



ATLAS Note

GROUP-2017-XX

10th November 2020



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 © 2020 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	7
23	4.1 Trigger Requirements	7
24	4.2 Light Leptons	7
25	4.3 Jets	8
26	4.4 Missing Transverse Energy	9
27	5 Higgs Momentum Reconstruction	9
28	5.1 Decay Candidate Reconstruction	9
29	5.2 b-jet Identification	10
30	5.3 Higgs Reconstruction	13
31	5.4 p_T Prediction	21
32	5.5 3l Decay Mode	26
33	6 Signal Region Definitions	28
34	6.1 Pre-MVA Event Selection	28
35	6.2 Event MVA	29
36	6.3 Signal Region Definitions	34
37	7 Systematic Uncertainties	36
38	8 Results	38
39	8.1 Results - 80 fb^{-1}	38
40	8.2 Projected Results - 140 fb^{-1}	39
41	9 Conclusion	39
42	Appendices	44
43	A Machine Learning Models	44
44	A.1 Higgs Reconstruction Models	44
45	A.2 Background Rejection MVAs	59
46	A.3 Alternate b-jet Identification Algorithm	65
47	A.4 Binary Classification of the Higgs p_T	65

⁴⁹ **1 Changes and outstanding items**

⁵⁰ **1.1 Changelog**

⁵¹ This is version 1

52 2 Introduction

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

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

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

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

76 This note is organized as follows: Section ?? describes the LHC and the ATLAS detector. The
 77 dataset and Monte Carlo (MC) simulations used in the analysis is outlined in section 3. Section 4
 78 describes the identification and reconstruction of the various physics objects. The models used to
 79 reconstruct the momentum spectrum of the Higgs is discussed in section 5. The selection and
 80 categorisation of events comprises section 6, and the theoretical and experimental systematic
 81 uncertainties considered are described in section 7. Finally, the results of the study are summarized
 82 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 a modified version of
 94 AnalysisBase version 21.2.127.

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 of the following Good Run Lists:

- 99 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
 100 _PHYSGRL_All_Good_25ns.xml
 101 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
 102 _PHYSGRL_All_Good_25ns.xml
 103 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYSGRL
 104 _All_Good_25ns_Triggerno17e33prim.xml
 105 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYSGRL
 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

4 Object Reconstruction

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

4.1 Trigger Requirements

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

4.2 Light Leptons

Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that are associated with charged particle tracks reconstructed in the inner detector [[ATLAS-CONF-2016-024](#)]. Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the transition region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant combining shower shape and track information is used to distinguish prompt electrons from nonprompt leptons, such as those originating from hadronic showers.

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

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

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

127 Muon candidates are reconstructed by combining inner detector tracks with track segments or
128 full tracks in the muon spectrometer **[PERF-2014-05]**. Muon candidates are required to have
129 $p_T > 10$ GeV and $|\eta| < 2.5$. All leptons are required to be isolated, and pass a non-prompt BDT
130 selection described in detail in **[ttH_paper]**.

131 4.3 Jets

132 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
133 calorimeters **[ATL-PHYS-PUB-2015-015]**, using the anti- k_t algorithm with a radius parameter
134 $R = 0.4$. Jets with energy contributions likely arising from noise or detector effects are removed
135 from consideration **[ATLAS-CONF-2015-029]**, and only jets satisfying $p_T > 25$ GeV and
136 $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track association
137 algorithm is used to confirm that the jet originates from the selected primary vertex, in order to
138 reject jets arising from pileup collisions **[PERF-2014-03]**.

¹³⁹ **4.4 Missing Transverse Energy**

¹⁴⁰ Because all $t\bar{t}H$ – ML channels considered include multiple neutrinos, missing transverse energy
¹⁴¹ (E_T^{miss}) is present in each event. The missing transverse momentum vector is defined as the
¹⁴² inverse of the sum of the transverse momenta of all reconstructed physics objects as well as
¹⁴³ remaining unclustered energy, the latter of which is estimated from low- p_T tracks associated with
¹⁴⁴ the primary vertex but not assigned to a hard object [ATL-PHYS-PUB-2015-027].

¹⁴⁵ **5 Higgs Momentum Reconstruction**

¹⁴⁶ Reconstructing the momentum of the Higgs boson is a particular challenge for channels with
¹⁴⁷ leptons in the final state: Because all channels include at least two neutrinos in the final state, the
¹⁴⁸ Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted
¹⁴⁹ by a neural network when provided with the four-vectors of the Higgs Boson decay products, as
¹⁵⁰ shown in section 5.1. With this in mind, several layers of MVAs are used to reconstruction the
¹⁵¹ Higgs momentum.

¹⁵² The first layer is a model designed to select which jets are most likely to be the b-jets that came
¹⁵³ from the top decay, detailed in section 5.2. As described in section 5.3, the kinematics of these
¹⁵⁴ jets are fed into the second layer, which is designed to identify the decay products of the Higgs
¹⁵⁵ Boson itself. The kinematics of these particles are then fed into yet another neural-network, which
¹⁵⁶ predicts the momentum of the Higgs (5.4). MVAs are also used in the analysis to determine the
¹⁵⁷ decay of the Higgs boson in the 3l channel (5.5).

¹⁵⁸ For all of these models, the Keras neural network framework, with Tensorflow as the backend, is
¹⁵⁹ used, and the number of hidden layers and nodes are determined using grid search optimization.
¹⁶⁰ Each neural network uses the LeakyReLU activation function, a learning rate of 0.01, and the
¹⁶¹ Adam optimization algorithm, as alternatives are found to either decrease or have no impact on
¹⁶² performance. Batch normalization is applied after each layer. For the classification algorithms
¹⁶³ (b-jet matching, Higgs reconstruction, and 3l decay identification) binary-cross entropy is used as
¹⁶⁴ the loss function, while the p_T reconstruction algorithm uses MSE.

¹⁶⁵ The specific inputs features used for each model are arrived at through a process of trial and error
¹⁶⁶ - features considered potentially useful are tried, and those that are found to increase performance
¹⁶⁷ are included. While each model includes a relatively large number of features, some using
¹⁶⁸ upwards of 30, this inclusive approach is found to maximize the performance of each model while
¹⁶⁹ decreasing the variance compared to a reduced number of inputs. Each input feature is validated
¹⁷⁰ by comparing MC simulations to 80 fb^{-1} of data, as shown in the sections below.

¹⁷¹ **5.1 Decay Candidate Reconstruction**

¹⁷² Machine Learning algorithms are trained to identify the decay products of the Higgs Boson using
¹⁷³ MC simulations of $t\bar{t}H$ events. These include light leptons and jets. Reconstructed physics

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

177 Leptons considered as possible Higgs and top decay candidates are required to pass the selection
 178 described in section 4.2. For jets, however, it is found that a large fraction that originate from either
 179 the top decay or the Higgs decay fall outside the selection described in section 4.3. Specifically,
 180 jets from the Higgs decay tend to be soft, with 32% having $p_T < 25$ GeV. Therefore jets with
 181 $p_T < 15$ GeV are considered as possible candidates in the models described below. By contrast,
 182 less than 5% of the jets originating from the Higgs fall below this p_T . The jets are found to be
 183 well modeled even down to this low p_T threshold, as shown in section 6.1. The impact of using
 184 different p_T selection for the jet candidates is considered in detail in section ???. As they are
 185 expected to originate from the primary vertex, jets are also required to pass a JVT cut.

186 5.2 b-jet Identification

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

192 Exactly two b-jets are expected in the final state of $t\bar{t}H - ML$ events. However, in both the 3l and
 193 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r b-tag working
 194 point is used). Therefore, for events which have exactly one, or more than two, b-tagged jets,
 195 deciding which combination of jets correspond to the top decay is non-trivial. Further, events
 196 with 1 b-tagged jet represent just over half of all $t\bar{t}H - ML$ events. Of those, both b-jets are
 197 reconstructed by the detector 75% of the time. Therefore, rather than adjusting the selection to
 198 require exactly 2 b-tagged jets, and losing more than half of the signal events, a neural network is
 199 used to predict which pair of jets is most likely to correspond to truth b-jets.

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

202 5.2.1 2lSS Channel

203 For the 2lSS channel, the input features shown in table 3 are used for training. Here j_0 and j_1 are
 204 the two jet candidates, while l_0 and l_1 are the two leptons in the event, ordered by p_T . jet DL1r is
 205 an integer corresponding to the calibrated b-tagging working points reached by each jet, where 5
 206 represents the tightest working point and 1 represents the loosest. The variables nJets DL1r 60%
 207 and nJets DL1r 85% represent the number of jets in the event passing the 60% and 85% b-tag
 208 working points, respectively.

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

Table 3: Input features used in the 2ISS b-jet identification algorithm

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

213 The difference between the distributions for a few of these features for the correct (i.e. both jets
 214 are truth b-jets), and incorrect combinations are shown in figure 5.1. The correct and incorrect
 215 contributions are scaled to the same integral, so as to better demonstrate the differences in the
 216 distributions.

217 The modeling of these inputs is validated against data, with figure 5.2 showing good general
 218 agreement between data and MC. Plots for the complete list of features can found in section A.

219 Based on the results of grid search evaluation, the optimal architecture is found to include 5
 220 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as overfitting
 221 is found to not be an issue. The output score distribution as well as the ROC curve for the trained
 222 model are shown in figure 5.2.1. The model is found to identify the correct pairing of jets for
 223 73% of 2ISS signal events on test data.

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

228 The accuracy of the model for different values of n-bjets, compared to this naive approach, is
 229 shown in table 4.

230 5.2.2 3l Channel

231 The input features used in the 3l channel are listed in table 5, with the same naming convention as
 232 the 2ISS channel.

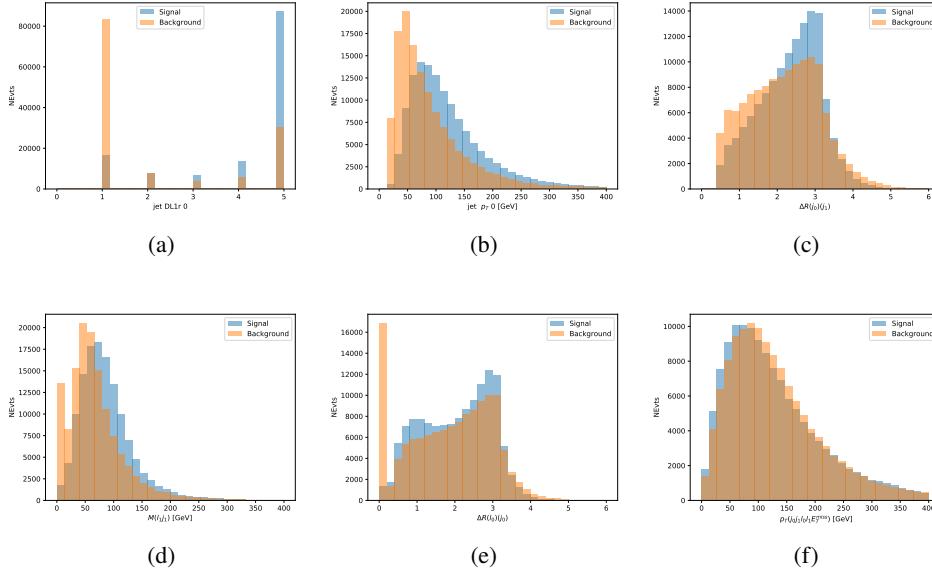


Figure 5.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

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

233 A few of these features are shown in figure 5.4, comparing the distributions for correct and incorrect
234 combinations of jets.

235 The modeling of these inputs is validated against data, with figure 5.5 showing good general
236 agreement between data and MC. Plots for the complete list of features can be found in section A.

237 Again, the dataset is downsized to reduce the ratio of correct and incorrect combination from 20:1,
238 to 5:1. Around 7 million events are used for training, with 10% set aside for testing. Based on the
239 results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with
240 60 nodes each. The output score distribution as well as the ROC curve for the trained model are
241 shown in figure 5.2.2.

242 This procedure is found to identify the correct pairing of jets for nearly 80% of 31 signal events.
243 The accuracy of the model is summarized in table 6.

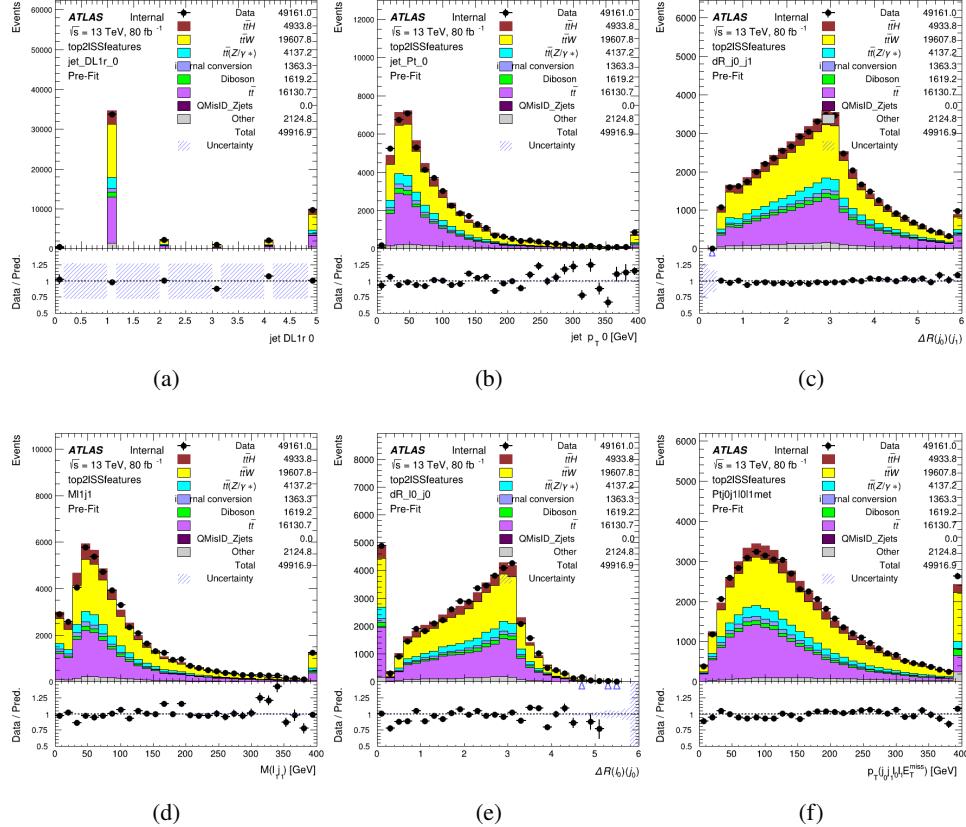


Figure 5.2: Data/MC comparisons of input features for top2lSS training for 80 fb^{-1} of data.

5.3 Higgs Reconstruction

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

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

3l events are therefore categorized as either semi-leptonic or fully-leptonic. In the semi-leptonic case the reconstructed decay products consist of two jets and a single lepton. For the fully-leptonic case, the decay products include 2 of the three leptons associated with the event. For training the models, events are separated into these two categories using truth level information. A separate MVA, described in section 5.5, is used to make this distinction at reco level and determine which model to use.

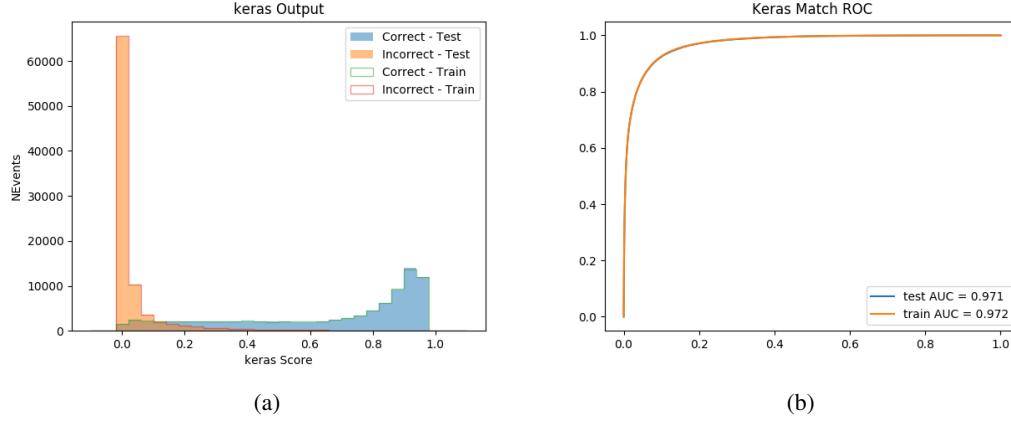


Figure 5.3: (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

Table 5: Input features

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

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

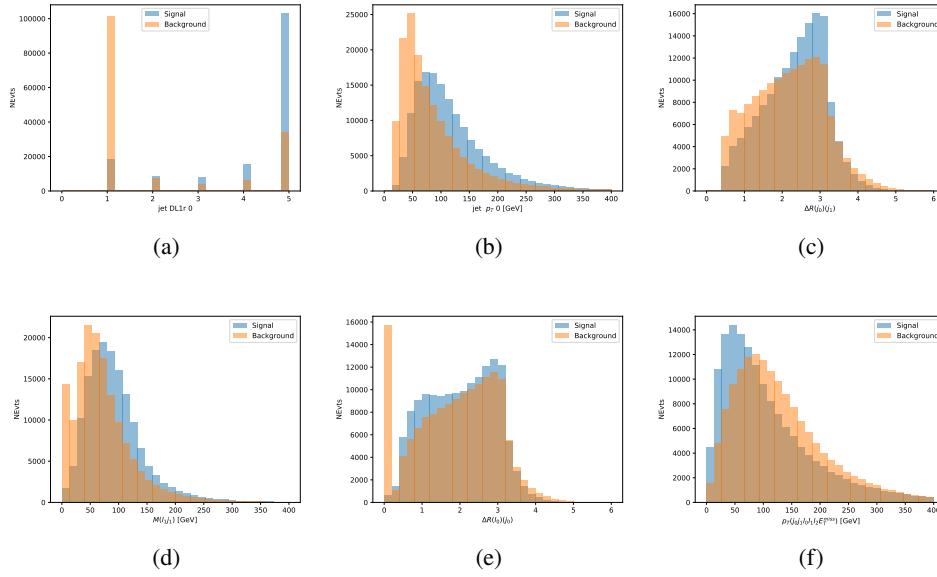


Figure 5.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

262 5.3.1 2ISS Channel

263 For the 2ISS channel, the Higgs decay products include one light lepton and two jets. The neural
 264 network is trained on the kinematics of different combinations of leptons and jets, as well as the
 265 b-jets identified in section 5.2, with the specific input features listed in table ??.

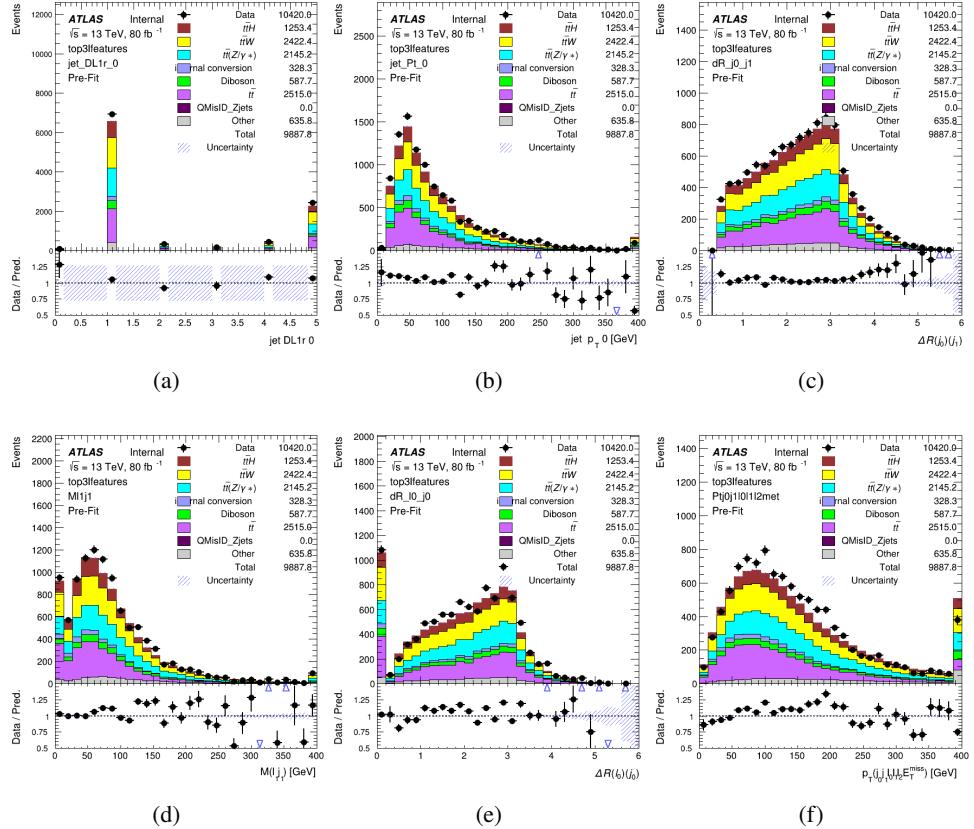
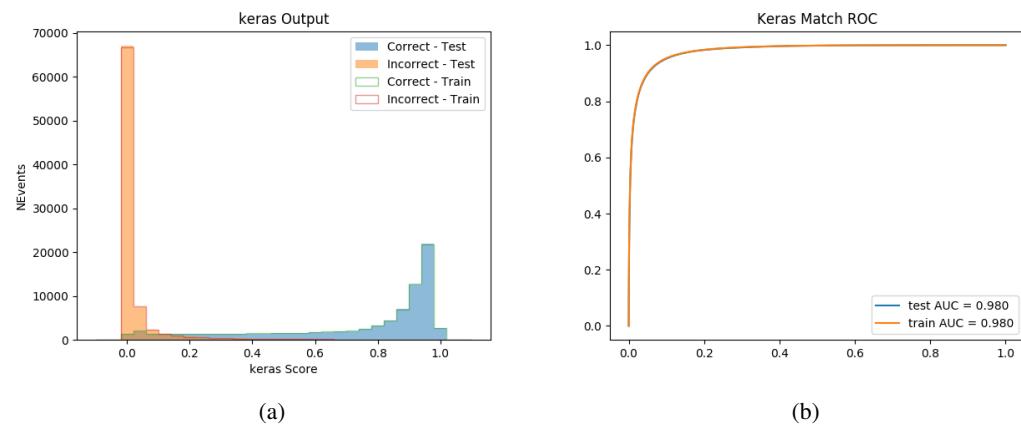
Figure 5.5: Data/MC comparisons of input features for top3l training for 80 fb^{-1} of data.

Figure 5.6: tmp

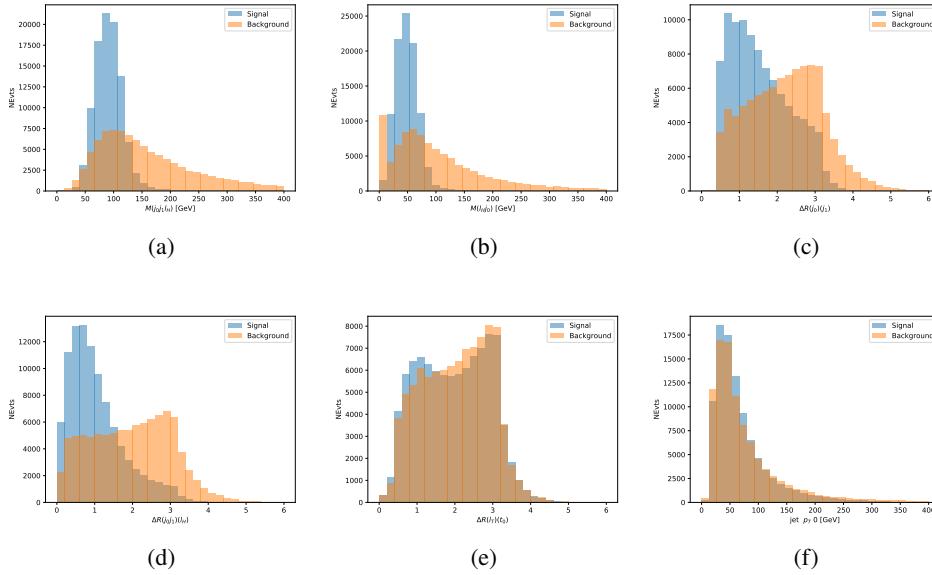


Figure 5.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

- 266 The modeling of these inputs is validated against data, with figure 5.2 showing good general
267 agreement between data and MC. Plots for the complete list of features can found in section A.
268 The neural network identifies the correct combination 55% of the time. It identifies the correct
269 lepton 85% of the time, and selects the correct lepton and at least one of the correct jets 81% of
270 the time.

271 5.3.2 3l Semi-leptonic Channel

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

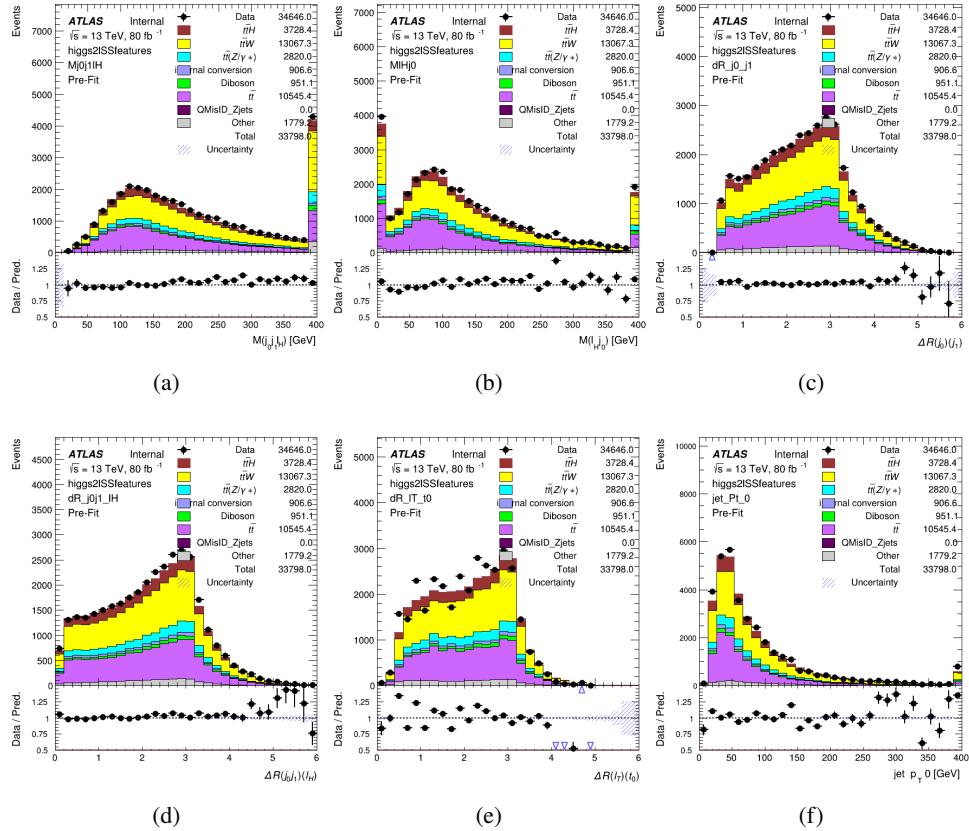


Figure 5.8: Data/MC comparisons of input features for higgs2lSS training for 80 fb^{-1} of data.

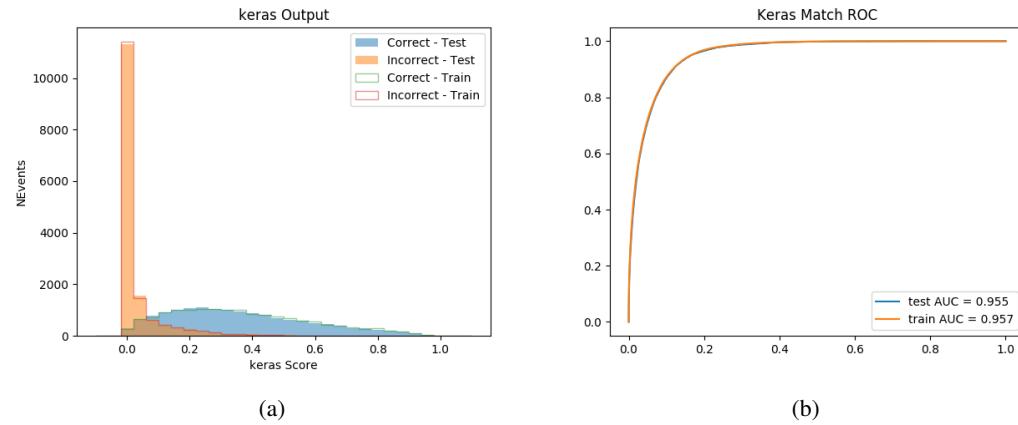


Figure 5.9: (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

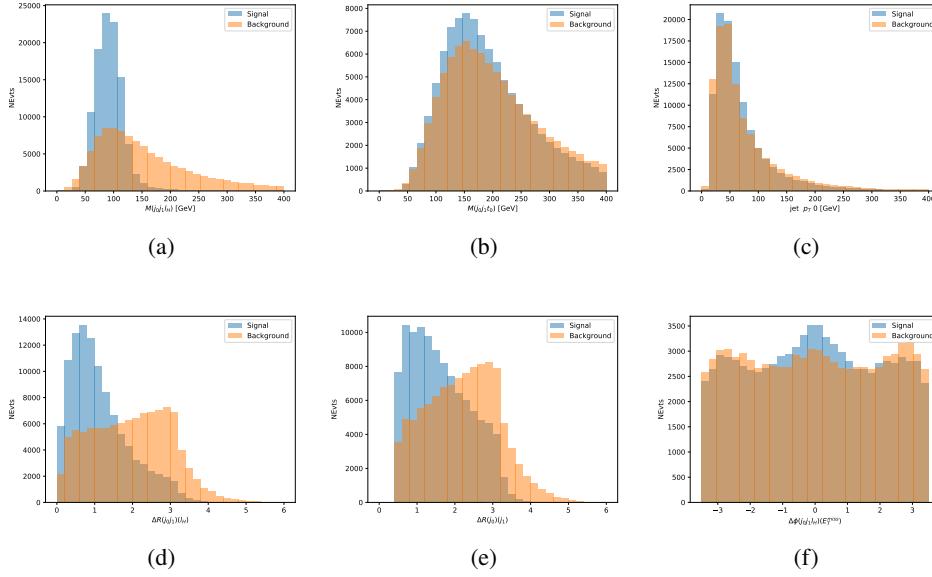


Figure 5.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

- 276 The modeling of these inputs is validated against data, with figure 5.11 showing good general
277 agreement between data and MC. Plots for the complete list of features can found in section A.
278 The neural network identifies the correct combination 65% of the time. It identifies the correct
279 lepton 85% of the time, an selects the correct lepton and at least one of the correct jets 83% of the
280 time.

281 5.3.3 3l Fully-leptonic Channel

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

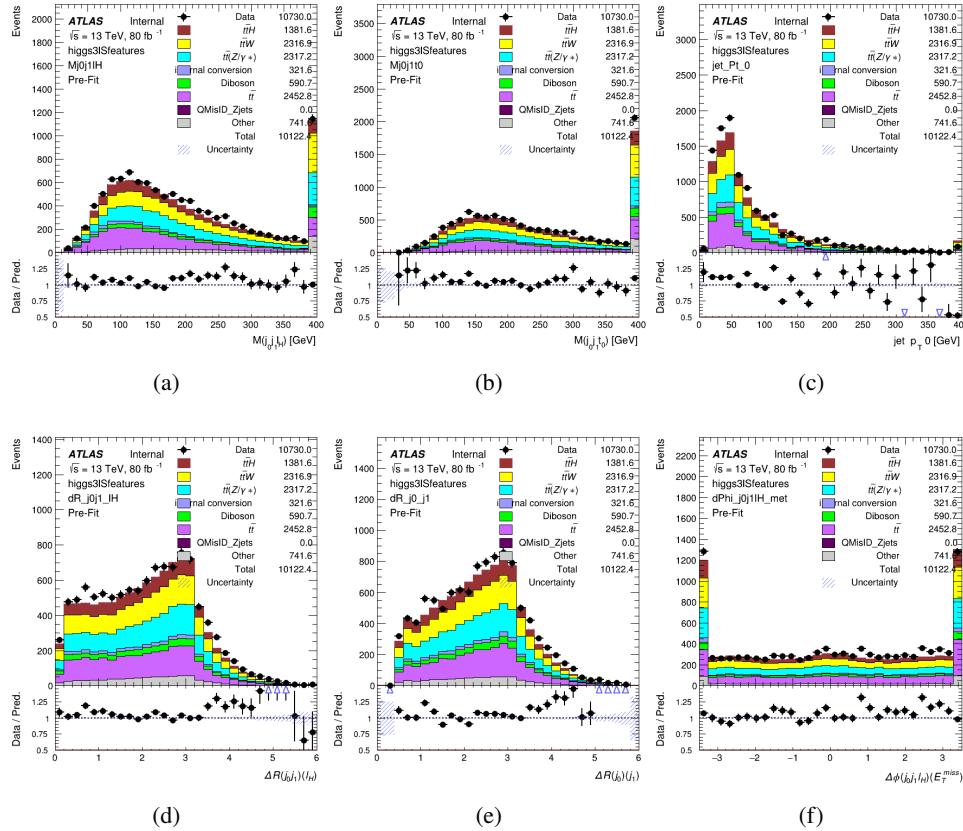


Figure 5.11: Data/MC comparisons of input features for higgs3IS training for 80 fb^{-1} of data.

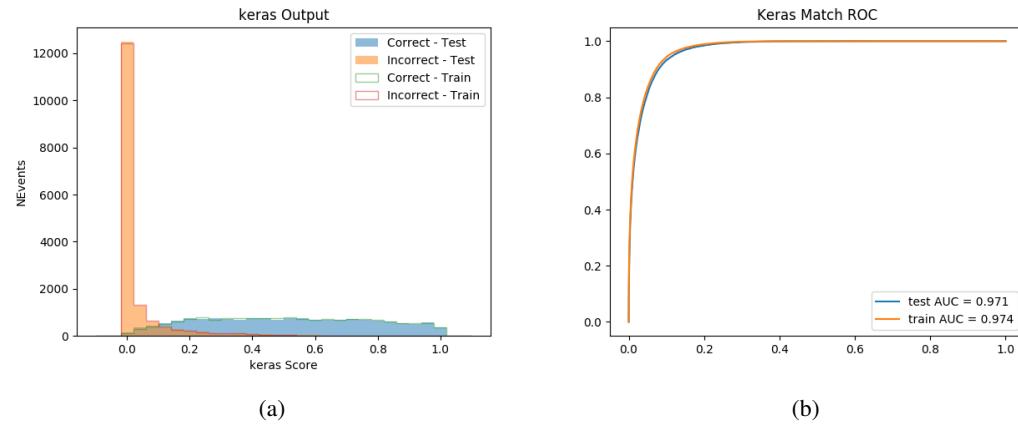


Figure 5.12: (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

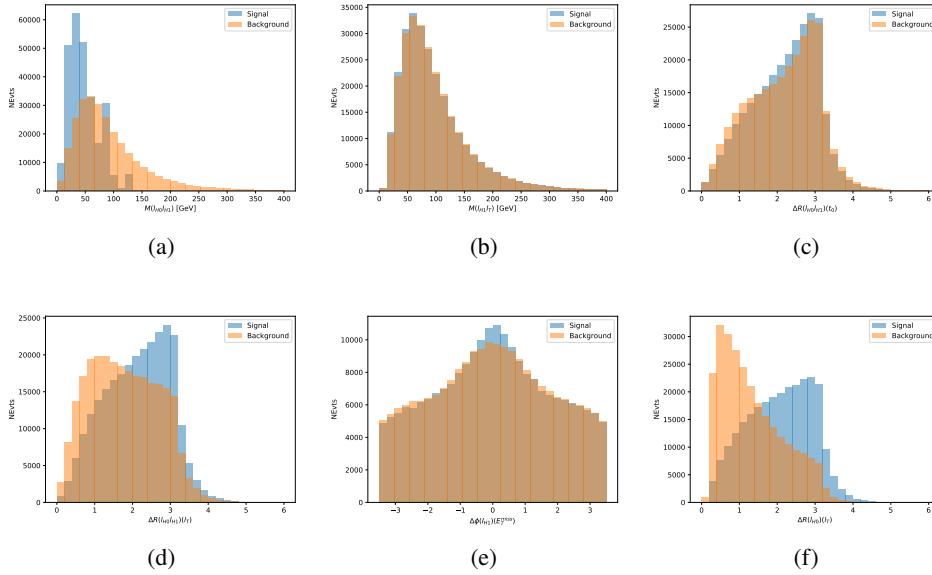


Figure 5.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

287 The modeling of these inputs is validated against data, with figure 5.14 showing good general
 288 agreement between data and MC. Plots for the complete list of features can found in section A.

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

290 5.4 p_T Prediction

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

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

300 Because the analysis uses a two bin fit of the Higgs p_T , the momentum reconstruction could be
 301 treated as a binary classification problem, rather than a regression problem. This approach is
 302 explored in detail in section A.4, and is found not to provide any significant increase in sensitivity.
 303 The regression approach is used because it provides more flexibility for future analyses, as it is

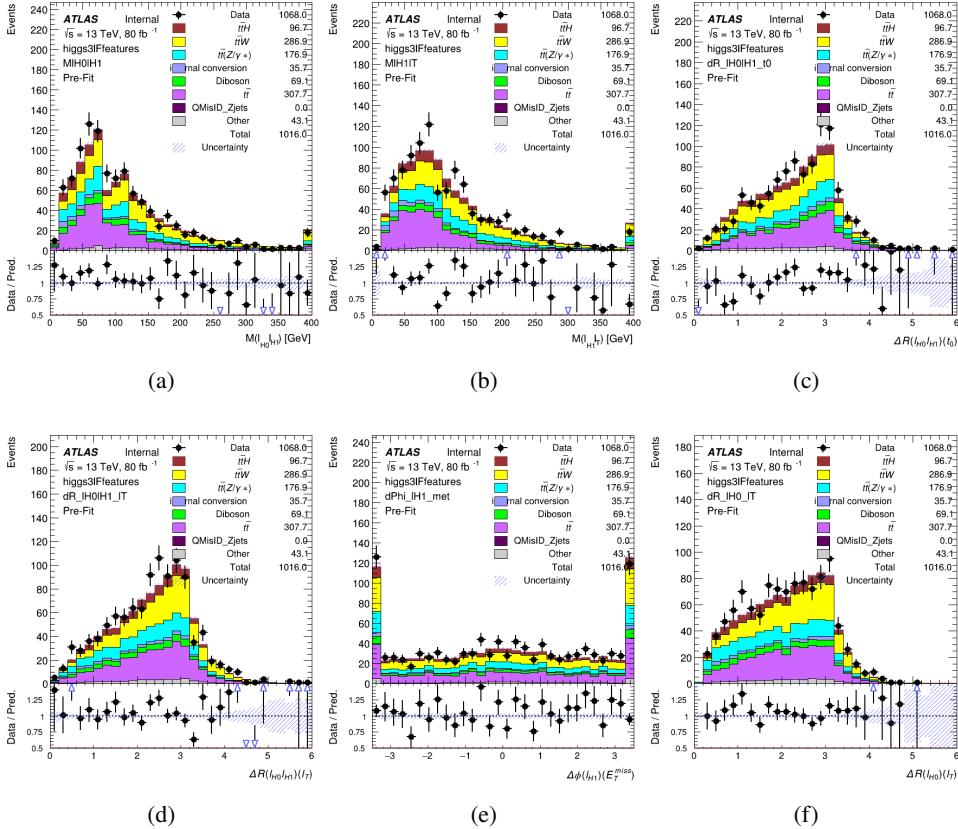


Figure 5.14: Data/MC comparisons of input features for higgs3IF training for 80 fb^{-1} of data.

independent of the cutoff between high and low p_T , as well as the number of bins. Further, a regression allows the output of the neural network to be more clearly understood, as it can be directly compared to a physics observable.

5.4.1 2ISS Channel

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

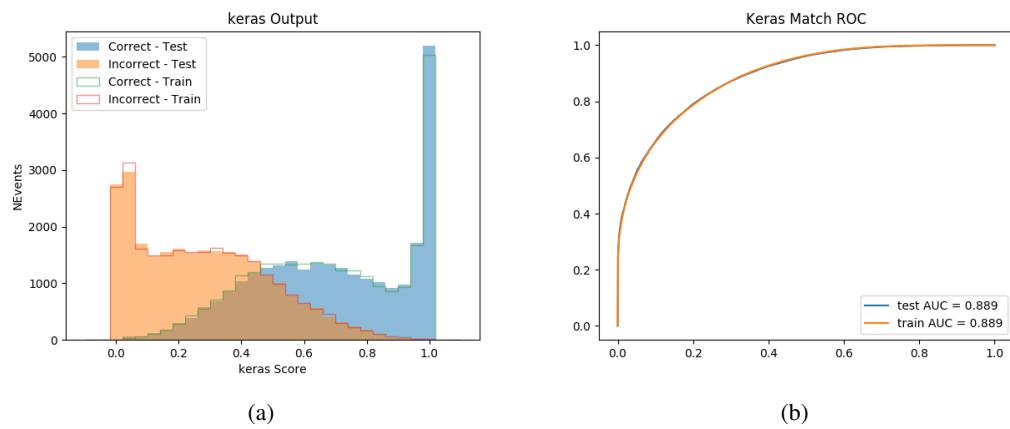


Figure 5.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

313 The optimal neural network architecture for this channel is found to consist of 5 hidden layers with
 314 40 nodes each. The input data set includes 1.2 million events, 10% of which is used for testing,
 315 the other 90% for training. Training is found to converge after around 150 epochs.

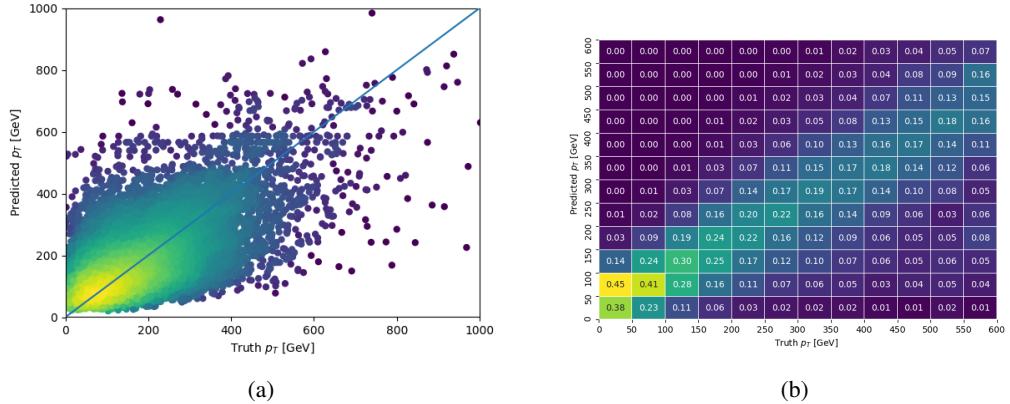


Figure 5.16:

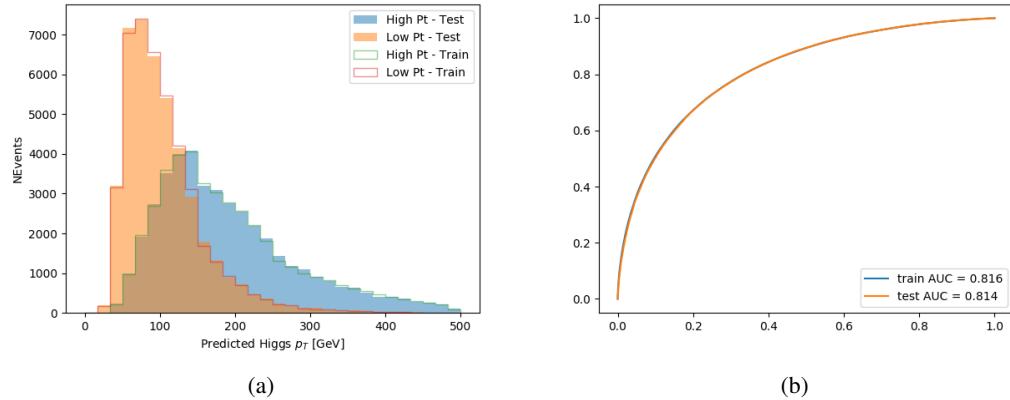


Figure 5.17:

316 5.4.2 3l Semi-leptonic Channel

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

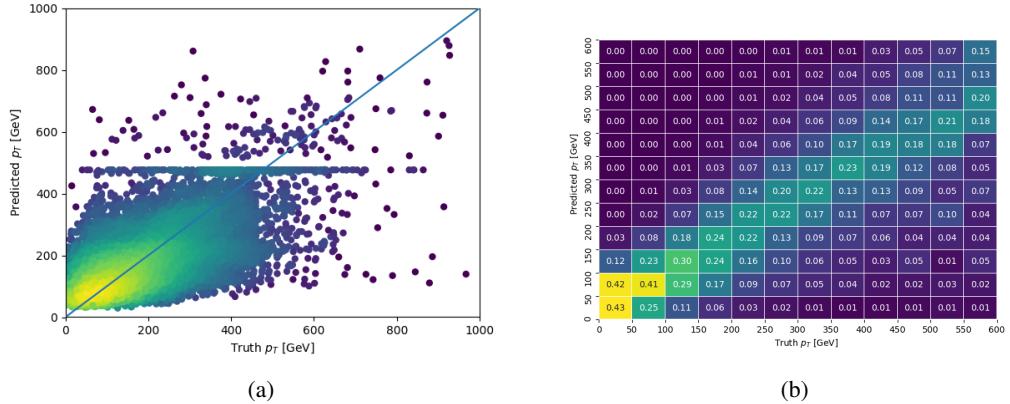


Figure 5.18:

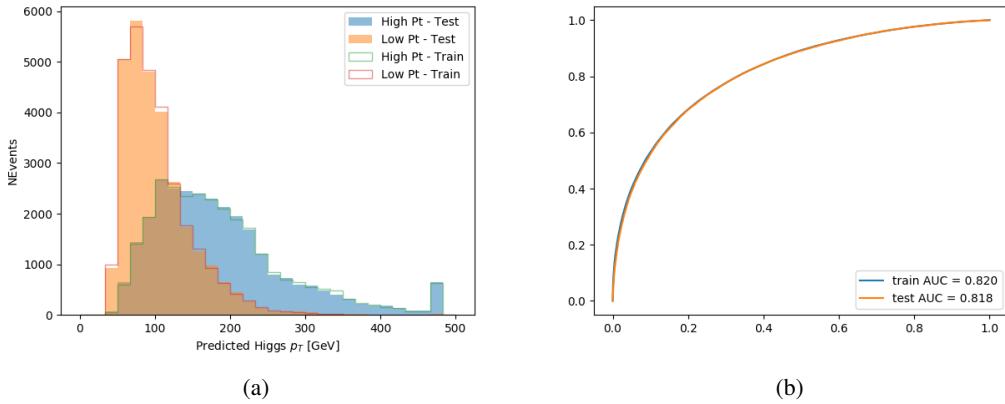


Figure 5.19:

320 **5.4.3 3l Fully-leptonic Channel**

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

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

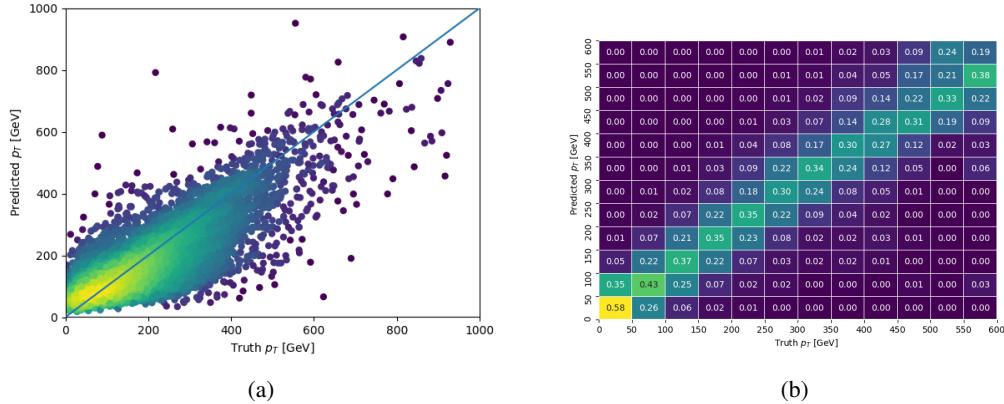


Figure 5.20:

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

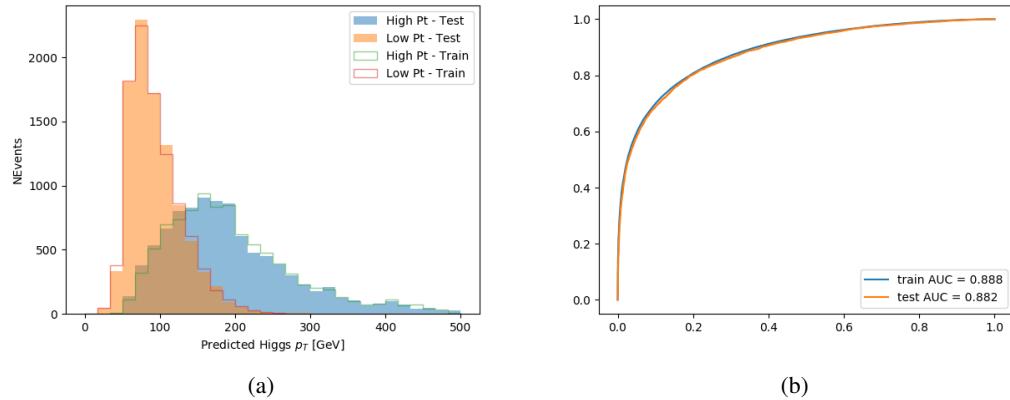


Figure 5.21:

326 5.5 3l Decay Mode

327 In the 3l channel, there are two possible ways for the Higgs to decay, both involving intermediate
 328 W boson pairs: Either both W bosons decay leptonically, in which case the reconstructed decay

329 consists of two leptons (referred as the fully-leptonic 3l channel), or one W decays leptonically
330 and the other hadronically, giving two jets and one lepton in the final state (referred to as the
331 semi-leptonic 3l channel). In order to accurately reconstruct the Higgs, it is necessary to identify
332 which of these decays took place for each 3l event.

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

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

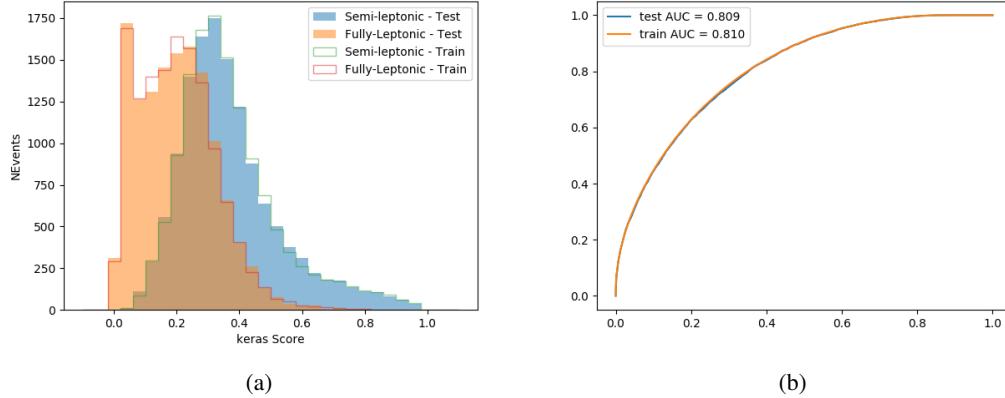


Figure 5.22:

338 6 Signal Region Definitions

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

342 6.1 Pre-MVA Event Selection

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

- 345 • Two very tight, same-charge, light leptons with $p_T > 20 \text{ GeV}$
- 346 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets
- 347 • No reconstructed tau candidates

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

- 349 • Three light leptons with total charge ± 1
- 350 • Same charge leptons are required to be very tight, with $p_T > 20 \text{ GeV}$
- 351 • Opposite charge lepton must be loose, with $p_T > 10 \text{ GeV}$

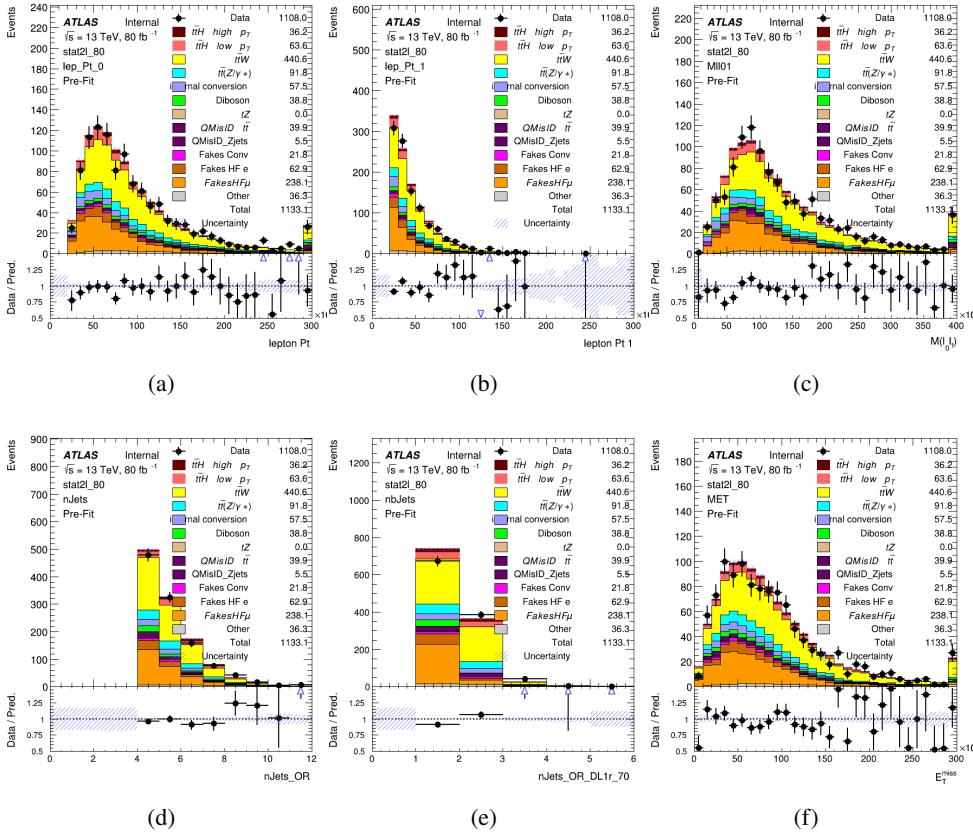


Figure 6.1:

- ≥ 2 reconstructed jets, ≥ 1 b-tagged jets
- No reconstructed tau candidates
- $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs

6.2 Event MVA

Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost [xgboost] are trained using the kinematics of signal and background events derived from Monte Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo event. Because the background composition differs for events with a high reconstructed Higgs p_T compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and

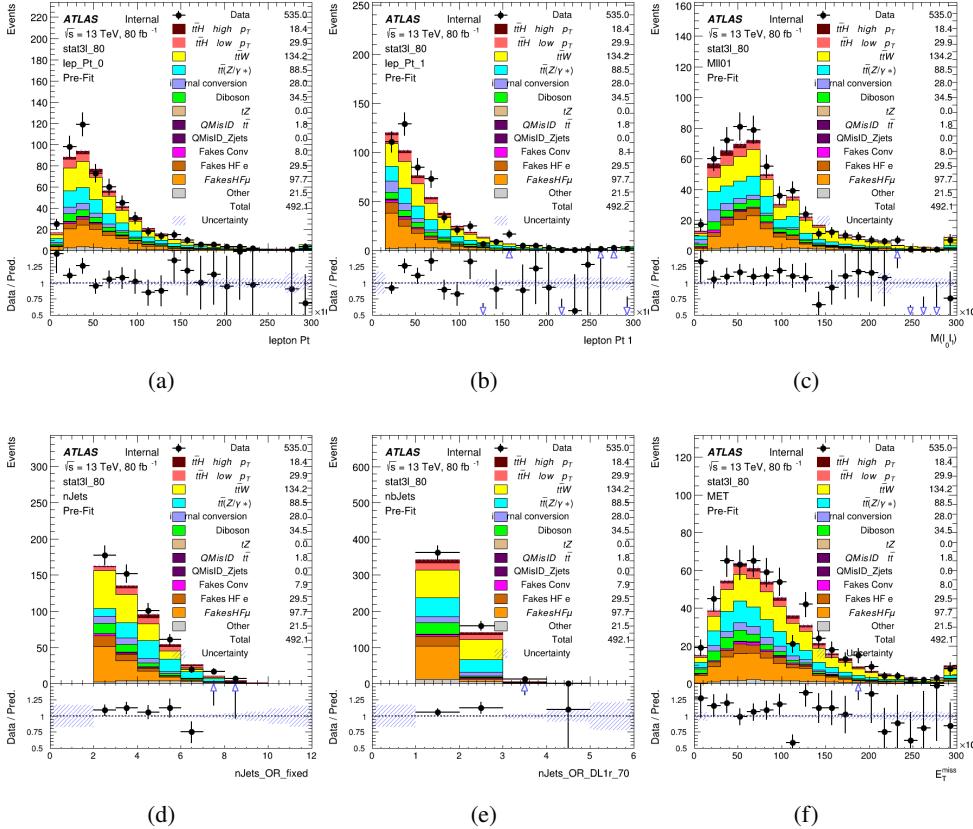


Figure 6.2:

363 low p_T regions. This is found to provide better significance than attempting to build an inclusive
 364 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6
 365 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
 366 3lF high p_T , and 3lF low p_T .

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

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

Table 7: Input features

³⁶⁸ While for each of the 31 BDTs, the features listed below are used for training:

$M(\text{lep}, E_T^{\text{miss}})$	$M(l_0 l_1)$	$M(l_0 l_1 l_2)$
$M(l_0 l_2)$	$M(l_1 l_2)$	$\text{binHiggs } p_T \text{ 3lF}$
$\Delta R(l_0)(l_1)$	$\Delta R(l_0)(l_2)$	$\Delta R(l_0)(l_2)$
$\Delta R(l_1)(l_2)$	decayScore	higgsRecoScore3lF
higgsRecoScore3lS	$\text{jet } \eta \ 0$	$\text{jet } \eta \ 1$
$\text{jet } \phi \ 0$	$\text{jet } \phi \ 1$	$\text{jet } p_T \ 0$
$\text{jet } p_T \ 1$	$\text{Lepton } \eta \ 0$	$\text{Lepton } \eta \ 1$
$\text{Lepton } \eta \ 2$	$\text{Lepton } \phi \ 0$	$\text{Lepton } \phi \ 1$
$\text{Lepton } \phi \ 2$	$\text{Lepton } p_T \ 0$	$\text{Lepton } p_T \ 1$
$\text{Lepton } p_T \ 2$	E_T^{miss}	$\min \Delta R(l_0)(\text{jet})$
$\min \Delta R(l_1)(\text{jet})$	$\min \Delta R(l_2)(\text{jet})$	$\min \Delta R(\text{Lepton})(\text{bjet})$
$mjj\text{Max frwdJet}$	$n\text{Jets}$	$n\text{Jets OR DL1r } 60$
$n\text{Jets OR DL1r } 70$	$n\text{Jets OR DL1r } 85$	signal
topScore	triLepton type	weight

Table 8: Input features

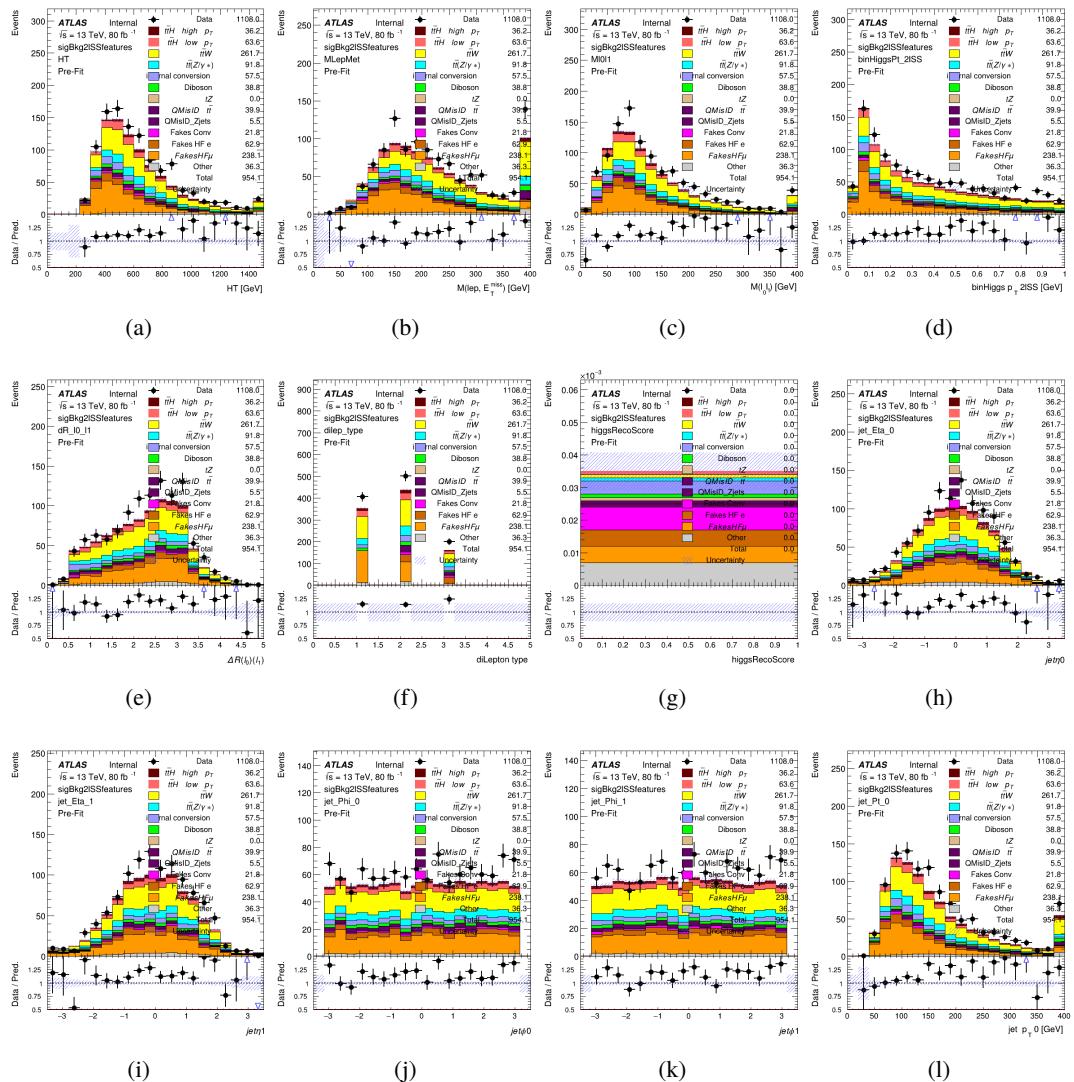


Figure 6.3:

- 369 The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss function.
370 Output distributions of each MVA are shown in figure 6.2.

371 6.3 Signal Region Definitions

- Once pre-selection has been applied, channels are further refined based on the MVAs described above. The output of the model described in section 5.5 is used to separate the three channel into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson.

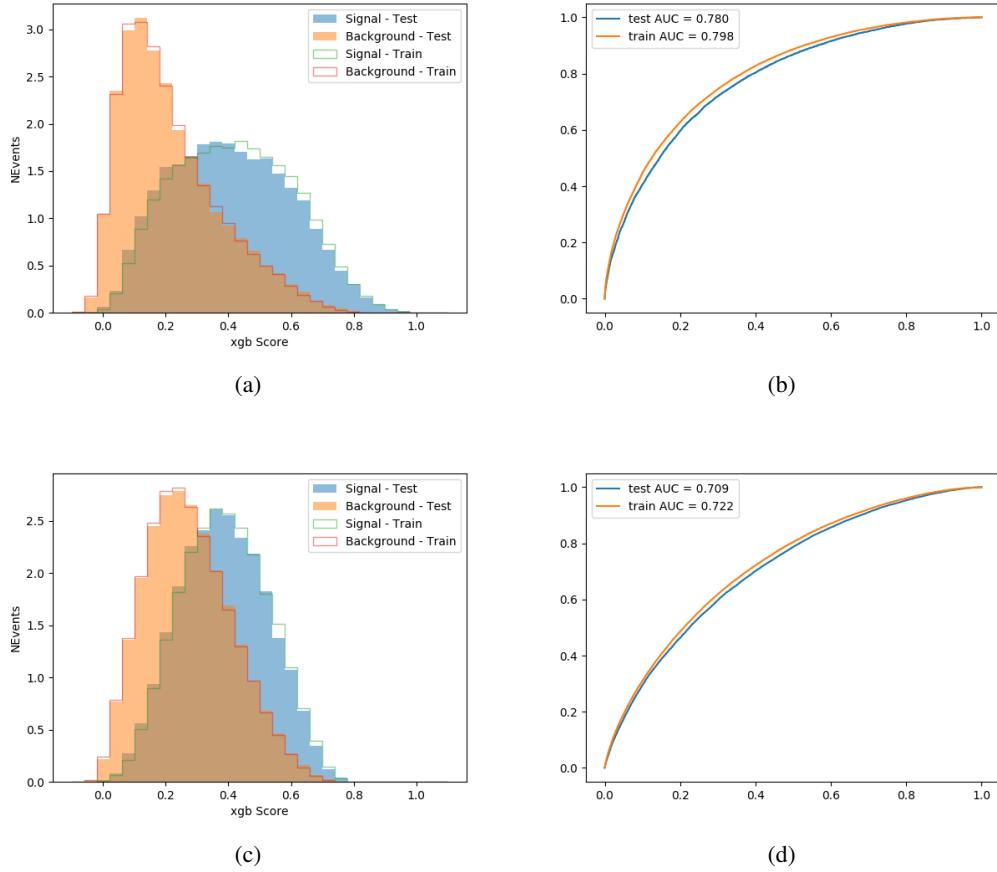


Figure 6.4:

375 For each event, depending on the channel as well as the predicted p_T of the Higgs derived from
 376 the algorithm described in section 5.4, a cut on the appropriate background rejection algorithm is
 377 applied. The specific selection used, and the event yield in each channel after this selection has
 378 been applied, is summarized below.

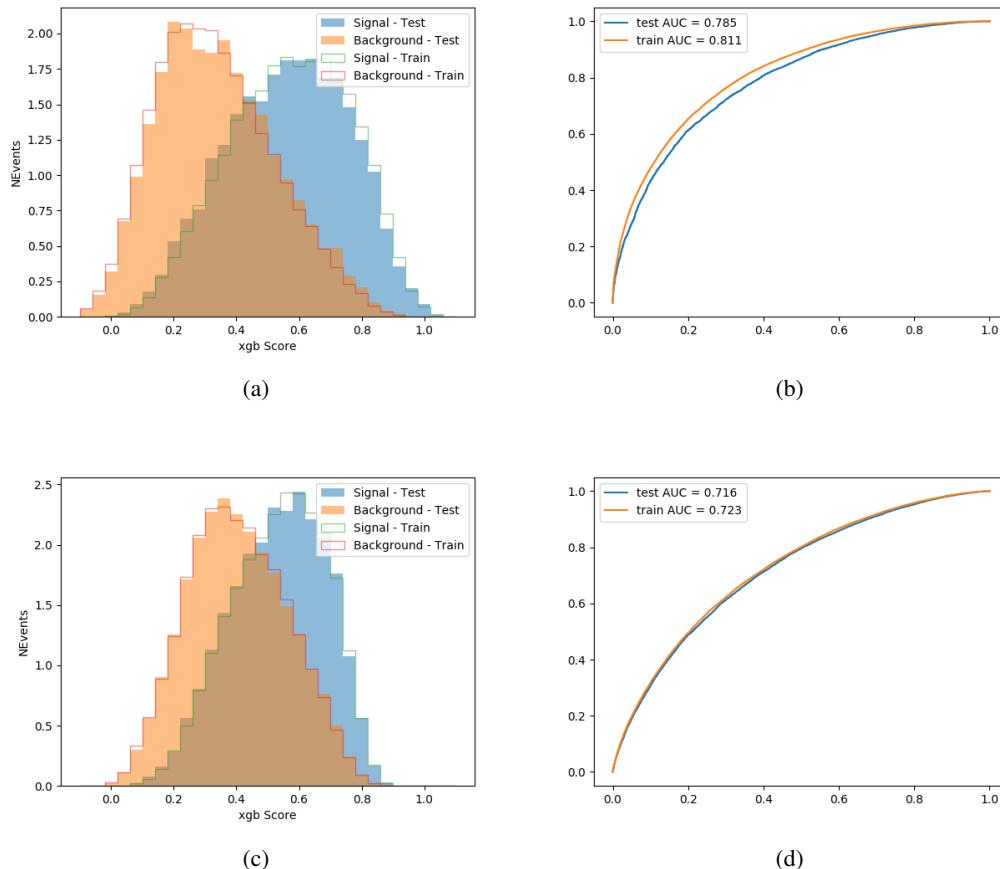


Figure 6.5:

379 **6.3.1 2lSS**380 **6.3.2 3l – Semi – leptonic**381 **6.3.3 3l – Fully – leptonic**382 **7 Systematic Uncertainties**

383 The systematic uncertainties that are considered are summarized in table ???. These are imple-
 384 mented in the fit either as a normalization factors or as a shape variation or both in the signal
 385 and background estimations. The numerical impact of each of these uncertainties is outlined in
 386 section 8.

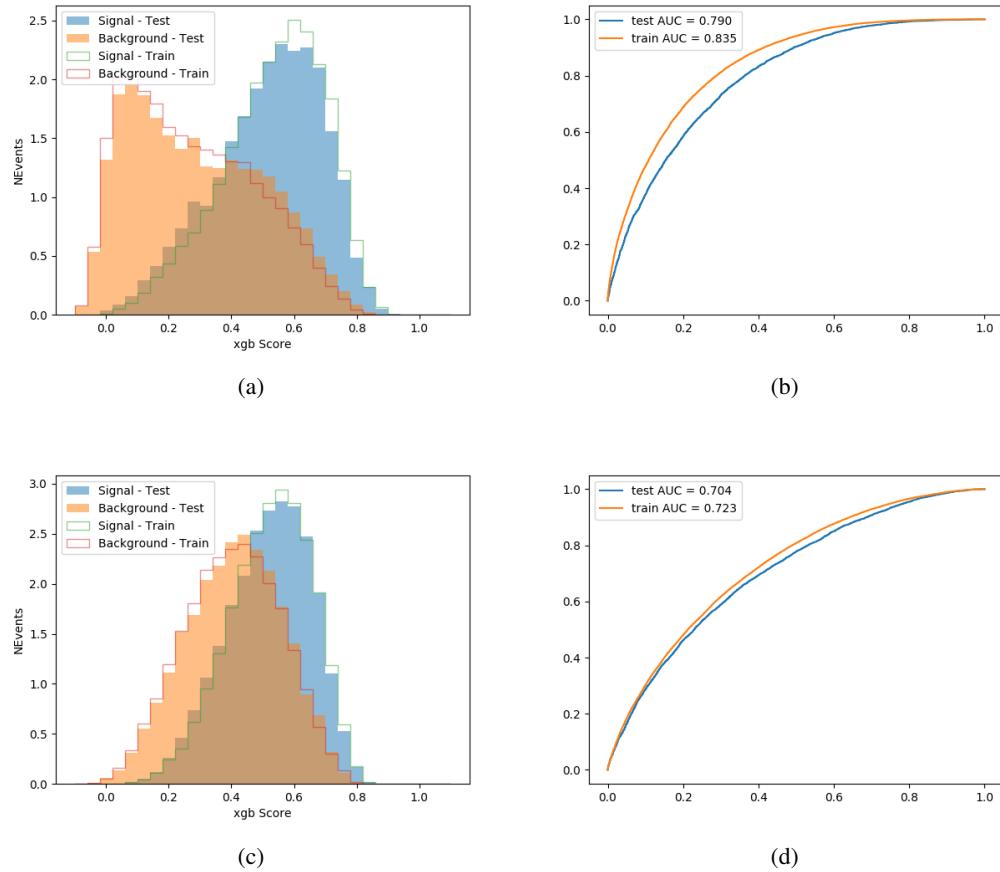


Figure 6.6:

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

390 The experimental uncertainties are related to the reconstruction and identification of light leptons
 391 and b-tagging of jets, and to the reconstruction of E_T^{miss} . The sources which contribute to the
 392 uncertainty in the jet energy scale [**jes**] are decomposed into uncorrelated components and treated
 393 as independent sources in the analysis.

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

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

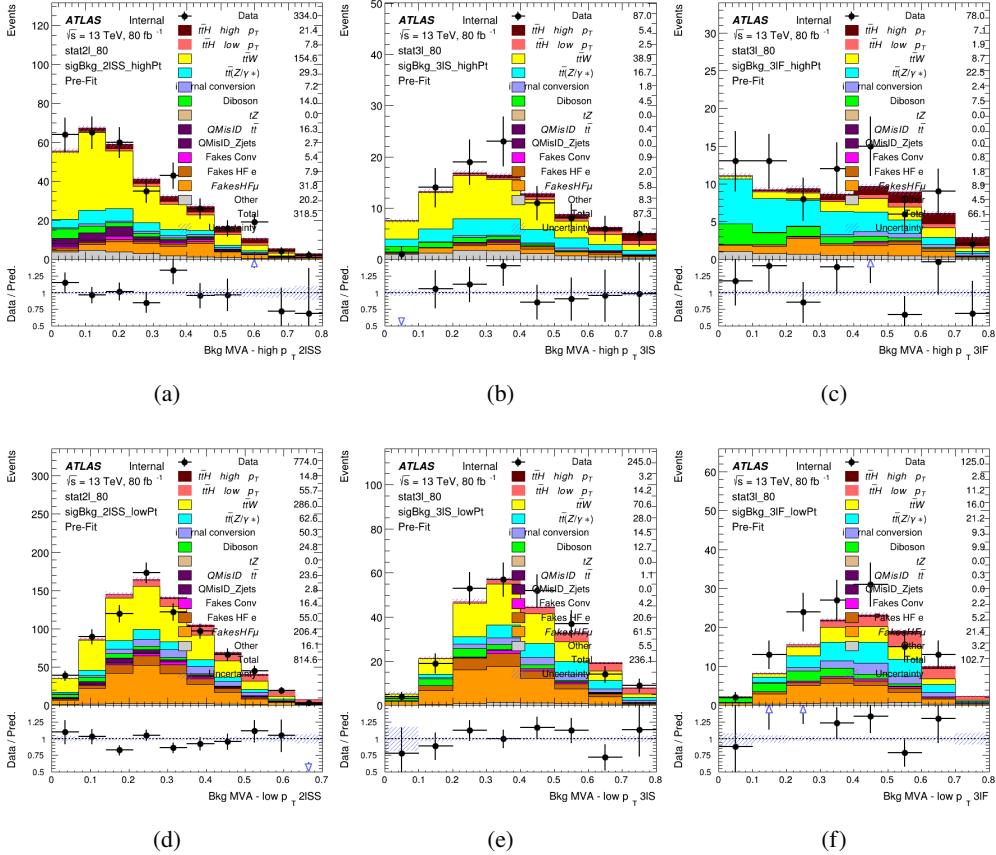


Figure 6.7: scores

399 8 Results

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

402 8.1 Results - 80 fb^{-1}

403 A maximum likelihood fit is performed simultaneously over the regions shown in figure 8.1.

Table 9: 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

404 8.2 Projected Results - 140 fb^{-1}

405 9 Conclusion

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

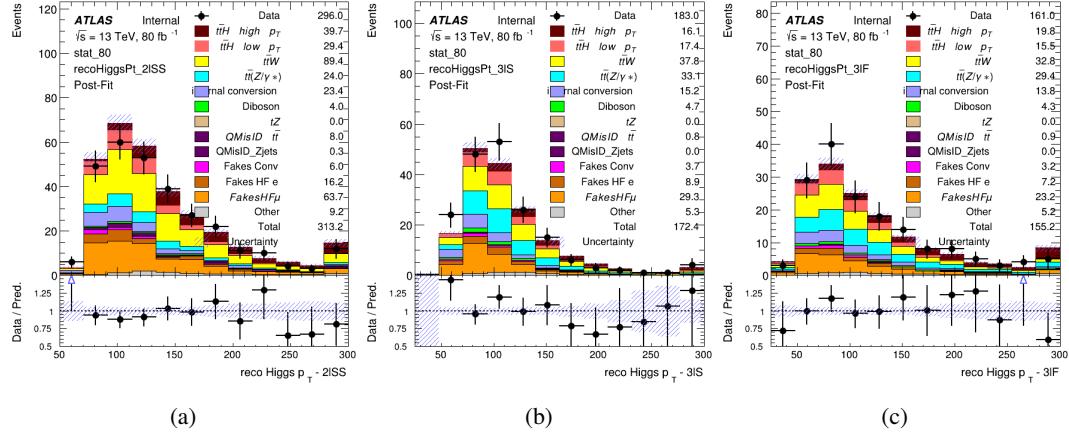


Figure 8.1:

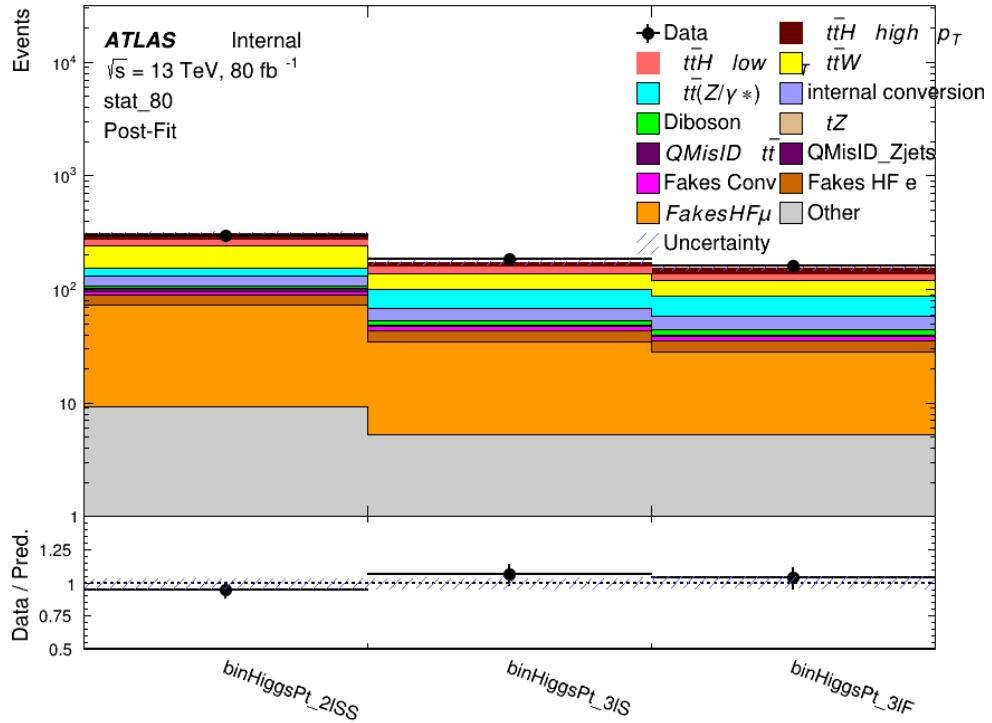


Figure 8.2: Post-fit summary of fit.

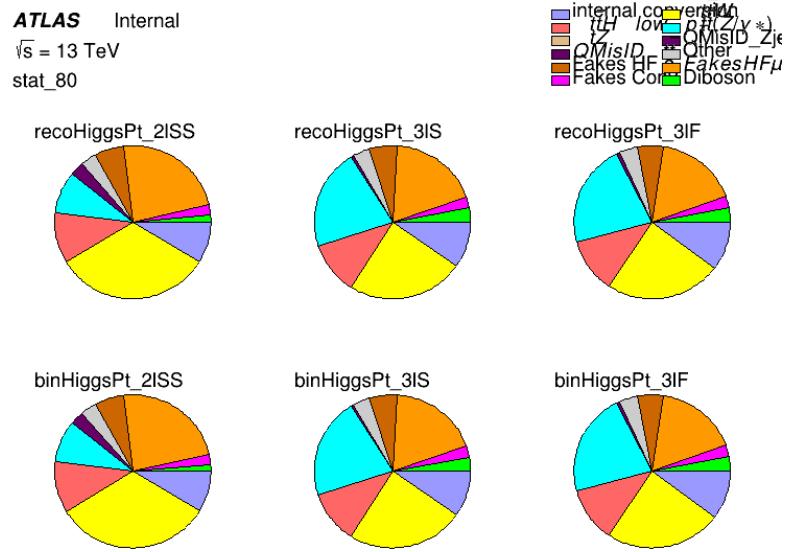


Figure 8.3: Background composition of the fit regions.

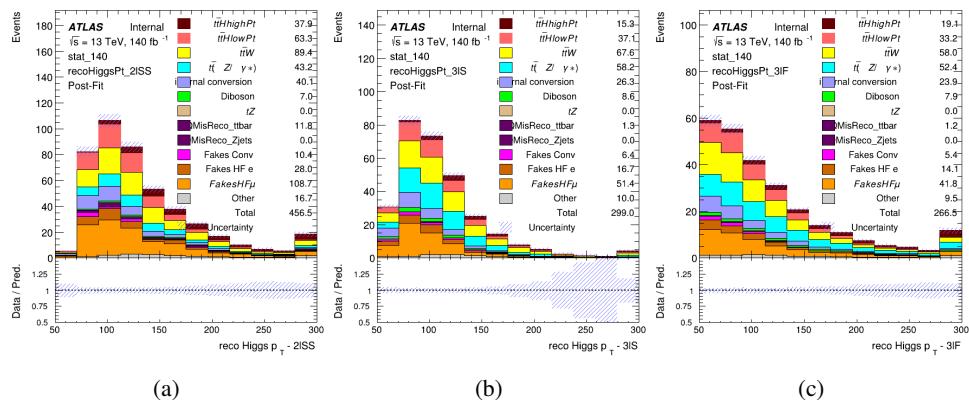


Figure 8.4:

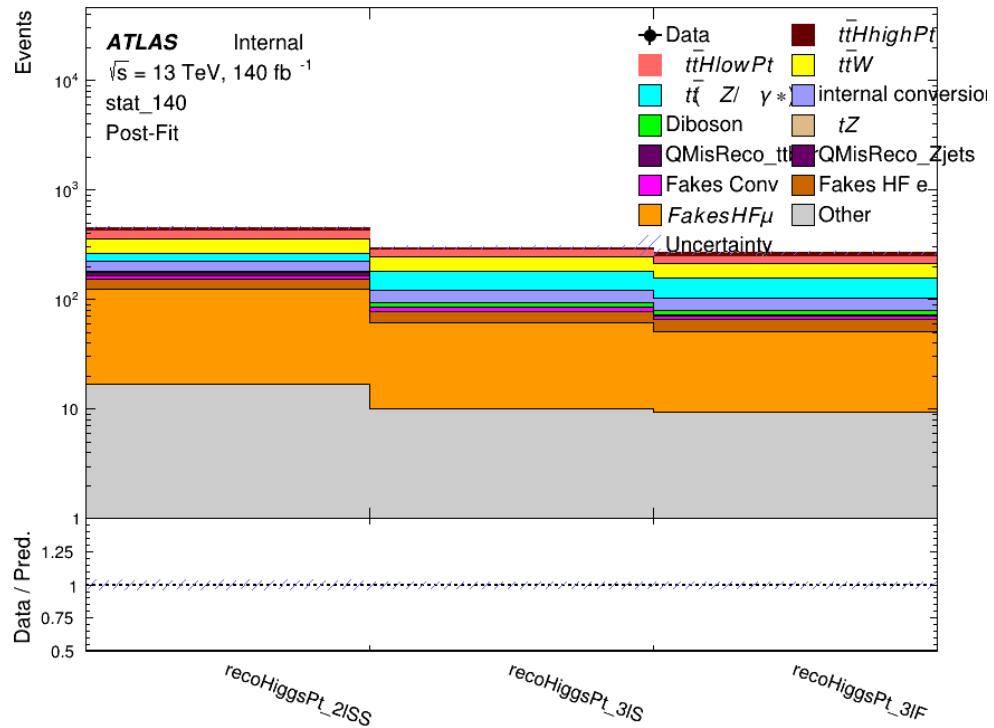


Figure 8.5: Post-fit summary of fit.

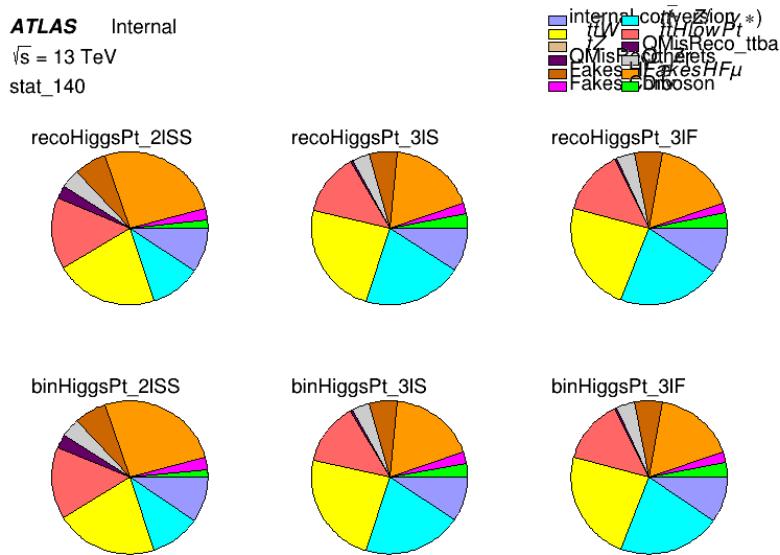


Figure 8.6: Background composition of the fit regions.

410 **List of contributions**

411

⁴¹² **Appendices**

⁴¹³ **A Machine Learning Models**

⁴¹⁴ The following section provides details of the various MVAs as well as a few studies performed in
⁴¹⁵ support of this analysis, exploring alternate decisions and strategies.

⁴¹⁶ **A.1 Higgs Reconstruction Models**

⁴¹⁷ **A.1.1 b-jet Identification Features - 2lSS**

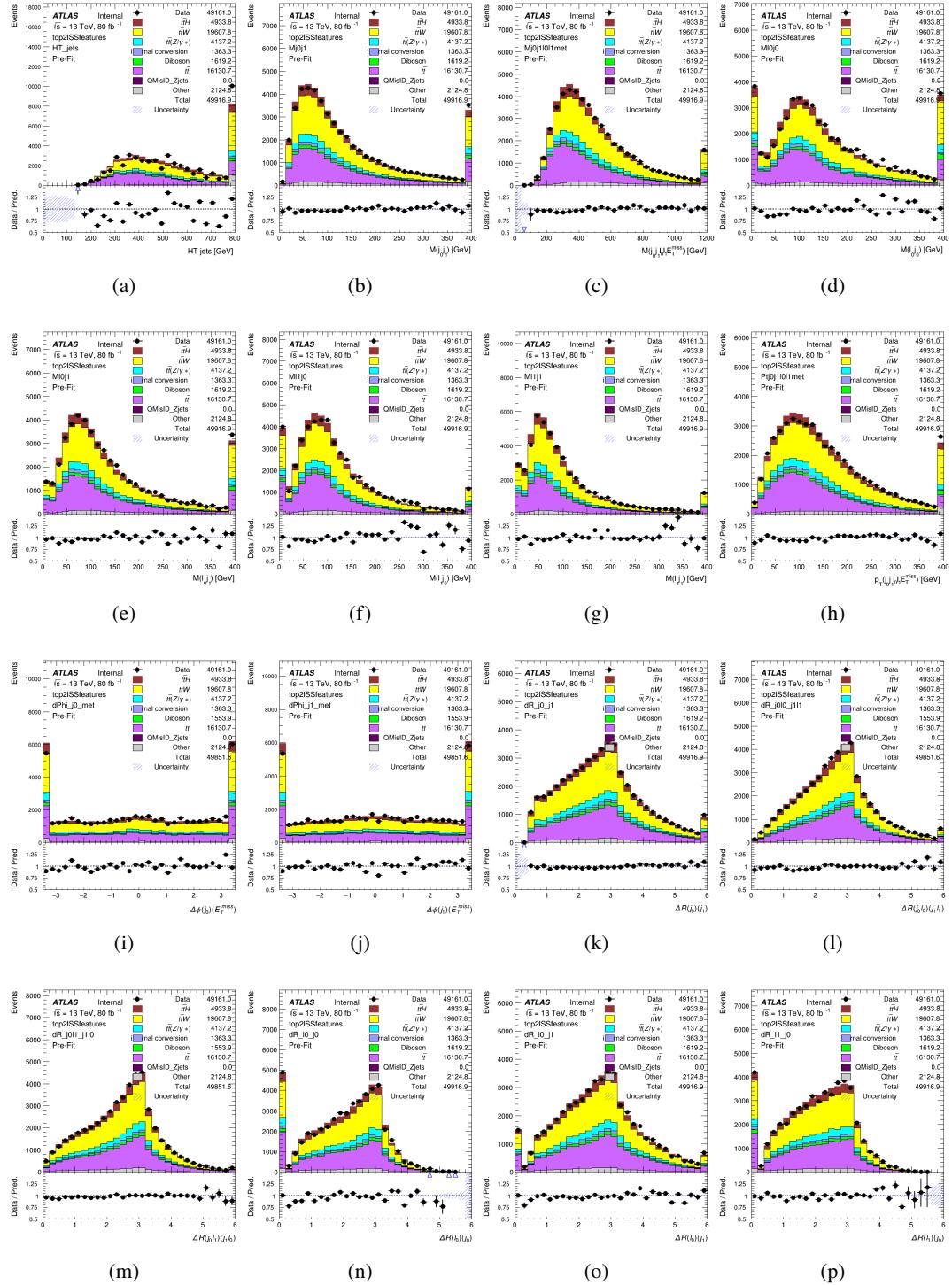


Figure A.1: Input features for top2lSS

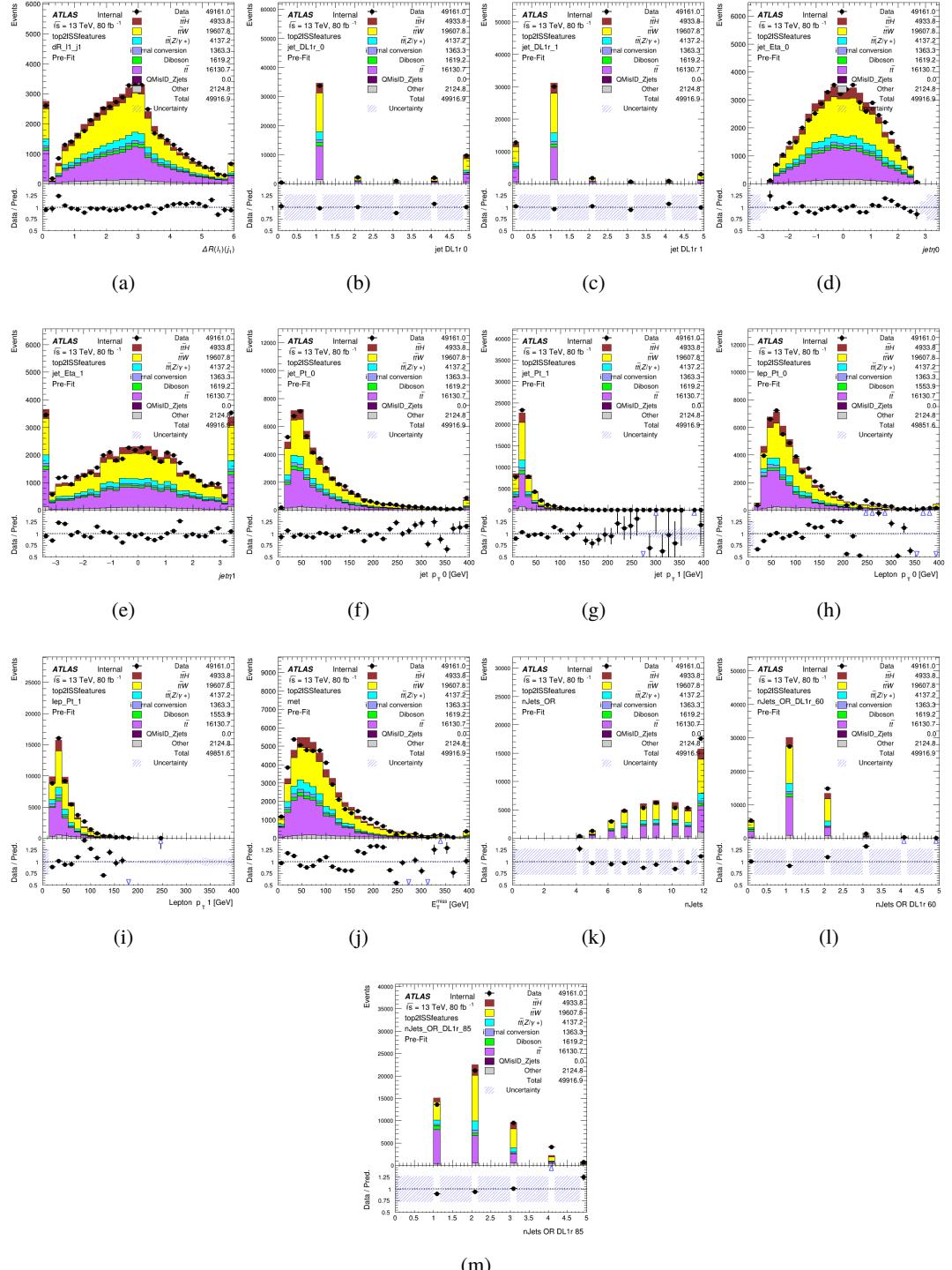


Figure A.2: Input features for top2lSS

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

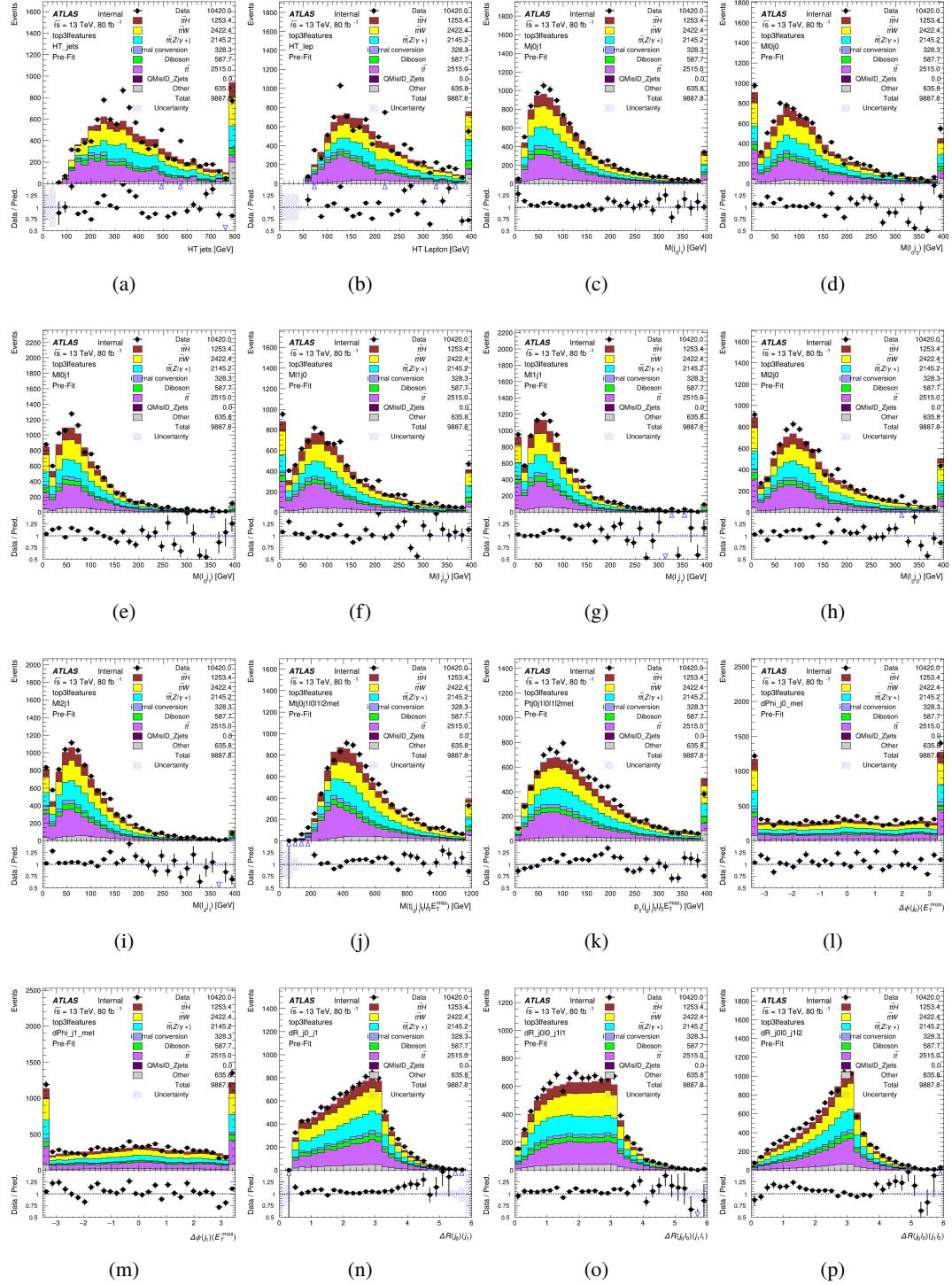


Figure A.3: Input features for top31

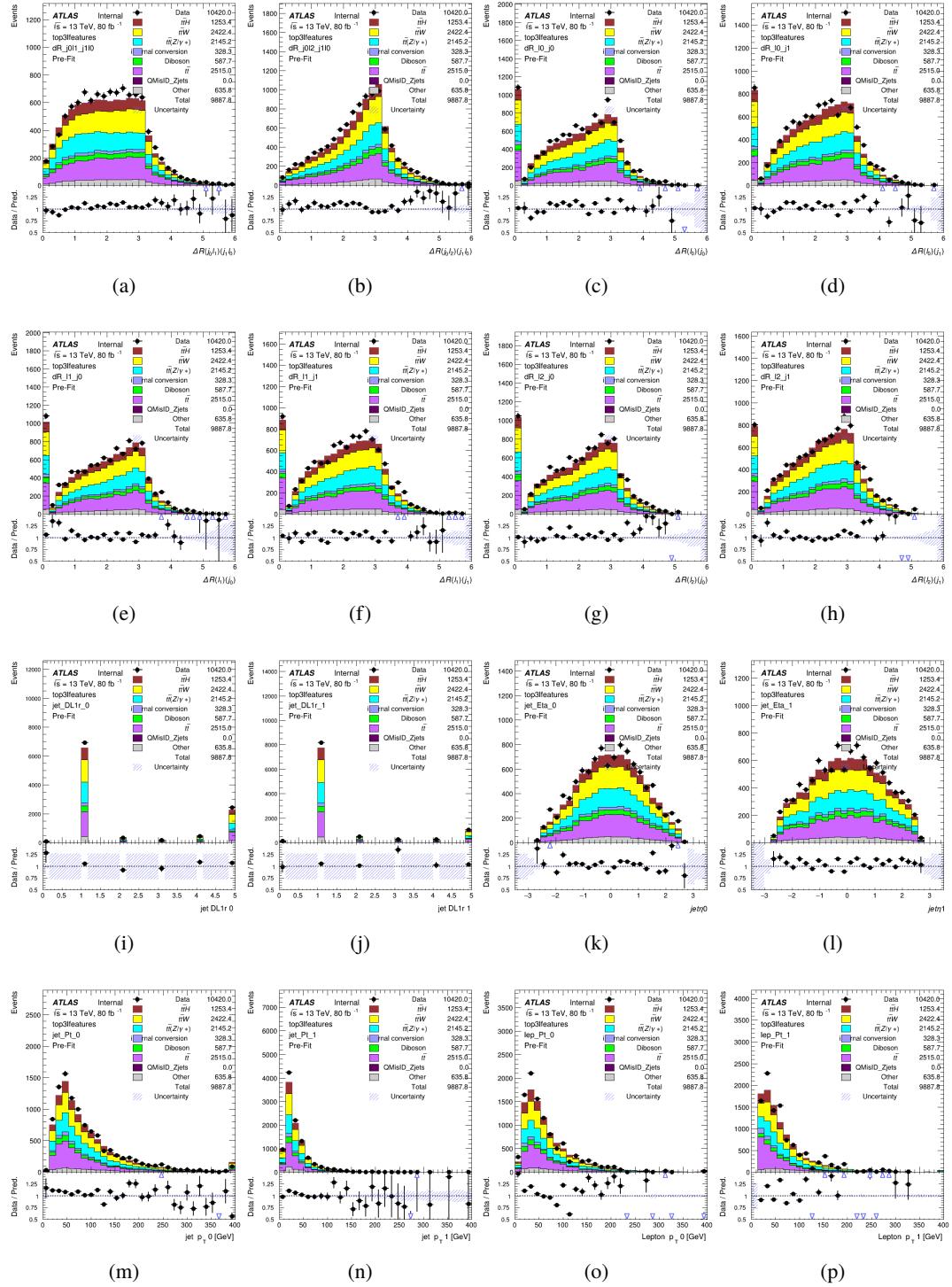


Figure A.4: Input features for top31

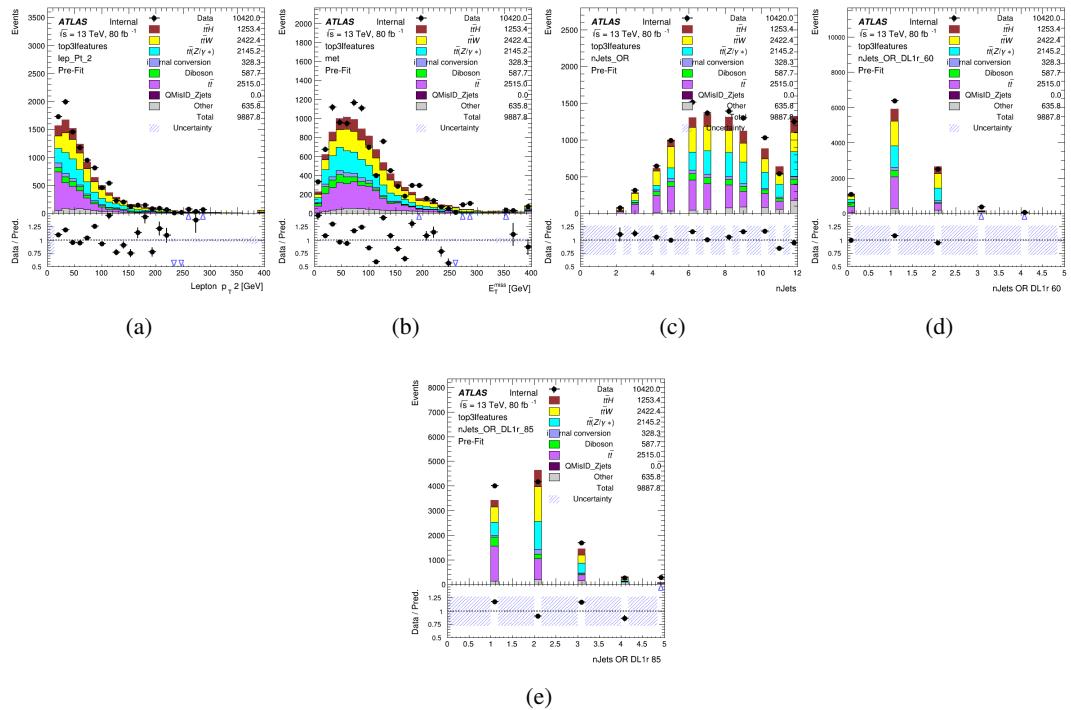


Figure A.5: Input features for top31

⁴¹⁹ **A.1.3 Higgs Reconstruction Features - 2lSS**

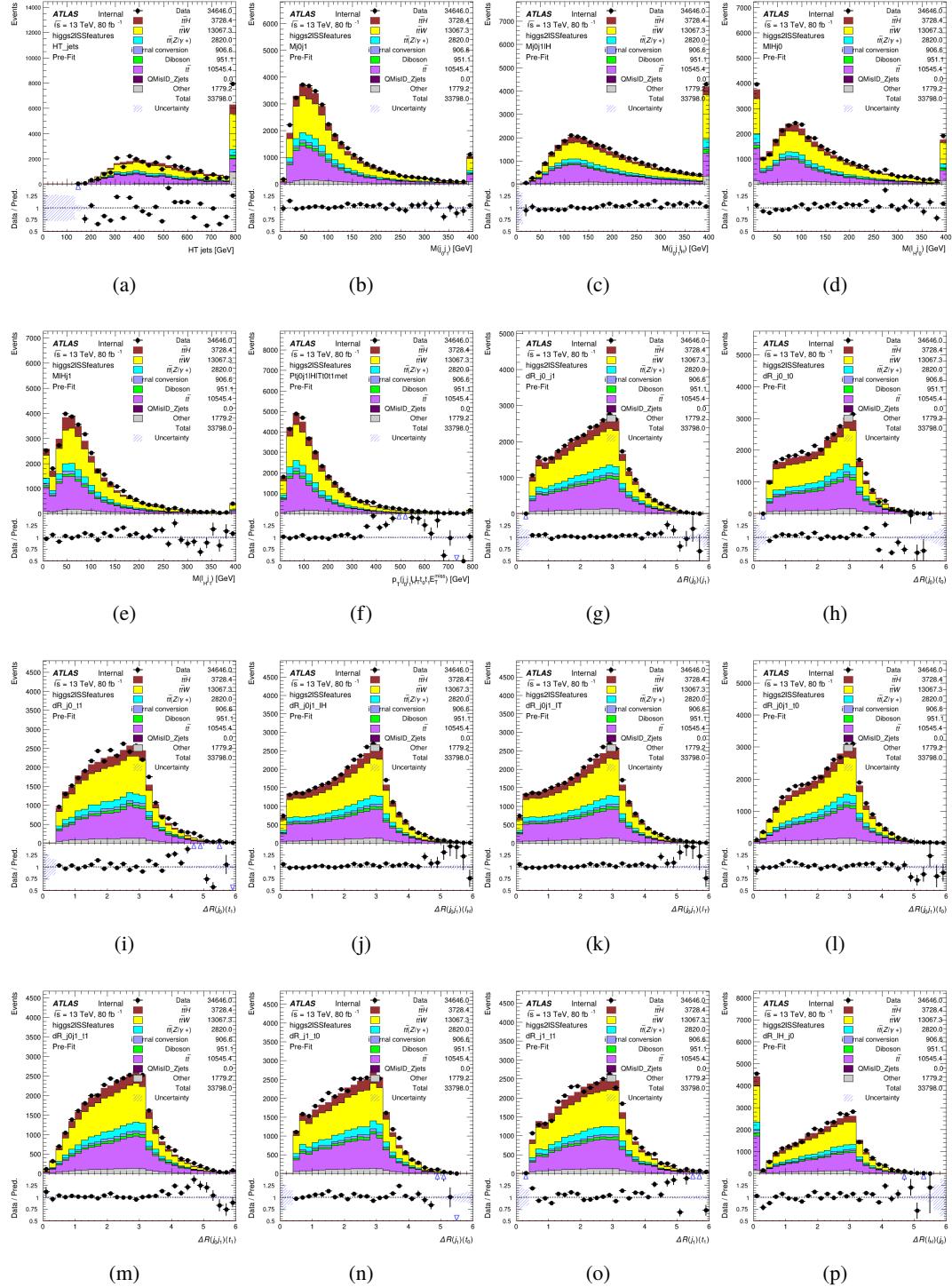


Figure A.6: Input features for higgs2lSS

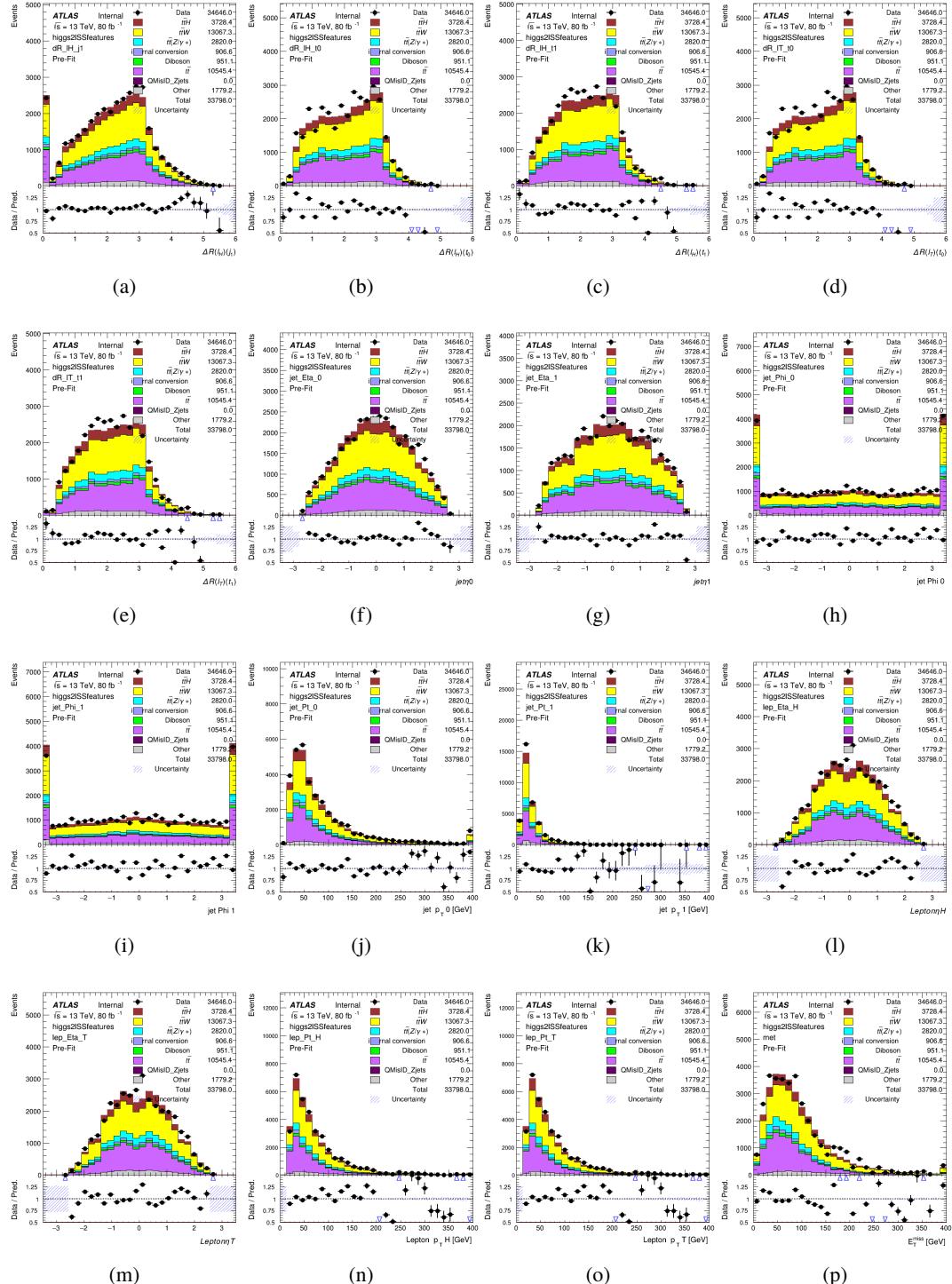


Figure A.7: Input features for higgs2lSS

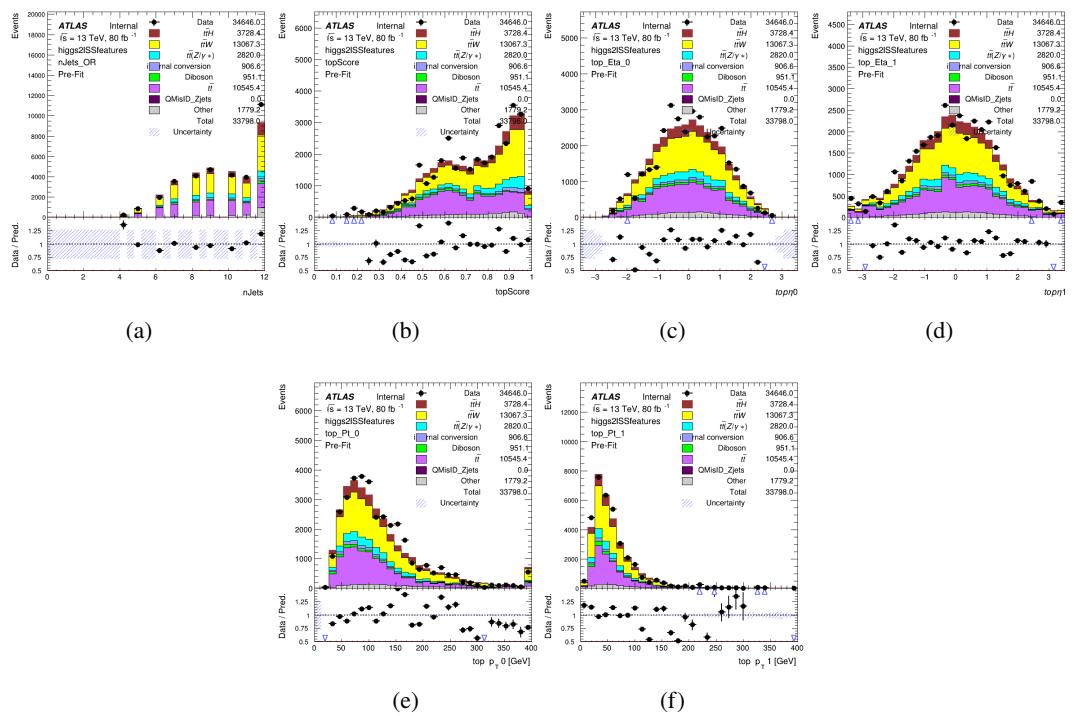


Figure A.8: Input features for higgs2lSS

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

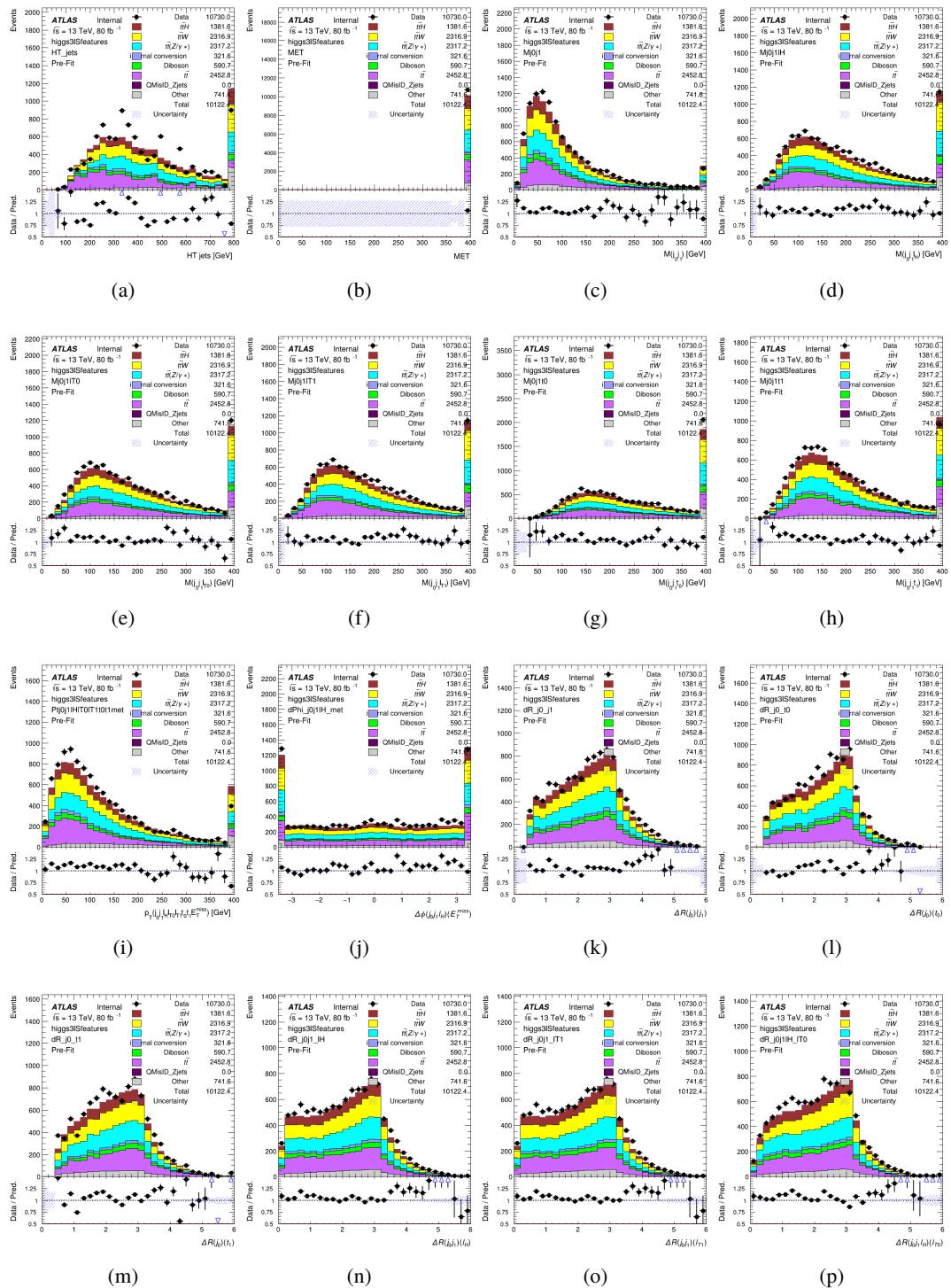


Figure A.9: Input features for higgs3lS

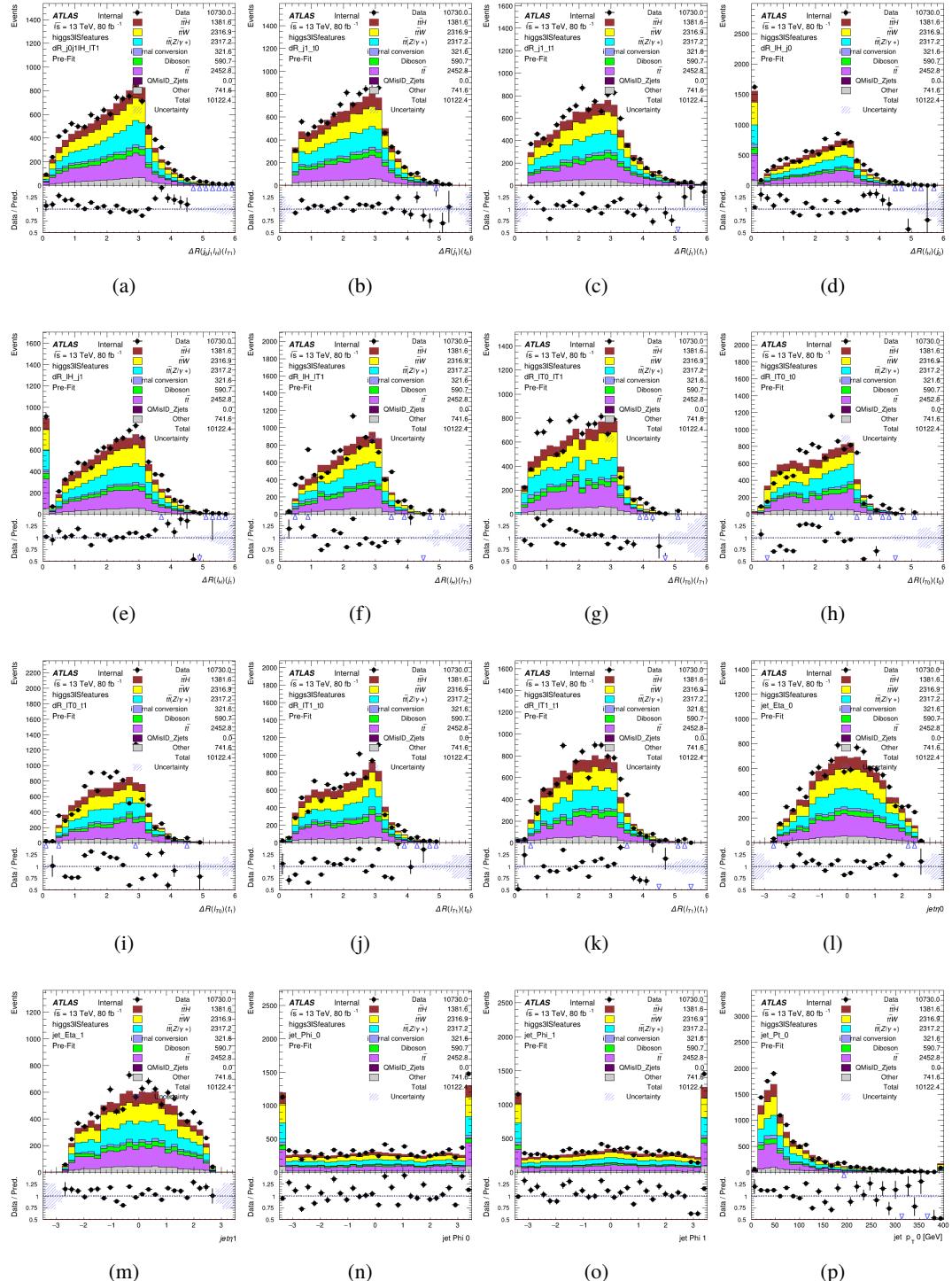


Figure A.10: Input features for higgs3l

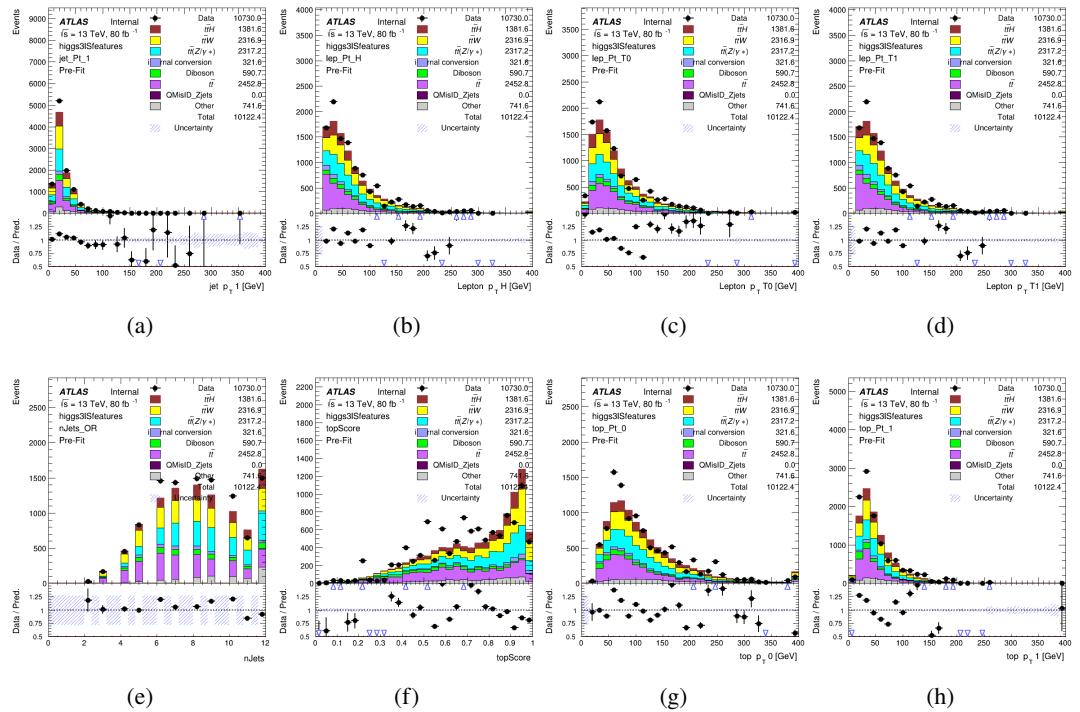


Figure A.11: Input features for higgs3lS

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

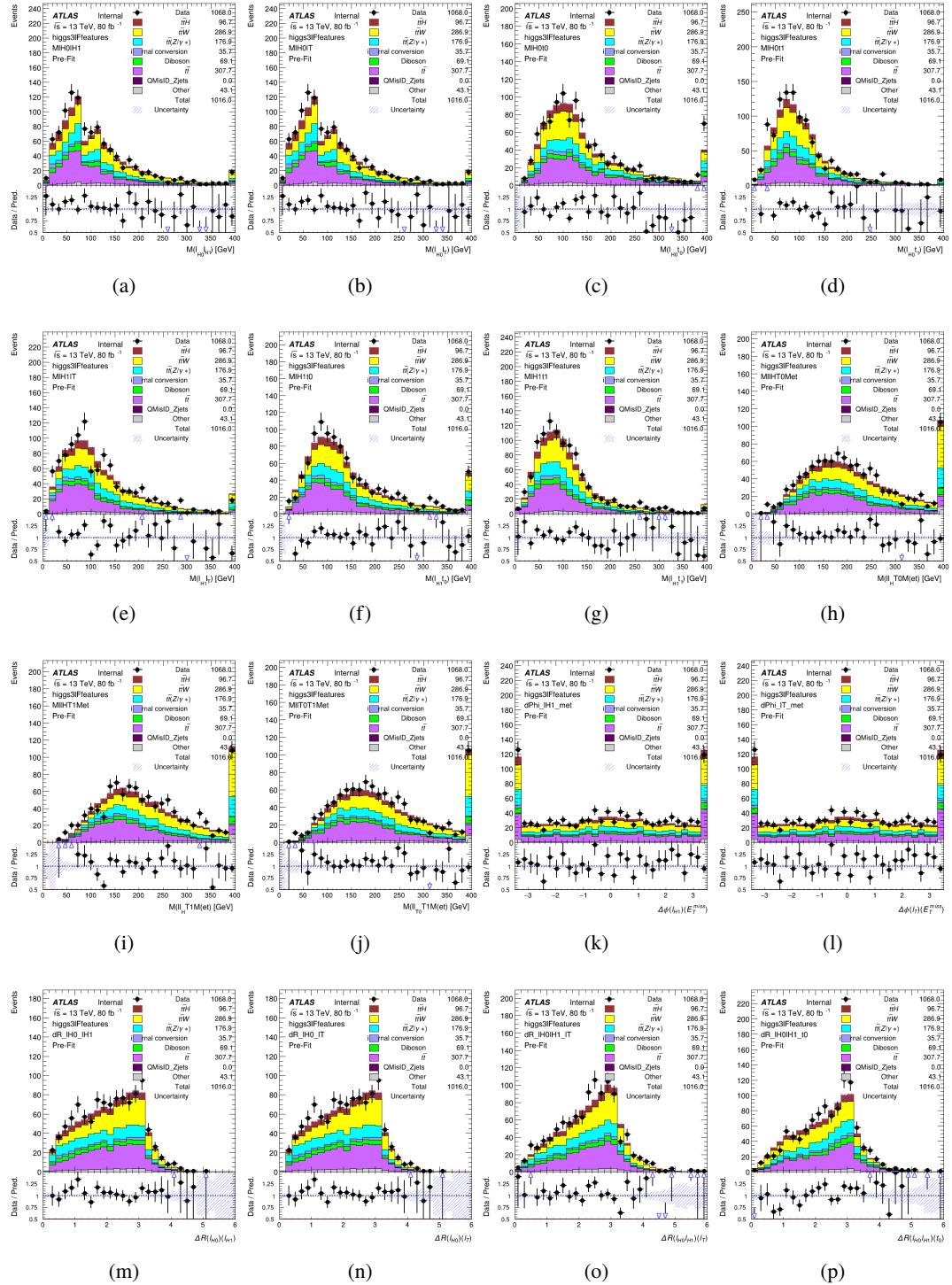


Figure A.12: Input features for higgs3lF

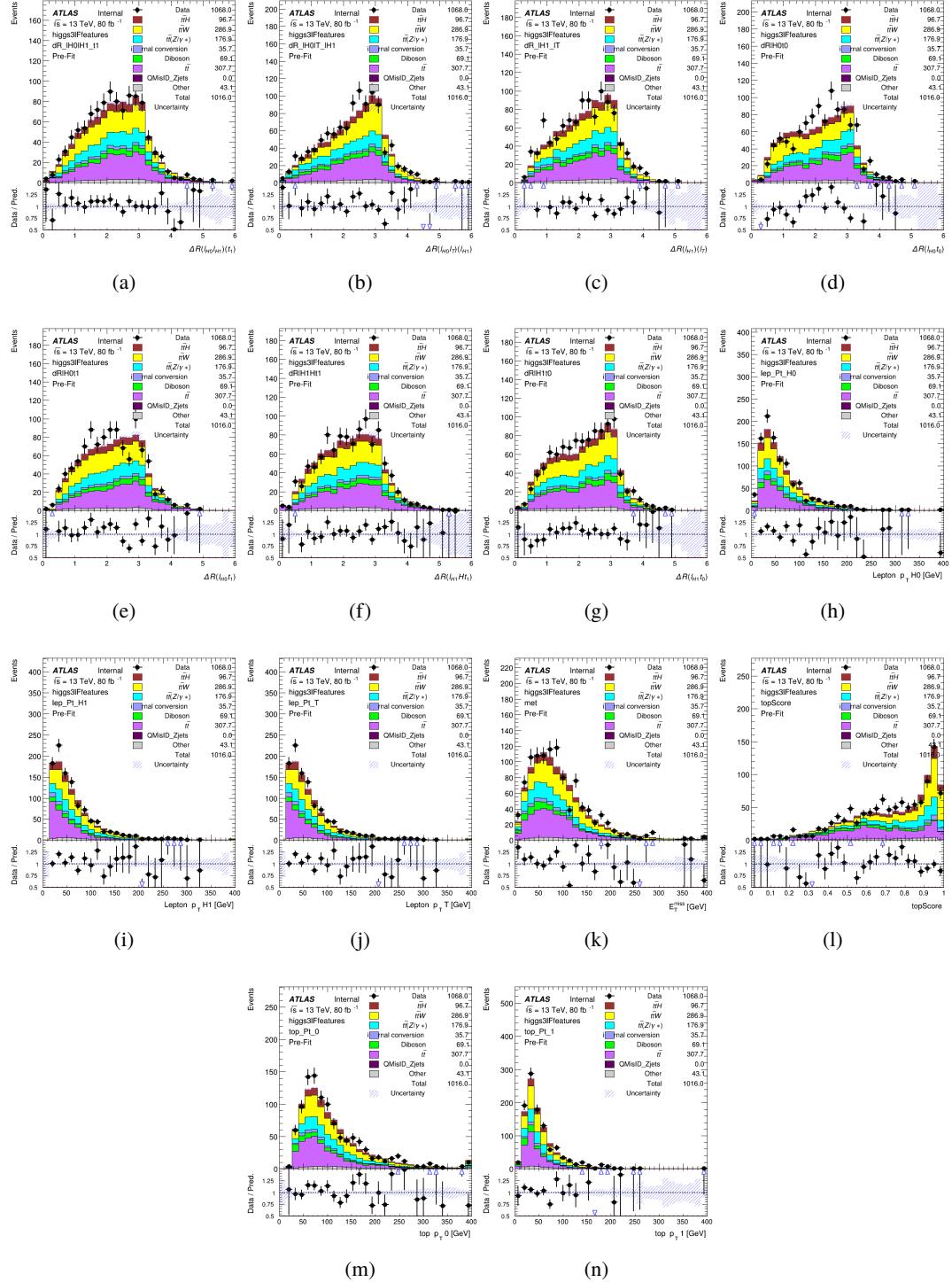


Figure A.13: Input features for higgs3lF

⁴²² **A.2 Background Rejection MVAs**

⁴²³ **A.2.1 Background Rejection MVA Features - 2lSS**

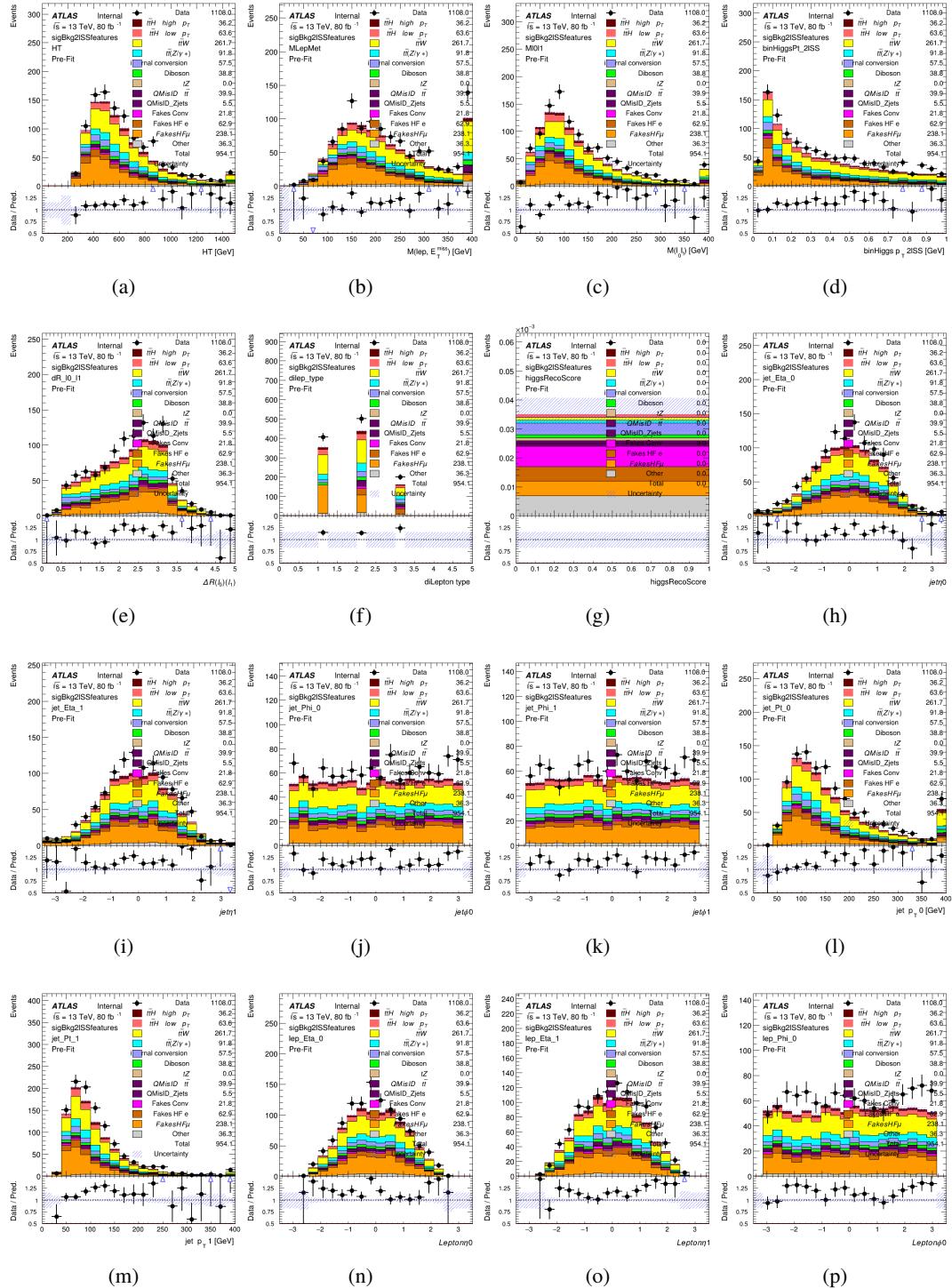


Figure A.14: Input features for sigBkg2lSS

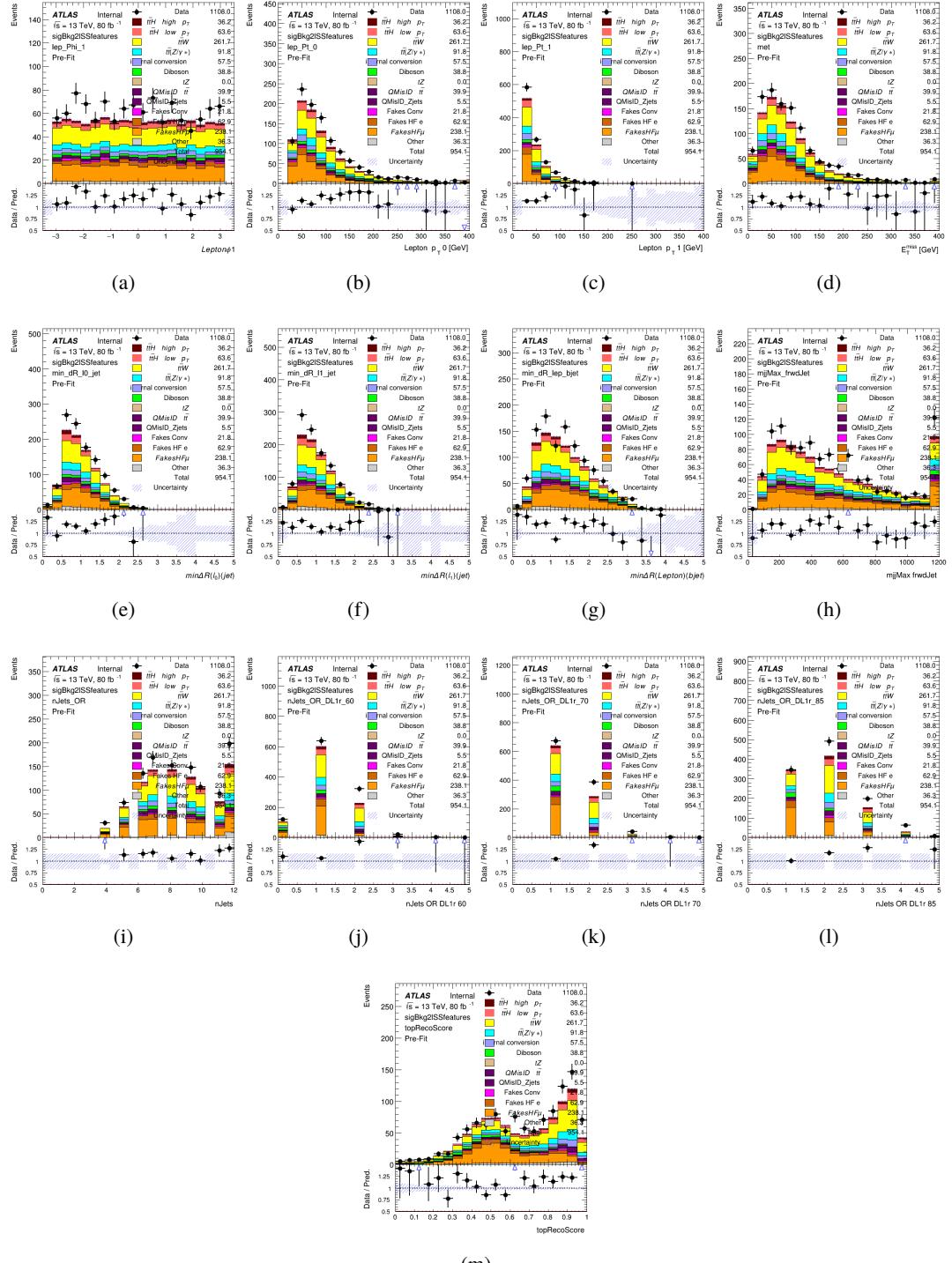
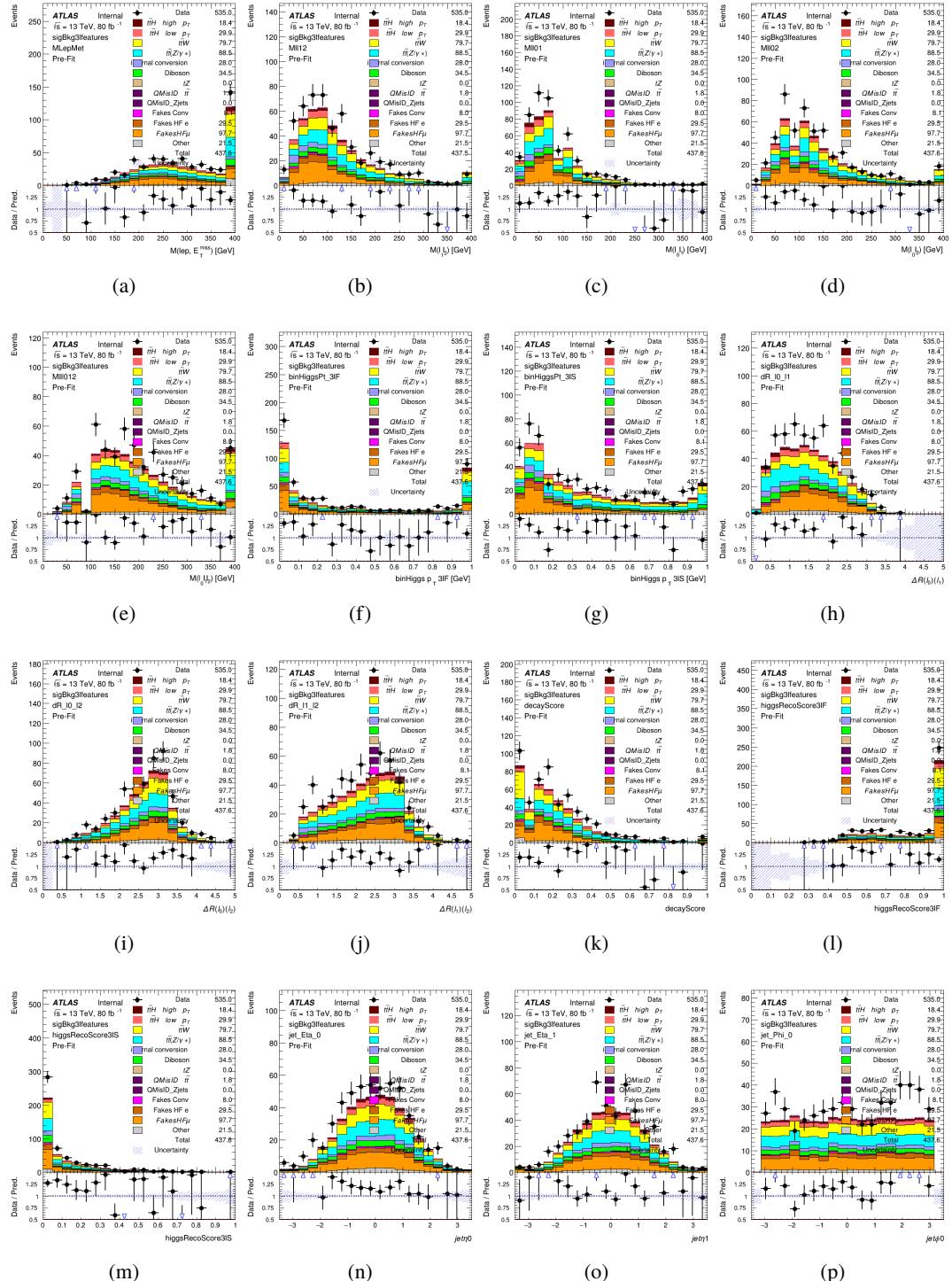
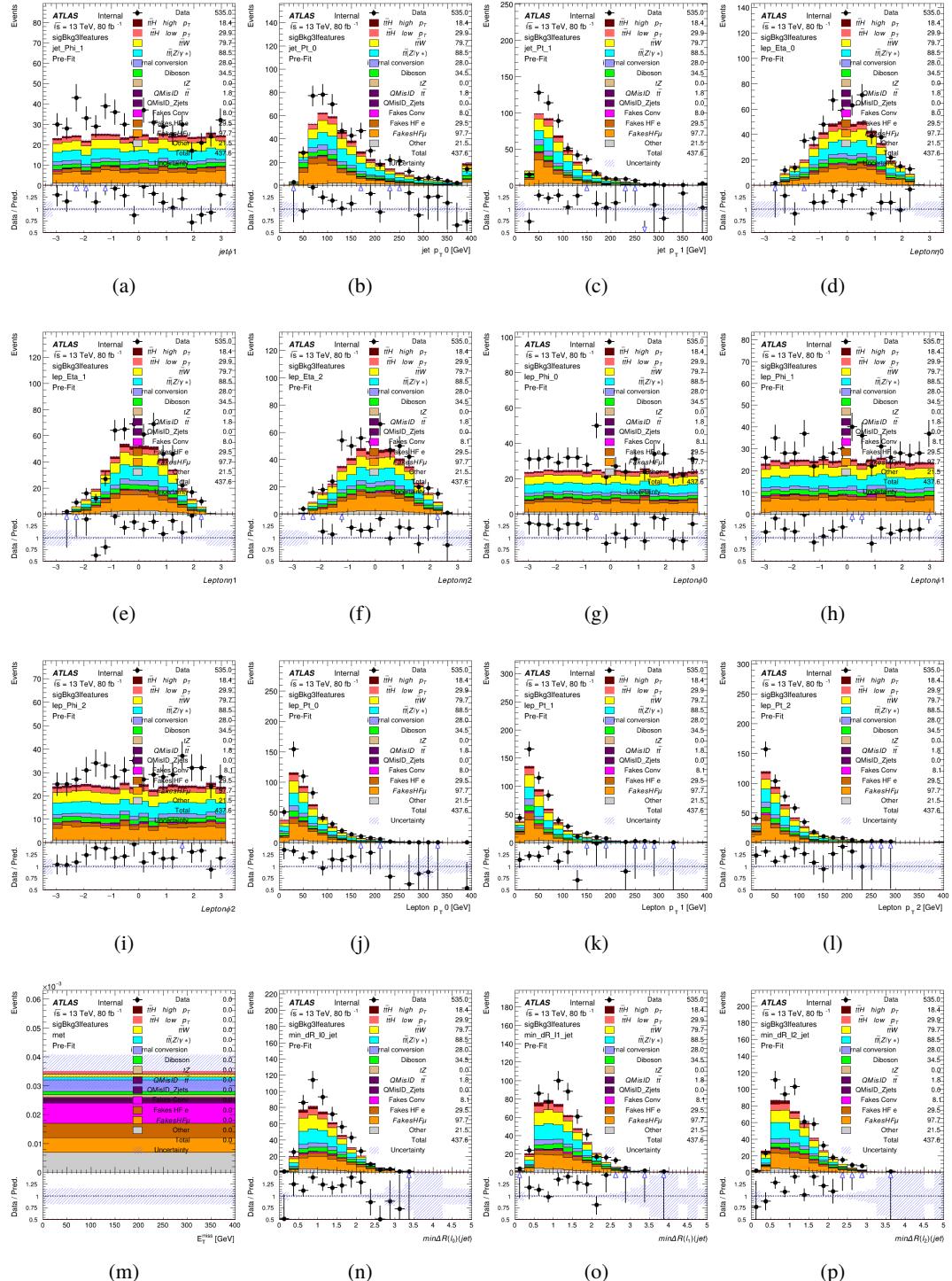


Figure A.15: Input features for sigBkg2lSS

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

Figure A.16: Input features for sigBkg3l

Figure A.17: Input features for `sigBkg3l`

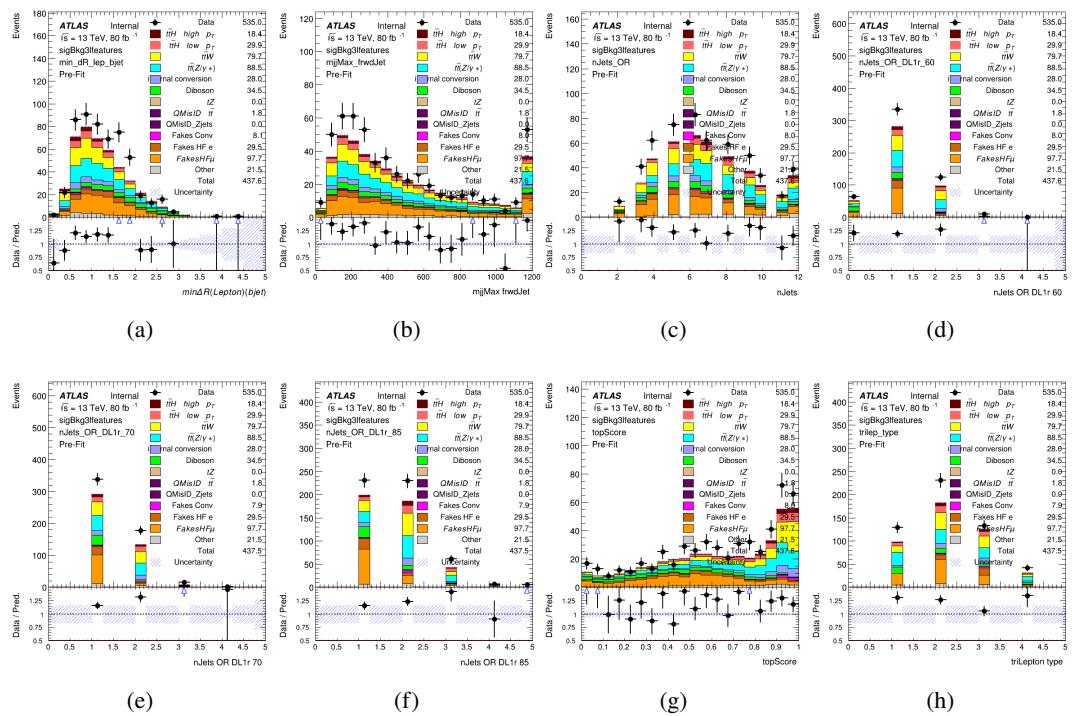


Figure A.18: Input features for sigBkg3l

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

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

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

Table 10: 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%

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

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

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

446 **A.5 Impact of Alternative Jet Selection**

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

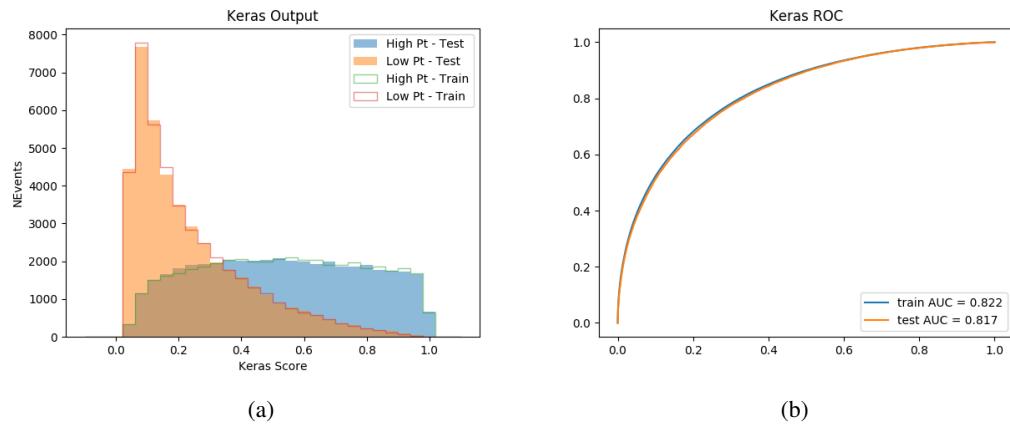


Figure A.19:

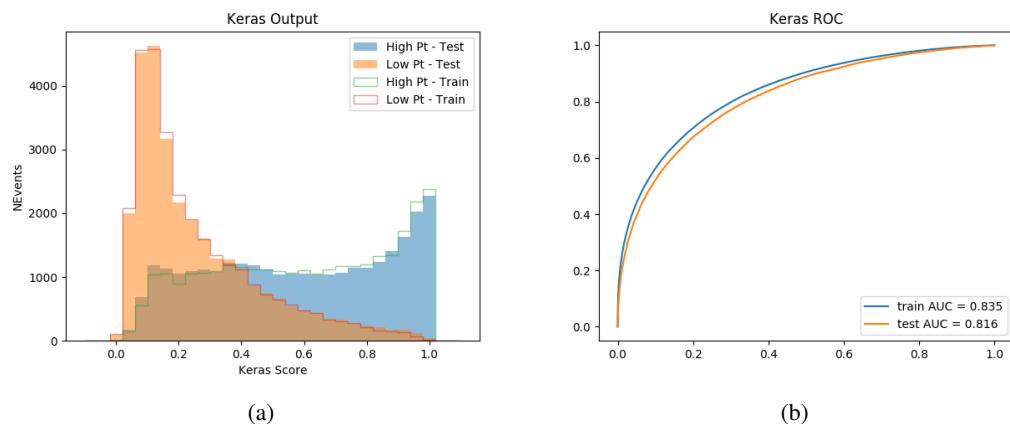


Figure A.20:

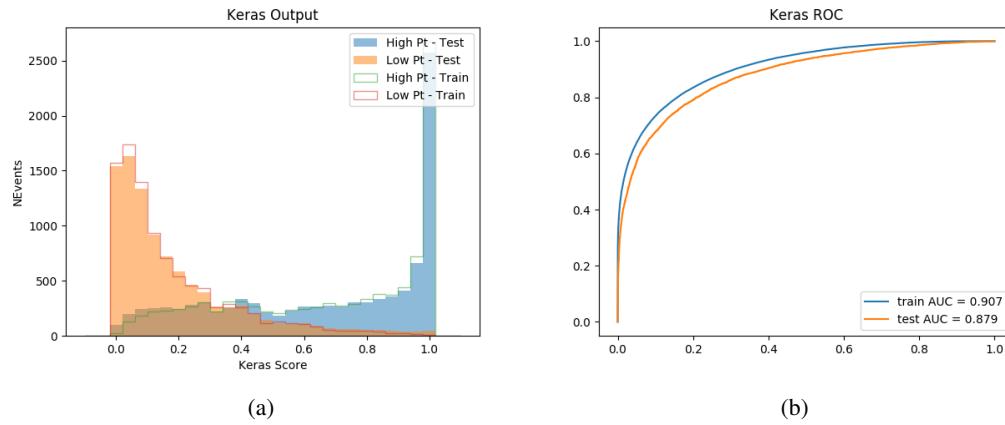


Figure A.21:

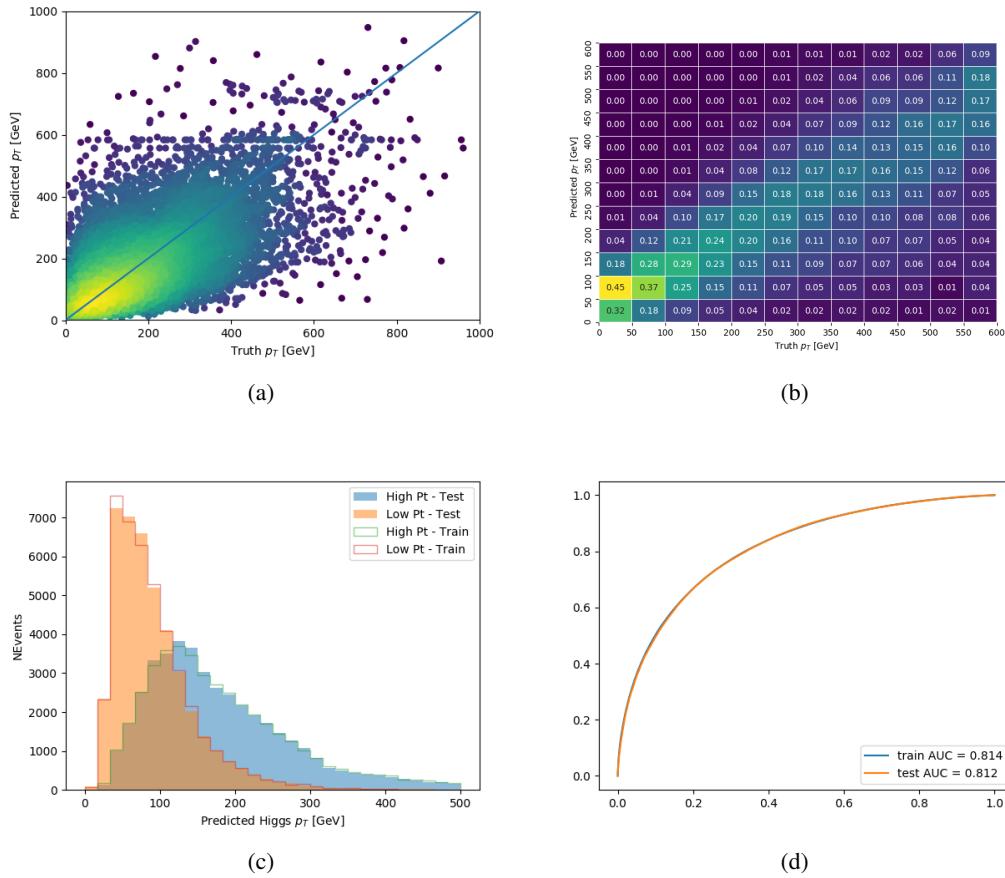


Figure A.22: