ATLAS Note

GROUP-2017-XX

3rd June 2021

Draft version 0.1

A Deep Learning Approach to Differential Measurements of Higgs - Top Interactions in Multilepton Final States using the ATLAS Detector at the LHC

The ATLAS Collaboration

Several theories Beyond the Standard Model predict a modification of the momentum spectrum of the Higgs Boson, without a significantly altered rate of Higgs produced in association with top quark pairs ($t\bar{t}H$). This provides a physical observable that can be used to search for new physics based on data collected by the LHC. This thesis presents techniques and preliminary results for a differential measurement of the Higgs transverse momentum in $t\bar{t}H$ events with multiple leptons in the final state, using data collected at an energy of $\sqrt{s}$ = 13 TeV by the ATLAS detector at the LHC.

Because of the challenges inherent in reconstructing the Higgs in multilepton final states, a deep learning approach is used to predict of the Higgs. The regressed Higgs p_T spectrum is fit to data for events with two same-sign leptons and three leptons in the final state, in order to extract normalization factors for high ($p_T(H) > 150$ GeV) and low ($p_T(H) < 150$ GeV) momentum $t\bar{t}H$ events. Preliminary results are presented for 80 fb^{-1} of data, with projected results shown for 140 fb^{-1}.

This thesis also details a measurement of WZ + heavy flavor production, a significant background to $t\bar{t}H$ that is poorly understood. This study targets events with three leptons and one or two jets in the final state, using 140 fb^{-1} of $\sqrt{s}$ = 13 TeV data. A measured cross-section of X ± X fb (X ± X fb) is observed for WZ + b (WZ + c) with 1 associated jet and X ± X fb (X ± X fb) for WZ + b (WZ + c) with 2 assoicated jets.

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 [1].

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.

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 in order to search for evidence of these BSM effects.

An additional study of WZ produced in association with a heavy flavor jet (including both b-jets and charm jets) is also included. This process mimics the final state of $t\bar{t}H$ multilpeton events, making it an irreducible background for that analysis. However, this process is poorly understood, and difficult to simulate accurately, introducing large systematic uncertainties for analyses that include it as a background. A measurement of WZ + heavy flavor in the fully leptonic decay mode is performed in an attempt to reduce this uncertainty.

This dissertation begins with a brief explanation of the SM, its limitations, and the theoretical motivation behind this workin Part II. This is followed by a description of the LHC and the ATLAS detector in Part III. Part IV details a measurement of WZ + heavy flavor. Studies of differential measurements of $t\bar{t}H$ are then described in Part V, and preliminary results are presented. Finally, the results of these studies are summarized in the conclusion, Part VI.

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.

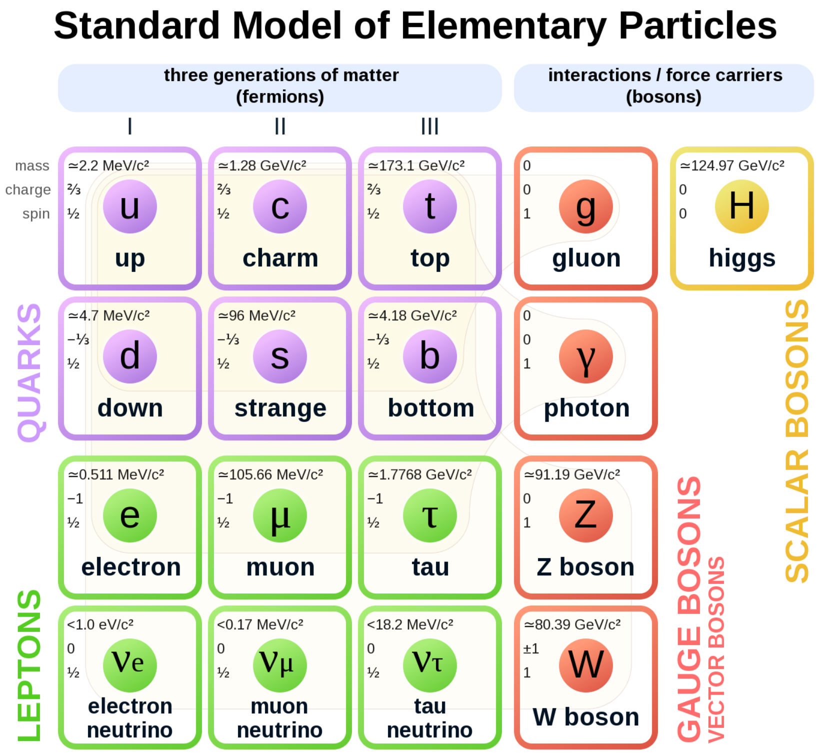


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. [2]

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 which 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, by contrast, 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 this 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 [3]. 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 [4].

2.2.1 The Higgs Field

The Higgs mechanism introduces a complex scalar $\mathrm{SU}(2)$ 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.

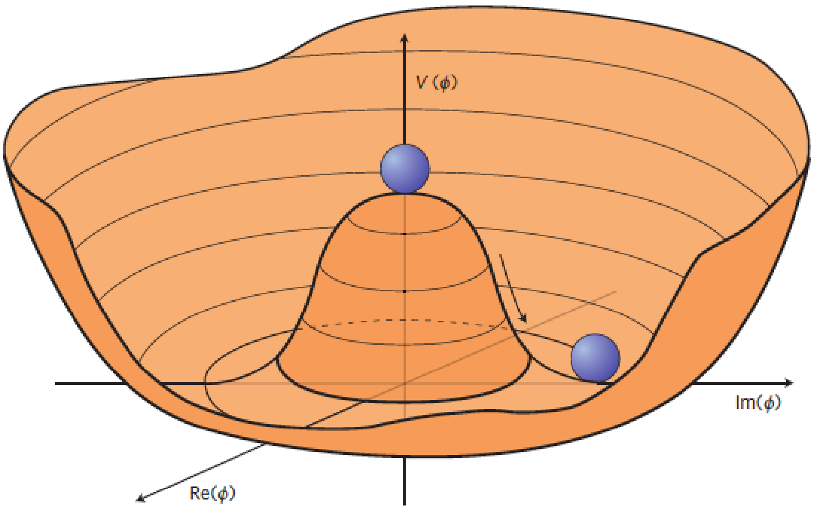


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ, for the case that $\mu^2 < 0$ [5].

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 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:

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

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_μ^1, W_μ^2, W_μ^3, and B_μ:

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

And the masses of these fields are given by:

$$\begin{aligned} M_W^2 &= \frac{1}{4}g^2\nu^2 \\ M_Z^2 &= \frac{1}{4}(g^2+g'^2)\nu^2 \\ M_A^2 &= 0 \end{aligned} \tag{2.10}$$

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 an excitation 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 fermions' Yakawa couplings to the VEV take the same form as the fermionss 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 $t\bar{t}H$ Production

The strength of a particles interaction with the Higgs, given by its Yakawa coupling, is proportionate to its mass. The top quark - as the heaviest known particle - has the strongest interaction, making this interaction particularly interesting to study. While several processes involve interactions between the Higgs and the top, some Higgs production modes include the top interaction only as a part of a loop diagram, such as the gluon-gluon fusion diagram shown in Figure 2.3. This process therefore only allows for an indirect probe of the Higgs-top Yakawa coupling, as the flavor of the quark in this diagram is not unique.

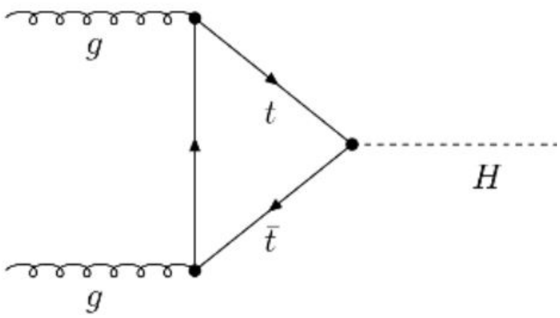


Figure 2.3: Diagram of a Higgs boson produced via gluon-gluon fusion.

Studying the Higgs produced in association with top quark pairs, $t\bar{t}H$, allows this interaction to be measured directly. This process, as shown in Figure 2.4, involves a unique coupling between the Higgs and the top, which can be identified by the top quark pair in the final state.

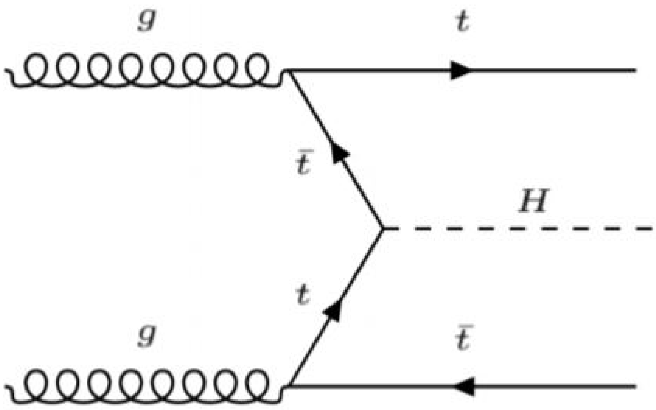


Figure 2.4: Diagram of a Higgs boson produced in association with a pair of top quarks.

The Higgs boson, as well as the top quarks, have very short lifetimes - on the order of 10^{-22} s and 10^{-25} s respectively - meaning they can only be observed via their decay products. Measuring this process is therefore a matter of identifying events with final states consistent with $t\bar{t}H$ production.

Studies of $t\bar{t}H$ production have been reported by the ATLAS collaboration for $H \to b\bar{b}$, $H \to \gamma\gamma$ and multilepton (encompassing $H \to W^+W^-$, $H \to ZZ$ and $H \to \tau^-\tau^+$, with $H \to ZZ \to 4l$ as a separate analysis) decay modes. While the branching ratio of $H \to W^+W^-$ is smaller than $H \to b\bar{b}$ (see Table 2.3), it produces a clearer signal, as $H \to b\bar{b}$ suffers from large $t\bar{t}$ backgrounds. On the other hand, $H \to \gamma\gamma$ produces the most easily identifiable signal, but has a much smaller branching ratio than $H \to W^+W^-$. Therefore, compared with other final states of $t\bar{t}H$, the $t\bar{t}H - ML$ channel is an attractive candidate for study, as it involves a good balance between statistical power and identifiability.

Decay Mode	Branching Ratio (%)
$H \to b\bar{b}$	58.2
$H \to WW^*$	21.4
$H \to gg$	8.19
$H \to \tau\tau$	6.27
$H \to c\bar{c}$	2.89
$H \to ZZ^*$	2.62
$H \to \gamma\gamma$	0.227

Table 1: Summary of the predominant SM Higgs ($m_H = 125$ GeV) branching ratios. Particles with a star imply off-shell decays.

Searches for $t\bar{t}H$ production typically target a measurement of the signal strength parameter, $\mu_{t\bar{t}H}$, which measures the ratio of the observed cross-section and the expected cross-section according to the SM.

$$\mu_{t\bar{t}H} = \frac{\sigma^{obs.}_{t\bar{t}H}}{\sigma^{SM}_{t\bar{t}H}} \tag{2.12}$$

$t\bar{t}H$ production was observed by ATLAS using up to 79.8 fb^{-1} of data collected at $\sqrt{s}$ = 13 TeV, based on a combination of five Higgs decay modes: $b\bar{b}$, WW^*, $\tau^-\tau^+$, $\gamma\gamma$, and ZZ^* [6]. A significance of 5.8σ was observed, compared to a 4.9σ expected significance. Since then, two analyses have published updated results ($H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$) with the full Run 2 dataset, representing 139 fb^{-1}. Studies are still ongoing in the remaining channels.

This thesis focuses on $t\bar{t}H$ events with multiple leptons in the final state, $t\bar{t}H - ML$, specifically targeting events with two same-sign leptons ($2lSS$) or three leptons ($3l$) in the final state. This includes $\mathrm{H} \to W^{+}W^{-}$ events, where at least one of the W bosons decays leptonically.

2.4 WZ + Heavy Flavor Production

Part IV is dedicated to a measurement of WZ produced in association with a heavy flavor jet - namely, a charm or b-jet - in the fully leptonic channel. In the instance that both the W and Z bosons decay leptonically, this process produces a final state similar to $\mathrm{t\bar{t}H}$, making it an irreducible background for $\mathrm{t\bar{t}H - ML}$ specifically, and any analysis that includes multiple leptons and b-tagged jets in the final state more broadly.

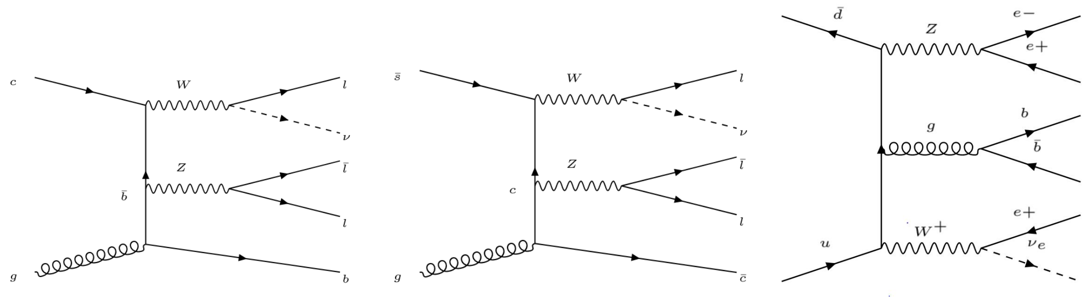


Figure 2.5: Example Feynman diagrams of WZ + heavy flavor production

The b-jets produced in this process can be thought of in two different ways: either as originating from the quark "sea" of the initial state hadrons, or as the result of a gluon from one the colliding protons splitting into $\mathrm{b\bar{b}}$ pairs. However, the heavy flavor contribution to the parton distribution function (PDF) of the proton is uncertain, and simulations of this process disagree depending on which of these two approaches one considers. This makes WZ + heavy flavor difficult to accurately simulate, and introduces a large uncertainty for any analysis which includes it as a background, motivating a measurement of this process.

2.5 Extensions to the Standard Model

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. Further, there are several experimental observations that the SM fails to explain. For example, the SM predicts neutrinos to be massless, despite experimental observation to the contrary, and fails to explain the observation of dark matter and dark energy.

Another example, revelant to the Higgs sector, is known as the hierarchy problem: large quantum corrections to the Higgs mass from loop diagrams, such as those shown in Figure 2.6, are many orders of magnitude larger than the Higgs mass itself. The observed value of the Higgs mass therefore requires extremely precise cancellation between these corrections and the bare mass of the Higgs, a cancellation which seems unnatural and suggests something missing in our theoretical picture.

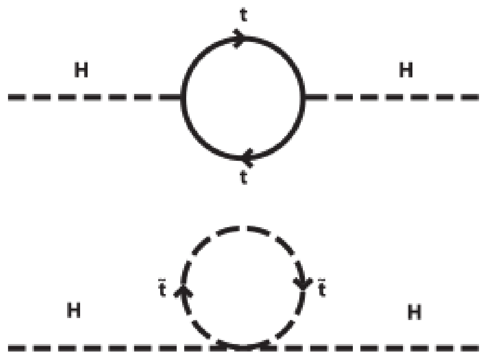


Figure 2.6: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram.

Because so many of the properties of the Higgs boson have not yet been studied, its interactions are a promising place to search for new physics that could resolve some of the limitations of the SM. As explained above, the interactions of the Higgs with the top quark, as in $t\bar{t}H$ production, are particularly interesting: As the most massive particle in the Standard Model, the top quark is the most strongly interacting with the Higgs. Therefore, any new physics effects are likely to be seen most prominently in this interaction.

These interactions can be measured directly by studying the production of a Higgs Boson in association with a pair of Top Quarks ($t\bar{t}H$). While this process has been observed by both the ATLAS [7] and CMS [8] collaborations, these analyses have focused on measuring the overall rate of $t\bar{t}H$ production. There are several theories of physics Beyond the Standard Model (BSM), however, that would affect the kinematics of $t\bar{t}H$ production without altering its overall rate [9].

An Effective Field Theory approach can be used to model the low energy effects of new, high energy physics, by paramaterizing BSM effects as higher dimensional operators. These additional operators can then be added to the SM Lagrangian to write an effective Lagrangian that accounts for the effects of these higher energy physics. The lowest order of these that could contribute to Higgs-top coupings are dimension-six, as represented in Equation 2.13.

$$\mathcal{L}_{\mathrm{eff}} = \mathcal{L}_{\mathrm{SM}} + \frac{\mathrm{f}}{\Lambda}\mathcal{O}^{6} \tag{2.13}$$

Here Λ represents the energy scale of the new physics, and f is a Wilson coefficient which represents the strength of the effective coupling. An experimental observation of any non-zero value of f would be a sign of BSM physics.

The addition of these operators can be shown to modify the transverse momentum ($\mathrm{p_T}$) spectrum of the Higgs Boson in Higg-top interactions, without effecting the overall rate of $\mathrm{t\bar{t}H}$ production [10]. The possible impact of these higher order effects on the Higgs $\mathrm{p_T}$ spectrum are shown in Figure 2.7.

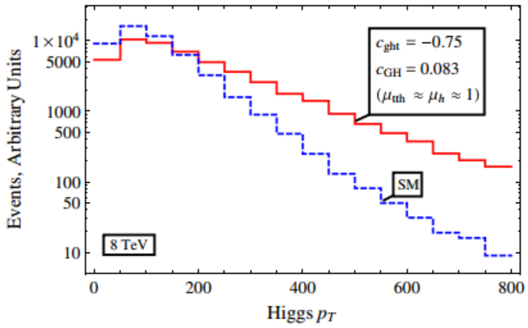
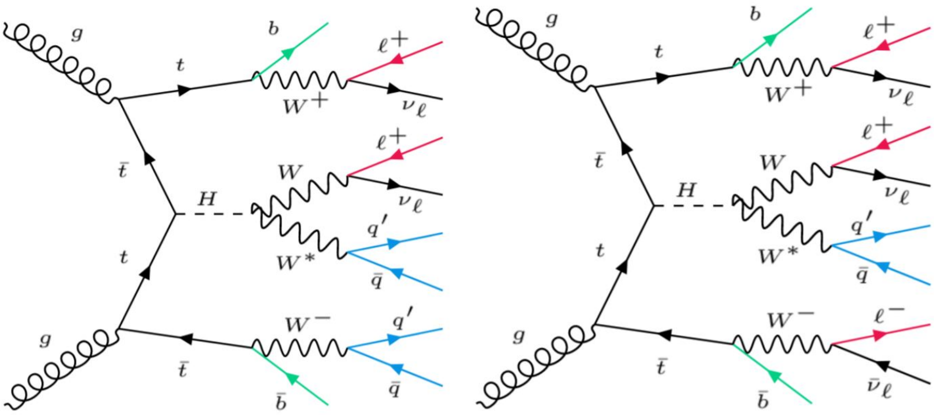


Figure 2.7: The momentum spectrum of the Higgs boson produced via top-quarks with (red) and without (blue) the presence of dimension-six operators.

This provides a clear, physics observable that could be used to search for evidence of BSM physics. The energy and luminosity produced by the LHC now make such a measurement possible. Reconstructing the momentum spectrum of the Higgs in $t\bar{t}H$ events therefore provides a means to search for new physics in the Higgs sector.

Reconstructing the Higgs is a particular challenge in the multilepton channels of $t\bar{t}H$, due to an ambiguity arising from multiple sources of missing energy. In the $H \to \gamma\gamma$ channel, the kinematics of the Higgs can be fully reconstructed from the two photons. The same is true of $H \to b\bar{b}$, though with the additional challenge of identifying which two of the four b-quarks in the final state originated from the Higgs. By contrast, the two channels ($2\ell SS$ and 3ℓ) targeted by this analysis include at least one neutrino originating from the Higgs decay.



(a) (b)

Figure 2.8: Feynman diagrams of $t\bar{t}H$ production with (a) two same-sign leptons and (b) three leptons in the final state.

Neutrinos are not detected by ATLAS; instead, their presence is inferred from missing transverse energy in the detector, E^{T}_{miss}. The two channels targeted here include not just a neutrino from the Higgs decay, but at least one additional neutrino from the decay of the top-quarks. This makes disentangling the contribution of the Higgs decay to E^{T}_{miss}, and thereby fully reconstructing the Higgs, impossible.

This challenge motivates the use of more sophisticated machine learning techniques when attempting to perform differential measurements of the Higgs p_T spectrum in the multi-lepton channels of $t\bar{t}H$.

Part III

The LHC and the ATLAS Detector

3 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 3.1. 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 and 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.

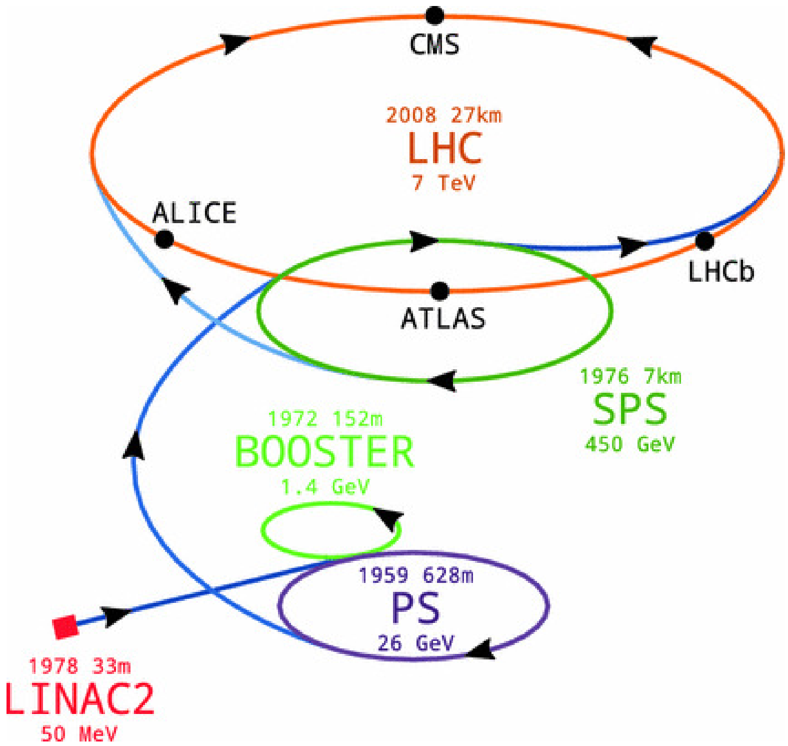


Figure 3.1: A summary of the accelerator chain used to feed protons into the LHC [11].

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.

$$\mathcal{L} = \frac{1}{\sigma}\frac{dN}{dt} \tag{3.1}$$

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

$$\mathcal{L}_{\mathrm{int}} = \int \mathcal{L}\,dt \tag{3.2}$$

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}.

4 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 concentric 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.

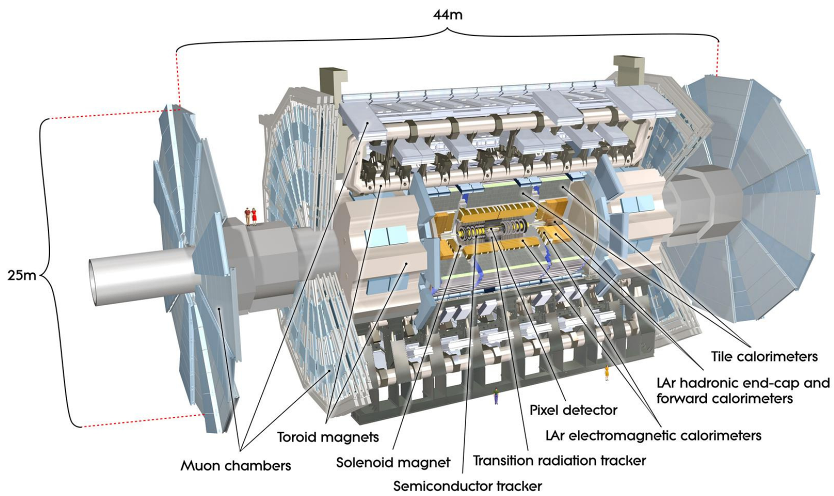


Figure 4.1: Cutaway view of the ATLAS detector, with labels of its major components [12].

4.1 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.

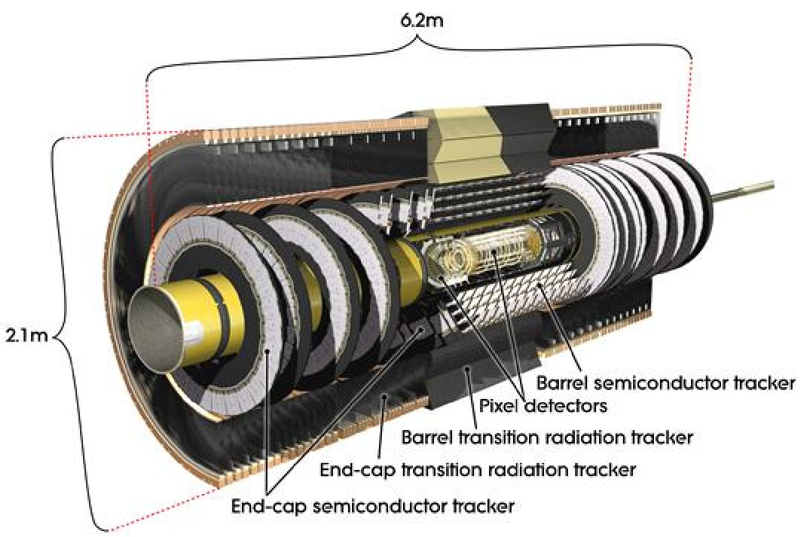


Figure 4.2: Cutaway view of the Inner Detector [13].

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 of silicon rather than small pixels to cover a larger spatial area. It includes over 6 million readout strips, allowing the position of charged particles to be measured to an accuracey of 17 μm.

The outermost component of the inner detector, the TRT consists of around 300,000 straw tubes filled with ionizable gas, which produces current through a wire in the center of each tube when a charged particle passes through. Between these staws are layers of material designed to produce transition radiation from ultrarelativistic particles as they pass through each material boundary, amplifying the signal. The position uncertainty in the TRT is higher than the other two, on the order of 200 μm, but covering a much larger area.

4.2 Calorimeters

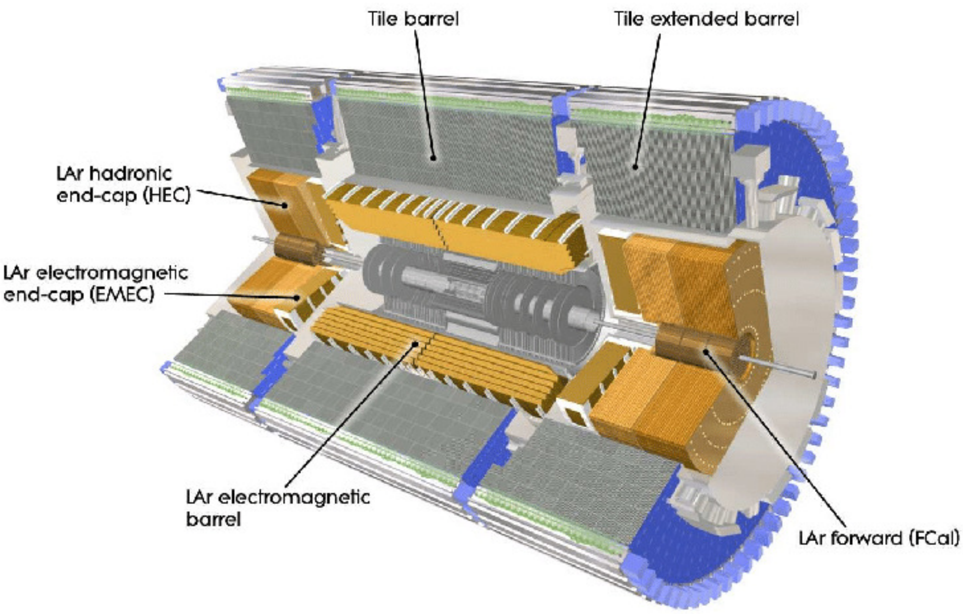


Figure 4.3: Cutaway view of the calorimeter system of the ATLAS detector [13].

Situated outside the Innder Detector are two concentric calorimeters. The inner calorimeter uses liquid argon (LAr) to measure the energy of particles that 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 the 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 designed to cause hadronic showering and around 500,000 scintillating tiles as the active material. The signals from the hadronic calorimeter are read out by photomultiplier tubes (PMTs).

4.3 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 systems. The largest of the subdetectors, 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. The path of the muons are tracked by Monitored Drift Tubes, which are drift chambers formed by aluminum tubes and filled with ionizing gas. These tubes produce a multi-layer spatial resolution on the order of 50 μm.

4.4 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

Measurement of WZ + Heavy Flavor

5 Introduction

The production of WZ in association with a heavy flavor jet represents an important background for many major analyses. This includes any process with multiple leptons and b-jets in the final state, such as $t\bar{t}H$, $t\bar{t}W$, and $t\bar{t}Z$. While precise measurements have been made of inclusive WZ production [14], WZ + heavy flavor remains poorly understood. This is largely because the QCD processes involved in the production of the b-jet make it difficult to simulate accurately. This introduces a large uncertainty for analyses that include this process as a background.

We perform a study of the fully leptonic decay mode of this channel; that is, events where both the W and Z decay leptonically. Because WZ has no associated jets at leading order, while the major backgrounds for this channel tend to have high jet multiplicity, events with more than two jets are rejected. This gives a final state signature of three leptons and one or two jets.

Events that meet a preselection criteria are sorted into regions based on the b-taggin score of their associated jets. This is done to separate WZ + b-jet events from WZ + charm and WZ + light jets. These regions are fit to data in order make a more accurate estimate of the contribution of WZ + heavy-flavor, where heavy-flavor jets include b-jets and charm jets. The full Run-2 dataset collected by the ATLAS detector, representing 139 fb^{-1} of data from pp collisions at $\sqrt{s} = 13$ TeV, is used for this study.

The fiducial volume at particle level is defined based on the number of stable leptons and jets in each event. Three light leptons with total charge ±1 and one or two associated jets are required. Only leptons which do not originate from hadron or τ decays are considered. The phase space definitions use dressed kinematics of the final state particles. Leptons are dressed by summing the momentum of photons within a cone of $\Delta R < 0.1$ of the lepton to correct the leptons energy. Particle level jets are reconstructed using the anti-k_t algorithm with a radius of $R = 0.4$. The kinematic selection applied to these objects is summarized below:

- Three light leptons with total charge ±1, $|\eta| < 2.5$
- OS lepton with p_T >10 GeV, SS leptons with p_T >20 GeV
- One OSSF lepton pair with $|M(ll) - 91.2\ GeV| < 10\ GeV$
- One or two associated truth jets with p_T >25 GeV, $|\eta| < 2.5$, $R < 0.4$

The result of the fit is used to extract the cross-section in this fiducial region for WZ + b and WZ + c with one associated jet, and WZ + b and WZ + c with two associated jets, where the number and flavor of the jets is determined at particle level. Events with both charm and b-jets are counted as WZ + b. The analysis reports a cross-section measurement of WZ + b and WZ + charm, along with their correlations, for both 1-jet and 2-jet exclusive regions.

Section 6 details the data and Monte Carlo (MC) samples used in the analysis. The reconstruction of various physics objects is described in Section 7. Section 8 describes the event selection applied to these samples, along the definitions of the various regions used in the fit. The multivariate analysis techniques used to separate the tZ background from WZ + heavy flavor are described in Section 9. Section 17 describes the various sources of systematic uncertainties considered in the fit. Finally, the results of the analysis are summarized in Section 18, followed by a brief conclusion in Section 12.

6 Data and Monte Carlo Samples

6.1 Data Samples

This study uses a sample of proton-proton collision data collected by the ATLAS detector from 2015 through 2018 at an energy of $\sqrt{s} = 13$ TeV, which represents an integrated luminosity of 139 fb^{-1} [15]. This data set was collected with a bunch-crossing rate of 25 ns. All data used in this analysis was verified by data quality checks [16].

6.2 Monte Carlo Samples

Several different generators were used to produce Monte Carlo simulations of the signal and background processes. For all samples, the response of the ATLAS detector is simulated using

GEANT4 [17]. The WZ signal samples are simulated using Sherpa 2.2.2 [18]. Specific information about the Monte Carlo samples being used can be found in Table 14.

Table 2: 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
WZ, ZZ, WW	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10 [19]
tZ	MG5_aMC [20]	NLO	PYTHIA 8	CTEQ6L1
$t\bar{t}W$	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
$t\bar{t}H$	MG5_aMC	NLO	PYTHIA 8	NNPDF 3.0 NLO [21]
	(MG5_aMC)	(NLO)	(HERWIG++) [22]	(CT10 [19])
$tHqb$	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 [23]
$t\bar{t}W^+W^-$	MG5_aMC	LO	PYTHIA 8	NNPDF 2.3 LO
$t\bar{t}$	POWHEG-BOX v2 [24]	NLO	PYTHIA 8	NNPDF 3.0 NLO
$t\bar{t}\gamma$	MG5_aMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	POWHEG-BOX v1 [25]	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 regions defined in this analysis share a common lepton, jet, and overall event preselection. The selection applied to each physics object is detailed here; the event preselection, and the selection used to define the various fit regions, is described in Section 8.

All events are required to be selected by dilepton triggers. The p_T thresholds of the dilepton trigger on two electrons were 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017 and 2018, while for the dimuon triggers the p_T thresholds on the leading (sub-leading) muon were 18 GeV (8 GeV) in 2015, and 22 GeV (8 GeV) in 2016-2018. For the electron+muon triggers, the p_T thresholds on the electron (muon) were 17 GeV (14 GeV) for all datasets.

Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that are associated with charged particle tracks reconstructed in the inner detector [26]. 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 real electrons from hadronic showers (fake electrons). To further reduce the non-prompt electron contribution, the track is required to be consistent with originating from the primary vertex; requirements are imposed on the transverse impact parameter significance ($|d_0|/\sigma_{d_0} < 5$) and the longitudinal impact parameter ($|\Delta z_0 \sin\theta_\ell| < 0.5$ mm). Electron candidates are required to pass the TightLH identification requirement detailed in [27].

Muon candidates are reconstructed by combining inner detector tracks with track segments or full tracks in the muon spectrometer [28]. Muon candidates are required to have $p_T > 10$ GeV and $|\eta| < 2.5$. The longitutinal impact parameter is the same for both electrons and muons, while muons are required to pass a slightly tighter transverse imapact parameter selection, $|d_0|/\sigma_{d_0} < 3$. Muons are also required to pass Medium ID requirements, as detailed in [27]. Leptons are additionally required to pass a non-prompt BDT selection, described in detail in

[29]. Optimized working points and scale factors for this BDT are taken from that analysis.

Jets are reconstructed from calibrated topological clusters built from energy deposits in the calorimeters using the anti-k_t algorithm [30], as well as information from the inner tracking detector, with a radius parameter $R = 0.4$. Jets with energy contributions likely arising from noise or detector effects are removed from consideration, 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 [31].

In order to make a measurement of WZ + heavy flavor it is necessary to distinguish these events from WZ + light jets. For this purpose, the DL1r b-tagging algorithm is used to distinguish heavy flavor jets from lighter ones [32]. The DL1r algorithm uses jet kinematics, particularly jet vertex information, as input for a neural network which assigns each jet a score designed to reflect how likely that jet is to have originated from a b-quark.

From the output of the BDT, calibrated working points (WPs) are developed based on the efficiency of truth b-jets at particular values of the DL1r algorithm. The working points used in this analysis are summarized in Table 3.

WP	Rejection	
b-jet eff.	c-jet	light jet
85%	2.6	29
77%	4.9	130
70%	9.4	390
60%	27	1300

Table 3: c-jet and light-flavor jet rejections corresponding to each b-tagging Working Point by b-jet efficiency, evaluated on $t\bar{t}$ events.

As shown in table 3, a tighter WP will accept fewer b-jets, but reject a higher fraction of charm and light jets. Generally, analyses that include b-jets will use a fixed working point, for example, requiring that a jet pass the 70% threshold. By instead treating these working point as bins, e.g. events with jets that fall between the 85% and 77% WPs fall into one bin, while events with jets passing the 60% WP fall into another, additional information can be gained. This analysis uses each of these working points to form orthogonal regions in order to provide separation between WZ + b, WZ + c, and WZ + light.

Missing transverse momentum (E_T^{miss}) is used as part of the event selection. The missing transverse momentum vector is defined as the negative of the vector 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, with object definitions taken from [33]. Light leptons considered in the E_T^{miss} reconstruction are required to have p_T>10 GeV, while jets are required to have p_T>20 GeV.

To avoid double counting objects and remove leptons originating from decays of hadrons, overlap removal is performed in the following order: any electron candidate within $\Delta R = 0.1$ of another electron candidate with higher p_T is removed; any electron candidate within $\Delta R = 0.1$ of a muon candidate is removed; any jet within $\Delta R = 0.2$ of an electron candidate is removed; if a muon candidate and a jet lie within $\Delta R = min(0.4, 0.04+10[GeV]/p_T(muon))$ of each other, the jet is kept and the muon is removed if the jet has at least three associated tracks, otherwise the jet is removed and the muon is kept. This algorithm is applied to the preselected objects.

8 Event Selection and Signal Region Definitions

Event are required to pass a preselection described in Section 8.1. Those that pass this preselection are divided into various fit regions described in Section 8.2, based on the number of jets in the event, and the b-tag score of those jets.

8.1 Event Preselection

Events are required to include exactly three reconstructed light leptons passing the requirement described in 7, which have a total charge of ±1. As the opposite sign lepton is found to be prompt the vast majority of the time [29], it is required to have $p_T > 10$ GeV, while the same sign leptons are required to have $p_T > 20$ GeV to reduce the contribution of non-prompt leptons.

The invariant mass of at least one pair of opposite sign, same flavor leptons is required to fall within 10 GeV of the mass of the Z boson, 91.2 GeV. Events where one of the opposite sign pairs have an invariant mass less than 12 GeV are rejected in order to suppress low mass resonances.

An additional requirement is placed on the missing transverse energy, $E_T^{miss} > 20$ GeV. The transverse mass of the W candidate, defined as $\sqrt{2p_T^{lep}E_T^{miss} * (1 - \cos(\phi_{lep} - \phi_{E_T^{miss}}))}$, is required to be greater than 30 GeV. Here E_T^{miss} is the missing transverse energy, and the lepton considered is the lepton not included in the Z-candidate.

Events are required to have exactly one or two reconstructed. Events with more than two jets are rejected in order to reduce the contribution of backgrounds such as $t\bar{t}Z$ and $t\bar{t}W$, which tend to have higher jet multiplicity.

The WZ events are split into WZ + b, WZ + c, and WZ + light based on the truth flavor of the associated jet in the event, as determined by the presence of a b- or c-hadron within $R = 0.3$ of the jet. In this ordering b-jet supersede charm, which supersides light. That is, WZ + light events contain no charm and no b jets at truth level, WZ + c contain at least one truth charm and no b-jets, and WZ + b contains at least one truth b-jet.

8.2 Fit Regions

Once preselection has been applied, the remaining events are categorized into one of twelve orthogonal regions. The regions used in the fit are summarized in Table 4.

Table 4: A list of the regions used in the fit and the selection used for each.

Region	Selection
1j, <85%	N_{jets} = 1, nJets_DL1r_85 = 0
1j, 85%-77%	N_{jets} = 1, nJets_DL1r_85 = 1, nJets_DL1r_77=0
1j, 77%-70%	N_{jets} = 1, nJets_DL1r_77 = 1, nJets_DL1r_70=0
1j, 70%-60%	N_{jets} = 1, nJets_DL1r_70 = 1, nJets_DL1r_60=0
1j, >60%	N_{jets} = 1, nJets_DL1r_60 = 1, tZ BDT > 0.12
1j tZ CR	N_{jets} = 1, nJets_DL1r_60 = 1, tZ BDT < 0.12
2j, <85%	N_{jets} = 2, nJets_DL1r_85 = 0
2j, 85%-77%	N_{jets} = 2, nJets_DL1r_85 >= 1, nJets_DL1r_77=0
2j, 77%-70%	N_{jets} = 2, nJets_DL1r_77 >= 1, nJets_DL1r_70=0
2j, 70%-60%	N_{jets} = 2, nJets_DL1r_70 >= 1, nJets_DL1r_60=0
2j, >60%	N_{jets} = 2, nJets_DL1r_60 >= 1, tZ BDT > 0.12
2j tZ CR	N_{jets} = 2, nJets_DL1r_60 >= 1, tZ BDT < 0.12

The working points discussed in Section 7 are used to separate events into fit regions based on the highest working point reached by a jet in each event. Because the background composition differs significantly based on the number of b-jets, events are further subdivided into 1-jet and 2-jet regions in order to minimize the impact of background uncertainties.

An unfolding procedure is performed to account for differences in the number of reconstructed jets compared to the number of truth jets in each event. In order to account for migration of WZ+1-jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal samples are separated based on the number of truth jets. Events with 0 jets or more than 3 jets at truth level, yet fall within one of the categories listed in Table 4, are categorized as WZ + other, and treated as background. The composition of the number of truth jets in each reco jet bin is taken from MC, with uncertainties in these estimates described in detail in Section 17.

An additional tZ control region is created based on the BDT described in Section 9. The region with 1-jet passing the 60% working point is split in two - a signal enriched region of events with a BDT score greater than 0.12, and a tZ control region including events with less than 0.12. This cutoff is optimized for significance of WZ + b.

8.3 Non-Prompt Lepton Estimation

Two processes that act as sources of non-prompt leptons appear in the analysis: $\mathrm{t}\bar{\mathrm{t}}$ and Z+jet production both produce two prompt leptons, but can meet the selection of this analysis when an additional non-prompt lepton appears in the event. The contribution of these processes is estimated with Monte Carlo simulations, which are validated using non-prompt enriched regions. These validation regions are used to derive correction factors and uncertainties for the non-prompt contribution.

$\mathrm{t}\bar{\mathrm{t}}$ events can produce two prompt leptons from the decay of each of the tops. These top decays produce two b-quarks, the decay of which can produce additional non-prompt leptons, which occasionally pass the event preselection. In order to validate that the Monte Carlo accurately simulates this process accurately, the MC prediction in a non-prompt $\mathrm{t}\bar{\mathrm{t}}$ enriched validation region is compared to data.

The $\mathrm{t}\bar{\mathrm{t}}$ validation region is similar to the preselection region - three leptons meeting the criteria described in Section 8 are required, and the requirements on $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ remain the same. However, the selection requiring that a lepton pair form a Z-candidate are reversed. Events where the invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of

91.2 GeV are rejected. This ensures the $t\bar{t}$ validation region is orthogonal to the preselection region.

Further, because the jet multiplicity of $t\bar{t}$ events tends to be higher than WZ + jets, the number of jets in each event is required to be greater than 1. As b-jets are almost invariably produced from top decays, at least one b-tagged jet passing the 70% DL1r WP in each event is required.

Data is compared to MC predictions in the region for a variety of kinematic variable, as well as various b-tag WPs. A constant normalization discrepency between data and MC predictions of approximately 10% is found, which is accounted for by applying a constant correction factor of 0.9 to the $t\bar{t}$ MC prediction. Once this correction factor has been applied, no significant modelling discrepencies, either in terms of shape or overall yield, are found in any of the kinematic distributions considered. As data and MC are found to agree within 20% for each of the b-tag WPs considered, a 20% systematic uncertainty on the $t\bar{t}$ prediction is included for the analysis.

Similar to $t\bar{t}$, a Z+jets validation region is produced in order to validate the MC predictions. The lepton requirements remain the same as the preselection region. Because no neutrinos are present for this process, the E_T^{miss} cut is reversed, requiring E_T^{miss} < 30 GeV. This also ensures this validation region is orthogonal to the preselection region. Further, the number of jets in each event is required to be greater than or equal to one.

While there is general agreement between data and MC, the shape of the p_T spectrum of the lepton from the W candidate is found to differ. This is the lepton not included in the Z-candidate, and in the case of Z+jets, this lepton is most often the non-prompt lepton. To account for this discrepency, a variable correction factor is applied to Z+jets. χ^2 minimization of the W lepton p_T spectrum is performed to derive a correction factor.

The systematic uncertainty in the Z + jets prediction is evaluated by comparing data to MC for each of the continuous b-tag WPs. For each of the regions considered, the data falls within 25% of the MC prediction once this correction factor has been applied. Therefore, a 25% systematic uncertainty is applied to Z + jets in the analysis.

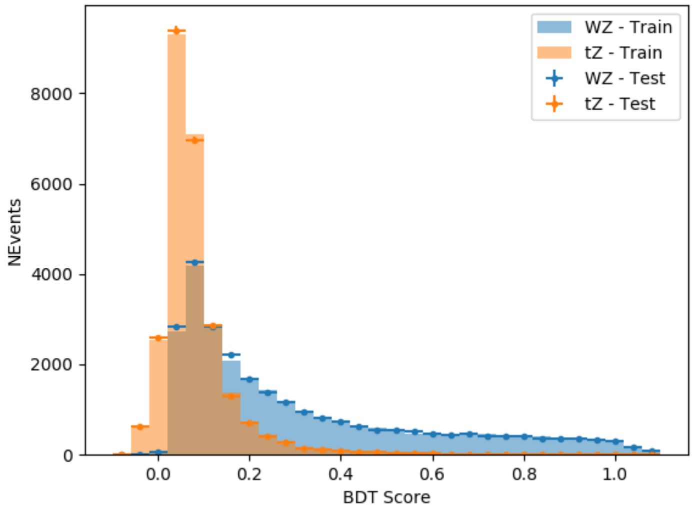
9 tZ Separation Multivariate Analysis

An important process to consider in this analysis is tZ: the top almost always decays into a W boson and b-quark, and when both the W and Z decay leptonically, this gives three leptons and a heavy flavor jet in the final state. Because tZ can produce a final state identical to the signal, it represents a predominant background in the most signal enriched regions. That is, the region with one jet passing the 60% DL1r WP. Therefore, a boosted decision tree (BDT) algorithm is trained using XGBoost [34] to separate WZ + heavy flavor from tZ using kinematic quantities. The result of this BDT is used to create a tZ enriched region in the fit, reducing its impact on the measurement of WZ + heavy flavor.

The kinematic variables used as inputs to train this BDT include the invariant mass of the reconstructed top candidate, the p_T of each of the leptons and associated jets, the invariant mass of each combination of lepton pairs, E_T^{miss}, the distance between each combination of leptons, $\Delta R(ll)$, and the distance between each lepton and the jet, $\Delta R(lj)$.

Here the top candidate is reconstructed based on the procedure described in section 6.1 of [35]. Broadly, the mass of the top quark candidate is reconstructed from the jet, the lepton not included in the Z-candidate, and a reconstructed neutrino. In the case that there is one jet in the event, there is only possible b-jet candidate. For events with two jets, the jet with the highest DL1r score is used.

The training samples included only events meeting the requirements of the 1-jet, >60% region, i.e. passing all the selection described in section 8 and having exactly one jet which passes the tightest (60%) DL1r working point.A sample of 20,000 background (tZ) and signal (WZ+b) Monte Carlo events are used to train the BDT. And additional 5,000 events are reserved for testing the model, in order to prevent over-fitting. A total of 750 decision trees with a maximum depth of 6 branches are used to build the model. These parameters are chosen empirically, by training several models with different parameters and selecting the one that gave the best separation for the test sample. The results of the BDT training are shown in figure 9.1.



(a)

Figure 9.1: Distribution of the BDT response for WZ+b (blue) and tZ (orange) events, for both training and testing samples.

A BDT score of 0.12 is selected as a cutoff, where events with scores higher than this form a signal enriched region, and events with scores lower than this form a tZ control region. This cutoff is selected by varying the value of this cutoff in stat-only Asimov fits, and selecting the value that minimizes the statistical uncertainty on WZ + b.

10 Systematic Uncertainties

The systematic uncertainties that are considered are summarized in Table 34. 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 18.

Table 5: Sources of systematic uncertainty considered in the analysis.

Systematic uncertainty	Components
Luminosity	1
Pileup reweighting	1
Physics Objects	
Electron	6
Muon	15
Jet energy scale	28
Jet energy resolution	8
Jet vertex fraction	1
Jet flavor tagging	131
E_T^{miss}	3
Total (Experimental)	194
Signal Modeling	
Shape modelling	3
Renormalization and factorization scales	5
nJet Migration	5
Background Modeling	
Cross section	15
Renormalization and factorization scales	12
Total (Signal and background modeling)	35
Total (Overall)	230

The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [36], obtained using the LUCID-2 detector [37] for the primary luminosity measurements.

The experimental uncertainties are related to the reconstruction and identification of light leptons and b-tagging of jets, and to the reconstruction of E_T^{miss}. The sources which contribute to the uncertainty in the jet energy scale (JES) [38] are decomposed into uncorrelated components and treated as independent sources of uncertainty in the analysis. A similar approach is used for the jet energy resolution (JER) uncertainty.

The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses [39] are also decomposed into uncorrelated components. The large number of components for b-tagging is due to the calibration of the distribution of the MVA discriminant for each individual WP bin.

The fit involves varying the overall normalization of signal templates over the regions described in Section 8.2, which are defined by the flavor and number of associated jets at truth-level. The modelling of these template shapes therefore significantly impacts the final result. Additional signal uncertainties, probing the shape of the signal templates as well as the rate of migrations between the number of truth-jets and reconstructed jets, are estimated by comparing estimates from the nominal Sherpa WZ samples with alternative WZ samples generated with Powheg+Pythia8. Separate systematics are included in the fit for WZ + b, WZ + c and WZ + light, where the distribution among each of the fit regions is varied based on the prediction of the Powheg sample.

A similar approach is taken to account for uncertainties in migrations between the number of reco and truth jets. The fraction of events with 1 truth jet which fall into the 1 jet bin versus the 2 jet bin at reco level is compared for Sherpa and Powheg. The same is done for events with 2 truth jets. A systematic is included where events are shifted between the 1-jet and 2-jet regions based on the differences between these two shapes. This is done independently for each of the WZ + b, WZ + c, and WZ + light templates.

Additional systematics are included to account for the uncertainty in the contamination of 0 jet and 3 or more jet events (as defined at truth level) in the 1 and 2 reco jet bins. Because these events fall outside the scope of this measurement, these events are included as a background. As such, a normalization, rather than a shape, uncertainty is applied for this background. The number of WZ events with 0-jets and >=3-jets in the reconstructed 1-jet and 2-jet regions are compared for Sherpa and Powheg, and these differences are taken as separate normalization systematics on the yield of WZ+0-jet and WZ+>=3-jet events.

Theoretical uncertainties applied to MC background predictions, including cross section, PDF, and scale uncertainties are taken from theory calculations, with the exception of non-prompt and diboson backgrounds. The cross-section uncertainty on tZ is taken from [40]. Derivation of the non-prompt background uncertainties, Z+jets and $t\bar{t}$, are explained in Section 8.3. These normalization uncertainties are chosen so as to account for the complete uncertainty in the non-prompt contribution, and therefore no additional modelling uncertainties are considered for Z+jets and $t\bar{t}$.

Due to its importance as a background, additional modelling uncertainties are considered for tZ. Alternative tZ samples with variations in scale and shower modelling are included as systematics. The other VV + heavy flavor processes (namely VV+b and VV+charm, which primarily consist of ZZ events) are also poorly understood, because these processes involve the same physics as WZ + heavy flavor, and have also not been measured. Therefore, a conservative 50% uncertainty is applied to those samples. While this uncertainty is large, it is found to have little impact on the significance of the final result.

747 The theory uncertainties applied to the MC estimates are summarized in Table 36.

Process	X-section [%]
WZ	QCD Scale: $^{+3.7}_{-3.4}$ PDF($+\alpha_S$): ± 3.1
tZ	X-sec: ± 15.2
$t\bar{t}$ H (aMC@NLO+Pythia8)	QCD Scale: $^{+5.8}_{-9.2}$ PDF($+\alpha_S$): $\pm$ 3.6
$t\bar{t}$ Z (aMC@NLO+Pythia8)	QCD Scale: $^{+9.6}_{-11.3}$ PDF($+\alpha_S$): $\pm$ 4
$t\bar{t}$ W (aMC@NLO+Pythia8)	QCD Scale: $^{+12.9}_{-11.5}$ PDF($+\alpha_S$): $\pm$ 3.4
VV + b/charm (Sherpa 2.2.1)	± 50
VV + light (Sherpa 2.2.1)	± 6
$t\bar{t}$	± 20
Z + jets	± 25
Others	± 50

Table 6: Summary of theoretical uncertainties for normalization of MC predictions in the analysis.

748 # 11 Results

749 ## 11.1 Fit Procedure

750 A maximum-likelihood fit is performed over the total yields in the various fit regions described
751 in Section 8 in order to extract the best-fit value of the WZ + b and WZ + c jet contributions for
752 events with both 1 and 2 associated jets.

753 Because the fit regions are defined by the number of associated jets at reco-level, the signal
754 is split into separate samples based on the number of truth jets in order to account for differences
755 in the number of truth jets compared to the number of reco-jets. The WZ + b, WZ + c and *WZ*

+ light contributions are separated into independent samples based on the number of truth jets in each event. WZ + 1 truth-jet and WZ + 2 truth-jets are treated as signal samples, while WZ + 0 truth-jets and WZ + >=3 truth-jets are treated as an additional background.

A maximum likelihood fit to data is performed simultaneously in the regions described in Section 8, summarized in figure 11.1. The six signal templates, which include WZ + b 1-jet, WZ + c 1-jet, WZ+ light 1-jet, WZ + b 2-jets, WZ + c 2-jets, WZ + light 2-jets, are allowed to float, while the remaining background contributions are held fixed. Normalization factors for each of these templates are extracted from the fit. A simultaneous fit is performed over all 1-jet and 2-jet regions.

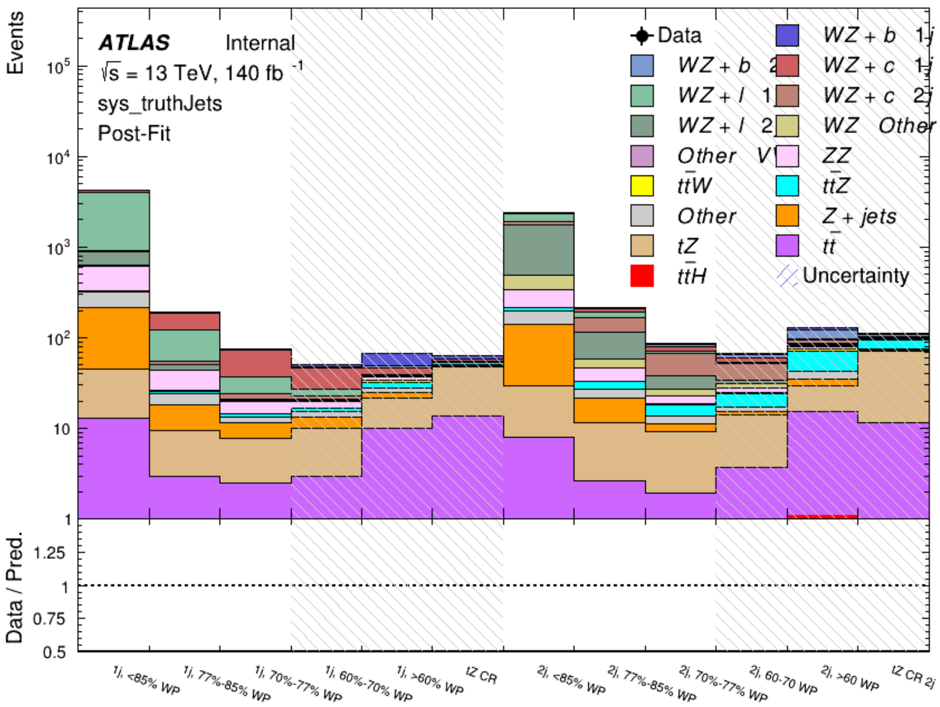


Figure 11.1: Post-fit summary of the fit regions.

Several alternative fit strategies are reported as well, including a measurement of WZ + 1 or 2 jets inclusively, a fit where tZ is allowed to float, and a case where tZ is included as part of the signal.

As described in Section 17, there are 230 systematic uncertainties that are considered as NPs in the fit. These NPs are constrained by Gaussian or log-normal probability density functions. The latter are used for normalisation factors to ensure that they are always positive. The expected number of signal and background events are functions of the likelihood. The prior for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its nominal value.

11.2 Results of the Simultaneous Fit

The results of the fit to an Asimov dataset for the fiducial regions considered, including both the normalization factors as well as the expected cross-sections, along with their uncertainties, are summarized in Table 7.

Process	μ	σ
WZ + b - 1-jet	$1.00^{+0.47}_{-0.43}(\text{stat})^{+0.32}_{-0.27}(\text{sys})$	$1.74^{+0.82}_{-0.75}(\text{stat})^{+0.53}_{-0.48}(\text{sys})$ fb
WZ + c - 1-jet	$1.00^{+0.18}_{-0.17}(\text{stat})^{+0.19}_{-0.17}(\text{sys})$	$14.6^{+2.5}_{-2.3}(\text{stat})^{+2.6}_{-2.3}(\text{sys})$ fb
WZ + b - 2-jet	$1.00^{+0.53}_{-0.51}(\text{stat})^{+0.39}_{-0.34}(\text{sys})$	$2.5^{+1.3}_{-1.3}(\text{stat})^{+0.95}_{-0.83}(\text{sys})$ fb
WZ + c - 2-jet	$1.00^{+0.25}_{-0.24}(\text{stat})^{+0.32}_{-0.27}(\text{sys})$	$12.7^{+3.3}_{-3.2}(\text{stat})^{+3.9}_{-3.4}(\text{sys})$ fb

Table 7: Normalization factors and cross-sections extracted from the fit for each of the fiducial regions considered

An expected significance of 2.0σ is observed for WZ + b with 1-jet, and 1.7σ for WZ + b with two jets. A summary of the correlations between these various measurements is shown in Figure 11.2.

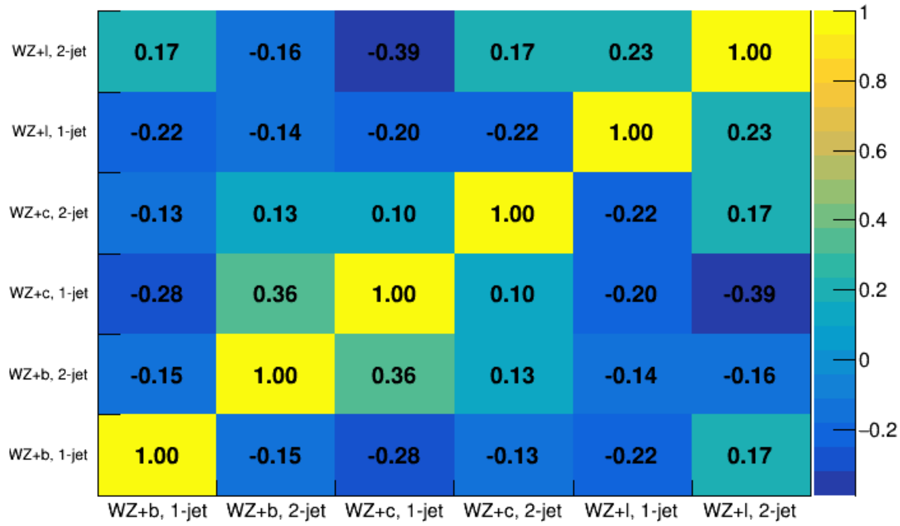


Figure 11.2: Correlations between the various measured components of WZ.

The impact of each NP is calculated by performing the fit with the parameter of interest held fixed, varied from its fitted value by its uncertainty, and calculating Δμ relative to the baseline fit. The impact of the most significant sources of systematic uncertainties on WZ + b and WZ + c with one associated jet are summarized in Table 8-9.

Uncertainty Source	$\Delta\sigma/\sigma_{nominal}$	
Jet Energy Scale	0.14	-0.15
WZ + light, 1-jet cross-section	0.12	-0.14
WZ + c, 1-jet cross-section	-0.09	0.11
tZ Modelling (shower tune)	-0.07	0.08
Other Diboson + b cross-section	-0.07	0.07
tZ cross-section	-0.06	0.08
Jet Energy Resolution	-0.07	0.07
WZ + b 1j/2j Migration	0.08	-0.07
Luminosity	-0.06	0.07
Flavor tagging	0.05	0.05
$t\bar{t}$ cross-section	-0.05	0.05
Total Systematic Uncertainty	0.28	0.33

Table 8: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

Uncertainty Source	$\Delta\sigma/\sigma_{nominal}$	
WZ + c 1j/2j migration	0.12	-0.09
Flavor Tagging	0.09	0.08
WZ + b, 1-jet cross-section	-0.04	0.05
Luminosity	-0.04	0.04
Jet Energy Resolution	0.04	0.04
WZ + b, 2-jet cross-section	0.04	-0.03
WZ cross-section - QCD scale	-0.04	0.04
Jet Energy Scaling	0.04	0.02
WZ cross-section - PDF	-0.03	0.03
WZ + light, 1-jet cross-section	0.03	-0.03
total	0.19	0.17

Table 9: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + c with exactly one associated jet.

The impact of the most significant systematic uncertainties on the 2-jet fiducial regions are summarized in Table 10-11.

Uncertainty Source	$\Delta\sigma/\sigma_{nominal}$	
WZ + c 2-jet cross-section	-0.13	0.16
WZ + l 2-jet cross-section	0.12	-0.09
ttZ cross-section - QCD scale	-0.10	0.13
Luminosity	-0.11	0.12
WZ + b 1-jet cross-section	-0.11	0.10
Jet Energy Scale	-0.11	0.11
tZ cross-section	-0.11	0.11
WtZ cross-section	-0.07	0.07
Flavor tagging	0.05	0.05
Other VV + b cross-section	-0.05	0.05
Total	0.36	0.37

Table 10: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with 2 associated jets.

Uncertainty Source	$\Delta\sigma/\sigma_{nominal}$	
WZ + c 1j/2j migration	-0.17	0.25
Flavor Tagging	0.14	0.13
WZ + b, 1-jet cross-section	-0.09	0.09
Jet Energy Scale	0.06	0.08
Jet Energy Resolution	0.05	0.05
WZ >=3j/2j migration	-0.04	0.04
WZ + c 2j/1j migration	-0.04	0.04
WZ cross-section - QCD scale	-0.04	0.04
WZ + light modelling	0.04	-0.03
Luminosity	-0.03	0.03
total	0.27	0.32

Table 11: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + c with 2 associated jets.

788 ### 11.3 Inclusive 1-2 Jet Fit

An alternative fit is performed which combines the WZ + 1-jet and WZ + 2-jet samples rather than fitting them independently. This is done primarily as a cross-check of the nominal analysis, to see if measuring 1-jet and 2-jet events separately and combining them gives drastically different results than measuring them together.

For this study, three signal templates, WZ + b, WZ + c and WZ + light, are fit to data, and the systematics accounting for migrations between 1-jet and 2-jet bins are removed. All other background and nuiscance parameters remain the same as the nominal fit.

The measured μ value for WZ + b is $\mu = 1.00^{+0.30}_{-0.29}(stat)^{+0.25}_{-0.23}(sys)$, with a significance of 2.8σ, and the uncertainty on WZ + c is $\mu = 1.00 \pm 0.12(stat) \pm 0.13(sys)$. This is compared to combined uncertainty of $\mu = 1.00^{+0.32}_{-0.30}(stat)^{+0.24}_{-0.23}(sys)$ for WZ + b when 1-jet and 2-jet events are measured separately and then combined.

A post-fit summary plot of the fit regions is shown in Figure 11.3:

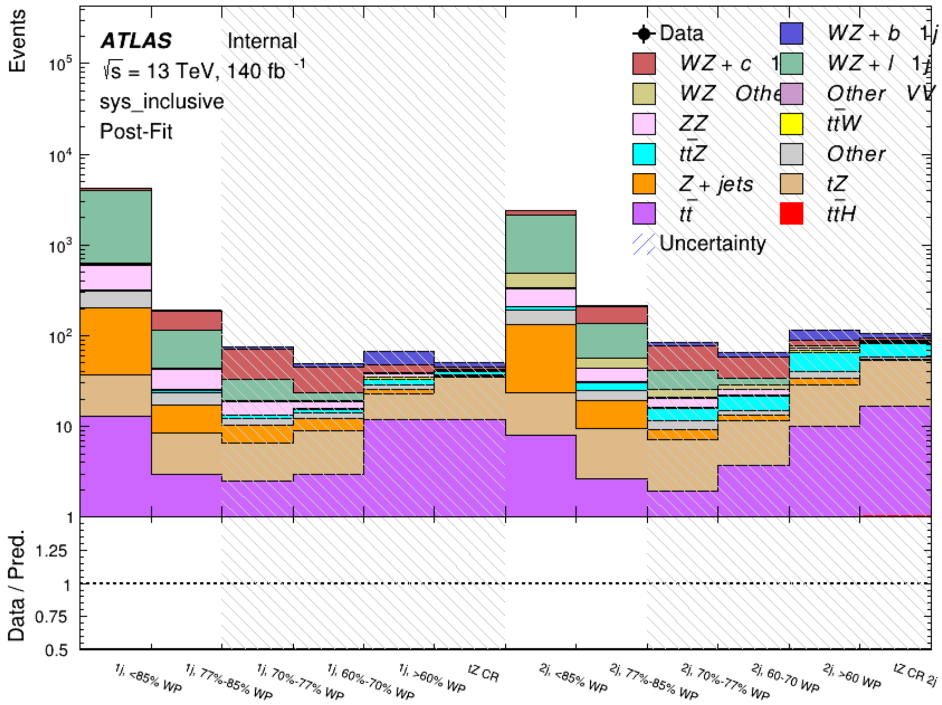


Figure 11.3: Post-fit summary of the fit regions.

801 The impact of the most significant sources of systematic uncertainties on the measurement
802 of WZ + b is summarized in Table 12.

Uncertainty Source	$\Delta\mu$	
WZ + light cross-section	0.13	-0.12
WZ + c cross-section	-0.10	0.12
Jet Energy Scale	0.08	0.13
tZ cross-section	-0.10	0.10
Jet Energy Resolution	-0.10	0.10
Luminosity	-0.08	0.09
Other Diboson + b cross-section	-0.07	0.07
Flavor tagging	0.05	0.05
$t\bar{t}$ cross-section	-0.05	0.05
WZ cross-section - QCD scale	-0.04	0.03
Total Systematic Uncertainty	0.28	0.32

Table 12: Summary of the most significant sources of systematic uncertainty on the measurement of $WZ + b$ with one or two associated jets.

11.4 Alternate tZ Inclusive Fit

11.4.1 tZ Inclusive Fit

While tZ is often considered as a distinct process form $WZ + b$, this could also be considered part of the signal. Alternative studies are performed where, using the same framework as the nominal analysis, a measurement of $WZ + b$ is performed that includes tZ as part of WZ+b.

Because of this change, the tZ CR is no longer necessary, and only the five pseudo-continuous b-tag regions are used in the fit. Further, systematics related to the tZ cross-section are removed from the fit, as they are now encompassed by the normalization measurement of $WZ + b$. All other systematic uncertainties are carried over from the nominal analysis.

An expected $WZ + b$ cross-section of $4.1^{+0.78}_{-0.74}(stat)^{+0.53}_{-0.52}(sys)$ fb is extracted from the fit, with an expected significance of 4.0σ.

The impact of the predominate systematics are summarized in Table 44.

Uncertainty Source	$\Delta\mu$	
WZ + light cross-section	0.08	-0.08
Jet Energy Scale	-0.06	0.08
Luminosity	-0.05	0.06
WZ + c cross-section	-0.04	0.05
Other Diboson + b cross-section	-0.04	0.04
WZ cross-section - QCD scale	-0.04	0.03
$t\bar{t}$ cross-section	-0.03	0.03
Jet Energy Resolution	-0.03	0.03
Flavor tagging	-0.03	0.03
Z+jets cross section	-0.02	0.02
Total Systematic Uncertainty	-0.15	0.16

Table 13: Summary of the most significant sources of systematic uncertainty on the measurement of $WZ + b$ with exactly one associated jet.

11.4.2 Floating tZ

In order to quantify the impact of the tZ uncertainty on the fit, an alternate fit strategy is used where the tZ normalization is allowed to float. This normalization factor replaces the cross-section uncertainty on tZ, and all other parameters of the fit remain the same.

An uncertainty of 17% on the normalization of tZ is extracted from the fit, compared to a theory uncertainty of 15% applied to the tZ cross-section. The measured uncertainties on WZ remain the same.

12 Conclusion

A measurement of WZ + heavy flavor is performed using 140 fb^{-1} of $\sqrt{s} = 13$ TeV proton-proton collision data collected by the ATLAS detector at the LHC. The expected cross-section of WZ + b with 1-jet is $1.74^{+0.82}_{-0.75}(stat)^{+0.53}_{-0.48}(sys)$ fb, and $14.6 \pm 2.5(stat) \pm 2.3(sys)$ fb for WZ + c, with a correlation of -0.22 between them. An expected significance of 2.0 is observed for WZ + b in this region.

For the 2-jet regions, an expected significance of 1.7 is observed for WZ + b, with an expected cross-section of $2.5^{+1.3}_{-1.3}(stat)^{+0.95}_{-0.83}(sys)$ fb. For WZ + c, a cross-section of $12.7 \pm 3.2(stat) \pm 2.7(sys)$ fb is expected for 2-jet events. A correlation of -0.26 is observed for WZ + b and WZ + c.

This section will be include final results once unblinded.

Part V

Differential Studies of $t\bar{t}H$ Multilepton

13 Data and Monte Carlo Samples

13.1 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} [15] and an energy of $\sqrt{s} = 13$ TeV. All data used in this analysis was included in one of the Good Run Lists verified by Data Quality checks [16].

13.2 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 [17]. The specific event generator used for each of these MC samples is listed in Table 14. A Higgs mass of 125 GeV is assumed in all simulations.

Table 14: 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 [21]
	(MG5_aMC)	(NLO)	(Herwig++)	(CT10 [19])
$t\bar{t}W$	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 [24]	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 [25]	NLO	Pythia 6	CT10
$qqVV$, VVV				
$Z \to l^+l^-$	Sherpa 2.2.1	MEPS NLO	Sherpa	NNPDF 3.0 NLO

The signal sample ($t\bar{t}H$) is modelled at NLO with Powheg-BOX v2 using the NNPDF2.0 parton distribution function (PDF) [41]. Parton showering and hadronisation were modelled with Pythia 8.2 [42]. The $t\bar{t}H$ sample is normalized to a cross-section of 507^{+35}_{-50} fb based on NLO calculations. Uncertainties are based on varying the QCD factorisation and renormalisation scale, as well as uncertainties in the PDF and α_s.

The $t\bar{t}W$ background is simulated using Sherpa 2.2.1 with the NNPDF3.0 NLO PDF. The matrix element is calculated with up to one additional parton at NLO, and up to two at LO. As explained in detail in [7], the $t\bar{t}W$ contribution predicted by MC is found disagree significantly with what is observed in data. While an effort is currently being undertaken to measure $t\bar{t}W$ more accurately, the approached used by the 79.8 fb^{-1} $t\bar{t}H$ analysis is used here: A normalization factor of 1.68 is applied to the MC estimate of $t\bar{t}W$ and additional systematic uncertainties on $t\bar{t}W$ are included to account for this modelling descrepency, as outlined in Section 17.

The $t\bar{t}(Z/\gamma^{*})$ process is simulated with the MadGraph5_aMC@NLO generator, using NNPDF3.0. Diboson processes are generated with Sherpa 2.2.2 at NLO precision for one extra parton, and at LO for up to three extra partons.

The "fake", or non-prompt, background comes primarily from leptons originating from hadron decays, leptons with missidentified charge,and photon conversions. While the main $t\bar{t}H$ analysis is currently refining a data-driven approach for estimating the contribution of events with non-prompt leptons, at the time of this note this strategy has not been completely developed for the full Run-2 dataset. Therefore, the non-prompt contribution is estimated with MC, while applying normalization corrections and systematic uncertainties derived from data driven techniques developed for the 79.8 fb^{-1} $t\bar{t}H/t\bar{t}W$ analysis [7].

