



## ATLAS Note

GROUP-2017-XX

10th November 2020



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

The ATLAS Collaboration

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

<sup>13</sup> © 2020 CERN for the benefit of the ATLAS Collaboration.

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

## 15 **Contents**

16	<b>1 Changes and outstanding items</b>	3
17	1.1 Changelog	3
18	<b>2 Introduction</b>	4
19	<b>3 Data and Monte Carlo Samples</b>	4
20	3.1 Data Samples	5
21	3.2 Monte Carlo Samples	5
22	<b>4 Object Reconstruction</b>	6
23	4.1 Trigger Requirements	6
24	4.2 Light Leptons	6
25	4.3 Jets	7
26	4.4 Missing Transverse Energy	8
27	<b>5 Higgs Momentum Reconstruction</b>	8
28	5.1 Decay Candidate Reconstruction	9
29	5.2 b-jet Identification	9
30	5.3 Higgs Reconstruction	12
31	5.4 $p_T$ Prediction	26
32	5.5 3l Decay Mode	36
33	<b>6 Signal Region Definitions</b>	39
34	6.1 Pre-MVA Event Selection	39
35	6.2 Event MVA	40
36	6.3 Signal Region Definitions	45
37	<b>7 Systematic Uncertainties</b>	47
38	<b>8 Results</b>	49
39	8.1 Results - $80 \text{ fb}^{-1}$	49
40	8.2 Projected Results - $140 \text{ fb}^{-1}$	50
41	<b>9 Conclusion</b>	50
42	<b>Appendices</b>	55
43	<b>A Machine Learning Models</b>	55
44	A.1 Higgs Reconstruction Models	55
45	A.2 Background Rejection MVAs	78
46	A.3 Alternate b-jet Identification Algorithm	87
47	A.4 Binary Classification of the Higgs $p_T$	87



<sup>49</sup> **1 Changes and outstanding items**

<sup>50</sup> **1.1 Changelog**

<sup>51</sup> This is version 1

---

## 52 2 Introduction

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

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

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

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

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

## 82 3 Data and Monte Carlo Samples

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

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

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

### 94     **3.1 Data Samples**

95     The study uses proton-proton collision data collected by the ATLAS detector from 2015 through  
96     2018, which represents an integrated luminosity of  $139 \text{ fb}^{-1}$  and an energy of  $\sqrt{s} = 13 \text{ TeV}$ . All  
97     data used in this analysis was included in one the following Good Run Lists:

- 98     • data15\_13TeV.periodAllYear\_DetStatus-v79-repro20-02\_DQDefects-00-02-02  
99       \_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- 100     • data16\_13TeV.periodAllYear\_DetStatus-v88-pro20-21\_DQDefects-00-02-04  
101       \_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- 102     • data17\_13TeV.periodAllYear\_DetStatus-v97-pro21-13\_Unknown\_PHYS\_StandardGRL  
103       \_All\_Good\_25ns\_Triggerno17e33prim.xml
- 104     • data18\_13TeV.periodAllYear\_DetStatus-v102-pro22-04\_Unknown\_PHYS\_StandardGRL  
105       \_All\_Good\_25ns\_Triggerno17e33prim.xml

### 106    **3.2 Monte Carlo Samples**

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

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

Process	Event generator	ME order	Parton Shower	PDF
t̄H	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [ <a href="#">Ball:2014uwa</a> ] (CT10 [ <a href="#">ct10</a> ])
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 [ <a href="#">powheggtt</a> ]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄γ	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
tZ	MG5_AMC	LO	PYTHIA 6	CTEQ6L1
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t, t̄t̄t̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	Powheg-Box v1 [ <a href="#">powhegstp</a> ]	NLO	PYTHIA 6	CT10
qqVV, VVV				
Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

## 4 Object Reconstruction

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

### 4.1 Trigger Requirements

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

### 4.2 Light Leptons

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

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

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

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

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

### 130 4.3 Jets

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

<sup>138</sup> **4.4 Missing Transverse Energy**

<sup>139</sup> Because all  $t\bar{t}H$  – ML channels considered include multiple neutrinos, missing transverse energy  
<sup>140</sup> ( $E_T^{\text{miss}}$ ) is present in each event. The missing transverse momentum vector is defined as the  
<sup>141</sup> inverse of the sum of the transverse momenta of all reconstructed physics objects as well as  
<sup>142</sup> remaining unclustered energy, the latter of which is estimated from low- $p_T$  tracks associated with  
<sup>143</sup> the primary vertex but not assigned to a hard object [ATL-PHYS-PUB-2015-027].

<sup>144</sup> **5 Higgs Momentum Reconstruction**

<sup>145</sup> Reconstructing the momentum of the Higgs boson is a particular challenge for channels with  
<sup>146</sup> leptons in the final state: Because all channels include at least two neutrinos in the final state, the  
<sup>147</sup> Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted  
<sup>148</sup> by a neural network when provided with the four-vectors of the Higgs Boson decay products, as  
<sup>149</sup> shown in section 5.1. With this in mind, several layers of MVAs are used to reconstruction the  
<sup>150</sup> Higgs momentum.

<sup>151</sup> The first layer is a model designed to select which jets are most likely to be the b-jets that came  
<sup>152</sup> from the top decay, detailed in section 5.2. As described in section 5.3, the kinematics of these  
<sup>153</sup> jets are fed into the second layer, which is designed to identify the decay products of the Higgs  
<sup>154</sup> Boson itself. The kinematics of these particles are then fed into yet another neural-network, which  
<sup>155</sup> predicts the momentum of the Higgs (5.4). MVAs are also used in the analysis to determine the  
<sup>156</sup> decay of the Higgs boson in the 3l channel (5.5).

<sup>157</sup> Models are trained on Monte Carlo simulations of  $t\bar{t}H$  events generated using MG5\_AMC.  
<sup>158</sup> Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training.

<sup>159</sup> For all of these models, the Keras neural network framework, with Tensorflow as the backend, is  
<sup>160</sup> used, and the number of hidden layers and nodes are determined using grid search optimization.  
<sup>161</sup> Each neural network uses the LeakyReLU activation function, a learning rate of 0.01, and the  
<sup>162</sup> Adam optimization algorithm, as alternatives are found to either decrease or have no impact on  
<sup>163</sup> performance. Batch normalization is applied after each layer. For the classification algorithms  
<sup>164</sup> (b-jet matching, Higgs reconstruction, and 3l decay identification) binary-cross entropy is used as  
<sup>165</sup> the loss function, while the  $p_T$  reconstruction algorithm uses MSE.

<sup>166</sup> The specific inputs features used for each model are arrived at through a process of trial and error  
<sup>167</sup> - features considered potentially useful are tried, and those that are found to increase performance  
<sup>168</sup> are included. While each model includes a relatively large number of features, some using  
<sup>169</sup> upwards of 30, this inclusive approach is found to maximize the performance of each model while  
<sup>170</sup> decreasing the variance compared to a reduced number of inputs. Each input feature is validated  
<sup>171</sup> by comparing MC simulations to  $80 \text{ fb}^{-1}$  of data, as shown in the sections below.

---

<sup>172</sup> **5.1 Decay Candidate Reconstruction**

<sup>173</sup> Machine Learning algorithms are trained to identify the decay products of the Higgs Boson using  
<sup>174</sup> MC simulations of  $t\bar{t}H$  events. These include light leptons and jets. Reconstructed physics  
<sup>175</sup> objects are matched to truth level particles, in order to identify the parents of these reconstructed  
<sup>176</sup> objects. The kinematics of the decay product candidates as well as event level variables are used  
<sup>177</sup> as inputs.

<sup>178</sup> Leptons considered as possible Higgs and top decay candidates are required to pass the selection  
<sup>179</sup> described in section 4.2. For jets, however, it is found that a large fraction that originate from either  
<sup>180</sup> the top decay or the Higgs decay fall outside the selection described in section 4.3. Specifically,  
<sup>181</sup> jets from the Higgs decay tend to be soft, with 32% having  $p_T < 25$  GeV. Therefore jets with  
<sup>182</sup>  $p_T < 15$  GeV are considered as possible candidates in the models described below. By contrast,  
<sup>183</sup> less than 5% of the jets originating from the Higgs fall below this  $p_T$ . The jets are found to be  
<sup>184</sup> well modeled even down to this low  $p_T$  threshold, as shown in section 6.1. The impact of using  
<sup>185</sup> different  $p_T$  selection for the jet candidates is considered in detail in section A.5. As they are  
<sup>186</sup> expected to originate from the primary vertex, jets are also required to pass a JVT cut.

<sup>187</sup> **5.2 b-jet Identification**

<sup>188</sup> Including the kinematics of the b-jets that originate from the top decay is found to improve the  
<sup>189</sup> identification of the Higgs decay products, and improve the accuracy with which the Higgs  
<sup>190</sup> momentum can be reconstructed. Because these b-jets are reconstructed by the detector with high  
<sup>191</sup> efficiency (just over 90% of the time), and can be identified relatively consistently, the first step in  
<sup>192</sup> reconstructing the Higgs is selecting the b-jets from the top decay.

<sup>193</sup> Exactly two b-jets are expected in the final state of  $t\bar{t}H - ML$  events. However, in both the 3l and  
<sup>194</sup> 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r b-tag working  
<sup>195</sup> point is used). Therefore, for events which have exactly one, or more than two, b-tagged jets,  
<sup>196</sup> deciding which combination of jets correspond to the top decay is non-trivial. Further, events  
<sup>197</sup> with 1 b-tagged jet represent just over half of all  $t\bar{t}H - ML$  events. Of those, both b-jets are  
<sup>198</sup> reconstructed by the detector 75% of the time. Therefore, rather than adjusting the selection to  
<sup>199</sup> require exactly 2 b-tagged jets, and losing more than half of the signal events, a neural network is  
<sup>200</sup> used to predict which pair of jets is most likely to correspond to truth b-jets.

<sup>201</sup> Once the network is trained, all possible pairings of jets are fed into the model, and the pair of jets  
<sup>202</sup> with the highest output score are taken to be b-jets in successive steps of the analysis.

<sup>203</sup> **5.2.1 2lSS Channel**

<sup>204</sup> For the 2lSS channel, the input features shown in table 3 are used for training. Here  $j_0$  and  $j_1$  are  
<sup>205</sup> the two jet candidates, while  $l_0$  and  $l_1$  are the two leptons in the event, both ordered by  $p_T$ . jet  
<sup>206</sup> DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,

207 where 5 represents the tightest working point and 1 represents the loosest. The variables nJets  
 208 DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and  
 209 85% b-tag working points, respectively.

jet p <sub>T</sub> 0	jet p <sub>T</sub> 1	Lepton p <sub>T</sub> 0
Lepton p <sub>T</sub> 1	jet η 0	jet η 1
ΔR(j <sub>0</sub> )(j <sub>1</sub> )	M(j <sub>0</sub> j <sub>1</sub> )	ΔR(l <sub>0</sub> )(j <sub>0</sub> )
ΔR(l <sub>0</sub> )(j <sub>1</sub> )	ΔR(l <sub>1</sub> )(j <sub>0</sub> )	ΔR(l <sub>1</sub> )(j <sub>1</sub> )
M(l <sub>0</sub> j <sub>0</sub> )	M(l <sub>0</sub> j <sub>1</sub> )	M(l <sub>1</sub> j <sub>0</sub> )
M(l <sub>1</sub> j <sub>1</sub> )	jet DL1r 0	jet DL1r 1
nJets OR DL1r 85	nJets OR DL1r 60	ΔR(j <sub>0</sub> l <sub>0</sub> )(j <sub>1</sub> l <sub>1</sub> )
ΔR(j <sub>0</sub> l <sub>1</sub> )(j <sub>1</sub> l <sub>0</sub> )	p <sub>T</sub> (j <sub>0</sub> j <sub>1</sub> l <sub>0</sub> l <sub>1</sub> E <sub>T</sub> <sup>miss</sup> )	M(j <sub>0</sub> j <sub>1</sub> l <sub>0</sub> l <sub>1</sub> E <sub>T</sub> <sup>miss</sup> )
ΔΦ(j <sub>0</sub> )(E <sub>T</sub> <sup>miss</sup> )	ΔΦ(j <sub>1</sub> )(E <sub>T</sub> <sup>miss</sup> )	HT jets
nJets	E <sub>T</sub> <sup>miss</sup>	

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

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

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

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

220 Based on the results of grid search evaluation, the optimal architecture is found to include 5  
 221 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as overfitting  
 222 is found to not be an issue. The output score distribution as well as the ROC curve for the trained  
 223 model are shown in figure 5.2.1. The model is found to identify the correct pairing of jets for  
 224 73% of 2lSS signal events on test data.

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

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

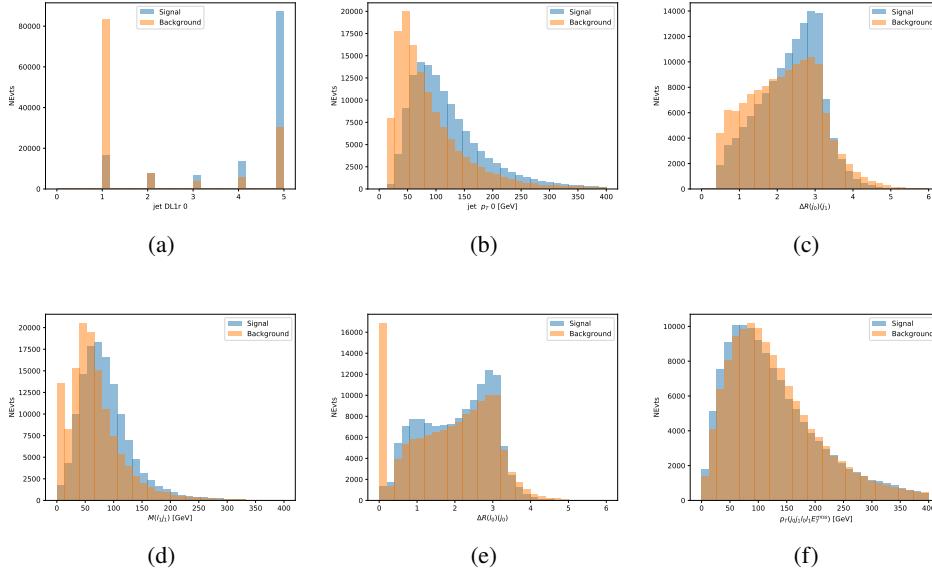


Figure 5.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the  $p_T$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

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

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

### 5.2.2 3l Channel

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

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

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

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

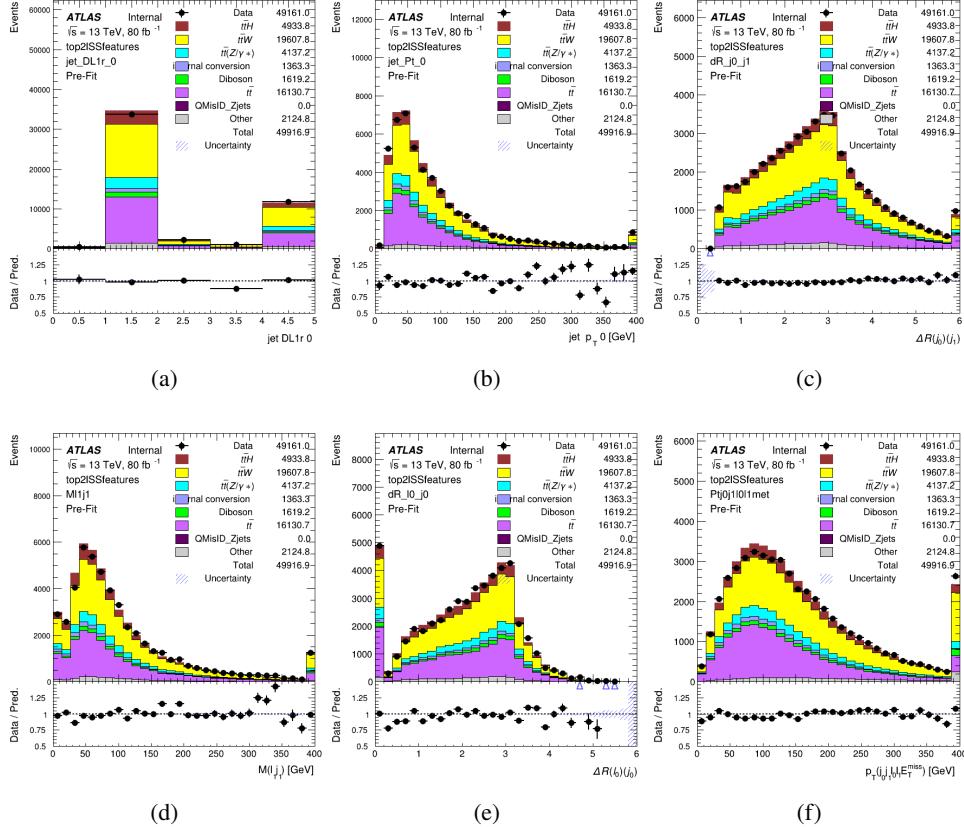


Figure 5.2: Data/MC comparisons of input features for top2lSS training for  $80 \text{ fb}^{-1}$  of data. (a) shows the DL1r working point of leading jet, (b) shows the  $p_T$  of the leading jet, (c) shows the  $\Delta R$  of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the  $\Delta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

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

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

### 5.3 Higgs Reconstruction

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

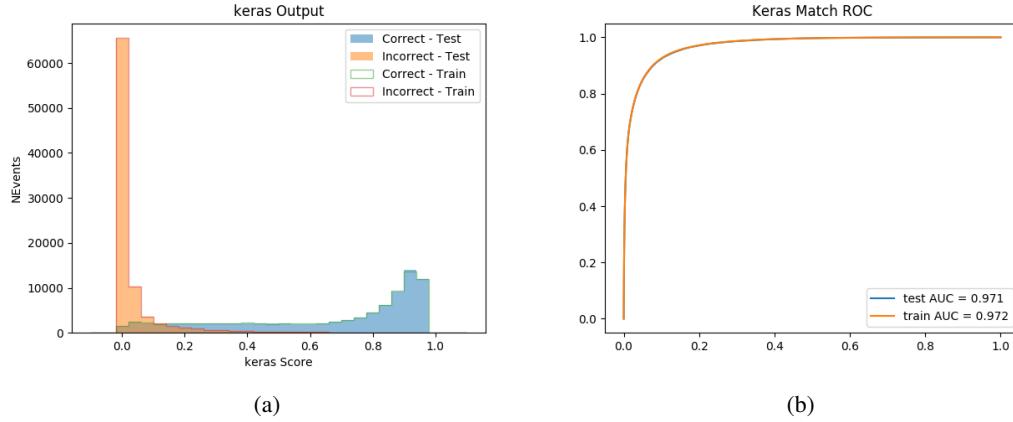


Figure 5.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

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

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

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

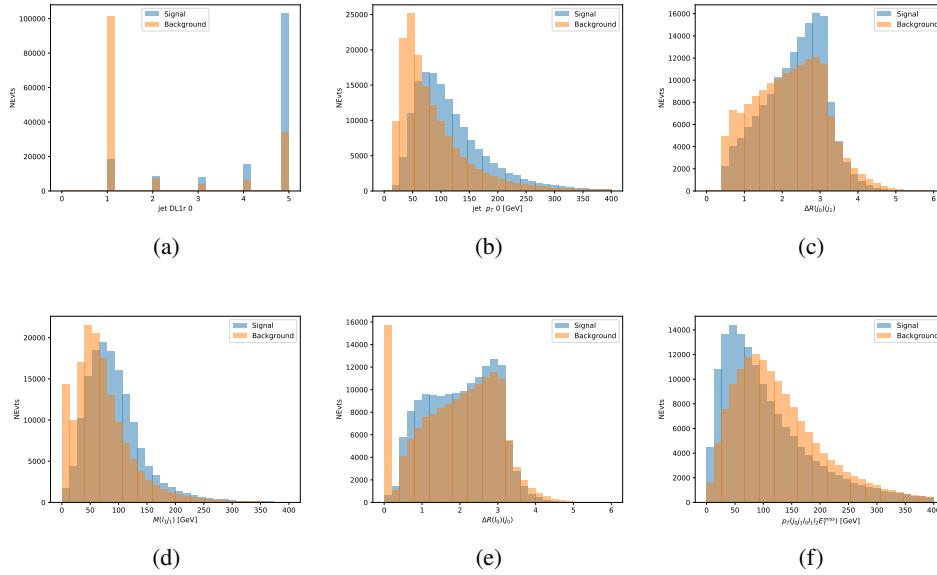


Figure 5.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

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

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

### 5.3.1 2ISS Channel

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

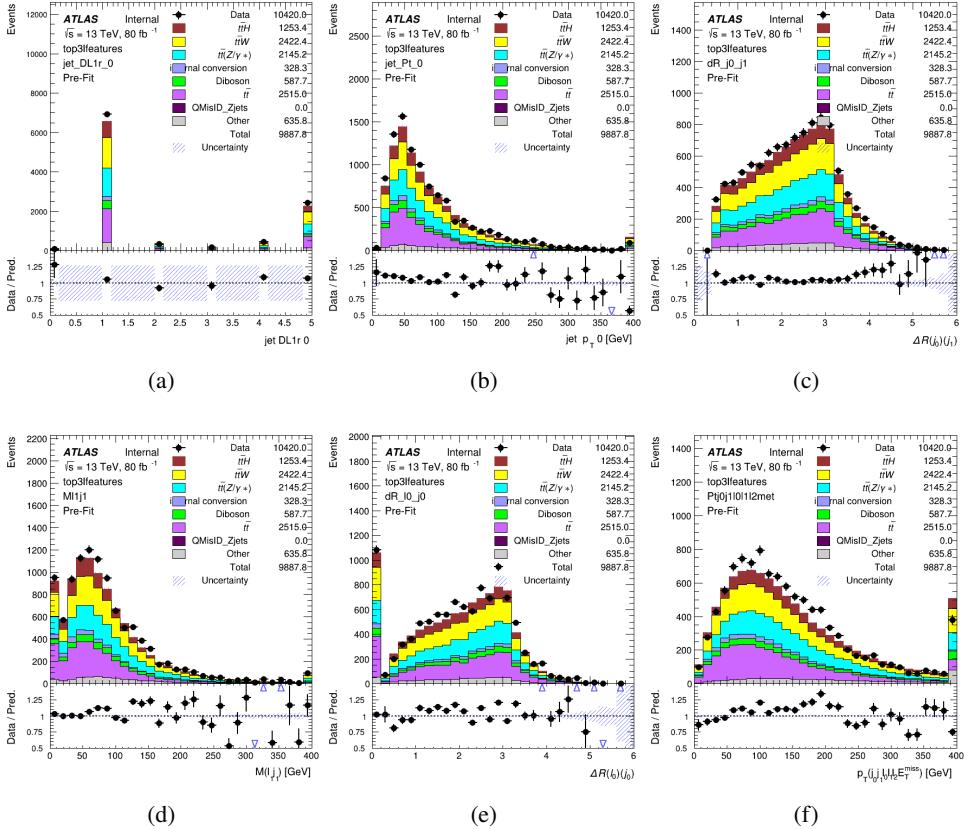


Figure 5.5: Data/MC comparisons of input features for top3l training for  $80 \text{ fb}^{-1}$  of data.

