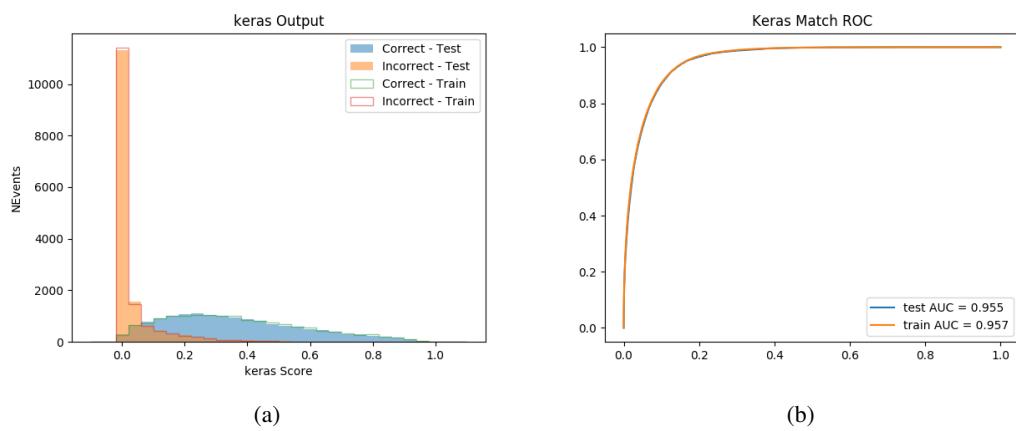


Figure 8.9: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

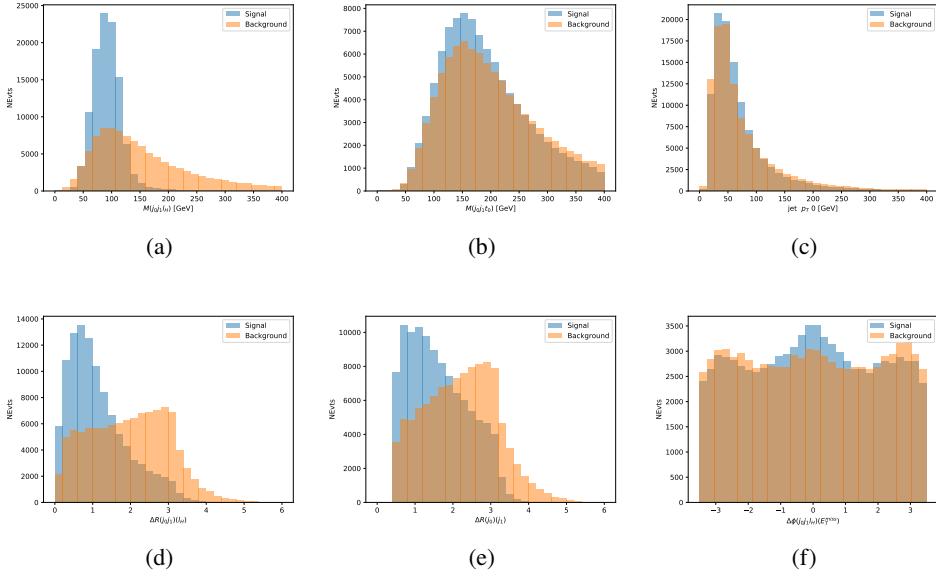


Figure 8.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

544        The modeling of these inputs is validated against data, with figure 8.11 showing good  
 545        general agreement between data and MC. Plots for the complete list of features can found in  
 546        section A.

547        The neural network identifies the correct combination 65% of the time. It identifies the  
 548        correct lepton 85% of the time, anselects the correct lepton and at least one of the correct jets  
 549        83% of the time.

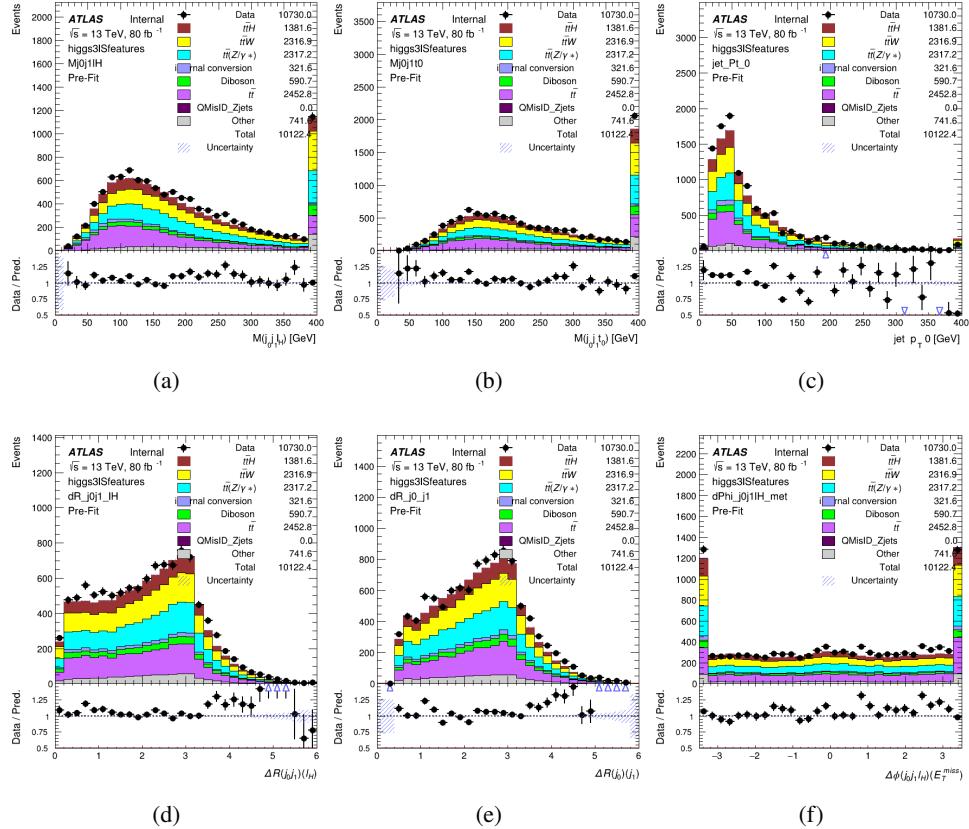


Figure 8.11: Data/MC comparisons of input features for higgs3lS training for  $80 \text{ fb}^{-1}$  of data.

