ATLAS Note

GROUP-2017-XX

10th November 2020

Draft version 0.1

A Deep Learning Approach to Differential Measurements of Higgs - Top Interactions in Multilepton Final States

The ATLAS Collaboration

Several theories Beyond the Standard Model modify the momentum spectrum of the Higgs Boson, without significantly altering the overall rate of Higgs produced from top quark pairs. A differential measurement of the Higgs transverse momentum provides a way to search for these effects at the LHC. Because of the challenges inherent to reconstructing the Higgs in multilepton final states, a deep learning approach is used to predict the momentum spectrum of the Higgs for events where the Higgs Boson is produced from top quark pairs and decays to final states that include multiple leptons. The regressed Higgs p_T is fit to data for events with two or three leptons in the final state, and estimates of the sensitivy to variations in the Higgs p_T spectrum are given.

Contents

Part I

Introduction

1 Introduction

Particle physics is an attempt to describe the fundamental building blocks of the universe and their interactions. The Standard Model (SM) - our best current theory of fundamental particle physics - does a remarkable job of that. All known fundamental particles and (almost) all of the forces underlying their interactions can be explained by the SM, and the predictions from this theory agree with experiment to an incredibly precise degree. This is especially true since the Higgs Boson, the last piece of the SM predicted decades before, was finally discovered at the Large Hadron Collider (LHC) in 2012.

Despite the success of the SM, there remains significant work to be done. For one, the SM is incomplete: it fails to provide a description of gravity, to give an explanation for the observation of Dark Matter, or to provide a mechanism for neutrinos to gain mass. Further, a Higgs Boson with a mass of around 125 GeV, as observed at the LHC, gives rise to what is known a hierarchy problem - such a low mass Higgs requires a seemingly unnatural level of “fine tuning” that is unexplained by the SM.

A promising avenue for addressing these problems is to study the properties of the Higgs Boson and the way it interacts with other particles, in part simply because these interactions have not been measured before. Its interactions with the Top Quark are a particularly promising place to look. Because the Higgs Field is responsible for allowing particle to acquire mass, the strength of a particle's interaction with the Higgs Boson is proportional to its mass. As the most massive of the fundamental particles, the Top Quark has the strongest coupling to the Higgs Boson, meaning any new physics in the Higgs sector is likely to present itself most prominantly in its interaction with the Top Quark.

These interactions can be measured by directly by studying the production of a Higgs Boson in association with a pair of Top Quarks ($t\bar{t}H$). While studies have been done measuring the overall rate of $t\bar{t}H$ production, there are several theories of physics Beyond the Standard Model (BSM) that would affect the kinematics of $t\bar{t}H$ production without altering its overall rate. This dissertation attempts to make a differential measurement of the kinematics of the Higgs Boson in $t\bar{t}H$ events in order to search for these BSM effects.

An Effective Field Theory model can be used to model the low energy effects of high energy physics.

The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-2018 provides the oppurtunity to make this measurement for the first time. The unprecedented energy acheived by the LHC during this period greatly increase the rate at which $t\bar{t}H$ events are produced, and the large amount of data collected provides the necessary statistics for a differential measurement to be performed.

A study of $t\bar{t}H$ events with multiple leptons in the final state is performed, using 139 fb^{-1}

of data from proton-proton collisions at an energy $\sqrt{s} = 13$ TeV collected by the ATLAS detector from 2015-2018. Events are separated into channels based on the number of light leptons in the final state - either two same-sign leptons, or three leptons. A deep neural network is used to reconstruct the momentum of the Higgs Boson in each event. This momentum spectrum is fit to data for each analysis channel, the result of which is used to place limits on BSM effects.

This dissertation begins with a brief explanation of the SM, its limitations, and the theoretical motivation behind this work. This is followed by a description of the LHC and the ATLAS detector. The analysis strategy is then described, and the results are presented. Finally, the results of the study are summarized in the conclusion.

Part II

Theoretical Motivation

2 The Standard Model and the Higgs Boson

The Standard Model of particle physics (SM) is a Quantum Field Theory (QFT) describing the known fundamental particles and their interactions. It accounts for three of the four known fundamental force - electromagnetism, the weak nuclear force, and the strong nuclear force, but not gravity. Further, the SM describes a mechanism for combining the weak and electromagnetic forces into a singular interaction, known as the electroweak force. It is a non-Abelian gauge theory, invariant under the Lie Group $SU(3)_C \bigotimes SU(2)_L \bigotimes U(1)_Y$, where C refers to color charge, L, the helicity of the particle, and Y, the hypercharge.

2.1 The Forces and Particles of the Standard Model

The SM particles, summarized in figure 2.1, can be classified into two general categories based on their spin: fermions, and bosons.

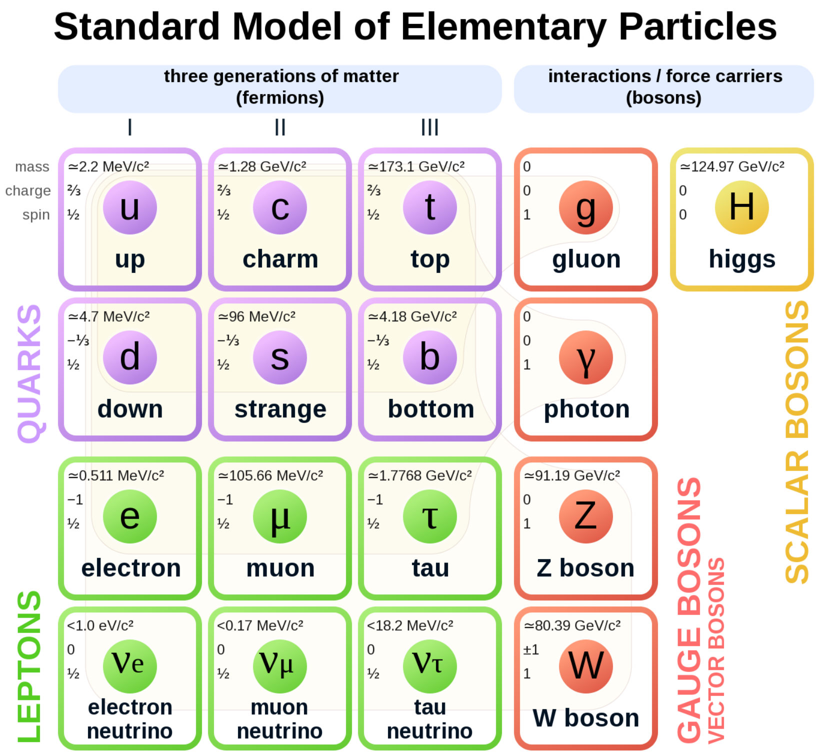


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. []

Fermions are particles with $\frac{1}{2}$-integer spin, which according to the spin-statistics theorem, causes them to comply with the Pauli-exclusion principle []. They can be separated into two groups, leptons and quarks, each of which consist of three generations of particles with increasing mass.

Leptons are fermions interact via the electroweak force, but not the strong force. The three generation of leptons consist of the electron and electron neutrino, the muon and muon neutrino, the tau and tau neutrino. The quarks, which do interact via the strong force - which is to say they have color charge - in addition to the electroweak force. The three generations include the up and down quarks, the strange and charm quarks, and the top and bottom quarks. Each of these generations form left-handed doublets invariant under $SU(2)$ transformations. For the leptons these doublets are:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}_L, \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}_L, \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}_L \tag{2.1}$$

And for the quarks:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} s \\ c \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L \tag{2.2}$$

For both leptons and quarks, the heavier generations can decay into the lighter generation of particles, while the first generation does not decay. Hence, ordinary matter generally consists of this first generation of fermions - electrons, up quarks, and down quarks. Each of these fermions has a corresponding anti-particle, which has an equal mass as its partner but oppsite charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos, whose mass has been experimentally confirmed but is not accounted for in the SM.

Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1 particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0, particle - the Higgs Boson. Of the gauge bosons, the W^{+}, W^{-} and Z bosons - which are the mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon mediates the electric force, and the gluon mediates the strong force.

2.2 The Higgs Mechanism

A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to the Lagrangian giving mass to the the gauge bosons would violate the underlying symmetry of the theory. This presents a clear problem with the theory: The experimental observation that the W and Z bosons have mass seems to contradict the basic structure of the SM.

Rather than abandoning gauge invariance, an alternative way for particles to acquire mass beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout in 1964 []. This procedure for introducing masses for the gauge bosons while preserving local gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory by Weinberg in 1967 [].

2.2.1 The Higgs Field

The Higgs mechanism introduces a complex scalar $SU(2)$ doublet, Φ, with the form:

$$\Phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix}_L \tag{2.3}$$

This field introduces a scalar potential to the Lagrangian of the form:

$$V(\Phi) = \mu^2 |\Phi^\dagger \Phi| + \lambda(|\Phi^\dagger \Phi|)^2 \tag{2.4}$$

Where μ and λ are free parameters of the new field. This represents the most general potential allowed while preversing $SU(2)_L$ invariance and renormalizability. In the case that $\mu^2 < 0$, this potential takes the form shown in figure 2.2.

The significant feature of this potential is that its minimum does not occur for a value of $\Phi = 0$. Instead, it is minimized when $|\Phi^\dagger \Phi| = -\mu^2/\lambda$. This means that in its ground state, the Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

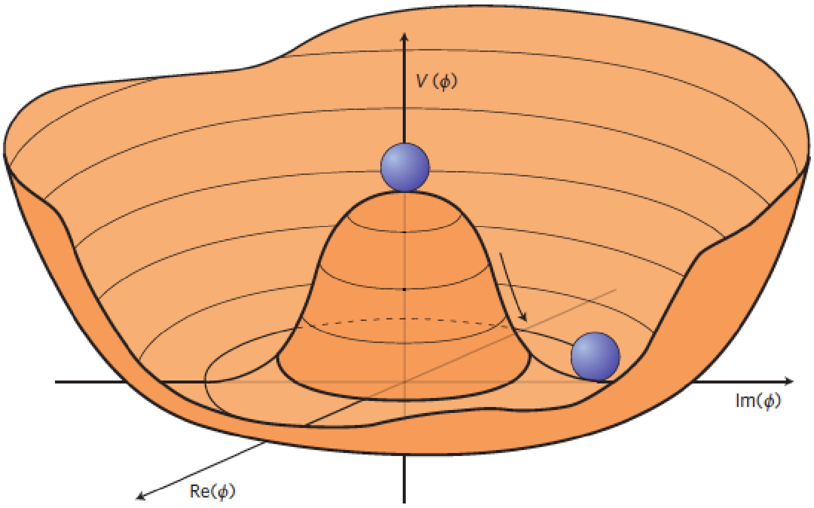


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ, for the case that $\mu^2 < 0$ [].

the minimum is determined only by $\Phi^\dagger\Phi$, there is some ambiguity in the particular definition of the VEV, but it is generally represented as

$$\langle\Phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu \end{pmatrix} \tag{2.5}$$

The full value of Φ can be written as

$$\langle\Phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu + H/\sqrt{2} \end{pmatrix} \tag{2.6}$$

with ν being the value of the VEV, and H being the real value of the scalar field.

2.2.2 Electroweak Symmetry Breaking

The Electroweak (EWK) interaction is described in the SM by a $SU(2)_L \bigotimes U(1)_Y$ gauge theory. This theory predicts three $SU(2)_L$ gauge boson, W^1_μ, W^2_μ, W^3_μ, and a single $U(1)_Y$ gauge boson, B_μ. The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar field Φ in the Lagrangian:

$$(D_\mu \Phi)^\dagger (D^\mu \Phi) = |(\partial_\mu - \frac{ig}{2} W^a_\mu \sigma^a - \frac{ig'}{2} B_\mu Y)\phi|^2 \tag{2.7}$$

Here D_μ represents the covariant derivative required to preserve gauge invariance, g and g' represent coupling constant of the gauge bosons, σ^a denotes the Pauli matrices of $SU(2)$, and Y represents the hypercharge of $U(1)$. The terms in this interaction which contribute to the masses of the gauge bosons can be written as:

$$\frac{1}{2}(0, \nu)(\frac{g}{2} W^a_\mu \sigma^a - \frac{g'}{2} B_\mu)^2 \begin{pmatrix} 0 \\ \nu \end{pmatrix} \tag{2.8}$$

Expanding these terms into the mass eigenstates of the electroweak interaction yields four physical gauge bosons, two charged and two neutral, which are linear combinations of the fields

W^1_μ, W^2_μ, W^3_μ, and B_μ:

$$
\begin{aligned}
W^{\pm}_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu \pm iW^2_\mu) \\
Z^\mu &= \frac{1}{\sqrt{(g^2+g'^2)}}(-g'B_\mu + gW^3_\mu) \\
A^\mu &= \frac{1}{\sqrt{(g^2+g'^2)}}(gB_\mu + g'W^3_\mu)
\end{aligned}
\tag{2.9}
$$

And the masses of these fields are given by:

$$
\begin{aligned}
M^2_W &= \frac{1}{4}g^2v^2 \\
M^2_Z &= \frac{1}{4}(g^2+g'^2)v^2 \\
M^2_A &= 0
\end{aligned}
\tag{2.10}
$$

This produces exactly the particles we observe - three massive gauge bosons and a single massless photon. The massless photon represents the portion of the gauge symmetry, a single $U(1)$ of the electromagnetic force, that remains unbroken by the VEV.

Interactions with the Higgs field also lead to the generation of the fermion masses, which in the Lagrangian take the form:

$$
-\lambda_\psi(\bar{\psi}_L\phi\psi_R + \bar{\psi}_R\phi^\dagger\psi_L)
\tag{2.11}
$$

After symmetry breaking has occured and ϕ has taken on the value of the VEV as written in equation 2.5, the mass terms or the fermions become $\lambda_\psi v$. Written this way, the fermion masses are proportional to their Yukawa coupling to the VEV, λ_ψ.

Based on the equation 2.6, an additional mass term, $\mu^2 H^2$ arises from the potential $V(\Phi)$. This term can be understood as anexcitation of the Higgs field, a scalar boson with mass $M_H = \mu$. This is the Higgs boson, which comes about as a natural prediction of electroweak symmetry breaking.

The fermion's Yakawa coupling to the VEV take the same form as the fermion's coupling to the Higgs boson - λ_ψ. Therefore, the strength of a fermion's interaction with the Higgs is directly proportional to its mass. We now have a model that predicts a Higgs boson with mass $M_H = \mu$, which interacts with the fermions with coupling strength λ_ψ. Because μ and λ_ψ are free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions must be measured experimentally.

2.3 Limitations of the Standard Model

While the SM has great predictive power, there are still several experimental observation that the SM fails to explain. For example, the SM predicts neutrinos to be massless, despite experimental observation to the contrary.

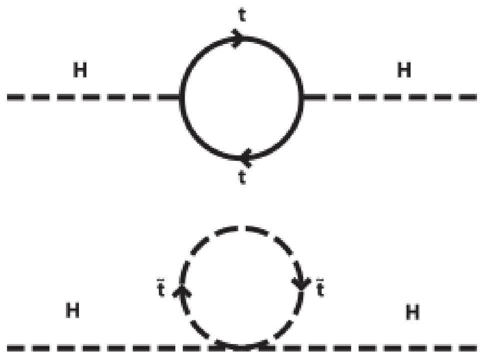


Figure 2.3: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram [].

3 Effective Field Theory in $t\bar{t}H$ Production

Higher dimension operators are a common way to paramaterize the effects of physics at very high energies into

3.1 Extensions to the Higgs Sector

3.2 Six Dimensional Operators

While the SM has been tested to great precision, particularly at the LHC, it is generally accepted that it is only valid up to a certain energy scale. It is assumed that above a certain energy, at the scale where something like a Grand Unified Theory (GUT) or quantum gravity become relevant, the SM will not be applicable.

Part III

The LHC and the ATLAS Detector

4 The LHC

The Large Hadron Collider (LHC) is a particle accelerator consisting of a 27 km ring, designed to collide protons at high energy. Located outside of Geneva, Switzerland and buried about 100 m undergound, it consists of a ring of superconducting magnets which are used to accelerate opposing beams of protons - or lead ions - which collide at the center of one of the various detectors located around the LHC ring which record the result of these collisions. These detectors include two general purpose detectors, ATLAS and CMS, which are designed to make precision measurements of a broad range of physics phenomenon, and two more specialized experiments, LHCb and ALICE, which are optimized to study b-quarks and heavy-ion physics, respectively.

The LHC first began running in 2009 at a proton-proton center of mass energy of $\sqrt{s} = 8$ TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during this period was used in discovering the Higgs Boson. The LHC began running again in 2015, and collected data at an increased energy of $\sqrt{s} = 13$ TeV until 2018, a period referred to as Run 2.

The LHC consists of a chain of accelerators, which accelerate the protons to higher and higher energies until they are injected into the main ring. This process is summarized in figure 4.1. Protons extracted from a tank of ionized hydrogen are fed into a linear accelerator, LINAC2, where they reach an energy of 50 MeV. From there, they enter a series of three separate circular accelerators, before being injected into the main accelerator ring at an energy of 450 GeV. Within the main ring protons are seperated into two seperate beams moving in opposite directions, and their energy is increased to their full collision energy. Radiofrequency cavities are used to accelerate these particles are sort them into bunches. From 2015-2018, these bunches consisted of around 100 billion protons each with an energy of 6.5 TeV per proton, which collided at a rate of 40 MHz, or every 25 ns.

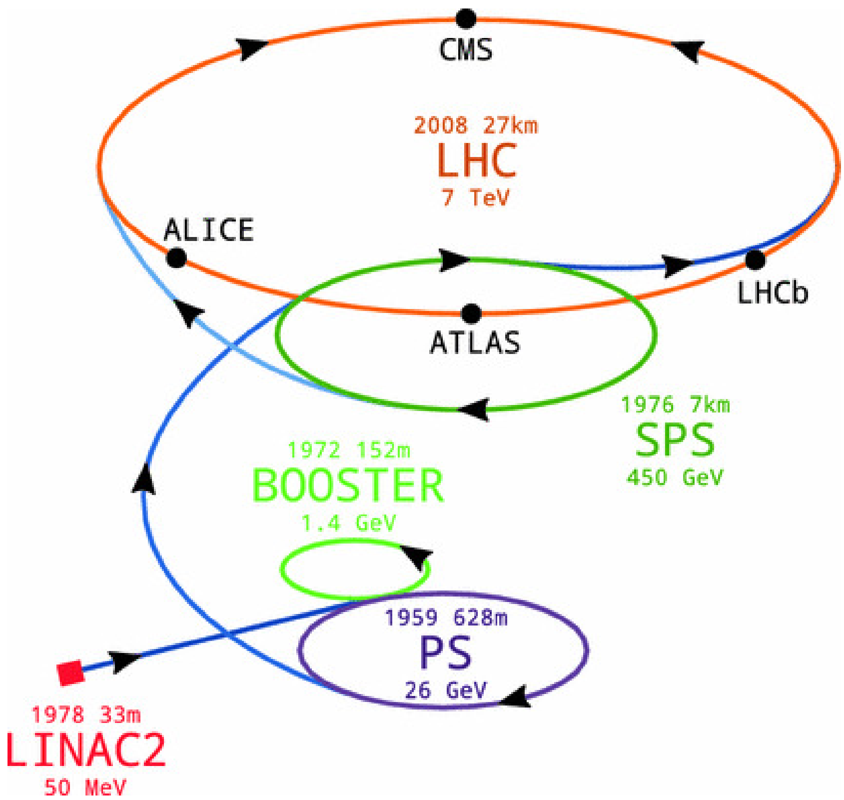


Figure 4.1: A summary of the accelerator chain used to feed protons into the LHC [].

Because these proton bunches consist of a large number of particles, each bunch crossing consists of not just one, but several direct proton-proton collisions. The number of interactions that occur per bunch crossing, μ, is known as pileup. During Run 2, the average pileup for bunch crossings was around $\langle\mu\rangle = 35$, with values typically ranging between 10 and 70.

The amount of data collected by the LHC is measured in terms of luminosity, which is the ratio of the number of events detected per unit time, $\frac{dN}{dt}$, and the interaction cross-section, σ.

$$\mathcal{L} = \frac{1}{\sigma}\frac{dN}{dt} \tag{4.1}$$

The design luminosity of the LHC is $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, however the LHC has acheived a luminosity of over $2 \times 10^34\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The total luminosity is then this instantaneous luminosity integrated over time.

$$\mathcal{L}_{\mathrm{int}} = \int \mathcal{L}dt \tag{4.2}$$

The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around 140 fb^{-1}, exceeding the expected integrated luminosity of 100 fb^{-1}.

5 The ATLAS Detector

ATLAS (a not terribly natural acronym for "A Toroidal LHC Apparatus") is a general purpose detector designed to maximize the detection efficiency of all physics objects, including leptons, jets, and photons. This means it is capable of measuring all SM particles, with the exception of neutrinos, the presence of which can be inferred based on missing transverse momentum. The detector measures 44 m long, and 25 m tall.

The ATLAS detector consists of multiple layers, each of which serves a different purpose in reconstructing collisions. At the very center of the detector is the interaction point where the proton beams of the LHC collide.

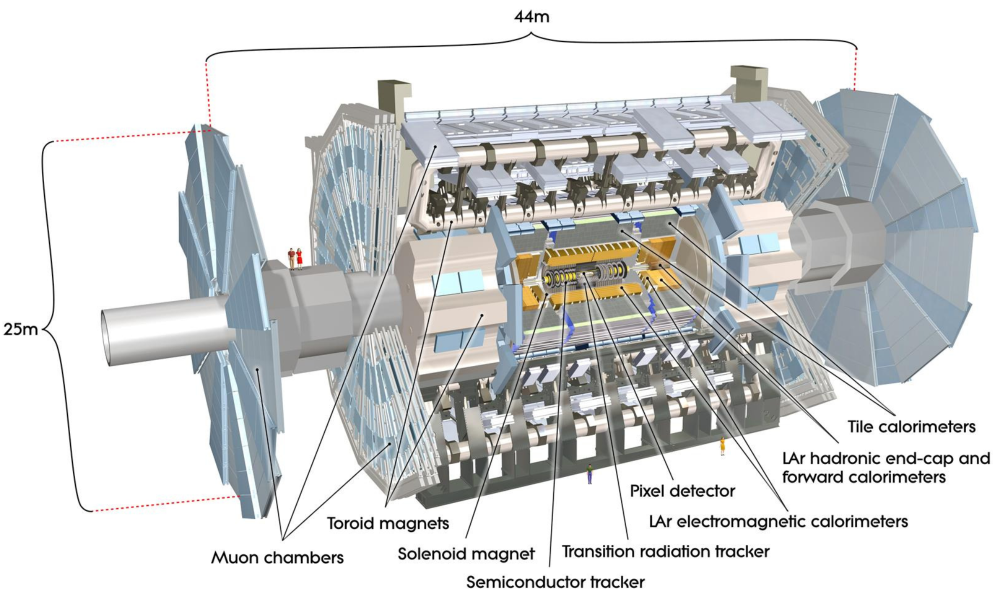


Figure 5.1: Cutaway view of the ATLAS detector, with labels of its major components [].

5.1 Inner Detector

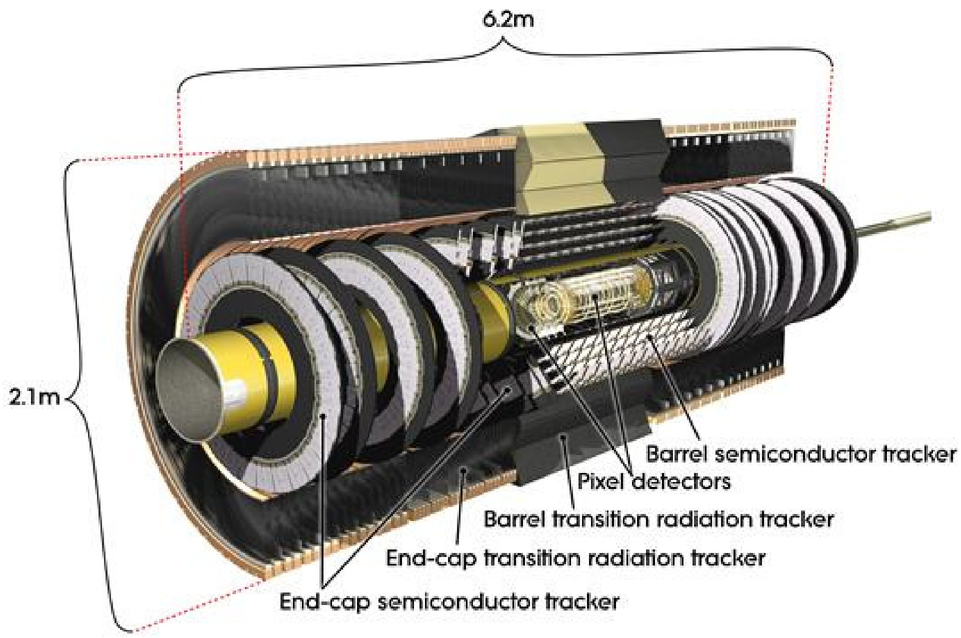


Figure 5.2: Cutaway view of the Inner Detector [].