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

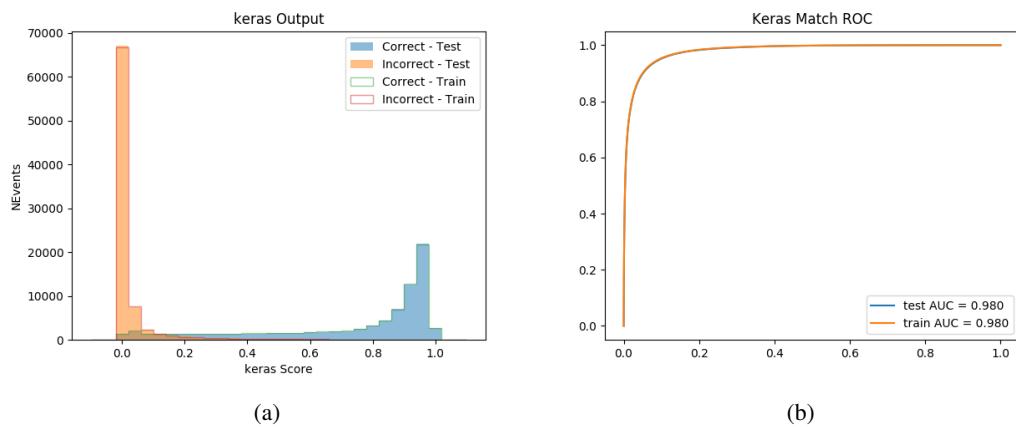


Figure 5.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

Lepton $p_T$ H	Lepton $p_T$ T	jet $p_T$ 0
jet $p_T$ 1	top $p_T$ 0	top $p_T$ 1
top $\eta$ 0	top $\eta$ 1	jet $\eta$ 0
jet $\eta$ 1	jet Phi 0	jet Phi 1
Lepton $\eta$ H	Lepton $\eta$ 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	$E_T^{\text{miss}}$
nJets	HT jets	

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

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

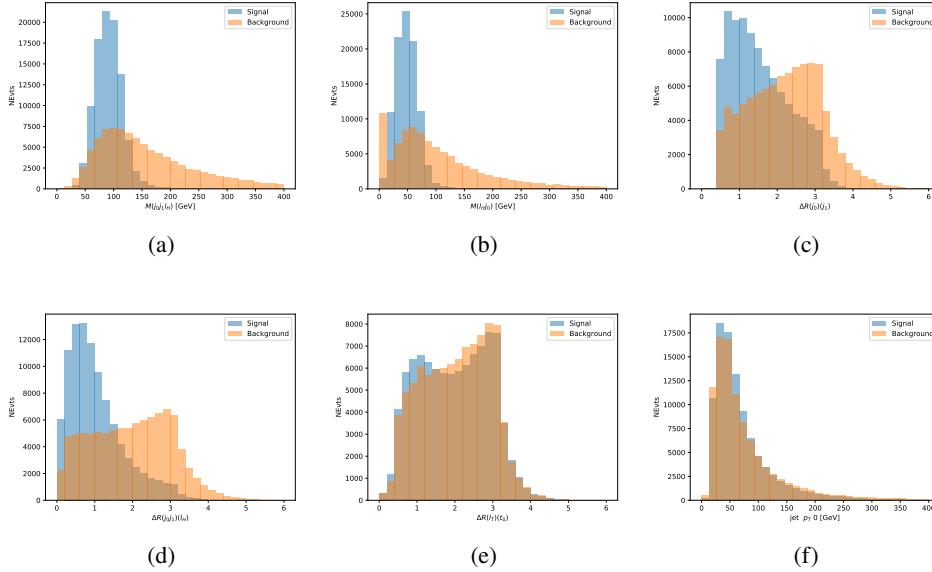


Figure 5.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

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

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

### 281 5.3.2 3l Semi-leptonic Channel

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

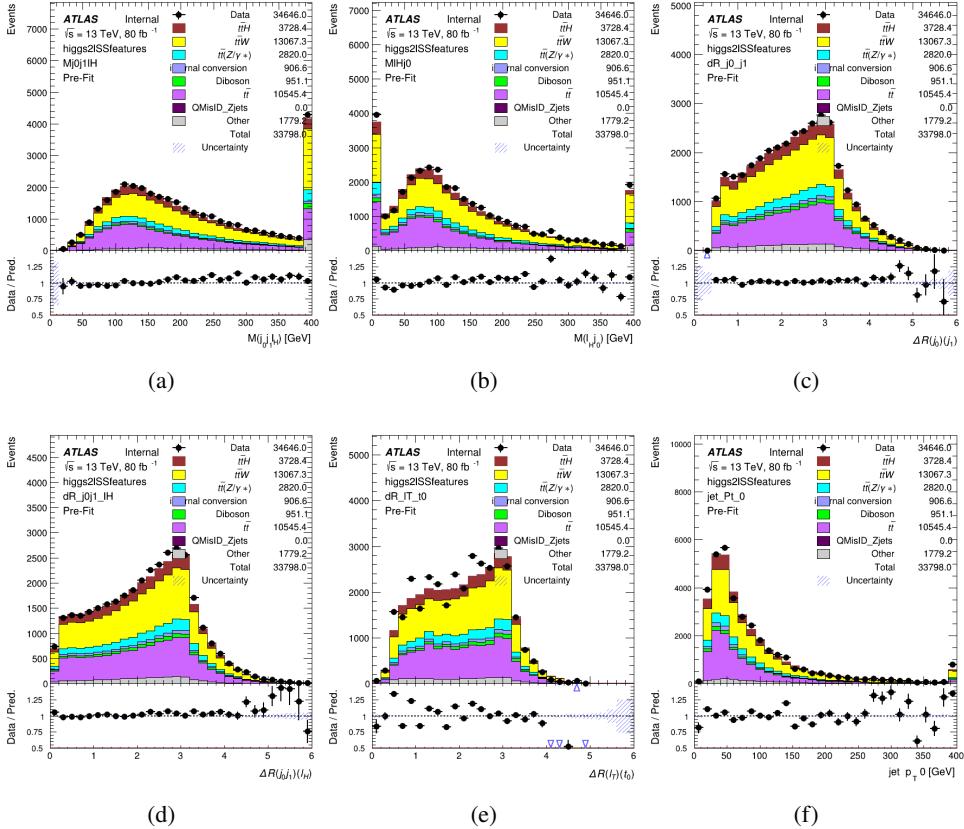


Figure 5.8: Data/MC comparisons of input features for higgs2lSS training for  $80 \text{ fb}^{-1}$  of data.

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

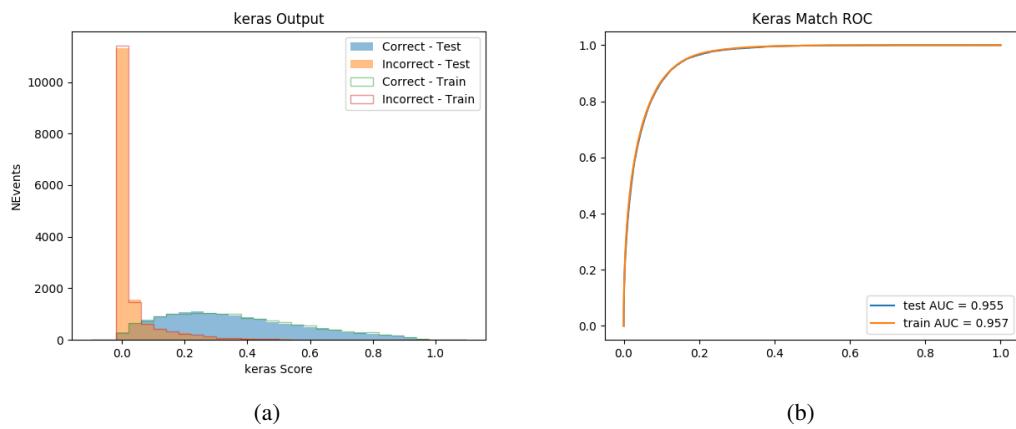


Figure 5.9: Result of the Higgs reconstruction algorithm in the 2ISS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

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

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

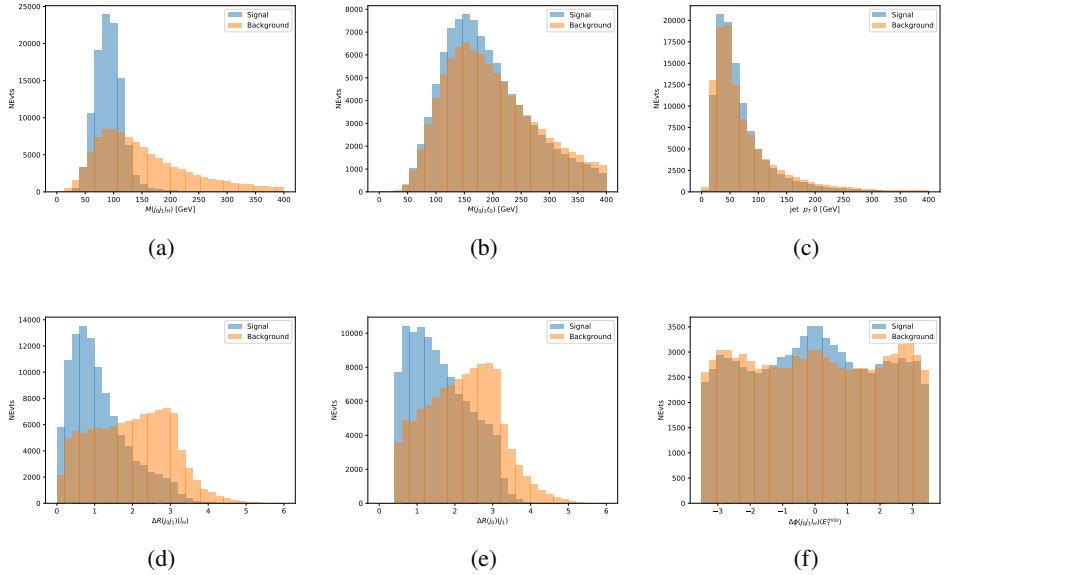


Figure 5.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

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

293 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events. Once  
 294 again, incorrect combinations are downsampled, such that the correct combinations are around  
 295 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN is  
 296 summarized in figure 5.3.2.

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

### 300 5.3.3 3l Fully-leptonic Channel

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

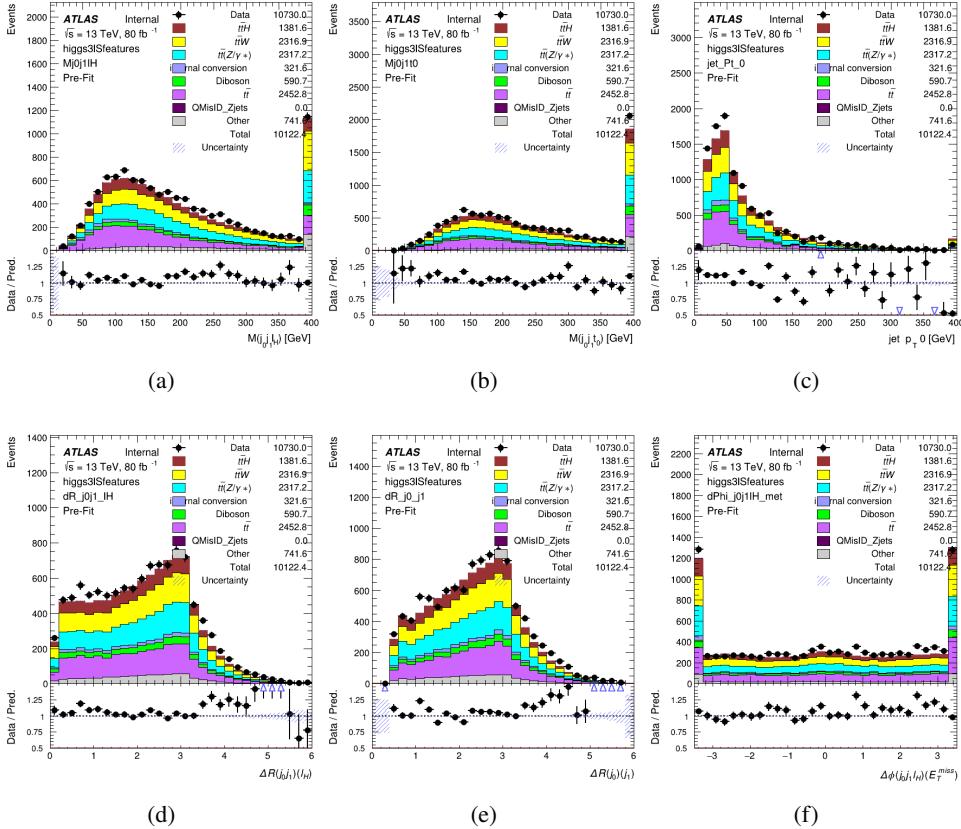


Figure 5.11: Data/MC comparisons of input features for higgs3IS training for  $80 \text{ fb}^{-1}$  of data.

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

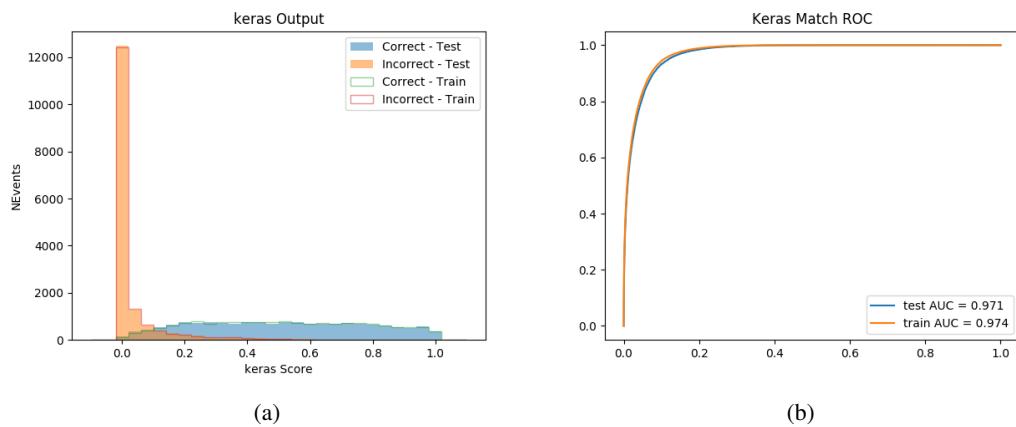


Figure 5.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries., (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

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

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

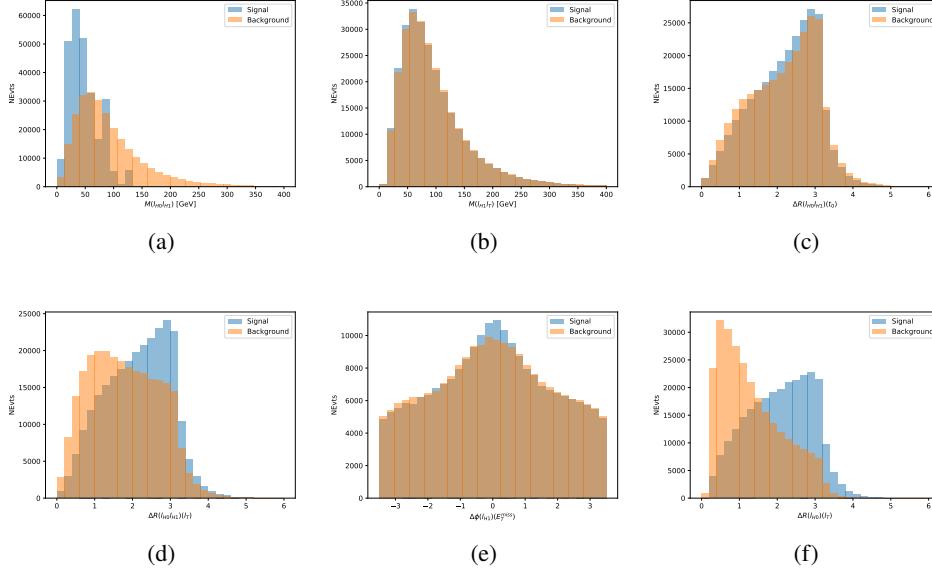


Figure 5.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

309 The modeling of these inputs is validated against data, with figure 5.14 showing good general  
 310 agreement between data and MC. Plots for the complete list of features can found in section A.  
 311 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events, with  
 312 10% of the dataset reserved for testing. The output of the model is summarized in figure 5.3.3.  
 313 The correct lepton is identified by the model for 80% of events in the testing data set.

#### 314 5.4 $p_T$ Prediction

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

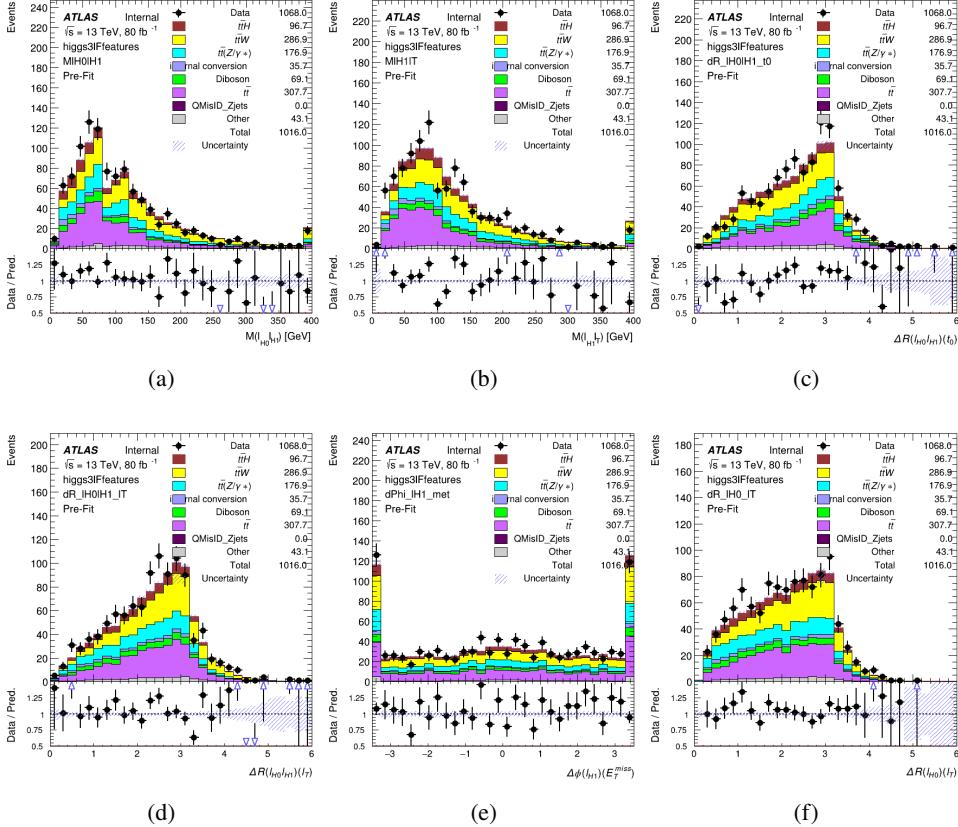


Figure 5.14: Data/MC comparisons of input features for higgs3IF training for  $80 \text{ fb}^{-1}$  of data.

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

Because the analysis uses a two bin fit of the Higgs  $p_T$ , the momentum reconstruction could be treated as a binary classification problem, rather than a regression problem. This approach is explored in detail in section A.4, and is found not to provide any significant increase in sensitivity. The regression approach is used because it provides more flexibility for future analyses, as it is independent of the cutoff between high and low  $p_T$ , as well as the number of bins. Further, a regression allows the output of the neural network to be more clearly understood, as it can be directly compared to a physics observable.

#### 5.4.1 2ISS Channel

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

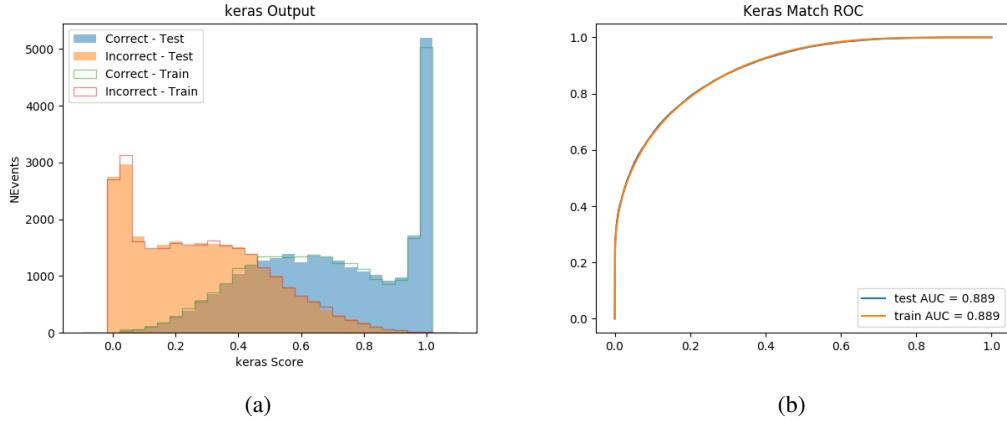


Figure 5.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

334 from the Higgs is labeled  $l_H$ , while the other lepton is labeled  $l_T$ , as it is assumed to have come  
 335 from the decay of one of the top quarks.  $b_0$  and  $b_1$  are the two b-jets identified by the b-jet  
 336 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs  
 337 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 $\eta$ 0
jet $\eta$ 1	jet Phi 0	jet Phi 1
jet $p_T$ 0	jet $p_T$ 1	Lepton $\eta$ H
Lepton $\phi$ H	Lepton $p_T$ H	Lepton $p_T$ T
$E_T^{\text{miss}}$	nJets	b-jet Reco Score
b-jet $p_T$ 0	b-jet $p_T$ 1	

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

338 The optimal neural network architecture for this channel is found to consist of 7 hidden layers with  
 339 60 nodes each. The input data set includes 1.2 million events, 10% of which is used for testing,  
 340 the other 90% for training. Training is found to converge after around 150 epochs.

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

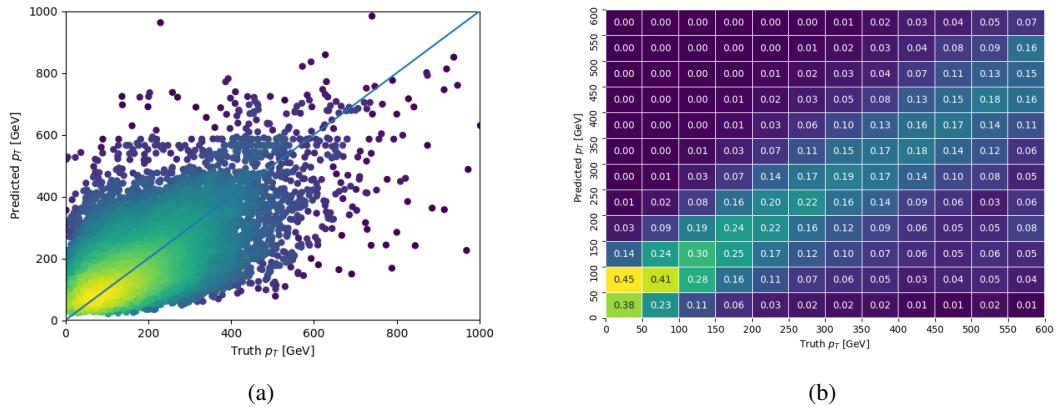


Figure 5.16: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 2lSS  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been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

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

#### 351 5.4.2 3l Semi-leptonic Channel

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

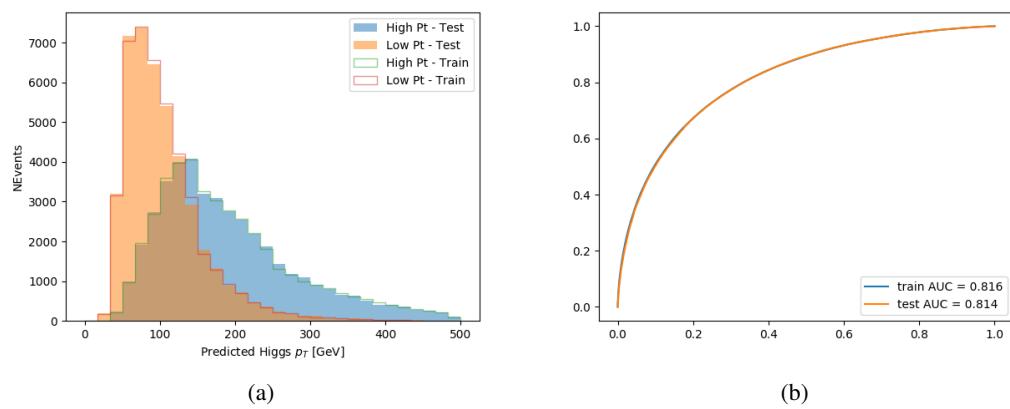


Figure 5.17: (a) shows the reconstructed Higgs  $p_T$  for 2lSS events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

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

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

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

The optimal neural network architecture for this channel is found to consist of 7 hidden layers with 80 nodes each. The 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) shows a scatterplot 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  $p_T$ .

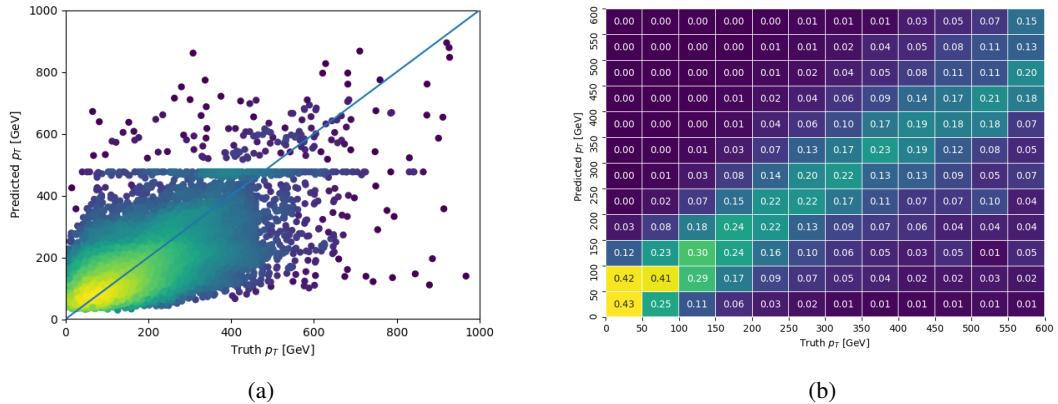


Figure 5.18: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 3lS  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

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

### 5.4.3 3l Fully-leptonic Channel

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

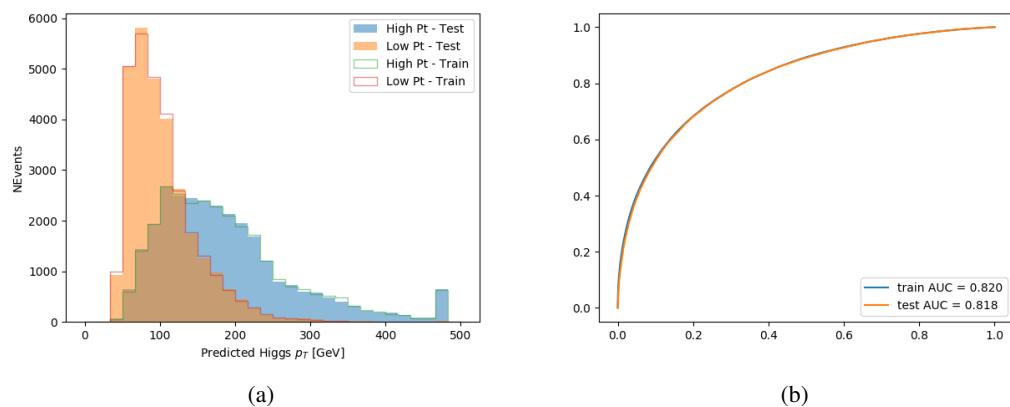


Figure 5.19: (

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.

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 $\eta$ $H_0$	Lepton $\eta$ $H_1$	Lepton $\eta$ T
Lepton $p_T$ $H_0$	Lepton $p_T$ $H_1$	Lepton $p_T$ T
$E_T^{\text{miss}}$	topScore	b-jet $p_T$ 0
b-jet $p_T$ 1		

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

<sup>371</sup>  $l_{H0}$  and  $l_{H1}$  represent the two leptons identified by the Higgs reconstruction model as originating  
<sup>372</sup> from the Higgs, while  $l_T$  is the other lepton in the event. The Higgs Reco Score and b-jet Reco  
<sup>373</sup> Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification algorithm,  
<sup>374</sup> respectively.

<sup>375</sup> The optimal neural network architecture for this channel is found to consist of 5 hidden layers  
<sup>376</sup> with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for testing,  
<sup>377</sup> the other 90% for training. Training is found to converge after around 150 epochs.

<sup>378</sup> The predicted transverse momentum, as a function of the truth  $p_T$ , is shown in figure ??.

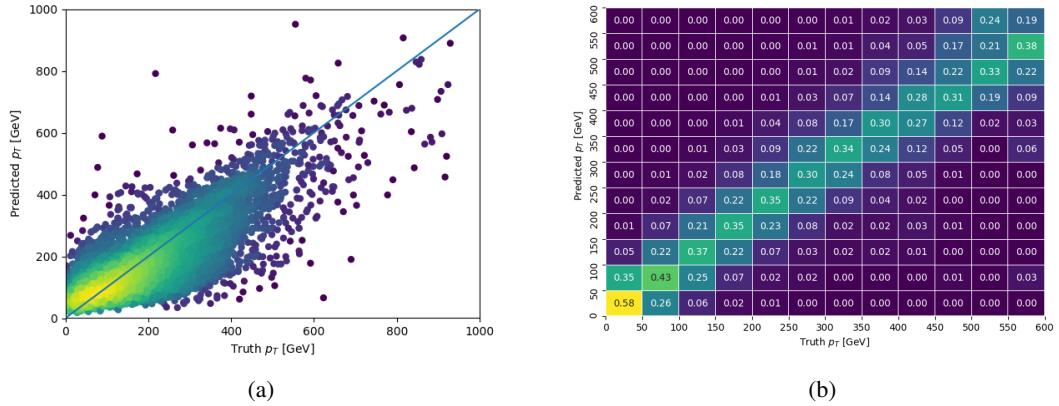


Figure 5.20: The regressed Higgs momentum spectrum as a function of the truth  $p_T$  for 3lF  $t\bar{t}H$  events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

<sup>379</sup> When split into high and low  $p_T$ , based on a cutoff of 150 GeV, the

## <sup>380</sup> 5.5 3l Decay Mode

<sup>381</sup> In the 3l channel, there are two possible ways for the Higgs to decay, both involving intermediate  
<sup>382</sup> W boson pairs: Either both W bosons decay leptonically, in which case the reconstructed decay  
<sup>383</sup> consists of two leptons (referred as the fully-leptonic 3l channel), or one W decays leptonically  
<sup>384</sup> and the other hadronically, giving two jets and one lepton in the final state (referred to as the  
<sup>385</sup> semi-leptonic 3l channel). In order to accurately reconstruct the Higgs, it is necessary to identify  
<sup>386</sup> which of these decays took place for each 3l event.

<sup>387</sup> The kinematics of each event, along with the output scores of the Higgs and top reconstruction  
<sup>388</sup> algorithms, are used to distinguish these two possible decay modes. The particular inputs used  
<sup>389</sup> are listed in table 13.

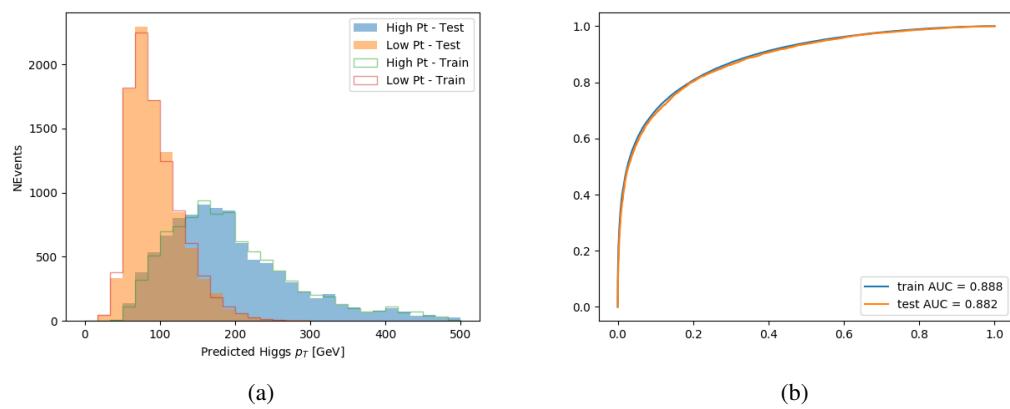


Figure 5.21: (a) shows the reconstructed Higgs  $p_T$  for 3lF events with truth  $p_T > 150$  GeV and  $< 150$  GeV, while (b) shows the ROC curve for those two sets of events.

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

Table 13: Input features

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

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

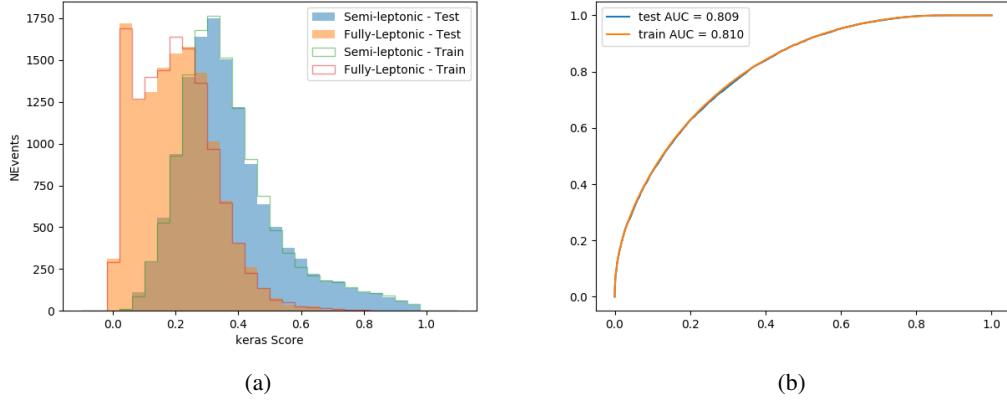


Figure 5.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

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

## 396 6 Signal Region Definitions

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

### 400 6.1 Pre-MVA Event Selection

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

- 403 • Two very tight, same-charge, light leptons with  $p_T > 20$  GeV
- 404 •  $>=4$  reconstructed jets,  $>=1$  b-tagged jets
- 405 • No reconstructed tau candidates

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

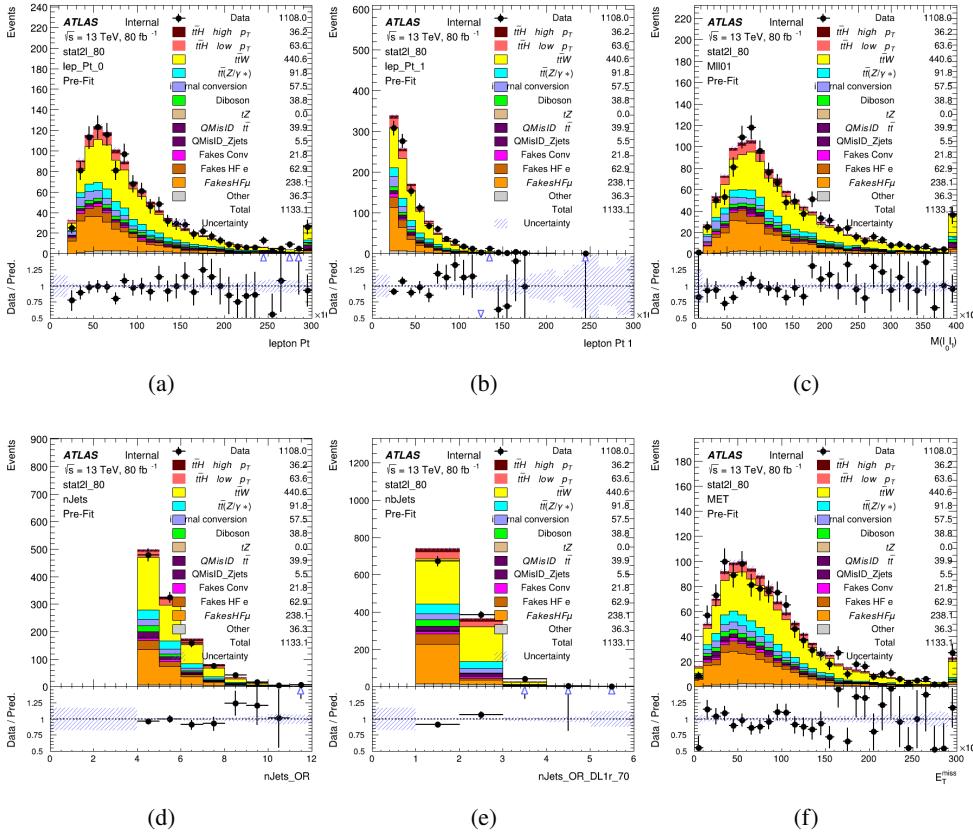


Figure 6.1:

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

## 413 6.2 Event MVA

414 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events  
 415 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic (3lF).  
 416 In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost [xgboost]

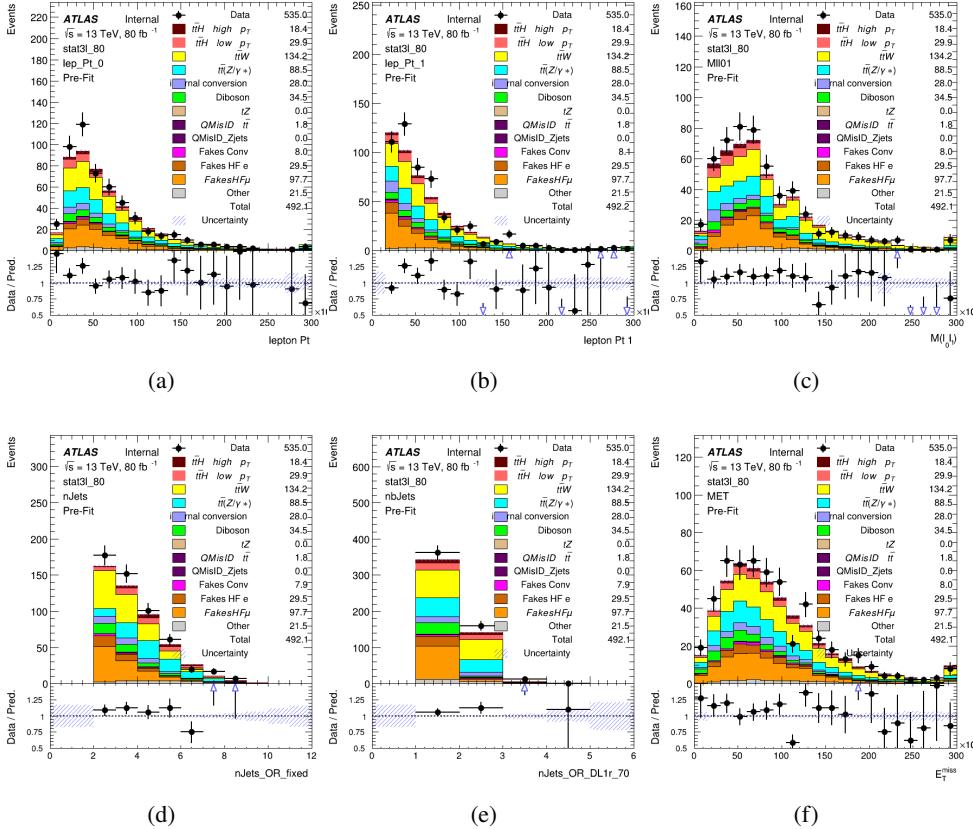


Figure 6.2:

417 are trained using the kinematics of signal and background events derived from Monte Carlo  
 418 simulations. Events are weighted in the BDT training by the weight of each Monte Carlo event.

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

425 The following features are used in both the high and low  $p_T$  2ISS BDTs:

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

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

<sup>426</sup> While for each of the 31 BDTs, the features listed below are used for training:

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

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

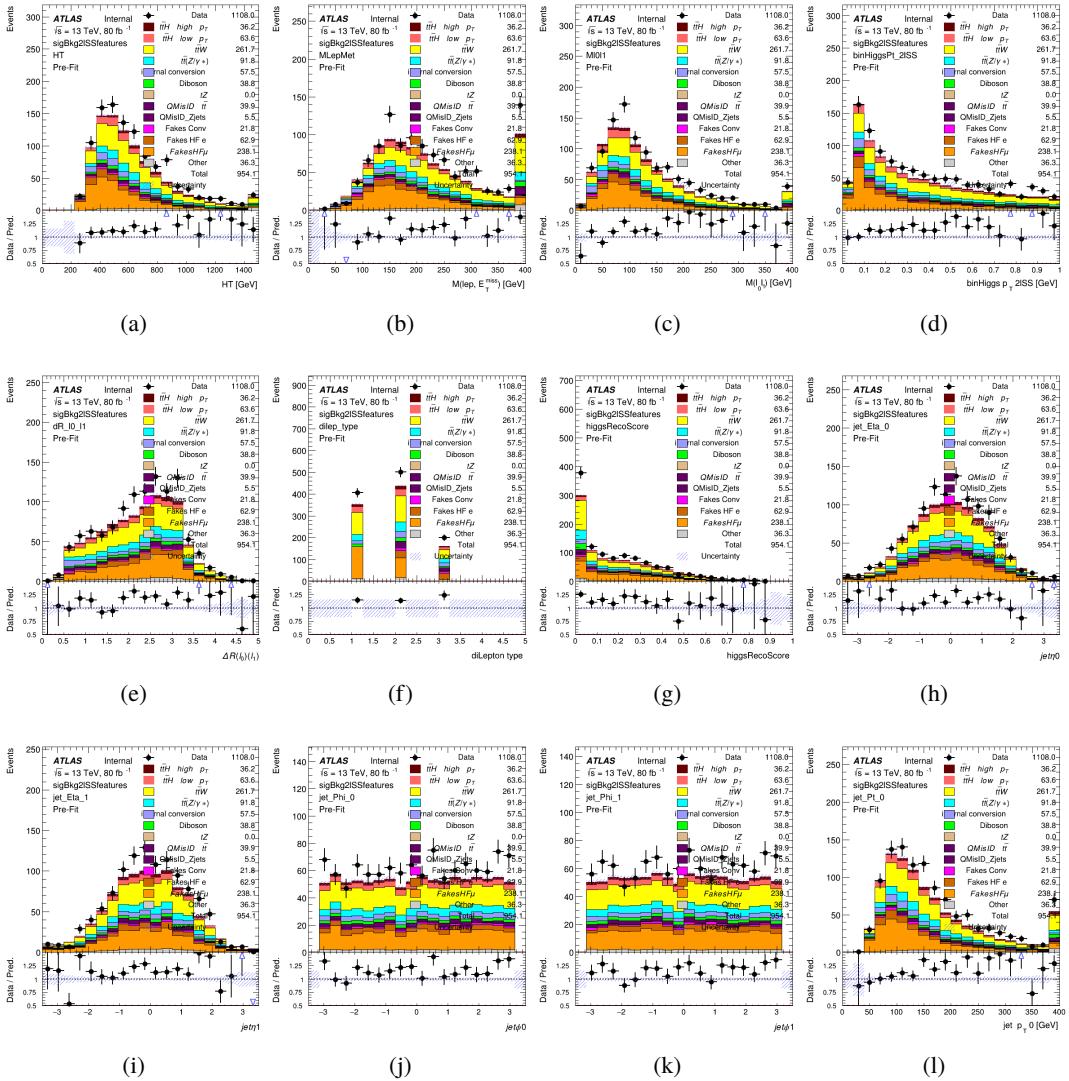


Figure 6.3:

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

### 429 6.3 Signal Region Definitions

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

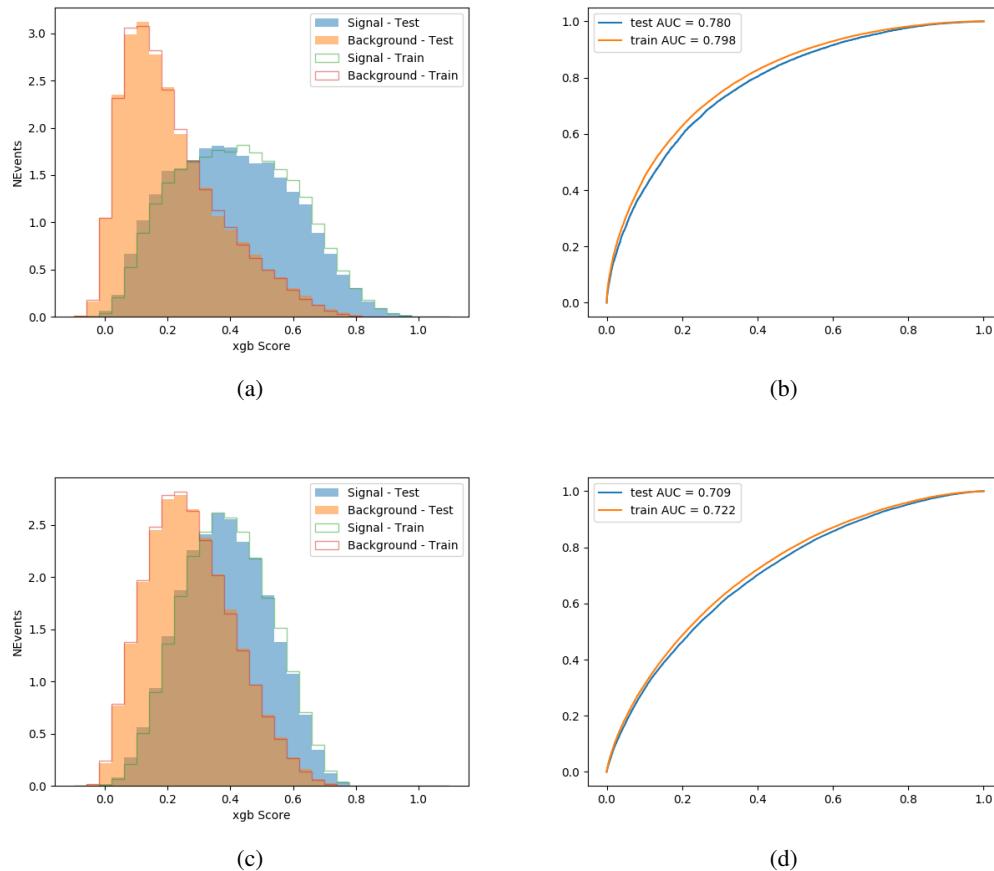


Figure 6.4:

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

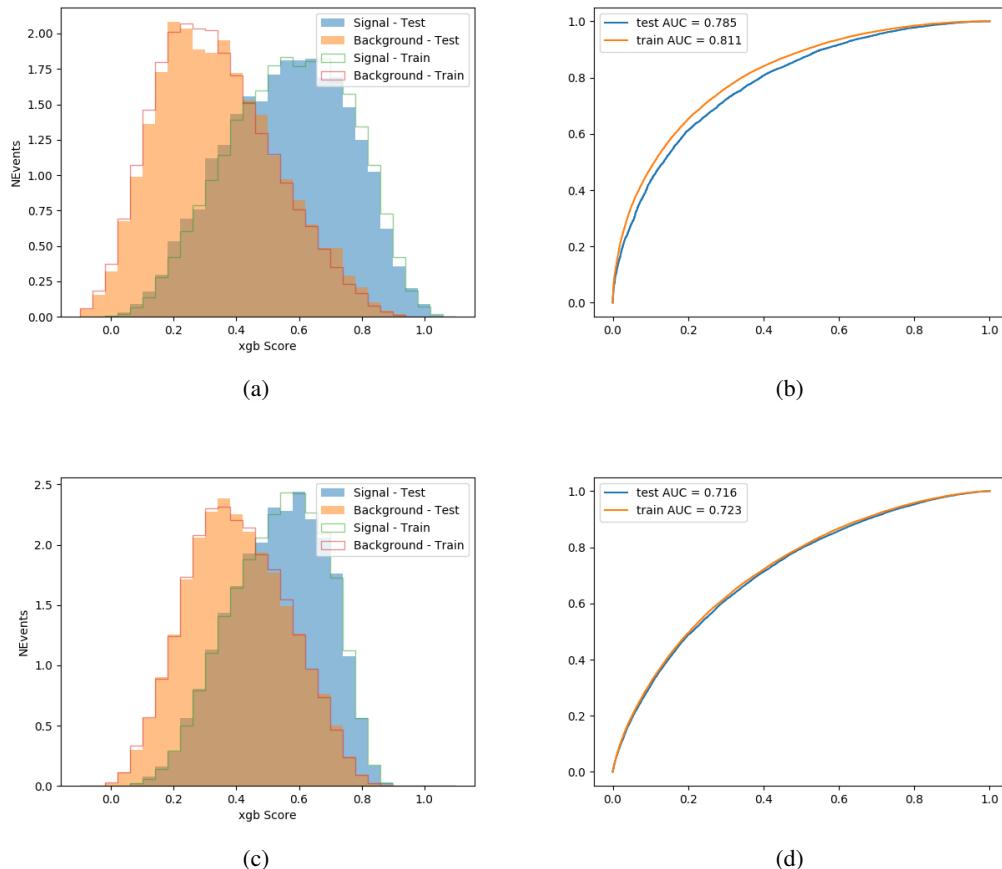


Figure 6.5:

437 **6.3.1 2lSS**438 **6.3.2 3l – Semi – leptonic**439 **6.3.3 3l – Fully – leptonic**440 

## 7 Systematic Uncertainties

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

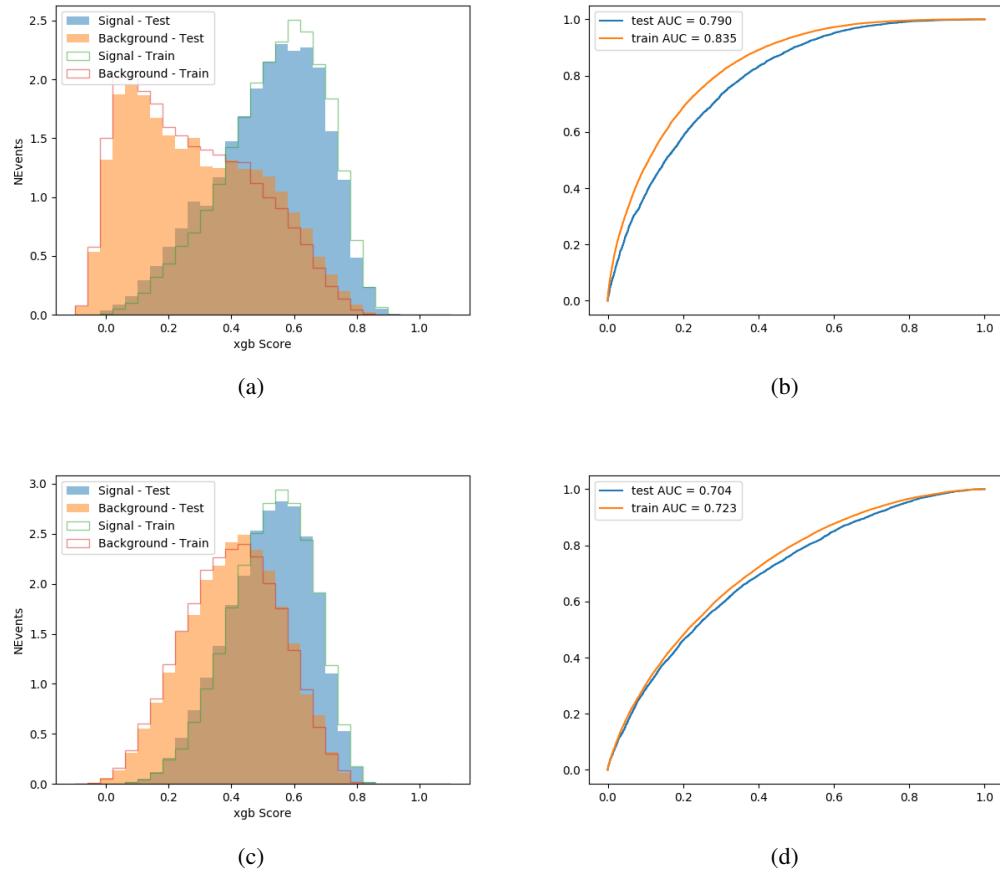


Figure 6.6:

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

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

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

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

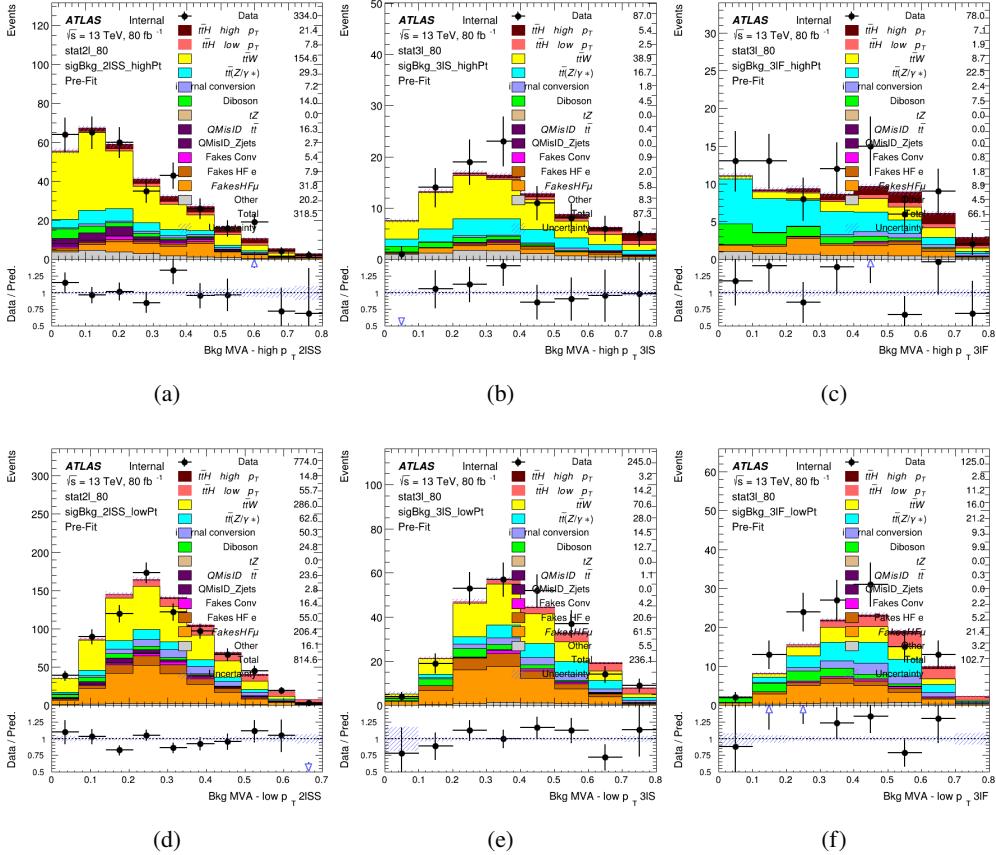


Figure 6.7: scores

## 457 8 Results

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

### 460 8.1 Results - $80 \text{ fb}^{-1}$

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

Table 16: Sources of systematic uncertainty considered in the analysis. Some of the systematic uncertainties are split into several components, as indicated by the number in the rightmost column.

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

## 462 8.2 Projected Results - $140 \text{ fb}^{-1}$

## 463 9 Conclusion

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

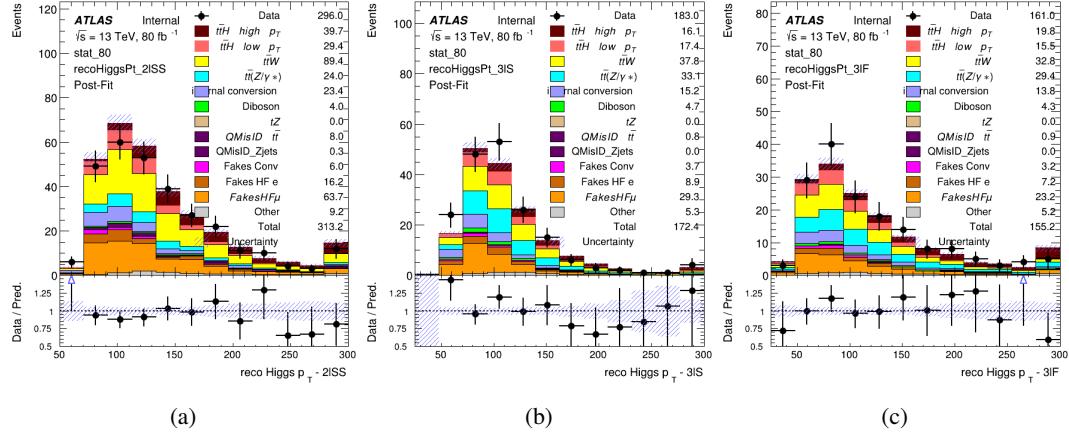


Figure 8.1:

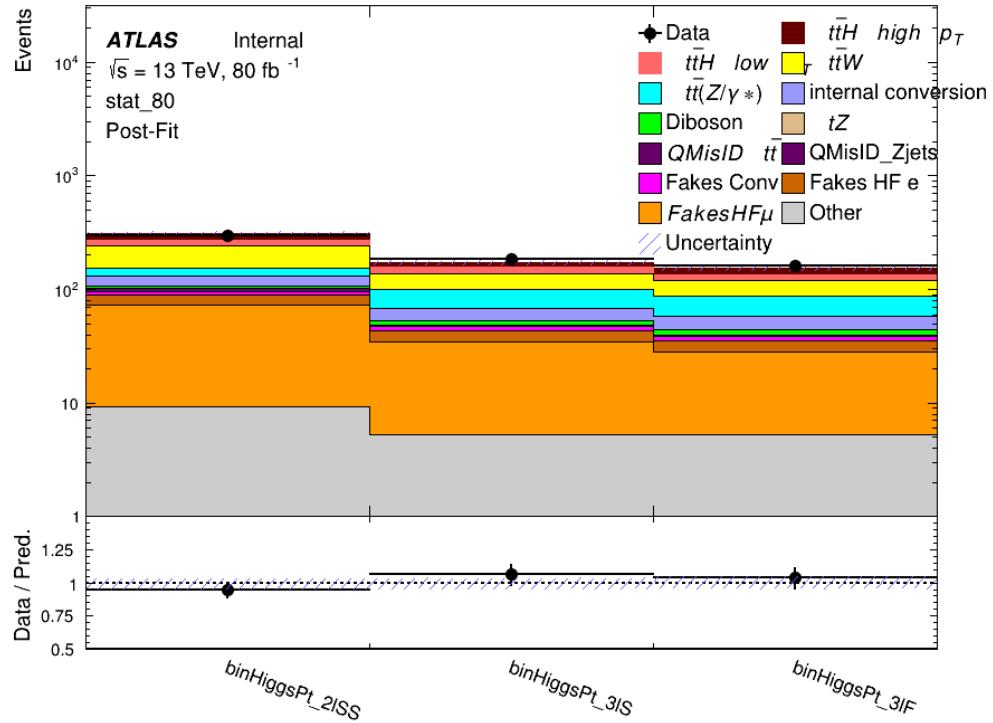


Figure 8.2: Post-fit summary of fit.

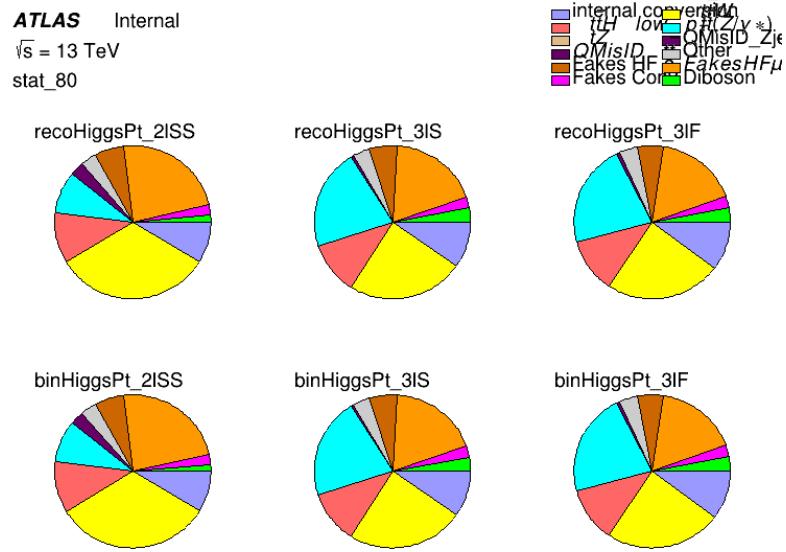


Figure 8.3: Background composition of the fit regions.

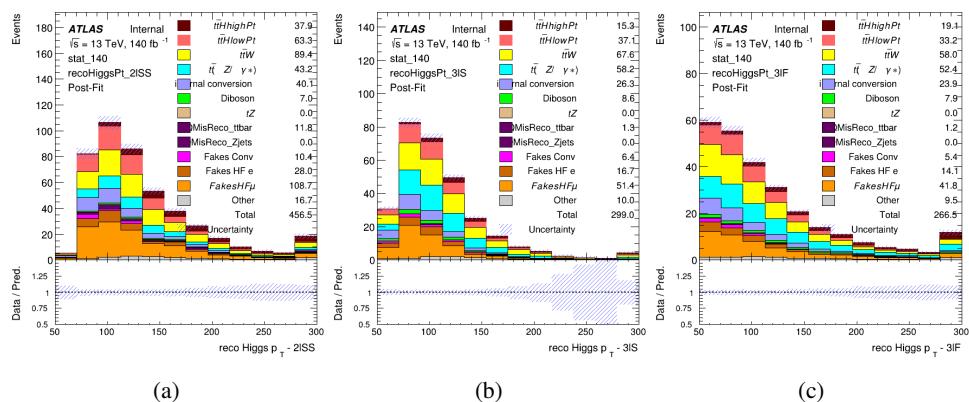


Figure 8.4:

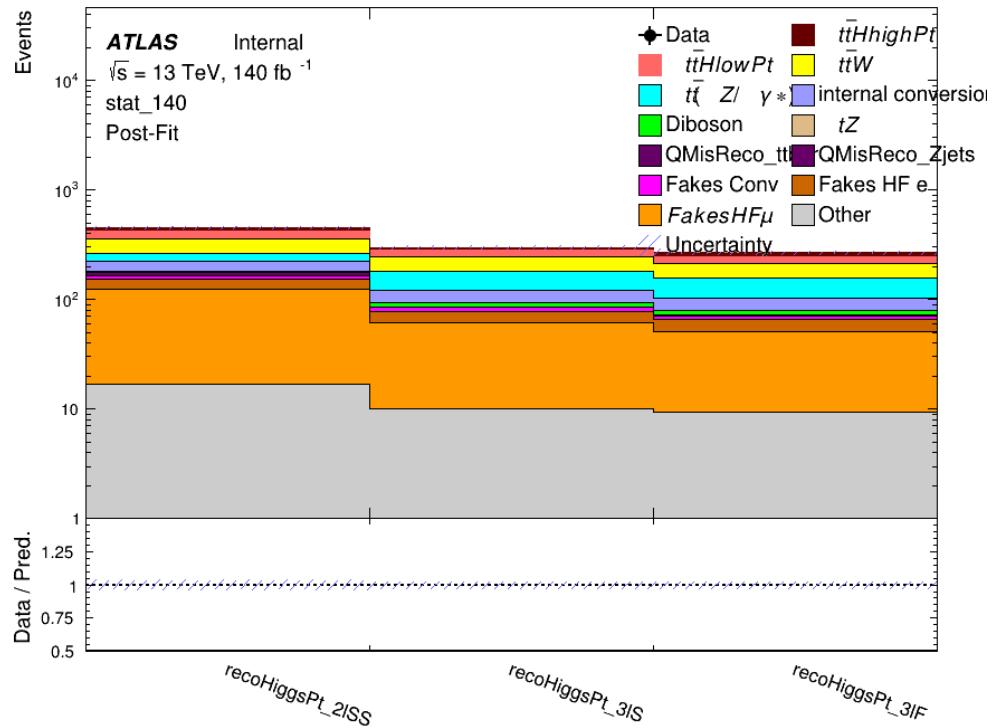


Figure 8.5: Post-fit summary of fit.

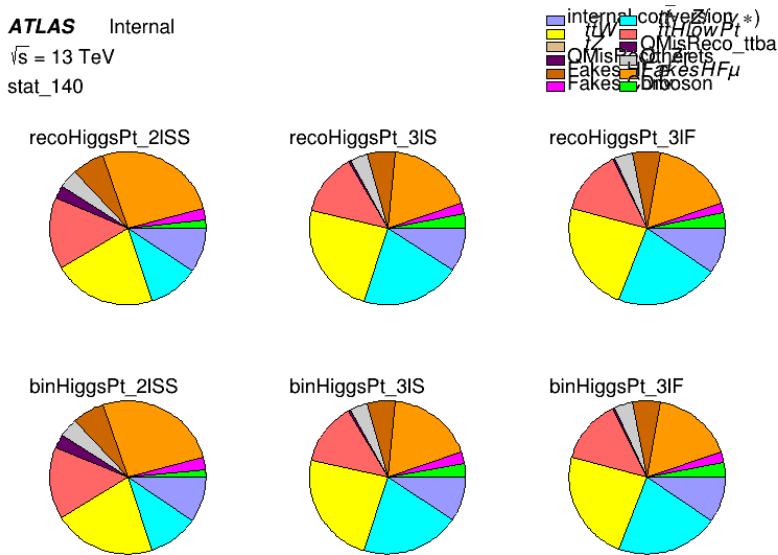


Figure 8.6: Background composition of the fit regions.

## 468 List of contributions

469

<sup>470</sup> **Appendices**

<sup>471</sup> **A Machine Learning Models**

<sup>472</sup> The following section provides details of the various MVAs as well as a few studies performed in  
<sup>473</sup> support of this analysis, exploring alternate decisions and strategies.

<sup>474</sup> **A.1 Higgs Reconstruction Models**

<sup>475</sup> **A.1.1 b-jet Identification Features - 2lSS**

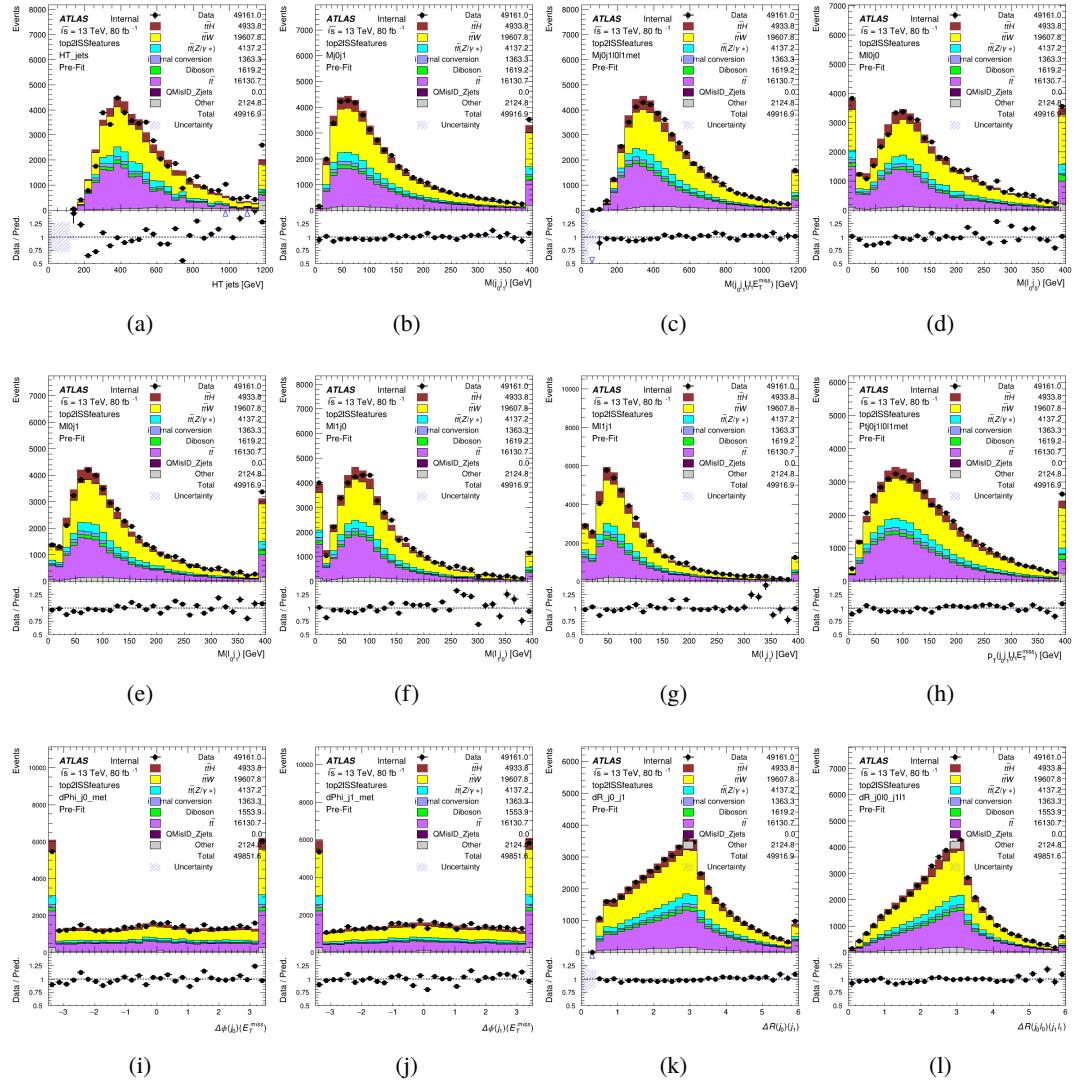


Figure A.1: Input features for top2lSS

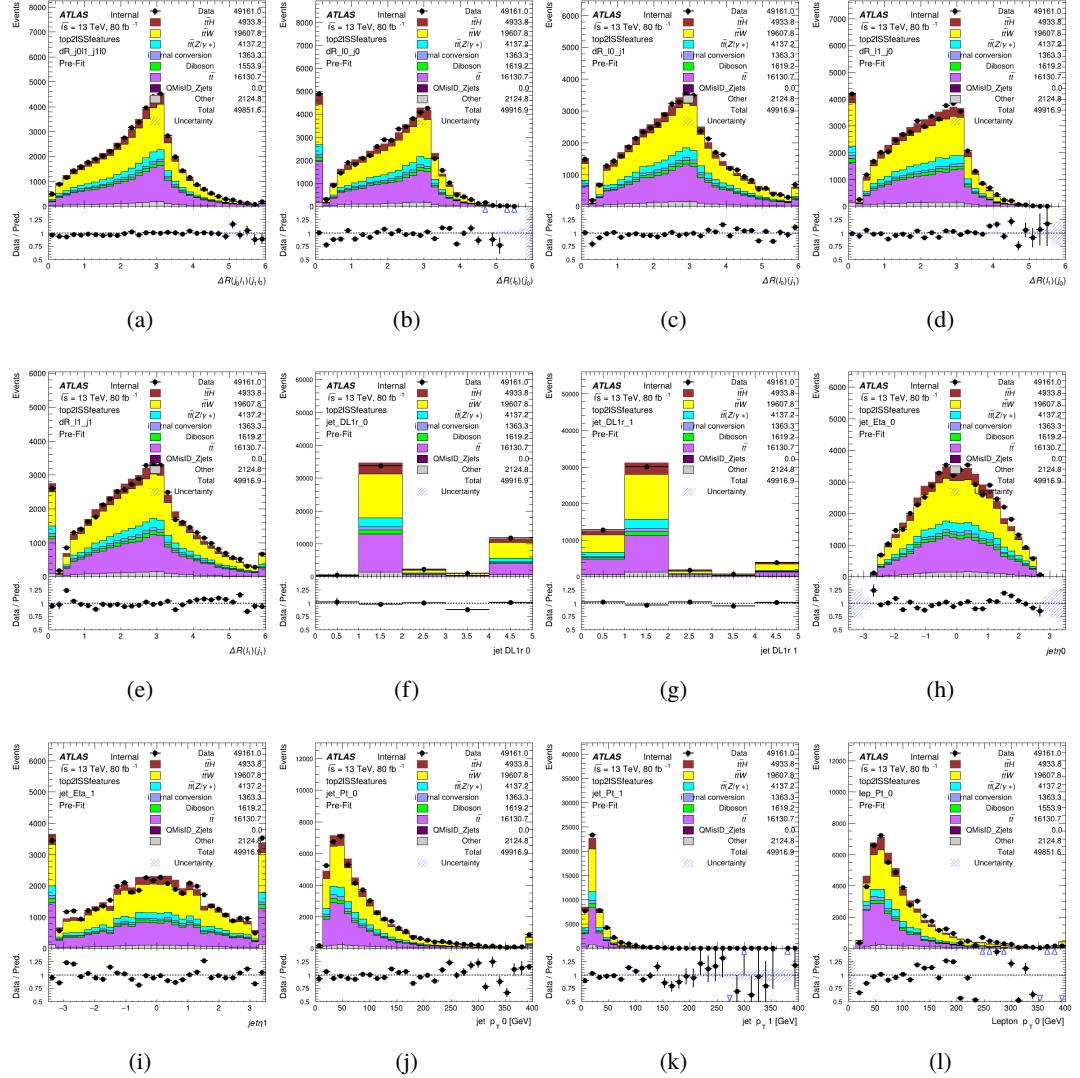


Figure A.2: Input features for top2lSS

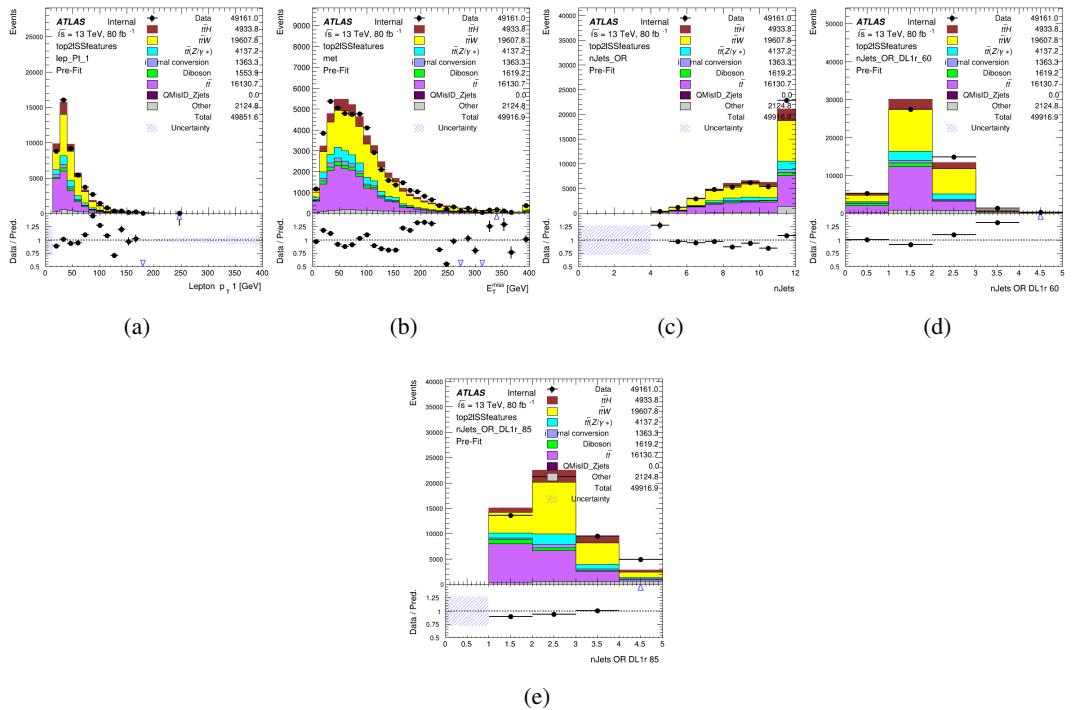


Figure A.3: Input features for top21SS

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

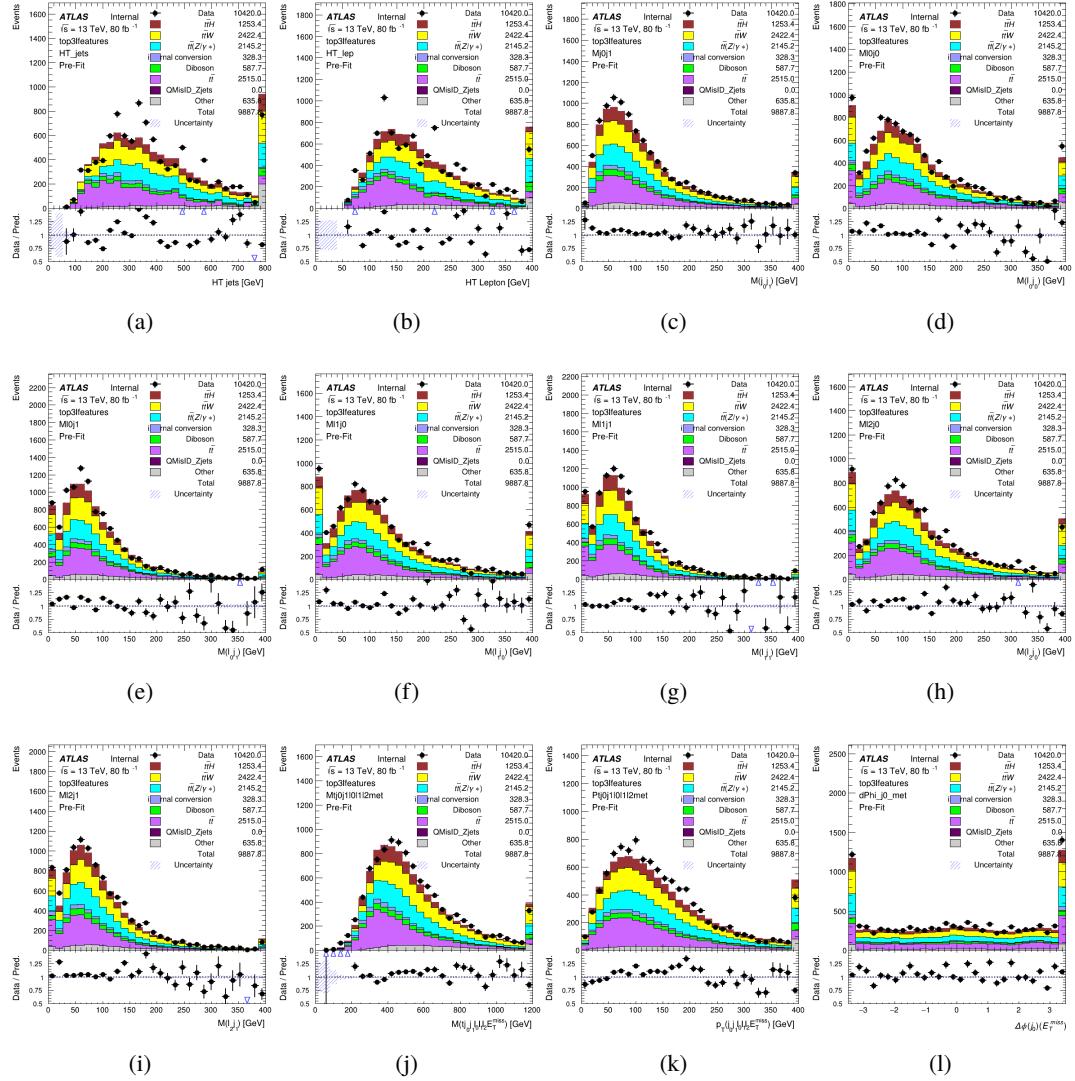


Figure A.4: Input features for top31

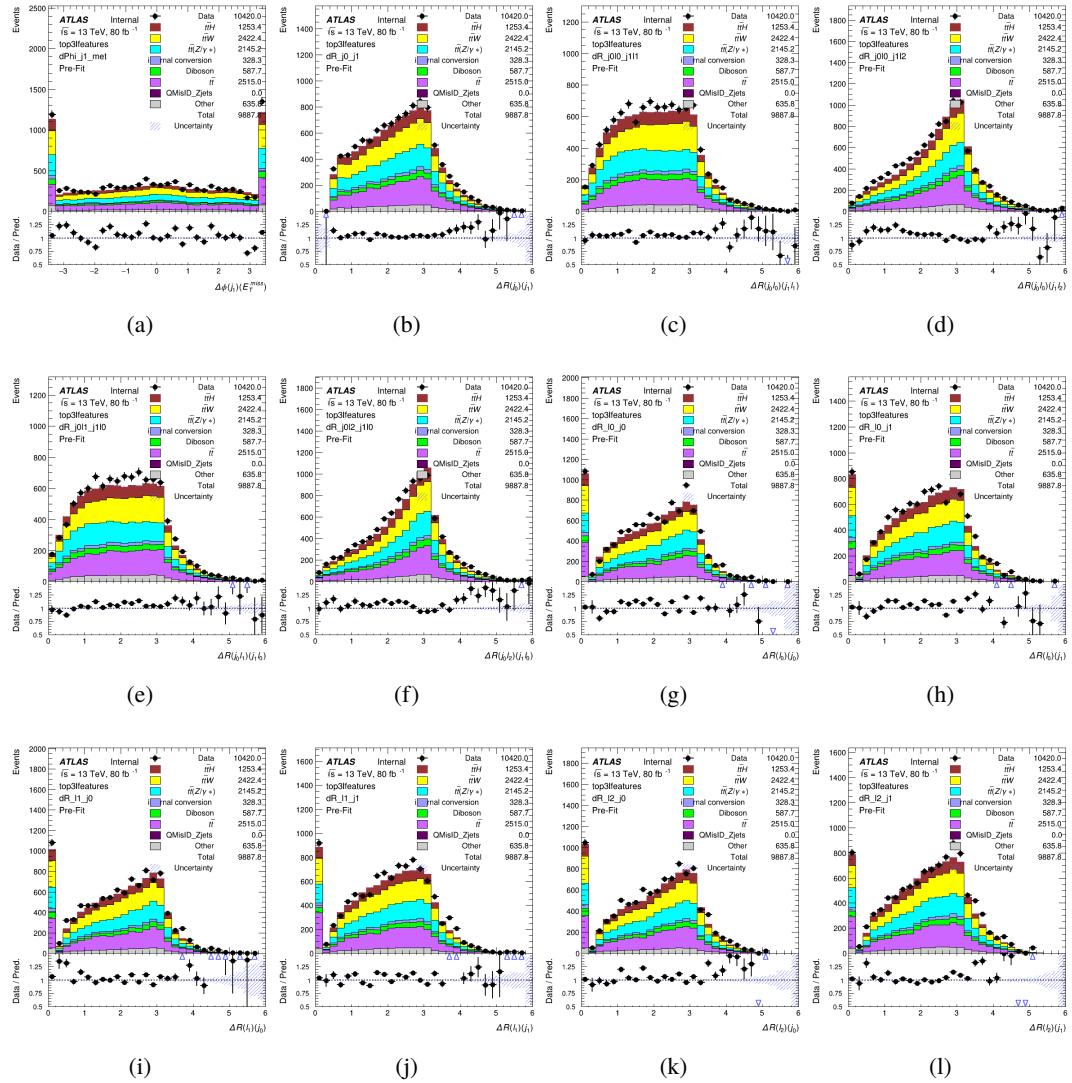


Figure A.5: Input features for top31

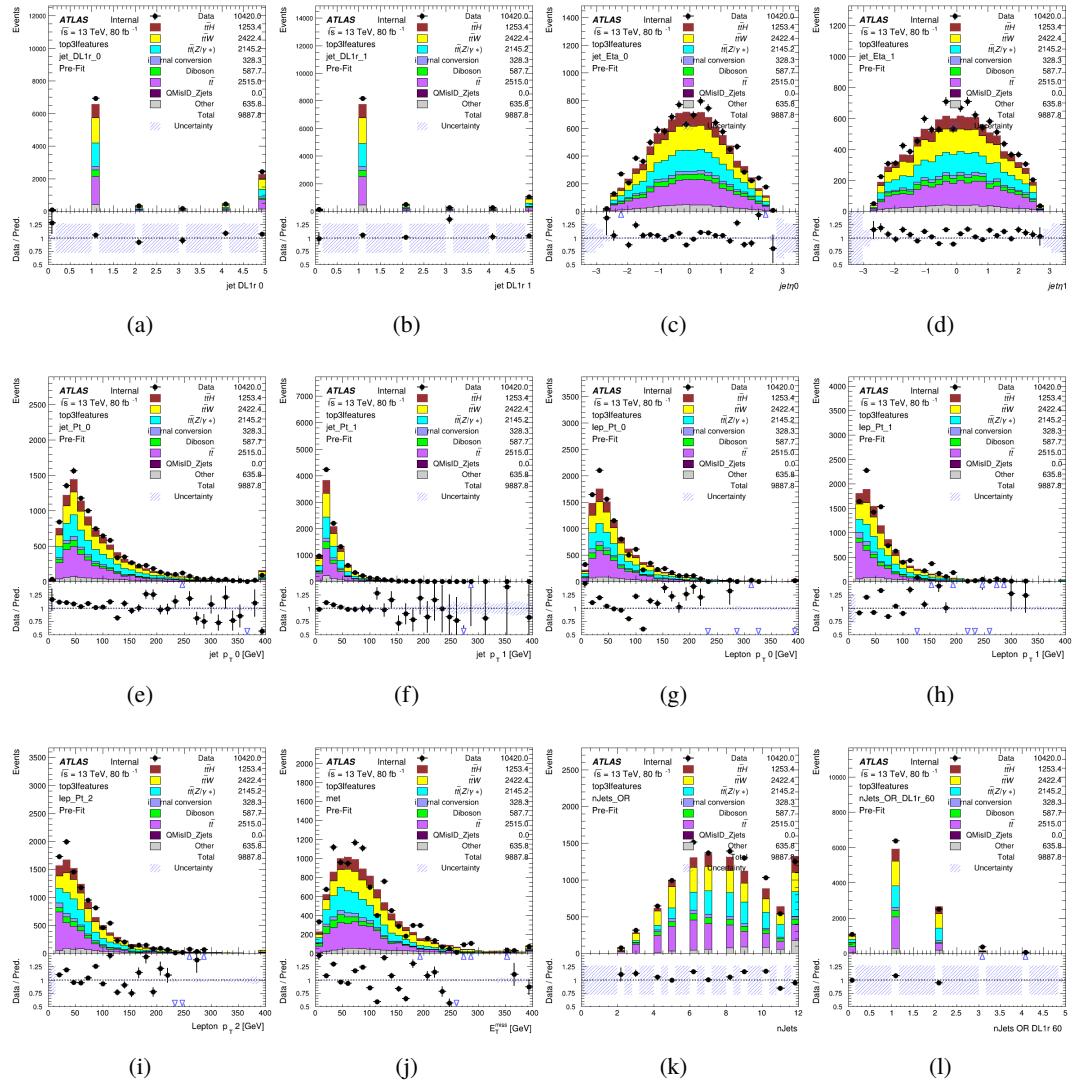
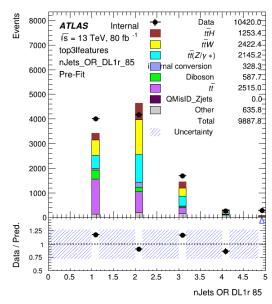


Figure A.6: Input features for top31



(a)

Figure A.7: Input features for top3l

<sup>477</sup> **A.1.3 Higgs Reconstruction Features - 2lSS**

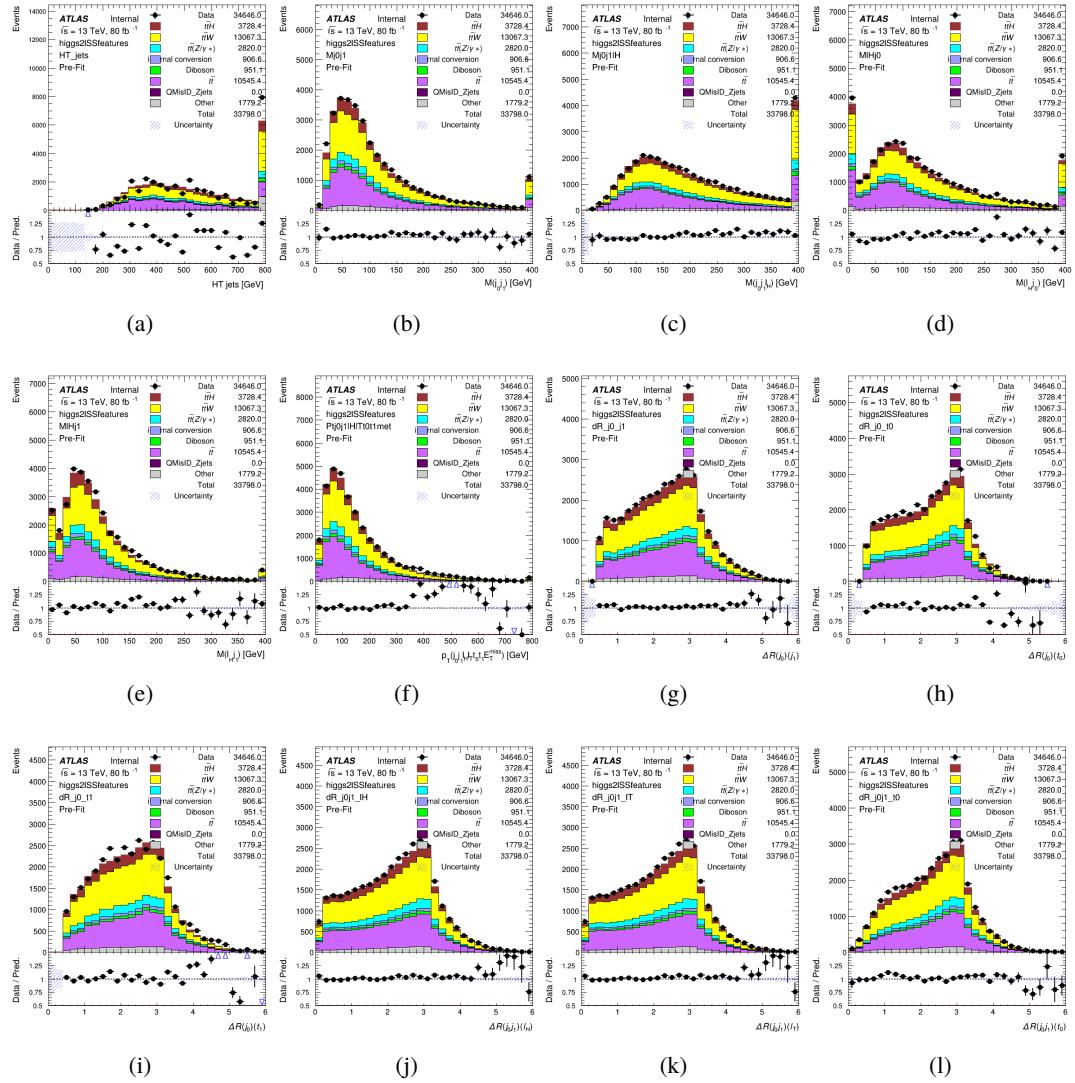


Figure A.8: Input features for higgs2lSS

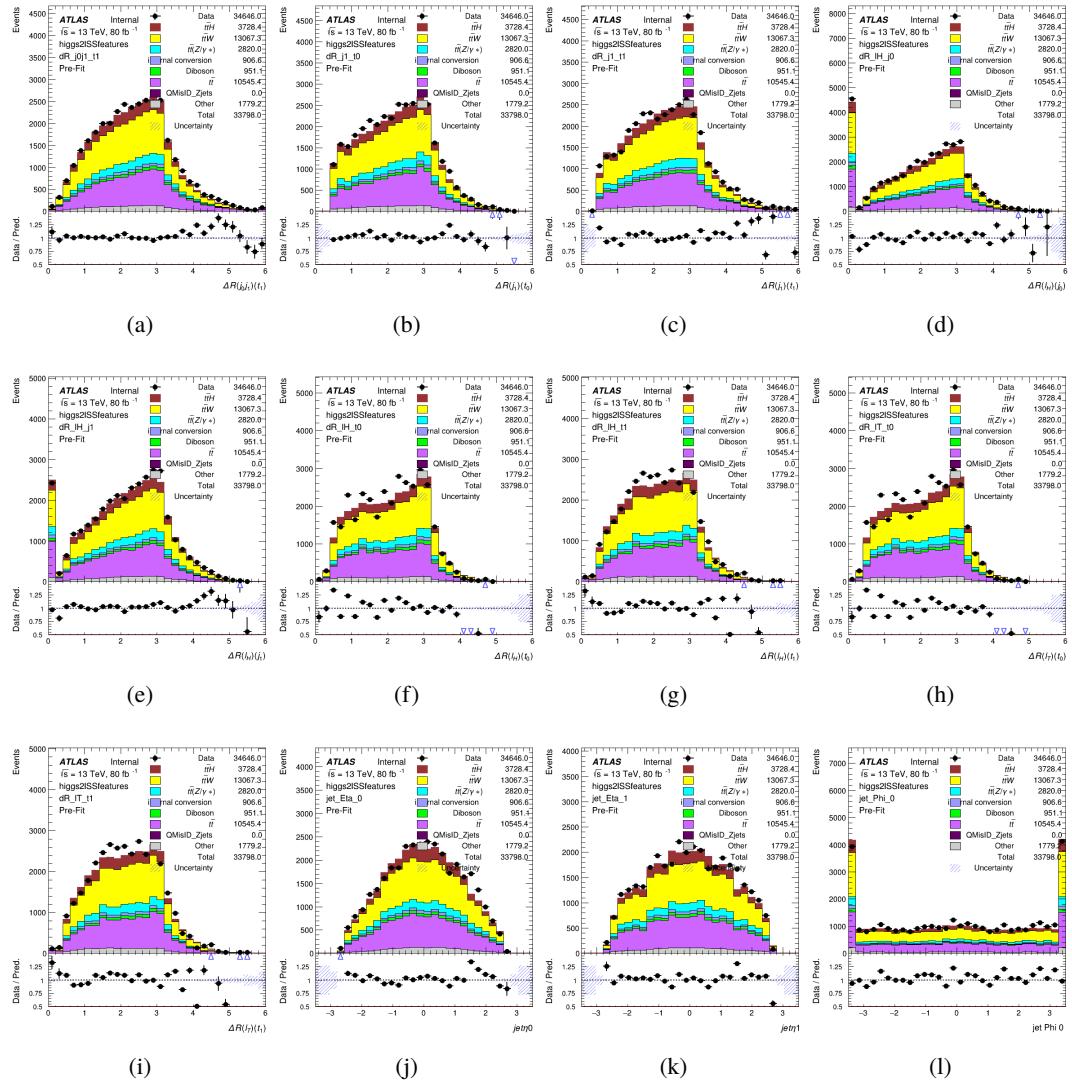


Figure A.9: Input features for higgs2lSS

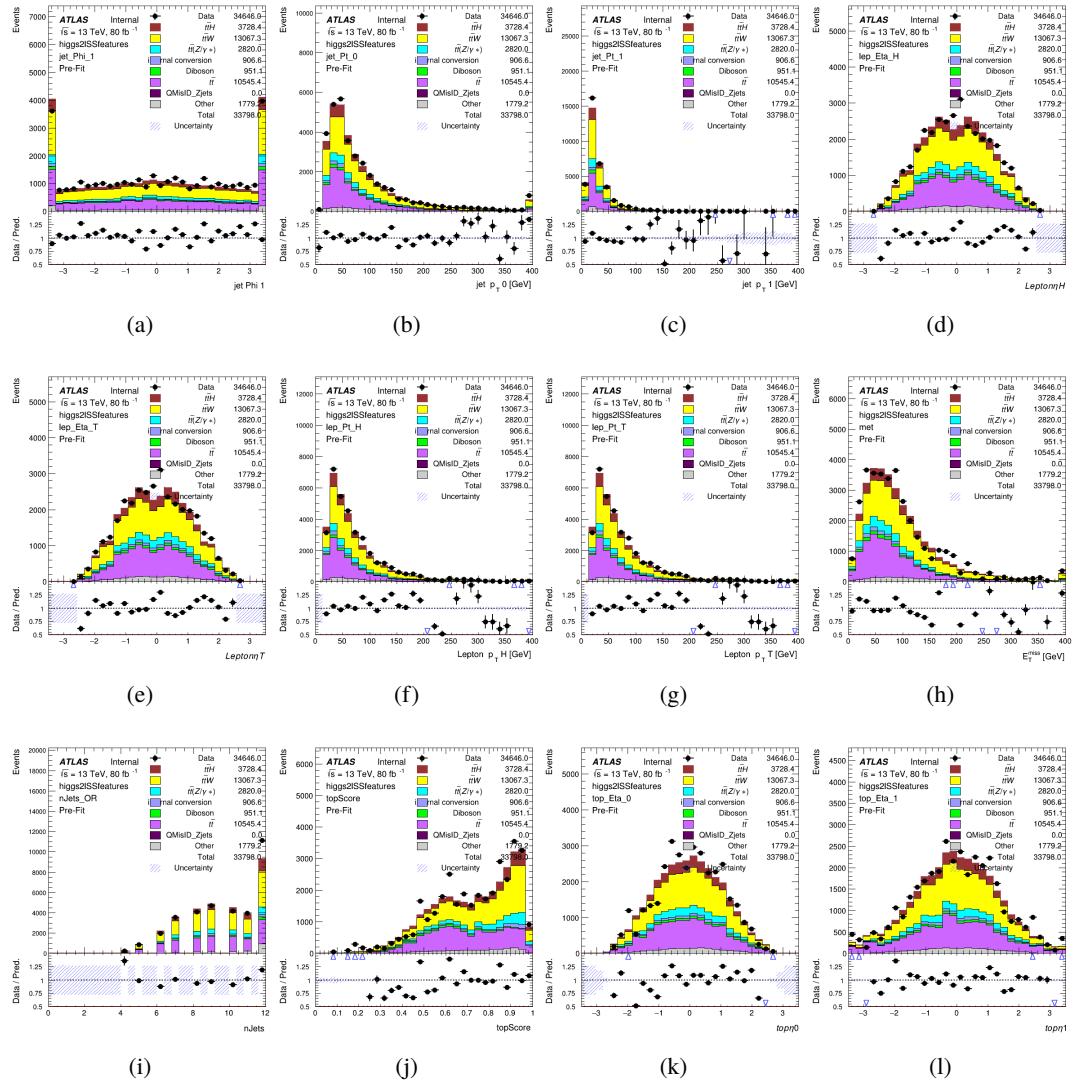


Figure A.10: Input features for higgs2ISS

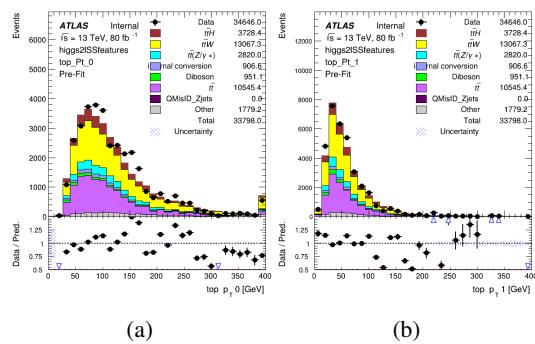


Figure A.11: Input features for higgs2ISS

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

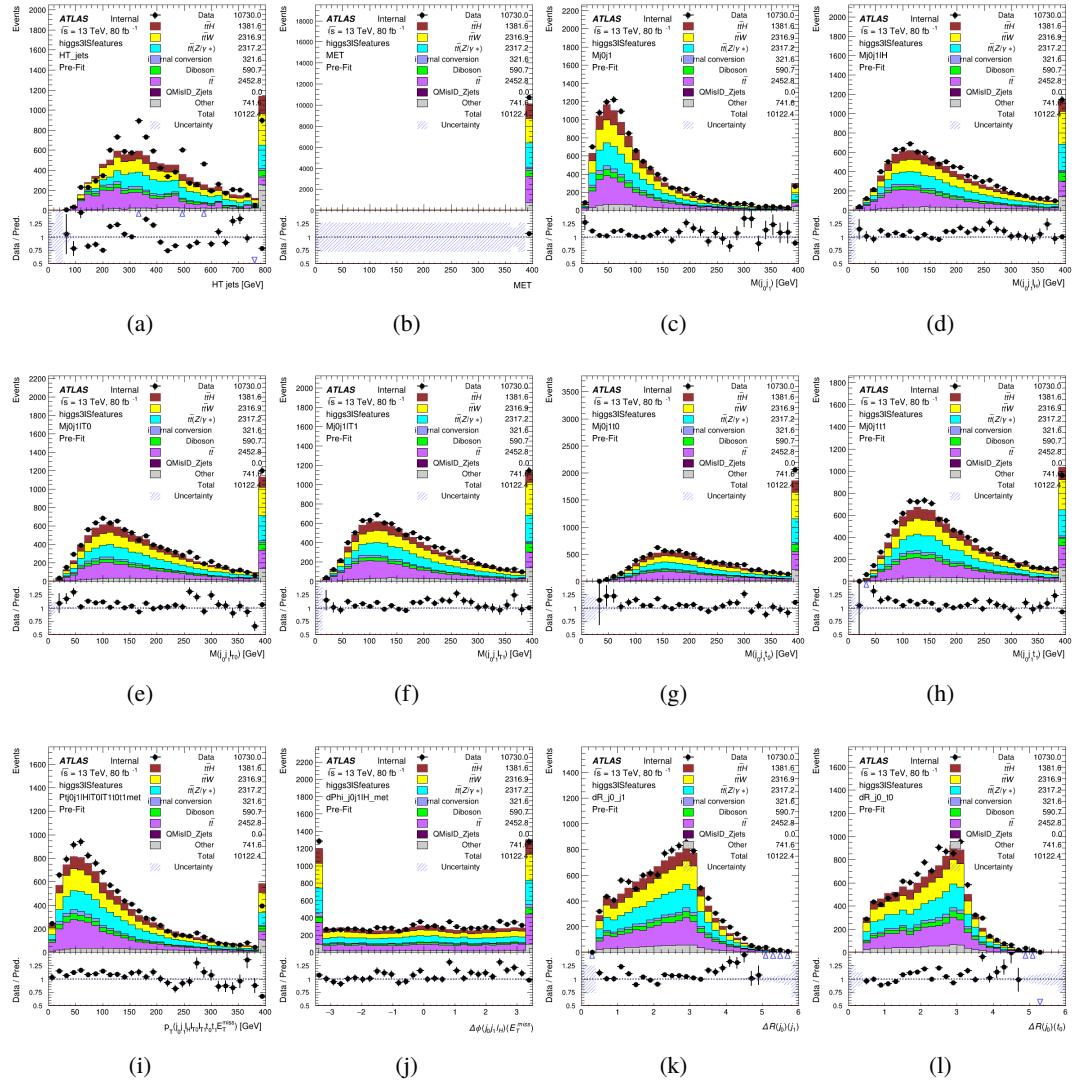


Figure A.12: Input features for higgs3IS

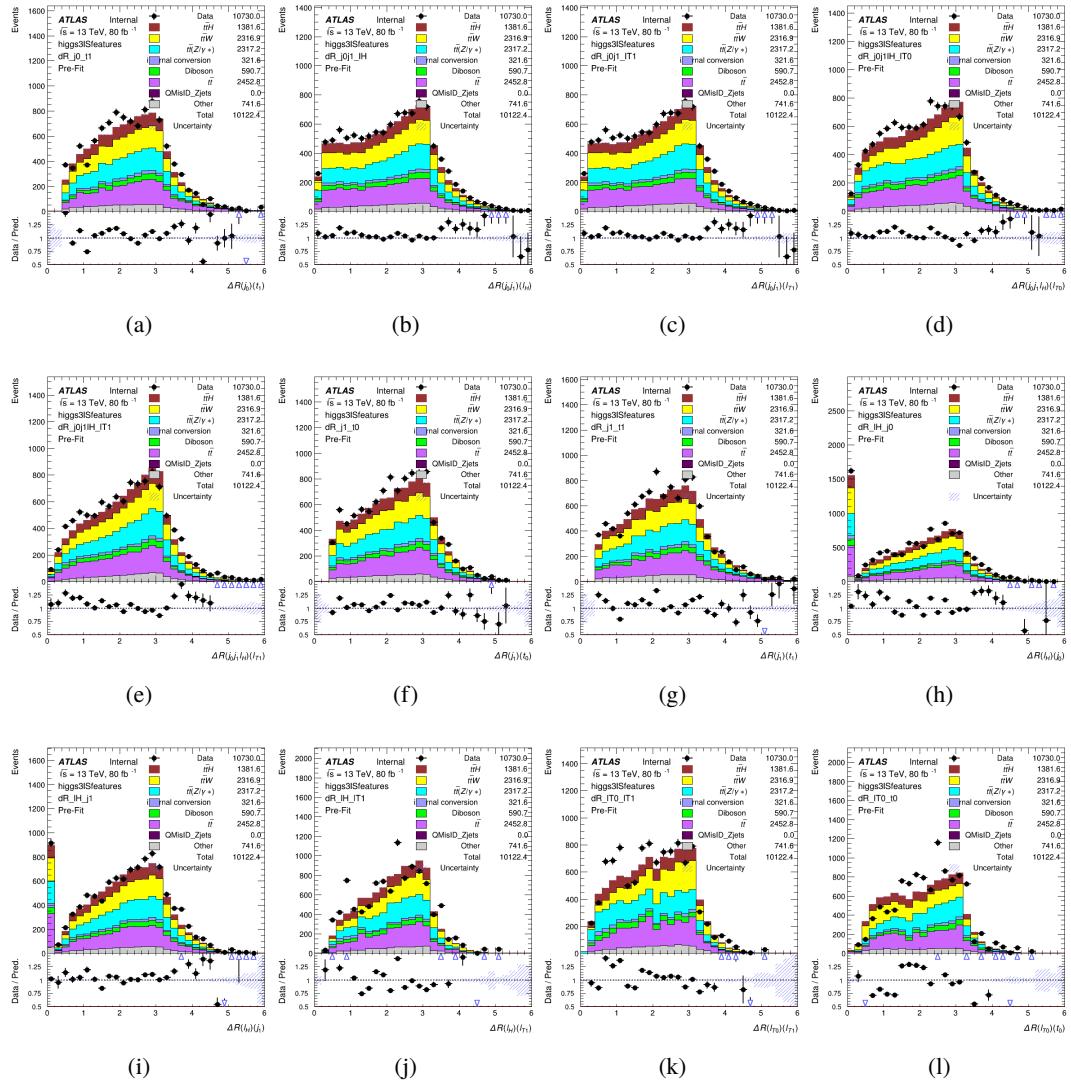


Figure A.13: Input features for higgs31S

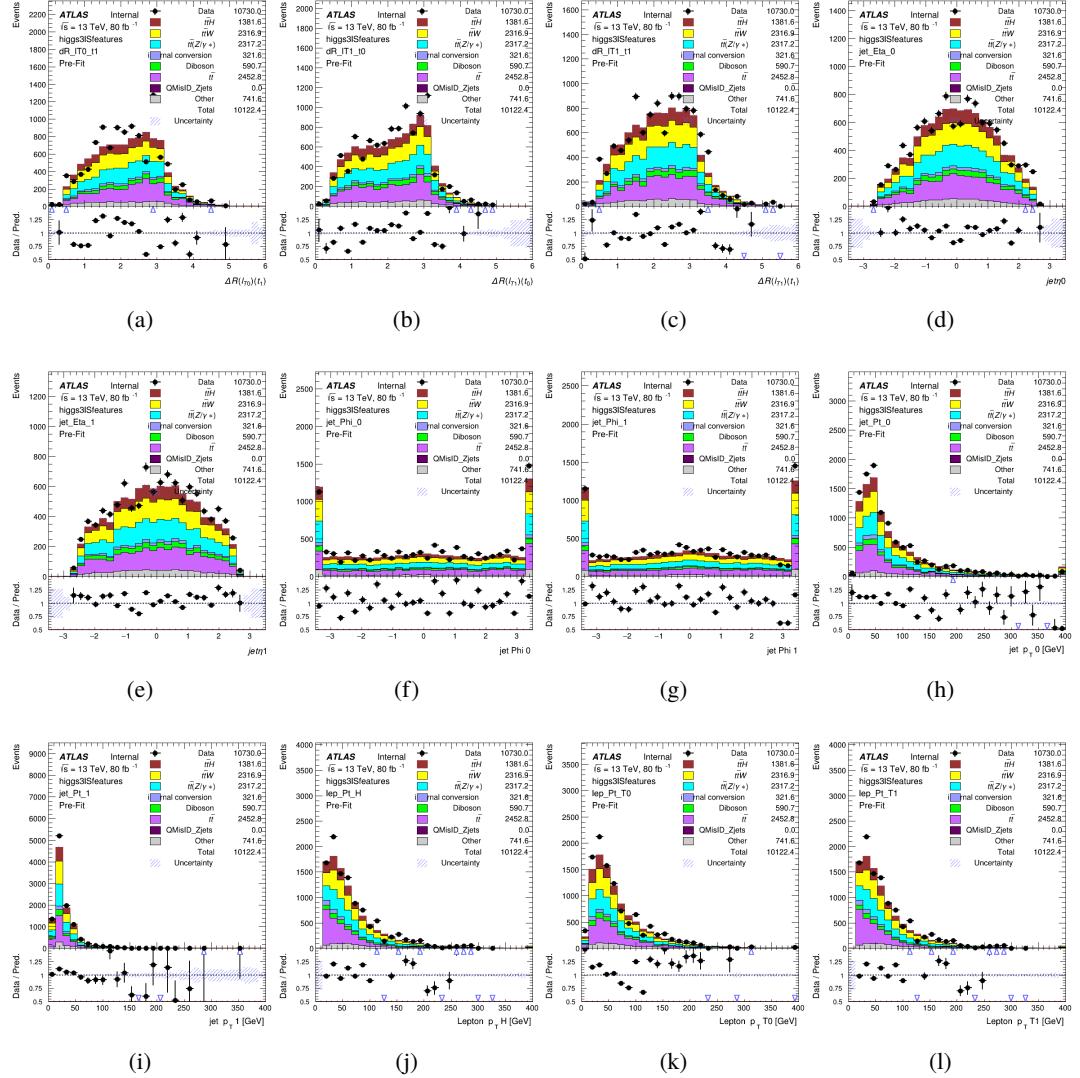


Figure A.14: Input features for higgs3lS

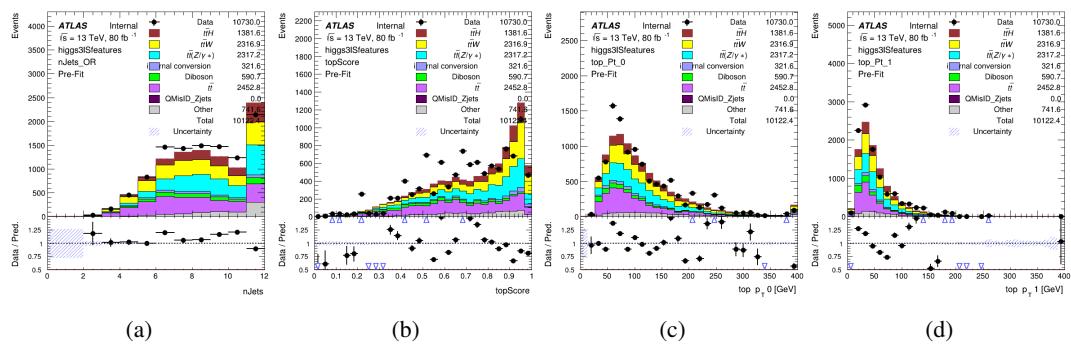


Figure A.15: Input features for higgs31S

<sup>479</sup> **A.1.5 Higgs Reconstruction Features - 3lF**

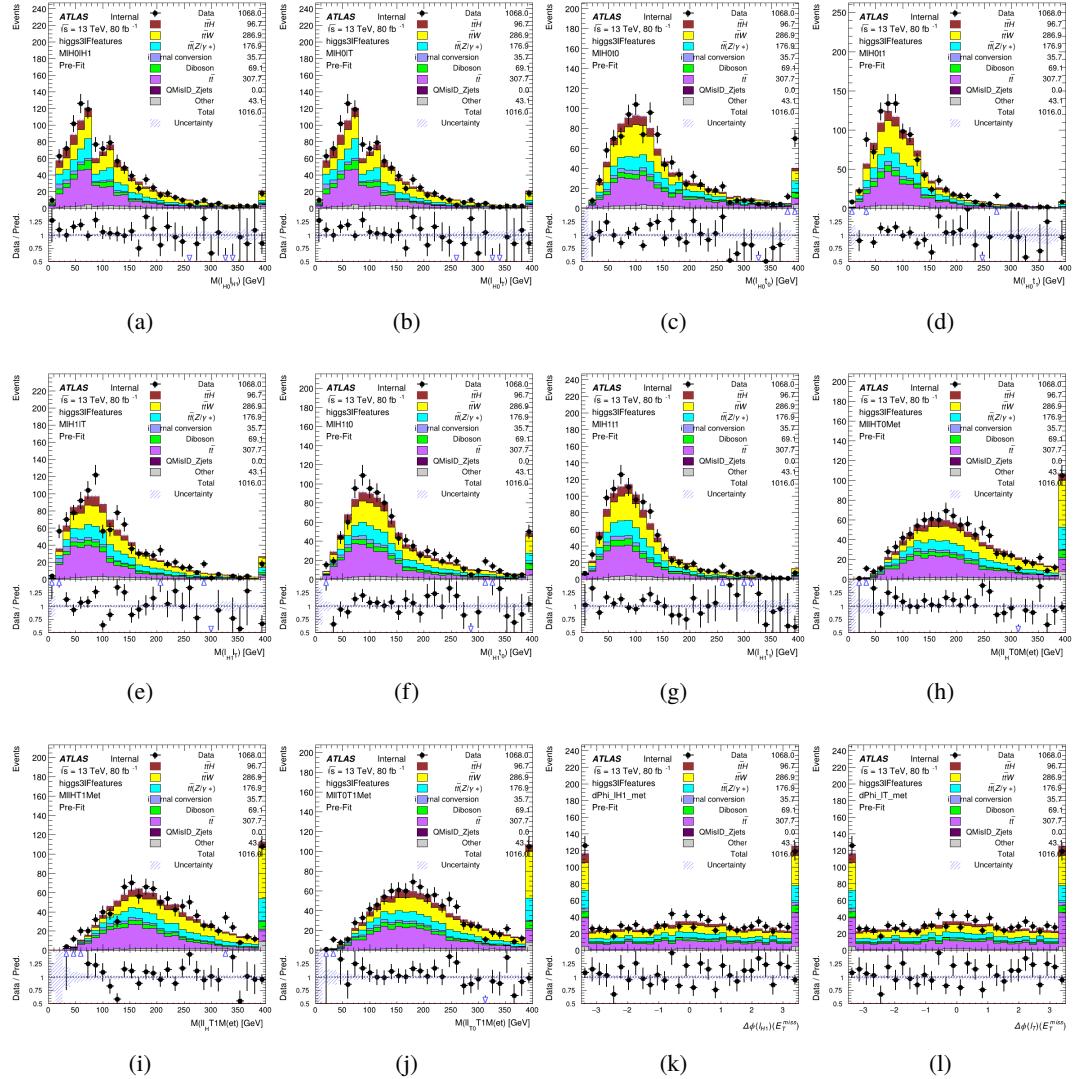


Figure A.16: Input features for higgs3IF

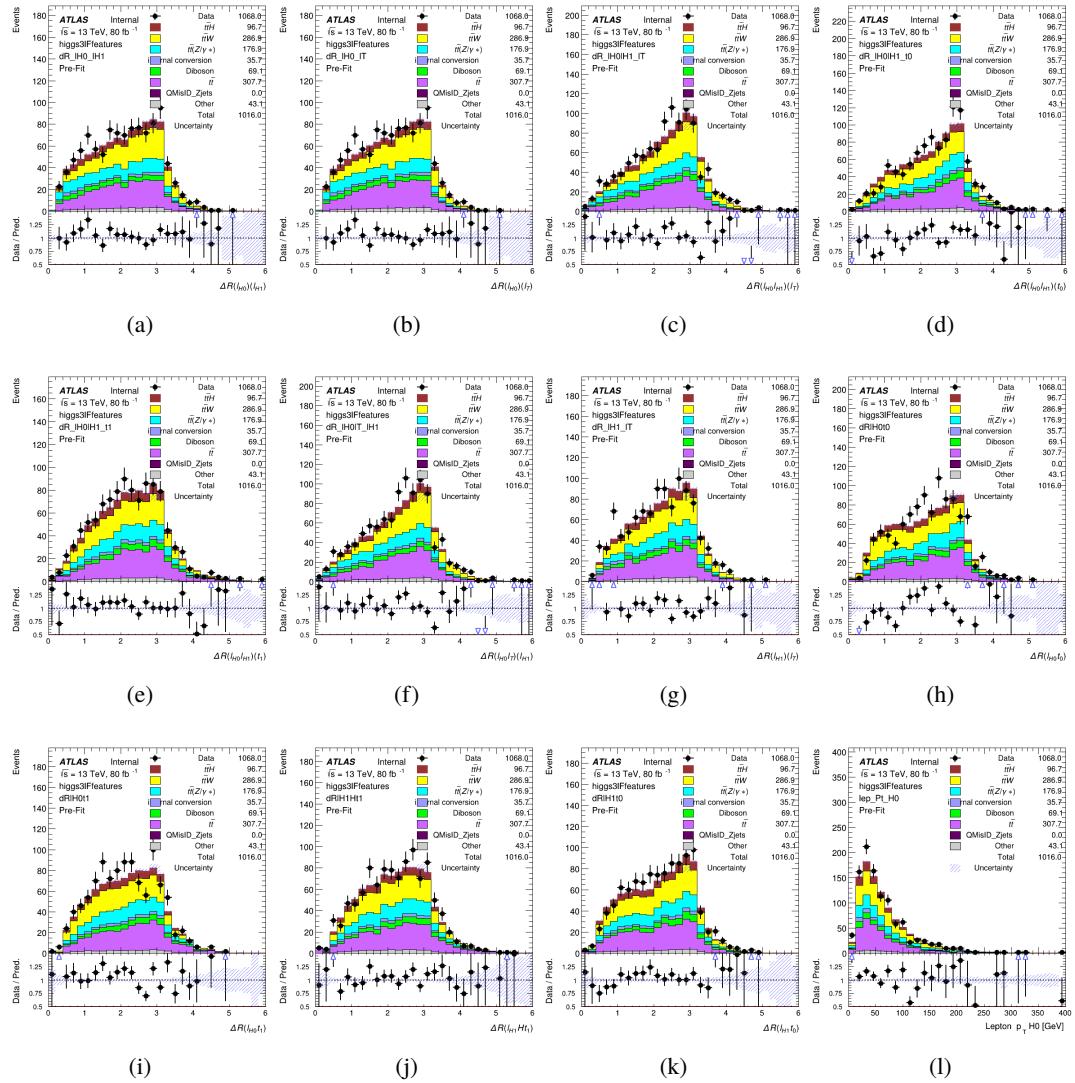


Figure A.17: Input features for higgs3lF

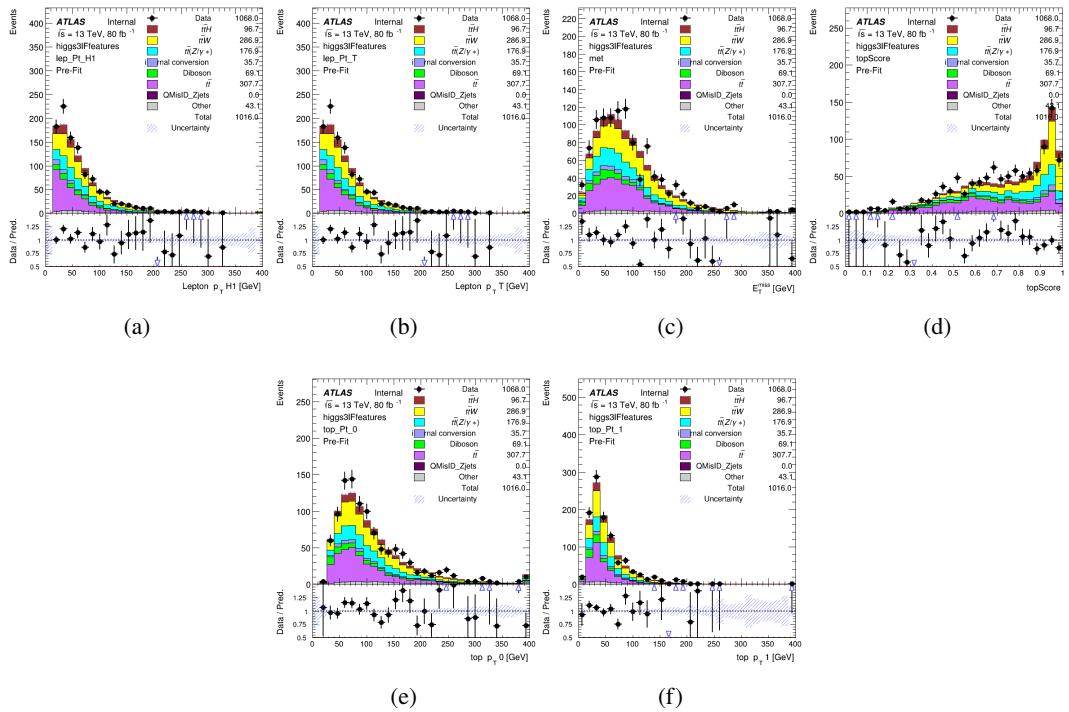


Figure A.18: Input features for higgs3IF

<sup>480</sup> **A.2 Background Rejection MVAs**

<sup>481</sup> **A.2.1 Background Rejection MVA Features - 2lSS**

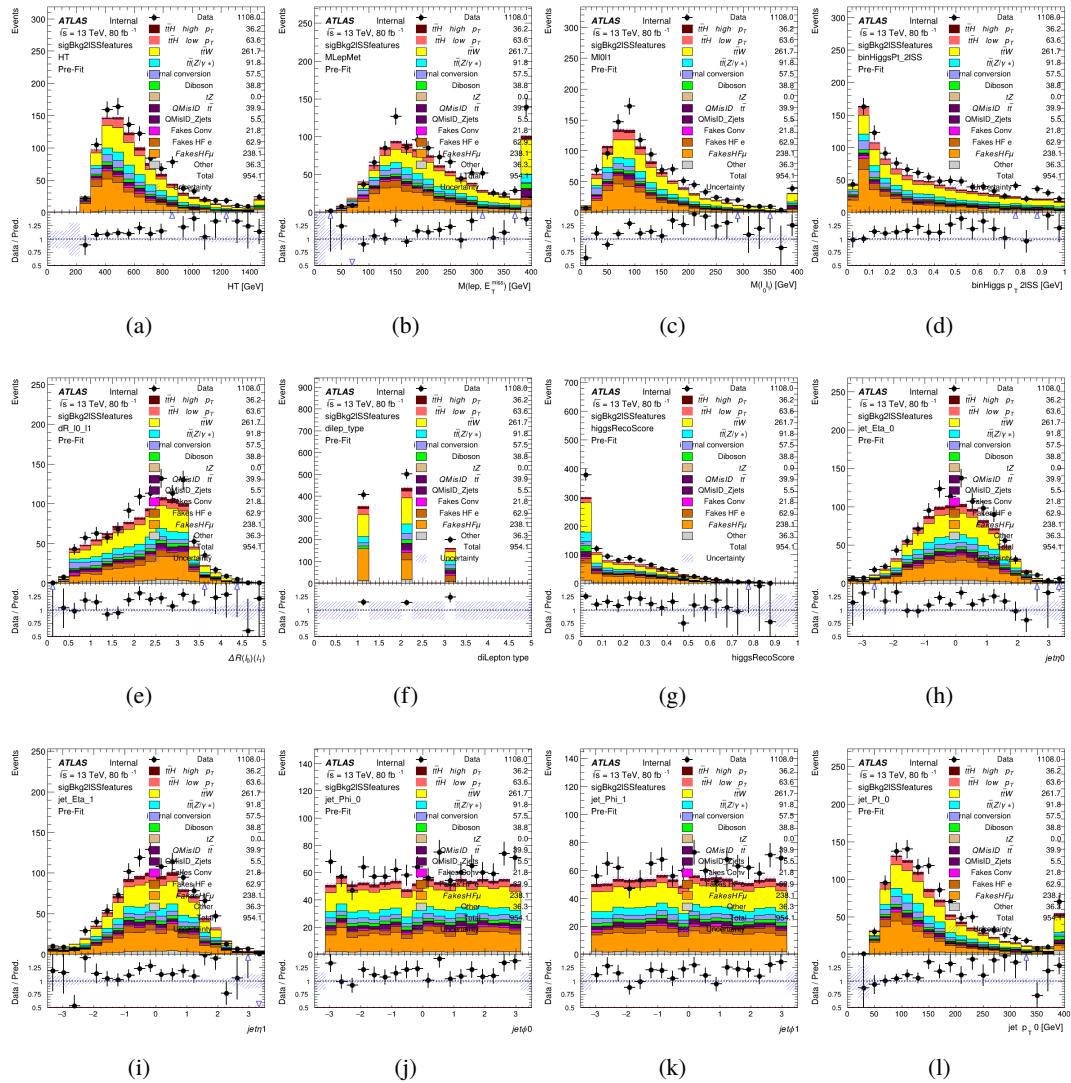


Figure A.19: Input features for sigBkg2lSS

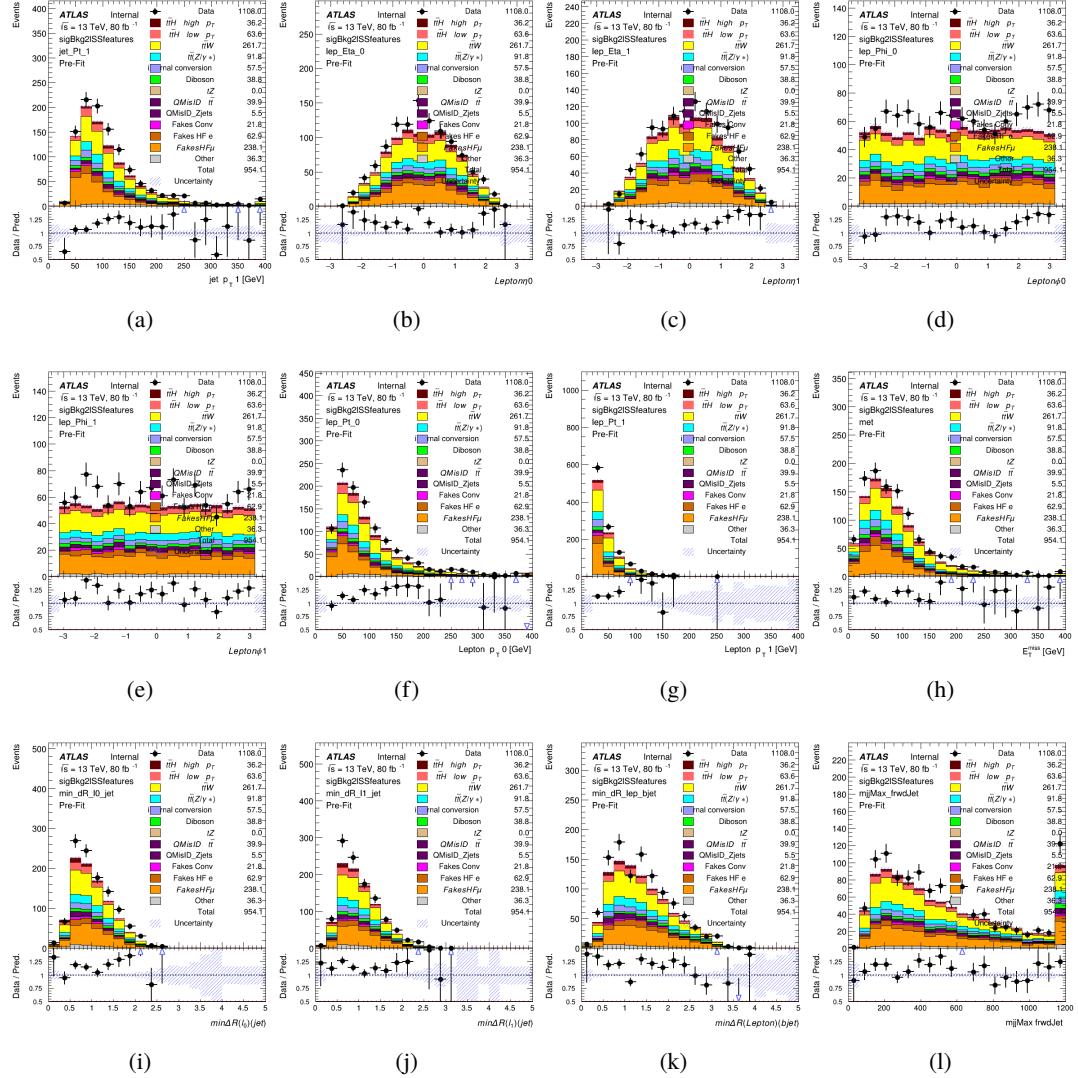


Figure A.20: Input features for sigBkg2lSS

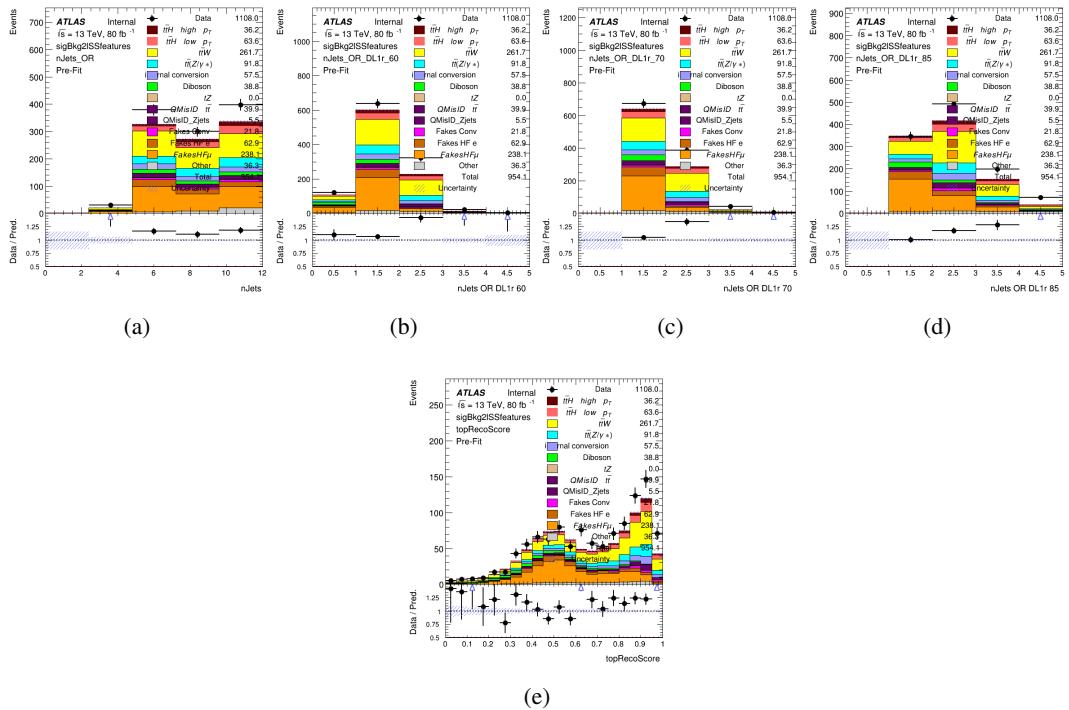
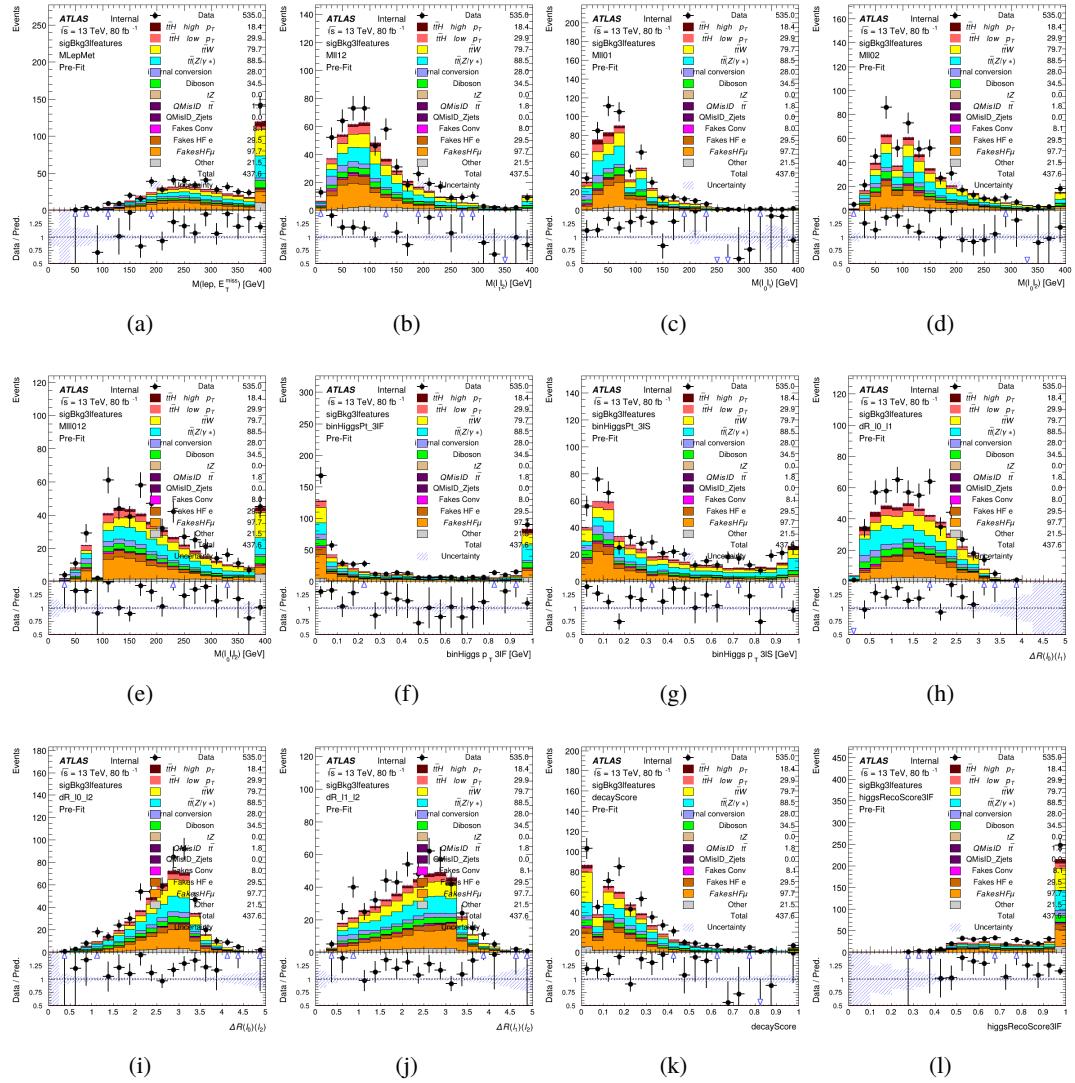


Figure A.21: Input features for sigBkg2ISS

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

Figure A.22: Input features for `sigBkg3l`

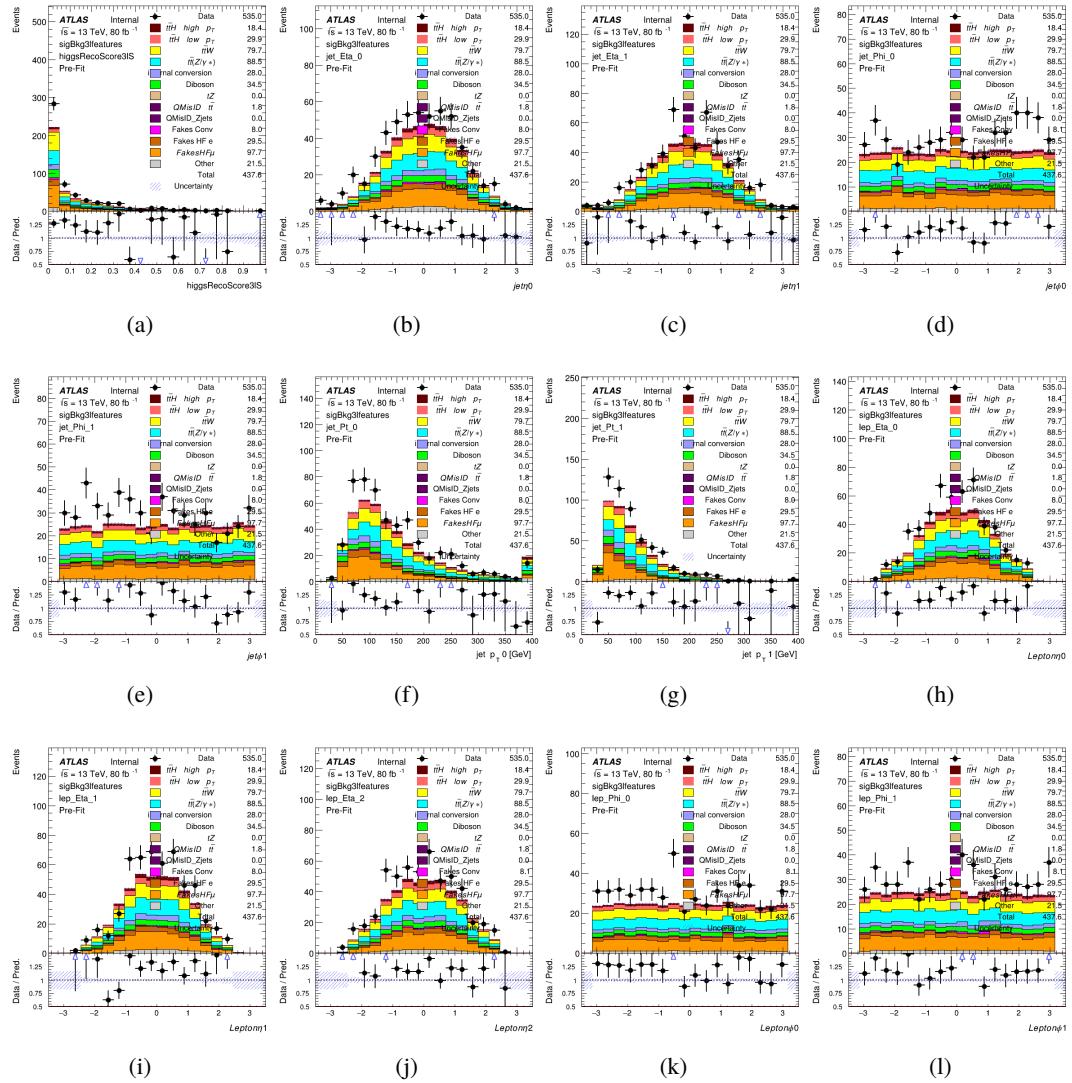


Figure A.23: Input features for sigBkg3l

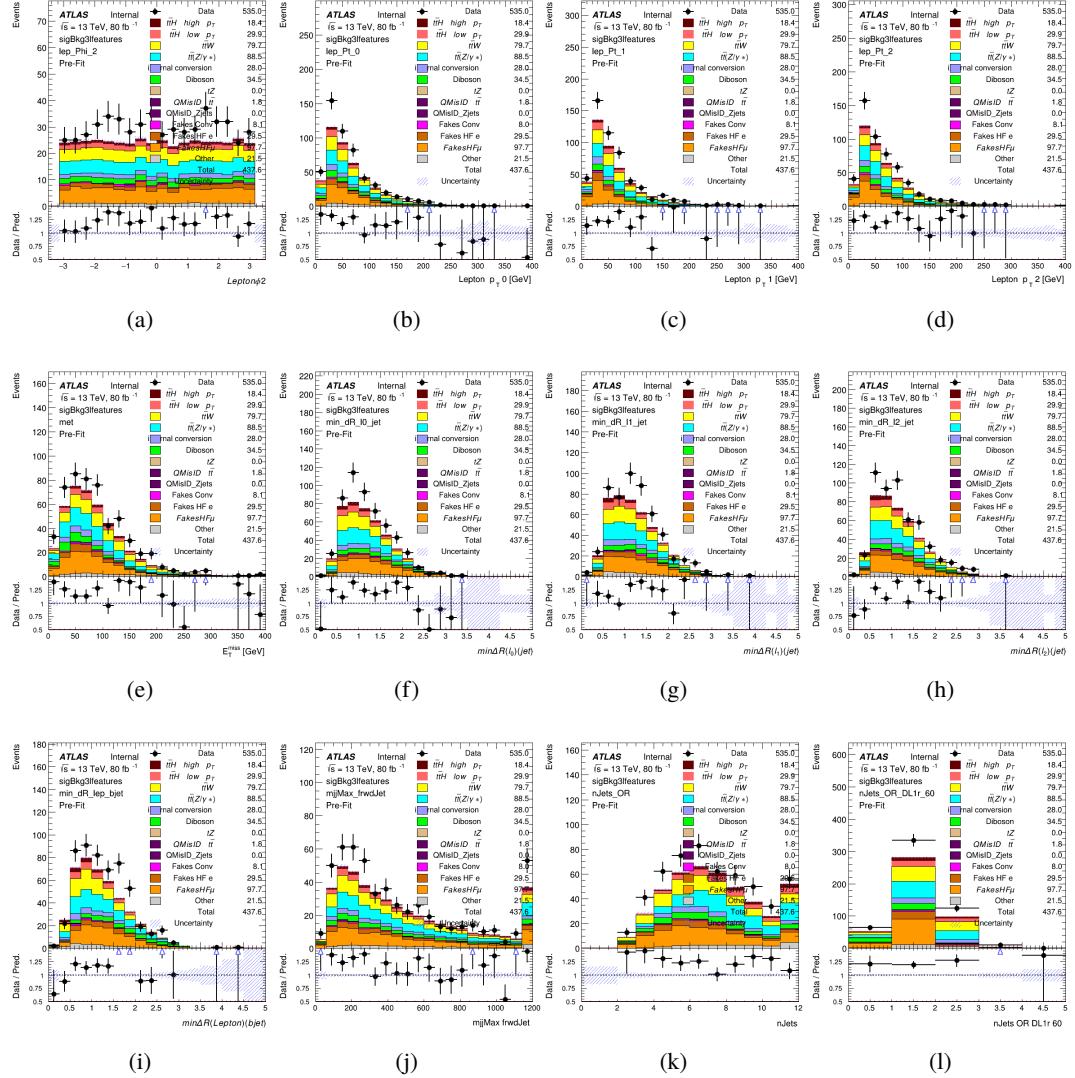


Figure A.24: Input features for sigBkg3l

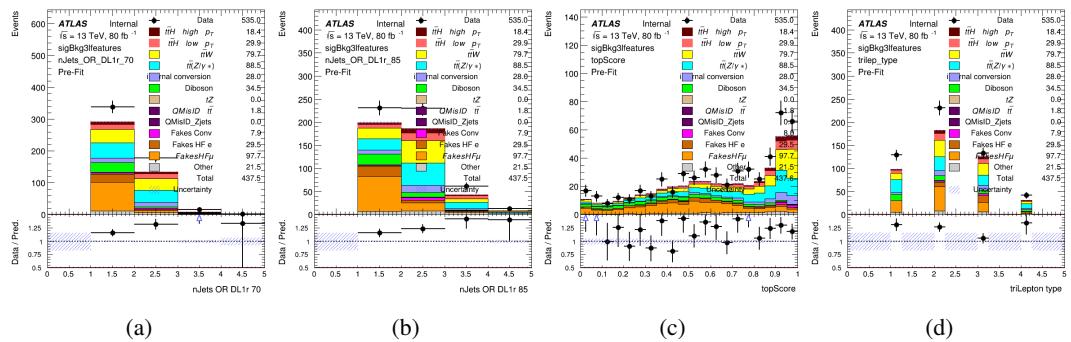


Figure A.25: Input features for sigBkg3l

---

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

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

491 For these studies, the kinematics of the 10 highest  $p_T$  jets in each event are used for training. This  
492 includes the vast majority of truth b-jets. Specifically the  $p_T$ ,  $\eta$ ,  $\phi$ ,  $E$ , and DL1r score of each jet  
493 are used. For events with fewer than 10 jets, these values are substituted with 0. The  $p_T$ ,  $\eta$ ,  $\phi$ ,  
494 and  $E$  of the leptons and  $E_T^{\text{miss}}$  are included as well. Categorical cross entropy is used as the loss  
495 function.

Table 17: 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%

---

496 **A.4 Binary Classification of the Higgs  $p_T$** 

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

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

---

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

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

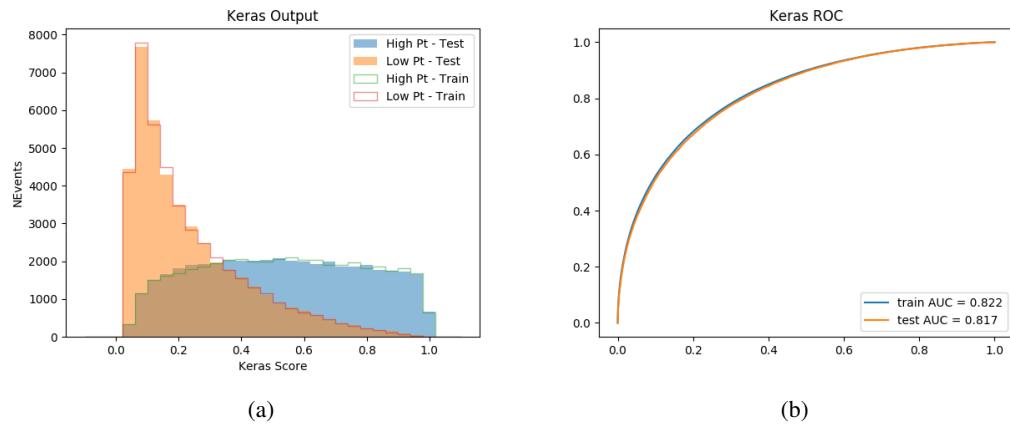


Figure A.26:

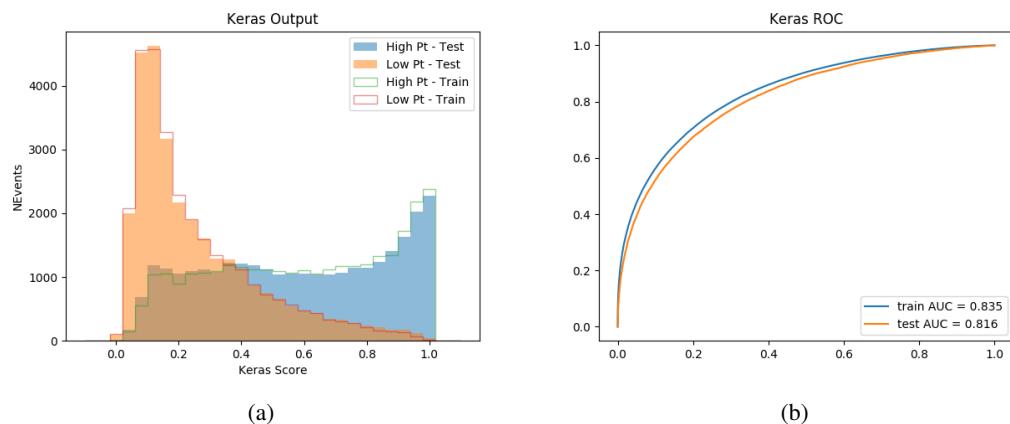


Figure A.27:

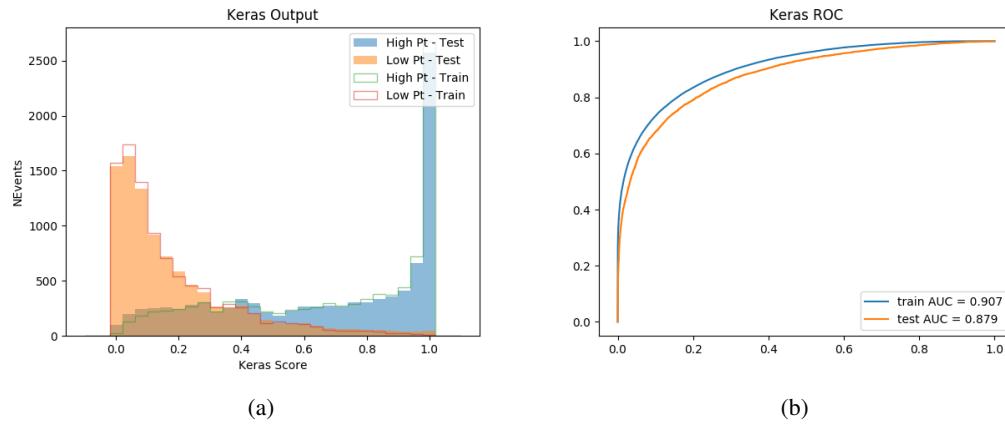


Figure A.28:

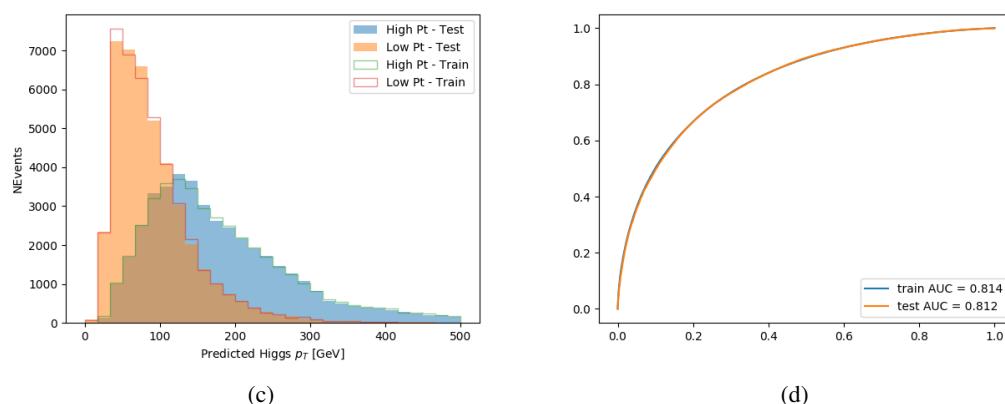
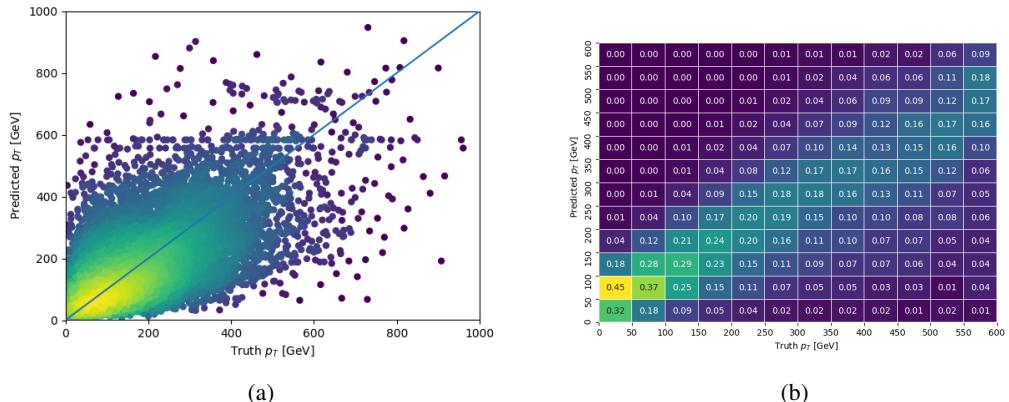


Figure A.29: