



ATLAS Note

GROUP-2017-XX



11th November 2020

Draft version 0.1

1

2 **A Deep Learning Approach to Differential**
3 **Measurements of Higgs - Top Interactions in**
4 **Multilepton Final States**

5

The ATLAS Collaboration

6 Several theories Beyond the Standard Model modify the momentum spectrum of the Higgs
7 Boson, without significantly altering the overall rate of Higgs produced from top quark pairs.
8 A differential measurement of the Higgs transverse momentum provides a way to search for
9 these effects at the LHC. Because of the challenges inherent to reconstructing the Higgs in
10 multilepton final states, a deep learning approach is used to predict the momentum spectrum
11 of the Higgs for events where the Higgs Boson is produced from top quark pairs and decays
12 to final states that include multiple leptons. The regressed Higgs p_T is fit to data for events
13 with two or three leptons in the final state, and estimates of the sensitivity to variations in the
14 Higgs p_T spectrum are given.

© 2020 CERN for the benefit of the ATLAS Collaboration.

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

16 **Contents**

17	I Introduction	6
18	1 Introduction	6
19	II Theoretical Motivation	8
20	2 The Standard Model and the Higgs Boson	8
21	2.1 The Forces and Particles of the Standard Model	9
22	2.2 The Higgs Mechanism	11
23	2.2.1 The Higgs Field	12
24	2.2.2 Electroweak Symmetry Breaking	14
25	2.3 Limitations of the Standard Model	16
26	3 Effective Field Theory in $t\bar{t}H$ Production	17
27	3.1 Extensions to the Higgs Sector	17
28	3.2 Six Dimensional Operators	17
29	III The LHC and the ATLAS Detector	18
30	4 The LHC	18
31	5 The ATLAS Detector	21
32	5.1 Inner Detector	22
33	5.2 Calorimeters	23
34	5.3 Muon Spectrometer	24
35	5.4 Trigger System	25
36	IV Search for Dimension-Six Operators	26
37	6 Data and Monte Carlo Samples	26
38	6.1 Data Samples	27
39	6.2 Monte Carlo Samples	28
40	7 Object Reconstruction	28
41	7.1 Trigger Requirements	29
42	7.2 Light Leptons	29
43	7.3 Jets	30
44	7.4 Missing Transverse Energy	31
45	8 Higgs Momentum Reconstruction	31

46	8.1	Decay Candidate Reconstruction	33
47	8.2	b-jet Identification	33
48	8.2.1	2lSS Channel	34
49	8.2.2	3l Channel	38
50	8.3	Higgs Reconstruction	41
51	8.3.1	2lSS Channel	42
52	8.3.2	3l Semi-leptonic Channel	46
53	8.3.3	3l Fully-leptonic Channel	50
54	8.4	p_T Prediction	53
55	8.4.1	2lSS Channel	55
56	8.4.2	3l Semi-leptonic Channel	58
57	8.4.3	3l Fully-leptonic Channel	60
58	8.5	3l Decay Mode	64
59	9	Signal Region Definitions	66
60	9.1	Pre-MVA Event Selection	67
61	9.2	Event MVA	68
62	9.3	Signal Region Definitions	74
63	9.3.1	2lSS	77
64	9.3.2	3l – Semi – leptonic	77
65	9.3.3	3l – Fully – leptonic	77
66	10	Background Rejection MVA	77
67	10.1	Background Rejection MVAs	77
68	10.1.1	2lSS - High p_T	78
69	10.1.2	2lSS - Low p_T	79
70	10.1.3	3l Semi-Leptonic - High p_T	79
71	10.1.4	3l Semi-Leptonic - Low p_T	79
72	10.1.5	3l Fully Leptonic - High p_T	79
73	10.1.6	3l Fully Leptonic - Low p_T	79
74	11	Systematic Uncertainties	79
75	12	Results	81
76	12.1	Results - 80 fb^{-1}	81
77	12.2	Projected Results - 140 fb^{-1}	84
78	V	Conclusion	84
79	Appendices		86
80	A	Machine Learning Models	86
81	A.1	Higgs Reconstruction Models	86

82	A.1.1	b-jet Identification Features - 2lSS	86
83	A.1.2	b-jet Identification Features - 3l	90
84	A.1.3	Higgs Reconstruction Features - 2lSS	95
85	A.1.4	Higgs Reconstruction Features - 3lS	100
86	A.1.5	Higgs Reconstruction Features - 3lF	105
87	A.2	Background Rejection MVAs	109
88	A.2.1	Background Rejection MVA Features - 2lSS	109
89	A.2.2	Background Rejection MVA Features - 3l	113
90	A.3	Alternate b-jet Identification Algorithm	118
91	A.4	Binary Classification of the Higgs p_T	118
92	A.5	Impact of Alternative Jet Selection	120

93 Part I**94 Introduction****95 1 Introduction**

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

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

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

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

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

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

125 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-
126 2018 provides the opportunity to make this measurement for the first time. The unprecedented
127 energy achieved by the LHC during this period greatly increase the rate at which $t\bar{t}H$ events are
128 produced, and the large amount of data collected provides the necessary statistics for a differential
129 measurement to be performed.

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

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

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

140 Part II

141 Theoretical Motivation

142 2 The Standard Model and the Higgs Boson

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

¹⁴⁸ theory, invariant under the Lie Group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where C refers to color

¹⁴⁹ charge, L, the helicity of the particle, and Y, the hypercharge.

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

¹⁵¹ The SM particles, summarized in figure 2.1, can be classified into two general categories based

¹⁵² on their spin: fermions, and bosons.

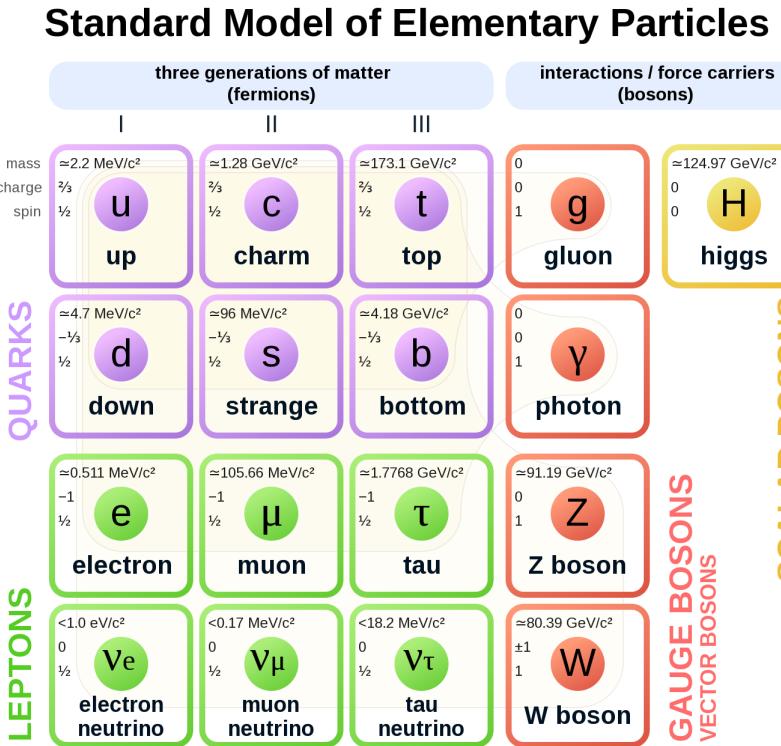


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. []

¹⁵³ Fermions are particles with $\frac{1}{2}$ -integer spin, which according to the spin-statistics theorem,

¹⁵⁴ causes them to comply with the Pauli-exclusion principle []. They can be separated into two

155 groups, leptons and quarks, each of which consist of three generations of particles with increasing
 156 mass.

157 Leptons are fermions interact via the electroweak force, but not the strong force. The three
 158 generation of leptons consist of the electron and electron neutrino, the muon and muon neutrino,
 159 the tau and tau neutrino. The quarks, which do interact via the strong force - which is to say they
 160 have color charge - in addition to the electroweak force. The three generations include the up
 161 and down quarks, the strange and charm quarks, and the top and bottom quarks. Each of these
 162 generations form left-handed doublets invariant under SU(2) transformations. For the leptons
 163 these doublets are:

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

164 And for the quarks:

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

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

¹⁶⁹ charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos,

¹⁷⁰ whose mass has been experimentally confirmed but is not accounted for in the SM.

¹⁷¹ Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-

¹⁷² exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1

¹⁷³ particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0,

¹⁷⁴ particle - the Higgs Boson. Of the gauge bosons, the W^+ , W^- and Z bosons - which are the

¹⁷⁵ mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon

¹⁷⁶ mediates the electric force, and the gluon mediates the strong force.

¹⁷⁷ 2.2 The Higgs Mechanism

¹⁷⁸ A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to

¹⁷⁹ the Lagrangian giving mass to the the gauge bosons would violate the underlying symmetry of

¹⁸⁰ the theory. This presents a clear problem with the theory: The experimental observation that the

¹⁸¹ W and Z bosons have mass seems to contradict the basic structure of the SM.

¹⁸² Rather than abandoning gauge invariance, an alternative way for particles to acquire mass

¹⁸³ beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout

¹⁸⁴ in 1964 []. This procedure for introducing masses for the gauge bosons while preserving local

¹⁸⁵ gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory

¹⁸⁶ by Weinberg in 1967 [].

¹⁸⁷ **2.2.1 The Higgs Field**

¹⁸⁸ The Higgs mechanism introduces a complex scalar $SU(2)$ doublet, Φ , with the form:

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

¹⁸⁹ This field introduces a scalar potential to the Lagrangian of the form:

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

¹⁹⁰ Where μ and λ are free parameters of the new field. This represents the most general
¹⁹¹ potential allowed while preserving $SU(2)_L$ invariance and renormalizability. In the case that
¹⁹² $\mu^2 < 0$, this potential takes the form shown in figure 2.2.

¹⁹³ The significant feature of this potential is that its minimum does not occur for a value of
¹⁹⁴ $\Phi = 0$. Instead, it is minimized when $|\Phi^\dagger \Phi| = -\mu^2/\lambda$. This means that in its ground state, the
¹⁹⁵ Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while
¹⁹⁶ the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

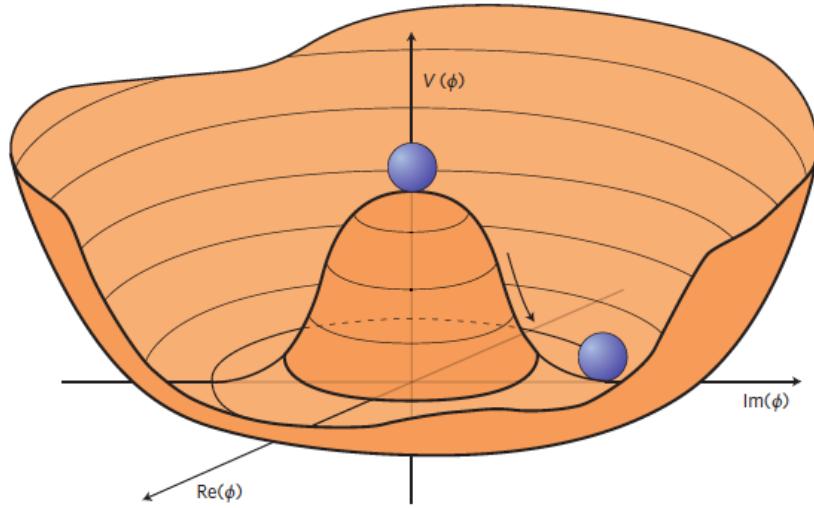


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ , for the case that $\mu^2 < 0$ [].

₁₉₇ the minimum is determined only by $\Phi^\dagger \Phi$, there is some ambiguity in the particular definition of
₁₉₈ the VEV, but it is generally represented as

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

₁₉₉ The full value of Φ can be written as

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

₂₀₀ with v being the value of the VEV, and H being the real value of the scalar field.

201 **2.2.2 Electroweak Symmetry Breaking**

202 The Electroweak (EWK) interaction is described in the SM by a $SU(2)_L \otimes U(1)_Y$ gauge theory.
 203 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,
 204 B_μ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar
 205 field Φ in the Lagrangian:

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

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

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

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

²¹² $W_\mu^1, W_\mu^2, W_\mu^3$, and B_μ :

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

²¹³ And the masses of these fields are given by:

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

²¹⁴ This produces exactly the particles we observe - three massive gauge bosons and a single
²¹⁵ massless photon. The massless photon represents the portion of the gauge symmetry, a single
²¹⁶ $U(1)$ of the electromagnetic force, that remains unbroken by the VEV.

²¹⁷ Interactions with the Higgs field also lead to the generation of the fermion masses, which
²¹⁸ in the Lagrangian take the form:

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

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

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

226 The fermion's Yukawa coupling to the VEV take the same form as the fermion's coupling
 227 to the Higgs boson - λ_ψ . Therefore, the strength of a fermion's interaction with the Higgs is
 228 directly proportional to its mass. We now have a model that predicts a Higgs boson with mass
 229 $M_H = \mu$, which interacts with the fermions with coupling strength λ_ψ . Because μ and λ_ψ are
 230 free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions
 231 must be measured experimentally.

232 2.3 Limitations of the Standard Model

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

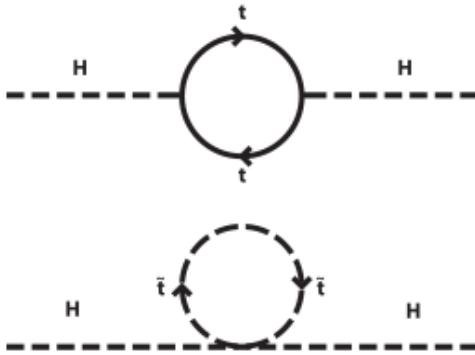


Figure 2.3: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram [6].

²³⁶ **3 Effective Field Theory in $t\bar{t}H$ Production**

²³⁷ Higher dimension operators are a common way to paramaterize the effects of physics at very
²³⁸ high energies into

²³⁹ **3.1 Extensions to the Higgs Sector**

²⁴⁰ **3.2 Six Dimensional Operators**

²⁴¹ While the SM has been tested to great precision, particularly at the LHC, it is generally accepted
²⁴² that it is only valid up to a certain energy scale. It is assumed that above a certain energy, at the
²⁴³ scale where something like a Grand Unified Theory (GUT) or quantum gravity become relevant,
²⁴⁴ the SM will not be applicable.

Part III**246 The LHC and the ATLAS Detector****247 4 The LHC**

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

257 The LHC first began running in 2009 at a proton-proton center of mass energy of $\sqrt{s} = 8$
258 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during
259 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,
260 and collected data at an increased energy of $\sqrt{s} = 13$ TeV until 2018, a period referred to as Run
261 2.

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

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

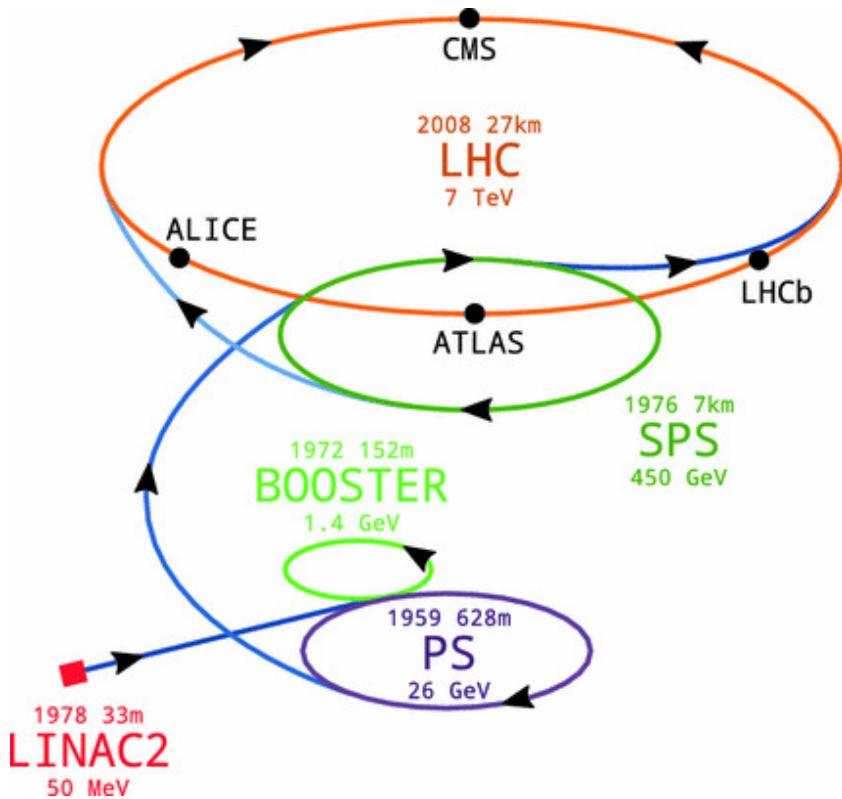


Figure 4.1: A summary of the accelerator chain used to feed protons into the LHC [].

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

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

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

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

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

281 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around
 282 140 fb^{-1} , exceeding the expected integrated luminosity of 100 fb^{-1} .

283 5 The ATLAS Detector

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

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

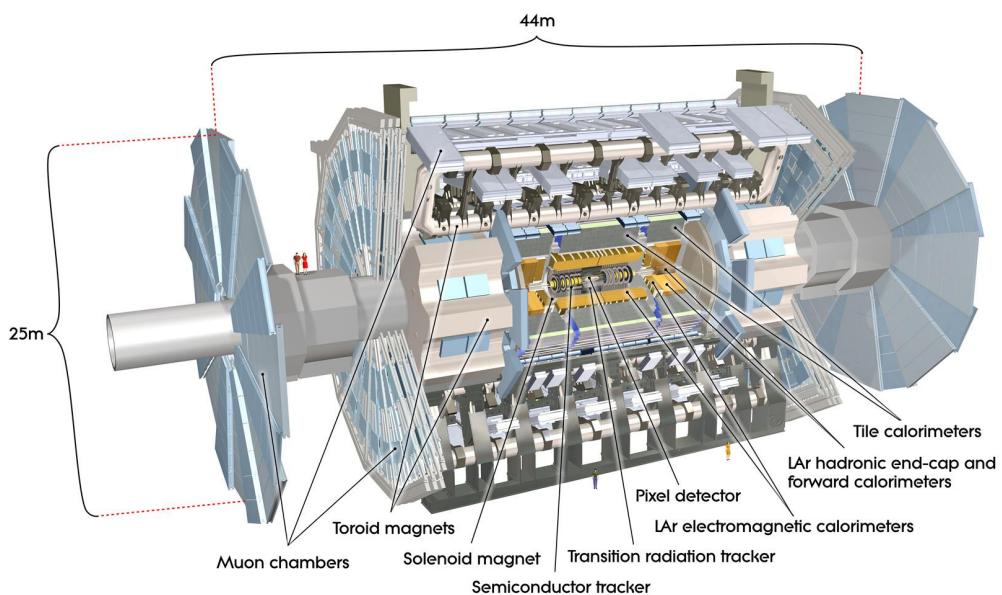


Figure 5.1: Cutaway view of the ATLAS detector, with labels of its major components [].

292 **5.1 Inner Detector**

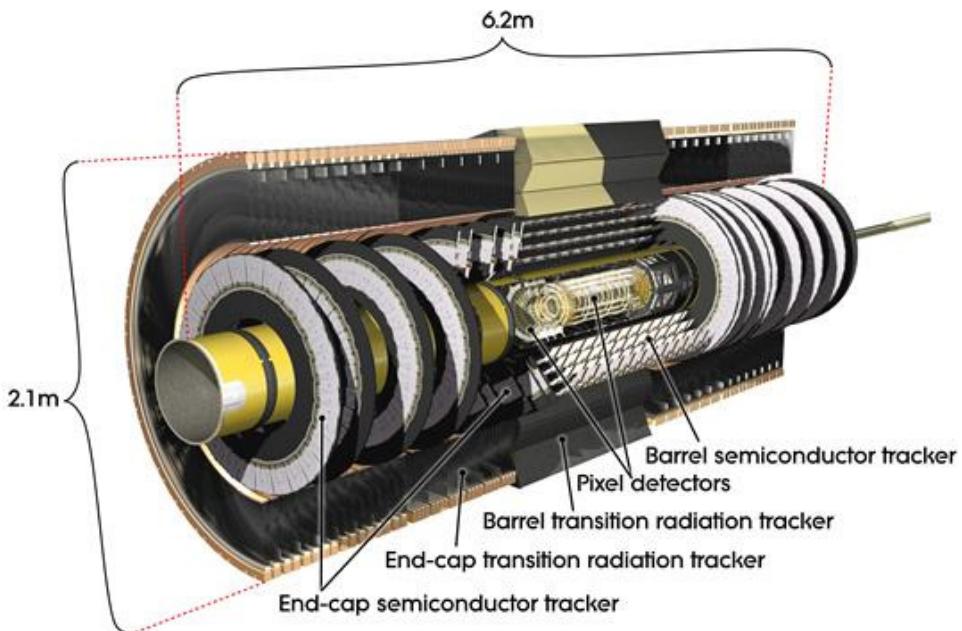


Figure 5.2: Cutaway view of the Inner Detector [].

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

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

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

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

305 **5.2 Calorimeters**

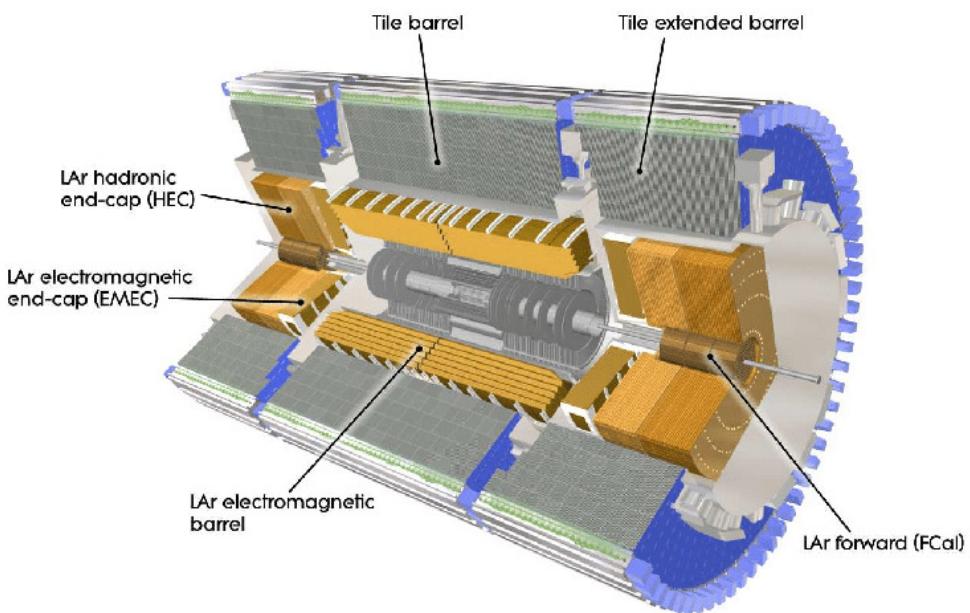


Figure 5.3: Cutaway view of the calorimeter system of the ATLAS detector [].

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

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

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

318 **5.3 Muon Spectrometer**

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

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

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

333 1200 tracking chambers are placed in the muon system in order to precisely measure the
334 tracks of muons with high spatial resolution.

335 **5.4 Trigger System**

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

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

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

354 **Part IV**

355 **Search for Dimension-Six Operators**

356 **6 Data and Monte Carlo Samples**

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

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

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

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

368 **6.1 Data Samples**

369 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through
370 2018, which represents an integrated luminosity of 139 fb^{-1} and an energy of $\sqrt{s} = 13 \text{ TeV}$. All
371 data used in this analysis was included in one the following Good Run Lists:

- 372 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
373 _PHYS_StandardGRL_All_Good_25ns.xml
- 374 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
375 _PHYS_StandardGRL_All_Good_25ns.xml
- 376 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL
377 _All_Good_25ns_Triggerno17e33prim.xml
- 378 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL
379 _All_Good_25ns_Triggerno17e33prim.xml

380 **6.2 Monte Carlo Samples**

381 Several Monte Carlo (MC) generators were used to simulate both signal and background pro-
382 cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific
383 event generator used for each of these MC samples is listed in table 1.

Table 1: The configurations used for event generation of signal and background processes, including the event generator, matrix element (ME) order, parton shower algorithm, and parton distribution function (PDF).

Process	Event generator	ME order	Parton Shower	PDF
ttH	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [Ball:2014uwa] (CT10 [ct10])
t̄W	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
t̄t(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
t̄t	POWHEG-BOX v2 [powhegtt]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄tγ	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
tZ	MG5_AMC	LO	PYTHIA 6	CTEQ6L1
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t̄t, t̄t̄t̄t	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	POWHEG-BOX v1 [powhegstp]	NLO	PYTHIA 6	CT10
qqVV, VVV				
Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

384 **7 Object Reconstruction**

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

387 **7.1 Trigger Requirements**

388 Events are required to be selected by dilepton triggers, as summarized in table 2.

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

Table 2: List of lowest p_T -threshold, un-prescaled dilepton triggers used for 2015-2018 data taking.

389 **7.2 Light Leptons**

390 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that
391 are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].
392 Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the
393 transition region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| <$
394 1.52, are rejected. A multivariate likelihood discriminant combining shower shape and track

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

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

400 Muon candidates are reconstructed by combining inner detector tracks with track segments
401 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
402 $p_T > 10$ GeV and $|\eta| < 2.5$. All leptons are required to be isolated, and pass a non-prompt BDT
403 selection described in detail in [**ttH_paper**].

404 7.3 Jets

405 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
406 calorimeters [**ATL-PHYS-PUB-2015-015**], using the anti- k_t algorithm with a radius parameter
407 $R = 0.4$. Jets with energy contributions likely arising from noise or detector effects are removed
408 from consideration [**ATLAS-CONF-2015-029**], and only jets satisfying $p_T > 25$ GeV and
409 $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
410 association algorithm is used to confirm that the jet originates from the selected primary vertex,
411 in order to reject jets arising from pileup collisions [**PERF-2014-03**].

⁴¹² **7.4 Missing Transverse Energy**

⁴¹³ Because all $t\bar{t}H - ML$ channels considered include multiple neutrinos, missing transverse
⁴¹⁴ energy (E_T^{miss}) is present in each event. The missing transverse momentum vector is defined as
⁴¹⁵ the inverse of the sum of the transverse momenta of all reconstructed physics objects as well
⁴¹⁶ as remaining unclustered energy, the latter of which is estimated from low- p_T tracks associated
⁴¹⁷ with the primary vertex but not assigned to a hard object [ATL-PHYS-PUB-2015-027].

⁴¹⁸ **8 Higgs Momentum Reconstruction**

⁴¹⁹ Reconstructing the momentum of the Higgs boson is a particular challenge for channels with
⁴²⁰ leptons in the final state: Because all channels include at least two neutrinos in the final state, the
⁴²¹ Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted
⁴²² by a neural network when provided with the four-vectors of the Higgs Boson decay products, as
⁴²³ shown in section 8.1. With this in mind, several layers of MVAs are used to reconstruction the
⁴²⁴ Higgs momentum.

⁴²⁵ The first layer is a model designed to select which jets are most likely to be the b-jets
⁴²⁶ that came from the top decay, detailed in section 8.2. As described in section 8.3, the kinematics
⁴²⁷ of these jets are fed into the second layer, which is designed to identify the decay products of
⁴²⁸ the Higgs Boson itself. The kinematics of these particles are then fed into yet another neural-

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

431 Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_AMC.
432 Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training.

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

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

447 8.1 Decay Candidate Reconstruction

448 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson
 449 using MC simulations of $t\bar{t}H$ events. These include light leptons and jets. Reconstructed
 450 physics objects are matched to truth level particles, in order to identify the parents of these
 451 reconstructed objects. The kinematics of the decay product candidates as well as event level
 452 variables are used as inputs.

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

462 8.2 b-jet Identification

463 Including the kinematics of the b-jets that originate from the top decay is found to improve the
 464 identification of the Higgs decay products, and improve the accuracy with which the Higgs
 465 momentum can be reconstructed. Because these b-jets are reconstructed by the detector with

⁴⁶⁶ high efficiency (just over 90% of the time), and can be identified relatively consistently, the first
⁴⁶⁷ step in reconstructing the Higgs is selecting the b-jets from the top decay.

⁴⁶⁸ Exactly two b-jets are expected in the final state of $t\bar{t}H - ML$ events. However, in both
⁴⁶⁹ the 3l and 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r b-tag
⁴⁷⁰ working point is used). Therefore, for events which have exactly one, or more than two, b-tagged
⁴⁷¹ jets, deciding which combination of jets correspond to the top decay is non-trivial. Further,
⁴⁷² events with 1 b-tagged jet represent just over half of all $t\bar{t}H - ML$ events. Of those, both b-jets
⁴⁷³ are reconstructed by the detector 75% of the time. Therefore, rather than adjusting the selection
⁴⁷⁴ to require exactly 2 b-tagged jets, and losing more than half of the signal events, a neural network
⁴⁷⁵ is used to predict which pair of jets is most likely to correspond to truth b-jets.

⁴⁷⁶ Once the network is trained, all possible pairings of jets are fed into the model, and the pair
⁴⁷⁷ of jets with the highest output score are taken to be b-jets in successive steps of the analysis.

⁴⁷⁸ 8.2.1 2lSS Channel

⁴⁷⁹ For the 2lSS channel, the input features shown in table 3 are used for training. Here j_0 and j_1 are
⁴⁸⁰ the two jet candidates, while l_0 and l_1 are the two leptons in the event, both ordered by p_T . jet
⁴⁸¹ DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,
⁴⁸² where 5 represents the tightest working point and 1 represents the loosest. The variables nJets
⁴⁸³ DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and
⁴⁸⁴ 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 3: Input features used in the b-jet identification algorithm for the 2lSS channel

485 As there are far more incorrect combinations than correct ones, by a factor of more than
 486 20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random
 487 sample of 5 million incorrect entries are used for training, along with close 1 million correct
 488 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for
 489 training.

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

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

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

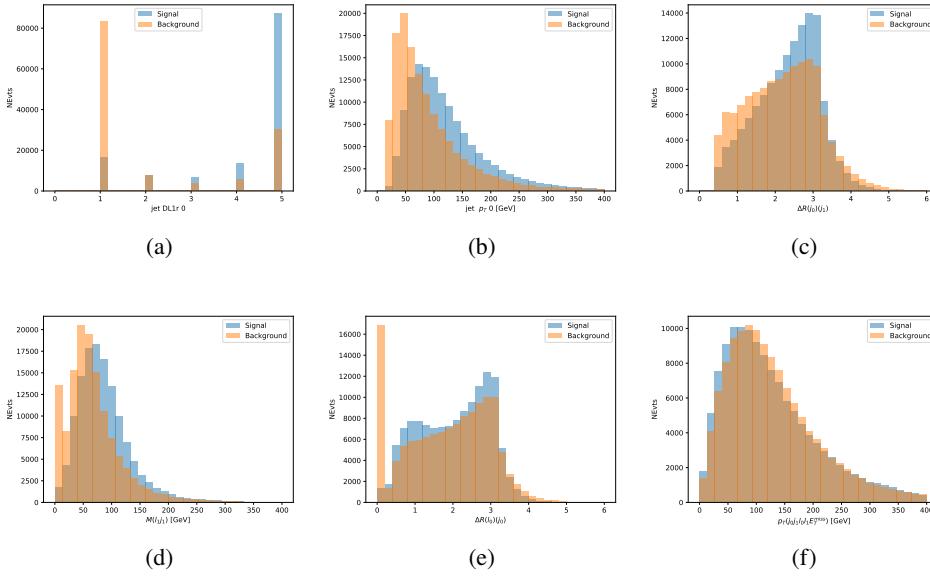


Figure 8.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T 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} .

498 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as
 499 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for
 500 the trained model are shown in figure 8.2.1. The model is found to identify the correct pairing
 501 of jets for 73% of 2ISS signal events on test data.

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

506 The accuracy of the model for different values of n-bjets, compared to this naive approach,

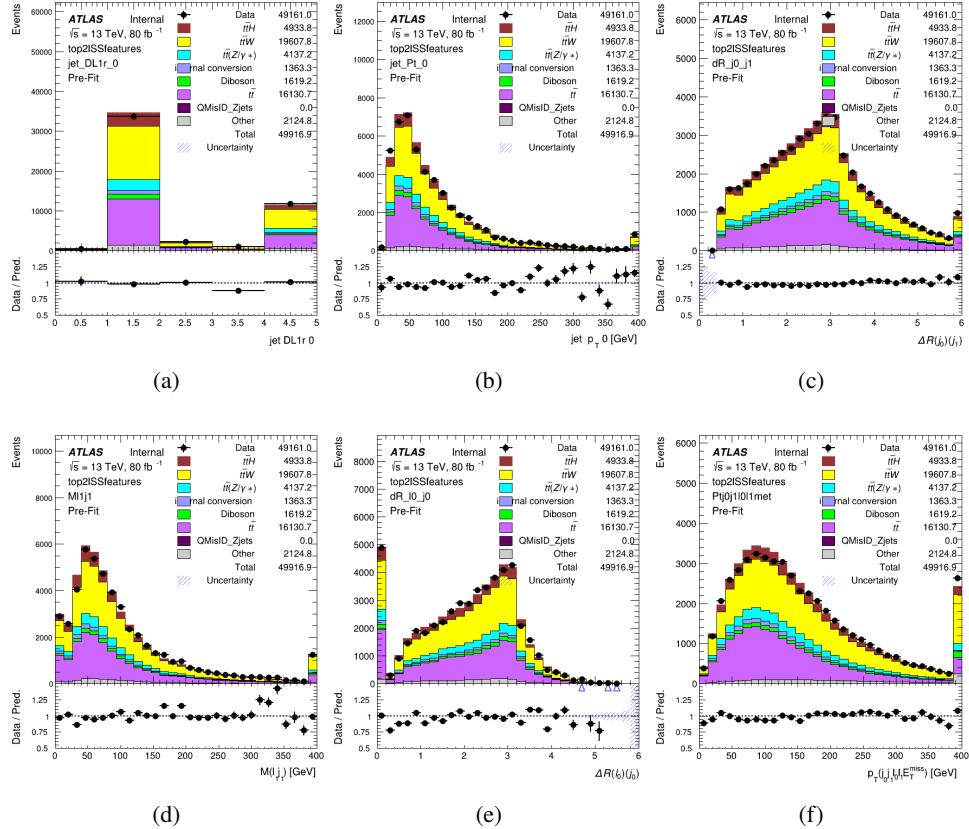


Figure 8.2: Data/MC comparisons of input features for top2ISS training for 80 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

is shown in table 4.

Table 4: Accuracy of the NN in identifying b-jets from tops in 2ISS events for, compared to the accuracy of taking the two highest b-tagged jets.

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

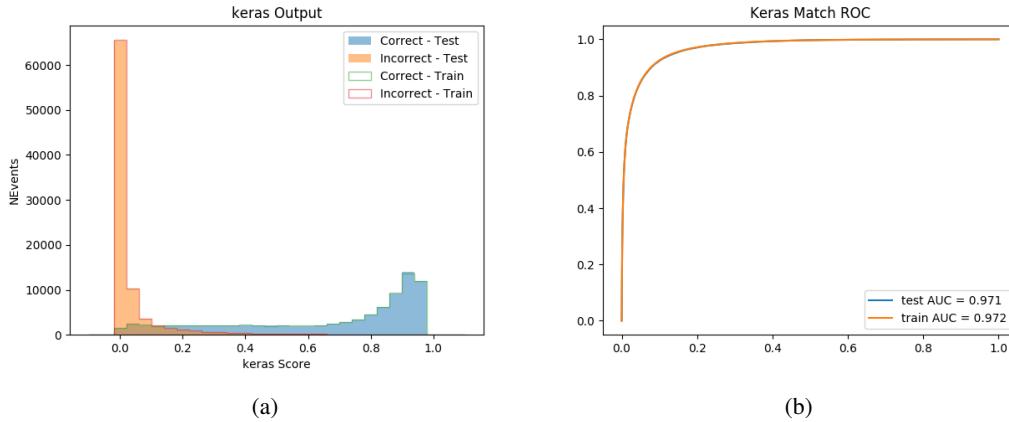


Figure 8.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

508 8.2.2 3l Channel

509 The input features used in the 3l channel are listed in table 5, with the same naming convention
 510 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 5: Input features for the b-jet identification algorithm in the 3l channel.

511 A few of these features are shown in figure 8.4, comparing the distributions for correct and

512 incorrect combinations of jets.

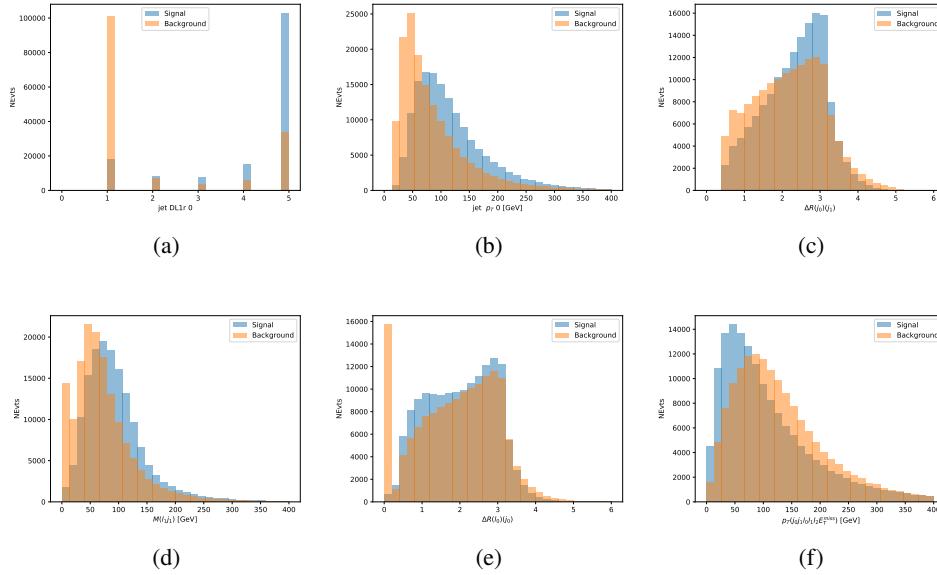


Figure 8.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

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

521 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal

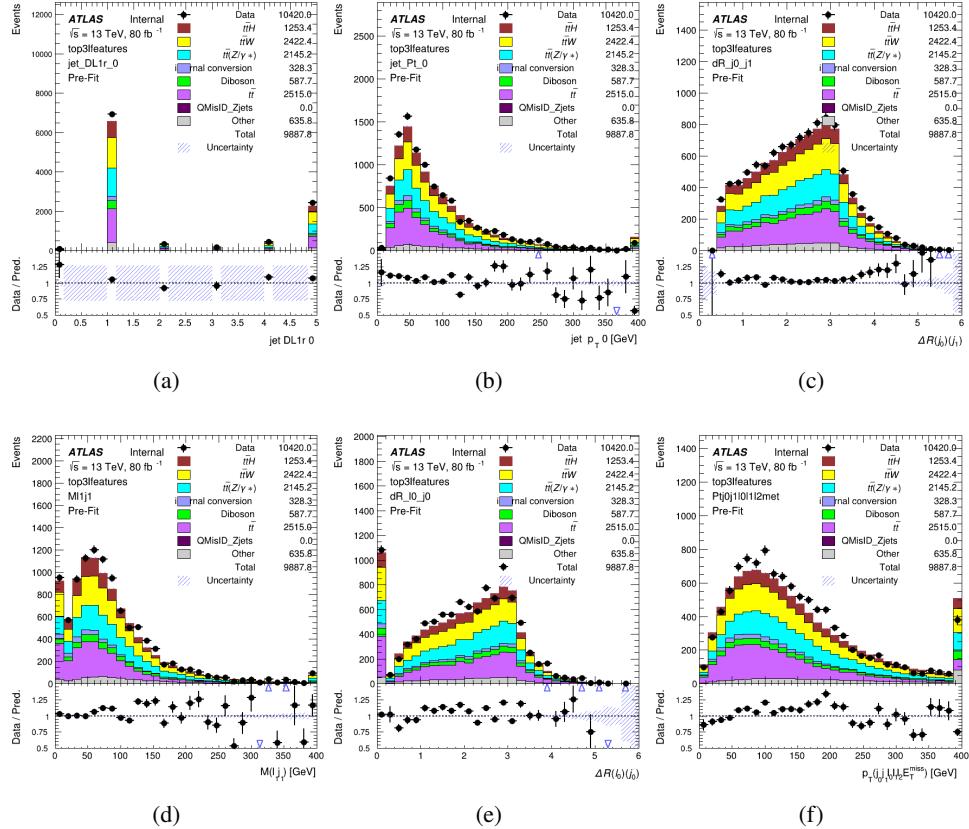


Figure 8.5: Data/MC comparisons of input features for top3l training for 80 fb^{-1} of data.

522 events. The accuracy of the model is summarized in table 6.

Table 6: Accuracy of the NN in identifying b-jets from tops, compared to the naive method of taking the highest b-tagged jets.

	NN	Naive
1 b-jet	69.0%	48.9%
2 b-jets	89.6%	88.3%
>=3 b-jets	55.7%	52.3%
Overall	79.8%	70.2%

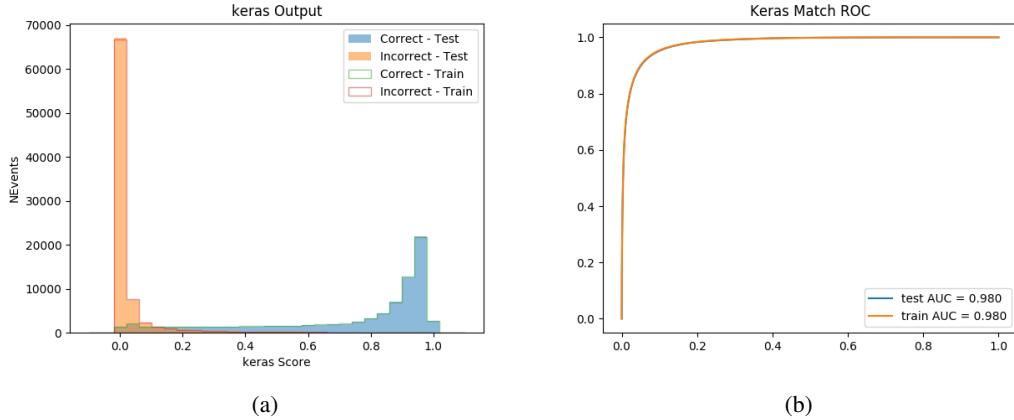


Figure 8.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

523 8.3 Higgs Reconstruction

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

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

531 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In
 532 the semi-leptonic case the reconstructed decay products consist of two jets and a single leptons.
 533 For the fully-leptonic case, the decay products include 2 of the three leptons associated with the

534 event. For training the models, events are separated into these two categories using truth level
535 information. A separate MVA, described in section 8.5, is used to make this distinction at reco
536 level and determine which model to use.

537 For all channels, the models described in section 8.2 are used to identify b-jet candidates,
538 whose kinematics are used to identify the Higgs decay products. These jets are not considered
539 as possible candidates for the Higgs decay, justified by the fact that these models are found to
540 misidentify jets from the Higgs decay as jets from the top decay less than 1% of the time.

541 **8.3.1 2lSS Channel**

542 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural
543 network is trained on the kinematics of different combinations of leptons and jets, as well as the
544 b-jets identified in section 8.2, with the specific input features listed in table 7.

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

Table 7: Input features used to identify the Higgs decay products in 2lSS events

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

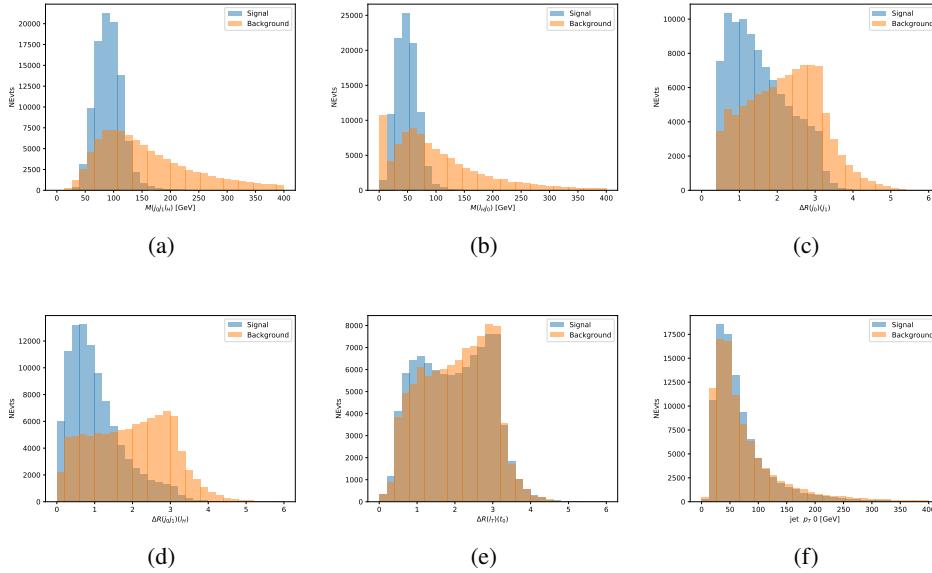


Figure 8.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

552 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2
 553 million events, with an additional 200,000 reserved for testing the model. In order to compensate
 554 for large number of incorrect combinations, these have been downsampled such that the correct

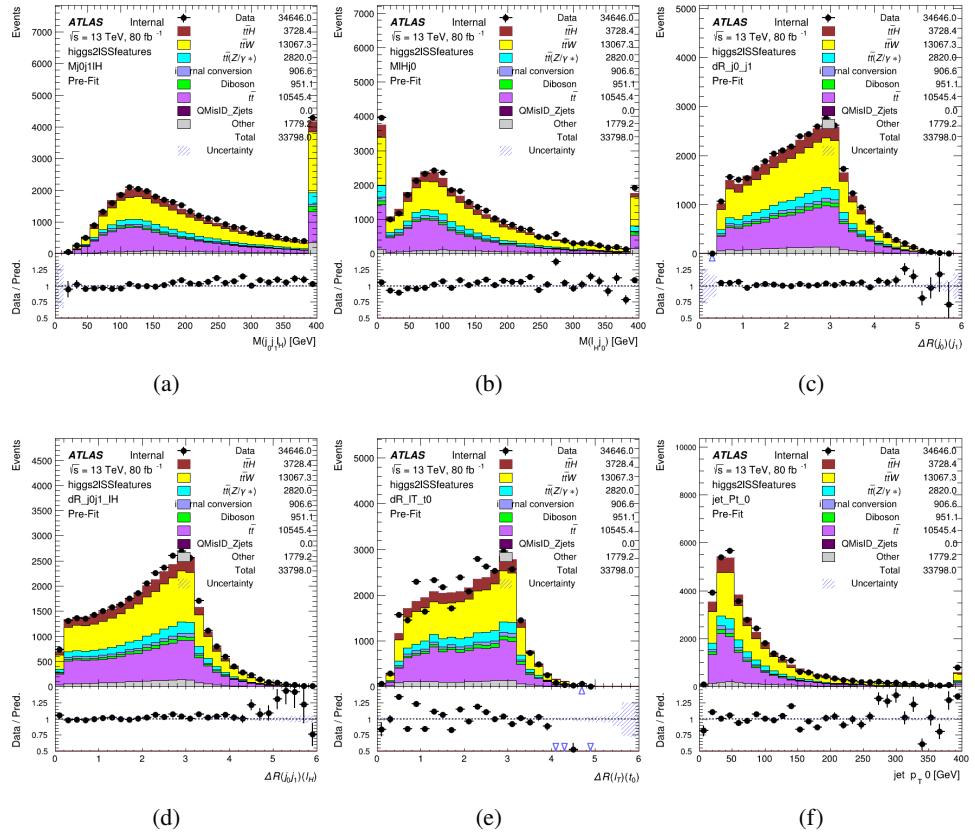


Figure 8.8: Data/MC comparisons of input features for higgs2lSS training for 80 fb^{-1} of data.

556 combinations represent over 10% of the training set. The output of the NN is summarized in
556 figure 8.3.1.

557 The neural network identifies the correct combination 55% of the time. It identifies the
558 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
559 81% of the time.

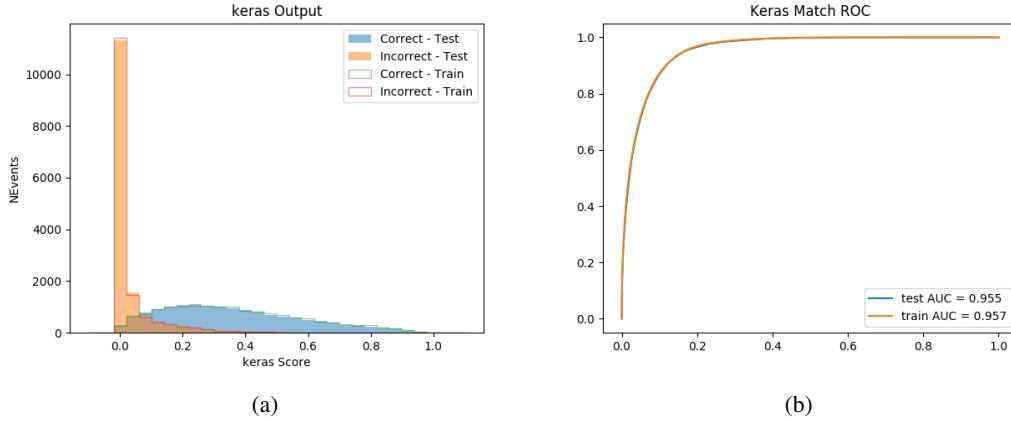


Figure 8.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

560 8.3.2 3l Semi-leptonic Channel

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

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

Table 8: Input features used to identify the Higgs decay products in 3lS events

565 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 566 two leptons in the event are labeled as l_{T0} and l_{T1} . b_0 and b_1 are the two b-jets identified by
 567 the b-jet identification algorithm. The b-jet Reco Score is the output of the Higgs reconstruction
 568 algorithm.

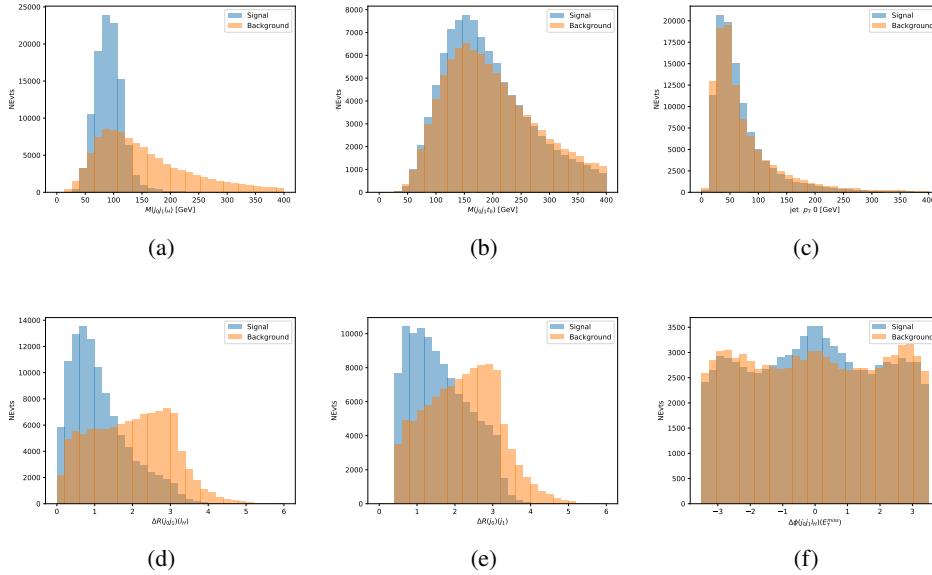


Figure 8.10: Input features for higgs3IS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

569 The modeling of these inputs is validated against data, with figure 8.11 showing good
 570 general agreement between data and MC. Plots for the complete list of features can found in
 571 appendix A.1.

572 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.
 573 Once again, incorrect combinations are downsampled, such that the correct combinations are
 574 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN

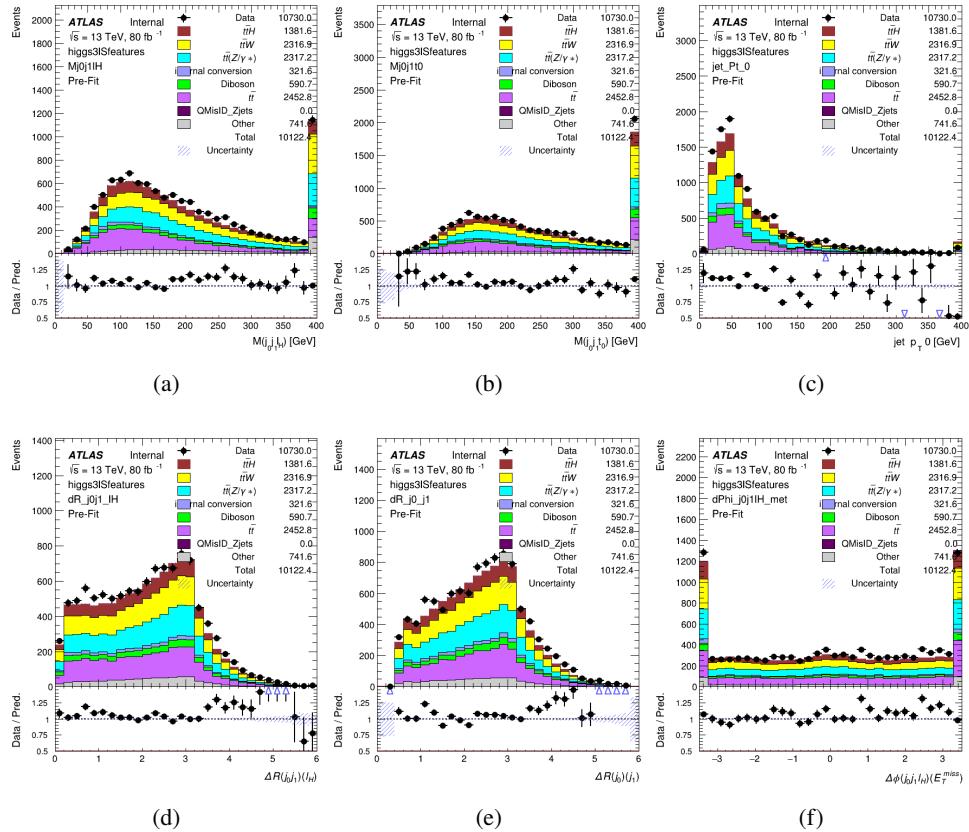


Figure 8.11: Data/MC comparisons of input features for higgs3IS training for 80 fb^{-1} of data.

575 is summarized in figure 8.3.2.

576 The neural network identifies the correct combination 64% of the time. It identifies the
 577 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
 578 83% of the time.

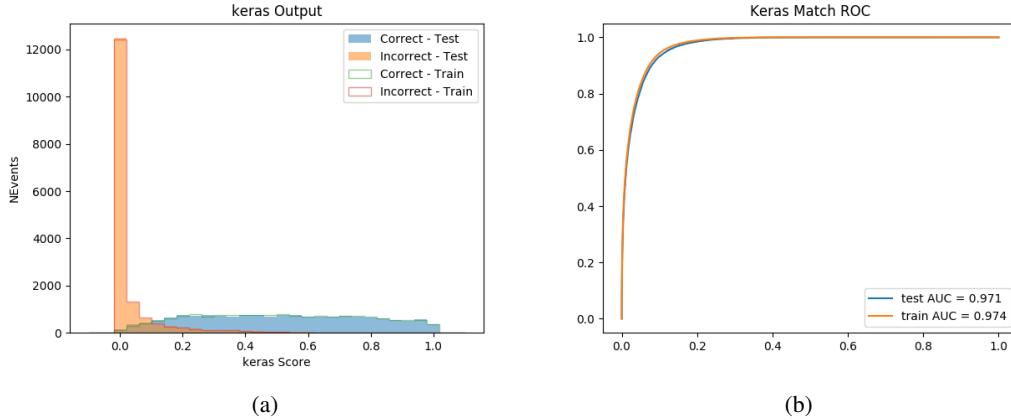


Figure 8.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

579 8.3.3 3l Fully-leptonic Channel

580 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 581 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to
 582 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 583 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 584 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}	topScore

Table 9: Input features used to identify the Higgs decay products in 3IF events

585 Here l_{H0} and l_{H1} are the Higgs decay candidates. The other lepton in the event is labeled
 586 l_T . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco
 587 Score is the output of the Higgs reconstruction algorithm.

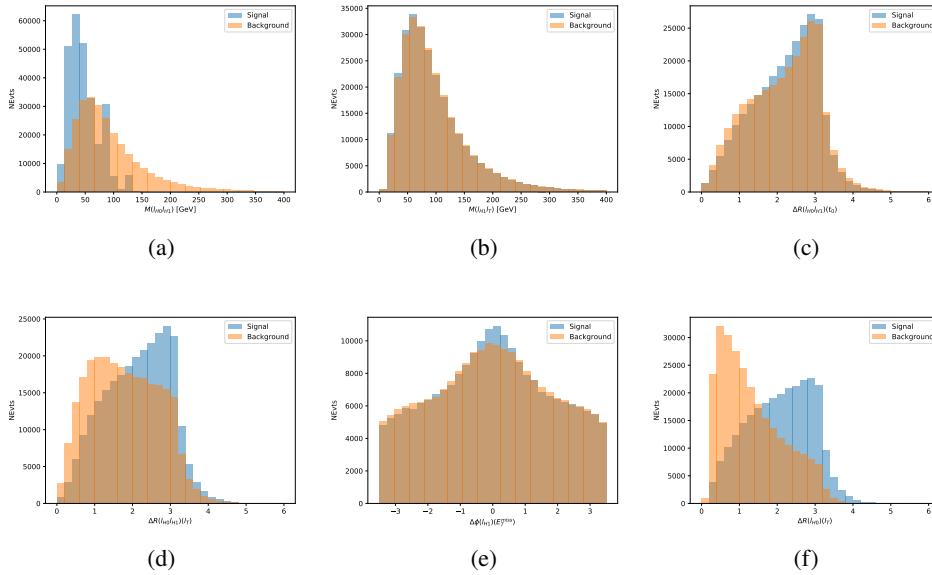


Figure 8.13: Input features for higgs3IF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

591 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,
 592 with 10% of the dataset reserved for testing. The output of the model is summarized in figure
 593 8.3.3.

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

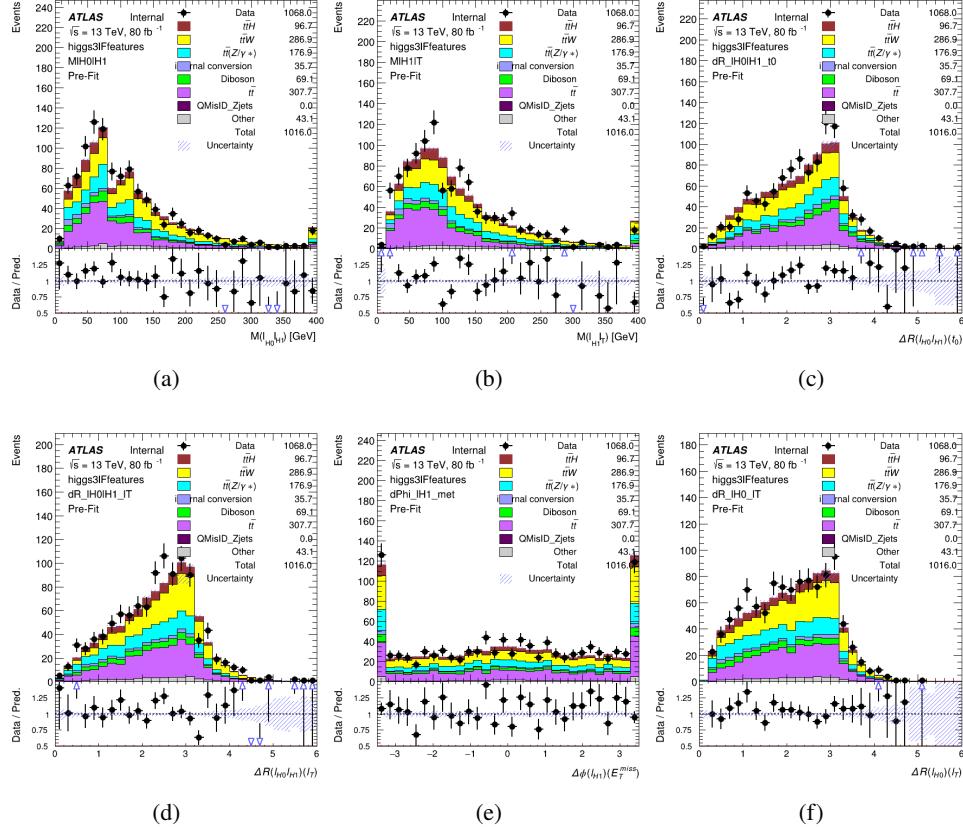


Figure 8.14: Data/MC comparisons of input features for higgs3IF training for 80 fb^{-1} of data.

595 8.4 p_T Prediction

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

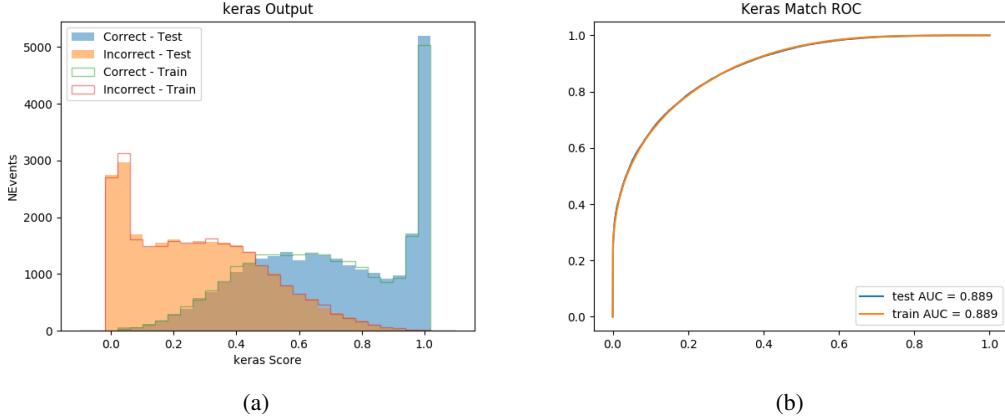


Figure 8.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

605 Because the analysis uses a two bin fit of the Higgs p_T , the momentum reconstruction
 606 could be treated as a binary classification problem, rather than a regression problem. This
 607 approach is explored in detail in section A.4, and is found not to provide any significant increase
 608 in sensitivity. The regression approach is used because it provides more flexibility for future
 609 analyses, as it is independent of the cutoff between high and low p_T , as well as the number of
 610 bins. Further, a regression allows the output of the neural network to be more clearly understood,
 611 as it can be directly compared to a physics observable.

612 **8.4.1 2ISS Channel**

613 The input variables listed in table 10 are used to predict the Higgs p_T in the 2ISS channel. Here
614 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
615 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
616 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
617 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
618 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 10: Input features for reconstructing the Higgs p_T spectrum for 2lSS events

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

622 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 623 truth Higgs p_T in figure 8.16. In order to visualize the model performance more clearly, in (a)
 624 of that figure, the color of each point is determined by Kernal Density Estimation (KDE). The
 625 color shown represents the logarithm of the output from KDE, to counteract the large number of
 626 low p_T events. For that same reason, each column of the histogram shown in (b) of figure 8.16
 627 is normalized to unity. This plot therefore demonstrates what the model predicts for each slice
 628 of truth p_T .

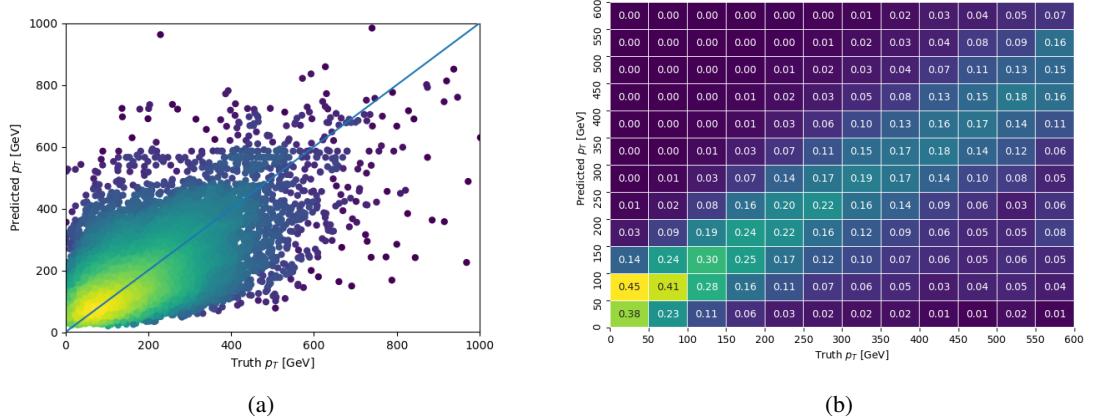


Figure 8.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2ISS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

629 We are also interested in how well the model distinguishes between events with $p_T < 150$
 630 GeV and > 150 GeV. Figure 8.17 demonstrates the NN output for high and low p_T events based

631 on this cutoff.

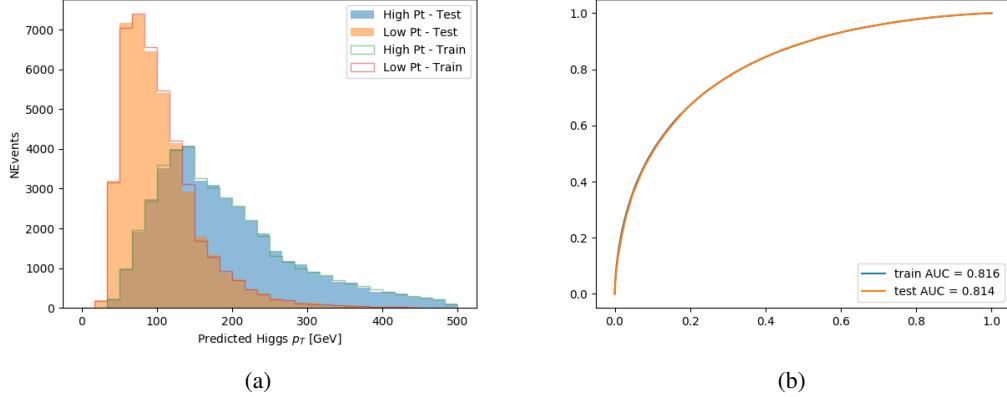


Figure 8.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

632 8.4.2 3l Semi-leptonic Channel

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

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

Table 11: Input features for reconstructing the Higgs p_T spectrum for 3lS events

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

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

642 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 643 truth Higgs p_T in figure 8.18. Once again, (a) of 8.18 shows a scatterplots of predicted vs truth
 644 p_T , where the color of each point corresponds to the log of the relative KDE at that point. Each
 645 column of the the histogram in (b) is normalized to unity, to better demonstrate the output of the
 646 NN for each slice of truth p_T .

647 Figure 8.19 shows (a) the output of the NN for events with truth p_T less than and greater
 648 than 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low
 649 p_T events.

650 **8.4.3 3l Fully-leptonic Channel**

651 The features listed in 12 are used to construct a model for predictin the Higgs p_T for 3lF events.

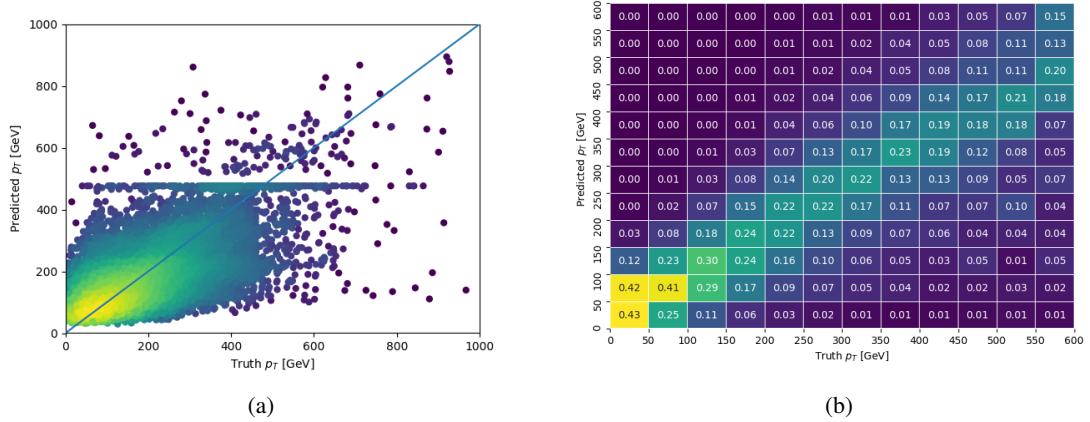


Figure 8.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

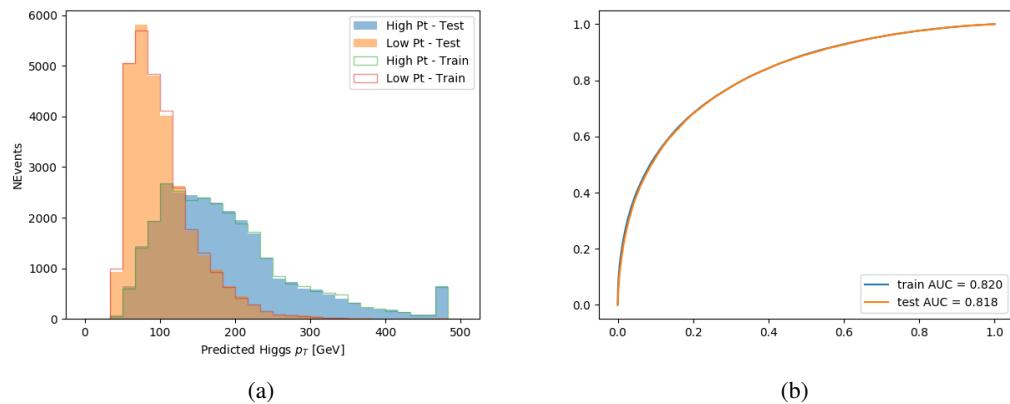


Figure 8.19: (a) shows the reconstructed Higgs p_T for 3lS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

HT	$M(l_{H0}l_{H1})$	$M(l_{H0}l_T)$
$M(l_{H0}b_0)$	$M(l_{H0}b_1)$	$M(l_{H1}l_T)$
$M(l_{H1}b_0)$	$M(l_{H1}b_1)$	$\Delta R(l_{H0})(l_{H1})$
$\Delta R(l_{H0})(l_T)$	$\Delta R(l_{H0}l_{H1})(l_T)$	$\Delta R(l_{H0}l_T)(l_{H1})$
$\Delta R(l_{H1})(l_T)$	$\Delta R(l_{H0}b_0)$	$\Delta R(l_{H0}b_1)$
$\Delta R(l_{H1}b_1)$	$\Delta R(l_{H1}b_0)$	higgsScore
Lepton 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}	topScore	b-jet p_T 0
b-jet p_T 1		

Table 12: Input features for reconstructing the Higgs p_T spectrum for 3lF events

652 l_{H0} and l_{H1} represent the two leptons identified by the Higgs reconstruction model as
 653 originating from the Higgs, while l_T is the other lepton in the event. The Higgs Reco Score and
 654 b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification
 655 algorithm, respectively.

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

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

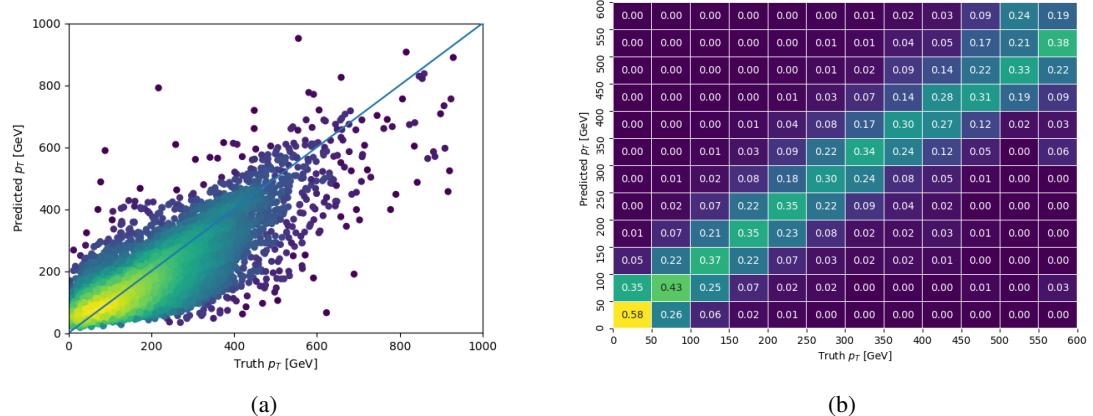


Figure 8.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

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

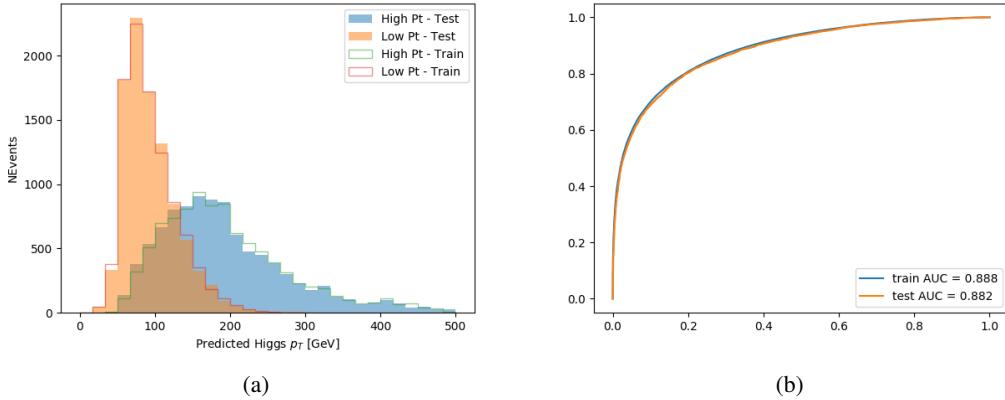


Figure 8.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

661 8.5 3l Decay Mode

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

668 The kinematics of each event, along with the output scores of the Higgs and top recon-
 669 struction algorithms, are used to distinguish these two possible decay modes. The particular
 670 inputs used are listed in table 13.

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 13: Input features

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

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

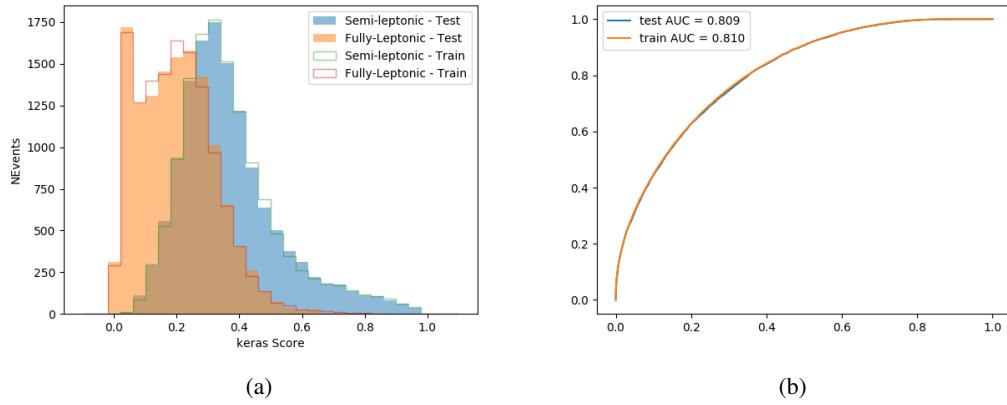


Figure 8.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

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

677 9 Signal Region Definitions

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

681 9.1 Pre-MVA Event Selection

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

- 684 • Two very tight, same-charge, light leptons with $p_T > 20 \text{ GeV}$
- 685 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets
- 686 • No reconstructed tau candidates

687 The event yield after the 2lSS preselection has been applied, for MC and data at 80 fb^{-1} ,
 688 is shown in figure 9.1. Good general agreement is found.

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

- 690 • Three light leptons with total charge ± 1
- 691 • Same charge leptons are required to be very tight, with $p_T > 20 \text{ GeV}$
- 692 • Opposite charge lepton must be loose, with $p_T > 10 \text{ GeV}$
- 693 • ≥ 2 reconstructed jets, ≥ 1 b-tagged jets
- 694 • No reconstructed tau candidates
- 695 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs

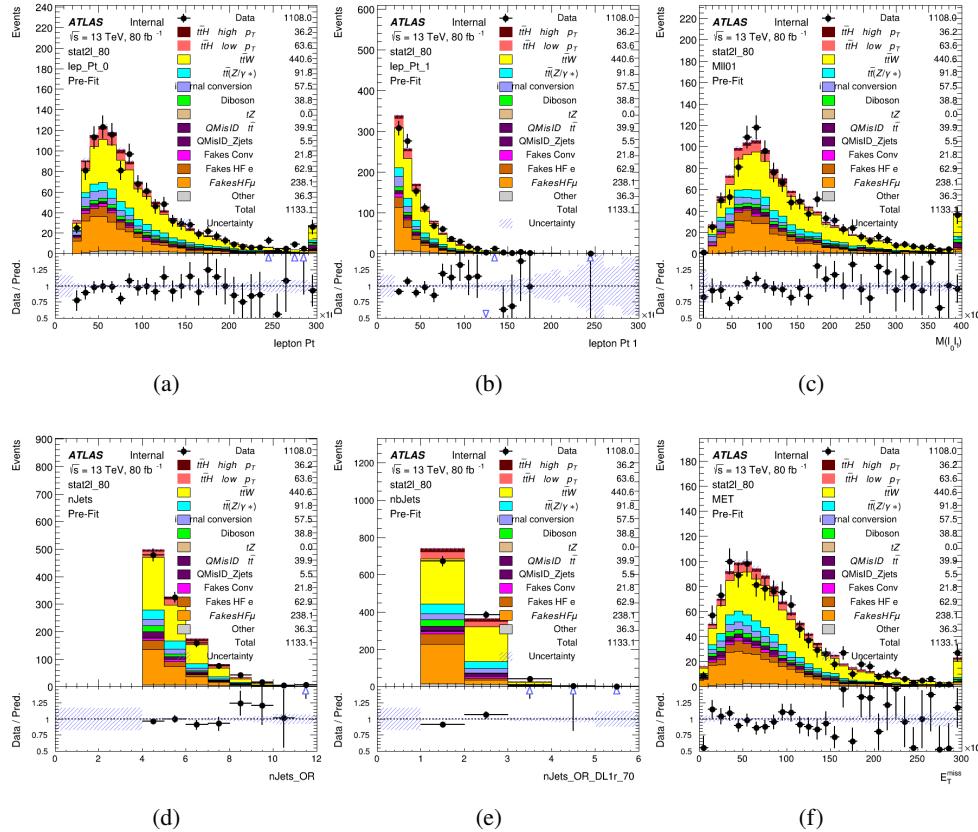


Figure 9.1: Data/MC comparisons of the 2lSS pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

696 Comparisons of data and MC event yields after this selection has been applied are shown
697 in figure 9.2.

698 9.2 Event MVA

699 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
700 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
701 (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

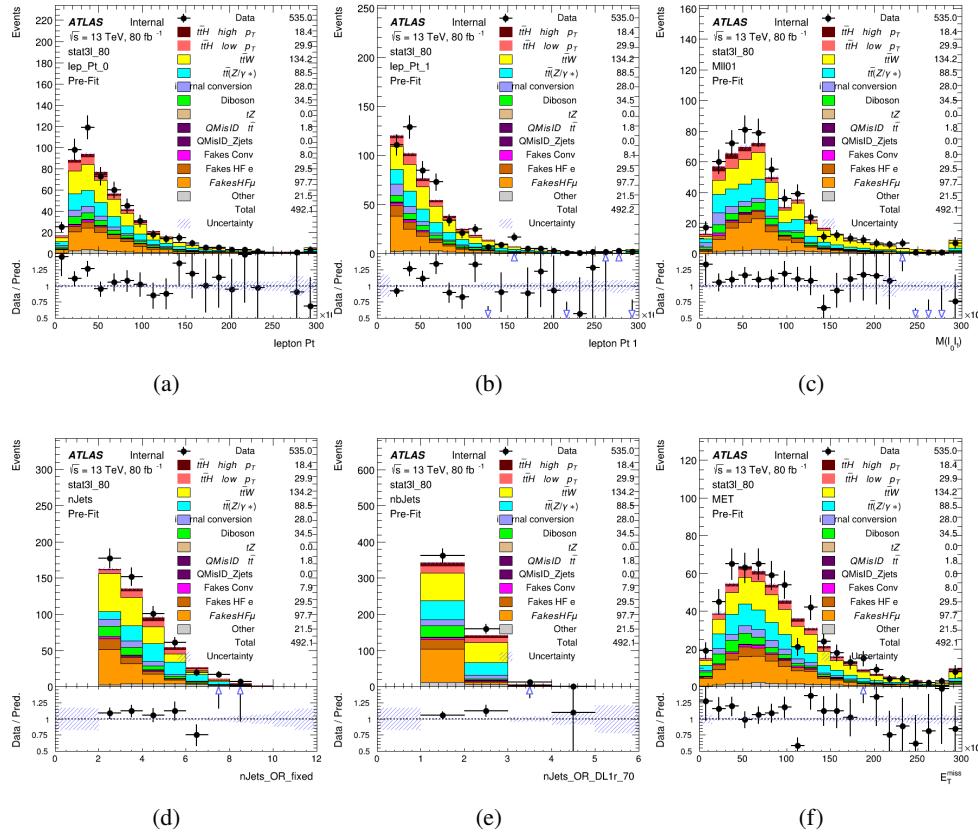


Figure 9.2: Data/MC comparisons of the 31 pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

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

705 Because the background composition differs for events with a high reconstructed Higgs p_T
 706 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and
 707 low p_T regions. This is found to provide better significance than attempting to build an inclusive
 708 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6

⁷⁰⁹ background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,

⁷¹⁰ 3lF high p_T , and 3lF low p_T .

⁷¹¹ The following features are used in both the high and low p_T 2lSS BDTs:

HT	$M(\text{lep}, E_T^{\text{miss}})$	$M(l_0 l_1)$
binHiggs p_T 2lSS	$\Delta R(l_0)(l_1)$	diLepton type
higgsRecoScore	jet η 0	jet η 1
jet ϕ 0	jet ϕ 1	jet p_T 0
jet p_T 1	Lepton η 0	Lepton η 1
Lepton ϕ 0	Lepton ϕ 1	Lepton p_T 0
Lepton p_T 1	E_T^{miss}	$\min \Delta R(l_0)(\text{jet})$
$\min \Delta R(l_1)(\text{jet})$	$\min \Delta R(\text{Lepton})(\text{bjet})$	mjjMax frwdJet
nJets	nJets OR DL1r 60	nJets OR DL1r 70
nJets OR DL1r 85	topRecoScore	

Table 14: Input features used to distinguish signal and background events in the 2lSS channel.

712

While for each of the 31 BDTs, the features listed below are used for training:

$M(\text{lep}, E_T^{\text{miss}})$	$M(l_0 l_1)$	$M(l_0 l_1 l_2)$
$M(l_0 l_2)$	$M(l_1 l_2)$	$\text{binHiggs } p_T \text{ 3lF}$
$\text{binHiggs } p_T \text{ 3lS}$	$\Delta R(l_0)(l_1)$	$\Delta R(l_0)(l_2)$
$\Delta R(l_1)(l_2)$	decayScore	higgsRecoScore3lF
higgsRecoScore3lS	$\text{jet } \eta \text{ 0}$	$\text{jet } \eta \text{ 1}$
$\text{jet } \phi \text{ 0}$	$\text{jet } \phi \text{ 1}$	$\text{jet } p_T \text{ 0}$
$\text{jet } p_T \text{ 1}$	$\text{Lepton } \eta \text{ 0}$	$\text{Lepton } \eta \text{ 1}$
$\text{Lepton } \eta \text{ 2}$	$\text{Lepton } \phi \text{ 0}$	$\text{Lepton } \phi \text{ 1}$
$\text{Lepton } \phi \text{ 2}$	$\text{Lepton } p_T \text{ 0}$	$\text{Lepton } p_T \text{ 1}$
$\text{Lepton } p_T \text{ 2}$	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 15: Input features used to distinguish signal and background events in the 3l channel.

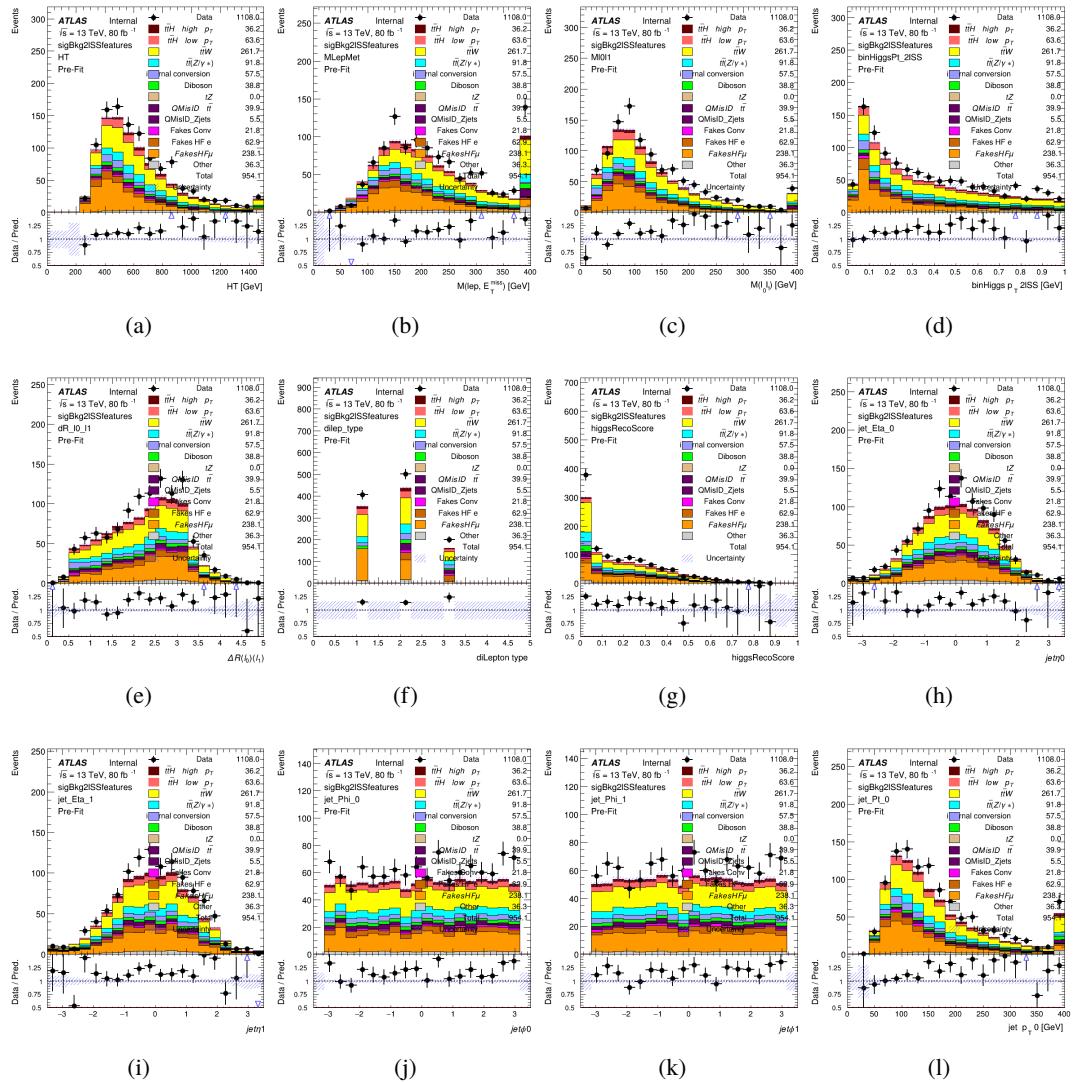


Figure 9.3:

713 The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss
714 function.

Output distributions of each MVA comparing MC prediction to data at 80 fb^{-1} are shown in figure 9.2.

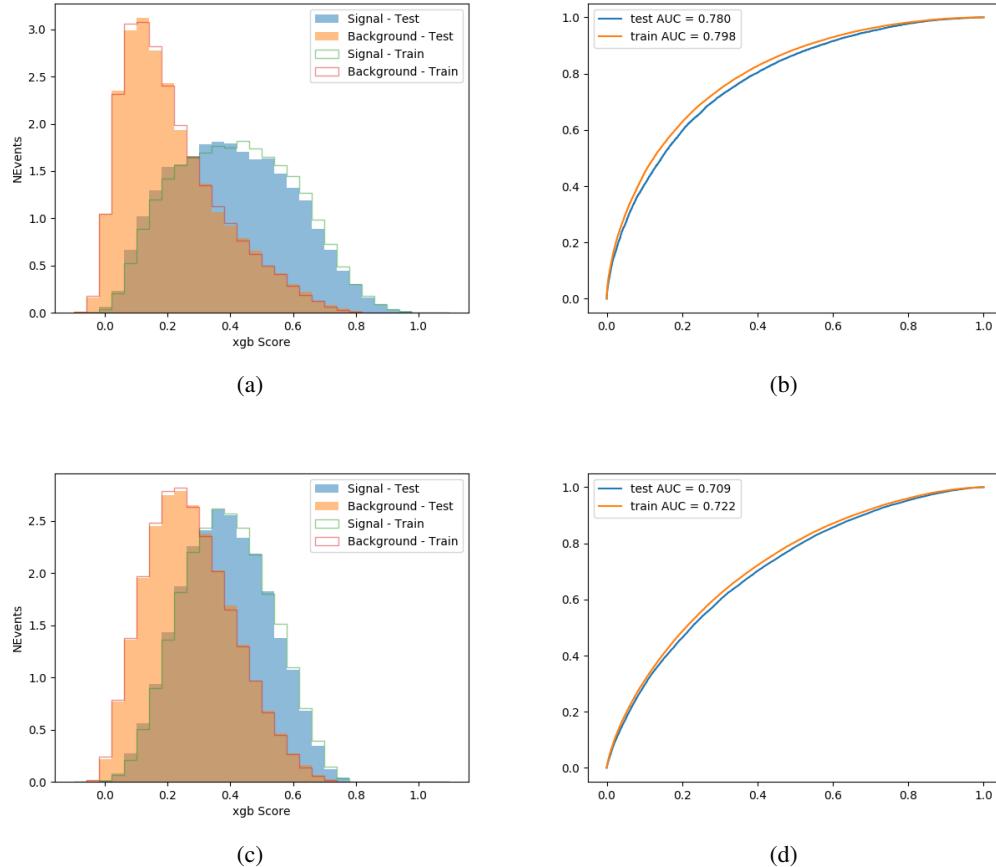


Figure 9.4:

717 9.3 Signal Region Definitions

718 Once pre-selection has been applied, channels are further refined based on the MVAs described
 719 above. The output of the model described in section 8.5 is used to separate the three channel into
 720 two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson.
 721 This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

722 For each event, depending on the number of leptons as well as whether the p_T of the Higgs

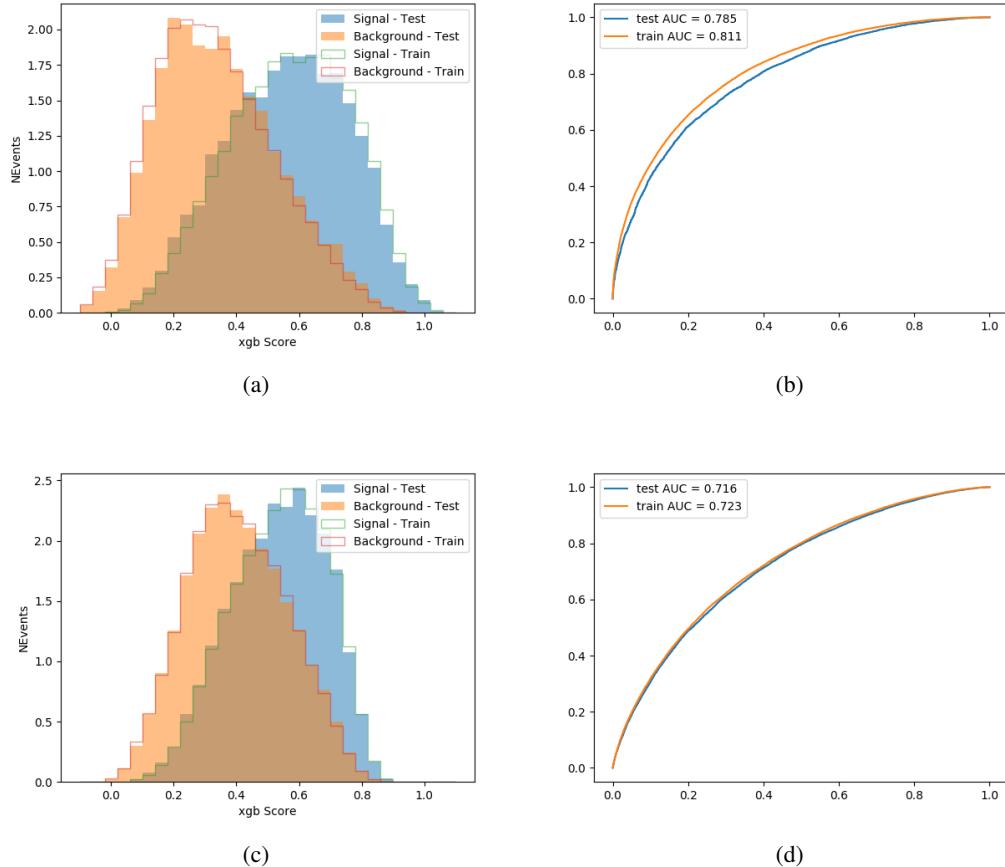


Figure 9.5:

is predicted to be high (> 150 GeV) or low (< 150 GeV), a cut on the appropriate background rejection MVA is applied. The particular cut values, listed in table ??, are determined by maximizing S/\sqrt{B} in each region.

The event preselection and MVA selection define the three signal regions. These signal region definitions are summarized in table ??.

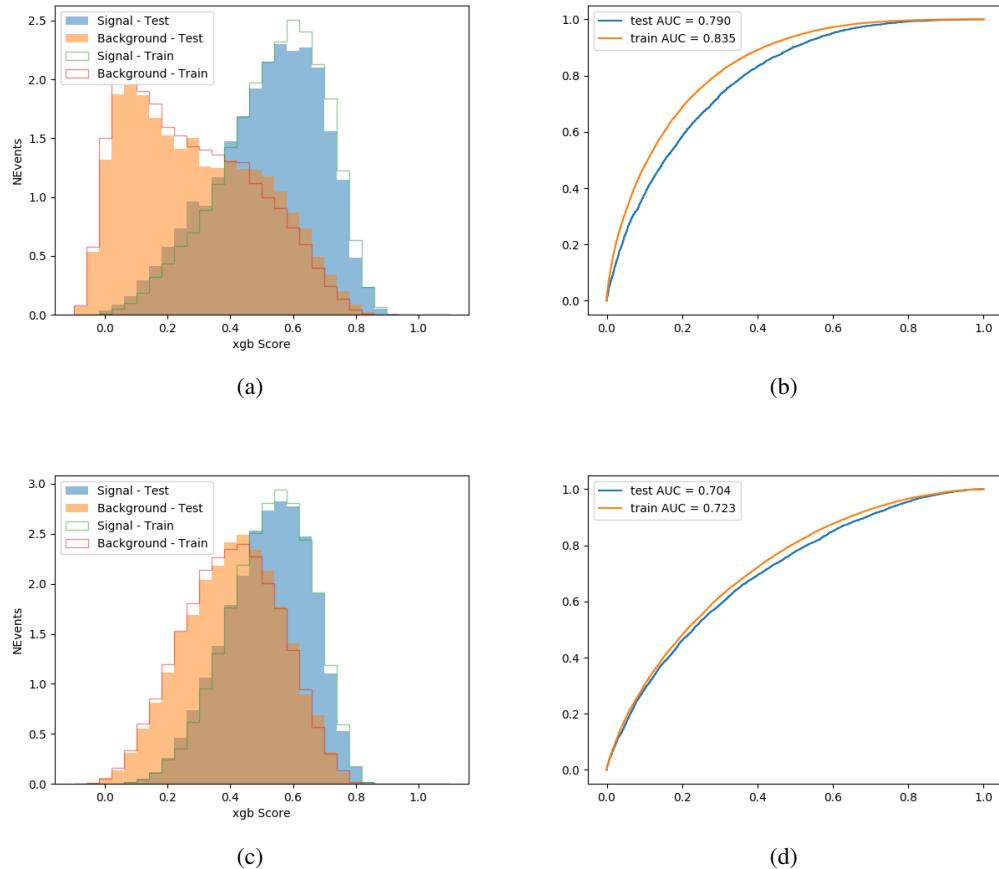


Figure 9.6:

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

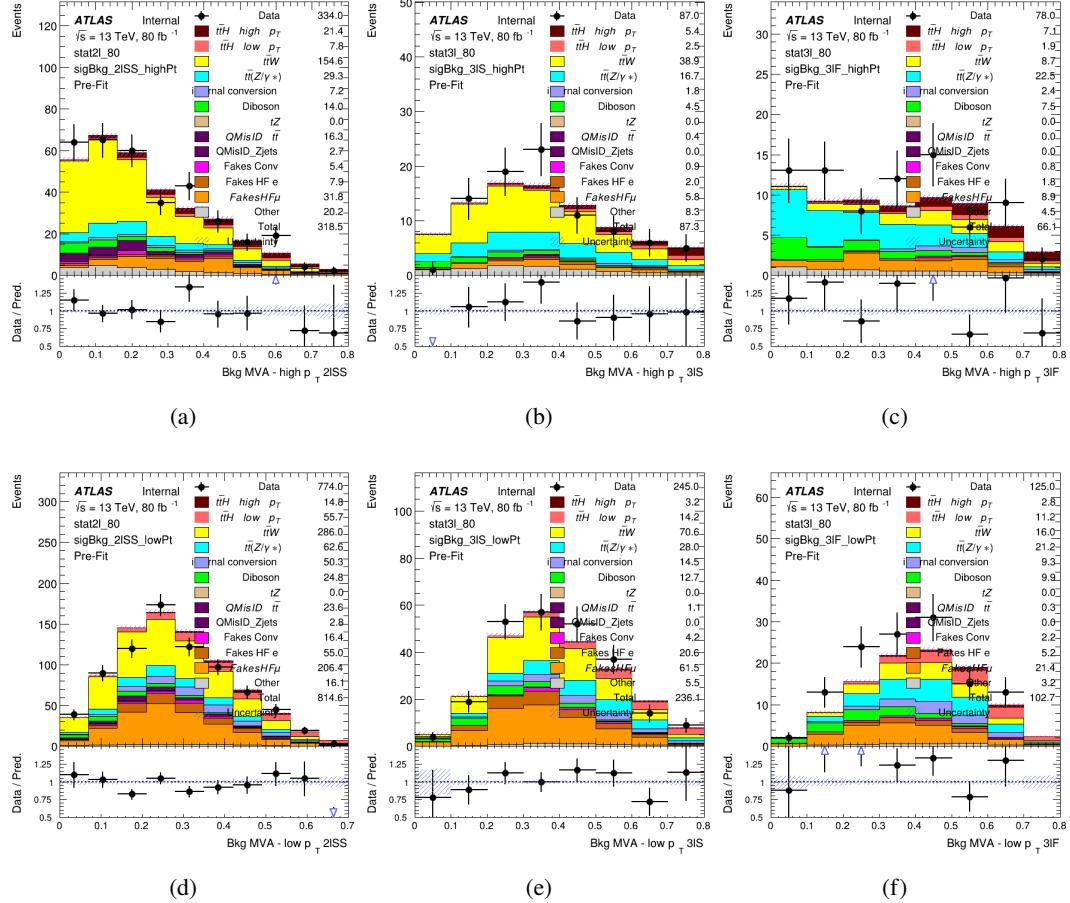


Figure 9.7: scores

10 Background Rejection MVA

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

Region	Selection
2ISS	Two same charge tight leptons with $p_T > 20$ GeV $N_{\text{jets}} \geq 4$, $N_{\text{b-jets}} \geq 1$ b-tagged jets Zero τ_{had} $H_{p_T}^{\text{pred}} > 150$ GeV and BDT score > 0.36 or $H_{p_T}^{\text{pred}} < 150$ GeV and BDT score > 0.34
3IS	Three light leptons with total charge ± 1 Two tight SS leptons, $p_T > 20$ GeV One loose OS lepton, $p_T > 10$ 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 $H_{p_T}^{\text{pred}} > 150$ GeV and BDT score > 0.51 or $H_{p_T}^{\text{pred}} > 150$ GeV and BDT score > 0.43
3IF	Three light leptons with total charge ± 1 Two tight SS leptons, $p_T > 20$ GeV One loose OS lepton, $p_T > 10$ 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 $H_{p_T}^{\text{pred}} > 150$ GeV and BDT score > 0.33 or $H_{p_T}^{\text{pred}} > 150$ GeV and BDT score > 0.41

Table 17: Selection applied to define the three signal regions used in the fit.

732 **10.1 Pre-MVA Event Selection**

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

- 735 • Two very tight, same-charge, light leptons with $p_T > 20$ GeV
 736 • $>=4$ reconstructed jets, $>=1$ b-tagged jets
 737 • No reconstructed tau candidates

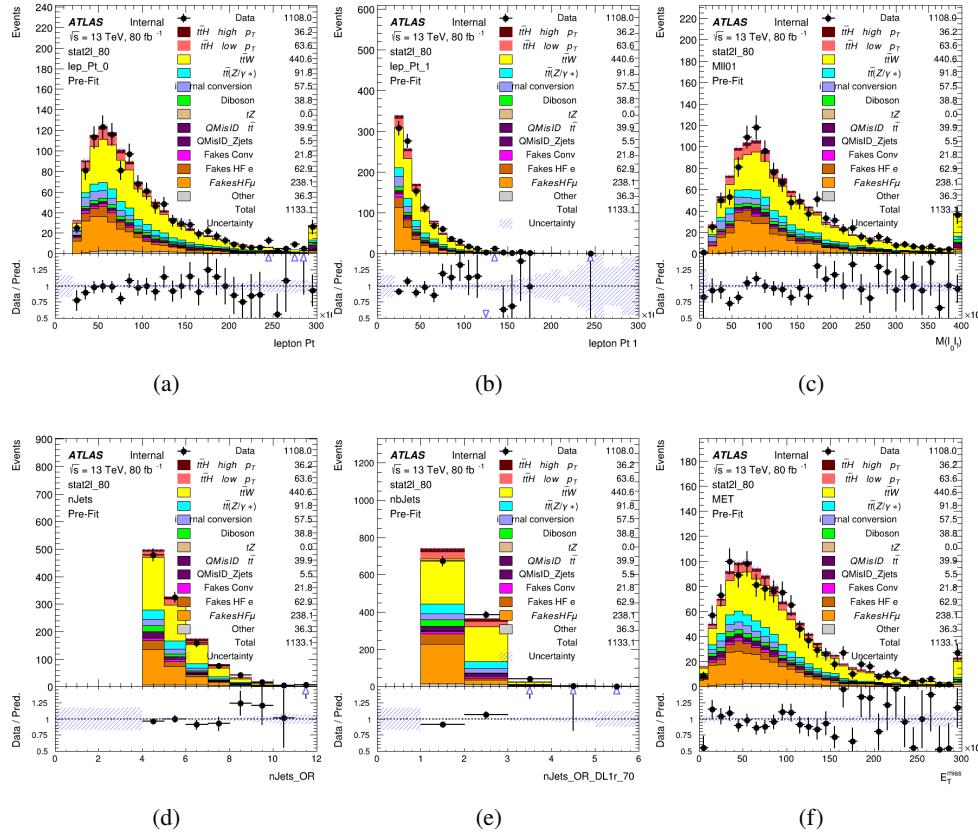


Figure 10.1:

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

- 739 • Three light leptons with total charge ± 1
- 740 • Same charge leptons are required to be very tight, with $p_T > 20$ GeV
- 741 • Opposite charge lepton must be loose, with $p_T > 10$ GeV
- 742 • $>=2$ reconstructed jets, $>=1$ b-tagged jets
- 743 • No reconstructed tau candidates

744

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

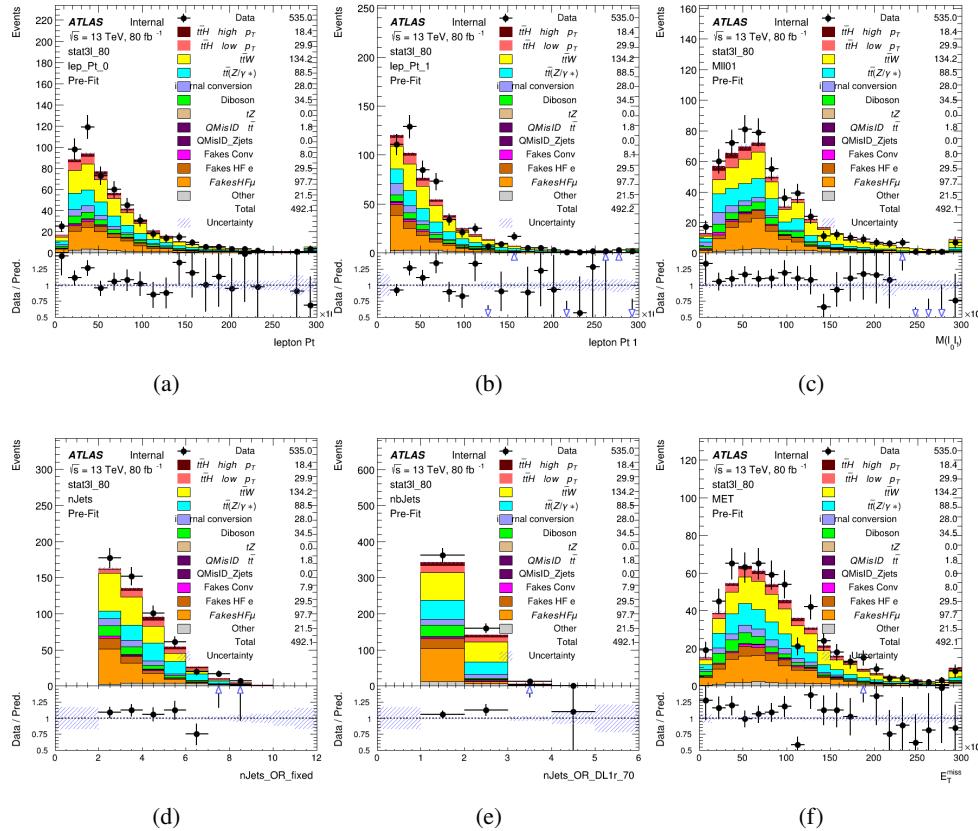


Figure 10.2:

745

10.2 Event MVA

746 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
 747 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
 748 (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost
 749 [xgboost] are trained using the kinematics of signal and background events derived from Monte
 750 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo

751 event.

752 Because the background composition differs for events with a high reconstructed Higgs p_T
753 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and
754 low p_T regions. This is found to provide better significance than attempting to build an inclusive
755 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6
756 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
757 3lF high p_T , and 3lF low p_T .

758 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 2ISS	$\Delta R(l_0)(l_1)$	diLepton type
higgsRecoScore	jet η 0	jet η 1
jet ϕ 0	jet ϕ 1	jet p_T 0
jet p_T 1	Lepton η 0	Lepton η 1
Lepton ϕ 0	Lepton ϕ 1	Lepton p_T 0
Lepton p_T 1	E_T^{miss}	$\min \Delta R(l_0)(\text{jet})$
$\min \Delta R(l_1)(\text{jet})$	$\min \Delta R(\text{Lepton})(\text{bjet})$	mjjMax frwdJet
nJets	nJets OR DL1r 60	nJets OR DL1r 70
nJets OR DL1r 85	topRecoScore	

Table 18: Input features used to distinguish signal and background events in the 2ISS channel.

759

While for each of the 31 BDTs, the features listed below are used for training:

$M(\text{lep}, E_T^{\text{miss}})$	$M(l_0 l_1)$	$M(l_0 l_1 l_2)$
$M(l_0 l_2)$	$M(l_1 l_2)$	$\text{binHiggs } p_T \text{ 3lF}$
$\text{binHiggs } p_T \text{ 3lS}$	$\Delta R(l_0)(l_1)$	$\Delta R(l_0)(l_2)$
$\Delta R(l_1)(l_2)$	decayScore	higgsRecoScore3lF
higgsRecoScore3lS	$\text{jet } \eta \text{ 0}$	$\text{jet } \eta \text{ 1}$
$\text{jet } \phi \text{ 0}$	$\text{jet } \phi \text{ 1}$	$\text{jet } p_T \text{ 0}$
$\text{jet } p_T \text{ 1}$	$\text{Lepton } \eta \text{ 0}$	$\text{Lepton } \eta \text{ 1}$
$\text{Lepton } \eta \text{ 2}$	$\text{Lepton } \phi \text{ 0}$	$\text{Lepton } \phi \text{ 1}$
$\text{Lepton } \phi \text{ 2}$	$\text{Lepton } p_T \text{ 0}$	$\text{Lepton } p_T \text{ 1}$
$\text{Lepton } p_T \text{ 2}$	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 19: Input features used to distinguish signal and background events in the 3l channel.

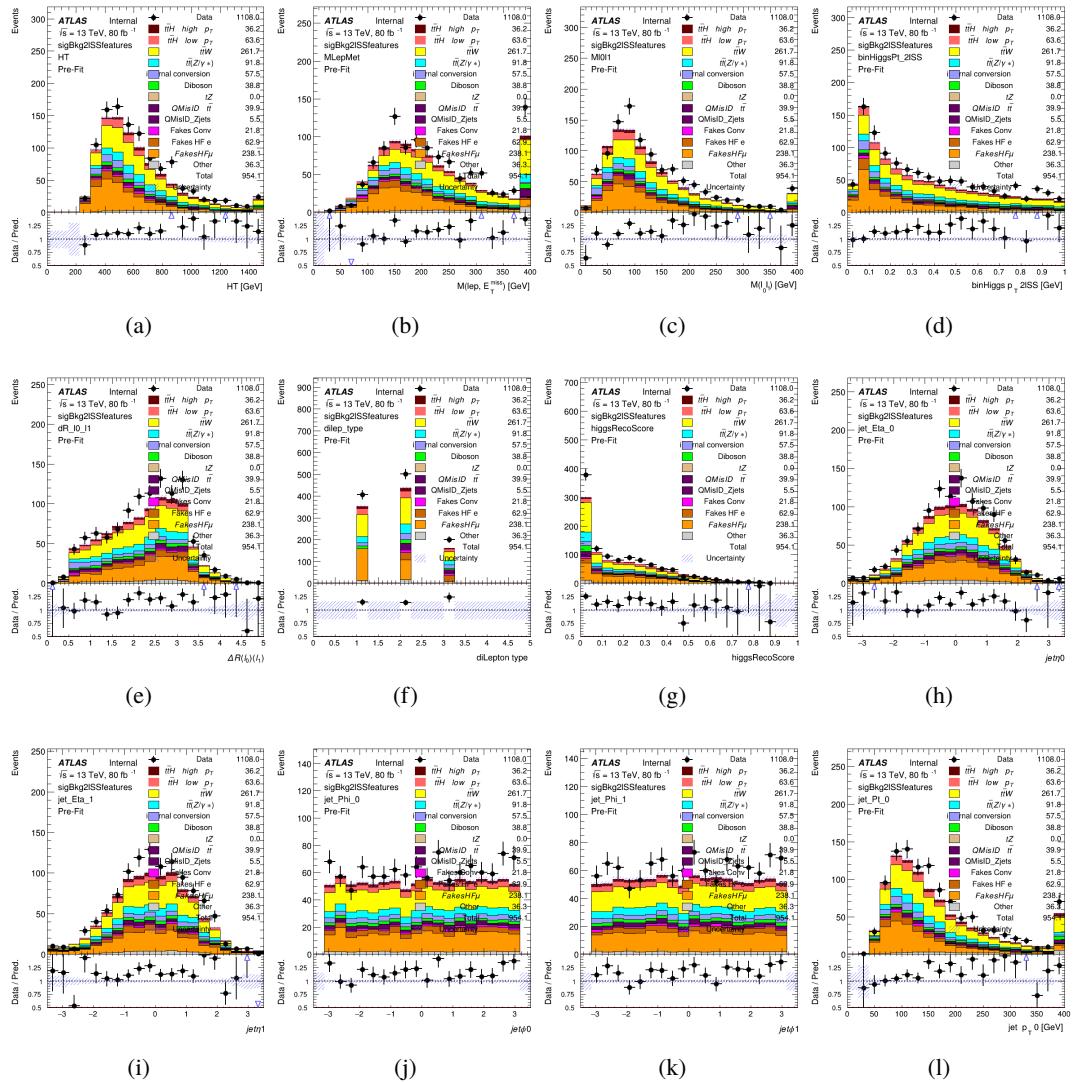


Figure 10.3:

760 The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss
 761 function.

762 Output distributions of each MVA comparing MC prediction to data at 80 fb^{-1} are shown
 763 in figure 9.2.

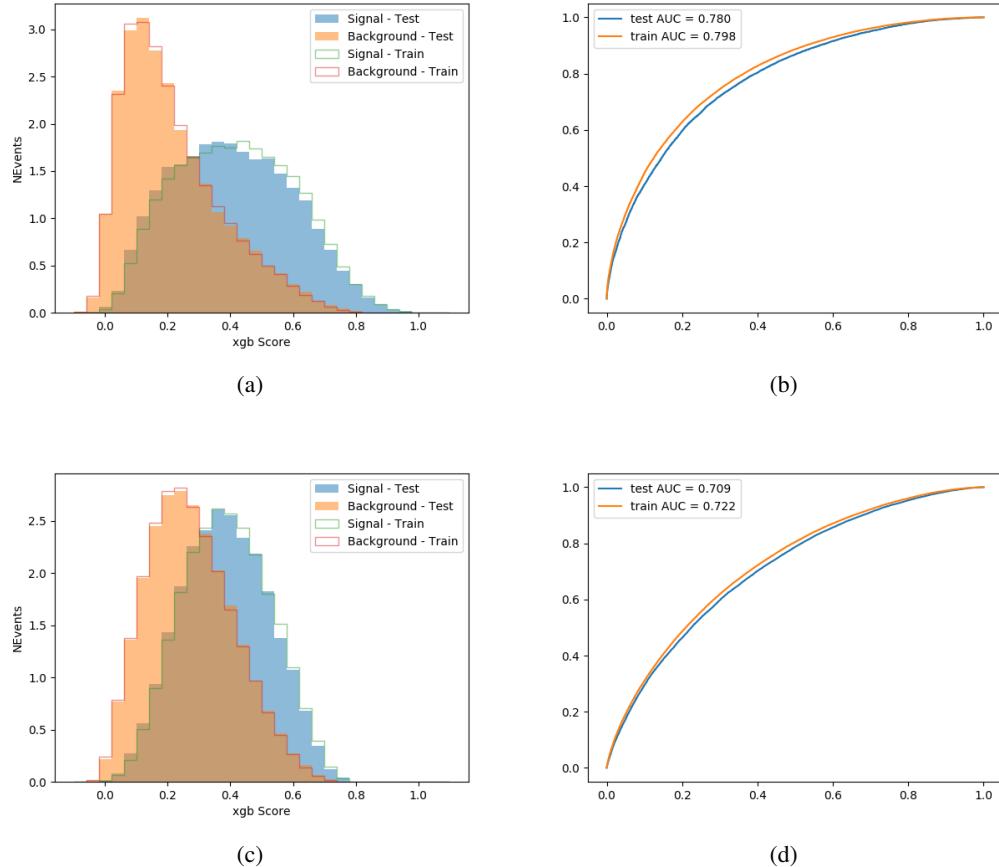


Figure 10.4:

764 **10.3 Signal Region Definitions**

765 Once pre-selection has been applied, channels are further refined based on the MVAs described
 766 above. The output of the model described in section 8.5 is used to separate the three channel
 767 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs
 768 boson.

769 For each event, depending on the channel as well as the predicted p_T of the Higgs derived

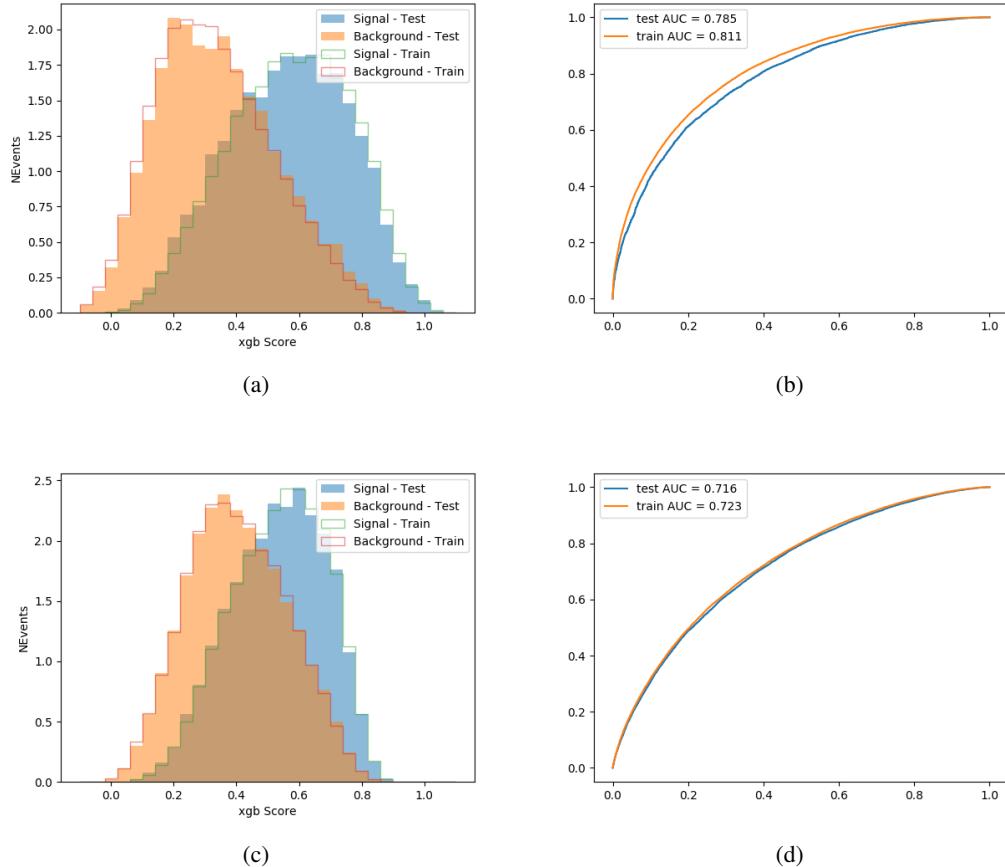


Figure 10.5:

from the algorithm described in section 8.4, a cut on the appropriate background rejection algorithm is applied. The specific selection used, and the event yield in each channel after this selection has been applied, is summarized below.

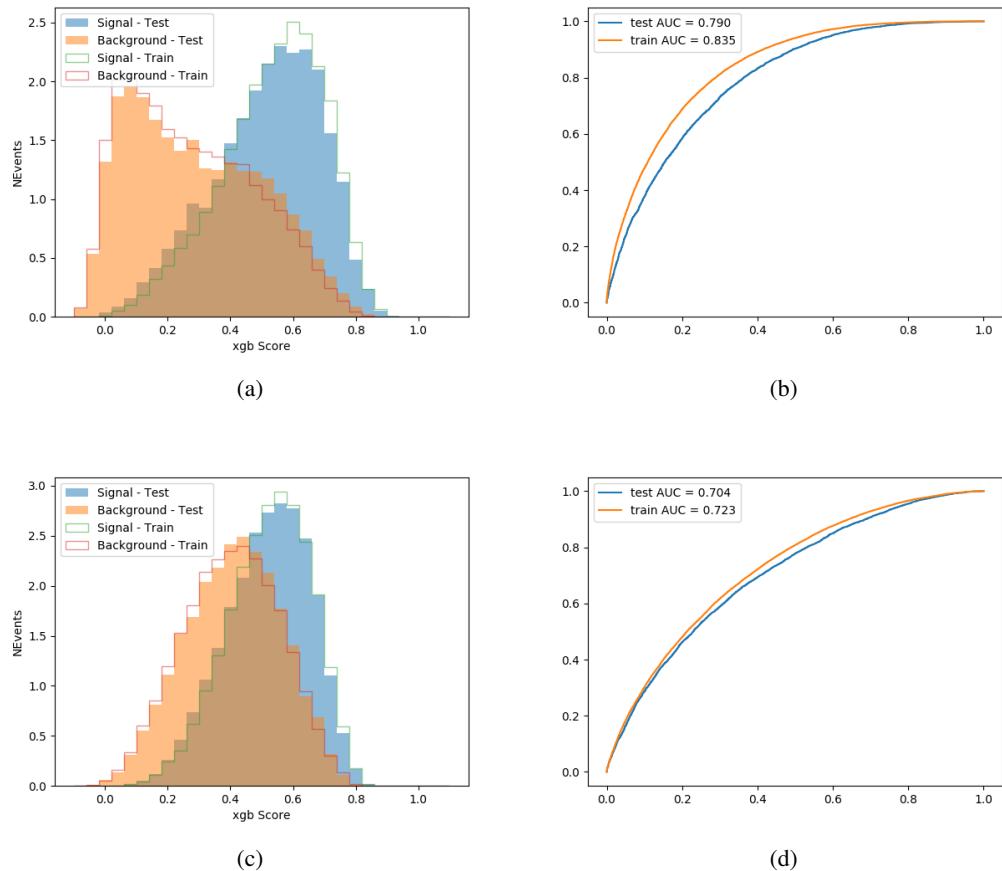


Figure 10.6:

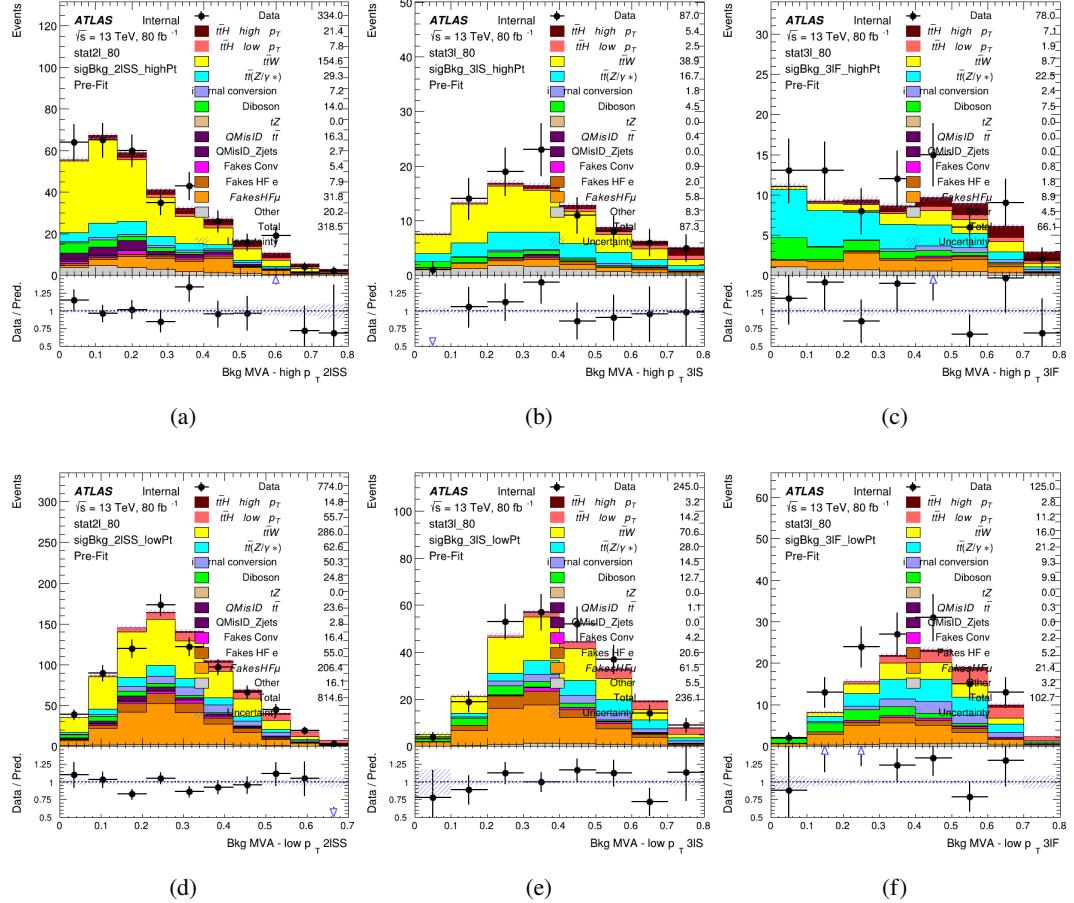


Figure 10.7: scores

773 **10.3.1 2ISS**774 **10.3.2 3l – Semi – leptonic**775 **10.3.3 3l – Fully – leptonic**776 **11 Systematic Uncertainties**

777 The systematic uncertainties that are considered are summarized in table 16. These are implemented in the fit either as a normalization factors or as a shape variation or both in the signal
 778 11th November 2020 – 14:26

⁷⁷⁹ and background estimations. The numerical impact of each of these uncertainties is outlined in
⁷⁸⁰ section 12.

Table 20: 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

⁷⁸¹ The uncertainty in the combined 2015+2016 integrated luminosity is derived from a
⁷⁸² calibration of the luminosity scale using x-y beam-separation scans performed in August 2015
⁷⁸³ and May 2016 [**lumi**].

⁷⁸⁴ The experimental uncertainties are related to the reconstruction and identification of light
⁷⁸⁵ leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The sources which contribute

786 to the uncertainty in the jet energy scale [**jes**] are decomposed into uncorrelated components and
787 treated as independent sources in the analysis.

788 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
789 [**btag_cal**] are also decomposed into uncorrelated components. The large number of components
790 for b-tagging is due to the calibration of the distribution of the BDT discriminant.

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

793 **12 Results**

794 Unblinded results are shown for the 80 fb^{-1} data set, as well as MC only projections of results
795 using the full Run-2, 140 fb^{-1} dataset.

796 **12.1 Results - 80 fb^{-1}**

797 A maximum likelihood fit is performed simultaneously over the regions shown in figure 12.1.

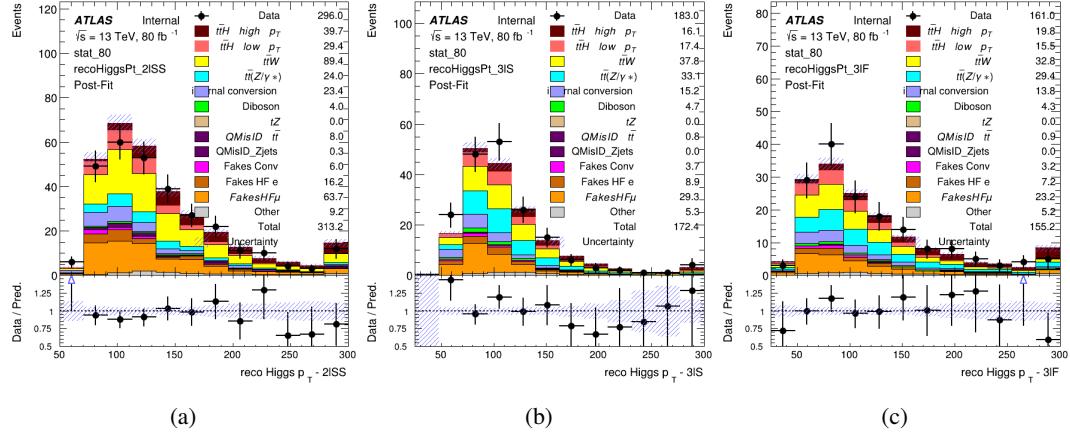


Figure 12.1:

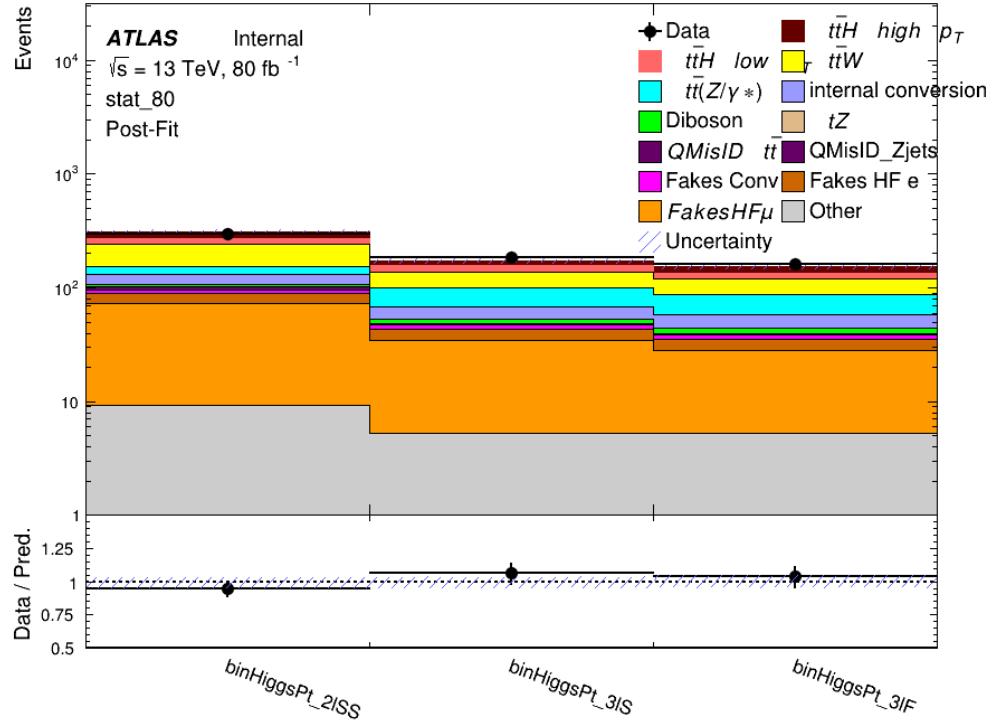


Figure 12.2: Post-fit summary of fit.

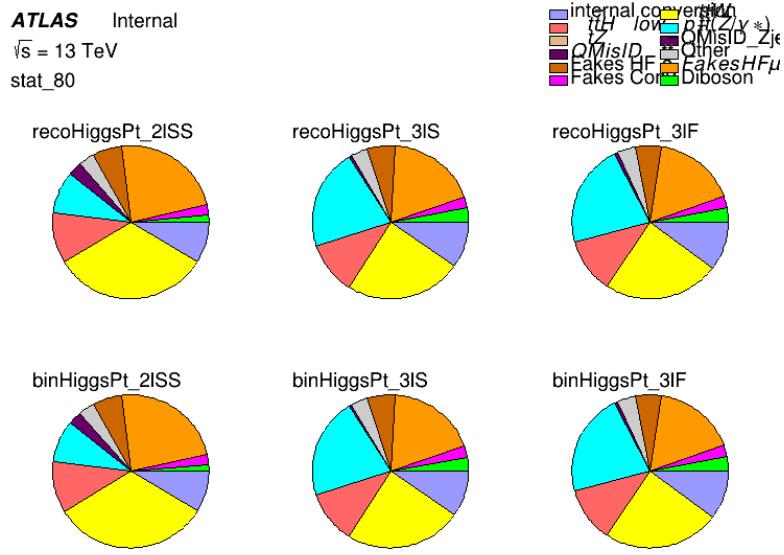


Figure 12.3: Background composition of the fit regions.

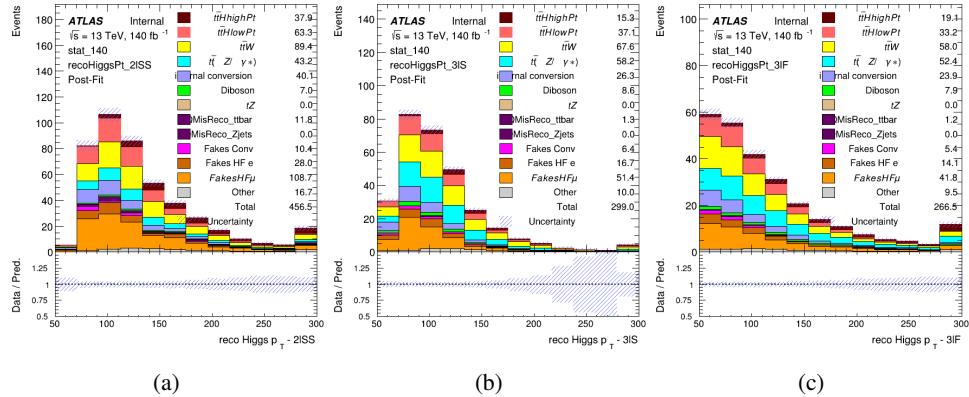


Figure 12.4:

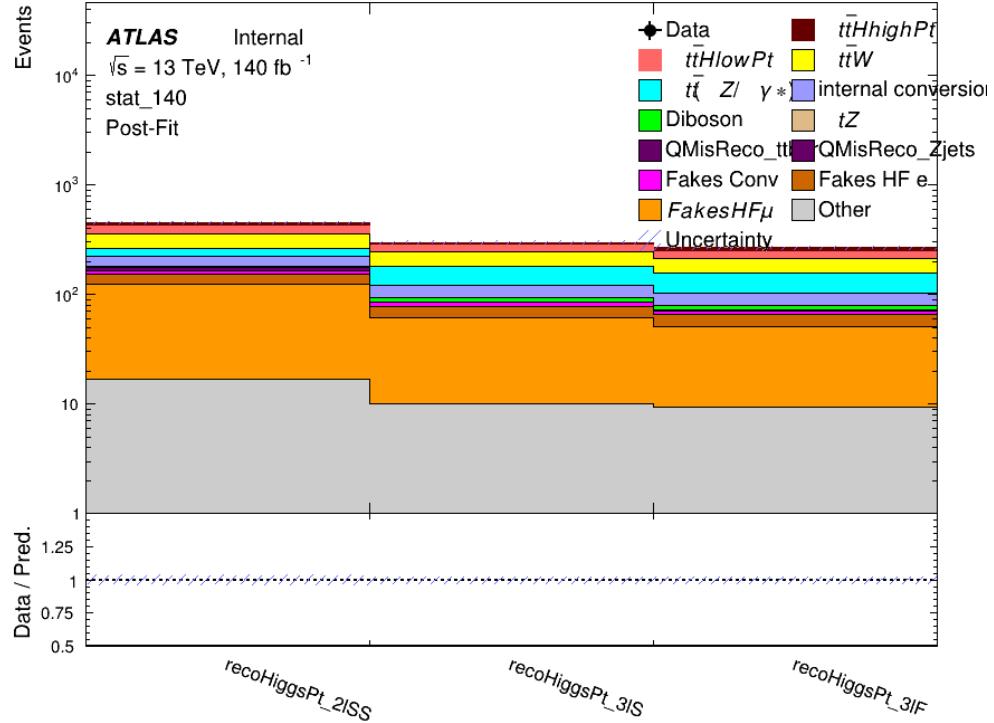


Figure 12.5: Post-fit summary of fit.

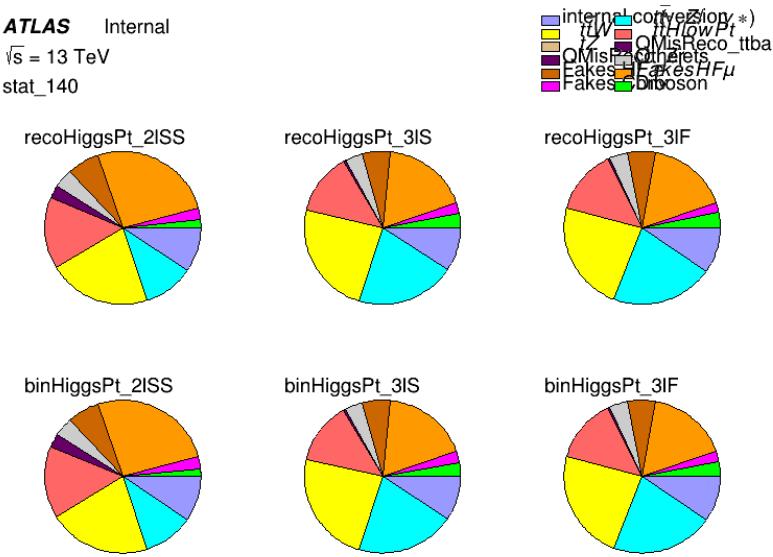


Figure 12.6: Background composition of the fit regions.

798 **12.2 Projected Results - 140 fb^{-1}**

799 **Part V**

800 **Conclusion**

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

805 **List of contributions**

806

807 **Appendices**

808 **A Machine Learning Models**

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

811 **A.1 Higgs Reconstruction Models**

812 **A.1.1 b-jet Identification Features - 2lSS**

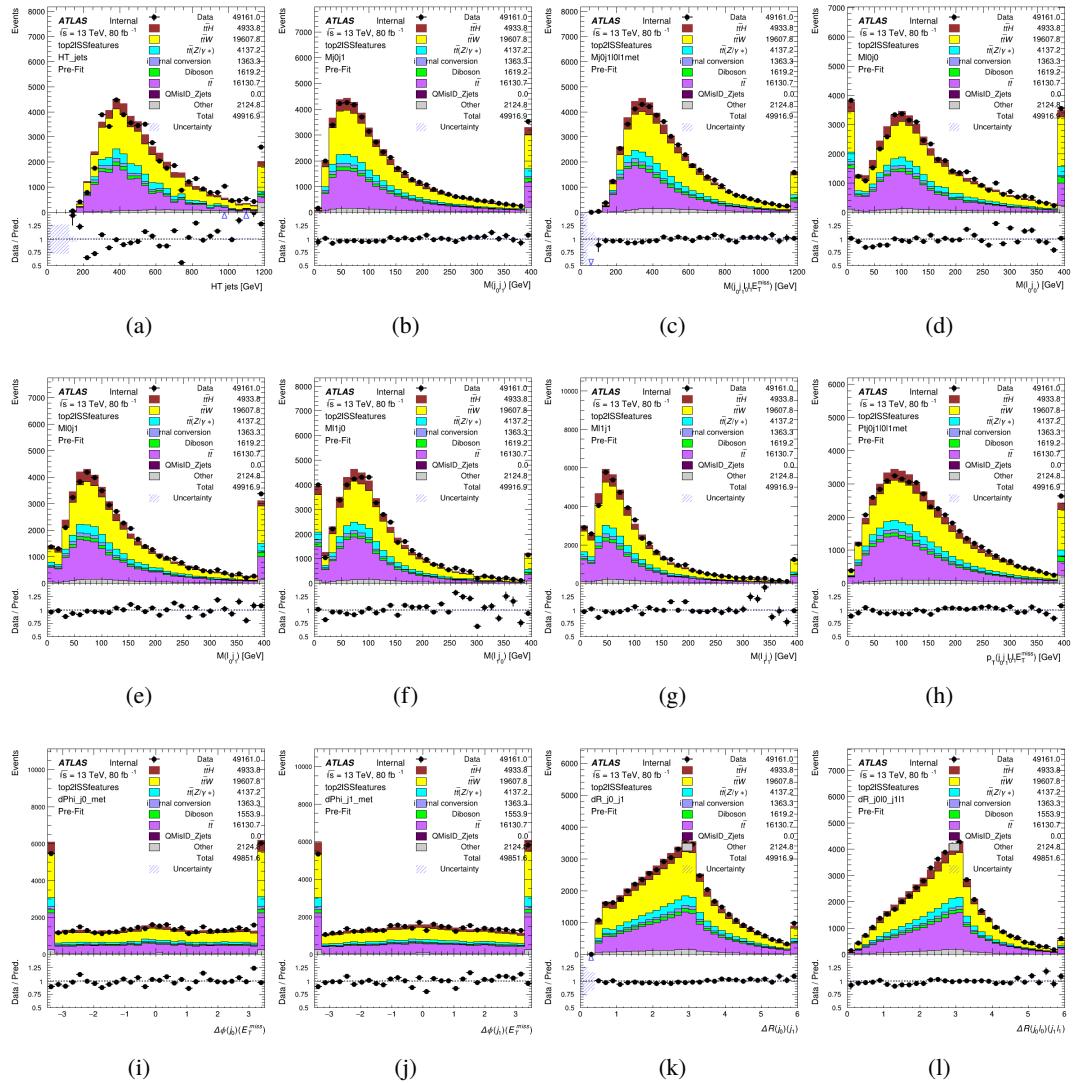


Figure A.1: Input features for top2ISS

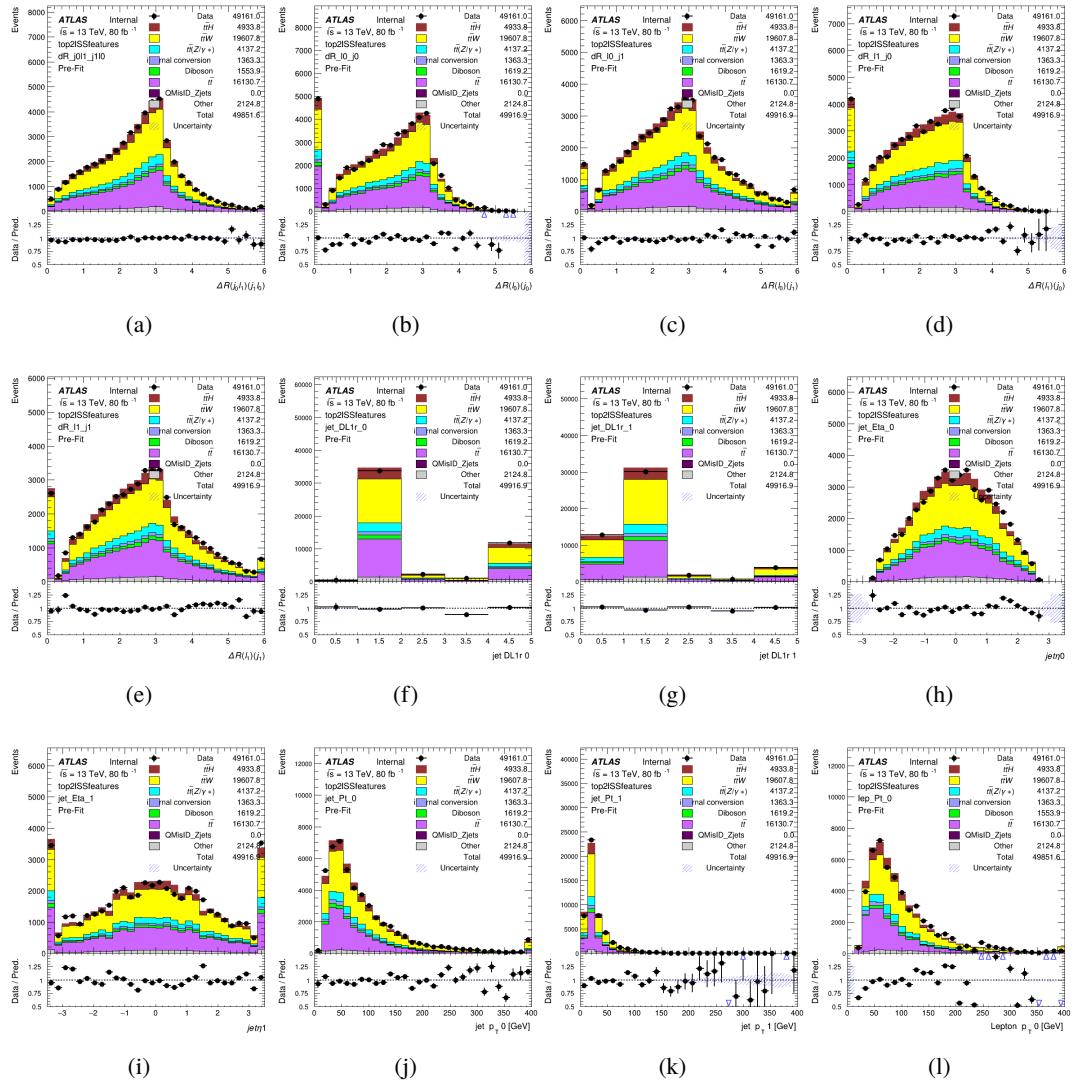


Figure A.2: Input features for top2lSS

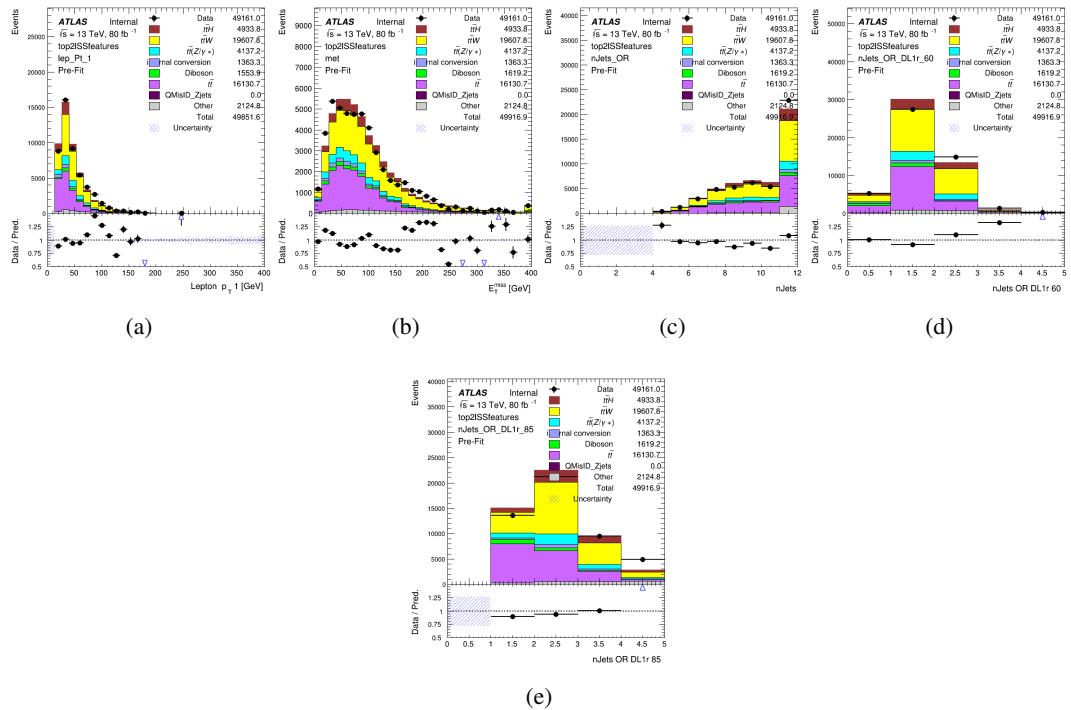


Figure A.3: Input features for top21SS

813 **A.1.2 b-jet Identification Features - 3l**

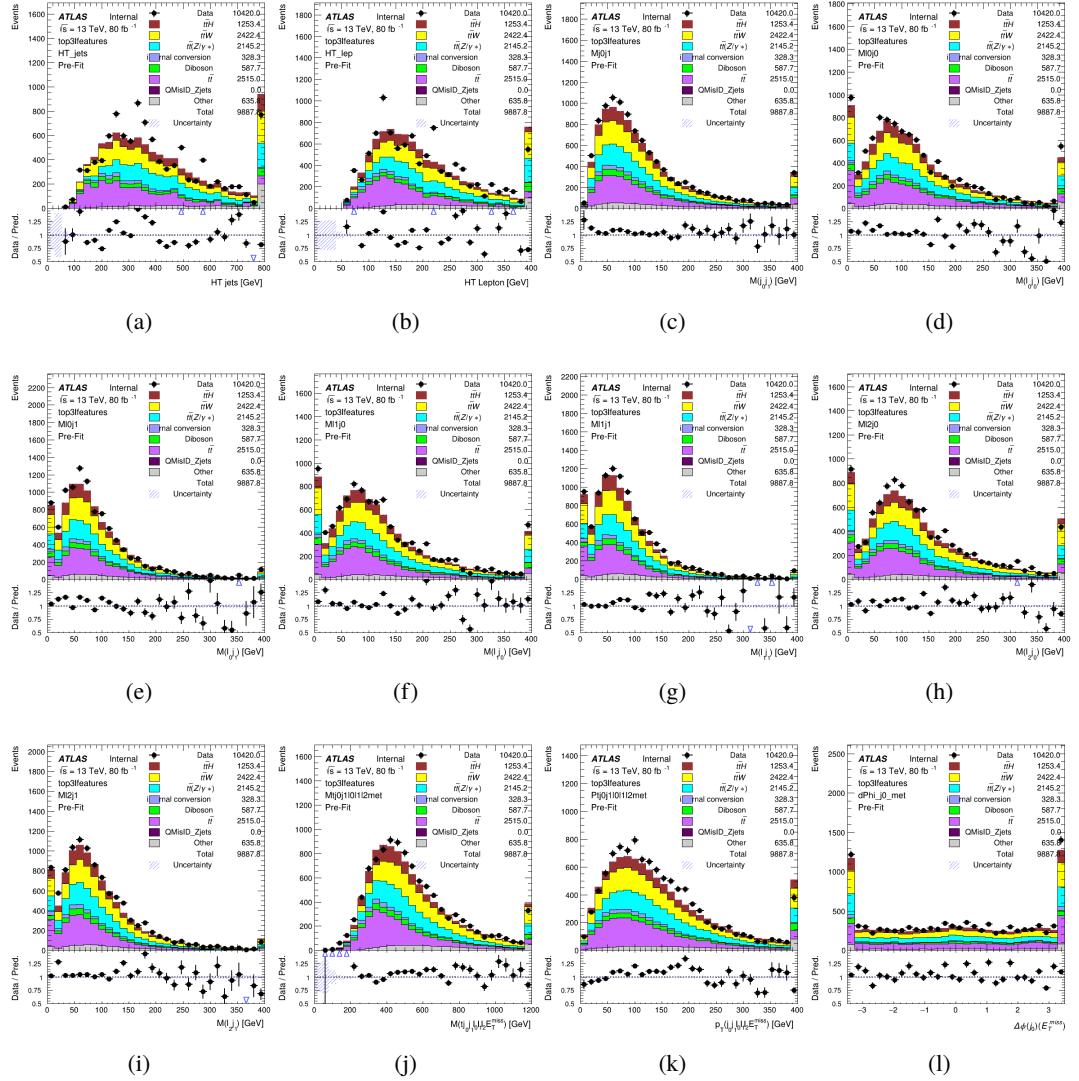


Figure A.4: Input features for top31

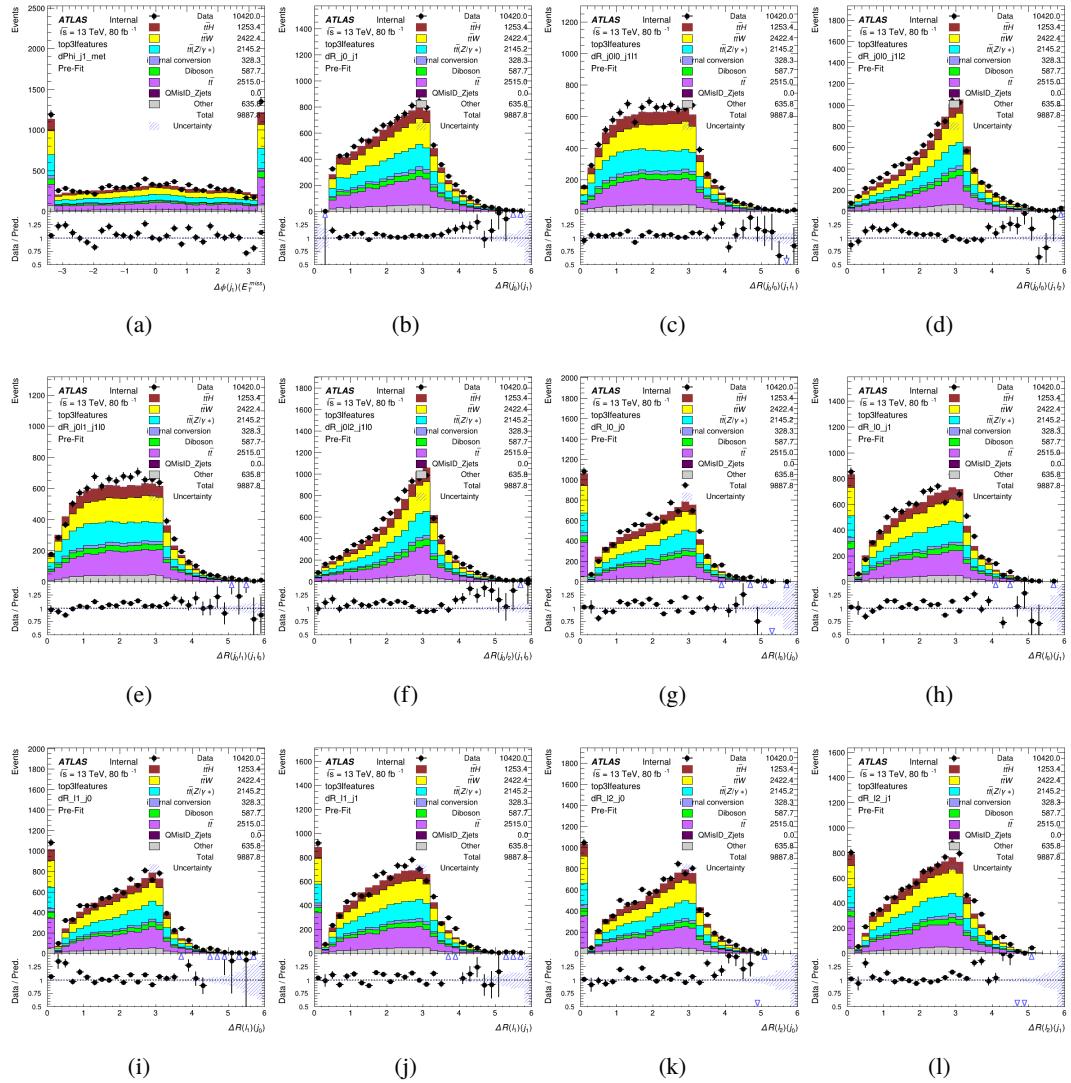


Figure A.5: Input features for top31

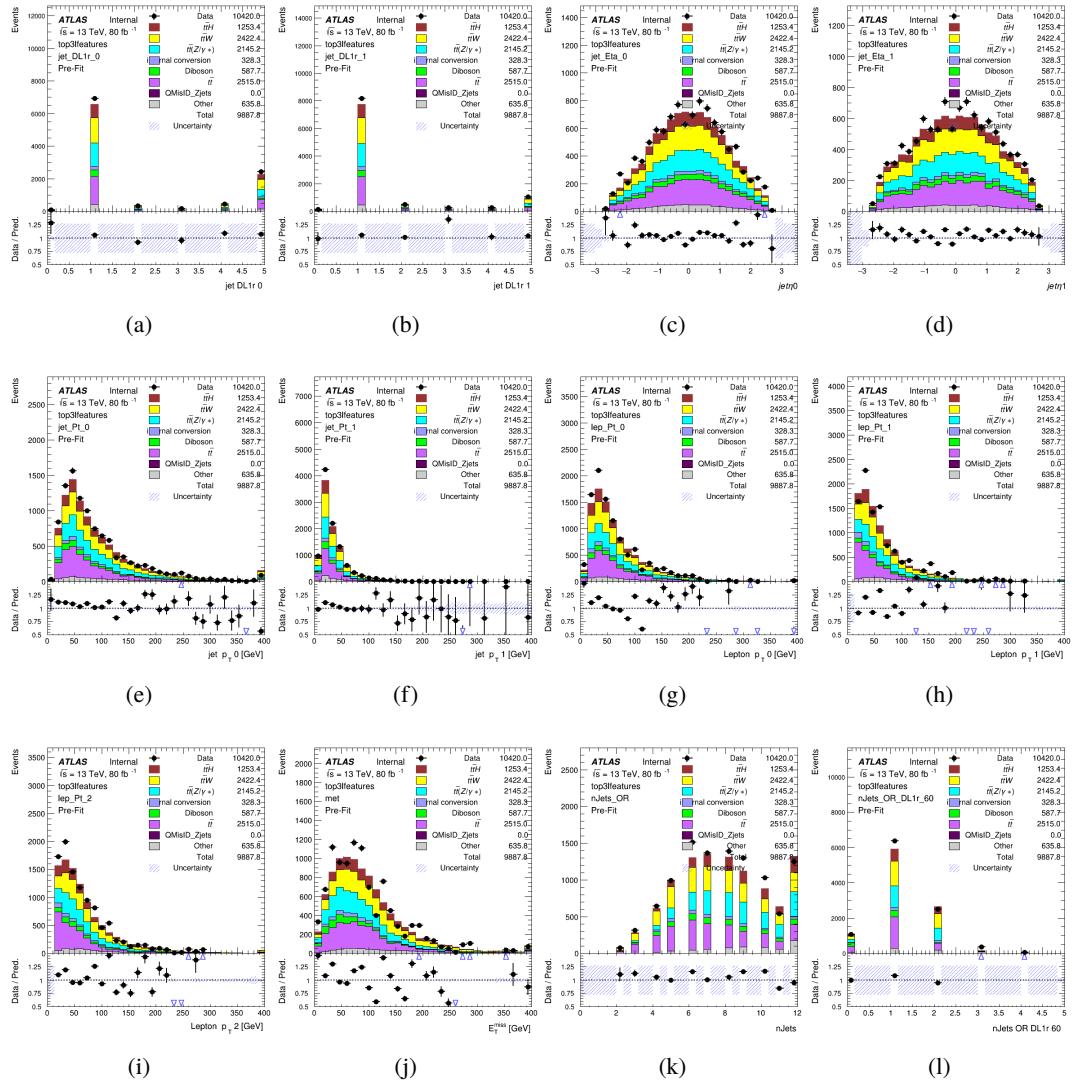
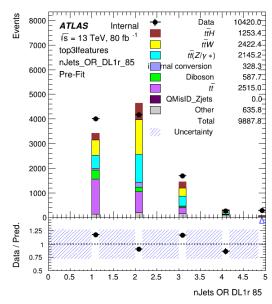


Figure A.6: Input features for top31



(a)

Figure A.7: Input features for top3l

814 **A.1.3 Higgs Reconstruction Features - 2lSS**

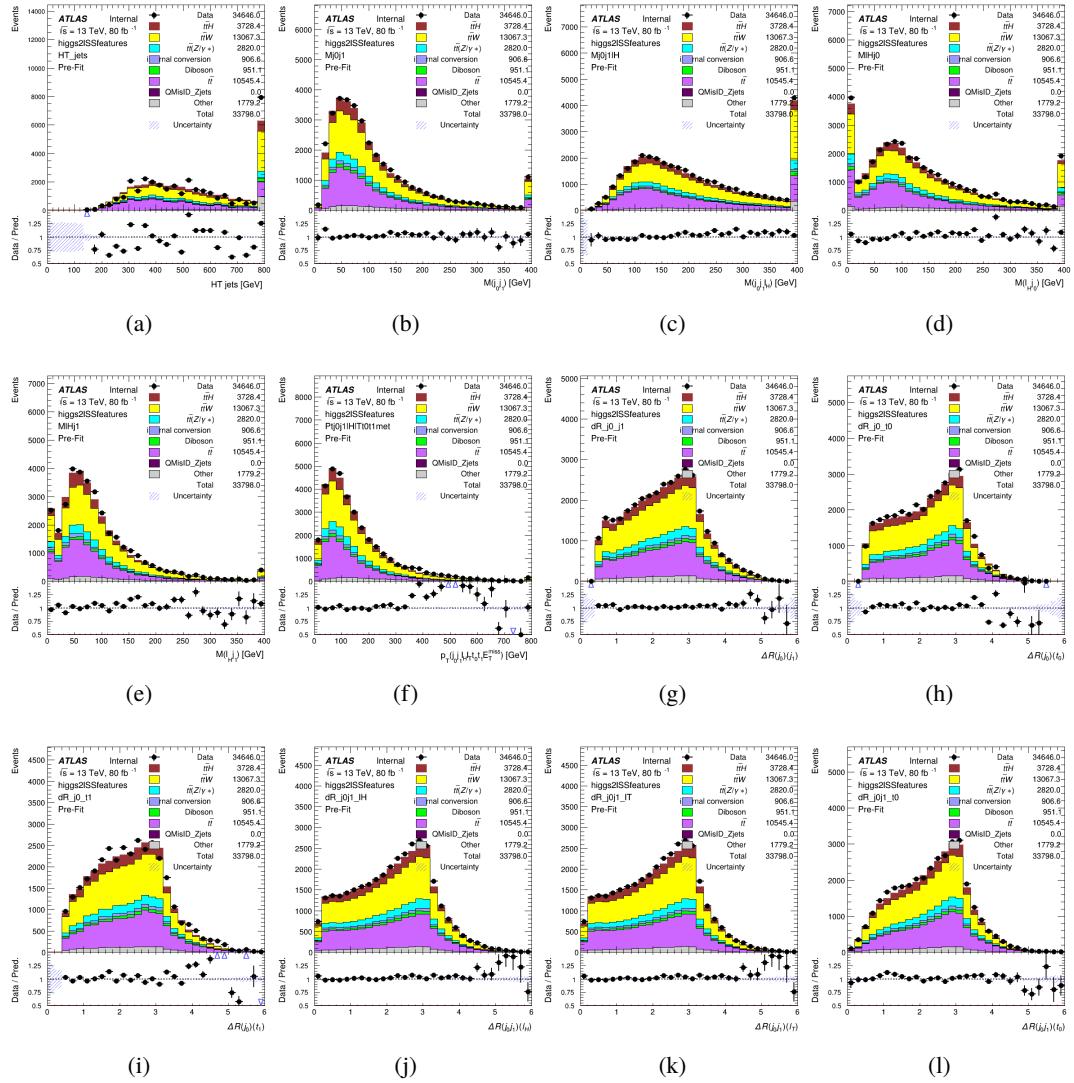


Figure A.8: Input features for higgs2lSS

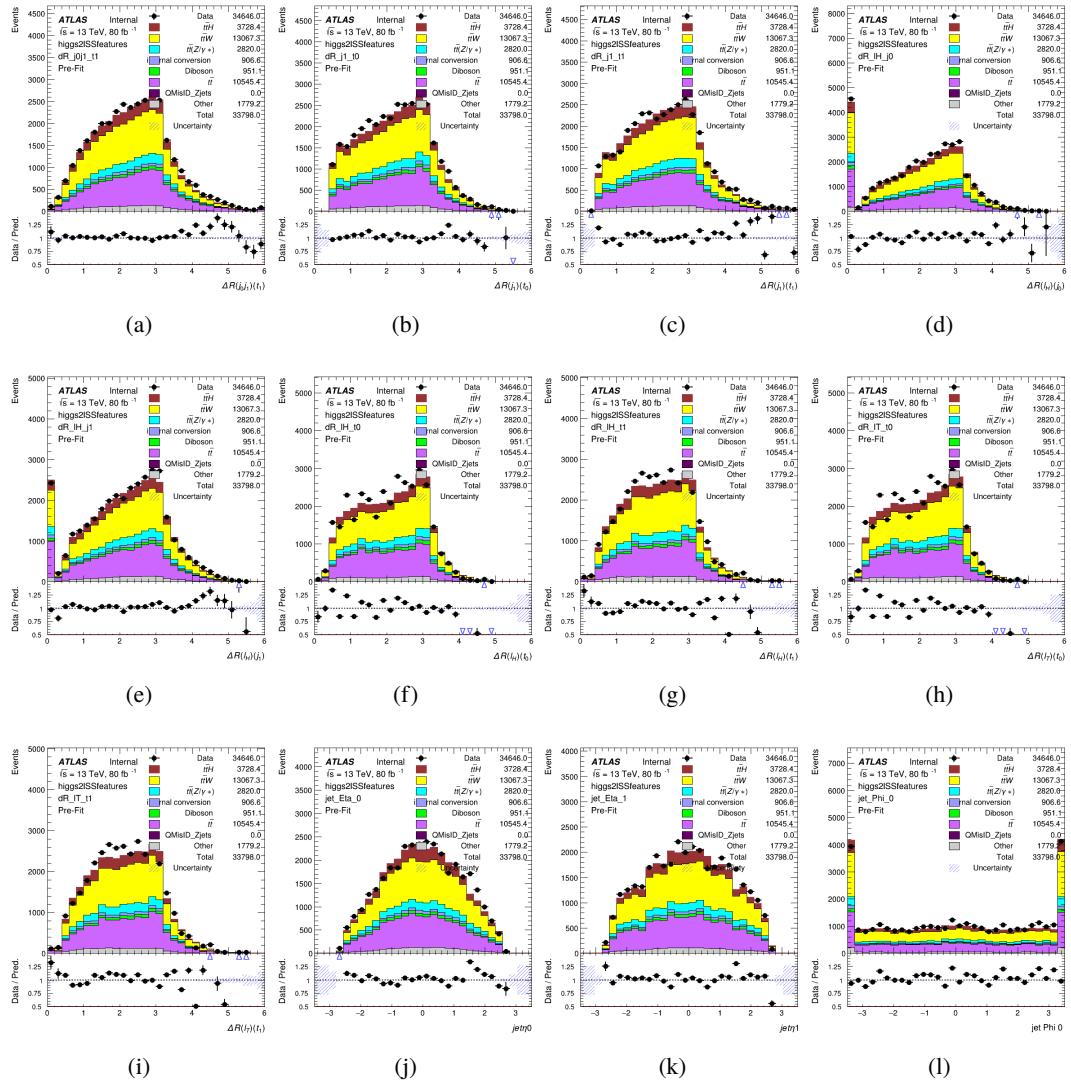


Figure A.9: Input features for higgs2lSS

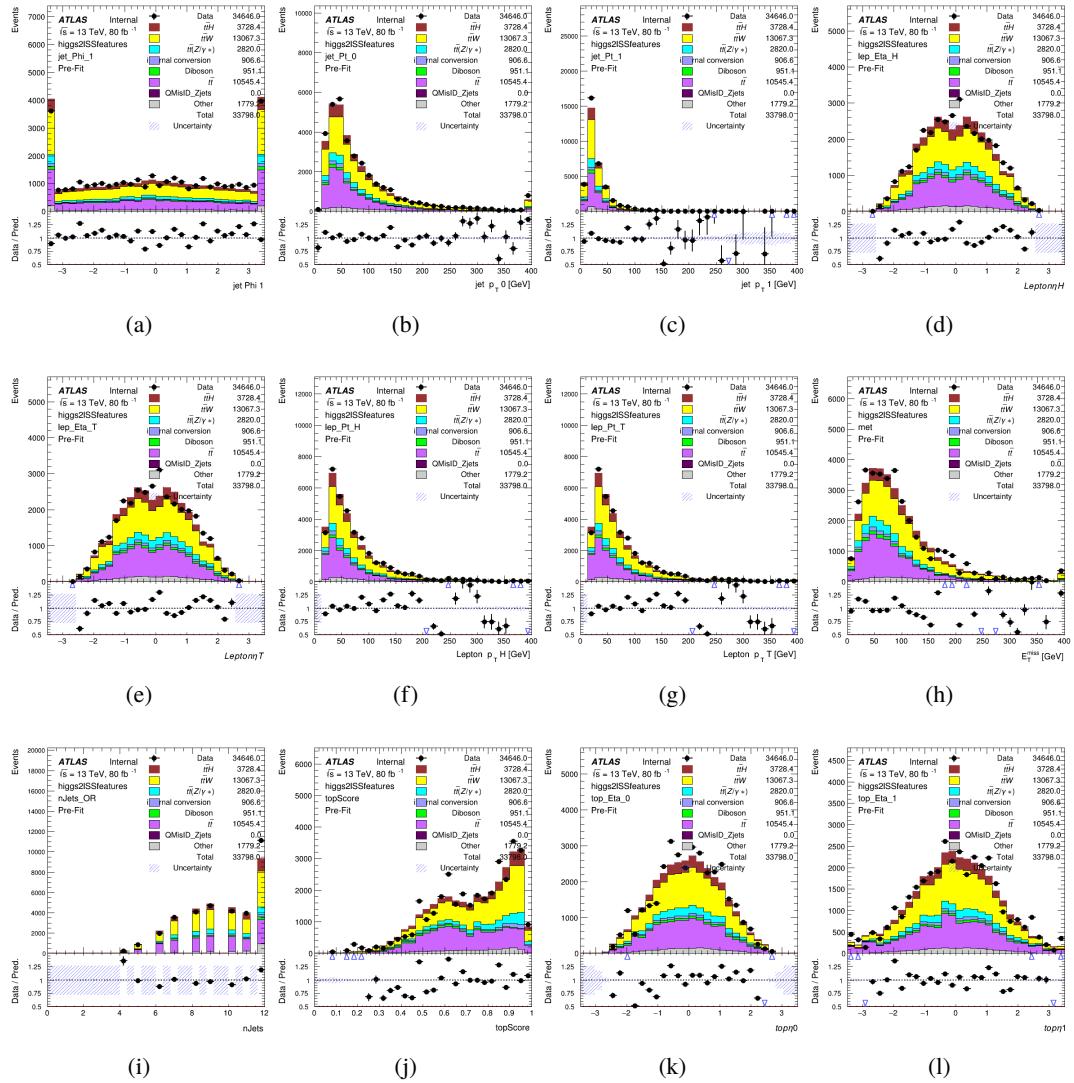


Figure A.10: Input features for higgs2ISS

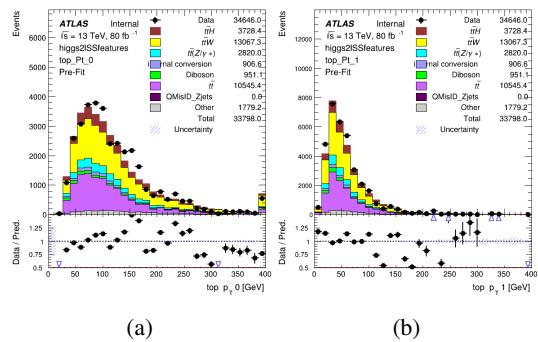


Figure A.11: Input features for higgs2ISS

815 **A.1.4 Higgs Reconstruction Features - 3lS**

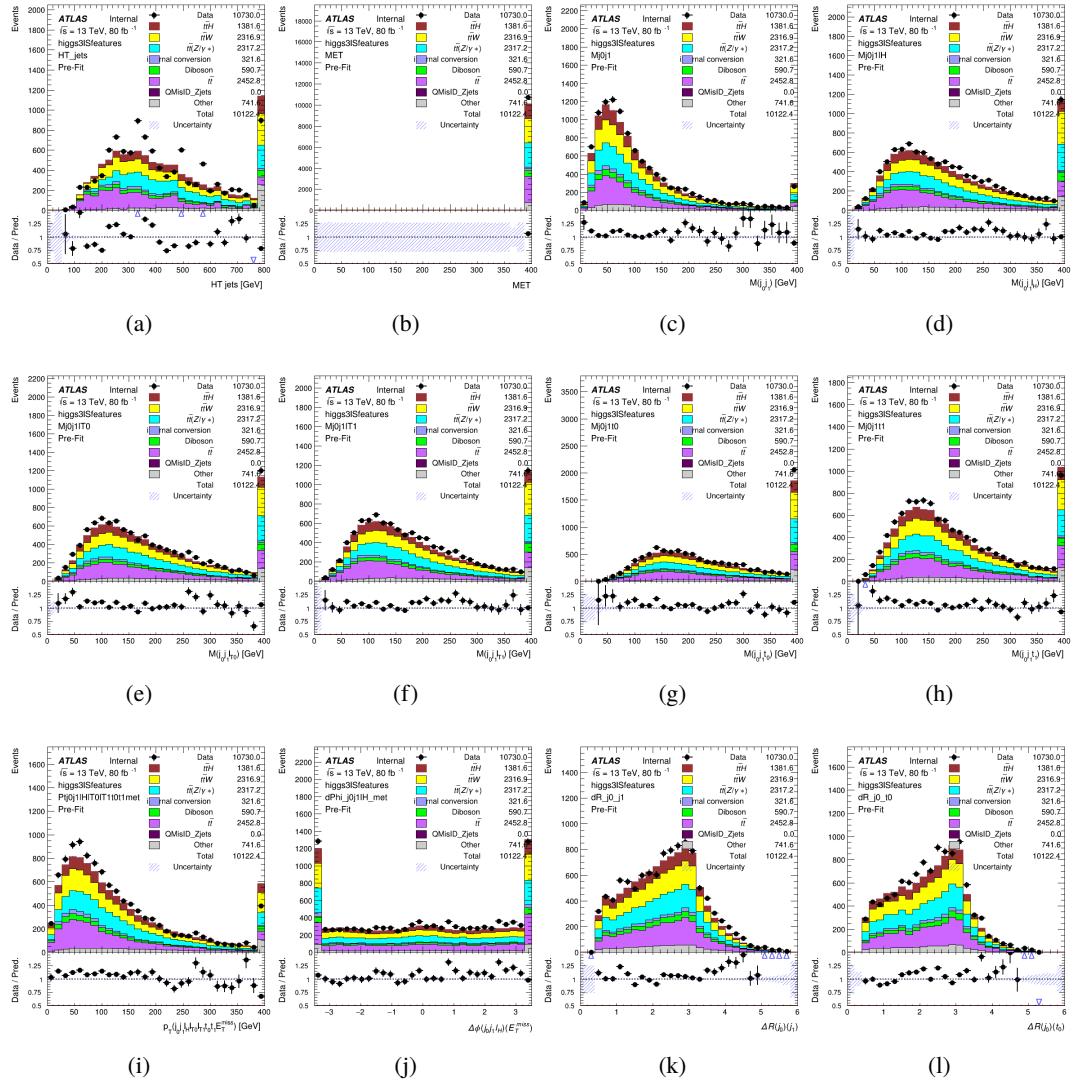


Figure A.12: Input features for higgs3IS

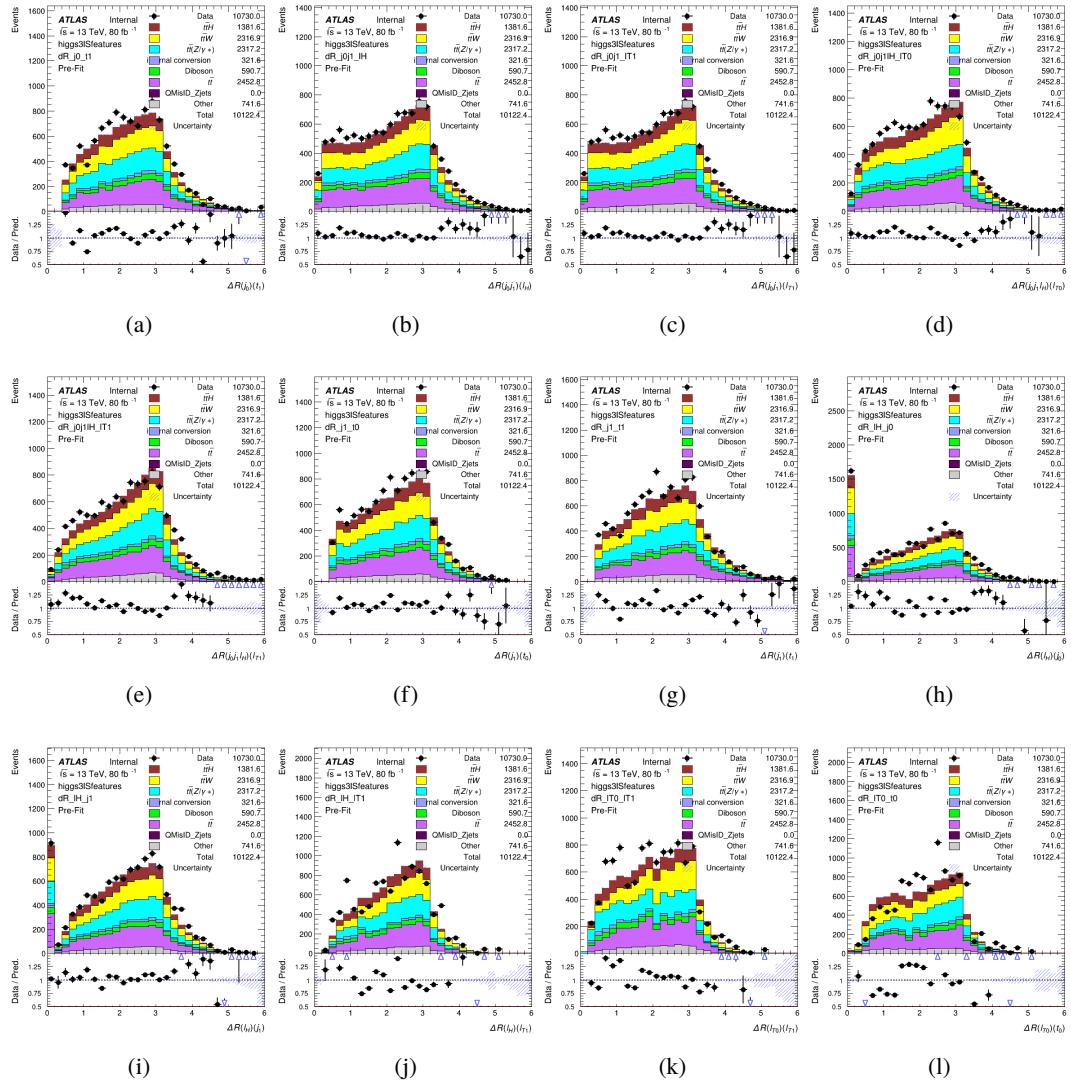


Figure A.13: Input features for higgs31S

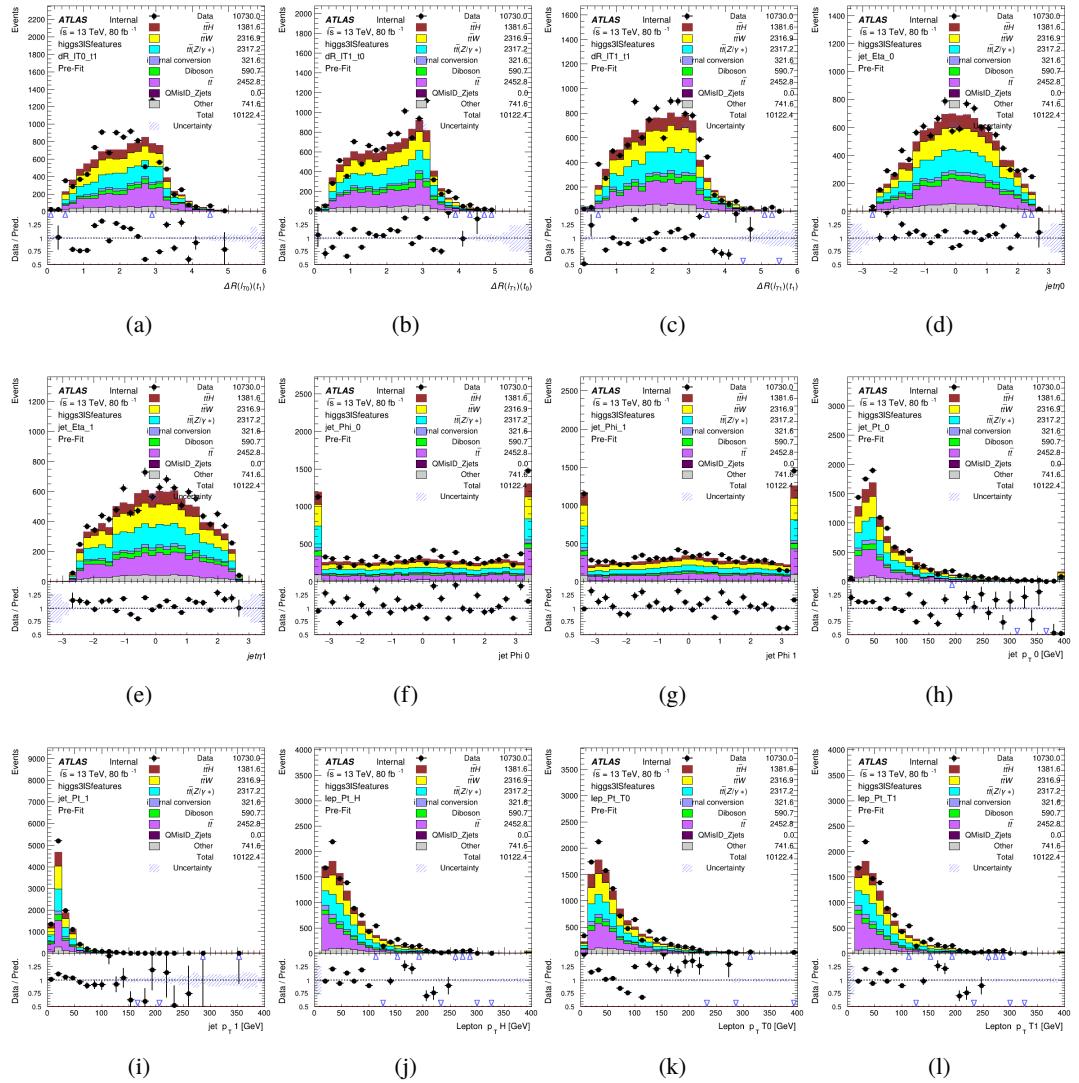


Figure A.14: Input features for higgs3lS

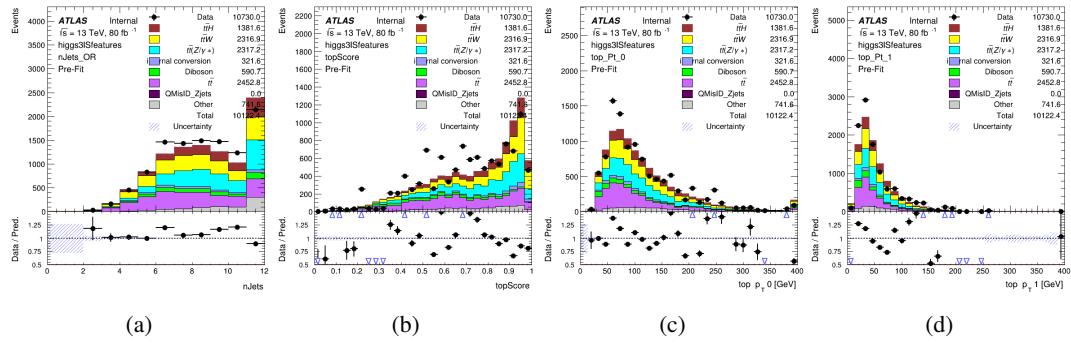


Figure A.15: Input features for higgs3lS

816 **A.1.5 Higgs Reconstruction Features - 3lF**

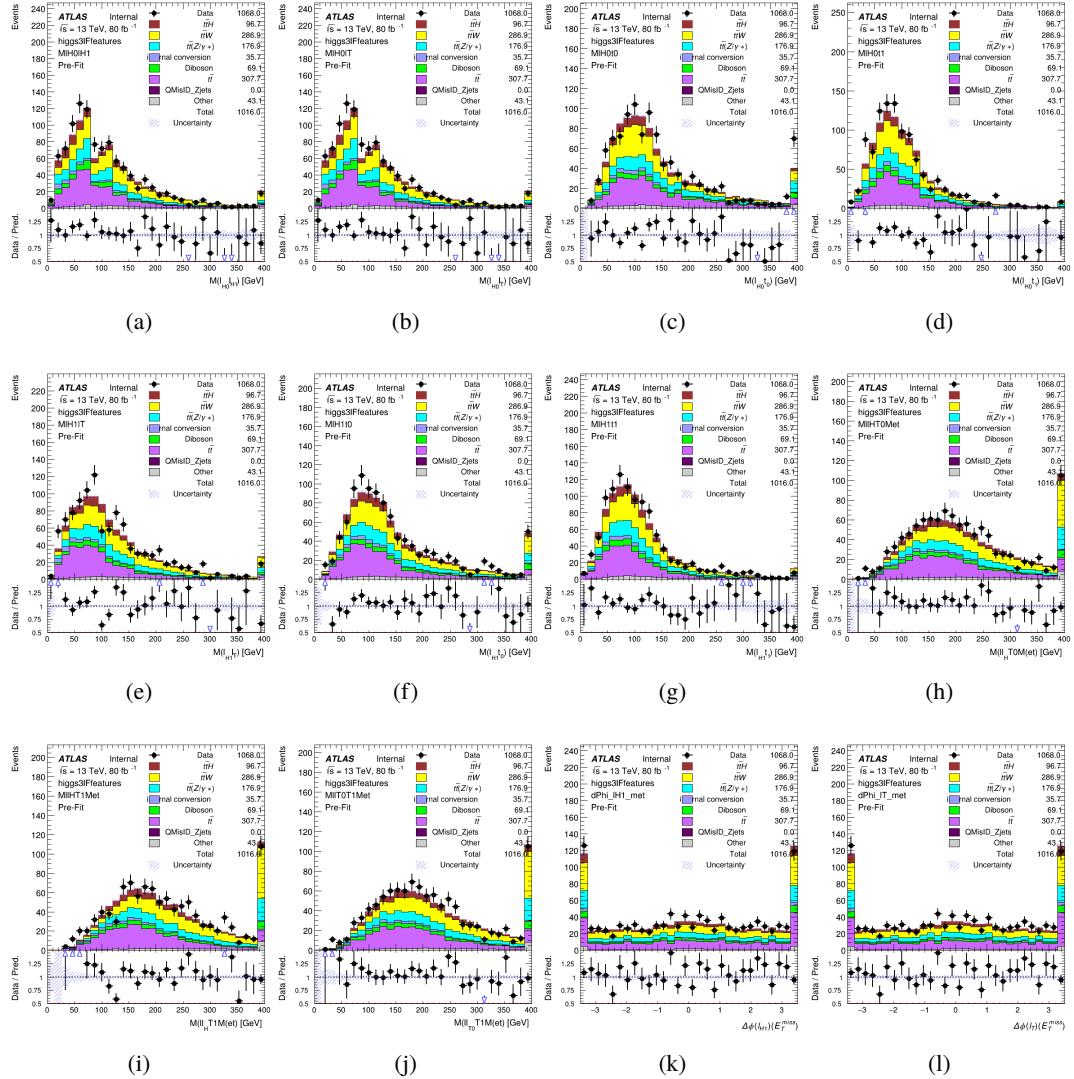


Figure A.16: Input features for higgs3IF

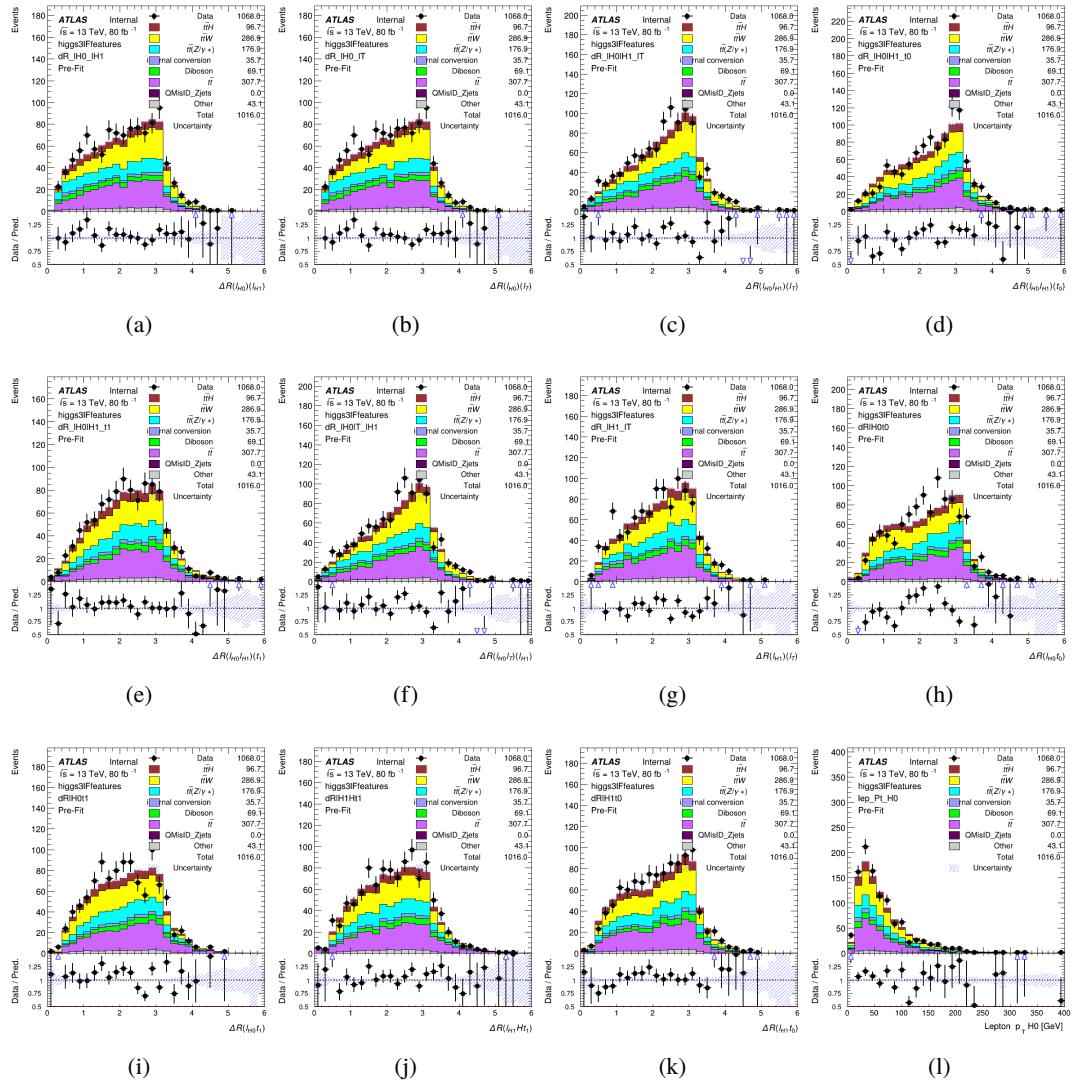


Figure A.17: Input features for higgs3lF

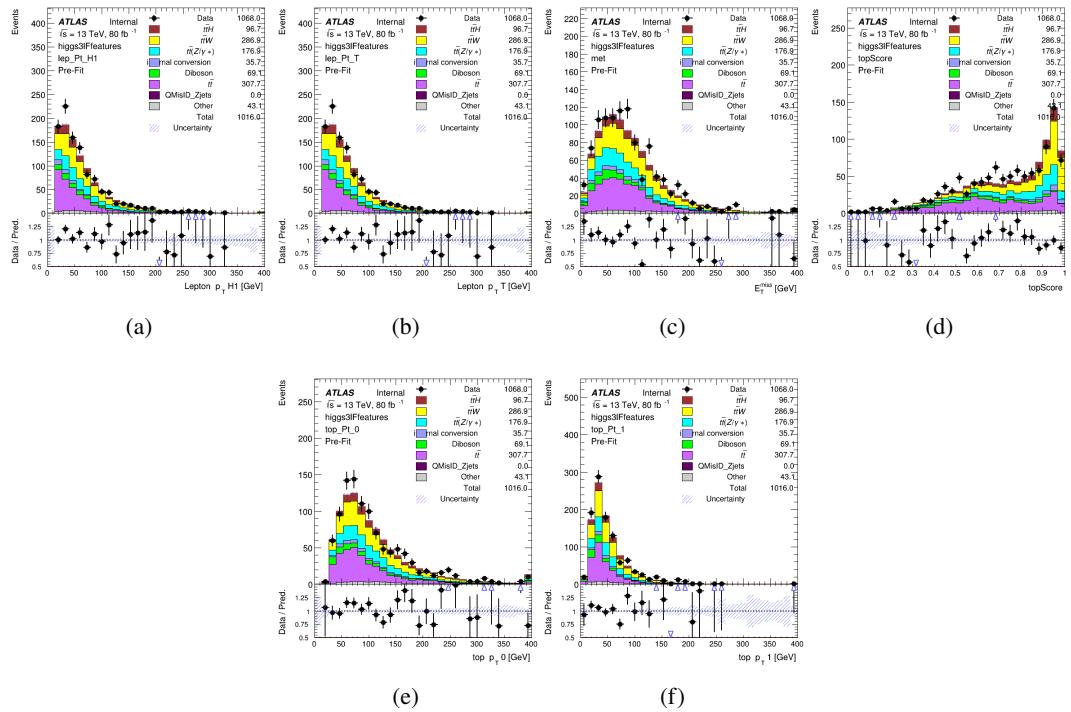


Figure A.18: Input features for higgs3IF

⁸¹⁷ **A.2 Background Rejection MVAs**

⁸¹⁸ **A.2.1 Background Rejection MVA Features - 2lSS**

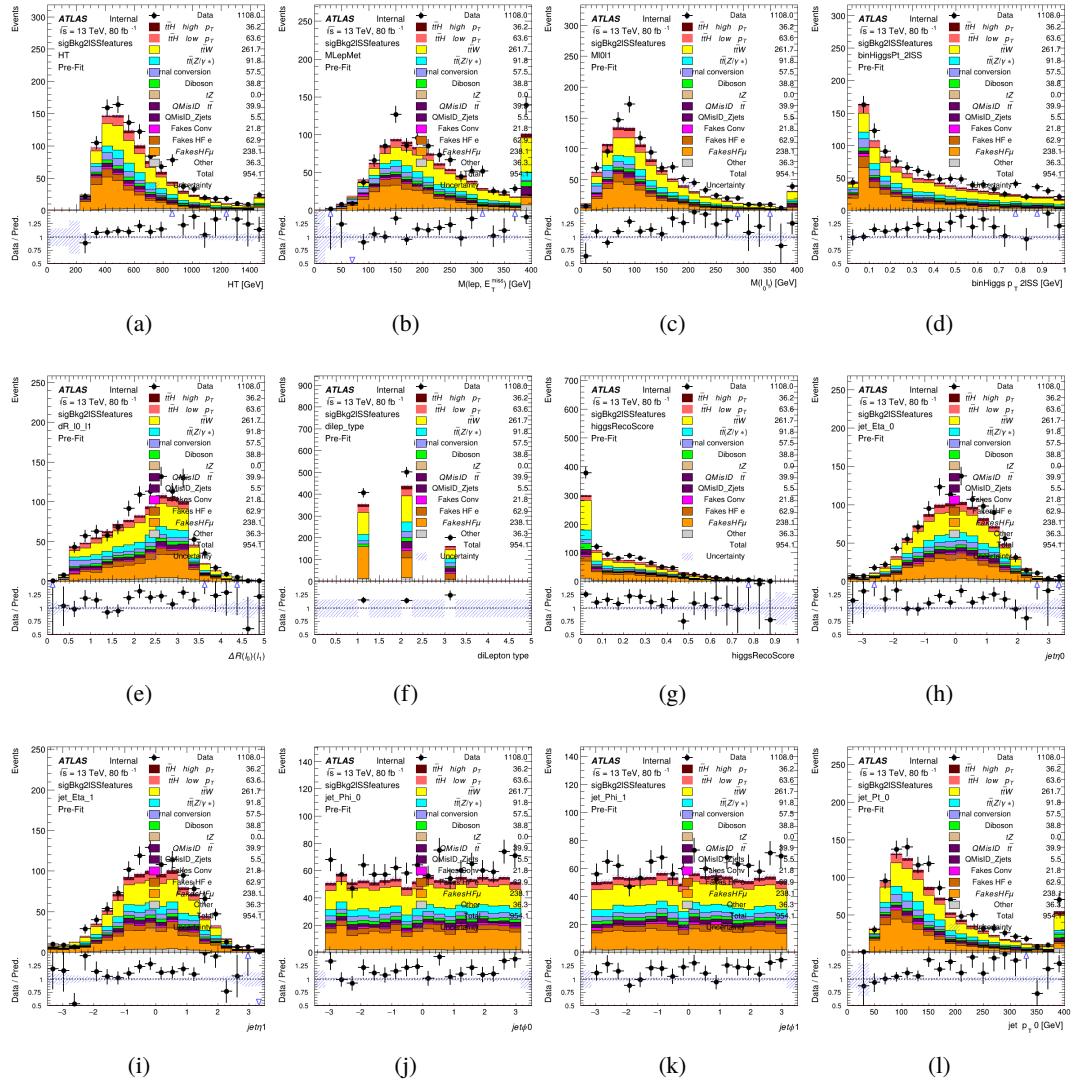


Figure A.19: Input features for sigBkg2lSS

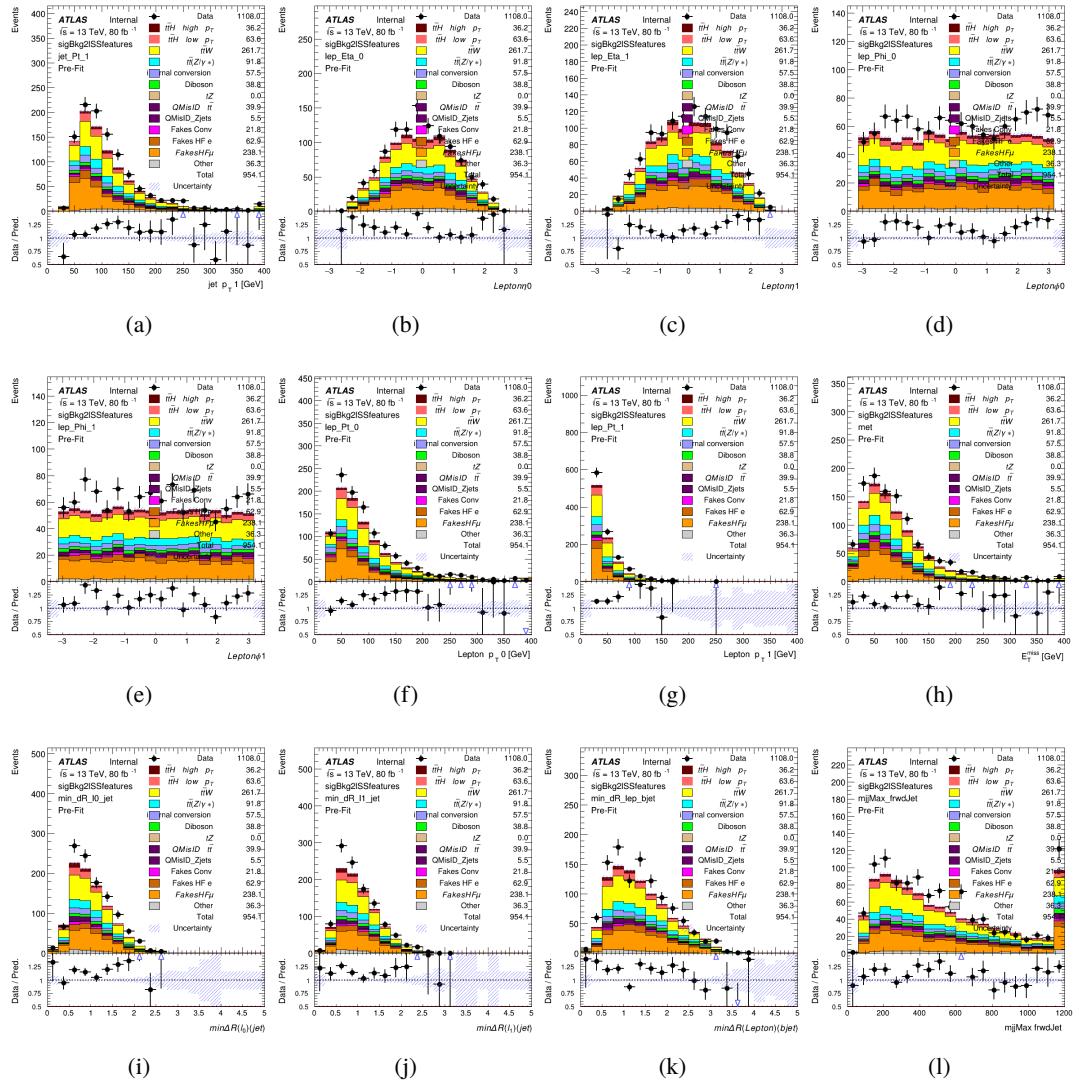


Figure A.20: Input features for sigBkg2ISS

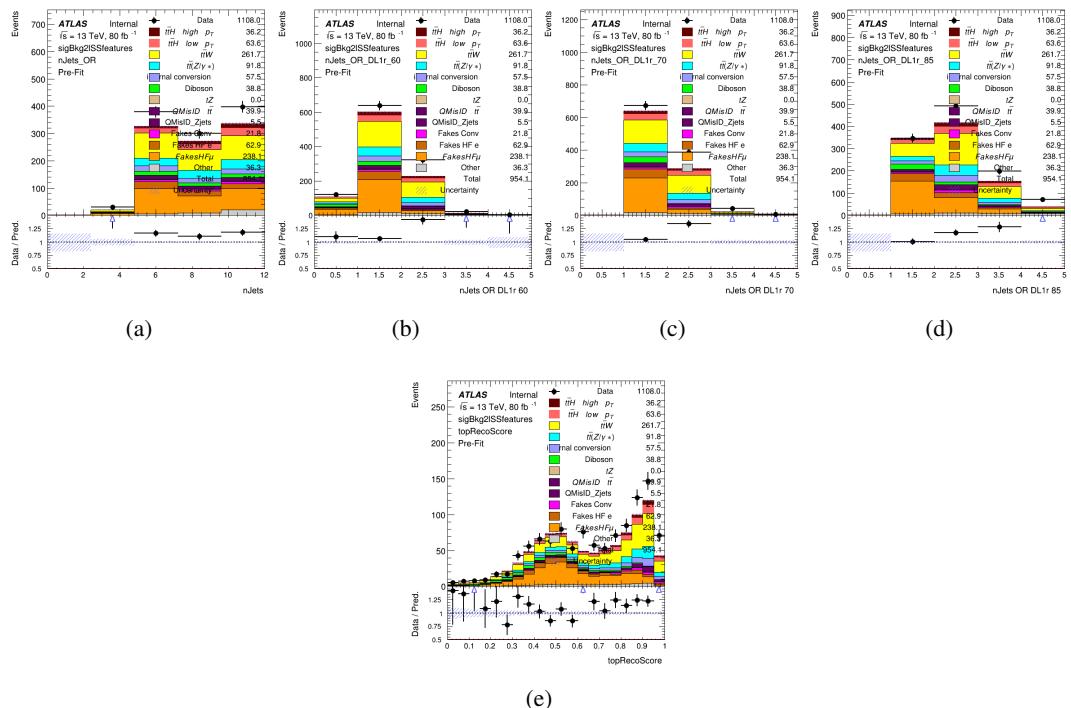
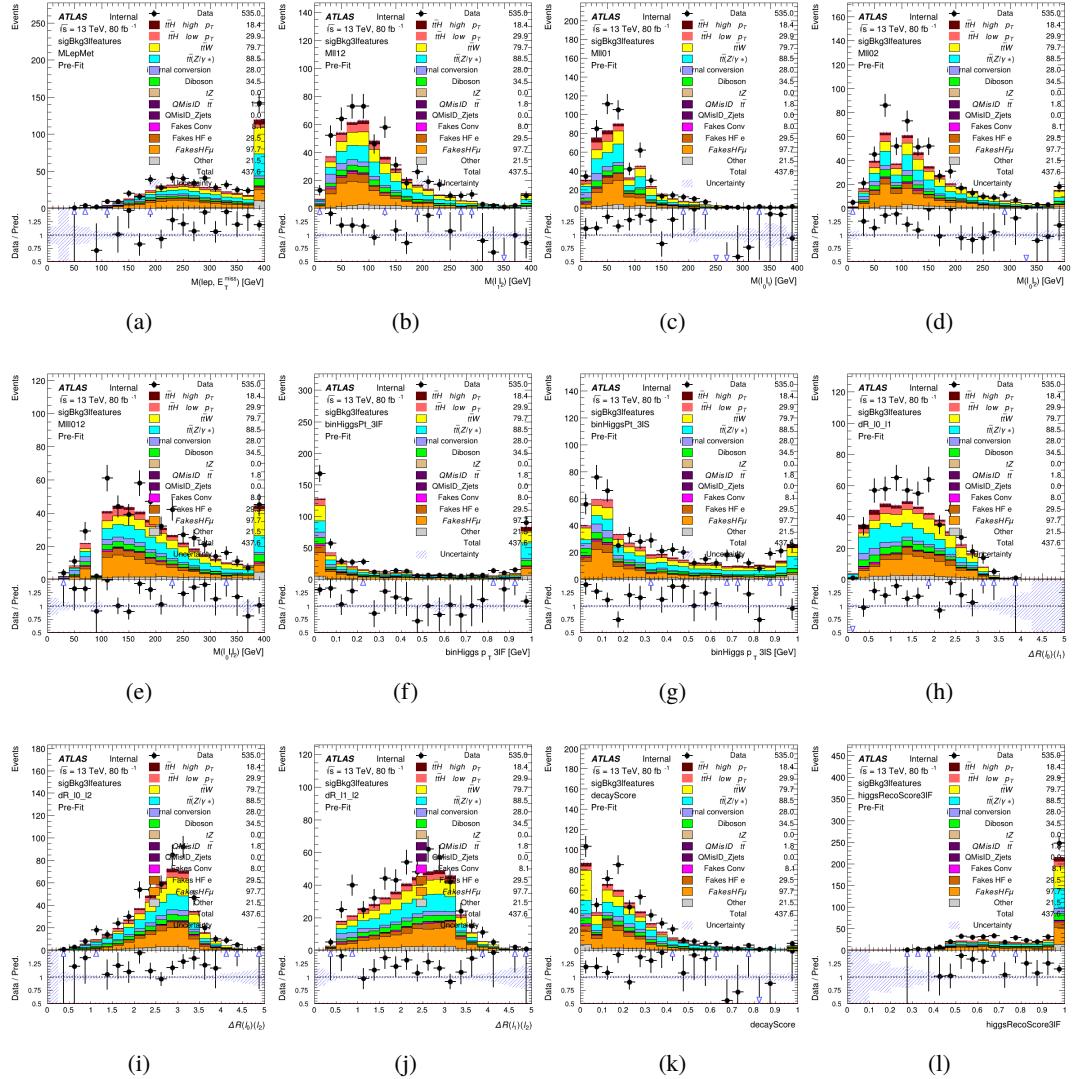


Figure A.21: Input features for sigBkg2ISS

819 **A.2.2 Background Rejection MVA Features - 3l**

Figure A.22: Input features for `sigBkg3l`

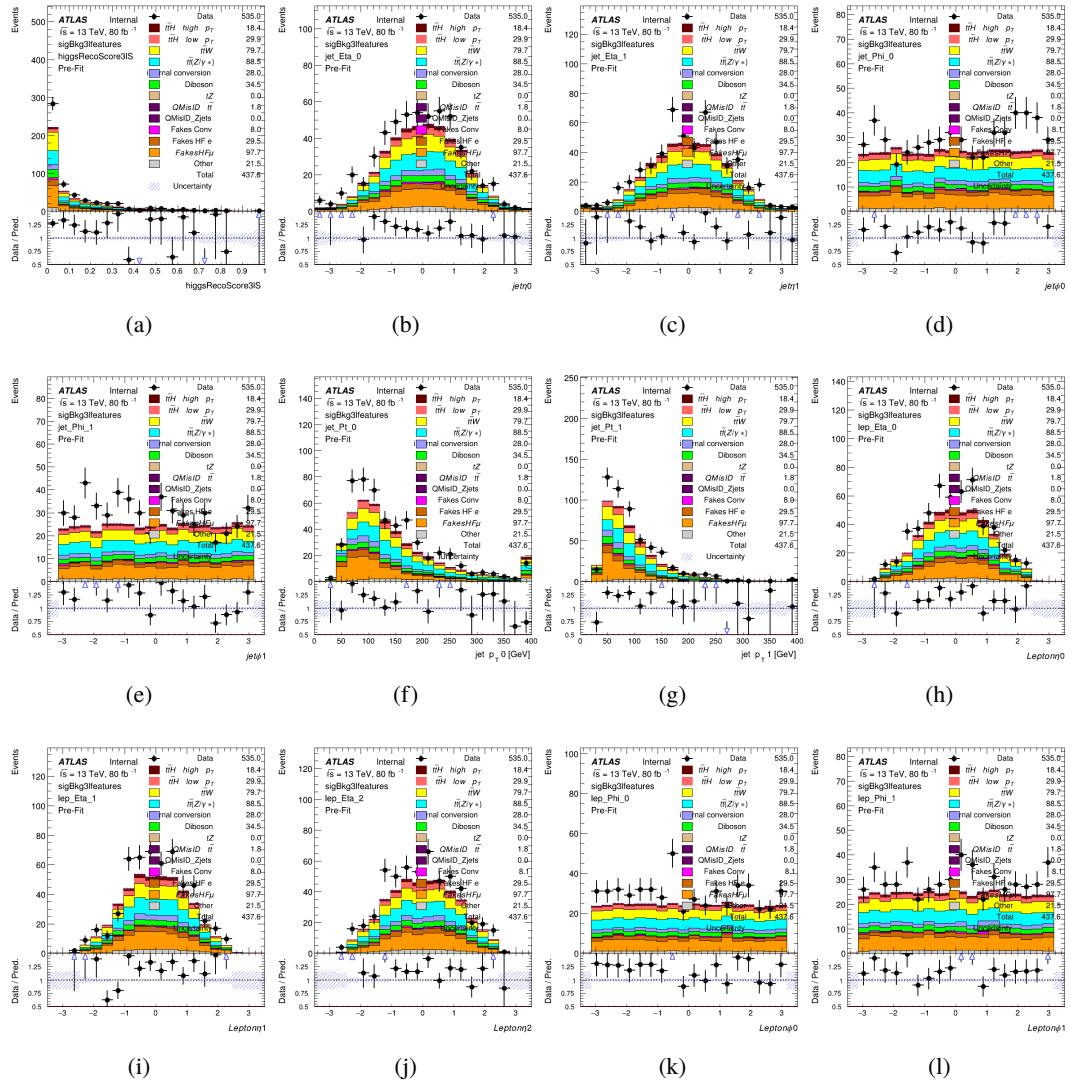


Figure A.23: Input features for sigBkg3l

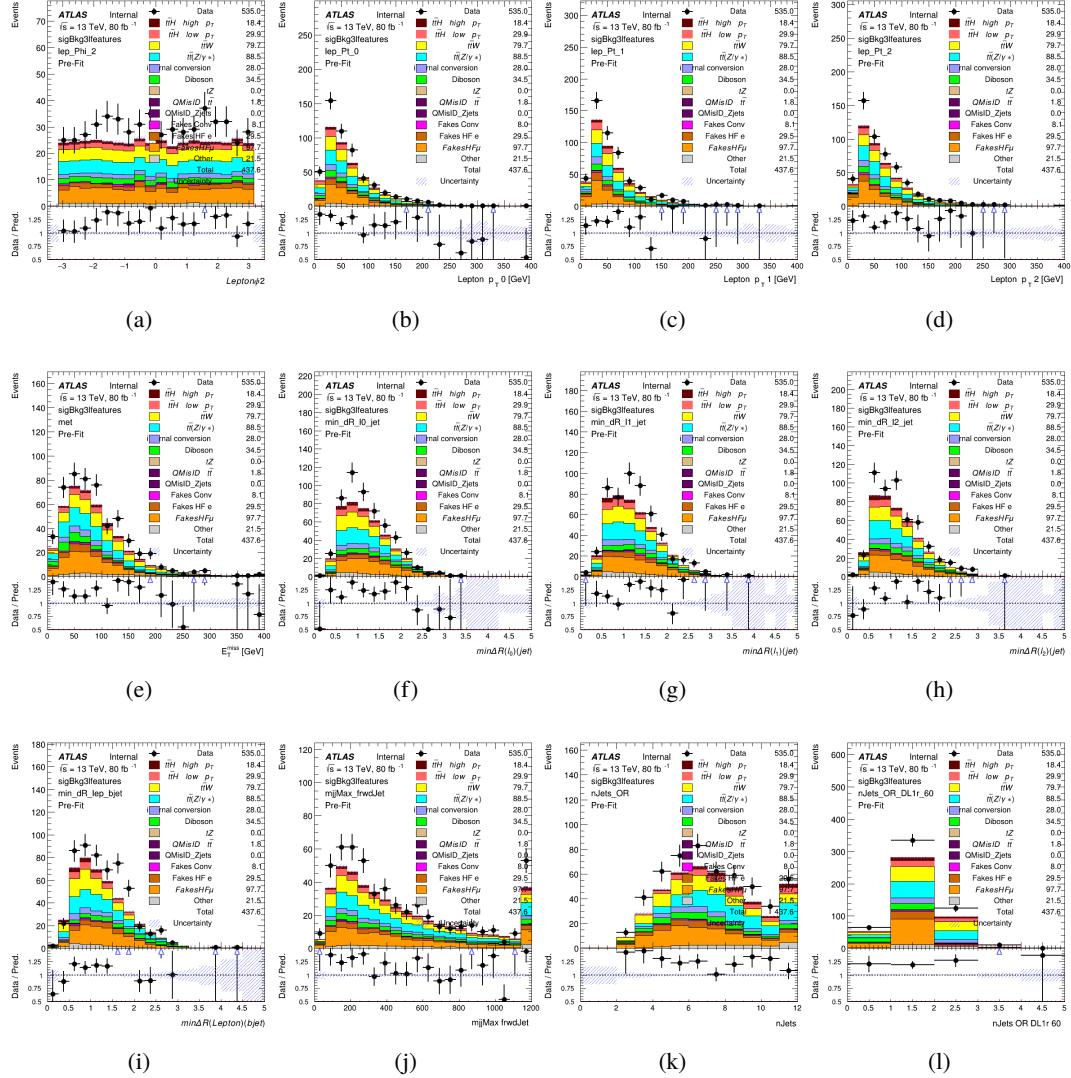
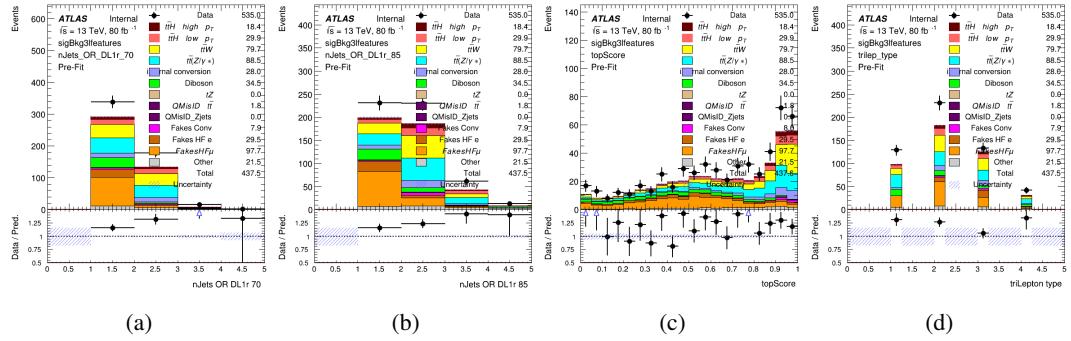


Figure A.24: Input features for sigBkg3l

Figure A.25: Input features for `sigBkg3l`

820 **A.3 Alternate b-jet Identification Algorithm**

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

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

Table 21: 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%

833 **A.4 Binary Classification of the Higgs p_T**

834 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer
 835 resolution. This means separating high and low p_T events is sufficient for this analysis. As

such, rather than attempting the reconstruct the full Higgs p_T spectrum, a binary classification approach is explored.

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

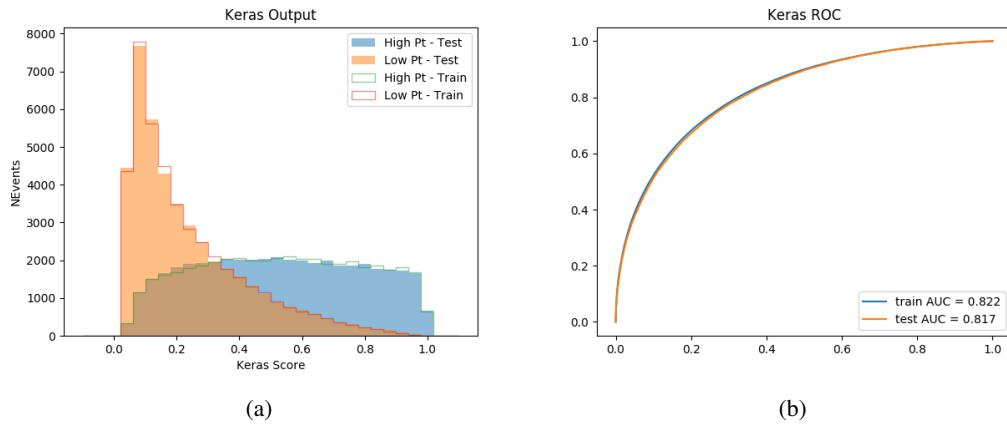


Figure A.26:

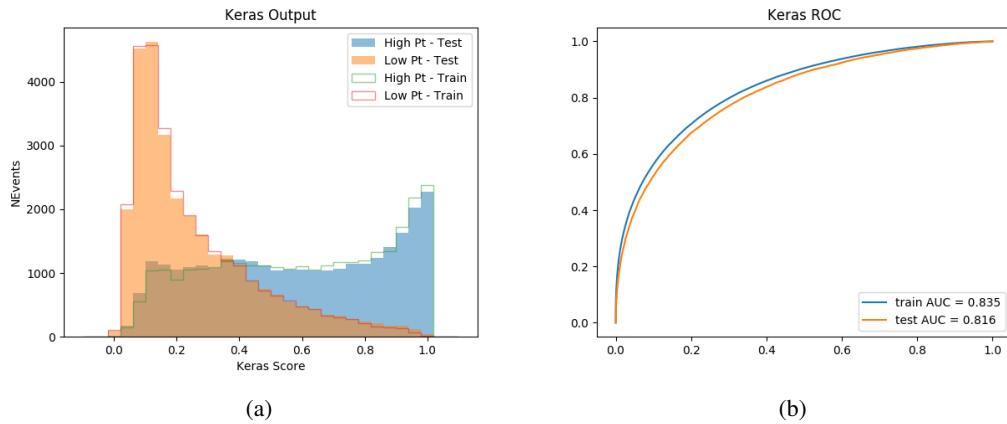


Figure A.27:

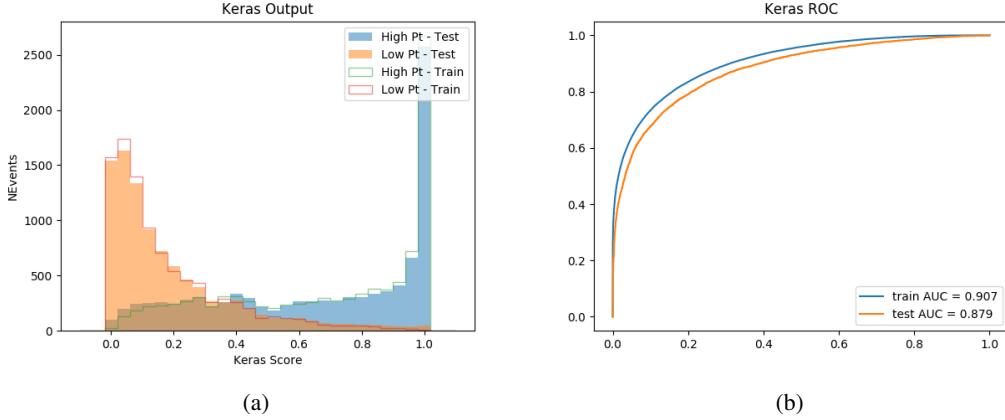


Figure A.28:

841 A.5 Impact of Alternative Jet Selection

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

847 B

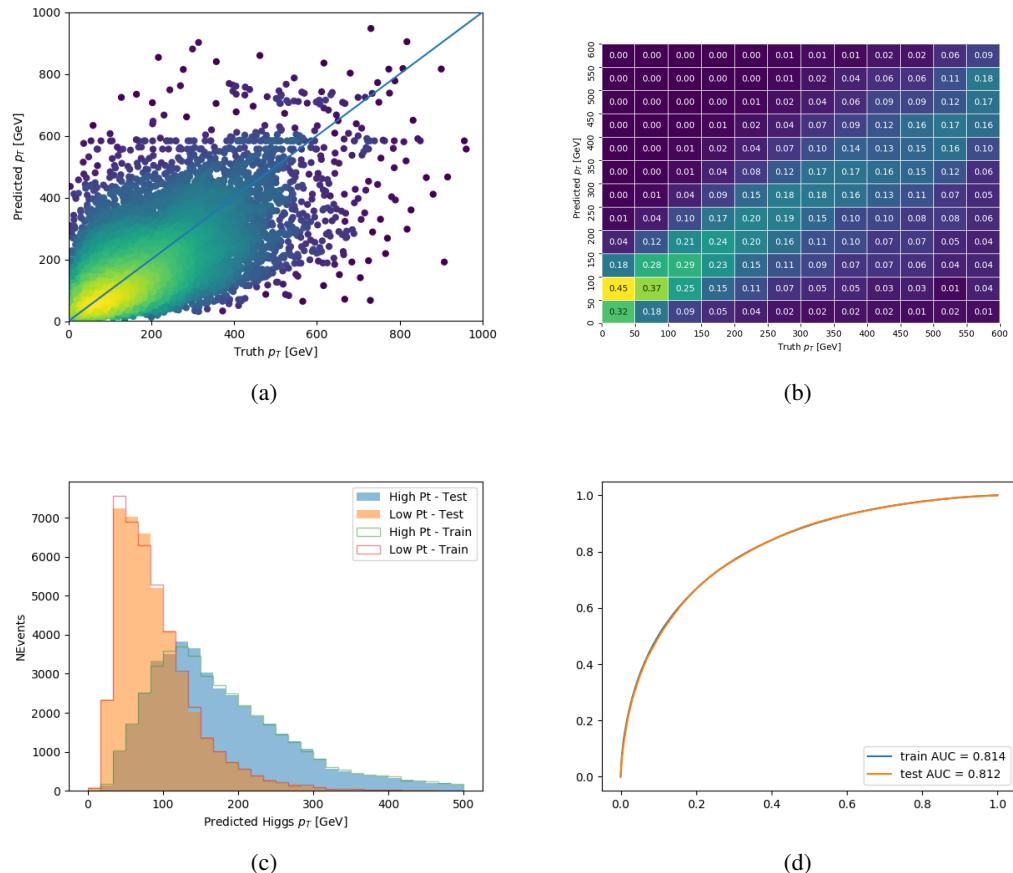


Figure A.29: