



**ATLAS Note**

**GROUP-2017-XX**

27th October 2020



Draft version 0.1

1

2

3

4

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

5

The ATLAS Collaboration

6

7

8

9

10

11

12

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  $139 \text{ fb}^{-1}$  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.

## Contents

16	<b>1 Changes and outstanding items</b>	<b>3</b>
17	1.1 Changelog	3
18	<b>2 Introduction</b>	<b>4</b>
19	<b>3 The ATLAS Detector</b>	<b>4</b>
20	<b>4 Data and Monte Carlo Samples</b>	<b>5</b>
21	4.1 Data Samples	6
22	4.2 Monte Carlo Samples	6
23	<b>5 Object Reconstruction</b>	<b>7</b>
24	5.1 Trigger Requirements	7
25	5.2 Light Leptons	7
26	5.3 Jets	8
27	5.4 Missing Transverse Energy	9
28	<b>6 Higgs Momentum Reconstruction</b>	<b>9</b>
29	6.1 Decay Candidate Reconstruction	9
30	6.2 b-jet Identification	10
31	6.3 Higgs Reconstruction	14
32	6.4 $p_T$ Prediction	21
33	6.5 3l Decay Mode	26
34	<b>7 Signal Region Definitions</b>	<b>27</b>
35	7.1 Pre-MVA Event Selection	27
36	7.2 Event MVA	27
37	7.3 Signal Region Definitions	27
38	<b>8 Systematic Uncertainties</b>	<b>28</b>
39	<b>9 Results</b>	<b>30</b>
40	<b>10 Conclusion</b>	<b>30</b>
41	<b>Appendices</b>	<b>32</b>
42	<b>A Machine Learning Models</b>	<b>32</b>
43	A.1 Background Rejection MVAs	32

## 44 **1 Changes and outstanding items**

### 45 **1.1 Changelog**

46 This is version 1

## 2 Introduction

Since the discovery of a Higgs boson compatible 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 parameterizing 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 feasibility 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  $\sqrt{s} = 13 \text{ TeV}$ . 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: Section 3 describes the LHC and the ATLAS detector. The dataset and Monte Carlo (MC) simulations used in the analysis is outlined in section 4. Section 5 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 6. The selection and categorisation of events comprises section 7, and the theoretical and experimental systematic uncertainties considered are described in section 8. Finally, the results of the study are summarized in section 9.

## 3 The ATLAS Detector

ATLAS is a general purpose detector designed to maximize the detection efficiency of nearly all physics objects, including leptons, jets, and photons, while covering nearly the entire solid angle around the collision point. Just surrounding the interaction point is the Inner Detector, designed

to track the path of charged particles moving through the detector. An inner solenoid surrounding the Inner Detector is used to produce a magnetic field of 2 T.

The Inner Detector consists of three components - the Pixel Detector, the Semi-Conductor Tracker (SCT), and the Transition Radiation Tracker (TRT). The Pixel Detector is the innermost of these, beginning just 33.25 mm away from the beam line. It consists of three silicon layers along the barrel, as well as three endcap layers, covering a range of  $|\eta| < 2.5$ . The Semiconductor Tracker (SCT) is similar to the Pixel detector, but uses long strips rather than small pixel to cover a larger spatial area.

Situated outside the Inner Detector are two concentric calorimeters, covering a range of  $|\eta| < 4.9$ . The inner calorimeter uses liquid argon (LAr) to measure energy of particles that interact electromagnetically within the region  $|\eta| < 3.2$ , and consists of around 180,000 readout channels. The outer calorimeter, or hadronic calorimeter, is composed of steel plates, with scintillating tiles as the active material. It covers a range of  $|\eta| < 1.7$ , and the signals from the hadronic calorimeter are read out by photomultiplier tubes (PMTs). The remaining pseudorapidity range is covered by forward calorimeter modules.

The outermost layer of the detector, the muon spectrometer, consists of tracking and triggering system. It extends from the outside of the calorimeter system, about a 4.25 m radius from the beam line, to a radius of 11 m. Two large toroidal magnets within the muon system generate a large magnetic ranging between 2 T and 8 T. 1200 tracking chambers are placed in the muon system in order to precisely measure the tracks of muons within  $|\eta| < 2.7$  with high spatial resolution.

A two-level trigger system is used to select out events to be recorded. The level-1 trigger uses hardware information from the calorimeters and muon spectrometer to select events that contain candidates for particles commonly used in analysis, and reduces the rate of events from 40 MHz to around 100 kHz. Events that pass the level-1 trigger move to the High-Level Trigger (HLT). The HLT takes place outside of the detector in software, and looks for properties such as a large amount of missing transverse energy, well defined leptons, and multiple high energy jets. Events that pass the HLT are stored and used for analysis. Because the specifics of the HLT are determined by software rather than hardware, the thresholds can be changed throughout the run of the detector in response to run conditions such as changes to pileup and luminosity. After the HLT is applied, the event rate is reduced to around 1 kHz, which are recorded for analysis.

## 4 Data and Monte Carlo Samples

For both data and Monte Carlo (MC) simulations, samples were prepared in the xAOD format, which was used to produce 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.

## 4.1 Data Samples

The study uses proton-proton collision data collected by the ATLAS detector from 2015 through 2018, which represents an integrated luminosity of  $139 \text{ fb}^{-1}$  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

## 4.2 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 function (PDF).

Process	Event generator	ME order	Parton Shower	PDF
$t\bar{t}H$	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [Ball:2014uwa] (CT10 [ct10])
$t\bar{t}W$	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
$t\bar{t}(Z/\gamma^* \rightarrow ll)$	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
$VV$	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
$t\bar{t}$	POWHEG-BOX v2 [powheggt]	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 (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (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
$s-, t\text{-channel},$ $Wt$ single top	POWHEG-BOX v1 [powhegstp]	NLO	PYTHIA 6	CT10
$q\bar{q}VV, VVV$				
$Z \rightarrow l^+l^-$	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

## 5 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.

### 5.1 Trigger Requirements

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

### 5.2 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.

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 ( $|\Delta z_0 \sin \theta_\ell|$ ), as shown in table ??.

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].

### 5.3 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].



## 5.4 Missing Transverse Energy

Because all  $t\bar{t}H$  — ML channels considered include multiple neutrinos, missing transverse energy ( $E_T^{\text{miss}}$ ) 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].

## 6 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 6.1. 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 6.2. As described in section 6.3, 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 (6.4). MVAs are also used in the analysis to determine the decay of the Higgs boson in the 3l channel (6.5).

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  $80 \text{ fb}^{-1}$  of data, as shown in the sections below.

### 6.1 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 5.2. 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 5.3. Specifically, jets from the Higgs decay tend to be soft, with 37% having  $p_T < 25$  GeV. Therefore jets with  $p_T < 15$  GeV are considered as possible candidates in the models described below. The spectrum is found to be well modeled even down to this low  $p_T$  threshold, as shown in section 7.1. As they are expected to originate from the primary vertex, jets are also required to pass a JVT cut.

## 6.2 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 accuracy 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.

### 6.2.1 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, 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.

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

jet $p_T$ 0	jet $p_T$ 1	Lepton $p_T$ 0
Lepton $p_T$ 1	jet $\eta$ 0	jet $\eta$ 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	$E_T^{\text{miss}}$	

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

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 correct (i.e. both jets are truth b-jets), and incorrect combinations are shown in figure 6.1. The correct and incorrect contributions are scaled to the same integral, so as to better demonstrate the differences in the distributions.

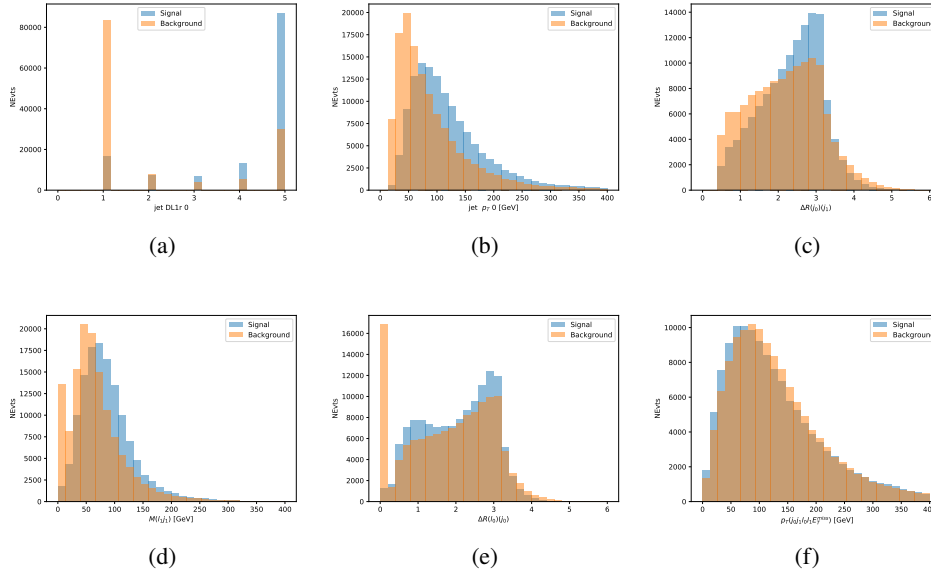


Figure 6.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

The modeling of these inputs is validated against data, with figure 6.2 showing good general

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

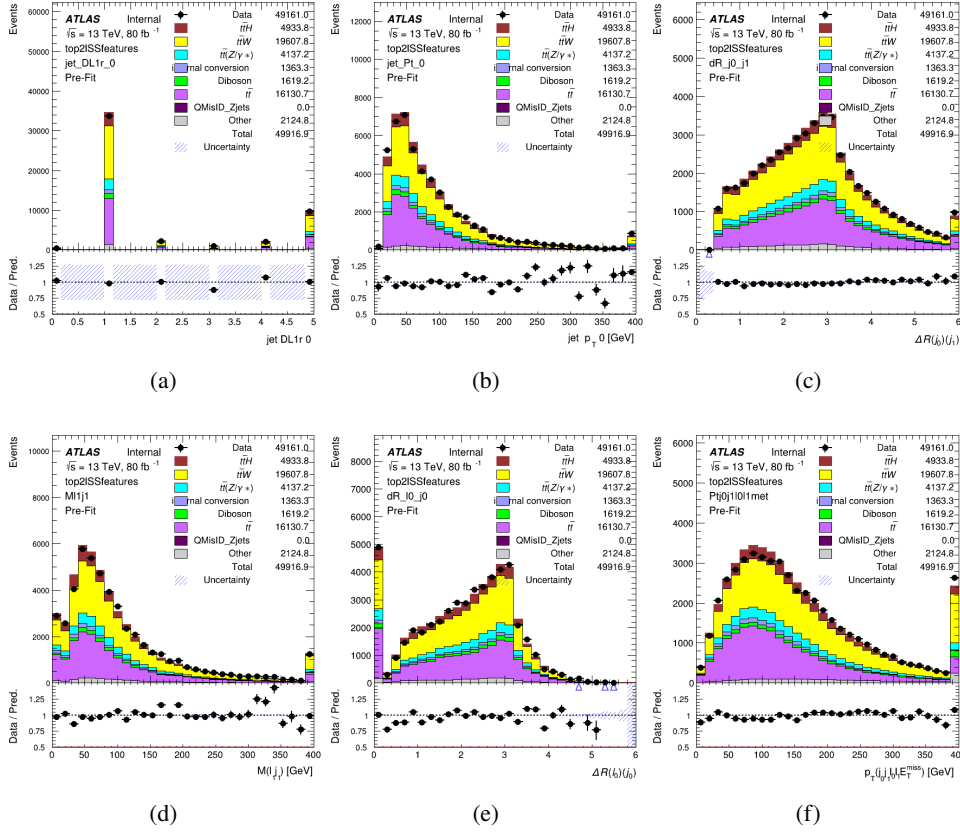


Figure 6.2: Data/MC comparisons of input features for top2ISS training for 80 fb<sup>-1</sup> of data.

246 Based on the results of grid search evaluation, the optimal architecture is found to include 5  
 247 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as overfitting  
 248 is found to not be an issue. The output score distribution as well as the ROC curve for the trained  
 249 model are shown in figure 6.2.1. The model is found to identify the correct pairing of jets for  
 250 73% of 2ISS signal events on test data.

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

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

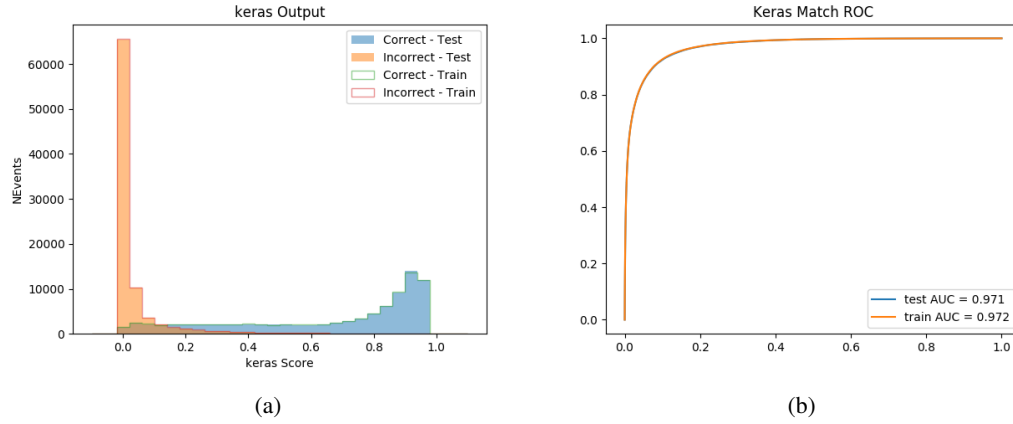


Figure 6.3: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

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

## 6.2.2 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 6.4, comparing the distributions for correct and incorrect combinations of jets.

The modeling of these inputs is validated against data, with figure 6.5 showing good general agreement between data and MC. Plots for the complete list of features can be found in section A.

Again, the dataset is downsized to reduce the ratio of correct and incorrect combination from 20:1, 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 6.2.2.

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.

jet $p_T$ 0	jet $p_T$ 1	jet $\eta$ 0
jet $\eta$ 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	$E_T^{\text{miss}}$	nJets OR DL1r 85
nJets OR DL1r 60		

Table 5: Input features

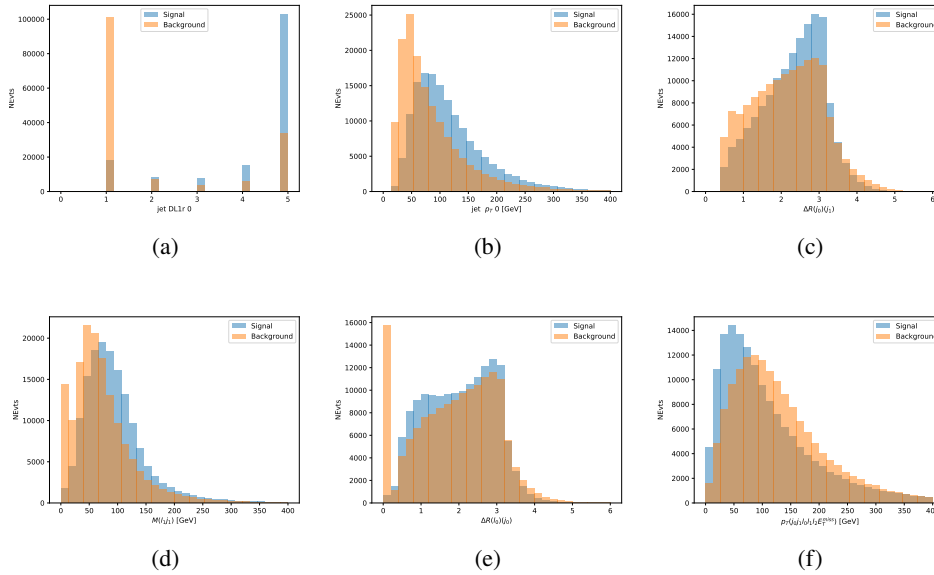


Figure 6.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

### 6.3 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.

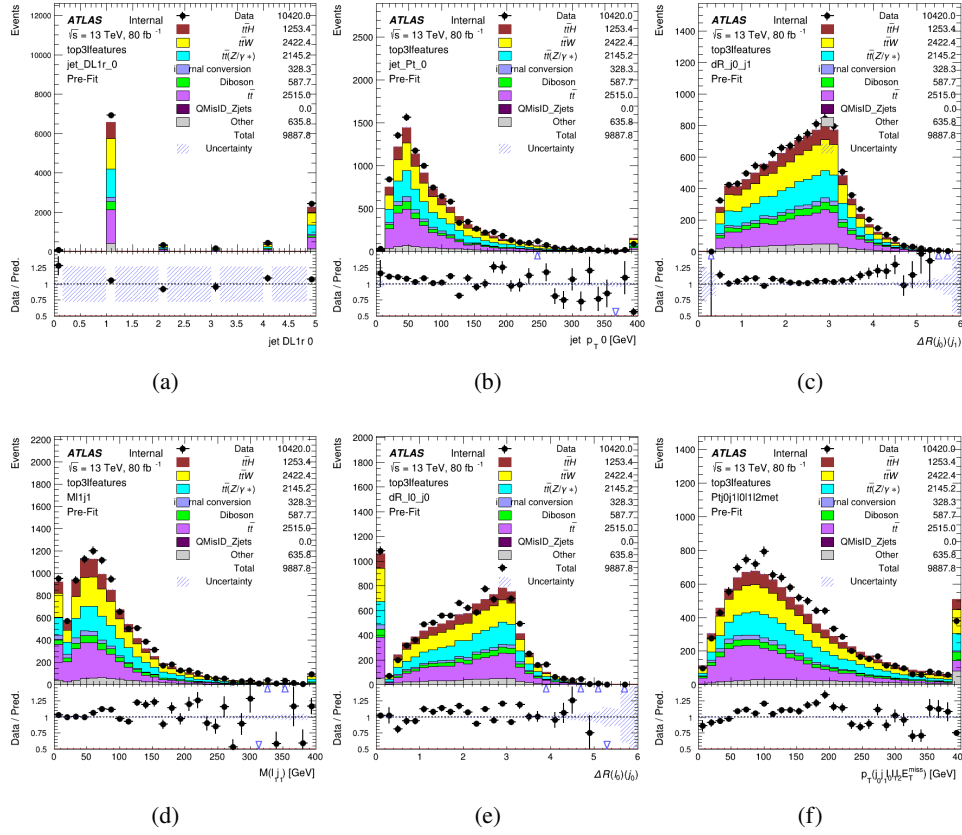


Figure 6.5: Data/MC comparisons of input features for top31 training for  $80 \text{ fb}^{-1}$  of data.

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%
$\geq 3$ b-jets	55.7%	52.3%
Overall	79.8%	70.2%

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 leptons. For the fully-leptonic case, the decay products include 2 of the three leptons associated with the event. For training the

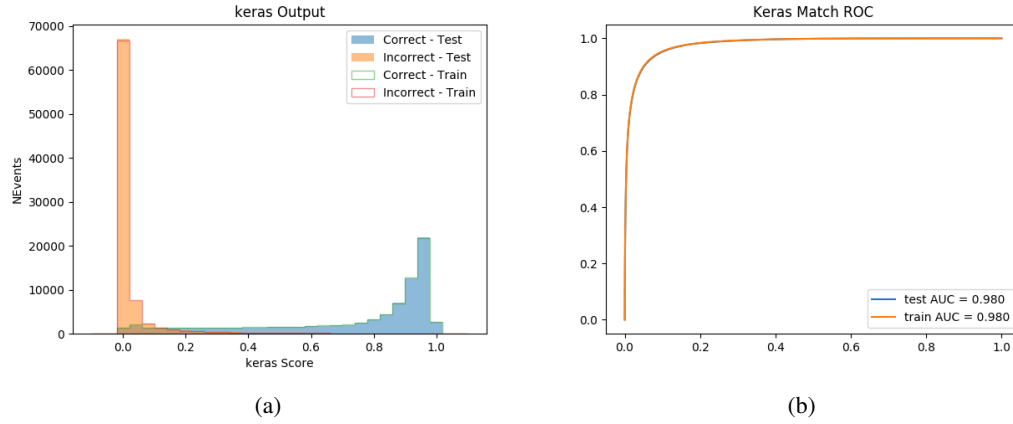


Figure 6.6: tmp

models, events are separated into these two categories using truth level information. A separate MVA, described in section 6.5, is used to make this distinction at reco level and determine which model to use.

For all channels, the models described in section 6.2 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.

### 6.3.1 2ISS Channel

For the 2ISS channel, the Higgs decay products include one light lepton and two jets. The neural network is trained on the kinematics of different combinations of leptons and jets, as well as the b-jets identified in section 6.2, with the specific input features listed in table ??.



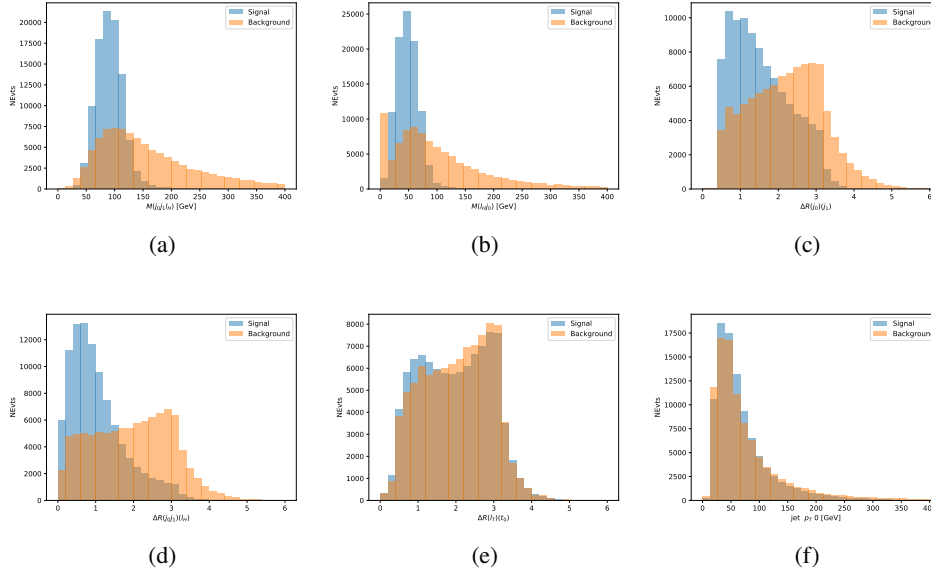


Figure 6.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

293 The modeling of these inputs is validated against data, with figure 6.2 showing good general  
 294 agreement between data and MC. Plots for the complete list of features can found in section A.

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

### 298 6.3.2 3l Semi-leptonic Channel

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

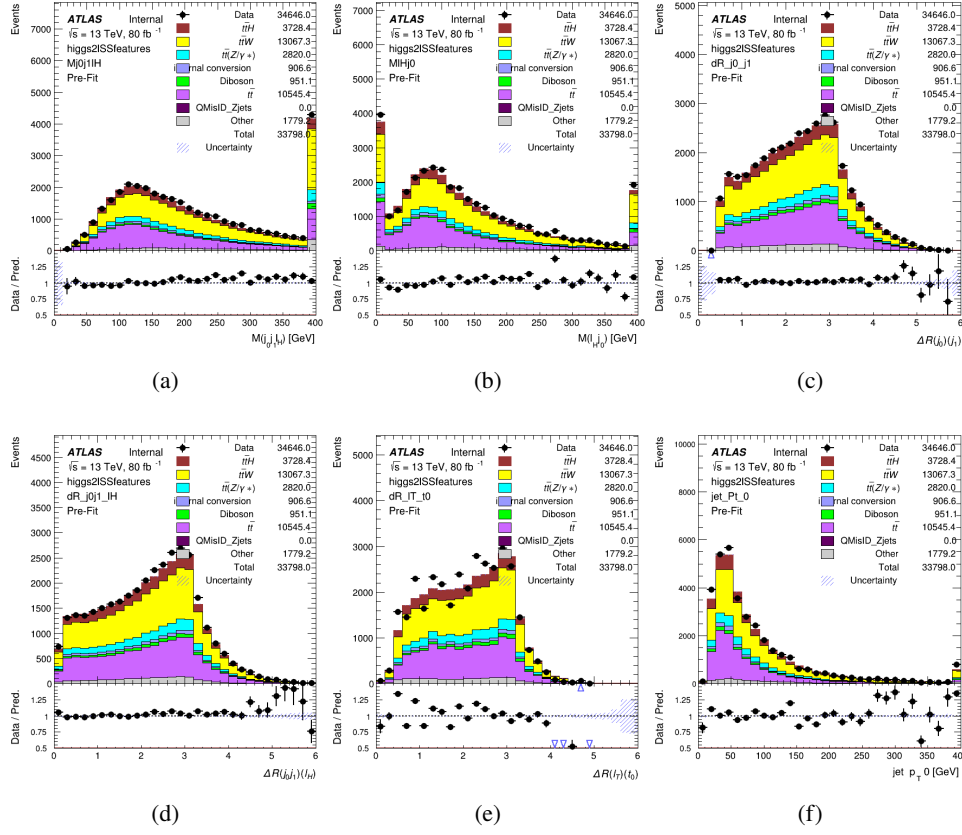


Figure 6.8: Data/MC comparisons of input features for higgs2lSS training for  $80 \text{ fb}^{-1}$  of data.

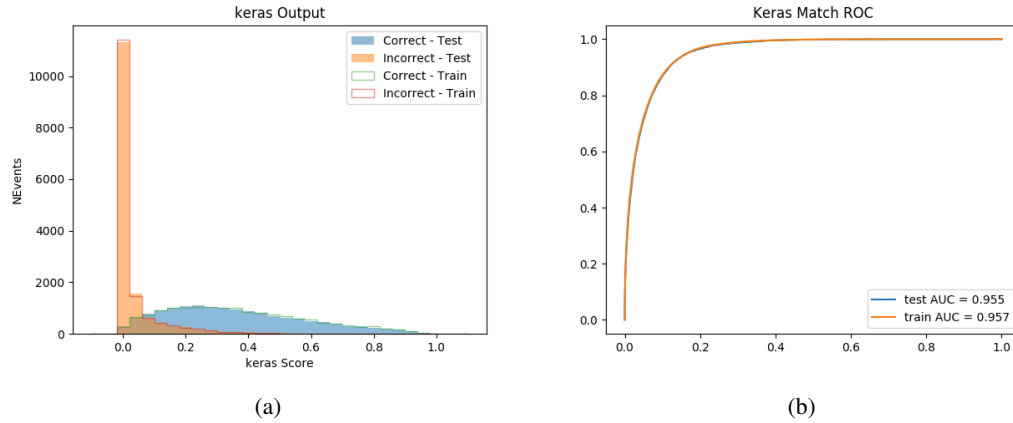


Figure 6.9: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

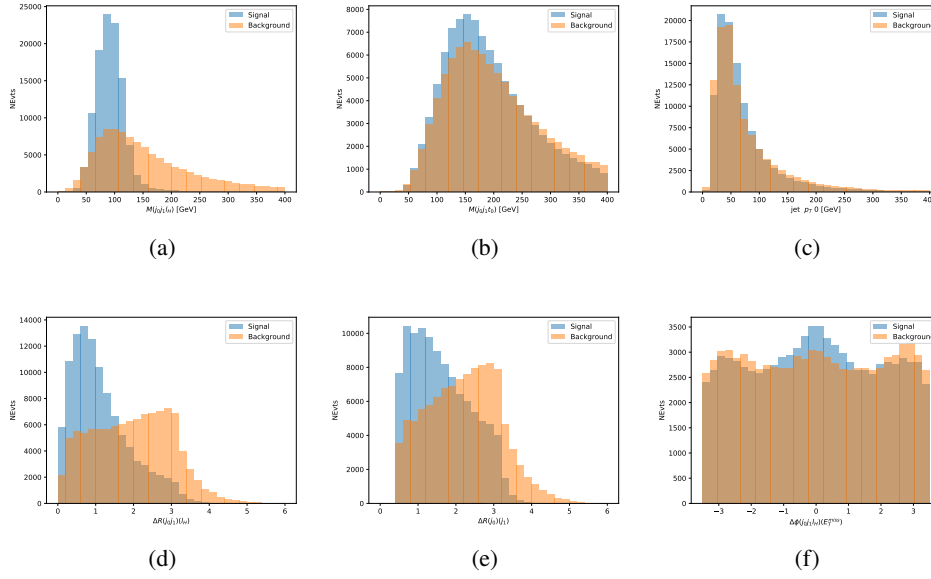


Figure 6.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

303 The modeling of these inputs is validated against data, with figure 6.11 showing good general  
 304 agreement between data and MC. Plots for the complete list of features can found in section A.

### 305 6.3.3 3l Fully-leptonic Channel

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

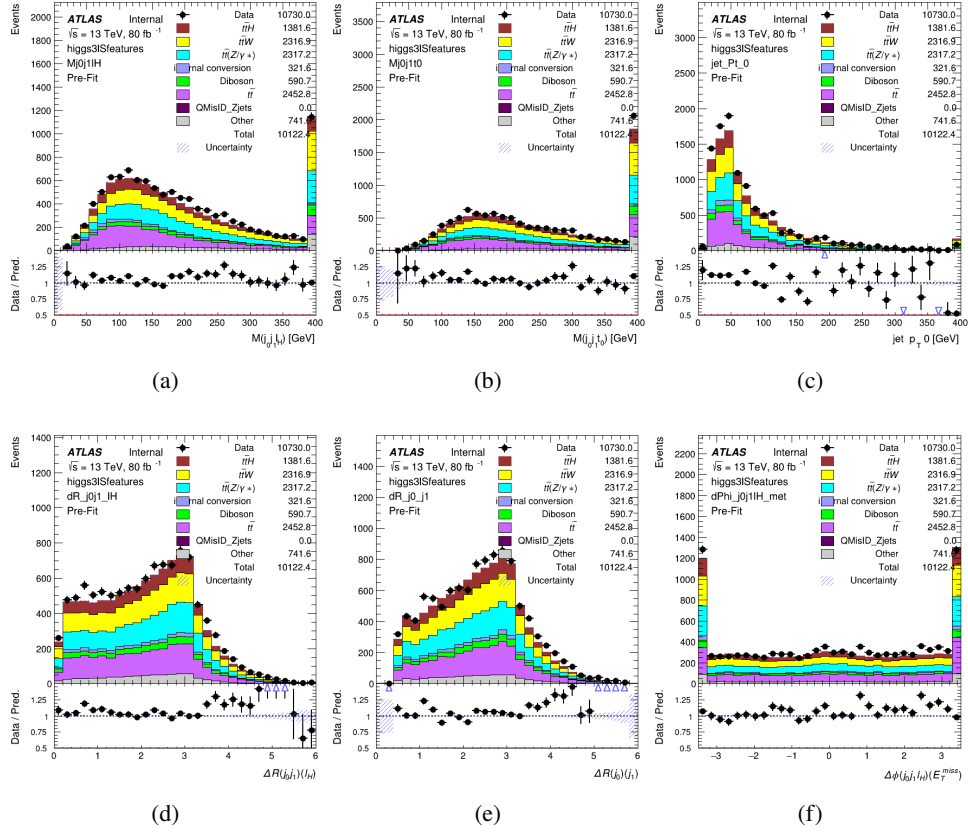


Figure 6.11: Data/MC comparisons of input features for higgs3lS training for  $80 \text{ fb}^{-1}$  of data.

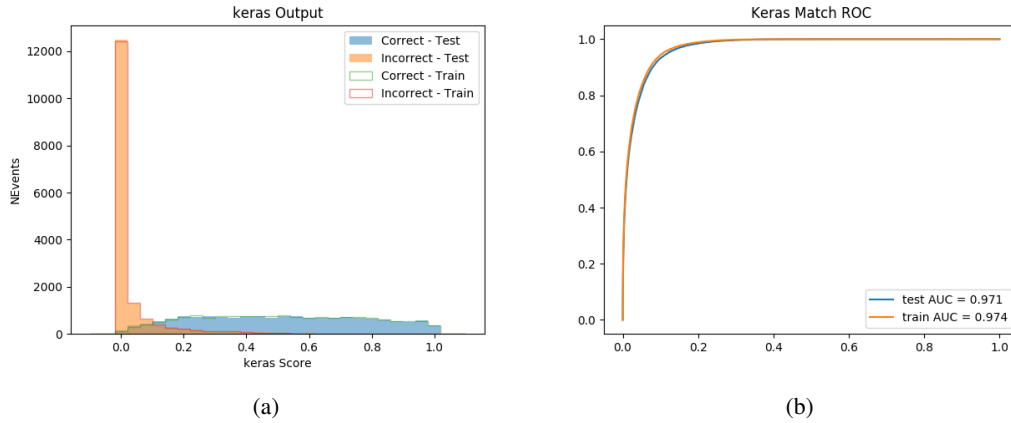


Figure 6.12: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

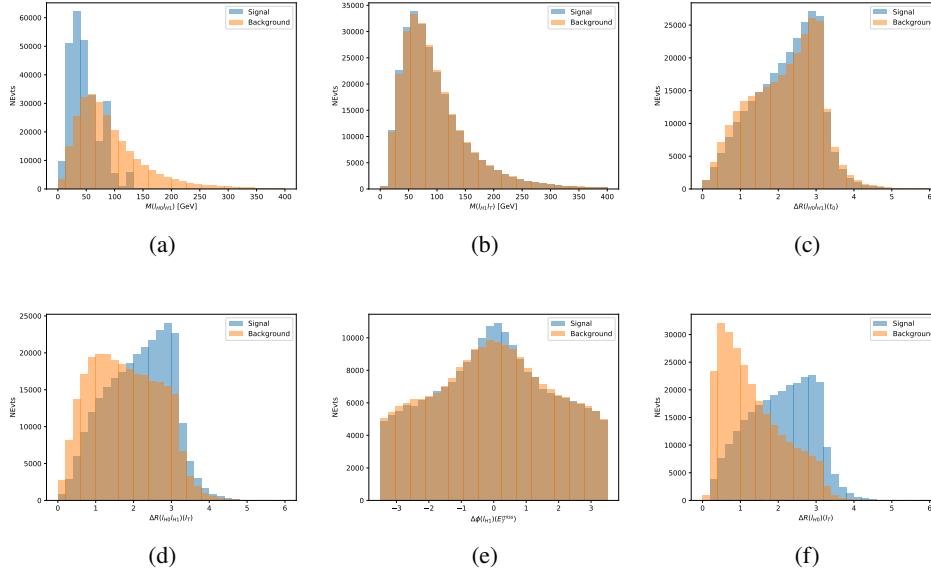


Figure 6.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

311 The modeling of these inputs is validated against data, with figure 6.14 showing good general  
 312 agreement between data and MC. Plots for the complete list of features can found in section A.

## 313 6.4 $p_T$ Prediction

314 Once the most probable decay products have been identified, their kinematics are used as inputs  
 315 to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,  
 316 a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay  
 317 are included as well, as these are found to improve performance. The truth  $p_T$  of the Higgs, as  
 318 predicted by MC, are used as labels.

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

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

### 323 6.4.1 2lSS Channel

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

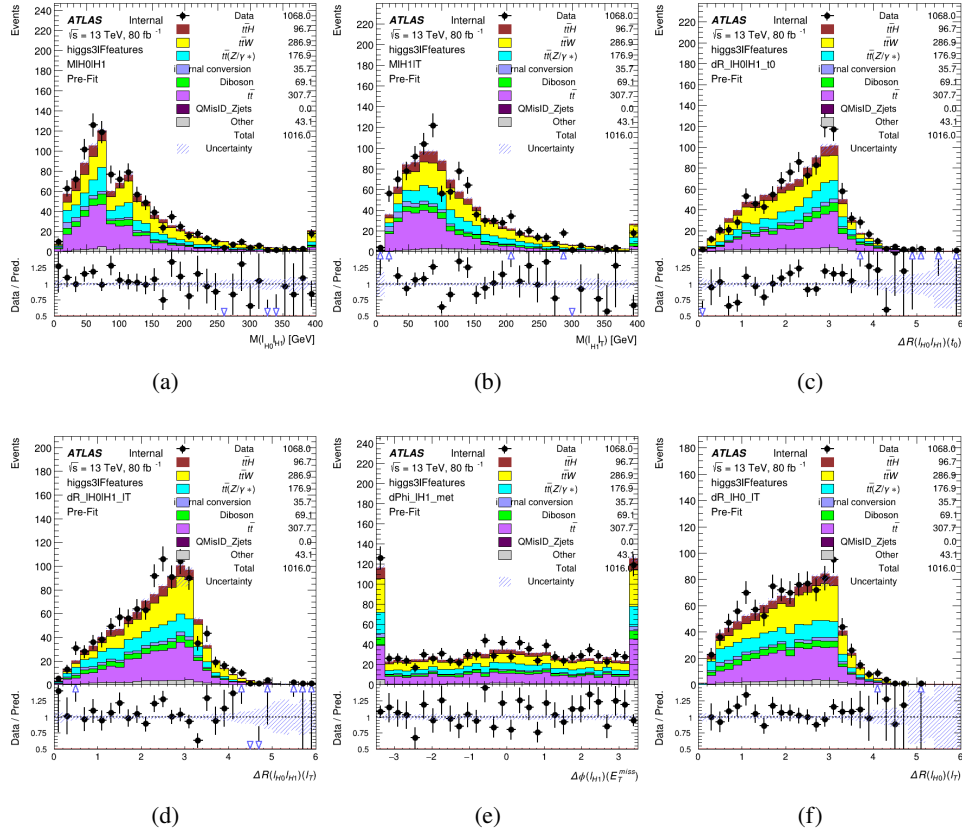


Figure 6.14: Data/MC comparisons of input features for higgs3lF training for 80 fb<sup>-1</sup> of data.

326 from the Higgs is labeled  $l_{\text{H}}$ , while the other lepton is labeled  $l_{\text{T}}$ , as it most have come from the  
 327 decay of one of the top quarks. The Higgs Reco Score and b-jet Reco Score are the outputs of the  
 328 Higgs reconstruction algorithm, and the b-jet identification algorithm, respectively.

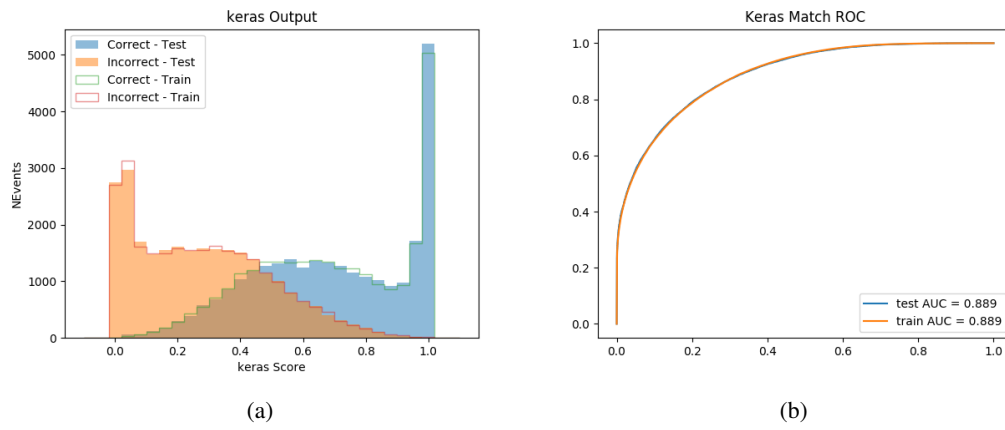


Figure 6.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

329 The optimal neural network architecture for this channel is found to consist of 5 hidden layers with  
 330 40 nodes each. The input data set includes 1.2 million events, 10% of which is used for testing,  
 331 the other 90% for training. Training is found to converge after around 150 epochs.

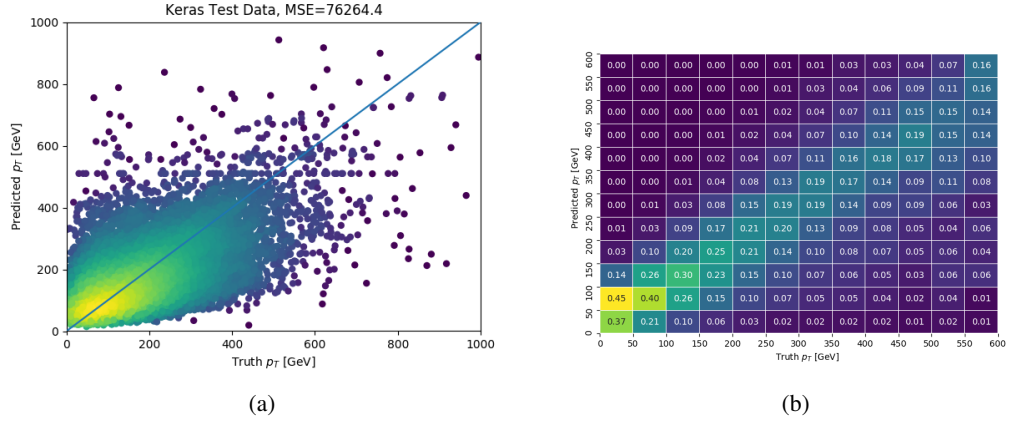


Figure 6.16:

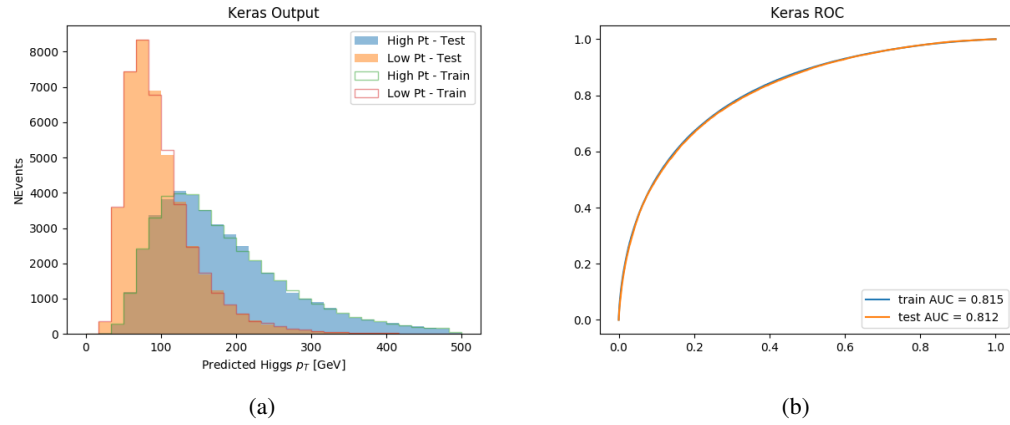


Figure 6.17:

## 332 6.4.2 3l Semi-leptonic Channel



333 The optimal neural network architecture for this channel is found to consist of 5 hidden layers with  
 334 40 nodes each. The input data set includes 1.2 million events, 10% of which is used for testing,  
 335 the other 90% for training. Training is found to converge after around 150 epochs.

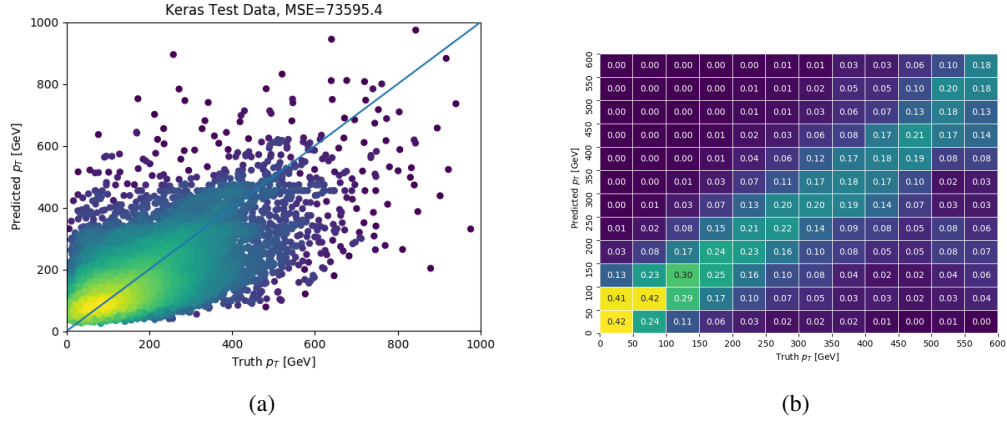


Figure 6.18:

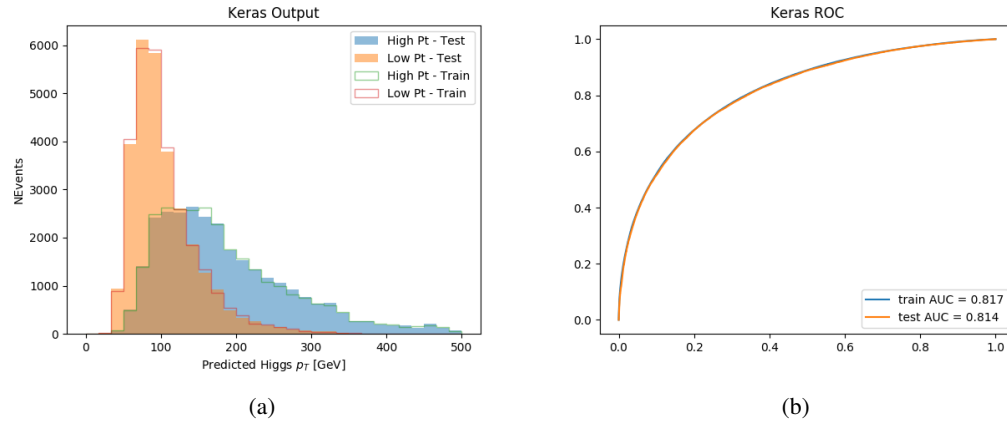


Figure 6.19:

### 336 6.4.3 3l Fully-leptonic Channel

337 The optimal neural network architecture for this channel is found to consist of 5 hidden layers with  
 338 40 nodes each. The input data set includes 1.2 million events, 10% of which is used for testing,  
 339 the other 90% for training. Training is found to converge after around 150 epochs.

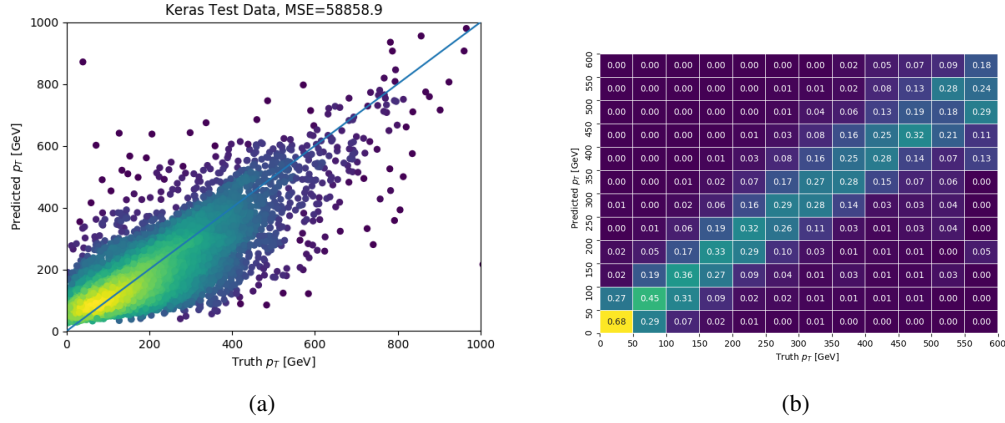


Figure 6.20:

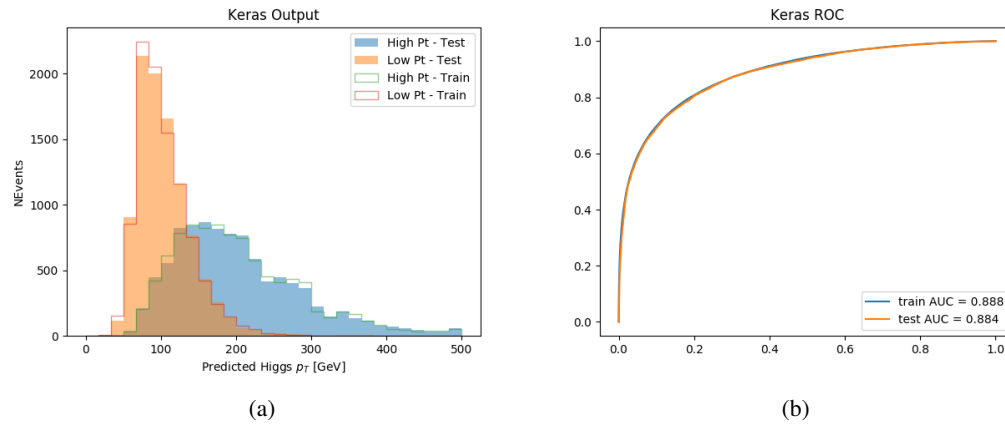


Figure 6.21:

## 340 6.5 3l Decay Mode

341 In the 3l channel, there are two possible ways for the Higgs to decay, both involving intermediate  
 342 W boson pairs: Either both W bosons decay leptonically, in which case the reconstructed decay  
 343 consists of two leptons (referred as the fully-leptonic 3l channel), or one W decays leptonically  
 344 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.

## 7 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

### 7.1 Pre-MVA Event Selection

A preselection is applied to define orthogonal analysis channels based on the number of leptons in each event.

#### 7.1.1 2lSS Channel

#### 7.1.2 3l Channel

### 7.2 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, and 3l fully leptonic. In particular, Neural Networks produced with Tensorflow are trained using the kinematics of signal and background events derived from Monte Carlo simulations. Further, 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.

Output distributions of each MVA are shown in figure 7.2. Detailed explanations of each of the models can be found in section A.

### 7.3 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 6.5 is used to separate the three channel into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson.

For each event, depending on the channel as well as the predicted  $p_T$  of the Higgs derived from the algorithm described in section 6.4, a cut on the appropriate background rejection algorithm is



and background estimations. The numerical impact of each of these uncertainties is outlined in section 9.

Table 7: Sources of systematic uncertainty considered in the analysis. Some of the systematic uncertainties are split into several components, as indicated by the number in the rightmost column.

Systematic uncertainty	Components
Luminosity	1
Pileup reweighting	1
<b>Physics Objects</b>	
Electron	6
Muon	15
Jet energy scale and resolution	28
Jet vertex fraction	1
Jet flavor tagging	131
$E_T^{\text{miss}}$	3
Total (Experimental)	186
<b>Background Modeling</b>	
Cross section	24
Renormalization and factorization scales	10
Parton shower and hadronization model	2
Shower tune	4
Total (Signal and background modeling)	40
<b>Background Modeling</b>	
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

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 b-tagging of jets, and to the reconstruction of  $E_T^{\text{miss}}$ . 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.

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.

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

## 395 **9 Results**

396 A maximum likelihood fit is performed simultaneously over the regions described in section  
397 ??.

## 398 **10 Conclusion**

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

403 **List of contributions**

404

## Appendices

### A Machine Learning Models

#### A.1 Background Rejection MVAs

Separate models are used in order to distinguish signal events from background for each analysis channel - 2ISS, 3l semi-leptonic, and 3l fully leptonic. In particular, Neural Networks produced with Tensorflow are trained using the kinematics of signal and background events derived from Monte Carlo simulations. Further, 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.

##### A.1.1 2ISS - High $p_T$

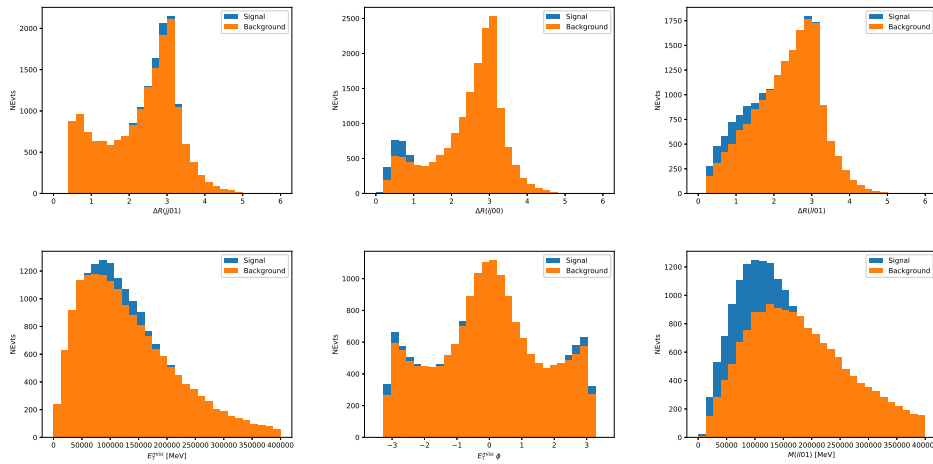


Figure A.1:



415 **A.1.2 2lSS - Low  $p_T$**

416 **A.1.3 3l Semi-Leptonic - High  $p_T$**

417 **A.1.4 3l Semi-Leptonic - Low  $p_T$**

418 **A.1.5 3l Fully Leptonic - High  $p_T$**

419 **A.1.6 3l Fully Leptonic - Low  $p_T$**