



ATLAS Note

GROUP-2017-XX

4th May 2021



Draft version 0.1

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

The ATLAS Collaboration

The possibility of performing differential measurements of $t\bar{t}H$ events with multiple leptons in the final state is investigated. 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. Preliminary results using 79.8 fb^{-1} , as well as projected results for 139 fb^{-1} , at $\sqrt{s} = 13 \text{ TeV}$ are presented, providing estimates of the sensitivity to variations in the Higgs p_T spectrum in this channel.

© 2021 CERN for the benefit of the ATLAS Collaboration.

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

16 **Contents**

17	1 Changes and outstanding items	4
18	1.1 Changelog	4
19	2 Introduction	5
20	3 Data and Monte Carlo Samples	6
21	3.1 Data Samples	6
22	3.2 Monte Carlo Samples	6
23	4 Object Reconstruction	9
24	4.1 Trigger Requirements	9
25	4.2 Light Leptons	10
26	4.3 Jets	11
27	4.4 B-tagged Jets	11
28	4.5 Missing Transverse Energy	11
29	4.6 Overlap removal	11
30	5 Higgs Momentum Reconstruction	12
31	5.1 Physics Object Truth Matching	13
32	5.2 Truth Level Studies	13
33	5.3 b-jet Identification	13
34	5.4 Higgs Reconstruction	20
35	5.5 p_T Prediction	31
36	5.6 3l Decay Mode	37
37	6 Signal Region Definitions	39
38	6.1 Pre-MVA Event Selection	39
39	6.2 Event MVA	43
40	6.3 Signal Region Definitions	47
41	7 Systematic Uncertainties	50
42	8 Results	55
43	8.1 Results - 79.8 fb^{-1}	56
44	8.2 Projected Results - 139 fb^{-1}	60
45	9 Conclusion	64
46	Appendices	68
47	.1 Non-prompt lepton MVA	68

48	A Machine Learning Models	70
49	A.1 Higgs Reconstruction Model Details	70
50	A.2 Background Rejection MVA Details	93
51	A.3 Truth Level Studies	102
52	A.4 Alternate b-jet Identification Algorithm	103
53	A.5 Binary Classification of the Higgs p_T	104
54	A.6 Impact of Alternative Jet Selection	105

⁵⁵ **1 Changes and outstanding items**

⁵⁶ **1.1 Changelog**

⁵⁷ This is version 1

58 2 Introduction

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

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

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 [3]. Therefore, reconstructing the momentum spectrum of the Higgs provides a means to observe
 73 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 p_T ($> 150 \text{ GeV}$) and low p_T ($< 150 \text{ GeV}$) 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 [4] (CT10 [5])
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 [6]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄t̄γ	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
t̄t̄, t̄tt̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄t̄W+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	POWHEG-BOX v1 [7]	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) [8]. Parton showering and hadronisation were modelled with Pythia
 123 8.2 [9]. The $t\bar{t}H$ sample is normalized to a cross-section of 507^{+35}_{-50} fb based on NLO calculations.
 124 Uncertainties are based on varying the QCD factorisation and renormalisation scale, as well as
 125 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 [10], the $t\bar{t}W$ contribution predicted by MC is found disagree significantly with
 129 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 The $t\bar{t}(Z/\gamma^* \rightarrow ll)$ process is simulated with the MADGRAPH5_AMC@NLO generator, using NNPDF3.0.
 134 Diboson processes are generated with SHERPA 2.2.2 at NLO precision for one extra parton, and at
 135 LO for up to three extra partons.

¹³⁶ The estimation of the “fake” or non-prompt background - with leptons from hadron decays or
¹³⁷ photon conversions - is done primarily using an inclusive $t\bar{t}$ sample. This sample is generated
¹³⁸ using POWHEG, with PYTHIA8 performing the parton shower and fragmentation.

¹³⁹ While the main $t\bar{t}H$ analysis is currently refining a data-driven approach for estimating the contribution
¹⁴⁰ of events with non-prompt leptons, at the time of this note this strategy has not been completely developed for the full Run-2 dataset. Therefore, the non-prompt contribution
¹⁴¹ is estimated with MC, while applying normalization corrections and systematic uncertainties
¹⁴² derived from data driven techniques developed for the $79.8 \text{ fb}^{-1} t\bar{t}H/t\bar{t}W$ analysis [10]. The primary contribution to the non-prompt lepton background is from $t\bar{t}$ production, with V+jets
¹⁴³ and single-top as much smaller sources. Likelihood fits over several control regions enriched with
¹⁴⁴ these non-prompt backgrounds are fit to data in order to derive normalization factors for these
¹⁴⁵ backgrounds. The specific normalization factors and uncertainties applied to the non-prompt
¹⁴⁶ contributions are listed in Section 7.
¹⁴⁷

¹⁴⁸ Other background processes, such as tH , tZ , $t\bar{t}WW$ and $t\bar{t}t\bar{t}$, are expected to make minor contributions to the total background. The generators and setting used for these backgrounds are summarized in Table ??.

¹⁵¹ The specific DSIDs used in the analysis are listed below:

Sample	DSID
tH	345873-5, 346343-5
VV	364250-364254, 364255, 363355-60, 364890
tW	413008
tZ	410156, 410157, 410218-20
low mass tZ	410276-8
Rare Top	410397, 410398, 410399
single Top	410658-9, 410644-5
three Top	304014
four Top	410080
tWW	410081
Z + jets	364100-41
low mass Z + jets	364198-215
W + jets	364156-97
V γ	364500-35
tZ	410560
tW	410013-4
WtZ	410408
VVV	364242-9
VH	342284-5
WtH	341998
t $\bar{t}\gamma$	410389
t \bar{t}	410470

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

153 4 Object Reconstruction

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

156 4.1 Trigger Requirements

157 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.

158 4.2 Light Leptons

159 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter
 160 that are associated with charged particle tracks reconstructed in the inner detector [11]. Electron
 161 candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the transition
 162 region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are
 163 rejected. A multivariate likelihood discriminant combining shower shape and track information
 164 is used to distinguish prompt electrons from nonprompt leptons, such as those originating from
 165 hadronic showers. Electron candidate are also required to pass TightLH identification.

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

169 Muon candidates are reconstructed by combining inner detector tracks with track segments or full
 170 tracks in the muon spectrometer [12]. Muon candidates are required to have $p_T > 10$ GeV and
 171 $|\eta| < 2.5$. Muons are required to Medium ID requirements.

172 All leptons are required to pass a non-prompt BDT selection developed by the main $t\bar{t}H - ML/t\bar{t}W$
 173 analysis, described in detail in [10]. Optimized working points and scale factors for this BDT are
 174 taken from that analysis. This BDT and the WPs used are summarized in Appendix .1,

175 **4.3 Jets**

176 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
 177 calorimeters [13], using the anti- k_t algorithm with a radius parameter $R = 0.4$. Particle Flow, or
 178 PFlow, jets are used in the analysis, which are hadronic objects reconstructed using information
 179 from both the tracker and the calorimeter. Jets with energy contributions likely arising from noise
 180 or detector effects are removed from consideration [14], and only jets satisfying $p_T > 25$ GeV
 181 and $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
 182 association algorithm is used to confirm that the jet originates from the selected primary vertex,
 183 in order to reject jets arising from pileup collisions [15].

184 **4.4 B-tagged Jets**

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

191 **4.5 Missing Transverse Energy**

192 Because all $t\bar{t}H - ML$ channels considered include multiple neutrinos, missing transverse energy
 193 (E_T^{miss}) is present in each event. The missing transverse momentum vector is defined as the
 194 inverse of the sum of the transverse momenta of all reconstructed physics objects as well as
 195 remaining unclustered energy, the latter of which is estimated from low- p_T tracks associated with
 196 the primary vertex but not assigned to a hard object [16].

197 **4.6 Overlap removal**

198 To avoid double counting objects and remove leptons originating from decays of hadrons, overlap
 199 removal is performed in the following order: any electron candidate within $\Delta R = 0.1$ of another
 200 electron candidate with higher p_T is removed; any electron candidate within $\Delta R = 0.1$ of a muon
 201 candidate is removed; any jet within $\Delta R = 0.3$ of an electron candidate is removed; if a muon
 202 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
 203 is kept and the muon is removed.

204 This algorithm is applied to the preselected objects. The overlap removal procedure is summarized
 205 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.

206 5 Higgs Momentum Reconstruction

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

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

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

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

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

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

237 **5.1 Physics Object Truth Matching**

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

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

249 Leptons considered as possible Higgs and top decay candidates are required to pass the selection
 250 described in Section 4.2. For jets, however, it is found that a large fraction that originate from either
 251 the top decay or the Higgs decay fall outside the selection described in Section 4.3. Specifically,
 252 jets from the Higgs decay tend to be soft, with 32% having $p_T < 25 \text{ GeV}$. Therefore jets with p_T
 253 $< 15 \text{ GeV}$ are considered as possible candidates in the models described below. By contrast, less
 254 than 5% of the jets originating from the Higgs fall below this p_T threshold. The jets are found to
 255 be well modeled even down to this low p_T threshold, as shown in Section 6.1. The impact of
 256 using different p_T selection for the jet candidates is considered in detail in Section A.6. As they
 257 are expected to originate from the primary vertex, jets are also required to pass a JVT cut.

258 **5.2 Truth Level Studies**

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

264 **5.3 b-jet Identification**

265 Including the kinematics of the b-jets that originate from the top decay is found to improve the
 266 identification of the Higgs decay products, and improve the accuracy with which the Higgs
 267 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.

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

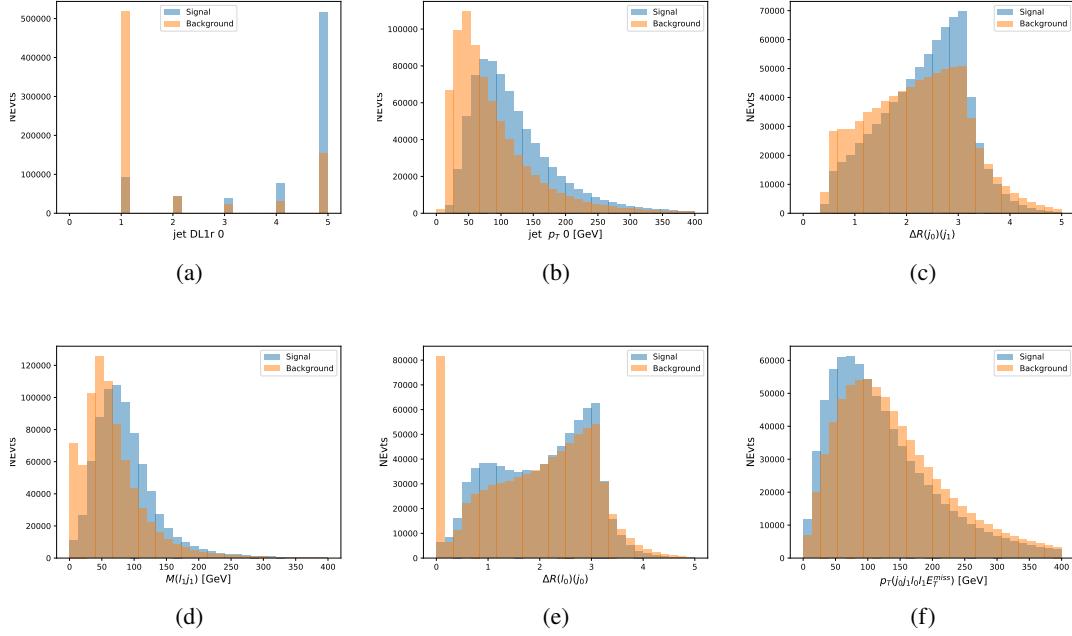


Figure 5.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

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

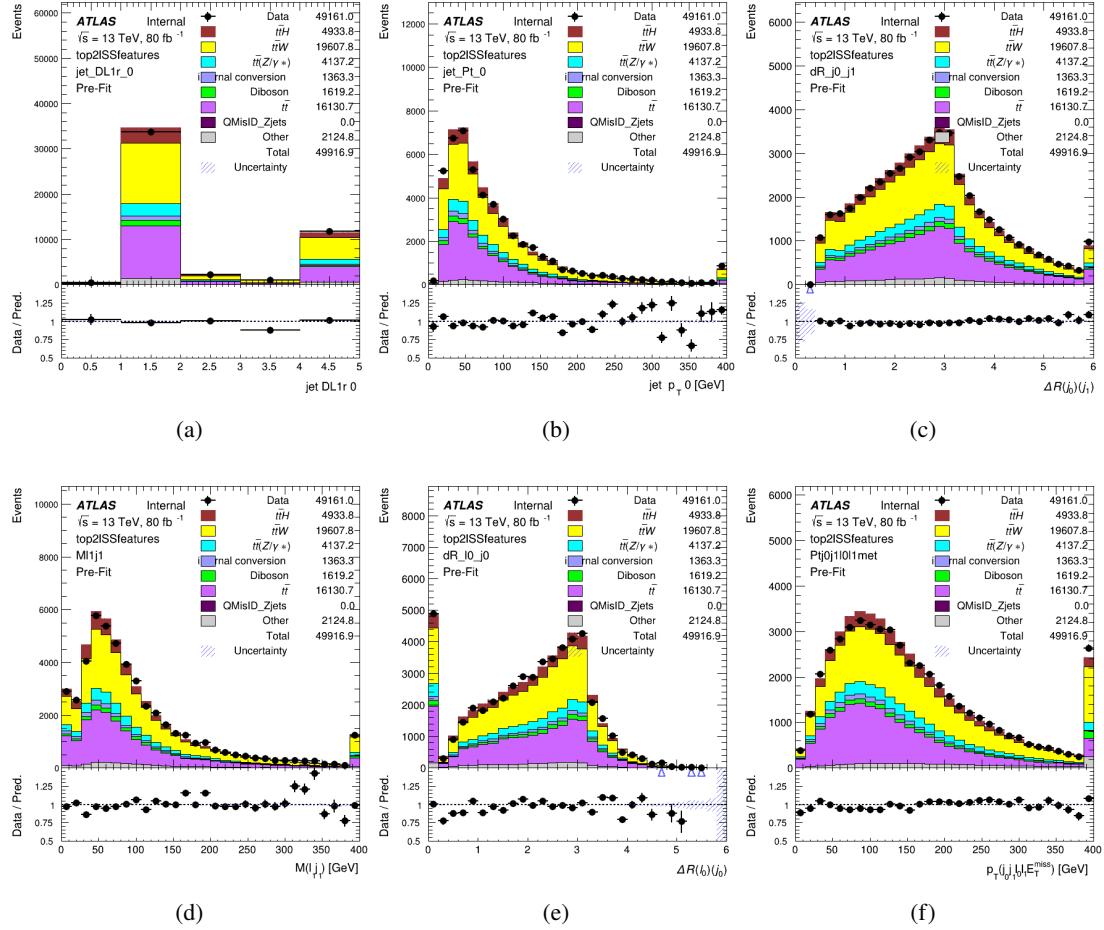


Figure 5.2: Data/MC comparisons of input features for top2ISS training for 79.8 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

297 Based on the results of grid search evaluation, the optimal architecture is found to include 5
 298 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as overfitting
 299 is found to not be an issue. The output score distribution as well as the ROC curve for the trained
 300 model are shown in Figure 5.3.1. The model is found to identify the correct pairing of jets for
 301 73% of 2ISS signal events on test data.

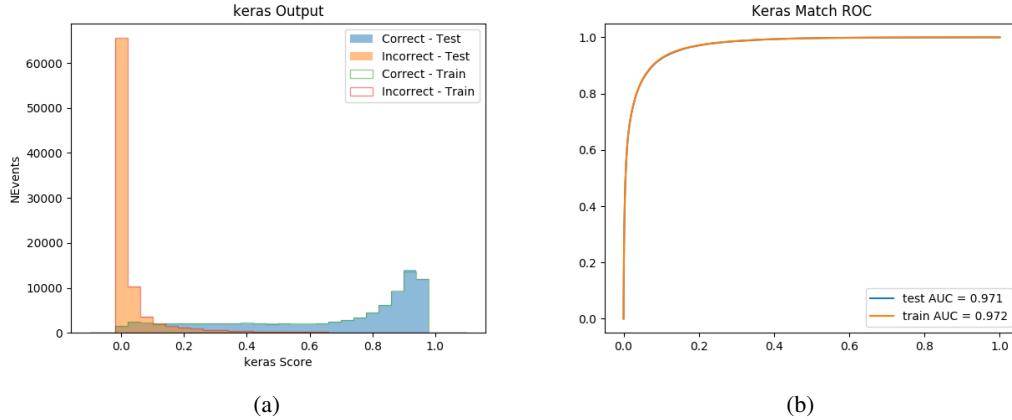


Figure 5.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

306 The accuracy of the model for different values of n-bjets, compared to this naive approach, is
 307 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.

308 5.3.2 3l Channel

309 The input features used in the 3l channel are listed in Table 7, with the same naming convention
 310 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.

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

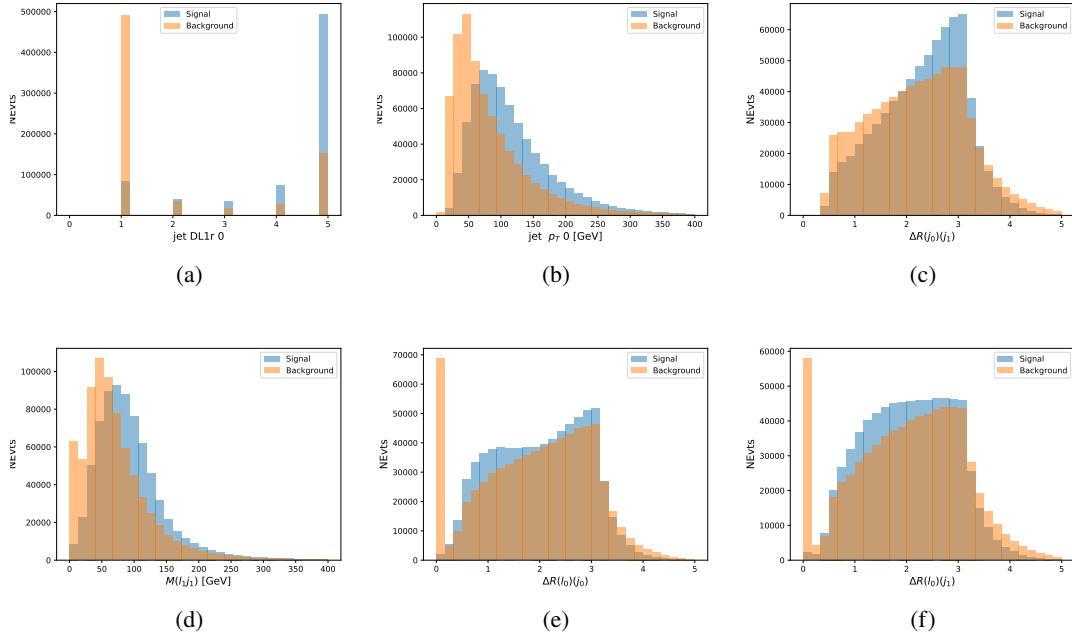


Figure 5.4: Input features for top31 training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and 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

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

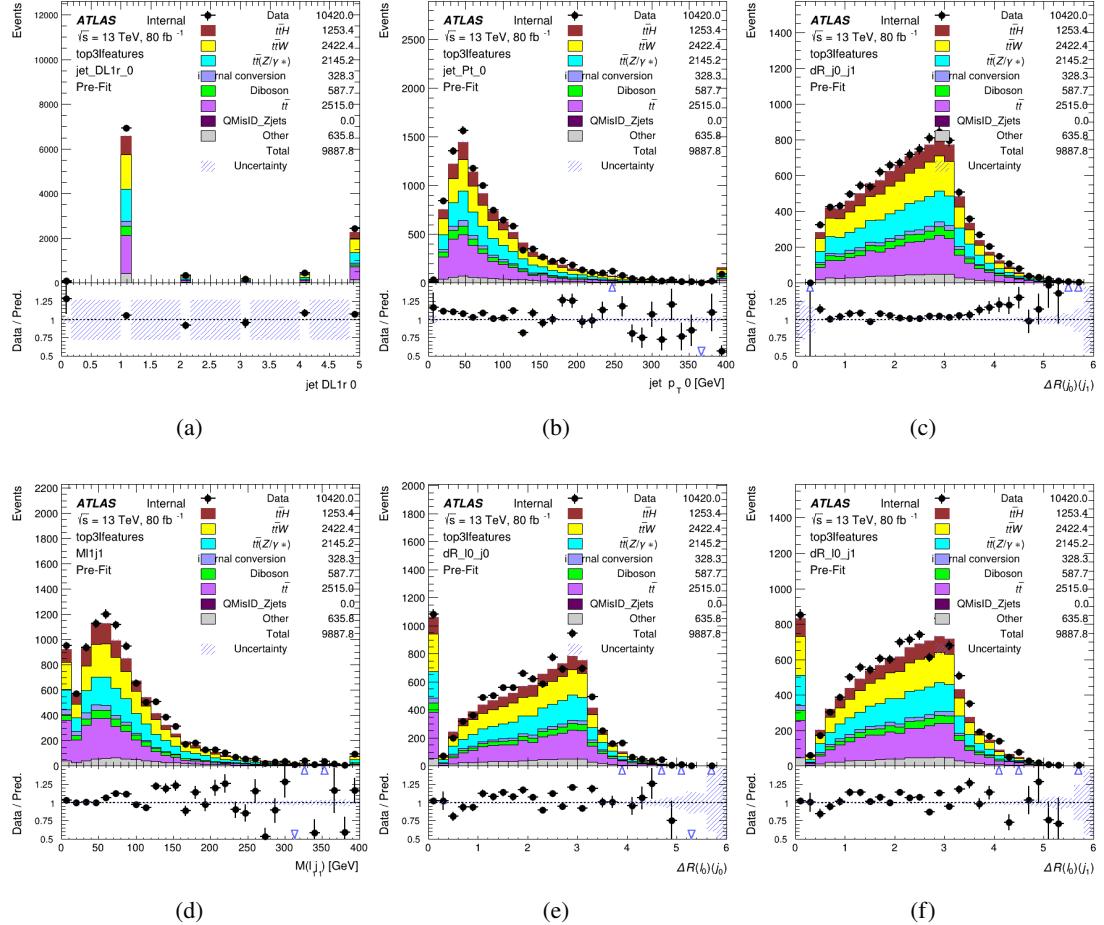


Figure 5.5: Data/MC comparisons of input features for top3l training for 79.8 fb^{-1} of data. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

315 Again, the dataset is downsized to reduce the ratio of correct and incorrect combination from 20:1,
 316 to 5:1. Around 7 million events are used for training, with 10% set aside for testing. Based on the
 317 results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with
 318 60 nodes each. The output score distribution as well as the ROC curve for the trained model are
 319 shown in Figure 5.3.2.

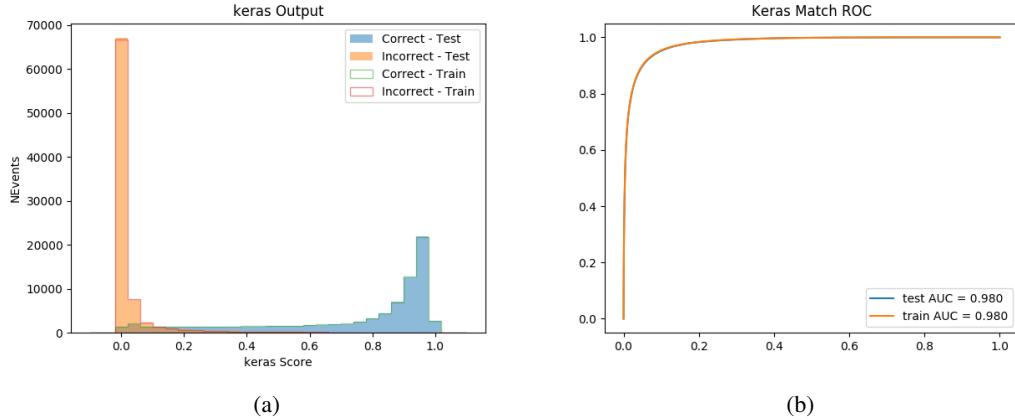


Figure 5.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

320 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal events.
 321 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%

322 5.4 Higgs Reconstruction

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

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

330 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In the
 331 semi-leptonic case the reconstructed decay products consist of two jets and a single leptons. For

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

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

340 5.4.1 2ISS Channel

341 For the 2ISS channel, the Higgs decay products include one light lepton and two jets. The neural
 342 network is trained on the kinematics of different combinations of leptons and jets, as well as the
 343 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 2ISS events

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

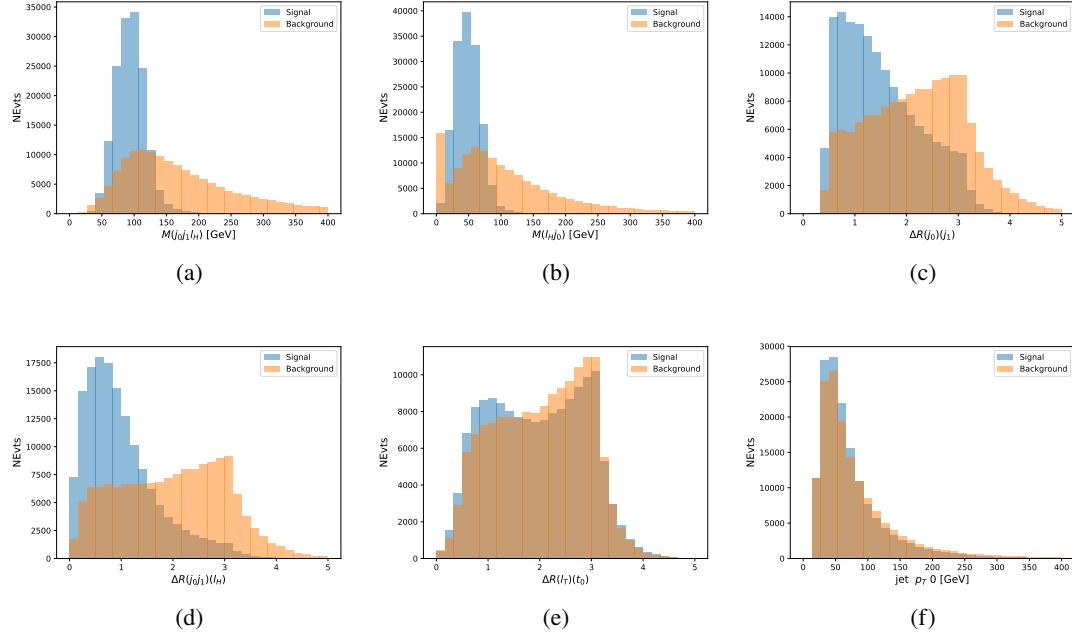


Figure 5.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate,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

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

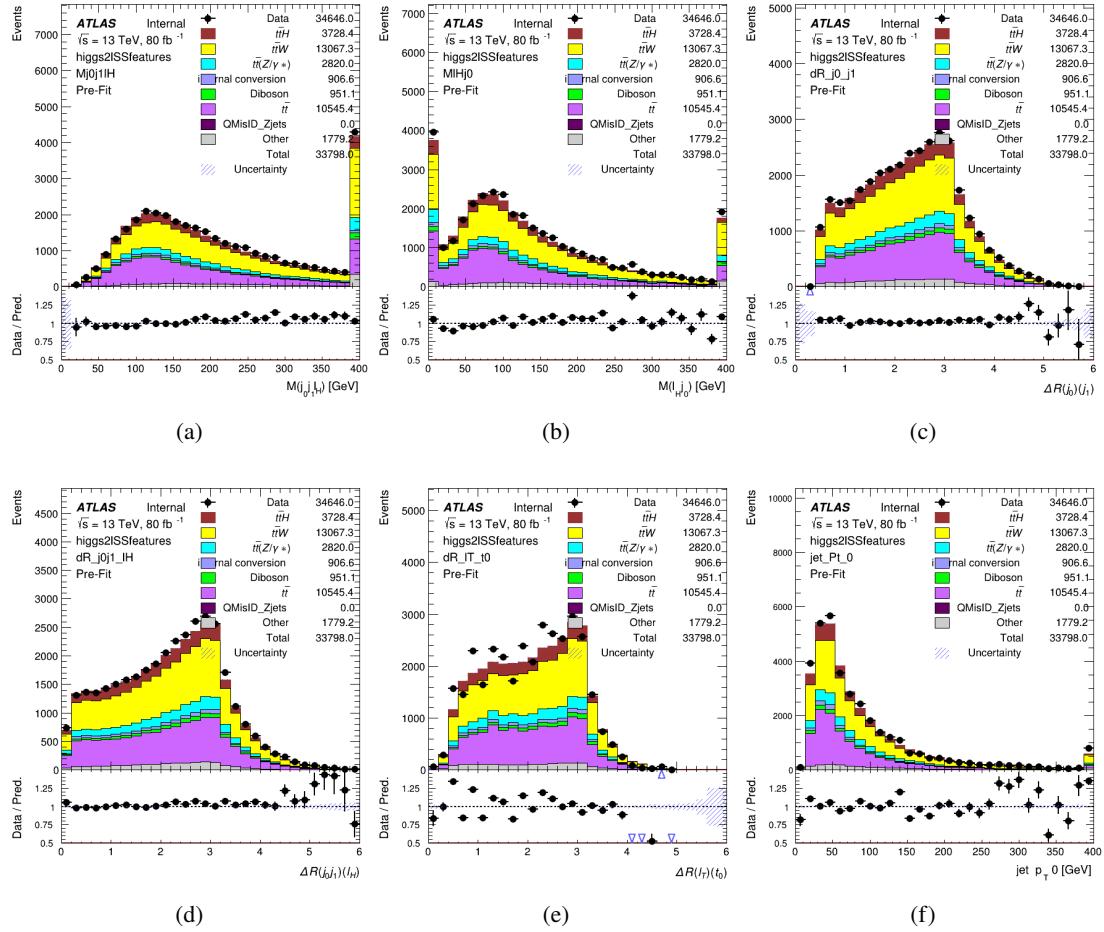


Figure 5.8: Data/MC comparisons of input features for higgs2lSS training for 79.8 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1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

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

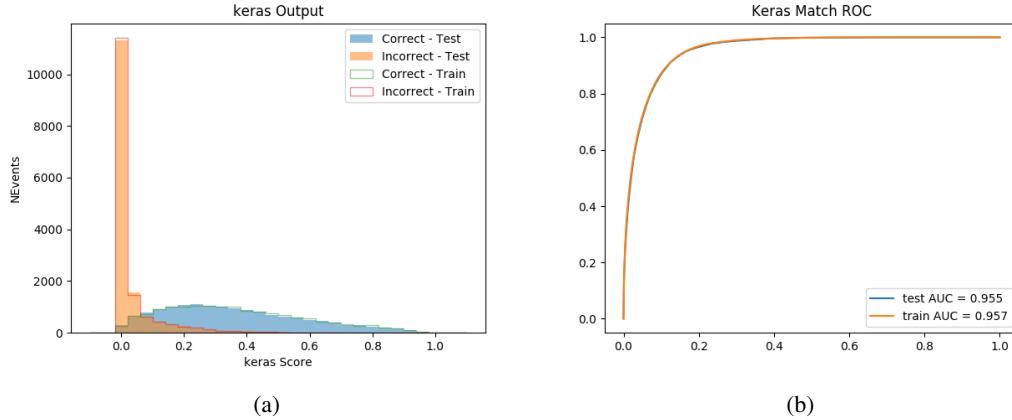


Figure 5.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

358 **5.4.2 3l Semi-leptonic Channel**

359 For 3l $t\bar{t}H$ where the Higgs decay semi-leptonically, the decay products include one of the three
 360 leptons and two jets. In this case, the other two leptons originated from the decay of the tops,
 361 meaning the opposite-sign (OS) lepton cannot have come from the Higgs. This leaves only the two
 362 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 $\eta 0$	jet $\eta 1$
jet $\phi 0$	jet $\phi 1$	$\Delta R(j_0)(j_1)$
$M(j_0 j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1)(l_{T_1})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0 j_1 l_{T_0})$	$M(j_0 j_1 l_{T_1})$
$M(j_0 j_1 l_H)$	$\Delta R(j_0 j_1 l_H)(l_{T_0})$	$\Delta R(j_0 j_1 l_H)(l_{T_1})$
$\Delta\phi(j_0 j_1 l_H)(E_T^{\text{miss}})$	$p_T(j_0 j_1 l_H l_{T_0} l_{T_1} b_0 b_1 E_T^{\text{miss}})$	$M(j_0 j_1 b_0)$
$M(j_0 j_1 b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
topScore	MET	HT jets
nJets		

Table 10: Input features used to identify the Higgs decay products in 31 semi-leptonic events

363 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other two
 364 leptons in the event are labeled as l_{T_0} and l_{T_1} . b_0 and b_1 are the two b-jets identified by the
 365 b-jet identification algorithm. The b-jet Reco Score is the output of the Higgs reconstruction
 366 algorithm.

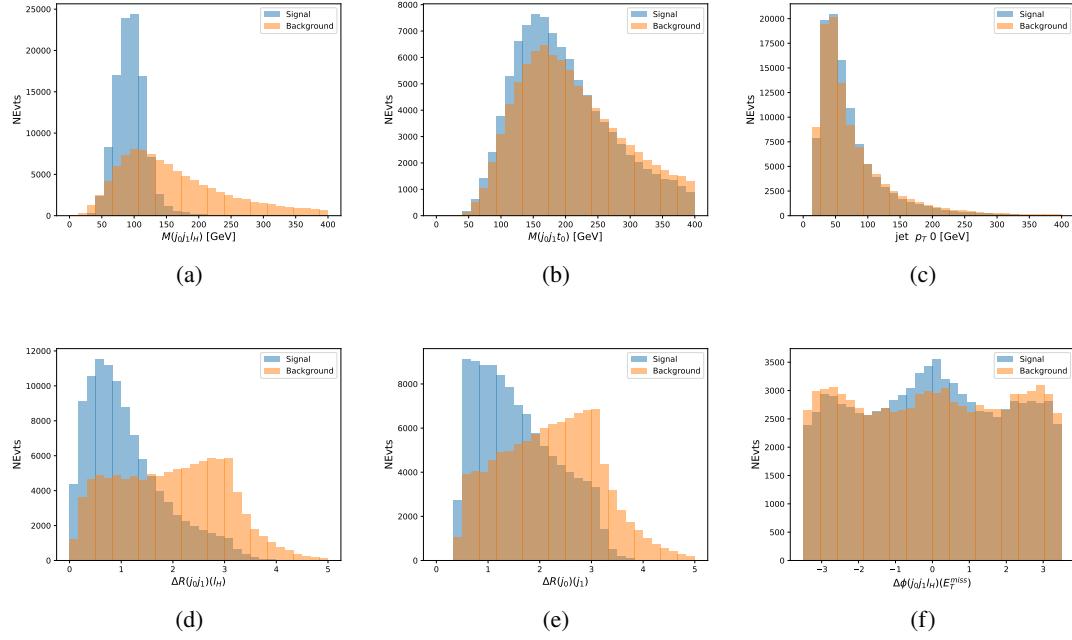


Figure 5.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

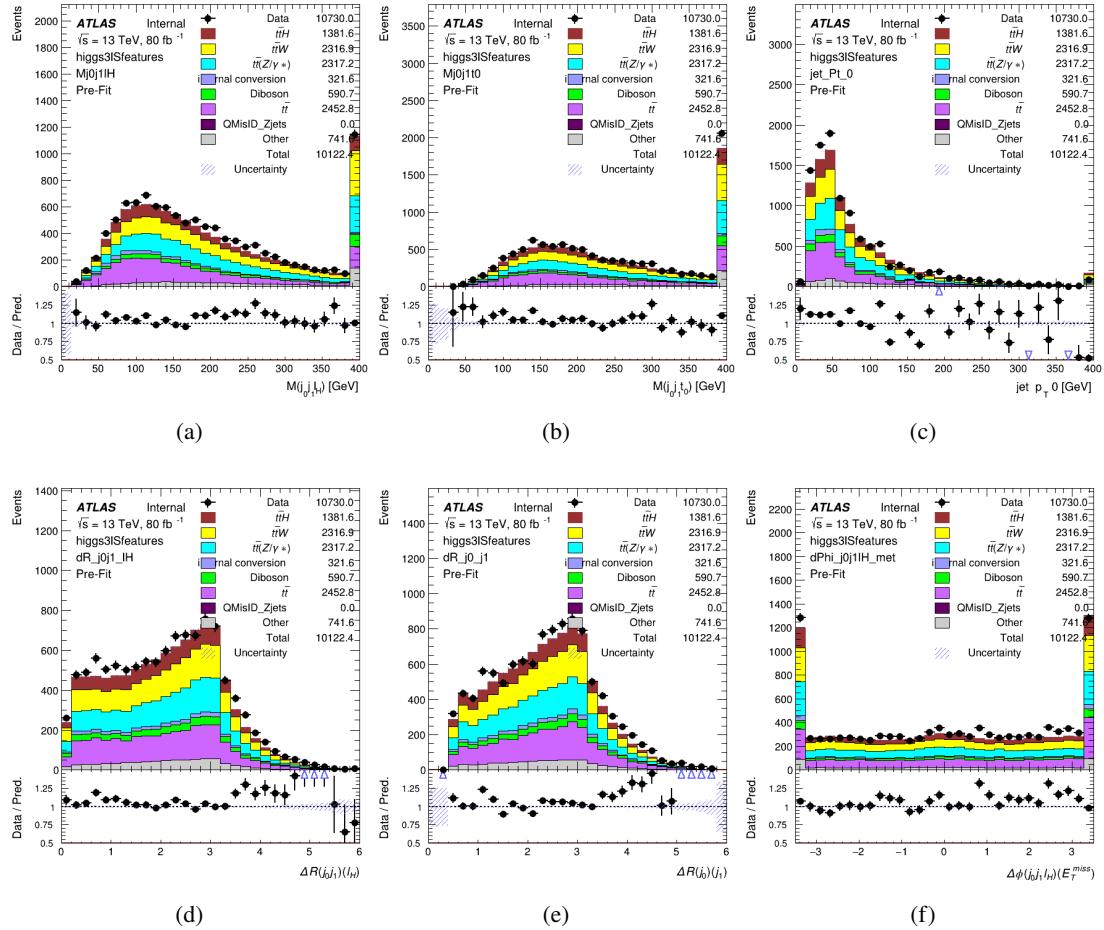


Figure 5.11: Data/MC comparisons of input features for higgs3lS training for 79.8 fb^{-1} of data.

370 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events. Once
 371 again, incorrect combinations are downsampled, such that the correct combinations are around
 372 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN is
 373 summarized in Figure 5.4.2.

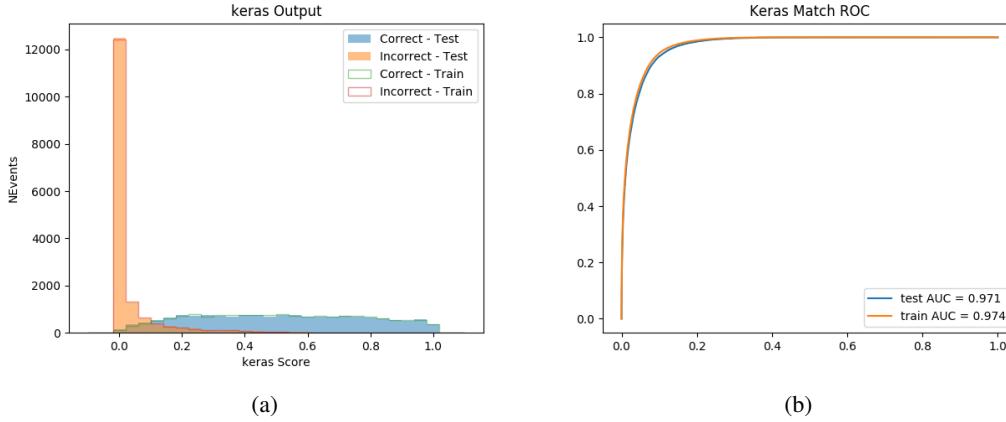


Figure 5.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

377 **5.4.3 3l Fully-leptonic Channel**

378 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 379 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to
 380 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 381 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 382 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

Table 12: Input features used to identify the Higgs decay products in 3l fully leptonic events

383 Here l_{H_0} and l_{H_1} are the Higgs decay candidates. The other lepton in the event is labeled l_T . b_0
 384 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco Score is
 385 the output of the Higgs reconstruction algorithm.

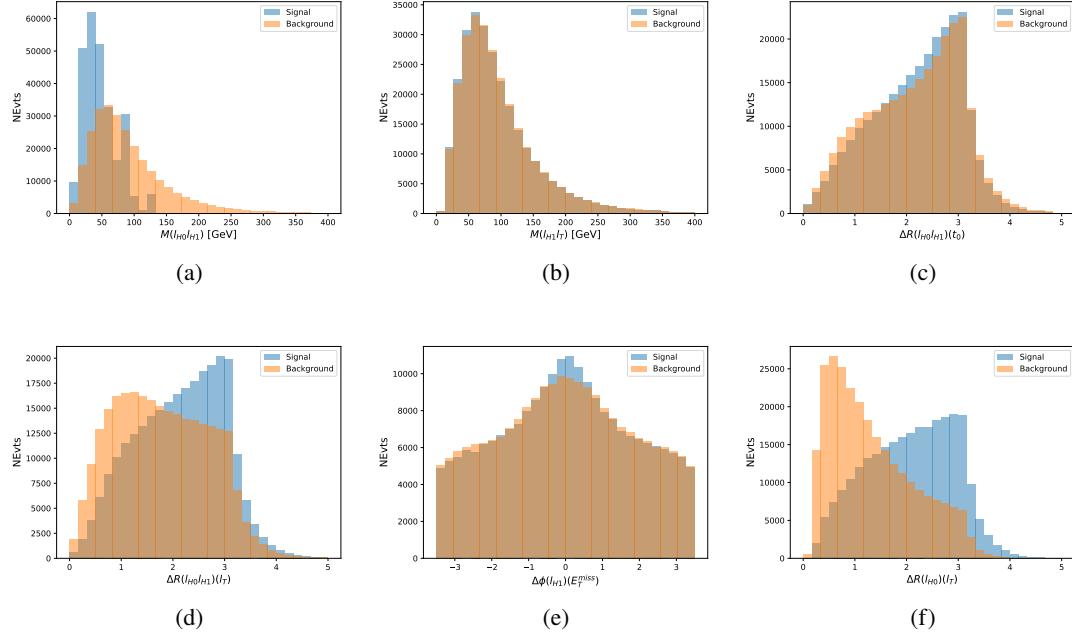


Figure 5.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

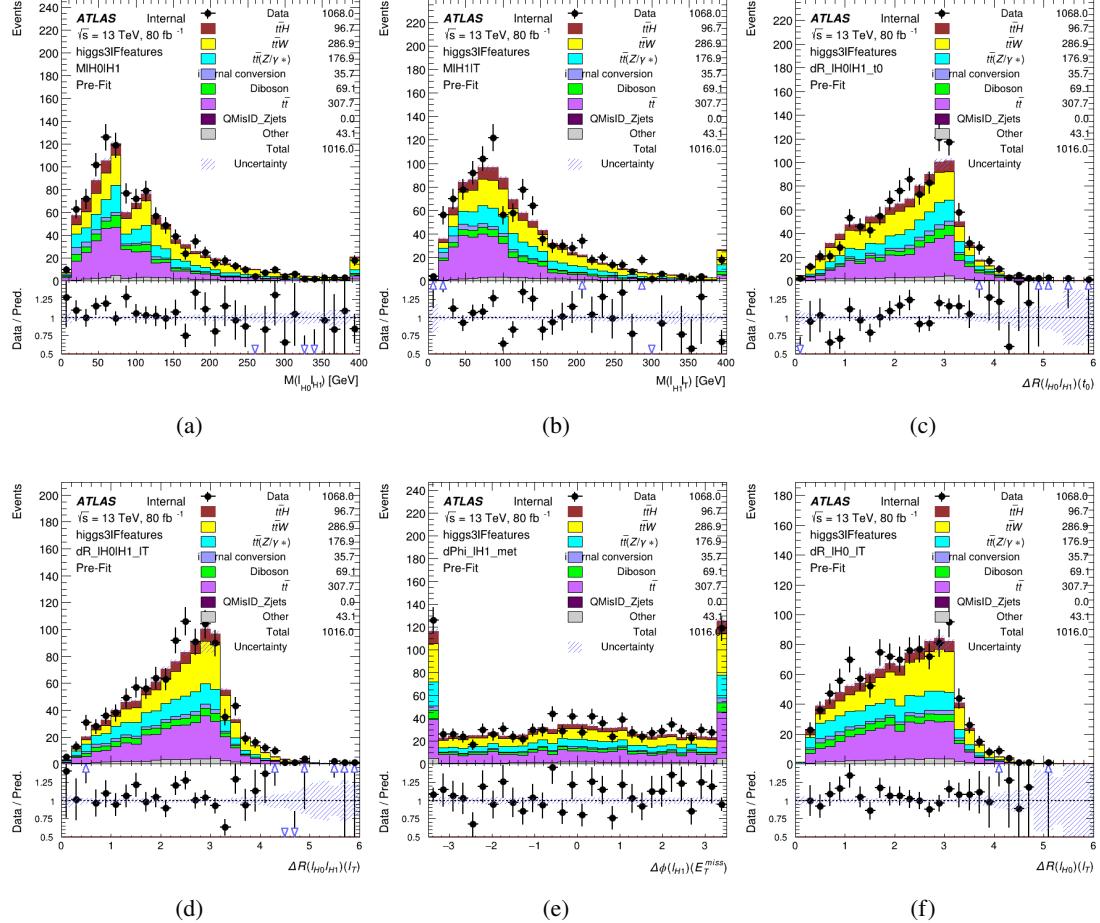


Figure 5.14: Data/MC comparisons of input features for higgs3IF training for 79.8 fb^{-1} of data.

388 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events, with
 389 10% of the dataset reserved for testing. The output of the model is summarized in Figure 5.4.3.

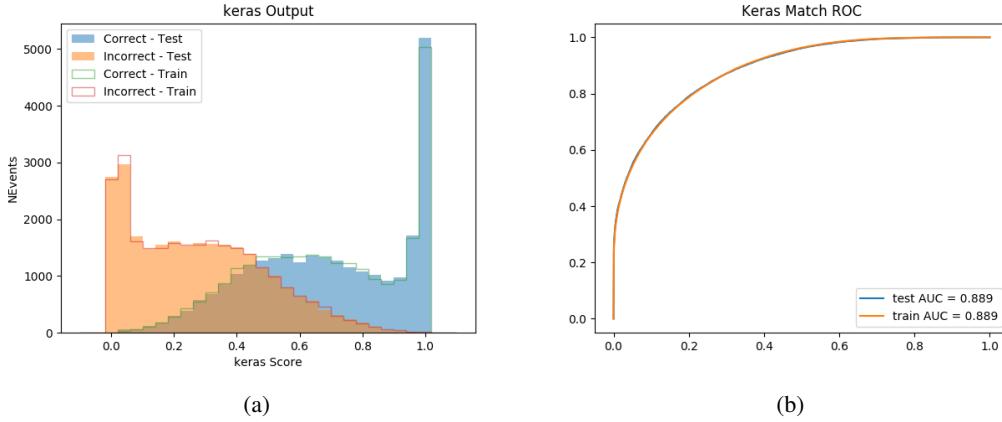


Figure 5.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

391 5.5 p_T Prediction

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

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

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

408 **5.5.1 2ISS Channel**

409 The input variables listed in Table 13 are used to predict the Higgs p_T in the 2ISS channel. Here
 410 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
 411 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
 412 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
 413 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
 414 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 13: Input features for reconstructing the Higgs p_T spectrum for 2ISS events

415 The optimal neural network architecture for this channel is found to consist of 7 hidden layers with
 416 60 nodes each. The input data set includes 1.2 million events, 10% of which is used for testing,
 417 the other 90% for training. Training is found to converge after around 150 epochs.

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

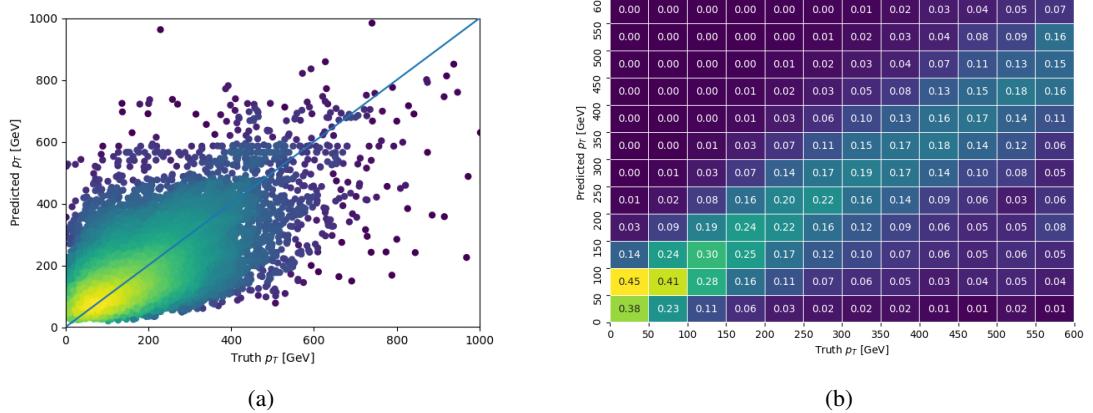


Figure 5.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized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

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

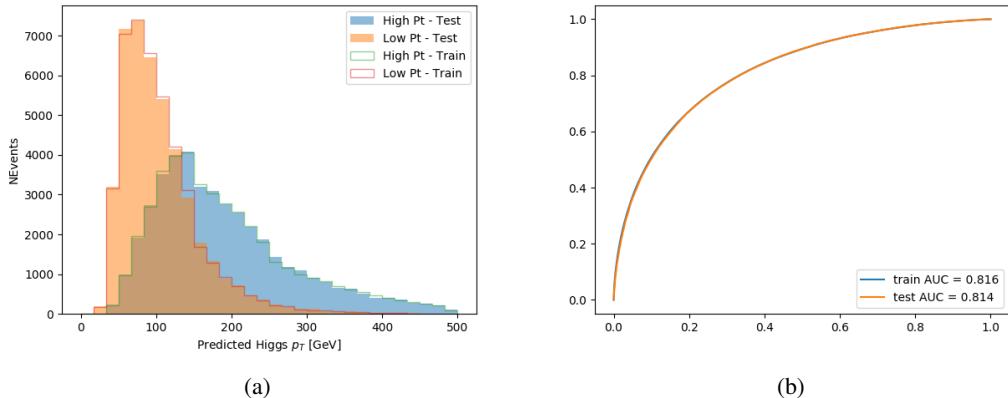


Figure 5.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

428 5.5.2 3l Semi-leptonic Channel

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

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

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

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

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

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

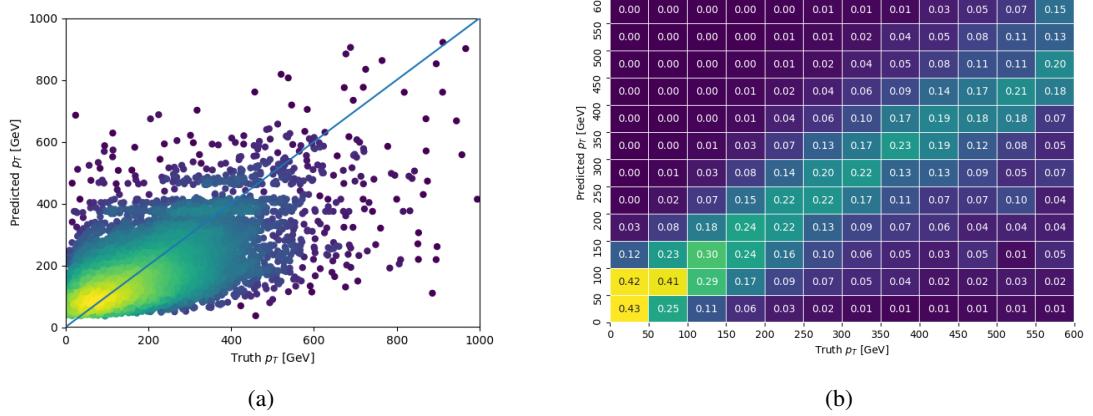


Figure 5.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized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

443 Figure 5.19 shows (a) the output of the NN for events with truth p_T less than and greater than
 444 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low p_T
 445 events.

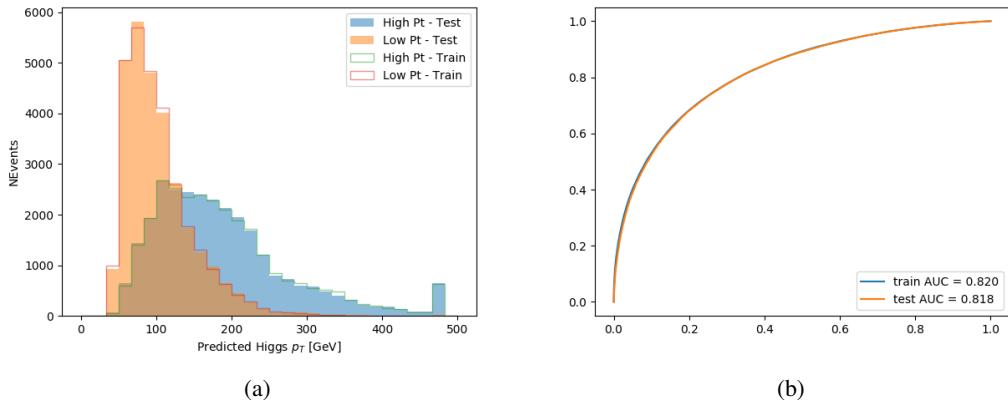


Figure 5.19: (a) shows the reconstructed Higgs p_T for 3lS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

446 5.5.3 3l Fully-leptonic Channel

447 The features listed in 15 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 15: Input features for reconstructing the Higgs p_T spectrum for 3lF events

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

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

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

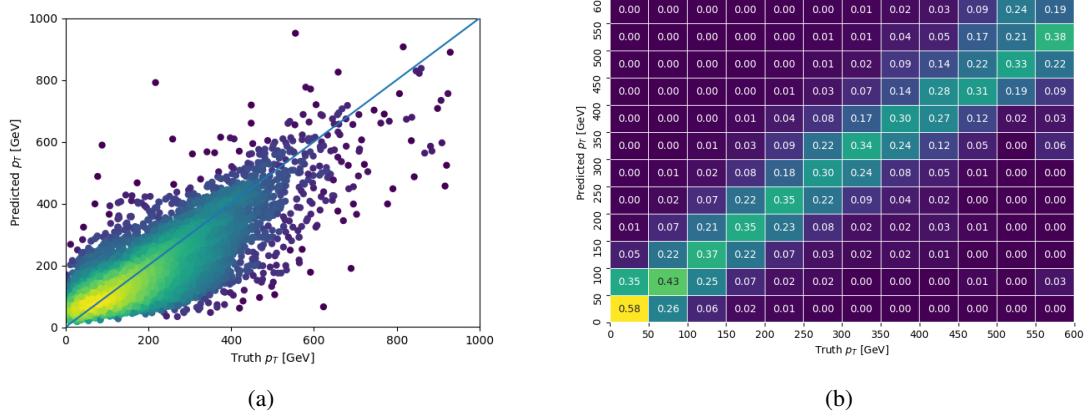


Figure 5.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized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

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

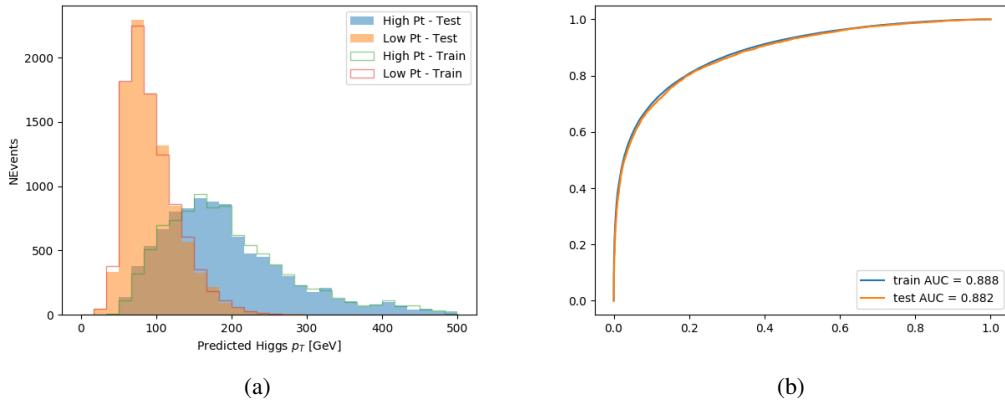


Figure 5.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

5.6 3l Decay Mode

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

461 and the other hadronically, giving two jets and one lepton in the final state (referred to as the
 462 semi-leptonic 3l channel). In order to accurately reconstruct the Higgs, it is necessary to identify
 463 which of these decays took place for each 3l event.

464 The kinematics of each event, along with the output scores of the Higgs and top reconstruction
 465 algorithms, are used to distinguish these two possible decay modes. The particular inputs used
 466 are listed in Table 16.

HT jets	$M(l_0 t_0)$	$M(l_0 t_1)$
$M(l_1 t_0)$	$M(l_1 t_1)$	$M(l_0 l_1)$
$M(l_0 l_2)$	$M(l_1 l_2)$	$\Delta R(l_0 t_0)$
$\Delta R(l_0 t_1)$	$\Delta R(l_1 t_0)$	$\Delta R(l_1 t_1)$
$\Delta R(l_0 l_1)$	$\Delta R(l_0 l_2)$	$\Delta R(l_1 l_2)$
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 16: Input features used to distinguish semi-leptonic and fully-leptonic Higgs decays in the 3l channel.

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

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

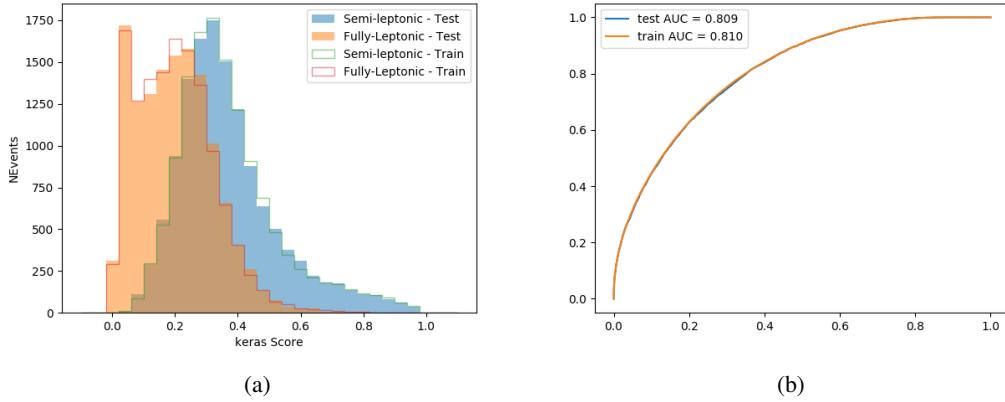


Figure 5.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

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

473 6 Signal Region Definitions

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

477 6.1 Pre-MVA Event Selection

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

- 480 • Two very tight, same-charge, light leptons with $p_T > 20$ GeV
- 481 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets
- 482 • No reconstructed tau candidates

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

Process	Yield	
t̄tH high p _T	41	±5
t̄tH low p _T	71	±8
t̄tW	450	±70
t̄t(Z/γ*)	91	±11
t̄tll low mass	10	±6
Rare Top	20	±12
VV	42	±22
tZ	10	±5
QMisID	44.7	±2.7
Fakes int. conv	47	±26
Fakes ext. conv	46	±44
Fakes HF e	45	±23
Fakes HF μ	250	±50
Three top	2.2	±1.1
Four top	5.64	±0.31
t̄tWW	10.9	±0.6
tW	0.0	±0.0
WtZ	9.1	±0.8
VVV	0.30	±0.05
VH	0.6	±1.0
Total	1170	±120
Data	1108	

Table 17: Yields of the 2lSS preselection region

485 Figure 6.1. Good general agreement is found.

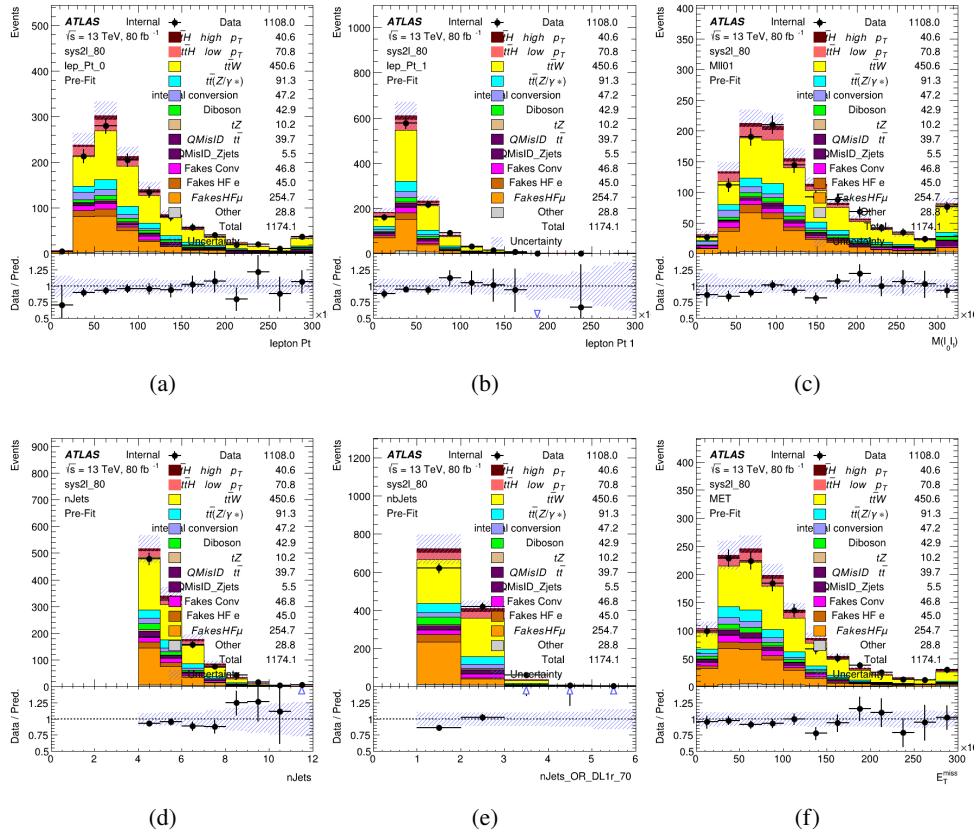


Figure 6.1: Data/MC comparisons of the 2LSS pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

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

- 487 • Three light leptons with total charge ± 1
- 488 • Same charge leptons are required to be very tight, with $p_T > 20$ GeV
- 489 • Opposite charge lepton must be loose, with $p_T > 10$ GeV
- 490 • ≥ 2 reconstructed jets, ≥ 1 b-tagged jets
- 491 • No reconstructed tau candidates
- 492 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs

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

Process	Yield
t̄tH high p _T	20.5 ±2.3
t̄tH low p _T	33.6 ±3.8
t̄tW	138 ±18
t̄tZ/γ	80 ±9
t̄tlowmass	3.5±2.0
rareTop	22 ±12
VV	39 ±19
tZ	9.2±4.5
QMisID	1.8±0.6
Fakes int. conv	31 ±17
Fakes ext. conv	14 ±11
Fakes HF e	20 ±10
Fakes HF μ	102 ±22
Three top	0.96±0.48
Four top	6.17±0.35
t̄tWW	5.46±0.33
tW	0.0±0.0
WtZ	8.7±0.6
VVV	0.81±0.11
VH	0.0±0.0
Total	512 ±48
Data	535

Table 18: Yields of the 3l preselection region.

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

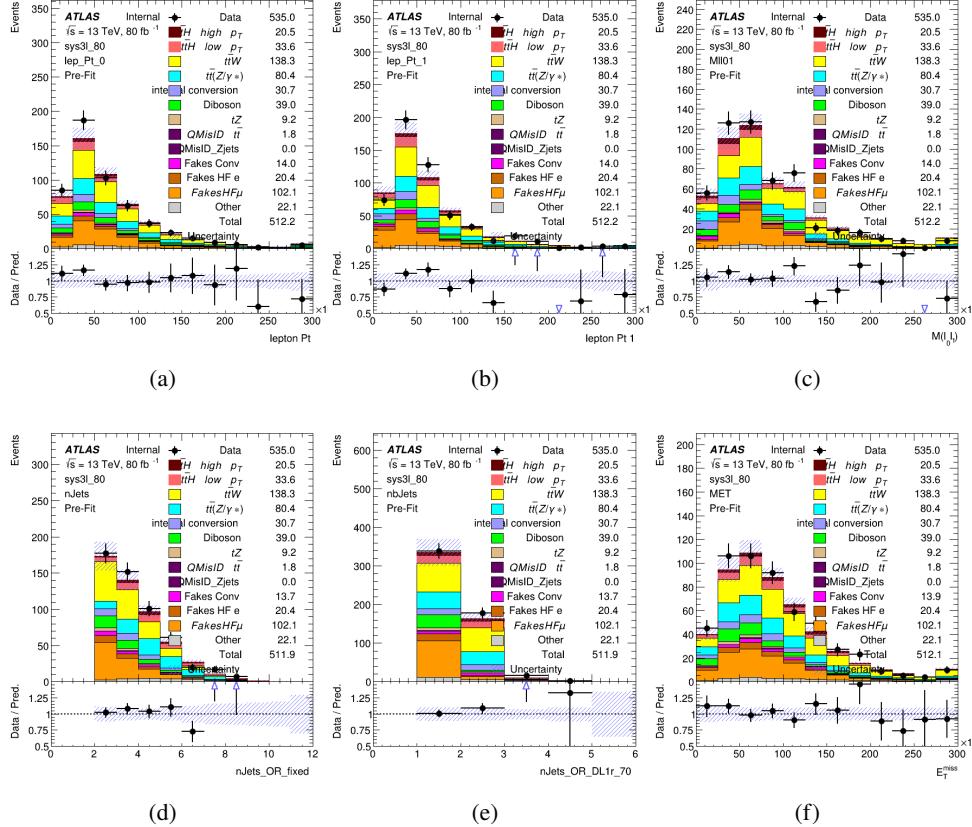


Figure 6.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

496 6.2 Event MVA

497 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
 498 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
 499 (3lF). Here events with three leptons are split into 3lS and 3lF based on the model described in 5.6.
 500 In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost [xgboost]
 501 are trained using the kinematics of signal and background events derived from Monte Carlo
 502 simulations. Events are weighted in the BDT training by the weight of each Monte Carlo event.

503 Because the background composition differs for events with a high reconstructed Higgs p_T
 504 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and
 505 low p_T regions. This is found to provide better significance than attempting to build an inclusive
 506 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6
 507 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
 508 3lF high p_T , and 3lF low p_T .

509 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 19: Input features used to distinguish signal and background events in the 2lSS channel.

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

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

Table 20: Input features used to distinguish signal and background events in the 3l channel.

511 Modelling of each of these input features is verified in Appendix A.2 by comparing data and MC
 512 for 79.8 fb^{-1} . The BDTs are produced with a maximum tree depth of 6, using AUC as the target
 513 loss function. The BDT response distribution and ROC curve for each model is shown in Figures
 514 6.3-6.5.

2LSS

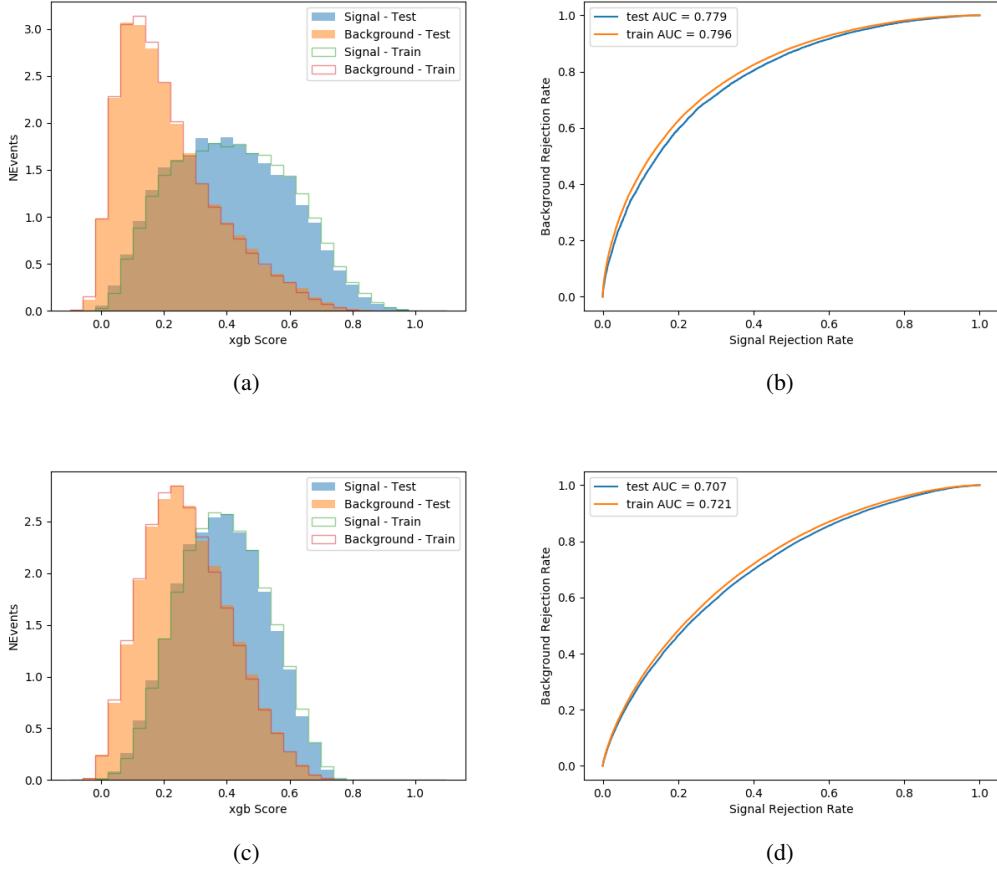


Figure 6.3: Output BDT scores of training and testing data for signal (blue) and background (orange) for 2LSS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 2LSS high and low p_T models, respectively.

3l - Semileptonic

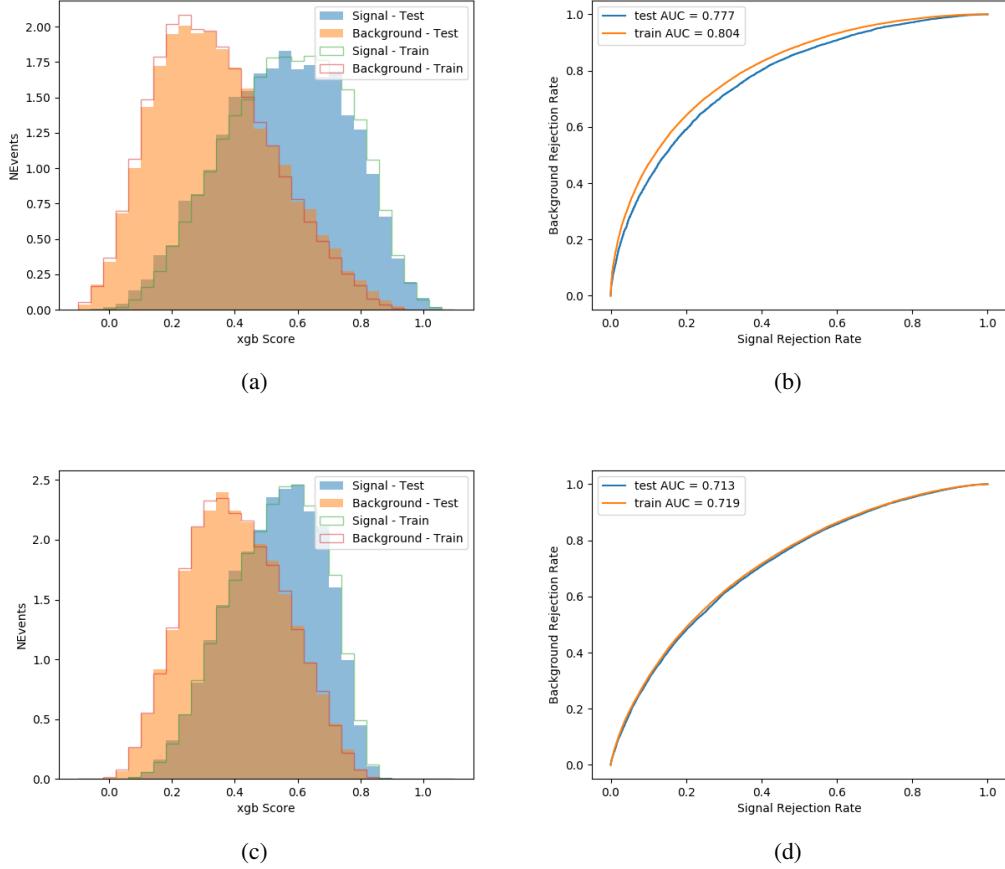


Figure 6.4: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lS high and low p_T models, respectively.

3l - Fully Leptonic

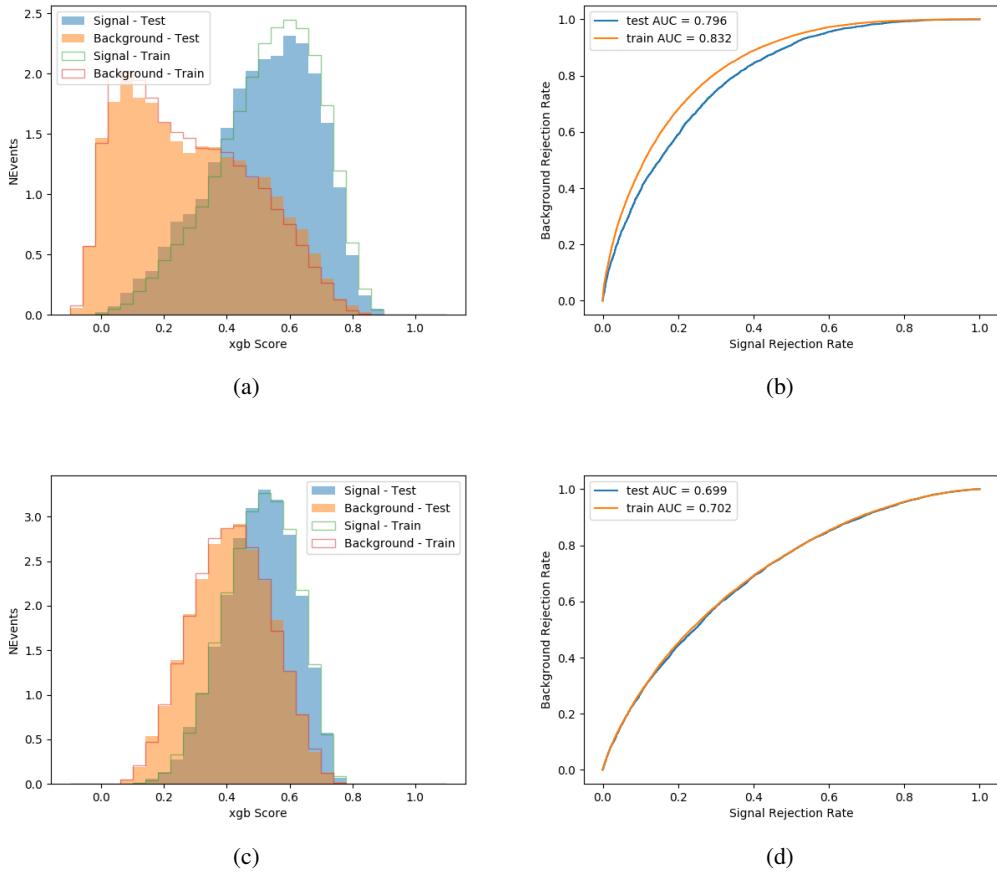


Figure 6.5: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lF events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lF high and low p_T models, respectively.

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

517 6.3 Signal Region Definitions

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

High p_T Background Rejection BDTs

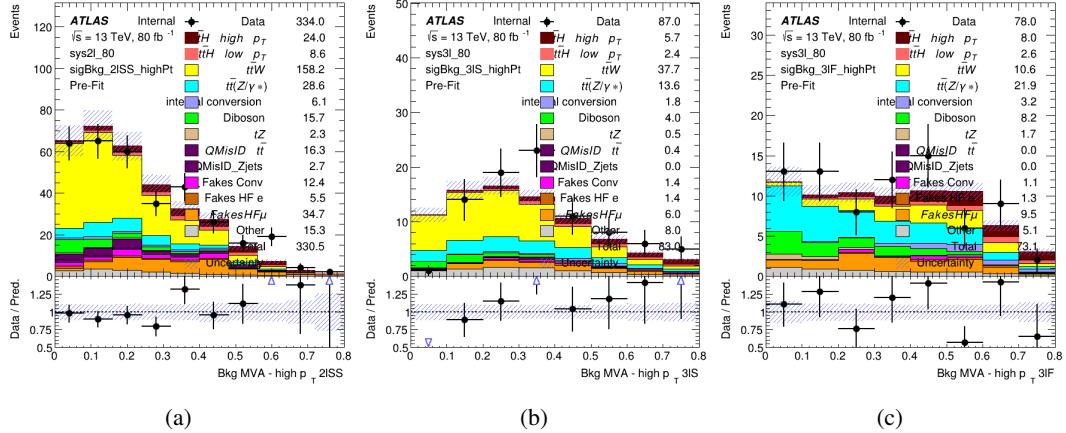


Figure 6.6: Output score of the high p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

Low p_T Background Rejection BDTs

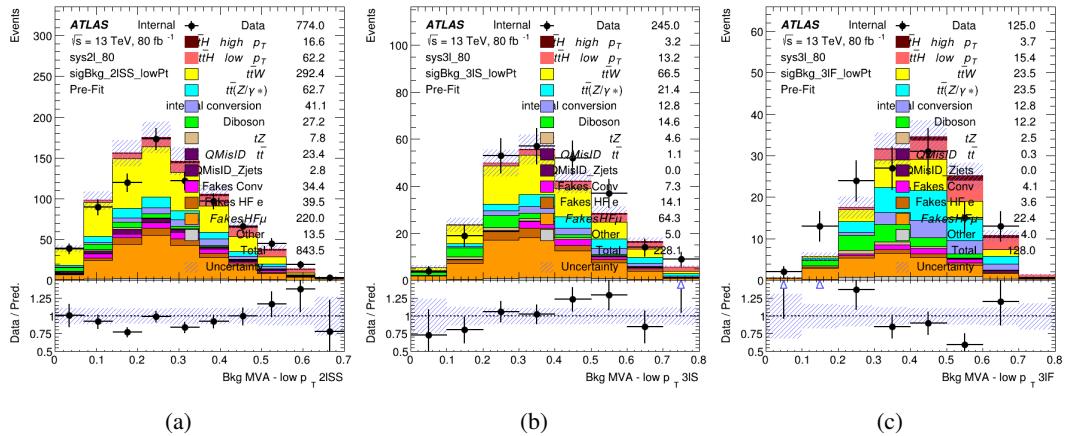


Figure 6.7: Output score of the low p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

522 For each event, depending on the number of leptons as well as whether the p_T of the Higgs is
 523 predicted to be high (> 150 GeV) or low (< 150 GeV), a cut on the appropriate background
 524 rejection MVA is applied. The particular cut values, listed in Table 21, are determined by
 525 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 21: Cutoff values on background rejection MVA score applied to signal regions.

526 The event preselection and MVA selection listed in Table 21 are used define the three signal
 527 regions used in the fit. These signal region definitions are summarized in Table 22.

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

528 7 Systematic Uncertainties

529 The systematic uncertainties that are considered are summarized in Table 23. These are
 530 implemented in the fit either as a normalization factors or as a shape variation or both in the signal
 531 and background estimations. The numerical impact of each of these uncertainties is outlined in
 532 section 8.

Table 23: 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

533 The uncertainty in the combined 2015+2016 integrated luminosity is derived from a calibration
 534 of the luminosity scale using x-y beam-separation scans performed in August 2015 and May 2016
 535 [17].

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

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

540 The sources which contribute to the uncertainty in the jet energy scale [18] are decomposed into
541 uncorrelated components and treated as independent sources in the analysis. The CategoryReduction
542 model is used to account for JES uncertainties, which decomposes the uncertainties into 30
543 nuisance parameters included in the fit. The SimpleJER model is used to account for jet energy
544 resolution (JER) uncertainties, and 8 JER uncertainty components unclued as NPs in the fit.

545 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses [19] are
546 also decomposed into uncorrelated components. The large number of components for b-tagging
547 is due to the calibration of the distribution of the BDT discriminant.

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

550 The full list of systematic uncertainties considered in the analysis is summarized in Tables 24, 25
551 and 26.

552

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 24: 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 25: 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 26: 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.

553 As mentioned in Section 3.2, a normalization corrections and uncertainties on the estimates of
 554 non-prompt leptons backgrounds are derived using data driven techniques, described in detail in
 555 [10]. These are derived from a likelihood fit over various non-prompt enriched control regions,
 556 targeting several sources of non-prompt light leptons separately: external conversion electrons,
 557 internal conversion electrons, electrons from heavy flavor decays, and muons from heavy flavor
 558 decays.

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

Processs	Normalization Factor
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

Table 27: Normalization factors - with statistical and systematic uncertainties - derived from the fit over fake control regions for each source of non-prompt leptons considered.

561 In addition to those derived from the control regions, several additional uncertainties are assigned
 562 to the non-prompt lepton background. An additional 25% uncertainty on material conversions is
 563 assigned, based on the comparison between data and MC in a region where a loose electron fails
 564 the photon conversion veto. A shape uncertainty of 15% (6%) is assigned to the HF non-prompt

565 electron (muon) background based on a comparison between data and MC where the second
 566 leading electron (muon) is only required to be loose. As the contribution from light non-prompt
 567 leptons is small, about 10% percent of the contribution from HF non-prompt leptons, it is derived
 568 from the agreement between data and simulation in a LF enriched region at low values of the
 569 non-prompt lepton BDT. The resulting uncertainty is 100%, and is taken to be uncorrelated
 570 between internal and material conversions.

571 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and scale
 572 uncertainties are taken from theory calculations for the predominate prompt backgrounds.
 573 Following the nominal $t\bar{t}H$ – ML analysis, a 50% uncertainty is applied to Diboson to account
 574 for the large uncertainty in estimating $VV + \text{heavy flavor}$. The other “rare” background processes
 575 - including tZ , rare top processes, $ttWW$, WtZ , VVV , $tHjb$ and WtH - are assigned an overall
 576 50% normalization uncertainty as well. The theory uncertainties applied to the MC estimates are
 577 summarized in Table 28.

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

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

582 8 Results

583 A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p_T spectrum
 584 in the three signal regions, 2IS, 3IS, and 3IF. The signal is split into high and low p_T samples,
 585 based on whether the truth p_T of the Higgs is above or below 150 GeV. The parameters $\mu_{t\bar{t}H\text{high}p_T}$
 586 and $\mu_{t\bar{t}H\text{low}p_T}$, where $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$, are extracted from the fit, signifying the difference

587 between the observed value and the theory prediction. Unblinded results are shown for the 79.8
 588 fb^{-1} data set, as well as MC only projections of results using the full Run-2, 139 fb^{-1} dataset.

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

594 8.1 Results - 79.8 fb^{-1}

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

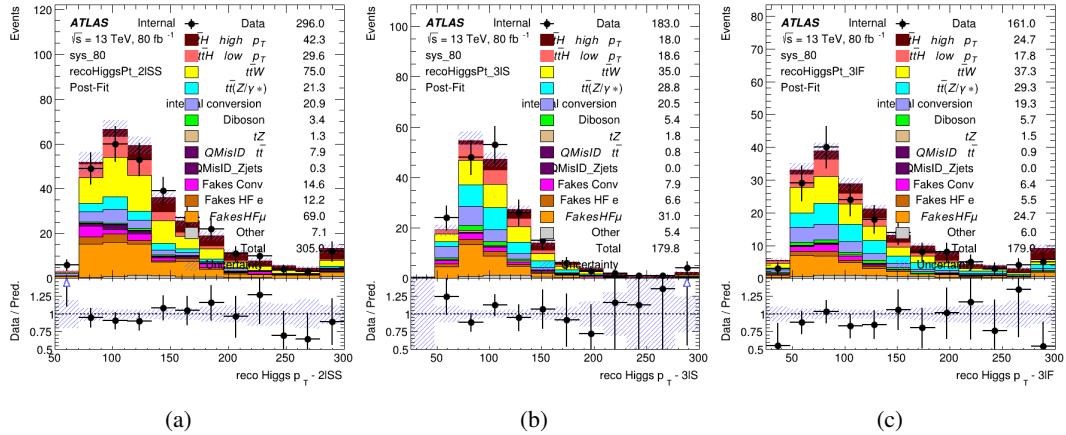


Figure 8.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2ISS, (b) 3IS, and (c) 3IF, for 79.8 fb^{-1} of MC

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

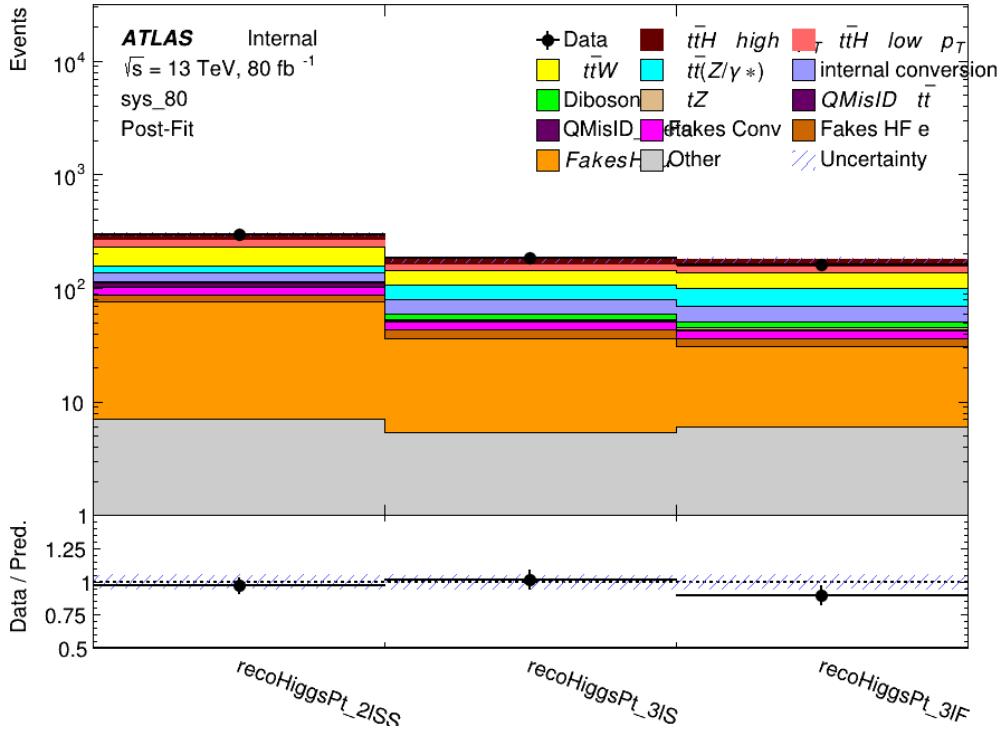


Figure 8.2: Post-fit summary of the yields in each signal region.

599 The measured μ values for high and low p_T Higgs production obtained from the fit are
 600 shown in 29. A significance of 1.7σ is observed for $t\bar{t}H$ high p_T , and 2.1σ is measured for
 601 $t\bar{t}H$ 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 29: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

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

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

604 The most significant sources of uncertainty on the measurement of t̄H - low p_T are shown in
 605 Table 31.

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

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

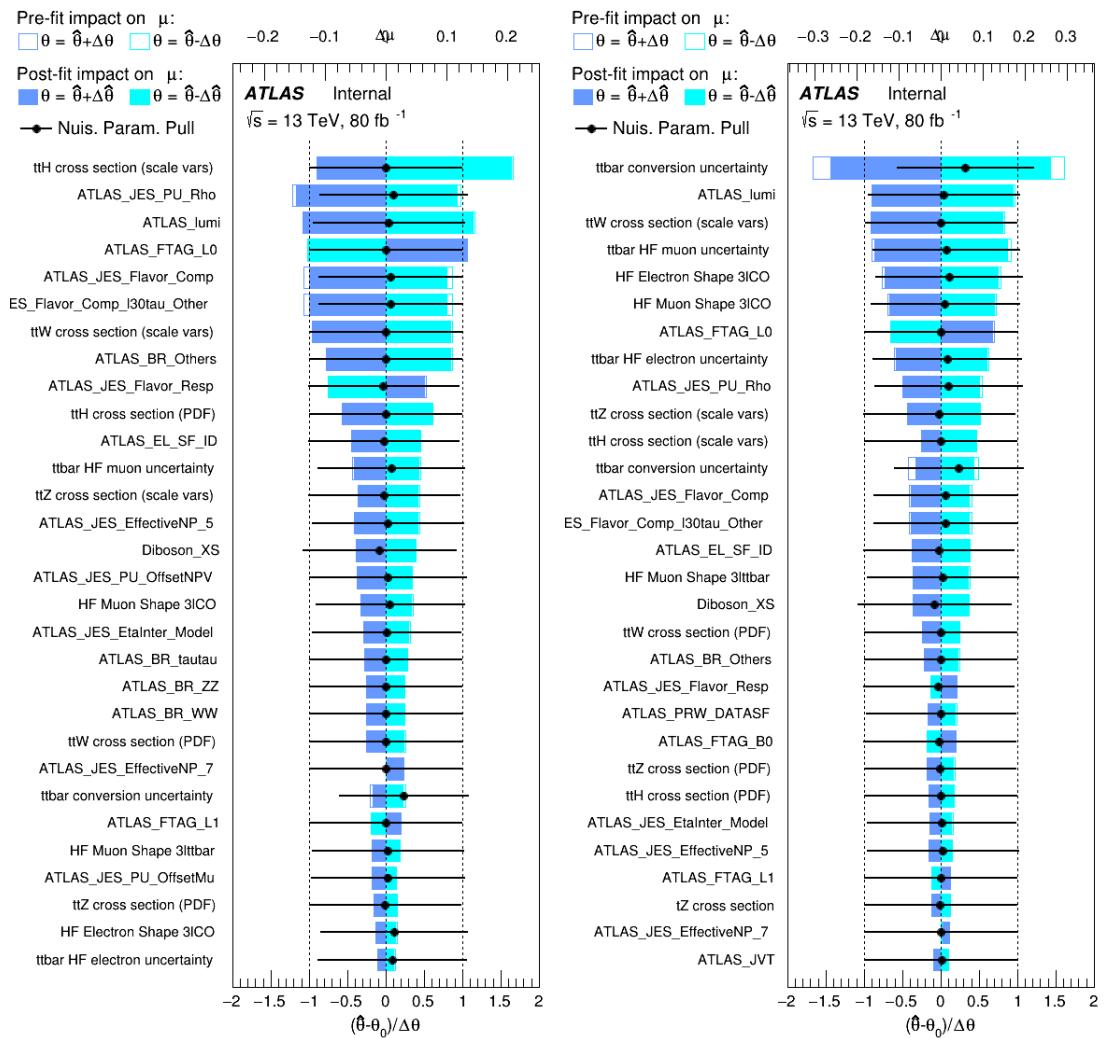


Figure 8.3: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

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

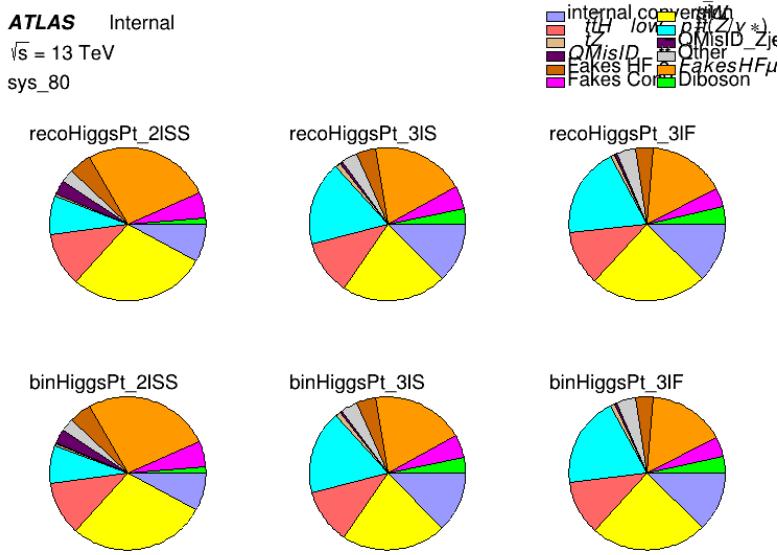
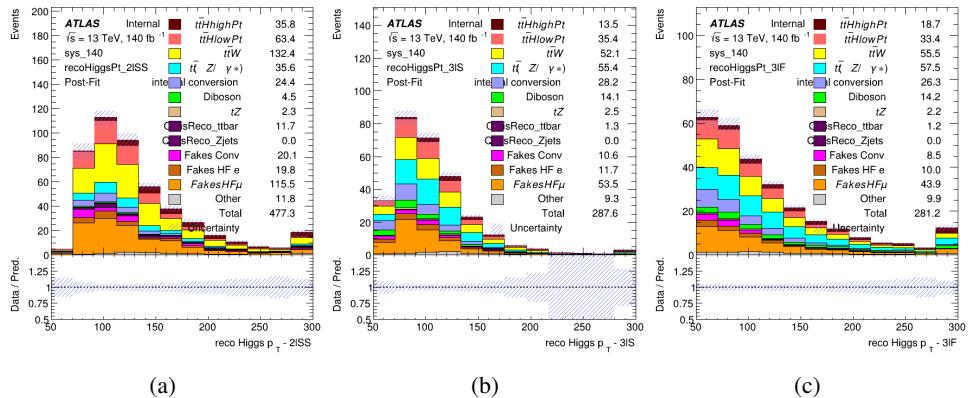


Figure 8.4: Background composition of the fit regions.

609 8.2 Projected Results - 139 fb^{-1}

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

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

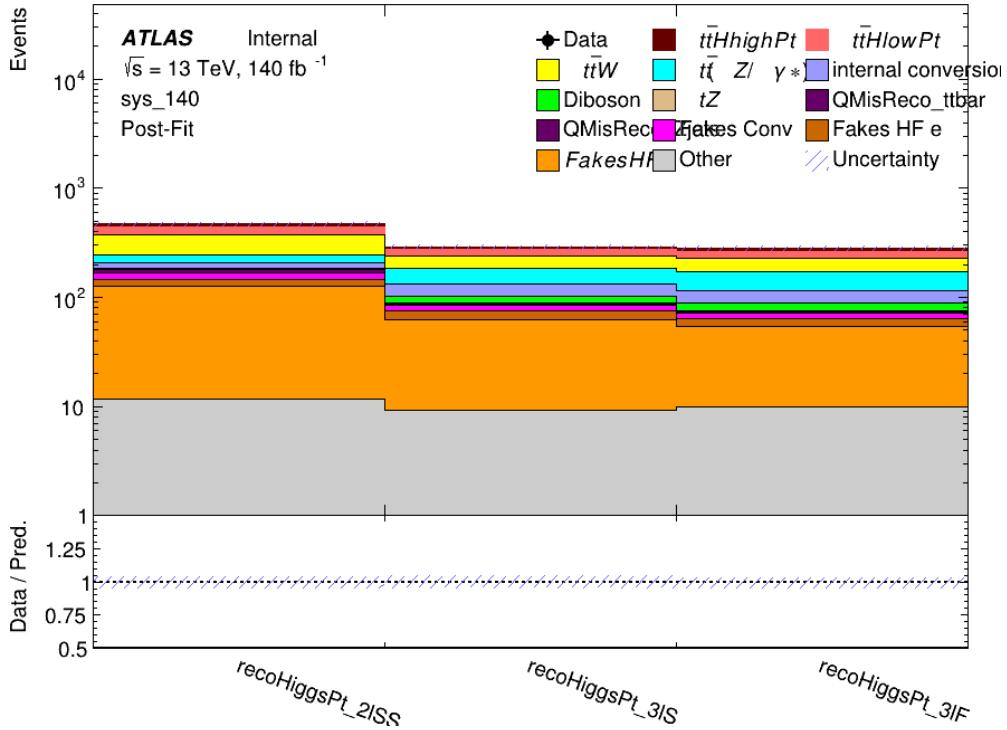


Figure 8.6: Post-fit summary of fit.

614 Projected uncertainties on the μ values extracted from the fit for high and low p_T Higgs are shown
 615 in 32. A significance of 2.0σ is expected for $t\bar{t}H$ high p_T , and a projected significance 2.3σ is
 616 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 32: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

617 The most prominent sources of systematic uncertainty, as measured by their impact on $\mu_{t\bar{t}H \text{ high } p_T}$,
 618 are summarized in Table 33.

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

619 The most significant sources of systematic uncertainty on t̄H low p_T are summarized in Table
 620 34.

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

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

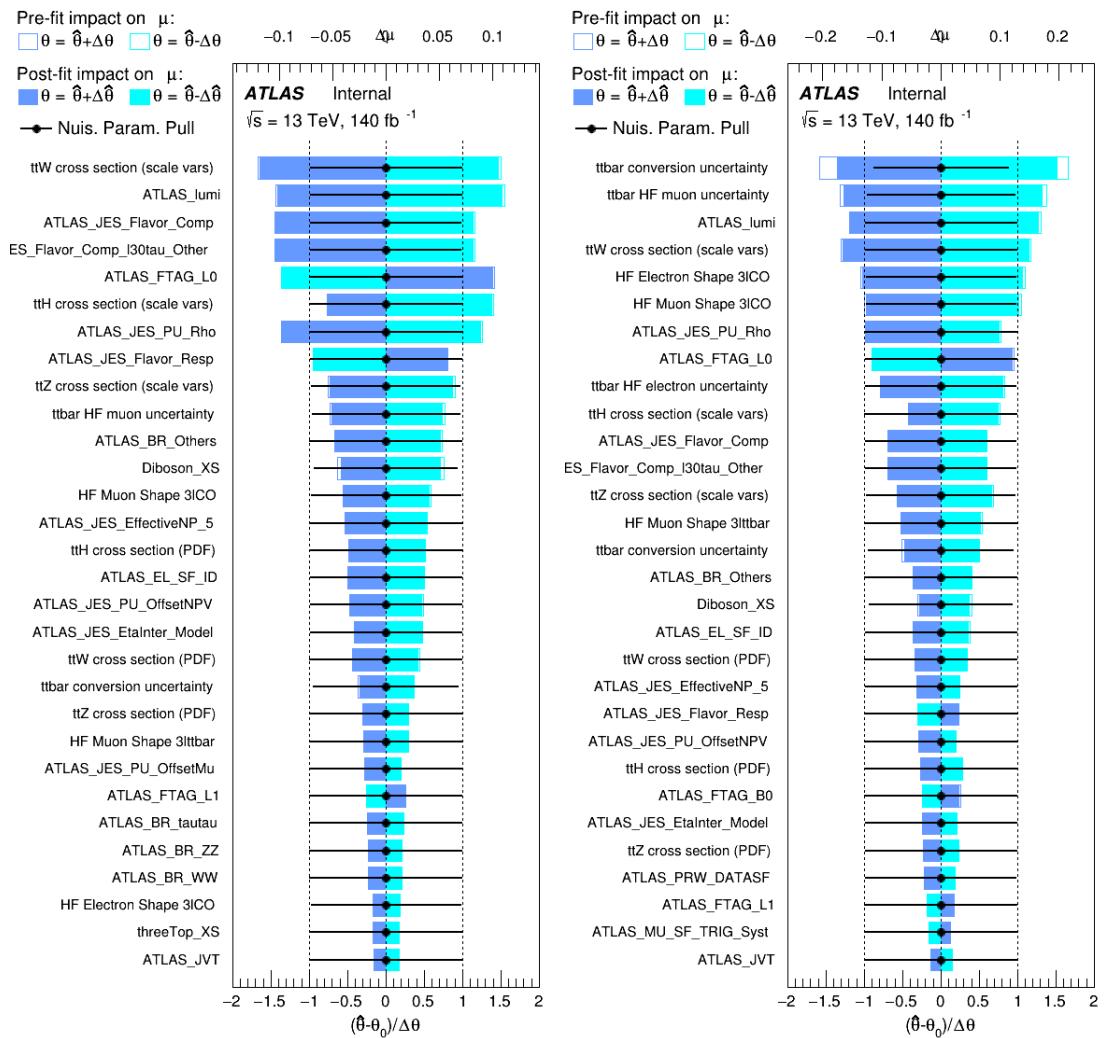


Figure 8.7: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

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

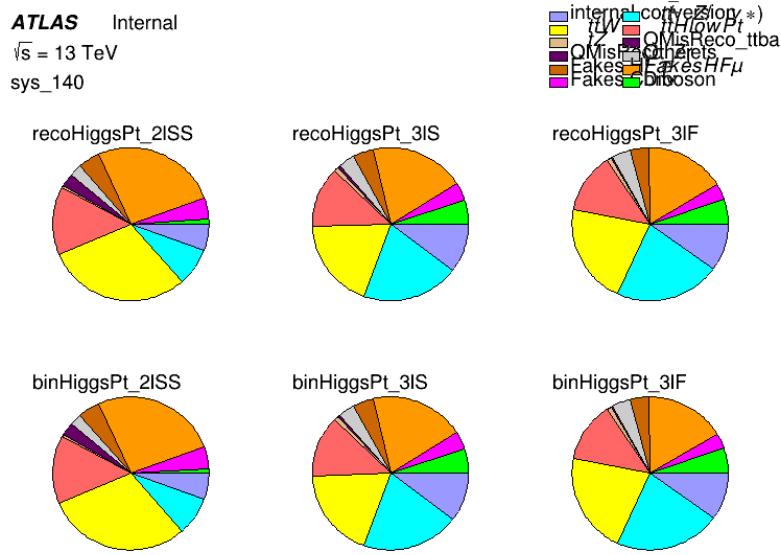


Figure 8.8: Background composition of the fit regions.

624 9 Conclusion

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

629 References

- 630 [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model*
 631 *Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, arXiv:
 632 [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214) (cit. on p. 5).
- 633 [2] B. Dumont, S. Fichet and G. von Gersdorff, *A Bayesian view of the Higgs sector with higher*
 634 *dimensional operators*, *Journal of High Energy Physics* **2013** (2013), ISSN: 1029-8479,
 635 URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)065](http://dx.doi.org/10.1007/JHEP07(2013)065) (cit. on p. 5).
- 636 [3] S. Banerjee, S. Mukhopadhyay and B. Mukhopadhyaya, *Higher dimensional operators and*
 637 *the LHC Higgs data: The role of modified kinematics*, *Physical Review D* **89** (2014), ISSN:
 638 1550-2368, URL: <http://dx.doi.org/10.1103/PhysRevD.89.053010> (cit. on p. 5).
- 639 [4] R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, arXiv:
 640 [1410.8849 \[hep-ph\]](https://arxiv.org/abs/1410.8849) (cit. on p. 7).
- 641 [5] H.-L. Lai et al., *New parton distributions for collider physics*, *Phys. Rev. D* **82** (2010) 074024,
 642 arXiv: [1007.2241 \[hep-ph\]](https://arxiv.org/abs/1007.2241) (cit. on p. 7).
- 643 [6] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo*
 644 *for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126, arXiv: [0707.3088 \[hep-ph\]](https://arxiv.org/abs/0707.3088)
 645 (cit. on p. 7).
- 646 [7] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG*
 647 *method*, *Eur. Phys. J. C* **71** (2011) 1547, arXiv: [1009.2450 \[hep-ph\]](https://arxiv.org/abs/1009.2450) (cit. on p. 7).
- 648 [8] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions from Powheg*
 649 *and Sherpa to ATLAS measurements of top pair production at 7 TeV*, ATL-PHYS-PUB-
 650 2015-011, 2015, URL: <https://cds.cern.ch/record/2020602> (cit. on p. 7).
- 651 [9] ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11*, ATL-PHYS-
 652 PUB-2011-009, 2011, URL: <https://cds.cern.ch/record/1363300> (cit. on p. 7).
- 653 [10] *Evidence for the associated production of the Higgs boson and a top quark pair with*
 654 *the ATLAS detector*, tech. rep. ATLAS-CONF-2017-077, CERN, 2017, URL: <https://cds.cern.ch/record/2291405> (cit. on pp. 7, 8, 10, 54).
- 656 [11] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using*
 657 *the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, URL: <https://cds.cern.ch/record/2157687> (cit. on p. 10).
- 659 [12] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the*
 660 *ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, *Eur. Phys. J. C*
 661 **74** (2014) 3130, arXiv: [1407.3935 \[hep-ex\]](https://arxiv.org/abs/1407.3935) (cit. on p. 10).
- 662 [13] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed*
 663 *in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-015, 2015, URL: <https://cds.cern.ch/record/2037613> (cit. on p. 11).
- 665 [14] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton–proton collisions with*
 666 *the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: <https://cds.cern.ch/record/2037702> (cit. on p. 11).

- 668 [15] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions*
669 *at $\sqrt{s} = 8 \text{ TeV}$ using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: [1510.03823](https://arxiv.org/abs/1510.03823) [hep-ex] (cit. on p. 11).
- 671 [16] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
672 *with the ATLAS detector in the first proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$* , ATL-
673 PHYS-PUB-2015-027, 2015, URL: <https://cds.cern.ch/record/2037904> (cit. on
674 p. 11).
- 675 [17] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ using the*
676 *ATLAS detector at the LHC*, Eur. Phys. J. C **71** (2011) 1630, arXiv: [1101.2185](https://arxiv.org/abs/1101.2185) [hep-ex]
677 (cit. on p. 50).
- 678 [18] G. Aad, T. Abajyan, B. Abbott and et al, *Jet energy resolution in proton-proton collisions at*
679 *$\sqrt{s} = 7 \text{ TeV}$ recorded in 2010 with the ATLAS detector*, The European Physical Journal
680 C **73** (2013) 2306, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-013-2306-0> (cit. on p. 51).
- 682 [19] A. Collaboration, *Performance of b -jet identification in the ATLAS experiment*, Journal of
683 Instrumentation **11** (2016) P04008, URL: <http://stacks.iop.org/1748-0221/11/i=04/a=P04008> (cit. on p. 51).
- 685 [20] R. Narayan et al., *Measurement of the total and differential cross sections of a top-quark*
686 *antiquark pair in association with a W boson in proton-proton collisions at a centre-of-mass*
687 *energy of 13 TeV with ATLAS detector at the Large Hadron Collider*, tech. rep. ATL-COM-
688 PHYS-2020-217, CERN, 2020, URL: <https://cds.cern.ch/record/2712986> (cit. on
689 p. 68).

690 **List of contributions**

691

692 Appendices

693 .1 Non-prompt lepton MVA

694 A lepton MVA has been developed to better reject non-prompt leptons than standard cut
 695 based selections based upon impact parameter, isolation and PID. The name of this MVA is
 696 `PromptLeptonIso`. The full set of studies and detailed explanation can be found in [20].

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

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

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

717 `PromptLeptonIso` is a gradient boosted BDT. The training of the BDT is performed on leptons
 718 selected from the Powheg+Pythia6 non-allhad $t\bar{t}$ MC sample. Eight variables are used to train
 719 the BDT in order to discriminate between prompt and non-prompt leptons. The track jets that
 720 are matched to the non-prompt leptons correspond to jets initiated by b or c quarks, and may
 721 contain a displaced vertex. Consequently, three of the selected variables are used to identify
 722 b-tag jets by standard ATLAS flavour tagging algorithms. Two variables use the relationship
 723 between the track jet and lepton: the ratio of the lepton p_T with respect to the track jet p_T and
 724 ΔR between the lepton and the track jet axis. Finally three additional variables test whether the
 725 reconstructed lepton is isolated: the number of tracks collected by the track jet and the lepton
 726 track and calorimeter isolation variables. Table 35 describes the variables used to train the BDT
 727 algorithm. The choice of input variables has been extensively discussed with Egamma, Muon,
 728 Tracking, and Flavour Tagging CP groups.

729 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
$\Delta R(\text{lepton}, \text{track jet})$	The ratio of the lepton p_T and the track jet p_T
$p_T^{\text{VarCone30}}/p_T$	ΔR between the lepton and the track jet axis
$E_T^{\text{TopoCone30}}/p_T$	Lepton track isolation, with track collecting radius of $\Delta R < 0.3$

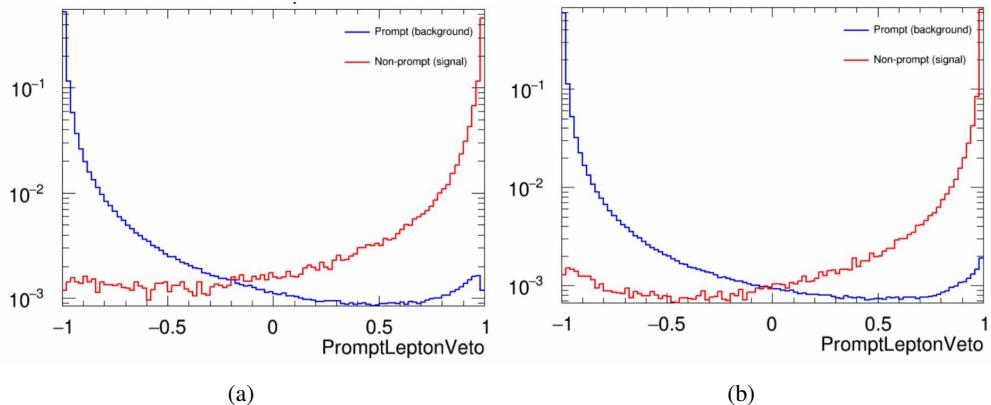
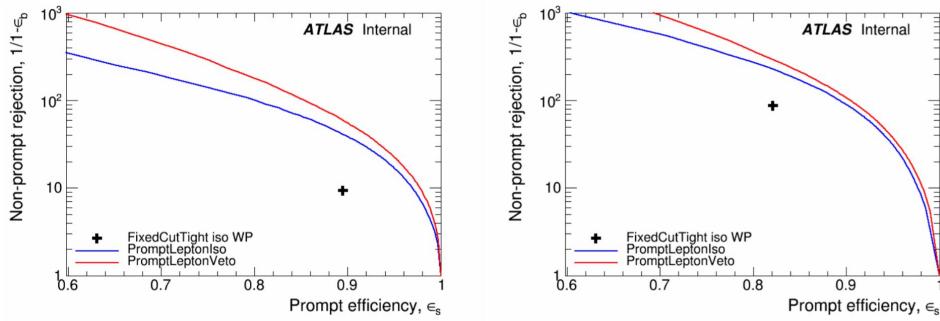
Table 35: A table of the variables used in the training of `PromptLeptonIso`.

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

730 The ROC curve for the BDT response, compared to the standard `FixedCutTight` WP, is shown
 731 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

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

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

743 **A Machine Learning Models**

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

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

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

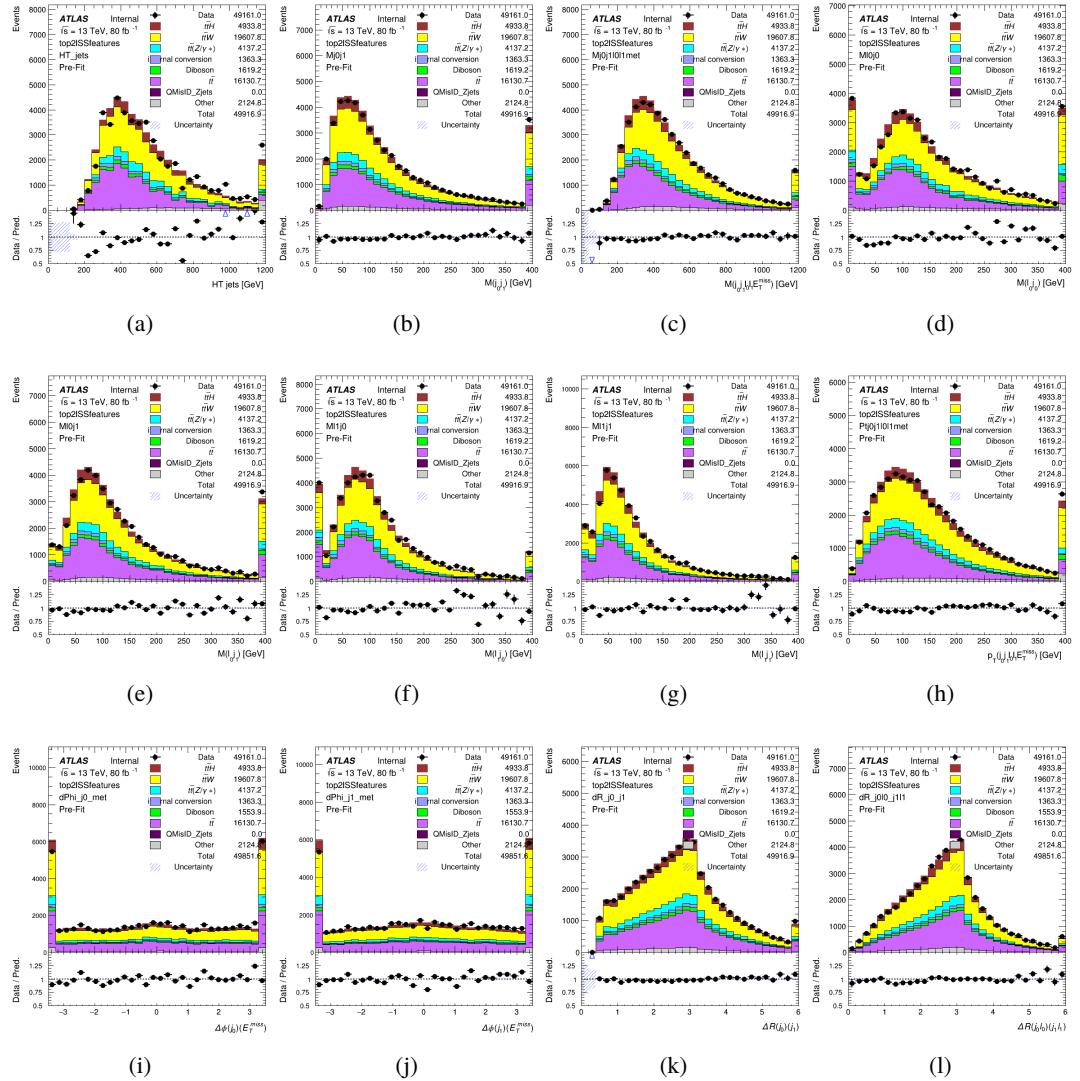


Figure A.1: Input features for top2lSS

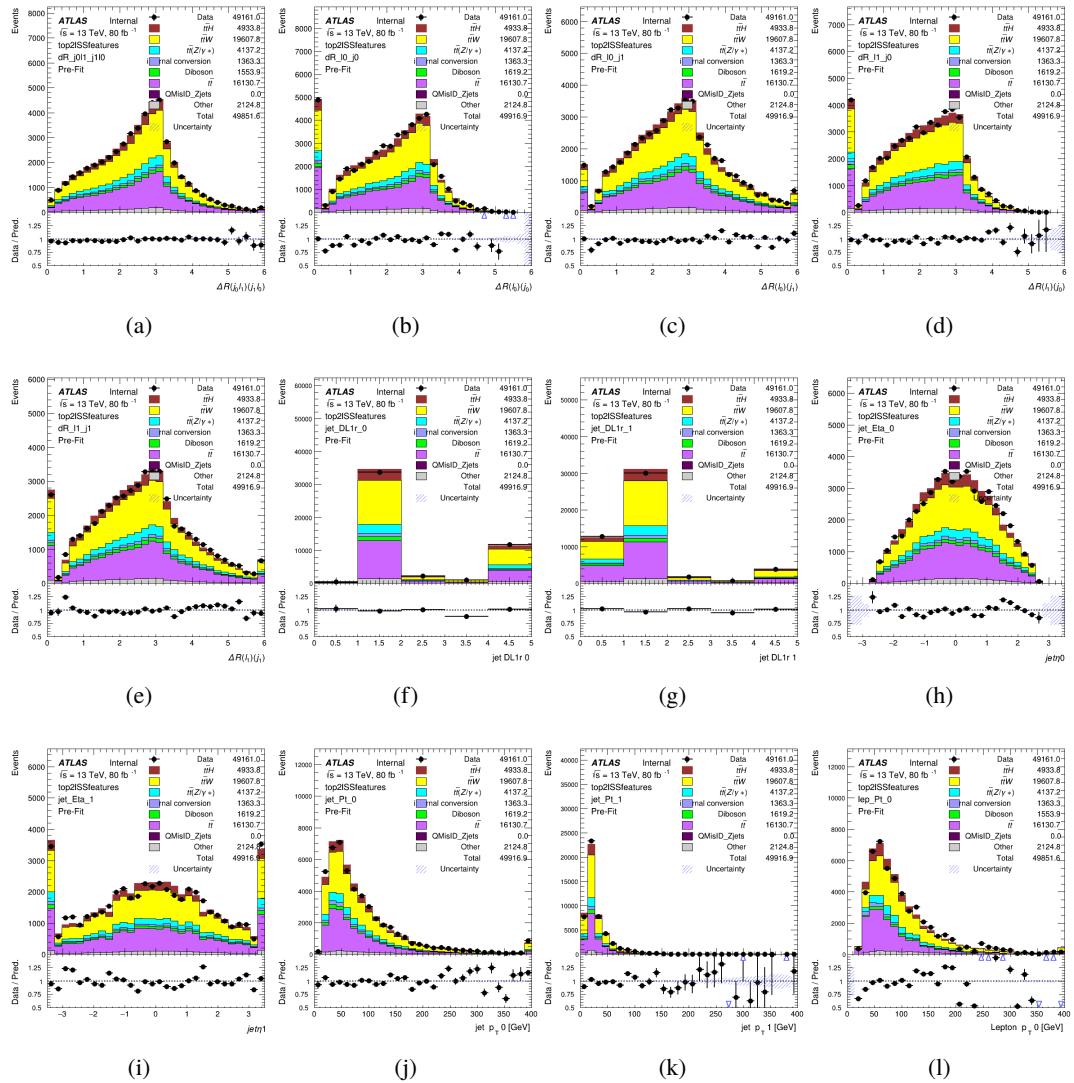


Figure A.2: Input features for top21SS

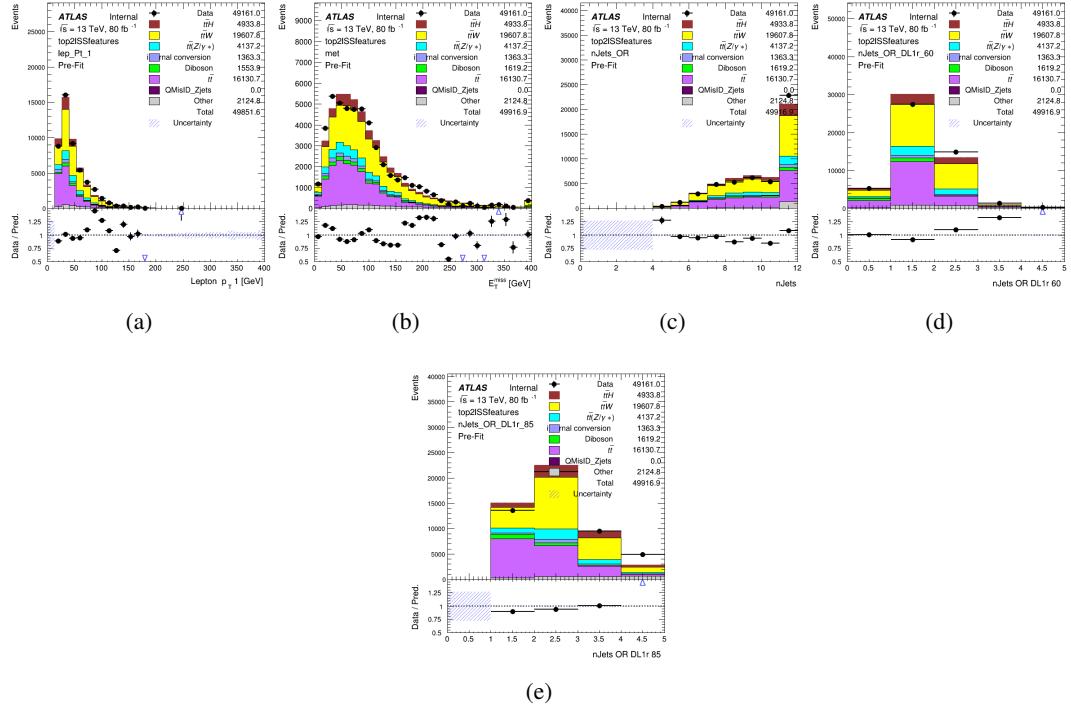


Figure A.3: Input features for top2ISS

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

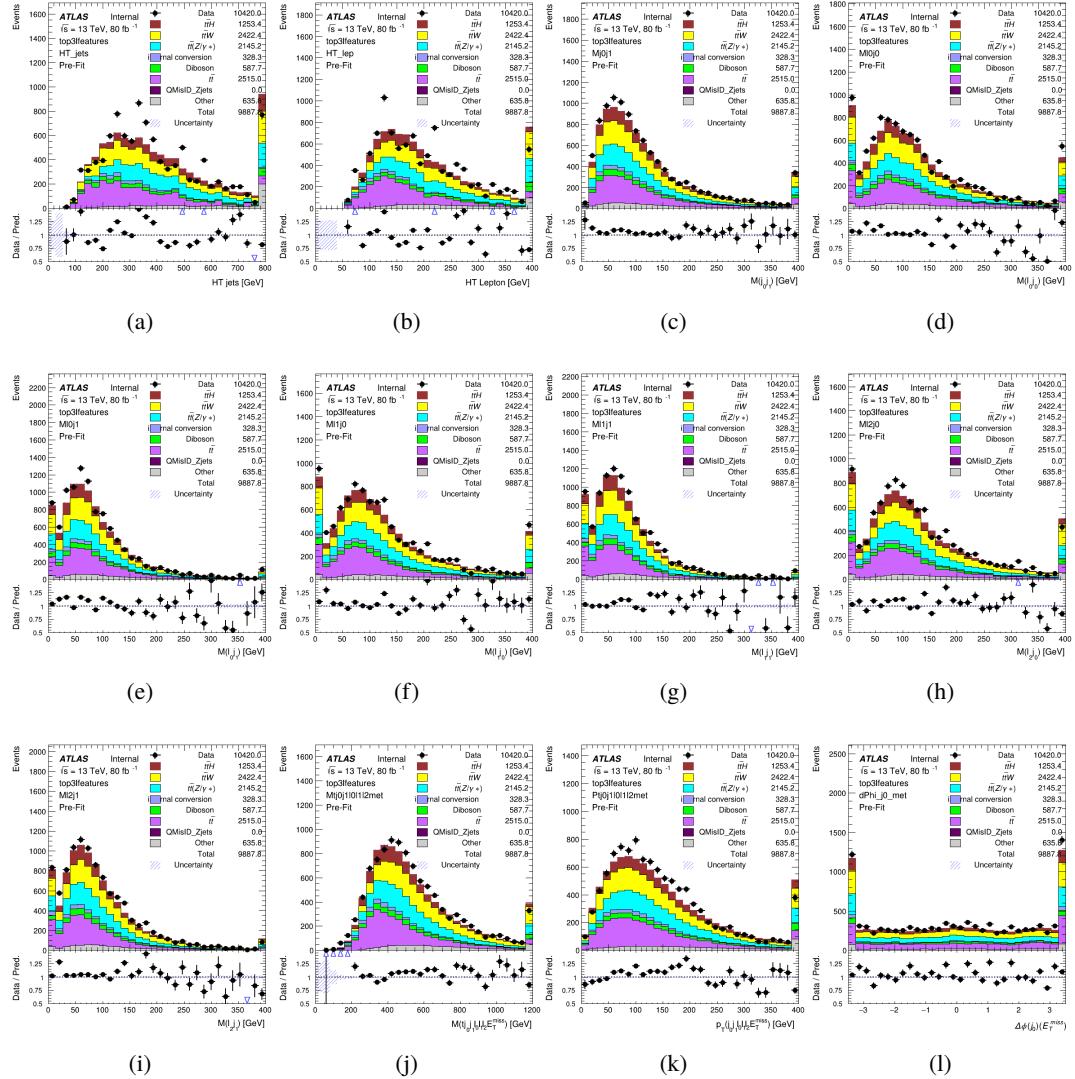


Figure A.4: Input features for top31

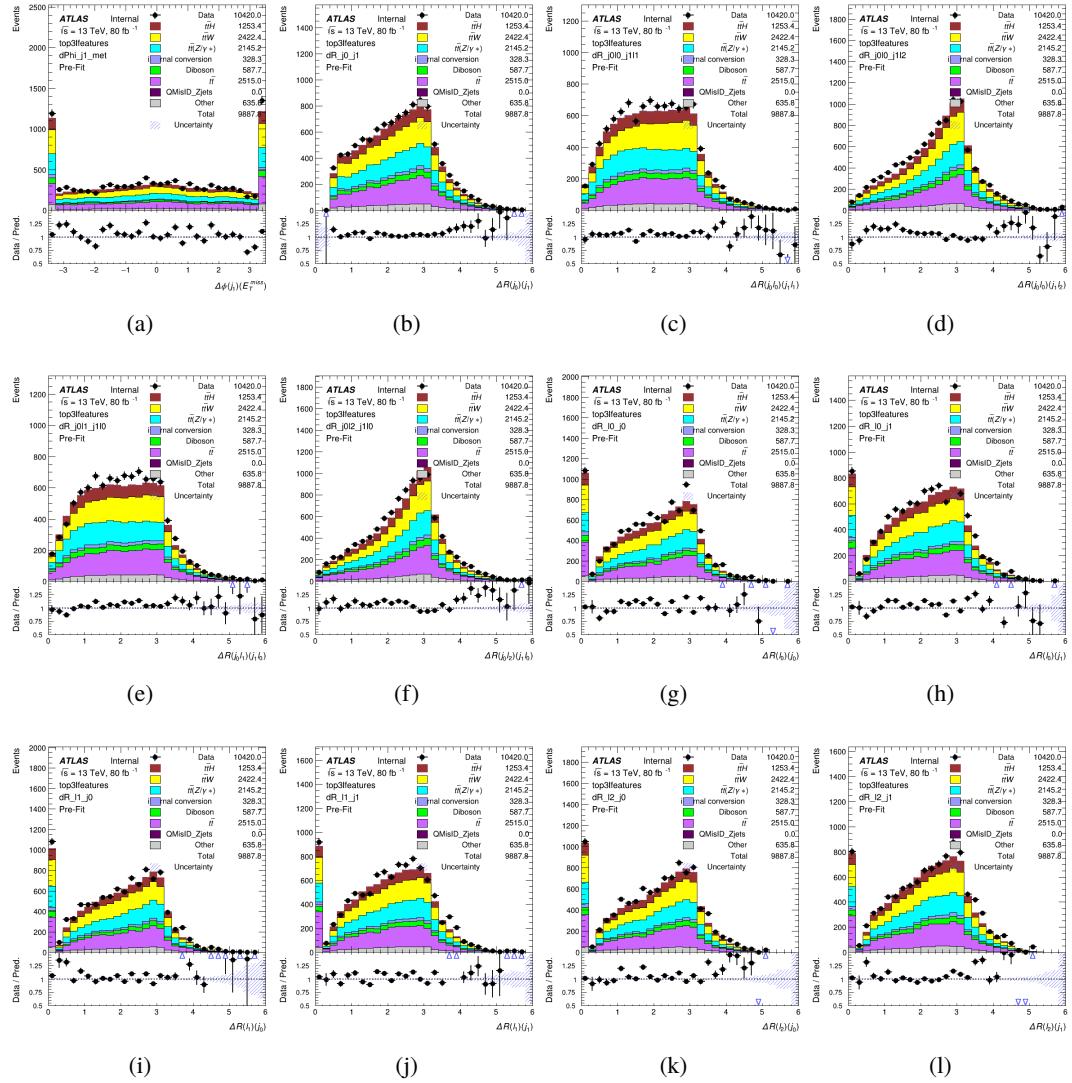


Figure A.5: Input features for top31

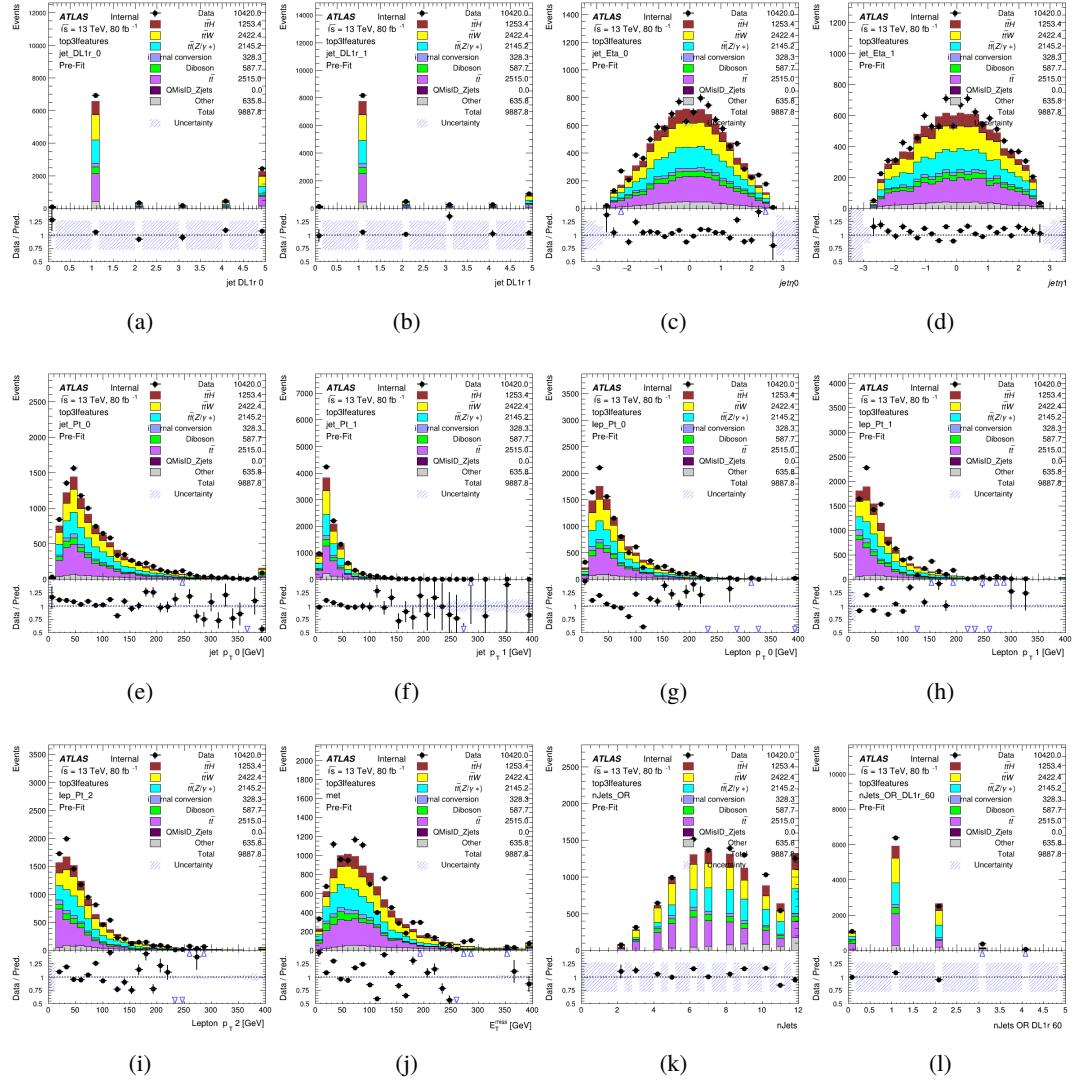
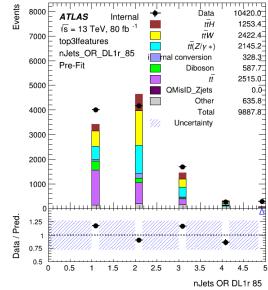


Figure A.6: Input features for top31



(a)

Figure A.7: Input features for top31

⁷⁴⁹ **A.1.3 Higgs Reconstruction Features - 2lSS**

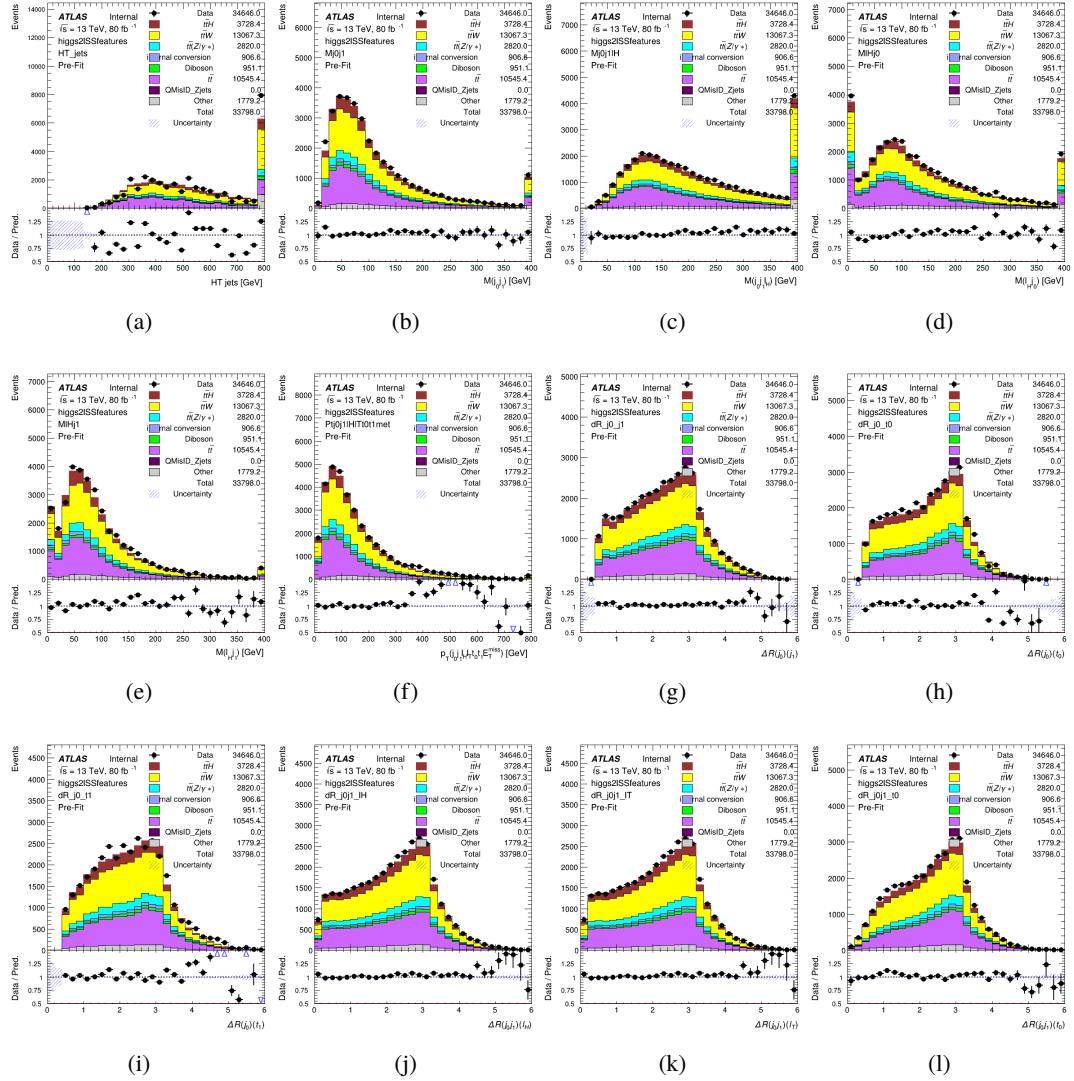


Figure A.8: Input features for higgs2lSS

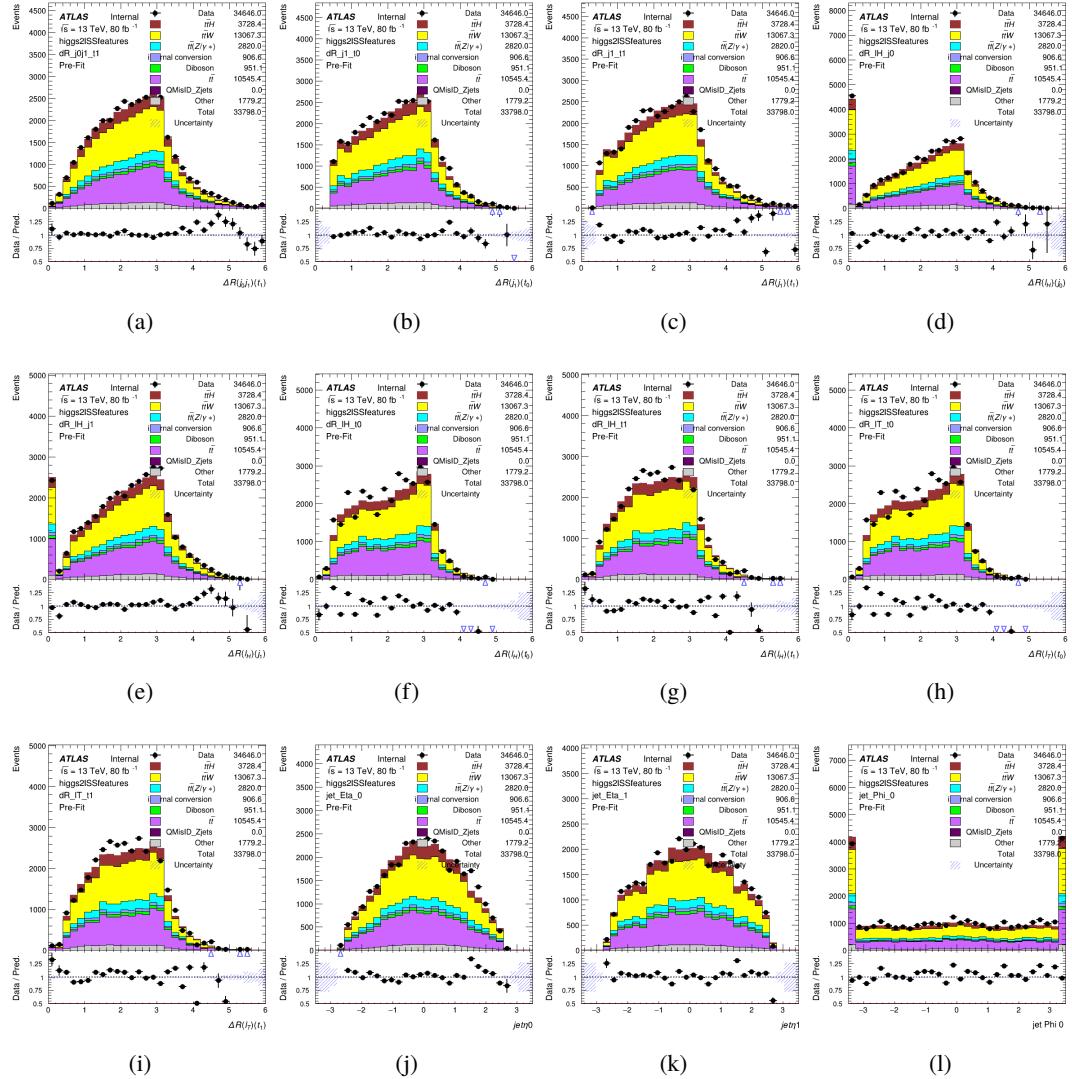


Figure A.9: Input features for higgs2lSS

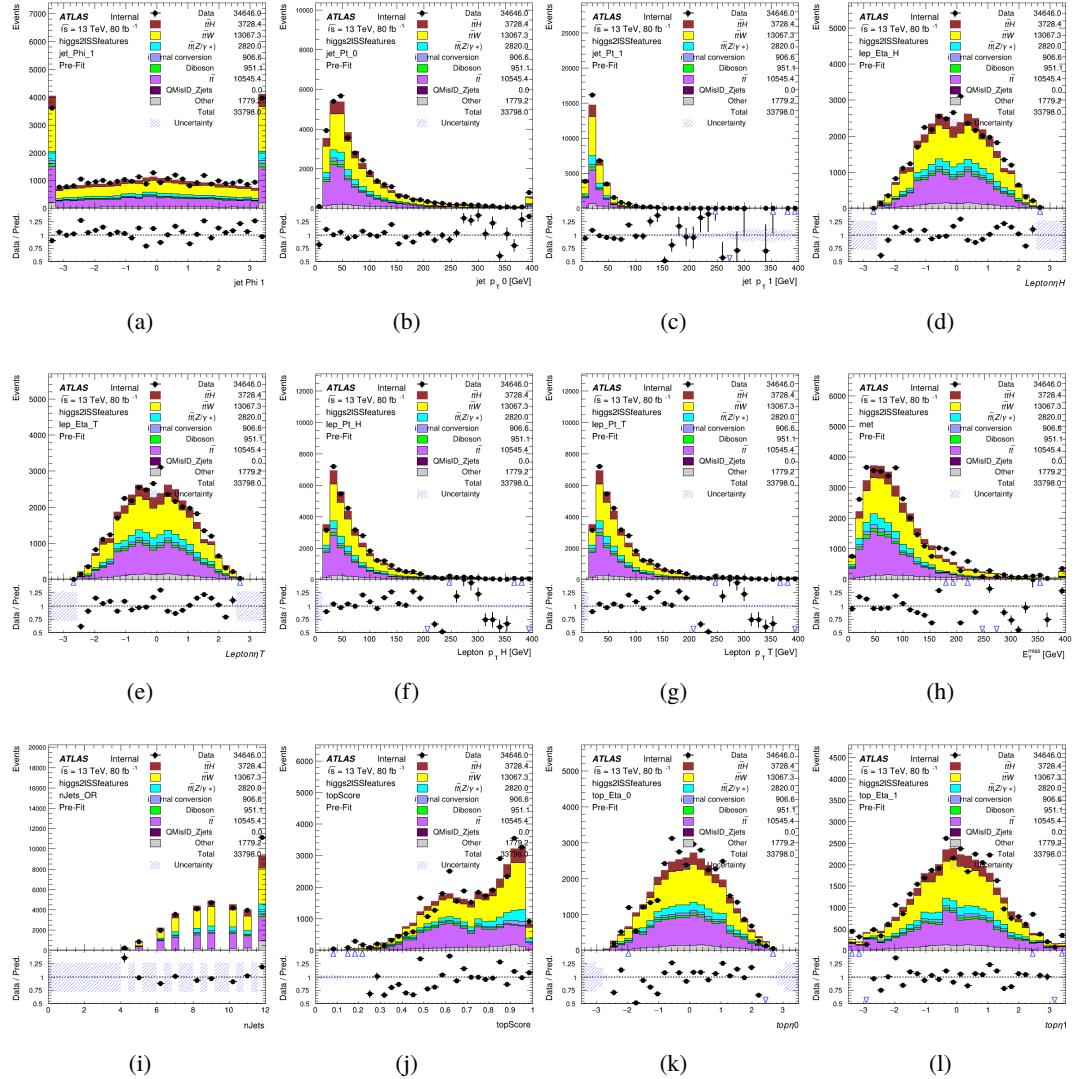


Figure A.10: Input features for higgs2IS

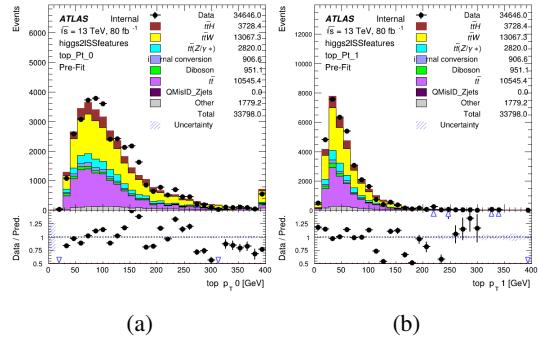


Figure A.11: Input features for higgs2lSS

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

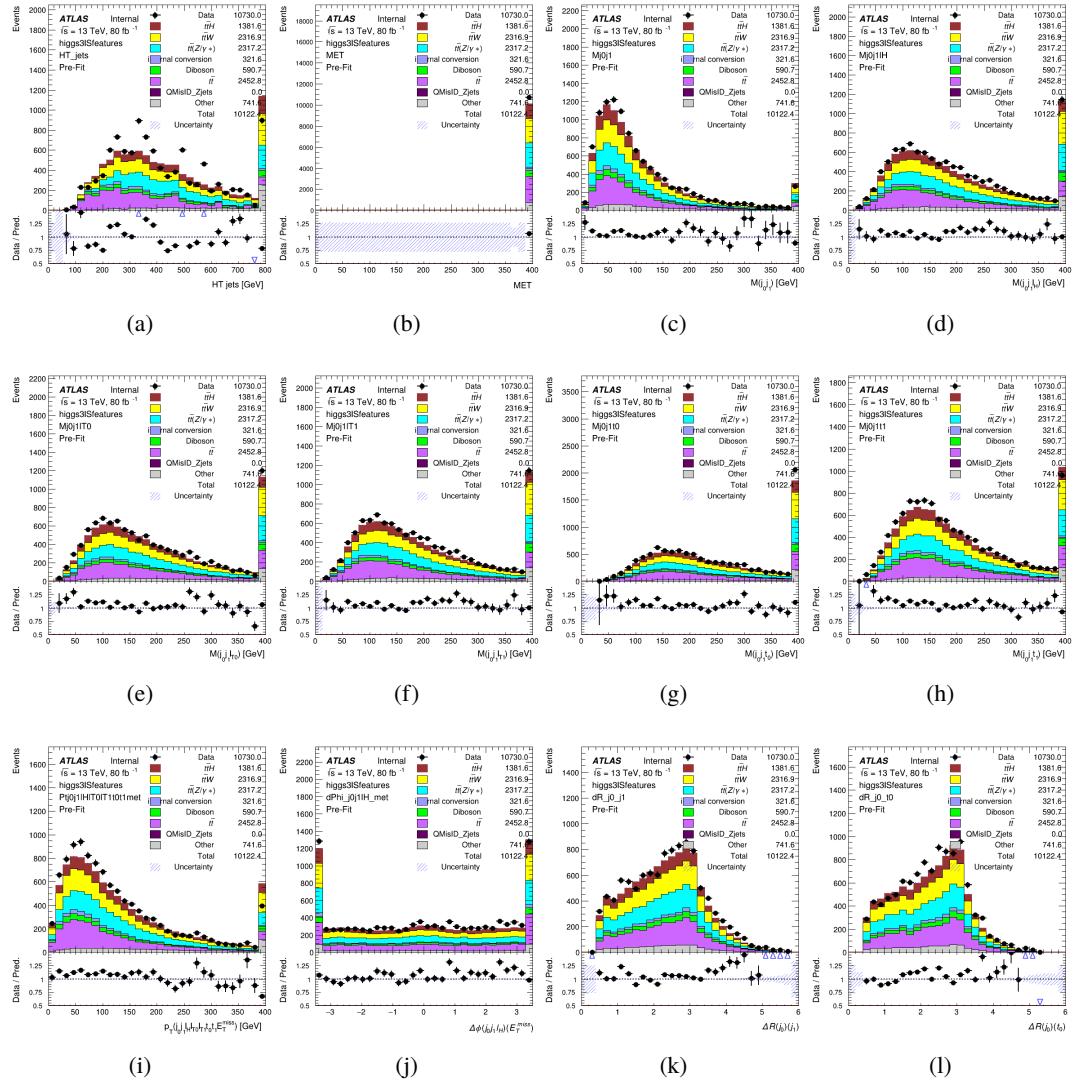


Figure A.12: Input features for higgs3IS

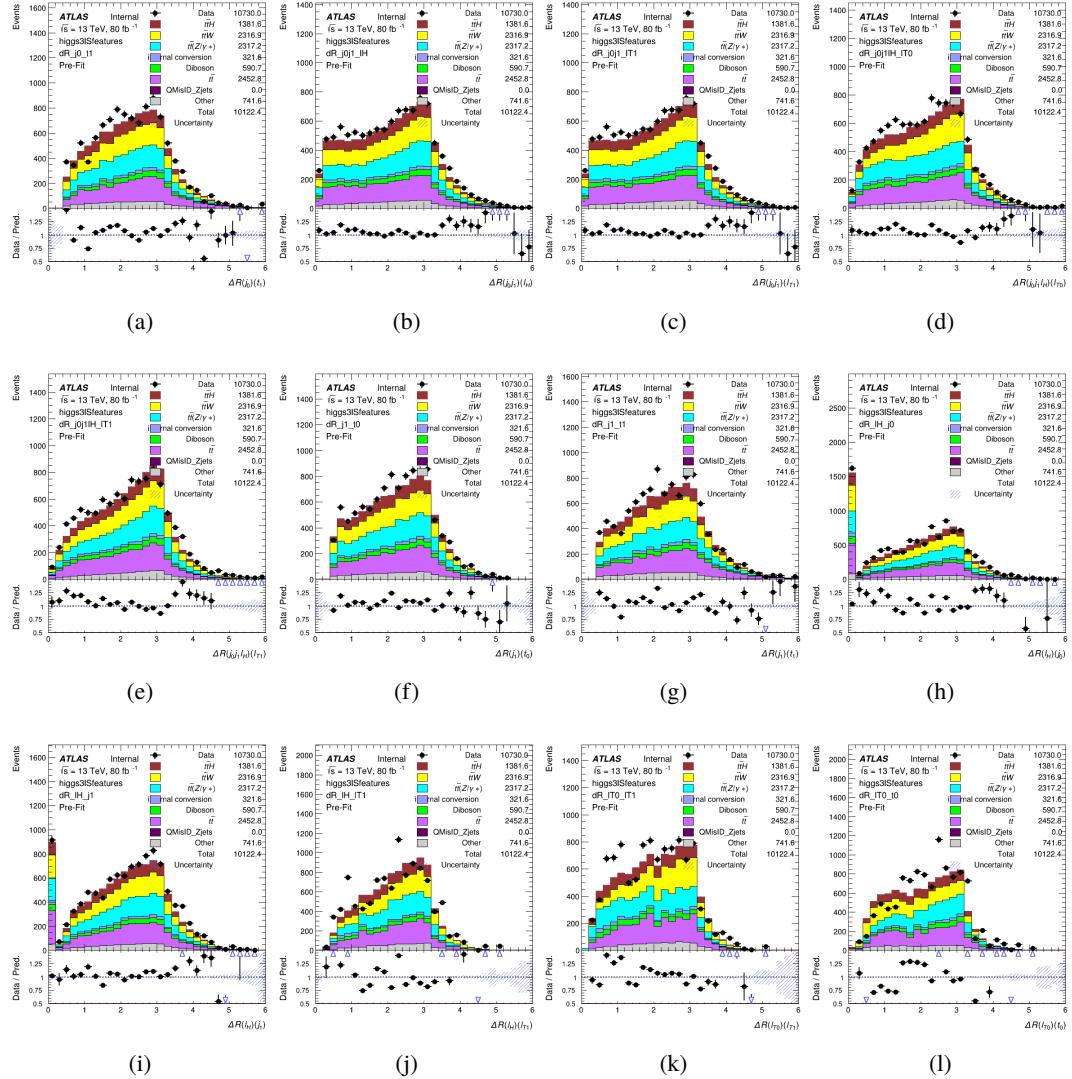


Figure A.13: Input features for higgs3IS

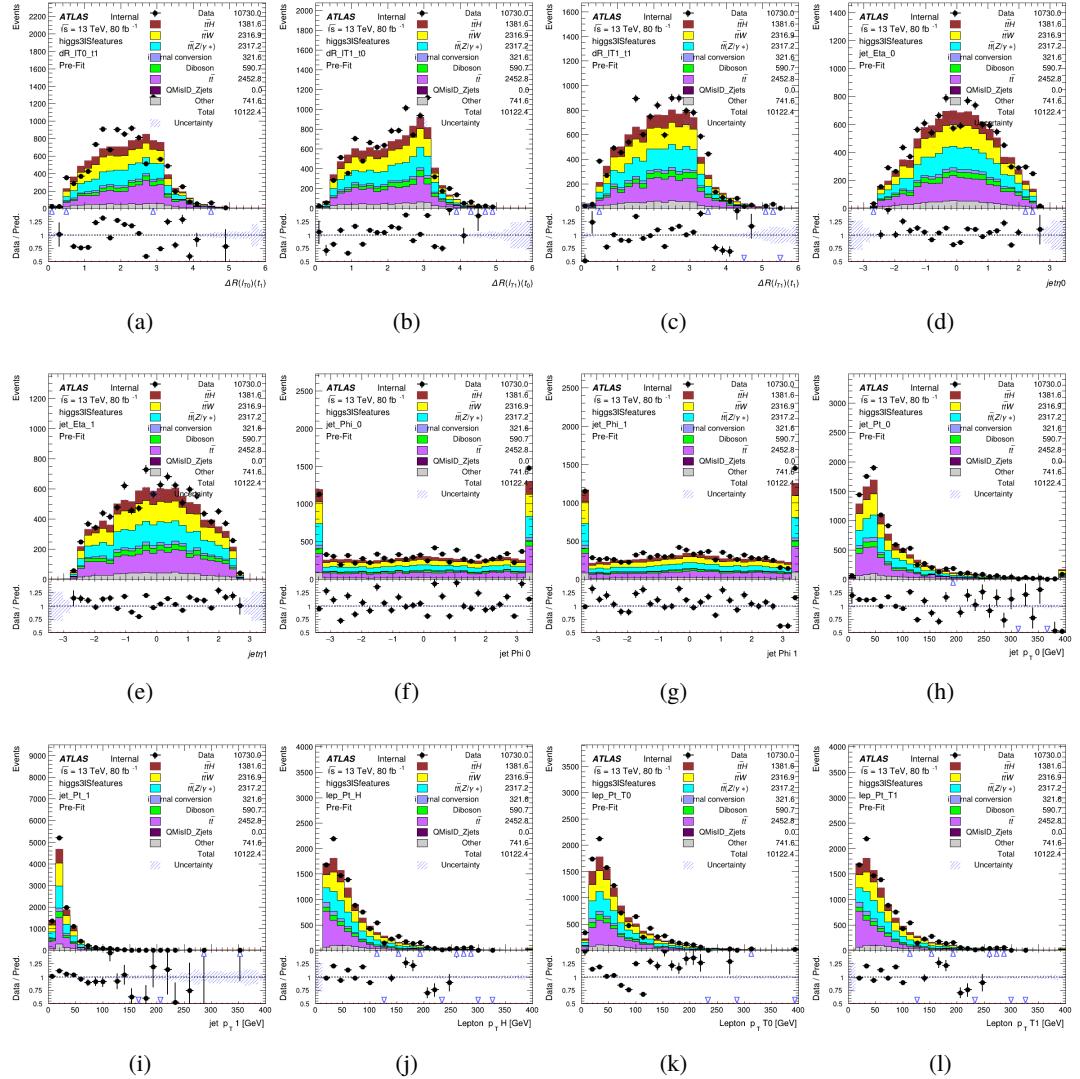


Figure A.14: Input features for higgs3IS

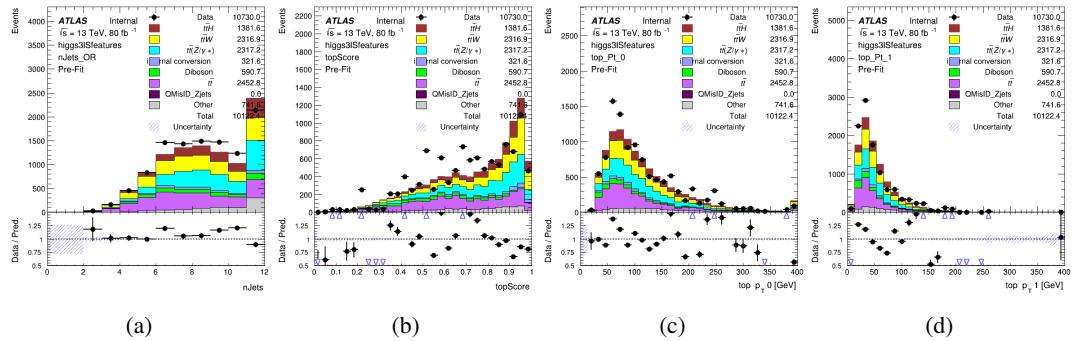


Figure A.15: Input features for higgs3IS

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

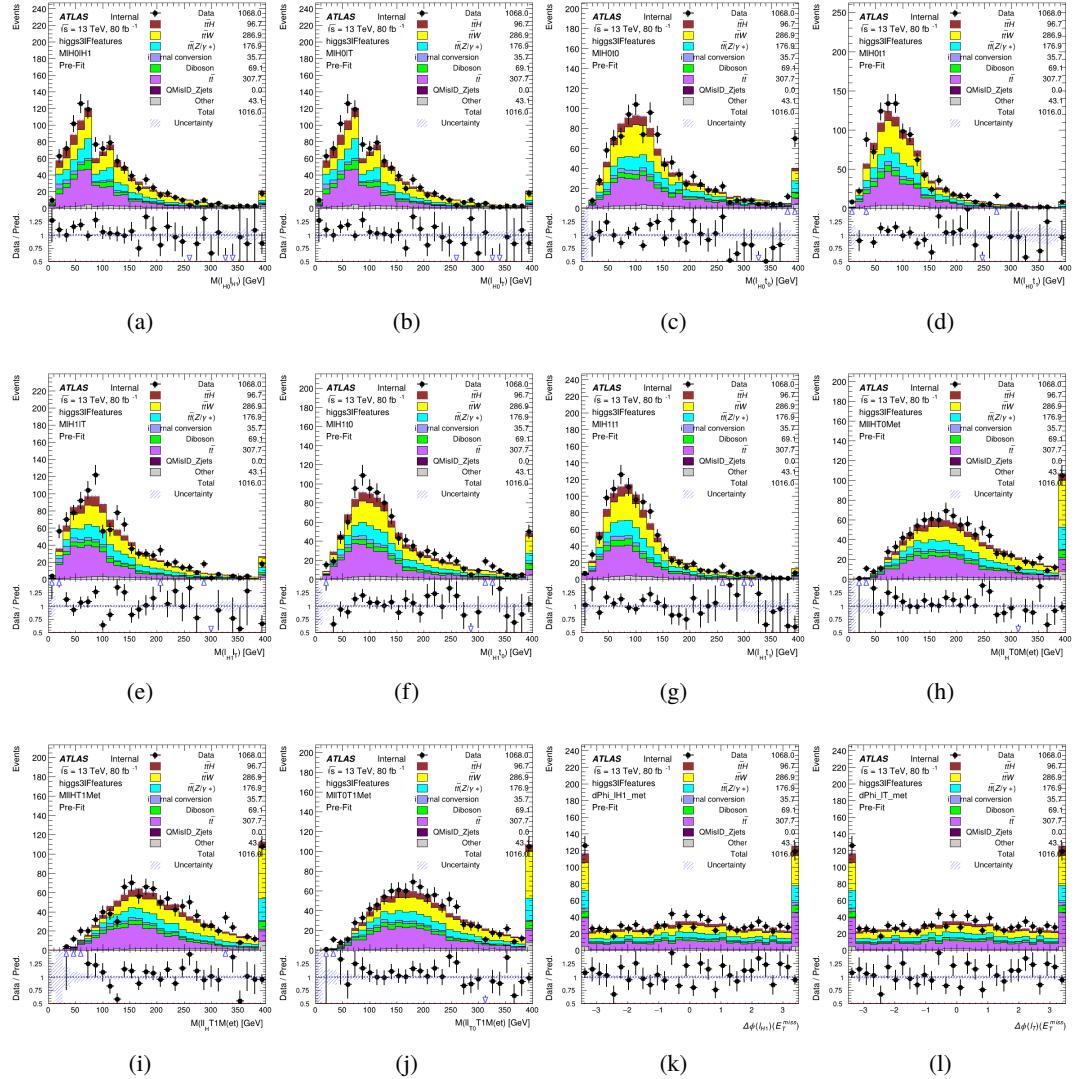


Figure A.16: Input features for higgs3IF

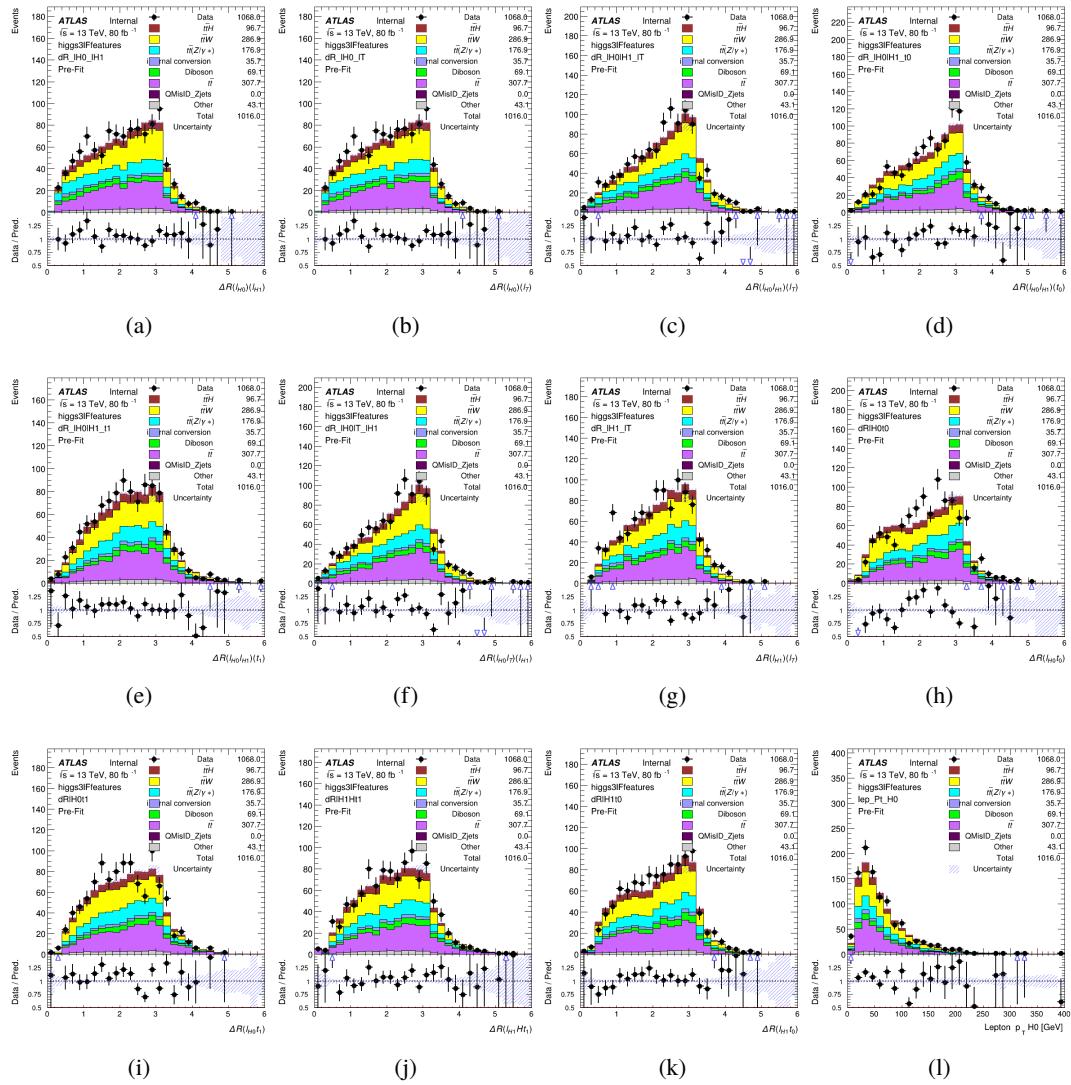


Figure A.17: Input features for higgs3lF

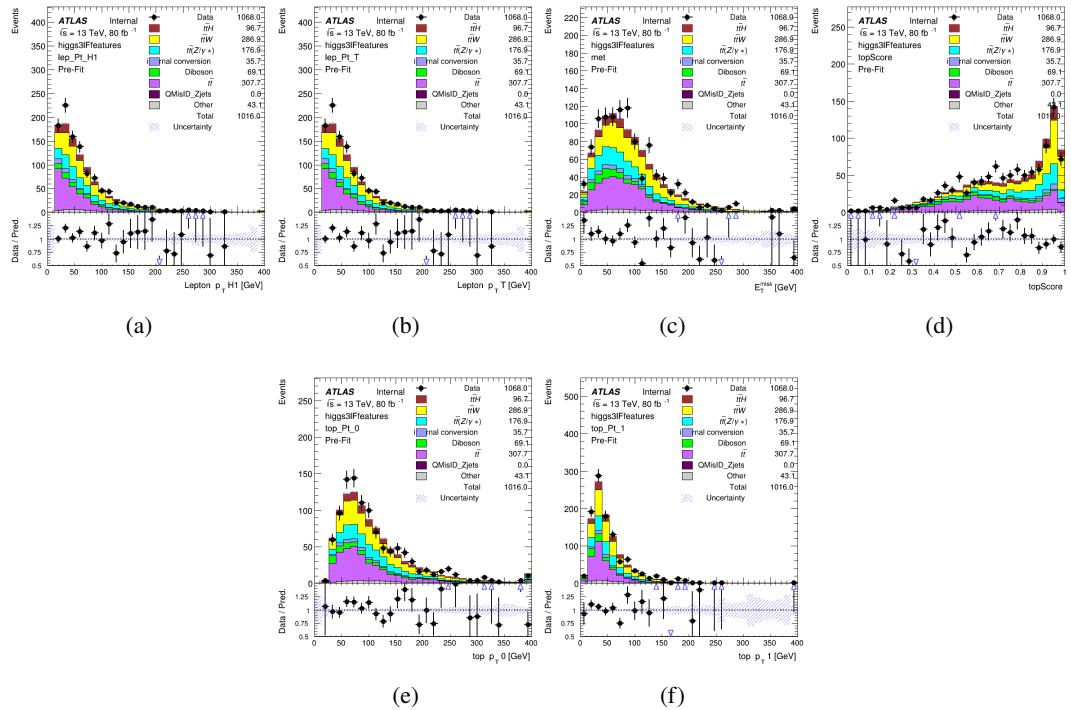


Figure A.18: Input features for higgs3lF

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

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

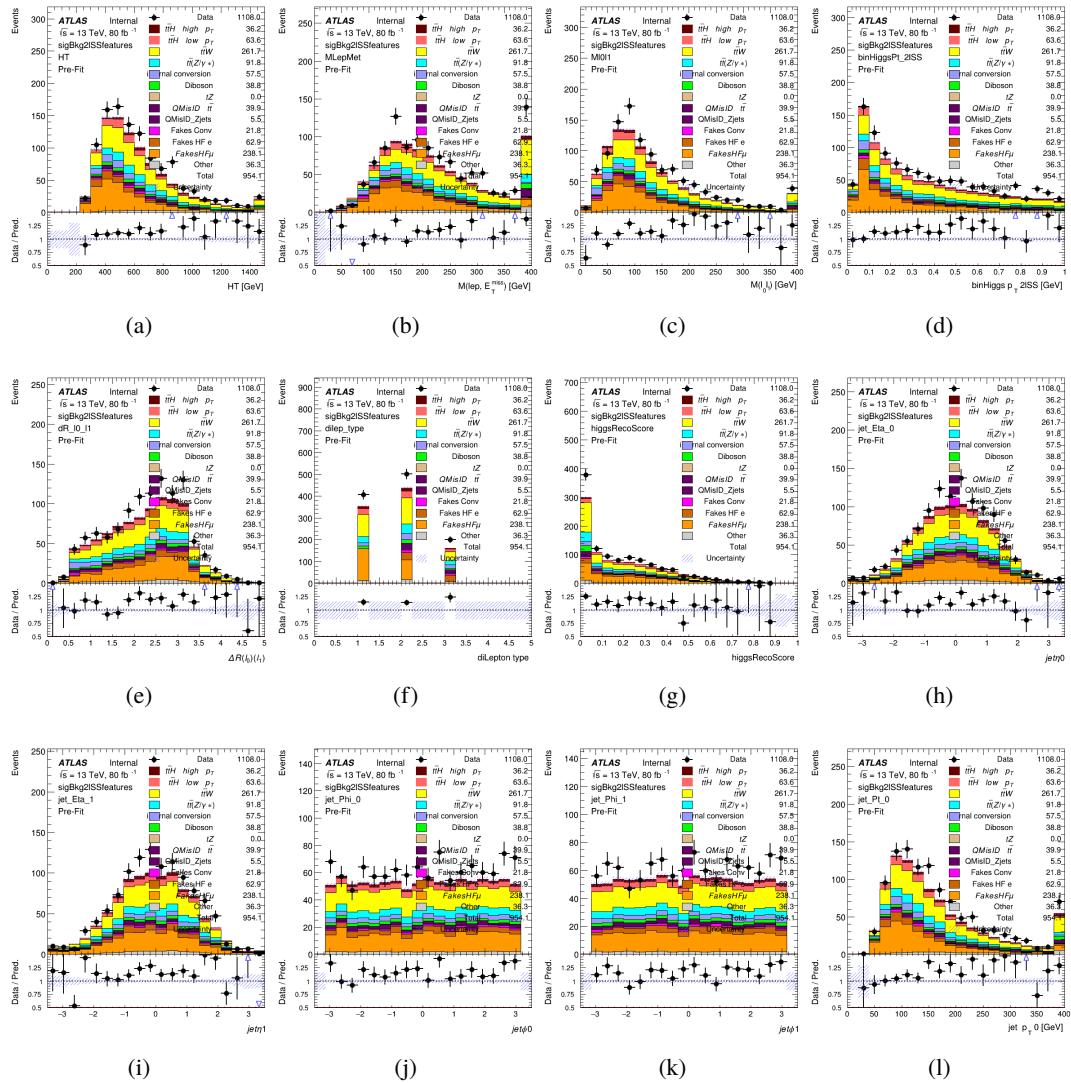


Figure A.19: Input features for sigBkg2lSS

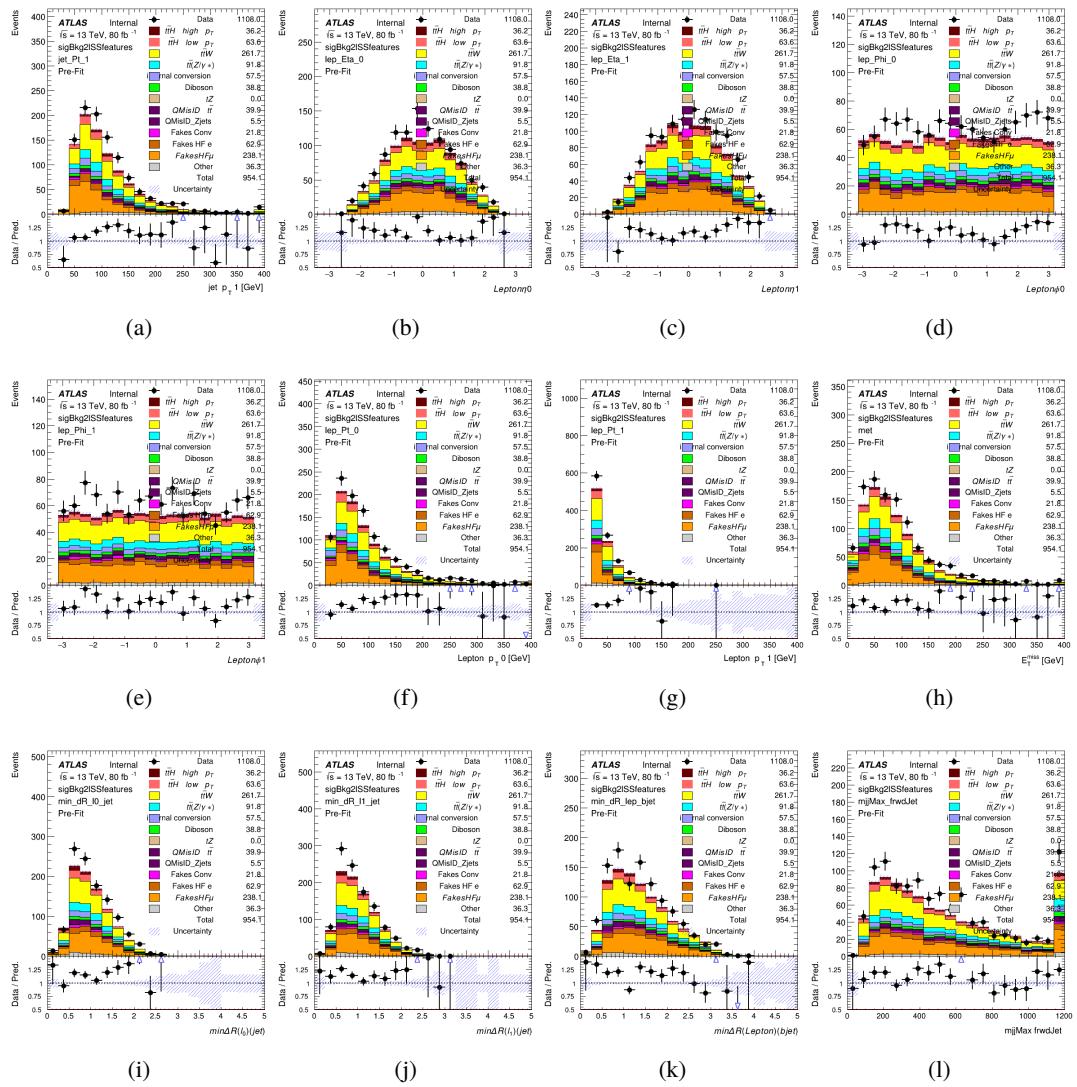


Figure A.20: Input features for sigBkg2lSS

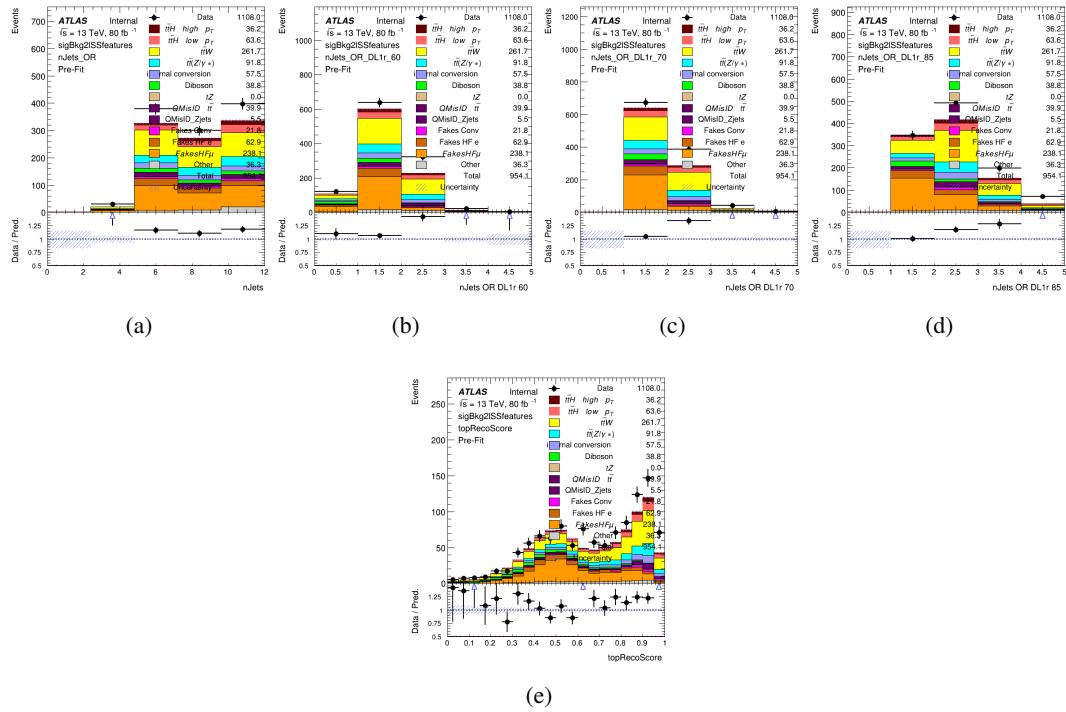
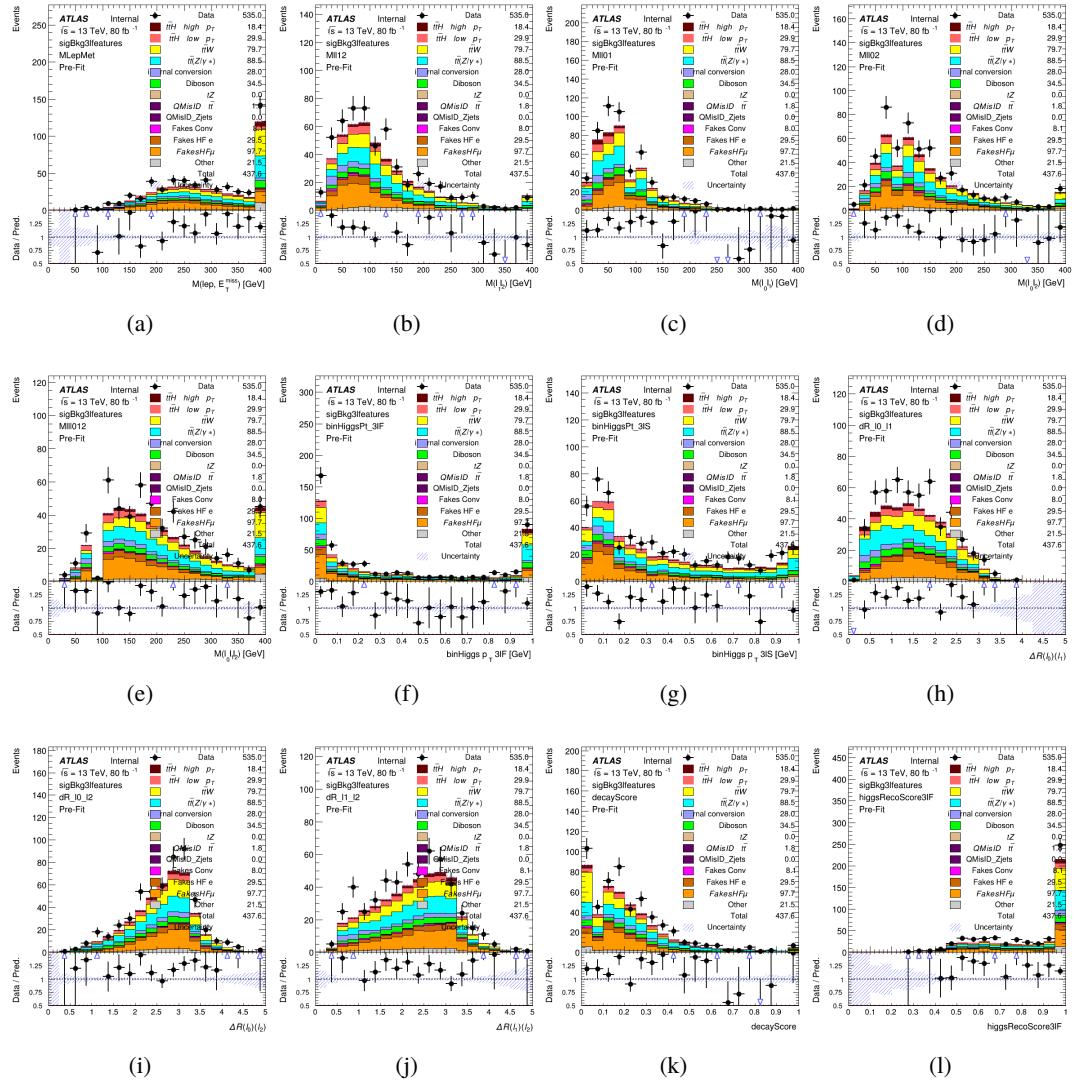


Figure A.21: Input features for sigBkg2lSS

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

Figure A.22: Input features for `sigBkg3l`

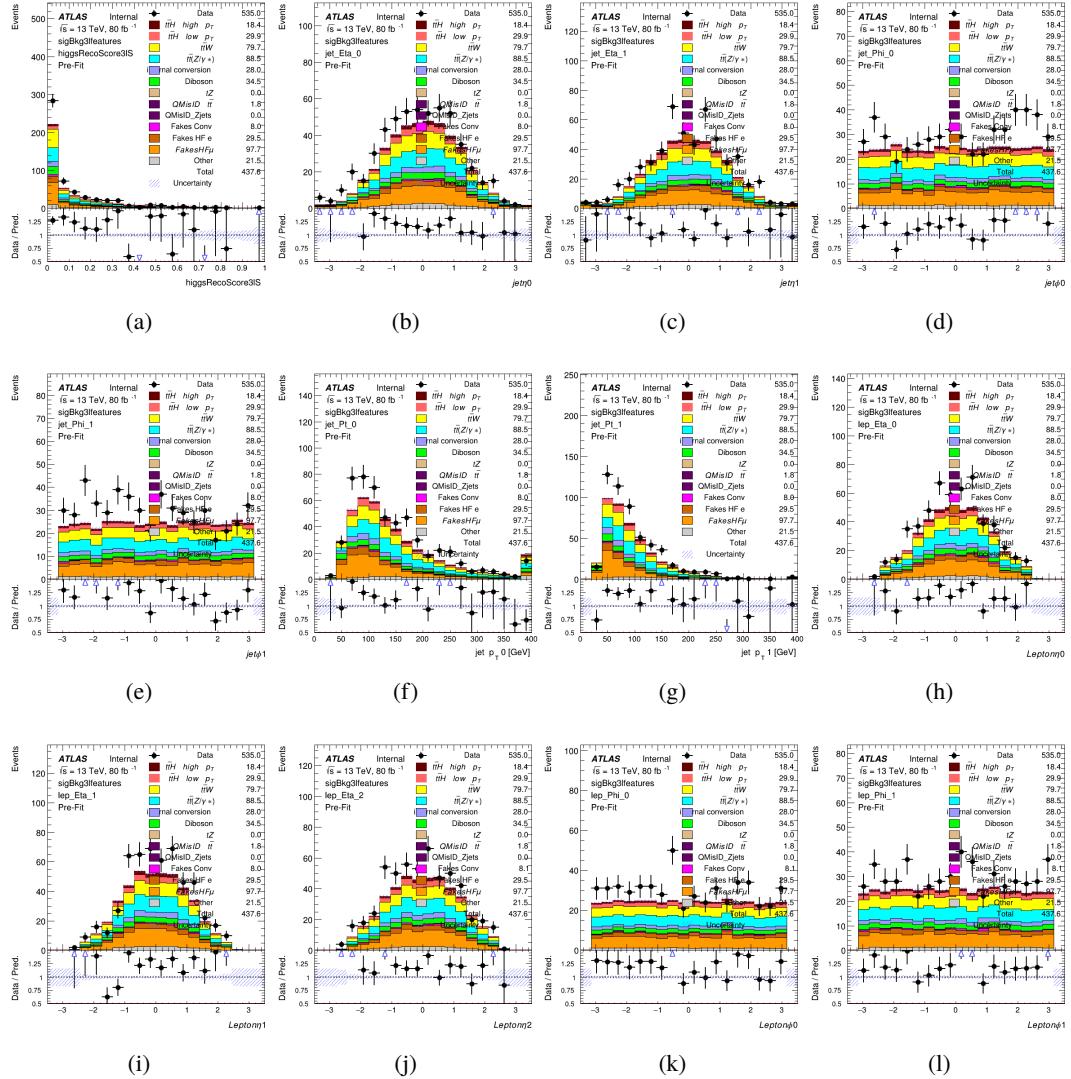
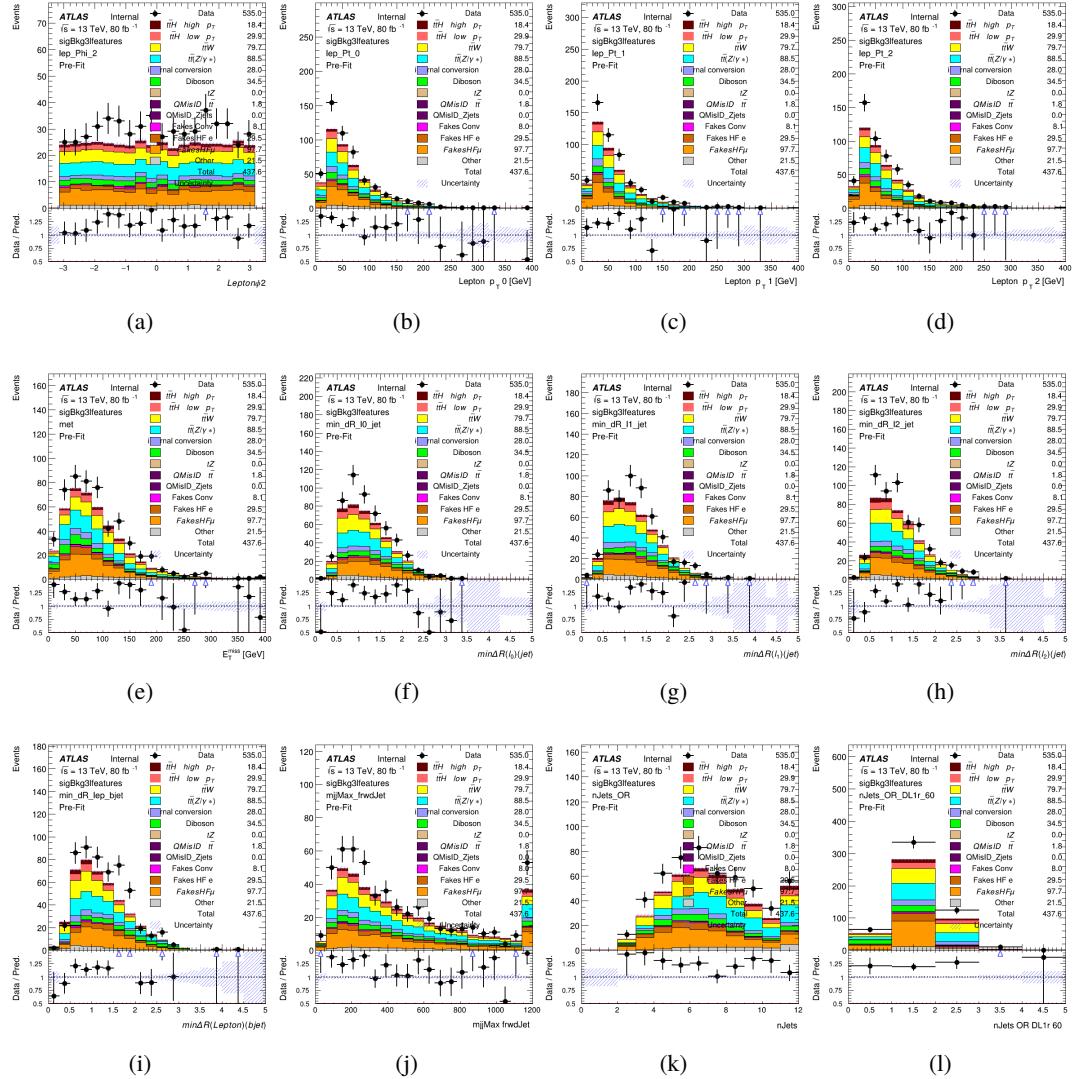


Figure A.23: Input features for sigBkg3l

Figure A.24: Input features for `sigBkg3l`

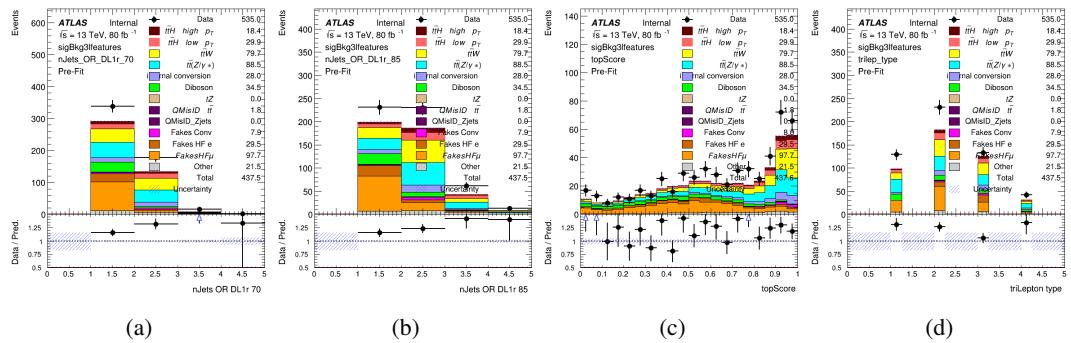


Figure A.25: Input features for sigBkg3l

755 **A.3 Truth Level Studies**

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

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

765 The results of the model are summarized below:

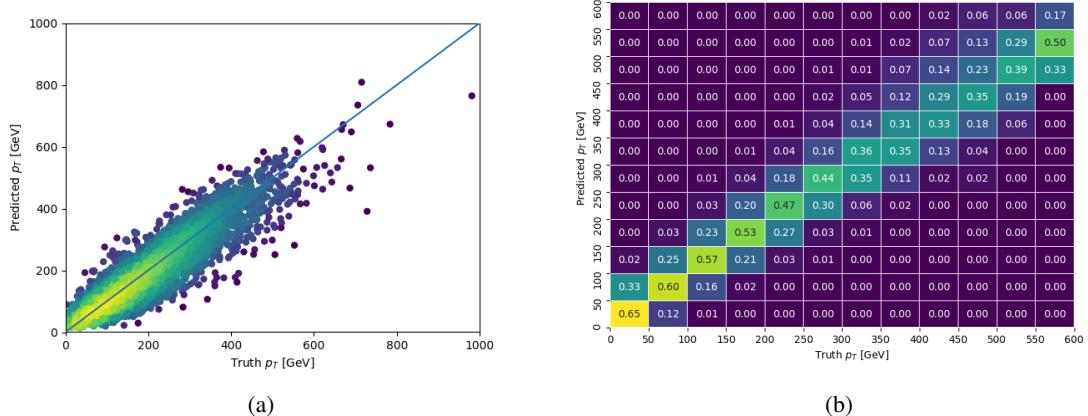


Figure A.26: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

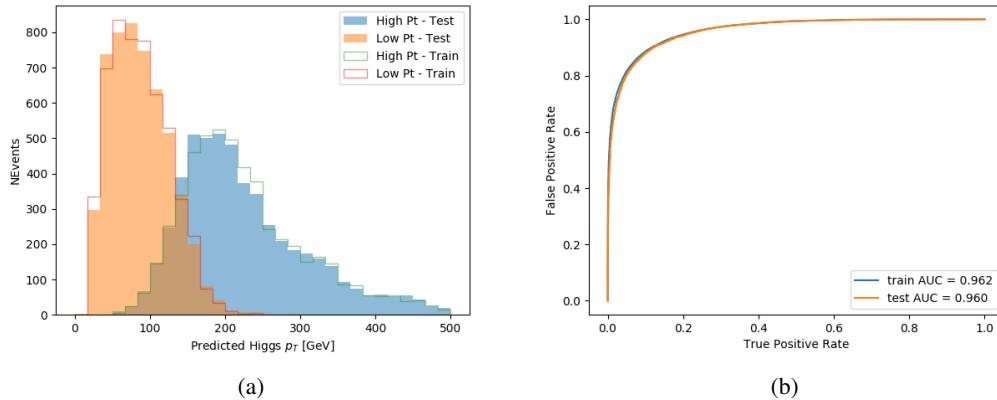


Figure A.27: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

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

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

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

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

Table 36: 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%

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

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

786 A model is built to determine whether t̄tH events include a high p_T (>150 GeV) or low p_T (<150
 787 GeV) Higgs Boson. While this is now a classification model, it uses the same input features
 788 described in section 5.5. Binary crossentropy is used as the loss function.

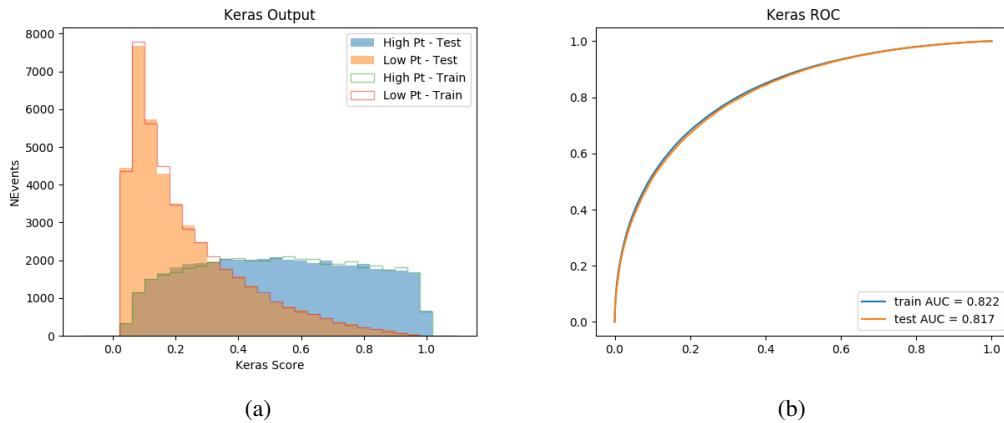


Figure A.28: Output distribution of the NN score for the binary high/low p_T separation model in the 2lSS channel.

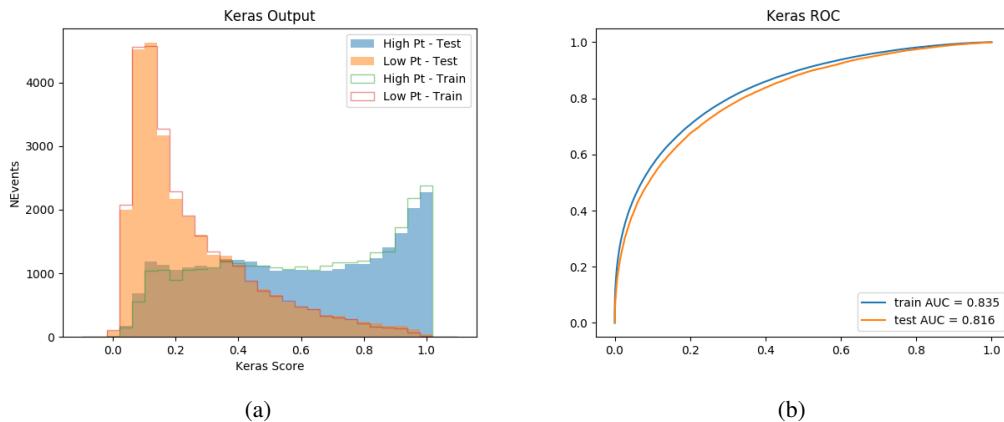


Figure A.29: Output distribution of the NN score for the binary high/low p_T separation model in the 3lS channel.

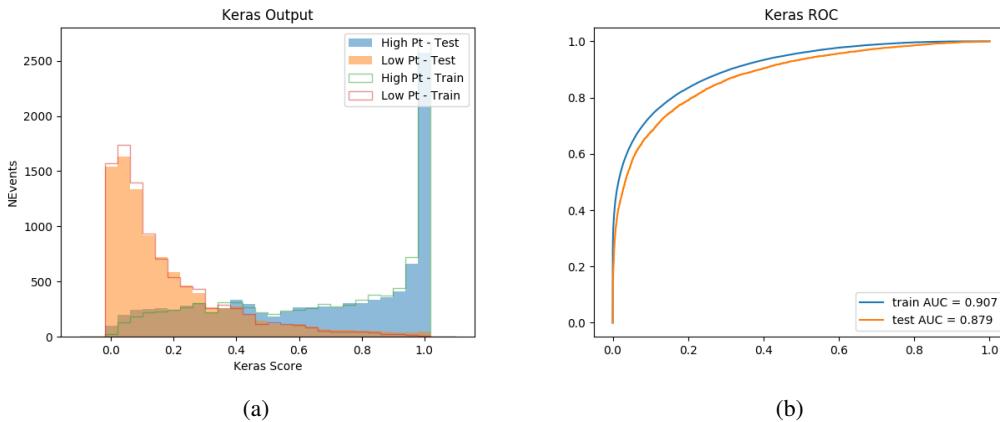


Figure A.30: Output distribution of the NN score for the binary high/low p_T separation model in the 3lS channel.

789 **A.6 Impact of Alternative Jet Selection**

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

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

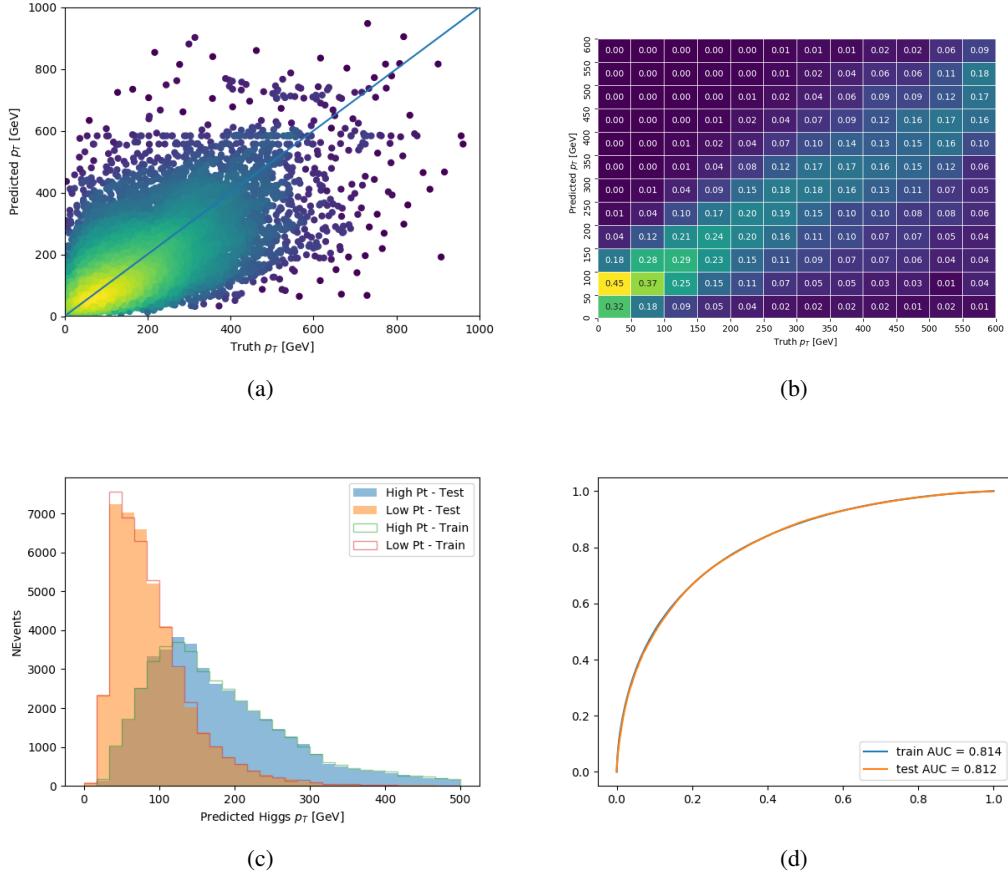
Jet $p_T > 20$ GeV

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.

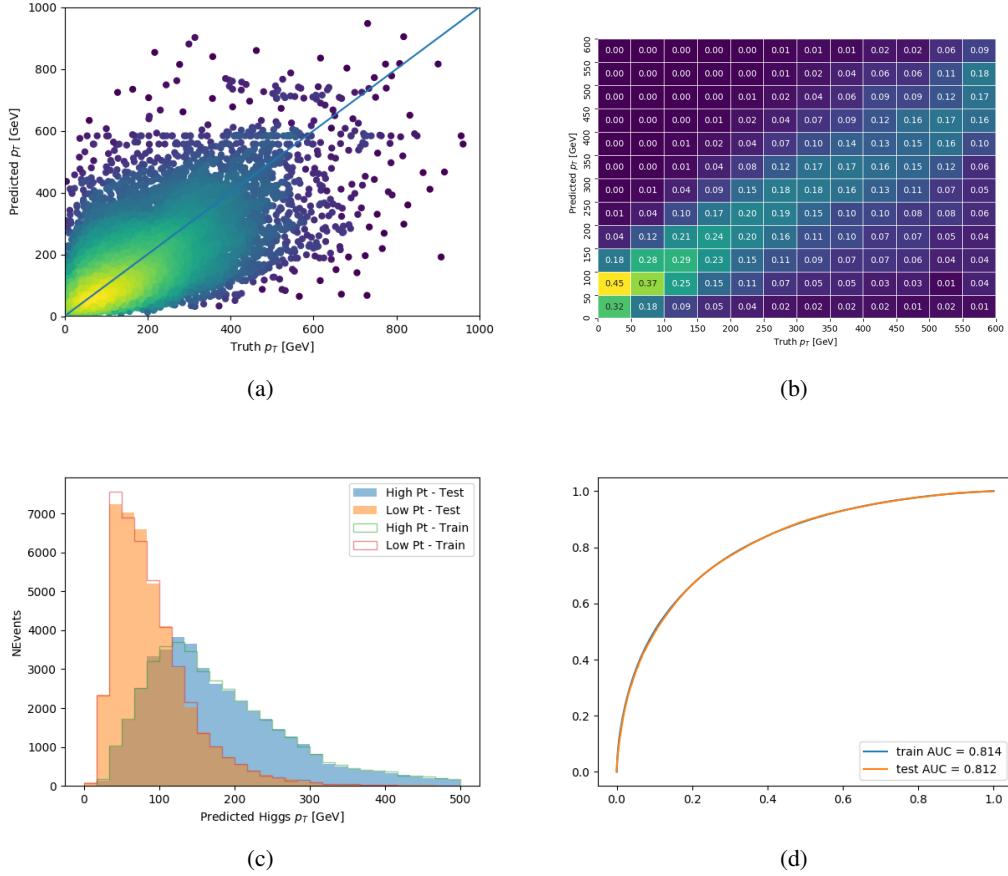
Jet $p_T > 25 \text{ GeV}$ 

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.