



ATLAS Note

GROUP-2017-XX



11th May 2021

Draft version 0.1

1

2

3

4

5

6

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

The ATLAS Collaboration

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

14 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	6
28	1 Introduction	6
29	II Theoretical Motivation	9
30	2 The Standard Model and the Higgs Boson	9
31	2.1 The Forces and Particles of the Standard Model	9
32	2.2 The Higgs Mechanism	12
33	2.2.1 The Higgs Field	12
34	2.2.2 Electroweak Symmetry Breaking	14
35	2.3 $t\bar{t}H$ Production	17
36	2.4 WZ + Heavy Flavor Production	20
37	2.5 Extensions to the Standard Model	21
38	III The LHC and the ATLAS Detector	26
39	3 The LHC	26
40	4 The ATLAS Detector	29
41	4.1 Inner Detector	30
42	4.2 Calorimeters	31
43	4.3 Muon Spectrometer	32
44	4.4 Trigger System	33
45	IV Measurement of WZ + Heavy Flavor	35
46	5 Introduction	35
47	6 Data and Monte Carlo Samples	36
48	6.1 Data Samples	36
49	6.2 Monte Carlo Samples	37
50	7 Object Reconstruction	38
51	7.1 Light leptons	38
52	7.2 Jets	39
53	7.3 B-tagged Jets	40
54	7.4 Missing transverse energy	41
55	7.5 Overlap removal	42

56	8 Event Selection and Signal Region Definitions	42
57	8.1 Event Preselection	43
58	8.2 Fit Regions	45
59	8.3 Non-Prompt Lepton Estimation	47
60	8.3.1 $t\bar{t}$ Validation	47
61	8.3.2 Z+jets Validation	48
62	9 tZ Separation Multivariate Analysis	49
63	10 Systematic Uncertainties	52
64	11 Results	56
65	11.1 1-jet Fit Results	57
66	11.2 2-jet Fit Results	66
67	V Differential Studies of $t\bar{t}H$ Multilepton	75
68	12 Data and Monte Carlo Samples	75
69	12.1 Data Samples	75
70	12.2 Monte Carlo Samples	75
71	13 Object Reconstruction	78
72	13.1 Trigger Requirements	78
73	13.2 Light Leptons	79
74	13.3 Jets	80
75	13.4 B-tagged Jets	80
76	13.5 Missing Transverse Energy	81
77	13.6 Overlap removal	81
78	14 Higgs Momentum Reconstruction	82
79	14.1 Physics Object Truth Matching	84
80	14.2 b-jet Identification	85
81	14.2.1 2lSS Channel	86
82	14.2.2 3l Channel	90
83	14.3 Higgs Reconstruction	93
84	14.3.1 2lSS Channel	94
85	14.3.2 3l Semi-leptonic Channel	98
86	14.3.3 3l Fully-leptonic Channel	102
87	14.4 p_T Prediction	106
88	14.4.1 2lSS Channel	107
89	14.4.2 3l Semi-leptonic Channel	110
90	14.4.3 3l Fully-leptonic Channel	112
91	14.5 3l Decay Mode	114

92	15 Signal Region Definitions	116
93	15.1 Pre-MVA Event Selection	116
94	15.2 Event MVA	120
95	15.3 Signal Region Definitions	127
96	16 Systematic Uncertainties	128
97	17 Results	132
98	17.1 Results - 79.8 fb^{-1}	133
99	17.2 Projected Results - 139 fb^{-1}	137
100	VI Conclusion	142
101	VII Appendices	150
102	Appendices	150
103	A Non-prompt lepton MVA	150
104	B Supplementary WZ + Heavy Flavor Studies	155
105	B.1 Non-prompt CR Modelling	155
106	B.2 tZ Interference Studies	160
107	B.3 Alternate tZ Inclusive Fit	164
108	B.3.1 tZ Inclusive Fit	165
109	B.3.2 Floating tZ	166
110	C Supplementary t̄tH Differential Analysis Studies	167
111	C.1 Higgs Reconstruction Model Details	167
112	C.1.1 b-jet Identification Features - 2ISS	167
113	C.1.2 b-jet Identification Features - 3l	171
114	C.1.3 Higgs Reconstruction Features - 2ISS	176
115	C.1.4 Higgs Reconstruction Features - 3IS	181
116	C.1.5 Higgs Reconstruction Features - 3IF	186
117	C.2 Background Rejection MVA Details	190
118	C.2.1 Background Rejection MVA Features - 2ISS	190
119	C.2.2 Background Rejection MVA Features - 3l	194
120	C.3 Truth Level Studies	199
121	C.4 Alternate b-jet Identification Algorithm	200
122	C.5 Binary Classification of the Higgs p_T	201
123	C.6 Impact of Alternative Jet Selection	203

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

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

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

154 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-
155 2018 provides the opportunity to make this measurement for the first time. The unprecedented
156 energy achieved by the LHC during this period greatly increase the rate at which $t\bar{t}H$ events are
157 produced, and the large amount of data collected provides the necessary statistics for a differential
158 measurement to be performed.

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

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

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

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

Part II**177 Theoretical Motivation****178 2 The Standard Model and the Higgs Boson**

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

186 2.1 The Forces and Particles of the Standard Model

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

Standard Model of Elementary Particles

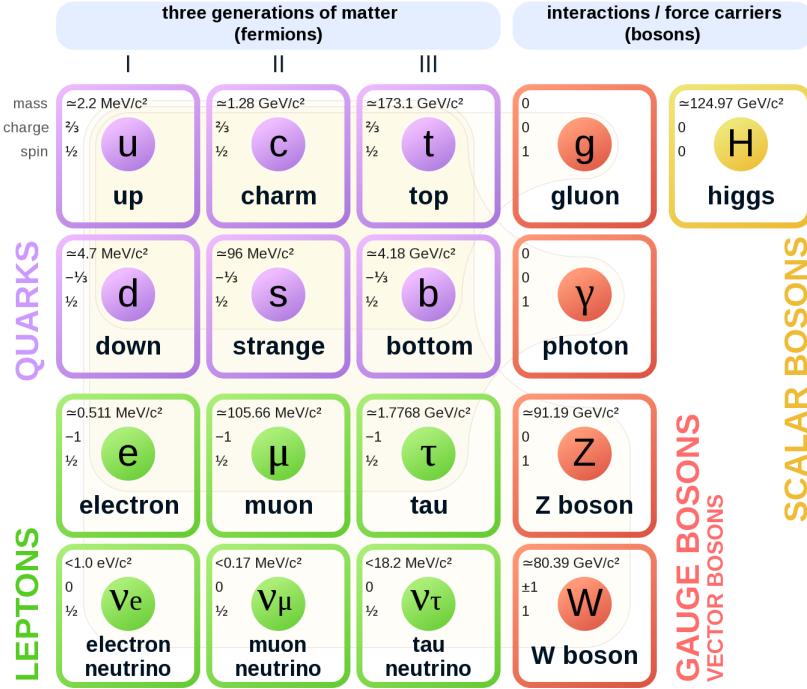


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. [2]

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

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

₁₉₇ Each of these generations form left-handed doublets invariant under SU(2) transfor-

₁₉₈ mations. For the leptons these doublets are:

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

₁₉₉ And for the quarks:

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

₂₀₀ For both leptons and quarks, the heavier generations can decay into the lighter generation

₂₀₁ of particles, while the first generation does not decay. Hence, ordinary matter generally consists

₂₀₂ of this first generation of fermions - electrons, up quarks, and down quarks. Each of these

₂₀₃ fermions has a corresponding anti-particle, which has an equal mass as its partner but opposite

₂₀₄ charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos,

₂₀₅ whose mass has been experimentally confirmed but is not accounted for in the SM.

₂₀₆ Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-

₂₀₇ exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1

₂₀₈ particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0,

₂₀₉ particle - the Higgs Boson. Of the gauge bosons, the W^+ , W^- and Z bosons - which are the

²¹⁰ mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon
²¹¹ mediates the electric force, and the gluon mediates the strong force.

²¹² **2.2 The Higgs Mechanism**

²¹³ A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to
²¹⁴ the Lagrangian giving mass to the the gauge bosons would violate this underlying symmetry of
²¹⁵ the theory. This presents a clear problem with the theory: The experimental observation that the
²¹⁶ W and Z bosons have mass seems to contradict the basic structure of the SM.

²¹⁷ Rather than abandoning gauge invariance, an alternative way for particles to acquire mass
²¹⁸ beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout
²¹⁹ in 1964 [3]. This procedure for introducing masses for the gauge bosons while preserving local
²²⁰ gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory
²²¹ by Weinberg in 1967 [4].

²²² **2.2.1 The Higgs Field**

²²³ The Higgs mechanism introduces a complex scalar SU(2) doublet, Φ , with the form:

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

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

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

225 Where μ and λ are free parameters of the new field. This represents the most general
226 potential allowed while preserving $SU(2)_L$ invariance and renormalizability. In the case that
227 $\mu^2 < 0$, this potential takes the form shown in Figure 2.2.

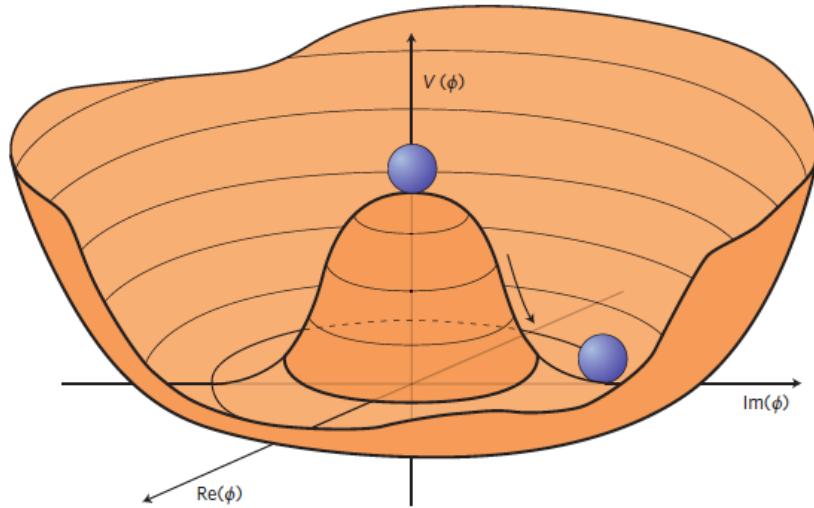


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ , for the case that $\mu^2 < 0$ [5].

228 The significant feature of this potential is that its minimum does not occur for a value of
229 $\Phi = 0$. Instead, it is minimized when $|\Phi^\dagger \Phi| = -\mu^2/\lambda$. This means that in its ground state, the
230 Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while
231 the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

²³² the minimum is determined only by $\Phi^\dagger \Phi$, there is some ambiguity in the particular definition of

²³³ the VEV, but it is generally represented as

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

²³⁴ The full value of Φ can be written as

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

²³⁵ with v being the value of the VEV, and H being the real value of the scalar field.

²³⁶ 2.2.2 Electroweak Symmetry Breaking

²³⁷ The Electroweak (EWK) interaction is described in the SM by a $SU(2)_L \otimes U(1)_Y$ gauge theory.

²³⁸ 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,

²³⁹ B_μ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar

²⁴⁰ field Φ in the Lagrangian:

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

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

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

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

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

248 And the masses of these fields are given by:

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

249 This produces exactly the particles we observe - three massive gauge bosons and a single
 250 massless photon. The massless photon represents the portion of the gauge symmetry, a single
 251 $U(1)$ of the electromagnetic force, that remains unbroken by the VEV.

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

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

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

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

261 The fermions' Yukawa couplings to the VEV take the same form as the fermions' coupling
 262 to the Higgs boson - λ_ψ . Therefore, the strength of a fermion's interaction with the Higgs is
 263 directly proportional to its mass. We now have a model that predicts a Higgs boson with mass
 264 $M_H = \mu$, which interacts with the fermions with coupling strength λ_ψ . Because μ and λ_ψ are

²⁶⁵ free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions

²⁶⁶ must be measured experimentally.

²⁶⁷ **2.3 $t\bar{t}H$ Production**

²⁶⁸ The strength of a particles interaction with the Higgs, given by its Yukawa coupling, is proportionate to its mass. The top quark - as the heaviest known particle - has the strongest interaction, making this interaction particularly interesting to study. While several processes involve interactions between the Higgs and the top, some Higgs production modes include the top interaction only as a part of a loop diagram, such as the gluon-gluon fusion diagram shown in Figure 2.3.

²⁷³ This process therefore only allows for an indirect probe of the Higgs-top Yukawa coupling, as
²⁷⁴ the flavor of the quark in this diagram is not unique.

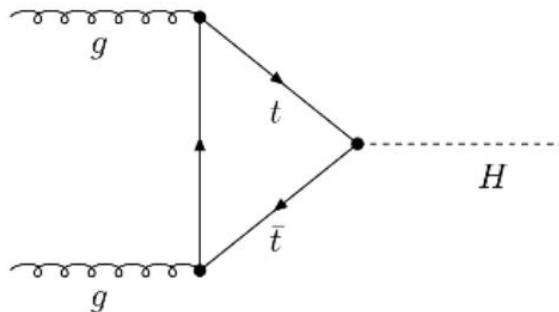


Figure 2.3: Diagram of a Higgs boson produced via gluon-gluon fusion.

²⁷⁵ Studying the Higgs produced in association with top quark pairs, $t\bar{t}H$, allows this interaction to be measured directly. This process, as shown in Figure 2.4, involves a unique coupling
²⁷⁷ between the Higgs and the top, which can be identified by the top quark pair in the final state.

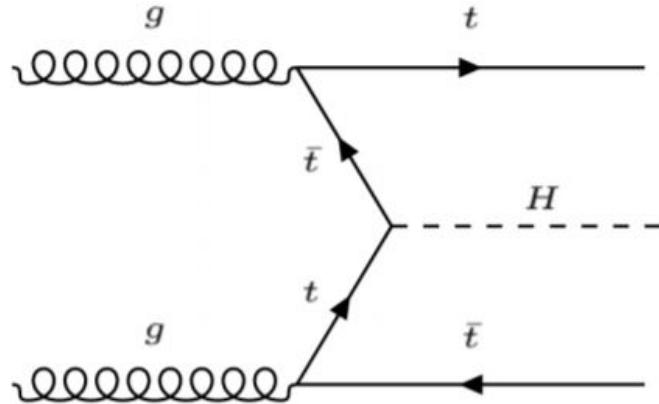


Figure 2.4: Diagram of a Higgs boson produced in association with a pair of top quarks.

278 The Higgs boson, as well as the top quarks, have very short lifetimes - on the order of
 279 10^{-22} s and 10^{-25} s respectively - meaning they can only be observed via their decay products.
 280 Measuring this process is therefore a matter of identifying events with final states consistent with
 281 $t\bar{t}H$ production.

282 Studies of $t\bar{t}H$ production have been reported by the ATLAS collaboration for $H \rightarrow b\bar{b}$,
 283 $H \rightarrow \gamma\gamma$ and multilepton (encompassing $H \rightarrow W^+W^-$, $H \rightarrow ZZ$ and $H \rightarrow \tau^-\tau^+$, with
 284 $H \rightarrow ZZ \rightarrow 4l$ as a separate analysis) decay modes. While the branching ratio of $H \rightarrow W^+W^-$
 285 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
 286 large $t\bar{t}$ backgrounds. On the other hand, $H \rightarrow \gamma\gamma$ produces the most easily identifiable signal,
 287 but has a much smaller branching ratio than $H \rightarrow W^+W^-$. Therefore, compared with other final
 288 states of $t\bar{t}H$, the $t\bar{t}H$ – ML channel is an attractive candidate for study, as it involves a good
 289 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.

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

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

293 $t\bar{t}H$ production was observed by ATLAS using up to 79.8 fb^{-1} of data collected at \sqrt{s}
 294 $= 13 \text{ TeV}$, based on a combination of five Higgs decay modes: $b\bar{b}$, WW^* , $\tau^-\tau^+$, $\gamma\gamma$, and ZZ^*
 295 [6]. A significance of 5.8σ was observed, compared to a 4.9σ expected significance. Since then,
 296 two analyses have published updated results ($H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4l$) with the full Run 2
 297 dataset, representing 139 fb^{-1} . Studies are still ongoing in the remaining channels.

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

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

302 **2.4 WZ + Heavy Flavor Production**

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

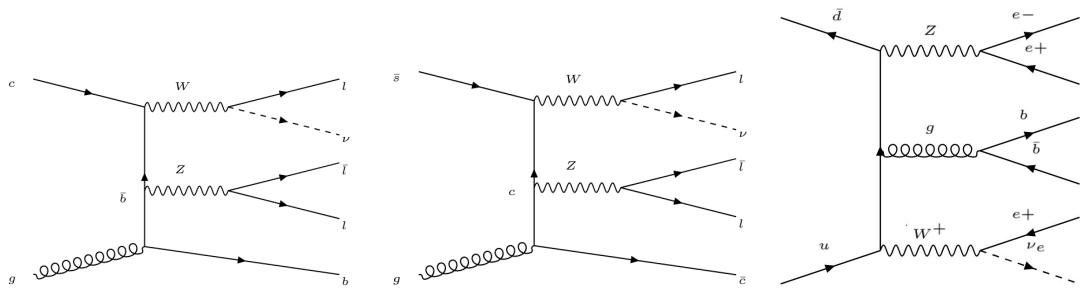


Figure 2.5: Example Feynman diagrams of WZ + heavy flavor production

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

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

315 **2.5 Extensions to the Standard Model**

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

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

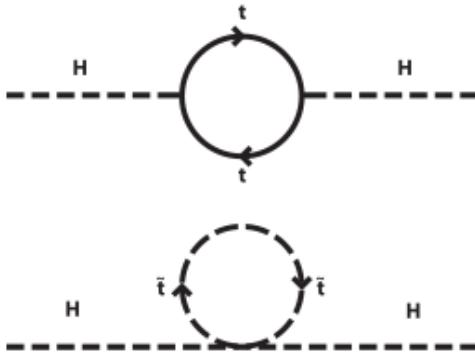


Figure 2.6: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram.

328 Because so many of the properties of the Higgs boson have not yet been studied, its
 329 interactions are a promising place to search for new physics that could resolve some of the
 330 limitations of the SM. As explained above, the interactions of the Higgs with the top quark, as in
 331 $t\bar{t}H$ production, are particularly interesting: As the most massive particle in the Standard Model,
 332 the top quark is the most strongly interacting with the Higgs. Therefore, any new physics effects
 333 are likely to be seen most prominently in this interaction.

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

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

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

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

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

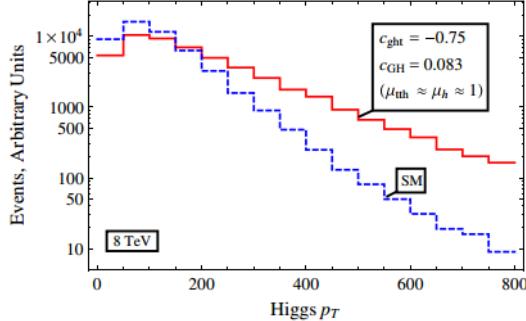


Figure 2.7: The momentum spectrum of the Higgs boson produced via top-quarks with (red) and without (blue) the presence of dimension-six operators.

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

356 Reconstructing the Higgs is a particular challenge in the multilepton channels of $t\bar{t}H$, due
 357 to an ambiguity arising from multiple sources of missing energy. In the $H \rightarrow \gamma\gamma$ channel, the
 358 kinematics of the Higgs can be fully reconstructed from the two photons. The same is true of
 359 $H \rightarrow b\bar{b}$, though with the additional challenge of identifying which two of the four b-quarks in
 360 the final state originated from the Higgs. By contrast, the two channels (2lSS and 3l) targeted
 361 by this analysis include at least one neutrino originating from the Higgs decay.

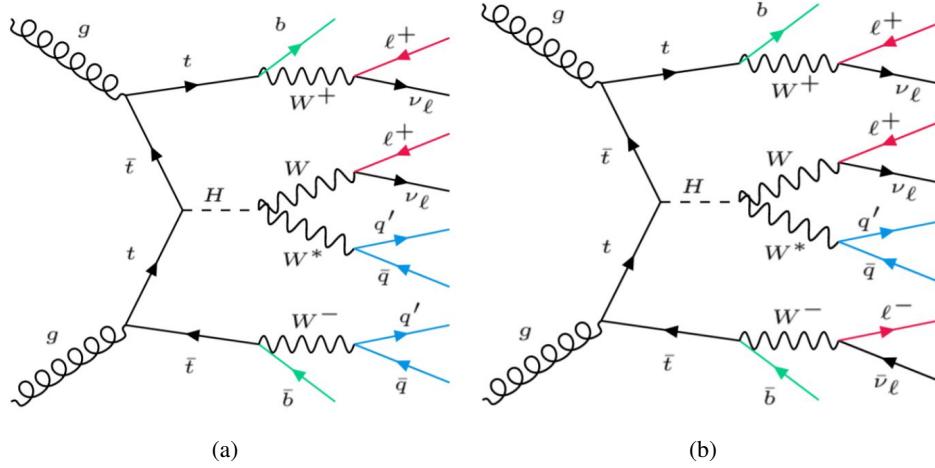


Figure 2.8: Feynman diagrams of $t\bar{t}H$ production with (a) two same-sign leptons and (b) three leptons in the final state.

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

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

Part III**371 The LHC and the ATLAS Detector****372 3 The LHC**

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

382 The LHC first began running in 2009 at a proton-proton center of mass energy of $\sqrt{s} = 8$
383 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during
384 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,
385 and collected data at an increased energy of $\sqrt{s} = 13$ TeV until 2018, a period referred to as Run
386 2.

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

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

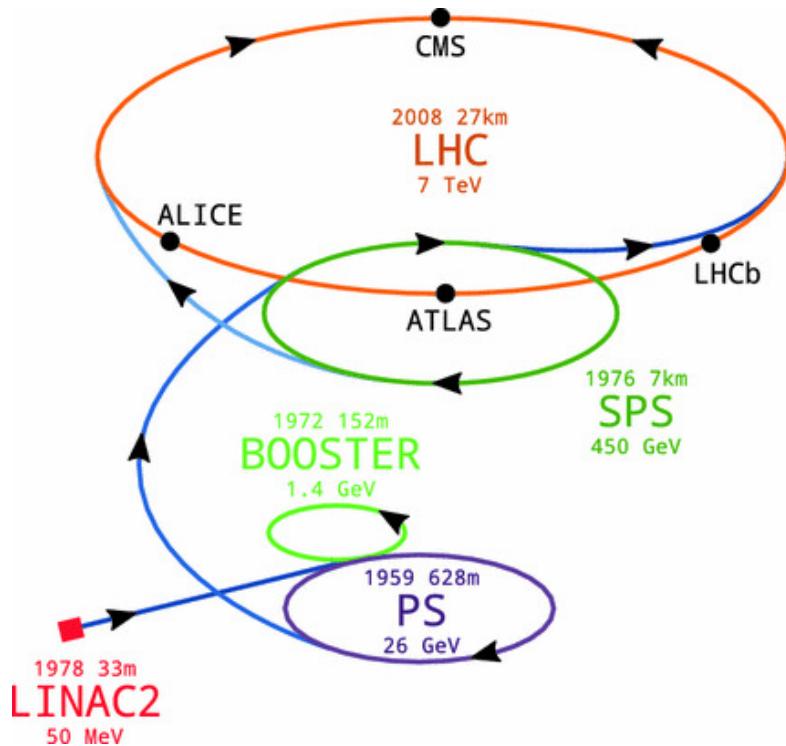


Figure 3.1: A summary of the accelerator chain used to feed protons into the LHC [11].

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

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

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

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

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

406 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around
 407 140 fb^{-1} , exceeding the expected integrated luminosity of 100 fb^{-1} .

408 4 The ATLAS Detector

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

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

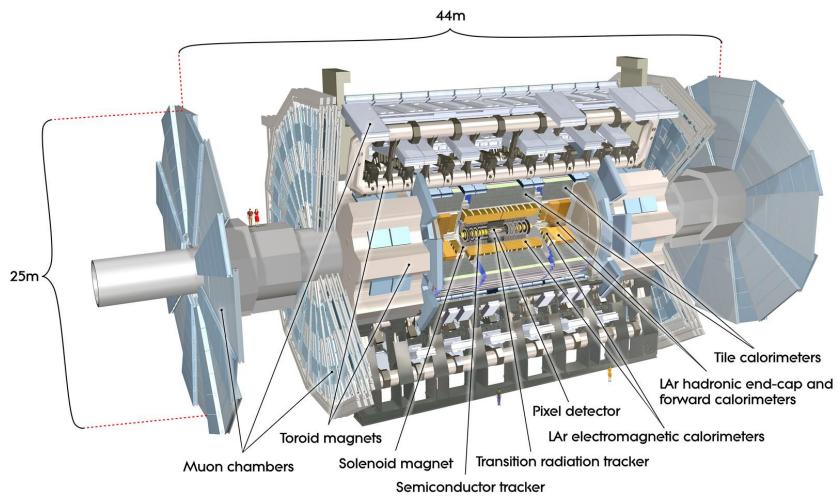


Figure 4.1: Cutaway view of the ATLAS detector, with labels of its major components [12].

417 **4.1 Inner Detector**

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

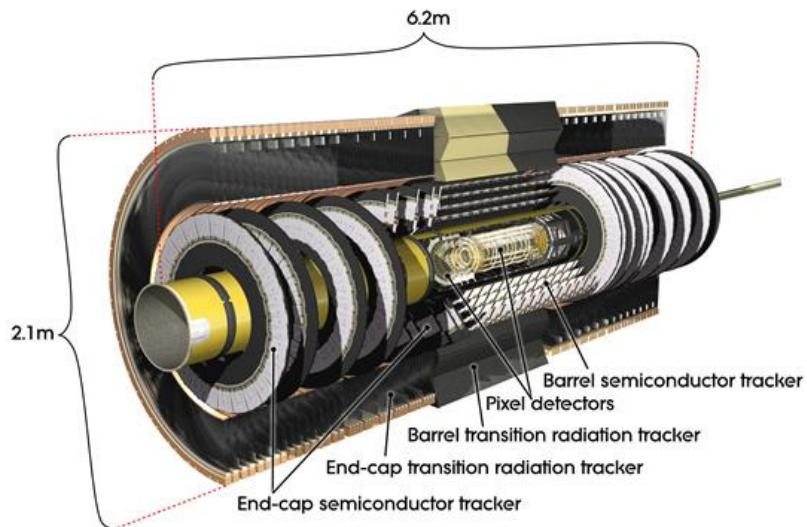


Figure 4.2: Cutaway view of the Inner Detector [13].

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

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

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

437 4.2 Calorimeters

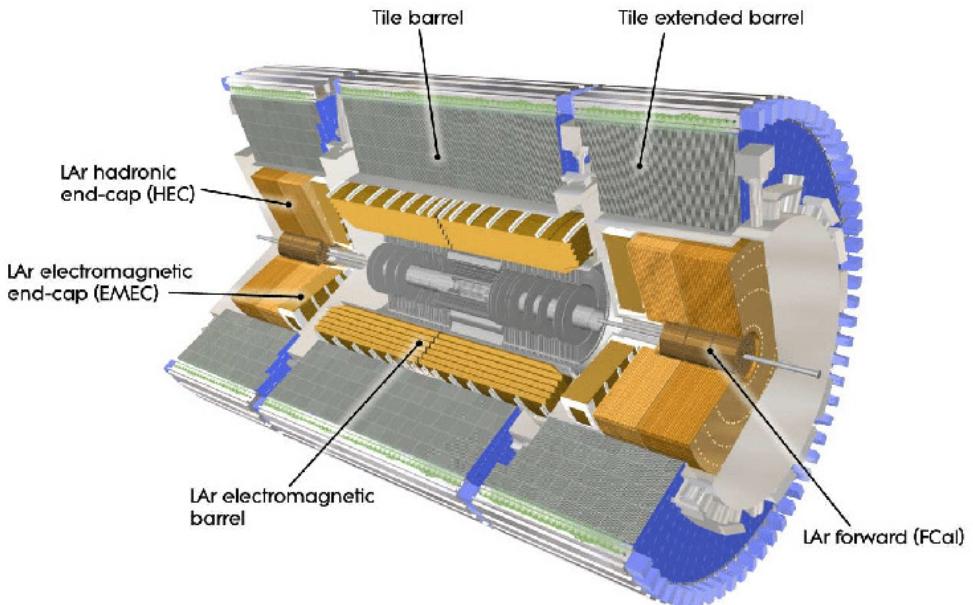


Figure 4.3: Cutaway view of the calorimeter system of the ATLAS detector [13].

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

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

450 **4.3 Muon Spectrometer**

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

455 The largest of the subdetectors, it extends from the outside of the calorimeter system, about a
456 4.25 m radius from the beam line, to a radius of 11 m. This large detector system is necessary

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

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

466 1200 tracking chambers are placed in the muon system in order to precisely measure the
467 tracks of muons with high spatial resolution. The path of the muons are tracked by Monitored
468 Drift Tubes, which are drift chambers formed by aluminum tubes and filled with ionizing gas.
469 These tubes produce a multi-layer spatial resolution on the order of 50 μm .

470 **4.4 Trigger System**

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

⁴⁷⁵ therefore a sophisticated trigger system is employed in real time to determine whether events are
⁴⁷⁶ sufficiently interesting to be worth storing.

⁴⁷⁷ The trigger system in ATLAS involves multiple levels, each of which select out which
⁴⁷⁸ events move on to the next level of scrutiny. The level-1 trigger uses hardware information from
⁴⁷⁹ the calorimeters and muon spectrometer to select events that contain candidates for particles
⁴⁸⁰ commonly used in analysis, such as energetic leptons and jets. The level-1 trigger reduces the
⁴⁸¹ rate of events from 40 MHz to around 100 kHz.

⁴⁸² Events that pass the level-1 trigger move to the High-Level Trigger (HLT). The HLT takes
⁴⁸³ place outside of the detector in software, and looks for properties such as a large amount of
⁴⁸⁴ missing transverse energy, well defined leptons, and multiple high energy jets. Events that pass
⁴⁸⁵ the HLT are stored and used for analysis. Because the specifics of the HLT are determined by
⁴⁸⁶ software rather than hardware, the thresholds can be changed throughout the run of the detector
⁴⁸⁷ in response to run conditions such as changes to pileup and luminosity. After the HLT is applied,
⁴⁸⁸ 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 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 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.

Motivated by its relevance to the $t\bar{t}H$ multilepton analysis, 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 this selection criteria are sorted into pseudo-continuous b-tagging regions based on the DL1r b-tag 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

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

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

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

519 6 Data and Monte Carlo Samples

520 6.1 Data Samples
521 This study uses a sample of proton-proton collision data collected by the ATLAS detector from
522 2015 through 2018 at an energy of $\sqrt{s} = 13 \text{ TeV}$, which represents an integrated luminosity of
523 139 fb^{-1} . This data set was collected with a bunch crossing rate of 25 ns. All data used in this
524 analysis was verified by data quality checks.

525 **6.2 Monte Carlo Samples**

526 Several different generators were used to produce Monte Carlo simulations of the signal and
 527 background processes. For all samples, the response of the ATLAS detector is simulated using
 528 Geant4. The WZ signal samples are simulated using Sherpa 2.2.2 [15]. Specific information
 529 about the Monte Carlo samples being used can be found in Table 16.

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

Process	Event generator	ME order	Parton Shower	PDF
WZ, VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
tZ	MG5_AMC	NLO	PYTHIA 8	CTEQ6L1
t̄tW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
t̄t(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄tH	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [16] (CT10 [17])
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t̄t, t̄t̄t̄t̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄t	POWHEG-BOX v2 [18]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄t̄γ	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	POWHEG-BOX v1 [19]	NLO	PYTHIA 6	CT10
qqVV, VVV Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

530 7 Object Reconstruction

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

534 All events are required to be selected by dilepton triggers. The p_T thresholds of the
 535 dilepton trigger on two electrons were 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017 and
 536 2018, while for the dimuon triggers the p_T thresholds on the leading (sub-leading) muon were
 537 18 GeV (8 GeV) in 2015, and 22 GeV (8 GeV) in 2016-2018. For the electron+muon triggers,
 538 the p_T thresholds on the electron (muon) were 17 GeV (14 GeV) for all datasets.

539 7.1 Light leptons

540 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter
 541 that are associated with charged particle tracks reconstructed in the inner detector [20]. Electron
 542 candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Muon candidates are
 543 reconstructed by combining inner detector tracks with track segments or full tracks in the muon
 544 spectrometer [21]. Muon candidates are required to have $p_T > 10$ GeV and $|\eta| < 2.5$. Candidates
 545 in the transition region between different electromagnetic calorimeter components, $1.37 <$
 546 $|\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant combining shower shape
 547 and track information is used to distinguish real electrons from hadronic showers (fake electrons).
 548 To further reduce the non-prompt electron contribution, the track is required to be consistent

549 with originating from the primary vertex; requirements are imposed on the transverse impact
 550 parameter 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 TightLH identification.

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

557 Leptons are additionally required to pass a non-prompt BDT selection developed by the
 558 $t\bar{t}H$ multilepton/ $t\bar{t}W$ analysis group. This BDT and the WPs used are summarized in Appendix
 559 A, and described in detail in [22]. Optimized working points and scale factors for this BDT are
 560 taken from that analysis.

561 7.2 Jets

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

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

570 **7.3 B-tagged Jets**

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

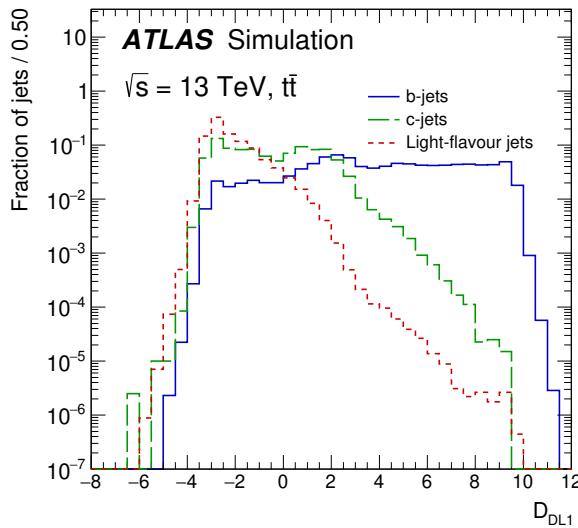


Figure 7.1: Output distribution of the DL1r algorithm for b-jets, charm jets, and light jets

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

578 this analysis are summarized in Table 3.

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

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

579 A tighter WP will accept fewer b-jets, but reject a higher fraction of charm and light jets.
 580 Generally, analyses that include b-jets will use a fixed working point, for example, requiring that
 581 a jet pass the 70% threshold. By instead treating these working point as bins, e.g. events with
 582 jets that fall between the 85% and 77% WPs fall into one bin, while events with jets passing the
 583 60% WP fall into another, and looking at the full psuedo-continuous DL1r spectrum of the jets,
 584 additional information can be gained. The psuedo-continuous b-tag spectrum is used in this case
 585 to separate out WZ + b, WZ + charm, and WZ + light.

586 7.4 Missing transverse energy

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

593 **7.5 Overlap removal**

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

600 This algorithm is applied to the preselected objects. The overlap removal procedure is
 601 summarized in Table 17.

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

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

602 **8 Event Selection and Signal Region Definitions**

603 Event are required to pass a preselection described in Section 8.1 and summarized in Table 5.
 604 Those that pass this preselection are divided into various fit regions described in Section 8.2,
 605 based on the number of jets in the event, and the b-tag score of those jets.

606 **8.1 Event Preselection**

607 Events are required to include exactly three reconstructed light leptons passing the requirement
 608 described in [7.1](#), which have a total charge of ± 1 .

609 The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose
 610 charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.
 611 the smallest ΔR , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton
 612 0 is required to have $p_T > 10$ GeV, while the same sign leptons, 1 and 2, are required to have
 613 $p_T > 20$ GeV to reduce the contribution of non-prompt leptons.

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

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

621 Events are required to have one or two reconstructed jets passing the selection described
 622 in [Section 7.2](#). Events with more than two jets are rejected in order to reduce the contribution of
 623 backgrounds such as $t\bar{t}Z$ and $t\bar{t}W$, which tend to have higher jet multiplicity.

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

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

624 The event yields in the preselection region for both data and Monte Carlo are summarized
 625 in Table 8.1, which shows good agreement between data and Monte Carlo, and demonstrates that
 626 this region consists primarily of WZ events. The WZ events are split into WZ + b, WZ + c, and
 627 WZ + l based on the truth flavor of the associated jet in the event. In this ordering b-jet supersedes
 628 charm, which supersedes light. That is, WZ + l events contain no charm and no b jets at truth
 629 level, WZ + c contain at least one truth charm and no b-jets, and WZ + b contains at least one
 630 truth b-jet.

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

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

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

633 8.2 Fit Regions

634 Once preselection has been applied, the remaining events are categorized into one of twelve
 635 orthogonal regions. The regions used in the fit are summarized in Table 7.

Table 7: A list of the regions used in the fit and the selection used for each.

Region	Selection
1j, <85%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_85} = 0$
1j, 85%-77%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_85} = 1, n\text{Jets_DL1r_77}=0$
1j, 77%-70%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_77} = 1, n\text{Jets_DL1r_70}=0$
1j, 70%-60%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_70} = 1, n\text{Jets_DL1r_60}=0$
1j, >60%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_60} = 1, tZ \text{ BDT} > 0.725$
1j tZ CR	$N_{\text{jets}} = 1, n\text{Jets_DL1r_60} = 1, tZ \text{ BDT} < 0.725$
2j, <85%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} = 0$
2j, 85%-77%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} >= 1, n\text{Jets_DL1r_77}=0$
2j, 77%-70%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_77} >= 1, n\text{Jets_DL1r_70}=0$
2j, 70%-60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_70} >= 1, n\text{Jets_DL1r_60}=0$
2j, >60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} >= 1, tZ \text{ BDT} > 0.725$
2j tZ CR	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} >= 1, tZ \text{ BDT} < 0.725$

636 The working points discussed in Section 7.3 are used to separate events into fit regions
 637 based on the highest working point reached by a jet in each event. Because the background
 638 composition differs significantly based on the number of b-jets, events are further subdivided
 639 into 1-jet and 2-jet regions in order to minimize the impact of background uncertainties.

640 An unfolding procedure is performed to account for differences in the number of recon-
 641 structed jets compared to the number of truth jets in each event. In order to account for migration
 642 of WZ+1-jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal
 643 samples are separated based on the number of truth jets. Events with 0 jets or more than 3 jets at
 644 truth level, yet fall within one of the categories listed in Table 7, are categorized as WZ + other,
 645 and treated as a background. The composition of the number of truth jets in each reco jet bin is
 646 taken from MC, with uncertainties in these estimates described in detail in Section 16.

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

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

652 **8.3 Non-Prompt Lepton Estimation**

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

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

661 **8.3.1 $t\bar{t}$ Validation**

662 $t\bar{t}$ events can produce two prompt leptons from the decay of each of the tops. These top decays
663 produce two b-quarks, the decay of which can produce additional non-prompt leptons, which
664 occasionally pass the event preselection. In order to validate that the Monte Carlo accurately

665 simulates this process accurately, the MC prediction in a non-prompt $t\bar{t}$ enriched control region

666 is compared to data.

667 The $t\bar{t}$ control region is similar to the preselection region - three leptons meeting the

668 criteria described in Section 8 are required, and the requirements on E_T^{miss} remain the same.

669 However, the selection requiring a lepton pair form a Z-candidate are reversed. Events where the

670 invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of 91.2 GeV are

671 rejected. This ensures the $t\bar{t}$ control region is orthogonal to the preselection region.

672 Further, because the jet multiplicity of $t\bar{t}$ events tends to be higher than WZ, the number

673 of jets in each event is required to be greater than 1. As b-jets are almost invariably produced

674 from top decays, at least one b-tagged jet passing the 70% DL1r WP in each event is required.

675 Data is compared to MC predictions in the region for a variety of kinematic variable, as well

676 as various b-tag WPs. A constant normalization discrepancy between data and MC predictions

677 of approximately 10% is found, which is accounted for by applying a constant correction factor

678 of 0.9 to the $t\bar{t}$ MC prediction. As data and MC are found to agree within 20% for each of

679 the b-tag WPs considered, a 20% systematic uncertainty on the $t\bar{t}$ prediction is included for the

680 analysis.

681 **8.3.2 Z+jets Validation**

682 Similar to $t\bar{t}$, a non-prompt Z+jets control region is produced in order to validate the MC

683 predictions. The lepton requirements remain the same as the preselection region. Because no

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

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

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

695 9 tZ Separation Multivariate Analysis

696 An important process to consider in this analysis is tZ: the top almost always decays into a W
 697 boson and b-quark, and when both the W and Z decay leptonically, this gives three leptons and
 698 a heavy flavor jet in the final state. Because tZ can produce a final state identical to the signal,
 699 it represents a predominant background in the most signal enriched regions. That is, the region
 700 with one jet passing the 60% DL1r WP. Therefore, a boosted decision tree (BDT) algorithm is
 701 trained using XGBoost [27] to separate WZ + heavy flavor from tZ. The result of this BDT is

702 used to create a tZ enriched region in the fit, reducing its impact on the measurement of WZ +
703 heavy flavor.

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

705 • The invariant mass of the reconstructed top candidate

706 • p_T of each of the leptons, jet

707 • The invariant mass of each combination of lepton pairs, $M(l\bar{l})$

708 • E_T^{miss}

709 • Distance between each combination of leptons, $\Delta R(l\bar{l})$

710 • Distance between each lepton and the jet, $\Delta R(lj)$

711 Here the top candidate is reconstructed based on the procedure described in section 6.1 of
712 [28]. Broadly, the mass of the top quark candidate is reconstructed from the jet, the lepton not
713 included in the Z-candidate, and a reconstructed neutrino. In the case that there is one jet in the
714 event, there is only possible b-jet candidate. For events with two jets, the jet with the highest
715 DL1r score is used.

716 The training samples included only events meeting the requirements of the 1-jet, >60%
717 region, i.e. passing all the selection described in section 8 and having exactly one jet which passes
718 the tightest (60%) DL1r working point. A sample of 20,000 background (tZ) and signal (WZ+b)

719 Monte Carlo events are used to train the BDT. And additional 5,000 events are reserved for testing
 720 the model, in order to prevent over-fitting. A total of 750 decision trees with a maximum depth
 721 of 6 branches are used to build the model. These parameters are chosen empirically, by training
 722 several models with different parameters and selecting the one that gave the best separation for
 723 the test sample.

724 The results of the BDT training are shown in figure 9.1. The output scores for both signal
 725 and background events is shown on the left. The right shows the receiving operating characteristic
 726 (ROC) curve that results from the MVA. The ROC curve represents the background rejection
 727 as a function of signal efficiency, where each point on the curve represents a different response
 728 score. The ROC curve of the BDT is compared to the performance of using an optimal set of flat
 729 selections on the same set of input variables.

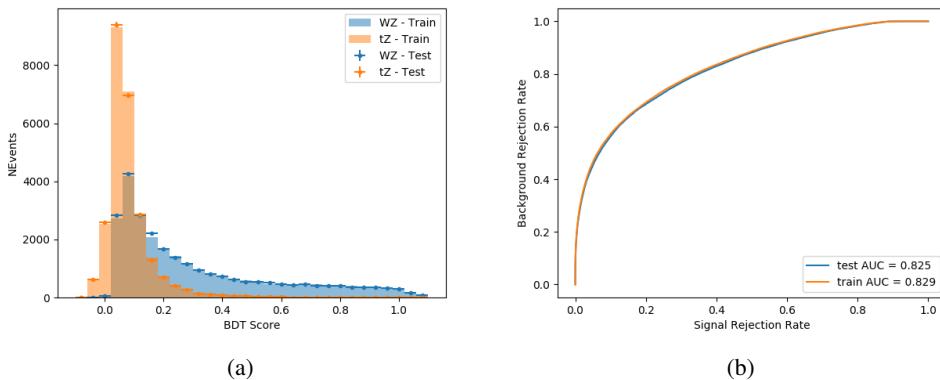


Figure 9.1: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

730 The relative important of each input feature in the model, measured by how often they
 731 appeared in the decision trees, is shown in figure 9.2.

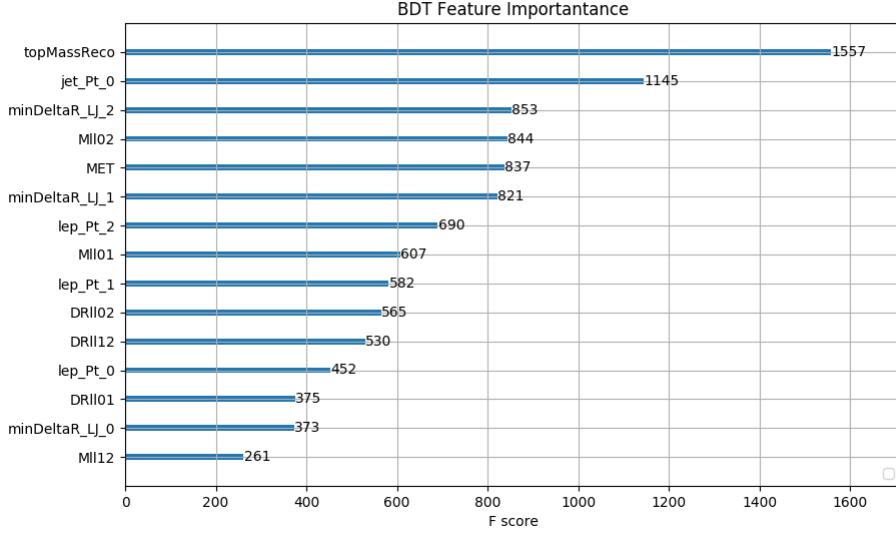


Figure 9.2: Relative importance of each input feature in the model.

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

738 10 Systematic Uncertainties

739 The systematic uncertainties that are considered are summarized in Table 37. These are imple-
 740 mented in the fit either as a normalization factors or as a shape variation or both in the signal

⁷⁴¹ and background estimations. The numerical impact of each of these uncertainties is outlined in

⁷⁴² Section 17.

Table 8: 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

⁷⁴³ The uncertainty in the combined integrated luminosity is derived from a calibration of the

⁷⁴⁴ luminosity scale performed for 13 TeV proton-proton collisions [29], [30].

⁷⁴⁵ The experimental uncertainties are related to the reconstruction and identification of light

⁷⁴⁶ leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The sources which contribute to

⁷⁴⁷ the uncertainty in the jet energy scale (JES) [31] are decomposed into uncorrelated components

⁷⁴⁸ and treated as independent sources of uncertainty in the analysis. These are treated as 30

749 nuisance parameters included in the fit. A similar approach is used for the jet energy resolution
750 (JER) uncertainty, which is decomposed into 8 JER uncertainty components included as NPs in
751 the fit.

752 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
753 [32] are also decomposed into uncorrelated components. The large number of components for
754 b-tagging is due to the calibration of the distribution of the MVA discriminant.

755 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and scale
756 uncertainties are taken from theory calculations, with the exception of non-prompt and diboson
757 backgrounds. The cross-section uncertainty on tZ is taken from [33]. Derivation of the non-
758 prompt background uncertainties, Z+jets and tt>, are explained in Section 8.3. These normaliz-
759 ation uncertainties are chosen so as to account for the complete uncertainty in the non-prompt
760 contribution, and therefore no additional modelling uncertainties are considered for Z+jets and
761 tt>.

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

767 The theory uncertainties applied to the predominate background estimates are summarized

768 in Table 39.

Process	X-section [%]
tZ	X-sec: ± 15.2 QCD Scale: $^{+5.2}_{-1.3}$ PDF($+\alpha_S$): ± 1.2
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 9: Summary of theoretical uncertainties for MC predictions in the analysis.

769 Additional signal uncertainties are estimated by comparing estimates from the nominal Sherpa

770 WZ samples with alternate WZ samples generated with Powheg+PYTHIA8. Separate systemati-

771 cs are included in the fit for WZ + b, WZ + charm and WZ + light, where the distribution among

772 each of the fit regions is varied based on the prediction of the Powheg sample.

773 A similar approach is taken to account for uncertainties in migrations between the number of

774 reco and truth jets. The fraction of events with 1 truth jet which fall into the 1 jet bin versus the

775 2 jet bin at reco level is compared for Sherpa and Powheg. The same is done for events with 2

776 truth jets. A systematic is included where events are shifted between the 1-jet and 2-jet regions

777 based on the differences between these two shapes. This is done independently for each of the
778 WZ + b, WZ + charm, and WZ + light templates.

779 Additional systematics are included to account for the uncertainty in the contamination of
780 0 jet and 3 or more jet events (as defined at truth level) in the 1 and 2 reco jet bins. Because these
781 events fall outside the scope of this measurement, these events are included as a background.
782 As such, a normalization, rather than a shape, uncertainty is applied for this background. The
783 number of WZ events with 0-jets and $>=3$ -jets in the reconstructed 1-jet and 2-jet regions are
784 compared for Sherpa and Powheg, and these differences are taken as separate normalization
785 systematics on the yield of WZ+0-jet and WZ+ $>=3$ -jet events.

786 11 Results

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

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

799 The Asimov fit for 1-jet events gives an expected μ value of $1.00^{+0.37}_{-0.35}(\text{stat})^{+0.24}_{-0.22}(\text{sys})$ for
 800 $WZ + b$. The fitted cross-section modifiers for $WZ + \text{charm}$ and $WZ + \text{light}$ are $1.00 \pm 0.17 \pm 0.15$
 801 and $1.00 \pm 0.04 \pm 0.07$, respectively.

802 The expected cross-section of $WZ+b$ with 1 associated jet obtained from the fit is
 803 $1.74^{+0.65}_{-0.61}(\text{stat})^{+0.42}_{-0.37}(\text{sys}) \text{ fb}$ with an expected significance of 2.2σ . The expected cross-section
 804 of $WZ + \text{charm}$ is measured to be $14.6 \pm 2.5(\text{stat}) \pm 2.1(\text{sys}) \text{ fb}$, with a correlation of -0.23.

805 For 2-jet events, the fit gives an expected μ value of $1.00^{+0.34}_{-0.33}(\text{stat})^{+0.29}_{-0.28}(\text{sys})$ for $WZ +$
 806 b . The fitted cross-section modifiers for $WZ + \text{charm}$ and $WZ + \text{light}$ are $1.00 \pm 0.17 \pm 0.15$ and
 807 $1.00 \pm 0.04 \pm 0.08$, respectively.

808 The expected $WZ + b$ cross-section in the 2-jet region is $2.46^{+0.83}_{-0.81}(\text{stat})^{+0.073}_{-0.68}(\text{sys}) \text{ fb}$
 809 with an expected significance of 2.6σ . The 2-jet expected cross-section of $WZ + \text{charm}$ is
 810 $12.7 \pm 2.2(\text{stat}) \pm 2.0(\text{sys}) \text{ fb}$, and the correlation between $WZ + \text{charm}$ and $WZ + b$ is -0.26.

811 **11.1 1-jet Fit Results**

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

⁸¹³ The pre-fit yields in each of the regions used in the fit are shown in Table 11.1, and

⁸¹⁴ summarized in Figure 11.1.

⁸¹⁵

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

Table 10: Pre-fit yields in each of the 1-jet fit regions.

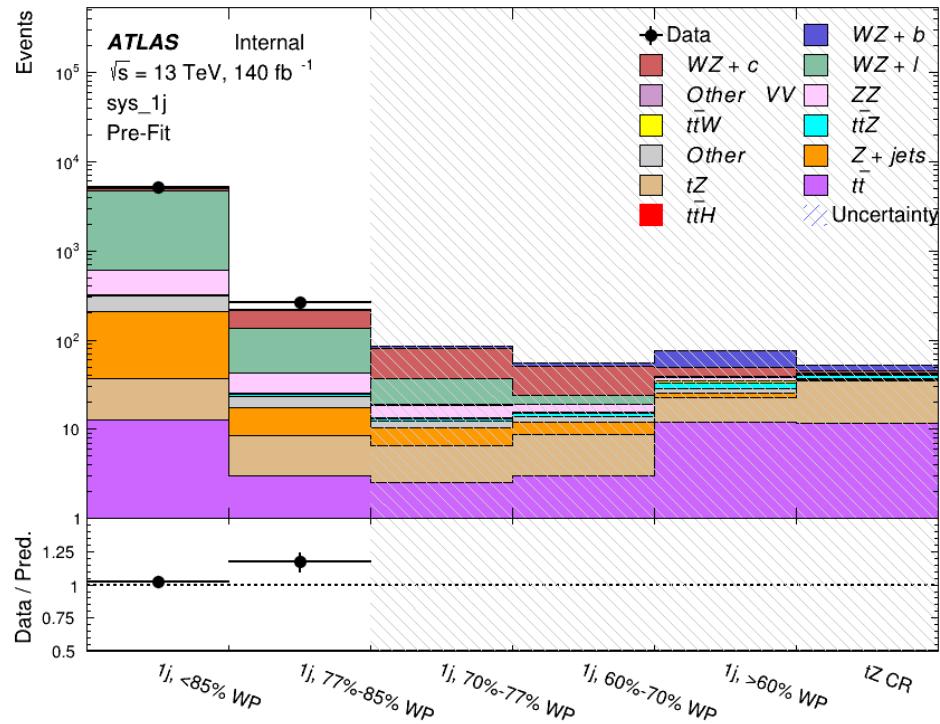


Figure 11.1: Pre-fit summary of the 1-jet fit regions.

816

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

817

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

Table 11: Post-fit yields in each of the 1-jet fit regions.

818

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

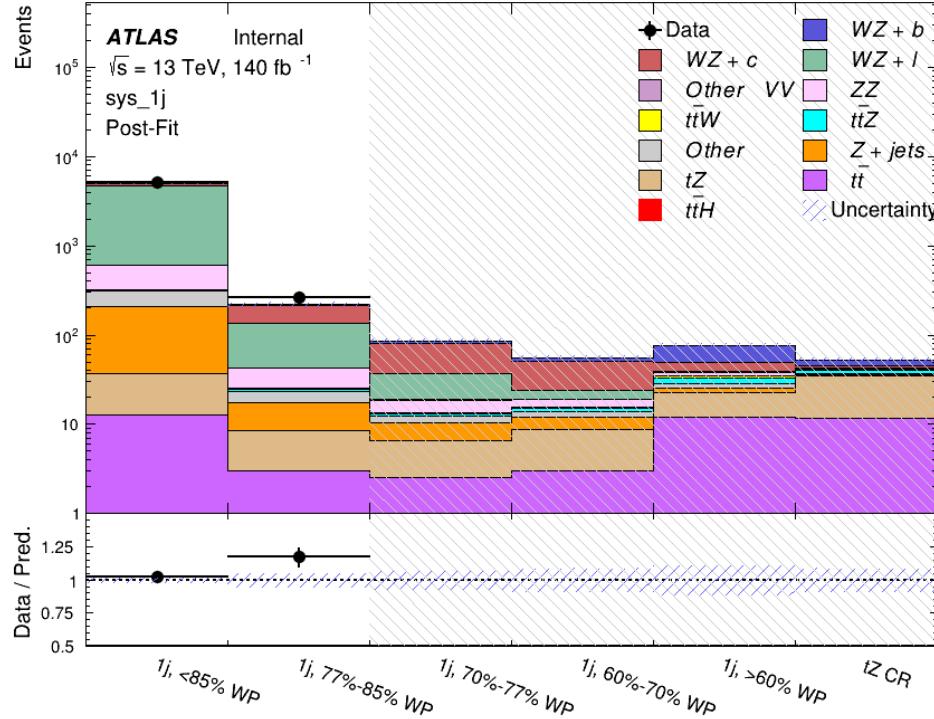


Figure 11.2: Post-fit summary of the 1-jet fit regions.

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

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

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

828 [12](#).

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

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

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

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

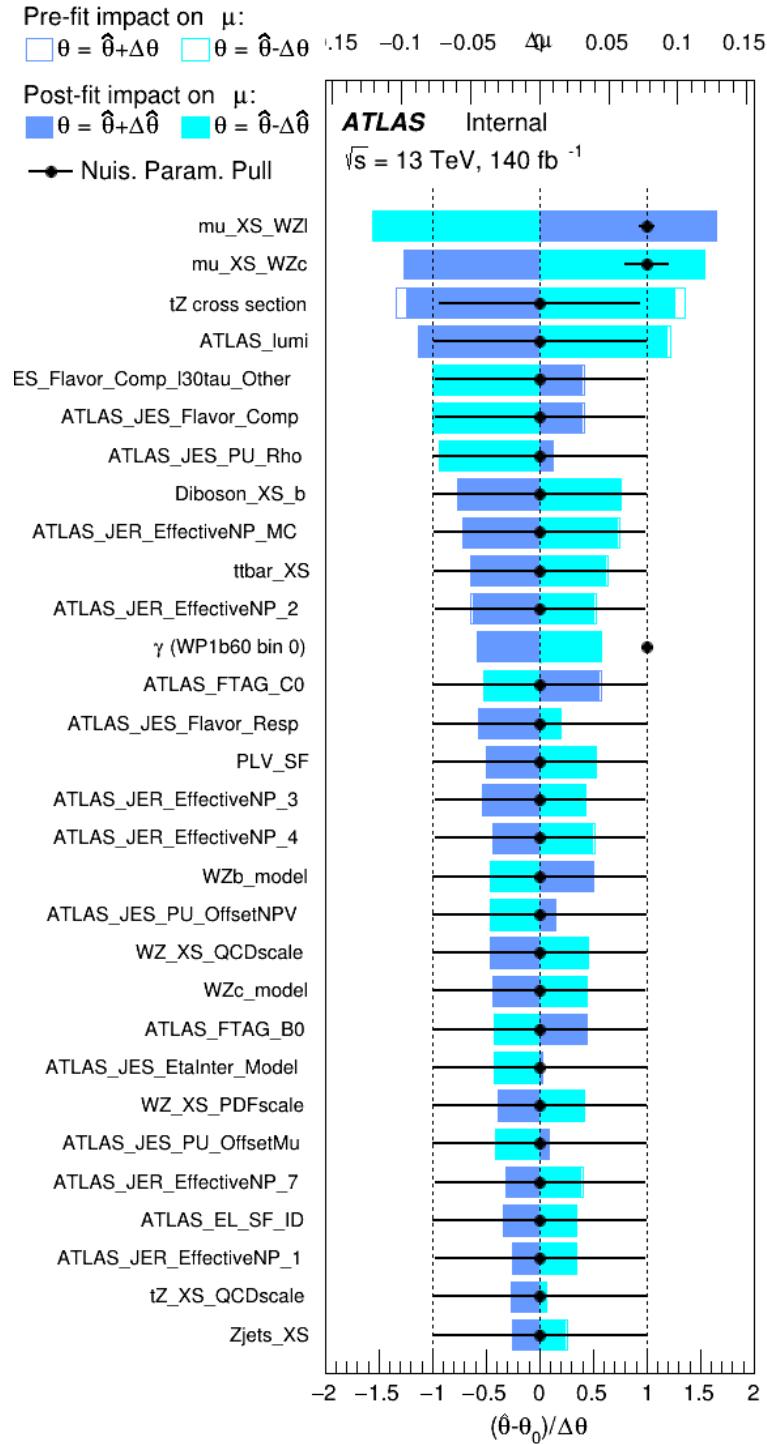


Figure 11.3: Impact of systematic uncertainties on the signal-strength of $WZ + b$ for events with exactly one jet

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

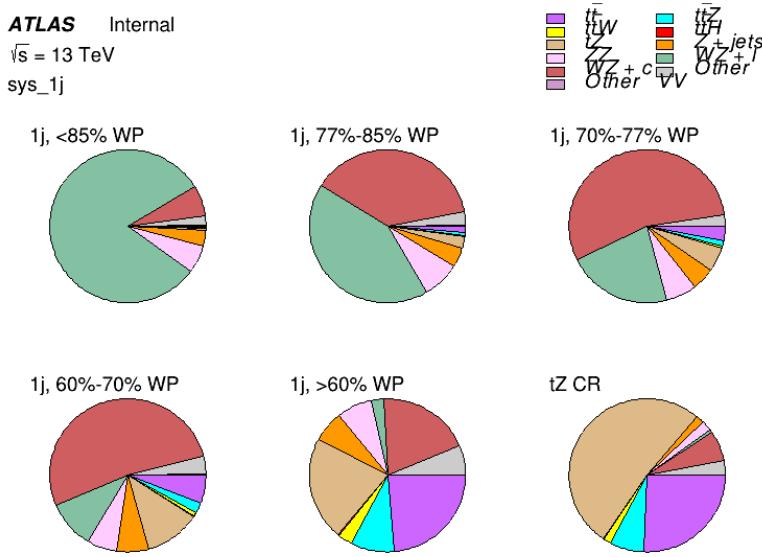


Figure 11.4: Post-fit background composition of the fit regions.

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

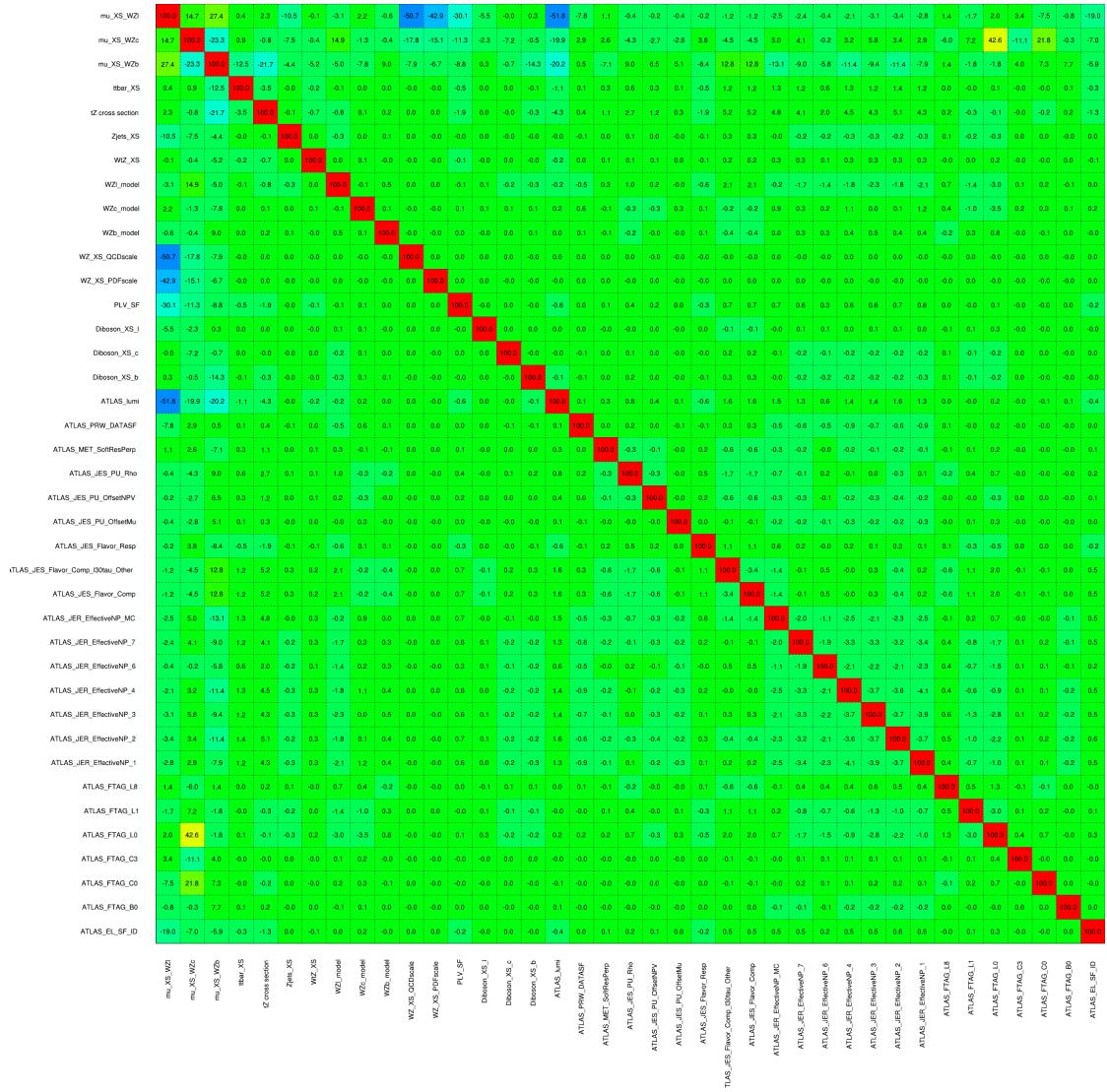


Figure 11.5: Correlations between nuisance parameters

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

841 Two of the major backgrounds in the region with the highest purity of WZ + b are tZ and
842 Other VV + b, explaining the negative correlations between μ_{WZ+b} and the tZ cross section, and
843 the VV + b cross section.

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

848 **11.2 2-jet Fit Results**

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

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

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

Table 13: Pre-fit yields in each of the 2-jet fit regions.

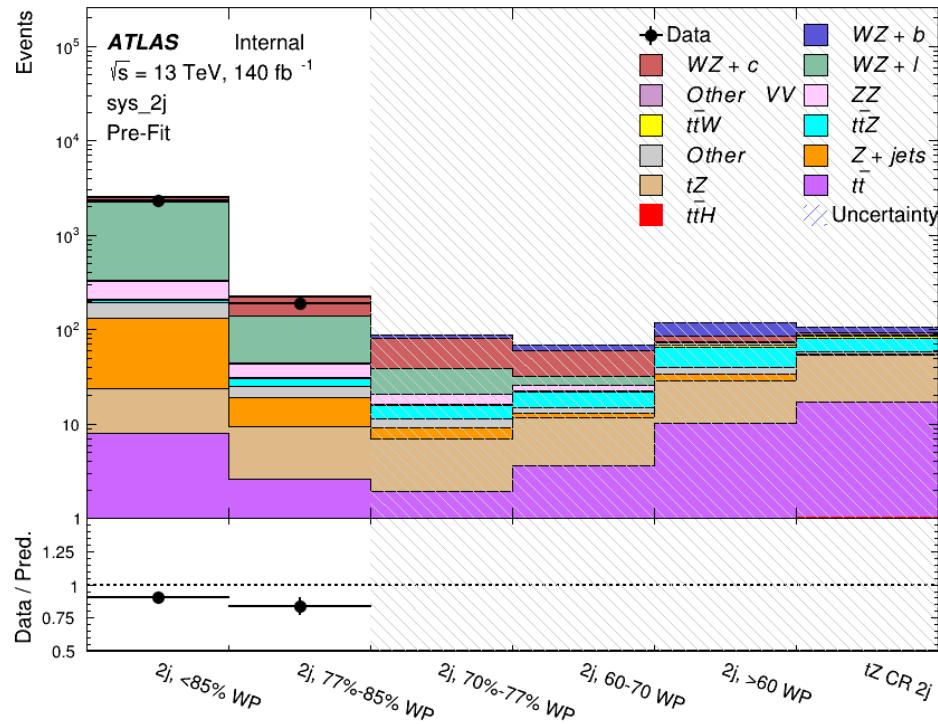


Figure 11.6: Pre-fit summary of the 2-jet fit regions.

851

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

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

Table 14: Post-fit yields in each of the 2-jet fit regions.

852

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

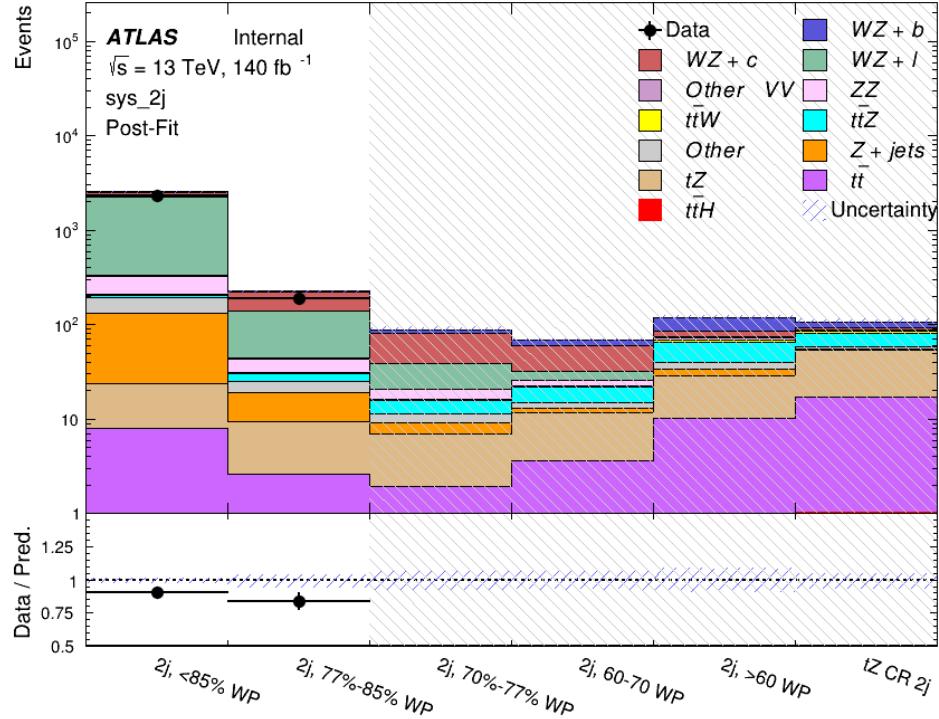


Figure 11.7: Post-fit summary of the fit over 2-jet regions.

853 The same set of systematic uncertainties consider for the 1-jet fit are included in the 2-jet
 854 fit as well. The impact of the most significant systematic uncertainties is summarized in Table
 855 15.

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

Table 15: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b 2-jet events.

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

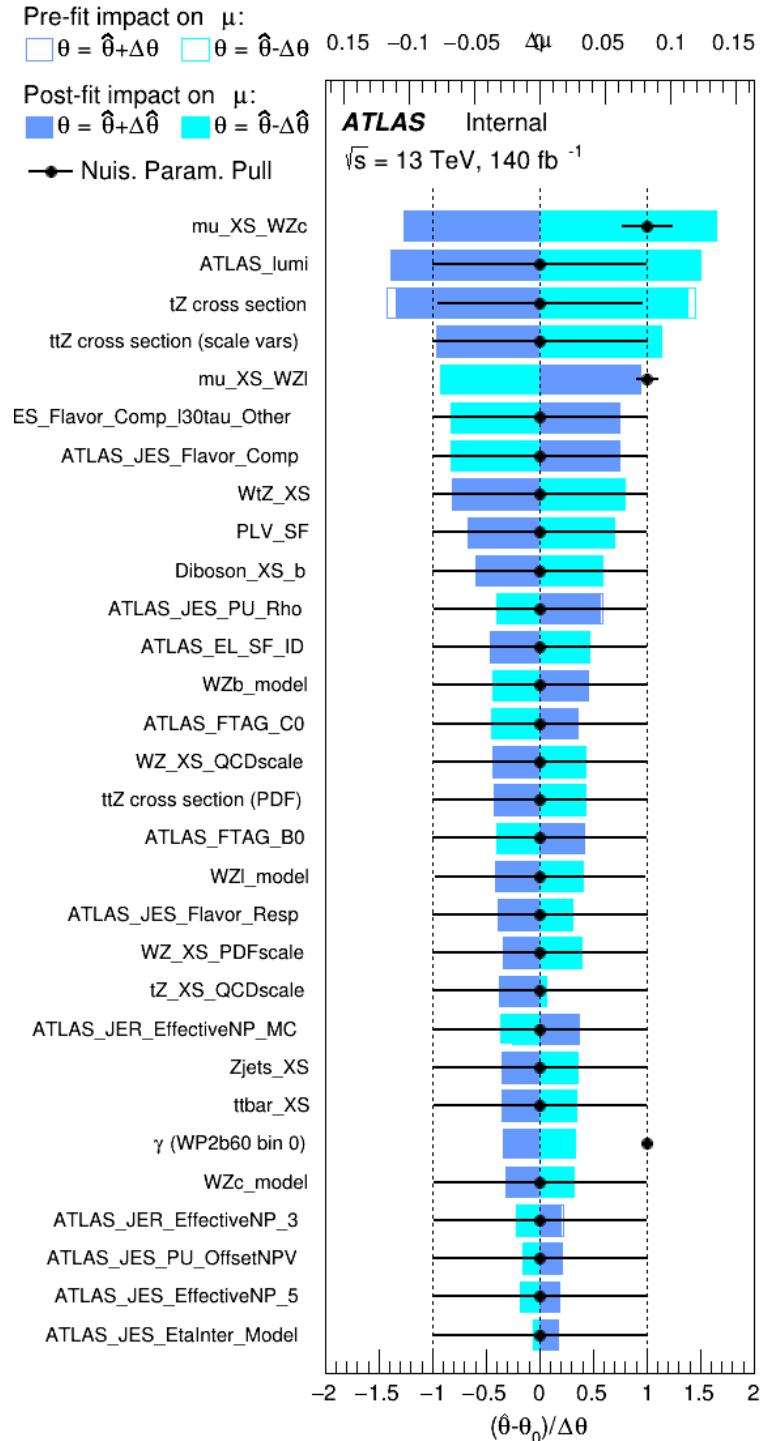


Figure 11.8: Impact of systematic uncertainties on the signal-strength of $WZ + b$ in 2-jet events.

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

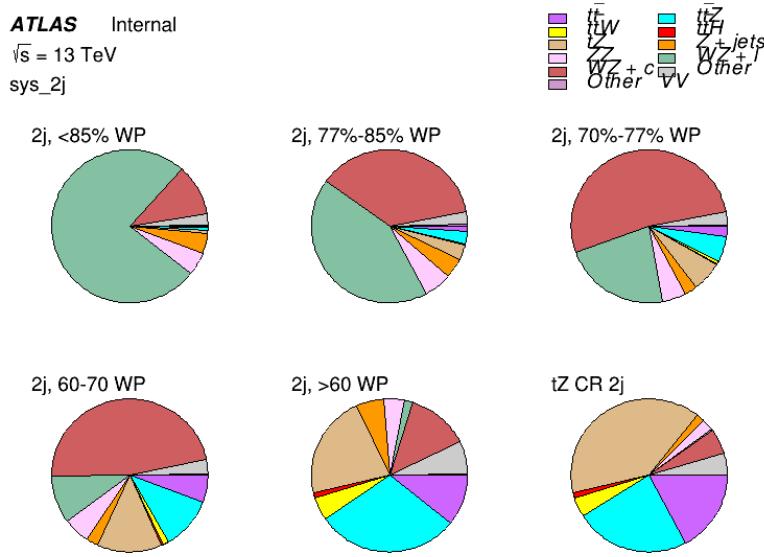


Figure 11.9: Post-fit background composition of the 2-jet fit regions.

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

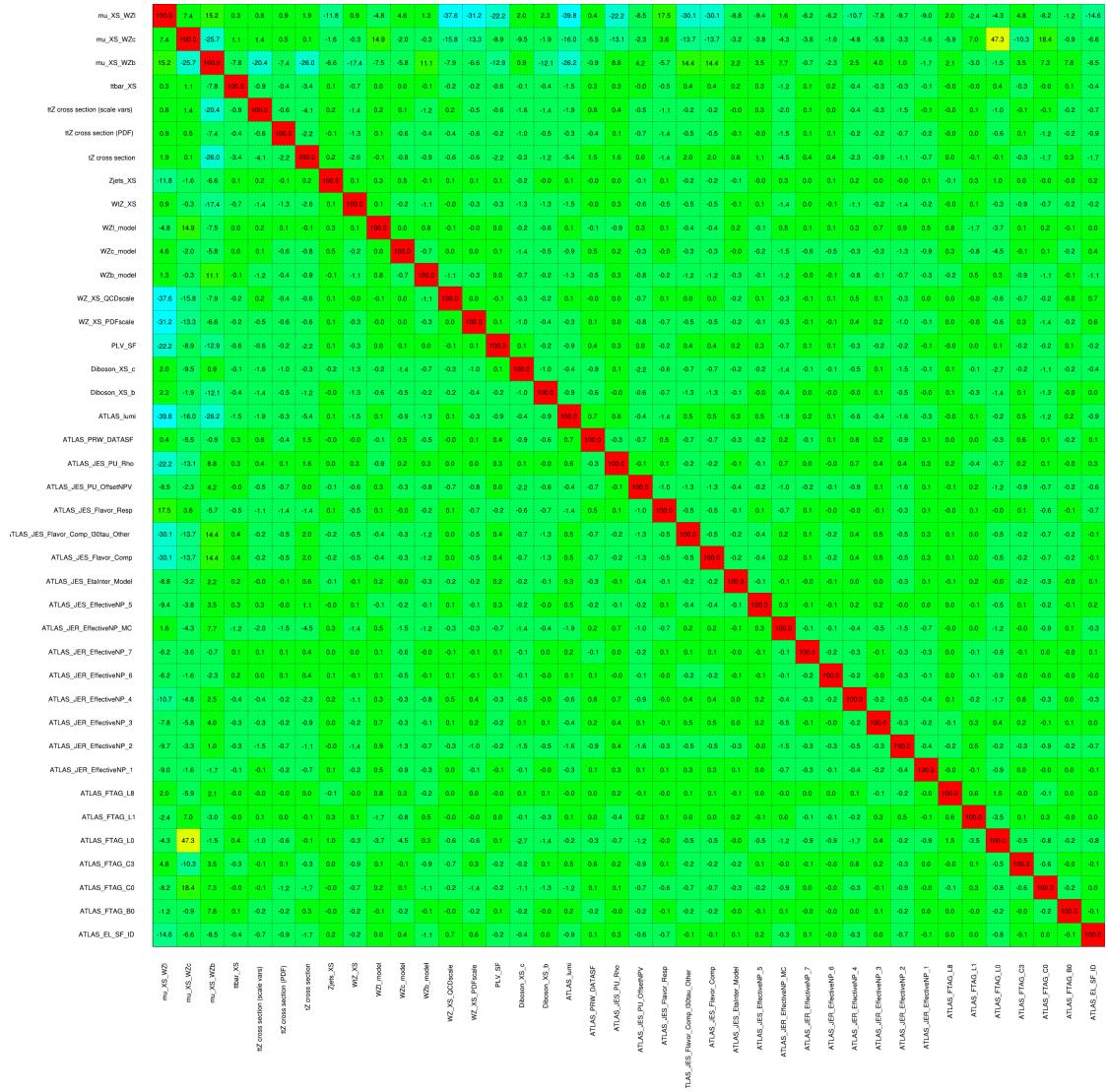


Figure 11.10: Correlations between nuisance parameters in the 2-jet fit

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

866 Part V**867 Differential Studies of $t\bar{t}H$ Multilepton****868 12 Data and Monte Carlo Samples****869 12.1 Data Samples**

870 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through
871 2018, which represents an integrated luminosity of 139 fb^{-1} and an energy of $\sqrt{s} = 13 \text{ TeV}$. All
872 data used in this analysis was included in one of the Good Run Lists verified by Data Quality
873 checks.

874 12.2 Monte Carlo Samples

875 Several Monte Carlo (MC) generators were used to simulate both signal and background pro-
876 cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific
877 event generator used for each of these MC samples is listed in Table 16. A Higgs mass of 125
878 GeV is assumed in all simulations.

Table 16: 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
t̄H	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [16] (CT10 [17])
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
VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
t̄t	POWHEG-BOX v2 [18]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄γ	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
tZ	MG5_AMC	LO	PYTHIA 6	CTEQ6L1
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t̄, t̄t̄t̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄W+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top qqVV, VVV Z → l+l-	POWHEG-BOX v1 [19]	NLO	PYTHIA 6	CT10
	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

879 The signal sample ($t\bar{t}H$) is modelled at NLO with POWHEG-BOX v2 using the NNPDF2.0
 880 parton distribution function (PDF) [34]. Parton showering and hadronisation were modelled
 881 with PYTHIA 8.2 [35]. The $t\bar{t}H$ sample is normalized to a cross-section of 507^{+35}_{-50} fb based on
 882 NLO calculations. Uncertainties are based on varying the QCD factorisation and renormalisation
 883 scale, as well as uncertainties in the PDF and α_s .

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

887 with what is observed in data. While an effort is currently being undertaken to measure $t\bar{t}W$ more
 888 accurately, the approach used by the 79.8 fb^{-1} $t\bar{t}H$ analysis is used here: A normalization
 889 factor of 1.68 is applied to the MC estimate of $t\bar{t}W$ and additional systematic uncertainties on
 890 $t\bar{t}W$ are included to account for this modelling discrepancy, as outlined in Section 16.

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

894 The “fake”, or non-prompt, background comes primarily from leptons originating from
 895 hadron decays, leptons with missidentified charge, and photon conversions. While the main
 896 $t\bar{t}H$ analysis is currently refining a data-driven approach for estimating the contribution of
 897 events with non-prompt leptons, at the time of this note this strategy has not been completely
 898 developed for the full Run-2 dataset. Therefore, the non-prompt contribution is estimated with
 899 MC, while applying normalization corrections and systematic uncertainties derived from data
 900 driven techniques developed for the 79.8 fb^{-1} $t\bar{t}H/t\bar{t}W$ analysis [7].

901 The primary contribution to the non-prompt lepton background is from $t\bar{t}$ production, with
 902 $V+jets$ and single-top as much smaller sources. Estimation of this background is done primarily
 903 using an inclusive $t\bar{t}$ sample, with corrections applied based on data driven methods. This sample
 904 is generated using **POWHEG**, with **PYTHIA8** performing the parton shower and fragmentation.
 905 Likelihood fits over several control regions enriched with these non-prompt backgrounds are fit
 906 to data in order to derive normalization factors for these backgrounds. The specific normalization

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

908 Other processes, such as tH , tZ , $t\bar{t}WW$ and $t\bar{t}t\bar{t}$, are expected to make minor contributions
909 to the total background. The generators and setting used for these backgrounds are summarized
910 in Table 16.

911 **13 Object Reconstruction**

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

914 **13.1 Trigger Requirements**

915 Events are required to be selected by dilepton triggers. The p_T thresholds of the dilepton trigger
916 on two electrons were 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017 and 2018, while for
917 the dimuon triggers the p_T thresholds on the leading (sub-leading) muon were 18 GeV (8 GeV)
918 in 2015, and 22 GeV (8 GeV) in 2016-2018. For the electron+muon triggers, the p_T thresholds
919 on the electron (muon) were 17 GeV (14 GeV) for all datasets.

920 13.2 Light Leptons

921 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter
 922 that are associated with charged particle tracks reconstructed in the inner detector [20]. Electron
 923 candidates are required to have $p_T > 10 \text{ GeV}$ and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the transition
 924 region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are
 925 rejected. A multivariate likelihood discriminant combining shower shape and track information
 926 is used to distinguish prompt electrons from nonprompt leptons, such as those originating from
 927 hadronic showers.

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

931 Muon candidates are reconstructed by combining inner detector tracks with track segments
 932 or full tracks in the muon spectrometer [21]. Muon candidates are required to have $p_T > 10 \text{ GeV}$
 933 and $|\eta| < 2.5$.

934 All leptons are required to pass a non-prompt BDT selection developed by the main
 935 $t\bar{t}H/t\bar{t}W$ analysis, described in detail in [7]. Optimized working points and scale factors for this
 936 BDT are taken from that analysis. This BDT and the WPs used are summarized in Appendix
 937 [A](#),

938 13.3 Jets

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

947 13.4 B-tagged Jets

948 Each analysis channel used in this analysis includes b-jets in the final state. These are identified
949 using the DL1r b-tagging algorithm, which uses jet vertex and kinematic information to distin-
950 guish heavy and light flavored jets. These features are used as inputs to a neural network, the
951 output of which is used to form calibrated working points (WPs) based on how likely a jet is to
952 have originated from a b-quark. This analysis uses the 70% DL1r WP - implying an efficiency of
953 70% for truth b-jets - for selecting b-tagged jets.

954 **13.5 Missing Transverse Energy**

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

960 **13.6 Overlap removal**

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

967 This algorithm is applied to the preselected objects. The overlap removal procedure is
968 summarized in Table 17.

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

Table 17: Summary of the overlap removal procedure between electrons, muons, and jets.

969 14 Higgs Momentum Reconstruction

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

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

984 Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_AMC.

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

986 increase the statistics of the training sample.

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

988 the backend [**tensorflow**], is used, and the number of hidden layers and nodes are determined

989 using grid search optimization. Each neural network uses the LeakyReLU activation function,

990 a learning rate of 0.01, and the Adam optimization algorithm, as alternatives are found to either

991 decrease or have no impact on performance. Batch normalization is applied after each layer in

992 order to stabilize the model and decrease training time. For the classification algorithms (b-jet

993 matching, Higgs reconstruction, and 3l decay identification) binary-cross entropy is used as the

994 loss function, while the p_T reconstruction algorithm uses MSE.

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

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

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

998 using upwards of 30, this inclusive approach is found to maximize the performance of each model

999 while decreasing the variance compared to a reduced number of inputs. Each input feature is

1000 validated by comparing MC simulations to 79.8 fb^{-1} of data, with the full set of features shown

1001 in Section C..

1002 14.1 Physics Object Truth Matching

1003 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson using
1004 MC simulations of $t\bar{t}H$ events. The kinematics of the reconstructed physics objects, as well as
1005 event level variables such the jet multiplicity and missing energy, used as inputs, with the parent
1006 ID taken from the truth record used to label the data. The objects considered include light leptons
1007 and jets.

1008 Reconstructed physics objects are matched to particle level objects, in order to identify the
1009 parent particle of these reconstructed objects. Reconstructed jets are matched to truth jets based
1010 on the requirements that the reco jet and truth jet fall within $\Delta R < 0.4$, and the two objects have
1011 a p_T that agrees within 10%. Truth level and reco level leptons are required to have the same
1012 flavor, a $\Delta R < 0.1$, and p_T that agree within 10%. Events where no match can be found between
1013 the particle level decay products and the reconstructed objects are not included in training.

1014 Leptons considered as possible Higgs and top decay candidates are required to pass the
1015 selection described in Section 13.2. For jets, however, it is found that a large fraction that
1016 originate from either the top decay or the Higgs decay fall outside the selection described in
1017 Section 13.3. Specifically, jets from the Higgs decay tend to be soft, with 32% having $p_T <$
1018 25 GeV. Therefore jets with $p_T < 15$ GeV are considered as possible candidates in the models
1019 described below. By contrast, less than 5% of the jets originating from the Higgs fall below this
1020 p_T threshold. The jets are found to be well modeled even down to this low p_T threshold, as shown
1021 in Section 15.1. The impact of using different p_T selection for the jet candidates is considered

1022 in detail in Section C.6. As they are expected to originate from the primary vertex, jets are also
 1023 required to pass a JVT cut. The overlap removal selection is not applied to the objects considered
 1024 in the models.

1025 **14.2 b-jet Identification**

1026 Including the kinematics of the b-jets that originate from the top decay is found to improve the
 1027 identification of the Higgs decay products, and improve the accuracy with which the Higgs
 1028 momentum can be reconstructed. Because these b-jets are reconstructed by the detector with
 1029 high efficiency (just over 90% of the time), and can be identified relatively consistently, the first
 1030 step in reconstructing the Higgs is selecting the b-jets from the top decay.

1031 Exactly two b-jets are expected in the final state of $t\bar{t}H - \text{ML}$ events. However, in both
 1032 the 3l and 2LSS channels, only one or more b-tagged jets are required (where the 70% DL1r
 1033 b-tag working point is used). Therefore, for events which have exactly one, or more than two,
 1034 b-tagged jets, deciding which combination of jets correspond to the top decay is non-trivial.
 1035 Further, events with 1 b-tagged jet represent just over half of all $t\bar{t}H - \text{ML}$ events. Of those,
 1036 both b-jets are reconstructed by the detector 75% of the time. Therefore, rather than adjusting
 1037 the selection to require exactly 2 b-tagged jets, and losing more than half of the signal events, a
 1038 neural network is used to predict which pair of jets is most likely to correspond to truth b-jets.

1039 Once the network is trained, kinematic variables for all possible pairings of jets are fed into
 1040 the model, and the pair of jets with the highest output score are taken to be b-jets in successive

1041 steps of the analysis.

1042 An alternate approach is considered, where information about all jets in each event are
 1043 used as the feature set, and the model is tasked with identifying which two originated from the
 1044 top decay. While this approach is found to underperform the nominal approach, and therefore
 1045 not used in the analysis, the results are documented in Appendix C.4.

1046 **14.2.1 2lSS Channel**

1047 For the 2lSS channel, the input features shown in Table 18 are used for training. Here j_0 and j_1
 1048 are the two jet candidates, while l_0 and l_1 are the two leptons in the event, both ordered by p_T . jet
 1049 DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,
 1050 where 5 represents the tightest working point and 1 represents the loosest. The variables nJets
 1051 DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and
 1052 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 18: Input features used in the b-jet identification algorithm for the 2lSS channel

1053 As there are far more incorrect combinations than correct ones, by a factor of more than
 1054 20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random
 1055 sample of 5 million incorrect entries are used for training, along with around 1 million correct
 1056 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for
 1057 training.

1058 The difference between the distributions for a few of these features for the "correct" (i.e.
 1059 both jets are truth b-jets), and "incorrect" combinations are shown in Figure 14.1. The correct and
 1060 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences
 1061 in the distributions.

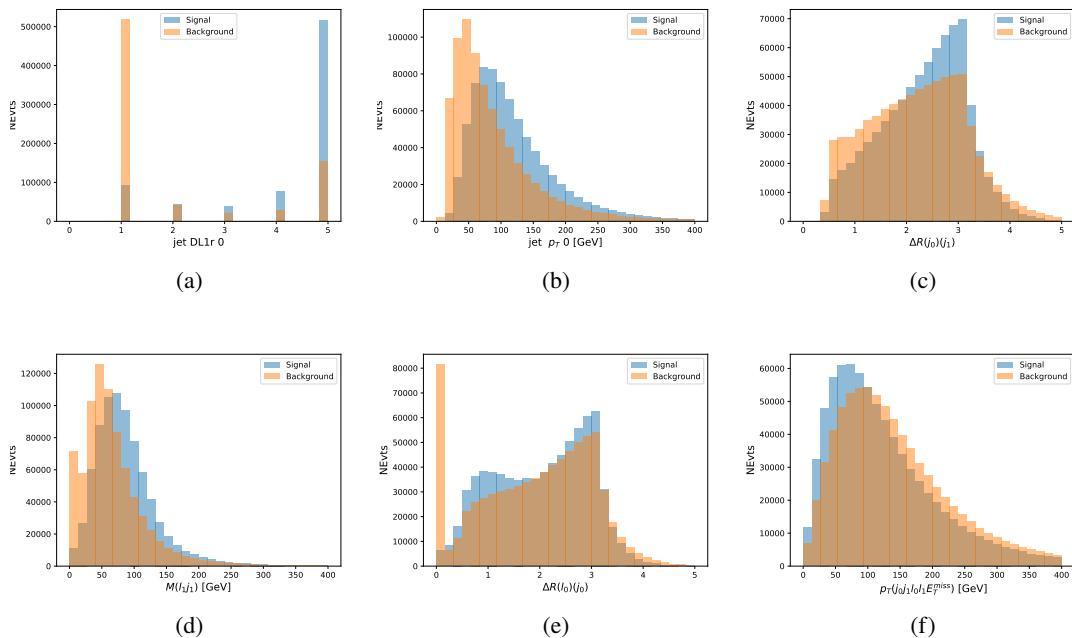


Figure 14.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the 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} .

1062 The modeling of these inputs is validated against data, with Figure 14.2 showing good
 1063 general agreement between data and MC. Plots for the complete list of features can found in
 1064 Appendix C.

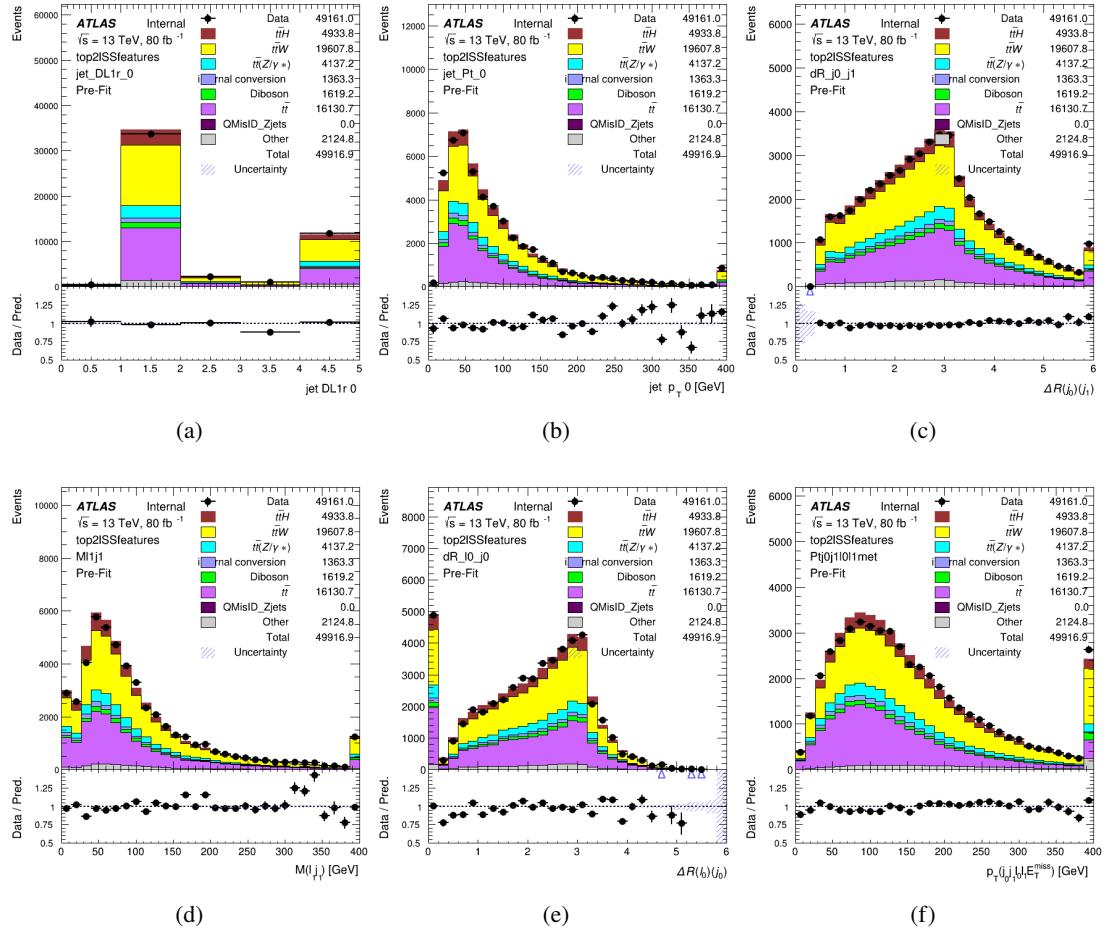


Figure 14.2: Data/MC comparisons of input features for top2ISS training for 79.8 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

1065 Based on the results of grid search evaluation, the optimal architecture is found to include
 1066 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as

1067 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for
 1068 the trained model are shown in Figure 14.2.1. The model is found to identify the correct pairing
 1069 of jets for 73% of 2lSS signal events on test data.

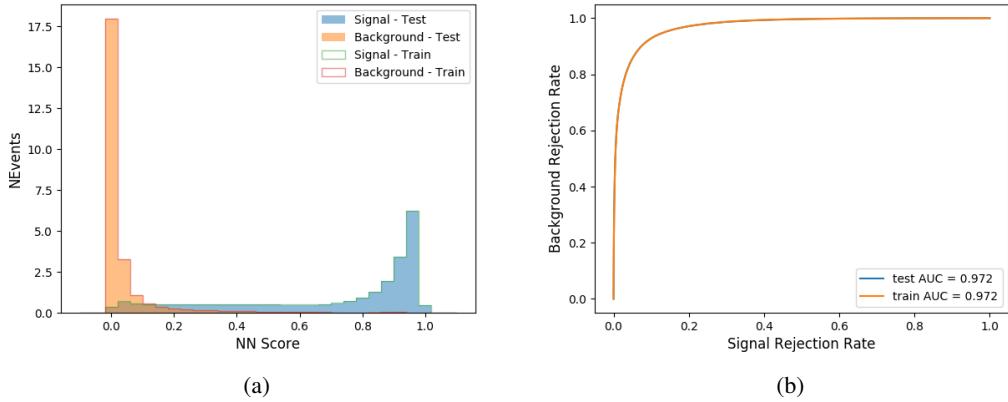


Figure 14.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1070 For point of comparison, a "naive" approach to identifying b-jets is used as well: The two
 1071 jets which pass the highest DL1r b-tag working point are assumed to be the b-jets from the top
 1072 decay. In the case that multiple jets meet the same b-tag working point, the jet with higher p_T is
 1073 used. This method identifies the correct jet pair 65% of the time.

1074 The accuracy of the model for different b-tagged jet multiplicities, compared to this naive
 1075 approach, is shown in Table 19.

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 19: 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.

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

14.2.2 3l Channel

The input features used in the 3l channel are listed in Table 20, with the same naming convention 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 20: Input features for the b-jet identification algorithm in the 3l channel.

1082 A few of these features are shown in Figure 14.4, comparing the distributions for correct
 1083 and incorrect combinations of jets.

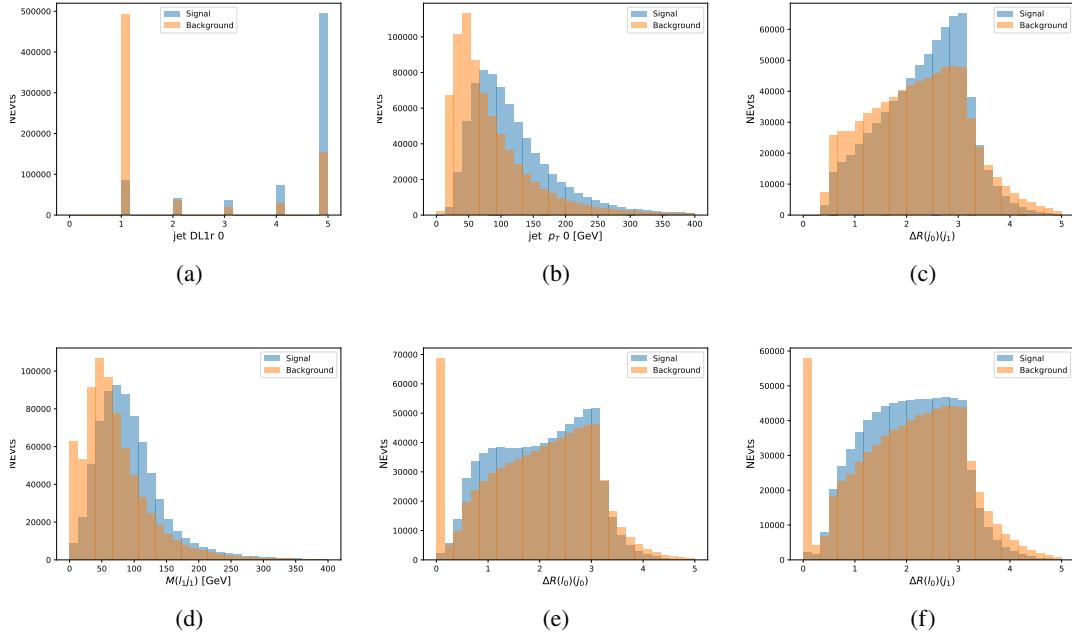


Figure 14.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and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

1084 The modeling of these inputs is validated against data, with Figure 14.5 showing good
 1085 general agreement between data and MC.

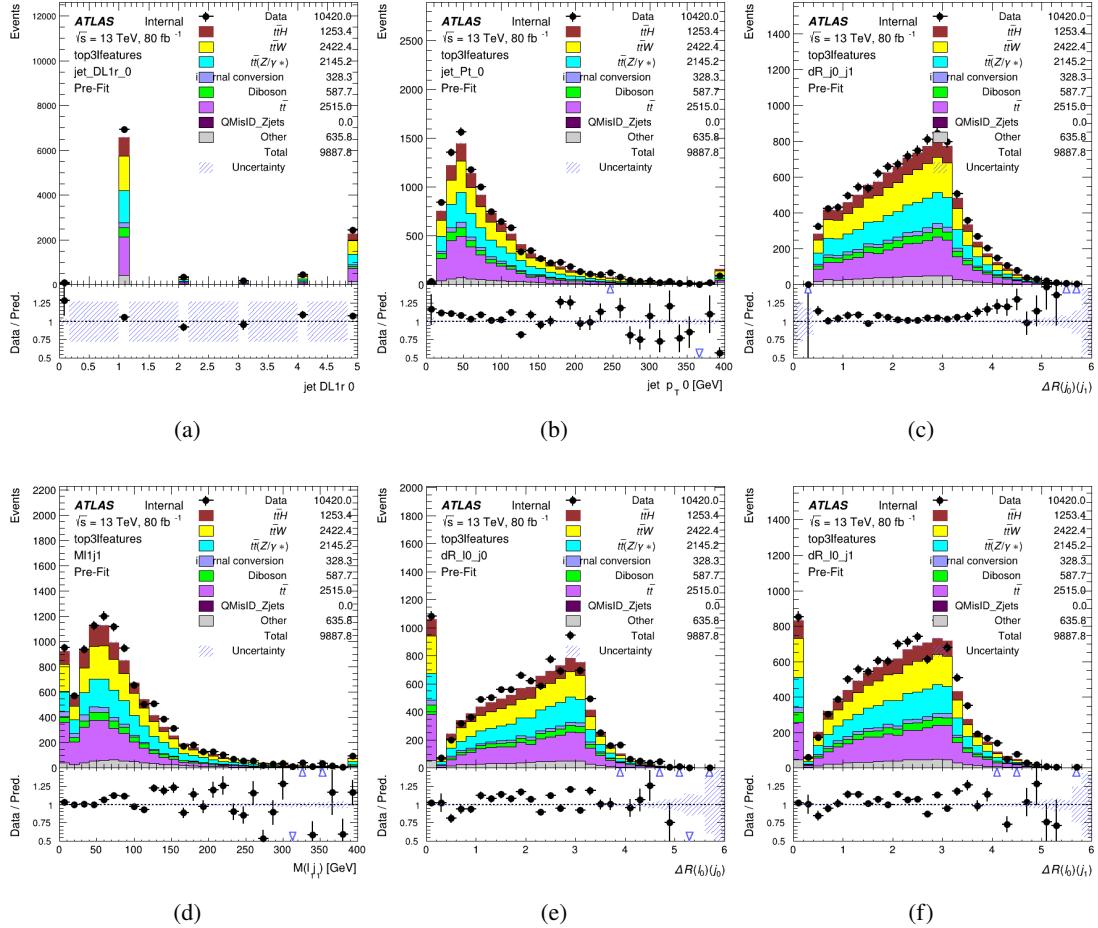


Figure 14.5: Data/MC comparisons of input features for top3l training for 79.8 fb^{-1} of data. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

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.

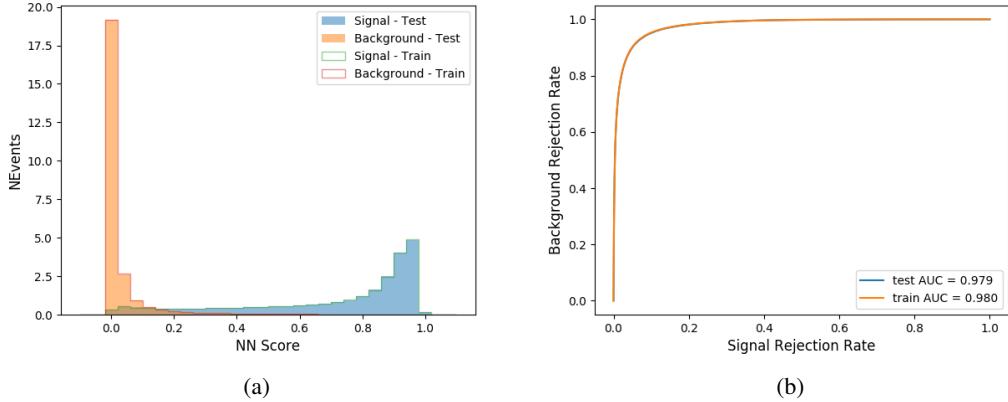


Figure 14.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1091 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal
 1092 events. The accuracy of the model is summarized in Table 21, once again compared to the naive
 1093 approach described above.

Table 21: 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%

1094 14.3 Higgs Reconstruction

1095 Techniques similar to the b-jet identification algorithms are employed to select the decay products
 1096 of the Higgs: kinematics of all possible combinations of reconstructed objects are fed into a neural

1097 network to determine which of those is most mostly to be the decay products of the Higgs.

1098 Again separate models are used for the 2lSS and 3l channels, while the 3l channel has
1099 now been split into two: $t\bar{t}H$ events with three leptons in the final state include both instances
1100 where the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the
1101 Higgs decays to two leptons (and two neutrinos which are not reconstructed).

1102 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In
1103 the semi-leptonic case the reconstructed decay products consist of two jets and a single lepton.
1104 For the fully-leptonic case, the decay products include 2 of the three leptons associated with
1105 the event. For training these models, events are separated into these two categories using truth
1106 level information. A separate MVA, described in Section 14.5, is used to make this distinction
1107 at reconstructed level, and determine which model to use.

1108 For all channels, the models described in Section 14.2 are used to identify b-jet candidates,
1109 whose kinematics are used as additional input features to help identify the Higgs decay products.
1110 These jets are not considered as possible candidates for the Higgs decay, justified by the fact that
1111 these models are found to misidentify jets from the Higgs decay as jets from the top decay less
1112 than 1% of the time.

1113 **14.3.1 2lSS Channel**

1114 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural
1115 network is trained on the kinematics of different combinations of leptons and jets, as well as the

¹¹¹⁶ b-jets identified in Section 14.2, with the specific input features listed in Table 22.

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

¹¹¹⁷ Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
¹¹¹⁸ lepton in the event is labeled l_T , as it is assumed to have come from the decay of one of the top
¹¹¹⁹ quarks. b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet
¹¹²⁰ Reco Score is the output of the b-jet reconstruction algorithm.

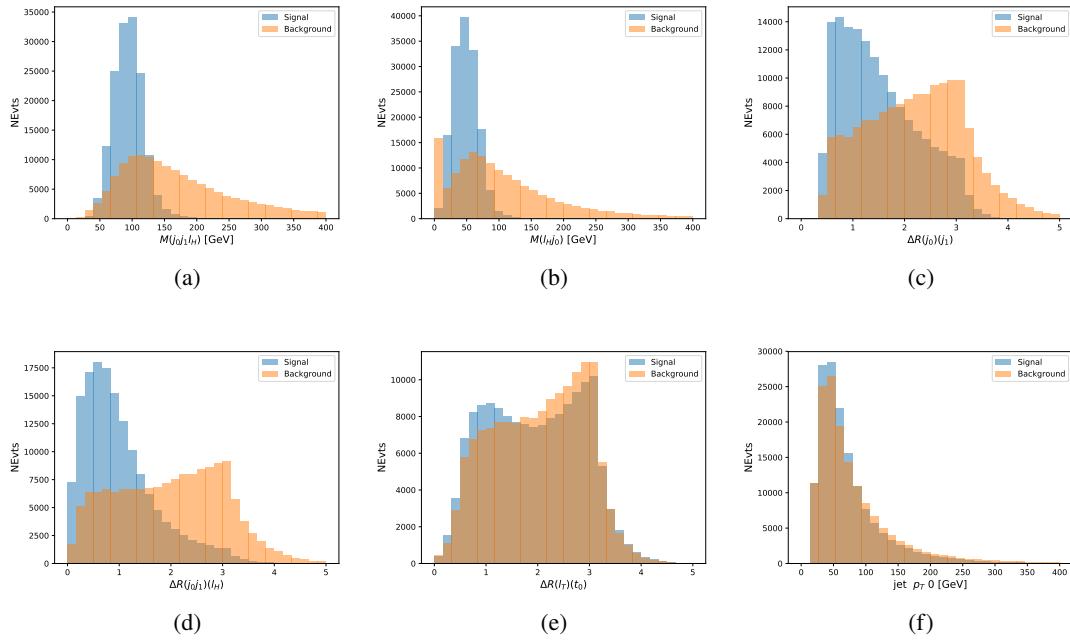


Figure 14.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate,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

1121 The modeling of these inputs is validated against data, with Figure 14.2 showing good
 1122 general agreement between data and MC. Plots for the complete list of features can found in
 1123 Section C.

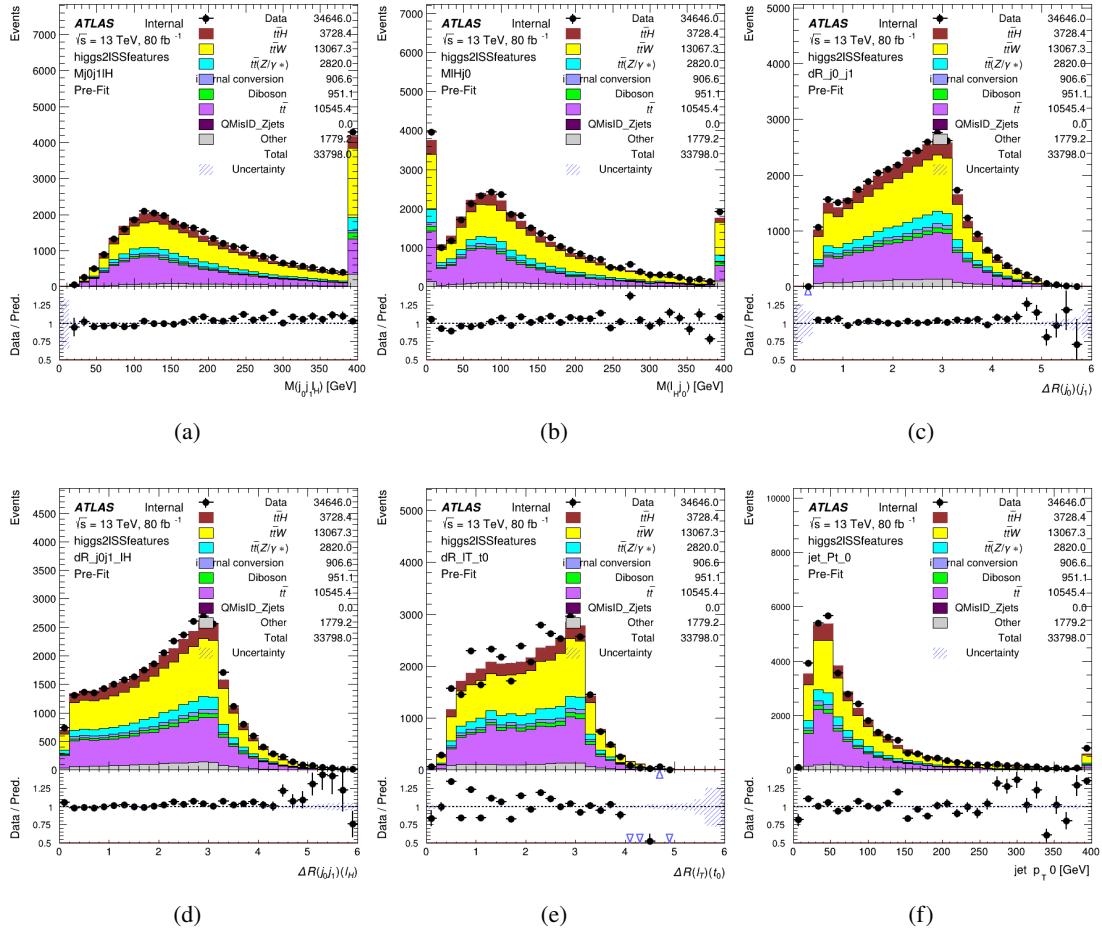


Figure 14.8: Data/MC comparisons of input features for higgs2ISS training for 79.8 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet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

1124 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2
 1125 million events, with an additional 200,000 reserved for testing the model. In order to compensate
 1126 for the large number of incorrect combinations, these have been downsampled such that the correct
 1127 combinations represent over 10% of the training set. The output of the NN is summarized in
 1128 Figure 14.3.1.

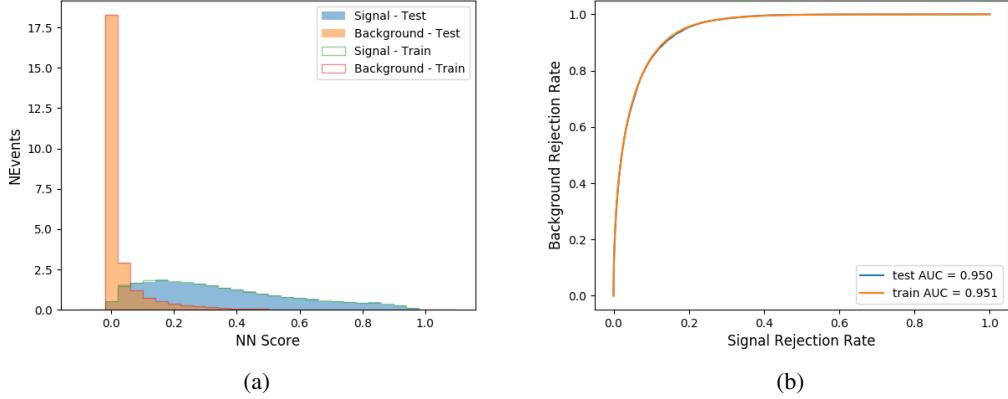


Figure 14.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1129 The neural network identifies the correct combination 55% of the time. It identifies the
 1130 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
 1131 81% of the time.

1132 14.3.2 3l Semi-leptonic Channel

1133 For 3l $t\bar{t}H$ where the Higgs decay semi-leptonically, the decay products include one of the three
 1134 leptons and two jets. In this case, the other two leptons originated from the decay of the tops,
 1135 meaning the opposite-sign (OS) lepton cannot have come from the Higgs. This leaves only the two
 1136 same-sign (SS) leptons as possible Higgs decay products.

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

Table 23: Input features used to identify the Higgs decay products in 3l semi-leptonic events

1137 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 1138 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
 1139 the b-jet identification algorithm. The b-jet Reco Score is the output of the b-jet identification
 1140 algorithm.

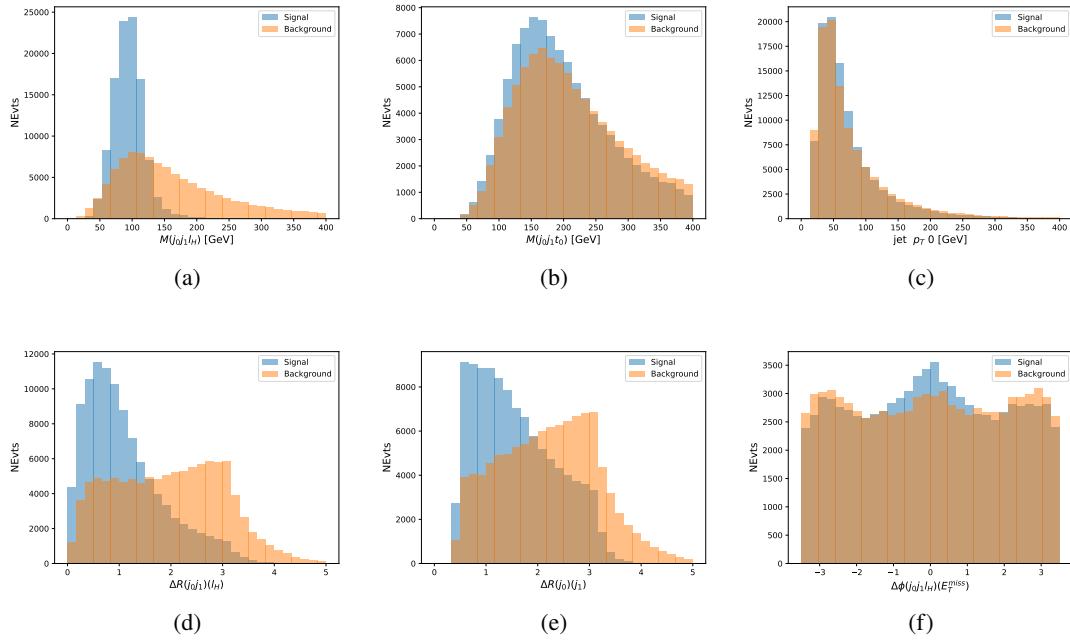


Figure 14.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1141 The modeling of these inputs is validated against data, with Figure 14.11 showing good
 1142 general agreement between data and MC. Plots for the complete list of features can found in
 1143 appendix C.1.

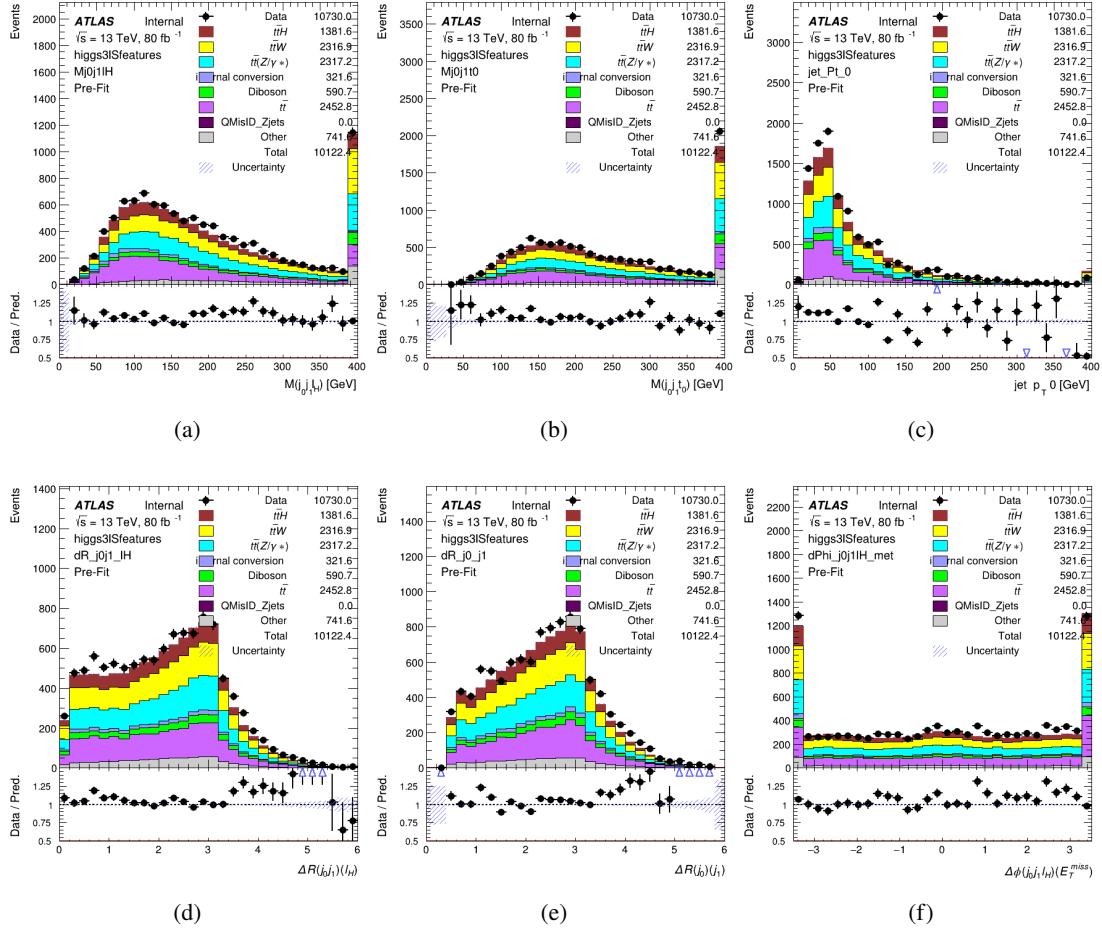


Figure 14.11: Data/MC comparisons of input features for higgs3lS training for 79.8 fb^{-1} of data.

1144 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.
 1145 Once again, incorrect combinations are downsampled, such that the correct combinations are
 1146 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN
 1147 is summarized in Figure 14.3.2.

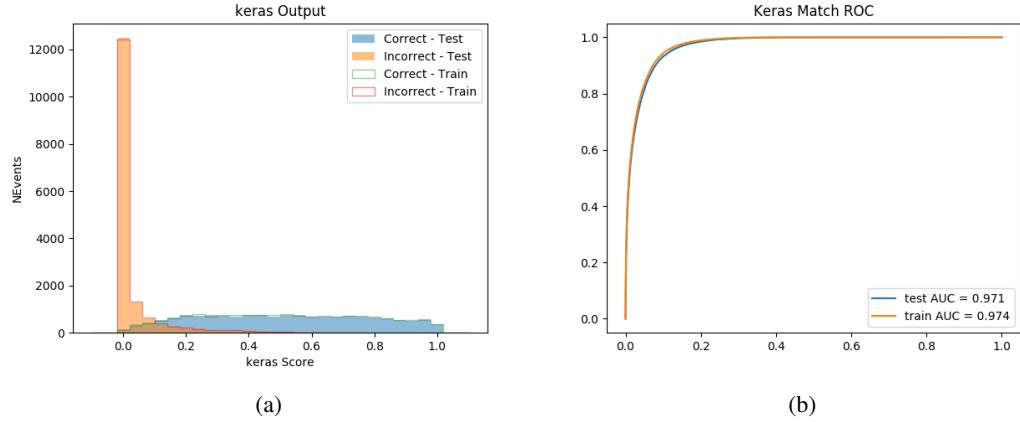


Figure 14.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1148 The neural network identifies the correct combination 64% of the time. It identifies the
 1149 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
 1150 83% of the time.

1151 14.3.3 3l Fully-leptonic Channel

1152 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 1153 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to
 1154 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 1155 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 1156 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 24: Input features used to identify the Higgs decay products in 3lF events

Table 25: Input features used to identify the Higgs decay products in 3l fully leptonic events

1157 Here l_{H_0} and l_{H_1} are the Higgs decay candidates. The other lepton in the event is labeled
 1158 l_T . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco
 1159 Score is the output of the Higgs reconstruction algorithm.

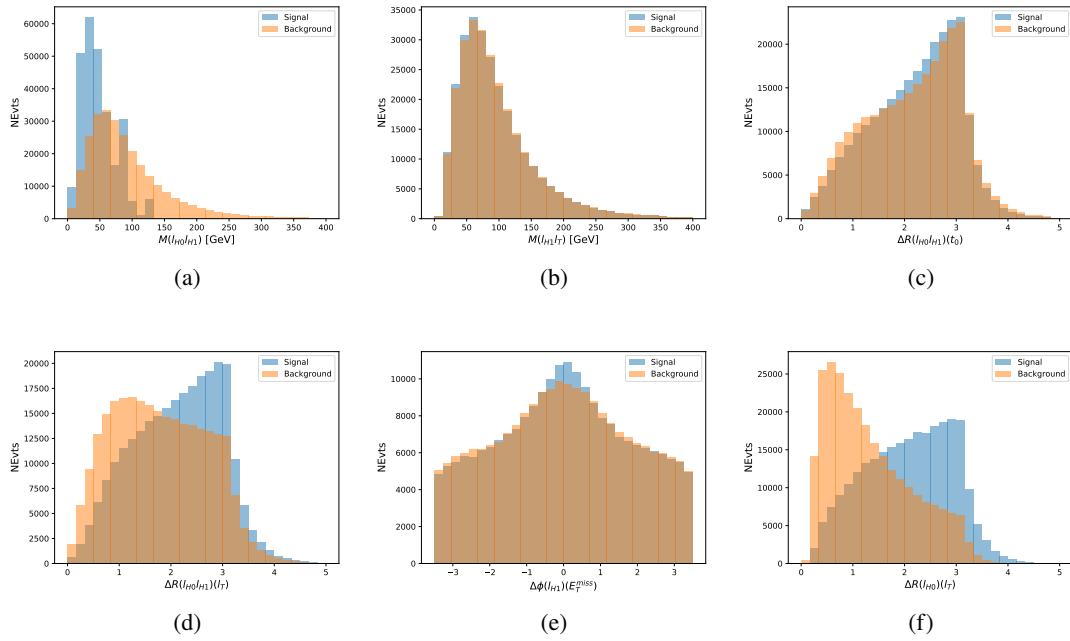


Figure 14.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1160 The modeling of these inputs is validated against data, with Figure 14.14 showing good
1161 general agreement between data and MC. Plots for the complete list of features can found in
1162 Section C.

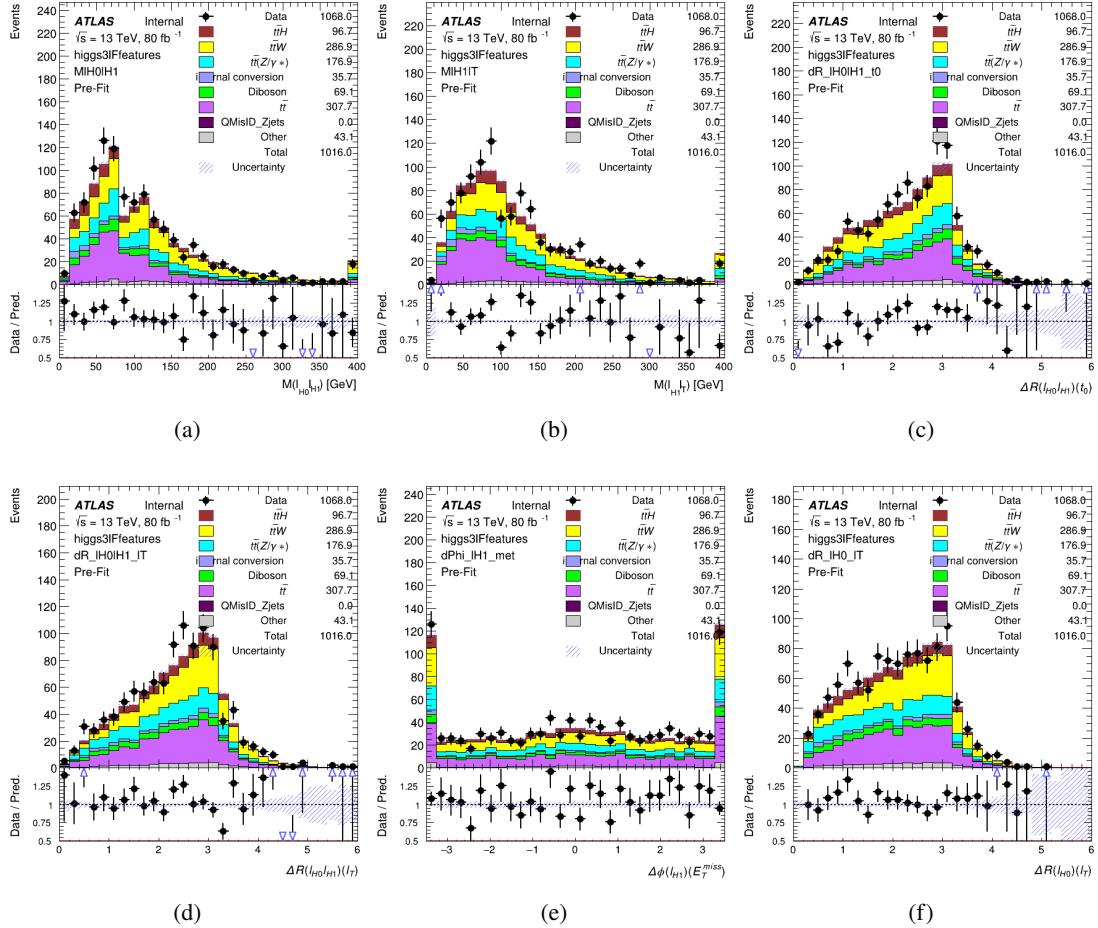


Figure 14.14: Data/MC comparisons of input features for higgs3lF training for 79.8 fb^{-1} of data.

1163

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

1164

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

1165

14.3.3.

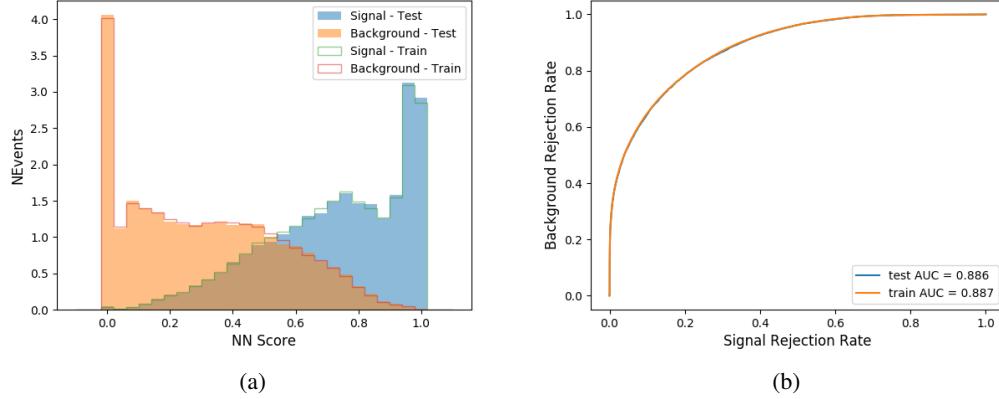


Figure 14.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

1166 The correct lepton is identified by the model for 80% of events in the testing data set.

1167 14.4 p_T Prediction

1168 Once the most probable decay products have been identified, their kinematics are used as inputs
 1169 to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,
 1170 a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay
 1171 are included as well, as these are found to improve performance. The truth p_T of the Higgs,
 1172 as predicted by MC, are used as labels. Separate models are built for each channel - 2lSS, 3l
 1173 Semi-leptonic and 3l Fully-leptonic.

1174 As a two-bin fit is targeted for the final result, some metrics evaluating the performance
 1175 of the models aim to show how well it distinguished between "high p_T " and "low p_T " events. A

1176 cutoff point of 150 GeV is used to define these two categories.

1177 Because the analysis uses a two bin fit of the Higgs p_T , the momentum reconstruction
1178 could be treated as a binary classification problem, rather than a regression problem. This
1179 approach is explored in detail in Section C.5, and is found not to provide any significant increase
1180 in sensitivity. The regression approach is used because it provides more flexibility for future
1181 analyses, as it is independent of the cutoff between high and low p_T , as well as the number of
1182 bins. Further, a regression allows the output of the neural network to be more clearly understood,
1183 as it can be directly compared to a physics observable.

1184 **14.4.1 2lSS Channel**

1185 The input variables listed in Table 26 are used to predict the Higgs p_T in the 2lSS channel. Here
1186 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
1187 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
1188 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
1189 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
1190 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 26: Input features for reconstructing the Higgs p_T spectrum for 2lSS events

1191 The optimal neural network architecture for this channel is found to consist of 7 hidden
 1192 layers with 60 nodes each. The input data set includes 1.2 million events, 10% of which is used
 1193 for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1194 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 1195 truth Higgs p_T in Figure 14.16. In order to visualize the model performance more clearly, in (a)
 1196 of that figure, the color of each point is determined by Kernel Density Estimation (KDE). The
 1197 color shown represents the logarithm of the output from KDE, to counteract the large number of
 1198 low p_T events. For that same reason, each column of the histogram shown in (b) of Figure 14.16
 1199 is normalized to unity. This plot therefore demonstrates what the model predicts for each slice
 1200 of truth p_T .

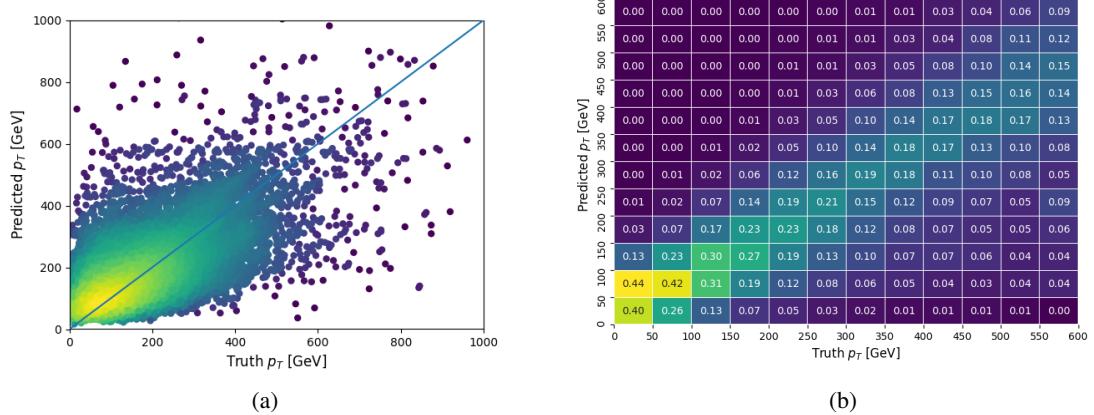


Figure 14.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized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

1201 We are also interested in how well the model distinguishes between events with $p_T < 150$
 1202 GeV and > 150 GeV. Figure 14.17 demonstrates the NN output for high and low p_T events based
 1203 on this cutoff.

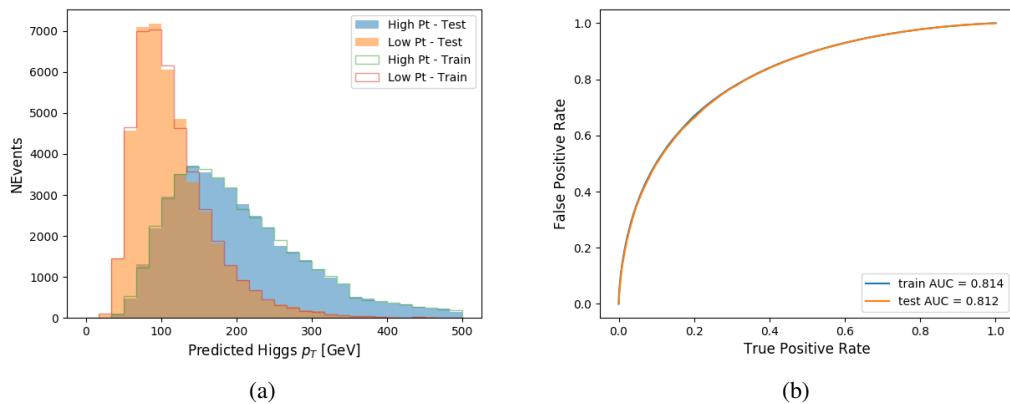


Figure 14.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

1204 **14.4.2 3l Semi-leptonic Channel**

1205 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 27: Input features for reconstructing the Higgs p_T spectrum for 3lS events

1206 Again, j_0 and j_1 are the two jets identified as Higgs decay products, ordered by p_T . The
 1207 lepton identified as originating from the Higgs is labeled l_H , while the other two leptons are
 1208 labeled l_{T0} and l_{T1} . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm.
 1209 The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs reconstruction algorithm,
 1210 and the b-jet identification algorithm, respectively.

1211 The optimal neural network architecture for this channel is found to consist of 7 hidden
 1212 layers with 80 nodes each. The inputdata set includes one million events, 10% of which is used
 1213 for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1214 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 1215 truth Higgs p_T in Figure 14.18. Once again, (a) of 14.18 shows a scatterplots of predicted vs
 1216 truth p_T , where the color of each point corresponds to the log of the relative KDE at that point.
 1217 Each column of the the histogram in (b) is normalized to unity, to better demonstrate the output
 1218 of the NN for each slice of truth p_T .

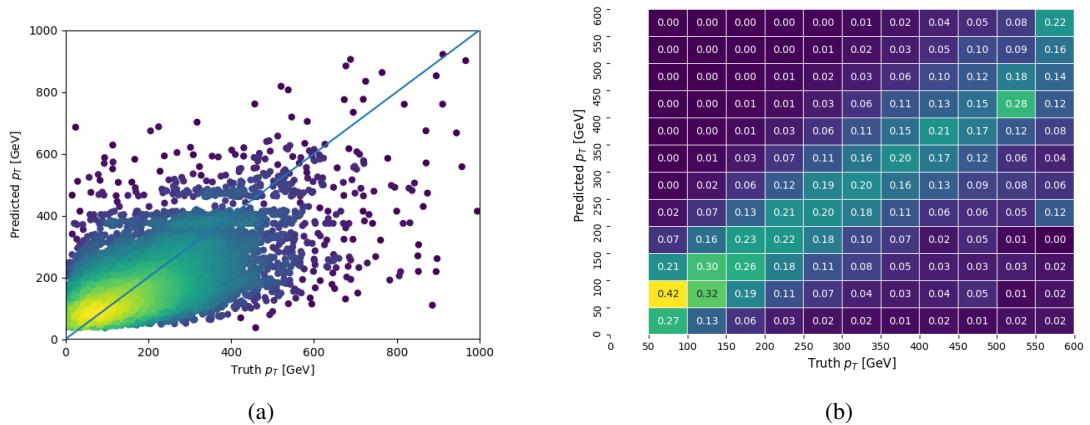


Figure 14.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

1219 Figure 14.19 shows (a) the output of the NN for events with truth p_T less than and greater
 1220 than 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low
 1221 p_T events.

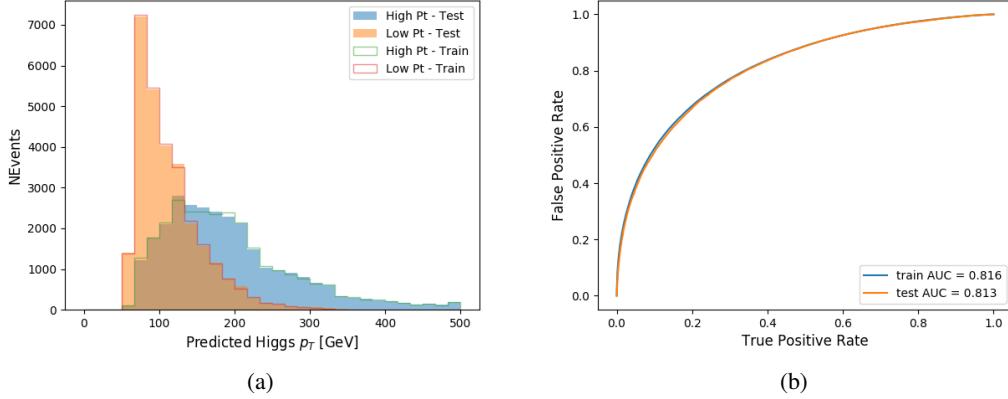


Figure 14.19: (

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 28 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 28: 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

1225 originating from the Higgs, while \mathbf{l}_T is the other lepton in the event. The Higgs Reco Score and
 1226 b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification
 1227 algorithm, respectively.

1228 The optimal neural network architecture for this channel is found to consist of 5 hidden
 1229 layers with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for
 1230 testing, the other 90% for training. Training is found to converge after around 150 epochs.

1231 The predicted transverse momentum, as a function of the truth p_T , is shown in Figure
 1232 [14.20](#).

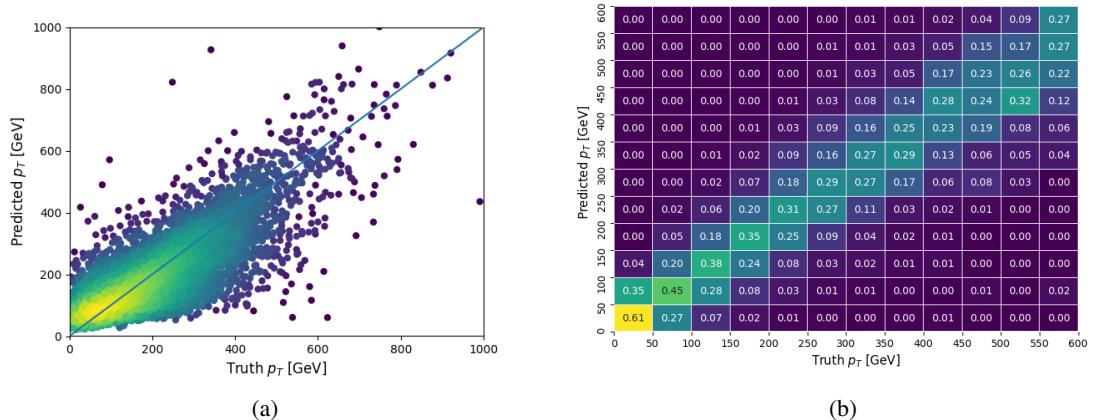


Figure 14.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

1233 When split into high and low p_T , based on a cutoff of 150 GeV, the

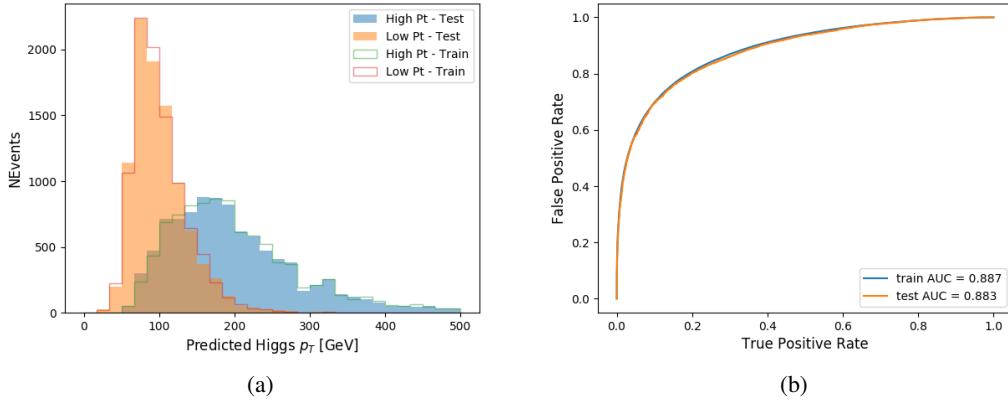


Figure 14.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

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 29.

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

1244 Here l_0 is the opposite charge lepton, l_1 and l_2 are the two SS leptons order by ΔR
 1245 from lepton 0. score3lF and score3lS are the outputs of the 3lS and 3lF Higgs reconstruction
 1246 algorithms, while topScore is the output of the b-jet identification algorithm.

1247 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these
 1248 two decay modes. The output of the model is summarized in Figure 14.22.

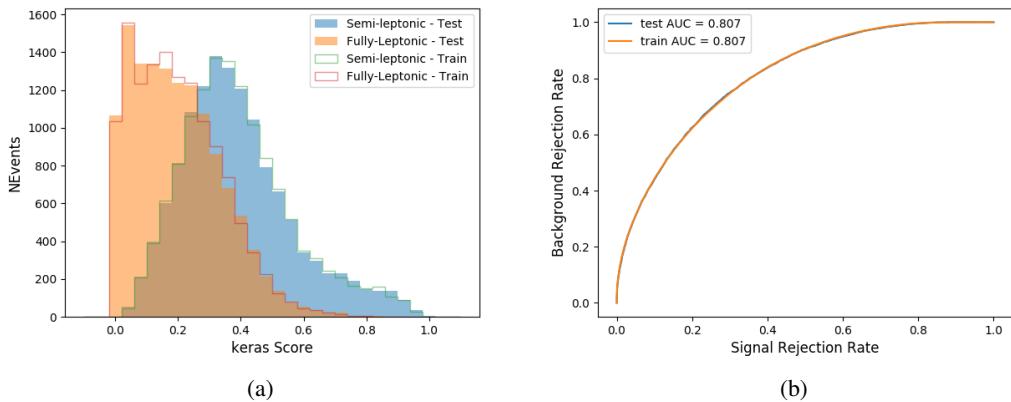


Figure 14.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

1249 A cutoff of 0.23 is determined to be optimal for separating 3lS and 3lF in the fit.

1250 15 Signal Region Definitions

1251 Events are divided into two channels based on the number of leptons in the final state: one
1252 with two same-sign leptons, the other with three leptons. The 3l channel includes events where
1253 two leptons originated from the Higgs boson as well as events where only one of the leptons
1254 originated from the Higgs. This motivates splitting the 3l channel into semi-leptonic, and fully
1255 leptonic channels, after an event preselection has been applied.

1256 15.1 Pre-MVA Event Selection

1257 A preselection is applied to define orthogonal analysis channels based on the number of leptons
1258 in each event. For the 2lSS channel, the following preselection is used:

- 1259 • Two very tight, same-charge, light leptons with $p_T > 20 \text{ GeV}$
- 1260 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets
- 1261 • No reconstructed tau candidates

1262 The event yield after the 2lSS preselection has been applied, for MC and data at 79.8 fb^{-1} ,
1263 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 30: Yields of the 2lSS preselection region

1264

Figure 15.1. Good general agreement is found.

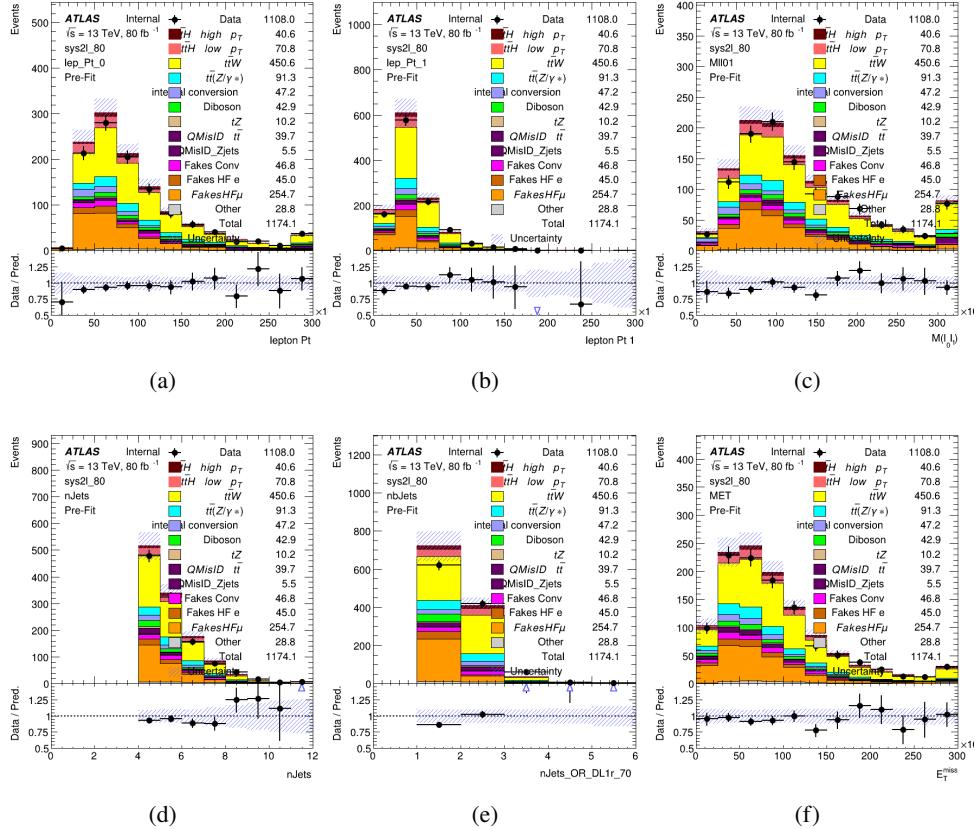


Figure 15.1: Data/MC comparisons of the 2LSS pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1265 For the 3l channel, the following selection is applied:

- 1266 • Three light leptons with total charge ± 1
- 1267 • Same charge leptons are required to be very tight, with $p_T > 20$ GeV
- 1268 • Opposite charge lepton must be loose, with $p_T > 10$ GeV
- 1269 • ≥ 2 reconstructed jets, ≥ 1 b-tagged jets

- 1270 • No reconstructed tau candidates
- 1271 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs
- 1272 The event yield after the 3l preselection has been applied, for MC and data at 79.8 fb^{-1} ,
- 1273 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 31: Yields of the 3l preselection region.

Table 32: Yields of the 3l preselection region.

1274 Comparisons of kinematic distributions for data and MC in this region are shown in Figure

1275 15.2.

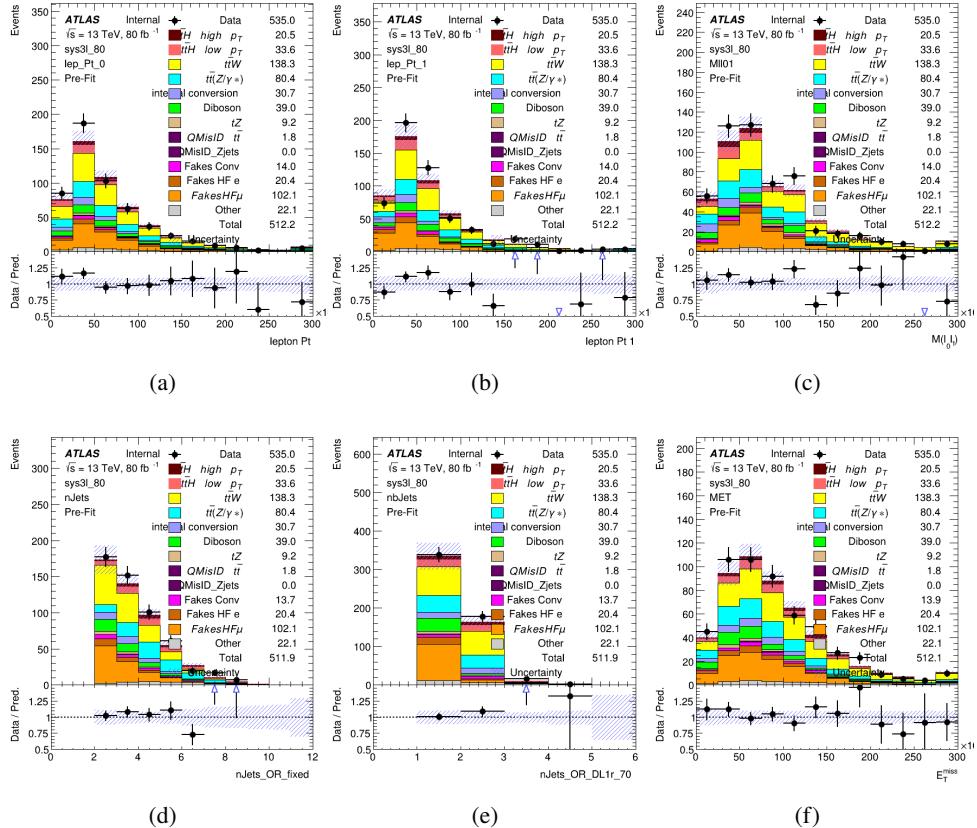


Figure 15.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1276 15.2 Event MVA

1277 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
 1278 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
 1279 (3lF). Here events with three leptons are split into 3lS and 3lF based on the model described
 1280 in 14.5. In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

1281 [27] are trained using the kinematics of signal and background events derived from Monte Carlo
1282 simulations. Events are weighted in the BDT training by the weight of each Monte Carlo event.

1283 Because the background composition differs for events with a high reconstructed Higgs p_T
1284 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and
1285 low p_T regions. This is found to provide better significance than attempting to build an inclusive
1286 model, as demonstrated in appendix C.2. A cutoff of 150 GeV is used. This gives a total of 6
1287 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
1288 3lF high p_T , and 3lF low p_T .

1289 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}	$\min \Delta R(l_0)(\text{jet})$
$\min \Delta R(l_1)(\text{jet})$	$\min \Delta R(\text{Lepton})(\text{bjet})$	$mjj\text{Max frwdJet}$
nJets	nJets OR DL1r 60	nJets OR DL1r 70
nJets OR DL1r 85	topRecoScore	

Table 33: Input features used to distinguish signal and background events in the 2lSS channel.

1290 While for each of the 3l BDTs, the features listed below are used for training:

$M(\text{lep}, E_T^{\text{miss}})$	$M(l_0 l_1)$	$M(l_0 l_1 l_2)$
$M(l_0 l_2)$	$M(l_1 l_2)$	$\text{binHiggs } p_T \text{ 3lS}$
$\Delta R(l_0)(l_2)$	$\Delta R(l_0)(l_1)$	$\Delta R(l_0)(l_2)$
higgsRecoScore3lS	decayScore	higgsRecoScore3lF
$\text{jet } \eta \text{ 0}$	$\text{jet } \eta \text{ 1}$	$\text{jet } \eta \text{ 1}$
$\text{jet } \phi \text{ 0}$	$\text{jet } \phi \text{ 1}$	$\text{jet } p_T \text{ 0}$
$\text{jet } p_T \text{ 1}$	$\text{Lepton } \eta \text{ 0}$	$\text{Lepton } \eta \text{ 1}$
$\text{Lepton } \eta \text{ 2}$	$\text{Lepton } \phi \text{ 0}$	$\text{Lepton } \phi \text{ 1}$
$\text{Lepton } \phi \text{ 2}$	$\text{Lepton } p_T \text{ 0}$	$\text{Lepton } p_T \text{ 1}$
$\text{Lepton } p_T \text{ 2}$	E_T^{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 34: Input features used to distinguish signal and background events in the 3l channel.

1291 Modelling of each of these input features is verified in Appendix C.2 by comparing data
 1292 and MC for 79.8 fb^{-1} . The BDTs are produced with a maximum tree depth of 6, using AUC as
 1293 the target loss function. The BDT response distribution and ROC curve for each model is shown
 1294 in Figures 15.3-15.5.

2LSS

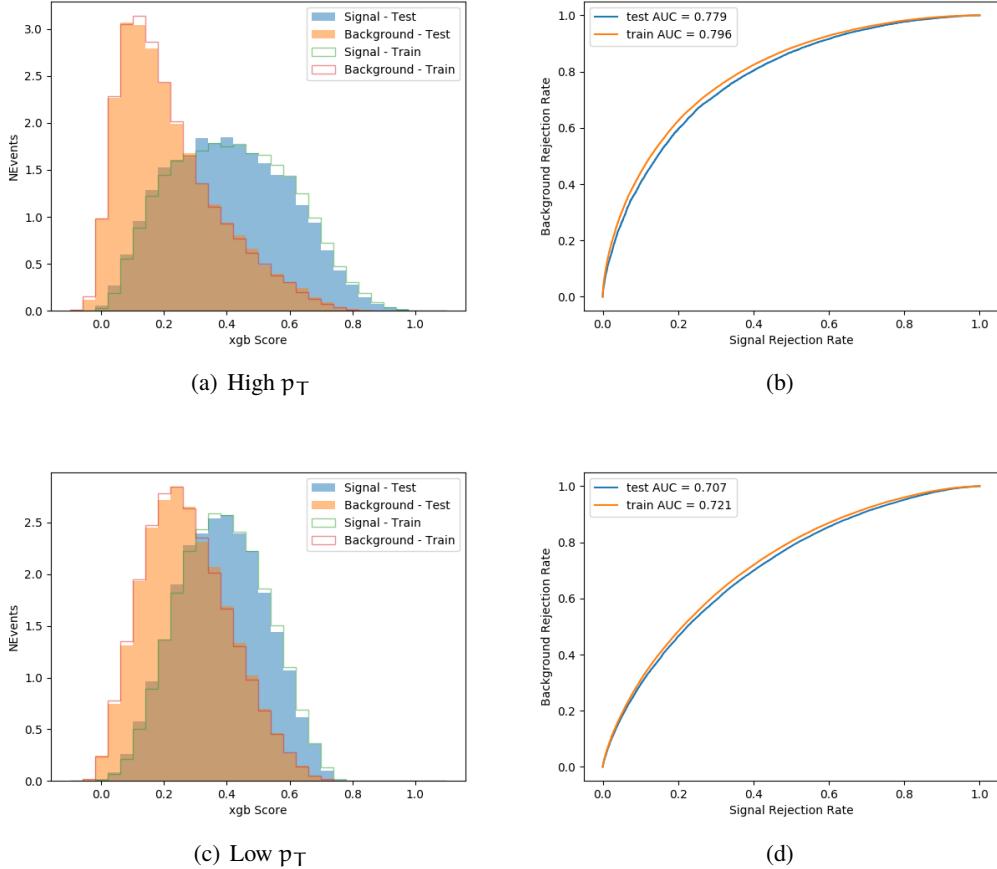


Figure 15.3: Output BDT scores of training and testing data for signal (blue) and background (orange) for 2LSS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 2LSS high and low p_T models, respectively.

3l - Semileptonic

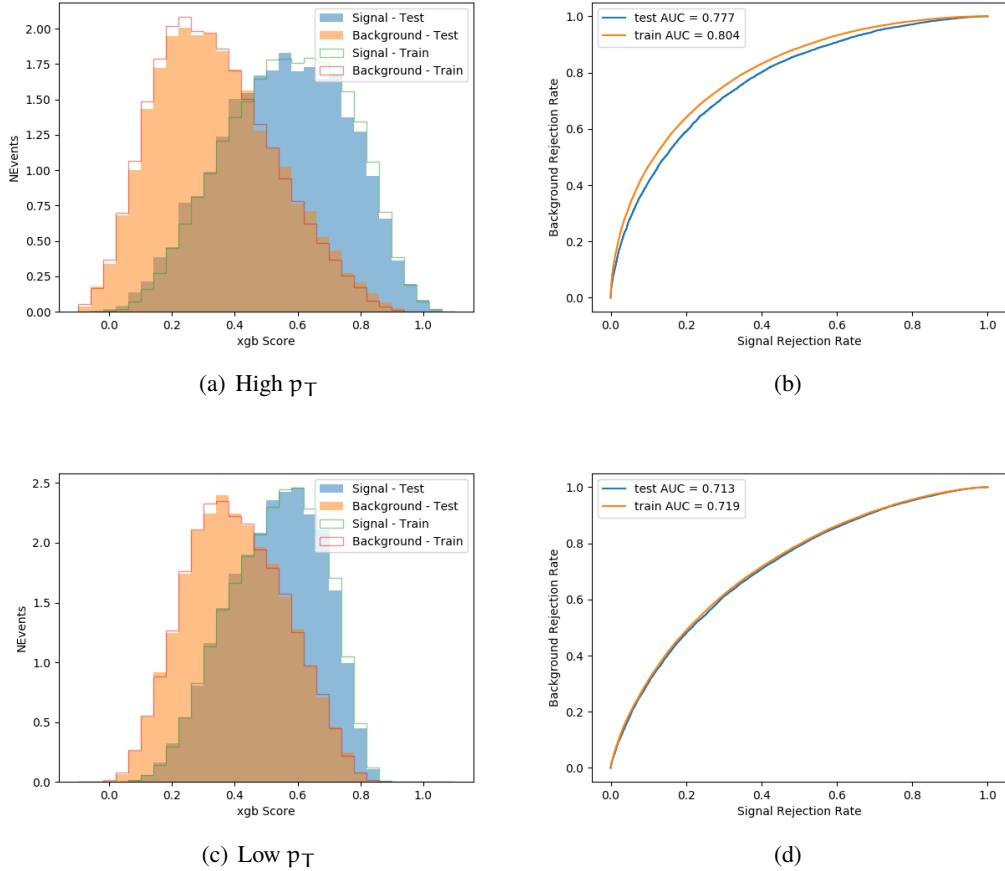


Figure 15.4: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lS high and low p_T models, respectively.

3l - Fully Leptonic

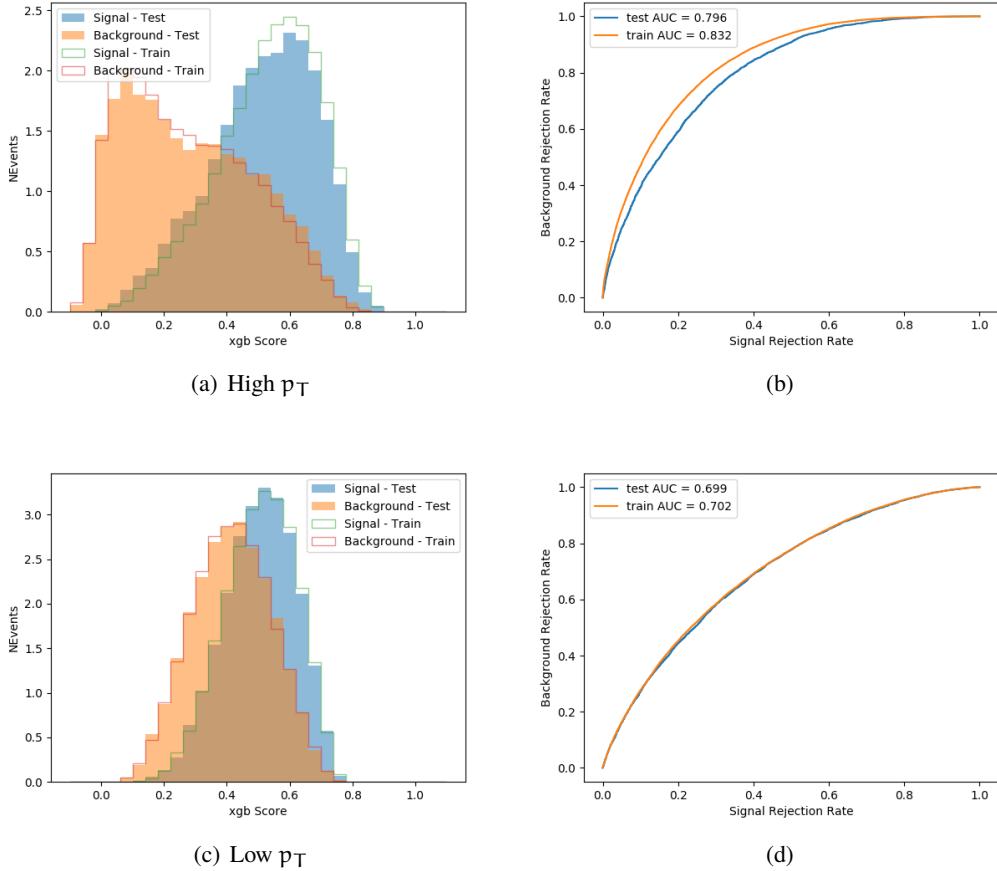


Figure 15.5: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lF events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lF high and low p_T models, respectively.

1295 Output distributions of each MVA, comparing MC predictions to data at 79.8 fb^{-1} are

1296 shown in figures [15.6-15.2](#).

High p_T Background Rejection BDTs

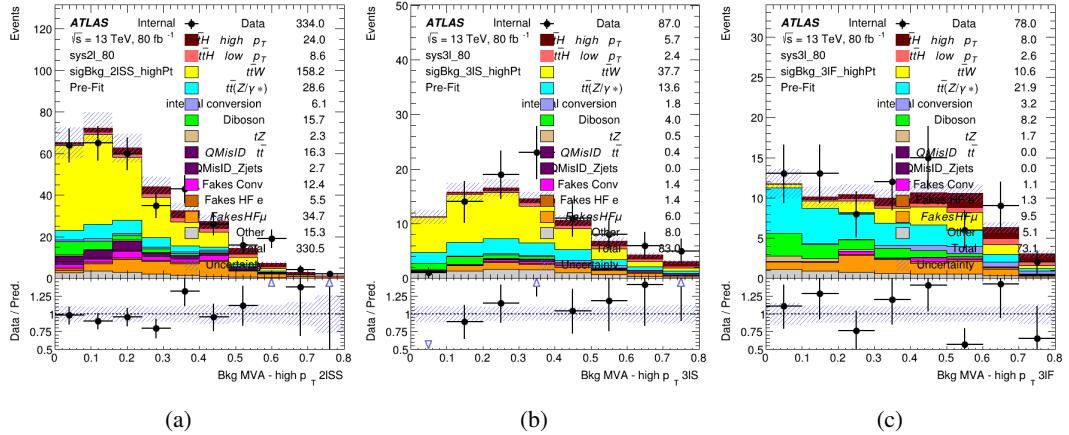


Figure 15.6: Output score of the high p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

Low p_T Background Rejection BDTs

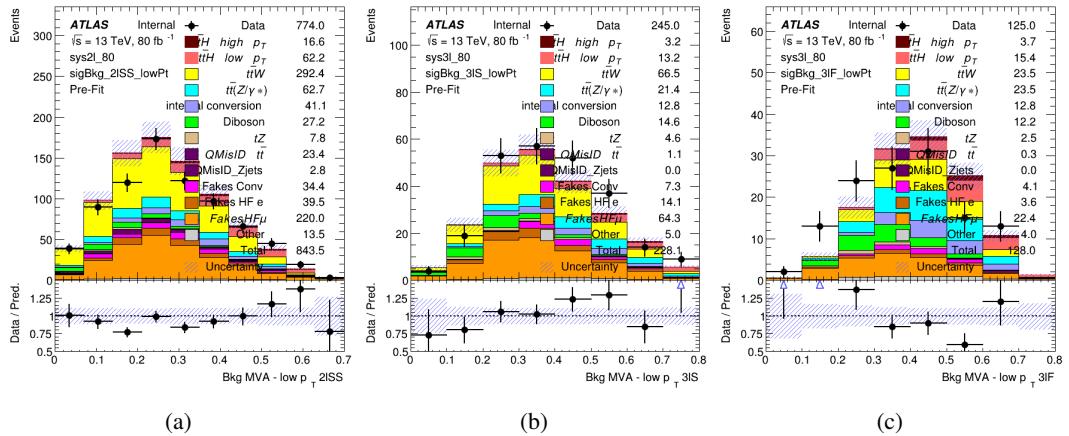


Figure 15.7: Output score of the low p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

1297 **15.3 Signal Region Definitions**

1298 Once pre-selection has been applied, channels are further refined based on the MVAs described
 1299 above. The output of the model described in Section 14.5 is used to separate the three channel
 1300 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs
 1301 boson. This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

1302 For each event, depending on the number of leptons as well as whether the p_T of the Higgs
 1303 is predicted to be high (> 150 GeV) or low (< 150 GeV), a cut on the appropriate background
 1304 rejection MVA is applied. The particular cut values, listed in Table 35, are determined by
 1305 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 35: Cutoff values on background rejection MVA score applied to signal regions.

1306 The event preselection and MVA selection listed in Table 35 are used define the three
 1307 signal regions used in the fit. These signal region definitions are summarized in Table 36.

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

16 Systematic Uncertainties

1308 The systematic uncertainties that are considered are summarized in Table 37. These are imple-
 1309 mented in the fit either as a normalization factors or as a shape variation or both in the signal
 1310 and background estimations. The numerical impact of each of these uncertainties is outlined in
 1311 section 17.
 1312

Table 37: 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

¹³¹³ The uncertainty in the combined integrated luminosity is derived from a calibration of the
¹³¹⁴ luminosity scale using x-y beam-separation scans performed for 13 TeV proton-proton data [29],
¹³¹⁵ [30].

¹³¹⁶ The experimental uncertainties are related to the reconstruction and identification of light
¹³¹⁷ leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} .

1318 The sources which contribute to the uncertainty in the jet energy scale [31] are decomposed
 1319 into uncorrelated components and treated as independent sources in the analysis. This method
 1320 decomposes the uncertainties into 30 nuisance parameters included in the fit. A similar method
 1321 is used to account for jet energy resolution (JER) uncertainties, and 8 JER uncertainty components
 1322 are included as NPs in the fit.

1323 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
 1324 [32] are also decomposed into uncorrelated components. The large number of components for
 1325 b-tagging is due to the calibration of the distribution of the BDT discriminant.

1326 As mentioned in Section 12.2, a normalization corrections and uncertainties on the estim-
 1327 ates of non-prompt leptons backgrounds are derived using data driven techniques, described in
 1328 detail in [7]. These are derived from a likelihood fit over various non-prompt enriched control
 1329 regions, targeting several sources of non-prompt light leptons separately: external conversion
 1330 electrons, internal conversion electrons, electrons from heavy flavor decays, and muons from
 1331 heavy flavor decays.

1332 The normalization factor and uncertainty applied to each source of non-prompt leptons is
 1333 summarized in Table 16

Process	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 38: Normalization factors - with statistical and systematic uncertainties - derived from the fit over fake control regions for each source of non-prompt leptons considered.

1334 In addition to those derived from the control regions, several additional uncertainties are
1335 assigned to the non-prompt lepton background. An additional 25% uncertainty on material
1336 conversions is assigned, based on the comparison between data and MC in a region where a
1337 loose electron fails the photon conversion veto. A shape uncertainty of 15% (6%) is assigned to
1338 the HF non-prompt electron (muon) background based on a comparison between data and MC
1339 where the second leading electron (muon) is only required to be loose. As the contribution from
1340 light non-prompt leptons is small, about 10% percent of the contribution from HF non-prompt
1341 leptons, it is derived from the agreement between data and simulation in a LF enriched region at
1342 low values of the non-prompt lepton BDT. The resulting uncertainty is 100%, and is taken to be
1343 uncorrelated between internal and material conversions.

1344 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and
1345 scale uncertainties are taken from theory calculations for the predominate prompt backgrounds.
1346 Following the nominal $t\bar{t}H - ML$ analysis, a 50% uncertainty is applied to Diboson to account
1347 for the large uncertainty in estimating $VV +$ heavy flavor. The other “rare” background processes
1348 - including tZ , rare top processes, $ttWW$, WtZ , VVV , $tHjb$ and WtH - are assigned an overall
1349 50% normalization uncertainty as well. The theory uncertainties applied to the MC estimates
1350 are summarized in Table 39.

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

1351 Additional uncertainties to account for t̄W mismodelling are also applied. These include
 1352 a “Generator” uncertainty, based on a comparison between the nominal Sherpa 2.2.5 sample,
 1353 and the formerly used aMC@NLO sample, and an “Extra radiation” uncertainty, which includes
 1354 renormalisation and factorisation scale variations of the Sherpa 2.2.5 sample.

1355 17 Results

1356 A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p_T spectrum
 1357 in the three signal regions, 2lSS, 3lS, and 3lF. The signal is split into high and low p_T samples,
 1358 based on whether the truth p_T of the Higgs is above or below 150 GeV. The parameters $\mu_{t\bar{t}H \text{ high } p_T}$
 1359 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
 1360 between the observed value and the theory prediction. Unblinded results are shown for the 79.8
 1361 fb⁻¹ data set, as well as MC only projections of results using the full Run-2, 139 fb⁻¹ dataset.

1362 As described in Section 16, there are 229 systematic uncertainties that are considered
 1363 as NPs in the fit. These NP s are constrained by Gaussian or log-normal probability density
 1364 functions. The latter are used for normalisation factors to ensure that they are always positive.
 1365 The expected number of signal and background events are functions of the likelihood. The prior
 1366 for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its
 1367 nominal value.

1368 17.1 Results - 79.8 fb^{-1}

1369 As the data collected from 2015-2017 has been unblinded for $t\bar{t}H$ – ML channels, represent-
 1370 ing 79.8 fb^{-1} , those events are unblinded. The predicted Higgs p_T spectrum is fit to data
 1371 simultaneously in each of the three signal regions shown in Figure 17.1.

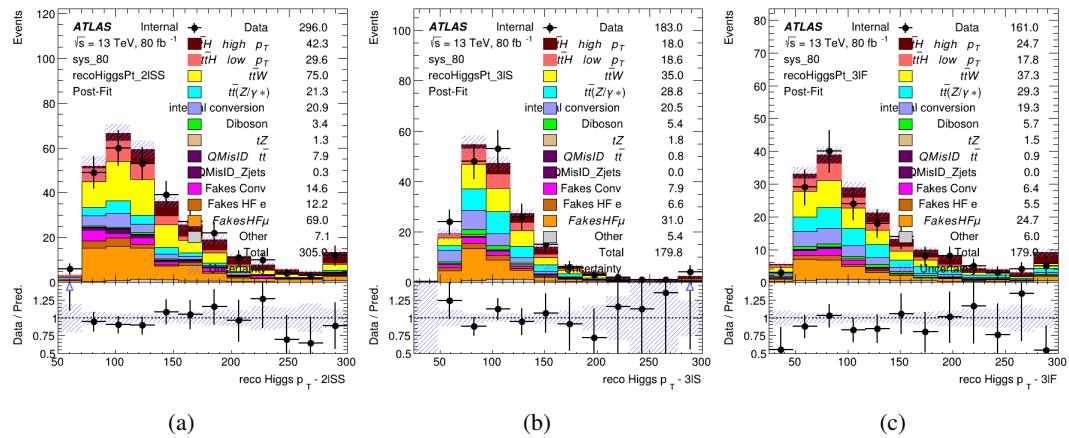


Figure 17.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 79.8 fb^{-1} of MC

1372 A post-fit summary of the fitted regions is shown in Figure 17.2.

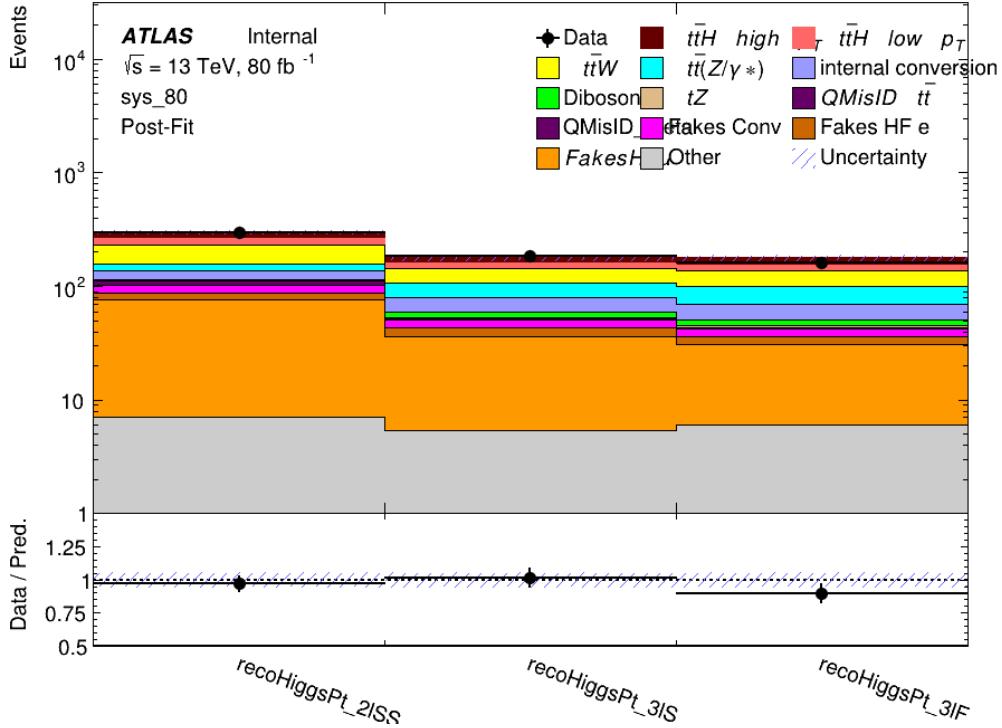


Figure 17.2: Post-fit summary of the yields in each signal region.

1373 The the measured μ values for high and low p_T Higgs production obtained from the fit
 1374 are shown in 40. A significance of 1.7σ is observed for $t\bar{t}H$ high p_T , and 2.1σ is measured for
 1375 $t\bar{t}H$ 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 40: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1376 The most prominent sources of systematic uncertainty, as measured by their impact on
 1377 $\mu_{t\bar{t}H \text{ high } p_T}$, are summarized in Table 41.

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

¹³⁷⁸ The most significant sources of uncertainty on the measurement of t̄H - low p_T are shown
¹³⁷⁹ in Table 42.

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

¹³⁸⁰ The ranking and impact of those nuisance parameters with the largest contribution to the

1381 overall uncertainty is shown in Figure 17.3.

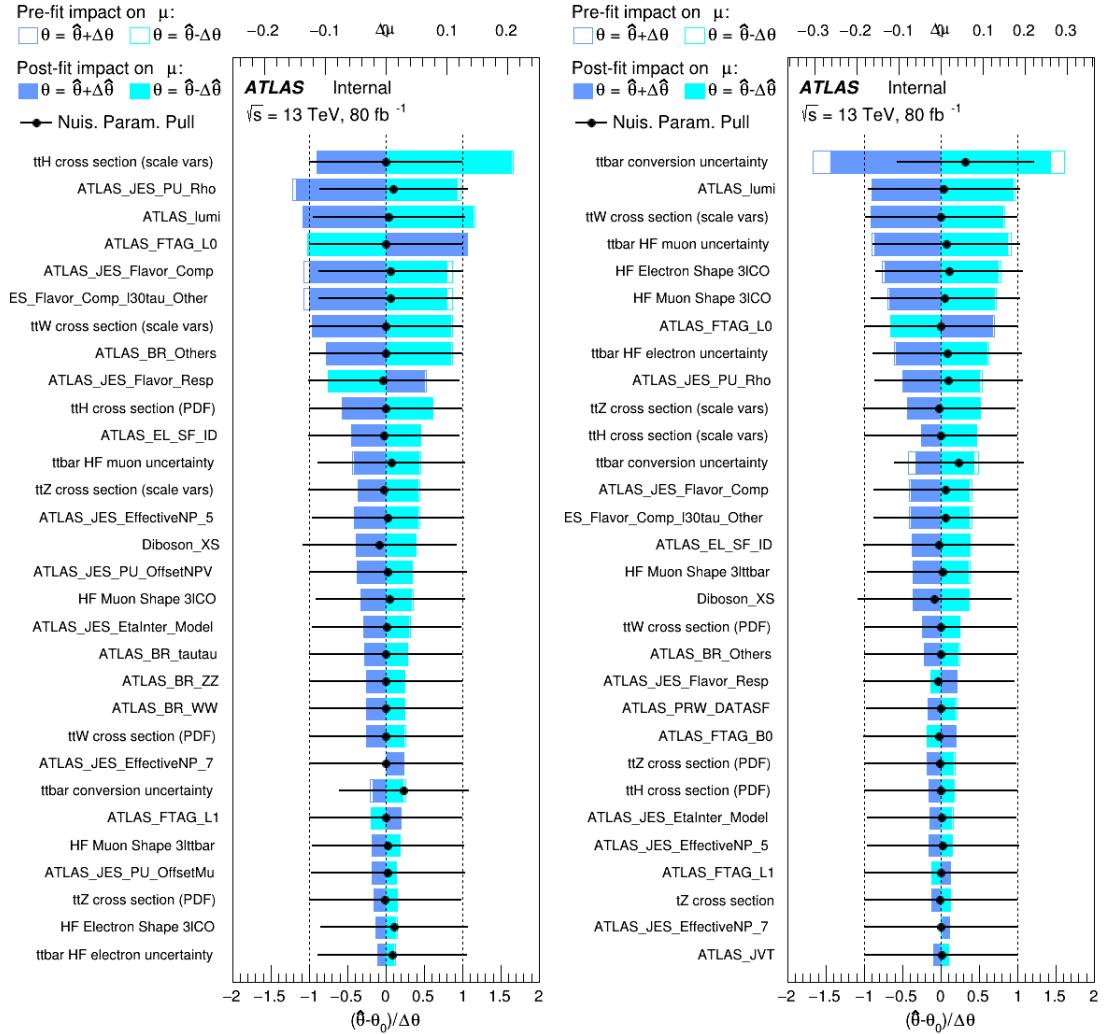


Figure 17.3: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

1382

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

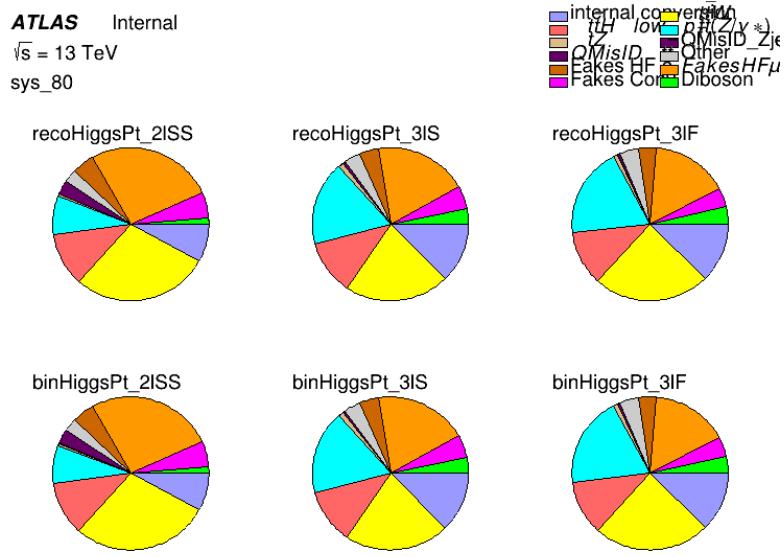


Figure 17.4: Background composition of the fit regions.

1383 17.2 Projected Results - 139 fb^{-1}

1384 As data collected in 2018 has not yet been unblinded for $t\bar{t}H - \text{ML}$ at the time of this note, data
 1385 from that year remains blinded. Instead, an Asimov fit is performed - with the MC prediction
 1386 being used both as the SM prediction as well as the data in the fit - in order to give expected
 1387 results.

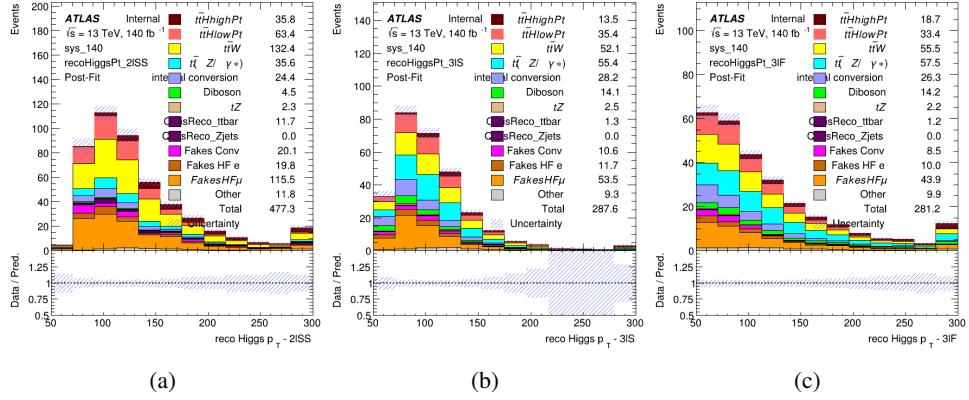


Figure 17.5: Blinded post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 139 fb^{-1} of data

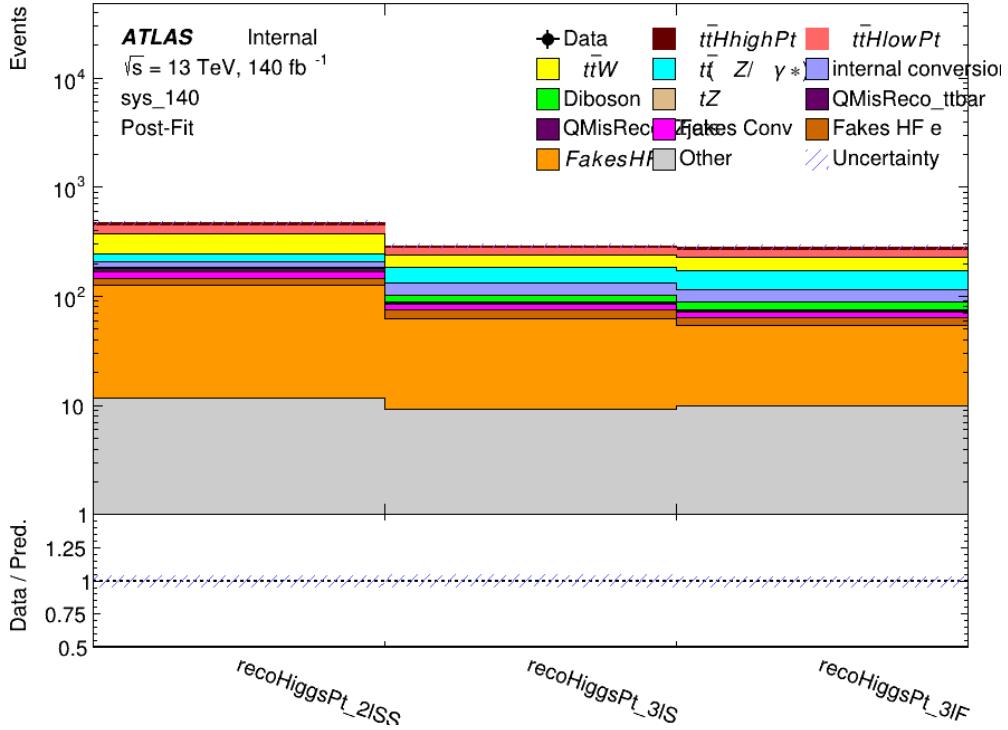


Figure 17.6: Post-fit summary of fit.

1389 shown in 43. A significance of 2.0σ is expected for $t\bar{t}H$ high p_T , and a projected significance

1390 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 43: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1391 The most prominent sources of systematic uncertainty, as measured by their impact on

1392 $\mu_{t\bar{t}H \text{ high } p_T}$, are summarized in Table 44.

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

1393 The most significant sources of systematic uncertainty on $t\bar{t}H$ low p_T are summarized in

1394 Table 45.

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

Table 45: Summary of the most significant sources of systematic uncertainty on the measurement of t̄tH low p_T.

¹³⁹⁵ The ranking and impact of those nuisance parameters with the largest contribution to the
¹³⁹⁶ overall uncertainty is shown in Figure 17.7.

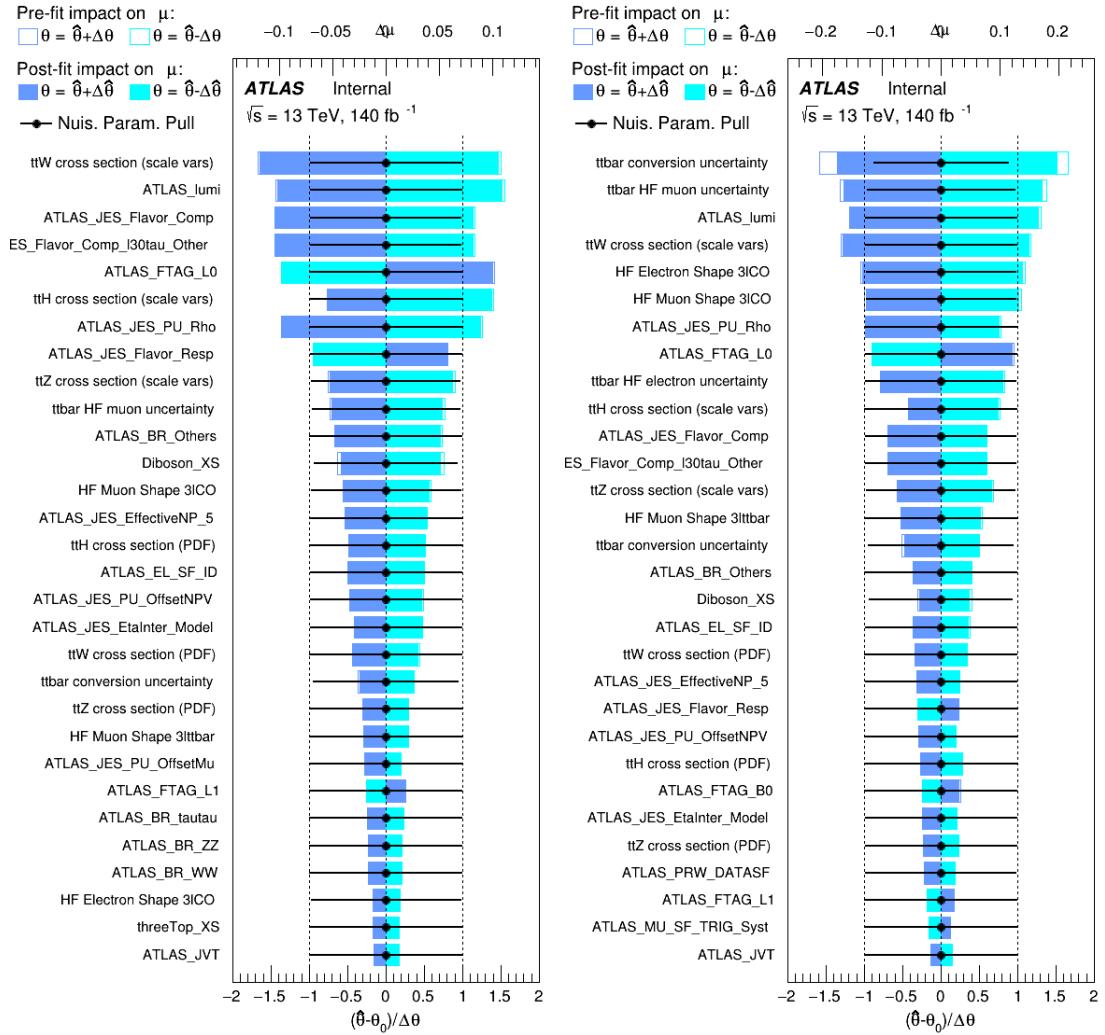


Figure 17.7: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

1397

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

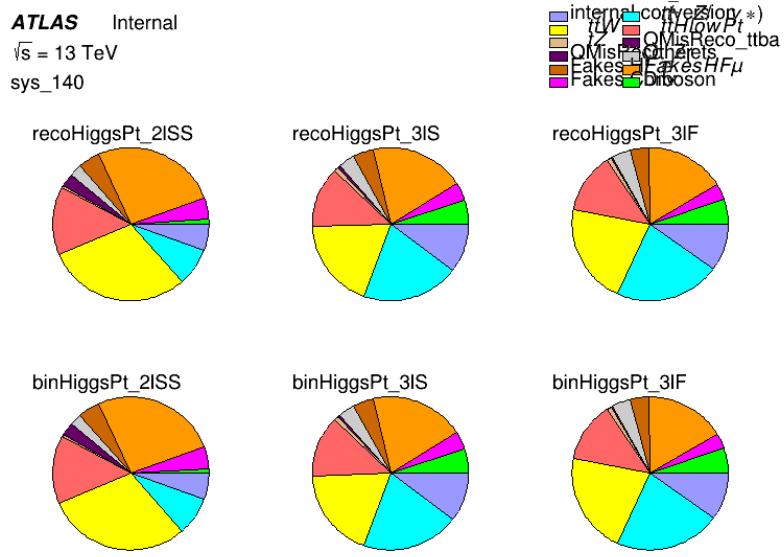


Figure 17.8: Background composition of the fit regions.

Part VI

Conclusion

1398 As search for the effects of dimension-six operators on $t\bar{t}H$ production is performed. An effective
 1400 field theory approached is used to parameterize the effects of high energy physics on the Higgs
 1401 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,
 1402 and the limits on dimension-six operators are limited to X.
 1403

1404 References

- 1405 [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard
1406 Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1,
1407 arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- 1408 [2] R. Oerter, *The theory of almost everything : the Standard Model, the unsung triumph of
1409 modern physics*, Pi Press, 2006, ISBN: 978-0-13-236678-6.
- 1410 [3] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,
1411 *Phys. Rev. Lett.* **13** (Oct, 1964) 508-509 ().
- 1412 [4] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*,
1413 *Phys. Rev.* **127** (3 1962) 965,
1414 URL: <https://link.aps.org/doi/10.1103/PhysRev.127.965>.
- 1415 [5] J. Ellis, *Higgs Physics*,
1416 (2013) 117, 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of
1417 High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report,
1418 URL: <https://cds.cern.ch/record/1638469>.
- 1419 [6] T. A. Collaboration, *Observation of Higgs boson production in association with a top
1420 quark pair at the LHC with the ATLAS detector*, *Physics Letters B* **784** (2018) 173,
1421 ISSN: 0370-2693, URL:
1422 <https://www.sciencedirect.com/science/article/pii/S0370269318305732>.

- 1423 [7] *Evidence for the associated production of the Higgs boson and a top quark pair with the*
1424 *ATLAS detector*, tech. rep. ATLAS-CONF-2017-077, CERN, 2017,
1425 URL: <https://cds.cern.ch/record/2291405>.
- 1426 [8] A. M. Sirunyan et al., *Observation of $t\bar{t}H$ Production*,
1427 *Physical Review Letters* **120** (2018), ISSN: 1079-7114,
1428 URL: <http://dx.doi.org/10.1103/PhysRevLett.120.231801>.
- 1429 [9] B. Dumont, S. Fichet and G. von Gersdorff,
1430 *A Bayesian view of the Higgs sector with higher dimensional operators*,
1431 *Journal of High Energy Physics* **2013** (2013), ISSN: 1029-8479,
1432 URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)065](http://dx.doi.org/10.1007/JHEP07(2013)065).
- 1433 [10] S. Banerjee, S. Mukhopadhyay and B. Mukhopadhyaya,
1434 *Higher dimensional operators and the LHC Higgs data: The role of modified kinematics*,
1435 *Physical Review D* **89** (2014), ISSN: 1550-2368,
1436 URL: <http://dx.doi.org/10.1103/PhysRevD.89.053010>.
- 1437 [11] M. Zinser, ‘The Large Hadron Collider’, *Search for New Heavy Charged Bosons and*
1438 *Measurement of High-Mass Drell-Yan Production in Proton—Proton Collisions*,
1439 Springer International Publishing, 2018 47, ISBN: 978-3-030-00650-1,
1440 URL: https://doi.org/10.1007/978-3-030-00650-1_4.
- 1441 [12] *Detector and Technology*, URL: <https://atlas.cern/discover/detector>.
- 1442 [13] M. Marjanovic, *ATLAS Tile calorimeter calibration and monitoring systems*, 2018.

- 1443 [14] M. Aaboud et al., *Observation of electroweak $W^\pm Z$ boson pair production in association*
 1444 *with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,*
 1445 **Phys. Lett. B** **793** (2019) 469, arXiv: [1812.09740 \[hep-ex\]](#).
- 1446 [15] T. Gleisberg et al., *Event generation with SHERPA 1.1*, **JHEP** **02** (2009) 007,
 1447 arXiv: [0811.4622 \[hep-ph\]](#).
- 1448 [16] R. D. Ball et al., *Parton distributions for the LHC Run II*, **JHEP** **04** (2015) 040,
 1449 arXiv: [1410.8849 \[hep-ph\]](#).
- 1450 [17] H.-L. Lai et al., *New parton distributions for collider physics*,
 1451 **Phys. Rev. D** **82** (2010) 074024, arXiv: [1007.2241 \[hep-ph\]](#).
- 1452 [18] S. Frixione, G. Ridolfi and P. Nason,
 1453 *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*,
 1454 **JHEP** **09** (2007) 126, arXiv: [0707.3088 \[hep-ph\]](#).
- 1455 [19] E. Re, *Single-top Wt-channel production matched with parton showers using the*
 1456 *POWHEG method*, **Eur. Phys. J. C** **71** (2011) 1547, arXiv: [1009.2450 \[hep-ph\]](#).
- 1457 [20] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using*
 1458 *the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016,
 1459 URL: <https://cds.cern.ch/record/2157687>.
- 1460 [21] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the*
 1461 *ATLAS detector using 2011 and 2012 LHC proton–proton collision data*,
 1462 **Eur. Phys. J. C** **74** (2014) 3130, arXiv: [1407.3935 \[hep-ex\]](#).

- 1463 [22] R. Narayan et al., *Measurement of the total and differential cross sections of a*
 1464 *top-quark-antiquark pair in association with a W boson in proton-proton collisions at a*
 1465 *centre-of-mass energy of 13 TeV with ATLAS detector at the Large Hadron Collider,*
 1466 tech. rep. ATL-COM-PHYS-2020-217, CERN, 2020,
 1467 URL: <https://cds.cern.ch/record/2712986>.
- 1468 [23] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets*
 1469 *Reconstructed in the ATLAS Detector at $\sqrt{s} = 13 \text{ TeV}$* , ATL-PHYS-PUB-2015-015,
 1470 2015, URL: <https://cds.cern.ch/record/2037613>.
- 1471 [24] ATLAS Collaboration,
 1472 *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector,*
 1473 ATLAS-CONF-2015-029, 2015, URL: <https://cds.cern.ch/record/2037702>.
- 1474 [25] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp*
 1475 *collisions at $\sqrt{s} = 8 \text{ TeV}$ using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581,
 1476 arXiv: [1510.03823 \[hep-ex\]](https://arxiv.org/abs/1510.03823).
- 1477 [26] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
 1478 *with the ATLAS detector in the first proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$* ,
 1479 ATL-PHYS-PUB-2015-027, 2015, URL: <https://cds.cern.ch/record/2037904>.
- 1480 [27] T. Chen and C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’,
 1481 *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*
 1482 *Discovery and Data Mining*, KDD ’16, ACM, 2016 785, ISBN: 978-1-4503-4232-2,
 1483 URL: <http://doi.acm.org/10.1145/2939672.2939785>.

- 1484 [28] F. Cardillo et al., *Measurement of the fiducial and differential cross-section of a top*
 1485 *quark pair in association with a Z boson at 13 TeV with the ATLAS detector*, (2019),
 1486 URL: <https://cds.cern.ch/record/2672207>.
- 1487 [29] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using*
 1488 *the ATLAS detector at the LHC*, *Eur. Phys. J. C* **71** (2011) 1630,
 1489 arXiv: [1101.2185 \[hep-ex\]](https://arxiv.org/abs/1101.2185).
- 1490 [30] G. Avoni et al.,
 1491 *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*,
 1492 *JINST* **13** (2018) P07017.
- 1493 [31] G. Aad, T. Abajyan, B. Abbott and etal, *Jet energy resolution in proton-proton collisions*
 1494 *at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*,
 1495 *The European Physical Journal C* **73** (2013) 2306, ISSN: 1434-6052,
 1496 URL: <https://doi.org/10.1140/epjc/s10052-013-2306-0>.
- 1497 [32] A. Collaboration, *Performance of b -jet identification in the ATLAS experiment*,
 1498 *Journal of Instrumentation* **11** (2016) P04008,
 1499 URL: <http://stacks.iop.org/1748-0221/11/i=04/a=P04008>.
- 1500 [33] *Observation of the associated production of a top quark and a Z boson at 13 TeV with*
 1501 *ATLAS*, (2020), URL: <https://cds.cern.ch/record/2722504>.
- 1502 [34] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions from Powheg*
 1503 *and Sherpa to ATLAS measurements of top pair production at 7 TeV*,
 1504 ATL-PHYS-PUB-2015-011, 2015, URL: <https://cds.cern.ch/record/2020602>.

- 1505 [35] ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11,*
1506 ATL-PHYS-PUB-2011-009, 2011, URL: <https://cds.cern.ch/record/1363300>.

List of contributions

1508

₁₅₀₉ **Part VII**

₁₅₁₀ **Appendices**

₁₅₁₁ **A Non-prompt lepton MVA**

1512 A lepton MVA has been developed to better reject non-prompt leptons than standard cut
1513 based selections based upon impact parameter, isolation and PID. The name of this MVA is
1514 **PromptLeptonIso**. The full set of studies and detailed explanation can be found in [22].

1515 The decays of W and Z bosons are commonly selected by the identification of one or two
1516 electrons or muons. The negligible lifetimes of these bosons mean that the leptons produced in the
1517 decay originate from the interaction vertex and are thus labelled “prompt”. Analyses using these
1518 light leptons impose strict reconstruction quality, isolation and impact parameter requirements
1519 to remove “fake” leptons. A significant source of the fake light leptons are non-prompt leptons
1520 produced in decays of hadrons that contain bottom (b) or charm (c) quarks. Such hadrons
1521 typically have microscopically significant lifetimes that can be detected experimentally.

1522 These non-prompt leptons can also pass the tight selection criteria. In analyses that
1523 involve top (t) quarks, which decay almost exclusively into a W boson and a b quark, non-
1524 prompt leptons from the semileptonic decay of bottom and charm hadrons can be a significant
1525 source of background events. This is particularly the case in the selection of same-sign dilepton
1526 and multilepton final states.

1527 The main idea is to identify non-prompt light leptons using lifetime information associated
1528 with a track jet that matches the selected light lepton. This lifetime information is computed
1529 using tracks contained within the jet. Typically, lepton lifetime is determined using the impact
1530 parameter of the track reconstructed by the inner tracking detector which is matched to the recon-
1531 structed lepton. Using additional reconstructed charged particle tracks increases the precision of

1532 identifying the displaced decay vertex of bottom or charm hadrons that produced a non-prompt
 1533 light lepton. The MVA also includes information related to the isolation of the lepton to reject
 1534 non-prompt leptons.

1535 **PromptLeptonIso** is a gradient boosted BDT. The training of the BDT is performed on
 1536 leptons selected from the PowHEG+PYTHIA6 non-allhad $t\bar{t}$ MC sample. Eight variables are used
 1537 to train the BDT in order to discriminate between prompt and non-prompt leptons. The track
 1538 jets that are matched to the non-prompt leptons correspond to jets initiated by b or c quarks, and
 1539 may contain a displaced vertex. Consequently, three of the selected variables are used to identify
 1540 b-tag jets by standard ATLAS flavour tagging algorithms. Two variables use the relationship
 1541 between the track jet and lepton: the ratio of the lepton p_T with respect to the track jet p_T and
 1542 ΔR between the lepton and the track jet axis. Finally three additional variables test whether the
 1543 reconstructed lepton is isolated: the number of tracks collected by the track jet and the lepton
 1544 track and calorimeter isolation variables. Table 46 describes the variables used to train the BDT
 1545 algorithm. The choice of input variables has been extensively discussed with Egamma, Muon,
 1546 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 46: A table of the variables used in the training of **PromptLeptonIso**.

1547 The output distribution of the BDT is shown in Figure A.

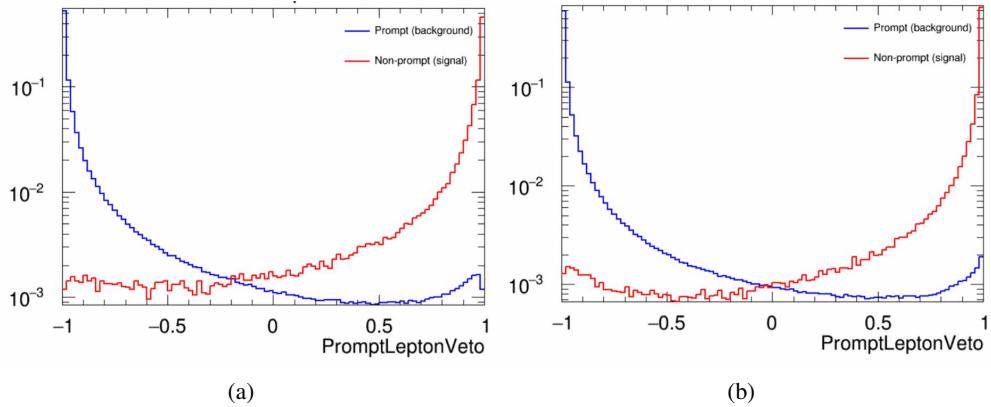


Figure A.1: Distribution of the PLV BDT discriminant for (a) electrons and (b) muons

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 alternate training.

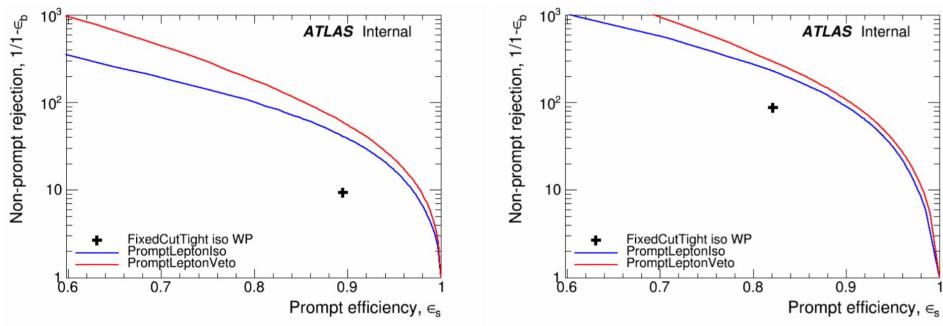


Figure A.2: ROC curves for the PLV as well as the performance of the standard FixedCutTight WP for (left) electrons and (right) muons

1550 A cutoff value of -0.7 for electrons and -0.5 for muons are chosen as the WPs for this
 1551 MVA, based on an optimisation of S/\sqrt{B} performed in the preselection regions of the $t\bar{t}H - ML$
 1552 analysis, which have a signature similar to that of this analysis.

1553 The efficiency of the tight `PromptLeptonIso` working point is measured using the tag
1554 and probe method with $Z \rightarrow \ell^+ \ell^-$ events. Such calibration are performed by analysers from
1555 this analysis in communication with the Egamma and Muon combined performance groups. The
1556 scale factor are approximately 0.92 for $10 < p_T < 15$ GeV, and averaging at 0.98 to 0.99 for
1557 higher p_T leptons. An extra systematic is applied to muons within $\Delta R < 0.6$ of a calorimeter
1558 jet, since there is a strong dependence on the scale factor due to the presence of these jets. For
1559 electrons, the dominant systematics is coming from pile-up dependence. Overall the systematics
1560 are a maximum of 3% at low p_T and decreasing at a function of p_T .

¹⁵⁶¹ **B Supplementary WZ + Heavy Flavor Studies**

¹⁵⁶² **B.1 Non-prompt CR Modelling**

1563 In order to further validate the modeling in each of the non-prompt CRs, additional
 1564 kinematic plots are made in the Z+jets CR and $t\bar{t}$ CR in each of the continuous b-tag regions,
 1565 after the correction factors detailed in Section 8.3 have been applied.

1566 In the case of the Z+jets CR, the p_T spectrum of the lepton originating from the W
 1567 candidate is shown, as this is the distribution used to extract the scale factor applied to Z+jets.
 1568 These plots are shown in Figures B.1 and B.2.

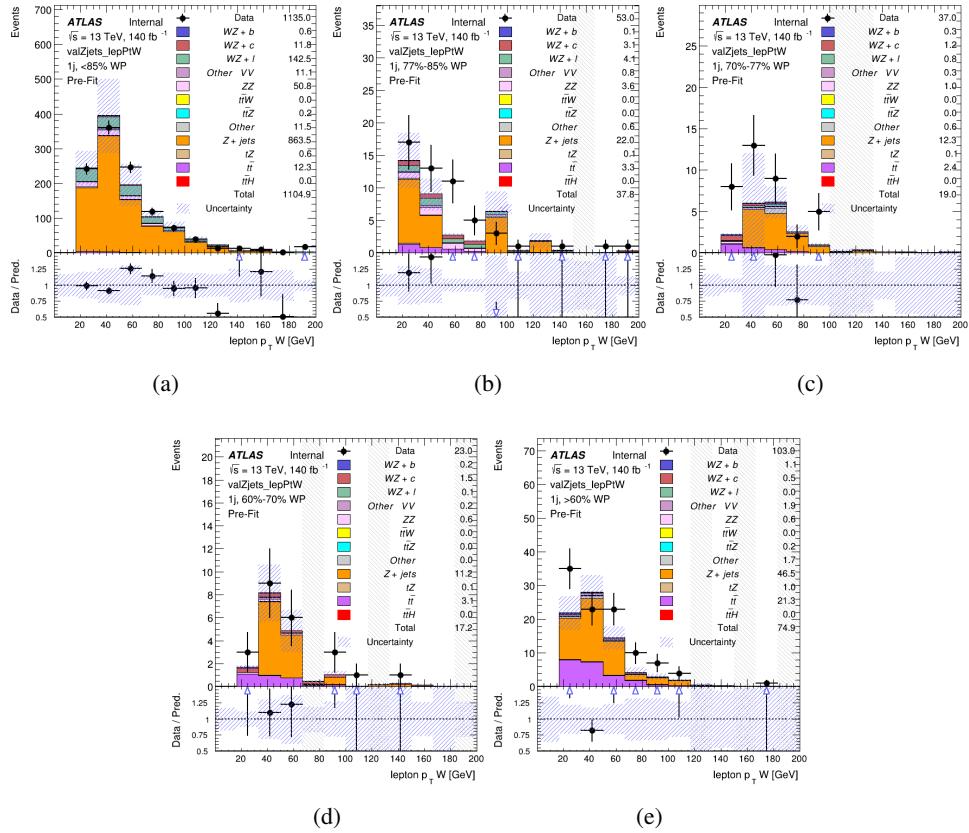


Figure B.1: Comparisons between the data and MC distributions of the p_T of the lepton originating from the W-candidate in the Z+jets CR for each of the 1-jet b-tag working point regions

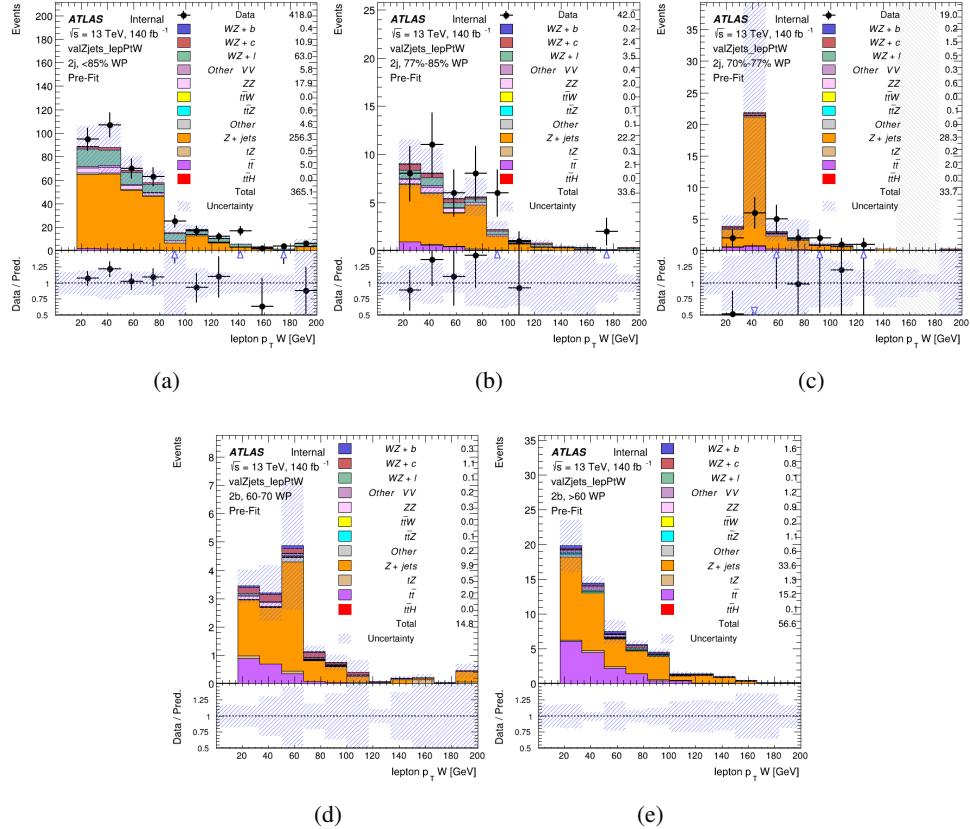


Figure B.2: Comparisons between the data and MC distributions of the p_T of the lepton originating from the W-candidate in the Z+jets CR for each of the 2-jet b-tag working point regions

1569 The same is shown for the $t\bar{t}$ CR, but the p_T of the OS lepton is used instead as a
 1570 representation of the modeling, as the lepton from the W is not well defined for $t\bar{t}$ events. These
 1571 plots are shown in Figures B.3 and B.4.

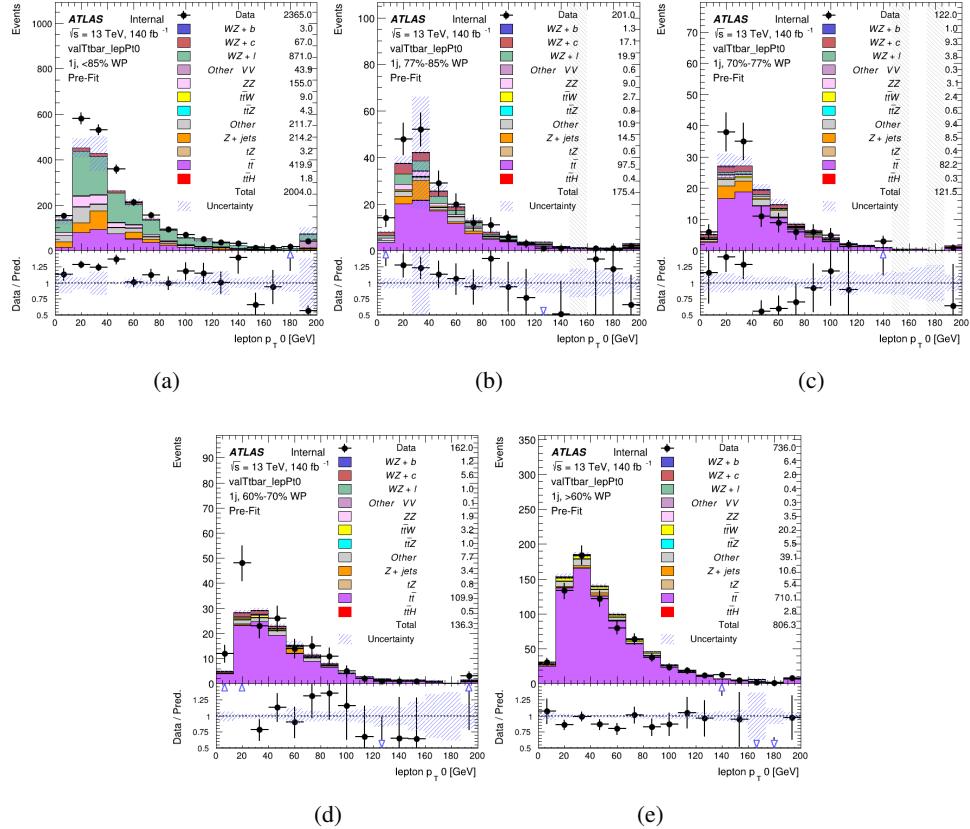


Figure B.3: Comparisons between the data and MC distributions of the p_T of the OS lepton in the $t\bar{t}$ CR for each of the 1-jet b-tag working point regions

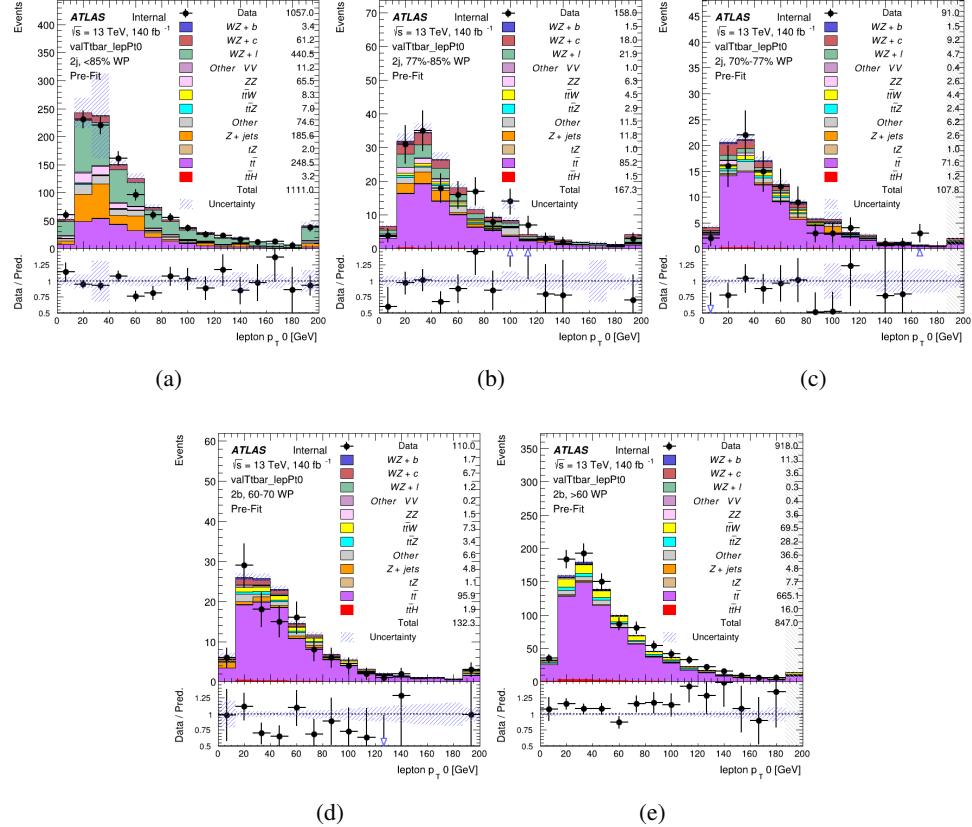


Figure B.4: Comparisons between the data and MC distributions of the p_T of the OS lepton in the $t\bar{t}$ CR for each of the 2-jet b-tag working point regions

¹⁵⁷² **B.2 tZ Interference Studies**

1573 Because it includes an on-shell Z boson as well as a b-jet and W from the top decay, tZ
1574 production represents an identical final state to WZ + b-jet. This implies the possibility of matrix
1575 level interference between these two processes not accounted for in the Monte Carlo simulations,
1576 which consider the two processes independently. Truth level studies are performed in order to
1577 estimate the impact of these interference effects.

1578 In order to estimate the matrix level interference effects between tZ and WZ + b-jet, two
1579 different sets of simulations are produced using MadGraph 5 [**Madgraph**] - one which simulates
1580 these two processes independently, and another where they are produced simultaneously, such
1581 that interference effects are present. These two sets of samples are then compared, and the
1582 difference between them can be taken to represent any interference effects.

1583 MadGraph simulations of 10,000 tZ and 10,000 WZ + b-jet events are produced, along
1584 with 20,000 events where both are present, in the fiducial region where three leptons and at least
1585 one jet are produced.

1586 A selection mimicking the preselection used in the main analysis is applied to the samples:
1587 The SS leptons are required to have $p_T > 20$ GeV, and > 10 GeV is required for the OS lepton.
1588 The associated b-jet is required to have $p_T > 25$ GeV, and all physics objects are required to fall
1589 in a range of $|\eta| < 2.5$.

1590 The kinematics of these samples after the selection has been applied are shown below:

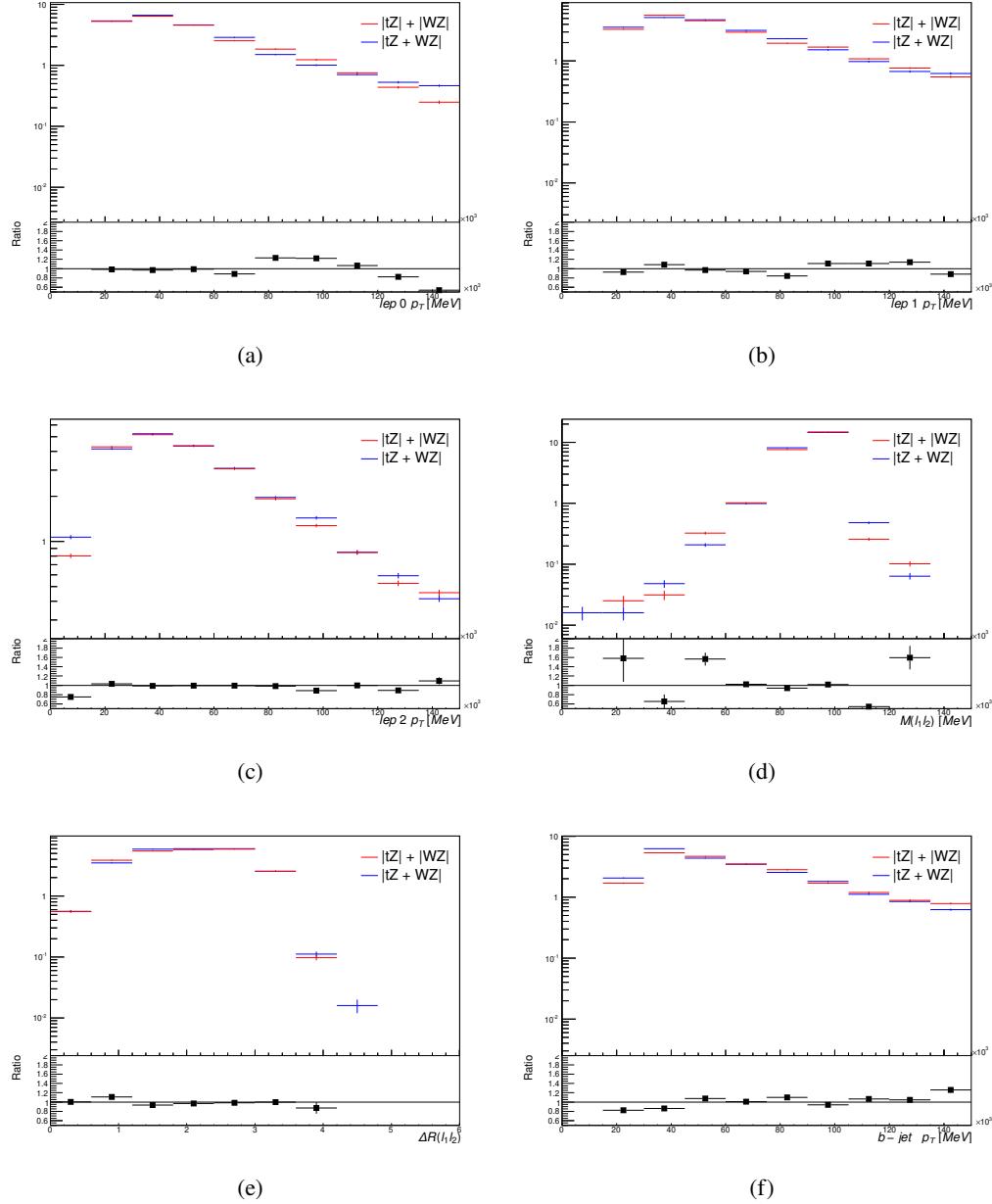


Figure B.5: Comparisons between (a) the p_T of the lepton from the W, (b) and (c) show the p_T of the lepton from the Z, (d) the invariant mass of the Z-candidate, (e) ΔR of the leptons from the Z, and (f) the p_T of the b-jet, for WZ and tZ events generated with interference effects (blue) and without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

1591 The overall cross-section of the two methods agree within error, and no significant differ-
1592 ences in the kinematic distributions are seen. It is therefore concluded that interference effects
1593 do not significantly impact the results.

1594 **B.3 Alternate tZ Inclusive Fit**

1595 **B.3.1 tZ Inclusive Fit**

1596 While tZ is often considered as a distinct process from WZ + b, this could also be considered part
 1597 of the signal. Alternate studies are performed where, using the same framework as the nominal
 1598 analysis, a measurement of WZ + b is performed that includes tZ as part of WZ+b.

1599 Because of this change, the tZ CR is no longer necessary, and only the five pseudo-
 1600 continuous b-tag regions are used in the fit. Further, systematics related to the tZ cross-section
 1601 are removed from the fit, as they are now encompassed by the normalization measurement of
 1602 WZ + b. All other systematic uncertainties are carried over from the nominal analysis.

1603 An expected WZ + b cross-section of $4.1^{+0.78}_{-0.74}(\text{stat})^{+0.53}_{-0.52}(\text{sys}) \text{ fb}$ is extracted from the fit,
 1604 with an expected significance of 4.0σ .

1605 The impact of the predominate systematics are summarized in Table 47.

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

Table 47: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

1606 **B.3.2 Floating tZ**

1607 In order to quantify the impact of the tZ uncertainty on the fit, an alternate fit strategy is used
1608 where the tZ normalization is allowed to float. This normalization factor replaces the cross-
1609 section uncertainty on tZ, and all other parameters of the fit remain the same.

1610 An uncertainty of 17% on the normalization of tZ is extracted from the fit, compared to a
1611 theory uncertainty of 15% applied to the tZ cross-section. The measured uncertainties on WZ
1612 remain the same.

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

1613 The following section provides details of the various MVAs as well as a few studies performed
1614 in support of this analysis, exploring alternate decisions and strategies.

1615 **C.1 Higgs Reconstruction Model Details**

1616 **C.1.1 b-jet Identification Features - 2lSS**

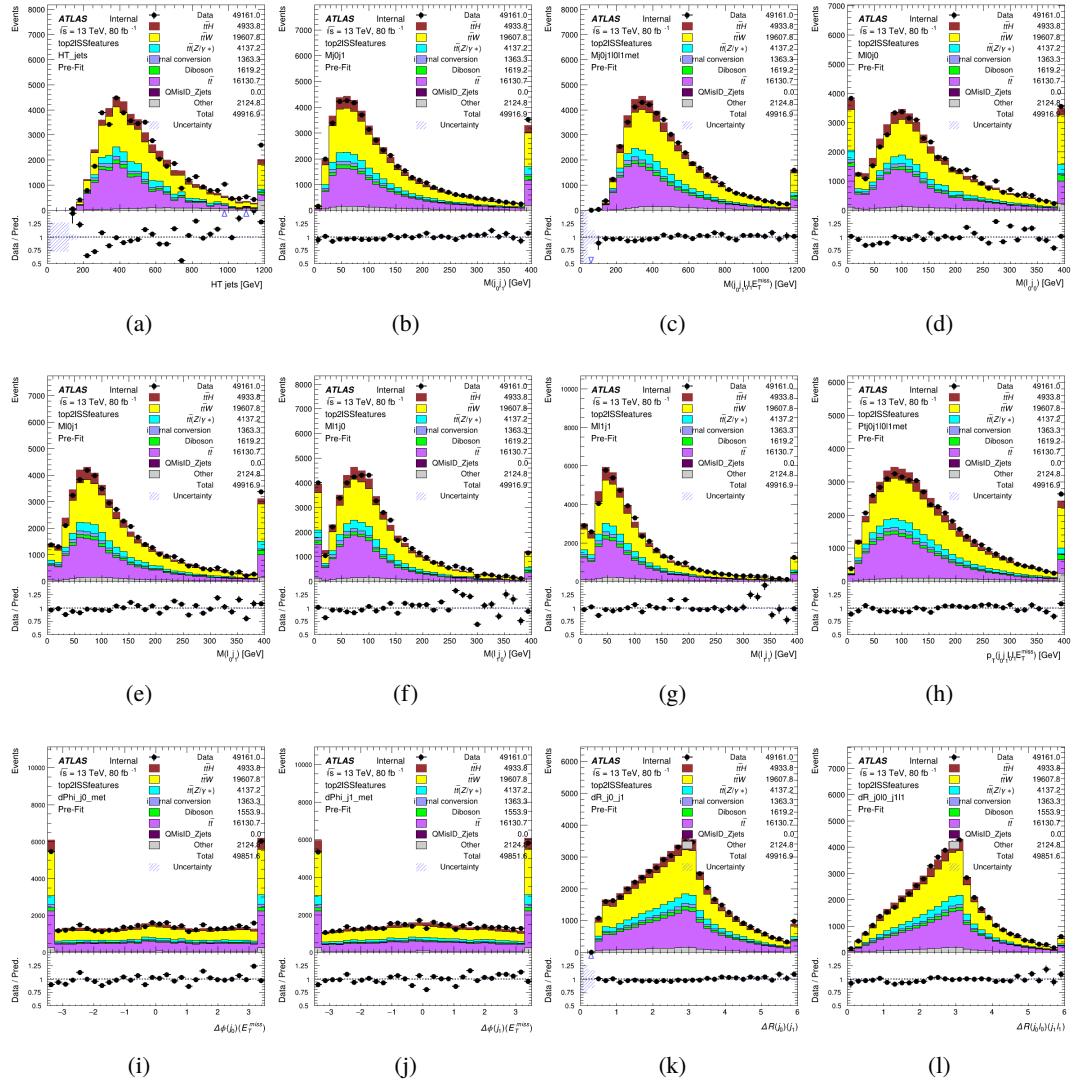


Figure C.1: Input features for top2ISS

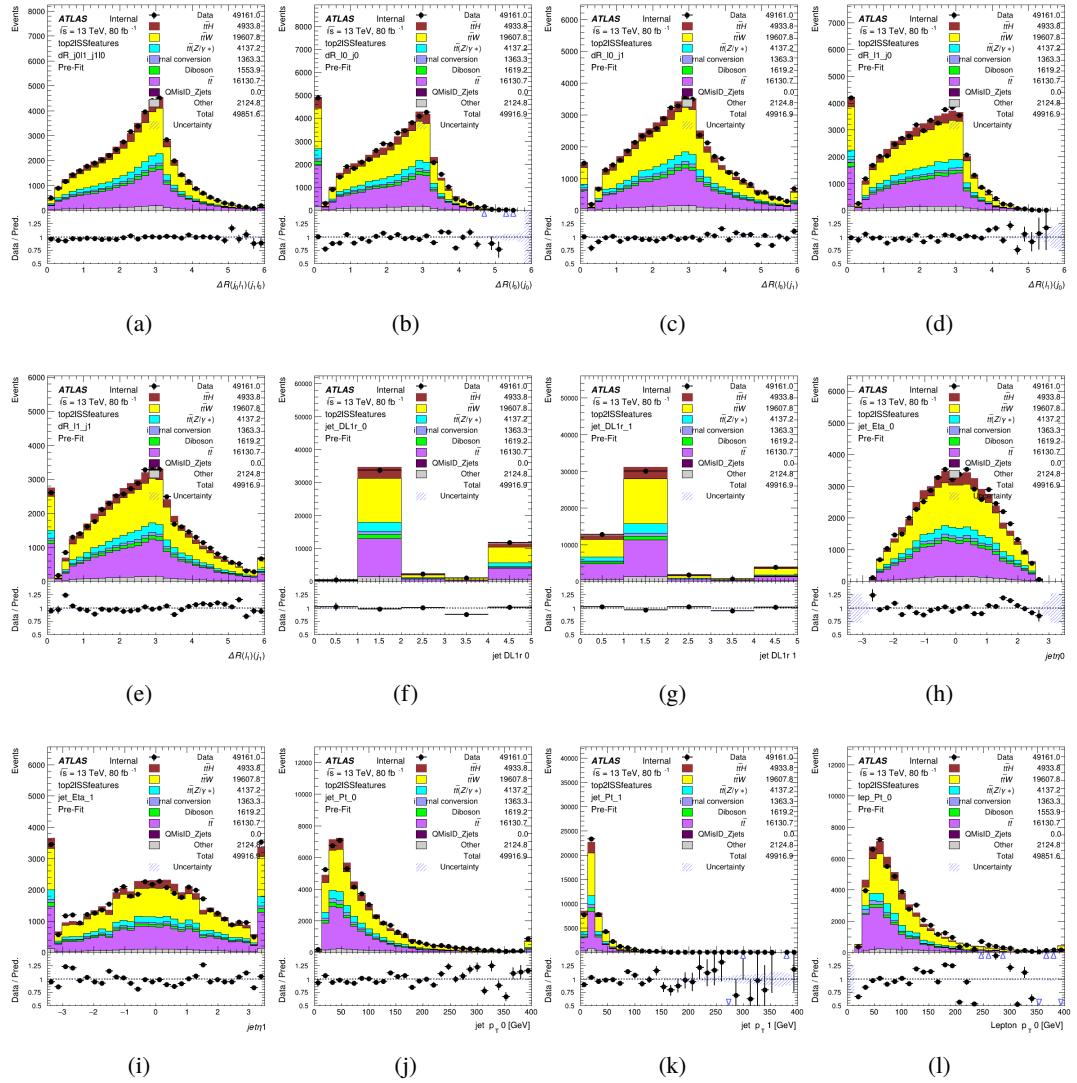


Figure C.2: Input features for top2lSS

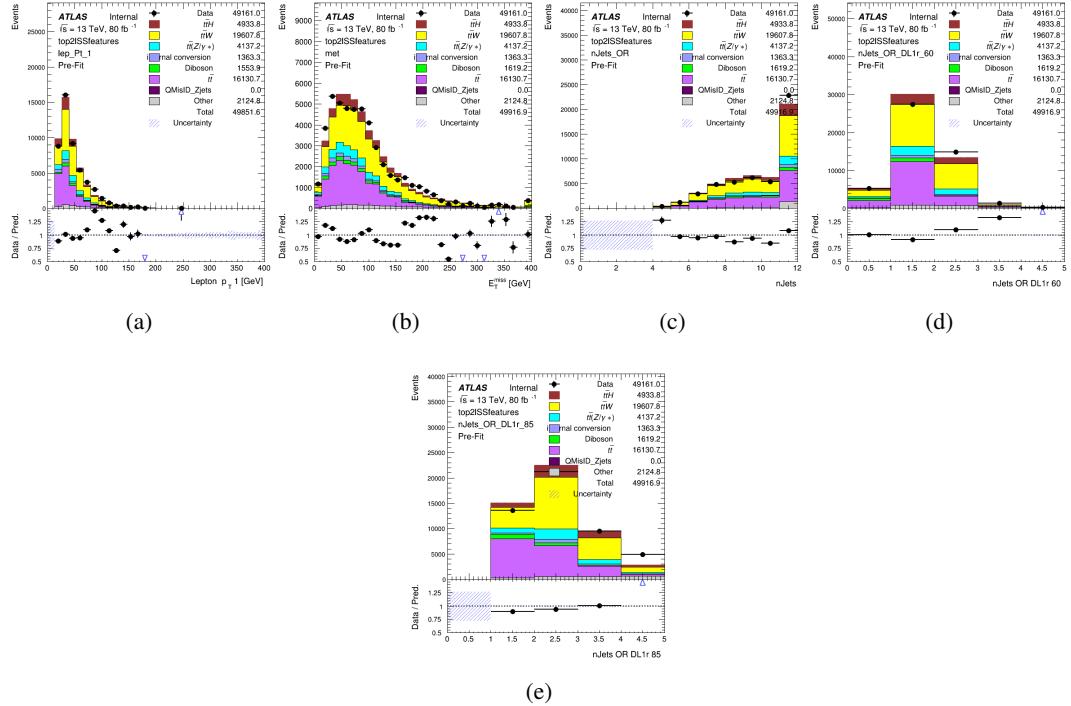


Figure C.3: Input features for top2ISS

1618 **C.1.2 b-jet Identification Features - 3l**

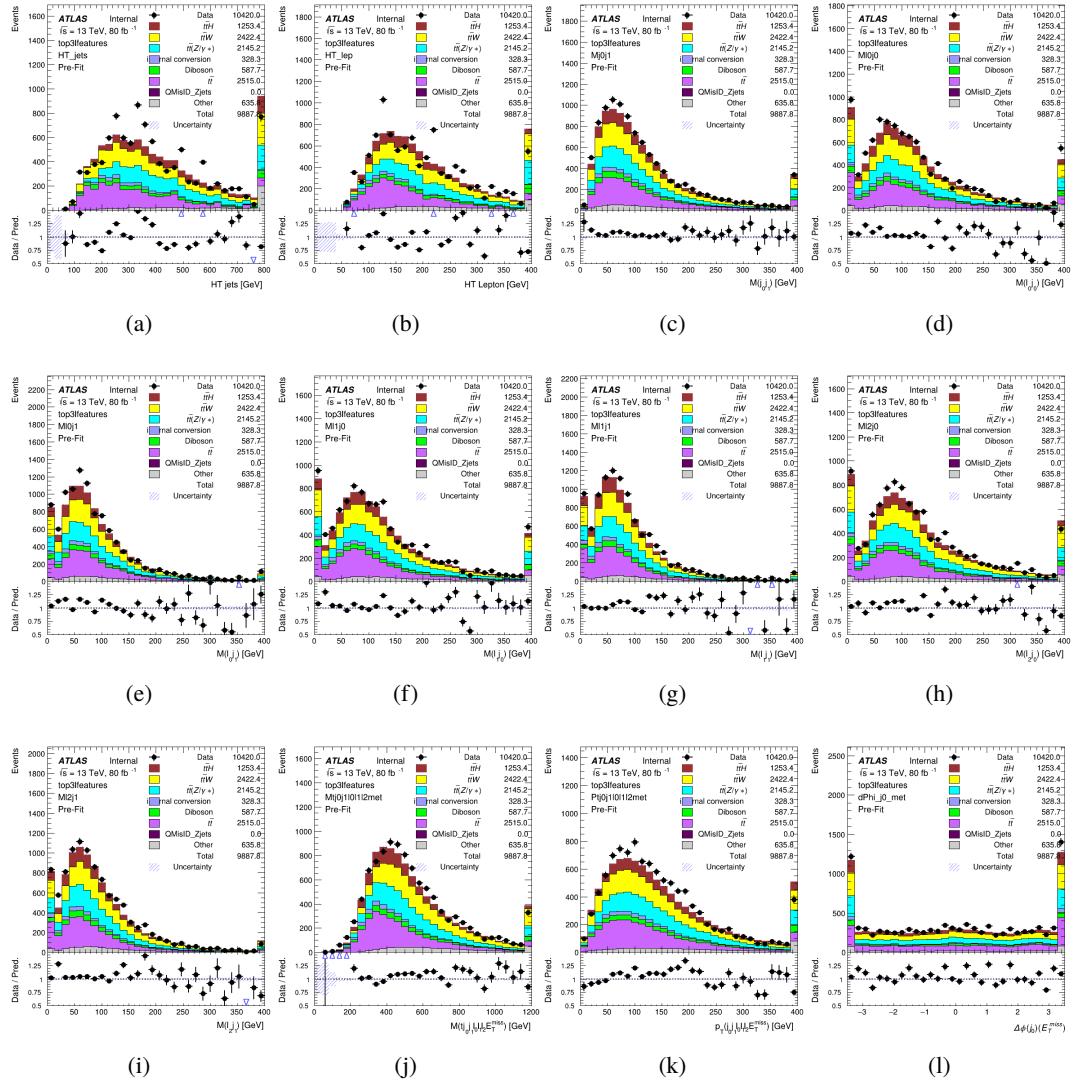


Figure C.4: Input features for top31

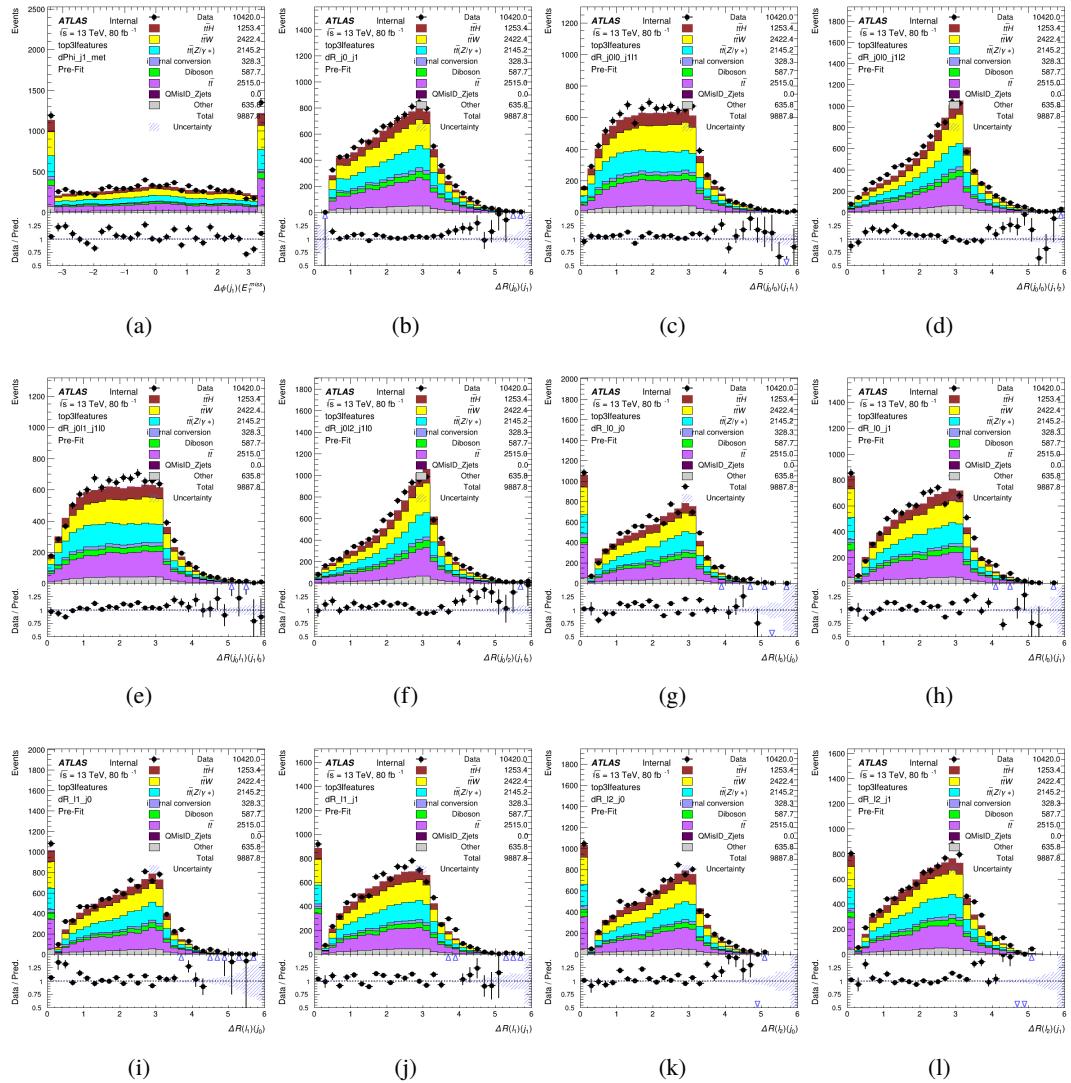


Figure C.5: Input features for top31

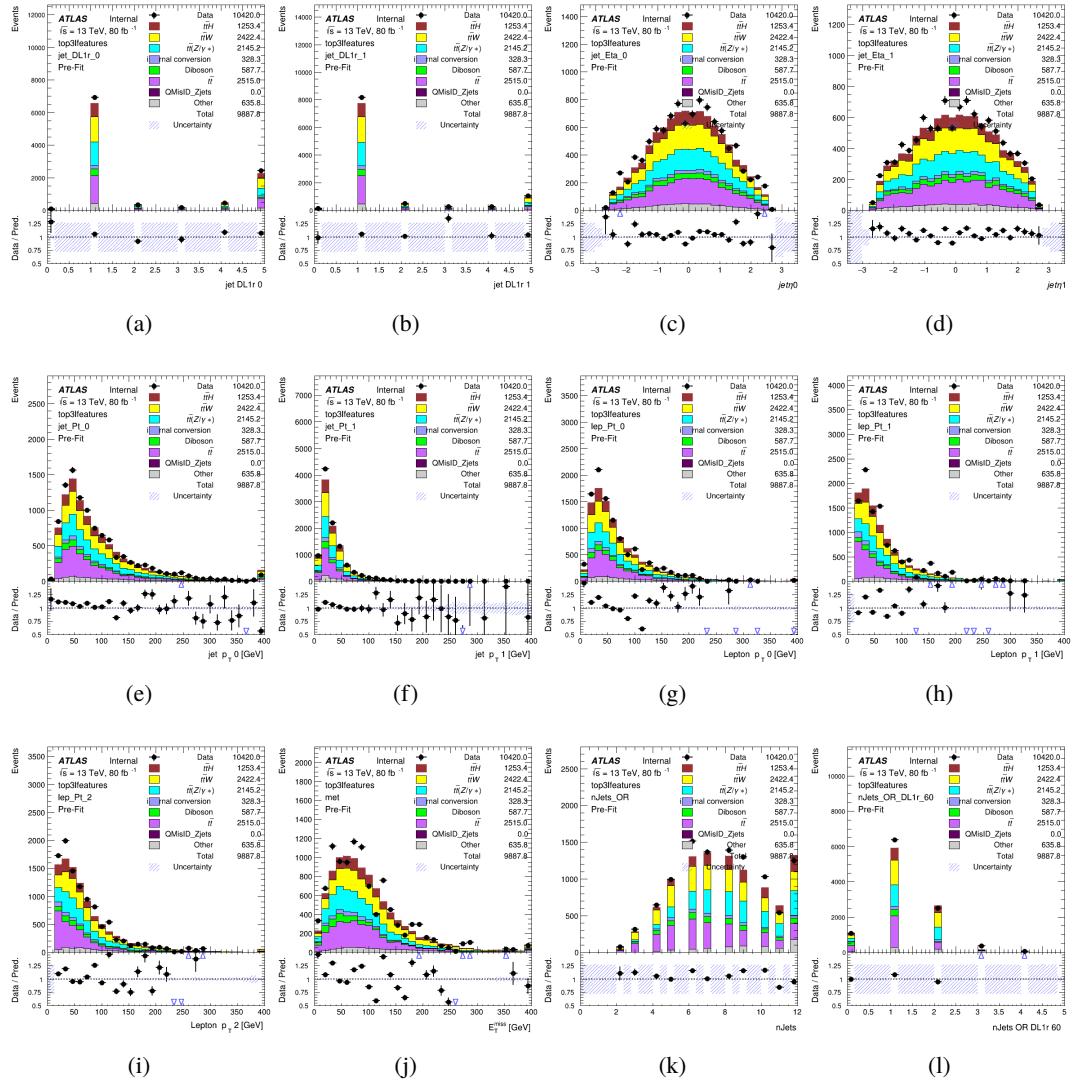
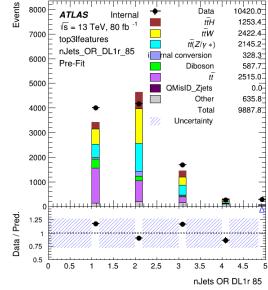


Figure C.6: Input features for top3l



(a)

Figure C.7: Input features for top3l

1619 **C.1.3 Higgs Reconstruction Features - 2lSS**

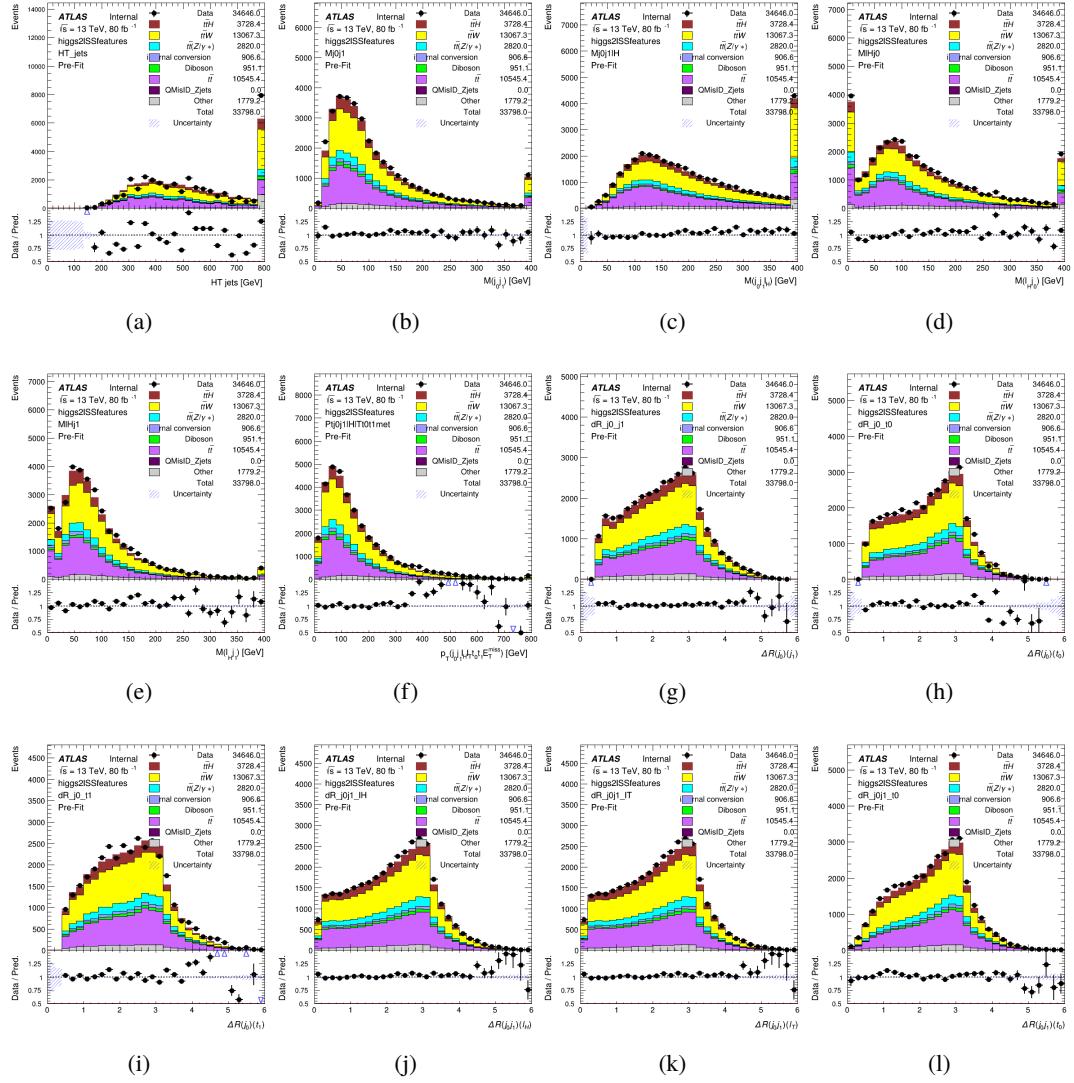


Figure C.8: Input features for higgs2ISST

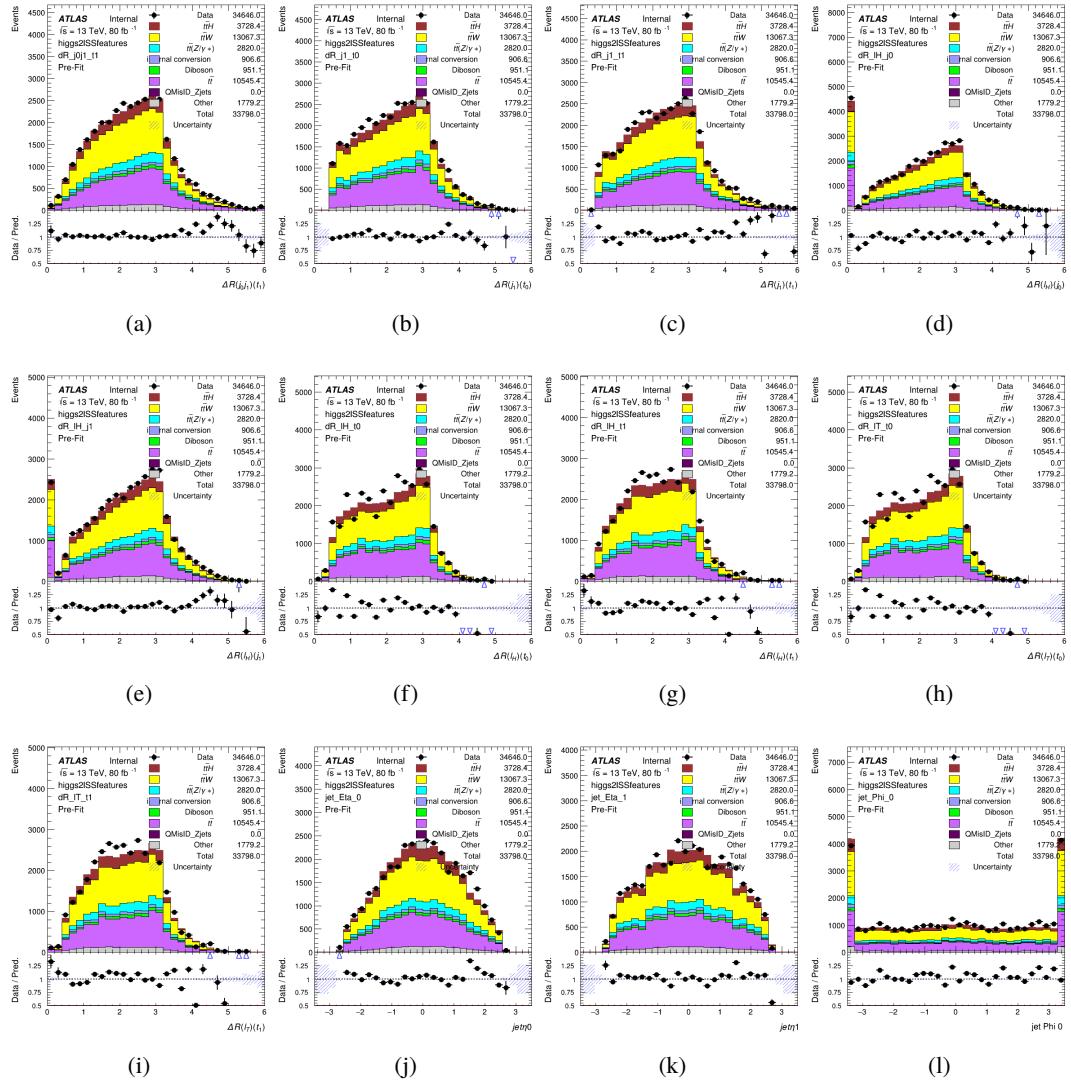


Figure C.9: Input features for higgs2lSS

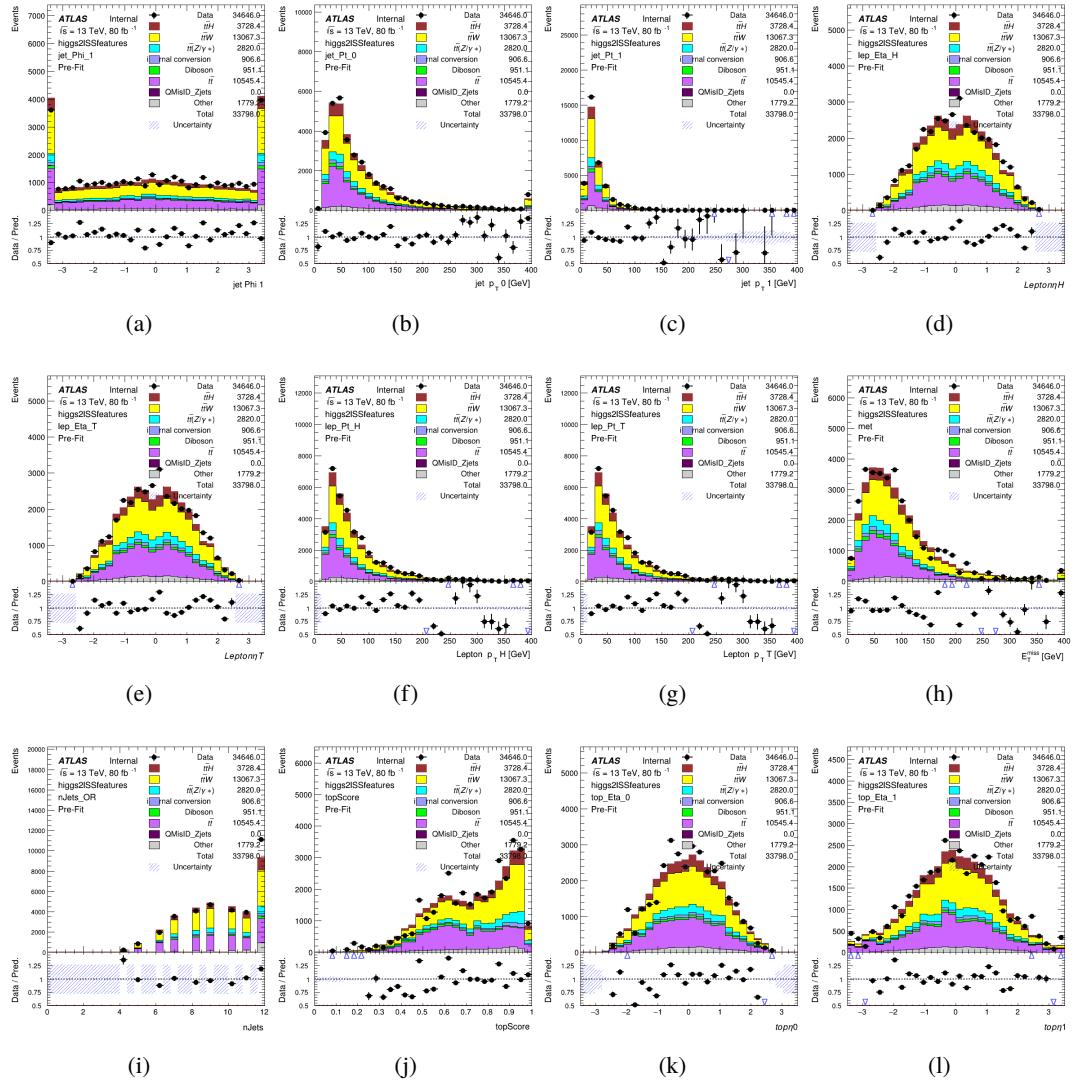


Figure C.10: Input features for higgs2lSS

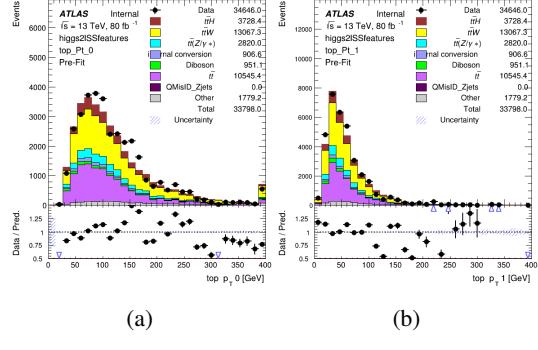


Figure C.11: Input features for higgs2lSS

1620 **C.1.4 Higgs Reconstruction Features - 3lS**

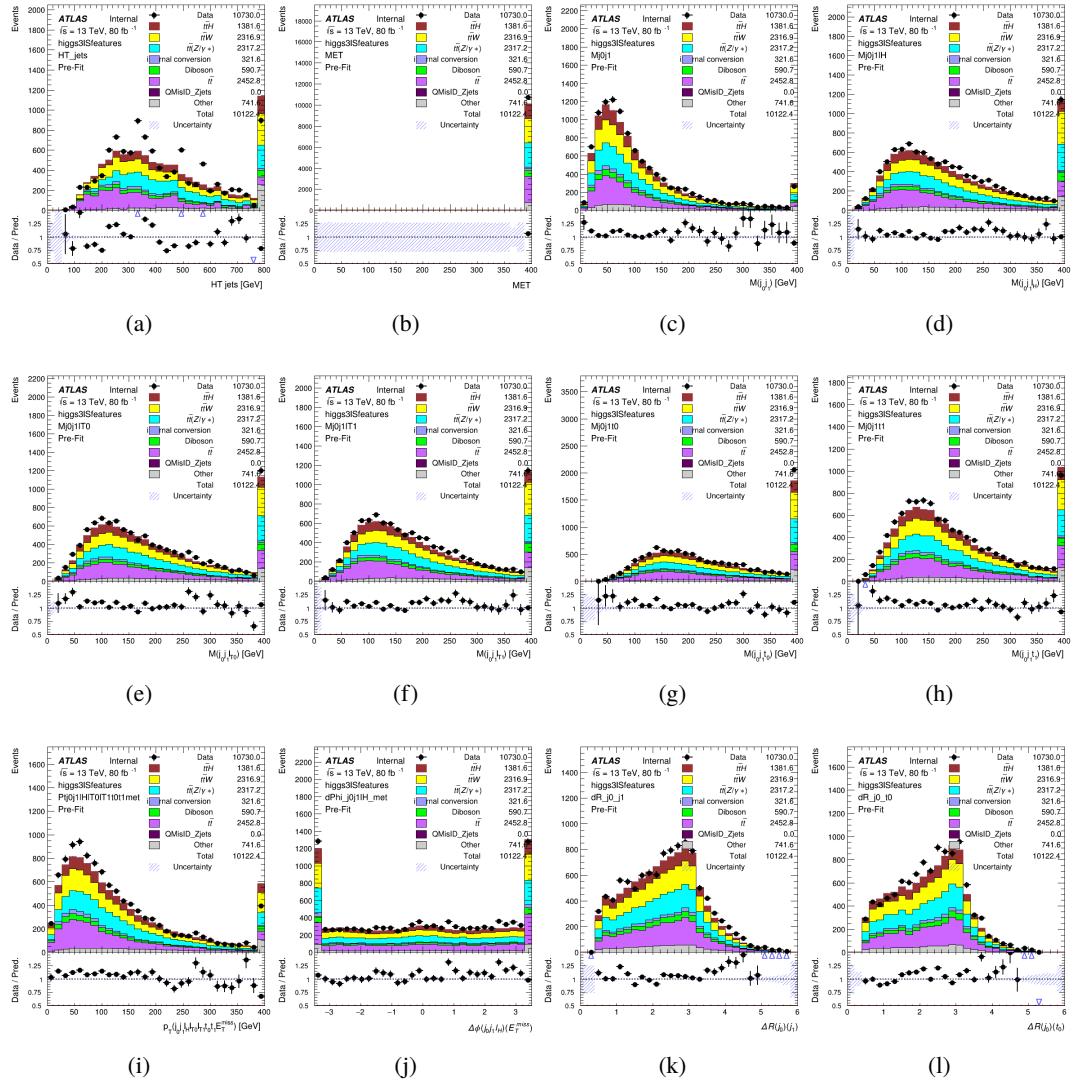


Figure C.12: Input features for higgs3IS

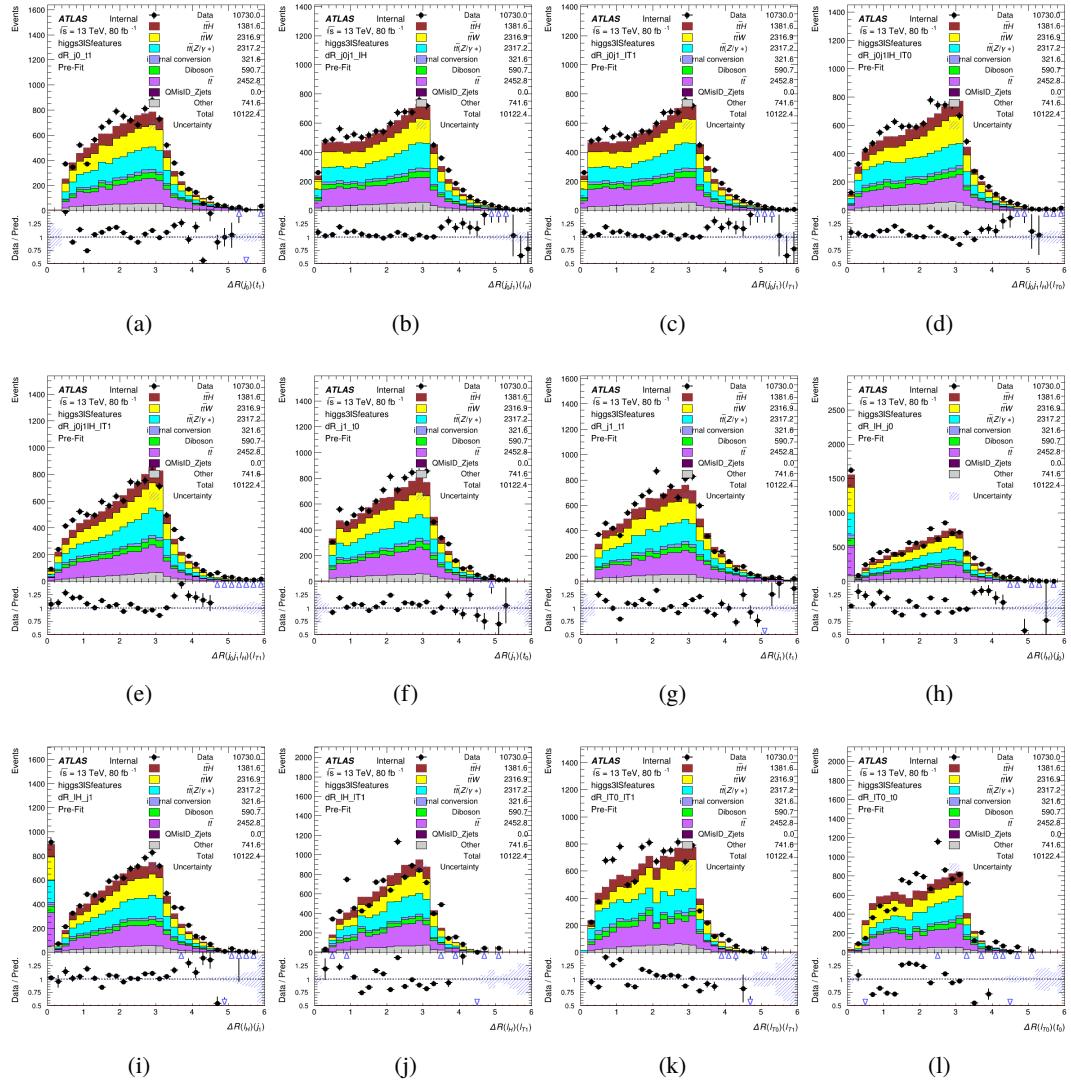


Figure C.13: Input features for higgs3lS

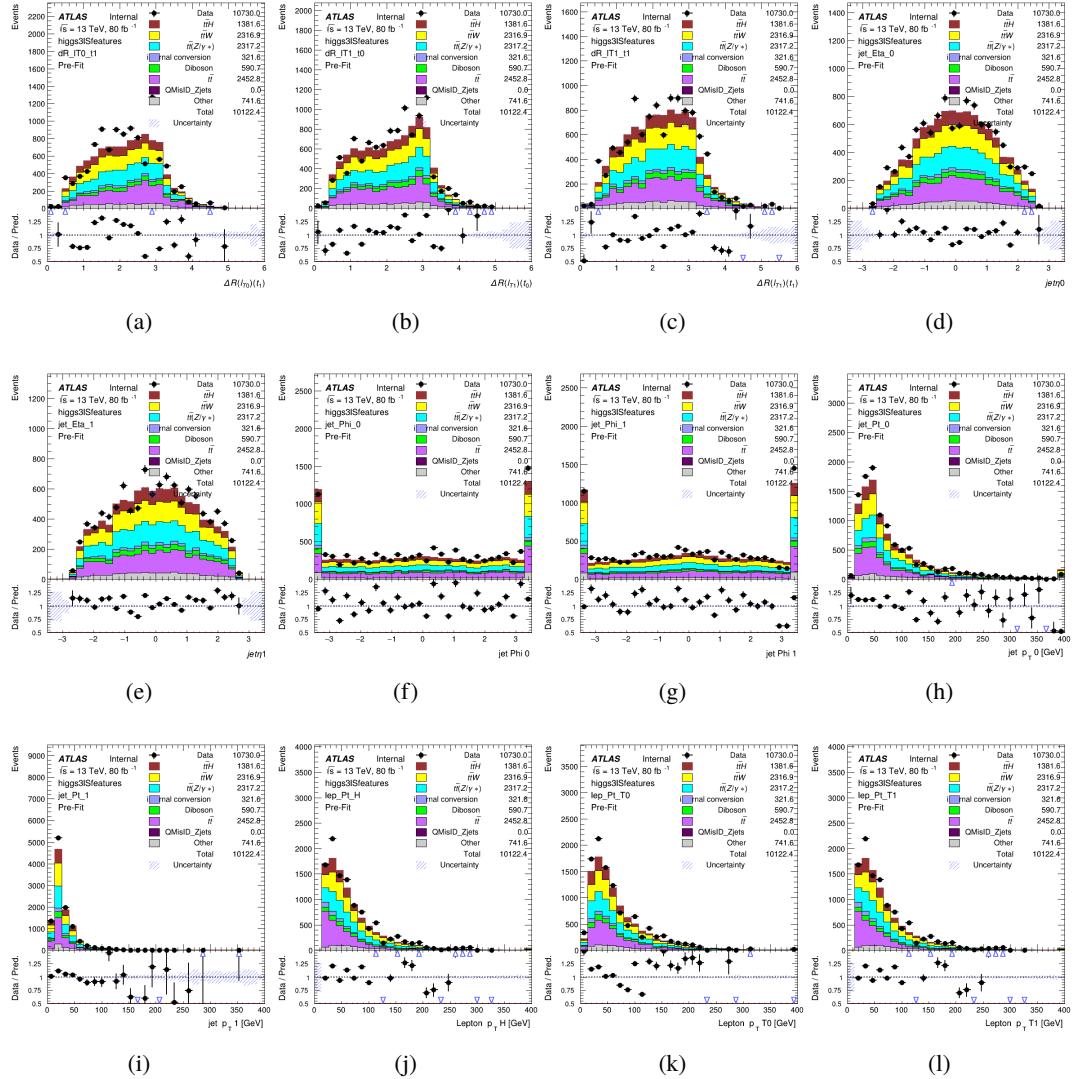


Figure C.14: Input features for higgs3IS

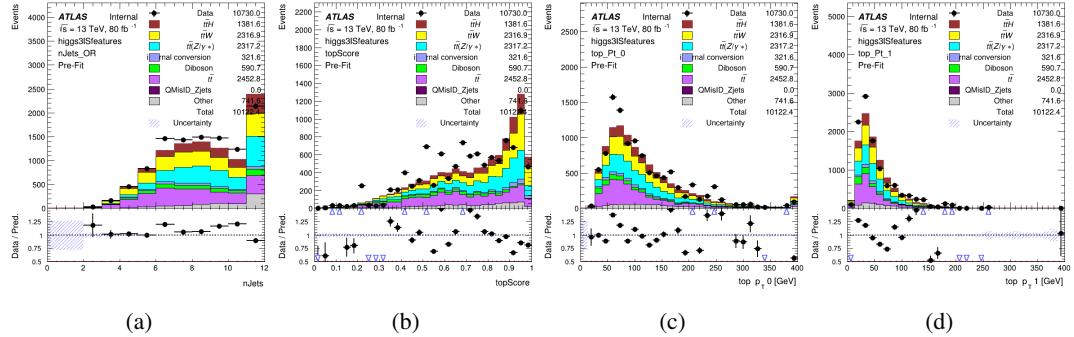


Figure C.15: Input features for higgs3IS

1621 **C.1.5 Higgs Reconstruction Features - 3lF**

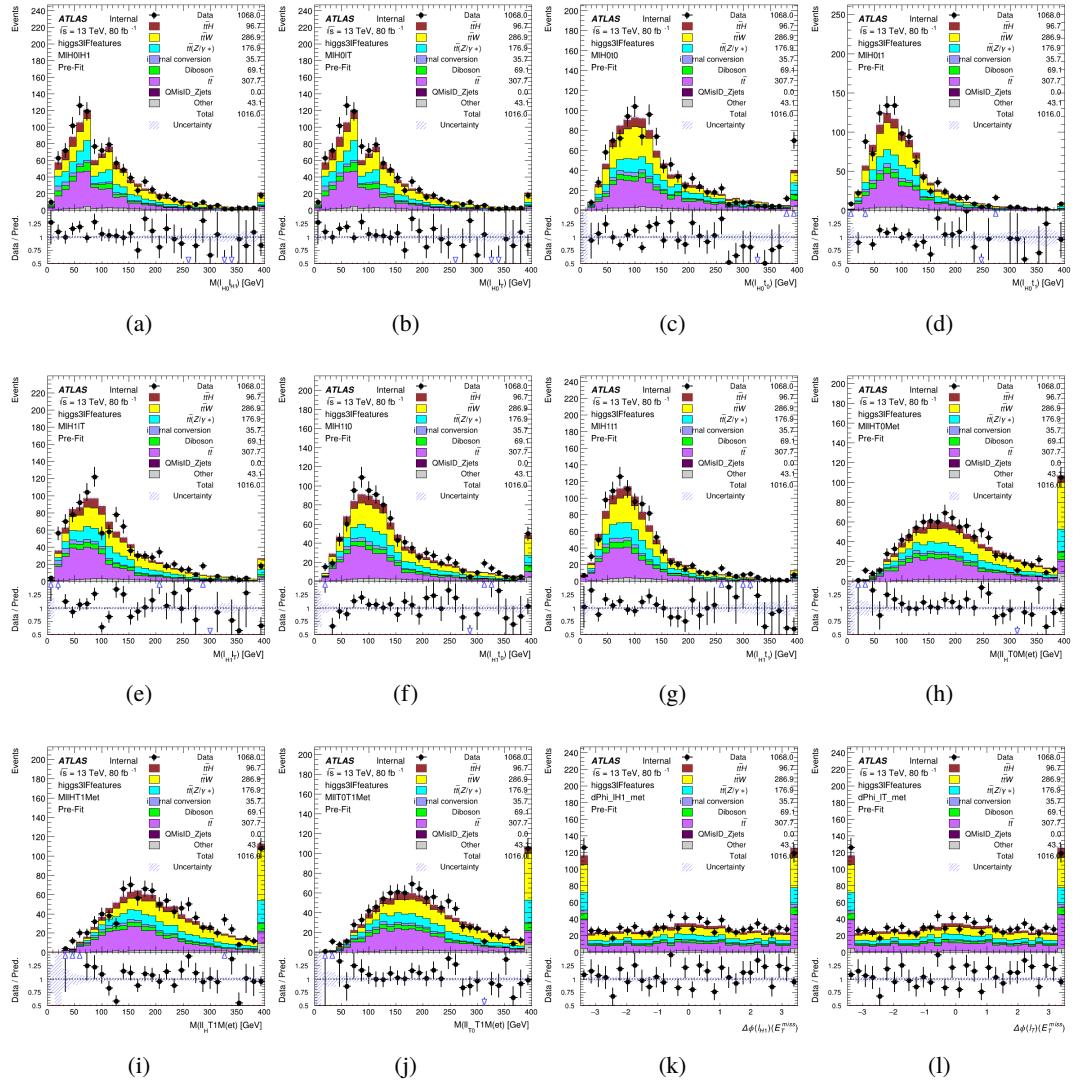


Figure C.16: Input features for higgs3lF

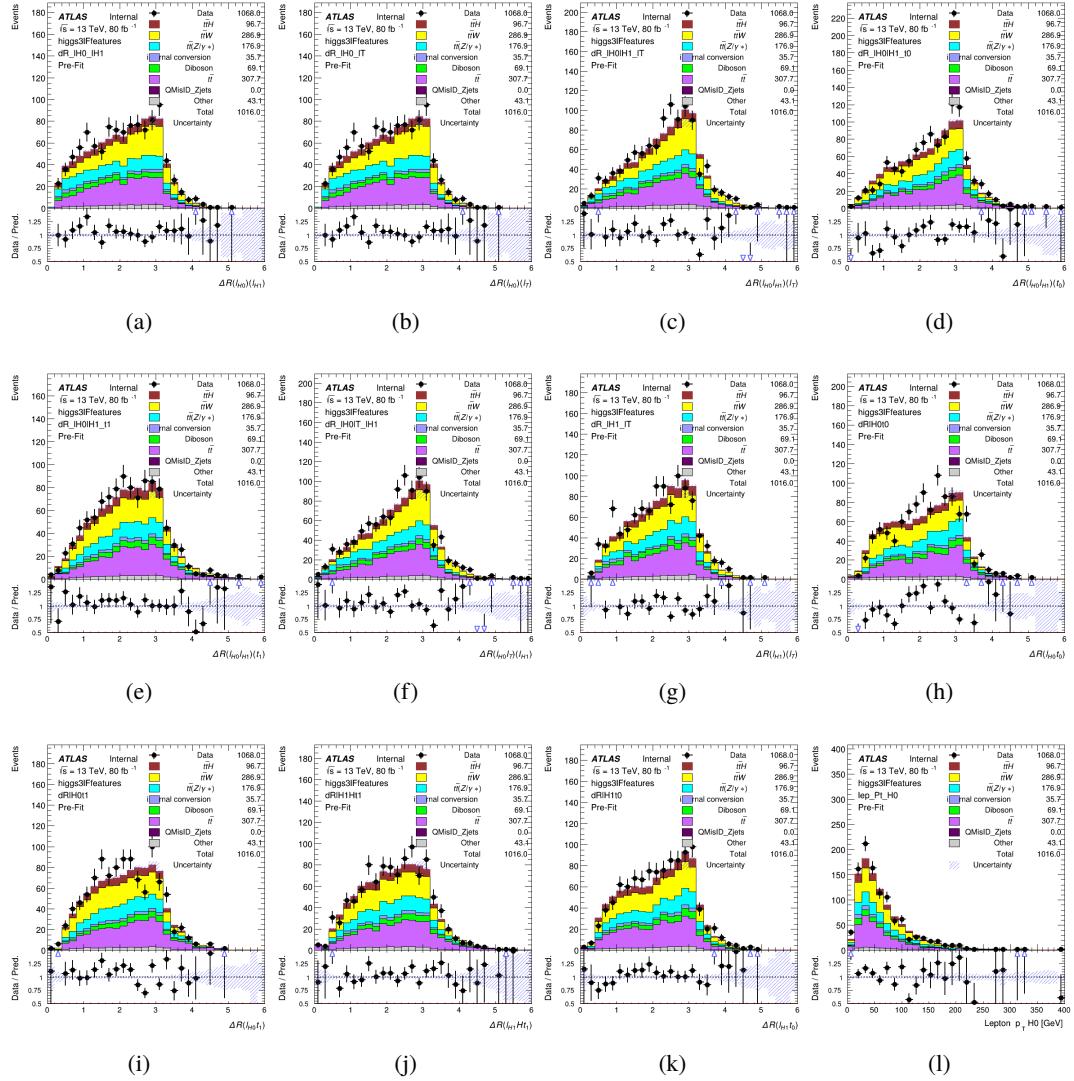


Figure C.17: Input features for higgs3IF

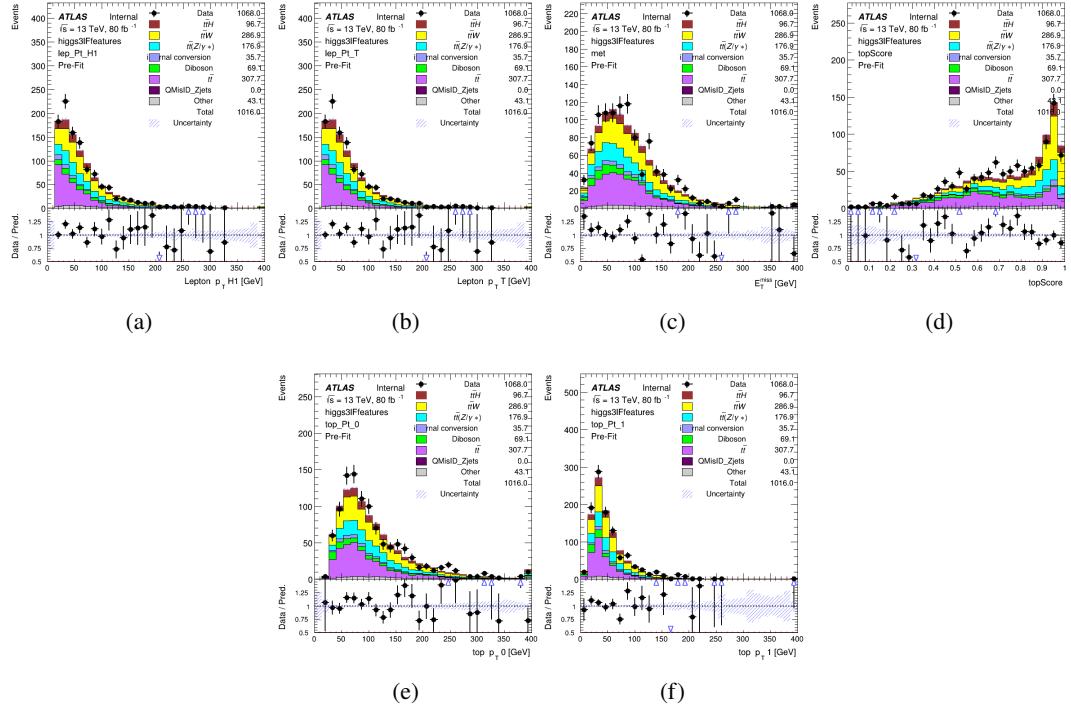


Figure C.18: Input features for higgs3lF

¹⁶²² **C.2 Background Rejection MVA Details**

¹⁶²³ **C.2.1 Background Rejection MVA Features - 2ISS**

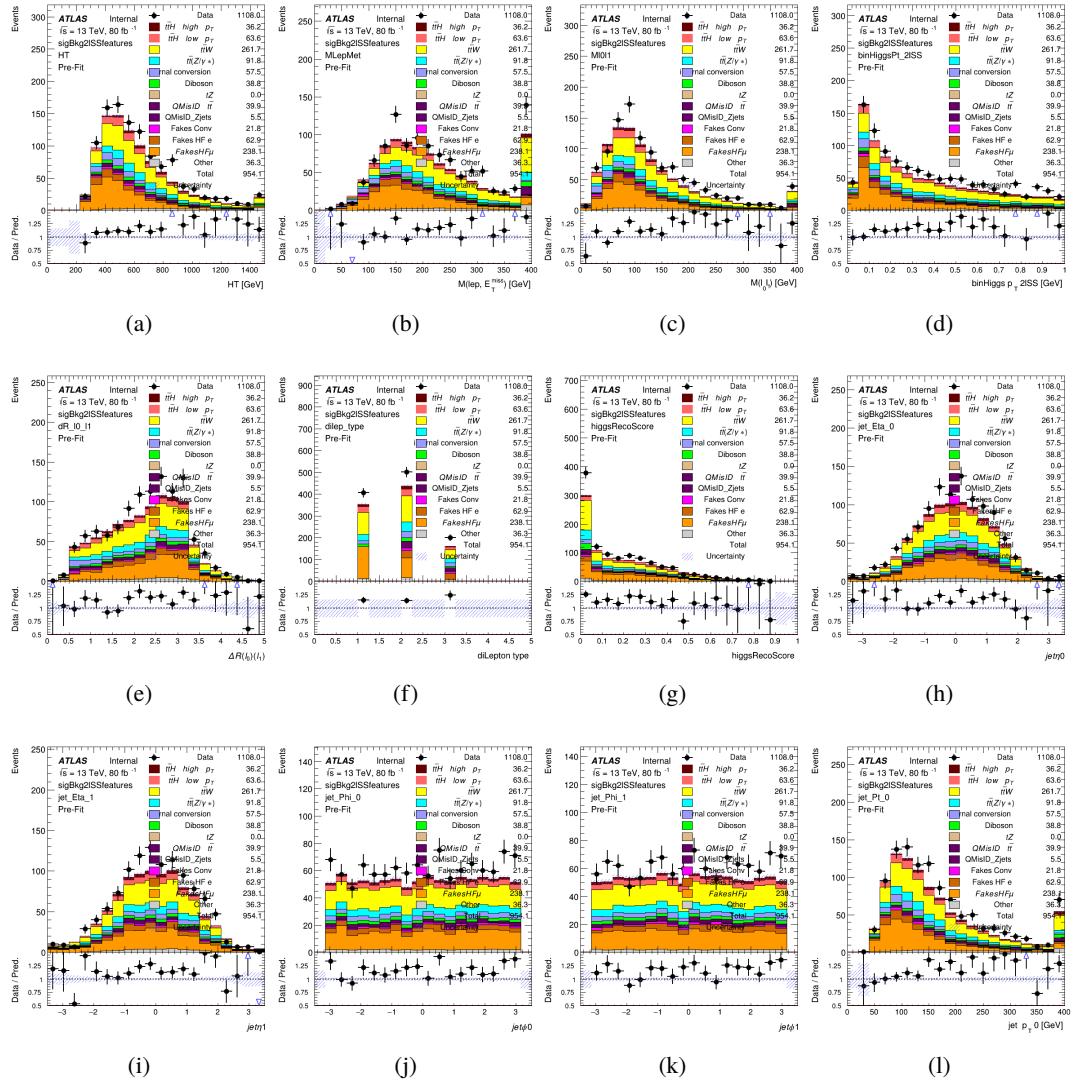


Figure C.19: Input features for sigBkg2ISS

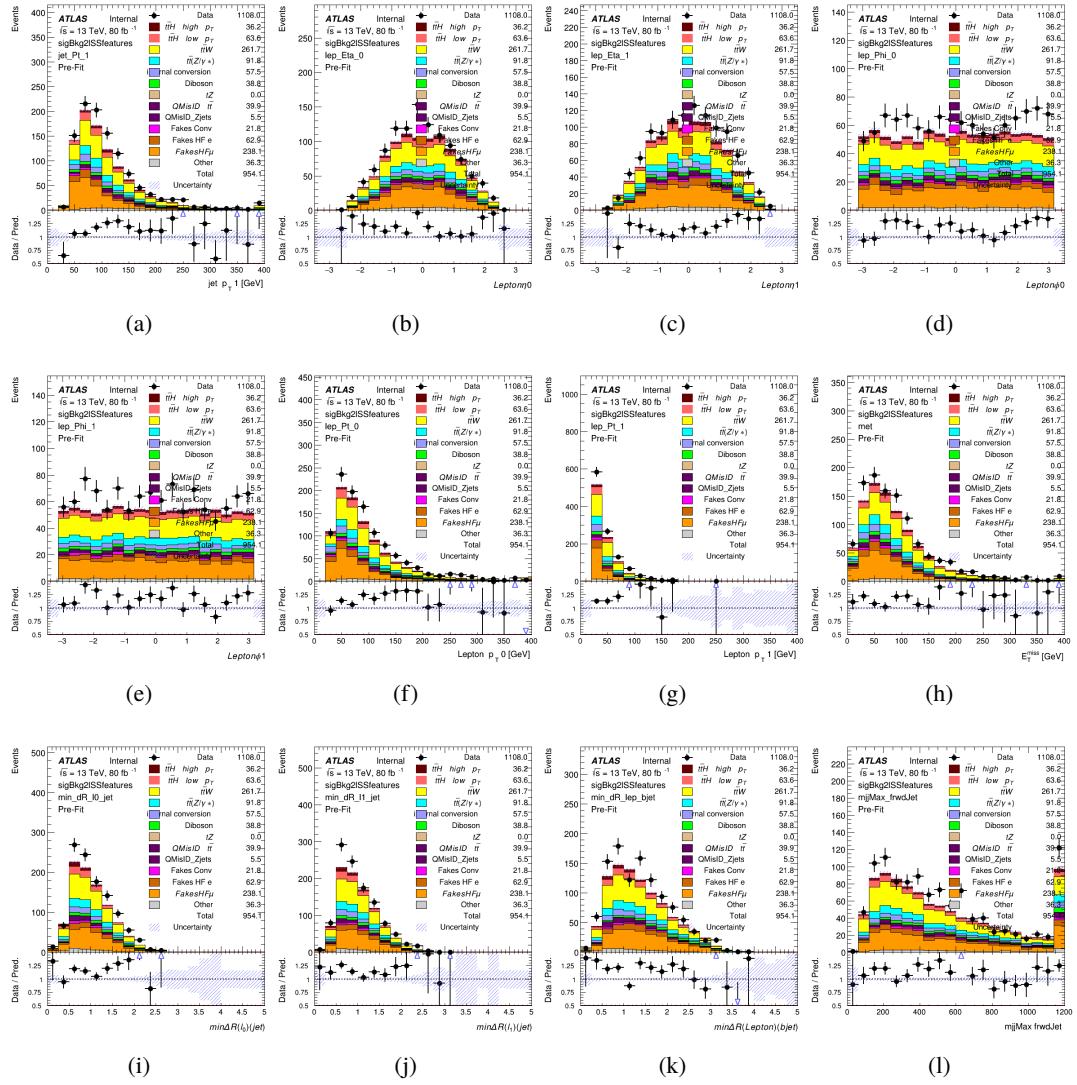


Figure C.20: Input features for sigBkg2lSS

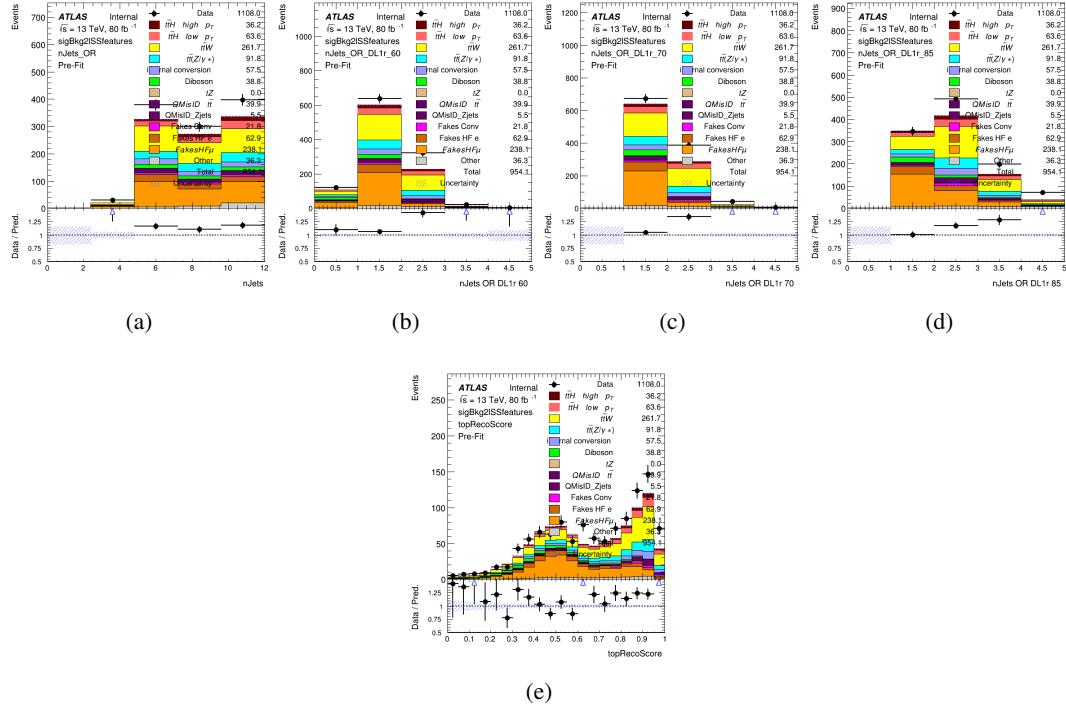


Figure C.21: Input features for sigBkg2lSS

1624 **C.2.2 Background Rejection MVA Features - 3l**

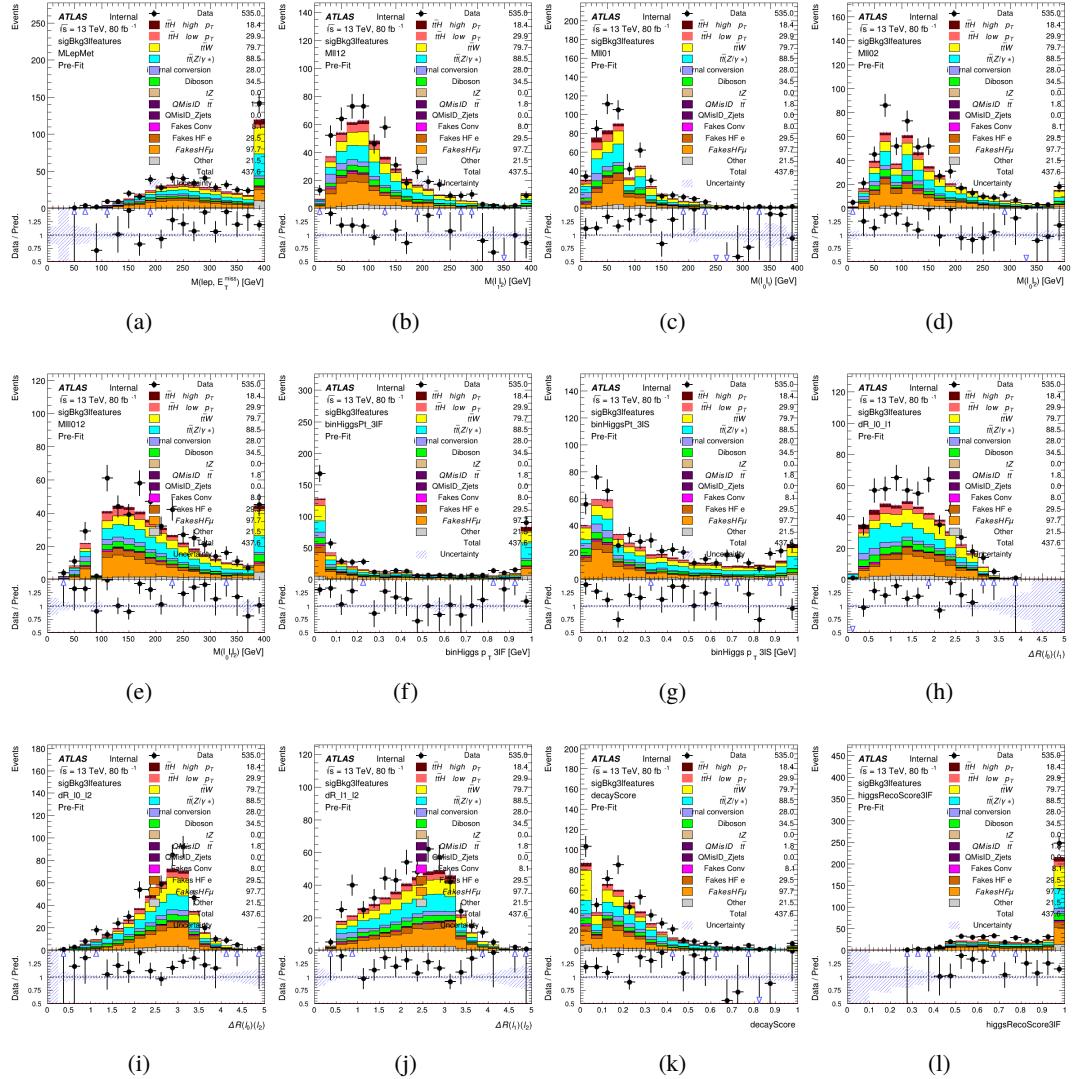


Figure C.22: Input features for sigBkg3l

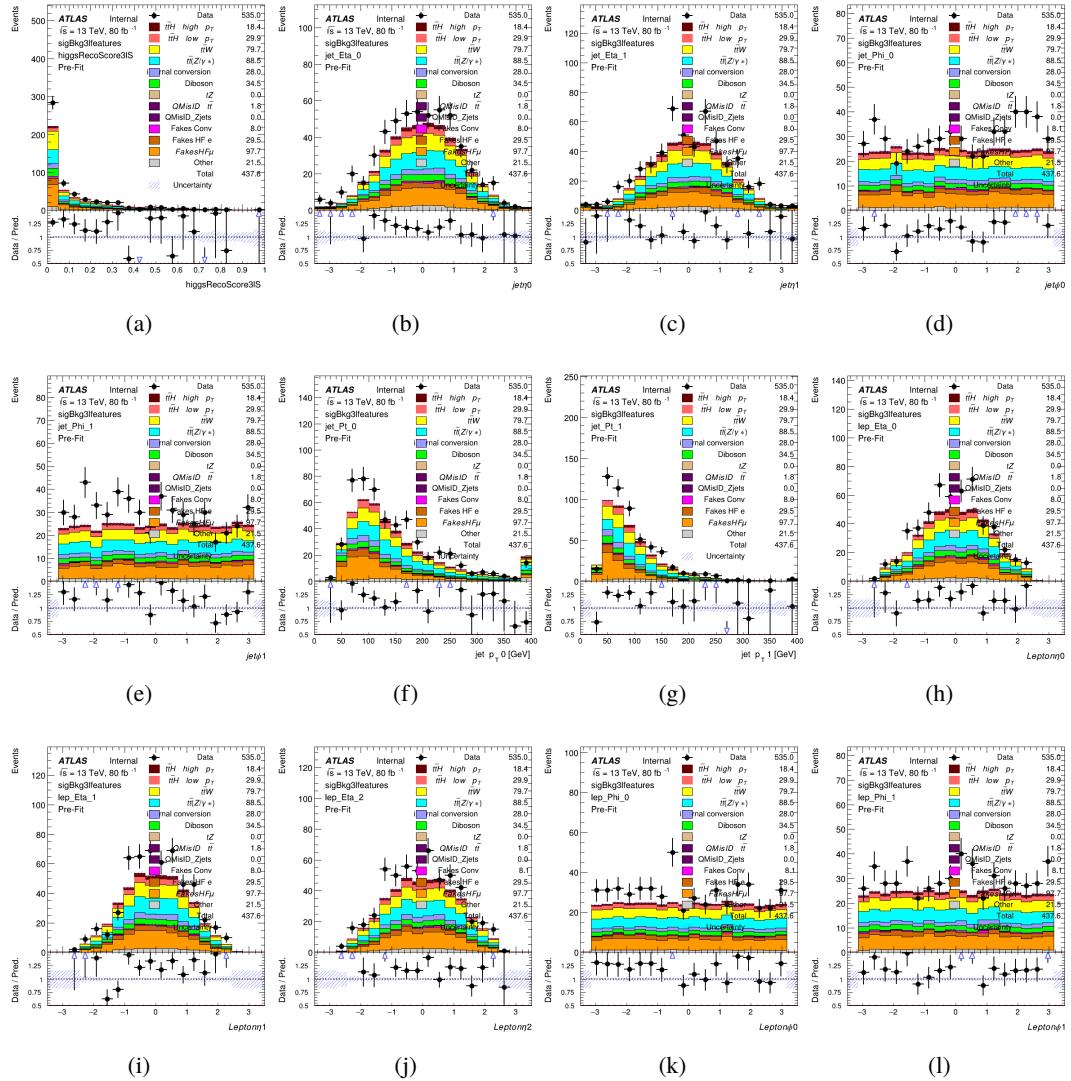


Figure C.23: Input features for sigBkg3l

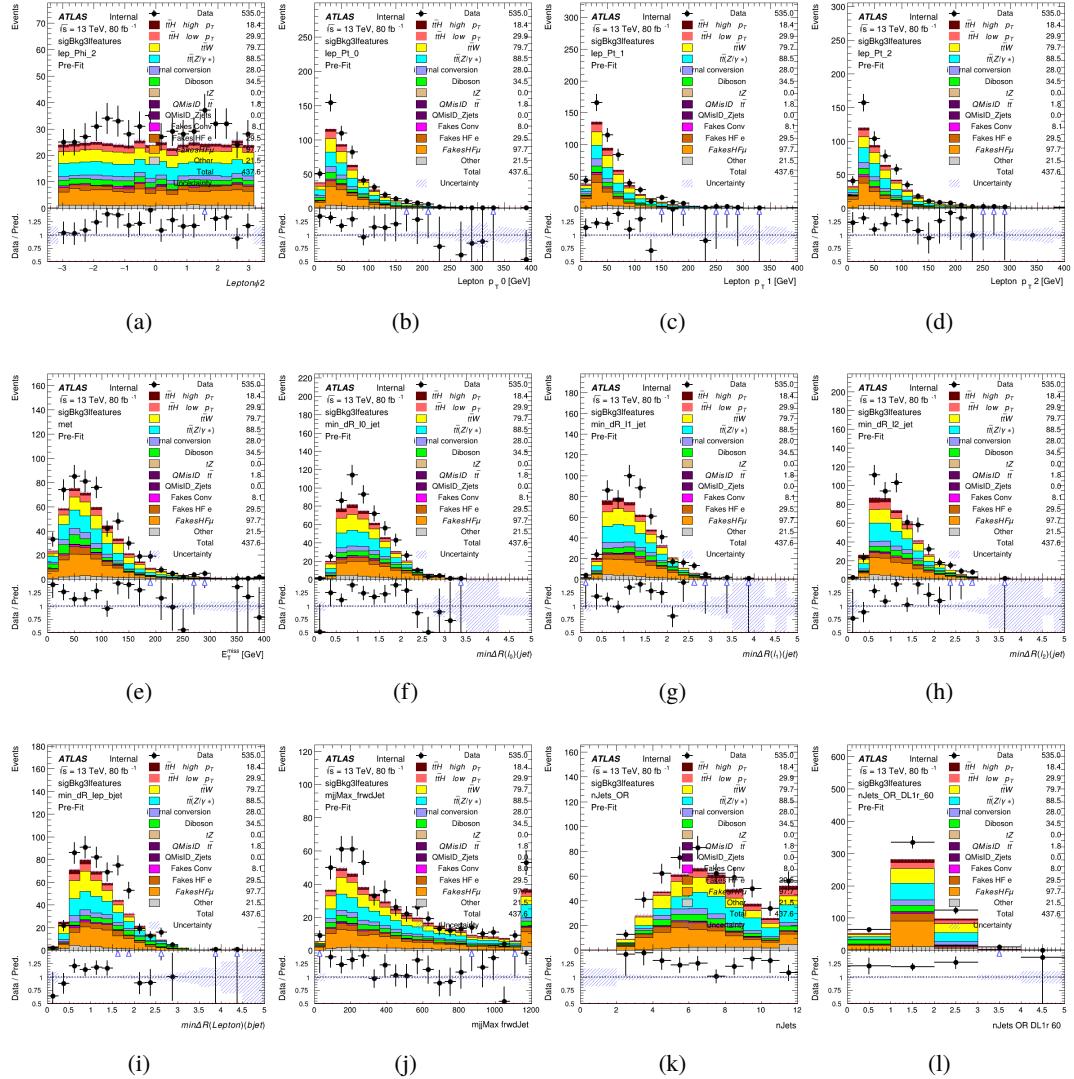


Figure C.24: Input features for sigBkg3l

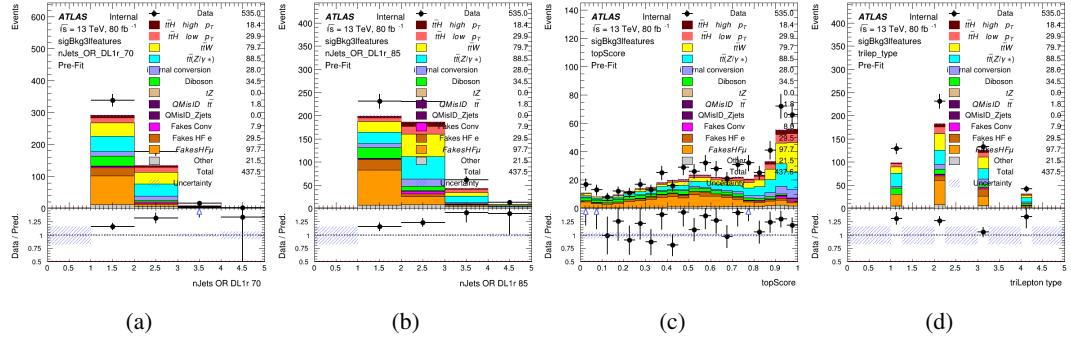


Figure C.25: Input features for sigBkg3l

1625 **C.3 Truth Level Studies**

1626 Attempts to identify the decay products of the Higgs are motivated by the ability to reconstruct
 1627 the Higgs momentum based on their kinematics. In order to demonstrate that this is case, the
 1628 kinematics of reconstructed objects that are truth matched to the Higgs decay are used as inputs
 1629 to a neural network which is designed to predict of the momentum of the Higgs. This is done in
 1630 the 2lSS channel, as it proves to be the most challenging for p_T reconstruction.

1631 Only leptons and jets which are truth matched to the Higgs are used as inputs for the
 1632 model; events where the lepton and both jets are not reconstructed are not included. The model
 1633 uses the same feature set and network architecture as the p_T prediction model used in the main
 1634 analysis, as described in Section 14.4.1.

1635 The results of the model are summarized below:

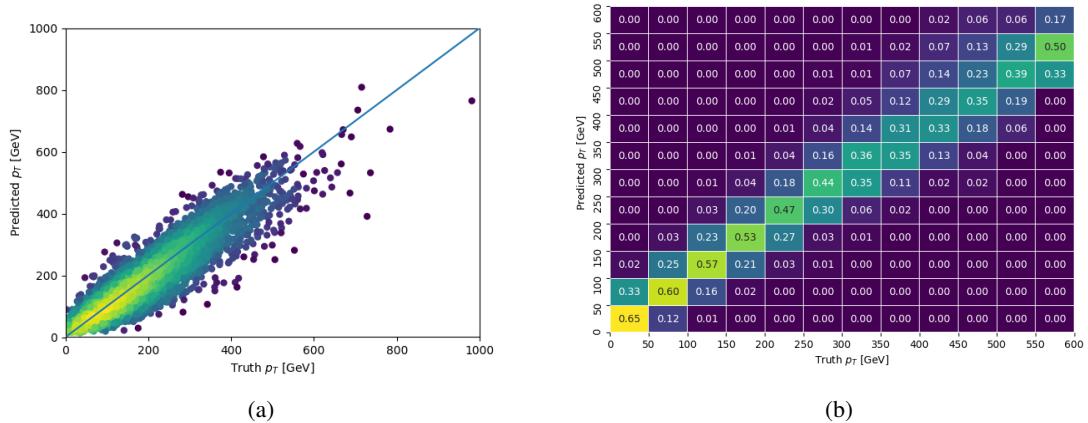


Figure C.26: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{H}$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

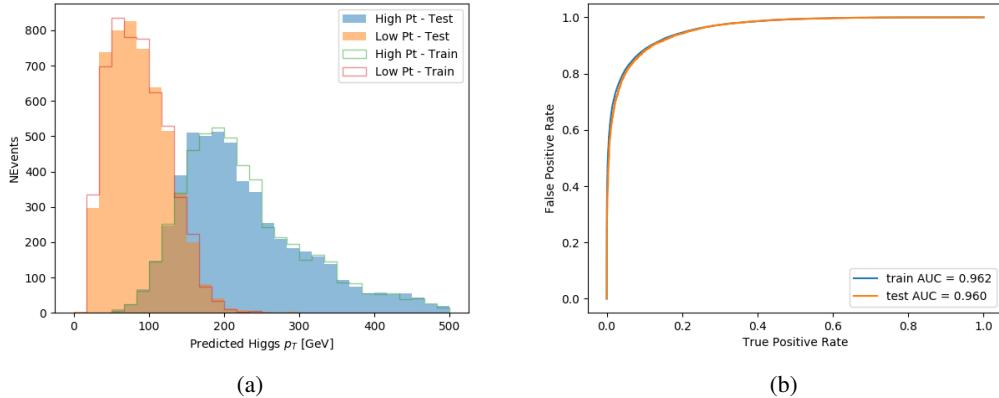


Figure C.27: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

1636 Based on the performance of the model, as shown Figures C.26 and C.27, the Higgs
 1637 momentum can be reconstructed with fairly high precision when its decay products are correctly
 1638 identified.

1639 C.4 Alternate b-jet Identification Algorithm

1640 The nominal analysis reconstructs the b-jets by considering different combinations of jets, and
 1641 asking a neural network to determine whether each combination consists of b-jets from top quark
 1642 decays. An alternate approach would be to give the neural network about all of the jets in an event
 1643 at once, and train it to select which two are most likely to be the b-jets from top decay. It was
 1644 hypothesized that this could perform better than considering each combination independently, as
 1645 the neural network could consider the event as a whole. While this is not found to be the case,
 1646 these studies are documented here as a point of interest and comparison.

1647 For these studies, the kinematics of the 10 highest p_T jets in each event are used for
 1648 training. This includes the vast majority of truth b-jets. Specifically the p_T , η , ϕ , E , and DL1r
 1649 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.
 1650 The p_T , η , ϕ , and E of the leptons and E_T^{miss} are included as well. Categorical cross entropy is
 1651 used as the loss function.

Table 48: 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%

1652 C.5 Binary Classification of the Higgs p_T

1653 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer
 1654 resolution. This means separating high and low p_T events is sufficient for this analysis. As
 1655 such, rather than attempting to reconstruct the full Higgs p_T spectrum, a binary classification
 1656 approach is explored.

1657 A model is built to determine whether $t\bar{t}H$ events include a high p_T (>150 GeV) or low
 1658 p_T (<150 GeV) Higgs Boson. While this is now a classification model, it uses the same input
 1659 features described in section 14.4. Binary crossentropy is used as the loss function.

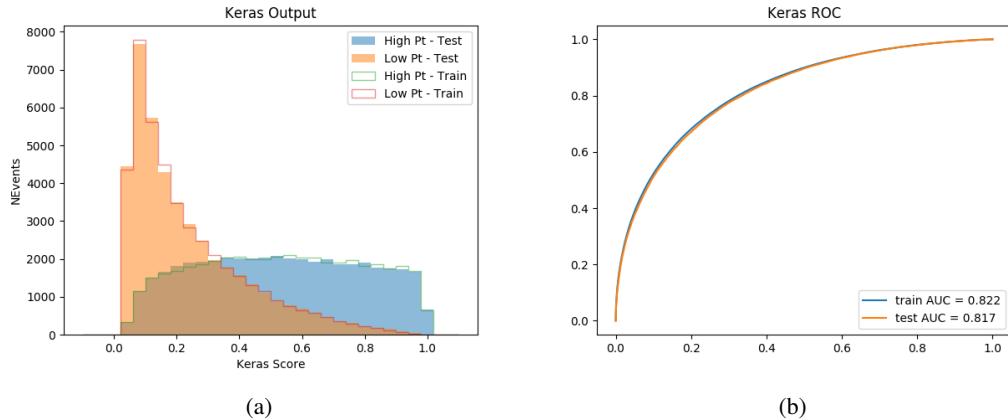


Figure C.28: Output distribution of the NN score for the binary high/low p_T separation model in the 2lSS channel.

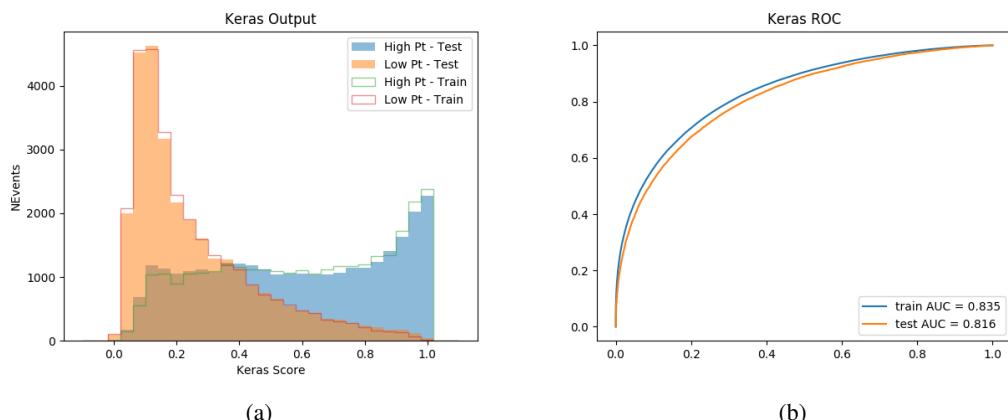


Figure C.29: Output distribution of the NN score for the binary high/low p_T separation model in the 3lS channel.

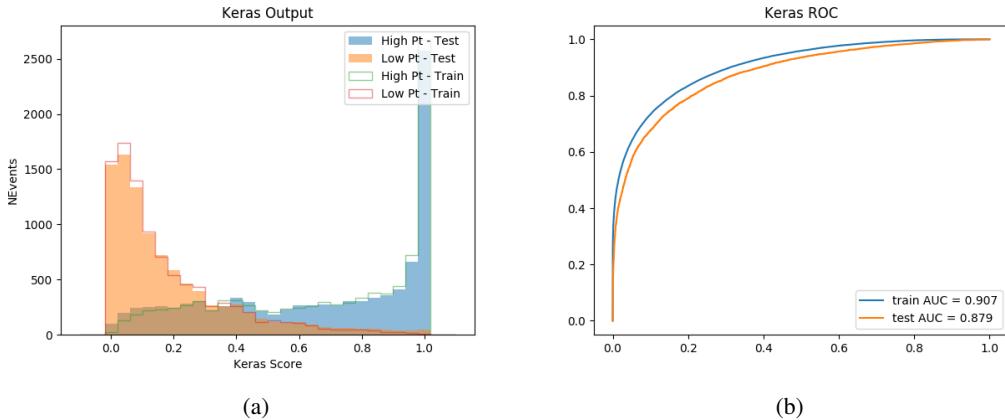


Figure C.30: Output distribution of the NN score for the binary high/low p_T separation model in the 3LS channel.

C.6 Impact of Alternative Jet Selection

1661 A relatively low p_T threshold of 15 GeV is used to determine jet candidates, as the jets originating
1662 from the Higgs decay are found to fall between 15 and 25 GeV a large fraction of the time. The
1663 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.

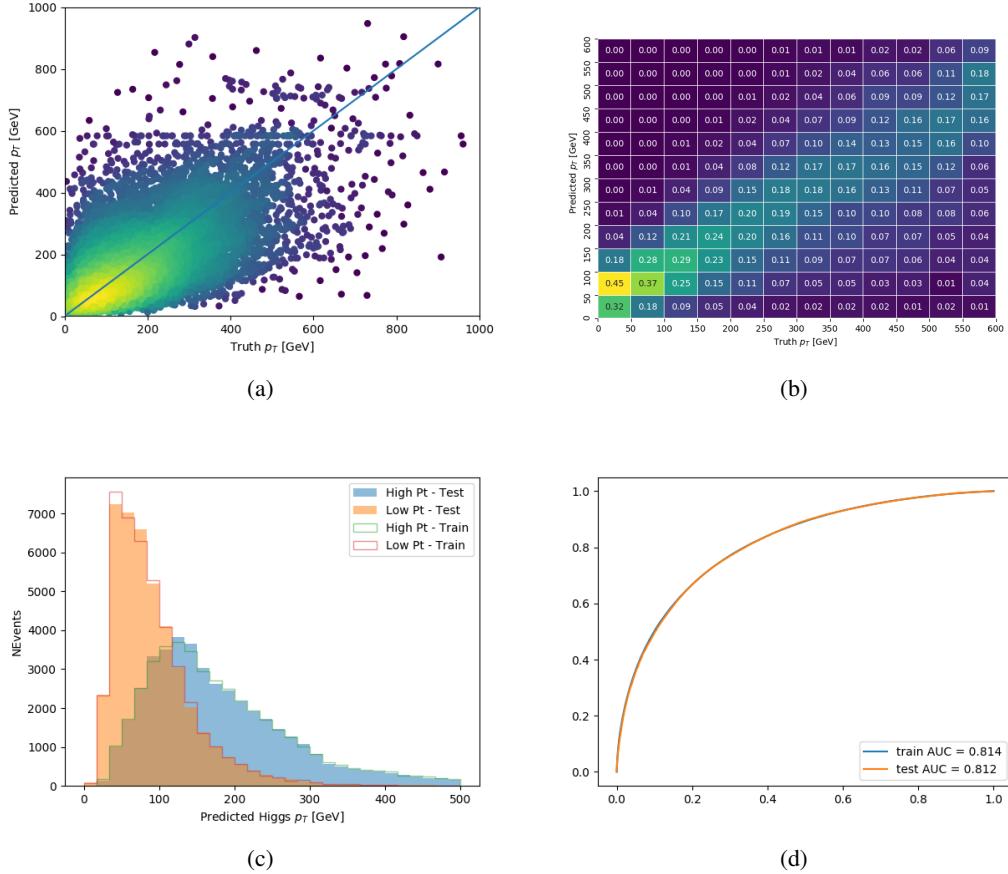
Jet $p_T > 20$ GeV

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.

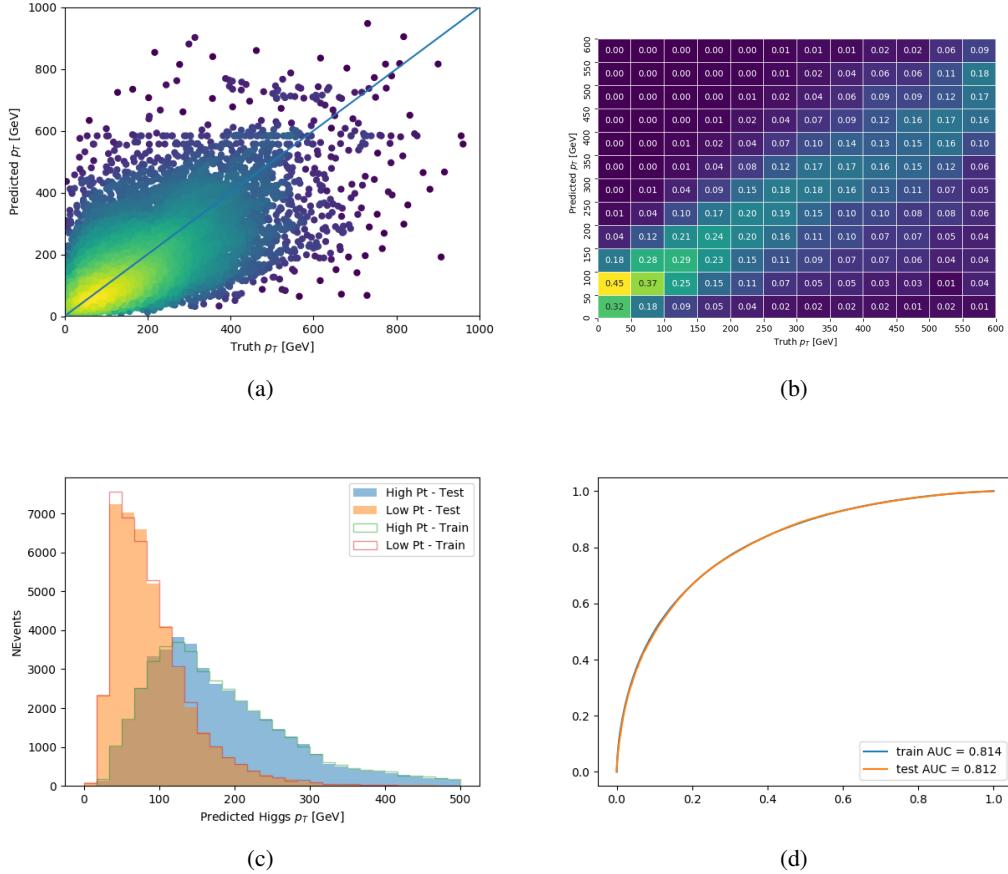
Jet $p_T > 25 \text{ GeV}$ 

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.

1667 **D**