



**ATLAS Note**

GROUP-2017-XX



2nd June 2021

Draft version 0.1

1

2

**A Deep Learning Approach to Differential**

3

**Measurements of Higgs - Top Interactions in**

4

**Multilepton Final States using the ATLAS**

5

**Detector at the LHC**

6

The ATLAS Collaboration

7 Several theories Beyond the Standard Model predict a modification of the momentum spec-  
8 trum of the Higgs Boson, without a significantly altered rate of Higgs produced in association  
9 with top quark pairs ( $t\bar{t}H$ ). This provides a physical observable that can be used to search  
10 for new physics based on data collected by the LHC. This thesis presents techniques and  
11 preliminary results for a differential measurement of the Higgs transverse momentum in  $t\bar{t}H$   
12 events with multiple leptons in the final state, using data collected at an energy of  $\sqrt{s} = 13$   
13 TeV by the ATLAS detector at the LHC.

14 Because of the challenges inherent in reconstructing the Higgs in multilepton final states, a  
15 deep learning approach is used to predict of the Higgs. The regressed Higgs  $p_T$  spectrum is  
16 fit to data for events with two same-sign leptons and three leptons in the final state, in order  
17 to extract normalization factors for high ( $p_T(H) > 150$  GeV) and low ( $p_T(H) < 150$  GeV)  
18 momentum  $t\bar{t}H$  events. Preliminary results are presented for  $80 \text{ fb}^{-1}$  of data, with projected  
19 results shown for  $140 \text{ fb}^{-1}$ .

20 This thesis also details a measurement of  $WZ +$  heavy flavor production, a significant back-  
21 ground to  $t\bar{t}H$  that is poorly understood. This study targets events with three leptons and one  
22 or two jets in the final state, using  $140 \text{ fb}^{-1}$  of  $\sqrt{s} = 13$  TeV data. A measured cross-section  
23 of  $X \pm X \text{ fb}$  ( $X \pm X \text{ fb}$ ) is observed for  $WZ + b$  ( $WZ + \text{charm}$ ) with 1 associated jet and  $X \pm X$   
24  $\text{fb}$  ( $X \pm X \text{ fb}$ ) for  $WZ + b$  ( $WZ + \text{charm}$ ) with 2 assoicated jets.

© 2021 CERN for the benefit of the ATLAS Collaboration.

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

# 26 **Contents**

27	<b>I Introduction</b>	<b>6</b>
28	<b>1 Introduction</b>	<b>6</b>
29	<b>II Theoretical Motivation</b>	<b>9</b>
30	<b>2 The Standard Model and the Higgs Boson</b>	<b>9</b>
31	2.1 The Forces and Particles of the Standard Model	9
32	2.2 The Higgs Mechanism	12
33	2.2.1 The Higgs Field	12
34	2.2.2 Electroweak Symmetry Breaking	14
35	2.3 $t\bar{t}H$ Production	17
36	2.4 WZ + Heavy Flavor Production	20
37	2.5 Extensions to the Standard Model	21
38	<b>III The LHC and the ATLAS Detector</b>	<b>26</b>
39	<b>3 The LHC</b>	<b>26</b>
40	<b>4 The ATLAS Detector</b>	<b>29</b>
41	4.1 Inner Detector	30
42	4.2 Calorimeters	31
43	4.3 Muon Spectrometer	32
44	4.4 Trigger System	33
45	<b>IV Measurement of WZ + Heavy Flavor</b>	<b>35</b>
46	<b>5 Introduction</b>	<b>35</b>
47	<b>6 Data and Monte Carlo Samples</b>	<b>37</b>
48	6.1 Data Samples	37
49	6.2 Monte Carlo Samples	37
50	<b>7 Object Reconstruction</b>	<b>38</b>
51	<b>8 Event Selection and Signal Region Definitions</b>	<b>42</b>
52	8.1 Event Preselection	42
53	8.2 Fit Regions	43
54	8.3 Non-Prompt Lepton Estimation	45

---

55	<b>9 tZ Separation Multivariate Analysis</b>	<b>47</b>
56	<b>10 Systematic Uncertainties</b>	<b>49</b>
57	<b>11 Results</b>	<b>53</b>
58	11.1 1-jet Fit Results	54
59	11.2 2-jet Fit Results	63
60	<b>V Differential Studies of t̄tH Multilepton</b>	<b>72</b>
61	<b>12 Data and Monte Carlo Samples</b>	<b>72</b>
62	12.1 Data Samples	72
63	12.2 Monte Carlo Samples	72
64	<b>13 Object Reconstruction</b>	<b>75</b>
65	<b>14 Higgs Momentum Reconstruction</b>	<b>77</b>
66	14.1 Physics Object Truth Matching	79
67	14.2 b-jet Identification	80
68	14.2.1 2lSS Channel	81
69	14.2.2 3l Channel	86
70	14.3 Higgs Reconstruction	89
71	14.3.1 2lSS Channel	90
72	14.3.2 3l Semi-leptonic Channel	94
73	14.3.3 3l Fully-leptonic Channel	98
74	14.4 p <sub>T</sub> Prediction	102
75	14.4.1 2lSS Channel	103
76	14.4.2 3l Semi-leptonic Channel	106
77	14.4.3 3l Fully-leptonic Channel	108
78	14.5 3l Decay Mode	110
79	<b>15 Signal Region Definitions</b>	<b>112</b>
80	15.1 Pre-MVA Event Selection	112
81	15.2 Event MVA	116
82	15.3 Signal Region Definitions	123
83	<b>16 Systematic Uncertainties</b>	<b>124</b>
84	<b>17 Results</b>	<b>128</b>
85	17.1 Results - 79.8 fb <sup>-1</sup>	129
86	17.2 Projected Results - 139 fb <sup>-1</sup>	133

---

87	<b>VI Conclusion</b>	<b>138</b>
88	<b>VII Appendices</b>	<b>147</b>
89	<b>Appendices</b>	<b>147</b>
90	<b>A Non-prompt lepton MVA</b>	<b>147</b>
91	<b>B Supplementary WZ + Heavy Flavor Studies</b>	<b>152</b>
92	B.1 Non-prompt CR Modelling	152
93	B.2 tZ Interference Studies	157
94	B.3 Alternate tZ Inclusive Fit	159
95	B.3.1 tZ Inclusive Fit	160
96	B.3.2 Floating tZ	161
97	<b>C Supplementary t̄H Differential Analysis Studies</b>	<b>162</b>
98	C.1 Higgs Reconstruction Model Details	162
99	C.1.1 b-jet Identification Features - 2lSS	162
100	C.1.2 b-jet Identification Features - 3l	166
101	C.1.3 Higgs Reconstruction Features - 2lSS	171
102	C.1.4 Higgs Reconstruction Features - 3lS	176
103	C.1.5 Higgs Reconstruction Features - 3lF	181
104	C.2 Background Rejection MVA Details	185
105	C.2.1 Background Rejection MVA Features - 2lSS	185
106	C.2.2 Background Rejection MVA Features - 3l	189
107	C.3 Truth Level Studies	194
108	C.4 Alternate b-jet Identification Algorithm	195
109	C.5 Binary Classification of the Higgs $p_T$	196
110	C.6 Impact of Alternative Jet Selection	198

---

# 111 Part I

## 112 Introduction

### 113 1 Introduction

114 Particle physics is an attempt to describe the fundamental building blocks of the universe and  
115 their interactions. The Standard Model (SM) - our best current theory of fundamental particle  
116 physics - does a remarkable job of that. All known fundamental particles and (almost) all of the  
117 forces underlying their interactions can be explained by the SM, and the predictions from this  
118 theory agree with experiment to an incredibly precise degree. This is especially true since the  
119 Higgs Boson, the last piece of the SM predicted decades before, was finally discovered at the  
120 Large Hadron Collider (LHC) in 2012 [1].

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

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

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

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

141 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-  
142 2018 provides the opportunity to make this measurement for the first time. The unprecedented  
143 energy achieved by the LHC during this period greatly increase the rate at which  $t\bar{t}H$  events are  
144 produced, and the large amount of data collected provides the necessary statistics for a differential  
145 measurement to be performed.

146 A study of  $t\bar{t}H$  events with multiple leptons in the final state is performed, using  $139 \text{ fb}^{-1}$   
147 of data from proton-proton collisions at an energy  $\sqrt{s} = 13 \text{ TeV}$  collected by the ATLAS detector  
148 from 2015-2018. Events are separated into channels based on the number of light leptons in the

149 final state - either two same-sign leptons, or three leptons. A deep neural network is used to  
150 reconstruct the momentum of the Higgs Boson in each event. This momentum spectrum is fit to  
151 data for each analysis channel in order to search for evidence of these BSM effects.

152 An additional study of WZ produced in association with a heavy flavor jet (including both  
153 b-jets and charm jets) is also included. This process mimics the final state of  $t\bar{t}H$  multilpjet  
154 events, making it an irreducible background for that analysis. However, this process is poorly  
155 understood, and difficult to simulate accurately, introducing large systematic uncertainties for  
156 analyses that include it as a background. A measurement of WZ + heavy flavor in the fully  
157 leptonic decay mode is performed in an attempt to reduce this uncertainty.

158 This dissertation begins with a brief explanation of the SM, its limitations, and the theor-  
159 etical motivation behind this work in Part II. This is followed by a description of the LHC and  
160 the ATLAS detector in Part III. Part IV details a measurement of WZ + heavy flavor. Studies  
161 of differential measurements of  $t\bar{t}H$  are then described in Part V, and preliminary results are  
162 presented. Finally, the results of these studies are summarized in the conclusion, Part VI.

---

**Part II****Theoretical Motivation****2 The Standard Model and the Higgs Boson**

The Standard Model of particle physics (SM) is a Quantum Field Theory (QFT) describing the known fundamental particles and their interactions. It accounts for three of the four known fundamental force - electromagnetism, the weak nuclear force, and the strong nuclear force, but not gravity. Further, the SM describes a mechanism for combining the weak and electromagnetic forces into a singular interaction, known as the electroweak force. It is a non-Abelian gauge theory, invariant under the Lie Group  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ , where C refers to color charge, L, the helicity of the particle, and Y, the hypercharge.

**2.1 The Forces and Particles of the Standard Model**

The SM particles, summarized in Figure 2.1, can be classified into two general categories based on their spin: fermions, and bosons.

## Standard Model of Elementary Particles

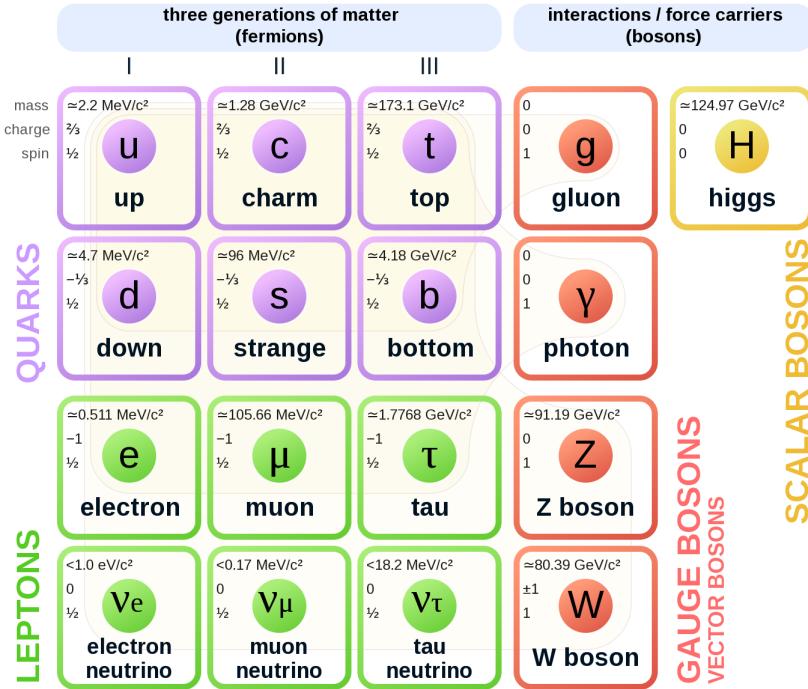


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. [2]

176 Fermions are particles with  $\frac{1}{2}$ -integer spin, which according to the spin-statistics theorem,  
 177 causes them to comply with the Pauli-exclusion principle. They can be separated into two groups,  
 178 leptons and quarks, each of which consist of three generations of particles with increasing mass.

179 Leptons are fermions which interact via the electroweak force, but not the strong force.  
 180 The three generation of leptons consist of the electron and electron neutrino, the muon and muon  
 181 neutrino, the tau and tau neutrino. The quarks, by contrast, do interact via the strong force - which  
 182 is to say they have color charge - in addition to the electroweak force. The three generations  
 183 include the up and down quarks, the strange and charm quarks, and the top and bottom quarks.

<sub>184</sub>      Each of these generations form left-handed doublets invariant under SU(2) transfor-

<sub>185</sub>      mations. For the leptons these doublets are:

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

<sub>186</sub>      And for the quarks:

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

<sub>187</sub>      For both leptons and quarks, the heavier generations can decay into the lighter generation

<sub>188</sub>      of particles, while the first generation does not decay. Hence, ordinary matter generally consists

<sub>189</sub>      of this first generation of fermions - electrons, up quarks, and down quarks. Each of these

<sub>190</sub>      fermions has a corresponding anti-particle, which has an equal mass as its partner but opposite

<sub>191</sub>      charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos,

<sub>192</sub>      whose mass has been experimentally confirmed but is not accounted for in the SM.

<sub>193</sub>      Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-

<sub>194</sub>      exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1

<sub>195</sub>      particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0,

<sub>196</sub>      particle - the Higgs Boson. Of the gauge bosons, the  $W^+$ ,  $W^-$  and  $Z$  bosons - which are the

<sup>197</sup> mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon

<sup>198</sup> mediates the electric force, and the gluon mediates the strong force.

<sup>199</sup> **2.2 The Higgs Mechanism**

<sup>200</sup> A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to

<sup>201</sup> the Lagrangian giving mass to the the gauge bosons would violate this underlying symmetry of

<sup>202</sup> the theory. This presents a clear problem with the theory: The experimental observation that the

<sup>203</sup> W and Z bosons have mass seems to contradict the basic structure of the SM.

<sup>204</sup> Rather than abandoning gauge invariance, an alternative way for particles to acquire mass

<sup>205</sup> beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout

<sup>206</sup> in 1964 [3]. This procedure for introducing masses for the gauge bosons while preserving local

<sup>207</sup> gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory

<sup>208</sup> by Weinberg in 1967 [4].

<sup>209</sup> **2.2.1 The Higgs Field**

<sup>210</sup> The Higgs mechanism introduces a complex scalar SU(2) doublet,  $\Phi$ , with the form:

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

211 This field introduces a scalar potential to the Lagrangian of the form:

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

212 Where  $\mu$  and  $\lambda$  are free parameters of the new field. This represents the most general  
 213 potential allowed while preserving  $SU(2)_L$  invariance and renormalizability. In the case that  
 214  $\mu^2 < 0$ , this potential takes the form shown in Figure 2.2.

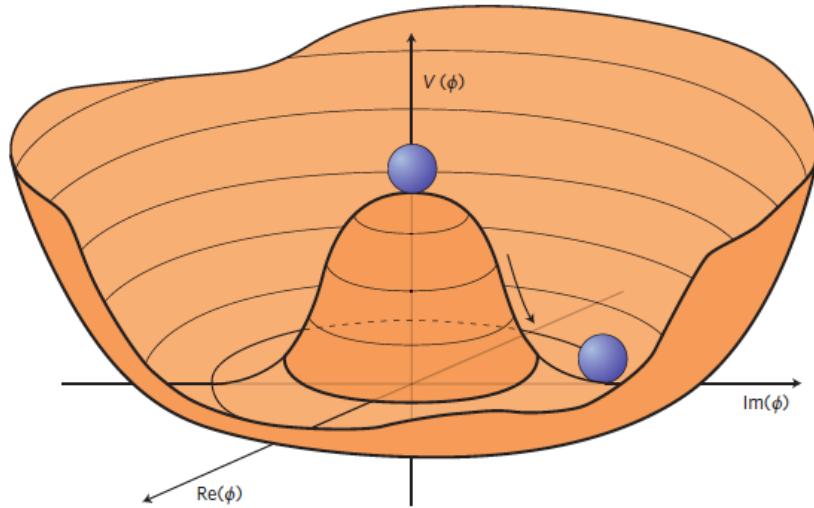


Figure 2.2: The value of the Higgs potential,  $V(\Phi)$  as a function of  $\Phi$ , for the case that  $\mu^2 < 0$  [5].

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

<sup>219</sup> the minimum is determined only by  $\Phi^\dagger \Phi$ , there is some ambiguity in the particular definition of

<sup>220</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>221</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>222</sup> with  $v$  being the value of the VEV, and  $H$  being the real value of the scalar field.

### <sup>223</sup> 2.2.2 Electroweak Symmetry Breaking

<sup>224</sup> The Electroweak (EWK) interaction is described in the SM by a  $SU(2)_L \otimes U(1)_Y$  gauge theory.

<sup>225</sup> 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,

<sup>226</sup>  $B_\mu$ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar

<sup>227</sup> field  $\Phi$  in the Lagrangian:

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

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

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

232 Expanding these terms into the mass eigenstates of the electroweak interaction yields four  
 233 physical gauge bosons, two charged and two neutral, which are linear combinations of the fields  
 234  $W_\mu^1$ ,  $W_\mu^2$ ,  $W_\mu^3$ , and  $B_\mu$ :

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

235 And the masses of these fields are given by:

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

236 This produces exactly the particles we observe - three massive gauge bosons and a single  
 237 massless photon. The massless photon represents the portion of the gauge symmetry, a single  
 238  $U(1)$  of the electromagnetic force, that remains unbroken by the VEV.

239 Interactions with the Higgs field also lead to the generation of the fermion masses, which  
 240 in the Lagrangian take the form:

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

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

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

248 The fermions' Yukawa couplings to the VEV take the same form as the fermions' coupling  
 249 to the Higgs boson -  $\lambda_\psi$ . Therefore, the strength of a fermion's interaction with the Higgs is  
 250 directly proportional to its mass. We now have a model that predicts a Higgs boson with mass  
 251  $M_H = \mu$ , which interacts with the fermions with coupling strength  $\lambda_\psi$ . Because  $\mu$  and  $\lambda_\psi$  are

252 free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions

253 must be measured experimentally.

254 **2.3  $t\bar{t}H$  Production**

255 The strength of a particles interaction with the Higgs, given by its Yukawa coupling, is propor-  
 256 tionate to its mass. The top quark - as the heaviest known particle - has the strongest interaction,  
 257 making this interaction particularly interesting to study. While several processes involve interac-  
 258 tions between the Higgs and the top, some Higgs production modes include the top interaction  
 259 only as a part of a loop diagram, such as the gluon-gluon fusion diagram shown in Figure 2.3.

260 This process therefore only allows for an indirect probe of the Higgs-top Yukawa coupling, as  
 261 the flavor of the quark in this diagram is not unique.

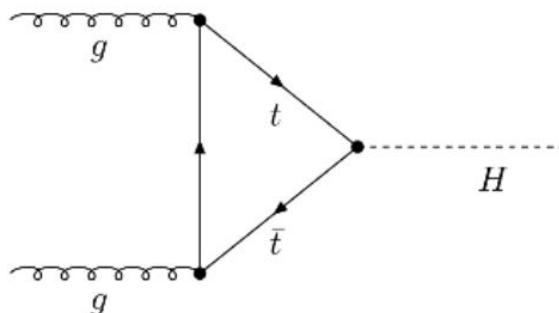


Figure 2.3: Diagram of a Higgs boson produced via gluon-gluon fusion.

262 Studying the Higgs produced in association with top quark pairs,  $t\bar{t}H$ , allows this interac-  
 263 tion to be measured directly. This process, as shown in Figure 2.4, involves a unique coupling  
 264 between the Higgs and the top, which can be identified by the top quark pair in the final state.

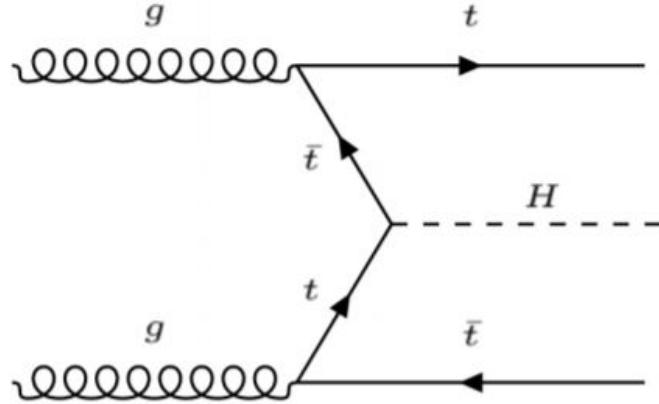


Figure 2.4: Diagram of a Higgs boson produced in association with a pair of top quarks.

265        The Higgs boson, as well as the top quarks, have very short lifetimes - on the order of  
 266         $10^{-22}$  s and  $10^{-25}$  s respectively - meaning they can only be observed via their decay products.  
 267        Measuring this process is therefore a matter of identifying events with final states consistent with  
 268         $t\bar{t}H$  production.

269        Studies of  $t\bar{t}H$  production have been reported by the ATLAS collaboration for  $H \rightarrow b\bar{b}$ ,  
 270         $H \rightarrow \gamma\gamma$  and multilepton (encompassing  $H \rightarrow W^+W^-$ ,  $H \rightarrow ZZ$  and  $H \rightarrow \tau^-\tau^+$ , with  
 271         $H \rightarrow ZZ \rightarrow 4l$  as a separate analysis) decay modes. While the branching ratio of  $H \rightarrow W^+W^-$   
 272        is smaller than  $H \rightarrow b\bar{b}$  (see Table 2.3), it produces a clearer signal, as  $H \rightarrow b\bar{b}$  suffers from  
 273        large  $t\bar{t}$  backgrounds. On the other hand,  $H \rightarrow \gamma\gamma$  produces the most easily identifiable signal,  
 274        but has a much smaller branching ratio than  $H \rightarrow W^+W^-$ . Therefore, compared with other final  
 275        states of  $t\bar{t}H$ , the  $t\bar{t}H$  – ML channel is an attractive candidate for study, as it involves a good  
 276        balance between statistical power and identifiability.

Decay Mode	Branching Ratio (%)
$H \rightarrow b\bar{b}$	58.2
$H \rightarrow WW^*$	21.4
$H \rightarrow gg$	8.19
$H \rightarrow \tau\tau$	6.27
$H \rightarrow c\bar{c}$	2.89
$H \rightarrow ZZ^*$	2.62
$H \rightarrow \gamma\gamma$	0.227

Table 1: Summary of the predominant SM Higgs ( $m_H = 125$  GeV) branching ratios. Particles with a star imply off-shell decays.

277        Searches for  $t\bar{t}H$  production typically target a measurement of the signal strength para-  
 278        meter,  $\mu_{t\bar{t}H}$ , which measures the ratio of the observed cross-section and the expected cross-section  
 279        according to the SM.

$$\mu_{t\bar{t}H} = \frac{\sigma_{t\bar{t}H}^{\text{obs.}}}{\sigma_{t\bar{t}H}^{\text{SM}}} \quad (2.12)$$

280         $t\bar{t}H$  production was observed by ATLAS using up to  $79.8 \text{ fb}^{-1}$  of data collected at  $\sqrt{s}$   
 281         $= 13 \text{ TeV}$ , based on a combination of five Higgs decay modes:  $b\bar{b}$ ,  $WW^*$ ,  $\tau^-\tau^+$ ,  $\gamma\gamma$ , and  $ZZ^*$   
 282        [6]. A significance of  $5.8\sigma$  was observed, compared to a  $4.9\sigma$  expected significance. Since then,  
 283        two analyses have published updated results ( $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$ ) with the full Run 2  
 284        dataset, representing  $139 \text{ fb}^{-1}$ . Studies are still ongoing in the remaining channels.

285        This thesis focuses on  $t\bar{t}H$  events with multiple leptons in the final state,  $t\bar{t}H - ML$ ,  
 286        specifically targeting events with two same-sign leptons (2lSS) or three leptons (3l) in the

287 final state. This includes  $H \rightarrow W^+W^-$  events, where at least one of the  $W$  bosons decays  
 288 leptonically.

289 **2.4 WZ + Heavy Flavor Production**

290 Part IV is dedicated to a measurement of  $WZ$  produced in association with a heavy flavor jet  
 291 - namely, a charm or  $b$ -jet - in the fully leptonic channel. In the instance that both the  $W$   
 292 and  $Z$  bosons decay leptonically, this process produces a final state similar to  $t\bar{t}H$ , making it  
 293 an irreducible background for  $t\bar{t}H$  – ML specifically, and any analysis that includes multiple  
 294 leptons and  $b$ -tagged jets in the final state more broadly.

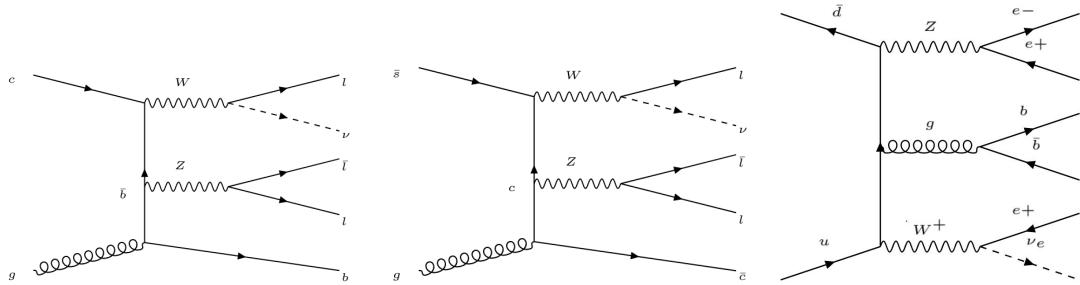


Figure 2.5: Example Feynman diagrams of  $WZ$  + heavy flavor production

295 The  $b$ -jets produced in this process can be thought of in two different ways: either as  
 296 originating from the quark “sea” of the initial state hadrons, or as the result of a gluon from  
 297 one the colliding protons splitting into  $b\bar{b}$  pairs. However, the heavy flavor contribution to the  
 298 parton distribution function (PDF) of the proton is uncertain, and simulations of this process  
 299 disagree depending on which of these two approaches one considers. This makes  $WZ$  + heavy

300 flavor difficult to accurately simulate, and introduces a large uncertainty for any analysis which  
301 includes it as a background, motivating a measurement of this process.

302 **2.5 Extensions to the Standard Model**

303 While the SM has been tested to great precision, particularly at the LHC, it is generally accepted  
304 that it is only valid up to a certain energy scale. It is assumed that above a certain energy, at the  
305 scale where something like a Grand Unified Theory (GUT) or quantum gravity become relevant,  
306 the SM will not be applicable. Further, there are several experimental observations that the SM  
307 fails to explain. For example, the SM predicts neutrinos to be massless, despite experimental  
308 observation to the contrary, and fails to explain the observation of dark matter and dark energy.

309 Another example, relevant to the Higgs sector, is known as the hierarchy problem: large  
310 quantum corrections to the Higgs mass from loop diagrams, such as those shown in Figure 2.6,  
311 are many orders of magnitude larger than the Higgs mass itself. The observed value of the Higgs  
312 mass therefore requires extremely precise cancellation between these corrections and the bare  
313 mass of the Higgs, a cancellation which seems unnatural and suggests something missing in our  
314 theoretical picture.

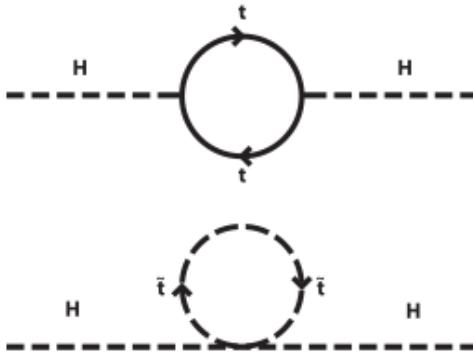


Figure 2.6: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram.

Because so many of the properties of the Higgs boson have not yet been studied, its interactions are a promising place to search for new physics that could resolve some of the limitations of the SM. As explained above, the interactions of the Higgs with the top quark, as in  $t\bar{t}H$  production, are particularly interesting: As the most massive particle in the Standard Model, the top quark is the most strongly interacting with the Higgs. Therefore, any new physics effects are likely to be seen most prominently in this interaction.

These interactions can be measured directly by studying the production of a Higgs Boson in association with a pair of Top Quarks ( $t\bar{t}H$ ). While this process has been observed by both the ATLAS [7] and CMS [8] collaborations, these analyses have focused on measuring the overall rate of  $t\bar{t}H$  production. There are several theories of physics Beyond the Standard Model (BSM), however, that would affect the kinematics of  $t\bar{t}H$  production without altering its overall rate [9].

327 An Effective Field Theory approach can be used to model the low energy effects of new,  
 328 high energy physics, by parameterizing BSM effects as higher dimensional operators. These  
 329 additional operators can then be added to the SM Lagrangian to write an effective Lagrangian  
 330 that accounts for the effects of these higher energy physics. The lowest order of these that could  
 331 contribute to Higgs-top couplings are dimension-six, as represented in Equation 2.13.

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \frac{f}{\Lambda} O^6 \quad (2.13)$$

332 Here  $\Lambda$  represents the energy scale of the new physics, and  $f$  is a Wilson coefficient which  
 333 represents the strength of the effective coupling. An experimental observation of any non-zero  
 334 value of  $f$  would be a sign of BSM physics.

335 The addition of these operators can be shown to modify the transverse momentum ( $p_T$ )  
 336 spectrum of the Higgs Boson in Higg-top interactions, without effecting the overall rate of  $t\bar{t}H$   
 337 production [10]. The possible impact of these higher order effects on the Higgs  $p_T$  spectrum are  
 338 shown in Figure 2.7.

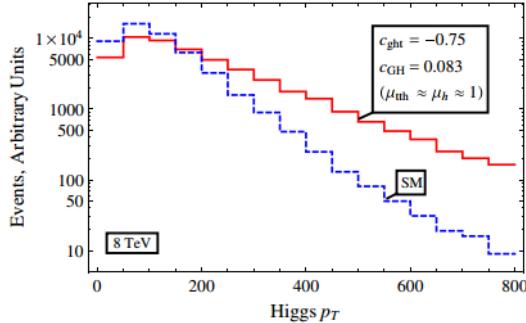


Figure 2.7: The momentum spectrum of the Higgs boson produced via top-quarks with (red) and without (blue) the presence of dimension-six operators.

339        This provides a clear, physics observable that could be used to search for evidence of  
 340        BSM physics. The energy and luminosity produced by the LHC now make such a measurement  
 341        possible. Reconstructing the momentum spectrum of the Higgs in  $t\bar{t}H$  events therefore provides  
 342        a means to search for new physics in the Higgs sector.

343        Reconstructing the Higgs is a particular challenge in the multilepton channels of  $t\bar{t}H$ , due  
 344        to an ambiguity arising from multiple sources of missing energy. In the  $H \rightarrow \gamma\gamma$  channel, the  
 345        kinematics of the Higgs can be fully reconstructed from the two photons. The same is true of  
 346         $H \rightarrow b\bar{b}$ , though with the additional challenge of identifying which two of the four b-quarks in  
 347        the final state originated from the Higgs. By contrast, the two channels (2lSS and 3l) targeted  
 348        by this analysis include at least one neutrino originating from the Higgs decay.

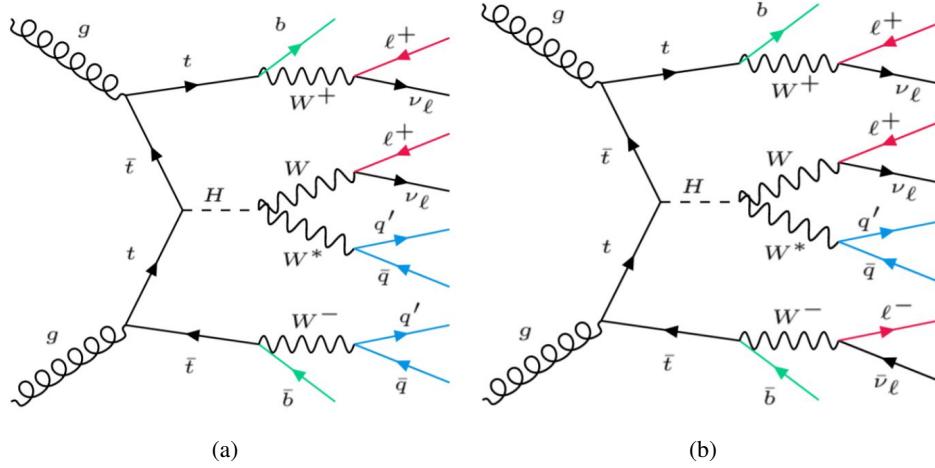


Figure 2.8: Feynman diagrams of  $t\bar{t}H$  production with (a) two same-sign leptons and (b) three leptons in the final state.

349            Neutrinos are not detected by ATLAS; instead, their presence is inferred from missing  
 350            transverse energy in the detector,  $E_{\text{miss}}^T$ . The two channels targeted here include not just a  
 351            neutrino from the Higgs decay, but at least one additional neutrino from the decay of the top-  
 352            quarks. This makes disentangling the contribution of the Higgs decay to  $E_{\text{miss}}^T$ , and thereby fully  
 353            reconstructing the Higgs, impossible.

354            This challenge motivates the use of more sophisticated machine learning techniques when  
 355            attempting to perform differential measurements of the Higgs  $p_T$  spectrum in the multi-lepton  
 356            channels of  $t\bar{t}H$ .

---

# 357 Part III

## 358 The LHC and the ATLAS Detector

### 359 3 The LHC

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

369 The LHC first began running in 2009 at a proton-proton center of mass energy of  $\sqrt{s} = 8$   
370 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during  
371 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,  
372 and collected data at an increased energy of  $\sqrt{s} = 13$  TeV until 2018, a period referred to as Run  
373 2.

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

375 higher energies until they are injected into the main ring. This process is summarized in figure  
 376 [3.1](#). Protons extracted from a tank of ionized hydrogen are fed into a linear accelerator, LINAC2,  
 377 where they reach an energy of 50 MeV. From there, they enter a series of three separate circular  
 378 accelerators, before being injected into the main accelerator ring at an energy of 450 GeV. Within  
 379 the main ring protons are separated into two separate beams moving in opposite directions,  
 380 and their energy is increased to their full collision energy. Radiofrequency cavities are used to  
 381 accelerate these particles and sort them into bunches. From 2015-2018, these bunches consisted  
 382 of around 100 billion protons each with an energy of 6.5 TeV per proton, which collided at a rate  
 383 of 40 MHz, or every 25 ns.

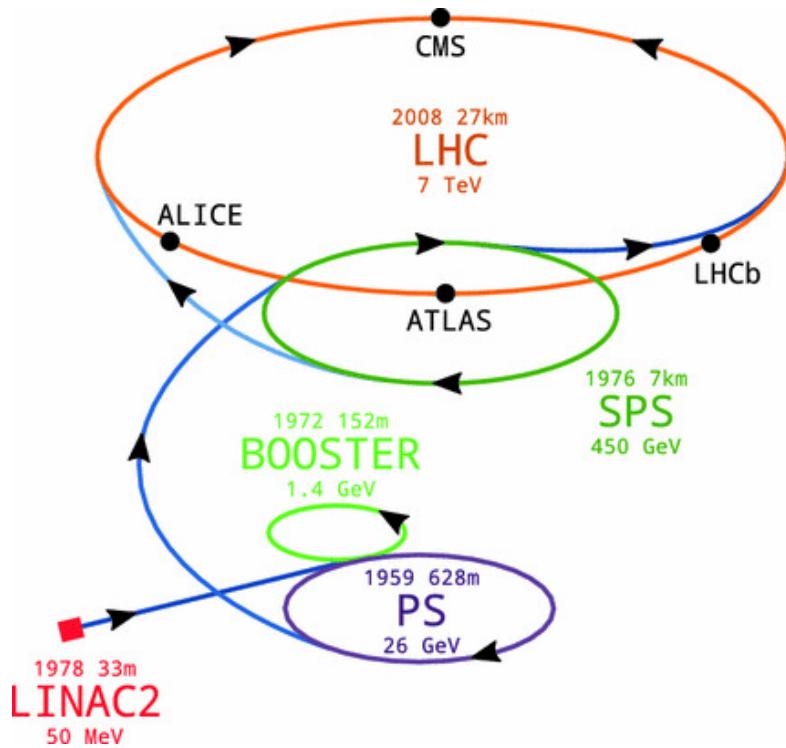


Figure 3.1: A summary of the accelerator chain used to feed protons into the LHC [11].

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

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

$$\mathcal{L} = \frac{1}{\sigma} \frac{dN}{dt} \quad (3.1)$$

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

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} dt \quad (3.2)$$

393 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around  
 394  $140 \text{ fb}^{-1}$ , exceeding the expected integrated luminosity of  $100 \text{ fb}^{-1}$ .

## 395 4 The ATLAS Detector

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

401 The ATLAS detector consists of multiple concentric layers, each of which serves a different  
 402 purpose in reconstructing collisions. At the very center of the detector is the interaction point  
 403 where the proton beams of the LHC collide.

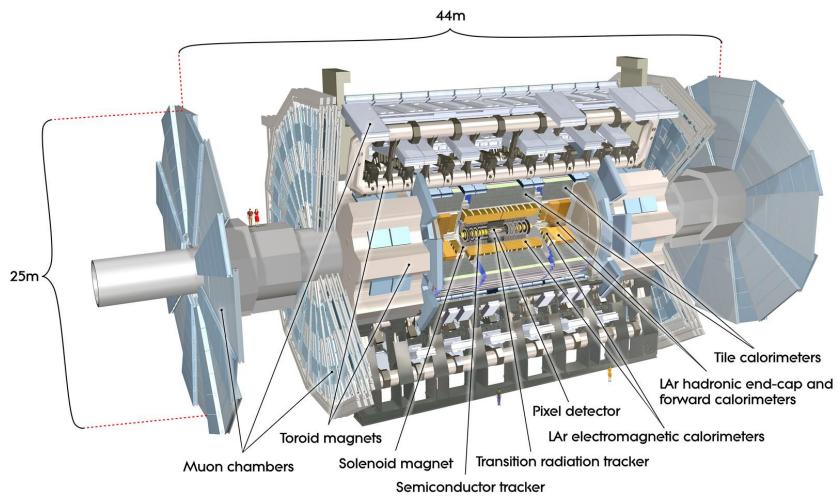


Figure 4.1: Cutaway view of the ATLAS detector, with labels of its major components [12].

404 **4.1 Inner Detector**

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

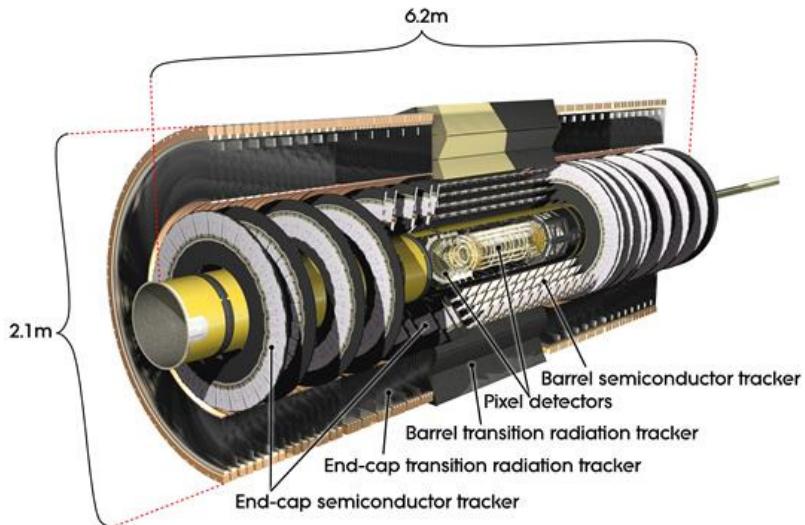


Figure 4.2: Cutaway view of the Inner Detector [13].

411 The Inner Detector consists of three components - the Pixel Detector, the Semi-Conductor  
 412 Tracker (SCT), and the Transition Radiation Tracker (TRT). The Pixel Detector is the innermost  
 413 of these, beginning just 33.25 mm away from the beam line. It consists of three silicon layers  
 414 along the barrel, as well as three endcap layers, covering a range of  $|\eta| < 2.5$ .

415        The Semiconductor Tracker (SCT) is similar to the Pixel detector, but uses long strips of  
 416        silicon rather than small pixels to cover a larger spatial area. It includes over 6 million readout  
 417        strips, allowing the position of charged particles to be measured to an accuracy of 17  $\mu\text{m}$ .

418        The outermost component of the inner detector, the TRT consists of around 300,000 straw  
 419        tubes filled with ionizable gas, which produces current through a wire in the center of each tube  
 420        when a charged particle passes through. Between these staws are layers of material designed  
 421        to produce transition radiation from ultrarelativistic particles as they pass through each material  
 422        boundary, amplifying the signal. The position uncertainty in the TRT is higher than the other  
 423        two, on the order of 200  $\mu\text{m}$ , but covering a much larger area.

## 424        4.2 Calorimeters

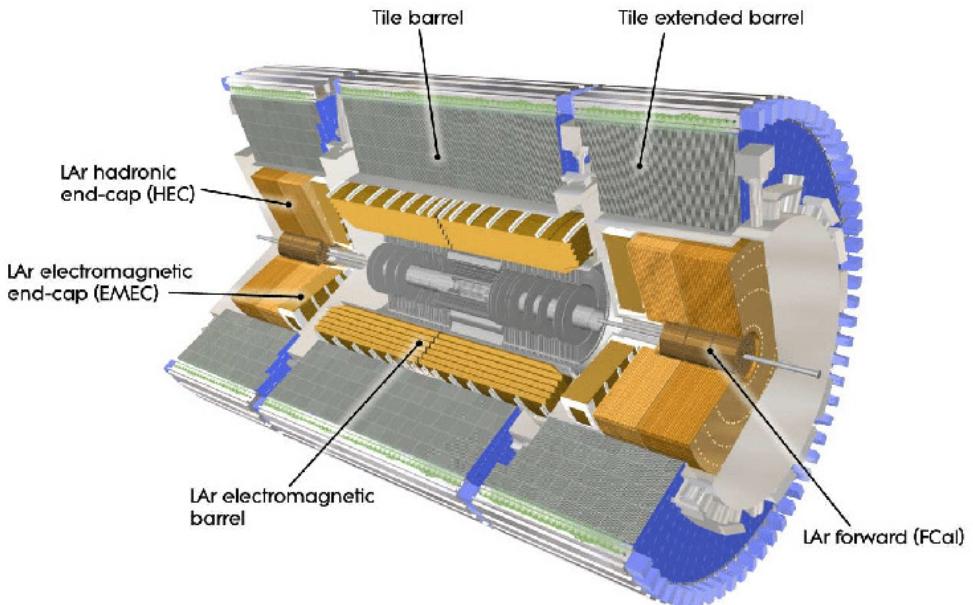


Figure 4.3: Cutaway view of the calorimeter system of the ATLAS detector [13].

425 Situated outside the Inner Detector are two concentric calorimeters. The inner calorimeter  
426 uses liquid argon (LAr) to measure the energy of particles that interact electromagnetically,  
427 which includes photons and any charged particle. The LAr calorimeter is made of heavy metals,  
428 primarily lead and copper, which causes electromagnetically interacting particles to shower,  
429 depositing their energy in the detector. The showering of the high energy particles that pass  
430 through the calorimeter cause the liquid argon to ionize, and the ionized electrons are detected  
431 by electronic readouts. The LAr calorimeter consists of around 180,000 readout channels.

432 The outer calorimeter measures the energy from particles that pass through the EM  
433 calorimeter, and measures the energy of particles that interact via the strong force. This is  
434 primarily hadrons. It is composed of steel plates designed to cause hadronic showering and  
435 around 500,000 scintillating tiles as the active material. The signals from the hadronic calorimeter  
436 are read out by photomultiplier tubes (PMTs).

### 437 **4.3 Muon Spectrometer**

438 Because muons are heavier than electrons and photons, and do not interact via the strong force,  
439 they generally pass through the detector without being stopped by the calorimeters. The outermost  
440 components of the detector are designed specifically to measure the energy and momentum of  
441 muons produced in the LHC. The muon spectrometer consists of tracking and triggering systems.

442 The largest of the subdetectors, it extends from the outside of the calorimeter system, about a  
443 4.25 m radius from the beam line, to a radius of 11 m. This large detector system is necessary

444 to accurately measure the momentum of muons, which is essential not only for measurements  
445 involving the muons themselves, but also to accurately estimate the missing energy in each  
446 event.

447 Two large toroidal magnets within the muon system generate a large magnetic field which  
448 covers an area 26 m long with a radius of 10 m. Because the area covered by this magnet system  
449 is so large, a uniform magnetic field like the one produced in the Inner Detector is impractical.  
450 Instead, the magnetic field that exists in the muon spectrometer ranges between 2 T and 8 T, and  
451 is much less uniform. The path of the muons passing through the spectrometer is bent by this  
452 field, allowing their charge to be determined.

453 1200 tracking chambers are placed in the muon system in order to precisely measure the  
454 tracks of muons with high spatial resolution. The path of the muons are tracked by Monitored  
455 Drift Tubes, which are drift chambers formed by aluminum tubes and filled with ionizing gas.  
456 These tubes produce a multi-layer spatial resolution on the order of 50  $\mu\text{m}$ .

457 **4.4 Trigger System**

458 Because of the high collision rate and large amount of data collected by the various subdetectors,  
459 ATLAS produces far more data than can actually be stored. Each event produces around 25 Mb  
460 of raw data, which multiplied by the bunch crossing rate of 40 MHz, comes out to around a  
461 petabyte of data every second. The information from every event cannot practically be stored,

462 therefore a sophisticated trigger system is employed in real time to determine whether events are  
463 sufficiently interesting to be worth storing.

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

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

---

## 476 Part IV

### 477 Measurement of WZ + Heavy Flavor

#### 478 5 Introduction

479 The production of  $WZ$  in association with a heavy flavor jet represents an important background  
480 for many major analyses. This includes any process with multiple leptons and b-jets in the final  
481 state, such as  $t\bar{t}H$ ,  $t\bar{t}W$ , and  $t\bar{t}Z$ . While precise measurements have been made of inclusive  
482  $WZ$  production [14],  $WZ$  + heavy flavor remains poorly understood. This is largely because the  
483 QCD processes involved in the production of the b-jet make it difficult to simulate accurately.  
484 This introduces a large uncertainty for analyses that include this process as a background.

485 We perform a study of the fully leptonic decay mode of this channel; that is, events where  
486 both the  $W$  and  $Z$  decay leptonically. Because  $WZ$  has no associated jets at leading order, while  
487 the major backgrounds for this channel tend to have high jet multiplicity, events with more than  
488 two jets are rejected. This gives a final state signature of three leptons and one or two jets.

489 Events that meet a preselection criteria are sorted into regions based on the b-tagging score  
490 of their associated jets. This is done to separate  $WZ$  + b-jet events from  $WZ$  + charm and  $WZ$  +  
491 light jets. These regions are fit to data in order to make a more accurate estimate of the contribution  
492 of  $WZ$  + heavy-flavor, where heavy-flavor jets include b-jets and charm jets. The full Run-2

<sup>493</sup> dataset collected by the ATLAS detector, representing  $139 \text{ fb}^{-1}$  of data from pp collisions at  
<sup>494</sup>  $\sqrt{s} = 13 \text{ TeV}$ , is used for this study.

<sup>495</sup> The fiducial volume at particle level is defined based on the number of stable leptons and  
<sup>496</sup> jets in each event. Three light leptons with total charge  $\pm 1$  and one or two associated jets are  
<sup>497</sup> required. Only leptons which do not originate from hadron or  $\tau$  decays are considered. The  
<sup>498</sup> phase space definitions use dressed kinematics of the final state particles. Leptons are dressed  
<sup>499</sup> by summing the momentum of photons within a cone of  $\Delta R < 0.1$  of the lepton to correct the  
<sup>500</sup> leptons energy. Particle level jets are reconstructed using the anti- $k_t$  algorithm with a radius of  
<sup>501</sup>  $R = 0.4$ . The kinematic selection applied to these objects is summarized below:

- <sup>502</sup> • Three light leptons with total charge  $\pm 1$ ,  $|\eta| < 2.5$
- <sup>503</sup> • OS lepton with  $p_T > 10 \text{ GeV}$ , SS leptons with  $p_T > 20 \text{ GeV}$
- <sup>504</sup> • One OSSF lepton pair with  $|M(l\bar{l}) - 91.2 \text{ GeV}| < 10 \text{ GeV}$
- <sup>505</sup> • One or two associated truth jets with  $p_T > 25 \text{ GeV}$ ,  $|\eta| < 2.5$ ,  $R < 0.4$

<sup>506</sup> The result of the fit is used to extract the cross-section in this fiducial region for WZ + b  
<sup>507</sup> and WZ + c with one associated jet, and WZ + b and WZ + c with two associated jets, where the  
<sup>508</sup> number and flavor of the jets is determined at particle level. Events with both charm and b-jets  
<sup>509</sup> are counted as WZ + b. The analysis reports a cross-section measurement of WZ + b and WZ +  
<sup>510</sup> charm, along with their correlations, for both 1-jet and 2-jet exclusive regions.

511       Section 6 details the data and Monte Carlo (MC) samples used in the analysis. The  
512       reconstruction of various physics objects is described in Section 7. Section 8 describes the event  
513       selection applied to these samples, along the definitions of the various regions used in the fit.  
514       The multivariate analysis techniques used to separate the tZ background from WZ + heavy flavor  
515       are described in Section 9. Section 16 describes the various sources of systematic uncertainties  
516       considered in the fit. Finally, the results of the analysis are summarized in Section 17, followed  
517       by a brief conclusion in Section ??.

## 518       **6 Data and Monte Carlo Samples**

### 519       **6.1 Data Samples**

520       This study uses a sample of proton-proton collision data collected by the ATLAS detector from  
521       2015 through 2018 at an energy of  $\sqrt{s} = 13$  TeV, which represents an integrated luminosity of  
522        $139 \text{ fb}^{-1}$  [15]. This data set was collected with a bunch-crossing rate of 25 ns. All data used in  
523       this analysis was verified by data quality checks [16].

### 524       **6.2 Monte Carlo Samples**

525       Several different generators were used to produce Monte Carlo simulations of the signal and  
526       background processes. For all samples, the response of the ATLAS detector is simulated using

<sup>527</sup> GEANT4 [17]. The WZ signal samples are simulated using Sherpa 2.2.2 [18]. Specific information  
<sup>528</sup> about the Monte Carlo samples being used can be found in Table 13.

Table 2: 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
WZ, ZZ, WW	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10 [19]
tZ	MG5_AMC [20]	NLO	PYTHIA 8	CTEQ6L1
t̄W	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
t̄(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄H	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++) [22]	NNPDF 3.0 NLO [21] (CT10 [19])
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
t̄t, t̄tt̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO [23]
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄t	POWHEG-BOX v2 [24]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄t̄γ	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	POWHEG-BOX v1 [25]	NLO	PYTHIA 6	CT10
qqVV, VVV Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

## 7 Object Reconstruction

<sup>530</sup> All regions defined in this analysis share a common lepton, jet, and overall event preselection.  
<sup>531</sup> The selection applied to each physics object is detailed here; the event preselection, and the  
<sup>532</sup> selection used to define the various fit regions, is described in Section 8.

<sup>533</sup> All events are required to be selected by dilepton triggers. The  $p_T$  thresholds of the

534 dilepton trigger on two electrons were 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017 and  
 535 2018, while for the dimuon triggers the  $p_T$  thresholds on the leading (sub-leading) muon were  
 536 18 GeV (8 GeV) in 2015, and 22 GeV (8 GeV) in 2016-2018. For the electron+muon triggers,  
 537 the  $p_T$  thresholds on the electron (muon) were 17 GeV (14 GeV) for all datasets.

538 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter  
 539 that are associated with charged particle tracks reconstructed in the inner detector [[ele\\_eff](#)].  
 540 Electron candidates are required to have  $p_T > 10$  GeV and  $|\eta_{\text{cluster}}| < 2.47$ . Candidates in the  
 541 transition region between different electromagnetic calorimeter components,  $1.37 < |\eta_{\text{cluster}}| <$   
 542 1.52, are rejected. A multivariate likelihood discriminant combining shower shape and track  
 543 information is used to distinguish real electrons from hadronic showers (fake electrons). To  
 544 further reduce the non-prompt electron contribution, the track is required to be consistent with  
 545 originating from the primary vertex; requirements are imposed on the transverse impact para-  
 546 meter significance ( $|d_0|/\sigma_{d_0} < 5$ ) and the longitudinal impact parameter ( $|\Delta z_0 \sin \theta_\ell| < 0.5$   
 547 mm). Electron candidates are required to pass the TIGHTLH identification requirement detailed  
 548 in [26].

549 Muon candidates are reconstructed by combining inner detector tracks with track segments  
 550 or full tracks in the muon spectrometer [27]. Muon candidates are required to have  $p_T > 10$  GeV  
 551 and  $|\eta| < 2.5$ . The longitudinal impact parameter is the same for both electrons and muons,  
 552 while muons are required to pass a slightly tighter transverse impact parameter selection,  
 553  $|d_0|/\sigma_{d_0} < 3$ . Muons are also required to pass Medium ID requirements, as detailed in [26].  
 554 Leptons are additionally required to pass a non-prompt BDT selection, described in detail in

555 [28]. Optimized working points and scale factors for this BDT are taken from that analysis.

556        Jets are reconstructed from calibrated topological clusters built from energy deposits in  
557        the calorimeters using the anti- $k_t$  algorithm [29], as well as information from the inner tracking  
558        detector, with a radius parameter  $R = 0.4$ . Jets with energy contributions likely arising from  
559        noise or detector effects are removed from consideration, and only jets satisfying  $p_T > 25$  GeV  
560        and  $|\eta| < 2.5$  are used in this analysis. For jets with  $p_T < 60$  GeV and  $|\eta| < 2.4$ , a jet-track  
561        association algorithm is used to confirm that the jet originates from the selected primary vertex,  
562        in order to reject jets arising from pileup collisions [30].

563        In order to make a measurement of  $WZ +$  heavy flavor it is necessary to distinguish these  
564        events from  $WZ +$  light jets. For this purpose, the DL1r b-tagging algorithm is used to distinguish  
565        heavy flavor jets from lighter ones [31]. The DL1r algorithm uses jet kinematics, particularly  
566        jet vertex information, as input for a neural network which assigns each jet a score designed to  
567        reflect how likely that jet is to have originated from a b-quark.

568        From the output of the BDT, calibrated working points (WPs) are developed based on the  
569        efficiency of truth b-jets at particular values of the DL1r algorithm. The working points used in  
570        this analysis are summarized in Table 3.

WP	Rejection	
	b-jet eff.	c-jet
85%	2.6	29
77%	4.9	130
70%	9.4	390
60%	27	1300

Table 3: c-jet and light-flavor jet rejections corresponding to each b-tagging Working Point by b-jet efficiency, evaluated on  $t\bar{t}$  events.

571 As shown in table 3, a tighter WP will accept fewer b-jets, but reject a higher fraction of  
 572 charm and light jets. Generally, analyses that include b-jets will use a fixed working point, for  
 573 example, requiring that a jet pass the 70% threshold. By instead treating these working point  
 574 as bins, e.g. events with jets that fall between the 85% and 77% WPs fall into one bin, while  
 575 events with jets passing the 60% WP fall into another, additional information can be gained.  
 576 This analysis uses each of these working points to form orthogonal regions in order to provide  
 577 separation between  $WZ + b$ ,  $WZ + c$ , and  $WZ + \text{light}$ .

578 Missing transverse momentum ( $E_T^{\text{miss}}$ ) is used as part of the event selection. The missing  
 579 transverse momentum vector is defined as the negative of the vector of the transverse momenta  
 580 of all reconstructed physics objects as well as remaining unclustered energy, the latter of which is  
 581 estimated from low- $p_T$  tracks associated with the primary vertex but not assigned to a hard object,  
 582 with object definitions taken from [32]. Light leptons considered in the  $E_T^{\text{miss}}$  reconstruction are  
 583 required to have  $p_T > 10$  GeV, while jets are required to have  $p_T > 20$  GeV.

584 To avoid double counting objects and remove leptons originating from decays of hadrons,  
 585 overlap removal is performed in the following order: any electron candidate within  $\Delta R = 0.1$  of

586 another electron candidate with higher  $p_T$  is removed; any electron candidate within  $\Delta R = 0.1$   
587 of a muon candidate is removed; any jet within  $\Delta R = 0.2$  of an electron candidate is removed; if  
588 a muon candidate and a jet lie within  $\Delta R = \min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$  of each other,  
589 the jet is kept and the muon is removed if the jet has at least three associated tracks, otherwise  
590 the jet is removed and the muon is kept. This algorithm is applied to the preselected objects.

## 591 **8 Event Selection and Signal Region Definitions**

592 Event are required to pass a preselection described in Section 8.1. Those that pass this preselection  
593 are divided into various fit regions described in Section 8.2, based on the number of jets in the  
594 event, and the b-tag score of those jets.

### 595 **8.1 Event Preselection**

596 Events are required to include exactly three reconstructed light leptons passing the requirement  
597 described in 7, which have a total charge of  $\pm 1$ . As the opposite sign lepton is found to be prompt  
598 the vast majority of the time [28], it is required to have  $p_T > 10 \text{ GeV}$ , while the same sign leptons  
599 are required to have  $p_T > 20 \text{ GeV}$  to reduce the contribution of non-prompt leptons.

600 The invariant mass of at least one pair of opposite sign, same flavor leptons is required  
601 to fall within 10 GeV of the mass of the Z boson, 91.2 GeV. Events where one of the opposite

602 sign pairs have an invariant mass less than 12 GeV are rejected in order to suppress low mass  
 603 resonances.

604 An additional requirement is placed on the missing transverse energy,  $E_T^{\text{miss}} > 20 \text{ GeV}$ .  
 605 The transverse mass of the  $W$  candidate, defined as  $\sqrt{2p_T^{\text{lep}} E_T^{\text{miss}} * (1 - \cos(\phi_{\text{lep}} - \phi_{E_T^{\text{miss}}}))}$ , is  
 606 required to be greater than 30 GeV. Here  $E_T^{\text{miss}}$  is the missing transverse energy, and the lepton  
 607 considered is the lepton not included in the  $Z$ -candidate.

608 Events are required to have exactly one or two reconstructed. Events with more than two  
 609 jets are rejected in order to reduce the contribution of backgrounds such as  $t\bar{t}Z$  and  $t\bar{t}W$ , which  
 610 tend to have higher jet multiplicity.

611 The  $WZ$  events are split into  $WZ + b$ ,  $WZ + c$ , and  $WZ + \text{light}$  based on the truth flavor of  
 612 the associated jet in the event, as determined by the presence of a  $b$ - or  $c$ -hadron within  $R = 0.3$   
 613 of the jet. In this ordering  $b$ -jet supersedes charm, which supersedes light. That is,  $WZ + \text{light}$   
 614 events contain no charm and no  $b$  jets at truth level,  $WZ + c$  contain at least one truth charm and  
 615 no  $b$ -jets, and  $WZ + b$  contains at least one truth  $b$ -jet.

## 616 8.2 Fit Regions

617 Once preselection has been applied, the remaining events are categorized into one of twelve  
 618 orthogonal regions. The regions used in the fit are summarized in Table 4.

Table 4: A list of the regions used in the fit and the selection used for each.

Region	Selection
1j, <85%	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_85} = 0$
1j, 85%-77%	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_85} = 1, n\text{Jets\_DL1r\_77}=0$
1j, 77%-70%	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_77} = 1, n\text{Jets\_DL1r\_70}=0$
1j, 70%-60%	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_70} = 1, n\text{Jets\_DL1r\_60}=0$
1j, >60%	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_60} = 1, tZ \text{ BDT} > 0.12$
1j tZ CR	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_60} = 1, tZ \text{ BDT} < 0.12$
2j, <85%	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_85} = 0$
2j, 85%-77%	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_85} >= 1, n\text{Jets\_DL1r\_77}=0$
2j, 77%-70%	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_77} >= 1, n\text{Jets\_DL1r\_70}=0$
2j, 70%-60%	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_70} >= 1, n\text{Jets\_DL1r\_60}=0$
2j, >60%	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_60} >= 1, tZ \text{ BDT} > 0.12$
2j tZ CR	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_60} >= 1, tZ \text{ BDT} < 0.12$

619        The working points discussed in Section 7 are used to separate events into fit regions based  
 620        on the highest working point reached by a jet in each event. Because the background composition  
 621        differs significantly based on the number of b-jets, events are further subdivided into 1-jet and  
 622        2-jet regions in order to minimize the impact of background uncertainties.

623        An unfolding procedure is performed to account for differences in the number of recon-  
 624        structed jets compared to the number of truth jets in each event. In order to account for migration  
 625        of WZ+1-jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal samples  
 626        are separated based on the number of truth jets. Events with 0 jets or more than 3 jets at truth  
 627        level, yet fall within one of the categories listed in Table 4, are categorized as WZ + other, and  
 628        treated as background. The composition of the number of truth jets in each reco jet bin is taken  
 629        from MC, with uncertainties in these estimates described in detail in Section 16.

630        An additional tZ control region is created based on the BDT described in Section 9. The

631 region with 1-jet passing the 60% working point is split in two - a signal enriched region of  
632 events with a BDT score greater than 0.12, and a tZ control region including events with less  
633 than 0.12. This cutoff is optimized for significance of WZ + b.

### 634 8.3 Non-Prompt Lepton Estimation

635 Two processes that act as sources of non-prompt leptons appear in the analysis:  $t\bar{t}$  and  $Z+jet$   
636 production both produce two prompt leptons, but can meet the selection of this analysis when  
637 an additional non-prompt lepton appears in the event. The contribution of these processes is  
638 estimated with Monte Carlo simulations, which are validated using non-prompt enriched regions.  
639 These validation regions are used to derive correction factors and uncertainties for the non-prompt  
640 contribution.

641  $t\bar{t}$  events can produce two prompt leptons from the decay of each of the tops. These top  
642 decays produce two b-quarks, the decay of which can produce additional non-prompt leptons,  
643 which occasionally pass the event preselection. In order to validate that the Monte Carlo  
644 accurately simulates this process accurately, the MC prediction in a non-prompt  $t\bar{t}$  enriched  
645 validation region is compared to data.

646 The  $t\bar{t}$  validation region is similar to the preselection region - three leptons meeting the  
647 criteria described in Section 8 are required, and the requirements on  $E_T^{\text{miss}}$  remain the same.  
648 However, the selection requiring that a lepton pair form a Z-candidate are reversed. Events  
649 where the invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of

650 91.2 GeV are rejected. This ensures the  $t\bar{t}$  validation region is orthogonal to the preselection  
651 region.

652 Further, because the jet multiplicity of  $t\bar{t}$  events tends to be higher than WZ + jets, the  
653 number of jets in each event is required to be greater than 1. As b-jets are almost invariably  
654 produced from top decays, at least one b-tagged jet passing the 70% DL1r WP in each event is  
655 required.

656 Data is compared to MC predictions in the region for a variety of kinematic variable,  
657 as well as various b-tag WPs. A constant normalization discrepancy between data and MC  
658 predictions of approximately 10% is found, which is accounted for by applying a constant  
659 correction factor of 0.9 to the  $t\bar{t}$  MC prediction. Once this correction factor has been applied, no  
660 significant modelling discrepancies, either in terms of shape or overall yield, are found in any of  
661 the kinematic distributions considered. As data and MC are found to agree within 20% for each  
662 of the b-tag WPs considered, a 20% systematic uncertainty on the  $t\bar{t}$  prediction is included for  
663 the analysis.

664 Similar to  $t\bar{t}$ , a Z+jets validation region is produced in order to validate the MC predictions.  
665 The lepton requirements remain the same as the preselection region. Because no neutrinos are  
666 present for this process, the  $E_T^{\text{miss}}$  cut is reversed, requiring  $E_T^{\text{miss}} < 30$  GeV. This also ensures  
667 this validation region is orthogonal to the preselection region. Further, the number of jets in each  
668 event is required to be greater than or equal to one.

669 While there is general agreement between data and MC, the shape of the  $p_T$  spectrum

670 of the lepton from the W candidate is found to differ. This is the lepton not included in the  
 671 Z-candidate, and in the case of Z+jets, this lepton is most often the non-prompt lepton. To  
 672 account for this discrepancy, a variable correction factor is applied to Z+jets.  $\chi^2$  minimization  
 673 of the W lepton  $p_T$  spectrum is performed to derive a correction factor.

674 The systematic uncertainty in the Z + jets prediction is evaluated by comparing data to  
 675 MC for each of the continuous b-tag WPs. For each of the regions considered, the data falls  
 676 within 25% of the MC prediction once this correction factor has been applied. Therefore, a 25%  
 677 systematic uncertainty is applied to Z + jets in the analysis.

## 678 **9 tZ Separation Multivariate Analysis**

679 An important process to consider in this analysis is tZ: the top almost always decays into a W  
 680 boson and b-quark, and when both the W and Z decay leptonically, this gives three leptons and  
 681 a heavy flavor jet in the final state. Because tZ can produce a final state identical to the signal,  
 682 it represents a predominant background in the most signal enriched regions. That is, the region  
 683 with one jet passing the 60% DL1r WP. Therefore, a boosted decision tree (BDT) algorithm is  
 684 trained using XGBoost [33] to separate WZ + heavy flavor from tZ using kinematic quantities.  
 685 The result of this BDT is used to create a tZ enriched region in the fit, reducing its impact on the  
 686 measurement of WZ + heavy flavor.

687 The kinematic variables used as inputs to train this BDT include the invariant mass of the  
 688 reconstructed top candidate, the  $p_T$  of each of the leptons and associated jets, the invariant mass

689 of each combination of lepton pairs,  $E_T^{\text{miss}}$ , the distance between each combination of leptons,  
690  $\Delta R(l\bar{l})$ , and the distance between each lepton and the jet,  $\Delta R(lj)$ .

691 Here the top candidate is reconstructed based on the procedure described in section 6.1 of  
692 [34]. Broadly, the mass of the top quark candidate is reconstructed from the jet, the lepton not  
693 included in the Z-candidate, and a reconstructed neutrino. In the case that there is one jet in the  
694 event, there is only possible b-jet candidate. For events with two jets, the jet with the highest  
695 DL1r score is used.

696 The training samples included only events meeting the requirements of the 1-jet, >60%  
697 region, i.e. passing all the selection described in section 8 and having exactly one jet which passes  
698 the tightest (60%) DL1r working point. A sample of 20,000 background (tZ) and signal (WZ+b)  
699 Monte Carlo events are used to train the BDT. And additional 5,000 events are reserved for testing  
700 the model, in order to prevent over-fitting. A total of 750 decision trees with a maximum depth  
701 of 6 branches are used to build the model. These parameters are chosen empirically, by training  
702 several models with different parameters and selecting the one that gave the best separation for  
703 the test sample. The results of the BDT training are shown in figure 9.1.

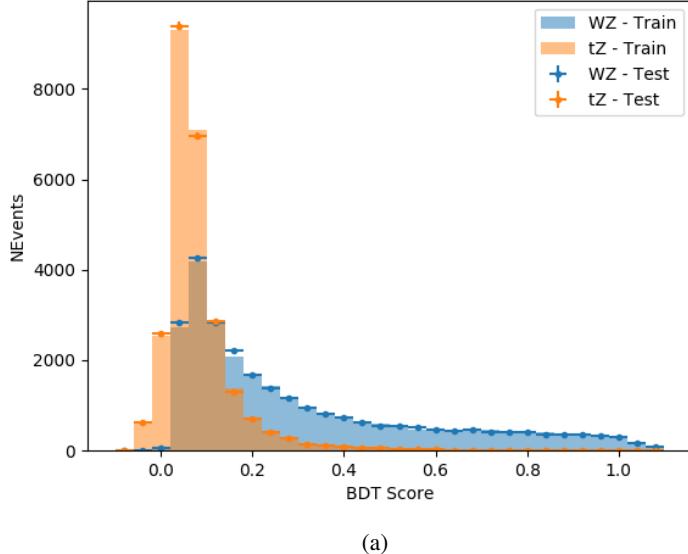


Figure 9.1: Distribution of the BDT response for WZ+b (blue) and tZ (orange) events, for both training and testing samples.

704        A BDT score of 0.12 is selected as a cutoff, where events with scores higher than this form  
 705        a signal enriched region, and events with scores lower than this form a tZ control region. This  
 706        cutoff is selected by varying the value of this cutoff in stat-only Asimov fits, and selecting the  
 707        value that minimizes the statistical uncertainty on WZ + b.

## 708        10 Systematic Uncertainties

709        The systematic uncertainties that are considered are summarized in Table 33. These are imple-  
 710        mented in the fit either as a normalization factors or as a shape variation or both in the signal

<sup>711</sup> and background estimations. The numerical impact of each of these uncertainties is outlined in

<sup>712</sup> Section 17.

Table 5: Sources of systematic uncertainty considered in the analysis.

Systematic uncertainty	Components
Luminosity	1
Pileup reweighting	1
<b>Physics Objects</b>	
Electron	6
Muon	15
Jet energy scale	28
Jet energy resolution	8
Jet vertex fraction	1
Jet flavor tagging	131
$E_T^{\text{miss}}$	3
Total (Experimental)	194
<b>Signal Modeling</b>	
Shape modelling	3
Renormalization and factorization scales	5
nJet Migration	5
<b>Background Modeling</b>	
Cross section	15
Renormalization and factorization scales	12
Total (Signal and background modeling)	35
Total (Overall)	230

<sup>713</sup> The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [35], obtained

<sup>714</sup> using the LUCID-2 detector [36] for the primary luminosity measurements.

<sup>715</sup> The experimental uncertainties are related to the reconstruction and identification of light

<sup>716</sup> leptons and b-tagging of jets, and to the reconstruction of  $E_T^{\text{miss}}$ . The sources which contribute to

<sup>717</sup> the uncertainty in the jet energy scale (JES) [37] are decomposed into uncorrelated components

<sup>718</sup> and treated as independent sources of uncertainty in the analysis. A similar approach is used for

719 the jet energy resolution (JER) uncertainty.

720 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses  
721 [38] are also decomposed into uncorrelated components. The large number of components for  
722 b-tagging is due to the calibration of the distribution of the MVA discriminant for each individual  
723 WP bin.

724 The fit involves varying the overall normalization of signal templates over the regions  
725 described in Section 8.2, which are defined by the flavor and number of associated jets at truth-  
726 level. The modelling of these template shapes therefore significantly impacts the final result.  
727 Additional signal uncertainties, probing the shape of the signal templates as well as the rate of  
728 migrations between the number of truth-jets and reconstructed jets, are estimated by comparing  
729 estimates from the nominal Sherpa WZ samples with alternative WZ samples generated with  
730 POWHEG+PYTHIA8. Separate systematics are included in the fit for WZ + b, WZ + c and WZ +  
731 light, where the distribution among each of the fit regions is varied based on the prediction of  
732 the Powheg sample.

733 A similar approach is taken to account for uncertainties in migrations between the number  
734 of reco and truth jets. The fraction of events with 1 truth jet which fall into the 1 jet bin versus  
735 the 2 jet bin at reco level is compared for Sherpa and Powheg. The same is done for events with  
736 2 truth jets. A systematic is included where events are shifted between the 1-jet and 2-jet regions  
737 based on the differences between these two shapes. This is done independently for each of the  
738 WZ + b, WZ + c, and WZ + light templates.

739 Additional systematics are included to account for the uncertainty in the contamination of  
740 0 jet and 3 or more jet events (as defined at truth level) in the 1 and 2 reco jet bins. Because these  
741 events fall outside the scope of this measurement, these events are included as a background.  
742 As such, a normalization, rather than a shape, uncertainty is applied for this background. The  
743 number of WZ events with 0-jets and  $>=3$ -jets in the reconstructed 1-jet and 2-jet regions are  
744 compared for Sherpa and Powheg, and these differences are taken as separate normalization  
745 systematics on the yield of WZ+0-jet and WZ+ $>=3$ -jet events.

746 Theoretical uncertainties applied to MC background predictions, including cross section,  
747 PDF, and scale uncertainties are taken from theory calculations, with the exception of non-prompt  
748 and diboson backgrounds. The cross-section uncertainty on tZ is taken from [39]. Derivation  
749 of the non-prompt background uncertainties, Z+jets and t $\bar{t}$ , are explained in Section 8.3. These  
750 normalization uncertainties are chosen so as to account for the complete uncertainty in the  
751 non-prompt contribution, and therefore no additional modelling uncertainties are considered for  
752 Z+jets and t $\bar{t}$ .

753 Due to its importance as a background, additional modelling uncertainties are considered  
754 for tZ. Alternative tZ samples with variations in scale and shower modelling are included as  
755 systematics. The other VV + heavy flavor processes (namely VV+b and VV+charm, which  
756 primarily consist of ZZ events) are also poorly understood, because these processes involve the  
757 same physics as WZ + heavy flavor, and have also not been measured. Therefore, a conservative  
758 50% uncertainty is applied to those samples. While this uncertainty is large, it is found to have  
759 little impact on the significance of the final result.

760 The theory uncertainties applied to the MC estimates are summarized in Table 35.

Process	X-section [%]
WZ	QCD Scale: $^{+3.7}_{-3.4}$ PDF( $+\alpha_S$ ): $\pm 3.1$
tZ	X-sec: $\pm 15.2$
t <bar>t} H (aMC@NLO+Pythia8)</bar>	QCD Scale: $^{+5.8}_{-9.2}$ PDF( $+\alpha_S$ ): $\pm 3.6$
t <bar>t} Z (aMC@NLO+Pythia8)</bar>	QCD Scale: $^{+9.6}_{-11.3}$ PDF( $+\alpha_S$ ): $\pm 4$
t <bar>t} W (aMC@NLO+Pythia8)</bar>	QCD Scale: $^{+12.9}_{-11.5}$ PDF( $+\alpha_S$ ): $\pm 3.4$
VV + b/charm (Sherpa 2.2.1)	$\pm 50$
VV + light (Sherpa 2.2.1)	$\pm 6$
t <bar>t}</bar>	$\pm 20$
Z + jets	$\pm 25$
Others	$\pm 50$

Table 6: Summary of theoretical uncertainties for normalization of MC predictions in the analysis.

## 761 11 Results

762 A separate maximum-likelihood fit is performed over the 1-jet and 2-jet fit regions in order to  
 763 extract the best-fit value of the WZ + b and WZ + c jet contributions. The WZ + b, WZ + charm  
 764 and WZ + light contributions are allowed to float, with the remaining background contributions  
 765 are held fixed. **The current fit strategy treats the WZ + b contribution as the parameter**  
 766 **of interest, with the normalization of the WZ + charm and the WZ + light contributions**  
 767 **taken as systematic uncertainties. This could however be adjusted, depending on whether it**  
 768 **is decided the goal of the analysis should be to measure WZ + b specifically or WZ + heavy**  
 769 **flavor overall.** The result of the fit is used to extract the cross-section of WZ + heavy-flavor  
 770 production.

771 A maximum likelihood fit to data is performed simultaneously in the regions described  
 772 in Section 8. The parameters  $\mu_{WZ+b}$ ,  $\mu_{WZ+charm}$ ,  $\mu_{WZ+light}$ , where  $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$ , are  
 773 extracted from the fit.

774 The Asimov fit for 1-jet events gives an expected  $\mu$  value of  $1.00^{+0.37}_{-0.35}(\text{stat})^{+0.24}_{-0.22}(\text{sys})$  for  
 775  $WZ + b$ . The fitted cross-section modifiers for  $WZ + \text{charm}$  and  $WZ + \text{light}$  are  $1.00 \pm 0.17 \pm 0.15$   
 776 and  $1.00 \pm 0.04 \pm 0.07$ , respectively.

777 The expected cross-section of  $WZ + b$  with 1 associated jet obtained from the fit is  
 778  $1.74^{+0.65}_{-0.61}(\text{stat})^{+0.42}_{-0.37}(\text{sys}) \text{ fb}$  with an expected significance of  $2.2\sigma$ . The expected cross-section  
 779 of  $WZ + c$  is measured to be  $14.6 \pm 2.5(\text{stat}) \pm 2.1(\text{sys}) \text{ fb}$ , with a correlation of -0.23.

780 For 2-jet events, the fit gives an expected  $\mu$  value of  $1.00^{+0.34}_{-0.33}(\text{stat})^{+0.29}_{-0.28}(\text{sys})$  for  $WZ +$   
 781  $b$ . The fitted cross-section modifiers for  $WZ + \text{charm}$  and  $WZ + \text{light}$  are  $1.00 \pm 0.17 \pm 0.15$  and  
 782  $1.00 \pm 0.04 \pm 0.08$ , respectively.

783 The expected  $WZ + b$  cross-section in the 2-jet region is  $2.46^{+0.83}_{-0.81}(\text{stat})^{+0.073}_{-0.68}(\text{sys}) \text{ fb}$   
 784 with an expected significance of  $2.6\sigma$ . The 2-jet expected cross-section of  $WZ + \text{charm}$  is  
 785  $12.7 \pm 2.2(\text{stat}) \pm 2.0(\text{sys}) \text{ fb}$ , and the correlation between  $WZ + c$  and  $WZ + b$  is -0.26.

## 786 **11.1 1-jet Fit Results**

787 **The results of the fit are currently blinded.**

<sup>788</sup> The pre-fit yields in each of the regions used in the fit are shown in Table 11.1, and

<sup>789</sup> summarized in Figure 11.1.

<sup>790</sup>

Sample	1j, <85% WP	1j, 77%-85% WP	1j, 70%-77% WP	1j, 60%-70% WP	1j, >60% WP	tZ CR
WZ + b - 1j	$8.1 \pm 1.6$	$4.7 \pm 0.5$	$4.6 \pm 0.4$	$5.1 \pm 0.4$	$18.2 \pm 2.4$	$4.8 \pm 0.6$
WZ + c - 1j	$260 \pm 22$	$81 \pm 6$	$43.1 \pm 3.6$	$25.8 \pm 2.6$	$9.4 \pm 1.8$	$2.9 \pm 0.6$
WZ + l - 1j	$3090 \pm 250$	$91 \pm 13$	$17 \pm 3$	$4.9 \pm 1.6$	$1.3 \pm 0.4$	$0.2 \pm 0.1$
WZ + b - 2j	$1.10 \pm 0.37$	$0.44 \pm 0.11$	$0.39 \pm 0.06$	$0.62 \pm 0.14$	$2.1 \pm 0.5$	$0.59 \pm 0.14$
WZ + c - 2j	$21 \pm 5$	$5.6 \pm 1.2$	$3.0 \pm 0.7$	$2.0 \pm 0.5$	$0.70 \pm 0.20$	$0.30 \pm 0.08$
WZ + l - 2j	$250 \pm 60$	$5.7 \pm 1.6$	$0.73 \pm 0.53$	$0.31 \pm 0.15$	$0.07 \pm 0.06$	$0.01 \pm 0.01$
WZ - Other	$13 \pm 5$	$1.4 \pm 0.4$	$0.42 \pm 0.08$	$0.2 \pm 0.01$	$0.30 \pm 0.05$	$0.67 \pm 0.15$
Other VV	$6.2 \pm 0.6$	$0.2 \pm 0.4$	$0.2 \pm 0.04$	$0.07 \pm 0.1$	$0.1 \pm 0.1$	$0.1 \pm 0.2$
ZZ	$336 \pm 26$	$17.8 \pm 2.1$	$4.3 \pm 0.6$	$1.7 \pm 0.5$	$0.36 \pm 0.08$	$0.10 \pm 0.03$
t̄tW	$1.1 \pm 0.2$	$0.2 \pm 0.1$	$0.3 \pm 0.1$	$0.4 \pm 0.1$	$1.5 \pm 0.3$	$0.7 \pm 0.2$
t̄tZ	$6.8 \pm 1.2$	$1.4 \pm 0.3$	$1.0 \pm 0.2$	$1.2 \pm 0.2$	$4.4 \pm 0.8$	$3.2 \pm 0.6$
Z + jets	$169 \pm 38$	$8.9 \pm 1.9$	$3.7 \pm 0.8$	$3.3 \pm 0.7$	$3.2 \pm 0.7$	$0.8 \pm 0.17$
V + $\gamma$	$45 \pm 28$	$1.9 \pm 2.4$	$0.1 \pm 0.1$	$0.02 \pm 0.01$	$1.0 \pm 0.9$	$0.02 \pm 0.03$
tZ	$24.3 \pm 4.3$	$5.5 \pm 1.1$	$4.1 \pm 0.8$	$5.9 \pm 1.1$	$10.7 \pm 2.0$	$23 \pm 4$
tW	$1.4 \pm 0.8$	$0.2 \pm 0.5$	$0.0 \pm 0.2$	$0.7 \pm 0.6$	$0.26 \pm 0.42$	$0.39 \pm 0.41$
WtZ	$2.3 \pm 1.2$	$0.6 \pm 0.3$	$0.3 \pm 0.21$	$0.27 \pm 0.2$	$1.1 \pm 0.7$	$0.6 \pm 0.5$
VVV	$12.4 \pm 0.5$	$0.93 \pm 0.06$	$0.35 \pm 0.03$	$0.13 \pm 0.02$	$0.14 \pm 0.03$	$0.02 \pm 0.01$
VH	$40 \pm 6$	$2.6 \pm 1.4$	$0.9 \pm 0.8$	$0.7 \pm 0.8$	$0.5 \pm 0.6$	$0.0 \pm 0.0$
t̄t	$12.1 \pm 1.6$	$2.9 \pm 0.6$	$2.5 \pm 0.5$	$2.8 \pm 0.5$	$11.2 \pm 1.4$	$10.9 \pm 1.5$
t̄tH	$0.24 \pm 0.03$	$0.05 \pm 0.01$	$0.04 \pm 0.01$	$0.06 \pm 0.01$	$0.20 \pm 0.03$	$0.13 \pm 0.02$
Total	$5010 \pm 260$	$227 \pm 24$	$88 \pm 12$	$57 \pm 8$	$76 \pm 16$	$53 \pm 8$

Table 7: Pre-fit yields in each of the 1-jet fit regions.

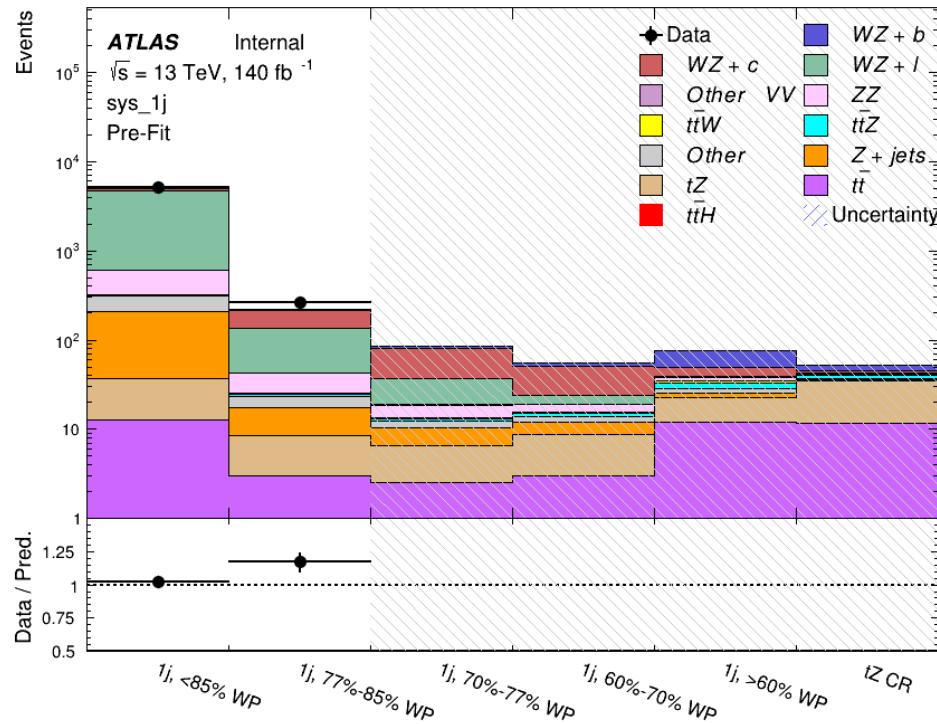


Figure 11.1: Pre-fit summary of the 1-jet fit regions.

791

The post-fit yields in each region are summarized in Figure 11.1.

792

	1j, <85% WP	1j, 77%-85% WP	1j, 70%-77% WP	1j, 60%-70% WP	1j, >60% WP	tZ CR
WZ + b	$8.1 \pm 4.9$	$4.7 \pm 2.0$	$4.6 \pm 2.0$	$5.1 \pm 2.1$	$18 \pm 10$	$5.0 \pm 2.5$
WZ + c	$260 \pm 60$	$80 \pm 14$	$43 \pm 7$	$26 \pm 5$	$9.4 \pm 2.3$	$2.9 \pm 0.7$
WZ + l	$3090 \pm 130$	$90 \pm 11$	$17.3 \pm 2.8$	$4.9 \pm 1.6$	$1.3 \pm 0.4$	$0.23 \pm 0.1$
WZ + b - 2j	$1.10 \pm 0.37$	$0.44 \pm 0.11$	$0.39 \pm 0.06$	$0.62 \pm 0.14$	$2.1 \pm 0.5$	$0.59 \pm 0.1$
WZ + c - 2j	$21 \pm 5$	$5.6 \pm 1.2$	$3.0 \pm 0.7$	$2.0 \pm 0.5$	$0.70 \pm 0.20$	$0.30 \pm 0.0$
WZ + l - 2j	$250 \pm 60$	$5.7 \pm 1.6$	$0.73 \pm 0.53$	$0.31 \pm 0.15$	$0.07 \pm 0.06$	$0.01 \pm 0.0$
WZ - Other	$13 \pm 5$	$1.4 \pm 0.4$	$0.42 \pm 0.08$	$0.2 \pm 0.01$	$0.30 \pm 0.05$	$0.67 \pm 0.1$
Other VV	$6.2 \pm 0.6$	$0.92 \pm 0.07$	$0.02 \pm 0.01$	$0.01 \pm 0.01$	$0.01 \pm 0.01$	$0.01 \pm 0.0$
ZZ	$346 \pm 57$	$19 \pm 5$	$4.3 \pm 0.8$	$2.7 \pm 0.5$	$2.4 \pm 0.1$	$2.1 \pm 0.6$
t̄tW	$1.09 \pm 0.21$	$0.2 \pm 0.1$	$0.1 \pm 0.1$	$0.4 \pm 0.1$	$1.5 \pm 0.3$	$0.1 \pm 0.2$
t̄tZ	$6.8 \pm 1.2$	$1.4 \pm 0.3$	$1.0 \pm 0.2$	$1.2 \pm 0.2$	$4.4 \pm 0.7$	$3.2 \pm 0.5$
rare Top	$0.14 \pm 0.04$	$0.04 \pm 0.02$	$0.04 \pm 0.0$	$0.1 \pm 0.03$	$0.14 \pm 0.04$	$0.15 \pm 0.0$
t̄tWW	$0.04 \pm 0.03$	$0.01 \pm 0.02$	$0.01 \pm 0.01$	$0.01 \pm 0.01$	$0.01 \pm 0.02$	$0.01 \pm 0.0$
Z + jets	$169 \pm 37$	$8.9 \pm 1.9$	$3.7 \pm 0.8$	$3.3 \pm 0.7$	$3.2 \pm 0.7$	$0.8 \pm 0.2$
W + jets	$0.01 \pm 0.01$	$0.01 \pm 0.0$				
V + $\gamma$	$46 \pm 28$	$1.9 \pm 2.4$	$0.1 \pm 0.1$	$0.0 \pm 0.2$	$1.0 \pm 0.9$	$0.0 \pm 0.0$
tZ	$24 \pm 4$	$5.5 \pm 1.0$	$4.1 \pm 0.8$	$5.9 \pm 1.1$	$10.7 \pm 1.8$	$23.3 \pm 3.7$
tW	$1.37 \pm 0.82$	$0.18 \pm 0.26$	$0.01 \pm 0.12$	$0.67 \pm 0.64$	$0.26 \pm 0.42$	$0.39 \pm 0.4$
WtZ	$2.3 \pm 1.2$	$0.6 \pm 0.3$	$0.3 \pm 0.2$	$0.3 \pm 0.2$	$1.1 \pm 0.6$	$0.6 \pm 0.3$
VVV	$12.4 \pm 0.4$	$0.9 \pm 0.1$	$0.4 \pm 0.1$	$0.13 \pm 0.02$	$0.14 \pm 0.03$	$0.02 \pm 0.0$
VH	$40 \pm 6$	$2.6 \pm 1.4$	$0.9 \pm 0.8$	$0.7 \pm 0.8$	$0.4 \pm 0.6$	$0.01 \pm 0.0$
t̄t	$12.1 \pm 1.6$	$2.9 \pm 0.6$	$2.5 \pm 0.5$	$2.8 \pm 0.5$	$11.2 \pm 1.5$	$10.9 \pm 1.4$
t̄tH	$0.24 \pm 0.03$	$0.05 \pm 0.01$	$0.04 \pm 0.01$	$0.06 \pm 0.01$	$0.20 \pm 0.03$	$0.13 \pm 0.0$
Total	$5100 \pm 110$	$227 \pm 12$	$87 \pm 6$	$56.7 \pm 4.4$	$76 \pm 9$	$52.5 \pm 4.2$

Table 8: Post-fit yields in each of the 1-jet fit regions.

793

A post-fit summary plot of the 1-jet fitted regions is shown in Figure 11.2:

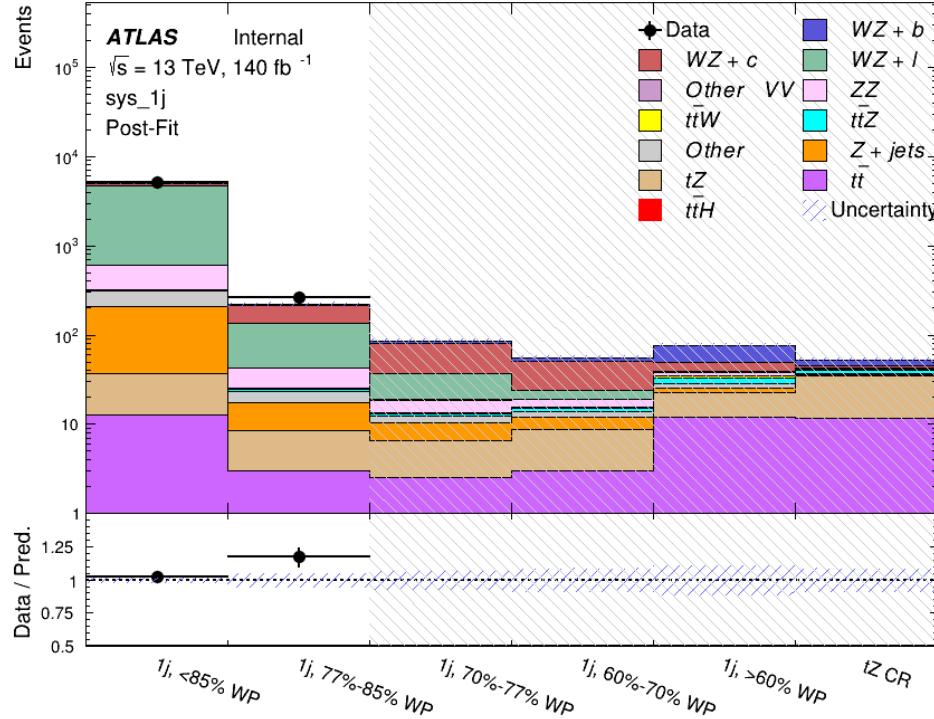


Figure 11.2: Post-fit summary of the 1-jet fit regions.

794 As described in Section 16, there are 226 systematic uncertainties that are considered  
 795 as NPs in the fit. These NPs are constrained by Gaussian or log-normal probability density  
 796 functions. The latter are used for normalisation factors to ensure that they are always positive.  
 797 The expected number of signal and background events are functions of the likelihood. The prior  
 798 for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its  
 799 nominal value.

800 The impact of each NP is calculated by performing the fit with the parameter of interest held  
 801 fixed, varied from its fitted value by its uncertainty, and calculating  $\Delta\mu$  relative to the baseline

802 fit. The impact of the most significant sources of systematic uncertainties is summarized in Table

803 [9](#).

Uncertainty Source	$\Delta\mu$	
WZ + light cross-section	0.13	-0.12
WZ + c cross-section	-0.10	0.12
Jet Energy Scale	0.08	0.13
tZ cross-section	-0.10	0.10
Jet Energy Resolution	-0.10	0.10
Luminosity	-0.08	0.09
Other Diboson + b cross-section	-0.07	0.07
Flavor tagging	0.05	0.05
t <bar>t} cross-section</bar>	-0.05	0.05
WZ cross-section - QCD scale	-0.04	0.03
Total Systematic Uncertainty	0.28	0.32

Table 9: Summary of the most significant sources of systematic uncertainty on the measurement of WZ+b with exactly one associated jet.

804 The ranking and impact of those nuisance parameters with the largest contribution to the

805 overall uncertainty is shown in Figure [11.3](#).

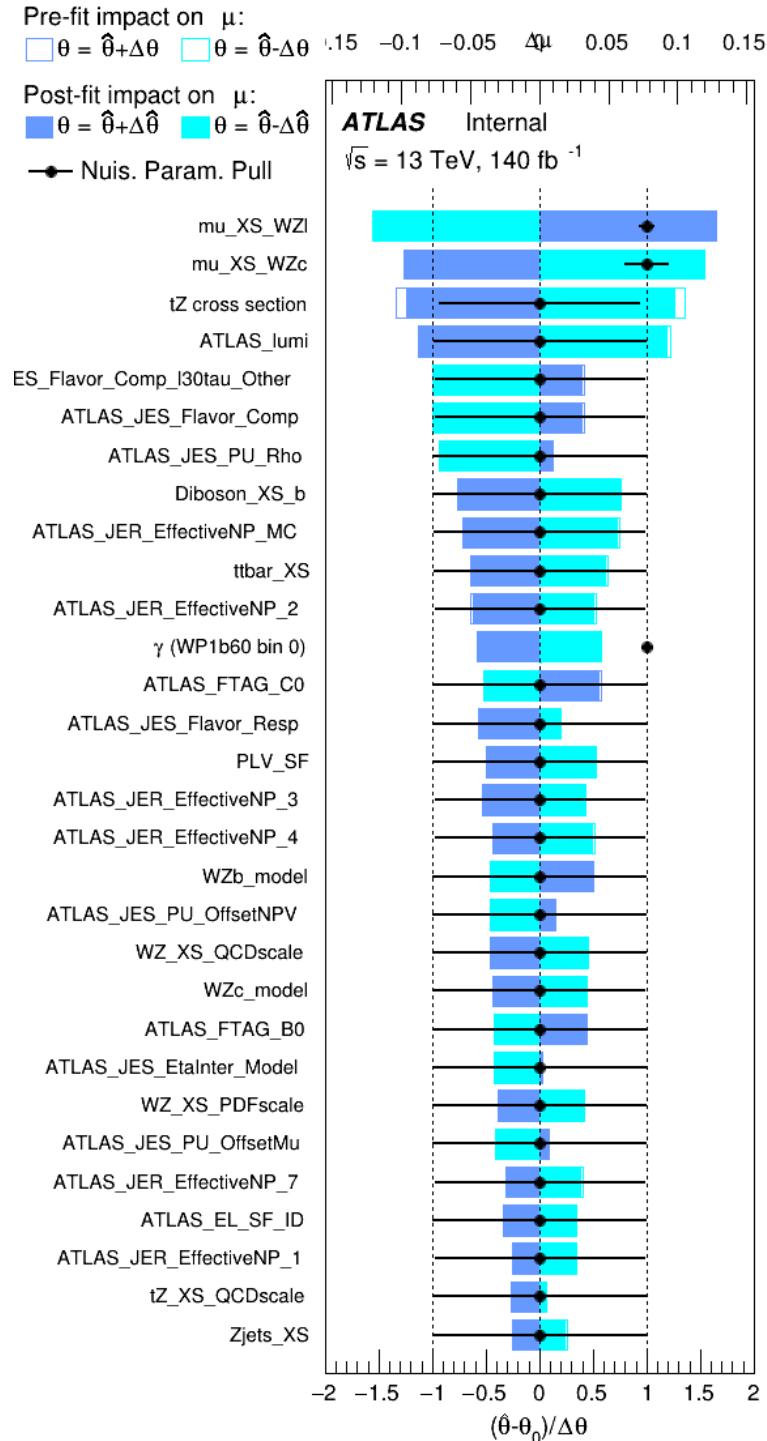


Figure 11.3: Impact of systematic uncertainties on the signal-strength of  $WZ + b$  for events with exactly one jet

806        The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the  
 807        shape of the fit regions depends heavily on the modeling of the jets. The other major sources  
 808        of uncertainty come from background modelling and cross-section uncertainty. The pie charts  
 809        in Figure 11.4 show that for the modelling uncertainties that contribute most correspond to the  
 810        most significant backgrounds.

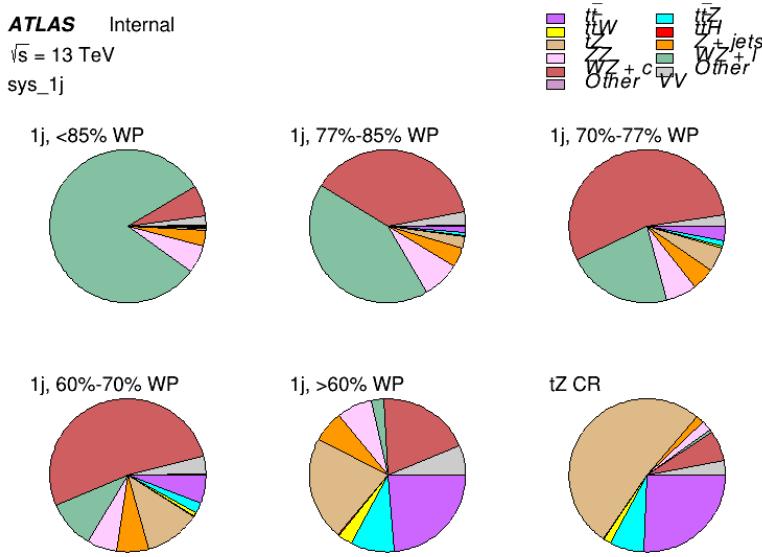


Figure 11.4: Post-fit background composition of the fit regions.

811        The correlations between these nuisance parameters are summarized in Figure 11.5.

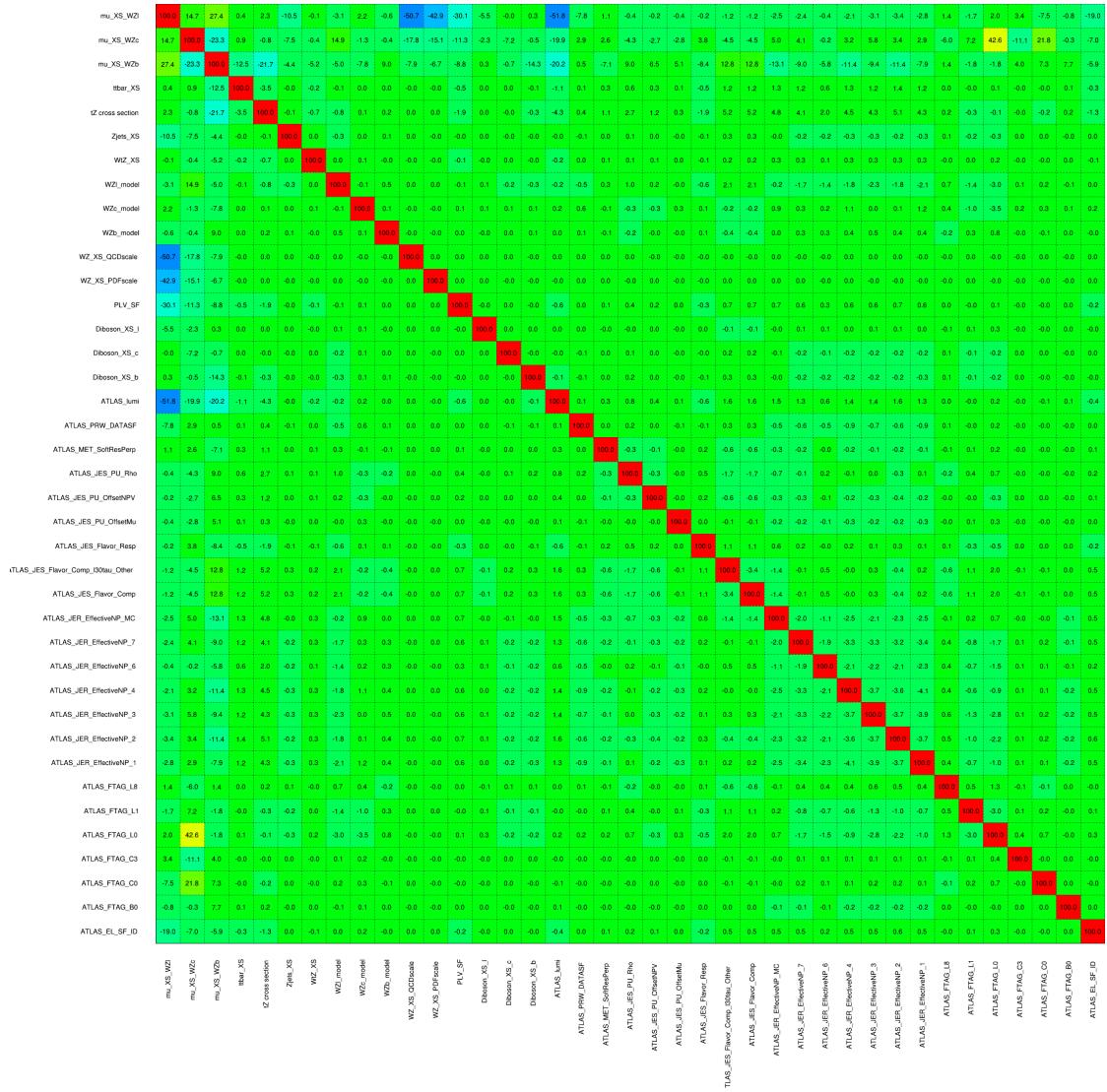


Figure 11.5: Correlations between nuisance parameters

812     The negative correlations between  $\mu_{WZ+charm}$  and  $\mu_{WZ+b}$  and  $\mu_{WZ+light}$  are expected:  
813     WZ + charm is present in both the WZ + b and WZ + light enriched regions, therefore increasing  
814     the fraction of charm requires increasing the fraction of WZ + b and WZ + light. This reasoning  
815     also explains the positive correlation between  $\mu_{WZ+b}$  and  $\mu_{WZ+light}$ .

816 Two of the major backgrounds in the region with the highest purity of WZ + b are tZ and  
817 Other VV + b, explaining the negative correlations between  $\mu_{WZ+b}$  and the tZ cross section, and  
818 the VV + b cross section.

819 The high correlation between the luminosity and  $\mu_{WZ+light}$  arises from the fact that the  
820 uncertainty on  $\mu_{WZ+light}$  is very low (around 4%). Small changes in luminosity cause a change  
821 in the yield of WZ + light that is large compared to its uncertainty, producing a large correlation  
822 between these two parameters.

823 **11.2 2-jet Fit Results**

824 **The results of the fit are currently blinded.**

825 Pre-fit yields in each of the 2-jet fit regions are shown in Figure 11.2.

	2j, <85% WP	2j, 77%-85% WP	2j, 70%-77% WP	2j, 60%-70% WP	2j, >60% WP	tZ CR 2j
WZ + b - 2j	$3.1 \pm 1.6$	$6.7 \pm 0.5$	$5.6 \pm 0.4$	$8.0 \pm 0.6$	$31 \pm 2$	$14 \pm 1$
WZ + c - 2j	$180 \pm 20$	$54 \pm 6$	$41 \pm 3$	$24 \pm 3$	$11 \pm 2$	$4.8 \pm 0.6$
WZ + l - 2j	$1250 \pm 150$	$90 \pm 14$	$18 \pm 3$	$5.8 \pm 1.4$	$1.4 \pm 0.4$	$0.25 \pm 0.1$
WZ + b - 1j	$3.4 \pm 0.6$	$1.52 \pm 0.35$	$1.58 \pm 0.23$	$1.95 \pm 0.39$	$6.7 \pm 1.1$	$1.9 \pm 0.6$
WZ + c - 1j	$56 \pm 14$	$17.6 \pm 4.0$	$8.6 \pm 2.2$	$6.3 \pm 1.8$	$3.0 \pm 0.9$	$0.7 \pm 0.2$
WZ + l - 1j	$427 \pm 120$	$24 \pm 7$	$4.7 \pm 2.3$	$1.6 \pm 0.7$	$0.3 \pm 0.2$	$0.01 \pm 0.0$
WZ - Other	$129 \pm 29$	$6.1 \pm 4.6$	$1.2 \pm 0.3$	$0.3 \pm 0.2$	$2.9 \pm 0.5$	$3.6 \pm 0.6$
Other VV	$7.63 \pm 0.63$	$0.6 \pm 0.5$	$0.16 \pm 0.03$	$0.01 \pm 0.01$	$0.1 \pm 0.1$	$0.1 \pm 0.1$
ZZ	$135 \pm 20$	$14.1 \pm 3.2$	$4.7 \pm 0.8$	$4.0 \pm 0.6$	$4.1 \pm 0.7$	$3.1 \pm 0.5$
t̄tW	$0.8 \pm 0.2$	$0.4 \pm 0.1$	$0.5 \pm 0.1$	$0.7 \pm 0.2$	$4.3 \pm 0.6$	$3.9 \pm 0.6$
t̄tZ	$14.7 \pm 2.2$	$5.6 \pm 0.8$	$4.5 \pm 0.7$	$6.5 \pm 1.1$	$25.4 \pm 4.0$	$21.9 \pm 3.4$
rare Top	$0.14 \pm 0.04$	$0.07 \pm 0.03$	$0.03 \pm 0.02$	$0.09 \pm 0.03$	$0.37 \pm 0.07$	$0.6 \pm 0.1$
t̄tWW	$0.04 \pm 0.03$	$0.02 \pm 0.02$	$0.01 \pm 0.01$	$0.01 \pm 0.00$	$0.03 \pm 0.03$	$0.01 \pm 0.0$
Z + jets	$110.0 \pm 22.9$	$9.6 \pm 2.0$	$2.1 \pm 0.50$	$1.6 \pm 0.4$	$5.1 \pm 1.1$	$1.5 \pm 0.3$
W + jets	$0.0 \pm 0.0$	$0.0 \pm 0.0$	$0.0 \pm 0.0$	$0.0 \pm 0.0$	$0.0 \pm 0.0$	$0.0 \pm 0.0$
V + $\gamma$	$25 \pm 18$	$0.5 \pm 0.2$	$0.1 \pm 0.1$	$0.1 \pm 0.1$	$0.0 \pm 0.02$	$0.05 \pm 0.0$
tZ	$15.9 \pm 2.9$	$6.9 \pm 1.3$	$5.1 \pm 1.0$	$8.0 \pm 1.5$	$18.7 \pm 3.2$	$36.4 \pm 6.1$
tW	$0.9 \pm 0.7$	$0.2 \pm 0.3$	$0.0 \pm 0.1$	$0.0 \pm 0.0$	$0.8 \pm 0.6$	$0.2 \pm 0.2$
WtZ	$4.9 \pm 2.5$	$1.5 \pm 0.8$	$1.1 \pm 0.6$	$1.3 \pm 0.7$	$4.6 \pm 2.4$	$3.3 \pm 1.7$
VVV	$7.4 \pm 0.3$	$1.0 \pm 0.1$	$0.4 \pm 0.1$	$0.2 \pm 0.1$	$0.13 \pm 0.03$	$0.04 \pm 0.0$
VH	$19.5 \pm 4.2$	$2.8 \pm 1.6$	$0.7 \pm 0.7$	$0.1 \pm 0.2$	$0.0 \pm 0.0$	$0.0 \pm 0.0$
t̄t	$0.7 \pm 0.4$	$0.1 \pm 0.1$	$0.05 \pm 0.06$	$0.15 \pm 0.13$	$0.8 \pm 0.5$	$2.3 \pm 1.2$
t̄t	$6.8 \pm 1.0$	$2.4 \pm 0.5$	$1.8 \pm 0.4$	$3.3 \pm 0.6$	$8.4 \pm 1.2$	$13.6 \pm 1.1$
t̄tH	$0.4 \pm 0.1$	$0.2 \pm 0.1$	$0.16 \pm 0.02$	$0.23 \pm 0.03$	$0.94 \pm 0.11$	$1.03 \pm 0.1$
Total	$2580 \pm 160$	$229 \pm 24$	$89 \pm 13$	$69 \pm 11$	$120 \pm 15$	$108 \pm 11$

Table 10: Pre-fit yields in each of the 2-jet fit regions.

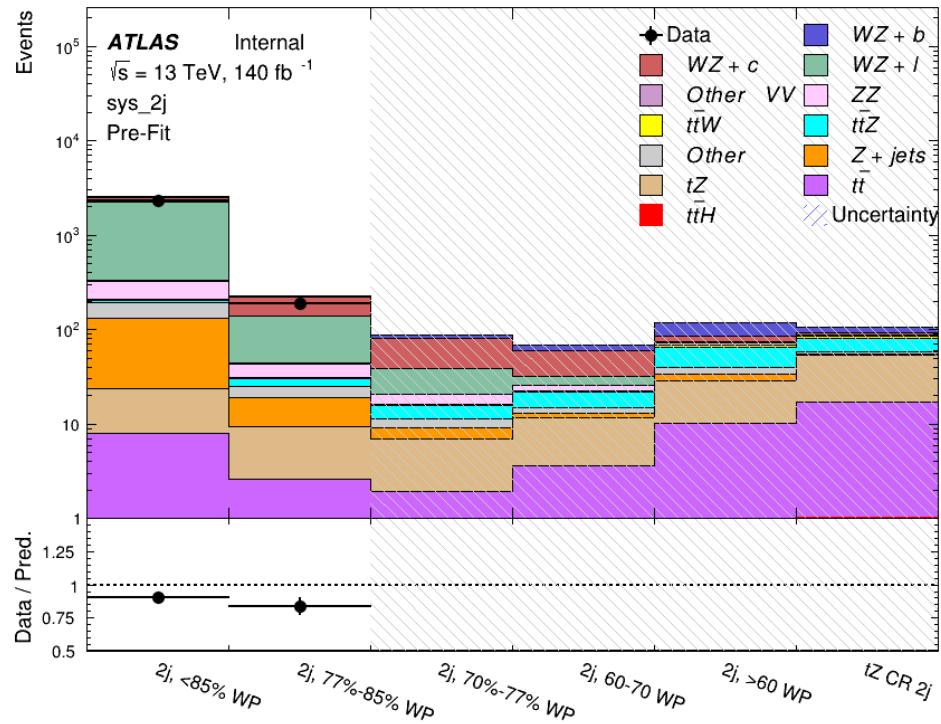


Figure 11.6: Pre-fit summary of the 2-jet fit regions.

826

The post-fit yields in each region are summarized in Figure 11.2.

	2j, <85% WP	2j, 77%-85% WP	2j, 70%-77% WP	2j, 60%-70% WP	2j, >60% WP	tZ CR 2j
WZ + b	$13 \pm 6$	$6.7 \pm 2.9$	$5.8 \pm 2.5$	$8.0 \pm 3.5$	$31 \pm 13$	$14 \pm 5$
WZ + c	$260 \pm 60$	$77 \pm 15$	$41 \pm 8$	$26 \pm 5$	$10.9 \pm 2.4$	$4.8 \pm 1.1$
WZ + l	$1860 \pm 90$	$90 \pm 12$	$17.6 \pm 2.8$	$5.8 \pm 1.3$	$1.4 \pm 0.4$	$0.3 \pm 0.2$
WZ + b - 1j	$3.4 \pm 0.6$	$1.52 \pm 0.35$	$1.58 \pm 0.23$	$1.95 \pm 0.39$	$6.7 \pm 1.1$	$1.9 \pm 0.6$
WZ + c - 1j	$56 \pm 14$	$17.6 \pm 4.0$	$8.6 \pm 2.2$	$6.3 \pm 1.8$	$3.0 \pm 0.9$	$0.7 \pm 0.2$
WZ + l - 1j	$427 \pm 120$	$24 \pm 7$	$4.7 \pm 2.3$	$1.6 \pm 0.7$	$0.3 \pm 0.2$	$0.01 \pm 0.0$
WZ - Other	$129 \pm 29$	$6.1 \pm 4.6$	$1.2 \pm 0.3$	$0.3 \pm 0.2$	$2.9 \pm 0.5$	$3.6 \pm 0.6$
Other VV	$7.6 \pm 0.6$	$0.3 \pm 0.3$	$0.3 \pm 0.1$	$0.1 \pm 0.06$	$0.03 \pm 0.02$	$0.1 \pm 0.1$
ZZ	$145 \pm 30$	$11.3 \pm 4.4$	$2.7 \pm 1.6$	$1.0 \pm 0.3$	$4.0 \pm 0.1$	$2.4 \pm 0.1$
t̄tW	$0.8 \pm 0.2$	$0.4 \pm 0.1$	$0.54 \pm 0.12$	$0.74 \pm 0.15$	$4.3 \pm 0.6$	$3.9 \pm 0.6$
t̄tZ	$14.7 \pm 2.2$	$5.6 \pm 0.8$	$4.5 \pm 0.7$	$6.5 \pm 1.0$	$25.4 \pm 3.9$	$21.9 \pm 3.0$
rare Top	$0.14 \pm 0.04$	$0.07 \pm 0.03$	$0.03 \pm 0.02$	$0.09 \pm 0.03$	$0.4 \pm 0.1$	$0.6 \pm 0.1$
t̄tWW	$0.04 \pm 0.03$	$0.02 \pm 0.02$	$0.01 \pm 0.01$	$0.01 \pm 0.01$	$0.03 \pm 0.03$	$0.01 \pm 0.0$
Z + jets	$110 \pm 23$	$9.6 \pm 2.0$	$2.1 \pm 0.5$	$1.6 \pm 0.4$	$5.1 \pm 1.1$	$1.5 \pm 0.3$
W + jets	$0.0 \pm 0.0$	$0.0 \pm 0.0$				
V + γ	$25 \pm 19$	$0.5 \pm 0.2$	$0.1 \pm 0.1$	$0.13 \pm 0.14$	$0.0 \pm 0.02$	$0.05 \pm 0.0$
tZ	$15.9 \pm 2.7$	$6.9 \pm 1.2$	$5.1 \pm 0.9$	$8.0 \pm 1.4$	$18.7 \pm 3.0$	$36 \pm 6$
tW	$0.1 \pm 0.7$	$0.2 \pm 0.3$	$0.0 \pm 0.1$	$0.0 \pm 0.0$	$0.8 \pm 0.6$	$0.2 \pm 0.2$
WtZ	$4.9 \pm 2.5$	$1.5 \pm 0.8$	$1.1 \pm 0.6$	$1.3 \pm 0.7$	$4.6 \pm 2.3$	$3.3 \pm 1.7$
VVV	$7.4 \pm 0.3$	$1.0 \pm 0.1$	$0.36 \pm 0.03$	$0.19 \pm 0.03$	$0.13 \pm 0.03$	$0.04 \pm 0.0$
VH	$19 \pm 4$	$2.8 \pm 1.6$	$0.7 \pm 0.7$	$0.1 \pm 0.2$	$0.0 \pm 0.0$	$0.0 \pm 0.0$
t̄t	$6.8 \pm 1.0$	$2.4 \pm 0.5$	$1.8 \pm 0.4$	$3.3 \pm 0.6$	$8.4 \pm 1.2$	$13.6 \pm 1.7$
t̄tH	$0.40 \pm 0.05$	$0.19 \pm 0.03$	$0.16 \pm 0.02$	$0.23 \pm 0.03$	$0.94 \pm 0.11$	$1.03 \pm 0.1$
Total	$2580 \pm 60$	$229 \pm 11$	$89 \pm 6$	$69.1 \pm 4.1$	$120 \pm 10$	$108 \pm 6$

Table 11: Post-fit yields in each of the 2-jet fit regions.

827

A post-fit summary of the fitted regions is shown in Figure 11.7:

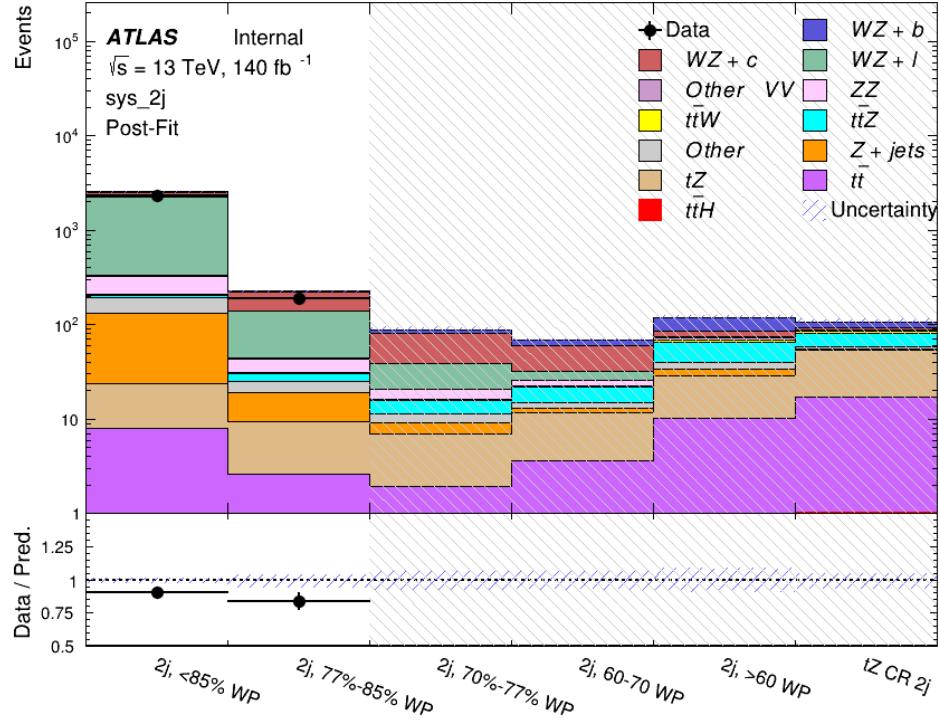


Figure 11.7: Post-fit summary of the fit over 2-jet regions.

828        The same set of systematic uncertainties consider for the 1-jet fit are included in the 2-jet  
 829        fit as well. The impact of the most significant systematic uncertainties is summarized in Table  
 830        [12](#).

Uncertainty Source	$\Delta\mu$	
WZ + c cross-section	-0.1	0.14
Luminosity	-0.11	0.12
tZ cross-section	-0.11	0.11
Jet Energy Scale	-0.11	0.11
ttZ cross-section - QCD scale	-0.08	0.09
WZ + l cross-section	0.08	-0.07
WtZ cross-section	-0.07	0.07
Flavor tagging	0.05	0.05
Other VV + b cross-section	-0.05	0.05
Jet Energy Resolution	-0.04	0.04
Total	0.29	0.31

Table 12: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b 2-jet events.

831        The ranking and impact of those nuisance parameters with the largest contribution to the  
 832        overall uncertainty is shown in Figure 11.8.

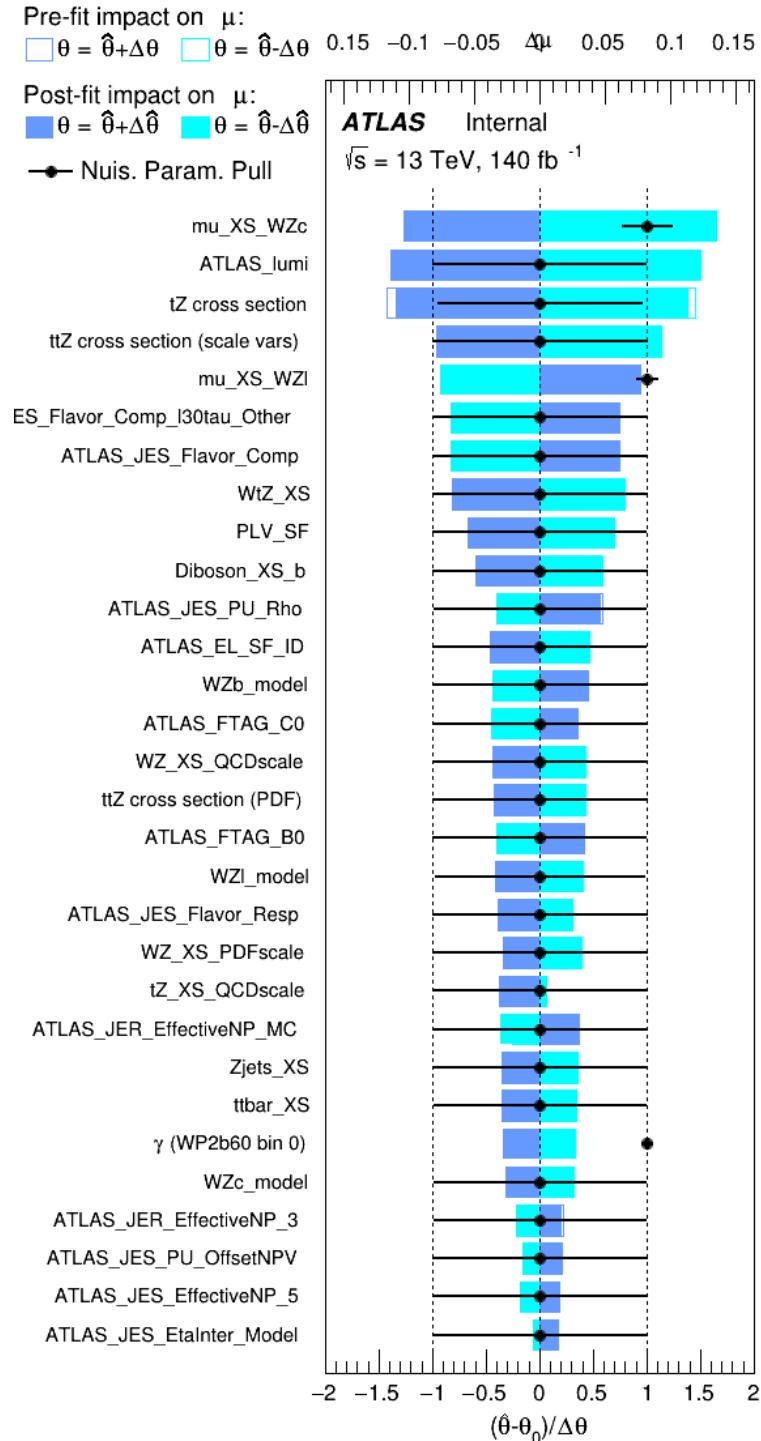


Figure 11.8: Impact of systematic uncertainties on the signal-strength of  $WZ + b$  in 2-jet events.

833        The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the  
 834        shape of the fit regions depends heavily on the modeling of the jets. The other major sources  
 835        of uncertainty come from background modelling and cross-section uncertainty. The pie charts  
 836        in Figure 11.9 show that for the modelling uncertainties that contribute most correspond to the  
 837        most significant backgrounds.

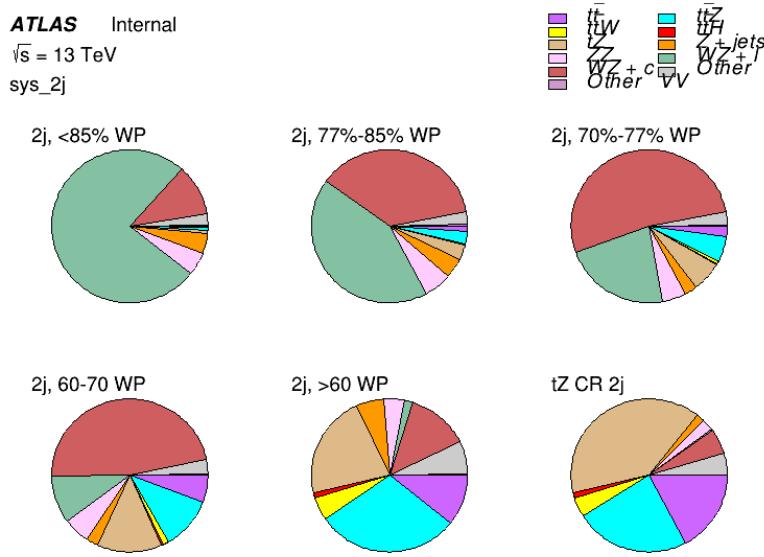


Figure 11.9: Post-fit background composition of the 2-jet fit regions.

838        The correlations between these nuisance parameters are summarized in Figure 11.10.



Figure 11.10: Correlations between nuisance parameters in the 2-jet fit

As in the 1-jet case, no significant, unexpected correlations are found between nuisance

---

**Part V****842 Differential Studies of  $t\bar{t}H$  Multilepton****843 12 Data and Monte Carlo Samples****844 12.1 Data Samples**

845 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through  
846 2018, which represents an integrated luminosity of  $139 \text{ fb}^{-1}$  [15] and an energy of  $\sqrt{s} = 13$   
847 TeV. All data used in this analysis was included in one of the Good Run Lists verified by Data  
848 Quality checks [16].

**849 12.2 Monte Carlo Samples**

850 Several Monte Carlo (MC) generators were used to simulate both signal and background pro-  
851 cesses. For all of these, the effects of the ATLAS detector are simulated in GEANT4 [17]. The  
852 specific event generator used for each of these MC samples is listed in Table 13. A Higgs mass  
853 of 125 GeV is assumed in all simulations.

Table 13: 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
$t\bar{t}H$	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO [21]
	(MG5_AMC)	(NLO)	(HERWIG++)	(CT10 [19])
$t\bar{t}W$	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
	(SHERPA 2.1.1)	(LO multileg)	(SHERPA)	(NNPDF 3.0 NLO)
$t\bar{t}(Z/\gamma^* \rightarrow ll)$	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
$t\bar{t}$	POWHEG-BOX v2 [24]	NLO	PYTHIA 8	NNPDF 3.0 NLO
$t\bar{t}\gamma$	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	NLO	HERWIG++	CT10
	(SHERPA 2.1.1)	(LO multileg)	(SHERPA)	(NNPDF 3.0 NLO)
$tWZ$	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
$t\bar{t}t, t\bar{t}t\bar{t}$	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
$t\bar{t}W^+W^-$	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	POWHEG-BOX v1 [25]	NLO	PYTHIA 6	CT10
qqVV, VVV				
$Z \rightarrow l^+l^-$	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

854       The signal sample ( $t\bar{t}H$ ) is modelled at NLO with POWHEG-BOX v2 using the NNPDF2.0  
 855       parton distribution function (PDF) [40]. Parton showering and hadronisation were modelled  
 856       with PYTHIA 8.2 [41]. The  $t\bar{t}H$  sample is normalized to a cross-section of  $507^{+35}_{-50}$  fb based on  
 857       NLO calculations. Uncertainties are based on varying the QCD factorisation and renormalisation  
 858       scale, as well as uncertainties in the PDF and  $\alpha_s$ .

859       The  $t\bar{t}W$  background is simulated using Sherpa 2.2.1 with the NNPDF3.0 NLO PDF. The  
 860       matrix element is calculated with up to one additional parton at NLO, and up to two at LO. As  
 861       explained in detail in [7], the  $t\bar{t}W$  contribution predicted by MC is found disagree significantly

862 with what is observed in data. While an effort is currently being undertaken to measure  $t\bar{t}W$  more  
 863 accurately, the approach used by the  $79.8 \text{ fb}^{-1}$   $t\bar{t}H$  analysis is used here: A normalization  
 864 factor of 1.68 is applied to the MC estimate of  $t\bar{t}W$  and additional systematic uncertainties on  
 865  $t\bar{t}W$  are included to account for this modelling discrepancy, as outlined in Section 16.

866 The  $t\bar{t}(Z/\gamma^*)$  process is simulated with the **MADGRAPH5\_AMC@NLO** generator, using  
 867 NNPDF3.0. Diboson processes are generated with **SHERPA 2.2.2** at NLO precision for one extra  
 868 parton, and at LO for up to three extra partons.

869 The “fake”, or non-prompt, background comes primarily from leptons originating from  
 870 hadron decays, leptons with missidentified charge, and photon conversions. While the main  
 871  $t\bar{t}H$  analysis is currently refining a data-driven approach for estimating the contribution of  
 872 events with non-prompt leptons, at the time of this note this strategy has not been completely  
 873 developed for the full Run-2 dataset. Therefore, the non-prompt contribution is estimated with  
 874 MC, while applying normalization corrections and systematic uncertainties derived from data  
 875 driven techniques developed for the  $79.8 \text{ fb}^{-1}$   $t\bar{t}H/t\bar{t}W$  analysis [7].

876 The primary contribution to the non-prompt lepton background is from  $t\bar{t}$  production, with  
 877  $V+jets$  and single-top as much smaller sources. Estimation of this background is done primarily  
 878 using an inclusive  $t\bar{t}$  sample, with corrections applied based on data driven methods. This sample  
 879 is generated using **POWHEG**, with **PYTHIA8** performing the parton shower and fragmentation.  
 880 Likelihood fits over several control regions enriched with these non-prompt backgrounds are fit  
 881 to data in order to derive normalization factors for these backgrounds. The specific normalization

882 factors and uncertainties applied to the non-prompt contributions are listed in Section 16.

883 Other processes, such as  $tH$ ,  $tZ$ ,  $t\bar{t}WW$  and  $t\bar{t}t\bar{t}$ , are expected to make minor contributions  
884 to the total background. The generators and setting used for these backgrounds are summarized  
885 in Table 13.

## 886 13 Object Reconstruction

887 All analysis channels considered in this note share a common object selection for leptons and  
888 jets, as well as a shared trigger selection. Events are required to be selected by dilepton triggers.  
889 The  $p_T$  thresholds of the dilepton trigger on two electrons were 12 GeV in 2015, 17 GeV in 2016,  
890 and 24 GeV in 2017 and 2018, while for the dimuon triggers the  $p_T$  thresholds on the leading  
891 (sub-leading) muon were 18 GeV (8 GeV) in 2015, and 22 GeV (8 GeV) in 2016-2018. For the  
892 electron+muon triggers, the  $p_T$  thresholds on the electron (muon) were 17 GeV (14 GeV) for all  
893 datasets.

894 Electron candidates are reconstructed from energy clusters in the electromagnetic calor-  
895 imeter that are associated with charged particle tracks reconstructed in the inner detector  
896 [42]. Electron candidates are required to have  $p_T > 10$  GeV and  $|\eta_{\text{cluster}}| < 2.47$ . Can-  
897 didates in the transition region between different electromagnetic calorimeter components,  
898  $1.37 < |\eta_{\text{cluster}}| < 1.52$ , are rejected. A multivariate likelihood discriminant combining shower  
899 shape and track information is used to distinguish prompt electrons from nonprompt leptons,  
900 such as those originating from hadronic showers.

901 To further reduce the non-prompt contribution, the track of each electron is required to  
 902 originate from the primary vertex; requirements are imposed on the transverse impact parameter  
 903 significance ( $|d_0|/\sigma_{d_0}$ ) and the longitudinal impact parameter ( $|\Delta z_0 \sin \theta_\ell|$ ). Muon candidates  
 904 are reconstructed by combining inner detector tracks with track segments or full tracks in the  
 905 muon spectrometer [27]. Muon candidates are required to have  $p_T > 10$  GeV and  $|\eta| < 2.5$ .

906 All leptons are required to pass a non-prompt BDT selection developed by the main  
 907  $t\bar{t}H/t\bar{t}W$  analysis, described in detail in [7]. Optimized working points and scale factors for this  
 908 BDT are taken from that analysis. This BDT and the WPs used are summarized in Appendix  
 909 [A](#),

910 Jets are reconstructed from calibrated topological clusters built from energy deposits in the  
 911 calorimeters [43], using the anti- $k_t$  algorithm with a radius parameter  $R = 0.4$ . Particle Flow, or  
 912 PFlow, jets are used in the analysis, which are hadronic objects reconstructed using information  
 913 from both the tracker and the calorimeter. Jets with energy contributions likely arising from noise  
 914 or detector effects are removed from consideration [44], and only jets satisfying  $p_T > 25$  GeV  
 915 and  $|\eta| < 2.5$  are used in this analysis. For jets with  $p_T < 60$  GeV and  $|\eta| < 2.4$ , a jet-track  
 916 association algorithm is used to confirm that the jet originates from the selected primary vertex,  
 917 in order to reject jets arising from pileup collisions [30].

918 Each analysis channel used in this analysis includes b-jets in the final state. These are  
 919 identified using the DL1r b-tagging algorithm, which uses jet vertex and kinematic information  
 920 to distinguish heavy and light flavored jets. These features are used as inputs to a neural network,

921 the output of which is used to form calibrated working points (WPs) based on how likely a jet is  
922 to have originated from a b-quark. This analysis uses the 70% DL1r WP - implying an efficiency  
923 of 70% for truth b-jets - for selecting b-tagged jets.

924 Because all  $t\bar{t}H$  – ML channels considered include multiple neutrinos, missing transverse  
925 energy ( $E_T^{\text{miss}}$ ) is present in each event. The missing transverse momentum vector is defined as  
926 the inverse of the sum of the transverse momenta of all reconstructed physics objects as well  
927 as remaining unclustered energy, the latter of which is estimated from low- $p_T$  tracks associated  
928 with the primary vertex but not assigned to a hard object [45].

929 To avoid double counting objects and remove leptons originating from decays of hadrons,  
930 overlap removal is performed in the following order: any electron candidate within  $\Delta R = 0.1$  of  
931 another electron candidate with higher  $p_T$  is removed; any electron candidate within  $\Delta R = 0.1$   
932 of a muon candidate is removed; any jet within  $\Delta R = 0.2$  of an electron candidate is removed; if  
933 a muon candidate and a jet lie within  $\Delta R = \min(0.4, 0.04 + 10[\text{GeV}] / p_T(\text{muon}))$  of each other,  
934 the jet is kept and the muon is removed if the jet has three or more tracks, otherwise the muon is  
935 kept and the jet is removed..

## 936 14 Higgs Momentum Reconstruction

937 Reconstructing the momentum of the Higgs boson is a particular challenge for channels with  
938 leptons in the final state: Because all channels include at least two neutrinos in the final state, the

939 Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted  
940 by a neural network when provided with the kinematics of the Higgs Boson decay products - as  
941 verified by studies detailed in Appendix C.3. With this in mind, several layers of MVAs are used  
942 to reconstruction the Higgs momentum:

943       The first layer is a model designed to select which jets are most likely to be the b-jets that  
944       came from the top decay, detailed in Section 14.2. As described in Section 14.3, the kinematics  
945       of these jets and possible Higgs decay products are fed into the second layer, which is designed to  
946       identify the decay products of the Higgs Boson itself. The kinematics of the particles this layer  
947       identifies as most likely to have originated from the Higgs decay are then fed into yet another  
948       neural-network, which predicts the momentum of the Higgs (14.4). For the 3l channel, because  
949       the Higgs can decay into either one lepton and two jets or two leptons, an additional MVA is  
950       used to determine the decay mode of the Higgs boson in the 3l channel (14.5).

951       Models are trained on Monte Carlo simulations of  $t\bar{t}H$  events generated using MG5\_AMC.  
952       Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training in order to  
953       increase the statistics of the training sample.

954       For all of these models, the Keras neural network framework, with Tensorflow 2.0 as  
955       the backend [**tensorflow**], is used, and the number of hidden layers and nodes are determined  
956       using grid search optimization. Each neural network uses the LeakyReLU activation function,  
957       a learning rate of 0.01, and the Adam optimization algorithm, as alternatives are found to either  
958       decrease or have no impact on performance. Batch normalization is applied after each layer in

order to stabilize the model and decrease training time. For the classification algorithms (b-jet matching, Higgs reconstruction, and 3l decay identification) binary-cross entropy is used as the loss function, while the  $p_T$  reconstruction algorithm uses MSE.

The specific inputs features used for each model are arrived at through a process of trial and error - features considered potentially useful are tried, and those that are found to increase performance are included. While each model includes a relatively large number of features, some using upwards of 30, this inclusive approach is found to maximize the performance of each model while decreasing the variance compared to a reduced number of inputs. Each input feature is validated by comparing MC simulations to  $79.8 \text{ fb}^{-1}$  of data, with the full set of features shown in Section C..

## 14.1 Physics Object Truth Matching

Machine Learning algorithms are trained to identify the decay products of the Higgs Boson using MC simulations of  $t\bar{t}H$  events. The kinematics of the reconstructed physics objects, as well as event level variables such the jet multiplicity and missing energy, used as inputs, with the parent ID taken from the truth record used to label the data. The objects considered include light leptons and jets.

Reconstructed physics objects are matched to particle level objects, in order to identify the parent particle of these reconstructed objects. Reconstructed jets are matched to truth jets based on the requirements that the reco jet and truth jet fall within  $\Delta R < 0.4$ , and the two objects have

978 a  $p_T$  that agrees within 10%. Truth level and reco level leptons are required to have the same  
979 flavor, a  $\Delta R < 0.1$ , and  $p_T$  that agree within 10%. Events where no match can be found between  
980 the particle level decay products and the reconstructed objects are not included in training.

981 Leptons considered as possible Higgs and top decay candidates are required to pass the  
982 selection described in Section ???. For jets, however, it is found that a large fraction that originate  
983 from either the top decay or the Higgs decay fall outside the selection described in Section ???.  
984 Specifically, jets from the Higgs decay tend to be soft, with 32% having  $p_T < 25$  GeV. Therefore  
985 jets with  $p_T < 15$  GeV are considered as possible candidates in the models described below. By  
986 contrast, less than 5% of the jets originating from the Higgs fall below this  $p_T$  threshold. The  
987 jets are found to be well modeled even down to this low  $p_T$  threshold, as shown in Section 15.1.  
988 The impact of using different  $p_T$  selection for the jet candidates is considered in detail in Section  
989 C.6. The overlap removal selection is not applied to the objects considered in the models.

## 990 14.2 b-jet Identification

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

996        Exactly two b-jets are expected in the final state of  $t\bar{t}H - ML$  events. However, in both  
 997        the 3l and 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r  
 998        b-tag working point is used). Therefore, for events which have exactly one, or more than two,  
 999        b-tagged jets, deciding which combination of jets correspond to the top decay is non-trivial.  
 1000      Further, events with 1 b-tagged jet represent just over half of all  $t\bar{t}H - ML$  events. Of those,  
 1001      both b-jets are reconstructed by the detector 75% of the time. Therefore, rather than adjusting  
 1002      the selection to require exactly 2 b-tagged jets, and losing more than half of the signal events, a  
 1003      neural network is used to predict which pair of jets is most likely to correspond to truth b-jets.

1004      Once the network is trained, kinematic variables for all possible pairings of jets are fed into  
 1005      the model, and the pair of jets with the highest output score are taken to be b-jets in successive  
 1006      steps of the analysis.

1007      An alternate approach is considered, where information about all jets in each event are  
 1008      used as the feature set, and the model is tasked with identifying which two originated from the  
 1009      top decay. While this approach is found to underperform the nominal approach, and therefore  
 1010      not used in the analysis, the results are documented in Appendix [C.4](#).

### 1011 **14.2.1 2lSS Channel**

1012     For the 2lSS channel, the input features shown in Table [14](#) are used for training. Here  $j_0$  and  $j_1$   
 1013     are the two jet candidates, while  $l_0$  and  $l_1$  are the two leptons in the event, both ordered by  $p_T$ . jet  
 1014     DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,

1015 where 5 represents the tightest working point and 1 represents the loosest. The variables nJets  
 1016 DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and  
 1017 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 14: Input features used in the b-jet identification algorithm for the 2lSS channel

1018 As there are far more incorrect combinations than correct ones, by a factor of more than  
 1019 20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random  
 1020 sample of 5 million incorrect entries are used for training, along with around 1 million correct  
 1021 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for  
 1022 training.

1023 The difference between the distributions for a few of these features for the "correct" (i.e.  
 1024 both jets are truth b-jets), and "incorrect" combinations are shown in Figure 14.1. The correct and  
 1025 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences  
 1026 in the distributions.

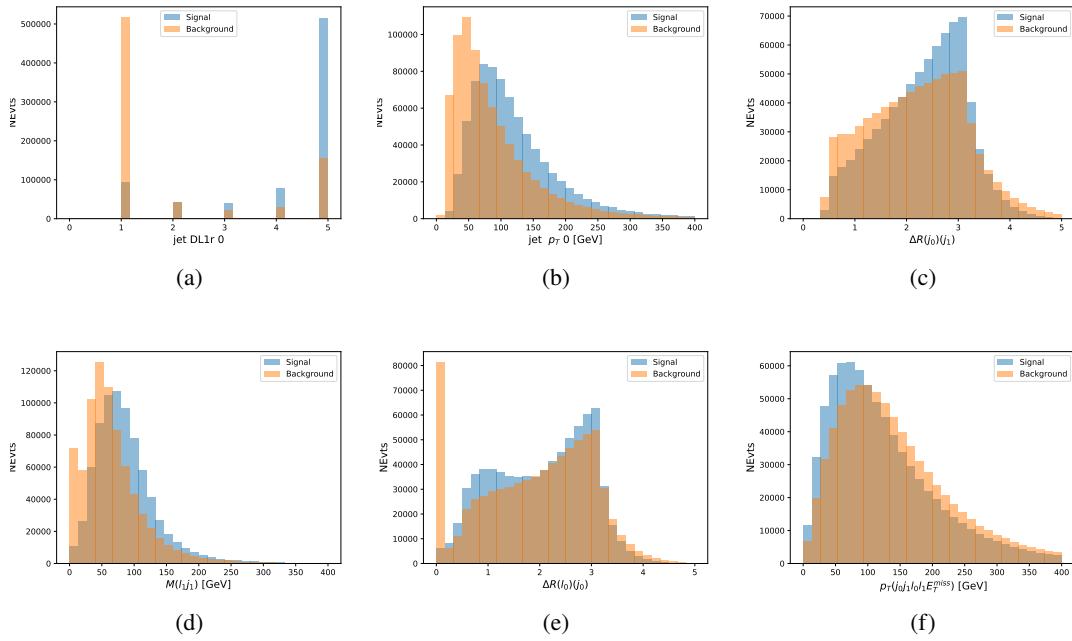


Figure 14.1: Input features for top2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the  $p_T$  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  $E_T^{\text{miss}}$ .

1027            The modeling of these inputs is validated against data, with Figure 14.2 showing good  
 1028            general agreement between data and MC. Plots for the complete list of features can found in  
 1029            Appendix C.

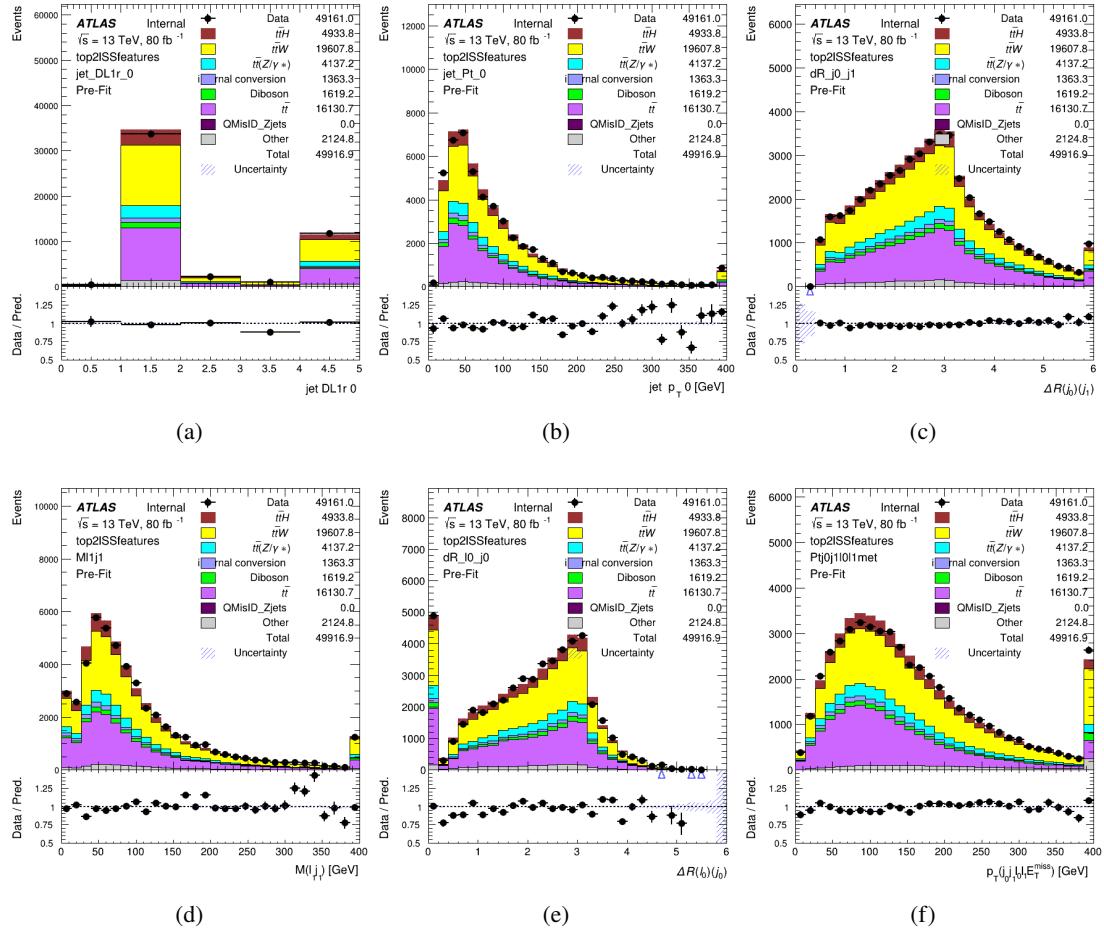


Figure 14.2: Data/MC comparisons of input features for top2lSS training for  $79.8 \text{ fb}^{-1}$  of data. (a) shows the DL1r working point of leading jet, (b) shows the  $p_T$  of the leading jet, (c) shows the  $\Delta R$  of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the  $\Delta R$  of lepton 0 and jet 0, and (f) the  $p_T$  of both jets, both leptons, and the  $E_T^{\text{miss}}$ .

1030 Based on the results of grid search evaluation, the optimal architecture is found to include  
 1031 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as  
 1032 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for  
 1033 the trained model are shown in Figure 14.2.1. The model is found to identify the correct pairing  
 1034 of jets for 73% of 2lSS signal events on test data.

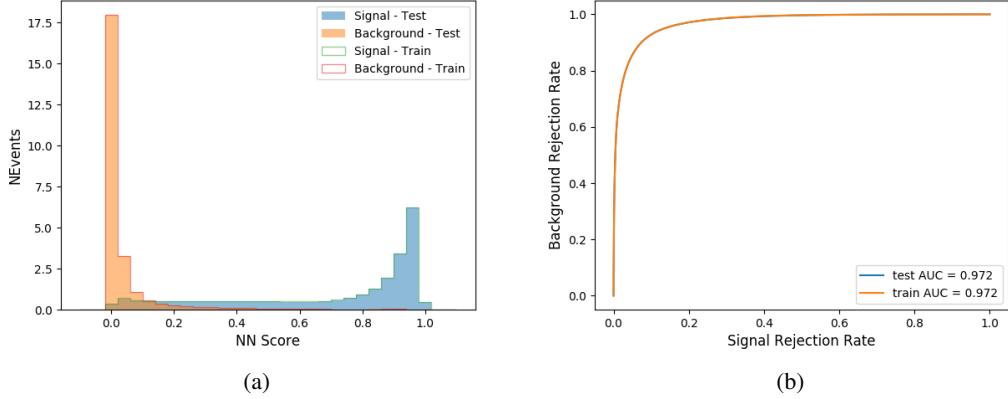


Figure 14.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1035 For point of comparison, a "naive" approach to identifying b-jets is used as well: The two  
 1036 jets which pass the highest DL1r b-tag working point are assumed to be the b-jets from the top  
 1037 decay. In the case that multiple jets meet the same b-tag working point, the jet with higher  $p_T$  is  
 1038 used. This method identifies the correct jet pair 65% of the time.

1039 The accuracy of the model for different b-tagged jet multiplicities, compared to this naive  
 1040 approach, is shown in Table 15.

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 15: Accuracy of the NN in identifying b-jets from tops in 2lSS events, overall and split by the number of b-tagged jets in the event, compared to the accuracy of taking the two highest b-tagged jets.

1041 This suggests that when there are exactly two b-tagged jets in an event, little is gained by  
 1042 using this more sophisticated approach, while for events with 1 or  $\geq 3$  b-tagged jets, the model  
 1043 does provide significant improvements.

1044 **14.2.2 3l Channel**

1045 The input features used in the 3l channel are listed in Table 16, with the same naming convention  
 1046 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 16: Input features for the b-jet identification algorithm in the 3l channel.

1047 A few of these features are shown in Figure 14.4, comparing the distributions for correct  
 1048 and incorrect combinations of jets.

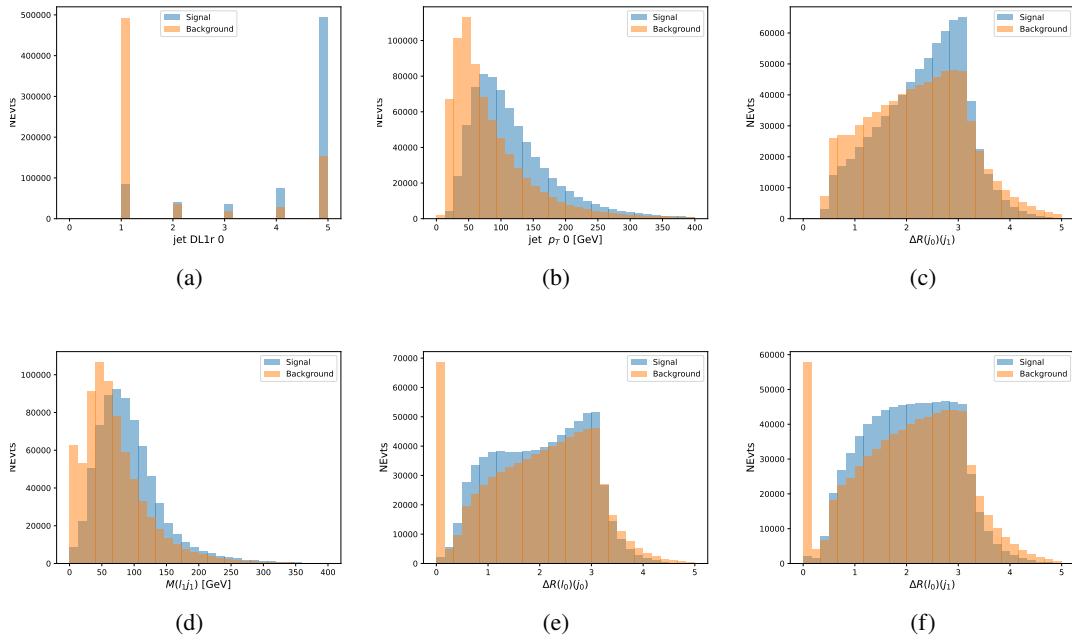


Figure 14.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the  $p_T$  of jet 0, (c)  $\Delta R$  between 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

1049            The modeling of these inputs is validated against data, with Figure 14.5 showing good  
 1050            general agreement between data and MC.

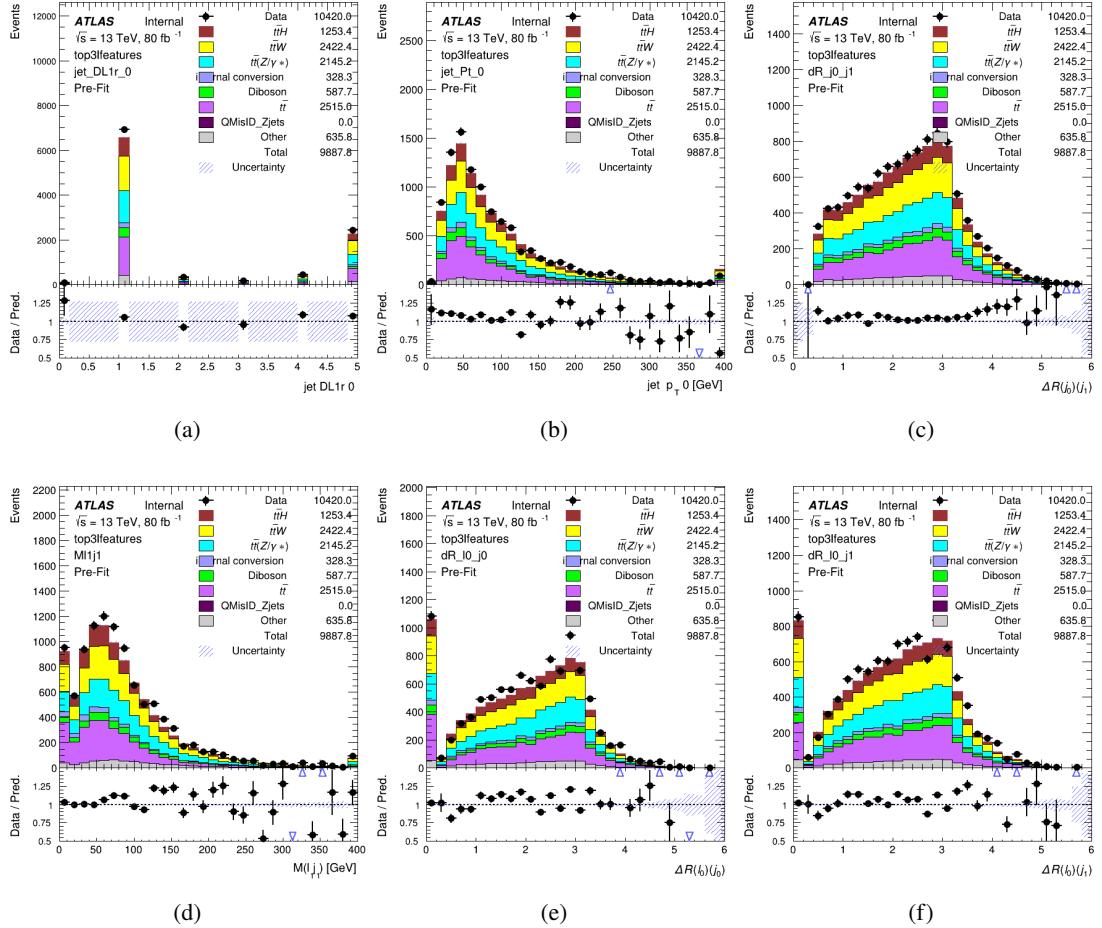


Figure 14.5: Data/MC comparisons of input features for top3l training for  $79.8 \text{ fb}^{-1}$  of data. (a) show the DL1r WP of jet 0, (b) the  $p_T$  of jet 0, (c)  $\Delta R$  between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e)  $\Delta R$  between lepton 0 and jet 0, and (f)  $\Delta R$  between lepton 0 and jet 1

1051 Again, the dataset is downsized to reduce the ratio of correct and incorrect combination

1052 from 20:1, to 5:1. Around 7 million events are used for training, with 10% set aside for testing.

1053 Based on the results of grid search evaluation, the optimal architecture is found to include 5

1054 hidden layers with 60 nodes each. The output score distribution as well as the ROC curve for the

1055 trained model are shown in Figure 14.2.2.

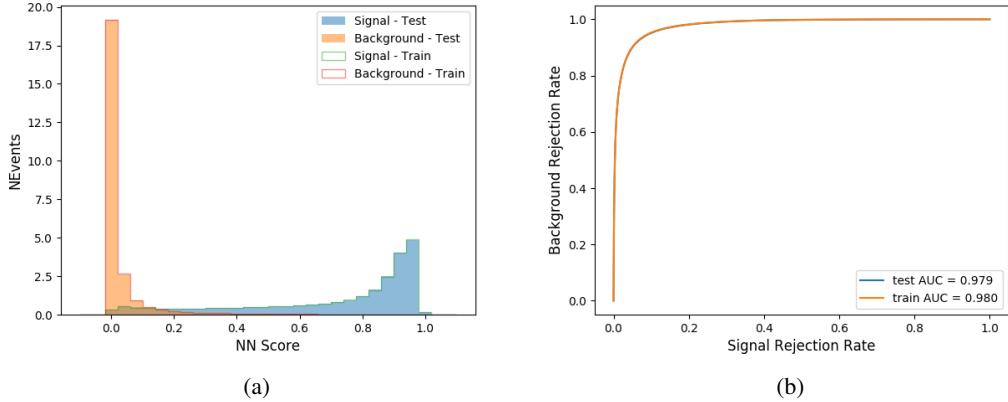


Figure 14.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1056 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal  
 1057 events. The accuracy of the model is summarized in Table 17, once again compared to the naive  
 1058 approach described above.

Table 17: 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%

### 1059 14.3 Higgs Reconstruction

1060 Techniques similar to the b-jet identification algorithms are employed to select the decay products  
 1061 of the Higgs: kinematics of all possible combinations of reconstructed objects are fed into a neural

1062 network to determine which of those is most mostly to be the decay products of the Higgs.

1063 Again separate models are used for the 2lSS and 3l channels, while the 3l channel has  
1064 now been split into two:  $t\bar{t}H$  events with three leptons in the final state include both instances  
1065 where the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the  
1066 Higgs decays to two leptons (and two neutrinos which are not reconstructed).

1067 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In  
1068 the semi-leptonic case the reconstructed decay products consist of two jets and a single lepton.  
1069 For the fully-leptonic case, the decay products include 2 of the three leptons associated with  
1070 the event. For training these models, events are separated into these two categories using truth  
1071 level information. A separate MVA, described in Section 14.5, is used to make this distinction  
1072 at reconstructed level, and determine which model to use.

1073 For all channels, the models described in Section 14.2 are used to identify b-jet candidates,  
1074 whose kinematics are used as additional input features to help identify the Higgs decay products.  
1075 These jets are not considered as possible candidates for the Higgs decay, justified by the fact that  
1076 these models are found to misidentify jets from the Higgs decay as jets from the top decay less  
1077 than 1% of the time.

### 1078 **14.3.1 2lSS Channel**

1079 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural  
1080 network is trained on the kinematics of different combinations of leptons and jets, as well as the

<sup>1081</sup> b-jets identified in Section 14.2, with the specific input features listed in Table 18.

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 heta T	$\Delta R(j_0)(j_1)$
$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$	$M(j_0 j_1)$
$M(l_H j_0)$	$M(l_H j_1)$	$\Delta R(l_H)(b_0)$
$\Delta R(l_H)(b_1)$	$\Delta R(l_T)(b_0)$	$\Delta R(l_T)(b_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1)(l_T)$	$\Delta R(j_0 j_1)(b_0)$
$\Delta R(j_0 j_1)(b_1)$	$\Delta R(j_0)(b_0)$	$\Delta R(j_0)(b_1)$
$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$	$M(j_0 j_1 l_H)$
$p_T(j_0 j_1 l_H l_T b_0 b_1 E_T^{\text{miss}})$	b-jet Reco Score	$E_T^{\text{miss}}$
nJets	HT jets	

Table 18: Input features used to identify the Higgs decay products in 2LSS events

<sup>1082</sup> Here  $j_0$  and  $j_1$ , and  $l_H$  are the jet and lepton decay candidates, respectively. The other  
<sup>1083</sup> lepton in the event is labeled  $l_T$ , as it is assumed to have come from the decay of one of the top  
<sup>1084</sup> quarks.  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet identification algorithm. The b-jet  
<sup>1085</sup> Reco Score is the output of the b-jet reconstruction algorithm.

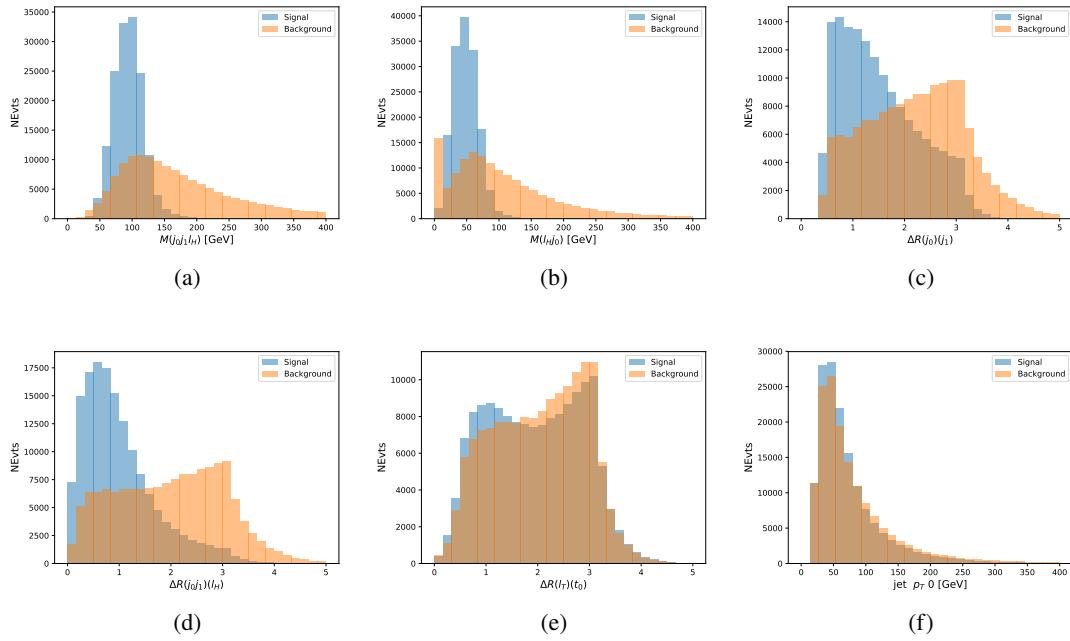


Figure 14.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (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.

1086           The modeling of these inputs is validated against data, with Figure 14.2 showing good  
1087           general agreement between data and MC. Plots for the complete list of features can found in  
1088           Section C.

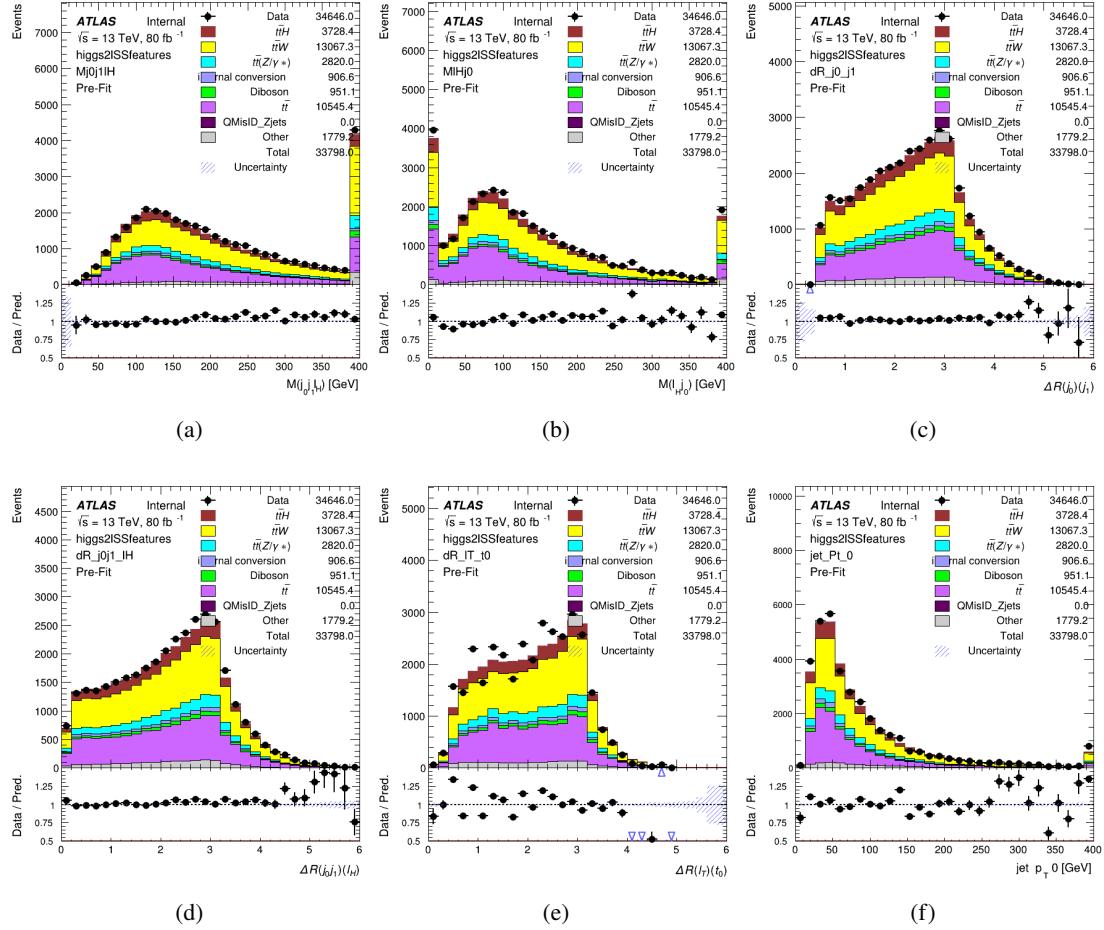


Figure 14.8: Data/MC comparisons of input features for higgs2ISS training for  $79.8 \text{ fb}^{-1}$  of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c)  $\Delta R$  between the jet candidates, (d)  $\Delta R$  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.

1089 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2  
 1090 million events, with an additional 200,000 reserved for testing the model. In order to compensate  
 1091 for the large number of incorrect combinations, these have been downsampled such that the correct  
 1092 combinations represent over 10% of the training set. The output of the NN is summarized in  
 1093 Figure 14.3.1.

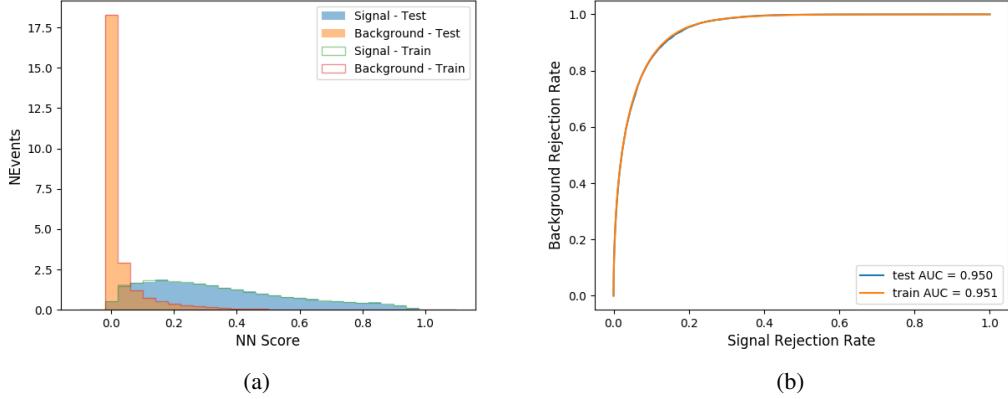


Figure 14.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1094        The neural network identifies the correct combination 55% of the time. It identifies the  
 1095        correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets  
 1096        81% of the time.

### 1097        14.3.2 3l Semi-leptonic Channel

1098        For 3l  $t\bar{t}H$  where the Higgs decay semi-leptonically, the decay products include one of the three  
 1099        leptons and two jets. In this case, the other two leptons originated from the decay of the tops,  
 1100        meaning the opposite-sign (OS) lepton cannot have come from the Higgs. This leaves only the two  
 1101        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	$\Delta R(j_0)(j_1)$
$M(j_0j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0j_1)(l_H)$	$\Delta R(j_0j_1)(l_{T_0})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0j_1l_{T_0})$	$M(j_0j_1l_{T_1})$
$M(j_0j_1l_H)$	$\Delta R(j_0j_1l_H)(l_{T_0})$	$\Delta R(j_0j_1l_H)(l_{T_1})$
$\Delta\phi(j_0j_1l_H)(E_T^{\text{miss}})$	$p_T(j_0j_1l_Hl_{T_0}l_{T_1}b_0b_1E_T^{\text{miss}})$	$M(j_0j_1b_0)$
$M(j_0j_1b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
b-jet Reco Score	$E_T^{\text{miss}}$	HT jets
nJets		

Table 19: Input features used to identify the Higgs decay products in 3l semi-leptonic events

1102 Here  $j_0$  and  $j_1$ , and  $l_H$  are the jet and lepton decay candidates, respectively. The other  
 1103 two leptons in the event are labeled as  $l_{T_0}$  and  $l_{T_1}$ .  $b_0$  and  $b_1$  are the two b-jets identified by  
 1104 the b-jet identification algorithm. The b-jet Reco Score is the output of the b-jet identification  
 1105 algorithm.

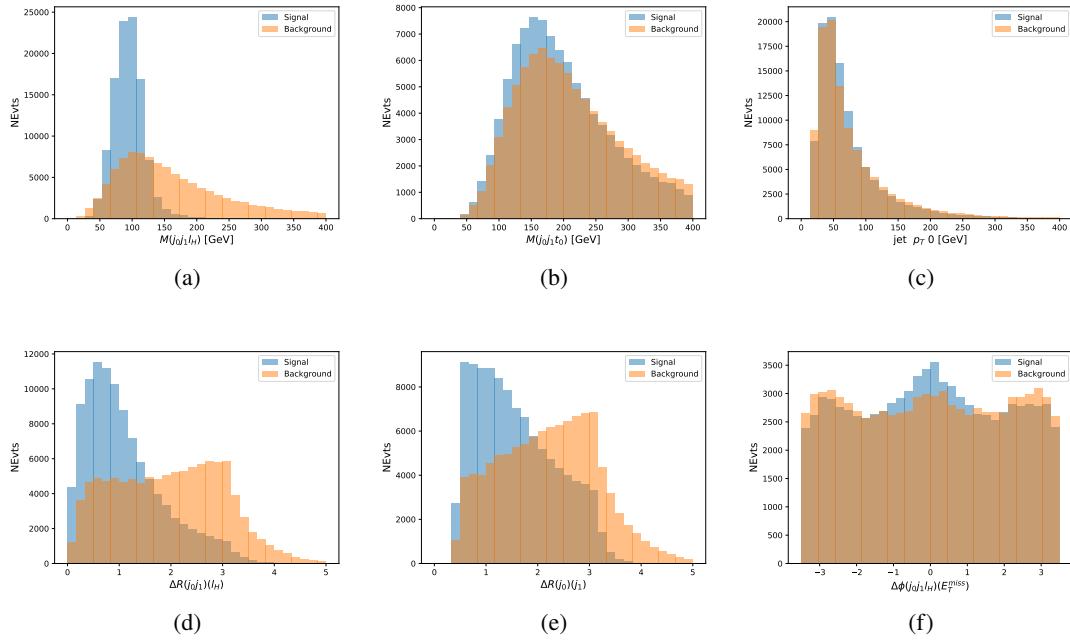


Figure 14.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1106           The modeling of these inputs is validated against data, with Figure 14.11 showing good  
 1107           general agreement between data and MC. Plots for the complete list of features can found in  
 1108           appendix C.1.

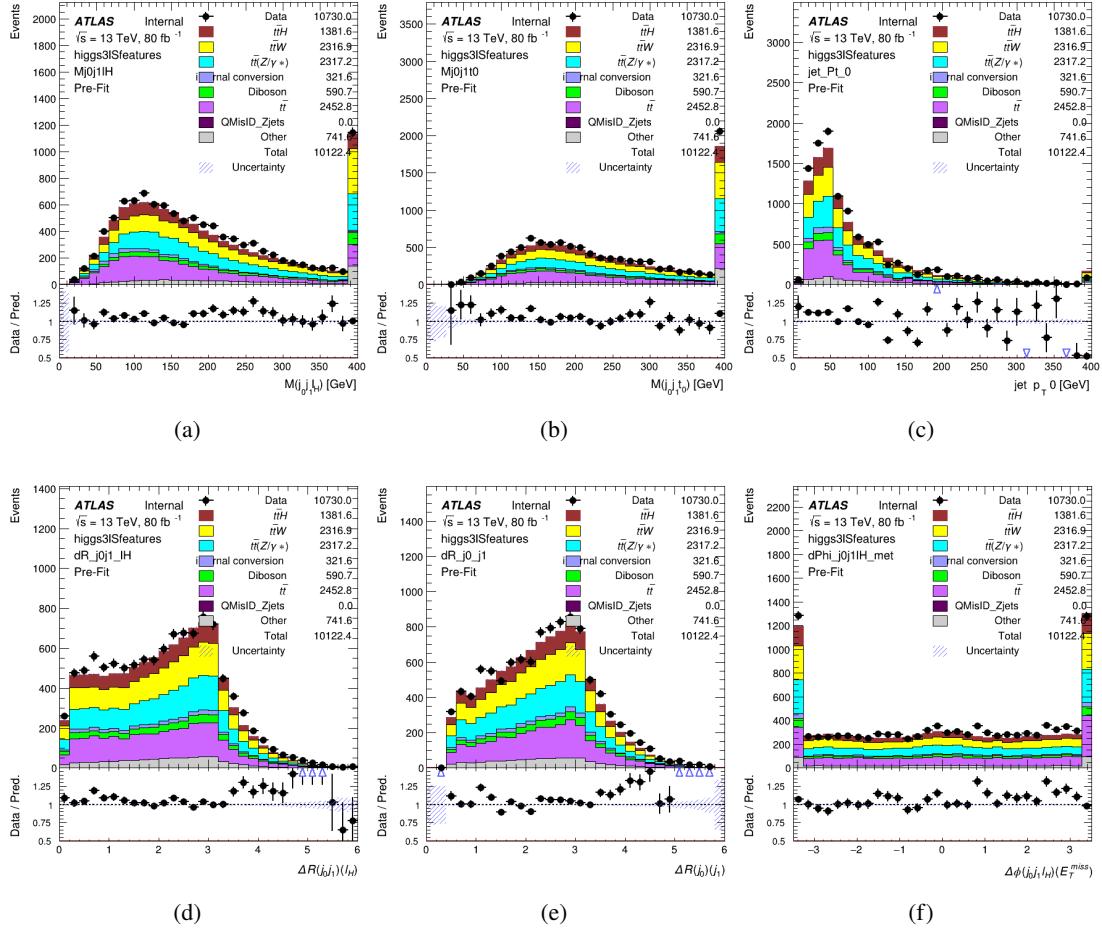


Figure 14.11: Data/MC comparisons of input features for higgs3lS training for  $79.8 \text{ fb}^{-1}$  of data.

1109 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.  
 1110 Once again, incorrect combinations are downsampled, such that the correct combinations are  
 1111 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN  
 1112 is summarized in Figure 14.3.2.

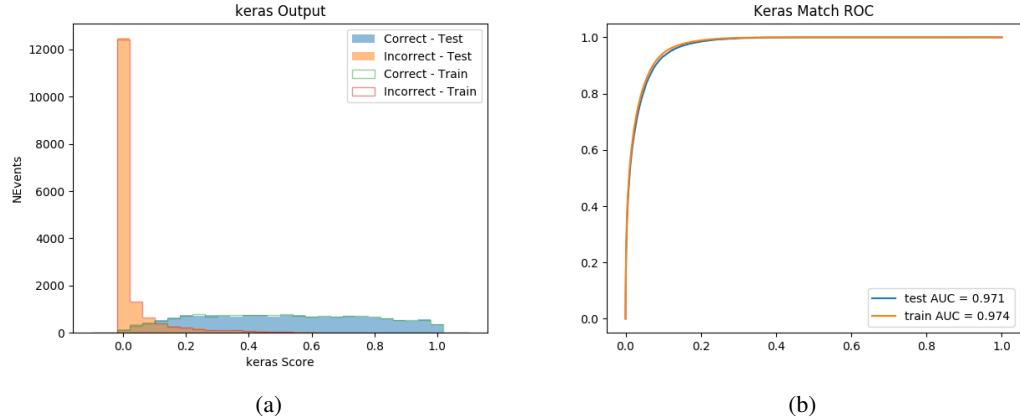


Figure 14.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1113        The neural network identifies the correct combination 64% of the time. It identifies the  
 1114        correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets  
 1115        83% of the time.

### 1116        14.3.3 3l Fully-leptonic Channel

1117        In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from  
 1118        the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to  
 1119        determining which of the two SS leptons originated from the Higgs. The kinematics of both  
 1120        possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,  
 1121        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 $p_T 1$	$\Delta\phi(l_{H_1})(E_T^{\text{miss}})$
$\Delta\phi(l_T)(E_T^{\text{miss}})$	$M(l_{H_0}l_{H_1})$	$M(l_{H_1}l_T)$
$M(l_{H_0}l_T)$	$\Delta R(l_{H_0})(l_{H_1})$	$\Delta R(l_{H_1})(l_T)$
$\Delta R(l_{H_0})(l_T)$	$\Delta R(l_{H_0}l_{H_1})(l_T)$	$\Delta R(l_{H_0}l_T)(l_{H_1})$
$\Delta R(l_{H_0}l_{H_1})(b_0)$	$\Delta R(l_{H_0}l_{H_1})(b_1)$	$\Delta R(l_{H_0}b_0)$
$M(l_{H_0}b_0)$	$\Delta R(l_{H_0}b_1)$	$M(l_{H_0}b_1)$
$\Delta R(l_{H_1}b_0)$	$M(l_{H_1}b_0)$	$\Delta R(l_{H_1}b_1)$
$M(l_{H_1}b_1)$	$E_T^{\text{miss}}$	b-jet Reco Score

Table 20: Input features used to identify the Higgs decay products in 3lF events

Table 21: Input features used to identify the Higgs decay products in 3l fully leptonic events

1122 Here  $l_{H_0}$  and  $l_{H_1}$  are the Higgs decay candidates. The other lepton in the event is labeled  
 1123  $l_T$ .  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco  
 1124 Score is the output of the Higgs reconstruction algorithm.

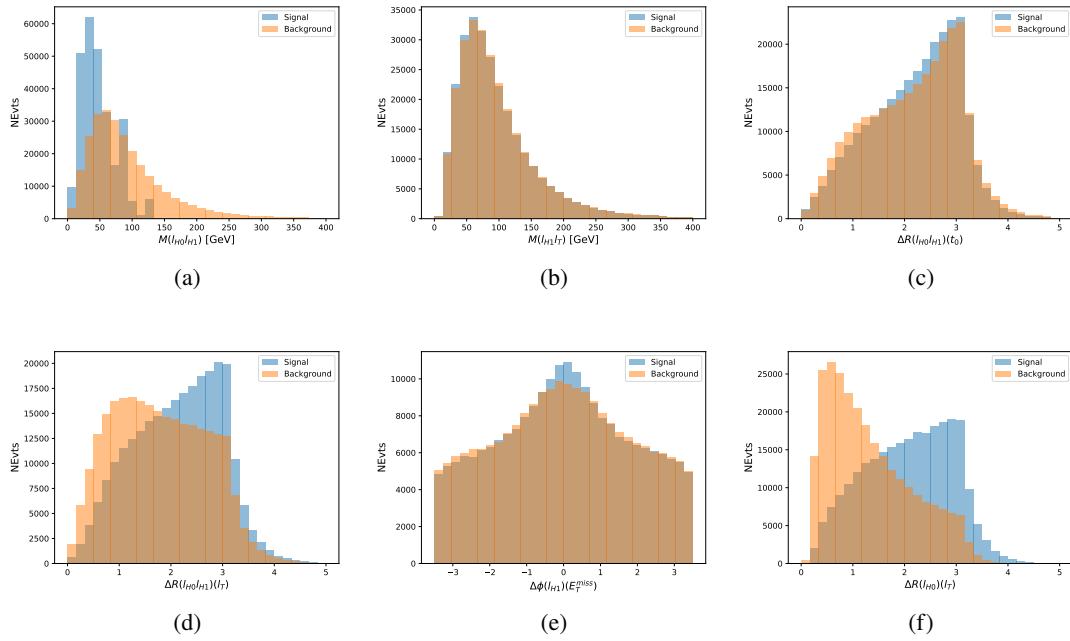


Figure 14.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1125            The modeling of these inputs is validated against data, with Figure 14.14 showing good  
 1126            general agreement between data and MC. Plots for the complete list of features can found in  
 1127            Section C.

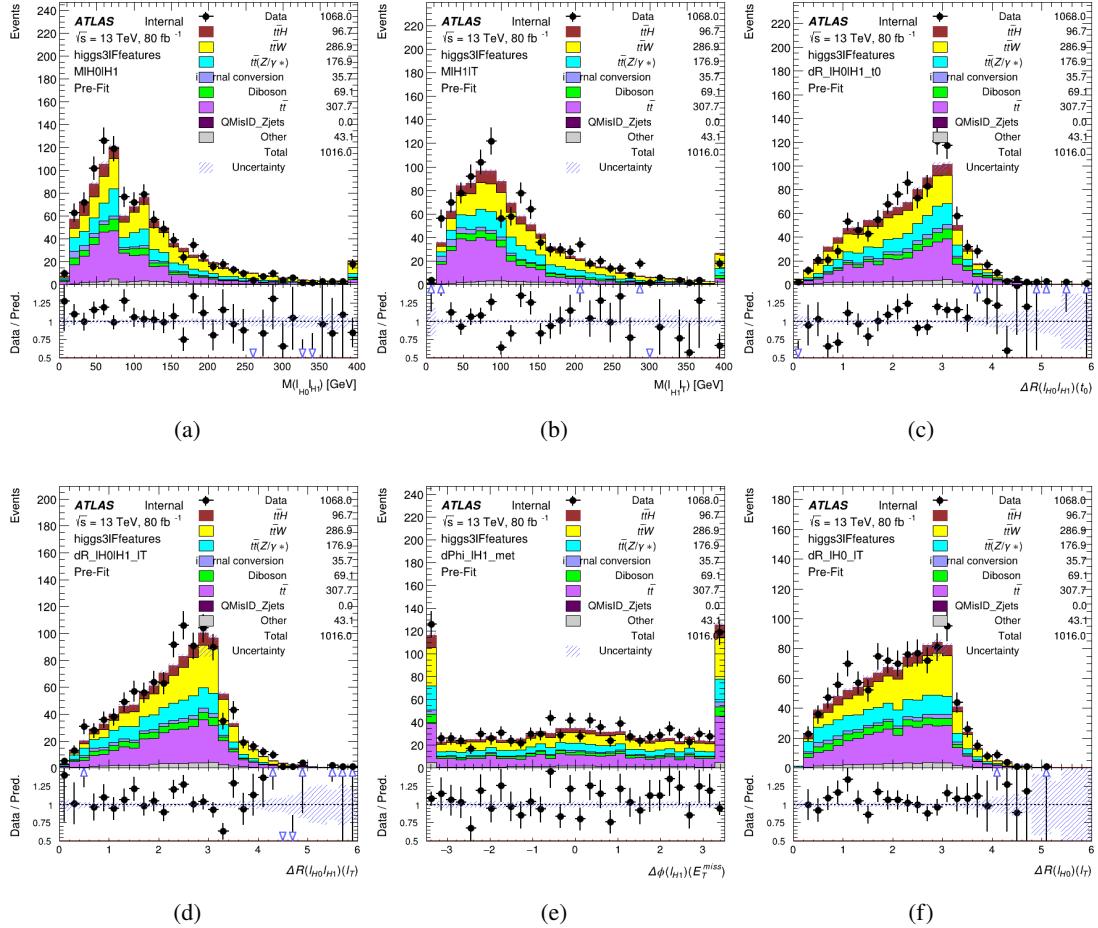


Figure 14.14: Data/MC comparisons of input features for higgs3lF training for  $79.8 \text{ fb}^{-1}$  of data.

1128 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,  
 1129 with 10% of the dataset reserved for testing. The output of the model is summarized in Figure  
 1130 [14.3.3.](#).

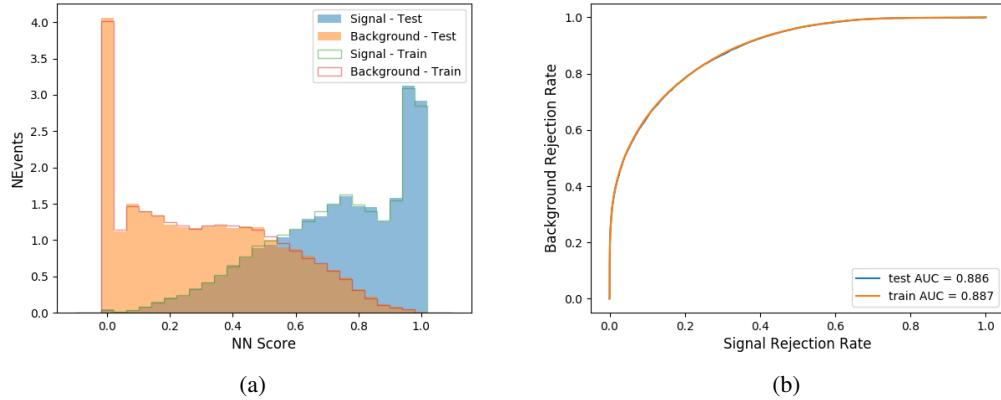


Figure 14.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

1131           The correct lepton is identified by the model for 80% of events in the testing data set.

## 1132       **14.4 $p_T$ Prediction**

1133       Once the most probable decay products have been identified, their kinematics are used as inputs  
 1134       to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,  
 1135       a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay  
 1136       are included as well, as these are found to improve performance. The truth  $p_T$  of the Higgs,  
 1137       as predicted by MC, are used as labels. Separate models are built for each channel - 2lSS, 3l  
 1138       Semi-leptonic and 3l Fully-leptonic.

1139       As a two-bin fit is targeted for the final result, some metrics evaluating the performance  
 1140       of the models aim to show how well it distinguished between "high  $p_T$ " and "low  $p_T$ " events. A

1141 cutoff point of 150 GeV is used to define these two categories.

1142 Because the analysis uses a two bin fit of the Higgs  $p_T$ , the momentum reconstruction  
1143 could be treated as a binary classification problem, rather than a regression problem. This  
1144 approach is explored in detail in Section C.5, and is found not to provide any significant increase  
1145 in sensitivity. The regression approach is used because it provides more flexibility for future  
1146 analyses, as it is independent of the cutoff between high and low  $p_T$ , as well as the number of  
1147 bins. Further, a regression allows the output of the neural network to be more clearly understood,  
1148 as it can be directly compared to a physics observable.

#### 1149 **14.4.1 2lSS Channel**

1150 The input variables listed in Table 22 are used to predict the Higgs  $p_T$  in the 2lSS channel. Here  
1151  $j_0$  and  $j_1$  are the two jets identified as Higgs decay products. The lepton identified as originating  
1152 from the Higgs is labeled  $l_H$ , while the other lepton is labeled  $l_T$ , as it is assumed to have come  
1153 from the decay of one of the top quarks.  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet  
1154 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs  
1155 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 22: Input features for reconstructing the Higgs  $p_T$  spectrum for 2lSS events

1156        The optimal neural network architecture for this channel is found to consist of 7 hidden  
 1157        layers with 60 nodes each. The input data set includes 1.2 million events, 10% of which is used  
 1158        for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1159        To evaluate the performance of the model, the predicted  $p_T$  spectrum is compared to the  
 1160        truth Higgs  $p_T$  in Figure 14.16. In order to visualize the model performance more clearly, in (a)  
 1161        of that figure, the color of each point is determined by Kernel Density Estimation (KDE). The  
 1162        color shown represents the logarithm of the output from KDE, to counteract the large number of  
 1163        low  $p_T$  events. For that same reason, each column of the histogram shown in (b) of Figure 14.16  
 1164        is normalized to unity. This plot therefore demonstrates what the model predicts for each slice  
 1165        of truth  $p_T$ .

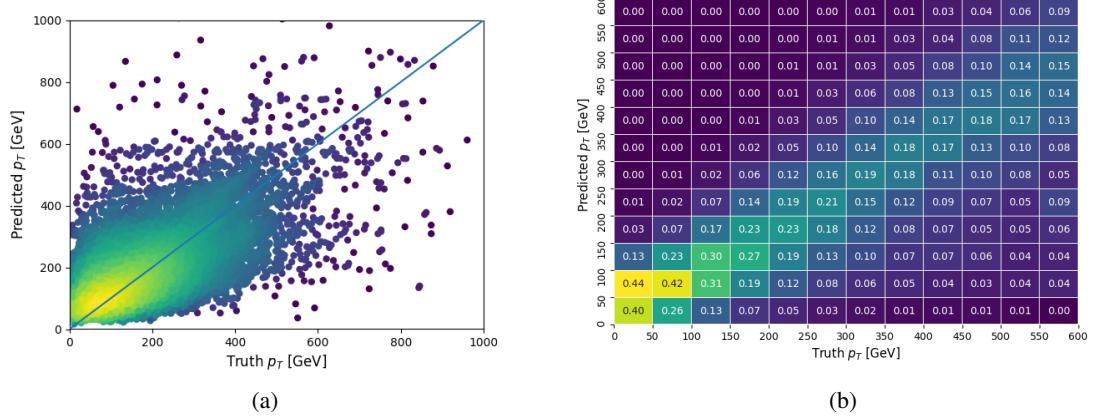


Figure 14.16: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 2lSS  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized to one.

1166 We are also interested in how well the model distinguishes between events with  $p_T < 150$   
 1167 GeV and  $> 150$  GeV. Figure 14.17 demonstrates the NN output for high and low  $p_T$  events based  
 1168 on this cutoff.

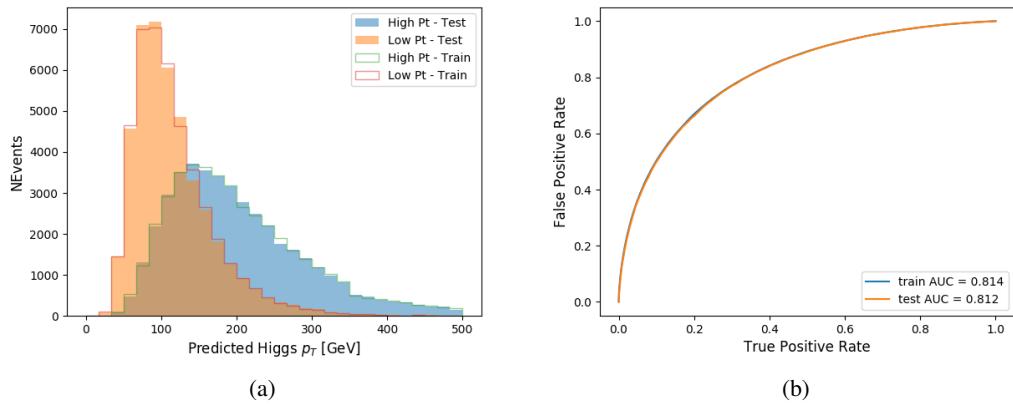


Figure 14.17: (a) shows the reconstructed Higgs  $p_T$  for 2lSS events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

<sup>1169</sup> **14.4.2 3l Semi-leptonic Channel**

<sup>1170</sup> The following input features are used to predict the Higgs  $p_T$  for events in the 3lS channel:

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

Table 23: Input features for reconstructing the Higgs  $p_T$  spectrum for 3lS events

<sup>1171</sup> Again,  $j_0$  and  $j_1$  are the two jets identified as Higgs decay products, ordered by  $p_T$ . The  
<sup>1172</sup> lepton identified as originating from the Higgs is labeled  $l_H$ , while the other two leptons are  
<sup>1173</sup> labeled  $l_{T0}$  and  $l_{T1}$ .  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet identification algorithm.  
<sup>1174</sup> The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs reconstruction algorithm,  
<sup>1175</sup> and the b-jet identification algorithm, respectively.

<sup>1176</sup> The optimal neural network architecture for this channel is found to consist of 7 hidden  
<sup>1177</sup> layers with 80 nodes each. The inputdata set includes one million events, 10% of which is used  
<sup>1178</sup> for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1179 To evaluate the performance of the model, the predicted  $p_T$  spectrum is compared to the  
 1180 truth Higgs  $p_T$  in Figure 14.18. Once again, (a) of 14.18 shows a scatterplots of predicted vs  
 1181 truth  $p_T$ , where the color of each point corresponds to the log of the relative KDE at that point.  
 1182 Each column of the the histogram in (b) is normalized to unity, to better demonstrate the output  
 1183 of the NN for each slice of truth  $p_T$ .

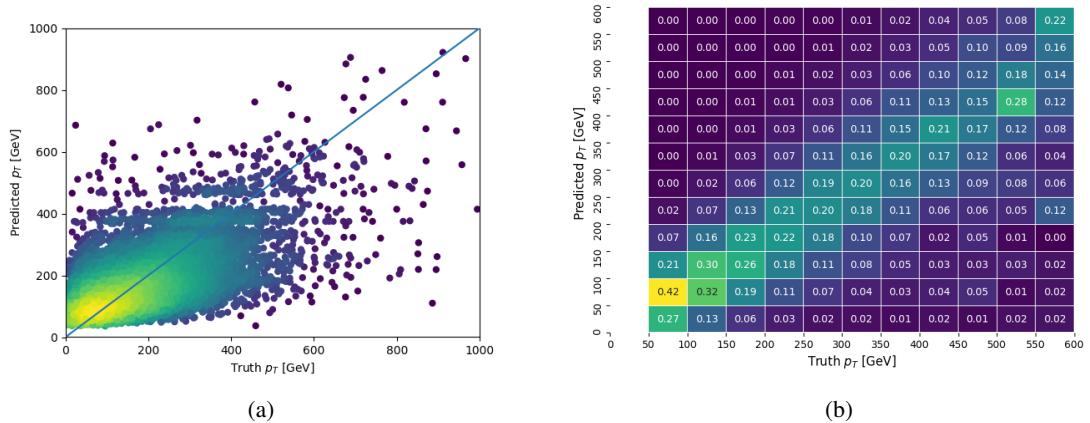


Figure 14.18: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 3lS  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1184 Figure 14.19 shows (a) the output of the NN for events with truth  $p_T$  less than and greater  
 1185 than 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low  
 1186  $p_T$  events.

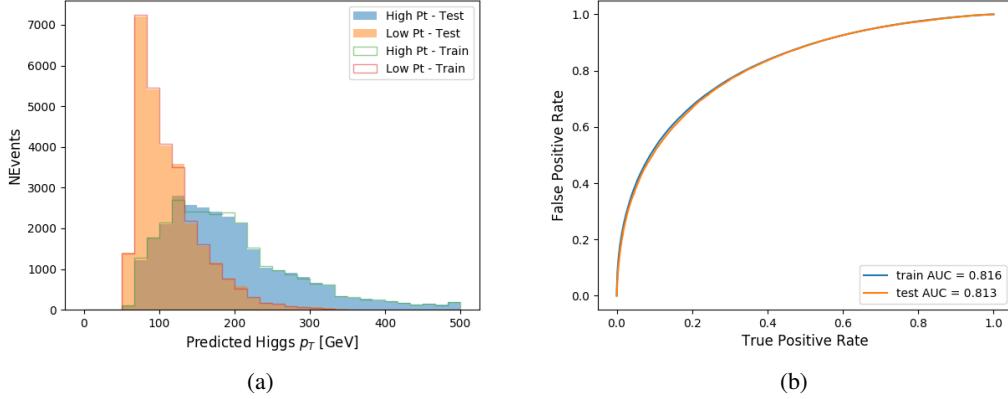


Figure 14.19: (

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.

### 14.4.3 3l Fully-leptonic Channel

The features listed in 24 are used to construct a model for predictin the Higgs  $p_T$  for 3lF events.

HT	$M(l_{H0} l_{H1})$	$M(l_{H0} l_T)$
$M(l_{H0} b_0)$	$M(l_{H0} b_1)$	$M(l_{H1} l_T)$
$M(l_{H1} b_0)$	$M(l_{H1} b_1)$	$\Delta R(l_{H0})(l_{H1})$
$\Delta R(l_{H0})(l_T)$	$\Delta R(l_{H0} l_{H1})(l_T)$	$\Delta R(l_{H0} l_T)(l_{H1})$
$\Delta R(l_{H1})(l_T)$	$\Delta R(l_{H0} b_0)$	$\Delta R(l_{H0} b_1)$
$\Delta R(l_{H1} b_1)$	$\Delta R(l_{H1} b_0)$	Higgs Reco Score
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
$E_T^{\text{miss}}$	b-jet Reco Score	b-jet $p_T$ 0
b-jet $p_T$ 1		

Table 24: Input features for reconstructing the Higgs  $p_T$  spectrum for 3lF events

$l_{H0}$  and  $l_{H1}$  respresent the two leptons identified by the Higgs reconstruction model as

1190 originating from the Higgs, while  $l_T$  is the other lepton in the event. The Higgs Reco Score and  
 1191 b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification  
 1192 algorithm, respectively.

1193 The optimal neural network architecture for this channel is found to consist of 5 hidden  
 1194 layers with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for  
 1195 testing, the other 90% for training. Training is found to converge after around 150 epochs.

1196 The predicted transverse momentum, as a function of the truth  $p_T$ , is shown in Figure  
 1197 [14.20](#).

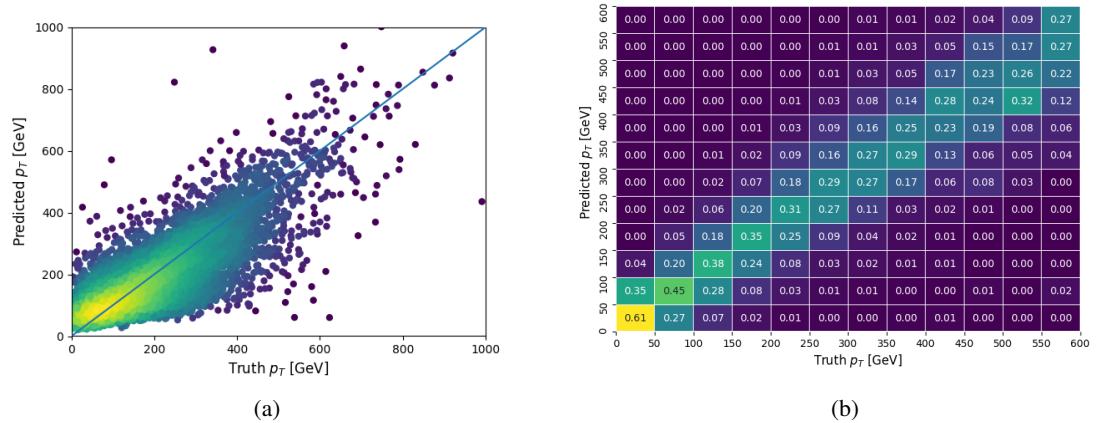


Figure 14.20: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 3LF  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1198 When split into high and low  $p_T$ , based on a cutoff of 150 GeV, the

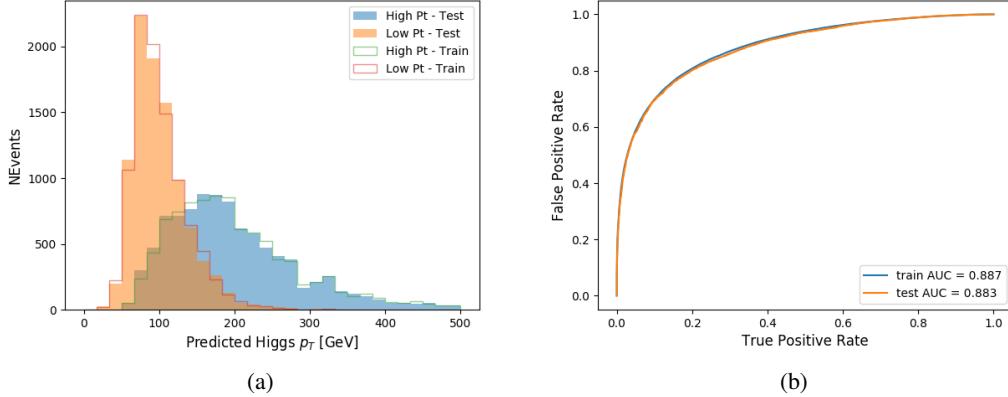


Figure 14.21: (a) shows the reconstructed Higgs  $p_T$  for 3lF events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

## 1199 14.5 3l Decay Mode

1200 In the 3l channel, there are two possible ways for the Higgs to decay, both involving intermediate  
 1201 W boson pairs: Either both W bosons decay leptonically, in which case the reconstructed decay  
 1202 consists of two leptons (referred as the fully-leptonic 3l channel), or one W decays leptonically  
 1203 and the other hadronically, giving two jets and one lepton in the final state (referred to as the  
 1204 semi-leptonic 3l channel). In order to accurately reconstruct the Higgs, it is necessary to identify  
 1205 which of these decays took place for each 3l event.

1206 The kinematics of each event, along with the output scores of the Higgs and top recon-  
 1207 struction algorithms, are used to distinguish these two possible decay modes. The particular  
 1208 inputs used are listed in Table 25.

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

1209 Here  $l_0$  is the opposite charge lepton,  $l_1$  and  $l_2$  are the two SS leptons order by  $\Delta R$   
 1210 from lepton 0. score3lF and score3lS are the outputs of the 3lS and 3lF Higgs reconstruction  
 1211 algorithms, while topScore is the output of the b-jet identification algorithm.

1212 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these  
 1213 two decay modes. The output of the model is summarized in Figure 14.22.

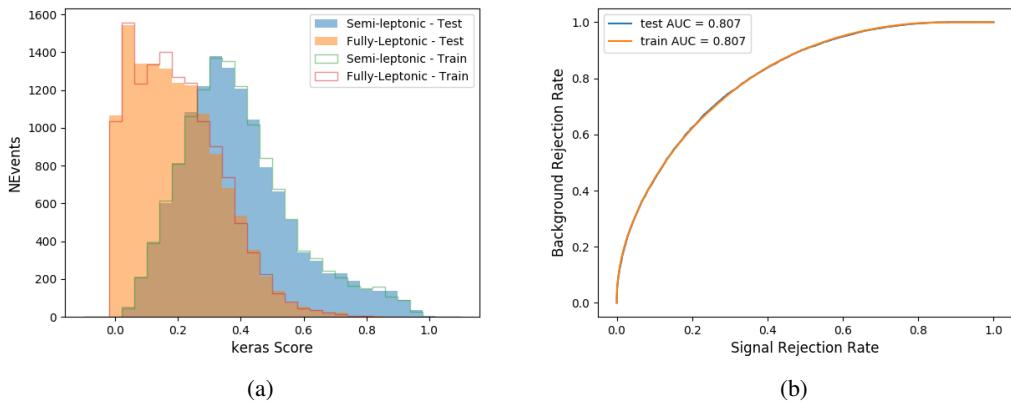


Figure 14.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

1214 A cutoff of 0.23 is determined to be optimal for separating 3lS and 3lF in the fit.

## 1215 15 Signal Region Definitions

1216 Events are divided into two channels based on the number of leptons in the final state: one  
1217 with two same-sign leptons, the other with three leptons. The 3l channel includes events where  
1218 two leptons originated from the Higgs boson as well as events where only one of the leptons  
1219 originated from the Higgs. This motivates splitting the 3l channel into semi-leptonic, and fully  
1220 leptonic channels, after an event preselection has been applied.

### 1221 15.1 Pre-MVA Event Selection

1222 A preselection is applied to define orthogonal analysis channels based on the number of leptons  
1223 in each event. For the 2lSS channel, the following preselection is used:

- 1224 • Two very tight, same-charge, light leptons with  $p_T > 20 \text{ GeV}$
- 1225 •  $\geq 4$  reconstructed jets,  $\geq 1$  b-tagged jets
- 1226 • No reconstructed tau candidates

1227 The event yield after the 2lSS preselection has been applied, for MC and data at  $79.8 \text{ fb}^{-1}$ ,  
1228 is shown in Table 15.1.

Process	Yield
t̄tH high p <sub>T</sub>	41 ± 5
t̄tH low p <sub>T</sub>	71 ± 8
t̄tW	450 ± 70
t̄t(Z/γ*)	91 ± 11
t̄tll low mass	10 ± 6
Rare Top	20 ± 12
VV	42 ± 22
tZ	10 ± 5
QMisID	44.7 ± 2.7
Fakes int. conv	47 ± 26
Fakes ext. conv	46 ± 44
Fakes HF e	45 ± 23
Fakes HF μ	250 ± 50
Three top	2.2 ± 1.1
Four top	5.64 ± 0.31
t̄tWW	10.9 ± 0.6
tW	0.0 ± 0.0
WtZ	9.1 ± 0.8
VVV	0.30 ± 0.05
VH	0.6 ± 1.0
Total	1170 ± 120
Data	1108

Table 26: Yields of the 2lSS preselection region

1229

Figure 15.1. Good general agreement is found.

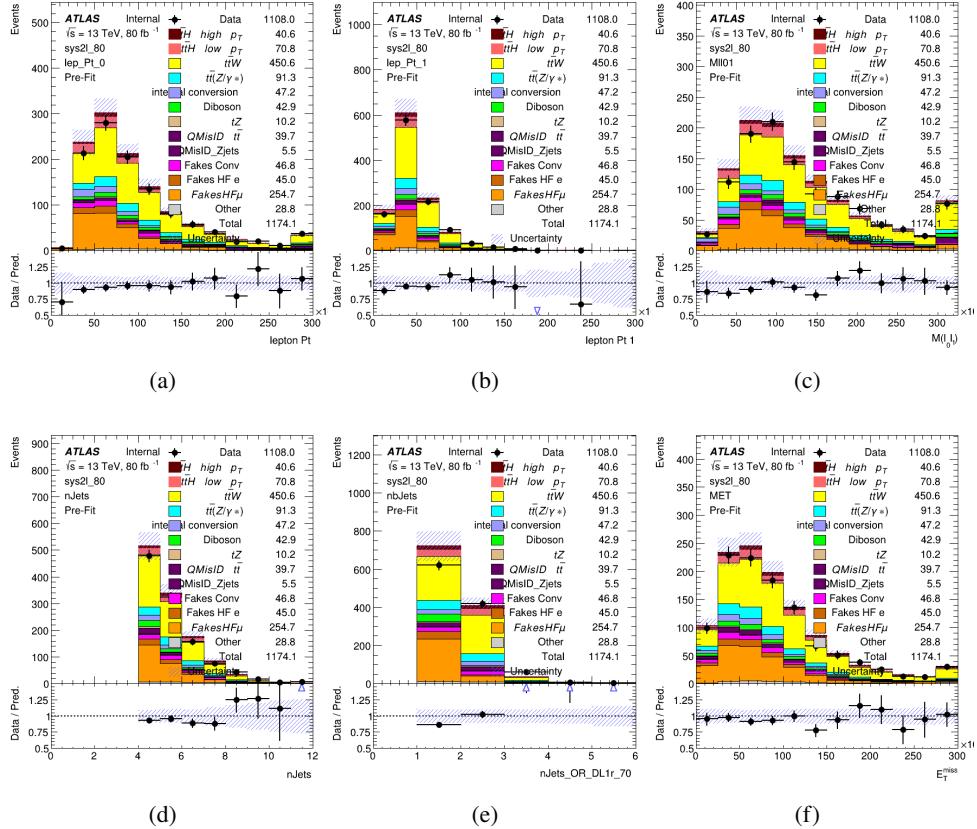


Figure 15.1: Data/MC comparisons of the 2lSS pre-selection region. (a) and (b) show the  $p_T$  of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1230

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

1231

- Three light leptons with total charge  $\pm 1$
- Same charge leptons are required to be very tight, with  $p_T > 20$  GeV
- Opposite charge lepton must be loose, with  $p_T > 10$  GeV
- $\geq 2$  reconstructed jets,  $\geq 1$  b-tagged jets

1233

- 1235      • No reconstructed tau candidates
- 1236      •  $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$  for all opposite-charge, same-flavor lepton pairs
- 1237      The event yield after the 3l preselection has been applied, for MC and data at  $79.8 \text{ fb}^{-1}$ ,
- 1238      is shown in Table 15.1.

Process	Yield
t̄tH high p <sub>T</sub>	$20.5 \pm 2.3$
t̄tH low p <sub>T</sub>	$33.6 \pm 3.8$
t̄tW	$138 \pm 18$
t̄tZ/γ	$80 \pm 9$
t̄tlllowmass	$3.5 \pm 2.0$
rareTop	$22 \pm 12$
VV	$39 \pm 19$
tZ	$9.2 \pm 4.5$
QMisID	$1.8 \pm 0.6$
Fakes int. conv	$31 \pm 17$
Fakes ext. conv	$14 \pm 11$
Fakes HF e	$20 \pm 10$
Fakes HF μ	$102 \pm 22$
Three top	$0.96 \pm 0.48$
Four top	$6.17 \pm 0.35$
t̄tWW	$5.46 \pm 0.33$
tW	$0.0 \pm 0.0$
WtZ	$8.7 \pm 0.6$
VVV	$0.81 \pm 0.11$
VH	$0.0 \pm 0.0$
Total	$512 \pm 48$
Data	535

Table 27: Yields of the 3l preselection region.

Table 28: Yields of the 3l preselection region.

1239      Comparisons of kinematic distributions for data and MC in this region are shown in Figure

1240 15.2.

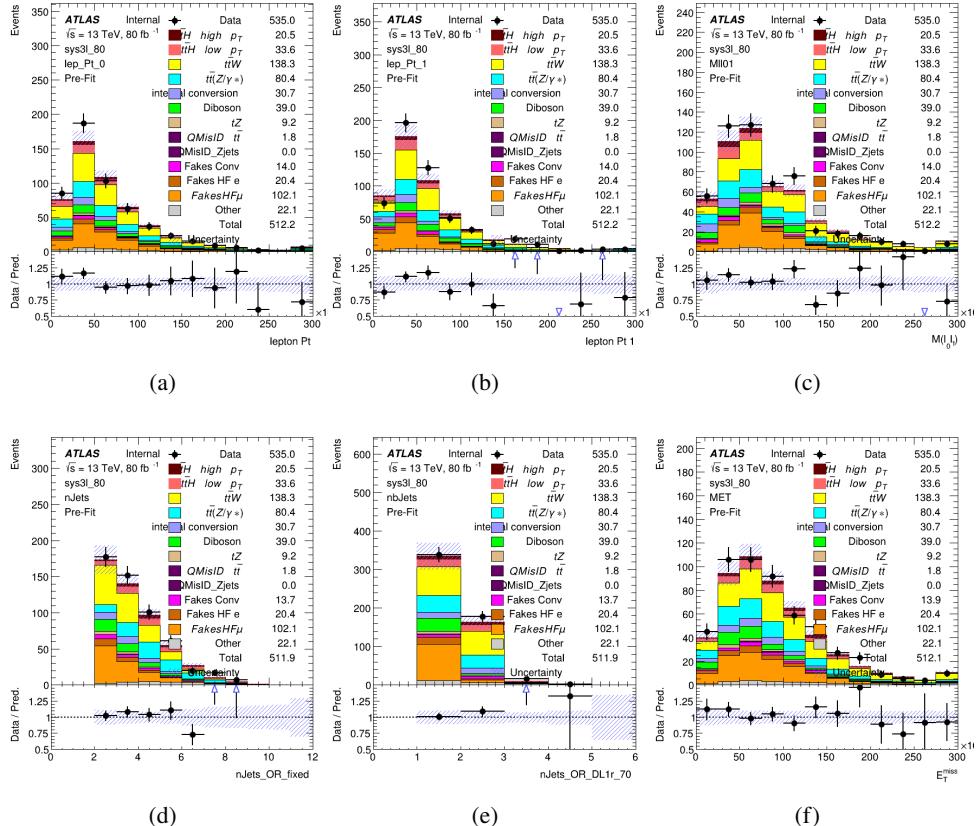


Figure 15.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the  $p_T$  of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1241 15.2 Event MVA

1242 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events  
 1243 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic  
 1244 (3lF). Here events with three leptons are split into 3lS and 3lF based on the model described  
 1245 in 14.5. In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

1246 [33] are trained using the kinematics of signal and background events derived from Monte Carlo  
1247 simulations. Events are weighted in the BDT training by the weight of each Monte Carlo event.

1248 Because the background composition differs for events with a high reconstructed Higgs  $p_T$   
1249 compared to events with low reconstructed Higgs  $p_T$ , separate MVAs are produced for high and  
1250 low  $p_T$  regions. This is found to provide better significance than attempting to build an inclusive  
1251 model, as demonstrated in appendix C.2. A cutoff of 150 GeV is used. This gives a total of 6  
1252 background rejection MVAs - explicitly, 2lSS high  $p_T$ , 2lSS low  $p_T$ , 3lS high  $p_T$ , 3lS low  $p_T$ ,  
1253 3lF high  $p_T$ , and 3lF low  $p_T$ .

1254 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 $p_T$ 2lSS	$\Delta R(l_0)(l_1)$	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 29: Input features used to distinguish signal and background events in the 2lSS channel.

1255 While for each of the 3l BDTs, the features listed below are used for training:

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

Table 30: Input features used to distinguish signal and background events in the 3l channel.

1256 Modelling of each of these input features is verified in Appendix C.2 by comparing data  
 1257 and MC for  $79.8 \text{ fb}^{-1}$ . The BDTs are produced with a maximum tree depth of 6, using AUC as  
 1258 the target loss function. The BDT response distribution and ROC curve for each model is shown  
 1259 in Figures 15.3-15.5.

## 2lSS

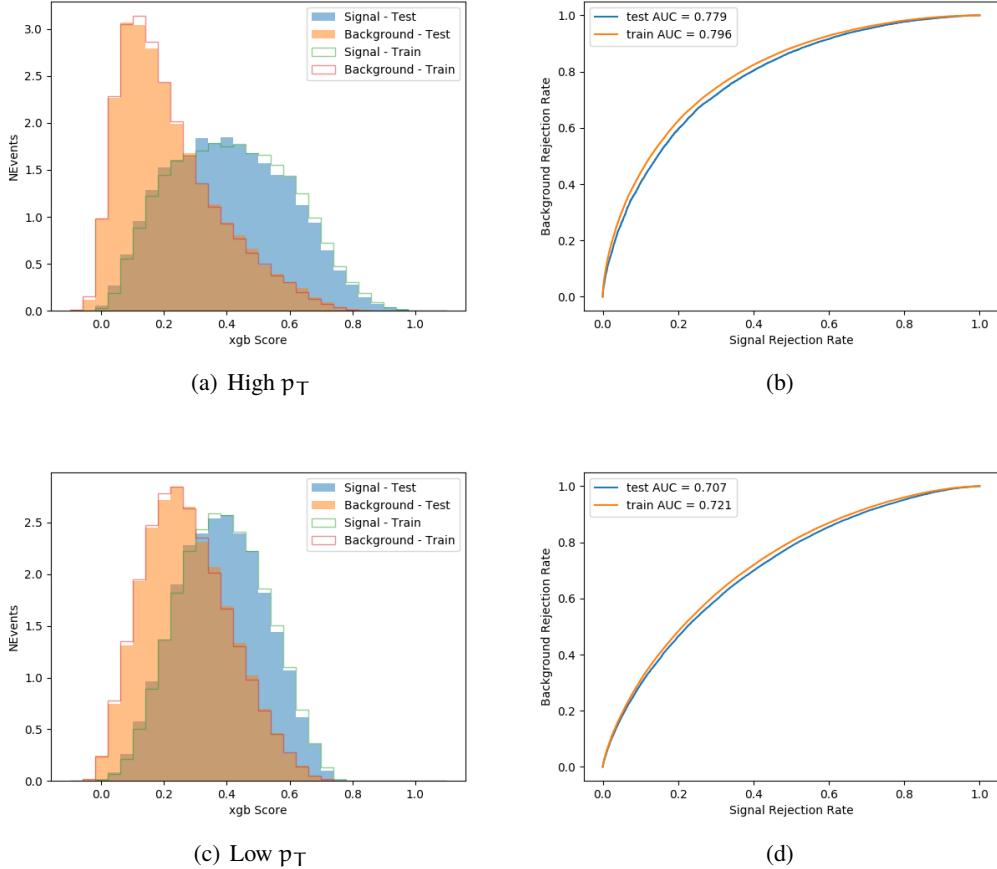


Figure 15.3: Output BDT scores of training and testing data for signal (blue) and background (orange) for 2lSS events with (a) high regressed Higgs  $p_T$  and (b) low regressed Higgs  $p_T$ . (b) and (d) show the ROC curve for the 2lSS high and low  $p_T$  models, respectively.

## 3l - Semileptonic

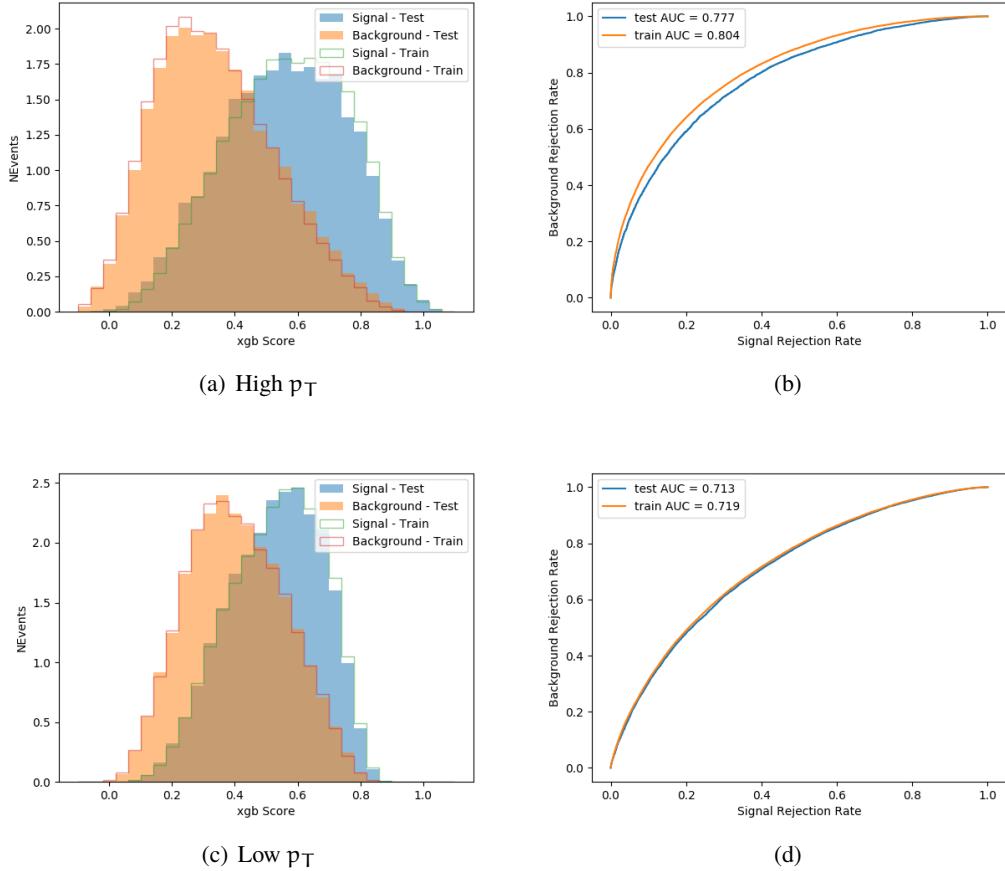


Figure 15.4: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lS events with (a) high regressed Higgs  $p_T$  and (b) low regressed Higgs  $p_T$ . (b) and (d) show the ROC curve for the 3lS high and low  $p_T$  models, respectively.

### 3l - Fully Leptonic

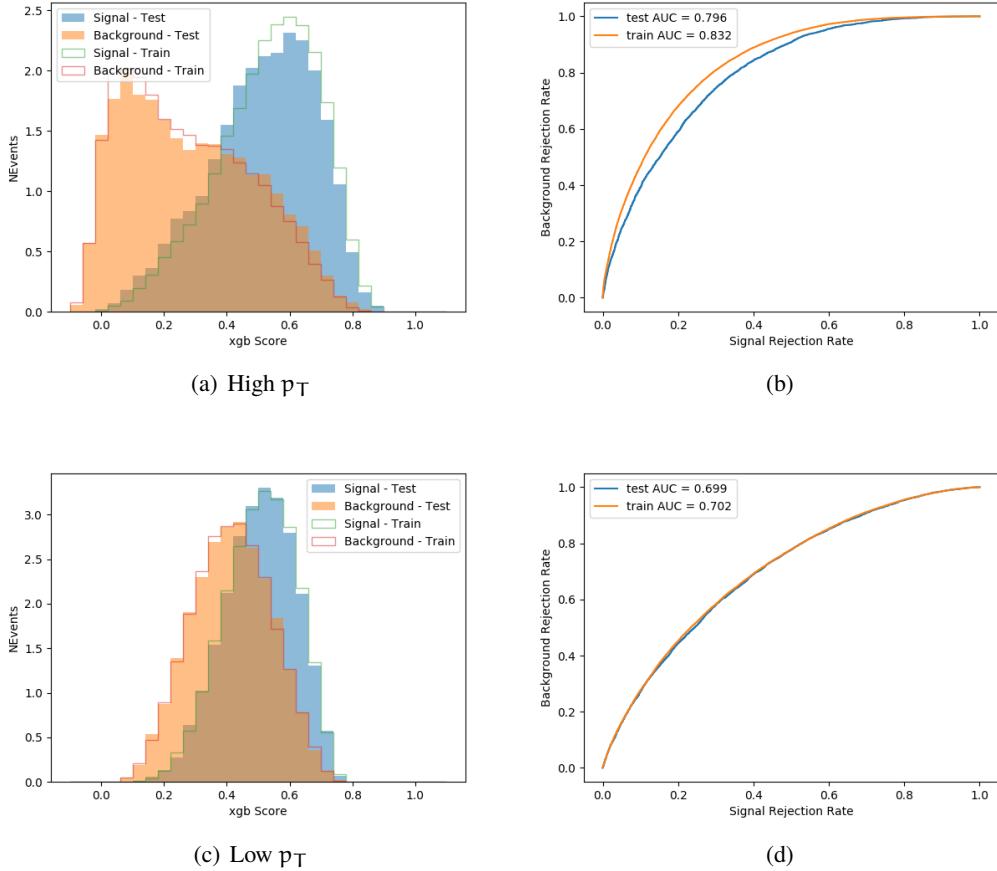


Figure 15.5: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lF events with (a) high regressed Higgs  $p_T$  and (b) low regressed Higgs  $p_T$ . (b) and (d) show the ROC curve for the 3lF high and low  $p_T$  models, respectively.

1260      Output distributions of each MVA, comparing MC predictions to data at  $79.8 \text{ fb}^{-1}$  are  
 1261      shown in figures [15.6-15.2](#).

### High $p_T$ Background Rejection BDTs

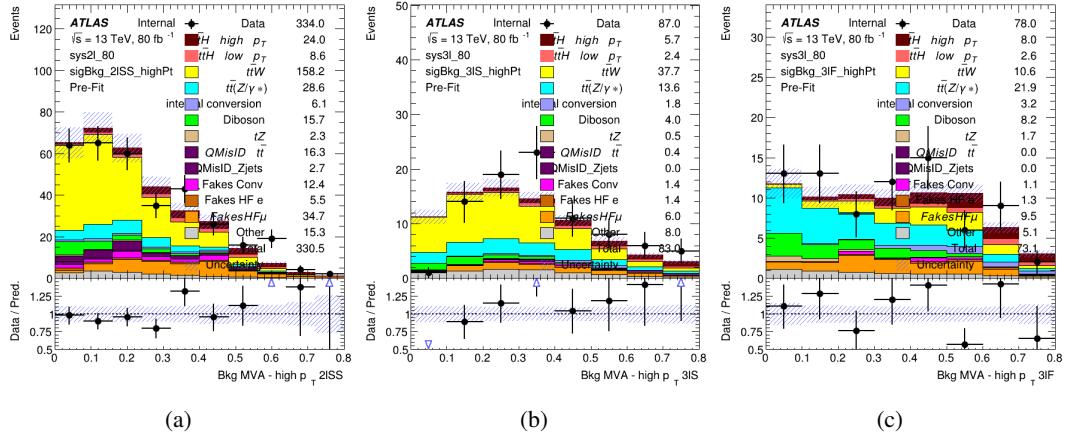


Figure 15.6: Output score of the high  $p_T$  BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

### Low $p_T$ Background Rejection BDTs

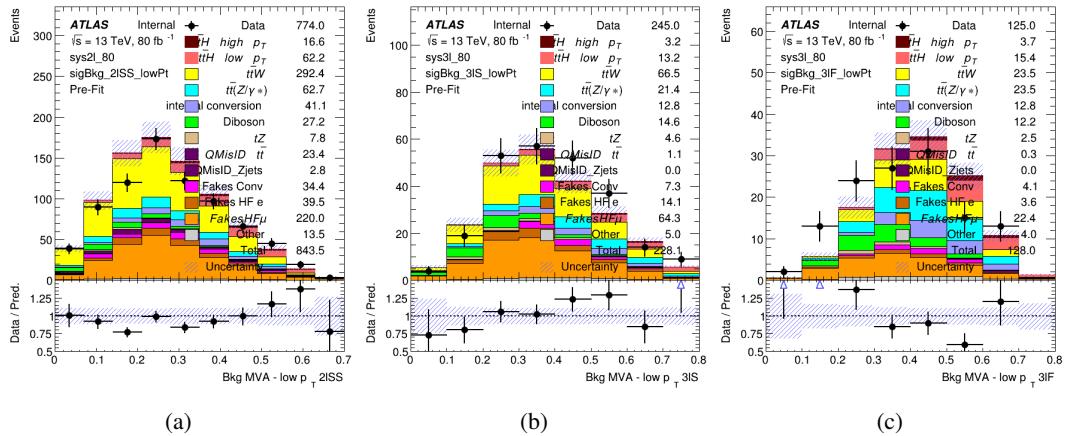


Figure 15.7: Output score of the low  $p_T$  BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

---

### 1262 15.3 Signal Region Definitions

1263 Once pre-selection has been applied, channels are further refined based on the MVAs described  
 1264 above. The output of the model described in Section 14.5 is used to separate the three channel  
 1265 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs  
 1266 boson. This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

1267 For each event, depending on the number of leptons as well as whether the  $p_T$  of the Higgs  
 1268 is predicted to be high ( $> 150$  GeV) or low ( $< 150$  GeV), a cut on the appropriate background  
 1269 rejection MVA is applied. The particular cut values, listed in Table 31, are determined by  
 1270 maximizing  $S/\sqrt{B}$  in each region.

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

1271 The event preselection and MVA selection listed in Table 31 are used define the three  
 1272 signal regions used in the fit. These signal region definitions are summarized in Table 32.

Region	Selection
2lSS	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 $> 0.34$
3lS	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 $> 0.43$
3lF	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 32: Selection applied to define the three signal regions used in the fit.

## 1273 16 Systematic Uncertainties

1274 The systematic uncertainties that are considered are summarized in Table 33. These are imple-  
 1275 mented in the fit either as a normalization factors or as a shape variation or both in the signal  
 1276 and background estimations. The numerical impact of each of these uncertainties is outlined in  
 1277 section 17.

Table 33: 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>1278</sup> The uncertainty in the combined integrated luminosity is derived from a calibration of the  
<sup>1279</sup> luminosity scale using x-y beam-separation scans performed for 13 TeV proton-proton data [15],  
<sup>1280</sup> [36].

<sup>1281</sup> The experimental uncertainties are related to the reconstruction and identification of light  
<sup>1282</sup> leptons and b-tagging of jets, and to the reconstruction of  $E_T^{\text{miss}}$ .

1283        The sources which contribute to the uncertainty in the jet energy scale [46] are decomposed  
 1284        into uncorrelated components and treated as independent sources in the analysis. This method  
 1285        decomposes the uncertainties into 30 nuisance parameters included in the fit. A similar method  
 1286        is used to account for jet energy resolution (JER) uncertainties, and 8 JER uncertainty components  
 1287        are included as NPs in the fit.

1288        The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses  
 1289        [38] are also decomposed into uncorrelated components. The large number of components for  
 1290        b-tagging is due to the calibration of the distribution of the BDT discriminant.

1291        As mentioned in Section 12.2, a normalization corrections and uncertainties on the estim-  
 1292        ates of non-prompt leptons backgrounds are derived using data driven techniques, described in  
 1293        detail in [7]. These are derived from a likelihood fit over various non-prompt enriched control  
 1294        regions, targeting several sources of non-prompt light leptons separately: external conversion  
 1295        electrons, internal conversion electrons, electrons from heavy flavor decays, and muons from  
 1296        heavy flavor decays.

1297        The normalization factor and uncertainty applied to each source of non-prompt leptons is  
 1298        summarized in Table 16

Processss	Normalization Factor
$NF_e^{\text{ExtCO}}$	$1.70 \pm 0.51$
$NF_e^{\text{IntCO}}$	$0.75 \pm 0.26$
$NF_e^{\text{HF}}$	$1.09 \pm 0.32$
$NF_{\mu}^{\text{HF}}$	$1.28 \pm 0.17$

Table 34: Normalization factors - with statistical and systematic uncertainties - derived from the fit over fake control regions for each source of non-prompt leptons considered.

1299 In addition to those derived from the control regions, several additional uncertainties are  
1300 assigned to the non-prompt lepton background. An additional 25% uncertainty on material  
1301 conversions is assigned, based on the comparison between data and MC in a region where a  
1302 loose electron fails the photon conversion veto. A shape uncertainty of 15% (6%) is assigned to  
1303 the HF non-prompt electron (muon) background based on a comparison between data and MC  
1304 where the second leading electron (muon) is only required to be loose. As the contribution from  
1305 light non-prompt leptons is small, about 10% percent of the contribution from HF non-prompt  
1306 leptons, it is derived from the agreement between data and simulation in a LF enriched region at  
1307 low values of the non-prompt lepton BDT. The resulting uncertainty is 100%, and is taken to be  
1308 uncorrelated between internal and material conversions.

1309 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and  
1310 scale uncertainties are taken from theory calculations for the predominate prompt backgrounds.  
1311 Following the nominal  $t\bar{t}H - ML$  analysis, a 50% uncertainty is applied to Diboson to account  
1312 for the large uncertainty in estimating  $VV +$  heavy flavor. The other “rare” background processes  
1313 - including  $tZ$ , rare top processes,  $ttWW$ ,  $WtZ$ ,  $VVV$ ,  $tHjb$  and  $WtH$  - are assigned an overall  
1314 50% normalization uncertainty as well. The theory uncertainties applied to the MC estimates  
1315 are summarized in Table 35.

Process	X-section [%]
t̄H (aMC@NLO+Pythia8)	QCD Scale: $^{+5.8}_{-9.2}$ PDF( $+\alpha_S$ ): $\pm 3.6$
t̄Z (aMC@NLO+Pythia8)	QCD Scale: $^{+9.6}_{-11.3}$ PDF( $+\alpha_S$ ): $\pm 4$
t̄W (aMC@NLO+Pythia8)	QCD Scale: $^{+12.9}_{-11.5}$ PDF( $+\alpha_S$ ): $\pm 3.4$
tHjb (aMC@NLO+Pythia8)	QCD Scale: $^{+6.5}_{-14.9}$ PDF( $+\alpha_S$ ): $\pm 3.7$
WtH (aMC@NLO+Pythia8)	QCD Scale: $^{+5.0}_{-6.7}$ PDF( $+\alpha_S$ ): $\pm 6.3$
VV (Sherpa 2.2.1)	$\pm 50$
Others	$\pm 50$

Table 35: Summary of theoretical uncertainties for MC predictions in the analysis.

1316 Additional uncertainties to account for t̄W mismodelling are also applied. These include  
 1317 a “Generator” uncertainty, based on a comparison between the nominal Sherpa 2.2.5 sample,  
 1318 and the formerly used aMC@NLO sample, and an “Extra radiation” uncertainty, which includes  
 1319 renormalisation and factorisation scale variations of the Sherpa 2.2.5 sample.

## 1320 17 Results

1321 A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p<sub>T</sub> spectrum  
 1322 in the three signal regions, 2lSS, 3lS, and 3lF. The signal is split into high and low p<sub>T</sub> samples,  
 1323 based on whether the truth p<sub>T</sub> of the Higgs is above or below 150 GeV. The parameters  $\mu_{t\bar{t}H \text{ high } p_T}$   
 1324 and  $\mu_{t\bar{t}H \text{ low } p_T}$ , where  $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$ , are extracted from the fit, signifying the difference  
 1325 between the observed value and the theory prediction. Unblinded results are shown for the 79.8  
 1326 fb<sup>-1</sup> data set, as well as MC only projections of results using the full Run-2, 139 fb<sup>-1</sup> dataset.

1327 As described in Section 16, there are 229 systematic uncertainties that are considered  
 1328 as NPs in the fit. These NP s are constrained by Gaussian or log-normal probability density  
 1329 functions. The latter are used for normalisation factors to ensure that they are always positive.  
 1330 The expected number of signal and background events are functions of the likelihood. The prior  
 1331 for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its  
 1332 nominal value.

## 1333 17.1 Results - $79.8 \text{ fb}^{-1}$

1334 As the data collected from 2015-2017 has been unblinded for  $t\bar{t}H$  – ML channels, represent-  
 1335 ing  $79.8 \text{ fb}^{-1}$ , those events are unblinded. The predicted Higgs  $p_T$  spectrum is fit to data  
 1336 simultaneously in each of the three signal regions shown in Figure 17.1.

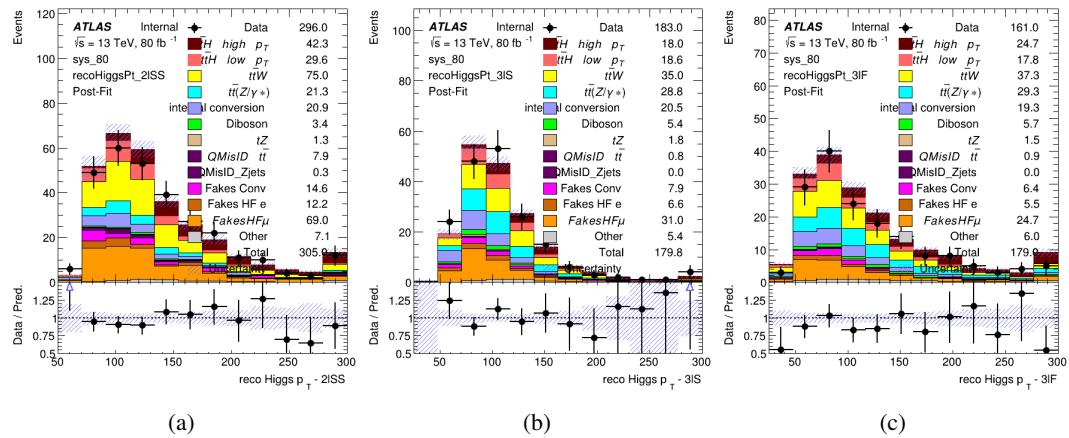


Figure 17.1: Post-fit distributions of the reconstructed Higgs  $p_T$  in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for  $79.8 \text{ fb}^{-1}$  of MC

1337 A post-fit summary of the fitted regions is shown in Figure 17.2.

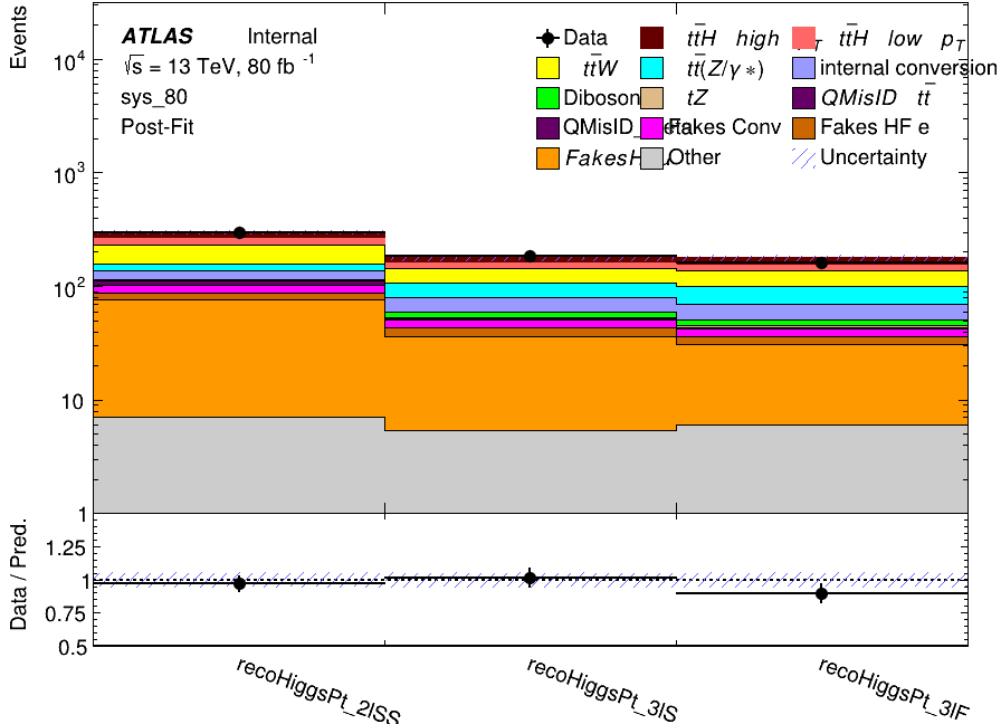


Figure 17.2: Post-fit summary of the yields in each signal region.

1338 The measured  $\mu$  values for high and low  $p_T$  Higgs production obtained from the fit  
 1339 are shown in 36. A significance of  $1.7\sigma$  is observed for  $t\bar{t}H$  high  $p_T$ , and  $2.1\sigma$  is measured for  
 1340  $t\bar{t}H$  low  $p_T$ .

$$\mu_{t\bar{t}H \text{ high } p_T} = 2.1^{+0.62}_{-0.59} (\text{stat})^{+0.40}_{-0.43} (\text{sys})$$

$$\mu_{t\bar{t}H \text{ low } p_T} = 0.83^{+0.39}_{-0.40} (\text{stat})^{+0.51}_{-0.53} (\text{sys})$$

Table 36: Best fit  $\mu$  values for  $t\bar{t}H$  high  $p_T$  and  $t\bar{t}H$  low  $p_T$ , where  $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1341 The most prominent sources of systematic uncertainty, as measured by their impact on  
 1342  $\mu_{t\bar{t}H \text{ high } p_T}$ , are summarized in Table 37.

Uncertainty Source	$\Delta\mu$	
Jet Energy Scale	0.25	0.23
t̄H cross-section (QCD scale)	-0.11	0.21
Luminosity	-0.13	0.14
Flavor Tagging	0.14	0.13
t̄W cross-section (QCD scale)	-0.12	0.11
Higgs Branching Ratio	-0.1	0.11
t̄H cross-section (PDF)	-0.07	0.08
Electron ID	-0.06	0.06
Non-prompt Muon Normalization	-0.05	0.06
t̄Z cross-section (QCD scale)	-0.05	0.05
Diboson cross-section	-0.05	0.05
Fake muon modelling	-0.04	0.04
Total	0.40	0.43

Table 37: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H high p<sub>T</sub>.

<sup>1343</sup> The most significant sources of uncertainty on the measurement of t̄H - low p<sub>T</sub> are shown  
<sup>1344</sup> in Table 38.

Uncertainty Source	$\Delta\mu$	
Internal Conversions	-0.26	0.26
Luminosity	-0.16	0.17
Non-prompt Muon Normalization	-0.16	0.16
t̄W cross-section (QCD scale)	-0.17	0.15
Jet Energy Scal	0.15	0.15
Non-prompt Electron Modelling	-0.13	0.14
Flavor Tagging	0.13	0.13
Non-prompt Muon Modelling	-0.12	0.13
Non-prompt Electron Normalization	-0.11	0.11
t̄Z cross-section (QCD scale)	-0.08	0.09
Diboson Cross-section	-0.07	0.07
Total	0.51	0.53

Table 38: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H low p<sub>T</sub>.

<sup>1345</sup> The ranking and impact of those nuisance parameters with the largest contribution to the

1346 overall uncertainty is shown in Figure 17.3.

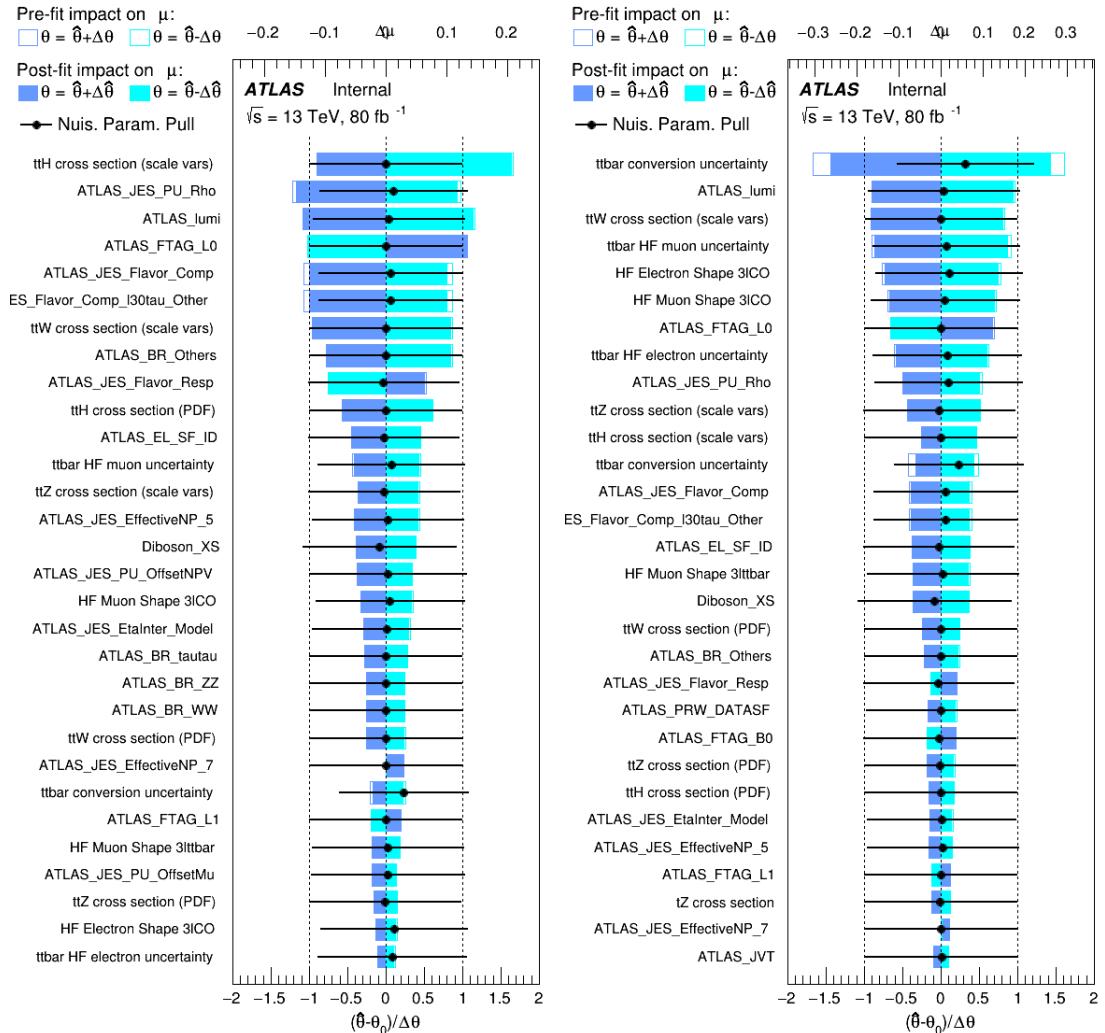


Figure 17.3: Impact of systematic uncertainties on the measurement of high  $p_T$  (left) and low  $p_T$  (right)  $t\bar{t}H$  events

1347

The background composition of each of the fit regions is shown in Figure 17.4.

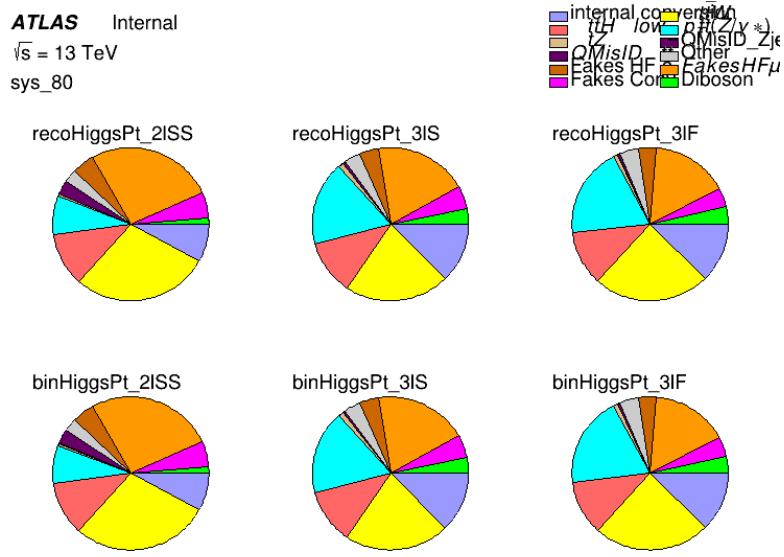


Figure 17.4: Background composition of the fit regions.

## 1348 17.2 Projected Results - $139 \text{ fb}^{-1}$

1349 As data collected in 2018 has not yet been unblinded for  $t\bar{t}H$  – ML at the time of this note, data  
1350 from that year remains blinded. Instead, an Asimov fit is performed - with the MC prediction  
1351 being used both as the SM prediction as well as the data in the fit - in order to give expected  
1352 results.

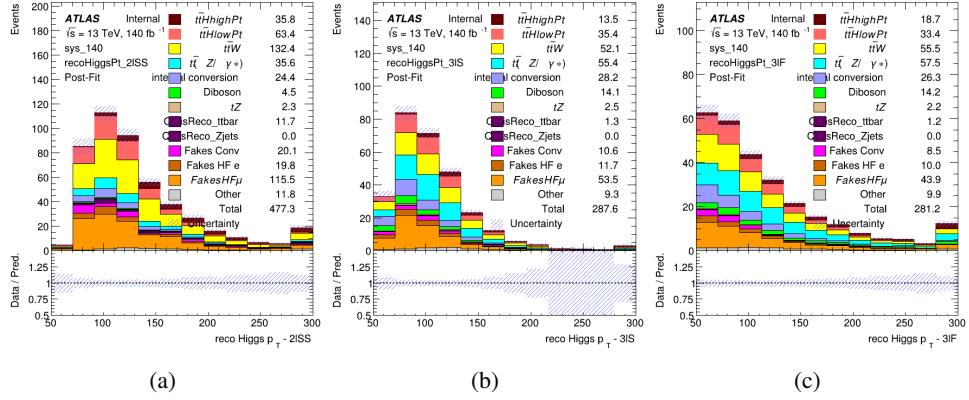


Figure 17.5: Blinded post-fit distributions of the reconstructed Higgs  $p_T$  in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for  $139 \text{ fb}^{-1}$  of data

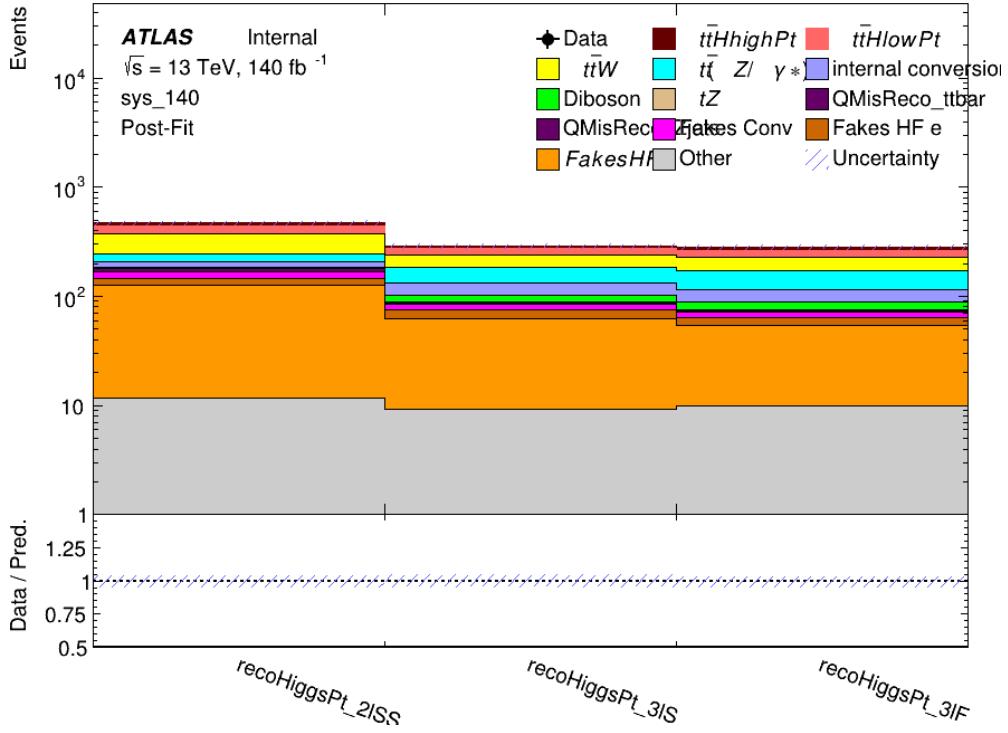


Figure 17.6: Post-fit summary of fit.

Projected uncertainties on the  $\mu$  values extracted from the fit for high and low  $p_T$  Higgs are

1354 shown in 39. A significance of  $2.0\sigma$  is expected for  $t\bar{t}H$  high  $p_T$ , and a projected significance

1355  $2.3\sigma$  is extracted for  $t\bar{t}H$  low  $p_T$ .

$$\mu_{t\bar{t}H \text{ high } p_T} = 1.00^{+0.45}_{-0.43}(\text{stat})^{+0.30}_{-0.31}(\text{sys})$$

$$\mu_{t\bar{t}H \text{ low } p_T} = 1.00^{+0.29}_{-0.30}(\text{stat})^{+0.48}_{-0.50}(\text{sys})$$

Table 39: Best fit  $\mu$  values for  $t\bar{t}H$  high  $p_T$  and  $t\bar{t}H$  low  $p_T$ , where  $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1356 The most prominent sources of systematic uncertainty, as measured by their impact on

1357  $\mu_{t\bar{t}H \text{ high } p_T}$ , are summarized in Table 40.

Uncertainty Source	$\Delta\mu$	
Jet Energy Scale	0.19	0.17
$t\bar{t}W$ Cross-section (QCD Scale)	-0.12	0.11
Luminosity	-0.1	0.11
Flavor Tagging	0.1	0.1
$t\bar{t}H$ Cross-section (QCD Scale)	-0.05	0.1
$t\bar{t}Z$ Cross-section (QCD Scale)	-0.05	0.06
Non-prompt Muon Normalization	-0.05	0.05
Higgs Branching Ratio	-0.05	0.05
Diboson Cross-section	-0.04	0.05
Non-prompt Muon Modelling	-0.04	0.04
$t\bar{t}H$ Cross-section (PDF)	-0.03	0.04
Electron ID	-0.04	0.04
$t\bar{t}W$ Cross-section (PDF)	-0.03	0.03
Total	0.30	0.31

Table 40: Summary of the most significant sources of systematic uncertainty on the measurement of  $t\bar{t}H$  high  $p_T$ .

1358 The most significant sources of systematic uncertainty on  $t\bar{t}H$  low  $p_T$  are summarized in

1359 Table 41.

Uncertainty Source	$\Delta\mu$	
Internal Conversions	-0.18	0.2
Jet Energy Scale	0.19	0.16
Non-prompt Muon Normalization	-0.16	0.17
Luminosity	-0.15	0.17
t̄tW Cross-section (QCD Scale)	-0.17	0.15
Non-prompt Electron Modelling	-0.13	0.14
Non-prompt Muon Modelling	-0.13	0.13
Flavor Tagging	0.13	0.12
Non-prompt Electron Normalization	-0.1	0.11
t̄tZ Cross-section (QCD Scale)	-0.07	0.09
t̄tH Cross-section (QCD Scale)	-0.05	0.1
Total	0.48	0.50

Table 41: Summary of the most significant sources of systematic uncertainty on the measurement of t̄tH low p<sub>T</sub>.

<sup>1360</sup> The ranking and impact of those nuisance parameters with the largest contribution to the  
<sup>1361</sup> overall uncertainty is shown in Figure 17.7.

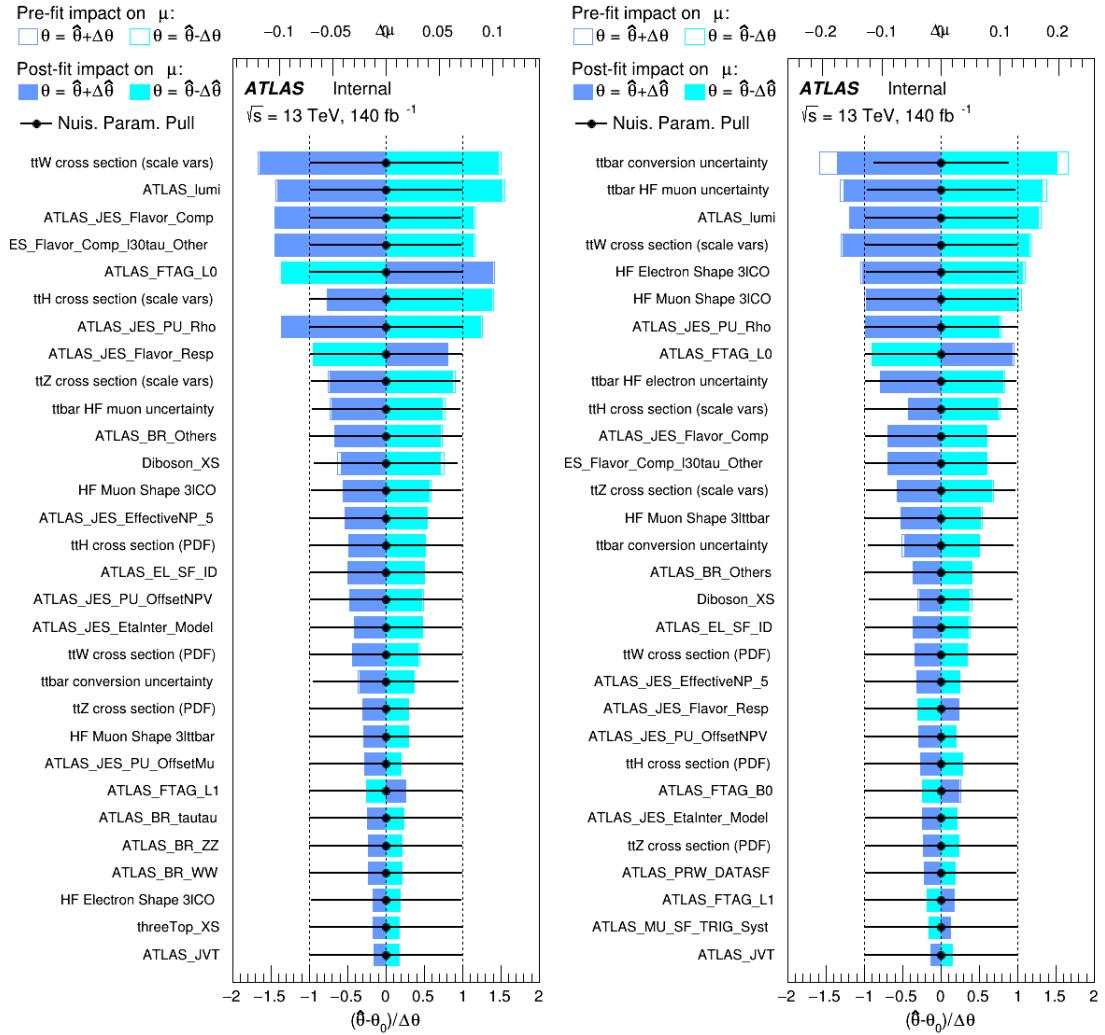


Figure 17.7: Impact of systematic uncertainties on the measurement of high  $p_T$  (left) and low  $p_T$  (right)  $t\bar{t}H$  events

1362

The background composition of each of the fit regions is shown in Figure 17.8.

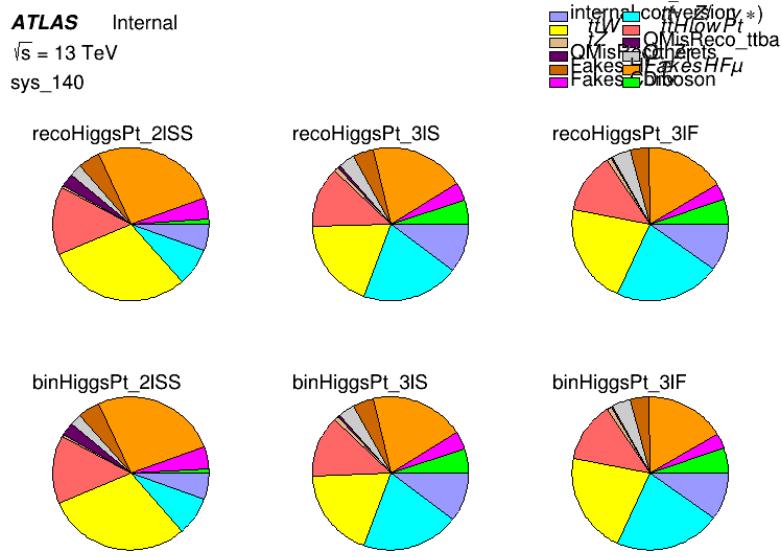


Figure 17.8: Background composition of the fit regions.

## Part VI

### Conclusion

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

## References

- 1369 [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*  
1370 *Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,  
1371 arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- 1372 [2] R. Oerter, *The theory of almost everything : the Standard Model, the unsung triumph of*  
1373 *modern physics*, Pi Press, 2006, ISBN: 978-0-13-236678-6.
- 1374 [3] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,  
1375 Phys. Rev. Lett. 13 (Oct, 1964) 508-509 ().
- 1376 [4] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*,  
1377 Phys. Rev. **127** (3 1962) 965,  
1378 URL: <https://link.aps.org/doi/10.1103/PhysRev.127.965>.
- 1379 [5] J. Ellis, *Higgs Physics*,  
1380 (2013) 117, 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of  
1381 High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report,  
1382 URL: <https://cds.cern.ch/record/1638469>.
- 1383 [6] T. A. Collaboration, *Observation of Higgs boson production in association with a top*  
1384 *quark pair at the LHC with the ATLAS detector*, Physics Letters B **784** (2018) 173,  
1385 ISSN: 0370-2693, URL:  
1386 <https://www.sciencedirect.com/science/article/pii/S0370269318305732>.

- 1388 [7] *Evidence for the associated production of the Higgs boson and a top quark pair with the*  
1389 *ATLAS detector*, tech. rep. ATLAS-CONF-2017-077, CERN, 2017,  
1390 URL: <https://cds.cern.ch/record/2291405>.
- 1391 [8] A. M. Sirunyan et al., *Observation of  $t\bar{t}H$  Production*,  
1392 *Physical Review Letters* **120** (2018), ISSN: 1079-7114,  
1393 URL: <http://dx.doi.org/10.1103/PhysRevLett.120.231801>.
- 1394 [9] B. Dumont, S. Fichet and G. von Gersdorff,  
1395 *A Bayesian view of the Higgs sector with higher dimensional operators*,  
1396 *Journal of High Energy Physics* **2013** (2013), ISSN: 1029-8479,  
1397 URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)065](http://dx.doi.org/10.1007/JHEP07(2013)065).
- 1398 [10] S. Banerjee, S. Mukhopadhyay and B. Mukhopadhyaya,  
1399 *Higher dimensional operators and the LHC Higgs data: The role of modified kinematics*,  
1400 *Physical Review D* **89** (2014), ISSN: 1550-2368,  
1401 URL: <http://dx.doi.org/10.1103/PhysRevD.89.053010>.
- 1402 [11] M. Zinser, ‘The Large Hadron Collider’, *Search for New Heavy Charged Bosons and*  
1403 *Measurement of High-Mass Drell-Yan Production in Proton—Proton Collisions*,  
1404 Springer International Publishing, 2018 47, ISBN: 978-3-030-00650-1,  
1405 URL: [https://doi.org/10.1007/978-3-030-00650-1\\_4](https://doi.org/10.1007/978-3-030-00650-1_4).
- 1406 [12] *Detector and Technology*, URL: <https://atlas.cern/discover/detector>.
- 1407 [13] M. Marjanovic, *ATLAS Tile calorimeter calibration and monitoring systems*, 2018.

- 1408 [14] ATLAS Collaboration, *Observation of electroweak  $W^\pm Z$  boson pair production in*  
 1409 *association with two jets in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector,*  
 1410 [Phys. Lett. B793 \(2019\) 469](#), arXiv: [1812.09740 \[hep-ex\]](#).
- 1411 [15] ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 7$  TeV using*  
 1412 *the ATLAS detector at the LHC*, [Eur. Phys. J. C 71 \(2011\) 1630](#),  
 1413 arXiv: [1101.2185 \[hep-ex\]](#).
- 1414 [16] ATLAS Collaboration, *Performance of the ATLAS detector using first collision data,*  
 1415 [JHEP 09 \(2010\) 056](#), arXiv: [1005.5254 \[hep-ex\]](#).
- 1416 [17] S. Agostinelli et al., *GEANT4: A Simulation toolkit,*  
 1417 [Nucl. Instrum. Meth. A506 \(2003\) 250](#).
- 1418 [18] T. Gleisberg et al., *Event generation with SHERPA 1.1*, [JHEP 02 \(2009\) 007](#),  
 1419 arXiv: [0811.4622 \[hep-ph\]](#).
- 1420 [19] H.-L. Lai et al., *New parton distributions for collider physics*,  
 1421 [Phys. Rev. D 82 \(2010\) 074024](#), arXiv: [1007.2241 \[hep-ph\]](#).
- 1422 [20] J. Alwall et al., *The automated computation of tree-level and next-to-leading order*  
 1423 *differential cross sections, and their matching to parton shower simulations*,  
 1424 [JHEP 07 \(2014\) 079](#), arXiv: [1405.0301 \[hep-ph\]](#).
- 1425 [21] R. D. Ball et al., *Parton distributions for the LHC Run II*, [JHEP 04 \(2015\) 040](#),  
 1426 arXiv: [1410.8849 \[hep-ph\]](#).
- 1427 [22] M. Bahr et al., *Herwig++ Physics and Manual*, [Eur. Phys. J. C 58 \(2008\) 639](#),  
 1428 arXiv: [0803.0883 \[hep-ph\]](#).

- 1429 [23] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244,  
1430 arXiv: [1207.1303 \[hep-ph\]](#).
- 1431 [24] S. Frixione, G. Ridolfi and P. Nason,  
1432 *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*,  
1433 JHEP **09** (2007) 126, arXiv: [0707.3088 \[hep-ph\]](#).
- 1434 [25] E. Re, *Single-top Wt-channel production matched with parton showers using the*  
1435 *POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: [1009.2450 \[hep-ph\]](#).
- 1436 [26] ATLAS Collaboration,  
1437 *Electron reconstruction and identification efficiency measurements with the ATLAS*  
1438 *detector using the 2011 LHC proton-proton collision data*,  
1439 Eur. Phys. J. C **74** (2014) 2941, arXiv: [1404.2240 \[hep-ex\]](#).
- 1440 [27] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the*  
1441 *ATLAS detector using 2011 and 2012 LHC proton–proton collision data*,  
1442 Eur. Phys. J. C **74** (2014) 3130, arXiv: [1407.3935 \[hep-ex\]](#).
- 1443 [28] ATLAS Collaboration, *Observation of the Higgs boson production in association with a*  
1444 *t̄t pair with the ATLAS detector*, PoS EPS-HEP2019 (2020) 325.
- 1445 [29] M. Cacciari, G. P. Salam and G. Soyez, *The anti-ktjet clustering algorithm*,  
1446 Journal of High Energy Physics **2008** (2008) 063, ISSN: 1029-8479,  
1447 URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/063>.

- 1448 [30] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp*  
1449 *collisions at  $\sqrt{s} = 8 \text{ TeV}$  using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 581,  
1450 arXiv: [1510.03823 \[hep-ex\]](https://arxiv.org/abs/1510.03823).
- 1451 [31] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector*  
1452 *using  $t\bar{t}$  events at  $\sqrt{s} = 13 \text{ TeV}$* , *Journal of High Energy Physics* **2018** (2018),  
1453 ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP08\(2018\)089](http://dx.doi.org/10.1007/JHEP08(2018)089).
- 1454 [32] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*  
1455 *with the ATLAS detector using proton–proton collisions at  $\sqrt{s} = 13 \text{ TeV}$* ,  
1456 *The European Physical Journal C* **78** (2018), ISSN: 1434-6052,  
1457 URL: <http://dx.doi.org/10.1140/epjc/s10052-018-6288-9>.
- 1458 [33] T. Chen and C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’,  
1459 *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*  
1460 *Discovery and Data Mining*, KDD ’16, ACM, 2016 785, ISBN: 978-1-4503-4232-2,  
1461 URL: <http://doi.acm.org/10.1145/2939672.2939785>.
- 1462 [34] ATLAS Collaboration, *Measurement of the fiducial and differential cross-section of a*  
1463 *top quark pair in association with a Z boson at 13 TeV with the ATLAS detector*, (2019),  
1464 URL: <https://cds.cern.ch/record/2672207>.
- 1465 [35] *Luminosity determination in pp collisions at  $\sqrt{s} = 13 \text{ TeV}$  using the ATLAS detector at*  
1466 *the LHC*, (2019).

- 1467 [36] ATLAS Collaboration,  
1468     *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS,*  
1469     **JINST** **13** (2018) P07017.
- 1470 [37] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties*  
1471     *in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector,*  
1472     **Phys. Rev. D** **96** (2017) 072002, arXiv: [1703.09665 \[hep-ex\]](https://arxiv.org/abs/1703.09665).
- 1473 [38] A. Collaboration, *Performance of  $b$ -jet identification in the ATLAS experiment,*  
1474     **Journal of Instrumentation** **11** (2016) P04008,  
1475     URL: <http://stacks.iop.org/1748-0221/11/i=04/a=P04008>.
- 1476 [39] ATLAS Collaboration, *Observation of the associated production of a top quark and a Z*  
1477     *boson in pp collisions at  $\sqrt{s}= 13$  TeV with the ATLAS detector,*  
1478     **Journal of High Energy Physics** **2020** (2020), ISSN: 1029-8479,  
1479     URL: [https://dx.doi.org/10.1007/JHEP07\(2020\)124](https://dx.doi.org/10.1007/JHEP07(2020)124).
- 1480 [40] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions from Powheg*  
1481     *and Sherpa to ATLAS measurements of top pair production at 7 TeV,*  
1482     ATL-PHYS-PUB-2015-011, 2015, URL: <https://cds.cern.ch/record/2020602>.
- 1483 [41] ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11,*  
1484     ATL-PHYS-PUB-2011-009, 2011, URL: <https://cds.cern.ch/record/1363300>.
- 1485 [42] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using*  
1486     *the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016,  
1487     URL: <https://cds.cern.ch/record/2157687>.

- 1488 [43] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets*  
1489       *Reconstructed in the ATLAS Detector at  $\sqrt{s} = 13 \text{ TeV}$* , ATL-PHYS-PUB-2015-015,  
1490       2015, URL: <https://cds.cern.ch/record/2037613>.
- 1491 [44] ATLAS Collaboration,  
1492       *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector,*  
1493       ATLAS-CONF-2015-029, 2015, URL: <https://cds.cern.ch/record/2037702>.
- 1494 [45] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*  
1495       *with the ATLAS detector in the first proton–proton collisions at  $\sqrt{s} = 13 \text{ TeV}$ ,*  
1496       ATL-PHYS-PUB-2015-027, 2015, URL: <https://cds.cern.ch/record/2037904>.
- 1497 [46] G. Aad, T. Abajyan, B. Abbott and et al, *Jet energy resolution in proton-proton collisions*  
1498       *at  $\sqrt{s} = 7 \text{ TeV}$  recorded in 2010 with the ATLAS detector,*  
1499       The European Physical Journal C **73** (2013) 2306, ISSN: 1434-6052,  
1500       URL: <https://doi.org/10.1140/epjc/s10052-013-2306-0>.
- 1501 [47] R. Narayan et al., *Measurement of the total and differential cross sections of a*  
1502       *top-quark-antiquark pair in association with a W boson in proton-proton collisions at a*  
1503       *centre-of-mass energy of 13 TeV with ATLAS detector at the Large Hadron Collider,*  
1504       tech. rep. ATL-COM-PHYS-2020-217, CERN, 2020,  
1505       URL: <https://cds.cern.ch/record/2712986>.
- 1506 [48] R. Frederix et al., *The automation of next-to-leading order electroweak calculations,*  
1507       Journal of High Energy Physics **2018** (2018), ISSN: 1029-8479,  
1508       URL: [http://dx.doi.org/10.1007/JHEP07\(2018\)185](http://dx.doi.org/10.1007/JHEP07(2018)185).

<sup>1509</sup> **List of contributions**

<sup>1510</sup>

<sub>1511</sub> **Part VII**

<sub>1512</sub> **Appendices**

<sub>1513</sub> **A Non-prompt lepton MVA**

1514        A lepton MVA has been developed to better reject non-prompt leptons than standard cut  
1515        based selections based upon impact parameter, isolation and PID. The name of this MVA is  
1516        **PromptLeptonVeto**. The full set of studies and detailed explanation can be found in [47].

1517        The decays of  $W$  and  $Z$  bosons are commonly selected by the identification of one or two  
1518        electrons or muons. The negligible lifetimes of these bosons mean that the leptons produced in the  
1519        decay originate from the interaction vertex and are thus labelled “prompt”. Analyses using these  
1520        light leptons impose strict reconstruction quality, isolation and impact parameter requirements  
1521        to remove “fake” leptons. A significant source of the fake light leptons are non-prompt leptons  
1522        produced in decays of hadrons that contain bottom (b) or charm (c) quarks. Such hadrons  
1523        typically have microscopically significant lifetimes that can be detected experimentally.

1524        These non-prompt leptons can also pass the tight selection criteria. In analyses that  
1525        involve top ( $t$ ) quarks, which decay almost exclusively into a  $W$  boson and a b quark, non-  
1526        prompt leptons from the semileptonic decay of bottom and charm hadrons can be a significant  
1527        source of background events. This is particularly the case in the selection of same-sign dilepton  
1528        and multilepton final states.

1529        The main idea is to identify non-prompt light leptons using lifetime information associated  
1530        with a track jet that matches the selected light lepton. This lifetime information is computed  
1531        using tracks contained within the jet. Typically, lepton lifetime is determined using the impact  
1532        parameter of the track reconstructed by the inner tracking detector which is matched to the recon-  
1533        structed lepton. Using additional reconstructed charged particle tracks increases the precision of

1534 identifying the displaced decay vertex of bottom or charm hadrons that produced a non-prompt  
 1535 light lepton. The MVA also includes information related to the isolation of the lepton to reject  
 1536 non-prompt leptons.

1537 **PromptLeptonVeto** is a gradient boosted BDT. The training of the BDT is performed on  
 1538 leptons selected from the PowHEG+PYTHIA6 non-allhad  $t\bar{t}$  MC sample. Eight variables are used  
 1539 to train the BDT in order to discriminate between prompt and non-prompt leptons. The track  
 1540 jets that are matched to the non-prompt leptons correspond to jets initiated by b or c quarks, and  
 1541 may contain a displaced vertex. Consequently, three of the selected variables are used to identify  
 1542 b-tag jets by standard ATLAS flavour tagging algorithms. Two variables use the relationship  
 1543 between the track jet and lepton: the ratio of the lepton  $p_T$  with respect to the track jet  $p_T$  and  
 1544  $\Delta R$  between the lepton and the track jet axis. Finally three additional variables test whether the  
 1545 reconstructed lepton is isolated: the number of tracks collected by the track jet and the lepton  
 1546 track and calorimeter isolation variables. Table 42 describes the variables used to train the BDT  
 1547 algorithm. The choice of input variables has been extensively discussed with Egamma, Muon,  
 1548 Tracking, and Flavour Tagging CP groups.

Variable	Description
$N_{\text{track}}$ in track jet	Number of tracks collected by the track jet
$\text{IP2 log}(P_b/P_{\text{light}})$	Log-likelihood ratio between the b and light jet hypotheses with the IP2D algorithm
$\text{IP3 log}(P_b/P_{\text{light}})$	Log-likelihood ratio between the b and light jet hypotheses with the IP3D algorithm
$N_{\text{TrkAtVtx}} \text{ SV + JF}$	Number of tracks used in the secondary vertex found by the SV1 algorithm in addition to the number of tracks from secondary vertices found by the JetFitter algorithm with at least two tracks
$p_T^{\text{lepton}} / p_T^{\text{track jet}}$	The ratio of the lepton $p_T$ and the track jet $p_T$
$\Delta R(\text{lepton}, \text{track jet})$	$\Delta R$ between the lepton and the track jet axis
$p_T^{\text{VarCone30}} / p_T$	Lepton track isolation, with track collecting radius of $\Delta R < 0.3$
$E_T^{\text{TopoCone30}} / p_T$	Lepton calorimeter isolation, with topological cluster collecting radius of $\Delta R < 0.3$

Table 42: A table of the variables used in the training of **PromptLeptonVeto**.

1549 The output distribution of the BDT is shown in Figure A.

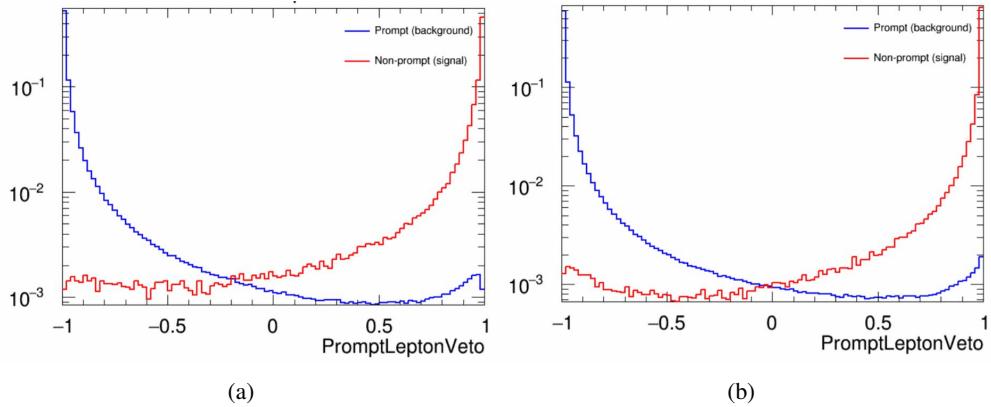


Figure A.1: Distribution of the PLV BDT discriminant for (a) electrons and (b) muons

1550 The ROC curve for the BDT response, compared to the standard `FixedCutTight` WP, is  
1551 shown in figure A, which shows a clear improvement when using this alternative training.

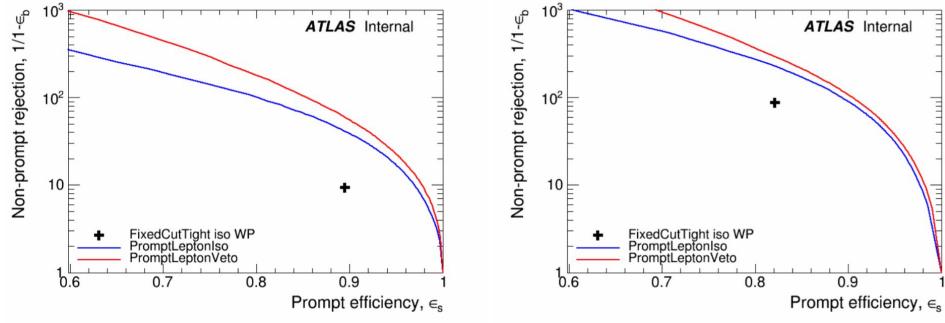


Figure A.2: ROC curves for the PLV as well as the performance of the standard FixedCutTight WP for (left) electrons and (right) muons

1552 A cutoff value of -0.7 for electrons and -0.5 for muons are chosen as the WPs for this  
 1553 MVA, based on an optimisation of  $S/\sqrt{B}$  performed in the preselection regions of the  $t\bar{t}H - ML$   
 1554 analysis, which have a signature similar to that of this analysis.

1555        The efficiency of the tight PromptLeptonVeto working point is measured using the tag  
1556        and probe method with  $Z \rightarrow \ell^+ \ell^-$  events. Such calibration are performed by analysers from  
1557        this analysis in communication with the Egamma and Muon combined performance groups. The  
1558        scale factor are approximately 0.92 for  $10 < p_T < 15$  GeV, and averaging at 0.98 to 0.99 for  
1559        higher  $p_T$  leptons. An extra systematic is applied to muons within  $\Delta R < 0.6$  of a calorimeter  
1560        jet, since there is a strong dependence on the scale factor due to the presence of these jets. For  
1561        electrons, the dominant systematics is coming from pile-up dependence. Overall the systematics  
1562        are a maximum of 3% at low  $p_T$  and decreasing at a function of  $p_T$ .

<sup>1563</sup> **B Supplementary WZ + Heavy Flavor Studies**

<sup>1564</sup> **B.1 Non-prompt CR Modelling**

1565 In order to further validate the modeling in each of the non-prompt CRs, additional  
 1566 kinematic plots are made in the Z+jets CR and  $t\bar{t}$  CR in each of the continuous b-tag regions,  
 1567 after the correction factors detailed in Section 8.3 have been applied.

1568 In the case of the Z+jets CR, the  $p_T$  spectrum of the lepton originating from the W  
 1569 candidate is shown, as this is the distribution used to extract the scale factor applied to Z+jets.

1570 These plots are shown in Figures B.1 and B.2.

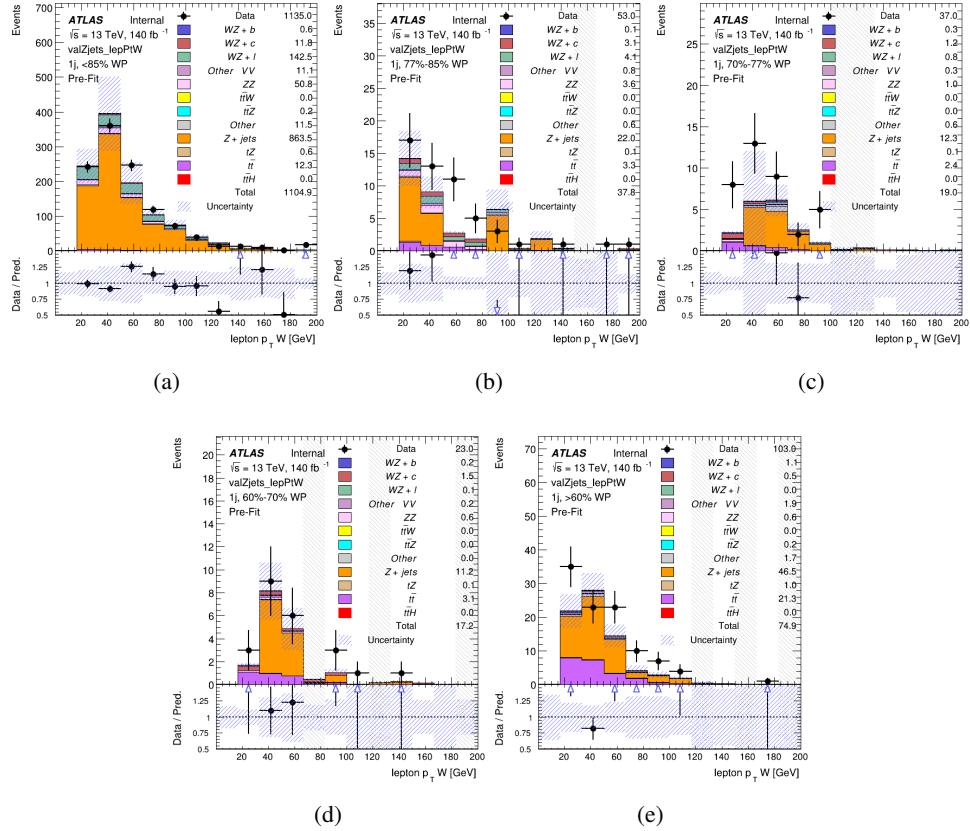


Figure B.1: Comparisons between the data and MC distributions of the  $p_T$  of the lepton originating from the W-candidate in the Z+jets CR for each of the 1-jet b-tag working point regions

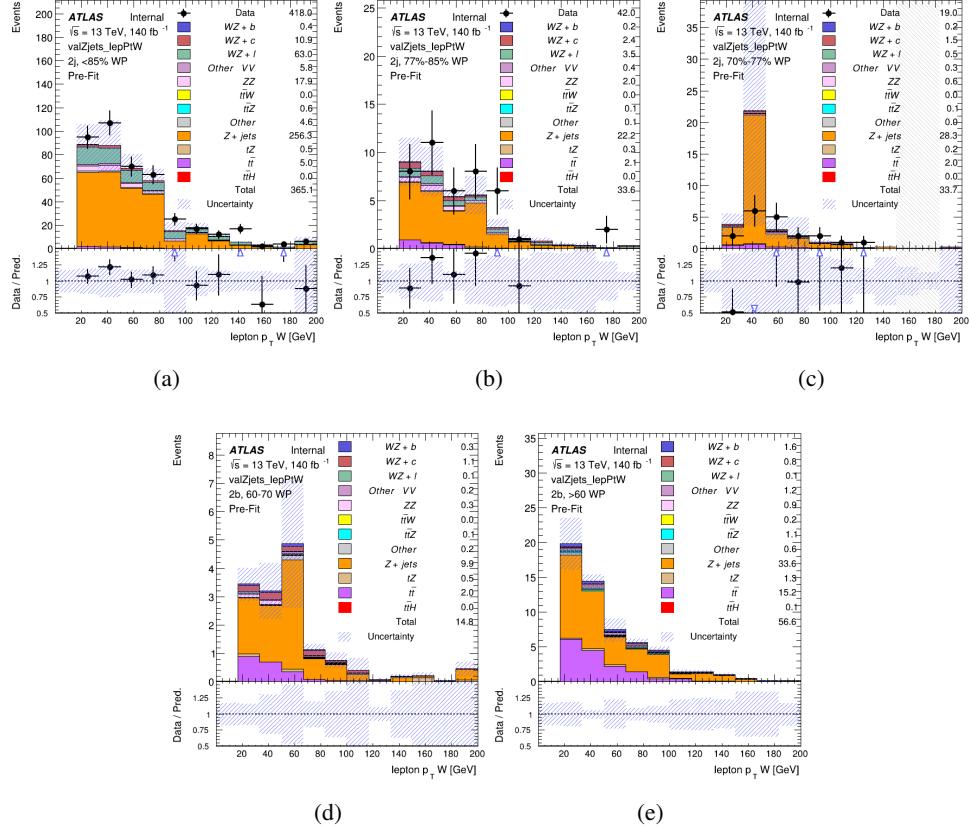


Figure B.2: Comparisons between the data and MC distributions of the  $p_T$  of the lepton originating from the W-candidate in the Z+jets CR for each of the 2-jet b-tag working point regions

1571 The same is shown for the  $t\bar{t}$  CR, but the  $p_T$  of the OS lepton is used instead as a  
 1572 representation of the modeling, as the lepton from the W is not well defined for  $t\bar{t}$  events. These  
 1573 plots are shown in Figures B.3 and B.4.

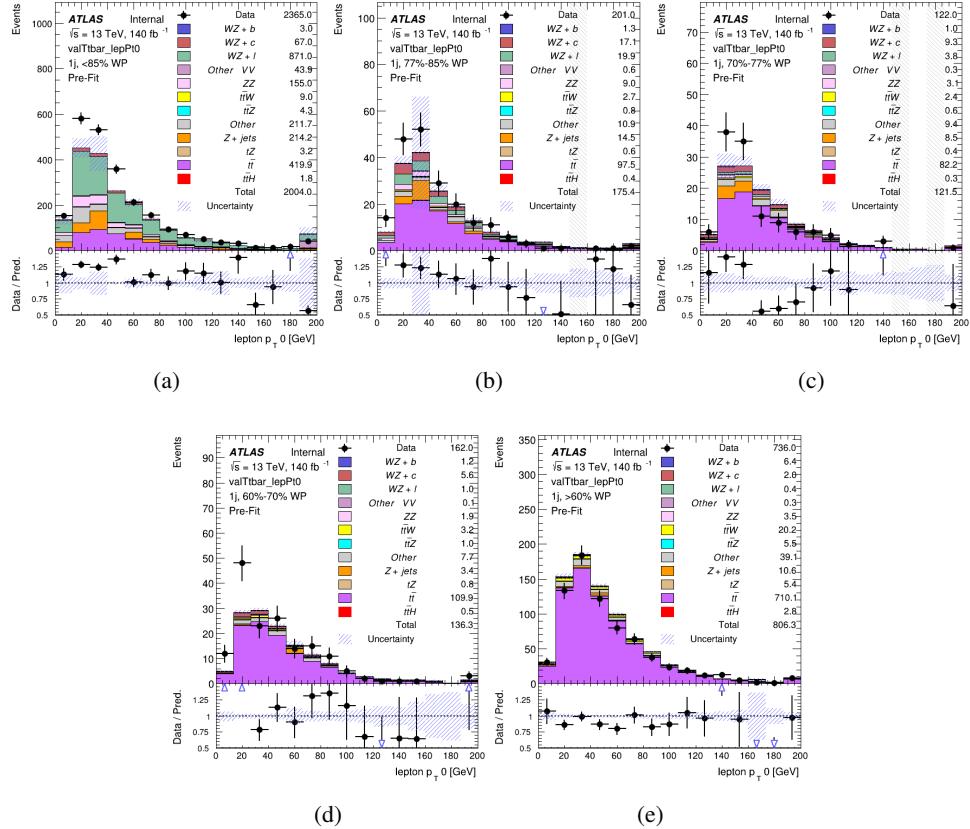


Figure B.3: Comparisons between the data and MC distributions of the  $p_T$  of the OS lepton in the  $t\bar{t}$  CR for each of the 1-jet b-tag working point regions

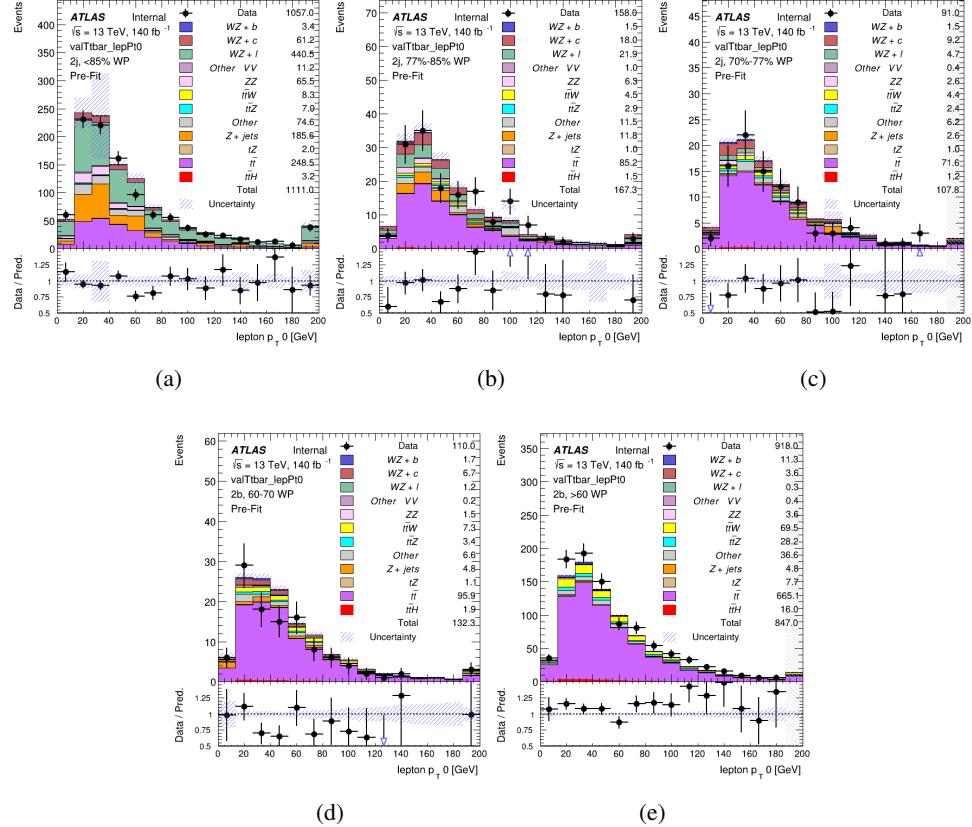


Figure B.4: Comparisons between the data and MC distributions of the  $p_T$  of the OS lepton in the  $t\bar{t}$  CR for each of the 2-jet b-tag working point regions

<sup>1574</sup> **B.2 tZ Interference Studies**

Because it includes an on-shell Z boson as well as a b-jet and W from the top decay, tZ production represents an identical final state to WZ + b-jet. This implies the possibility of matrix level interference between these two processes not accounted for in the Monte Carlo simulations, which consider the two processes independently. Truth level studies are performed in order to estimate the impact of these interference effects.

In order to estimate the matrix level interference effects between tZ and WZ + b-jet, two different sets of simulations are produced using **MADGRAPH 5** [48] - one which simulates these two processes independently, and another where they are produced simultaneously, such that interference effects are present. These two sets of samples are then compared, and the difference between them can be taken to represent any interference effects.

MadGraph simulations of 10,000 tZ and 10,000 WZ + b events are produced, along with 20,000 events where both are present, in the fiducial region where three leptons and at least one jet are produced.

A selection mimicking the preselection used in the main analysis is applied to the samples: The SS leptons are required to have  $p_T > 20$  GeV, and  $> 10$  GeV is required for the OS lepton. The associated b-jet is required to have  $p_T > 25$  GeV, and all physics objects are required to fall in a range of  $|\eta| < 2.5$ .

The overall cross-section with and without intereference effects agree within error, and no significant differences in the kinematic distributions are seen. It is therefore concluded that interference effects do not significantly impact the results.

1595 **B.3 Alternate tZ Inclusive Fit**

1596 **B.3.1 tZ Inclusive Fit**

1597 While tZ is often considered as a distinct process from WZ + b, this could also be considered  
 1598 part of the signal. Alternative studies are performed where, using the same framework as the  
 1599 nominal analysis, a measurement of WZ + b is performed that includes tZ as part of WZ+b.

1600 Because of this change, the tZ CR is no longer necessary, and only the five pseudo-  
 1601 continuous b-tag regions are used in the fit. Further, systematics related to the tZ cross-section  
 1602 are removed from the fit, as they are now encompassed by the normalization measurement of  
 1603 WZ + b. All other systematic uncertainties are carried over from the nominal analysis.

1604 An expected WZ + b cross-section of  $4.1^{+0.78}_{-0.74}(\text{stat})^{+0.53}_{-0.52}(\text{sys}) \text{ fb}$  is extracted from the  
 1605 fit, with an expected significance of  $4.0\sigma$ .

1606 The impact of the predominate systematics are summarized in Table 43.

Uncertainty Source	$\Delta\mu$	
WZ + light cross-section	0.08	-0.08
Jet Energy Scale	-0.06	0.08
Luminosity	-0.05	0.06
WZ + c cross-section	-0.04	0.05
Other Diboson + b cross-section	-0.04	0.04
WZ cross-section - QCD scale	-0.04	0.03
t̄ cross-section	-0.03	0.03
Jet Energy Resolution	-0.03	0.03
Flavor tagging	-0.03	0.03
Z+jets cross section	-0.02	0.02
Total Systematic Uncertainty	-0.15	0.16

Table 43: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

1607    **B.3.2 Floating tZ**

1608    In order to quantify the impact of the tZ uncertainty on the fit, an alternate fit strategy is used  
1609    where the tZ normalization is allowed to float. This normalization factor replaces the cross-  
1610    section uncertainty on tZ, and all other parameters of the fit remain the same.

1611           An uncertainty of 17% on the normalization of tZ is extracted from the fit, compared to a  
1612    theory uncertainty of 15% applied to the tZ cross-section. The measured uncertainties on WZ  
1613    remain the same.

## **C Supplementary $t\bar{t}H$ Differential Analysis Studies**

<sup>1614</sup> The following section provides details of the various MVAs as well as a few studies performed  
<sup>1615</sup> in support of this analysis, exploring alternate decisions and strategies.

### **<sup>1617</sup> C.1 Higgs Reconstruction Model Details**

#### **<sup>1618</sup> C.1.1 b-jet Identification Features - 2lSS**

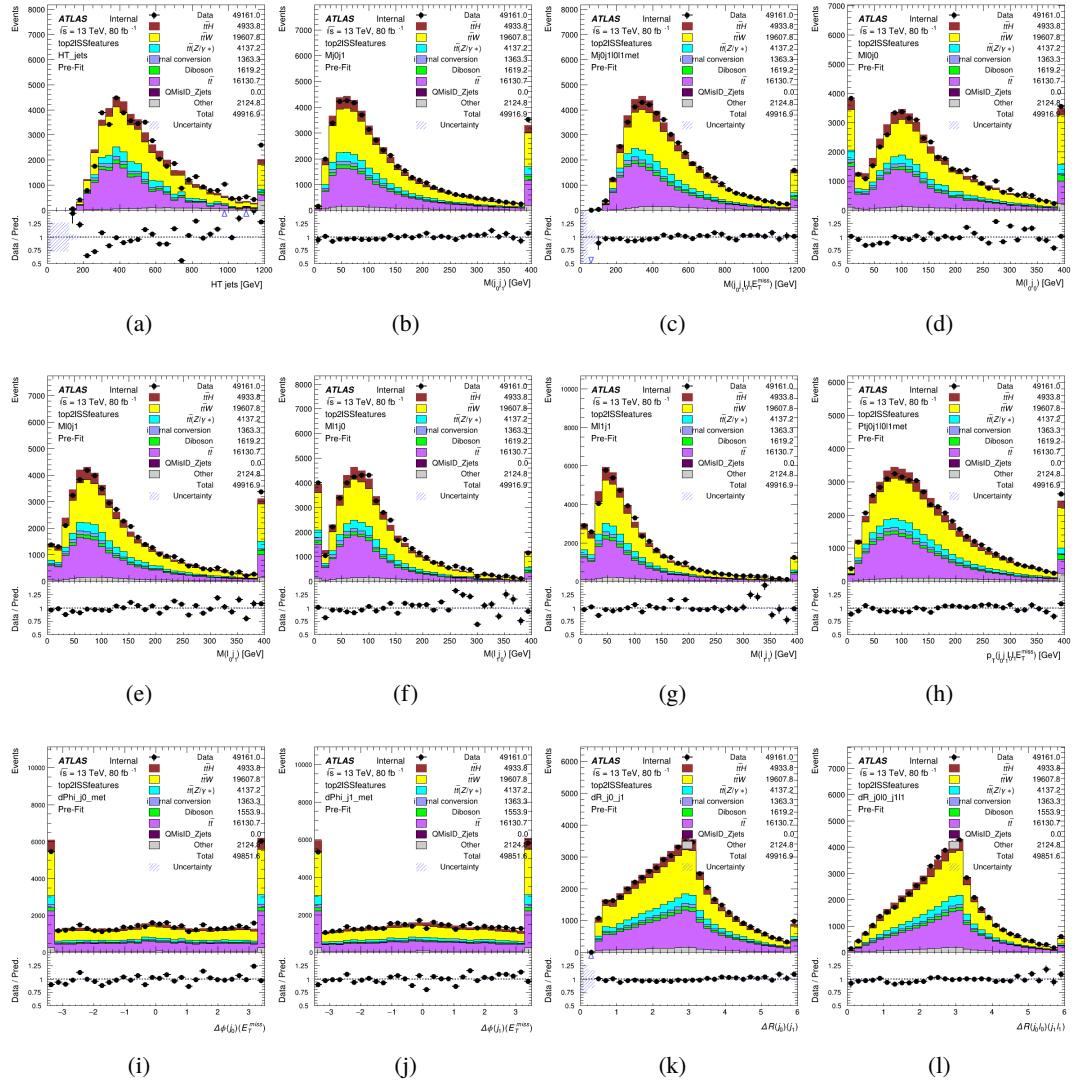


Figure C.1: Input features for top2lSS

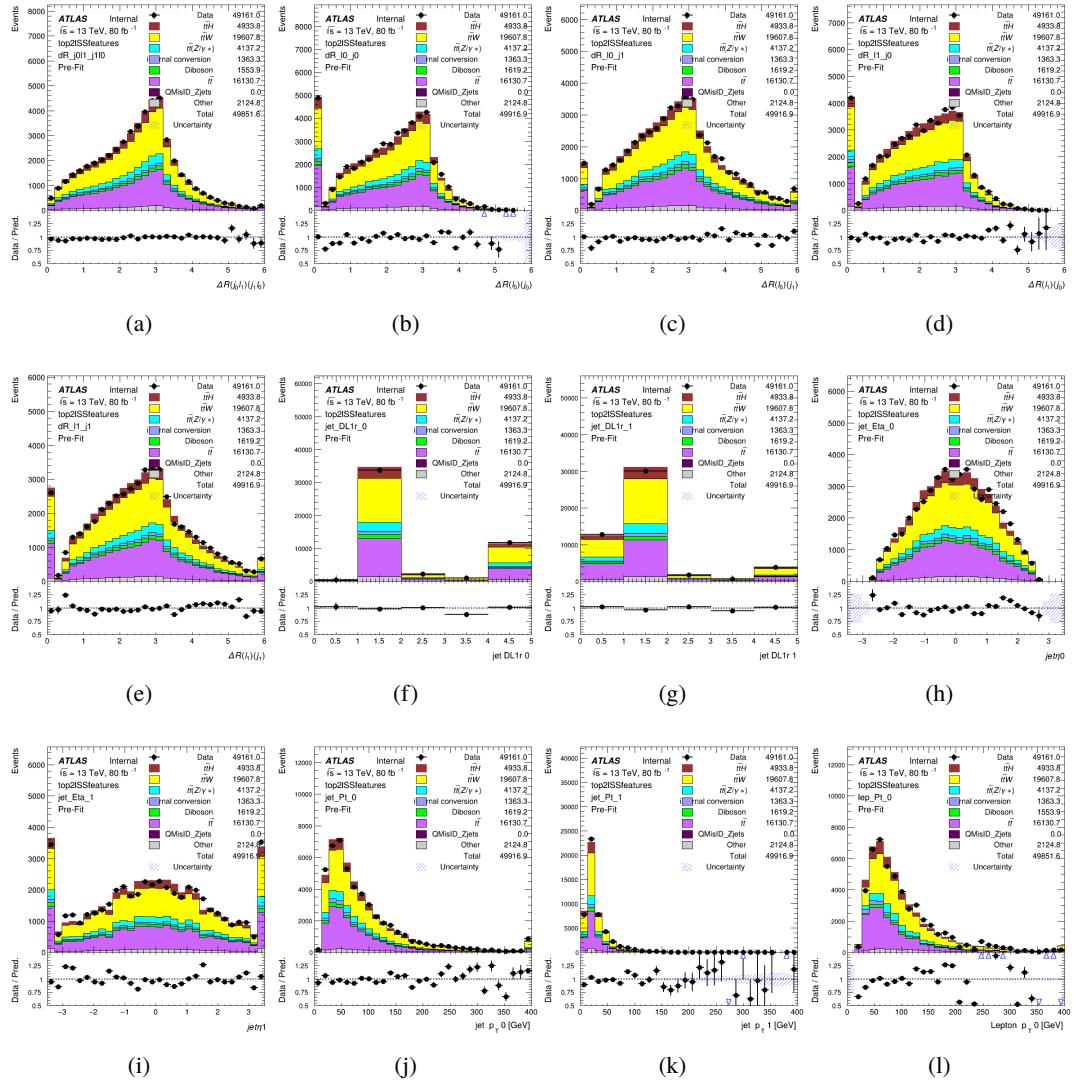


Figure C.2: Input features for top2lSS

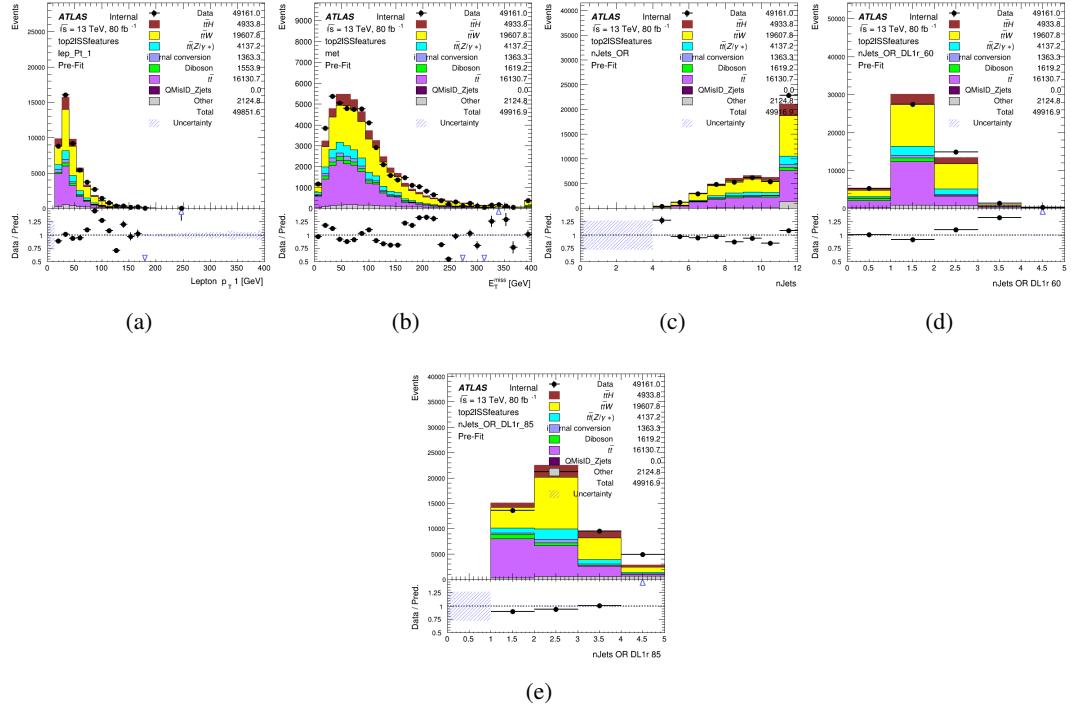


Figure C.3: Input features for top2ISS

1619 **C.1.2 b-jet Identification Features - 3l**

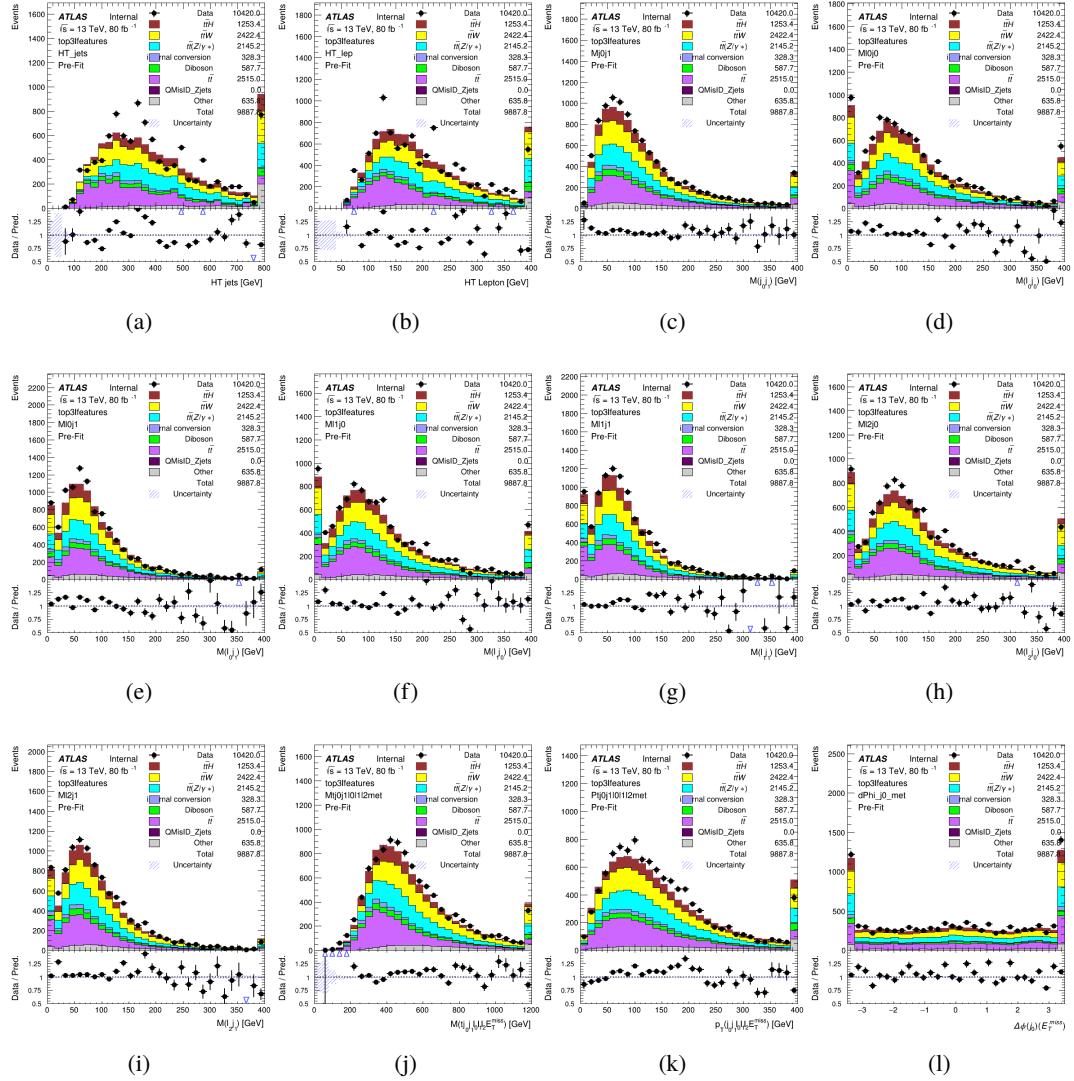


Figure C.4: Input features for top3l

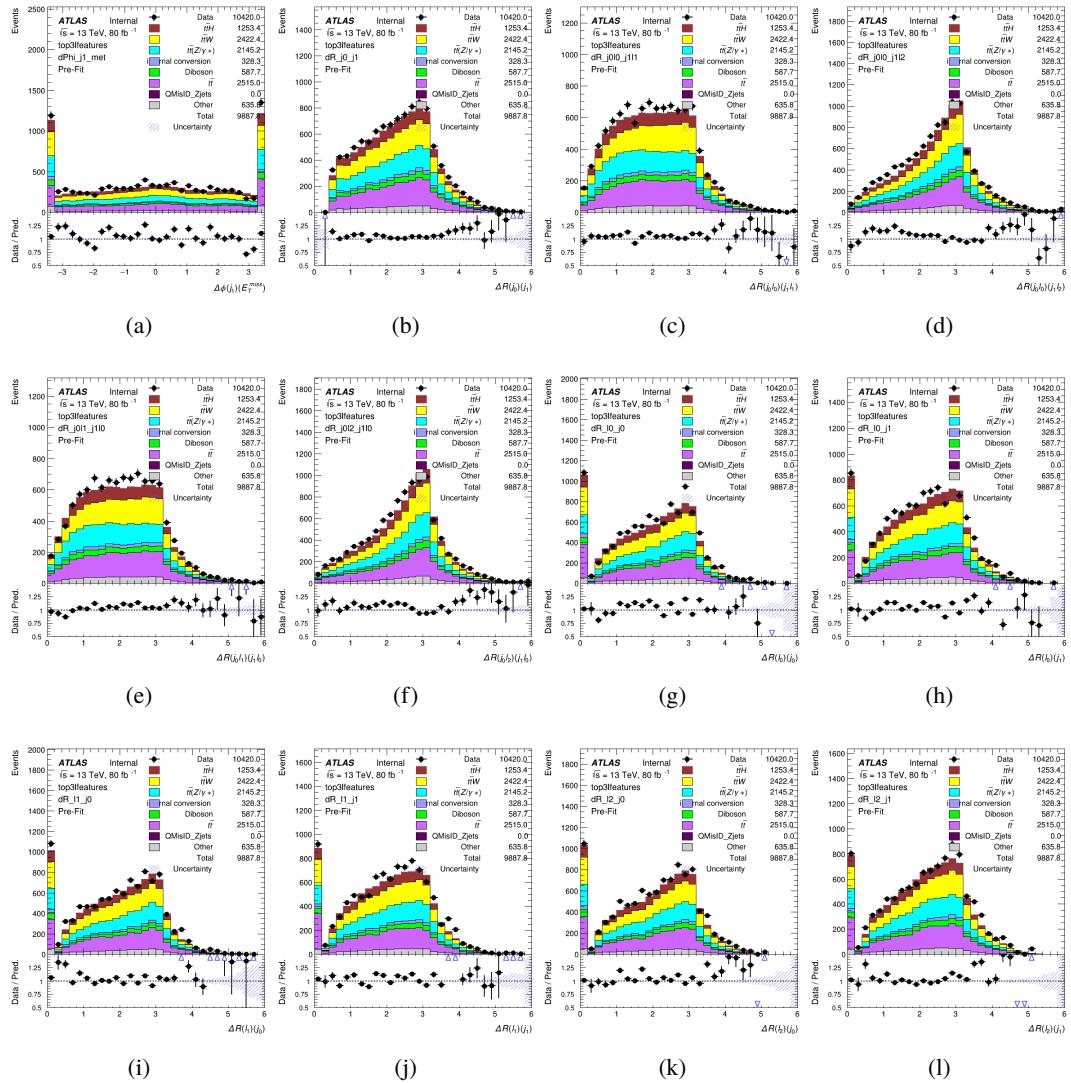


Figure C.5: Input features for top31

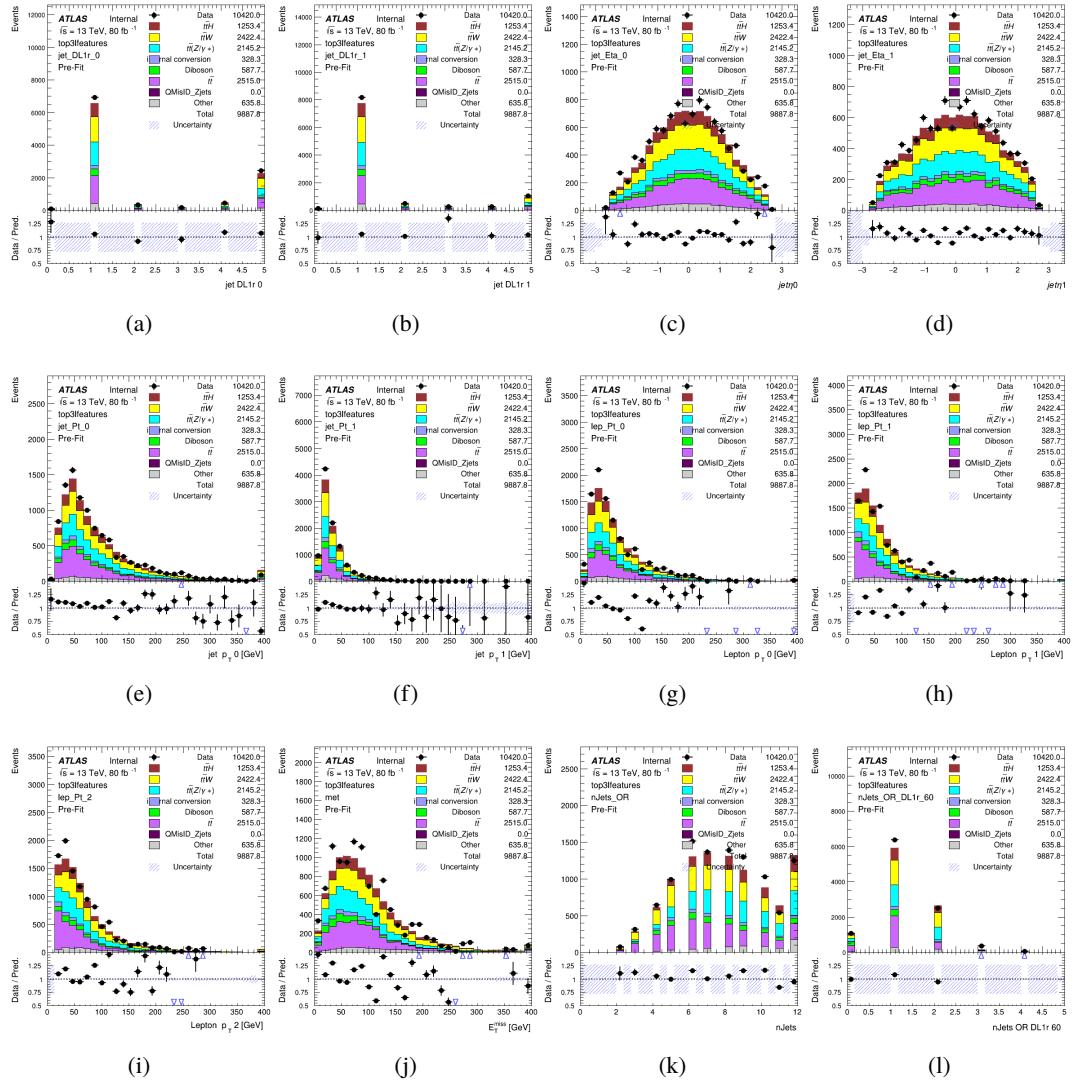
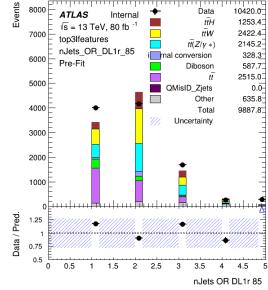


Figure C.6: Input features for top3l



(a)

Figure C.7: Input features for top3l

1620 **C.1.3 Higgs Reconstruction Features - 2lSS**

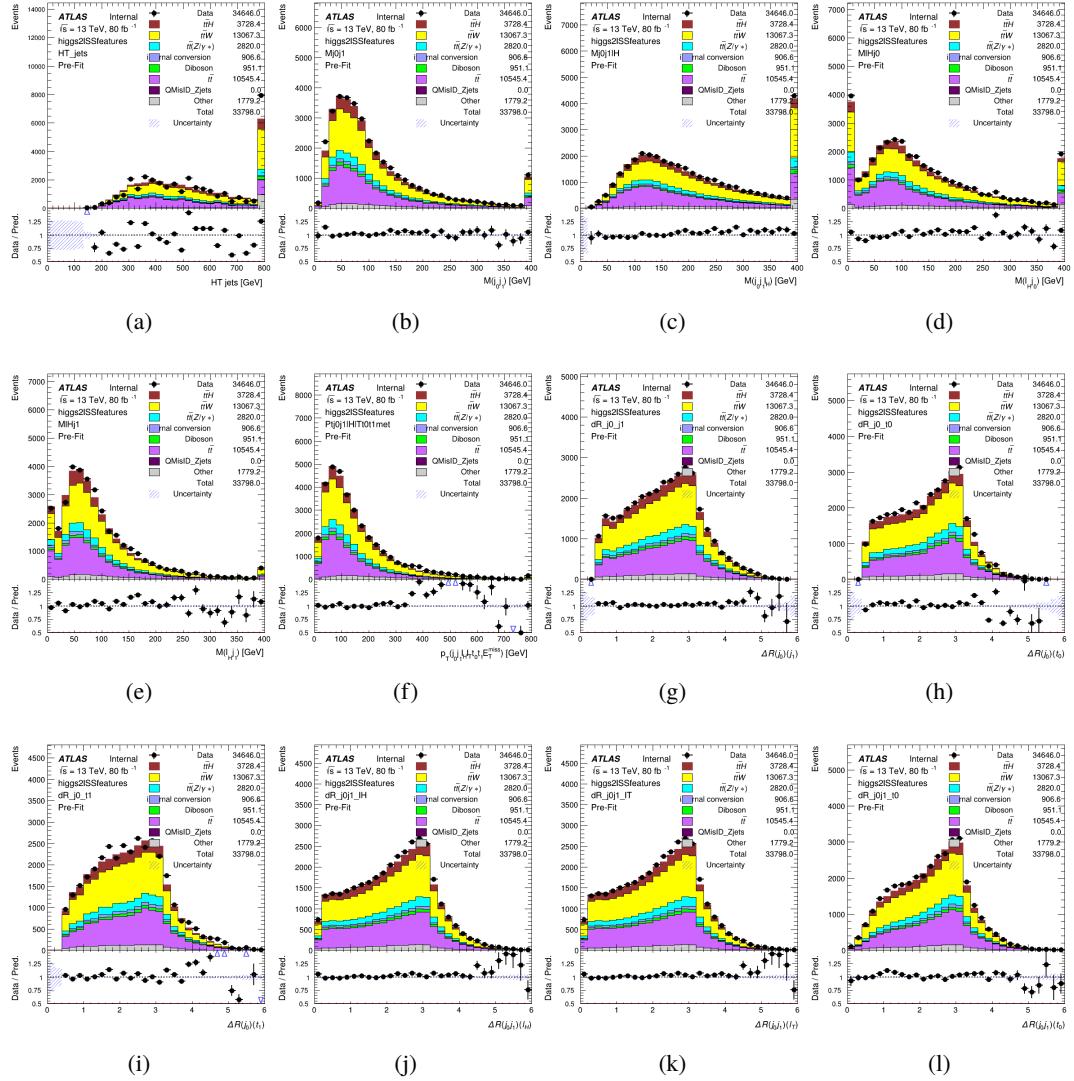


Figure C.8: Input features for higgs2ISST

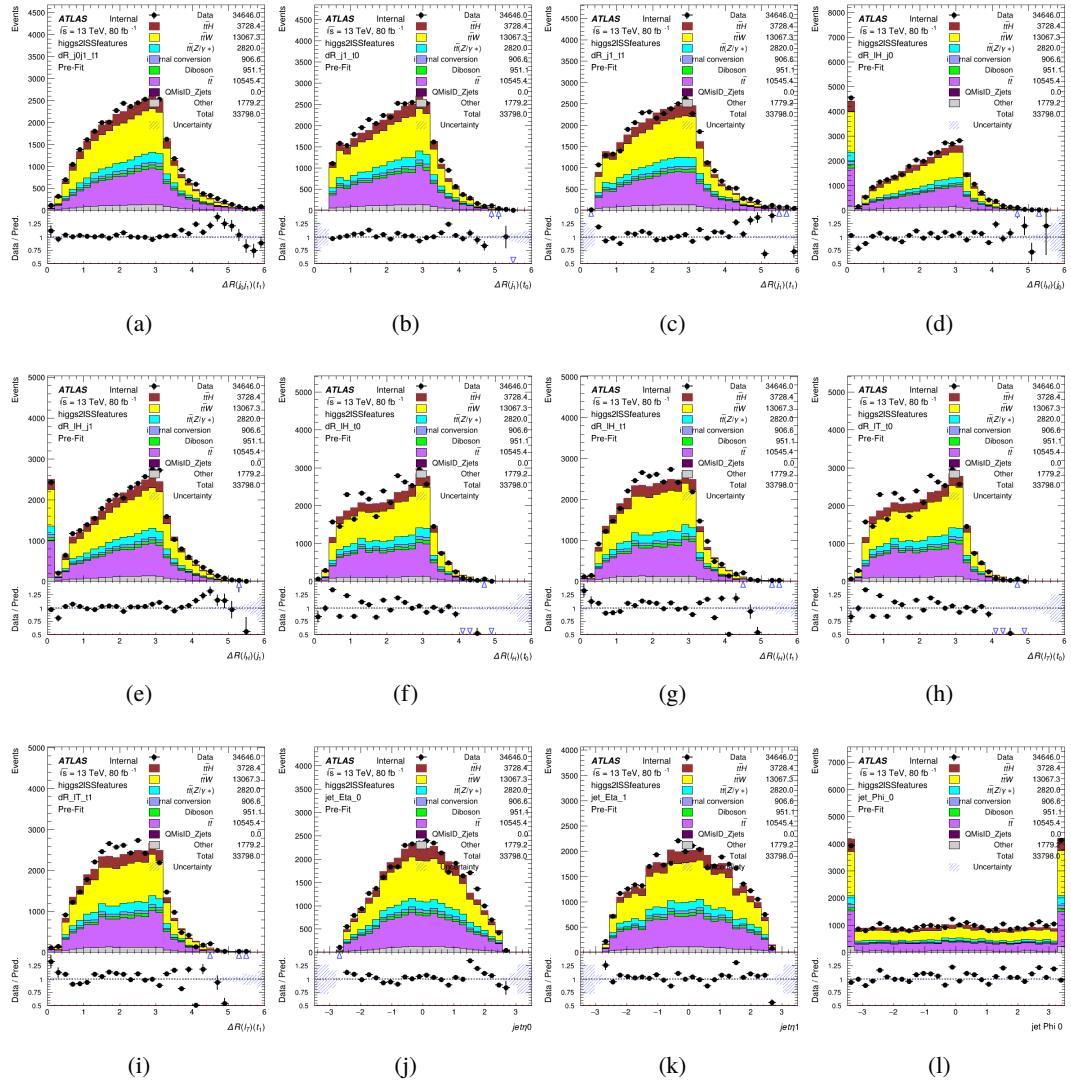


Figure C.9: Input features for higgs2ISSTest

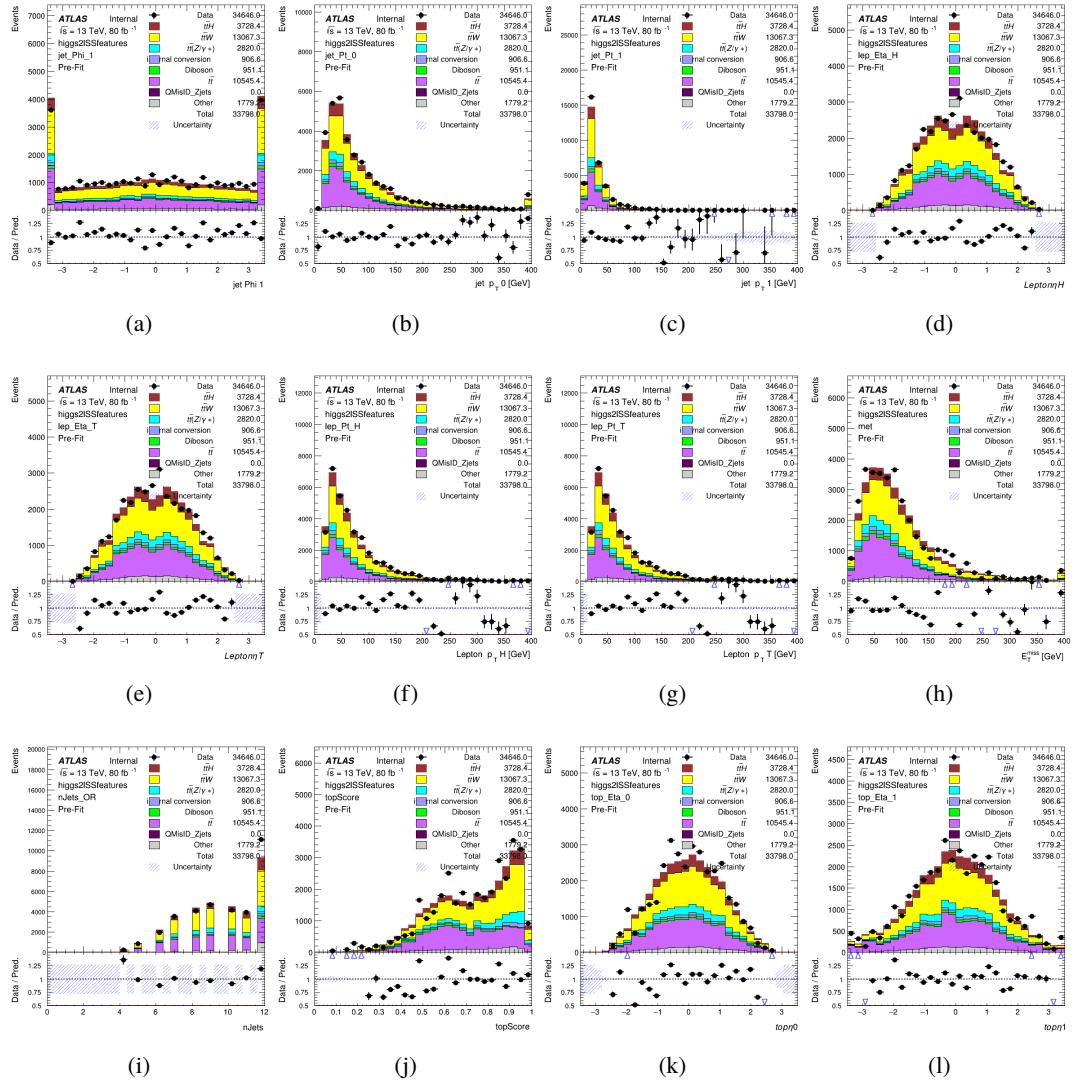


Figure C.10: Input features for higgs2lSS

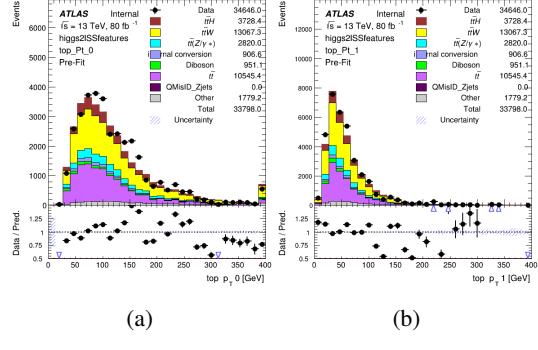


Figure C.11: Input features for higgs2lSS

<sub>1621</sub> **C.1.4 Higgs Reconstruction Features - 3lS**

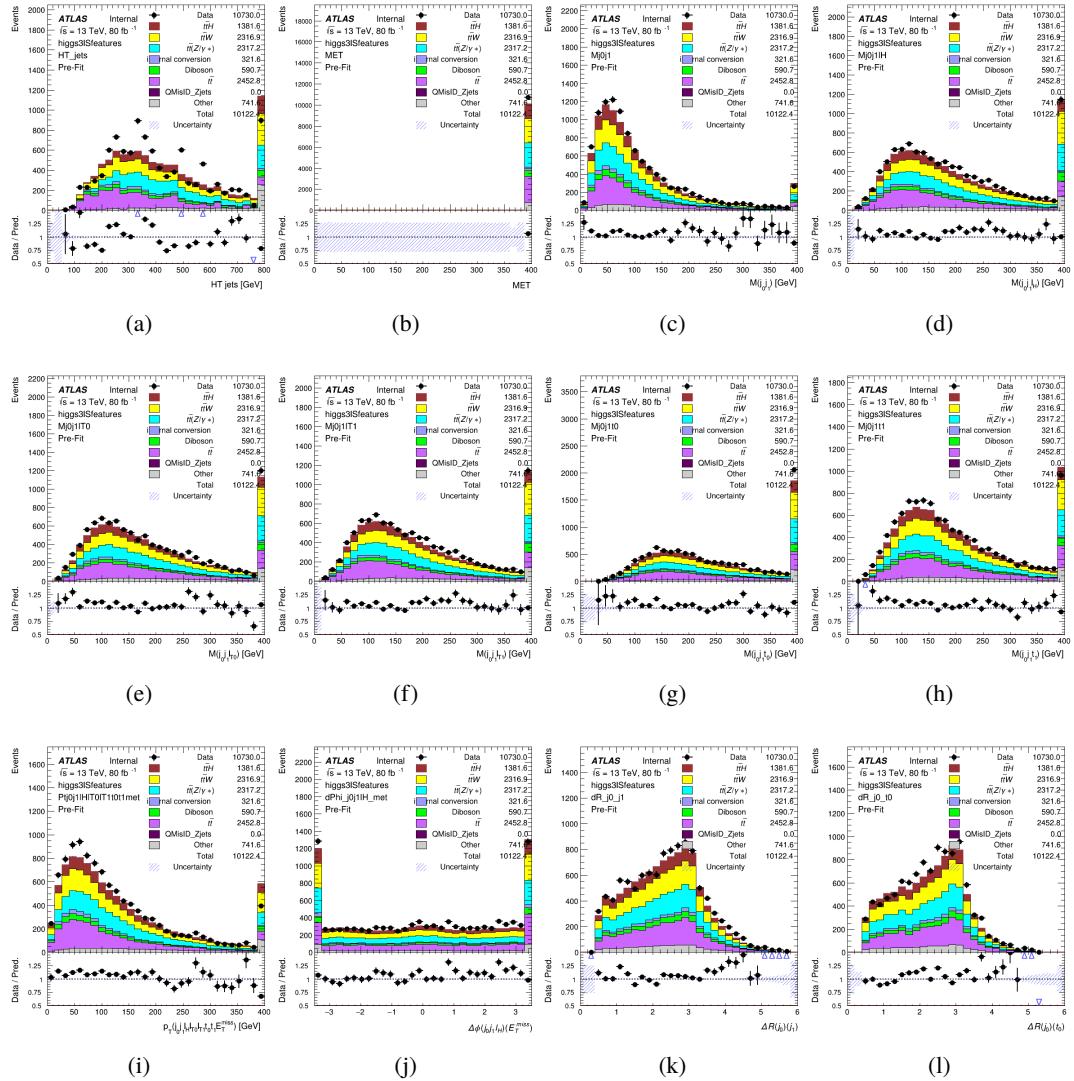


Figure C.12: Input features for higgs3IS

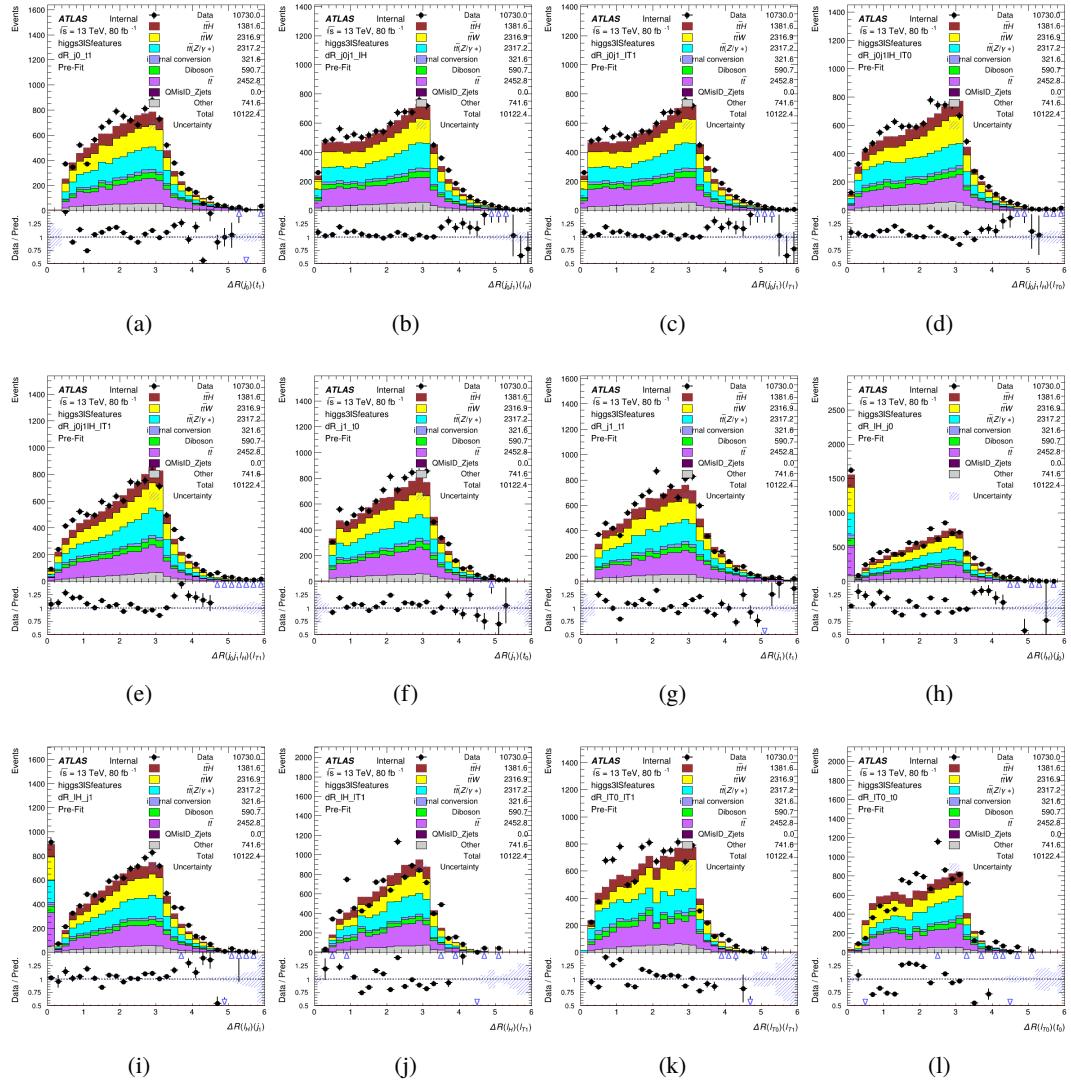


Figure C.13: Input features for higgs3lS

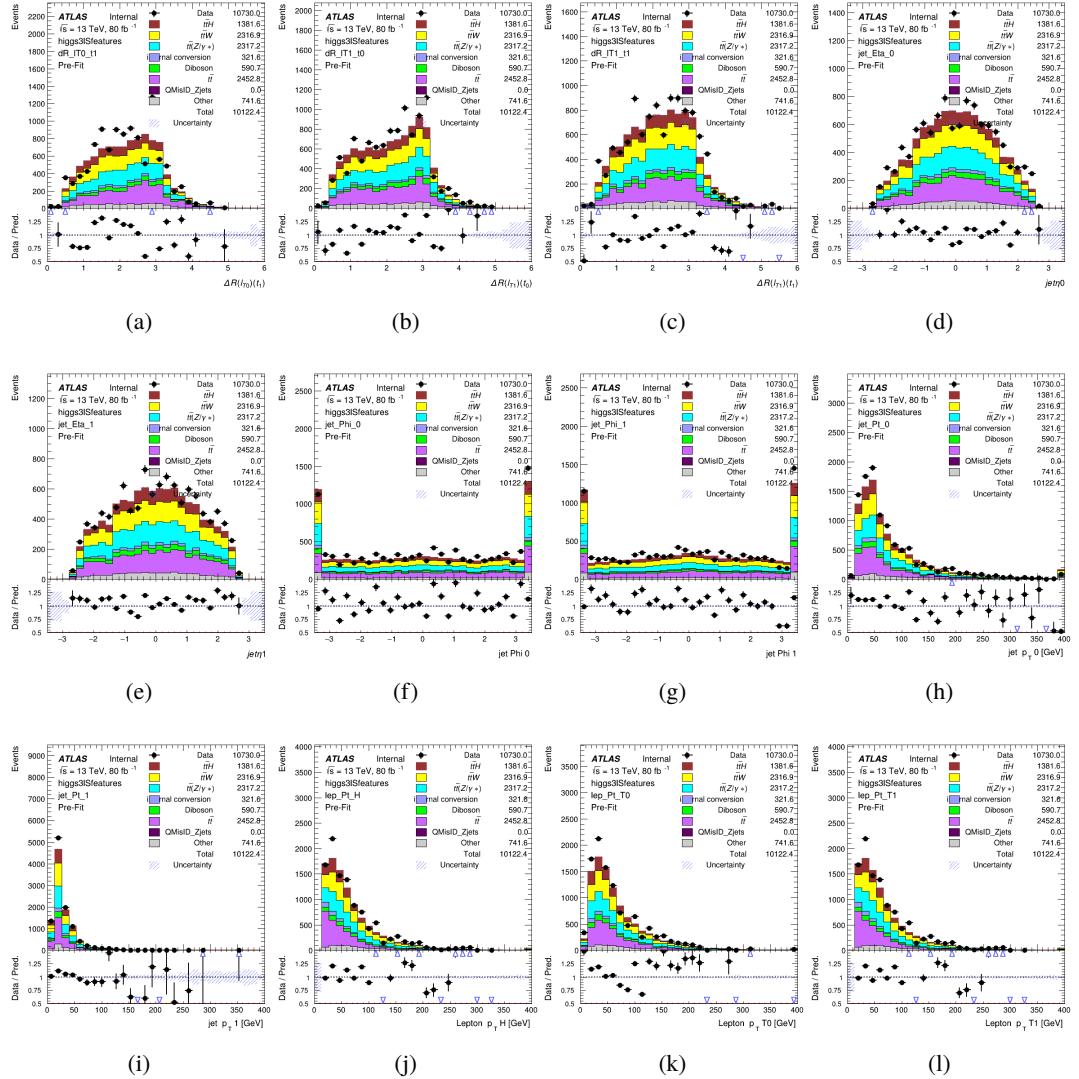


Figure C.14: Input features for higgs3IS

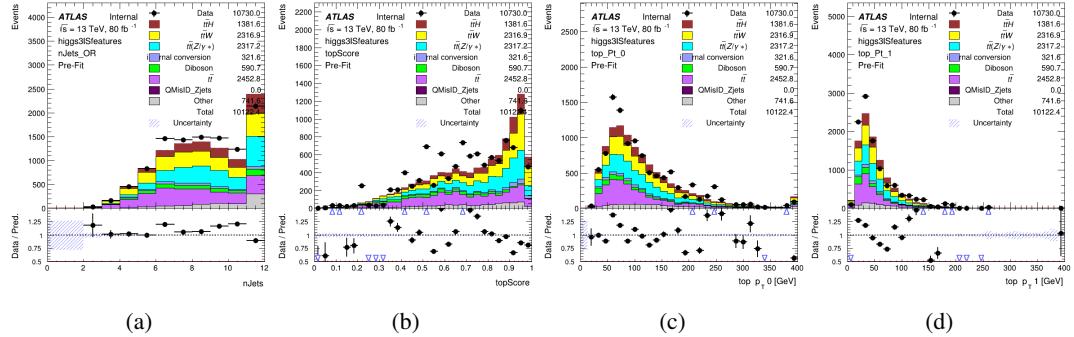


Figure C.15: Input features for higgs3IS

1622 **C.1.5 Higgs Reconstruction Features - 3lF**

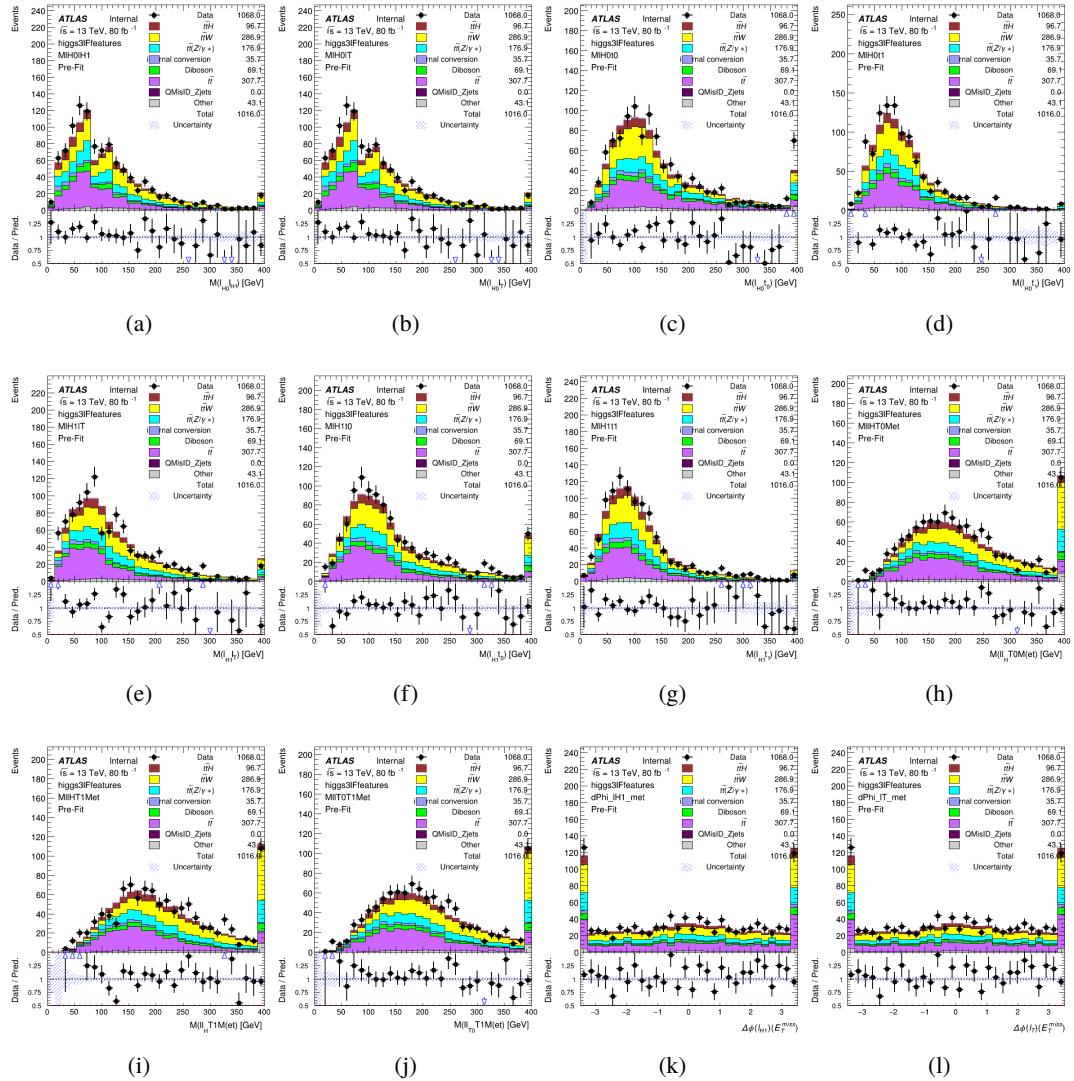


Figure C.16: Input features for higgs3lF

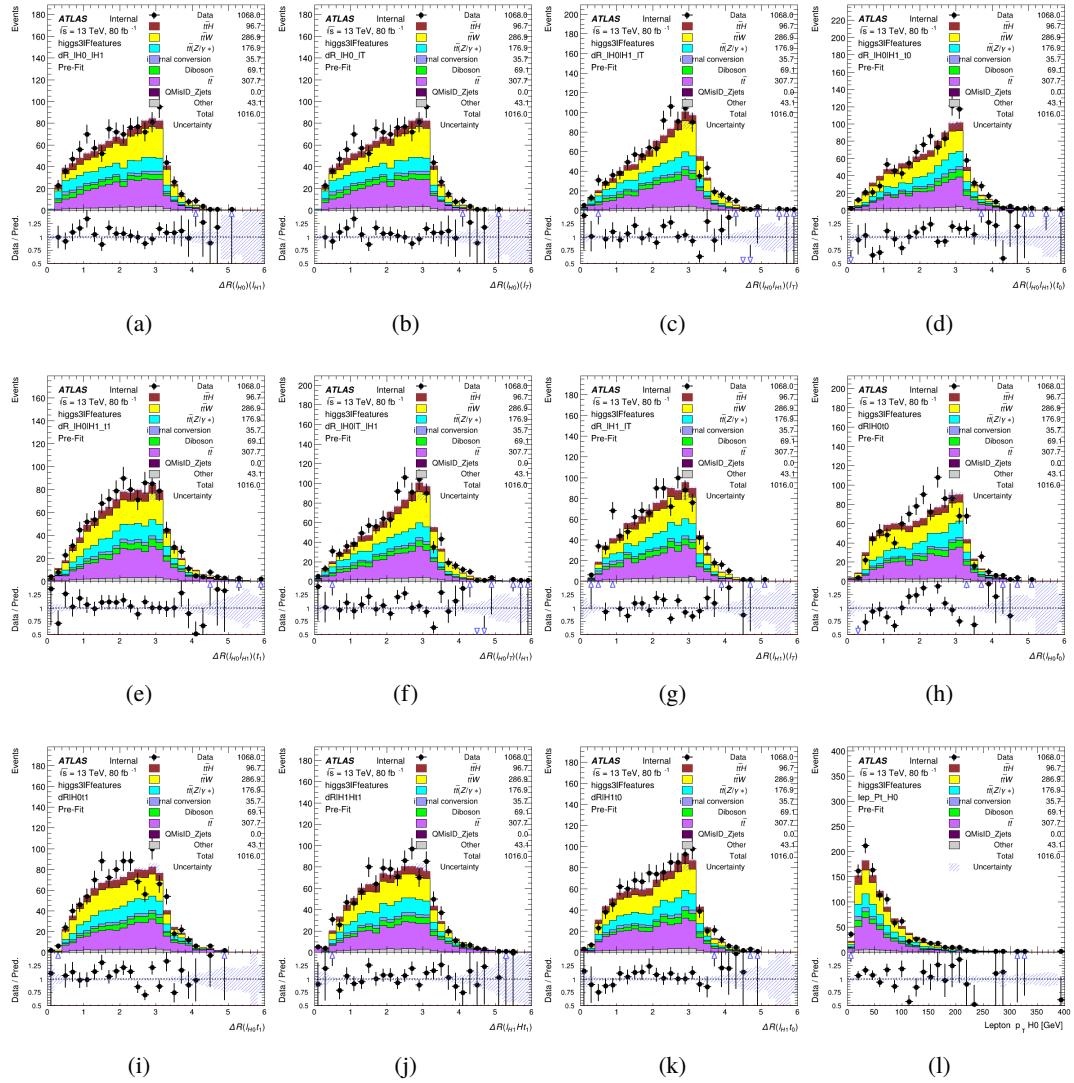


Figure C.17: Input features for higgs3lF

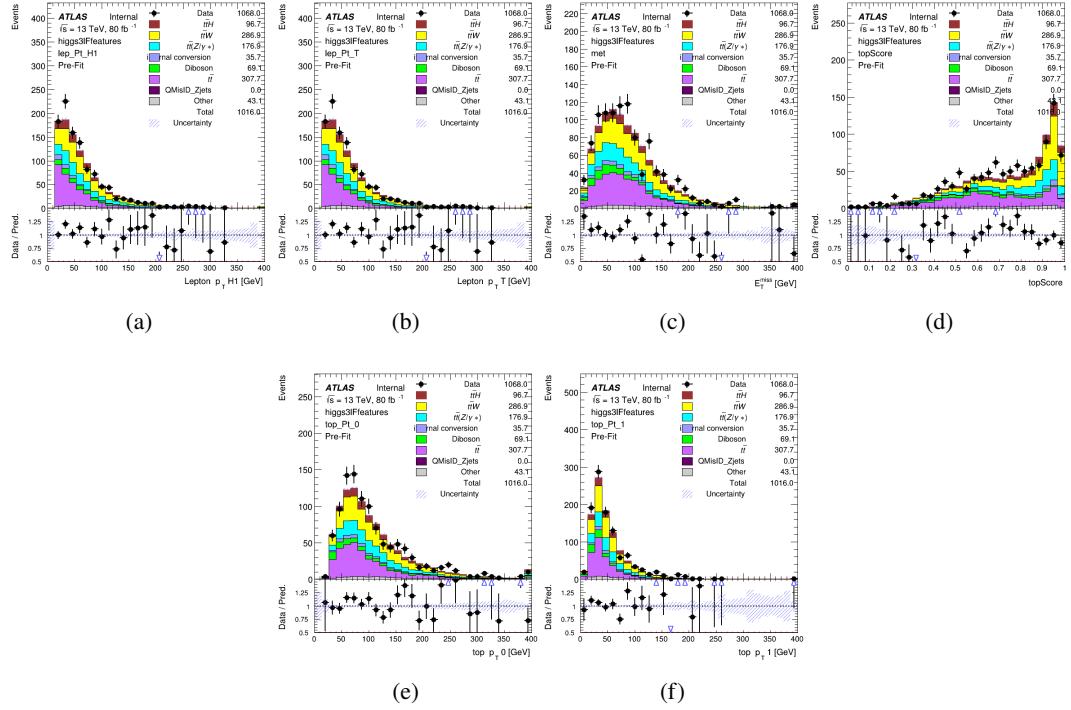


Figure C.18: Input features for higgs3lF

<sup>1623</sup> **C.2 Background Rejection MVA Details**

<sup>1624</sup> **C.2.1 Background Rejection MVA Features - 2ISS**

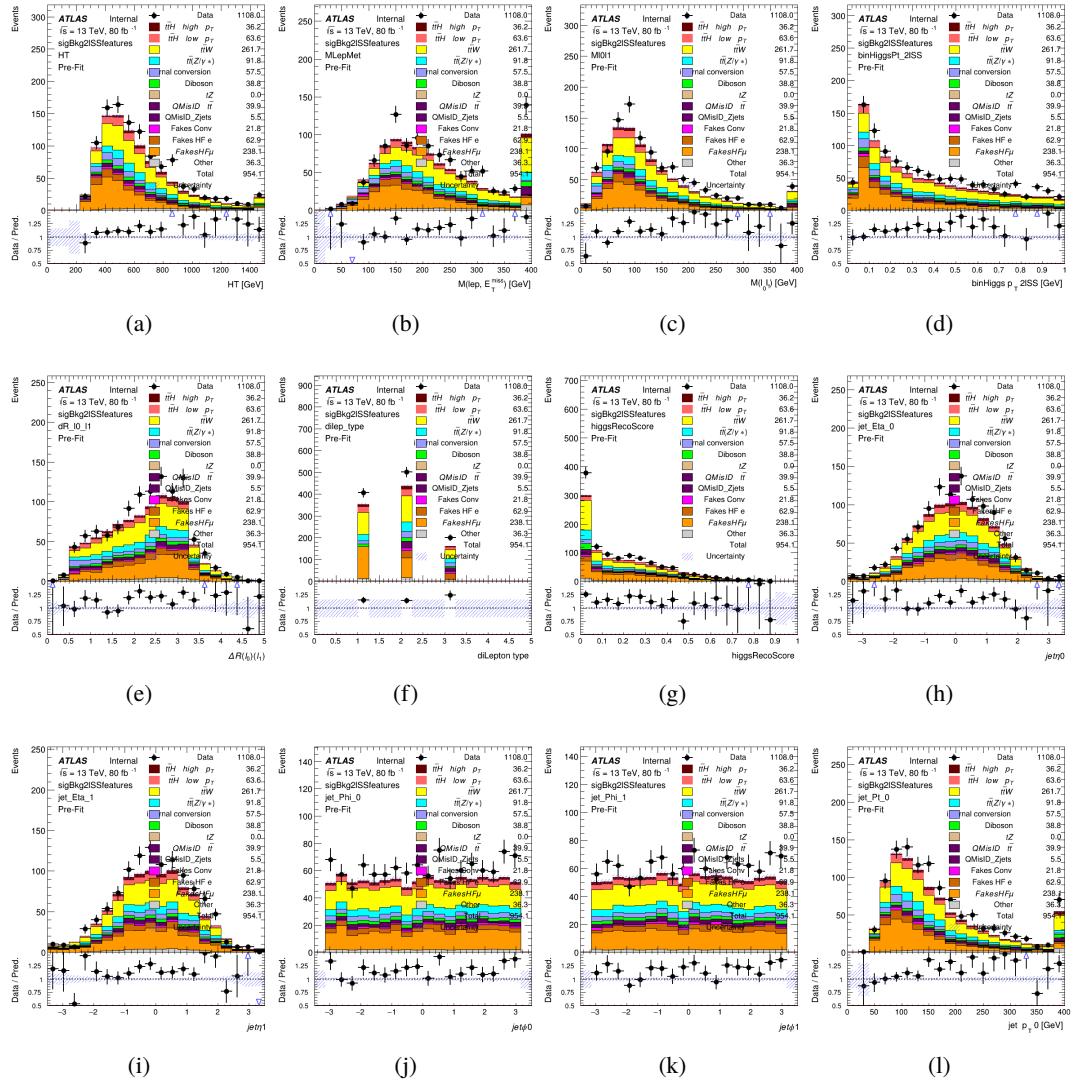


Figure C.19: Input features for sigBkg2ISS

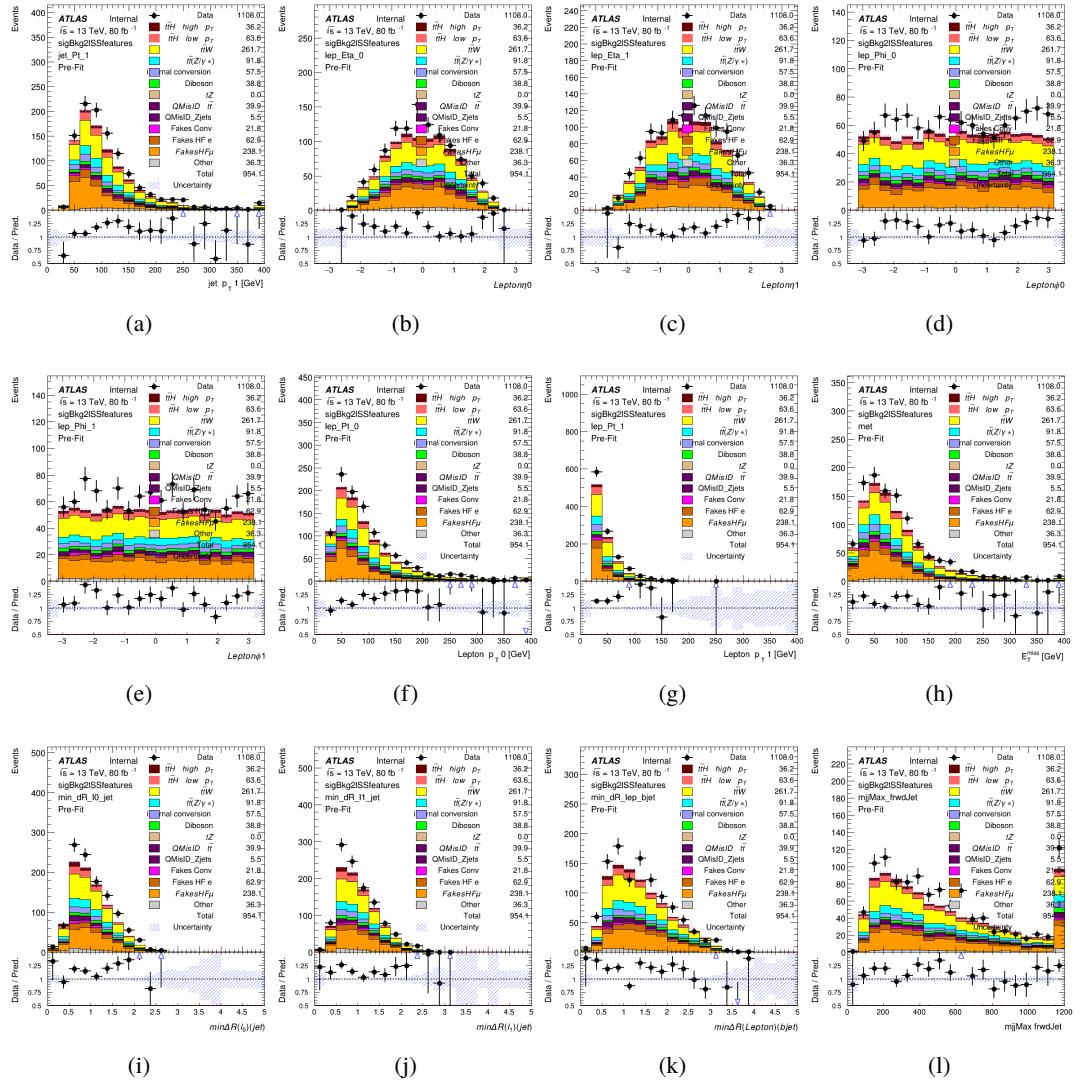


Figure C.20: Input features for sigBkg2lSS

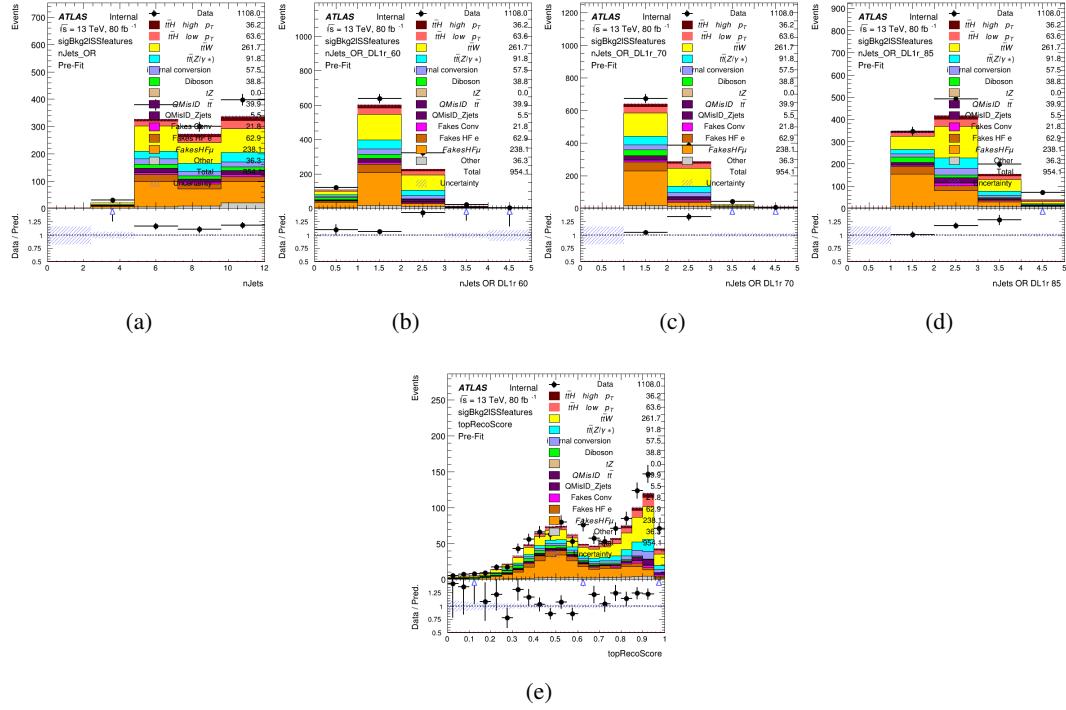


Figure C.21: Input features for sigBkg2lSS

1625 **C.2.2 Background Rejection MVA Features - 3l**

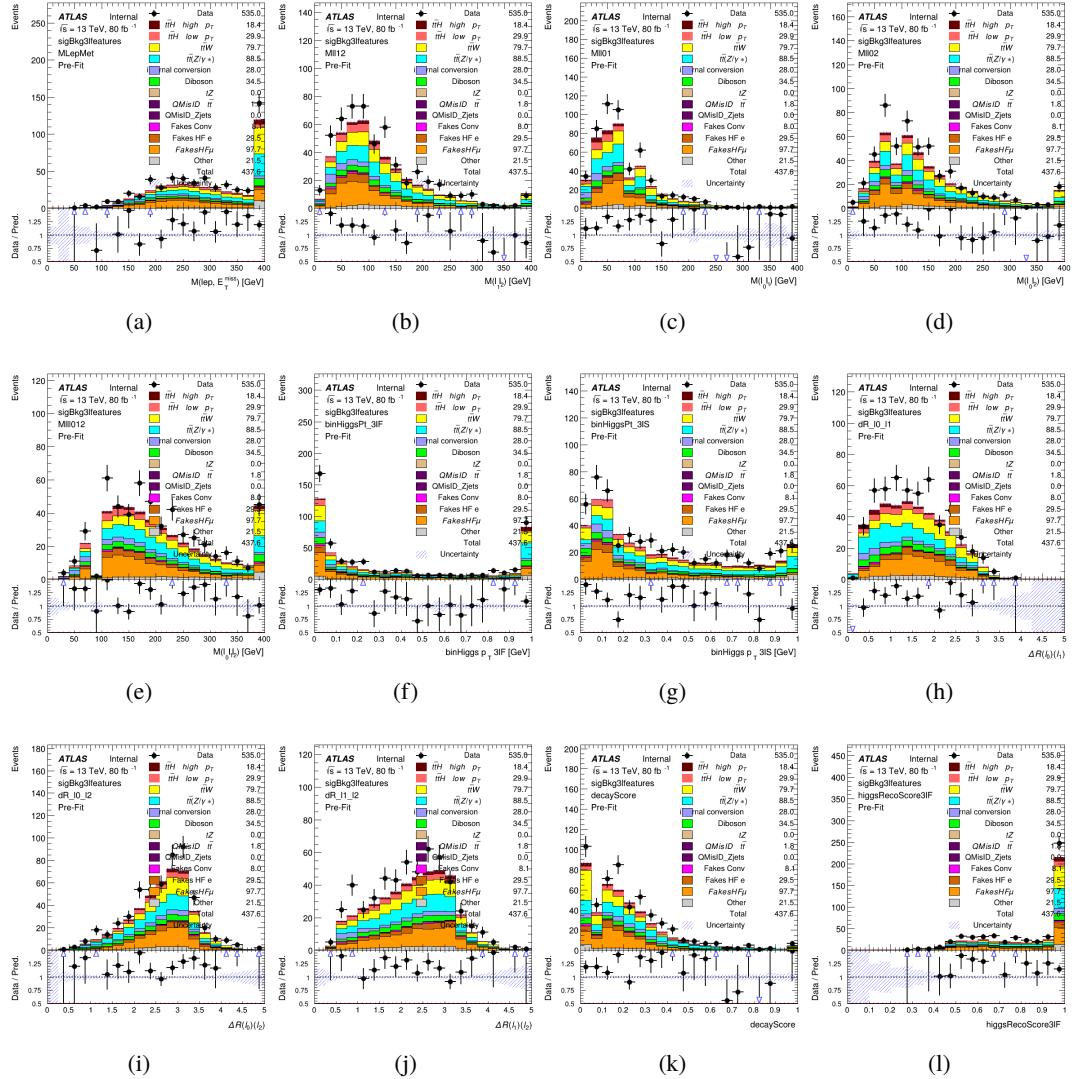


Figure C.22: Input features for sigBkg3l

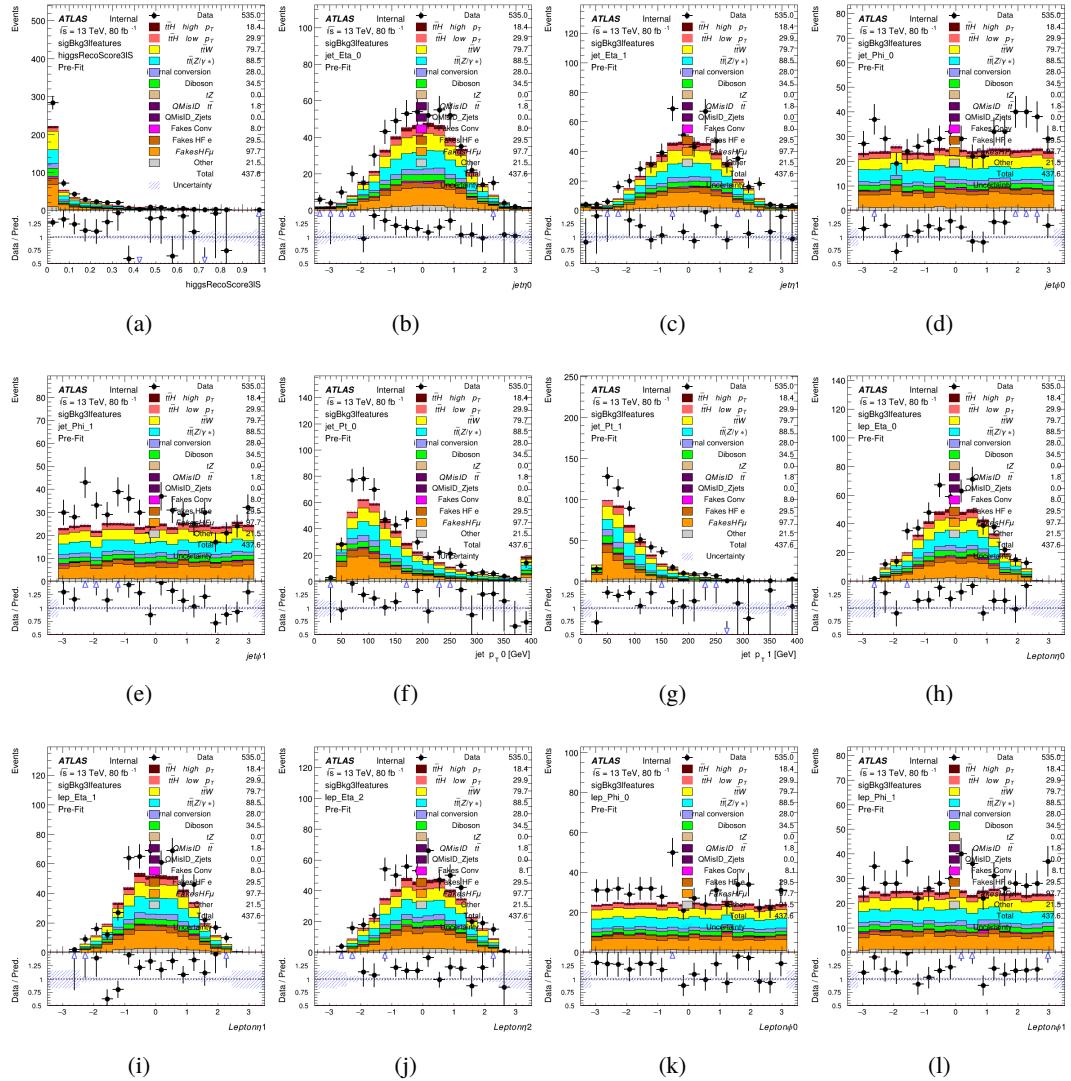
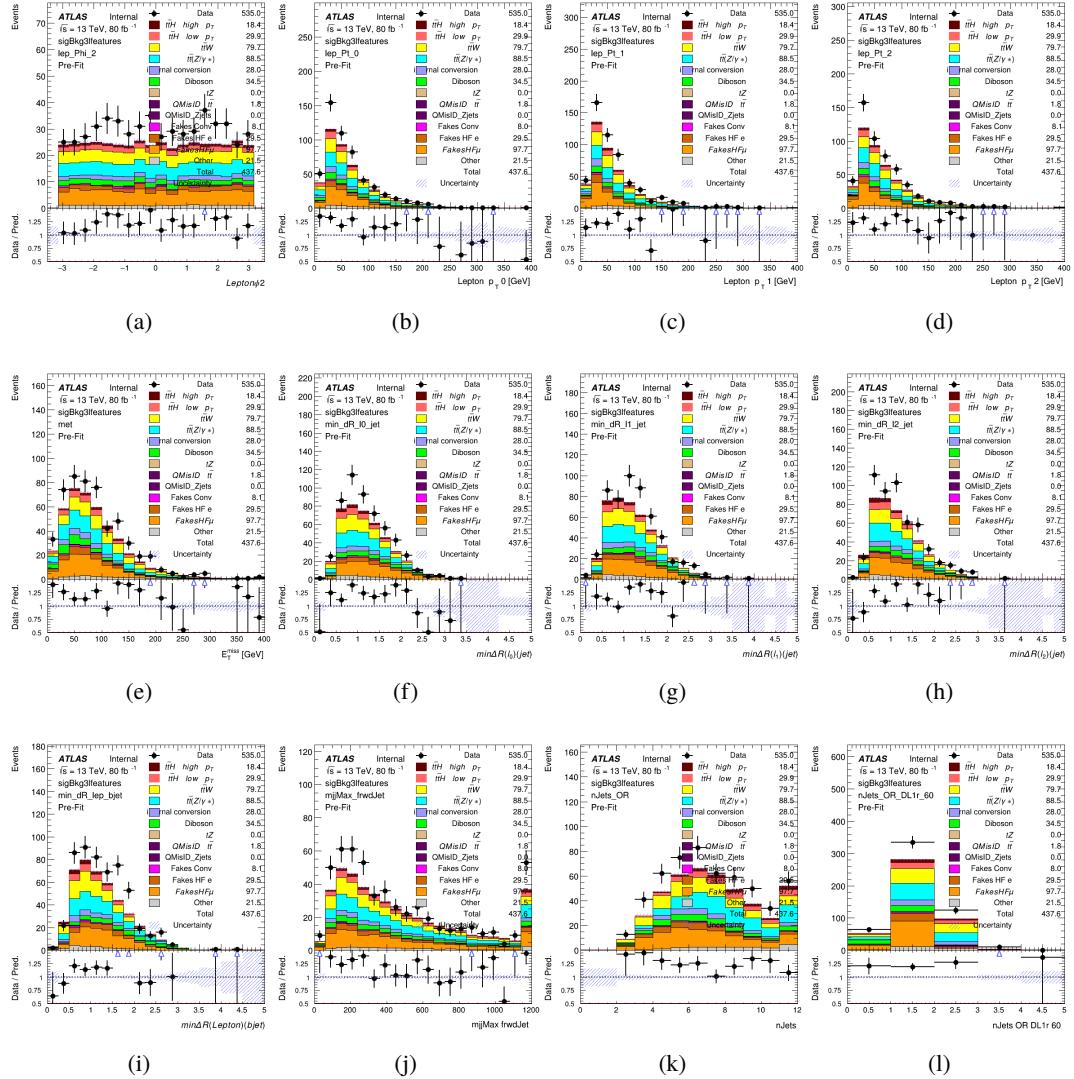


Figure C.23: Input features for sigBkg3l

Figure C.24: Input features for `sigBkg3l`

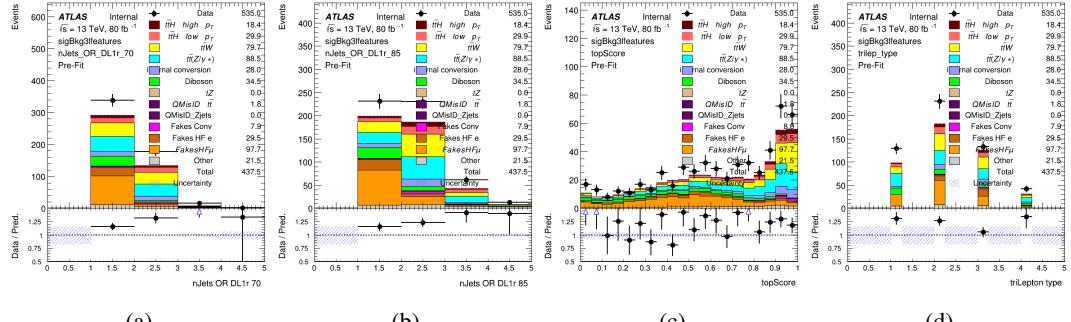


Figure C.25: Input features for sigBkg3l

1626 **C.3 Truth Level Studies**

1627 Attempts to identify the decay products of the Higgs are motivated by the ability to reconstruct  
 1628 the Higgs momentum based on their kinematics. In order to demonstrate that this is case, the  
 1629 kinematics of reconstructed objects that are truth matched to the Higgs decay are used as inputs  
 1630 to a neural network which is designed to predict of the momentum of the Higgs. This is done in  
 1631 the 2lSS channel, as it proves to be the most challenging for  $p_T$  reconstruction.

1632 Only leptons and jets which are truth matched to the Higgs are used as inputs for the  
 1633 model; events where the lepton and both jets are not reconstructed are not included. The model  
 1634 uses the same feature set and network architecture as the  $p_T$  prediction model used in the main  
 1635 analysis, as described in Section 14.4.1.

1636 The results of the model are summarized below:

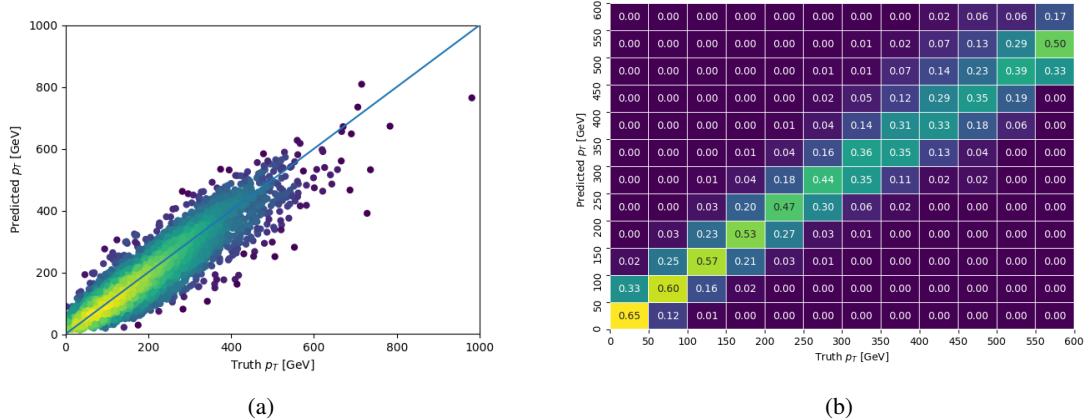


Figure C.26: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 2lSS  $t\bar{H}$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

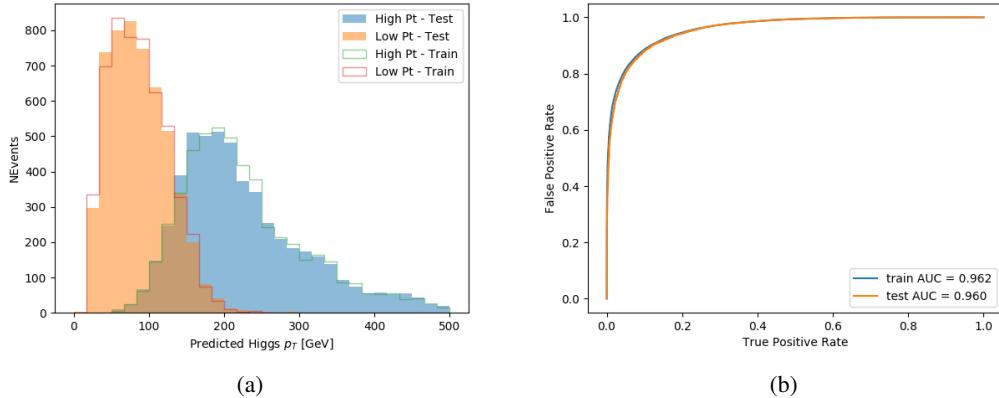


Figure C.27: (a) shows the reconstructed Higgs  $p_T$  for 2lSS events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

1637 Based on the performance of the model, as shown Figures C.26 and C.27, the Higgs  
 1638 momentum can be reconstructed with fairly high precision when its decay products are correctly  
 1639 identified.

#### 1640 C.4 Alternate b-jet Identification Algorithm

1641 The nominal analysis reconstructs the b-jets by considering different combinations of jets, and  
 1642 asking a neural network to determine whether each combination consists of b-jets from top quark  
 1643 decays. An alternate approach would be to give the neural network about all of the jets in an event  
 1644 at once, and train it to select which two are most likely to be the b-jets from top decay. It was  
 1645 hypothesized that this could perform better than considering each combination independently, as  
 1646 the neural network could consider the event as a whole. While this is not found to be the case,  
 1647 these studies are documented here as a point of interest and comparison.

1648 For these studies, the kinematics of the 10 highest  $p_T$  jets in each event are used for  
 1649 training. This includes the vast majority of truth b-jets. Specifically the  $p_T$ ,  $\eta$ ,  $\phi$ ,  $E$ , and DL1r  
 1650 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.  
 1651 The  $p_T$ ,  $\eta$ ,  $\phi$ , and  $E$  of the leptons and  $E_T^{\text{miss}}$  are included as well. Categorical cross entropy is  
 1652 used as the loss function.

Table 44: 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%

### 1653 C.5 Binary Classification of the Higgs $p_T$

1654 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer  
 1655 resolution. This means separating high and low  $p_T$  events is sufficient for this analysis. As  
 1656 such, rather than attempting to reconstruct the full Higgs  $p_T$  spectrum, a binary classification  
 1657 approach is explored.

1658 A model is built to determine whether  $t\bar{t}H$  events include a high  $p_T$  ( $>150$  GeV) or low  
 1659  $p_T$  ( $<150$  GeV) Higgs Boson. While this is now a classification model, it uses the same input  
 1660 features described in section 14.4. Binary crossentropy is used as the loss function.

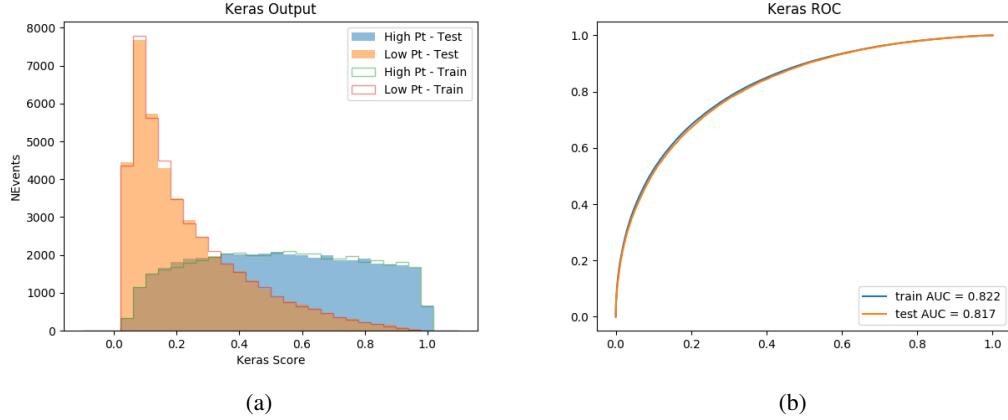


Figure C.28: Output distribution of the NN score for the binary high/low  $p_T$  separation model in the 2lSS channel.

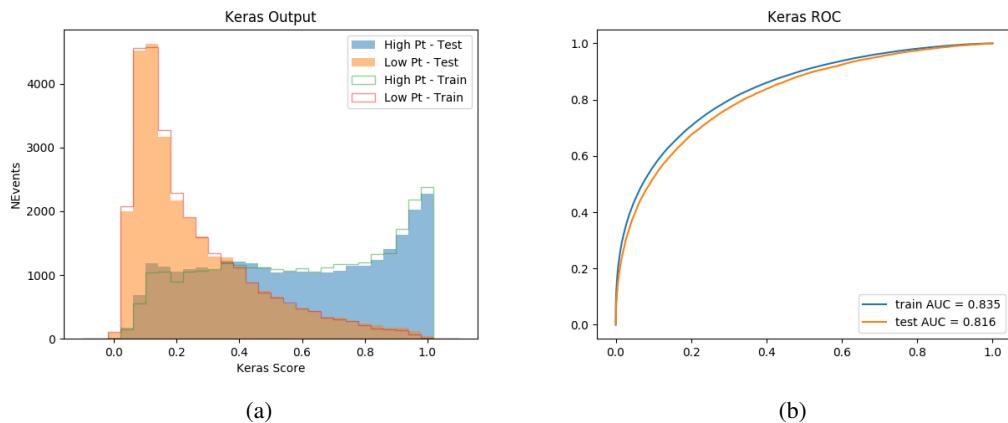


Figure C.29: Output distribution of the NN score for the binary high/low  $p_T$  separation model in the 3lS channel.

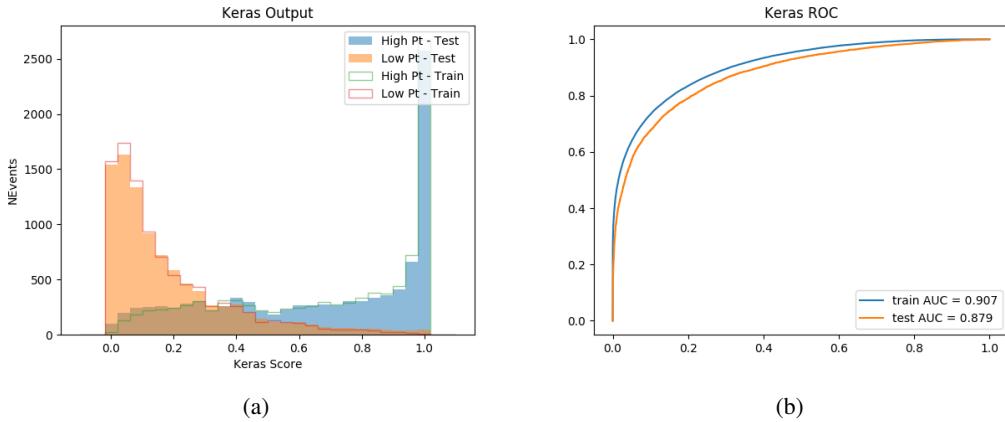


Figure C.30: Output distribution of the NN score for the binary high/low  $p_T$  separation model in the 3LS channel.

## C.6 Impact of Alternative Jet Selection

1662 A relatively low  $p_T$  threshold of 15 GeV is used to determine jet candidates, as the jets originating  
 1663 from the Higgs decay are found to fall between 15 and 25 GeV a large fraction of the time. The  
 1664 impact of different jet  $p_T$  cuts on our ability to reconstruct the Higgs  $p_T$  is explored here.

The models are retrained in the 2LSS channel with the same parameters as those used in the nominal analysis, but the jet  $p_T$  threshold is altered. The performance of the Higgs  $p_T$  prediction models for jet  $p_T$  cuts of 20 and 25 GeV are shown below.

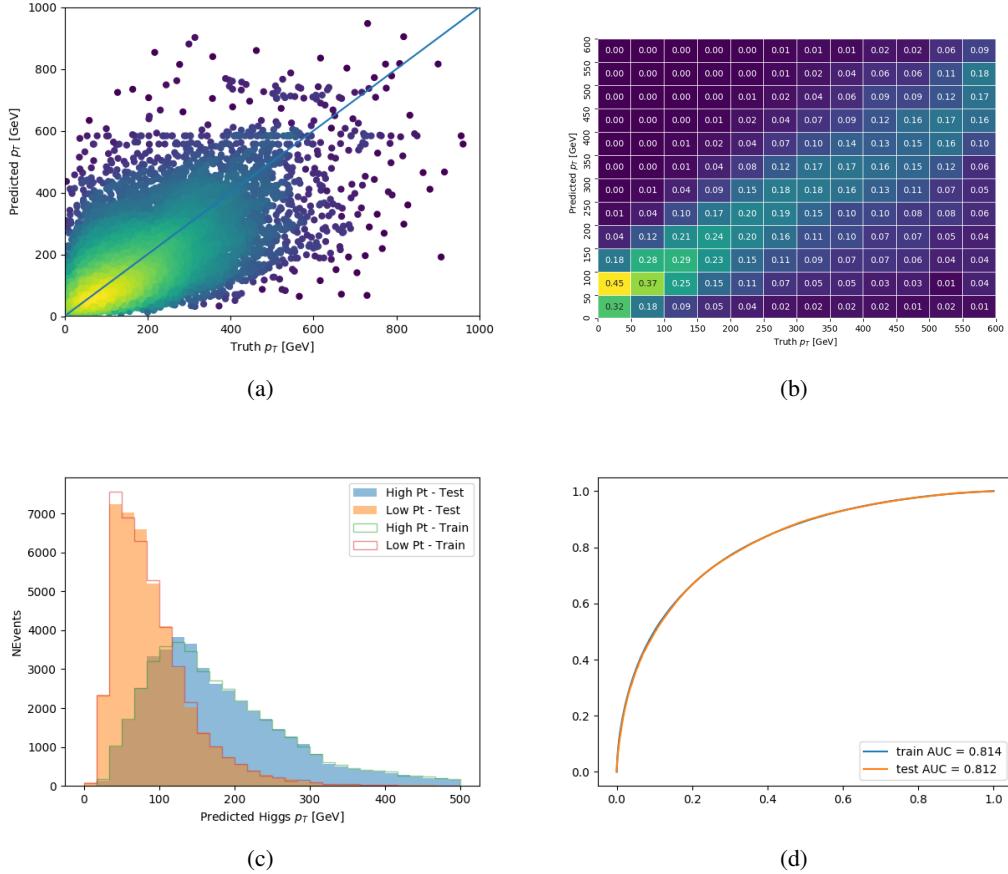
Jet  $p_T > 20$  GeV

Figure C.31: Output of the model designed to predict the Higgs momentum in the 2LSS channel, with the jet  $p_T$  cutoff used is raised to 20 GeV.

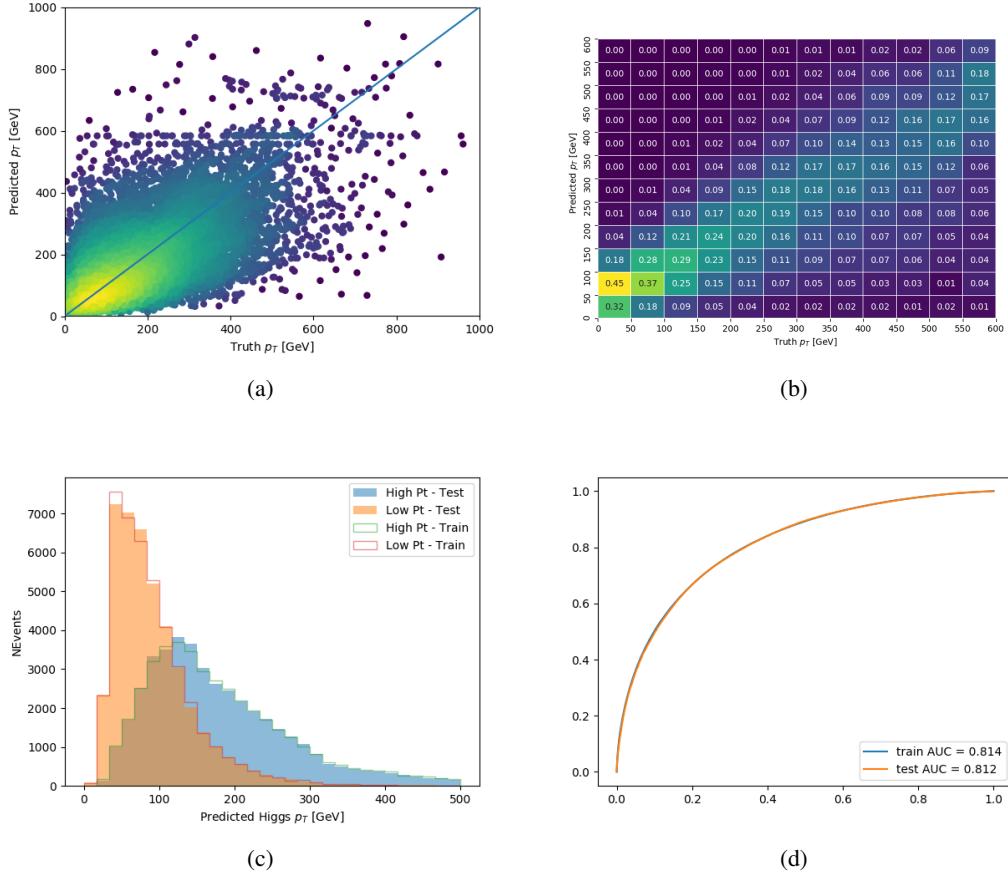
**Jet  $p_T > 25 \text{ GeV}$** 

Figure C.32: Output of the model designed to predict the Higgs momentum in the 2LSS channel, with the jet  $p_T$  cutoff used is raised to 25 GeV.

1668 **D**