



ATLAS Note

GROUP-2017-XX

8th April 2021



A Deep Learning Approach to Differential Measurements of $t\bar{t}H$ Production in Multilepton Final States

The ATLAS Collaboration

The possibility of using the kinematic properties of the Higgs boson to search for new physics is investigated using $t\bar{t}H$ events with multiple leptons in the final state. Because of the challenges inherent to reconstructing the Higgs in multilepton final states, a deep learning approach is used to reconstruct the momentum spectrum of the Higgs, which would be altered by the presence of new physics without affecting the overall rate of $t\bar{t}H$ production. Simulations representing 139 fb^{-1} at $\sqrt{s} = 13 \text{ TeV}$ are used to provide estimates of the sensitivity to variations in the Higgs p_T spectrum.

¹³ © 2021 CERN for the benefit of the ATLAS Collaboration.

¹⁴ Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

15 **Contents**

16	1 Changes and outstanding items	4
17	1.1 Changelog	4
18	2 Introduction	5
19	3 Data and Monte Carlo Samples	5
20	3.1 Data Samples	6
21	3.2 Monte Carlo Samples	6
22	4 Object Reconstruction	8
23	4.1 Trigger Requirements	8
24	4.2 Light Leptons	9
25	4.3 Jets	10
26	4.4 Missing Transverse Energy	10
27	4.5 Overlap removal	10
28	5 Higgs Momentum Reconstruction	11
29	5.1 Decay Candidate Reconstruction	11
30	5.2 b-jet Identification	12
31	5.3 Higgs Reconstruction	19
32	5.4 p_T Prediction	32
33	5.5 3l Decay Mode	42
34	6 Signal Region Definitions	44
35	6.1 Pre-MVA Event Selection	44
36	6.2 Event MVA	49
37	6.3 Signal Region Definitions	57
38	7 Systematic Uncertainties	60
39	8 Results	65
40	8.1 Results - 80 fb^{-1}	65
41	8.2 Projected Results - 140 fb^{-1}	69
42	9 Conclusion	73
43	Appendices	76
44	A Machine Learning Models	76
45	A.1 Higgs Reconstruction Models	76
46	A.2 Background Rejection MVAs	99
47	A.3 Alternate b-jet Identification Algorithm	108

48	A.4 Binary Classification of the Higgs p_T	108
49	A.5 Impact of Alternative Jet Selection	110

⁵⁰ **1 Changes and outstanding items**

⁵¹ **1.1 Changelog**

⁵² This is version 1

53 2 Introduction

54 Since the discovery of a Higgs boson compatible with the Standard Model (SM) in 2012 [1], its
 55 interactions with other particles have been studied using proton-proton collision data produced by
 56 the Large Hadron Collider (LHC). The strongest of these interactions is the coupling of the Higgs
 57 to the top quark, making the Yukawa coupling between these two particles of particular interest
 58 for study.

59 These interactions can be measured directly by studying the production of a Higgs Boson in
 60 association with a pair of Top Quarks ($t\bar{t}H$). While this process has been observed by both the
 61 ATLAS and CMS collaborations, these analyses have focused on measuring the overall rate of $t\bar{t}H$
 62 production. There are several theories of physics Beyond the Standard Model (BSM), however,
 63 that would affect the kinematics of $t\bar{t}H$ production without altering its overall rate [2].

64 An Effective Field Theory approach can be used to model the low energy effects of new, high
 65 energy physics, by parameterizing BSM effects as dimension-six operators. The addition of these
 66 operators can be shown to modify the transverse momentum (p_T) spectrum of the Higgs Boson
 67 [3]. Therefore, reconstructing the momentum spectrum of the Higgs provides a means to observe
 68 new physics in the Higgs sector.

69 This note reports on the feasibility of measuring the impact of dimension-six operators in $t\bar{t}H$
 70 events with multiple leptons in the final state, using Monte Carlo (MC) simulations scaled to
 71 139 fb^{-1} at an energy $\sqrt{s} = 13 \text{ TeV}$. Events are separated into channels based on the number
 72 of light leptons (electrons and muons) in the final state - either two same-sign leptons (2lSS),
 73 or three leptons (3l). A deep neural network is used to identify which objects originate from
 74 the decay of the Higgs, and reconstruct the momentum of the Higgs Boson in each event. This
 75 reconstructed momentum spectrum is used to place limits on BSM effects, and on the parameters
 76 of dimension-six operators.

77 This note is organized as follows: The dataset and Monte Carlo (MC) simulations used in the
 78 analysis is outlined in Section 3. Section 4 describes the identification and reconstruction of the
 79 various physics objects. The models used to reconstruct the momentum spectrum of the Higgs
 80 is discussed in Section 5. The selection and categorisation of events comprises Section 6, and
 81 the theoretical and experimental systematic uncertainties considered are described in Section 7.
 82 Finally, the results of the study are summarized in Section 8.

83 3 Data and Monte Carlo Samples

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

-
- 89 • at least two light leptons within a range $|\eta| < 2.6$, with leading lepton $p_T > 15$ GeV and
90 subleading lepton $p_T > 5$ GeV
- 91 • at least one light lepton with $p_T > 15$ GeV within a range $|\eta| < 2.6$, and at least two hadronic
92 taus with $p_T > 15$ GeV.

93 Samples were then generated from these HIGG8D1 derivations using AnalysisBase version
94 21.2.127. A ptag of p4133 was used for MC samples, and p4134 for data.

95 **3.1 Data Samples**

96 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through
97 2018, which represents an integrated luminosity of 139 fb^{-1} and an energy of $\sqrt{s} = 13 \text{ TeV}$. All
98 data used in this analysis was included in one the following Good Run Lists:

- 99 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
100 _PHYS_StandardGRL_All_Good_25ns.xml
- 101 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
102 _PHYS_StandardGRL_All_Good_25ns.xml
- 103 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL
104 _All_Good_25ns_Triggerno17e33prim.xml
- 105 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL
106 _All_Good_25ns_Triggerno17e33prim.xml

107 **3.2 Monte Carlo Samples**

108 Several Monte Carlo (MC) generators were used to simulate both signal and background processes.
109 For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific event
110 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
t̄H	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)
t̄(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
t̄t	POWHEG-BOX v2 [powheggtt]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄γ	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
tZ	MG5_AMC	LO	PYTHIA 6	CTEQ6L1
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t, t̄t̄t̄	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

111 While the main t̄H analysis uses a more sophisticated data-driven approach to estimating the
 112 contribution of events with non-prompt leptons (or "fakes"), at the time of this note this strategy has
 113 not been completely developed for the full Run-2 dataset. Therefore, the non-prompt contribution
 114 is estimated with MC, while applying conservative systematic uncertainties to these processes, as
 115 described in Section 7.

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

Sample	DSID
tH	345873-5, 346343-5
VV	364250-364254, 364255, 363355-60, 364890
tW	410155
tZ	410156, 410157, 410218-20
low mass tZ	410276-8
Rare Top	410397, 410398, 410399
single Top	410658-9, 410644-5
three Top	304014
four Top	410080
tWW	410081
Z + jets	364100-41
low mass Z + jets	364198-215
W + jets	364156-97
V γ	364500-35
tZ	410560
tW	410013-4
WtZ	410408
VVV	364242-9
VH	342284-5
WtH	341998
t $\bar{t}\gamma$	410389
t \bar{t}	410470

Table 2: List of Monte Carlo samples by data set ID used in the analysis.

¹¹⁷ 4 Object Reconstruction

¹¹⁸ All analysis channels considered in this note share a common object selection for leptons and jets,
¹¹⁹ as well as a shared trigger selection.

¹²⁰ 4.1 Trigger Requirements

¹²¹ Events are required to be selected by dilepton triggers, as summarized in Table 3.

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

122 4.2 Light Leptons

123 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter
 124 that are associated with charged particle tracks reconstructed in the inner detector [4]. Electron
 125 candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the transition
 126 region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are
 127 rejected. A multivariate likelihood discriminant combining shower shape and track information
 128 is used to distinguish prompt electrons from nonprompt leptons, such as those originating from
 129 hadronic showers.

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

133 Muon candidates are reconstructed by combining inner detector tracks with track segments or full
 134 tracks in the muon spectrometer [5]. Muon candidates are required to have $p_T > 10$ GeV and
 135 $|\eta| < 2.5$. All leptons are required to be isolated, and pass a non-prompt BDT selection described
 136 in detail in [6].

137 **4.3 Jets**

138 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
139 calorimeters [7], using the anti- k_t algorithm with a radius parameter $R = 0.4$. Jets with energy
140 contributions likely arising from noise or detector effects are removed from consideration [8],
141 and only jets satisfying $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$ are used in this analysis. For jets with
142 $p_T < 60 \text{ GeV}$ and $|\eta| < 2.4$, a jet-track association algorithm is used to confirm that the jet
143 originates from the selected primary vertex, in order to reject jets arising from pileup collisions
144 [9].

145 **4.4 Missing Transverse Energy**

146 Because all $t\bar{t}H - \text{ML}$ channels considered include multiple neutrinos, missing transverse energy
147 (E_T^{miss}) is present in each event. The missing transverse momentum vector is defined as the
148 inverse of the sum of the transverse momenta of all reconstructed physics objects as well as
149 remaining unclustered energy, the latter of which is estimated from low- p_T tracks associated with
150 the primary vertex but not assigned to a hard object [10].

151 **4.5 Overlap removal**

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

158 This algorithm is applied to the preselected objects. The overlap removal procedure is summarized
159 in Table 4.

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

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

160 5 Higgs Momentum Reconstruction

161 Reconstructing the momentum of the Higgs boson is a particular challenge for channels with
 162 leptons in the final state: Because all channels include at least two neutrinos in the final state, the
 163 Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted
 164 by a neural network when provided with the four-vectors of the Higgs Boson decay products, as
 165 shown in Section 5.1. With this in mind, several layers of MVAs are used to reconstruction the
 166 Higgs momentum.

167 The first layer is a model designed to select which jets are most likely to be the b-jets that came
 168 from the top decay, detailed in Section 5.2. As described in Section 5.3, the kinematics of these
 169 jets are fed into the second layer, which is designed to identify the decay products of the Higgs
 170 Boson itself. The kinematics of these particles are then fed into yet another neural-network, which
 171 predicts the momentum of the Higgs (5.4). MVAs are also used in the analysis to determine the
 172 decay of the Higgs boson in the 3l channel (5.5).

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

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

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

188 5.1 Decay Candidate Reconstruction

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

194 Leptons considered as possible Higgs and top decay candidates are required to pass the selection
 195 described in Section 4.2. For jets, however, it is found that a large fraction that originate from either
 196 the top decay or the Higgs decay fall outside the selection described in Section 4.3. Specifically,

197 jets from the Higgs decay tend to be soft, with 32% having $p_T < 25$ GeV. Therefore jets with
 198 $p_T < 15$ GeV are considered as possible candidates in the models described below. By contrast,
 199 less than 5% of the jets originating from the Higgs fall below this p_T . The jets are found to be
 200 well modeled even down to this low p_T threshold, as shown in Section 6.1. The impact of using
 201 different p_T selection for the jet candidates is considered in detail in Section A.5. As they are
 202 expected to originate from the primary vertex, jets are also required to pass a JVT cut.

203 5.2 b-jet Identification

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

209 Exactly two b-jets are expected in the final state of $t\bar{t}H - ML$ events. However, in both the 3l and
 210 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r b-tag working
 211 point is used). Therefore, for events which have exactly one, or more than two, b-tagged jets,
 212 deciding which combination of jets correspond to the top decay is non-trivial. Further, events
 213 with 1 b-tagged jet represent just over half of all $t\bar{t}H - ML$ events. Of those, both b-jets are
 214 reconstructed by the detector 75% of the time. Therefore, rather than adjusting the selection to
 215 require exactly 2 b-tagged jets, and losing more than half of the signal events, a neural network is
 216 used to predict which pair of jets is most likely to correspond to truth b-jets.

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

219 5.2.1 2lSS Channel

220 For the 2lSS channel, the input features shown in Table 5 are used for training. Here j_0 and j_1 are
 221 the two jet candidates, while l_0 and l_1 are the two leptons in the event, both ordered by p_T . jet
 222 DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,
 223 where 5 represents the tightest working point and 1 represents the loosest. The variables nJets
 224 DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and
 225 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 5: Input features used in the b-jet identification algorithm for the 2lSS channel

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

230 The difference between the distributions for a few of these features for the correct(i.e. both jets
 231 are truth b-jets), and incorrectcombinations are shown in Figure 5.1. The correct and incorrect
 232 contributions are scaled to the same integral, so as to better demonstrate the differences in the
 233 distributions.

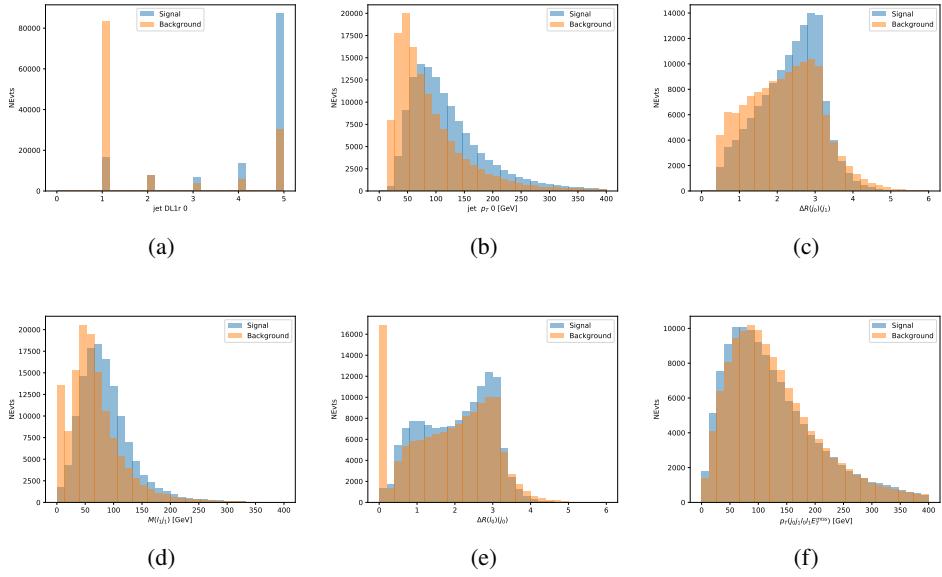


Figure 5.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T 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} .

234 The modeling of these inputs is validated against data, with Figure 5.2 showing good general
 235 agreement between data and MC. Plots for the complete list of features can be found in Section A.

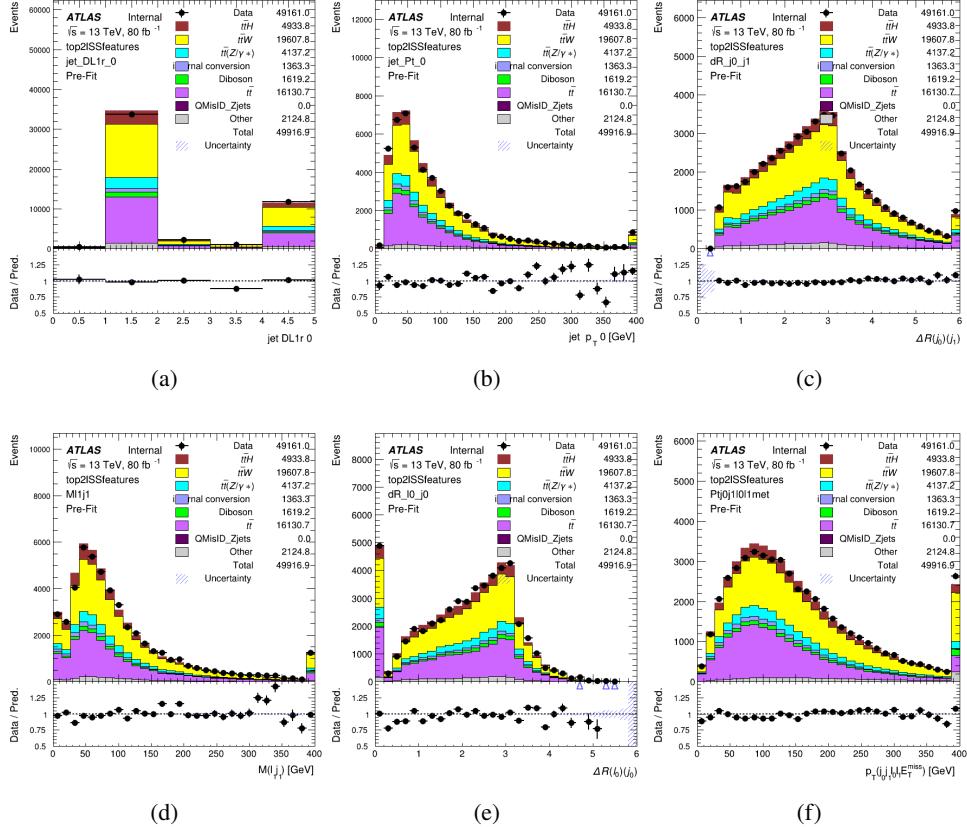


Figure 5.2: Data/MC comparisons of input features for top2lSS training for 80 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

Based on the results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as overfitting is found to not be an issue. The output score distribution as well as the ROC curve for the trained model are shown in Figure 5.2.1. The model is found to identify the correct pairing of jets for 73% of 2lSS signal events on test data.

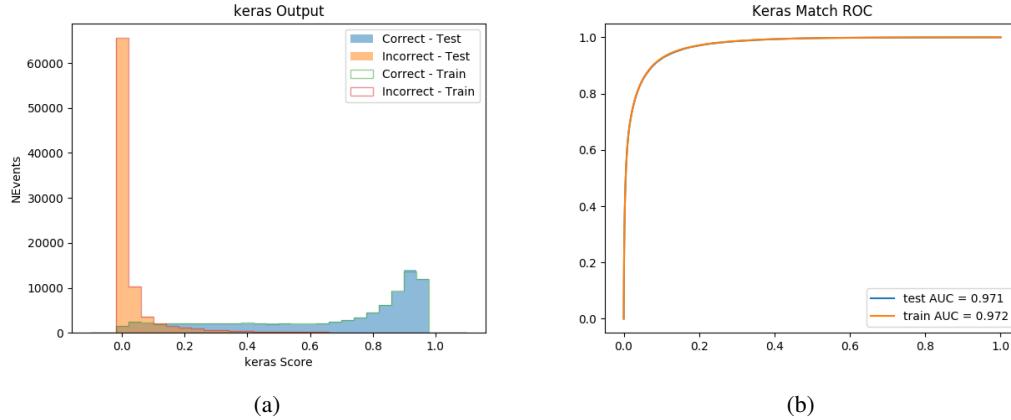


Figure 5.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

The accuracy of the model for different values of n-bjets, compared to this naive approach, is shown in Table 6.

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

Table 6: Accuracy of the NN in identifying b-jets from tops in 2lSS events for, compared to the accuracy of taking the two highest b-tagged jets.

5.2.2 3l Channel

The input features used in the 3l channel are listed in Table 7, with the same naming convention as the 2lSS channel.

jet p_T 0	jet p_T 1	jet η 0
jet η 1	Lepton p_T 0	Lepton p_T 1
Lepton p_T 2	$\Delta R(j_0)(j_1)$	$M(j_0 j_1)$
$\Delta R(l_0)(j_0)$	$\Delta R(l_1)(j_0)$	$\Delta R(l_2)(j_0)$
$\Delta R(l_0)(j_1)$	$\Delta R(l_1)(j_1)$	$\Delta R(l_2)(j_1)$
$M(l_0 j_0)$	$M(l_1 j_0)$	$M(l_2 j_0)$
$M(l_0 j_1)$	$M(l_1 j_1)$	$M(l_2 j_1)$
$\Delta R(j_0 l_0)(j_1 l_1)$	$\Delta R(j_0 l_0)(j_1 l_2)$	$\Delta R(j_0 l_1)(j_1 l_0)$
$\Delta R(j_0 l_2)(j_1 l_0)$	jet DL1r 0	jet DL1r 1
$p_T(j_0 j_1 l_0 l_1 l_2 E_T^{\text{miss}})$	$M(t j_0 j_1 l_0 l_1 l_2 E_T^{\text{miss}})$	$\Delta\phi(j_0)(E_T^{\text{miss}})$
$\Delta\phi(j_1)(E_T^{\text{miss}})$	HT Lepton	HT jets
nJets	E_T^{miss}	nJets OR DL1r 85
nJets OR DL1r 60		

Table 7: Input features for the b-jet identification algorithm in the 3l channel.

250 A few of these features are shown in Figure 5.4, comparing the distributions for correct and
251 incorrect combinations of jets.

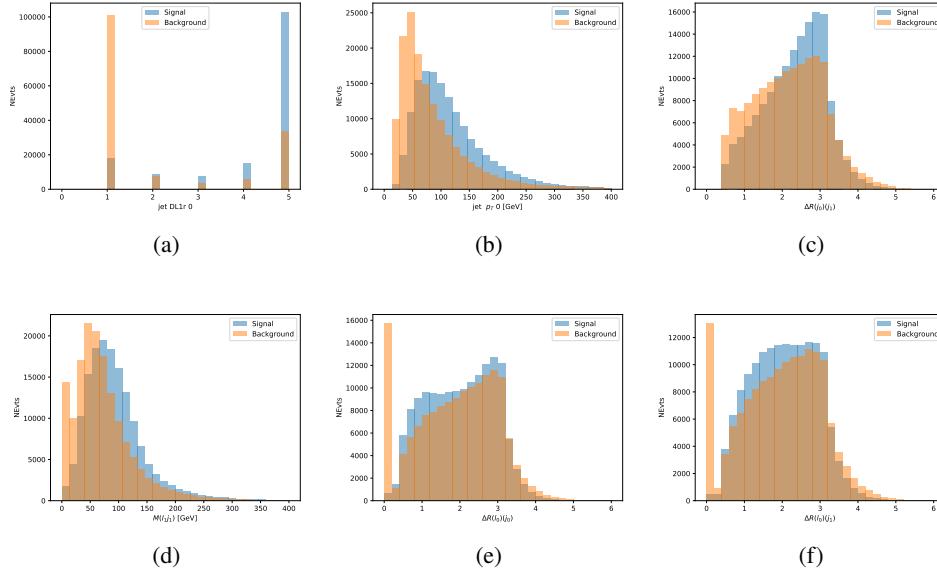


Figure 5.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

252 The modeling of these inputs is validated against data, with Figure 5.5 showing good general
 253 agreement between data and MC. Plots for the complete list of features can be found in Section A.

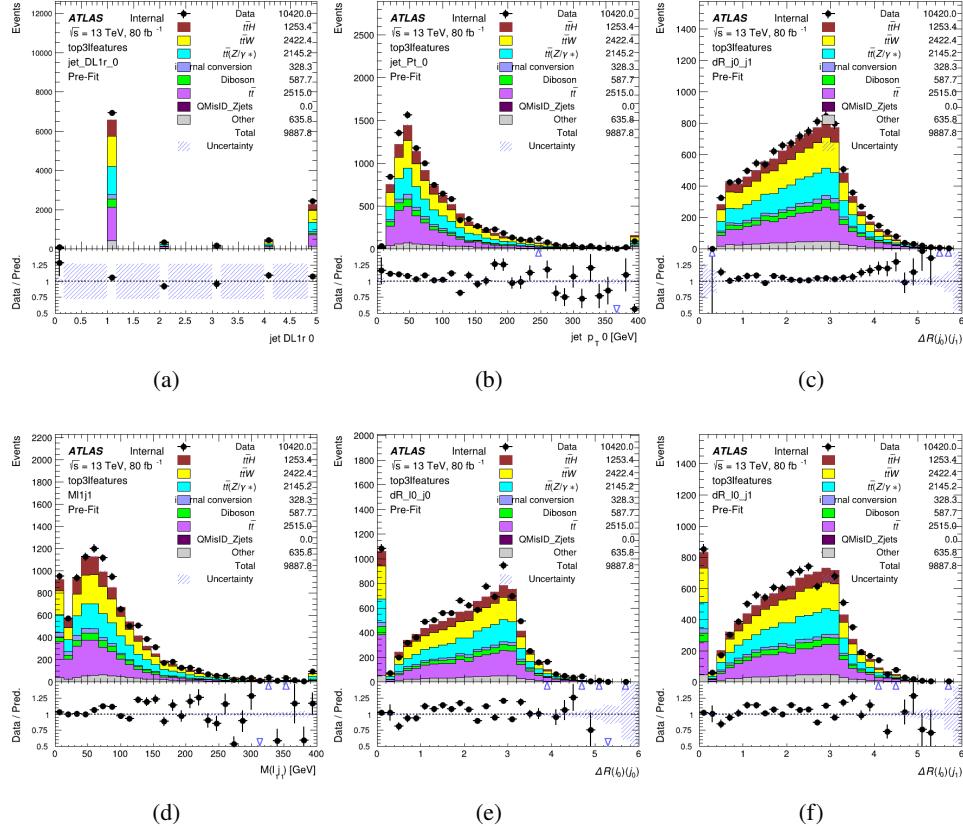


Figure 5.5: Data/MC comparisons of input features for top31 training for 80 fb^{-1} of data. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

254 Again, the dataset is downsized to reduce the ratio of correct and incorrect combination from 20:1,
 255 to 5:1. Around 7 million events are used for training, with 10% set aside for testing. Based on the
 256 results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with
 257 60 nodes each. The output score distribution as well as the ROC curve for the trained model are
 258 shown in Figure 5.2.2.

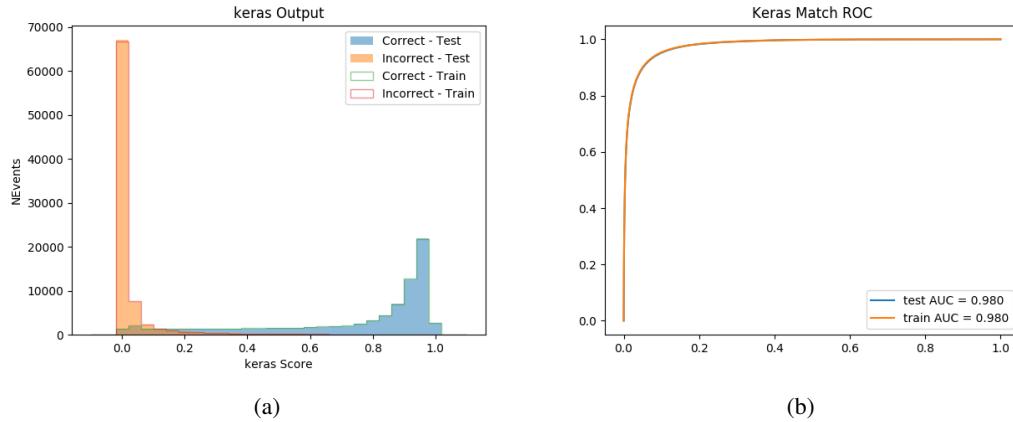


Figure 5.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

259 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal events.
 260 The accuracy of the model is summarized in Table 8.

Table 8: 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%

261 5.3 Higgs Reconstruction

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

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

269 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In the
 270 semi-leptonic case the reconstructed decay products consist of two jets and a single leptons. For

271 the fully-leptonic case, the decay products include 2 of the three leptons associated with the
272 event. For training the models, events are separated into these two categories using truth level
273 information. A separate MVA, described in Section 5.5, is used to make this distinction at reco
274 level and determine which model to use.

275 For all channels, the models described in Section 5.2 are used to identify b-jet candidates, whose
276 kinematics are used to identify the Higgs decay products. These jets are not considered as possible
277 candidates for the Higgs decay, justified by the fact that these models are found to misidentify jets
278 from the Higgs decay as jets from the top decay less than 1% of the time.

279 **5.3.1 2lSS Channel**

280 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural
281 network is trained on the kinematics of different combinations of leptons and jets, as well as the
282 b-jets identified in Section 5.2, with the specific input features listed in Table 9.

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

Table 9: Input features used to identify the Higgs decay products in 2lSS events

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

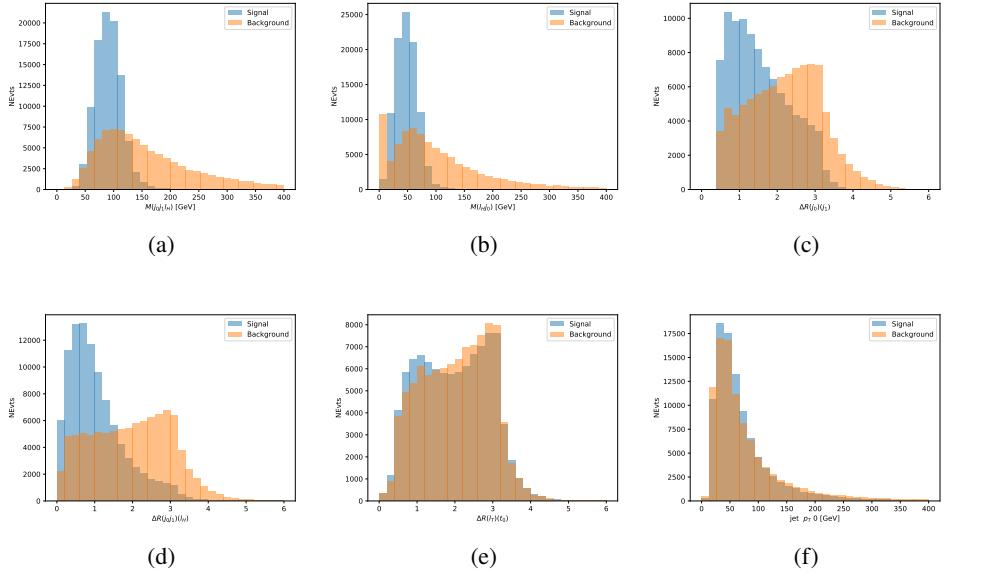


Figure 5.7: Input features for higgs2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 and the lepton candidate, (e) ΔR between the lepton from the top and the leading b-jet, (f) the p_T of jet 0.

287 The modeling of these inputs is validated against data, with Figure 5.2 showing good general
 288 agreement between data and MC. Plots for the complete list of features can be found in Section A.

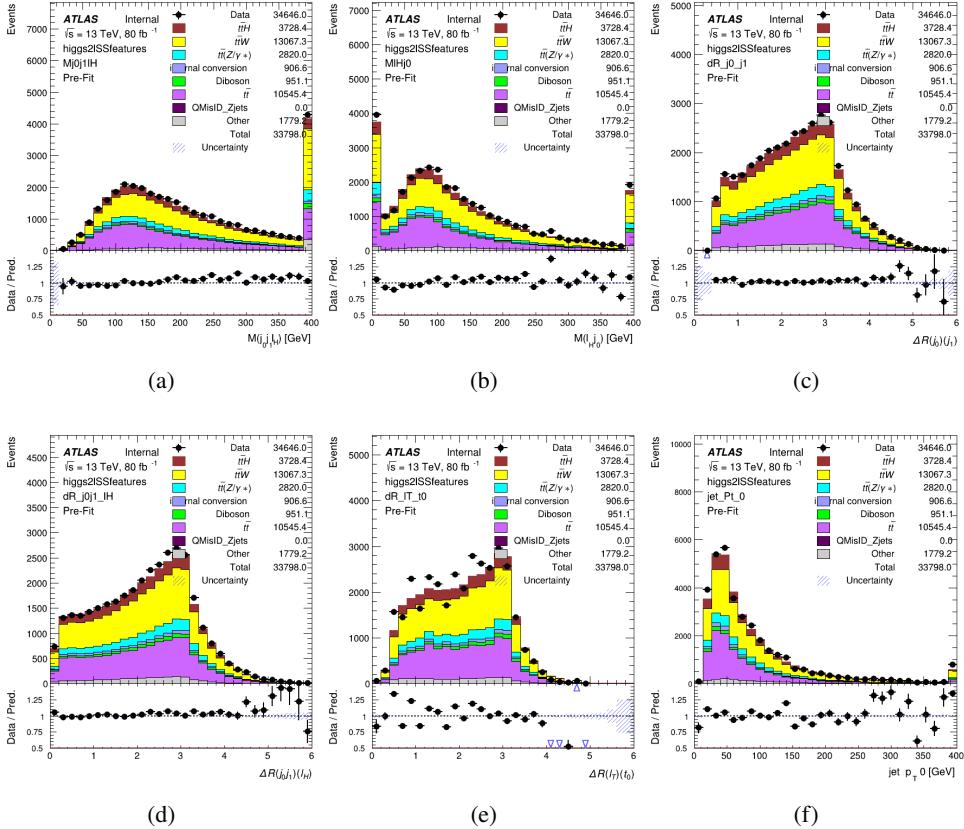


Figure 5.8: Data/MC comparisons of input features for higgs2ISS training for 80 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 and the lepton candidate, (e) ΔR between the lepton from the top and the leading b-jet, (f) the p_T of jet 0.

289 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2 million
 290 events, with an additional 200,000 reserved for testing the model. In order to compensate for
 291 large number of incorrect combinations, these have been downsampled such that the correct
 292 combinations represent over 10% of the training set. The output of the NN is summarized in
 293 Figure 5.3.1.

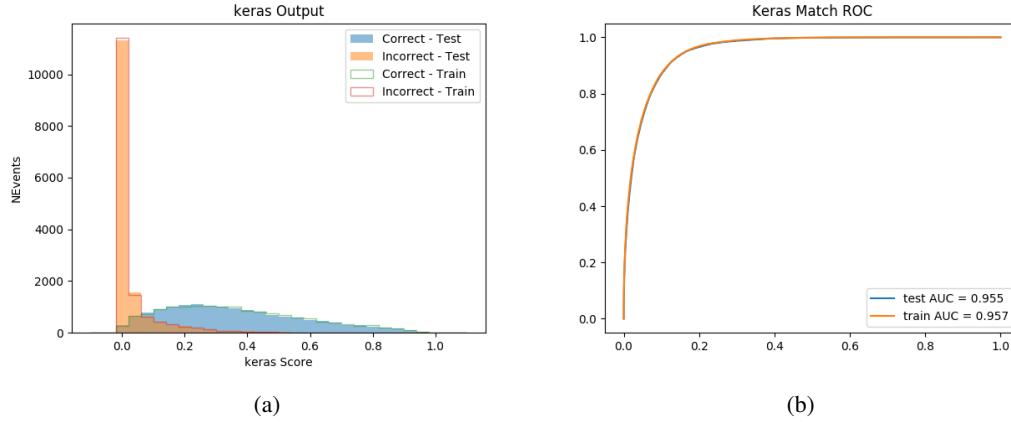


Figure 5.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

294 The neural network identifies the correct combination 55% of the time. It identifies the correct
 295 lepton 85% of the time, and selects the correct lepton and at least one of the correct jets 81% of
 296 the time.

297 **5.3.2 3l Semi-leptonic Channel**

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

Lepton p_T H	Lepton p_T T_0	Lepton p_T T_1
jet p_T 0	jet p_T 1	top p_T 0
top p_T 1	jet η 0	jet η 1
jet ϕ 0	jet ϕ 1	$\Delta R(j_0)(j_1)$
$M(j_0j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0j_1)(l_H)$	$\Delta R(j_0j_1)(l_{T_1})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0j_1l_{T_0})$	$M(j_0j_1l_{T_1})$
$M(j_0j_1l_H)$	$\Delta R(j_0j_1l_H)(l_{T_0})$	$\Delta R(j_0j_1l_H)(l_{T_1})$
$\Delta\phi(j_0j_1l_H)(E_T^{\text{miss}})$	$p_T(j_0j_1l_Hl_{T_0}l_{T_1}b_0b_1E_T^{\text{miss}})$	$M(j_0j_1b_0)$
$M(j_0j_1b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
topScore	MET	HT jets
nJets		

Table 10: Input features used to identify the Higgs decay products in 3lS events

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

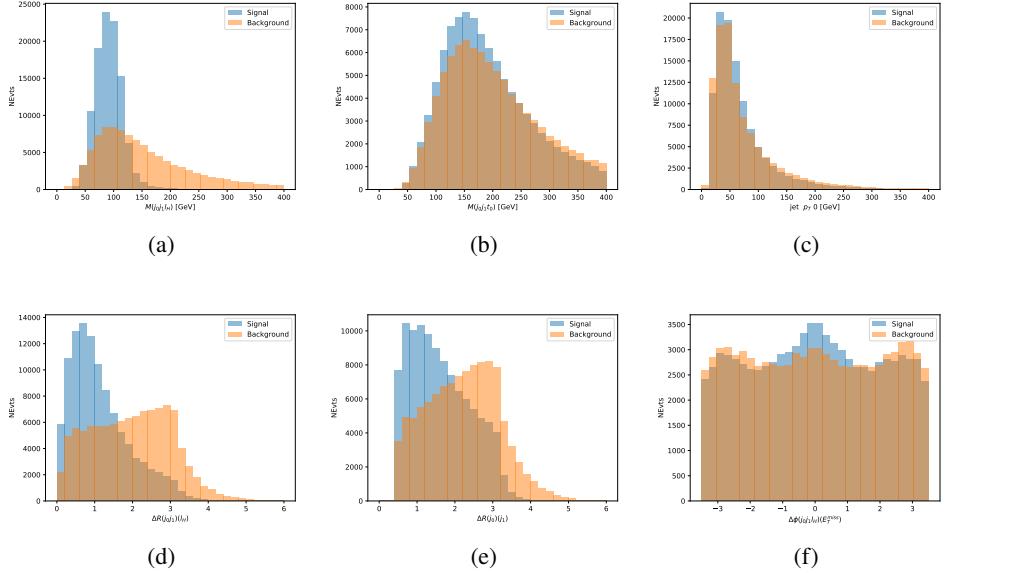


Figure 5.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

306 The modeling of these inputs is validated against data, with Figure 5.11 showing good general
 307 agreement between data and MC. Plots for the complete list of features can found in appendix
 308 [A.1](#).

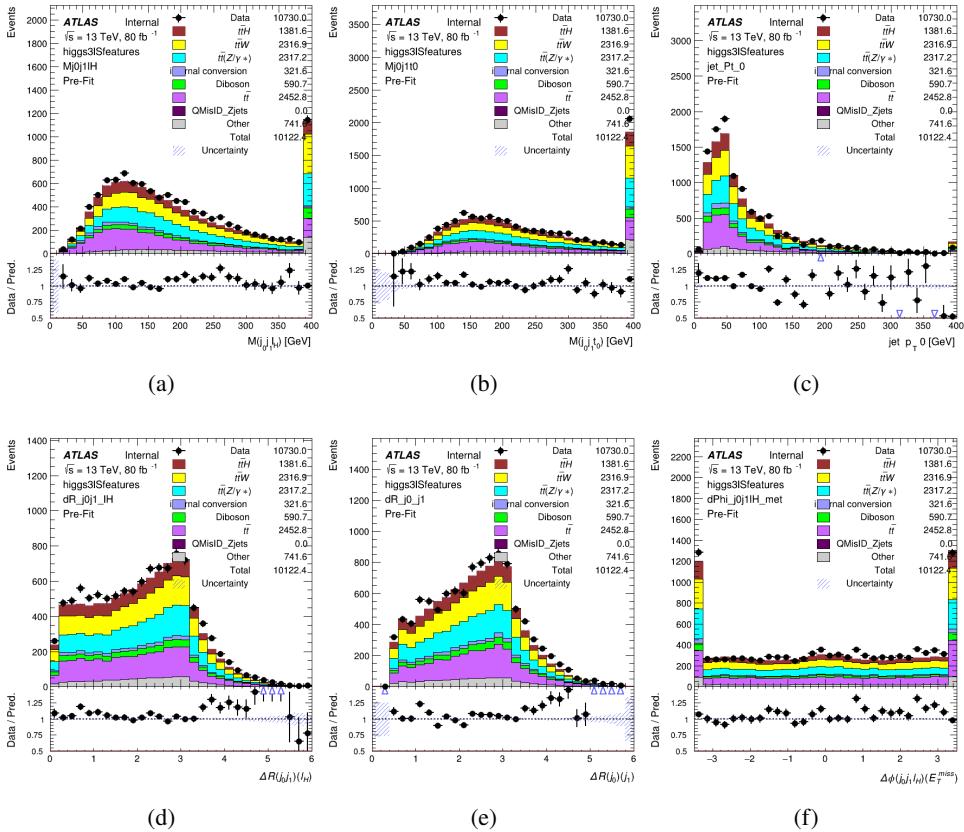


Figure 5.11: Data/MC comparisons of input features for higgs3IS training for 80 fb^{-1} of data.

309 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events. Once
 310 again, incorrect combinations are downsampled, such that the correct combinations are around
 311 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN is
 312 summarized in Figure 5.3.2.

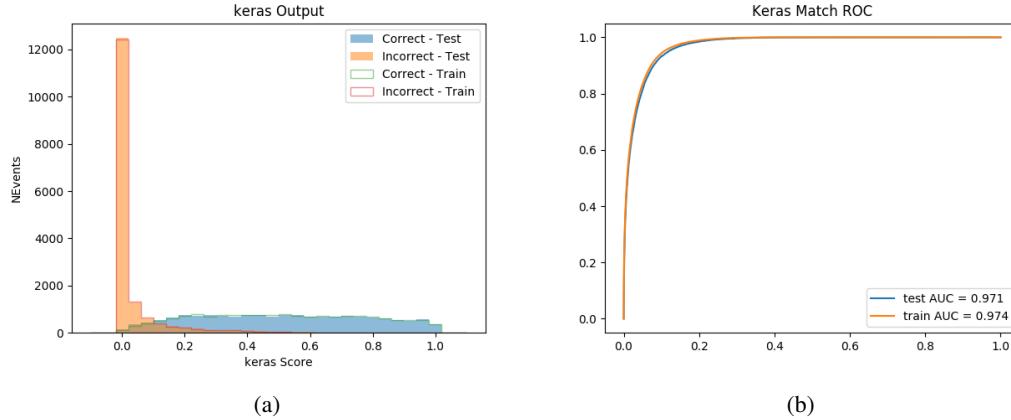


Figure 5.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

313 The neural network identifies the correct combination 64% of the time. It identifies the correct
 314 lepton 85% of the time, and selects the correct lepton and at least one of the correct jets 83% of
 315 the time.

316 **5.3.3 3l Fully-leptonic Channel**

317 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 318 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to
 319 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 320 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 321 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 11: Input features used to identify the Higgs decay products in 3lF events

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

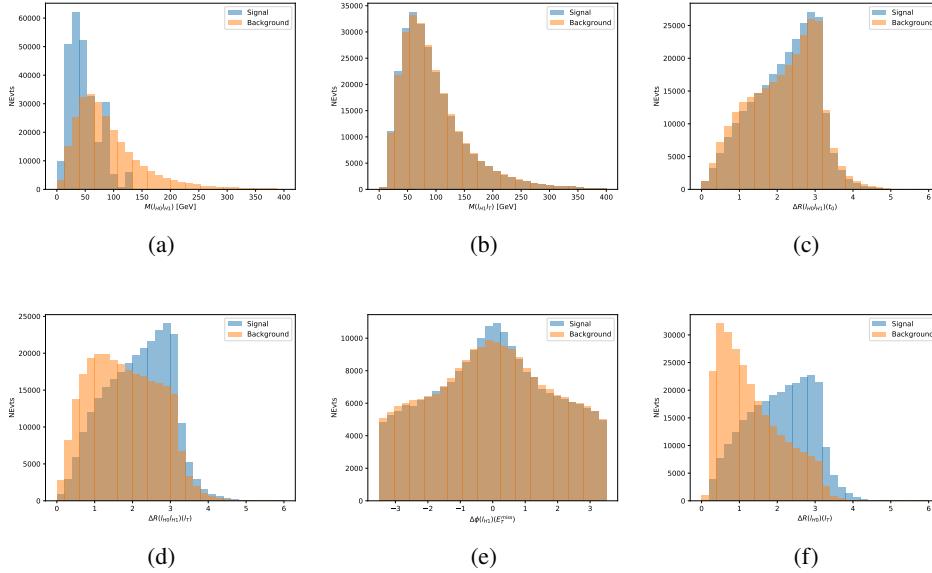


Figure 5.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

325 The modeling of these inputs is validated against data, with Figure 5.14 showing good general
 326 agreement between data and MC. Plots for the complete list of features can found in Section A.

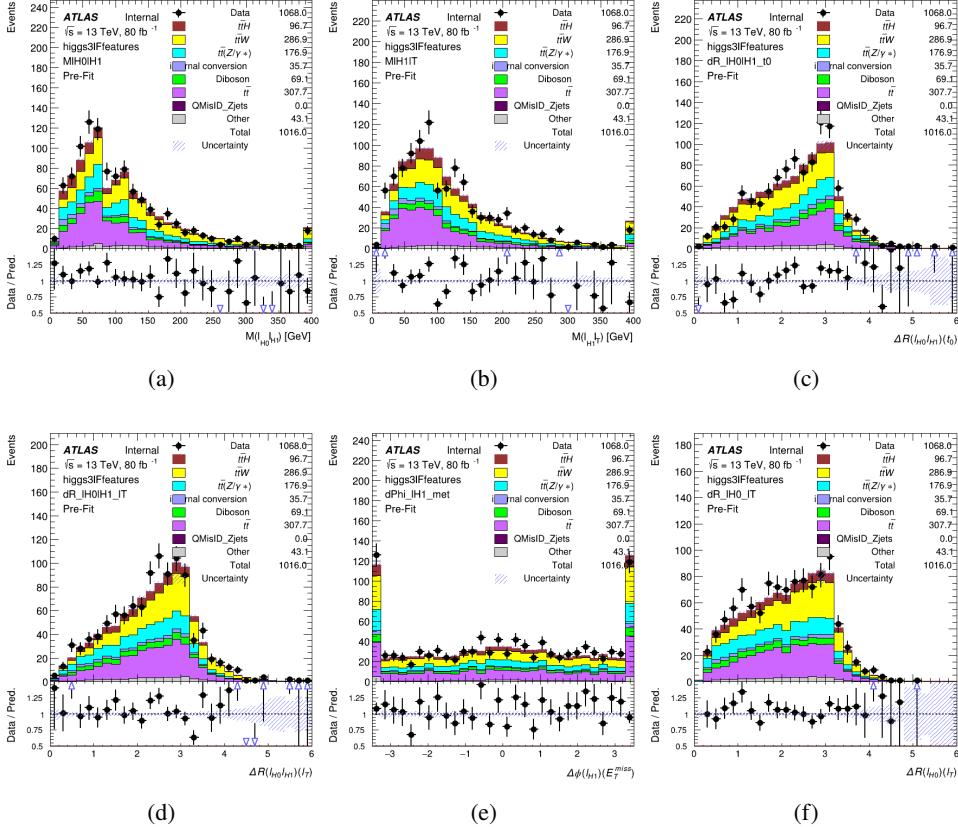


Figure 5.14: Data/MC comparisons of input features for higgs3IF training for 80 fb^{-1} of data.

327 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events, with
 328 10% of the dataset reserved for testing. The output of the model is summarized in Figure 5.3.3.

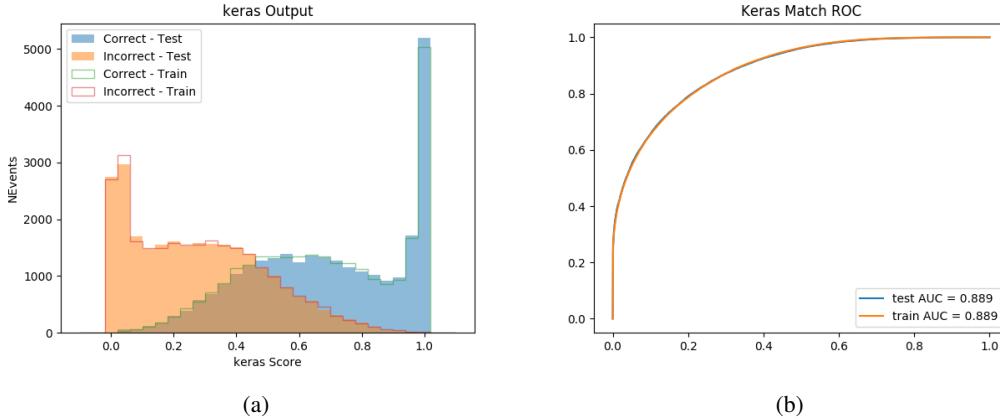


Figure 5.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

5.4 p_T Prediction

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

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

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

347 **5.4.1 2ISS Channel**

348 The input variables listed in Table 12 are used to predict the Higgs p_T in the 2ISS channel. Here
349 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
350 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
351 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
352 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
353 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 12: Input features for reconstructing the Higgs p_T spectrum for 2lSS events

354 The optimal neural network architecture for this channel is found to consist of 7 hidden layers with
 355 60 nodes each. The input data set includes 1.2 million events, 10% of which is used for testing,
 356 the other 90% for training. Training is found to converge after around 150 epochs.

357 To evaluate the performance of the model, the predicted p_T spectrum is compared to the truth
 358 Higgs p_T in Figure 5.16. In order to visualize the model performance more clearly, in (a) of that
 359 figure, the color of each point is determined by Kernel Density Estimation (KDE). The color
 360 shown represents the logarithm of the output from KDE, to counteract the large number of low
 361 p_T events. For that same reason, each column of the histogram shown in (b) of Figure 5.16 is
 362 normalized to unity. This plot therefore demonstrates what the model predicts for each slice of
 363 truth p_T .

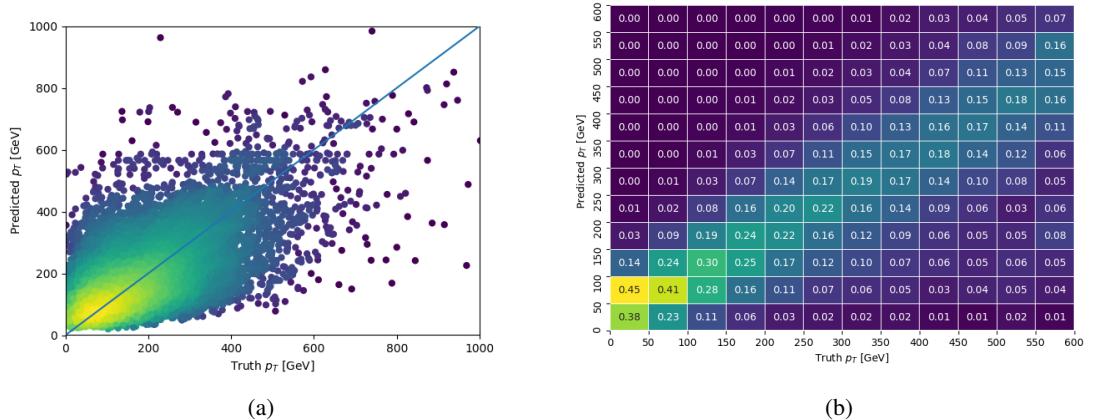


Figure 5.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

364 We are also interested in how well the model distinguishes between events with $p_T < 150$ GeV
 365 and > 150 GeV. Figure 5.17 demonstrates the NN output for high and low p_T events based on this
 366 cutoff.

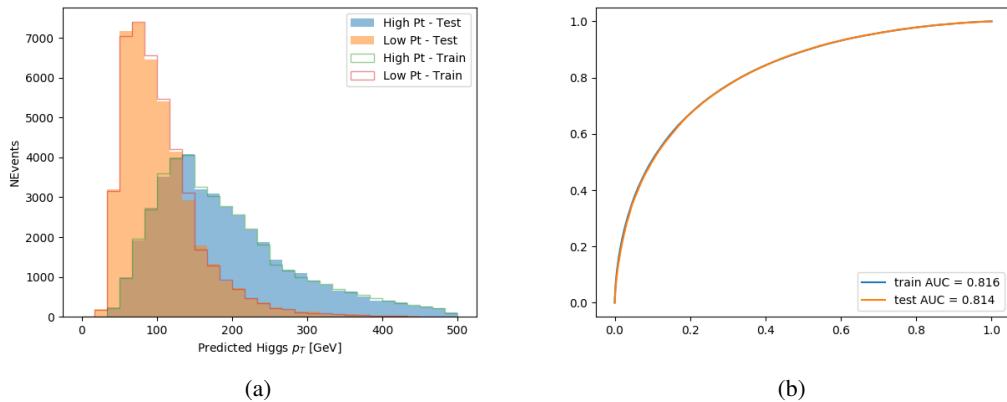


Figure 5.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

367 5.4.2 3l Semi-leptonic Channel

368 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 13: Input features for reconstructing the Higgs p_T spectrum for 3lS events

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

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

377 To evaluate the performance of the model, the predicted p_T spectrum is compared to the truth
 378 Higgs p_T in Figure 5.18. Once again, (a) of 5.18 shows a scatterplot of predicted vs truth p_T ,
 379 where the color of each point corresponds to the log of the relative KDE at that point. Each
 380 column of the histogram in (b) is normalized to unity, to better demonstrate the output of the
 381 NN for each slice of truth p_T .

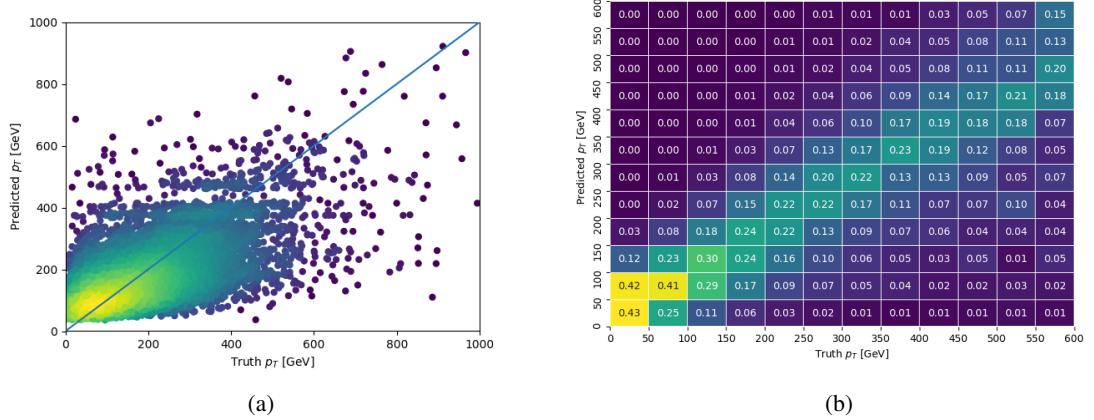


Figure 5.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in
 (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian
 Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

382 Figure 5.19 shows (a) the output of the NN for events with truth p_T less than and greater than
 383 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low p_T
 384 events.

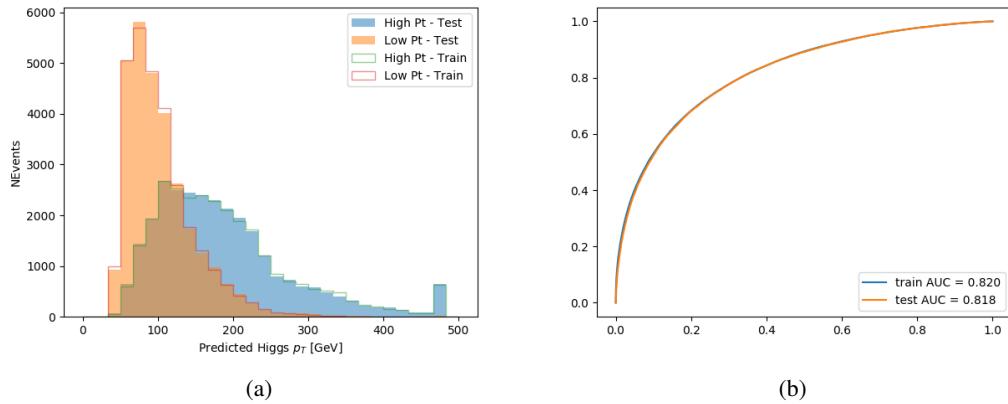


Figure 5.19: (

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.

385 **5.4.3 3l Fully-leptonic Channel**

386 The features listed in 14 are used to construct a model for predictin the Higgs p_T for 3lF events.

HT	$M(l_{H0}l_{H1})$	$M(l_{H0}l_T)$
$M(l_{H0}b_0)$	$M(l_{H0}b_1)$	$M(l_{H1}l_T)$
$M(l_{H1}b_0)$	$M(l_{H1}b_1)$	$\Delta R(l_{H0})(l_{H1})$
$\Delta R(l_{H0})(l_T)$	$\Delta R(l_{H0}l_{H1})(l_T)$	$\Delta R(l_{H0}l_T)(l_{H1})$
$\Delta R(l_{H1})(l_T)$	$\Delta R(l_{H0}b_0)$	$\Delta R(l_{H0}b_1)$
$\Delta R(l_{H1}b_1)$	$\Delta R(l_{H1}b_0)$	higgsScore
Lepton 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 14: Input features for reconstructing the Higgs p_T spectrum for 3lF events

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

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

394 The predicted transverse momentum, as a function of the truth p_T , is shown in Figure 5.20.

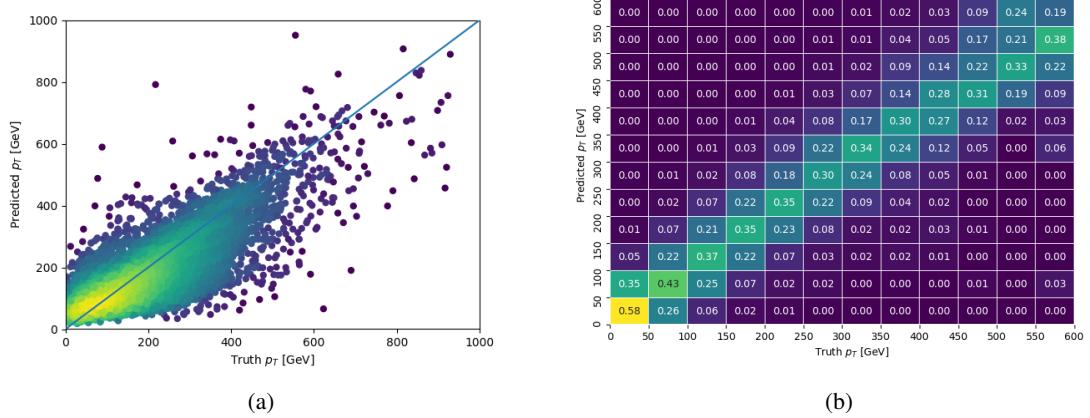


Figure 5.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

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

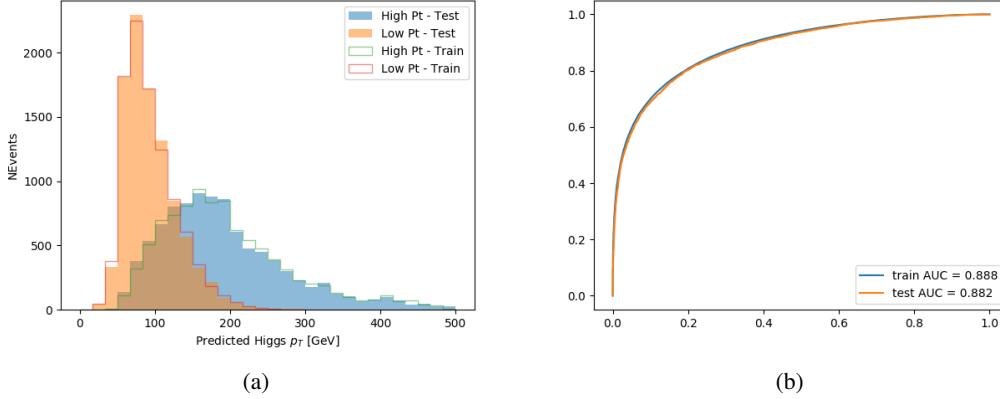


Figure 5.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

5.5 3l Decay Mode

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

The kinematics of each event, along with the output scores of the Higgs and top reconstruction algorithms, are used to distinguish these two possible decay modes. The particular inputs used are listed in Table 15.

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

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

409 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these two
 410 decay modes. The output of the model is summarized in Figure 5.22.

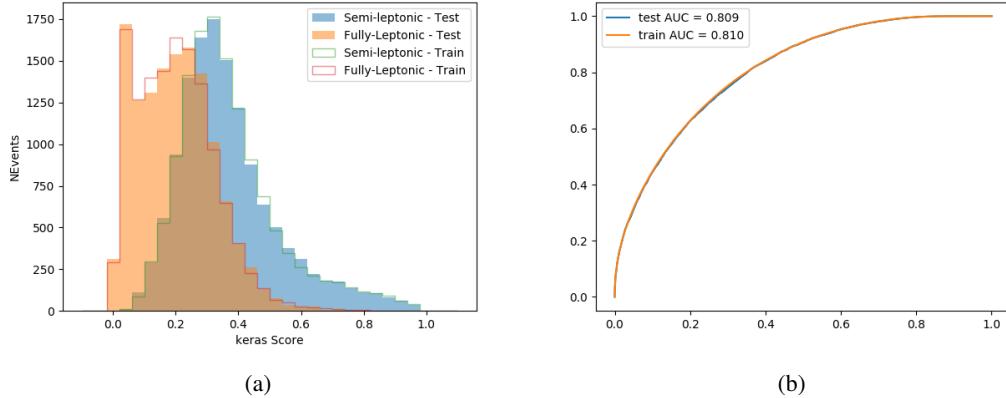


Figure 5.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

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

412 6 Signal Region Definitions

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

416 6.1 Pre-MVA Event Selection

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

- 419 • Two very tight, same-charge, light leptons with $p_T > 20$ GeV
- 420 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets
- 421 • No reconstructed tau candidates

⁴²² The event yield after the 2lSS preselection has been applied, for MC and data at 80 fb^{-1} , is shown
⁴²³ in Table 16.

	Yields
t̄H high p _T	36.19 ± 0.23
t̄H low p _T	63.58 ± 0.31
t̄W	440.64 ± 2.32
t̄Z/γ	91.84 ± 0.79
t̄lllowmass	8.47 ± 0.28
rareTop	24.2099 ± 0.40
VV	38.7927 ± 0.55
tZ	3e-05 ± 5.47-06
QMISID t̄	39.90 ± 2.36
QMISID Zjets	5.49 ± 0.67
t̄ int. conv.	12.74 ± 1.40
t̄ + γ int. conv.	12.09 ± 0.58
t̄ Conv.	13.55 ± 1.43
t̄ + γ Conv.	5.35 ± 0.38
t̄ HF e	59.92 ± 2.89
t̄ + γ HF e	0.51 ± 0.15
t̄ HF μ	224.57 ± 5.62
t̄ + γ HF μ	1.60 ± 0.23
Z + jets internal conv	3e-05 ± 5.47e-06
Z + jets conv	0.62 ± 0.21
Z + jets HF e	0.14 ± 0.13
Z + jets HF μ	0.82 ± 0.26
Single top Conv	2.27 ± 0.53
Single top HF e	2.33 ± 0.50
Single top HF μ	11.12 ± 1.07
Three top	2.22 ± 0.02
Four top	13.09 ± 0.16
t̄WW	10.985 ± 0.30
tW	3e-05 ± 5.47-06
WtZ	9.07 ± 0.44
VVV	0.30 ± 0.04
VH	0.59 ± 1.55
Total	1133.11 ± 7.69
Data	1108

Table 16: Event yield in the 2lSS preselection region.

424 Figure 6.1. Good general agreement is found.

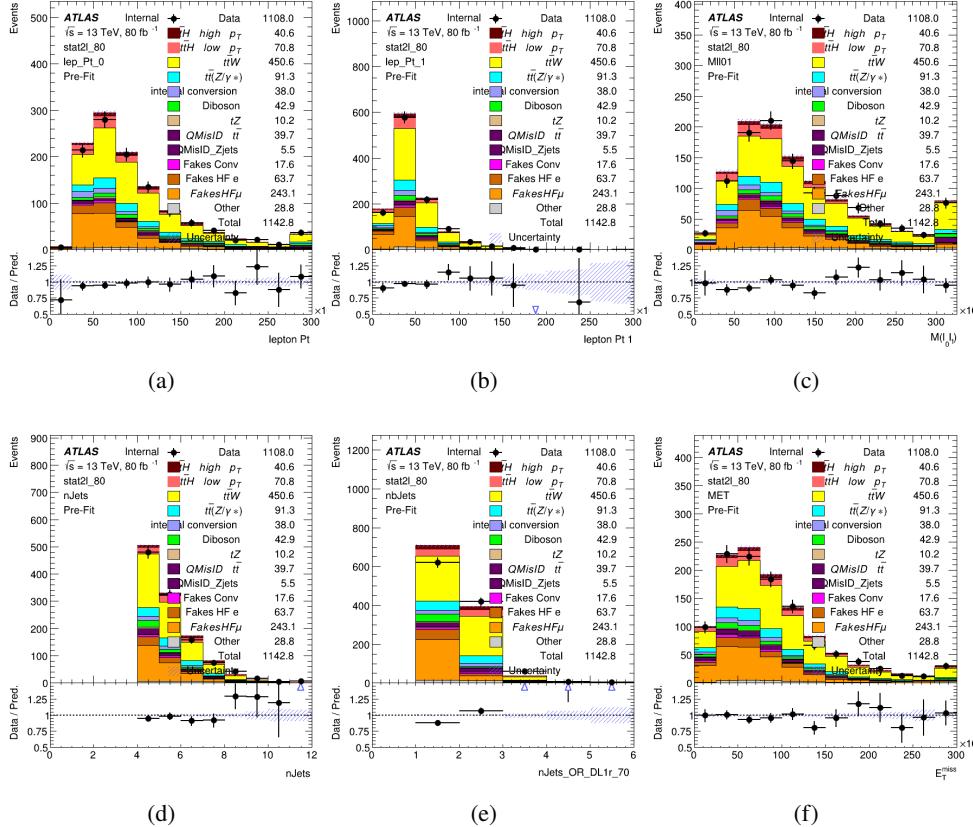


Figure 6.1: Data/MC comparisons of the 2lSS pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

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

- 426 • Three light leptons with total charge ± 1
- 427 • Same charge leptons are required to be very tight, with $p_T > 20 \text{ GeV}$
- 428 • Opposite charge lepton must be loose, with $p_T > 10 \text{ GeV}$
- 429 • ≥ 2 reconstructed jets, ≥ 1 b-tagged jets
- 430 • No reconstructed tau candidates
- 431 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs

432 The event yield after the 3l preselection has been applied, for MC and data at 80 fb^{-1} , is shown
 433 in Table 6.1.

	Yields
t̄H high p _T	18.40 ± 0.13
t̄H low p _T	29.91 ± 0.16
t̄W	134.22 ± 1.25
t̄Z/γ	88.47 ± 0.73
t̄lllowmass	2.77 ± 0.16
rareTop	15.05 ± 0.32
VV	34.54 ± 0.54
tZ	2e-05 ± 4.47-06
QMisID t̄t	1.80 ± 0.59
QMisID Zjets	0.02 ± 0.02
t̄t internal conversion	4.34 ± 0.43
t̄t + γ internal conversion	5.83 ± 0.42
t̄t Conv.	4.71 ± 0.45
t̄t + γ Conv.	2.64 ± 0.27
t̄t HF e	27.44 ± 1.05
t̄t + γ HF e	0.27 ± 0.11
t̄t HF μ	89.21 ± 1.92
t̄t + γ HF μ	0.94 ± 0.16
Z + jets conv	0.09 ± 0.19
Z + jets HF e	0.25 ± 0.15
Z + jets HF μ	2.41 ± 0.95
Single top Conv	0.58 ± 0.61
Single top HF e	1.50 ± 0.43
Single top HF μ	4.62 ± 0.85
Three top	0.96 ± 0.02
Four top	5.58 ± 0.10
t̄WW	5.45 ± 0.21
WtZ	8.71 ± 0.42
VVV	0.81 ± 0.02
Total	492.14 ± 3.22
Data	535

Table 17: Yields of the analysis

434 Comparisons of kinematic distributions for data and MC in this region are shown in Figure 6.2.

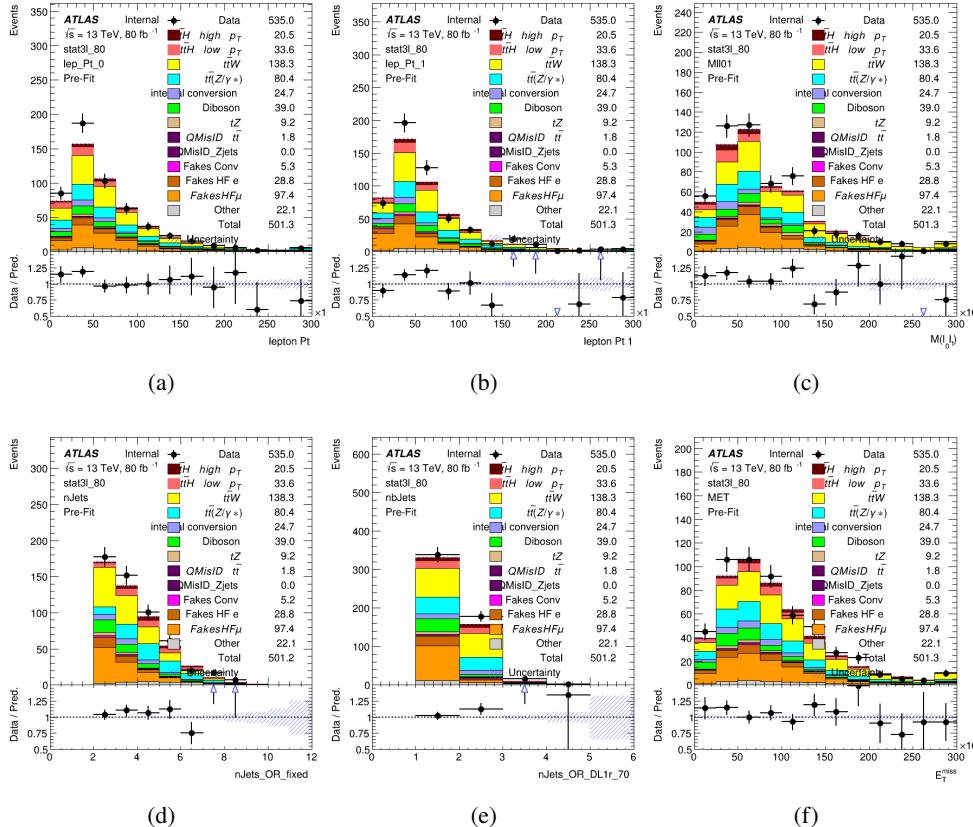


Figure 6.2: Data/MC comparisons of the 31 pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

435 **6.2 Event MVA**

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

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

⁴⁴⁵ background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
⁴⁴⁶ 3lF high p_T , and 3lF low p_T .

⁴⁴⁷ 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.

⁴⁴⁸ 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.

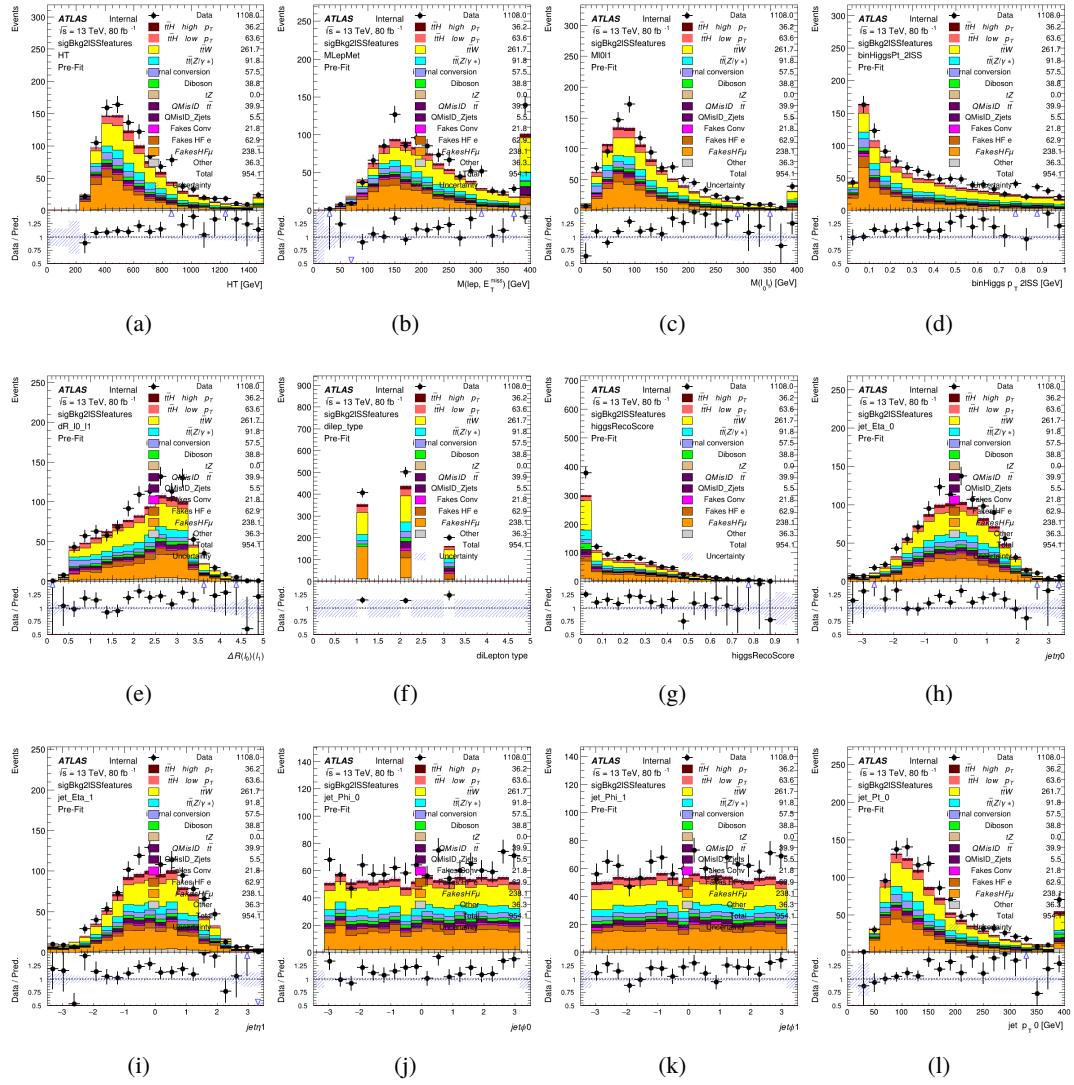


Figure 6.3:

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

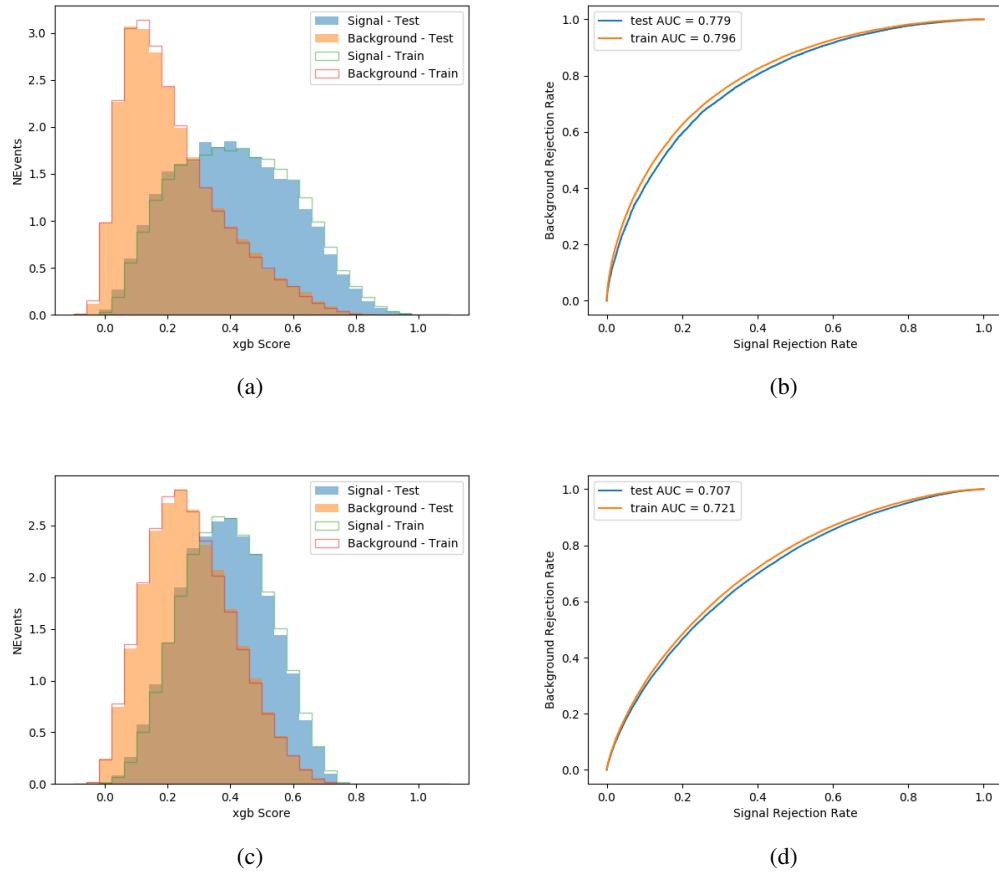


Figure 6.4:

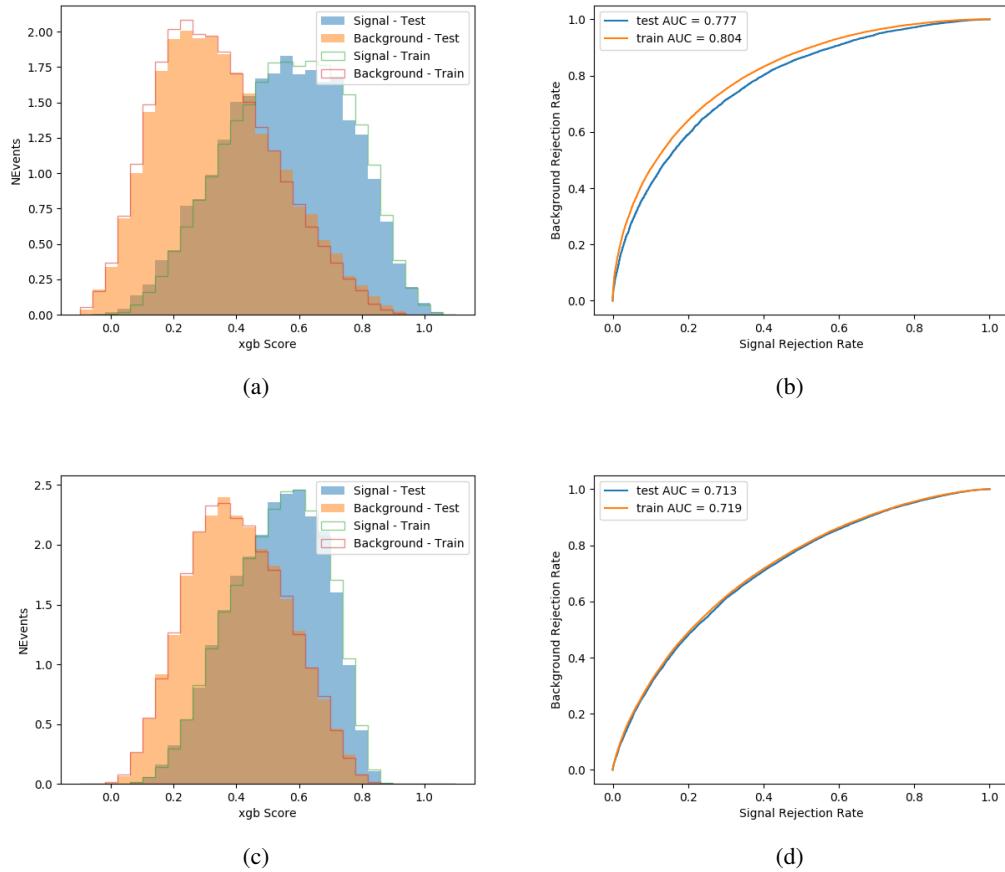


Figure 6.5:

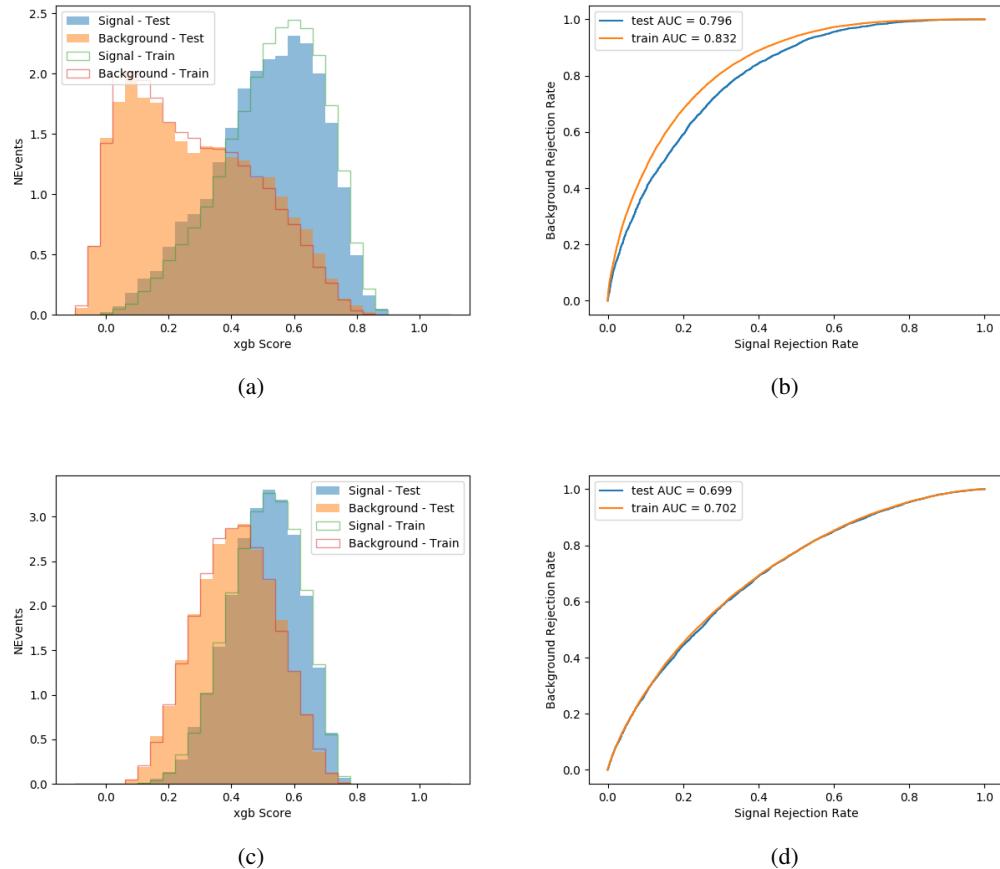


Figure 6.6:

450 Output distributions of each MVA comparing MC prediction to data at 80 fb^{-1} are shown in
 451 figures 6.7-6.2.

452 6.3 Signal Region Definitions

453 Once pre-selection has been applied, channels are further refined based on the MVAs described
 454 above. The output of the model described in Section 5.5 is used to separate the three channel into
 455 two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson.
 456 This leaves three orthogonal signal regions - 2ISS, 3IS, and 3IF.

457 For each event, depending on the number of leptons as well as whether the p_T of the Higgs is
 458 predicted to be high ($> 150 \text{ GeV}$) or low ($< 150 \text{ GeV}$), a cut on the appropriate background
 459 rejection MVA is applied. The particular cut values, listed in Table 20, are determined by
 460 maximizing S/\sqrt{B} in each region.

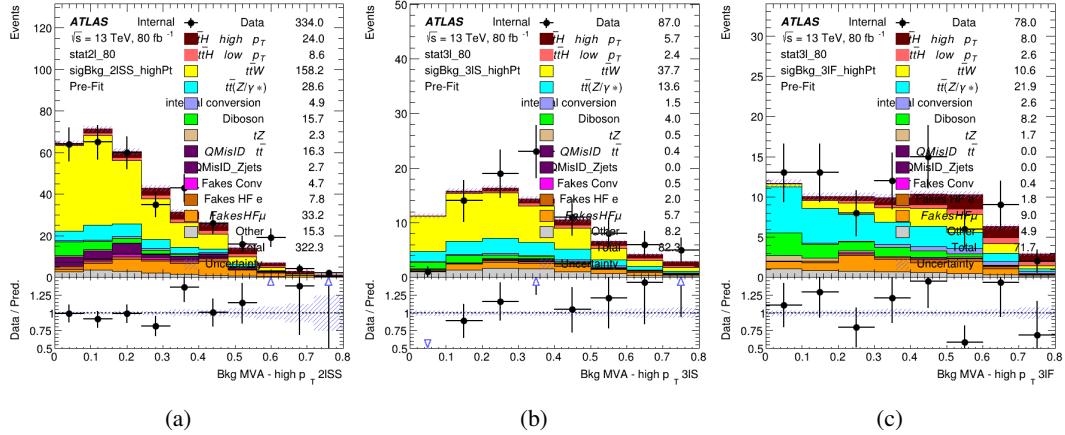


Figure 6.7: Output score of the high p_T BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channels

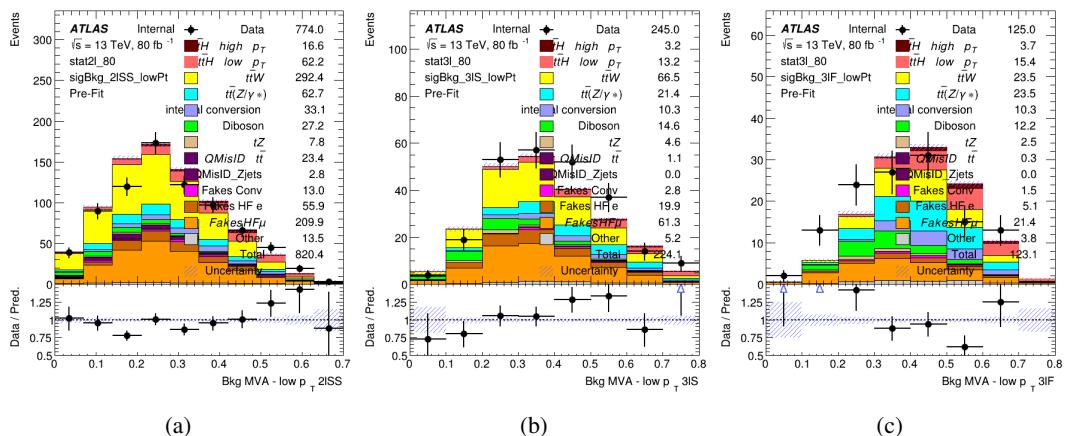


Figure 6.8: Output score of the low p_T BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channels

Channel	BDT Score
2ISS high p_T	0.36
2ISS low p_T	0.34
3IS high p_T	0.51
3IS low p_T	0.43
3IF high p_T	0.33
3IF low p_T	0.41

Table 20: Cutoff values on background rejection MVA score applied to signal regions.

The event preselection and MVA selection define the three signal regions. These signal region definitions are summarized in Table 21.

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

463 7 Systematic Uncertainties

464 The systematic uncertainties that are considered are summarized in Table 22. These are
 465 implemented in the fit either as a normalization factors or as a shape variation or both in the signal
 466 and background estimations. The numerical impact of each of these uncertainties is outlined in
 467 section 8.

Table 22: 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

468 The uncertainty in the combined 2015+2016 integrated luminosity is derived from a calibration
 469 of the luminosity scale using x-y beam-separation scans performed in August 2015 and May 2016
 470 [11].

471 The experimental uncertainties are related to the reconstruction and identification of light leptons
 472 and b-tagging of jets, and to the reconstruction of E_T^{miss} . The TOTAL electron ID correlation

473 model is used, corresponding to 1 electron ID systematic. Electron ID is found to be a subleading
474 systematic that is unconstrained by the fit, making it an appropriate choice for this analysis.

475 The sources which contribute to the uncertainty in the jet energy scale [12] are decomposed into
476 uncorrelated components and treated as independent sources in the analysis. The CategoryReduction
477 model is used to account for JES uncertainties, which decomposes the uncertainties into 30
478 nuisance parameters included in the fit. The SimpleJER model is used to account for jet energy
479 resolution (JER) uncertainties, and 8 JER uncertainty components unclued as NPs in the fit.

480 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses [13] are
481 also decomposed into uncorrelated components. The large number of components for b-tagging
482 is due to the calibration of the distribution of the BDT discriminant.

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

485 The full list of systematic uncertainties considered in the analysis is summarized in Tables 23, 24
486 and 25.

487

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

Table 23: Summary of experimental systematics considered for leptons and E_T^{miss} . Includes type, description, name of systematic as used in the fit, and mode of application. The mode of application indicates the systematic evaluation, e.g. as an overall event re-weighting (Event Weight) or rescaling (p_T Correction).

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

Table 24: Jet systematics take into account effects of jets calibration method, η inter-calibration, high p_T jets, pile-up, and flavor response. They are all diagonalised into effective parameters.

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

Table 25: Summary of experimental systematics to be included for b-tagging of jets in the analysis, using the continuous DL1r tagging algorithm. All of the b-tagging related systematics are applied as event weights. From left: type, description, and the name of systematic used in the fit.

488 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and scale
 489 uncertainties are taken from theory calculations. The theory uncertainties applied to the
 490 predominate background estimates are summarized in Table 26.

Process	X-section [%]
t̄t H (aMC@NLO+Pythia8)	QCD Scale: $^{+5.8}_{-9.2}$ PDF($+\alpha_S$): ± 3.6
t̄t Z (aMC@NLO+Pythia8)	QCD Scale: $^{+9.6}_{-11.3}$ PDF($+\alpha_S$): ± 4
t̄t W (aMC@NLO+Pythia8)	QCD Scale: $^{+12.9}_{-11.5}$ PDF($+\alpha_S$): ± 3.4
tHjb (aMC@NLO+Pythia8)	QCD Scale: $^{+6.5}_{-14.9}$ PDF($+\alpha_S$): ± 3.7
WtH (aMC@NLO+Pythia8)	QCD Scale: $^{+5.0}_{-6.7}$ PDF($+\alpha_S$): ± 6.3
VV (Sherpa 2.2.1)	± 50
t̄t	± 20
tZ	$\pm X$
c	
Z + jets	± 25
Others	± 50

Table 26: Summary of theoretical uncertainties for MC predictions in the analysis.

491 8 Results

- 492 A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p_T spectrum
 493 in the three signal regions, 2I Σ , 3I Σ , and 3I F . The signal is split into high and low p_T samples,
 494 based on whether the truth p_T of the Higgs is above or below 150 GeV. The parameters $\mu_{t\bar{t}H\text{high}p_T}$
 495 and $\mu_{t\bar{t}H\text{low}p_T}$, where $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$, are extracted from the fit, signifying the difference
 496 between the observed value and the theory prediction. Unblinded results are shown for the 80
 497 fb^{-1} data set, as well as MC only projections of results using the full Run-2, 140 fb^{-1} dataset.
 498 As described in Section 7, there are 229 systematic uncertainties that are considered as NPs in
 499 the fit. These NPs are constrained by Gaussian or log-normal probability density functions. The
 500 latter are used for normalisation factors to ensure that they are always positive. The expected
 501 number of signal and background events are functions of the likelihood. The prior for each NP is
 502 added as a penalty term, decreasing the likelihood as it is shifted away from its nominal value.

503 8.1 Results - 80 fb^{-1}

- 504 As the data collected from 2015-2017 has been unblinded for $t\bar{t}H - \text{ML}$ channels, representing 80
 505 fb^{-1} , those events are unblinded. The predicted Higgs p_T spectrum is fit to data simultaneously
 506 in each of the three signal regions shown in Figure 8.1.

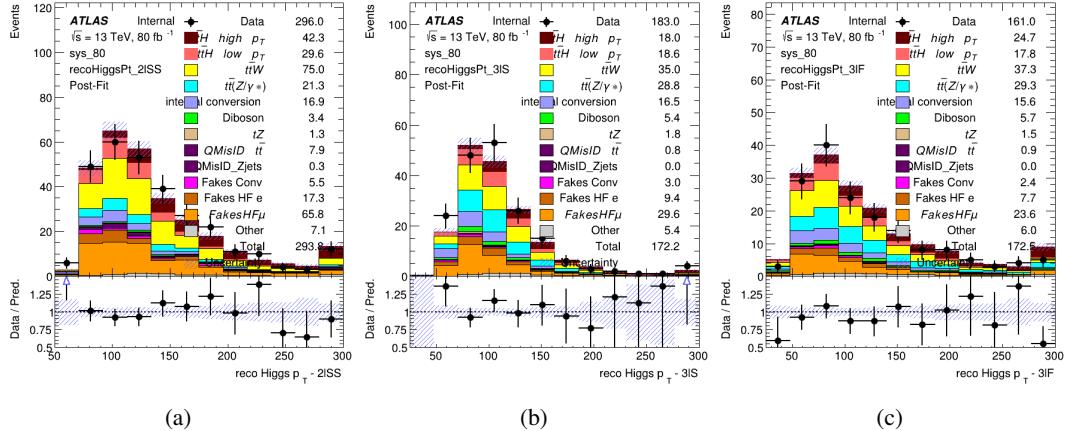


Figure 8.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2I Σ , (b) 3I Σ , and (c) 3I F , for 80 fb^{-1} of MC

- 507 A post-fit summary of the fitted regions is shown in Figure 8.2.

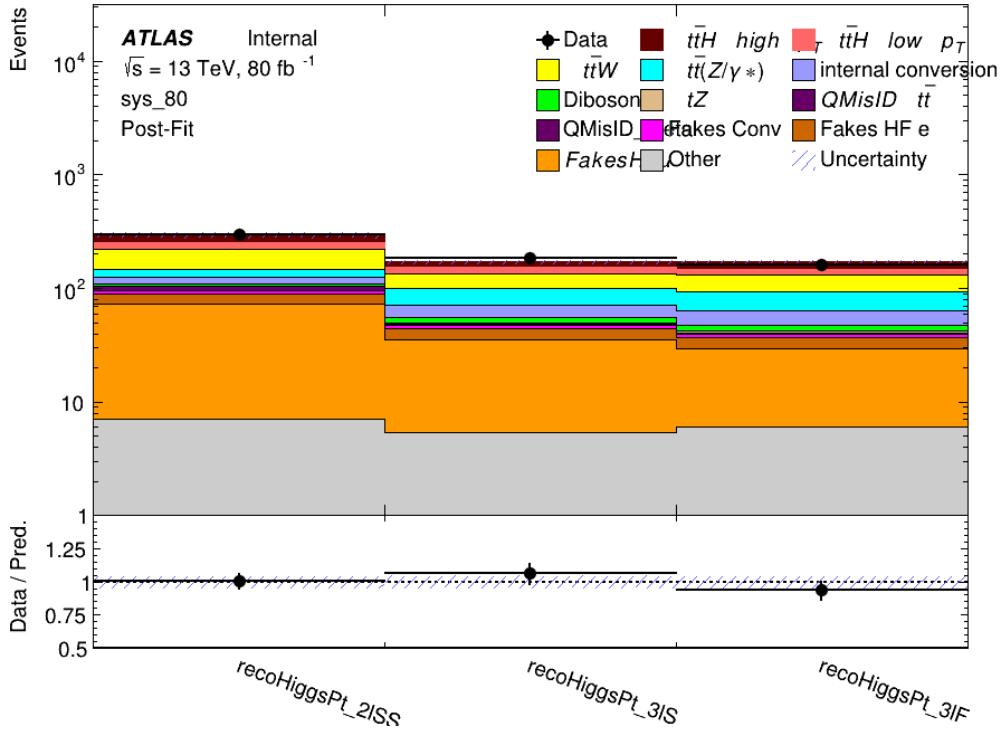


Figure 8.2: Post-fit summary of the yields in each signal region.

508 The measured μ values for high and low p_T Higgs production obtained from the fit are
 509 shown in 27. A significance of 1.7σ is observed for $t\bar{t}H$ high p_T , and 2.1σ is measured for
 510 $t\bar{t}H$ low p_T .

$$\mu_{t\bar{t}H \text{ high } p_T} = 2.1^{+0.62}_{-0.59} (\text{stat})^{+0.40}_{-0.43} (\text{sys})$$

$$\mu_{t\bar{t}H \text{ low } p_T} = 0.83^{+0.37}_{-0.37} (\text{stat})^{+0.48}_{-0.47} (\text{sys})$$

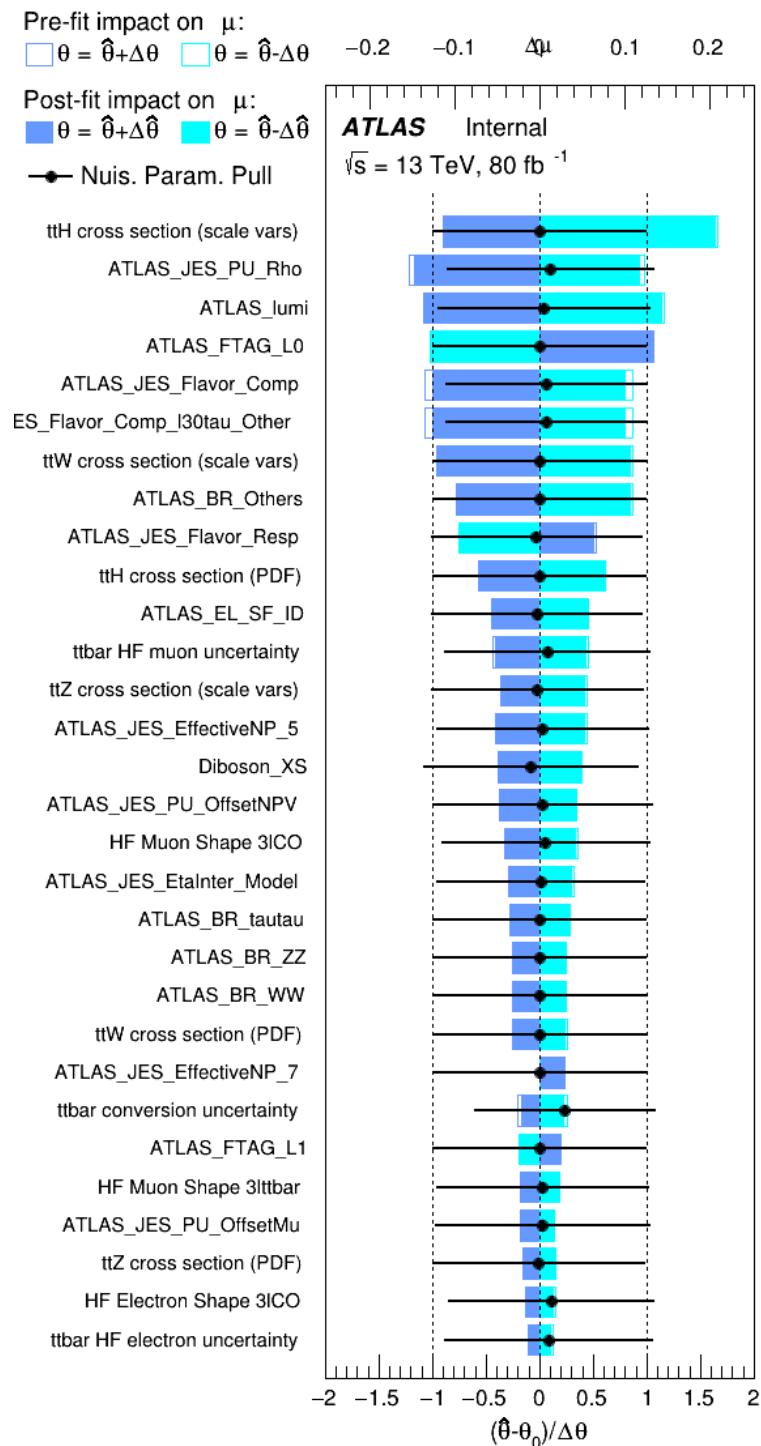
Table 27: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

511 The most prominent sources of systematic uncertainty, as measured by their impact on $\mu_{t\bar{t}H \text{ high } p_T}$,
 512 are summarized in Table 28.

Uncertainty Source	$\Delta\mu$	
Jet Energy Scale	0.25	0.23
t̄H cross-section (QCD Scale)	-0.11	0.21
ATLAS Luminosity	-0.13	0.14
Jet Flavor Tagging	0.14	0.13
t̄W cross-section (QCD Scale)	-0.12	0.11
Higgs Branching Ratio	-0.1	0.11
t̄H cross-section (PDF)	-0.07	0.08
Electron ID	-0.06	0.06
t̄ HF Muon Unc.	-0.05	0.06
t̄Z cross-section (QCD Scale)	-0.05	0.05
Diboson cross-section	-0.05	0.05
HF Muon Shape - 3l	-0.04	0.04
Total Systematic Uncertainty	0.40	0.43

Table 28: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H high p_T.

⁵¹³ The ranking and impact of those nuisance parameters with the largest contribution to the overall uncertainty is shown in Figure 8.3.
⁵¹⁴

Figure 8.3: Impact of systematic uncertainties on the measurement of high p_T $t\bar{t}H$ events

515 The background composition of each of the fit regions is shown in Figure 8.4.

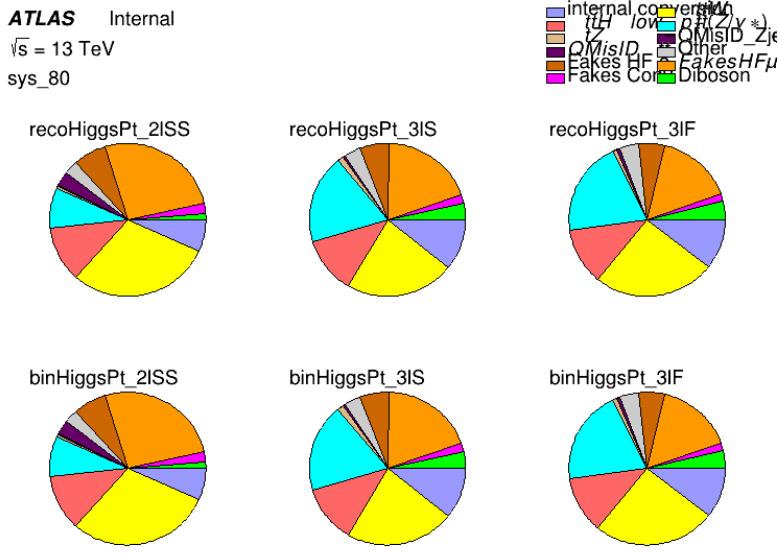


Figure 8.4: Background composition of the fit regions.

516 8.2 Projected Results - 140 fb^{-1}

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

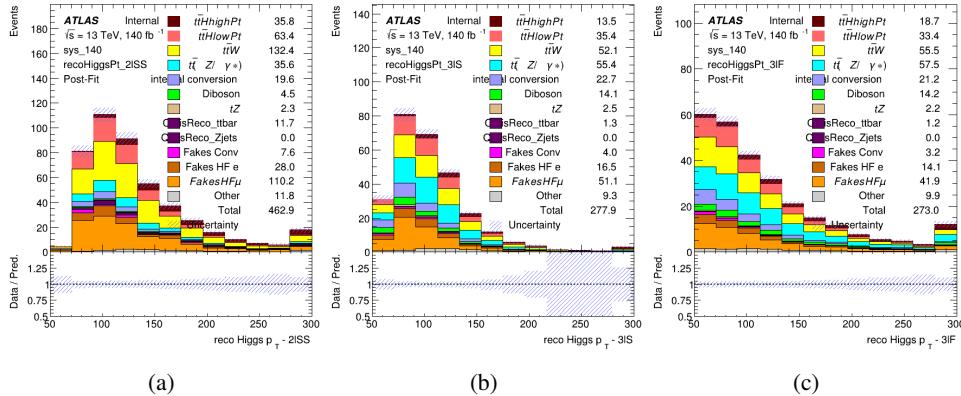


Figure 8.5: Blinded post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2ISSL, (b) 3ISL, and (c) 3IFL, for 140 fb^{-1} of data

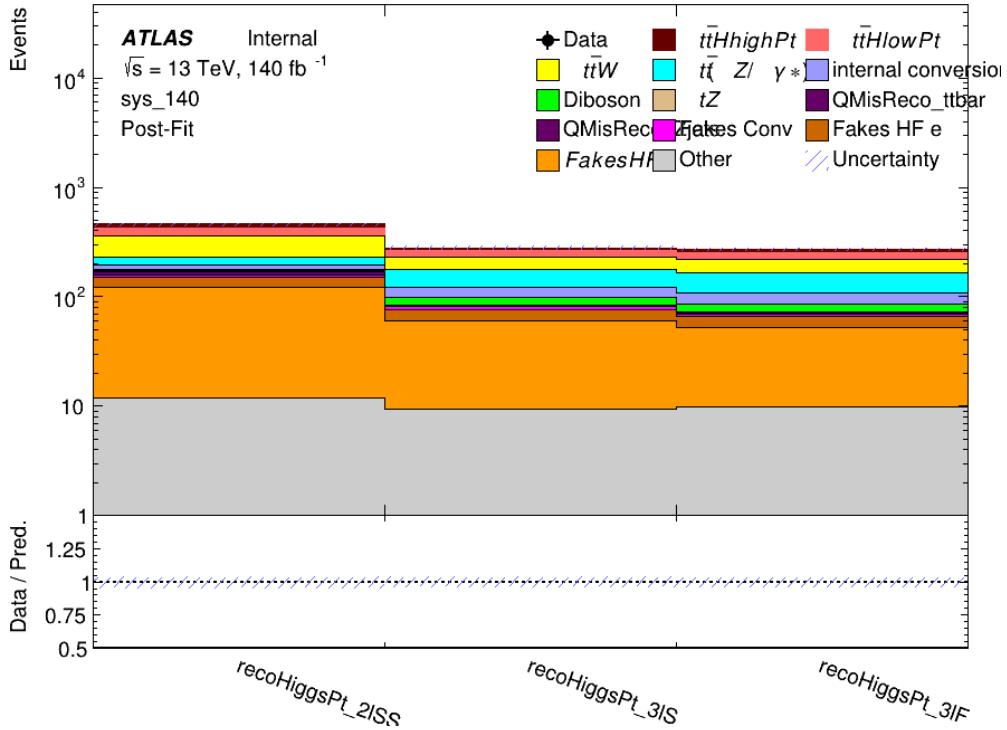


Figure 8.6: Post-fit summary of fit.

Projected uncertainties on the μ values extracted from the fit for high and low p_T Higgs are shown in 29. A significance of 2.0σ is expected for $t\bar{t}H$ high p_T , and a projected significance 2.3σ is extracted for $t\bar{t}H$ low p_T .

$$\mu_{t\bar{t}H \text{ high } p_T} = 1.00^{+0.45}_{-0.43} (\text{stat})^{+0.31}_{-0.31} (\text{sys})$$

$$\mu_{t\bar{t}H \text{ low } p_T} = 1.00^{+0.29}_{-0.30} (\text{stat})^{+0.47}_{-0.46} (\text{sys})$$

Table 29: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

The most prominent sources of systematic uncertainty, as measured by their impact on $\mu_{t\bar{t}H \text{ high } p_T}$, are summarized in Table 30.

Uncertainty Source	$\Delta\mu$	
Jet Energy Scale	0.2	0.18
t̄W cross-section (QCD Scale)	-0.12	0.11
ATLAS Luminosity	-0.11	0.11
Jet Flavor Tagging	0.11	0.10
t̄H cross-section (QCD Scale)	-0.06	0.06
Higgs Branching Ratio	-0.1	0.11
t̄H cross-section (PDF)	-0.07	0.08
Electron ID	-0.05	0.05
t̄ HF Muon Unc.	-0.04	0.06
t̄Z cross-section (QCD Scale)	-0.03	0.04
Diboson cross-section	-0.03	0.03
HF Muon Shape - 3l	-0.02	0.02
Total Systematic Uncertainty	0.31	0.31

Table 30: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H high p_T.

⁵²⁶ The ranking and impact of those nuisance parameters with the largest contribution to the overall
⁵²⁷ uncertainty is shown in Figure 8.7.

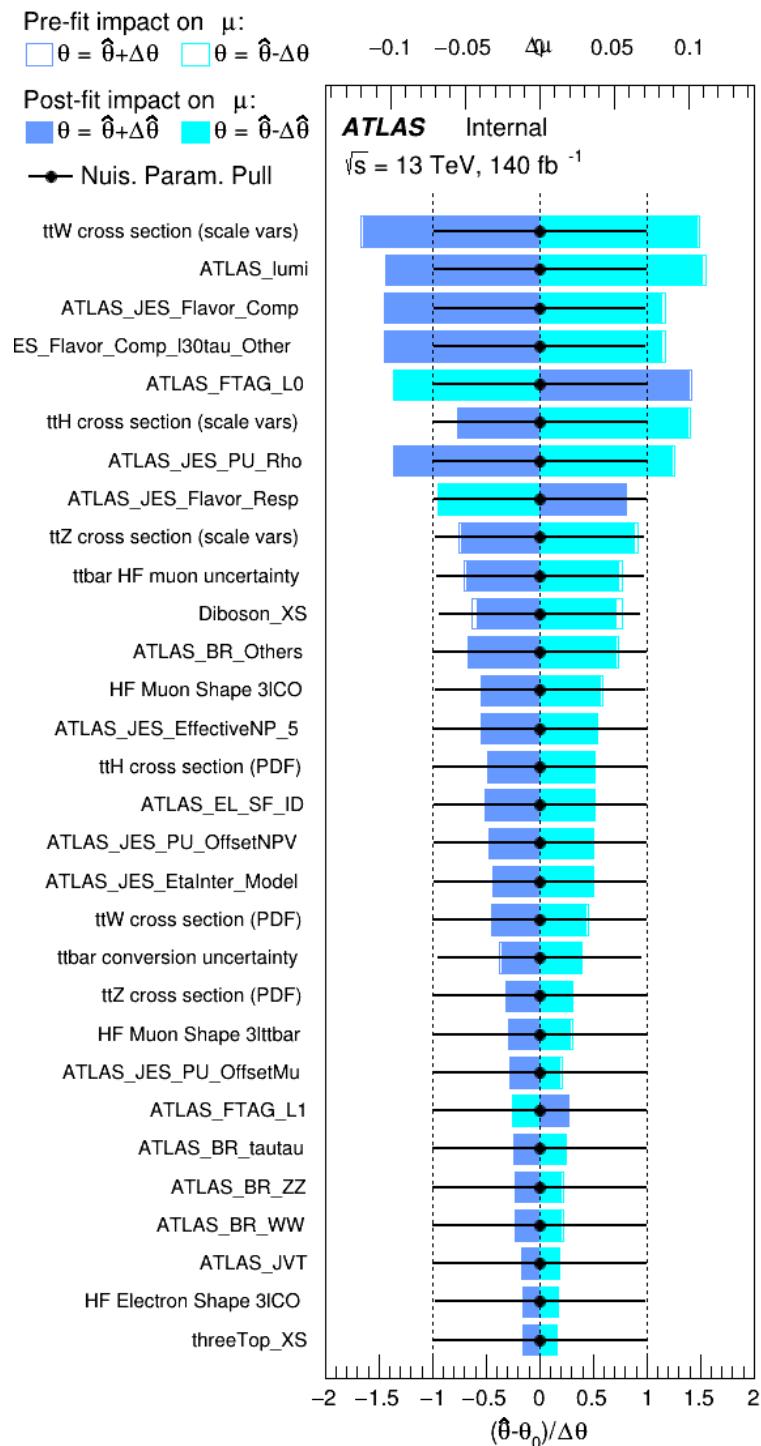


Figure 8.7: Impact of systematic uncertainties on the measurement of high p_T $t\bar{t}H$ events

528 The background composition of each of the fit regions is shown in Figure 8.8.

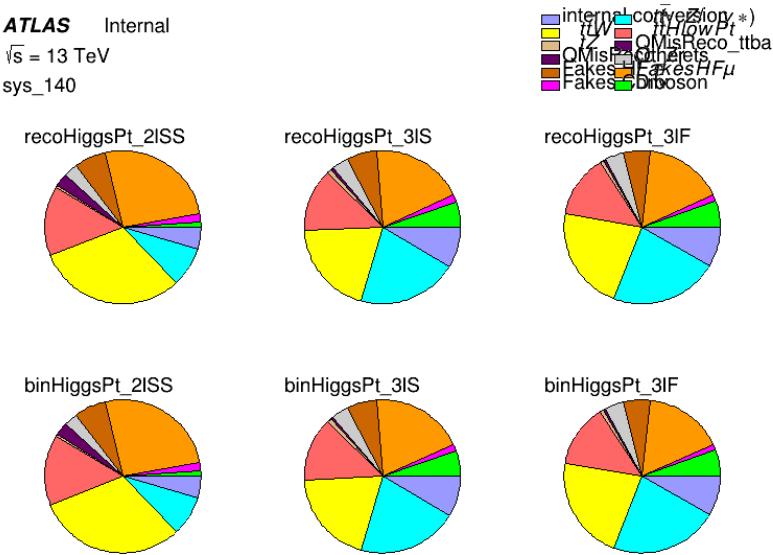


Figure 8.8: Background composition of the fit regions.

9 Conclusion

529 As search for the effects of dimension-six operators on $t\bar{t}H$ production is performed. An effective
 530 field theory approach is used to parameterize the effects of high energy physics on the Higgs
 531 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,
 532 and the limits on dimension-six operators are limited to X.
 533

References

- 535 [1] ATLAS Collaboration, *Observation of a Centrality-Dependent Dijet Asymmetry in Lead–*
 536 *Lead Collisions at $\sqrt{s_{NN}} = 2.76 \text{ TeV}$ with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.*
 537 **105** (2010) 252303, arXiv: [1011.6182 \[hep-ex\]](https://arxiv.org/abs/1011.6182) (cit. on p. 5).
- 538 [2] B. Dumont, S. Fichet and G. von Gersdorff, *A Bayesian view of the Higgs sector with higher*
 539 *dimensional operators*, *Journal of High Energy Physics* **2013** (2013), ISSN: 1029-8479,
 540 URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)065](http://dx.doi.org/10.1007/JHEP07(2013)065) (cit. on p. 5).
- 541 [3] S. Banerjee, S. Mukhopadhyay and B. Mukhopadhyaya, *Higher dimensional operators and*
 542 *the LHC Higgs data: The role of modified kinematics*, *Physical Review D* **89** (2014), ISSN:
 543 1550-2368, URL: <http://dx.doi.org/10.1103/PhysRevD.89.053010> (cit. on p. 5).

- 544 [4] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using*
 545 *the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, URL: <https://cds.cern.ch/record/2157687> (cit. on p. 9).
- 547 [5] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the*
 548 *ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, *Eur. Phys. J. C*
 549 **74** (2014) 3130, arXiv: [1407.3935 \[hep-ex\]](https://arxiv.org/abs/1407.3935) (cit. on p. 9).
- 550 [6] *Evidence for the associated production of the Higgs boson and a top quark pair with*
 551 *the ATLAS detector*, tech. rep. ATLAS-CONF-2017-077, CERN, 2017, URL: <https://cds.cern.ch/record/2291405> (cit. on p. 9).
- 553 [7] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed*
 554 *in the ATLAS Detector at $\sqrt{s} = 13 \text{ TeV}$* , ATL-PHYS-PUB-2015-015, 2015, URL: <https://cds.cern.ch/record/2037613> (cit. on p. 10).
- 556 [8] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton–proton collisions with*
 557 *the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: <https://cds.cern.ch/record/2037702> (cit. on p. 10).
- 559 [9] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions*
 560 *at $\sqrt{s} = 8 \text{ TeV}$ using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 581, arXiv: [1510.03823 \[hep-ex\]](https://arxiv.org/abs/1510.03823) (cit. on p. 10).
- 562 [10] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
 563 *with the ATLAS detector in the first proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$* , ATL-
 564 PHYS-PUB-2015-027, 2015, URL: <https://cds.cern.ch/record/2037904> (cit. on
 565 p. 10).
- 566 [11] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ using the*
 567 *ATLAS detector at the LHC*, *Eur. Phys. J. C* **71** (2011) 1630, arXiv: [1101.2185 \[hep-ex\]](https://arxiv.org/abs/1101.2185)
 568 (cit. on p. 60).
- 569 [12] G. Aad, T. Abajyan, B. Abbott and etal, *Jet energy resolution in proton-proton collisions at*
 570 *$\sqrt{s} = 7 \text{ TeV}$ recorded in 2010 with the ATLAS detector*, *The European Physical Journal C* **73** (2013) 2306, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-013-2306-0> (cit. on p. 61).
- 573 [13] A. Collaboration, *Performance of b -jet identification in the ATLAS experiment*, *Journal of*
 574 *Instrumentation* **11** (2016) P04008, URL: <http://stacks.iop.org/1748-0221/11/i=04/a=P04008> (cit. on p. 61).

576 **List of contributions**

577

⁵⁷⁸ **Appendices**

⁵⁷⁹ **A Machine Learning Models**

⁵⁸⁰ The following section provides details of the various MVAs as well as a few studies performed in
⁵⁸¹ support of this analysis, exploring alternate decisions and strategies.

⁵⁸² **A.1 Higgs Reconstruction Models**

⁵⁸³ **A.1.1 b-jet Identification Features - 2lSS**

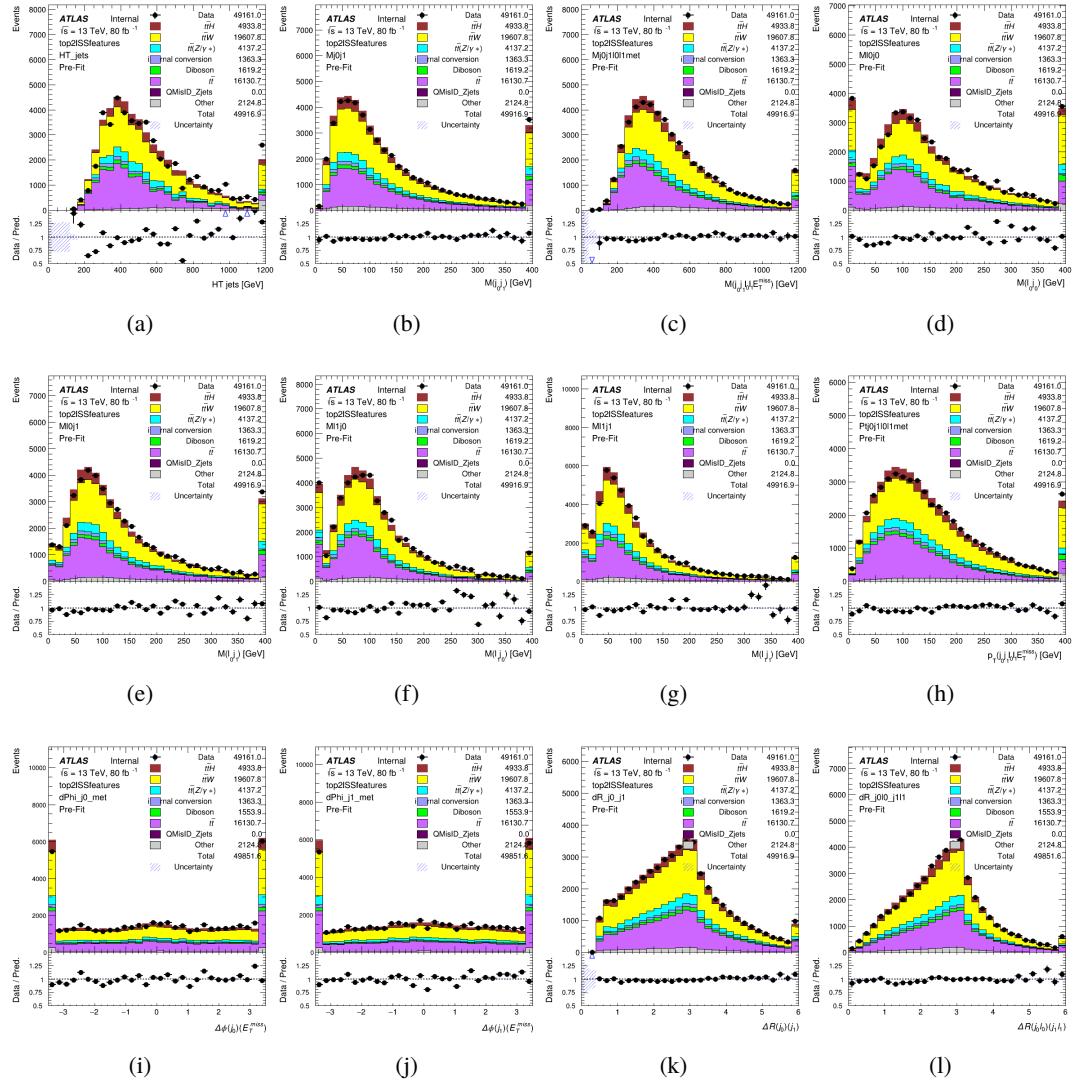


Figure A.1: Input features for top2lSS

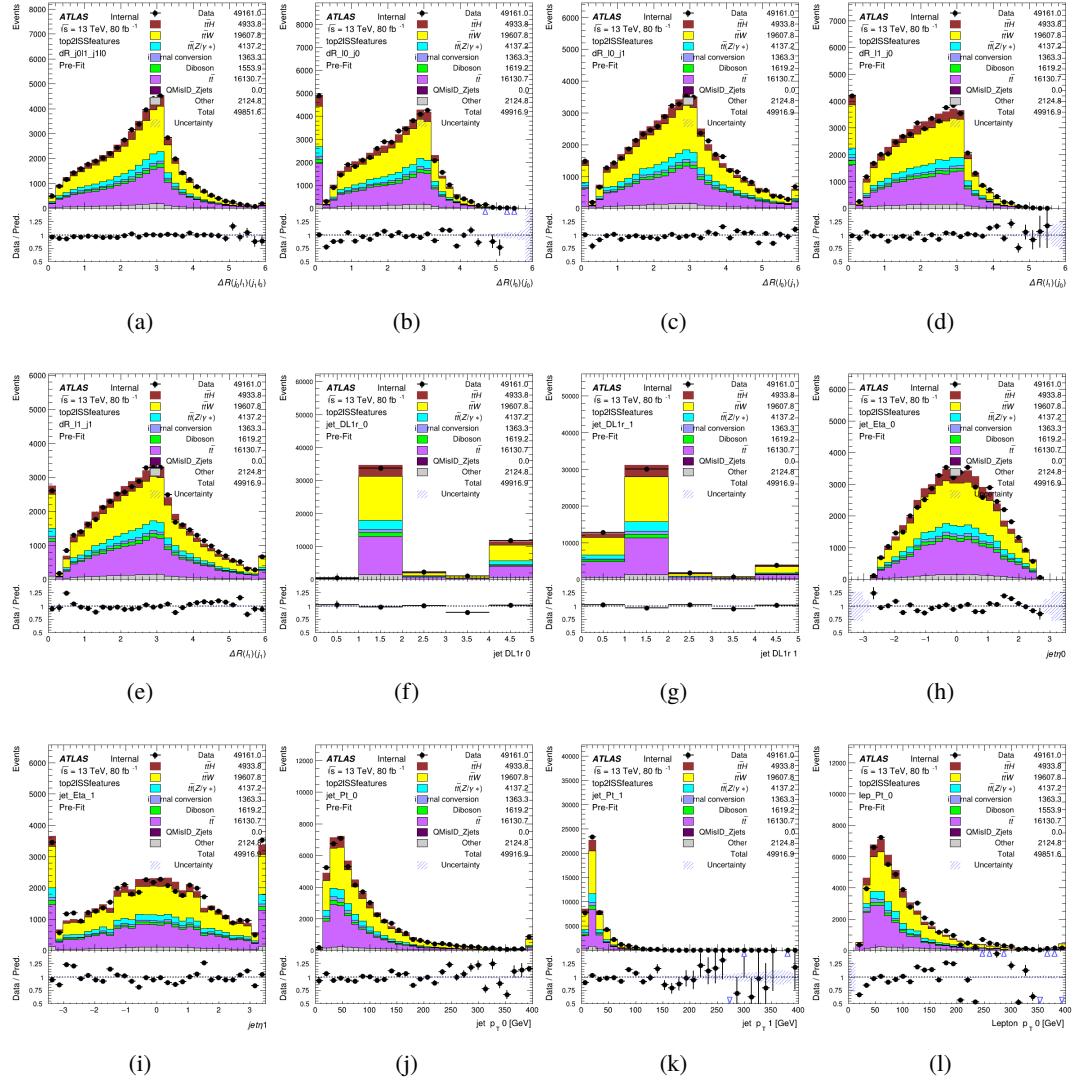


Figure A.2: Input features for top2lSS

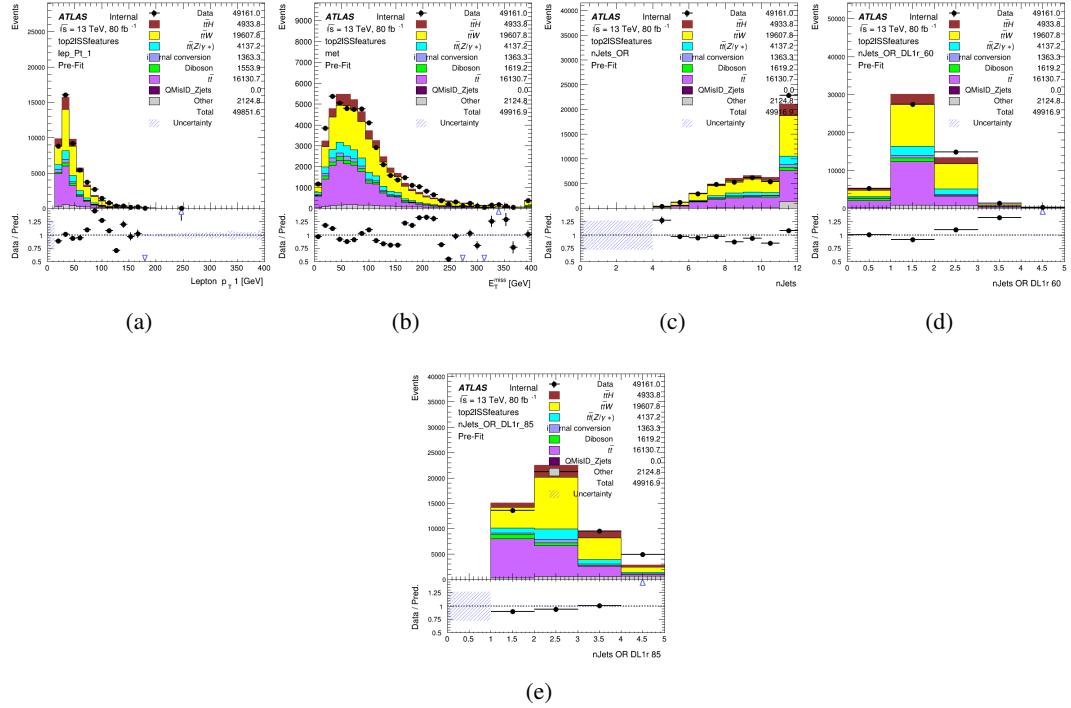


Figure A.3: Input features for top2ISS

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

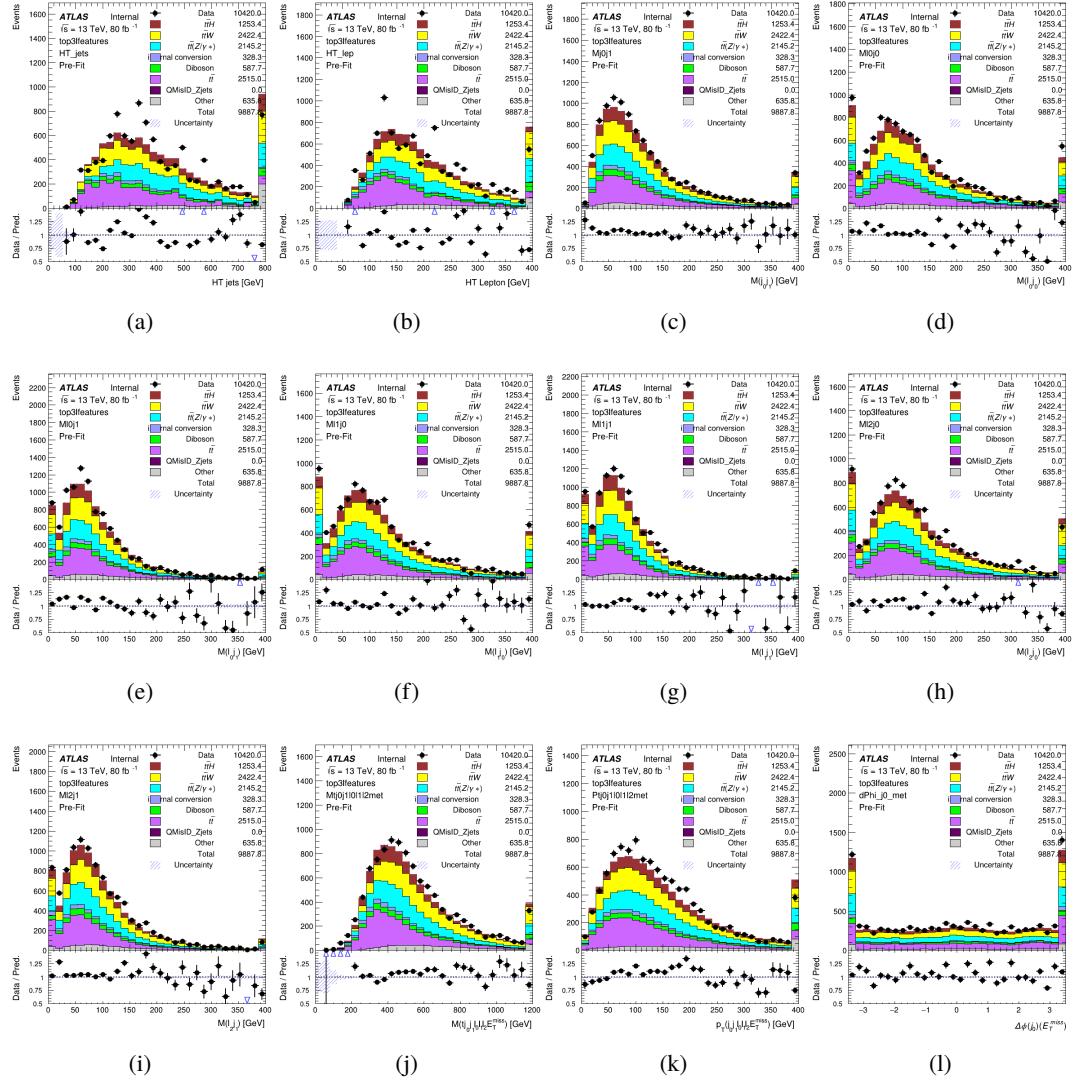


Figure A.4: Input features for top31

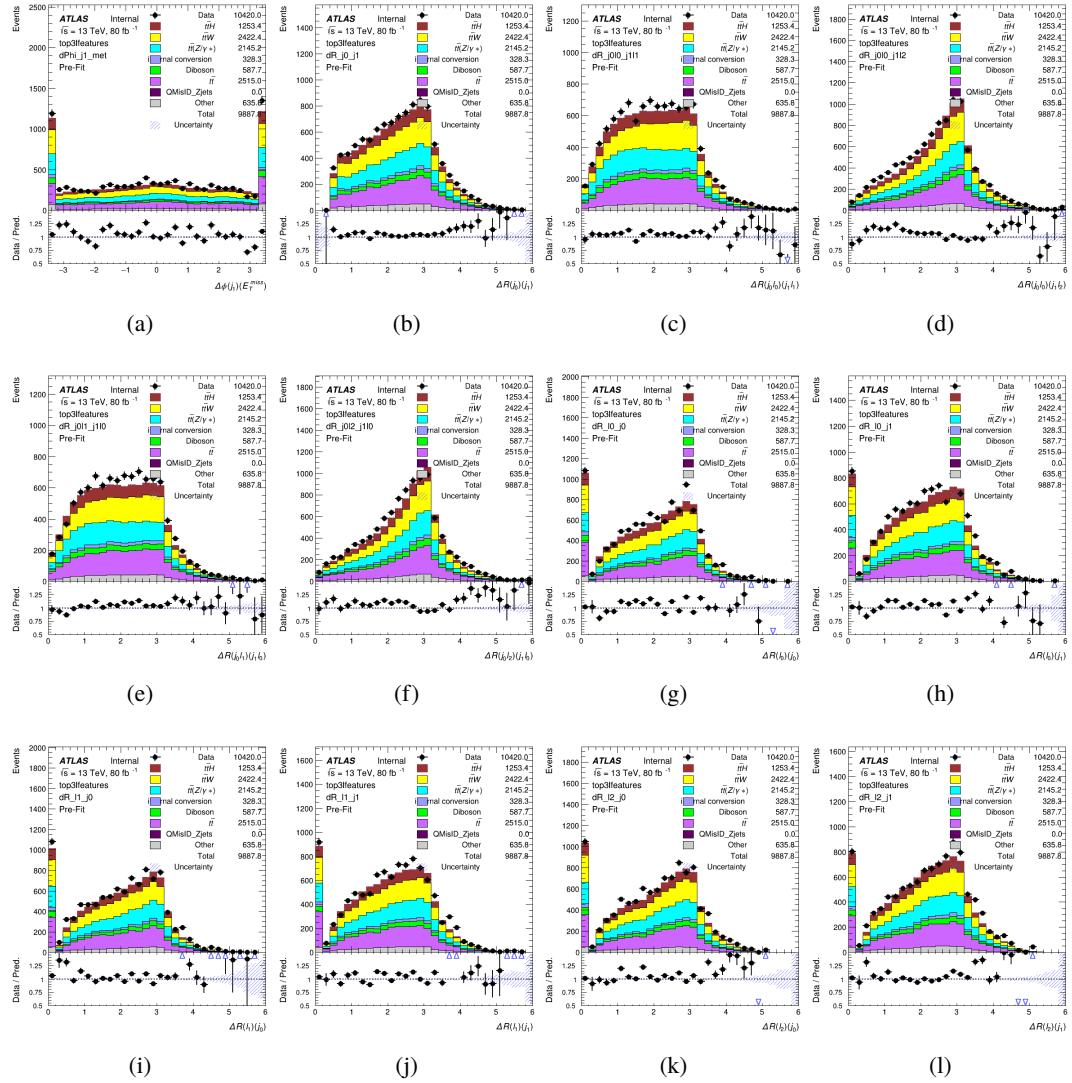


Figure A.5: Input features for top31

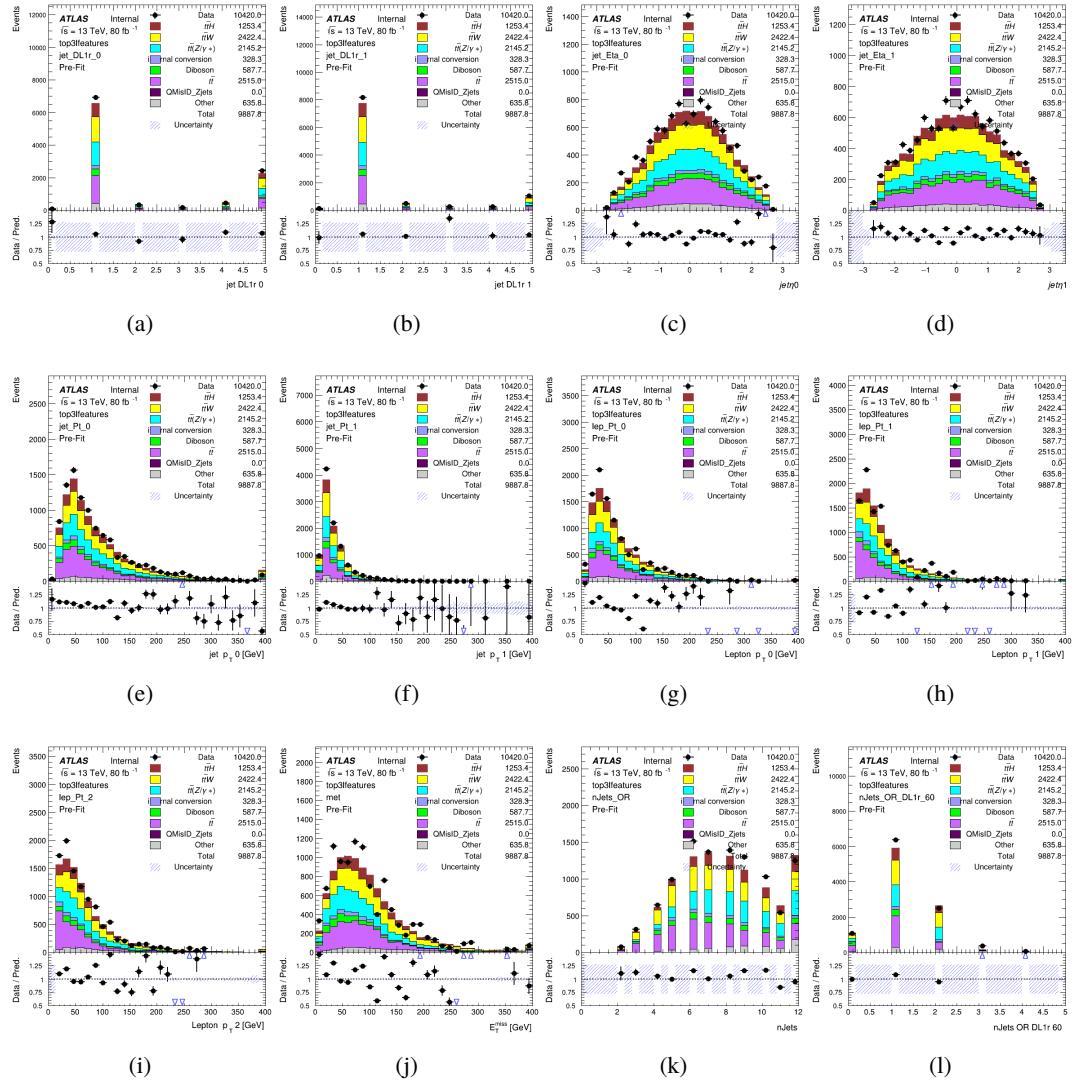
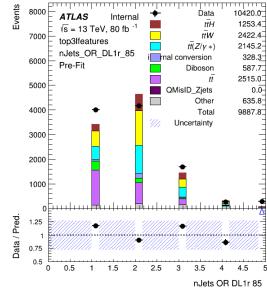


Figure A.6: Input features for top31



(a)

Figure A.7: Input features for top31

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

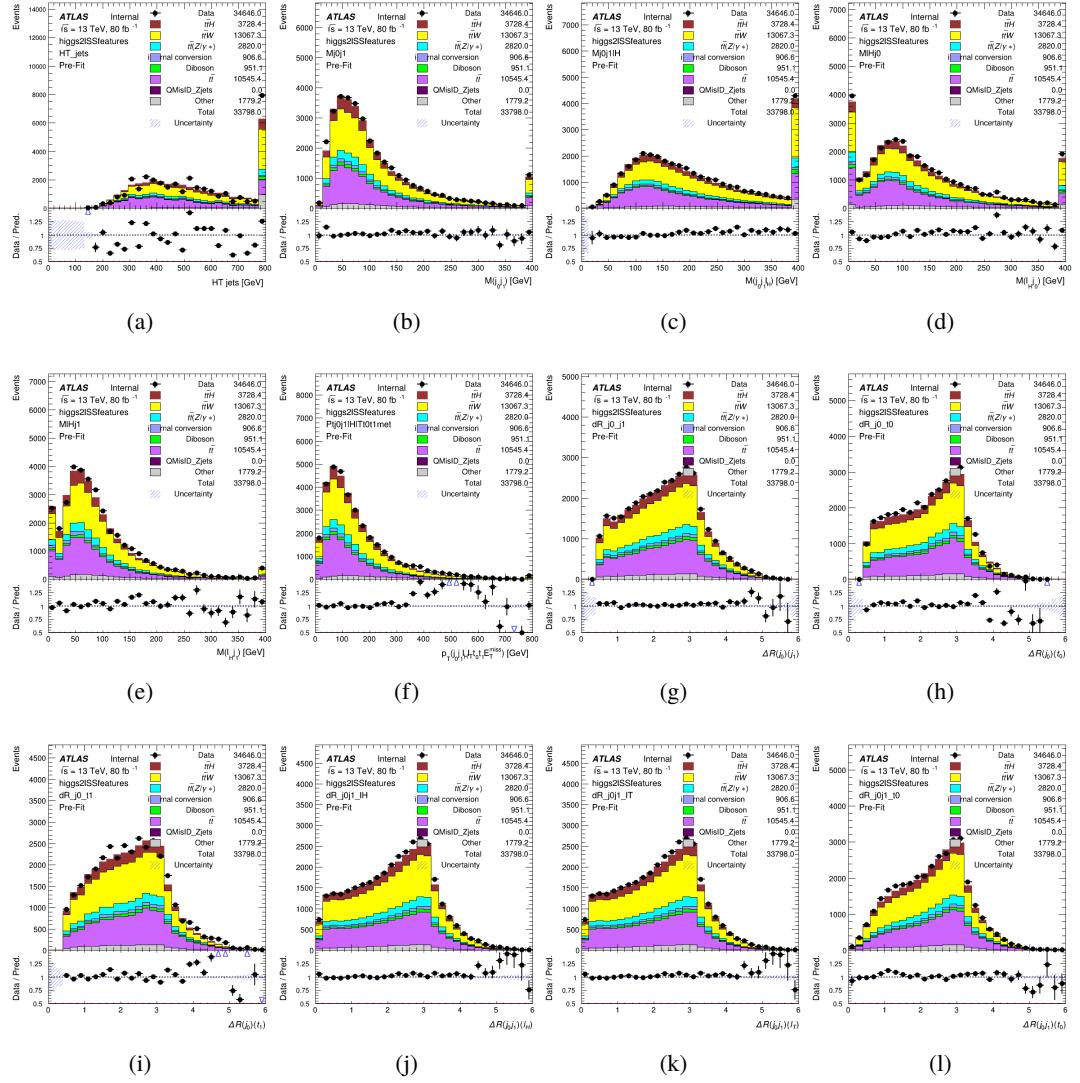


Figure A.8: Input features for higgs2lSS

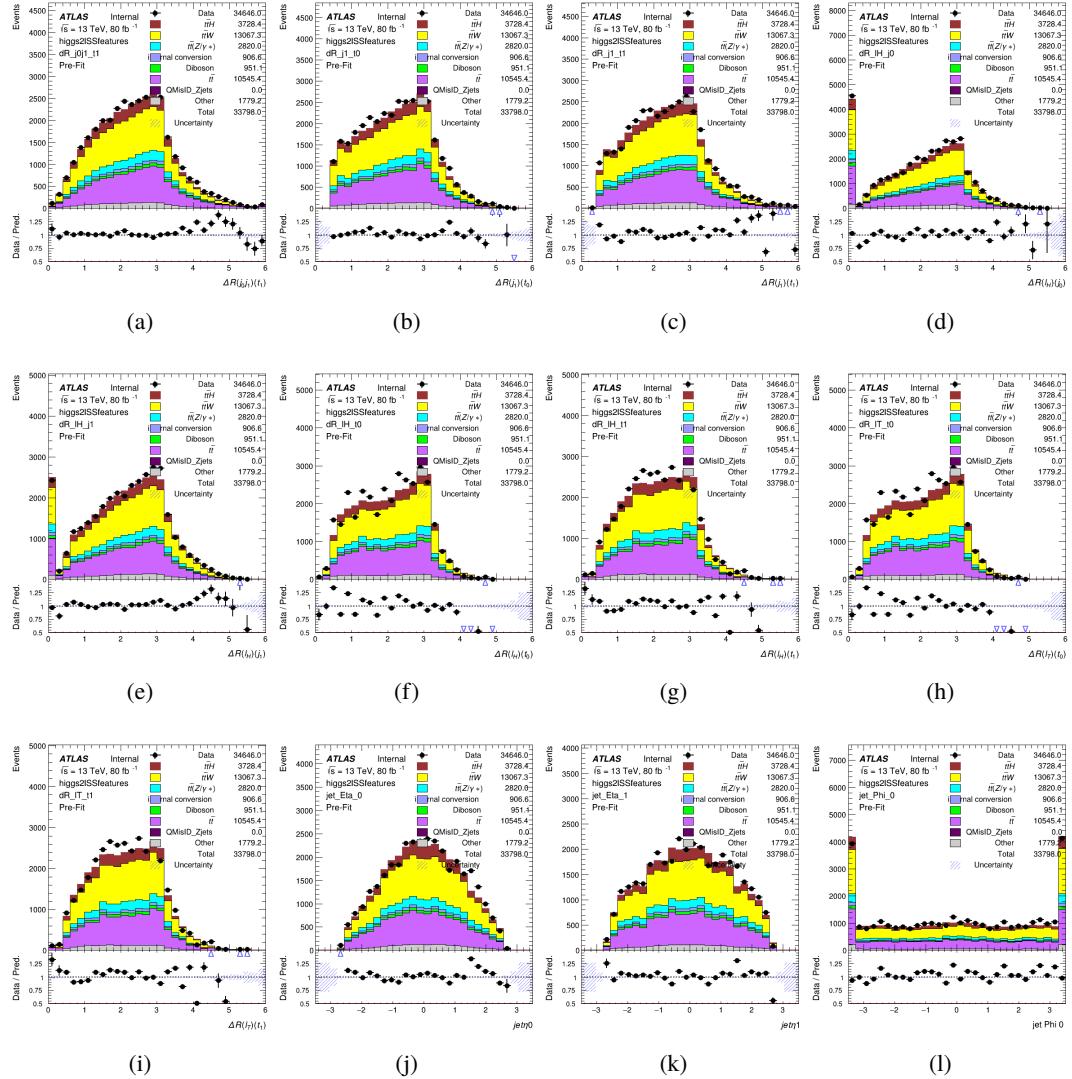


Figure A.9: Input features for higgs2lSS

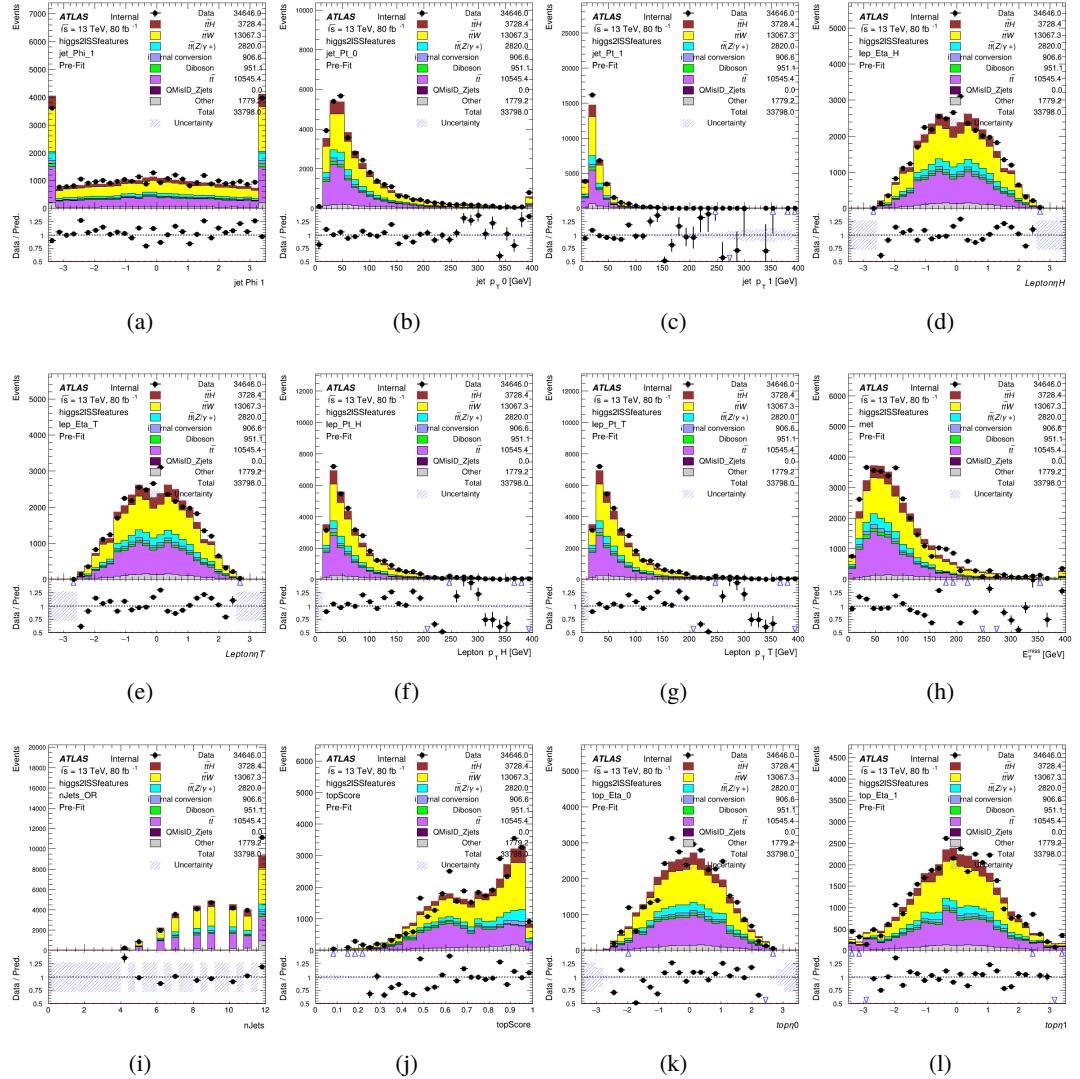


Figure A.10: Input features for higgs2IS

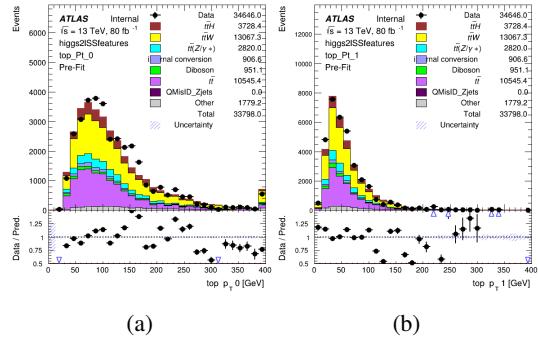


Figure A.11: Input features for higgs2lSS

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

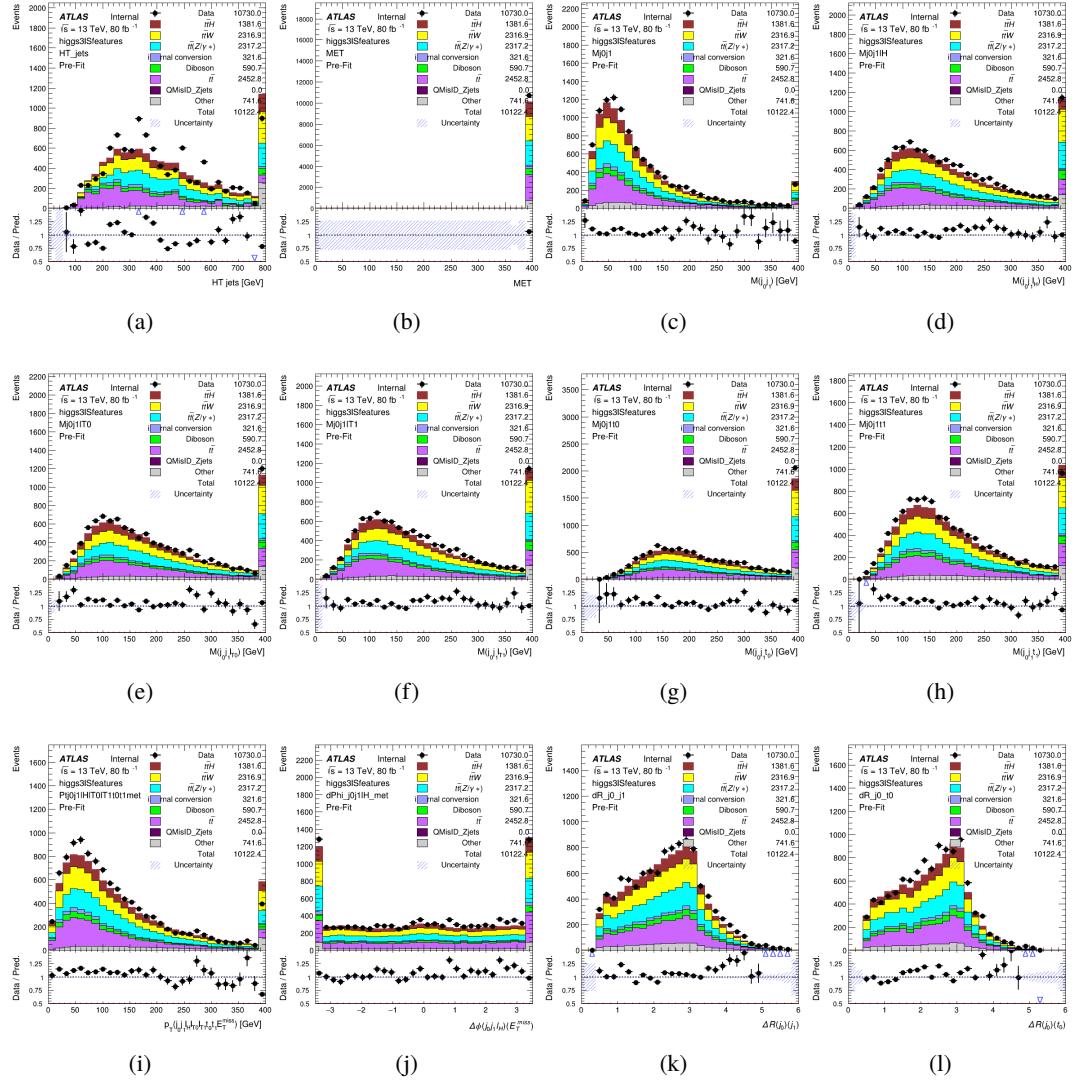


Figure A.12: Input features for higgs3IS

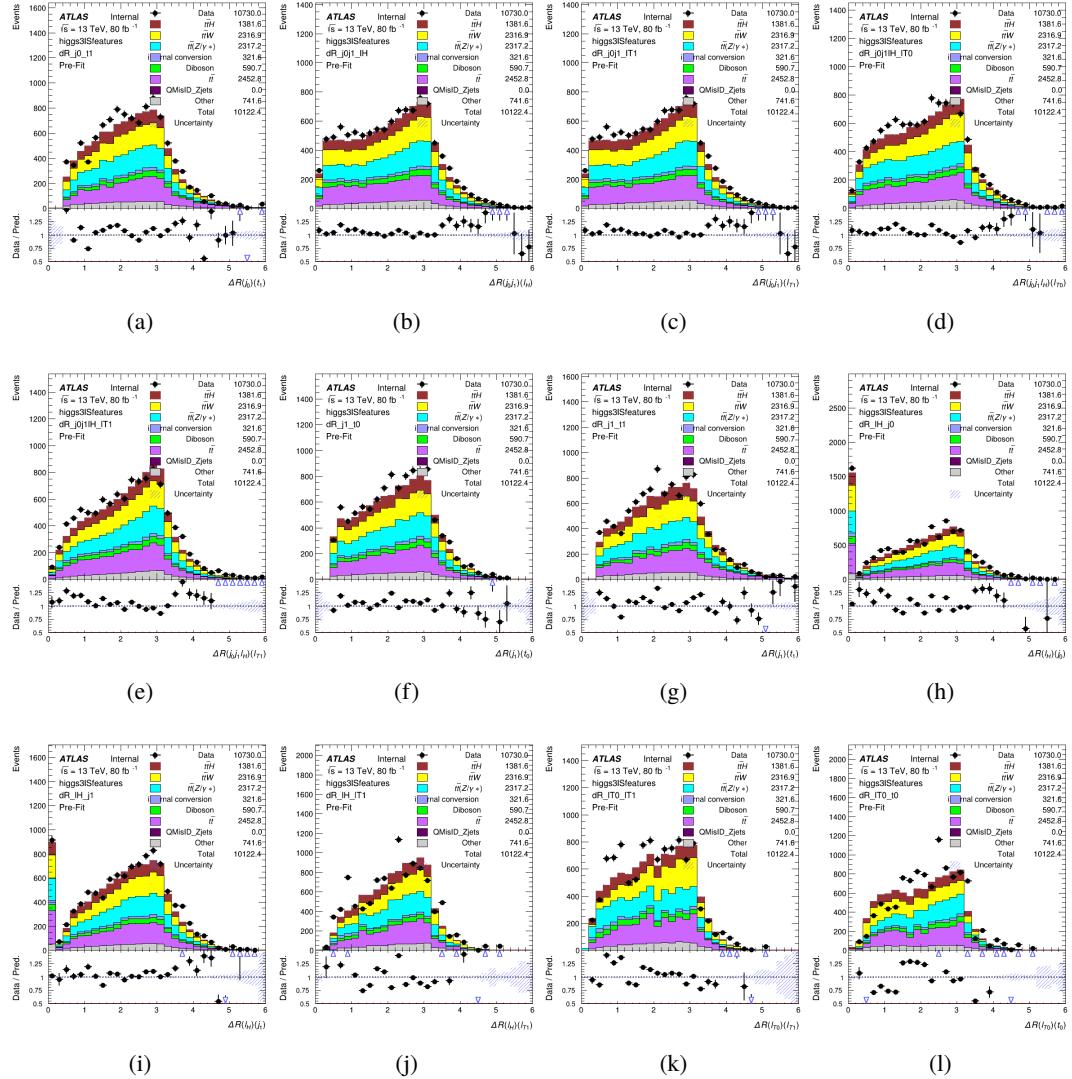


Figure A.13: Input features for higgs3IS

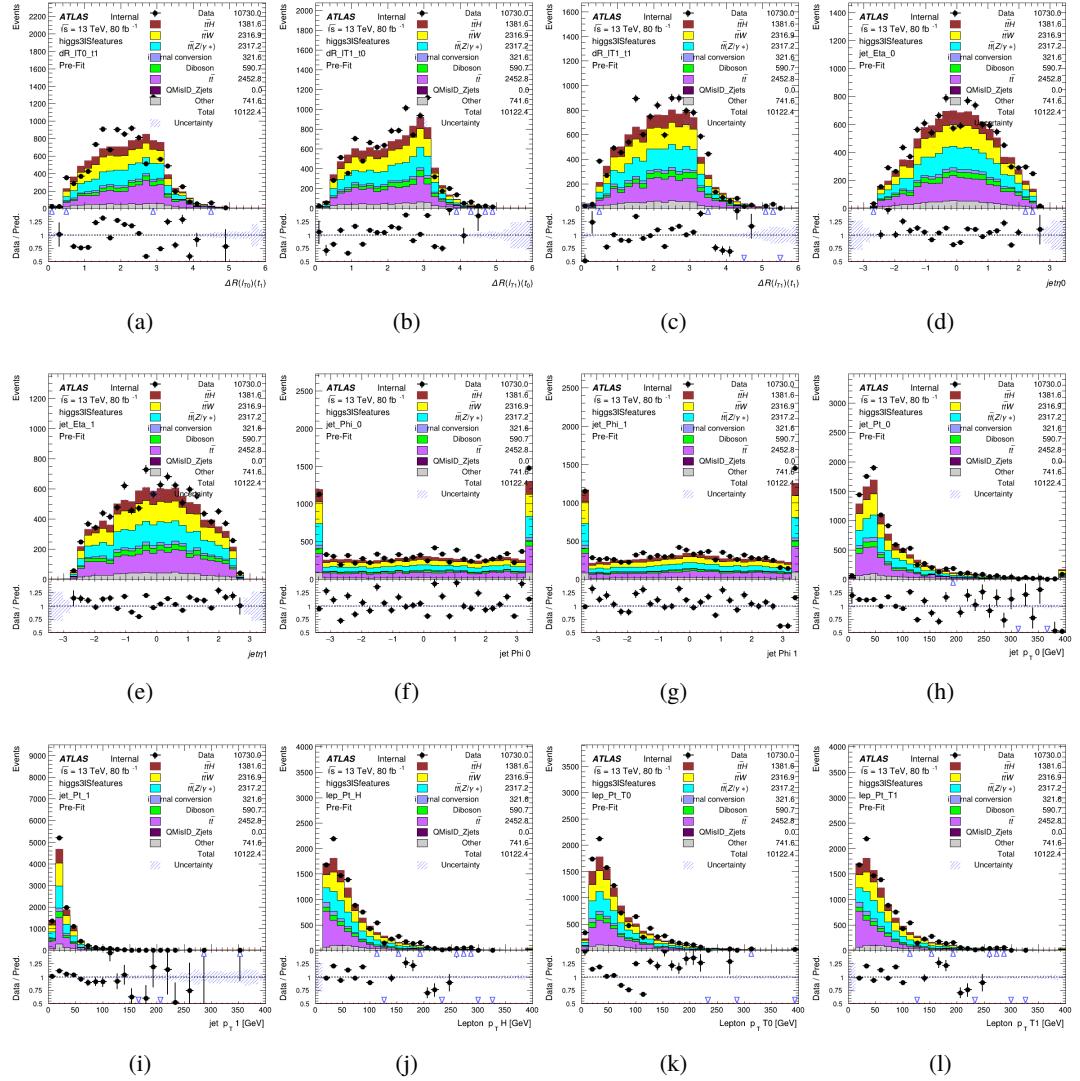


Figure A.14: Input features for higgs3IS

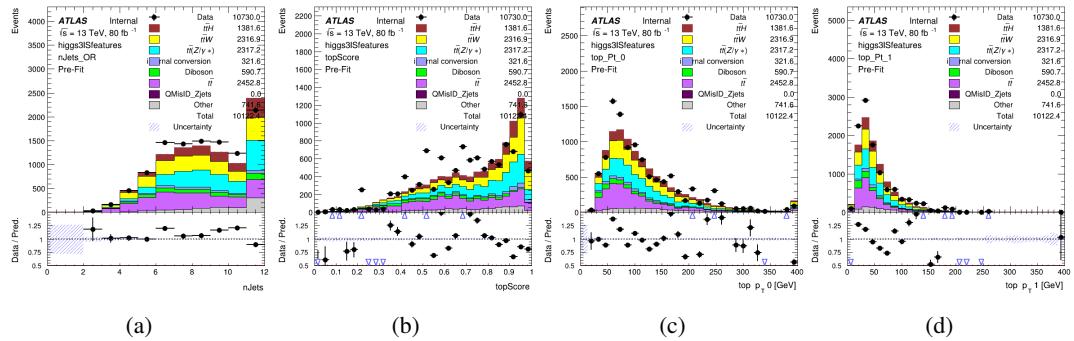


Figure A.15: Input features for higgs3IS

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

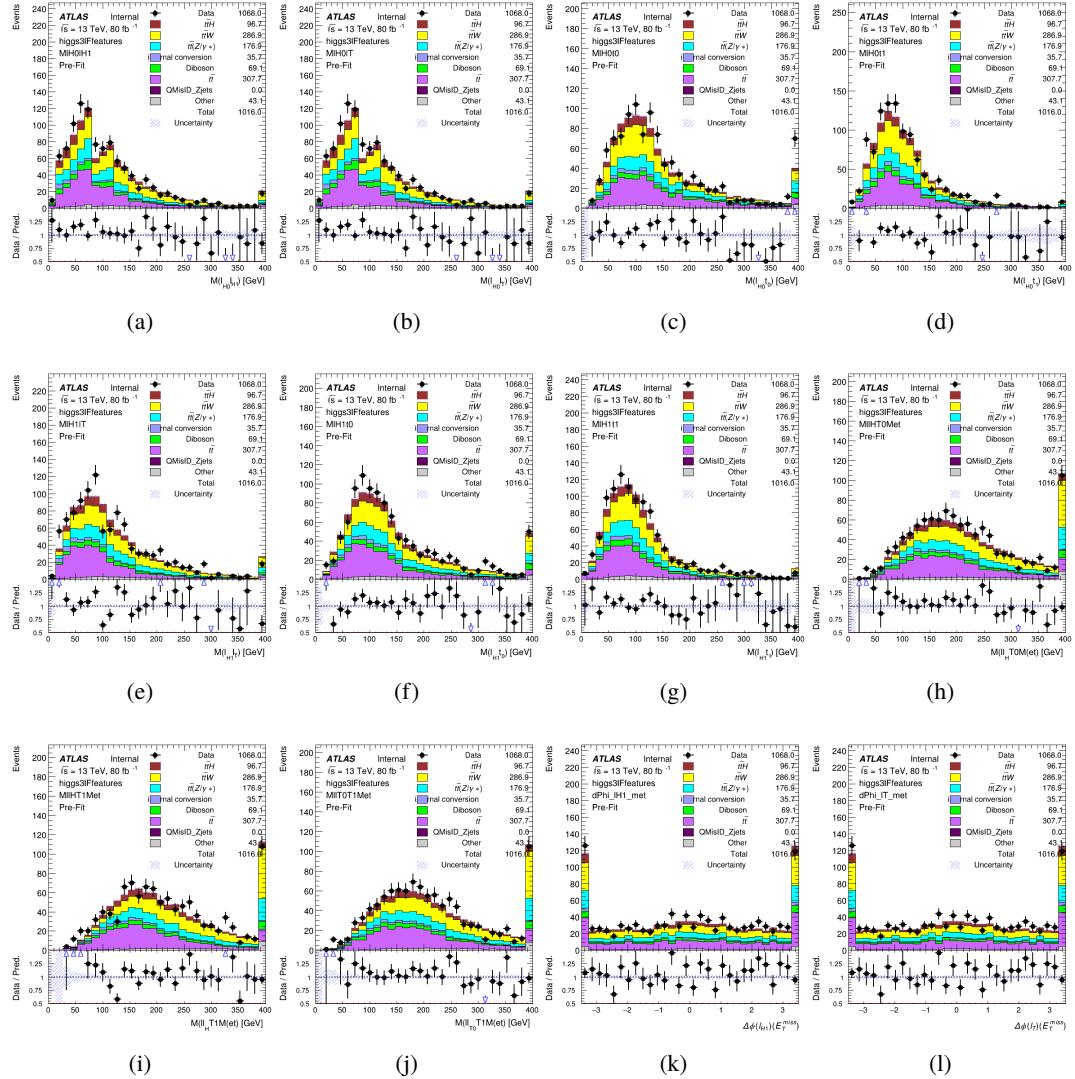


Figure A.16: Input features for higgs3IF

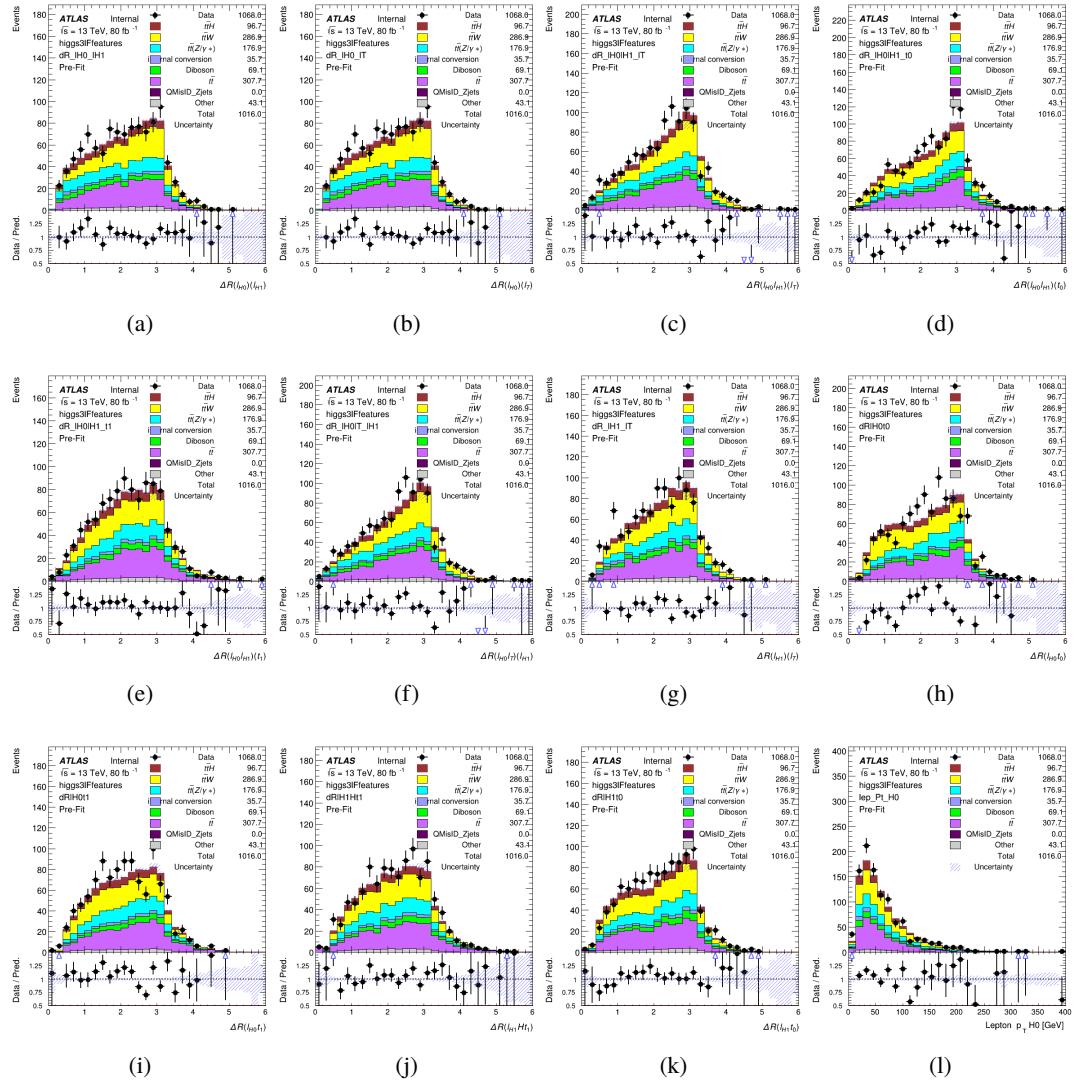


Figure A.17: Input features for higgs3IF

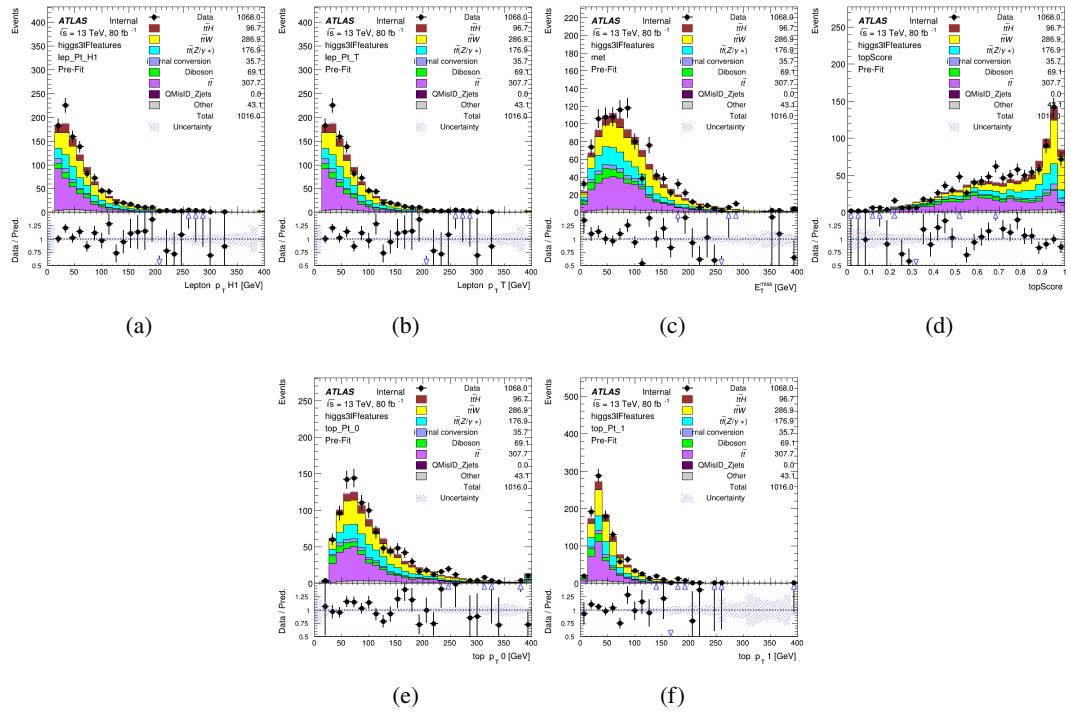


Figure A.18: Input features for higgs3lF

588 **A.2 Background Rejection MVAs**

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

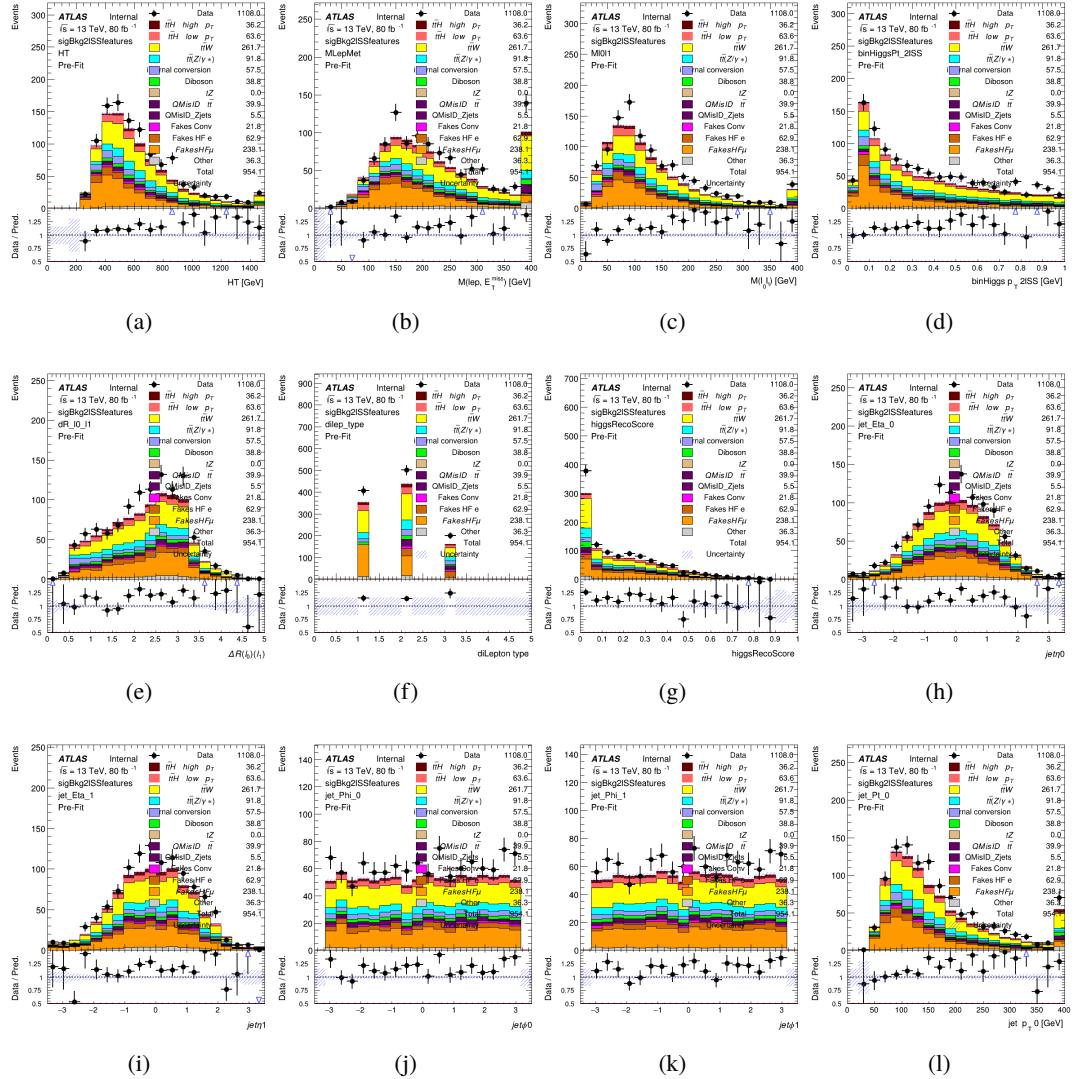


Figure A.19: Input features for `sigBkg2lSS`

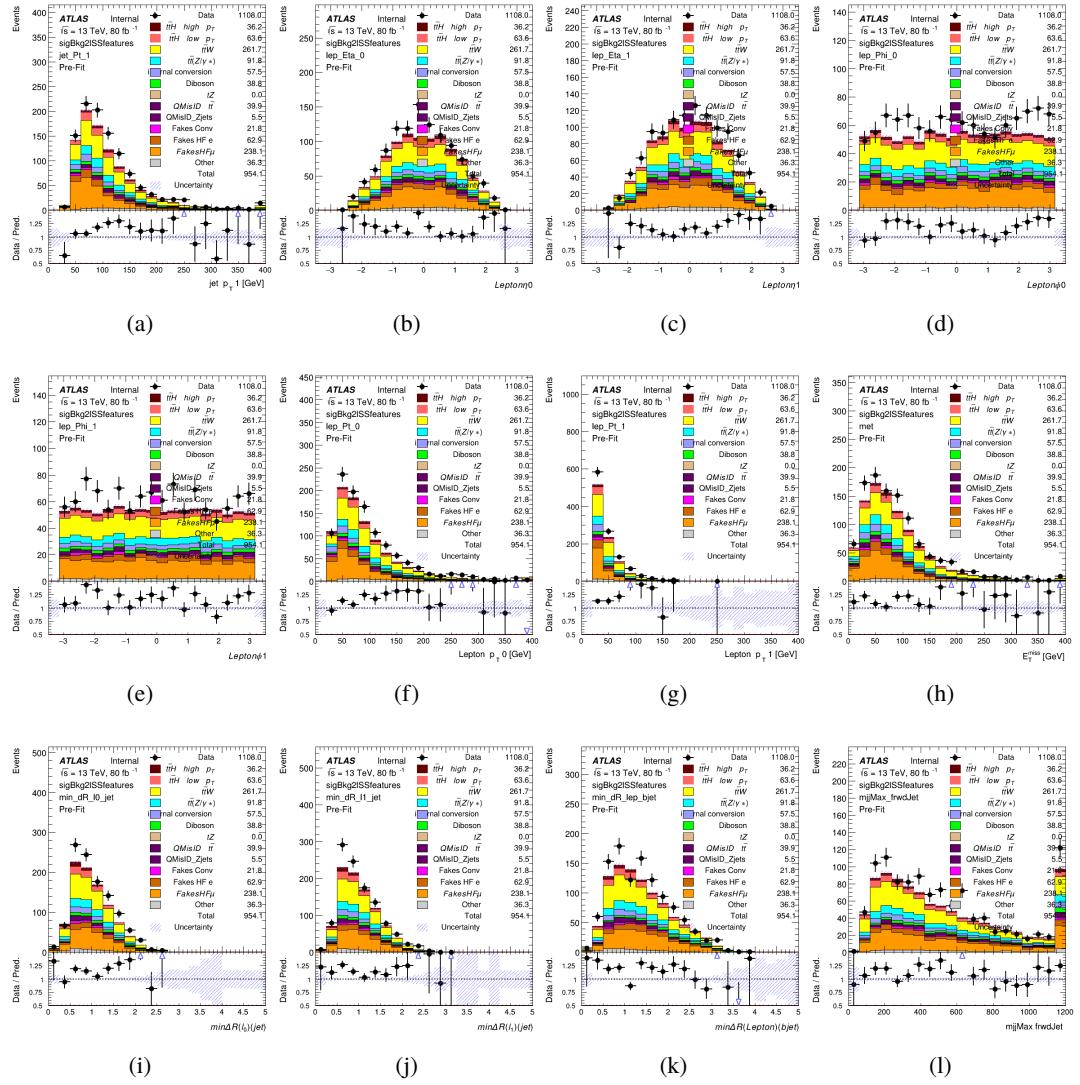


Figure A.20: Input features for sigBkg2lSS

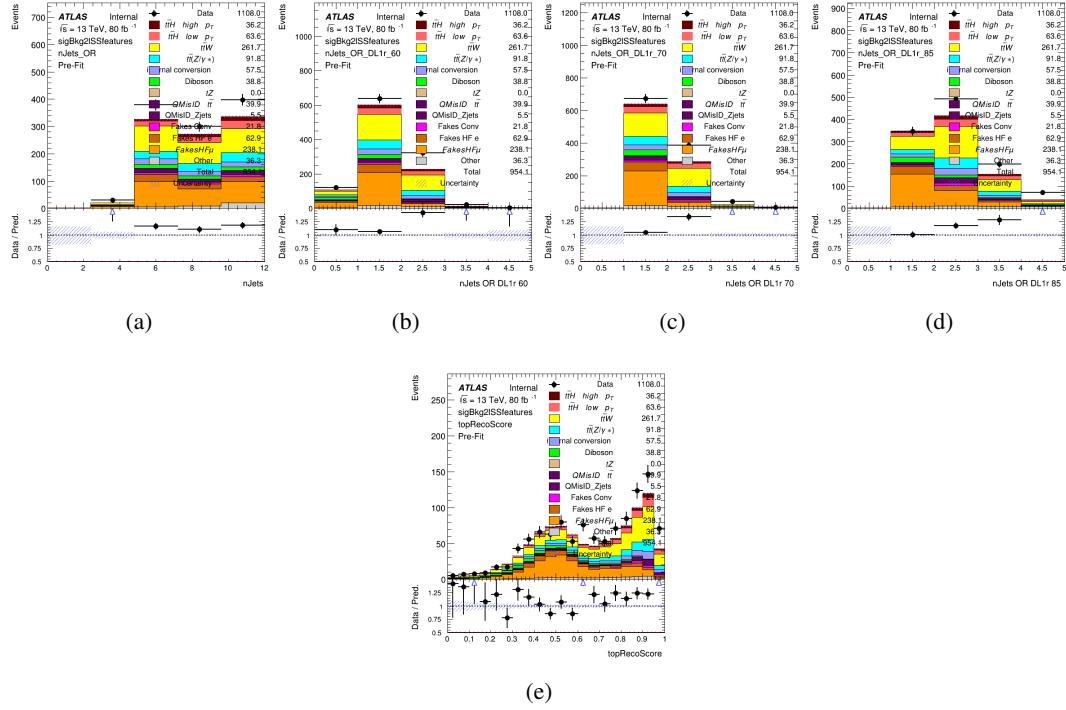


Figure A.21: Input features for sigBkg2lSS

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

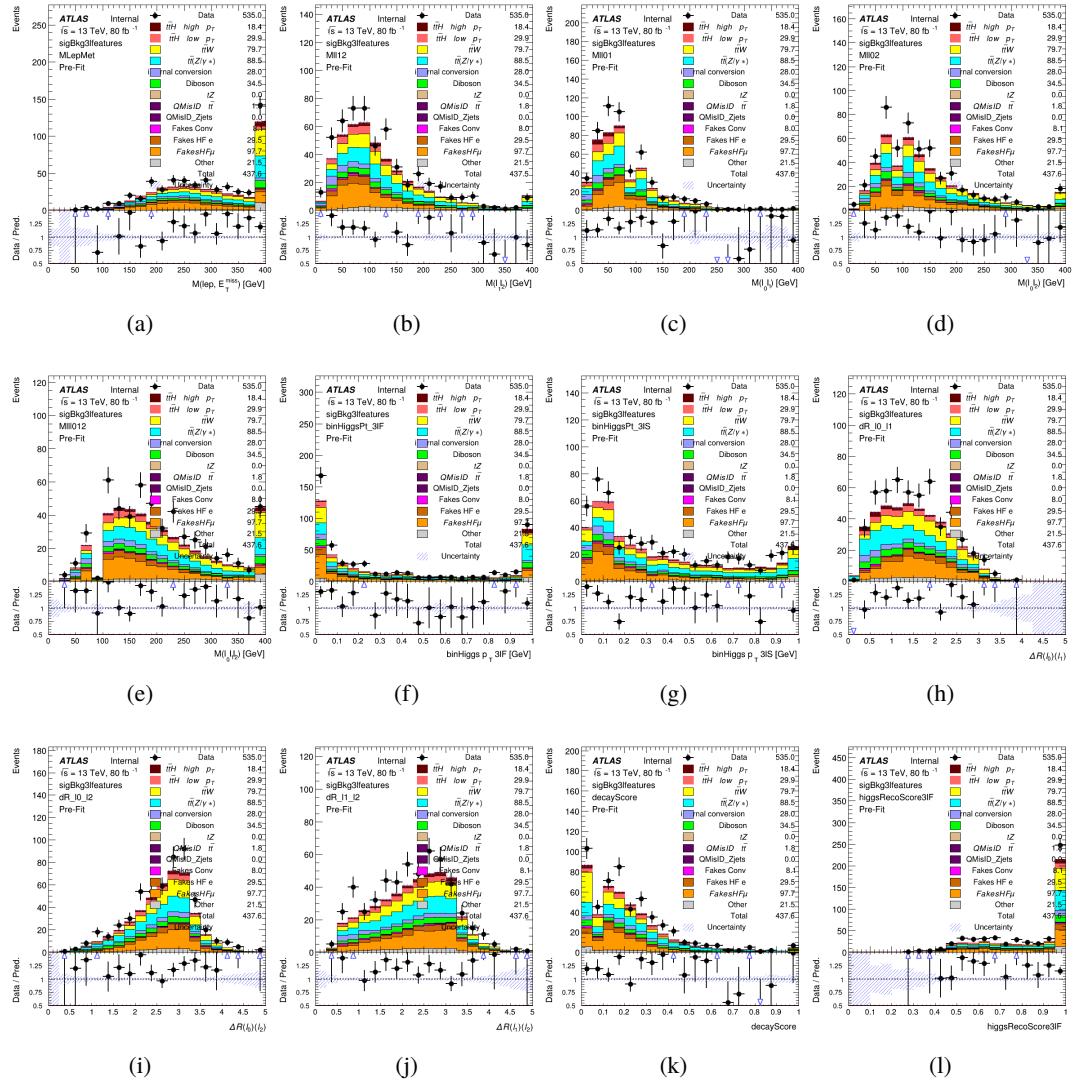


Figure A.22: Input features for sigBkg3l

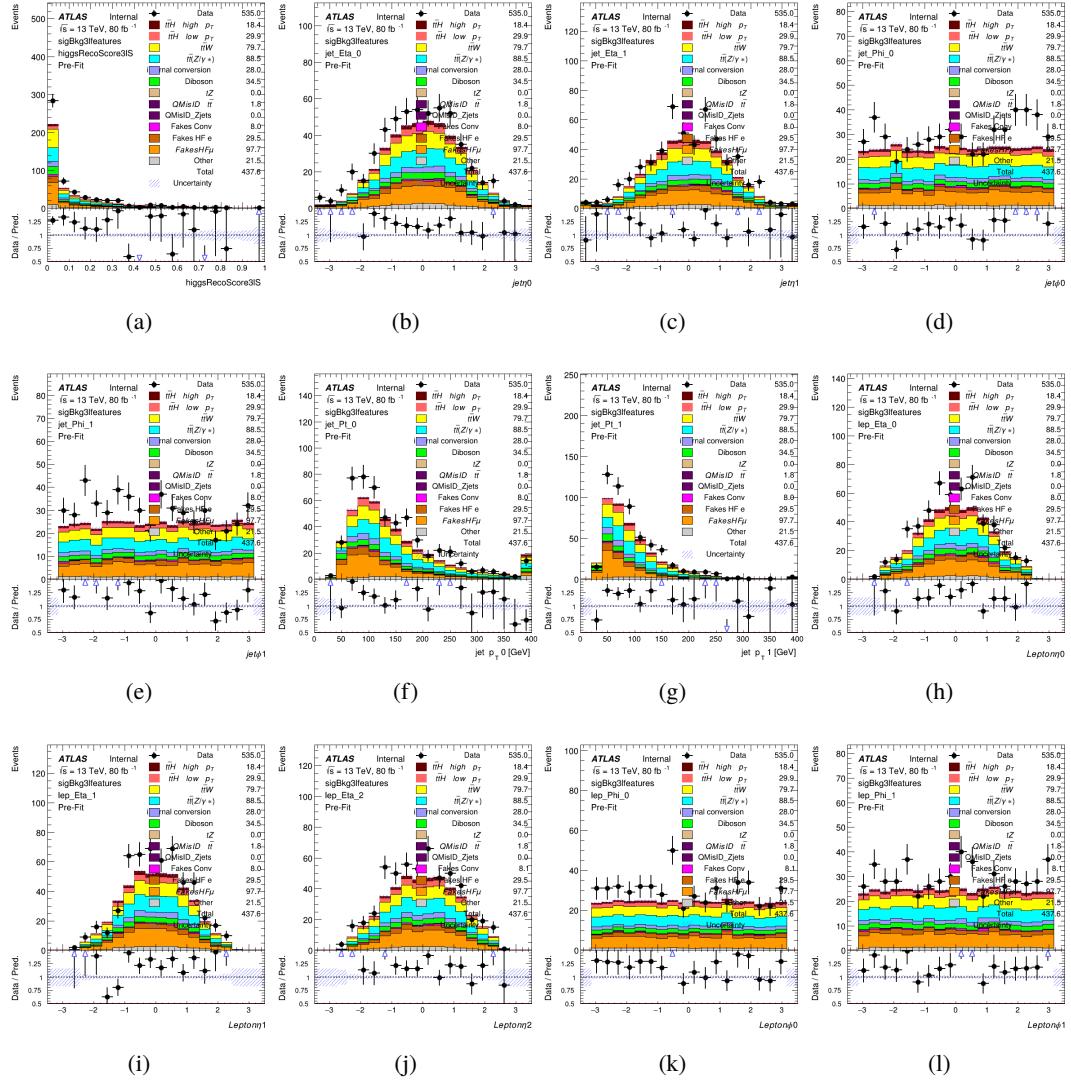


Figure A.23: Input features for sigBkg3l

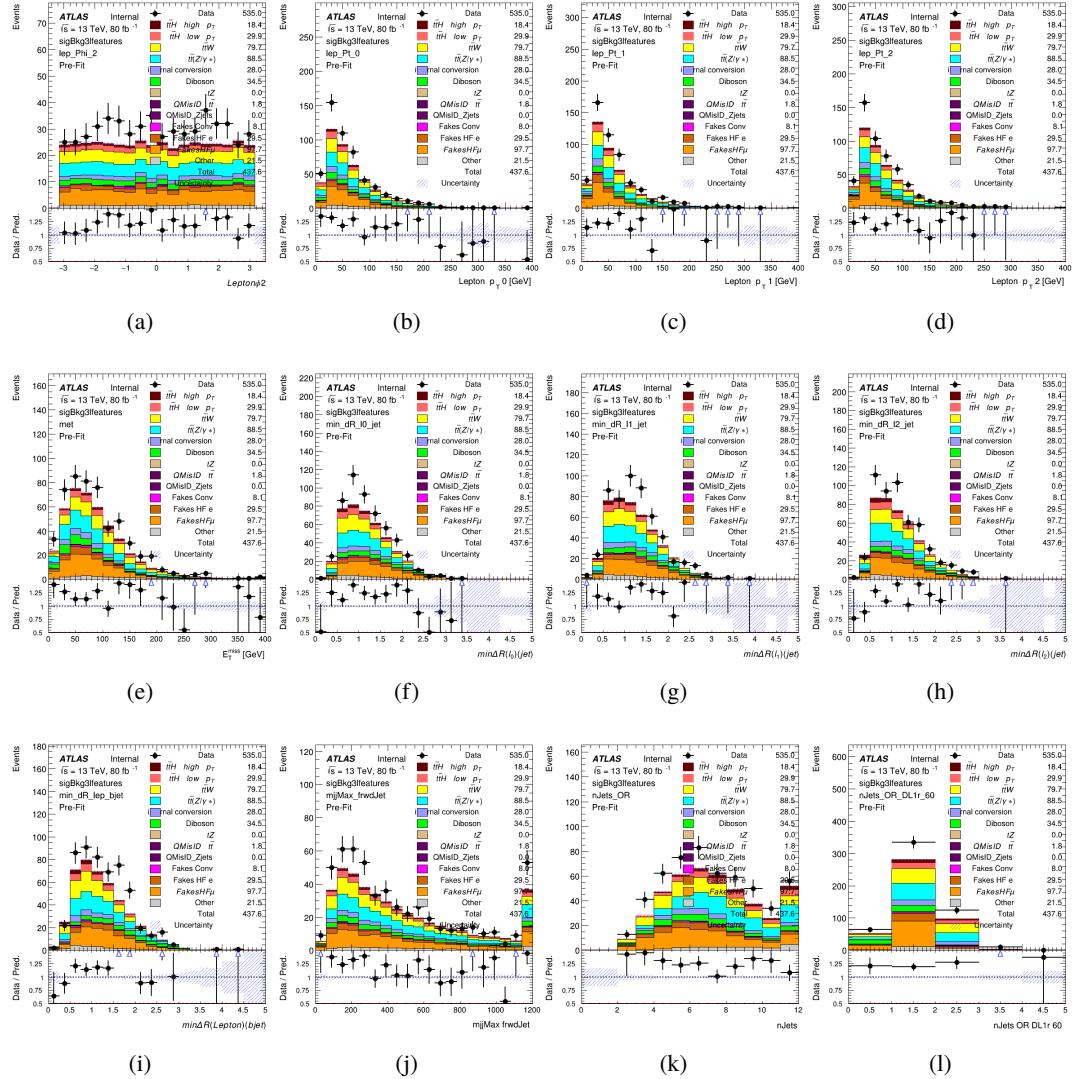


Figure A.24: Input features for sigBkg3l

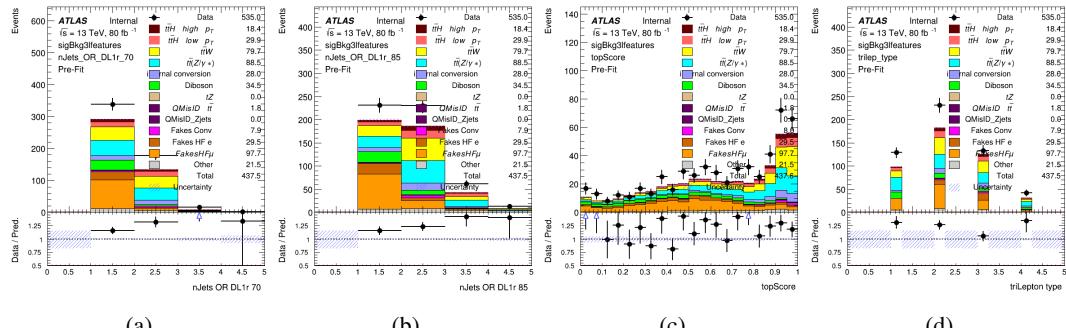


Figure A.25: Input features for sigBkg3l

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

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

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

Table 31: 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%

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

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

609 A model is built to determine whether $t\bar{t}H$ events include a high p_T (>150 GeV) or low p_T (<150
 610 GeV) Higgs Boson. While this is now a classification model, it uses the same input features
 611 described in section 5.4. Binary crossentropy is used as the loss function.

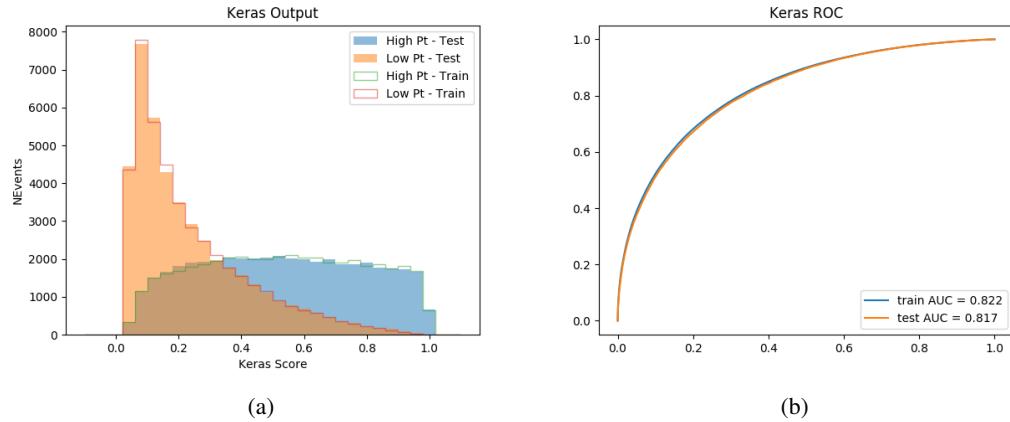


Figure A.26:

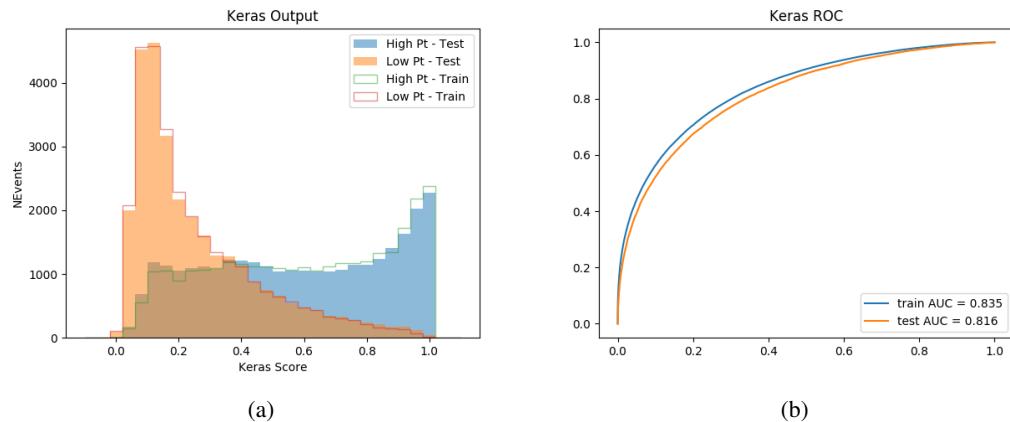


Figure A.27:

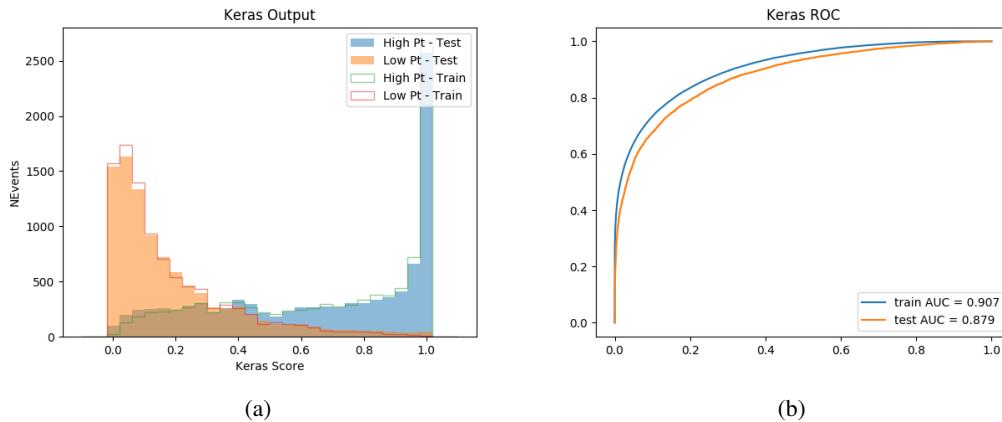


Figure A.28:

A.5 Impact of Alternative Jet Selection

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

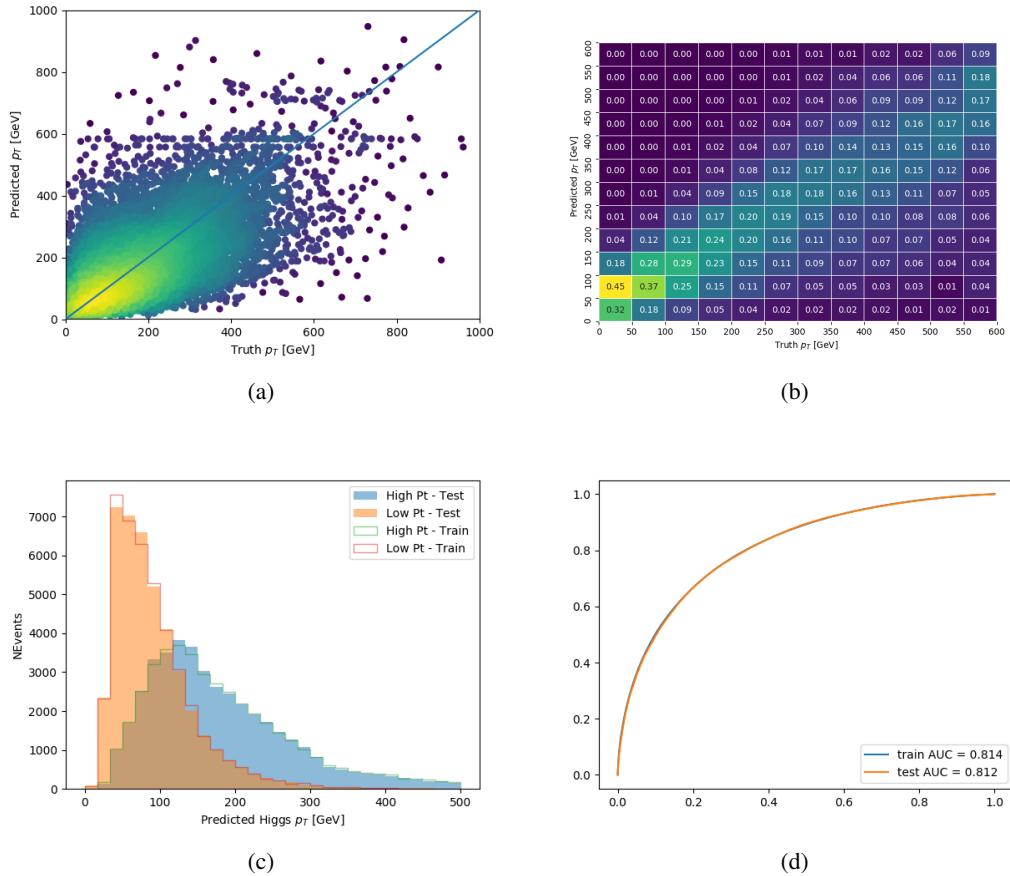


Figure A.29: