



ATLAS Note

GROUP-2017-XX

4th May 2021



1

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

2

The ATLAS Collaboration

3

4

5

6

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

7

8

9

10

11

12

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

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

15 **Contents**

16	1 Changes and outstanding items	4
17	1.1 Changelog	4
18	2 Introduction	5
19	3 Data and Monte Carlo Samples	6
20	3.1 Data Samples	6
21	3.2 Monte Carlo Samples	6
22	4 Object Reconstruction	8
23	4.1 Trigger Requirements	8
24	4.2 Light Leptons	9
25	4.3 Jets	10
26	4.4 B-tagged Jets	10
27	4.5 Missing Transverse Energy	10
28	4.6 Overlap removal	10
29	5 Higgs Momentum Reconstruction	11
30	5.1 Physics Object Truth Matching	12
31	5.2 Truth Level Studies	12
32	5.3 b-jet Identification	13
33	5.4 Higgs Reconstruction	19
34	5.5 p_T Prediction	32
35	5.6 3l Decay Mode	42
36	6 Signal Region Definitions	44
37	6.1 Pre-MVA Event Selection	44
38	6.2 Event MVA	49
39	6.3 Signal Region Definitions	56
40	7 Systematic Uncertainties	59
41	8 Results	64
42	8.1 Results - 79.8 fb^{-1}	65
43	8.2 Projected Results - 139 fb^{-1}	69
44	9 Conclusion	73
45	Appendices	75
46	.1 Non-prompt lepton MVA	75

47	A Machine Learning Models	77
48	A.1 Higgs Reconstruction Model Details	77
49	A.2 Background Rejection MVA Details	100
50	A.3 Truth Level Studies	109
51	A.4 Alternate b-jet Identification Algorithm	110
52	A.5 Binary Classification of the Higgs p_T	111
53	A.6 Impact of Alternative Jet Selection	112

⁵⁴ **1 Changes and outstanding items**

⁵⁵ **1.1 Changelog**

⁵⁶ This is version 1

57 2 Introduction

58 Since the discovery of a Higgs boson compatible with the Standard Model (SM) in 2012
 59 [**HION-2010-02**], its interactions with other particles have been studied using proton-proton
 60 collision data produced by the Large Hadron Collider (LHC). The strongest of these interactions
 61 is the coupling of the Higgs to the top quark, making the Yukawa coupling between these two
 62 particles of particular interest for study.

63 These interactions can be measured directly by studying the production of a Higgs Boson in
 64 association with a pair of Top Quarks ($t\bar{t}H$). While this process has been observed by both
 65 the ATLAS and CMS collaborations, these analyses have focused on measuring the overall rate
 66 of $t\bar{t}H$ production. There are several theories of physics Beyond the Standard Model (BSM),
 67 however, that would affect the kinematics involved in $t\bar{t}H$ production without altering its overall
 68 rate [**Dumont_2013**].

69 An Effective Field Theory approach can be used to model the low energy effects of new, high
 70 energy physics, by parameterizing BSM effects as dimension-six operators. The addition of these
 71 operators can be shown to modify the transverse momentum (p_T) spectrum of the Higgs Boson
 72 [**Banerjee_2014**]. Therefore, reconstructing the momentum spectrum of the Higgs provides a
 73 means to observe new physics in the Higgs sector.

74 This note reports on the feasibility of performing differential measurements in $t\bar{t}H$ events with
 75 multiple leptons in the final state, using Monte Carlo (MC) simulations scaled to 139 fb^{-1} at an
 76 energy $\sqrt{s} = 13 \text{ TeV}$. Events are separated into channels based on the number of light leptons
 77 (electrons and muons) in the final state - either two same-sign leptons (2lSS), or three leptons
 78 (3l), where the 3l channel is split into two based on the decay of the Higgs.

79 The presence of multiple neutrinos in the final state of the multilepton channels introduces an
 80 ambiguity that prevents the Higgs from being fully reconstructed. This motivates the use of
 81 sophisticated machine learning techniques to better predict the Higgs p_T spectrum for these events.
 82 A deep neural network is used to identify which objects originate from the decay of the Higgs,
 83 and reconstruct the momentum of the Higgs Boson in each event. This spectrum is fit to data in
 84 the three decay channels considered in order to extract normalization factors on $t\bar{t}H$ produced
 85 with high ($> 150 \text{ GeV}$) and low ($< 150 \text{ GeV}$) Higgs.

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

92 3 Data and Monte Carlo Samples

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

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

102 Samples were then generated from these HIGG8D1 derivations using AnalysisBase version
 103 21.2.127. A ptag of p4133 was used for MC samples, and p4134 for data.

104 3.1 Data Samples

105 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through
 106 2018, which represents an integrated luminosity of 139 fb^{-1} and an energy of $\sqrt{s} = 13 \text{ TeV}$. All
 107 data used in this analysis was included in one of the following Good Run Lists:

- 108 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
 109 _PHYS_StandardGRL_All_Good_25ns.xml
- 110 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
 111 _PHYS_StandardGRL_All_Good_25ns.xml
- 112 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL
 113 _All_Good_25ns_Triggerno17e33prim.xml
- 114 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL
 115 _All_Good_25ns_Triggerno17e33prim.xml

116 3.2 Monte Carlo Samples

117 Several Monte Carlo (MC) generators were used to simulate both signal and background processes.
 118 For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific event
 119 generator used for each of these MC samples is listed in Table 1. A Higgs mass of 125 GeV is
 120 assumed in all simulations.

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

Process	Event generator	ME order	Parton Shower	PDF
t̄H	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [Ball:2014uwa] (CT10 [ct10])
t̄W	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
t̄(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
t̄t	POWHEG-BOX v2 [powheggtt]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄γ	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
tZ	MG5_AMC	LO	PYTHIA 6	CTEQ6L1
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t, t̄t̄t̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄W+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	Powheg-Box v1 [powhegstp]	NLO	PYTHIA 6	CT10
qqVV, VVV				
Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

121 The signal sample ($t\bar{t}H$) is modelled at NLO with POWHEG-BOX v2 using the NNPDF2.0 parton
 122 distribution function (PDF) [[ATL-PHYS-PUB-2015-011](#)]. Parton showering and hadronisation
 123 were modelled with PYTHIA 8.2 [[ATL-PHYS-PUB-2011-009](#)]. The $t\bar{t}H$ sample is normalized
 124 to a cross-section of 507^{+35}_{-50} fb based on NLO calculations. Uncertainties are based on varying
 125 the QCD factorisation and renormalisation scale, as well as uncertainties in the PDF and α_s .

126 The $t\bar{t}W$ background is simulated using Sherpa 2.2.1 with the NNPDF3.0 NLO PDF. The matrix
 127 element is calculated with up to one additional parton at NLO, and up to two at LO. As explained
 128 in detail in [[ttH_paper](#)], the $t\bar{t}W$ contribution predicted by MC is found disagree significantly
 129 with what is observed in data. While an effort is currently being undertaken to measure $t\bar{t}W$ more
 130 accurately, the approach used by the 79.8 fb^{-1} $t\bar{t}H$ analysis is used here: A normalization
 131 factor of 1.68 is applied to the MC estimate of $t\bar{t}W$ and additional systematic uncertainties on
 132 $t\bar{t}W$ are included to account for this modelling discrepancy, as outlined in Section 7.

133 While the main $t\bar{t}H$ analysis uses a more sophisticated data-driven approach to estimating the
 134 contribution of events with non-prompt leptons (or “fakes”), at the time of this note this strategy has
 135 not been completely developed for the full Run-2 dataset. Therefore, the non-prompt contribution
 136 is estimated with MC, while applying normalization corrections and systematic uncertainties
 137 derived from data driven techniques developed for the 79.8 fb^{-1} $t\bar{t}H/t\bar{t}W$ analysis [[ttH_paper](#)].
 138 The primary contribution to the non-prompt lepton background is from $t\bar{t}$ production, with V+jets
 139 and single-top as much smaller sources. Likelihood fits over several control regions enriched with
 140 these non-prompt backgrounds are fit to data in order to derive normalization factors for these

¹⁴¹ backgrounds. The specific normalization factors and uncertainties applied to the non-prompt
¹⁴² contributions are listed in Section 7.

¹⁴³ The specific DSIDs used in the analysis are listed below:

Sample	DSID
t̄H	345873-5, 346343-5
VV	364250-364254, 364255, 363355-60, 364890
t̄W	413008
t̄Z	410156, 410157, 410218-20
low mass t̄Z	410276-8
Rare Top	410397, 410398, 410399
single Top	410658-9, 410644-5
three Top	304014
four Top	410080
t̄WW	410081
Z + jets	364100-41
low mass Z + jets	364198-215
W + jets	364156-97
Vγ	364500-35
tZ	410560
tW	410013-4
WtZ	410408
VVV	364242-9
VH	342284-5
WtH	341998
t̄tγ	410389
t̄t	410470

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

¹⁴⁴ 4 Object Reconstruction

¹⁴⁵ All analysis channels considered in this note share a common object selection for leptons and jets,
¹⁴⁶ as well as a shared trigger selection.

¹⁴⁷ 4.1 Trigger Requirements

¹⁴⁸ Events are required to be selected by dilepton triggers, as summarized in Table 3.

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

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

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. Electron candidate are also required to pass TightLH identification.

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 $|\eta| < 2.5$. Muons are required to Medium ID requirements.

All leptons are required to pass a non-prompt BDT selection developed by the main $t\bar{t}H - ML/t\bar{t}W$ analysis, described in detail in [ttH_paper]. Optimized working points and scale factors for this BDT are taken from that analysis. This BDT and the WPs used are summarized in Appendix .1,

167 **4.3 Jets**

168 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
169 calorimeters [**ATL-PHYS-PUB-2015-015**], using the anti- k_t algorithm with a radius parameter
170 $R = 0.4$. Particle Flow, or PFlow, jets are used in the analysis, which are hadronic objects
171 reconstructed using information from both the tracker and the calorimeter. Jets with energy
172 contributions likely arising from noise or detector effects are removed from consideration
173 [**ATLAS-CONF-2015-029**], and only jets satisfying $p_T > 25$ GeV and $|\eta| < 2.5$ are used in this
174 analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track association algorithm is used to
175 confirm that the jet originates from the selected primary vertex, in order to reject jets arising from
176 pileup collisions [**PERF-2014-03**].

177 **4.4 B-tagged Jets**

178 Each analysis channel used in this analysis includes b-jets in the final state. These are identified
179 using the DL1r b-tagging algorithm, which uses jet vertex and kinematic information to distinguish
180 heavy and light flavored jets. These features are used as inputs to a neural network, the output
181 of which is used to form calibrated working points (WPs) based on how likely a jet is to have
182 originated from a b-quark. This analysis uses the 70% DL1r WP - implying an efficiency of 70%
183 for truth b-jets - for selecting b-tagged jets.

184 **4.5 Missing Transverse Energy**

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

190 **4.6 Overlap removal**

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

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

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

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

199 5 Higgs Momentum Reconstruction

200 Reconstructing the momentum of the Higgs boson is a particular challenge for channels with
 201 leptons in the final state: Because all channels include at least two neutrinos in the final state, the
 202 Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted
 203 by a neural network when provided with the kinematics of the Higgs Boson decay products - as
 204 verified by studies detailed in Appendix A.3. With this in mind, several layers of MVAs are used
 205 to reconstruction the Higgs momentum:

206 The first layer is a model designed to select which jets are most likely to be the b-jets that came
 207 from the top decay, detailed in Section 5.3. As described in Section 5.4, the kinematics of these jets
 208 and possible Higgs decay products are fed into the second layer, which is designed to identify the
 209 decay products of the Higgs Boson itself. The kinematics of the particles this layer identifies as
 210 most likely to have originated from the Higgs decay are then fed into yet another neural-network,
 211 which predicts the momentum of the Higgs (5.5). For the 3l channel, an additional MVA is used
 212 to determine the decay mode of the Higgs boson in the 3l channel (5.6).

213 Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_AMC.
 214 Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training in order to
 215 increase the statistics of the training sample.

216 For all of these models, the Keras neural network framework, with Tensorflow 2.0 as the backend,
 217 is used, and the number of hidden layers and nodes are determined using grid search optimization.
 218 Each neural network uses the LeakyReLU activation function, a learning rate of 0.01, and the
 219 Adam optimization algorithm, as alternatives are found to either decrease or have no impact on
 220 performance. Batch normalization is applied after each layer in order to stabilize the model and
 221 decrease training time. For the classification algorithms (b-jet matching, Higgs reconstruction,
 222 and 3l decay identification) binary-cross entropy is used as the loss function, while the p_T
 223 reconstruction algorithm uses MSE.

224 The specific inputs features used for each model are arrived at through a process of trial and error
 225 - features considered potentially useful are tried, and those that are found to increase performance
 226 are included. While each model includes a relatively large number of features, some using
 227 upwards of 30, this inclusive approach is found to maximize the performance of each model while

228 decreasing the variance compared to a reduced number of inputs. Each input feature is validated
 229 by comparing MC simulations to 79.8 fb^{-1} of data, as shown in the sections below.

230 **5.1 Physics Object Truth Matching**

231 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson using
 232 MC simulations of $t\bar{t}H$ events. The kinematics of the reconstructed physics objects, as well as
 233 event level variables, are used as inputs, with the parent ID taken from the truth record used to
 234 label the data. The objects considered include light leptons and jets.

235 Reconstructed physics objects are matched to particle level objects, in order to identify the parent
 236 particle of these reconstructed objects. Reconstructed jets are matched to truth jets based on
 237 the requirements that the reco jet and truth jet fall within $\Delta R < 0.4$, and the two objects have
 238 a p_T that agrees within 10%. Truth level and reco level leptons are required to have the same
 239 flavor, a $\Delta R < 0.1$, and p_T that agree within 10%. Events where no match can be found between
 240 the particle level decay products of the Higgs and the reconstructed objects are not included in
 241 training.

242 Leptons considered as possible Higgs and top decay candidates are required to pass the selection
 243 described in Section 4.2. For jets, however, it is found that a large fraction that originate from either
 244 the top decay or the Higgs decay fall outside the selection described in Section 4.3. Specifically,
 245 jets from the Higgs decay tend to be soft, with 32% having $p_T < 25 \text{ GeV}$. Therefore jets with p_T
 246 $< 15 \text{ GeV}$ are considered as possible candidates in the models described below. By contrast, less
 247 than 5% of the jets originating from the Higgs fall below this p_T threshold. The jets are found to
 248 be well modeled even down to this low p_T threshold, as shown in Section 6.1. The impact of
 249 using different p_T selection for the jet candidates is considered in detail in Section A.6. As they
 250 are expected to originate from the primary vertex, jets are also required to pass a JVT cut.

251 **5.2 Truth Level Studies**

252 Attempts to identify the decay products of the Higgs are motivated by the ability to reconstruct
 253 the Higgs momentum based on their kinematics. In order to demonstrate that this is case, the
 254 kinematics of reconstructed objects that are truth matched to the Higgs decay are used as inputs
 255 to a neural network which is designed to predict the momentum of the Higgs. This is done in
 256 the 2lSS channel -

257 **5.3 b-jet Identification**

258 Including the kinematics of the b-jets that originate from the top decay is found to improve the
 259 identification of the Higgs decay products, and improve the accuracy with which the Higgs
 260 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.

5.3.1 2lSS Channel

For the 2lSS channel, the input features shown in Table 5 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	$\Delta R(j_0 l_0)(j_1 l_1)$
$\Delta R(j_0 l_1)(j_1 l_0)$	$p_T(j_0 j_1 l_0 l_1 E_T^{\text{miss}})$	$M(j_0 j_1 l_0 l_1 E_T^{\text{miss}})$
$\Delta\phi(j_0)(E_T^{\text{miss}})$	$\Delta\phi(j_1)(E_T^{\text{miss}})$	HT jets
nJets	E_T^{miss}	

Table 5: Input features used in the b-jet identification algorithm for the 2lSS channel

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.

284 The difference between the distributions for a few of these features for the "correct" (i.e. both jets
 285 are truth b-jets), and "incorrect" combinations are shown in Figure 5.1. The correct and incorrect
 286 contributions are scaled to the same integral, so as to better demonstrate the differences in the
 287 distributions.

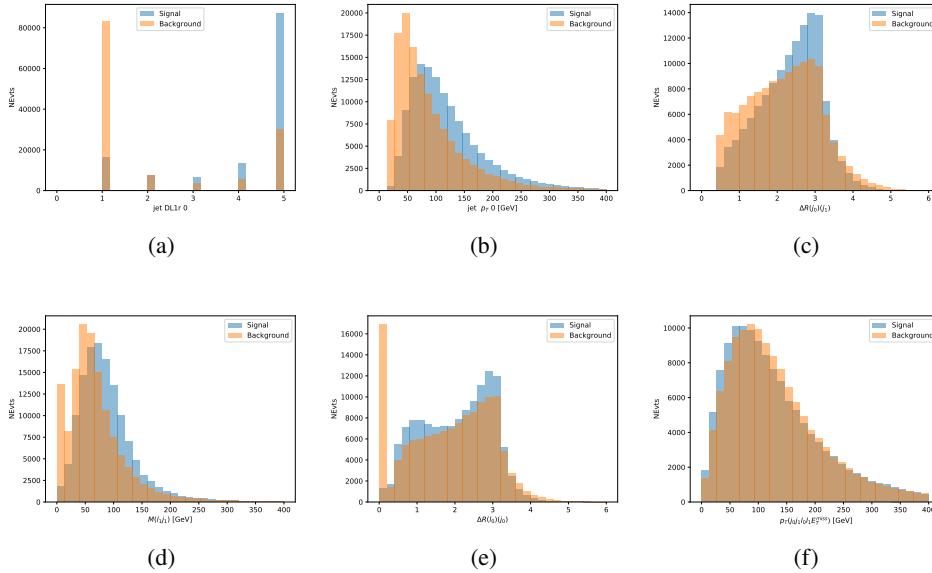


Figure 5.1: Input features for top2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T 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} .

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

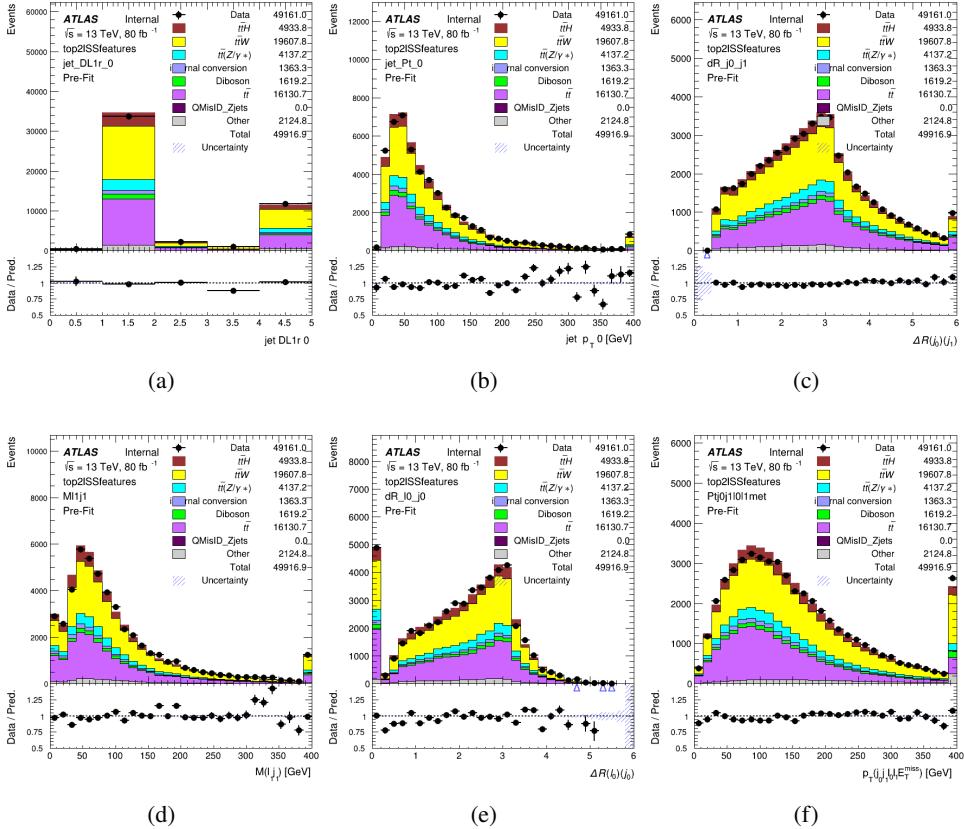


Figure 5.2: Data/MC comparisons of input features for top2lSS training for 79.8 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

Based on the results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as overfitting is found to not be an issue. The output score distribution as well as the ROC curve for the trained model are shown in Figure 5.3.1. The model is found to identify the correct pairing of jets for 73% of 2lSS signal events on test data.

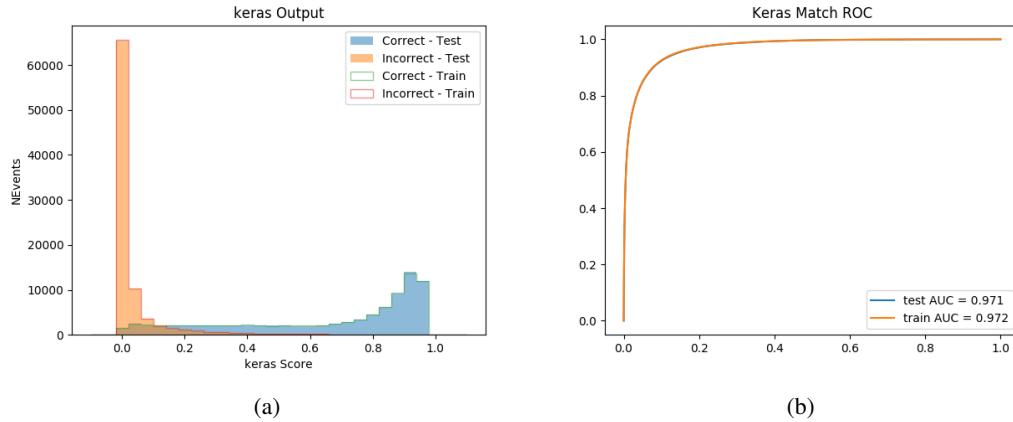


Figure 5.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

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

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

Table 6: Accuracy of the NN in identifying b-jets from tops in 2lSS events, overall and split by the number of b-tagged jets in the event, compared to the accuracy of taking the two highest b-tagged jets.

301 5.3.2 3l Channel

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

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

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

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

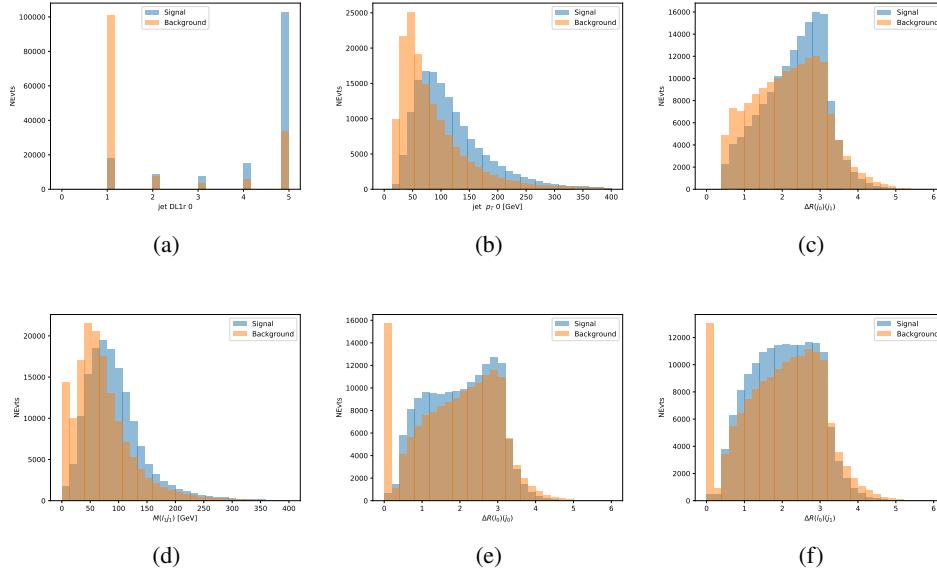


Figure 5.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and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

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

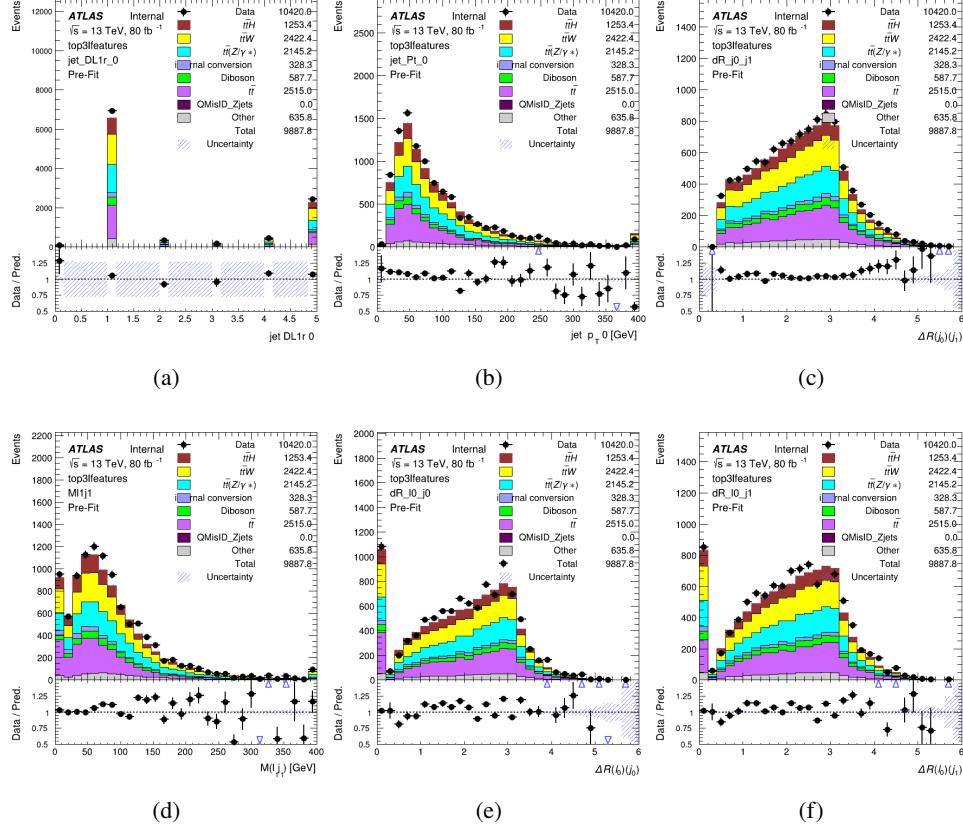


Figure 5.5: Data/MC comparisons of input features for top3l training for 79.8 fb^{-1} of data. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

308 Again, the dataset is downsized to reduce the ratio of correct and incorrect combination from 20:1,
 309 to 5:1. Around 7 million events are used for training, with 10% set aside for testing. Based on the
 310 results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with
 311 60 nodes each. The output score distribution as well as the ROC curve for the trained model are
 312 shown in Figure 5.3.2.

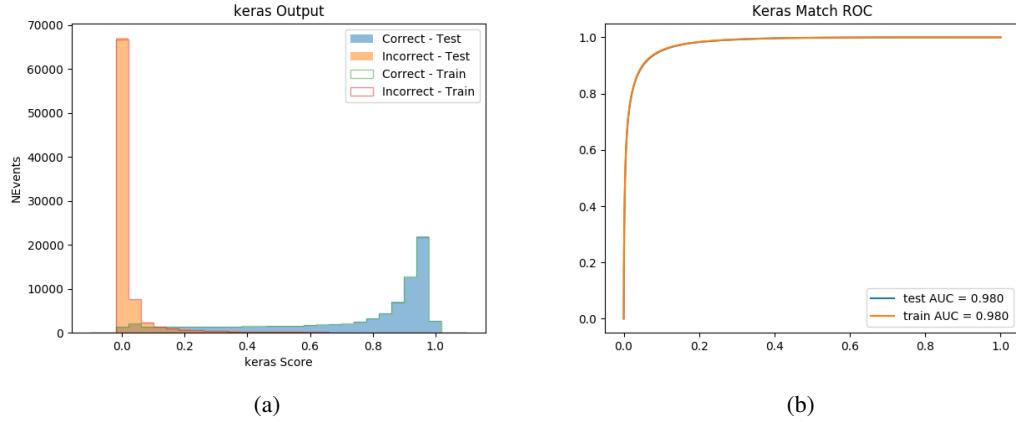


Figure 5.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

Table 8: Accuracy of the NN in identifying b-jets from tops, compared to the naive method of taking the highest b-tagged jets.

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

315 5.4 Higgs Reconstruction

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

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

323 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In the
 324 semi-leptonic case the reconstructed decay products consist of two jets and a single leptons. For

325 the fully-leptonic case, the decay products include 2 of the three leptons associated with the
326 event. For training the models, events are separated into these two categories using truth level
327 information. A separate MVA, described in Section 5.6, is used to make this distinction at reco
328 level and determine which model to use.

329 For all channels, the models described in Section 5.3 are used to identify b-jet candidates, whose
330 kinematics are used to identify the Higgs decay products. These jets are not considered as possible
331 candidates for the Higgs decay, justified by the fact that these models are found to misidentify jets
332 from the Higgs decay as jets from the top decay less than 1% of the time.

333 **5.4.1 2lSS Channel**

334 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural
335 network is trained on the kinematics of different combinations of leptons and jets, as well as the
336 b-jets identified in Section 5.3, with the specific input features listed in Table 9.

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

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

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

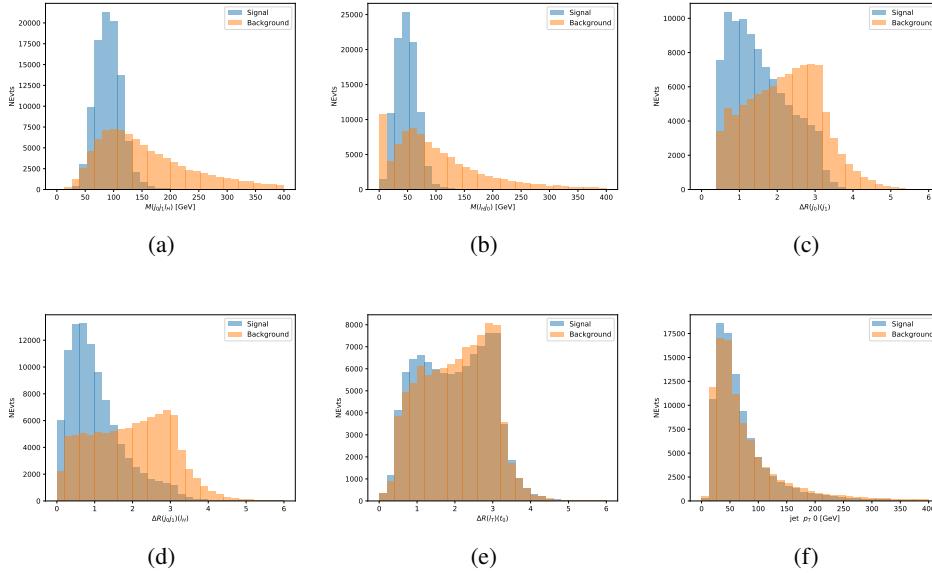


Figure 5.7: Input features for higgs2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate,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

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

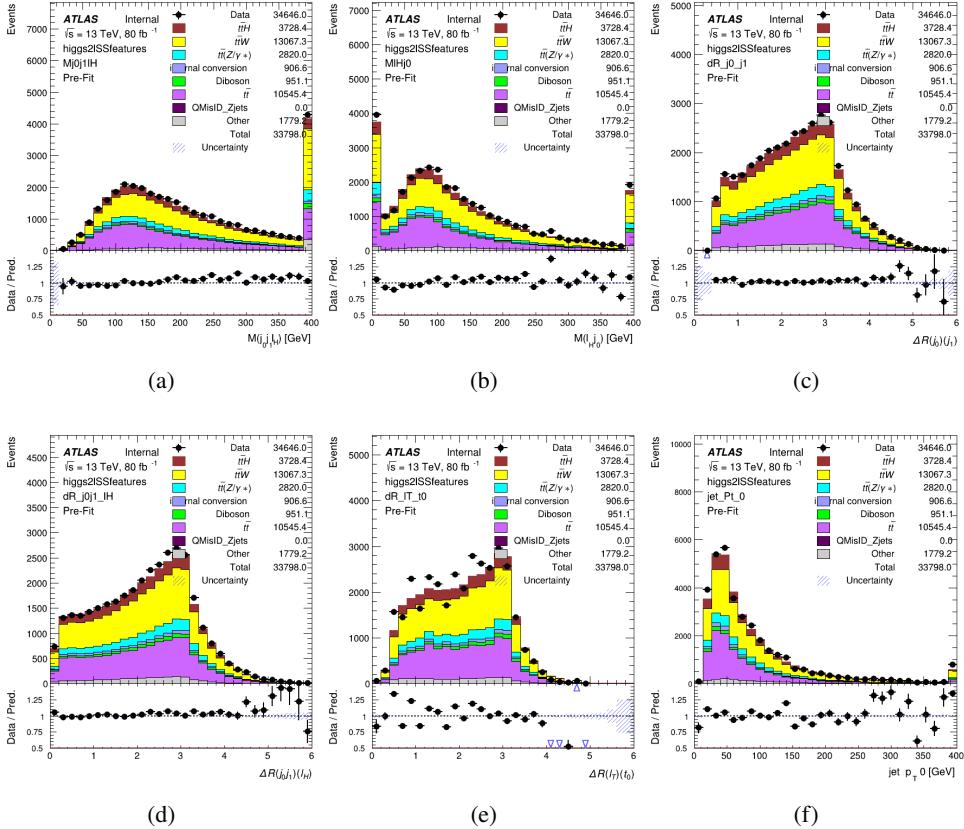


Figure 5.8: Data/MC comparisons of input features for higgs2lSS training for 79.8 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 and the lepton candidate, (e) ΔR between the lepton from the top and the leading b-jet, (f) the p_T of jet 0.

343 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2 million
 344 events, with an additional 200,000 reserved for testing the model. In order to compensate for
 345 large number of incorrect combinations, these have been downsampled such that the correct
 346 combinations represent over 10% of the training set. The output of the NN is summarized in
 347 Figure 5.4.1.

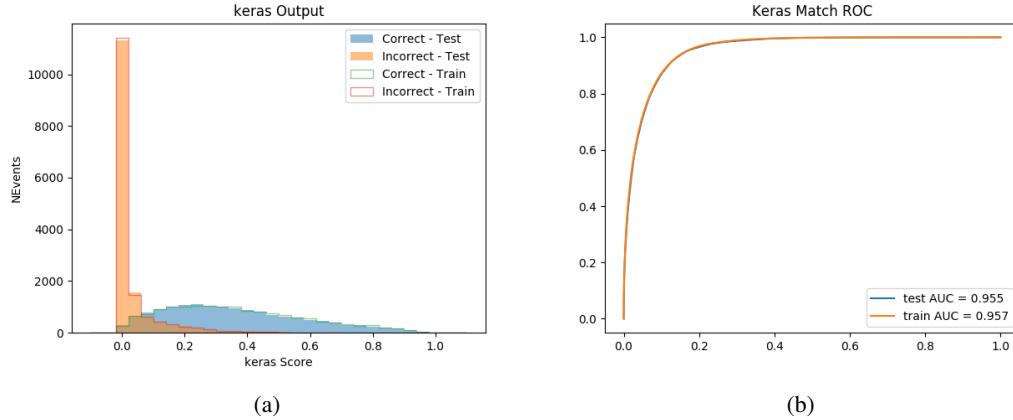


Figure 5.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

351 5.4.2 3l Semi-leptonic Channel

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

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

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

356 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other two
 357 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
 358 b-jet identification algorithm. The b-jet Reco Score is the output of the Higgs reconstruction
 359 algorithm.

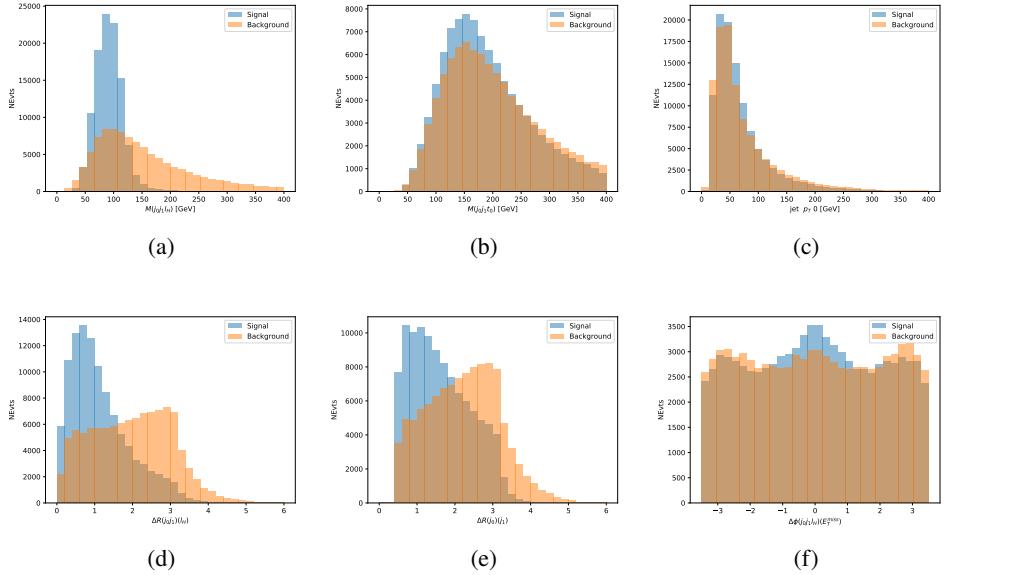


Figure 5.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

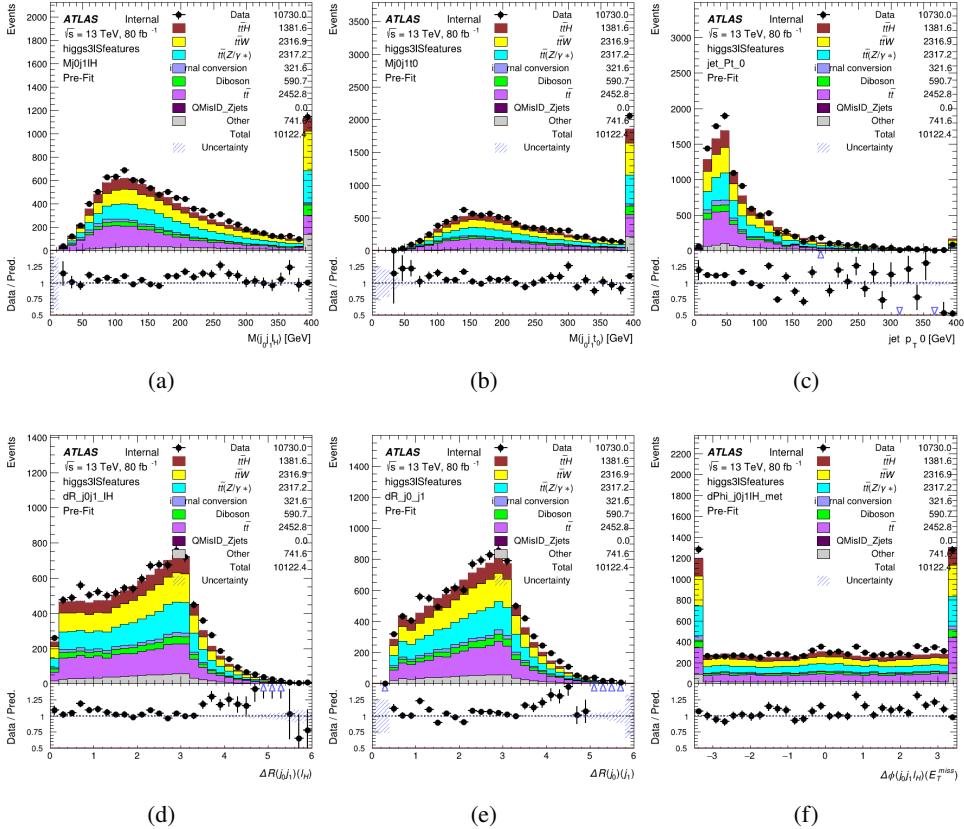


Figure 5.11: Data/MC comparisons of input features for higgs3IS training for 79.8 fb^{-1} of data.

363 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events. Once
 364 again, incorrect combinations are downsampled, such that the correct combinations are around
 365 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN is
 366 summarized in Figure 5.4.2.

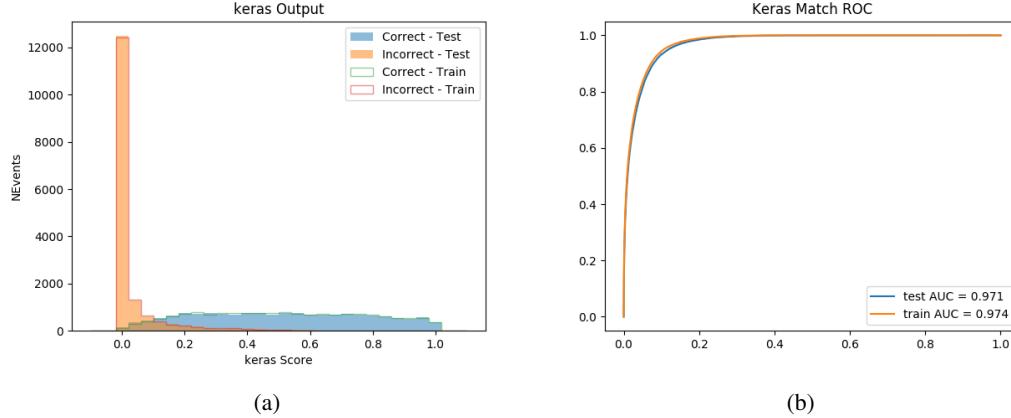


Figure 5.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries., (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

370 **5.4.3 3l Fully-leptonic Channel**

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

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

Table 11: Input features used to identify the Higgs decay products in 3lF events

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

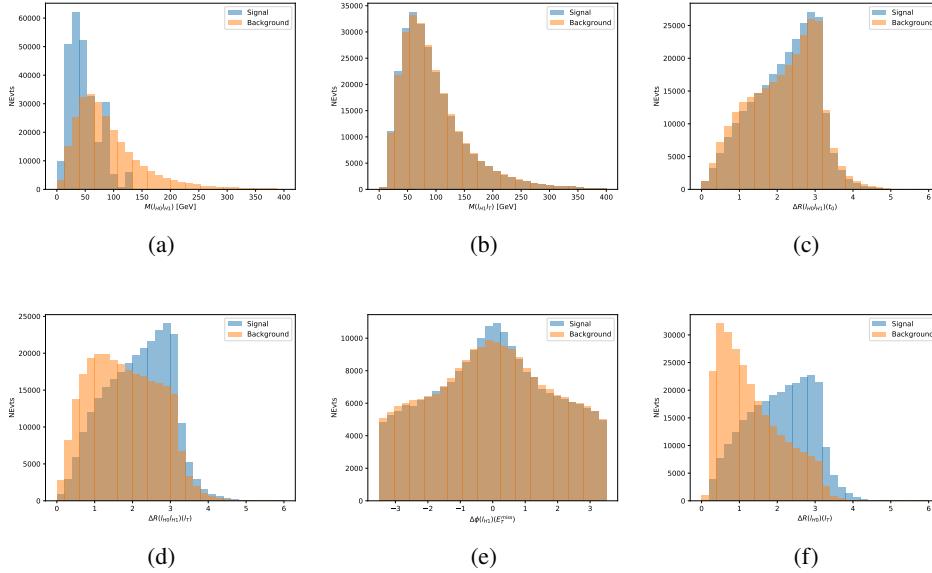


Figure 5.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

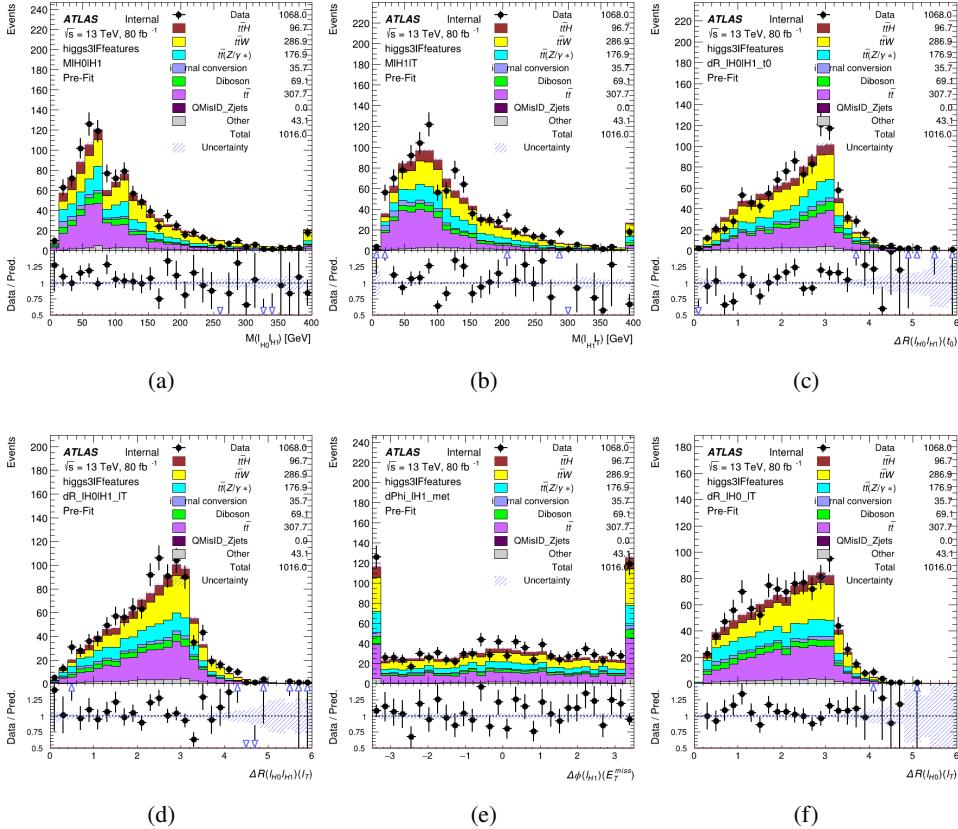


Figure 5.14: Data/MC comparisons of input features for higgs3lF training for 79.8 fb^{-1} of data.

381 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events, with
 382 10% of the dataset reserved for testing. The output of the model is summarized in Figure 5.4.3.

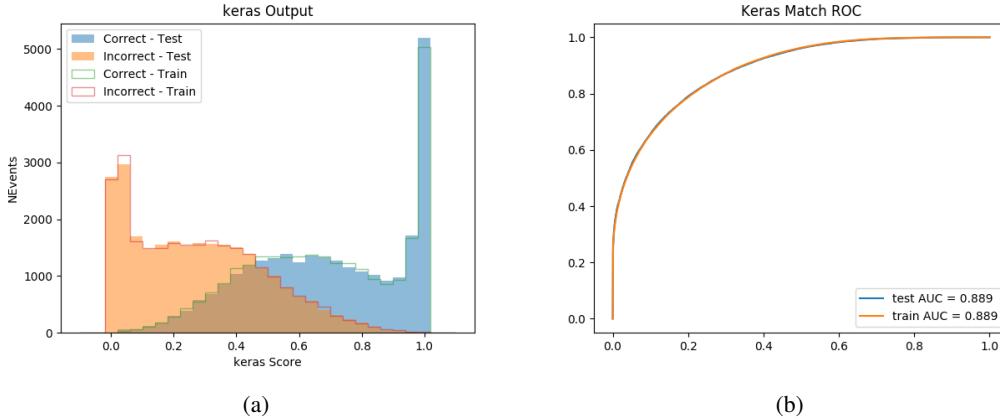


Figure 5.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

384 5.5 p_T Prediction

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

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

394 Because the analysis uses a two bin fit of the Higgs p_T , the momentum reconstruction could be
 395 treated as a binary classification problem, rather than a regression problem. This approach is
 396 explored in detail in Section A.5, and is found not to provide any significant increase in sensitivity.
 397 The regression approach is used because it provides more flexibility for future analyses, as it is
 398 independent of the cutoff between high and low p_T , as well as the number of bins. Further, a
 399 regression allows the output of the neural network to be more clearly understood, as it can be
 400 directly compared to a physics observable.

401 **5.5.1 2ISS Channel**

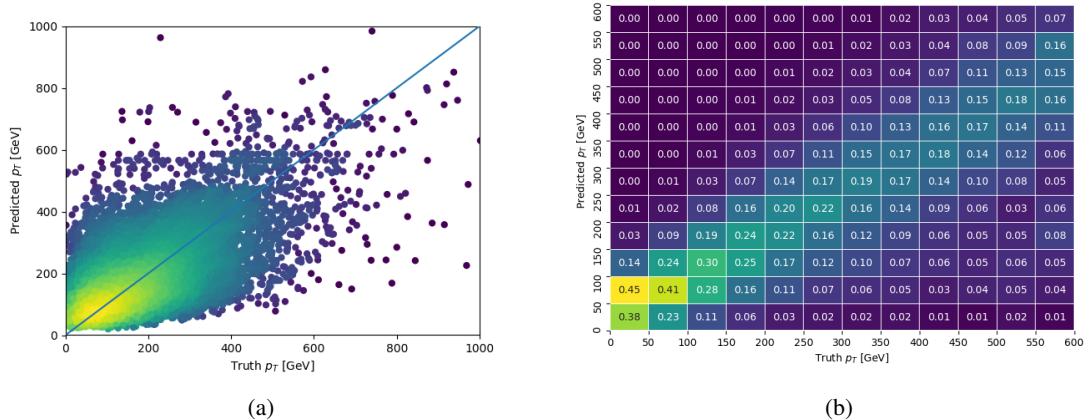
402 The input variables listed in Table 12 are used to predict the Higgs p_T in the 2ISS channel. Here
403 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
404 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
405 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
406 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
407 reconstruction algorithm, and the b-jet identification algorithm, respectively.

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

Table 12: Input features for reconstructing the Higgs p_T spectrum for 2lSS events

408 The optimal neural network architecture for this channel is found to consist of 7 hidden layers with
 409 60 nodes each. The input data set includes 1.2 million events, 10% of which is used for testing,
 410 the other 90% for training. Training is found to converge after around 150 epochs.

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

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

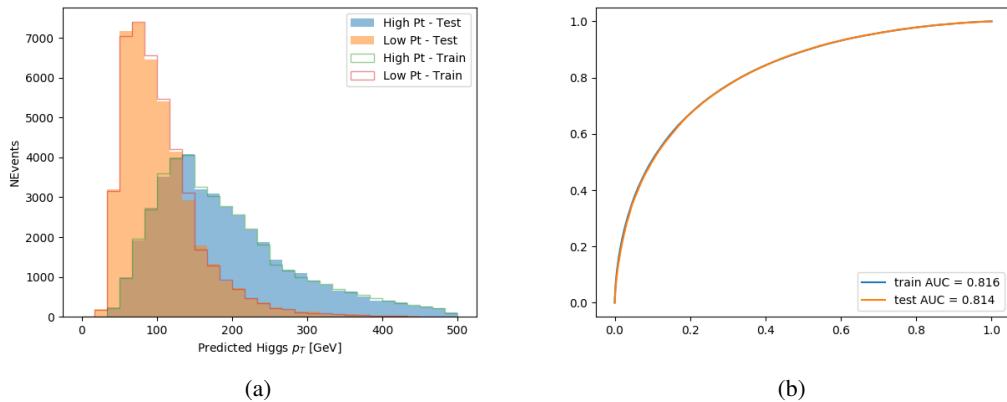


Figure 5.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

5.5.2 3l Semi-leptonic Channel

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

HT jets	MET	M(j ₀ j ₁)
M(j ₀ j ₁ l _H)	M(j ₀ j ₁ l _{T0})	M(j ₀ j ₁ l _{T1})
M(j ₀ j ₁ b ₀)	M(j ₀ j ₁ b ₁)	M(b ₀ l _{T0})
M(b ₀ l _{T1})	M(b ₁ l _{T0})	M(b ₁ l _{T1})
$\Delta\phi(j_0j_1l_H)(E_T^{\text{miss}})$	$\Delta R(j_0)(j_1)$	$\Delta R(j_0j_1)(l_H)$
$\Delta R(j_0j_1)(l_{T1})$	$\Delta R(j_0j_1)(b_0)$	$\Delta R(j_0j_1)(b_1)$
$\Delta R(j_0j_1l_H)(l_{T0})$	$\Delta R(j_0j_1l_H)(l_{T1})$	$\Delta R(j_0j_1l_H)(b_0)$
$\Delta R(j_0j_1l_H)(b_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(l_H)(l_{T1})$	$\Delta R(l_{T0})(l_{T1})$	$\Delta R(l_{T0})(b_0)$
$\Delta R(l_{T0})(b_1)$	$\Delta R(l_{T1})(b_0)$	$\Delta R(l_{T1})(b_1)$
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 13: Input features for reconstructing the Higgs p_T spectrum for 3lS events

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

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

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

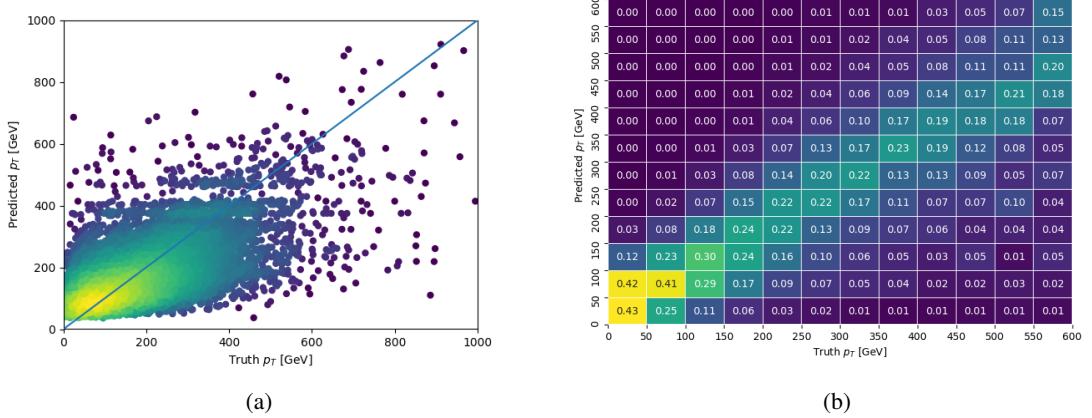


Figure 5.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

436 Figure 5.19 shows (a) the output of the NN for events with truth p_T less than and greater than
 437 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low p_T
 438 events.

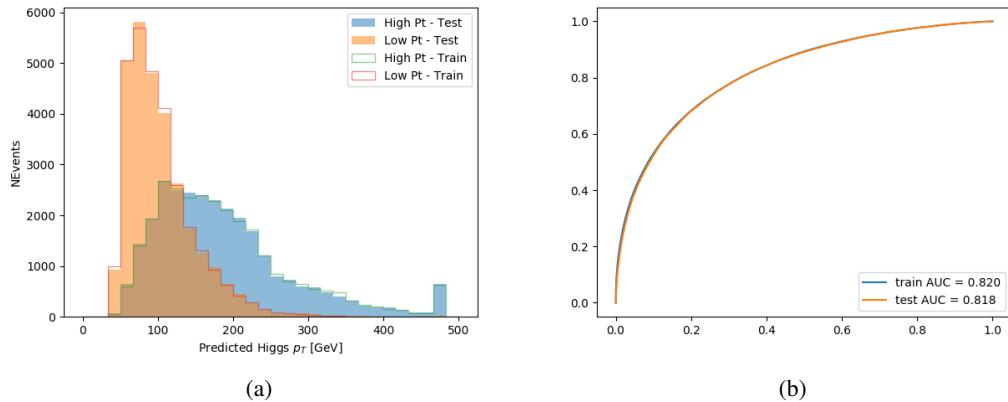


Figure 5.19: (

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

439 **5.5.3 3l Fully-leptonic Channel**

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

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

Table 14: Input features for reconstructing the Higgs p_T spectrum for 3lF events

⁴⁴¹ l_{H0} and l_{H1} represent the two leptons identified by the Higgs reconstruction model as originating
⁴⁴² from the Higgs, while l_T is the other lepton in the event. The Higgs Reco Score and b-jet Reco
⁴⁴³ Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification algorithm,
⁴⁴⁴ respectively.

⁴⁴⁵ The optimal neural network architecture for this channel is found to consist of 5 hidden layers
⁴⁴⁶ with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for testing,
⁴⁴⁷ the other 90% for training. Training is found to converge after around 150 epochs.

⁴⁴⁸ The predicted transverse momentum, as a function of the truth p_T , is shown in Figure 5.20.

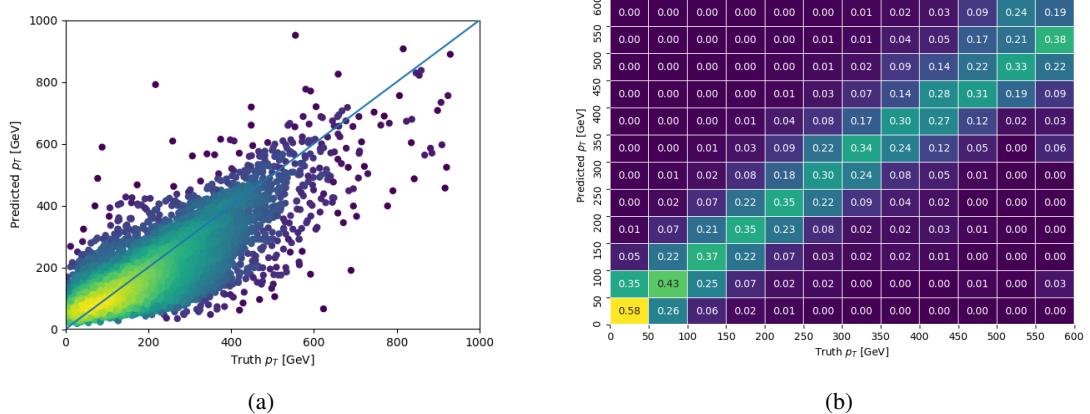


Figure 5.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

⁴⁴⁹ When split into high and low p_T , based on a cutoff of 150 GeV, the

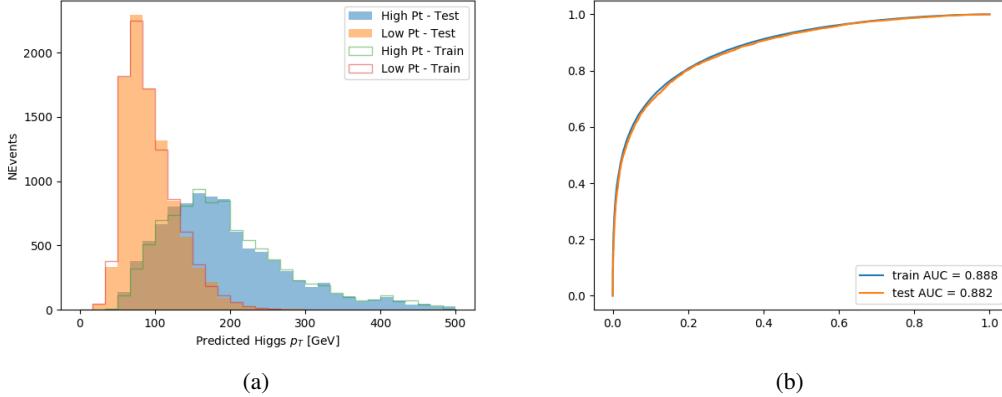


Figure 5.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

5.6 3l Decay Mode

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

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

HT jets	$M(l_0 t_0)$	$M(l_0 t_1)$
$M(l_1 t_0)$	$M(l_1 t_1)$	$M(l_0 l_1)$
$M(l_0 l_2)$	$M(l_1 l_2)$	$\Delta R(l_0 t_0)$
$\Delta R(l_0 t_1)$	$\Delta R(l_1 t_0)$	$\Delta R(l_1 t_1)$
$\Delta R(l l_0 1)$	$\Delta R(l l_0 2)$	$\Delta R(l l_1 2)$
Lepton η 0	Lepton η 1	Lepton η 2
Lepton ϕ 0	Lepton ϕ 1	Lepton ϕ 2
Lepton p_T 0	Lepton p_T 1	Lepton p_T 2
E_T^{miss}	nJets	nJets OR DL1r 60
nJets OR DL1r 85	score3lF	score3lS
topScore	total charge	

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

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

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

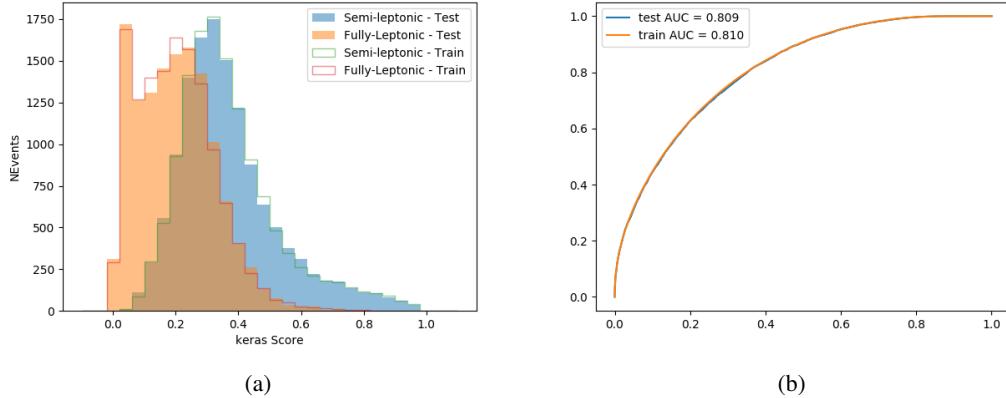


Figure 5.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

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

466 6 Signal Region Definitions

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

470 6.1 Pre-MVA Event Selection

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

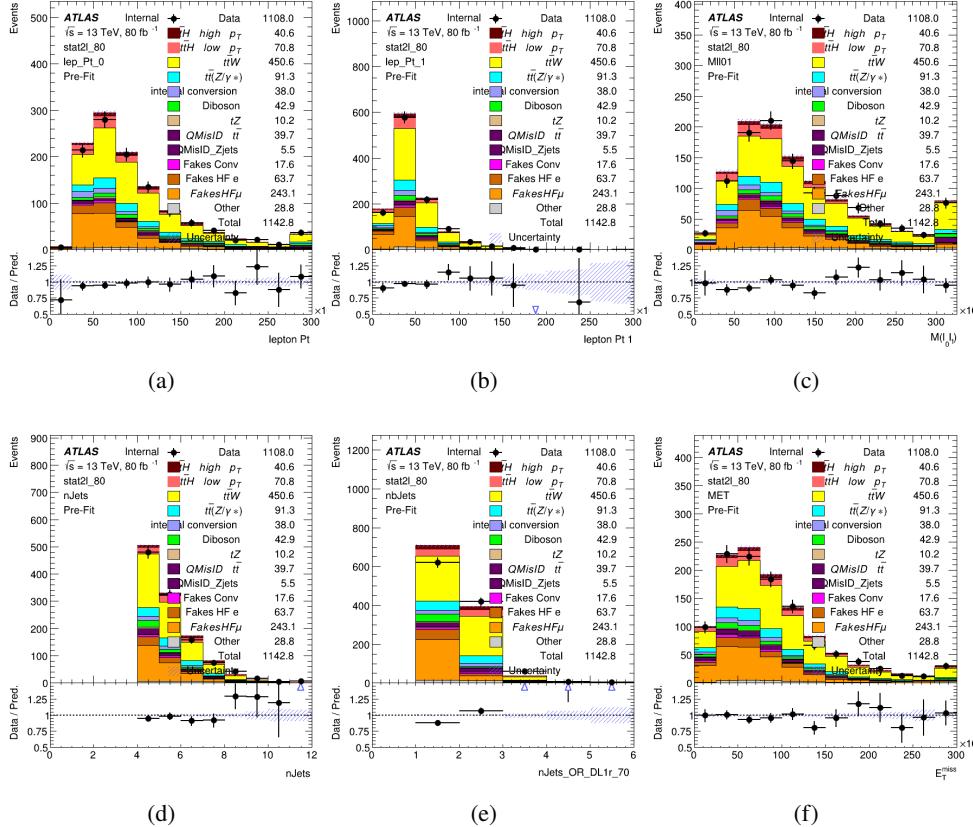
- 473 • Two very tight, same-charge, light leptons with $p_T > 20$ GeV
- 474 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets
- 475 • No reconstructed tau candidates

⁴⁷⁶ The event yield after the 2lSS preselection has been applied, for MC and data at 79.8 fb^{-1} , is
⁴⁷⁷ shown in Table 16.

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

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

478 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.

479 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
- $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs

486 The event yield after the 3l preselection has been applied, for MC and data at 79.8 fb^{-1} , is shown
487 in Table 6.1.

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

Table 17: Yields of the analysis

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

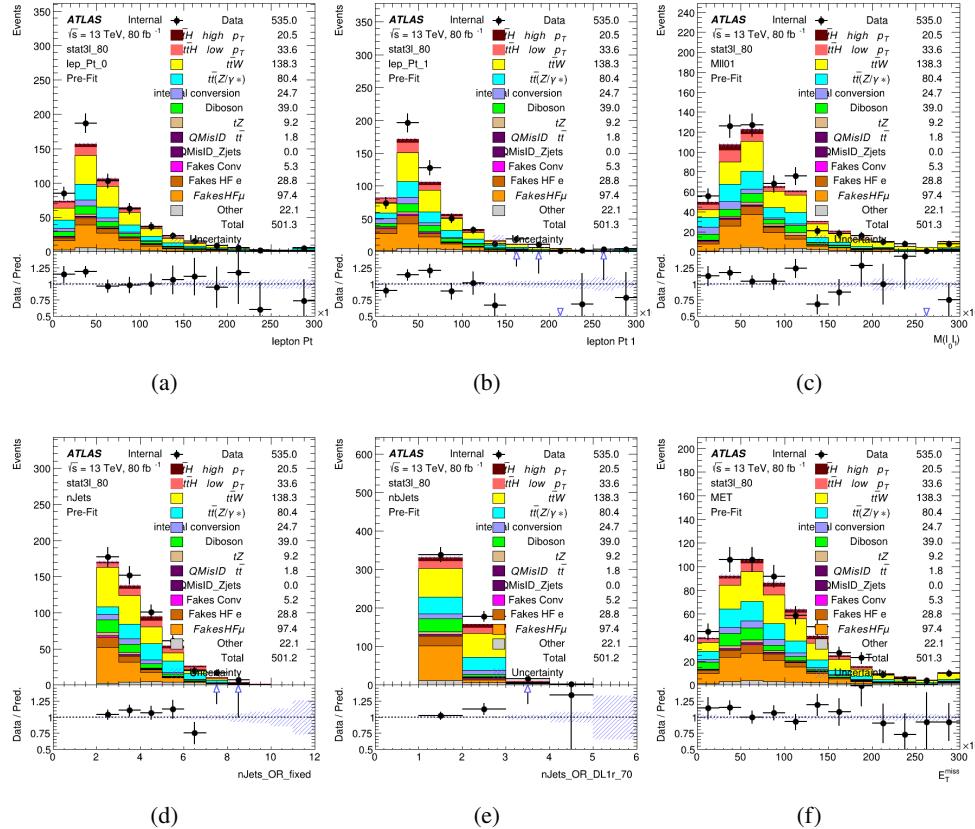


Figure 6.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

489 **6.2 Event MVA**

490 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
 491 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
 492 (3lF). Here events with three leptons are split into 3lS and 3lF based on the model described in 5.6.
 493 In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost [xgboost]
 494 are trained using the kinematics of signal and background events derived from Monte Carlo
 495 simulations. Events are weighted in the BDT training by the weight of each Monte Carlo event.

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

500 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
501 3lF high p_T , and 3lF low p_T .

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

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

Table 18: Input features used to distinguish signal and background events in the 2lSS channel.

⁵⁰³ While for each of the 31 BDTs, the features listed below are used for training:

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

Table 19: Input features used to distinguish signal and background events in the 3l channel.

504 Modelling of each of these input features is verified in Appendix A.2 by comparing data and MC
 505 for 79.8 fb^{-1} . The BDTs are produced with a maximum tree depth of 6, using AUC as the target
 506 loss function. The BDT response distribution and ROC curve for each model is shown in Figures
 507 **6.3-6.5**.

2lSS

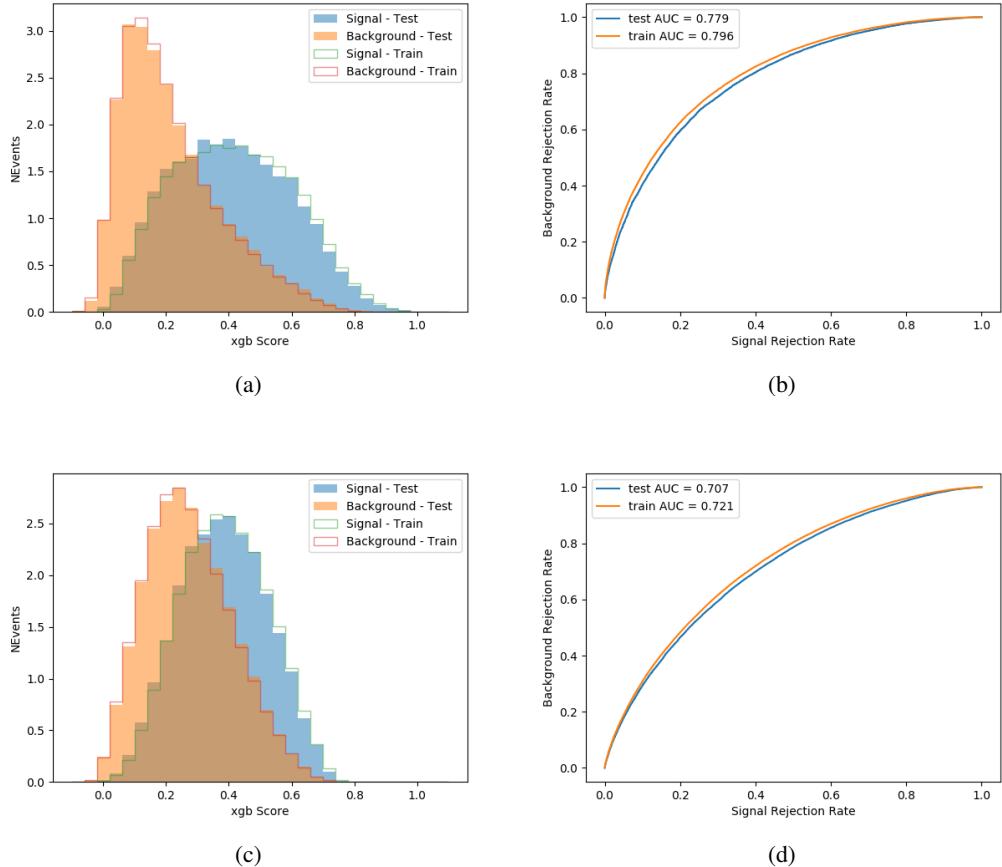


Figure 6.3: Output BDT scores of training and testing data for signal (blue) and background (orange) for 2lSS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 2lSS high and low p_T models, respectively.

3l - Semileptonic

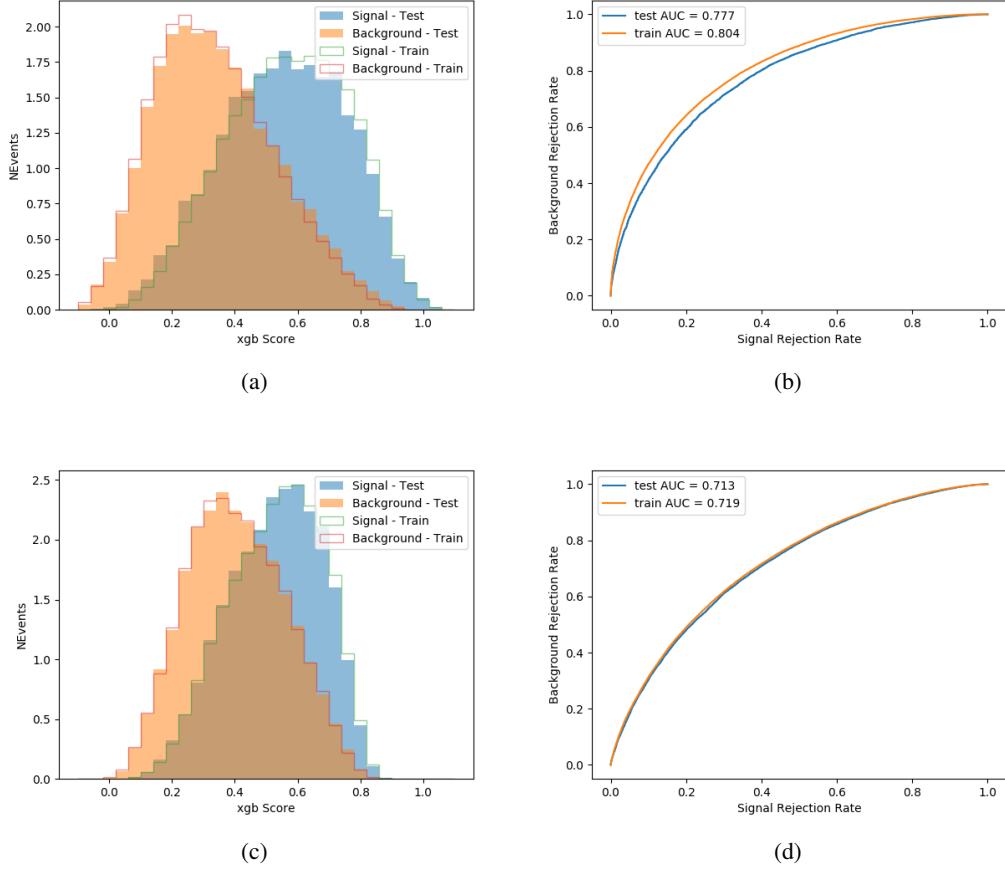


Figure 6.4: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lS high and low p_T models, respectively.

3l - Fully Leptonic

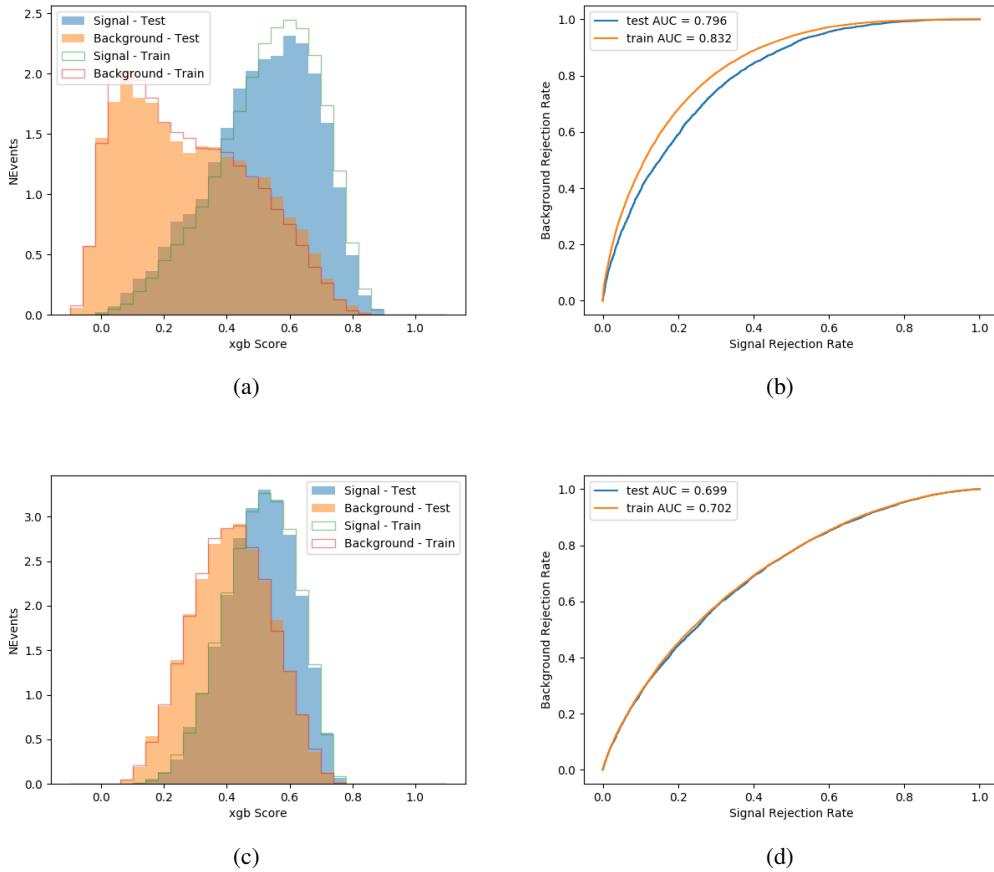


Figure 6.5: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lF events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lF high and low p_T models, respectively.

508 Output distributions of each MVA, comparing MC predictions to data at 79.8 fb^{-1} are shown in
 509 figures 6.6-6.2.

510 **6.3 Signal Region Definitions**

511 Once pre-selection has been applied, channels are further refined based on the MVAs described
 512 above. The output of the model described in Section 5.6 is used to separate the three channel into
 513 two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson.
 514 This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

High p_T Background Rejection BDTs

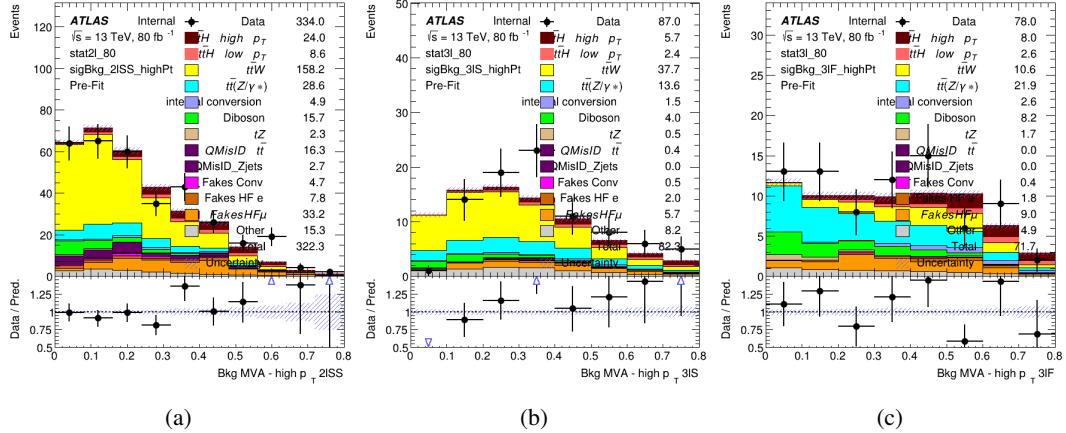


Figure 6.6: Output score of the high p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

Low p_T Background Rejection BDTs

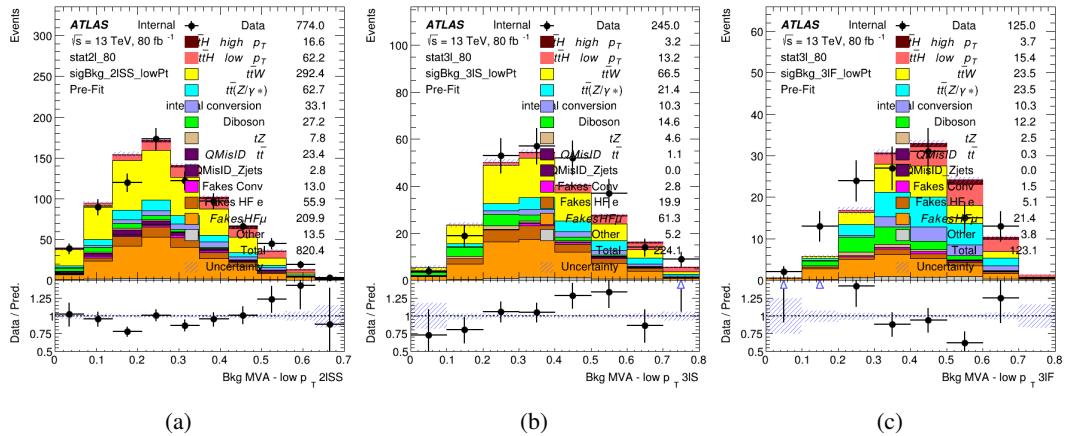


Figure 6.7: Output score of the low p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

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

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

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

519 The event preselection and MVA selection listed in Table 20 are used define the three signal
 520 regions used in the fit. These signal region definitions are summarized in Table 21.

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

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

521 7 Systematic Uncertainties

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

Table 22: Sources of systematic uncertainty considered in the analysis. Some of the systematic uncertainties are split into several components, as indicated by the number in the rightmost column.

Systematic uncertainty	Components
Luminosity	1
Pileup reweighting	1
Physics Objects	
Electron	6
Muon	15
Jet energy scale and resolution	28
Jet vertex fraction	1
Jet flavor tagging	131
E_T^{miss}	3
Total (Experimental)	186
Background Modeling	
Cross section	24
Renormalization and factorization scales	10
Parton shower and hadronization model	2
Shower tune	4
Total (Signal and background modeling)	40
Background Modeling	
Cross section	24
Renormalization and factorization scales	10
Parton shower and hadronization model	2
Shower tune	4
Total (Signal and background modeling)	40
Total (Overall)	226

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

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

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

533 The sources which contribute to the uncertainty in the jet energy scale [**jes**] are decomposed into
534 uncorrelated components and treated as independent sources in the analysis. The CategoryReduc-
535 tion model is used to account for JES uncertainties, which decomposes the uncertainties into 30
536 nuisance parameters included in the fit. The SimpleJER model is used to account for jet energy
537 resolution (JER) uncertainties, and 8 JER uncertainty components included as NPs in the fit.

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

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

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

545

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

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

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

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

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

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

546 As mentioned in Section 3.2, a normalization corrections and uncertainties on the estimates of
 547 non-prompt leptons backgrounds are derived using data driven techniques, described in detail in
 548 [ttH_paper]. These are derived from a likelihood fit over various non-prompt enriched control
 549 regions, targeting several sources of non-prompt light leptons separately: external conversion
 550 electrons, internal conversion electrons, electrons from heavy flavor decays, and muons from
 551 heavy flavor decays.

552 The normalization factor and uncertainty applied to each source of non-prompt leptons is
 553 summarized in Table 7

Processs	
NF_e^{ExtCO}	1.70 ± 0.51
NF_e^{IntCO}	0.75 ± 0.26
NF_e^{HF}	1.09 ± 0.32
NF_μ^{HF}	1.28 ± 0.17

554 In addition to those derived from the control regions, several additional uncertainties are assigned
 555 to the non-prompt lepton background. An additional 25% uncertainty on material conversions is
 556 assigned, based on the comparison between data and MC in a region where a loose electron fails
 557 the photon conversion veto. A shape uncertainty of 15% (6%) is assigned to the HF non-prompt
 558 electron (muon) background based on a comparison between data and MC where the second
 559 leading electron (muon) is only required to be loose. As the contribution from light non-prompt

leptons is small, about 10% percent of the contribution from HF non-prompt leptons, it is derived from the agreement between data and simulation in a LF enriched region at low values of the non-prompt lepton BDT. The resulting uncertainty is 100%, and is taken to be uncorrelated between internal and material conversions.

Theoretical uncertainties applied to MC predictions, including cross section, PDF, and scale uncertainties are taken from theory calculations for the predominate prompt backgrounds. Following the nominal $t\bar{t}H - ML$ analysis, a 50% uncertainty is applied to Diboson to account for the large uncertainty in estimating $VV +$ heavy flavor. The other “rare” background processes - including tZ , rare top processes, $ttWW$, WtZ , VVV , $tHjb$ and WtH - are assigned an overall 50% normalization uncertainty as well. The theory uncertainties applied to the MC estimates are summarized in Table 26.

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

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

Additional uncertainties to account for $t\bar{t}W$ mismodelling are also applied. These include a “Generator” uncertainty, based on a comparison between the nominal Sherpa 2.2.5 sample, and the formerly used aMC@NLO sample, and an “Extra radiation” uncertainty, which includes renormalisation and factorisation scale variations of the Sherpa 2.2.5 sample.

8 Results

A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p_T spectrum in the three signal regions, 2lSS, 3lS, and 3lF. The signal is split into high and low p_T samples, based on whether the truth p_T of the Higgs is above or below 150 GeV. The parameters $\mu_{t\bar{t}Hhighp_T}$ and $\mu_{t\bar{t}Hlowp_T}$, where $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$, are extracted from the fit, signifying the difference between the observed value and the theory prediction. Unblinded results are shown for the 79.8 fb^{-1} data set, as well as MC only projections of results using the full Run-2, 139 fb^{-1} dataset.

582 As described in Section 7, there are 229 systematic uncertainties that are considered as NPs in the
 583 fit. These NP s are constrained by Gaussian or log-normal probability density functions. The
 584 latter are used for normalisation factors to ensure that they are always positive. The expected
 585 number of signal and background events are functions of the likelihood. The prior for each NP is
 586 added as a penalty term, decreasing the likelihood as it is shifted away from its nominal value.

587 8.1 Results - 79.8 fb^{-1}

588 As the data collected from 2015-2017 has been unblinded for $t\bar{t}H$ –ML channels, representing 79.8
 589 fb^{-1} , those events are unblinded. The predicted Higgs p_T spectrum is fit to data simultaneously
 590 in each of the three signal regions shown in Figure 8.1.

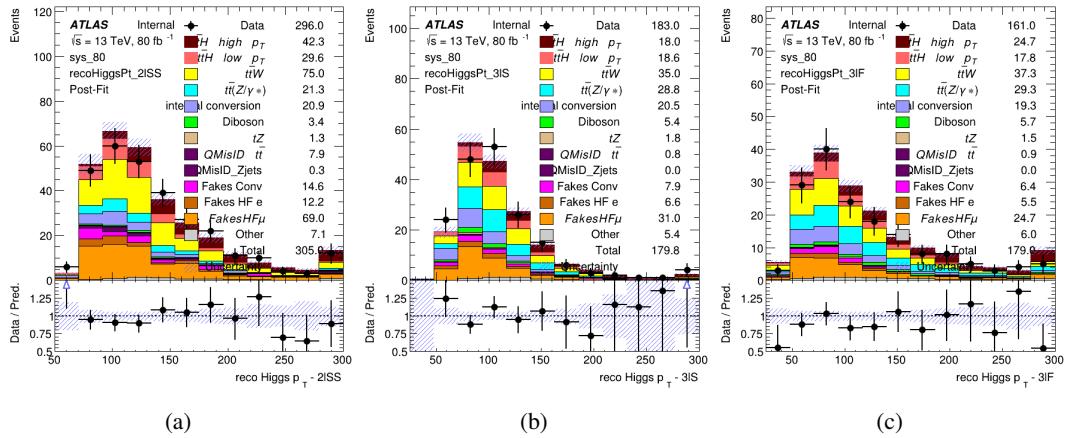


Figure 8.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2ISS, (b) 3IS, and (c) 3IF, for 79.8 fb^{-1} of MC

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

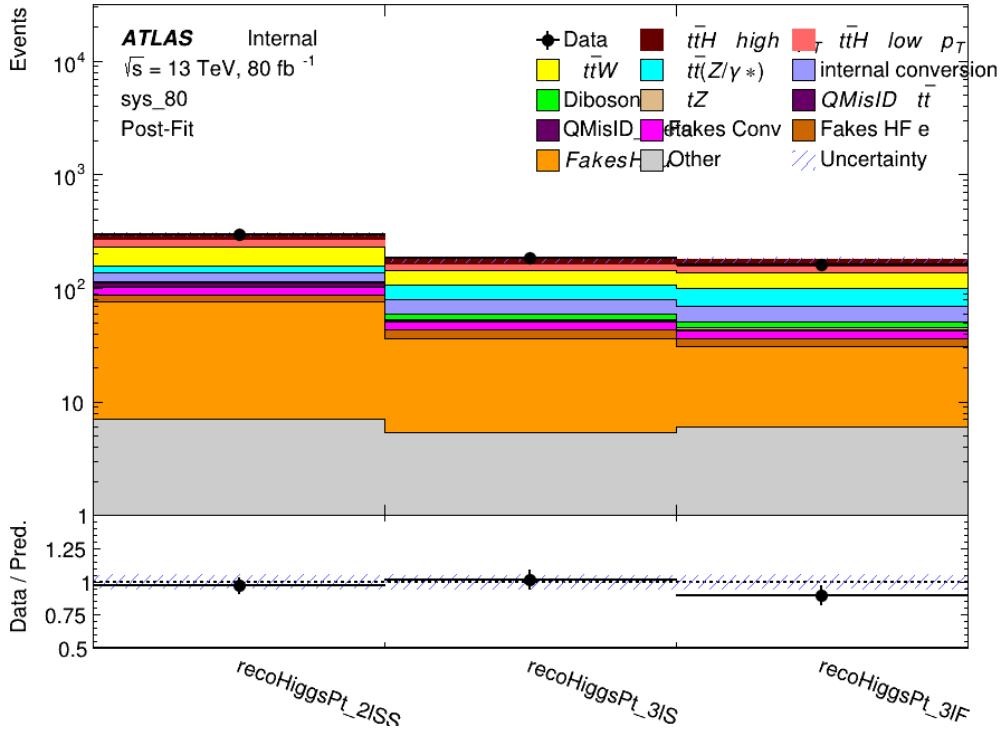


Figure 8.2: Post-fit summary of the yields in each signal region.

592 The measured μ values for high and low p_T Higgs production obtained from the fit are
 593 shown in 27. A significance of 1.7σ is observed for $t\bar{t}H$ high p_T , and 2.1σ is measured for
 594 $t\bar{t}H$ low p_T .

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

$$\mu_{t\bar{t}H \text{ low } p_T} = 0.83^{+0.39}_{-0.40} (\text{stat})^{+0.51}_{-0.53} (\text{sys})$$

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

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

Uncertainty Source	$\Delta\mu$	
Jet Energy Scale	0.25	0.23
t̄H cross-section (QCD scale)	-0.11	0.21
Luminosity	-0.13	0.14
Flavor Tagging	0.14	0.13
t̄W cross-section (QCD scale)	-0.12	0.11
Higgs Branching Ratio	-0.1	0.11
t̄H cross-section (PDF)	-0.07	0.08
Electron ID	-0.06	0.06
Non-prompt Muon Normalization	-0.05	0.06
t̄Z cross-section (QCD scale)	-0.05	0.05
Diboson cross-section	-0.05	0.05
Fake muon modelling	-0.04	0.04
Total	0.40	0.43

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

597 The most significant sources of uncertainty on the measurement of t̄H - low p_T are shown in
 598 Table 29.

Uncertainty Source	$\Delta\mu$	
Internal Conversions	-0.26	0.26
Luminosity	-0.16	0.17
Non-prompt Muon Normalization	-0.16	0.16
t̄W cross-section (QCD scale)	-0.17	0.15
Jet Energy Scal	0.15	0.15
Non-prompt Electron Modelling	-0.13	0.14
Flavor Tagging	0.13	0.13
Non-prompt Muon Modelling	-0.12	0.13
Non-prompt Electron Normalization	-0.11	0.11
t̄Z cross-section (QCD scale)	-0.08	0.09
Diboson Cross-section	-0.07	0.07
Total	0.51	0.53

Table 29: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H low p_T.

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

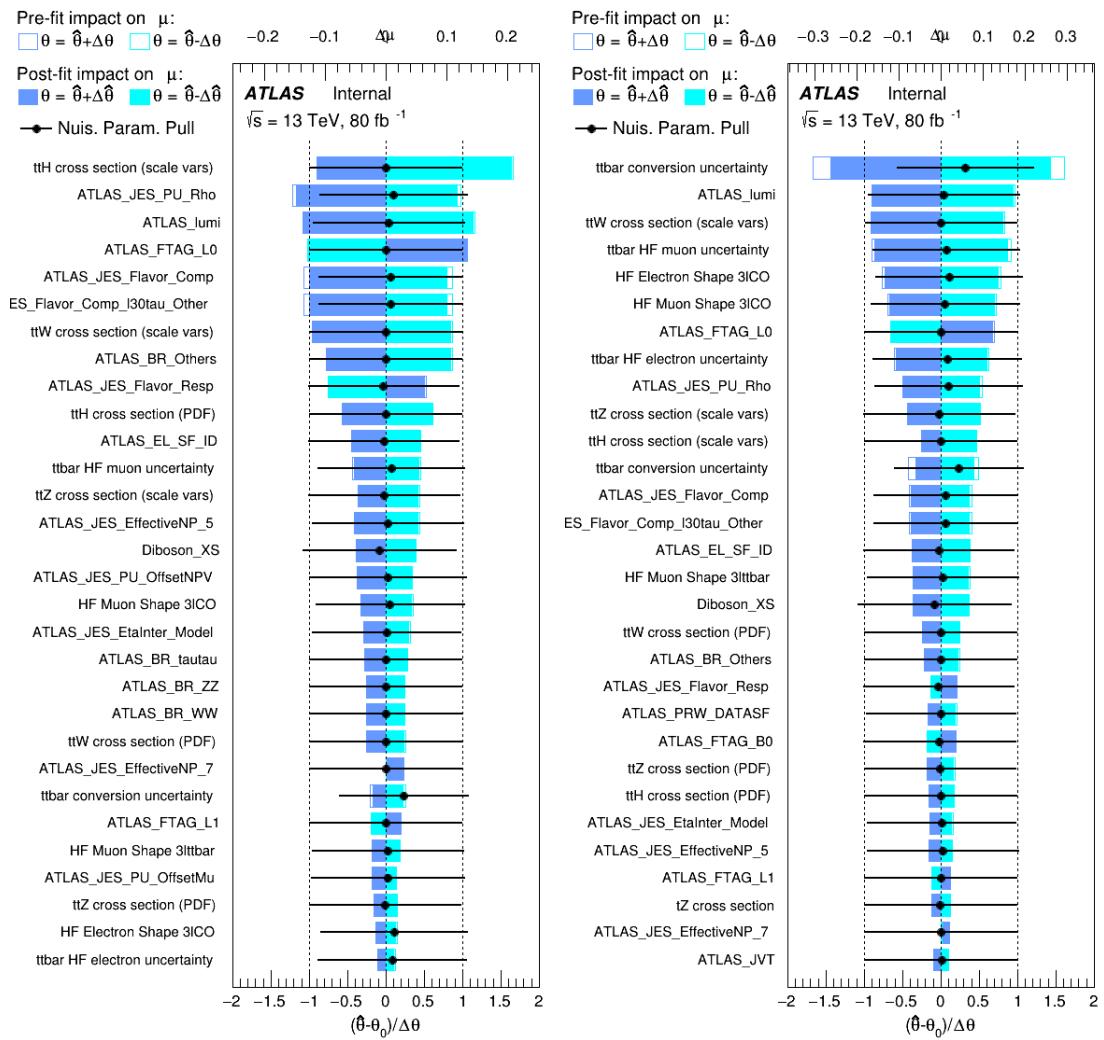


Figure 8.3: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

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

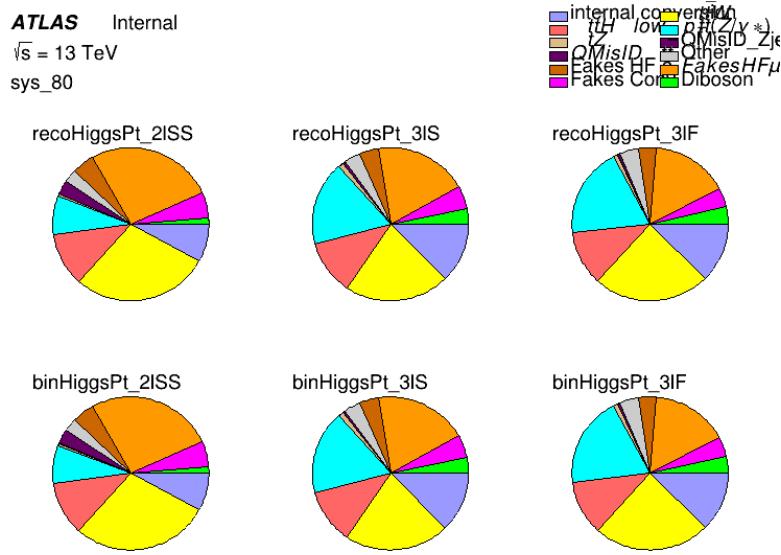
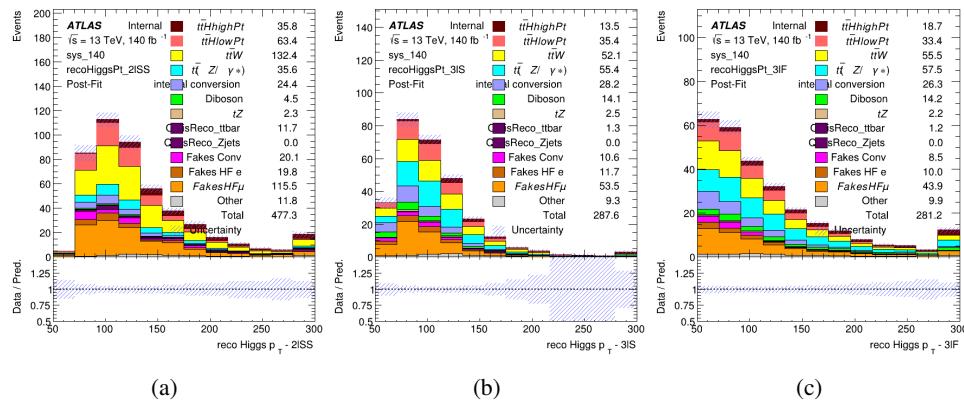


Figure 8.4: Background composition of the fit regions.

602 8.2 Projected Results - 139 fb^{-1}

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

Figure 8.5: Blinded post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 139 fb^{-1} of data

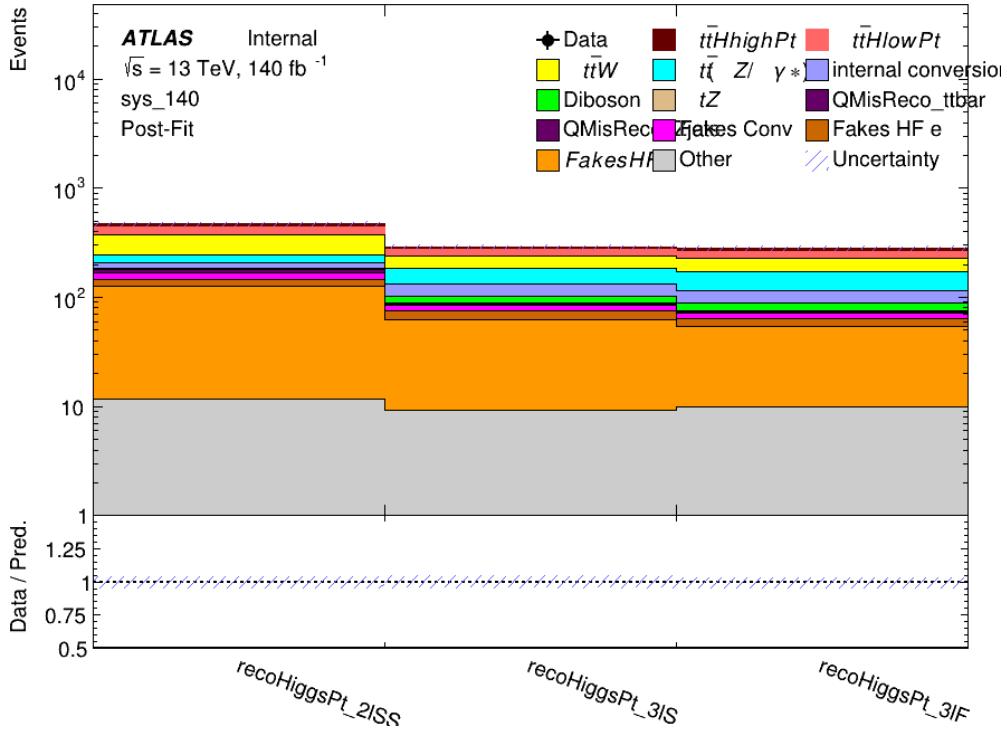


Figure 8.6: Post-fit summary of fit.

607 Projected uncertainties on the μ values extracted from the fit for high and low p_T Higgs are shown
 608 in 30. A significance of 2.0σ is expected for $t\bar{t}H$ high p_T , and a projected significance 2.3σ is
 609 extracted for $t\bar{t}H$ low p_T .

$$\mu_{t\bar{t}H \text{ high } p_T} = 1.00^{+0.45}_{-0.43} (\text{stat})^{+0.30}_{-0.31} (\text{sys})$$

$$\mu_{t\bar{t}H \text{ low } p_T} = 1.00^{+0.29}_{-0.30} (\text{stat})^{+0.48}_{-0.50} (\text{sys})$$

Table 30: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

610 The most prominent sources of systematic uncertainty, as measured by their impact on $\mu_{t\bar{t}H \text{ high } p_T}$,
 611 are summarized in Table 31.

Uncertainty Source	$\Delta\mu$	
Jet Energy Scale	0.19	0.17
t̄W Cross-section (QCD Scale)	-0.12	0.11
Luminosity	-0.1	0.11
Flavor Tagging	0.1	0.1
t̄H Cross-section (QCD Scale)	-0.05	0.1
t̄Z Cross-section (QCD Scale)	-0.05	0.06
Non-prompt Muon Normalization	-0.05	0.05
Higgs Branching Ratio	-0.05	0.05
Diboson Cross-section	-0.04	0.05
Non-prompt Muon Modelling	-0.04	0.04
t̄H Cross-section (PDF)	-0.03	0.04
Electron ID	-0.04	0.04
t̄W Cross-section (PDF)	-0.03	0.03
Total	0.30	0.31

Table 31: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H high p_T.

⁶¹² The most significant sources of systematic uncertainty on t̄H low p_T are summarized in Table
⁶¹³ 32.

Uncertainty Source	$\Delta\mu$	
Internal Conversions	-0.18	0.2
Jet Energy Scale	0.19	0.16
Non-prompt Muon Normalization	-0.16	0.17
Luminosity	-0.15	0.17
t̄W Cross-section (QCD Scale)	-0.17	0.15
Non-prompt Electron Modelling	-0.13	0.14
Non-prompt Muon Modelling	-0.13	0.13
Flavor Tagging	0.13	0.12
Non-prompt Electron Normalization	-0.1	0.11
t̄Z Cross-section (QCD Scale)	-0.07	0.09
t̄H Cross-section (QCD Scale)	-0.05	0.1
Total	0.48	0.50

Table 32: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H low p_T.

⁶¹⁴ The ranking and impact of those nuisance parameters with the largest contribution to the overall
⁶¹⁵ uncertainty is shown in Figure 8.7.

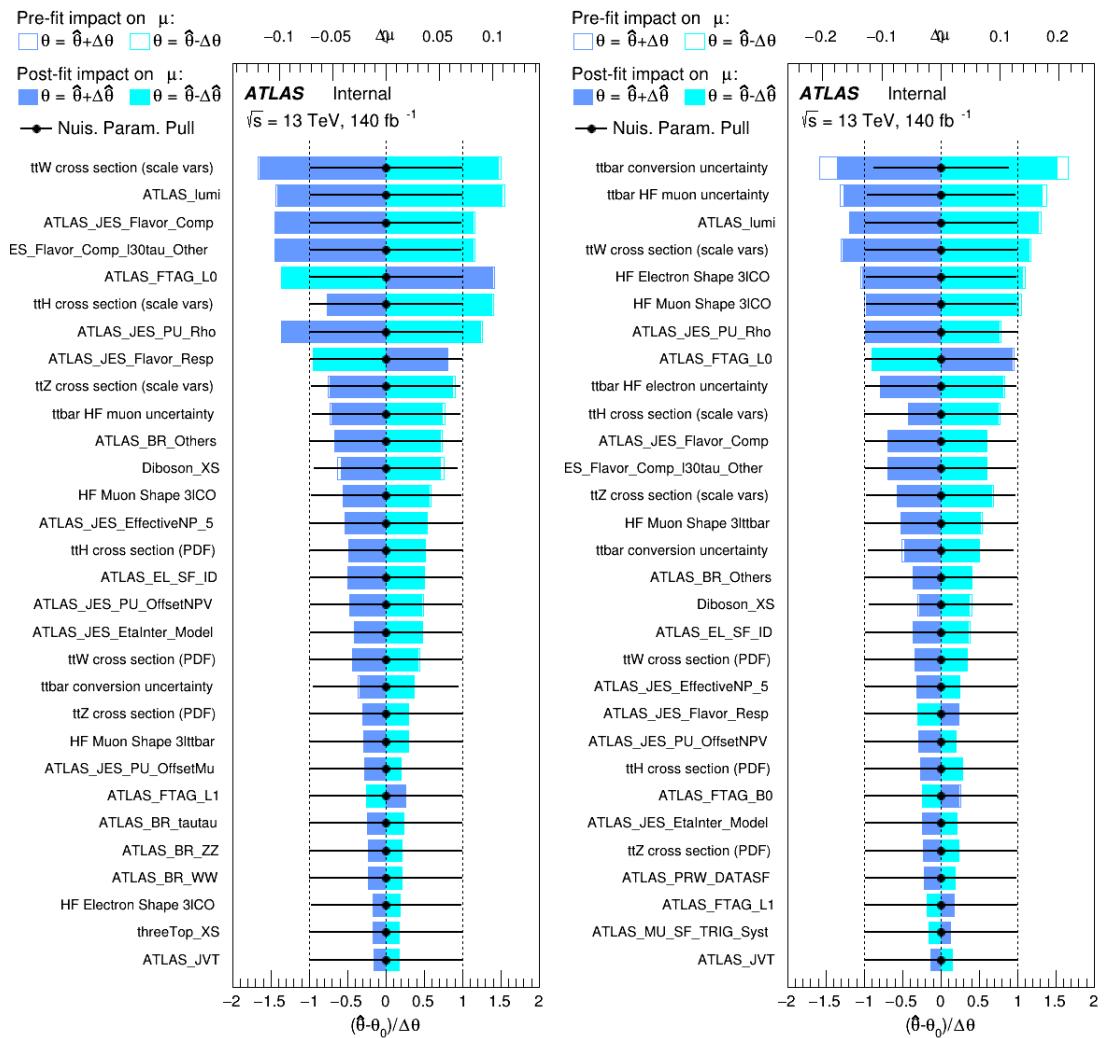


Figure 8.7: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

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

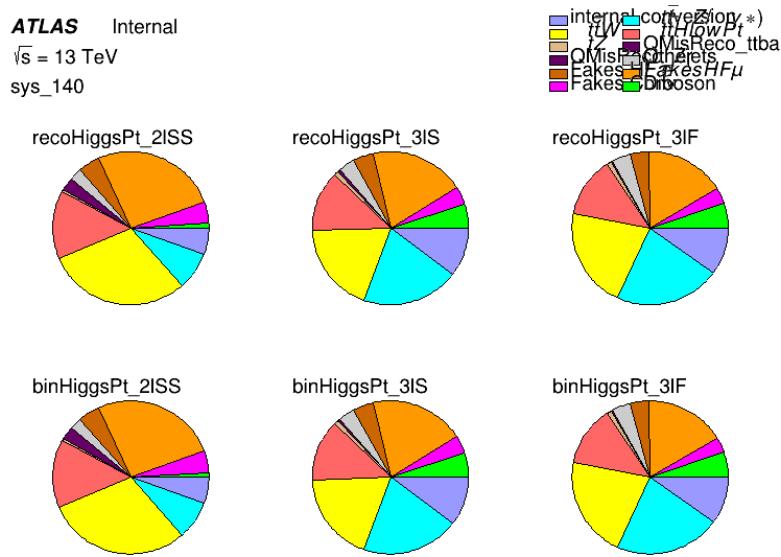


Figure 8.8: Background composition of the fit regions.

617 9 Conclusion

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

622 **List of contributions**

623

624 Appendices

625 .1 Non-prompt lepton MVA

626 A lepton MVA has been developed to better reject non-prompt leptons than standard cut
 627 based selections based upon impact parameter, isolation and PID. The name of this MVA is
 628 **PromptLeptonIso**. The full set of studies and detailed explanation can be found in [**ttW_140**].

629 The decays of W and Z bosons are commonly selected by the identification of one or two electrons
 630 or muons. The negligible lifetimes of these bosons mean that the leptons produced in the decay
 631 originate from the interaction vertex and are thus labelled “prompt”. Analyses using these light
 632 leptons impose strict reconstruction quality, isolation and impact parameter requirements to
 633 remove “fake” leptons. A significant source of the fake light leptons are non-prompt leptons
 634 produced in decays of hadrons that contain bottom (b) or charm (c) quarks. Such hadrons typically
 635 have microscopically significant lifetimes that can be detected experimentally.

636 These non-prompt leptons can also pass the tight selection criteria. In analyses that involve top (t)
 637 quarks, which decay almost exclusively into a W boson and a b quark, non-prompt leptons from
 638 the semileptonic decay of bottom and charm hadrons can be a significant source of background
 639 events. This is particularly the case in the selection of same-sign dilepton and multilepton final
 640 states.

641 The main idea is to identify non-prompt light leptons using lifetime information associated with a
 642 track jet that matches the selected light lepton. This lifetime information is computed using tracks
 643 contained within the jet. Typically, lepton lifetime is determined using the impact parameter of
 644 the track reconstructed by the inner tracking detector which is matched to the reconstructed lepton.
 645 Using additional reconstructed charged particle tracks increases the precision of identifying the
 646 displaced decay vertex of bottom or charm hadrons that produced a non-prompt light lepton.
 647 The MVA also includes information related to the isolation of the lepton to reject non-prompt
 648 leptons.

649 **PromptLeptonIso** is a gradient boosted BDT. The training of the BDT is performed on leptons
 650 selected from the Powheg+Pythia6 non-allhad $t\bar{t}$ MC sample. Eight variables are used to train
 651 the BDT in order to discriminate between prompt and non-prompt leptons. The track jets that
 652 are matched to the non-prompt leptons correspond to jets initiated by b or c quarks, and may
 653 contain a displaced vertex. Consequently, three of the selected variables are used to identify
 654 b-tag jets by standard ATLAS flavour tagging algorithms. Two variables use the relationship
 655 between the track jet and lepton: the ratio of the lepton p_T with respect to the track jet p_T and
 656 ΔR between the lepton and the track jet axis. Finally three additional variables test whether the
 657 reconstructed lepton is isolated: the number of tracks collected by the track jet and the lepton
 658 track and calorimeter isolation variables. Table 33 describes the variables used to train the BDT
 659 algorithm. The choice of input variables has been extensively discussed with Egamma, Muon,
 660 Tracking, and Flavour Tagging CP groups.

661 The output distribution of the BDT is shown in Figure .1.

Variable	Description
N_{track} in track jet	Number of tracks collected by the track jet
$\text{IP2 log}(P_b/P_{\text{light}})$	Log-likelihood ratio between the b and light jet hypotheses with the IP2D algorithm
$\text{IP3 log}(P_b/P_{\text{light}})$	Log-likelihood ratio between the b and light jet hypotheses with the IP3D algorithm
N_{TrkAtVtx} SV + JF	Number of tracks used in the secondary vertex found by the SV1 algorithm
$p_T^{\text{lepton}}/p_T^{\text{track jet}}$	in addition to the number of tracks from secondary vertices found by the JetFitter algorithm with at least two tracks The ratio of the lepton p_T and the track jet p_T
$\Delta R(\text{lepton}, \text{track jet})$	ΔR between the lepton and the track jet axis
$p_T^{\text{VarCone30}}/p_T$	Lepton track isolation, with track collecting radius of $\Delta R < 0.3$
$E_T^{\text{TopoCone30}}/p_T$	Lepton calorimeter isolation, with topological cluster collecting radius of $\Delta R < 0.3$

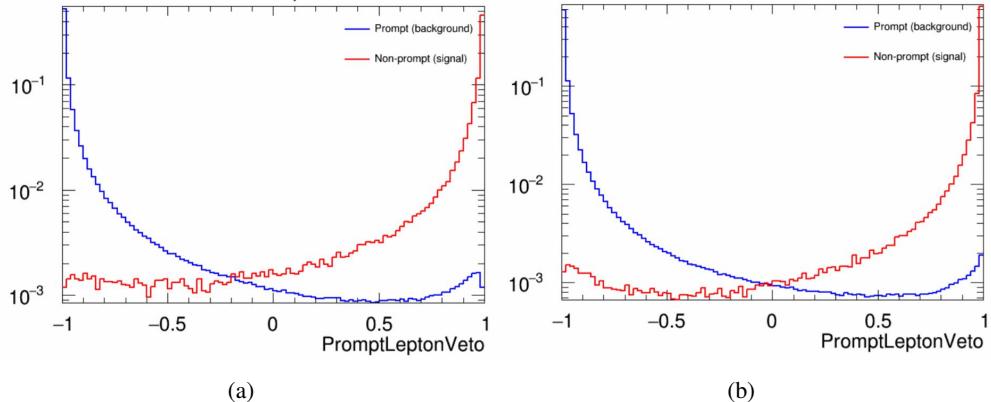
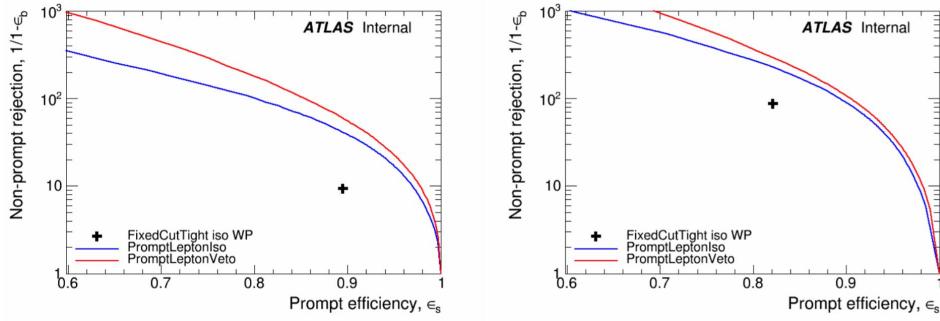
Table 33: A table of the variables used in the training of `PromptLeptonIso`.

Figure .1: Distribution of the PLV BDT discriminant for (a) electrons and (b) muons

662 The ROC curve for the BDT response, compared to the standard `FixedCutTight` WP, is shown
 663 in figure .1, which shows a clear improvement when using this alternate training.

Figure .2: ROC curves for the PLV as well as the performance of the standard `FixedCutTight` WP for (left) electrons and (right) muons

664 A cutoff value of -0.7 for electrons and -0.5 for muons are chosen as the WPs for this MVA, based
665 on an optimisation of S/\sqrt{B} performed in the preselection regions of the $t\bar{t}H$ – ML analysis,
666 which have a signature similar to that of this analysis.

667 The efficiency of the tight PromptLeptonIso working point is measured using the tag and probe
668 method with $Z \rightarrow \ell^+\ell^-$ events. Such calibration are performed by analysers from this analysis in
669 communication with the Egamma and Muon combined performance groups. The scale factor are
670 approximately 0.92 for $10 < p_T < 15$ GeV, and averaging at 0.98 to 0.99 for higher p_T leptons.
671 An extra systematic is applied to muons within $\Delta R < 0.6$ of a calorimeter jet, since there is a
672 strong dependence on the scale factor due to the presence of these jets. For electrons, the dominant
673 systematics is coming from pile-up dependence. Overall the systematics are a maximum of 3% at
674 low p_T and decreasing at a function of p_T .

675 **A Machine Learning Models**

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

678 **A.1 Higgs Reconstruction Model Details**

679 **A.1.1 b-jet Identification Features - 2ISS**

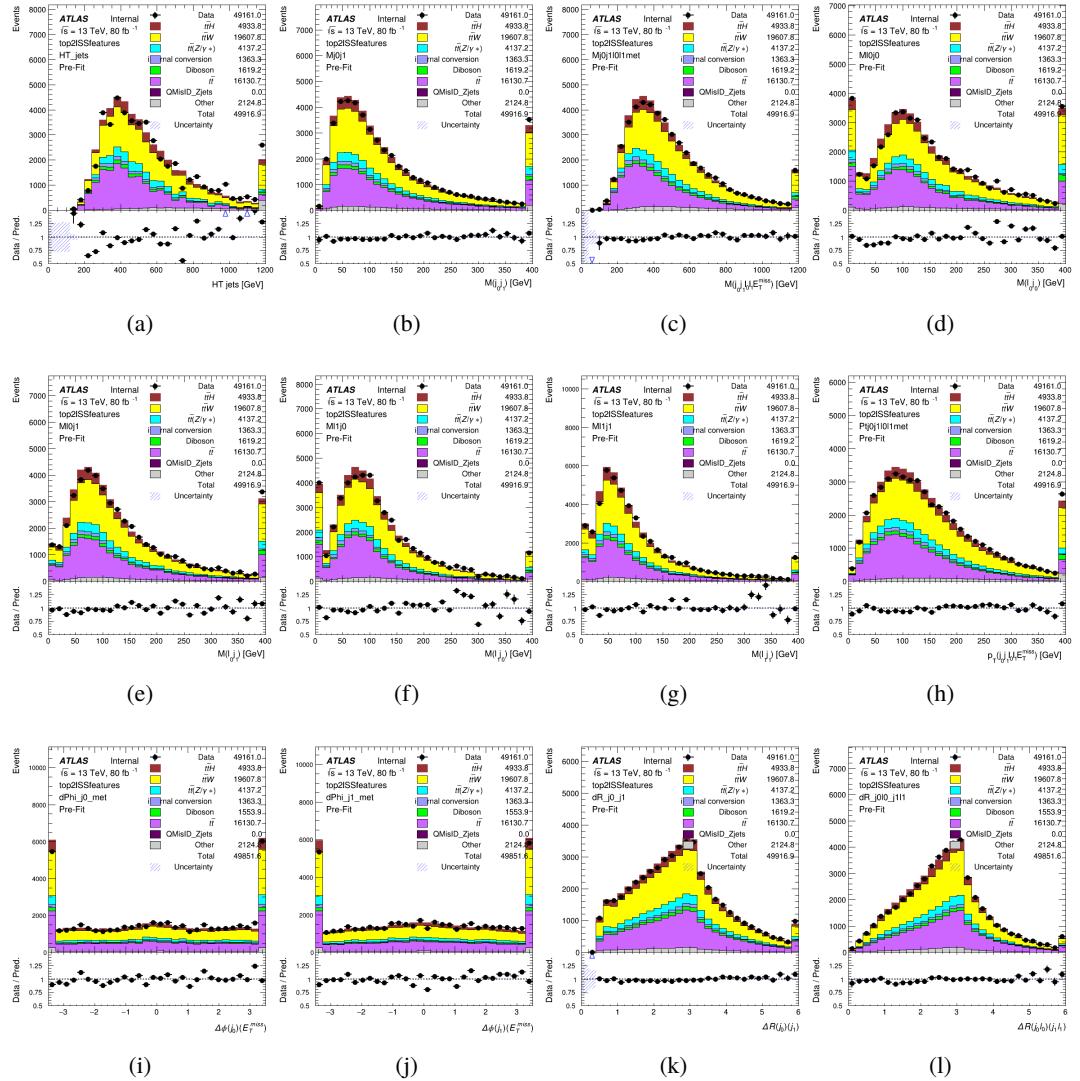


Figure A.1: Input features for top2lSS

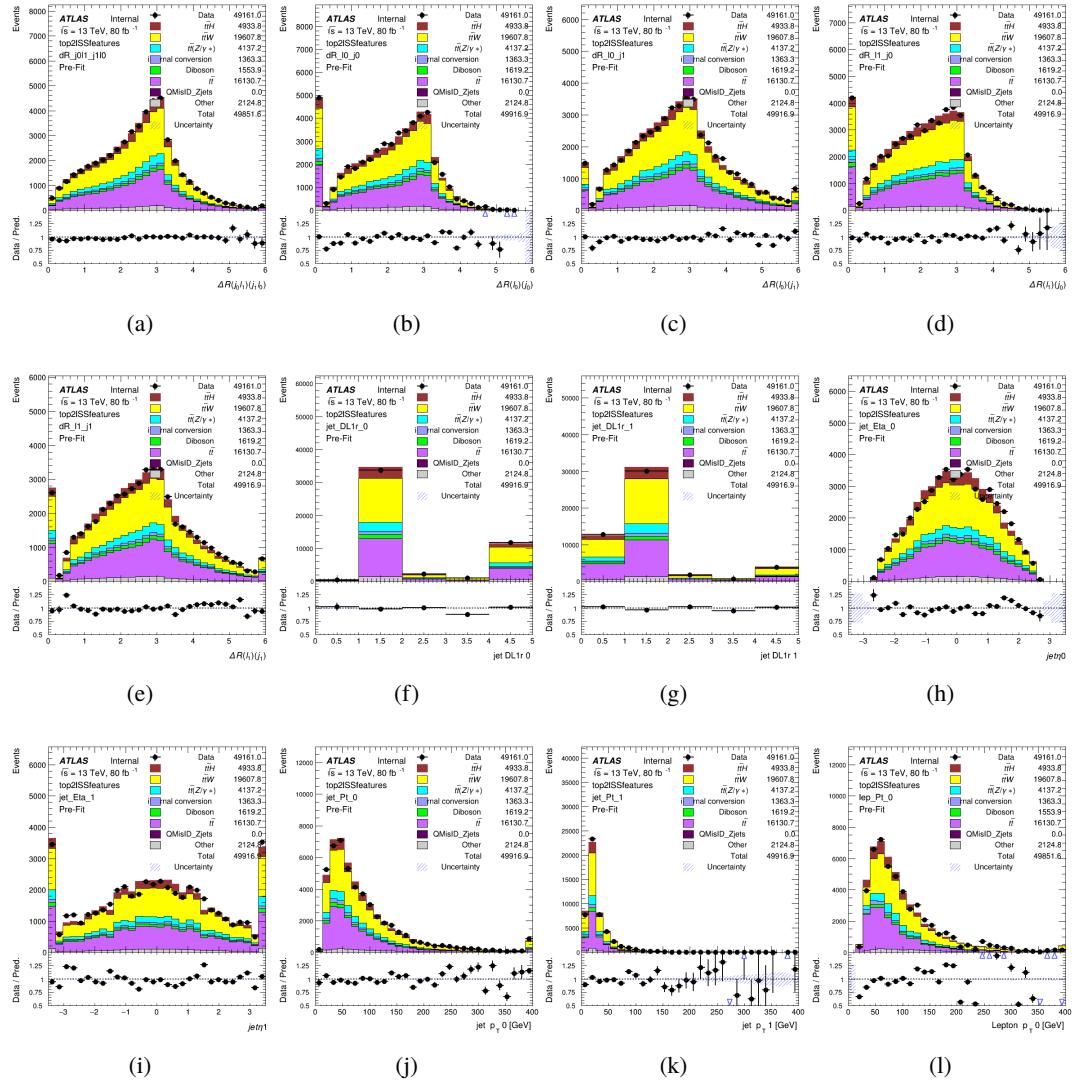


Figure A.2: Input features for top2lSS

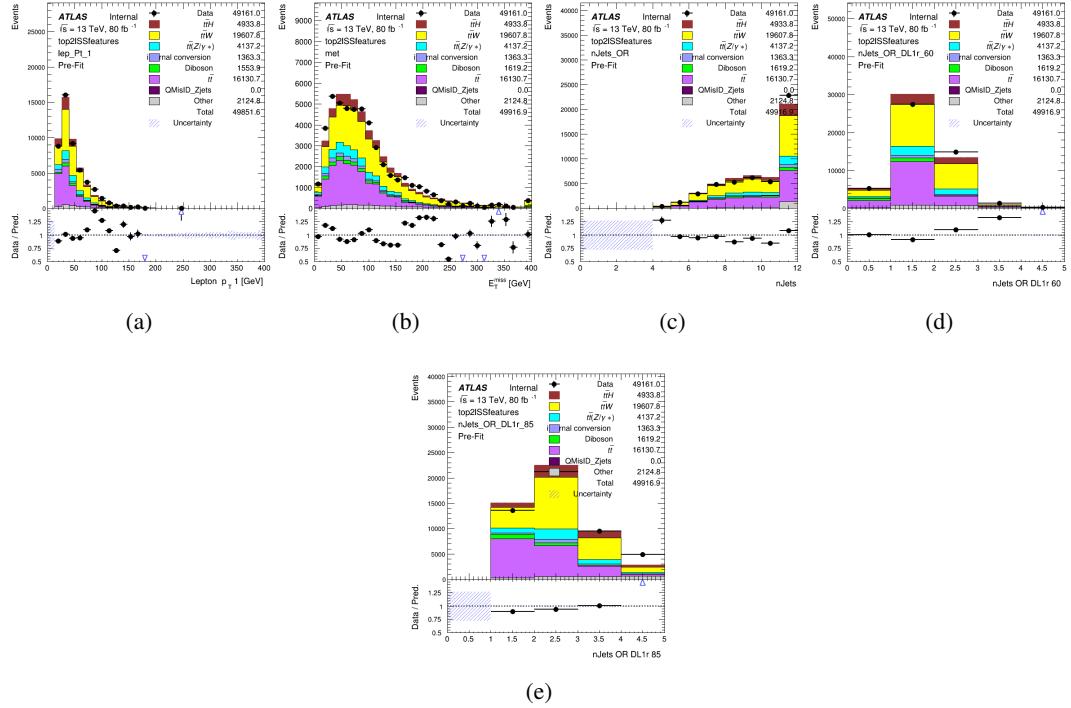


Figure A.3: Input features for top2ISS

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

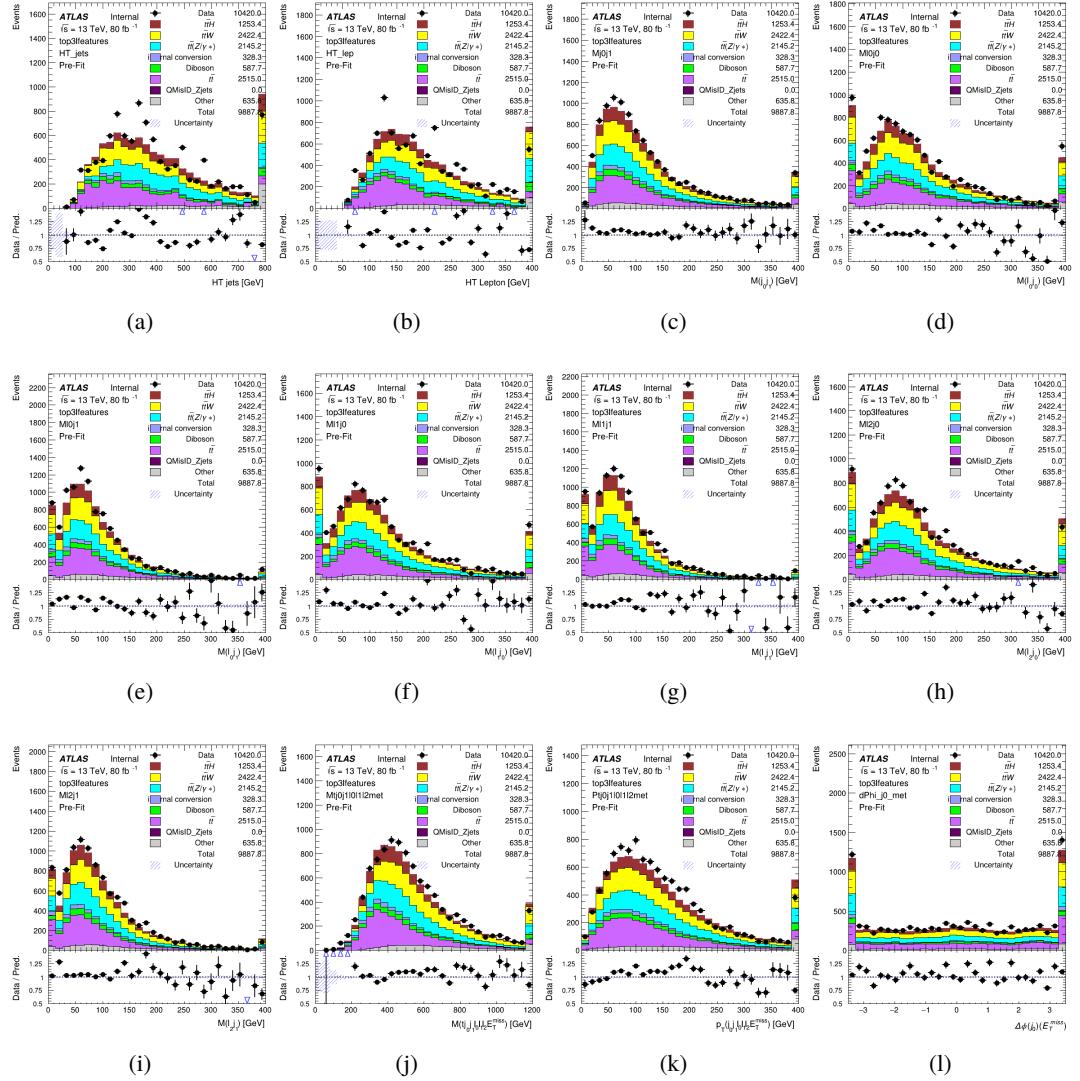


Figure A.4: Input features for top31

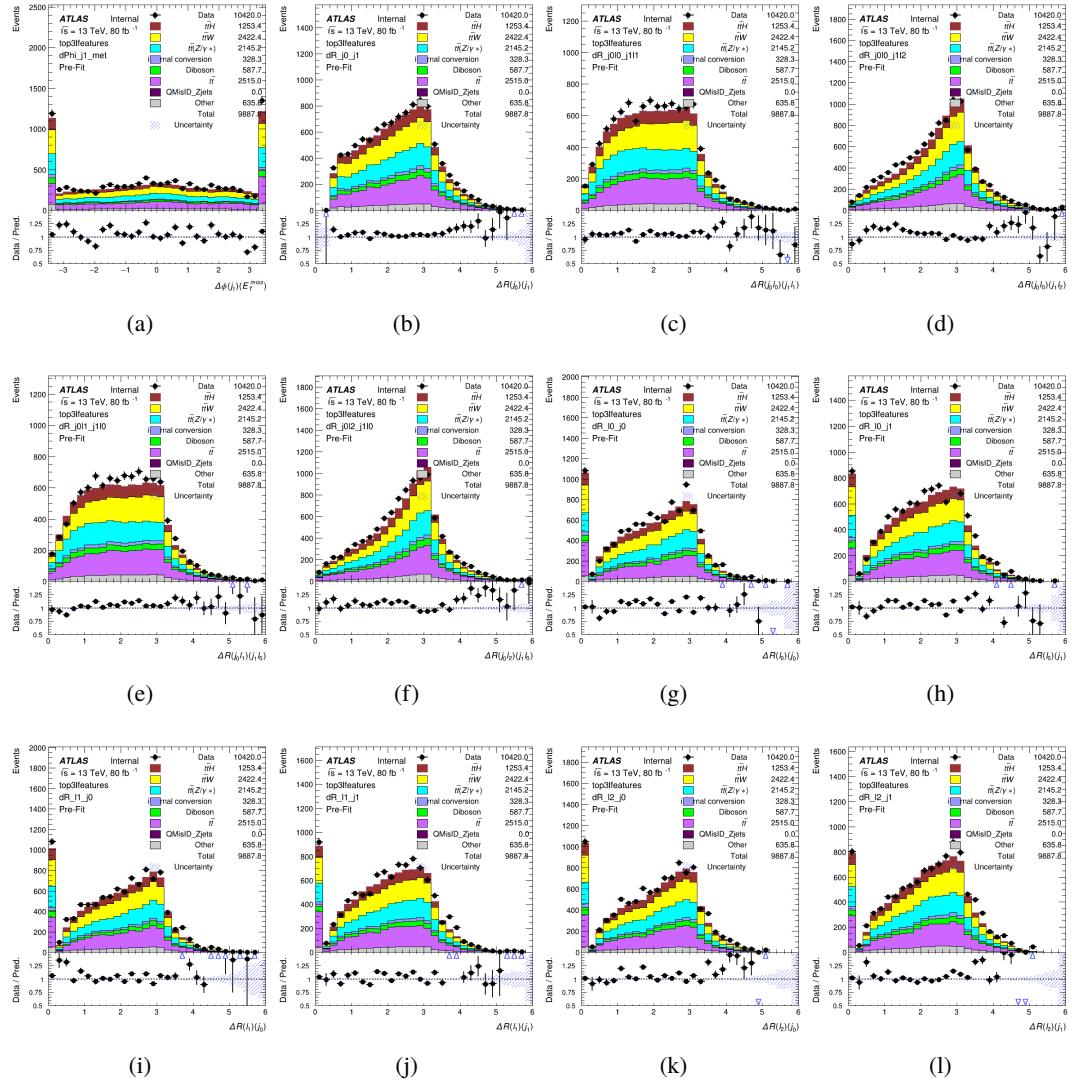


Figure A.5: Input features for top31

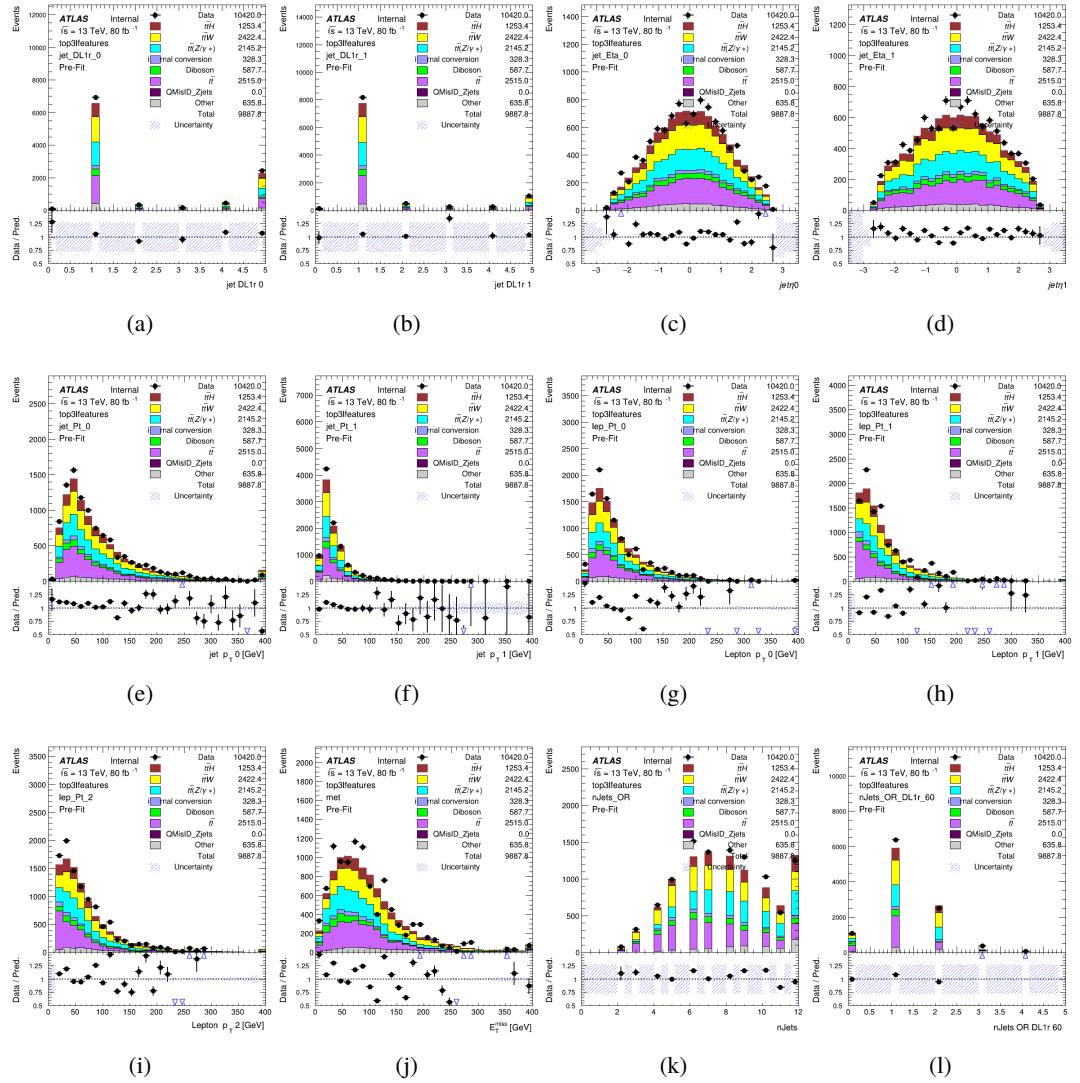
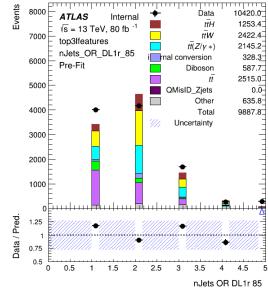


Figure A.6: Input features for top31



(a)

Figure A.7: Input features for top31

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

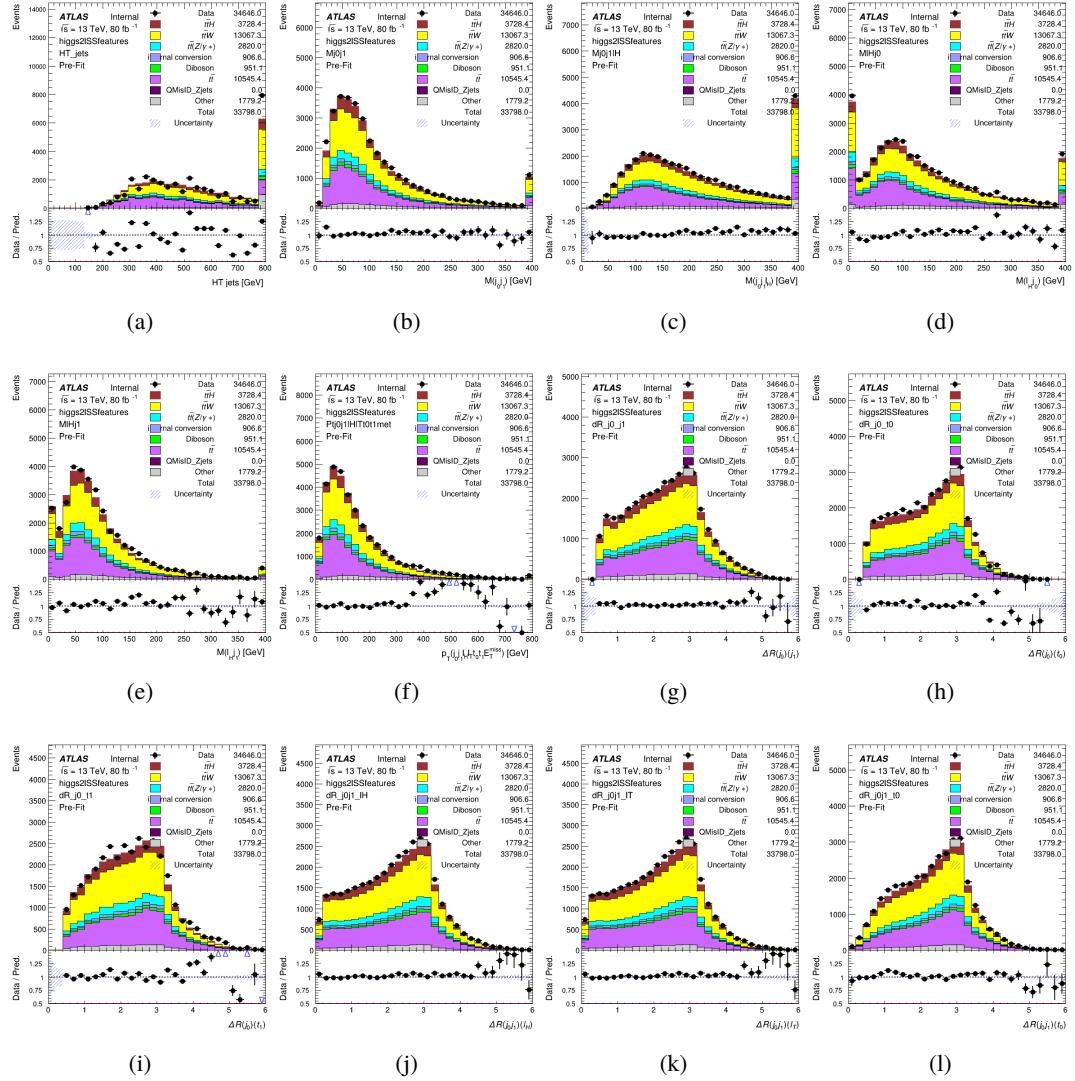


Figure A.8: Input features for higgs2lSS

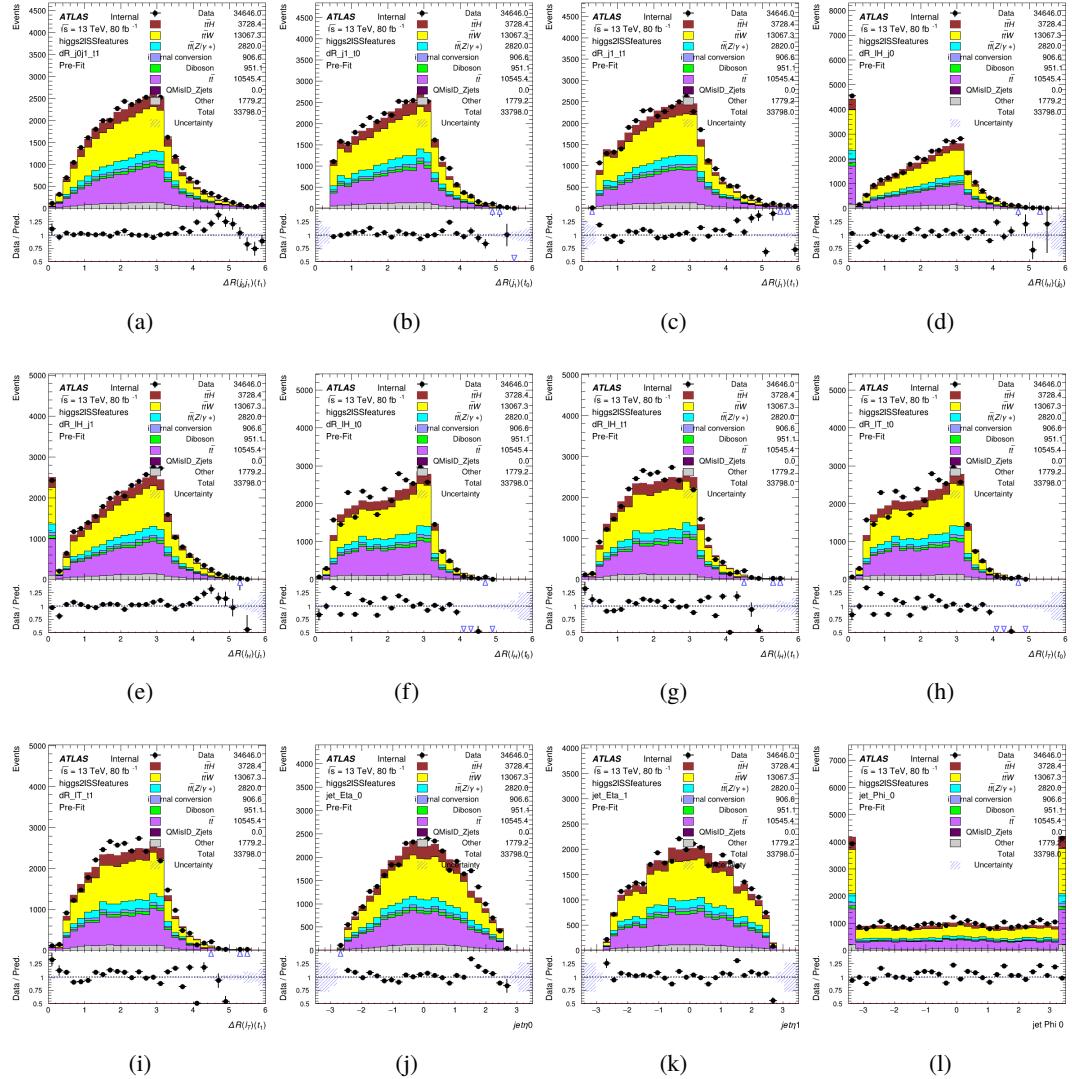


Figure A.9: Input features for higgs2lSS

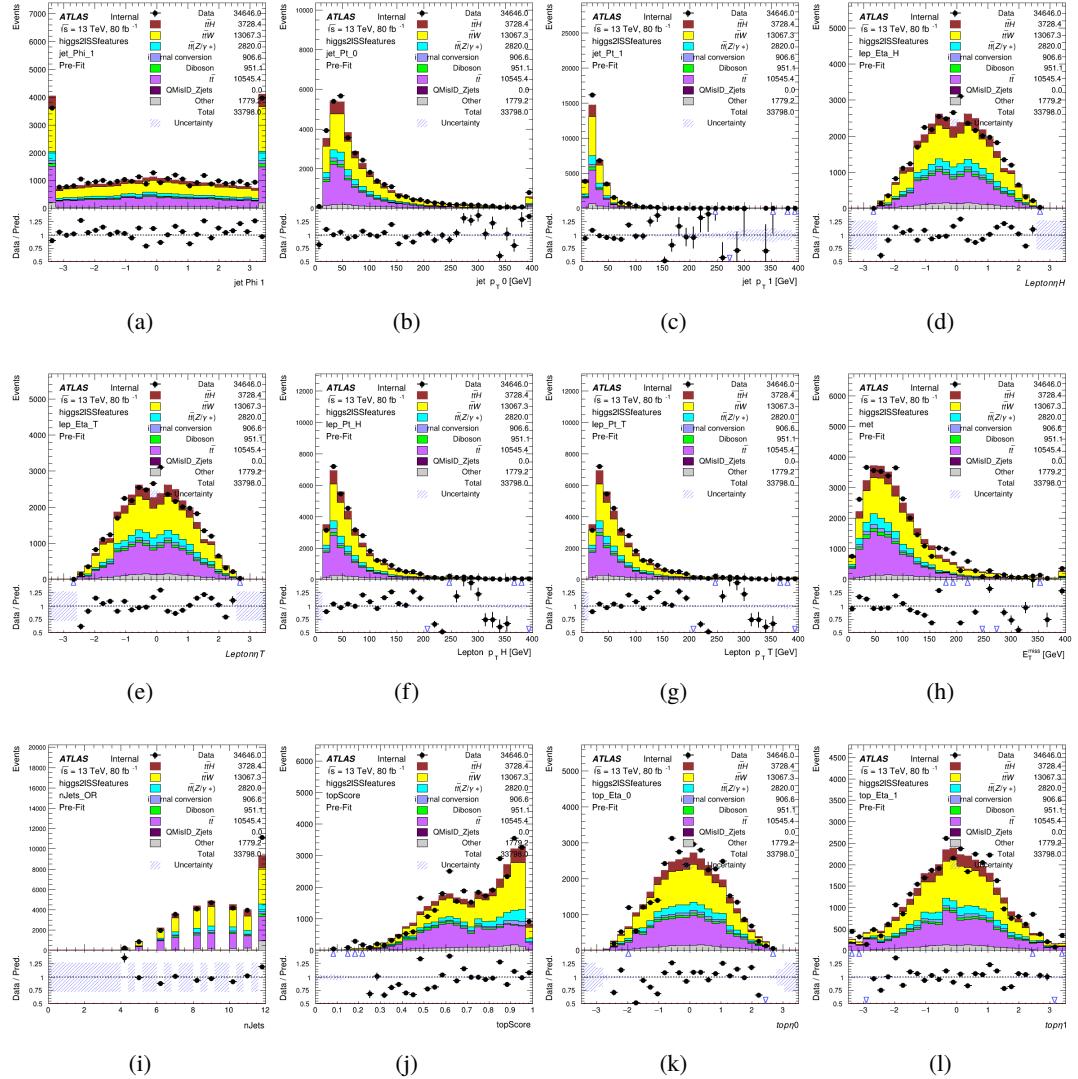


Figure A.10: Input features for higgs2IS

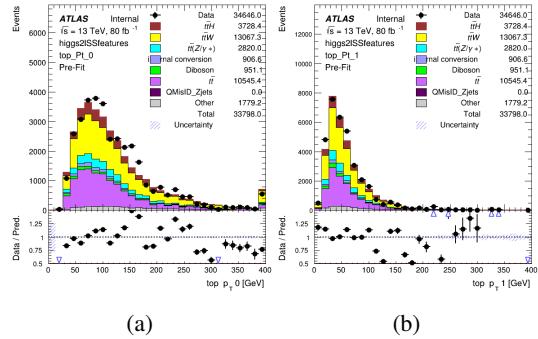


Figure A.11: Input features for higgs2lSS

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

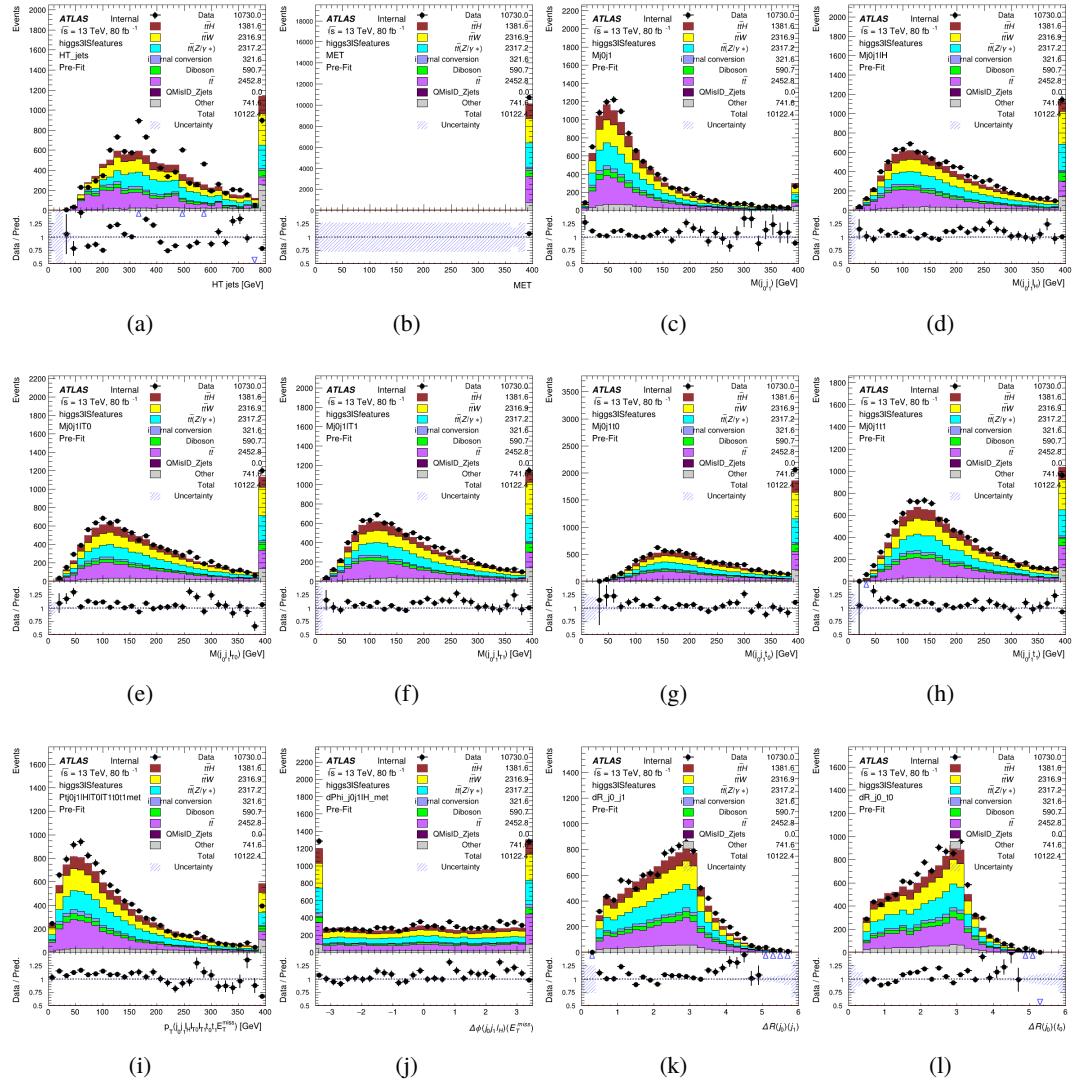


Figure A.12: Input features for higgs3SIS

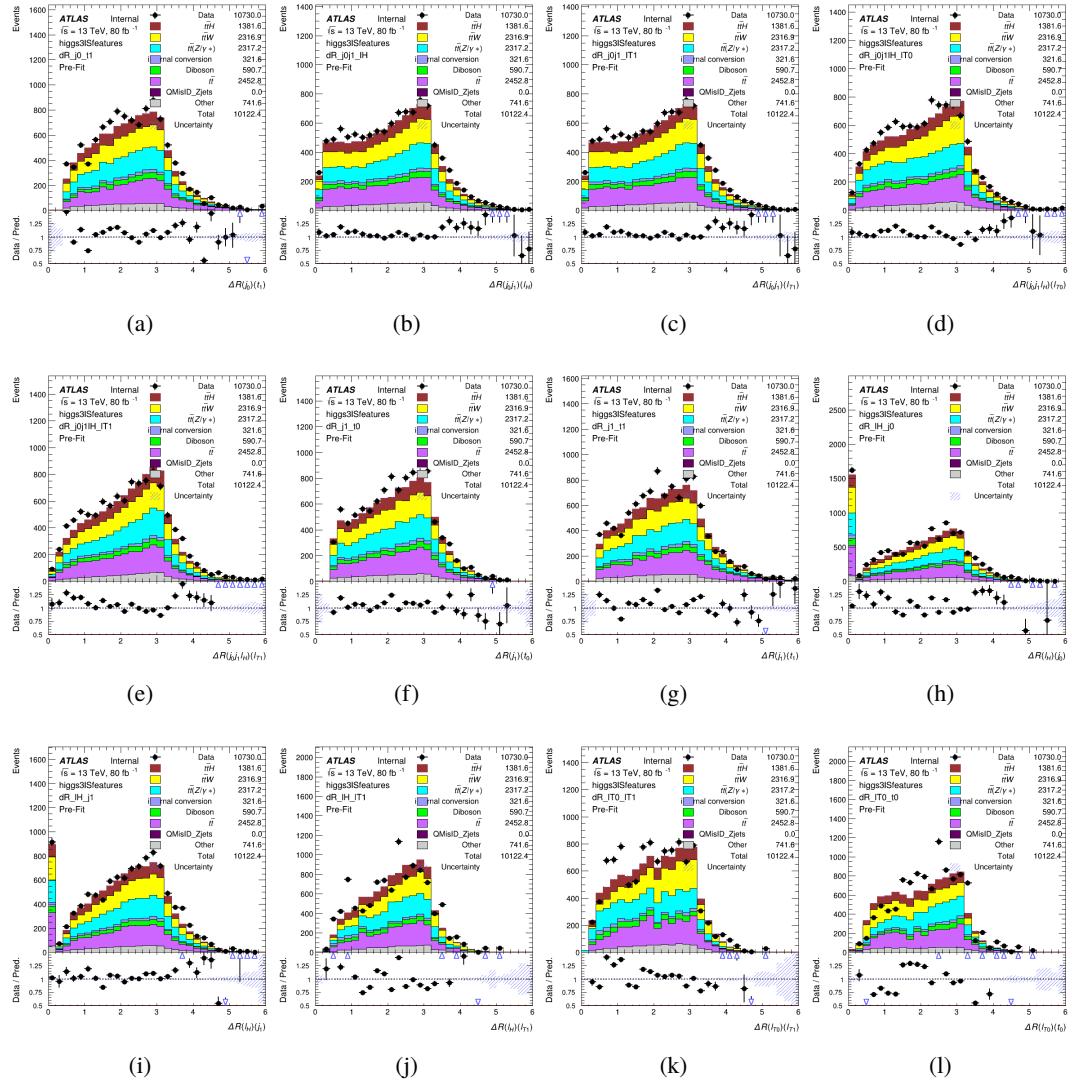


Figure A.13: Input features for higgs3S1S

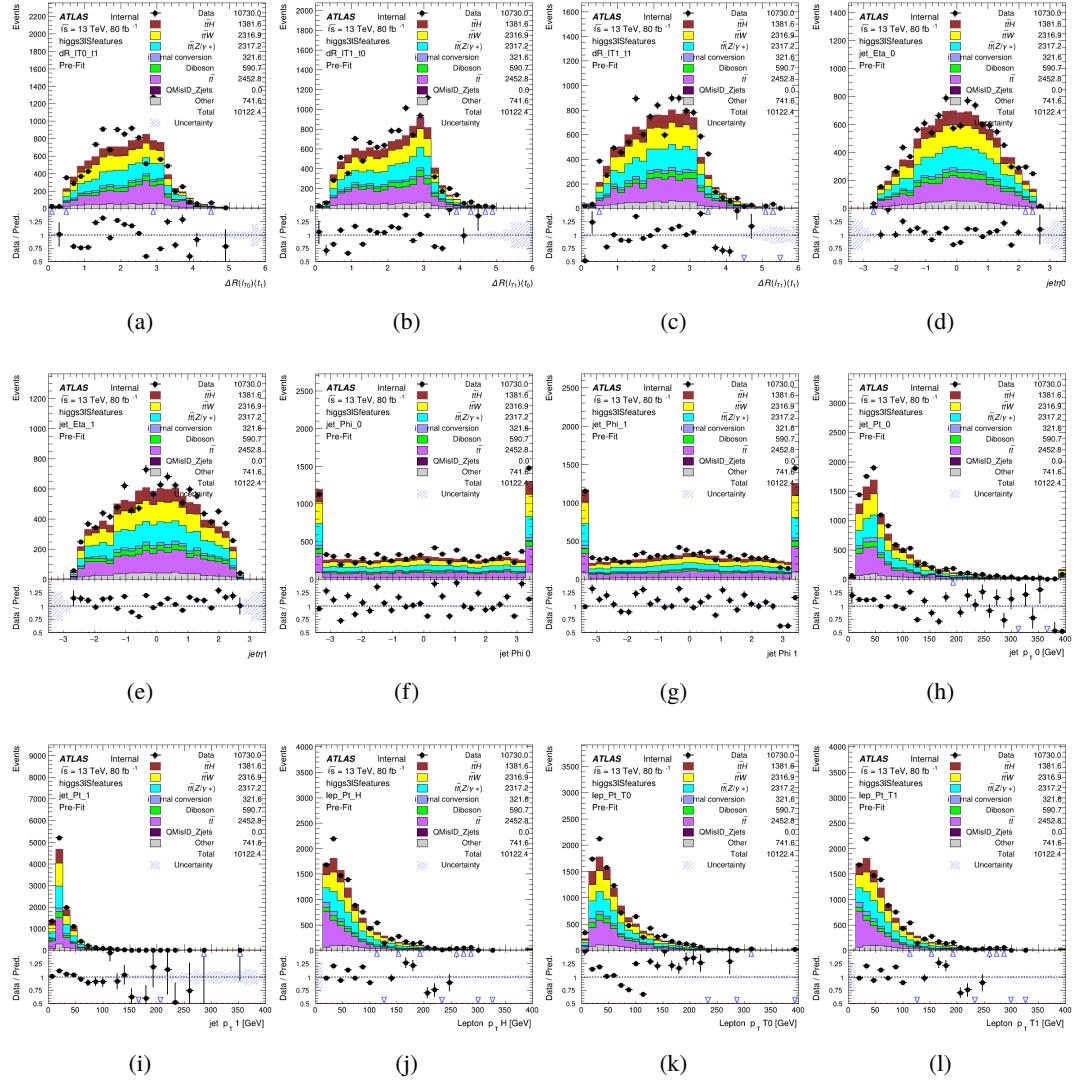


Figure A.14: Input features for higgs3IS

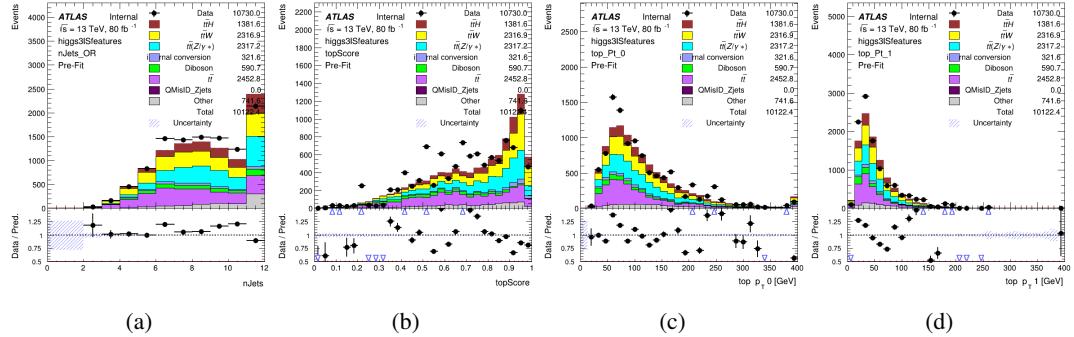


Figure A.15: Input features for higgs3IS

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

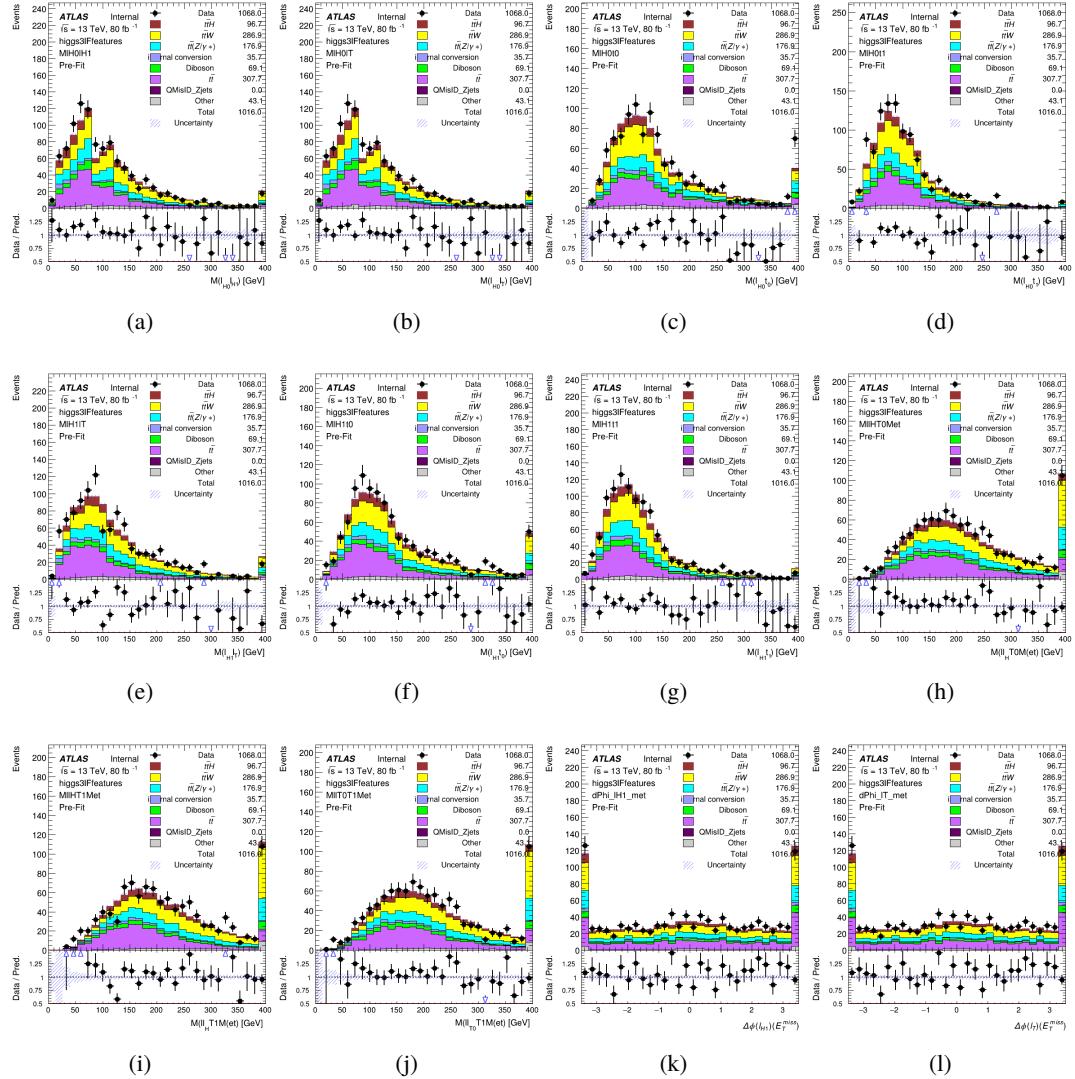


Figure A.16: Input features for higgs3IF

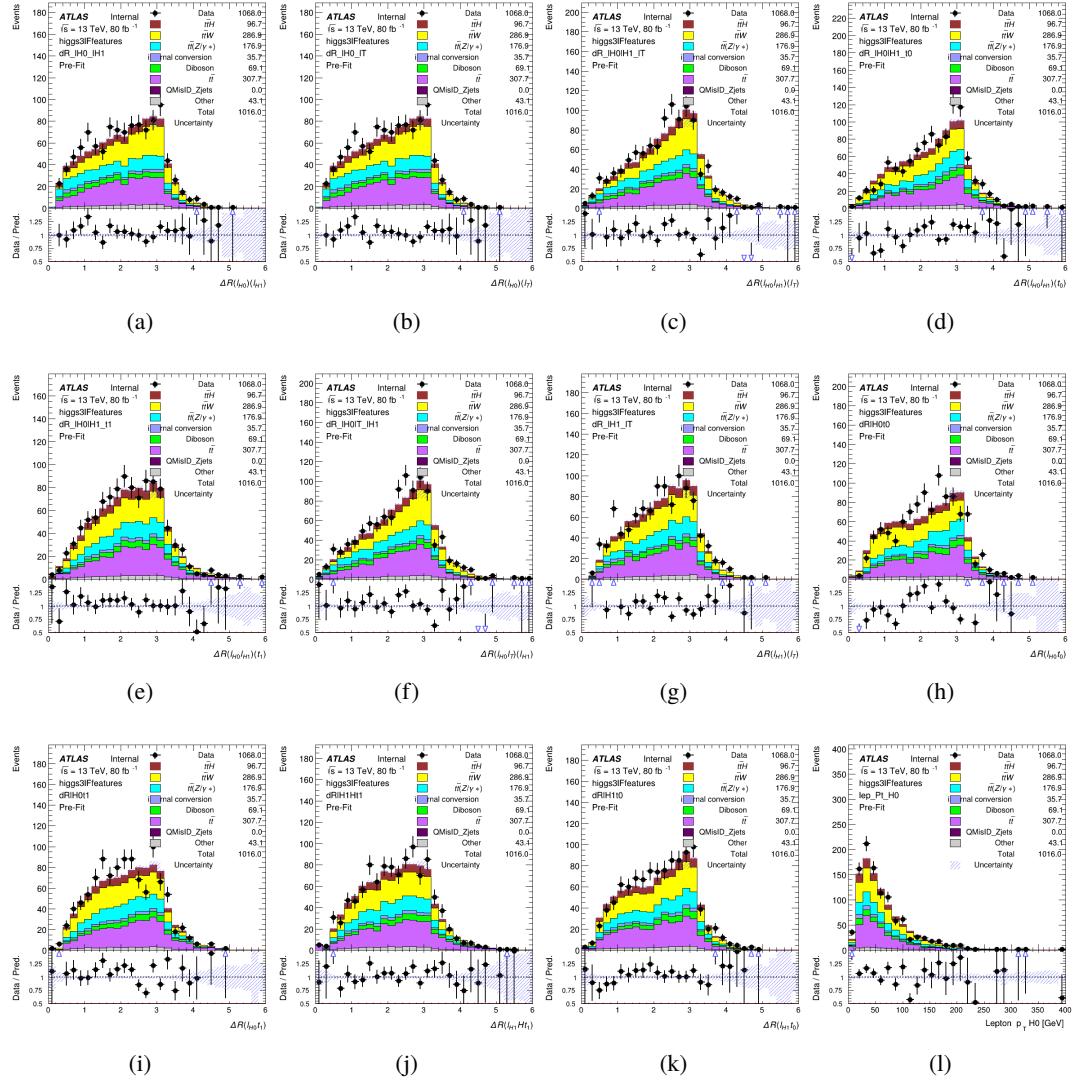


Figure A.17: Input features for higgs3IF

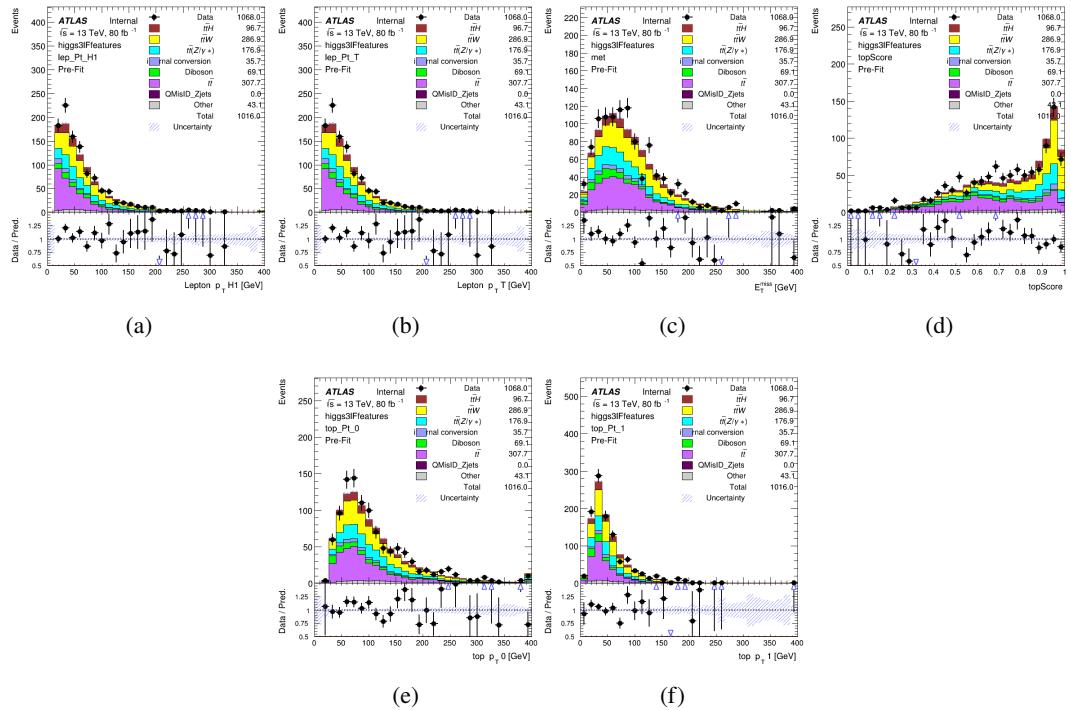


Figure A.18: Input features for higgs3lF

684 **A.2 Background Rejection MVA Details**

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

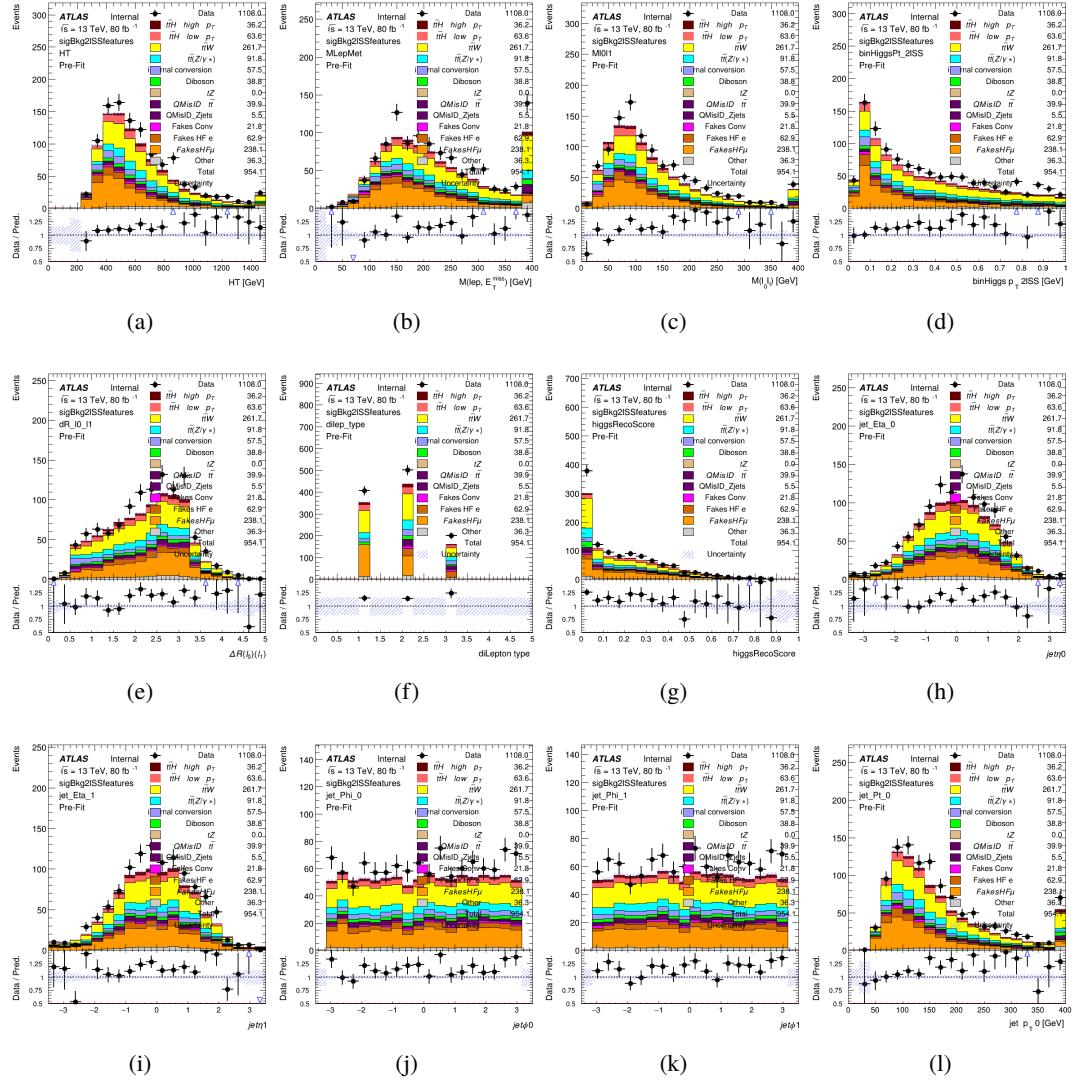


Figure A.19: Input features for sigBkg2lSS

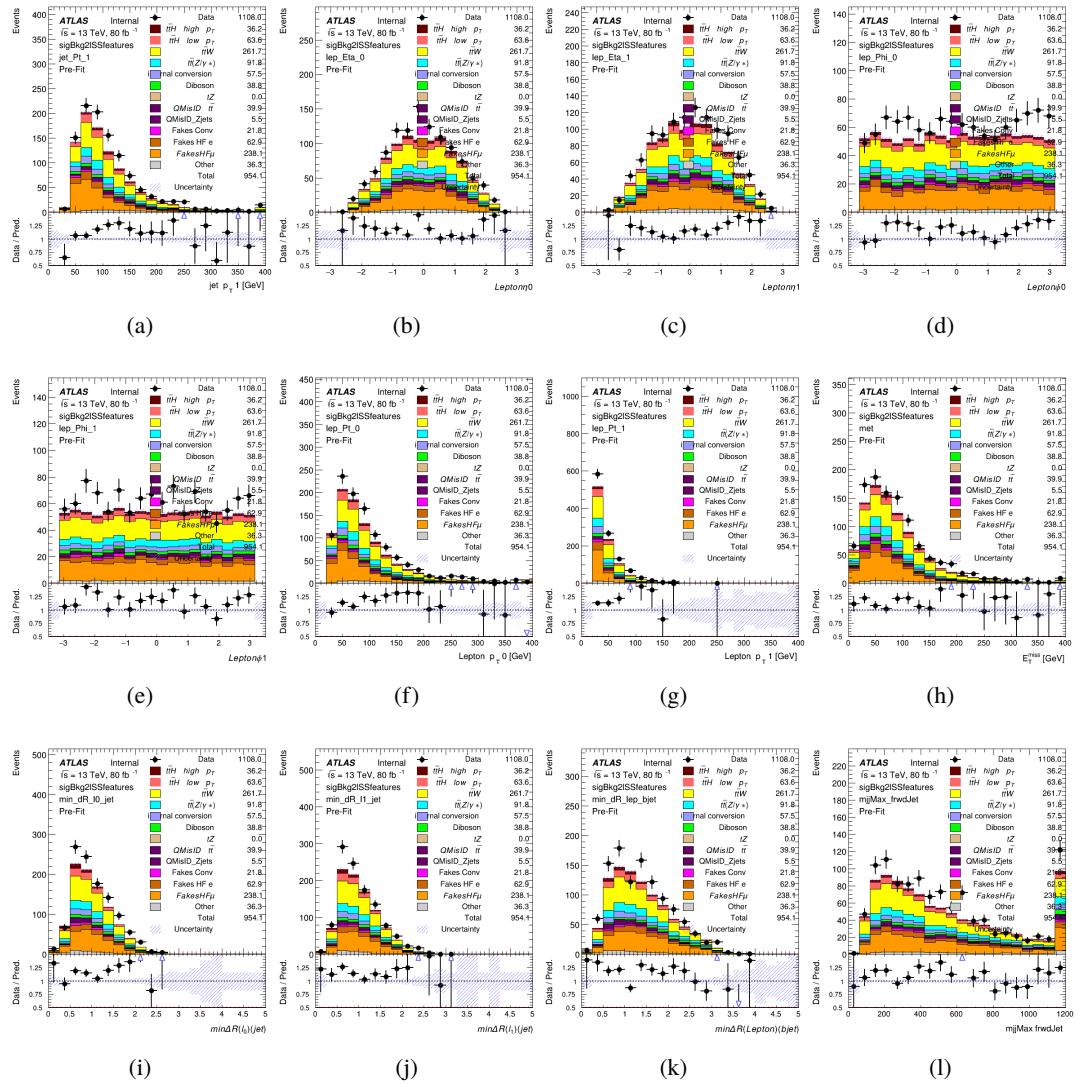


Figure A.20: Input features for sigBkg2lSS

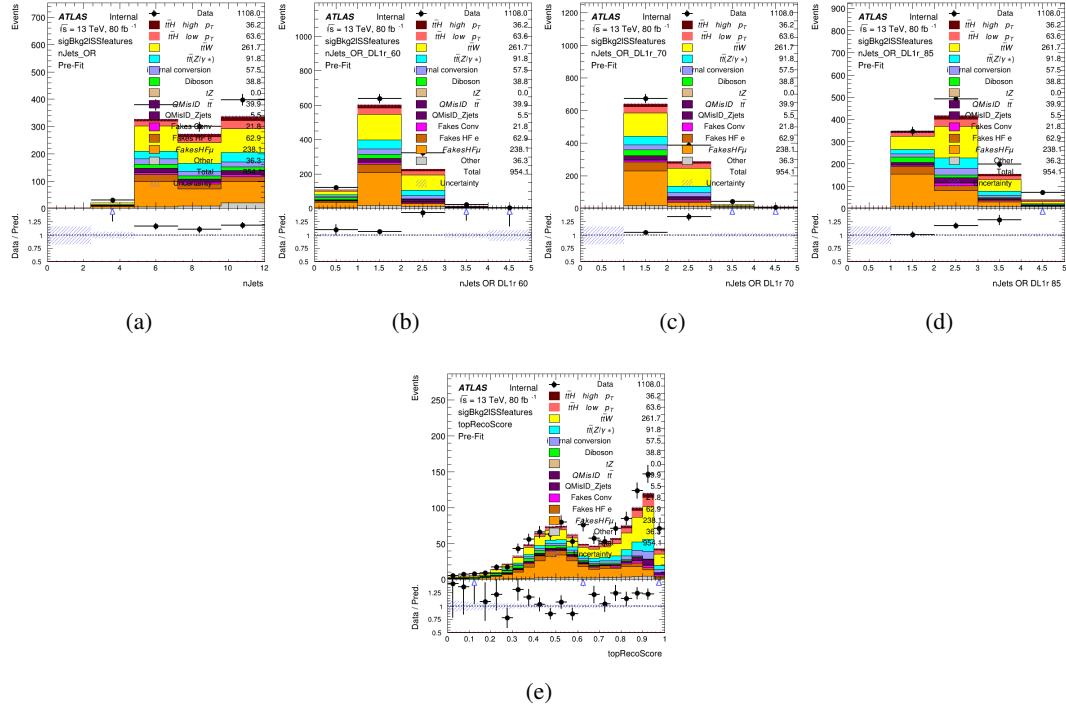


Figure A.21: Input features for sigBkg2lSS

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

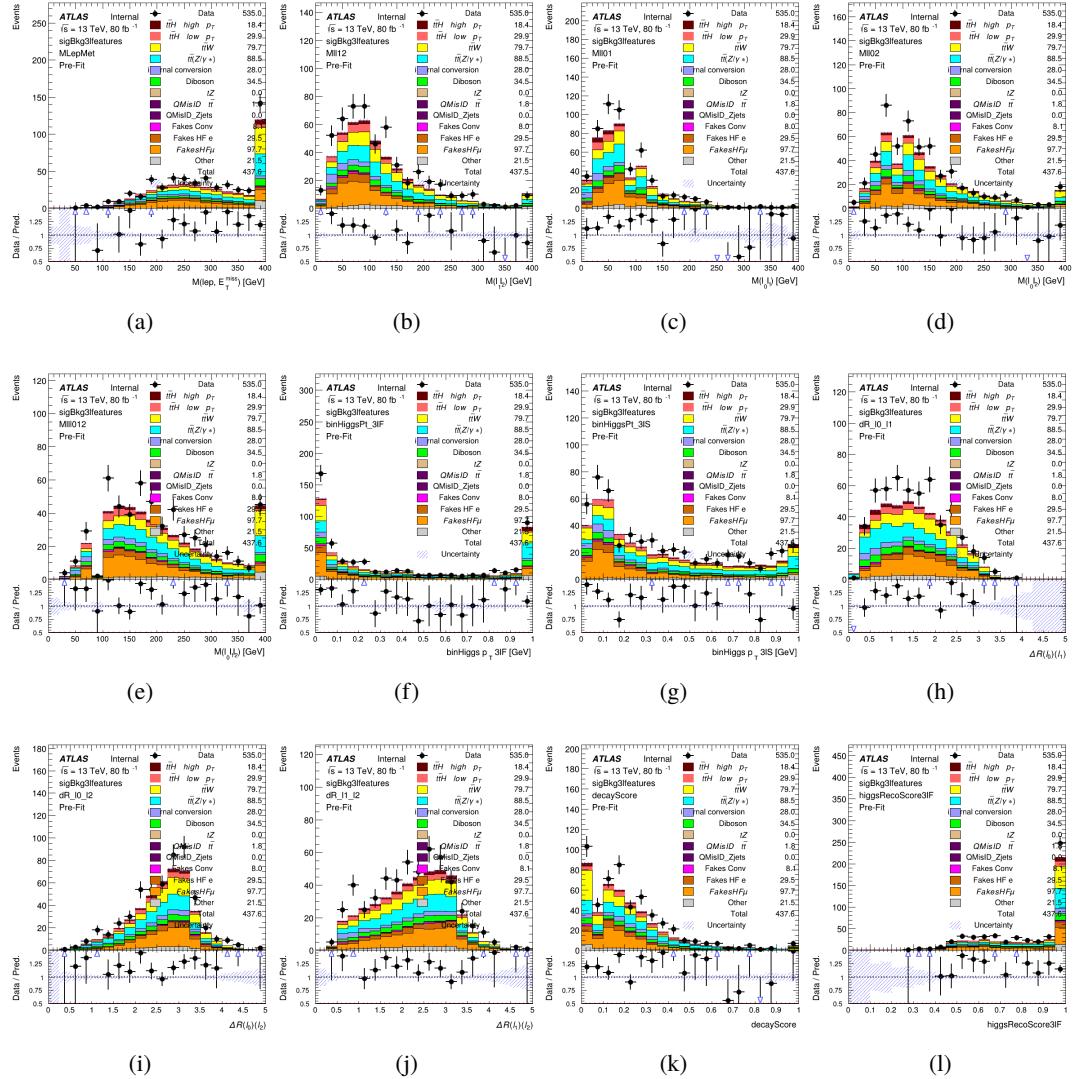


Figure A.22: Input features for sigBkg3l

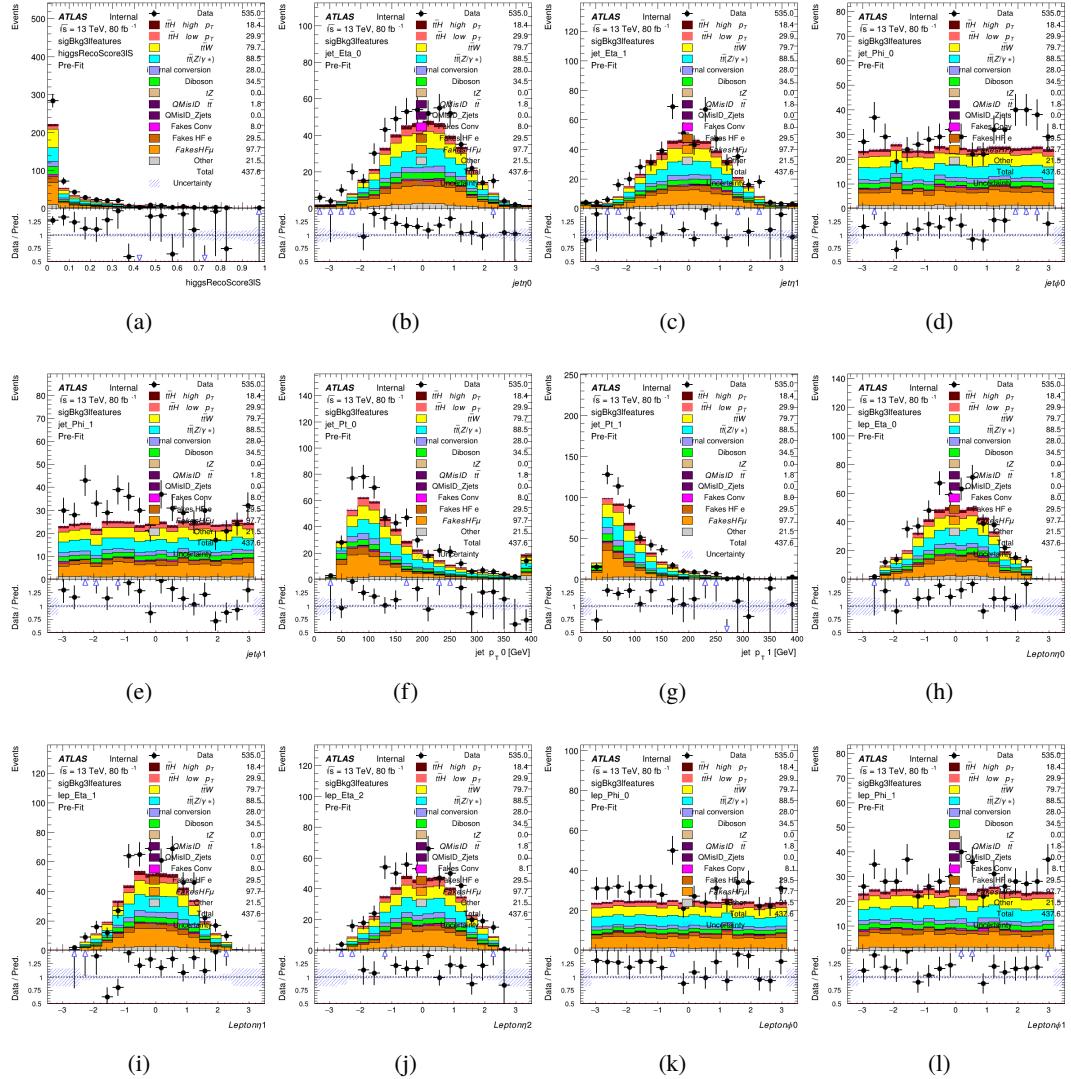
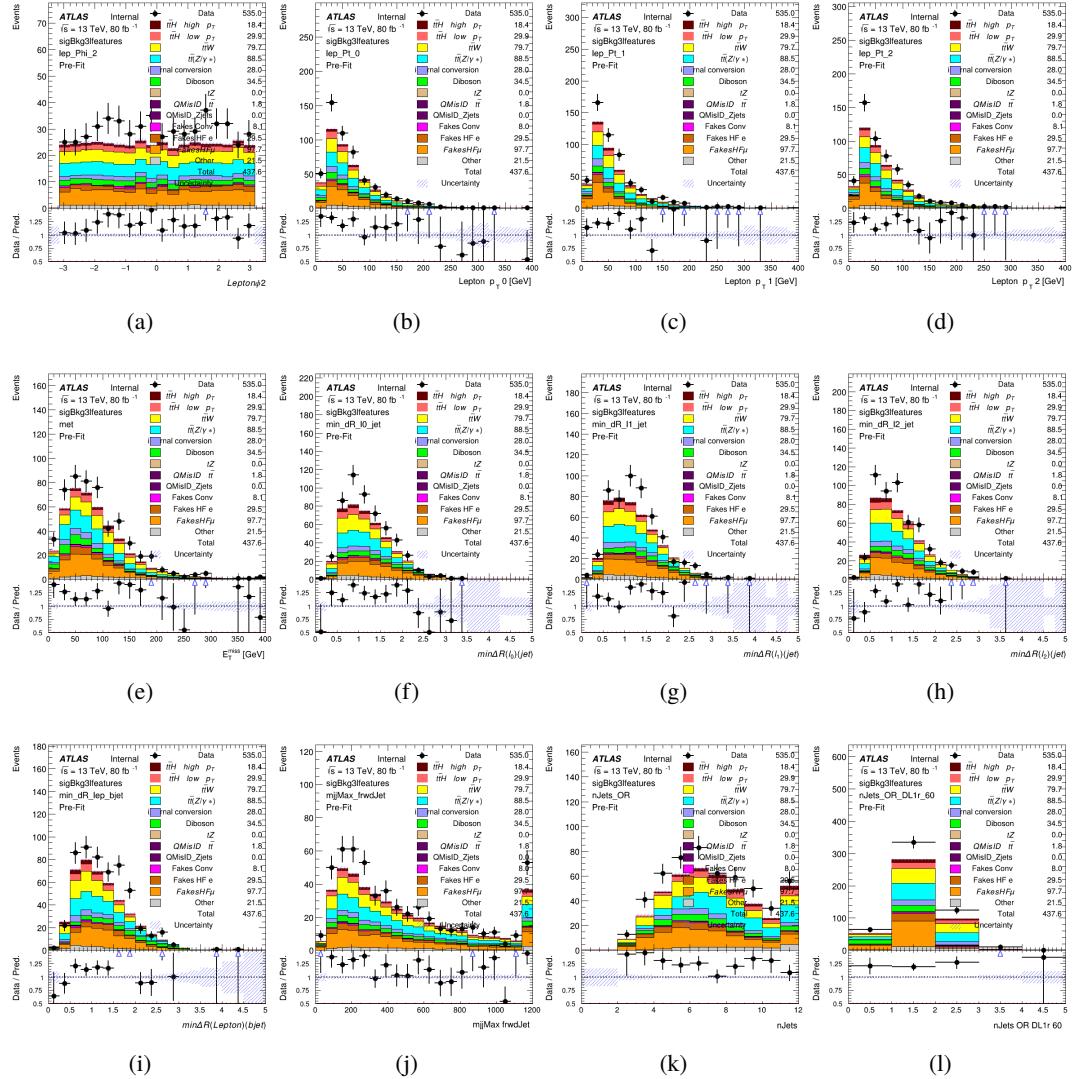


Figure A.23: Input features for sigBkg3l

Figure A.24: Input features for sigBkg3l

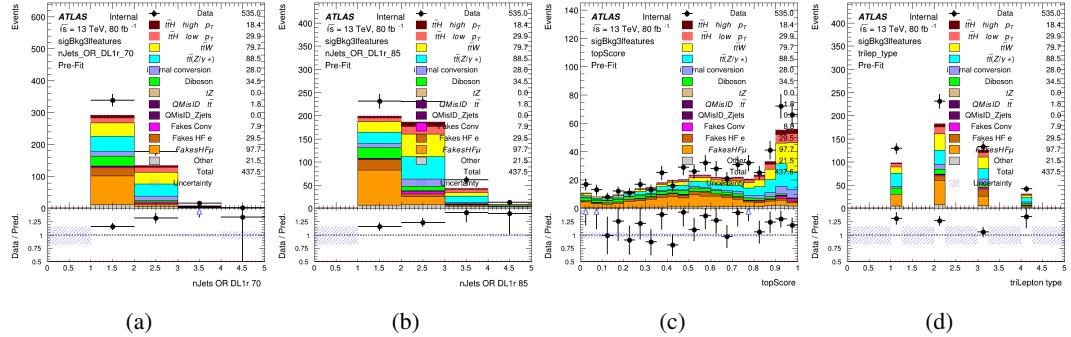


Figure A.25: Input features for sigBkg3l

687 **A.3 Truth Level Studies**

688 Attempts to identify the decay products of the Higgs are motivated by the ability to reconstruct
 689 the Higgs momentum based on their kinematics. In order to demonstrate that this is case, the
 690 kinematics of reconstructed objects that are truth matched to the Higgs decay are used as inputs
 691 to a neural network which is designed to predict the momentum of the Higgs. This is done in
 692 the 2lSS channel, as it proves to be the most challenging for p_T reconstruction.

693 Only leptons and jets which are truth matched to the Higgs are used as inputs for the model;
 694 events where the lepton and both jets are not reconstructed are not included. The model uses the
 695 same feature set and network architecture as the p_T prediction model used in the main analysis, as
 696 described in Section 5.5.1.

697 The results of the model are summarized below:

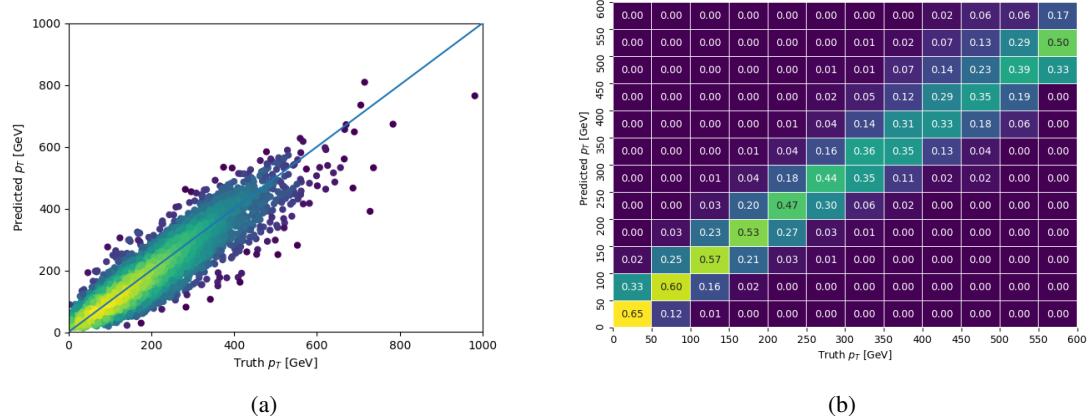


Figure A.26: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

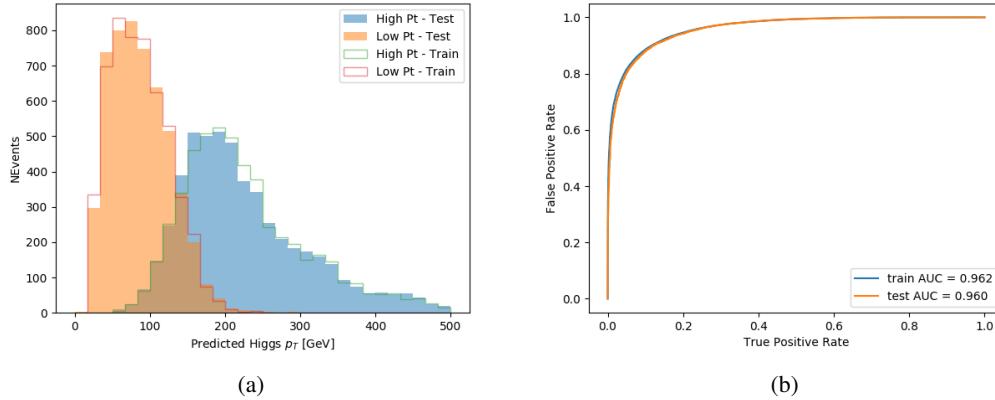


Figure A.27: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

698 Based on the performance of the model, as shown Figures A.26 and A.27, the Higgs momentum
 699 can be reconstructed with fairly high precision when its decay products are correctly identified.

700 **A.4 Alternate b-jet Identification Algorithm**

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

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

Table 34: Accuracy of the NN in identifying b-jets from tops in 2lSS events for the alternate categorical method compared to the nominal method.

Channel	Categorical	Nominal
2lSS	70.6%	73.9%
3l	76.1%	79.8%

713 **A.5 Binary Classification of the Higgs p_T**

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

718 A model is built to determine whether $t\bar{t}H$ events include a high p_T (>150 GeV) or low p_T (<150
 719 GeV) Higgs Boson. While this is now a classification model, it uses the same input features
 720 described in section 5.5. Binary crossentropy is used as the loss function.

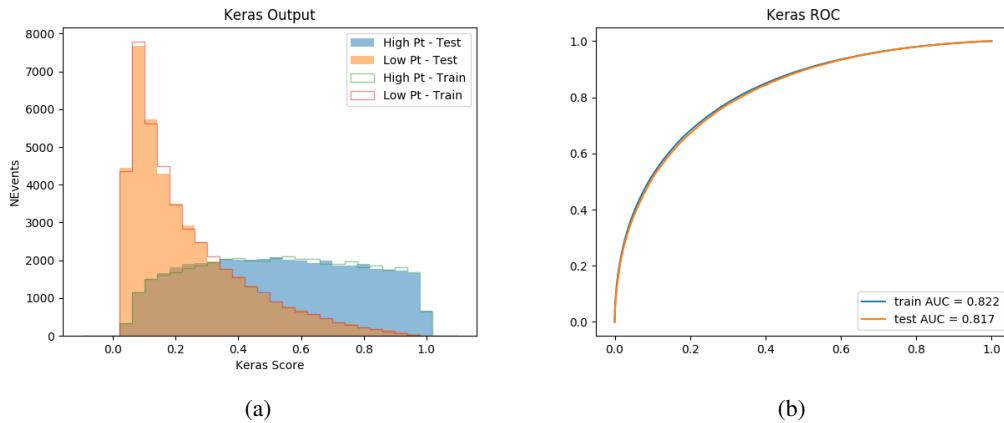


Figure A.28: Output distribution of the NN score for the binary high/low p_T separation model in the 2lSS channel.

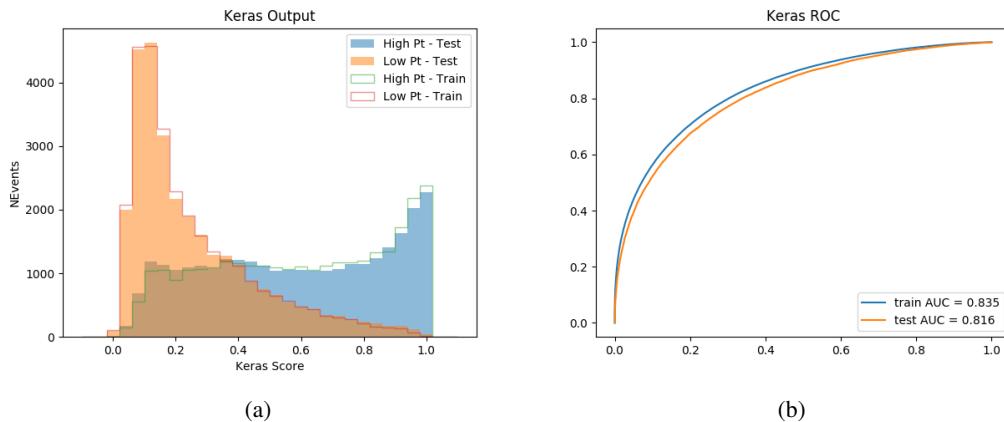


Figure A.29: Output distribution of the NN score for the binary high/low p_T separation model in the 3lS channel.

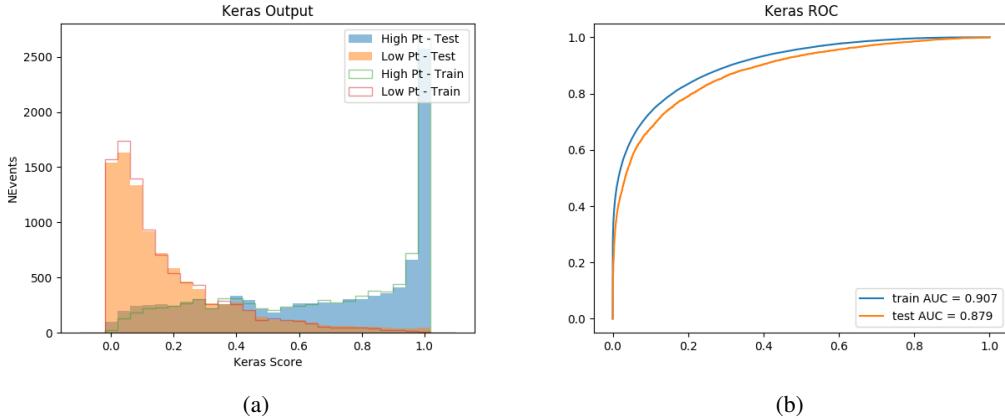


Figure A.30: Output distribution of the NN score for the binary high/low p_T separation model in the 3lS channel.

A.6 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 models are retrained in the 2lSS channel with the same parameters as those used in the nominal analysis, but the jet p_T threshold is altered. The performance of the Higgs p_T prediction models for jet p_T cuts of 20 and 25 GeV are shown below.

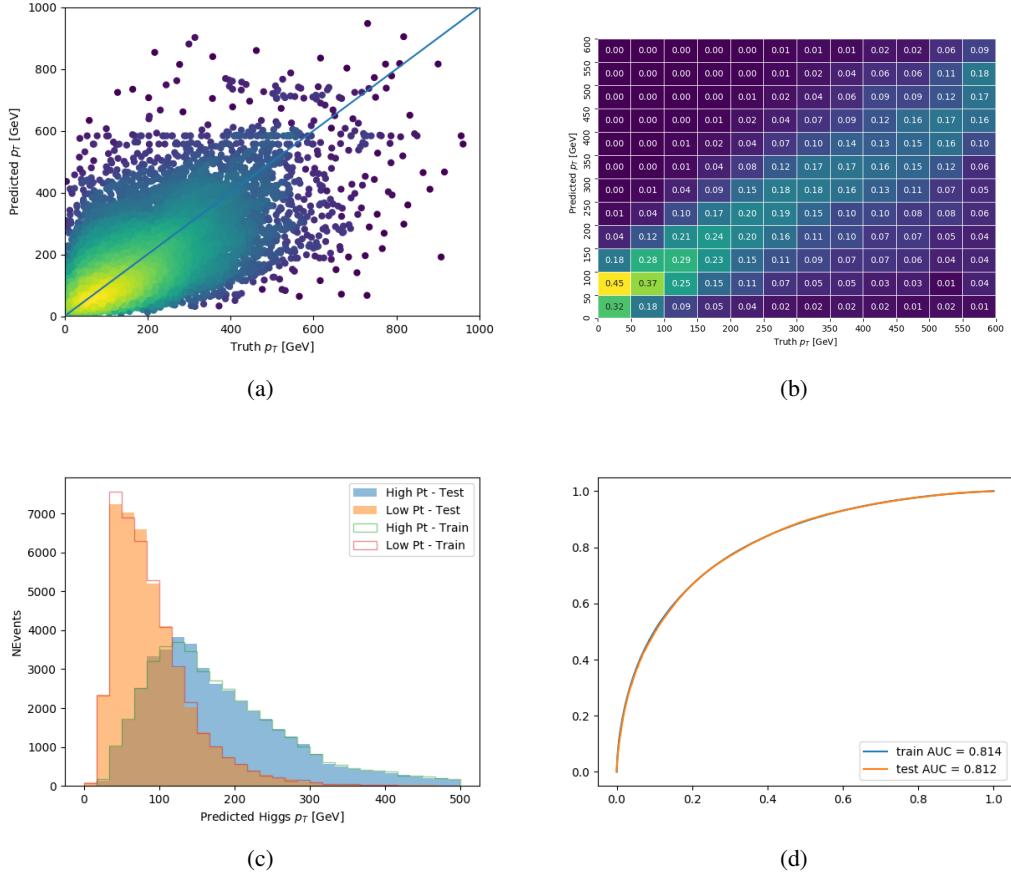
Jet $p_T > 20$ GeV

Figure A.31: Output of the model designed to predict the Higgs momentum in the 2LSS channel, with the jet p_T cutoff used is raised to 20 GeV.

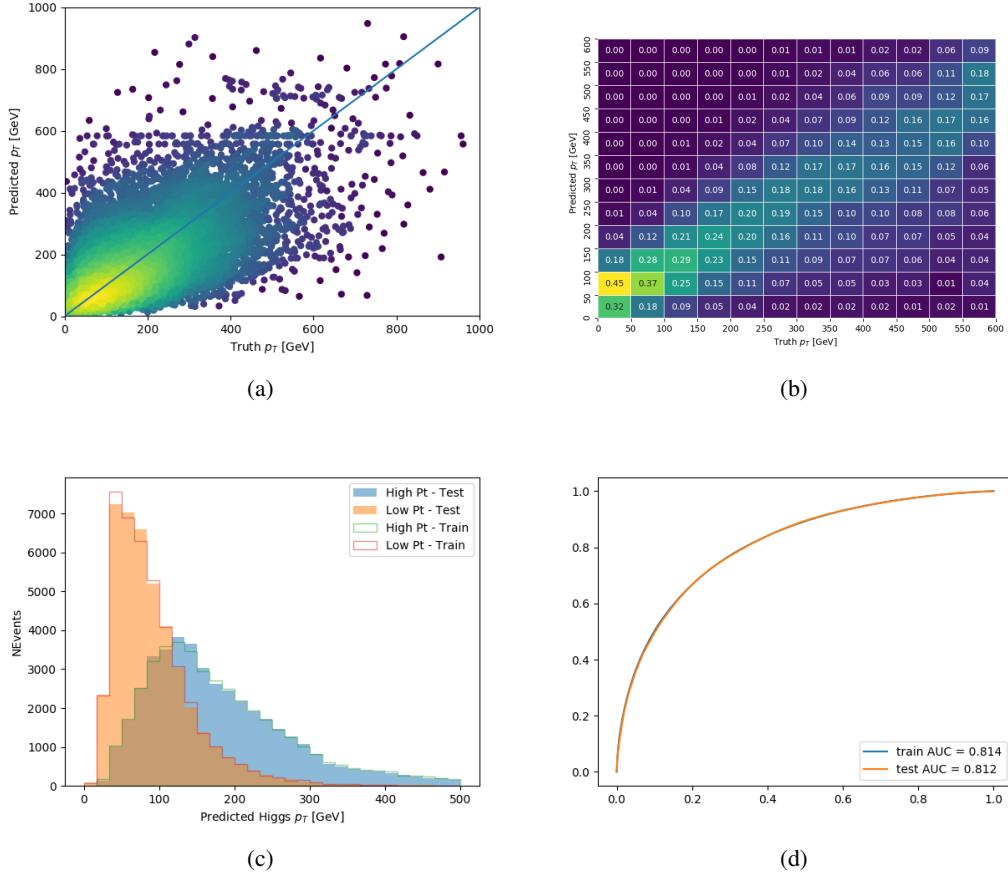
Jet $p_T > 25 \text{ GeV}$ 

Figure A.32: Output of the model designed to predict the Higgs momentum in the 2LSS channel, with the jet p_T cutoff used is raised to 25 GeV.