The primary contribution to the non-prompt lepton background is from $t\bar{t}$ production, with V+jets and single-top as much smaller sources. Estimation of this background is done primarily using an inclusive $t\bar{t}$ sample, with corrections applied based on data driven methods. This sample is generated using Powheg, with Pythia8 performing the parton shower and fragmentation. Likelihood fits over several control regions enriched with these non-prompt backgrounds are fit to data in order to derive normalization factors for these backgrounds. The specific normalization factors and uncertainties applied to the non-prompt contributions are listed in Section 17.

Other processes, such as tH, tZ, $t\bar{t}WW$ and $t\bar{t}t\bar{t}$, are expected to make minor contributions to the total background. The generators and setting used for these backgrounds are summarized in Table 14.

14 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. Events are required to be selected by dilepton triggers. The p_T thresholds of the dilepton trigger on two electrons were 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017 and 2018, while for the dimuon triggers the p_T thresholds on the leading (sub-leading) muon were 18 GeV (8 GeV) in 2015, and 22 GeV (8 GeV) in 2016-2018. For the electron+muon triggers, the p_T thresholds on the electron (muon) were 17 GeV (14 GeV) for all datasets.

Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that are associated with charged particle tracks reconstructed in the inner detector [43]. 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 ($|\Delta z_0 \sin\theta_\ell|$). Muon candidates are reconstructed by combining inner detector tracks with track segments or full tracks in the muon spectrometer [28]. Muon candidates are required to have $p_T > 10$ GeV and $|\eta| < 2.5$.

All leptons are required to pass a non-prompt BDT selection developed by the main $t\bar{t}H/t\bar{t}W$ analysis, described in detail in [7]. Optimized working points and scale factors for this BDT are taken from that analysis. This BDT and the WPs used are summarized in Appendix A,

Jets are reconstructed from calibrated topological clusters built from energy deposits in the calorimeters [44], using the anti-k_t algorithm with a radius parameter $R = 0.4$. Particle Flow, or PFlow, jets are used in the analysis, which are hadronic objects reconstructed using information from both the tracker and the calorimeter. Jets with energy contributions likely arising from noise or detector effects are removed from consideration [45], 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 [31].

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

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 [46].

To avoid double counting objects and remove leptons originating from decays of hadrons, overlap removal is performed in the following order: any electron candidate within $\Delta R = 0.1$ of another electron candidate with higher p_T is removed; any electron candidate within $\Delta R = 0.1$ of a muon candidate is removed; any jet within $\Delta R = 0.2$ of an electron candidate is removed; if a muon candidate and a jet lie within $\Delta R = min(0.4{,}0.04{+}10[GeV]/p_T(muon))$ of each other, the jet is kept and the muon is removed if the jet has three or more tracks, otherwise the muon is kept and the jet is removed.

15 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 kinematics of the Higgs Boson decay products - as verified by studies detailed in Appendix C.3. 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 15.2. As described in Section 15.3, the kinematics of these jets and possible Higgs decay products are fed into the second layer, which is designed to identify the decay products of the Higgs Boson itself. The kinematics of the particles this layer identifies as most likely to have originated from the Higgs decay are then fed into yet another neural-network, which predicts the momentum of the Higgs (15.4). For the $3l$ channel, because the Higgs can decay into either one lepton and two jets or two leptons, an additional MVA is used to determine the decay mode of the Higgs boson in the $3l$ channel (15.5).

Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_aMC. For all of these models, the Keras neural network framework, with Tensorflow 2.0 as the backend [**tensorflow**], 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 in order to stabilize the model and decrease training time. 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 79.8 fb^{-1} of data, with the full set of features shown in Section C.

15.1 Physics Object Truth Matching

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

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

Leptons considered as possible Higgs and top decay candidates are required to pass the selection described in Section 14. 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 14. 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 threshold. The jets are found to be well modeled even down to this low p_T threshold, as shown in Section 16.1. The impact of using different p_T selection for the jet candidates is considered in detail in Section C.6. The overlap removal selection is not applied to the objects considered in the models.

15.2 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, kinematic variables for 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.

An alternate approach is considered, where information about all jets in each event are used as the feature set, and the model is tasked with identifying which two originated from the top decay. While this approach is found to underperform the nominal approach, and therefore not used in the analysis, the results are documented in Appendix C.4.

15.2.1 2lSS Channel

For the $2lSS$ channel, the input features shown in Table 15 are used for training. Here j_0 and j_1 are the two jet candidates, while l_0 and l_1 are the two leptons in the event, both ordered by p_T. jet DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet, where 5 represents the tightest working point and 1 represents the loosest. The variables nJets

DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and 85% b-tag working points, respectively.

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

Table 15: Input features used in the b-jet identification algorithm for the $2lSS$ channel

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

The difference between the distributions for a few of these features for the "correct" (i.e. both jets are truth b-jets), and "incorrect" combinations are shown in Figure 15.1. The correct and incorrect contributions are scaled to the same integral, so as to better demonstrate the differences in the distributions.

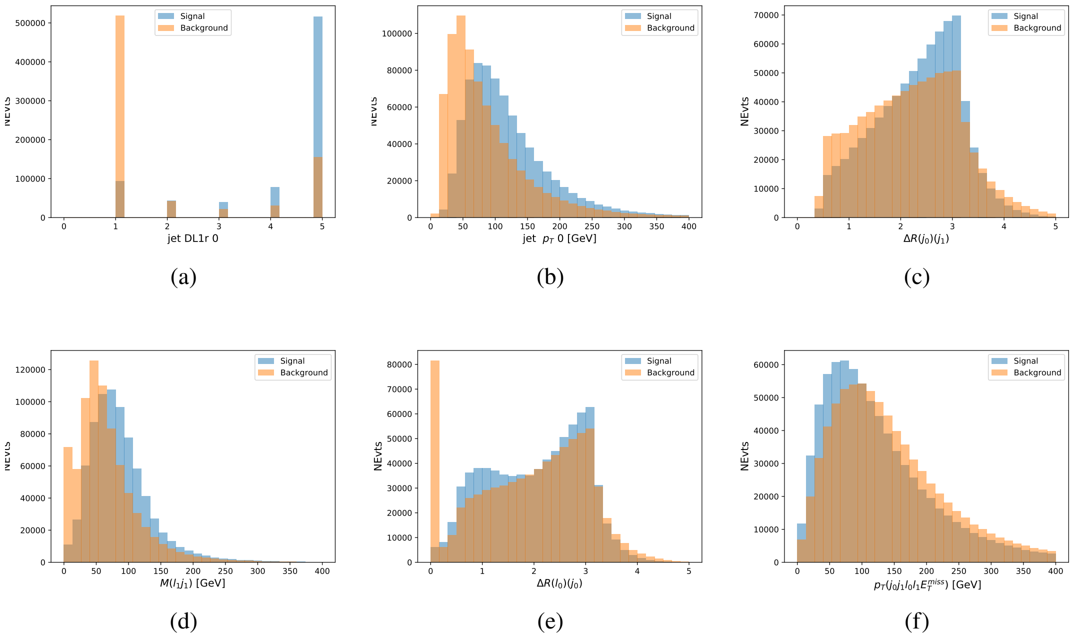


Figure 15.1: Input features for top2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss}.

The modeling of these inputs is validated against data, with Figure 15.2 showing good general agreement between data and MC. Plots for the complete list of features can found in Appendix C.

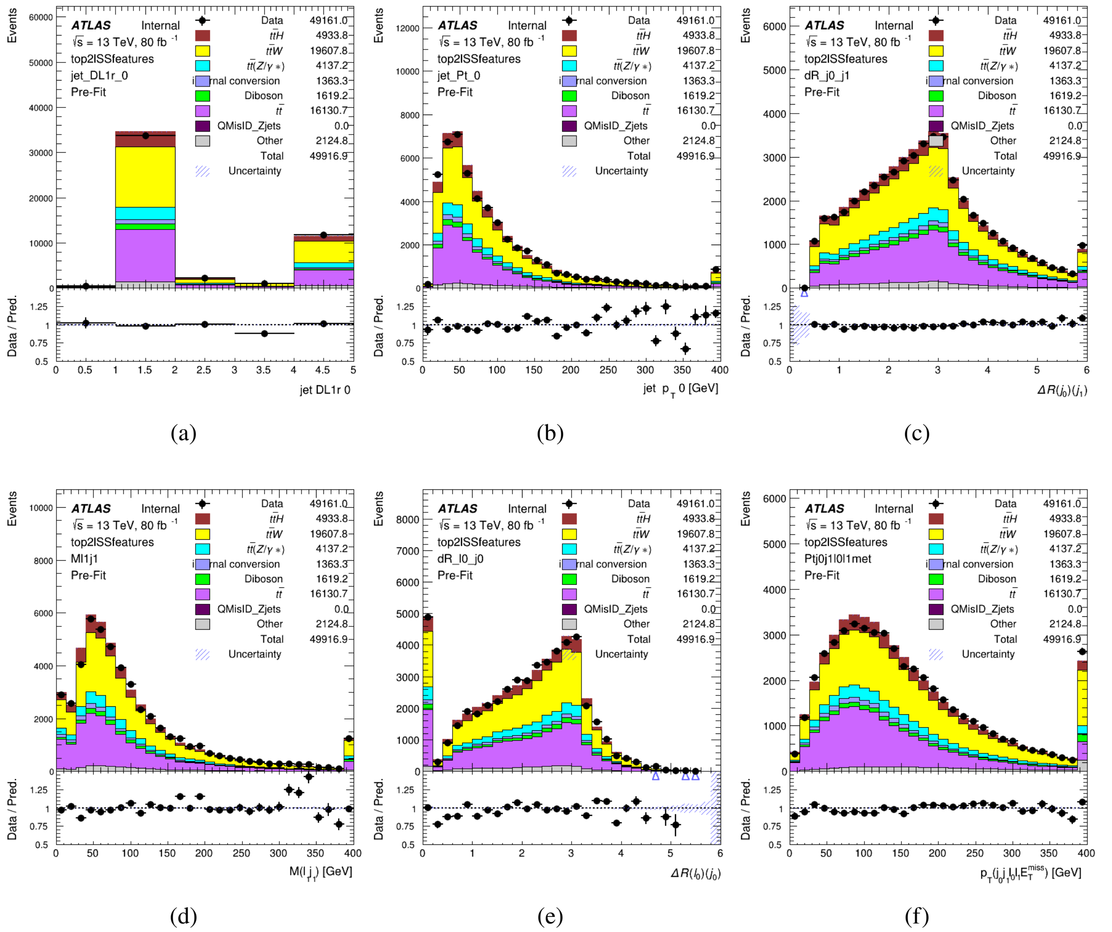
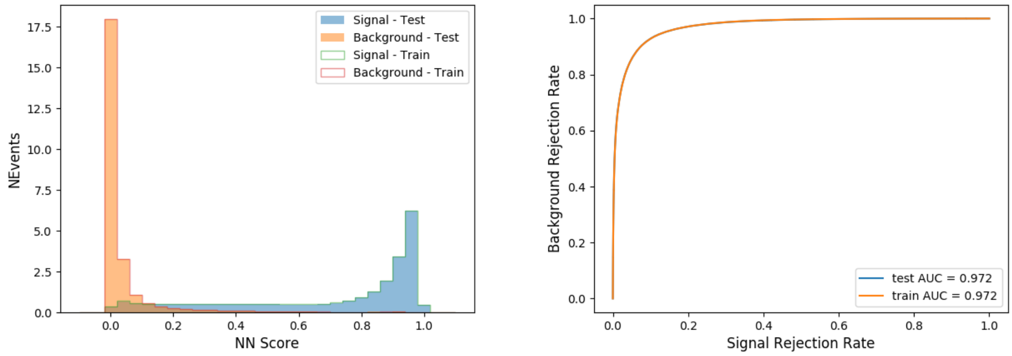


Figure 15.2: Data/MC comparisons of input features for top2lSS training for 79.8 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss}.

Based on the results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as overfitting is found to not be an issue. The output score distribution as well as the ROC curve for the trained model are shown in Figure 15.2.1. The model is found to identify the correct pairing of jets for 73% of 2lSS signal events on test data.



(a) (b)

Figure 15.3: 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

For point of comparison, a "naive" approach to identifying 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 b-tagged jet multiplicities, compared to this naive approach, is shown in Table 16.

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

Table 16: Accuracey of the NN in identifying b-jets from tops in 2lSS events, overall and split by the number of b-tagged jets in the event, compared to the accuracey of taking the two highest b-tagged jets.

This suggests that when there are exactly two b-tagged jets in an event, little is gained by using this more sophisticated approach, while for events with 1 or >= 3 b-tagged jets, the model does provide significant improvements.

15.2.2 3l Channel

The input features used in the $3l$ channel are listed in Table 17, with the same naming convention as the $2lSS$ channel.

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

Table 17: Input features for the b-jet identification algorithm in the $3l$ channel.

A few of these features are shown in Figure 15.4, comaring the distributions for correct and incorrect combinations of jets.

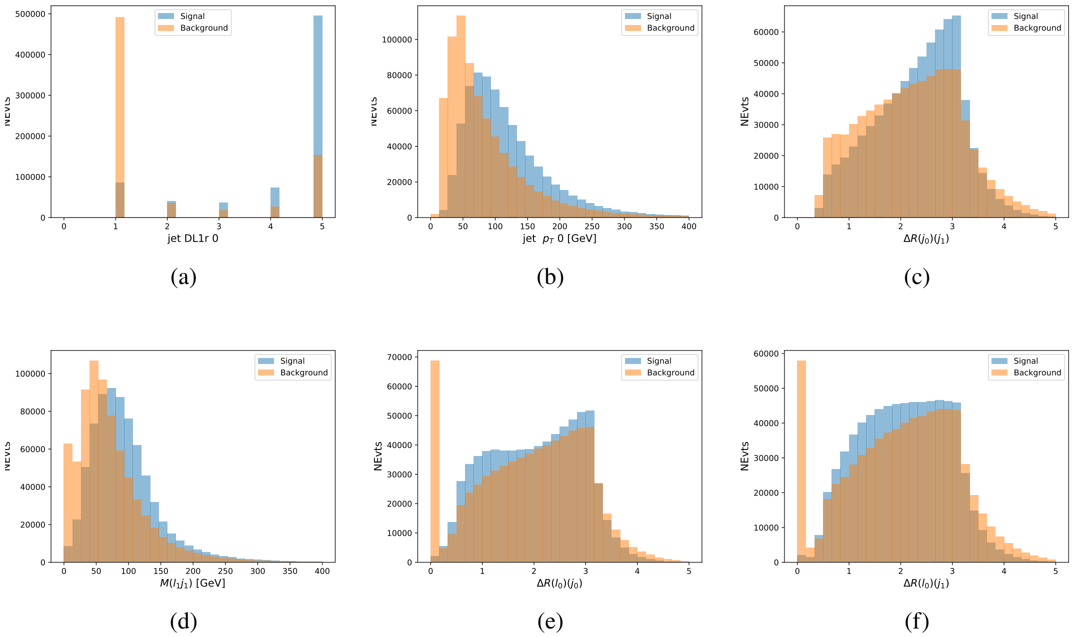


Figure 15.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

The modeling of these inputs is validated against data, with Figure 15.5 showing good general agreement between data and MC.

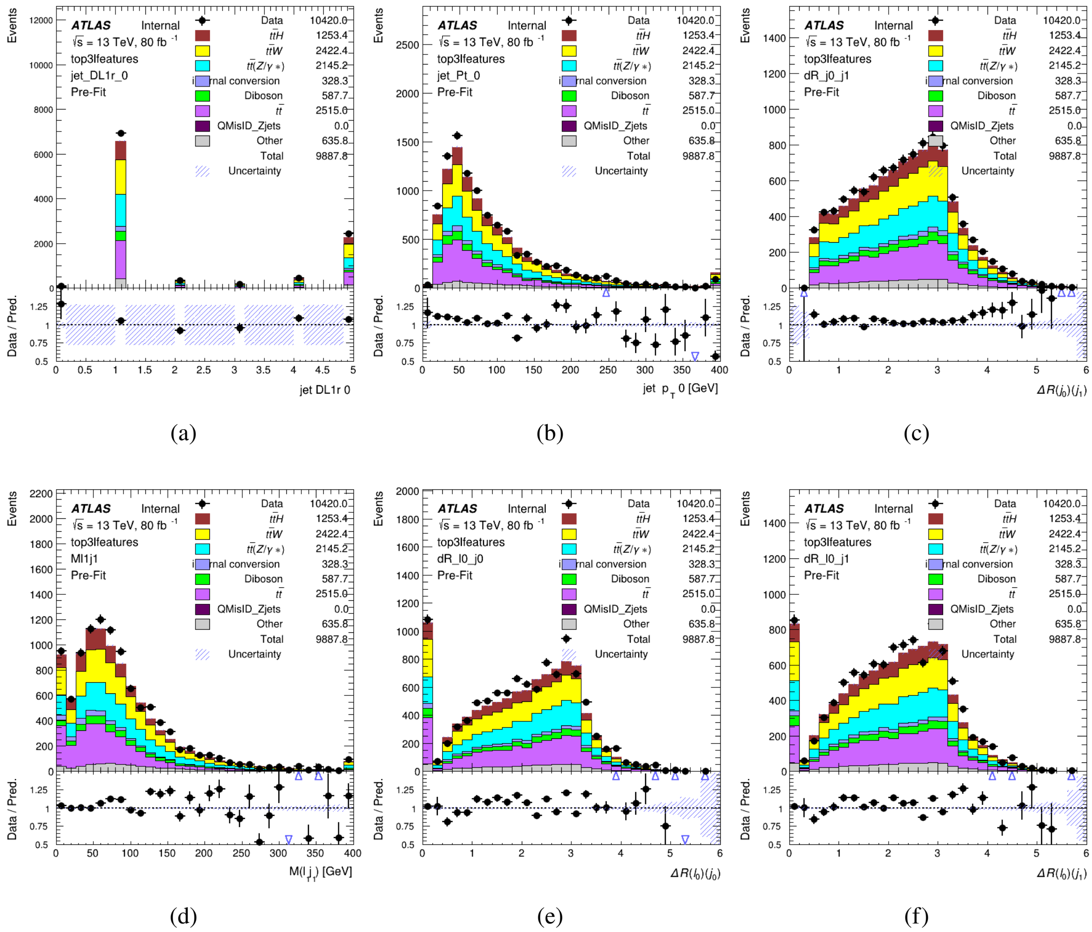
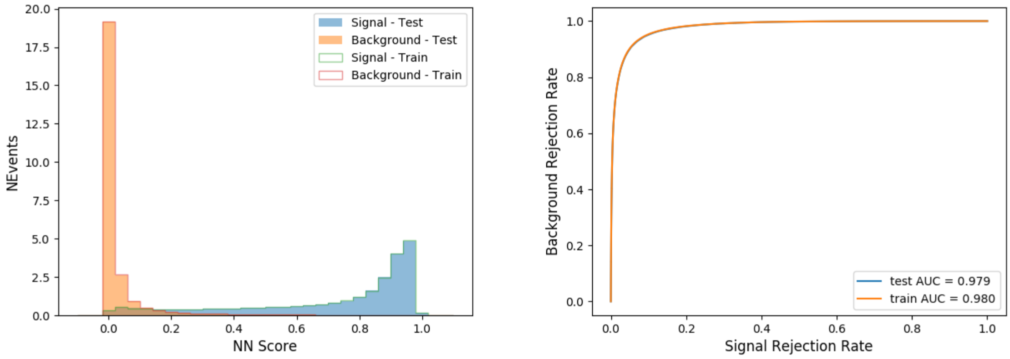


Figure 15.5: Data/MC comparisons of input features for top3l training for 79.8 fb^{-1} of data. (a) show the DL1r WP of je t 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

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 15.2.2.



(a) (b)

Figure 15.6: 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

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 18, once again compared to the naive approach described above.

Table 18: 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%

15.3 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 $2\ell SS$ 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 (and two neutrinos which are not reconstructed).

3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In the semi-leptonic case the reconstructed decay products consist of two jets and a single lepton. For the fully-leptonic case, the decay products include 2 of the three leptons associated with the event. For training these models, events are separated into these two categories using truth level information. A separate MVA, described in Section 15.5, is used to make this distinction at recontructed level, and determine which model to use.

For all channels, the models described in Section 15.2 are used to identify b-jet candidates, whose kinematics are used as additional input features to help 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.

15.3.1 2lSS Channel

For the $2\ell SS$ 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 15.2, with the specific input features listed in Table 19.

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

Table 19: Input features used to identify the Higgs decay products in 2lSS events

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

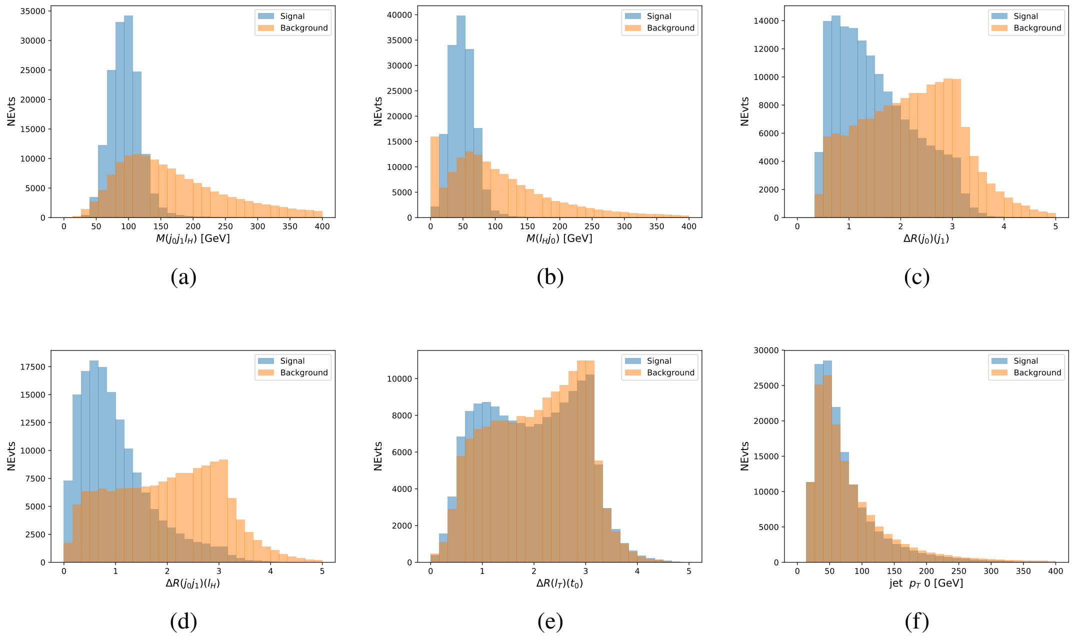
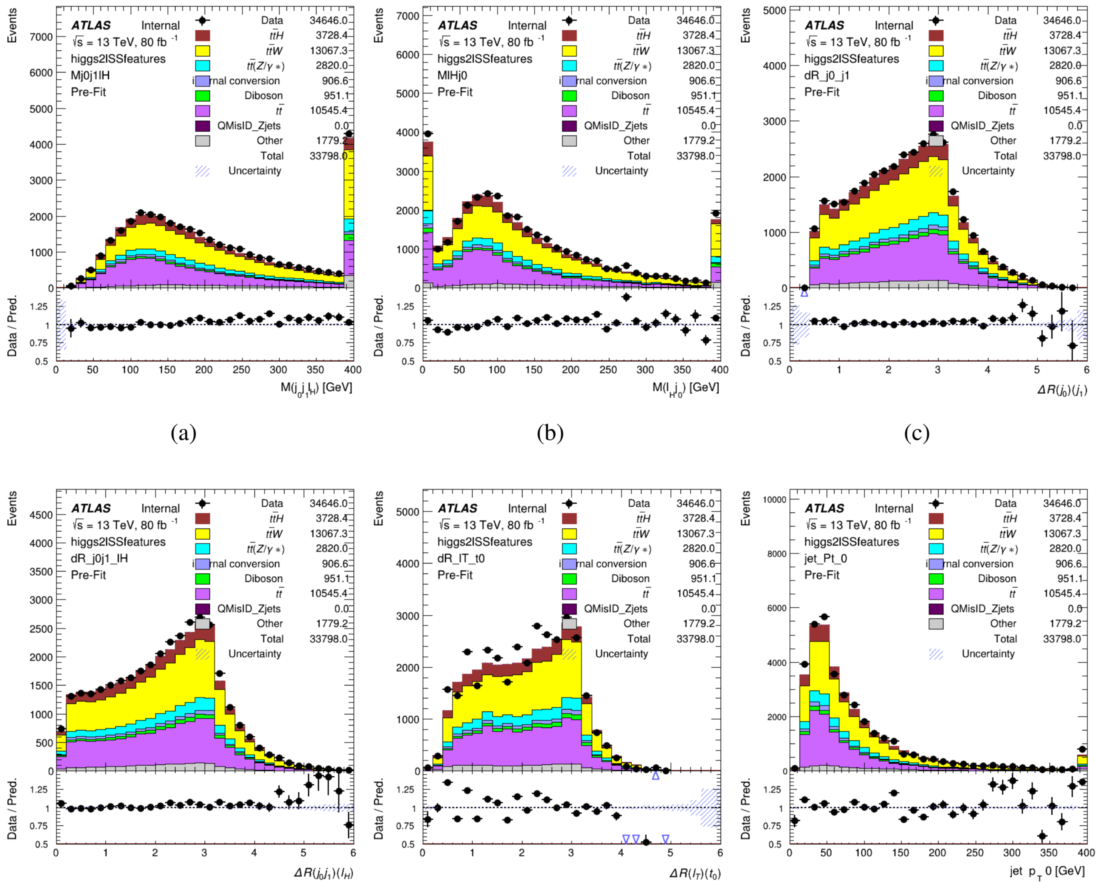


Figure 15.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 and the lepton candidate, (e) ΔR between the lepton from the top and the leading b-jet, (f) the p_T of jet 0.

The modeling of these inputs is validated against data, with Figure 15.2 showing good general agreement between data and MC. Plots for the complete list of features can found in Section C.

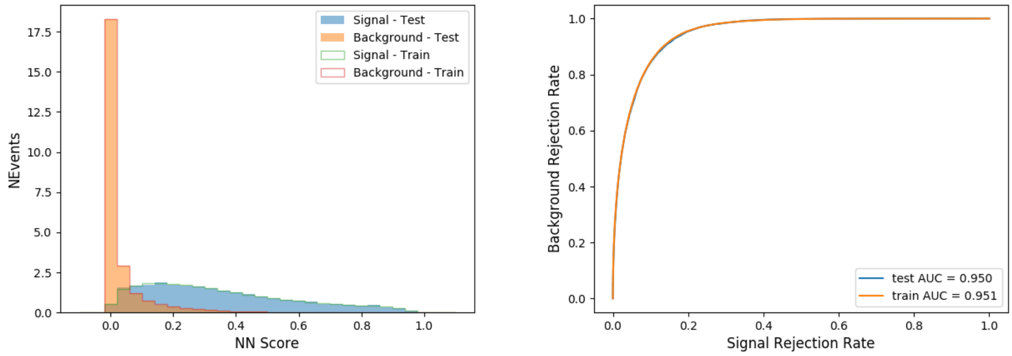


(a) (b) (c)

(d) (e) (f)

Figure 15.8: Data/MC comparisons of input features for higgs2lSS training for 79.8 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 and the lepton candidate, (e) ΔR between the lepton from the top and the leading b-jet, (f) the p_T of jet 0.

A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2 million events, with an additional 200,000 reserved for testing the model. In order to compensate for the large number of incorrect combinations, these have been downsampled such that the correct combinations represent over 10% of the training set. The output of the NN is summarized in Figure 15.3.1.



(a) (b)

Figure 15.9: Result of the Higgs reconstruction algorithm in the $2\ell SS$ 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

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.

15.3.2 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.

Lepton p_T H	Lepton p_T T_0	Lepton p_T T_1
jet p_T 0	jet p_T 1	top p_T 0
top p_T 1	jet η 0	jet η 1
jet ϕ 0	jet ϕ 1	$\Delta R(j_0)(j_1)$
$M(j_0 j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1)(l_{T_1})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0 j_1 l_{T_0})$	$M(j_0 j_1 l_{T_1})$
$M(j_0 j_1 l_H)$	$\Delta R(j_0 j_1 l_H)(l_{T_0})$	$\Delta R(j_0 j_1 l_H)(l_{T_1})$
$\Delta\phi(j_0 j_1 l_H)(E_T^{miss})$	$p_T(j_0 j_1 l_H l_{T_0} l_{T_1} b_0 b_1 E_T^{miss})$	$M(j_0 j_1 b_0)$
$M(j_0 j_1 b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
b-jet Reco Score	E_T^{miss}	HT jets
nJets		

Table 20: Input features used to identify the Higgs decay products in $3l$ semi-leptonic events

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

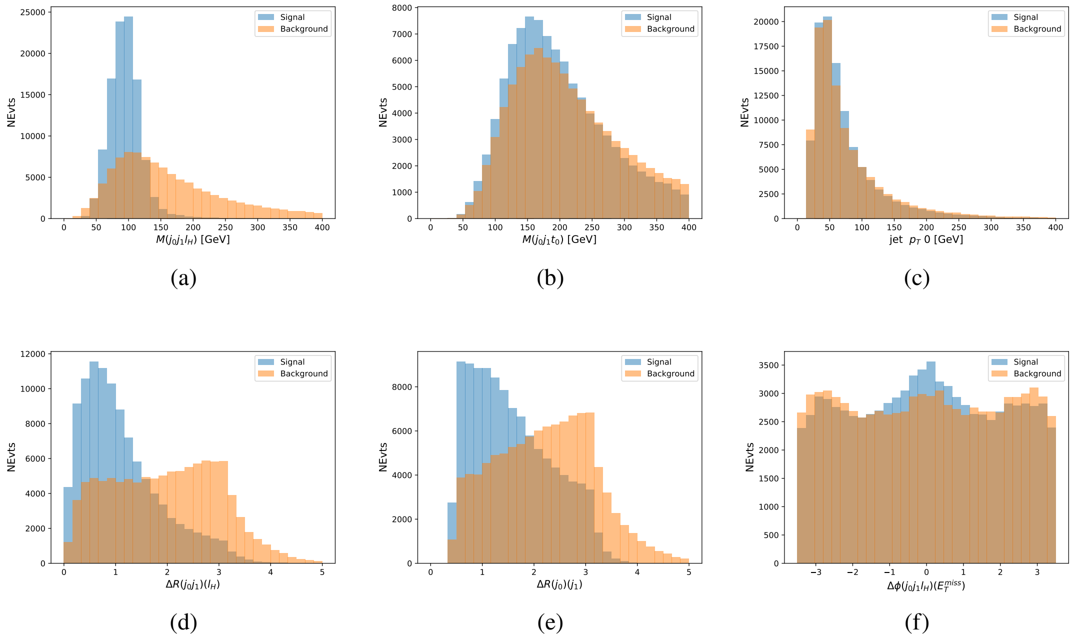


Figure 15.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

The modeling of these inputs is validated against data, with Figure 15.11 showing good general agreement between data and MC. Plots for the complete list of features can found in appendix C.1.

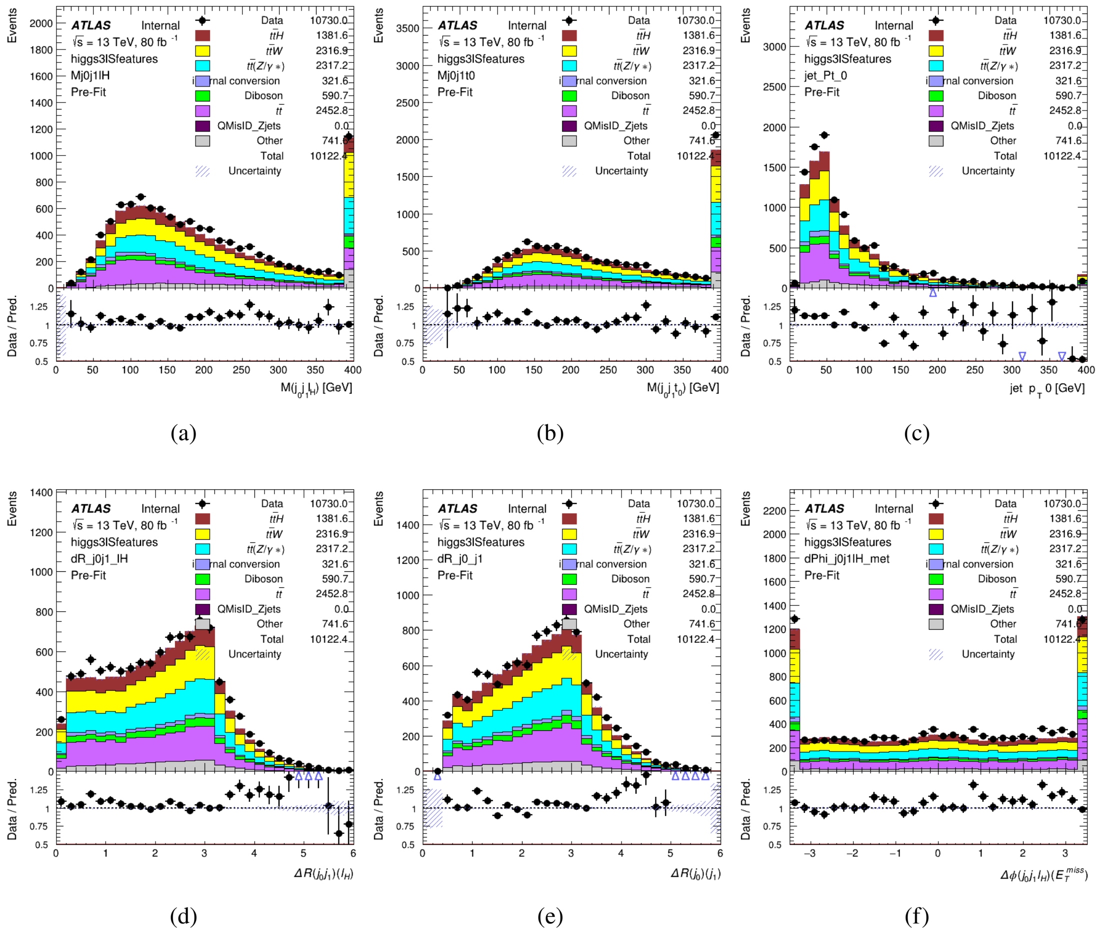
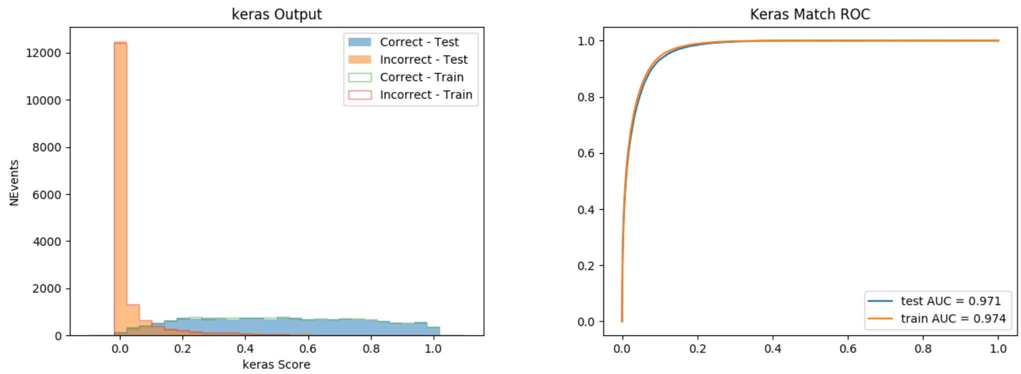


Figure 15.11: Data/MC comparisons of input features for higgs3lS training for 79.8 fb^{-1} of data.

A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events. Once again, incorrect combinations are downsampled, such that the correct combinations are around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN is summarized in Figure 15.3.2.



(a) (b)

Figure 15.12: 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

The neural network identifies the correct combination 64% 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 83% of the time.

15.3.3 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.

Lepton p_T H_1	Lepton p_T H_0	Lepton p_T T
top p_T 0	top p_T 1	$\Delta\phi(l_{H_1})(E_T^{miss})$
$\Delta\phi(l_T)(E_T^{miss})$	$M(l_{H_0}l_{H_1})$	$M(l_{H_1}l_T)$
$M(l_{H_0}l_T)$	$\Delta R(l_{H_0})(l_{H_1})$	$\Delta R(l_{H_1})(l_T)$
$\Delta R(l_{H_0})(l_T)$	$\Delta R(l_{H_0}l_{H_1})(l_T)$	$\Delta R(l_{H_0}l_T)(l_{H_1})$
$\Delta R(l_{H_0}l_{H_1})(b_0)$	$\Delta R(l_{H_0}l_{H_1})(b_1)$	$\Delta R(l_{H_0}b_0)$
$M(l_{H_0}b_0)$	$\Delta R(l_{H_0}b_1)$	$M(l_{H_0}b_1)$
$\Delta R(l_{H_1}b_0)$	$M(l_{H_1}b_0)$	$\Delta R(l_{H_1}b_1)$
$M(l_{H_1}b_1)$	E_T^{miss}	b-jet Reco Score

Table 21: Input features used to identify the Higgs decay products in 3lF events

Table 22: Input features used to identify the Higgs decay products in $3l$ fully leptonic events

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

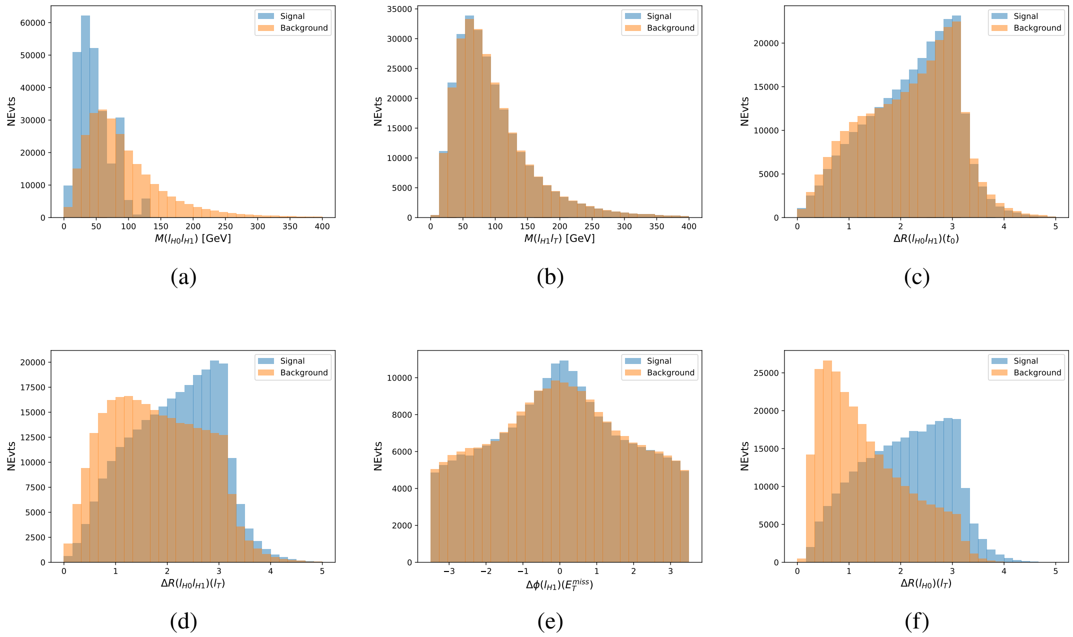


Figure 15.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

The modeling of these inputs is validated against data, with Figure 15.14 showing good general agreement between data and MC. Plots for the complete list of features can found in Section C.

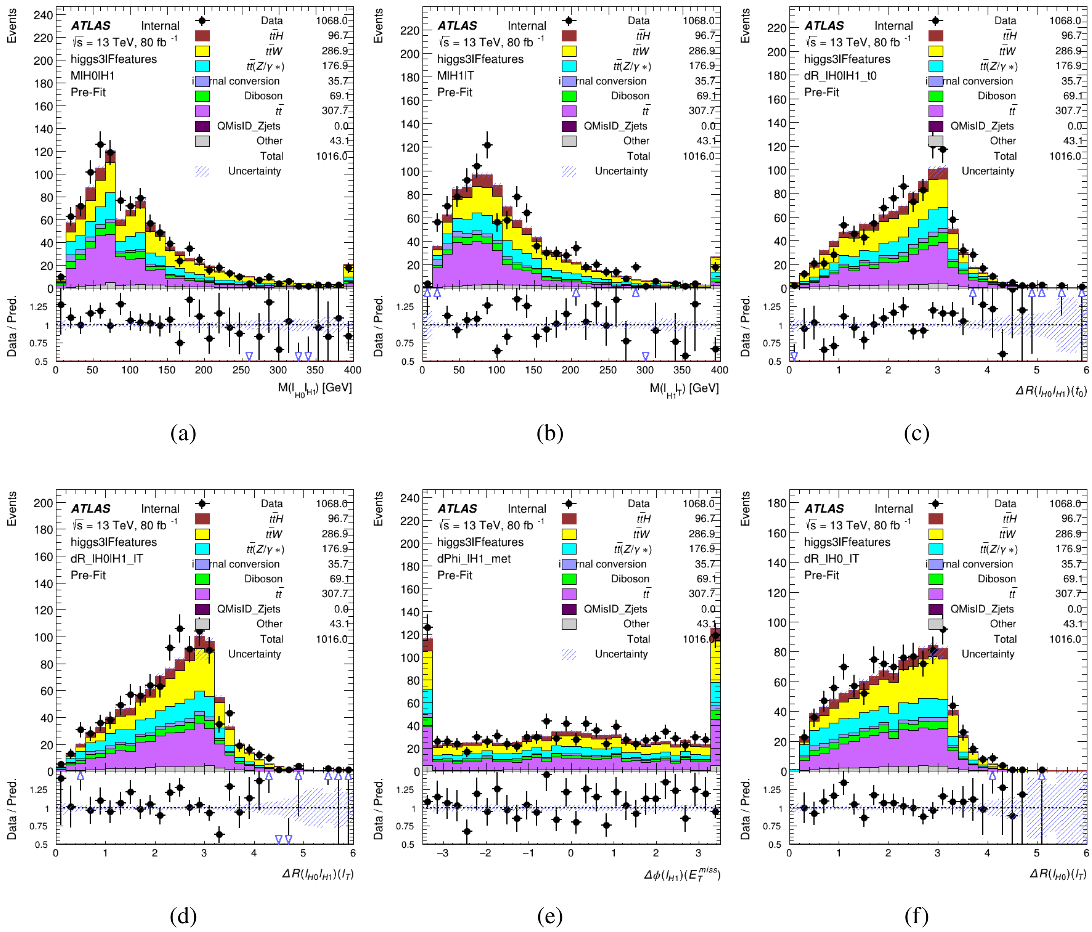
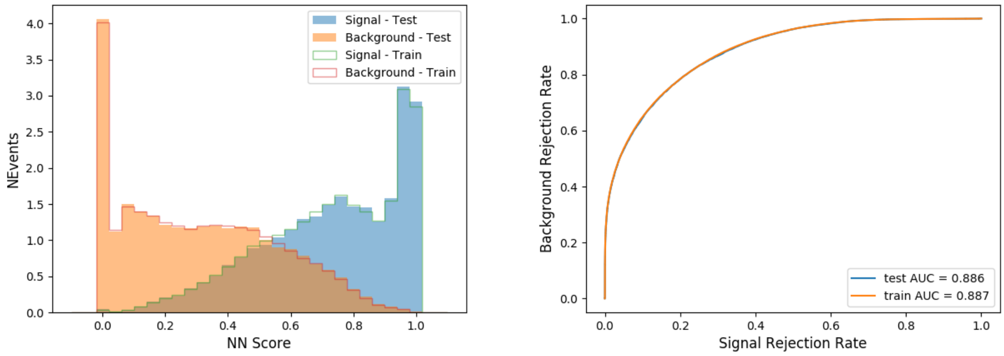


Figure 15.14: Data/MC comparisons of input features for higgs3lF training for 79.8 fb^{-1} of data.

1119 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,
1120 with 10% of the dataset reserved for testing. The output of the model is summarized in Figure
1121 15.3.3.



(a) (b)

Figure 15.15: (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

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

15.4 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 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 could be treated as a binary classification problem, rather than a regression problem. This approach is explored in detail in Section C.5, 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.

15.4.1 2lSS Channel

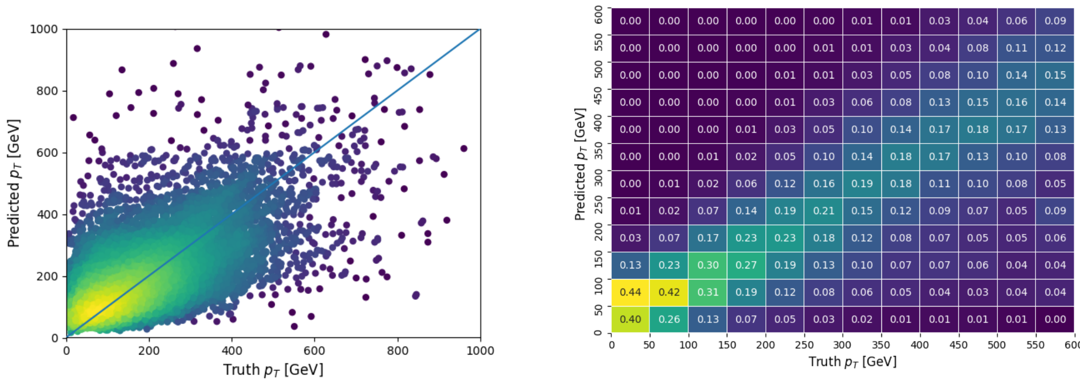
The input variables listed in Table 23 are used to predict the Higgs p_T in the $2lSS$ channel. Here j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating from the Higgs is labeled l_H, while the other lepton is labeled l_T, as it is assumed to have come from the decay of one of the top quarks. 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.

HT	$M(j_0j_1)$	$M(j_0j_1l_H)$
$M(l_Hj_0)$	$M(l_Hj_1)$	$p_T(b_0b_1)$
$p_T(j_0j_1l_H)$	$\Delta\phi(j_0j_1l_H)(E_T^{miss})$	$\Delta R(j_0)(j_1)$
$\Delta R(j_0j_1)(l_H)$	$\Delta R(j_0j_1l_H)(l_T)$	$\Delta R(j_0j_1l_H)(b_0)$
$\Delta R(j_0j_1l_H)(b_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(b_0)$
$\Delta R(l_H)(b_1)$	$\Delta R(l_T)(b_0)$	$\Delta R(l_T)(b_1)$
$\Delta R(b_0)(b_1)$	Higgs Reco Score	jet η 0
jet η 1	jet Phi 0	jet Phi 1
jet p_T 0	jet p_T 1	Lepton η H
Lepton ϕ H	Lepton p_T H	Lepton p_T T
E_T^{miss}	nJets	b-jet Reco Score
b-jet p_T 0	b-jet p_T 1	

Table 23: Input features for reconstructing the Higgs p_T spectrum for $2lSS$ events

The optimal neural network architecture for this channel is found to consist of 7 hidden layers with 60 nodes each. The input data 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.

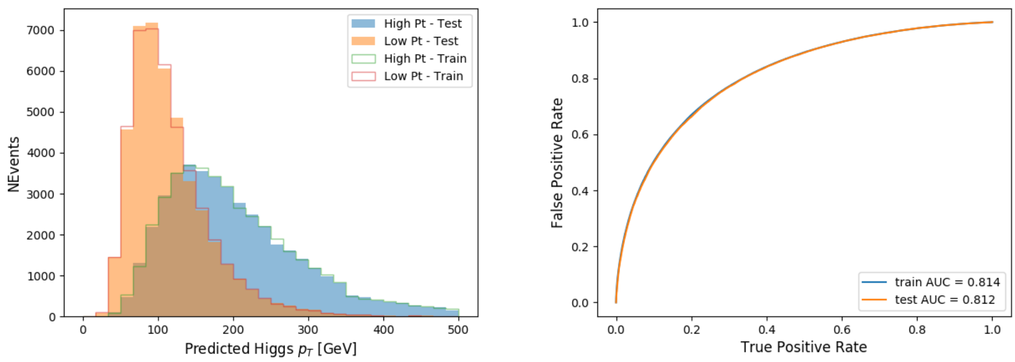
To evaluate the performance of the model, the predicted p_T spectrum is compared to the truth Higgs p_T in Figure 15.16. In order to visualize the model performance more clearly, in (a) of that figure, the color of each point is determined by Kernal Density Estimation (KDE). The color shown represents the logarithm of the output from KDE, to counteract the large number of low p_T events. For that same reason, each column of the histogram shown in (b) of Figure 15.16 is normalized to unity. This plot therefore demonstrates what the model predicts for each slice of truth p_T.



(a) (b)

Figure 15.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized to one.

We are also interested in how well the model distinguishes between events with p_T <150 GeV and >150 GeV. Figure 15.17 demonstrates the NN output for high and low p_T events based on this cutoff.



(a) (b)

Figure 15.17: (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.

15.4.2 3l Semi-leptonic Channel

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

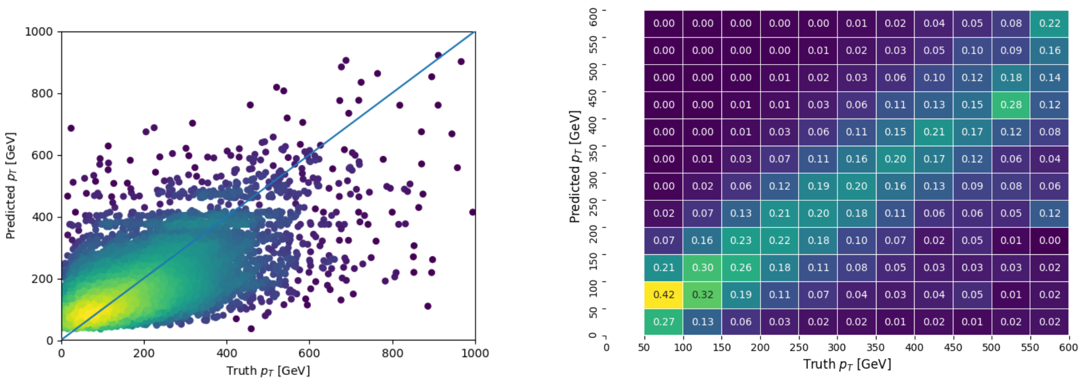
HT jets	MET	$M(j_0j_1)$
$M(j_0j_1l_H)$	$M(j_0j_1l_{T0})$	$M(j_0j_1l_{T1})$
$M(j_0j_1b_0)$	$M(j_0j_1b_1)$	$M(b_0l_{T0})$
$M(b_0l_{T1})$	$M(b_1l_{T0})$	$M(b_1l_{T1})$
$\Delta\phi(j_0j_1l_H)(E_T^{miss})$	$\Delta R(j_0)(j_1)$	$\Delta R(j_0j_1)(l_H)$
$\Delta R(j_0j_1)(l_{T1})$	$\Delta R(j_0j_1)(b_0)$	$\Delta R(j_0j_1)(b_1)$
$\Delta R(j_0j_1l_H)(l_{T0})$	$\Delta R(j_0j_1l_H)(l_{T1})$	$\Delta R(j_0j_1l_H)(b_0)$
$\Delta R(j_0j_1l_H)(b_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(l_H)(l_{T1})$	$\Delta R(l_{T0})(l_{T1})$	$\Delta R(l_{T0})(b_0)$
$\Delta R(l_{T0})(b_1)$	$\Delta R(l_{T1})(b_0)$	$\Delta R(l_{T1})(b_1)$
Higgs Reco Score	jet η 0	jet η 1
jet ϕ 0	jet ϕ 1	jet p_T 0
jet p_T 1	Lepton η H	Lepton ϕ H
Lepton p_T H	Lepton p_T T0	Lepton p_T T1
nJets	b-jet Reco Score	b-jet p_T 0
b-jet p_T 1		

Table 24: 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 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.

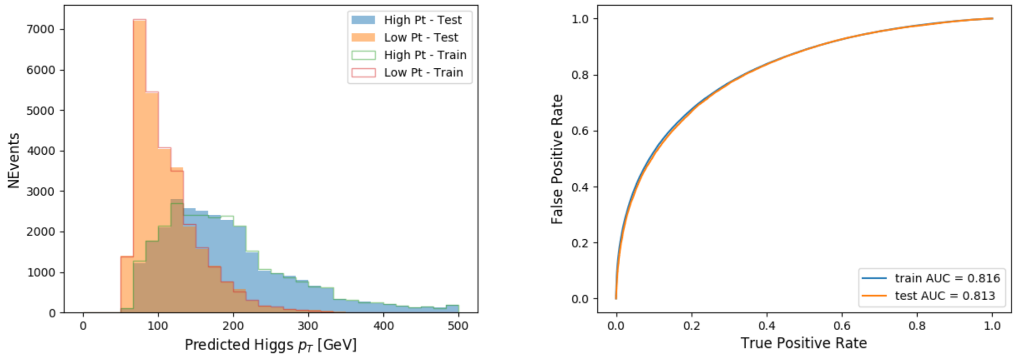
To evaluate the performance of the model, the predicted p_T spectrum is compared to the truth Higgs p_T in Figure 15.18. Once again, (a) of 15.18 shows a scatterplots of predicted vs truth p_T, where the color of each point corresponds to the log of the relative KDE at that point. Each column of the the histogram in (b) is normalized to unity, to better demonstrate the output of the NN for each slice of truth p_T.



(a) (b)

Figure 15.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized to one.

Figure 15.19 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.



(a) (b)

Figure 15.19: (
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.

15.4.3 3l Fully-leptonic Channel

The features listed in 25 are used to construct a model for predictin the Higgs p_T for $3lF$ events.

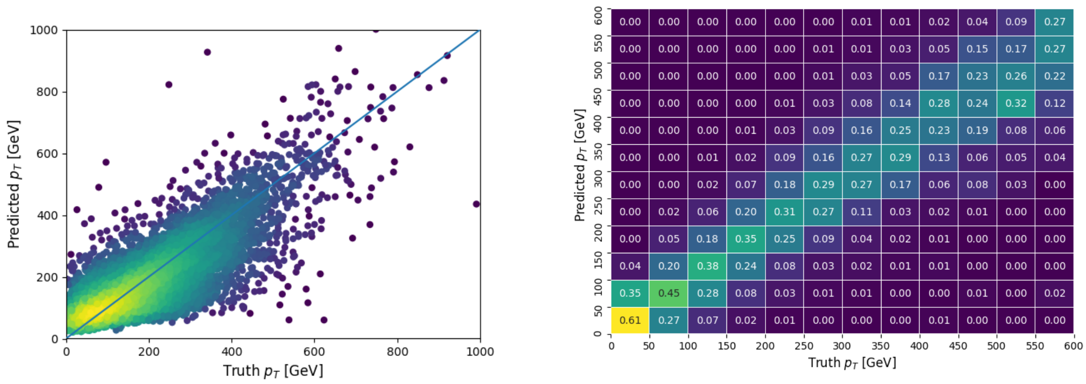
HT	$M(l_{H0}l_{H1})$	$M(l_{H0}l_T)$
$M(l_{H0}b_0)$	$M(l_{H0}b_1)$	$M(l_{H1}l_T)$
$M(l_{H1}b_0)$	$M(l_{H1}b_1)$	$\Delta R(l_{H0})(l_{H1})$
$\Delta R(l_{H0})(l_T)$	$\Delta R(l_{H0}l_{H1})(l_T)$	$\Delta R(l_{H0}l_T)(l_{H1})$
$\Delta R(l_{H1})(l_T)$	$\Delta R(l_{H0}b_0)$	$\Delta R(l_{H0}b_1)$
$\Delta R(l_{H1}b_1)$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

Table 25: Input features for reconstructing the Higgs p_T spectrum for $3lF$ events

l_{H0} and l_{H1} respresent the two leptons identified by the Higgs reconstruction model as originating from the Higgs, while l_T is the other lepton in the event. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and 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 400,000 events, 10% of which is used for testing, the other 90% for training. Training is found to converge after around 150 epochs.

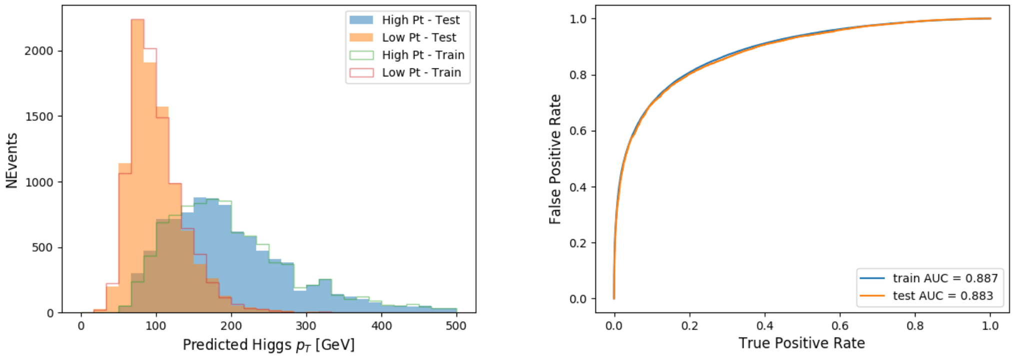
The predicted transverse momentum, as a function of the truth p_T, is shown in Figure 15.20.



(a) (b)

Figure 15.20: The regressed Higgs momentum spectrum as a function of the truth p_T for $3lF$ $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized to one.

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



(a) (b)

Figure 15.21: (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.

15.5 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.

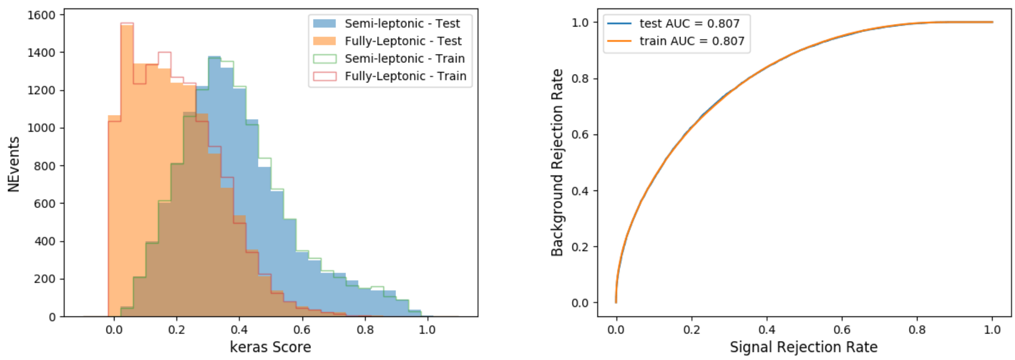
The kinematics of each event, along with the output scores of the Higgs and top reconstruction algorithms, are used to distinguish these two possible decay modes. The particular inputs used are listed in Table 26.

HT jets	$M(l_0t_0)$	$M(l_0t_1)$
$M(l_1t_0)$	$M(l_1t_1)$	$M(l_0l_1)$
$M(l_0l_2)$	$M(l_1l_2$	$\Delta R(l_0t_0)$
$\Delta R(l_0t_1)$	$\Delta R(l_1t_0)$	$\Delta R(l_1t_1)$
$\Delta R(ll_01)$	$\Delta R(ll_02)$	$\Delta R(ll_12)$
Lepton η 0	Lepton η 1	Lepton η 2
Lepton ϕ 0	Lepton ϕ 1	Lepton ϕ 2
Lepton p_T 0	Lepton p_T 1	Lepton p_T 2
E_T^{miss}	nJets	nJets OR DL1r 60
nJets OR DL1r 85	score3lF	score3lS
topScore	total charge	

Table 26: Input features used to distinguish semi-leptonic and fully-leptonic Higgs decays in the $3l$ channel.

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

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 15.22.



(a) (b)

Figure 15.22: (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.

A cutoff of 0.23 is determined to be optimal for separating $3\ell S$ and $3\ell F$ in the fit.

16 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 two leptons originated from the Higgs boson as well as events where only one of the leptons originated from the Higgs. This motivates splitting the 3ℓ channel into semi-leptonic, and fully leptonic channels, after an event preselection has been applied.

16.1 Pre-MVA Event Selection

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

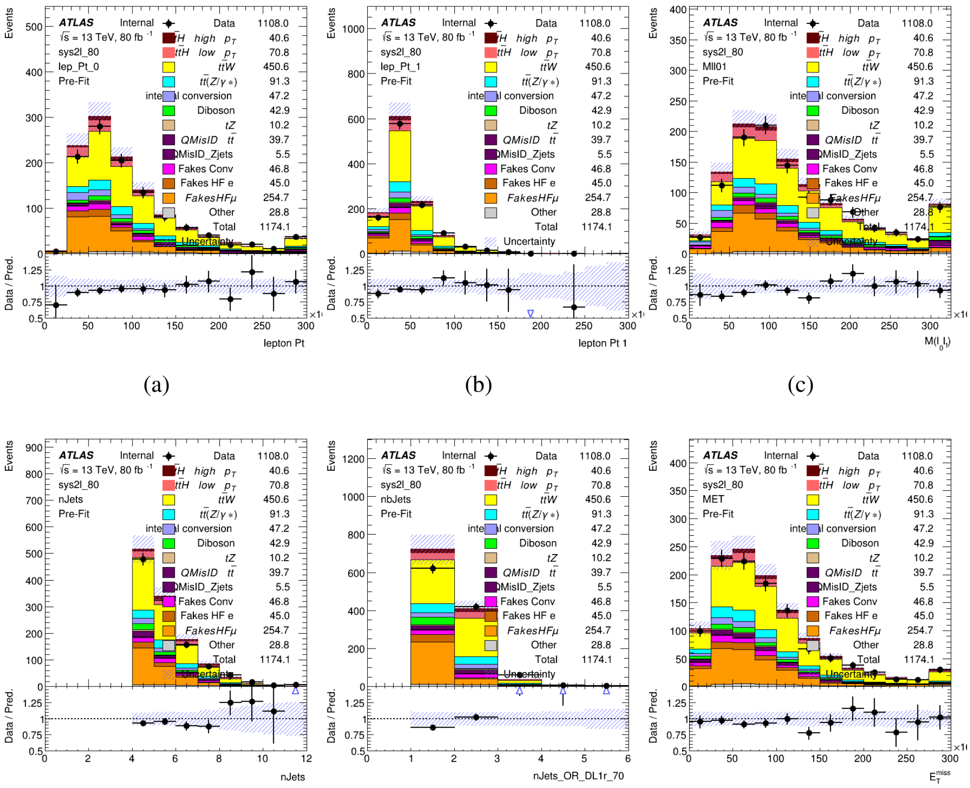
- Two very tight, same-charge, light leptons with $p_T > 20$ GeV
- $\geq$4 reconstructed jets, $\geq$1 b-tagged jets
- No reconstructed tau candidates

The event yield after the $2\ell SS$ preselection has been applied, for MC and data at 79.8 fb^{-1}, is shown in Table 16.1.

Process	Yield
$t\bar{t}H$ high p_T	41 ± 5
$t\bar{t}H$ low p_T	71 ± 8
$t\bar{t}W$	450 ± 70
$t\bar{t}(Z/\gamma^*)$	91 ± 11
$t\bar{t}ll$ low mass	10 ± 6
Rare Top	20 ± 12
VV	42 ± 22
tZ	10 ± 5
QMisID	44.7 ± 2.7
Fakes int. conv	47 ± 26
Fakes ext. conv	46 ± 44
Fakes HF e	45 ± 23
Fakes HF μ	250 ± 50
Three top	2.2 ± 1.1
Four top	5.64 ± 0.31
$t\bar{t}WW$	10.9 ± 0.6
tW	0.0 ± 0.0
WtZ	9.1 ± 0.8
VVV	0.30 ± 0.05
VH	0.6 ± 1.0
Total	1170 ± 120
Data	1108

Table 27: Yields of the $2\ell SS$ preselection region

1220 Figure 16.1. Good general agreement is found.



(a) (b) (c)

(d) (e) (f)

Figure 16.1: Data/MC comparisons of the $2\ell SS$ pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

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

The event yield after the $3l$ preselection has been applied, for MC and data at 79.8 fb^{-1}, is shown in Table 16.1.

Process	Yield
$t\bar{t}H$ high p_T	20.5 ± 2.3
$t\bar{t}H$ low p_T	33.6 ± 3.8
$t\bar{t}W$	138 ± 18
$t\bar{t}Z/\gamma$	80 ± 9
$t\bar{t}lllowmass$	3.5 ± 2.0
rareTop	22 ± 12
VV	39 ± 19
tZ	9.2 ± 4.5
QMisID	1.8 ± 0.6
Fakes int. conv	31 ± 17
Fakes ext. conv	14 ± 11
Fakes HF e	20 ± 10
Fakes HF μ	102 ± 22
Three top	0.96 ± 0.48
Four top	6.17 ± 0.35
$t\bar{t}WW$	5.46 ± 0.33
tW	0.0 ± 0.0
WtZ	8.7 ± 0.6
VVV	0.81 ± 0.11
VH	0.0 ± 0.0
Total	512 ± 48
Data	535

Table 28: Yields of the $3l$ preselection region.

Table 29: Yields of the $3l$ preselection region.

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

16.2.

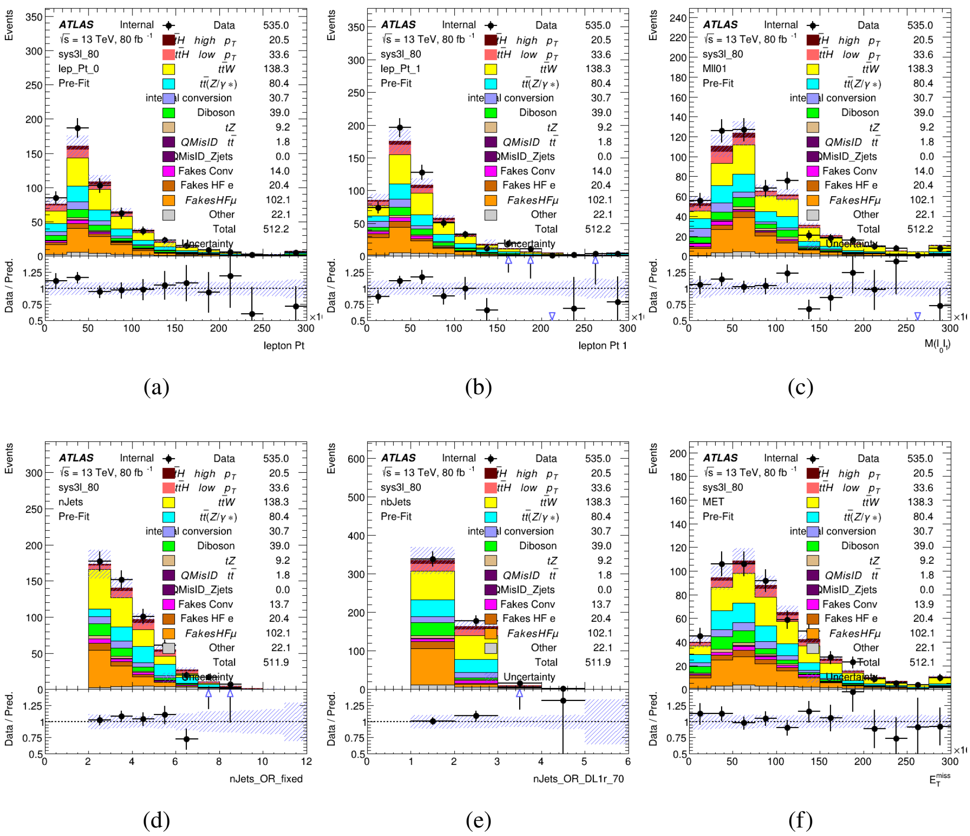


Figure 16.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

16.2 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). Here events with three leptons are split into 3lS and 3lF based on the model described in 15.5. In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

[34] 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 C.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	topRecoScore	

Table 30: Input features used to distinguish signal and background events in the 2lSS channel.

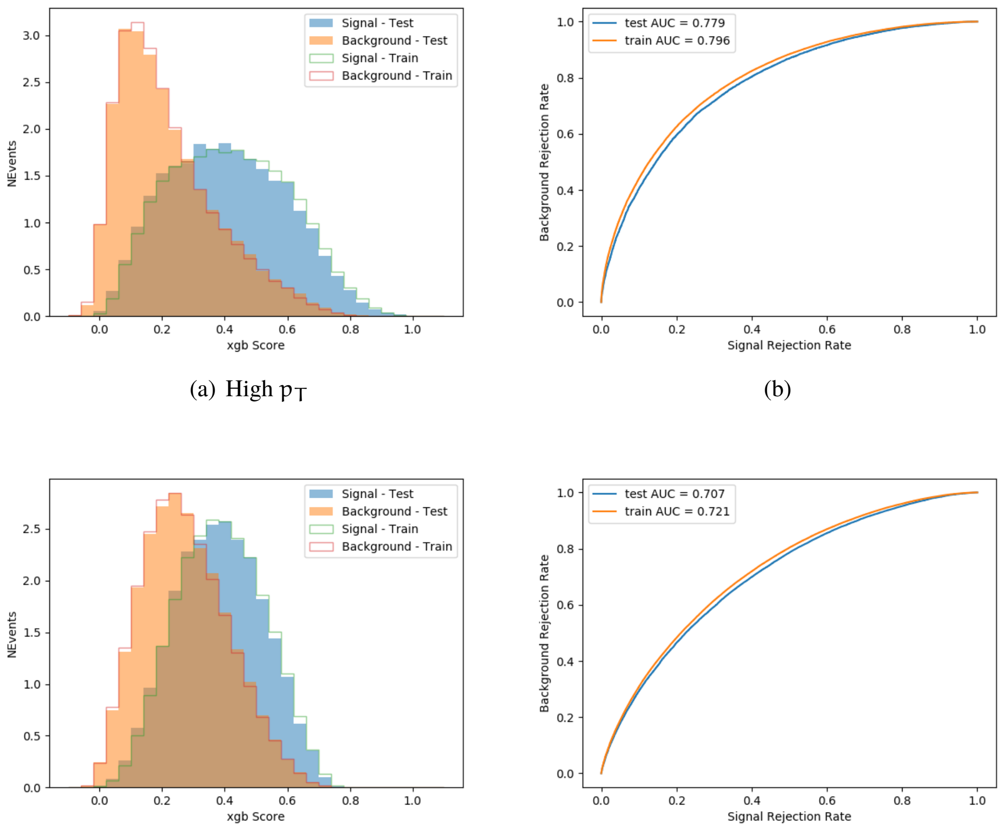
While for each of the $3l$ BDTs, the features listed below are used for training:

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	topScore

Table 31: Input features used to distinguish signal and background events in the 3l channel.

Modelling of each of these input features is verified in Appendix C.2 by comparing data and MC for 79.8 fb^{-1}. The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss function. The BDT response distribution and ROC curve for each model is shown in Figures 16.3-16.5.

2ℓSS

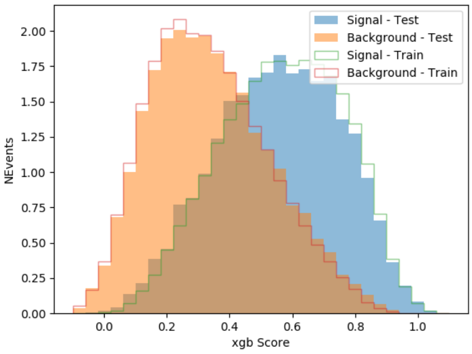


(a) High p_T (b)

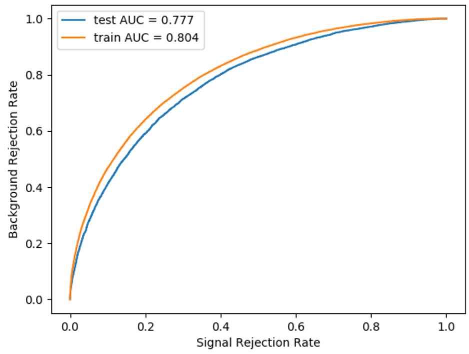
(c) Low p_T (d)

Figure 16.3: Output BDT scores of training and testing data for signal (blue) and background (orange) for 2ℓSS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T. (b) and (d) show the ROC curve for the 2ℓSS high and low p_T models, respectively.

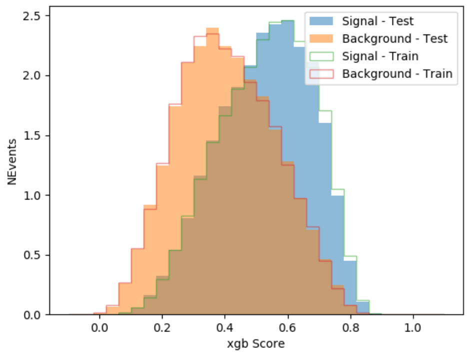
3l - Semileptonic



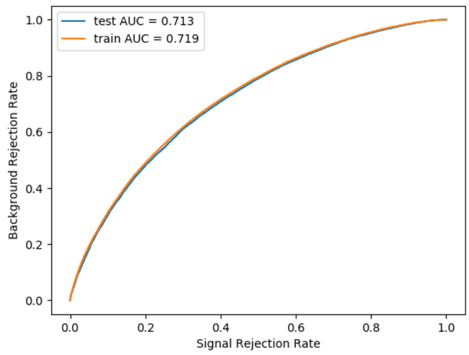
(a) High p_T



(b)



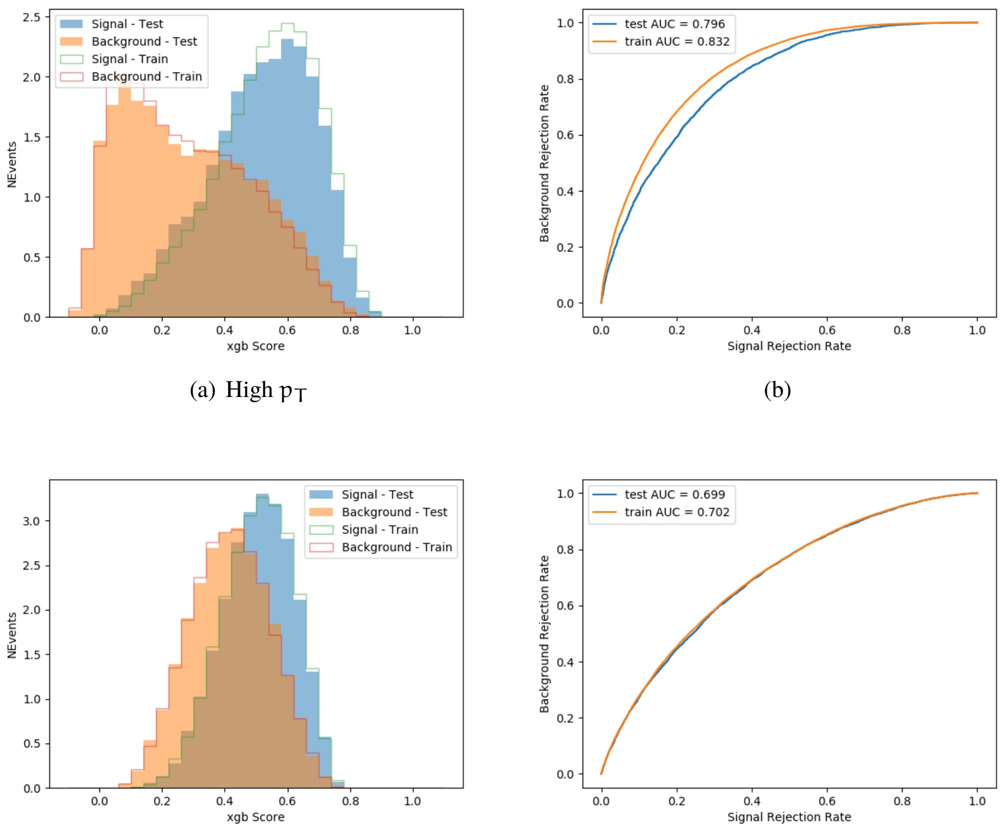
(c) Low p_T



(d)

Figure 16.4: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T. (b) and (d) show the ROC curve for the 3lS high and low p_T models, respectively.

3l - Fully Leptonic



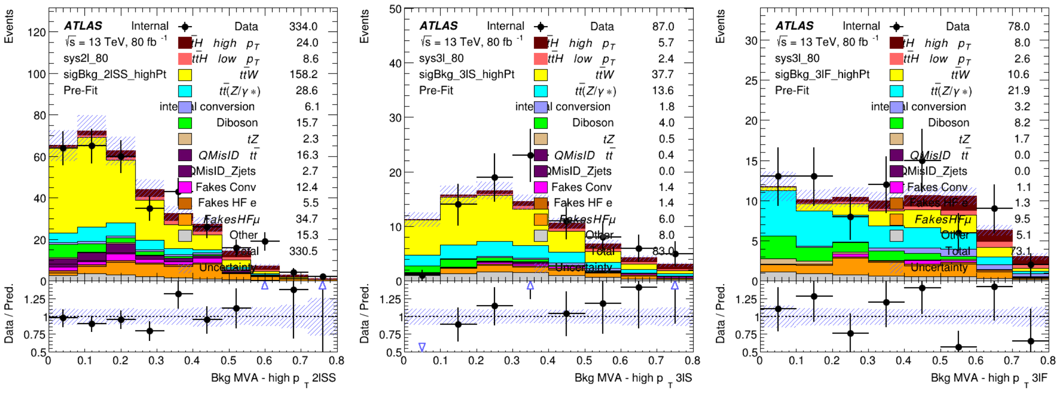
(a) High p_T (b)

(c) Low p_T (d)

Figure 16.5: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lF events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T. (b) and (d) show the ROC curve for the 3lF high and low p_T models, respectively.

Output distributions of each MVA, comparing MC predictions to data at 79.8 fb^{-1} are shown in figures 16.6-16.2.

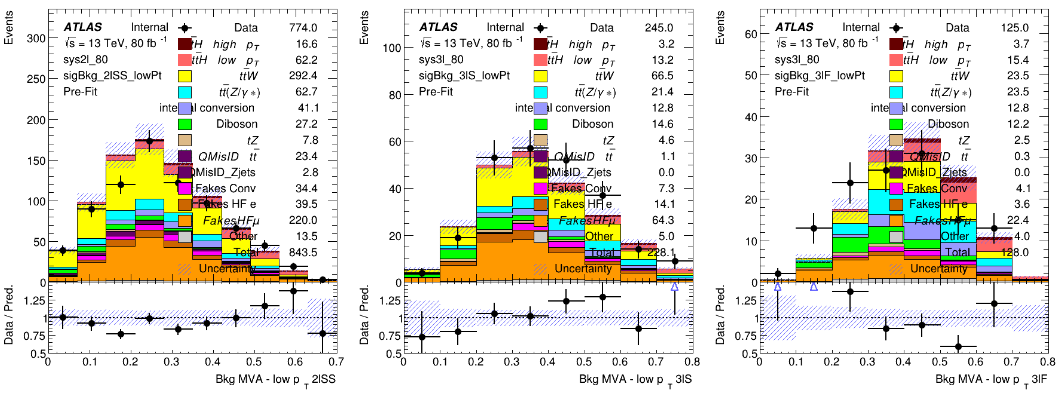
High p_T Background Rejection BDTs



(a) (b) (c)

Figure 16.6: Output score of the high p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

Low p_T Background Rejection BDTs



(a) (b) (c)

Figure 16.7: Output score of the low p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

16.3 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 15.5 is used to separate the three channel into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson. This leaves three orthogonal signal regions - $2\ell SS$, $3\ell S$, and $3\ell F$.

For each event, depending on the number of leptons as well as whether the p_T of the Higgs is predicted to be high ($>$ 150 GeV) or low ($<$ 150 GeV), a cut on the appropriate background rejection MVA is applied. The particular cut values, listed in Table 32, are determined by maximizing $S/\sqrt{B}$ in each region.

Channel	BDT Score
$2\ell SS$ high p_T	0.36
$2\ell SS$ low p_T	0.34
$3\ell S$ high p_T	0.51
$3\ell S$ low p_T	0.43
$3\ell F$ high p_T	0.33
$3\ell F$ low p_T	0.41

Table 32: Cutoff values on background rejection MVA score applied to signal regions.

The event preselection and MVA selection listed in Table 32 are used define the three signal regions used in the fit. These signal region definitions are summarized in Table 33.

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

Table 33: Selection applied to define the three signal regions used in the fit.

17 Systematic Uncertainties

The systematic uncertainties that are considered are summarized in Table 34. 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 18.

Table 34: 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

The uncertainty in the combined integrated luminosity is derived from a calibration of the luminosity scale using x-y beam-separation scans performed for 13 TeV proton-proton data [15], [37].

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 to the uncertainty in the jet energy scale [47] are decomposed into uncorrelated components and treated as independent sources in the analysis. This method decomposes the uncertainties into 30 nuiscance parameters included in the fit. A similar method is used to account for jet energy resolution (JER) uncertainties, and 8 JER uncertainty components are uncluded as NPs in the fit.

The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses [39] 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 for each of the b-tag Working Points considered in the analysis.

As mentioned in Section 13.2, a normalization corrections and uncertainties on the estimates of non-prompt leptons backgrounds are derived using data driven techniques, decribed in detail in [7]. These are derived from a likelihood fit over various non-prompt enriched control regions, targeting several sources of non-prompt light leptons separately: external conversion electrons, internal conversion electrons, electrons from heavy flavor decays, and muons from heavy flavor decays.

The normalization factor and uncertainty applied to each source of non-prompt leptons is summarized in Table 17

Processs	Normalization Factor
$\mathsf{NF}_{\mathsf{e}}^{\mathsf{ExtCO}}$	1.70 ± 0.51
$\mathsf{NF}_{\mathsf{e}}^{\mathsf{IntCO}}$	0.75 ± 0.26
$\mathsf{NF}_{\mathsf{e}}^{\mathsf{HF}}$	1.09 ± 0.32
$\mathsf{NF}_{\mu}^{\mathsf{HF}}$	1.28 ± 0.17

Table 35: Normalization factors - with statistical and systematic uncertainties - derived from the fit over fake control regions for each source of non-prompt leptons considered.

In addition to those derived from the control regions, several additional uncertainties are assigned to the non-prompt lepton background. An additional 25% uncertainty on material conversions is assigned, based on the comparison between data and MC in a region where a loose electron fails the photon conversion veto. A shape uncertainty of 15% (6%) is assigned to the HF non-prompt electron (muon) background based on a comparison between data and MC where the second leading electron (muon) is only required to be loose. As the contribution from light non-prompt leptons is small, about 10% percent of the contribution from HF non-prompt leptons, it is derived from the agreement between data and simulation in a LF enriched region at low values of the non-prompt lepton BDT. The resulting uncertainty is 100%, and is taken to be uncorrelated between internal and material conversions.

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

50% normalization uncertainty as well. The theory uncertainties applied to the MC estimates are summarized in Table 36.

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

Table 36: Summary of theoretical uncertainties for MC predictions in the analysis.

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

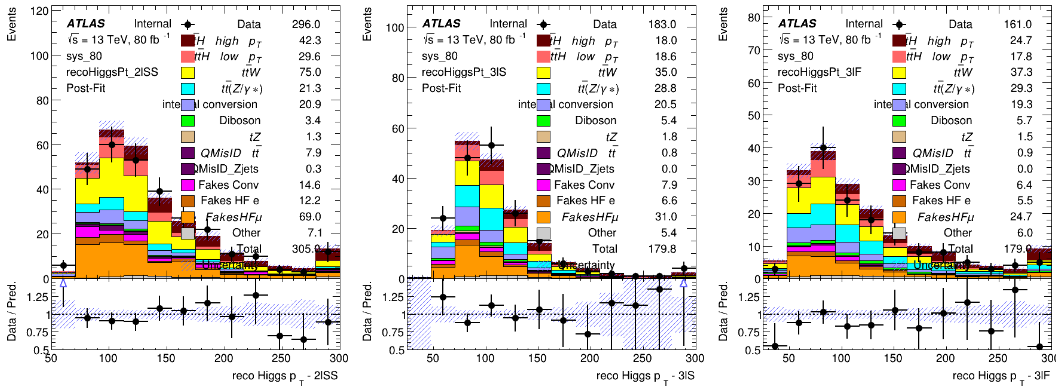
18 Results

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

As described in Section 17, there are 229 systematic uncertainties that are considered as NPs in the fit. These NP s are constrained by Gaussian or log-normal probability density functions. The latter are used for normalisation factors to ensure that they are always positive. The expected number of signal and background events are functions of the likelihood. The prior for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its nominal value.

18.1 Results - 79.8 fb^{-1}

As the data collected from 2015-2017 has been unblinded for $\mathrm{t\bar{t}H-ML}$ channels, representing 79.8 fb^{-1}, those events are unblinded. The predicted Higgs p_{T} spectrum is fit to data simultaneously in each of the three signal regions shown in Figure 18.1.



(a) (b) (c)

Figure 18.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) $2\ell SS$, (b) $3\ell S$, and (c) $3\ell F$, for 79.8 fb^{-1} of MC

A post-fit summary of the fitted regions is shown in Figure 18.2.

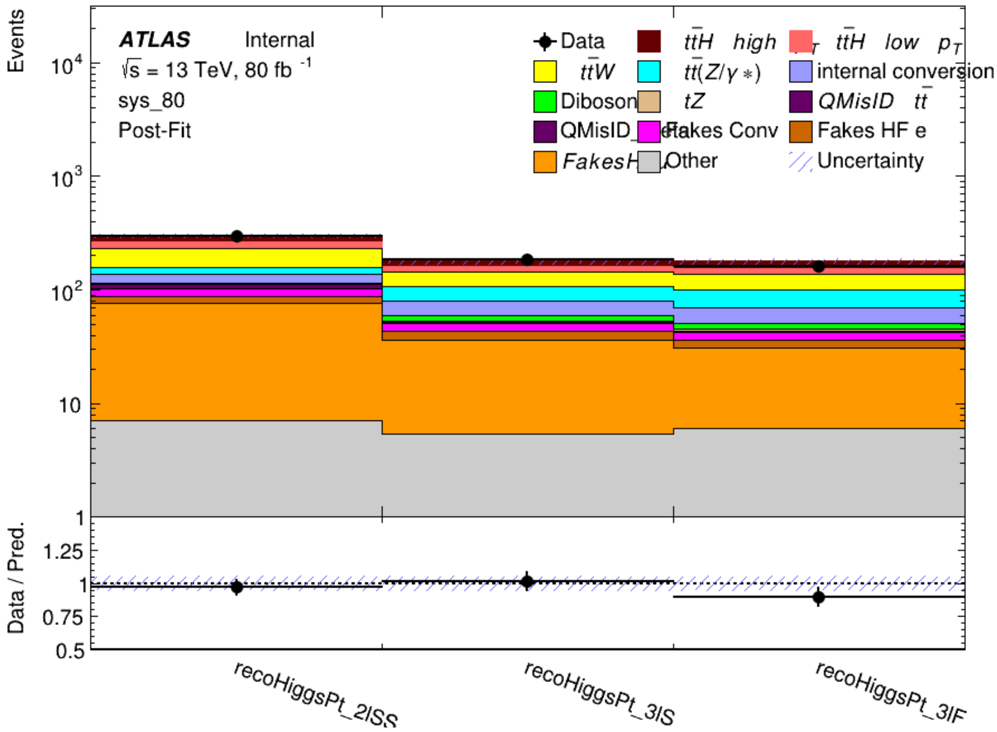


Figure 18.2: Post-fit summary of the yields in each signal region.

The the measured μ values for high and low p_T Higgs production obtained from the fit are shown in 37. A significance of 1.7σ is observed for $t\bar{t}H$ high p_T, and 2.1σ is measured for $t\bar{t}H$ low p_T.

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

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

Table 37: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T, where $\mu = \sigma_{obs}/\sigma_{pred}$

The most prominent sources of systematic uncertainty, as measured by their impact on $\mu_{t\bar{t}H \text{ high } p_T}$, are summarized in Table 38.

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

Table 38: Summary of the most significant sources of systematic uncertainty on the measurement of $t\bar{t}H$ high p_T.

The most significant sources of uncertainty on the measurement of $t\bar{t}H$ - low p_T are shown in Table 39.

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

Table 39: Summary of the most significant sources of systematic uncertainty on the measurement of $t\bar{t}H$ low p_T.

The ranking and impact of those nuisance parameters with the largest contribution to the overall uncertainty is shown in Figure 18.3.

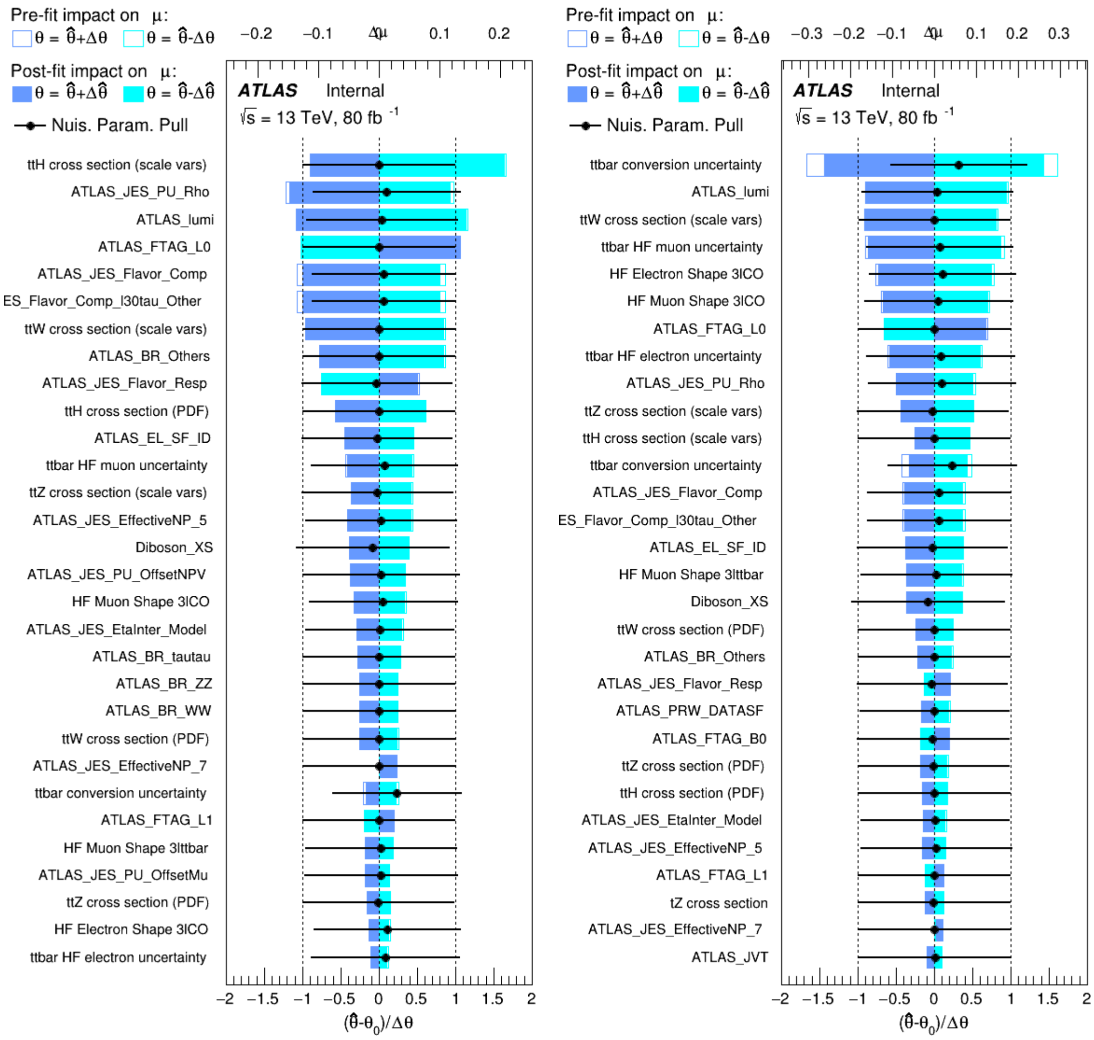
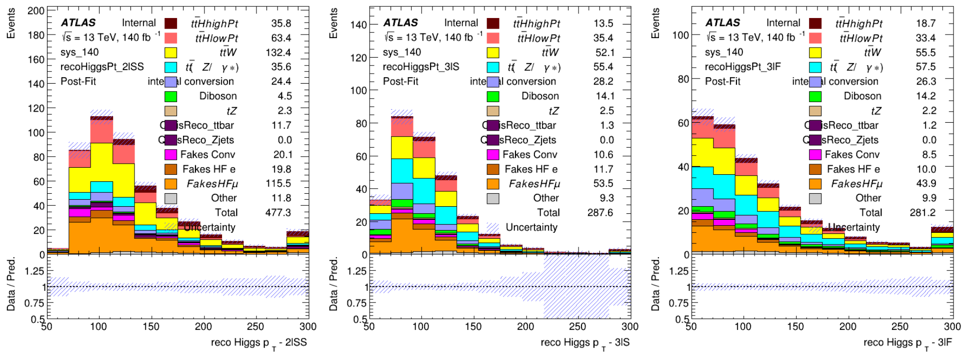


Figure 18.3: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

18.2 Projected Results - 139 fb^{-1}

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



(a) (b) (c)

Figure 18.4: Blinded post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 139 fb^{-1} of data

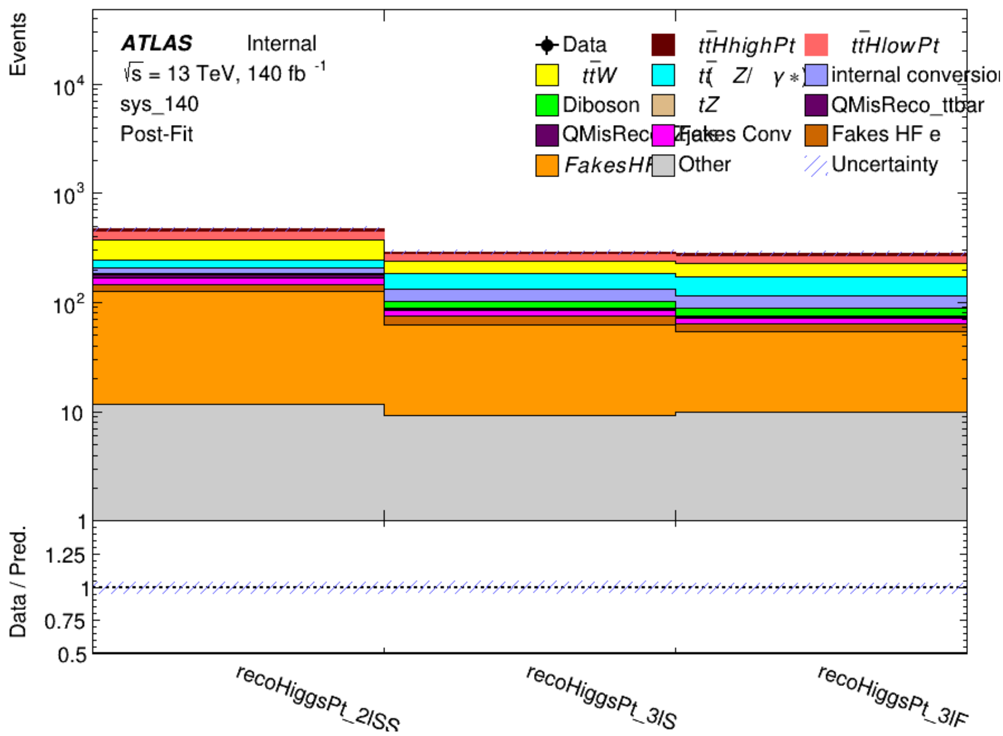


Figure 18.5: Post-fit summary of fit.

Projected uncertainties on the μ values extracted from the fit for high and low p_T Higgs are shown in 40. A significance of 2.0σ is expected for $t\bar{t}H$ high p_T, and a projected significance 2.3σ is extracted for $t\bar{t}H$ low p_T.

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

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

Table 40: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T, where $\mu = \sigma_{obs}/\sigma_{pred}$

The most prominent sources of systematic uncertainty, as measured by their impact on $\mu_{t\bar{t}H \text{ high } p_T}$, are summarized in Table 41.

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

Table 41: Summary of the most significant sources of systematic uncertainty on the measurement of $t\bar{t}H$ high p_T.

The most significant sources of systematic uncertainty on $t\bar{t}H$ low p_T are summarized in Table 42.

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

Table 42: Summary of the most significant sources of systematic uncertainty on the measurement of $t\bar{t}H$ low p_T.

1351 The ranking and impact of those nuisance parameters with the largest contribution to the
1352 overall uncertainty is shown in Figure 18.6.

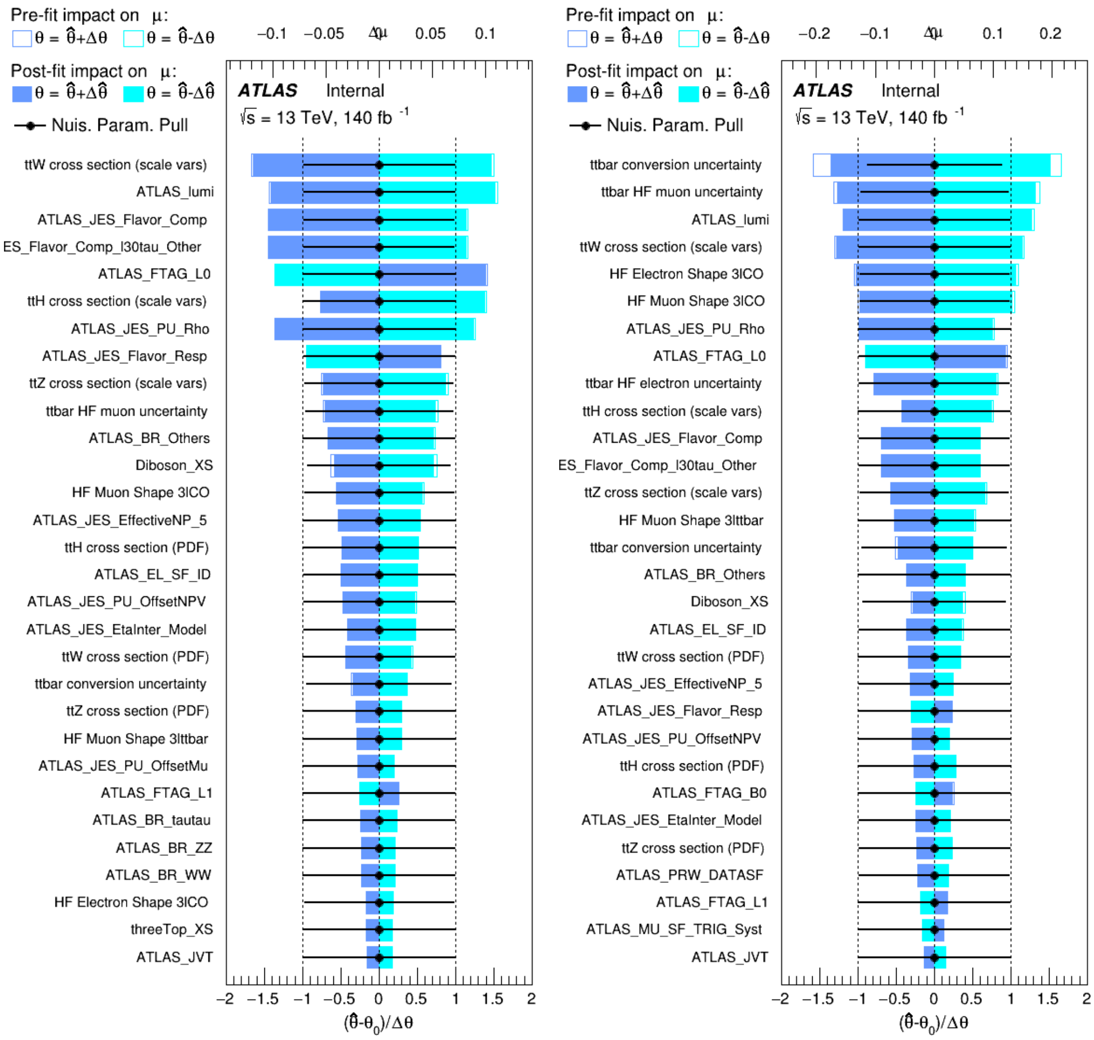


Figure 18.6: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

Part VI

Conclusion

A method of employing machine learning techniques in order to reconstruct the momentum sprectum of the Higgs boson produced in association with top quark pairs has been presented for events with multiple leptons in the final state. Preliminary results for using these techniques to perform differential measurements in this channel have been presented: Using 80 fb^{-1} of data, normalization factors of $\mu_{t\bar{t}H \text{ high } p_T} = 2.1^{+0.62}_{-0.59}(\mathrm{stat})^{+0.40}_{-0.43}(\mathrm{sys})$ and $\mu_{t\bar{t}H \text{ low } p_T} = 0.83^{+0.39}_{-0.40}(\mathrm{stat})^{+0.51}_{-0.53}(\mathrm{sys})$ are measured for events with Higgs $p_T >$150 GeV and $<$150 GeV, respectively. Projected results for 139 fb^{-1} of data give sentivity of $\mu_{t\bar{t}H \text{ high } p_T} = 1.00^{+0.45}_{-0.43}(\mathrm{stat})^{+0.30}_{-0.31}(\mathrm{sys})$ and $\mu_{t\bar{t}H \text{ low } p_T} = 1.00^{+0.29}_{-0.30}(\mathrm{stat})^{+0.48}_{-0.50}(\mathrm{sys})$.

References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: `1207.7214 [hep-ex]`.

[2] R. Oerter, *The theory of almost everything : the Standard Model, the unsung triumph of modern physics*, Pi Press, 2006, ISBN: 978-0-13-236678-6.

[3] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. 13 (Oct, 1964) 508-509 ().

[4] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*, Phys. Rev. **127** (3 1962) 965, URL: https://link.aps.org/doi/10.1103/PhysRev.127.965.

[5] J. Ellis, *Higgs Physics*, (2013) 117, 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report, URL: https://cds.cern.ch/record/1638469.

[6] T. A. Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Physics Letters B **784** (2018) 173, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/S0370269318305732.

[7] *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, tech. rep. ATLAS-CONF-2017-077, CERN, 2017, URL: https://cds.cern.ch/record/2291405.

[8] A. M. Sirunyan et al., *Observation of* $\mathrm{t\bar{t}H}$ *Production*, Physical Review Letters **120** (2018), ISSN: 1079-7114, URL: http://dx.doi.org/10.1103/PhysRevLett.120.231801.

[9] B. Dumont, S. Fichet and G. von Gersdorff, *A Bayesian view of the Higgs sector with higher dimensional operators*,

Journal of High Energy Physics **2013** (2013), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP07(2013)065.

[10] S. Banerjee, S. Mukhopadhyay and B. Mukhopadhyaya, *Higher dimensional operators and the LHC Higgs data: The role of modified kinematics*, Physical Review D **89** (2014), ISSN: 1550-2368, URL: http://dx.doi.org/10.1103/PhysRevD.89.053010.

[11] M. Zinser, 'The Large Hadron Collider', *Search for New Heavy Charged Bosons and Measurement of High-Mass Drell-Yan Production in Proton—Proton Collisions*, Springer International Publishing, 2018 47, ISBN: 978-3-030-00650-1, URL: https://doi.org/10.1007/978-3-030-00650-1_4.

[12] *Detector and Technology*, URL: https://atlas.cern/discover/detector.

[13] M. Marjanovic, *ATLAS Tile calorimeter calibration and monitoring systems*, 2018.

[14] ATLAS Collaboration, *Observation of electroweak* $W^{\pm}Z$ *boson pair production in association with two jets in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. **B793** (2019) 469, arXiv: 1812.09740 [hep-ex].

[15] ATLAS Collaboration, *Luminosity determination in* pp *collisions at* $\sqrt{s} = 7$ *TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **71** (2011) 1630, arXiv: 1101.2185 [hep-ex].

[16] ATLAS Collaboration, *Performance of the ATLAS detector using first collision data*, JHEP **09** (2010) 056, arXiv: 1005.5254 [hep-ex].

[17] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[18] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph].

[19] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv: 1007.2241 [hep-ph].

[20] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph].

[21] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph].

[22] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639, arXiv: 0803.0883 [hep-ph].

[23] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[24] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph].

[25] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: 1009.2450 [hep-ph].

[26] A. Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, The European Physical Journal C **79** (2019), ISSN: 1434-6052, URL: `http://dx.doi.org/10.1140/epjc/s10052-019-7140-6`.

[27] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur. Phys. J. C **74** (2014) 2941, arXiv: `1404.2240 [hep-ex]`.

[28] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, Eur. Phys. J. C **74** (2014) 3130, arXiv: `1407.3935 [hep-ex]`.

[29] ATLAS Collaboration, *Observation of the Higgs boson production in association with a* $t\bar{t}$ *pair with the ATLAS detector*, PoS **EPS-HEP2019** (2020) 325.

[30] M. Cacciari, G. P. Salam and G. Soyez, *The anti-ktjet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063, ISSN: 1029-8479, URL: `http://dx.doi.org/10.1088/1126-6708/2008/04/063`.

[31] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in* pp *collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: `1510.03823 [hep-ex]`.

[32] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, Journal of High Energy Physics **2018** (2018), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP08(2018)089.

[33] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, The European Physical Journal C **78** (2018), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-018-6288-9.

[34] T. Chen and C. Guestrin, 'XGBoost: A Scalable Tree Boosting System', *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, ACM, 2016 785, ISBN: 978-1-4503-4232-2, URL: http://doi.acm.org/10.1145/2939672.2939785.

[35] ATLAS Collaboration, *Measurement of the fiducial and differential cross-section of a top quark pair in association with a Z boson at 13 TeV with the ATLAS detector*, (2019), URL: https://cds.cern.ch/record/2672207.

[36] *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, (2019).

[37] ATLAS Collaboration, *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017.

[38] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, arXiv: 1703.09665 [hep-ex].

[39] A. Collaboration, *Performance of b -jet identification in the ATLAS experiment*, Journal of Instrumentation **11** (2016) P04008, URL: http://stacks.iop.org/1748-0221/11/i=04/a=P04008.

[40] ATLAS Collaboration, *Observation of the associated production of a top quark and a Z boson in pp collisions at* $\sqrt{s}$*= 13 TeV with the ATLAS detector*, Journal of High Energy Physics **2020** (2020), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP07(2020)124.

[41] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions from Powheg and Sherpa to ATLAS measurements of top pair production at 7 TeV*, ATL-PHYS-PUB-2015-011, 2015, URL: https://cds.cern.ch/record/2020602.

[42] ATLAS Collaboration, *ATLAS tunes of Pythia 6 and Pythia 8 for MC11*, ATL-PHYS-PUB-2011-009, 2011, URL: https://cds.cern.ch/record/1363300.

[43] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, URL: https://cds.cern.ch/record/2157687.

[44] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-015, 2015, URL: https://cds.cern.ch/record/2037613.

[45] ATLAS Collaboration, *Selection of jets produced in* 13 *TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: https://cds.cern.ch/record/2037702.

[46] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-027, 2015, URL: https://cds.cern.ch/record/2037904.

[47] G. Aad, T. Abajyan, B. Abbott and etal, *Jet energy resolution in proton-proton collisions at* $\sqrt{s} = 7$ *TeV recorded in 2010 with the ATLAS detector*, The European Physical Journal C **73** (2013) 2306, ISSN: 1434-6052, URL: https://doi.org/10.1140/epjc/s10052-013-2306-0.

[48] R. Narayan et al., *Measurement of the total and differential cross sections of a top-quark-antiquark pair in association with a W boson in proton-proton collisions at a centre-of-mass energy of 13 TeV with ATLAS detector at the Large Hadron Collider*, tech. rep. ATL-COM-PHYS-2020-217, CERN, 2020, URL: https://cds.cern.ch/record/2712986.

[49] R. Frederix et al., *The automation of next-to-leading order electroweak calculations*, Journal of High Energy Physics **2018** (2018), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP07(2018)185.

List of contributions

Part VII

Appendices

A Non-prompt lepton MVA

A lepton MVA has been developed to better reject non-prompt leptons than standard cut based selections based upon impact parameter, isolation and PID. The name of this MVA is `PromptLeptonVeto`. The full set of studies and detailed explanation can be found in [48].

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

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

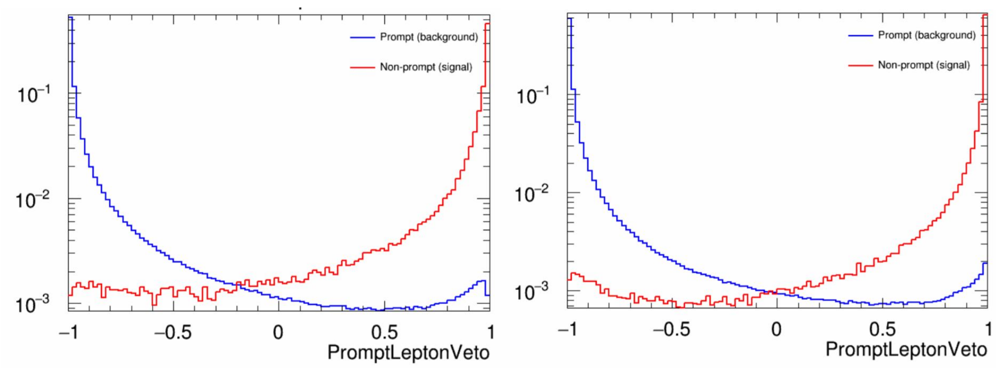
The main idea is to identify non-prompt light leptons using lifetime information associated with a track jet that matches the selected light lepton. This lifetime information is computed using tracks contained within the jet. Typically, lepton lifetime is determined using the impact parameter of the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

`PromptLeptonVeto` is a gradient boosted BDT. The training of the BDT is performed on leptons selected from the Powheg+Pythia6 non-allhad $t\bar{t}$ MC sample. Eight variables are used to train the BDT in order to discriminate between prompt and non-prompt leptons. The track jets that are matched to the non-prompt leptons correspond to jets initiated by b or c quarks, and may contain a displaced vertex. Consequently, three of the selected variables are used to identify b-tag jets by standard ATLAS flavour tagging algorithms. Two variables use the relationship between the track jet and lepton: the ratio of the lepton p_T with respect to the track jet p_T and ΔR between the lepton and the track jet axis. Finally three additional variables test whether the reconstructed lepton is isolated: the number of tracks collected by the track jet and the lepton track and calorimeter isolation variables. Table 43 describes the variables used to train the BDT algorithm. The choice of input variables has been extensively discussed with Egamma, Muon, Tracking, and Flavour Tagging CP groups.

The output distribution of the BDT is shown in Figure A.

Variable	Description
N_{track} in track jet	Number of tracks collected by the track jet
IP2 $\log(P_b/P_{\mathrm{light}})$	Log-likelihood ratio between the b and light jet hypotheses with the IP2D algorithm
IP3 $\log(P_b/P_{\mathrm{light}})$	Log-likelihood ratio between the b and light jet hypotheses with the IP3D algorithm
N_{TrkAtVtx} SV + JF	Number of tracks used in the secondary vertex found by the SV1 algorithm in addition to the number of tracks from secondary vertices found by the JetFitter algorithm with at least two tracks
$p_T^{\mathrm{lepton}}/p_T^{\mathrm{track\ jet}}$	The ratio of the lepton p_T and the track jet p_T
$\Delta R(\mathrm{lepton, track\ jet})$	ΔR between the lepton and the track jet axis
p_TVarCone30/p_T	Lepton track isolation, with track collecting radius of $\Delta R < 0.3$
E_TTopoCone30/p_T	Lepton calorimeter isolation, with topological cluster collecting radius of $\Delta R < 0.3$

Table 43: A table of the variables used in the training of `PromptLeptonVeto`.



(a) (b)

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

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

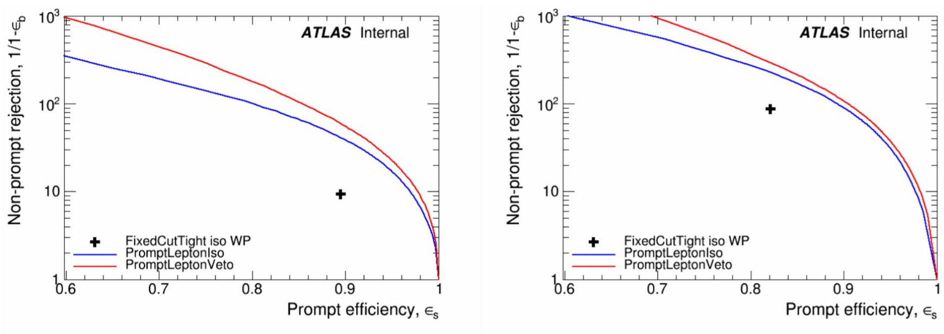


Figure A.2: ROC curves for the PLV as well as the peformance of the standard FixedCutTight WP for (left) electrons and (right) muons

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

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

B Supplementary WZ + Heavy Flavor Studies

B.1 Non-prompt CR Modelling

In order to further validate the modeling in each of the non-prompt CRs, additional kinematic plots are made in the Z+jets CR and $t\bar{t}$ CR in each of the continuous b-tag regions, after the correction factors detailed in Section 8.3 have been applied.

In the case of the Z+jets CR, the p_T spectrum of the lepton originating from the W candidate is shown, as this is the distribution used to extract the scale factor applied to Z+jets. These plots are shown in Figures B.1 and B.2.

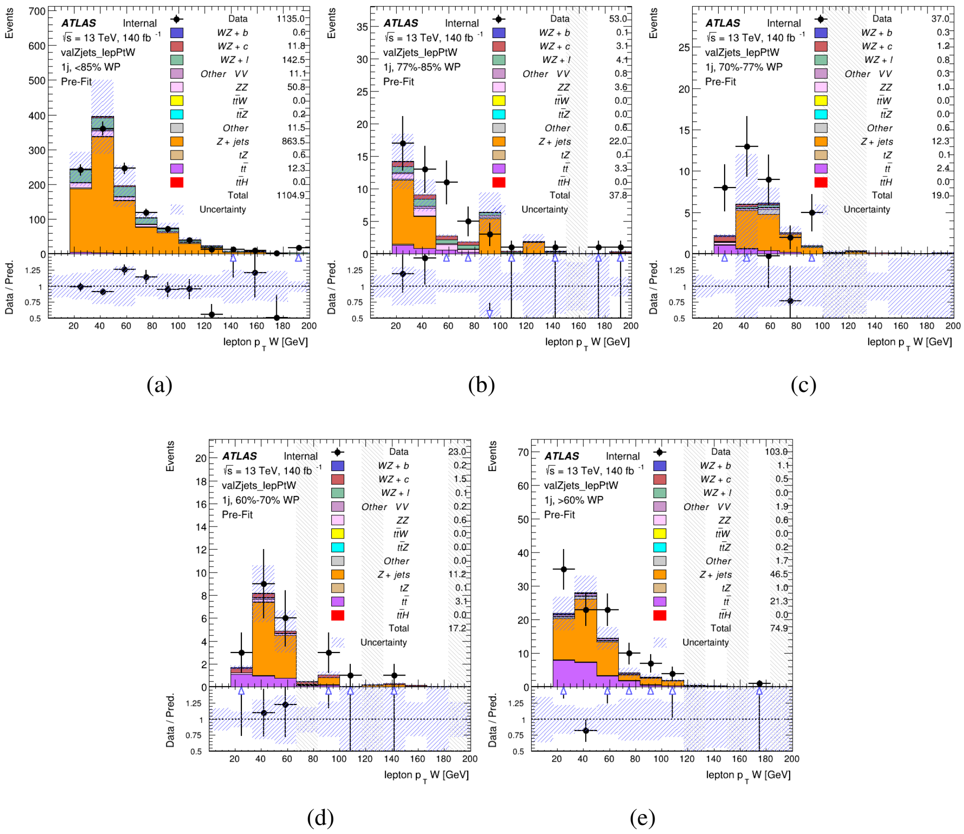


Figure B.1: Comparisons between the data and MC distributions of the p_T of the lepton originating from the W-candidate in the Z+jets CR for each of the 1-jet b-tag working point regions

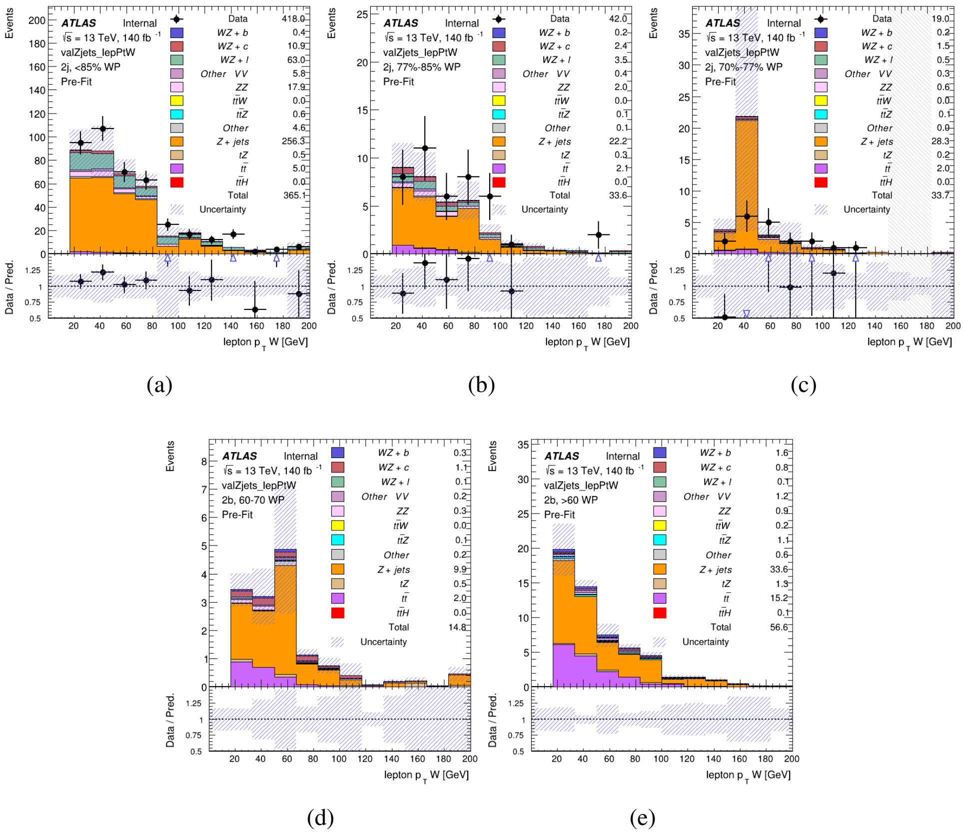
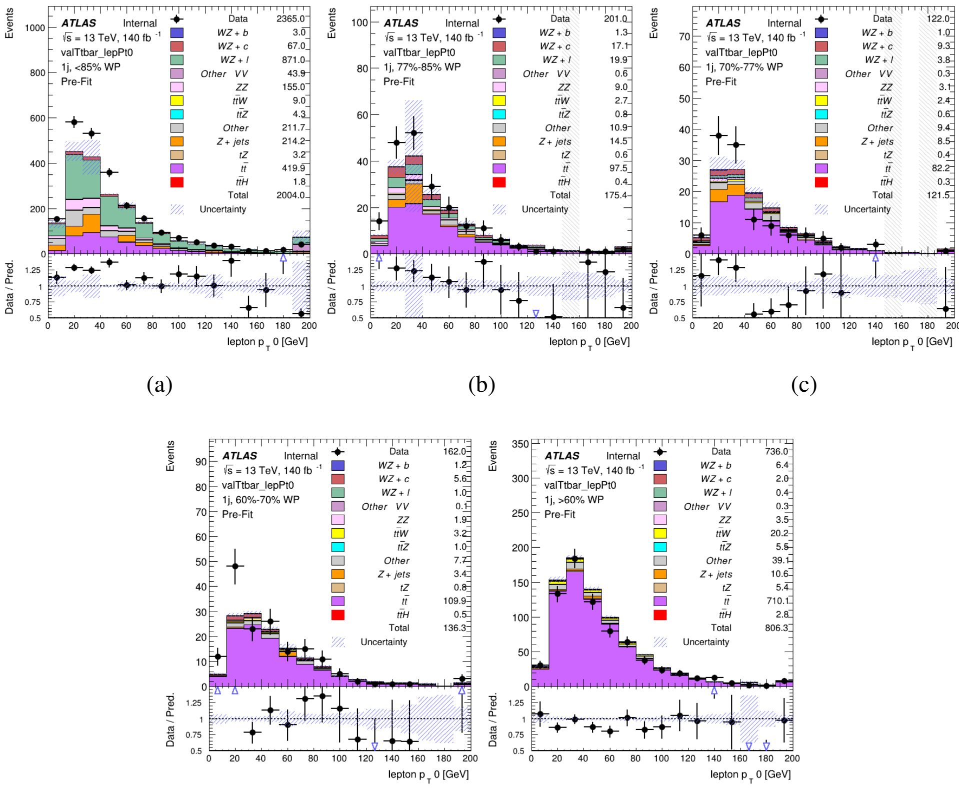


Figure B.2: Comparisons between the data and MC distributions of the p_T of the lepton originating from the W-candidate in the Z+jets CR for each of the 2-jet b-tag working point regions

The same is shown for the $t\bar{t}$ CR, but the p_T of the OS lepton is used instead as a representation of the modeling, as the lepton from the W is not well defined for $t\bar{t}$ events. These plots are shown in Figures B.3 and B.4.



(a) (b) (c)

(d) (e)

Figure B.3: Comparisons between the data and MC distributions of the p_T of the OS lepton in the $t\bar{t}$ CR for each of the 1-jet b-tag working point regions

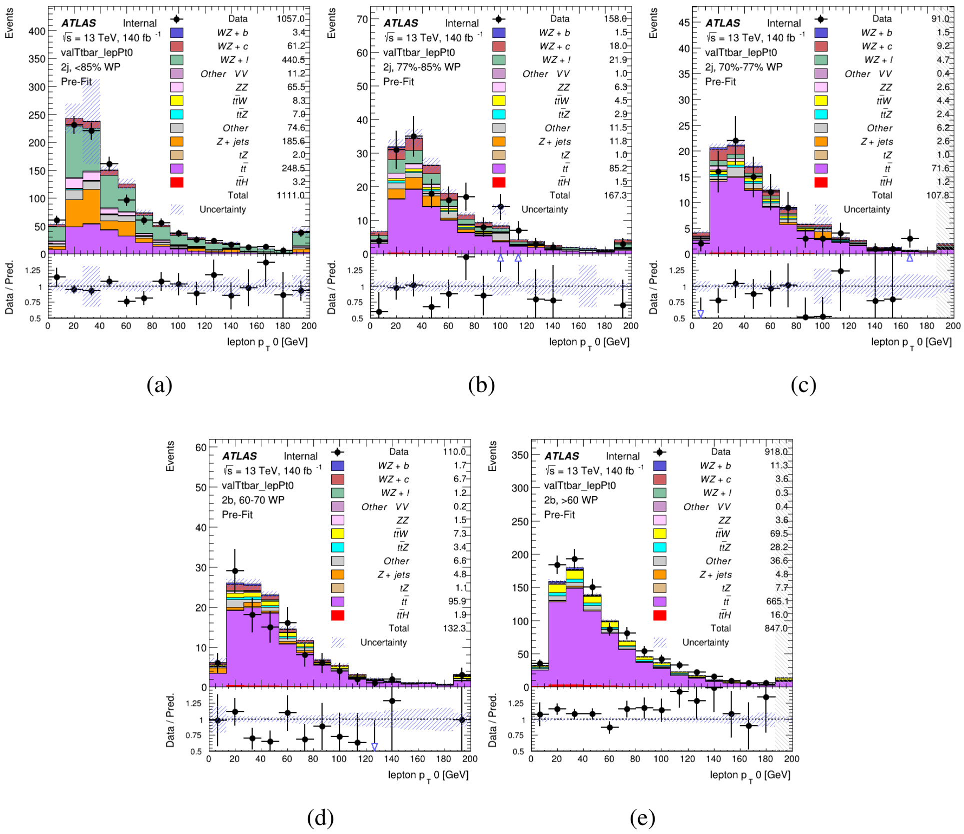


Figure B.4: Comparisons between the data and MC distributions of the p_T of the OS lepton in the $t\bar{t}$ CR for each of the 2-jet b-tag working point regions

B.2 tZ Interference Studies

Because it includes an on-shell Z boson as well as a b-jet and W from the top decay, tZ production represents an identical final state to WZ + b-jet. This implies the possibility of matrix level interference between these two processes not accounted for in the Monte Carlo simulations, which consider the two processes independently. Truth level studies are performed in order to estimate the impact of these interference effects.

In order to estimate the matrix level interference effects between tZ and WZ + b-jet, two different sets of simulations are produced using MadGraph 5 [49] - one which simulates these two processes independently, and another where they are produced simultaneously, such that interference effects are present. These two sets of samples are then compared, and the difference between them can be taken to represent any interference effects.

MadGraph simulations of 10,000 tZ and 10,000 WZ + b events are produced, along with 20,000 events where both are present, in the fiducial region where three leptons and at least one jet are produced.

A selection mimicking the preselection used in the main analysis is applied to the samples: The SS leptons are required to have p_T >20 GeV, and >10 GeV is required for the OS lepton. The associated b-jet is required to have p_T >25 GeV, and all physics objects are required to fall in a range of $|\eta|$ <2.5.

The overall cross-section with and without intereference effects agree within error, and no significant differences in the kinematic distributions are seen. It is therefore concluded that interference effects do not significantly impact the results.

B.3 Alternate tZ Inclusive Fit

B.3.1 tZ Inclusive Fit

While tZ is often considered as a distinct process form $WZ + b$, this could also be considered part of the signal. Alternative studies are performed where, using the same framework as the nominal analysis, a measurement of $WZ + b$ is performed that includes tZ as part of WZ+b.

Because of this change, the tZ CR is no longer necessary, and only the five pseudo-continuous b-tag regions are used in the fit. Further, systematics related to the tZ cross-section are removed from the fit, as they are now encompassed by the normalization measurement of $WZ + b$. All other systematic uncertainties are carried over from the nominal analysis.

An expected $WZ + b$ cross-section of $4.1^{+0.78}_{-0.74}(stat)^{+0.53}_{-0.52}(sys)$ fb is extracted from the fit, with an expected significance of 4.0σ.

The impact of the predominate systematics are summarized in Table 44.

Uncertainty Source	$\Delta\mu$	
WZ + light cross-section	0.08	-0.08
Jet Energy Scale	-0.06	0.08
Luminosity	-0.05	0.06
WZ + c cross-section	-0.04	0.05
Other Diboson + b cross-section	-0.04	0.04
WZ cross-section - QCD scale	-0.04	0.03
$t\bar{t}$ cross-section	-0.03	0.03
Jet Energy Resolution	-0.03	0.03
Flavor tagging	-0.03	0.03
Z+jets cross section	-0.02	0.02
Total Systematic Uncertainty	-0.15	0.16

Table 44: Summary of the most significant sources of systematic uncertainty on the measurement of $WZ + b$ with exactly one associated jet.

B.3.2 Floating tZ

In order to quantify the impact of the tZ uncertainty on the fit, an alternate fit strategy is used where the tZ normalization is allowed to float. This normalization factor replaces the cross-section uncertainty on tZ, and all other parameters of the fit remain the same.

An uncertainty of 17% on the normalization of tZ is extracted from the fit, compared to a theory uncertainty of 15% applied to the tZ cross-section. The measured uncertainties on WZ remain the same.

C Supplementary $t\bar{t}H$ Differential Analysis Studies

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.

C.1 Higgs Reconstruction Model Details

C.1.1 b-jet Identification Features - 2lSS

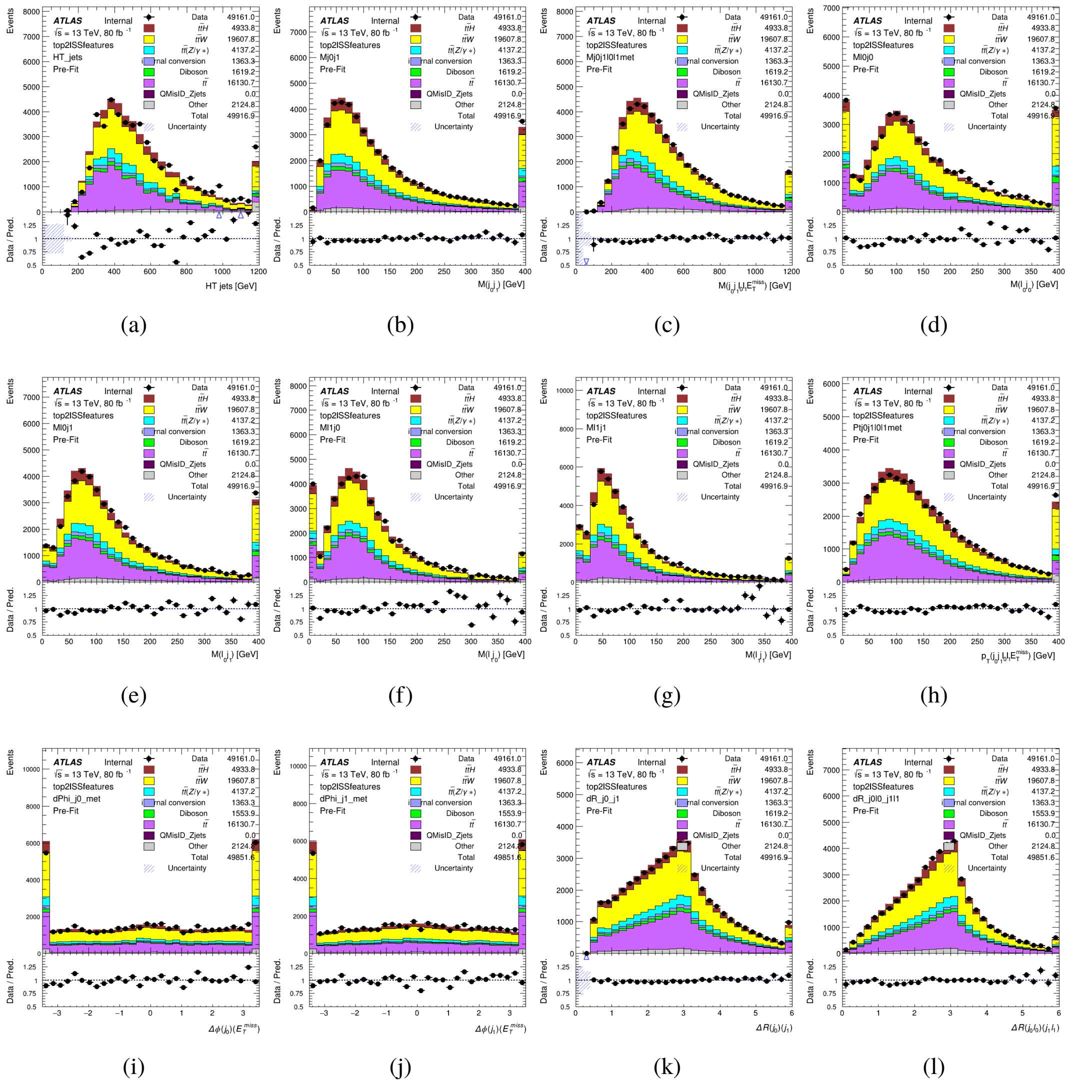


Figure C.1: Input features for top2lSS

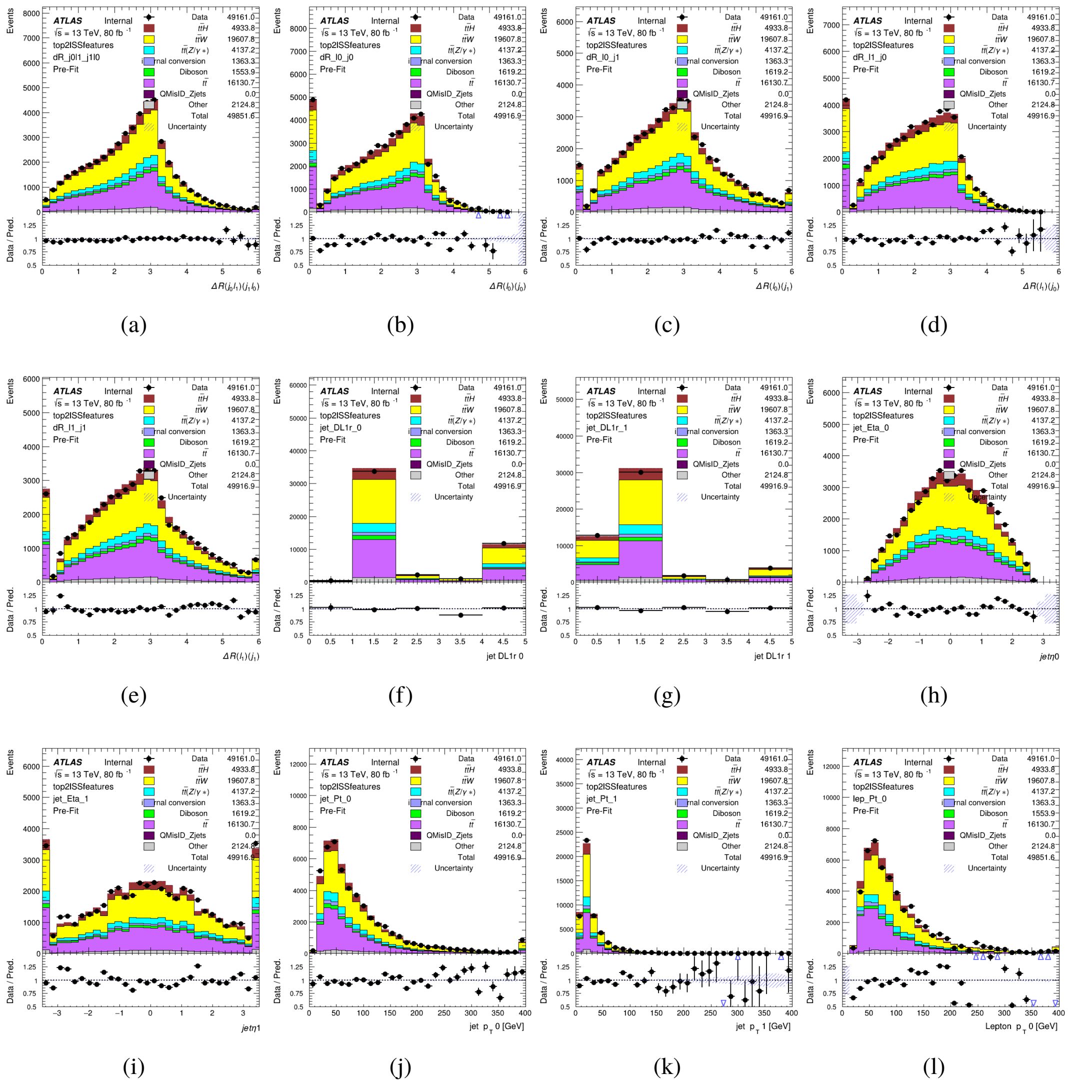
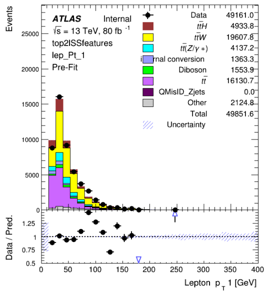
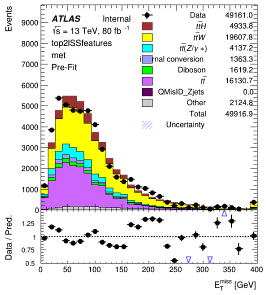


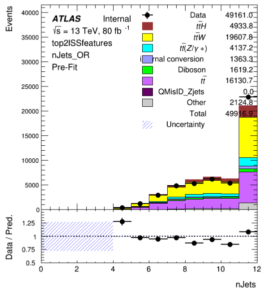
Figure C.2: Input features for top2lSS



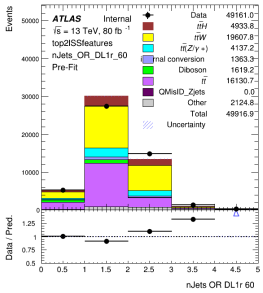
(a)



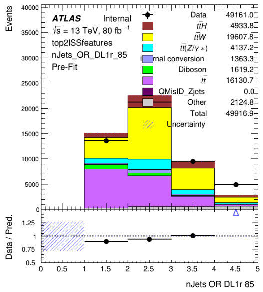
(b)



(c)



(d)



(e)

Figure C.3: Input features for top2lSS

C.1.2 b-jet Identification Features - 3l

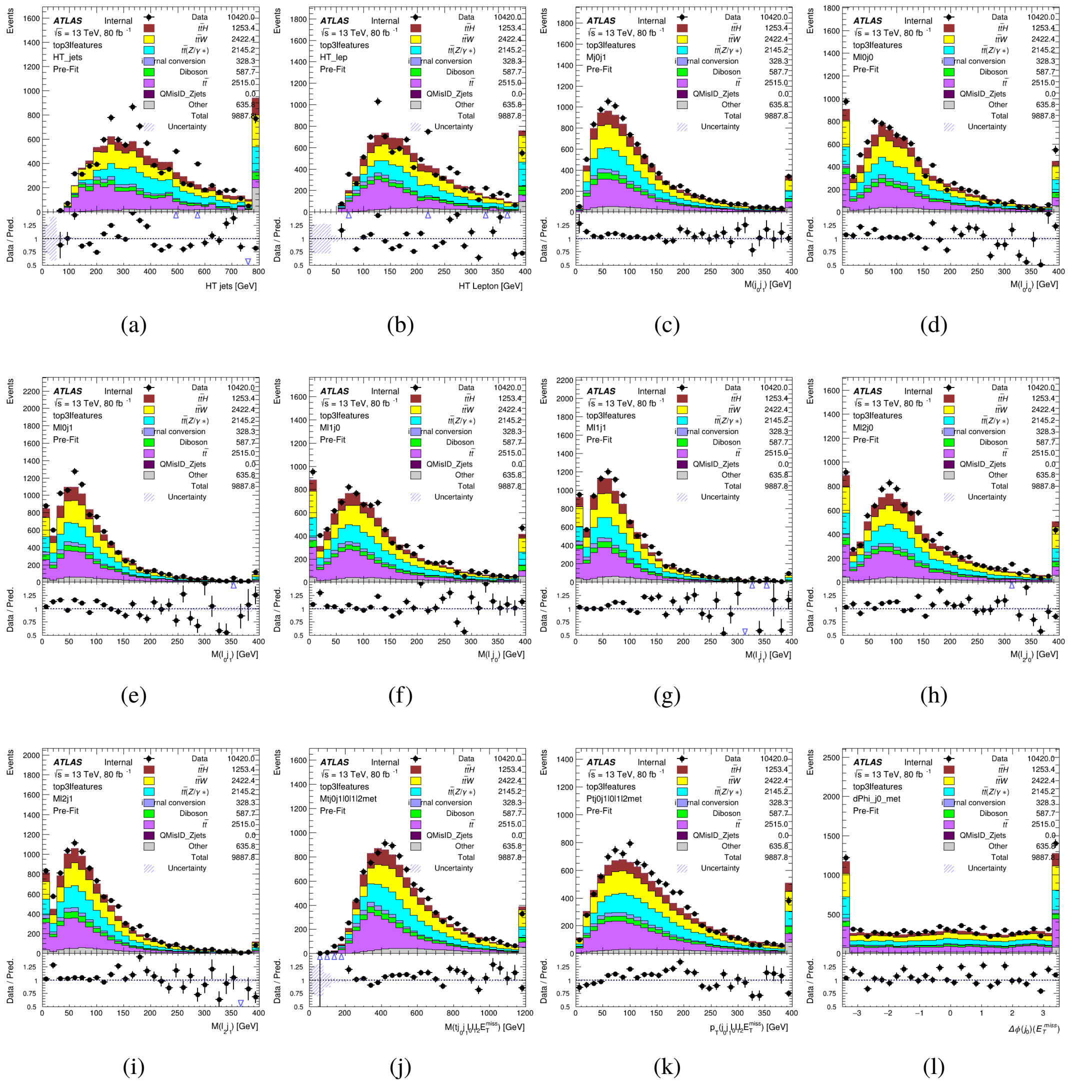
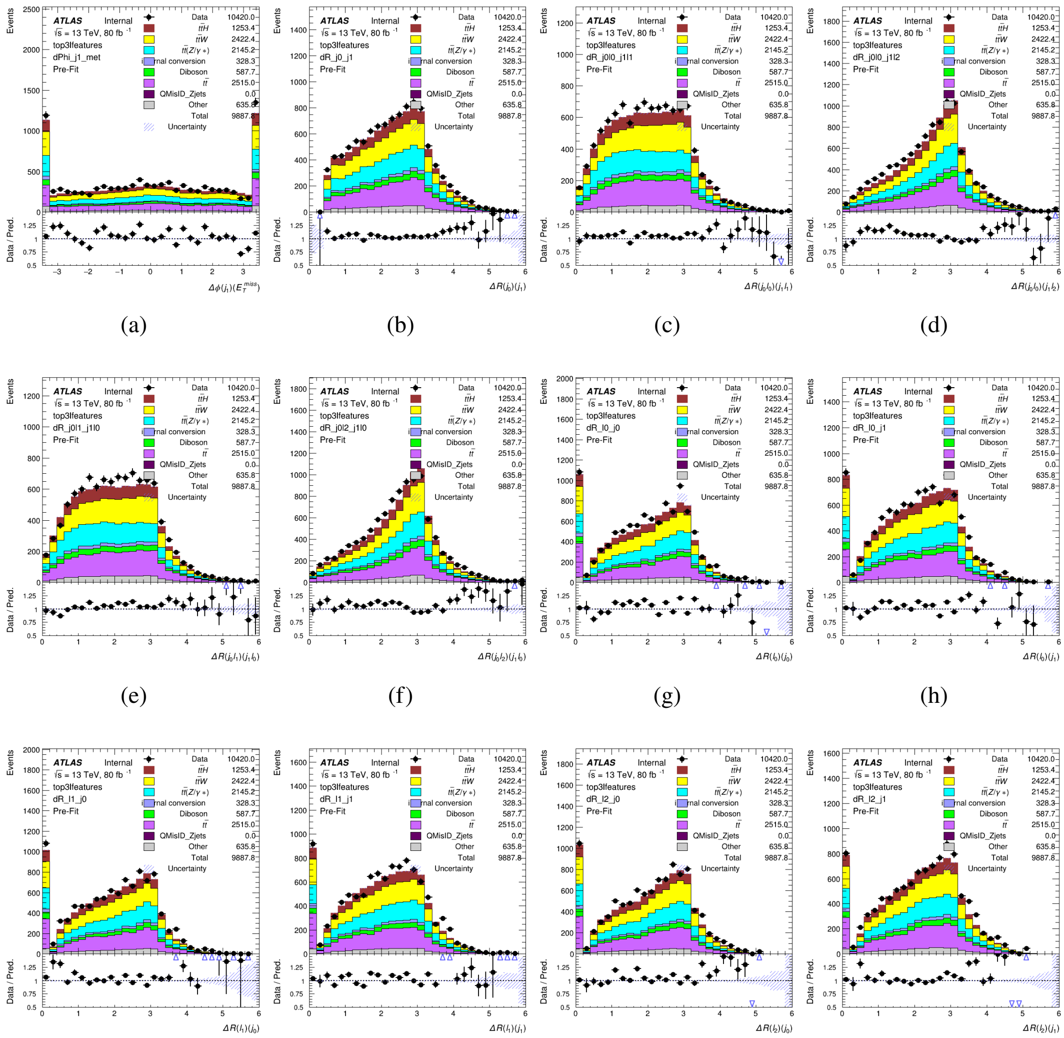


Figure C.4: Input features for top3l

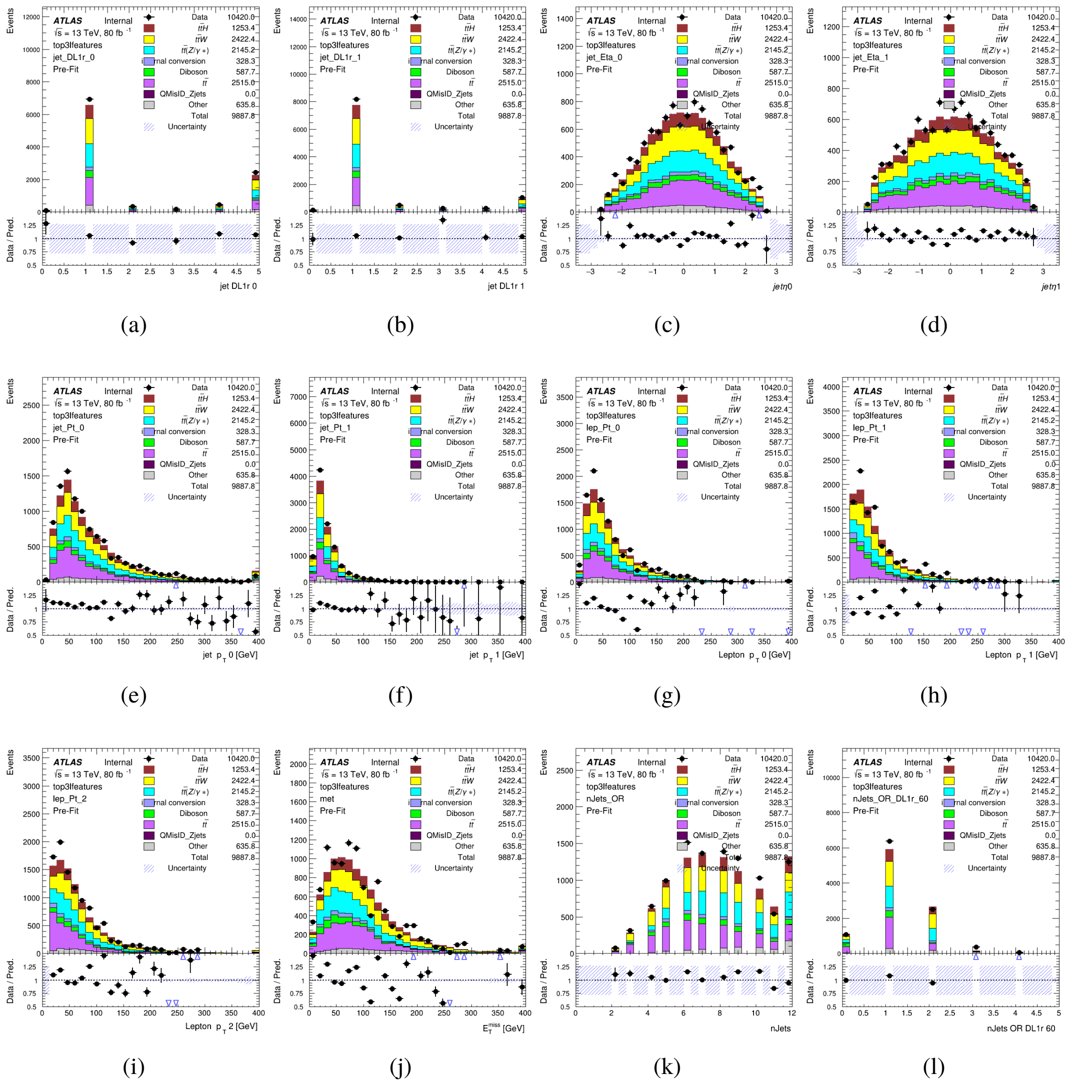


(a) (b) (c) (d)

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

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

Figure C.5: Input features for top3l

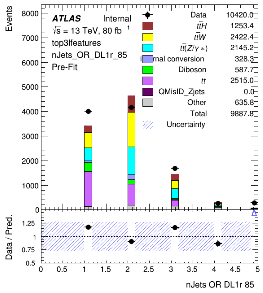


(a) (b) (c) (d)

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

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

Figure C.6: Input features for top3l



(a)

Figure C.7: Input features for top3l

1618 ### C.1.3 Higgs Reconstruction Features - 2lSS

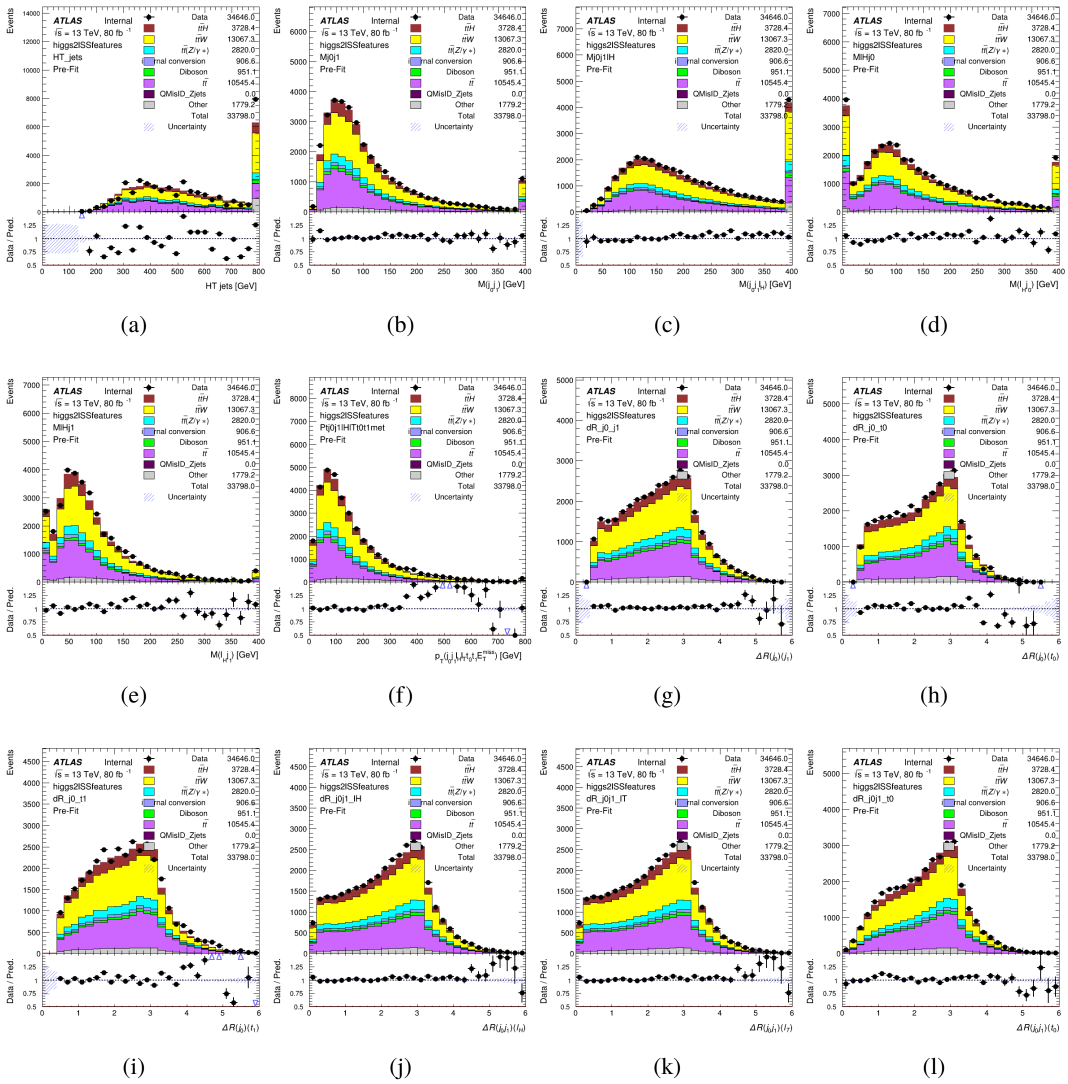
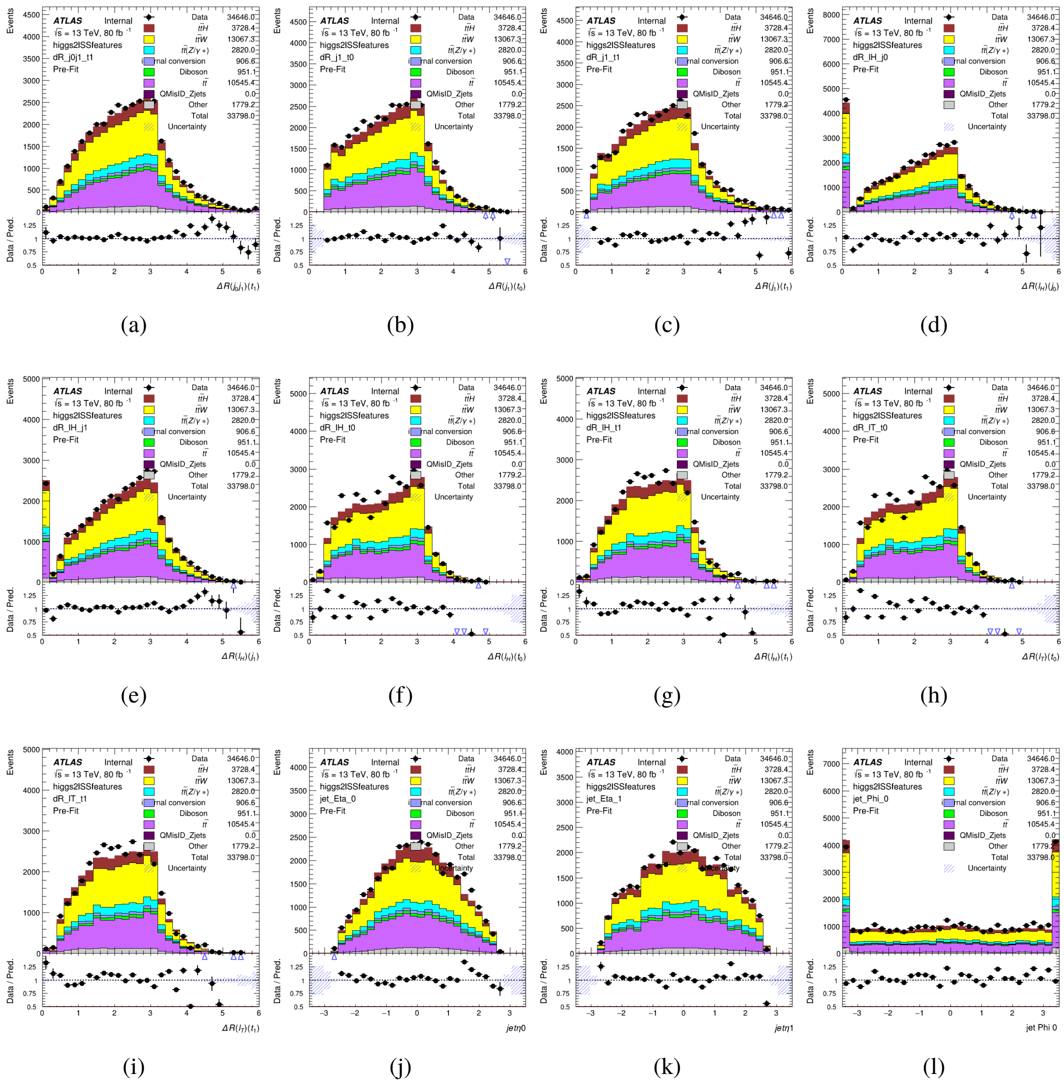
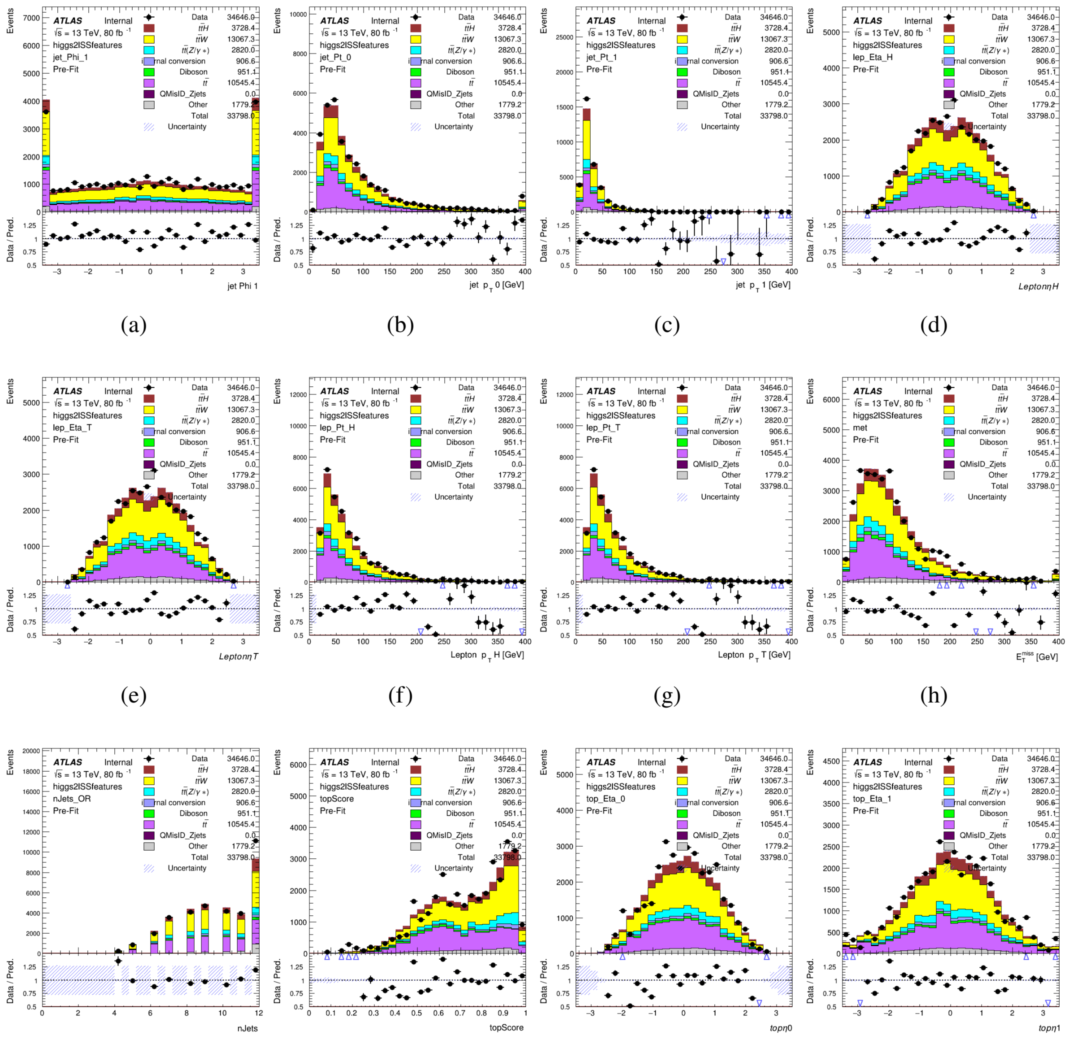


Figure C.8: Input features for higgs2lSS



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

Figure C.9: Input features for higgs2lSS

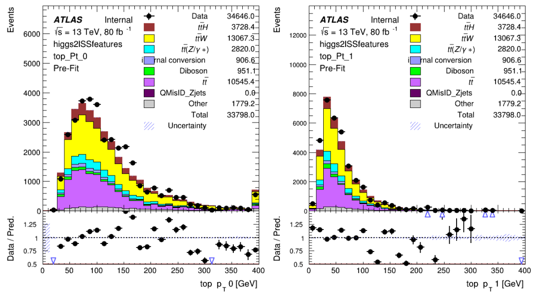


(a) (b) (c) (d)

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

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

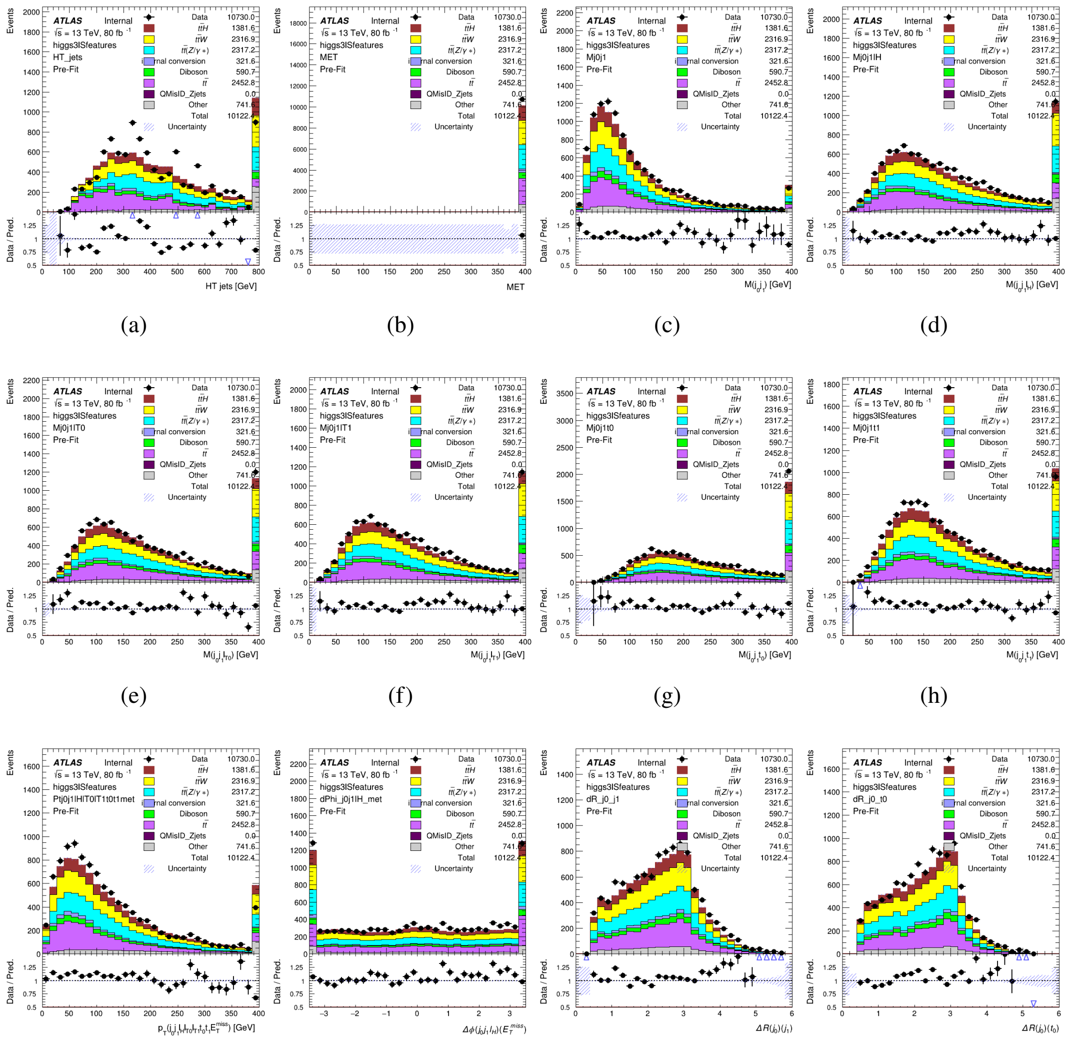
Figure C.10: Input features for higgs2lSS



(a) (b)

Figure C.11: Input features for higgs2lSS

C.1.4 Higgs Reconstruction Features - 3lS



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

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

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

Figure C.12: Input features for higgs3lS

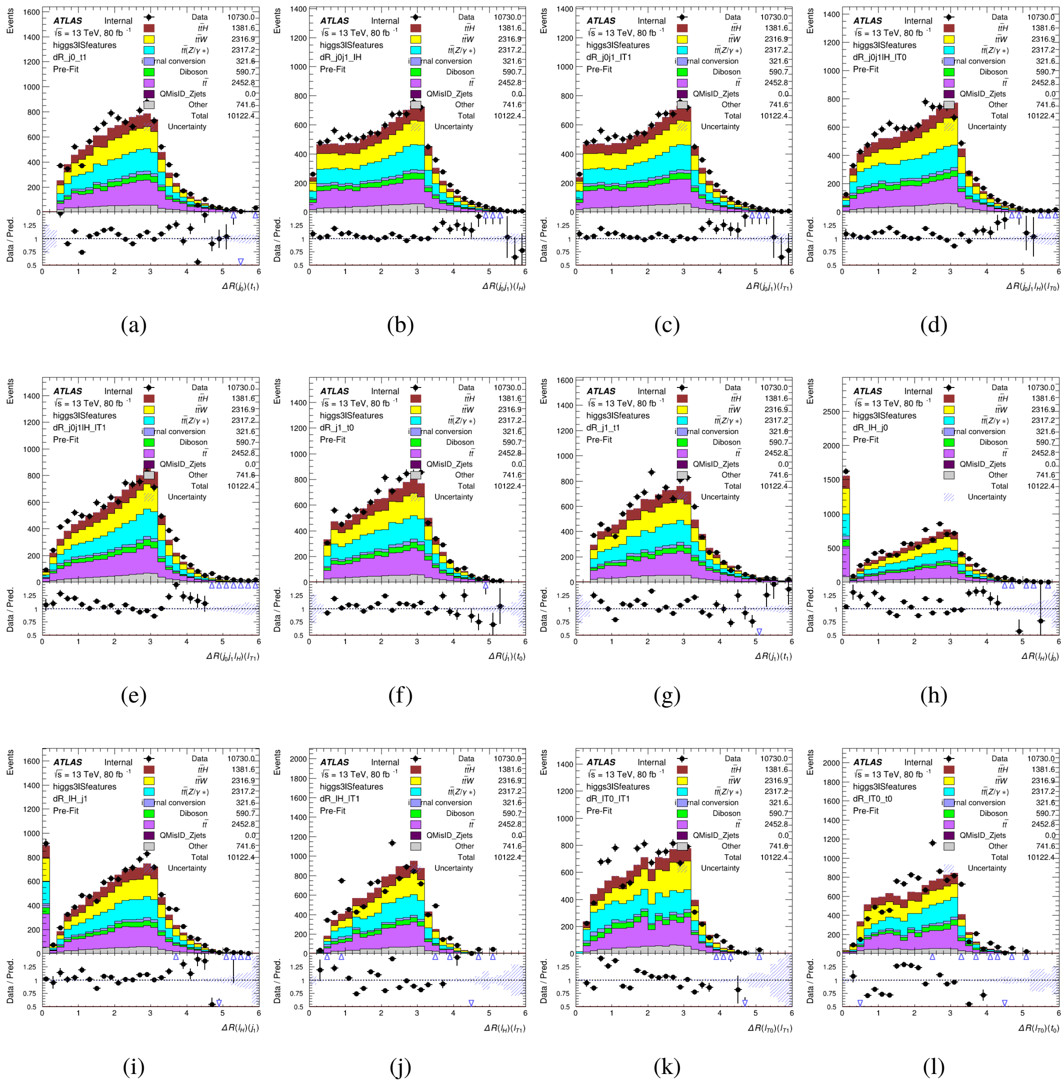
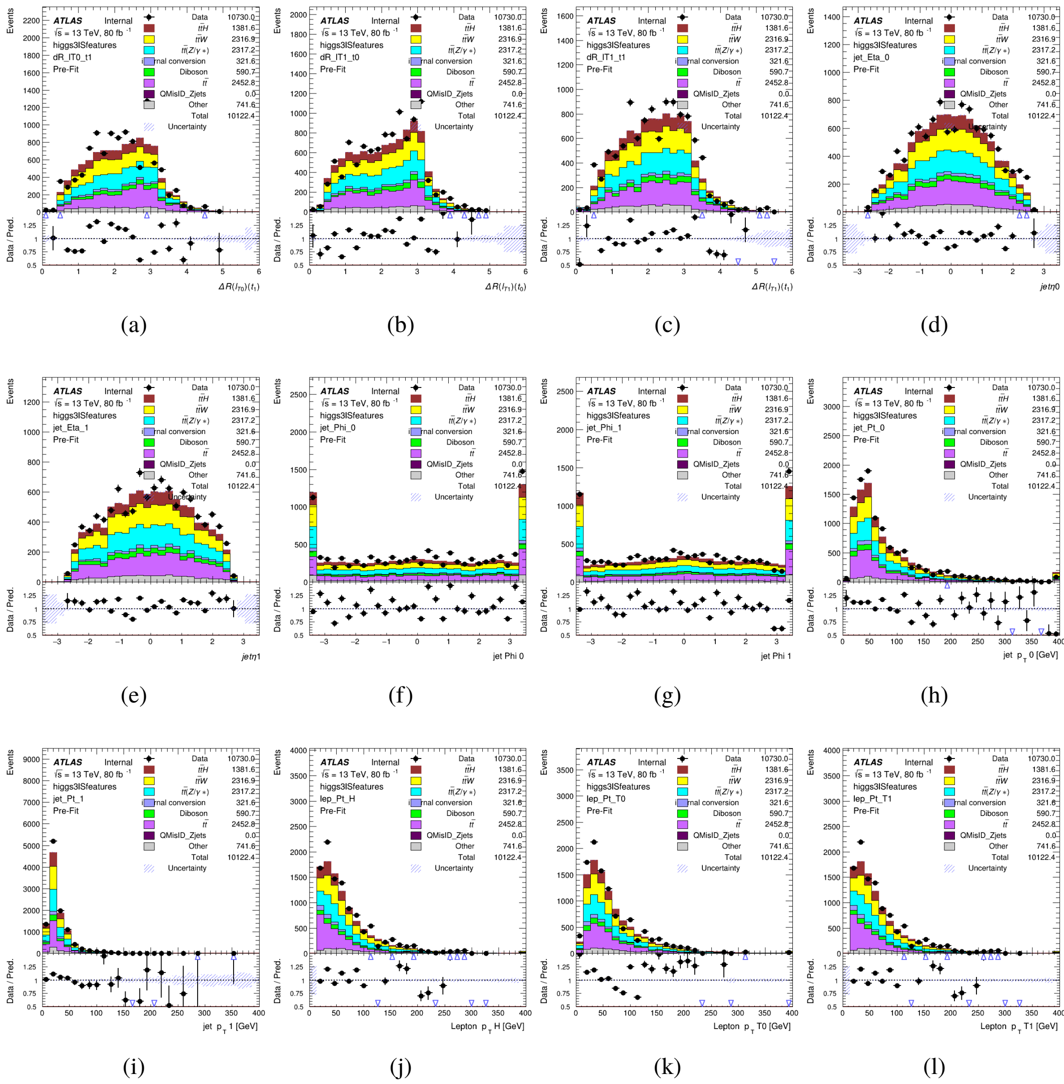


Figure C.13: Input features for higgs3lS

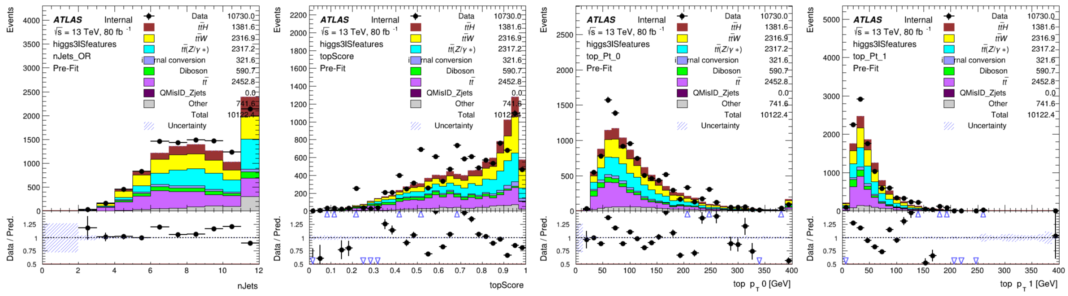


(a) (b) (c) (d)

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

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

Figure C.14: Input features for higgs3lS



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

Figure C.15: Input features for higgs3lS

1620 ### C.1.5 Higgs Reconstruction Features - 3lF

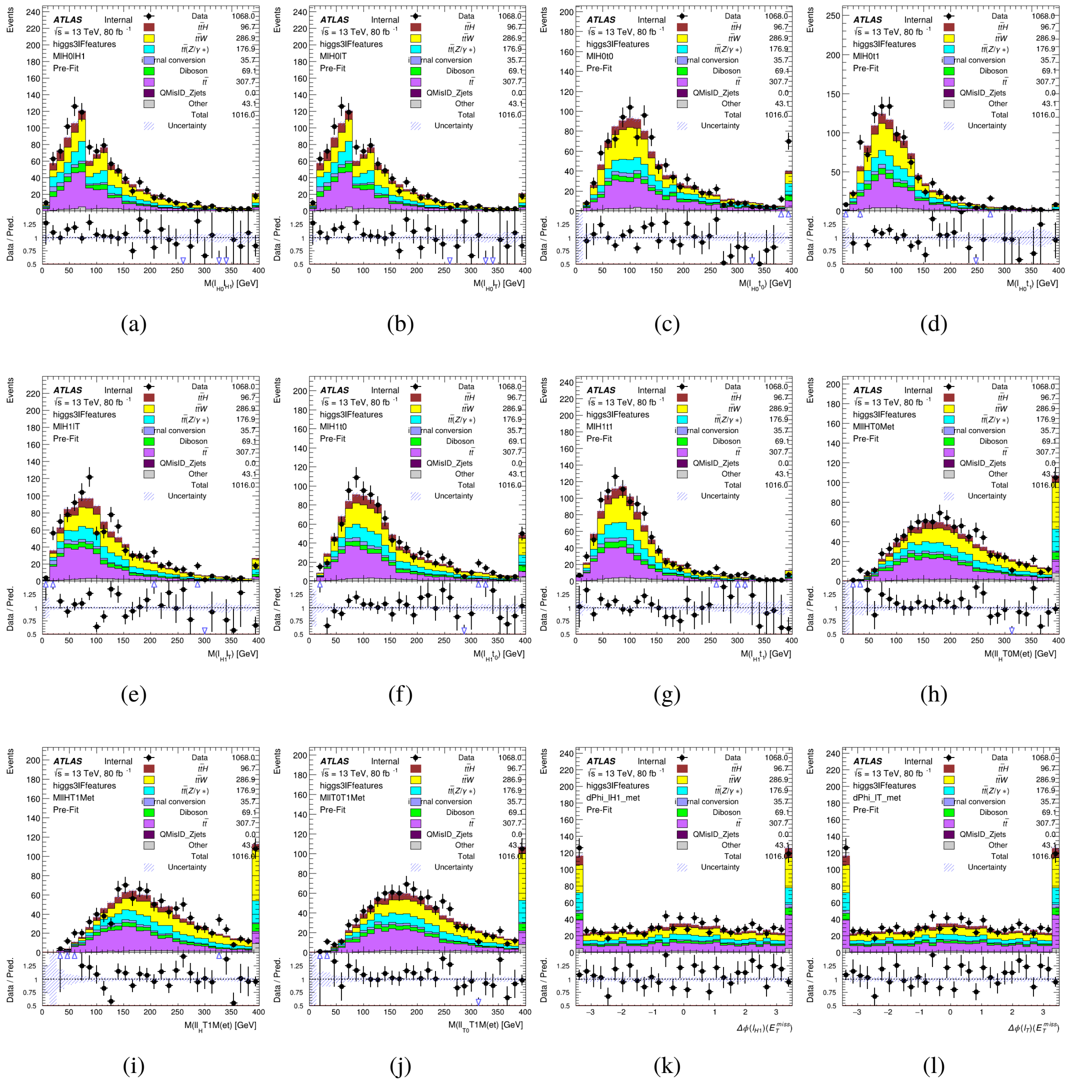


Figure C.16: Input features for higgs3lF

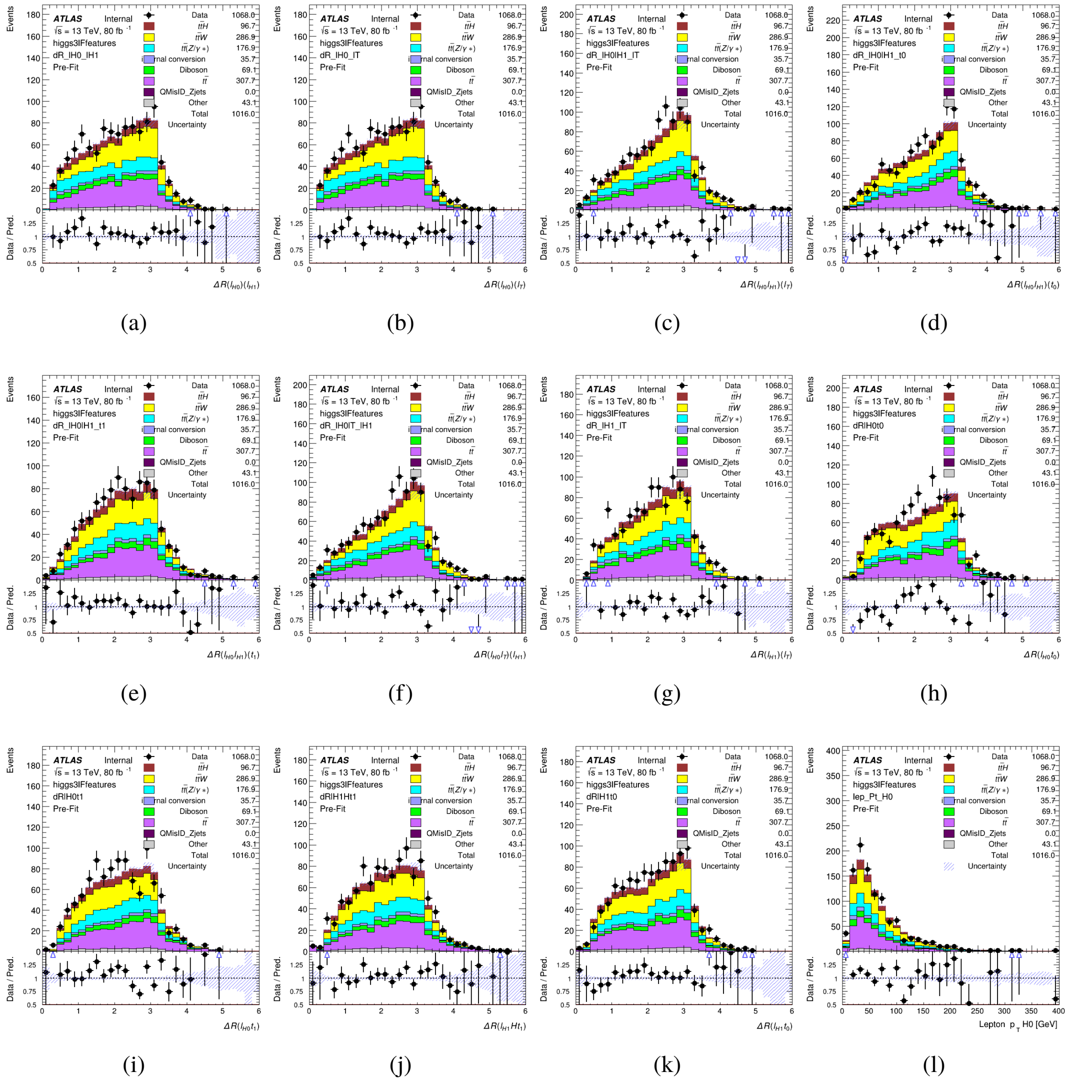
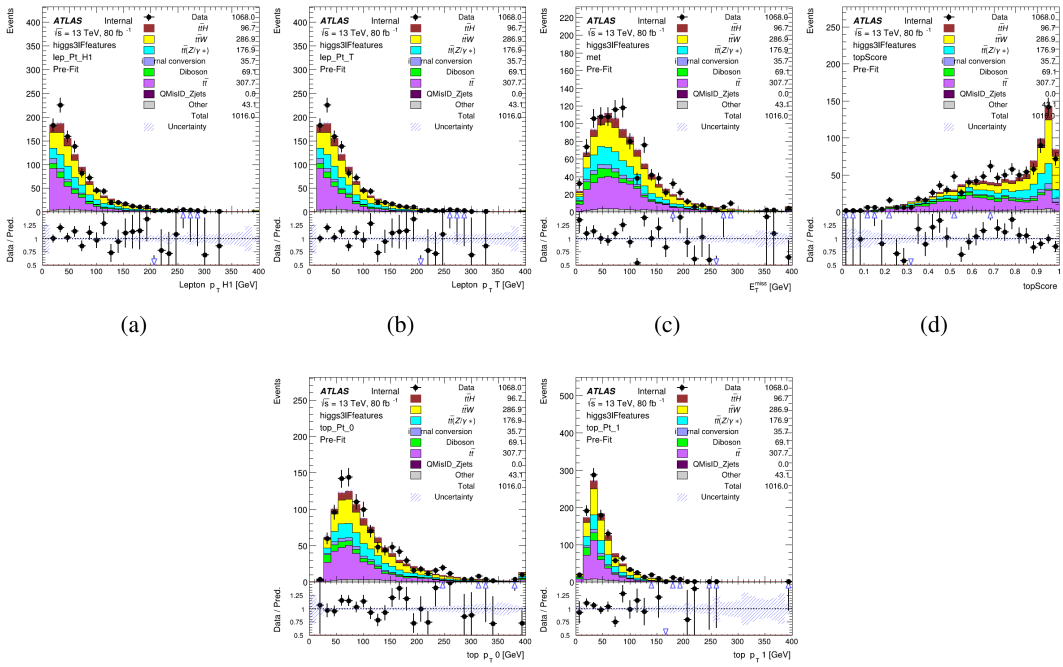


Figure C.17: Input features for higgs3lF



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

(e) (f)

Figure C.18: Input features for higgs3lF

C.2 Background Rejection MVA Details

C.2.1 Background Rejection MVA Features - 2lSS

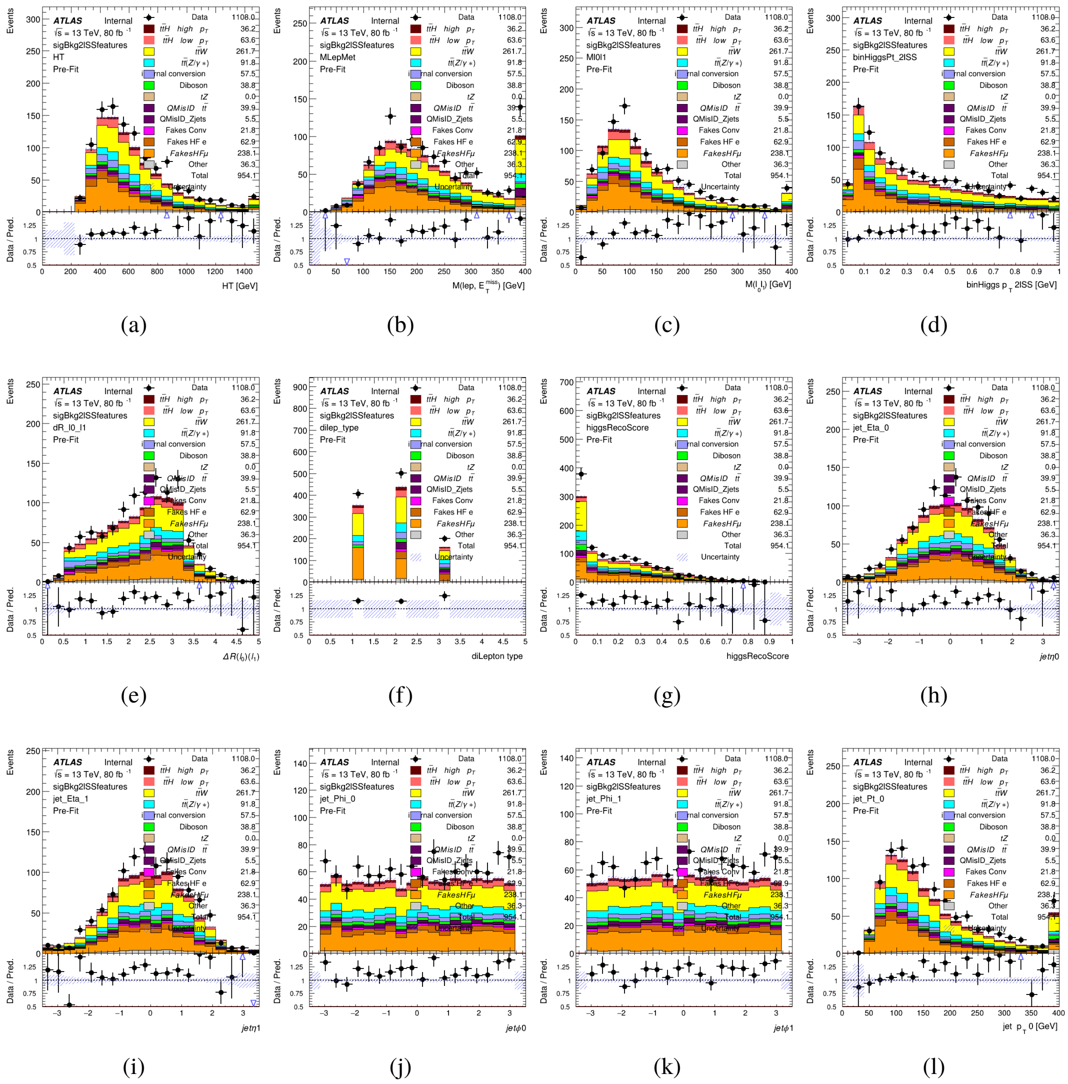
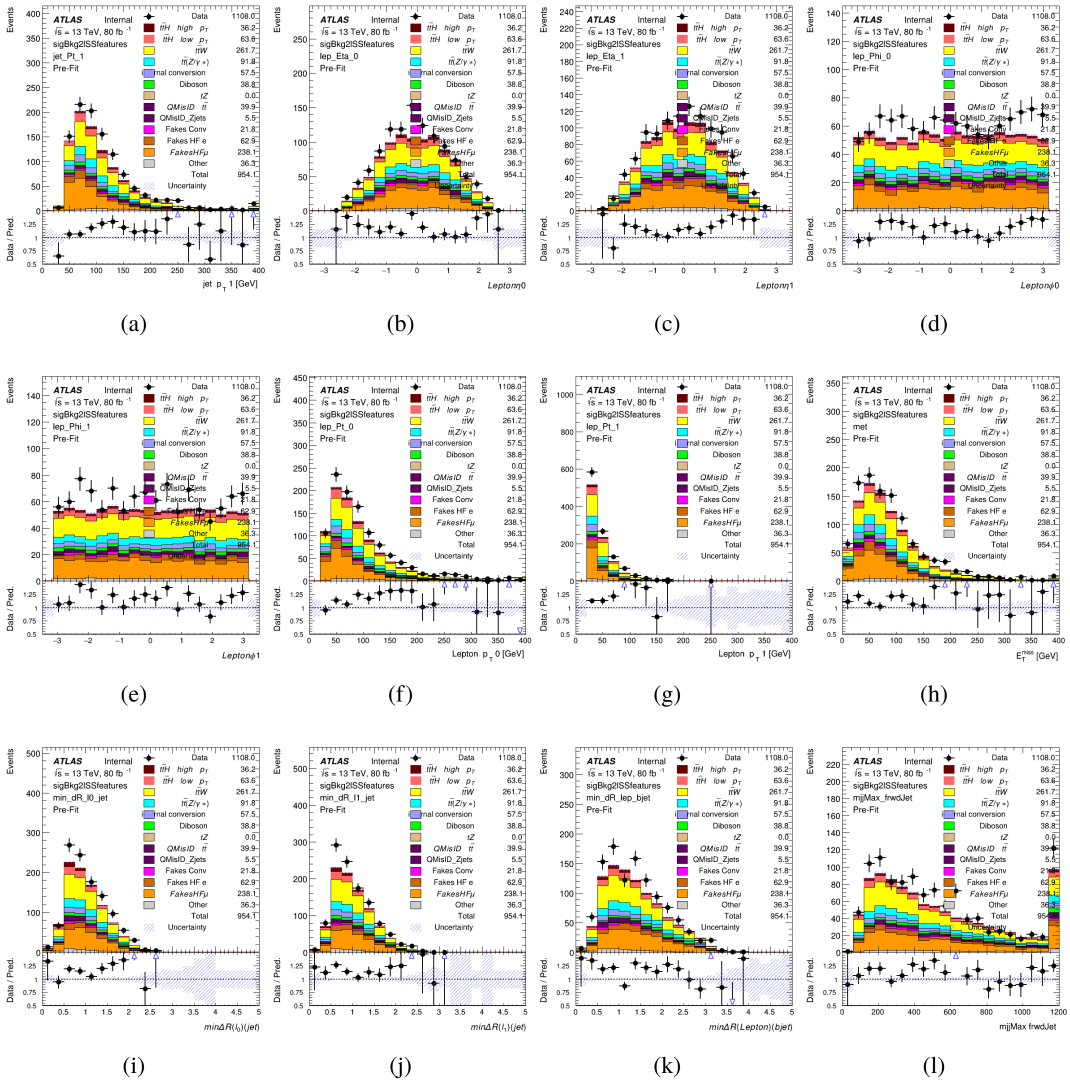


Figure C.19: Input features for sigBkg2lSS

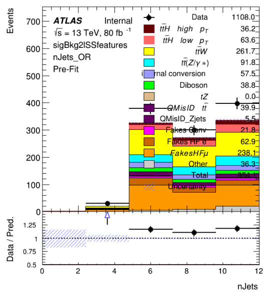


(a) (b) (c) (d)

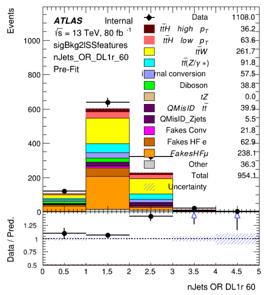
(e) (f) (g) (h)

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

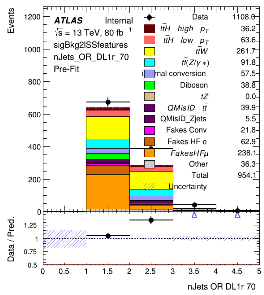
Figure C.20: Input features for sigBkg2lSS



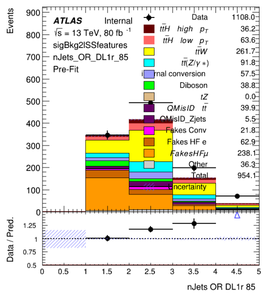
(a)



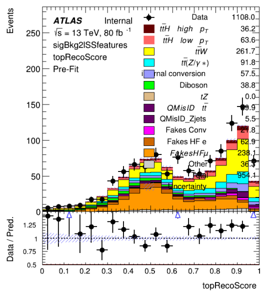
(b)



(c)



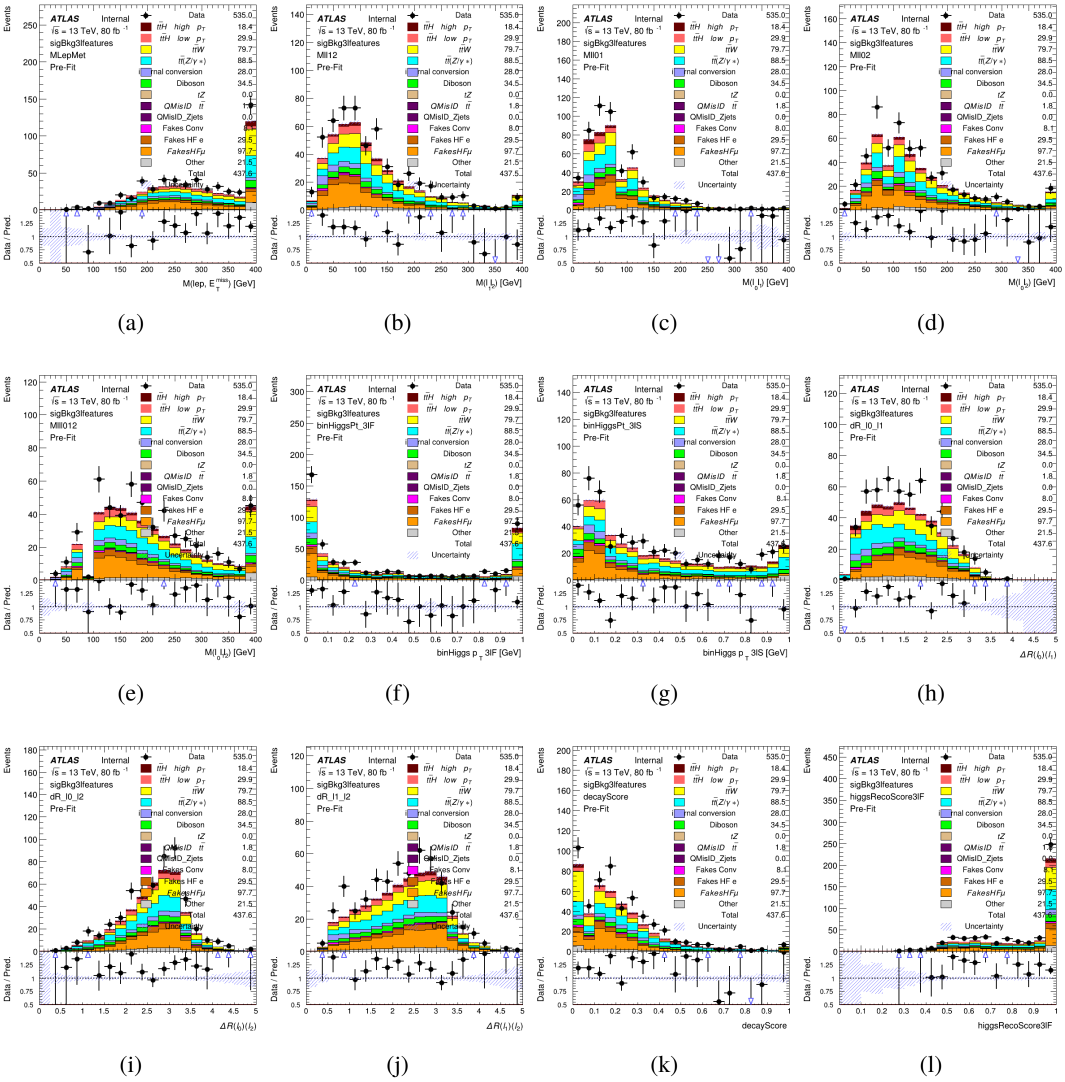
(d)



(e)

Figure C.21: Input features for sigBkg2lSS

C.2.2 Background Rejection MVA Features - 3l

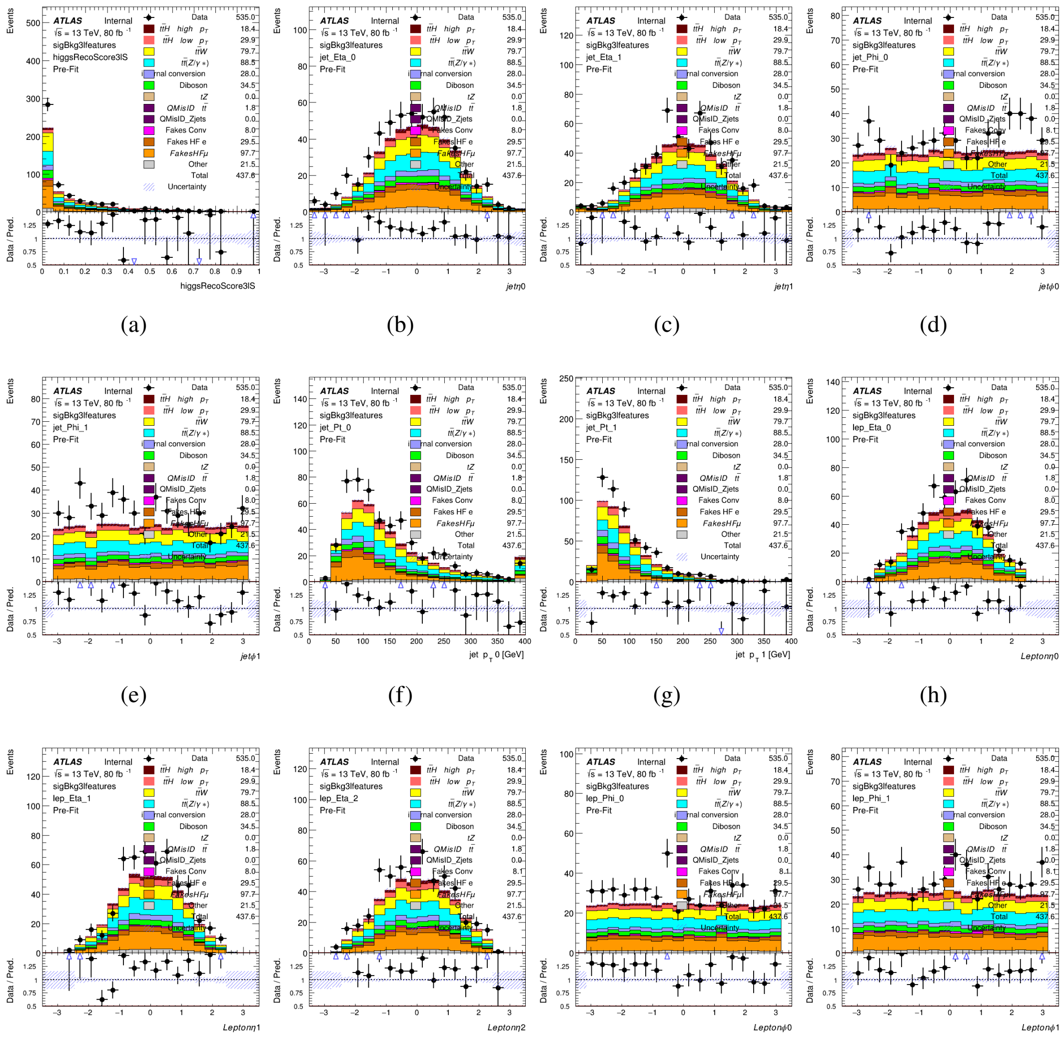


(a) (b) (c) (d)

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

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

Figure C.22: Input features for sigBkg3l

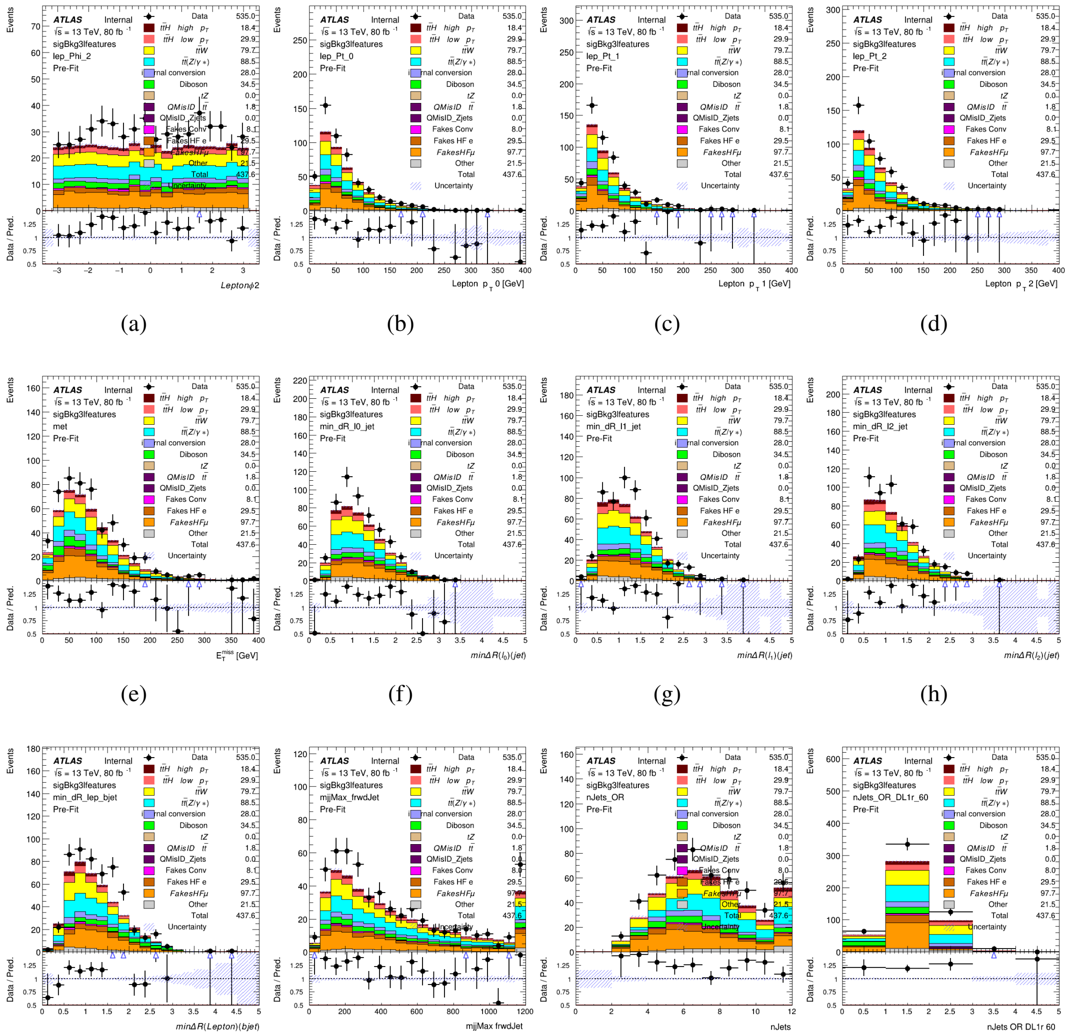


(a) (b) (c) (d)

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

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

Figure C.23: Input features for sigBkg3l

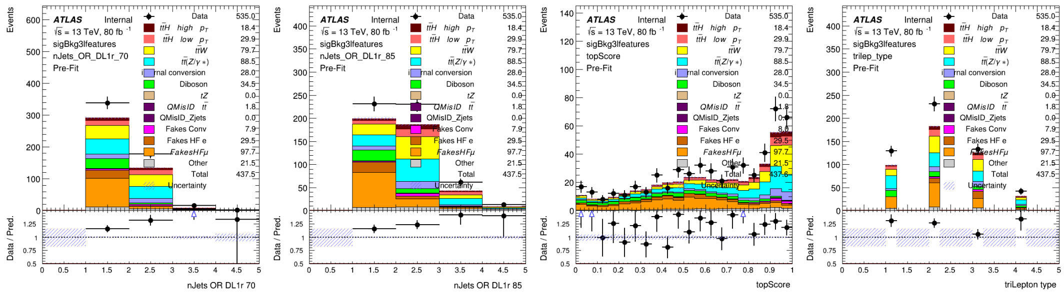


(a) (b) (c) (d)

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

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

Figure C.24: Input features for sigBkg3l



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

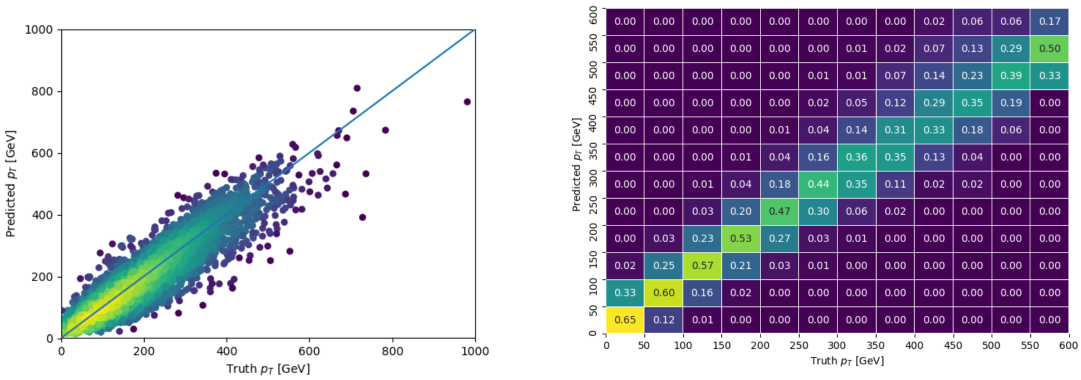
Figure C.25: Input features for sigBkg3l

C.3 Truth Level Studies

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

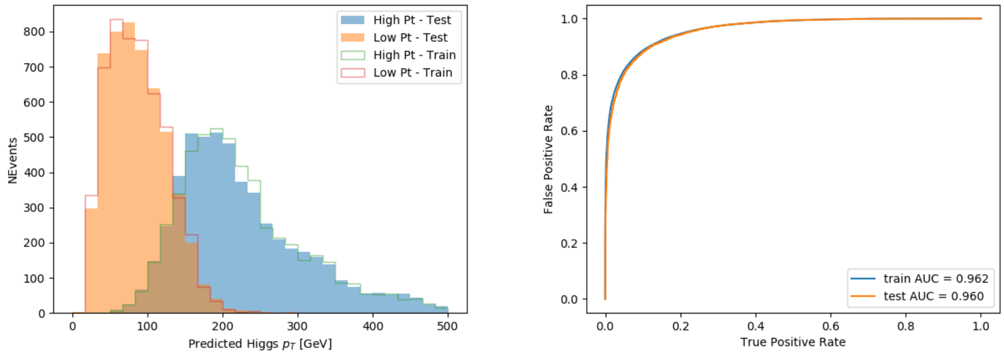
Only leptons and jets which are truth matched to the Higgs are used as inputs for the model; events where the lepton and both jets are not reconstructed are not included. The model uses the same feature set and network architecture as the p_T prediction model used in the main analysis, as described in Section 15.4.1.

The results of the model are summarized below:



(a) (b)

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



(a) (b)

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

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

C.4 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.

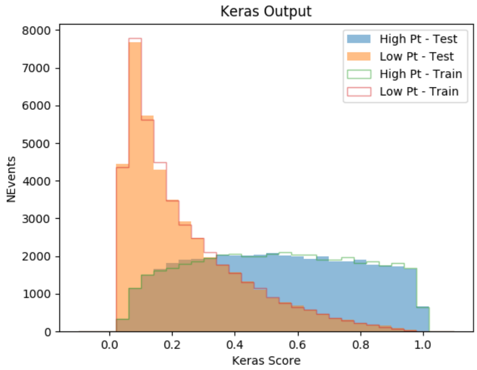
Table 45: Accuracey of the NN in identifying b-jets from tops in 2lSS events for the alternate categorical method compared to the nominal method.

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

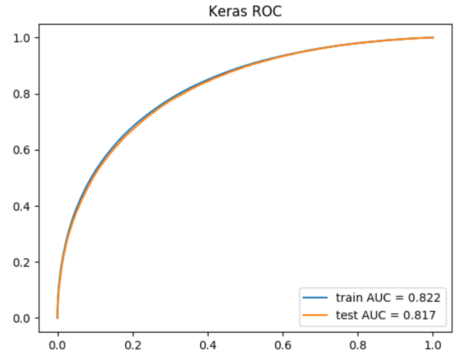
C.5 Binary Classification of the Higgs p_T

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 15.4. Binary crossentropy is used as the loss function.

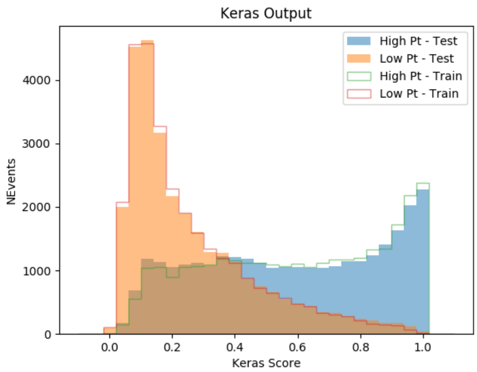


(a)

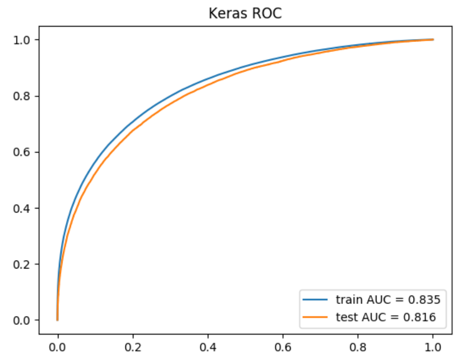


(b)

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

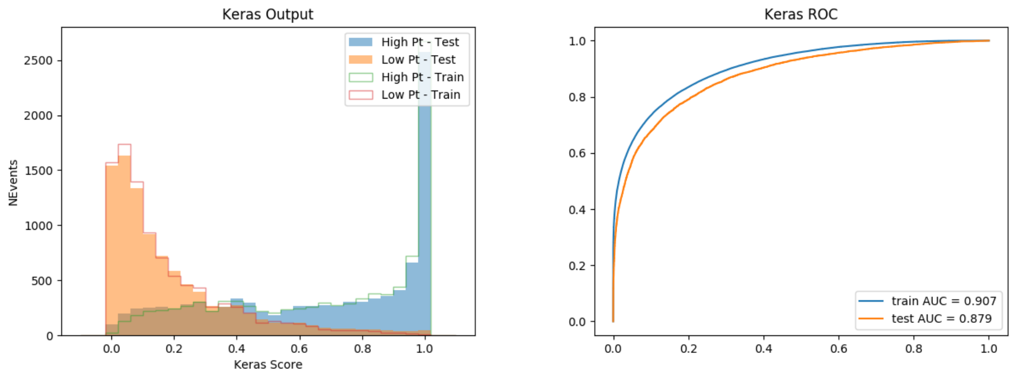


(a)



(b)

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



(a) (b)

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

C.6 Impact of Alternative Jet Selection

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

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

Jet p_T > 20 GeV

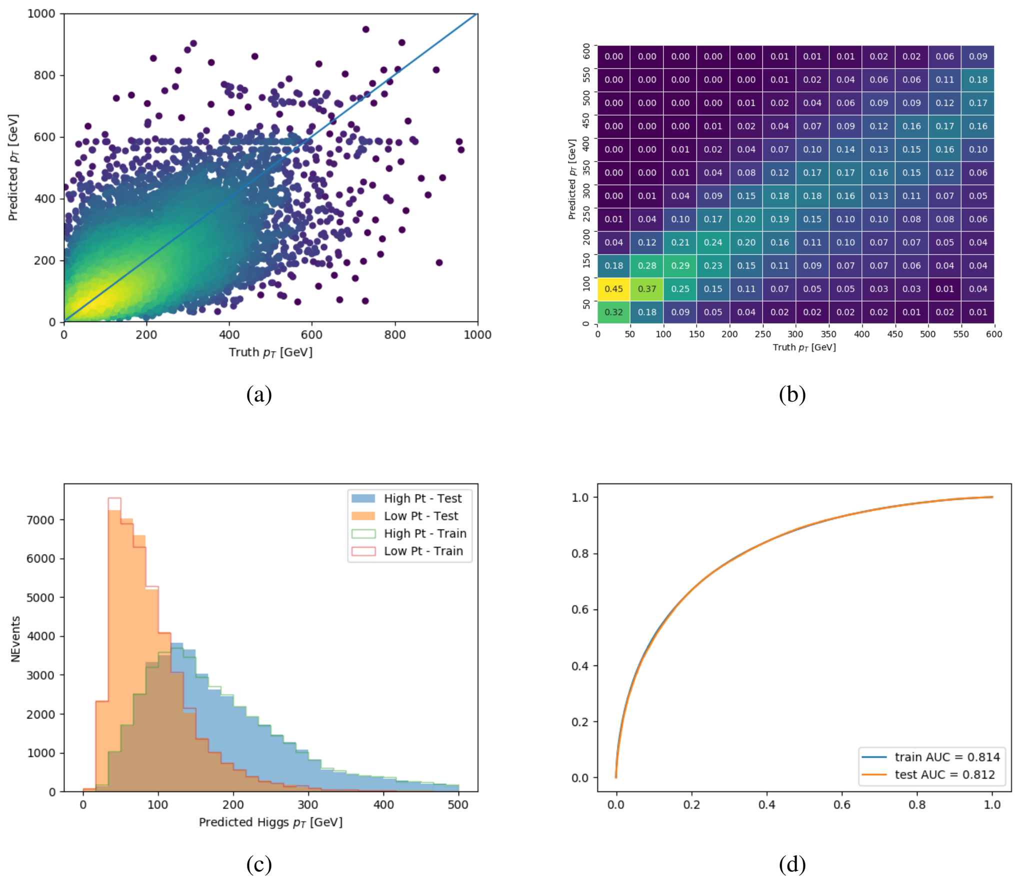


Figure C.31: Output of the model designed to predict the Higgs momentum in the 2lSS channel, with the jet p_T cutoff used is raised to 20 GeV.

Jet $p_T > 25$ GeV

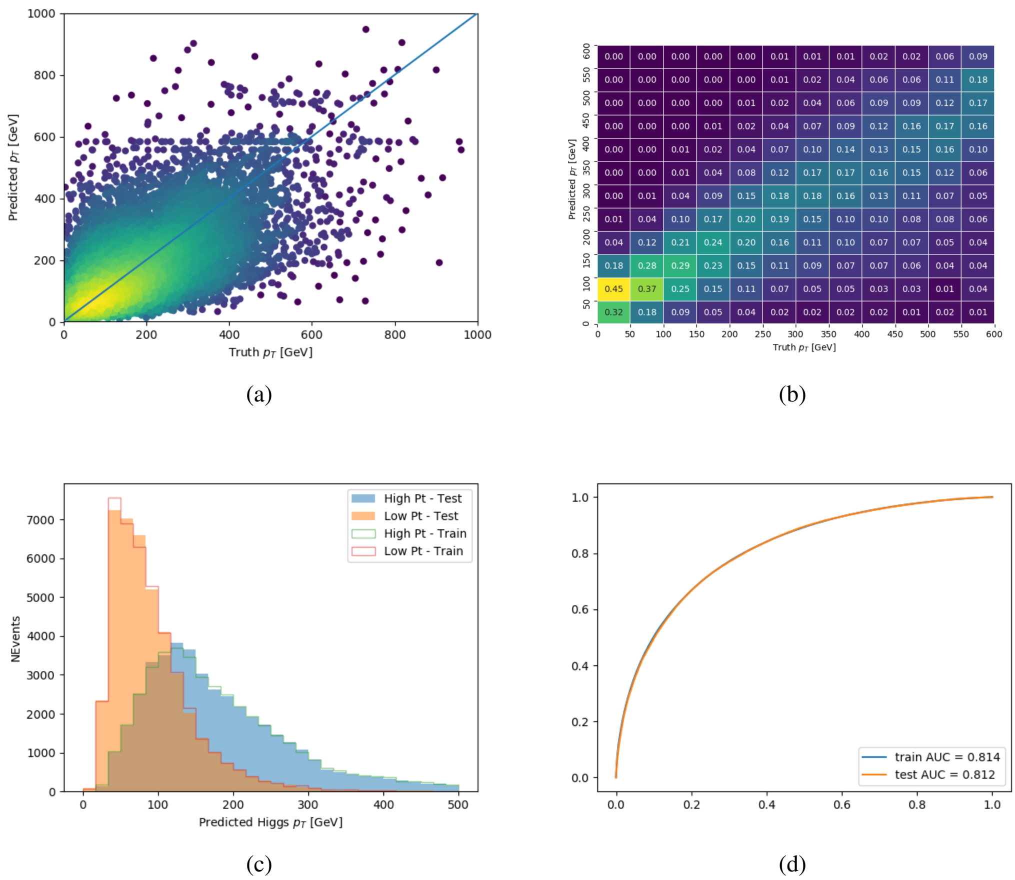


Figure C.32: Output of the model designed to predict the Higgs momentum in the 2ℓSS channel, with the jet p_T cutoff used is raised to 25 GeV.

D