### 550 8.3.3 3l Fully-leptonic Channel

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

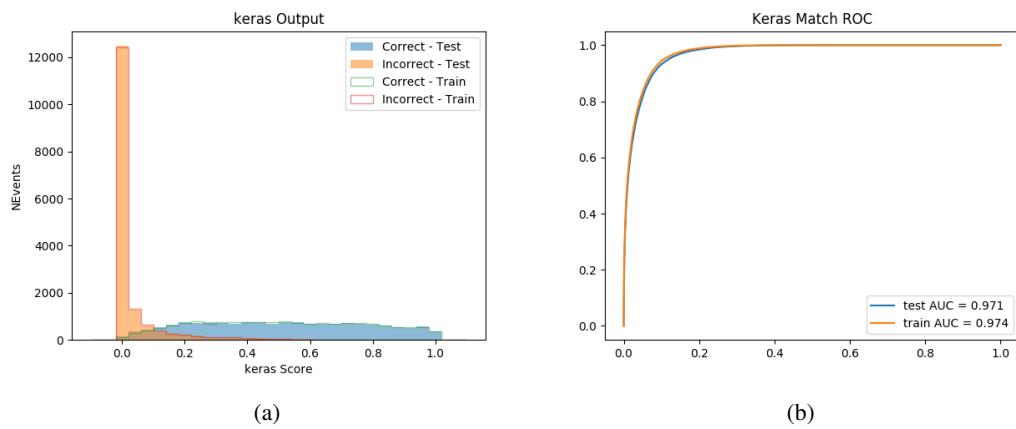


Figure 8.12: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

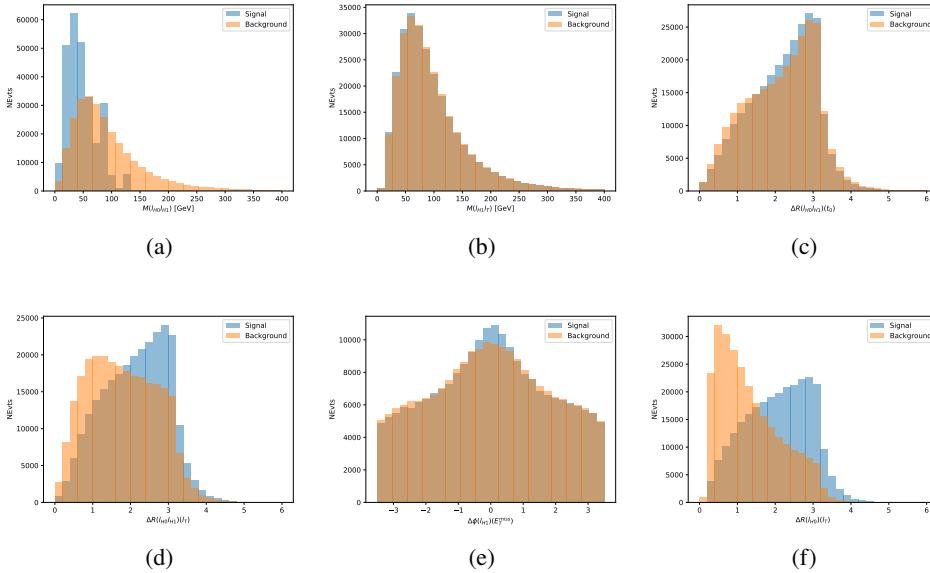


Figure 8.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

559        The correct lepton is identified 80% of the time.

## 560        8.4 $p_T$ Prediction

561        Once the most probable decay products have been identified, their kinematics are used as inputs  
 562        to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,  
 563        a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay

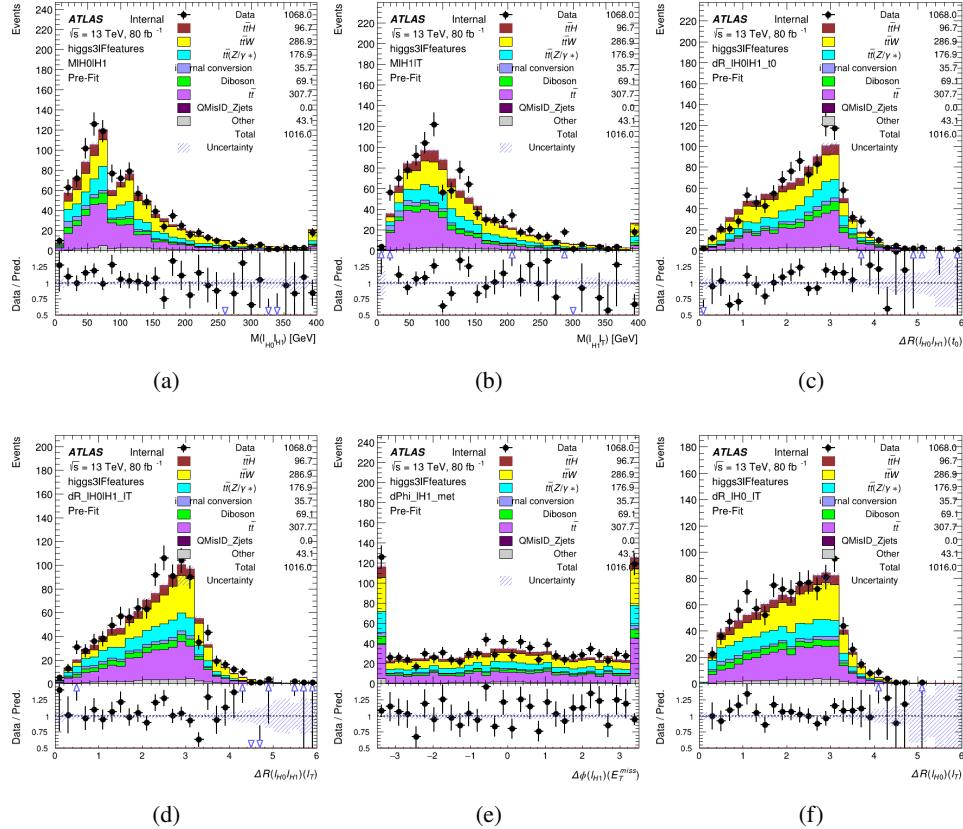


Figure 8.14: Data/MC comparisons of input features for higgs3IF training for  $80 \text{ fb}^{-1}$  of data.

564 are included as well, as these are found to improve performance. The truth  $p_T$  of the Higgs,  
565 as predicted by MC, are used as labels. Separate models are built for each channel - 2lSS, 3l  
566 Semi-leptonic and 3l Fully-leptonic.

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

570 Because the analysis uses a two bin fit of the Higgs  $p_T$ , the momentum reconstruction

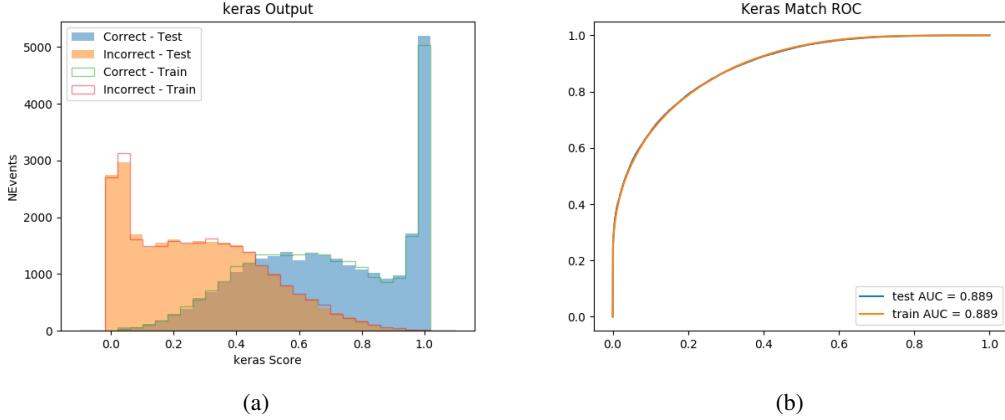


Figure 8.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

571 could be treated as a binary classification problem, rather than a regression problem. This  
 572 approach is explored in detail in section A.1.1, and is found not to provide any significant  
 573 increase in sensitivity. The regression approach is used because it provides more flexibility  
 574 for future analyses, as it is independent of the cutoff between high and low  $p_T$ , as well as the  
 575 number of bins. Further, a regression allows the output of the neural network to be more clearly  
 576 understood, as it can be directly compared to a physics observable.

#### 577 8.4.1 2ISS Channel

578 The input variables listed in table ?? are used to predict the Higgs  $p_T$  in the 2ISS channel. Here  
 579  $j_0$  and  $j_1$  are the two jets identified as Higgs decay products. The lepton identified as originating  
 580 from the Higgs is labeled  $l_H$ , while the other lepton is labeled  $l_T$ , as it most have come from the  
 581 decay of one of the top quarks. The Higgs Reco Score and b-jet Reco Score are the outputs of  
 582 the Higgs reconstruction algorithm, and the b-jet identification algorithm, respectively.

583        The optimal neural network architecture for this channel is found to consist of 5 hidden  
 584        layers with 40 nodes each. The input data set includes 1.2 million events, 10% of which is used  
 585        for testing, the other 90% for training. Training is found to converge after around 150 epochs.

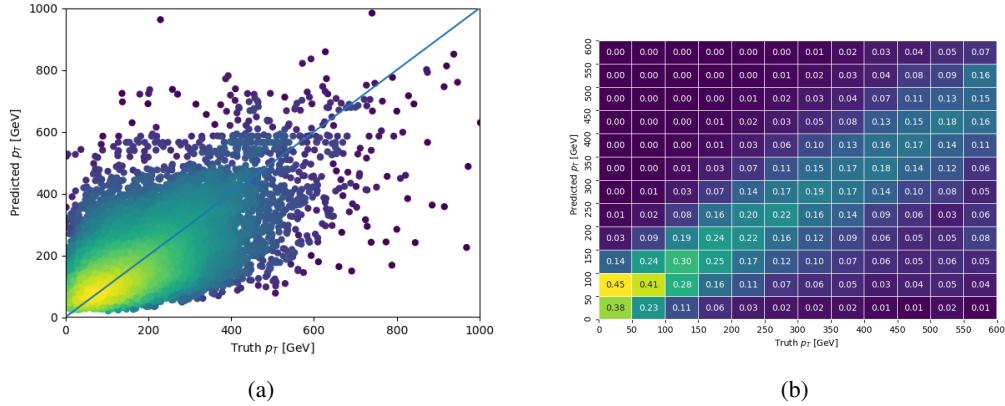


Figure 8.16:

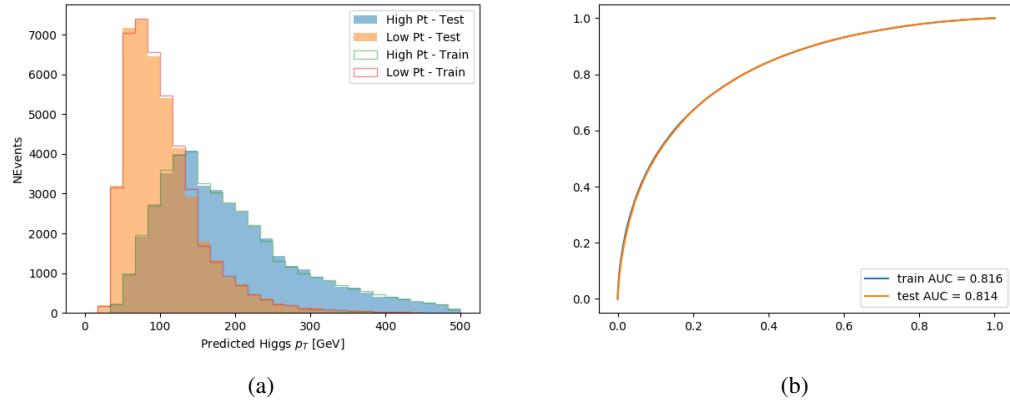


Figure 8.17:

#### 586        8.4.2 3l Semi-leptonic Channel

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

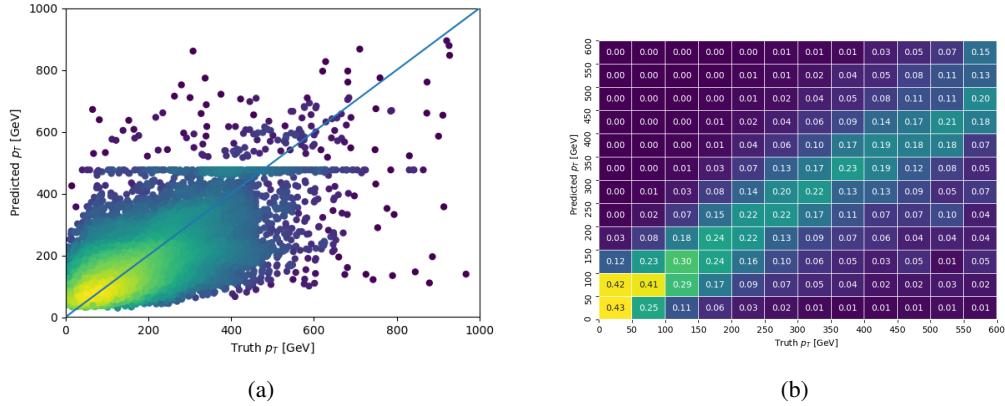


Figure 8.18:

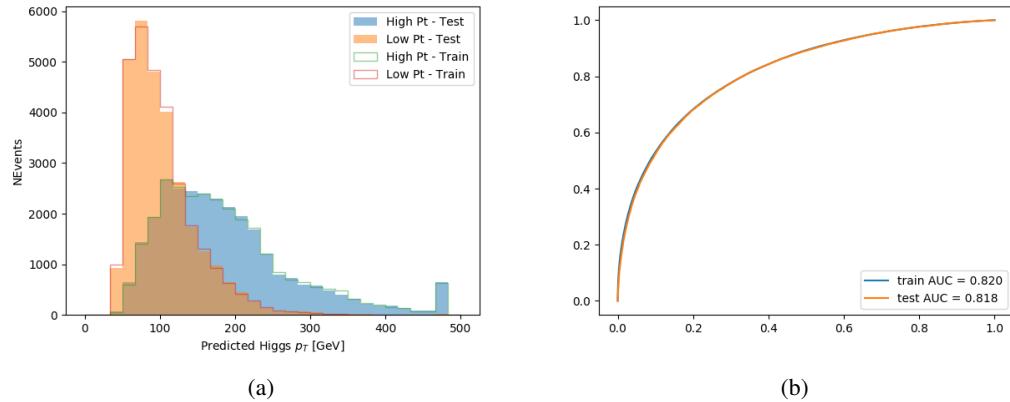


Figure 8.19:

### 590        8.4.3 3l Fully-leptonic Channel

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

594        The predicted transverse momentum, as a function of the truth  $p_T$ , is shown in figure ??.

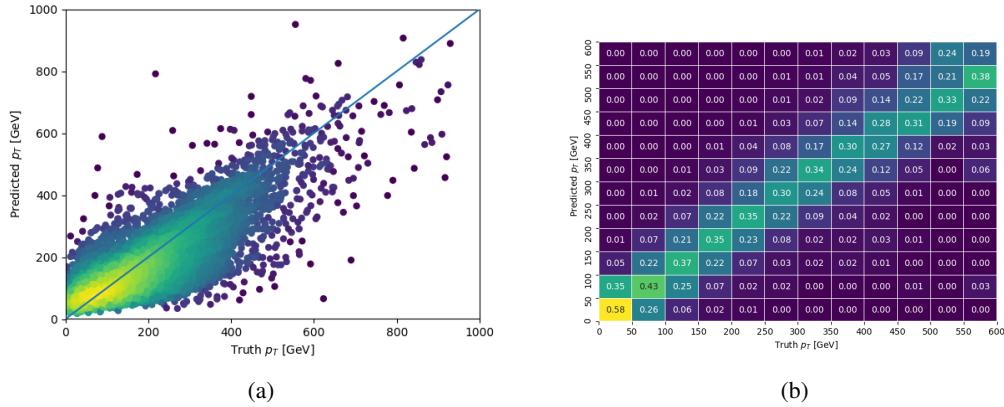


Figure 8.20:

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

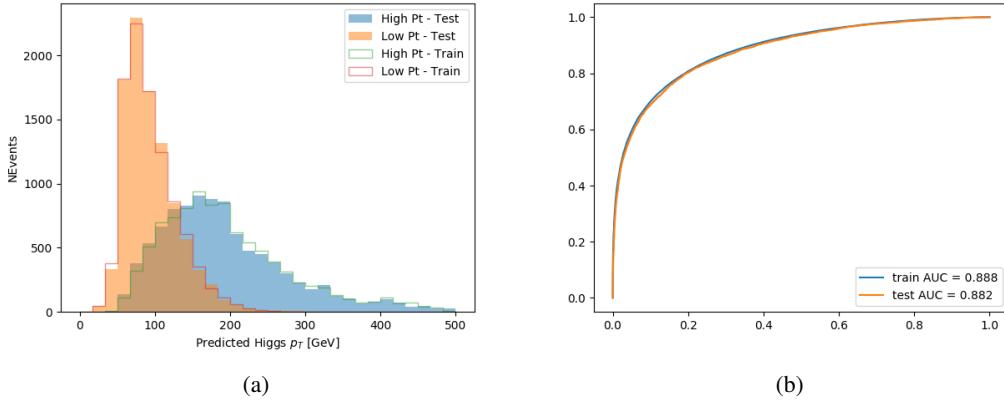


Figure 8.21:

596 **8.5 3l Decay Mode**

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

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

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

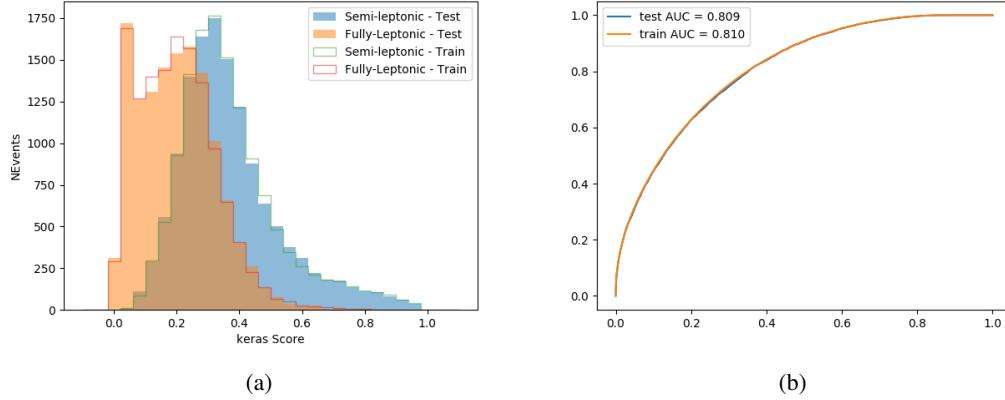


Figure 8.22:

## 608 9 Signal Region Definitions

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

### 612 9.1 Pre-MVA Event Selection

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

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

- 616 •  $>=4$  reconstructed jets,  $>=1$  b-tagged jets

- 617 • No reconstructed tau candidates

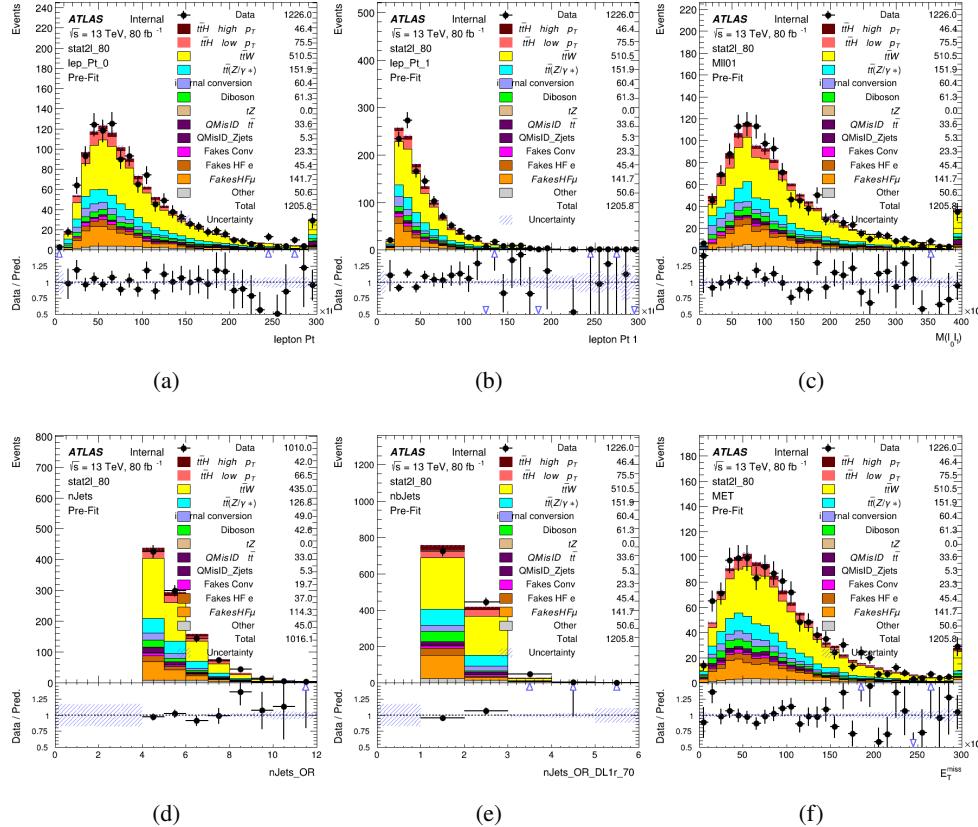


Figure 9.1:

- 618 For the 31 channel, the following selection is applied:

- 619 • Three light leptons with total charge  $\pm 1$   
 620 • Same charge leptons are required to be very tight, with  $p_T > 20$  GeV

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

- 622 •  $>=2$  reconstructed jets,  $>=1$  b-tagged jets

- 623 • No reconstructed tau candidates

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

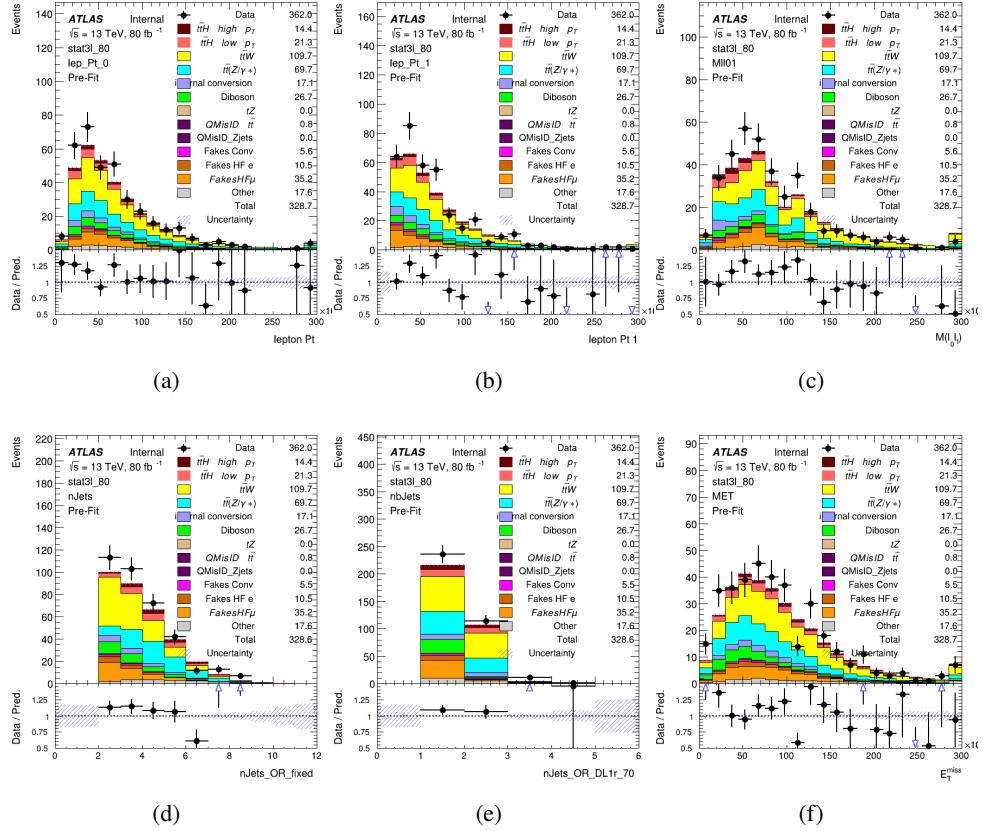


Figure 9.2:

625 **9.2 Event MVA**

626 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events  
627 from background for each analysis channel - 2lSS, 3l semi-leptonic, and 3l fully leptonic. In  
628 particular, Neural Networks produced with Tensorflow are trained using the kinematics of signal  
629 and background events derived from Monte Carlo simulations. Further, because the background  
630 composition differs for events with a high reconstructed Higgs  $p_T$  compared to events with low  
631 reconstructed Higgs  $p_T$ , separate MVAs are produced for high and low  $p_T$  regions.

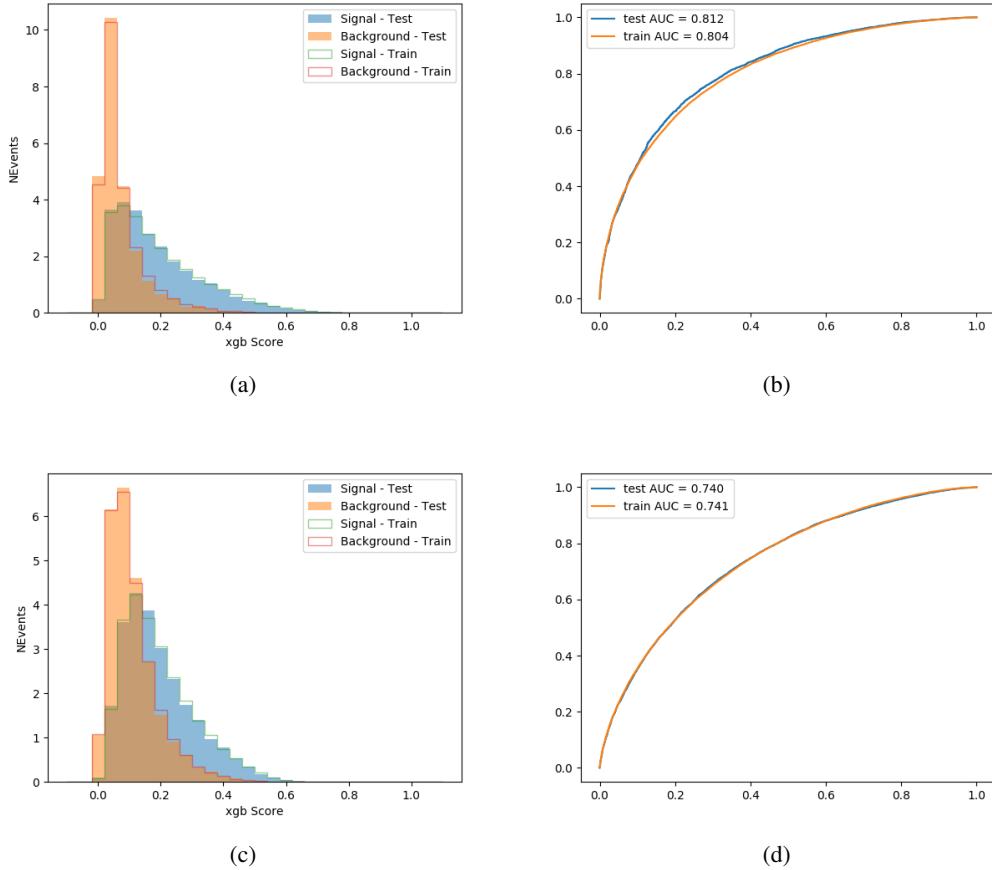


Figure 9.3:

632 Output distributions of each MVA are shown in figure 9.2.

### 633 9.3 Signal Region Definitions

634 Once pre-selection has been applied, channels are further refined based on the MVAs described

635 above. The output of the model described in section 8.5 is used to separate the three channel

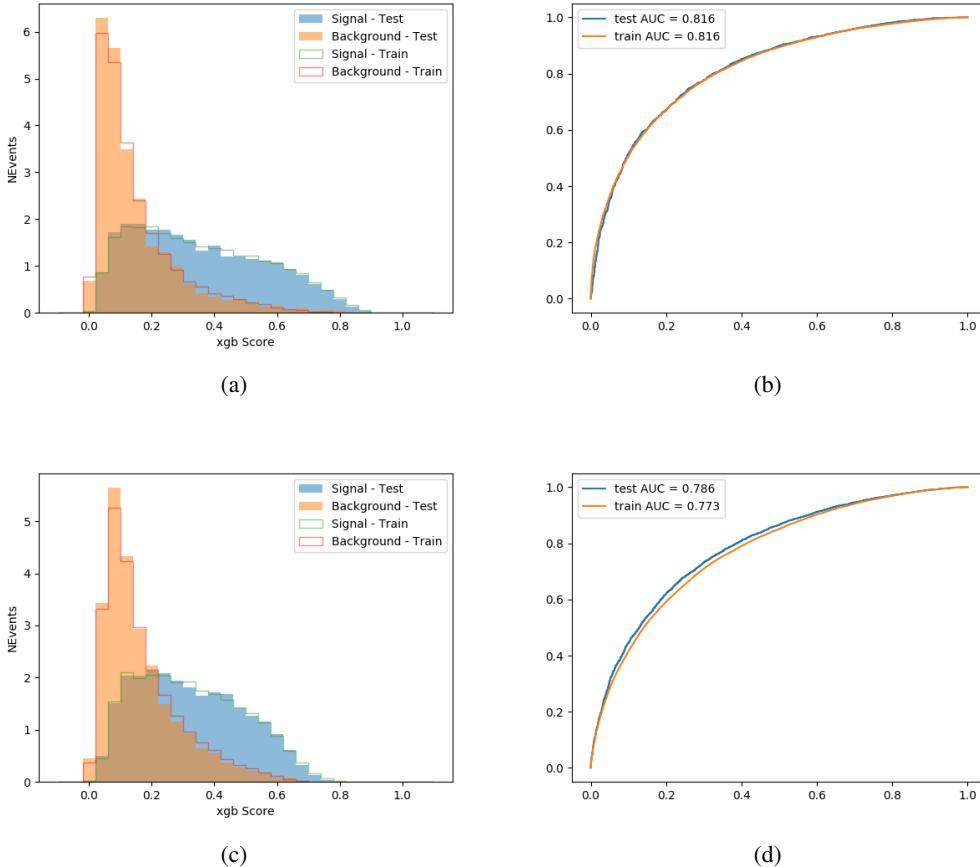


Figure 9.4:

636 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs  
 637 boson.

638 For each event, depending on the channel as well as the predicted  $p_T$  of the Higgs derived  
 639 from the algorithm described in section 8.4, a cut on the appropriate background rejection  
 640 algorithm is applied. The specific selection used, and the event yield in each channel after this  
 641 selection has been applied, is summarized below.

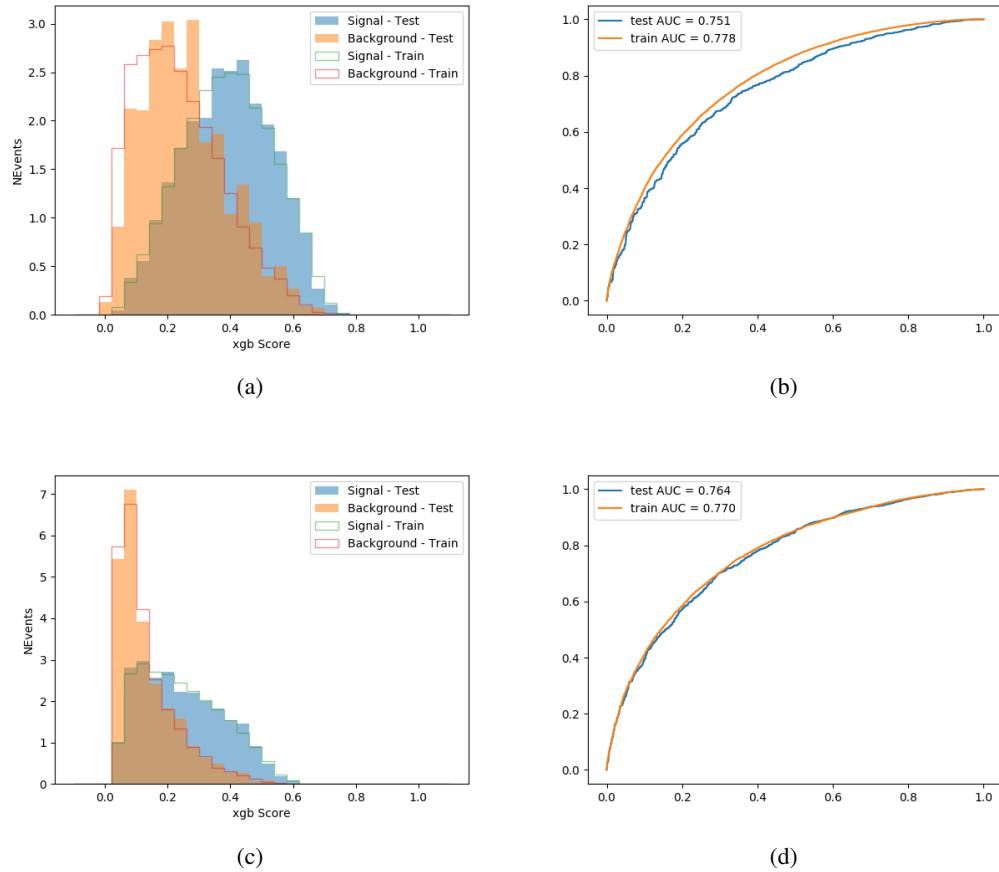


Figure 9.5:

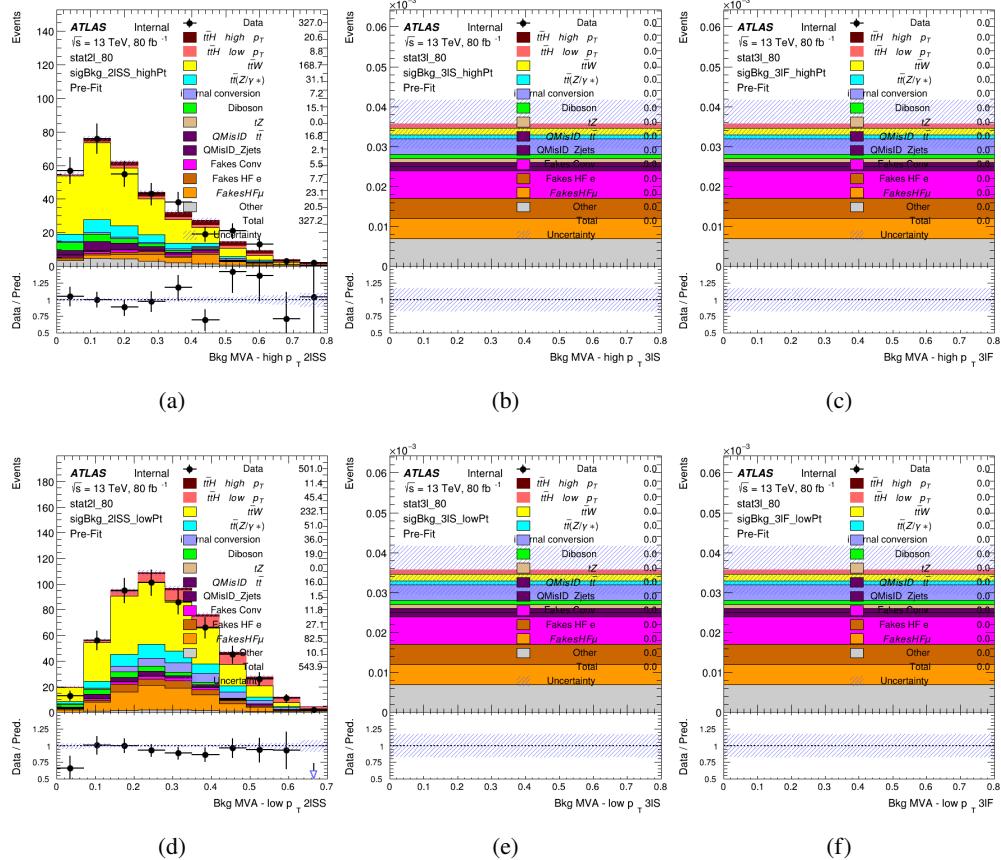


Figure 9.6: scores

642 **9.3.1 2lSS**643 **9.3.2 3l – Semi – leptonic**644 **9.3.3 3l – Fully – leptonic**645 **10 Background Rejection MVAs**

647 Separate models are used in order to distinguish signal events from background for each analysis  
 8th November 2020 – 21:42 62  
 648 channel - 2lSS, 3l semi-leptonic, and 3l fully leptonic. In particular, Neural Networks produced

with Tensorflow are trained using the kinematics of signal and background events derived from Monte Carlo simulations. Further, because the background composition differs for events with a high reconstructed Higgs  $p_T$  compared to events with low reconstructed Higgs  $p_T$ , separate MVAs are produced for high and low  $p_T$  regions.

### 10.1.1 2lSS - High $p_T$

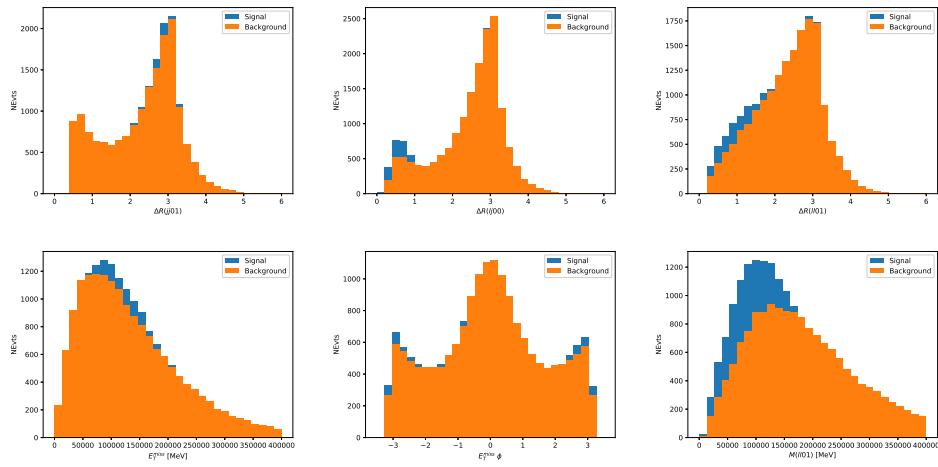


Figure 10.1:

654 **10.1.2 2lSS - Low  $p_T$**

655 **10.1.3 3l Semi-Leptonic - High  $p_T$**

656 **10.1.4 3l Semi-Leptonic - Low  $p_T$**

657 **10.1.5 3l Fully Leptonic - High  $p_T$**

658 **10.1.6 3l Fully Leptonic - Low  $p_T$**

## 659 **11 Systematic Uncertainties**

660 The systematic uncertainties that are considered are summarized in table ???. These are imple-  
661 mented in the fit either as a normalization factors or as a shape variation or both in the signal  
662 and background estimations. The numerical impact of each of these uncertainties is outlined in  
663 section 12.

664 The uncertainty in the combined 2015+2016 integrated luminosity is derived from a  
665 calibration of the luminosity scale using x-y beam-separation scans performed in August 2015  
666 and May 2016 [**lumi**].

667 The experimental uncertainties are related to the reconstruction and identification of light  
668 leptons and b-tagging of jets, and to the reconstruction of  $E_T^{\text{miss}}$ . The sources which contribute

Table 7: 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

669 to the uncertainty in the jet energy scale [`jes`] are decomposed into uncorrelated components and

670 treated as independent sources in the analysis.

671 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses

672 [`btag_cal`] are also decomposed into uncorrelated components. The large number of components

673 for b-tagging is due to the calibration of the distribution of the BDT discriminant.

674 The systematic uncertainties associated with the signal and background processes are

675 accounted for by varying the cross-section of each process within its uncertainty.

676 **12 Results**

677 A maximum likelihood fit is performed simultaneously over the regions described in section

678 ??.

679 **Part V**

680 **Conclusion**

681 As search for the effects of dimension-six operators on  $t\bar{t}H$  production is performed. An effective

682 field theory approached is used to parameterize the effects of high energy physics on the Higgs

683 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,

684 and the limits on dimension-six operators are limited to X.

685 **List of contributions**

686

## 687 Appendices

### 688 A Machine Learning Models

689 The following section provides details regarding various studies performed in support of this  
690 analysis, exploring alternate decisions and strategies.

#### 691 A.1 Alternate b-jet Identification Algorithm

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

699 For these studies, the kinematics of the 10 highest  $p_T$  jets in each event are used for  
700 training. This includes the vast majority of truth b-jets. Specifically the  $p_T$ ,  $\eta$ ,  $\phi$ ,  $E$ , and DL1r  
701 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.  
702 The  $p_T$ ,  $\eta$ ,  $\phi$ , and  $E$  of the leptons and  $E_T^{\text{miss}}$  are included as well.

703 Caterogical cross entropy is used as the loss function.

Table 8: 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		73.9%
3l		79.8%

704 **A.1.1 Binary Classification of the Higgs  $p_T$**

705 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer  
 706 resolution. This means separating high and low  $p_T$  events is sufficient for this analysis. As  
 707 such, rather than attempting the reconstruct the full Higgs  $p_T$  spectrum, a binary classification  
 708 approach is explored.

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

712 **A.1.2 Impact of Alternative Jet Selection**

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

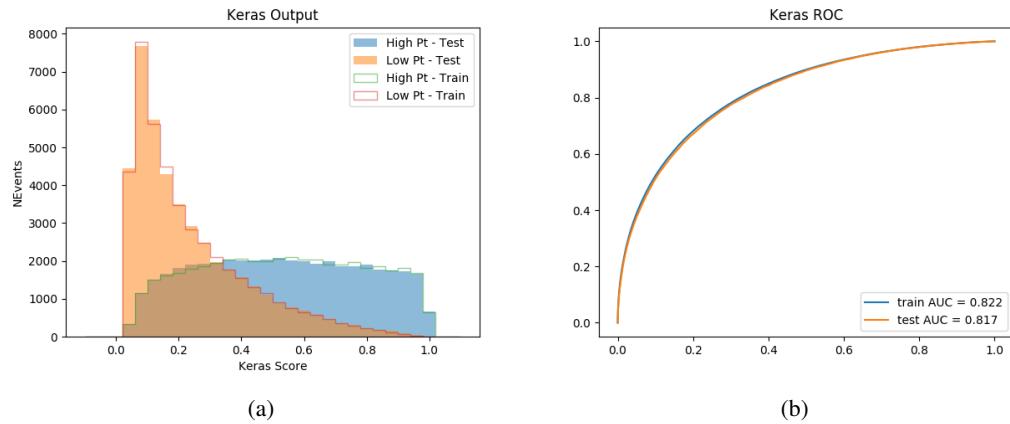


Figure A.1:

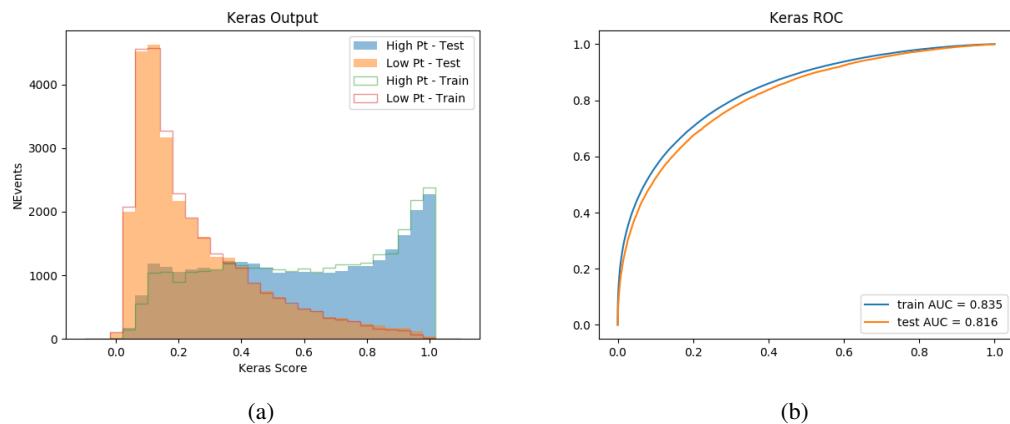


Figure A.2:

718 **B**

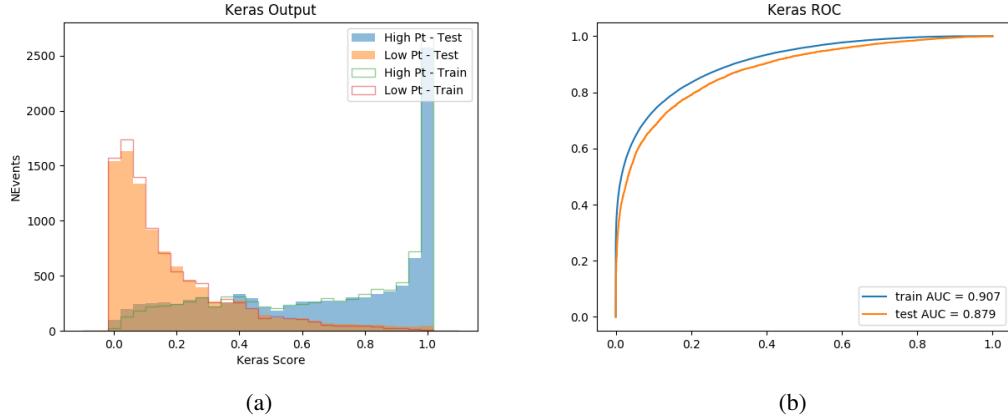


Figure A.3:

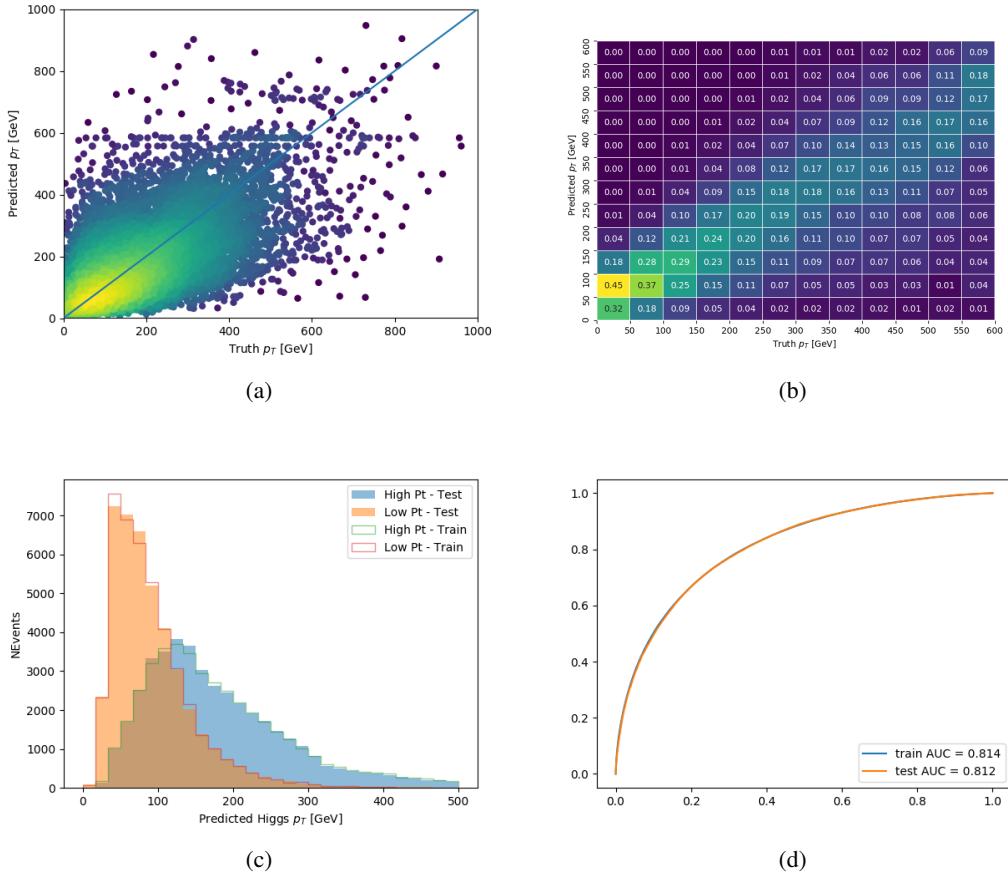


Figure A.4: