



**ATLAS Note**

GROUP-2017-XX



7th May 2021

Draft version 0.1

1

2

3

4

5

6

**A Deep Learning Approach to Differential  
Measurements of Higgs - Top Interactions in  
Multilepton Final States using the ATLAS  
Detector at the LHC**

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

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

© 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>	7
28	<b>1 Introduction</b>	7
29	<b>II Theoretical Motivation</b>	10
30	<b>2 The Standard Model and the Higgs Boson</b>	10
31	2.1 The Forces and Particles of the Standard Model	10
32	2.2 The Higgs Mechanism	13
33	2.2.1 The Higgs Field	13
34	2.2.2 Electroweak Symmetry Breaking	15
35	2.3 $t\bar{t}H$ Production	18
36	2.4 WZ + Heavy Flavor Production	21
37	2.5 Extensions to the Standard Model	22
38	<b>III The LHC and the ATLAS Detector</b>	27
39	<b>3 The LHC</b>	27
40	<b>4 The ATLAS Detector</b>	30
41	4.1 Inner Detector	31
42	4.2 Calorimeters	32
43	4.3 Muon Spectrometer	33
44	4.4 Trigger System	34
45	<b>IV Measurement of WZ + Heavy Flavor</b>	36
46	<b>5 Introduction</b>	36
47	<b>6 Data and Monte Carlo Samples</b>	37
48	6.1 Data Samples	37
49	6.2 Monte Carlo Samples	38
50	<b>7 Object Reconstruction</b>	39
51	7.1 Light leptons	39
52	7.2 Jets	40
53	7.3 B-tagged Jets	41
54	7.4 Missing transverse energy	42
55	7.5 Overlap removal	43

---

56	<b>8 Event Selection and Signal Region Definitions</b>	<b>43</b>
57	8.1 Event Preselection	44
58	8.2 Fit Regions	46
59	8.3 Non-Prompt Lepton Estimation	48
60	8.3.1 $t\bar{t}$ Validation	48
61	8.3.2 Z+jets Validation	49
62	<b>9 tZ Interference Studies and Separation Multivariate Analysis</b>	<b>50</b>
63	<b>10 Data and Monte Carlo Samples</b>	<b>54</b>
64	10.1 Data Samples	54
65	10.2 Monte Carlo Samples	54
66	<b>11 Object Reconstruction</b>	<b>55</b>
67	11.1 Light leptons	56
68	11.2 Jets	57
69	11.3 B-tagged Jets	57
70	11.4 Missing transverse energy	59
71	11.5 Overlap removal	59
72	<b>12 Event Selection and Signal Region Definitions</b>	<b>60</b>
73	12.1 Event Preselection	60
74	12.2 Fit Regions	63
75	12.3 Non-Prompt Lepton Estimation	64
76	12.3.1 $t\bar{t}$ Validation	65
77	12.3.2 Z+jets Validation	66
78	<b>13 tZ Interference Studies and Separation Multivariate Analysis</b>	<b>66</b>
79	<b>14 Systematic Uncertainties</b>	<b>70</b>
80	<b>15 Results</b>	<b>74</b>
81	15.1 1-jet Fit Results	75
82	15.2 2-jet Fit Results	84
83	<b>V Differential Studies of <math>t\bar{t}H</math> Multilepton</b>	<b>93</b>
84	<b>16 Data and Monte Carlo Samples</b>	<b>93</b>
85	16.1 Data Samples	94
86	16.2 Monte Carlo Samples	94
87	<b>17 Object Reconstruction</b>	<b>98</b>
88	17.1 Trigger Requirements	98
89	17.2 Light Leptons	99

90	17.3 Jets	100
91	17.4 B-tagged Jets	100
92	17.5 Missing Transverse Energy	101
93	17.6 Overlap removal	101
94	<b>18 Higgs Momentum Reconstruction</b>	<b>102</b>
95	18.1 Physics Object Truth Matching	104
96	18.2 Truth Level Studies	105
97	18.3 b-jet Identification	105
98	18.3.1 2lSS Channel	106
99	18.3.2 3l Channel	110
100	18.4 Higgs Reconstruction	113
101	18.4.1 2lSS Channel	114
102	18.4.2 3l Semi-leptonic Channel	118
103	18.4.3 3l Fully-leptonic Channel	122
104	18.5 $p_T$ Prediction	126
105	18.5.1 2lSS Channel	127
106	18.5.2 3l Semi-leptonic Channel	130
107	18.5.3 3l Fully-leptonic Channel	132
108	18.6 3l Decay Mode	134
109	<b>19 Signal Region Definitions</b>	<b>136</b>
110	19.1 Pre-MVA Event Selection	136
111	19.2 Event MVA	140
112	19.3 Signal Region Definitions	147
113	<b>20 Systematic Uncertainties</b>	<b>148</b>
114	<b>21 Results</b>	<b>155</b>
115	21.1 Results - $79.8 \text{ fb}^{-1}$	156
116	21.2 Projected Results - $139 \text{ fb}^{-1}$	161
117	<b>VI Conclusion</b>	<b>166</b>
118	<b>VII Appendices</b>	<b>174</b>
119	<b>Appendices</b>	<b>174</b>
120	<b>A Non-prompt lepton MVA</b>	<b>174</b>
121	<b>B Supplementary WZ + Heavy Flavor Studies</b>	<b>179</b>
122	B.1 Non-prompt CR Modelling	179

---

123	B.2 tZ Interference Studies	184
124	B.3 Alternate tZ Inclusive Fit	188
125	B.3.1 tZ Inclusive Fit	189
126	B.3.2 Floating tZ	190
127	<b>C Supplementary ttH Differential Analysis Studies</b>	<b>191</b>
128	C.1 Higgs Reconstruction Model Details	191
129	C.1.1 b-jet Identification Features - 2ISS	191
130	C.1.2 b-jet Identification Features - 3l	195
131	C.1.3 Higgs Reconstruction Features - 2ISS	200
132	C.1.4 Higgs Reconstruction Features - 3IS	205
133	C.1.5 Higgs Reconstruction Features - 3IF	210
134	C.2 Background Rejection MVA Details	214
135	C.2.1 Background Rejection MVA Features - 2ISS	214
136	C.2.2 Background Rejection MVA Features - 3l	218
137	C.3 Truth Level Studies	223
138	C.4 Alternate b-jet Identification Algorithm	224
139	C.5 Binary Classification of the Higgs p <sub>T</sub>	225
140	C.6 Impact of Alternative Jet Selection	227

---

# 141 Part I

## 142 Introduction

### 143 1 Introduction

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

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

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

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

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

171 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-  
172 2018 provides the opportunity to make this measurement for the first time. The unprecedented  
173 energy achieved by the LHC during this period greatly increase the rate at which  $t\bar{t}H$  events are  
174 produced, and the large amount of data collected provides the necessary statistics for a differential  
175 measurement to be performed.

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

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

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

188 This dissertation begins with a brief explanation of the SM, its limitations, and the theor-  
189 etical motivation behind this work in Part II. This is followed by a description of the LHC and  
190 the ATLAS detector in Part III. Part IV details a measurement of WZ + heavy flavor. Studies  
191 of differential measurements of  $t\bar{t}H$  are then described in Part V, and preliminary results are  
192 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

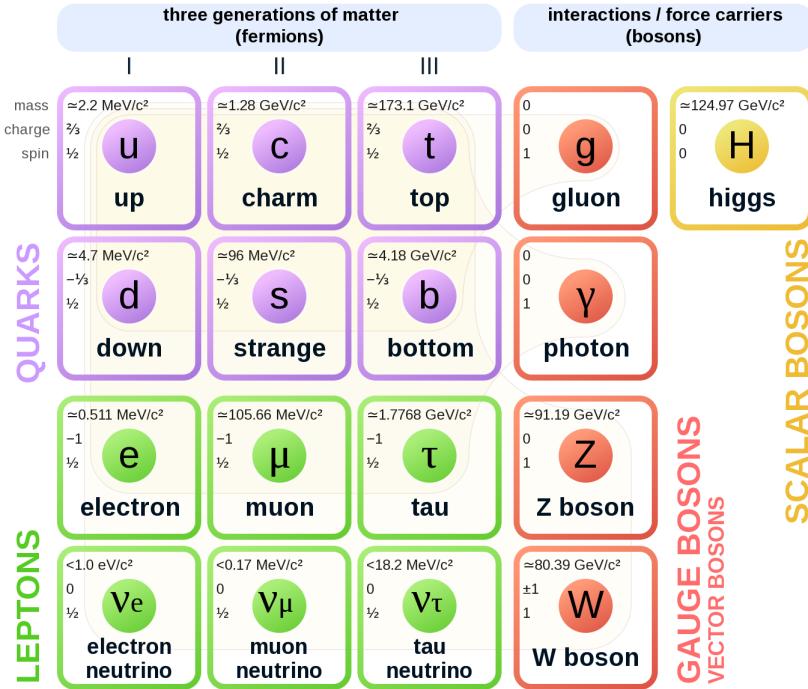


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. [2]

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

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

214 Each of these generations form left-handed doublets invariant under SU(2) transfor-

215 mations. For the leptons these doublets are:

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

216 And for the quarks:

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

217 For both leptons and quarks, the heavier generations can decay into the lighter generation

218 of particles, while the first generation does not decay. Hence, ordinary matter generally consists

219 of this first generation of fermions - electrons, up quarks, and down quarks. Each of these

220 fermions has a corresponding anti-particle, which has an equal mass as its partner but opposite

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

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

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

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

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

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

<sup>227</sup> mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon  
<sup>228</sup> mediates the electric force, and the gluon mediates the strong force.

<sup>229</sup> **2.2 The Higgs Mechanism**

<sup>230</sup> A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to  
<sup>231</sup> the Lagrangian giving mass to the the gauge bosons would violate this underlying symmetry of  
<sup>232</sup> the theory. This presents a clear problem with the theory: The experimental observation that the  
<sup>233</sup> W and Z bosons have mass seems to contradict the basic structure of the SM.

<sup>234</sup> Rather than abandoning gauge invariance, an alternative way for particles to acquire mass  
<sup>235</sup> beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout  
<sup>236</sup> in 1964 [3]. This procedure for introducing masses for the gauge bosons while preserving local  
<sup>237</sup> gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory  
<sup>238</sup> by Weinberg in 1967 [4].

<sup>239</sup> **2.2.1 The Higgs Field**

<sup>240</sup> The Higgs mechanism introduces a complex scalar SU(2) doublet,  $\Phi$ , with the form:

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

<sup>241</sup> This field introduces a scalar potential to the Lagrangian of the form:

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

<sup>242</sup> Where  $\mu$  and  $\lambda$  are free parameters of the new field. This represents the most general  
<sup>243</sup> potential allowed while preserving  $SU(2)_L$  invariance and renormalizability. In the case that  
<sup>244</sup>  $\mu^2 < 0$ , this potential takes the form shown in Figure 2.2.

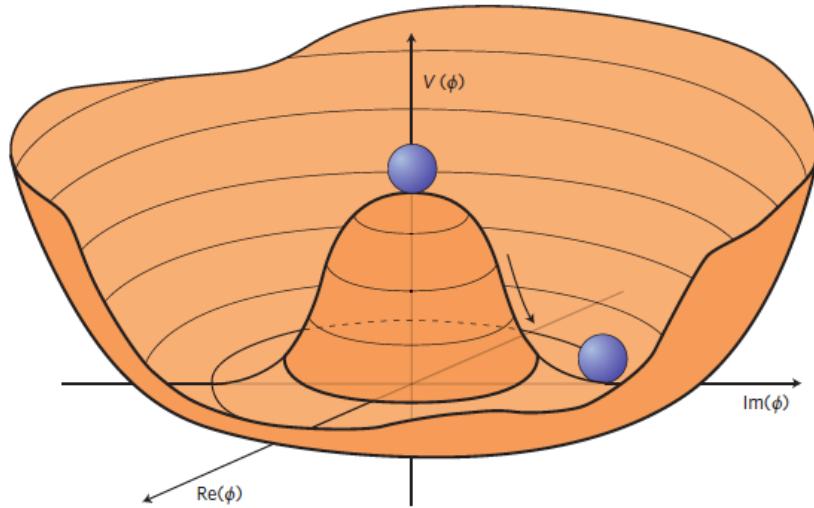


Figure 2.2: The value of the Higgs potential,  $V(\Phi)$  as a function of  $\Phi$ , for the case that  $\mu^2 < 0$  [5].

<sup>245</sup> The significant feature of this potential is that its minimum does not occur for a value of  
<sup>246</sup>  $\Phi = 0$ . Instead, it is minimized when  $|\Phi^\dagger \Phi| = -\mu^2/\lambda$ . This means that in its ground state, the  
<sup>247</sup> Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while  
<sup>248</sup> the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

<sup>249</sup> the minimum is determined only by  $\Phi^\dagger \Phi$ , there is some ambiguity in the particular definition of

<sup>250</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>251</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>252</sup> with  $v$  being the value of the VEV, and  $H$  being the real value of the scalar field.

### <sup>253</sup> 2.2.2 Electroweak Symmetry Breaking

<sup>254</sup> The Electroweak (EWK) interaction is described in the SM by a  $SU(2)_L \otimes U(1)_Y$  gauge theory.

<sup>255</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>256</sup>  $B_\mu$ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar

<sup>257</sup> field  $\Phi$  in the Lagrangian:

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

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

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

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

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

265 And the masses of these fields are given by:

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

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

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

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

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

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

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

<sup>282</sup> free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions

<sup>283</sup> must be measured experimentally.

<sup>284</sup> **2.3  $t\bar{t}H$  Production**

<sup>285</sup> The strength of a particles interaction with the Higgs, given by its Yukawa coupling, is proportionate to its mass. The top quark - as the heaviest known particle - has the strongest interaction, making this interaction particularly interesting to study. While several processes involve interactions between the Higgs and the top, some Higgs production modes include the top interaction only as a part of a loop diagram, such as the gluon-gluon fusion diagram shown in Figure 2.3.

<sup>290</sup> This process therefore only allows for an indirect probe of the Higgs-top Yukawa coupling, as <sup>291</sup> the flavor of the quark in this diagram is not unique.

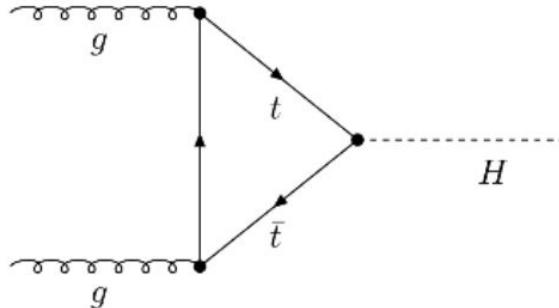


Figure 2.3: Diagram of a Higgs boson produced via gluon-gluon fusion.

<sup>292</sup> Studying the Higgs produced in association with top quark pairs,  $t\bar{t}H$ , allows this interaction to be measured directly. This process, as shown in Figure 2.4, involves a unique coupling <sup>293</sup> between the Higgs and the top, which can be identified by the top quark pair in the final state. <sup>294</sup>

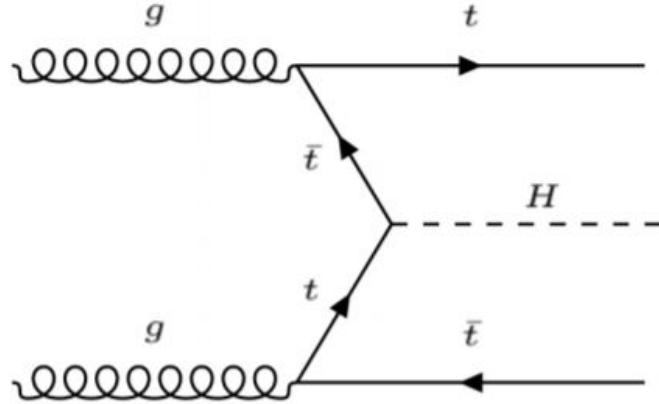


Figure 2.4: Diagram of a Higgs boson produced in association with a pair of top quarks.

295        The Higgs boson, as well as the top quarks, have very short lifetimes - on the order of  
 296         $10^{-22}$  s and  $10^{-25}$  s respectively - meaning they can only be observed via their decay products.  
 297        Measuring this process is therefore a matter of identifying events with final states consistent with  
 298         $t\bar{t}H$  production.

299        Studies of  $t\bar{t}H$  production have been reported by the ATLAS collaboration for  $H \rightarrow b\bar{b}$ ,  
 300         $H \rightarrow \gamma\gamma$  and multilepton (encompassing  $H \rightarrow W^+W^-$ ,  $H \rightarrow ZZ$  and  $H \rightarrow \tau^-\tau^+$ , with  
 301         $H \rightarrow ZZ \rightarrow 4l$  as a separate analysis) decay modes. While the branching ratio of  $H \rightarrow W^+W^-$   
 302        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  
 303        large  $t\bar{t}$  backgrounds. On the other hand,  $H \rightarrow \gamma\gamma$  produces the most easily identifiable signal,  
 304        but has a much smaller branching ratio than  $H \rightarrow W^+W^-$ . Therefore, compared with other final  
 305        states of  $t\bar{t}H$ , the  $t\bar{t}H$  – ML channel is an attractive candidate for study, as it involves a good  
 306        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.

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

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

310         $t\bar{t}H$  production was observed by ATLAS using up to  $79.8 \text{ fb}^{-1}$  of data collected at  $\sqrt{s}$   
 311         $= 13 \text{ TeV}$ , based on a combination of five Higgs decay modes:  $b\bar{b}$ ,  $WW^*$ ,  $\tau^-\tau^+$ ,  $\gamma\gamma$ , and  $ZZ^*$   
 312        [6]. A significance of  $5.8\sigma$  was observed, compared to a  $4.9\sigma$  expected significance. Since then,  
 313        two analyses have published updated results ( $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$ ) with the full Run 2  
 314        dataset, representing  $139 \text{ fb}^{-1}$ . Studies are still ongoing in the remaining channels.

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

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

319 **2.4 WZ + Heavy Flavor Production**

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

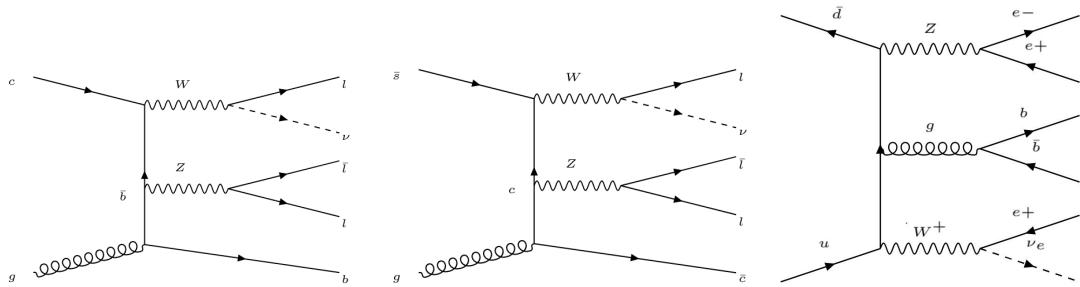


Figure 2.5: Example Feynman diagrams of  $WZ +$  heavy flavor production

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

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

332 **2.5 Extensions to the Standard Model**

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

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

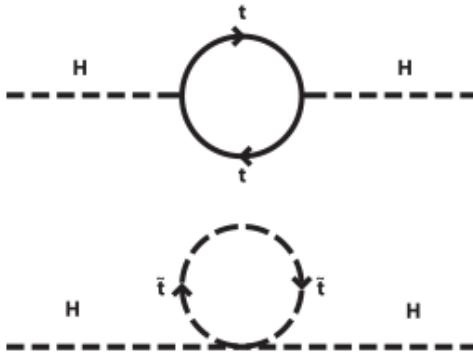


Figure 2.6: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram.

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].

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

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

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

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

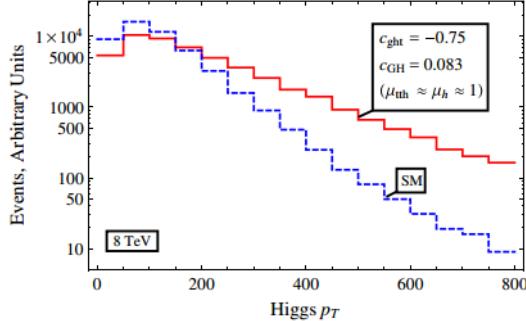


Figure 2.7: The momentum spectrum of the Higgs boson produced via top-quarks with (red) and without (blue) the presence of dimension-six operators.

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

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

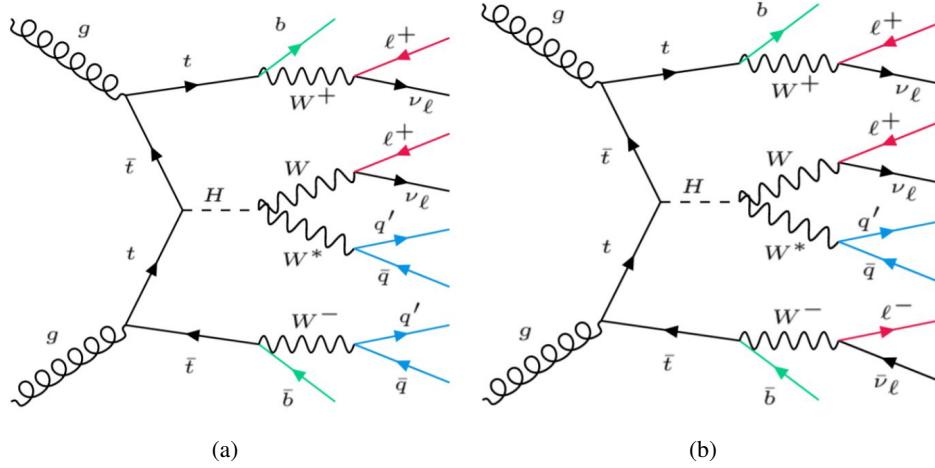


Figure 2.8: Feynman diagrams of  $t\bar{t}H$  production with (a) two same-sign leptons and (b) three leptons in the final state.

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

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

---

**Part III****388 The LHC and the ATLAS Detector****389 3 The LHC**

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

399 The LHC first began running in 2009 at a proton-proton center of mass energy of  $\sqrt{s} = 8$   
400 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during  
401 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,  
402 and collected data at an increased energy of  $\sqrt{s} = 13$  TeV until 2018, a period referred to as Run  
403 2.

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

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

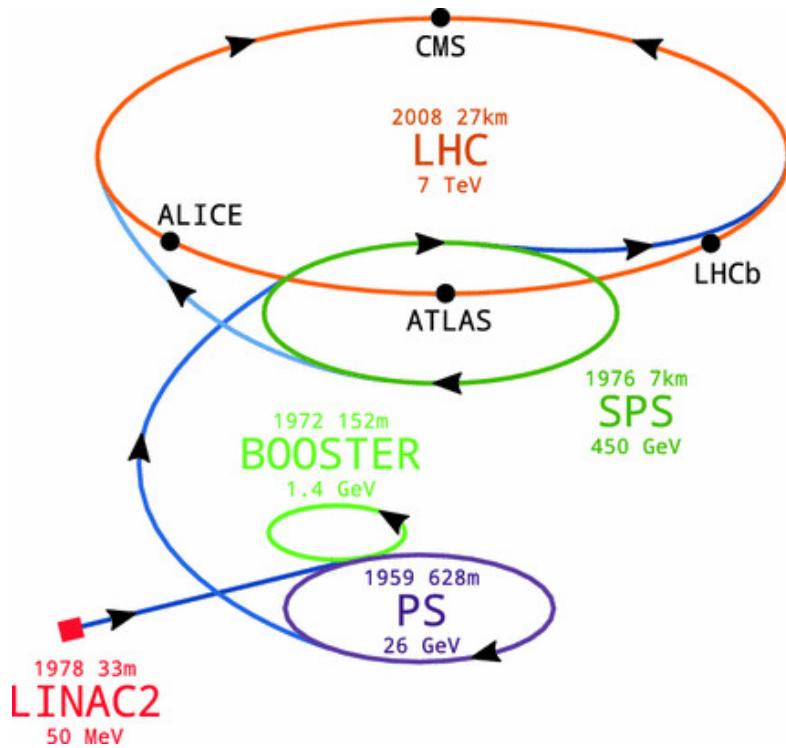


Figure 3.1: A summary of the accelerator chain used to feed protons into the LHC [11].

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

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

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

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

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

423        The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around  
 424       $140 \text{ fb}^{-1}$ , exceeding the expected integrated luminosity of  $100 \text{ fb}^{-1}$ .

## 425 4 The ATLAS Detector

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

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

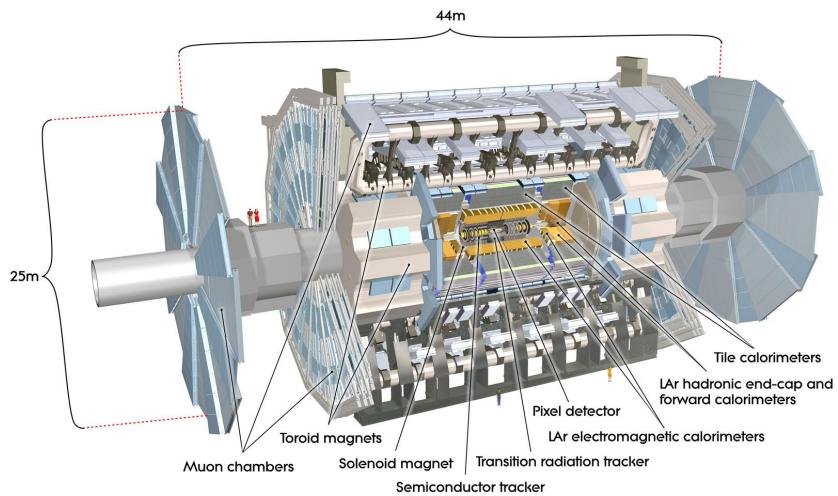


Figure 4.1: Cutaway view of the ATLAS detector, with labels of its major components [12].

434 **4.1 Inner Detector**

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

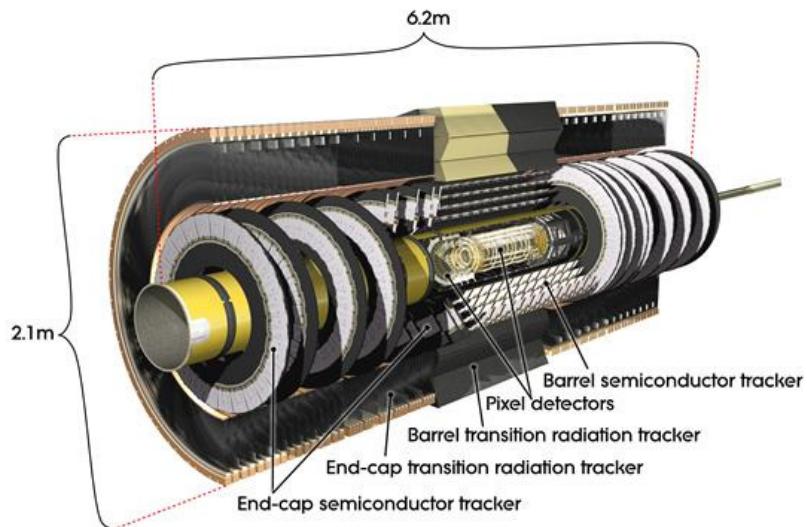


Figure 4.2: Cutaway view of the Inner Detector [13].

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

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

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

## 454        4.2 Calorimeters

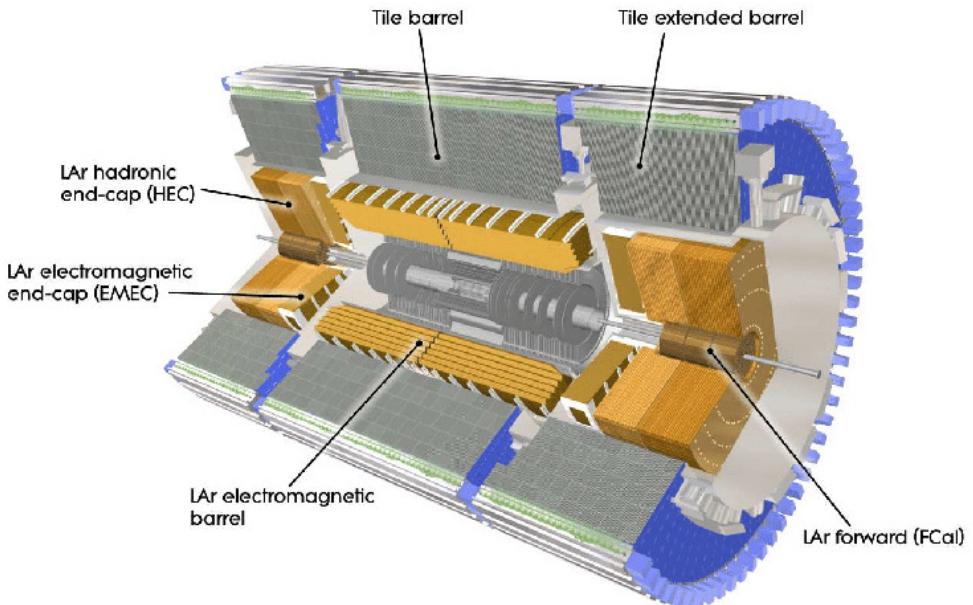


Figure 4.3: Cutaway view of the calorimeter system of the ATLAS detector [13].

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

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

### 467 **4.3 Muon Spectrometer**

468 Because muons are heavier than electrons and photons, and do not interact via the strong force,  
469 they generally pass through the detector without being stopped by the calorimeters. The outermost  
470 components of the detector are designed specifically to measure the energy and momentum of  
471 muons produced in the LHC. The muon spectrometer consists of tracking and triggering systems.  
472 The largest of the subdetectors, it extends from the outside of the calorimeter system, about a  
473 4.25 m radius from the beam line, to a radius of 11 m. This large detector system is necessary

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

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

483 1200 tracking chambers are placed in the muon system in order to precisely measure the  
484 tracks of muons with high spatial resolution. The path of the muons are tracked by Monitored  
485 Drift Tubes, which are drift chambers formed by aluminum tubes and filled with ionizing gas.  
486 These tubes produce a multi-layer spatial resolution on the order of 50  $\mu\text{m}$ .

#### 487 4.4 Trigger System

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

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

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

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

---

**506 Part IV****507 Measurement of WZ + Heavy Flavor****508 5 Introduction**

509 The production of  $WZ$  in association with a heavy flavor jet represents an important background  
510 for many major analyses. This includes any process with leptons and b-jets in the final state,  
511 such as  $t\bar{t}H$ ,  $t\bar{t}W$ , and  $t\bar{t}Z$ . While precise measurements have been made of  $WZ$  production  
512 [14],  $WZ$  + heavy flavor remains poorly understood. This is largely because the QCD processes  
513 involved in the production of the b-jet make it difficult to simulate accurately. This introduces a  
514 large uncertainty for analyses that include this process as a background.

515 Motivated by its relevance to the  $t\bar{t}H$  multilepton analysis, we perform a study of the fully  
516 leptonic decay mode of this channel; that is, events where both the  $W$  and  $Z$  decay leptonically.  
517 Because  $WZ$  has no associated jets at leading order, while the major backgrounds for this channel  
518 tend to have high jet multiplicity, events with more than two jets are rejected. This gives a final  
519 state signature of three leptons and one or two jets.

520 Events that meet this selection criteria are sorted into pseudo-continuous b-tagging regions  
521 based on the DL1r b-tag score of their associated jets. This is done to separate  $WZ$  + b-jet events  
522 from  $WZ$  + charm and  $WZ$  + light jets. These regions are fit to data in order to make a more  
523 accurate estimate of the contribution of  $WZ$  + heavy-flavor, where heavy-flavor jets include

524 b-jets and charm jets. The full Run-2 dataset collected by the ATLAS detector, representing 139  
525  $\text{fb}^{-1}$  of data from pp collisions at  $\sqrt{s} = 13 \text{ TeV}$ , is used for this study.

526       Section 10 details the data and Monte Carlo (MC) samples used in the analysis. The  
527 reconstruction of various physics objects is described in Section 11. Section 12 describes the  
528 event selection applied to these samples, along the definitions of the various regions used in  
529 the fit. The multivariate analysis techniques used to separate the tZ background from WZ +  
530 heavy flavor are described in Section 13. Section 20 describes the various sources of systematic  
531 uncertainties considered in the fit. Finally, the results of the analysis are summarized in Section  
532 21, followed by a brief conclusion in Section ??.

533       **The current state of the analysis shows blinded results for the full 2018 dataset.**  
534       **Regions containing >5% WZ+b events are blinded, and results are from Asimov, MC only**  
535       **fits.**

## 536       6 Data and Monte Carlo Samples

### 537       6.1 Data Samples

538       This study uses a sample of proton-proton collision data collected by the ATLAS detector from  
539 2015 through 2018 at an energy of  $\sqrt{s} = 13 \text{ TeV}$ , which represents an integrated luminosity of  
540  $139 \text{ fb}^{-1}$ . This data set was collected with a bunch crossing rate of 25 ns. All data used in this  
541 analysis was verified by data quality checks.

542 **6.2 Monte Carlo Samples**

543 Several different generators were used to produce Monte Carlo simulations of the signal and  
544 background processes. For all samples, the response of the ATLAS detector is simulated using  
545 Geant4. The WZ signal samples are simulated using Sherpa 2.2.2 [15]. Specific information  
546 about the Monte Carlo samples being used can be found in Table 22.

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, VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
tZ	MG5_AMC	NLO	PYTHIA 8	CTEQ6L1
t̄tW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
t̄t(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄tH	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [16] (CT10 [17])
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t̄t, t̄tt̄t̄t̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄t	POWHEG-BOX v2 [18]	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 [19]	NLO	PYTHIA 6	CT10
qqVV, VVV Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

## 547 7 Object Reconstruction

548 All regions defined in this analysis share a common lepton, jet, and overall event preselection.  
 549 The selection applied to each physics object is detailed here; the event preselection, and the  
 550 selection used to define the various fit regions, is described in Section 12.

551 All events are required to be selected by dilepton triggers. The  $p_T$  thresholds of the  
 552 dilepton trigger on two electrons were 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017 and  
 553 2018, while for the dimuon triggers the  $p_T$  thresholds on the leading (sub-leading) muon were  
 554 18 GeV (8 GeV) in 2015, and 22 GeV (8 GeV) in 2016-2018. For the electron+muon triggers,  
 555 the  $p_T$  thresholds on the electron (muon) were 17 GeV (14 GeV) for all datasets.

### 556 7.1 Light leptons

557 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter  
 558 that are associated with charged particle tracks reconstructed in the inner detector [20]. Electron  
 559 candidates are required to have  $p_T > 10$  GeV and  $|\eta_{\text{cluster}}| < 2.47$ . Muon candidates are  
 560 reconstructed by combining inner detector tracks with track segments or full tracks in the muon  
 561 spectrometer [21]. Muon candidates are required to have  $p_T > 10$  GeV and  $|\eta| < 2.5$ . Candidates  
 562 in the transition region between different electromagnetic calorimeter components,  $1.37 <$   
 563  $|\eta_{\text{cluster}}| < 1.52$ , are rejected. A multivariate likelihood discriminant combining shower shape  
 564 and track information is used to distinguish real electrons from hadronic showers (fake electrons).  
 565 To further reduce the non-prompt electron contribution, the track is required to be consistent

566 with originating from the primary vertex; requirements are imposed on the transverse impact  
 567 parameter significance ( $|d_0|/\sigma_{d_0} < 5$ ) and the longitudinal impact parameter ( $|\Delta z_0 \sin \theta_\ell| < 0.5$   
 568 mm). Electron candidates are required to pass TightLH identification.

569 Muon candidates are reconstructed by combining inner detector tracks with track segments  
 570 or full tracks in the muon spectrometer [21]. Muon candidates are required to have  $p_T > 10$  GeV  
 571 and  $|\eta| < 2.5$ . The longitudinal impact parameter is the same for both electrons and muons, while  
 572 muons are required to pass a slightly tighter transverse impact parameter,  $|d_0|/\sigma_{d_0} < 3$ . Muons  
 573 are also required to pass Medium ID requirements.

574 Leptons are additionally required to pass a non-prompt BDT selection developed by the  
 575  $t\bar{t}H$  multilepton/ $t\bar{t}W$  analysis group. This BDT and the WPs used are summarized in Appendix  
 576 A, and described in detail in [22]. Optimized working points and scale factors for this BDT are  
 577 taken from that analysis.

## 578 7.2 Jets

579 Jets are reconstructed from calibrated topological clusters built from energy deposits in the  
 580 calorimeters [23], using the anti- $k_t$  algorithm with a radius parameter  $R = 0.4$ . Particle Flow, or  
 581 PFlow, jets are used in the analysis, which are hadronic objects reconstructed using information  
 582 from both the tracker and the calorimeter. Jets with energy contributions likely arising from noise  
 583 or detector effects are removed from consideration [24], and only jets satisfying  $p_T > 25$  GeV  
 584 and  $|\eta| < 2.5$  are used in this analysis. For jets with  $p_T < 60$  GeV and  $|\eta| < 2.4$ , a jet-track

585 association algorithm is used to confirm that the jet originates from the selected primary vertex,  
 586 in order to reject jets arising from pileup collisions [25].

587 **7.3 B-tagged Jets**

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

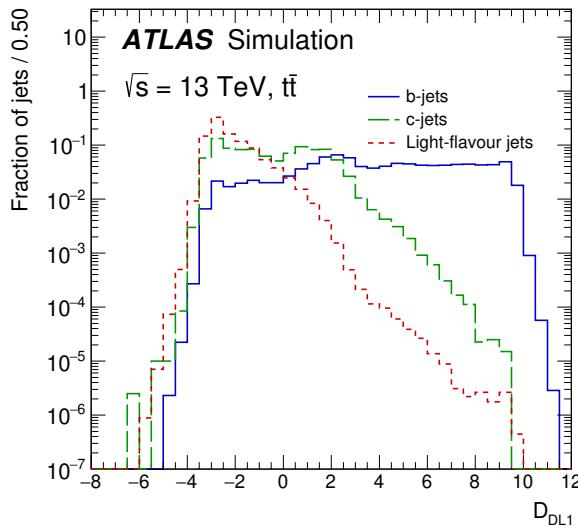


Figure 7.1: Output distribution of the DL1r algorithm for b-jets, charm jets, and light jets

593 From the output of the BDT, calibrated working points (WPs) are developed based on the  
 594 efficiency of truth b-jets at particular values of the DL1r algorithm. The working points used in

595 this analysis are summarized in Table 9.

WP	none	loose	medium	tight	tightest
b eff.	-	85%	77%	70%	60%

Table 3: B-tagging Working Points by tightness and b-jet efficiency

596 A tighter WP will accept fewer b-jets, but reject a higher fraction of charm and light jets.

597 Generally, analyses that include b-jets will use a fixed working point, for example, requiring that  
 598 a jet pass the 70% threshold. By instead treating these working point as bins, e.g. events with  
 599 jets that fall between the 85% and 77% WPs fall into one bin, while events with jets passing the  
 600 60% WP fall into another, and looking at the full psuedo-continuous DL1r spectrum of the jets,  
 601 additional information can be gained. The psuedo-continuous b-tag spectrum is used in this case  
 602 to separate out WZ + b, WZ + charm, and WZ + light.

## 603 7.4 Missing transverse energy

604 Missing transverse momentum ( $E_T^{\text{miss}}$ ) is used as part of the event selection. The missing  
 605 transverse momentum vector is defined as the inverse of the sum of the transverse momenta of  
 606 all reconstructed physics objects as well as remaining unclustered energy, the latter of which is  
 607 estimated from low- $p_T$  tracks associated with the primary vertex but not assigned to a hard object,  
 608 with object definitions taken from [26]. Light leptons considered in the  $E_T^{\text{miss}}$  reconstruction are  
 609 required to have  $p_T > 10 \text{ GeV}$ , while jets are required to have  $p_T > 20 \text{ GeV}$ .

<sup>610</sup> **7.5 Overlap removal**

<sup>611</sup> To avoid double counting objects and remove leptons originating from decays of hadrons, overlap  
<sup>612</sup> removal is performed in the following order: any electron candidate within  $\Delta R = 0.1$  of another  
<sup>613</sup> electron candidate with higher  $p_T$  is removed; any electron candidate within  $\Delta R = 0.1$  of a muon  
<sup>614</sup> candidate is removed; any jet within  $\Delta R = 0.3$  of an electron candidate is removed; if a muon  
<sup>615</sup> candidate and a jet lie within  $\Delta R = \min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$  of each other, the jet  
<sup>616</sup> is kept and the muon is removed.

<sup>617</sup> This algorithm is applied to the preselected objects. The overlap removal procedure is  
<sup>618</sup> summarized in Table 25.

Keep	Remove	Cone size ( $\Delta R$ )
electron	electron (low $p_T$ )	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	$\min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$
electron	tau	0.2

Table 4: Summary of the overlap removal procedure between electrons, muons, and jets.

<sup>619</sup> **8 Event Selection and Signal Region Definitions**

<sup>620</sup> Event are required to pass a preselection described in Section 12.1 and summarized in Table 11.  
<sup>621</sup> Those that pass this preselection are divided into various fit regions described in Section 12.2,  
<sup>622</sup> based on the number of jets in the event, and the b-tag score of those jets.

---

## 623 8.1 Event Preselection

624 Events are required to include exactly three reconstructed light leptons passing the requirement  
 625 described in 11.1, which have a total charge of  $\pm 1$ .

626 The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose  
 627 charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.  
 628 the smallest  $\Delta R$ , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton  
 629 0 is required to have  $p_T > 10$  GeV, while the same sign leptons, 1 and 2, are required to have  
 630  $p_T > 20$  GeV to reduce the contribution of non-prompt leptons.

631 The invariant mass of at least one pair of opposite sign, same flavor leptons is required  
 632 to fall within 10 GeV of the mass of the Z boson, 91.2 GeV. Events where one of the opposite  
 633 sign pairs have an invariant mass less than 12 GeV are rejected in order to suppress low mass  
 634 resonances.

635 An additional requirement is placed on the missing transverse energy,  $E_T^{\text{miss}} > 20$  GeV,  
 636 and the transverse mass of the W candidate,  $m(E_T^{\text{miss}} + l_{\text{other}}) > 30$  GeV, where  $E_T^{\text{miss}}$  is the  
 637 missing transverse energy, and  $l_{\text{other}}$  is the lepton not included in the Z-candidate.

638 Events are required to have one or two reconstructed jets passing the selection described  
 639 in Section 11.2. Events with more than two jets are rejected in order to reduce the contribution  
 640 of backgrounds such as  $t\bar{t}Z$  and  $t\bar{t}W$ , which tend to have higher jet multiplicity.

Event Selection
Exactly three leptons with charge $\pm 1$
Two same-charge leptons with $p_T > 20 \text{ GeV}$
One opposite charge lepton with $p_T > 10 \text{ GeV}$
$m(l^+l^-)$ within 10 GeV of 91.2 GeV
Transverse mass of W-candidate, $m_T(E_T^{\text{miss}} + \text{lep}_{\text{other}}) > 30 \text{ GeV}$
Missing transverse energy, $E_T^{\text{miss}} > 20 \text{ GeV}$
One or two jets with $p_T > 25 \text{ GeV}$

Table 5: Summary of the selection applied to events for inclusion in the fit

641        The event yields in the preselection region for both data and Monte Carlo are summarized  
 642        in Table 12.1, which shows good agreement between data and Monte Carlo, and demonstrates  
 643        that this region consists primarily of WZ events. The WZ events are split into WZ + b, WZ +  
 644        c, and WZ + l based on the truth flavor of the associated jet in the event. In this ordering b-jet  
 645        supersede charm, which supersedes light. That is, WZ + l events contain no charm and no b jets  
 646        at truth level, WZ + c contain at least one truth charm and no b-jets, and WZ + b contains at least  
 647        one truth b-jet.

Process	Events
WZ + b	$167.6 \pm 6.5$
WZ + c	$1080 \pm 40$
WZ + l	$7220 \pm 310$
Other VV	$850 \pm 140$
t̄tW	$16.8 \pm 2.3$
t̄tZ	$115 \pm 17$
rare Top	$2.2 \pm 0.1$
Single top	$0.10 \pm 0.45$
Three top	$0.01 \pm 0.01$
Four top	$0.02 \pm 0.01$
t̄tWW	$0.23 \pm 0.05$
Z + jets	$600 \pm 260$
V + $\gamma$	$37 \pm 54$
tZ	$190 \pm 70$
tW	$5.5 \pm 1.2$
WtZ	$25.8 \pm 1.1$
VVV	$26.2 \pm 0.9$
VH	$94 \pm 7$
t̄t	$108.68 \pm 8$
t̄tH	$4.3 \pm 0.5$
Total	$10600 \pm 530$
Data	10574

Table 6: Event yields in the preselection region at  $139.0 \text{ fb}^{-1}$ 

648 Here Other VV represents diboson processes other than WZ, and consists predominantly  
 649 of  $ZZ \rightarrow ll\bar{l}\bar{l}$  events where one of the leptons is not reconstructed.

## 650 8.2 Fit Regions

651 Once preselection has been applied, the remaining events are categorized into one of twelve  
 652 orthogonal regions. The regions used in the fit are summarized in Table 13.

Table 7: 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%	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_60} = 1, tZ \text{ BDT} > 0.725$
1j tZ CR	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_60} = 1, tZ \text{ BDT} < 0.725$
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%	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_60} >= 1, tZ \text{ BDT} > 0.725$
2j tZ CR	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_60} >= 1, tZ \text{ BDT} < 0.725$

653        The working points discussed in Section 11.3 are used to separate events into fit regions  
 654        based on the highest working point reached by a jet in each event. Because the background  
 655        composition differs significantly based on the number of b-jets, events are further subdivided  
 656        into 1-jet and 2-jet regions in order to minimize the impact of background uncertainties.

657        An unfolding procedure is performed to account for differences in the number of recon-  
 658        structed jets compared to the number of truth jets in each event. In order to account for migration  
 659        of WZ+1-jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal  
 660        samples are separated based on the number of truth jets. Events with 0 jets or more than 3 jets at  
 661        truth level, yet fall within one of the categories listed in Table 13, are categorized as WZ + other,  
 662        and treated as a background. The composition of the number of truth jets in each reco jet bin is  
 663        taken from MC, with uncertainties in these estimates described in detail in Section 20.

664        An additional tZ control region is created based on the BDT described in Section 13. The

665 region with 1-jet passing the 60% working point is split in two - a signal enriched region of  
666 events with a BDT score greater than 0.03, and a tZ control region including events with less  
667 than 0.03. This cutoff is arrived at by performing an Asimov fit with a variety of cutoffs, and  
668 selecting the value that produces the highest significance for the measurement of WZ + b.

### 669 **8.3 Non-Prompt Lepton Estimation**

670 Two processes act as sources of non-prompt leptons appear in the analysis:  $t\bar{t}$  and Z+jet  
671 production both produce two prompt leptons, and each contribute to the 3l region when an  
672 additional non-prompt lepton appears in the event. The contribution of these processes is  
673 estimated with Monte Carlo simulations, which are validated using enriched control regions.

674 The modelling in the Z+jets and  $t\bar{t}$  CRs is further validated for each of the pseudo-  
675 continuous b-tag regions used in the analysis. The relevant lepton  $p_T$  spectrum in each b-tag  
676 region is shown in Appendix B.1 for these CRs after the correction factors derived below have  
677 been applied.

#### 678 **8.3.1 $t\bar{t}$ Validation**

679  $t\bar{t}$  events can produce two prompt leptons from the decay of each of the tops. These top decays  
680 produce two b-quarks, the decay of which can produce additional non-prompt leptons, which  
681 occasionally pass the event preselection. In order to validate that the Monte Carlo accurately

682 simulates this process accurately, the MC prediction in a non-prompt  $t\bar{t}$  enriched control region  
683 is compared to data.

684 The  $t\bar{t}$  control region is similar to the preselection region - three leptons meeting the  
685 criteria described in Section 12 are required, and the requirements on  $E_T^{\text{miss}}$  remain the same.  
686 However, the selection requiring a lepton pair form a Z-candidate are reversed. Events where the  
687 invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of 91.2 GeV are  
688 rejected. This ensures the  $t\bar{t}$  control region is orthogonal to the preselection region.

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

692 Data is compared to MC predictions in the region for a variety of kinematic variable, as well  
693 as various b-tag WPs. A constant normalization discrepancy between data and MC predictions  
694 of approximately 10% is found, which is accounted for by applying a constant correction factor  
695 of 0.9 to the  $t\bar{t}$  MC prediction. As data and MC are found to agree within 20% for each of  
696 the b-tag WPs considered, a 20% systematic uncertainty on the  $t\bar{t}$  prediction is included for the  
697 analysis.

### 698 **8.3.2 Z+jets Validation**

699 Similar to  $t\bar{t}$ , a non-prompt Z+jets control region is produced in order to validate the MC  
700 predictions. The lepton requirements remain the same as the preselection region. Because no

701 neutrinos are present for this process, the  $E_T^{\text{miss}}$  cut is reversed, requiring  $E_T^{\text{miss}} < 30 \text{ GeV}$ . This  
 702 also ensures this control region is orthogonal to the preselection region. Further, the number of  
 703 jets in each event is required to be greater than or equal to one.

704 While there is general agreement between data and MC within statistical uncertainty, the  
 705 shape of the  $p_T$  spectrum of the lepton from the W is found to differ. To account for this  
 706 discrepancy, a variable correction factor is applied to Z+jets.  $\chi^2$  minimization of the W lepton  
 707  $p_T$  spectrum is performed to derive a correction factor of  $1.53 - 6.6 * 10^{-6}(\text{lep}_\text{Pt}_W)$ .

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

## 712 9 tZ Interference Studies and Separation Multivariate Analysis

713 An important process to consider in this analysis is tZ: the top almost always decays into a W  
 714 boson and b-quark, and when both the W and Z decay leptonically, this gives three leptons and  
 715 a heavy flavor jet in the final state. Because tZ can produce a final state identical to the signal,  
 716 it represents a predominant background in the most signal enriched regions. That is, the region  
 717 with one jet passing the 60% DL1r WP. Therefore, a boosted decision tree (BDT) algorithm is  
 718 trained using XGBoost [29] to separate WZ + heavy flavor from tZ. The result of this BDT is

<sup>719</sup> used to create a tZ enriched region in the fit, reducing its impact on the measurement of WZ +  
<sup>720</sup> heavy flavor.

<sup>721</sup> The following kinematic variables are used as inputs to train this BDT:

- <sup>722</sup> • The invariant mass of the reconstructed top candidate
- <sup>723</sup>     •  $p_T$  of each of the leptons, jet
- <sup>724</sup>     • The invariant mass of each combination of lepton pairs,  $M(l\bar{l})$
- <sup>725</sup>     •  $E_T^{\text{miss}}$
- <sup>726</sup>     • Distance between each combination of leptons,  $\Delta R(l\bar{l})$
- <sup>727</sup>     • Distance between each lepton and the jet,  $\Delta R(lj)$

<sup>728</sup> Here the top candidate is reconstructed based on the procedure described in section 6.1 of  
<sup>729</sup> [28]. Broadly, the mass of the top quark candidate is reconstructed from the jet, the lepton not  
<sup>730</sup> included in the Z-candidate, and a reconstructed neutrino. In the case that there is one jet in the  
<sup>731</sup> event, there is only possible b-jet candidate. For events with two jets, the jet with the highest  
<sup>732</sup> DL1r score is used.

<sup>733</sup> The training samples included only events meeting the requirements of the 1-jet, >60%  
<sup>734</sup> region, i.e. passing all the selection described in section 12 and having exactly one jet which  
<sup>735</sup> passes the tightest (60%) DL1r working point. A sample of 20,000 background (tZ) and signal

736 (WZ+b) Monte Carlo events are used to train the BDT. And additional 5,000 events are reserved  
 737 for testing the model, in order to prevent over-fitting. A total of 750 decision trees with a  
 738 maximum depth of 6 branches are used to build the model. These parameters are chosen  
 739 empirically, by training several models with different parameters and selecting the one that gave  
 740 the best separation for the test sample.

741 The results of the BDT training are shown in figure 13.1. The output scores for both signal  
 742 and background events is shown on the left. The right shows the receiving operating characteristic  
 743 (ROC) curve that results from the MVA. The ROC curve represents the background rejection  
 744 as a function of signal efficiency, where each point on the curve represents a different response  
 745 score. The ROC curve of the BDT is compared to the performance of using an optimal set of flat  
 746 selections on the same set of input variables.

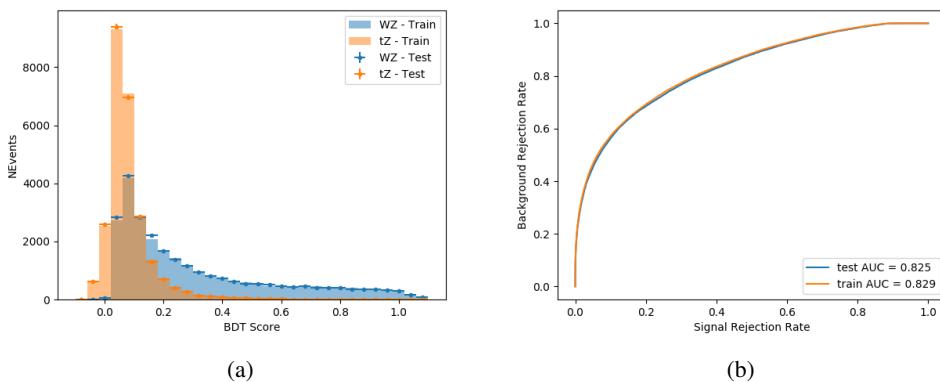


Figure 9.1: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

747 The relative important of each input feature in the model, measured by how often they  
 748 appeared in the decision trees, is shown in figure 13.2.

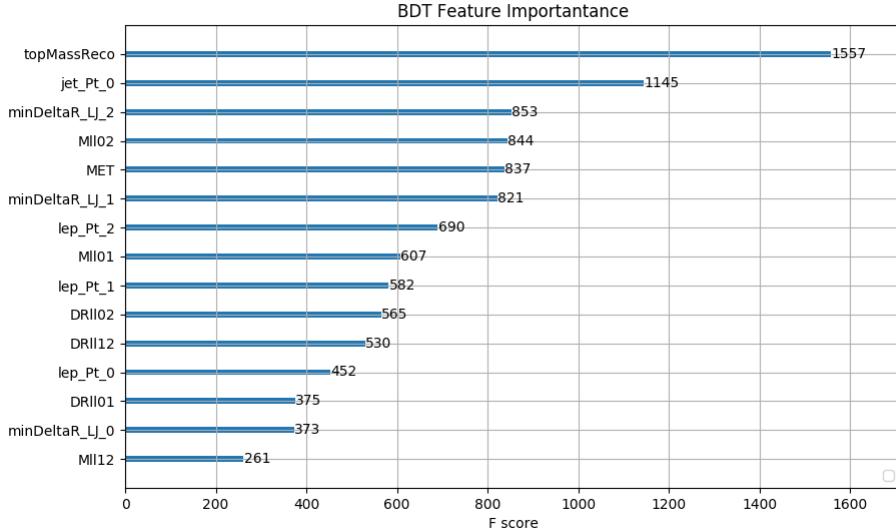


Figure 9.2: Relative importance of each input feature in the model.

749 These results suggest that some amount of separation can be achieved between these two  
 750 processes, with a high BDT score selecting a set of events that is pure in  $WZ + b$ . A BDT score  
 751 of 0.725 is selected as a cutoff, where events with scores higher than this form a signal enriched  
 752 region, and events with scores lower than this form a  $tZ$  control region. This cutoff is selected by  
 753 varying the value of this cutoff in stat-only Asimov fits, and selecting the value that minimizes  
 754 the statistical uncertainty on  $WZ + b$ .

---

**755 10 Data and Monte Carlo Samples****756 10.1 Data Samples**

757 This study uses a sample of proton-proton collision data collected by the ATLAS detector from  
758 2015 through 2018 at an energy of  $\sqrt{s} = 13$  TeV, which represents an integrated luminosity of  
759  $139 \text{ fb}^{-1}$ . This data set was collected with a bunch crossing rate of 25 ns. All data used in this  
760 analysis was verified by data quality checks.

**761 10.2 Monte Carlo Samples**

762 Several different generators were used to produce Monte Carlo simulations of the signal and  
763 background processes. For all samples, the response of the ATLAS detector is simulated using  
764 Geant4. The WZ signal samples are simulated using Sherpa 2.2.2 [15]. Specific information  
765 about the Monte Carlo samples being used can be found in Table 22.

Table 8: 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, VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
tZ	MG5_AMC	NLO	PYTHIA 8	CTEQ6L1
t̄tW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
t̄t(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄tH	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [16] (CT10 [17])
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t̄t, t̄t̄t̄t̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄t	POWHEG-BOX v2 [18]	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 [19]	NLO	PYTHIA 6	CT10
qqVV, VVV				
Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

## 11 Object Reconstruction

All regions defined in this analysis share a common lepton, jet, and overall event preselection.

768 The selection applied to each physics object is detailed here; the event preselection, and the

<sup>769</sup> selection used to define the various fit regions, is described in Section 12.

770 All events are required to be selected by dilepton triggers. The  $p_T$  thresholds of the

<sup>771</sup> dilepton trigger on two electrons were 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017 and

772 2018, while for the dimuon triggers the  $p_T$  thresholds on the leading (sub-leading) muon were

773 18 GeV (8 GeV) in 2015, and 22 GeV (8 GeV) in 2016-2018. For the electron+muon triggers,

774 the  $p_T$  thresholds on the electron (muon) were 17 GeV (14 GeV) for all datasets.

775 **11.1 Light leptons**

776 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter

777 that are associated with charged particle tracks reconstructed in the inner detector [20]. Electron

778 candidates are required to have  $p_T > 10$  GeV and  $|\eta_{\text{cluster}}| < 2.47$ . Muon candidates are

779 reconstructed by combining inner detector tracks with track segments or full tracks in the muon

780 spectrometer [21]. Muon candidates are required to have  $p_T > 10$  GeV and  $|\eta| < 2.5$ . Candidates

781 in the transition region between different electromagnetic calorimeter components,  $1.37 <$

782  $|\eta_{\text{cluster}}| < 1.52$ , are rejected. A multivariate likelihood discriminant combining shower shape

783 and track information is used to distinguish real electrons from hadronic showers (fake electrons).

784 To further reduce the non-prompt electron contribution, the track is required to be consistent

785 with originating from the primary vertex; requirements are imposed on the transverse impact

786 parameter significance ( $|d_0|/\sigma_{d_0} < 5$ ) and the longitudinal impact parameter ( $|\Delta z_0 \sin \theta_\ell| < 0.5$

787 mm). Electron candidates are required to pass TightLH identification.

788 Muon candidates are reconstructed by combining inner detector tracks with track segments

789 or full tracks in the muon spectrometer [21]. Muon candidates are required to have  $p_T > 10$  GeV

790 and  $|\eta| < 2.5$ . The longitudinal impact parameter is the same for both electrons and muons, while

791 muons are required to pass a slightly tighter transverse impact parameter,  $|d_0|/\sigma_{d_0} < 3$ . Muons  
792 are also required to pass Medium ID requirements.

793 Leptons are additionally required to pass a non-prompt BDT selection developed by the  
794  $t\bar{t}H$  multilepton/ $t\bar{t}W$  analysis group. This BDT and the WPs used are summarized in Appendix  
795 A, and described in detail in [22]. Optimized working points and scale factors for this BDT are  
796 taken from that analysis.

## 797 11.2 Jets

798 Jets are reconstructed from calibrated topological clusters built from energy deposits in the  
799 calorimeters [23], using the anti- $k_t$  algorithm with a radius parameter  $R = 0.4$ . Particle Flow, or  
800 PFlow, jets are used in the analysis, which are hadronic objects reconstructed using information  
801 from both the tracker and the calorimeter. Jets with energy contributions likely arising from noise  
802 or detector effects are removed from consideration [24], and only jets satisfying  $p_T > 25$  GeV  
803 and  $|\eta| < 2.5$  are used in this analysis. For jets with  $p_T < 60$  GeV and  $|\eta| < 2.4$ , a jet-track  
804 association algorithm is used to confirm that the jet originates from the selected primary vertex,  
805 in order to reject jets arising from pileup collisions [25].

## 806 11.3 B-tagged Jets

807 In order to make a measurement of  $WZ +$  heavy flavor it is necessary to distinguish these events  
808 from  $WZ +$  light jets. For this purpose, the DL1r b-tagging algorithm is used to distinguish

809 heavy flavor jets from lighter ones. The DL1r algorithm uses jet kinematics, particularly jet  
 810 vertex information, as input for a neural network which assigns each jet a score designed to  
 811 reflect how likely that jet is to have originated from a b-quark.

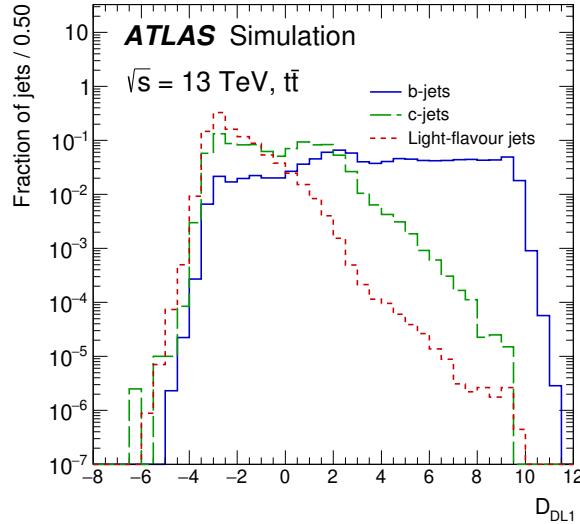


Figure 11.1: Output distribution of the DL1r algorithm for b-jets, charm jets, and light jets

812 From the output of the BDT, calibrated working points (WPs) are developed based on the  
 813 efficiency of truth b-jets at particular values of the DL1r algorithm. The working points used in  
 814 this analysis are summarized in Table 9.

WP	none	loose	medium	tight	tightest
b eff.	-	85%	77%	70%	60%

Table 9: B-tagging Working Points by tightness and b-jet efficiency

815 A tighter WP will accept fewer b-jets, but reject a higher fraction of charm and light jets.  
 816 Generally, analyses that include b-jets will use a fixed working point, for example, requiring that

817 a jet pass the 70% threshold. By instead treating these working point as bins, e.g. events with  
818 jets that fall between the 85% and 77% WPs fall into one bin, while events with jets passing the  
819 60% WP fall into another, and looking at the full psuedo-continuous DL1r spectrum of the jets,  
820 additional information can be gained. The psuedo-continuous b-tag spectrum is used in this case  
821 to separate out WZ + b, WZ + charm, and WZ + light.

## 822 **11.4 Missing transverse energy**

823 Missing transverse momentum ( $E_T^{\text{miss}}$ ) is used as part of the event selection. The missing  
824 transverse momentum vector is defined as the inverse of the sum of the transverse momenta of  
825 all reconstructed physics objects as well as remaining unclustered energy, the latter of which is  
826 estimated from low- $p_T$  tracks associated with the primary vertex but not assigned to a hard object,  
827 with object definitions taken from [26]. Light leptons considered in the  $E_T^{\text{miss}}$  reconstruction are  
828 required to have  $p_T > 10$  GeV, while jets are required to have  $p_T > 20$  GeV.

## 829 **11.5 Overlap removal**

830 To avoid double counting objects and remove leptons originating from decays of hadrons, overlap  
831 removal is performed in the following order: any electron candidate within  $\Delta R = 0.1$  of another  
832 electron candidate with higher  $p_T$  is removed; any electron candidate within  $\Delta R = 0.1$  of a muon  
833 candidate is removed; any jet within  $\Delta R = 0.3$  of an electron candidate is removed; if a muon

<sup>834</sup> candidate and a jet lie within  $\Delta R = \min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$  of each other, the jet  
<sup>835</sup> is kept and the muon is removed.

<sup>836</sup> This algorithm is applied to the preselected objects. The overlap removal procedure is  
<sup>837</sup> summarized in Table 25.

Keep	Remove	Cone size ( $\Delta R$ )
electron	electron (low $p_T$ )	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	$\min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$
electron	tau	0.2

Table 10: Summary of the overlap removal procedure between electrons, muons, and jets.

## <sup>838</sup> 12 Event Selection and Signal Region Definitions

<sup>839</sup> Event are required to pass a preselection described in Section 12.1 and summarized in Table 11.  
<sup>840</sup> Those that pass this preselection are divided into various fit regions described in Section 12.2,  
<sup>841</sup> based on the number of jets in the event, and the b-tag score of those jets.

### <sup>842</sup> 12.1 Event Preselection

<sup>843</sup> Events are required to include exactly three reconstructed light leptons passing the requirement  
<sup>844</sup> described in 11.1, which have a total charge of  $\pm 1$ .

845        The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose  
 846        charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.  
 847        the smallest  $\Delta R$ , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton  
 848        0 is required to have  $p_T > 10$  GeV, while the same sign leptons, 1 and 2, are required to have  
 849         $p_T > 20$  GeV to reduce the contribution of non-prompt leptons.

850        The invariant mass of at least one pair of opposite sign, same flavor leptons is required  
 851        to fall within 10 GeV of the mass of the Z boson, 91.2 GeV. Events where one of the opposite  
 852        sign pairs have an invariant mass less than 12 GeV are rejected in order to suppress low mass  
 853        resonances.

854        An additional requirement is placed on the missing transverse energy,  $E_T^{\text{miss}} > 20$  GeV,  
 855        and the transverse mass of the W candidate,  $m(E_T^{\text{miss}} + l_{\text{other}}) > 30$  GeV, where  $E_T^{\text{miss}}$  is the  
 856        missing transverse energy, and  $l_{\text{other}}$  is the lepton not included in the Z-candidate.

857        Events are required to have one or two reconstructed jets passing the selection described  
 858        in Section 11.2. Events with more than two jets are rejected in order to reduce the contribution  
 859        of backgrounds such as  $t\bar{t}Z$  and  $t\bar{t}W$ , which tend to have higher jet multiplicity.

---

#### Event Selection

---

- Exactly three leptons with charge  $\pm 1$
  - Two same-charge leptons with  $p_T > 20$  GeV
  - One opposite charge lepton with  $p_T > 10$  GeV
  - $m(l^+l^-)$  within 10 GeV of 91.2 GeV
  - Transverse mass of W-candidate,  $m_T(E_T^{\text{miss}} + \text{lep}_{\text{other}}) > 30$  GeV
  - Missing transverse energy,  $E_T^{\text{miss}} > 20$  GeV
  - One or two jets with  $p_T > 25$  GeV
- 

Table 11: Summary of the selection applied to events for inclusion in the fit

860        The event yields in the preselection region for both data and Monte Carlo are summarized  
 861        in Table 12.1, which shows good agreement between data and Monte Carlo, and demonstrates  
 862        that this region consists primarily of WZ events. The WZ events are split into WZ + b, WZ +  
 863        c, and WZ + l based on the truth flavor of the associated jet in the event. In this ordering b-jet  
 864        supersede charm, which supersides light. That is, WZ + l events contain no charm and no b jets  
 865        at truth level, WZ + c contain at least one truth charm and no b-jets, and WZ + b contains at least  
 866        one truth b-jet.

Process	Events
WZ + b	$167.6 \pm 6.5$
WZ + c	$1080 \pm 40$
WZ + l	$7220 \pm 310$
Other VV	$850 \pm 140$
$t\bar{t}W$	$16.8 \pm 2.3$
$t\bar{t}Z$	$115 \pm 17$
rare Top	$2.2 \pm 0.1$
Single top	$0.10 \pm 0.45$
Three top	$0.01 \pm 0.01$
Four top	$0.02 \pm 0.01$
$t\bar{t}WW$	$0.23 \pm 0.05$
Z + jets	$600 \pm 260$
V + $\gamma$	$37 \pm 54$
tZ	$190 \pm 70$
tW	$5.5 \pm 1.2$
WtZ	$25.8 \pm 1.1$
VVV	$26.2 \pm 0.9$
VH	$94 \pm 7$
t $\bar{t}$	$108.68 \pm 8$
t $\bar{t}H$	$4.3 \pm 0.5$
Total	$10600 \pm 530$
Data	10574

Table 12: Event yields in the preselection region at  $139.0 \text{ fb}^{-1}$

867        Here Other VV represents diboson processes other than WZ, and consists predominantly

868 of  $ZZ \rightarrow llll$  events where one of the leptons is not reconstructed.

869 **12.2 Fit Regions**

870 Once preselection has been applied, the remaining events are categorized into one of twelve  
871 orthogonal regions. The regions used in the fit are summarized in Table 13.

Table 13: 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%	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_60} = 1, tZ \text{ BDT} > 0.725$
1j tZ CR	$N_{\text{jets}} = 1, n\text{Jets\_DL1r\_60} = 1, tZ \text{ BDT} < 0.725$
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%	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_60} >= 1, tZ \text{ BDT} > 0.725$
2j tZ CR	$N_{\text{jets}} = 2, n\text{Jets\_DL1r\_60} >= 1, tZ \text{ BDT} < 0.725$

872 The working points discussed in Section 11.3 are used to separate events into fit regions  
873 based on the highest working point reached by a jet in each event. Because the background  
874 composition differs significantly based on the number of b-jets, events are further subdivided  
875 into 1-jet and 2-jet regions in order to minimize the impact of background uncertainties.

876 An unfolding procedure is performed to account for differences in the number of recon-  
877 structed jets compared to the number of truth jets in each event. In order to account for migration

878 of WZ+1-jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal  
 879 samples are separated based on the number of truth jets. Events with 0 jets or more than 3 jets at  
 880 truth level, yet fall within one of the categories listed in Table 13, are categorized as WZ + other,  
 881 and treated as a background. The composition of the number of truth jets in each reco jet bin is  
 882 taken from MC, with uncertainties in these estimates described in detail in Section 20.

883 An additional tZ control region is created based on the BDT described in Section 13. The  
 884 region with 1-jet passing the 60% working point is split in two - a signal enriched region of  
 885 events with a BDT score greater than 0.03, and a tZ control region including events with less  
 886 than 0.03. This cutoff is arrived at by performing an Asimov fit with a variety of cutoffs, and  
 887 selecting the value that produces the highest significance for the measurement of WZ + b.

### 888 **12.3 Non-Prompt Lepton Estimation**

889 Two processes act as sources of non-prompt leptons appear in the analysis:  $t\bar{t}$  and Z+jet  
 890 production both produce two prompt leptons, and each contribute to the 31 region when an  
 891 additional non-prompt lepton appears in the event. The contribution of these processes is  
 892 estimated with Monte Carlo simulations, which are validated using enriched control regions.

893 The modelling in the Z+jets and  $t\bar{t}$  CRs is further validated for each of the pseudo-  
 894 continuous b-tag regions used in the analysis. The relevant lepton  $p_T$  spectrum in each b-tag  
 895 region is shown in Appendix B.1 for these CRs after the correction factors derived below have  
 896 been applied.

---

897 **12.3.1  $t\bar{t}$  Validation**

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

903 The  $t\bar{t}$  control region is similar to the preselection region - three leptons meeting the  
904 criteria described in Section 12 are required, and the requirements on  $E_T^{\text{miss}}$  remain the same.  
905 However, the selection requiring a lepton pair form a Z-candidate are reversed. Events where the  
906 invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of 91.2 GeV are  
907 rejected. This ensures the  $t\bar{t}$  control region is orthogonal to the preselection region.

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

911 Data is compared to MC predictions in the region for a variety of kinematic variable, as well  
912 as various b-tag WPs. A constant normalization discrepancy between data and MC predictions  
913 of approximately 10% is found, which is accounted for by applying a constant correction factor  
914 of 0.9 to the  $t\bar{t}$  MC prediction. As data and MC are found to agree within 20% for each of  
915 the b-tag WPs considered, a 20% systematic uncertainty on the  $t\bar{t}$  prediction is included for the  
916 analysis.

917 **12.3.2 Z+jets Validation**

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

923 While there is general agreement between data and MC within statistical uncertainty, the  
 924 shape of the  $p_T$  spectrum of the lepton from the  $W$  is found to differ. To account for this  
 925 discrepancy, a variable correction factor is applied to  $Z$ +jets.  $\chi^2$  minimization of the  $W$  lepton  
 926  $p_T$  spectrum is performed to derive a correction factor of  $1.53 - 6.6 * 10^{-6}(\text{lep}_\text{Pt}_W)$ .

927 The uncertainty in the  $Z + \text{jets}$  prediction is evaluated by comparing data to MC for each  
 928 of the continuous b-tag WPs. For each of the regions considered, the data falls within 25% of  
 929 the MC prediction once this correction factor has been applied. Therefore, a 25% systematic  
 930 uncertainty is applied to  $Z + \text{jets}$  in the analysis.

931 **13 tZ Interference Studies and Separation Multivariate Analysis**

932 An important process to consider in this analysis is  $tZ$ : the top almost always decays into a  $W$   
 933 boson and b-quark, and when both the  $W$  and  $Z$  decay leptonically, this gives three leptons and  
 934 a heavy flavor jet in the final state. Because  $tZ$  can produce a final state identical to the signal,

935 it represents a predominant background in the most signal enriched regions. That is, the region  
 936 with one jet passing the 60% DL1r WP. Therefore, a boosted decision tree (BDT) algorithm is  
 937 trained using XGBoost [29] to separate WZ + heavy flavor from tZ. The result of this BDT is  
 938 used to create a tZ enriched region in the fit, reducing its impact on the measurement of WZ +  
 939 heavy flavor.

940 The following kinematic variables are used as inputs to train this BDT:

- 941 • The invariant mass of the reconstructed top candidate
- 942 •  $p_T$  of each of the leptons, jet
- 943 • The invariant mass of each combination of lepton pairs,  $M(l\bar{l})$
- 944 •  $E_T^{\text{miss}}$
- 945 • Distance between each combination of leptons,  $\Delta R(l\bar{l})$
- 946 • Distance between each lepton and the jet,  $\Delta R(lj)$

947 Here the top candidate is reconstructed based on the procedure described in section 6.1 of  
 948 [28]. Broadly, the mass of the top quark candidate is reconstructed from the jet, the lepton not  
 949 included in the Z-candidate, and a reconstructed neutrino. In the case that there is one jet in the  
 950 event, there is only possible b-jet candidate. For events with two jets, the jet with the highest  
 951 DL1r score is used.

952        The training samples included only events meeting the requirements of the 1-jet, >60%  
953        region, i.e. passing all the selection described in section 12 and having exactly one jet which  
954        passes the tightest (60%) DL1r working point. A sample of 20,000 background (tZ) and signal  
955        (WZ+b) Monte Carlo events are used to train the BDT. And additional 5,000 events are reserved  
956        for testing the model, in order to prevent over-fitting. A total of 750 decision trees with a  
957        maximum depth of 6 branches are used to build the model. These parameters are chosen  
958        empirically, by training several models with different parameters and selecting the one that gave  
959        the best separation for the test sample.

960        The results of the BDT training are shown in figure 13.1. The output scores for both signal  
961        and background events is shown on the left. The right shows the receiving operating characteristic  
962        (ROC) curve that results from the MVA. The ROC curve represents the background rejection  
963        as a function of signal efficiency, where each point on the curve represents a different response  
964        score. The ROC curve of the BDT is compared to the performance of using an optimal set of flat  
965        selections on the same set of input variables.

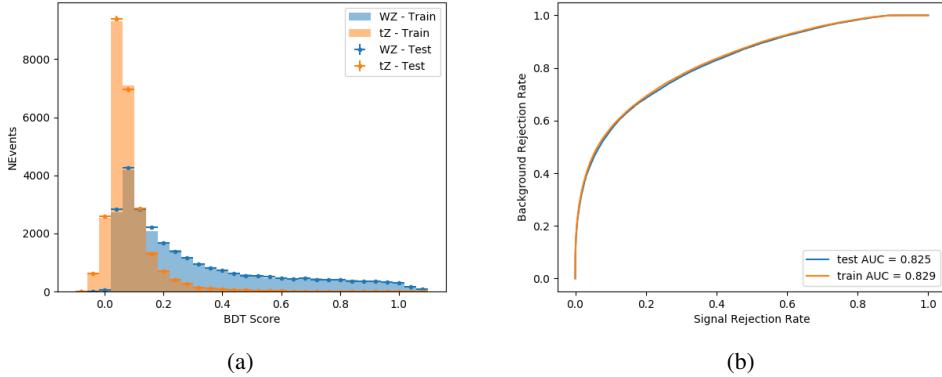


Figure 13.1: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

966        The relative important of each input feature in the model, measured by how often they  
 967        appeared in the decision trees, is shown in figure 13.2.

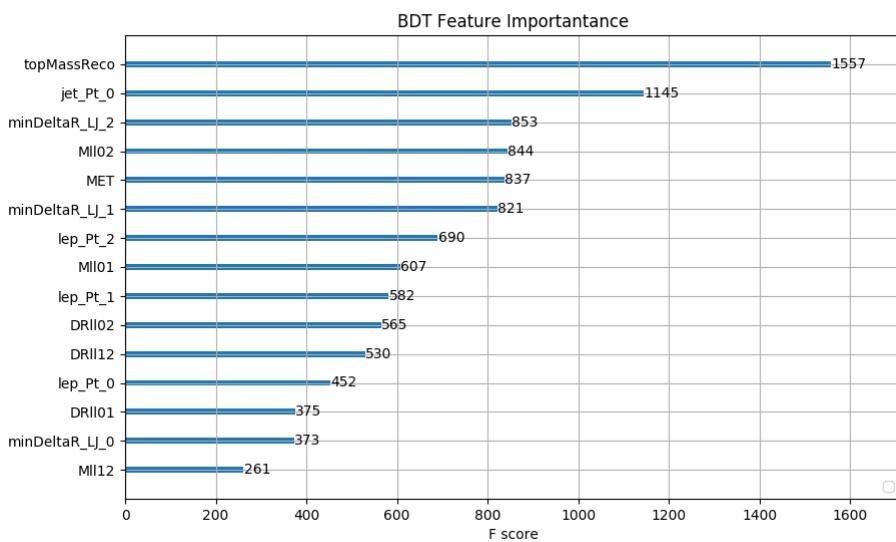


Figure 13.2: Relative importance of each input feature in the model.

968 These results suggest that some amount of separation can be achieved between these two  
969 processes, with a high BDT score selecting a set of events that is pure in  $WZ + b$ . A BDT score  
970 of 0.725 is selected as a cutoff, where events with scores higher than this form a signal enriched  
971 region, and events with scores lower than this form a  $t\bar{Z}$  control region. This cutoff is selected by  
972 varying the value of this cutoff in stat-only Asimov fits, and selecting the value that minimizes  
973 the statistical uncertainty on  $WZ + b$ .

## 974 **14 Systematic Uncertainties**

975 The systematic uncertainties that are considered are summarized in Table 45. These are imple-  
976 mented in the fit either as a normalization factors or as a shape variation or both in the signal  
977 and background estimations. The numerical impact of each of these uncertainties is outlined in  
978 Section 21.

Table 14: 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

979        The uncertainty in the combined integrated luminosity is derived from a calibration of the  
 980        luminosity scale performed for 13 TeV proton-proton collisions [30], [31].

981        The experimental uncertainties are related to the reconstruction and identification of light  
 982        leptons and b-tagging of jets, and to the reconstruction of  $E_T^{\text{miss}}$ . The sources which contribute to  
 983        the uncertainty in the jet energy scale (JES) [32] are decomposed into uncorrelated components  
 984        and treated as independent sources of uncertainty in the analysis. These are treated as 30  
 985        nuisance parameters included in the fit. A similar approach is used for the jet energy resolution

986 (JER) uncertainty, which is decomposed into 8 JER uncertainty components included as NPs in  
987 the fit.

988 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses  
989 [33] are also decomposed into uncorrelated components. The large number of components for  
990 b-tagging is due to the calibration of the distribution of the MVA discriminant.

991 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and scale  
992 uncertainties are taken from theory calculations, with the exception of non-prompt and dibo-  
993 son backgrounds. The cross-section uncertainty on tZ is taken from [34]. Derivation of the  
994 non-prompt background uncertainties, Z+jets and tt>, are explained in Section 12.3. These  
995 normalization uncertainties are chosen so as to account for the complete uncertainty in the  
996 non-prompt contribution, and therefore no additional modelling uncertainties are considered for  
997 Z+jets and tt>.

998 The other VV + heavy flavor processes (namely VV+b and VV+charm, which primarily  
999 consist of ZZ events) are also poorly understood, because these processes involve the same  
1000 physics as WZ + heavy flavor, and have also not been measured. Therefore, a conservative 50%  
1001 uncertainty is applied to those samples. While this uncertainty is large, it is found to have little  
1002 impact on the significance of the final result.

1003 The theory uncertainties applied to the predominate background estimates are summarized  
1004 in Table 50.

Process	X-section [%]
tZ	X-sec: $\pm 15.2$ QCD Scale: $^{+5.2}_{-1.3}$ PDF( $+\alpha_S$ ): $\pm 1.2$
t̄t H (aMC@NLO+Pythia8)	QCD Scale: $^{+5.8}_{-9.2}$ PDF( $+\alpha_S$ ): $\pm 3.6$
t̄t Z (aMC@NLO+Pythia8)	QCD Scale: $^{+9.6}_{-11.3}$ PDF( $+\alpha_S$ ): $\pm 4$
t̄t W (aMC@NLO+Pythia8)	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)	$\pm 6$
t̄t	$\pm 20$
Z + jets	$\pm 25$
Others	$\pm 50$

Table 15: Summary of theoretical uncertainties for MC predictions in the analysis.

1005 Additional signal uncertainties are estimated by comparing estimates from the nominal Sherpa  
 1006 WZ samples with alternate WZ samples generated with Powheg+PYTHIA8. Separate systemat-  
 1007 ics are included in the fit for WZ + b, WZ + charm and WZ + light, where the distribution among  
 1008 each of the fit regions is varied based on the prediction of the Powheg sample.

1009 A similar approach is taken to account for uncertainties in migrations between the number of  
 1010 reco and truth jets. The fraction of events with 1 truth jet which fall into the 1 jet bin versus the  
 1011 2 jet bin at reco level is compared for Sherpa and Powheg. The same is done for events with 2  
 1012 truth jets. A systematic is included where events are shifted between the 1-jet and 2-jet regions  
 1013 based on the differences between these two shapes. This is done independently for each of the  
 1014 WZ + b, WZ + charm, and WZ + light templates.

1015        Additional systematics are included to account for the uncertainty in the contamination of  
 1016      0 jet and 3 or more jet events (as defined at truth level) in the 1 and 2 reco jet bins. Because these  
 1017      events fall outside the scope of this measurement, these events are included as a background.  
 1018      As such, a normalization, rather than a shape, uncertainty is applied for this background. The  
 1019      number of WZ events with 0-jets and  $>=3$ -jets in the reconstructed 1-jet and 2-jet regions are  
 1020      compared for Sherpa and Powheg, and these differences are taken as separate normalization  
 1021      systematics on the yield of WZ+0-jet and WZ+ $>=3$ -jet events.

## 1022      15 Results

1023      A separate maximum-likelihood fit is performed over the 1-jet and 2-jet fit regions in order to  
 1024      extract the best-fit value of the WZ + b-jet and WZ + charm jet contributions. The WZ + b,  
 1025      WZ + charm and WZ + light contributions are allowed to float, with the remaining background  
 1026      contributions are held fixed. **The current fit strategy treats the WZ + b-jet contribution as**  
 1027      **the parameter of interest, with the normalization of the WZ + charm and the WZ + light**  
 1028      **contributions taken as systematic uncertainties. This could however be adjusted, depending**  
 1029      **on whether it is decided the goal of the analysis should be to measure WZ+b specifically or**  
 1030      **WZ + heavy flavor overall.** The result of the fit is used to extract the cross-section of WZ +  
 1031      heavy-flavor production.

1032      A maximum likelihood fit to data is performed simultaneously in the regions described  
 1033      in Section 12. The parameters  $\mu_{WZ+b}$ ,  $\mu_{WZ+charm}$ ,  $\mu_{WZ+light}$ , where  $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$ , are

1034 extracted from the fit.

1035 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  
 1036 WZ + b. The fitted cross-section modifiers for WZ + charm and WZ + light are  $1.00 \pm 0.17 \pm 0.15$   
 1037 and  $1.00 \pm 0.04 \pm 0.07$ , respectively.

1038 The expected cross-section of WZ+b with 1 associated jet obtained from the fit is  
 1039  $1.74^{+0.65}_{-0.61}(\text{stat})^{+0.42}_{-0.37}(\text{sys})$  fb with an expected significance of  $2.2\sigma$ . The expected cross-sectin  
 1040 of WZ + charm is measured to be  $14.6 \pm 2.5(\text{stat}) \pm 2.1(\text{sys})$  fb, with a correlation of -0.23.

1041 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 +  
 1042 b. The fitted cross-section modifiers for WZ + charm and WZ + light are  $1.00 \pm 0.17 \pm 0.15$  and  
 1043  $1.00 \pm 0.04 \pm 0.08$ , respectively.

1044 The expected WZ + b cross-section in the 2-jet region is  $2.46^{+0.83}_{-0.81}(\text{stat})^{+0.073}_{-0.68}(\text{sys})$  fb  
 1045 with an expected significance of  $2.6\sigma$ . The 2-jet expected cross-section of WZ + charm is  
 1046  $12.7 \pm 2.2(\text{stat}) \pm 2.0(\text{sys})$  fb, and the correlation between WZ + charm and WZ + b is -0.26.

## 1047 15.1 1-jet Fit Results

1048 **The results of the fit are currently blinded.**

1049 The pre-fit yields in each of the regions used in the fit are shown in Table 15.1, and  
 1050 summarized in Figure 15.1.

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	$336 \pm 26$	$17.8 \pm 2.1$	$4.3 \pm 0.6$	$1.7 \pm 0.5$	$0.36 \pm 0.08$	$0.10 \pm 0.03$
t̄W	$1.1 \pm 0.2$	$0.2 \pm 0.1$	$0.3 \pm 0.1$	$0.4 \pm 0.1$	$1.5 \pm 0.3$	$0.7 \pm 0.2$
t̄Z	$6.8 \pm 1.2$	$1.4 \pm 0.3$	$1.0 \pm 0.2$	$1.2 \pm 0.2$	$4.4 \pm 0.8$	$3.2 \pm 0.6$
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 + γ	$45 \pm 28$	$1.9 \pm 2.4$	$0.1 \pm 0.1$	$0.02 \pm 0.01$	$1.0 \pm 0.9$	$0.02 \pm 0.03$
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 16: Pre-fit yields in each of the 1-jet fit regions.

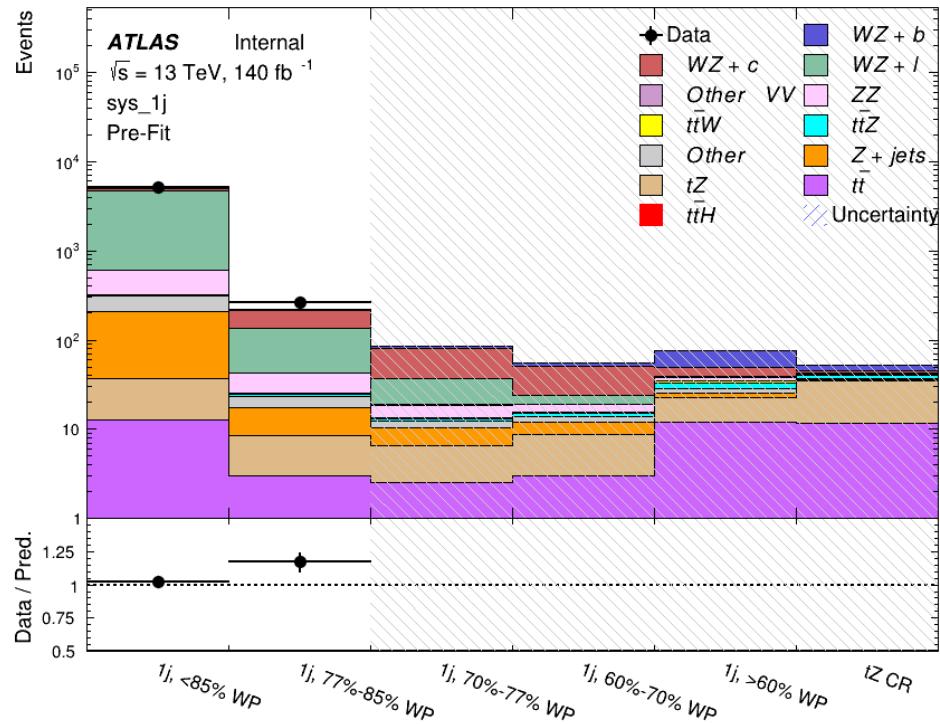


Figure 15.1: Pre-fit summary of the 1-jet fit regions.

1052

The post-fit yields in each region are summarized in Figure 15.1.

1053

	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 + γ	$46 \pm 28$	$1.9 \pm 2.4$	$0.1 \pm 0.1$	$0.0 \pm 0.2$	$1.0 \pm 0.9$	$0.0 \pm 0.0$
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 17: Post-fit yields in each of the 1-jet fit regions.

1054

A post-fit summary plot of the 1-jet fitted regions is shown in Figure 15.2:

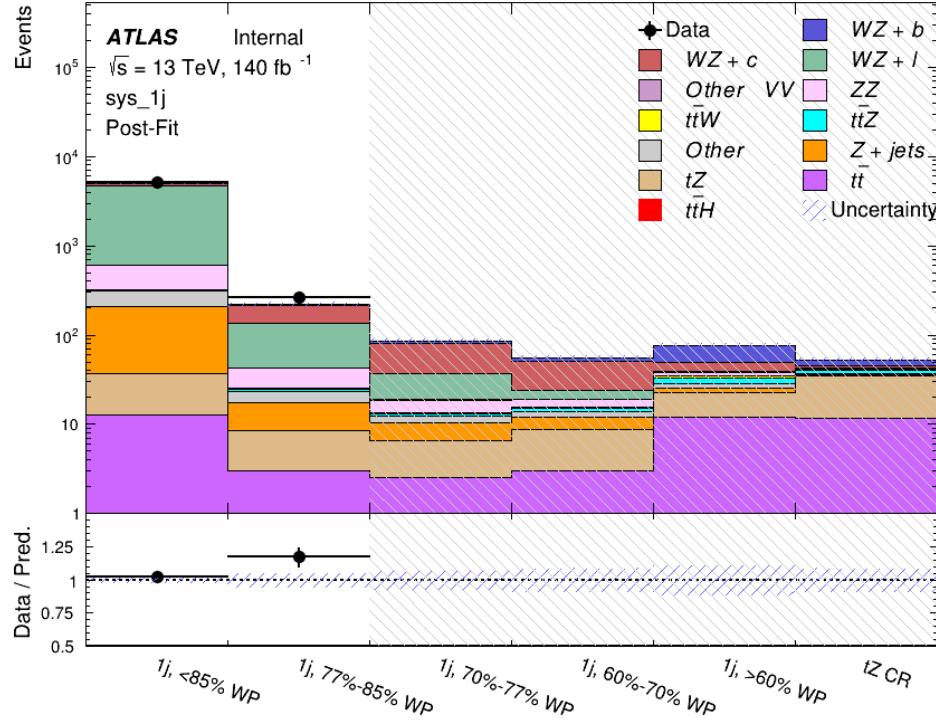


Figure 15.2: Post-fit summary of the 1-jet fit regions.

As described in Section 20, there are 226 systematic uncertainties that are considered as NPs in the fit. These NPs are constrained by Gaussian or log-normal probability density functions. The latter are used for normalisation factors to ensure that they are always positive. The expected number of signal and background events are functions of the likelihood. The prior for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its nominal value.

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

<sup>1063</sup> fit. The impact of the most significant sources of systematic uncertainties is summarized in Table

<sup>1064</sup> [18.](#)

Uncertainty Source	$\Delta\mu$	
WZ + light cross-section	0.13	-0.12
WZ + charm 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</bar>	-0.05	0.05
WZ cross-section - QCD scale	-0.04	0.03
Total Systematic Uncertainty	0.28	0.32

Table 18: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

<sup>1065</sup> The ranking and impact of those nuisance parameters with the largest contribution to the

<sup>1066</sup> overall uncertainty is shown in Figure [15.3](#).

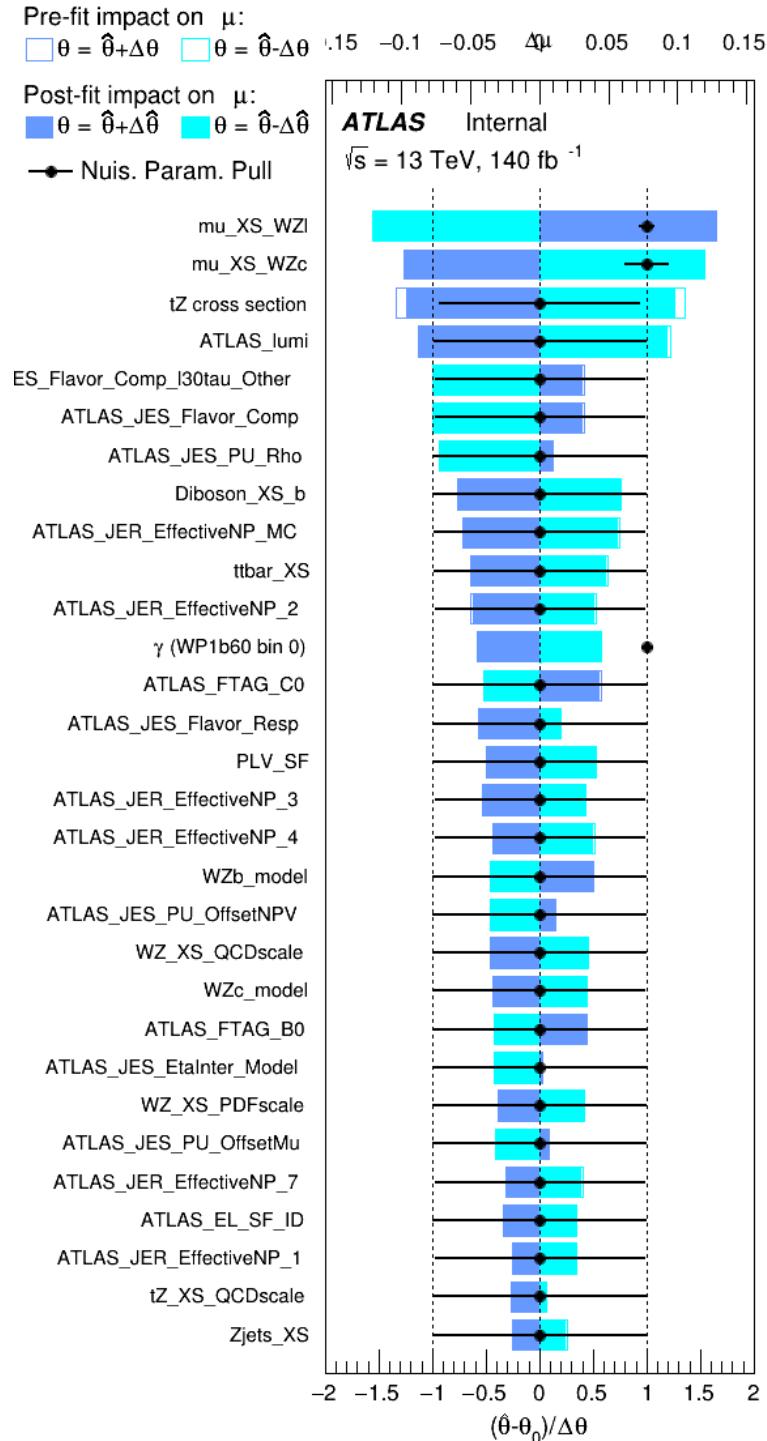


Figure 15.3: Impact of systematic uncertainties on the signal-strength of  $WZ + b$  for events with exactly one jet

1067 The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the  
 1068 shape of the fit regions depends heavily on the modeling of the jets. The other major sources  
 1069 of uncertainty come from background modelling and cross-section uncertainty. The pie charts  
 1070 in Figure 15.4 show that for the modelling uncertainties that contribute most correspond to the  
 1071 most significant backgrounds.

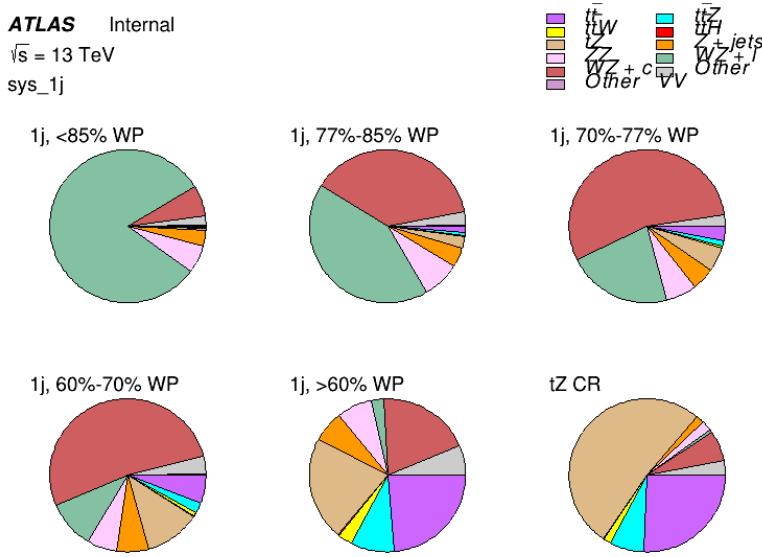


Figure 15.4: Post-fit background composition of the fit regions.

1072 The correlations between these nuisance parameters are summarized in Figure 15.5.

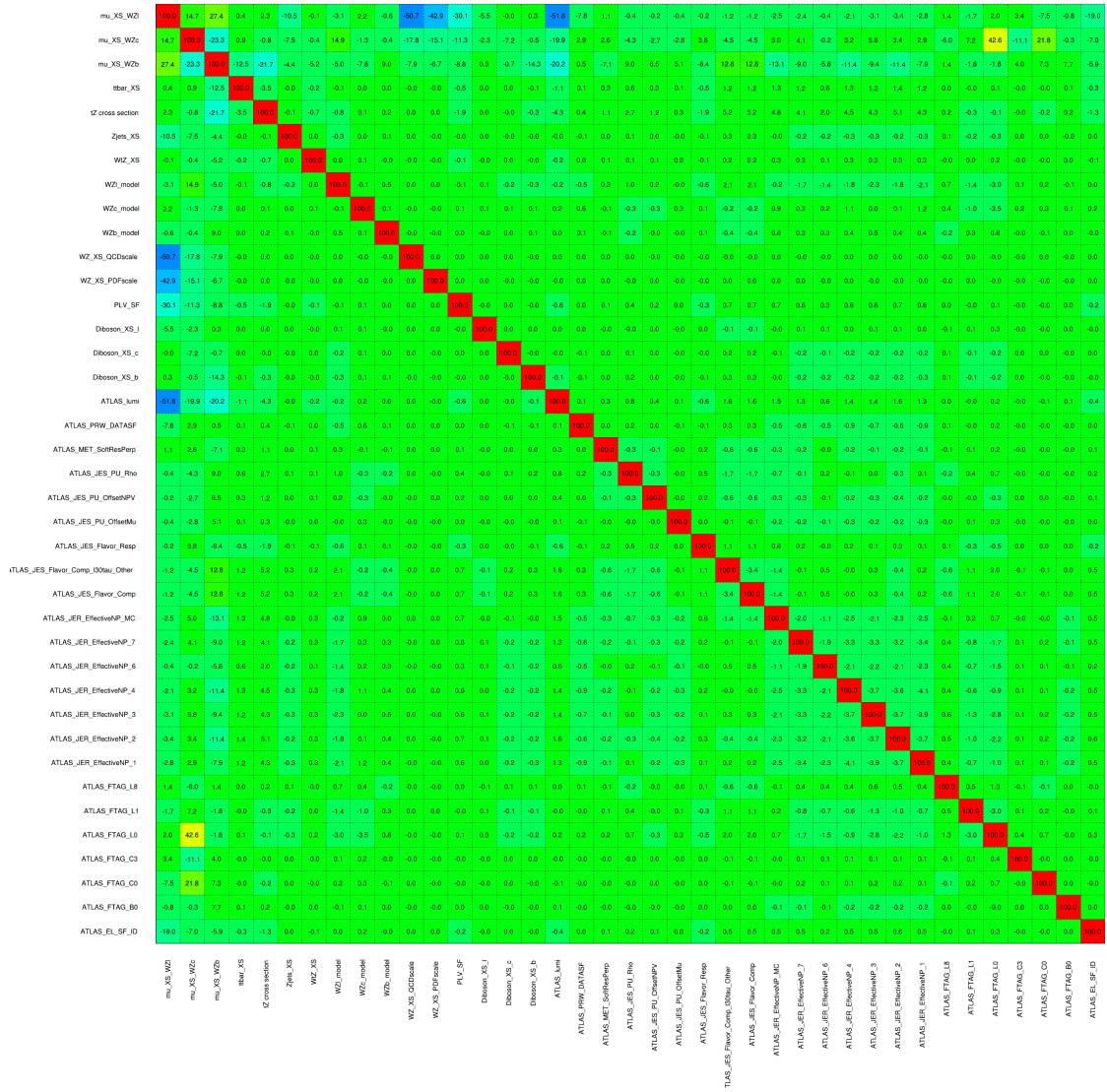


Figure 15.5: Correlations between nuisance parameters

1073 The negative correlations between  $\mu_{WZ+charm}$  and  $\mu_{WZ+b}$  and  $\mu_{WZ+light}$  are expected:  
1074  $WZ + charm$  is present in both the  $WZ + b$  and  $WZ + light$  enriched regions, therefore increasing  
1075 the fraction of charm requires increasing the fraction of  $WZ + b$  and  $WZ + light$ . This reasoning  
1076 also explains the positive correlation between  $\mu_{WZ+b}$  and  $\mu_{WZ+light}$ .

1077 Two of the major backgrounds in the region with the highest purity of  $WZ + b$  are  $tZ$  and  
1078 Other  $VV + b$ , explaining the negative correlations between  $\mu_{WZ+b}$  and the  $tZ$  cross section, and  
1079 the  $VV + b$  cross section.

1080 The high correlation between the luminosity and  $\mu_{WZ+light}$  arises from the fact that the  
1081 uncertainty on  $\mu_{WZ+light}$  is very low (around 4%). Small changes in luminosity cause a change  
1082 in the yield of  $WZ + light$  that is large compared to its uncertainty, producing a large correlation  
1083 between these two parameters.

1084 **15.2 2-jet Fit Results**

1085 **The results of the fit are currently blinded.**

1086 Pre-fit yields in each of the 2-jet fit regions are shown in Figure 15.2.

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

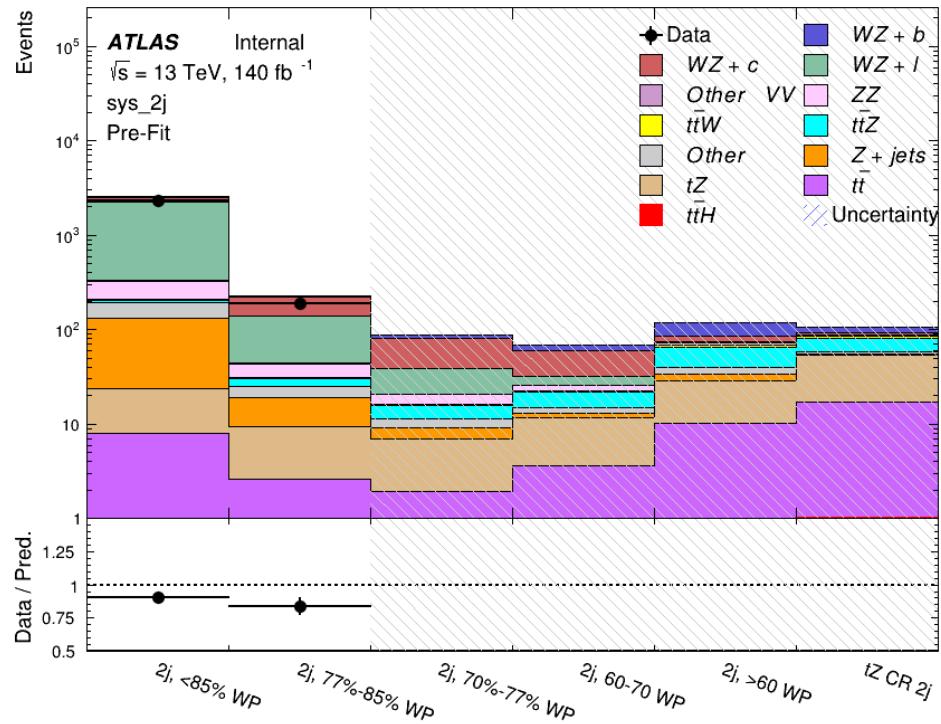


Figure 15.6: Pre-fit summary of the 2-jet fit regions.

1087

The post-fit yields in each region are summarized in Figure 15.2.

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

1088

A post-fit summary of the fitted regions is shown in Figure 15.7:

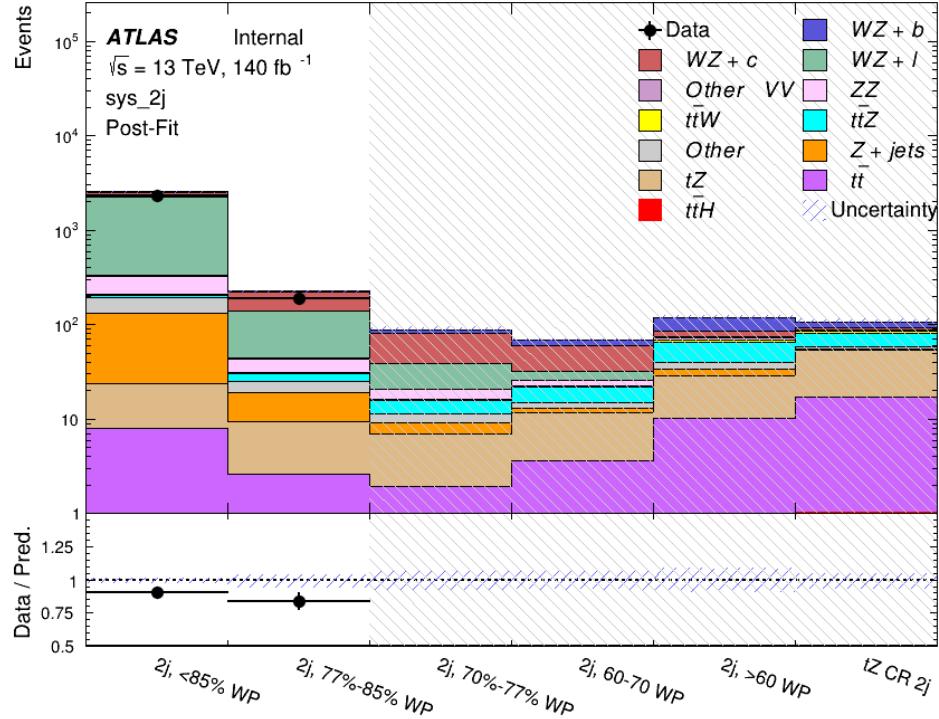


Figure 15.7: Post-fit summary of the fit over 2-jet regions.

1089        The same set of systematic uncertainties consider for the 1-jet fit are included in the 2-jet  
 1090        fit as well. The impact of the most significant systematic uncertainties is summarized in Table  
 1091        [21](#).

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

<sup>1092</sup> The ranking and impact of those nuisance parameters with the largest contribution to the  
<sup>1093</sup> overall uncertainty is shown in Figure 15.8.

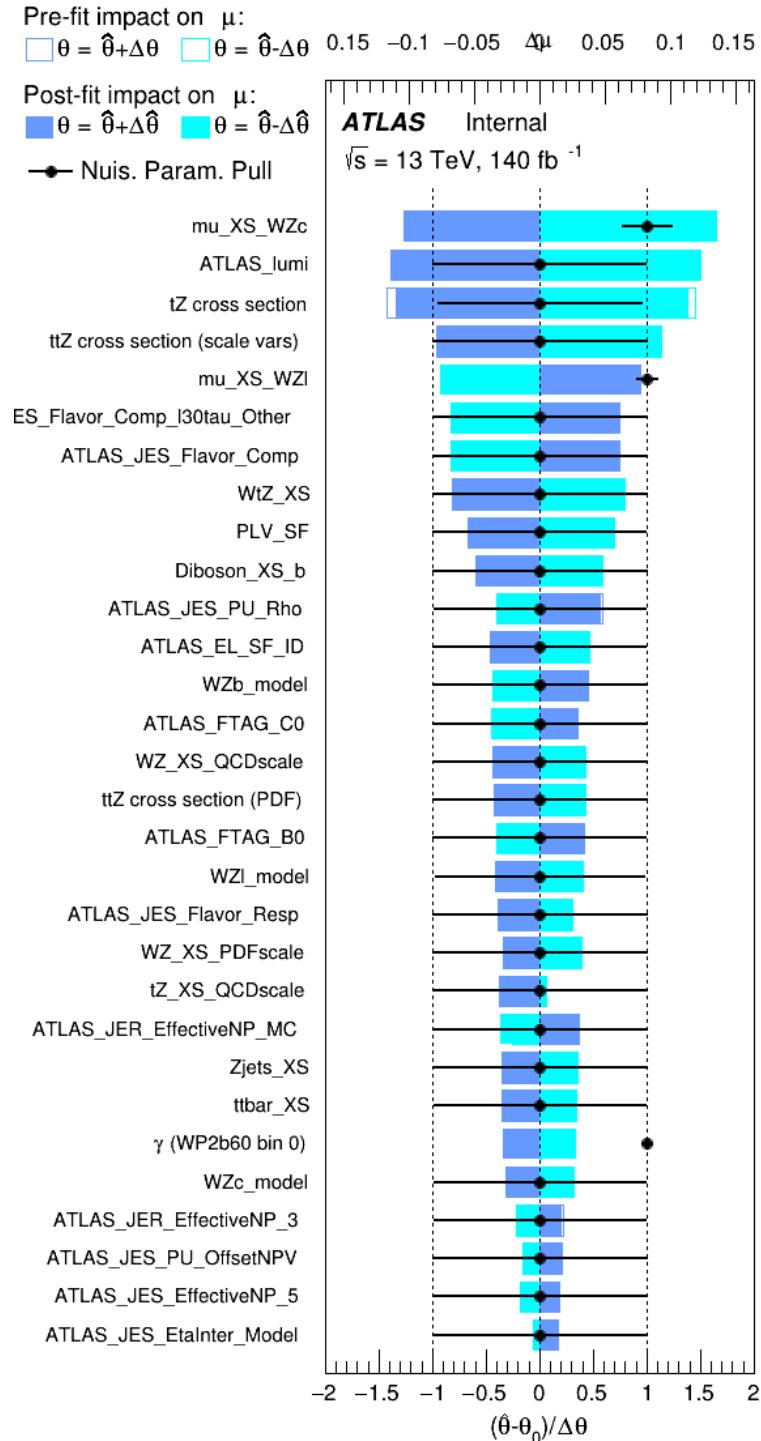


Figure 15.8: Impact of systematic uncertainties on the signal-strength of  $WZ + b$  in 2-jet events.

1094 The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the  
 1095 shape of the fit regions depends heavily on the modeling of the jets. The other major sources  
 1096 of uncertainty come from background modelling and cross-section uncertainty. The pie charts  
 1097 in Figure 15.9 show that for the modelling uncertainties that contribute most correspond to the  
 1098 most significant backgrounds.

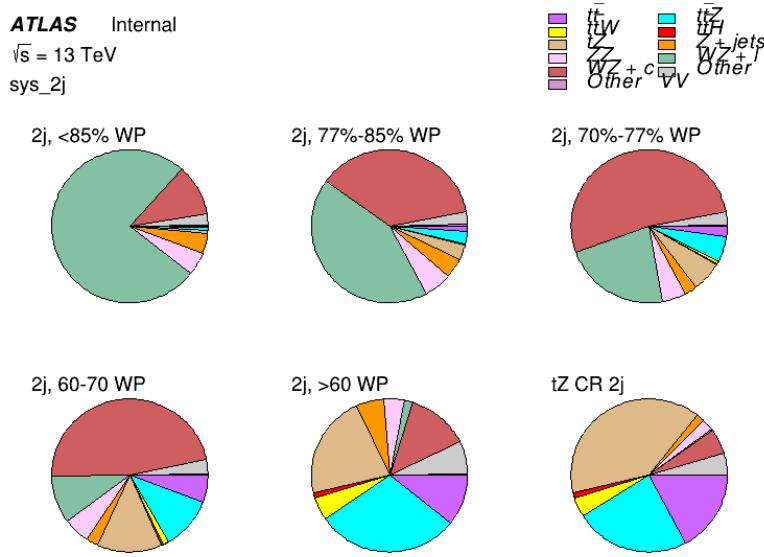


Figure 15.9: Post-fit background composition of the 2-jet fit regions.

1099 The correlations between these nuisance parameters are summarized in Figure 15.10.

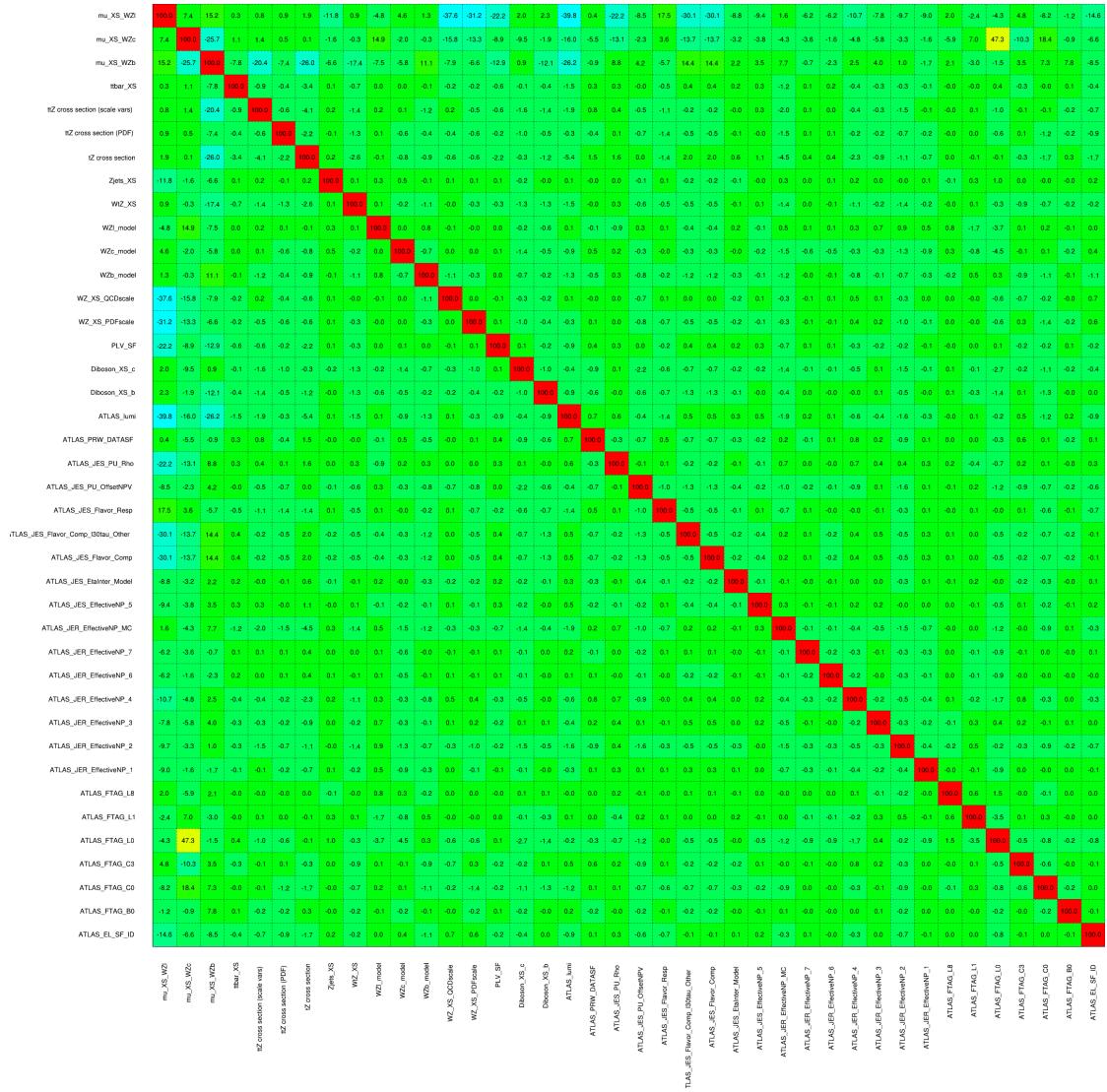


Figure 15.10: Correlations between nuisance parameters in the 2-jet fit

1100 As in the 1-jet case, no significant, unexpected correlations are found between nuisance  
parameters.

---

**Part V****Differential Studies of  $t\bar{t}H$  Multilepton****16 Data and Monte Carlo Samples**

1105 For both data and Monte Carlo (MC) simulations, samples were prepared in the `xAOD` format,  
1106 which was used to produce a `xAOD` based on the `HIGG8D1` derivation framework. This framework  
1107 was designed for the main  $t\bar{t}H$  multi-lepton analysis. Because this analysis targets events with  
1108 multiple light leptons, as well as tau hadrons, this framework skims the dataset of any events that  
1109 do not meet at least one of the following requirements:

- 1110 • at least two light leptons within a range  $|\eta| < 2.6$ , with leading lepton  $p_T > 15$  GeV and  
1111 subleading lepton  $p_T > 5$  GeV
- 1112 • at least one light lepton with  $p_T > 15$  GeV within a range  $|\eta| < 2.6$ , and at least two hadronic  
1113 taus with  $p_T > 15$  GeV.

1114 Samples were then generated from these `HIGG8D1` derivations using `AnalysisBase` version  
1115 21.2.127. A ptag of `p4133` was used for MC samples, and `p4134` for data.

---

## 1116 16.1 Data Samples

1117 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through  
 1118 2018, which represents an integrated luminosity of  $139 \text{ fb}^{-1}$  and an energy of  $\sqrt{s} = 13 \text{ TeV}$ . All  
 1119 data used in this analysis was included in one of the following Good Run Lists:

- 1120 • data15\_13TeV.periodAllYear\_DetStatus-v79-repro20-02\_DQDefects-00-02-02  
   1121   \_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- 1122 • data16\_13TeV.periodAllYear\_DetStatus-v88-pro20-21\_DQDefects-00-02-04  
   1123   \_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- 1124 • data17\_13TeV.periodAllYear\_DetStatus-v97-pro21-13\_Unknown\_PHYS\_StandardGRL  
   1125   \_All\_Good\_25ns\_Triggerno17e33prim.xml
- 1126 • data18\_13TeV.periodAllYear\_DetStatus-v102-pro22-04\_Unknown\_PHYS\_StandardGRL  
   1127   \_All\_Good\_25ns\_Triggerno17e33prim.xml

## 1128 16.2 Monte Carlo Samples

1129 Several Monte Carlo (MC) generators were used to simulate both signal and background pro-  
 1130 cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific  
 1131 event generator used for each of these MC samples is listed in Table 22. A Higgs mass of 125  
 1132 GeV is assumed in all simulations.

Table 22: 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 [16]
	(MG5_AMC)	(NLO)	(HERWIG++)	(CT10 [17])
$t\bar{t}W$	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 [18]	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
$t\bar{t}t, t\bar{t}\bar{t}$	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 qqVV, VVV $Z \rightarrow l^+l^-$	POWHEG-BOX v1 [19]	NLO	PYTHIA 6	CT10
	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

1133        The signal sample ( $t\bar{t}H$ ) is modelled at NLO with POWHEG-BOX v2 using the NNPDF2.0  
 1134        parton distribution function (PDF) [35]. Parton showering and hadronisation were modelled  
 1135        with PYTHIA 8.2 [36]. The  $t\bar{t}H$  sample is normalized to a cross-section of  $507^{+35}_{-50}$  fb based on  
 1136        NLO calculations. Uncertainties are based on varying the QCD factorisation and renormalisation  
 1137        scale, as well as uncertainties in the PDF and  $\alpha_s$ .

1138        The  $t\bar{t}W$  background is simulated using Sherpa 2.2.1 with the NNPDF3.0 NLO PDF. The  
 1139        matrix element is calculated with up to one additional parton at NLO, and up to two at LO. As  
 1140        explained in detail in [7], the  $t\bar{t}W$  contribution predicted by MC is found disagree significantly

1141 with what is observed in data. While an effort is currently being undertaken to measure  $t\bar{t}W$  more  
1142 accurately, the approach used by the  $79.8 \text{ fb}^{-1}$   $t\bar{t}H$  analysis is used here: A normalization  
1143 factor of 1.68 is applied to the MC estimate of  $t\bar{t}W$  and additional systematic uncertainties on  
1144  $t\bar{t}W$  are included to account for this modelling discrepancy, as outlined in Section 20.

1145 The  $t\bar{t}(Z/\gamma^*)$  process is simulated with the **MADGRAPH5\_AMC@NLO** generator, using  
1146 NNPDF3.0. Diboson processes are generated with **SHERPA 2.2.2** at NLO precision for one extra  
1147 parton, and at LO for up to three extra partons.

1148 The estimation of the “fake” or non-prompt background - with leptons from hadron decays  
1149 or photon conversions - is done primarily using an inclusive  $t\bar{t}$  sample. This sample is generated  
1150 using **POWHEG**, with **PYTHIA8** performing the parton shower and fragmentation.

1151 While the main  $t\bar{t}H$  analysis is currently refining a data-driven approach for estimating  
1152 the contribution of events with non-prompt leptons, at the time of this note this strategy has not  
1153 been completely developed for the full Run-2 dataset. Therefore, the non-prompt contribution  
1154 is estimated with MC, while applying normalization corrections and systematic uncertainties  
1155 derived from data driven techniques developed for the  $79.8 \text{ fb}^{-1}$   $t\bar{t}H/t\bar{t}W$  analysis [7]. The  
1156 primary contribution to the non-prompt lepton background is from  $t\bar{t}$  production, with V+jets  
1157 and single-top as much smaller sources. Likelihood fits over several control regions enriched with  
1158 these non-prompt backgrounds are fit to data in order to derive normalization factors for these  
1159 backgrounds. The specific normalization factors and uncertainties applied to the non-prompt  
1160 contributions are listed in Section 20.

1161 Other background processes, such as  $tH$ ,  $tZ$ ,  $t\bar{t}WW$  and  $t\bar{t}t\bar{t}$ , are expected to make minor  
1162 contributions to the total background. The generators and setting used for these backgrounds are  
1163 summarized in Table 22.

1164 The specific DSIDs used in the analysis are listed below:

Sample	DSID
$t\bar{t}H$	345873-5, 346343-5
$VV$	364250-364254, 364255, 363355-60, 364890
$t\bar{t}W$	413008
$t\bar{t}Z$	410156, 410157, 410218-20
low mass $t\bar{t}Z$	410276-8
Rare Top	410397, 410398, 410399
single Top	410658-9, 410644-5
three Top	304014
four Top	410080
$t\bar{t}WW$	410081
$Z + \text{jets}$	364100-41
low mass $Z + \text{jets}$	364198-215
$W + \text{jets}$	364156-97
$V\gamma$	364500-35
$tZ$	410560
$tW$	410013-4
$WtZ$	410408
$VVV$	364242-9
$VH$	342284-5
$WtH$	341998
$t\bar{t}\gamma$	410389
$t\bar{t}$	410470

Table 23: List of Monte Carlo samples by data set ID used in the analysis.

## 1165 17 Object Reconstruction

1166 All analysis channels considered in this note share a common object selection for leptons and  
 1167 jets, as well as a shared trigger selection.

### 1168 17.1 Trigger Requirements

1169 Events are required to be selected by dilepton triggers, as summarized in Table 24.

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

Table 24: List of lowest  $p_T$ -threshold, un-prescaled dilepton triggers used for 2015-2018 data taking.

1170 **17.2 Light Leptons**

1171 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter  
 1172 that are associated with charged particle tracks reconstructed in the inner detector [20]. Electron  
 1173 candidates are required to have  $p_T > 10 \text{ GeV}$  and  $|\eta_{\text{cluster}}| < 2.47$ . Candidates in the transition  
 1174 region between different electromagnetic calorimeter components,  $1.37 < |\eta_{\text{cluster}}| < 1.52$ , are  
 1175 rejected. A multivariate likelihood discriminant combining shower shape and track information  
 1176 is used to distinguish prompt electrons from nonprompt leptons, such as those originating from  
 1177 hadronic showers. Electron candidate are also required to pass TightLH identification.

1178 To further reduce the non-prompt contribution, the track of each electron is required to  
 1179 originate from the primary vertex; requirements are imposed on the transverse impact parameter  
 1180 significance ( $|d_0|/\sigma_{d_0}$ ) and the longitudinal impact parameter ( $|\Delta z_0 \sin \theta_\ell|$ ).

1181 Muon candidates are reconstructed by combining inner detector tracks with track segments  
 1182 or full tracks in the muon spectrometer [21]. Muon candidates are required to have  $p_T > 10 \text{ GeV}$   
 1183 and  $|\eta| < 2.5$ . Muons are required to Medium ID requirements.

1184 All leptons are required to pass a non-prompt BDT selection developed by the main  
 1185  $t\bar{t}H/t\bar{t}W$  analysis, described in detail in [7]. Optimized working points and scale factors for this  
 1186 BDT are taken from that analysis. This BDT and the WPs used are summarized in Appendix  
 1187 [A](#),

---

**1188 17.3 Jets**

1189 Jets are reconstructed from calibrated topological clusters built from energy deposits in the  
1190 calorimeters [23], using the anti- $k_t$  algorithm with a radius parameter  $R = 0.4$ . Particle Flow, or  
1191 PFlow, jets are used in the analysis, which are hadronic objects reconstructed using information  
1192 from both the tracker and the calorimeter. Jets with energy contributions likely arising from noise  
1193 or detector effects are removed from consideration [24], and only jets satisfying  $p_T > 25$  GeV  
1194 and  $|\eta| < 2.5$  are used in this analysis. For jets with  $p_T < 60$  GeV and  $|\eta| < 2.4$ , a jet-track  
1195 association algorithm is used to confirm that the jet originates from the selected primary vertex,  
1196 in order to reject jets arising from pileup collisions [25].

---

**1197 17.4 B-tagged Jets**

1198 Each analysis channel used in this analysis includes b-jets in the final state. These are identified  
1199 using the DL1r b-tagging algorithm, which uses jet vertex and kinematic information to distin-  
1200 guish heavy and light flavored jets. These features are used as inputs to a neural network, the  
1201 output of which is used to form calibrated working points (WPs) based on how likely a jet is to  
1202 have originated from a b-quark. This analysis uses the 70% DL1r WP - implying an efficiency of  
1203 70% for truth b-jets - for selecting b-tagged jets.

---

**1204 17.5 Missing Transverse Energy**

1205 Because all  $t\bar{t}H - ML$  channels considered include multiple neutrinos, missing transverse  
1206 energy ( $E_T^{\text{miss}}$ ) is present in each event. The missing transverse momentum vector is defined as  
1207 the inverse of the sum of the transverse momenta of all reconstructed physics objects as well  
1208 as remaining unclustered energy, the latter of which is estimated from low- $p_T$  tracks associated  
1209 with the primary vertex but not assigned to a hard object [26].

**1210 17.6 Overlap removal**

1211 To avoid double counting objects and remove leptons originating from decays of hadrons, overlap  
1212 removal is performed in the following order: any electron candidate within  $\Delta R = 0.1$  of another  
1213 electron candidate with higher  $p_T$  is removed; any electron candidate within  $\Delta R = 0.1$  of a muon  
1214 candidate is removed; any jet within  $\Delta R = 0.3$  of an electron candidate is removed; if a muon  
1215 candidate and a jet lie within  $\Delta R = \min(0.4, 0.04 + 10[\text{GeV}] / p_T(\text{muon}))$  of each other, the jet  
1216 is kept and the muon is removed.

1217 This algorithm is applied to the preselected objects. The overlap removal procedure is  
1218 summarized in Table 25.

<b>Keep</b>	<b>Remove</b>	<b>Cone size (<math>\Delta R</math>)</b>
electron	electron (low $p_T$ )	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	$\min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$
electron	tau	0.2

Table 25: Summary of the overlap removal procedure between electrons, muons, and jets.

## 1219 18 Higgs Momentum Reconstruction

1220 Reconstructing the momentum of the Higgs boson is a particular challenge for channels with  
 1221 leptons in the final state: Because all channels include at least two neutrinos in the final state, the  
 1222 Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted  
 1223 by a neural network when provided with the kinematics of the Higgs Boson decay products - as  
 1224 verified by studies detailed in Appendix C.3. With this in mind, several layers of MVAs are used  
 1225 to reconstruction the Higgs momentum:

1226 The first layer is a model designed to select which jets are most likely to be the b-jets that  
 1227 came from the top decay, detailed in Section 18.3. As described in Section 18.4, the kinematics  
 1228 of these jets and possible Higgs decay products are fed into the second layer, which is designed  
 1229 to identify the decay products of the Higgs Boson itself. The kinematics of the particles this  
 1230 layer identifies as most likely to have originated from the Higgs decay are then fed into yet  
 1231 another neural-network, which predicts the momentum of the Higgs (18.5). For the 3l channel,  
 1232 an additional MVA is used to determine the decay mode of the Higgs boson in the 3l channel  
 1233 (18.6).

1234 Models are trained on Monte Carlo simulations of  $t\bar{t}H$  events generated using MG5\_AMC.

1235 Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training in order to

1236 increase the statistics of the training sample.

1237 For all of these models, the Keras neural network framework, with Tensorflow 2.0 as the

1238 backend, is used, and the number of hidden layers and nodes are determined using grid search

1239 optimization. Each neural network uses the LeakyReLU activation function, a learning rate of

1240 0.01, and the Adam optimization algorithm, as alternatives are found to either decrease or have

1241 no impact on performance. Batch normalization is applied after each layer in order to stabilize

1242 the model and decrease training time. For the classification algorithms (b-jet matching, Higgs

1243 reconstruction, and 3l decay identification) binary-cross entropy is used as the loss function,

1244 while the  $p_T$  reconstruction algorithm uses MSE.

1245 The specific inputs features used for each model are arrived at through a process of trial

1246 and error - features considered potentially useful are tried, and those that are found to increase

1247 performance are included. While each model includes a relatively large number of features,

1248 some using upwards of 30, this inclusive approach is found to maximize the performance of

1249 each model while decreasing the variance compared to a reduced number of inputs. Each input

1250 feature is validated by comparing MC simulations to  $79.8 \text{ fb}^{-1}$  of data, as shown in the sections

1251 below.

---

**1252 18.1 Physics Object Truth Matching**

1253 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson using  
1254 MC simulations of  $t\bar{t}H$  events. The kinematics of the reconstructed physics objects, as well as  
1255 event level variables, are used as inputs, with the parent ID taken from the truth record used to  
1256 label the data. The objects considered include light leptons and jets.

1257 Reconstructed physics objects are matched to particle level objects, in order to identify the  
1258 parent particle of these reconstructed objects. Reconstructed jets are matched to truth jets based  
1259 on the requirements that the reco jet and truth jet fall within  $\Delta R < 0.4$ , and the two objects have  
1260 a  $p_T$  that agrees within 10%. Truth level and reco level leptons are required to have the same  
1261 flavor, a  $\Delta R < 0.1$ , and  $p_T$  that agree within 10%. Events where no match can be found between  
1262 the particle level decay products of the Higgs and the reconstructed objects are not included in  
1263 training.

1264 Leptons considered as possible Higgs and top decay candidates are required to pass the  
1265 selection described in Section 17.2. For jets, however, it is found that a large fraction that  
1266 originate from either the top decay or the Higgs decay fall outside the selection described in  
1267 Section 17.3. Specifically, jets from the Higgs decay tend to be soft, with 32% having  $p_T <$   
1268 25 GeV. Therefore jets with  $p_T < 15$  GeV are considered as possible candidates in the models  
1269 described below. By contrast, less than 5% of the jets originating from the Higgs fall below this  
1270  $p_T$  threshold. The jets are found to be well modeled even down to this low  $p_T$  threshold, as shown  
1271 in Section 19.1. The impact of using different  $p_T$  selection for the jet candidates is considered

1272 in detail in Section C.6. As they are expected to originate from the primary vertex, jets are also  
1273 required to pass a JVT cut.

## 1274 **18.2 Truth Level Studies**

1275 Attempts to identify the decay products of the Higgs are motivated by the ability to reconstruct  
1276 the Higgs momentum based on their kinematics. In order to demonstrate that this is case, the  
1277 kinematics of reconstructed objects that are truth matched to the Higgs decay are used as inputs  
1278 to a neural network which is designed to predict of the momentum of the Higgs. This is done in  
1279 the 2lSS channel -

## 1280 **18.3 b-jet Identification**

1281 Including the kinematics of the b-jets that originate from the top decay is found to improve the  
1282 identification of the Higgs decay products, and improve the accuracy with which the Higgs  
1283 momentum can be reconstructed. Because these b-jets are reconstructed by the detector with  
1284 high efficiency (just over 90% of the time), and can be identified relatively consistently, the first  
1285 step in reconstructing the Higgs is selecting the b-jets from the top decay.

1286 Exactly two b-jets are expected in the final state of  $t\bar{t}H$  – ML events. However, in both  
1287 the 3l and 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r b-tag  
1288 working point is used). Therefore, for events which have exactly one, or more than two, b-tagged  
1289 jets, deciding which combination of jets correspond to the top decay is non-trivial. Further,

1290 events with 1 b-tagged jet represent just over half of all  $t\bar{t}H - ML$  events. Of those, both b-jets  
 1291 are reconstructed by the detector 75% of the time. Therefore, rather than adjusting the selection  
 1292 to require exactly 2 b-tagged jets, and losing more than half of the signal events, a neural network  
 1293 is used to predict which pair of jets is most likely to correspond to truth b-jets.

1294 Once the network is trained, all possible pairings of jets are fed into the model, and the pair  
 1295 of jets with the highest output score are taken to be b-jets in successive steps of the analysis.

### 1296 18.3.1 2ISS Channel

1297 For the 2ISS channel, the input features shown in Table 26 are used for training. Here  $j_0$  and  $j_1$   
 1298 are the two jet candidates, while  $l_0$  and  $l_1$  are the two leptons in the event, both ordered by  $p_T$ . jet  
 1299 DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,  
 1300 where 5 represents the tightest working point and 1 represents the loosest. The variables nJets  
 1301 DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and  
 1302 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
$\Delta R(j_0)(j_1)$	$M(j_0j_1)$	$\Delta R(l_0)(j_0)$
$\Delta R(l_0)(j_1)$	$\Delta R(l_1)(j_0)$	$\Delta R(l_1)(j_1)$
$M(l_0j_0)$	$M(l_0j_1)$	$M(l_1j_0)$
$M(l_1j_1)$	jet DL1r 0	jet DL1r 1
nJets OR DL1r 85	nJets OR DL1r 60	$\Delta R(j_0l_0)(j_1l_1)$
$\Delta R(j_0l_1)(j_1l_0)$	$p_T(j_0j_1l_0l_1 E_T^{\text{miss}})$	$M(j_0j_1l_0l_1 E_T^{\text{miss}})$
$\Delta\phi(j_0)(E_T^{\text{miss}})$	$\Delta\phi(j_1)(E_T^{\text{miss}})$	HT jets
nJets	$E_T^{\text{miss}}$	

Table 26: Input features used in the b-jet identification algorithm for the 2ISS channel

1303 As there are far more incorrect combinations than correct ones, by a factor of more than  
 1304 20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random  
 1305 sample of 5 million incorrect entries are used for training, along with close 1 million correct  
 1306 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for  
 1307 training.

1308 The difference between the distributions for a few of these features for the "correct" (i.e.  
 1309 both jets are truth b-jets), and "incorrect" combinations are shown in Figure 18.1. The correct and  
 1310 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences  
 1311 in the distributions.

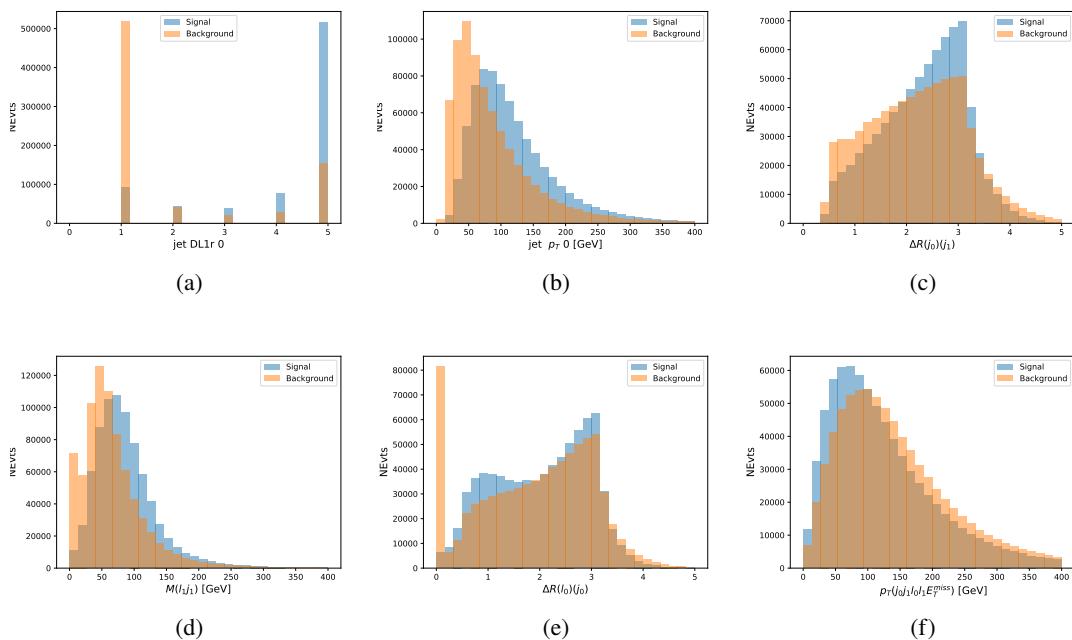


Figure 18.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the  $p_T$  of the leading jet, (c) shows the  $\Delta R$  of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the  $\Delta R$  of lepton 0 and jet 0, and (f) the  $p_T$  of both jets, both leptons, and the  $E_T^{\text{miss}}$ .

1312        The modeling of these inputs is validated against data, with Figure 18.2 showing good  
 1313        general agreement between data and MC. Plots for the complete list of features can found in  
 1314        Section C.

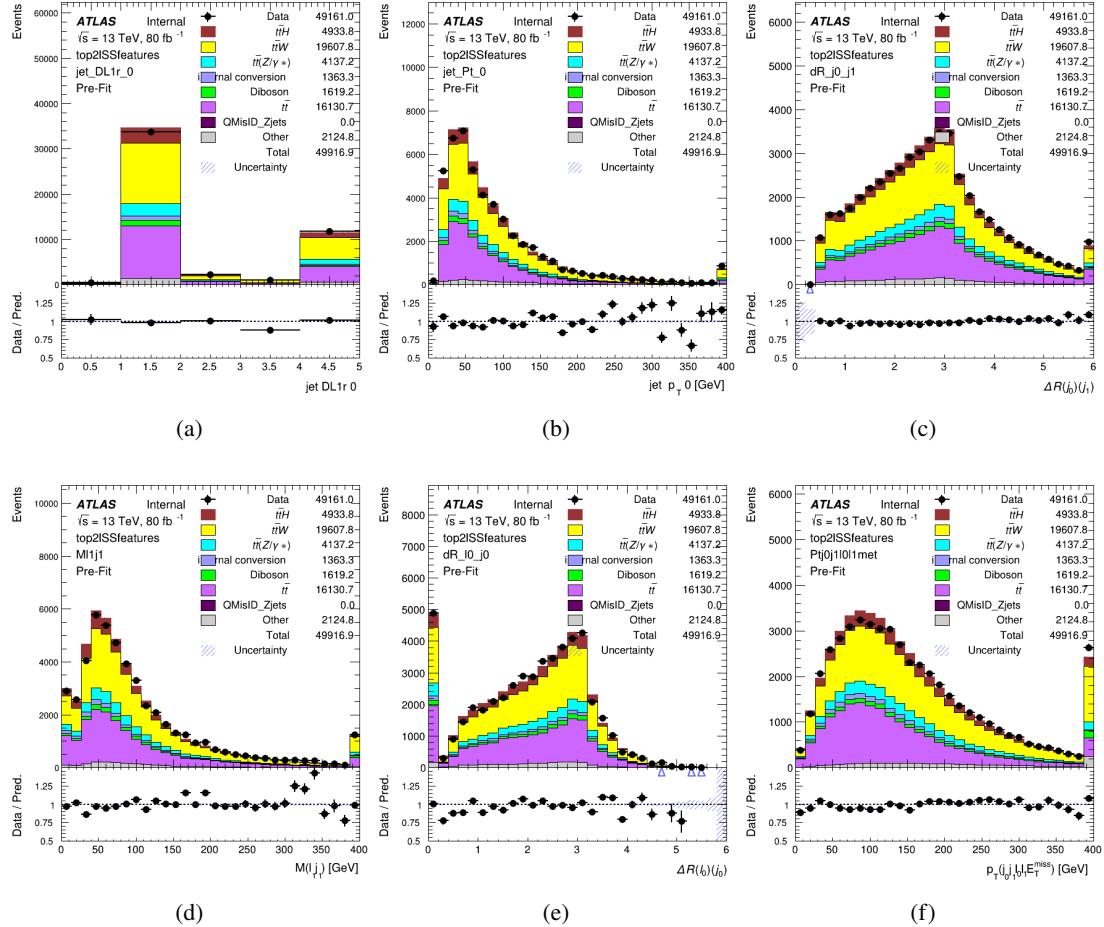


Figure 18.2: Data/MC comparisons of input features for top2ISS training for  $79.8 \text{ fb}^{-1}$  of data. (a) shows the DL1r working point of leading jet, (b) shows the  $p_T$  of the leading jet, (c) shows the  $\Delta R$  of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the  $\Delta R$  of lepton 0 and jet 0, and (f) the  $p_T$  of both jets, both leptons, and the  $E_T^{\text{miss}}$ .

1315        Based on the results of grid search evaluation, the optimal architecture is found to include  
 1316        5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as

1317 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for  
 1318 the trained model are shown in Figure 18.3.1. The model is found to identify the correct pairing  
 1319 of jets for 73% of 2lSS signal events on test data.

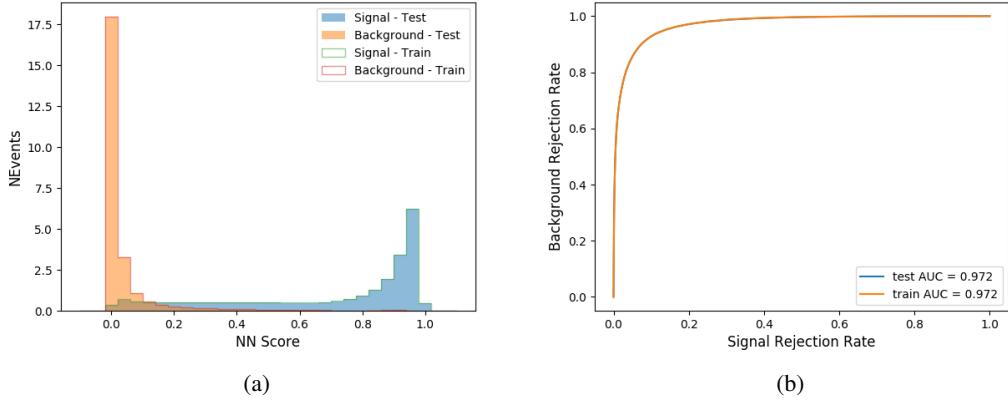


Figure 18.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1320 For point of comparison, a "naive" approach to identify b-jets is used as well: The two  
 1321 jets which pass the highest DL1r b-tag working point are assumed to be the b-jets from the top  
 1322 decay. In the case that multiple jets meet the same b-tag working point, the jet with higher  $p_T$  is  
 1323 used. This method identifies the correct jet pair 65% of the time.

1324 The accuracy of the model for different values of n-bjets, compared to this naive approach,  
 1325 is shown in Table 27.

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 27: Accuracy of the NN in identifying b-jets from tops in 2ISS 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.

1326 **18.3.2 3l Channel**

1327 The input features used in the 3l channel are listed in Table 28, with the same naming convention  
 1328 as the 2ISS 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 28: Input features for the b-jet identification algorithm in the 3l channel.

1329 A few of these features are shown in Figure 18.4, comparing the distributions for correct  
 1330 and incorrect combinations of jets.

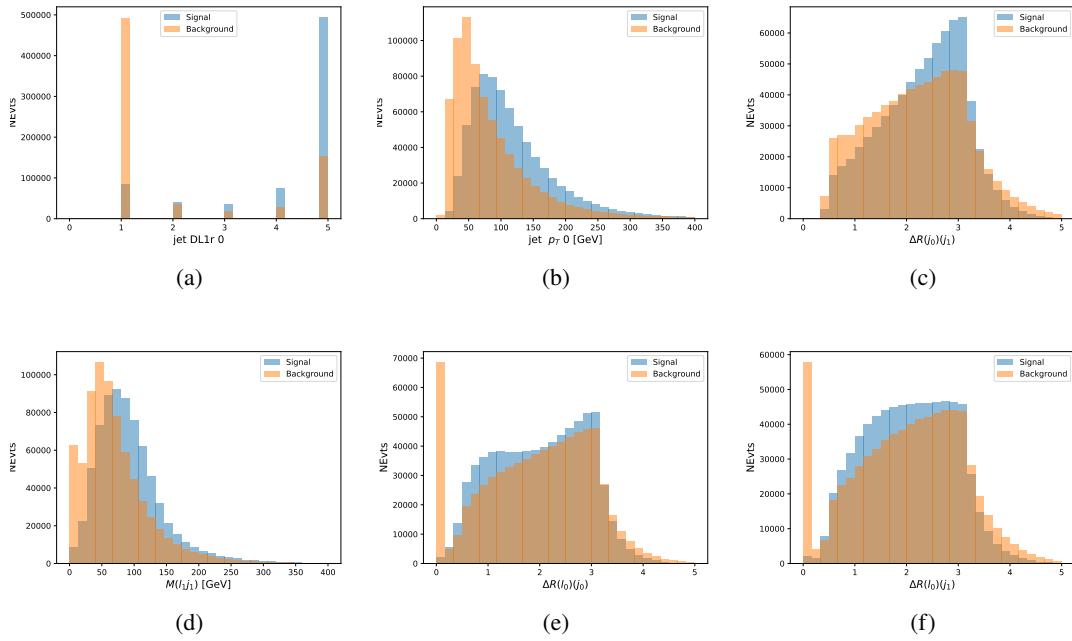


Figure 18.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the  $p_T$  of jet 0, (c)  $\Delta R$  between 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

1331        The modeling of these inputs is validated against data, with Figure 18.5 showing good  
 1332        general agreement between data and MC. Plots for the complete list of features can found in  
 1333        Section C.

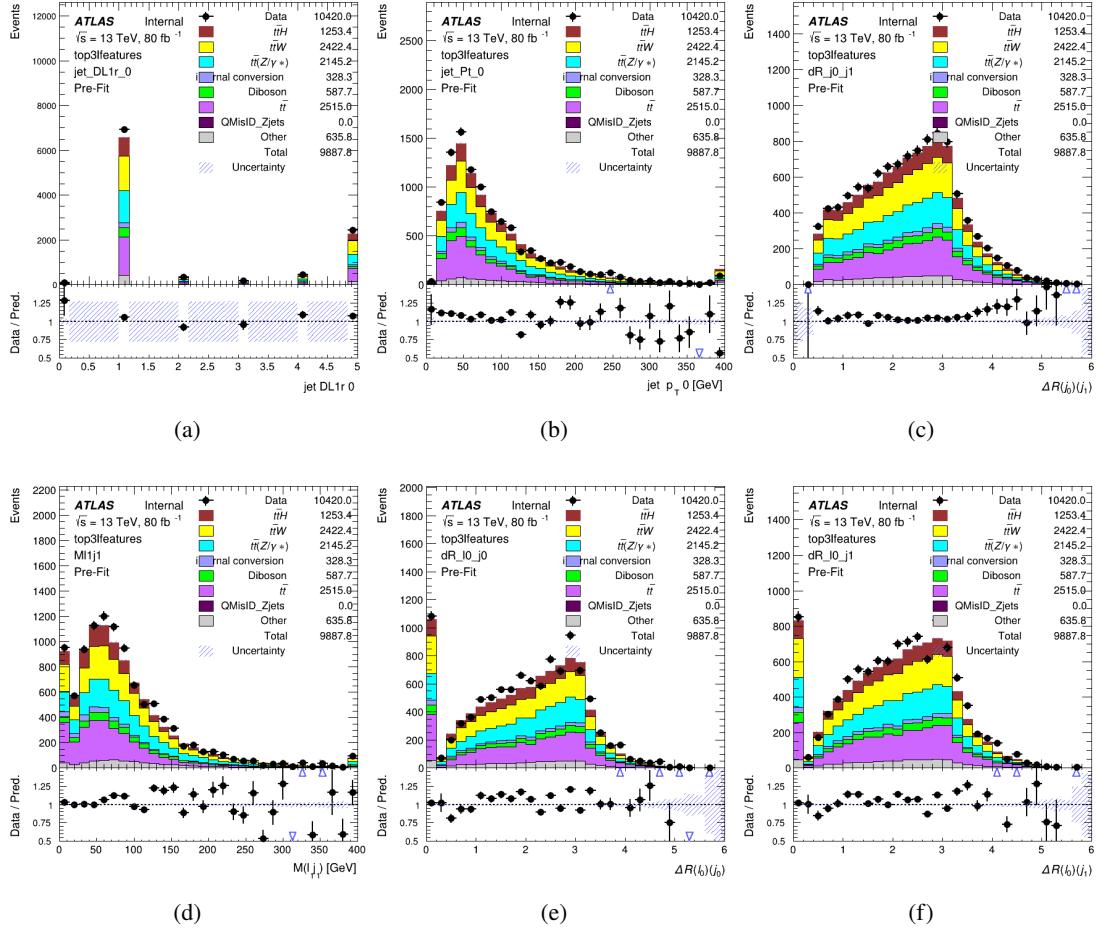


Figure 18.5: Data/MC comparisons of input features for top3l training for  $79.8 \text{ fb}^{-1}$  of data. (a) show the DL1r WP of jet 0, (b) the  $p_T$  of jet 0, (c)  $\Delta R$  between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e)  $\Delta R$  between lepton 0 and jet 0, and (f)  $\Delta R$  between lepton 0 and jet 1

Again, the dataset is downsized to reduce the ratio of correct and incorrect combination from 20:1, to 5:1. Around 7 million events are used for training, with 10% set aside for testing. Based on the results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with 60 nodes each. The output score distribution as well as the ROC curve for the trained model are shown in Figure 18.3.2.

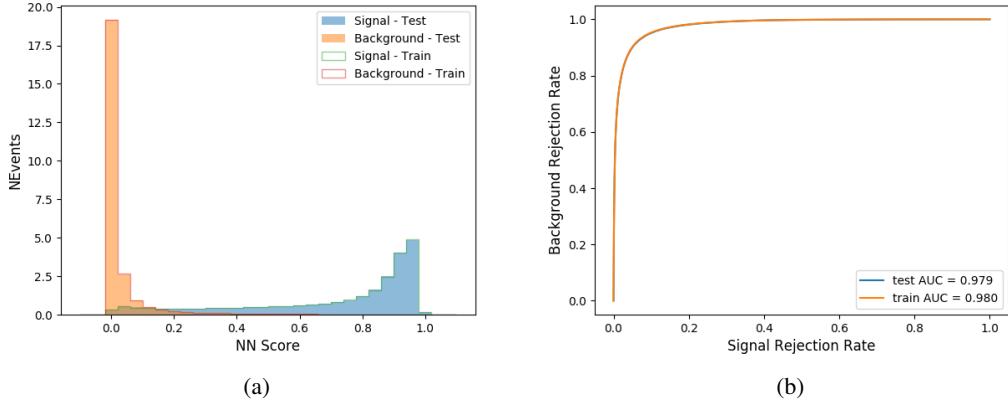


Figure 18.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1339 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal  
 1340 events. The accuracy of the model is summarized in Table 29.

Table 29: 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%

## 1341 18.4 Higgs Reconstruction

1342 Techniques similar to the b-jet identification algorithms are employed to select the decay products  
 1343 of the Higgs: kinematics of all possible combinations of reconstructed objects are fed into a neural  
 1344 network to determine which of those is most likely to be the decay products of the Higgs.

1345 Again separate models are used for the 2lSS and 3l channels, while the 3l channel has now  
 1346 been split into two:  $t\bar{t}H$  events with three leptons in the final state include both instances where  
 1347 the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the Higgs  
 1348 decays to two leptons.

1349 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In  
 1350 the semi-leptonic case the reconstructed decay products consist of two jets and a single lepton.  
 1351 For the fully-leptonic case, the decay products include 2 of the three leptons associated with the  
 1352 event. For training the models, events are separated into these two categories using truth level  
 1353 information. A separate MVA, described in Section 18.6, is used to make this distinction at reco  
 1354 level and determine which model to use.

1355 For all channels, the models described in Section 18.3 are used to identify b-jet candidates,  
 1356 whose kinematics are used to identify the Higgs decay products. These jets are not considered  
 1357 as possible candidates for the Higgs decay, justified by the fact that these models are found to  
 1358 misidentify jets from the Higgs decay as jets from the top decay less than 1% of the time.

#### 1359 **18.4.1 2lSS Channel**

1360 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural  
 1361 network is trained on the kinematics of different combinations of leptons and jets, as well as the  
 1362 b-jets identified in Section 18.3, with the specific input features listed in Table 30.

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	$\Delta R(j_0)(j_1)$
$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$	$M(j_0 j_1)$
$M(l_H j_0)$	$M(l_H j_1)$	$\Delta R(l_H)(b_0)$
$\Delta R(l_H)(b_1)$	$\Delta R(l_T)(b_0)$	$\Delta R(l_T)(b_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1)(l_T)$	$\Delta R(j_0 j_1)(b_0)$
$\Delta R(j_0 j_1)(b_1)$	$\Delta R(j_0)(b_0)$	$\Delta R(j_0)(b_1)$
$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$	$M(j_0 j_1 l_H)$
$p_T(j_0 j_1 l_H l_T b_0 b_1 E_T^{\text{miss}})$	topScore	$E_T^{\text{miss}}$
nJets	HT jets	

Table 30: Input features used to identify the Higgs decay products in 2ISS events

1363 Here  $j_0$  and  $j_1$ , and  $l_H$  are the jet and lepton decay candidates, respectively. The other  
 1364 lepton in the event is labeled  $l_T$ , as it is assumed to have come from the decay of one of the top  
 1365 quarks.  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet identification algorithm. The b-jet  
 1366 Reco Score is the output of the b-jet reconstruction algorithm.

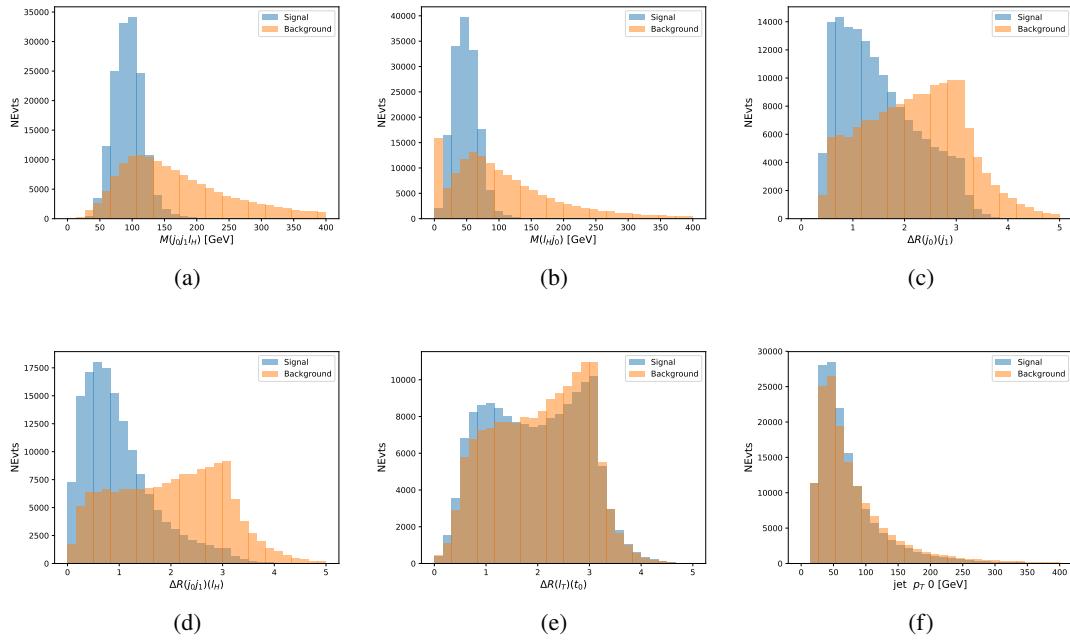


Figure 18.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (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.

1367

The modeling of these inputs is validated against data, with Figure 18.2 showing good

1368

general agreement between data and MC. Plots for the complete list of features can found in

1369

Section C.

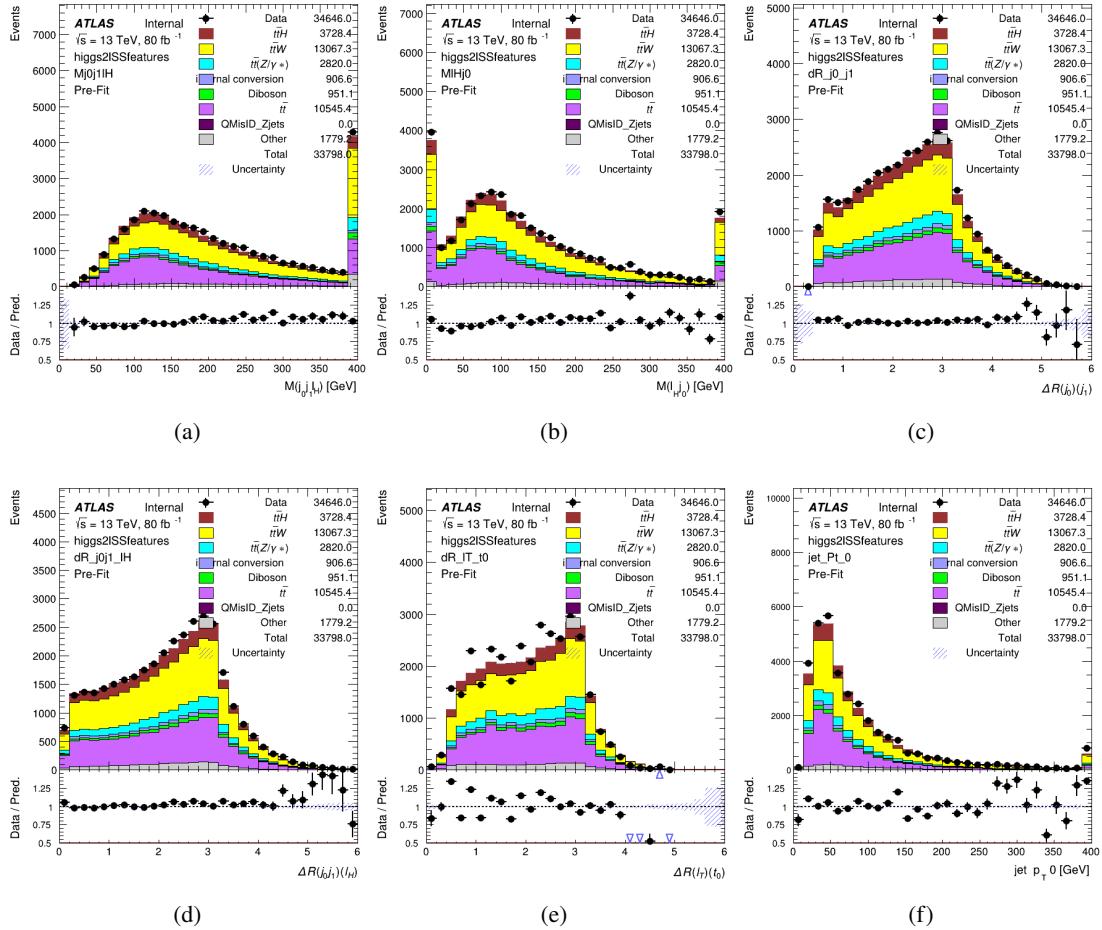


Figure 18.8: Data/MC comparisons of input features for higgs2ISS training for  $79.8 \text{ fb}^{-1}$  of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c)  $\Delta R$  between the jet candidates, (d)  $\Delta R$  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.

1370 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2  
 1371 million events, with an additional 200,000 reserved for testing the model. In order to compensate  
 1372 for large number of incorrect combinations, these have been downsampled such that the correct  
 1373 combinations represent over 10% of the training set. The output of the NN is summarized in  
 1374 Figure 18.4.1.

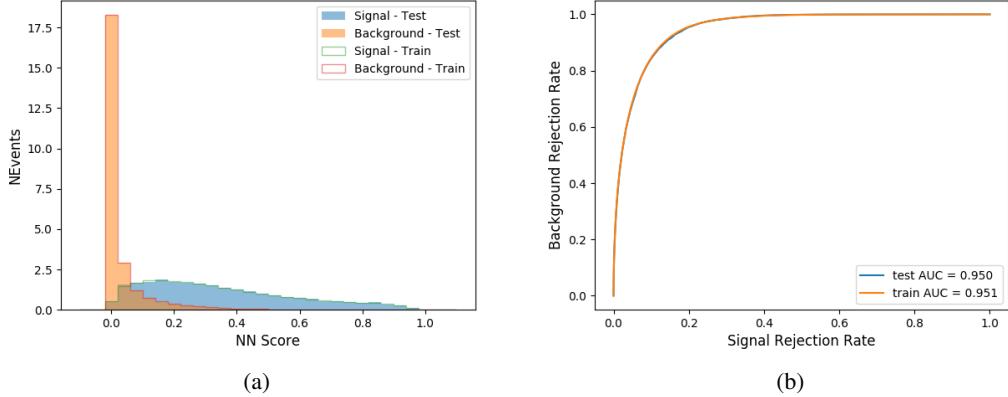


Figure 18.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1375        The neural network identifies the correct combination 55% of the time. It identifies the  
 1376        correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets  
 1377        81% of the time.

#### 1378        18.4.2 3l Semi-leptonic Channel

1379        For 3l  $t\bar{t}H$  where the Higgs decay semi-leptonically, the decay products include one of the three  
 1380        leptons and two jets. In this case, the other two leptons originated from the decay of the tops,  
 1381        meaning the opposite-sign (OS) lepton cannot have come from the Higgs. This leaves only the two  
 1382        same-sign (SS) leptons as possible Higgs decay products.

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

Table 31: Input features used to identify the Higgs decay products in 31 semi-leptonic events

1383 Here  $j_0$  and  $j_1$ , and  $l_H$  are the jet and lepton decay candidates, respectively. The other  
 1384 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  
 1385 the b-jet identification algorithm. The b-jet Reco Score is the output of the b-jet identification  
 1386 algorithm.

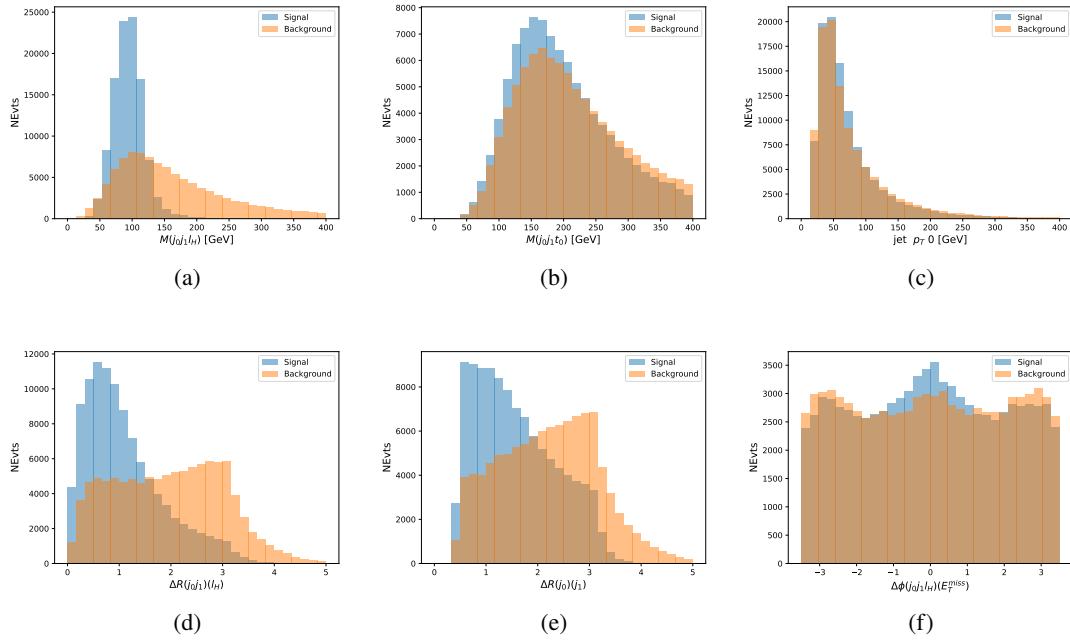


Figure 18.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1387

The modeling of these inputs is validated against data, with Figure 18.11 showing good

1388

general agreement between data and MC. Plots for the complete list of features can found in

1389

appendix C.1.

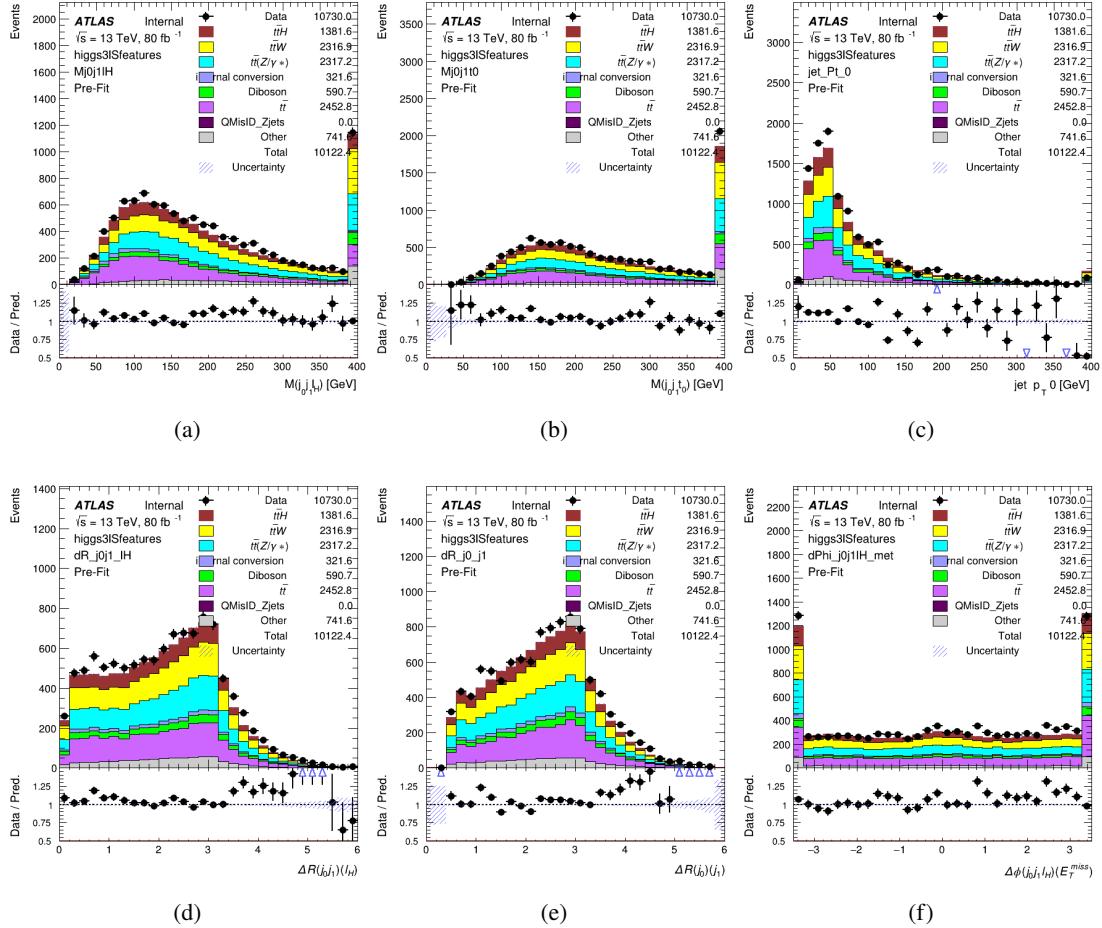


Figure 18.11: Data/MC comparisons of input features for higgs3lS training for  $79.8 \text{ fb}^{-1}$  of data.

1390 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.  
 1391 Once again, incorrect combinations are downsampled, such that the correct combinations are  
 1392 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN  
 1393 is summarized in Figure 18.4.2.

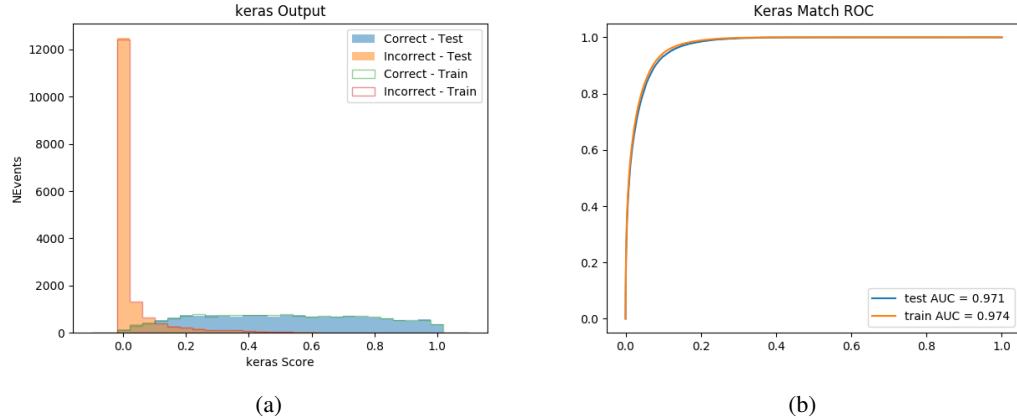


Figure 18.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1394        The neural network identifies the correct combination 64% of the time. It identifies the  
 1395        correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets  
 1396        83% of the time.

### 1397        18.4.3 3l Fully-leptonic Channel

1398        In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from  
 1399        the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to  
 1400        determining which of the two SS leptons originated from the Higgs. The kinematics of both  
 1401        possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,  
 1402        and one where it is not.

Lepton p <sub>T</sub> H <sub>1</sub>	Lepton p <sub>T</sub> H <sub>0</sub>	Lepton p <sub>T</sub> T
top p <sub>T</sub> 0	top p <sub>T</sub> 1	$\Delta\phi(l_{H_1})(E_T^{\text{miss}})$
$\Delta\phi(l_T)(E_T^{\text{miss}})$	$M(l_{H_0}l_{H_1})$	$M(l_{H_1}l_T)$
$M(l_{H_0}l_T)$	$\Delta R(l_{H_0})(l_{H_1})$	$\Delta R(l_{H_1})(l_T)$
$\Delta R(l_{H_0})(l_T)$	$\Delta R(l_{H_0}l_{H_1})(l_T)$	$\Delta R(l_{H_0}l_T)(l_{H_1})$
$\Delta R(l_{H_0}l_{H_1})(b_0)$	$\Delta R(l_{H_0}l_{H_1})(b_1)$	$\Delta R(l_{H_0}b_0)$
$M(l_{H_0}b_0)$	$\Delta R(l_{H_0}b_1)$	$M(l_{H_0}b_1)$
$\Delta R(l_{H_1}b_0)$	$M(l_{H_1}b_0)$	$\Delta R(l_{H_1}b_1)$
$M(l_{H_1}b_1)$	$E_T^{\text{miss}}$	topScore

Table 32: Input features used to identify the Higgs decay products in 3lF events

Table 33: Input features used to identify the Higgs decay products in 3l fully leptonic events

1403 Here  $l_{H_0}$  and  $l_{H_1}$  are the Higgs decay candidates. The other lepton in the event is labeled  
 1404  $l_T$ .  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco  
 1405 Score is the output of the Higgs reconstruction algorithm.

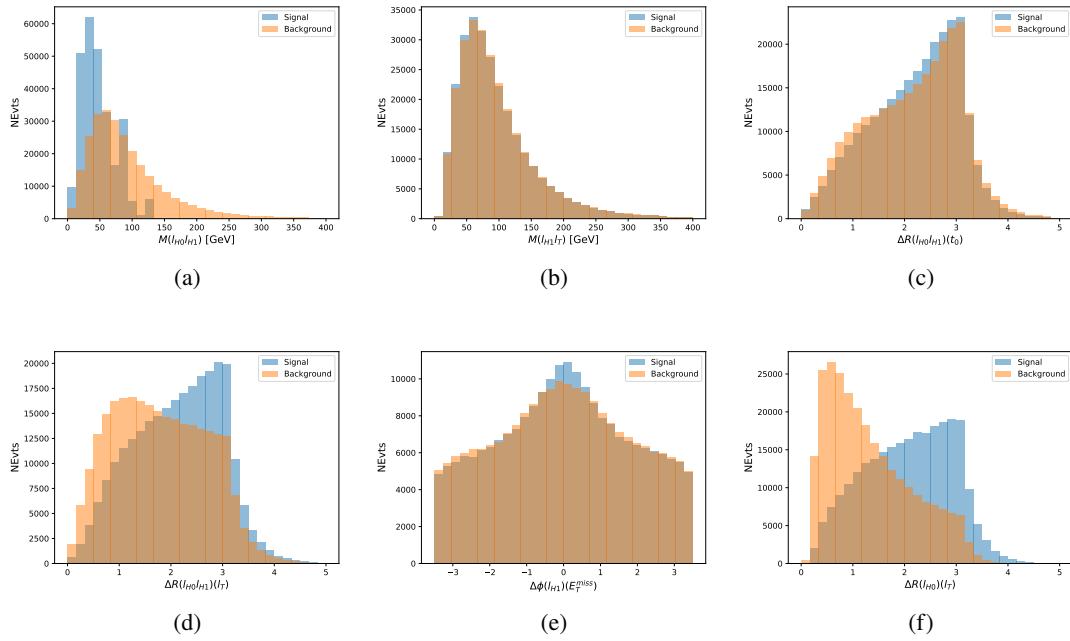


Figure 18.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1406            The modeling of these inputs is validated against data, with Figure 18.14 showing good  
 1407            general agreement between data and MC. Plots for the complete list of features can found in  
 1408            Section C.

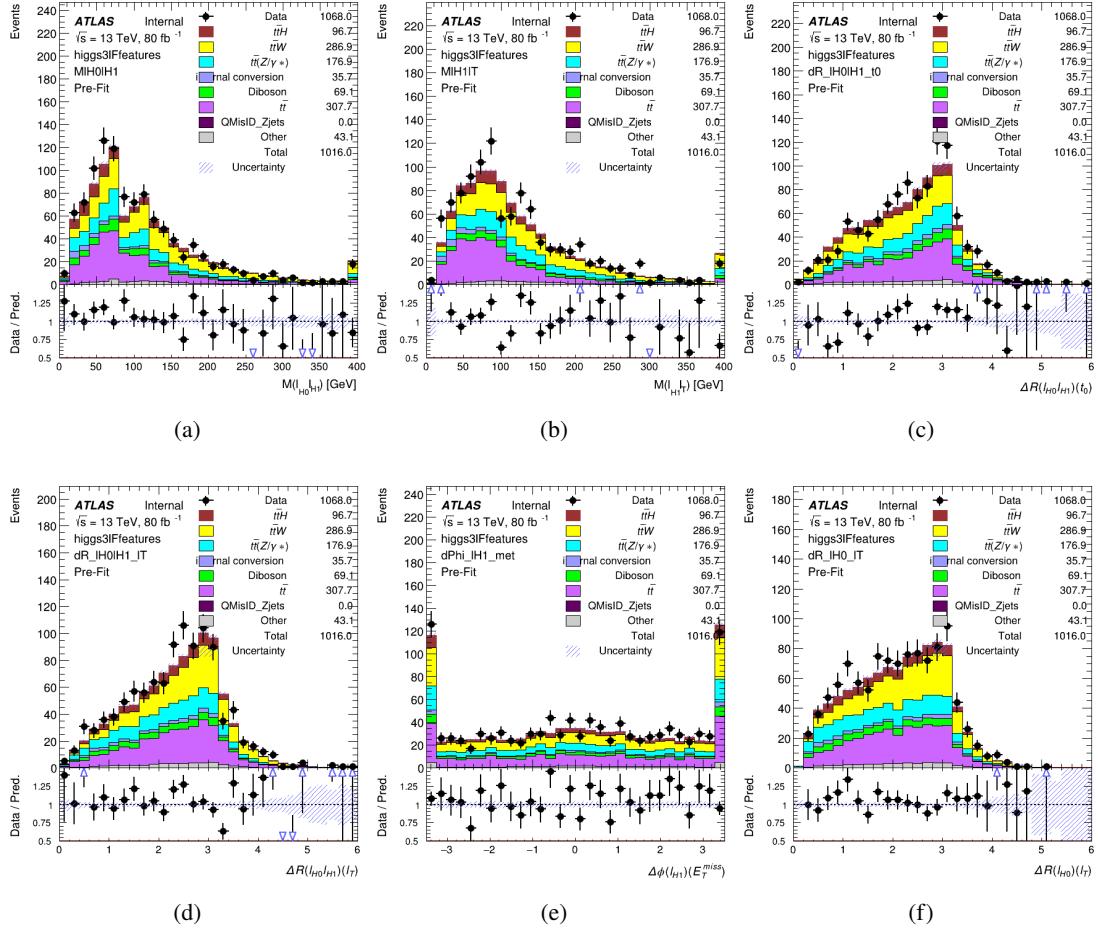


Figure 18.14: Data/MC comparisons of input features for higgs3lF training for  $79.8 \text{ fb}^{-1}$  of data.

1409 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,

1410 with 10% of the dataset reserved for testing. The output of the model is summarized in Figure

1411 **18.4.3.**

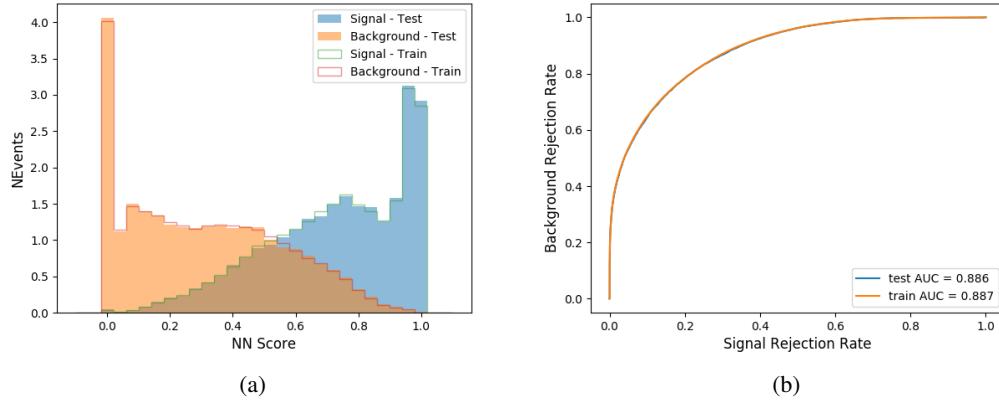


Figure 18.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

1412 The correct lepton is identified by the model for 80% of events in the testing data set.

## 1413 18.5 $p_T$ Prediction

1414 Once the most probable decay products have been identified, their kinematics are used as inputs  
1415 to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,  
1416 a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay  
1417 are included as well, as these are found to improve performance. The truth  $p_T$  of the Higgs,  
1418 as predicted by MC, are used as labels. Separate models are built for each channel - 2lSS, 3l  
1419 Semi-leptonic and 3l Fully-leptonic.

1420 As a two-bin fit is targeted for the final result, some metrics evaluating the performance  
1421 of the models aim to show how well it distinguished between "high  $p_T$ " and "low  $p_T$ " events. A

1422 cutoff point of 150 GeV is used to define these two categories.

1423 Because the analysis uses a two bin fit of the Higgs  $p_T$ , the momentum reconstruction  
1424 could be treated as a binary classification problem, rather than a regression problem. This  
1425 approach is explored in detail in Section C.5, and is found not to provide any significant increase  
1426 in sensitivity. The regression approach is used because it provides more flexibility for future  
1427 analyses, as it is independent of the cutoff between high and low  $p_T$ , as well as the number of  
1428 bins. Further, a regression allows the output of the neural network to be more clearly understood,  
1429 as it can be directly compared to a physics observable.

### 1430 18.5.1 2lSS Channel

1431 The input variables listed in Table 34 are used to predict the Higgs  $p_T$  in the 2lSS channel. Here  
1432  $j_0$  and  $j_1$  are the two jets identified as Higgs decay products. The lepton identified as originating  
1433 from the Higgs is labeled  $l_H$ , while the other lepton is labeled  $l_T$ , as it is assumed to have come  
1434 from the decay of one of the top quarks.  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet  
1435 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs  
1436 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 34: Input features for reconstructing the Higgs  $p_T$  spectrum for 2lSS events

1437        The optimal neural network architecture for this channel is found to consist of 7 hidden  
 1438        layers with 60 nodes each. The input data set includes 1.2 million events, 10% of which is used  
 1439        for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1440        To evaluate the performance of the model, the predicted  $p_T$  spectrum is compared to the  
 1441        truth Higgs  $p_T$  in Figure 18.16. In order to visualize the model performance more clearly, in (a)  
 1442        of that figure, the color of each point is determined by Kernel Density Estimation (KDE). The  
 1443        color shown represents the logarithm of the output from KDE, to counteract the large number of  
 1444        low  $p_T$  events. For that same reason, each column of the histogram shown in (b) of Figure 18.16  
 1445        is normalized to unity. This plot therefore demonstrates what the model predicts for each slice  
 1446        of truth  $p_T$ .

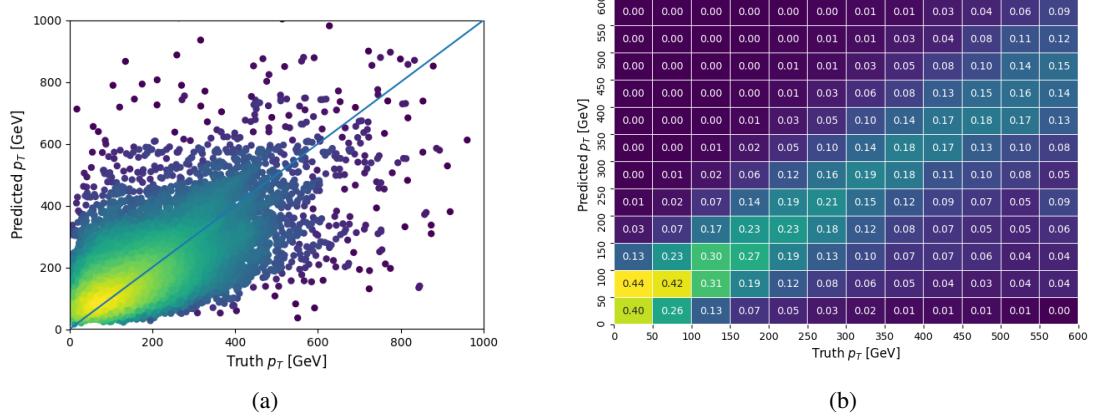


Figure 18.16: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 2lSS  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1447 We are also interested in how well the model distinguishes between events with  $p_T < 150$   
 1448 GeV and  $> 150$  GeV. Figure 18.17 demonstrates the NN output for high and low  $p_T$  events based  
 1449 on this cutoff.

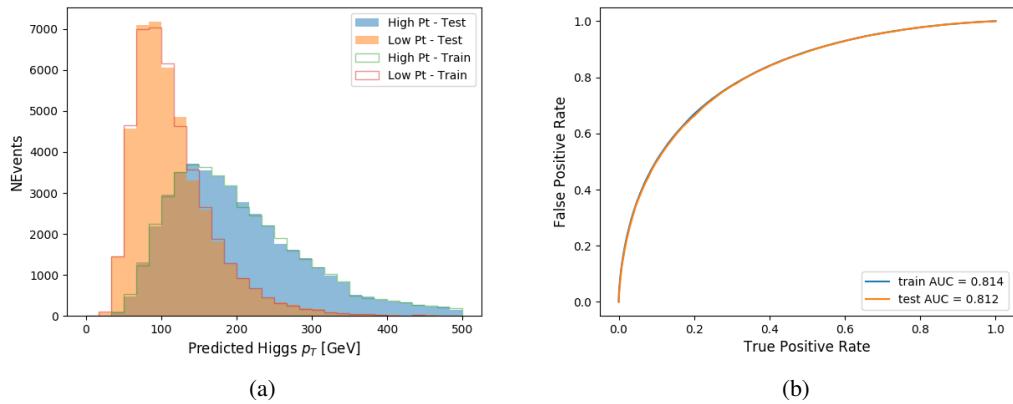


Figure 18.17: (a) shows the reconstructed Higgs  $p_T$  for 2lSS events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

1450 **18.5.2 3l Semi-leptonic Channel**

1451 The following input features are used to predict the Higgs  $p_T$  for events in the 3lS channel:

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

Table 35: Input features for reconstructing the Higgs  $p_T$  spectrum for 3lS events

1452 Again,  $j_0$  and  $j_1$  are the two jets identified as Higgs decay products, ordered by  $p_T$ . The

1453 lepton identified as originating from the Higgs is labeled  $l_H$ , while the other two leptons are

1454 labeled  $l_{T0}$  and  $l_{T1}$ .  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet identification algorithm.

1455 The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs reconstruction algorithm,

1456 and the b-jet identification algorithm, respectively.

1457 The optimal neural network architecture for this channel is found to consist of 7 hidden

1458 layers with 80 nodes each. The inputdata set includes one million events, 10% of which is used

1459 for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1460 To evaluate the performance of the model, the predicted  $p_T$  spectrum is compared to the  
 1461 truth Higgs  $p_T$  in Figure 18.18. Once again, (a) of 18.18 shows a scatterplots of predicted vs  
 1462 truth  $p_T$ , where the color of each point corresponds to the log of the relative KDE at that point.  
 1463 Each column of the the histogram in (b) is normalized to unity, to better demonstrate the output  
 1464 of the NN for each slice of truth  $p_T$ .

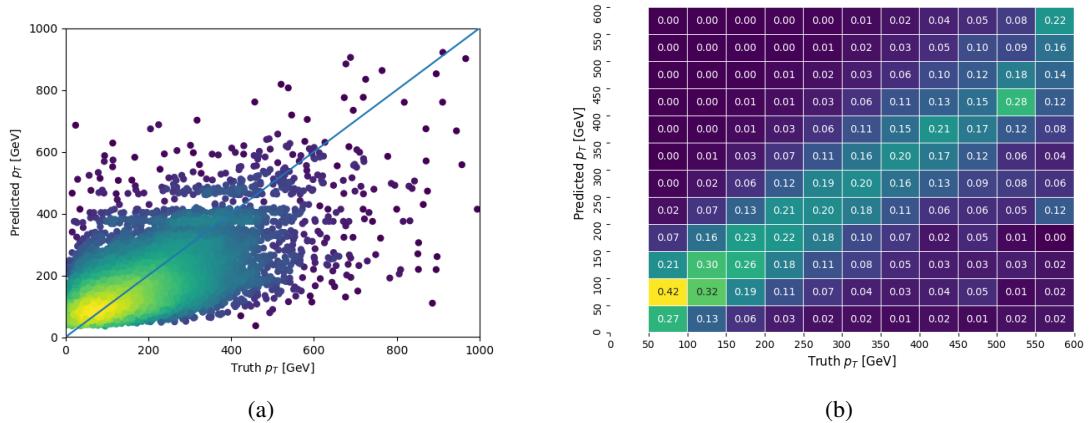


Figure 18.18: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 3lS  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1465 Figure 18.19 shows (a) the output of the NN for events with truth  $p_T$  less than and greater  
 1466 than 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low  
 1467  $p_T$  events.

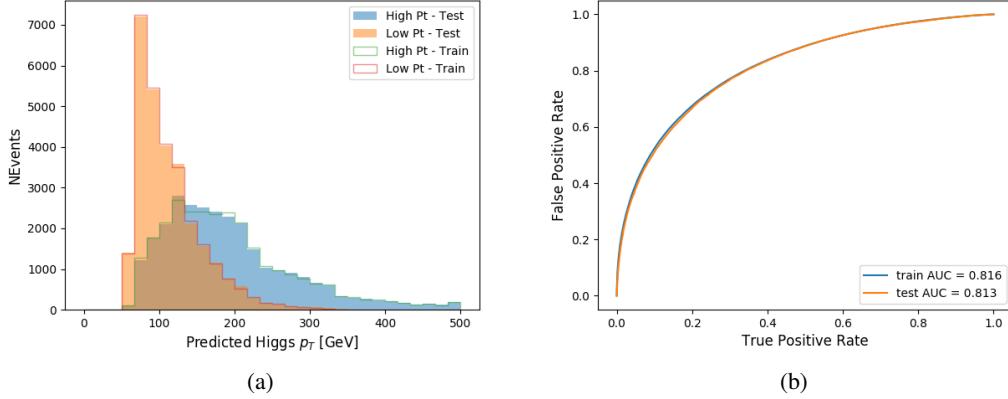


Figure 18.19: (

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.

### 1468 18.5.3 3l Fully-leptonic Channel

1469 The features listed in 36 are used to construct a model for predictin the Higgs  $p_T$  for 3lF events.

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

Table 36: Input features for reconstructing the Higgs  $p_T$  spectrum for 3lF events

1470  $l_{H0}$  and  $l_{H1}$  respresent the two leptons identified by the Higgs reconstruction model as

1471 originating from the Higgs, while  $l_T$  is the other lepton in the event. The Higgs Reco Score and  
 1472 b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification  
 1473 algorithm, respectively.

1474 The optimal neural network architecture for this channel is found to consist of 5 hidden  
 1475 layers with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for  
 1476 testing, the other 90% for training. Training is found to converge after around 150 epochs.

1477 The predicted transverse momentum, as a function of the truth  $p_T$ , is shown in Figure  
 1478 [18.20](#).

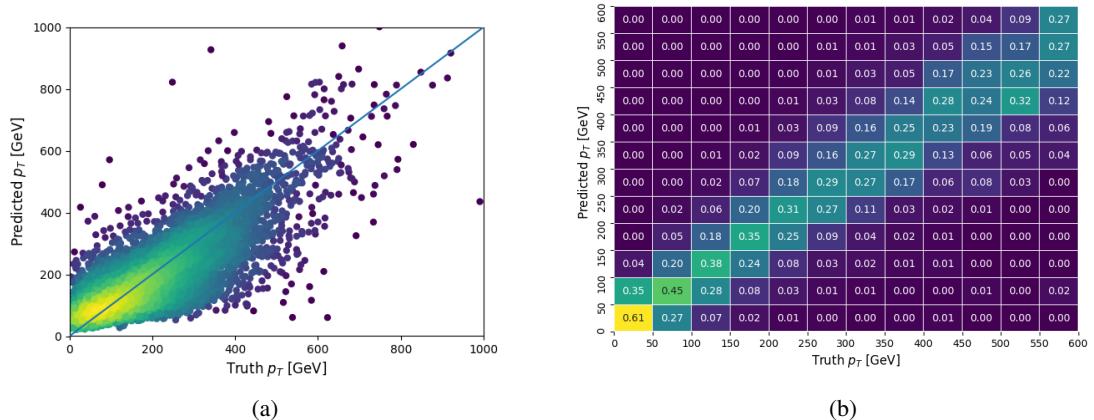


Figure 18.20: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 3lF  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1479 When split into high and low  $p_T$ , based on a cutoff of 150 GeV, the

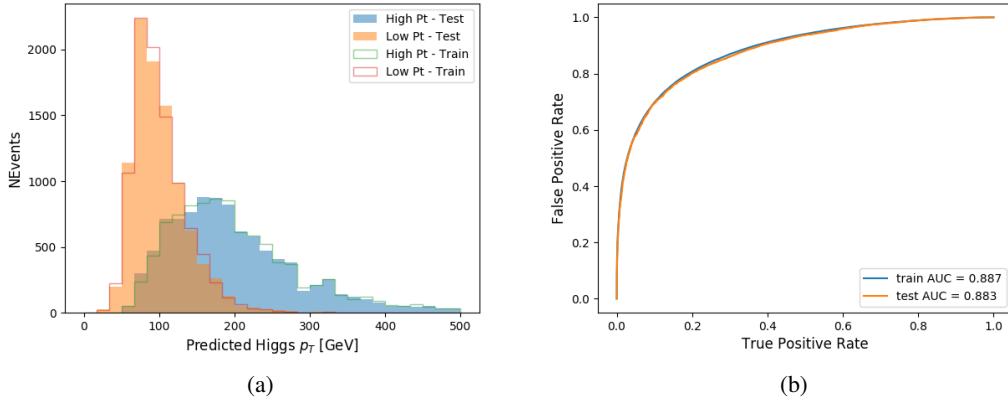


Figure 18.21: (a) shows the reconstructed Higgs  $p_T$  for 3lF events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

## 1480 18.6 3l Decay Mode

1481 In the 3l channel, there are two possible ways for the Higgs to decay, both involving intermediate  
 1482 W boson pairs: Either both W bosons decay leptonically, in which case the reconstructed decay  
 1483 consists of two leptons (referred as the fully-leptonic 3l channel), or one W decays leptonically  
 1484 and the other hadronically, giving two jets and one lepton in the final state (referred to as the  
 1485 semi-leptonic 3l channel). In order to accurately reconstruct the Higgs, it is necessary to identify  
 1486 which of these decays took place for each 3l event.

1487 The kinematics of each event, along with the output scores of the Higgs and top recon-  
 1488 struction algorithms, are used to distinguish these two possible decay modes. The particular  
 1489 inputs used are listed in Table 37.

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

1490 Here  $l_0$  is the opposite charge lepton,  $l_1$  and  $l_2$  are the two SS leptons order by  $\Delta R$   
 1491 from lepton 0. score3lF and score3lS are the outputs of the 3lS and 3lF Higgs reconstruction  
 1492 algorithms, while topScore is the output of the b-jet identification algorithm.

1493 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these  
 1494 two decay modes. The output of the model is summarized in Figure 18.22.

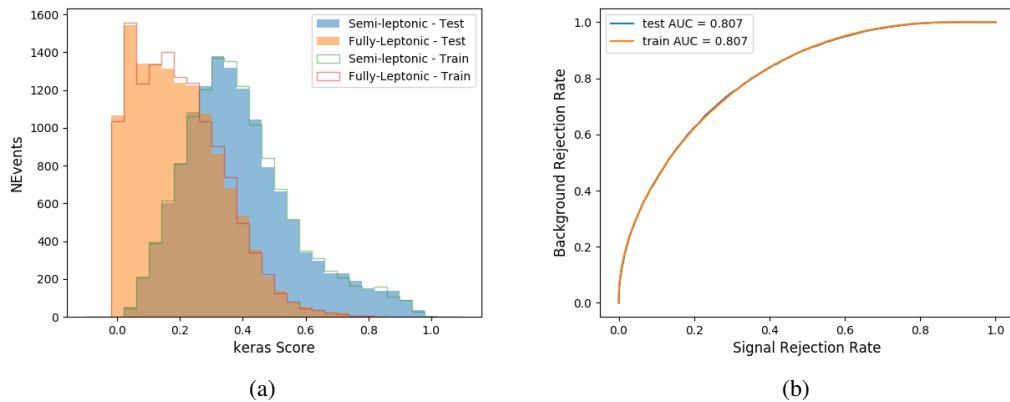


Figure 18.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

1495 A cutoff of 0.23 is determined to be optimal for separating 3lS and 3lF in the fit.

## 1496 19 Signal Region Definitions

1497 Events are divided into two channels based on the number of leptons in the final state: one with  
1498 two same-sign leptons, the other with three leptons. The 3l channel includes events where both  
1499 leptons originated from the Higgs boson as well as events where only one of the leptons

### 1500 19.1 Pre-MVA Event Selection

1501 A preselection is applied to define orthogonal analysis channels based on the number of leptons  
1502 in each event. For the 2lSS channel, the following preselection is used:

- 1503 • Two very tight, same-charge, light leptons with  $p_T > 20$  GeV
- 1504 •  $\geq 4$  reconstructed jets,  $\geq 1$  b-tagged jets
- 1505 • No reconstructed tau candidates

1506 The event yield after the 2lSS preselection has been applied, for MC and data at  $79.8 \text{ fb}^{-1}$ ,  
1507 is shown in Table 19.1.

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 38: Yields of the 2lSS preselection region

1508

Figure 19.1. Good general agreement is found.

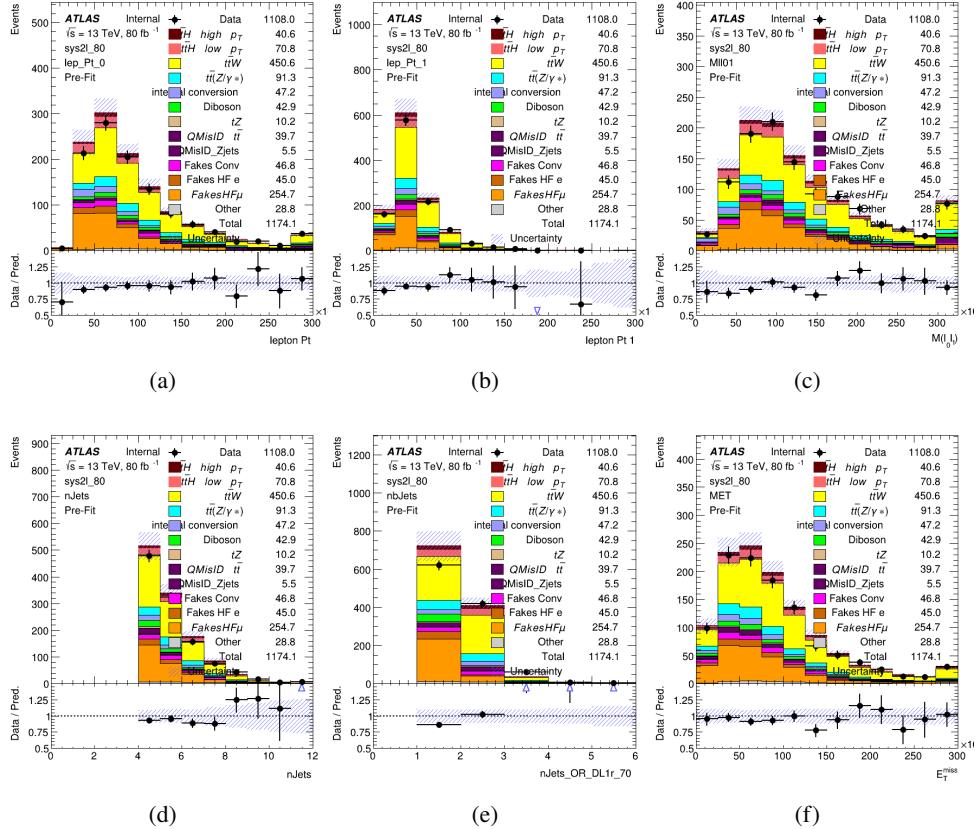


Figure 19.1: Data/MC comparisons of the 2LSS pre-selection region. (a) and (b) show the  $p_T$  of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1509 For the 31 channel, the following selection is applied:

- 1510 • Three light leptons with total charge  $\pm 1$
- 1511 • Same charge leptons are required to be very tight, with  $p_T > 20 \text{ GeV}$
- 1512 • Opposite charge lepton must be loose, with  $p_T > 10 \text{ GeV}$
- 1513 •  $\geq 2$  reconstructed jets,  $\geq 1$  b-tagged jets

- 1514     • No reconstructed tau candidates
- 1515     •  $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$  for all opposite-charge, same-flavor lepton pairs
- 1516     The event yield after the 3l preselection has been applied, for MC and data at  $79.8 \text{ fb}^{-1}$ ,  
 1517     is shown in Table 19.1.

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 39: Yields of the 3l preselection region.

Table 40: Yields of the 3l preselection region.

1518     Comparisons of kinematic distributions for data and MC in this region are shown in Figure

1519 19.2.

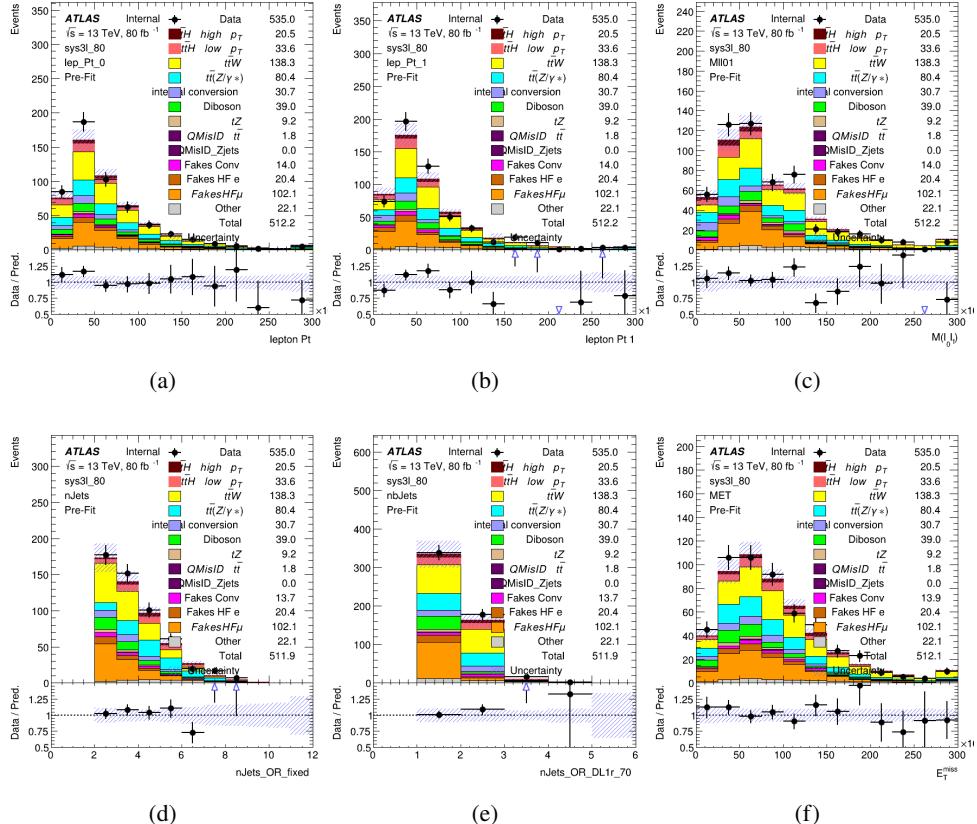


Figure 19.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the  $p_T$  of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1520 19.2 Event MVA

1521 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events  
 1522 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic  
 1523 (3lF). Here events with three leptons are split into 3lS and 3lF based on the model described  
 1524 in 18.6. In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

1525 [xgboost] are trained using the kinematics of signal and background events derived from Monte  
 1526 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo  
 1527 event.

1528 Because the background composition differs for events with a high reconstructed Higgs  $p_T$   
 1529 compared to events with low reconstructed Higgs  $p_T$ , separate MVAs are produced for high and  
 1530 low  $p_T$  regions. This is found to provide better significance than attempting to build an inclusive  
 1531 model, as demonstrated in appendix C.2. A cutoff of 150 GeV is used. This gives a total of 6  
 1532 background rejection MVAs - explicitly, 2lSS high  $p_T$ , 2lSS low  $p_T$ , 3lS high  $p_T$ , 3lS low  $p_T$ ,  
 1533 3lF high  $p_T$ , and 3lF low  $p_T$ .

1534 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 41: Input features used to distinguish signal and background events in the 2lSS channel.

1535 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 42: Input features used to distinguish signal and background events in the 3l channel.

1536 Modelling of each of these input features is verified in Appendix C.2 by comparing data  
 1537 and MC for  $79.8 \text{ fb}^{-1}$ . The BDTs are produced with a maximum tree depth of 6, using AUC as  
 1538 the target loss function. The BDT response distribution and ROC curve for each model is shown  
 1539 in Figures 19.3-19.5.

## 2LSS

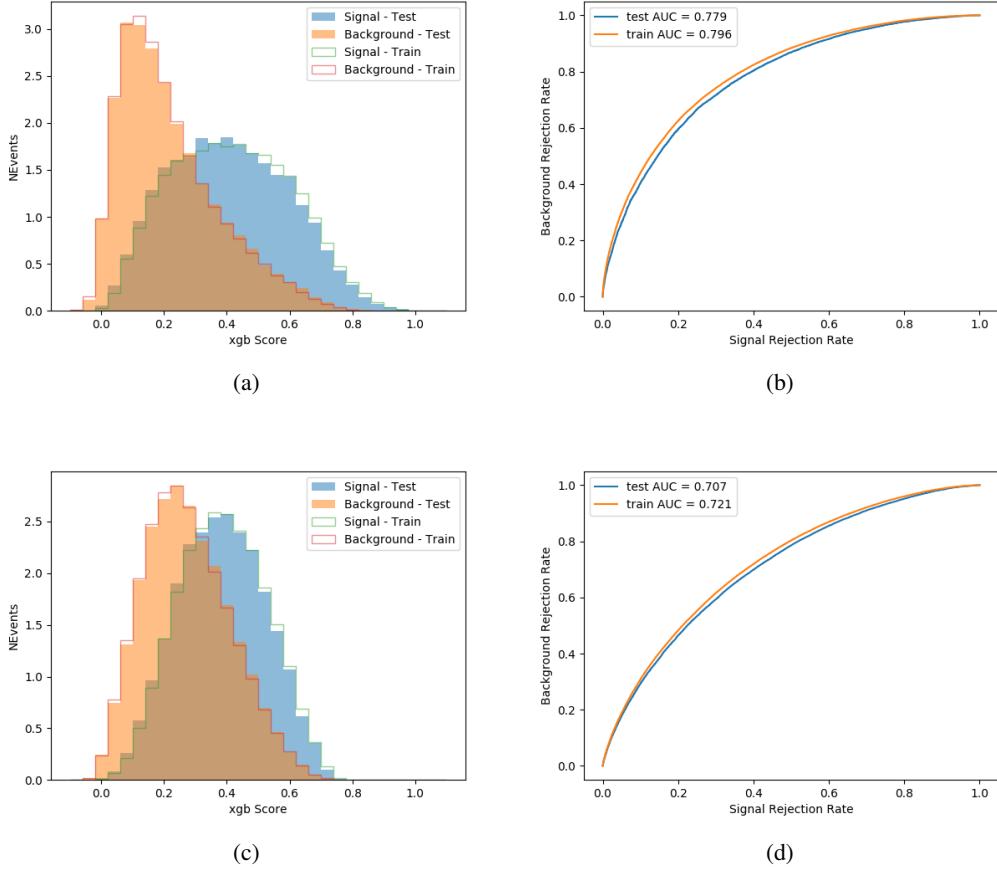


Figure 19.3: Output BDT scores of training and testing data for signal (blue) and background (orange) for 2LSS events with (a) high regressed Higgs  $p_T$  and (b) low regressed Higgs  $p_T$ . (b) and (d) show the ROC curve for the 2LSS high and low  $p_T$  models, respectively.

## 3l - Semileptonic

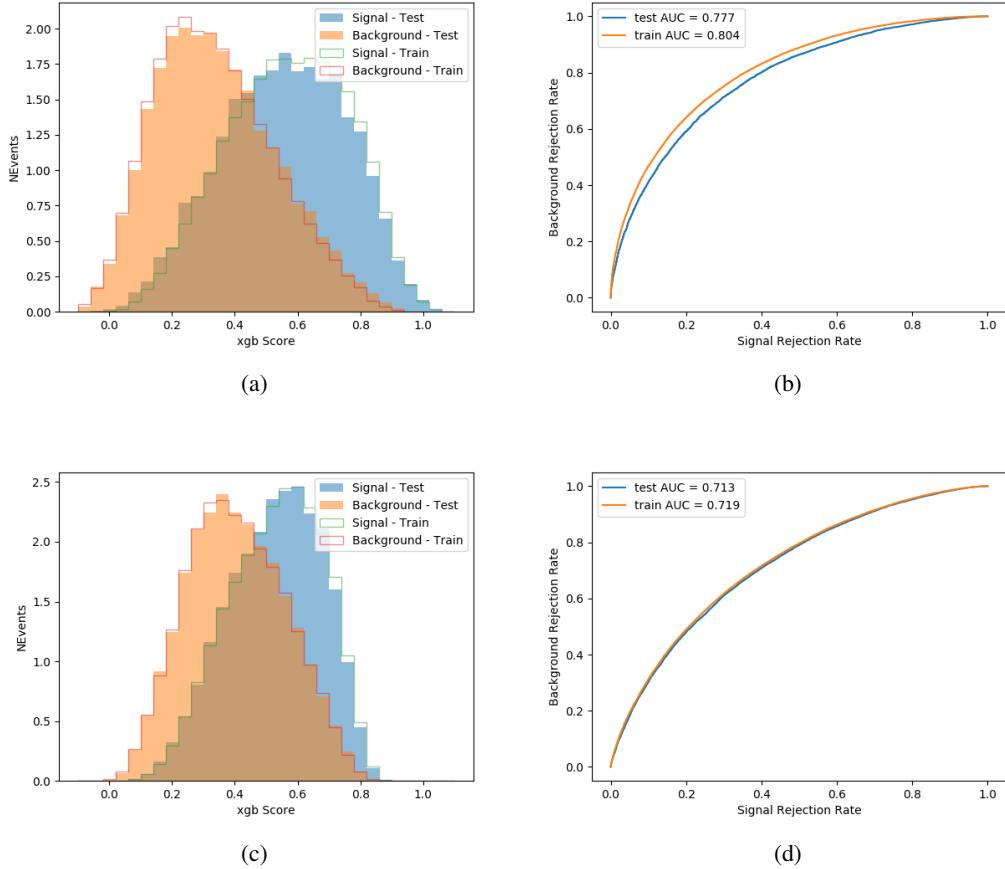


Figure 19.4: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lS events with (a) high regressed Higgs  $p_T$  and (b) low regressed Higgs  $p_T$ . (b) and (d) show the ROC curve for the 3lS high and low  $p_T$  models, respectively.

### 3l - Fully Leptonic

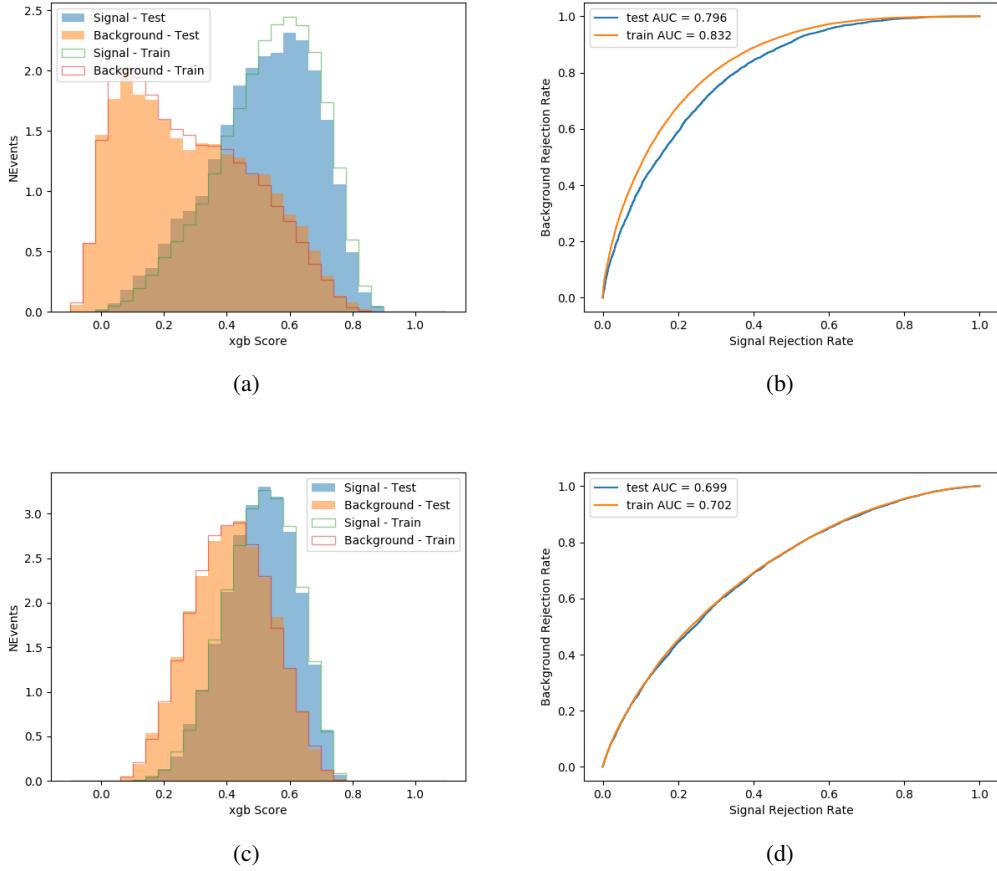


Figure 19.5: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lF events with (a) high regressed Higgs  $p_T$  and (b) low regressed Higgs  $p_T$ . (b) and (d) show the ROC curve for the 3lF high and low  $p_T$  models, respectively.

1540      Output distributions of each MVA, comparing MC predictions to data at  $79.8 \text{ fb}^{-1}$  are  
 1541      shown in figures 19.6-19.2.

### High $p_T$ Background Rejection BDTs

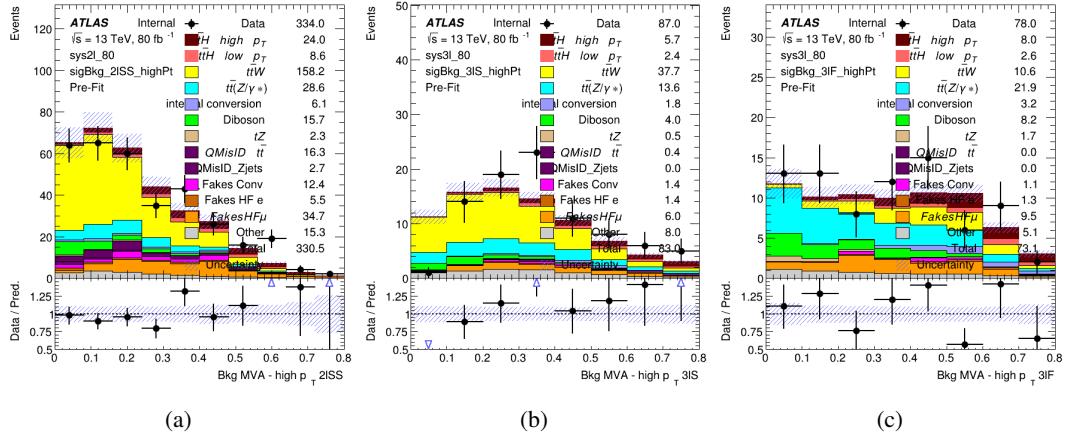


Figure 19.6: Output score of the high  $p_T$  BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channels

### Low $p_T$ Background Rejection BDTs

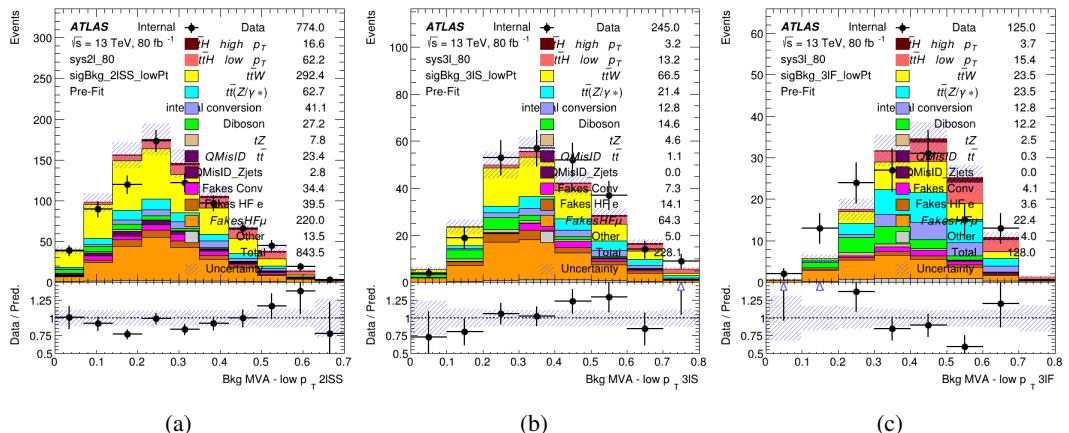


Figure 19.7: Output score of the low  $p_T$  BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channels

1542 **19.3 Signal Region Definitions**

1543 Once pre-selection has been applied, channels are further refined based on the MVAs described  
 1544 above. The output of the model described in Section 18.6 is used to separate the three channel  
 1545 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs  
 1546 boson. This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

1547 For each event, depending on the number of leptons as well as whether the  $p_T$  of the Higgs  
 1548 is predicted to be high ( $> 150$  GeV) or low ( $< 150$  GeV), a cut on the appropriate background  
 1549 rejection MVA is applied. The particular cut values, listed in Table 43, are determined by  
 1550 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 43: Cutoff values on background rejection MVA score applied to signal regions.

1551 The event preselection and MVA selection listed in Table 43 are used define the three  
 1552 signal regions used in the fit. These signal region definitions are summarized in Table 44.

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

Table 44: Selection applied to define the three signal regions used in the fit.

## 20 Systematic Uncertainties

1553 The systematic uncertainties that are considered are summarized in Table 45. These are imple-  
 1554 mented in the fit either as a normalization factors or as a shape variation or both in the signal  
 1555 and background estimations. The numerical impact of each of these uncertainties is outlined in  
 1556 section 21.  
 1557

Table 45: 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>1558</sup> The uncertainty in the combined integrated luminosity is derived from a calibration of the  
<sup>1559</sup> luminosity scale using x-y beam-separation scans performed for 13 TeV proton-proton data [30],  
<sup>1560</sup> [31].

<sup>1561</sup> The experimental uncertainties are related to the reconstruction and identification of light  
<sup>1562</sup> leptons and b-tagging of jets, and to the reconstruction of  $E_T^{\text{miss}}$ . The TOTAL electron ID  
<sup>1563</sup> correlation model is used, corresponding to 1 electron ID systematic. Electron ID is found to be

1564 a subleading systematic that is unconstrained by the fit, making it an appropriate choice for this  
1565 analysis.

1566 The sources which contribute to the uncertainty in the jet energy scale [32] are decom-  
1567 posed into uncorrelated components and treated as independent sources in the analysis. The  
1568 CategoryReduction model is used to account for JES uncertainties, which decomposes the uncer-  
1569 tainties into 30 nuisance parameters included in the fit. The SimpleJER model is used to account  
1570 for jet energy resolution (JER) uncertainties, and 8 JER uncertainty components unclued as  
1571 NPs in the fit.

1572 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses  
1573 [33] are also decomposed into uncorrelated components. The large number of components for  
1574 b-tagging is due to the calibration of the distribution of the BDT discriminant.

1575 The systematic uncertainties associated with the signal and background processes are  
1576 accounted for by varying the cross-section of each process within its uncertainty.

1577 The full list of systematic uncertainties considered in the analysis is summarized in Tables  
1578 46, 47 and 48.

1579

<b>Experimental Systematics on Leptons and <math>E_T^{\text{miss}}</math></b>			
Type	Description	Systematics Name	Application
<b>Trigger</b>			
Scale Factors	Trigger Efficiency	lepSFTrigTight_MU(EL)_SF_Trigger_STAT(SYST)	Event Weight
<b>Muons</b>			
Efficiencies	Reconstruction and Identification	lepSFObjTight_MU_SF_ID_STAT(SYST)	Event Weight
	Isolation	lepSFObjTight_MU_SF_Isol_STAT(SYST)	Event Weight
	Track To Vertex Association	lepSFObjTight_MU_SF_TTVA_STAT(SYST )	Event Weight
$p_T$ Scale	$p_T$ Scale	MUONS_SCALE	$p_T$ Correction
Resolution	Inner Detector Energy Resolution	MUONS_ID	$p_T$ Correction
	Muon Spectrometer Energy Resolution	MUONS_MS	$p_T$ Correction
<b>Electrons</b>			
Efficiencies	Reconstruction	lepSFObjTight_EL_SF_ID	Event Weight
	Identification	lepSFObjTight_EL_SF_Reco	Event Weight
	Isolation	lepSFObjTight_EL_SF_Isol	Event Weight
Scale Factor	Energy Scale	EG_SCALE_ALL	Energy Correction
Resolution	Energy Resolution	EG_RESOLUTION_ALL	Energy Correction
<b><math>E_T^{\text{miss}}</math></b>			
Soft Tracks Terms	Resolution	MET_SoftTrk_ResoPerp	$p_T$ Correction
	Resolution	MET_SoftTrk_ResoPara	$p_T$ Correction
	Scale	MET_SoftTrk_ScaleUp	$p_T$ Correction
	Scale	MET_SoftTrk_ScaleDown	$p_T$ Correction

Table 46: Summary of experimental systematics considered for leptons and  $E_T^{\text{miss}}$ . Includes type, description, name of systematic as used in the fit, and mode of application. The mode of application indicates the systematic evaluation, e.g. as an overall event re-weighting (Event Weight) or rescaling ( $p_T$  Correction).

Experimental Systematics on Jets			
Type	Origin	Systematics Name	Application
Jet Vertex Tagger		JVT	Event Weight
Energy Scale	Calibration Method	JET_21NP_ JET_EffectiveNP_1-19	$p_T$ Correction $p_T$ Correction
	$\eta$ inter-calibration	JET_EtaIntercalibration_Modelling JET_EtaIntercalibration_NonClosure JET_EtaIntercalibration_TotalStat	$p_T$ Correction $p_T$ Correction $p_T$ Correction
	High $p_T$ jets	JET_SingleParticle_HighPt	$p_T$ Correction
	Pile-Up	JET_Pileup_OffsetNPV JET_Pileup_OffsetMu JET_Pileup_PtTerm JET_Pileup_RhoTopology	$p_T$ Correction $p_T$ Correction $p_T$ Correction $p_T$ Correction
	Non Closure	JET_PunchThrough_MC15	$p_T$ Correction
	Flavour	JET_Flavor_Response JET_BJES_Response JET_Flavor_Composition	$p_T$ Correction $p_T$ Correction $p_T$ Correction
Resolution		JET_JER_SINGLE_NP	Event Weight

Table 47: Jet systematics take into account effects of jets calibration method,  $\eta$  inter-calibration, high  $p_T$  jets, pile-up, and flavor response. They are all diagonalised into effective parameters.

Experimental Systematics on b-tagging		
Type	Origin	Systematic Name
Scale Factors	DL1r b-tagger efficiency on b originated jets in bins of $\eta$	DL1r_Continuous_EventWeight_B0-29
	DL1r b-tagger efficiency on c originated jets in bins of $\eta$	DL1r_Continuous_EventWeight_C0-19
	DL1r b-tagger efficiency on light flavoured originated jets in bins of $\eta$ and $p_T$	DL1r_Continuous_EventWeight_Light0-79
	DL1r b-tagger extrapolation efficiency	DL1r_Continuous_EventWeight_extrapolation DL1r_Continuous_EventWeight_extrapolation_from_charm

Table 48: Summary of experimental systematics to be included for b-tagging of jets in the analysis, using the continuous DL1r tagging algorithm. All of the b-tagging related systematics are applied as event weights. From left: type, description, and the name of systematic used in the fit.

1580 As mentioned in Section 16.2, a normalization corrections and uncertainties on the estim-  
 1581 ates of non-prompt leptons backgrounds are derived using data driven techniques, decribed in  
 1582 detail in [7]. These are derived from a likelihood fit over various non-prompt enriched control  
 1583 regions, targeting several sources of non-prompt light leptons separately: external conversion  
 1584 electrons, internal conversion electrons, electrons from heavy flavor decays, and muons from  
 1585 heavy flavor decays.

1586 The normalization factor and uncertainty applied to each source of non-prompt leptons is  
 1587 summarized in Table 20

Processs	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 49: Normalization factors - with statistical and systematic uncertainties - derived from the fit over fake control regions for each source of non-prompt leptons considered.

1588        In addition to those derived from the control regions, several additional uncertainties are  
 1589        assigned to the non-prompt lepton background. An additional 25% uncertainty on material  
 1590        conversions is assigned, based on the comparison between data and MC in a region where a  
 1591        loose electron fails the photon conversion veto. A shape uncertainty of 15% (6%) is assigned to  
 1592        the HF non-prompt electron (muon) background based on a comparison between data and MC  
 1593        where the second leading electron (muon) is only required to be loose. As the contribution from  
 1594        light non-prompt leptons is small, about 10% percent of the contribution from HF non-prompt  
 1595        leptons, it is derived from the agreement between data and simulation in a LF enriched region at  
 1596        low values of the non-prompt lepton BDT. The resulting uncertainty is 100%, and is taken to be  
 1597        uncorrelated between internal and material conversions.

1598        Theoretical uncertainties applied to MC predictions, including cross section, PDF, and  
 1599        scale uncertainties are taken from theory calculations for the predominate prompt backgrounds.  
 1600        Following the nominal  $t\bar{t}H$  – ML analysis, a 50% uncertainty is applied to Diboson to account  
 1601        for the large uncertainty in estimating  $VV +$  heavy flavor. The other “rare” background processes  
 1602        - including  $tZ$ , rare top processes,  $ttWW$ ,  $WtZ$ ,  $VVV$ ,  $tHjb$  and  $WtH$  - are assigned an overall

<sup>1603</sup> 50% normalization uncertainty as well. The theory uncertainties applied to the MC estimates

<sup>1604</sup> are summarized in Table 50.

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

<sup>1605</sup> Additional uncertainties to account for t̄W mismodelling are also applied. These include  
<sup>1606</sup> a “Generator” uncertainty, based on a comparison between the nominal Sherpa 2.2.5 sample,  
<sup>1607</sup> and the formerly used aMC@NLO sample, and an “Extra radiation” uncertainty, which includes  
<sup>1608</sup> renormalisation and factorisation scale variations of the Sherpa 2.2.5 sample.

## <sup>1609</sup> 21 Results

<sup>1610</sup> A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p<sub>T</sub> spectrum  
<sup>1611</sup> in the three signal regions, 2lSS, 3lS, and 3lF. The signal is split into high and low p<sub>T</sub> samples,  
<sup>1612</sup> based on whether the truth p<sub>T</sub> of the Higgs is above or below 150 GeV. The parameters  $\mu_{t\bar{t}H\text{high}p_T}$   
<sup>1613</sup> 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

<sub>1614</sub> between the observed value and the theory prediction. Unblinded results are shown for the 79.8  
<sub>1615</sub>  $\text{fb}^{-1}$  data set, as well as MC only projections of results using the full Run-2, 139  $\text{fb}^{-1}$  dataset.

<sub>1616</sub> As described in Section 20, there are 229 systematic uncertainties that are considered  
<sub>1617</sub> as NPs in the fit. These NP s are constrained by Gaussian or log-normal probability density  
<sub>1618</sub> functions. The latter are used for normalisation factors to ensure that they are always positive.  
<sub>1619</sub> The expected number of signal and background events are functions of the likelihood. The prior  
<sub>1620</sub> for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its  
<sub>1621</sub> nominal value.

## <sub>1622</sub> **21.1 Results - 79.8 $\text{fb}^{-1}$**

<sub>1623</sub> As the data collected from 2015-2017 has been unblinded for  $t\bar{t}H$  – ML channels, represent-  
<sub>1624</sub> ing 79.8  $\text{fb}^{-1}$ , those events are unblinded. The predicted Higgs  $p_T$  spectrum is fit to data  
<sub>1625</sub> simultaneously in each of the three signal regions shown in Figure 21.1.

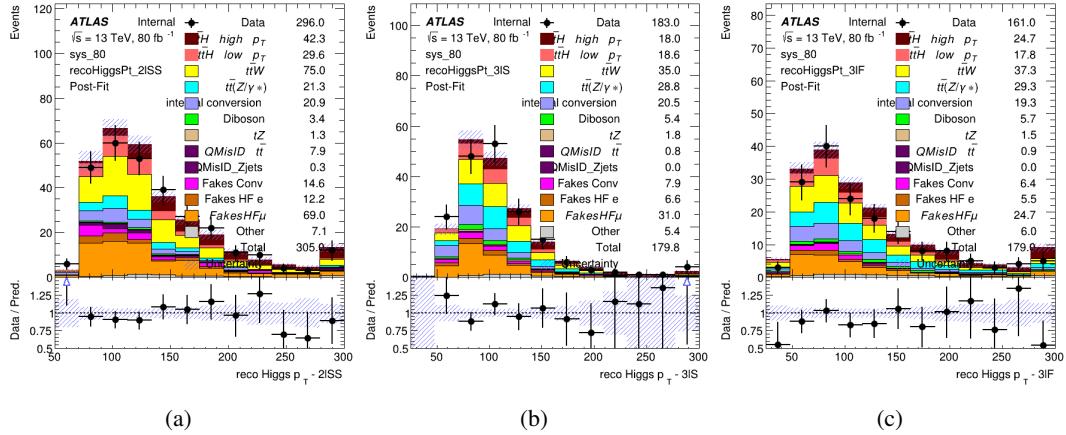


Figure 21.1: Post-fit distributions of the reconstructed Higgs  $p_T$  in the three signal regions, (a) 2ISS, (b) 3IS, and (c) 3IF, for  $79.8 \text{ fb}^{-1}$  of MC

1626

A post-fit summary of the fitted regions is shown in Figure 21.2.

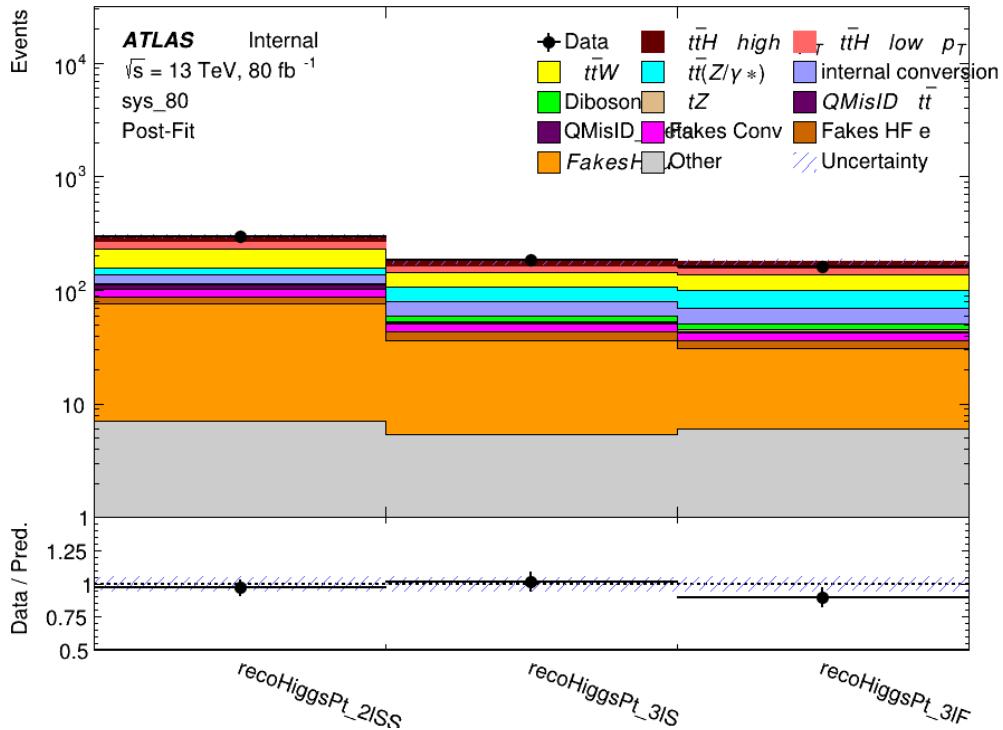


Figure 21.2: Post-fit summary of the yields in each signal region.

1627 The the measured  $\mu$  values for high and low  $p_T$  Higgs production obtained from the fit  
 1628 are shown in 51. A significance of  $1.7\sigma$  is observed for  $t\bar{t}H$  high  $p_T$ , and  $2.1\sigma$  is measured for  
 1629  $t\bar{t}H$  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 51: Best fit  $\mu$  values for  $t\bar{t}H$  high  $p_T$  and  $t\bar{t}H$  low  $p_T$ , where  $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1630 The most prominent sources of systematic uncertainty, as measured by their impact on  
 1631  $\mu_{t\bar{t}H \text{ high } p_T}$ , are summarized in Table 52.

Uncertainty Source	$\Delta\mu$	
Jet Energy Scale	0.25	0.23
$t\bar{t}H$ cross-section (QCD scale)	-0.11	0.21
Luminosity	-0.13	0.14
Flavor Tagging	0.14	0.13
$t\bar{t}W$ cross-section (QCD scale)	-0.12	0.11
Higgs Branching Ratio	-0.1	0.11
$t\bar{t}H$ cross-section (PDF)	-0.07	0.08
Electron ID	-0.06	0.06
Non-prompt Muon Normalization	-0.05	0.06
$t\bar{t}Z$ 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 52: Summary of the most significant sources of systematic uncertainty on the measurement of  $t\bar{t}H$  high  $p_T$ .

1632 The most significant sources of uncertainty on the measurement of  $t\bar{t}H$  - low  $p_T$  are shown  
 1633 in Table 53.

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

<sup>1634</sup> The ranking and impact of those nuisance parameters with the largest contribution to the overall uncertainty is shown in Figure 21.3.

<sup>1635</sup>

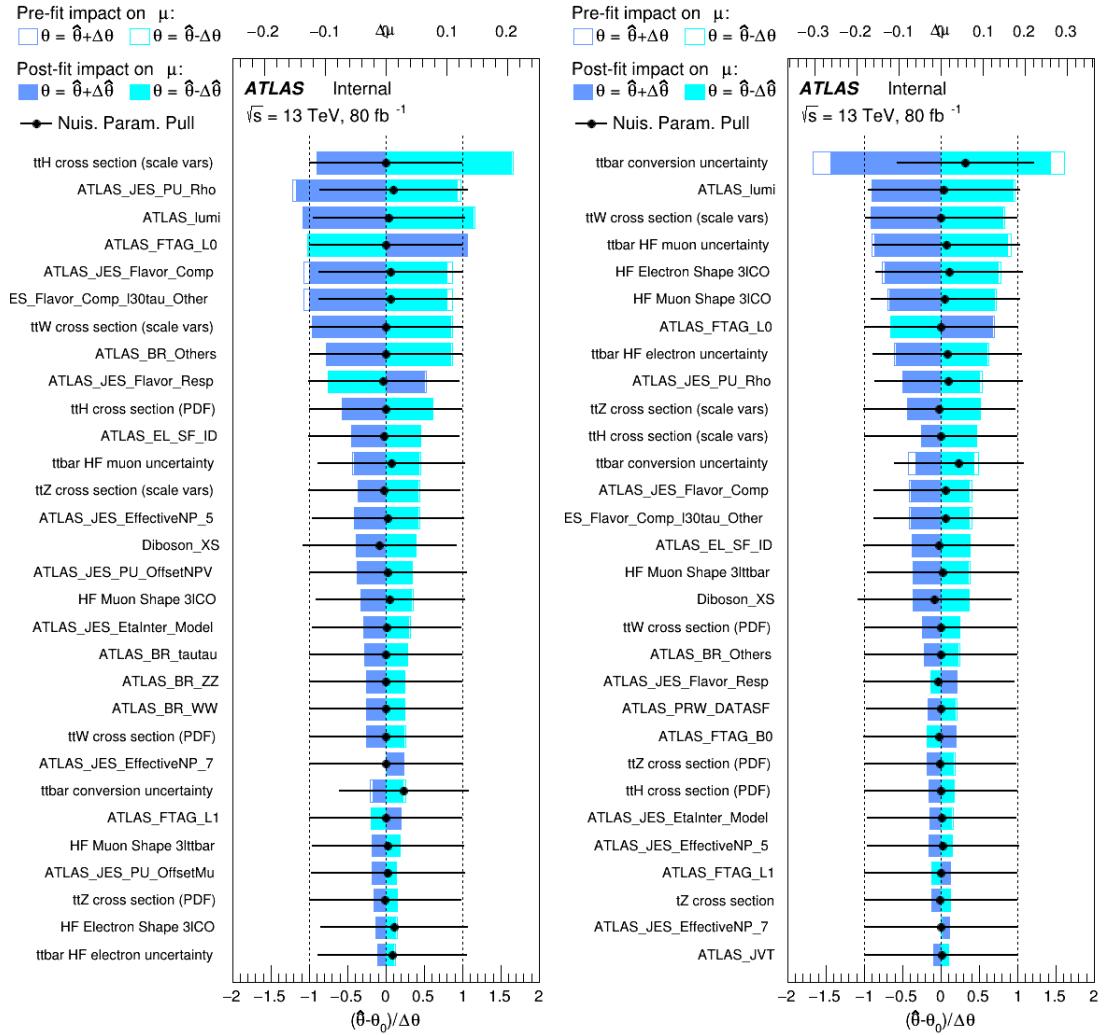


Figure 21.3: Impact of systematic uncertainties on the measurement of high  $p_T$  (left) and low  $p_T$  (right)  $t\bar{t}H$  events

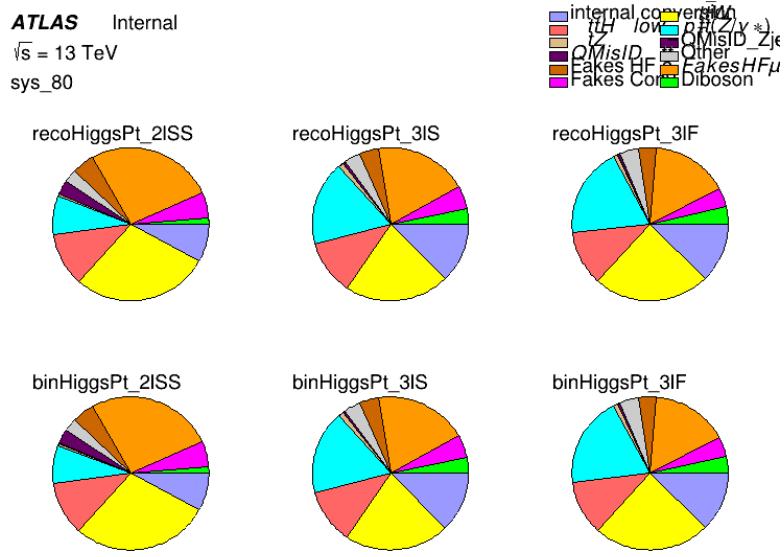


Figure 21.4: Background composition of the fit regions.

## 1637 21.2 Projected Results - $139 \text{ fb}^{-1}$

1638 As data collected in 2018 has not yet been unblinded for  $t\bar{t}H$  – ML at the time of this note, data  
1639 from that year remains blinded. Instead, an Asimov fit is performed - with the MC prediction  
1640 being used both as the SM prediction as well as the data in the fit - in order to give expected  
1641 results.

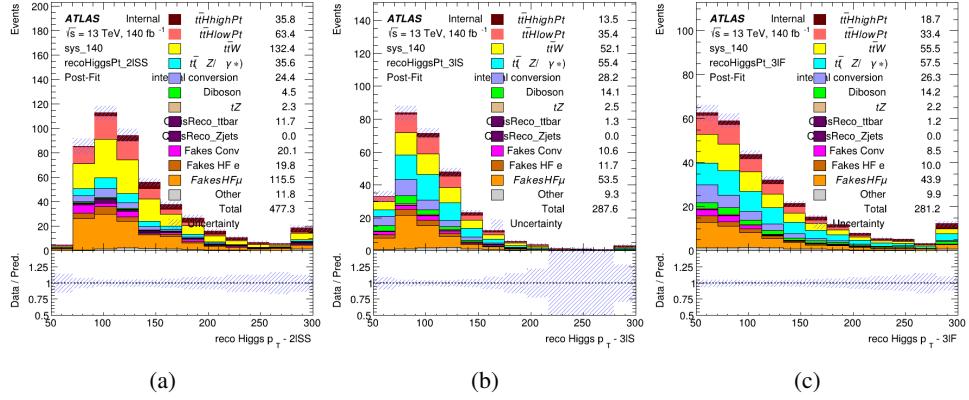


Figure 21.5: Blinded post-fit distributions of the reconstructed Higgs  $p_T$  in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for  $139 \text{ fb}^{-1}$  of data

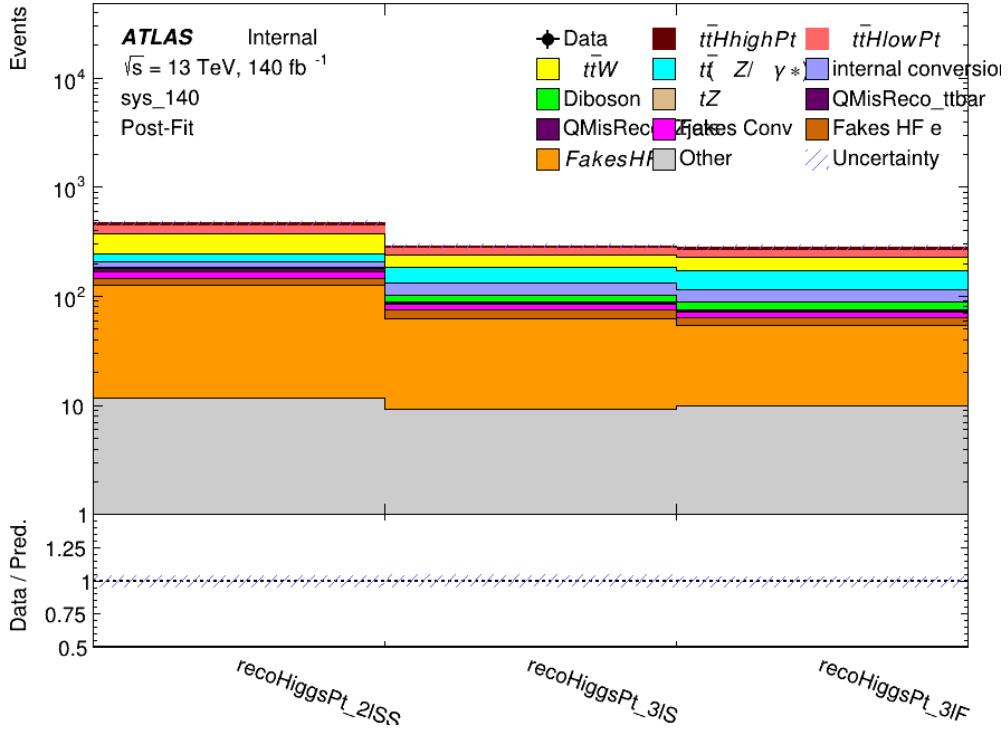


Figure 21.6: Post-fit summary of fit.

<sup>1643</sup> shown in [54](#). A significance of  $2.0\sigma$  is expected for  $t\bar{t}H$  high  $p_T$ , and a projected significance

<sup>1644</sup>  $2.3\sigma$  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 54: Best fit  $\mu$  values for  $t\bar{t}H$  high  $p_T$  and  $t\bar{t}H$  low  $p_T$ , where  $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

<sup>1645</sup> The most prominent sources of systematic uncertainty, as measured by their impact on

<sup>1646</sup>  $\mu_{t\bar{t}H \text{ high } p_T}$ , are summarized in Table [55](#).

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

<sup>1647</sup> The most significant sources of systematic uncertainty on  $t\bar{t}H$  low  $p_T$  are summarized in

<sup>1648</sup> Table [56](#).

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

<sup>1649</sup> The ranking and impact of those nuisance parameters with the largest contribution to the  
<sup>1650</sup> overall uncertainty is shown in Figure 21.7.

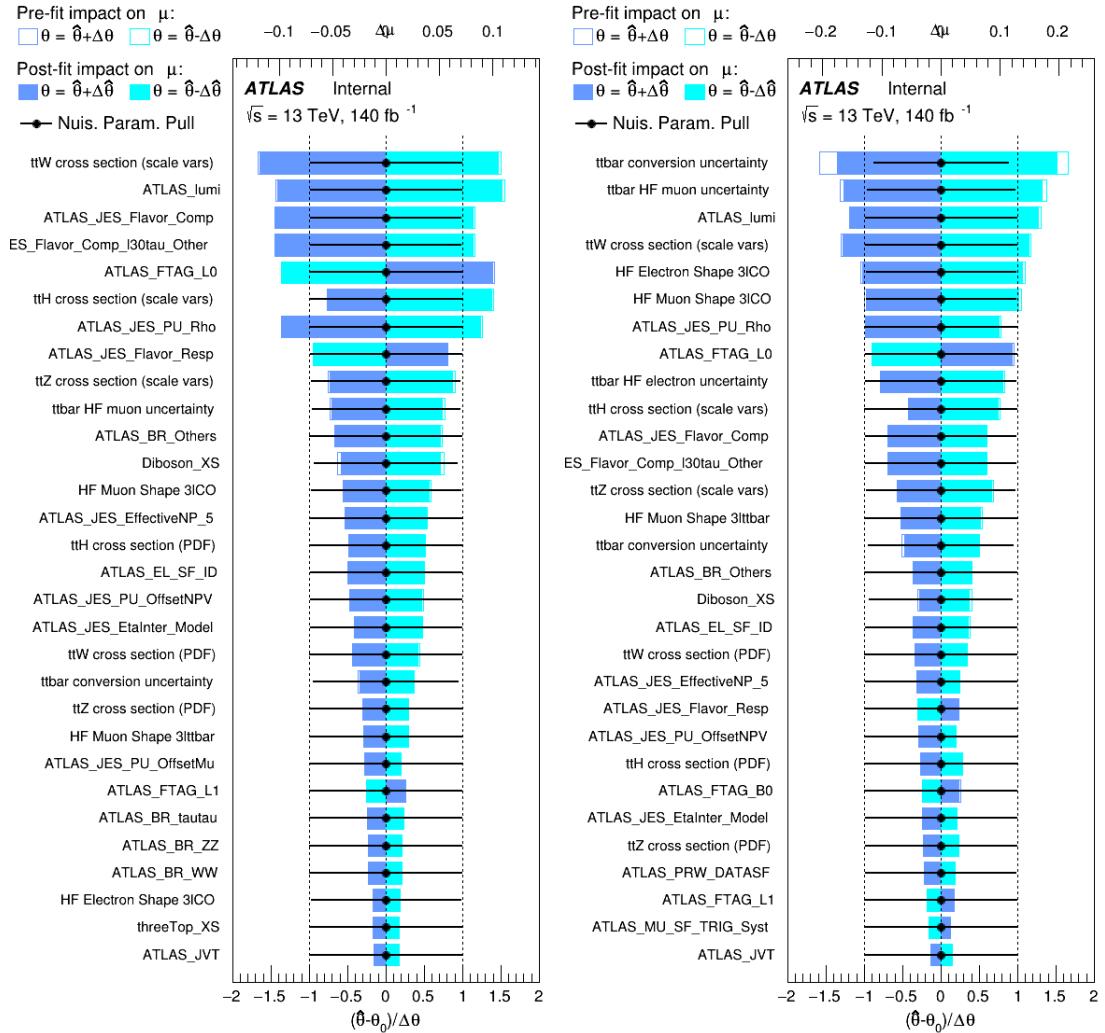


Figure 21.7: Impact of systematic uncertainties on the measurement of high  $p_T$  (left) and low  $p_T$  (right)  $t\bar{t}H$  events

1651

The background composition of each of the fit regions is shown in Figure 21.8.

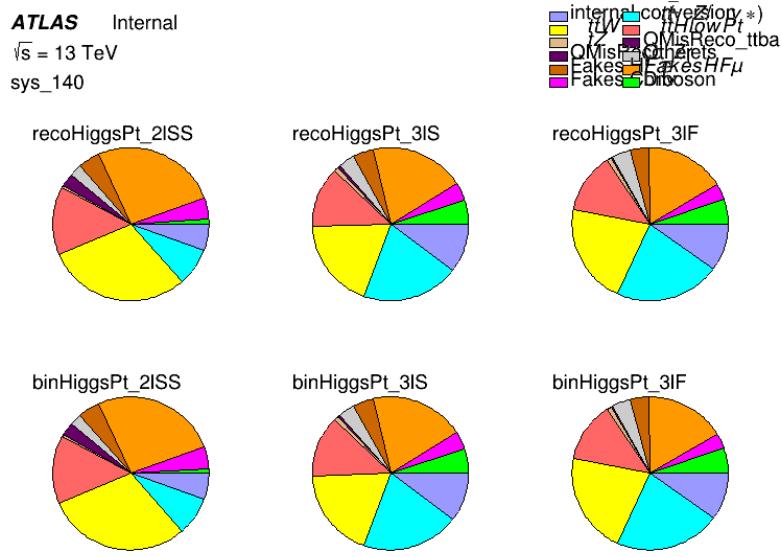


Figure 21.8: Background composition of the fit regions.

## 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

- 1658 [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*  
1660 *Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,  
1661 arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- 1662 [2] R. Oerter, *The theory of almost everything : the Standard Model, the unsung triumph of*  
1663 *modern physics*, Pi Press, 2006, ISBN: 978-0-13-236678-6.
- 1664 [3] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,  
1665 Phys. Rev. Lett. 13 (Oct, 1964) 508-509 ().
- 1666 [4] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*,  
1667 Phys. Rev. **127** (3 1962) 965,  
1668 URL: <https://link.aps.org/doi/10.1103/PhysRev.127.965>.
- 1669 [5] J. Ellis, *Higgs Physics*,  
1670 (2013) 117, 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of  
1671 High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report,  
1672 URL: <https://cds.cern.ch/record/1638469>.
- 1673 [6] T. A. Collaboration, *Observation of Higgs boson production in association with a top*  
1674 *quark pair at the LHC with the ATLAS detector*, Physics Letters B **784** (2018) 173,  
1675 ISSN: 0370-2693, URL:  
1676 <https://www.sciencedirect.com/science/article/pii/S0370269318305732>.

- 1677 [7] *Evidence for the associated production of the Higgs boson and a top quark pair with the*  
1678 *ATLAS detector*, tech. rep. ATLAS-CONF-2017-077, CERN, 2017,  
1679 URL: <https://cds.cern.ch/record/2291405>.
- 1680 [8] A. M. Sirunyan et al., *Observation of  $t\bar{t}H$  Production*,  
1681 *Physical Review Letters* **120** (2018), ISSN: 1079-7114,  
1682 URL: <http://dx.doi.org/10.1103/PhysRevLett.120.231801>.
- 1683 [9] B. Dumont, S. Fichet and G. von Gersdorff,  
1684 *A Bayesian view of the Higgs sector with higher dimensional operators*,  
1685 *Journal of High Energy Physics* **2013** (2013), ISSN: 1029-8479,  
1686 URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)065](http://dx.doi.org/10.1007/JHEP07(2013)065).
- 1687 [10] S. Banerjee, S. Mukhopadhyay and B. Mukhopadhyaya,  
1688 *Higher dimensional operators and the LHC Higgs data: The role of modified kinematics*,  
1689 *Physical Review D* **89** (2014), ISSN: 1550-2368,  
1690 URL: <http://dx.doi.org/10.1103/PhysRevD.89.053010>.
- 1691 [11] M. Zinser, ‘The Large Hadron Collider’, *Search for New Heavy Charged Bosons and*  
1692 *Measurement of High-Mass Drell-Yan Production in Proton—Proton Collisions*,  
1693 Springer International Publishing, 2018 47, ISBN: 978-3-030-00650-1,  
1694 URL: [https://doi.org/10.1007/978-3-030-00650-1\\_4](https://doi.org/10.1007/978-3-030-00650-1_4).
- 1695 [12] *Detector and Technology*, URL: <https://atlas.cern/discover/detector>.
- 1696 [13] M. Marjanovic, *ATLAS Tile calorimeter calibration and monitoring systems*, 2018.

- 1697 [14] M. Aaboud et al., *Observation of electroweak  $W^\pm Z$  boson pair production in association*  
 1698 *with two jets in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector,*  
 1699 [Phys. Lett. B793 \(2019\) 469](#), arXiv: [1812.09740 \[hep-ex\]](#).
- 1700 [15] T. Gleisberg et al., *Event generation with SHERPA 1.1*, [JHEP 02 \(2009\) 007](#),  
 1701 arXiv: [0811.4622 \[hep-ph\]](#).
- 1702 [16] R. D. Ball et al., *Parton distributions for the LHC Run II*, [JHEP 04 \(2015\) 040](#),  
 1703 arXiv: [1410.8849 \[hep-ph\]](#).
- 1704 [17] H.-L. Lai et al., *New parton distributions for collider physics*,  
 1705 [Phys. Rev. D 82 \(2010\) 074024](#), arXiv: [1007.2241 \[hep-ph\]](#).
- 1706 [18] S. Frixione, G. Ridolfi and P. Nason,  
 1707 *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*,  
 1708 [JHEP 09 \(2007\) 126](#), arXiv: [0707.3088 \[hep-ph\]](#).
- 1709 [19] E. Re, *Single-top Wt-channel production matched with parton showers using the*  
 1710 *POWHEG method*, [Eur. Phys. J. C 71 \(2011\) 1547](#), arXiv: [1009.2450 \[hep-ph\]](#).
- 1711 [20] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using*  
 1712 *the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016,  
 1713 URL: <https://cds.cern.ch/record/2157687>.
- 1714 [21] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the*  
 1715 *ATLAS detector using 2011 and 2012 LHC proton–proton collision data*,  
 1716 [Eur. Phys. J. C 74 \(2014\) 3130](#), arXiv: [1407.3935 \[hep-ex\]](#).

- 1717 [22] R. Narayan et al., *Measurement of the total and differential cross sections of a*  
1718 *top-quark-antiquark pair in association with a W boson in proton-proton collisions at a*  
1719 *centre-of-mass energy of 13 TeV with ATLAS detector at the Large Hadron Collider,*  
1720 *tech. rep. ATL-COM-PHYS-2020-217, CERN, 2020,*  
1721 *URL: <https://cds.cern.ch/record/2712986>.*
- 1722 [23] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets*  
1723 *Reconstructed in the ATLAS Detector at  $\sqrt{s} = 13 \text{ TeV}$ , ATL-PHYS-PUB-2015-015,*  
1724 *2015, URL: <https://cds.cern.ch/record/2037613>.*
- 1725 [24] ATLAS Collaboration,  
1726 *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector,*  
1727 *ATLAS-CONF-2015-029, 2015, URL: <https://cds.cern.ch/record/2037702>.*
- 1728 [25] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp*  
1729 *collisions at  $\sqrt{s} = 8 \text{ TeV}$  using the ATLAS detector, Eur. Phys. J. C **76** (2016) 581,*  
1730 *arXiv: [1510.03823 \[hep-ex\]](https://arxiv.org/abs/1510.03823).*
- 1731 [26] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*  
1732 *with the ATLAS detector in the first proton–proton collisions at  $\sqrt{s} = 13 \text{ TeV}$ ,*  
1733 *ATL-PHYS-PUB-2015-027, 2015, URL: <https://cds.cern.ch/record/2037904>.*
- 1734 [27] 2021, URL: [https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTagCalibrationRecommendationsRelease21#Tools\\_for\\_Flavor\\_Tagging\\_Calibra](https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTagCalibrationRecommendationsRelease21#Tools_for_Flavor_Tagging_Calibra).

- 1737 [28] F. Cardillo et al., *Measurement of the fiducial and differential cross-section of a top*  
1738 *quark pair in association with a Z boson at 13 TeV with the ATLAS detector*, (2019),  
1739 URL: <https://cds.cern.ch/record/2672207>.
- 1740 [29] T. Chen and C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’,  
1741 *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*  
1742 *Discovery and Data Mining*, KDD ’16, ACM, 2016 785, ISBN: 978-1-4503-4232-2,  
1743 URL: <http://doi.acm.org/10.1145/2939672.2939785>.
- 1744 [30] ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 7$  TeV using*  
1745 *the ATLAS detector at the LHC*, *Eur. Phys. J. C* **71** (2011) 1630,  
1746 arXiv: [1101.2185 \[hep-ex\]](https://arxiv.org/abs/1101.2185).
- 1747 [31] G. Avoni et al.,  
1748 *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*,  
1749 *JINST* **13** (2018) P07017.
- 1750 [32] G. Aad, T. Abajyan, B. Abbott and etal, *Jet energy resolution in proton-proton collisions*  
1751 *at  $\sqrt{s} = 7$  TeV recorded in 2010 with the ATLAS detector*,  
1752 *The European Physical Journal C* **73** (2013) 2306, ISSN: 1434-6052,  
1753 URL: <https://doi.org/10.1140/epjc/s10052-013-2306-0>.
- 1754 [33] A. Collaboration, *Performance of b -jet identification in the ATLAS experiment*,  
1755 *Journal of Instrumentation* **11** (2016) P04008,  
1756 URL: <http://stacks.iop.org/1748-0221/11/i=04/a=P04008>.

- 1757 [34] *Observation of the associated production of a top quark and a Z boson at 13 TeV with*  
1758 *ATLAS*, (2020), URL: <https://cds.cern.ch/record/2722504>.
- 1759 [35] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions from Powheg*  
1760 *and Sherpa to ATLAS measurements of top pair production at 7 TeV,*  
1761 ATL-PHYS-PUB-2015-011, 2015, URL: <https://cds.cern.ch/record/2020602>.
- 1762 [36] ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11,*  
1763 ATL-PHYS-PUB-2011-009, 2011, URL: <https://cds.cern.ch/record/1363300>.

**List of contributions**

1764

1765

<sup>1766</sup> **Part VII**

<sup>1767</sup> **Appendices**

<sup>1768</sup> **A Non-prompt lepton MVA**

1769 A lepton MVA has been developed to better reject non-prompt leptons than standard cut  
1770 based selections based upon impact parameter, isolation and PID. The name of this MVA is  
1771 **PromptLeptonIso**. The full set of studies and detailed explanation can be found in [22].

1772 The decays of  $W$  and  $Z$  bosons are commonly selected by the identification of one or two  
1773 electrons or muons. The negligible lifetimes of these bosons mean that the leptons produced in the  
1774 decay originate from the interaction vertex and are thus labelled “prompt”. Analyses using these  
1775 light leptons impose strict reconstruction quality, isolation and impact parameter requirements  
1776 to remove “fake” leptons. A significant source of the fake light leptons are non-prompt leptons  
1777 produced in decays of hadrons that contain bottom (b) or charm (c) quarks. Such hadrons  
1778 typically have microscopically significant lifetimes that can be detected experimentally.

1779 These non-prompt leptons can also pass the tight selection criteria. In analyses that  
1780 involve top ( $t$ ) quarks, which decay almost exclusively into a  $W$  boson and a b quark, non-  
1781 prompt leptons from the semileptonic decay of bottom and charm hadrons can be a significant  
1782 source of background events. This is particularly the case in the selection of same-sign dilepton  
1783 and multilepton final states.

1784 The main idea is to identify non-prompt light leptons using lifetime information associated  
1785 with a track jet that matches the selected light lepton. This lifetime information is computed  
1786 using tracks contained within the jet. Typically, lepton lifetime is determined using the impact  
1787 parameter of the track reconstructed by the inner tracking detector which is matched to the recon-  
1788 structed lepton. Using additional reconstructed charged particle tracks increases the precision of

1789 identifying the displaced decay vertex of bottom or charm hadrons that produced a non-prompt  
 1790 light lepton. The MVA also includes information related to the isolation of the lepton to reject  
 1791 non-prompt leptons.

1792 **PromptLeptonIso** is a gradient boosted BDT. The training of the BDT is performed on  
 1793 leptons selected from the PowHEG+PYTHIA6 non-allhad  $t\bar{t}$  MC sample. Eight variables are used  
 1794 to train the BDT in order to discriminate between prompt and non-prompt leptons. The track  
 1795 jets that are matched to the non-prompt leptons correspond to jets initiated by b or c quarks, and  
 1796 may contain a displaced vertex. Consequently, three of the selected variables are used to identify  
 1797 b-tag jets by standard ATLAS flavour tagging algorithms. Two variables use the relationship  
 1798 between the track jet and lepton: the ratio of the lepton  $p_T$  with respect to the track jet  $p_T$  and  
 1799  $\Delta R$  between the lepton and the track jet axis. Finally three additional variables test whether the  
 1800 reconstructed lepton is isolated: the number of tracks collected by the track jet and the lepton  
 1801 track and calorimeter isolation variables. Table 57 describes the variables used to train the BDT  
 1802 algorithm. The choice of input variables has been extensively discussed with Egamma, Muon,  
 1803 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 57: A table of the variables used in the training of **PromptLeptonIso**.

1804 The output distribution of the BDT is shown in Figure A.

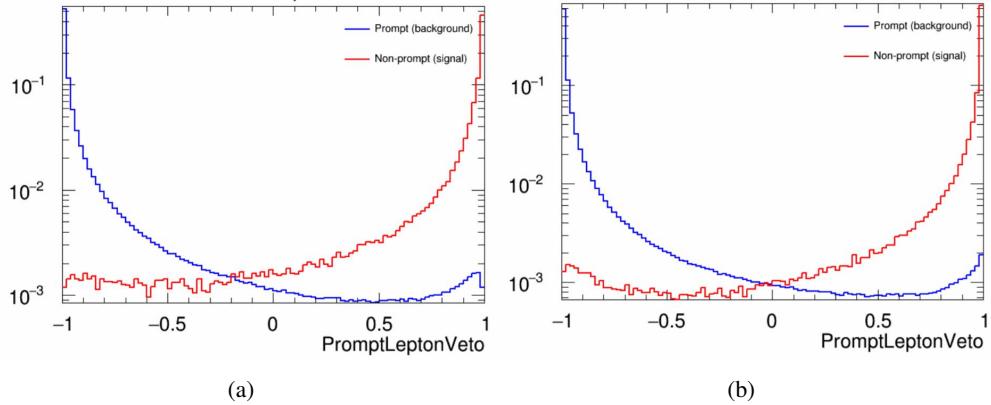


Figure A.1: Distribution of the PLV BDT discriminant for (a) electrons and (b) muons

1805 The ROC curve for the BDT response, compared to the standard `FixedCutTight` WP, is  
 1806 shown in figure A, which shows a clear improvement when using this alternate training.

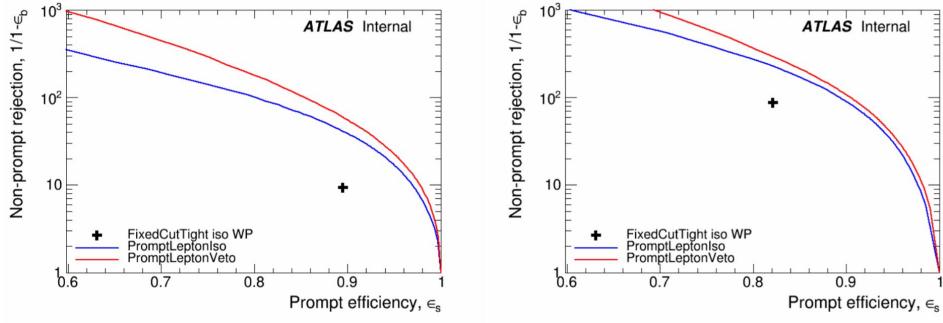


Figure A.2: ROC curves for the PLV as well as the performance of the standard `FixedCutTight` WP for (left) electrons and (right) muons

1807 A cutoff value of -0.7 for electrons and -0.5 for muons are chosen as the WPs for this  
 1808 MVA, based on an optimisation of  $S/\sqrt{B}$  performed in the preselection regions of the  $t\bar{t}H - ML$   
 1809 analysis, which have a signature similar to that of this analysis.

1810        The efficiency of the tight `PromptLeptonIso` working point is measured using the tag  
1811        and probe method with  $Z \rightarrow \ell^+ \ell^-$  events. Such calibration are performed by analysers from  
1812        this analysis in communication with the Egamma and Muon combined performance groups. The  
1813        scale factor are approximately 0.92 for  $10 < p_T < 15$  GeV, and averaging at 0.98 to 0.99 for  
1814        higher  $p_T$  leptons. An extra systematic is applied to muons within  $\Delta R < 0.6$  of a calorimeter  
1815        jet, since there is a strong dependence on the scale factor due to the presence of these jets. For  
1816        electrons, the dominant systematics is coming from pile-up dependence. Overall the systematics  
1817        are a maximum of 3% at low  $p_T$  and decreasing at a function of  $p_T$ .

<sup>1818</sup> **B Supplementary WZ + Heavy Flavor Studies**

<sup>1819</sup> **B.1 Non-prompt CR Modelling**

1820 In order to further validate the modeling in each of the non-prompt CRs, additional  
 1821 kinematic plots are made in the Z+jets CR and  $t\bar{t}$  CR in each of the continuous b-tag regions,  
 1822 after the correction factors detailed in Section 12.3 have been applied.

1823 In the case of the Z+jets CR, the  $p_T$  spectrum of the lepton originating from the W  
 1824 candidate is shown, as this is the distribution used to extract the scale factor applied to Z+jets.  
 1825 These plots are shown in Figures B.1 and B.2.

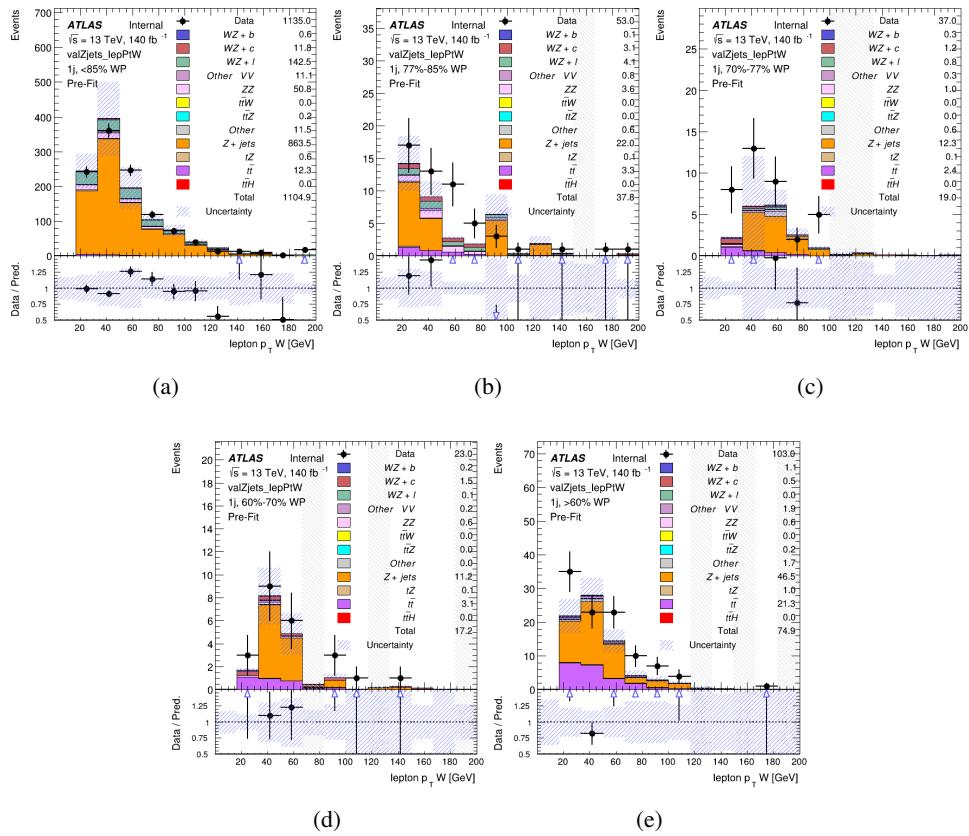


Figure B.1: Comparisons between the data and MC distributions of the  $p_T$  of the lepton originating from the W-candidate in the Z+jets CR for each of the 1-jet b-tag working point regions

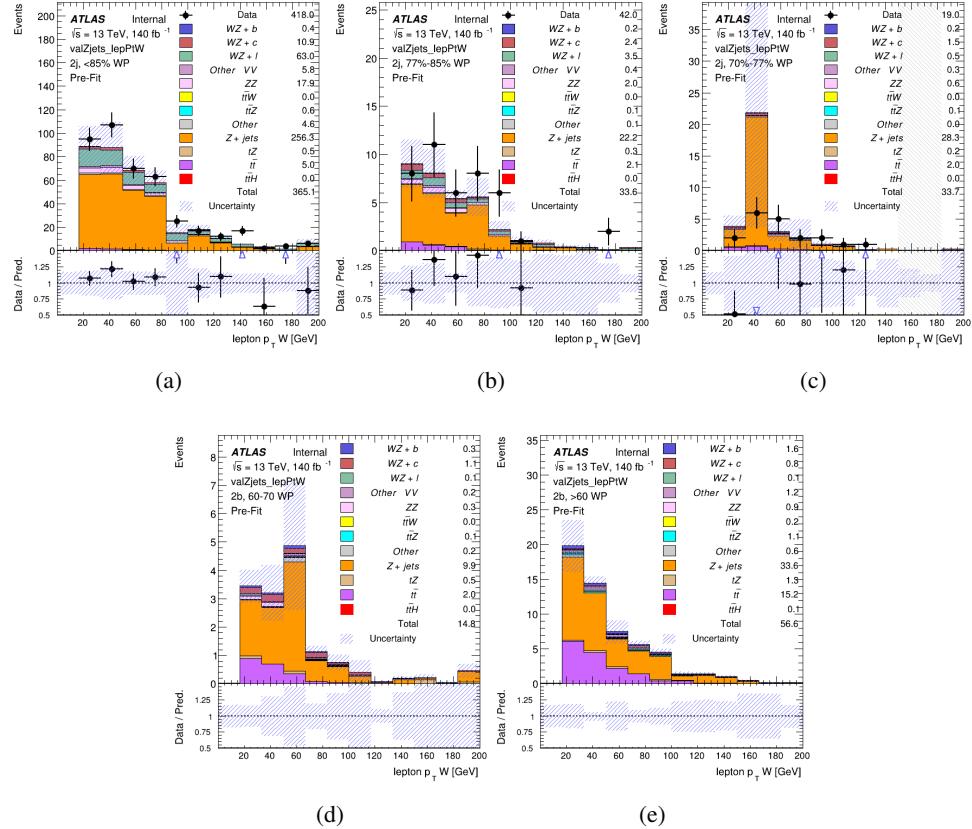


Figure B.2: Comparisons between the data and MC distributions of the  $p_T$  of the lepton originating from the W-candidate in the Z+jets CR for each of the 2-jet b-tag working point regions

1826 The same is shown for the  $t\bar{t}$  CR, but the  $p_T$  of the OS lepton is used instead as a  
 1827 representation of the modeling, as the lepton from the W is not well defined for  $t\bar{t}$  events. These  
 1828 plots are shown in Figures B.3 and B.4.

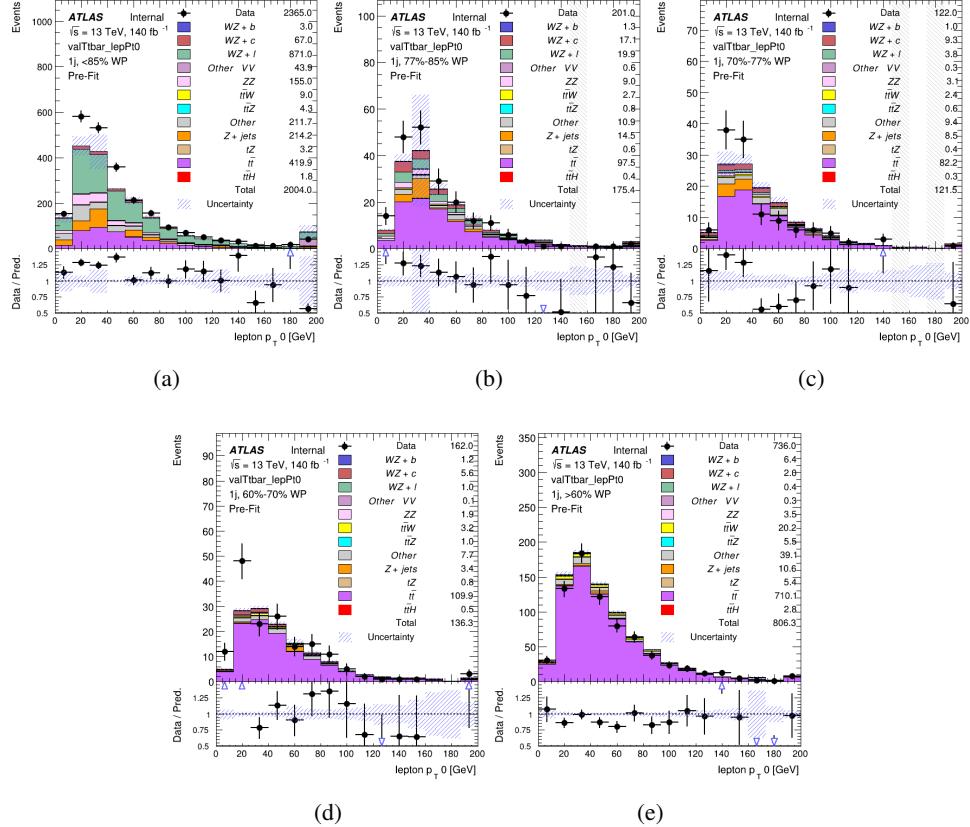


Figure B.3: Comparisons between the data and MC distributions of the  $p_T$  of the OS lepton in the  $t\bar{t}$  CR for each of the 1-jet b-tag working point regions

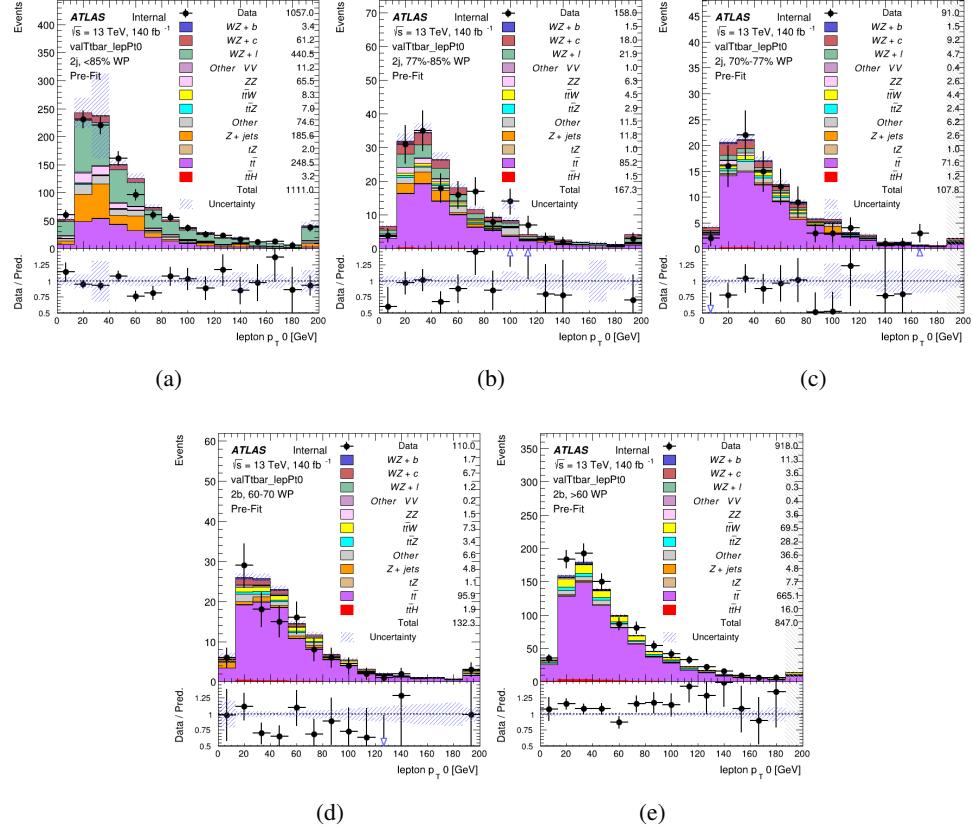


Figure B.4: Comparisons between the data and MC distributions of the  $p_T$  of the OS lepton in the  $t\bar{t}$  CR for each of the 2-jet b-tag working point regions

<sup>1829</sup> **B.2 tZ Interference Studies**

1830 Because it includes an on-shell Z boson as well as a b-jet and W from the top decay, tZ  
1831 production represents an identical final state to WZ + b-jet. This implies the possibility of matrix  
1832 level interference between these two processes not accounted for in the Monte Carlo simulations,  
1833 which consider the two processes independently. Truth level studies are performed in order to  
1834 estimate the impact of these interference effects.

1835 In order to estimate the matrix level interference effects between tZ and WZ + b-jet, two  
1836 different sets of simulations are produced using MadGraph 5 [**Madgraph**] - one which simulates  
1837 these two processes independently, and another where they are produced simultaneously, such  
1838 that interference effects are present. These two sets of samples are then compared, and the  
1839 difference between them can be taken to represent any interference effects.

1840 MadGraph simulations of 10,000 tZ and 10,000 WZ + b-jet events are produced, along  
1841 with 20,000 events where both are present, in the fiducial region where three leptons and at least  
1842 one jet are produced.

1843 A selection mimicking the preselection used in the main analysis is applied to the samples:  
1844 The SS leptons are required to have  $p_T > 20$  GeV, and  $> 10$  GeV is required for the OS lepton.  
1845 The associated b-jet is required to have  $p_T > 25$  GeV, and all physics objects are required to fall  
1846 in a range of  $|\eta| < 2.5$ .

1847 The kinematics of these samples after the selection has been applied are shown below:

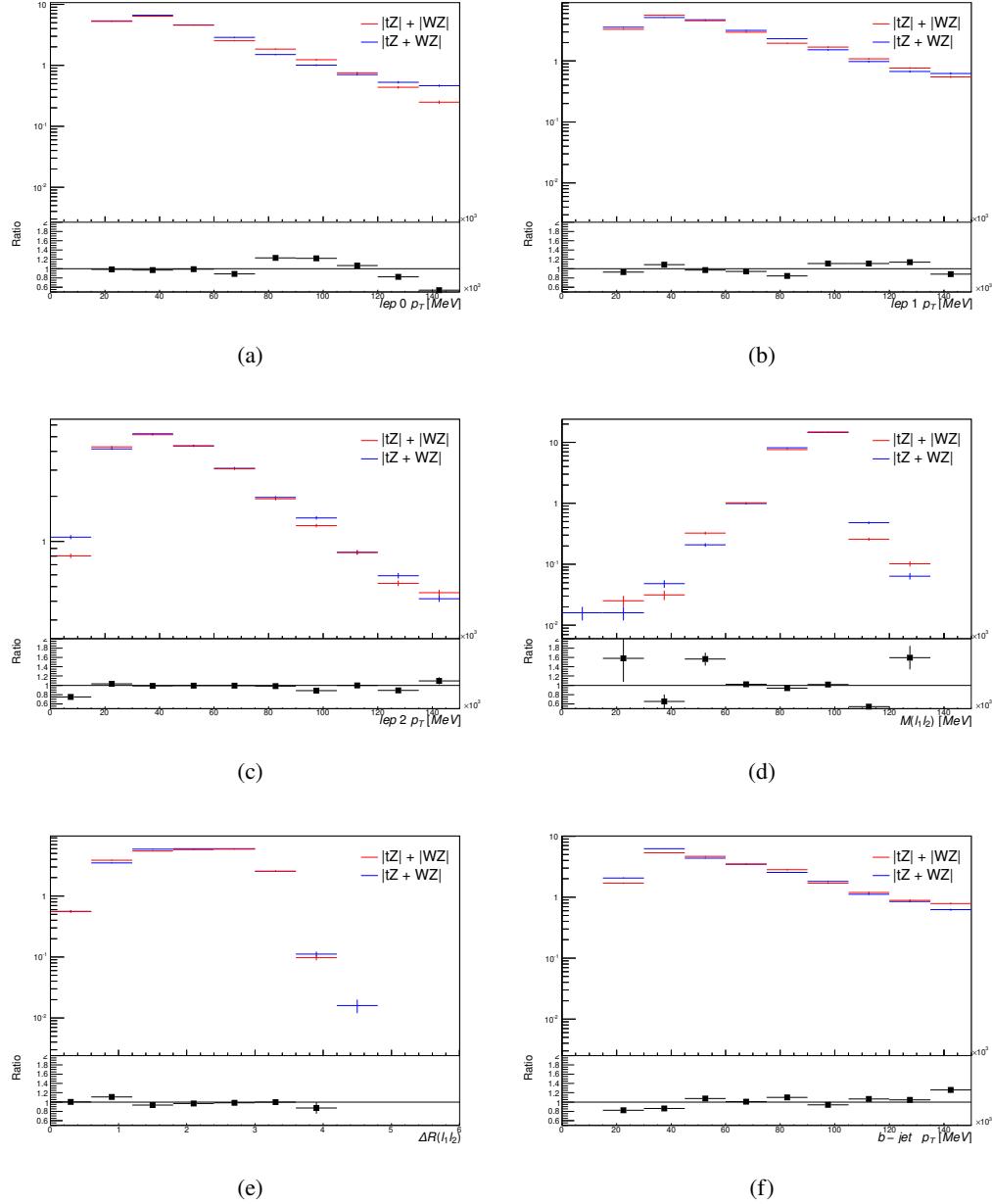


Figure B.5: Comparisons between (a) the  $p_T$  of the lepton from the W, (b) and (c) show the  $p_T$  of the lepton from the Z, (d) the invariant mass of the Z-candidate, (e)  $\Delta R$  of the leptons from the Z, and (f) the  $p_T$  of the b-jet, for WZ and tZ events generated with interference effects (blue) and without interference effects (red).

1848        The overall cross-section of the two methods agree within error, and no significant differ-  
1849        ences in the kinematic distributions are seen. It is therefore concluded that interference effects  
1850        do not significantly impact the results.

<sup>1851</sup> **B.3 Alternate tZ Inclusive Fit**

1852 **B.3.1 tZ Inclusive Fit**

1853 While tZ is often considered as a distinct process from WZ + b, this could also be considered part  
 1854 of the signal. Alternate studies are performed where, using the same framework as the nominal  
 1855 analysis, a measurement of WZ + b is performed that includes tZ as part of WZ+b.

1856 Because of this change, the tZ CR is no longer necessary, and only the five pseudo-  
 1857 continuous b-tag regions are used in the fit. Further, systematics related to the tZ cross-section  
 1858 are removed from the fit, as they are now encompassed by the normalization measurement of  
 1859 WZ + b. All other systematic uncertainties are carried over from the nominal analysis.

1860 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 fit,  
 1861 with an expected significance of  $4.0\sigma$ .

1862 The impact of the predominate systematics are summarized in Table 58.

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 + charm cross-section	-0.04	0.05
Other Diboson + b cross-section	-0.04	0.04
WZ cross-section - QCD scale	-0.04	0.03
t̄t 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 58: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

1863    **B.3.2 Floating tZ**

1864    In order to quantify the impact of the tZ uncertainty on the fit, an alternate fit strategy is used  
1865    where the tZ normalization is allowed to float. This normalization factor replaces the cross-  
1866    section uncertainty on tZ, and all other parameters of the fit remain the same.

1867           An uncertainty of 17% on the normalization of tZ is extracted from the fit, compared to a  
1868    theory uncertainty of 15% applied to the tZ cross-section. The measured uncertainties on WZ  
1869    remain the same.

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

<sup>1870</sup> The following section provides details of the various MVAs as well as a few studies performed  
<sup>1871</sup> in support of this analysis, exploring alternate decisions and strategies.

<sup>1873</sup> **C.1 Higgs Reconstruction Model Details**

<sup>1874</sup> **C.1.1 b-jet Identification Features - 2lSS**

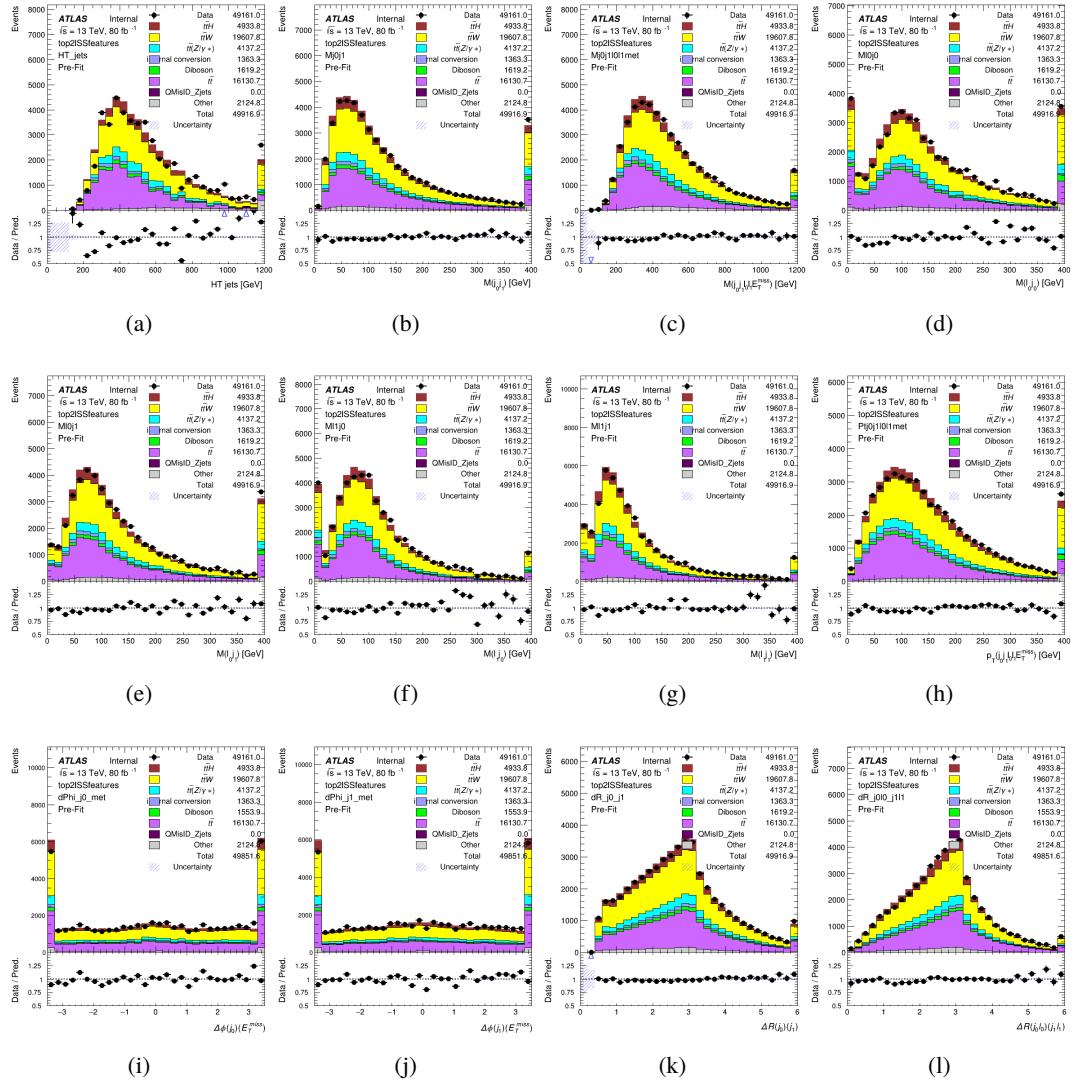


Figure C.1: Input features for top2ISS

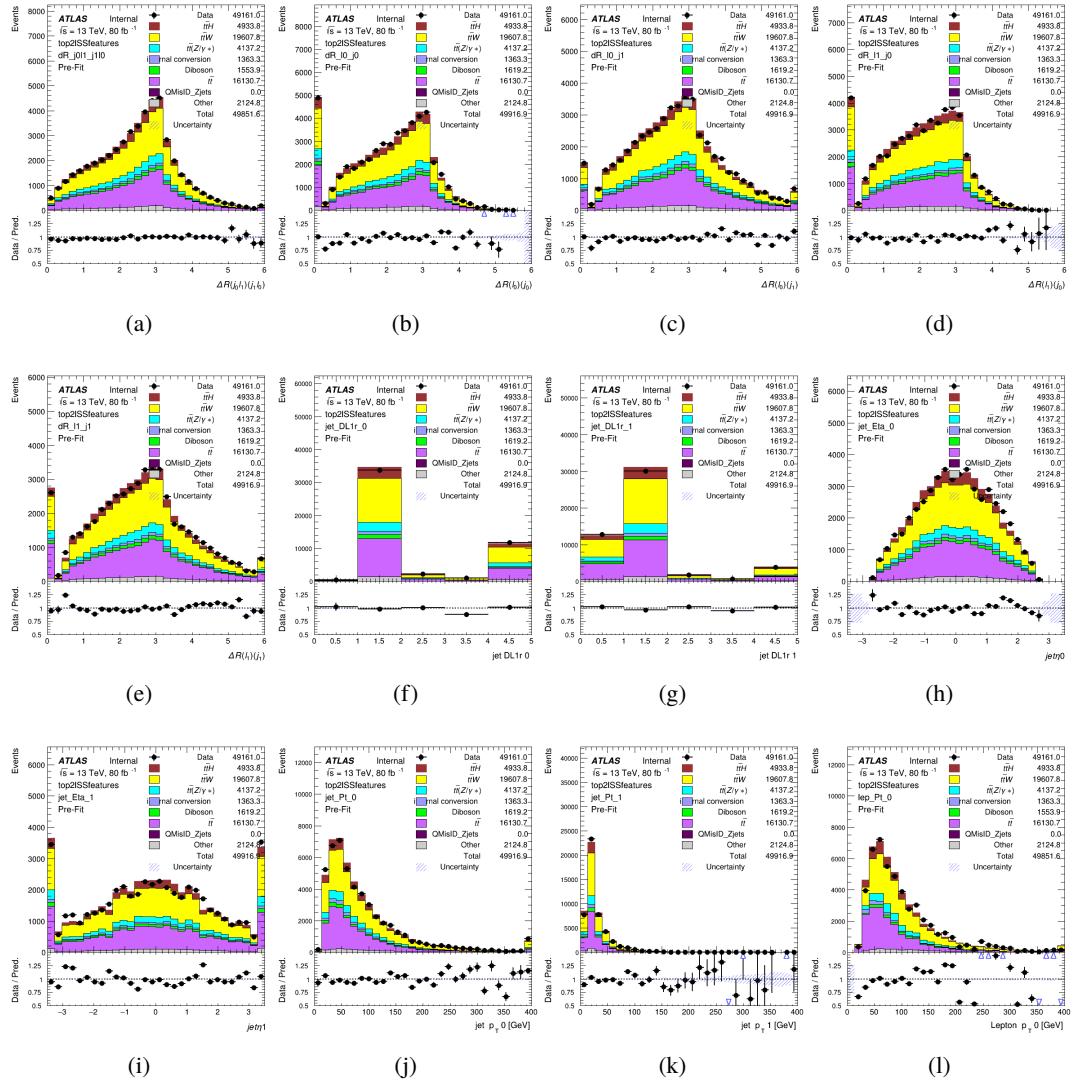


Figure C.2: Input features for top2lSS

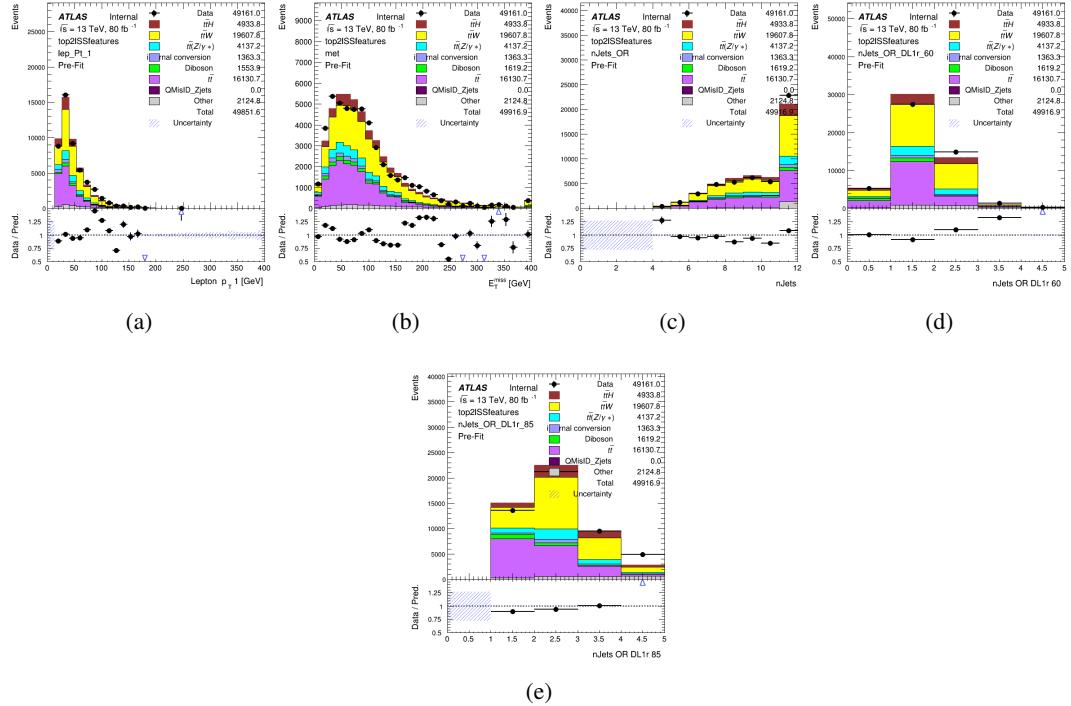


Figure C.3: Input features for top2ISS

1875 **C.1.2 b-jet Identification Features - 3l**

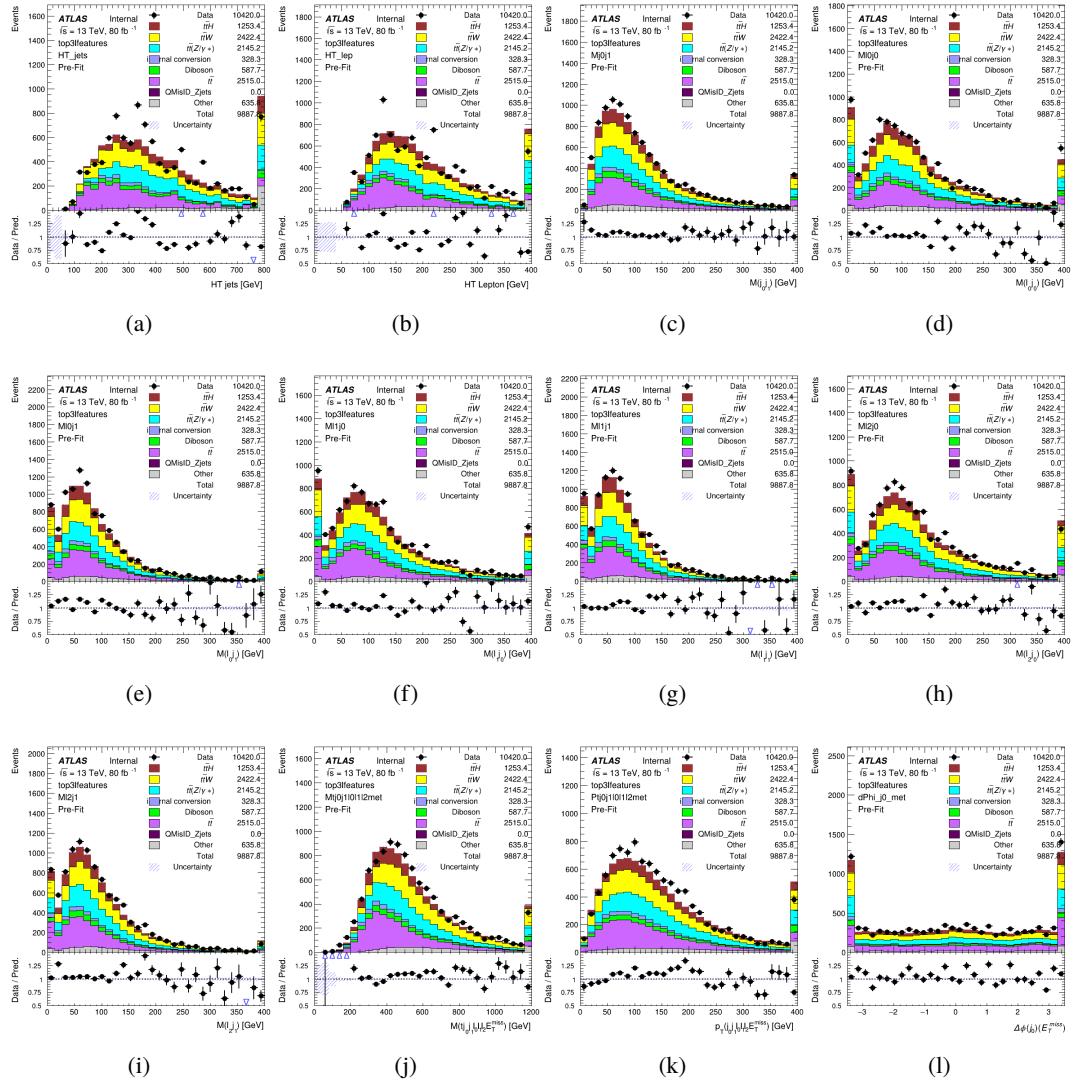


Figure C.4: Input features for top3l

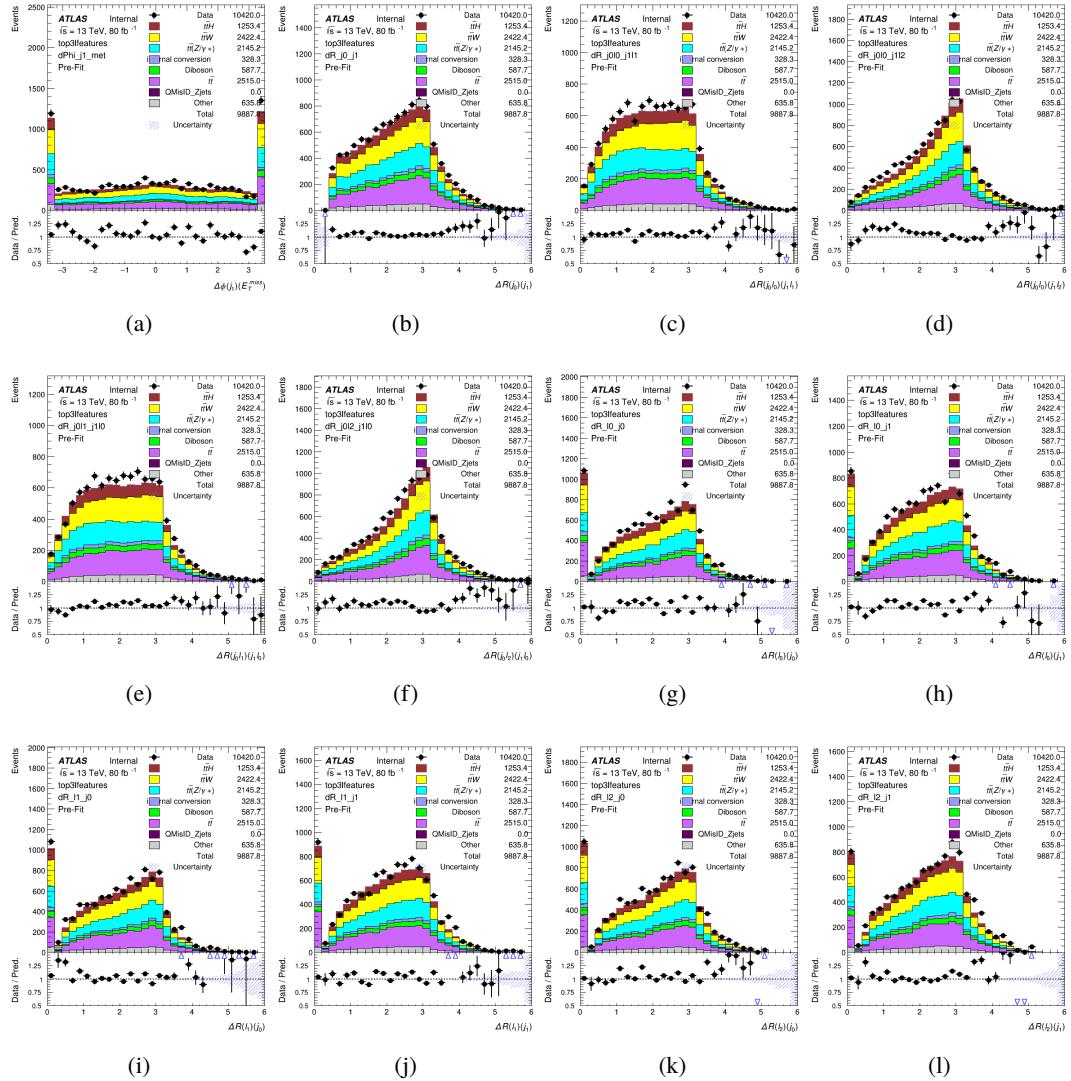


Figure C.5: Input features for top31

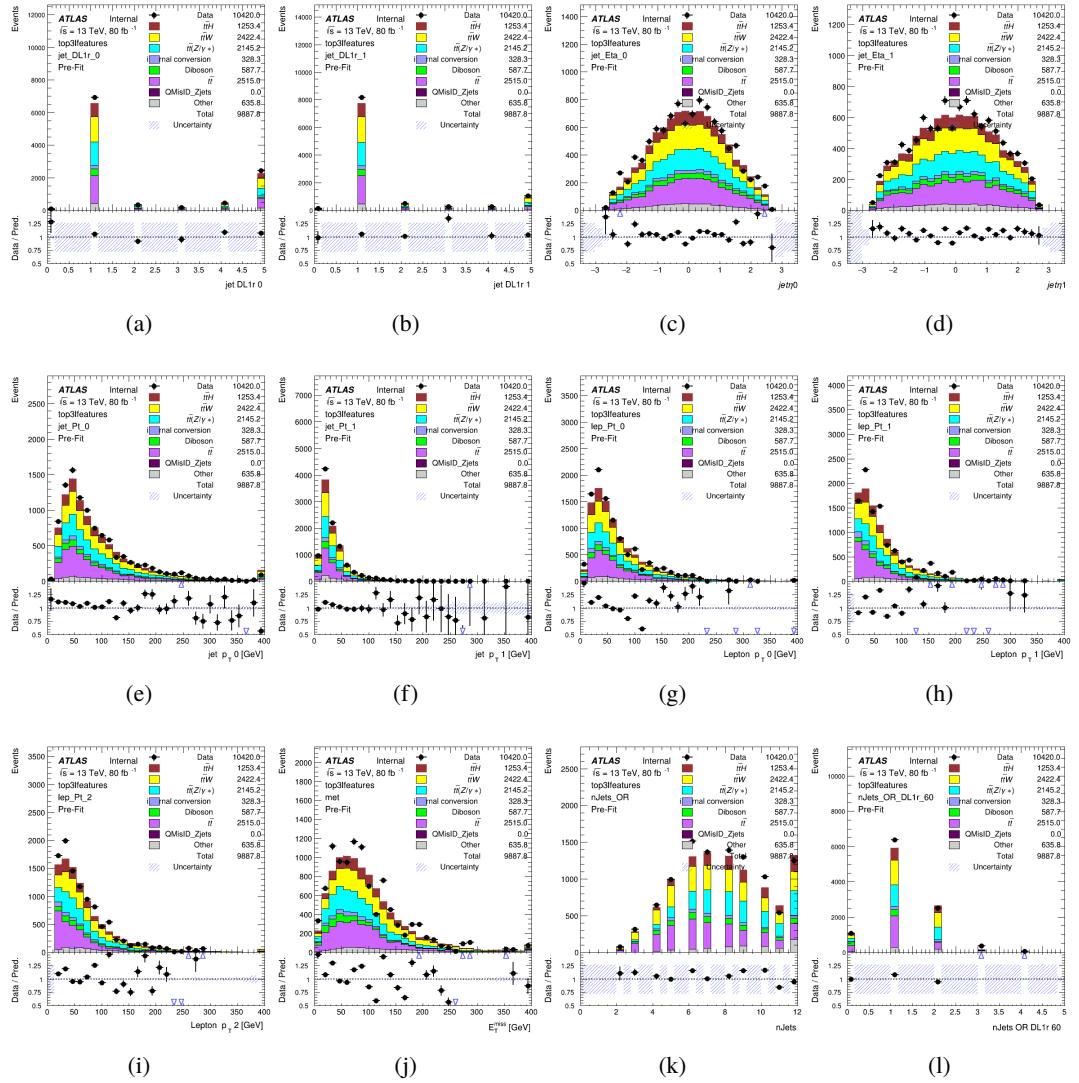
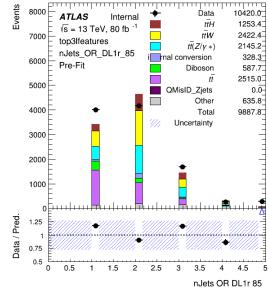


Figure C.6: Input features for top3l



(a)

Figure C.7: Input features for top3l

<sup>1876</sup> **C.1.3 Higgs Reconstruction Features - 2lSS**

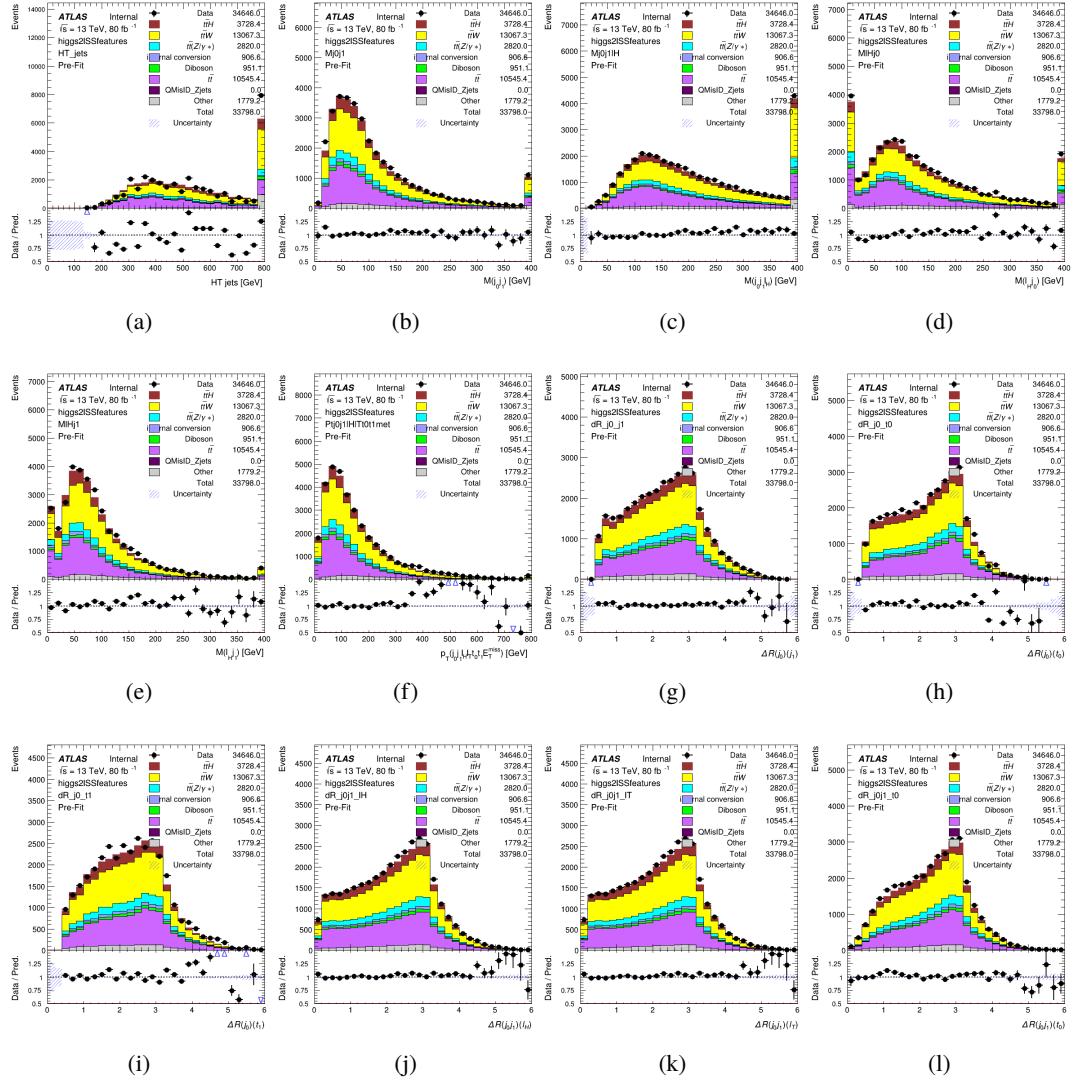


Figure C.8: Input features for higgs2ISST

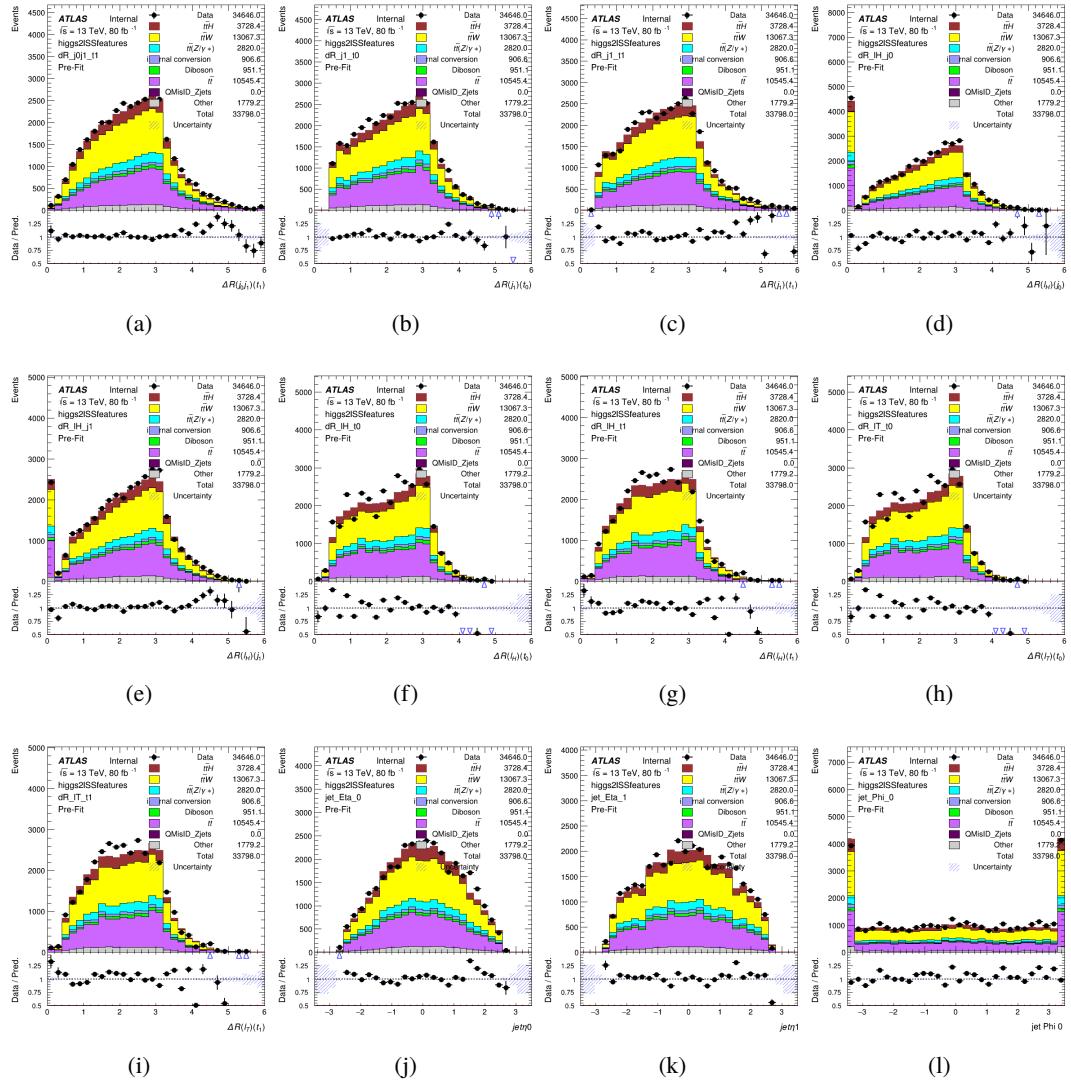


Figure C.9: Input features for higgs2lSS

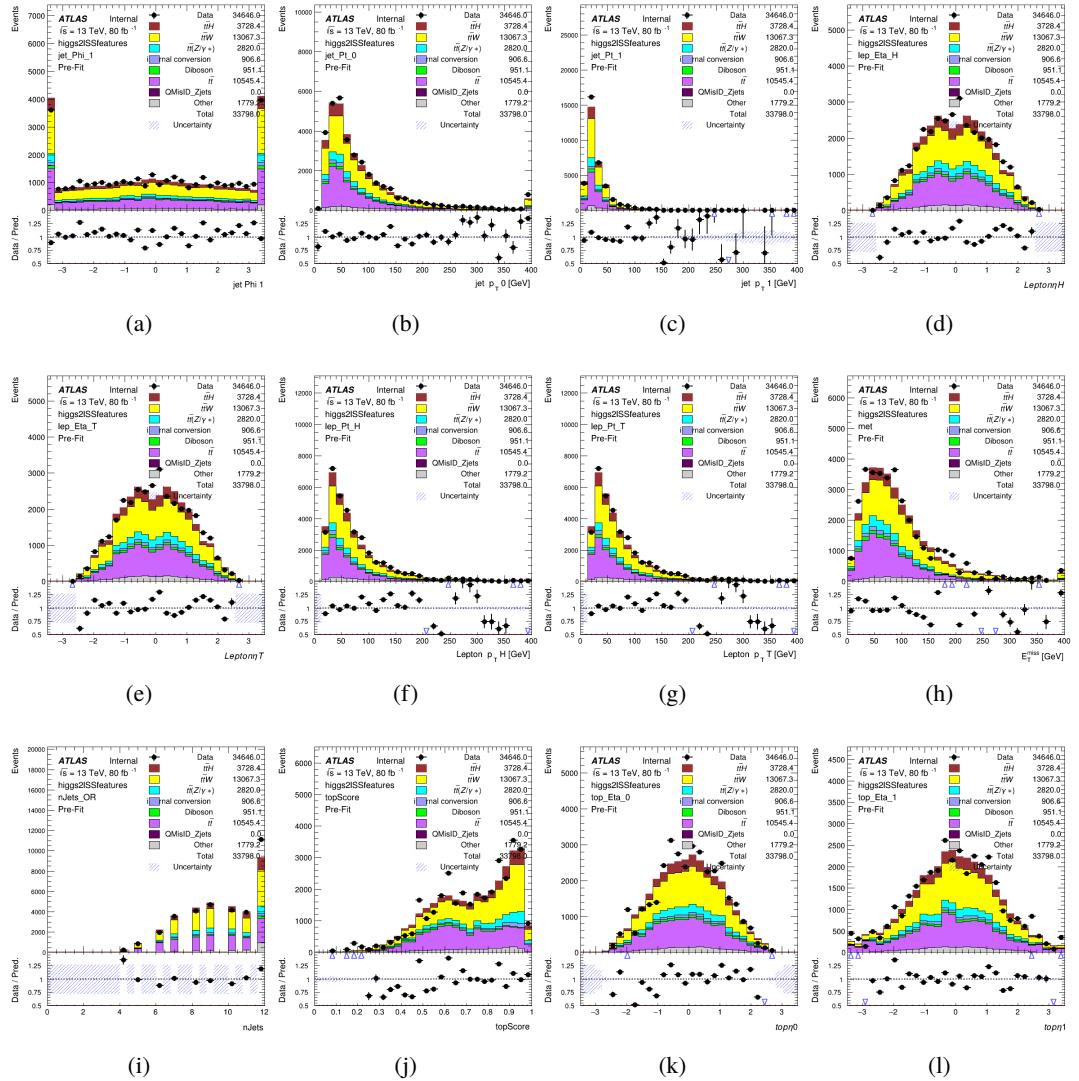


Figure C.10: Input features for higgs2lSS

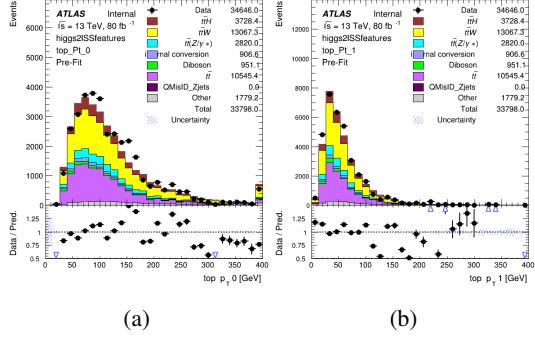


Figure C.11: Input features for higgs2lSS

<sup>1877</sup> **C.1.4 Higgs Reconstruction Features - 3lS**

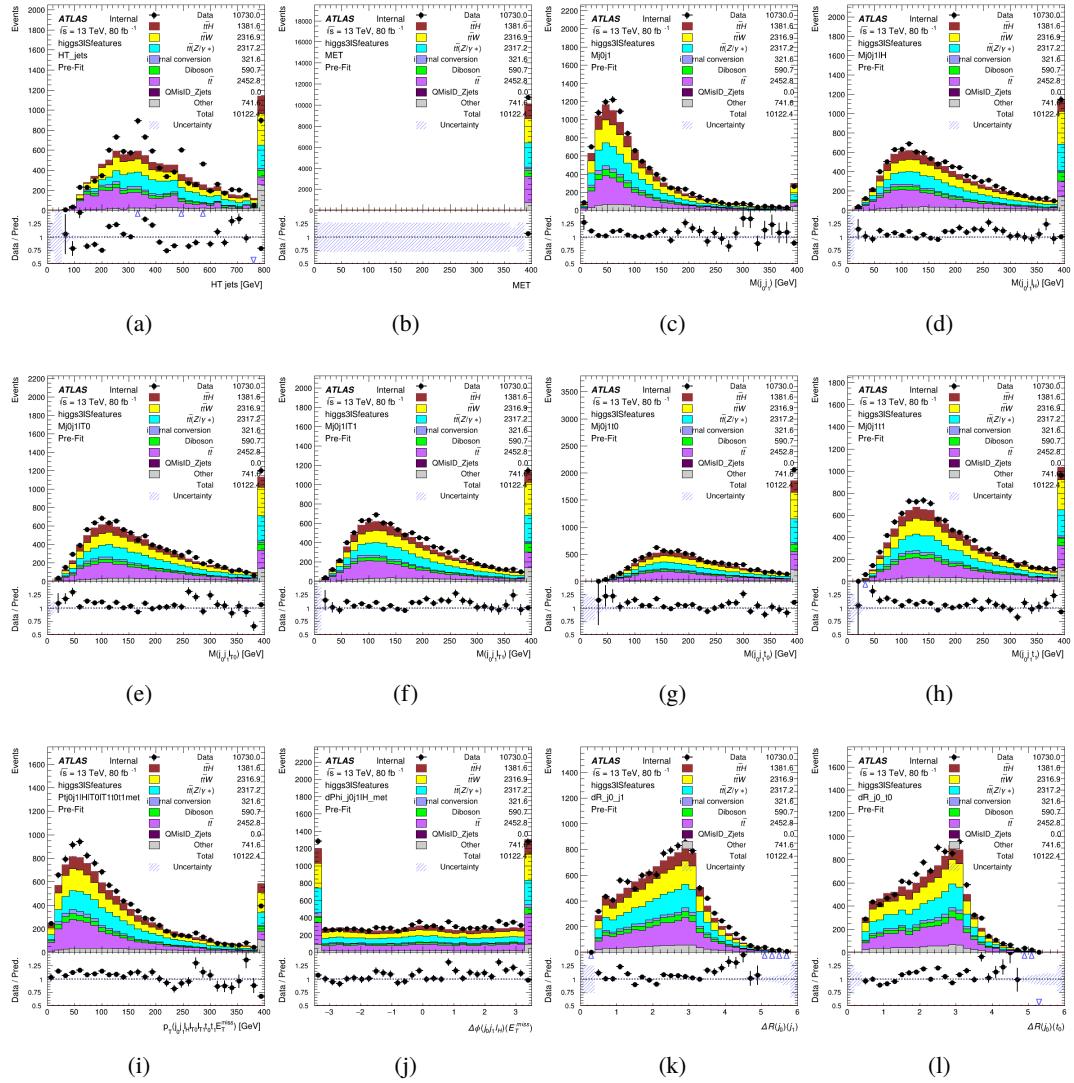


Figure C.12: Input features for higgs3S1S

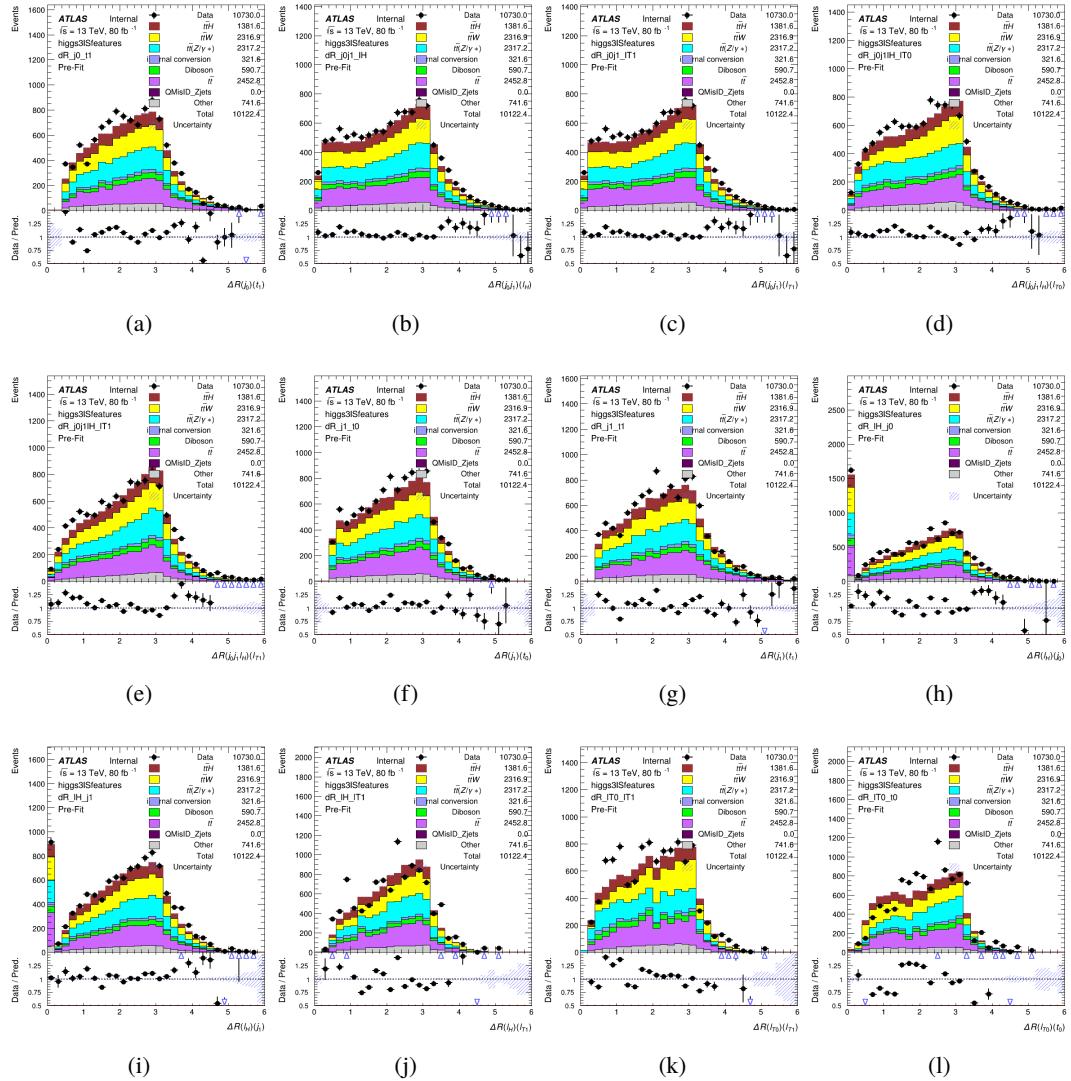


Figure C.13: Input features for higgs3lS

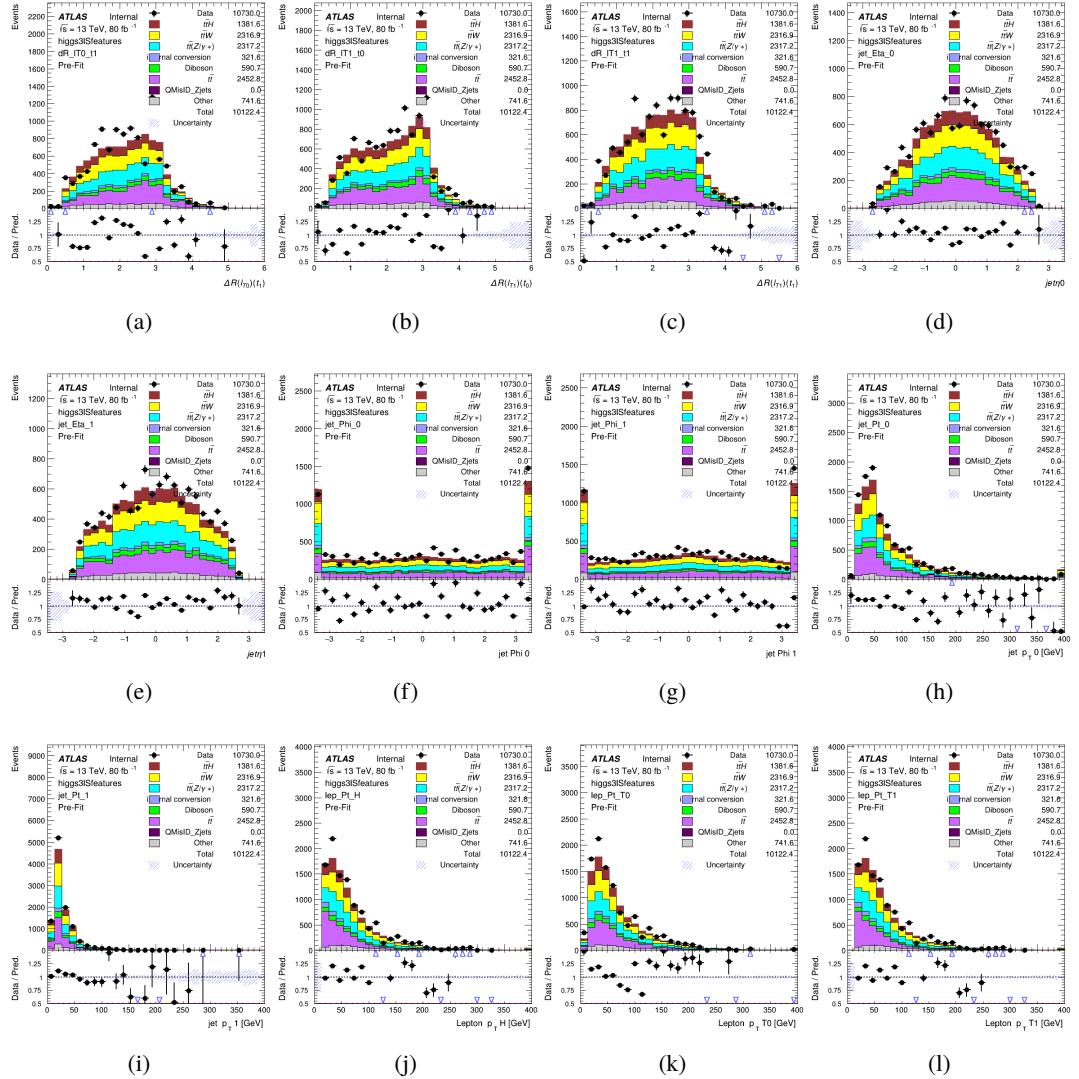


Figure C.14: Input features for higgs3lS

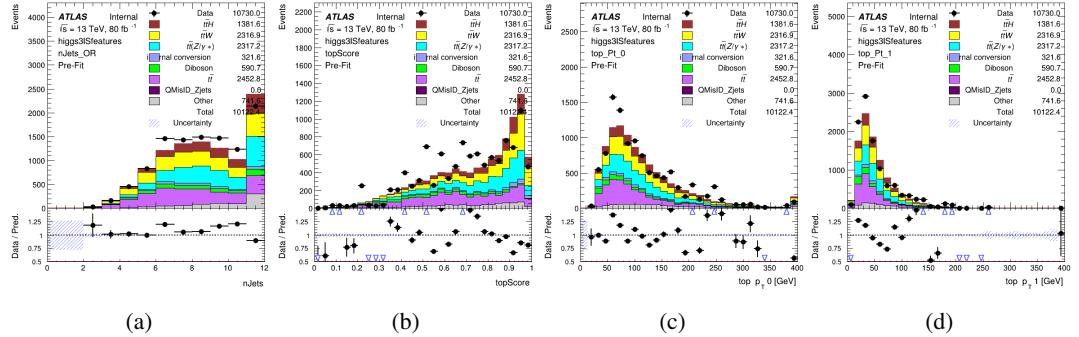


Figure C.15: Input features for higgs3IS

<sup>1878</sup> **C.1.5 Higgs Reconstruction Features - 3lF**

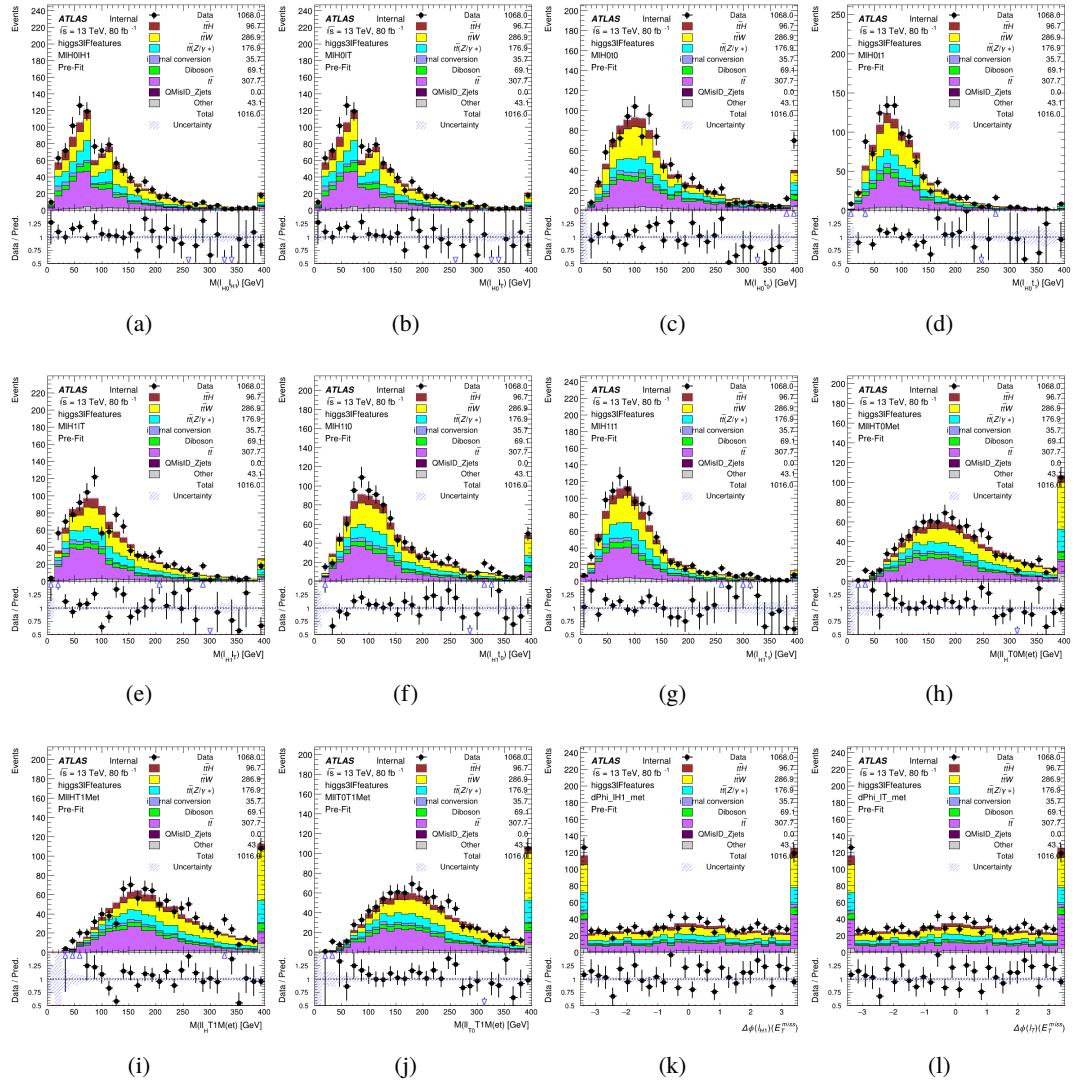


Figure C.16: Input features for higgs3IF

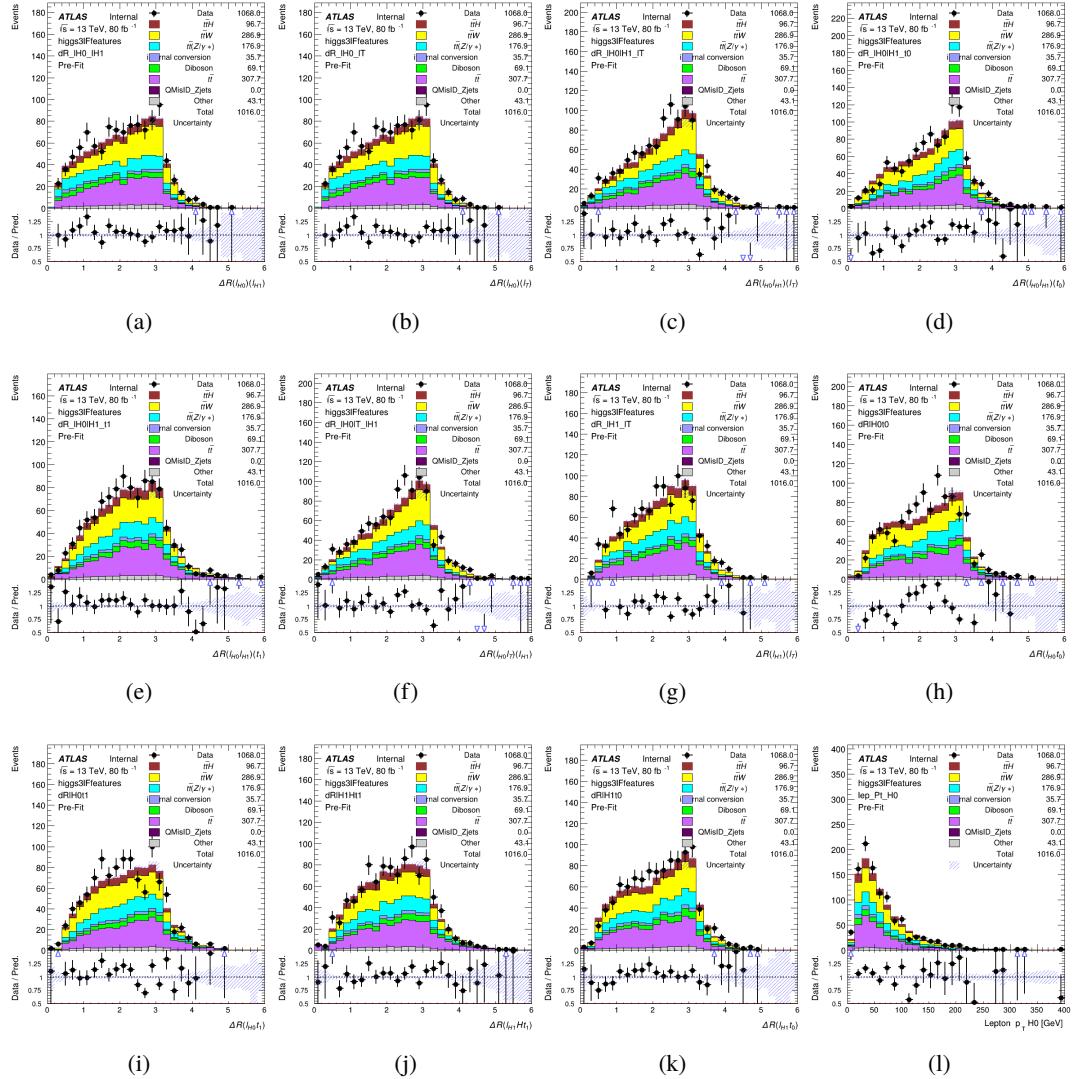


Figure C.17: Input features for higgs3IF

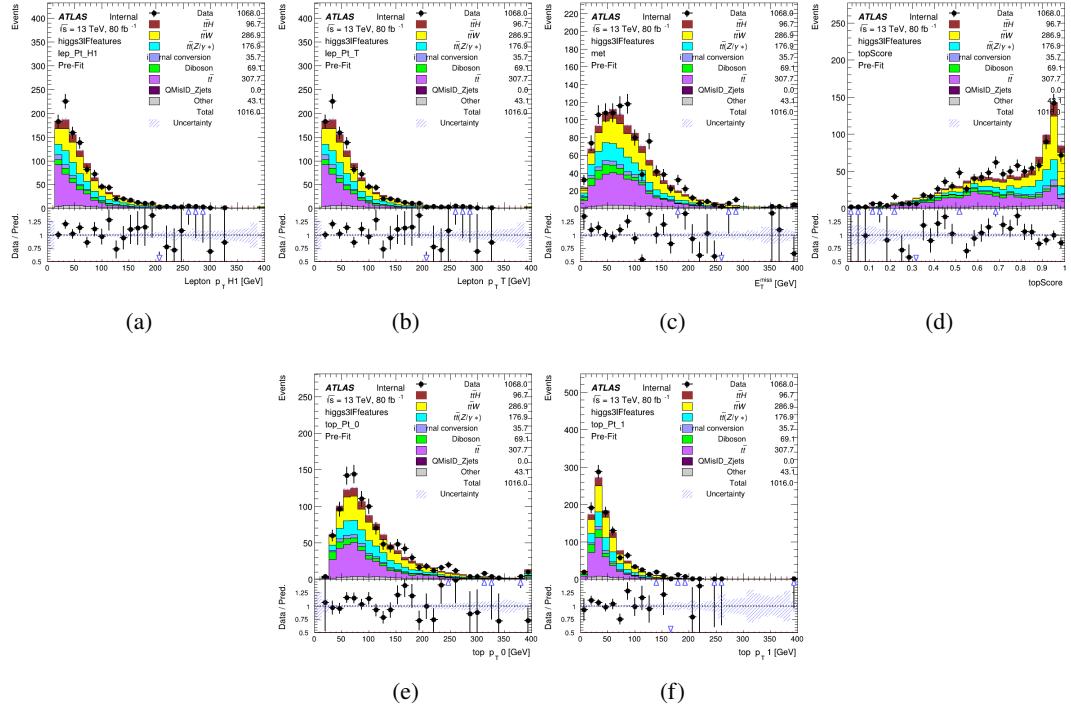


Figure C.18: Input features for higgs3lF

<sup>1879</sup> **C.2 Background Rejection MVA Details**

<sup>1880</sup> **C.2.1 Background Rejection MVA Features - 2ISS**

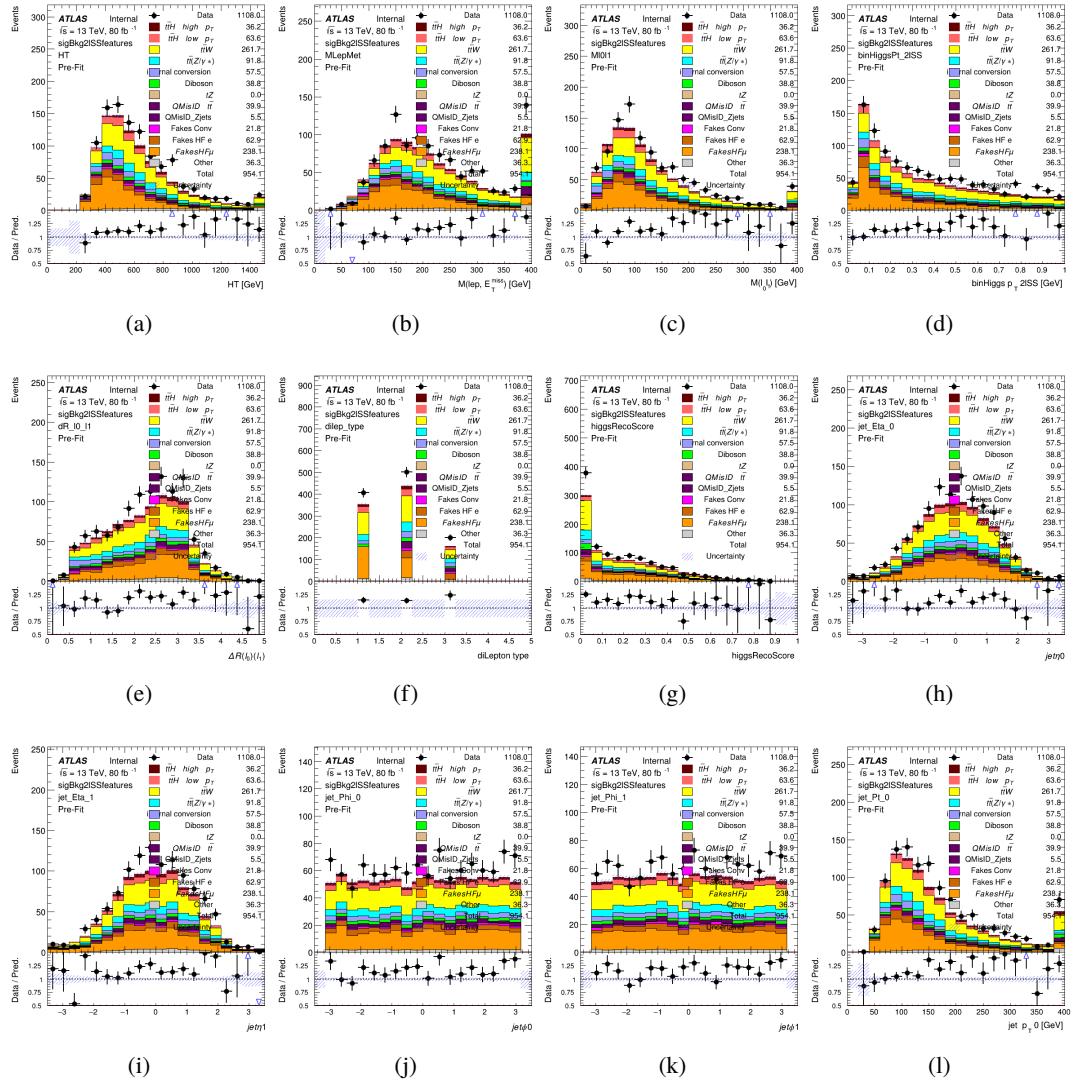


Figure C.19: Input features for sigBkg2ISS

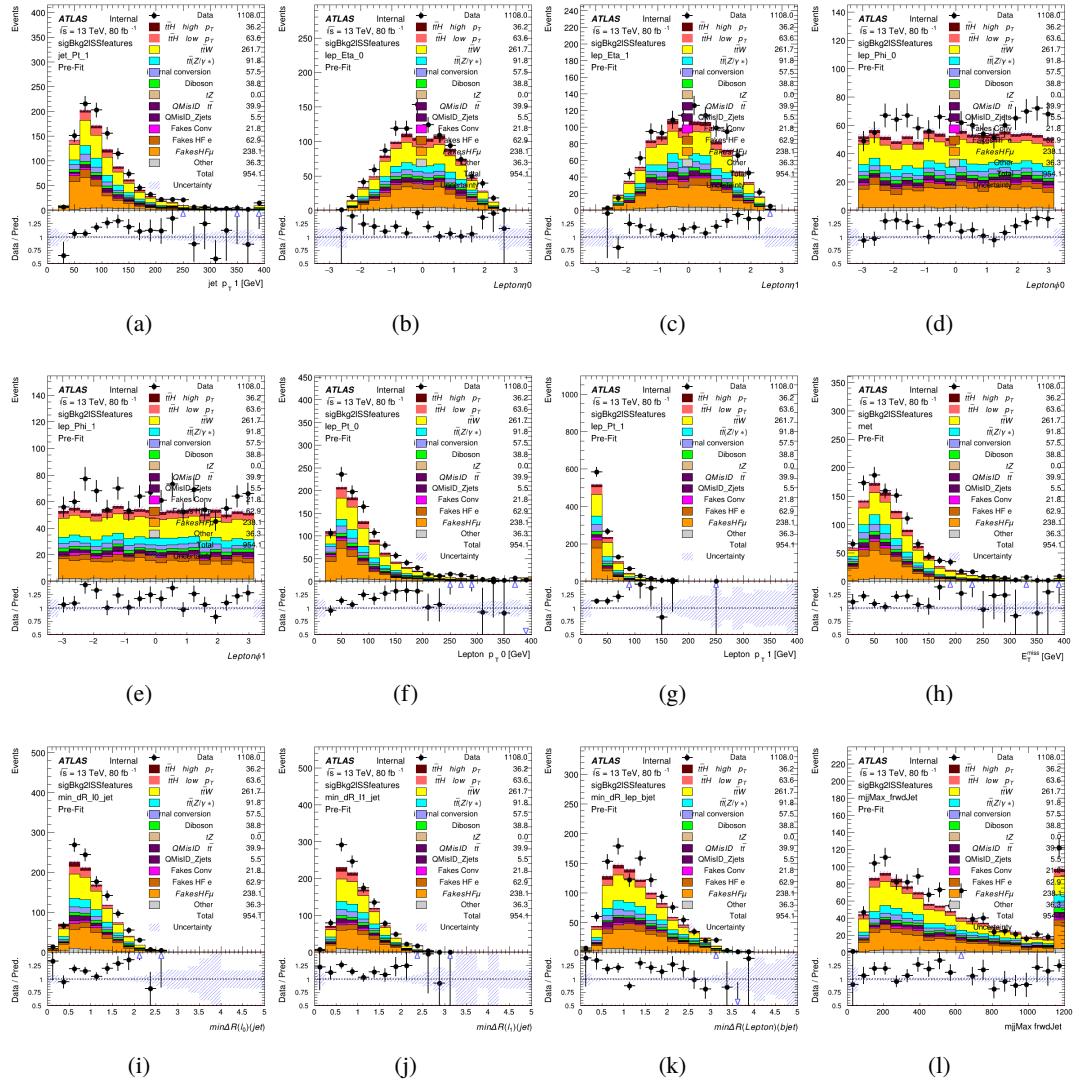


Figure C.20: Input features for sigBkg2lSS

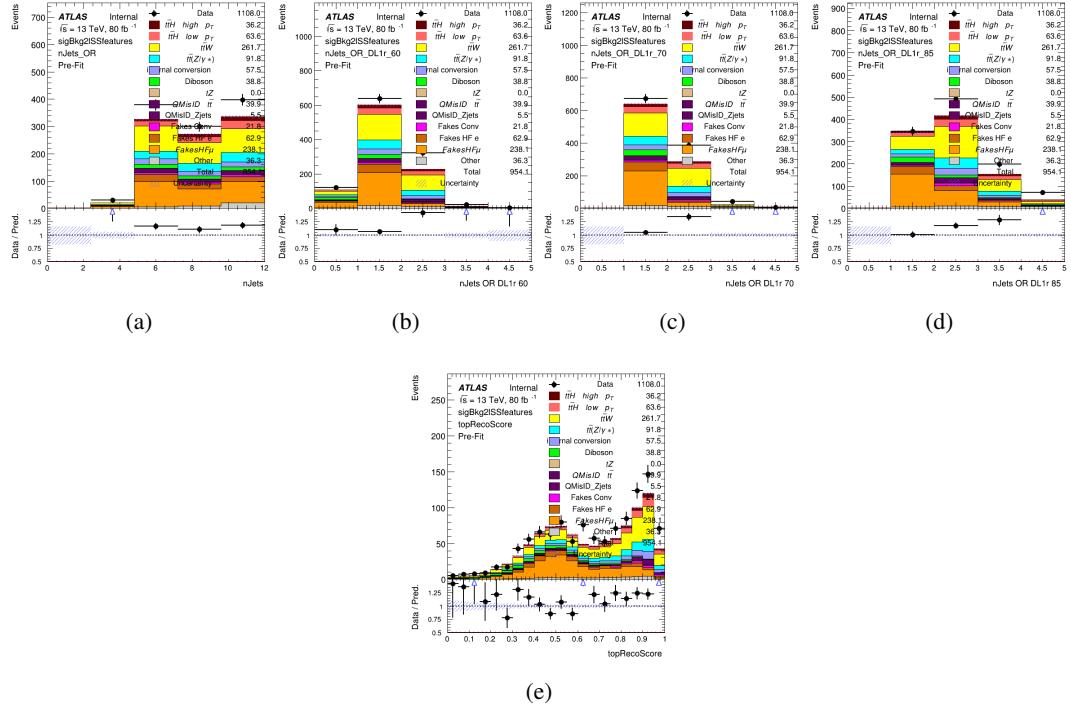


Figure C.21: Input features for sigBkg2lSS

1881 **C.2.2 Background Rejection MVA Features - 31**

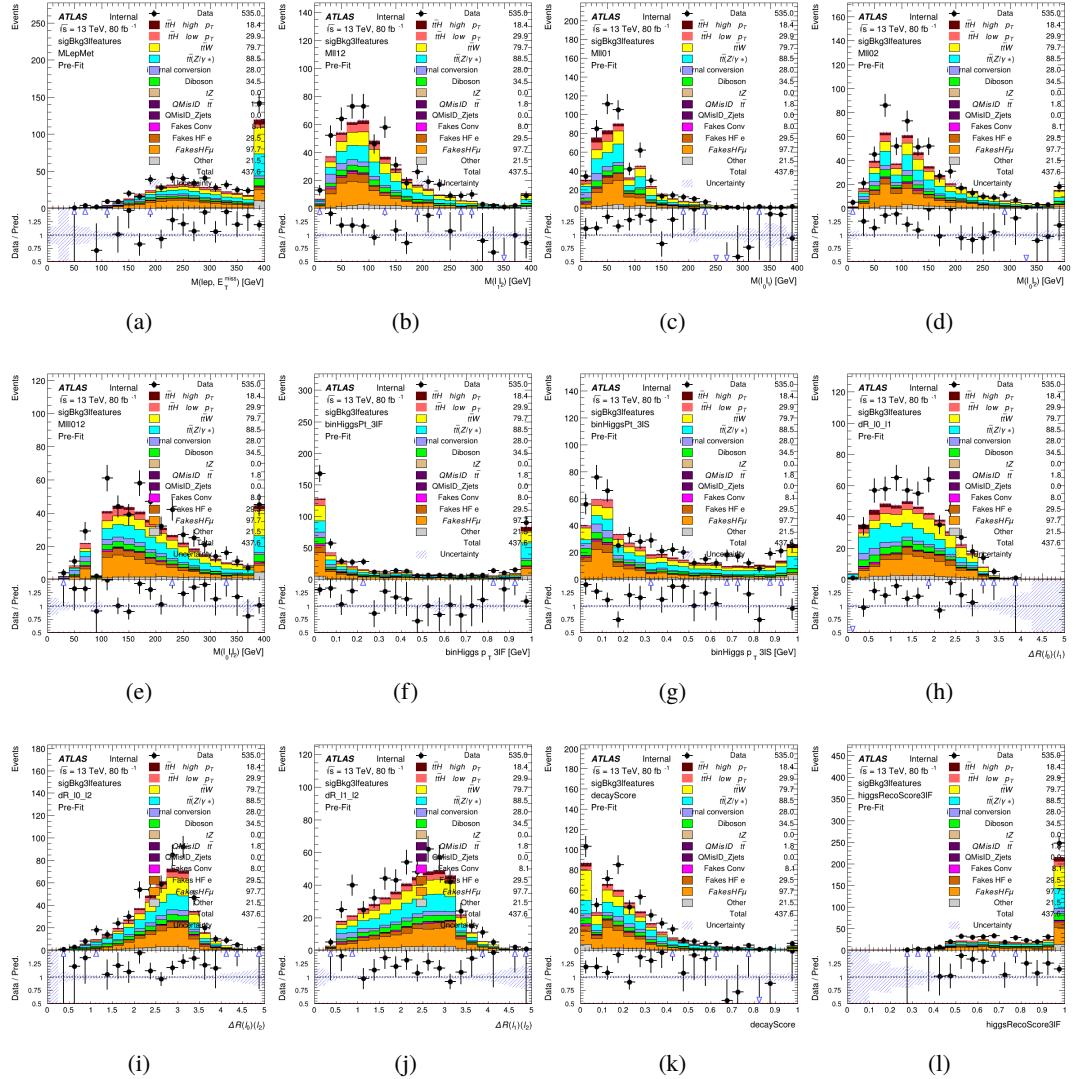


Figure C.22: Input features for sigBkg3l

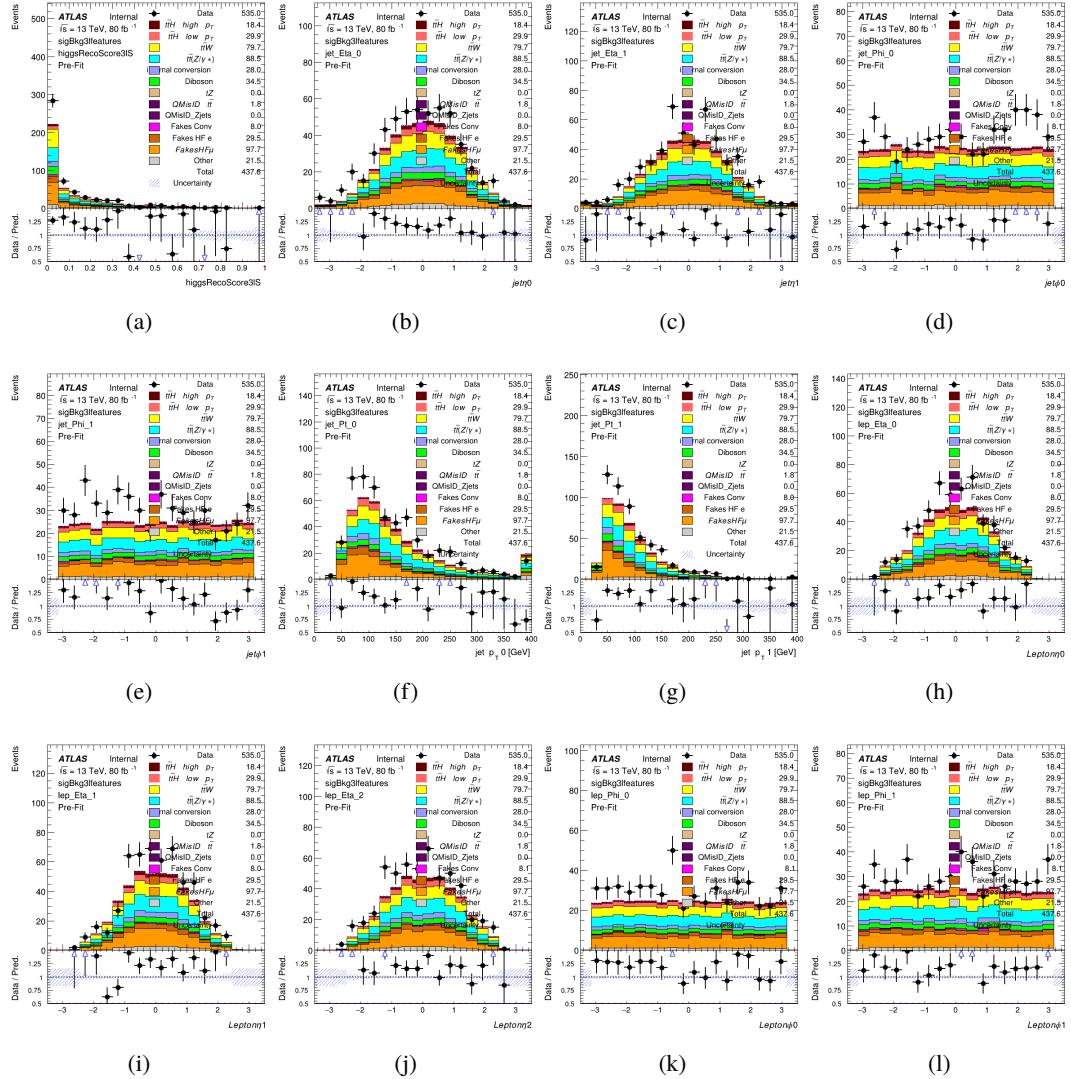


Figure C.23: Input features for sigBkg3l

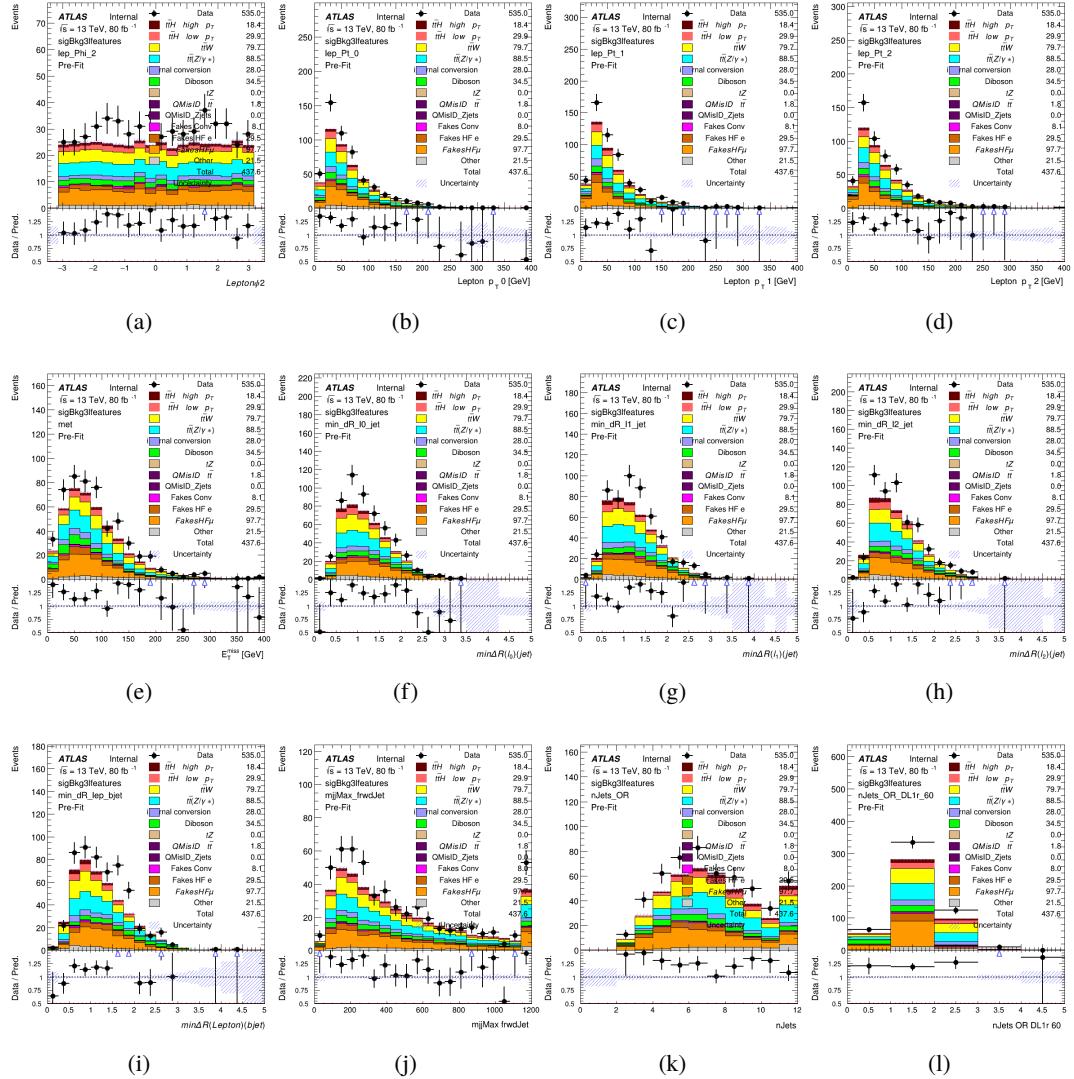


Figure C.24: Input features for sigBkg3l

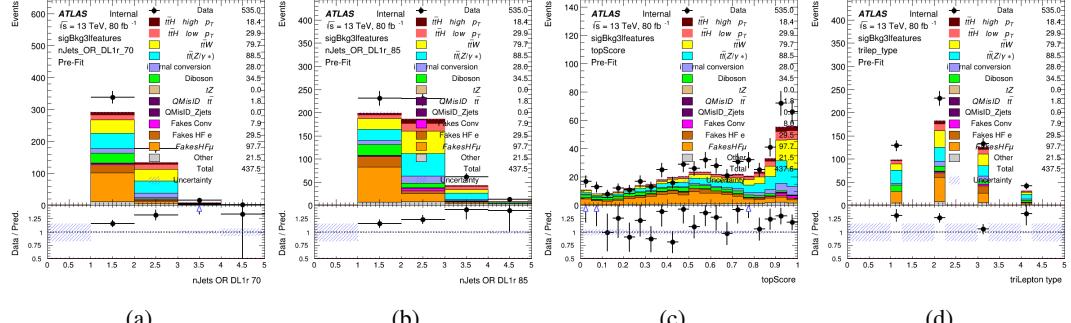


Figure C.25: Input features for sigBkg3l

### 1882 C.3 Truth Level Studies

1883 Attempts to identify the decay products of the Higgs are motivated by the ability to reconstruct  
 1884 the Higgs momentum based on their kinematics. In order to demonstrate that this is case, the  
 1885 kinematics of reconstructed objects that are truth matched to the Higgs decay are used as inputs  
 1886 to a neural network which is designed to predict of the momentum of the Higgs. This is done in  
 1887 the 2lSS channel, as it proves to be the most challenging for  $p_T$  reconstruction.

1888 Only leptons and jets which are truth matched to the Higgs are used as inputs for the  
 1889 model; events where the lepton and both jets are not reconstructed are not included. The model  
 1890 uses the same feature set and network architecture as the  $p_T$  prediction model used in the main  
 1891 analysis, as described in Section 18.5.1.

1892 The results of the model are summarized below:

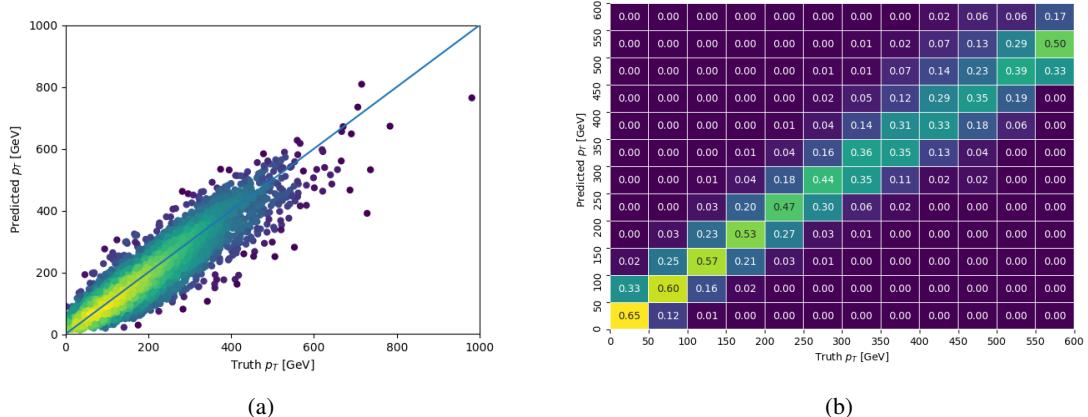


Figure C.26: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 2lSS  $t\bar{H}$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

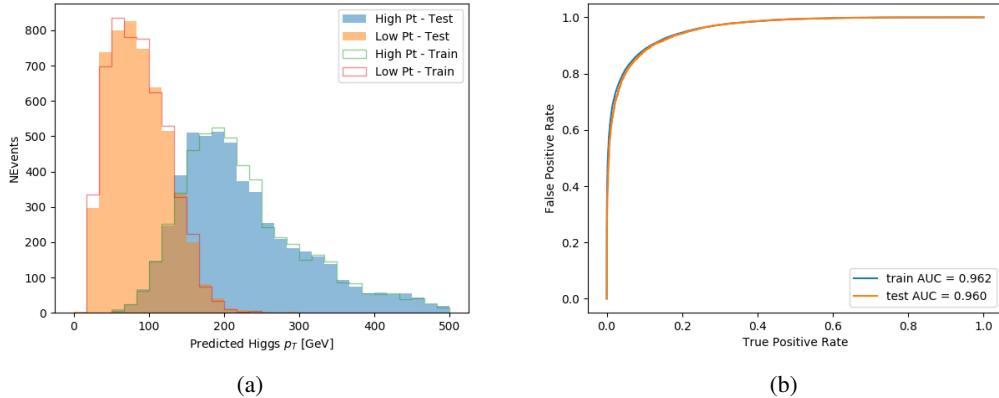


Figure C.27: (a) shows the reconstructed Higgs  $p_T$  for 2lSS events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

1893        Based on the performance of the model, as shown Figures C.26 and C.27, the Higgs  
 1894        momentum can be reconstructed with fairly high precision when its decay products are correctly  
 1895        identified.

#### 1896        C.4 Alternate b-jet Identification Algorithm

1897        The nominal analysis reconstructs the b-jets by considering different combinations of jets, and  
 1898        asking a neural network to determine whether each combination consists of b-jets from top quark  
 1899        decays. An alternate approach would be to give the neural network about all of the jets in an event  
 1900        at once, and train it to select which two are most likely to be the b-jets from top decay. It was  
 1901        hypothesized that this could perform better than considering each combination independently, as  
 1902        the neural network could consider the event as a whole. While this is not found to be the case,  
 1903        these studies are documented here as a point of interest and comparison.

1904 For these studies, the kinematics of the 10 highest  $p_T$  jets in each event are used for  
 1905 training. This includes the vast majority of truth b-jets. Specifically the  $p_T$ ,  $\eta$ ,  $\phi$ ,  $E$ , and DL1r  
 1906 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.  
 1907 The  $p_T$ ,  $\eta$ ,  $\phi$ , and  $E$  of the leptons and  $E_T^{\text{miss}}$  are included as well. Categorical cross entropy is  
 1908 used as the loss function.

Table 59: 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%

## 1909 C.5 Binary Classification of the Higgs $p_T$

1910 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer  
 1911 resolution. This means separating high and low  $p_T$  events is sufficient for this analysis. As  
 1912 such, rather than attempting to reconstruct the full Higgs  $p_T$  spectrum, a binary classification  
 1913 approach is explored.

1914 A model is built to determine whether  $t\bar{t}H$  events include a high  $p_T$  ( $>150$  GeV) or low  
 1915  $p_T$  ( $<150$  GeV) Higgs Boson. While this is now a classification model, it uses the same input  
 1916 features described in section 18.5. Binary crossentropy is used as the loss function.

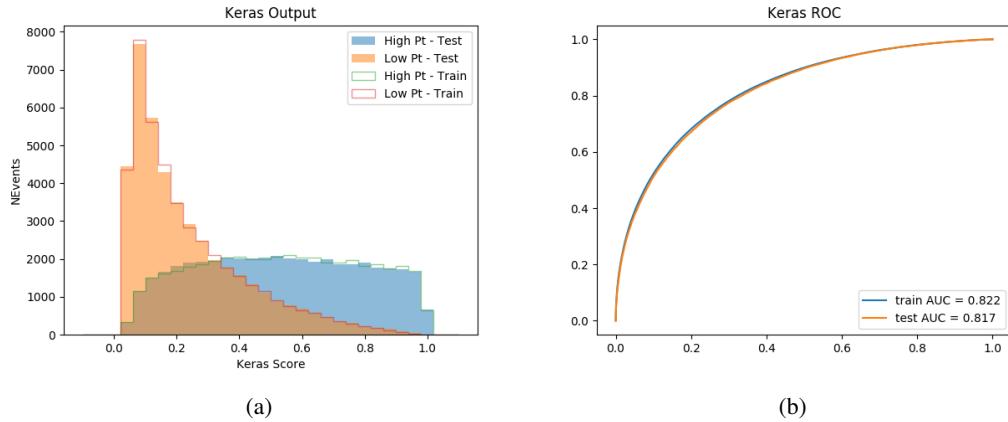


Figure C.28: Output distribution of the NN score for the binary high/low  $p_T$  separation model in the 2lSS channel.

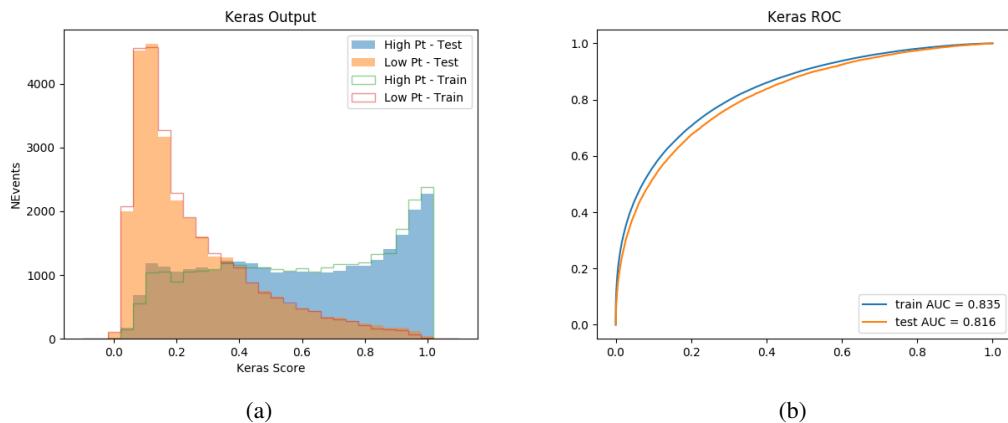


Figure C.29: Output distribution of the NN score for the binary high/low  $p_T$  separation model in the 3lS channel.

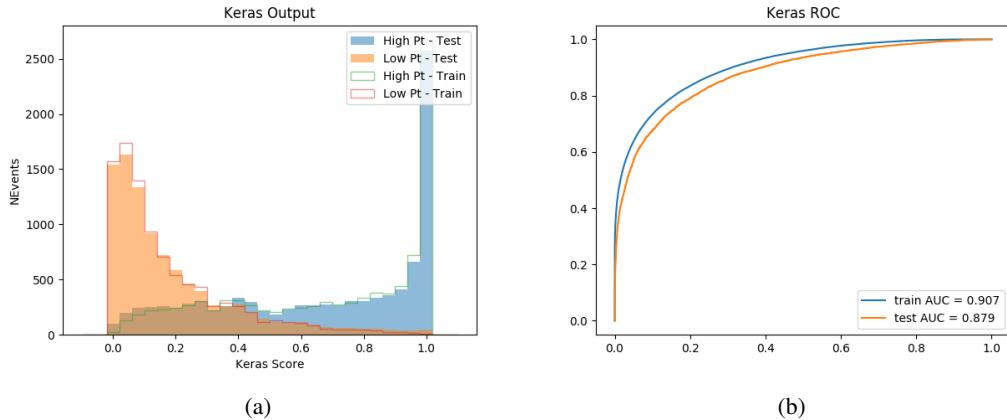


Figure C.30: Output distribution of the NN score for the binary high/low  $p_T$  separation model in the 3lS channel.

## 1917 C.6 Impact of Alternative Jet Selection

1918 A relatively low  $p_T$  threshold of 15 GeV is used to determine jet candidates, as the jets originating  
 1919 from the Higgs decay are found to fall between 15 and 25 GeV a large fraction of the time. The  
 1920 impact of different jet  $p_T$  cuts on our ability to reconstruct the Higgs  $p_T$  is explored here.

1921 The models are retrained in the 2lSS channel with the same parameters as those used in the  
 1922 nominal analysis, but the jet  $p_T$  threshold is altered. The performance of the Higgs  $p_T$  prediction  
 1923 models for jet  $p_T$  cuts of 20 and 25 GeV are shown below.

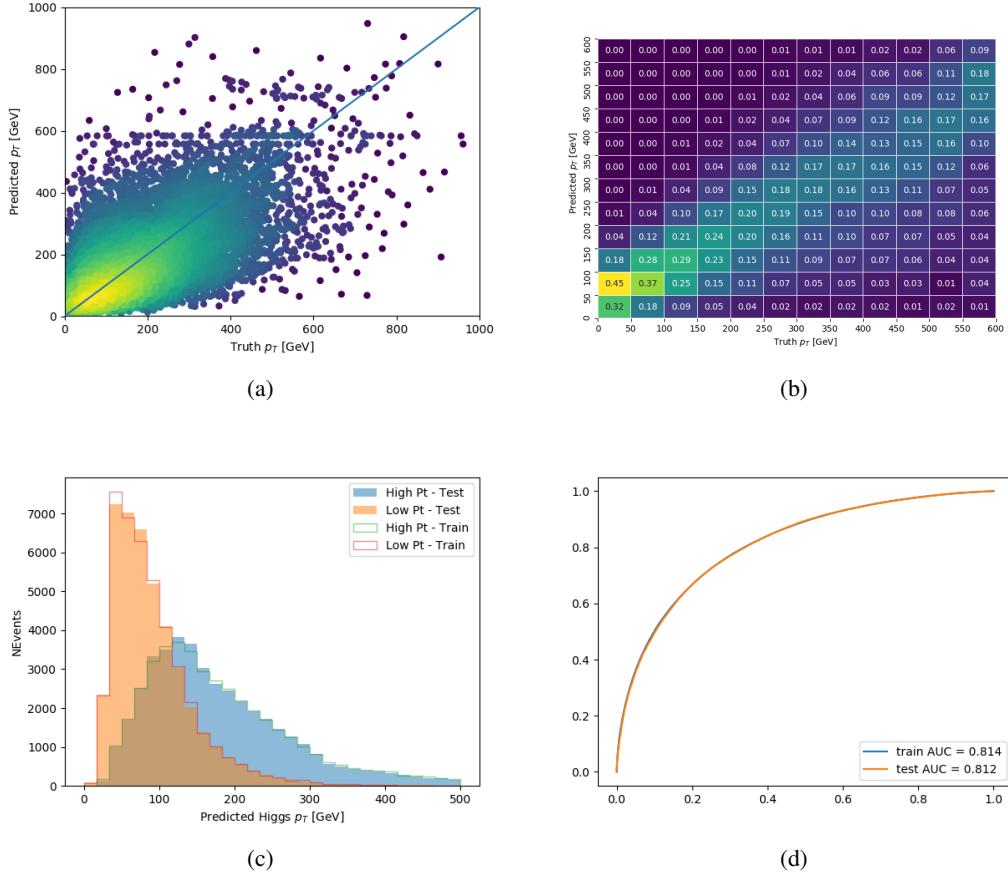
Jet  $p_T > 20$  GeV

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.

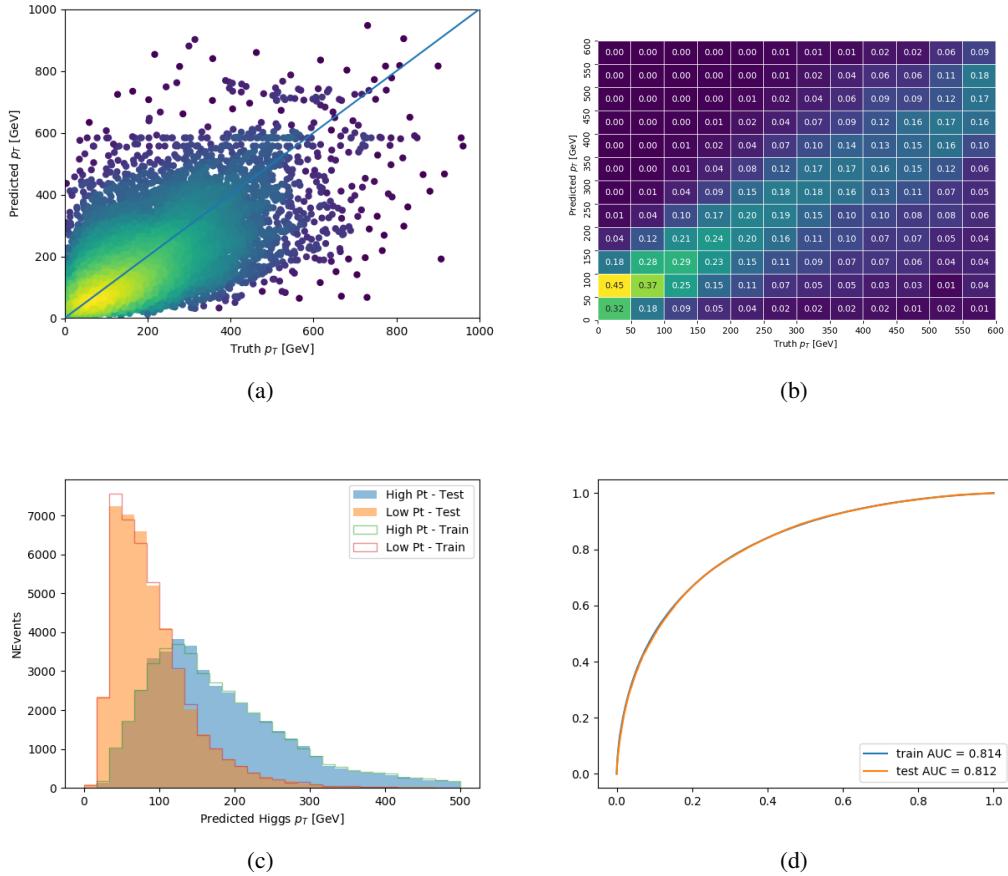
**Jet  $p_T > 25 \text{ GeV}$** 

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.