Just surrounding the interaction point is the Inner Detector, designed to track the path of charged particles moving through the detector. An inner solenoid surrounding the Innder Detector is used to produces a magnetic field of 2 T. This large magnetic field causes the path of charged particles moving through the Inner Detector to bend. Because this magnetic field is uniform and well known, it can be used in conjunction with the curvature of a particles path to measure its charge and momentum.

The Inner Detector consists of three components - the Pixel Detector, the Semi-Conductor Tracker (SCT), and the Transition Radiation Tracker (TRT). The Pixel Detector is the innermost of these, beginning just 33.25 mm away from the beam line. It consists of three silicon layers along the barrel, as well as three endcap layers, covering a range of $|\eta| < 2.5$.

The Semiconductor Tracker (SCT) is similar to the Pixel detector, but uses long strips rather than small pixel to cover a larger spatial area.

5.2 Calorimeters

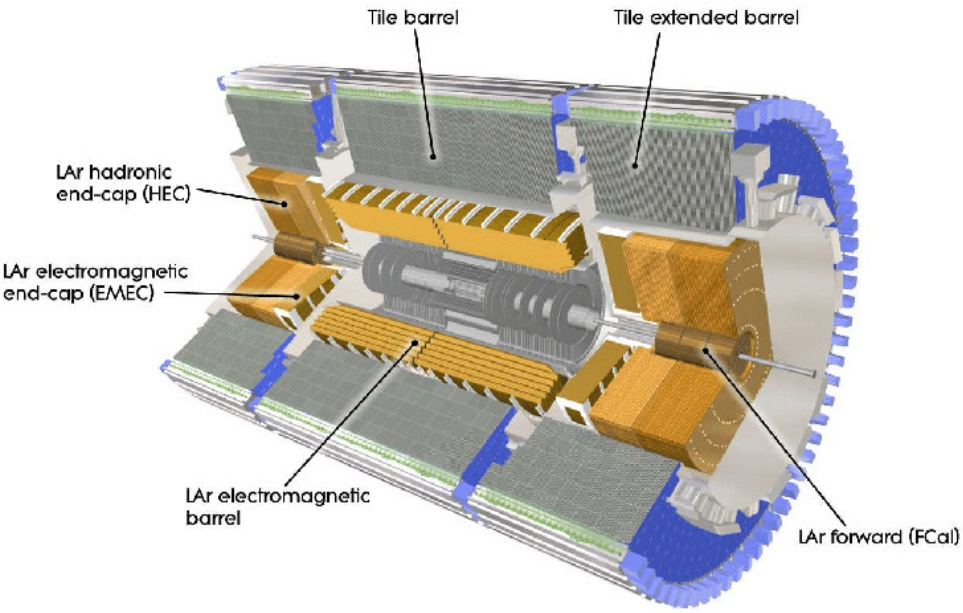


Figure 5.3: Cutaway view of the calorimeter system of the ATLAS detector [].

Situated outside the Innder Detector are two concentric calorimeters. The inner calorimeter uses liquid argon (LAr) to measure energy of particles the interact electromagnetically, which includes photons and any charged particle. The LAr calorimeter is made of heavy metals, primarily lead and copper, which causes electromagnetically interacting particles to shower, depositing their energy in the detector. The showering of the high energy particles that pass through calorimeter cause the liquid argon to ionize, and the ionized electrons are detected by electronic readouts. The LAr calorimeter consists of around 180,000 readout channels.

The outer calorimeter measures the energy from particles that pass through the EM calorimeter, and measures the energy of particles that interact via the strong force. This is primarily hadrons. It is composed of steel plates to cause hadronic showering and scintillating tiles as the active material. The signals from the hadronic calorimeter are read out by photomultiplier tubes (PMTs).

5.3 Muon Spectrometer

Because muons are heavier than electrons and photons, and do not interact via the strong force, they generally pass through the detector without being stopped by the calorimeters. The outermost components of the detector are designed specifically to measure the energy and momentum of muons produced in the LHC. The muon spectrometer consists of tracking and triggering system. It extends from the outside of the calormeter system, about a 4.25 m radius from the beam line, to a radius of 11 m. This large detector system is necessary to accurately measure the momentum of muons, which is essential not only for measurements involving the muons themselves, but also to accurately estimate the missing energy in each event.

Two large toroidal magnets within the muon system generate a large magnetic field which covers an area 26 m long with a radius of 10 m. Because the area covered by this magnet system is so large, a uniform magnetic field like the one produced in the Inner Detector is impractical.

Instead, the magnetic field that exists in the muon spectrometer ranges between 2 T and 8 T, and is much less uniform. The path of the muons passing through the spectrometer is bent by this field, allowing their charge to be determined.

1200 tracking chambers are placed in the muon system in order to precisely measure the tracks of muons with high spatial resolution.

5.4 Trigger System

Because of the high collision rate and large amount of data collected by the various subdetectors, ATLAS produces far more data than can actually be stored. Each event produces around 25 Mb of raw data, which multiplied by the bunch crossing rate of 40 MHz, comes out to around a petabyte of data every second. The information from every event cannot practically be stored, therefore a sophisticated trigger system is employed in real time to determine whether events are sufficiently interesting to be worth storing.

The trigger system in ATLAS involves multiple levels, each of which select out which events move on to the next level of scrutiny. The level-1 trigger uses hardware information from the calorimeters and muon spectrometer to select events that contain candidates for particles commonly used in analysis, such as energetic leptons and jets. The level-1 trigger reduces the rate of events from 40 MHz to around 100 kHz.

Events that pass the level-1 trigger move to the High-Level Trigger (HLT). The HLT takes place outside of the detector in software, and looks for properties such as a large amount of missing transverse energy, well defined leptons, and multiple high energy jets. Events that pass the HLT are stored and used for analysis. Because the specifics of the HLT are determined by software rather than hardware, the thresholds can be changed throughout the run of the detector in response to run conditions such as changes to pilup and luminosity. After the HLT is applied, the event rate is reduced to around 1000 per second, which are recorded for analysis.

Part IV

Search for Dimension-Six Operators

6 Data and Monte Carlo Samples

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

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

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

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

6.1 Data Samples

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

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

6.2 Monte Carlo Samples

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

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

Process	Event generator	ME order	Parton Shower	PDF
$t\bar{t}H$	MG5_aMC	NLO	Pythia 8	NNPDF 3.0 NLO [**Ball:2014uwa**]
	(MG5_aMC)	(NLO)	(Herwig++)	(CT10 [**ct10**])
$t\bar{t}W$	MG5_aMC	NLO	Pythia 8	NNPDF 3.0 NLO
	(Sherpa 2.1.1)	(LO multileg)	(Sherpa)	(NNPDF 3.0 NLO)
$t\bar{t}(Z/\gamma^* \to ll)$	MG5_aMC	NLO	Pythia 8	NNPDF 3.0 NLO
VV	Sherpa 2.2.2	MEPS NLO	Sherpa	CT10
$t\bar{t}$	Powheg-BOX v2 [**powhegtt**]	NLO	Pythia 8	NNPDF 3.0 NLO
$t\bar{t}\gamma$	MG5_aMC	LO	Pythia 8	NNPDF 2.3 LO
tZ	MG5_aMC	LO	Pythia 6	CTEQ6L1
$tHqb$	MG5_aMC	LO	Pythia 8	CT10
tHW	MG5_aMC	NLO	Herwig++	CT10
	(Sherpa 2.1.1)	(LO multileg)	(Sherpa)	(NNPDF 3.0 NLO)
tWZ	MG5_aMC	NLO	Pythia 8	NNPDF 2.3 LO
$t\bar{t}t$, $t\bar{t}t\bar{t}$	MG5_aMC	LO	Pythia 8	NNPDF 2.3 LO
$t\bar{t}W^+W^-$	MG5_aMC	LO	Pythia 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	Powheg-BOX v1 [**powhegstp**]	NLO	Pythia 6	CT10
$qqVV$, VVV				
$Z \to l^+l^-$	Sherpa 2.2.1	MEPS NLO	Sherpa	NNPDF 3.0 NLO

7 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.

7.1 Trigger Requirements

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

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

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

7.2 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.

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

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

7.3 Jets

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

7.4 Missing Transverse Energy

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

8 Higgs Momentum Reconstruction

Reconstructing the momentum of the Higgs boson is a particular challenge for channels with leptons in the final state: Because all channels include at least two neutrinos in the final state, the Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted by a neural network when provided with the four-vectors of the Higgs Boson decay products, as shown in section 8.1. With this in mind, several layers of MVAs are used to reconstruction the Higgs momentum.

The first layer is a model designed to select which jets are most likely to be the b-jets that came from the top decay, detailed in section 8.2. As described in section 8.3,the kinematics of these jets are fed into the second layer, which is designed to identify the decay products of the Higgs Boson itself. The kinematics of these particles are then fed into yet another neural-network, which predicts the momentum of the Higgs (8.4). MVAs are also used in the analysis to determine the decay of the Higgs boson in the 3l channel (8.5).

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

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

8.1 Decay Candidate Reconstruction

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

Leptons considered as possible Higgs and top decay candidates are required to pass the selection described in section 7.2. For jets, however, it is found that a large fraction that originate from either the top decay or the Higgs decay fall outside the selection described in section 7.3. Specifically, jets from the Higgs decay tend to be soft, with 32% having $p_T < 25$ GeV. Therefore jets with $p_T < 15$ GeV are considered as possible candidates in the models described below. By contrast, less than 5% of the jets originating from the Higgs fall below this p_T. The jets are found to be well modeled even down to this low p_T threshold, as shown in section 9.1. The impact of using different p_T selection for the jet candidates is considered in detail in section **??**. As they are expected to originate from the primary vertex, jets are also required to pass a JVT cut.

8.2 b-jet Identification

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

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

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

8.2.1 2lSS Channel

For the 2lSS channel, the input features shown in table 3 are used for training. Here j_0 and j_1 are the two jet candidates, while l_0 and l_1 are the two leptons in the event, ordered by p_T. jet DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet, where 5 represents the tightest working point and 1 represents the loosest. The variables nJets DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and 85% b-tag working points, respectively.

As there are far more incorrect combinations than correct ones, by a factor of more than 20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random

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

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

sample of 5 million incorrect entries are used for training, along with close 1 million correct entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for training.

The difference between the distributions for a few of these features for the c̈orrect(̈i.e. both jets are truth b-jets), and ïncorrectc̈ombinations are shown in figure 8.1. The correct and incorrect contributions are scaled to the same integral, so as to better demonstrate the differences in the distributions.

The modeling of these inputs is validated against data, with figure 8.2 showing good general agreement between data and MC. Plots for the complete list of features can found in section A.

Based on the results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as overfitting is found to not be an issue. The output score distribution as well as the ROC curve for

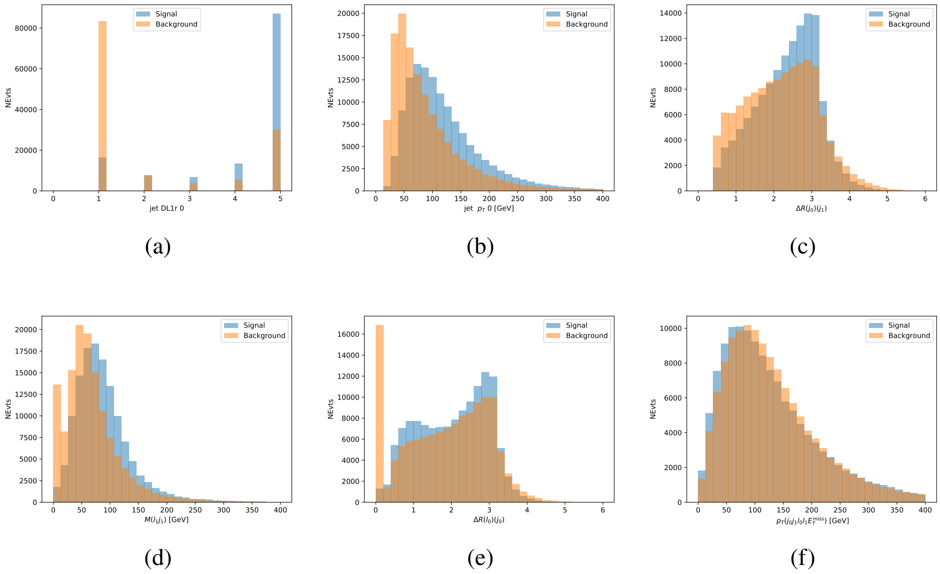


Figure 8.1: Input features for top2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

the trained model are shown in figure 8.2.1. The model is found to identify the correct pairing of jets for 73% of 2lSS signal events on test data.

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

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

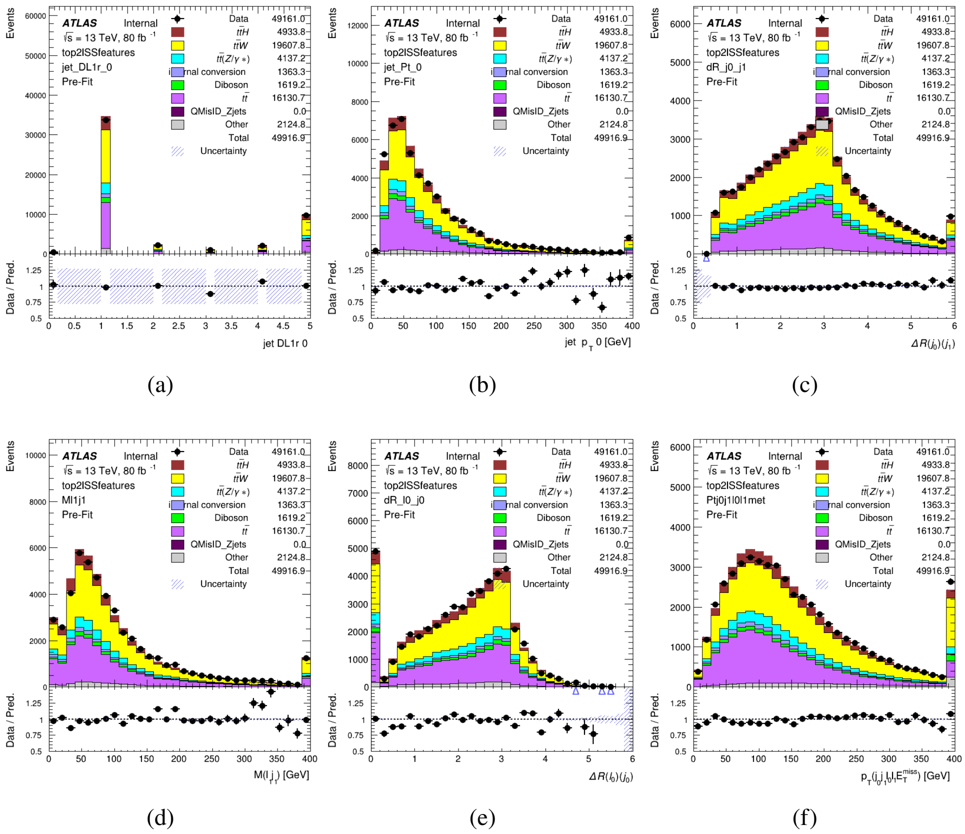


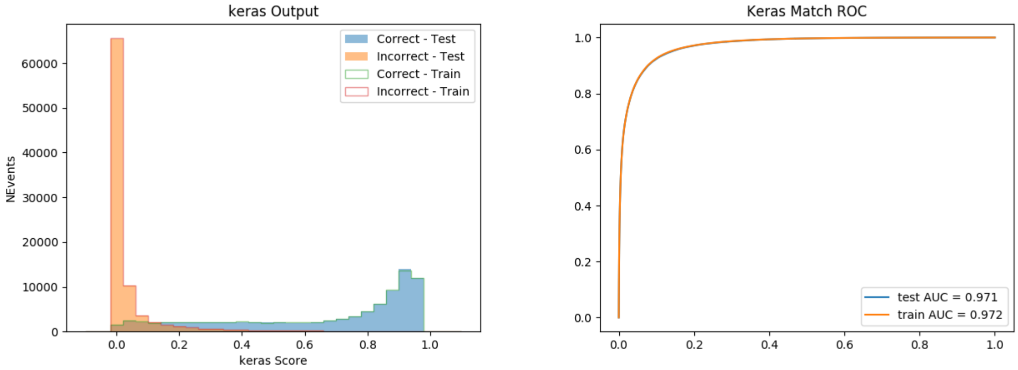
Figure 8.2: Data/MC comparisons of input features for top2lSS training for 80 fb^{-1} of data.

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

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

8.2.2 3l Channel

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



(a) (b)

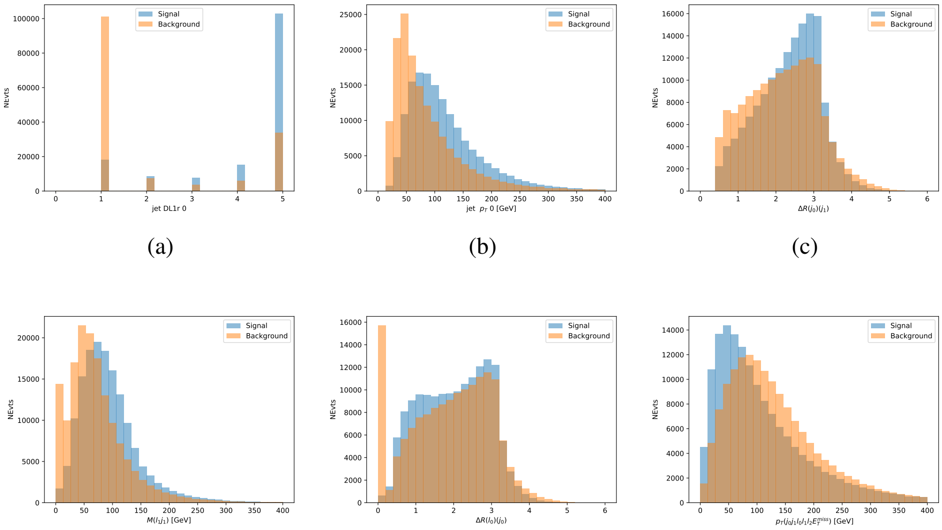
Figure 8.3: (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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

Table 5: Input features

A few of these features are shown in figure 8.4, comaring the distributions for correct and incorrect combinations of jets.

The modeling of these inputs is validated against data, with figure 8.5 showing good general agreement between data and MC. Plots for the complete list of features can found in section A.



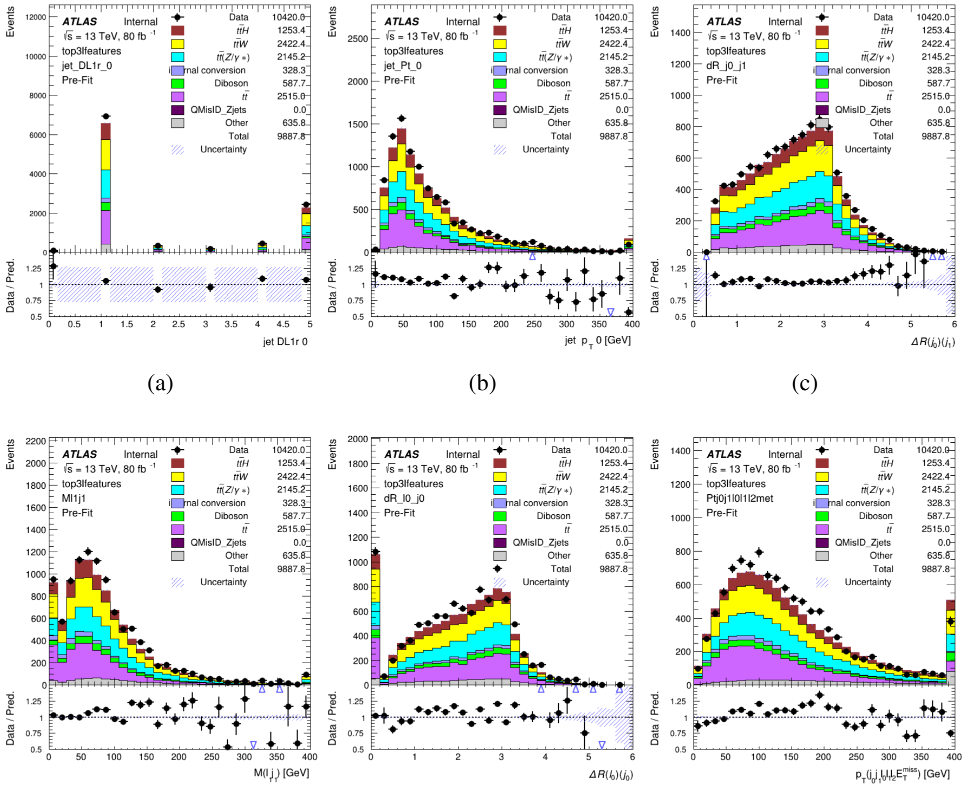
(a) (b) (c)

(d) (e) (f)

Figure 8.4: 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, the dataset is downsized to reduce the ratio of correct and incorrect combination from 20:1, to 5:1. Around 7 million events are used for training, with 10% set aside for testing. Based on the results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with 60 nodes each. The output score distribution as well as the ROC curve for the trained model are shown in figure 8.2.2.

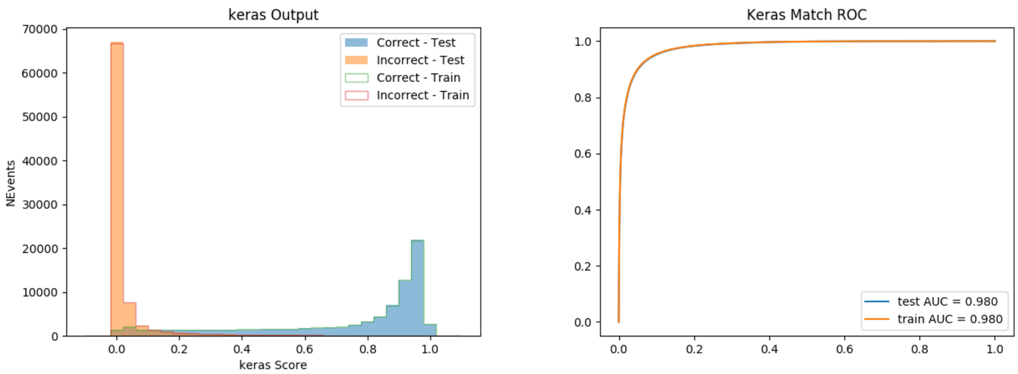
This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal events. The accuracy of the model is summarized in table 6.



(a) (b) (c)

(d) (e) (f)

Figure 8.5: Data/MC comparisons of input features for top3l training for 80 fb^{-1} of data.



(a) (b)

Figure 8.6: tmp

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

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

8.3 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.

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

3l events are therefore categorized as either semi-leptonic or fully-leptonic. In the semi-leptonic case the reconstructed decay products consist of two jets and a single leptons. For the fully-leptonic case, the decay products include 2 of the three leptons associated with the event. For training the models, events are separated into these two categories using truth level information. A separate MVA, described in section 8.5, is used to make this distinction at reco level and determine which model to use.

For all channels, the models described in section 8.2 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.

8.3.1 2lSS Channel

For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural network is trained on the kienmatics of different combinations of leptons and jets, as well as the b-jets identified in section 8.2, with the specific input features listed in table **??**.

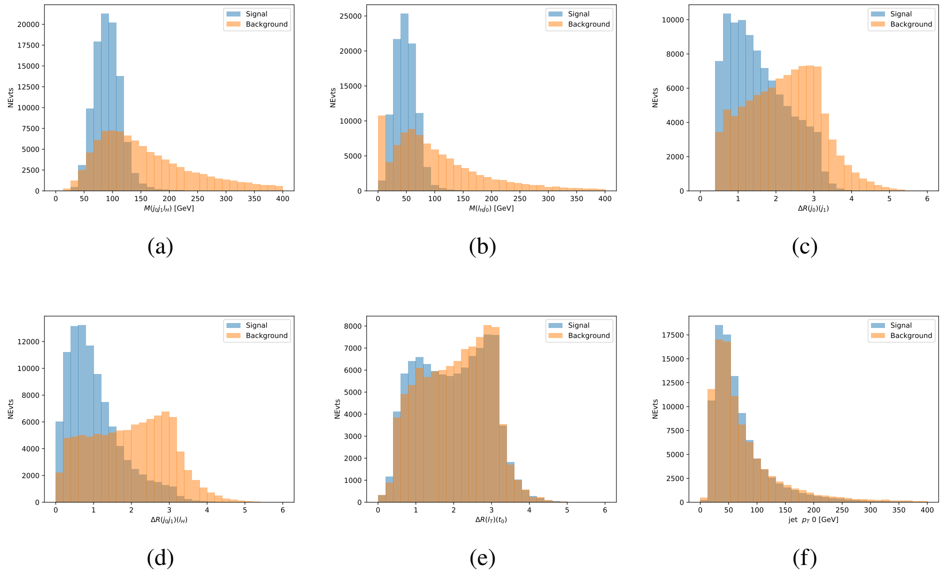
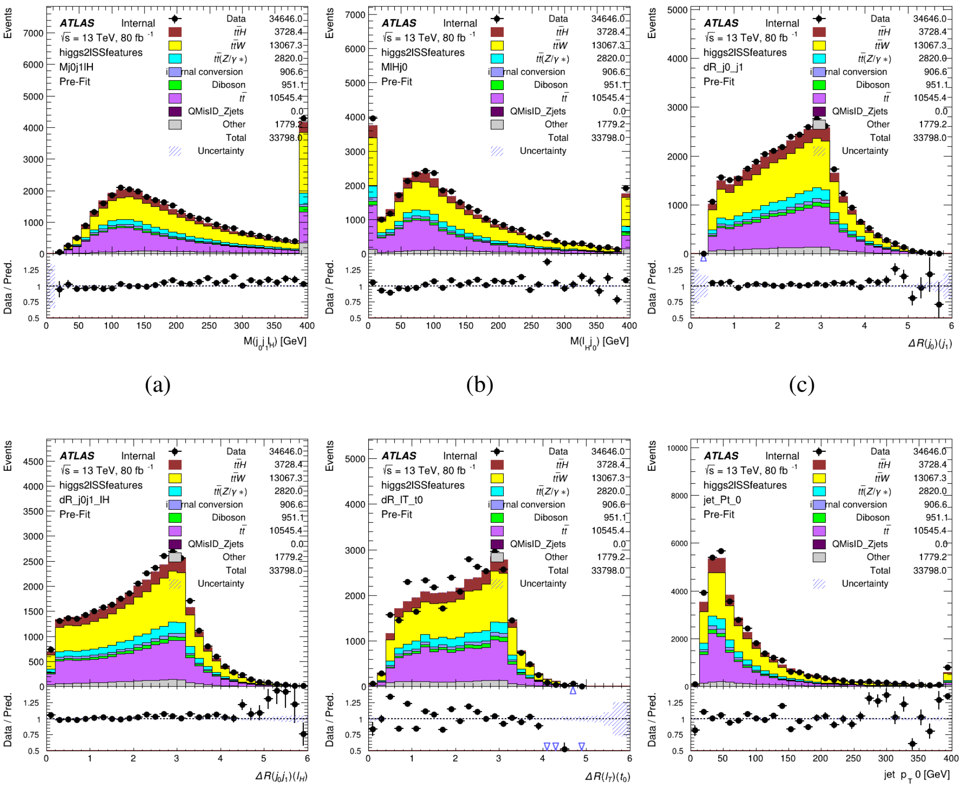


Figure 8.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

The modeling of these inputs is validated against data, with figure 8.2 showing good general agreement between data and MC. Plots for the complete list of features can found in section A.

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



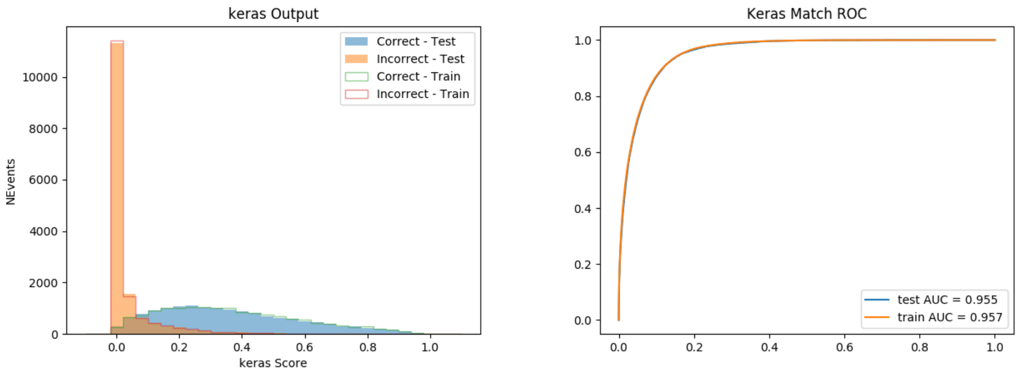
(a) (b) (c)

(d) (e) (f)

Figure 8.8: Data/MC comparisons of input features for higgs2lSS training for 80 fb^{-1} of data.

8.3.2 3l Semi-leptonic Channel

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



(a)

(b)

Figure 8.9: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

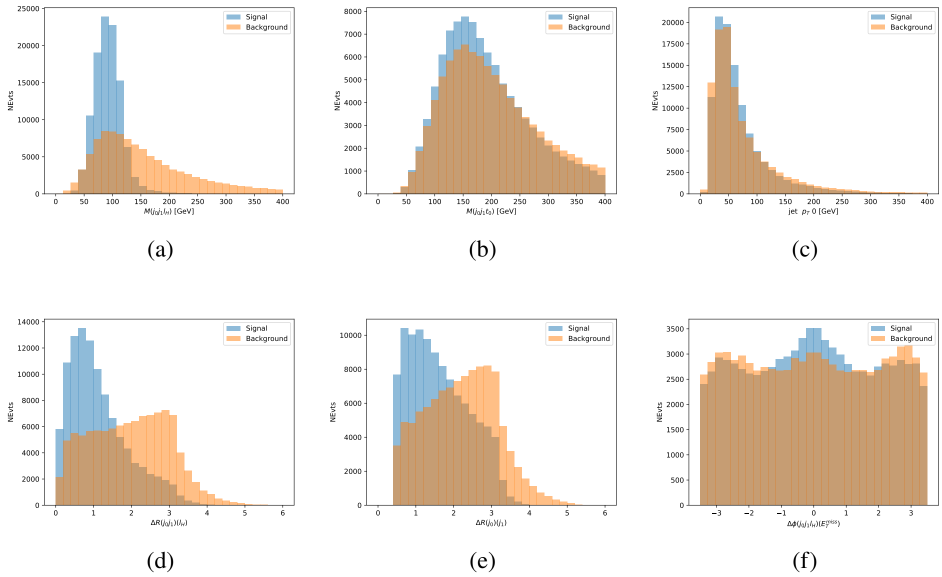
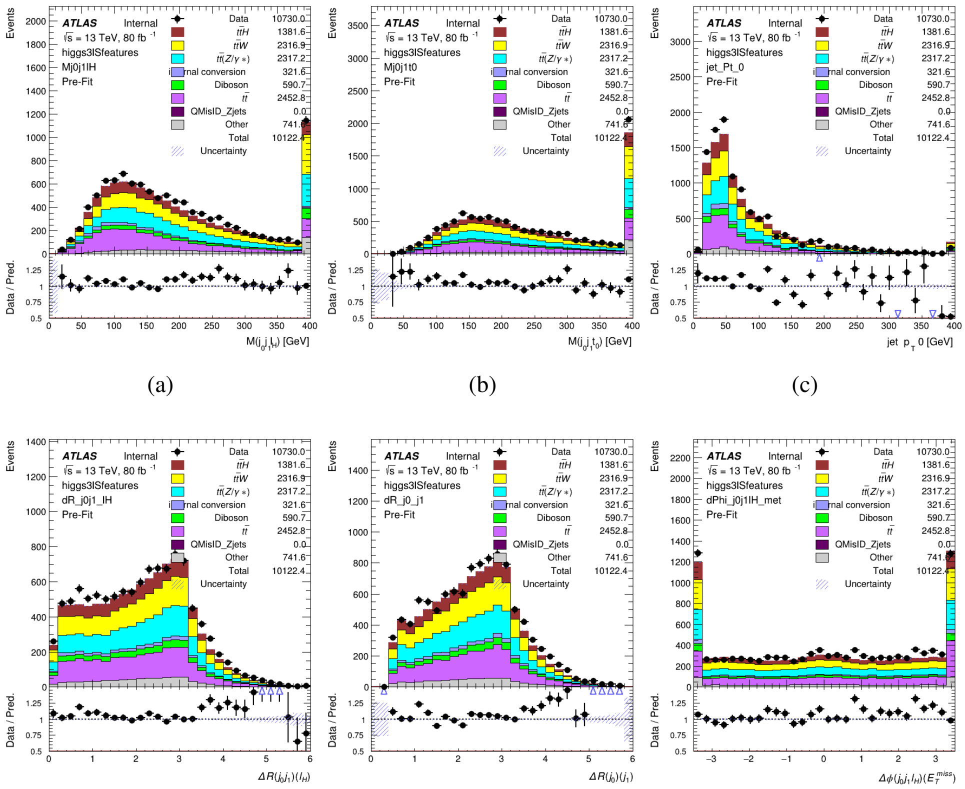


Figure 8.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

The modeling of these inputs is validated against data, with figure 8.11 showing good general agreement between data and MC. Plots for the complete list of features can found in section A.

The neural network identifies the correct combination 65% of the time. It identifies the correct lepton 85% of the time, anselects the correct lepton and at least one of the correct jets 83% of the time.



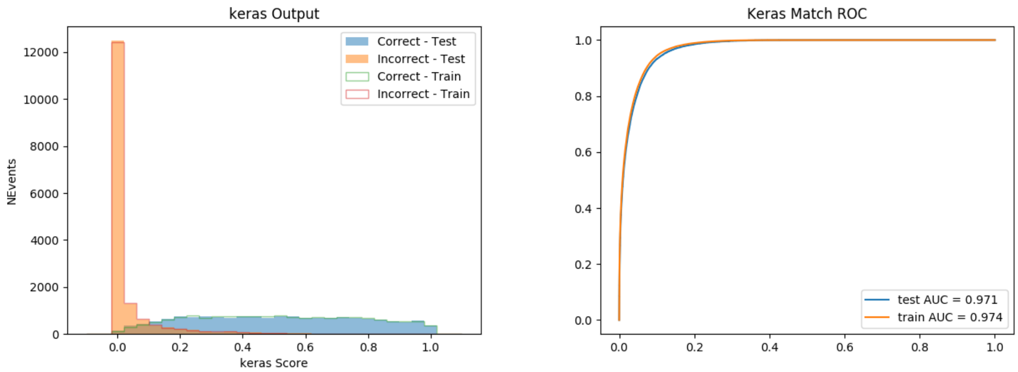
(a) (b) (c)

(d) (e) (f)

Figure 8.11: Data/MC comparisons of input features for higgs3lS training for 80 fb^{-1} of data.

8.3.3 3l Fully-leptonic Channel

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



(a) (b)

Figure 8.12: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

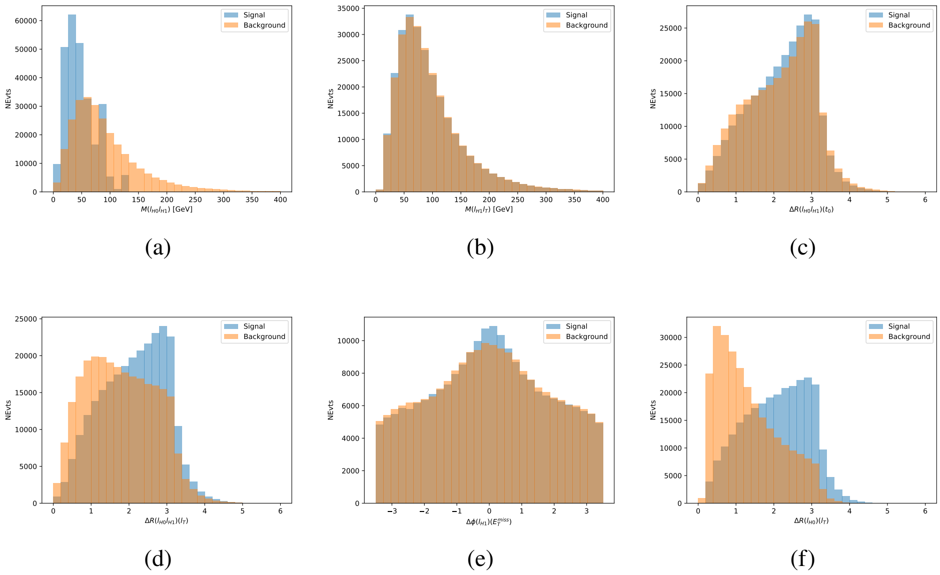


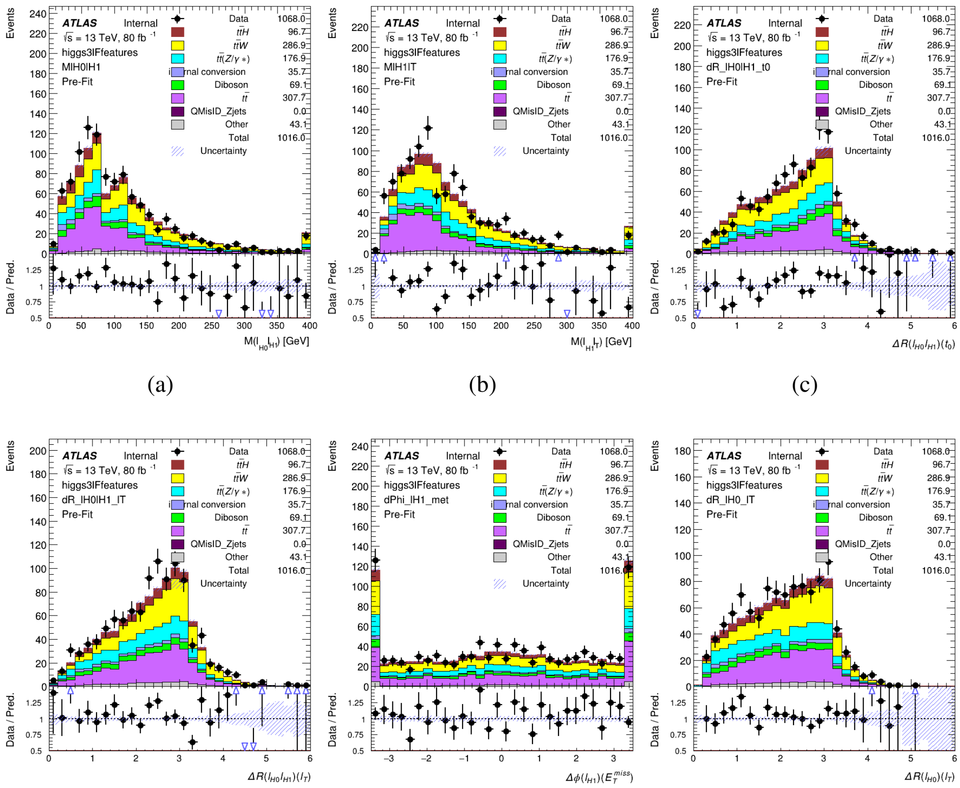
Figure 8.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

The modeling of these inputs is validated against data, with figure 8.14 showing good general agreement between data and MC. Plots for the complete list of features can found in section A.

The correct lepton is identified 80% of the time.

8.4 p_T Prediction

Once the most probable decay products have been identified, their kinematics are used as inputs to a regression model which attempts to predict the momentum of the Higgs Boson. Once again, a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay



(a) (b) (c)

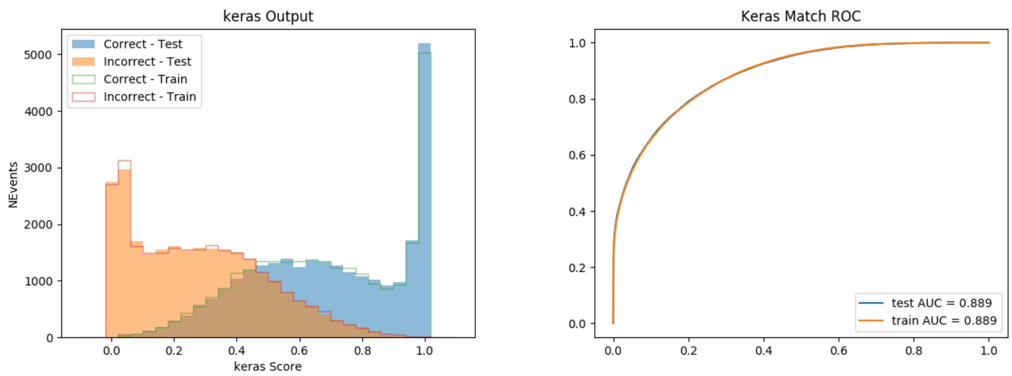
(d) (e) (f)

Figure 8.14: Data/MC comparisons of input features for higgs3lF training for 80 fb^{-1} of data.

are included as well, as these are found to improve performance. The truth p_T of the Higgs, as predicted by MC, are used as labels. Separate models are built for each channel - 2lSS, 3l Semi-leptonic and 3l Fully-leptonic.

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



(a) (b)

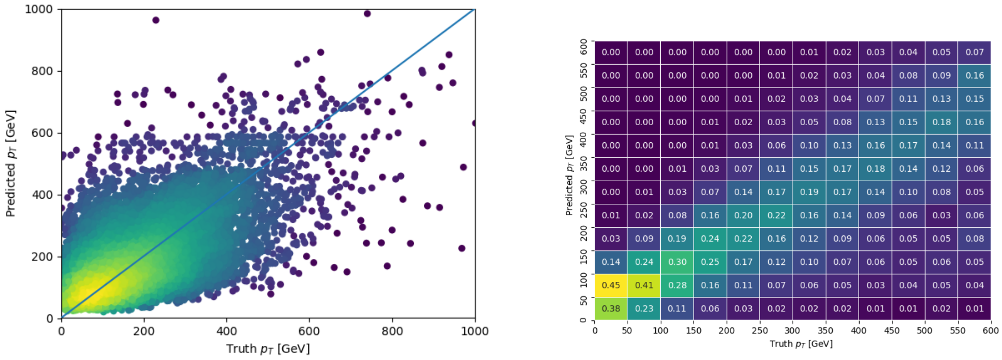
Figure 8.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

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.

8.4.1 2lSS Channel

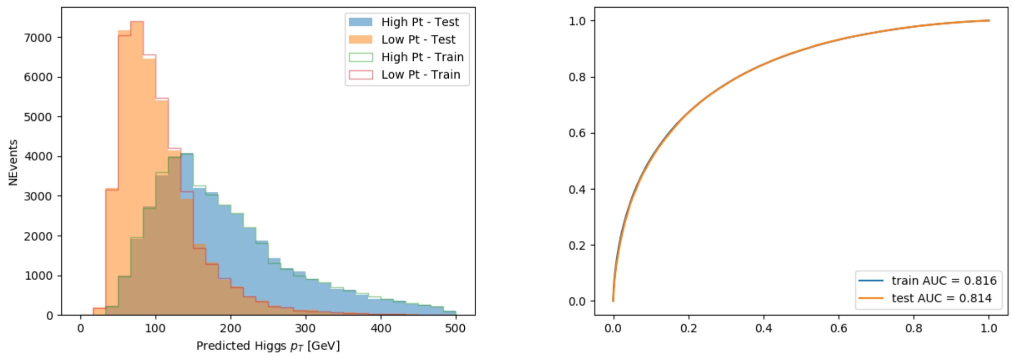
The input variables listed in table **??** are used to predict the Higgs p_T in the 2lSS channel. Here j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating from the Higgs is labeled l_H, while the other lepton is labeled l_T, as it most have come from the decay of one of the top quarks. 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 5 hidden layers with 40 nodes each. The inputdata set includes 1.2 million events, 10% of which is used for testing, the other 90% for training. Training is found to converge after around 150 epochs.



(a) (b)

Figure 8.16:

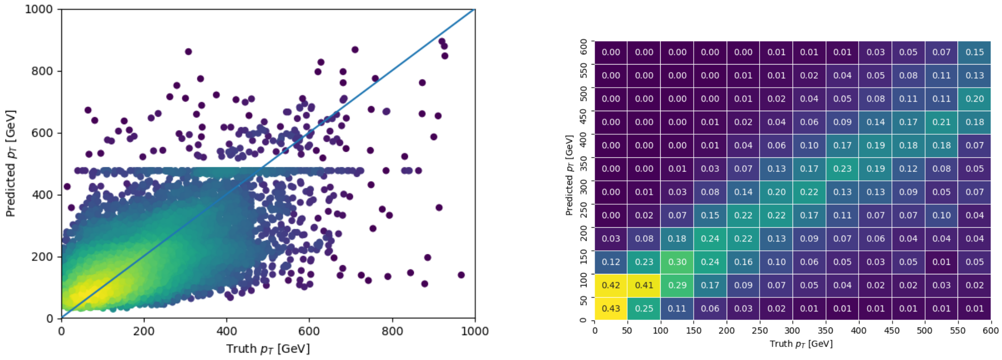


(a) (b)

Figure 8.17:

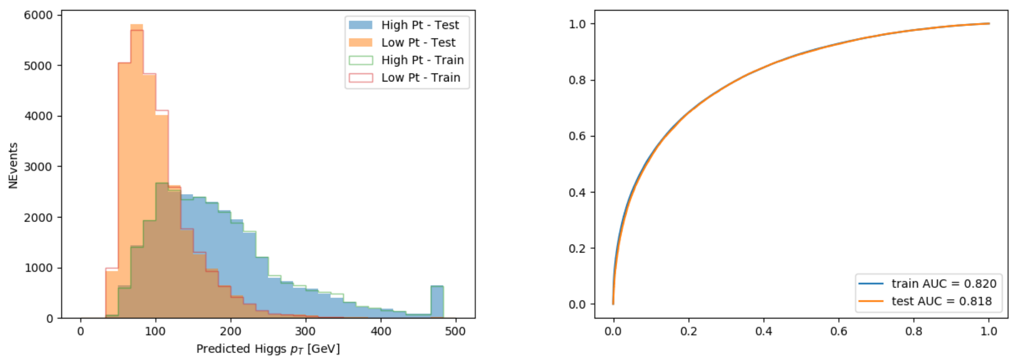
8.4.2 3l Semi-leptonic Channel

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



(a) (b)

Figure 8.18:



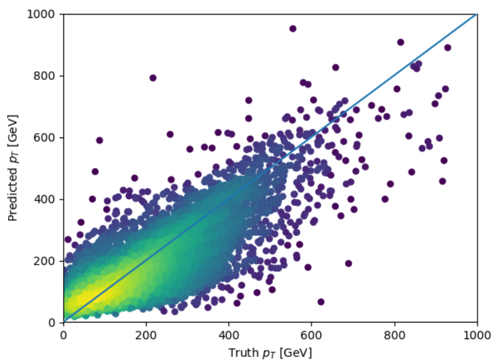
(a) (b)

Figure 8.19:

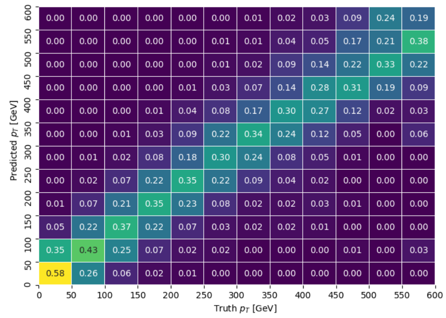
8.4.3 3l Fully-leptonic Channel

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

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



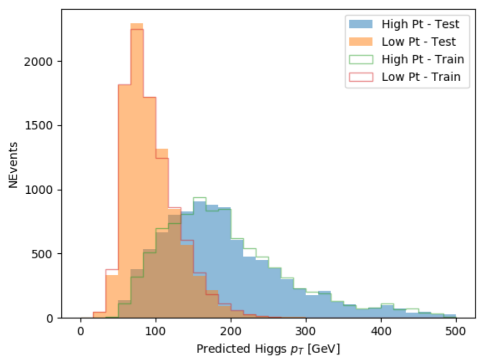
(a)



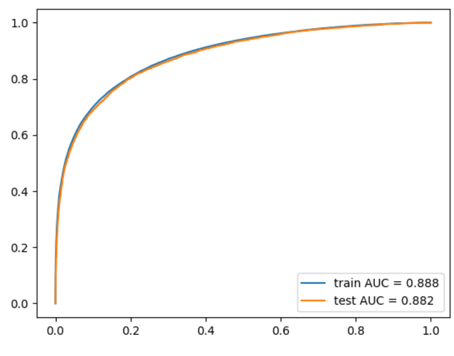
(b)

Figure 8.20:

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



(a)



(b)

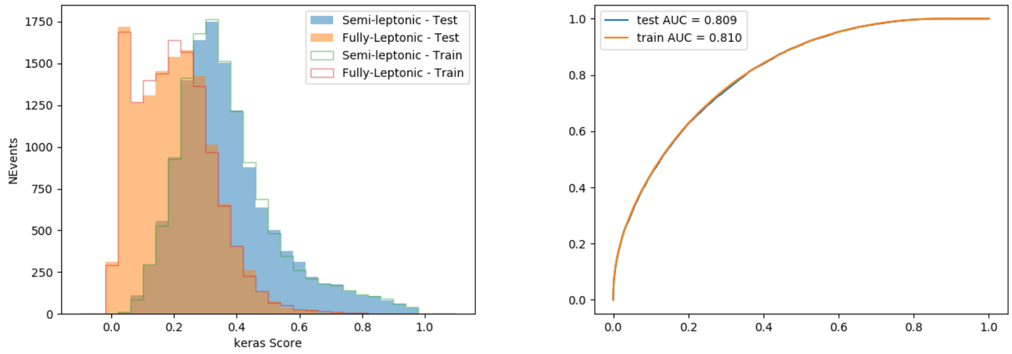
Figure 8.21:

8.5 3l Decay Mode

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

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

A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these two decay modes. The output of the model is summarized in figure 8.22.



(a) (b)

Figure 8.22:

9 Signal Region Definitions

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

9.1 Pre-MVA Event Selection

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

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

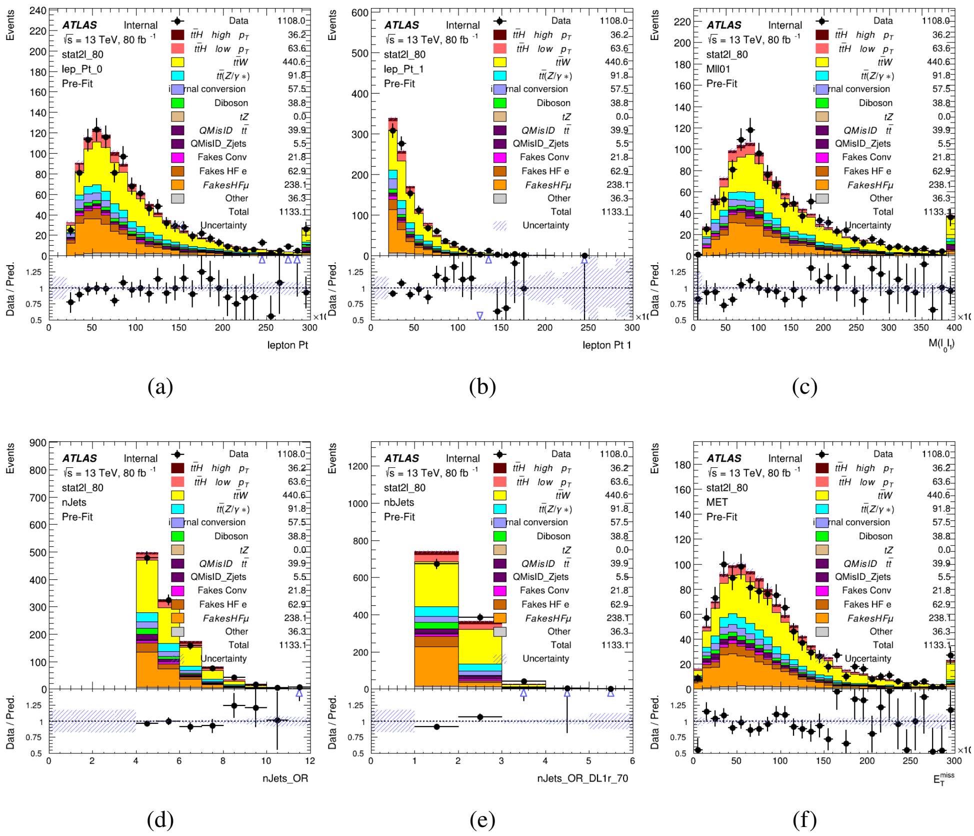


Figure 9.1:

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

- Three light leptons with total charge ±1
- Same charge leptons are required to be very tight, with $p_T > 20$ GeV

- Opposite charge lepton must be loose, with $p_T > 10$ GeV
- >=2 reconstructed jets, >=1 b-tagged jets
- No reconstructed tau candidates
- $|M(l^+l^-) - 91.2\text{ GeV}| > 10$ GeV for all opposite-charge, same-flavor lepton pairs

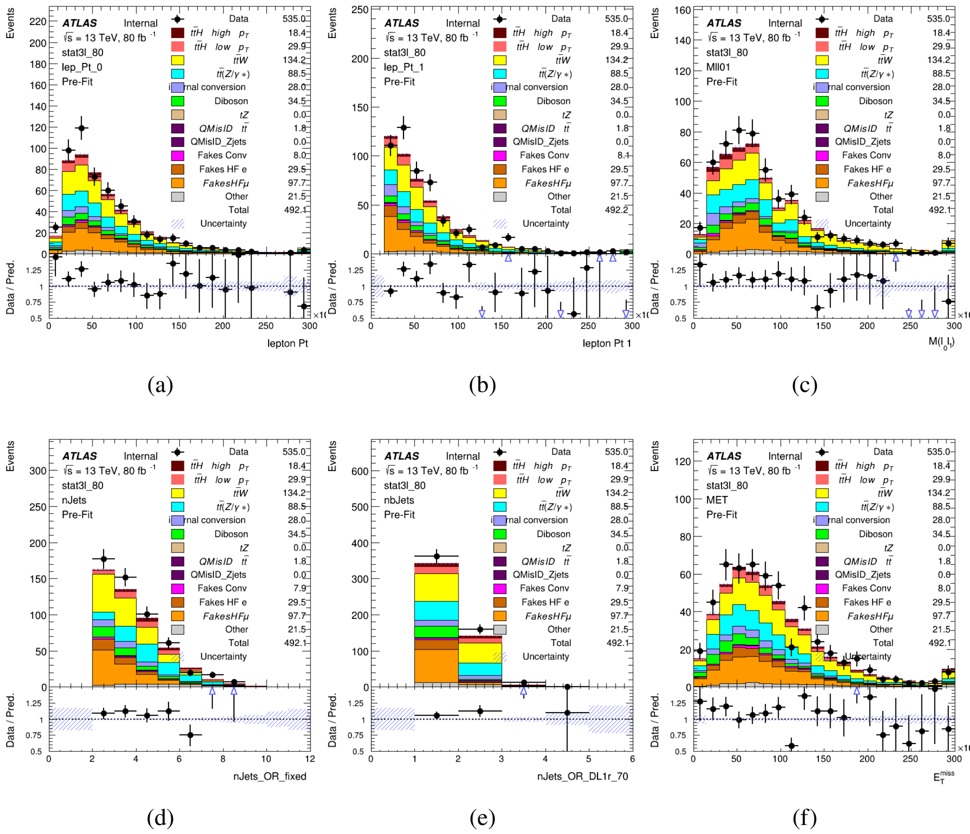


Figure 9.2:

9.2 Event MVA

Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost [**xgboost**] are trained using the kinematics of signal and background events derived from Monte Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo event.

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

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

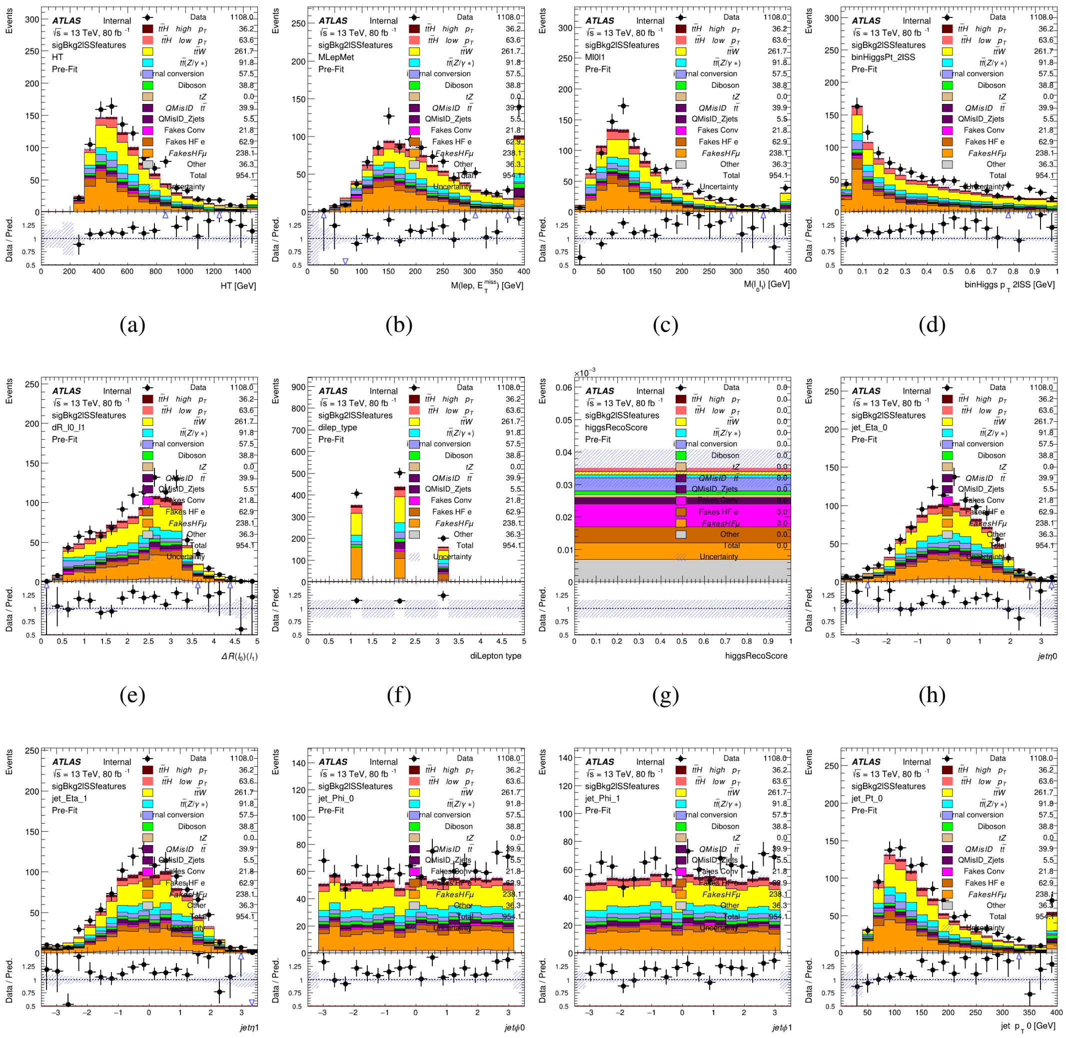
HT	M(lep, E_T^{miss})	$M(l_0 l_1)$
binHiggs p_T 2lSS	$\Delta R(l_0)(l_1)$	diLepton type
higgsRecoScore	jet η 0	jet η 1
jet ϕ 0	jet ϕ 1	jet p_T 0
jet p_T 1	Lepton η 0	Lepton η 1
Lepton ϕ 0	Lepton ϕ 1	Lepton p_T 0
Lepton p_T 1	E_T^{miss}	min $\Delta R(l_0)(jet)$
min $\Delta R(l_1)(jet)$	min $\Delta R(Lepton)(bjet)$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

Table 7: Input features

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

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

Table 8: Input features



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

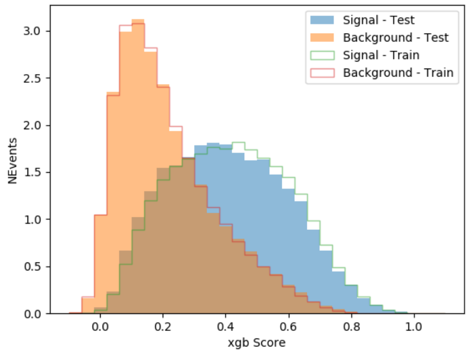
(e) (f) (g) (h)

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

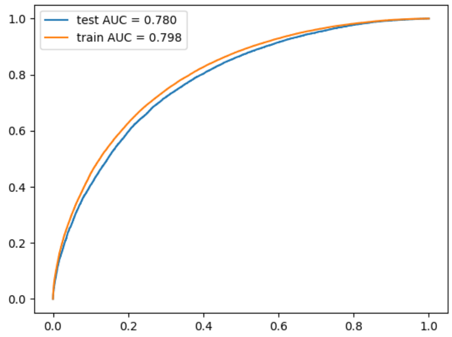
Figure 9.3:

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

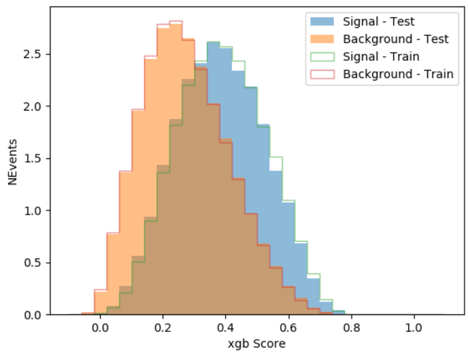
Output distributions of each MVA are shown in figure 9.2.



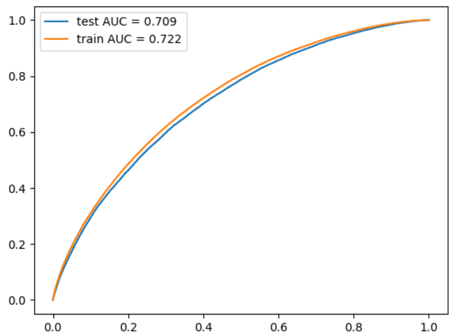
(a)



(b)



(c)



(d)

Figure 9.4:

9.3 Signal Region Definitions

Once pre-selection has been applied, channels are further refined based on the MVAs described above. The output of the model described in section 8.5 is used to separate the three channel into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson.

For each event, depending on the channel as well as the predicted p_T of the Higgs derived

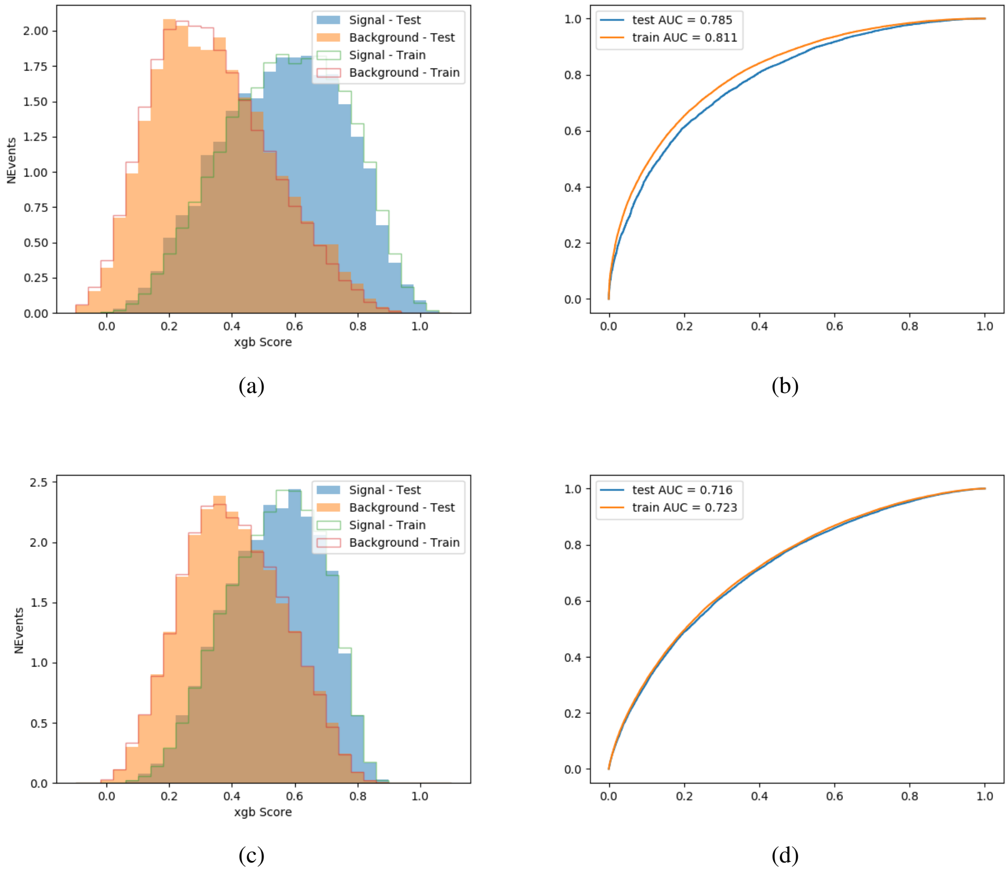


Figure 9.5:

from the algorithm described in section 8.4, a cut on the appropriate background rejection algorithm is applied. The specific selection used, and the event yield in each channel after this selection has been applied, is summarized below.

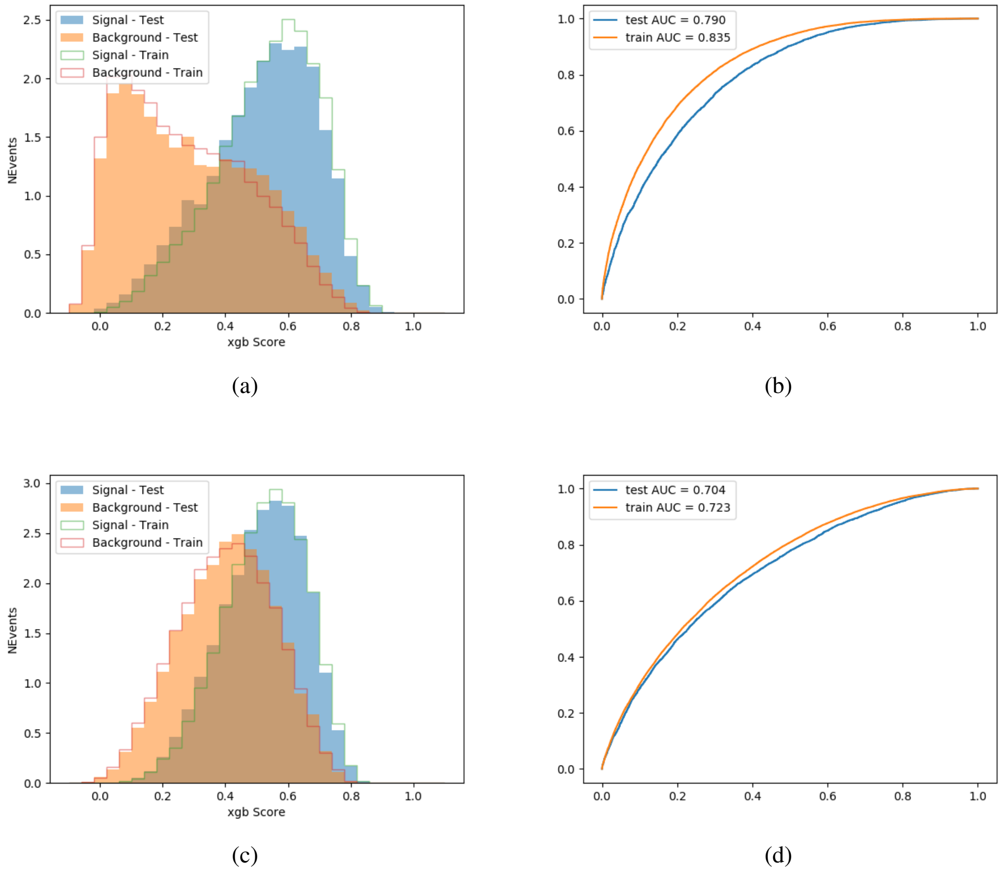

Signal - Test
Background - Test
Signal - Train
Background - Train
NEvents
xgb Score
(a)
test AUC = 0.790
train AUC = 0.835
(b)
Signal - Test
Background - Test
Signal - Train
Background - Train
NEvents
xgb Score
(c)
test AUC = 0.704
train AUC = 0.723
(d)

Figure 9.6:

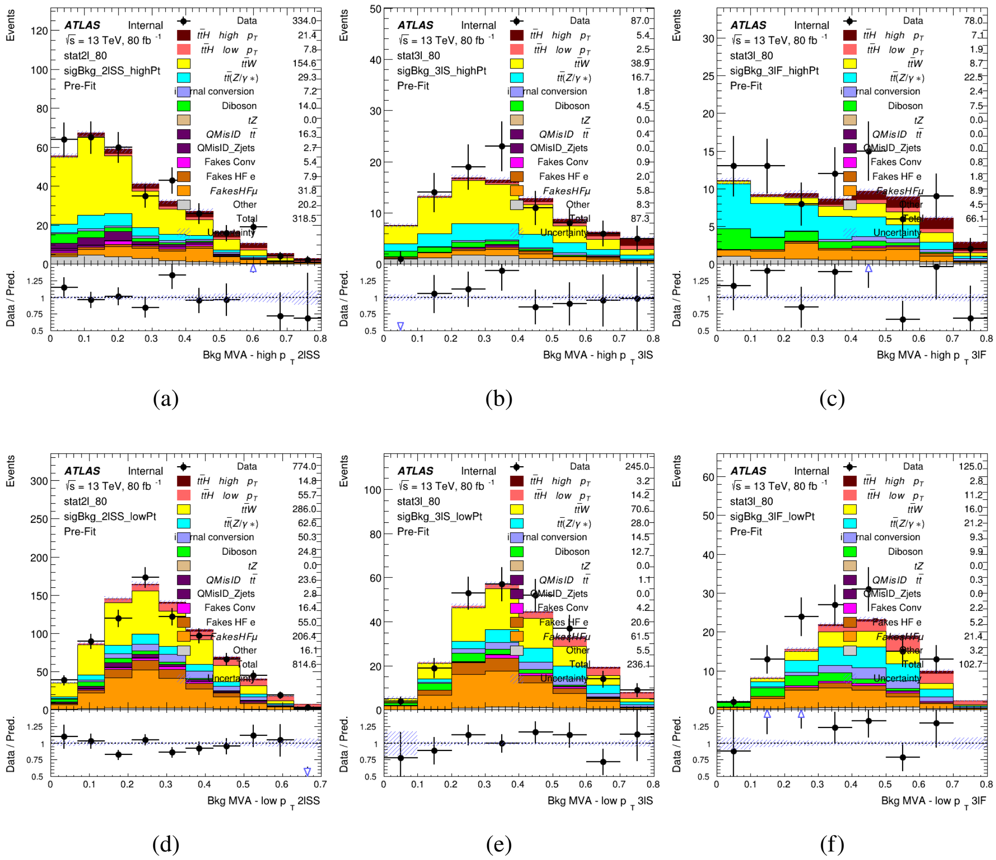


Figure 9.7: scores

9.3.1 $2lSS$

9.3.2 $3l - Semi - leptonic$

9.3.3 $3l - Fully - leptonic$

10 Background Rejection MVA

10.1 Background Rejection MVAs

Separate mdoels are used in order to distinguish signal events from background for each analysis channel - 2lSS, 3l semi-leptonic, and 3l fully leptonic. In particular, Neural Networks produced with Tensorflow are trained using the kinematics of signal and background events derived from Monte Carlo simulations. Further, because the background composition differs for events with a high reconstructed Higgs p_T compared to events with low reconstructed Higgs p_T, separate MVAs are produced for high and low p_T regions.

10.1.1 2lSS - High p_T

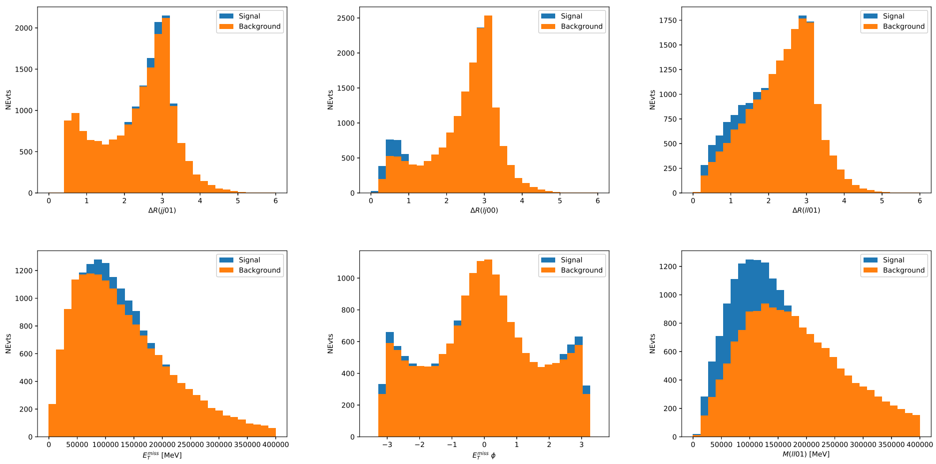


Figure 10.1:

10.1.2 2lSS - Low p_T

10.1.3 3l Semi-Leptonic - High p_T

10.1.4 3l Semi-Leptonic - Low p_T

10.1.5 3l Fully Leptonic - High p_T

10.1.6 3l Fully Leptonic - Low p_T

11 Systematic Uncertainties

The systematic uncertainties that are considered are summarized in table **??**. These are implemented in the fit either as a normalization factors or as a shape variation or both in the signal and background estimations. The numerical impact of each of these uncertainties is outlined in section 12.

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

The experimental uncertainties are related to the reconstruction and identification of light leptons and and b-tagging of jets, and to the reconstruction of E_T^{miss}. The sources which contribute

Table 9: 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

to the uncertainty in the jet energy scale [**jes**] are decomposed into uncorrelated components and treated as independent sources in the analysis.

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

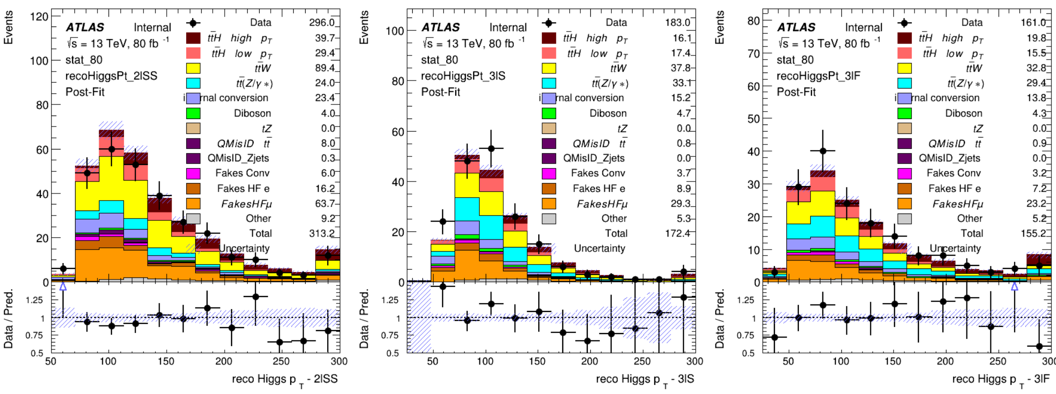
The systematic uncertainties associated with the signal and background processes are accounted for by varying the cross-section of each process within its uncertainty.

12 Results

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

12.1 Results - 80 fb^{-1}

A maximum likelihood fit is performed simultaneously over the regions shown in figure 12.1.



(a) (b) (c)

Figure 12.1:

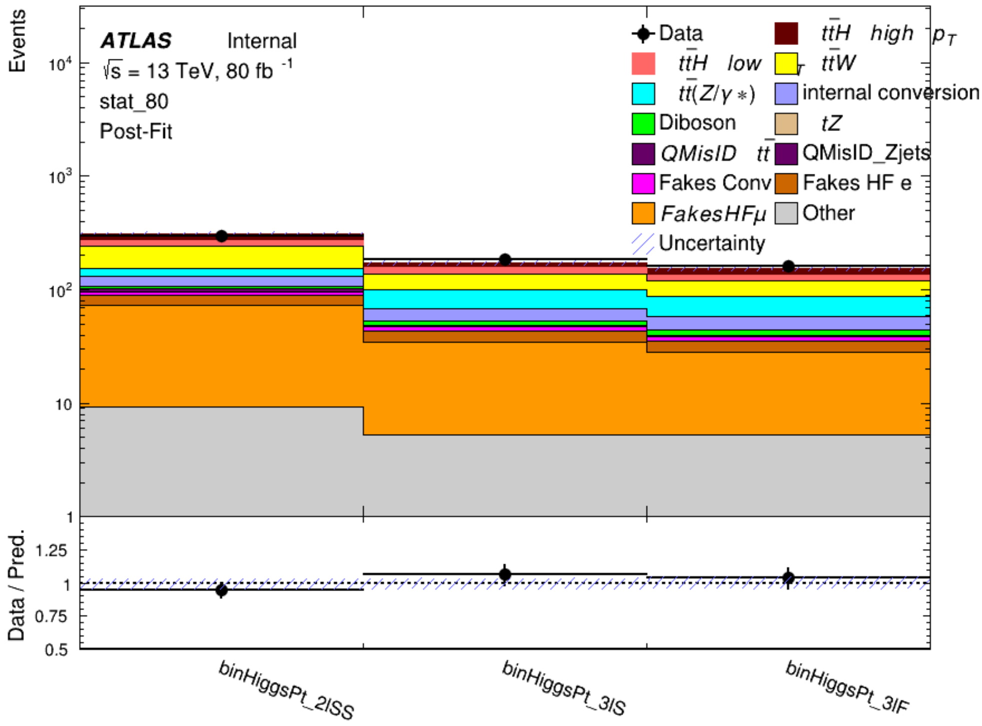


Figure 12.2: Post-fit summary of fit.

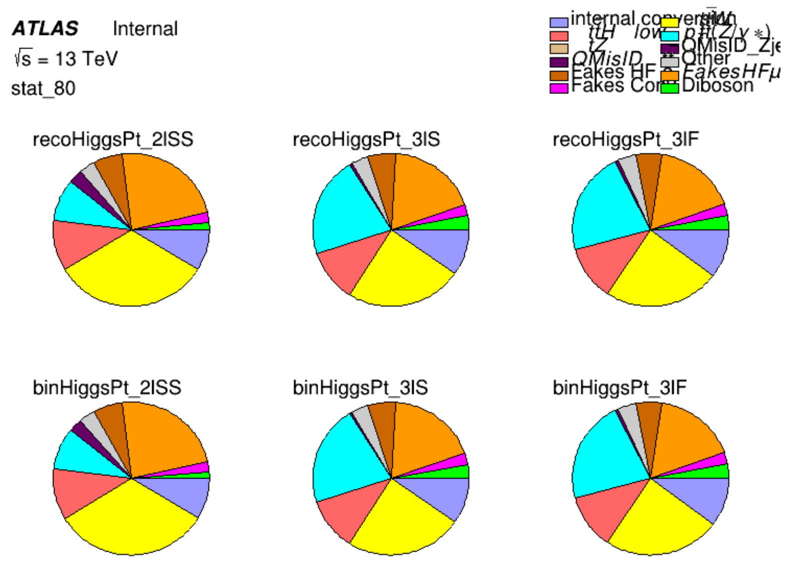
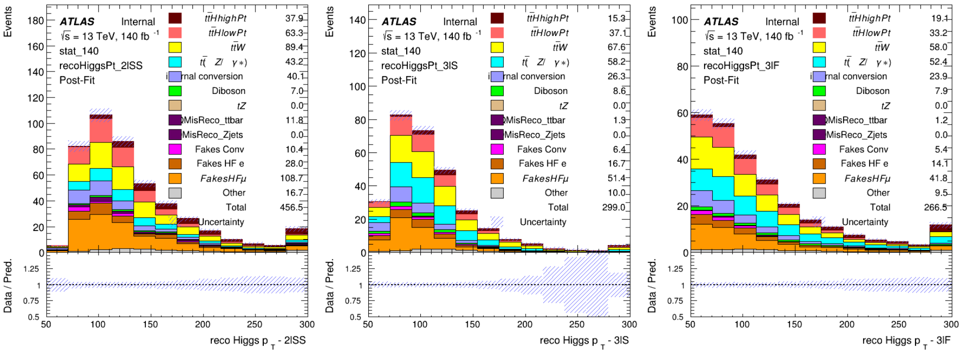


Figure 12.3: Background composition of the fit regions.



(a) (b) (c)

Figure 12.4:

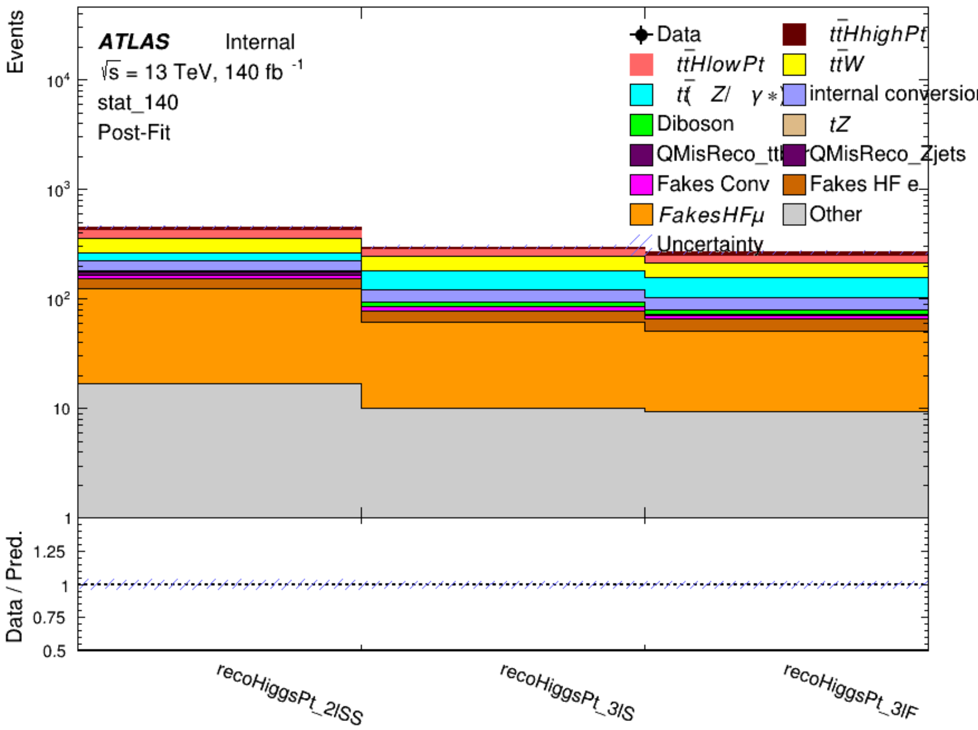


Figure 12.5: Post-fit summary of fit.

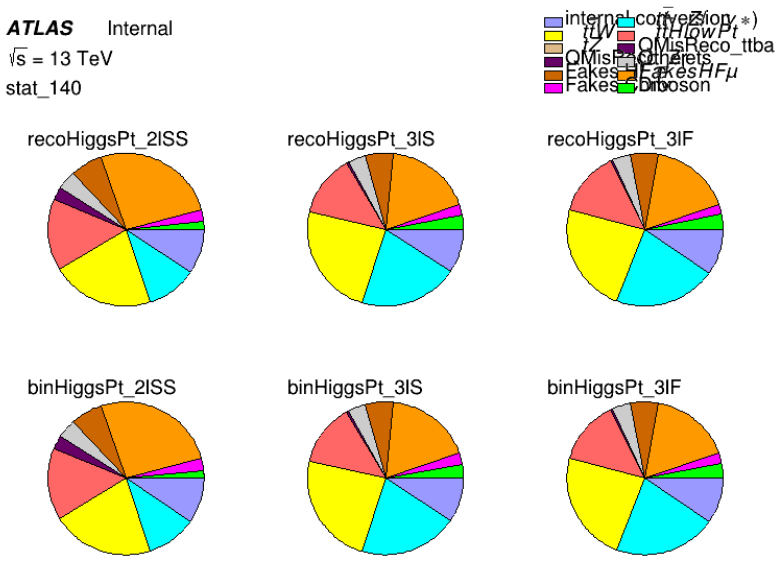


Figure 12.6: Background composition of the fit regions.

12.2 Projected Results - 140 fb^{-1}

Part V

Conclusion

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

List of contributions

Appendices

A Machine Learning Models

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

A.1 Higgs Reconstruction Models

A.1.1 b-jet Identification Features - 2lSS

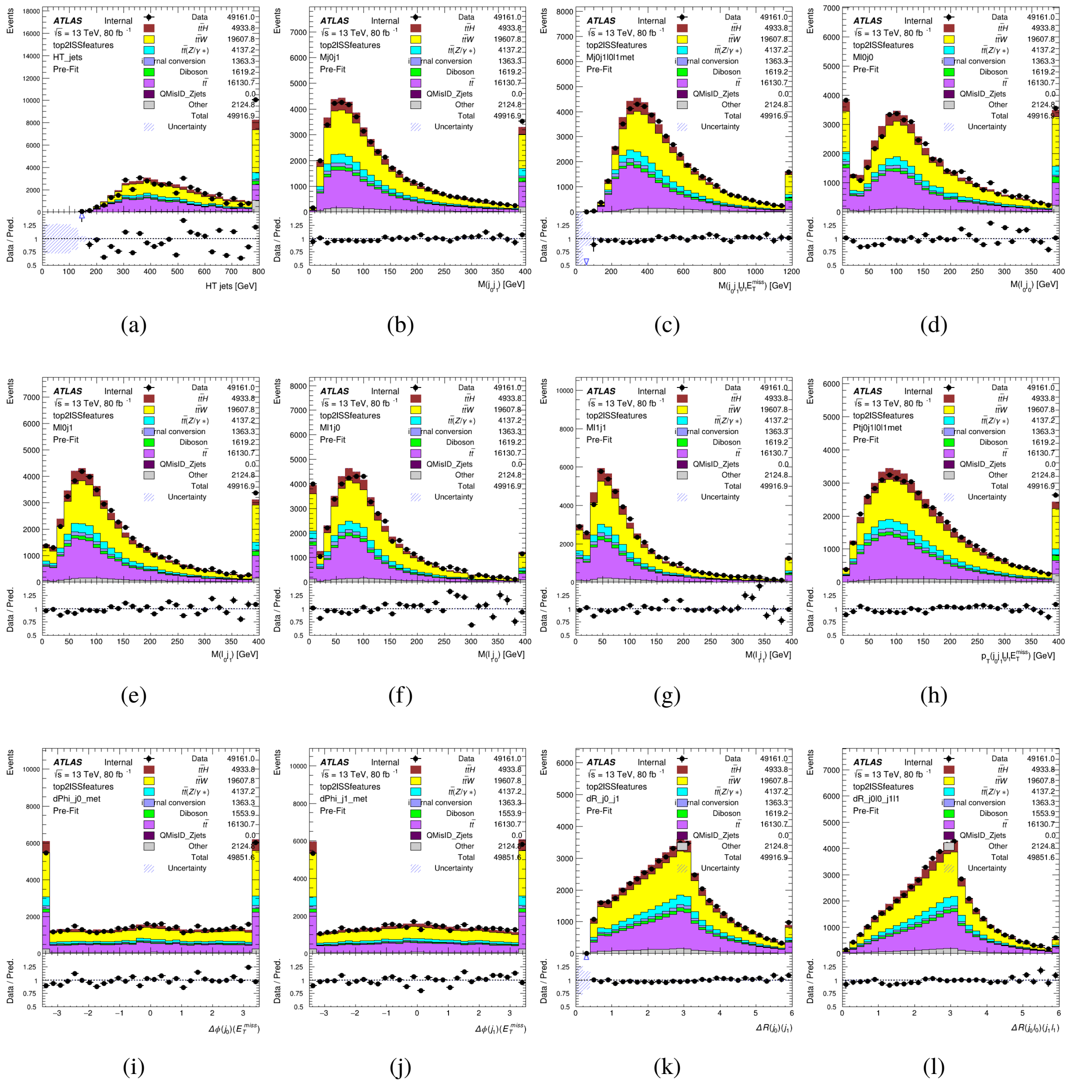
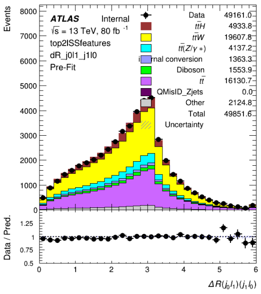
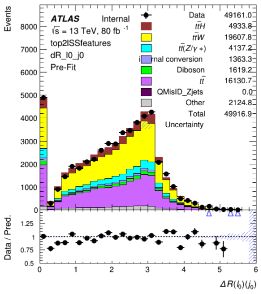
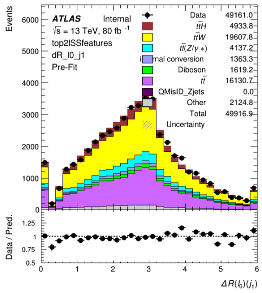
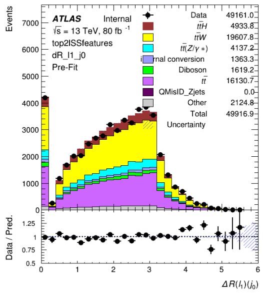


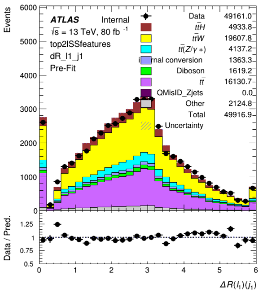
Figure A.1: Input features for top2lSS

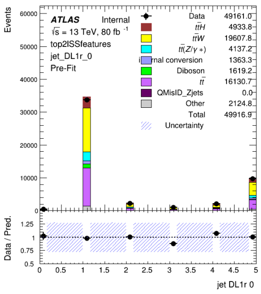

(a)

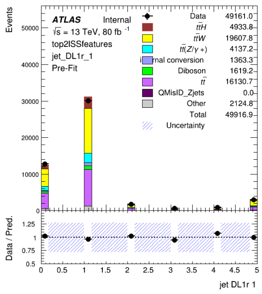

(b)

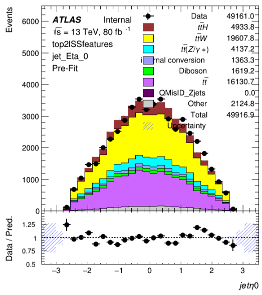

(c)

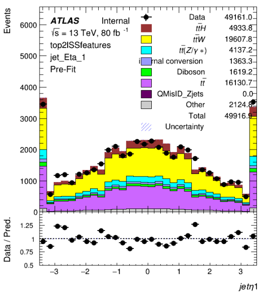

(d)

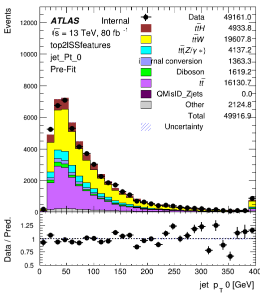

(e)

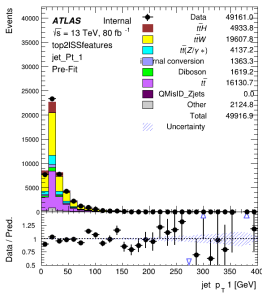

(f)


(g)


(h)


(i)


(j)


(k)

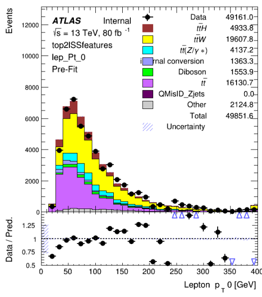
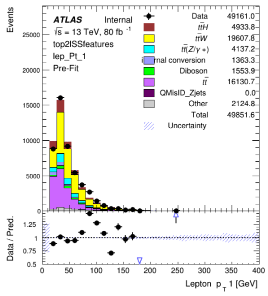
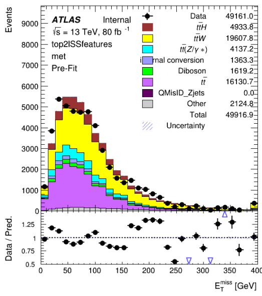

(l)

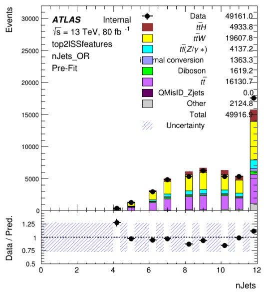
Figure A.2: Input features for top2lSS



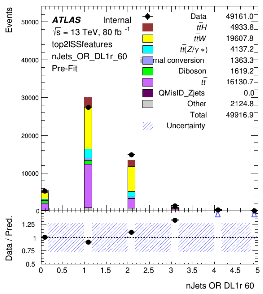
(a)



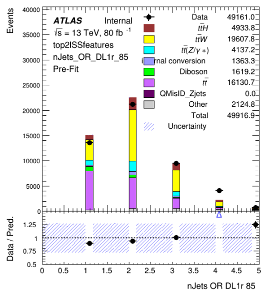
(b)



(c)



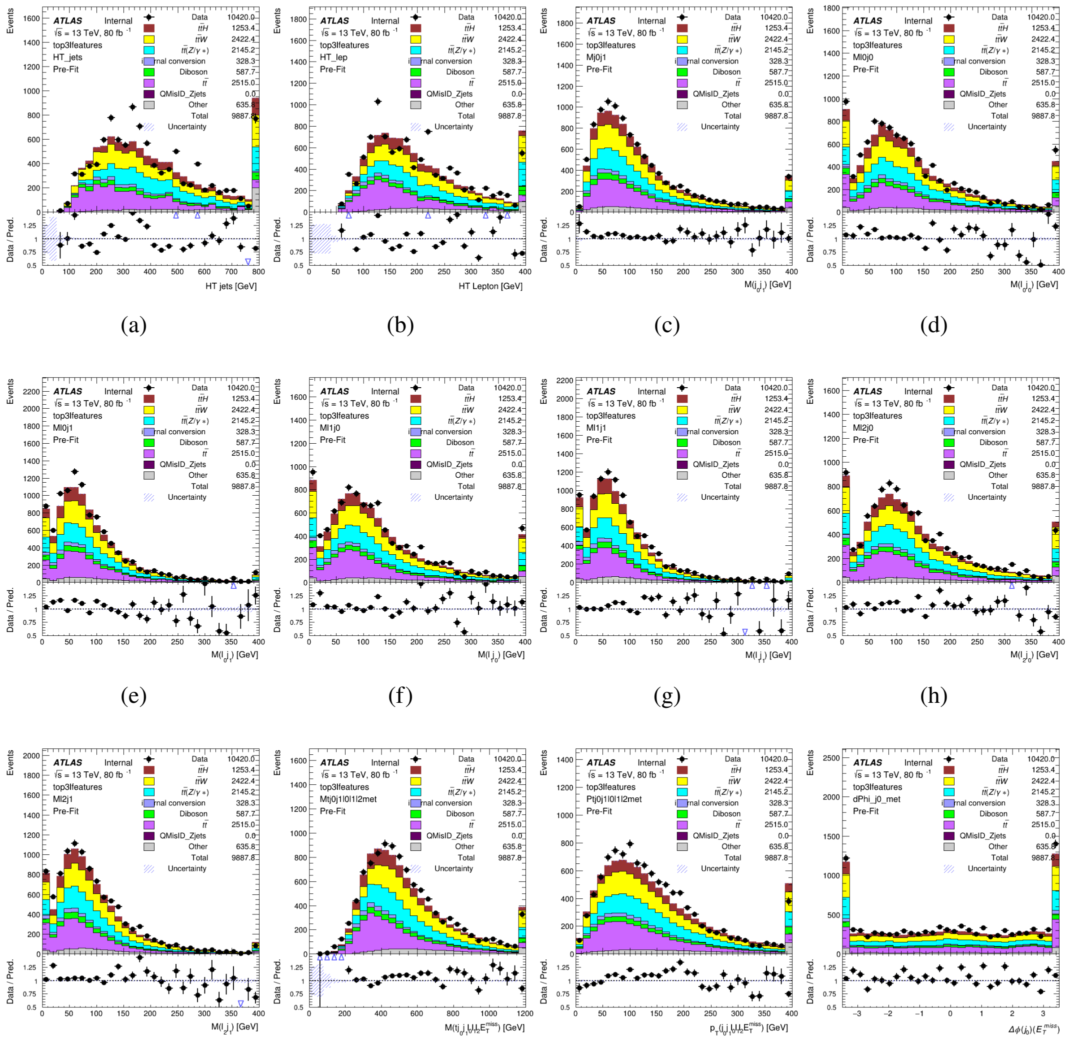
(d)



(e)

Figure A.3: Input features for top2lSS

A.1.2 b-jet Identification Features - 3l

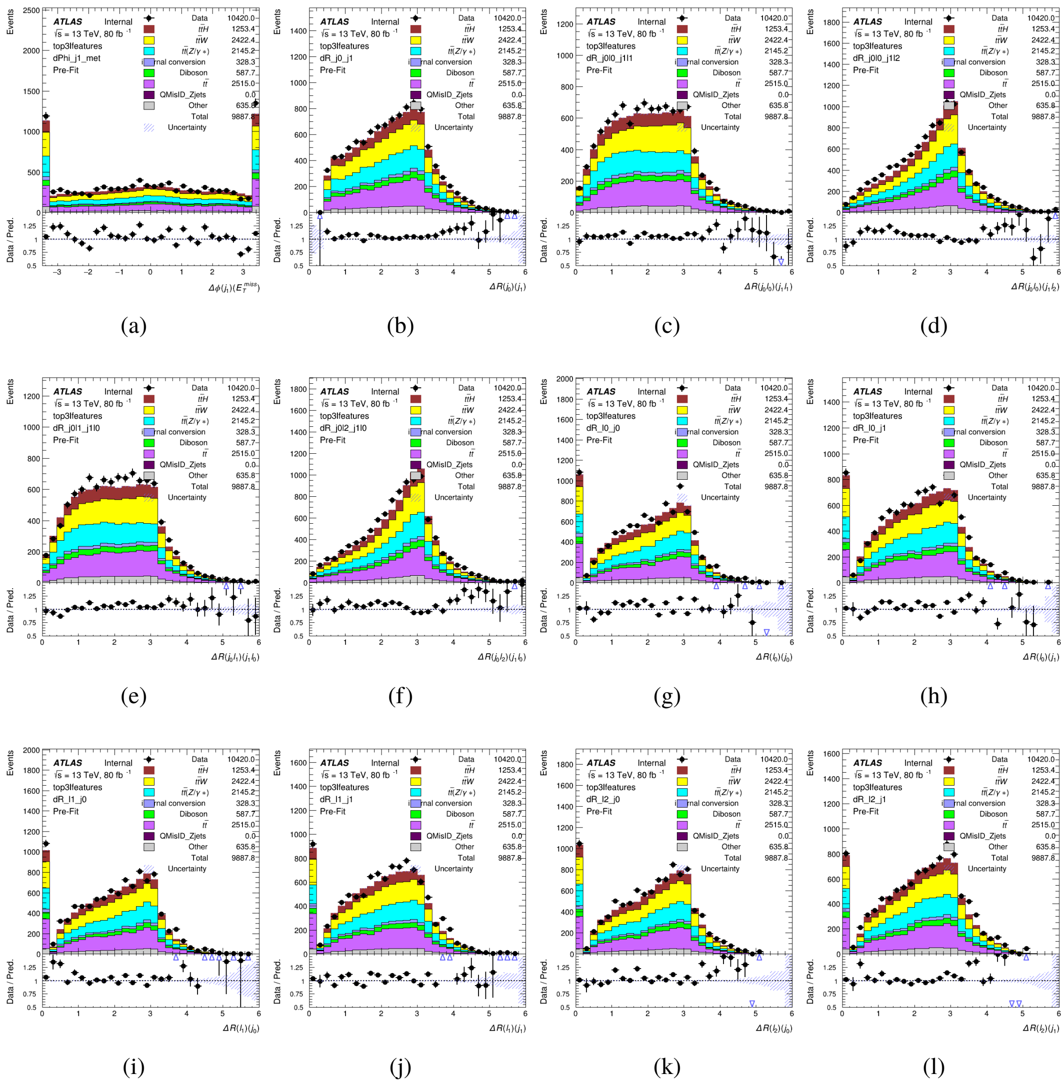


(a) (b) (c) (d)

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

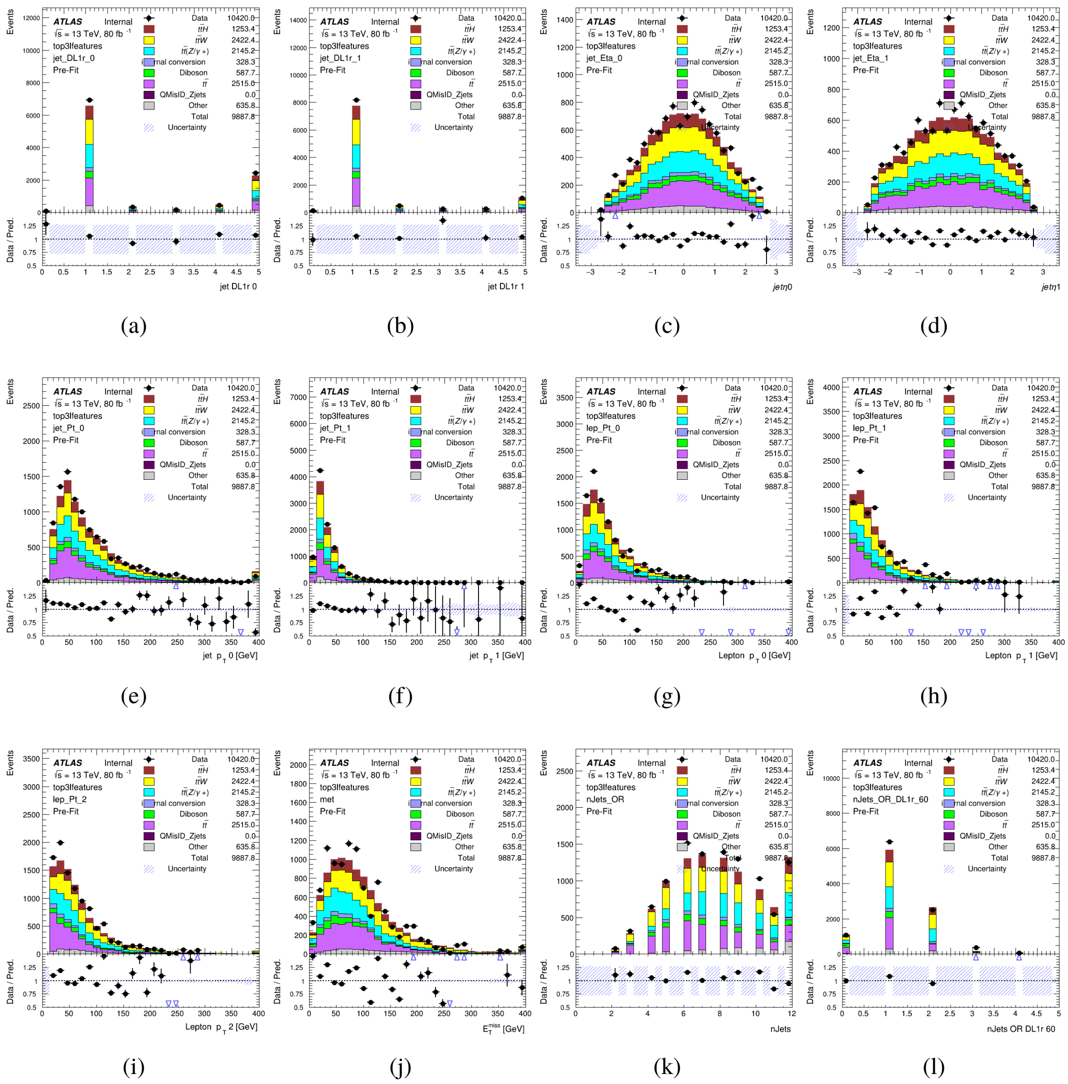
(i) (j) (k) (l)

Figure A.4: Input features for top3l



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

Figure A.5: Input features for top3l

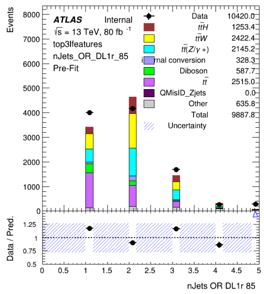


(a) (b) (c) (d)

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

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

Figure A.6: Input features for top3l



(a)

Figure A.7: Input features for top3l

A.1.3 Higgs Reconstruction Features - 2lSS

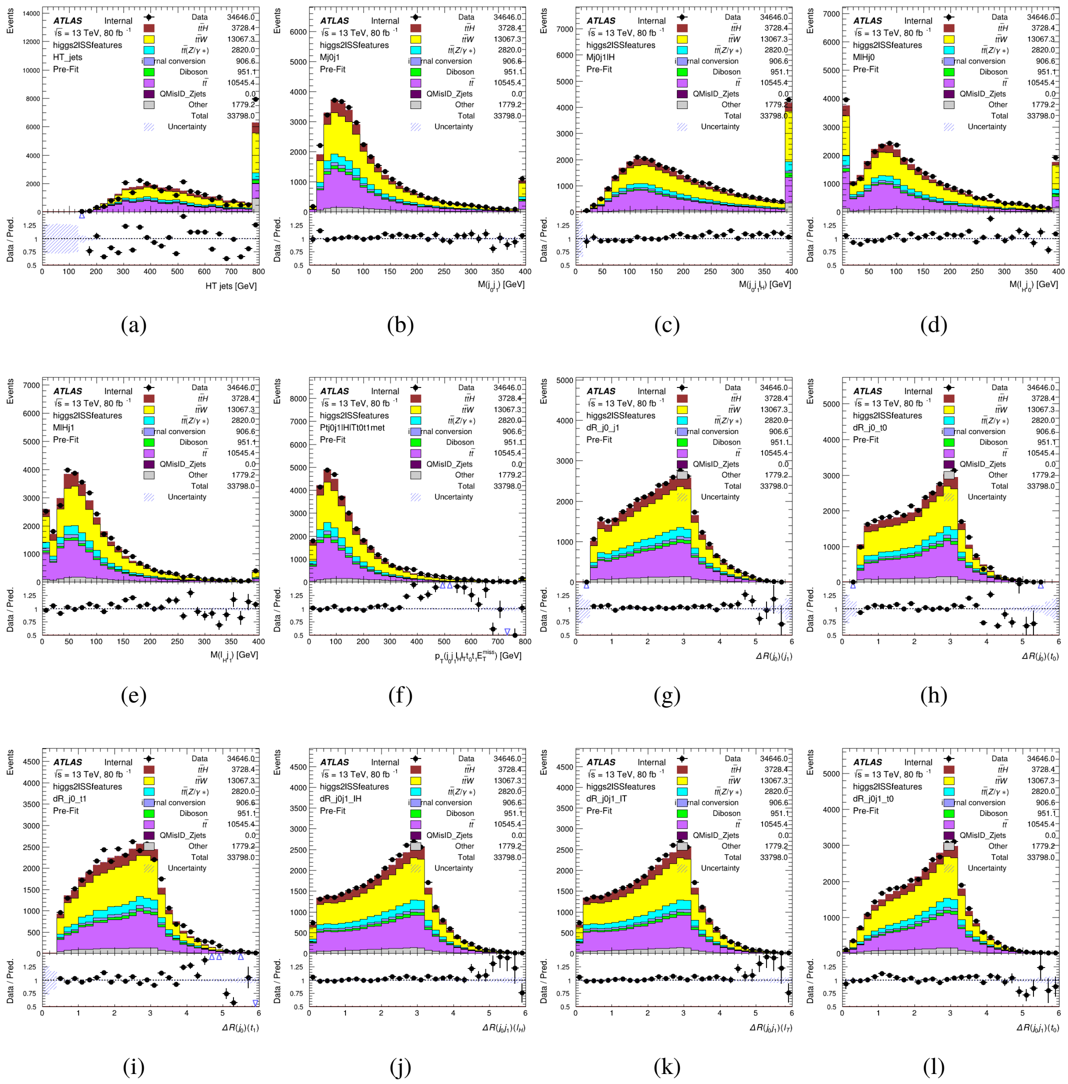


Figure A.8: Input features for higgs2lSS

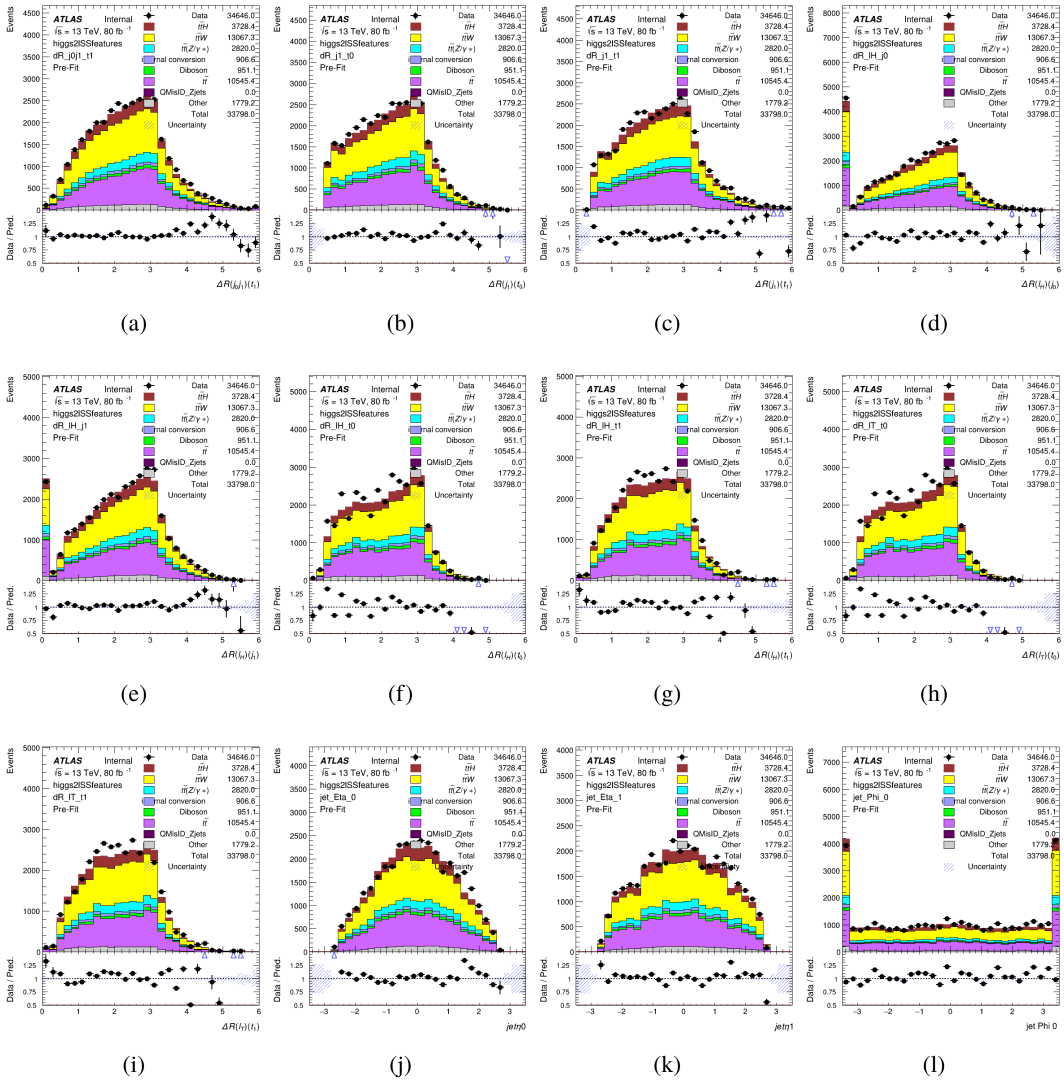


Figure A.9: Input features for higgs2lSS

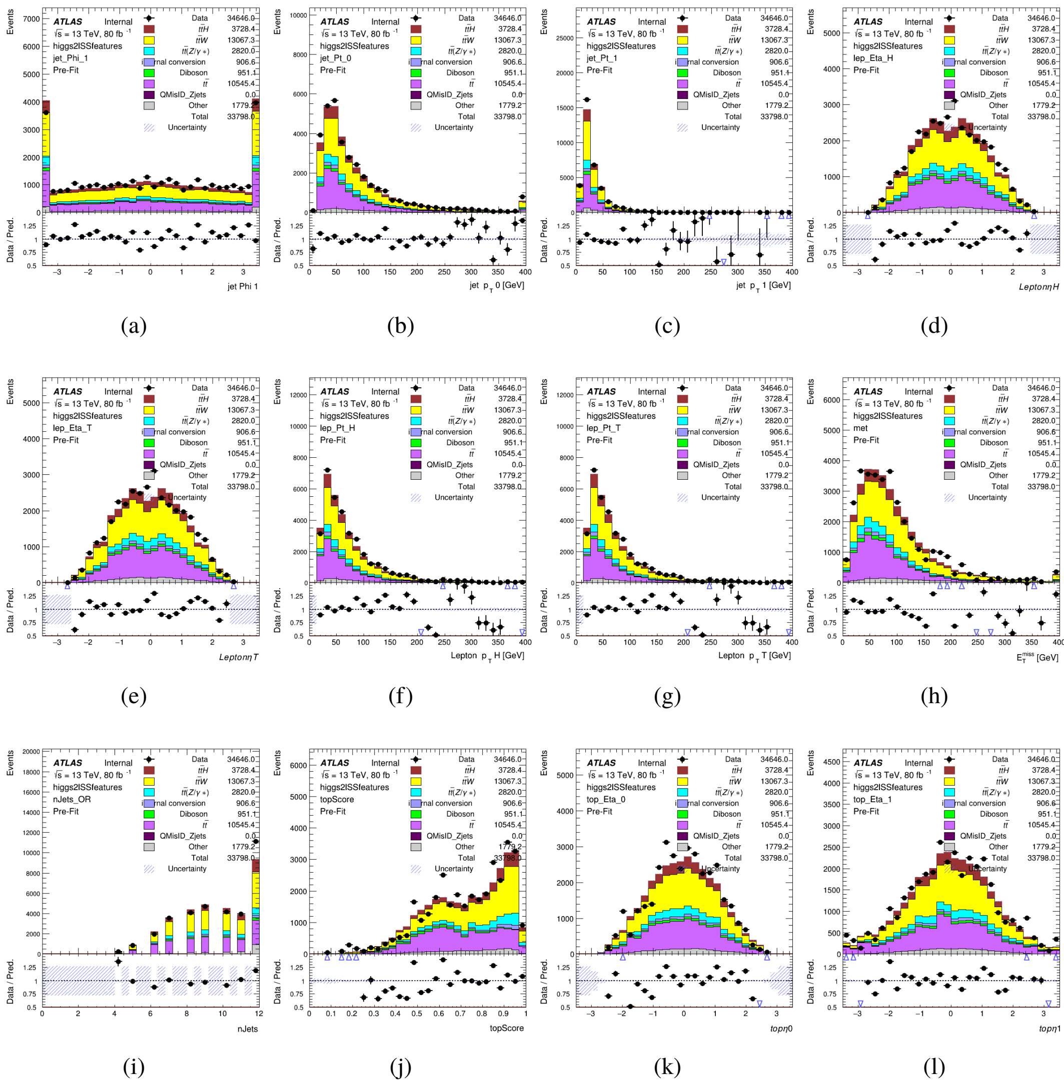
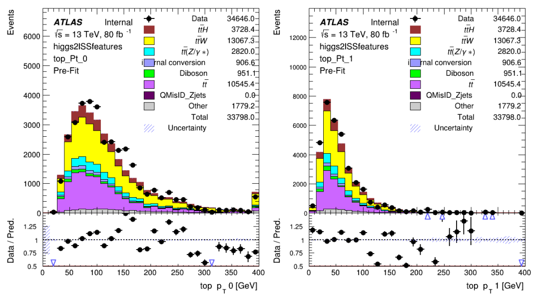


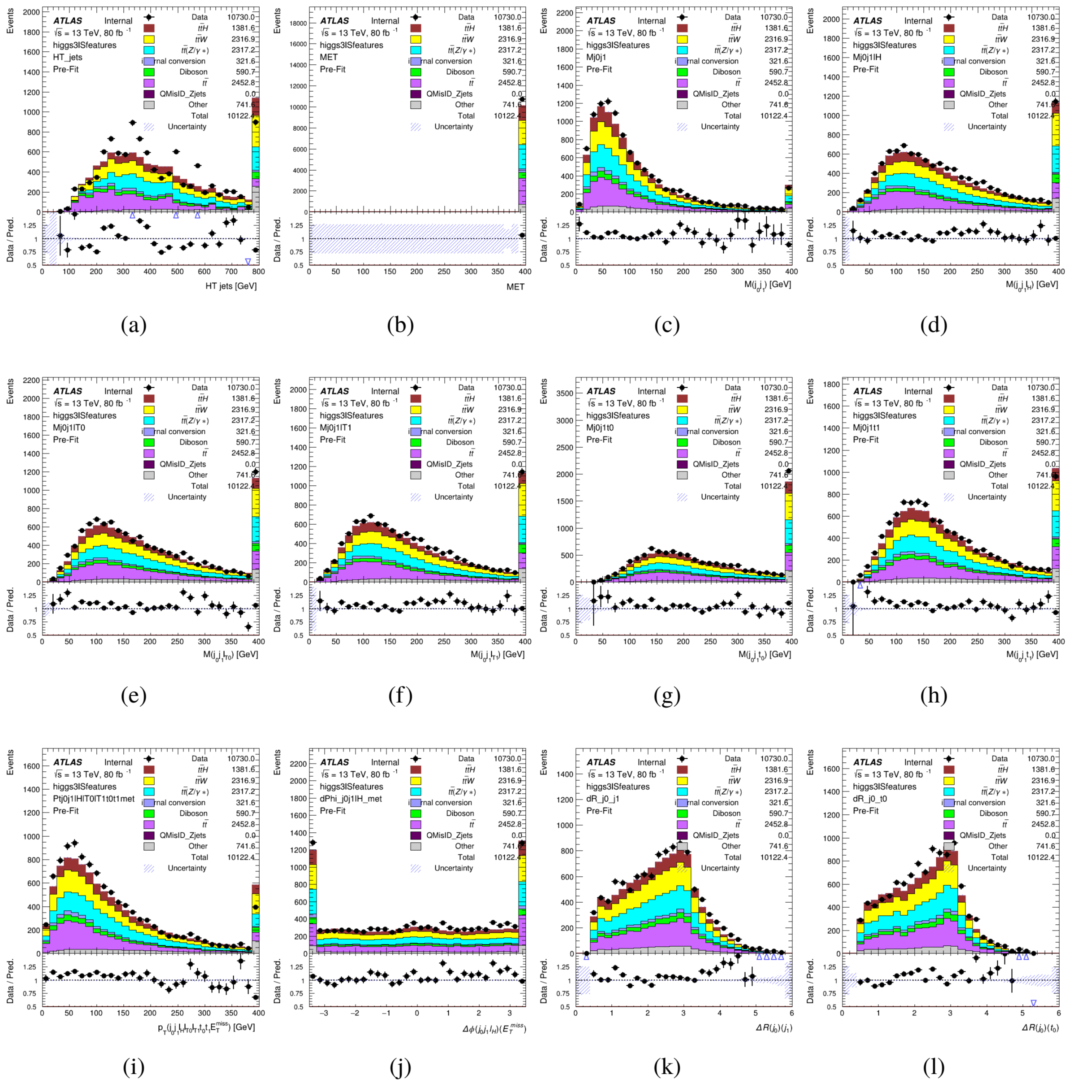
Figure A.10: Input features for higgs2lSS



(a) (b)

Figure A.11: Input features for higgs2lSS

A.1.4 Higgs Reconstruction Features - 3lS

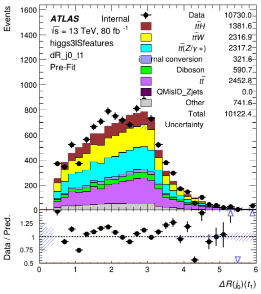


(a) (b) (c) (d)

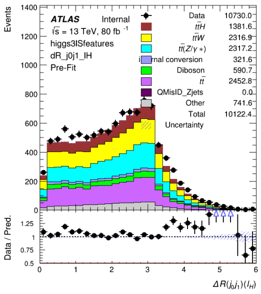
(e) (f) (g) (h)

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

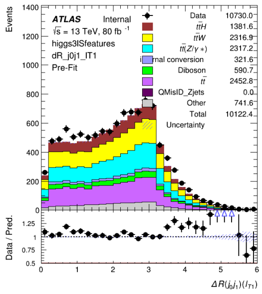
Figure A.12: Input features for higgs3lS



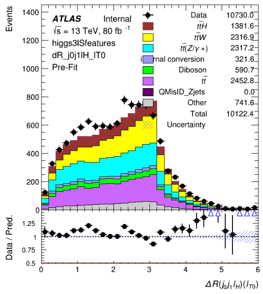
(a)



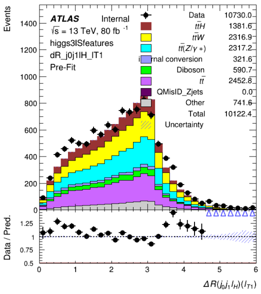
(b)



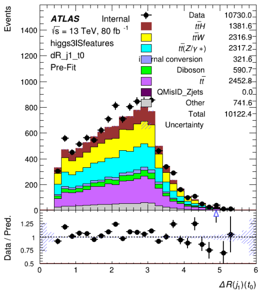
(c)



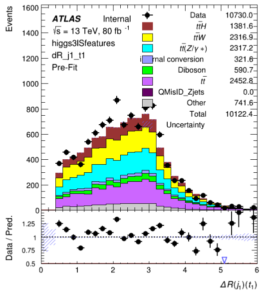
(d)



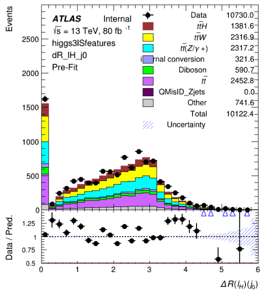
(e)



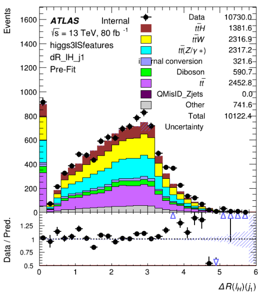
(f)



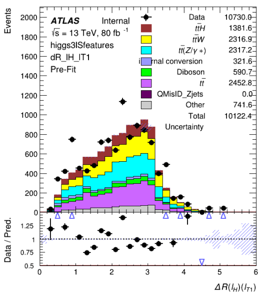
(g)



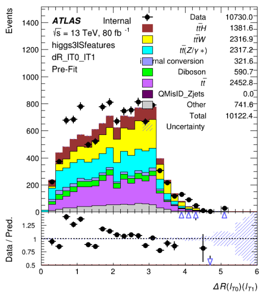
(h)



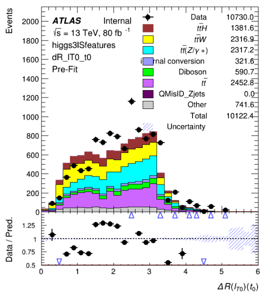
(i)



(j)



(k)



(l)

Figure A.13: Input features for higgs3lS

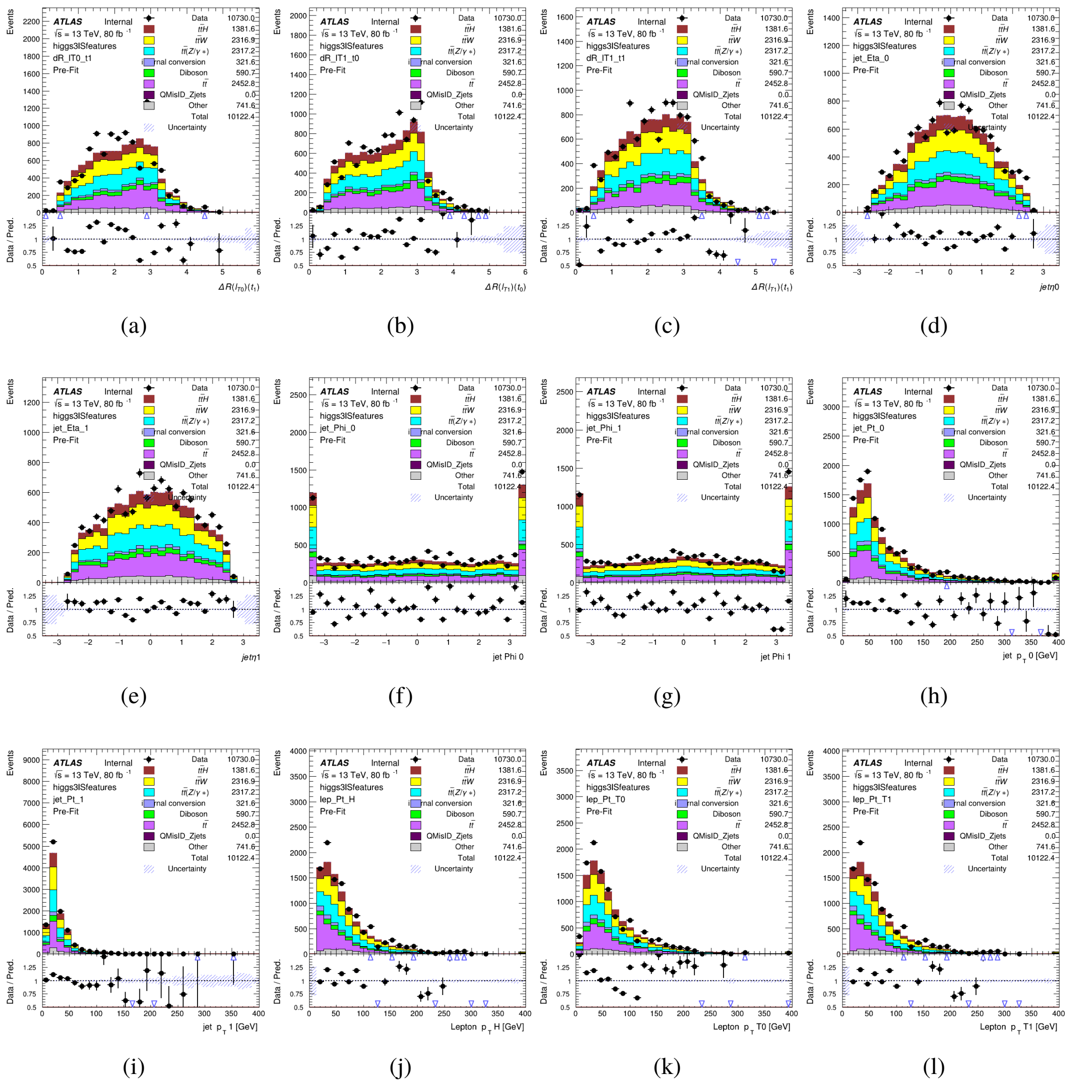
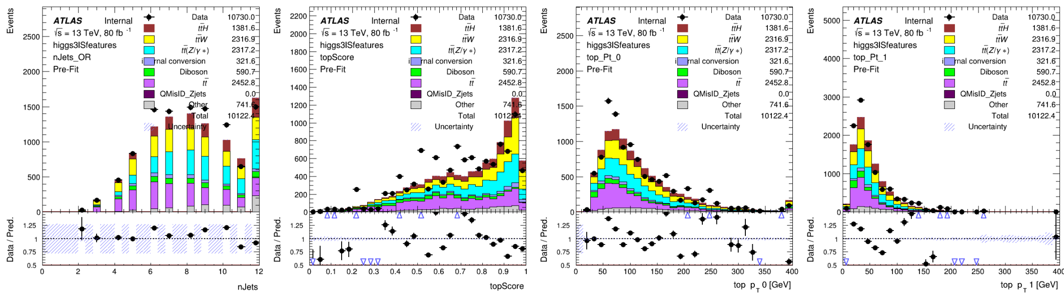


Figure A.14: Input features for higgs3lS



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

Figure A.15: Input features for higgs3lS

A.1.5 Higgs Reconstruction Features - 3lF

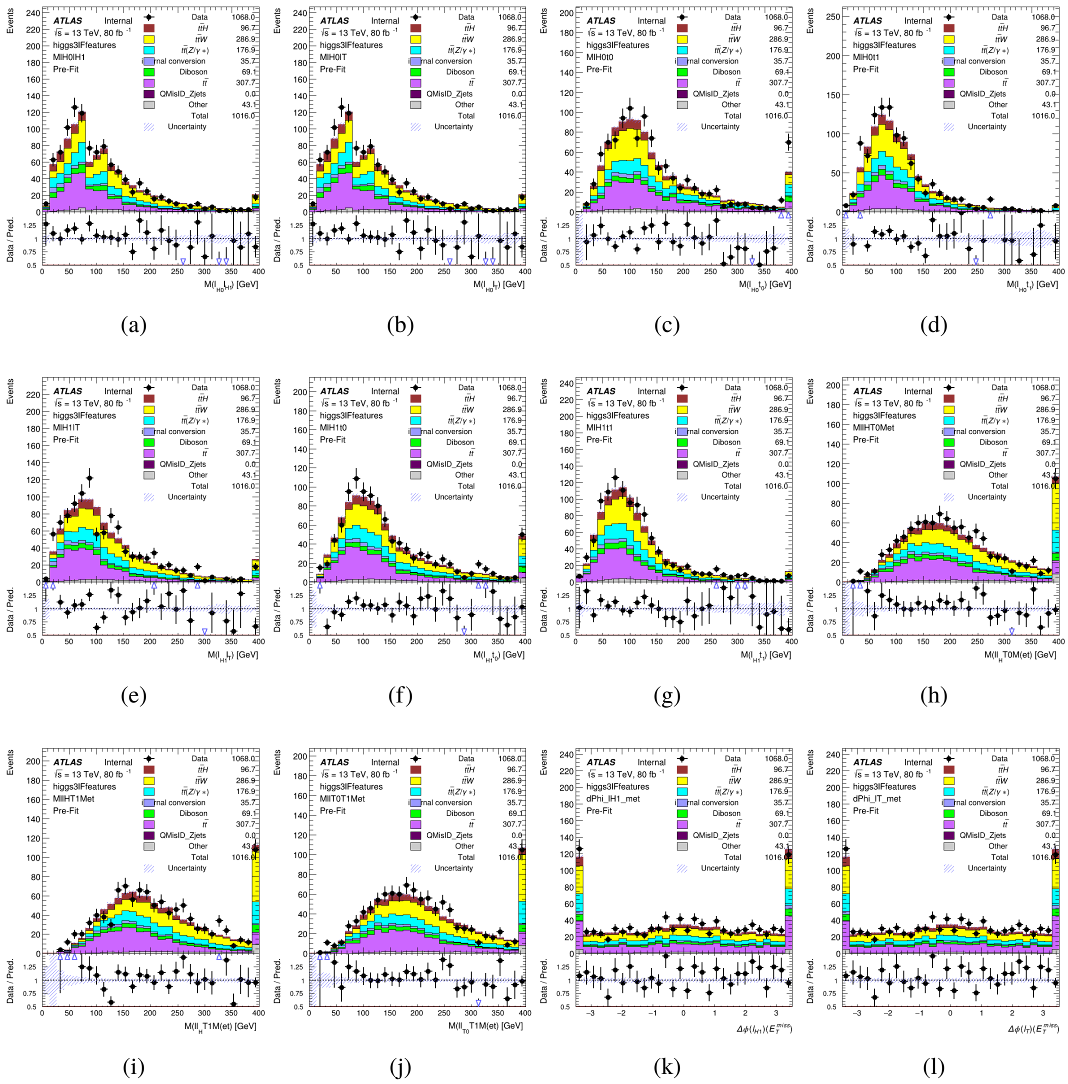
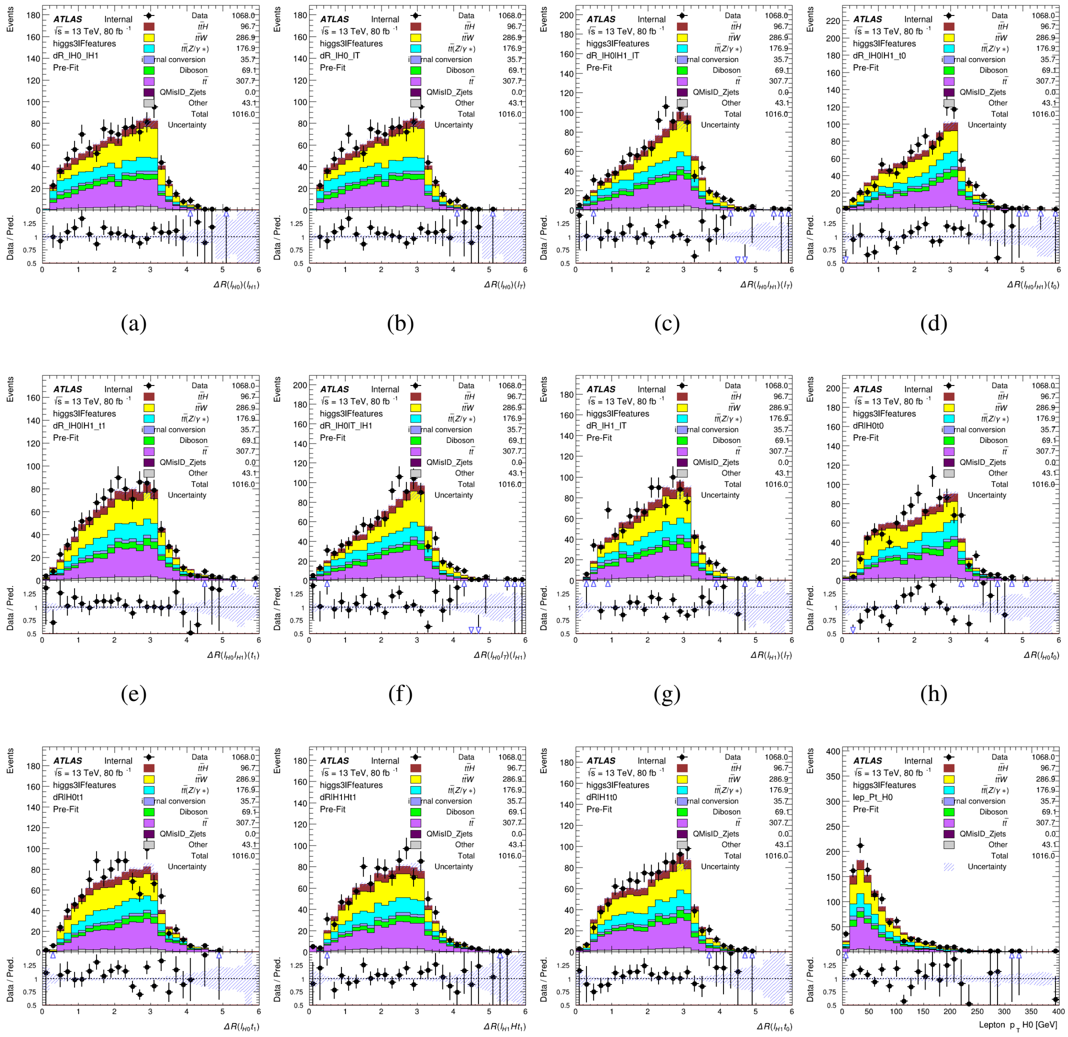


Figure A.16: Input features for higgs3lF

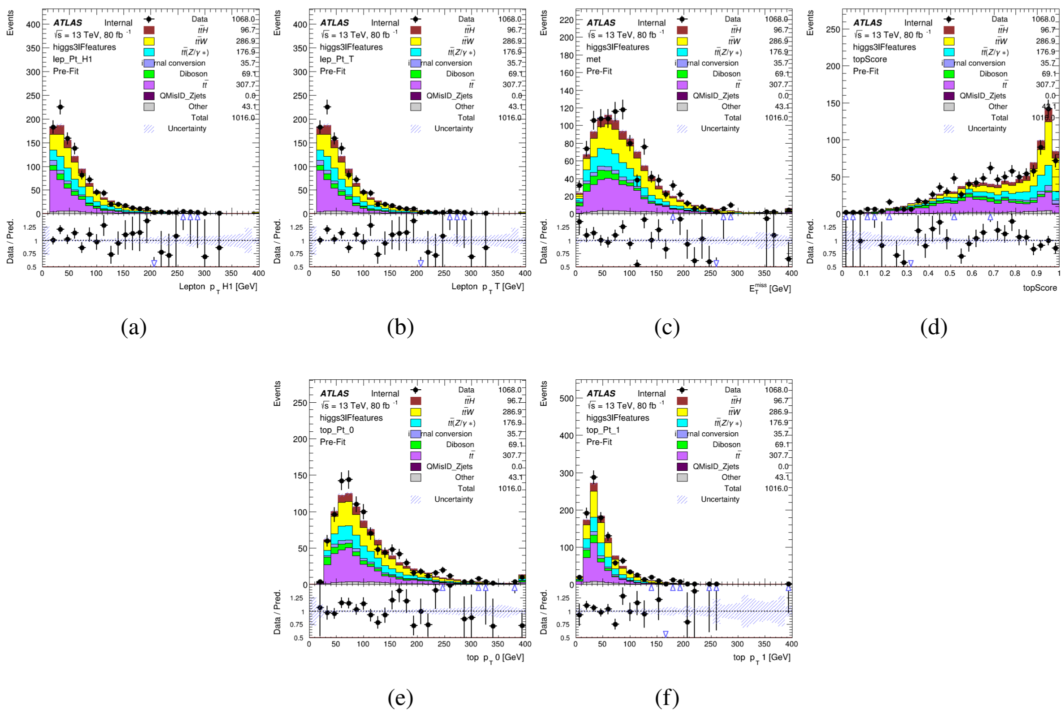


(a) (b) (c) (d)

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

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

Figure A.17: Input features for higgs3lF



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

(e) (f)

Figure A.18: Input features for higgs3lF

A.2 Background Rejection MVAs

A.2.1 Background Rejection MVA Features - 2lSS

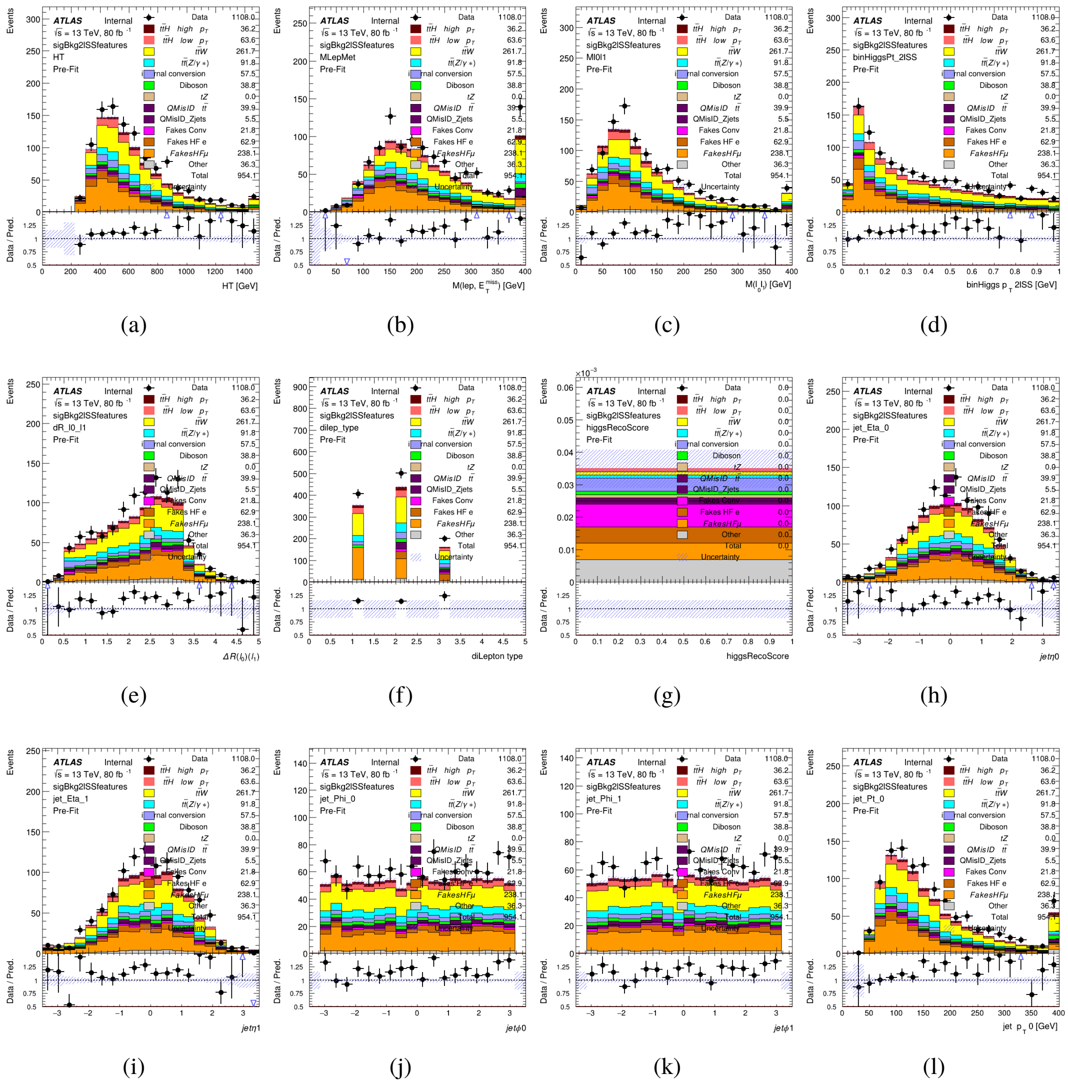
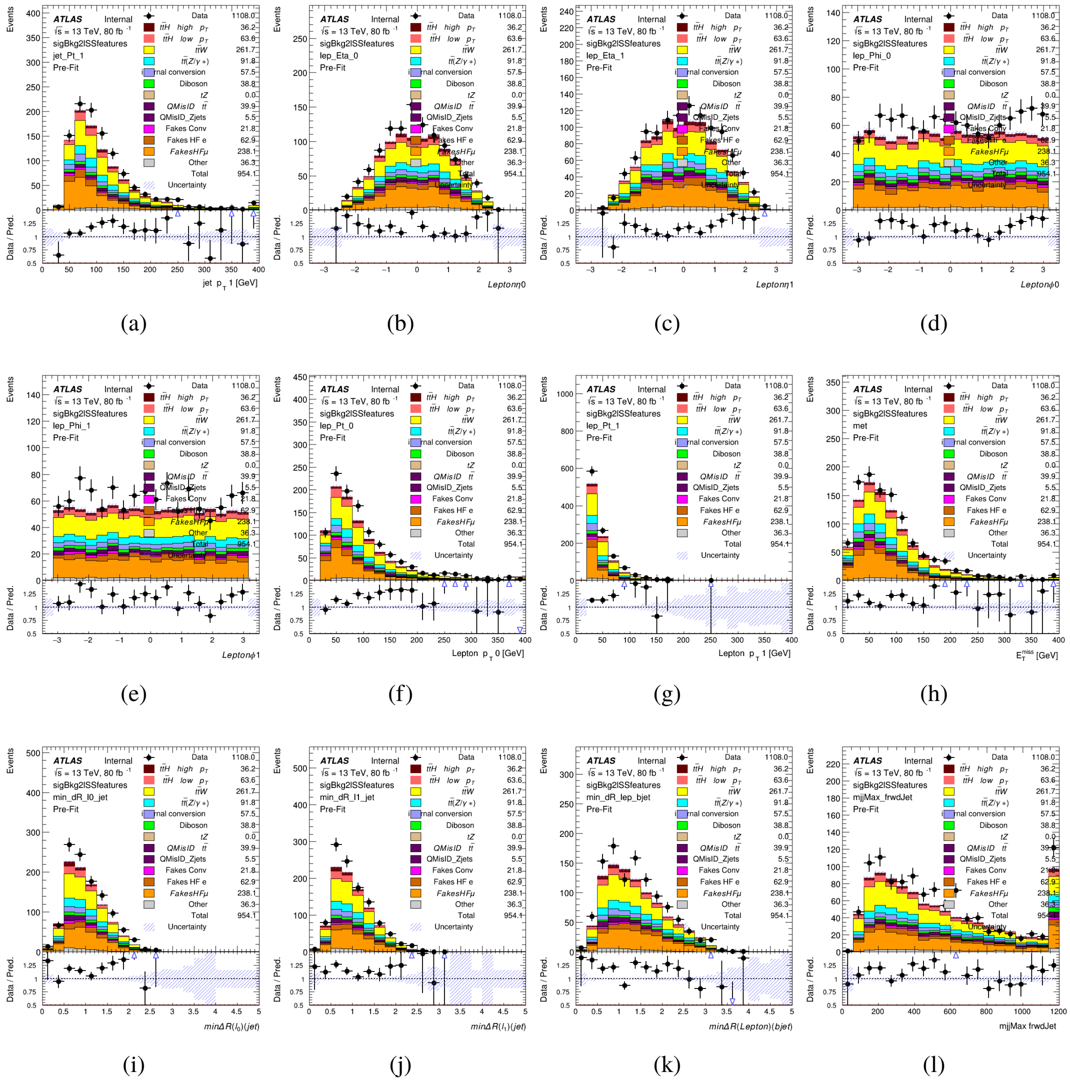


Figure A.19: Input features for sigBkg2lSS



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

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

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

Figure A.20: Input features for sigBkg2lSS

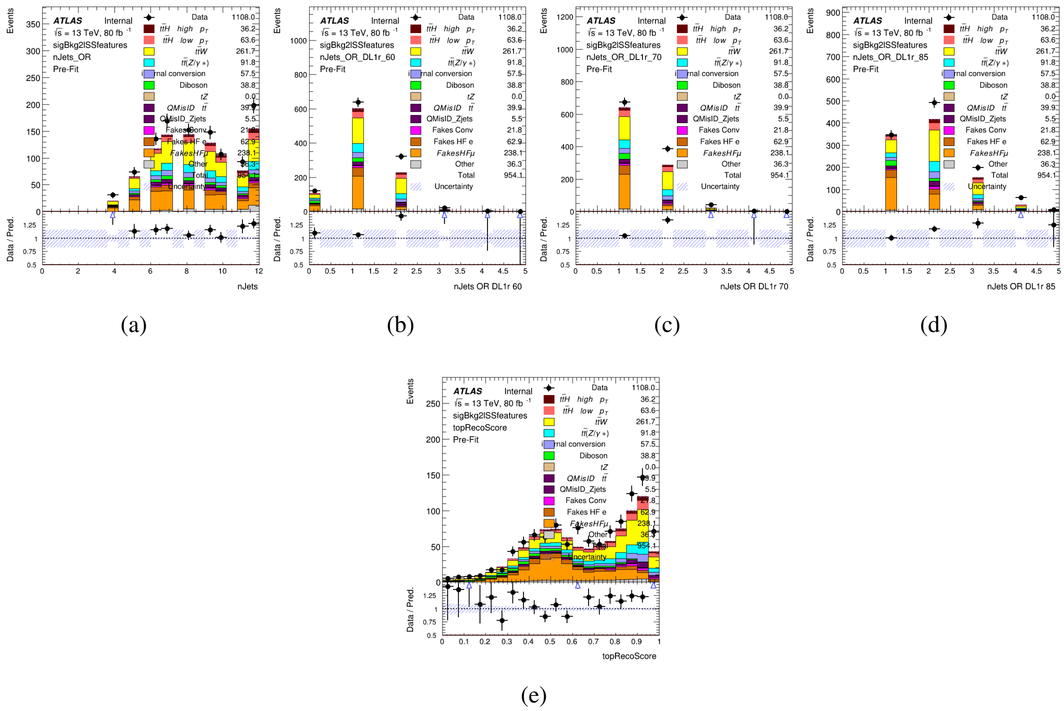


Figure A.21: Input features for sigBkg2lSS

A.2.2 Background Rejection MVA Features - 3l

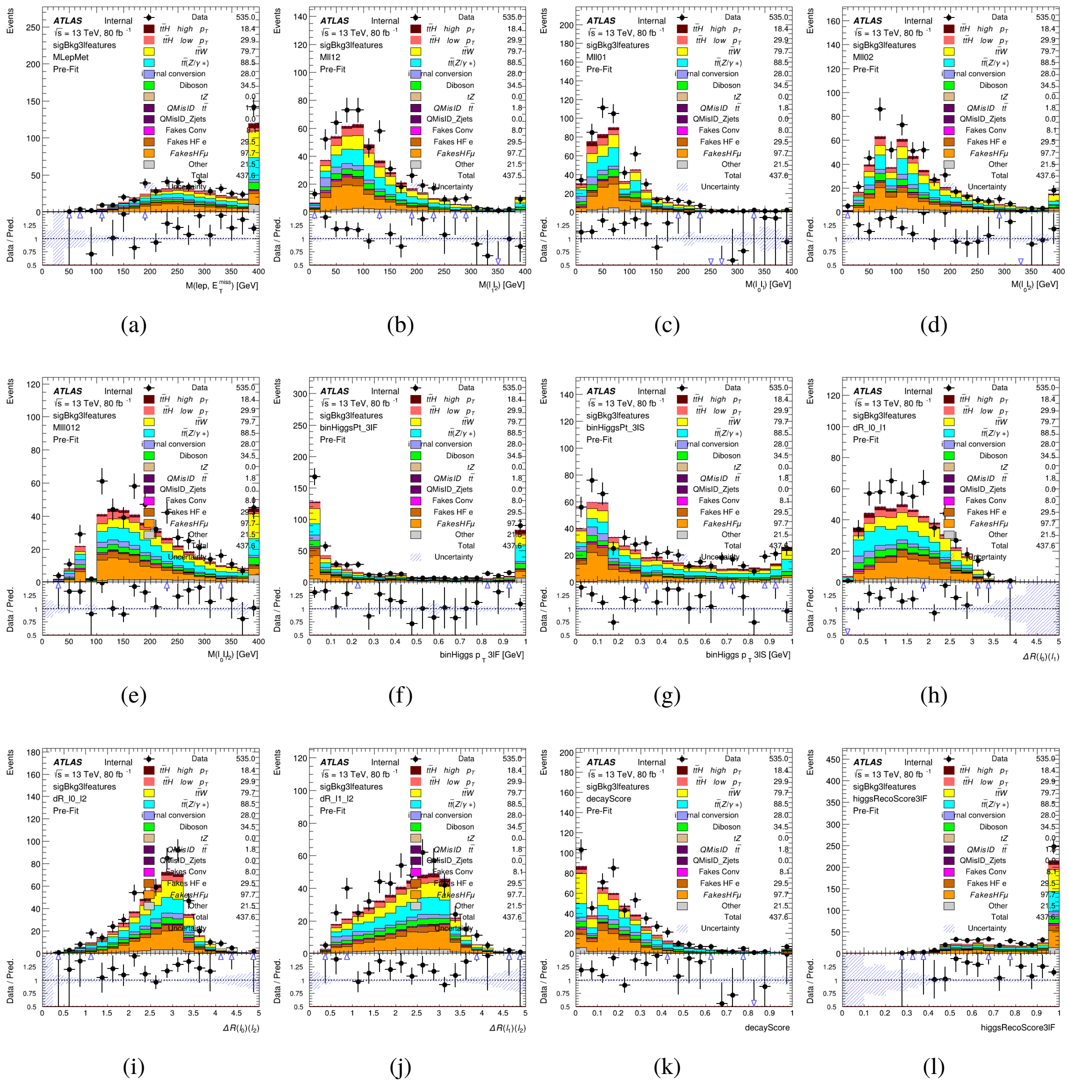


Figure A.22: Input features for sigBkg3l

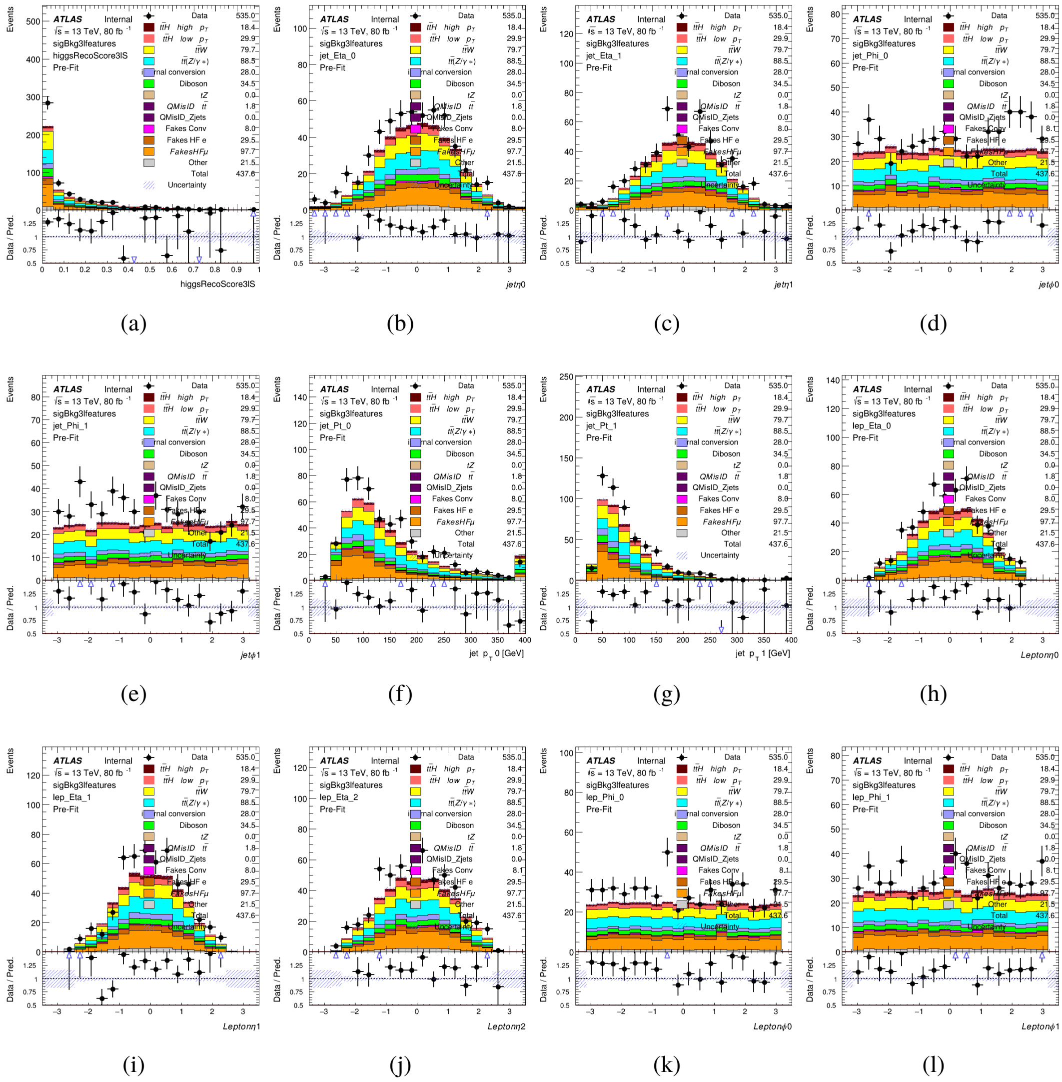


Figure A.23: Input features for sigBkg3l

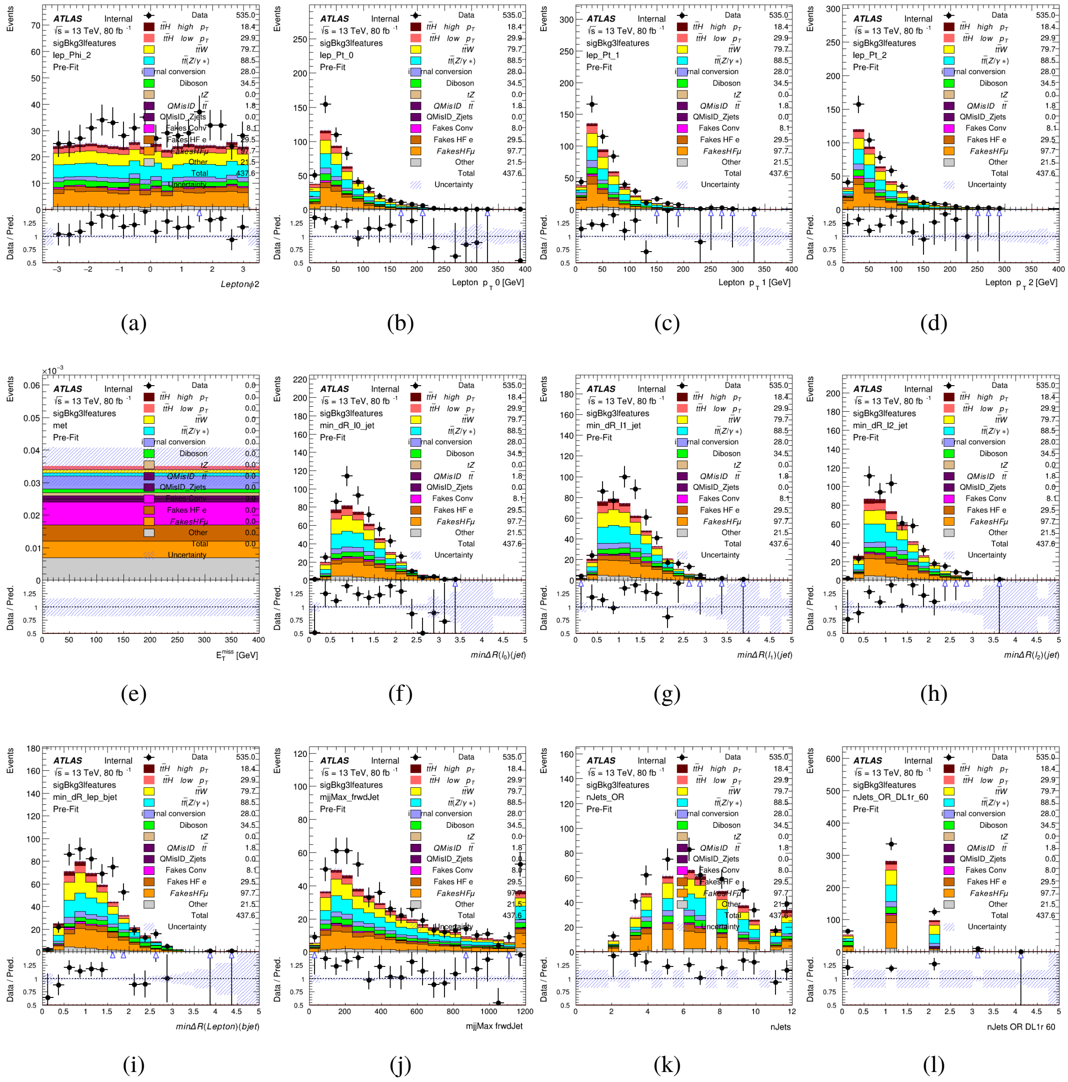
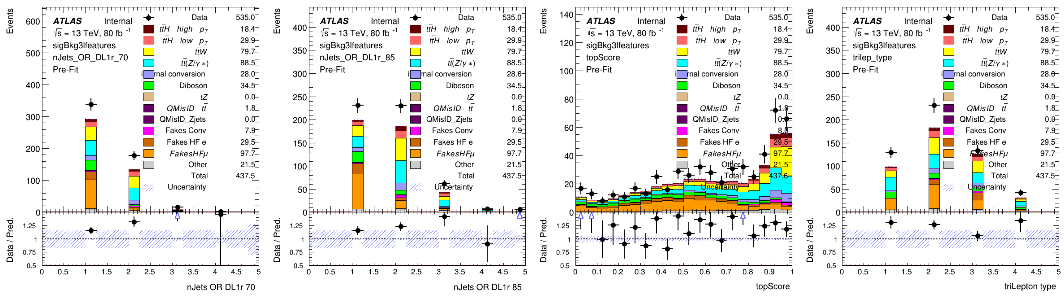


Figure A.24: Input features for sigBkg3l



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

Figure A.25: Input features for sigBkg3l

A.3 Alternate b-jet Identification Algorithm

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

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

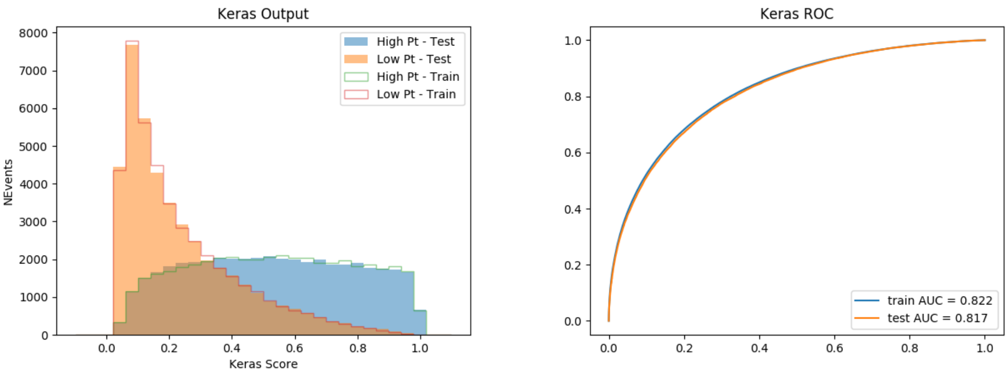
Table 10: Accuracey of the NN in identifying b-jets from tops in 2lSS events for the alternate categorical method compared to the nominal method.

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

A.4 Binary Classification of the Higgs p_T

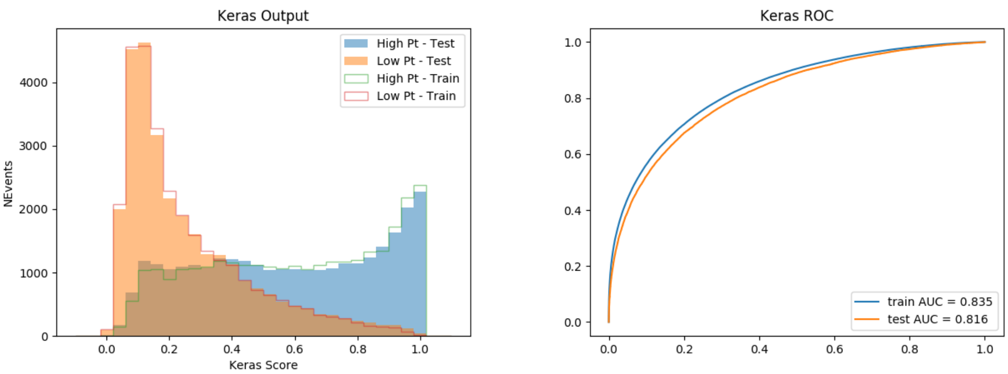
A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer resolution. This means separating high and low p_T events is sufficient for this analysis. As such, rather than attempting the reconstruct the full Higgs p_T spectrum, a binary classification approach is explored.

A model is built to determine whether $t\bar{t}H$ events include a high p_T (>150 GeV) or low p_T (<150 GeV) Higgs Boson. While this is now a classification model, it uses the same input features described in section 8.4. Binary crossentropy is used as the loss function.



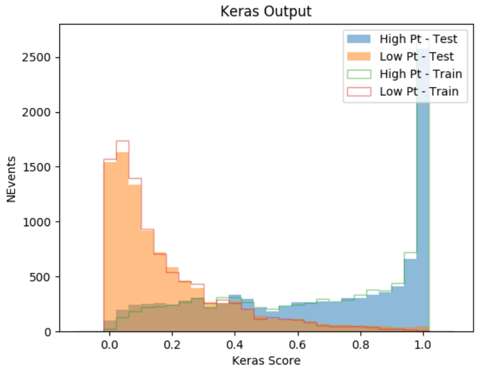
(a) (b)

Figure A.26:

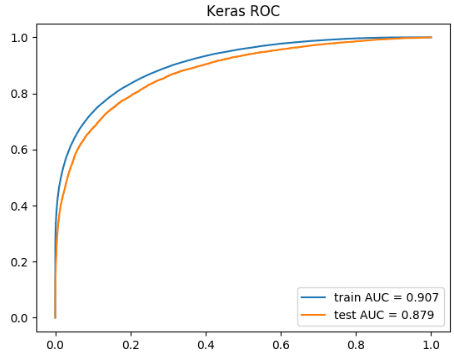


(a) (b)

Figure A.27:



(a)



(b)

Figure A.28:

A.5 Impact of Alternative Jet Selection

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

B

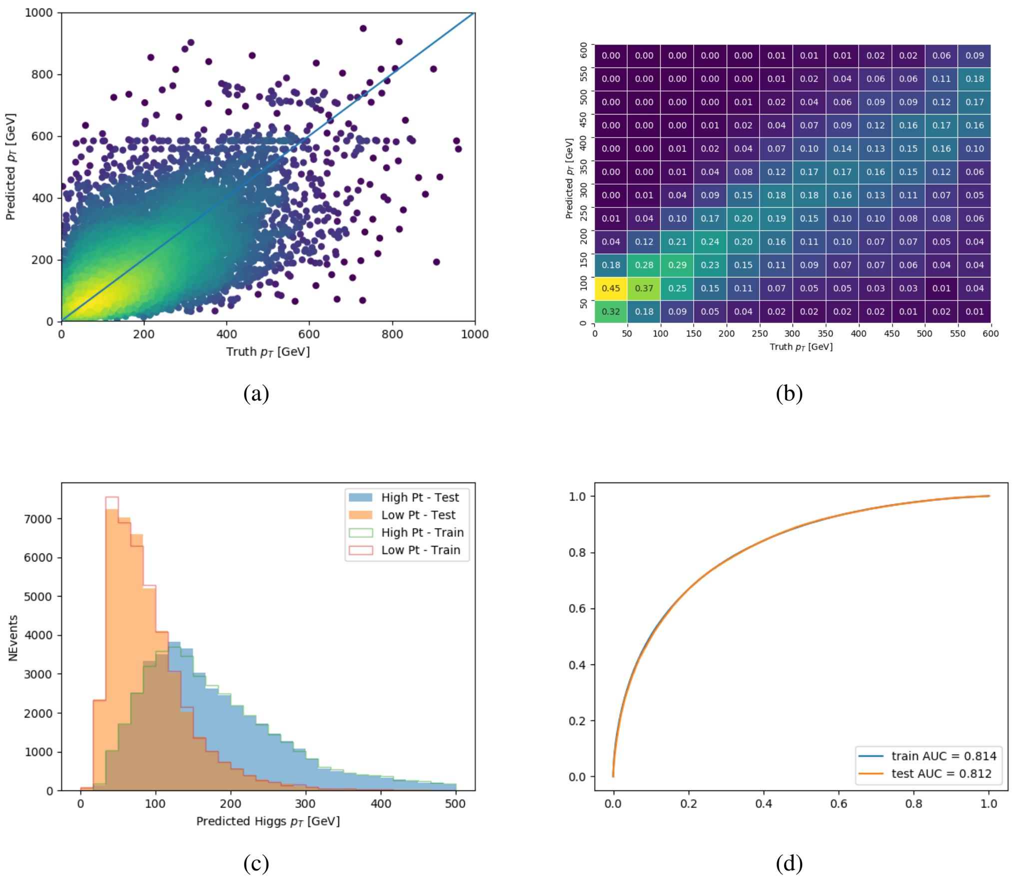

Predicted p_T [GeV]
Truth p_T [GeV]
(a)
Predicted p_T [GeV]
Truth p_T [GeV]
(b)
High Pt - Test
Low Pt - Test
High Pt - Train
Low Pt - Train
NEvents
Predicted Higgs p_T [GeV]
(c)
train AUC = 0.814
test AUC = 0.812
(d)

Figure A.29: