



ATLAS Note

GROUP-2017-XX



2nd 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	Truth Level Reconstruction	32
47	8.2	b-jet Identification	32
48	8.3	Higgs Reconstruction	33
49	8.4	p_T Prediction	33
50	8.5	3l Decay Mode	34
51	9	Signal Region Definitions	34
52	9.1	Pre-MVA Event Selection	34
53	9.1.1	2lSS Channel	35
54	9.1.2	3l Channel	35
55	9.2	Event MVA	35
56	9.3	Signal Region Definitions	35
57	9.3.1	2lSS	37
58	9.3.2	3l – Semi – leptonic	37
59	9.3.3	3l – Fully – leptonic	37
60	10	Background Rejection MVA	37
61	10.1	Background Rejection MVAs	37
62	10.1.1	2lSS - High p_T	39
63	10.1.2	2lSS - Low p_T	39
64	10.1.3	3l Semi-Leptonic - High p_T	39
65	10.1.4	3l Semi-Leptonic - Low p_T	39
66	10.1.5	3l Fully Leptonic - High p_T	39
67	10.1.6	3l Fully Leptonic - Low p_T	39
68	11	Systematic Uncertainties	39
69	12	Results	41
70	V	Conclusion	41
71	Appendices		43
72	A	Machine Learning Models	43
73	A.1	b-jet Identification Algorithms	43
74	A.1.1	2lSS Channel	43
75	A.1.2	3l Channel	45
76	A.2	Higgs Reconstruction Algorithms	46
77	A.2.1	2lSS Channel	46
78	A.2.2	3l Semi-leptonic Channel	46
79	A.2.3	3l Fully-leptonic Channel	46
80	A.3	p_T Prediction MVA	46
81	A.3.1	2lSS Channel	46

82	A.3.2 3l Semi-leptonic Channel	46
83	A.3.3 3l Fully-leptonic Channel	46
84	A.4 3l Decay MVA	46
85	A.5 Background Rejection MVAs	46
86	A.5.1 2lSS - High p_T	47
87	A.5.2 2lSS - Low p_T	48
88	A.5.3 3l Semi-Leptonic - High p_T	48
89	A.5.4 3l Semi-Leptonic - Low p_T	48
90	A.5.5 3l Fully Leptonic - High p_T	48
91	A.5.6 3l Fully Leptonic - Low p_T	48

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

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

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

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

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

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

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

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

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

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

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

139 **Part II**

140 **Theoretical Motivation**

141 **2 The Standard Model and the Higgs Boson**

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

¹⁴⁷ theory, invariant under the Lie Group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where C refers to color

¹⁴⁸ charge, L, the helicity of the particle, and Y, the hypercharge.

¹⁴⁹ 2.1 The Forces and Particles of the Standard Model

¹⁵⁰ The SM particles, summarized in figure 2.1, can be classified into two general categories based

¹⁵¹ on their spin: fermions, and bosons.

Standard Model of Elementary Particles

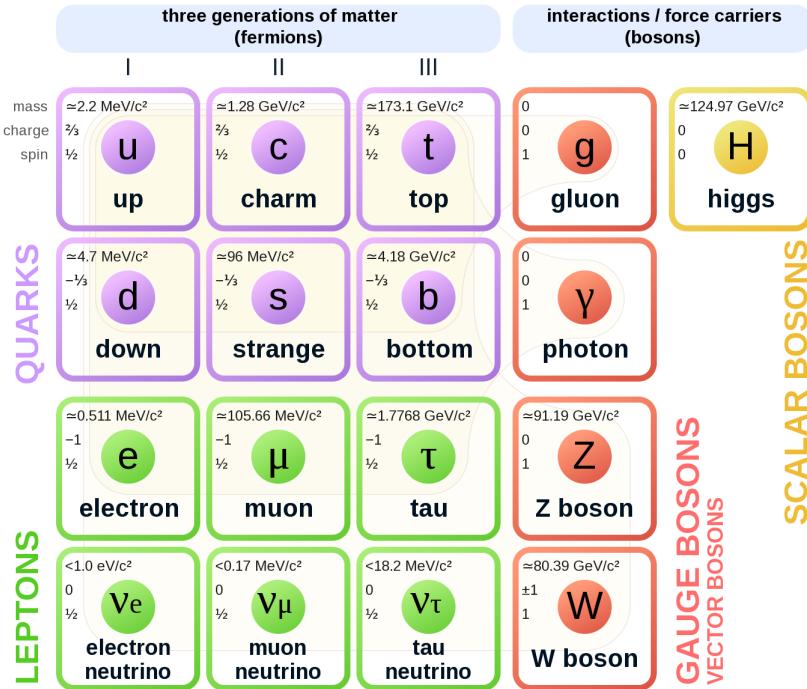


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. []

¹⁵² Fermions are particles with $\frac{1}{2}$ -integer spin, which according to the spin-statistics theorem,

¹⁵³ causes them to comply with the Pauli-exclusion principle []. They can be separated into two

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

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

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

163 And for the quarks:

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

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

¹⁶⁸ charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos,

¹⁶⁹ whose mass has been experimentally confirmed but is not accounted for in the SM.

¹⁷⁰ Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-

¹⁷¹ exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1

¹⁷² particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0,

¹⁷³ particle - the Higgs Boson. Of the gauge bosons, the W^+ , W^- and Z bosons - which are the

¹⁷⁴ mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon

¹⁷⁵ mediates the electric force, and the gluon mediates the strong force.

¹⁷⁶ 2.2 The Higgs Mechanism

¹⁷⁷ A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to

¹⁷⁸ the Lagrangian giving mass to the the gauge bosons would violate the underlying symmetry of

¹⁷⁹ the theory. This presents a clear problem with the theory: The experimental observation that the

¹⁸⁰ W and Z bosons have mass seems to contradict the basic structure of the SM.

¹⁸¹ Rather than abandoning gauge invariance, an alternative way for particles to acquire mass

¹⁸² beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout

¹⁸³ in 1964 []. This procedure for introducing masses for the gauge bosons while preserving local

¹⁸⁴ gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory

¹⁸⁵ by Weinberg in 1967 [].

¹⁸⁶ **2.2.1 The Higgs Field**

¹⁸⁷ The Higgs mechanism introduces a complex scalar $SU(2)$ doublet, Φ , with the form:

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

¹⁸⁸ This field introduces a scalar potential to the Lagrangian of the form:

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

¹⁸⁹ Where μ and λ are free parameters of the new field. This represents the most general
¹⁹⁰ potential allowed while preserving $SU(2)_L$ invariance and renormalizability. In the case that
¹⁹¹ $\mu^2 < 0$, this potential takes the form shown in figure 2.2.

¹⁹² The significant feature of this potential is that its minimum does not occur for a value of
¹⁹³ $\Phi = 0$. Instead, it is minimized when $|\Phi^\dagger \Phi| = -\mu^2/\lambda$. This means that in its ground state, the
¹⁹⁴ Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while
¹⁹⁵ the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

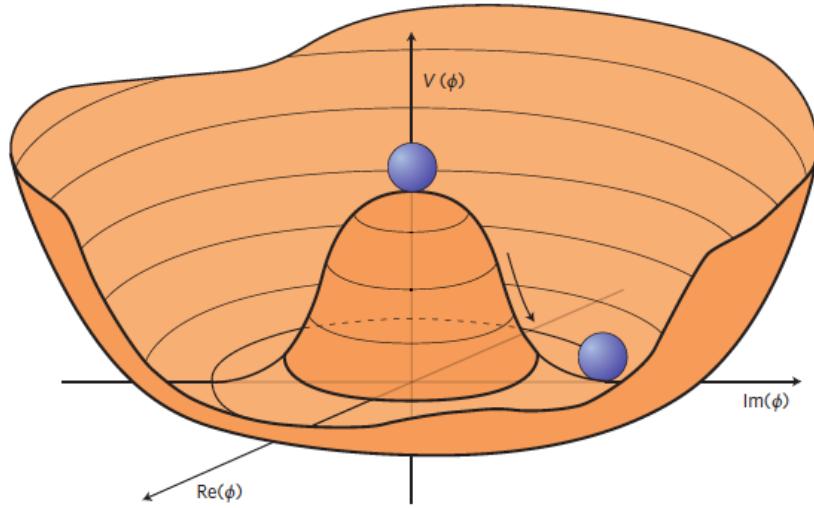


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ , for the case that $\mu^2 < 0$ [].

¹⁹⁶ the minimum is determined only by $\Phi^\dagger \Phi$, there is some ambiguity in the particular definition of
¹⁹⁷ the VEV, but it is generally represented as

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

¹⁹⁸ The full value of Φ can be written as

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

¹⁹⁹ with v being the value of the VEV, and H being the real value of the scalar field.

200 **2.2.2 Electroweak Symmetry Breaking**

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

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

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

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

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

²¹¹ $W_\mu^1, W_\mu^2, W_\mu^3$, and B_μ :

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

²¹² And the masses of these fields are given by:

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

²¹³ This produces exactly the particles we observe - three massive gauge bosons and a single
²¹⁴ massless photon. The massless photon represents the portion of the gauge symmetry, a single
²¹⁵ $U(1)$ of the electromagnetic force, that remains unbroken by the VEV.

²¹⁶ Interactions with the Higgs field also lead to the generation of the fermion masses, which
²¹⁷ in the Lagrangian take the form:

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

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

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

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

231 2.3 Limitations of the Standard Model

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

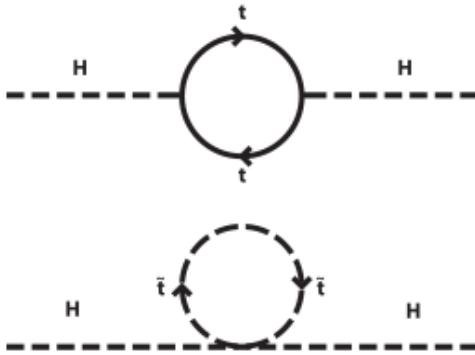


Figure 2.3: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram [6].

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

²³⁶ Higher dimension operators are a common way to paramaterize the effects of physics at very
²³⁷ high energies into

²³⁸ **3.1 Extensions to the Higgs Sector**

²³⁹ **3.2 Six Dimensional Operators**

²⁴⁰ While the SM has been tested to great precision, particularly at the LHC, it is generally accepted
²⁴¹ that it is only valid up to a certain energy scale. It is assumed that above a certain energy, at the
²⁴² scale where something like a Grand Unified Theory (GUT) or quantum gravity become relevant,
²⁴³ the SM will not be applicable.

244 Part III

245 The LHC and the ATLAS Detector

246 4 The LHC

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

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

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

262 higher energies until they are injected into the main ring. This process is summarized in figure
 263 [4.1](#). Protons extracted from a tank of ionized hydrogen are fed into a linear accelerator, LINAC2,
 264 where they reach an energy of 50 MeV. From there, they enter a series of three separate circular
 265 accelerators, before being injected into the main accelerator ring at an energy of 450 GeV. Within
 266 the main ring protons are separated into two separate beams moving in opposite directions,
 267 and their energy is increased to their full collision energy. Radiofrequency cavities are used to
 268 accelerate these particles and sort them into bunches. From 2015-2018, these bunches consisted
 269 of around 100 billion protons each with an energy of 6.5 TeV per proton, which collided at a rate
 270 of 40 MHz, or every 25 ns.

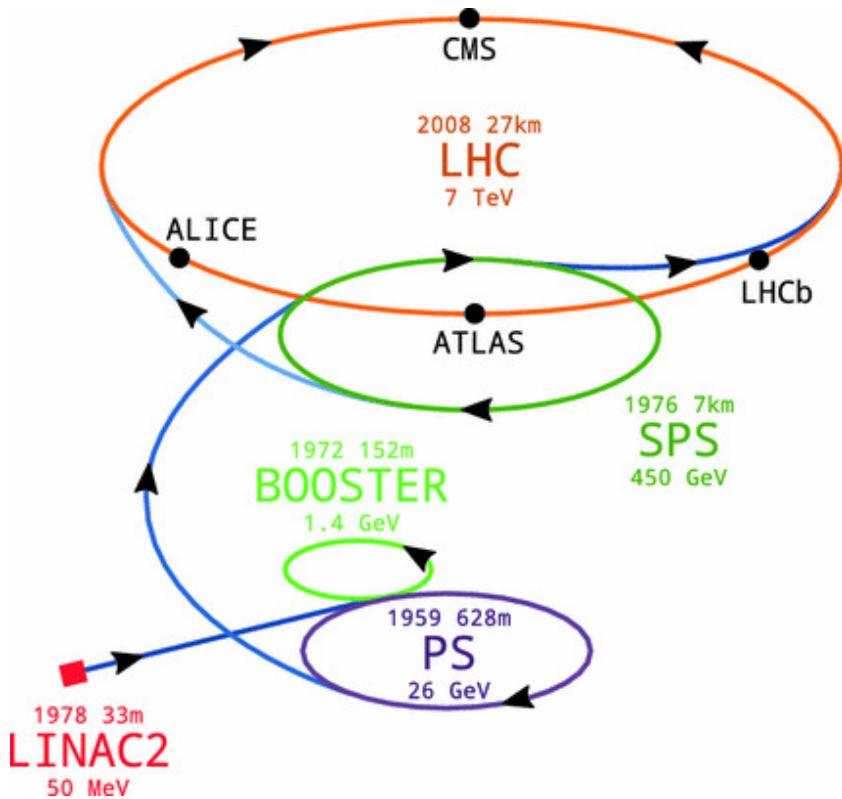


Figure 4.1: A summary of the accelerator chain used to feed protons into the LHC [].

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

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

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

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

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

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

²⁸² 5 The ATLAS Detector

²⁸³ ATLAS (a not terribly natural acronym for “A Toroidal LHC Apparatus”) is a general purpose
²⁸⁴ detector designed to maximize the detection efficiency of all physics objects, including leptons,
²⁸⁵ jets, and photons. This means it is capable of measuring all SM particles, with the exception of
²⁸⁶ neutrinos, the presence of which can be inferred based on missing transverse momentum. The
²⁸⁷ detector measures 44 m long, and 25 m tall.

²⁸⁸ The ATLAS detector consists of multiple layers, each of which serves a different purpose
²⁸⁹ in reconstructing collisions. At the very center of the detector is the interaction point where the
²⁹⁰ proton beams of the LHC collide.

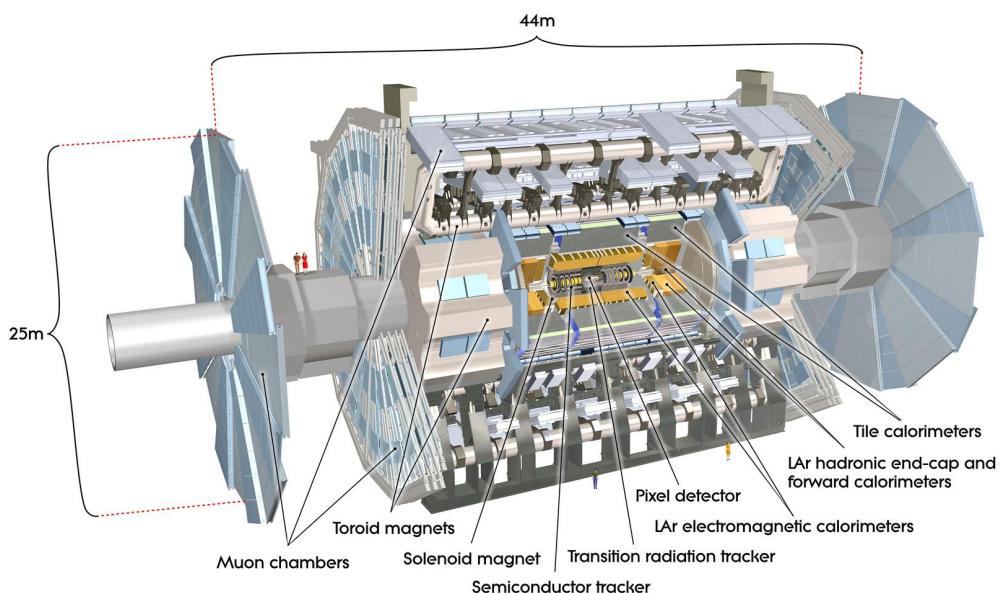


Figure 5.1: Cutaway view of the ATLAS detector, with labels of its major components [].

²⁹¹ **5.1 Inner Detector**

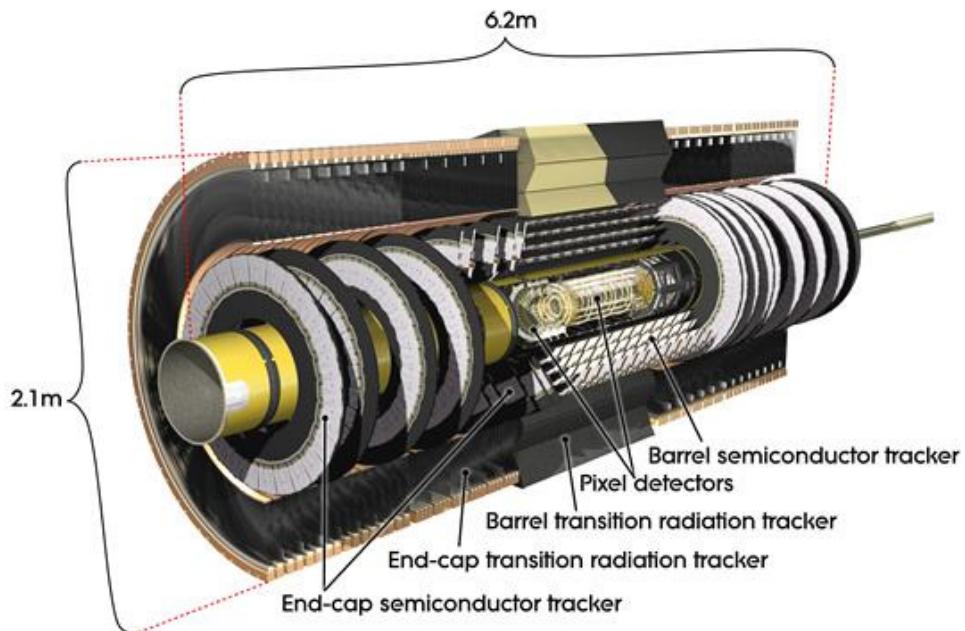


Figure 5.2: Cutaway view of the Inner Detector [].

²⁹² Just surrounding the interaction point is the Inner Detector, designed to track the path
²⁹³ of charged particles moving through the detector. An inner solenoid surrounding the Innder
²⁹⁴ Detector is used to produces a magnetic field of 2 T. This large magnetic field causes the path
²⁹⁵ of charged particles moving through the Inner Detector to bend. Because this magnetic field is
²⁹⁶ uniform and well known, it can be used in conjunction with the curvature of a particles path to
²⁹⁷ measure its charge and momentum.

²⁹⁸ The Inner Detector consists of three components - the Pixel Detector, the Semi-Conductor
²⁹⁹ Tracker (SCT), and the Transition Radiation Tracker (TRT). The Pixel Detector is the innermost
³⁰⁰ of these, beginning just 33.25 mm away from the beam line. It consists of three silicon layers

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

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

304 **5.2 Calorimeters**

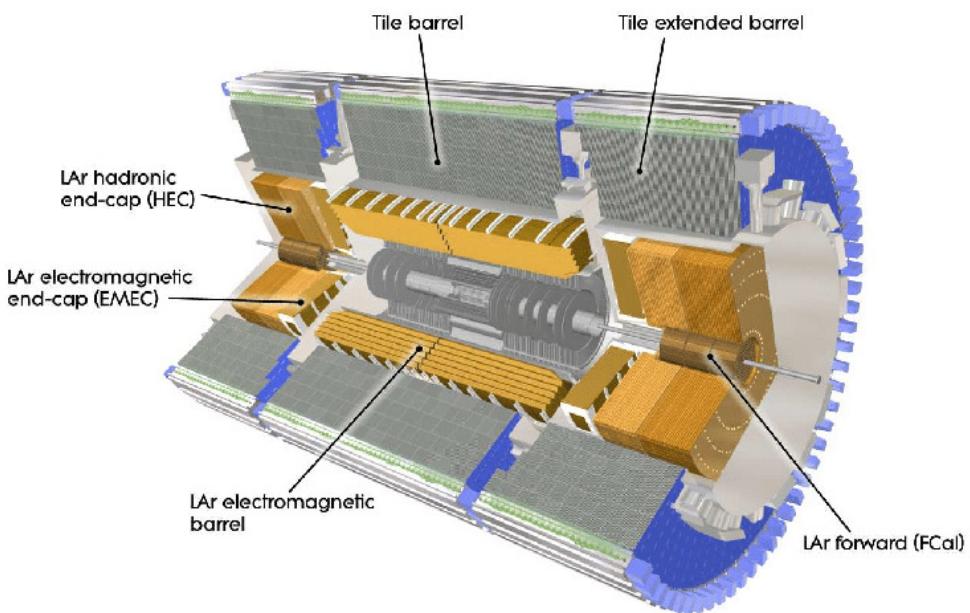


Figure 5.3: Cutaway view of the calorimeter system of the ATLAS detector [].

