ATLAS Note

GROUP-2017-XX

11th November 2020

Draft version 0.1

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 $\sqrt{s} = 13$ TeV are used to provide estimates of the sensitivy to variations in the Higgs p_{T} spectrum.

Contents

1 Changes and outstanding items

1.1 Changelog

This is version 1

2 Introduction

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

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

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

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

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

3 Data and Monte Carlo Samples

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

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

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

3.1 Data Samples

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

- data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02 _PHYS_StandardGRL_All_Good_25ns.xml
- data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04 _PHYS_StandardGRL_All_Good_25ns.xml
- data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL _All_Good_25ns_Triggerno17e33prim.xml
- data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL _All_Good_25ns_Triggerno17e33prim.xml

3.2 Monte Carlo Samples

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

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

Process	Event generator	ME order	Parton Shower	PDF
$t\bar{t}H$	MG5_aMC	NLO	Pythia 8	NNPDF 3.0 NLO [**Ball:2014uwa**]
	(MG5_aMC)	(NLO)	(Herwig++)	(CT10 [**ct10**])
$t\bar{t}W$	MG5_aMC	NLO	Pythia 8	NNPDF 3.0 NLO
	(Sherpa 2.1.1)	(LO multileg)	(Sherpa)	(NNPDF 3.0 NLO)
$t\bar{t}(Z/\gamma^* \to ll)$	MG5_aMC	NLO	Pythia 8	NNPDF 3.0 NLO
VV	Sherpa 2.2.2	MEPS NLO	Sherpa	CT10
$t\bar{t}$	Powheg-BOX v2 [**powhegtt**]	NLO	Pythia 8	NNPDF 3.0 NLO
$t\bar{t}\gamma$	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	NLO	Herwig++	CT10
	(Sherpa 2.1.1)	(LO multileg)	(Sherpa)	(NNPDF 3.0 NLO)
tWZ	MG5_aMC	NLO	Pythia 8	NNPDF 2.3 LO
$t\bar{t}t$, $t\bar{t}t\bar{t}$	MG5_aMC	LO	Pythia 8	NNPDF 2.3 LO
$t\bar{t}W^+W^-$	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				
$Z \to l^+l^-$	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
$e\mu$, μe ($\sim$symm.)	HLT_e17_lhloose_mu14
	Dilepton triggers (2016)
$\mu\mu$ (asymm.)	HLT_mu22_mu8noL1
ee (symm.)	HLT_2e17_lhvloose_nod0
$e\mu$, μe ($\sim$symm.)	HLT_e17_lhloose_nod0_mu14
	Dilepton triggers (2017)
$\mu\mu$ (asymm.)	HLT_mu22_mu8noL1
ee (symm.)	HLT_2e24_lhvloose_nod0
$e\mu$, μe ($\sim$symm.)	HLT_e17_lhloose_nod0_mu14
	Dilepton triggers (2018)
$\mu\mu$ (asymm.)	HLT_mu22_mu8noL1
ee (symm.)	HLT_2e24_lhvloose_nod0
$e\mu$, μe ($\sim$symm.)	HLT_e17_lhloose_nod0_mu14

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

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

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

4.3 Jets

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

4.4 Missing Transverse Energy

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

5 Higgs Momentum Reconstruction

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

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

Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_aMC. Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training.

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

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

5.1 Decay Candidate Reconstruction

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

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

5.2 b-jet Identification

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

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

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

5.2.1 2lSS Channel

For the 2lSS channel, the input features shown in table 3 are used for training. Here j_0 and j_1 are the two jet candidates, while l_0 and l_1 are the two leptons in the event, both ordered by p_T. jet DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet, where 5 represents the tightest working point and 1 represents the loosest. The variables nJets DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and 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	$\Delta R(j_0 l_0)(j_1 l_1)$
$\Delta R(j_0 l_1)(j_1 l_0)$	$p_T(j_0 j_1 l_0 l_1 E_T^{miss})$	$M(j_0 j_1 l_0 l_1 E_T^{miss})$
$\Delta\phi(j_0)(E_T^{miss})$	$\Delta\phi(j_1)(E_T^{miss})$	HT jets
nJets	E_T^{miss}	

Table 3: Input features used in the b-jet identification algorithm for the 2lSS channel

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

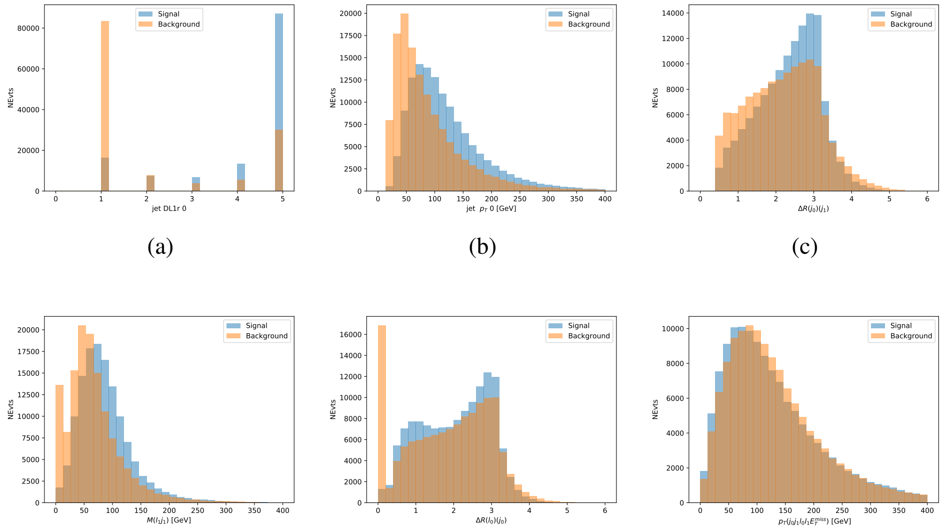
The difference between the distributions for a few of these features for the c̈orrect(̈i.e. both jets are truth b-jets), and ïncorrectc̈ombinations are shown in figure 5.1. The correct and incorrect contributions are scaled to the same integral, so as to better demonstrate the differences in the distributions.

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

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.

For point of comparison, a n̈aiveäpproach 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 accuracey of the model for different values of n-bjets, compared to this naive approach, is shown in table 4.



(a) (b) (c)

(d) (e) (f)

Figure 5.1: Input features for top2lSS 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}.

Table 4: Accuracey of the NN in identifying b-jets from tops in 2lSS events for, compared to the accuracey 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%

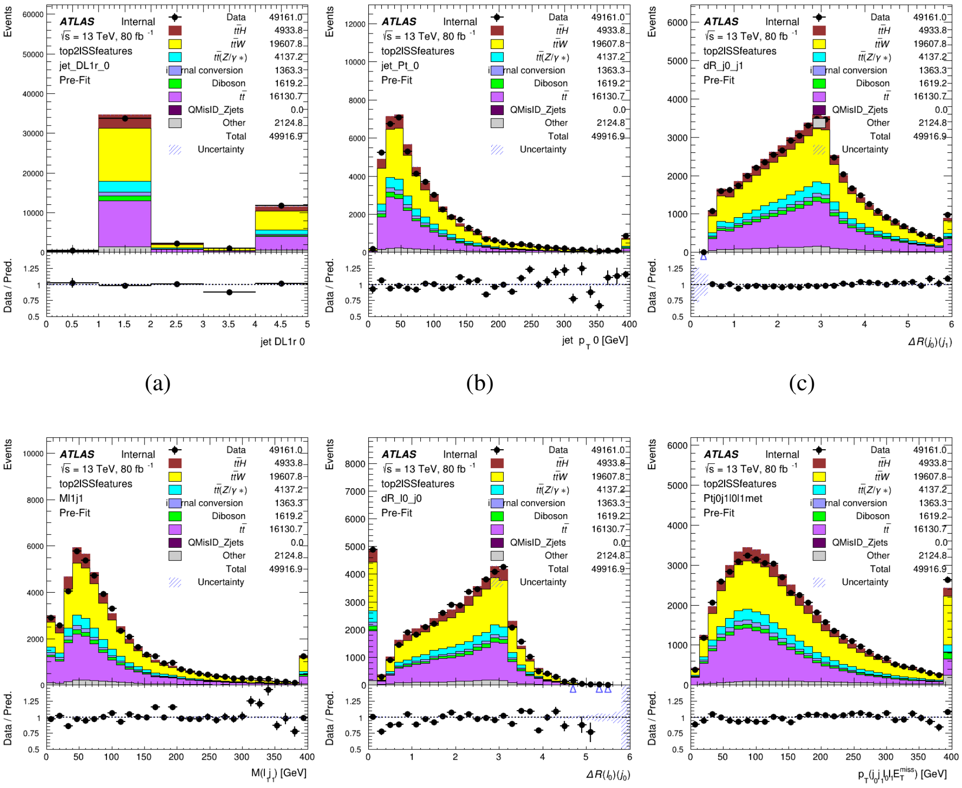
5.2.2 3l Channel

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

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

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

Again, the dataset is downsized to reduce the ratio of correct and incorrect combination from 20:1,



(a) (b) (c)

(d) (e) (f)

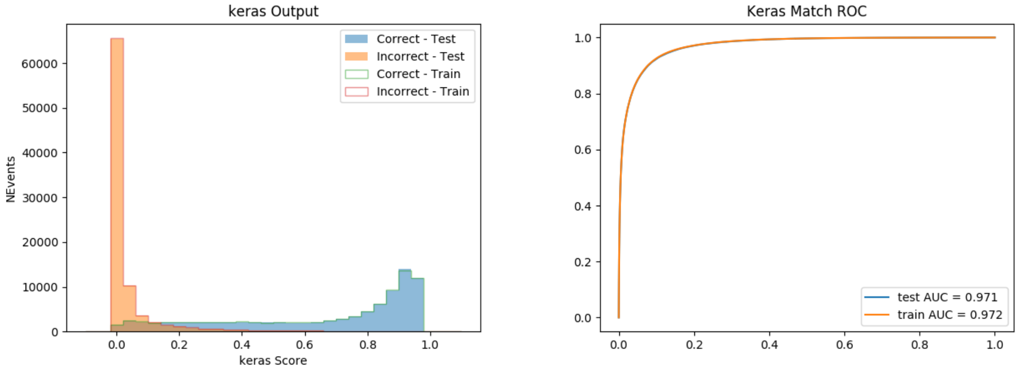
Figure 5.2: Data/MC comparisons of input features for top2lSS training for 80 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f)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

to 5:1. Around 7 million events are used for training, with 10% set aside for testing. Based on the results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with 60 nodes each. The output score distribution as well as the ROC curve for the trained model are shown in figure 5.2.2.

This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal events. The accuracy of the model is summarized in table 6.

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.



(a) (b)

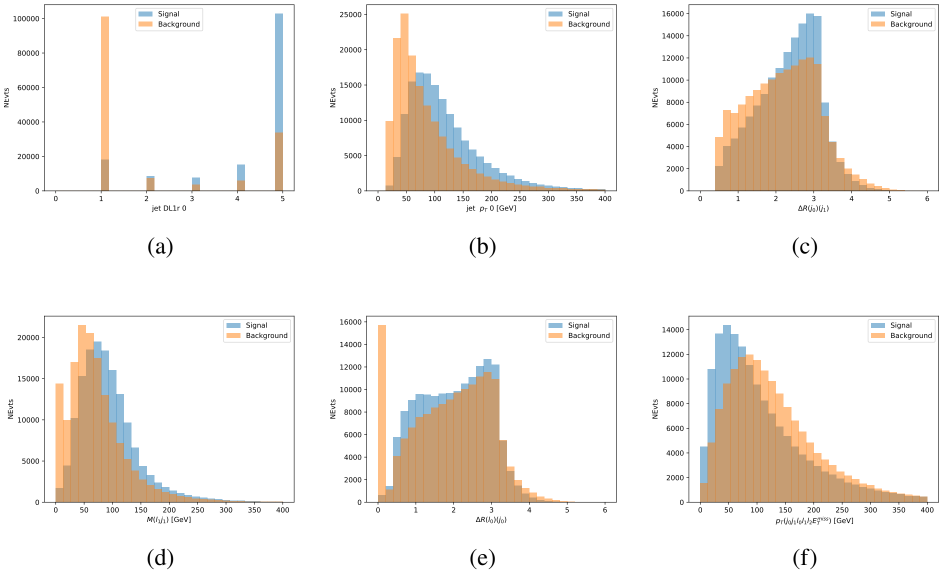
Figure 5.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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
$p_T(j_0 j_1 l_0 l_1 l_2 E_T^{miss})$	$M(t j_0 j_1 l_0 l_1 l_2 E_T^{miss})$	$\Delta\phi(j_0)(E_T^{miss})$
$\Delta\phi(j_1)(E_T^{miss})$	HT Lepton	HT jets
nJets	E_T^{miss}	nJets OR DL1r 85
nJets OR DL1r 60		

Table 5: Input features for the b-jet identification algorithm in the 3l channel.

Table 6: Accuracey 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%



(a) (b) (c)

(d) (e) (f)

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.

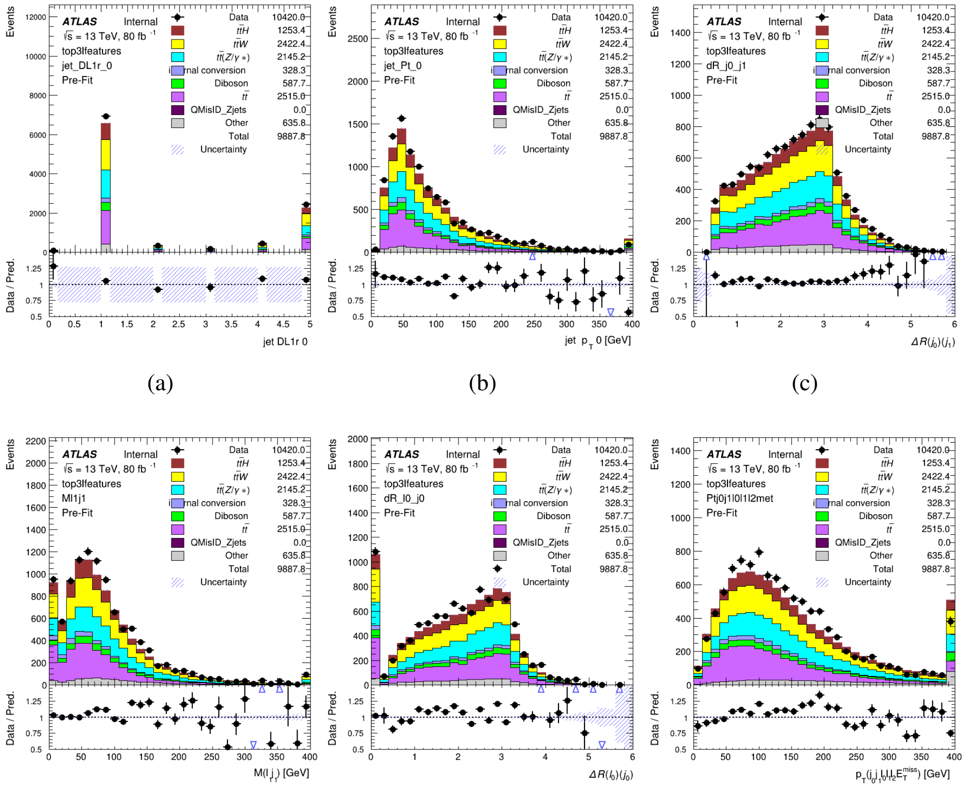
Again separate models are used for the 2lSS and 3l channels, while the 3l channel has now been split into two: $t\bar{t}H$ events with three leptons in the final state include both intances where the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the Higgs decays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

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

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

5.3.1 2lSS Channel

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

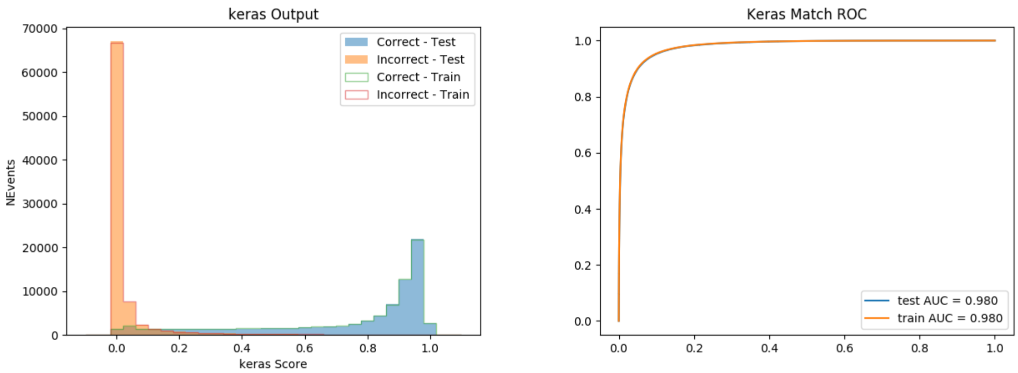


(a) (b) (c)

(d) (e) (f)

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

266 b-jets identified in section 5.2, with the specific input features listed in table 7.



(a) (b)

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

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

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

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

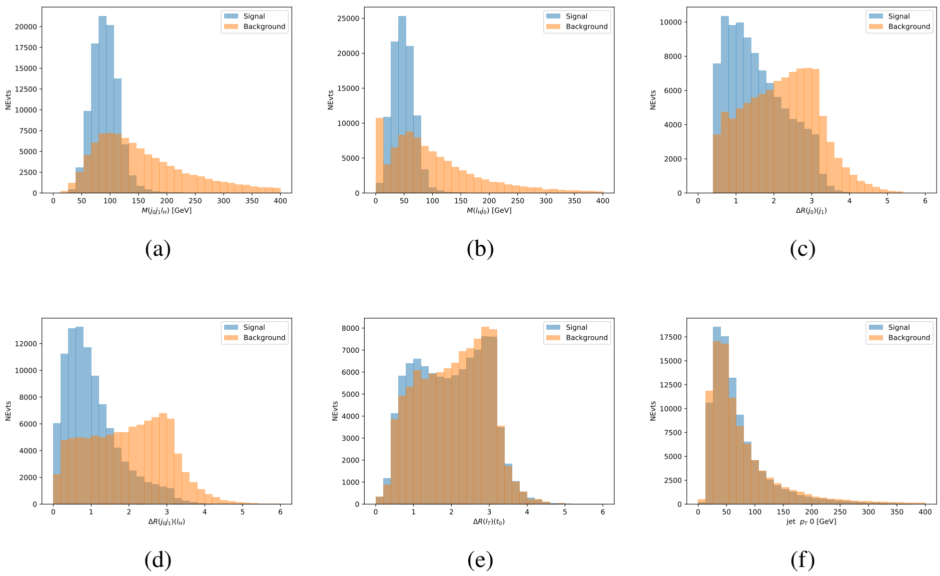


Figure 5.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

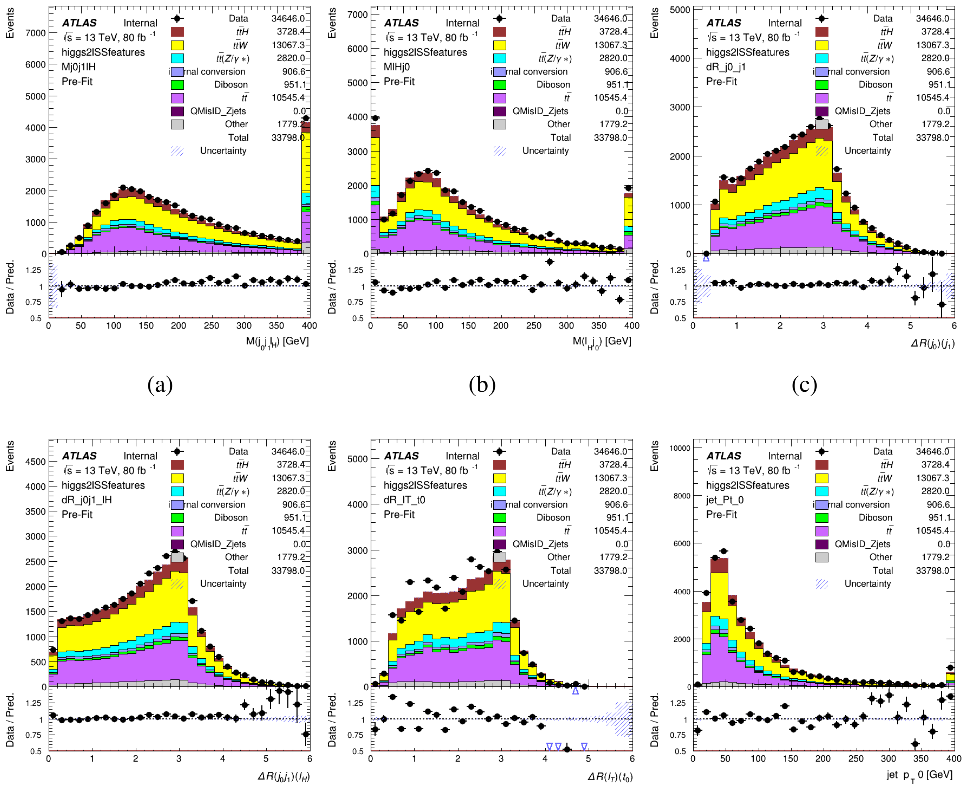
The modeling of these inputs is validated against data, with figure 5.2 showing good general agreement between data and MC. Plots for the complete list of features can found in section A.

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

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

5.3.2 3l Semi-leptonic Channel

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

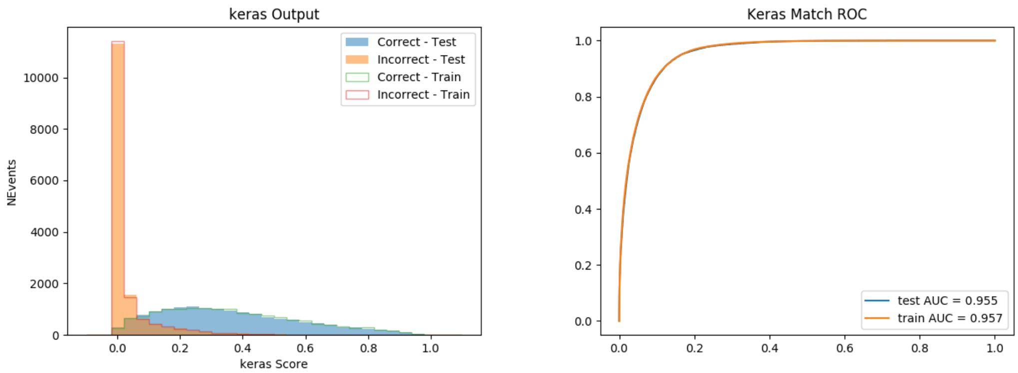


(a) (b) (c)

(d) (e) (f)

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

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



(a) (b)

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

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	$\Delta R(j_0)(j_1)$
$M(j_0 j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1)(l_{T_1})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0 j_1 l_{T_0})$	$M(j_0 j_1 l_{T_1})$
$M(j_0 j_1 l_H)$	$\Delta R(j_0 j_1 l_H)(l_{T_0})$	$\Delta R(j_0 j_1 l_H)(l_{T_1})$
$\Delta\phi(j_0 j_1 l_H)(E_T^{miss})$	$p_T(j_0 j_1 l_H l_{T_0} l_{T_1} b_0 b_1 E_T^{miss})$	$M(j_0 j_1 b_0)$
$M(j_0 j_1 b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
topScore	MET	HT jets
nJets		

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

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

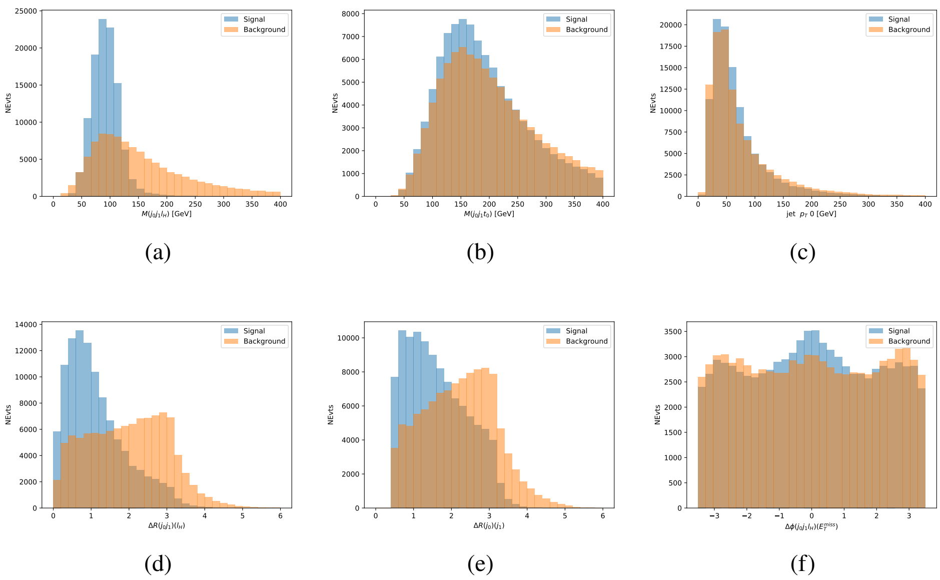


Figure 5.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

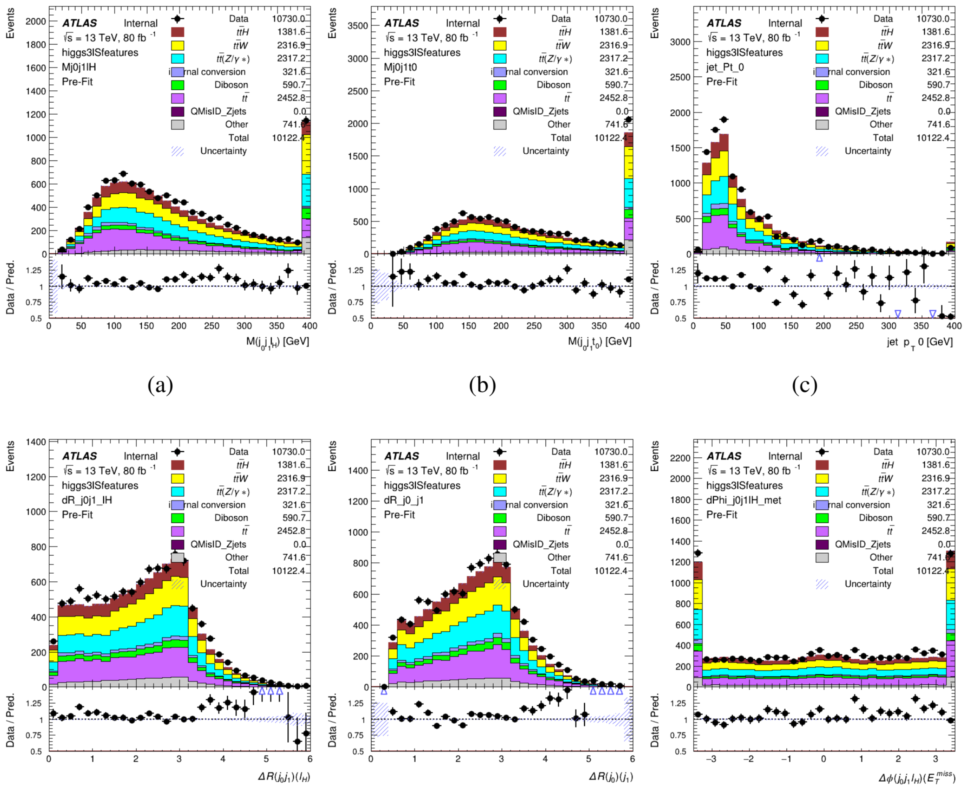
The modeling of these inputs is validated against data, with figure 5.11 showing good general agreement between data and MC. Plots for the complete list of features can found in appendix A.1.

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

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

5.3.3 3l Fully-leptonic Channel

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

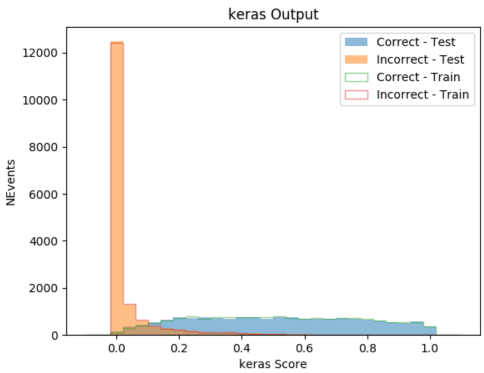


(a) (b) (c)

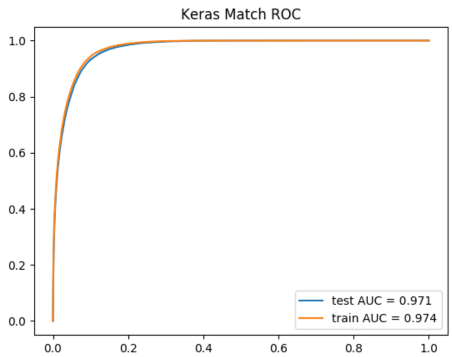
(d) (e) (f)

Figure 5.11: Data/MC comparisons of input features for higgs3lS training for 80 fb^{-1} of data.

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



(a)



(b)

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

Lepton p_T H_1	Lepton p_T H_0	Lepton p_T T
top p_T 0	top p_T 1	$\Delta\phi(l_{H_1})(E_T^{miss})$
$\Delta\phi(l_T)(E_T^{miss})$	$M(l_{H_0} l_{H_1})$	$M(l_{H_1} l_T)$
$M(l_{H_0} l_T)$	$\Delta R(l_{H_0})(l_{H_1})$	$\Delta R(l_{H_1})(l_T)$
$\Delta R(l_{H_0})(l_T)$	$\Delta R(l_{H_0} l_{H_1})(l_T)$	$\Delta R(l_{H_0} l_T)(l_{H_1})$
$\Delta R(l_{H_0} l_{H_1})(b_0)$	$\Delta R(l_{H_0} l_{H_1})(b_1)$	$\Delta R(l_{H_0} b_0)$
$M(l_{H_0} b_0)$	$\Delta R(l_{H_0} b_1)$	$M(l_{H_0} b_1)$
$\Delta R(l_{H_1} b_0)$	$M(l_{H_1} b_0)$	$\Delta R(l_{H_1} b_1)$
$M(l_{H_1} b_1)$	E_T^{miss}	topScore

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

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

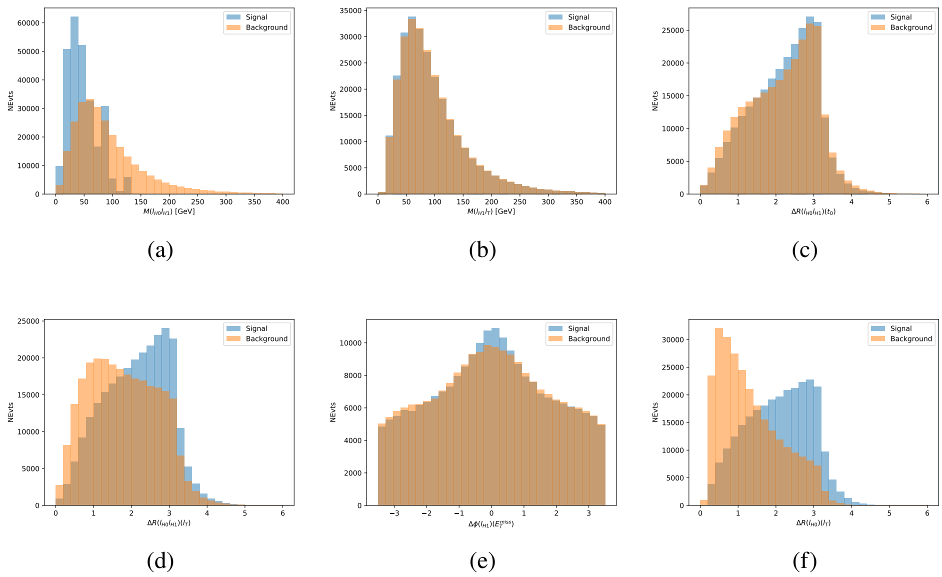


Figure 5.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

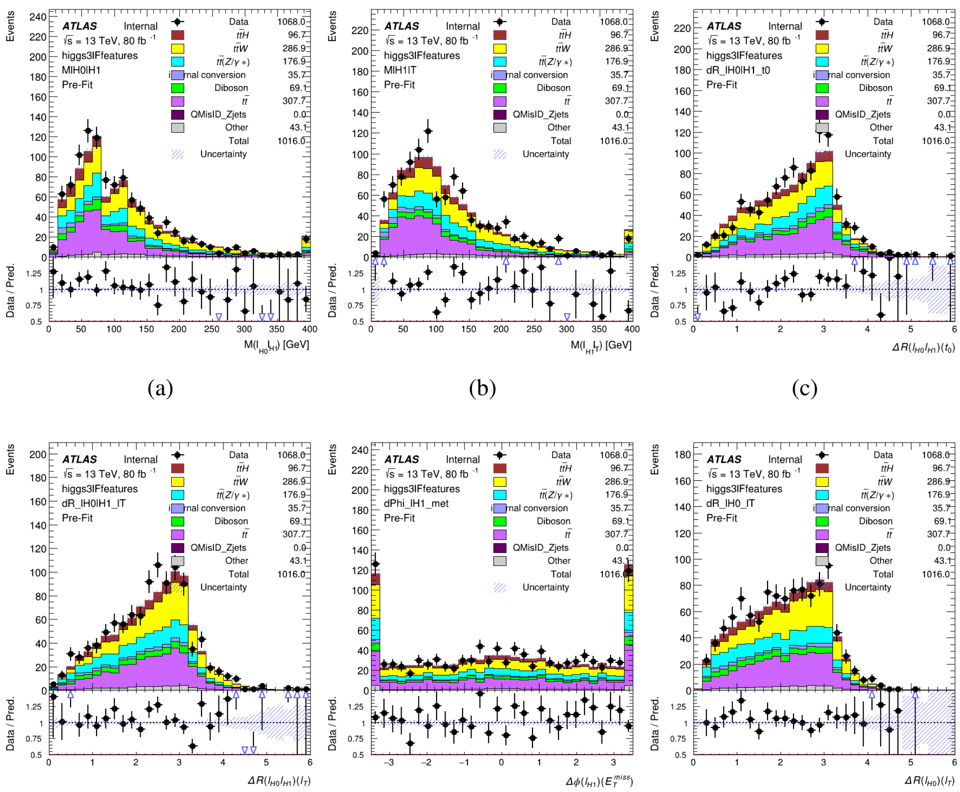
The modeling of these inputs is validated against data, with figure 5.14 showing good general agreement between data and MC. Plots for the complete list of features can found in section A.

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

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

5.4 p_T Prediction

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



(a) (b) (c)

(d) (e) (f)

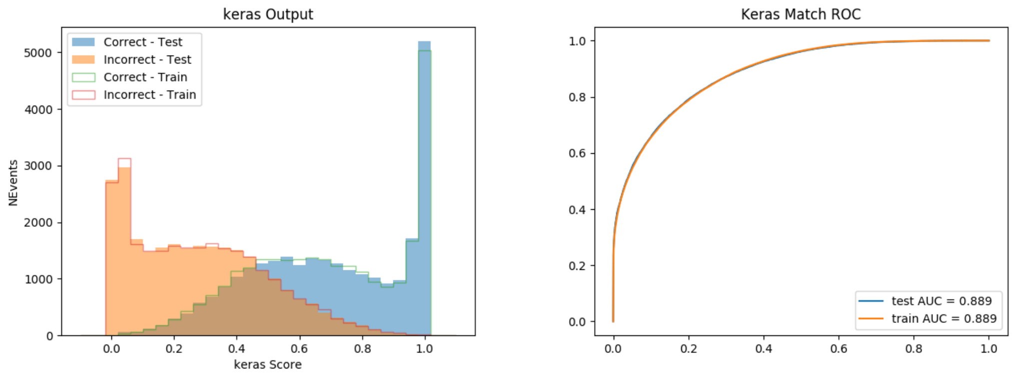
Figure 5.14: Data/MC comparisons of input features for higgs3lF training for 80 fb^{-1} of data.

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

Because the analysis uses a two bin fit of the Higgs p_T, the momentum reconstruction could be treated as a binary classification problem, rather than a regression problem. This approach is explored in detail in section A.4, and is found not to provide any significant increase in sensitivity. The regression approach is used because it provides more flexibility for future analyses, as it is 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 2lSS Channel

The input variables listed in table 10 are used to predict the Higgs p_T in the 2lSS channel. Here j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating



(a) (b)

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

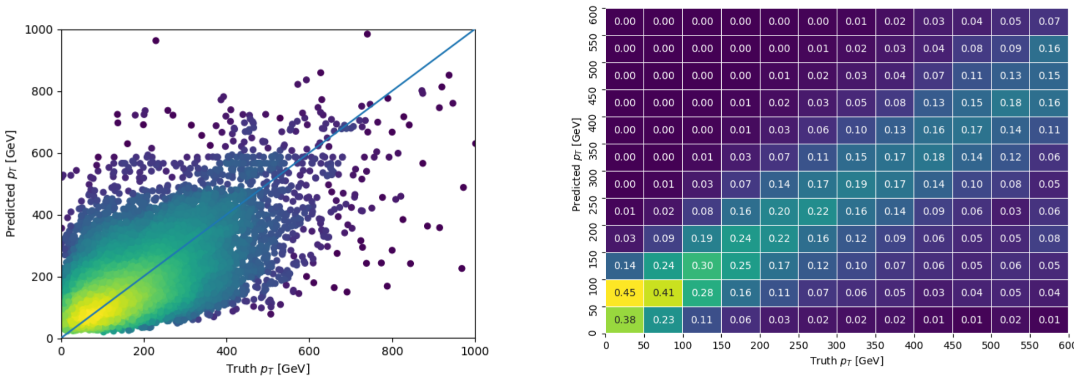
from the Higgs is labeled l_H, while the other lepton is labeled l_T, as it is assumed to have come from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and the b-jet identification algorithm, respectively.

HT	$M(j_0j_1)$	$M(j_0j_1l_H)$
$M(l_Hj_0)$	$M(l_Hj_1)$	$p_T(b_0b_1)$
$p_T(j_0j_1l_H)$	$\Delta\phi(j_0j_1l_H)(E_T^{miss})$	$\Delta R(j_0)(j_1)$
$\Delta R(j_0j_1)(l_H)$	$\Delta R(j_0j_1l_H)(l_T)$	$\Delta R(j_0j_1l_H)(b_0)$
$\Delta R(j_0j_1l_H)(b_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(b_0)$
$\Delta R(l_H)(b_1)$	$\Delta R(l_T)(b_0)$	$\Delta R(l_T)(b_1)$
$\Delta R(b_0)(b_1)$	Higgs Reco Score	jet η 0
jet η 1	jet Phi 0	jet Phi 1
jet p_T 0	jet p_T 1	Lepton η H
Lepton ϕ H	Lepton p_T H	Lepton p_T T
E_T^{miss}	nJets	b-jet Reco Score
b-jet p_T 0	b-jet p_T 1	

Table 10: Input features for reconstructing the Higgs p_T spectrum for 2lSS events

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

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



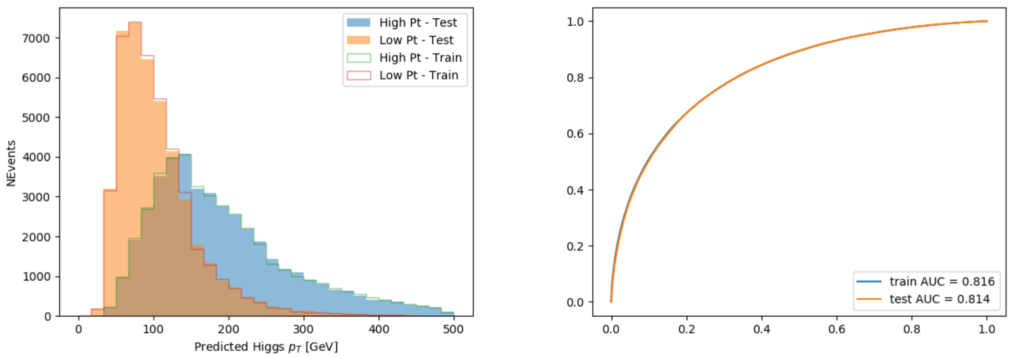
(a) (b)

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 Kernal Density Estimation, and (b) a histogram where each column has been normalized to one.

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

5.4.2 3l Semi-leptonic Channel

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



(a) (b)

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.

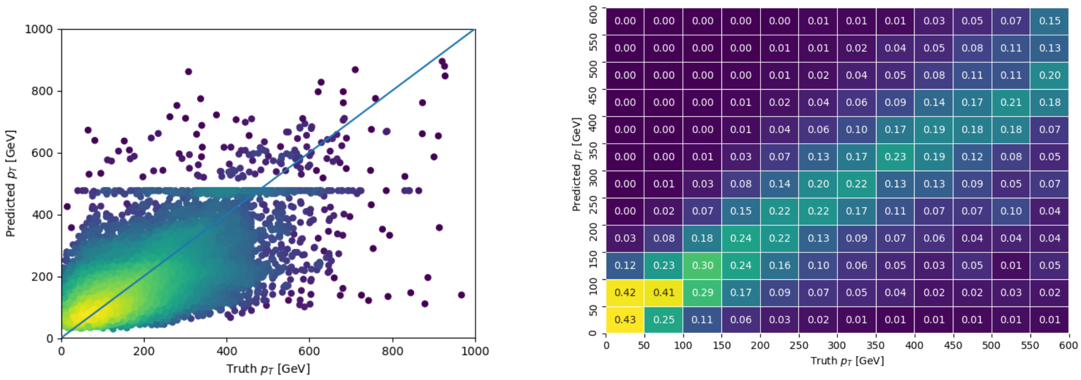
HT jets	MET	$M(j_0j_1)$
$M(j_0j_1l_H)$	$M(j_0j_1l_{T0})$	$M(j_0j_1l_{T1})$
$M(j_0j_1b_0)$	$M(j_0j_1b_1)$	$M(b_0l_{T0})$
$M(b_0l_{T1})$	$M(b_1l_{T0})$	$M(b_1l_{T1})$
$\Delta\phi(j_0j_1l_H)(E_T^{miss})$	$\Delta R(j_0)(j_1)$	$\Delta R(j_0j_1)(l_H)$
$\Delta R(j_0j_1)(l_{T1})$	$\Delta R(j_0j_1)(b_0)$	$\Delta R(j_0j_1)(b_1)$
$\Delta R(j_0j_1l_H)(l_{T0})$	$\Delta R(j_0j_1l_H)(l_{T1})$	$\Delta R(j_0j_1l_H)(b_0)$
$\Delta R(j_0j_1l_H)(b_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(l_H)(l_{T1})$	$\Delta R(l_{T0})(l_{T1})$	$\Delta R(l_{T0})(b_0)$
$\Delta R(l_{T0})(b_1)$	$\Delta R(l_{T1})(b_0)$	$\Delta R(l_{T1})(b_1)$
higgsScore	jet η 0	jet η 1
jet ϕ 0	jet ϕ 1	jet p_T 0
jet p_T 1	Lepton η H	Lepton ϕ H
Lepton p_T H	Lepton p_T T0	Lepton p_T T1
nJets	topScore	b-jet p_T 0
b-jet p_T 1		

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

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

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

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



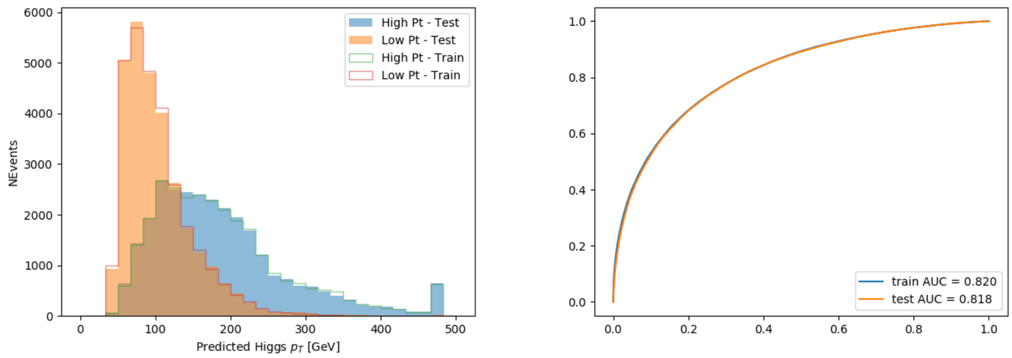
(a) (b)

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 Kernal Density Estimation, and (b) a histogram where each column has been normalized to one.

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

5.4.3 3l Fully-leptonic Channel

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



(a) (b)

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.

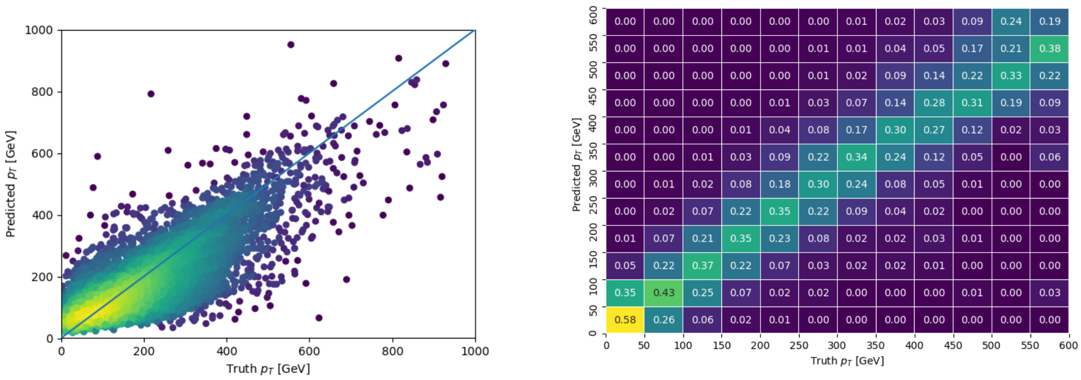
HT	$M(l_{H0}l_{H1})$	$M(l_{H0}l_T)$
$M(l_{H0}b_0)$	$M(l_{H0}b_1)$	$M(l_{H1}l_T)$
$M(l_{H1}b_0)$	$M(l_{H1}b_1)$	$\Delta R(l_{H0})(l_{H1})$
$\Delta R(l_{H0})(l_T)$	$\Delta R(l_{H0}l_{H1})(l_T)$	$\Delta R(l_{H0}l_T)(l_{H1})$
$\Delta R(l_{H1})(l_T)$	$\Delta R(l_{H0}b_0)$	$\Delta R(l_{H0}b_1)$
$\Delta R(l_{H1}b_1)$	$\Delta R(l_{H1}b_0)$	higgsScore
Lepton η H_0	Lepton η H_1	Lepton η T
Lepton p_T H_0	Lepton p_T H_1	Lepton p_T T
E_T^{miss}	topScore	b-jet p_T 0
b-jet p_T 1		

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

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

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

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



(a) (b)

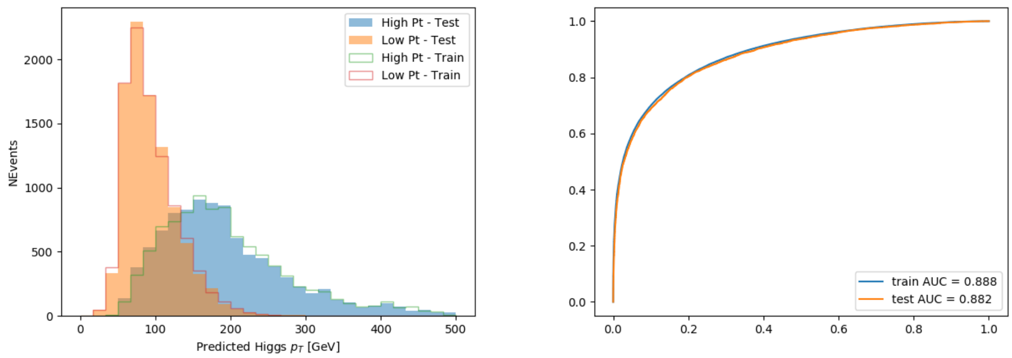
Figure 5.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized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

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

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 13.



(a) (b)

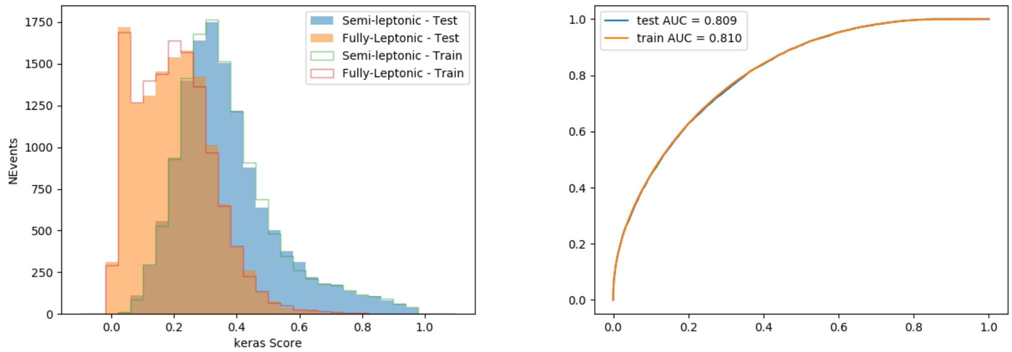
Figure 5.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

HT jets	$M(l_0t_0)$	$M(l_0t_1)$
$M(l_1t_0)$	$M(l_1t_1)$	$M(l_0l_1)$
$M(l_0l_2)$	$M(l_1l_2$	$\Delta R(l_0t_0)$
$\Delta R(l_0t_1)$	$\Delta R(l_1t_0)$	$\Delta R(l_1t_1)$
$\Delta R(ll_01)$	$\Delta R(ll_02)$	$\Delta R(ll_12)$
Lepton η 0	Lepton η 1	Lepton η 2
Lepton ϕ 0	Lepton ϕ 1	Lepton ϕ 2
Lepton p_T 0	Lepton p_T 1	Lepton p_T 2
E_T^{miss}	nJets	nJets OR DL1r 60
nJets OR DL1r 85	score3lF	score3lS
topScore	total charge	

Table 13: Input features

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

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



(a) (b)

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.

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

6 Signal Region Definitions

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

6.1 Pre-MVA Event Selection

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

- Two very tight, same-charge, light leptons with $p_T > 20$ GeV
- $\geq$4 reconstructed jets, $\geq$1 b-tagged jets
- No reconstructed tau candidates

The event yield after the 2lSS preselection has been applied, for MC and data at 80 fb^{-1}, is shown in figure 6.1. Good general agreement is found.

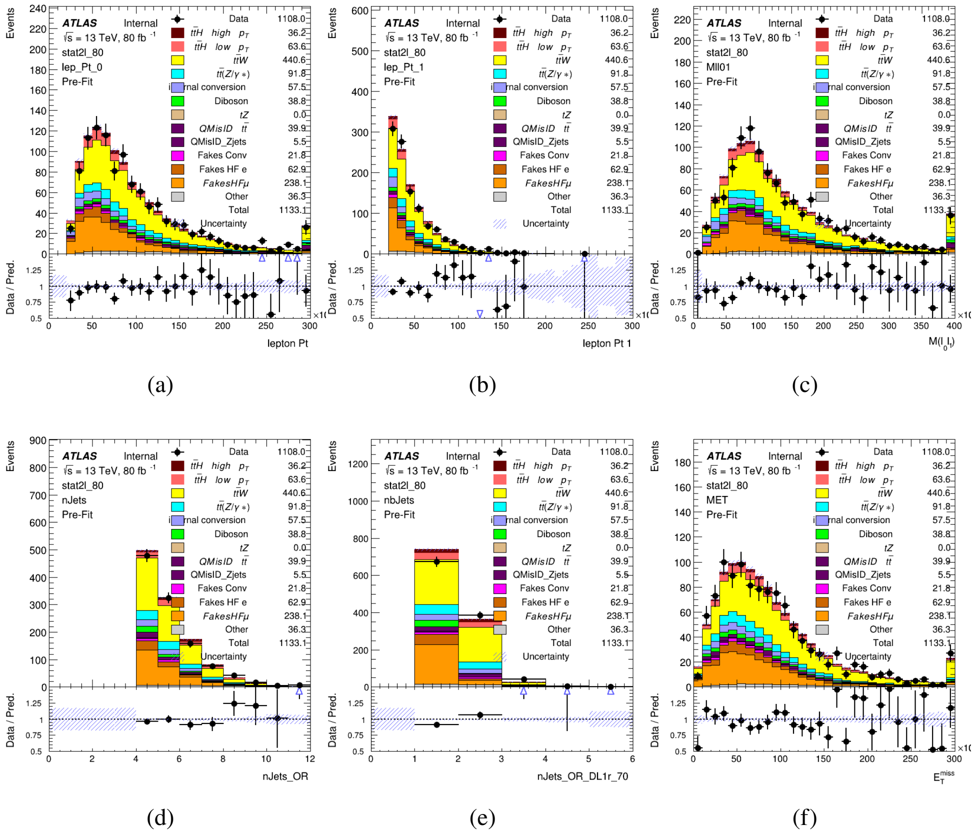
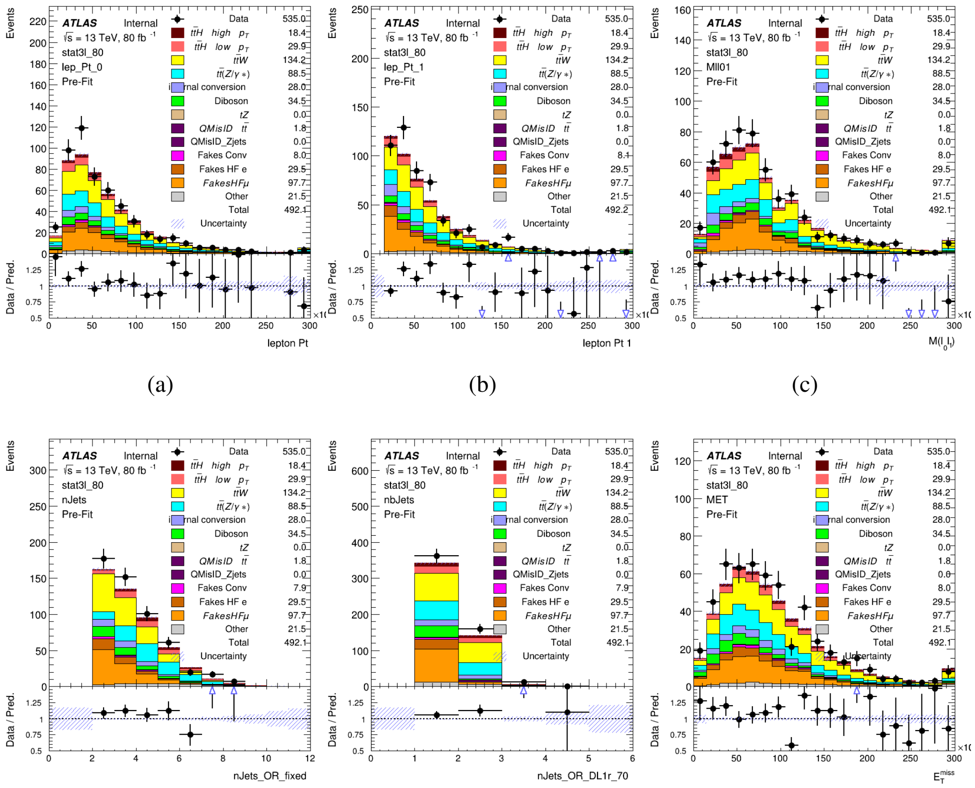


Figure 6.1: Data/MC comparisons of the 2lSS pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

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

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

Comparisons of data and MC event yields after this selection has been applied are shown in figure 6.2.



(a) (b) (c)

(d) (e) (f)

Figure 6.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

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

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

HT	M(lep, E_T^{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 $\Delta R(l_0)(jet)$
min $\Delta R(l_1)(jet)$	min $\Delta R(Lepton)(bjet)$	mjjMax frwdJet
nJets	nJets OR DL1r 60	nJets OR DL1r 70
nJets OR DL1r 85	topRecoScore	

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

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

M(lep, E_T^{miss})	$M(l_0 l_1)$	$M(l_0 l_1 l_2)$
$M(l_0 l_2)$	$M(l_1 l_2)$	binHiggs p_T 3lF
binHiggs p_T 3lS	$\Delta R(l_0)(l_1)$	$\Delta R(l_0)(l_2)$
$\Delta R(l_1)(l_2)$	decayScore	higgsRecoScore3lF
higgsRecoScore3lS	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 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}	min $\Delta R(l_0)(jet)$
min $\Delta R(l_1)(jet)$	min $\Delta R(l_2)(jet)$	min $\Delta R(Lepton)(bjet)$
mjjMax frwdJet	nJets	nJets OR DL1r 60
nJets OR DL1r 70	nJets OR DL1r 85	topScore

Table 15: Input features used to distinguish signal and background events in the 3l channel.

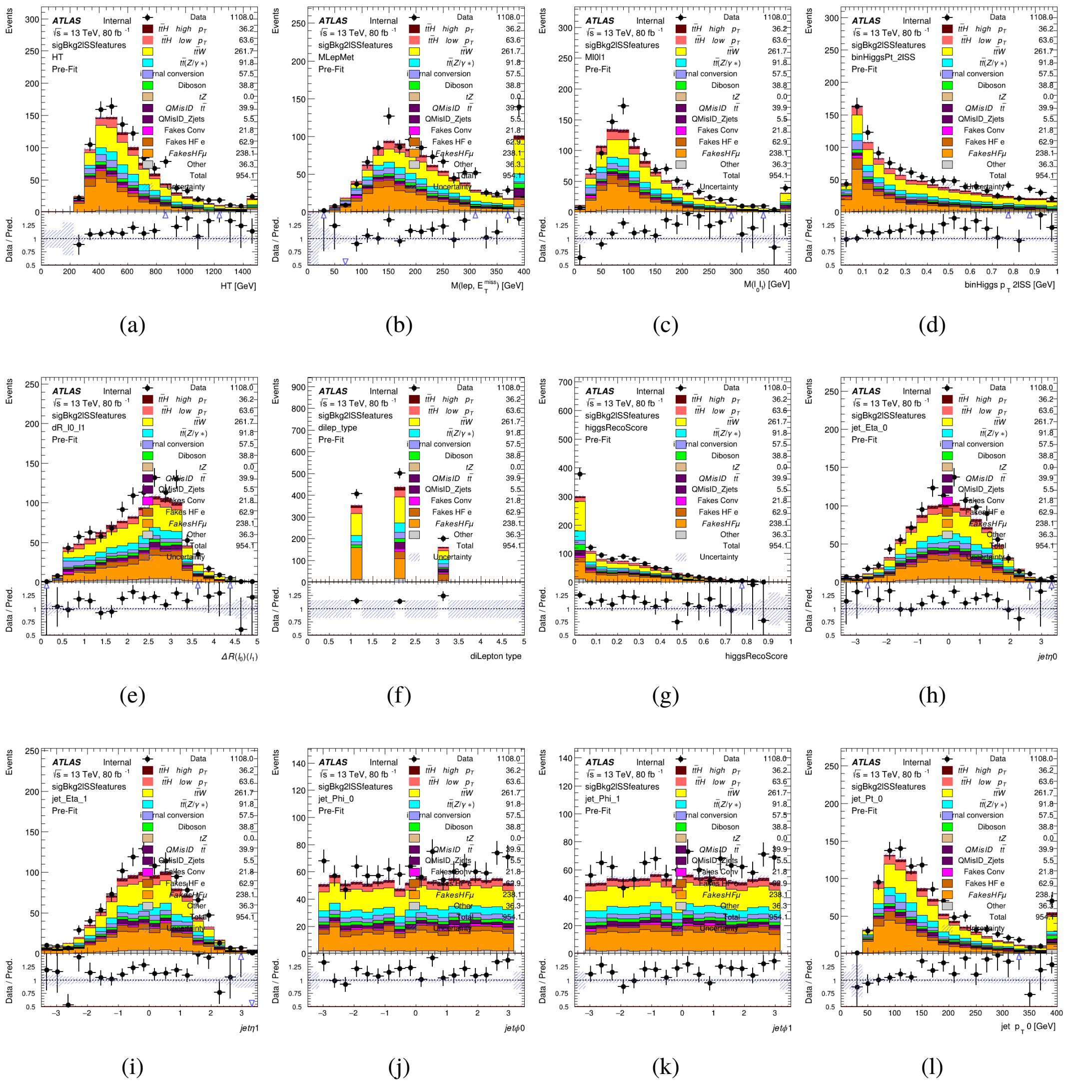


Figure 6.3:

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

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

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

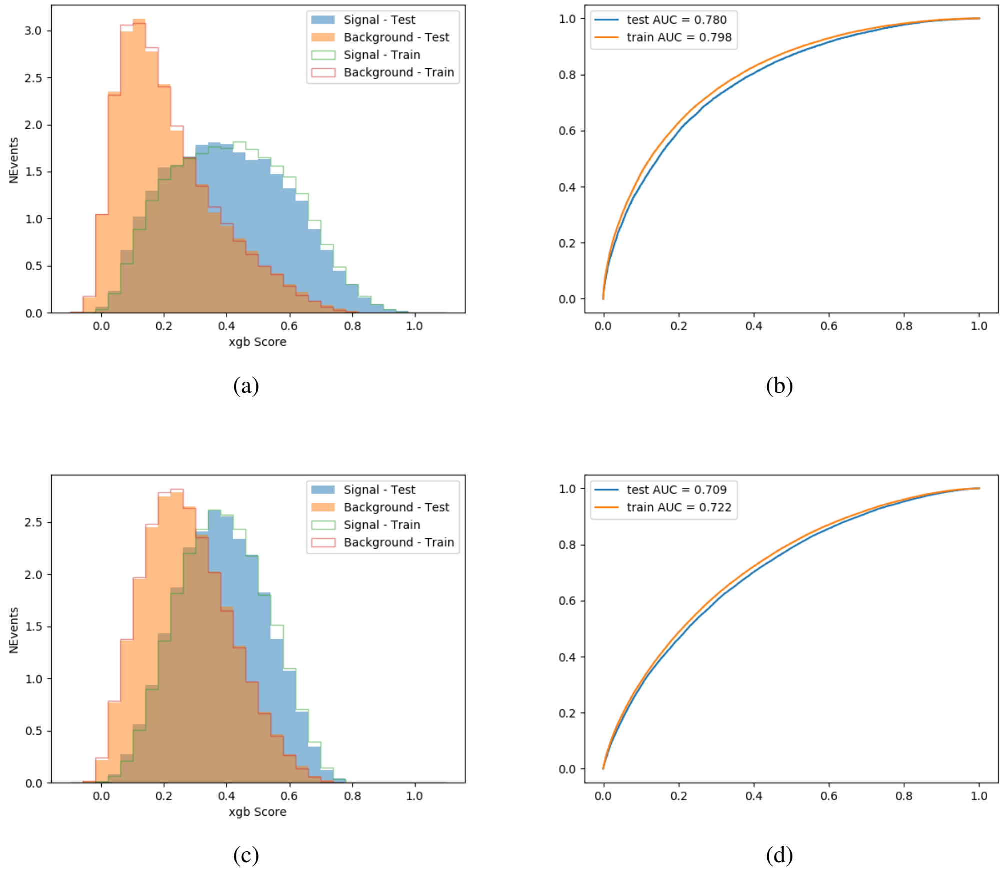
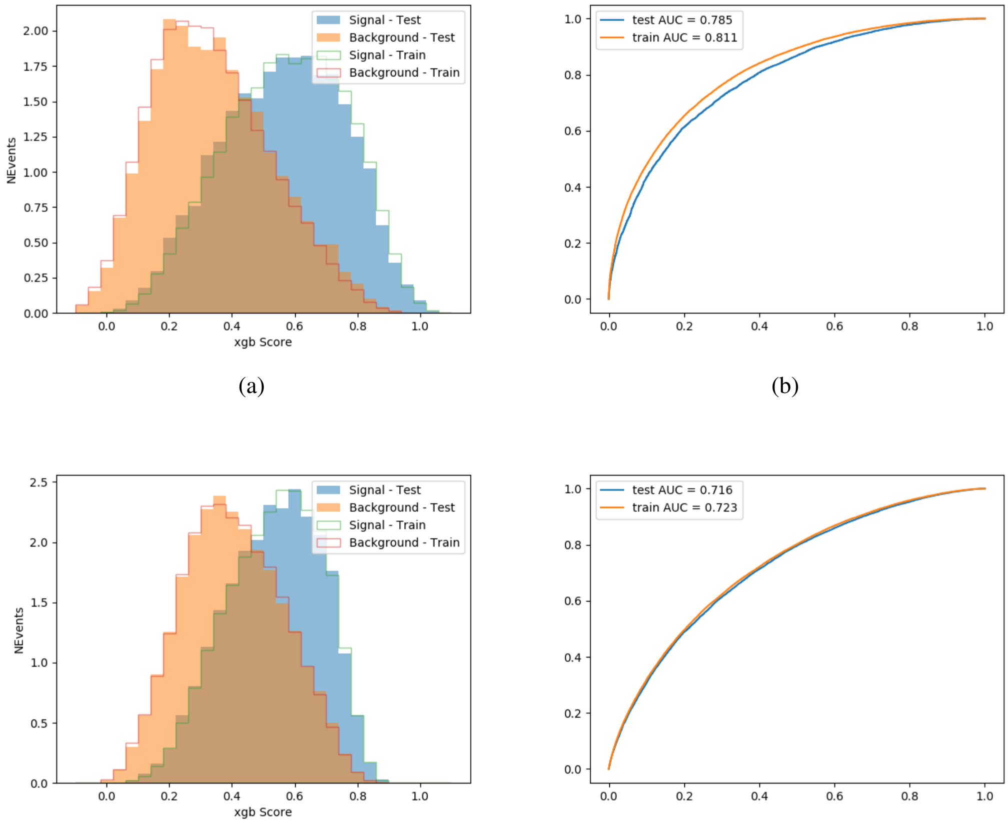


Figure 6.4:

two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson. This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

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

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

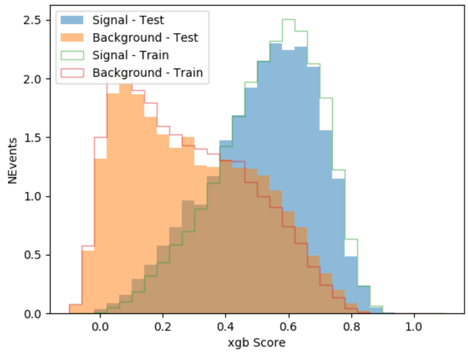


(a) (b) (c) (d)

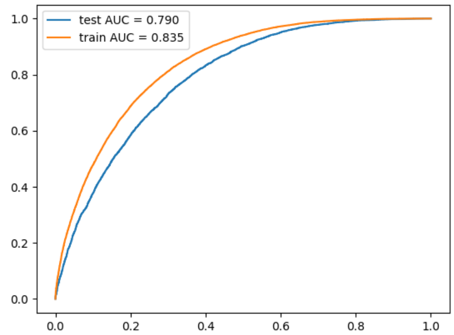
Figure 6.5:

Channel	BDT Score
2lSS high p_T	0.36
2lSS low p_T	0.34
3lS high p_T	0.51
3lS low p_T	0.43
3lF high p_T	0.33
3lF low p_T	0.41

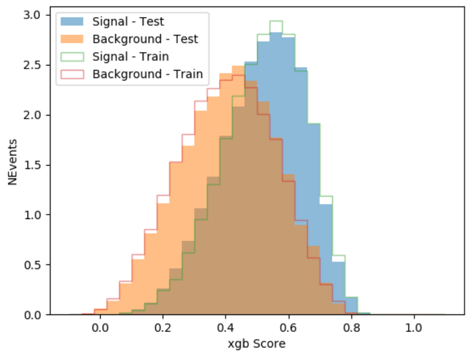
Table 16: Cutoff values on background rejection MVA score applied to signal regions.



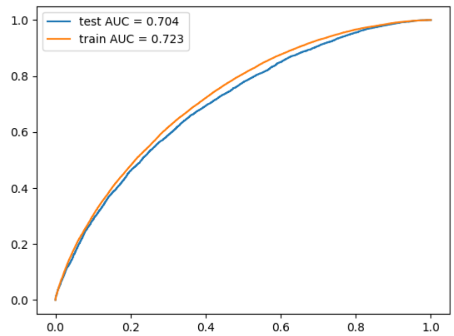
(a)



(b)



(c)



(d)

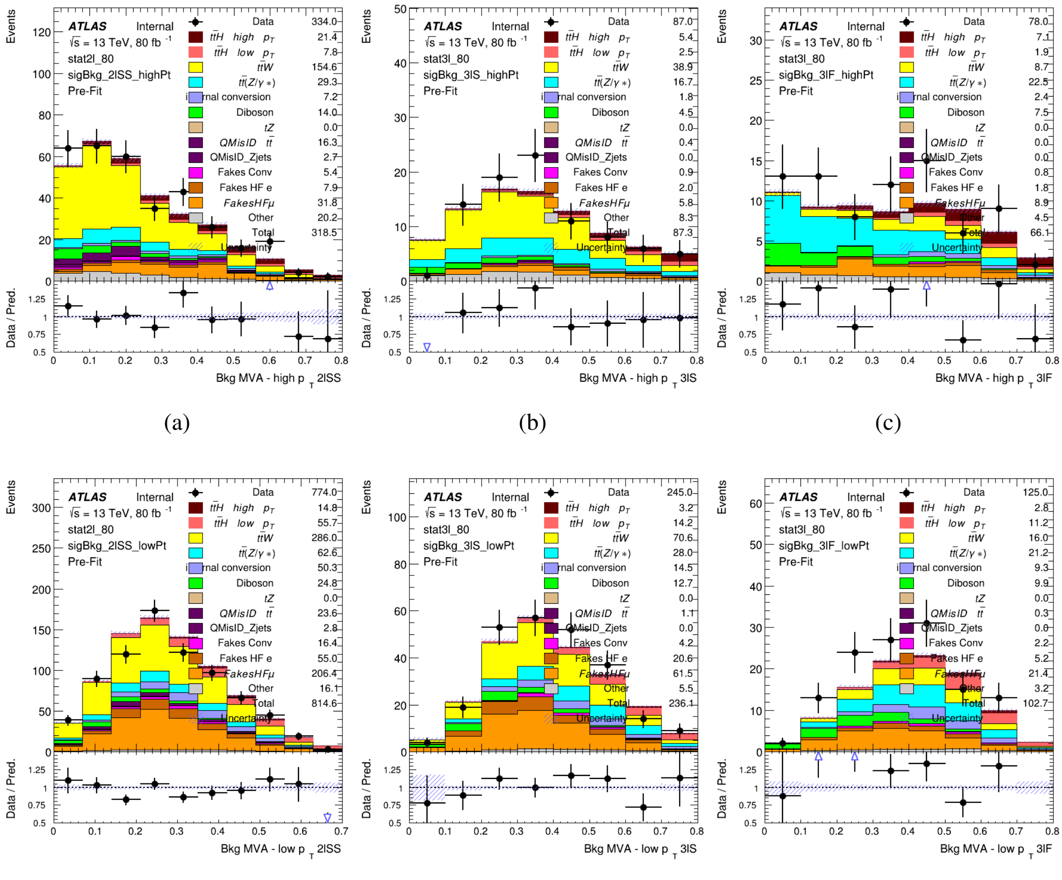
Figure 6.6:

7 Systematic Uncertainties

The systematic uncertainties that are considered are summarized in table 16. These are implemented in the fit either as a normalization factors or as a shape variation or both in the signal and background estimations. The numerical impact of each of these uncertainties is outlined in section 8.

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

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



(a) (b) (c)

(d) (e) (f)

Figure 6.7: scores

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

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

8 Results

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

Region	Selection
2lSS	Two same charge tight leptons with $p_T > 20$ GeV
	$N_{jets} \geq 4$, $N_{b-jets} \geq 1$ b-tagged jets
	Zero τ_{had}
	$H_{p_T}^{pred}$ >150 GeV and BDT score > 0.36 **or** $H_{p_T}^{pred}$ <150 GeV and BDT score > 0.34
3lS	Three light leptons with total charge ± 1
	Two tight SS leptons, $p_T > 20$ GeV
	One loose OS lepton, $p_T > 10$ GeV
	$N_{jets} \geq 2$, $N_{b-jets} \geq 1$ b-tagged jets
	Zero τ_{had}
	$\|M(l^+l^-) - 91.2\text{ GeV}\| > 10$ GeV for all OSSF lepton pairs
	Decay NN Score < 0.23
	$H_{p_T}^{pred}$ >150 GeV and BDT score > 0.51 **or** $H_{p_T}^{pred}$ >150 GeV and BDT score > 0.43
3lF	Three light leptons with total charge ± 1
	Two tight SS leptons, $p_T > 20$ GeV
	One loose OS lepton, $p_T > 10$ GeV
	$N_{jets} \geq 2$, $N_{b-jets} \geq 1$ b-tagged jets
	Zero τ_{had}
	$\|M(l^+l^-) - 91.2\text{ GeV}\| > 10$ GeV for all OSSF lepton pairs
	Decay NN Score > 0.23
	$H_{p_T}^{pred}$ >150 GeV and BDT score > 0.33 **or** $H_{p_T}^{pred}$ >150 GeV and BDT score > 0.41

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

8.1 Results - 80 fb^{-1}

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

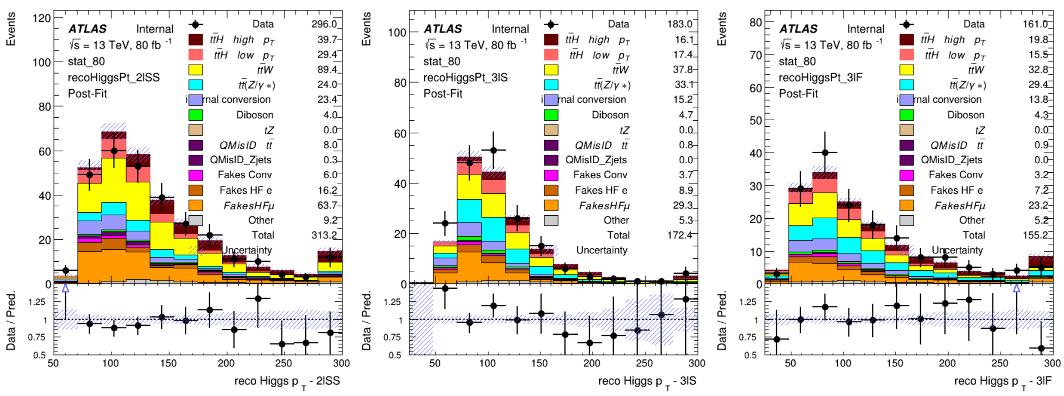
8.2 Projected Results - 140 fb^{-1}

9 Conclusion

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

Table 18: 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



(a) (b) (c)

Figure 8.1:

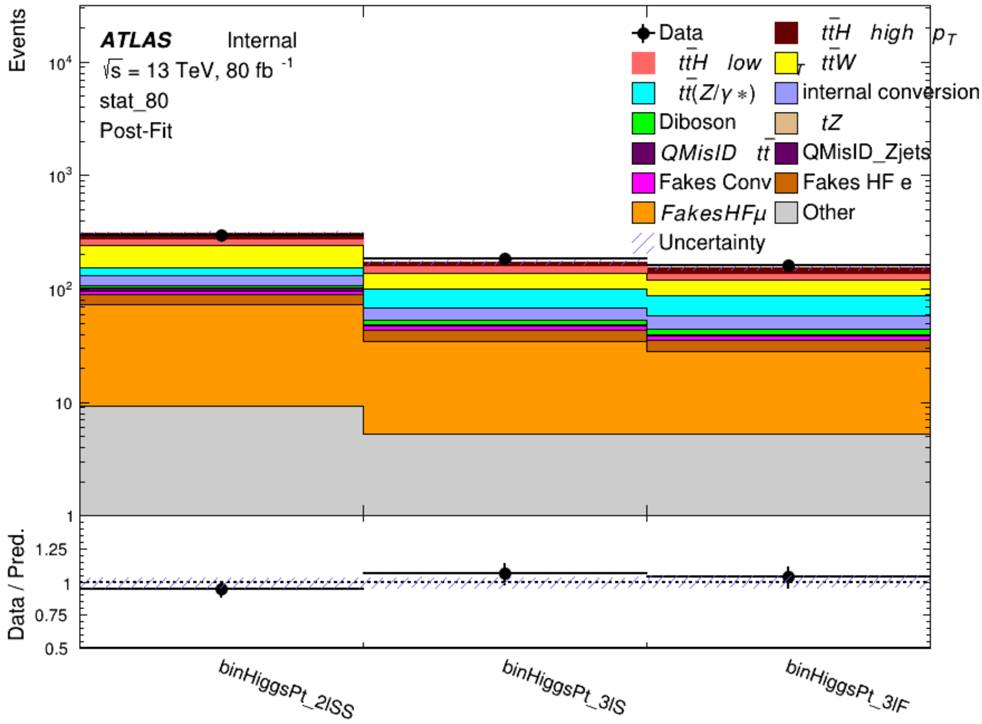


Figure 8.2: Post-fit summary of fit.

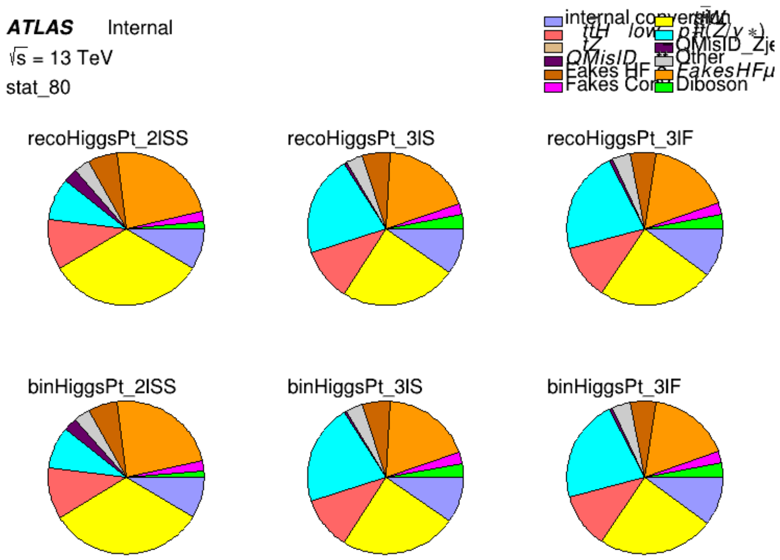
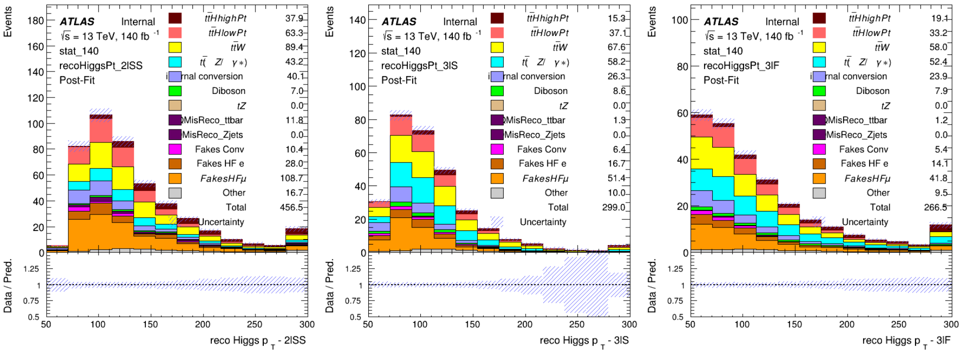


Figure 8.3: Background composition of the fit regions.



(a) (b) (c)

Figure 8.4:

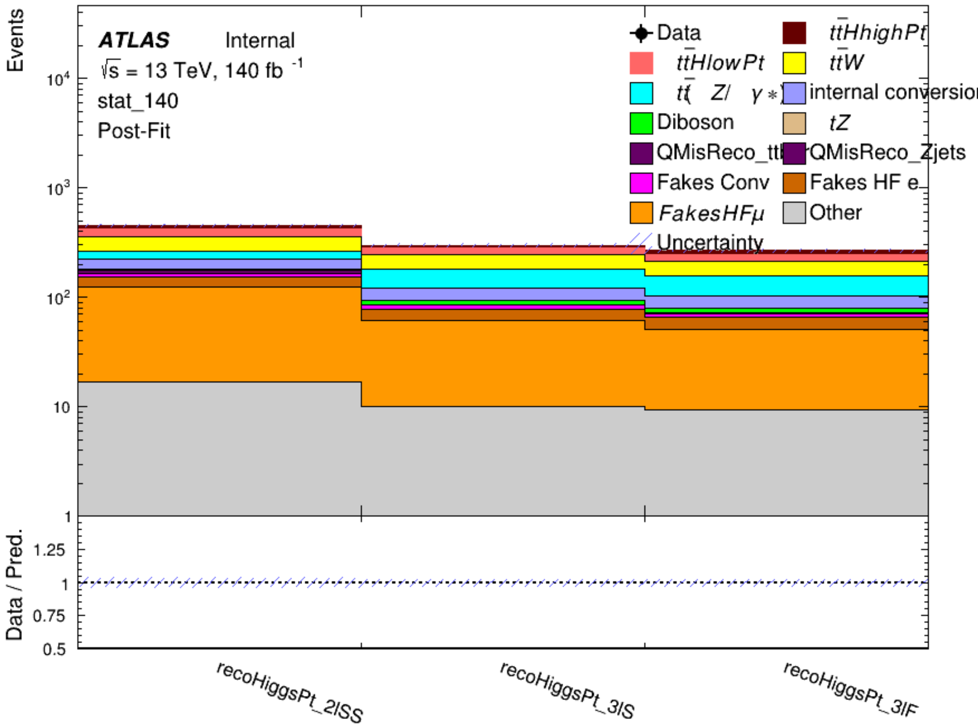


Figure 8.5: Post-fit summary of fit.

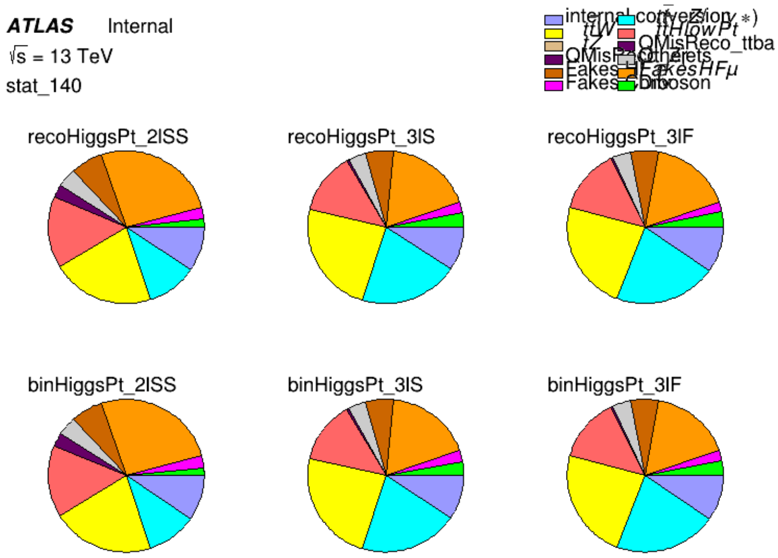


Figure 8.6: Background composition of the fit regions.

List of contributions

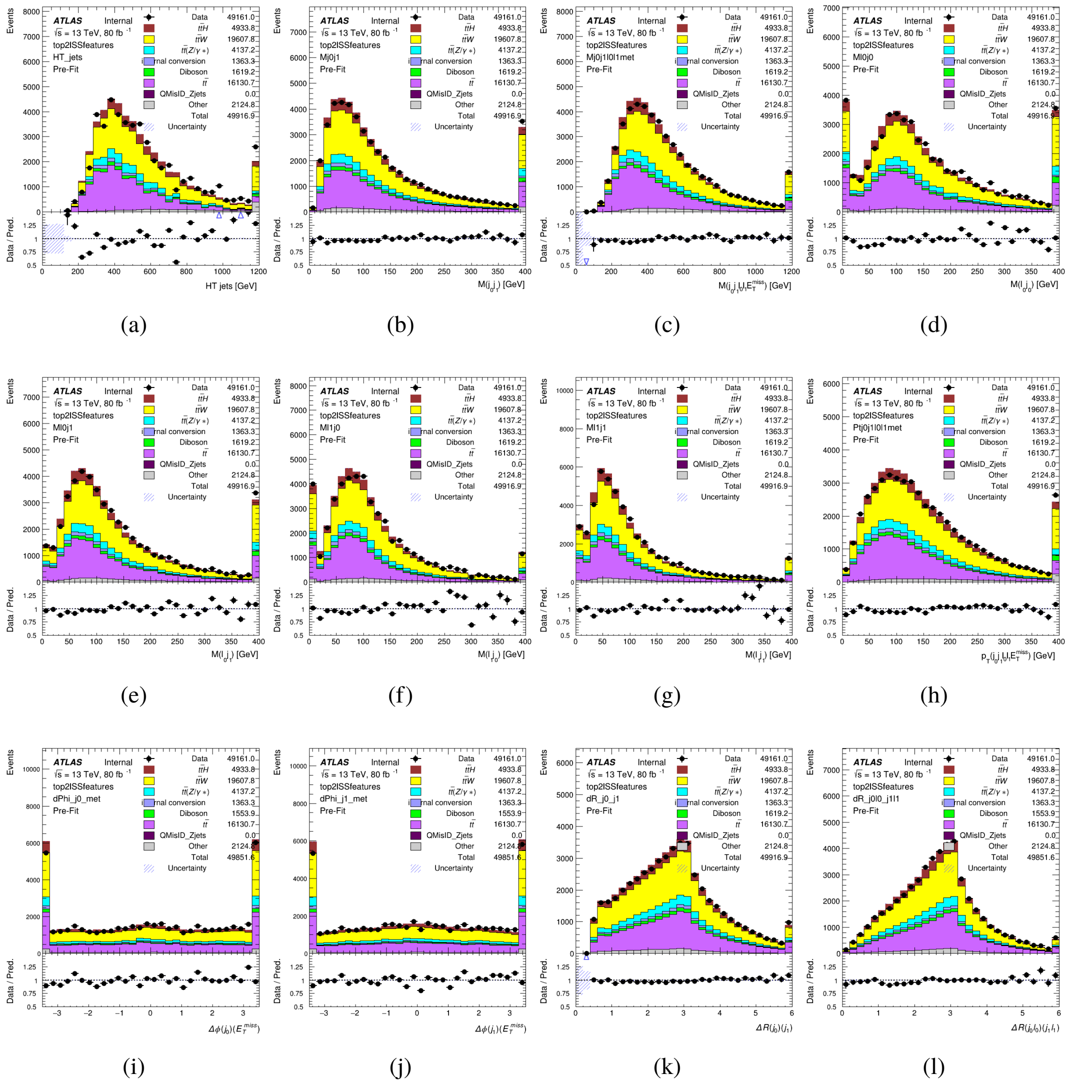
Appendices

A Machine Learning Models

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

A.1 Higgs Reconstruction Models

A.1.1 b-jet Identification Features - 2lSS



(a) (b) (c) (d)
(e) (f) (g) (h)
(i) (j) (k) (l)

Figure A.1: Input features for top2lSS

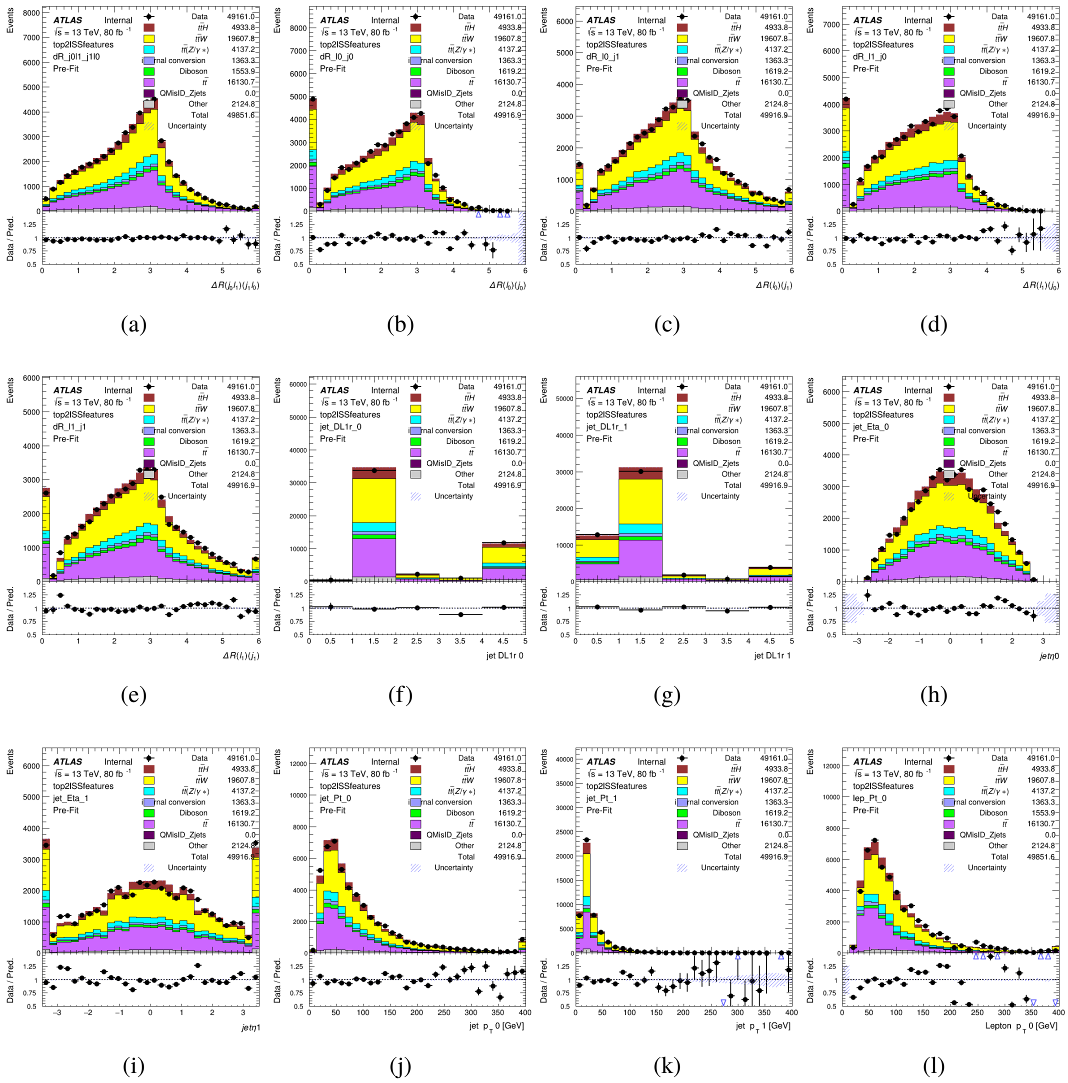
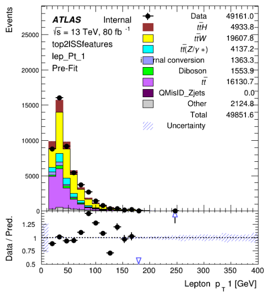
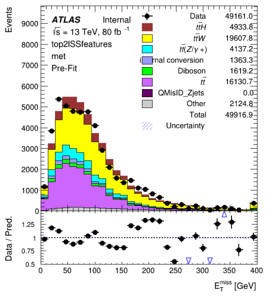


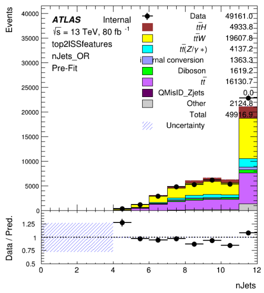
Figure A.2: Input features for top2lSS



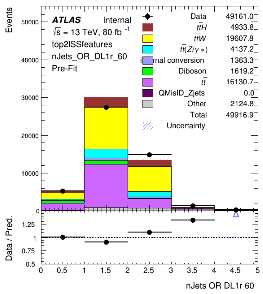
(a)



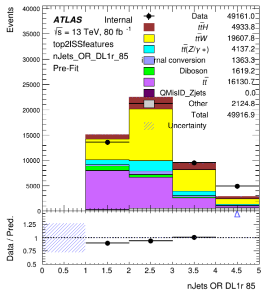
(b)



(c)



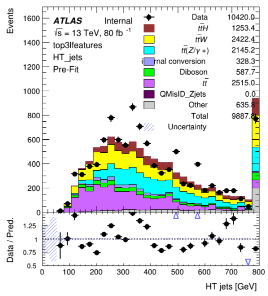
(d)



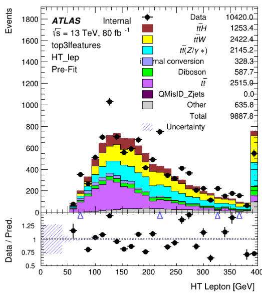
(e)

Figure A.3: Input features for top2lSS

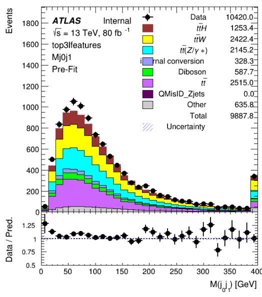
481 ### A.1.2 b-jet Identification Features - 3l



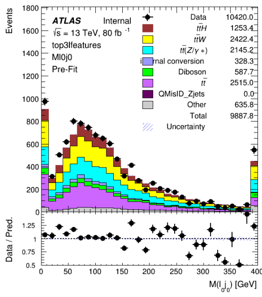
(a)



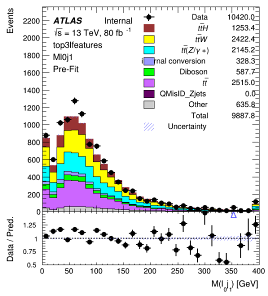
(b)



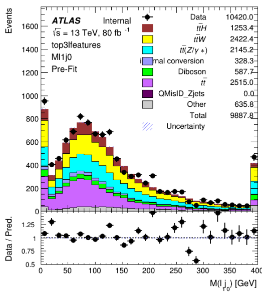
(c)



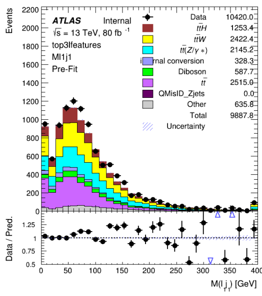
(d)



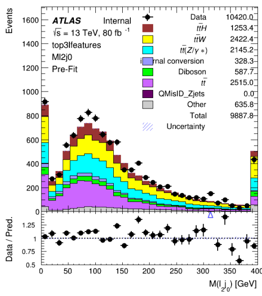
(e)



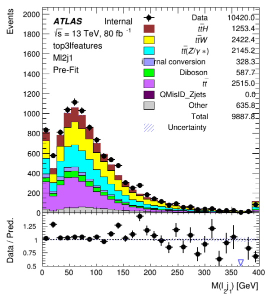
(f)



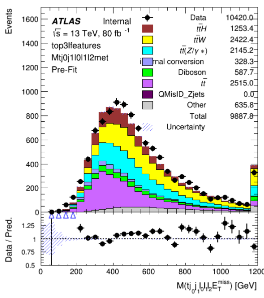
(g)



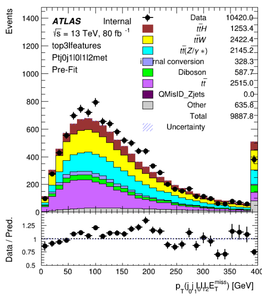
(h)



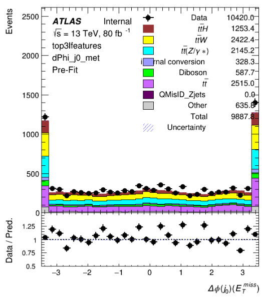
(i)



(j)

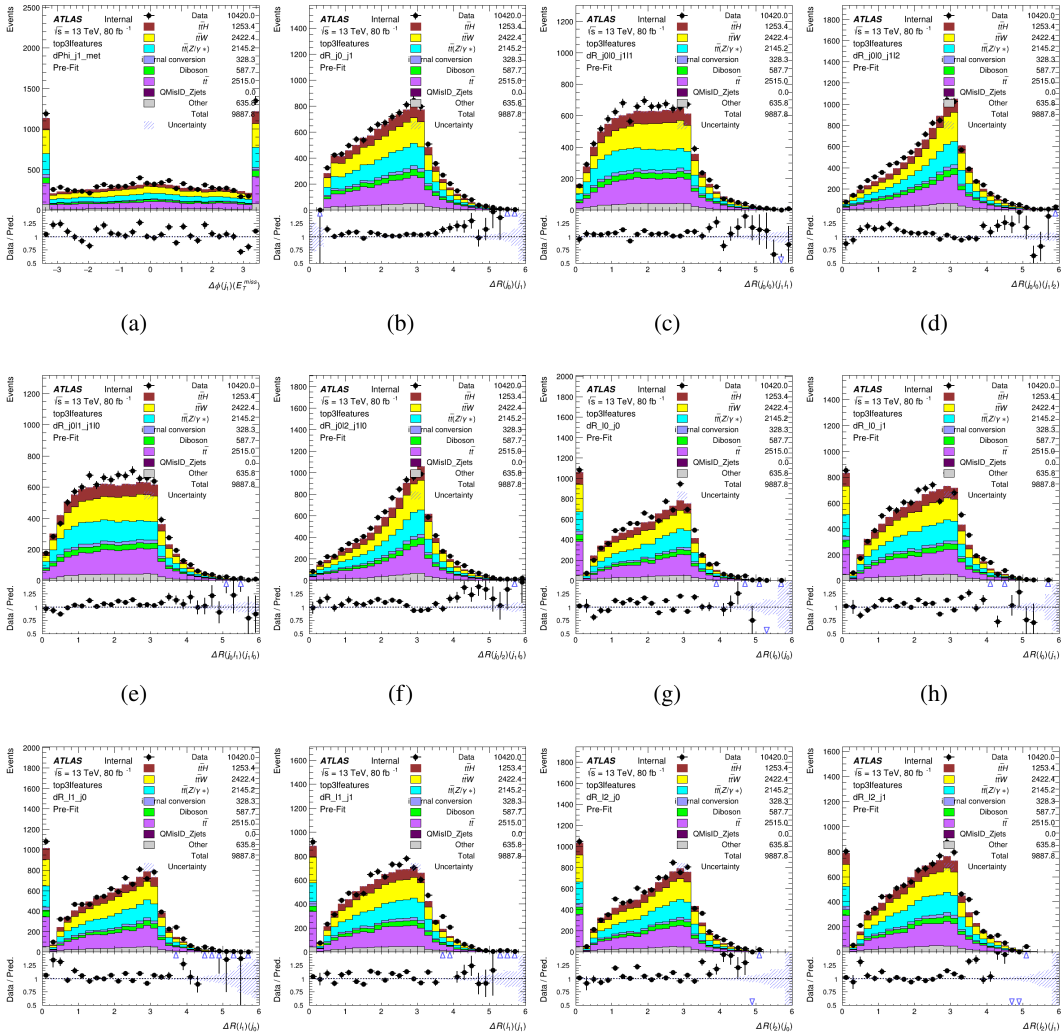


(k)



(l)

Figure A.4: Input features for top3l

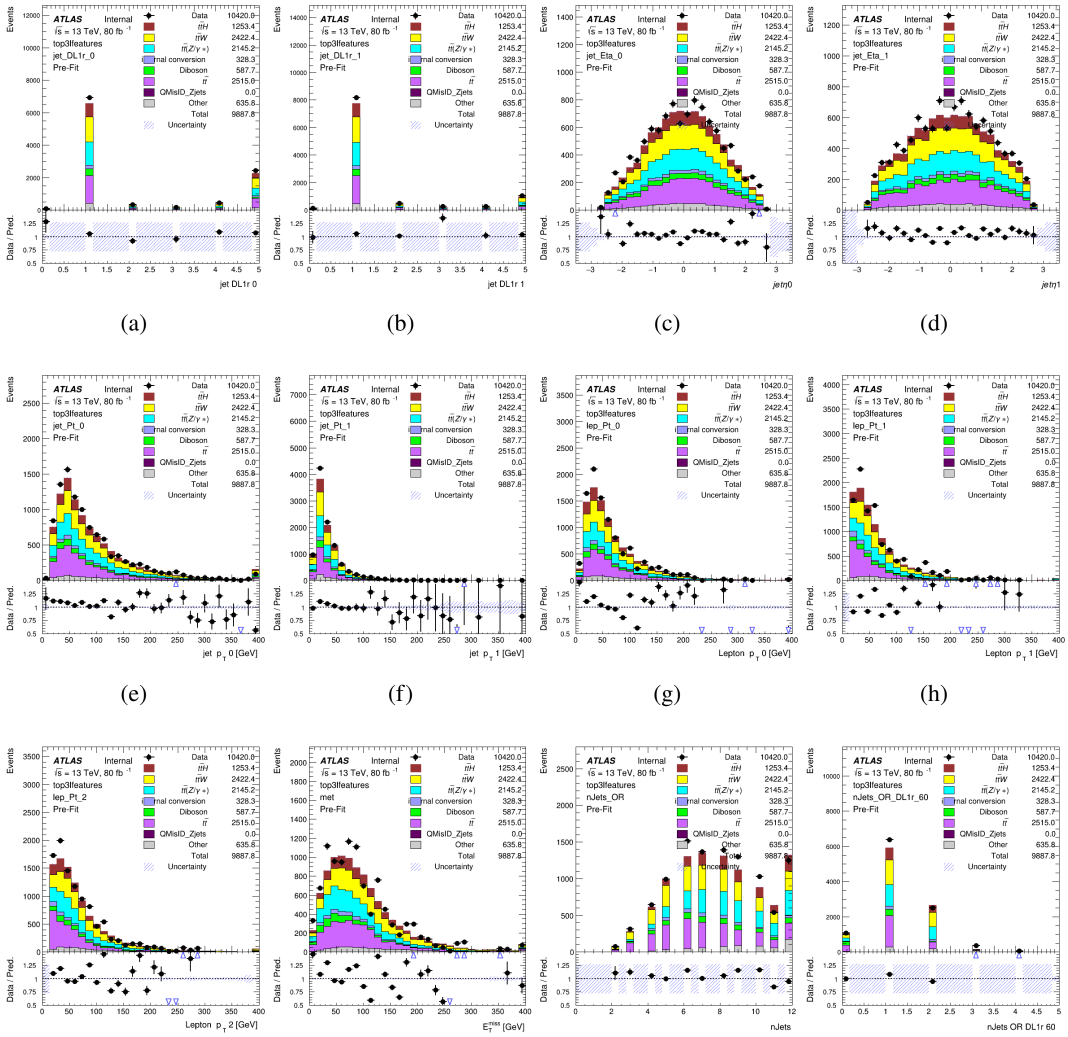


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure A.5: Input features for top3l

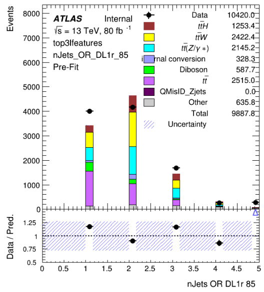


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure A.6: Input features for top3l



(a)

Figure A.7: Input features for top3l

A.1.3 Higgs Reconstruction Features - 2lSS

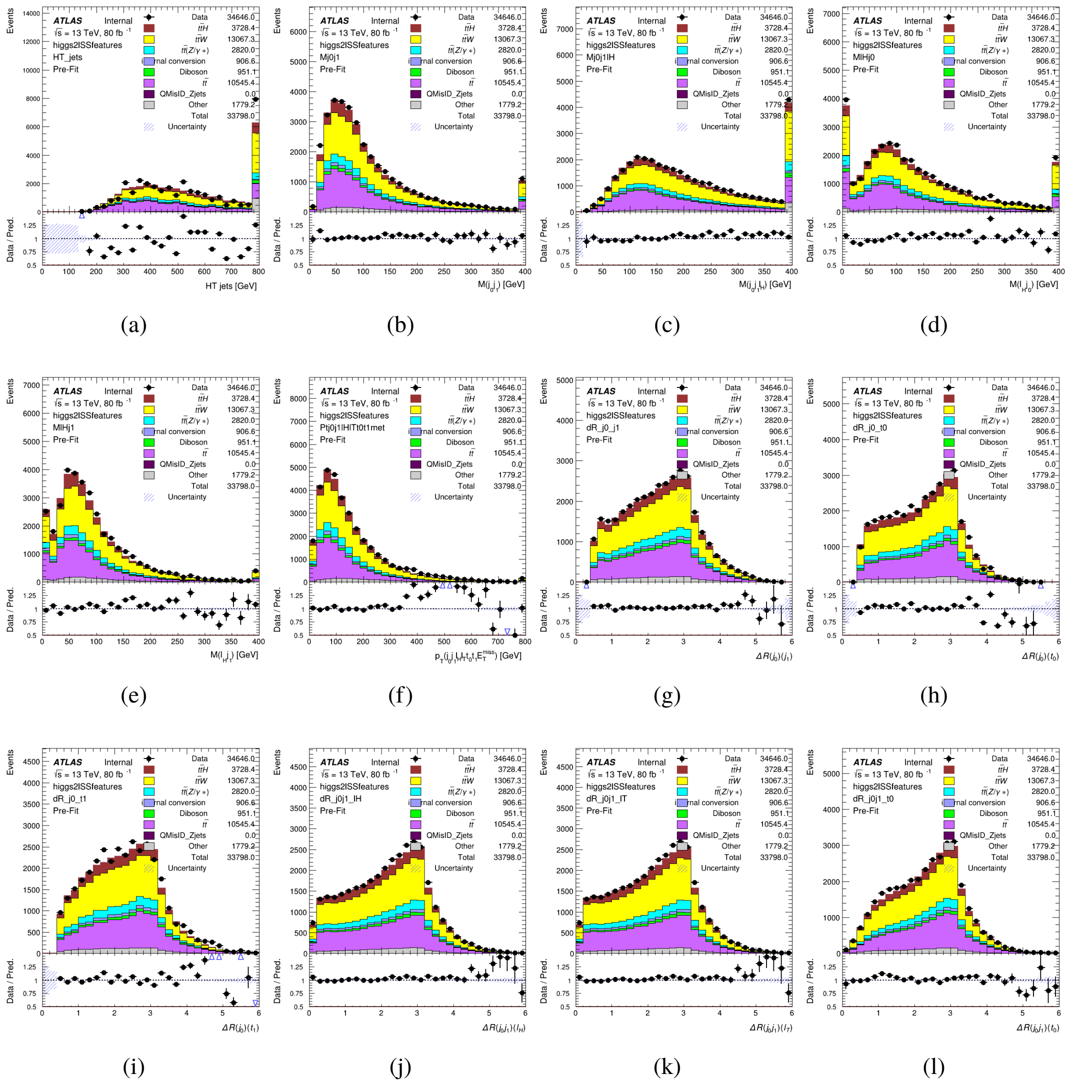


Figure A.8: Input features for higgs2lSS

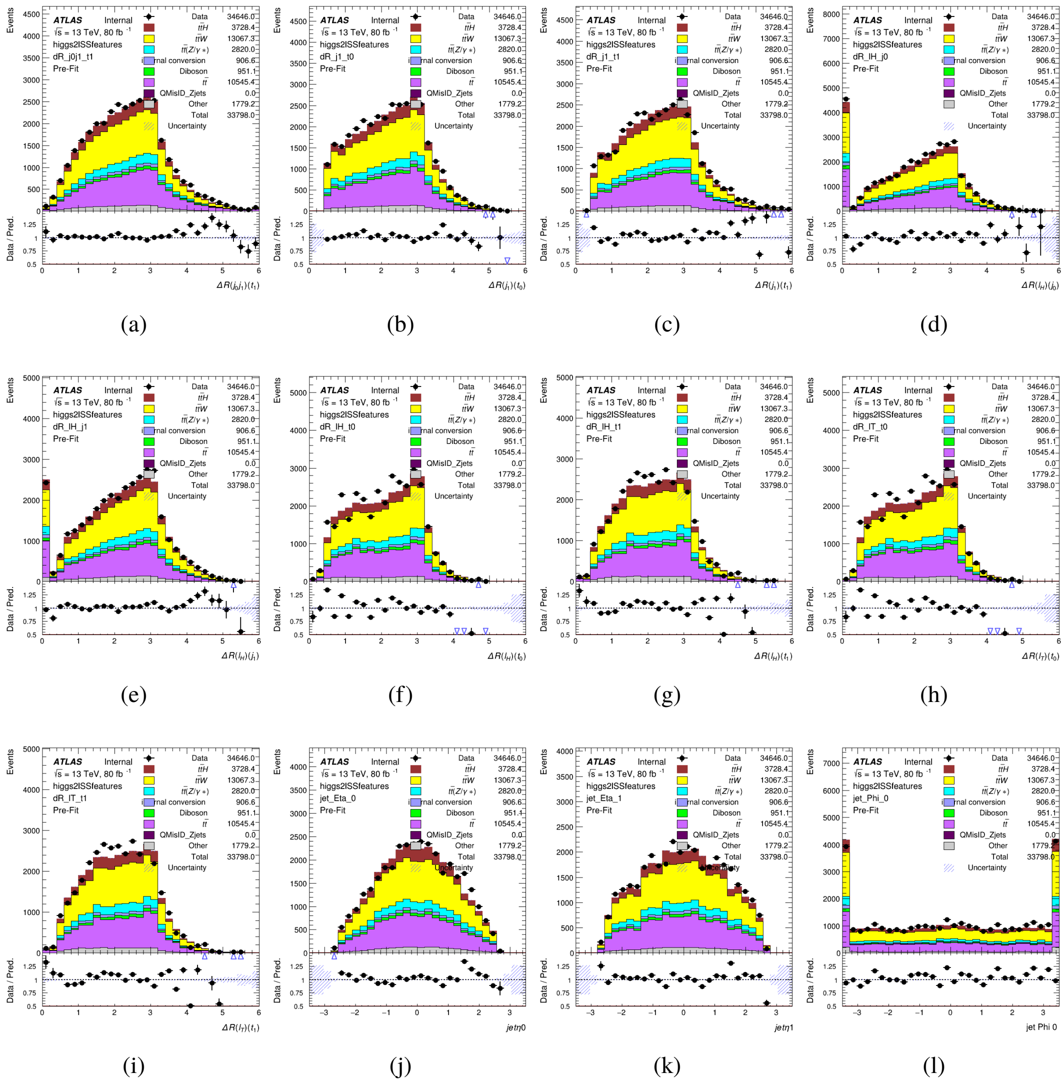
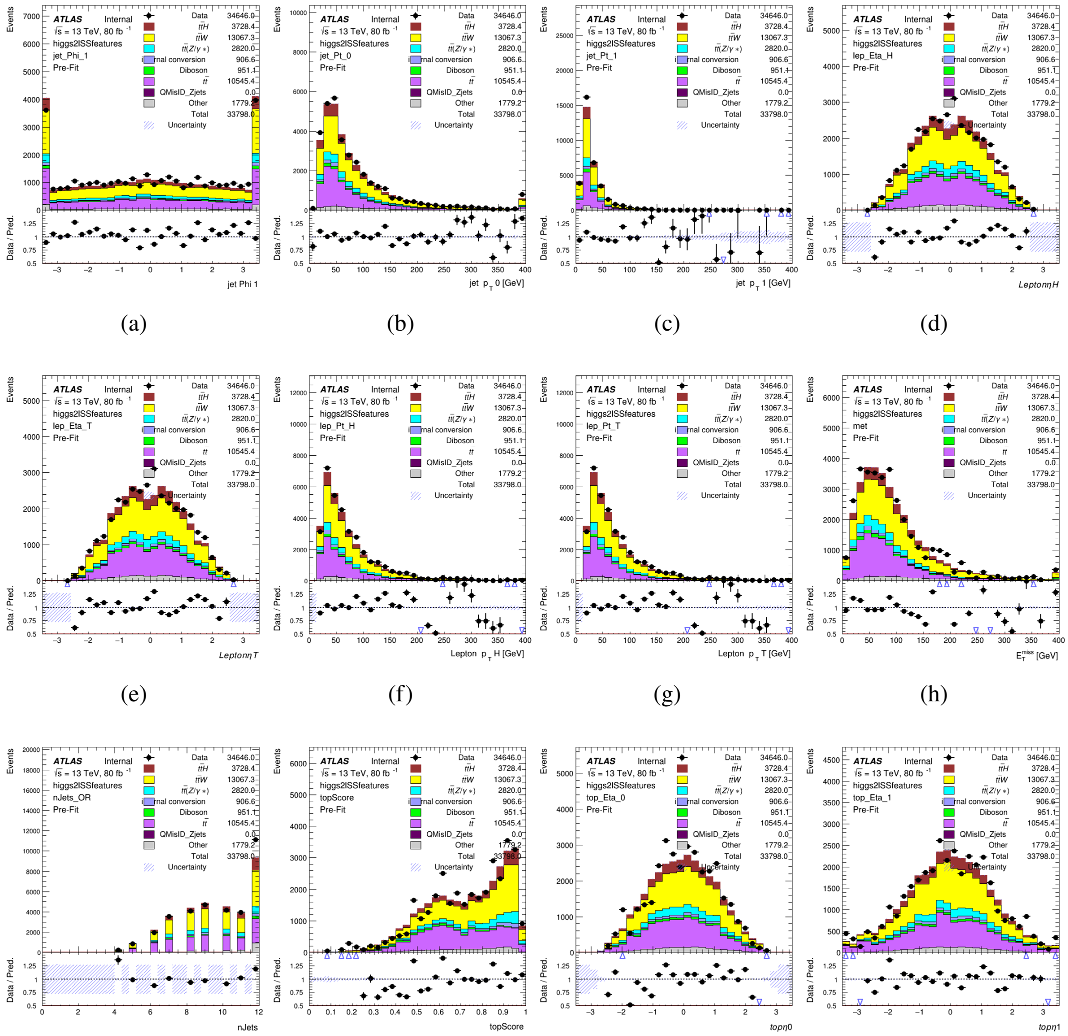


Figure A.9: Input features for higgs2lSS

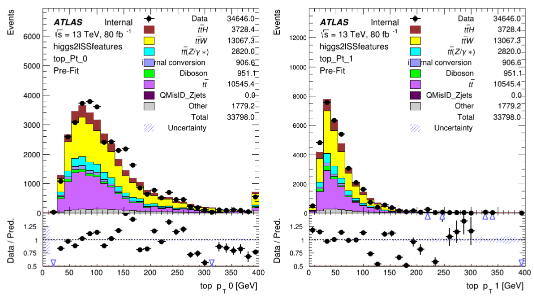


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure A.10: Input features for higgs2lSS



(a) (b)

Figure A.11: Input features for higgs2lSS

483 ### A.1.4 Higgs Reconstruction Features - 3lS

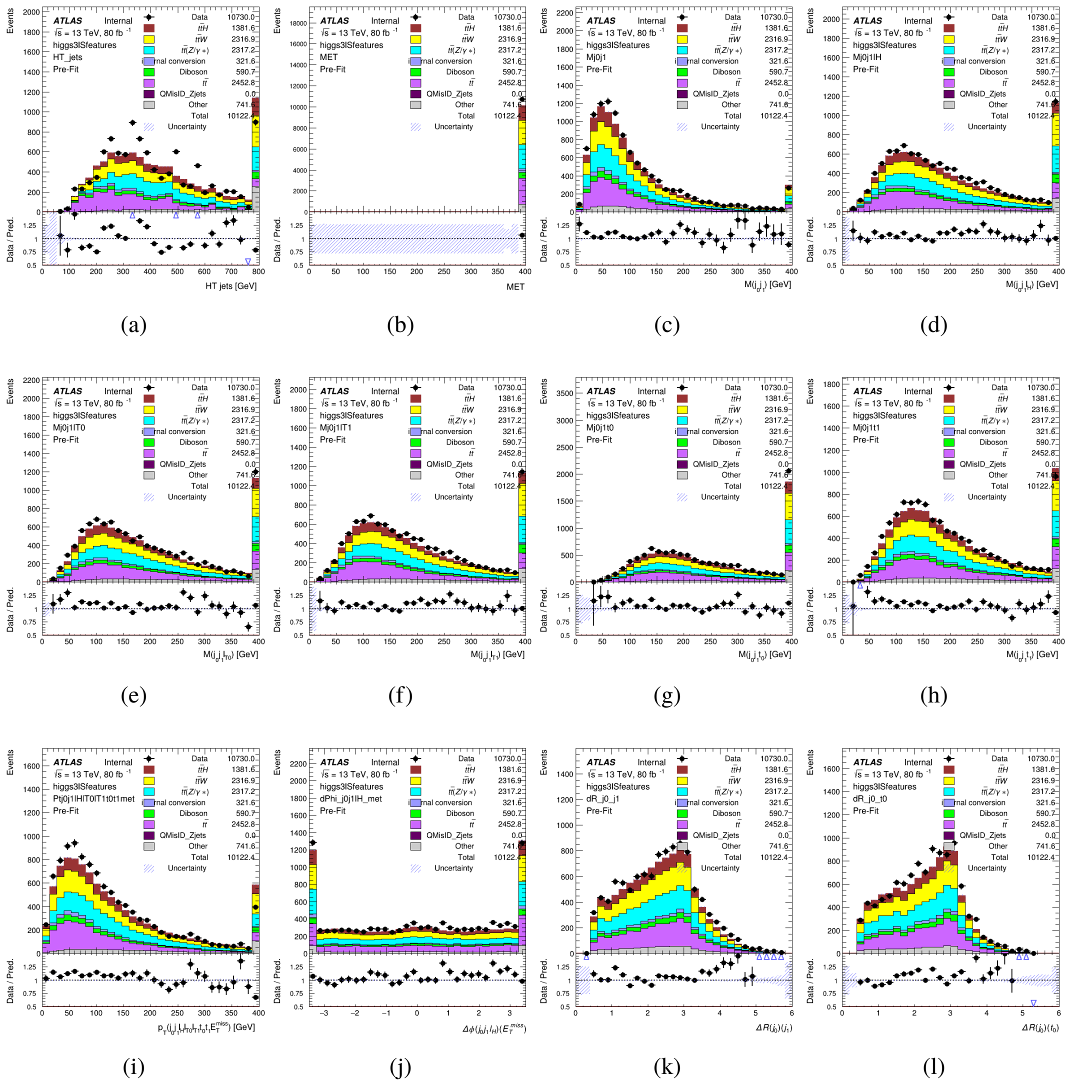
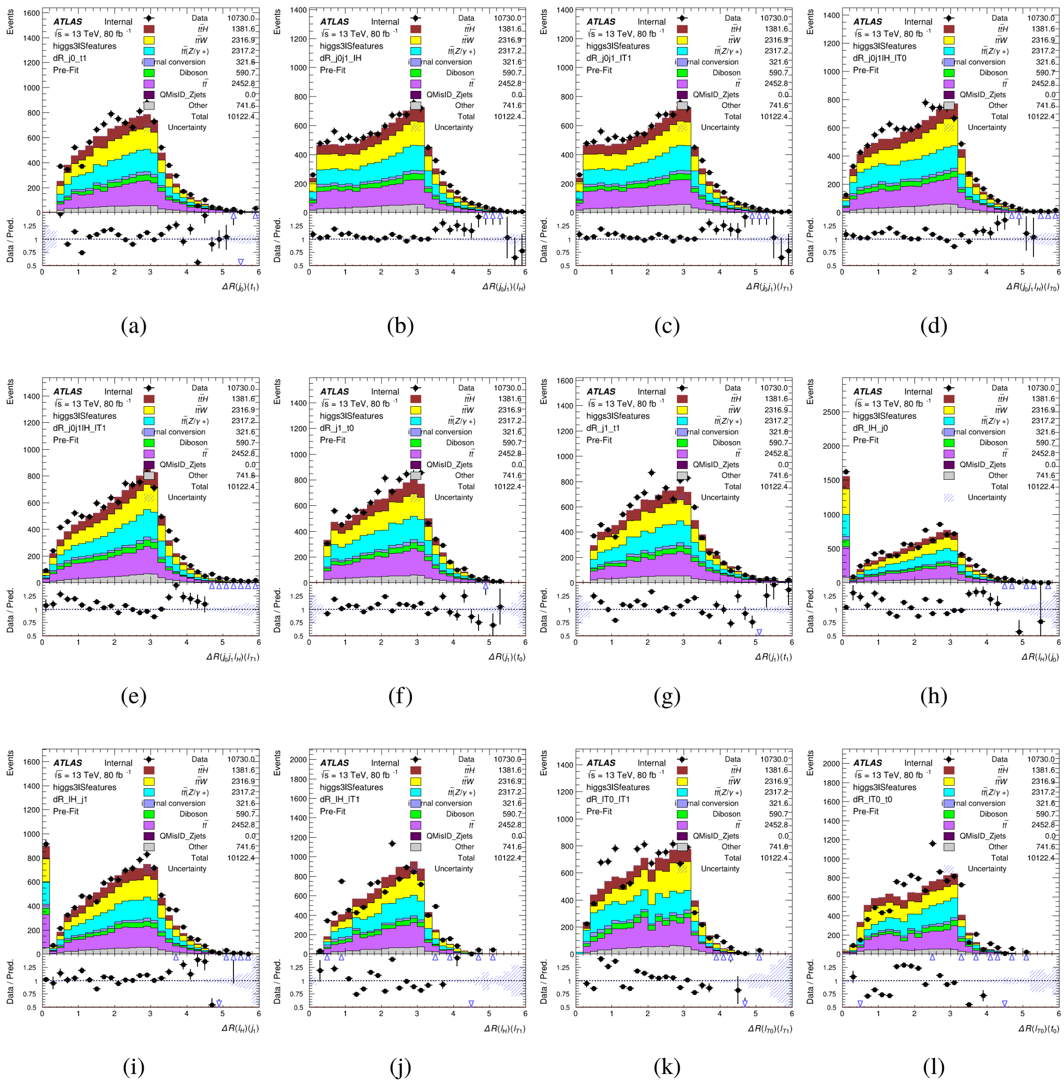


Figure A.12: Input features for higgs3lS

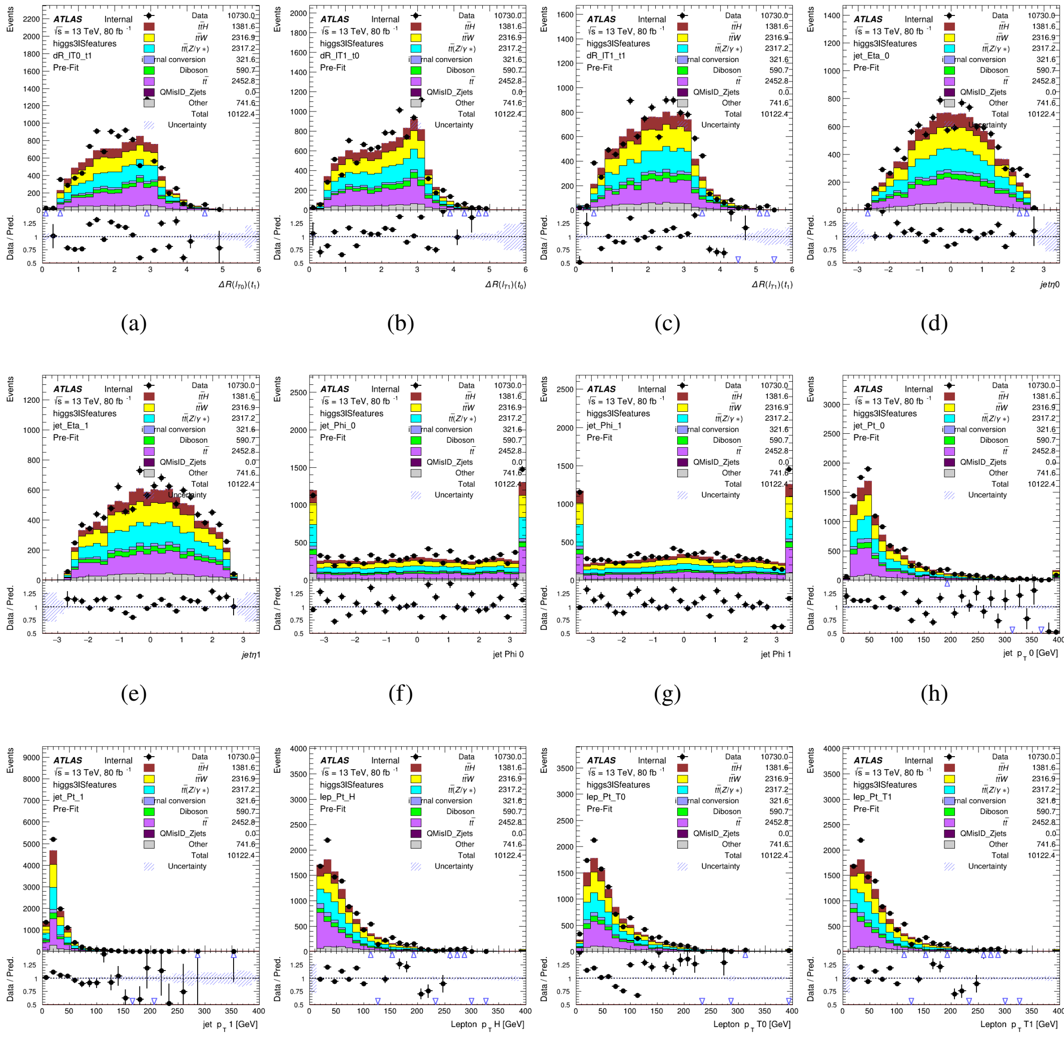


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure A.13: Input features for higgs3lS

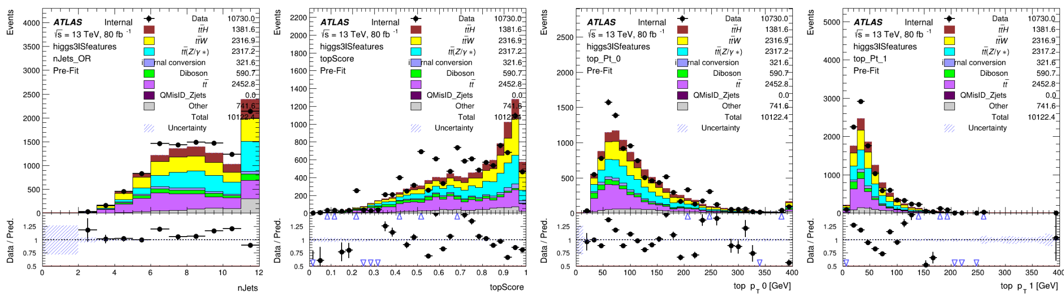


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure A.14: Input features for higgs3lS



(a) (b) (c) (d)

Figure A.15: Input features for higgs3lS

A.1.5 Higgs Reconstruction Features - 3lF

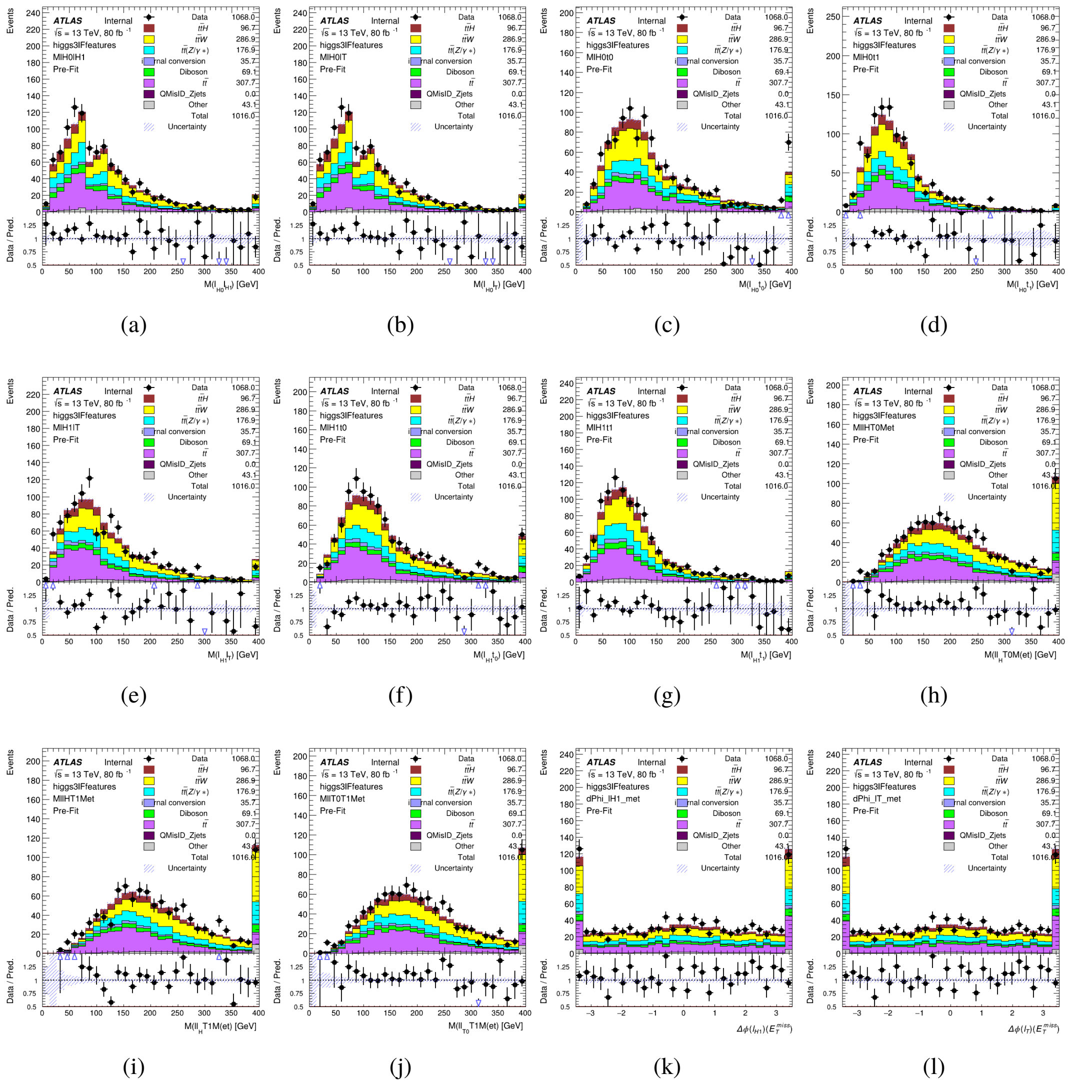


Figure A.16: Input features for higgs3lF

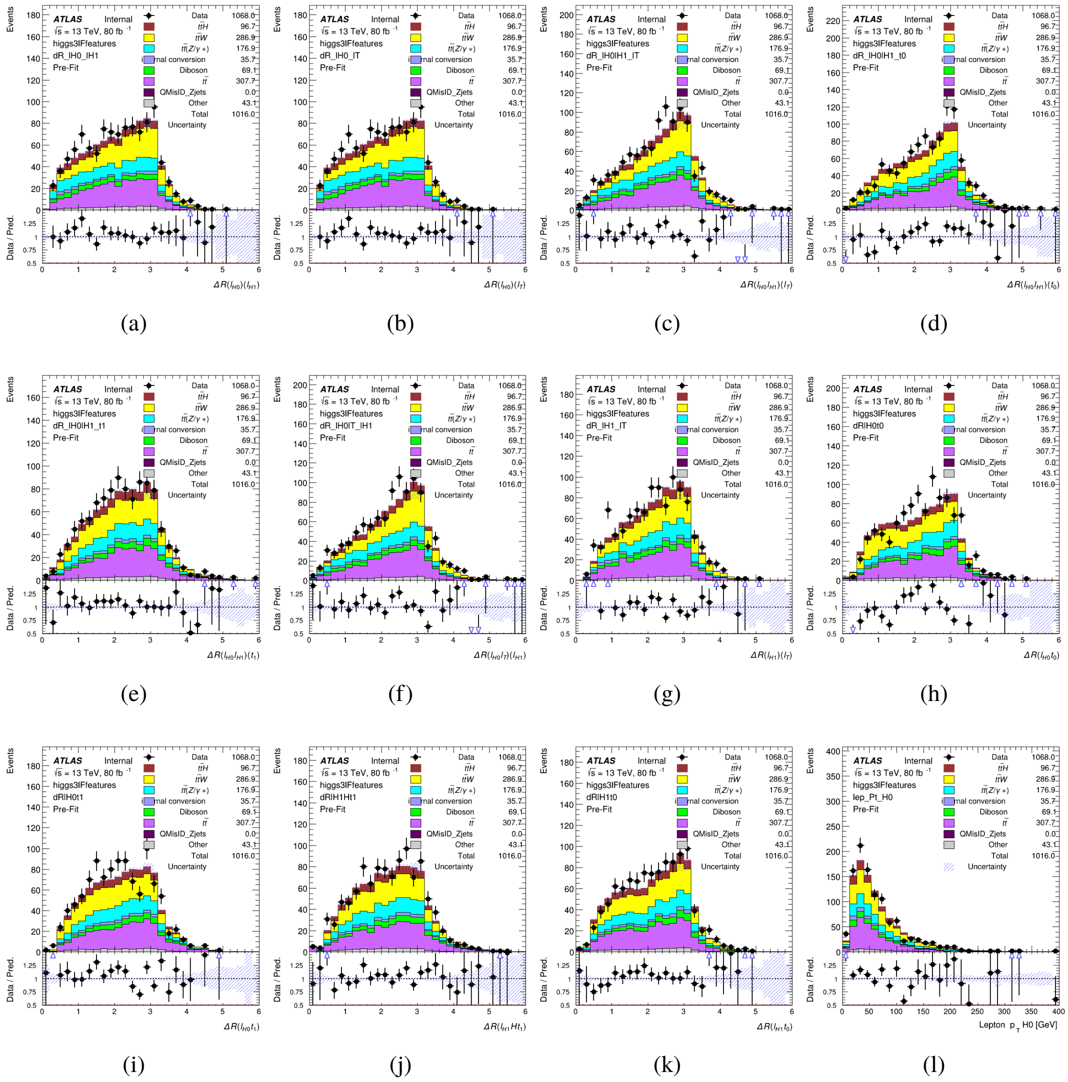
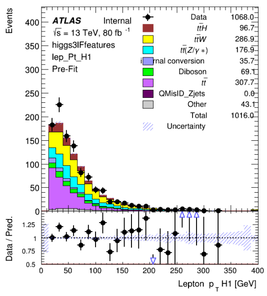
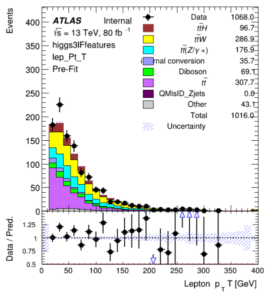


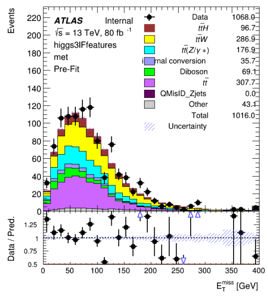
Figure A.17: Input features for higgs3lF



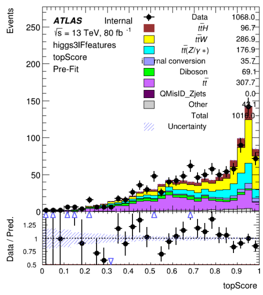
(a)



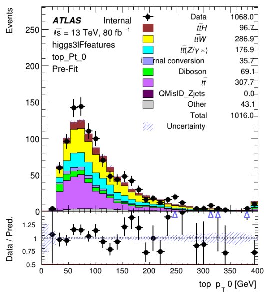
(b)



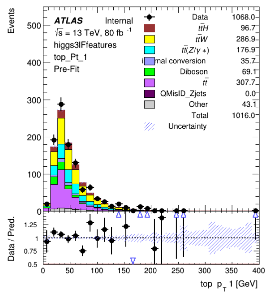
(c)



(d)



(e)



(f)

Figure A.18: Input features for higgs3lF

A.2 Background Rejection MVAs

A.2.1 Background Rejection MVA Features - 2lSS

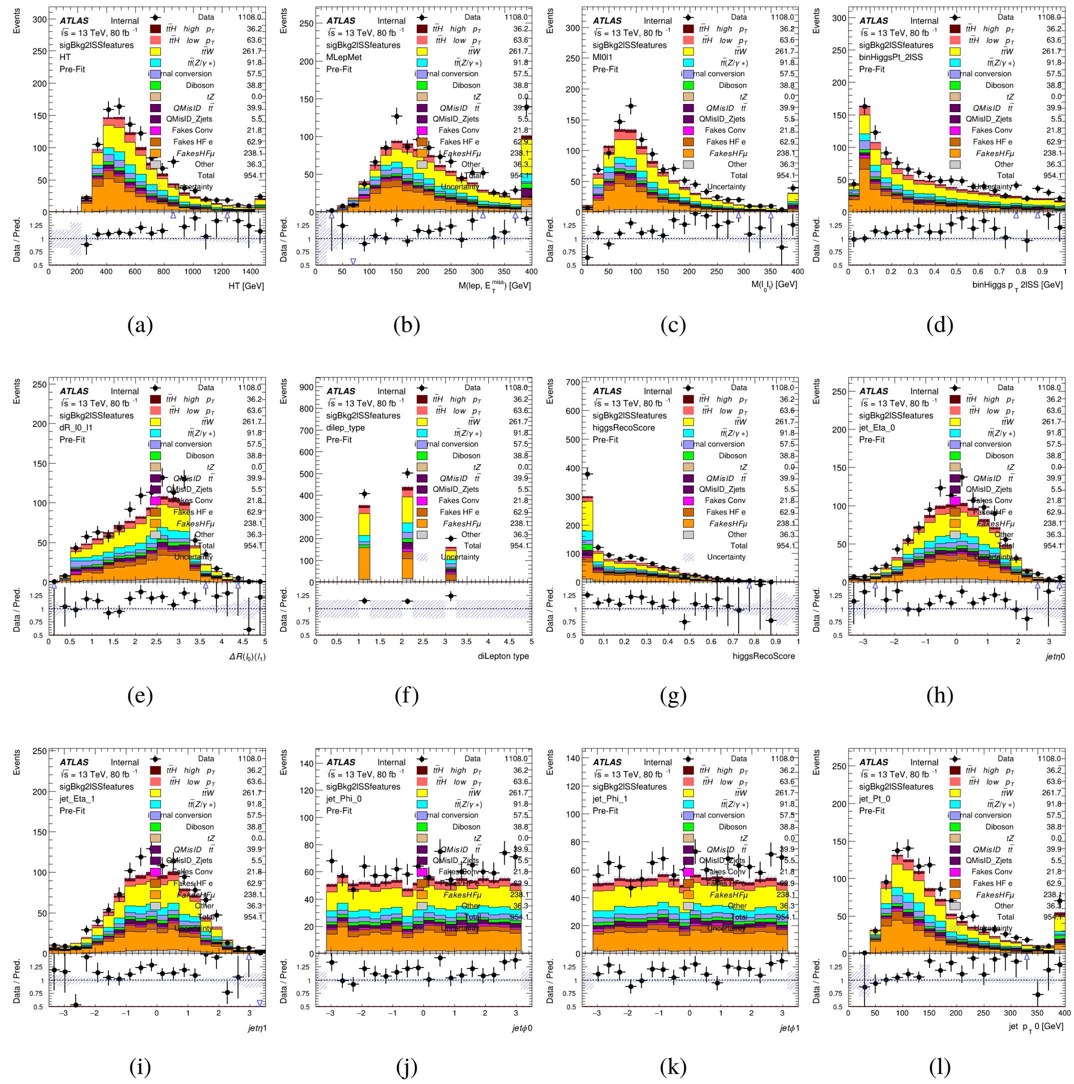
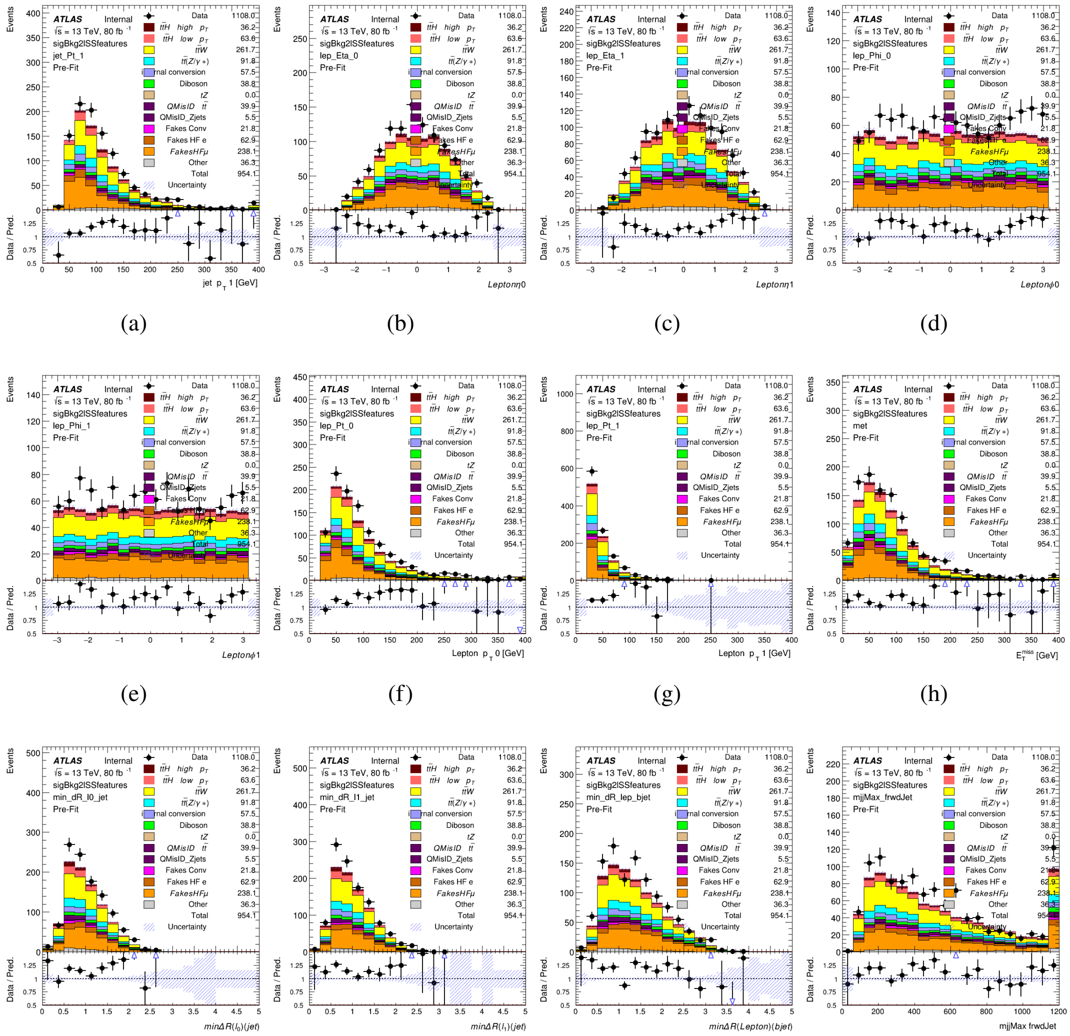


Figure A.19: Input features for sigBkg2lSS

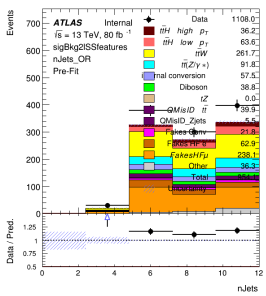


(a) (b) (c) (d)

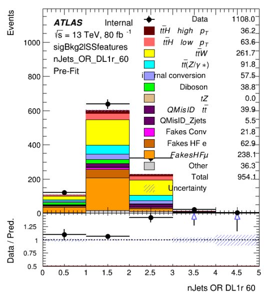
(e) (f) (g) (h)

(i) (j) (k) (l)

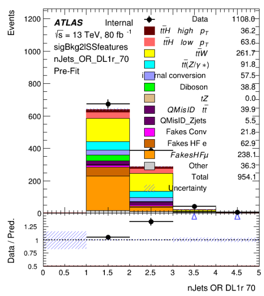
Figure A.20: Input features for sigBkg2lSS



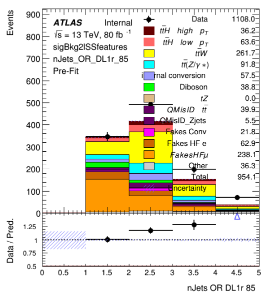
(a)



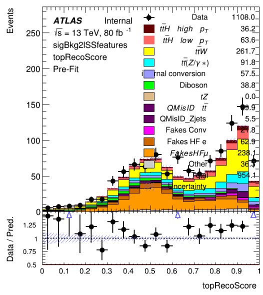
(b)



(c)



(d)



(e)

Figure A.21: Input features for sigBkg2lSS

487 ### A.2.2 Background Rejection MVA Features - 3l

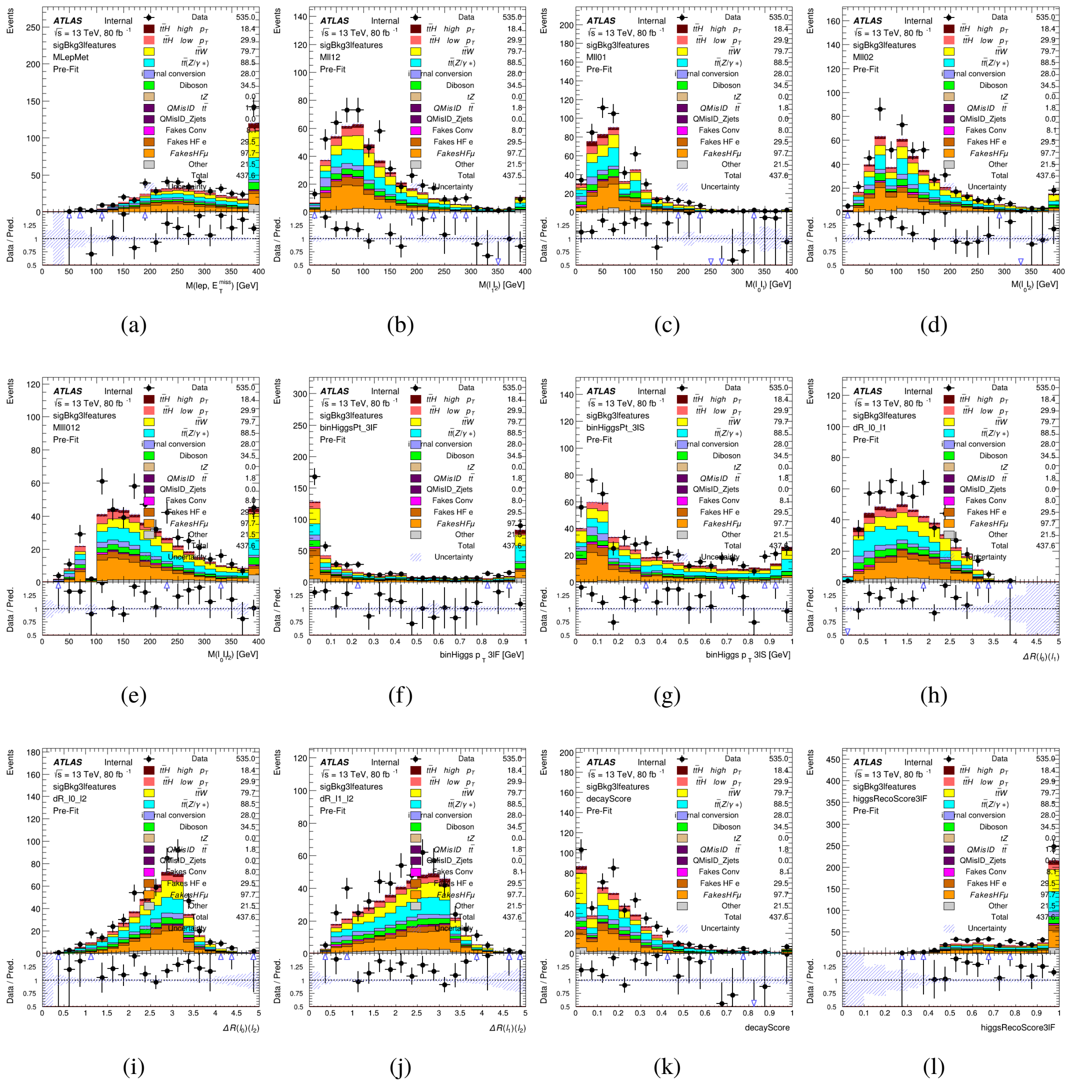


Figure A.22: Input features for sigBkg3l

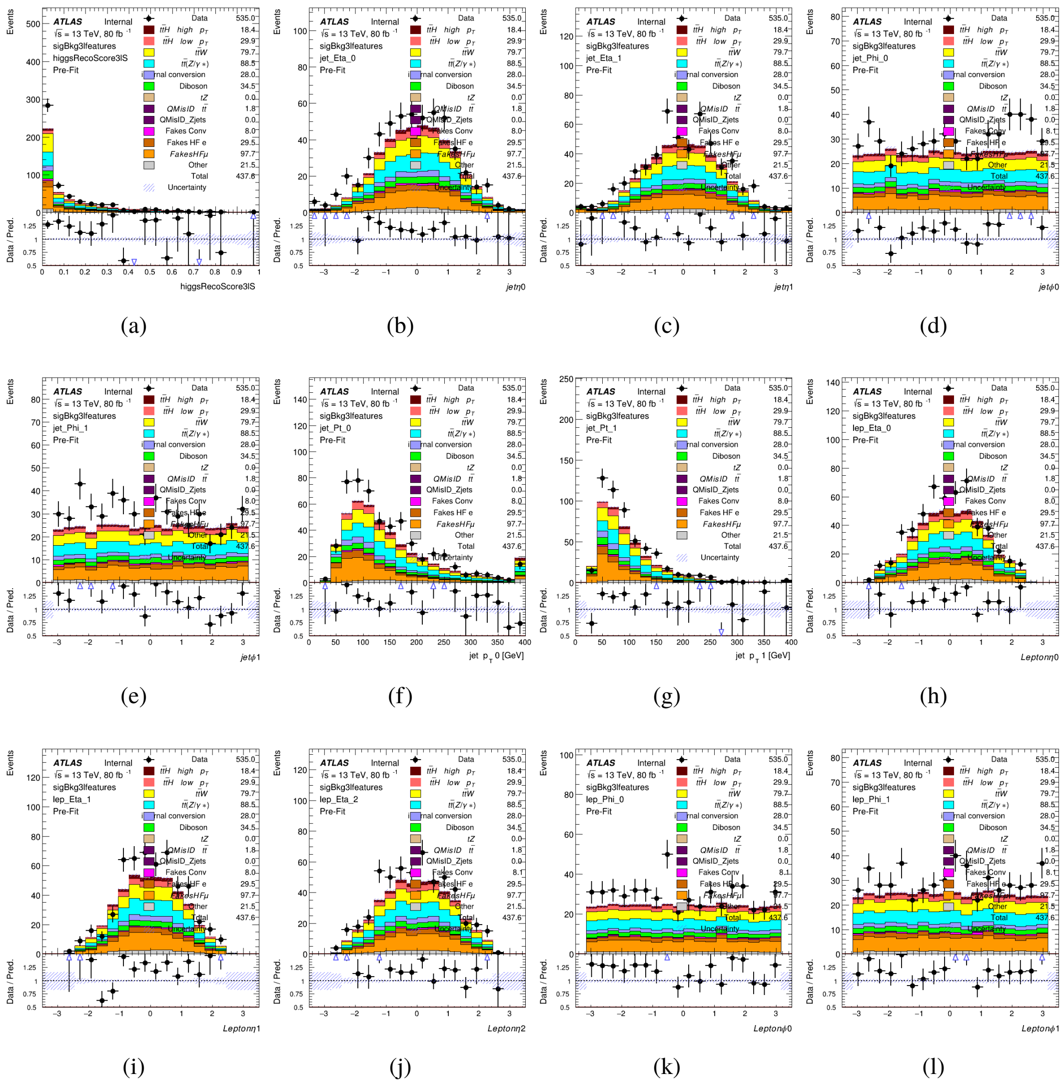
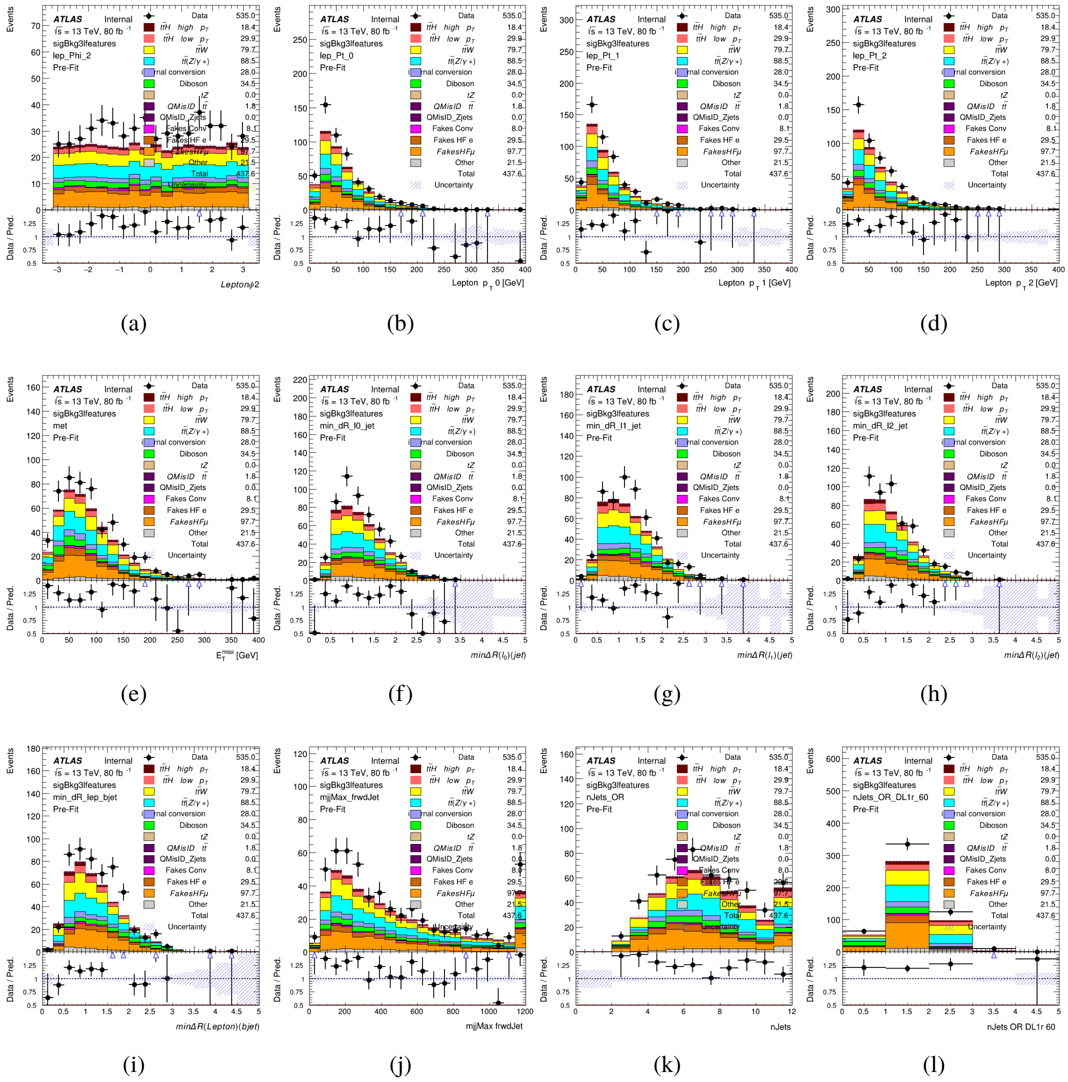


Figure A.23: Input features for sigBkg3l

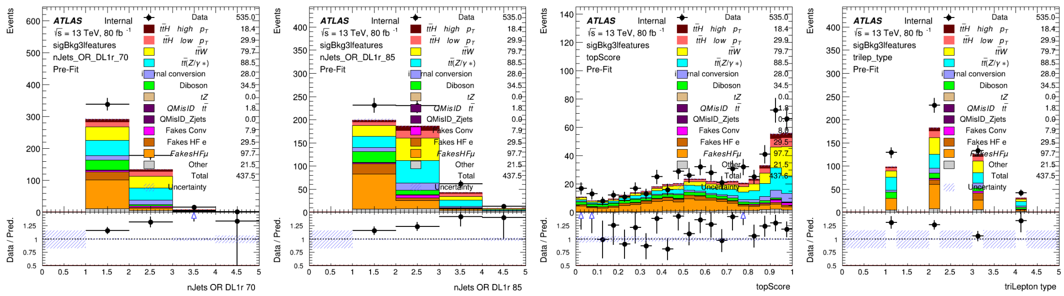


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure A.24: Input features for sigBkg3l



(a) (b) (c) (d)

Figure A.25: Input features for sigBkg3l

A.3 Alternate b-jet Identification Algorithm

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

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

Table 19: Accuracey 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%

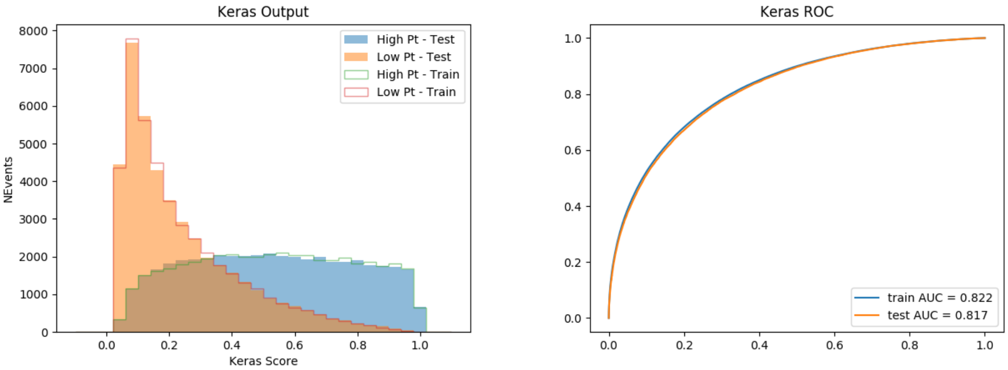
A.4 Binary Classification of the Higgs p_T

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

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

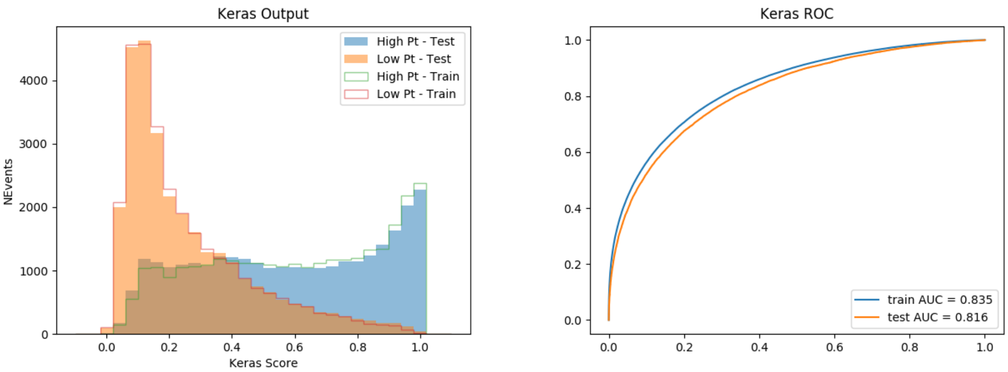
A.5 Impact of Alternative Jet Selection

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.



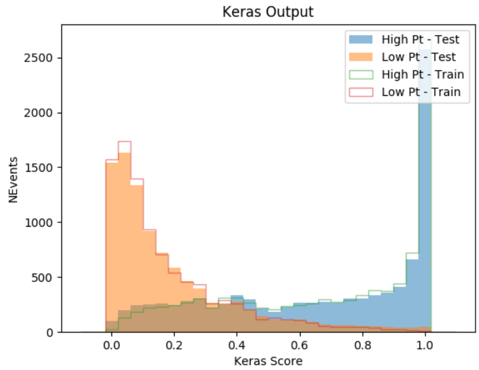
(a) (b)

Figure A.26:

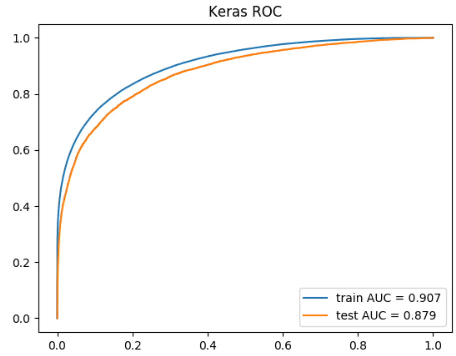


(a) (b)

Figure A.27:

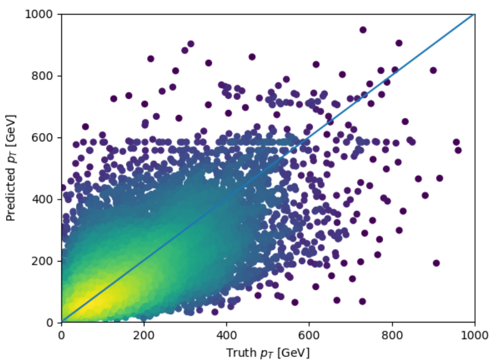


(a)

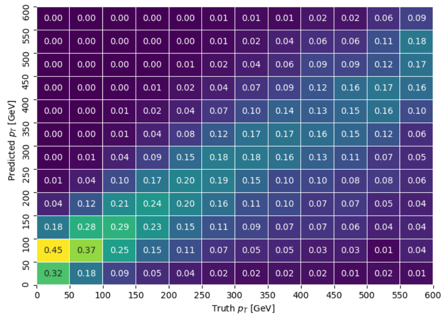


(b)

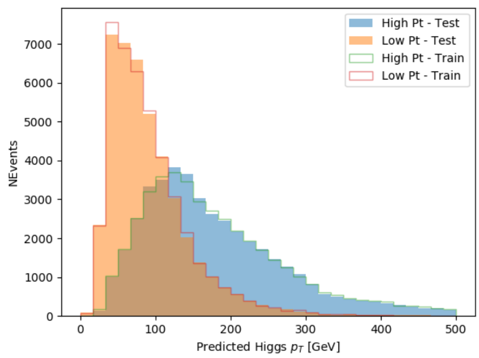
Figure A.28:



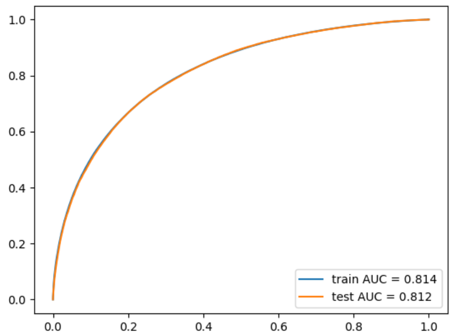
(a)



(b)



(c)



(d)

Figure A.29: