



ATLAS Paper

GROUP-2017-XX



2nd June 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 Because of the challenges inherent in reconstructing the Higgs in multilepton final states, a
15 deep learning approach is used to predict of the Higgs. The regressed Higgs p_T spectrum is
16 fit to data for events with two same-sign leptons and three leptons in the final state, in order
17 to extract normalization factors for high ($p_T(H) > 150$ GeV) and low ($p_T(H) < 150$ GeV)
18 momentum $t\bar{t}H$ events. Preliminary results are presented for 80 fb^{-1} of data, with projected
19 results shown for 140 fb^{-1} .

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

© 2021 CERN for the benefit of the ATLAS Collaboration.

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

26 **Contents**

27	I Introduction	4
28	1 Introduction	4
29	II Theoretical Motivation	7
30	2 The Standard Model and the Higgs Boson	7
31	2.1 The Forces and Particles of the Standard Model	7
32	2.2 The Higgs Mechanism	10
33	2.2.1 The Higgs Field	10
34	2.2.2 Electroweak Symmetry Breaking	12
35	2.3 $t\bar{t}H$ Production	15
36	2.4 WZ + Heavy Flavor Production	18
37	2.5 Extensions to the Standard Model	19
38	III The LHC and the ATLAS Detector	24
39	3 The LHC	24
40	4 The ATLAS Detector	27
41	4.1 Inner Detector	28
42	4.2 Calorimeters	29
43	4.3 Muon Spectrometer	30
44	4.4 Trigger System	31
45	IV Measurement of WZ + Heavy Flavor	33
46	5 Introduction	33
47	6 Data and Monte Carlo Samples	35
48	6.1 Data Samples	35
49	6.2 Monte Carlo Samples	35
50	7 Object Reconstruction	36
51	8 Event Selection and Signal Region Definitions	40
52	8.1 Event Preselection	40
53	8.2 Fit Regions	41
54	8.3 Non-Prompt Lepton Estimation	43

55	9 tZ Separation Multivariate Analysis	45
56	10 Systematic Uncertainties	47
57	11 Results	51
58	11.1 Fit Procedure	51
59	11.2 Results of the Simultaneous Fit	53
60	11.3 Inclusive 1-2 Jet Fit	56
61	11.4 Alternate tZ Inclusive Fit	60
62	11.4.1 tZ Inclusive Fit	60
63	11.4.2 Floating tZ	61
64	V Differential Studies of t̄tH Multilepton	62
65	12 Data and Monte Carlo Samples	62
66	12.1 Data Samples	62
67	12.2 Monte Carlo Samples	62
68	13 Object Reconstruction	65
69	14 Higgs Momentum Reconstruction	67
70	14.1 Physics Object Truth Matching	69
71	14.2 b-jet Identification	70
72	14.2.1 2lSS Channel	71
73	14.2.2 3l Channel	76
74	14.3 Higgs Reconstruction	79
75	14.3.1 2lSS Channel	80
76	14.3.2 3l Semi-leptonic Channel	84
77	14.3.3 3l Fully-leptonic Channel	88
78	14.4 p _T Prediction	92
79	14.4.1 2lSS Channel	93
80	14.4.2 3l Semi-leptonic Channel	96
81	14.4.3 3l Fully-leptonic Channel	98
82	14.5 3l Decay Mode	100
83	15 Signal Region Definitions	102
84	15.1 Pre-MVA Event Selection	102
85	15.2 Event MVA	106
86	15.3 Signal Region Definitions	113
87	16 Systematic Uncertainties	114
88	17 Results	118
89	17.1 Results - 79.8 fb ⁻¹	119

90	17.2 Projected Results - 139 fb^{-1}	123
91	VI Conclusion	128
92	VII Appendices	137
93	Appendices	137
94	A Non-prompt lepton MVA	137
95	B Supplementary WZ + Heavy Flavor Studies	142
96	B.1 Non-prompt CR Modelling	142
97	B.2 tZ Interference Studies	147
98	B.3 Alternate tZ Inclusive Fit	149
99	B.3.1 tZ Inclusive Fit	150
100	B.3.2 Floating tZ	151
101	C Supplementary $t\bar{t}H$ Differential Analysis Studies	152
102	C.1 Higgs Reconstruction Model Details	152
103	C.1.1 b-jet Identification Features - 2ISS	152
104	C.1.2 b-jet Identification Features - 3l	156
105	C.1.3 Higgs Reconstruction Features - 2ISS	161
106	C.1.4 Higgs Reconstruction Features - 3IS	166
107	C.1.5 Higgs Reconstruction Features - 3IF	171
108	C.2 Background Rejection MVA Details	175
109	C.2.1 Background Rejection MVA Features - 2ISS	175
110	C.2.2 Background Rejection MVA Features - 3l	179
111	C.3 Truth Level Studies	184
112	C.4 Alternate b-jet Identification Algorithm	185
113	C.5 Binary Classification of the Higgs p_T	186
114	C.6 Impact of Alternative Jet Selection	188

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

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

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

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

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

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

145 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-
146 2018 provides the opportunity to make this measurement for the first time. The unprecedented
147 energy achieved by the LHC during this period greatly increase the rate at which $t\bar{t}H$ events are
148 produced, and the large amount of data collected provides the necessary statistics for a differential
149 measurement to be performed.

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

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

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

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

167 Part II

168 Theoretical Motivation

169 2 The Standard Model and the Higgs Boson

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

177 2.1 The Forces and Particles of the Standard Model

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

Standard Model of Elementary Particles

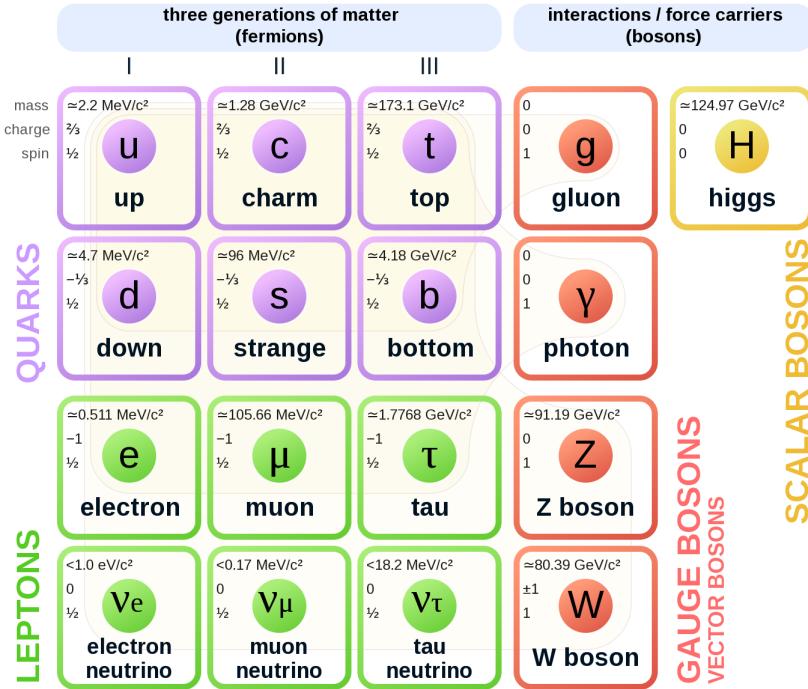


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. [2]

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

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

₁₈₈ Each of these generations form left-handed doublets invariant under SU(2) transfor-

₁₈₉ mations. For the leptons these doublets are:

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

₁₉₀ And for the quarks:

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

₁₉₁ For both leptons and quarks, the heavier generations can decay into the lighter generation

₁₉₂ of particles, while the first generation does not decay. Hence, ordinary matter generally consists

₁₉₃ of this first generation of fermions - electrons, up quarks, and down quarks. Each of these

₁₉₄ fermions has a corresponding anti-particle, which has an equal mass as its partner but opposite

₁₉₅ charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos,

₁₉₆ whose mass has been experimentally confirmed but is not accounted for in the SM.

₁₉₇ Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-

₁₉₈ exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1

₁₉₉ particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0,

₂₀₀ particle - the Higgs Boson. Of the gauge bosons, the W^+ , W^- and Z bosons - which are the

201 mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon
202 mediates the electric force, and the gluon mediates the strong force.

203 **2.2 The Higgs Mechanism**

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

208 Rather than abandoning gauge invariance, an alternative way for particles to acquire mass
209 beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout
210 in 1964 [3]. This procedure for introducing masses for the gauge bosons while preserving local
211 gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory
212 by Weinberg in 1967 [4].

213 **2.2.1 The Higgs Field**

214 The Higgs mechanism introduces a complex scalar SU(2) doublet, Φ , with the form:

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

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

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

216 Where μ and λ are free parameters of the new field. This represents the most general
 217 potential allowed while preserving $SU(2)_L$ invariance and renormalizability. In the case that
 218 $\mu^2 < 0$, this potential takes the form shown in Figure 2.2.

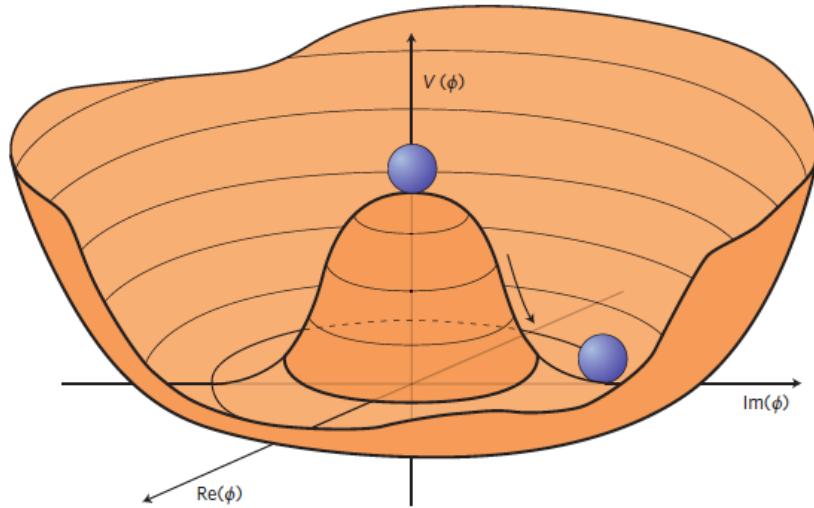


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ , for the case that $\mu^2 < 0$ [5].

219 The significant feature of this potential is that its minimum does not occur for a value of
 220 $\Phi = 0$. Instead, it is minimized when $|\Phi^\dagger \Phi| = -\mu^2/\lambda$. This means that in its ground state, the
 221 Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while
 222 the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

²²³ the minimum is determined only by $\Phi^\dagger \Phi$, there is some ambiguity in the particular definition of

²²⁴ the VEV, but it is generally represented as

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

²²⁵ The full value of Φ can be written as

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

²²⁶ with v being the value of the VEV, and H being the real value of the scalar field.

²²⁷ 2.2.2 Electroweak Symmetry Breaking

²²⁸ The Electroweak (EWK) interaction is described in the SM by a $SU(2)_L \otimes U(1)_Y$ gauge theory.

²²⁹ 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,

²³⁰ B_μ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar

²³¹ field Φ in the Lagrangian:

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

232 Here D_μ represents the covariant derivative required to preserve gauge invariance, g and
 233 g' represent coupling constant of the gauge bosons, σ^a denotes the Pauli matrices of $SU(2)$,
 234 and Y represents the hypercharge of $U(1)$. The terms in this interaction which contribute to the
 235 masses of the gauge bosons can be written as:

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

236 Expanding these terms into the mass eigenstates of the electroweak interaction yields four
 237 physical gauge bosons, two charged and two neutral, which are linear combinations of the fields
 238 W_μ^1 , W_μ^2 , W_μ^3 , and B_μ :

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

239 And the masses of these fields are given by:

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

240 This produces exactly the particles we observe - three massive gauge bosons and a single
 241 massless photon. The massless photon represents the portion of the gauge symmetry, a single
 242 $U(1)$ of the electromagnetic force, that remains unbroken by the VEV.

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

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

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

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

252 The fermions' Yukawa couplings to the VEV take the same form as the fermions' coupling
 253 to the Higgs boson - λ_ψ . Therefore, the strength of a fermion's interaction with the Higgs is
 254 directly proportional to its mass. We now have a model that predicts a Higgs boson with mass
 255 $M_H = \mu$, which interacts with the fermions with coupling strength λ_ψ . Because μ and λ_ψ are

256 free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions
 257 must be measured experimentally.

258 **2.3 $t\bar{t}H$ Production**

259 The strength of a particles interaction with the Higgs, given by its Yukawa coupling, is propor-
 260 tionate to its mass. The top quark - as the heaviest known particle - has the strongest interaction,
 261 making this interaction particularly interesting to study. While several processes involve interac-
 262 tions between the Higgs and the top, some Higgs production modes include the top interaction
 263 only as a part of a loop diagram, such as the gluon-gluon fusion diagram shown in Figure 2.3.
 264 This process therefore only allows for an indirect probe of the Higgs-top Yukawa coupling, as
 265 the flavor of the quark in this diagram is not unique.

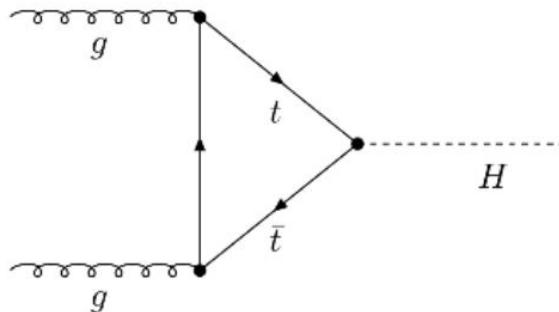


Figure 2.3: Diagram of a Higgs boson produced via gluon-gluon fusion.

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

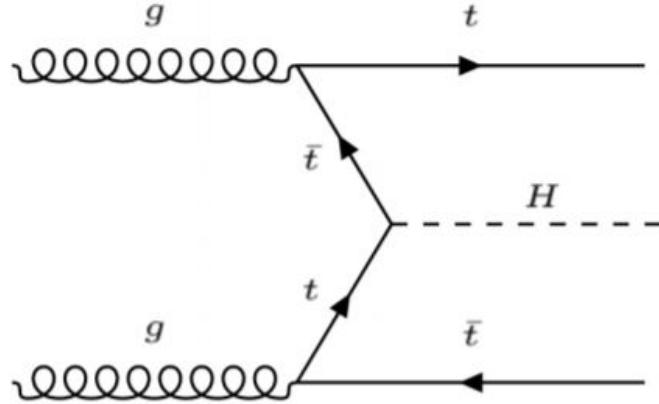


Figure 2.4: Diagram of a Higgs boson produced in association with a pair of top quarks.

269 The Higgs boson, as well as the top quarks, have very short lifetimes - on the order of
 270 10^{-22} s and 10^{-25} s respectively - meaning they can only be observed via their decay products.
 271 Measuring this process is therefore a matter of identifying events with final states consistent with
 272 $t\bar{t}H$ production.

273 Studies of $t\bar{t}H$ production have been reported by the ATLAS collaboration for $H \rightarrow b\bar{b}$,
 274 $H \rightarrow \gamma\gamma$ and multilepton (encompassing $H \rightarrow W^+W^-$, $H \rightarrow ZZ$ and $H \rightarrow \tau^-\tau^+$, with
 275 $H \rightarrow ZZ \rightarrow 4l$ as a separate analysis) decay modes. While the branching ratio of $H \rightarrow W^+W^-$
 276 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
 277 large $t\bar{t}$ backgrounds. On the other hand, $H \rightarrow \gamma\gamma$ produces the most easily identifiable signal,
 278 but has a much smaller branching ratio than $H \rightarrow W^+W^-$. Therefore, compared with other final
 279 states of $t\bar{t}H$, the $t\bar{t}H$ – ML channel is an attractive candidate for study, as it involves a good
 280 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.

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

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

284 $t\bar{t}H$ production was observed by ATLAS using up to 79.8 fb^{-1} of data collected at \sqrt{s}
 285 $= 13 \text{ TeV}$, based on a combination of five Higgs decay modes: $b\bar{b}$, WW^* , $\tau^-\tau^+$, $\gamma\gamma$, and ZZ^*
 286 [6]. A significance of 5.8σ was observed, compared to a 4.9σ expected significance. Since then,
 287 two analyses have published updated results ($H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4l$) with the full Run 2
 288 dataset, representing 139 fb^{-1} . Studies are still ongoing in the remaining channels.

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

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

293 **2.4 WZ + Heavy Flavor Production**

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

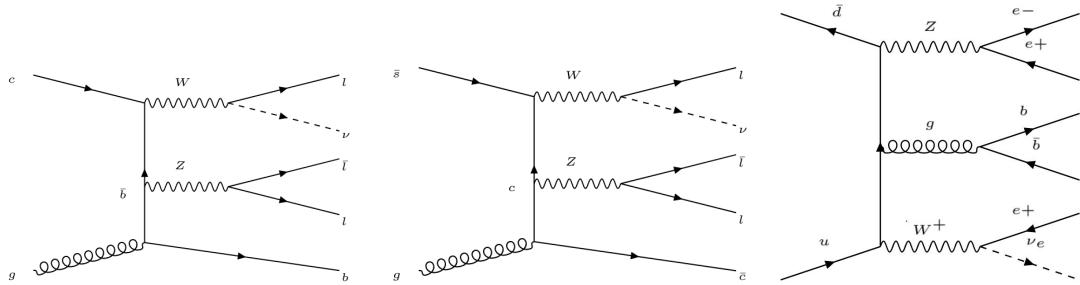


Figure 2.5: Example Feynman diagrams of $WZ +$ heavy flavor production

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

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

306 **2.5 Extensions to the Standard Model**

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

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

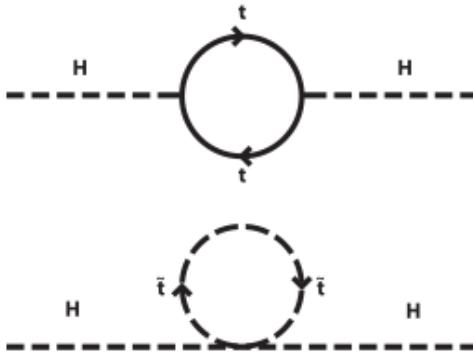


Figure 2.6: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram.

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].

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

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

336 Here Λ represents the energy scale of the new physics, and f is a Wilson coefficient which
 337 represents the strength of the effective coupling. An experimental observation of any non-zero
 338 value of f would be a sign of BSM physics.

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

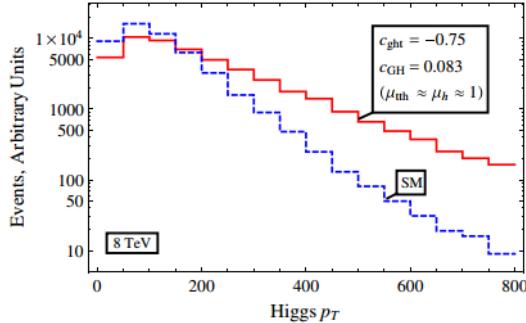


Figure 2.7: The momentum spectrum of the Higgs boson produced via top-quarks with (red) and without (blue) the presence of dimension-six operators.

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

347 Reconstructing the Higgs is a particular challenge in the multilepton channels of $t\bar{t}H$, due
 348 to an ambiguity arising from multiple sources of missing energy. In the $H \rightarrow \gamma\gamma$ channel, the
 349 kinematics of the Higgs can be fully reconstructed from the two photons. The same is true of
 350 $H \rightarrow b\bar{b}$, though with the additional challenge of identifying which two of the four b-quarks in
 351 the final state originated from the Higgs. By contrast, the two channels (2lSS and 3l) targeted
 352 by this analysis include at least one neutrino originating from the Higgs decay.

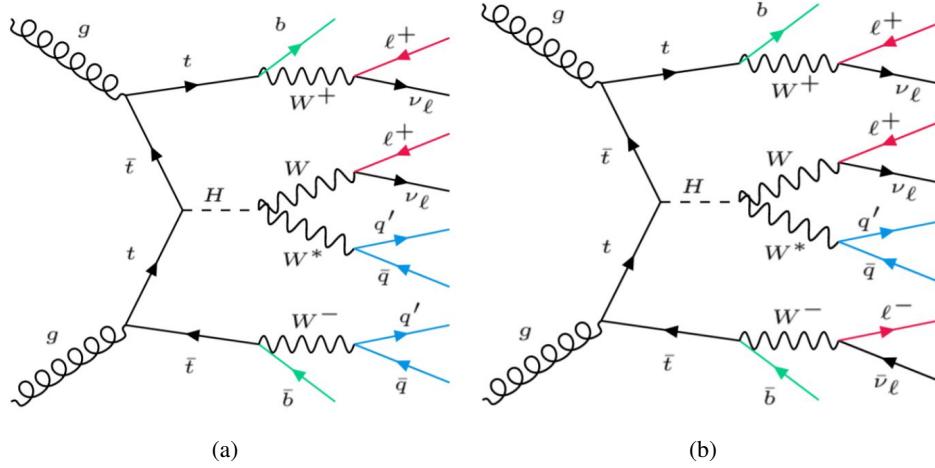


Figure 2.8: Feynman diagrams of $t\bar{t}H$ production with (a) two same-sign leptons and (b) three leptons in the final state.

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

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

Part III**362 The LHC and the ATLAS Detector****363 3 The LHC**

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

373 The LHC first began running in 2009 at a proton-proton center of mass energy of $\sqrt{s} = 8$
374 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during
375 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,
376 and collected data at an increased energy of $\sqrt{s} = 13$ TeV until 2018, a period referred to as Run
377 2.

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

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

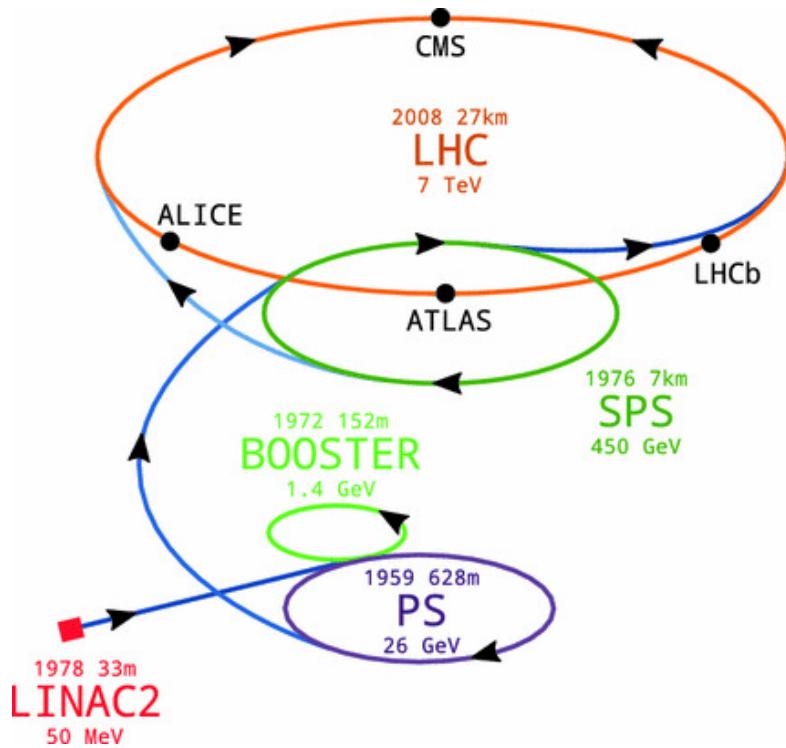


Figure 3.1: A summary of the accelerator chain used to feed protons into the LHC [11].

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

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

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

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

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

397 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around
 398 140 fb^{-1} , exceeding the expected integrated luminosity of 100 fb^{-1} .

399 4 The ATLAS Detector

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

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

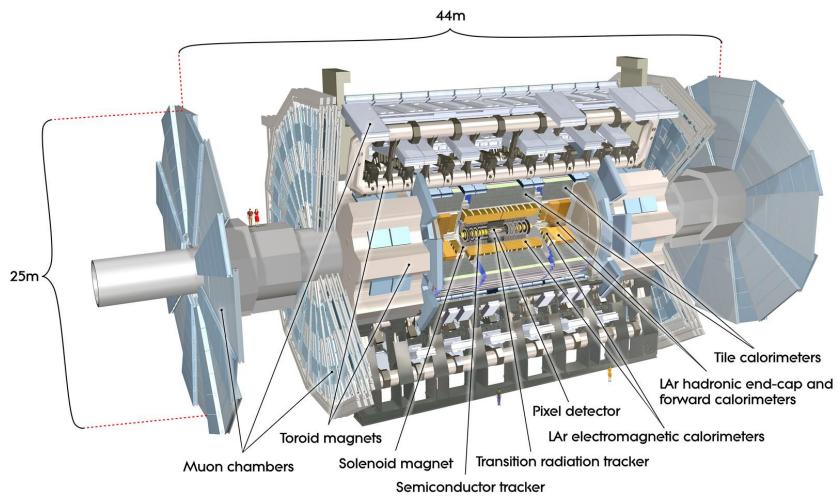


Figure 4.1: Cutaway view of the ATLAS detector, with labels of its major components [12].

408 **4.1 Inner Detector**

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

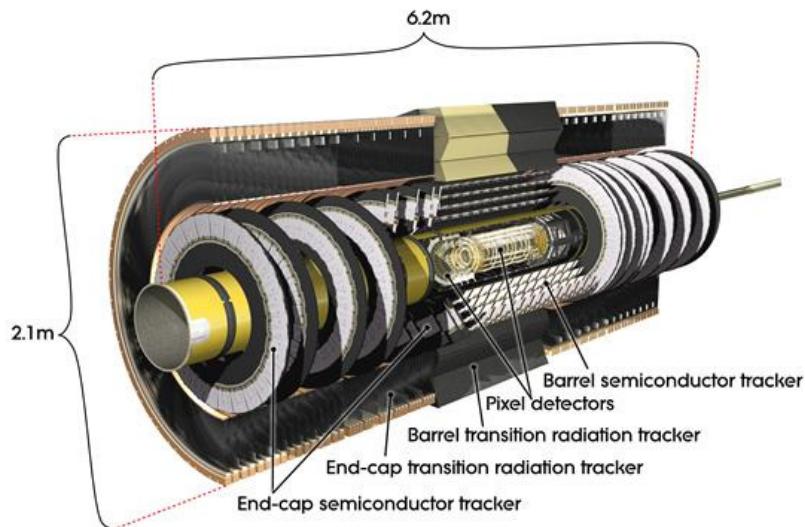


Figure 4.2: Cutaway view of the Inner Detector [13].

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

419 The Semiconductor Tracker (SCT) is similar to the Pixel detector, but uses long strips of
 420 silicon rather than small pixels to cover a larger spatial area. It includes over 6 million readout
 421 strips, allowing the position of charged particles to be measured to an accuracy of 17 μm .

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

428 **4.2 Calorimeters**

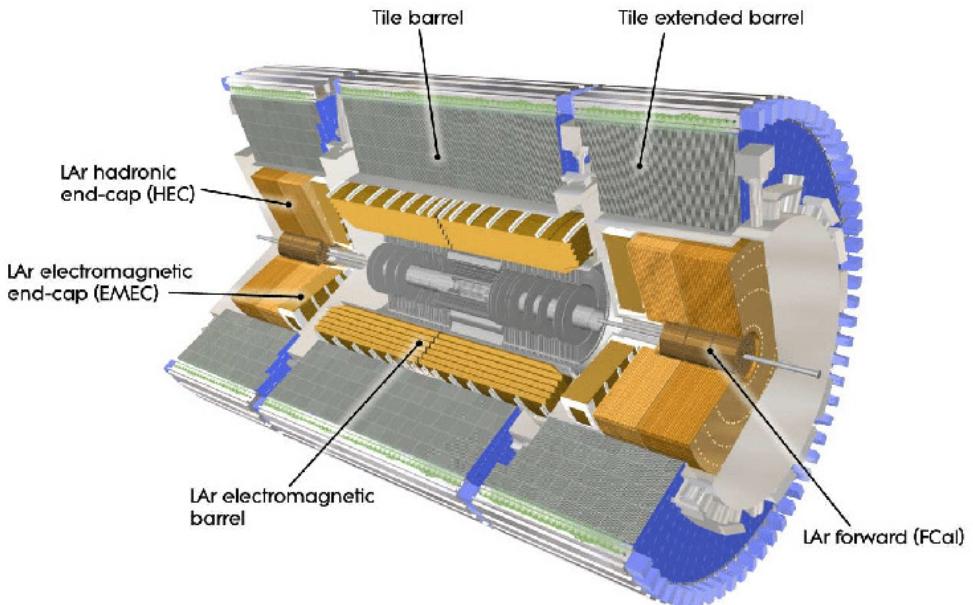


Figure 4.3: Cutaway view of the calorimeter system of the ATLAS detector [13].

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

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

441 **4.3 Muon Spectrometer**

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

446 The largest of the subdetectors, it extends from the outside of the calorimeter system, about a
447 4.25 m radius from the beam line, to a radius of 11 m. This large detector system is necessary

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

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

457 1200 tracking chambers are placed in the muon system in order to precisely measure the
458 tracks of muons with high spatial resolution. The path of the muons are tracked by Monitored
459 Drift Tubes, which are drift chambers formed by aluminum tubes and filled with ionizing gas.
460 These tubes produce a multi-layer spatial resolution on the order of 50 μm .

461 **4.4 Trigger System**

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

⁴⁶⁶ therefore a sophisticated trigger system is employed in real time to determine whether events are
⁴⁶⁷ sufficiently interesting to be worth storing.

⁴⁶⁸ The trigger system in ATLAS involves multiple levels, each of which select out which
⁴⁶⁹ events move on to the next level of scrutiny. The level-1 trigger uses hardware information from
⁴⁷⁰ the calorimeters and muon spectrometer to select events that contain candidates for particles
⁴⁷¹ commonly used in analysis, such as energetic leptons and jets. The level-1 trigger reduces the
⁴⁷² rate of events from 40 MHz to around 100 kHz.

⁴⁷³ Events that pass the level-1 trigger move to the High-Level Trigger (HLT). The HLT takes
⁴⁷⁴ place outside of the detector in software, and looks for properties such as a large amount of
⁴⁷⁵ missing transverse energy, well defined leptons, and multiple high energy jets. Events that pass
⁴⁷⁶ the HLT are stored and used for analysis. Because the specifics of the HLT are determined by
⁴⁷⁷ software rather than hardware, the thresholds can be changed throughout the run of the detector
⁴⁷⁸ in response to run conditions such as changes to pileup and luminosity. After the HLT is applied,
⁴⁷⁹ the event rate is reduced to around 1000 per second, which are recorded for analysis.

Part IV**Measurement of WZ + Heavy Flavor****5 Introduction**

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

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

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

497 dataset collected by the ATLAS detector, representing 139 fb^{-1} of data from pp collisions at
 498 $\sqrt{s} = 13 \text{ TeV}$, is used for this study.

499 The fiducial volume at particle level is defined based on the number of stable leptons and
 500 jets in each event. Three light leptons with total charge ± 1 and one or two associated jets are
 501 required. Only leptons which do not originate from hadron or τ decays are considered. The
 502 phase space definitions use dressed kinematics of the final state particles. Leptons are dressed
 503 by summing the momentum of photons within a cone of $\Delta R < 0.1$ of the lepton to correct the
 504 leptons energy. Particle level jets are reconstructed using the anti- k_t algorithm with a radius of
 505 $R = 0.4$. The kinematic selection applied to these objects is summarized below:

- 506 • Three light leptons with total charge ± 1 , $|\eta| < 2.5$
- 507 • OS lepton with $p_T > 10 \text{ GeV}$, SS leptons with $p_T > 20 \text{ GeV}$
- 508 • One OSSF lepton pair with $|M(l\bar{l}) - 91.2 \text{ GeV}| < 10 \text{ GeV}$
- 509 • One or two associated truth jets with $p_T > 25 \text{ GeV}$, $|\eta| < 2.5$, $R < 0.4$

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

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

522 **6 Data and Monte Carlo Samples**

523 **6.1 Data Samples**

524 This study uses a sample of proton-proton collision data collected by the ATLAS detector from
525 2015 through 2018 at an energy of $\sqrt{s} = 13$ TeV, which represents an integrated luminosity of
526 139 fb^{-1} [15]. This data set was collected with a bunch-crossing rate of 25 ns. All data used in
527 this analysis was verified by data quality checks [16].

528 **6.2 Monte Carlo Samples**

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

⁵³¹ GEANT4 [17]. The WZ signal samples are simulated using Sherpa 2.2.2 [18]. Specific information

⁵³² about the Monte Carlo samples being used can be found in Table 14.

Table 2: The configurations used for event generation of signal and background processes, including the event generator, matrix element (ME) order, parton shower algorithm, and parton distribution function (PDF).

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

7 Object Reconstruction

⁵³³ All regions defined in this analysis share a common lepton, jet, and overall event preselection.

⁵³⁴ The selection applied to each physics object is detailed here; the event preselection, and the

⁵³⁵ selection used to define the various fit regions, is described in Section 8.

⁵³⁶ All events are required to be selected by dilepton triggers. The p_T thresholds of the

538 dilepton trigger on two electrons were 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017 and
 539 2018, while for the dimuon triggers the p_T thresholds on the leading (sub-leading) muon were
 540 18 GeV (8 GeV) in 2015, and 22 GeV (8 GeV) in 2016-2018. For the electron+muon triggers,
 541 the p_T thresholds on the electron (muon) were 17 GeV (14 GeV) for all datasets.

542 Electron candidates are reconstructed from energy clusters in the electromagnetic calor-
 543 imeter that are associated with charged particle tracks reconstructed in the inner detector [26].
 544 Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the
 545 transition region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| <$
 546 1.52, are rejected. A multivariate likelihood discriminant combining shower shape and track
 547 information is used to distinguish real electrons from hadronic showers (fake electrons). To
 548 further reduce the non-prompt electron contribution, the track is required to be consistent with
 549 originating from the primary vertex; requirements are imposed on the transverse impact para-
 550 meter significance ($|d_0|/\sigma_{d_0} < 5$) and the longitudinal impact parameter ($|\Delta z_0 \sin \theta_\ell| < 0.5$
 551 mm). Electron candidates are required to pass the TIGHTLH identification requirement detailed
 552 in [27].

553 Muon candidates are reconstructed by combining inner detector tracks with track segments
 554 or full tracks in the muon spectrometer [28]. Muon candidates are required to have $p_T > 10$ GeV
 555 and $|\eta| < 2.5$. The longitudinal impact parameter is the same for both electrons and muons,
 556 while muons are required to pass a slightly tighter transverse impact parameter selection,
 557 $|d_0|/\sigma_{d_0} < 3$. Muons are also required to pass Medium ID requirements, as detailed in [27].
 558 Leptons are additionally required to pass a non-prompt BDT selection, described in detail in

559 [29]. Optimized working points and scale factors for this BDT are taken from that analysis.

560 Jets are reconstructed from calibrated topological clusters built from energy deposits in
561 the calorimeters using the anti- k_t algorithm [30], as well as information from the inner tracking
562 detector, with a radius parameter $R = 0.4$. Jets with energy contributions likely arising from
563 noise or detector effects are removed from consideration, and only jets satisfying $p_T > 25$ GeV
564 and $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
565 association algorithm is used to confirm that the jet originates from the selected primary vertex,
566 in order to reject jets arising from pileup collisions [31].

567 In order to make a measurement of $WZ +$ heavy flavor it is necessary to distinguish these
568 events from $WZ +$ light jets. For this purpose, the DL1r b-tagging algorithm is used to distinguish
569 heavy flavor jets from lighter ones [32]. The DL1r algorithm uses jet kinematics, particularly
570 jet vertex information, as input for a neural network which assigns each jet a score designed to
571 reflect how likely that jet is to have originated from a b-quark.

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

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

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

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

582 Missing transverse momentum (E_T^{miss}) is used as part of the event selection. The missing
 583 transverse momentum vector is defined as the negative of the vector of the transverse momenta
 584 of all reconstructed physics objects as well as remaining unclustered energy, the latter of which is
 585 estimated from low- p_T tracks associated with the primary vertex but not assigned to a hard object,
 586 with object definitions taken from [33]. Light leptons considered in the E_T^{miss} reconstruction are
 587 required to have $p_T > 10$ GeV, while jets are required to have $p_T > 20$ GeV.

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

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

595 8 Event Selection and Signal Region Definitions

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

599 8.1 Event Preselection

600 Events are required to include exactly three reconstructed light leptons passing the requirement
601 described in 7, which have a total charge of ± 1 . As the opposite sign lepton is found to be prompt
602 the vast majority of the time [29], it is required to have $p_T > 10 \text{ GeV}$, while the same sign leptons
603 are required to have $p_T > 20 \text{ GeV}$ to reduce the contribution of non-prompt leptons.

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

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

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

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

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

620 8.2 Fit Regions

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

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

Region	Selection
1j, <85%	$N_{\text{jets}} = 1, n\text{Jets}_{\text{DL1r}_85} = 0$
1j, 85%-77%	$N_{\text{jets}} = 1, n\text{Jets}_{\text{DL1r}_85} = 1, n\text{Jets}_{\text{DL1r}_77} = 0$
1j, 77%-70%	$N_{\text{jets}} = 1, n\text{Jets}_{\text{DL1r}_77} = 1, n\text{Jets}_{\text{DL1r}_70} = 0$
1j, 70%-60%	$N_{\text{jets}} = 1, n\text{Jets}_{\text{DL1r}_70} = 1, n\text{Jets}_{\text{DL1r}_60} = 0$
1j, >60%	$N_{\text{jets}} = 1, n\text{Jets}_{\text{DL1r}_60} = 1, tZ \text{ BDT} > 0.12$
1j tZ CR	$N_{\text{jets}} = 1, n\text{Jets}_{\text{DL1r}_60} = 1, tZ \text{ BDT} < 0.12$
2j, <85%	$N_{\text{jets}} = 2, n\text{Jets}_{\text{DL1r}_85} = 0$
2j, 85%-77%	$N_{\text{jets}} = 2, n\text{Jets}_{\text{DL1r}_85} \geq 1, n\text{Jets}_{\text{DL1r}_77} = 0$
2j, 77%-70%	$N_{\text{jets}} = 2, n\text{Jets}_{\text{DL1r}_77} \geq 1, n\text{Jets}_{\text{DL1r}_70} = 0$
2j, 70%-60%	$N_{\text{jets}} = 2, n\text{Jets}_{\text{DL1r}_70} \geq 1, n\text{Jets}_{\text{DL1r}_60} = 0$
2j, >60%	$N_{\text{jets}} = 2, n\text{Jets}_{\text{DL1r}_60} \geq 1, tZ \text{ BDT} > 0.12$
2j tZ CR	$N_{\text{jets}} = 2, n\text{Jets}_{\text{DL1r}_60} \geq 1, tZ \text{ BDT} < 0.12$

623 The working points discussed in Section 7 are used to separate events into fit regions based
 624 on the highest working point reached by a jet in each event. Because the background composition
 625 differs significantly based on the number of b-jets, events are further subdivided into 1-jet and
 626 2-jet regions in order to minimize the impact of background uncertainties.

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

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

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

638 8.3 Non-Prompt Lepton Estimation

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

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

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

654 91.2 GeV are rejected. This ensures the $t\bar{t}$ validation region is orthogonal to the preselection
655 region.

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

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

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

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

674 of the lepton from the W candidate is found to differ. This is the lepton not included in the
 675 Z-candidate, and in the case of Z+jets, this lepton is most often the non-prompt lepton. To
 676 account for this discrepancy, a variable correction factor is applied to Z+jets. χ^2 minimization
 677 of the W lepton p_T spectrum is performed to derive a correction factor.

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

682 9 tZ Separation Multivariate Analysis

683 An important process to consider in this analysis is tZ: the top almost always decays into a W
 684 boson and b-quark, and when both the W and Z decay leptonically, this gives three leptons and
 685 a heavy flavor jet in the final state. Because tZ can produce a final state identical to the signal,
 686 it represents a predominant background in the most signal enriched regions. That is, the region
 687 with one jet passing the 60% DL1r WP. Therefore, a boosted decision tree (BDT) algorithm is
 688 trained using XGBoost [34] to separate WZ + heavy flavor from tZ using kinematic quantities.
 689 The result of this BDT is used to create a tZ enriched region in the fit, reducing its impact on the
 690 measurement of WZ + heavy flavor.

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

693 of each combination of lepton pairs, E_T^{miss} , the distance between each combination of leptons,
694 $\Delta R(l\bar{l})$, and the distance between each lepton and the jet, $\Delta R(lj)$.

695 Here the top candidate is reconstructed based on the procedure described in section 6.1 of
696 [35]. Broadly, the mass of the top quark candidate is reconstructed from the jet, the lepton not
697 included in the Z-candidate, and a reconstructed neutrino. In the case that there is one jet in the
698 event, there is only possible b-jet candidate. For events with two jets, the jet with the highest
699 DL1r score is used.

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

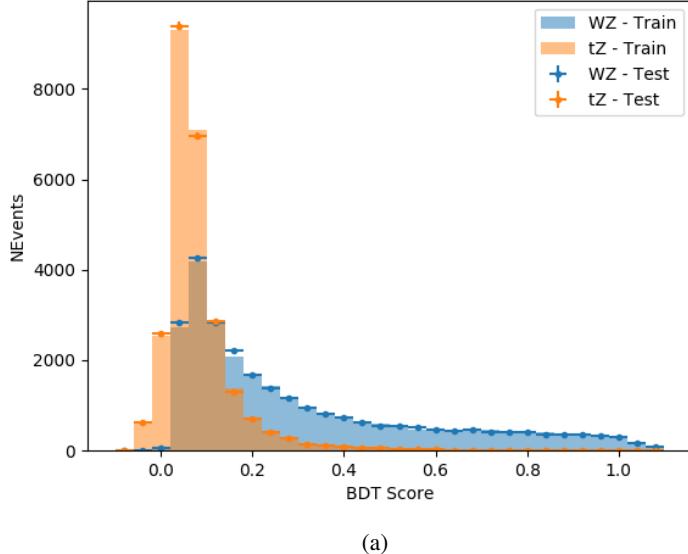


Figure 9.1: Distribution of the BDT response for WZ+b (blue) and tZ (orange) events, for both training and testing samples.

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

712 10 Systematic Uncertainties

713 The systematic uncertainties that are considered are summarized in Table 34. These are imple-
 714 mented in the fit either as a normalization factors or as a shape variation or both in the signal

⁷¹⁵ and background estimations. The numerical impact of each of these uncertainties is outlined in

⁷¹⁶ Section 17.

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

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

⁷¹⁷ The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [36], obtained

⁷¹⁸ using the LUCID-2 detector [37] for the primary luminosity measurements.

⁷¹⁹ The experimental uncertainties are related to the reconstruction and identification of light

⁷²⁰ leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The sources which contribute to

⁷²¹ the uncertainty in the jet energy scale (JES) [38] are decomposed into uncorrelated components

⁷²² and treated as independent sources of uncertainty in the analysis. A similar approach is used for

723 the jet energy resolution (JER) uncertainty.

724 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
725 [39] are also decomposed into uncorrelated components. The large number of components for
726 b-tagging is due to the calibration of the distribution of the MVA discriminant for each individual
727 WP bin.

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

737 A similar approach is taken to account for uncertainties in migrations between the number
738 of reco and truth jets. The fraction of events with 1 truth jet which fall into the 1 jet bin versus
739 the 2 jet bin at reco level is compared for Sherpa and Powheg. The same is done for events with
740 2 truth jets. A systematic is included where events are shifted between the 1-jet and 2-jet regions
741 based on the differences between these two shapes. This is done independently for each of the
742 WZ + b, WZ + c, and WZ + light templates.

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

750 Theoretical uncertainties applied to MC background predictions, including cross section,
751 PDF, and scale uncertainties are taken from theory calculations, with the exception of non-prompt
752 and diboson backgrounds. The cross-section uncertainty on tZ is taken from [40]. Derivation
753 of the non-prompt background uncertainties, Z+jets and t \bar{t} , are explained in Section 8.3. These
754 normalization uncertainties are chosen so as to account for the complete uncertainty in the
755 non-prompt contribution, and therefore no additional modelling uncertainties are considered for
756 Z+jets and t \bar{t} .

757 Due to its importance as a background, additional modelling uncertainties are considered
758 for tZ. Alternative tZ samples with variations in scale and shower modelling are included as
759 systematics. The other VV + heavy flavor processes (namely VV+b and VV+charm, which
760 primarily consist of ZZ events) are also poorly understood, because these processes involve the
761 same physics as WZ + heavy flavor, and have also not been measured. Therefore, a conservative
762 50% uncertainty is applied to those samples. While this uncertainty is large, it is found to have
763 little impact on the significance of the final result.

764 The theory uncertainties applied to the MC estimates are summarized in Table 36.

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

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

765 11 Results

766 11.1 Fit Procedure

767 A maximum-likelihood fit is performed over the total yields in the various fit regions described
768 in Section 8 in order to extract the best-fit value of the WZ + b and WZ + c jet contributions for
769 events with both 1 and 2 associated jets.

770 Because the fit regions are defined by the number of associated jets at reco-level, the signal
771 is split into separate samples based on the number of truth jets in order to account for differences
772 in the number of truth jets compared to the number of reco-jets. The WZ + b, WZ + c and WZ

773 + light contributions are separated into independent samples based on the number of truth jets
 774 in each event. WZ + 1 truth-jet and WZ + 2 truth-jets are treated as signal samples, while WZ +
 775 0 truth-jets and WZ + $>= 3$ truth-jets are treated as an additional background.

776 A maximum likelihood fit to data is performed simultaneously in the regions described in
 777 Section 8, summarized in figure 11.1. The six signal templates, which include WZ + b 1-jet,
 778 WZ + c 1-jet, WZ+ light 1-jet, WZ + b 2-jets, WZ + c 2-jets, WZ + light 2-jets, are allowed
 779 to float, while the remaining background contributions are held fixed. Normalization factors for
 780 each of these templates are extracted from the fit. A simultaneous fit is performed over all 1-jet
 781 and 2-jet regions.

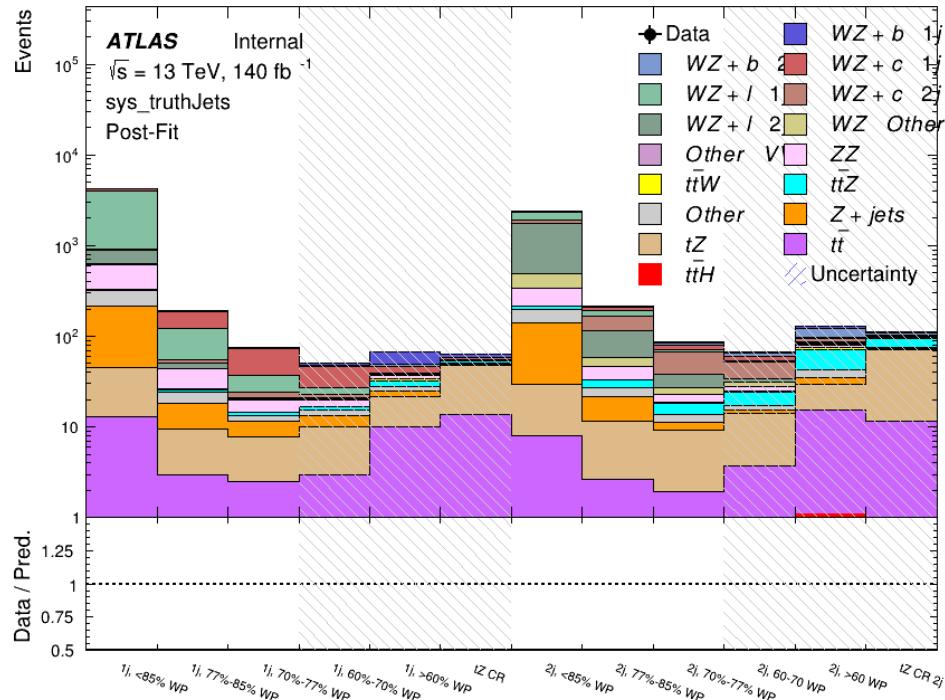


Figure 11.1: Post-fit summary of the fit regions.

Several alternative fit strategies are reported as well, including a measurement of WZ + 1 or 2 jets inclusively, a fit where tZ is allowed to float, and a case where tZ is included as part of the signal.

As described in Section 16, there are 230 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.

11.2 Results of the Simultaneous Fit

The results of the fit to an Asimov dataset for the fiducial regions considered, including both the normalization factors as well as the expected cross-sections, along with their uncertainties, are summarized in Table 7.

795

Process	μ	σ
WZ + b - 1-jet	$1.00^{+0.47}_{-0.43}(\text{stat})^{+0.32}_{-0.27}(\text{sys})$	$1.74^{+0.82}_{-0.75}(\text{stat})^{+0.53}_{-0.48}(\text{sys}) \text{ fb}$
WZ + c - 1-jet	$1.00^{+0.18}_{-0.17}(\text{stat})^{+0.19}_{-0.17}(\text{sys})$	$14.6^{+2.5}_{-2.3}(\text{stat})^{+2.6}_{-2.3}(\text{sys}) \text{ fb}$
WZ + b - 2-jet	$1.00^{+0.53}_{-0.51}(\text{stat})^{+0.39}_{-0.34}(\text{sys})$	$2.5^{+1.3}_{-1.3}(\text{stat})^{+0.95}_{-0.83}(\text{sys}) \text{ fb}$
WZ + c - 2-jet	$1.00^{+0.25}_{-0.24}(\text{stat})^{+0.32}_{-0.27}(\text{sys})$	$12.7^{+3.3}_{-3.2}(\text{stat})^{+3.9}_{-3.4}(\text{sys}) \text{ fb}$

Table 7: Normalization factors and cross-sections extracted from the fit for each of the fiducial regions considered

796 An expected significance of 2.0σ is observed for WZ + b with 1-jet, and 1.7σ for WZ + b
 797 with two jets. A summary of the correlations between these various measurements is shown in
 798 Figure 11.2.

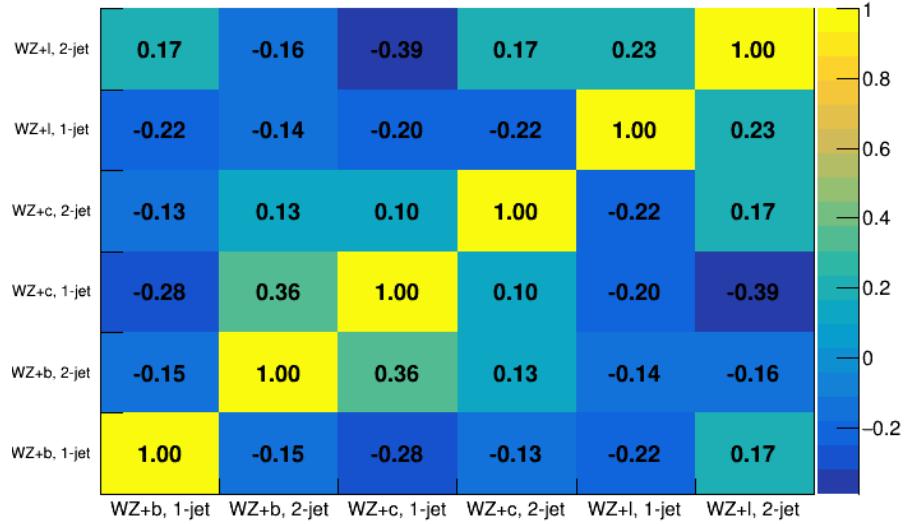


Figure 11.2: Correlations between the various measured components of WZ.

799 The impact of each NP is calculated by performing the fit with the parameter of interest
 800 held fixed, varied from its fitted value by its uncertainty, and calculating $\Delta\mu$ relative to the
 801 baseline fit. The impact of the most significant sources of systematic uncertainties on WZ + b
 802 and WZ + c with one associated jet are summarized in Table 8-9.

Uncertainty Source	$\Delta\sigma/\sigma_{\text{nominal}}$	
Jet Energy Scale	0.14	-0.15
WZ + light, 1-jet cross-section	0.12	-0.14
WZ + c, 1-jet cross-section	-0.09	0.11
tZ Modelling (shower tune)	-0.07	0.08
Other Diboson + b cross-section	-0.07	0.07
tZ cross-section	-0.06	0.08
Jet Energy Resolution	-0.07	0.07
WZ + b 1j/2j Migration	0.08	-0.07
Luminosity	-0.06	0.07
Flavor tagging	0.05	0.05
t <bar>t} cross-section</bar>	-0.05	0.05
Total Systematic Uncertainty	0.28	0.33

Table 8: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

Uncertainty Source	$\Delta\sigma/\sigma_{\text{nominal}}$	
WZ + c 1j/2j migration	0.12	-0.09
Flavor Tagging	0.09	0.08
WZ + b, 1-jet cross-section	-0.04	0.05
Luminosity	-0.04	0.04
Jet Energy Resolution	0.04	0.04
WZ + b, 2-jet cross-section	0.04	-0.03
WZ cross-section - QCD scale	-0.04	0.04
Jet Energy Scaling	0.04	0.02
WZ cross-section - PDF	-0.03	0.03
WZ + light, 1-jet cross-section	0.03	-0.03
total	0.19	0.17

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

803 The impact of the most significant systematic uncertainties on the 2-jet fiducial regions
 804 are summarized in Table 10-11.

Uncertainty Source	$\Delta\sigma/\sigma_{\text{nominal}}$	
WZ + c 2-jet cross-section	-0.13	0.16
WZ + l 2-jet cross-section	0.12	-0.09
ttZ cross-section - QCD scale	-0.10	0.13
Luminosity	-0.11	0.12
WZ + b 1-jet cross-section	-0.11	0.10
Jet Energy Scale	-0.11	0.11
tZ cross-section	-0.11	0.11
WtZ cross-section	-0.07	0.07
Flavor tagging	0.05	0.05
Other VV + b cross-section	-0.05	0.05
Total	0.36	0.37

Table 10: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with 2 associated jets.

Uncertainty Source	$\Delta\sigma/\sigma_{\text{nominal}}$	
WZ + c 1j/2j migration	-0.17	0.25
Flavor Tagging	0.14	0.13
WZ + b, 1-jet cross-section	-0.09	0.09
Jet Energy Scale	0.06	0.08
Jet Energy Resolution	0.05	0.05
WZ \geq 3j/2j migration	-0.04	0.04
WZ + c 2j/1j migration	-0.04	0.04
WZ cross-section - QCD scale	-0.04	0.04
WZ + light modelling	0.04	-0.03
Luminosity	-0.03	0.03
total	0.27	0.32

Table 11: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + c with 2 associated jets.

805 11.3 Inclusive 1-2 Jet Fit

806 An alternative fit is performed which combines the WZ + 1-jet and WZ + 2-jet samples
807 rather than fitting them independently. This is done primarily as a cross-check of the nominal
808 analysis, to see if measuring 1-jet and 2-jet events separately and combining them gives drastically
809 different results than measuring them together.

810 For this study, three signal templates, WZ + b, WZ + c and WZ + light, are fit to data, and
811 the systematics accounting for migrations between 1-jet and 2-jet bins are removed. All other
812 background and nuisance parameters remain the same as the nominal fit.

813 The measured μ value for WZ + b is $\mu = 1.00^{+0.30}_{-0.29}(\text{stat})^{+0.25}_{-0.23}(\text{sys})$, with a significance
814 of 2.8σ , and the uncertainty on WZ + c is $\mu = 1.00 \pm 0.12(\text{stat}) \pm 0.13(\text{sys})$. This is compared
815 to combined uncertainty of $\mu = 1.00^{+0.32}_{-0.30}(\text{stat})^{+0.24}_{-0.23}(\text{sys})$ for WZ + b when 1-jet and 2-jet
816 events are measured separately and then combined.

817 A post-fit summary plot of the fit regions is shown in Figure 11.3:

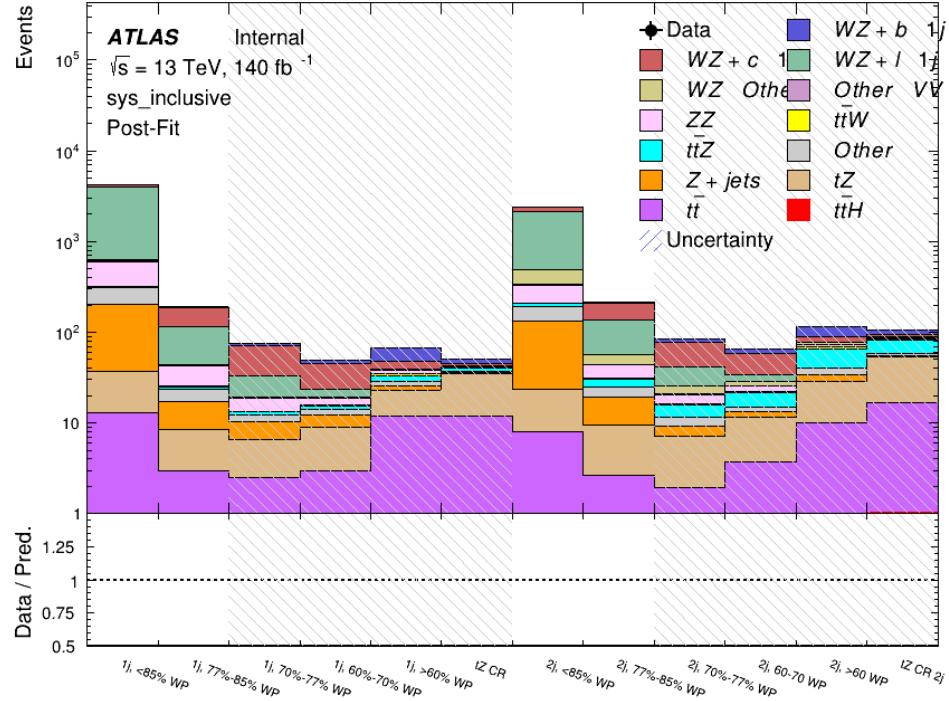


Figure 11.3: Post-fit summary of the fit regions.

818 The impact of the most significant sources of systematic uncertainties on the measurement
 819 of WZ + b is summarized in Table 12.

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

Table 12: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with one or two associated jets.

820 **11.4 Alternate tZ Inclusive Fit**

821 **11.4.1 tZ Inclusive Fit**

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

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

829 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
830 fit, with an expected significance of 4.0σ .

831 The impact of the predominate systematics are summarized in Table 44.

Uncertainty Source	$\Delta\mu$	
WZ + light cross-section	0.08	-0.08
Jet Energy Scale	-0.06	0.08
Luminosity	-0.05	0.06
WZ + c cross-section	-0.04	0.05
Other Diboson + b cross-section	-0.04	0.04
WZ cross-section - QCD scale	-0.04	0.03
t <bar>t</bar>	-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 13: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

832 11.4.2 Floating tZ

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

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

Part V**839 Differential Studies of $t\bar{t}H$ Multilepton****841 12 Data and Monte Carlo Samples****842 12.1 Data Samples**

843 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through
844 2018, which represents an integrated luminosity of 139 fb^{-1} [15] and an energy of $\sqrt{s} = 13$
845 TeV. All data used in this analysis was included in one of the Good Run Lists verified by Data
846 Quality checks [16].

847 12.2 Monte Carlo Samples

848 Several Monte Carlo (MC) generators were used to simulate both signal and background pro-
849 cesses. For all of these, the effects of the ATLAS detector are simulated in GEANT4 [17]. The
850 specific event generator used for each of these MC samples is listed in Table 14. A Higgs mass
851 of 125 GeV is assumed in all simulations.

Table 14: The configurations used for event generation of signal and background processes, including the event generator, matrix element (ME) order, parton shower algorithm, and parton distribution function (PDF).

Process	Event generator	ME order	Parton Shower	PDF
$t\bar{t}H$	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO [21]
	(MG5_AMC)	(NLO)	(HERWIG++)	(CT10 [19])
$t\bar{t}W$	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
	(SHERPA 2.1.1)	(LO multileg)	(SHERPA)	(NNPDF 3.0 NLO)
$t\bar{t}(Z/\gamma^* \rightarrow ll)$	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
$t\bar{t}$	POWHEG-BOX v2 [24]	NLO	PYTHIA 8	NNPDF 3.0 NLO
$t\bar{t}\gamma$	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
tZ	MG5_AMC	LO	PYTHIA 6	CTEQ6L1
$tHqb$	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC	NLO	HERWIG++	CT10
	(SHERPA 2.1.1)	(LO multileg)	(SHERPA)	(NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
$t\bar{t}t, t\bar{t}t\bar{t}$	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
$t\bar{t}W^+W^-$	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top qqVV, VVV $Z \rightarrow l^+l^-$	POWHEG-BOX v1 [25]	NLO	PYTHIA 6	CT10
	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

852 The signal sample ($t\bar{t}H$) is modelled at NLO with POWHEG-BOX v2 using the NNPDF2.0
 853 parton distribution function (PDF) [41]. Parton showering and hadronisation were modelled
 854 with PYTHIA 8.2 [42]. The $t\bar{t}H$ sample is normalized to a cross-section of 507^{+35}_{-50} fb based on
 855 NLO calculations. Uncertainties are based on varying the QCD factorisation and renormalisation
 856 scale, as well as uncertainties in the PDF and α_s .

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

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

864 The $t\bar{t}(Z/\gamma^*)$ process is simulated with the **MADGRAPH5_AMC@NLO** generator, using
865 NNPDF3.0. Diboson processes are generated with **SHERPA 2.2.2** at NLO precision for one extra
866 parton, and at LO for up to three extra partons.

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

874 The primary contribution to the non-prompt lepton background is from $t\bar{t}$ production, with
875 $V+jets$ and single-top as much smaller sources. Estimation of this background is done primarily
876 using an inclusive $t\bar{t}$ sample, with corrections applied based on data driven methods. This sample
877 is generated using **POWHEG**, with **PYTHIA8** performing the parton shower and fragmentation.
878 Likelihood fits over several control regions enriched with these non-prompt backgrounds are fit
879 to data in order to derive normalization factors for these backgrounds. The specific normalization

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

881 Other processes, such as tH , tZ , $t\bar{t}WW$ and $t\bar{t}t\bar{t}$, are expected to make minor contributions
 882 to the total background. The generators and setting used for these backgrounds are summarized
 883 in Table 14.

884 13 Object Reconstruction

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

892 Electron candidates are reconstructed from energy clusters in the electromagnetic calor-
 893 imeter that are associated with charged particle tracks reconstructed in the inner detector
 894 [43]. Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Can-
 895 didates in the transition region between different electromagnetic calorimeter components,
 896 $1.37 < |\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant combining shower
 897 shape and track information is used to distinguish prompt electrons from nonprompt leptons,
 898 such as those originating from hadronic showers.

899 To further reduce the non-prompt contribution, the track of each electron is required to
 900 originate from the primary vertex; requirements are imposed on the transverse impact parameter
 901 significance ($|d_0|/\sigma_{d_0}$) and the longitudinal impact parameter ($|\Delta z_0 \sin \theta_\ell|$). Muon candidates
 902 are reconstructed by combining inner detector tracks with track segments or full tracks in the
 903 muon spectrometer [28]. Muon candidates are required to have $p_T > 10$ GeV and $|\eta| < 2.5$.

904 All leptons are required to pass a non-prompt BDT selection developed by the main
 905 $t\bar{t}H/t\bar{t}W$ analysis, described in detail in [7]. Optimized working points and scale factors for this
 906 BDT are taken from that analysis. This BDT and the WPs used are summarized in Appendix
 907 [A](#),

908 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
 909 calorimeters [44], using the anti- k_t algorithm with a radius parameter $R = 0.4$. Particle Flow, or
 910 PFlow, jets are used in the analysis, which are hadronic objects reconstructed using information
 911 from both the tracker and the calorimeter. Jets with energy contributions likely arising from noise
 912 or detector effects are removed from consideration [45], and only jets satisfying $p_T > 25$ GeV
 913 and $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
 914 association algorithm is used to confirm that the jet originates from the selected primary vertex,
 915 in order to reject jets arising from pileup collisions [31].

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

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

922 Because all $t\bar{t}H$ – ML channels considered include multiple neutrinos, missing transverse
923 energy (E_T^{miss}) is present in each event. The missing transverse momentum vector is defined as
924 the inverse of the sum of the transverse momenta of all reconstructed physics objects as well
925 as remaining unclustered energy, the latter of which is estimated from low- p_T tracks associated
926 with the primary vertex but not assigned to a hard object [46].

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

934 14 Higgs Momentum Reconstruction

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

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

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

949 Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_AMC.
950 Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training in order to
951 increase the statistics of the training sample.

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

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

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

14.1 Physics Object Truth Matching

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

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

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

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

988 14.2 b-jet Identification

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

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

1002 Once the network is trained, kinematic variables for all possible pairings of jets are fed into
 1003 the model, and the pair of jets with the highest output score are taken to be b-jets in successive
 1004 steps of the analysis.

1005 An alternate approach is considered, where information about all jets in each event are
 1006 used as the feature set, and the model is tasked with identifying which two originated from the
 1007 top decay. While this approach is found to underperform the nominal approach, and therefore
 1008 not used in the analysis, the results are documented in Appendix [C.4](#).

1009 **14.2.1 2lSS Channel**

1010 For the 2lSS channel, the input features shown in Table [15](#) are used for training. Here j_0 and j_1
 1011 are the two jet candidates, while l_0 and l_1 are the two leptons in the event, both ordered by p_T . jet
 1012 DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,

1013 where 5 represents the tightest working point and 1 represents the loosest. The variables nJets
 1014 DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and
 1015 85% b-tag working points, respectively.

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

Table 15: Input features used in the b-jet identification algorithm for the 2lSS channel

1016 As there are far more incorrect combinations than correct ones, by a factor of more than
 1017 20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random
 1018 sample of 5 million incorrect entries are used for training, along with around 1 million correct
 1019 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for
 1020 training.

1021 The difference between the distributions for a few of these features for the "correct" (i.e.
 1022 both jets are truth b-jets), and "incorrect" combinations are shown in Figure 14.1. The correct and
 1023 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences
 1024 in the distributions.

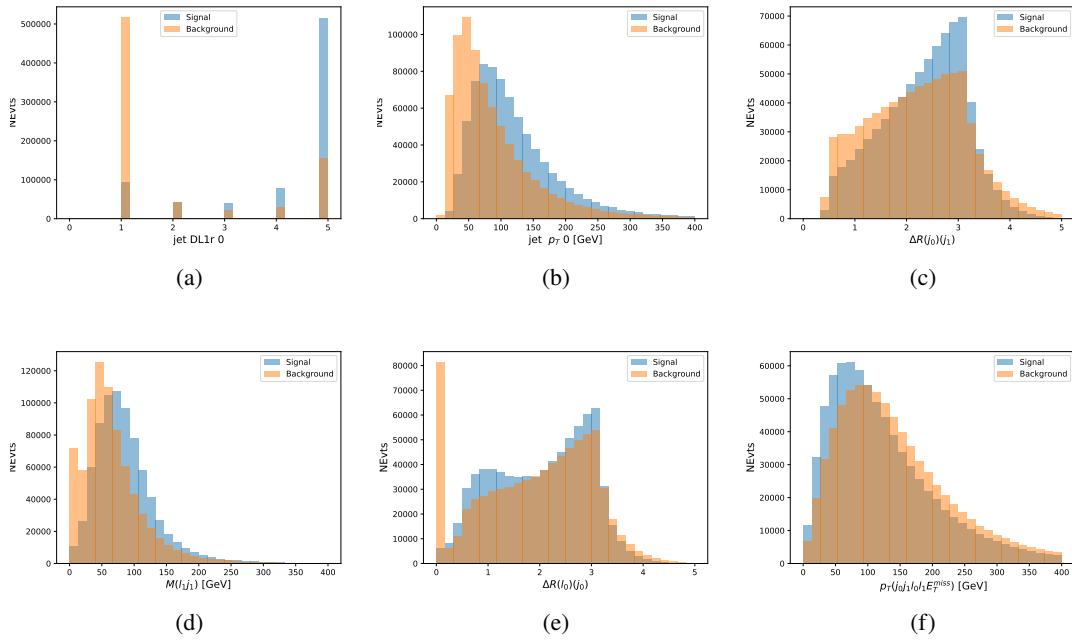


Figure 14.1: Input features for top2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

1025 The modeling of these inputs is validated against data, with Figure 14.2 showing good
1026 general agreement between data and MC. Plots for the complete list of features can found in
1027 Appendix C.

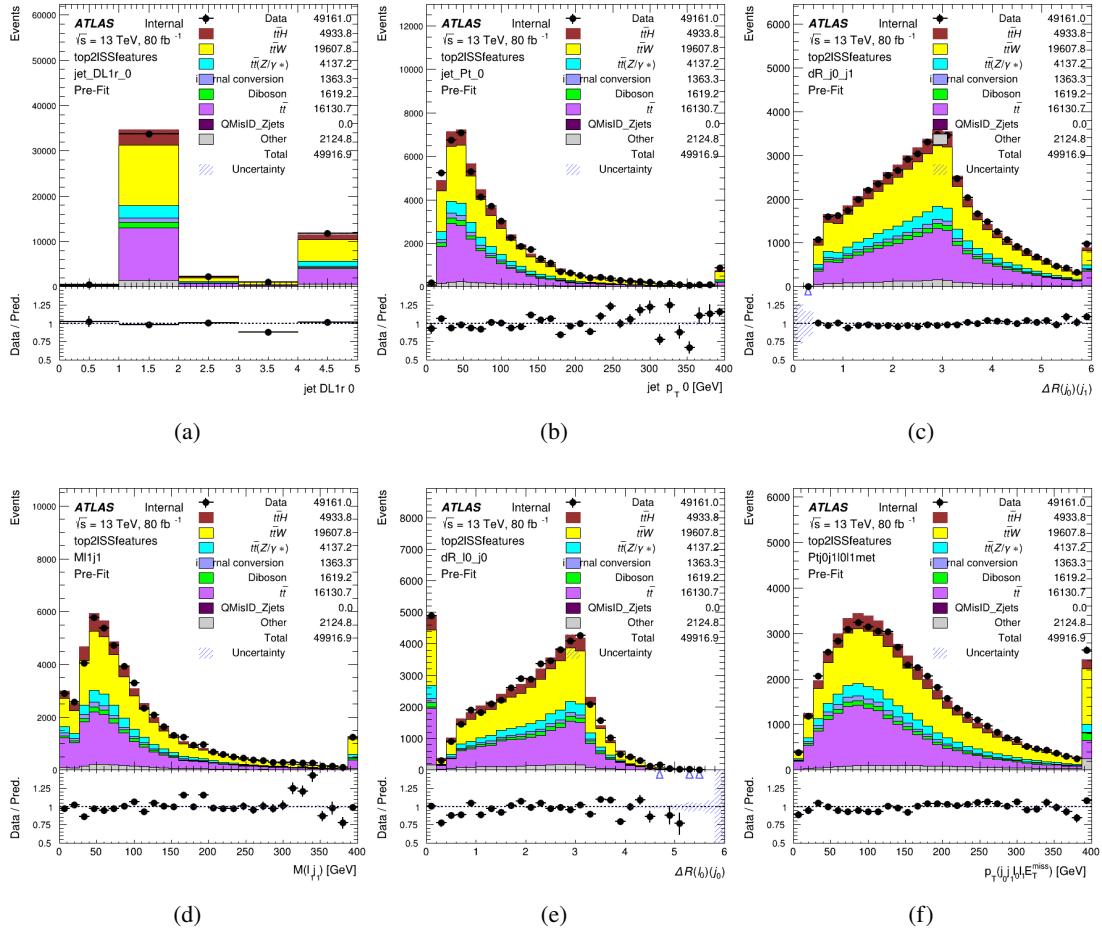


Figure 14.2: Data/MC comparisons of input features for top2lSS training for 79.8 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and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

1028 Based on the results of grid search evaluation, the optimal architecture is found to include
 1029 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as
 1030 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for
 1031 the trained model are shown in Figure 14.2.1. The model is found to identify the correct pairing
 1032 of jets for 73% of 2lSS signal events on test data.

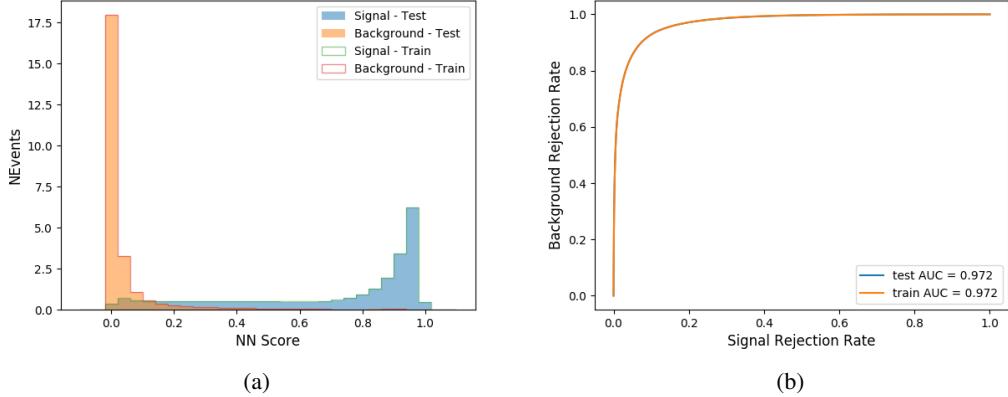


Figure 14.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1033 For point of comparison, a "naive" approach to identifying b-jets is used as well: The two
 1034 jets which pass the highest DL1r b-tag working point are assumed to be the b-jets from the top
 1035 decay. In the case that multiple jets meet the same b-tag working point, the jet with higher p_T is
 1036 used. This method identifies the correct jet pair 65% of the time.

1037 The accuracy of the model for different b-tagged jet multiplicities, compared to this naive
 1038 approach, is shown in Table 16.

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

Table 16: Accuracy of the NN in identifying b-jets from tops in 2lSS events, overall and split by the number of b-tagged jets in the event, compared to the accuracy of taking the two highest b-tagged jets.

1039 This suggests that when there are exactly two b-tagged jets in an event, little is gained by
 1040 using this more sophisticated approach, while for events with 1 or ≥ 3 b-tagged jets, the model
 1041 does provide significant improvements.

1042 **14.2.2 3l Channel**

1043 The input features used in the 3l channel are listed in Table 17, with the same naming convention
 1044 as the 2lSS channel.

jet p_T 0	jet p_T 1	jet η 0
jet η 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^{miss}	nJets OR DL1r 85
nJets OR DL1r 60		

Table 17: Input features for the b-jet identification algorithm in the 3l channel.

1045 A few of these features are shown in Figure 14.4, comparing the distributions for correct
 1046 and incorrect combinations of jets.

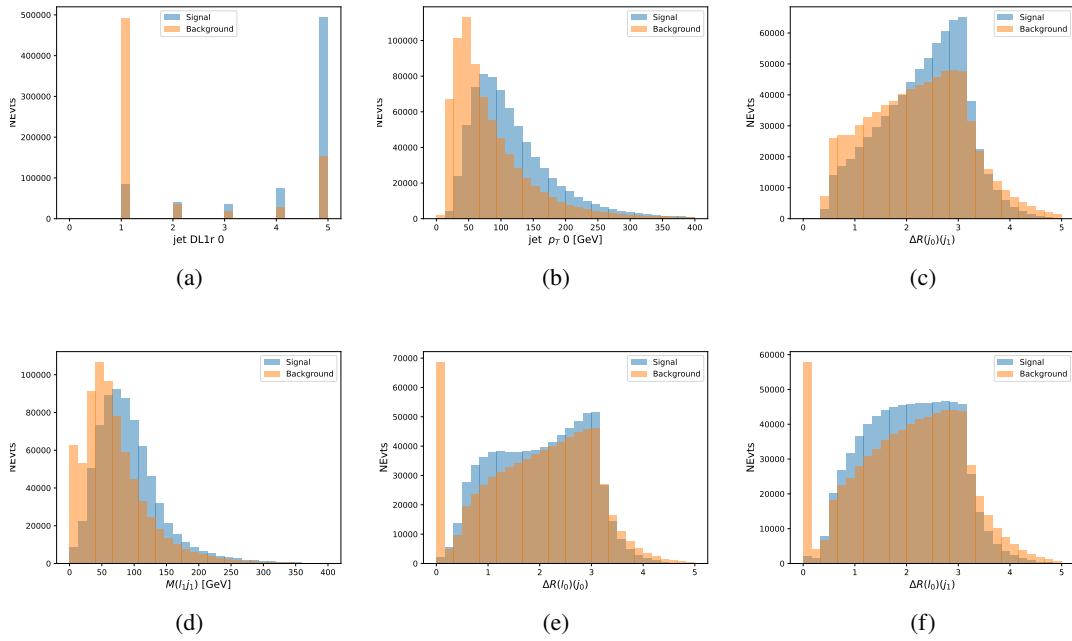


Figure 14.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

1047

The modeling of these inputs is validated against data, with Figure 14.5 showing good

1048

general agreement between data and MC.

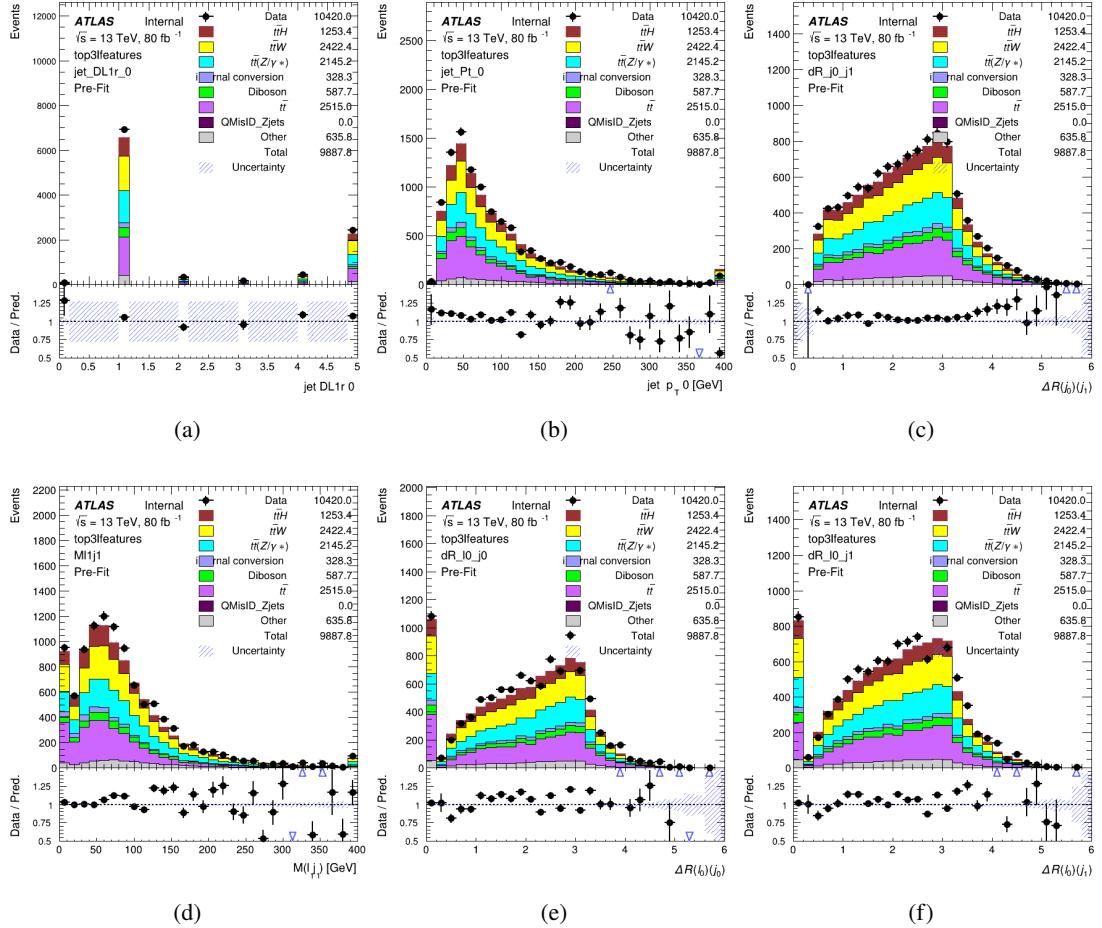


Figure 14.5: Data/MC comparisons of input features for top3l training for 79.8 fb^{-1} of data. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

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 14.2.2.

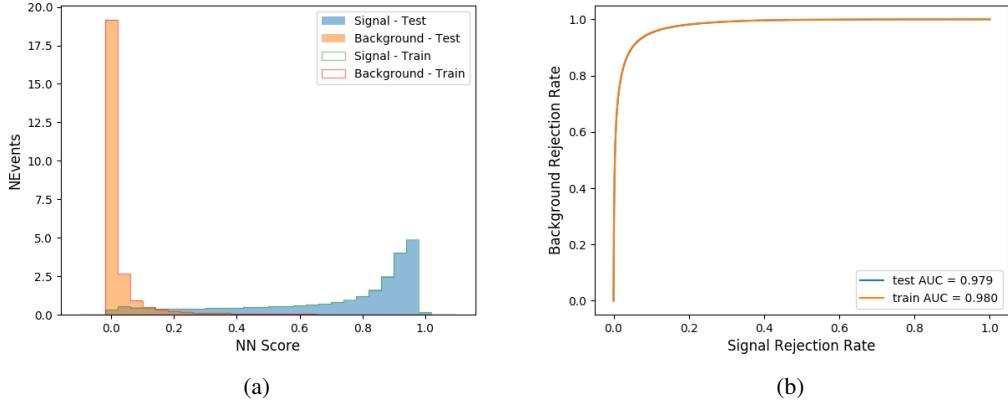


Figure 14.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1054 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal
 1055 events. The accuracy of the model is summarized in Table 18, once again compared to the naive
 1056 approach described above.

Table 18: 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%
≥ 3 b-jets	55.7%	52.3%
Overall	79.8%	70.2%

14.3 Higgs Reconstruction

1057 Techniques similar to the b-jet identification algorithms are employed to select the decay products
 1058 of the Higgs: kinematics of all possible combinations of reconstructed objects are fed into a neural
 1059

1060 network to determine which of those is most mostly to be the decay products of the Higgs.

1061 Again separate models are used for the 2lSS and 3l channels, while the 3l channel has
1062 now been split into two: $t\bar{t}H$ events with three leptons in the final state include both instances
1063 where the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the
1064 Higgs decays to two leptons (and two neutrinos which are not reconstructed).

1065 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In
1066 the semi-leptonic case the reconstructed decay products consist of two jets and a single lepton.
1067 For the fully-leptonic case, the decay products include 2 of the three leptons associated with
1068 the event. For training these models, events are separated into these two categories using truth
1069 level information. A separate MVA, described in Section 14.5, is used to make this distinction
1070 at reconstructed level, and determine which model to use.

1071 For all channels, the models described in Section 14.2 are used to identify b-jet candidates,
1072 whose kinematics are used as additional input features to help identify the Higgs decay products.
1073 These jets are not considered as possible candidates for the Higgs decay, justified by the fact that
1074 these models are found to misidentify jets from the Higgs decay as jets from the top decay less
1075 than 1% of the time.

1076 **14.3.1 2lSS Channel**

1077 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural
1078 network is trained on the kinematics of different combinations of leptons and jets, as well as the

¹⁰⁷⁹ b-jets identified in Section 14.2, with the specific input features listed in Table 19.

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

Table 19: Input features used to identify the Higgs decay products in 2LSS events

¹⁰⁸⁰ Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
¹⁰⁸¹ lepton in the event is labeled l_T , as it is assumed to have come from the decay of one of the top
¹⁰⁸² quarks. b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet
¹⁰⁸³ Reco Score is the output of the b-jet reconstruction algorithm.

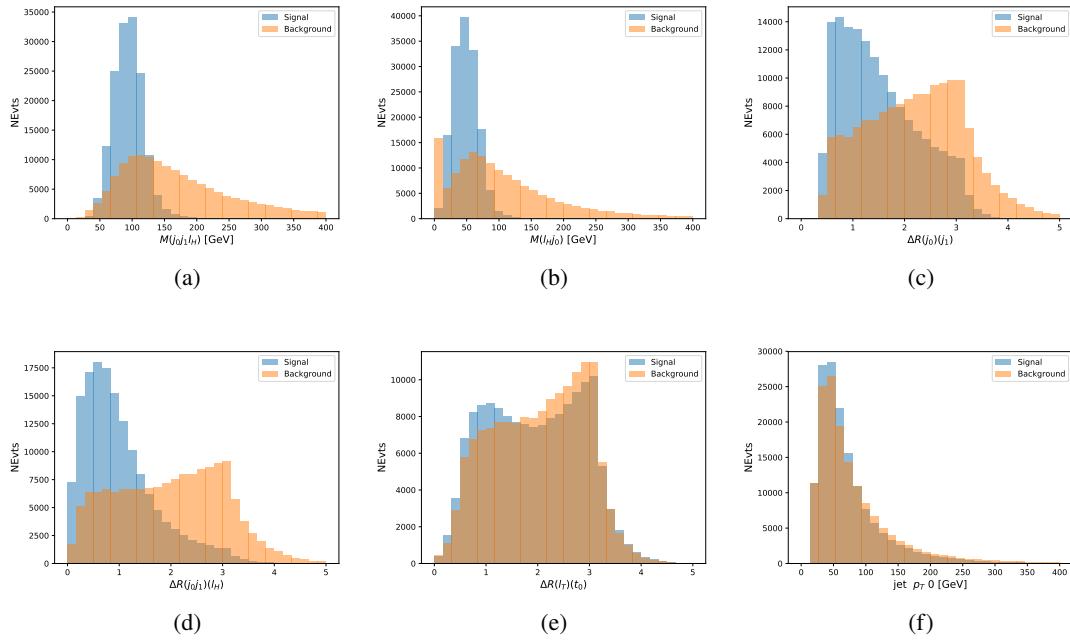


Figure 14.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 and the lepton candidate, (e) ΔR between the lepton from the top and the leading b-jet, (f) the p_T of jet 0.

1084 The modeling of these inputs is validated against data, with Figure 14.2 showing good
1085 general agreement between data and MC. Plots for the complete list of features can found in
1086 Section C.

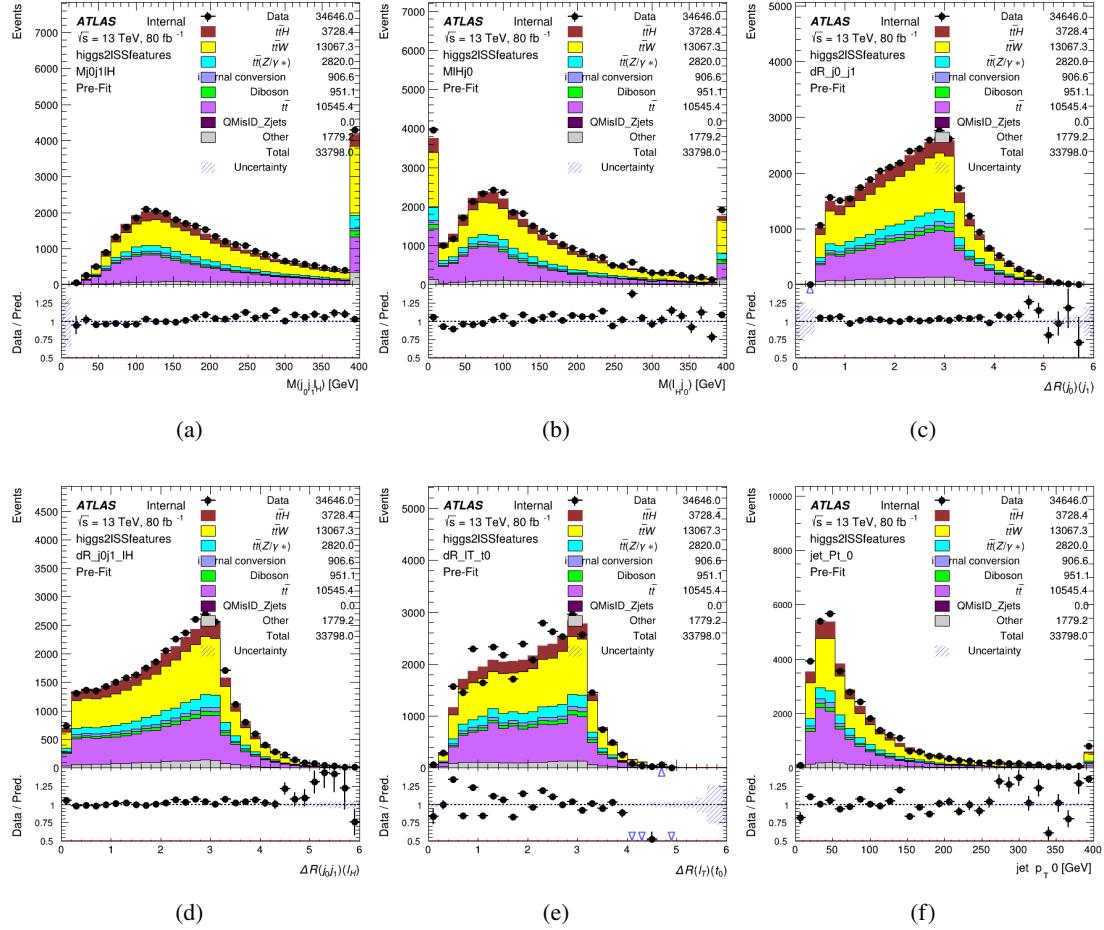


Figure 14.8: Data/MC comparisons of input features for higgs2ISS training for 79.8 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 and the lepton candidate, (e) ΔR between the lepton from the top and the leading b-jet, (f) the p_T of jet 0.

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

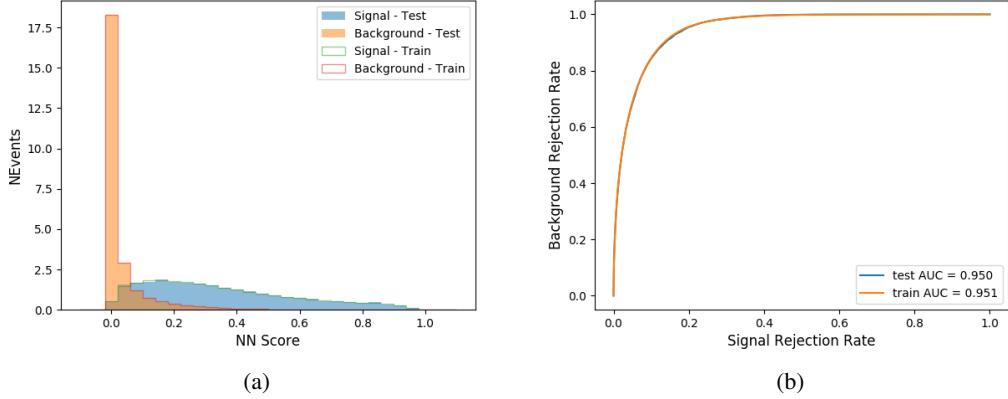


Figure 14.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1092 The neural network identifies the correct combination 55% of the time. It identifies the
 1093 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
 1094 81% of the time.

1095 14.3.2 3l Semi-leptonic Channel

1096 For 3l $t\bar{t}H$ where the Higgs decay semi-leptonically, the decay products include one of the three
 1097 leptons and two jets. In this case, the other two leptons originated from the decay of the tops,
 1098 meaning the opposite-sign (OS) lepton cannot have come from the Higgs. This leaves only the two
 1099 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 η 0	jet η 1
jet ϕ 0	jet ϕ 1	$\Delta R(j_0)(j_1)$
$M(j_0j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0j_1)(l_H)$	$\Delta R(j_0j_1)(l_{T_0})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0j_1l_{T_0})$	$M(j_0j_1l_{T_1})$
$M(j_0j_1l_H)$	$\Delta R(j_0j_1l_H)(l_{T_0})$	$\Delta R(j_0j_1l_H)(l_{T_1})$
$\Delta\phi(j_0j_1l_H)(E_T^{\text{miss}})$	$p_T(j_0j_1l_Hl_{T_0}l_{T_1}b_0b_1E_T^{\text{miss}})$	$M(j_0j_1b_0)$
$M(j_0j_1b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
b-jet Reco Score	E_T^{miss}	HT jets
nJets		

Table 20: Input features used to identify the Higgs decay products in 3l semi-leptonic events

1100 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 1101 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
 1102 the b-jet identification algorithm. The b-jet Reco Score is the output of the b-jet identification
 1103 algorithm.

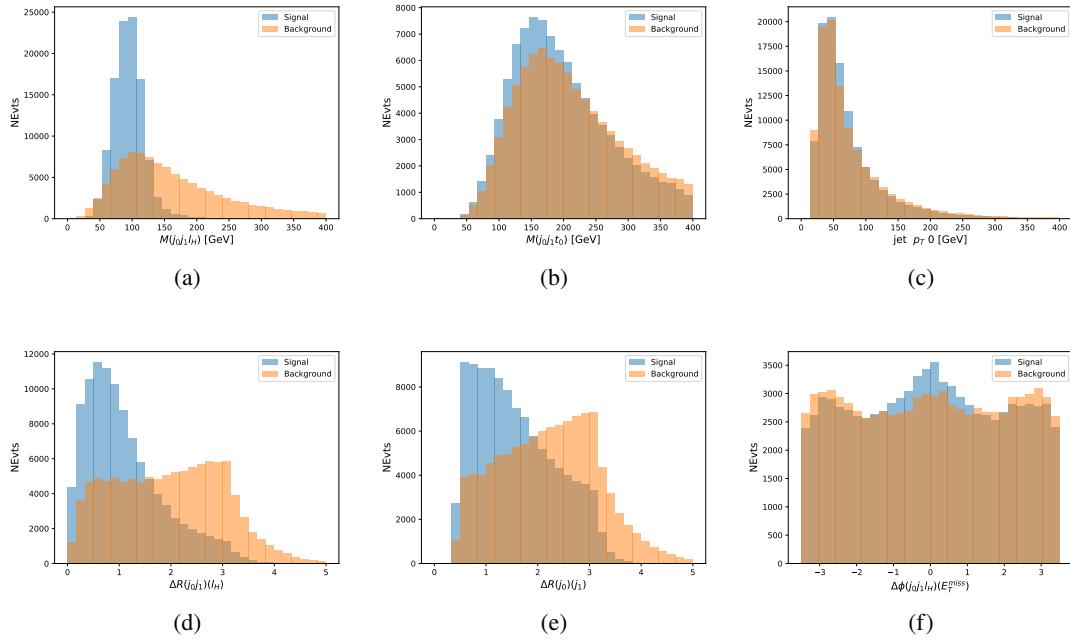


Figure 14.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1104 The modeling of these inputs is validated against data, with Figure 14.11 showing good
 1105 general agreement between data and MC. Plots for the complete list of features can found in
 1106 appendix C.1.

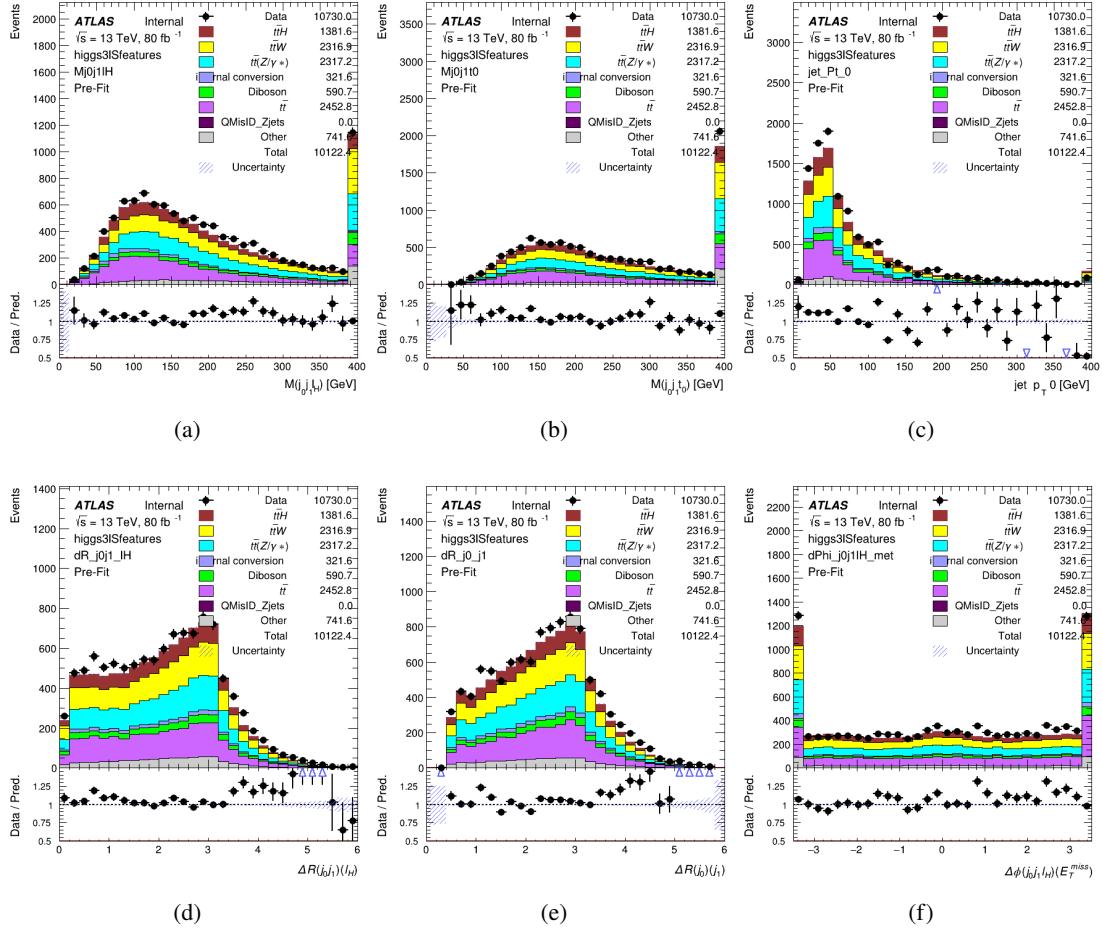


Figure 14.11: Data/MC comparisons of input features for higgs3lS training for 79.8 fb^{-1} of data.

1107 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.
 1108 Once again, incorrect combinations are downsampled, such that the correct combinations are
 1109 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN
 1110 is summarized in Figure 14.3.2.

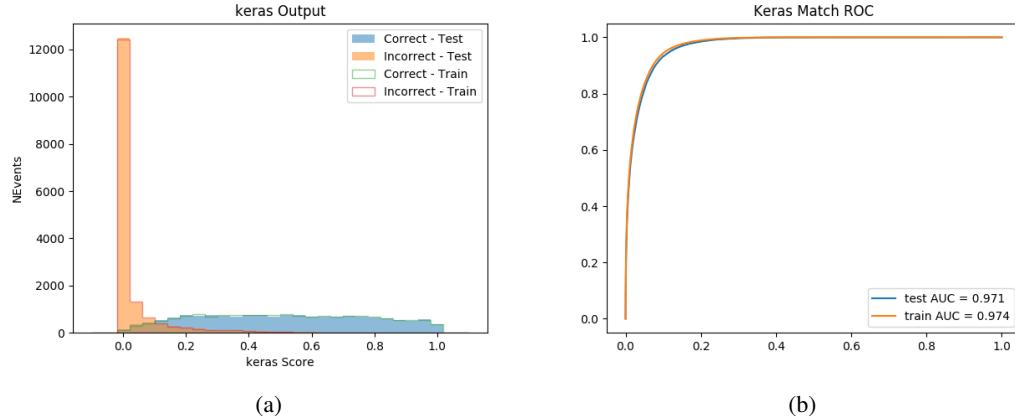


Figure 14.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1111 The neural network identifies the correct combination 64% of the time. It identifies the
 1112 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
 1113 83% of the time.

1114 14.3.3 3l Fully-leptonic Channel

1115 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 1116 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to
 1117 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 1118 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 1119 and one where it is not.

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

Table 21: Input features used to identify the Higgs decay products in 3lF events

Table 22: Input features used to identify the Higgs decay products in 3l fully leptonic events

1120 Here l_{H_0} and l_{H_1} are the Higgs decay candidates. The other lepton in the event is labeled
 1121 l_T . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco
 1122 Score is the output of the Higgs reconstruction algorithm.

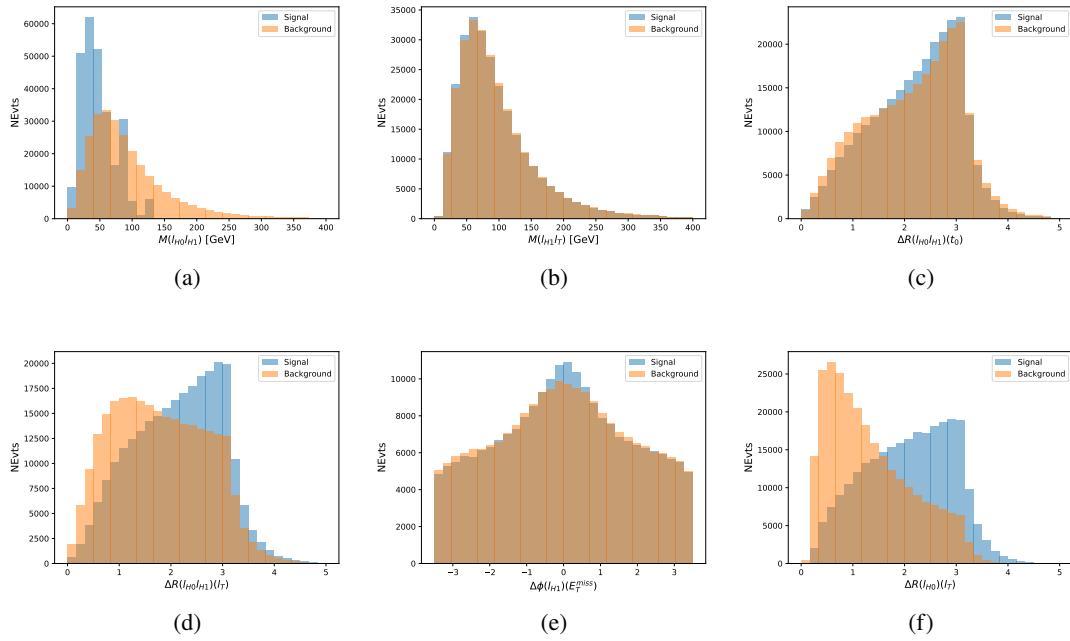


Figure 14.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1123 The modeling of these inputs is validated against data, with Figure 14.14 showing good
 1124 general agreement between data and MC. Plots for the complete list of features can found in
 1125 Section C.

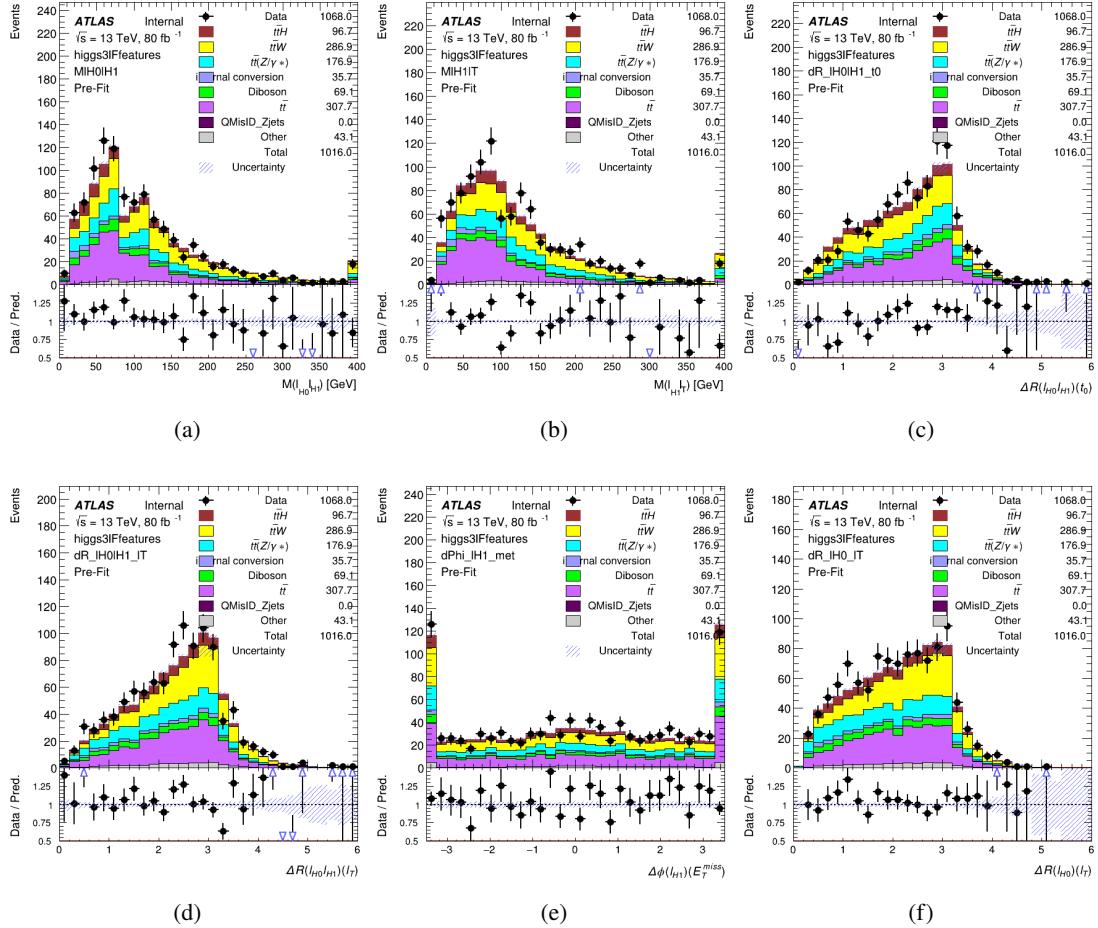


Figure 14.14: Data/MC comparisons of input features for higgs3lF training for 79.8 fb^{-1} of data.

1126 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,
 1127 with 10% of the dataset reserved for testing. The output of the model is summarized in Figure
 1128 [14.3.3.](#).

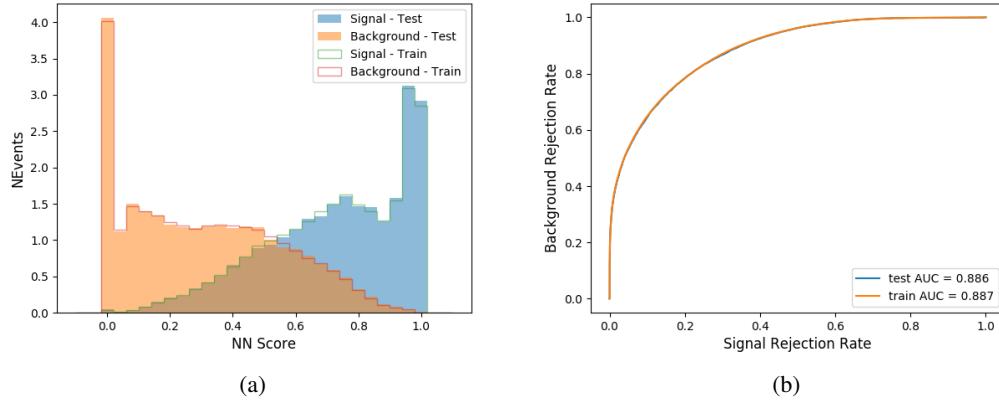


Figure 14.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

1129 The correct lepton is identified by the model for 80% of events in the testing data set.

1130 14.4 p_T Prediction

1131 Once the most probable decay products have been identified, their kinematics are used as inputs
 1132 to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,
 1133 a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay
 1134 are included as well, as these are found to improve performance. The truth p_T of the Higgs,
 1135 as predicted by MC, are used as labels. Separate models are built for each channel - 2lSS, 3l
 1136 Semi-leptonic and 3l Fully-leptonic.

1137 As a two-bin fit is targeted for the final result, some metrics evaluating the performance
 1138 of the models aim to show how well it distinguished between "high p_T " and "low p_T " events. A

1139 cutoff point of 150 GeV is used to define these two categories.

1140 Because the analysis uses a two bin fit of the Higgs p_T , the momentum reconstruction
1141 could be treated as a binary classification problem, rather than a regression problem. This
1142 approach is explored in detail in Section C.5, and is found not to provide any significant increase
1143 in sensitivity. The regression approach is used because it provides more flexibility for future
1144 analyses, as it is independent of the cutoff between high and low p_T , as well as the number of
1145 bins. Further, a regression allows the output of the neural network to be more clearly understood,
1146 as it can be directly compared to a physics observable.

1147 14.4.1 2lSS Channel

1148 The input variables listed in Table 23 are used to predict the Higgs p_T in the 2lSS channel. Here
1149 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
1150 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
1151 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
1152 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
1153 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 η 0
jet η 1	jet Phi 0	jet Phi 1
jet p_T 0	jet p_T 1	Lepton η H
Lepton ϕ H	Lepton p_T H	Lepton p_T T
E_T^{miss}	nJets	b-jet Reco Score
b-jet p_T 0	b-jet p_T 1	

Table 23: Input features for reconstructing the Higgs p_T spectrum for 2lSS events

1154 The optimal neural network architecture for this channel is found to consist of 7 hidden
 1155 layers with 60 nodes each. The input data set includes 1.2 million events, 10% of which is used
 1156 for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1157 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 1158 truth Higgs p_T in Figure 14.16. In order to visualize the model performance more clearly, in (a)
 1159 of that figure, the color of each point is determined by Kernel Density Estimation (KDE). The
 1160 color shown represents the logarithm of the output from KDE, to counteract the large number of
 1161 low p_T events. For that same reason, each column of the histogram shown in (b) of Figure 14.16
 1162 is normalized to unity. This plot therefore demonstrates what the model predicts for each slice
 1163 of truth p_T .

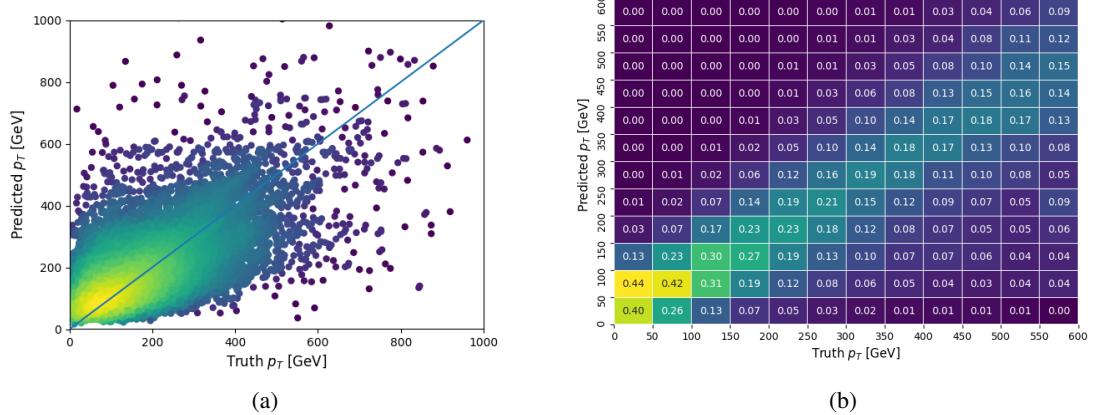


Figure 14.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized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

1164 We are also interested in how well the model distinguishes between events with $p_T < 150$
 1165 GeV and > 150 GeV. Figure 14.17 demonstrates the NN output for high and low p_T events based
 1166 on this cutoff.

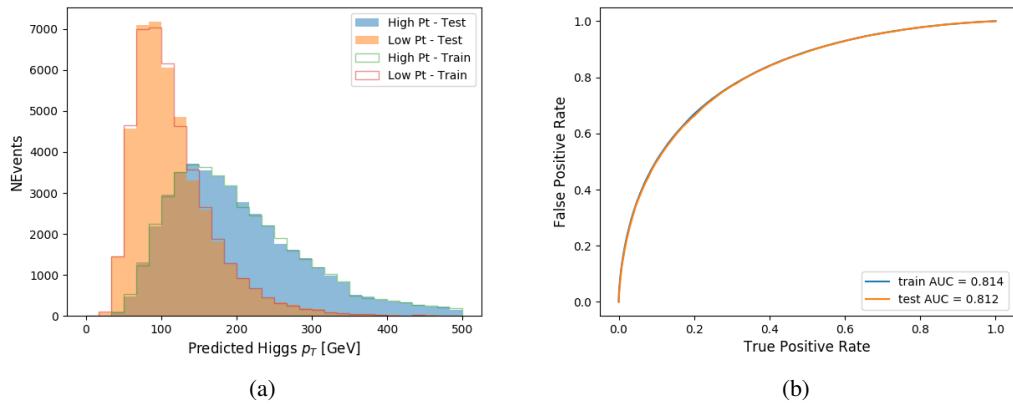


Figure 14.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

¹¹⁶⁷ **14.4.2 3l Semi-leptonic Channel**

¹¹⁶⁸ The following input features are used to predict the Higgs p_T for events in the 3lS channel:

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

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

¹¹⁶⁹ Again, j_0 and j_1 are the two jets identified as Higgs decay products, ordered by p_T . The
¹¹⁷⁰ lepton identified as originating from the Higgs is labeled l_H , while the other two leptons are
¹¹⁷¹ labeled l_{T0} and l_{T1} . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm.
¹¹⁷² The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs reconstruction algorithm,
¹¹⁷³ and the b-jet identification algorithm, respectively.

¹¹⁷⁴ The optimal neural network architecture for this channel is found to consist of 7 hidden
¹¹⁷⁵ layers with 80 nodes each. The inputdata set includes one million events, 10% of which is used
¹¹⁷⁶ for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1177 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 1178 truth Higgs p_T in Figure 14.18. Once again, (a) of 14.18 shows a scatterplots of predicted vs
 1179 truth p_T , where the color of each point corresponds to the log of the relative KDE at that point.
 1180 Each column of the the histogram in (b) is normalized to unity, to better demonstrate the output
 1181 of the NN for each slice of truth p_T .

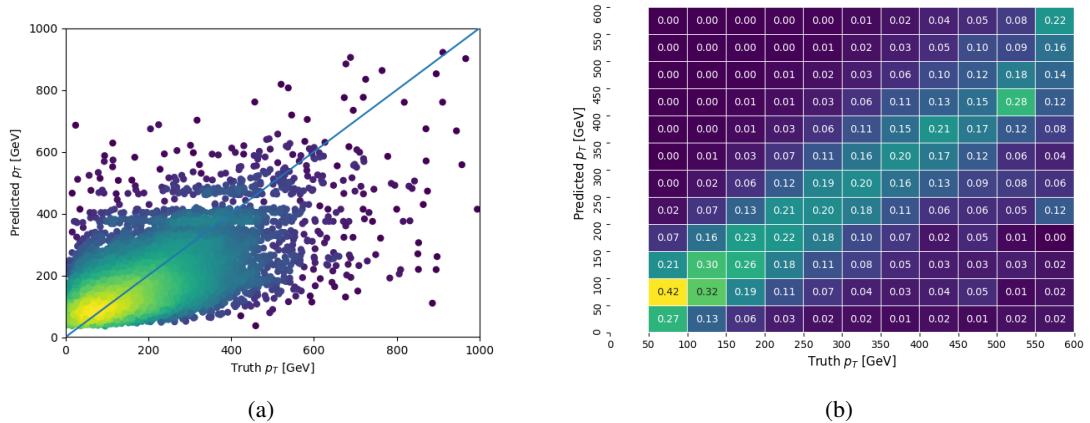


Figure 14.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

1182 Figure 14.19 shows (a) the output of the NN for events with truth p_T less than and greater
 1183 than 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low
 1184 p_T events.

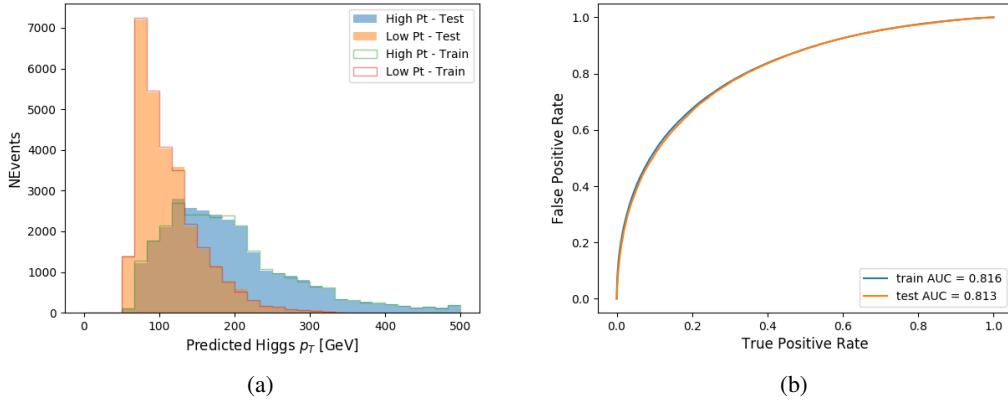


Figure 14.19: (

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

14.4.3 3l Fully-leptonic Channel

The features listed in 25 are used to construct a model for predictin the Higgs p_T for 3lF events.

HT	$M(l_{H0} l_{H1})$	$M(l_{H0} l_T)$
$M(l_{H0} b_0)$	$M(l_{H0} b_1)$	$M(l_{H1} l_T)$
$M(l_{H1} b_0)$	$M(l_{H1} b_1)$	$\Delta R(l_{H0})(l_{H1})$
$\Delta R(l_{H0})(l_T)$	$\Delta R(l_{H0} l_{H1})(l_T)$	$\Delta R(l_{H0} l_T)(l_{H1})$
$\Delta R(l_{H1})(l_T)$	$\Delta R(l_{H0} b_0)$	$\Delta R(l_{H0} b_1)$
$\Delta R(l_{H1} b_1)$	$\Delta R(l_{H1} b_0)$	Higgs Reco Score
Lepton η H_0	Lepton η H_1	Lepton η T
Lepton p_T H_0	Lepton p_T H_1	Lepton p_T T
E_T^{miss}	b-jet Reco Score	b-jet p_T 0
b-jet p_T 1		

Table 25: Input features for reconstructing the Higgs p_T spectrum for 3lF events

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

1188 originating from the Higgs, while l_T is the other lepton in the event. The Higgs Reco Score and
 1189 b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification
 1190 algorithm, respectively.

1191 The optimal neural network architecture for this channel is found to consist of 5 hidden
 1192 layers with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for
 1193 testing, the other 90% for training. Training is found to converge after around 150 epochs.

1194 The predicted transverse momentum, as a function of the truth p_T , is shown in Figure
 1195 [14.20](#).

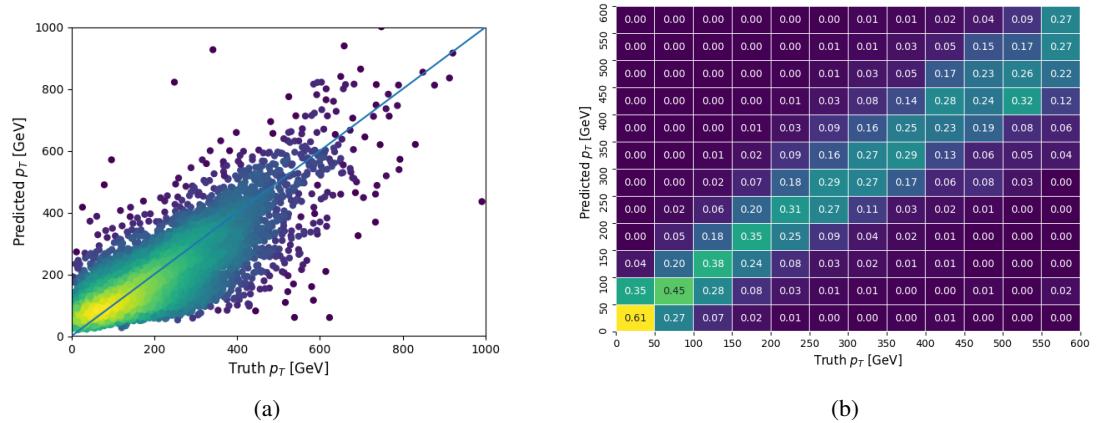


Figure 14.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

1196 When split into high and low p_T , based on a cutoff of 150 GeV, the

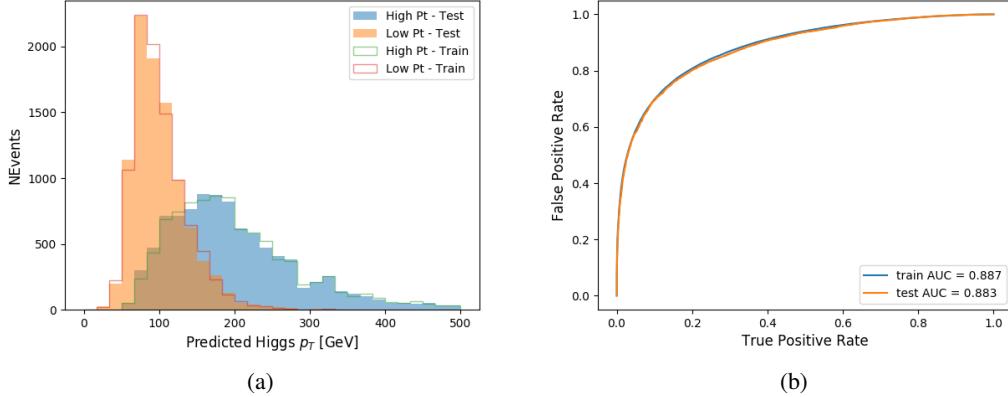


Figure 14.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

14.5 3l Decay Mode

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

The kinematics of each event, along with the output scores of the Higgs and top reconstruction algorithms, are used to distinguish these two possible decay modes. The particular inputs used are listed in Table 26.

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 η 0	Lepton η 1	Lepton η 2
Lepton ϕ 0	Lepton ϕ 1	Lepton ϕ 2
Lepton p_T 0	Lepton p_T 1	Lepton p_T 2
E_T^{miss}	nJets	nJets OR DL1r 60
nJets OR DL1r 85	score3lF	score3lS
topScore	total charge	

Table 26: Input features used to distinguish semi-leptonic and fully-leptonic Higgs decays in the 3l channel.

1207 Here l_0 is the opposite charge lepton, l_1 and l_2 are the two SS leptons order by ΔR
 1208 from lepton 0. score3lF and score3lS are the outputs of the 3lS and 3lF Higgs reconstruction
 1209 algorithms, while topScore is the output of the b-jet identification algorithm.

1210 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these
 1211 two decay modes. The output of the model is summarized in Figure 14.22.

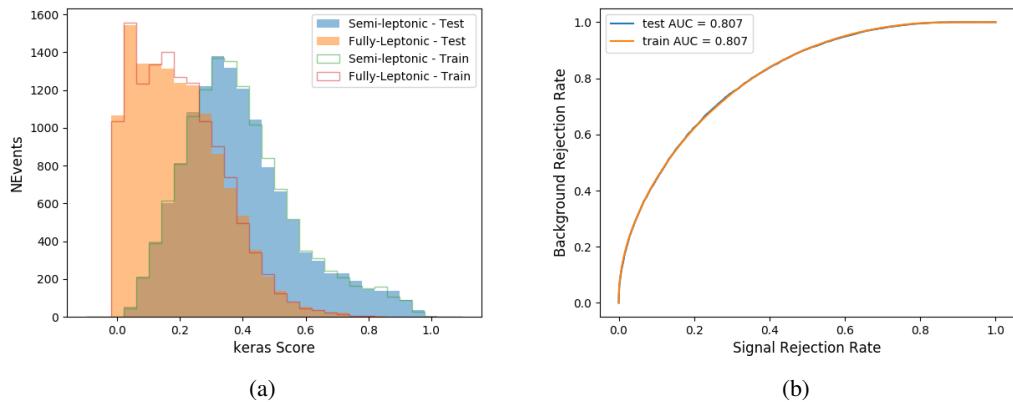


Figure 14.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

1212 A cutoff of 0.23 is determined to be optimal for separating 3lS and 3lF in the fit.

1213 15 Signal Region Definitions

1214 Events are divided into two channels based on the number of leptons in the final state: one
1215 with two same-sign leptons, the other with three leptons. The 3l channel includes events where
1216 two leptons originated from the Higgs boson as well as events where only one of the leptons
1217 originated from the Higgs. This motivates splitting the 3l channel into semi-leptonic, and fully
1218 leptonic channels, after an event preselection has been applied.

1219 15.1 Pre-MVA Event Selection

1220 A preselection is applied to define orthogonal analysis channels based on the number of leptons
1221 in each event. For the 2lSS channel, the following preselection is used:

- 1222 • Two very tight, same-charge, light leptons with $p_T > 20 \text{ GeV}$
- 1223 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets
- 1224 • No reconstructed tau candidates

1225 The event yield after the 2lSS preselection has been applied, for MC and data at 79.8 fb^{-1} ,
1226 is shown in Table 15.1.

Process	Yield
t̄tH high p _T	41 ± 5
t̄tH low p _T	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 27: Yields of the 2lSS preselection region

1227

Figure 15.1. Good general agreement is found.

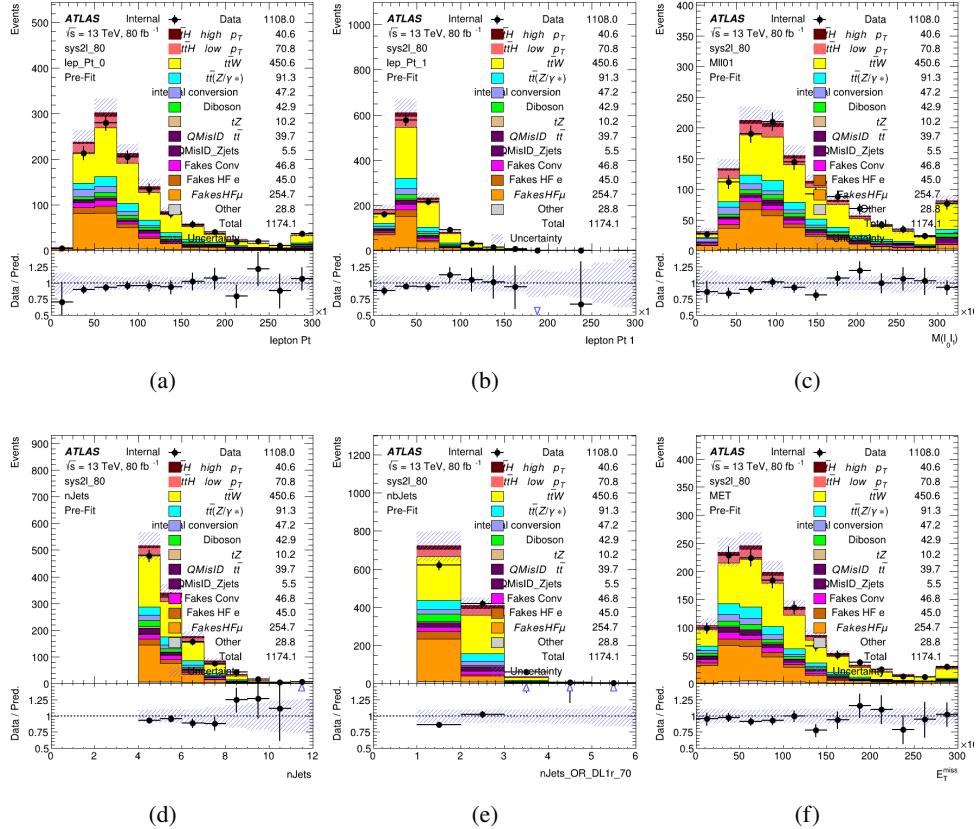


Figure 15.1: Data/MC comparisons of the 2lSS pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1228

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

1229

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

1231

- 1233 • No reconstructed tau candidates
- 1234 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs
- 1235 The event yield after the 3l preselection has been applied, for MC and data at 79.8 fb^{-1} ,
- 1236 is shown in Table 15.1.

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

Table 28: Yields of the 3l preselection region.

Table 29: Yields of the 3l preselection region.

1237 Comparisons of kinematic distributions for data and MC in this region are shown in Figure

1238 15.2.

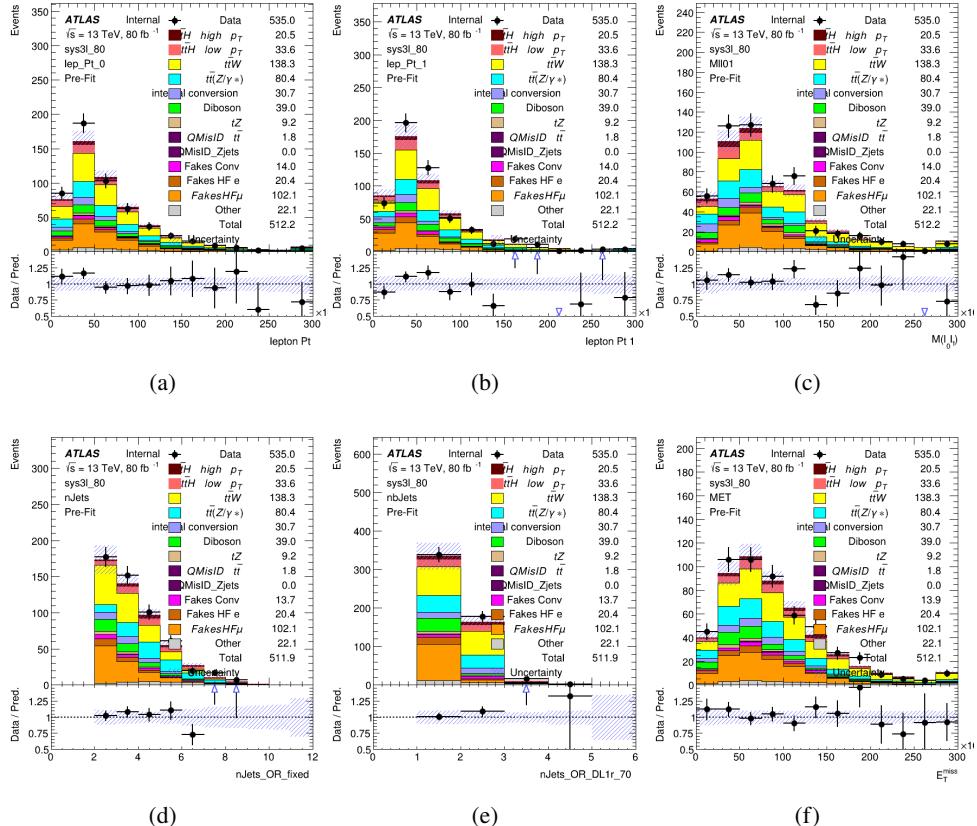


Figure 15.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1239 15.2 Event MVA

1240 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
 1241 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
 1242 (3lF). Here events with three leptons are split into 3lS and 3lF based on the model described
 1243 in 14.5. In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

1244 [34] are trained using the kinematics of signal and background events derived from Monte Carlo
1245 simulations. Events are weighted in the BDT training by the weight of each Monte Carlo event.

1246 Because the background composition differs for events with a high reconstructed Higgs p_T
1247 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and
1248 low p_T regions. This is found to provide better significance than attempting to build an inclusive
1249 model, as demonstrated in appendix C.2. A cutoff of 150 GeV is used. This gives a total of 6
1250 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
1251 3lF high p_T , and 3lF low p_T .

1252 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 η 0	jet η 1
jet ϕ 0	jet ϕ 1	jet p_T 0
jet p_T 1	Lepton η 0	Lepton η 1
Lepton ϕ 0	Lepton ϕ 1	Lepton p_T 0
Lepton p_T 1	E_T^{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 30: Input features used to distinguish signal and background events in the 2lSS channel.

1253 While for each of the 3l BDTs, the features listed below are used for training:

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

Table 31: Input features used to distinguish signal and background events in the 3l channel.

1254 Modelling of each of these input features is verified in Appendix C.2 by comparing data
 1255 and MC for 79.8 fb^{-1} . The BDTs are produced with a maximum tree depth of 6, using AUC as
 1256 the target loss function. The BDT response distribution and ROC curve for each model is shown
 1257 in Figures 15.3-15.5.

2lSS

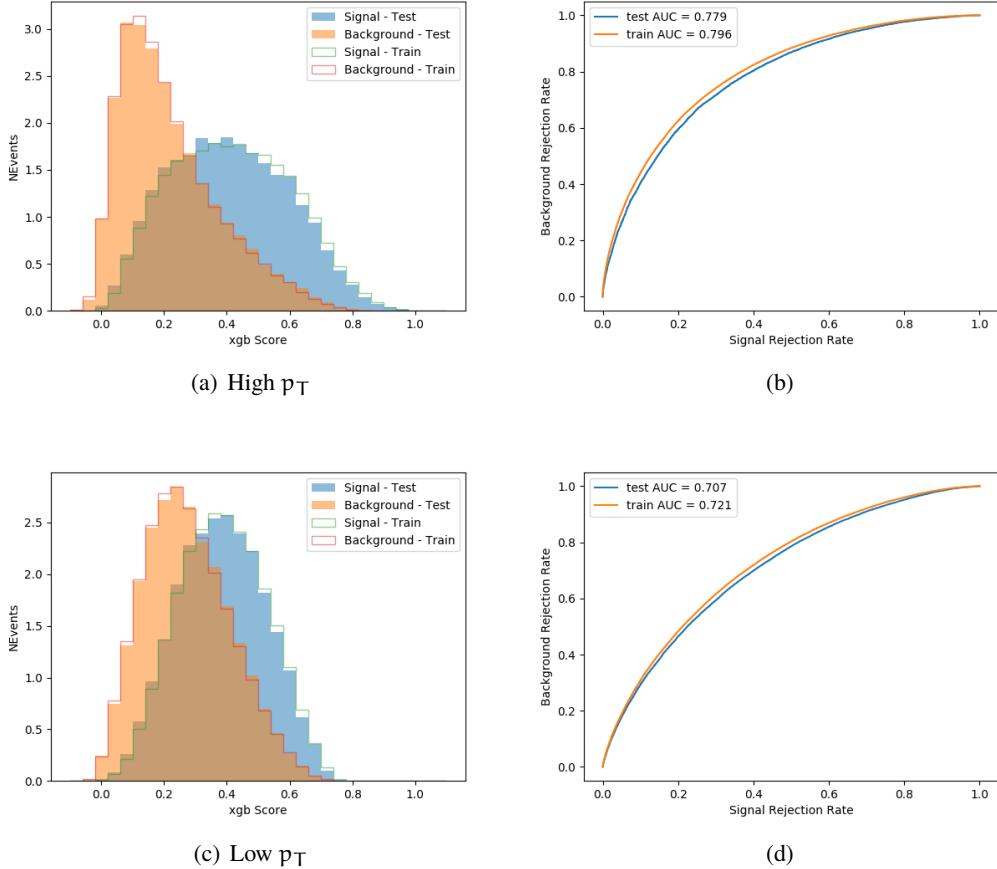


Figure 15.3: Output BDT scores of training and testing data for signal (blue) and background (orange) for 2lSS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 2lSS high and low p_T models, respectively.

3l - Semileptonic

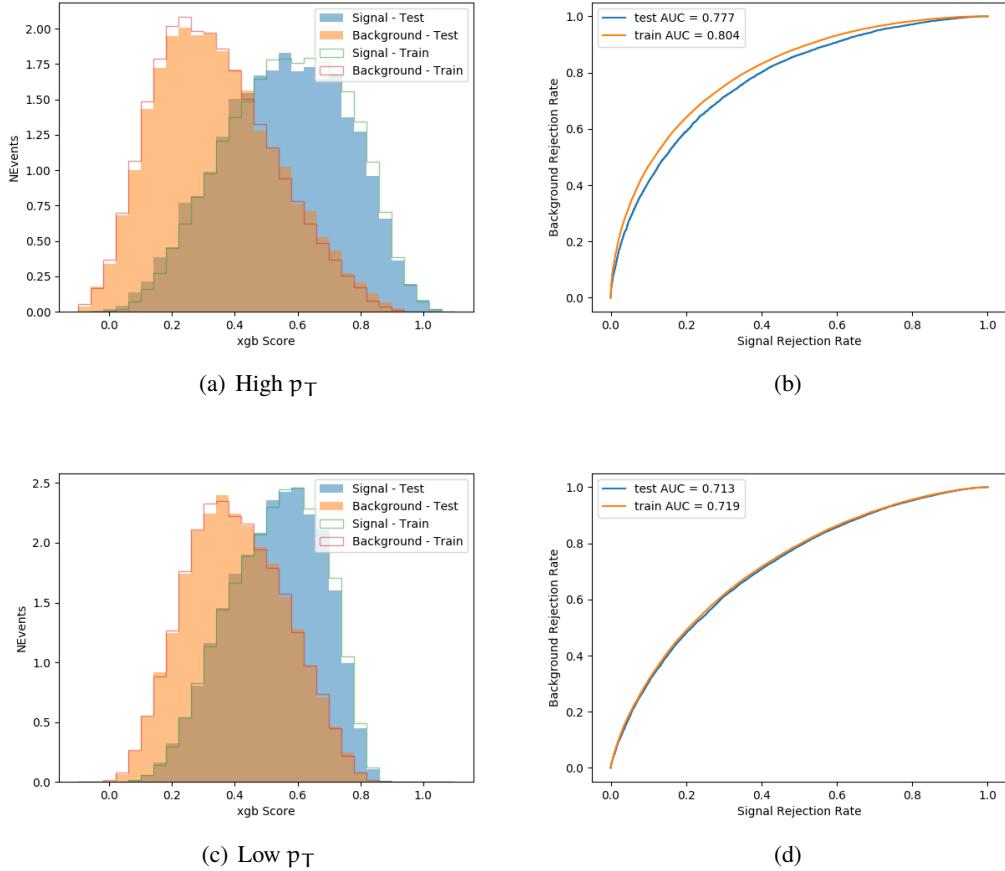


Figure 15.4: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lS high and low p_T models, respectively.

3l - Fully Leptonic

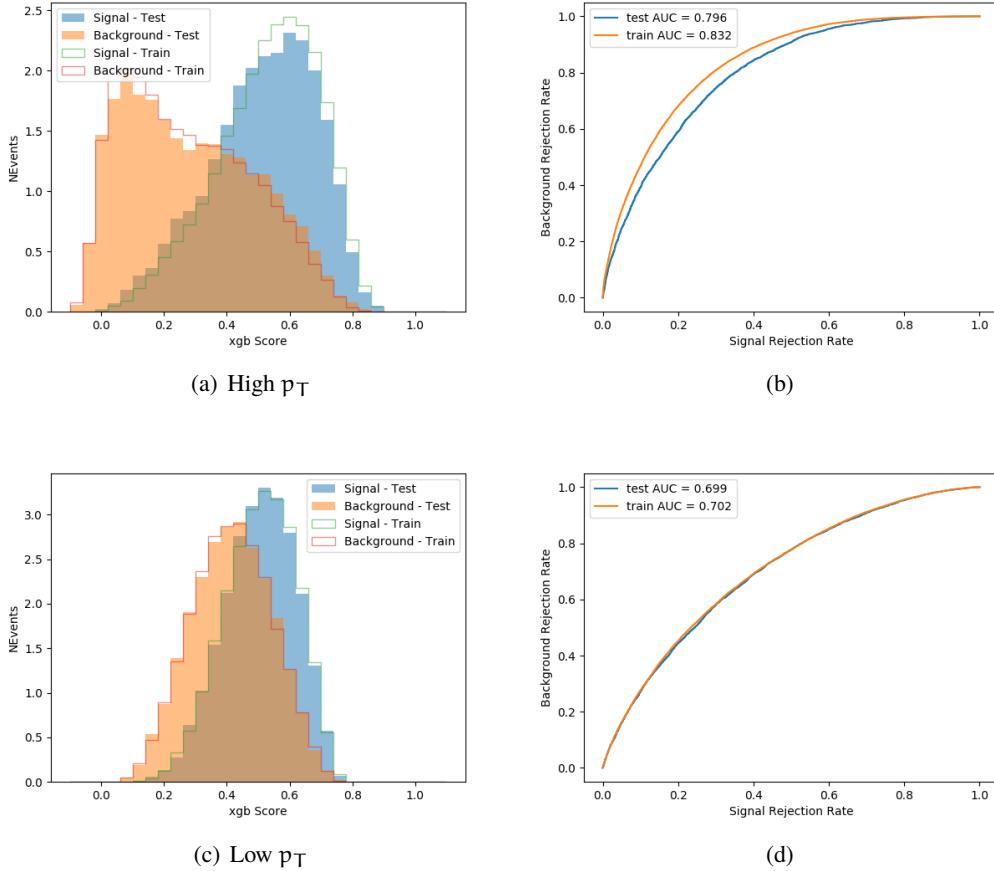


Figure 15.5: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lF events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lF high and low p_T models, respectively.

1258 Output distributions of each MVA, comparing MC predictions to data at 79.8 fb^{-1} are
 1259 shown in figures 15.6-15.2.

High p_T Background Rejection BDTs

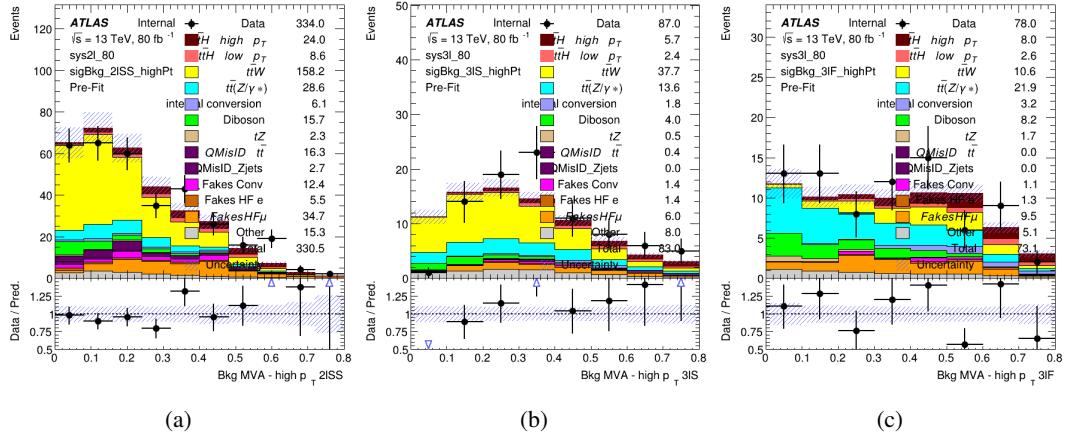


Figure 15.6: Output score of the high p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

Low p_T Background Rejection BDTs

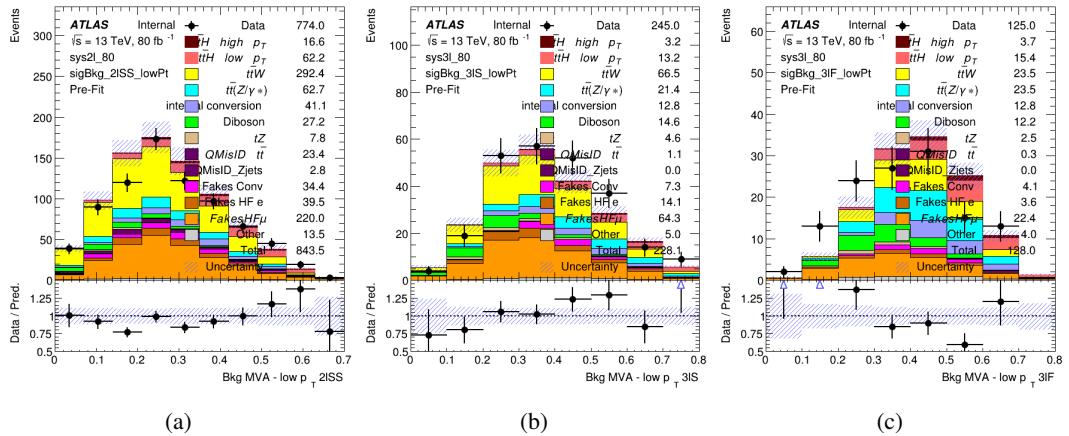


Figure 15.7: Output score of the low p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

1260 **15.3 Signal Region Definitions**

1261 Once pre-selection has been applied, channels are further refined based on the MVAs described
 1262 above. The output of the model described in Section 14.5 is used to separate the three channel
 1263 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs
 1264 boson. This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

1265 For each event, depending on the number of leptons as well as whether the p_T of the Higgs
 1266 is predicted to be high (> 150 GeV) or low (< 150 GeV), a cut on the appropriate background
 1267 rejection MVA is applied. The particular cut values, listed in Table 32, are determined by
 1268 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 32: Cutoff values on background rejection MVA score applied to signal regions.

1269 The event preselection and MVA selection listed in Table 32 are used define the three
 1270 signal regions used in the fit. These signal region definitions are summarized in Table 33.

Region	Selection
2lSS	Two same charge tight leptons with $p_T > 20 \text{ GeV}$ $N_{\text{jets}} \geq 4, N_{\text{b-jets}} \geq 1$ b-tagged jets Zero τ_{had} $H_{p_T}^{\text{pred}} > 150 \text{ GeV}$ and BDT score > 0.36 or $H_{p_T}^{\text{pred}} < 150 \text{ GeV}$ and BDT score > 0.34
3lS	Three light leptons with total charge ± 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 τ_{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 or $H_{p_T}^{\text{pred}} < 150 \text{ GeV}$ and BDT score > 0.43
3lF	Three light leptons with total charge ± 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 τ_{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 or $H_{p_T}^{\text{pred}} < 150 \text{ GeV}$ and BDT score > 0.41

Table 33: Selection applied to define the three signal regions used in the fit.

1271 16 Systematic Uncertainties

1272 The systematic uncertainties that are considered are summarized in Table 34. These are imple-
 1273 mented in the fit either as a normalization factors or as a shape variation or both in the signal
 1274 and background estimations. The numerical impact of each of these uncertainties is outlined in
 1275 section 17.

Table 34: 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
Physics Objects	
Electron	6
Muon	15
Jet energy scale and resolution	28
Jet vertex fraction	1
Jet flavor tagging	131
E_T^{miss}	3
Total (Experimental)	186
Background Modeling	
Cross section	24
Renormalization and factorization scales	10
Parton shower and hadronization model	2
Shower tune	4
Total (Signal and background modeling)	40
Background Modeling	
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

¹²⁷⁶ The uncertainty in the combined integrated luminosity is derived from a calibration of the
¹²⁷⁷ luminosity scale using x-y beam-separation scans performed for 13 TeV proton-proton data [15],
¹²⁷⁸ [37].

¹²⁷⁹ The experimental uncertainties are related to the reconstruction and identification of light
¹²⁸⁰ leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} .

1281 The sources which contribute to the uncertainty in the jet energy scale [47] are decomposed
 1282 into uncorrelated components and treated as independent sources in the analysis. This method
 1283 decomposes the uncertainties into 30 nuisance parameters included in the fit. A similar method
 1284 is used to account for jet energy resolution (JER) uncertainties, and 8 JER uncertainty components
 1285 are unclued as NPs in the fit.

1286 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
 1287 [39] are also decomposed into uncorrelated components. The large number of components for
 1288 b-tagging is due to the calibration of the distribution of the BDT discriminant.

1289 As mentioned in Section 12.2, a normalization corrections and uncertainties on the estim-
 1290 ates of non-prompt leptons backgrounds are derived using data driven techniques, decribed in
 1291 detail in [7]. These are derived from a likelihood fit over various non-prompt enriched control
 1292 regions, targeting several sources of non-prompt light leptons separately: external conversion
 1293 electrons, internal conversion electrons, electrons from heavy flavor decays, and muons from
 1294 heavy flavor decays.

1295 The normalization factor and uncertainty applied to each source of non-prompt leptons is
 1296 summarized in Table 16

Processss	Normalization Factor
NF_e^{ExtCO}	1.70 ± 0.51
NF_e^{IntCO}	0.75 ± 0.26
NF_e^{HF}	1.09 ± 0.32
NF_{μ}^{HF}	1.28 ± 0.17

Table 35: Normalization factors - with statistical and systematic uncertainties - derived from the fit over fake control regions for each source of non-prompt leptons considered.

1297 In addition to those derived from the control regions, several additional uncertainties are
1298 assigned to the non-prompt lepton background. An additional 25% uncertainty on material
1299 conversions is assigned, based on the comparison between data and MC in a region where a
1300 loose electron fails the photon conversion veto. A shape uncertainty of 15% (6%) is assigned to
1301 the HF non-prompt electron (muon) background based on a comparison between data and MC
1302 where the second leading electron (muon) is only required to be loose. As the contribution from
1303 light non-prompt leptons is small, about 10% percent of the contribution from HF non-prompt
1304 leptons, it is derived from the agreement between data and simulation in a LF enriched region at
1305 low values of the non-prompt lepton BDT. The resulting uncertainty is 100%, and is taken to be
1306 uncorrelated between internal and material conversions.

1307 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and
1308 scale uncertainties are taken from theory calculations for the predominate prompt backgrounds.
1309 Following the nominal $t\bar{t}H - ML$ analysis, a 50% uncertainty is applied to Diboson to account
1310 for the large uncertainty in estimating $VV +$ heavy flavor. The other “rare” background processes
1311 - including tZ , rare top processes, $ttWW$, WtZ , VVV , $tHjb$ and WtH - are assigned an overall
1312 50% normalization uncertainty as well. The theory uncertainties applied to the MC estimates
1313 are summarized in Table 36.

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

Table 36: Summary of theoretical uncertainties for MC predictions in the analysis.

1314 Additional uncertainties to account for t̄W mismodelling are also applied. These include
 1315 a “Generator” uncertainty, based on a comparison between the nominal Sherpa 2.2.5 sample,
 1316 and the formerly used aMC@NLO sample, and an “Extra radiation” uncertainty, which includes
 1317 renormalisation and factorisation scale variations of the Sherpa 2.2.5 sample.

1318 17 Results

1319 A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p_T spectrum
 1320 in the three signal regions, 2lSS, 3lS, and 3lF. The signal is split into high and low p_T samples,
 1321 based on whether the truth p_T of the Higgs is above or below 150 GeV. The parameters $\mu_{t\bar{t}H \text{ high } p_T}$
 1322 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
 1323 between the observed value and the theory prediction. Unblinded results are shown for the 79.8
 1324 fb⁻¹ data set, as well as MC only projections of results using the full Run-2, 139 fb⁻¹ dataset.

1325 As described in Section 16, there are 229 systematic uncertainties that are considered
 1326 as NPs in the fit. These NP s are constrained by Gaussian or log-normal probability density
 1327 functions. The latter are used for normalisation factors to ensure that they are always positive.
 1328 The expected number of signal and background events are functions of the likelihood. The prior
 1329 for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its
 1330 nominal value.

1331 17.1 Results - 79.8 fb^{-1}

1332 As the data collected from 2015-2017 has been unblinded for $t\bar{t}H$ – ML channels, represent-
 1333 ing 79.8 fb^{-1} , those events are unblinded. The predicted Higgs p_T spectrum is fit to data
 1334 simultaneously in each of the three signal regions shown in Figure 17.1.

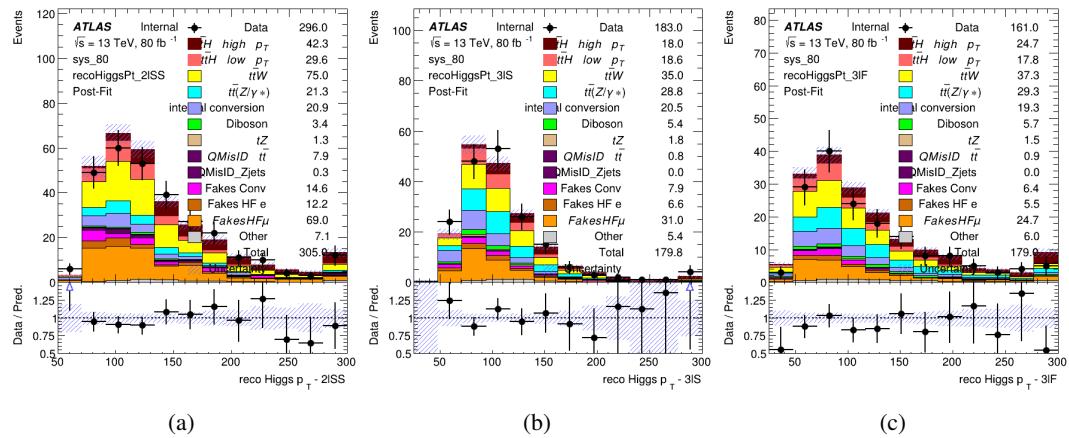


Figure 17.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 79.8 fb^{-1} of MC

1335 A post-fit summary of the fitted regions is shown in Figure 17.2.

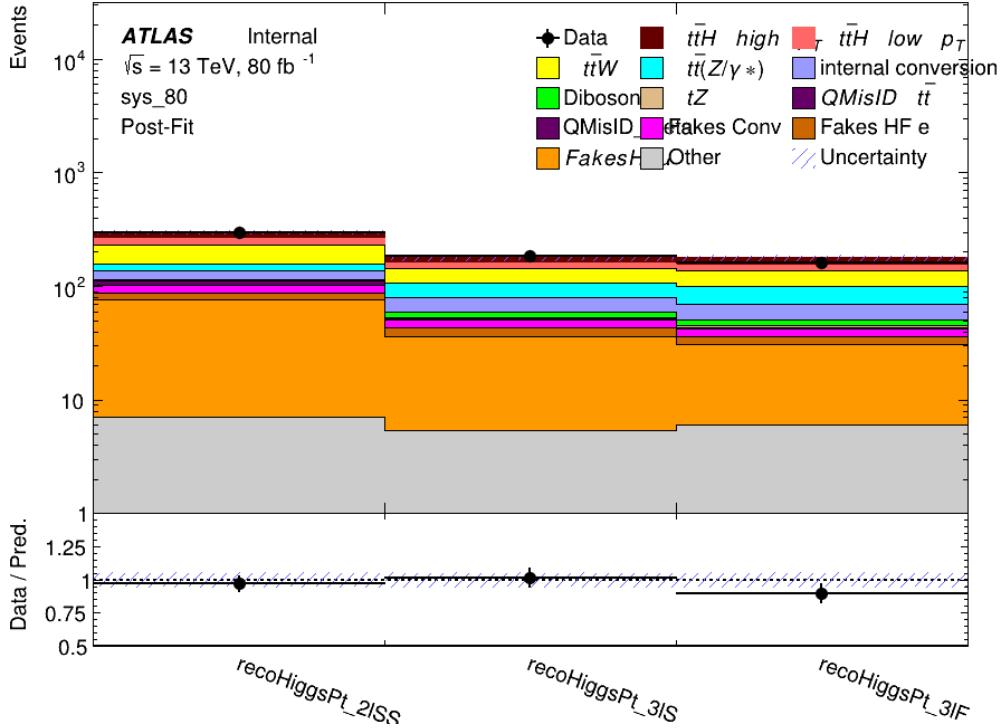


Figure 17.2: Post-fit summary of the yields in each signal region.

1336 The measured μ values for high and low p_T Higgs production obtained from the fit
 1337 are shown in 37. A significance of 1.7σ is observed for $t\bar{t}H$ high p_T , and 2.1σ is measured for
 1338 $t\bar{t}H$ 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 37: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1339 The most prominent sources of systematic uncertainty, as measured by their impact on
 1340 $\mu_{t\bar{t}H \text{ high } p_T}$, are summarized in Table 38.

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

Table 38: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H high p_T.

¹³⁴¹ The most significant sources of uncertainty on the measurement of t̄H - low p_T are shown
¹³⁴² in Table 39.

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 39: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H low p_T.

¹³⁴³ The ranking and impact of those nuisance parameters with the largest contribution to the

1344 overall uncertainty is shown in Figure 17.3.

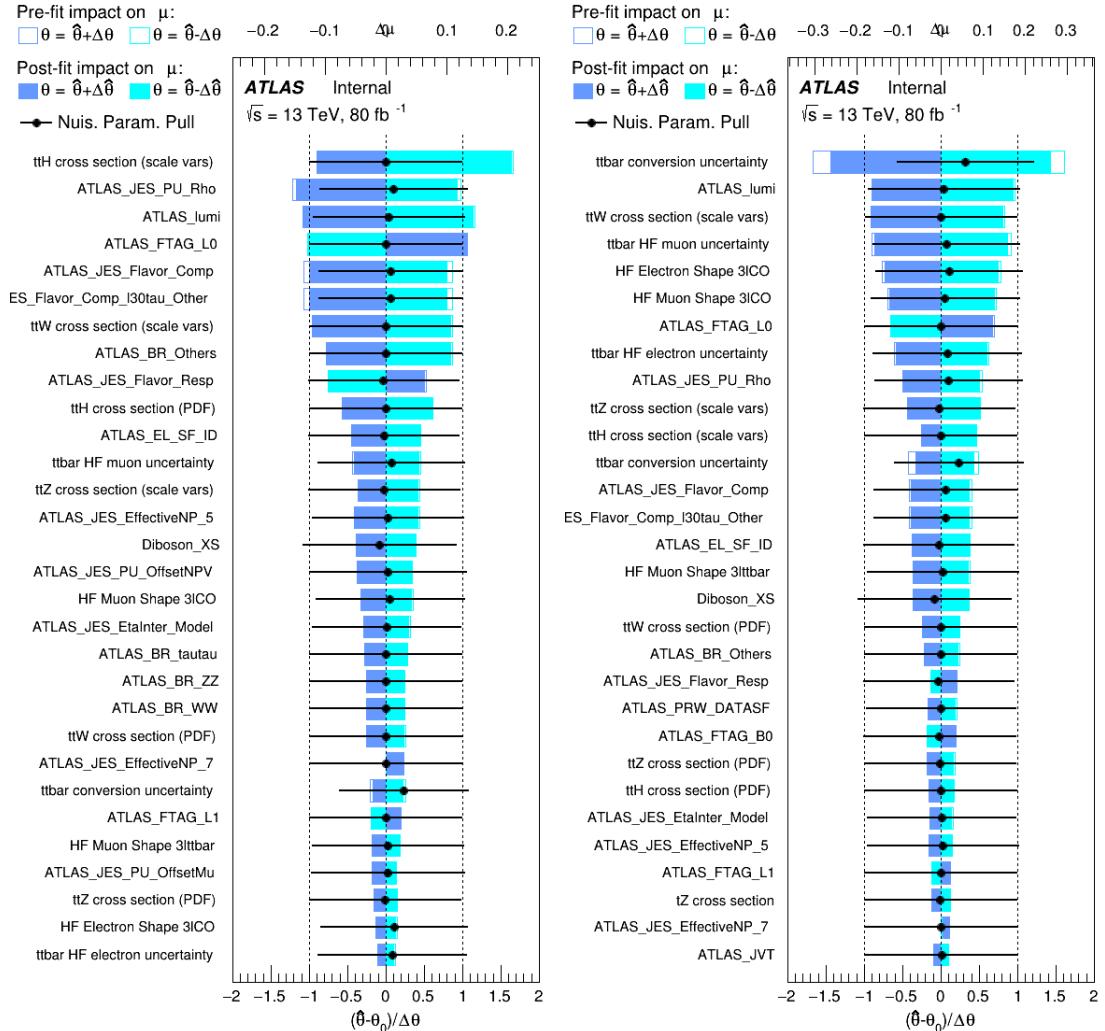


Figure 17.3: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

1345

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

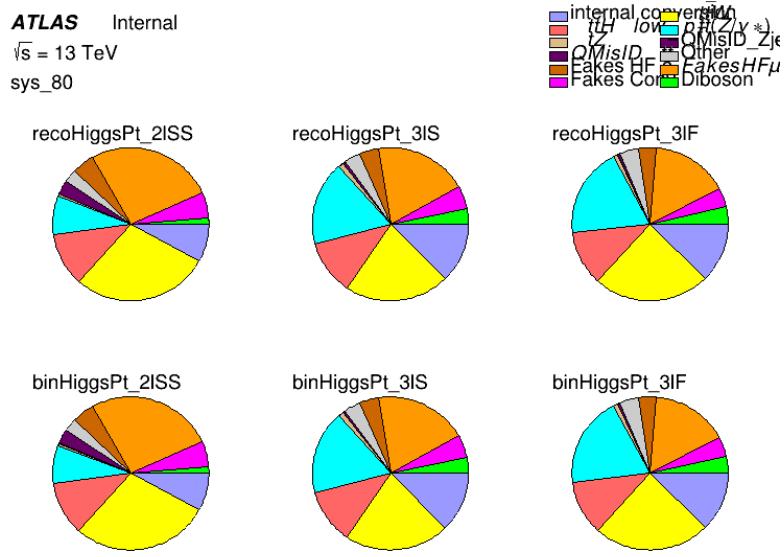


Figure 17.4: Background composition of the fit regions.

1346 17.2 Projected Results - 139 fb^{-1}

1347 As data collected in 2018 has not yet been unblinded for $t\bar{t}H$ – ML at the time of this note, data
 1348 from that year remains blinded. Instead, an Asimov fit is performed - with the MC prediction
 1349 being used both as the SM prediction as well as the data in the fit - in order to give expected
 1350 results.

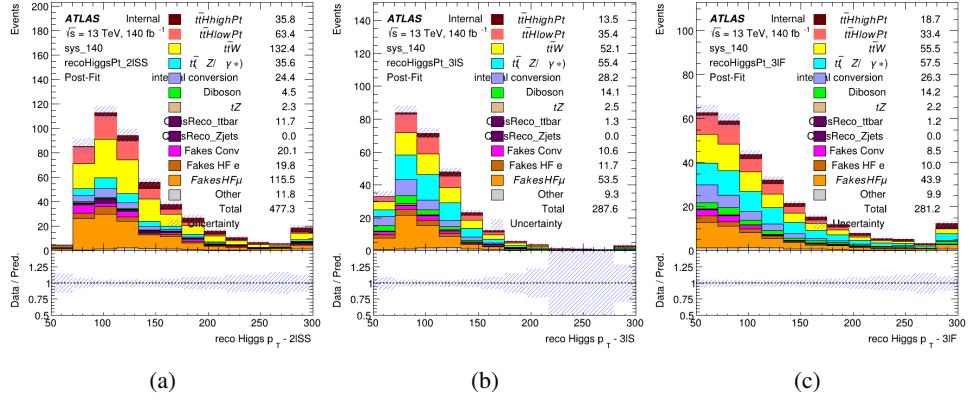


Figure 17.5: Blinded post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 139 fb^{-1} of data

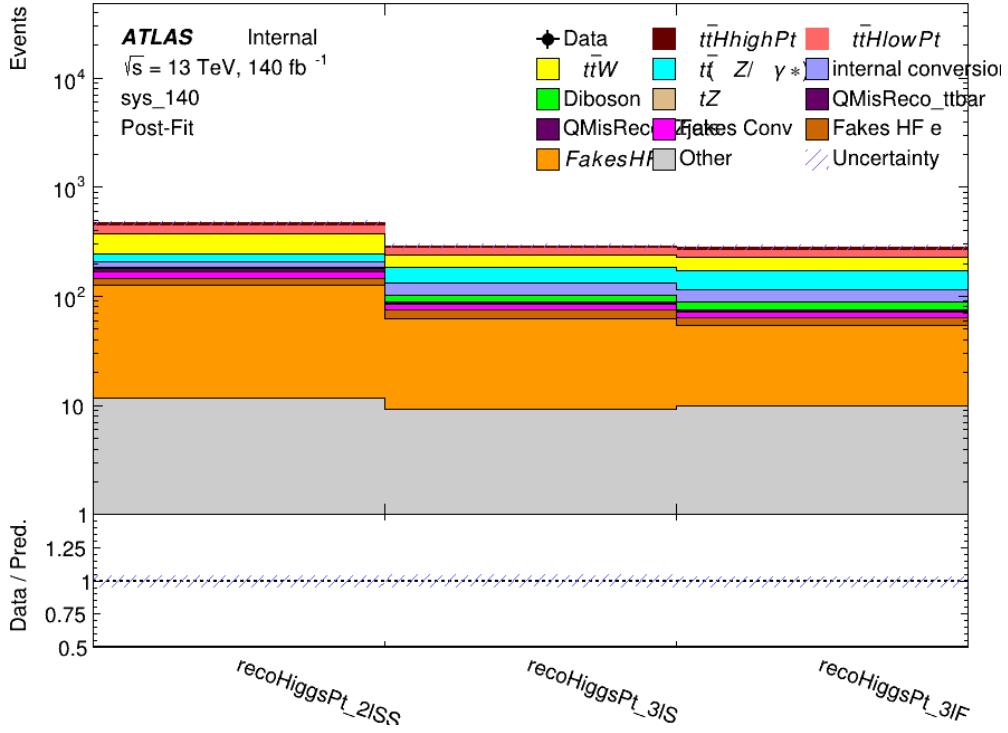


Figure 17.6: Post-fit summary of fit.

1352 shown in 40. A significance of 2.0σ is expected for $t\bar{t}H$ high p_T , and a projected significance

1353 2.3σ 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 40: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1354 The most prominent sources of systematic uncertainty, as measured by their impact on

1355 $\mu_{t\bar{t}H \text{ high } p_T}$, are summarized in Table 41.

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

1356 The most significant sources of systematic uncertainty on $t\bar{t}H$ low p_T are summarized in

1357 Table 42.

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
t̄W 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
t̄Z Cross-section (QCD Scale)	-0.07	0.09
t̄H Cross-section (QCD Scale)	-0.05	0.1
Total	0.48	0.50

Table 42: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H low p_T.

¹³⁵⁸ The ranking and impact of those nuisance parameters with the largest contribution to the
¹³⁵⁹ overall uncertainty is shown in Figure 17.7.

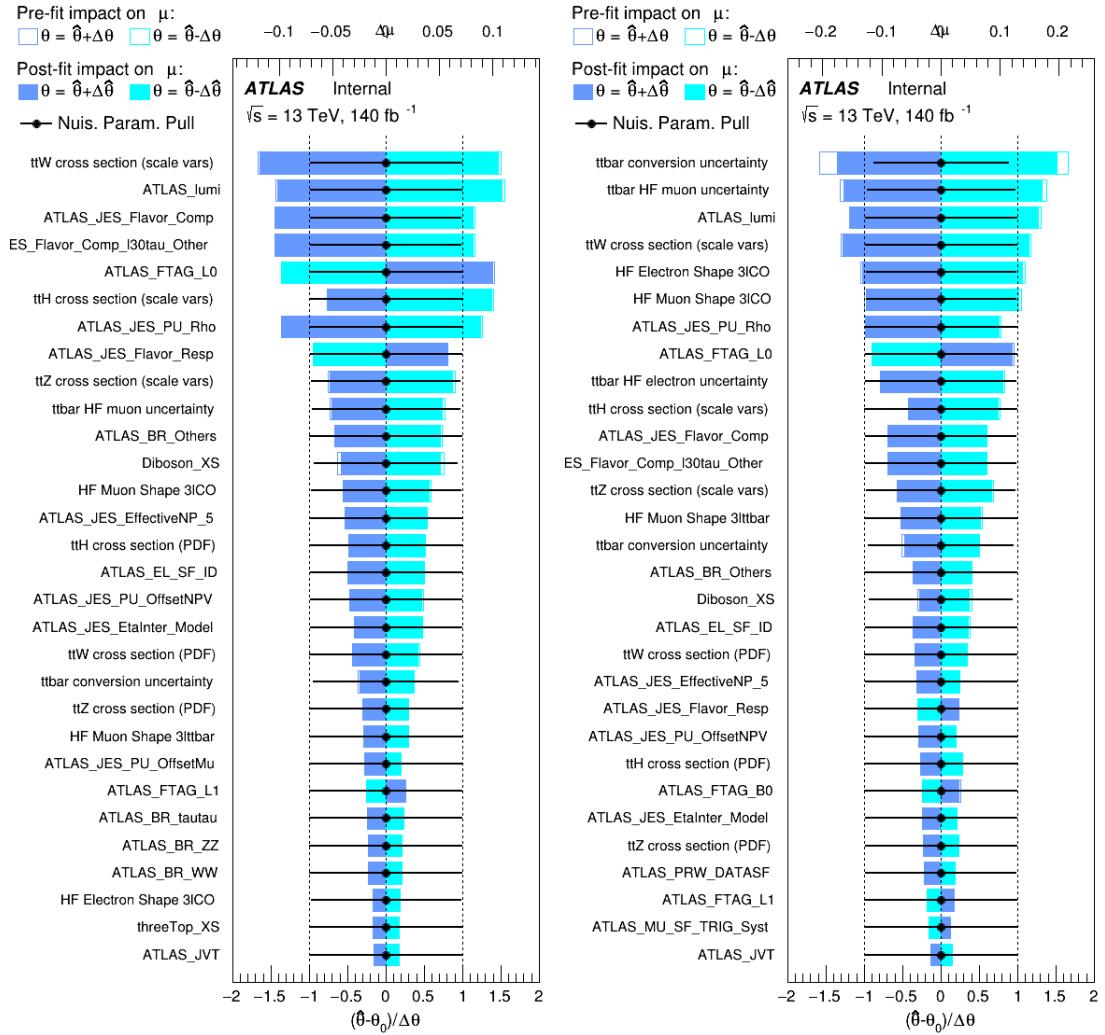


Figure 17.7: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

1360

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

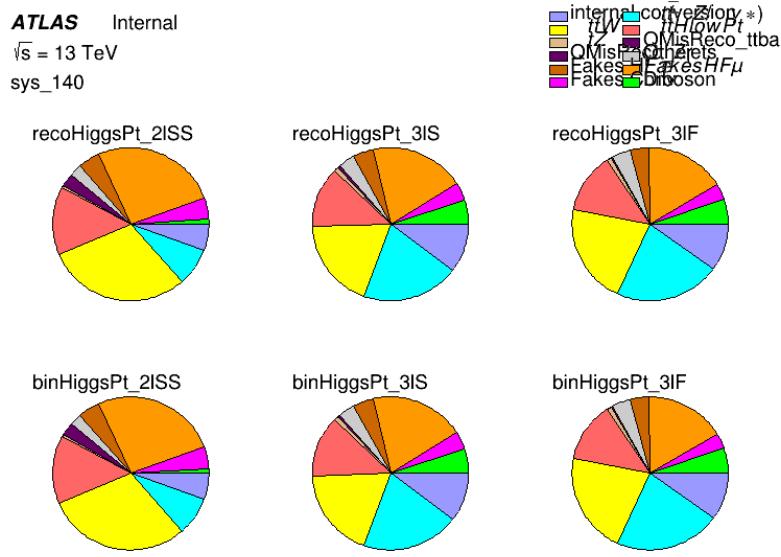


Figure 17.8: Background composition of the fit regions.

1361

Part VI

1362

Conclusion

1363 As search for the effects of dimension-six operators on $t\bar{t}H$ production is performed. An effective
 1364 field theory approached is used to parameterize the effects of high energy physics on the Higgs
 1365 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,
 1366 and the limits on dimension-six operators are limited to X.

References

- 1367 [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*
1368 *Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,
1369 arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- 1370 [2] R. Oerter, *The theory of almost everything : the Standard Model, the unsung triumph of*
1371 *modern physics*, Pi Press, 2006, ISBN: 978-0-13-236678-6.
- 1372 [3] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,
1373 Phys. Rev. Lett. 13 (Oct, 1964) 508-509 ().
- 1374 [4] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*,
1375 Phys. Rev. **127** (3 1962) 965,
1376 URL: <https://link.aps.org/doi/10.1103/PhysRev.127.965>.
- 1377 [5] J. Ellis, *Higgs Physics*,
1378 (2013) 117, 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of
1379 High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report,
1380 URL: <https://cds.cern.ch/record/1638469>.
- 1381 [6] T. A. Collaboration, *Observation of Higgs boson production in association with a top*
1382 *quark pair at the LHC with the ATLAS detector*, Physics Letters B **784** (2018) 173,
1383 ISSN: 0370-2693, URL:
1384 <https://www.sciencedirect.com/science/article/pii/S0370269318305732>.

- 1386 [7] *Evidence for the associated production of the Higgs boson and a top quark pair with the*
1387 *ATLAS detector*, tech. rep. ATLAS-CONF-2017-077, CERN, 2017,
1388 URL: <https://cds.cern.ch/record/2291405>.
- 1389 [8] A. M. Sirunyan et al., *Observation of $t\bar{t}H$ Production*,
1390 *Physical Review Letters* **120** (2018), ISSN: 1079-7114,
1391 URL: <http://dx.doi.org/10.1103/PhysRevLett.120.231801>.
- 1392 [9] B. Dumont, S. Fichet and G. von Gersdorff,
1393 *A Bayesian view of the Higgs sector with higher dimensional operators*,
1394 *Journal of High Energy Physics* **2013** (2013), ISSN: 1029-8479,
1395 URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)065](http://dx.doi.org/10.1007/JHEP07(2013)065).
- 1396 [10] S. Banerjee, S. Mukhopadhyay and B. Mukhopadhyaya,
1397 *Higher dimensional operators and the LHC Higgs data: The role of modified kinematics*,
1398 *Physical Review D* **89** (2014), ISSN: 1550-2368,
1399 URL: <http://dx.doi.org/10.1103/PhysRevD.89.053010>.
- 1400 [11] M. Zinser, ‘The Large Hadron Collider’, *Search for New Heavy Charged Bosons and*
1401 *Measurement of High-Mass Drell-Yan Production in Proton—Proton Collisions*,
1402 Springer International Publishing, 2018 47, ISBN: 978-3-030-00650-1,
1403 URL: https://doi.org/10.1007/978-3-030-00650-1_4.
- 1404 [12] *Detector and Technology*, URL: <https://atlas.cern/discover/detector>.
- 1405 [13] M. Marjanovic, *ATLAS Tile calorimeter calibration and monitoring systems*, 2018.

- 1406 [14] ATLAS Collaboration, *Observation of electroweak $W^\pm Z$ boson pair production in*
 1407 *association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,*
 1408 [Phys. Lett. B793 \(2019\) 469](#), arXiv: [1812.09740 \[hep-ex\]](#).
- 1409 [15] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using*
 1410 *the ATLAS detector at the LHC*, [Eur. Phys. J. C 71 \(2011\) 1630](#),
 1411 arXiv: [1101.2185 \[hep-ex\]](#).
- 1412 [16] ATLAS Collaboration, *Performance of the ATLAS detector using first collision data,*
 1413 [JHEP 09 \(2010\) 056](#), arXiv: [1005.5254 \[hep-ex\]](#).
- 1414 [17] S. Agostinelli et al., *GEANT4: A Simulation toolkit*,
 1415 [Nucl. Instrum. Meth. A506 \(2003\) 250](#).
- 1416 [18] T. Gleisberg et al., *Event generation with SHERPA 1.1*, [JHEP 02 \(2009\) 007](#),
 1417 arXiv: [0811.4622 \[hep-ph\]](#).
- 1418 [19] H.-L. Lai et al., *New parton distributions for collider physics*,
 1419 [Phys. Rev. D 82 \(2010\) 074024](#), arXiv: [1007.2241 \[hep-ph\]](#).
- 1420 [20] J. Alwall et al., *The automated computation of tree-level and next-to-leading order*
 1421 *differential cross sections, and their matching to parton shower simulations*,
 1422 [JHEP 07 \(2014\) 079](#), arXiv: [1405.0301 \[hep-ph\]](#).
- 1423 [21] R. D. Ball et al., *Parton distributions for the LHC Run II*, [JHEP 04 \(2015\) 040](#),
 1424 arXiv: [1410.8849 \[hep-ph\]](#).
- 1425 [22] M. Bahr et al., *Herwig++ Physics and Manual*, [Eur. Phys. J. C 58 \(2008\) 639](#),
 1426 arXiv: [0803.0883 \[hep-ph\]](#).

- 1427 [23] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244,
 1428 arXiv: [1207.1303 \[hep-ph\]](https://arxiv.org/abs/1207.1303).
- 1429 [24] S. Frixione, G. Ridolfi and P. Nason,
 1430 *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*,
 1431 JHEP **09** (2007) 126, arXiv: [0707.3088 \[hep-ph\]](https://arxiv.org/abs/0707.3088).
- 1432 [25] E. Re, *Single-top Wt-channel production matched with parton showers using the*
 1433 *POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: [1009.2450 \[hep-ph\]](https://arxiv.org/abs/1009.2450).
- 1434 [26] A. Collaboration, *Electron reconstruction and identification in the ATLAS experiment*
 1435 *using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13 \text{ TeV}$,*
 1436 The European Physical Journal C **79** (2019), ISSN: 1434-6052,
 1437 URL: <http://dx.doi.org/10.1140/epjc/s10052-019-7140-6>.
- 1438 [27] ATLAS Collaboration,
 1439 *Electron reconstruction and identification efficiency measurements with the ATLAS*
 1440 *detector using the 2011 LHC proton-proton collision data*,
 1441 Eur. Phys. J. C **74** (2014) 2941, arXiv: [1404.2240 \[hep-ex\]](https://arxiv.org/abs/1404.2240).
- 1442 [28] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the*
 1443 *ATLAS detector using 2011 and 2012 LHC proton–proton collision data*,
 1444 Eur. Phys. J. C **74** (2014) 3130, arXiv: [1407.3935 \[hep-ex\]](https://arxiv.org/abs/1407.3935).
- 1445 [29] ATLAS Collaboration, *Observation of the Higgs boson production in association with a*
 1446 *$t\bar{t}$ pair with the ATLAS detector*, PoS EPS-HEP**2019** (2020) 325.

- 1447 [30] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_T jet clustering algorithm*,
1448 *Journal of High Energy Physics* **2008** (2008) 063, issn: 1029-8479,
1449 URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- 1450 [31] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp*
1451 *collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 581,
1452 arXiv: [1510.03823 \[hep-ex\]](https://arxiv.org/abs/1510.03823).
- 1453 [32] ATLAS Collaboration, *Measurements of b -jet tagging efficiency with the ATLAS detector*
1454 *using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, *Journal of High Energy Physics* **2018** (2018),
1455 issn: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP08\(2018\)089](http://dx.doi.org/10.1007/JHEP08(2018)089).
- 1456 [33] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
1457 *with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*,
1458 *The European Physical Journal C* **78** (2018), issn: 1434-6052,
1459 URL: <http://dx.doi.org/10.1140/epjc/s10052-018-6288-9>.
- 1460 [34] T. Chen and C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’,
1461 *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*
1462 *Discovery and Data Mining*, KDD ’16, ACM, 2016 785, isbn: 978-1-4503-4232-2,
1463 URL: <http://doi.acm.org/10.1145/2939672.2939785>.
- 1464 [35] ATLAS Collaboration, *Measurement of the fiducial and differential cross-section of a*
1465 *top quark pair in association with a Z boson at 13 TeV with the ATLAS detector*, (2019),
1466 URL: <https://cds.cern.ch/record/2672207>.

- 1467 [36] *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at*
1468 *the LHC*, (2019).
- 1469 [37] ATLAS Collaboration,
1470 *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*,
1471 [JINST **13** \(2018\) P07017](#).
- 1472 [38] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties*
1473 *in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
1474 [Phys. Rev. D **96** \(2017\) 072002](#), arXiv: [1703.09665 \[hep-ex\]](#).
- 1475 [39] A. Collaboration, *Performance of b -jet identification in the ATLAS experiment*,
1476 [Journal of Instrumentation **11** \(2016\) P04008](#),
1477 URL: <http://stacks.iop.org/1748-0221/11/i=04/a=P04008>.
- 1478 [40] ATLAS Collaboration, *Observation of the associated production of a top quark and a Z*
1479 *boson in pp collisions at $\sqrt{s}= 13$ TeV with the ATLAS detector*,
1480 [Journal of High Energy Physics **2020** \(2020\)](#), ISSN: 1029-8479,
1481 URL: [http://dx.doi.org/10.1007/JHEP07\(2020\)124](http://dx.doi.org/10.1007/JHEP07(2020)124).
- 1482 [41] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions from Powheg*
1483 *and Sherpa to ATLAS measurements of top pair production at 7 TeV*,
1484 ATL-PHYS-PUB-2015-011, 2015, URL: <https://cds.cern.ch/record/2020602>.
- 1485 [42] ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11*,
1486 ATL-PHYS-PUB-2011-009, 2011, URL: <https://cds.cern.ch/record/1363300>.

- 1487 [43] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using*
1488 *the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016,
1489 URL: <https://cds.cern.ch/record/2157687>.
- 1490 [44] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets*
1491 *Reconstructed in the ATLAS Detector at $\sqrt{s} = 13 \text{ TeV}$* , ATL-PHYS-PUB-2015-015,
1492 2015, URL: <https://cds.cern.ch/record/2037613>.
- 1493 [45] ATLAS Collaboration,
1494 *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector*,
1495 ATLAS-CONF-2015-029, 2015, URL: <https://cds.cern.ch/record/2037702>.
- 1496 [46] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
1497 *with the ATLAS detector in the first proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$* ,
1498 ATL-PHYS-PUB-2015-027, 2015, URL: <https://cds.cern.ch/record/2037904>.
- 1499 [47] G. Aad, T. Abajyan, B. Abbott and et al, *Jet energy resolution in proton-proton collisions*
1500 *at $\sqrt{s} = 7 \text{ TeV}$ recorded in 2010 with the ATLAS detector*,
1501 *The European Physical Journal C* **73** (2013) 2306, ISSN: 1434-6052,
1502 URL: <https://doi.org/10.1140/epjc/s10052-013-2306-0>.
- 1503 [48] R. Narayan et al., *Measurement of the total and differential cross sections of a*
1504 *top-quark-antiquark pair in association with a W boson in proton-proton collisions at a*
1505 *centre-of-mass energy of 13 TeV with ATLAS detector at the Large Hadron Collider*,
1506 tech. rep. ATL-COM-PHYS-2020-217, CERN, 2020,
1507 URL: <https://cds.cern.ch/record/2712986>.

1508 [49] R. Frederix et al., *The automation of next-to-leading order electroweak calculations*,

1509 [Journal of High Energy Physics 2018 \(2018\)](#), ISSN: 1029-8479,

1510 URL: [http://dx.doi.org/10.1007/JHEP07\(2018\)185](http://dx.doi.org/10.1007/JHEP07(2018)185).

¹⁵¹¹ **List of contributions**

¹⁵¹²

₁₅₁₃ **Part VII**

₁₅₁₄ **Appendices**

₁₅₁₅ **A Non-prompt lepton MVA**

1516 A lepton MVA has been developed to better reject non-prompt leptons than standard cut
1517 based selections based upon impact parameter, isolation and PID. The name of this MVA is
1518 **PromptLeptonVeto**. The full set of studies and detailed explanation can be found in [48].

1519 The decays of W and Z bosons are commonly selected by the identification of one or two
1520 electrons or muons. The negligible lifetimes of these bosons mean that the leptons produced in the
1521 decay originate from the interaction vertex and are thus labelled “prompt”. Analyses using these
1522 light leptons impose strict reconstruction quality, isolation and impact parameter requirements
1523 to remove “fake” leptons. A significant source of the fake light leptons are non-prompt leptons
1524 produced in decays of hadrons that contain bottom (b) or charm (c) quarks. Such hadrons
1525 typically have microscopically significant lifetimes that can be detected experimentally.

1526 These non-prompt leptons can also pass the tight selection criteria. In analyses that
1527 involve top (t) quarks, which decay almost exclusively into a W boson and a b quark, non-
1528 prompt leptons from the semileptonic decay of bottom and charm hadrons can be a significant
1529 source of background events. This is particularly the case in the selection of same-sign dilepton
1530 and multilepton final states.

1531 The main idea is to identify non-prompt light leptons using lifetime information associated
1532 with a track jet that matches the selected light lepton. This lifetime information is computed
1533 using tracks contained within the jet. Typically, lepton lifetime is determined using the impact
1534 parameter of the track reconstructed by the inner tracking detector which is matched to the recon-
1535 structed lepton. Using additional reconstructed charged particle tracks increases the precision of

1536 identifying the displaced decay vertex of bottom or charm hadrons that produced a non-prompt
 1537 light lepton. The MVA also includes information related to the isolation of the lepton to reject
 1538 non-prompt leptons.

1539 **PromptLeptonVeto** is a gradient boosted BDT. The training of the BDT is performed on
 1540 leptons selected from the PowHEG+PYTHIA6 non-allhad $t\bar{t}$ MC sample. Eight variables are used
 1541 to train the BDT in order to discriminate between prompt and non-prompt leptons. The track
 1542 jets that are matched to the non-prompt leptons correspond to jets initiated by b or c quarks, and
 1543 may contain a displaced vertex. Consequently, three of the selected variables are used to identify
 1544 b-tag jets by standard ATLAS flavour tagging algorithms. Two variables use the relationship
 1545 between the track jet and lepton: the ratio of the lepton p_T with respect to the track jet p_T and
 1546 ΔR between the lepton and the track jet axis. Finally three additional variables test whether the
 1547 reconstructed lepton is isolated: the number of tracks collected by the track jet and the lepton
 1548 track and calorimeter isolation variables. Table 43 describes the variables used to train the BDT
 1549 algorithm. The choice of input variables has been extensively discussed with Egamma, Muon,
 1550 Tracking, and Flavour Tagging CP groups.

Variable	Description
N_{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
$\Delta R(\text{lepton}, \text{track jet})$	ΔR 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 43: A table of the variables used in the training of **PromptLeptonVeto**.

1551 The output distribution of the BDT is shown in Figure A.

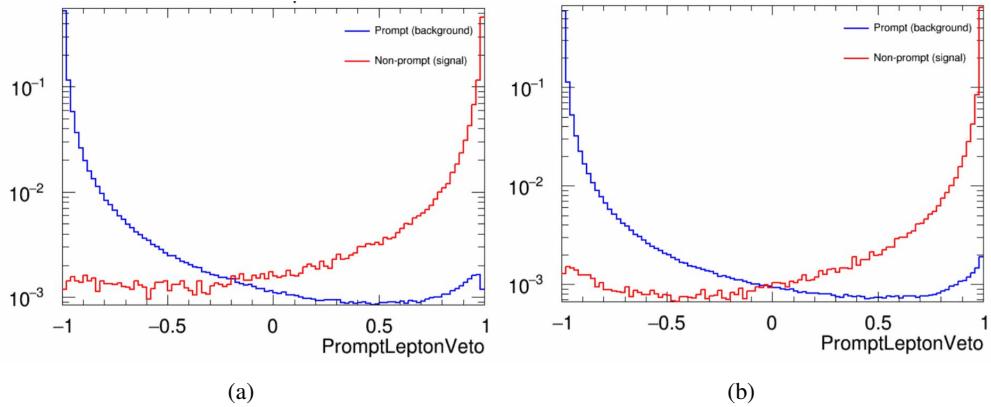


Figure A.1: Distribution of the PLV BDT discriminant for (a) electrons and (b) muons

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

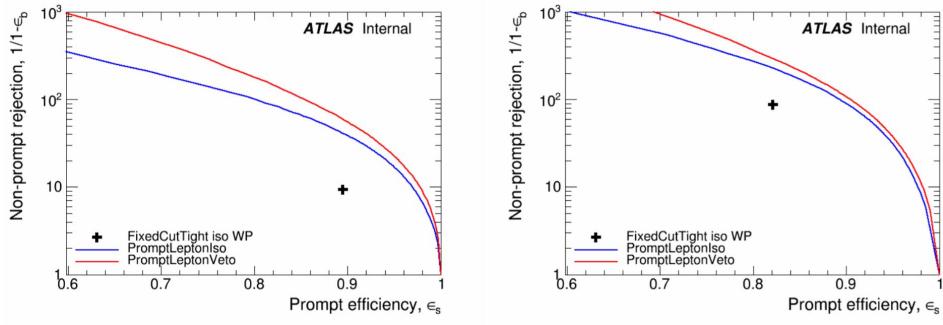


Figure A.2: ROC curves for the PLV as well as the performance of the standard FixedCutTight WP for (left) electrons and (right) muons

1554 A cutoff value of -0.7 for electrons and -0.5 for muons are chosen as the WPs for this
 1555 MVA, based on an optimisation of S/\sqrt{B} performed in the preselection regions of the $t\bar{t}H - ML$
 1556 analysis, which have a signature similar to that of this analysis.

1557 The efficiency of the tight PromptLeptonVeto working point is measured using the tag
1558 and probe method with $Z \rightarrow \ell^+ \ell^-$ events. Such calibration are performed by analysers from
1559 this analysis in communication with the Egamma and Muon combined performance groups. The
1560 scale factor are approximately 0.92 for $10 < p_T < 15$ GeV, and averaging at 0.98 to 0.99 for
1561 higher p_T leptons. An extra systematic is applied to muons within $\Delta R < 0.6$ of a calorimeter
1562 jet, since there is a strong dependence on the scale factor due to the presence of these jets. For
1563 electrons, the dominant systematics is coming from pile-up dependence. Overall the systematics
1564 are a maximum of 3% at low p_T and decreasing at a function of p_T .

¹⁵⁶⁵ **B Supplementary WZ + Heavy Flavor Studies**

¹⁵⁶⁶ **B.1 Non-prompt CR Modelling**

1567 In order to further validate the modeling in each of the non-prompt CRs, additional
 1568 kinematic plots are made in the Z+jets CR and $t\bar{t}$ CR in each of the continuous b-tag regions,
 1569 after the correction factors detailed in Section 8.3 have been applied.

1570 In the case of the Z+jets CR, the p_T spectrum of the lepton originating from the W
 1571 candidate is shown, as this is the distribution used to extract the scale factor applied to Z+jets.

1572 These plots are shown in Figures B.1 and B.2.

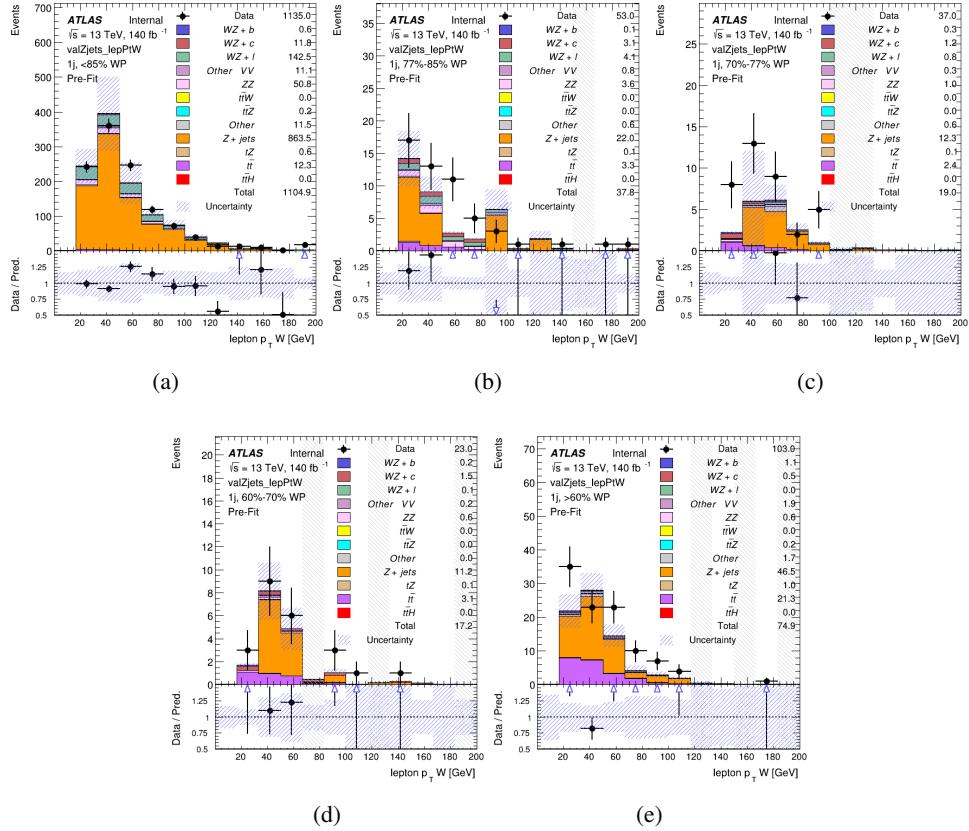


Figure B.1: Comparisons between the data and MC distributions of the p_T of the lepton originating from the W-candidate in the Z+jets CR for each of the 1-jet b-tag working point regions

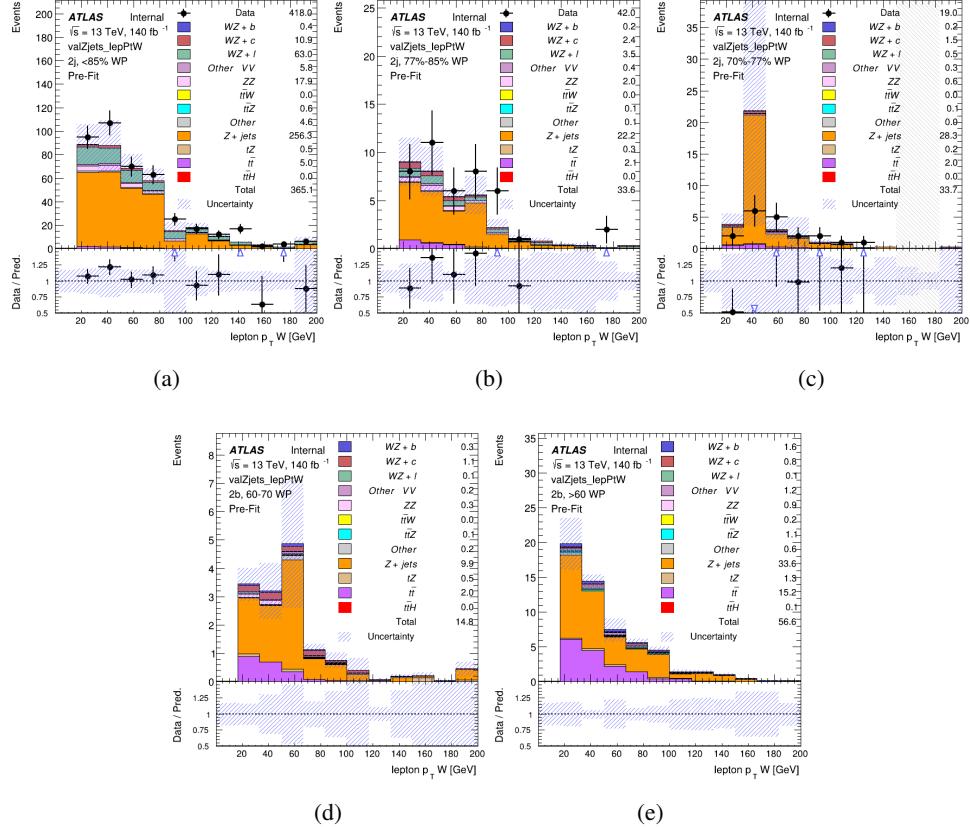


Figure B.2: Comparisons between the data and MC distributions of the p_T of the lepton originating from the W-candidate in the Z+jets CR for each of the 2-jet b-tag working point regions

1573 The same is shown for the $t\bar{t}$ CR, but the p_T of the OS lepton is used instead as a
 1574 representation of the modeling, as the lepton from the W is not well defined for $t\bar{t}$ events. These
 1575 plots are shown in Figures B.3 and B.4.

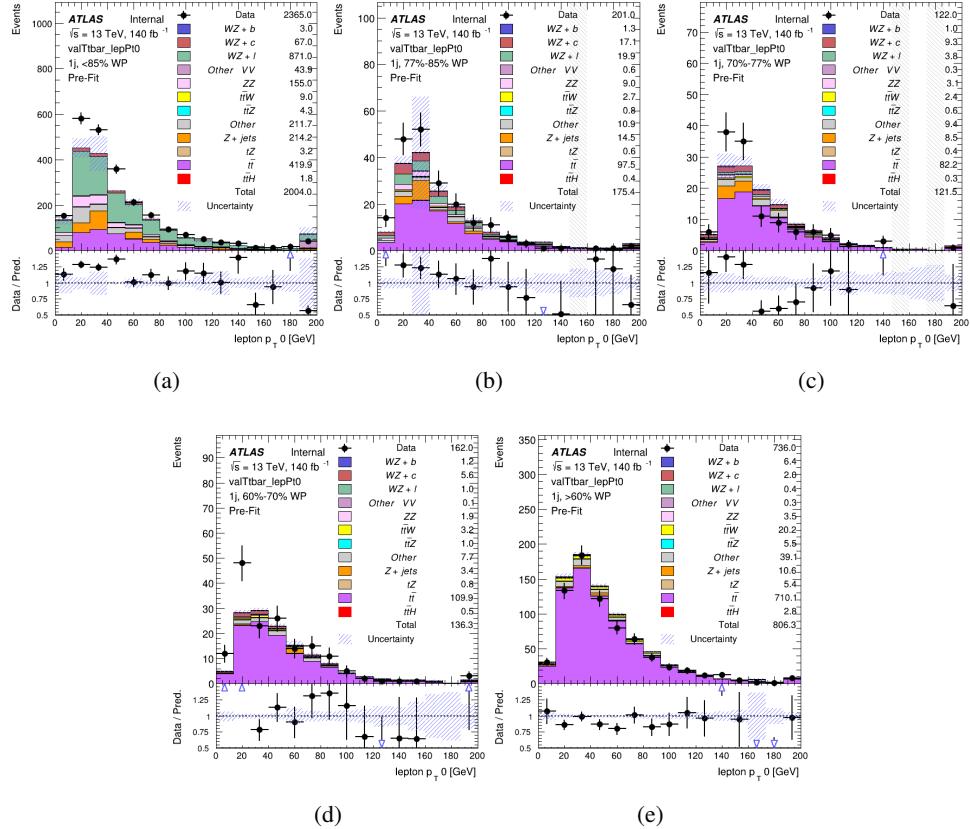


Figure B.3: Comparisons between the data and MC distributions of the p_T of the OS lepton in the $t\bar{t}$ CR for each of the 1-jet b-tag working point regions

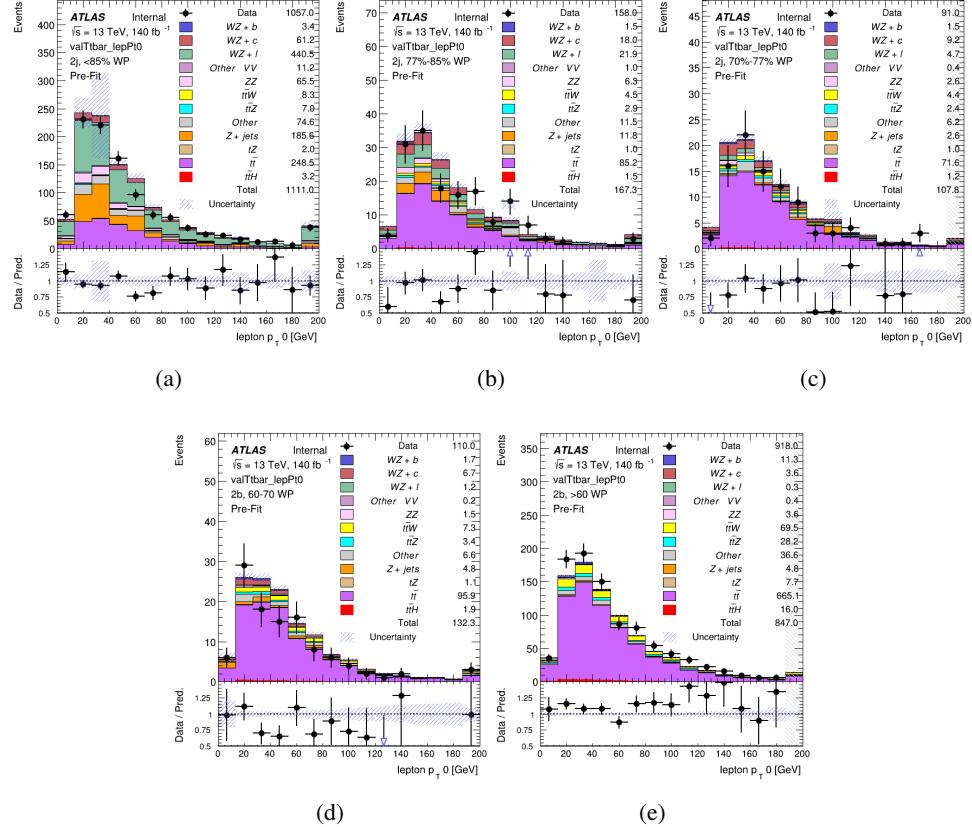


Figure B.4: Comparisons between the data and MC distributions of the p_T of the OS lepton in the $t\bar{t}$ CR for each of the 2-jet b-tag working point regions

¹⁵⁷⁶ **B.2 tZ Interference Studies**

1577 Because it includes an on-shell Z boson as well as a b-jet and W from the top decay, tZ
1578 production represents an identical final state to WZ + b-jet. This implies the possibility of matrix
1579 level interference between these two processes not accounted for in the Monte Carlo simulations,
1580 which consider the two processes independently. Truth level studies are performed in order to
1581 estimate the impact of these interference effects.

1582 In order to estimate the matrix level interference effects between tZ and WZ + b-jet, two
1583 different sets of simulations are produced using **MADGRAPH 5** [49] - one which simulates these
1584 two processes independently, and another where they are produced simultaneously, such that
1585 interference effects are present. These two sets of samples are then compared, and the difference
1586 between them can be taken to represent any interference effects.

1587 MadGraph simulations of 10,000 tZ and 10,000 WZ + b events are produced, along with
1588 20,000 events where both are present, in the fiducial region where three leptons and at least one
1589 jet are produced.

1590 A selection mimicking the preselection used in the main analysis is applied to the samples:
1591 The SS leptons are required to have $p_T > 20$ GeV, and > 10 GeV is required for the OS lepton.
1592 The associated b-jet is required to have $p_T > 25$ GeV, and all physics objects are required to fall
1593 in a range of $|\eta| < 2.5$.

1594 The overall cross-section with and without intereference effects agree within error, and
1595 no significant differences in the kinematic distributions are seen. It is therefore concluded that
1596 interference effects do not significantly impact the results.

¹⁵⁹⁷ **B.3 Alternate tZ Inclusive Fit**

1598 **B.3.1 tZ Inclusive Fit**

1599 While tZ is often considered as a distinct process from WZ + b, this could also be considered
 1600 part of the signal. Alternative studies are performed where, using the same framework as the
 1601 nominal analysis, a measurement of WZ + b is performed that includes tZ as part of WZ+b.

1602 Because of this change, the tZ CR is no longer necessary, and only the five pseudo-
 1603 continuous b-tag regions are used in the fit. Further, systematics related to the tZ cross-section
 1604 are removed from the fit, as they are now encompassed by the normalization measurement of
 1605 WZ + b. All other systematic uncertainties are carried over from the nominal analysis.

1606 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
 1607 fit, with an expected significance of 4.0σ .

1608 The impact of the predominate systematics are summarized in Table 44.

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

Table 44: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

1609 **B.3.2 Floating tZ**

1610 In order to quantify the impact of the tZ uncertainty on the fit, an alternate fit strategy is used
1611 where the tZ normalization is allowed to float. This normalization factor replaces the cross-
1612 section uncertainty on tZ, and all other parameters of the fit remain the same.

1613 An uncertainty of 17% on the normalization of tZ is extracted from the fit, compared to a
1614 theory uncertainty of 15% applied to the tZ cross-section. The measured uncertainties on WZ
1615 remain the same.

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

¹⁶¹⁶ The following section provides details of the various MVAs as well as a few studies performed
¹⁶¹⁷ in support of this analysis, exploring alternate decisions and strategies.

¹⁶¹⁹ **C.1 Higgs Reconstruction Model Details**

¹⁶²⁰ **C.1.1 b-jet Identification Features - 2lSS**

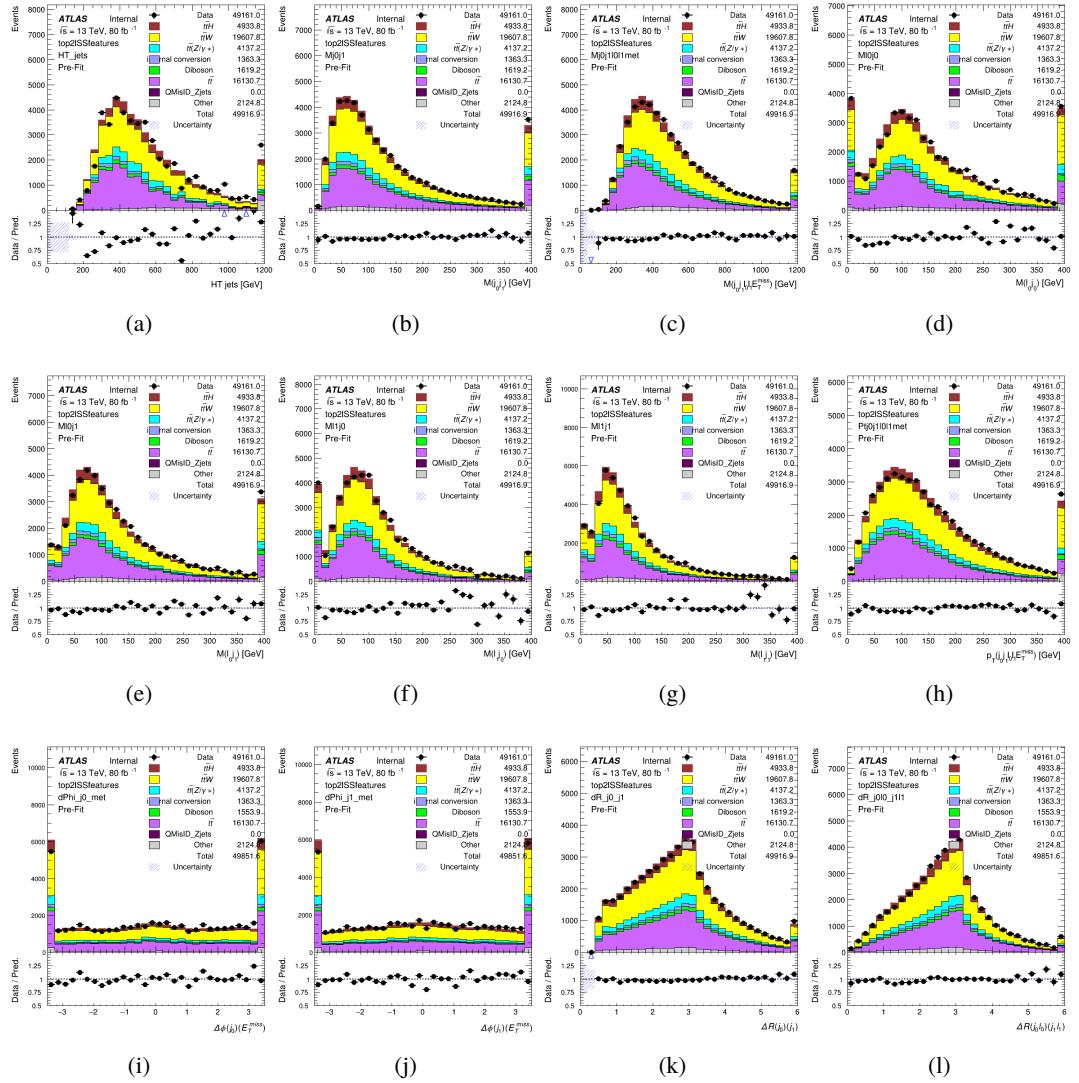


Figure C.1: Input features for top2lSS

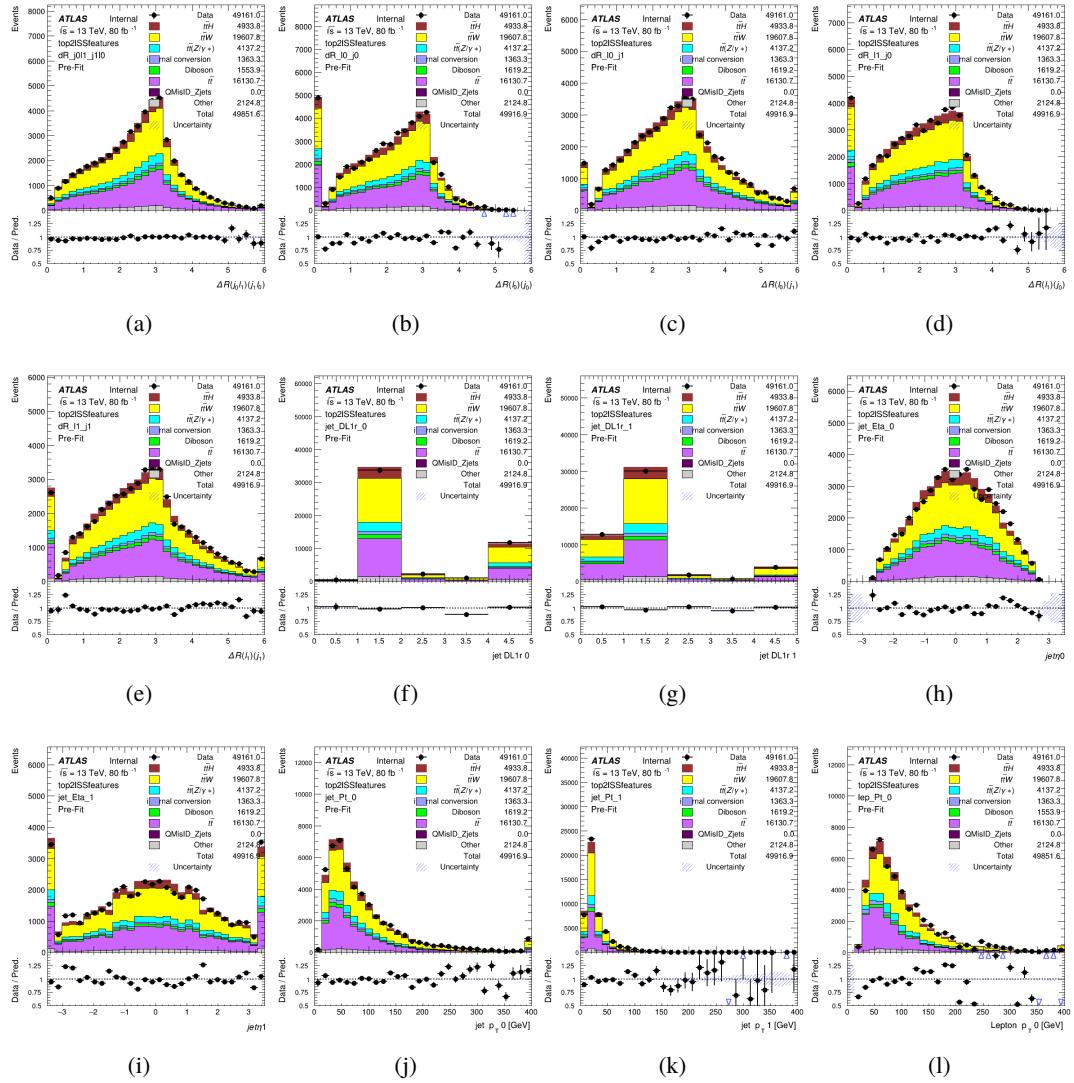


Figure C.2: Input features for top2lSS

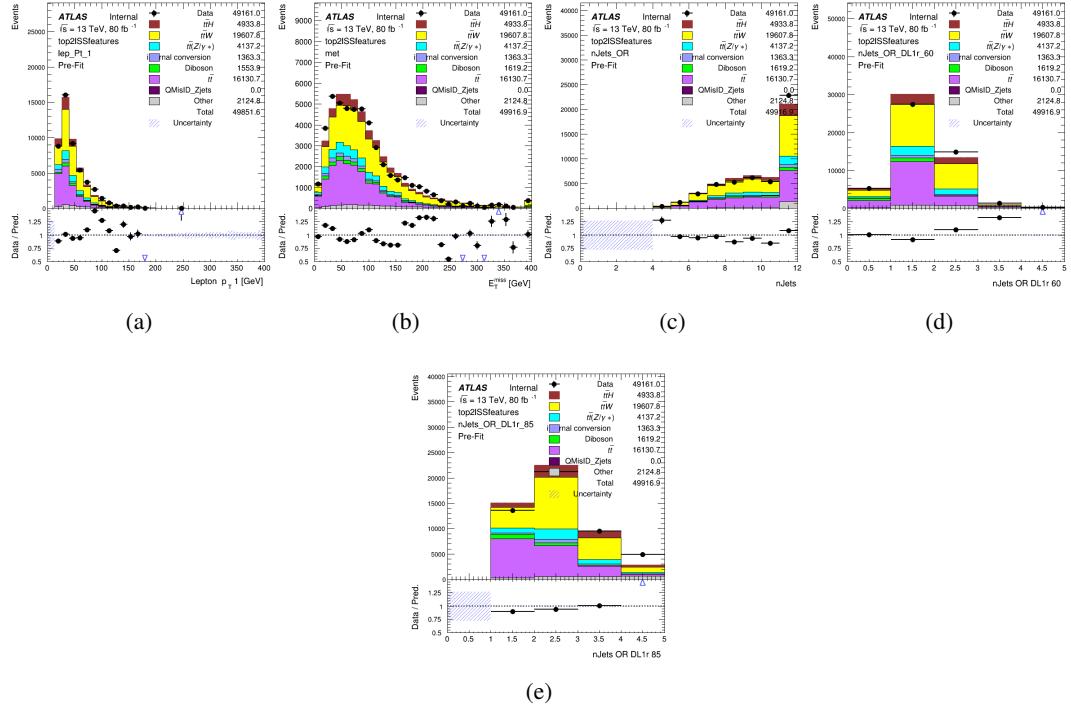


Figure C.3: Input features for top2ISS

1621 **C.1.2 b-jet Identification Features - 3l**

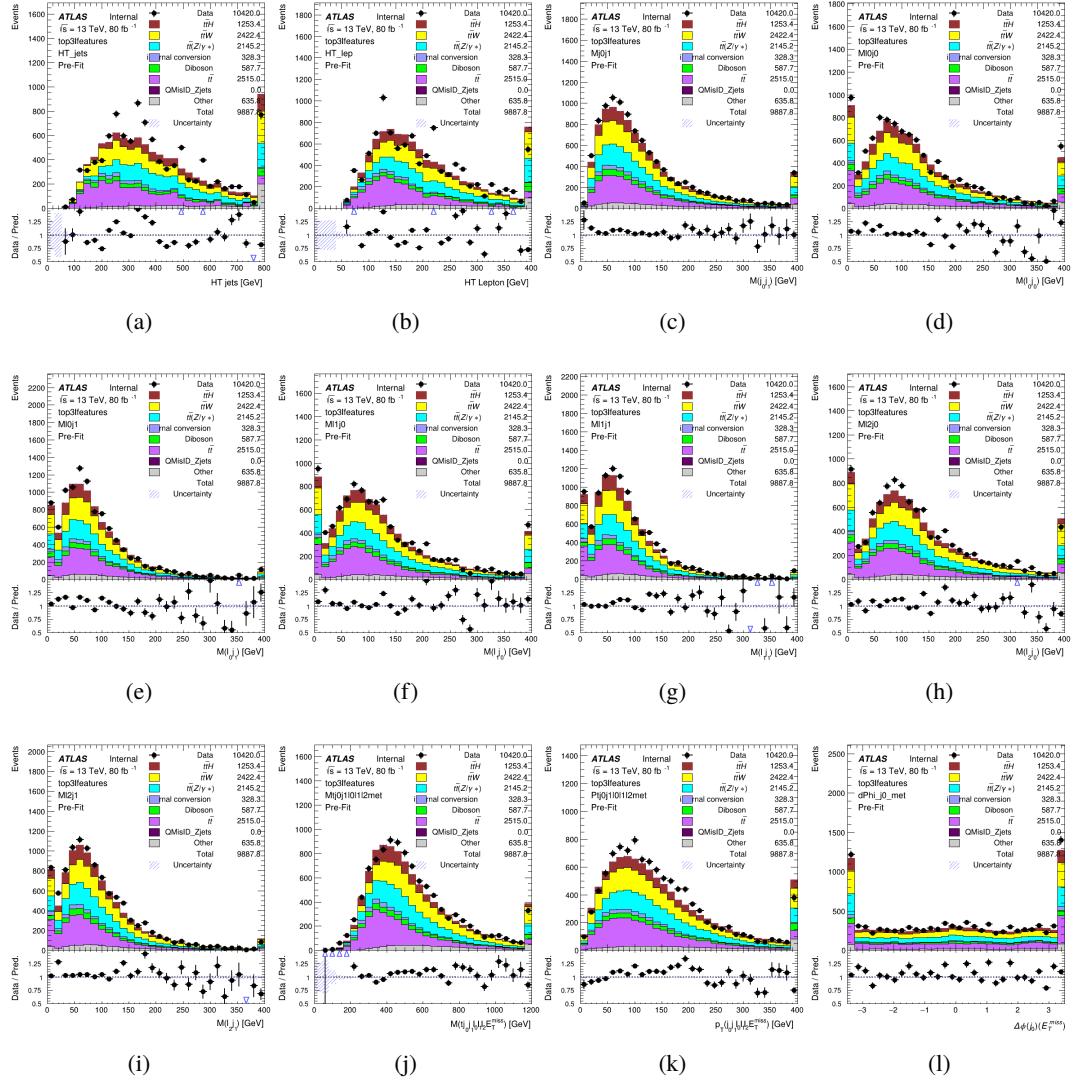


Figure C.4: Input features for top3l

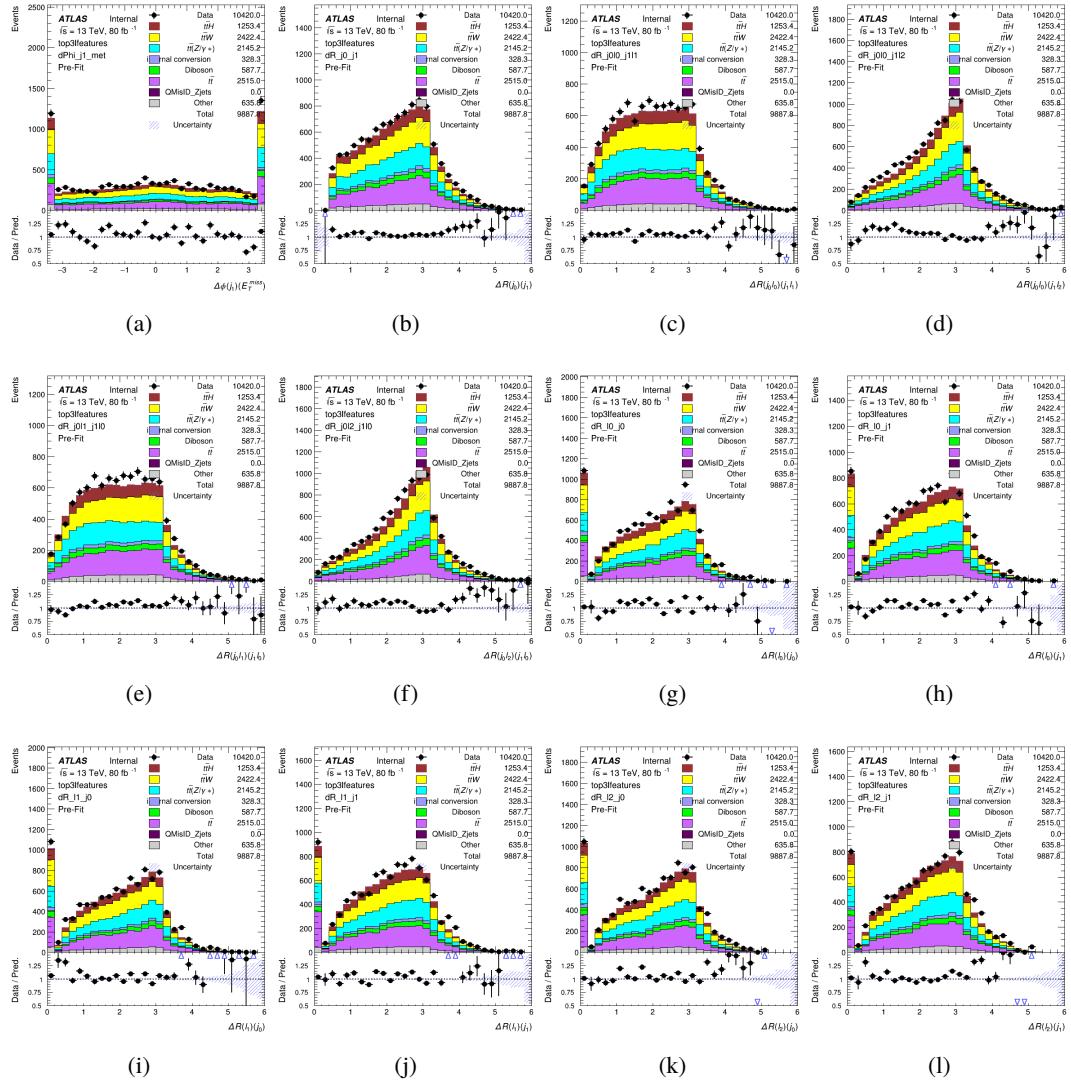


Figure C.5: Input features for top31

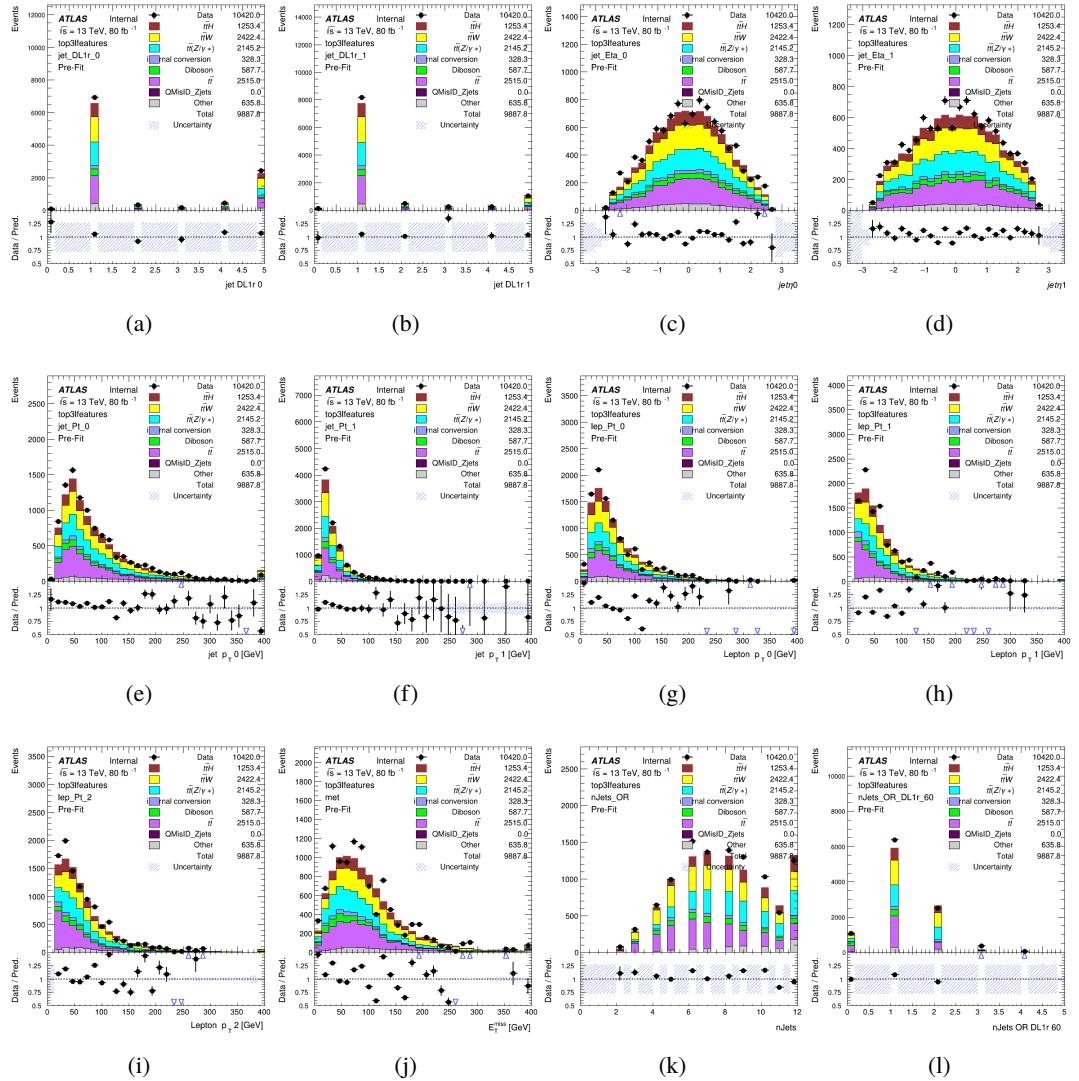
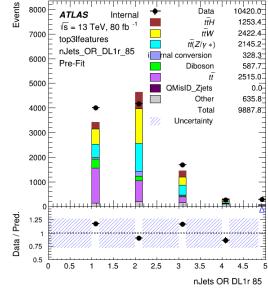


Figure C.6: Input features for top3l



(a)

Figure C.7: Input features for top3l

1622 **C.1.3 Higgs Reconstruction Features - 2lSS**

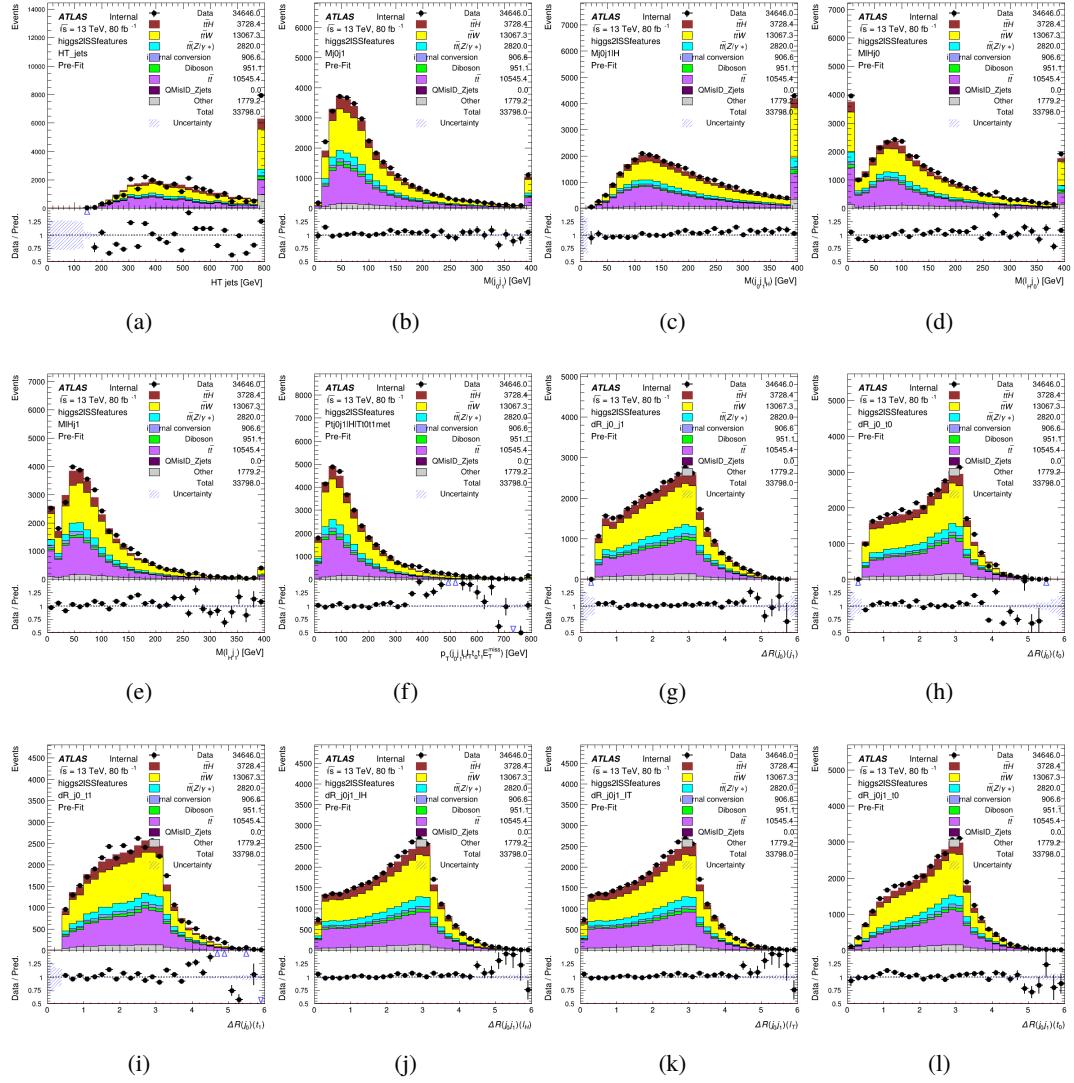


Figure C.8: Input features for higgs2ISST

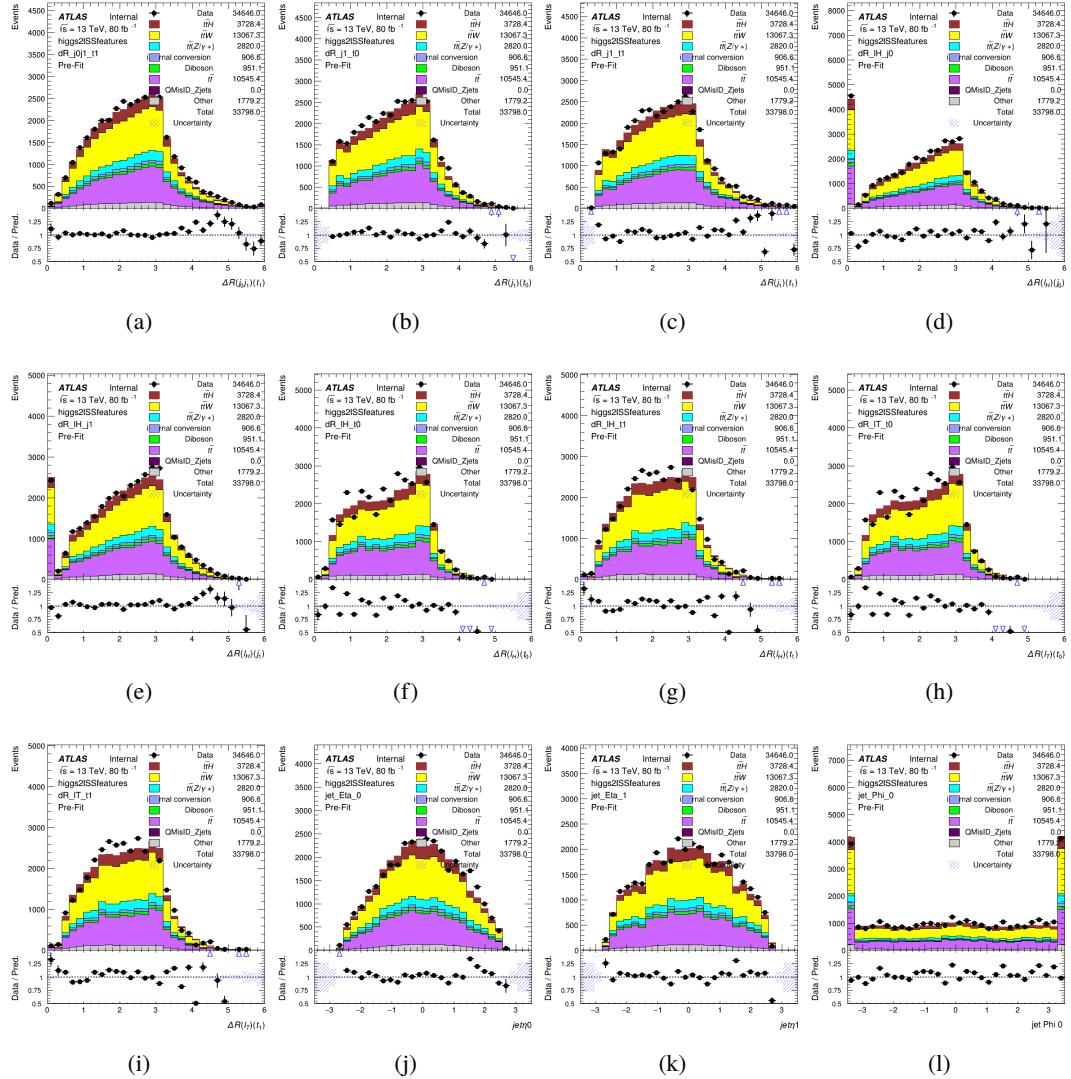


Figure C.9: Input features for higgs2ISSTest

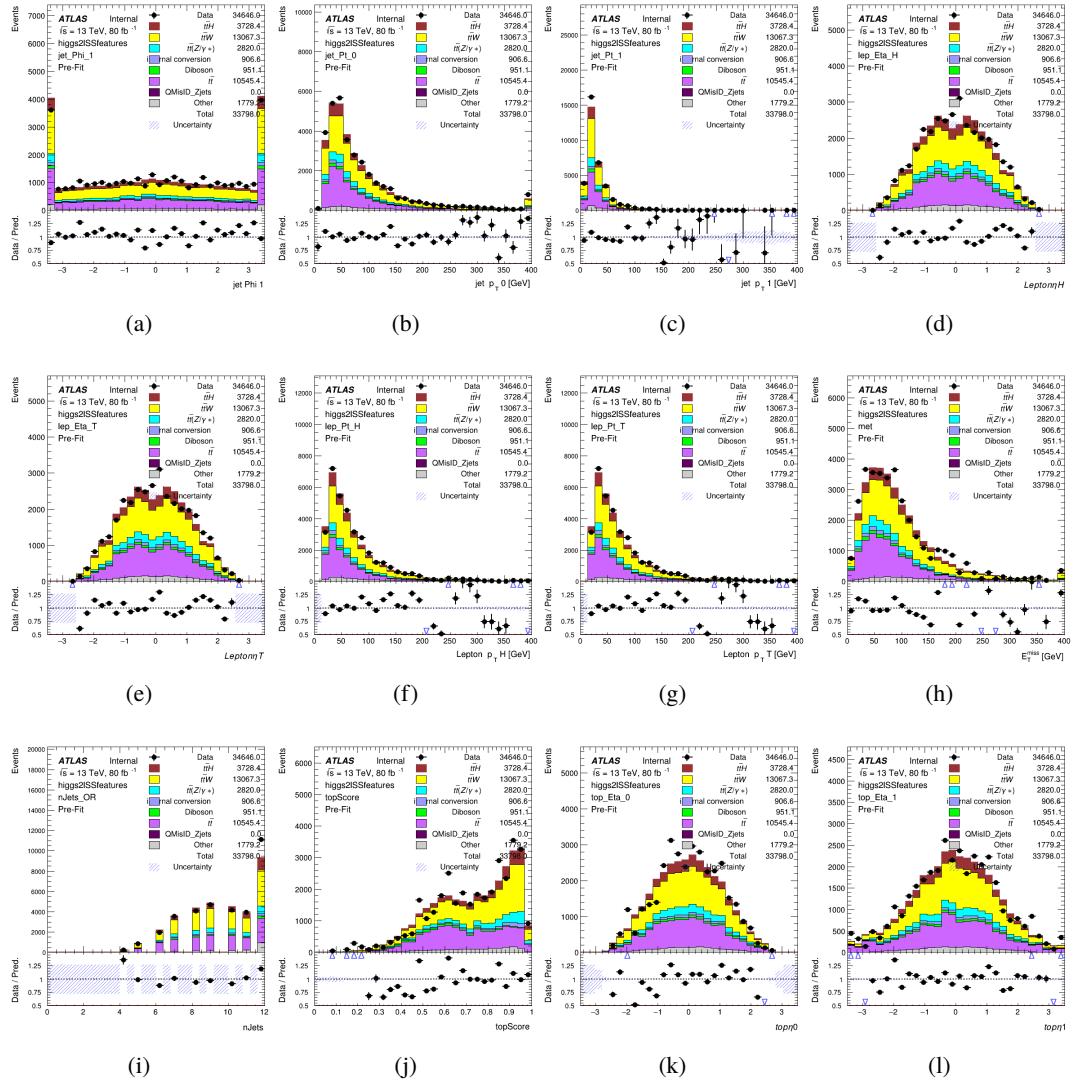


Figure C.10: Input features for higgs2lSS

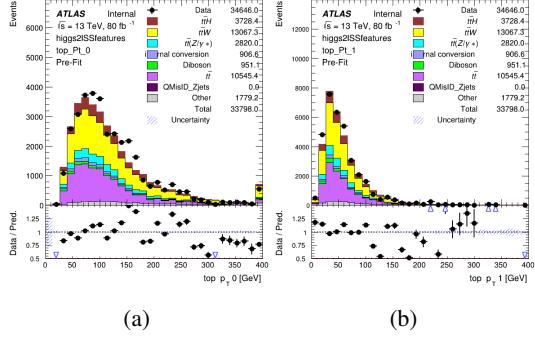


Figure C.11: Input features for higgs2lSS

1623 **C.1.4 Higgs Reconstruction Features - 3lS**

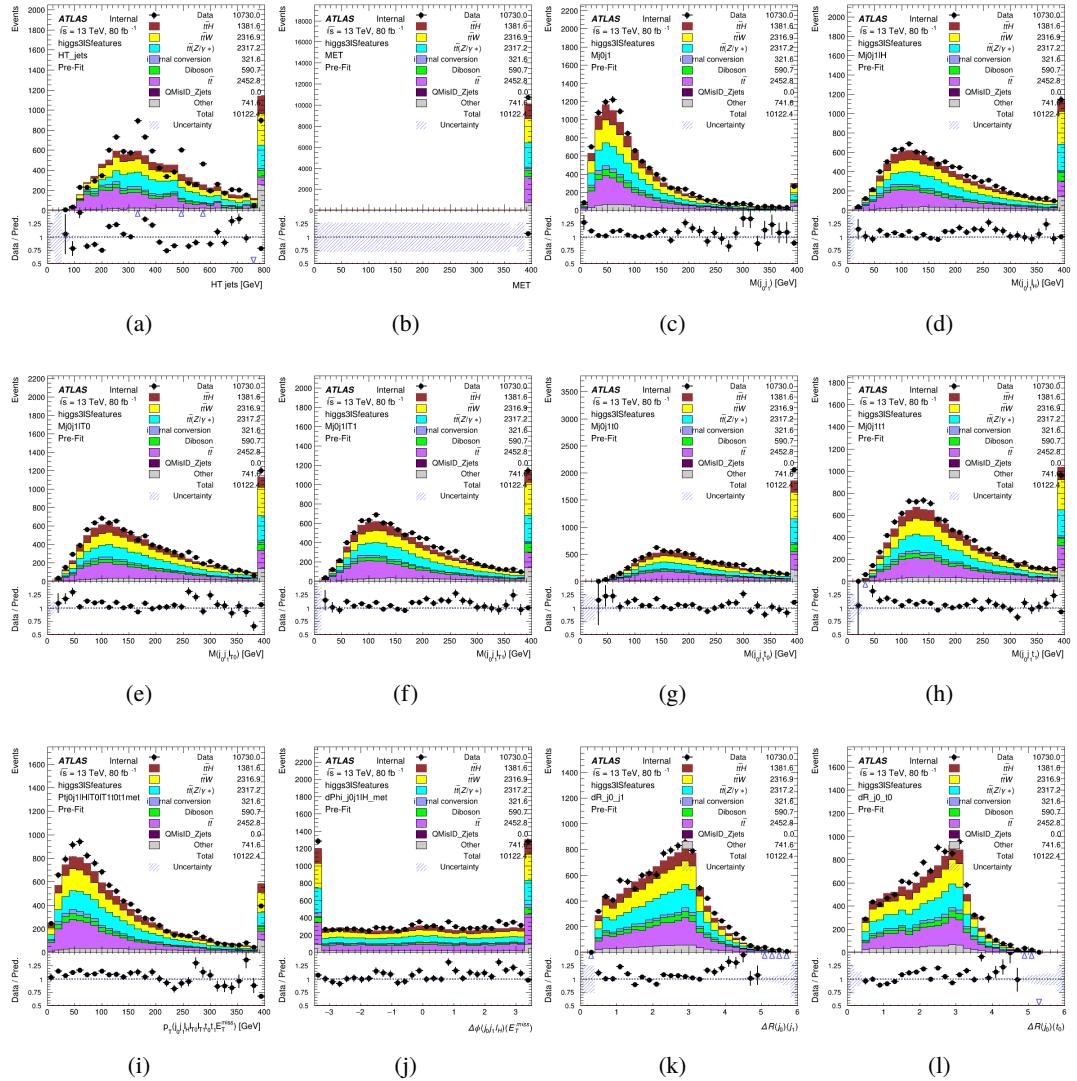


Figure C.12: Input features for higgs3lS

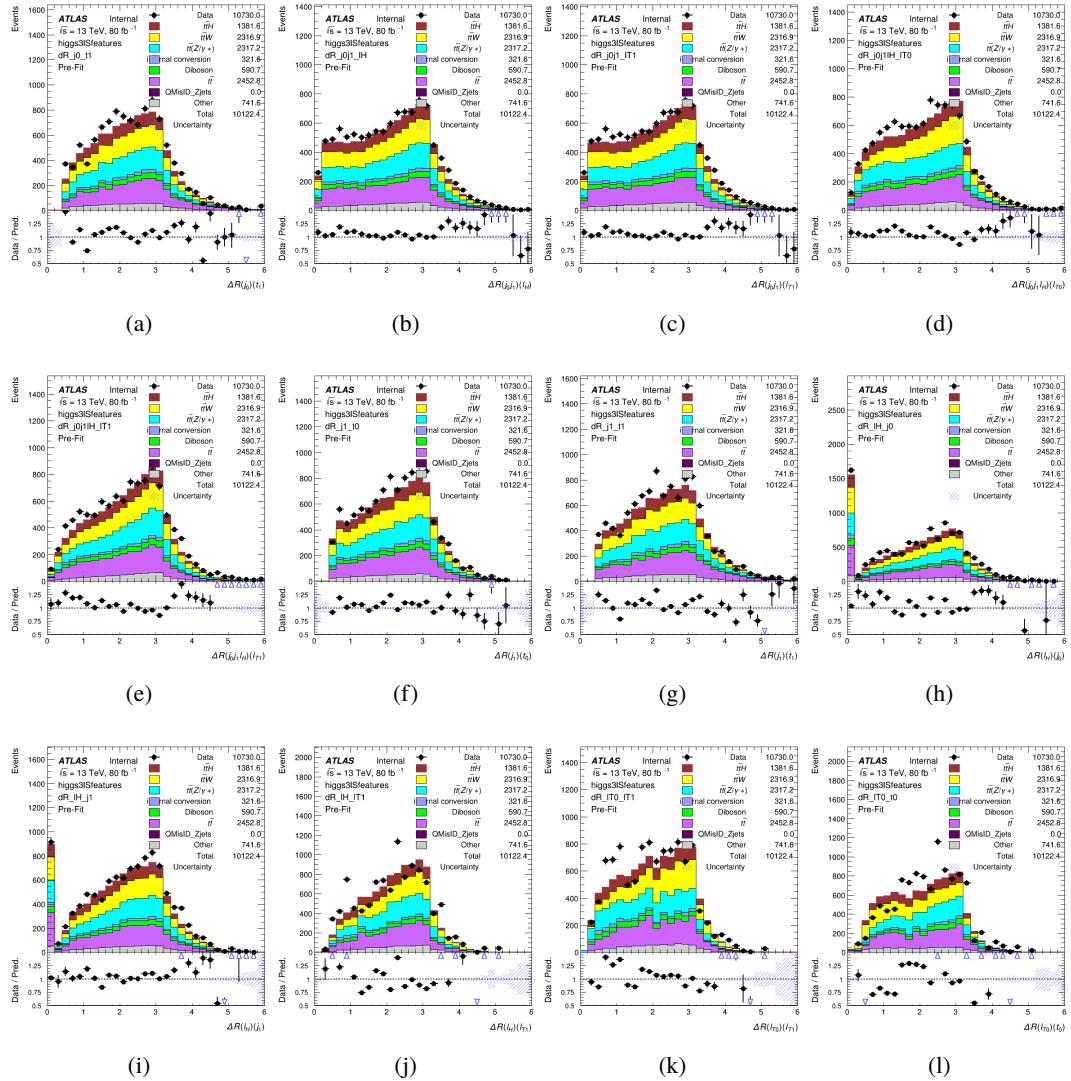


Figure C.13: Input features for higgs3lS

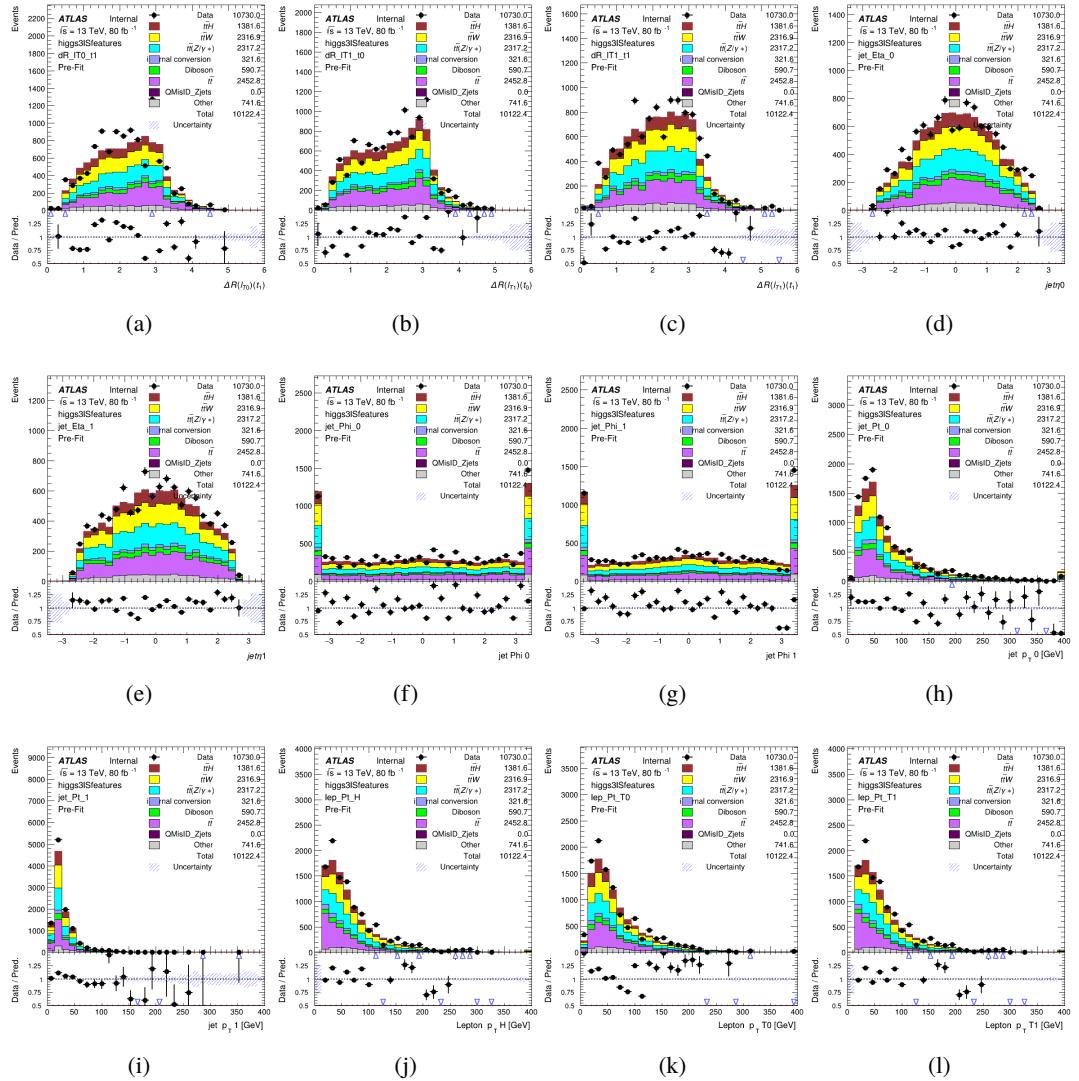


Figure C.14: Input features for higgs3lS

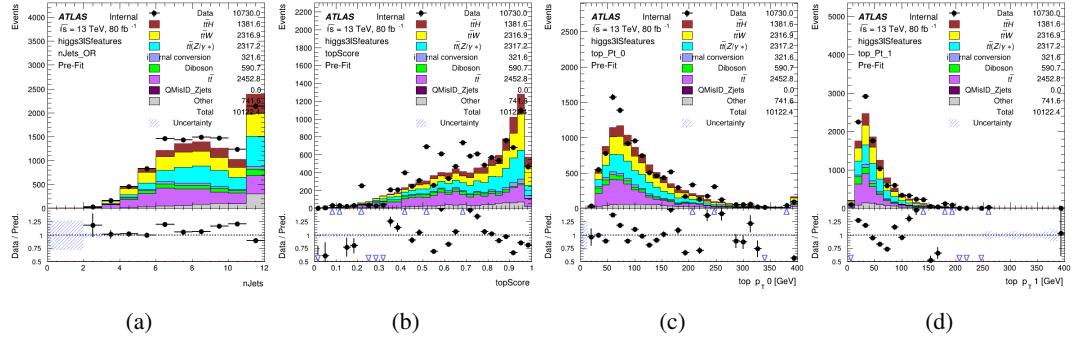


Figure C.15: Input features for higgs3IS

1624 **C.1.5 Higgs Reconstruction Features - 3lF**

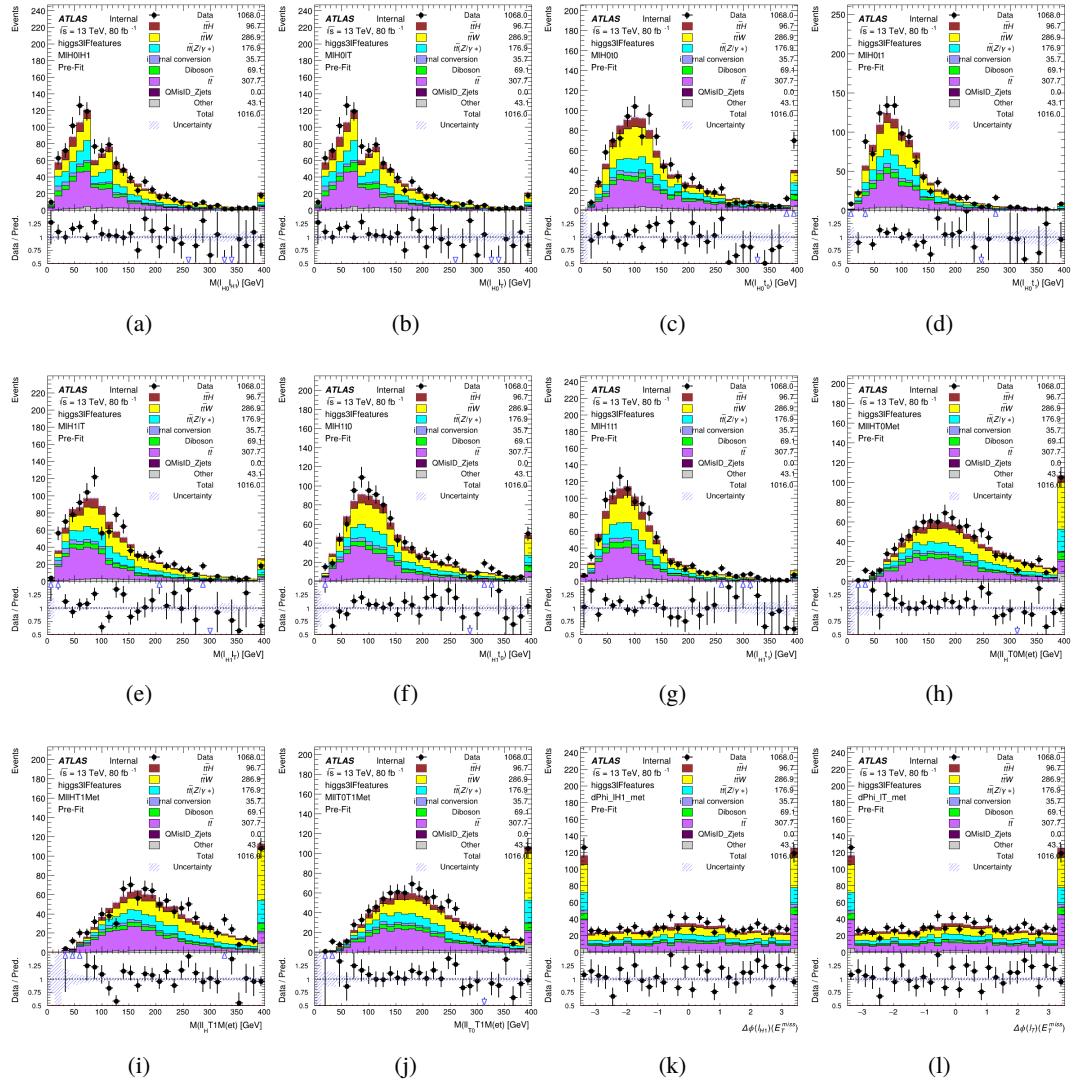


Figure C.16: Input features for higgs3lF

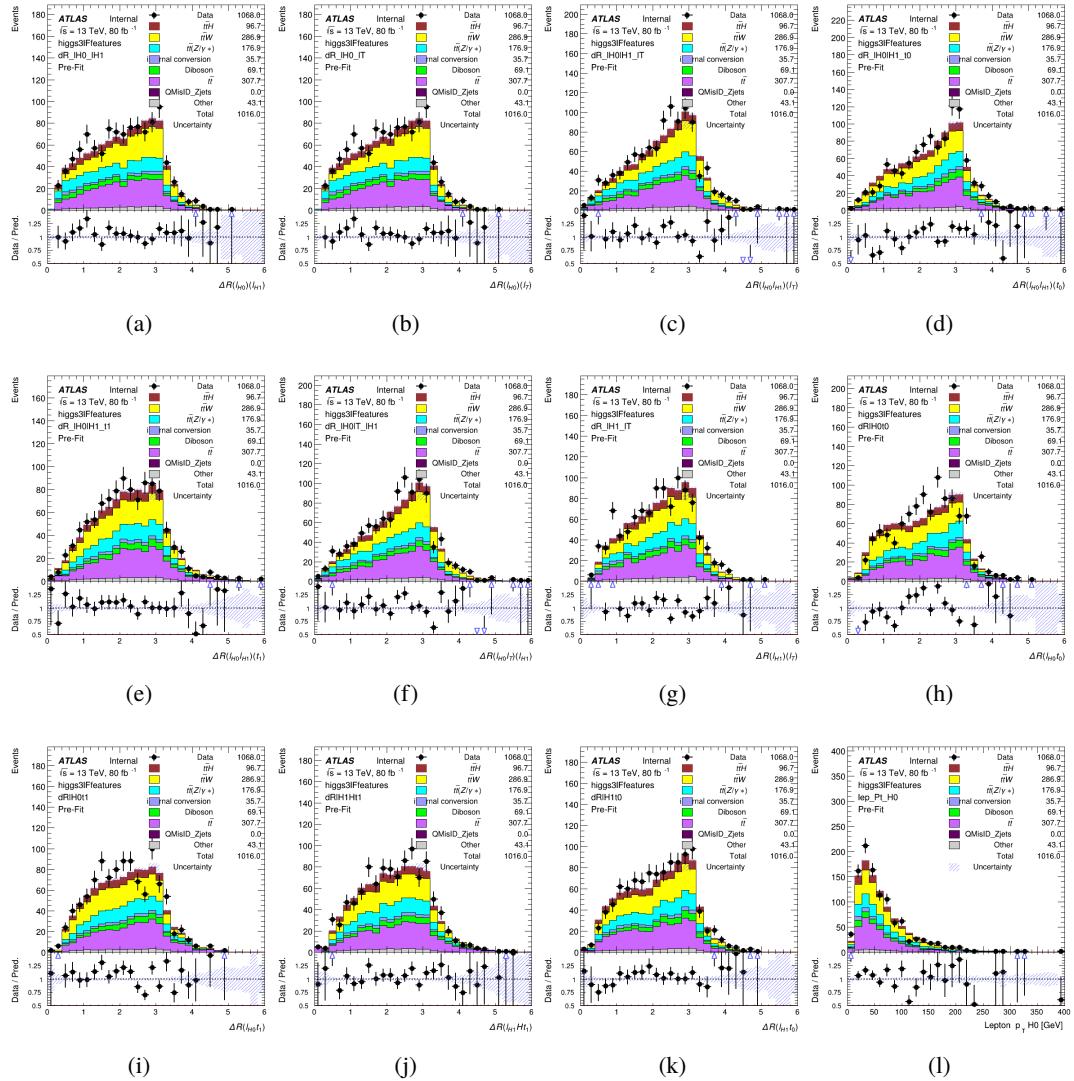


Figure C.17: Input features for higgs3lF

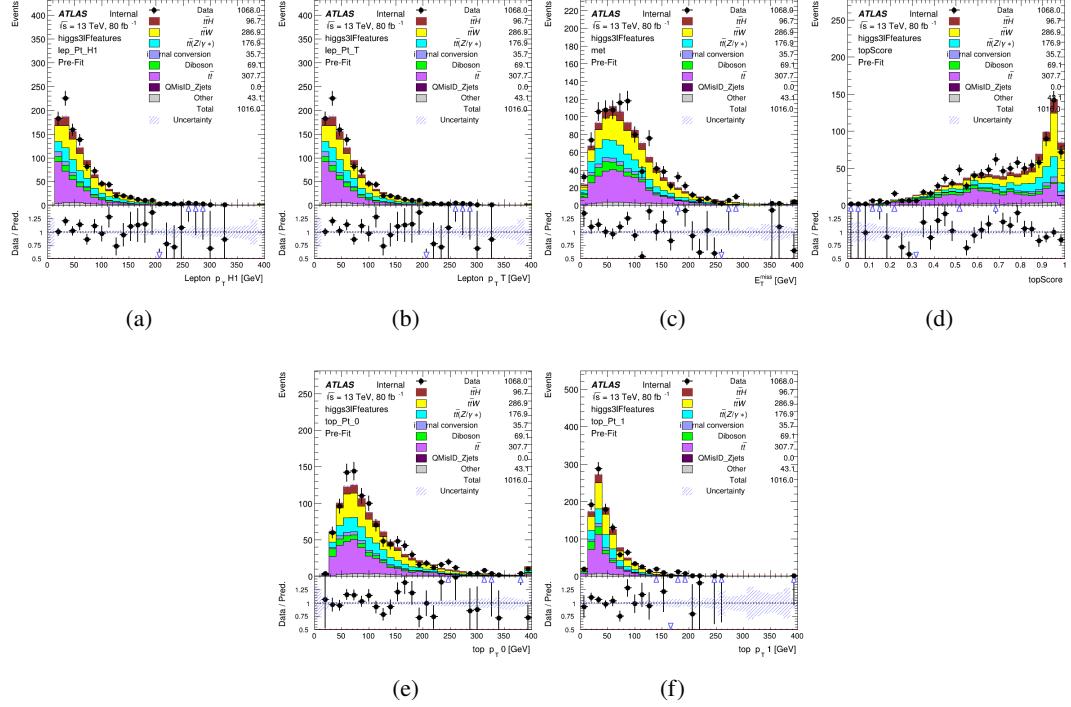


Figure C.18: Input features for higgs3lF

1625 **C.2 Background Rejection MVA Details**

1626 **C.2.1 Background Rejection MVA Features - 2ISS**

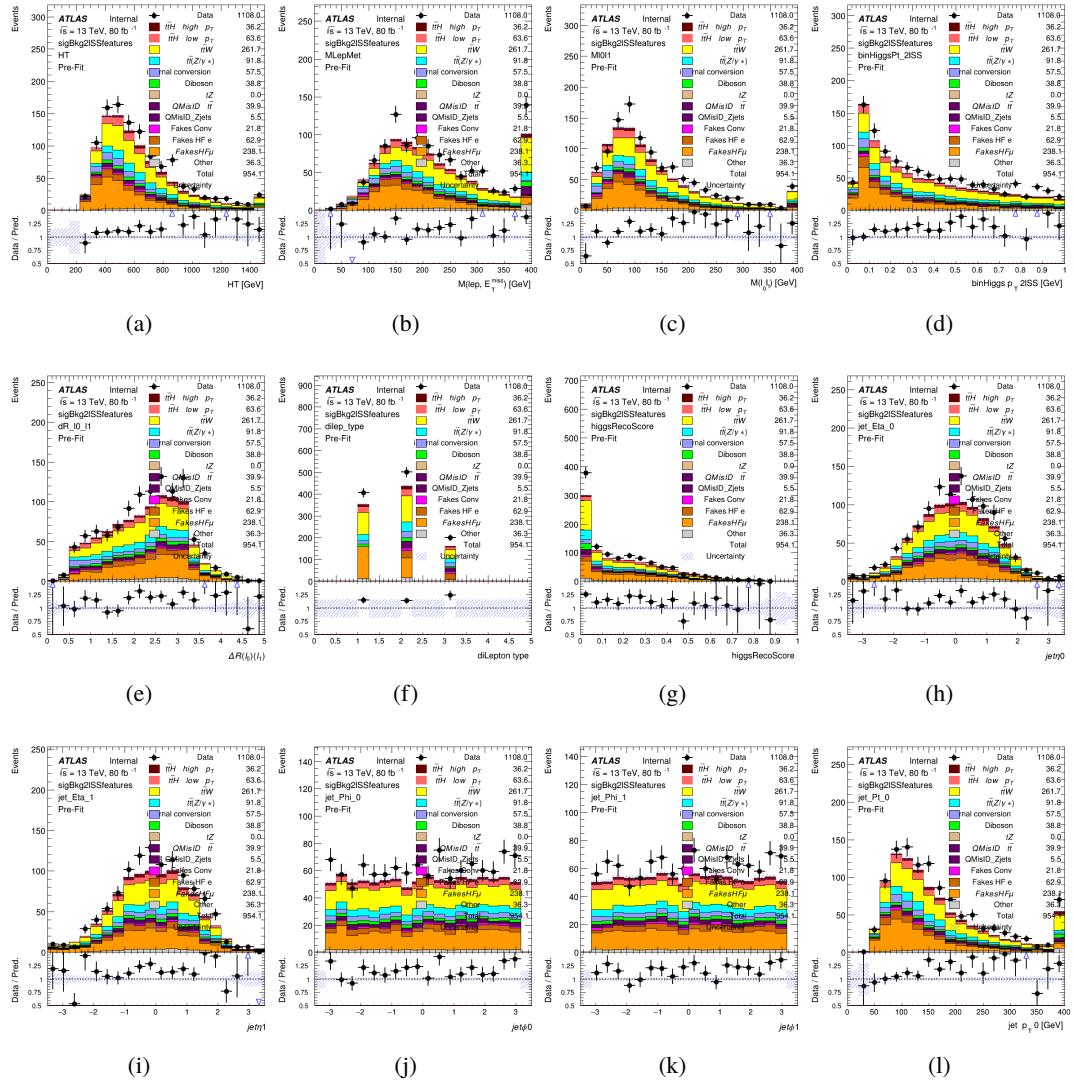


Figure C.19: Input features for sigBkg2ISS

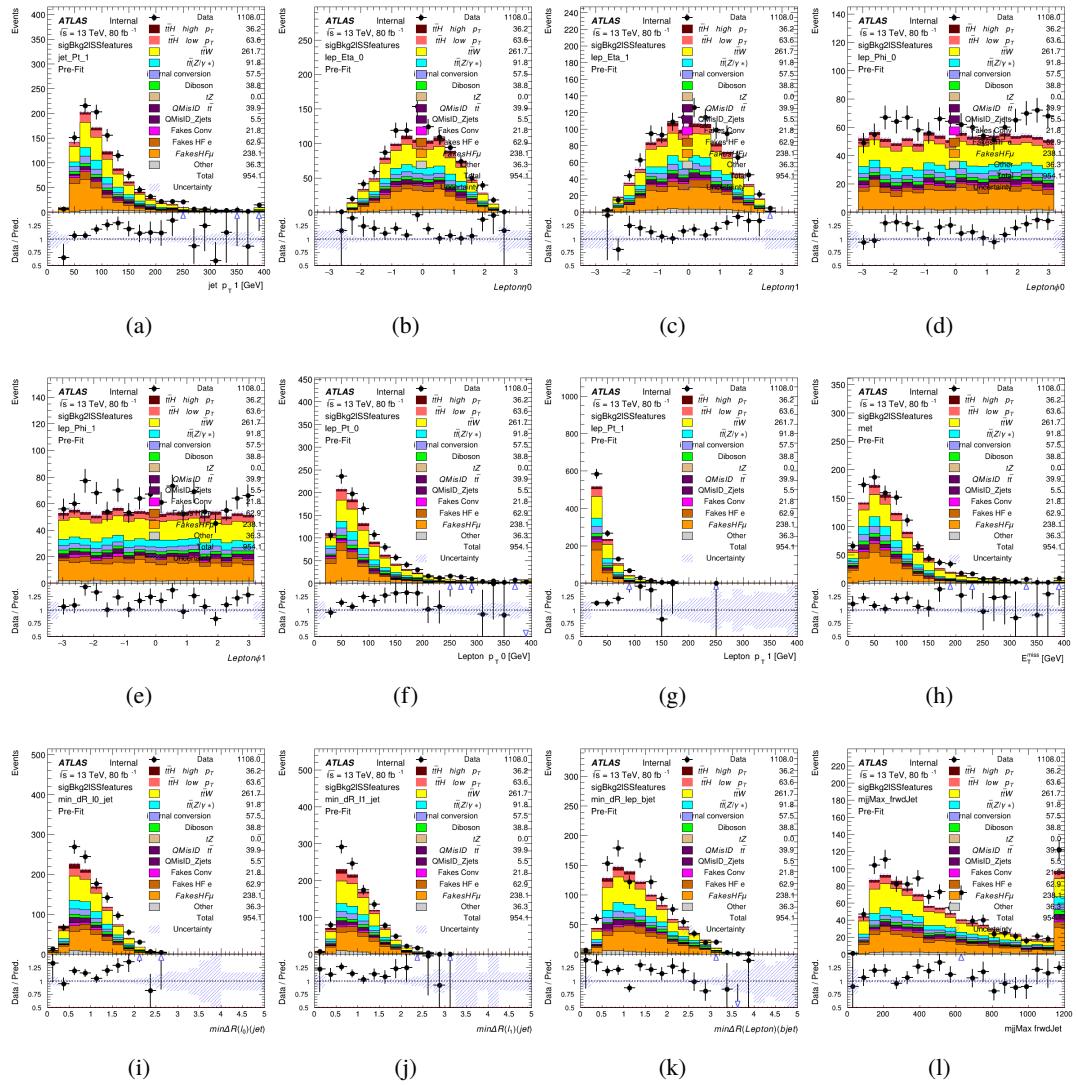


Figure C.20: Input features for sigBkg2lSS

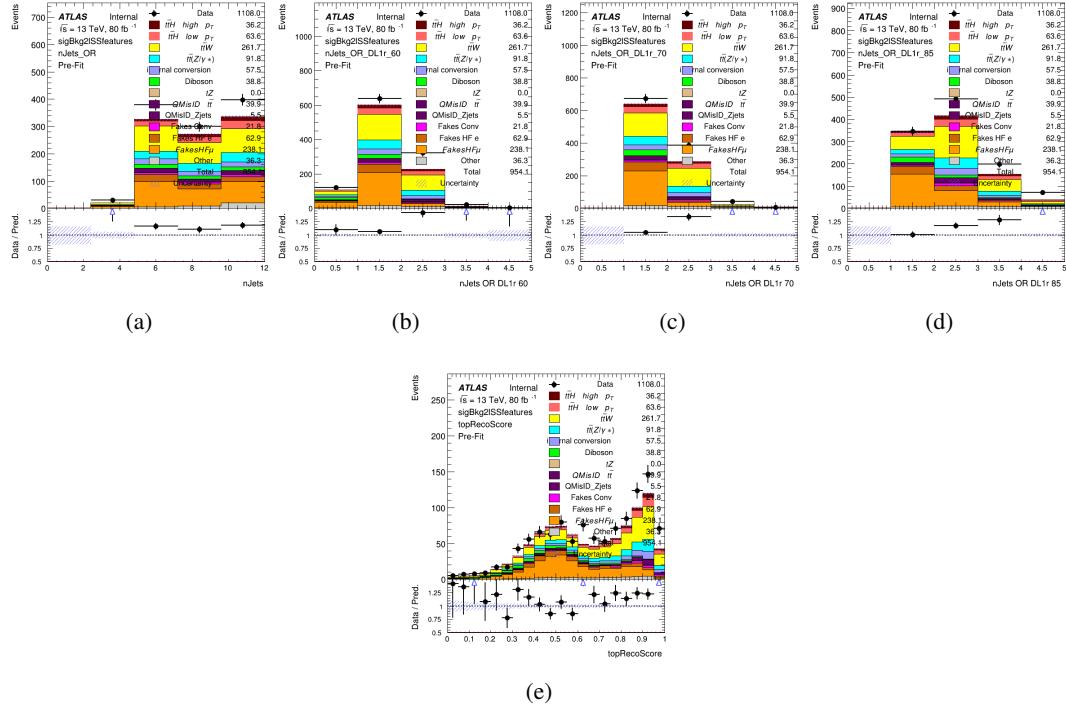


Figure C.21: Input features for sigBkg2lSS

1627 **C.2.2 Background Rejection MVA Features - 3l**

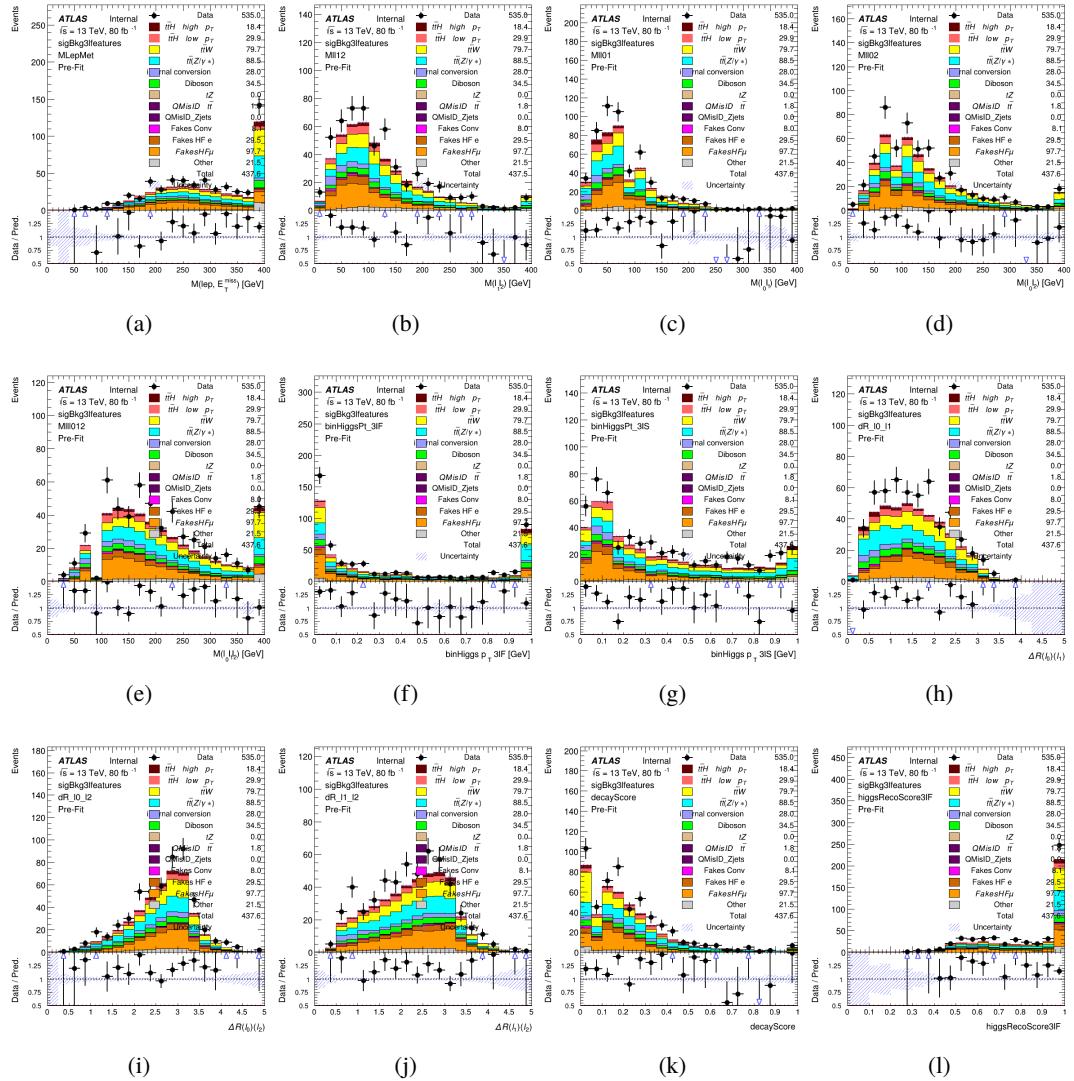


Figure C.22: Input features for sigBkg3l

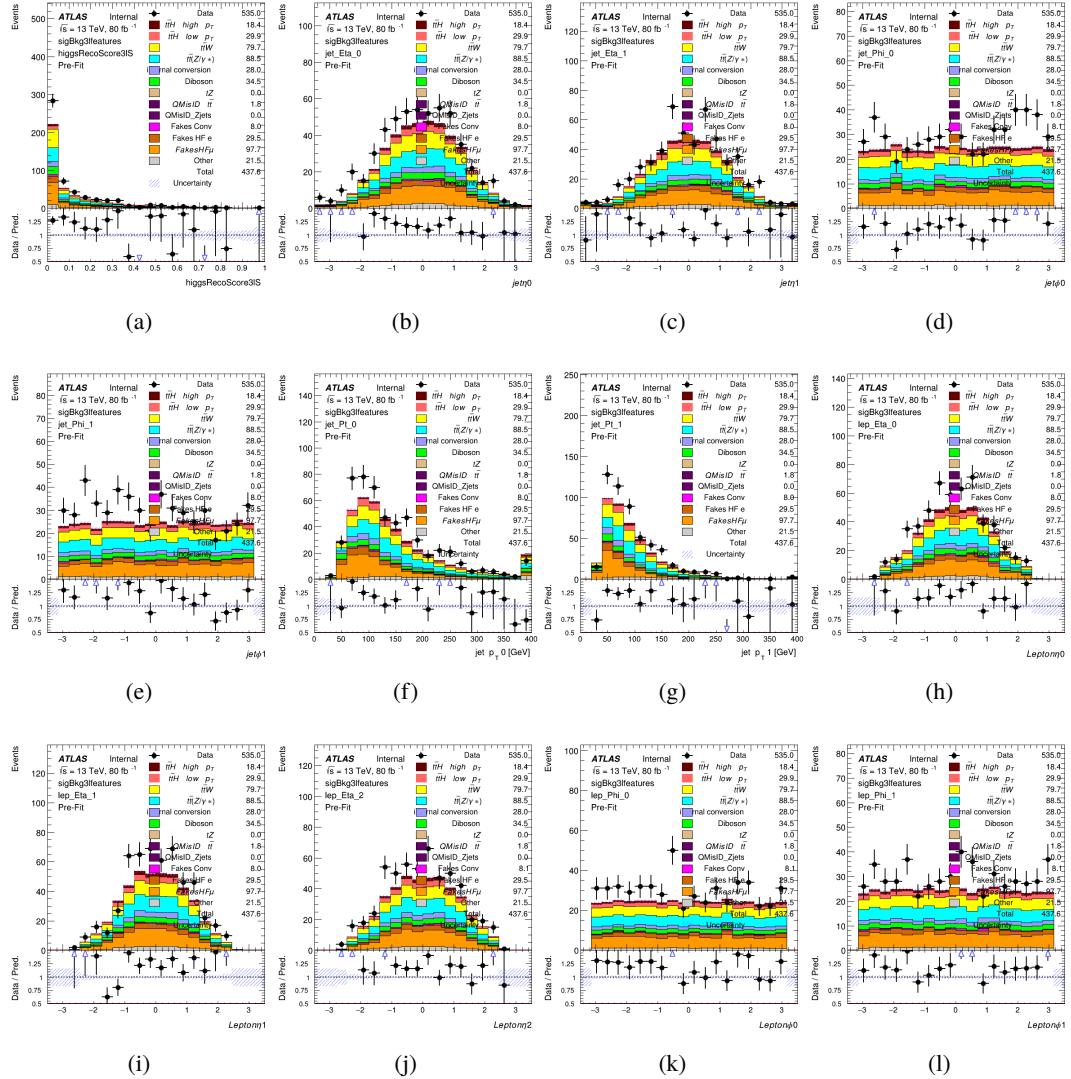


Figure C.23: Input features for sigBkg3l

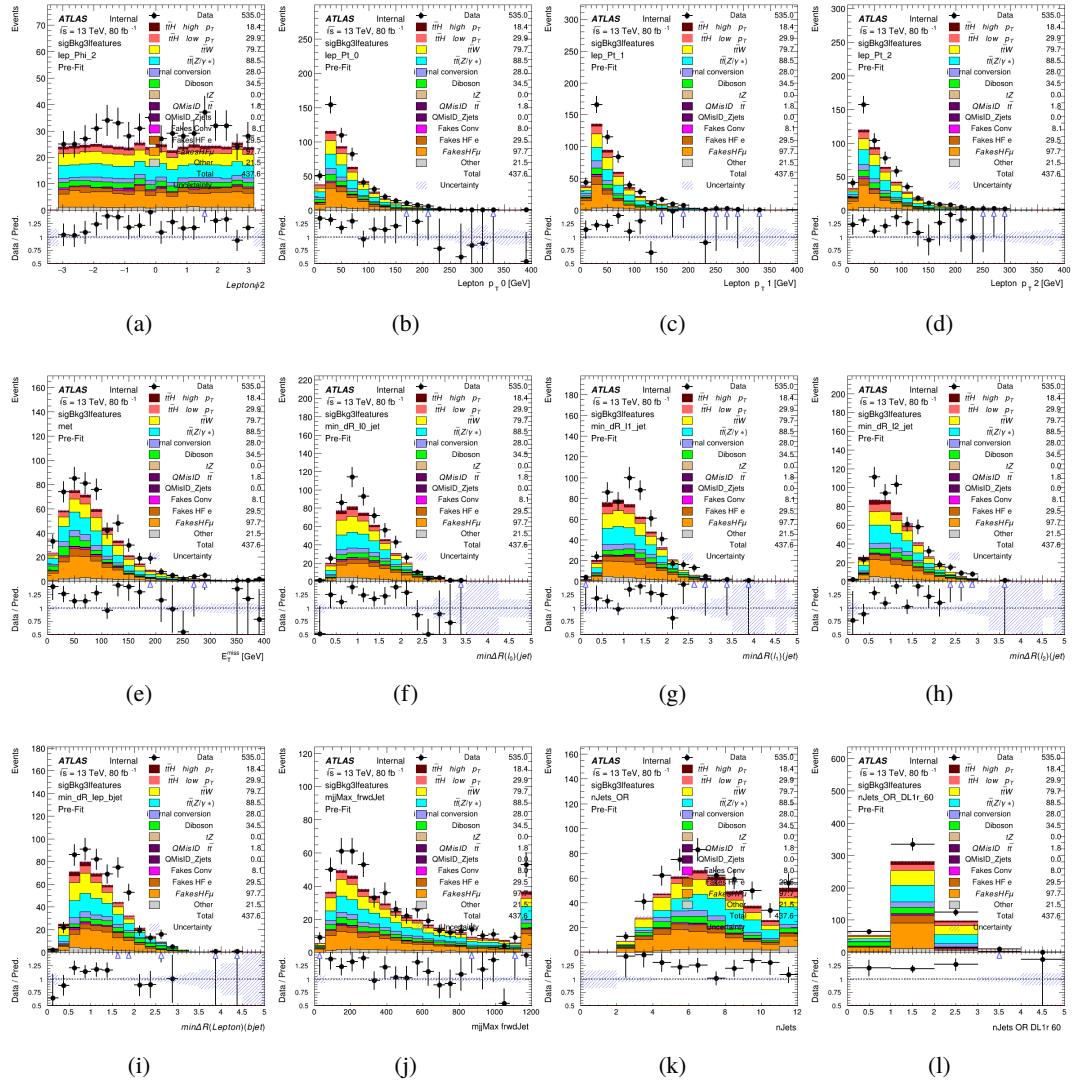


Figure C.24: Input features for sigBkg3l

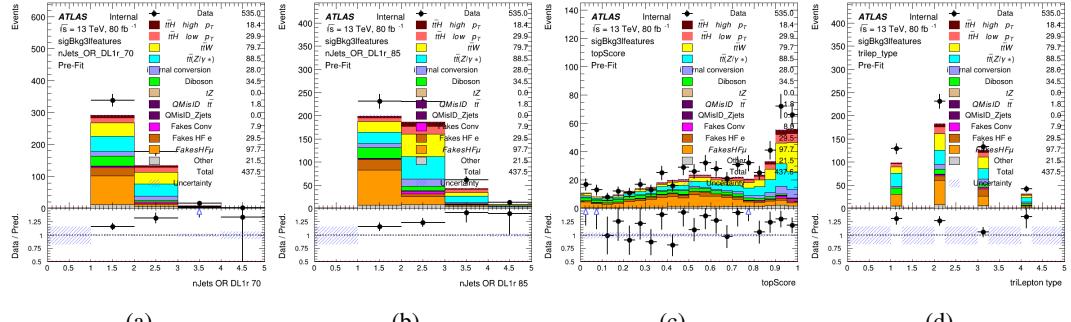


Figure C.25: Input features for sigBkg3l

1628 C.3 Truth Level Studies

1629 Attempts to identify the decay products of the Higgs are motivated by the ability to reconstruct
 1630 the Higgs momentum based on their kinematics. In order to demonstrate that this is case, the
 1631 kinematics of reconstructed objects that are truth matched to the Higgs decay are used as inputs
 1632 to a neural network which is designed to predict of the momentum of the Higgs. This is done in
 1633 the 2lSS channel, as it proves to be the most challenging for p_T reconstruction.

1634 Only leptons and jets which are truth matched to the Higgs are used as inputs for the
 1635 model; events where the lepton and both jets are not reconstructed are not included. The model
 1636 uses the same feature set and network architecture as the p_T prediction model used in the main
 1637 analysis, as described in Section 14.4.1.

1638 The results of the model are summarized below:

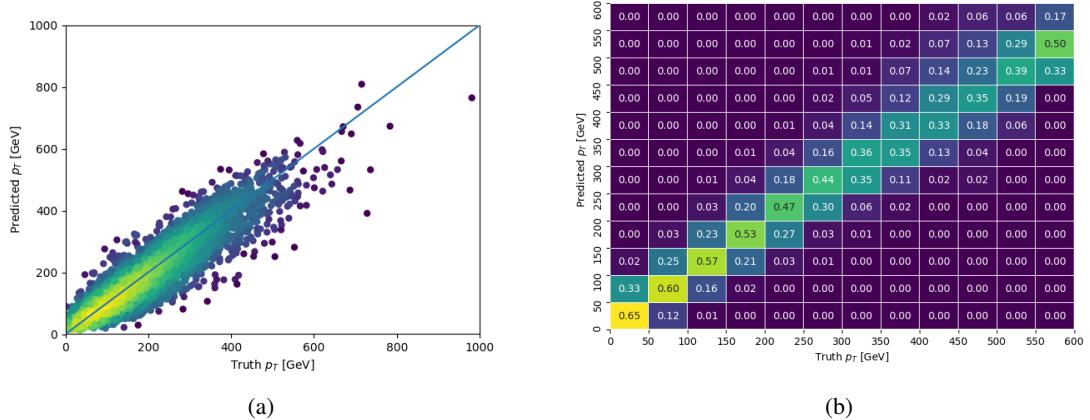


Figure C.26: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

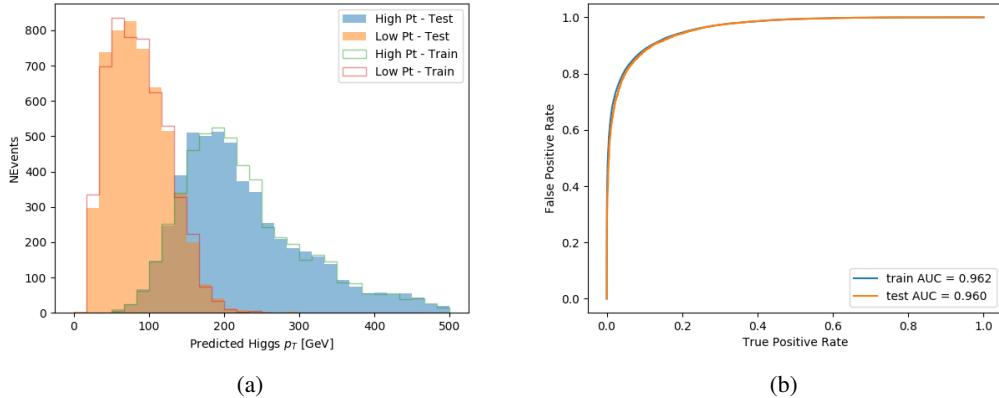


Figure C.27: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

1639 Based on the performance of the model, as shown Figures C.26 and C.27, the Higgs
 1640 momentum can be reconstructed with fairly high precision when its decay products are correctly
 1641 identified.

1642 C.4 Alternate b-jet Identification Algorithm

1643 The nominal analysis reconstructs the b-jets by considering different combinations of jets, and
 1644 asking a neural network to determine whether each combination consists of b-jets from top quark
 1645 decays. An alternate approach would be to give the neural network about all of the jets in an event
 1646 at once, and train it to select which two are most likely to be the b-jets from top decay. It was
 1647 hypothesized that this could perform better than considering each combination independently, as
 1648 the neural network could consider the event as a whole. While this is not found to be the case,
 1649 these studies are documented here as a point of interest and comparison.

1650 For these studies, the kinematics of the 10 highest p_T jets in each event are used for
 1651 training. This includes the vast majority of truth b-jets. Specifically the p_T , η , ϕ , E , and DL1r
 1652 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.
 1653 The p_T , η , ϕ , and E of the leptons and E_T^{miss} are included as well. Categorical cross entropy is
 1654 used as the loss function.

Table 45: 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%

1655 C.5 Binary Classification of the Higgs p_T

1656 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer
 1657 resolution. This means separating high and low p_T events is sufficient for this analysis. As
 1658 such, rather than attempting to reconstruct the full Higgs p_T spectrum, a binary classification
 1659 approach is explored.

1660 A model is built to determine whether $t\bar{t}H$ events include a high p_T (>150 GeV) or low
 1661 p_T (<150 GeV) Higgs Boson. While this is now a classification model, it uses the same input
 1662 features described in section 14.4. Binary crossentropy is used as the loss function.

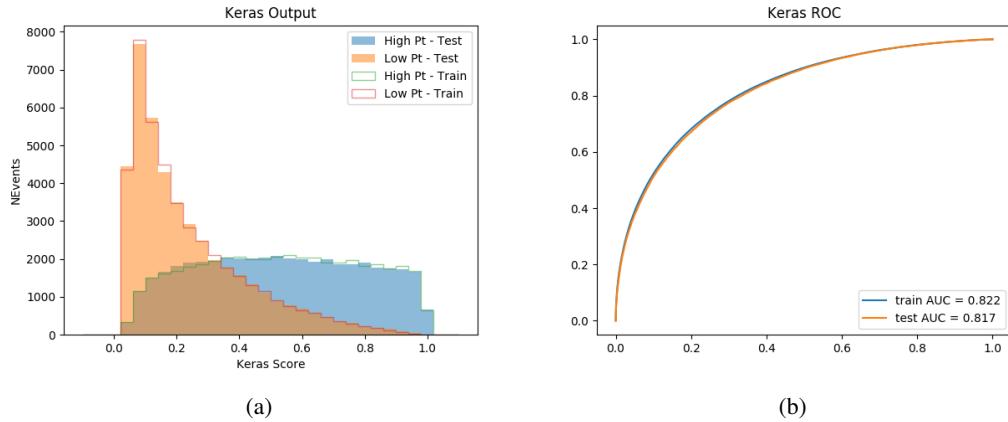


Figure C.28: Output distribution of the NN score for the binary high/low p_T separation model in the 2lSS channel.

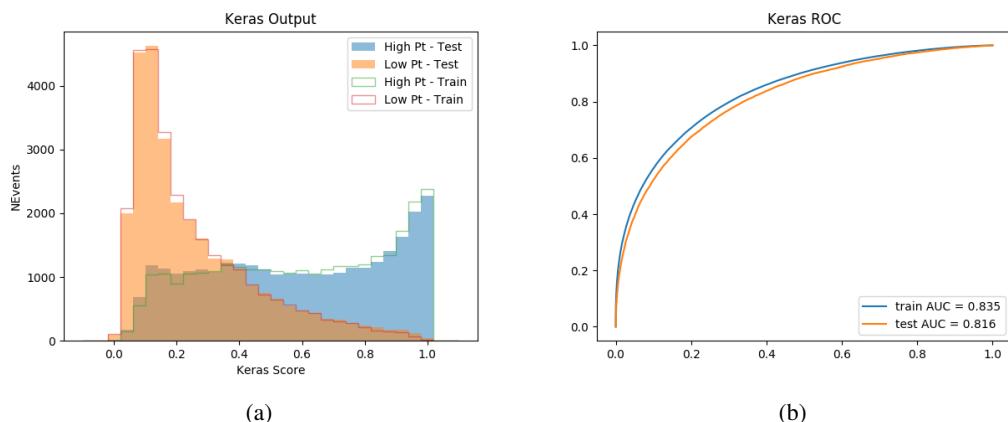


Figure C.29: Output distribution of the NN score for the binary high/low p_T separation model in the 3lS channel.

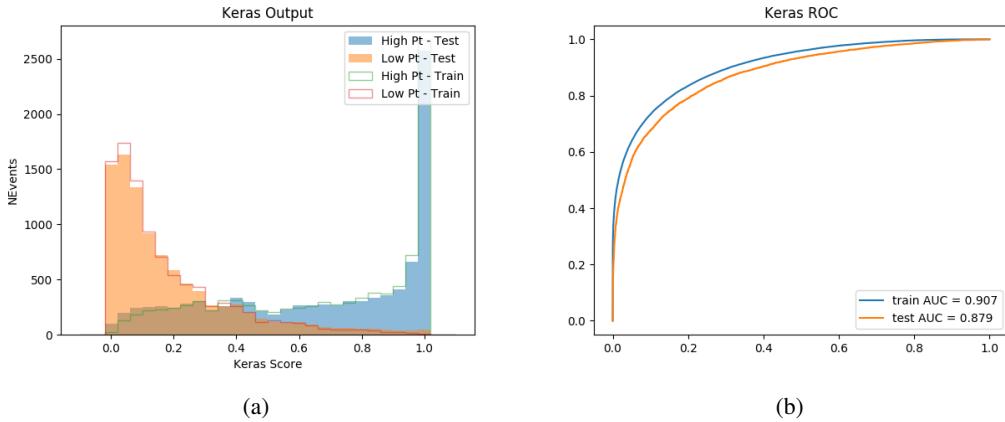


Figure C.30: Output distribution of the NN score for the binary high/low p_T separation model in the 3LS channel.

C.6 Impact of Alternative Jet Selection

1664 A relatively low p_T threshold of 15 GeV is used to determine jet candidates, as the jets originating
 1665 from the Higgs decay are found to fall between 15 and 25 GeV a large fraction of the time. The
 1666 impact of different jet p_T cuts on our ability to reconstruct the Higgs p_T is explored here.

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

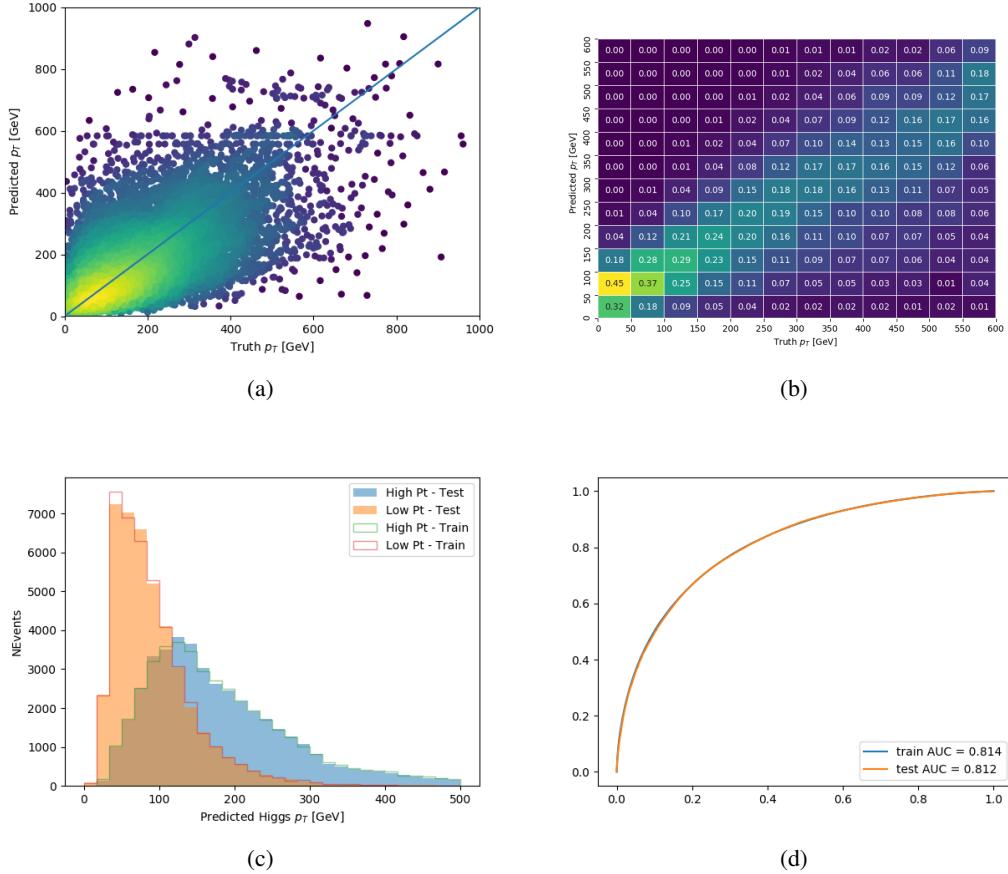
Jet $p_T > 20$ GeV

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

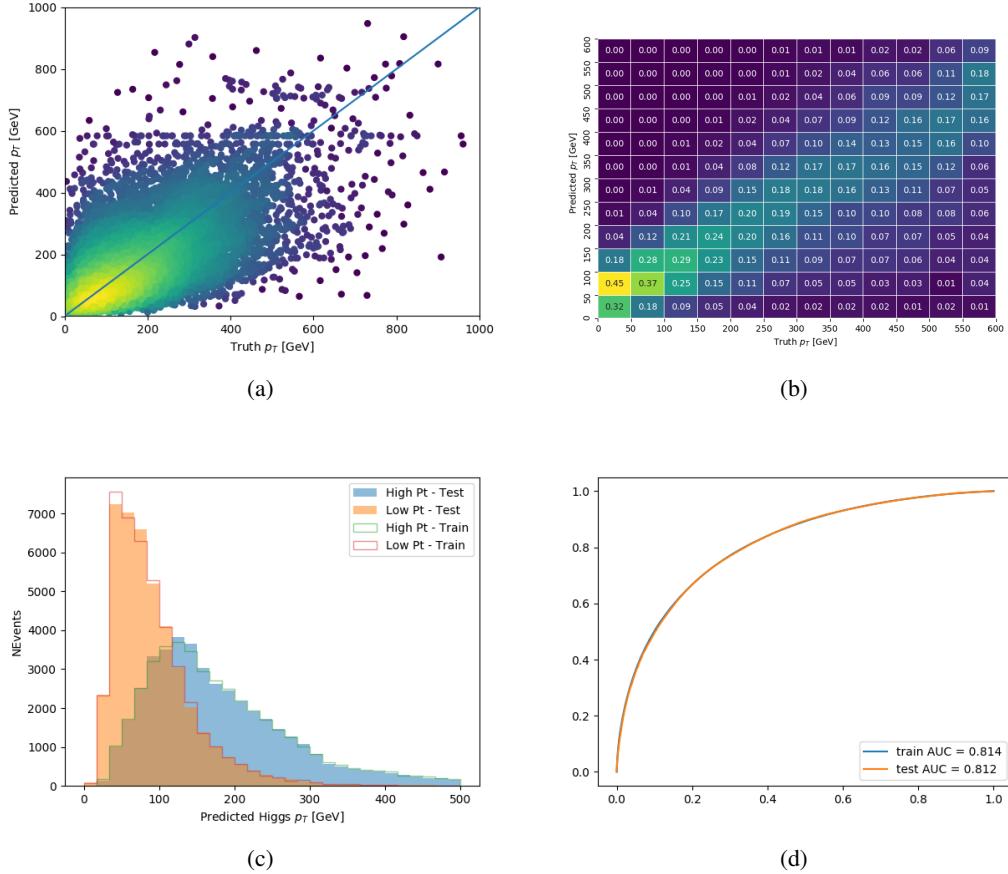
Jet $p_T > 25 \text{ GeV}$ 

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

1670 **D**