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

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

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

317 **5.3 Muon Spectrometer**

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

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

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

332 1200 tracking chambers are placed in the muon system in order to precisely measure the
333 tracks of muons with high spatial resolution.

334 **5.4 Trigger System**

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

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

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

353 **Part IV**

354 **Search for Dimension-Six Operators**

355 **6 Data and Monte Carlo Samples**

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

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

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

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

367 **6.1 Data Samples**

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

- 371 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
372 _PHYS_StandardGRL_All_Good_25ns.xml

- 373 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
374 _PHYS_StandardGRL_All_Good_25ns.xml

- 375 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL
376 _All_Good_25ns_Triggerno17e33prim.xml

- 377 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL
378 _All_Good_25ns_Triggerno17e33prim.xml

³⁷⁹ **6.2 Monte Carlo Samples**

³⁸⁰ Several Monte Carlo (MC) generators were used to simulate both signal and background pro-
³⁸¹ cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific
³⁸² 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

³⁸³ **7 Object Reconstruction**

³⁸⁴ All analysis channels considered in this note share a common object selection for leptons and
³⁸⁵ jets, as well as a shared trigger selection.

386 **7.1 Trigger Requirements**

387 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.

388 **7.2 Light Leptons**

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

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

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

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

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

445 8.1 Decay Candidate Reconstruction

446 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson
447 using MC simulations of $t\bar{t}H$ events. These include light leptons and jets. Reconstructed

448 physics objects are matched to truth level particles, in order to identify the parents of these
449 reconstructed objects. The kinematics of the decay product candidates as well as event level
450 variables are used as inputs.

451 Leptons considered as possible Higgs and top decay candidates are required to pass the
452 selection described in section ???. For jets, however, it is found that a large fraction that originate
453 from either the top decay or the Higgs decay fall outside the selection described in section ???.
454 Specifically, jets from the Higgs decay tend to be soft, with 37% having $p_T < 25$ GeV. Therefore
455 jets with $p_T < 15$ GeV are considered as possible candidates in the models described below. The
456 spectrum is found to be well modeled even down to this low p_T threshold, as shown in section
457 9.1. As they are expected to originate from the primary vertex, jets are also required to pass a
458 JVT cut.

459 8.2 b-jet Identification

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

465 Exactly two b-jets are expected in the final state of $t\bar{t}H - ML$ events. However, in both
466 the 3l and 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r b-tag

467 working point is used). Therefore, for events which have exactly one, or more than two, b-tagged
 468 jets, deciding which combination of jets correspond to the top decay is non-trivial. Further,
 469 events with 1 b-tagged jet represent just over half of all $t\bar{t}H - ML$ events. Of those, both b-jets
 470 are reconstructed by the detector 75% of the time. Therefore, rather than adjusting the selection
 471 to require exactly 2 b-tagged jets, and losing more than half of the signal events, a neural network
 472 is used to predict which pair of jets is most likely to correspond to truth b-jets.

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

475 **8.2.1 2lSS Channel**

476 For the 2lSS channel, the input features shown in table 4 are used for training. Here j_0 and j_1
 477 are the two jet candidates, while l_0 and l_1 are the two leptons in the event, ordered by p_T . jet
 478 DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,
 479 where 5 represents the tightest working point and 1 represents the loosest. The variables nJets
 480 DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and
 481 85% b-tag working points, respectively.

482 As there are far more incorrect combinations than correct ones, by a factor of more than
 483 20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random
 484 sample of 5 million incorrect entries are used for training, along with close 1 million correct

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 2ISS b-jet identification algorithm

485 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for
 486 training.

487 The difference between the distributions for a few of these features for the correct (i.e.
 488 both jets are truth b-jets), and incorrect combinations are shown in figure A.1. The correct and
 489 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences
 490 in the distributions.

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

493 Based on the results of grid search evaluation, the optimal architecture is found to include
 494 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as
 495 overfitting is found to not be an issue. The output score distribution as well as the ROC curve
 496 for the trained model are shown in figure ?? . The model is found to identify the correct pairing
 497 of jets for 73% of 2ISS signal events on test data.

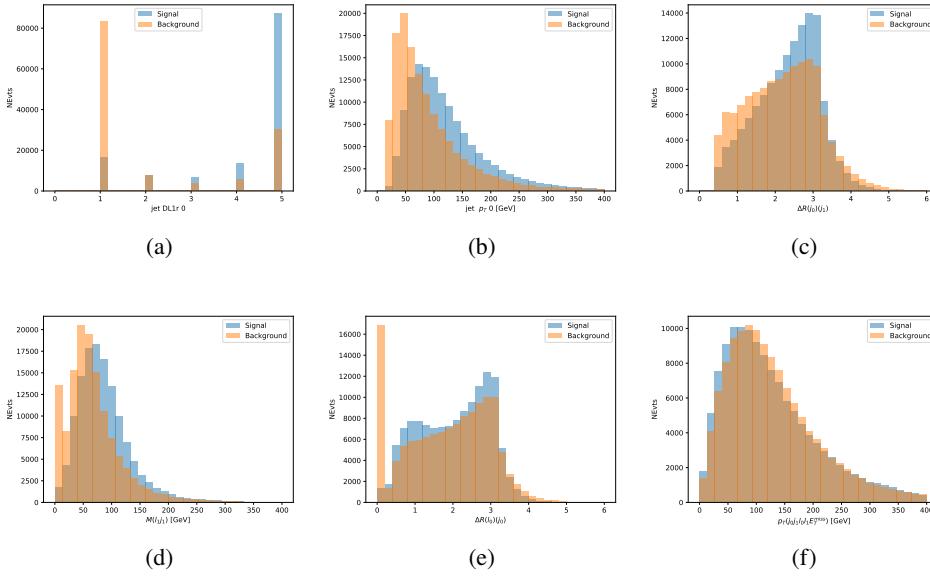


Figure 8.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

502 The accuracy of the model for different values of n-bjets, compared to this naive approach,
 503 is shown in table ??.

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%

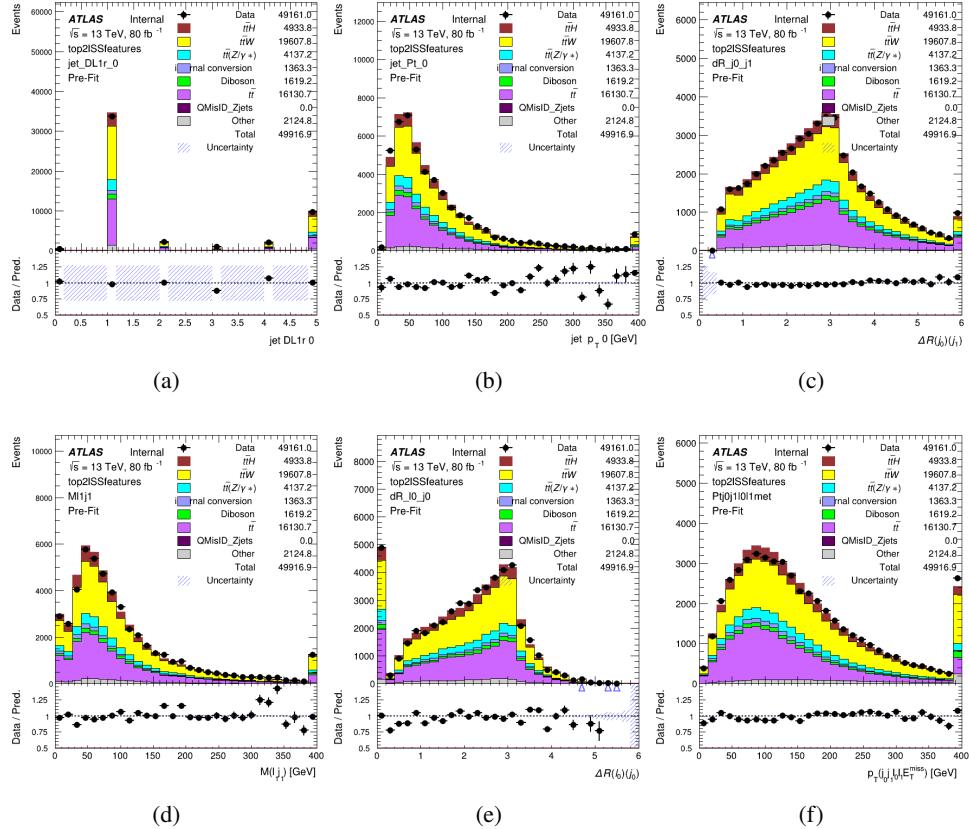


Figure 8.2: Data/MC comparisons of input features for top2lSS training for 80 fb^{-1} of data.

504 8.2.2 3l Channel

505 The input features used in the 3l channel are listed in table ??, with the same naming convention
506 as the 2lSS channel.

507 A few of these features are shown in figure ??, comparing the distributions for correct and
508 incorrect combinations of jets.

509 The modeling of these inputs is validated against data, with figure ?? showing good general

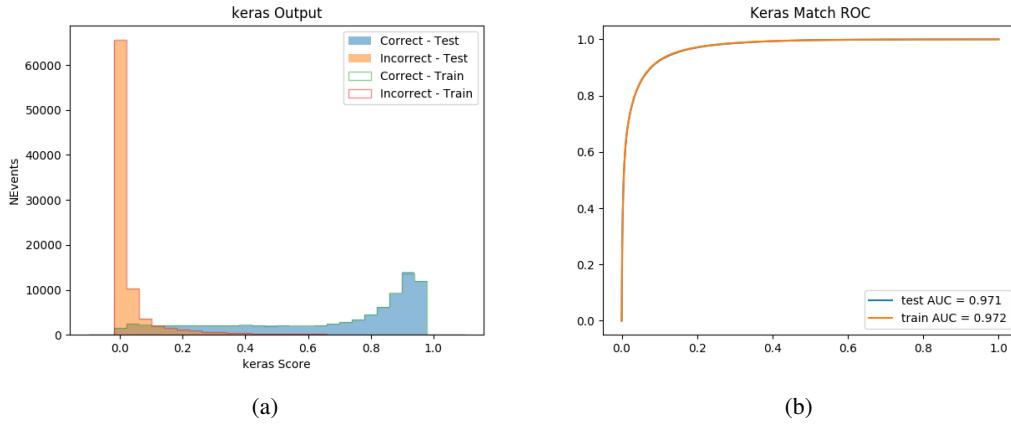


Figure 8.3: (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

510 agreement between data and MC. Plots for the complete list of features can be found in section A.

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

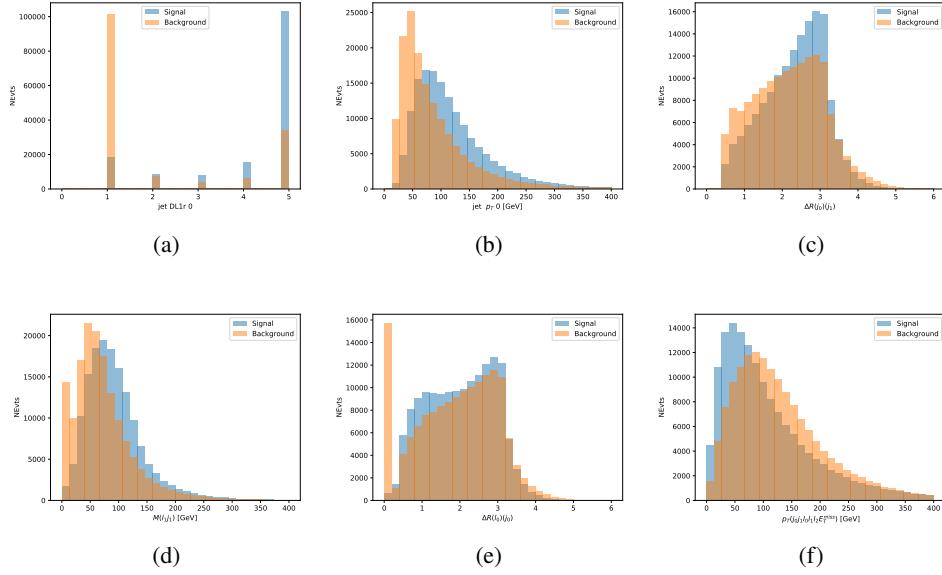


Figure 8.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

515 trained model are shown in figure A.1.2.

516 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal
517 events. The accuracy of the model is summarized in table ??.

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%

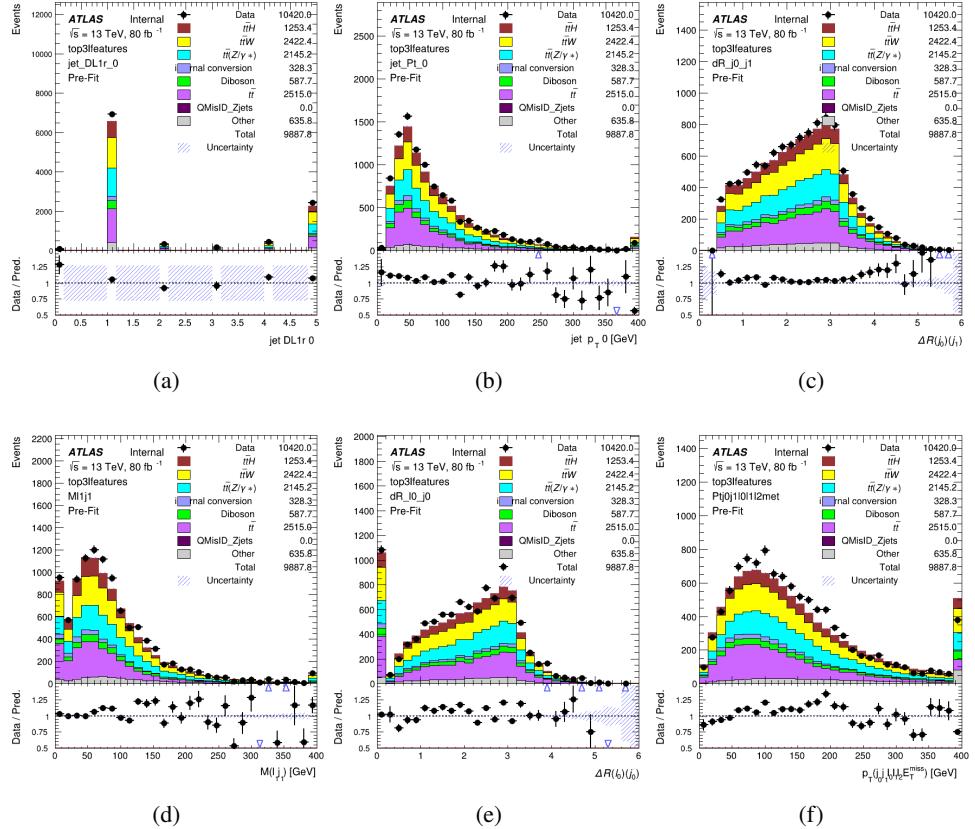
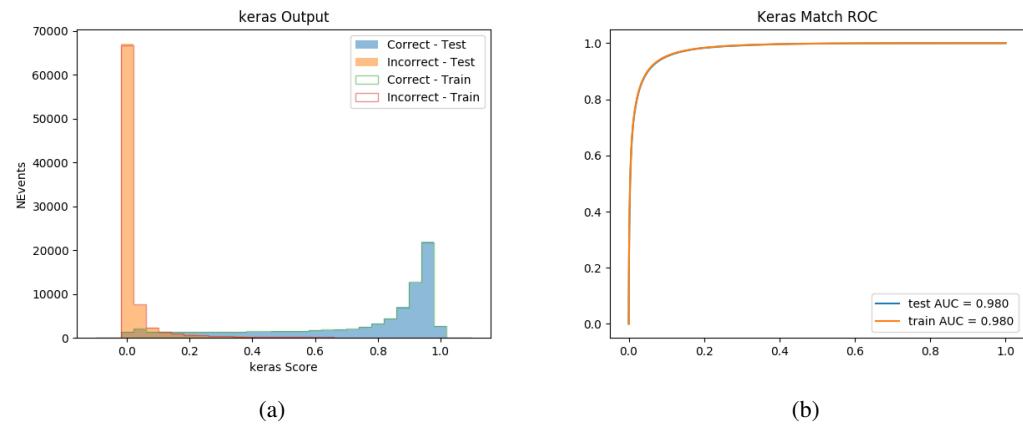
Figure 8.5: Data/MC comparisons of input features for top3l training for 80 fb^{-1} of data.

Figure 8.6: tmp

518 8.3 Higgs Reconstruction

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

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

526 3l events are therefore categorized as either semi-leptonic or fully-leptonic. In the semi-
527 leptonic case the reconstructed decay products consist of two jets and a single leptons. For
528 the fully-leptonic case, the decay products include 2 of the three leptons associated with the
529 event. For training the models, events are separated into these two categories using truth level
530 information. A separate MVA, described in section 8.5, is used to make this distinction at reco
531 level and determine which model to use.

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

536 **8.3.1 2lSS Channel**

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

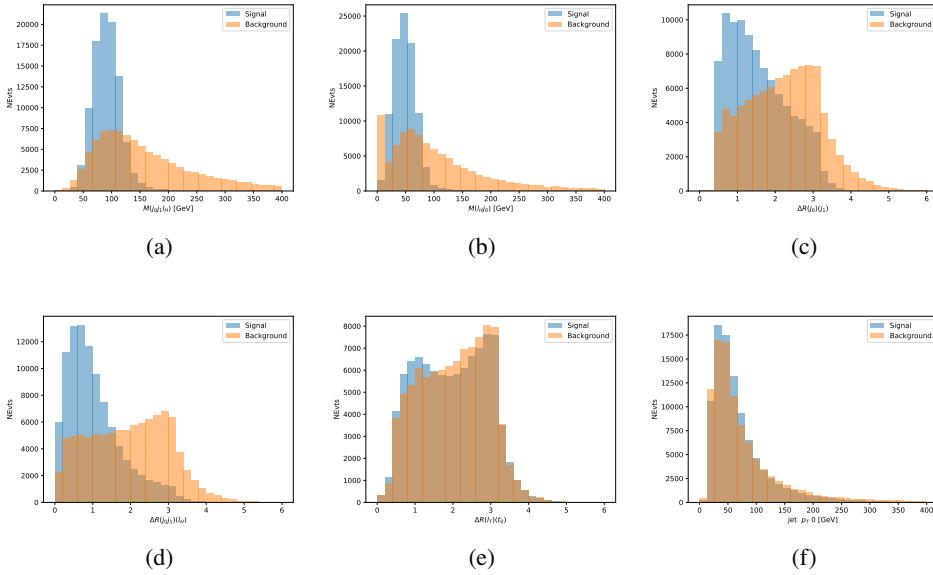


Figure 8.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

542 The neural network identifies the correct combination 55% of the time. It identifies the
543 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
544 81% of the time.

545 8.3.2 3l Semi-leptonic Channel

546 For 3l $t\bar{t}H$ where the Higgs decay semi-leptonically, the decay products include one of the three
547 leptons and two jets. In this case, the other two leptons originated from the decay of the tops,

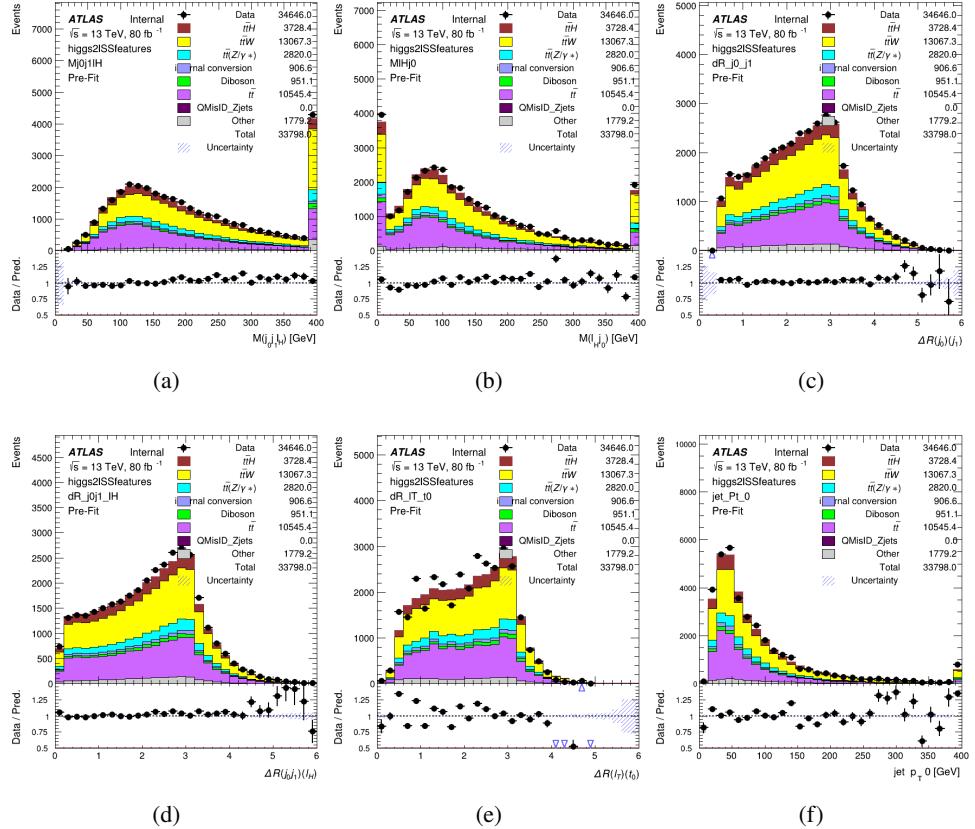


Figure 8.8: Data/MC comparisons of input features for higgs2lSS training for 80 fb^{-1} of data.

meaning the opposite-sign (OS) lepton cannot have come from the Higgs. This leaves only the two same-sign (SS) leptons as possible Higgs decay products.

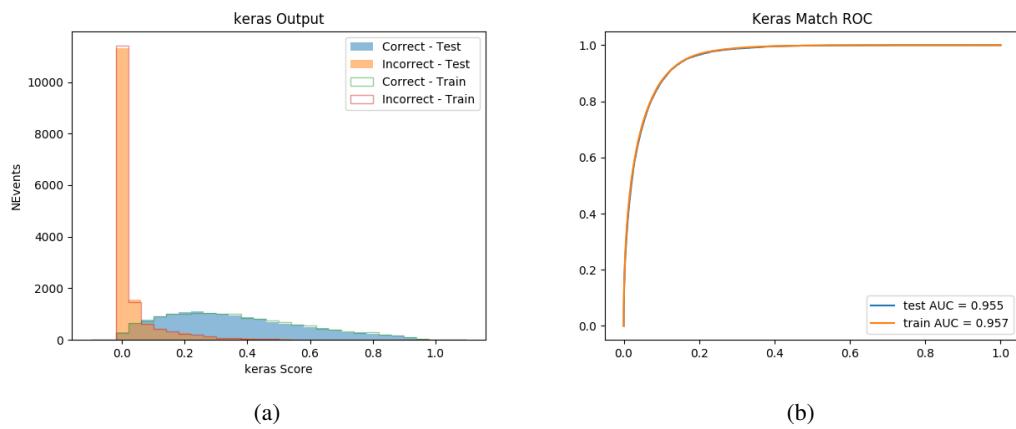


Figure 8.9: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

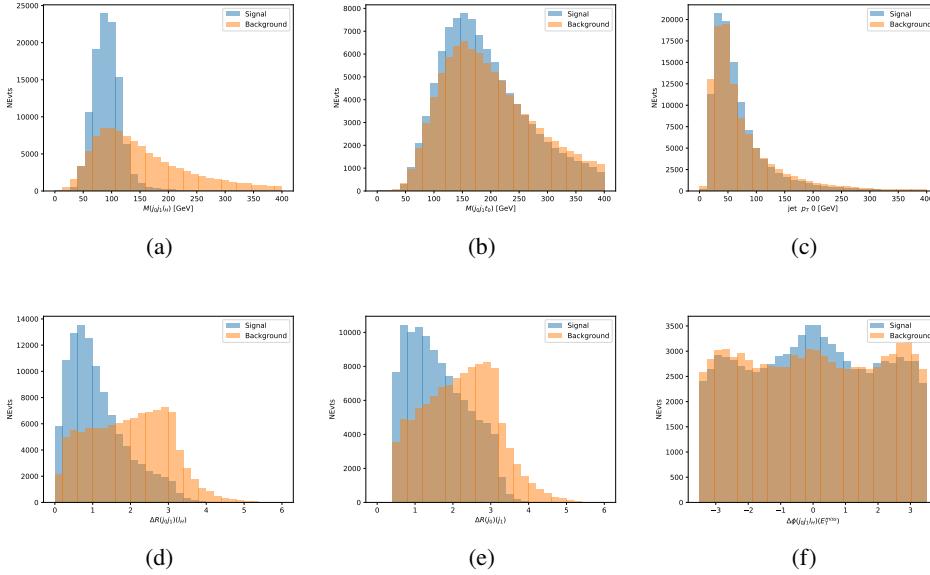


Figure 8.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

552 The neural network identifies the correct combination 65% of the time. It identifies the
 553 correct lepton 85% of the time, an selects the correct lepton and at least one of the correct jets
 554 83% of the time.

555 8.3.3 3l Fully-leptonic Channel

556 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 557 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to

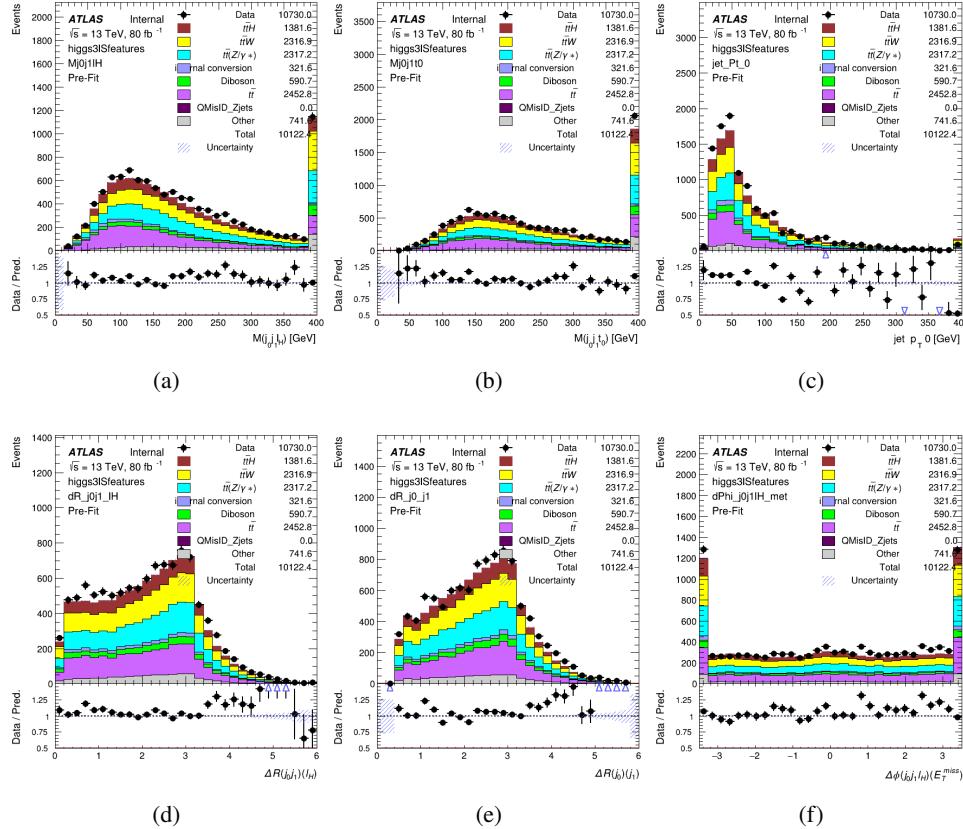


Figure 8.11: Data/MC comparisons of input features for higgs3IS training for 80 fb^{-1} of data.

558 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 559 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 560 and one where it is not.

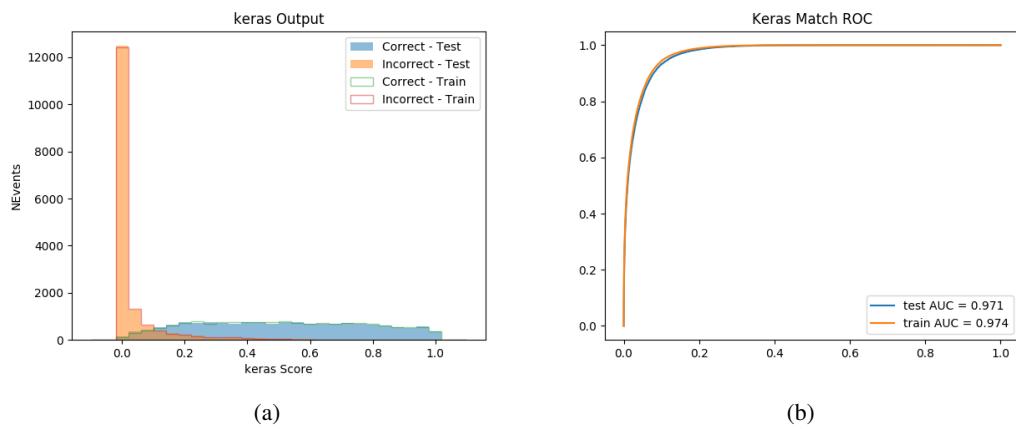


Figure 8.12: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

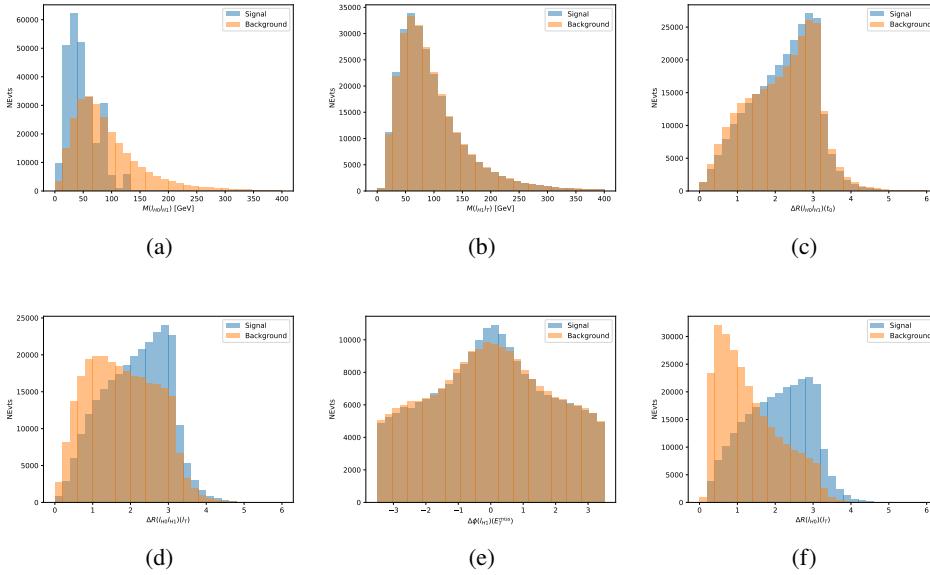


Figure 8.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

563 The correct lepton is identified 80% of the time.

564 8.4 p_T Prediction

565 Once the most probable decay products have been identified, their kinematics are used as inputs
 566 to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,
 567 a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay

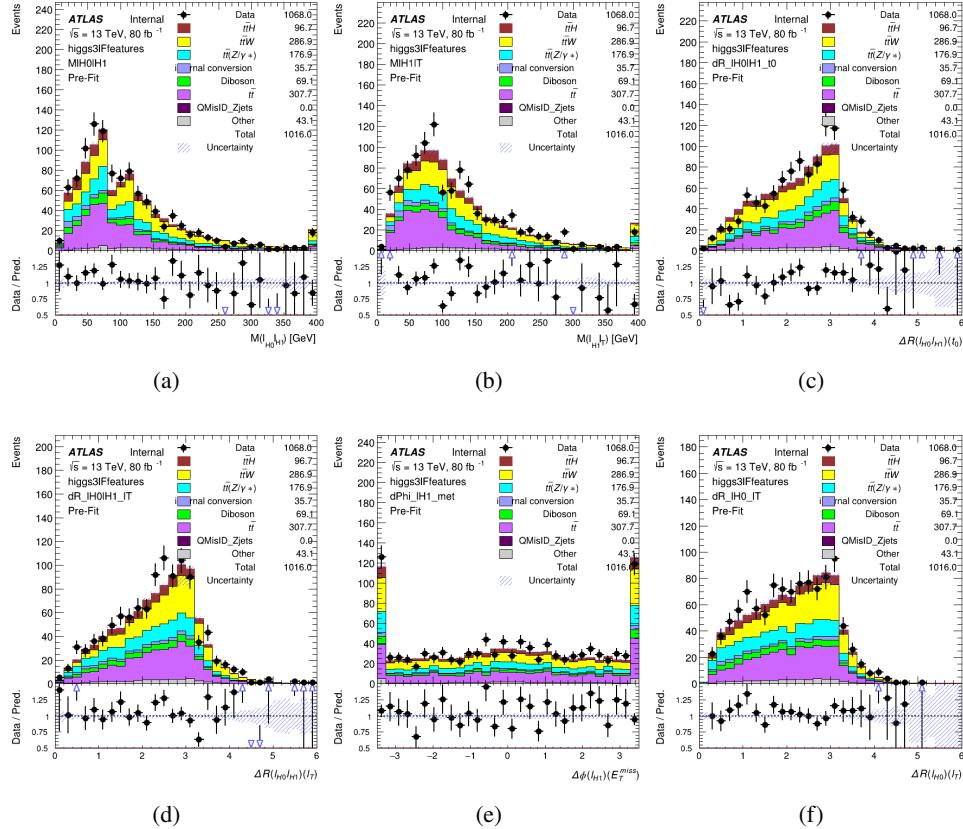


Figure 8.14: Data/MC comparisons of input features for higgs3IF training for 80 fb^{-1} of data.

568 are included as well, as these are found to improve performance. The truth p_T of the Higgs, as
 569 predicted by MC, are used as labels.

570 Separate models are built for each channel - 2lSS, 3l Semi-leptonic and 3l Fully-leptonic.

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

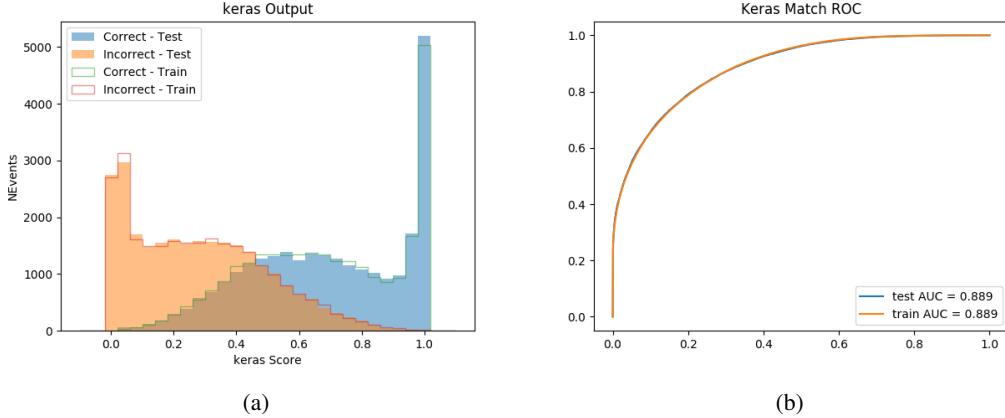


Figure 8.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

574 **8.4.1 2ISS Channel**

575 The input variables listed in table ?? are used to predict the Higgs p_T in the 2ISS channel. Here
 576 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
 577 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it most have come from the
 578 decay of one of the top quarks. The Higgs Reco Score and b-jet Reco Score are the outputs of
 579 the Higgs reconstruction algorithm, and the b-jet identification algorithm, respectively.

580 The optimal neural network architecture for this channel is found to consist of 5 hidden
 581 layers with 40 nodes each. The input data set includes 1.2 million events, 10% of which is used
 582 for testing, the other 90% for training. Training is found to converge after around 150 epochs.

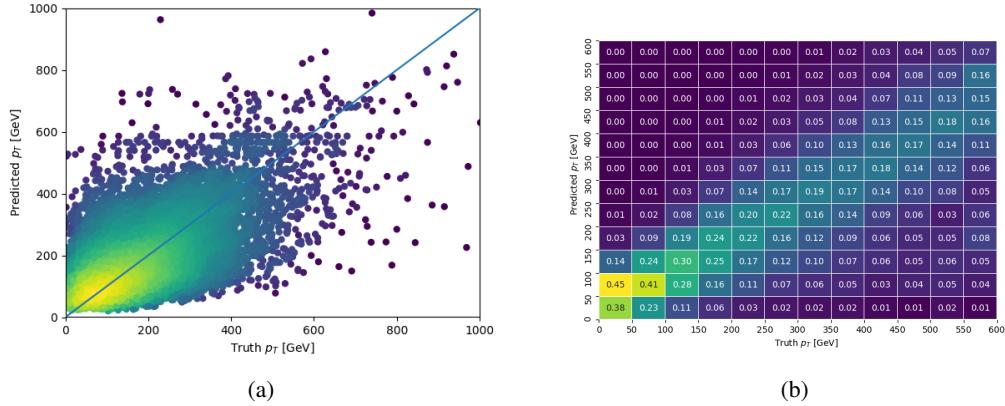


Figure 8.16:

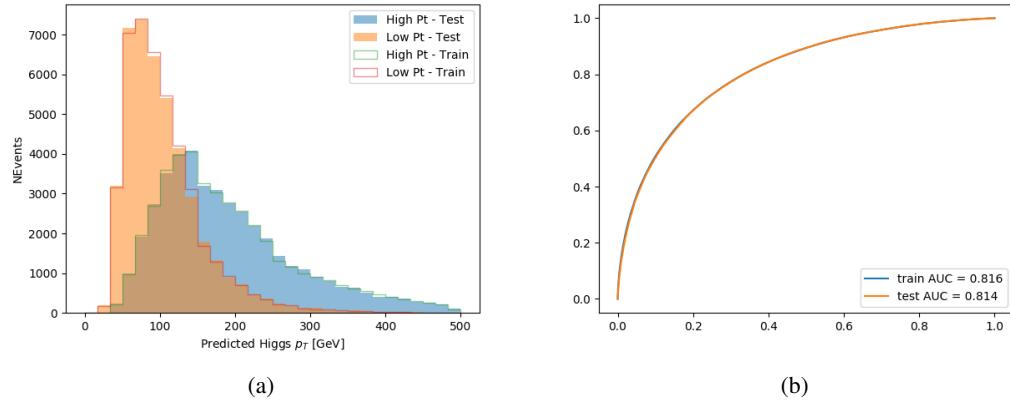


Figure 8.17:

583 8.4.2 3l Semi-leptonic Channel

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

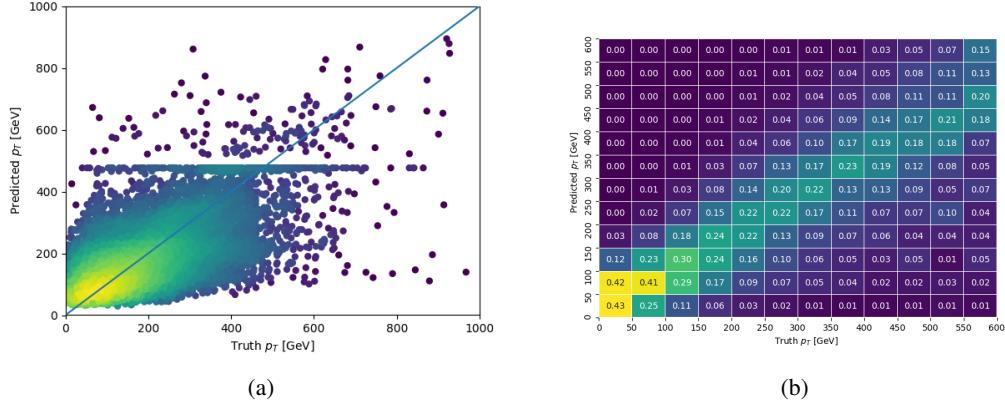


Figure 8.18:

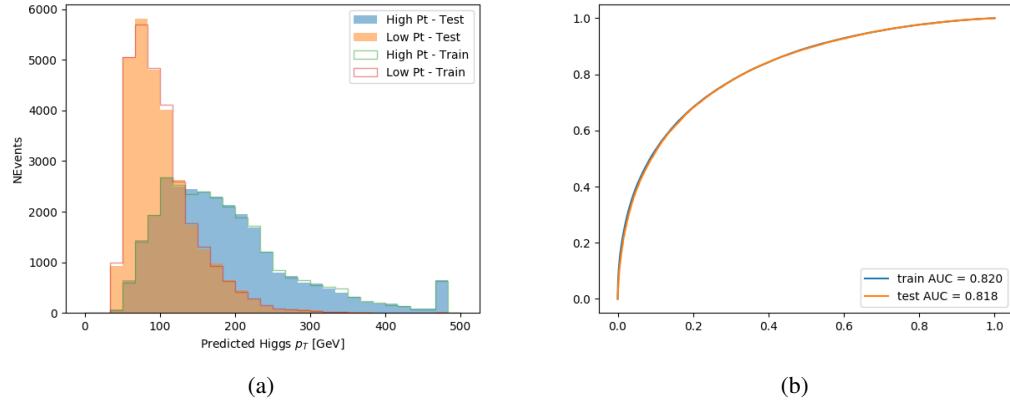


Figure 8.19:

587 8.4.3 3l Fully-leptonic Channel

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

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

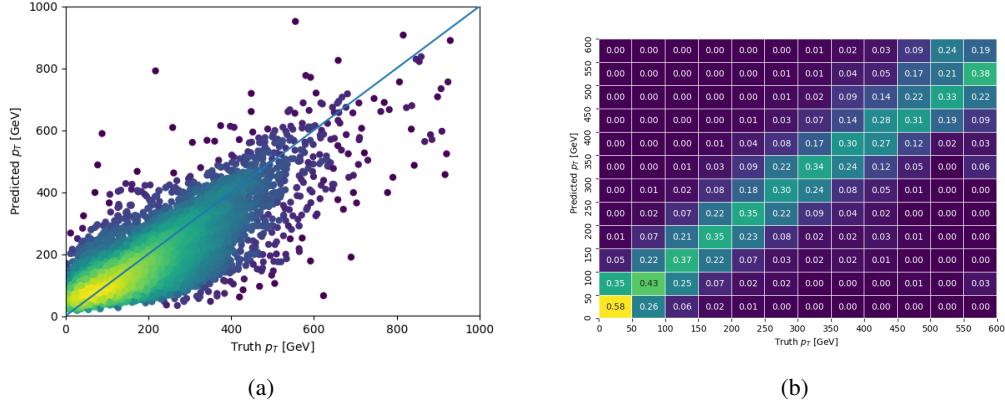


Figure 8.20:

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

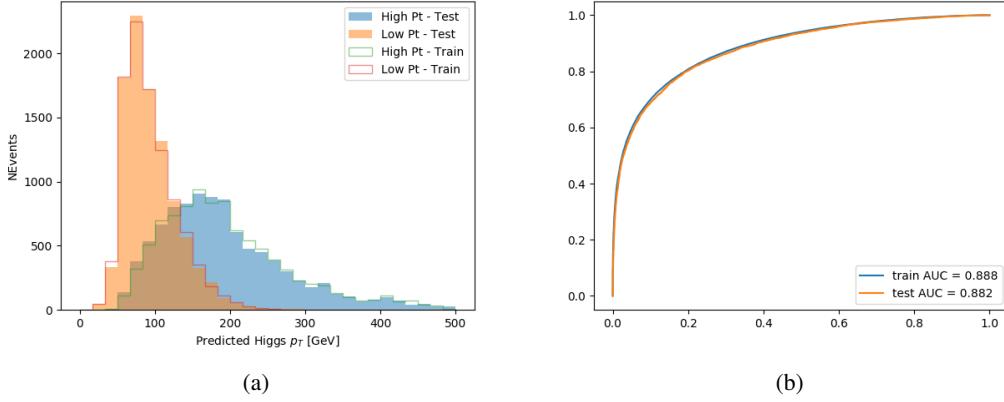


Figure 8.21:

593 8.5 3l Decay Mode

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

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

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

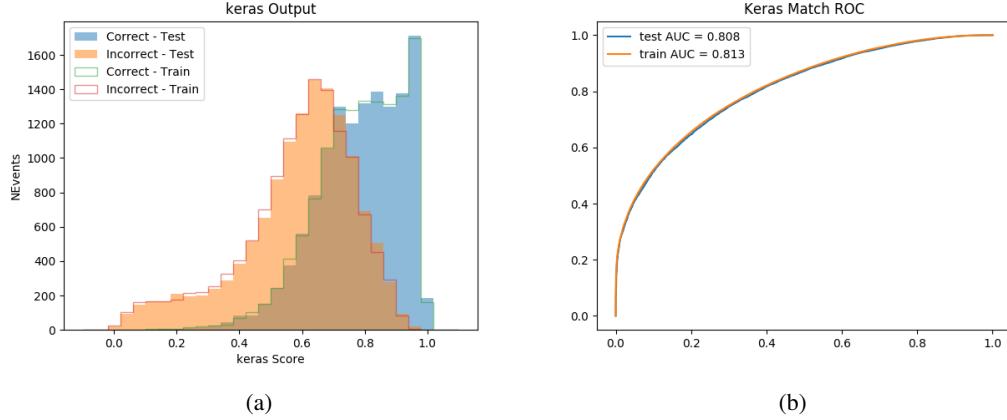


Figure 8.22:

605 9 Signal Region Definitions

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

609 9.1 Pre-MVA Event Selection

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

- 612 • Two very tight, same-charge, light leptons with $p_T > 20$ GeV

- 613 • $>=4$ reconstructed jets, $>=1$ b-tagged jets

- 614 • No reconstructed tau candidates

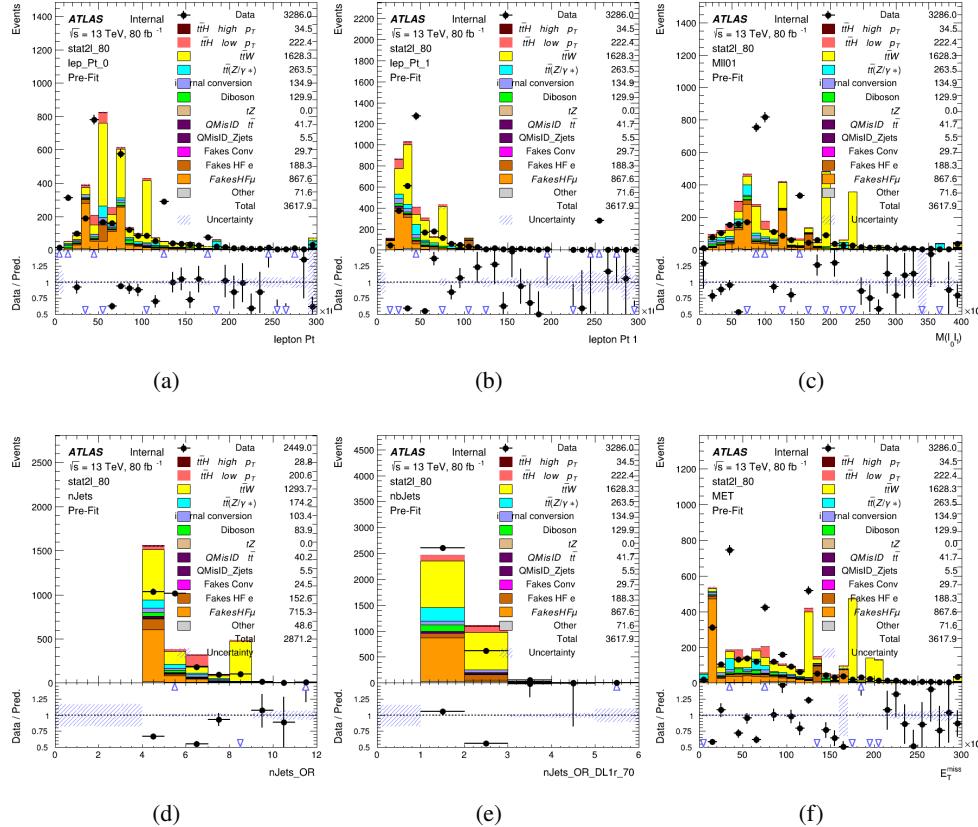


Figure 9.1:

- 615 For the 31 channel, the following selection is applied:

- 616 • Three light leptons with total charge ± 1
 617 • Same charge leptons are required to be very tight, with $p_T > 20$ GeV

- 618 • Opposite charge lepton must be loose, with $p_T > 10 \text{ GeV}$

- 619 • $>=2$ reconstructed jets, $>=1$ b-tagged jets

- 620 • No reconstructed tau candidates

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

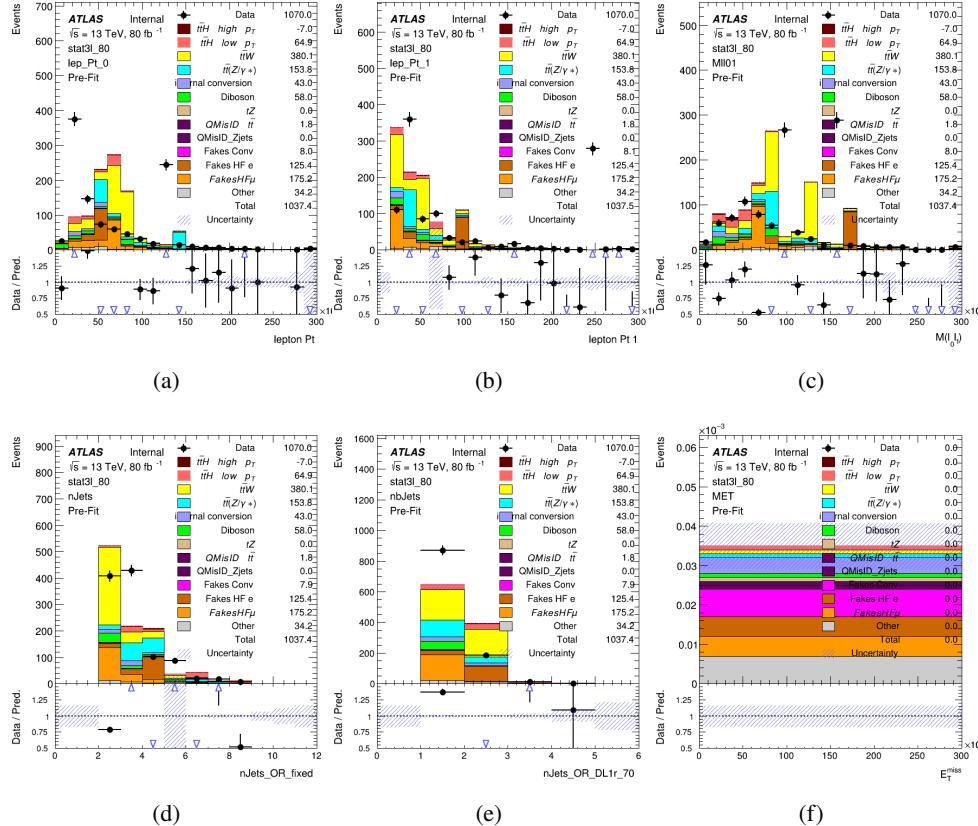


Figure 9.2:

622 9.2 Event MVA

623 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
624 from background for each analysis channel - 2lSS, 3l semi-leptonic, and 3l fully leptonic. In
625 particular, Neural Networks produced with Tensorflow are trained using the kinematics of signal
626 and background events derived from Monte Carlo simulations. Further, because the background
627 composition differs for events with a high reconstructed Higgs p_T compared to events with low
628 reconstructed Higgs p_T , separate MVAs are produced for high and low p_T regions.

629 Output distributions of each MVA are shown in figure 9.2. Detailed explanations of each
630 of the models can be found in section A.

631 9.3 Signal Region Definitions

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

636 For each event, depending on the channel as well as the predicted p_T of the Higgs derived
637 from the algorithm described in section 8.4, a cut on the appropriate background rejection
638 algorithm is applied. The specific selection used, and the event yield in each channel after this
639 selection has been applied, is summarized below.

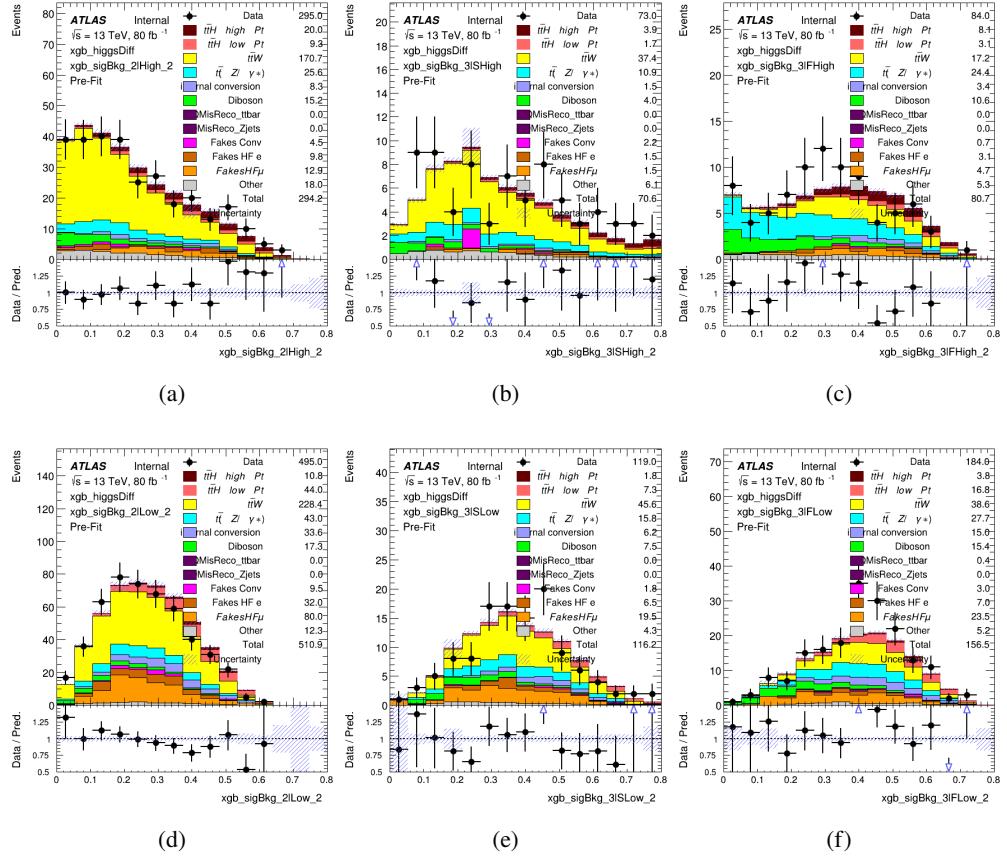


Figure 9.3: scores

640 **9.3.1 2lSS**

641 **9.3.2 3l – Semi – leptonic**

642 **9.3.3 3l – Fully – leptonic**

643 **10 Background Rejection MVA**

644 **10.1 Background Rejection MVAs**

645 Separate mdoels are used in order to distinguish signal events from background for each analysis
646 channel - 2lSS, 3l semi-leptonic, and 3l fully leptonic. In particular, Neural Networks produced
647 with Tensorflow are trained using the kinematics of signal and background events derived from
648 Monte Carlo simulations. Further, because the background composition differs for events with
649 a high reconstructed Higgs p_T compared to events with low reconstructed Higgs p_T , separate
650 MVAs are produced for high and low p_T regions.

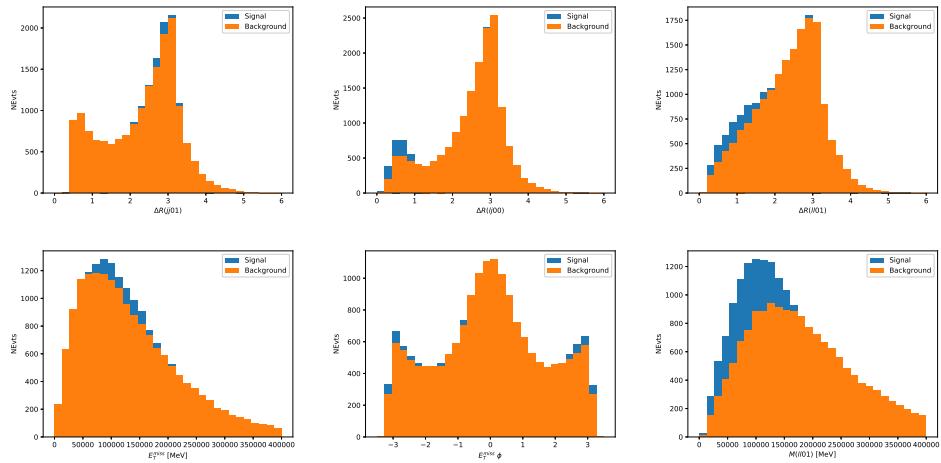


Figure 10.1:

651 **10.1.1 2lSS - High p_T**

652 **10.1.2 2lSS - Low p_T**

653 **10.1.3 3l Semi-Leptonic - High p_T**

654 **10.1.4 3l Semi-Leptonic - Low p_T**

655 **10.1.5 3l Fully Leptonic - High p_T**

656 **10.1.6 3l Fully Leptonic - Low p_T**

657 **11 Systematic Uncertainties**

658 The systematic uncertainties that are considered are summarized in table ???. These are imple-
659 mented in the fit either as a normalization factors or as a shape variation or both in the signal
660 and background estimations. The numerical impact of each of these uncertainties is outlined in
661 section 12.

662 The uncertainty in the combined 2015+2016 integrated luminosity is derived from a
663 calibration of the luminosity scale using x-y beam-separation scans performed in August 2015
664 and May 2016 [[lumi](#)].

Table 7: 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

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

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

667 to the uncertainty in the jet energy scale [**jes**] are decomposed into uncorrelated components and

668 treated as independent sources in the analysis.

669 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses

670 [**btag_cal**] are also decomposed into uncorrelated components. The large number of components

671 for b-tagging is due to the calibration of the distribution of the BDT discriminant.

672 The systematic uncertainties associated with the signal and background processes are

673 accounted for by varying the cross-section of each process within its uncertainty.

674 **12 Results**

675 A maximum likelihood fit is performed simultaneously over the regions described in section

676 **??.**

677 **Part V**

678 **Conclusion**

679 As search for the effects of dimension-six operators on $t\bar{t}H$ production is performed. An effective

680 field theory approached is used to parameterize the effects of high energy physics on the Higgs

681 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,

682 and the limits on dimension-six operators are limited to X.

683 List of contributions

684

685 Appendices

686 A Machine Learning Models

687 The following section provides details regarding the various machine learning models used in
688 the analysis. The Keras neural network framework, with Tensorflow as the backend, is used
689 to create each model, and the number of hidden layers and nodes are determined using grid
690 search optimization. For each model, a LeakyReLU activation function is used, along with a
691 learning rate of 0.01, and the Adam optimization algorithm. For the classification algorithms
692 (b-jet matching, Higgs reconstruction, and background separation) binary-cross entropy is used
693 as the loss function.

694 A.1 b-jet Identification Algorithms

695 A.1.1 2lSS Channel

696 For the 2lSS channel, the following input features as used for training:

697 **A.1.2 3l Channel**

698 **A.2 Alternate b-jet Identification Approach**

699 An alternative approach

700 **A.3 Higgs Reconstruction Algorithms**

701 **A.3.1 2lSS Channel**

702 **A.3.2 3l Semi-leptonic Channel**

703 **A.3.3 3l Fully-leptonic Channel**

704 **A.4 p_T Prediction MVA**

705 **A.4.1 2lSS Channel**

706 **A.4.2 3l Semi-leptonic Channel**

707 **A.4.3 3l Fully-leptonic Channel**

708 **A.5 3l Decay MVA**

709 **A.6 Background Rejection MVAs**

710 Separate mdoels are used in order to distinguish signal events from background for each analysis
711 channel - 2lSS, 3l semi-leptonic, and 3l fully leptonic. In particular, Neural Networks produced
712 with Tensorflow are trained using the kinematics of signal and background events derived from
713 Monte Carlo simulations. Further, because the background composition differs for events with

- 714 a high reconstructed Higgs p_T compared to events with low reconstructed Higgs p_T , separate
 715 MVAs are produced for high and low p_T regions.

716 **A.6.1 2lSS - High p_T**

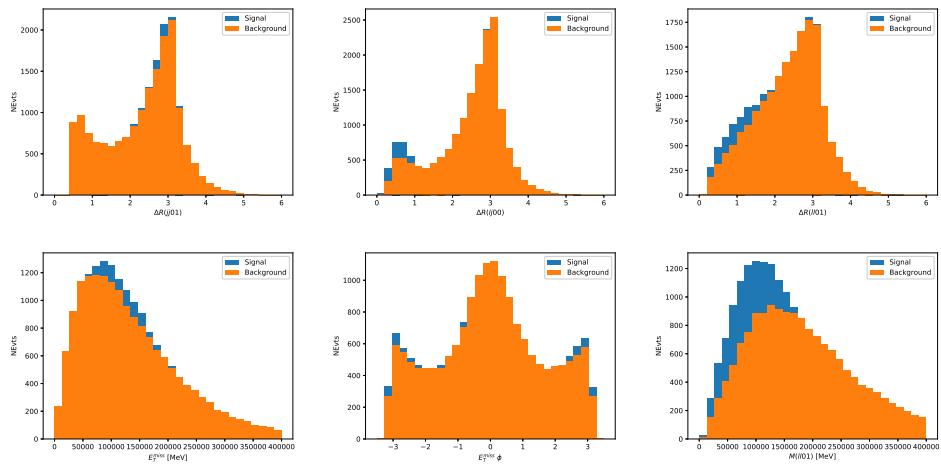


Figure A.1:

⁷¹⁷ **A.6.2 2lSS - Low p_T**

⁷¹⁸ **A.6.3 3l Semi-Leptonic - High p_T**

⁷¹⁹ **A.6.4 3l Semi-Leptonic - Low p_T**

⁷²⁰ **A.6.5 3l Fully Leptonic - High p_T**

⁷²¹ **A.6.6 3l Fully Leptonic - Low p_T**

⁷²² **B**