



ATLAS Note

GROUP-2017-XX



24th February 2021

Draft version 0.1

1

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

5

The ATLAS Collaboration

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

© 2021 CERN for the benefit of the ATLAS Collaboration.

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

16 **Contents**

17	I Introduction	4
18	1 Introduction	4
19	II Theoretical Motivation	6
20	2 The Standard Model and the Higgs Boson	6
21	2.1 The Forces and Particles of the Standard Model	7
22	2.2 The Higgs Mechanism	9
23	2.2.1 The Higgs Field	10
24	2.2.2 Electroweak Symmetry Breaking	12
25	2.3 Limitations of the Standard Model	14
26	3 Effective Field Theory in $t\bar{t}H$ Production	15
27	3.1 Extensions to the Higgs Sector	15
28	3.2 Six Dimensional Operators	15
29	III The LHC and the ATLAS Detector	16
30	4 The LHC	16
31	5 The ATLAS Detector	19
32	5.1 Inner Detector	20
33	5.2 Calorimeters	21
34	5.3 Muon Spectrometer	22
35	5.4 Trigger System	23
36	IV Measurement of WZ + Heavy Flavor	24
37	6 Introduction	24
38	7 Data and Monte Carlo Samples	26
39	7.1 Data Samples	26
40	7.2 Monte Carlo Samples	27
41	8 Object Reconstruction	29
42	8.1 Trigger	30
43	8.2 Light leptons	30
44	8.3 Jets	31
45	8.4 B-tagged Jets	32

46	8.5	Missing transverse energy	33
47	9	Event Selection and Signal Region Definitions	34
48	9.1	Event Preselection	34
49	9.2	Fit Regions	37
50	9.3	Non-Prompt Lepton Estimation	52
51	9.3.1	$t\bar{t}$ Validation	53
52	9.3.2	Z+jets Validation	58
53	10	tZ Interference Studies and Separation Multivariate Analysis	63
54	10.1	Top Mass Reconstruction	63
55	10.2	tZ BDT	65
56	11	Data and Monte Carlo Samples	69
57	11.1	Data Samples	70
58	11.2	Monte Carlo Samples	71
59	12	Object Reconstruction	73
60	12.1	Trigger	74
61	12.2	Light leptons	74
62	12.3	Jets	75
63	12.4	B-tagged Jets	76
64	12.5	Missing transverse energy	77
65	13	Event Selection and Signal Region Definitions	78
66	13.1	Event Preselection	78
67	13.2	Fit Regions	81
68	13.3	Non-Prompt Lepton Estimation	96
69	13.3.1	$t\bar{t}$ Validation	97
70	13.3.2	Z+jets Validation	102
71	14	tZ Interference Studies and Separation Multivariate Analysis	107
72	14.1	Top Mass Reconstruction	107
73	14.2	tZ BDT	109
74	15	Systematic Uncertainties	113
75	16	Results	120
76	16.1	1-jet Fit Results	122
77	16.2	2-jet Fit Results	130
78	V	Differential Studies of $t\bar{t}H$ Multilepton	139
79	17	Data and Monte Carlo Samples	139

80	17.1 Data Samples	140
81	17.2 Monte Carlo Samples	140
82	18 Object Reconstruction	141
83	18.1 Trigger Requirements	141
84	18.2 Light Leptons	142
85	18.3 Jets	143
86	18.4 Missing Transverse Energy	143
87	19 Higgs Momentum Reconstruction	144
88	19.1 Decay Candidate Reconstruction	145
89	19.2 b-jet Identification	146
90	19.2.1 2lSS Channel	147
91	19.2.2 3l Channel	151
92	19.3 Higgs Reconstruction	154
93	19.3.1 2lSS Channel	155
94	19.3.2 3l Semi-leptonic Channel	159
95	19.3.3 3l Fully-leptonic Channel	163
96	19.4 p_T Prediction	166
97	19.4.1 2lSS Channel	168
98	19.4.2 3l Semi-leptonic Channel	171
99	19.4.3 3l Fully-leptonic Channel	173
100	19.5 3l Decay Mode	177
101	20 Signal Region Definitions	179
102	20.1 Pre-MVA Event Selection	180
103	20.2 Event MVA	185
104	20.3 Signal Region Definitions	191
105	21 Background Rejection MVA	195
106	21.1 Pre-MVA Event Selection	196
107	21.2 Event MVA	197
108	21.3 Signal Region Definitions	203
109	21.3.1 2lSS	206
110	21.3.2 3l – Semi – leptonic	206
111	21.3.3 3l – Fully – leptonic	206
112	22 Systematic Uncertainties	206
113	23 Results	208
114	23.1 Results - 80 fb^{-1}	208
115	23.2 Projected Results - 140 fb^{-1}	211

116	VI Conclusion	213
117	Appendices	215
118	A Machine Learning Models	215
119	A.1 Higgs Reconstruction Models	215
120	A.1.1 b-jet Identification Features - 2lSS	215
121	A.1.2 b-jet Identification Features - 3l	219
122	A.1.3 Higgs Reconstruction Features - 2lSS	224
123	A.1.4 Higgs Reconstruction Features - 3lS	229
124	A.1.5 Higgs Reconstruction Features - 3lF	234
125	A.2 Background Rejection MVAs	238
126	A.2.1 Background Rejection MVA Features - 2lSS	238
127	A.2.2 Background Rejection MVA Features - 3l	242
128	A.3 Alternate b-jet Identification Algorithm	247
129	A.4 Binary Classification of the Higgs p_T	247
130	A.5 Impact of Alternative Jet Selection	249

Part I**Introduction****1 Introduction**

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

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

147 A promising avenue for addressing these problems is to study the properties of the Higgs
148 Boson and the way it interacts with other particles, in part simply because these interactions

149 have not been measured before. Its interactions with the Top Quark are a particularly promising
150 place to look. Because the Higgs Field is responsible for allowing particle to acquire mass, the
151 strength of a particle's interaction with the Higgs Boson is proportional to its mass. As the most
152 massive of the fundamental particles, the Top Quark has the strongest coupling to the Higgs
153 Boson, meaning any new physics in the Higgs sector is likely to present itself most prominently
154 in its interaction with the Top Quark.

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

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

163 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-
164 2018 provides the opportunity to make this measurement for the first time. The unprecedented
165 energy achieved by the LHC during this period greatly increase the rate at which $t\bar{t}H$ events are
166 produced, and the large amount of data collected provides the necessary statistics for a differential
167 measurement to be performed.

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

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

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

178 Part II

179 Theoretical Motivation

180 2 The Standard Model and the Higgs Boson

181 The Standard Model of particle physics (SM) is a Quantum Field Theory (QFT) describing the
182 known fundamental particles and their interactions. It accounts for three of the four known
183 fundamental force - electromagnetism, the weak nuclear force, and the strong nuclear force, but
184 not gravity. Further, the SM describes a mechanism for combining the weak and electromagnetic
185 forces into a singular interaction, known as the electroweak force. It is a non-Abelian gauge

¹⁸⁶ theory, invariant under the Lie Group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where C refers to color

¹⁸⁷ charge, L, the helicity of the particle, and Y, the hypercharge.

¹⁸⁸ **2.1 The Forces and Particles of the Standard Model**

¹⁸⁹ The SM particles, summarized in figure 2.1, can be classified into two general categories based

¹⁹⁰ on their spin: fermions, and bosons.

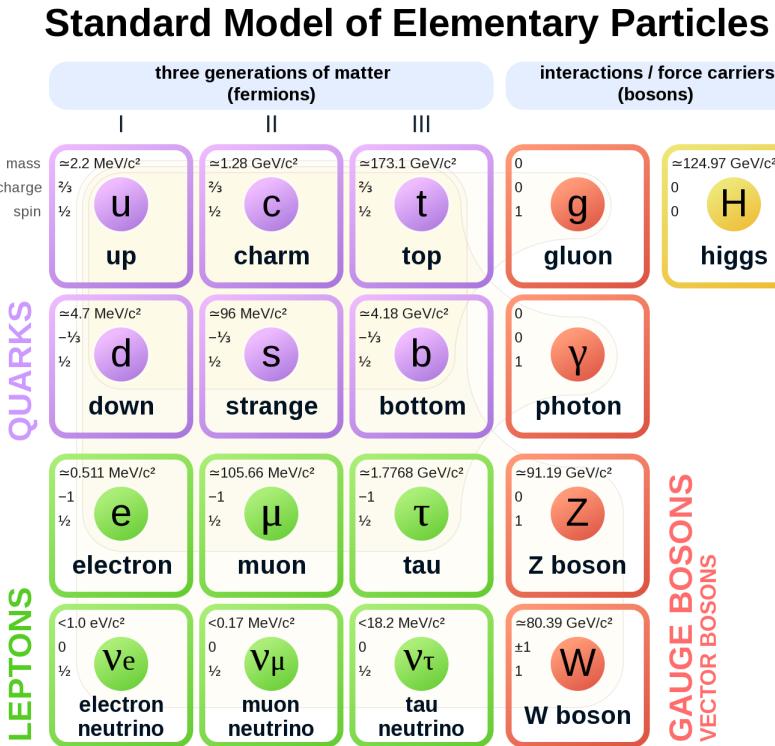


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. []

¹⁹¹ Fermions are particles with $\frac{1}{2}$ -integer spin, which according to the spin-statistics theorem,

¹⁹² causes them to comply with the Pauli-exclusion principle []. They can be separated into two

¹⁹³ groups, leptons and quarks, each of which consist of three generations of particles with increasing
¹⁹⁴ mass.

¹⁹⁵ Leptons are fermions interact via the electroweak force, but not the strong force. The three
¹⁹⁶ generation of leptons consist of the electron and electron neutrino, the muon and muon neutrino,
¹⁹⁷ the tau and tau neutrino. The quarks, which do interact via the strong force - which is to say they
¹⁹⁸ have color charge - in addition to the electroweak force. The three generations include the up
¹⁹⁹ and down quarks, the strange and charm quarks, and the top and bottom quarks. Each of these
²⁰⁰ generations form left-handed doublets invariant under SU(2) transformations. For the leptons
²⁰¹ these doublets are:

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

²⁰² And for the quarks:

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

²⁰³ For both leptons and quarks, the heavier generations can decay into the lighter generation
²⁰⁴ of particles, while the first generation does not decay. Hence, ordinary matter generally consists
²⁰⁵ of this first generation of fermions - electrons, up quarks, and down quarks. Each of these
²⁰⁶ fermions has a corresponding anti-particle, which has an equal mass as its partner but opposite

207 charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos,
208 whose mass has been experimentally confirmed but is not accounted for in the SM.

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

215 **2.2 The Higgs Mechanism**

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

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

225 **2.2.1 The Higgs Field**

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

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}_L \quad (2.3)$$

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

$$V(\Phi) = \mu^2 |\Phi^\dagger \Phi| + \lambda (|\Phi^\dagger \Phi|)^2 \quad (2.4)$$

228 Where μ and λ are free parameters of the new field. This represents the most general
229 potential allowed while preserving $SU(2)_L$ invariance and renormalizability. In the case that
230 $\mu^2 < 0$, this potential takes the form shown in figure 2.2.

231 The significant feature of this potential is that its minimum does not occur for a value of
232 $\Phi = 0$. Instead, it is minimized when $|\Phi^\dagger \Phi| = -\mu^2/\lambda$. This means that in its ground state, the
233 Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while
234 the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

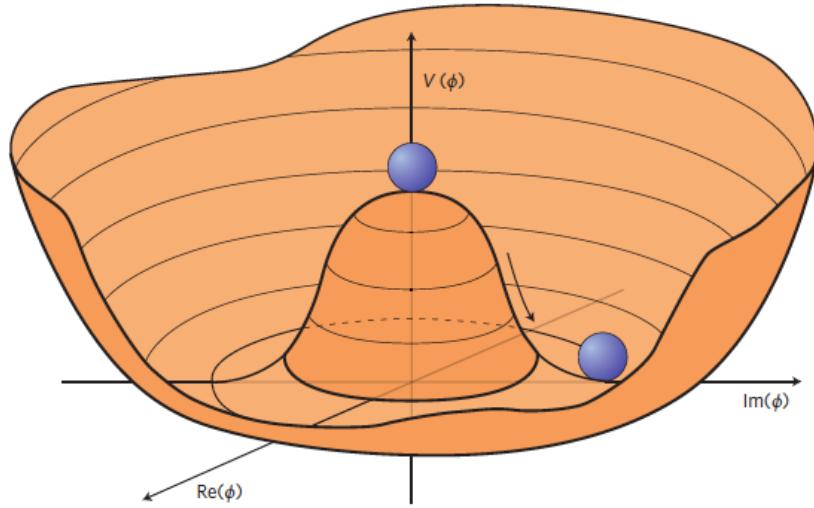


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ , for the case that $\mu^2 < 0$ [].

²³⁵ the minimum is determined only by $\Phi^\dagger \Phi$, there is some ambiguity in the particular definition of
²³⁶ the VEV, but it is generally represented as

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (2.5)$$

²³⁷ The full value of Φ can be written as

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H/\sqrt{2} \end{pmatrix} \quad (2.6)$$

²³⁸ with v being the value of the VEV, and H being the real value of the scalar field.

²³⁹ **2.2.2 Electroweak Symmetry Breaking**

²⁴⁰ The Electroweak (EWK) interaction is described in the SM by a $SU(2)_L \otimes U(1)_Y$ gauge theory.
²⁴¹ This theory predicts three $SU(2)_L$ gauge boson, $W_\mu^1, W_\mu^2, W_\mu^3$, and a single $U(1)_Y$ gauge boson,
²⁴² B_μ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar
²⁴³ field Φ in the Lagrangian:

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

²⁴⁴ Here D_μ represents the covariant derivative required to preserve gauge invariance, g and
²⁴⁵ g' represent coupling constant of the gauge bosons, σ^a denotes the Pauli matrices of $SU(2)$,
²⁴⁶ and Y represents the hypercharge of $U(1)$. The terms in this interaction which contribute to the
²⁴⁷ masses of the gauge bosons can be written as:

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

²⁴⁸ Expanding these terms into the mass eigenstates of the electroweak interaction yields four
²⁴⁹ physical gauge bosons, two charged and two neutral, which are linear combinations of the fields

²⁵⁰ $W_\mu^1, W_\mu^2, W_\mu^3$, and B_μ :

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

²⁵¹ And the masses of these fields are given by:

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

²⁵² This produces exactly the particles we observe - three massive gauge bosons and a single
²⁵³ massless photon. The massless photon represents the portion of the gauge symmetry, a single
²⁵⁴ $U(1)$ of the electromagnetic force, that remains unbroken by the VEV.

²⁵⁵ Interactions with the Higgs field also lead to the generation of the fermion masses, which
²⁵⁶ in the Lagrangian take the form:

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

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

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

264 The fermion's Yukawa coupling to the VEV take the same form as the fermion's coupling
265 to the Higgs boson - λ_ψ . Therefore, the strength of a fermion's interaction with the Higgs is
266 directly proportional to its mass. We now have a model that predicts a Higgs boson with mass
267 $M_H = \mu$, which interacts with the fermions with coupling strength λ_ψ . Because μ and λ_ψ are
268 free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions
269 must be measured experimentally.

270 2.3 Limitations of the Standard Model

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

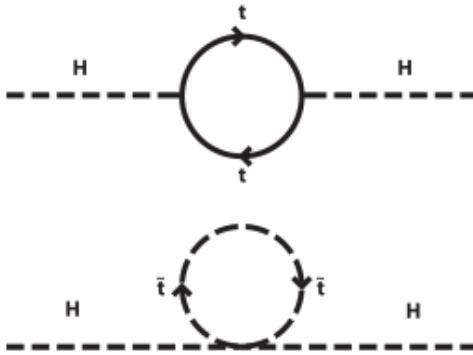


Figure 2.3: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram [6].

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

²⁷⁵ Higher dimension operators are a common way to paramaterize the effects of physics at very
²⁷⁶ high energies into

²⁷⁷ **3.1 Extensions to the Higgs Sector**

²⁷⁸ **3.2 Six Dimensional Operators**

²⁷⁹ While the SM has been tested to great precision, particularly at the LHC, it is generally accepted
²⁸⁰ that it is only valid up to a certain energy scale. It is assumed that above a certain energy, at the
²⁸¹ scale where something like a Grand Unified Theory (GUT) or quantum gravity become relevant,
²⁸² the SM will not be applicable.

Part III**284 The LHC and the ATLAS Detector****285 4 The LHC**

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

295 The LHC first began running in 2009 at a proton-proton center of mass energy of $\sqrt{s} = 8$
296 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during
297 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,
298 and collected data at an increased energy of $\sqrt{s} = 13$ TeV until 2018, a period referred to as Run
299 2.

300 The LHC consists of a chain of accelerators, which accelerate the protons to higher and

higher energies until they are injected into the main ring. This process is summarized in figure 4.1. Protons extracted from a tank of ionized hydrogen are fed into a linear accelerator, LINAC2, where they reach an energy of 50 MeV. From there, they enter a series of three separate circular accelerators, before being injected into the main accelerator ring at an energy of 450 GeV. Within the main ring protons are separated into two separate 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.

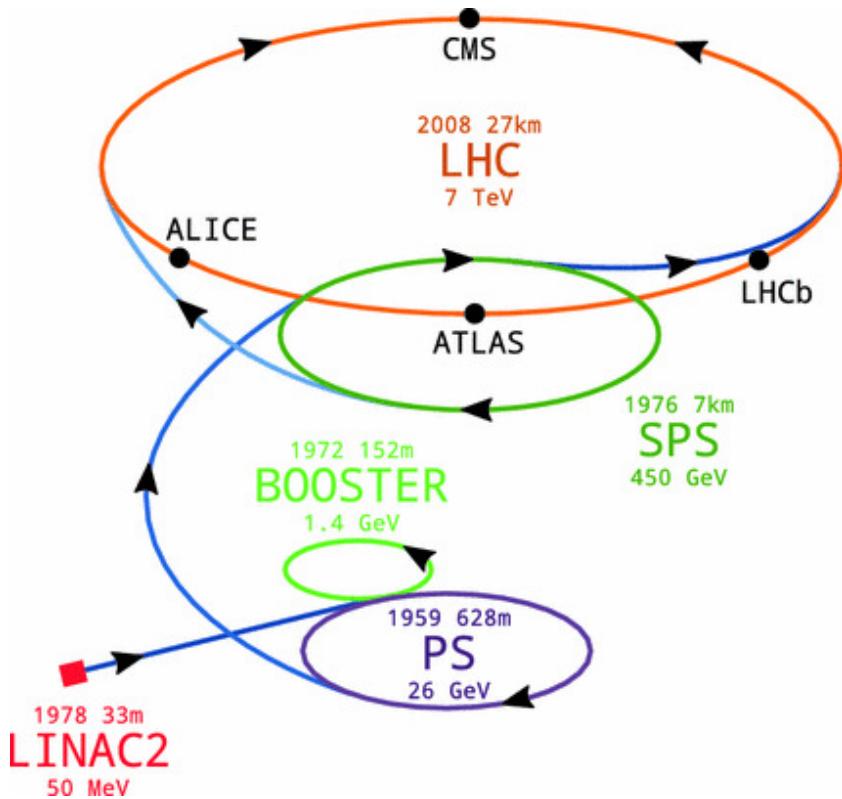


Figure 4.1: A summary of the accelerator chain used to feed protons into the LHC [].

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

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

$$\mathcal{L} = \frac{1}{\sigma} \frac{dN}{dt} \quad (4.1)$$

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

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} dt \quad (4.2)$$

319 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around
 320 140 fb^{-1} , exceeding the expected integrated luminosity of 100 fb^{-1} .

321 5 The ATLAS Detector

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

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

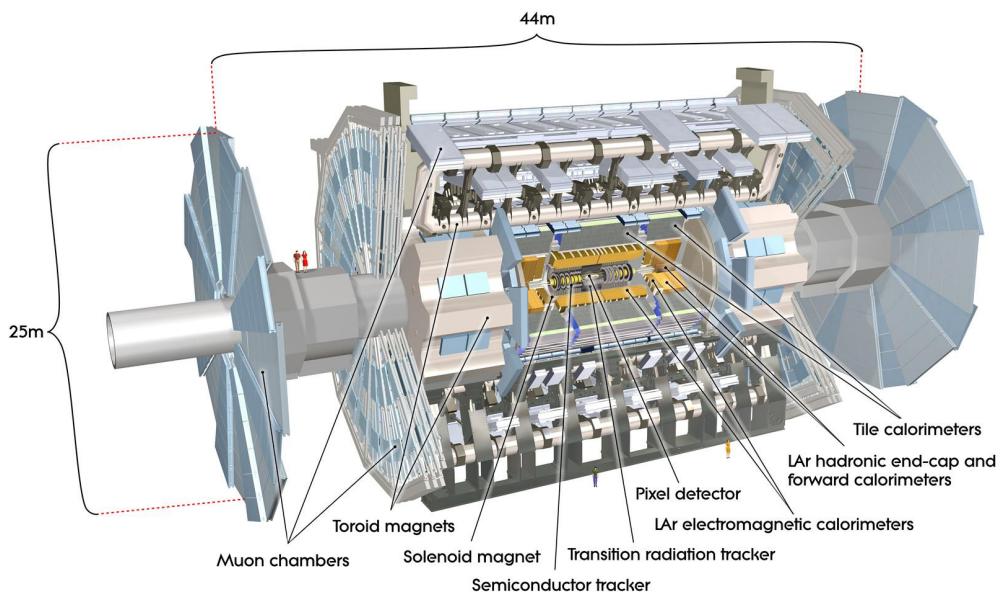


Figure 5.1: Cutaway view of the ATLAS detector, with labels of its major components [].

³³⁰ **5.1 Inner Detector**

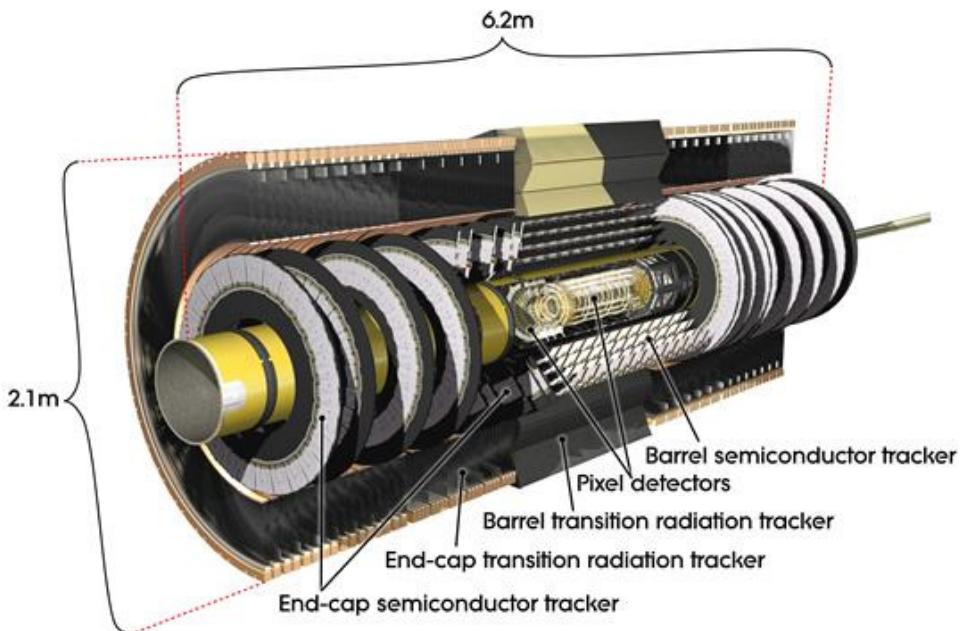


Figure 5.2: Cutaway view of the Inner Detector [].

³³¹ Just surrounding the interaction point is the Inner Detector, designed to track the path
³³² of charged particles moving through the detector. An inner solenoid surrounding the Innder
³³³ Detector is used to produces a magnetic field of 2 T. This large magnetic field causes the path
³³⁴ of charged particles moving through the Inner Detector to bend. Because this magnetic field is
³³⁵ uniform and well known, it can be used in conjunction with the curvature of a particles path to
³³⁶ measure its charge and momentum.

³³⁷ The Inner Detector consists of three components - the Pixel Detector, the Semi-Conductor
³³⁸ Tracker (SCT), and the Transition Radiation Tracker (TRT). The Pixel Detector is the innermost
³³⁹ of these, beginning just 33.25 mm away from the beam line. It consists of three silicon layers

³⁴⁰ along the barrel, as well as three endcap layers, covering a range of $|\eta| < 2.5$.

³⁴¹ The Semiconductor Tracker (SCT) is similar to the Pixel detector, but uses long strips
³⁴² rather than small pixel to cover a larger spatial area.

³⁴³ **5.2 Calorimeters**

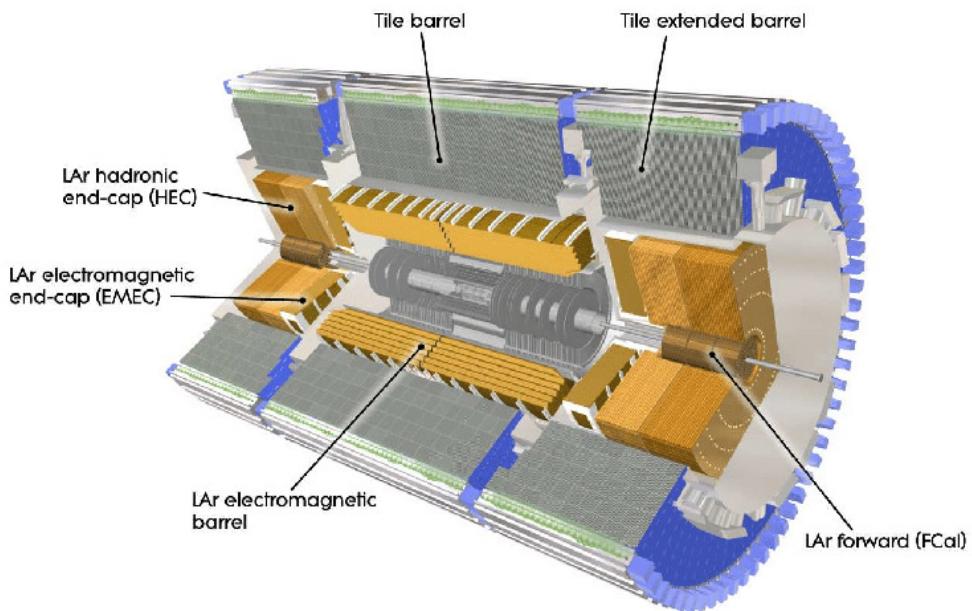


Figure 5.3: Cutaway view of the calorimeter system of the ATLAS detector [].

³⁴⁴ Situated outside the Inner Detector are two concentric calorimeters. The inner calorimeter
³⁴⁵ uses liquid argon (LAr) to measure energy of particles that interact electromagnetically, which
³⁴⁶ includes photons and any charged particle. The LAr calorimeter is made of heavy metals,
³⁴⁷ primarily lead and copper, which causes electromagnetically interacting particles to shower,
³⁴⁸ depositing their energy in the detector. The showering of the high energy particles that pass

349 through calorimeter cause the liquid argon to ionize, and the ionized electrons are detected by
350 electronic readouts. The LAr calorimeter consists of around 180,000 readout channels.

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

356 **5.3 Muon Spectrometer**

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

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

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

371 1200 tracking chambers are placed in the muon system in order to precisely measure the
372 tracks of muons with high spatial resolution.

373 **5.4 Trigger System**

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

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

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

392 **Part IV**

393 **Measurement of WZ + Heavy Flavor**

394 **6 Introduction**

395 The production of WZ in association with a heavy flavor jet represents an important background
396 for many major analyses. This includes any process with leptons and b-jets in the final state,
397 such as $t\bar{t}H$, $t\bar{t}W$, and $t\bar{t}Z$. While precise measurements have been made of WZ production
398 [WZ_36], WZ + heavy flavor remains poorly understood. This is largely because the QCD
399 processes involved in the production of the b-jet make it difficult to simulate accurately. This
400 introduces a large uncertainty for analyses that include this process as a background.

401 Motivated by its relevance to the $t\bar{t}H$ multilepton analysis, we perform a study of the fully

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

406 Events that meet this selection criteria are sorted into pseudo-continuous b-tagging regions
407 based on the DL1r b-tag score of their associated jets. This is done to separate WZ + b-jet events
408 from WZ + charm and WZ + light jets. These regions are fit to data in order make a more
409 accurate estimate of the contribution of WZ + heavy-flavor, where heavy-flavor jets include
410 b-jets and charm jets. The full Run-2 dataset collected by the ATLAS detector, representing 139
411 fb^{-1} of data from pp collisions at $\sqrt{s} = 13 \text{ TeV}$, is used for this study.

412 Section 11 details the data and Monte Carlo (MC) samples used in the analysis. The
413 reconstruction of various physics objects is described in section 12. Section 13 describes the
414 event selection applied to these samples, along the definitions of the various regions used in
415 the fit. The multivariate analysis techniques used to separate the tZ background from WZ +
416 heavy flavor are described in section 14. Section 22 describes the various sources of systematic
417 uncertainties considered in the fit. Finally, the results of the analysis are summarized in section
418 23, followed by a brief conclusion in section ??.

419 **The current state of the analysis shows blinded results for the full 2018 dataset.**
420 **Regions containing >5% WZ+b events are blinded, and results are from Asimov, MC only**
421 **fits.**

422 7 Data and Monte Carlo Samples

423 Both data and Monte Carlo samples used in this analysis were prepared in the `xAOD` format,
 424 which was used to produce a `DxAOD` sample in the HIGG8D1 derivation framework. The HIGG8D1
 425 framework is designed for the $t\bar{t}H$ multi-lepton analysis, which targets events with multiple
 426 leptons as well as tau hadrons. This framework skims the dataset to remove unneeded variables
 427 as well as entire events. Events are removed from the derivations that do not meet the following
 428 selection:

- 429 • at least two light leptons within a range $|\eta| < 2.6$, with leading lepton $p_T > 15$ GeV and
 430 subleading lepton $p_T > 5$ GeV
- 431 • at least one light lepton with $p_T > 15$ GeV within a range $|\eta| < 2.6$, and at least two hadronic
 432 taus with $p_T > 15$ GeV.

433 Samples were then generated from these HIGG8D1 derivations with p-tag of p4134 using
 434 AnalysisBase version 21.2.127 modified to include custom variables..

435 7.1 Data Samples

436 The study uses a sample of proton-proton collision data collected by the ATLAS detector from
 437 2015 through 2018 at an energy of $\sqrt{s} = 13$ TeV, which represents an integrated luminosity of
 438 139 fb^{-1} . This data set was collected with a bunch crossing rate of 25 ns. All data used in this

439 analysis was verified by data quality checks, having been included in the following Good Run

440 Lists:

441 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02

442 _PHYS_StandardGRL_All_Good_25ns.xml

443 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04

444 _PHYS_StandardGRL_All_Good_25ns.xml

445 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL

446 _All_Good_25ns_Triggerno17e33prim.xml

447 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL

448 _All_Good_25ns_Triggerno17e33prim.xml

449 Runs included from the AllYear period containers are included.

450 7.2 Monte Carlo Samples

451 Several different generators were used to produce Monte Carlo simulations of the signal and

452 background processes. For all samples, the response of the ATLAS detector is simulated using

453 Geant4. The WZ signal samples are simulated using Sherpa 2.2.2 [[sherpa](#)]. Specific information

454 about the Monte Carlo samples being used can be found in table [26](#). A list of the specific samples

455 used by data set ID is shown in table [9](#).

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

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

Sample	DSID
WZ	364253, 364739-42
VV	364250, 364254, 364255, 363355-60, 364890
t̄W	410155
t̄Z	410156, 410157, 410218-20
low mass t̄Z	410276-8
Rare Top	410397, 410398, 410399
single Top	410658-9, 410644-5
three Top	304014
four Top	410080
t̄WW	410081
Z + jets	364100-41
low mass Z + jets	364198-215
W + jets	364156-97
Vγ	364500-35
tZ	410560
tW	410013-4
WtZ	410408
VVV	364242-9
VH	342284-5
WtH	341998
t̄tγ	410389
t̄t	410470
t̄tH	345873-5, 346343-5

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

456 8 Object Reconstruction

- 457 All regions defined in this analysis share a common lepton, jet, and overall event preselection.
- 458 The selection applied to each physics object is detailed here; the event preselection, and the
- 459 selection used to define the various fit regions, is described in section 13.

⁴⁶⁰ **8.1 Trigger**

⁴⁶¹ Events are required to be selected by dilepton triggers, as summarized in table 27.

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

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

⁴⁶² **8.2 Light leptons**

⁴⁶³ Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that

⁴⁶⁴ are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].

⁴⁶⁵ Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Muon candidates

⁴⁶⁶ are reconstructed by combining inner detector tracks with track segments or full tracks in the

⁴⁶⁷ muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have $p_T > 10$ GeV and

468 $|\eta| < 2.5$. Candidates in the transition region between different electromagnetic calorimeter
 469 components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant com-
 470 bining shower shape and track information is used to distinguish real electrons from hadronic
 471 showers (fake electrons). To further reduce the non-prompt electron contribution, the track is
 472 required to be consistent with originating from the primary vertex; requirements are imposed
 473 on the transverse impact parameter significance ($|d_0|/\sigma_{d_0} < 5$) and the longitudinal impact
 474 parameter ($|\Delta z_0 \sin \theta_\ell| < 0.5 \text{ mm}$).

475 Muon candidates are reconstructed by combining inner detector tracks with track segments
 476 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
 477 $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$. The longitudinal impact parameter is the same for both electrons
 478 and muons, while muons are required to pass a slightly tighter transverse impact parameter,
 479 $|d_0|/\sigma_{d_0} < 3$.

480 All leptons are required to be isolated, as defined through the standard `isolationFixedCutLoose`
 481 working point supported by combined performance groups. Leptons are additionally required to
 482 pass a non-prompt BDT selection described in detail in [**ttH_paper**]. Optimized working points
 483 and scale factors for this BDT are taken from that analysis.

484 8.3 Jets

485 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
 486 calorimeters [**ATL-PHYS-PUB-2015-015**], using the anti- k_t algorithm with a radius parameter

487 $R = 0.4$. Jets with energy contributions likely arising from noise or detector effects are removed
 488 from consideration [ATLAS-CONF-2015-029], and only jets satisfying $p_T > 25$ GeV and
 489 $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
 490 association algorithm is used to confirm that the jet originates from the selected primary vertex,
 491 in order to reject jets arising from pileup collisions [PERF-2014-03].

492 **8.4 B-tagged Jets**

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

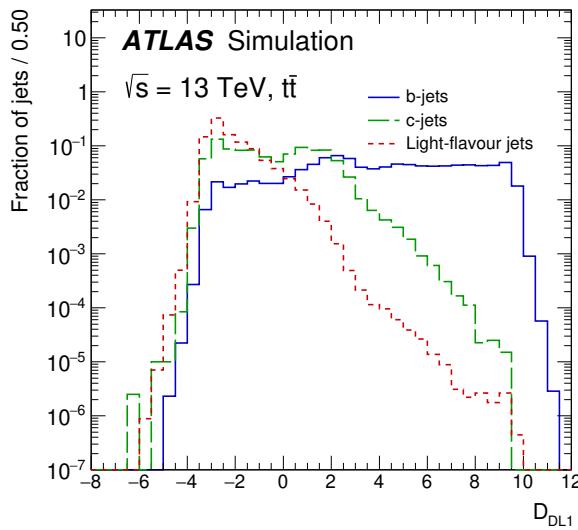


Figure 8.1: Output distribution of the DL1r algorithm for b-jets, charm jets, and light jets

498 From the output of the BDT, working points (WPs) are developed based on the efficiency
 499 of truth b-jets at particular values of the DL1r algorithm. The working points used in this analysis
 500 are summarized in table 11.

WP	none	loose	medium	tight	tightest
b eff.	-	85%	77%	70%	60%

Table 4: B-tagging Working Points by tightness and b-jet efficiency

501 A tighter WP will accept fewer b-jets, but reject a higher fraction of charm and light jets.
 502 Generally, analyses that include b-jets will use a fixed working point, for example, requiring that
 503 a jet pass the 70% threshold. By instead treating these working point as bins, e.g. events with
 504 jets that fall between the 85% and 77% WPs fall into one bin, while events with jets passing the
 505 60% WP fall into another, and looking at the full psuedo-continuous DL1r spectrum of the jets,
 506 additional information can be gained. The psuedo-continuous b-tag spectrum is used in this case
 507 to separate out WZ + b, WZ + charm, and WZ + light.

508 8.5 Missing transverse energy

509 Missing transverse momentum (E_T^{miss}) is used as part of the event selection. The missing
 510 transverse momentum vector is defined as the inverse of the sum of the transverse momenta of
 511 all reconstructed physics objects as well as remaining unclustered energy, the latter of which is
 512 estimated from low- p_T tracks associated with the primary vertex but not assigned to a hard object,
 513 with object definitions taken from [ATL-PHYS-PUB-2015-027]. Light leptons considered in

514 the E_T^{miss} reconstruction are required to have $p_T > 10 \text{ GeV}$, while jets are required to have $p_T > 20$
515 GeV .

516 **9 Event Selection and Signal Region Definitions**

517 Event are required to pass a preselection described in section 13.1 and summarized in table 12.
518 Those that pass this preselection are divided into various fit regions described in section 13.2,
519 based on the number of jets in the event, and the b-tag score of those jets.

520 **9.1 Event Preselection**

521 Events are required to include exactly three reconstructed light leptons passing the requirement
522 described in 12.2, which have a total charge of ± 1 . As the opposite sign lepton is found to be
523 prompt the vast majority of the time [ttH_paper], it is required to be loose and isolated, as
524 defined though the standard `isolationFixedCutLoose` working point supported by combined
525 performance groups. The same sign leptons are required to be very tightly isolated, as per the
526 recommended `isolationFixedCutTight`.

527 The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose
528 charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.
529 the smallest ΔR , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton

530 0 is required to have $p_T > 10$ GeV, while the same sign leptons, 1 and 2, are required to have

531 $p_T > 20$ GeV to reduce the contribution of non-prompt leptons.

532 The invariant mass of at least one pair of opposite sign, same flavor leptons is required

533 to fall within 10 GeV of the mass of the Z boson, 91.2 GeV. Events where one of the opposite

534 sign pairs have an invariant mass less than 12 GeV are rejected in order to suppress low mass

535 resonances.

536 An additional requirement is placed on the missing transverse energy, $E_T^{\text{miss}} > 20$ GeV,

537 and the transverse mass of the W candidate, $m(E_T^{\text{miss}} + l_{\text{other}}) > 30$ GeV, where E_T^{miss} is the

538 missing transverse energy, and l_{other} is the lepton not included in the Z-candidate.

539 Events are required to have one or two reconstructed jets passing the selection described

540 in section 12.3. Events with more than two jets are rejected in order to reduce the contribution

541 of backgrounds such as $t\bar{t}Z$ and $t\bar{t}W$, which tend to have higher jet multiplicity.

Event Selection

Exactly three leptons with charge ± 1

Two tight Iso, tight ID same-charge leptons with $p_T > 20$ GeV

One loose Iso, medium ID opposite charge lepton with $p_T > 10$ GeV

$m(l^+l^-)$ within 10 GeV of 91.2 GeV

Transverse mass of W-candidate, $m_T(E_T^{\text{miss}} + \text{lep}_{\text{other}}) > 30$ GeV

Missing transverse energy, $E_T^{\text{miss}} > 20$ GeV

One or two jets with $p_T > 25$ GeV

Table 5: Summary of the selection applied to events for inclusion in the fit

542 The event yields in the preselection region for both data and Monte Carlo are summarized

543 in table 13.1, which shows good agreement between data and Monte Carlo, and demonstrates that

544 this region consists primarily of WZ events. The WZ events are split into WZ + b, WZ + c, and
 545 WZ + l based on the truth flavor of the associated jet in the event. Specifically, this determination
 546 is made based on the HadronConeExclTruthLabelID of the jet, as recommended by the b-
 547 tagging working group [**BtagWG**]. In this ordering b-jet supersedes charm, which supersedes
 548 light. That is, WZ + l events contain no charm and no b jets at truth level, WZ + c contain at
 549 least one truth charm and no b-jets, and WZ + b contains at least one truth b-jet.

Process	Events
WZ + b	167.6 ± 6.5
WZ + c	1080 ± 40
WZ + l	7220 ± 310
Other VV	850 ± 140
$t\bar{t}W$	16.8 ± 2.3
$t\bar{t}Z$	115 ± 17
rare Top	2.2 ± 0.1
Single top	0.10 ± 0.45
Three top	0.01 ± 0.01
Four top	0.02 ± 0.01
$t\bar{t}WW$	0.23 ± 0.05
Z + jets	600 ± 260
$V + \gamma$	37 ± 54
tZ	190 ± 70
tW	5.5 ± 1.2
WtZ	25.8 ± 1.1
VVV	26.2 ± 0.9
VH	94 ± 7
$t\bar{t}$	108.68 ± 8
$t\bar{t}H$	4.3 ± 0.5
Total	10600 ± 530
Data	10574

Table 6: Event yields in the preselection region at 139.0 fb^{-1}

550 Here Other VV represents diboson processes other than WZ, and consists predominantly
 551 of $ZZ \rightarrow ll\bar{l}\bar{l}$ events where one of the leptons is not reconstructed.

552 Simulations are further validated by comparing the kinematic distributions of the Monte
 553 Carlo with data, which are shown in figure 13.1. Here, bins with 5% or more WZ+b are blinded.

WZ Fit Region - Inclusive

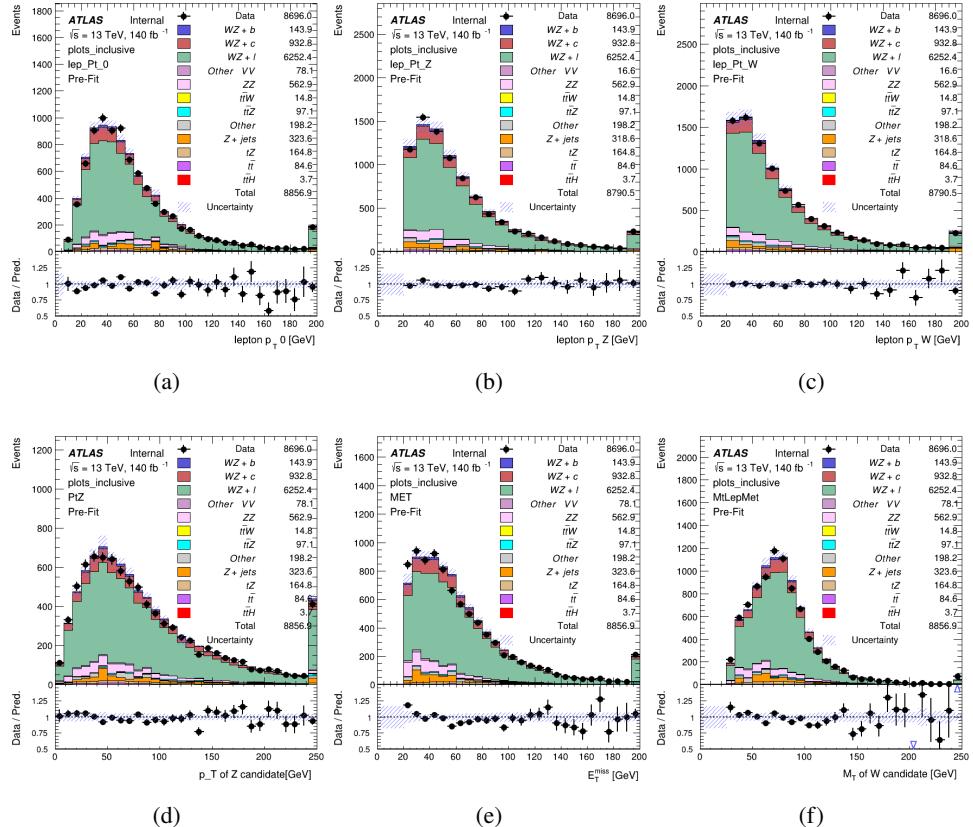


Figure 9.1: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the M_T of the W candidate.

9.2 Fit Regions

554 Once preselection has been applied, the remaining events are categorized into one of twelve
 555 orthogonal regions. The regions used in the fit are summarized in table 14.

Table 7: A list of the regions used in the fit and the selection used for each.

Region	Selection
1j, <85%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_85} = 0$
1j, 85%-77%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_85} = 1, n\text{Jets_DL1r_77}=0$
1j, 77%-70%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_77} = 1, n\text{Jets_DL1r_70}=0$
1j, 70%-60%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_70} = 1, n\text{Jets_DL1r_60}=0$
1j, >60%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_60} = 1, tZ \text{ BDT} > 0.725$
1j tZ CR	$N_{\text{jets}} = 1, n\text{Jets_DL1r_60} = 1, tZ \text{ BDT} < 0.725$
2j, <85%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} = 0$
2j, 85%-77%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} >= 1, n\text{Jets_DL1r_77}=0$
2j, 77%-70%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_77} >= 1, n\text{Jets_DL1r_70}=0$
2j, 70%-60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_70} >= 1, n\text{Jets_DL1r_60}=0$
2j, >60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} >= 1, tZ \text{ BDT} > 0.725$
2j tZ CR	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} >= 1, tZ \text{ BDT} < 0.725$

557 The working points discussed in section 12.4 are used to separate events into fit regions
 558 based on the highest working point reached by a jet in each event. Because the background
 559 composition differs significantly based on the number of b-jets, events are further subdivided
 560 into 1 jet and 2 jet regions in order to minimize the impact of background uncertainties.

561 An additional tZ control region is created based on the BDT described in section 14. The
 562 region with 1-jet passing the 60% working point is split in two - a signal enriched region of
 563 events with a BDT score greater than 0.03, and a tZ control region including events with less
 564 than 0.03. This cutoff is arrived at by performing an Asimov fit with a variety of cutoffs, and
 565 selecting the value that produces the highest significance for the measurement of $WZ + b$.

566 The modeling in each region is validated by comparing data and MC predictions for
 567 various kinematic distributions. These plot are shown in figures 13.2-13.15.

WZ Fit Region - 1j Inclusive

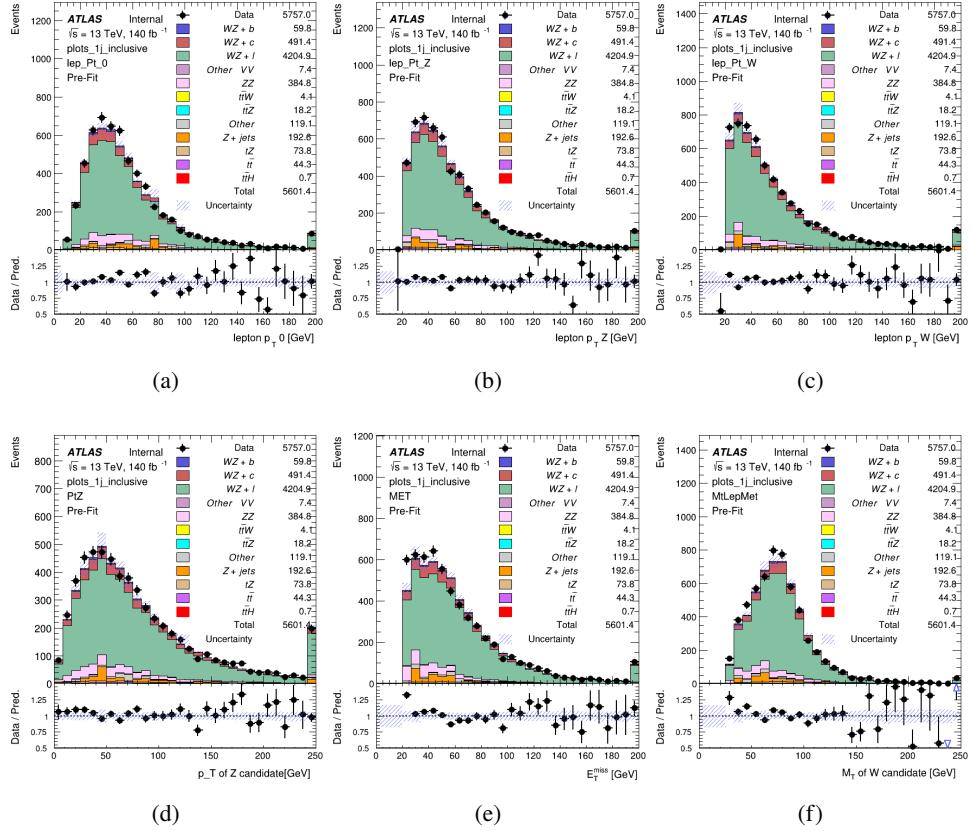


Figure 9.2: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 1j < 85% WP

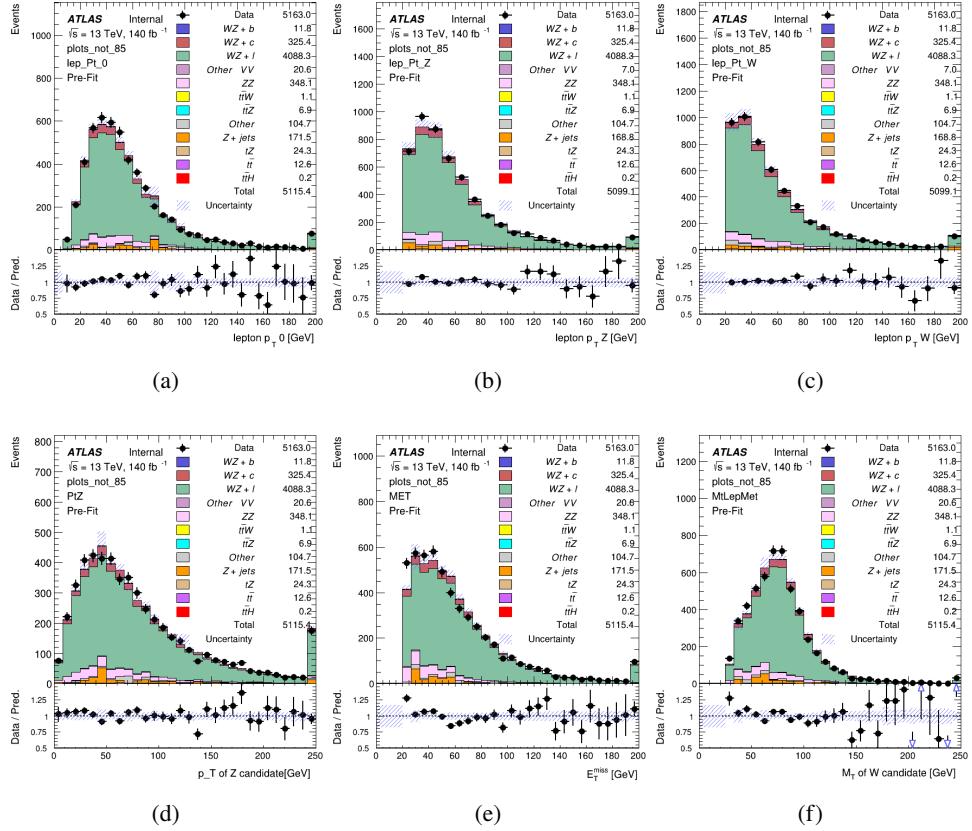


Figure 9.3: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 1j 77-85% WP

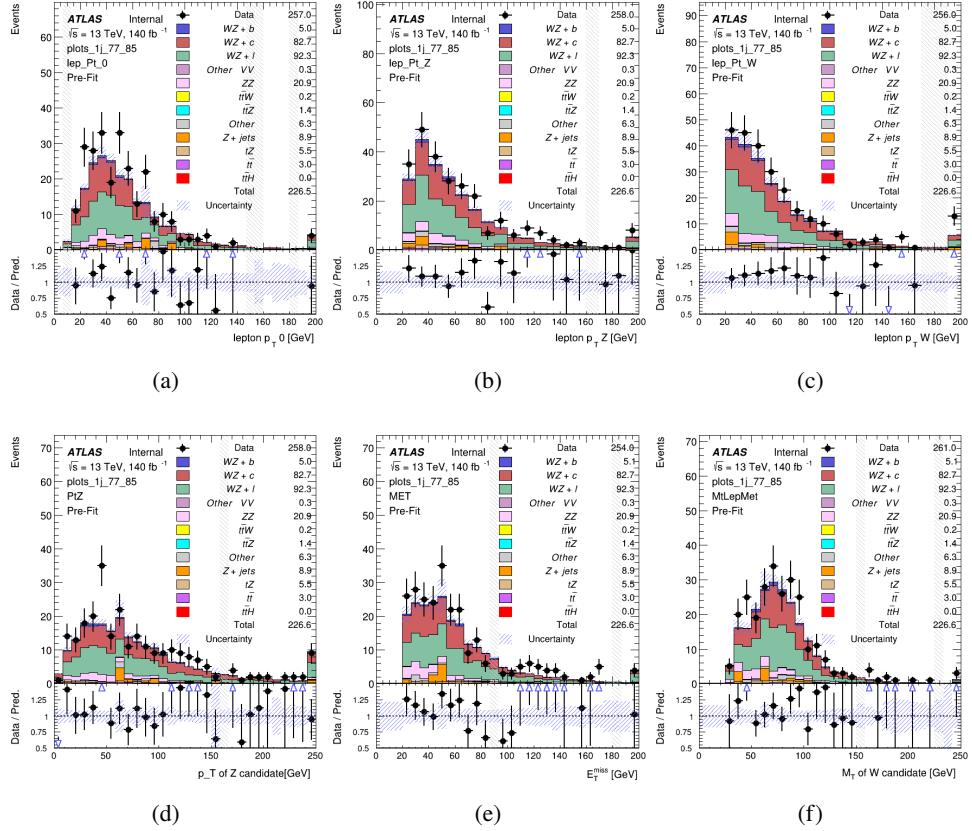


Figure 9.4: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 1j 70-77% WP

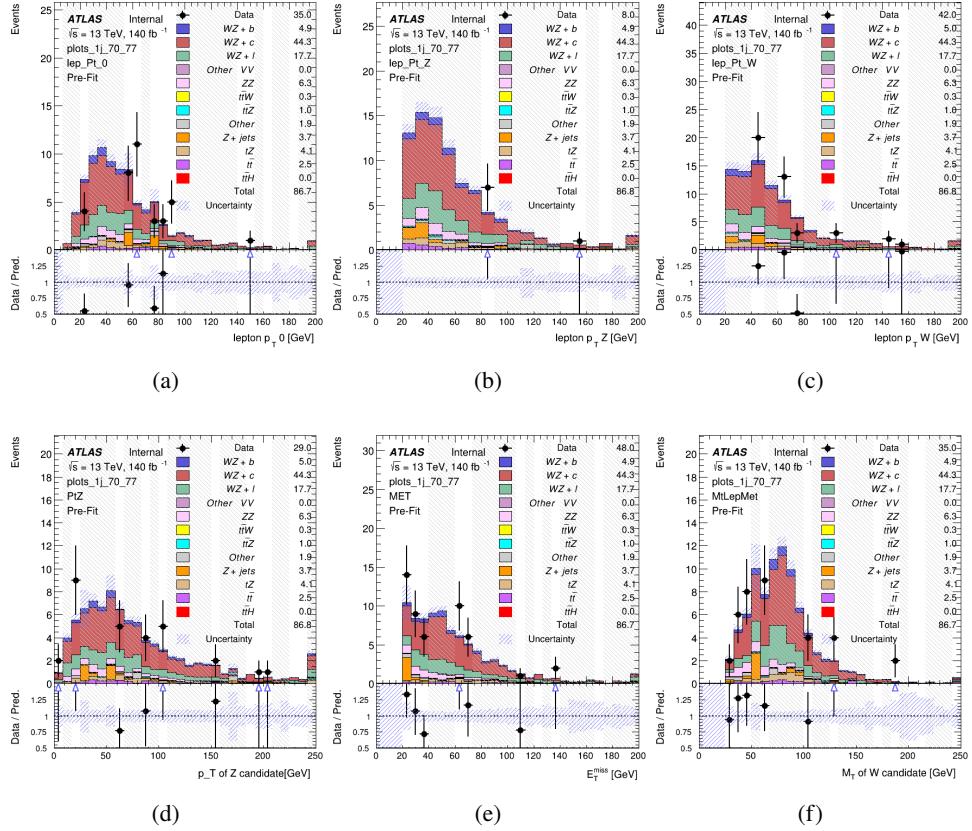


Figure 9.5: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 1j 60-70% WP

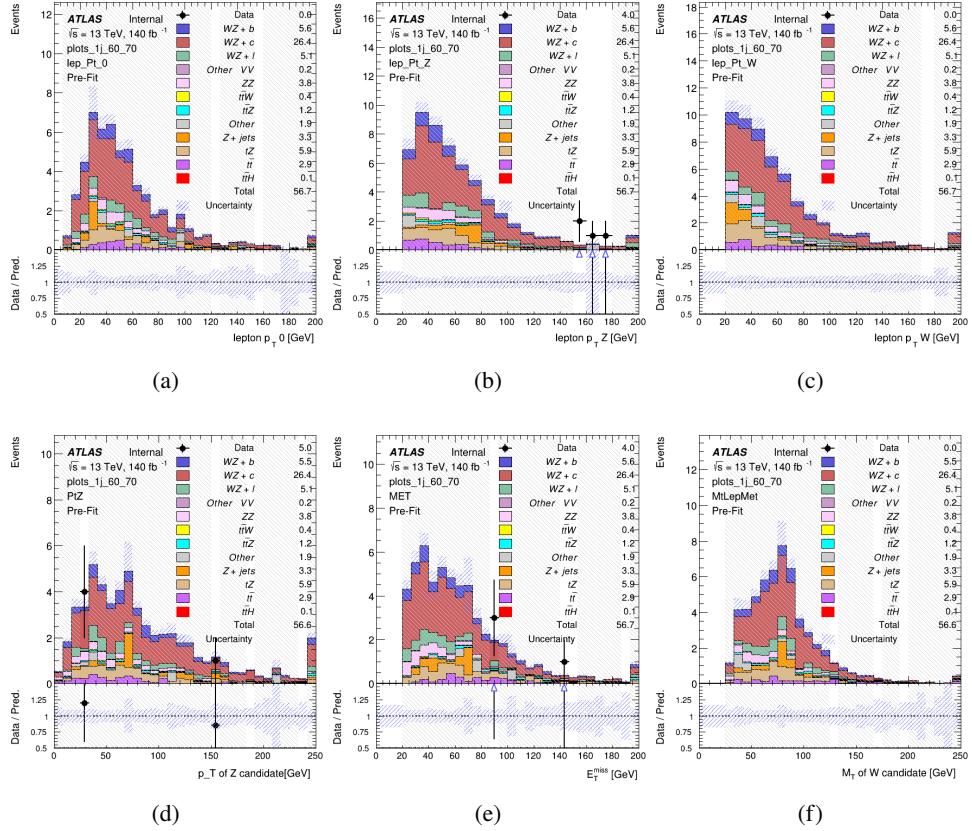


Figure 9.6: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

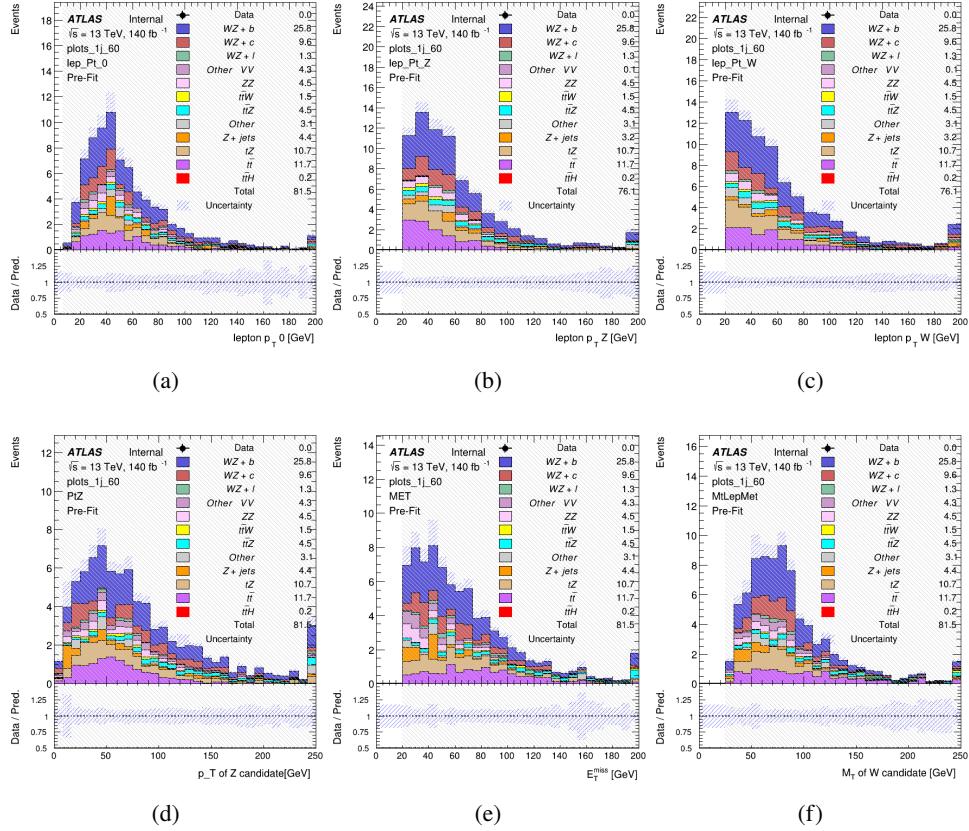
WZ Fit Region - 1j 60% WP

Figure 9.7: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - tZ-CR

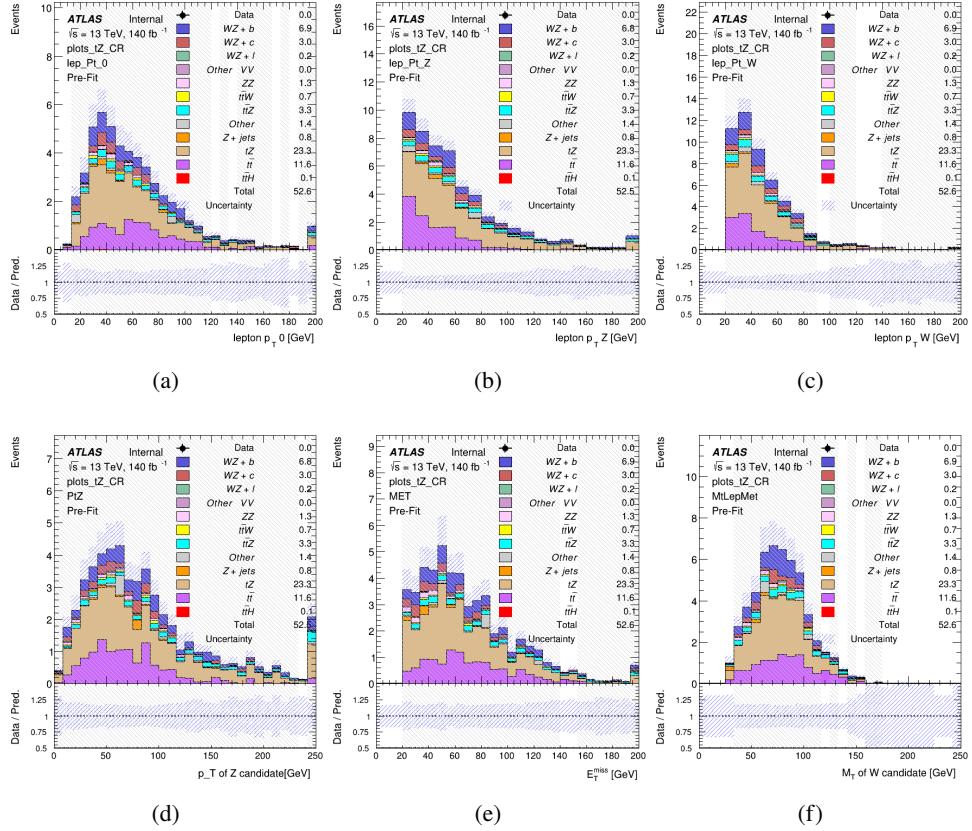


Figure 9.8: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 2j Inclusive

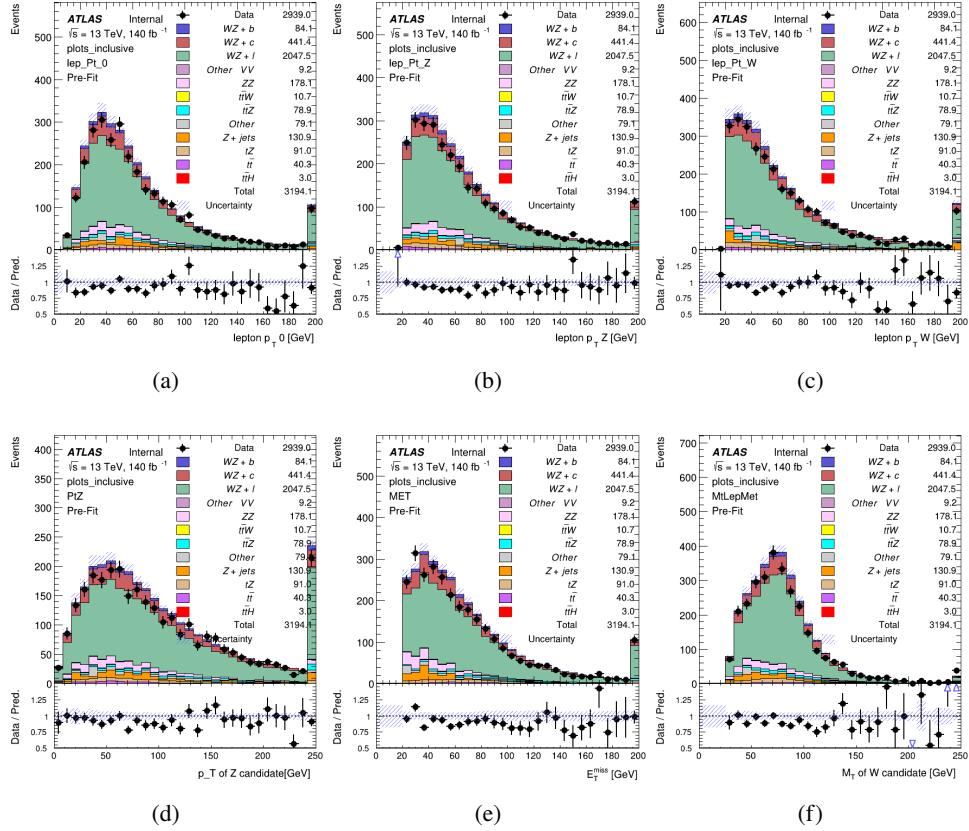


Figure 9.9: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the M_T of the W candidate.

WZ Fit Region - 2j < 85% WP

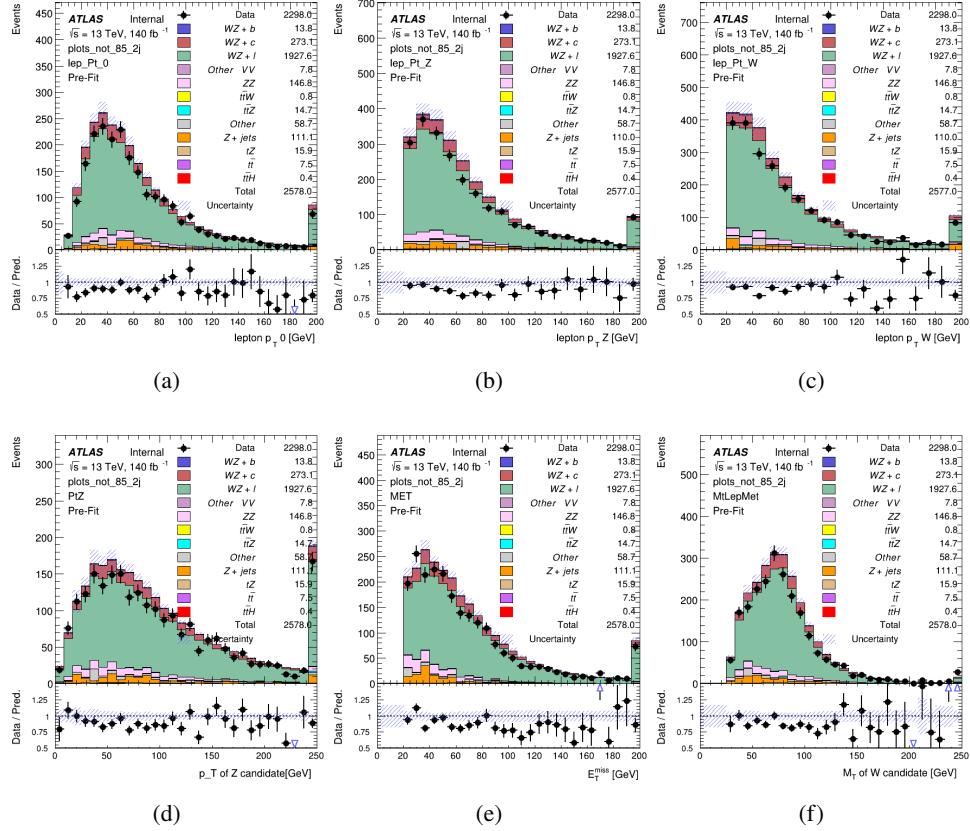


Figure 9.10: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the M_T of the W candidate.

WZ Fit Region - 2j 77-85% WP

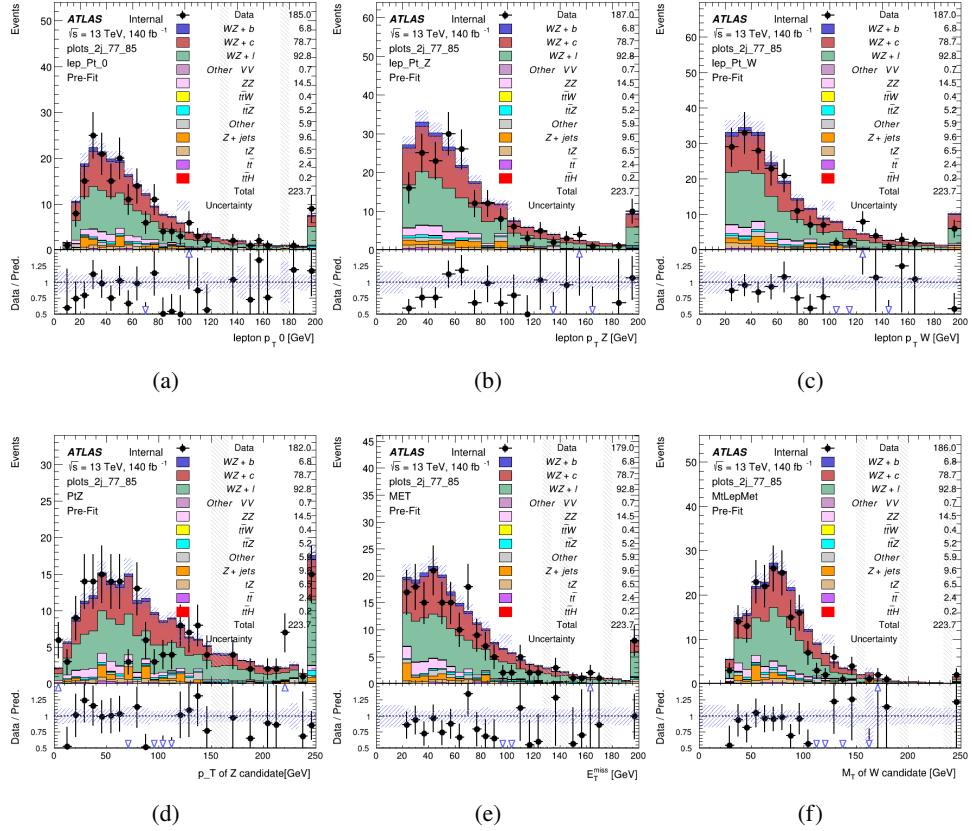


Figure 9.11: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 2j 70-77% WP

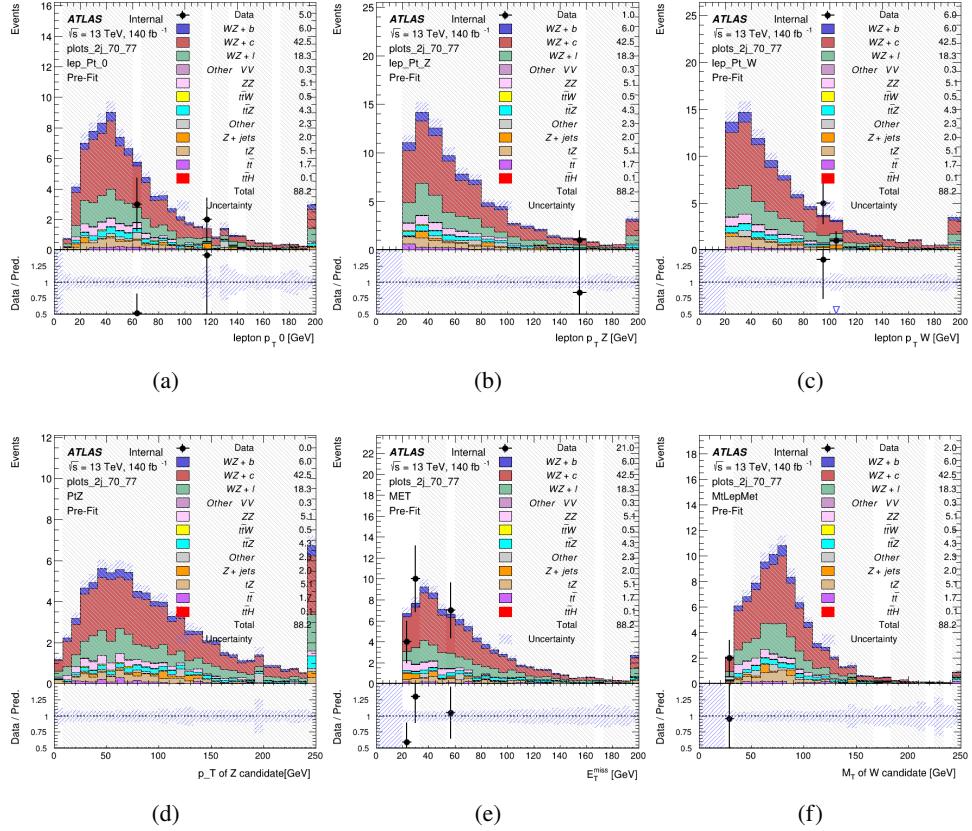


Figure 9.12: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 2j 60-70% WP

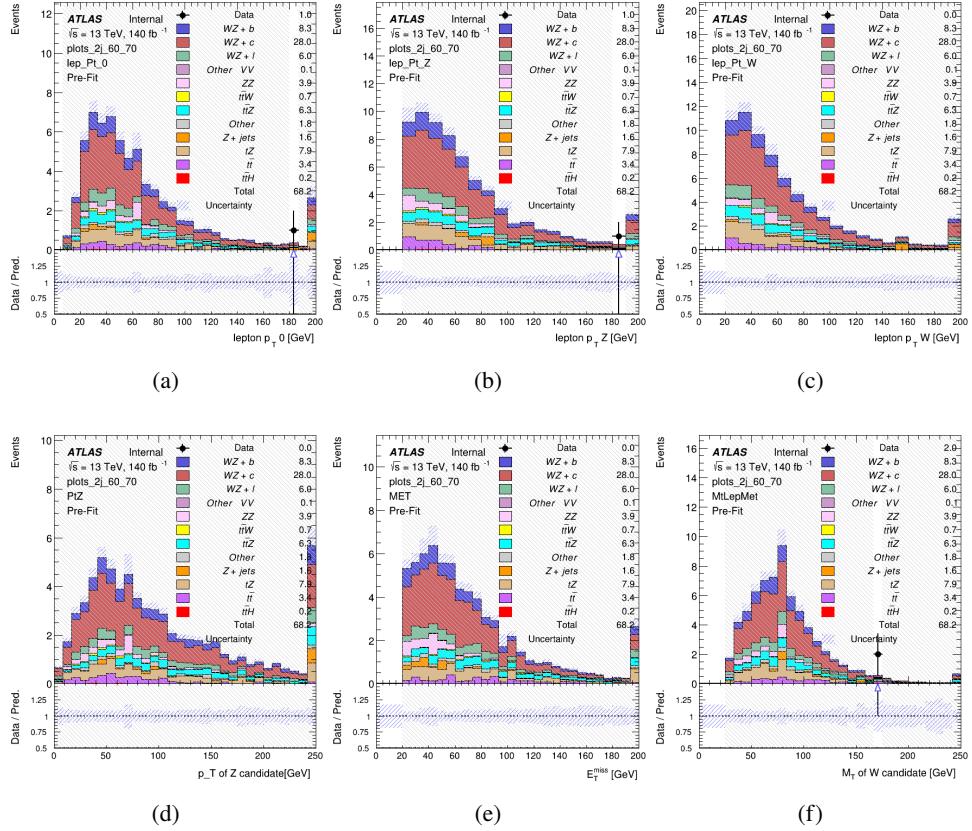


Figure 9.13: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

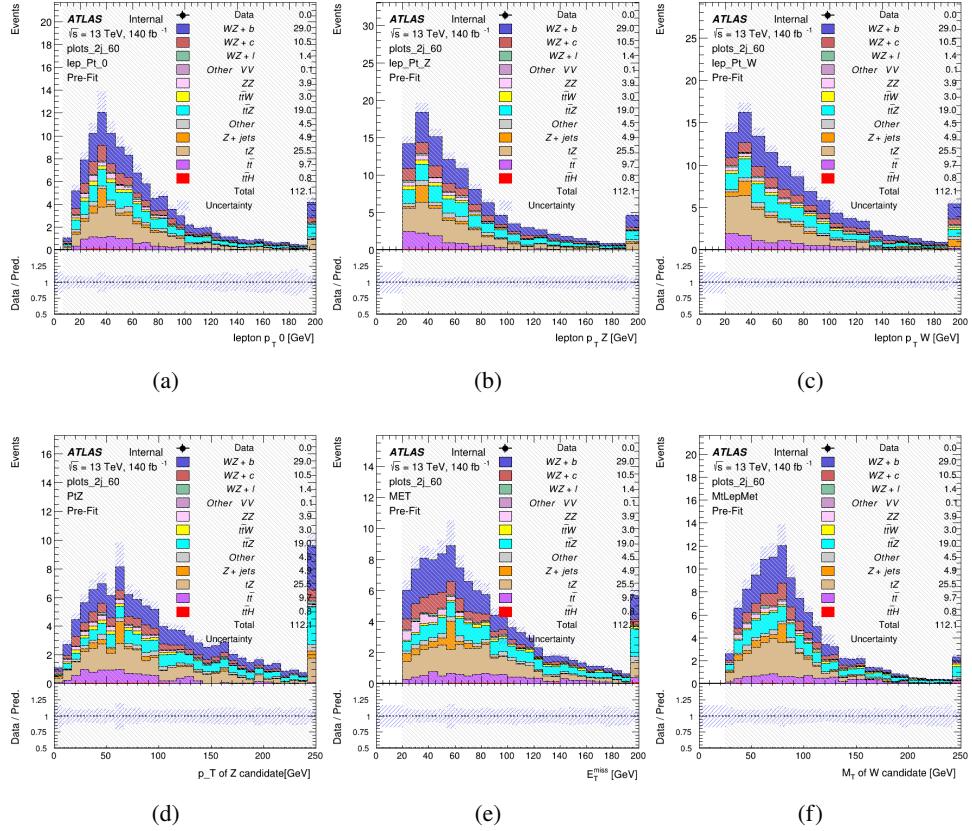
WZ Fit Region - 2j 60% WP

Figure 9.14: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - tZ-CR-2j

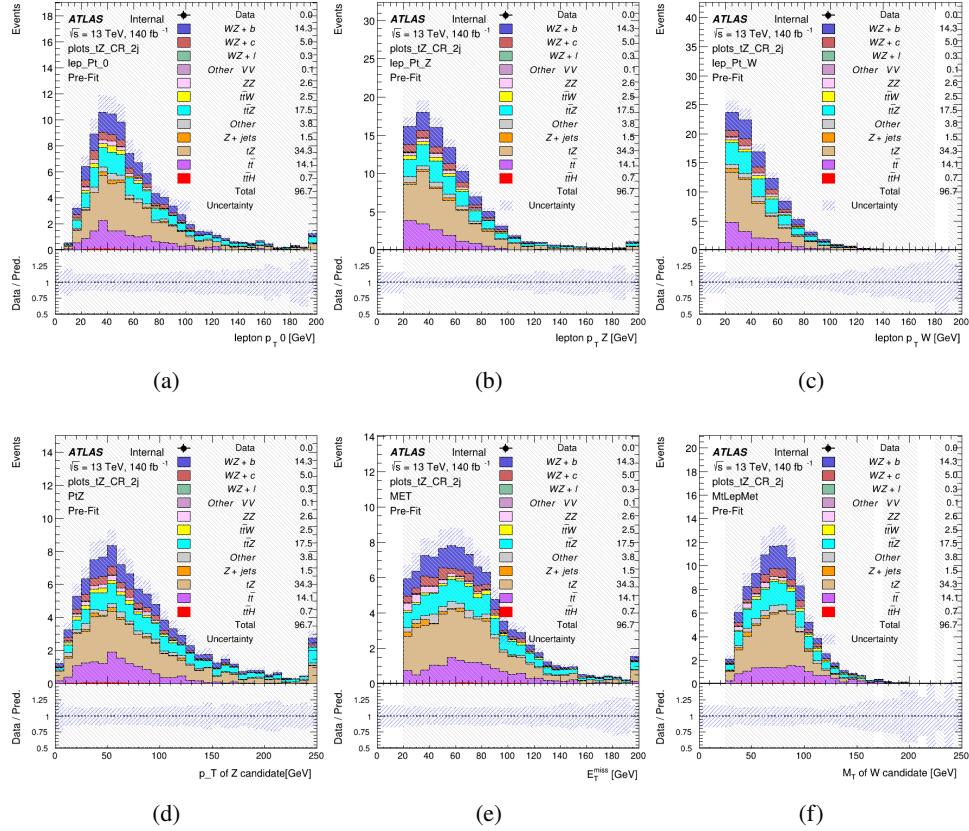


Figure 9.15: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

568 9.3 Non-Prompt Lepton Estimation

569 Two processes act as sources of non-prompt leptons appear in the analysis: $t\bar{t}$ and $Z+jet$
570 production both produce two prompt leptons, and each contribute to the 31 region when an
571 additional non-prompt lepton appears in the event. The contribution of these processes is
572 estimated with Monte Carlo simulations, which are validated using enriched validation regions.

573 **9.3.1 $t\bar{t}$ Validation**

574 $t\bar{t}$ events can produce two prompt leptons from the decay of each of the tops. These top decays
 575 produce two b-quarks, the decay of which can produce additional non-prompt leptons, which
 576 occasionally pass the event preselection. In order to validate that the Monte Carlo accurately
 577 simulates this process accurately, the MC prediction in a non-prompt $t\bar{t}$ enriched validation
 578 region is compared to data.

579 The $t\bar{t}$ validation region is similar to the preselection region - three leptons meeting the
 580 criteria described in section 13 are required, and the requirements on E_T^{miss} remain the same.
 581 However, the selection requiring a lepton pair form a Z-candidate are reversed. Events where the
 582 invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of 91.2 GeV are
 583 rejected. This ensures the $t\bar{t}$ validation region is orthogonal to the preselection region.

584 Further, because the jet multiplicity of $t\bar{t}$ events tends to be higher than WZ, the number
 585 of jets in each event is required to be greater than 1. As b-jets are almost invariably produced
 586 from top decays, at least one b-tagged jet passing the 70% DL1r WP in each event is required.
 587 Various kinematic plots of this region are shown in figure 13.16.

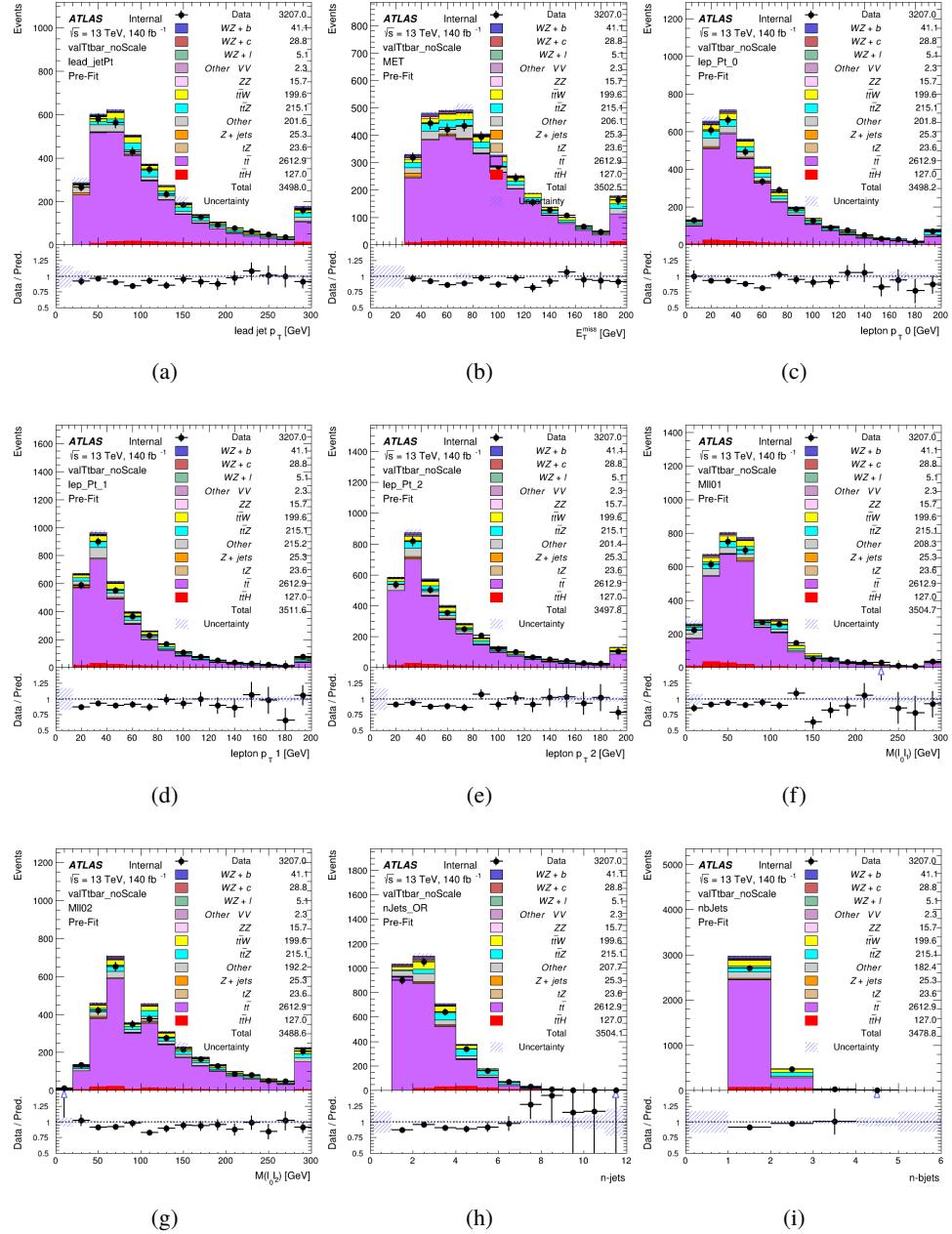


Figure 9.16: Comparisons between the data and MC distributions in the $t\bar{t}$ validation region for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) the number of b-tagged jets.

588 The shape of each distribution agrees quite well between data and MC, with a constant
589 offset between the two. This is accounted for by applying a constant correction factor of 0.883
590 to the $t\bar{t}$ MC prediction. Plots showing the kinematics of the $t\bar{t}$ VR after this correction factor
591 has been applied are shown in figure 13.17.

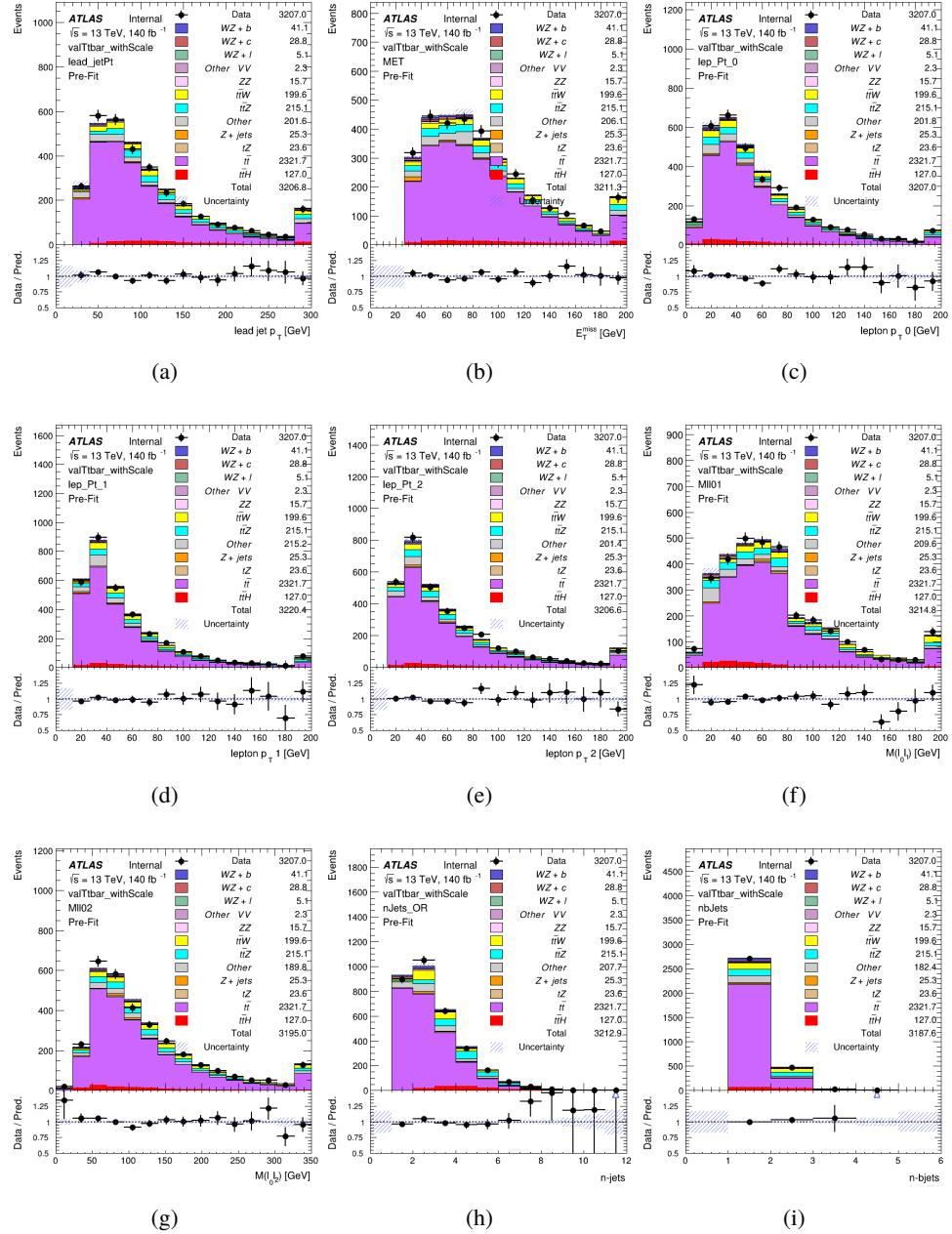


Figure 9.17: Comparisons between the data and MC distributions in the $t\bar{t}$ validation region after the correction factor has been applied for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) the number of b-tagged jets.

592 The modeling is further validated by looking at the yield in the $t\bar{t}$ VR for each DL1r WP,
 593 giving a clearer correspondence to the signal regions used in the fit. Each region shown in figure
 594 13.21 requires one or more jets pass the listed WP, with no jets passing the next highest WP.

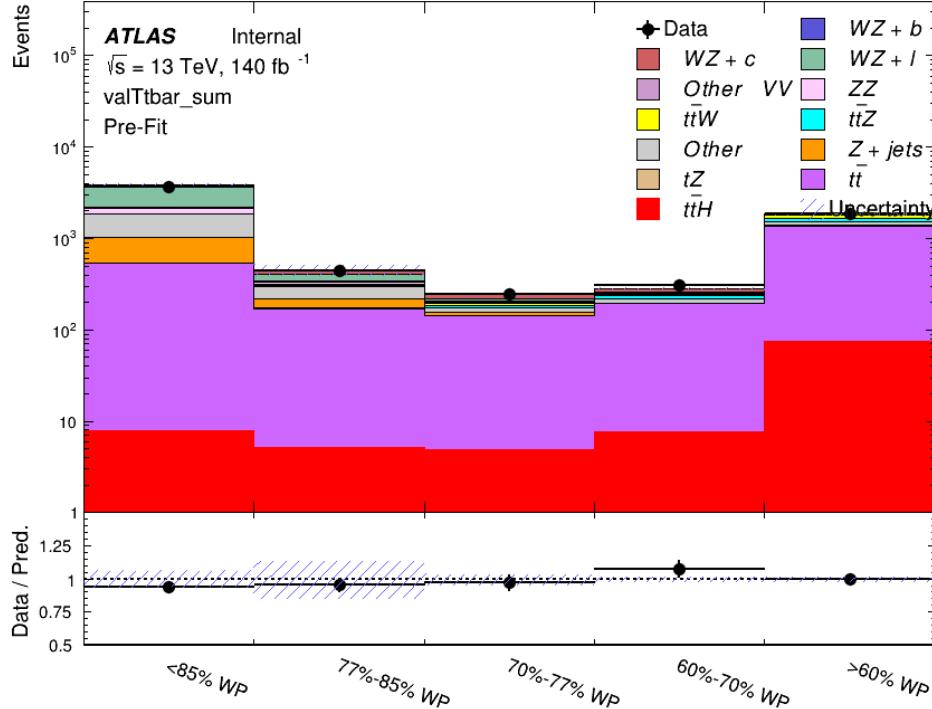


Figure 9.18: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the $t\bar{t}$ VR selection and correction factor have been applied

595 As data and MC are found to agree within 10% for each of these working points, a 10%
 596 systematic uncertainty on the $t\bar{t}$ prediction is included for the analysis.

597 9.3.2 Z+jets Validation

598 Similar to $t\bar{t}$, a non-prompt Z +jets validation region is produced in order to validate the MC
599 predictions. The lepton requirements remain the same as the preselection region. Because no
600 neutrinos are present for this process, the E_T^{miss} cut is reversed, requiring $E_T^{\text{miss}} < 30 \text{ GeV}$. This
601 also ensures this validation region is orthogonal to the preselection region. Further, the number
602 of jets in each event is required to be greater than or equal to one. Various kinematic plots of this
603 region are shown below. The general agreement between data and MC in each of these suggests
604 that the non-prompt contribution of Z +jets is well modeled by Monte Carlo.

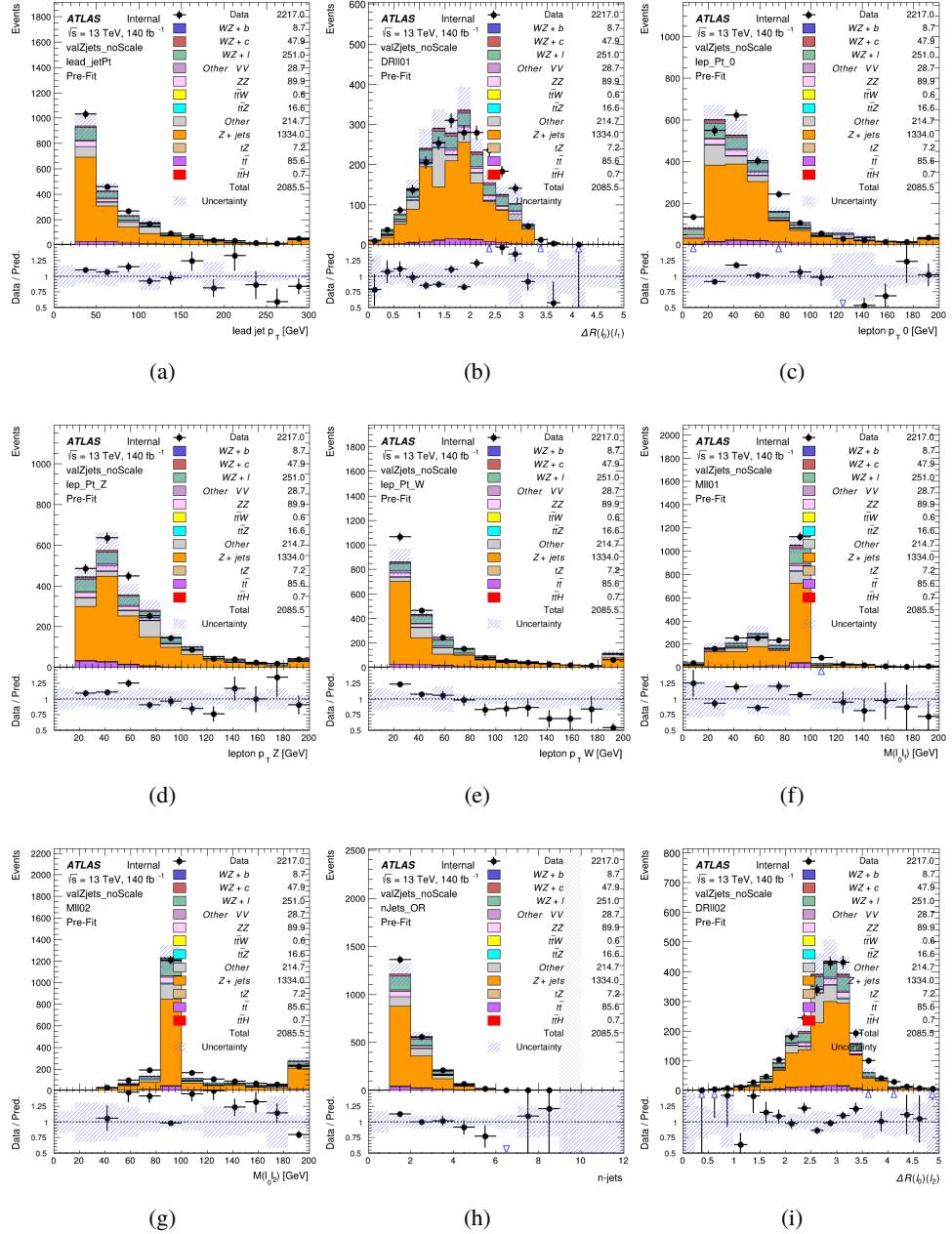


Figure 9.19: Comparisons between the data and MC distributions in the Z+jets validation region for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) ΔR between leptons 0 and 2. Includes only statistical uncertainties

605 While there is general agreement between data and MC within statistical uncertainty, the
606 shape of the p_T spectrum of the lepton from the W is found to differ. To account for this
607 discrepancy, a variable correction factor is applied to Z+jets. χ^2 minimization of the lepton 2 p_T
608 spectrum is performed to derive a correction factor of $1.53 - 6.6 * 10^{-6}(\text{lep}_\text{Pt}_W)$. Kinematic
609 plots of the Z + jets validation region after this correction factor has been aplied are shown in
610 figure 13.20.

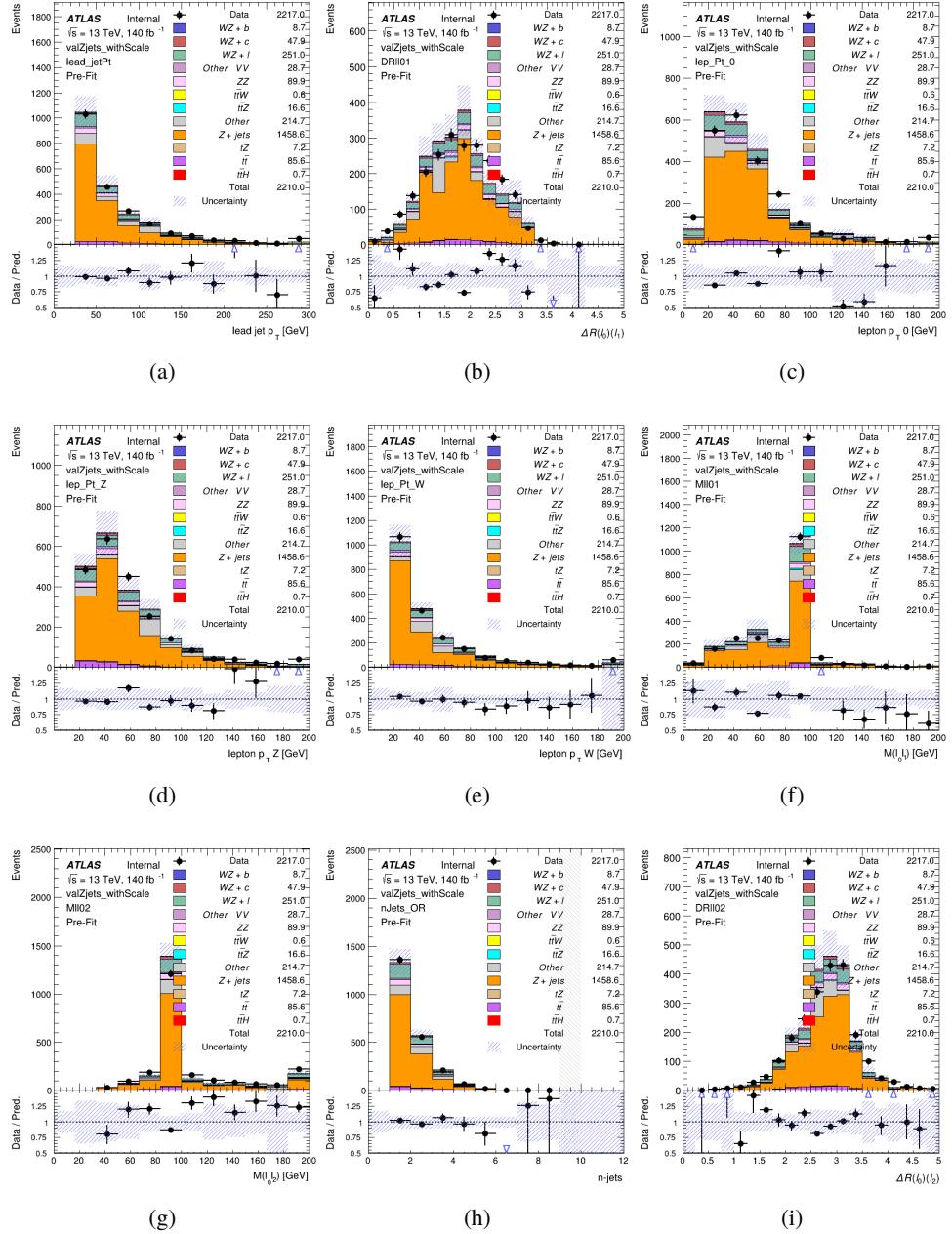


Figure 9.20: Comparisons between the data and MC distributions in the $Z + \text{jets}$ validation region after the correction factor has been applied for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) ΔR between leptons 0 and 2

611 The modeling is further validated by looking at the yield in the Z+jets VR for each DL1r
 612 WP, giving a clearer correspondence to the signal regions used in the fit. Each region shown in
 613 figure 13.21 requires one or more jets pass the listed WP, with no jets passing the next highest
 614 WP.

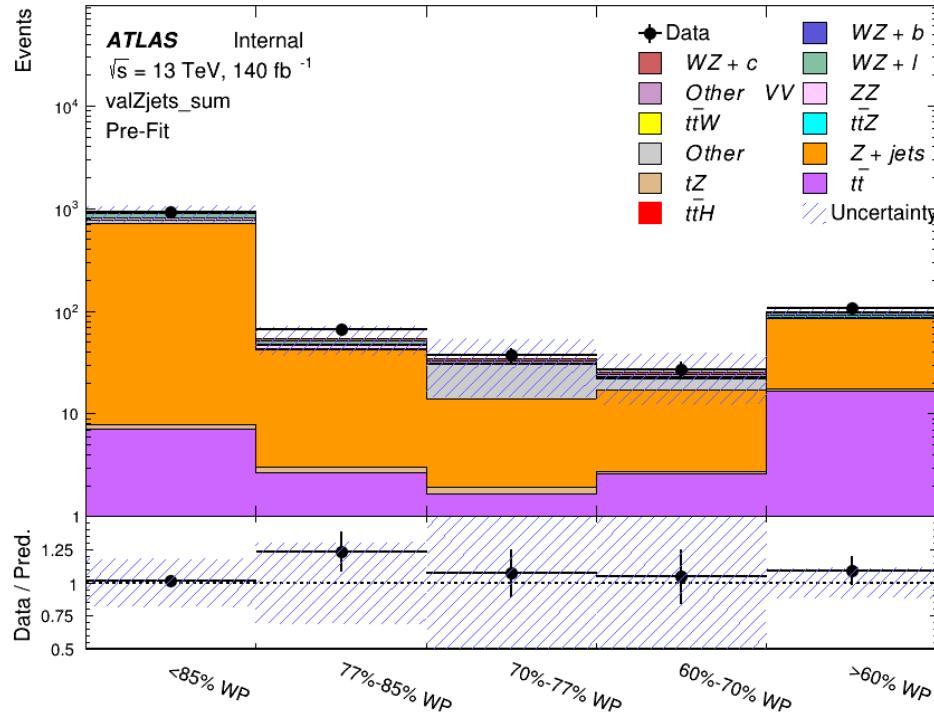


Figure 9.21: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the Z+jets VR selection and correction factor have been applied

615 For each of the b-tagging working points considered, the data falls within 20% of the MC
 616 prediction once this correction factor has been applied. Therefore, a 20% systematic uncertainty
 617 is applied to Z + jets in the analysis.

618 **10 tZ Interference Studies and Separation Multivariate Analysis**

619 Because tZ produces a final state identical to signal, it represents a predominant background in
620 the most signal enriched regions. That is, the region with one jet passing the 60% DL1r WP.
621 Therefore, a boosted decision tree (BDT) algorithm is trained using TMVA [[TMVA_guide](#)] to
622 separate WZ + heavy flavor from tZ.

623 Separation between tZ and WZ + heavy flavor is achieved in part by reconstructing the
624 invariant mass of the top candidate, which clusters more closely to the top mass for tZ than WZ
625 + heavy flavor.

626 The result of this BDT is used to create a tZ enriched region in the fit, reducing its impact
627 on the measurement of WZ + heavy flavor.

628 **10.1 Top Mass Reconstruction**

629 The reconstruction of the top mass follows the procedure described in detail in section 6.1 of
630 [[ttZ_paper](#)]. The mass of the top quark candidate is reconstructed from the jet, the lepton not
631 included in the Z-candidate, and a reconstructed neutrino. In the case that there is one jet in the
632 event, there is only possible b-jet candidate. For events with two jets, the jet with the highest
633 DL1r score is used.

634 The neutrino from the W decay is expected to be the only source of E_T^{miss} . Therefore, the
 635 E_T and ϕ of the neutrino are taken from the E_T^{miss} measurement. This leaves the z-component
 636 of the neutrino momentum, $p_{\nu z}$ as the only unknown.

637 This unknown is solved for by taking the combined invariant mass of the lepton and
 638 neutrino to give the invariant mass of the W boson:

$$639 \quad (p_l + p_\nu)^2 = m_W^2$$

640 Expanding this out into components, this equation gives:

$$641 \quad \sqrt{p_{T\nu}^2 + p_{z\nu}^2} E_l = \frac{m_W^2 - m_l^2}{2} + p_{T\nu}(p_{lx}\cos\phi_\nu + p_{ly}\sin\phi_\nu) + p_{lz}p_{\nu z}$$

642 This equation gives two solutions for $p_{\nu z}$. For cases where only one of these solutions is
 643 real, that is taken as the value of $p_{\nu z}$. For instances with two real solutions, the one which is
 644 shown to be correct in the largest fraction of simulations is taken. For cases when no real solution
 645 is found, often because of detector effects, the value of E_T^{miss} is varied in decreasing increments
 646 of 100 MeV until a real solution is found.

647 The reconstructed top mass distribution for tZ and WZ + b can be seen in figure 14.1.

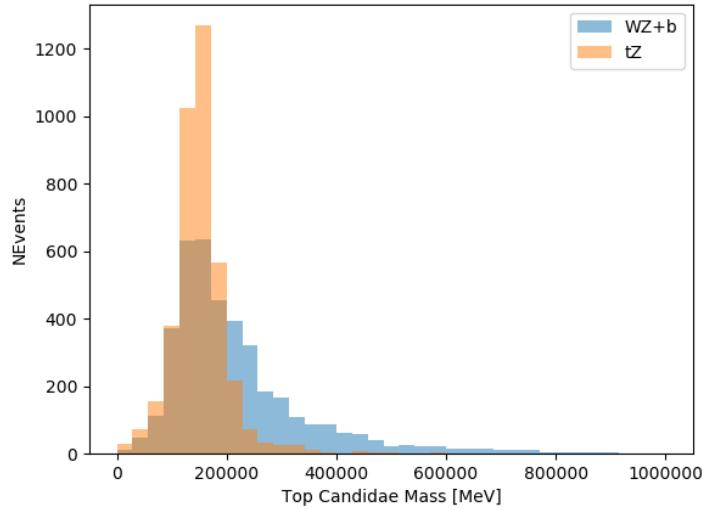


Figure 10.1: Reconstructed top mass distributions for tZ and WZ + b, measured in MeV.

648 10.2 tZ BDT

649 A Boosted Decision Tree (BDT), specifically XGBoost [xgboost_cite], is used to provide separ-
 650 ation between tZ and WZ+b. The following kinematic variables are used as inputs:

651 • The invariant mass of the reconstructed top candidate

652 • p_T of each of the leptons, jet

653 • The invariant mass of each combination of lepton pairs, $M(l\bar{l})$

654 • E_T^{miss}

655 • Distance between each combination of leptons, $\Delta R(l\bar{l})$

- 656 • Distance between each lepton and the jet, $\Delta R(lj)$

657 The training samples included only events meeting the requirements of the 1-jet, $>60\%$
658 region, i.e. passing all the selection described in section 13 and having exactly one jet which
659 passes the tightest (60%) DL1r working point.

660 The distributions of a few of these features for both signal and background is shown in
661 figure 14.2.

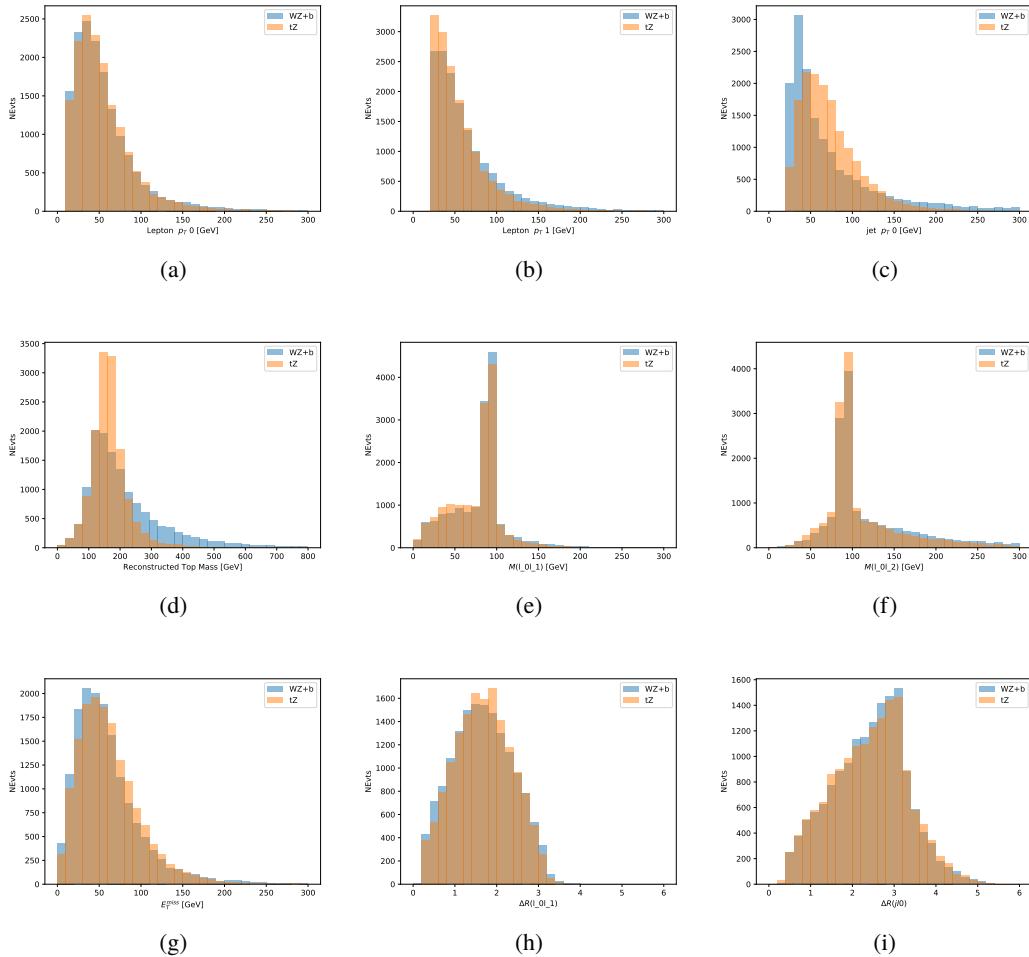


Figure 10.2: Distribution of input features of the BDT for signal (WZ) and background (tZ). Both are scaled to an equal number of events. (a), (b) and (c) show the p_T of lepton 0, lepton 1, and the jet, (d) show the reconstructed top mass, (e) and (f) show the invariant mass of leptons 0 and 1, leptons 0 and 2. (g) shows the E_T^{miss} of each event. (h) and (i) show the ΔR between lepton 0 and lepton 1, and the jet.

662 A sample of 20,000 background (tZ) and signal (WZ+b) Monte Carlo events are used to
 663 train the BDT. And additional 5,000 events are reserved for testing the model, in order to prevent
 664 over-fitting. A total of 750 decision trees with a maximum depth of 6 branches are used to build
 665 the model. These parameters are chosen empirically, by training several models with different

666 parameters and selecting the one that gave the best separation for the test sample.

667 The results of the BDT training are shown in figure 14.3. The output scores for both signal
 668 and background events is shown on the left. The right shows the receiving operating characteristic
 669 (ROC) curve that results from the MVA. The ROC curve represents the background rejection
 670 as a function of signal efficiency, where each point on the curve represents a different response
 671 score. The ROC curve of the BDT is compared to the performance of using an optimal set of flat
 672 selections on the same set of input variables.

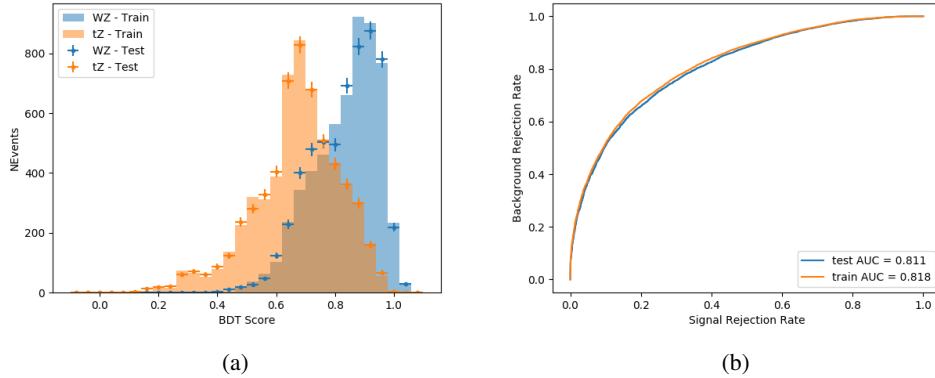


Figure 10.3: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

673 The relative important of each input feature in the model, measured by how often they
 674 appeared in the decision trees, is shown in figure 14.4.

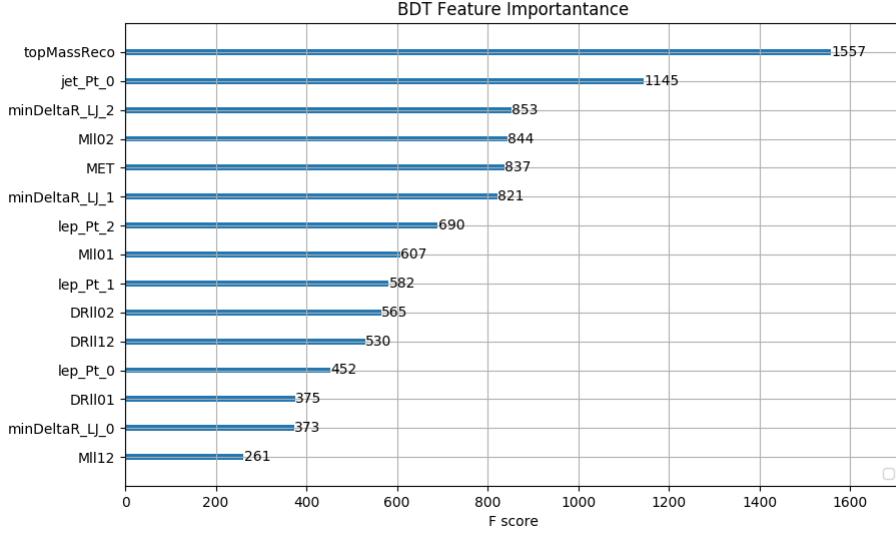


Figure 10.4: Relative importance of each input feature in the model.

675 These results suggest that some amount of separation can be achieved between these two
 676 processes, with a high BDT score selecting a set of events that is pure in $WZ + b$. A BDT score
 677 of 0.725 is selected as a cutoff, where events with scores higher than this form a signal enriched
 678 region, and events with scores lower than this form a $t\bar{Z}$ control region. This cutoff is selected by
 679 varying the value of this cutoff in stat-only Asimov fits, and selecting the value that minimizes
 680 the statistical uncertainty on $WZ + b$.

681 11 Data and Monte Carlo Samples

682 Both data and Monte Carlo samples used in this analysis were prepared in the xAOD format,
 683 which was used to produce a DxAOD sample in the HIGG8D1 derivation framework. The HIGG8D1

684 framework is designed for the $t\bar{t}H$ multi-lepton analysis, which targets events with multiple
 685 leptons as well as tau hadrons. This framework skims the dataset to remove unneeded variables
 686 as well as entire events. Events are removed from the derivations that do not meet the following
 687 selection:

- 688 • at least two light leptons within a range $|\eta| < 2.6$, with leading lepton $p_T > 15$ GeV and
 689 subleading lepton $p_T > 5$ GeV
- 690 • at least one light lepton with $p_T > 15$ GeV within a range $|\eta| < 2.6$, and at least two hadronic
 691 taus with $p_T > 15$ GeV.

692 Samples were then generated from these HIGG8D1 derivations with p-tag of p4134 using
 693 AnalysisBase version 21.2.127 modified to include custom variables..

694 11.1 Data Samples

695 The study uses a sample of proton-proton collision data collected by the ATLAS detector from
 696 2015 through 2018 at an energy of $\sqrt{s} = 13$ TeV, which represents an integrated luminosity of
 697 139 fb^{-1} . This data set was collected with a bunch crossing rate of 25 ns. All data used in this
 698 analysis was verified by data quality checks, having been included in the following Good Run
 699 Lists:

- 700 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
 701 _PHYS_StandardGRL_All_Good_25ns.xml

- 702 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
703 _PHYS_StandardGRL_All_Good_25ns.xml

704 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_UncKnown_PHYS_StandardGRL
705 _All_Good_25ns_Triggerno17e33prim.xml

706 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_UncKnown_PHYS_StandardGRL
707 _All_Good_25ns_Triggerno17e33prim.xml

708 Runs included from the AllYear period containers are included.

709 **11.2 Monte Carlo Samples**

710 Several different generators were used to produce Monte Carlo simulations of the signal and
711 background processes. For all samples, the response of the ATLAS detector is simulated using
712 Geant4. The WZ signal samples are simulated using Sherpa 2.2.2 [[sherpa](#)]. Specific information
713 about the Monte Carlo samples being used can be found in table 26. A list of the specific samples
714 used by data set ID is shown in table 9.

Table 8: 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, VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
tZ	MG5_AMC	LO	PYTHIA 6	CTEQ6L1
t̄W	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
t̄t(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄tH	MG5_AMC (MG5_AMC)	NLO (NLO)	PYTHIA 8 (HERWIG++)	NNPDF 3.0 NLO [Ball:2014uwa] (CT10 [ct10])
tHqb	MG5_AMC	LO	PYTHIA 8	CT10
tHW	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	HERWIG++ (SHERPA)	CT10 (NNPDF 3.0 NLO)
tWZ	MG5_AMC	NLO	PYTHIA 8	NNPDF 2.3 LO
t̄t, t̄t̄t̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄t	Powheg-BOX v2 [powheggtt]	NLO	PYTHIA 8	NNPDF 3.0 NLO
t̄tγ	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
s-, t-channel, Wt single top	Powheg-BOX v1 [powhegstp]	NLO	PYTHIA 6	CT10
qqVV, VVV Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

Sample	DSID
WZ	364253, 364739-42
VV	364250, 364254, 364255, 363355-60, 364890
t̄W	410155
t̄Z	410156, 410157, 410218-20
low mass t̄Z	410276-8
Rare Top	410397, 410398, 410399
single Top	410658-9, 410644-5
three Top	304014
four Top	410080
t̄WW	410081
Z + jets	364100-41
low mass Z + jets	364198-215
W + jets	364156-97
Vγ	364500-35
tZ	410560
tW	410013-4
WtZ	410408
VVV	364242-9
VH	342284-5
WtH	341998
t̄tγ	410389
t̄t	410470
t̄tH	345873-5, 346343-5

Table 9: List of Monte Carlo samples by data set ID used in the analysis.

715 12 Object Reconstruction

- 716 All regions defined in this analysis share a common lepton, jet, and overall event preselection.
- 717 The selection applied to each physics object is detailed here; the event preselection, and the
- 718 selection used to define the various fit regions, is described in section 13.

⁷¹⁹ **12.1 Trigger**

⁷²⁰ Events are required to be selected by dilepton triggers, as summarized in table 27.

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

Table 10: List of lowest p_T -threshold, un-prescaled dilepton triggers used for 2015-2018 data taking.

⁷²¹ **12.2 Light leptons**

⁷²² Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that

⁷²³ are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].

⁷²⁴ Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Muon candidates

⁷²⁵ are reconstructed by combining inner detector tracks with track segments or full tracks in the

⁷²⁶ muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have $p_T > 10$ GeV and

727 $|\eta| < 2.5$. Candidates in the transition region between different electromagnetic calorimeter
 728 components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant com-
 729 bining shower shape and track information is used to distinguish real electrons from hadronic
 730 showers (fake electrons). To further reduce the non-prompt electron contribution, the track is
 731 required to be consistent with originating from the primary vertex; requirements are imposed
 732 on the transverse impact parameter significance ($|d_0|/\sigma_{d_0} < 5$) and the longitudinal impact
 733 parameter ($|\Delta z_0 \sin \theta_\ell| < 0.5 \text{ mm}$).

734 Muon candidates are reconstructed by combining inner detector tracks with track segments
 735 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
 736 $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$. The longitudinal impact parameter is the same for both electrons
 737 and muons, while muons are required to pass a slightly tighter transverse impact parameter,
 738 $|d_0|/\sigma_{d_0} < 3$.

739 All leptons are required to be isolated, as defined through the standard `isolationFixedCutLoose`
 740 working point supported by combined performance groups. Leptons are additionally required to
 741 pass a non-prompt BDT selection described in detail in [**ttH_paper**]. Optimized working points
 742 and scale factors for this BDT are taken from that analysis.

743 12.3 Jets

744 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
 745 calorimeters [**ATL-PHYS-PUB-2015-015**], using the anti- k_t algorithm with a radius parameter

746 $R = 0.4$. Jets with energy contributions likely arising from noise or detector effects are removed
 747 from consideration [ATLAS-CONF-2015-029], and only jets satisfying $p_T > 25$ GeV and
 748 $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
 749 association algorithm is used to confirm that the jet originates from the selected primary vertex,
 750 in order to reject jets arising from pileup collisions [PERF-2014-03].

751 12.4 B-tagged Jets

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

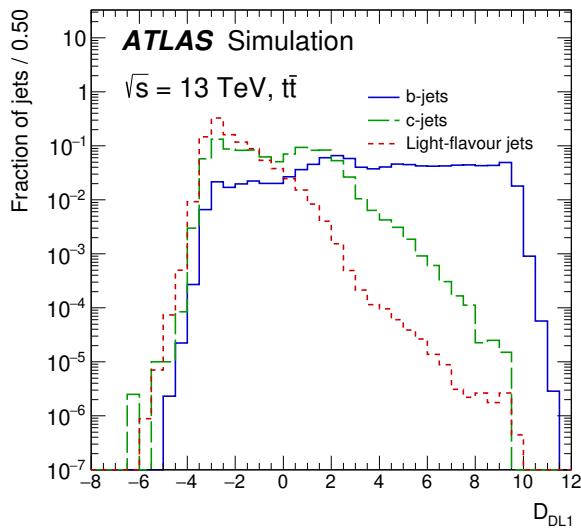


Figure 12.1: Output distribution of the DL1r algorithm for b-jets, charm jets, and light jets

757 From the output of the BDT, working points (WPs) are developed based on the efficiency
 758 of truth b-jets at particular values of the DL1r algorithm. The working points used in this analysis
 759 are summarized in table 11.

WP	none	loose	medium	tight	tightest
b eff.	-	85%	77%	70%	60%

Table 11: B-tagging Working Points by tightness and b-jet efficiency

760 A tighter WP will accept fewer b-jets, but reject a higher fraction of charm and light jets.
 761 Generally, analyses that include b-jets will use a fixed working point, for example, requiring that
 762 a jet pass the 70% threshold. By instead treating these working point as bins, e.g. events with
 763 jets that fall between the 85% and 77% WPs fall into one bin, while events with jets passing the
 764 60% WP fall into another, and looking at the full psuedo-continuous DL1r spectrum of the jets,
 765 additional information can be gained. The psuedo-continuous b-tag spectrum is used in this case
 766 to separate out WZ + b, WZ + charm, and WZ + light.

767 12.5 Missing transverse energy

768 Missing transverse momentum (E_T^{miss}) is used as part of the event selection. The missing
 769 transverse momentum vector is defined as the inverse of the sum of the transverse momenta of
 770 all reconstructed physics objects as well as remaining unclustered energy, the latter of which is
 771 estimated from low- p_T tracks associated with the primary vertex but not assigned to a hard object,
 772 with object definitions taken from [ATL-PHYS-PUB-2015-027]. Light leptons considered in

773 the E_T^{miss} reconstruction are required to have $p_T > 10 \text{ GeV}$, while jets are required to have $p_T > 20$
774 GeV .

775 **13 Event Selection and Signal Region Definitions**

776 Event are required to pass a preselection described in section 13.1 and summarized in table 12.
777 Those that pass this preselection are divided into various fit regions described in section 13.2,
778 based on the number of jets in the event, and the b-tag score of those jets.

779 **13.1 Event Preselection**

780 Events are required to include exactly three reconstructed light leptons passing the requirement
781 described in 12.2, which have a total charge of ± 1 . As the opposite sign lepton is found to be
782 prompt the vast majority of the time [ttH_paper], it is required to be loose and isolated, as
783 defined though the standard `isolationFixedCutLoose` working point supported by combined
784 performance groups. The same sign leptons are required to be very tightly isolated, as per the
785 recommended `isolationFixedCutTight`.

786 The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose
787 charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.
788 the smallest ΔR , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton

⁷⁸⁹ 0 is required to have $p_T > 10$ GeV, while the same sign leptons, 1 and 2, are required to have
⁷⁹⁰ $p_T > 20$ GeV to reduce the contribution of non-prompt leptons.

⁷⁹¹ The invariant mass of at least one pair of opposite sign, same flavor leptons is required
⁷⁹² to fall within 10 GeV of the mass of the Z boson, 91.2 GeV. Events where one of the opposite
⁷⁹³ sign pairs have an invariant mass less than 12 GeV are rejected in order to suppress low mass
⁷⁹⁴ resonances.

⁷⁹⁵ An additional requirement is placed on the missing transverse energy, $E_T^{\text{miss}} > 20$ GeV,
⁷⁹⁶ and the transverse mass of the W candidate, $m(E_T^{\text{miss}} + l_{\text{other}}) > 30$ GeV, where E_T^{miss} is the
⁷⁹⁷ missing transverse energy, and l_{other} is the lepton not included in the Z-candidate.

⁷⁹⁸ Events are required to have one or two reconstructed jets passing the selection described
⁷⁹⁹ in section 12.3. Events with more than two jets are rejected in order to reduce the contribution
⁸⁰⁰ of backgrounds such as $t\bar{t}Z$ and $t\bar{t}W$, which tend to have higher jet multiplicity.

Event Selection

- Exactly three leptons with charge ± 1
 - Two tight Iso, tight ID same-charge leptons with $p_T > 20$ GeV
 - One loose Iso, medium ID opposite charge lepton with $p_T > 10$ GeV
 - $m(l^+l^-)$ within 10 GeV of 91.2 GeV
 - Transverse mass of W-candidate, $m_T(E_T^{\text{miss}} + \text{lep}_{\text{other}}) > 30$ GeV
 - Missing transverse energy, $E_T^{\text{miss}} > 20$ GeV
 - One or two jets with $p_T > 25$ GeV
-

Table 12: Summary of the selection applied to events for inclusion in the fit

⁸⁰¹ The event yields in the preselection region for both data and Monte Carlo are summarized
⁸⁰² in table 13.1, which shows good agreement between data and Monte Carlo, and demonstrates that

803 this region consists primarily of WZ events. The WZ events are split into WZ + b, WZ + c, and
 804 WZ + l based on the truth flavor of the associated jet in the event. Specifically, this determination
 805 is made based on the HadronConeExclTruthLabelID of the jet, as recommended by the b-
 806 tagging working group [BtagWG]. In this ordering b-jet supersedes charm, which supersedes
 807 light. That is, WZ + l events contain no charm and no b jets at truth level, WZ + c contain at
 808 least one truth charm and no b-jets, and WZ + b contains at least one truth b-jet.

Process	Events
WZ + b	167.6 ± 6.5
WZ + c	1080 ± 40
WZ + l	7220 ± 310
Other VV	850 ± 140
$t\bar{t}W$	16.8 ± 2.3
$t\bar{t}Z$	115 ± 17
rare Top	2.2 ± 0.1
Single top	0.10 ± 0.45
Three top	0.01 ± 0.01
Four top	0.02 ± 0.01
$t\bar{t}WW$	0.23 ± 0.05
Z + jets	600 ± 260
$V + \gamma$	37 ± 54
tZ	190 ± 70
tW	5.5 ± 1.2
WtZ	25.8 ± 1.1
VVV	26.2 ± 0.9
VH	94 ± 7
$t\bar{t}$	108.68 ± 8
$t\bar{t}H$	4.3 ± 0.5
Total	10600 ± 530
Data	10574

Table 13: Event yields in the preselection region at 139.0 fb^{-1}

809 Here Other VV represents diboson processes other than WZ, and consists predominantly
 810 of $ZZ \rightarrow ll\bar{l}\bar{l}$ events where one of the leptons is not reconstructed.

811 Simulations are further validated by comparing the kinematic distributions of the Monte
 812 Carlo with data, which are shown in figure 13.1. Here, bins with 5% or more WZ+b are blinded.

WZ Fit Region - Inclusive

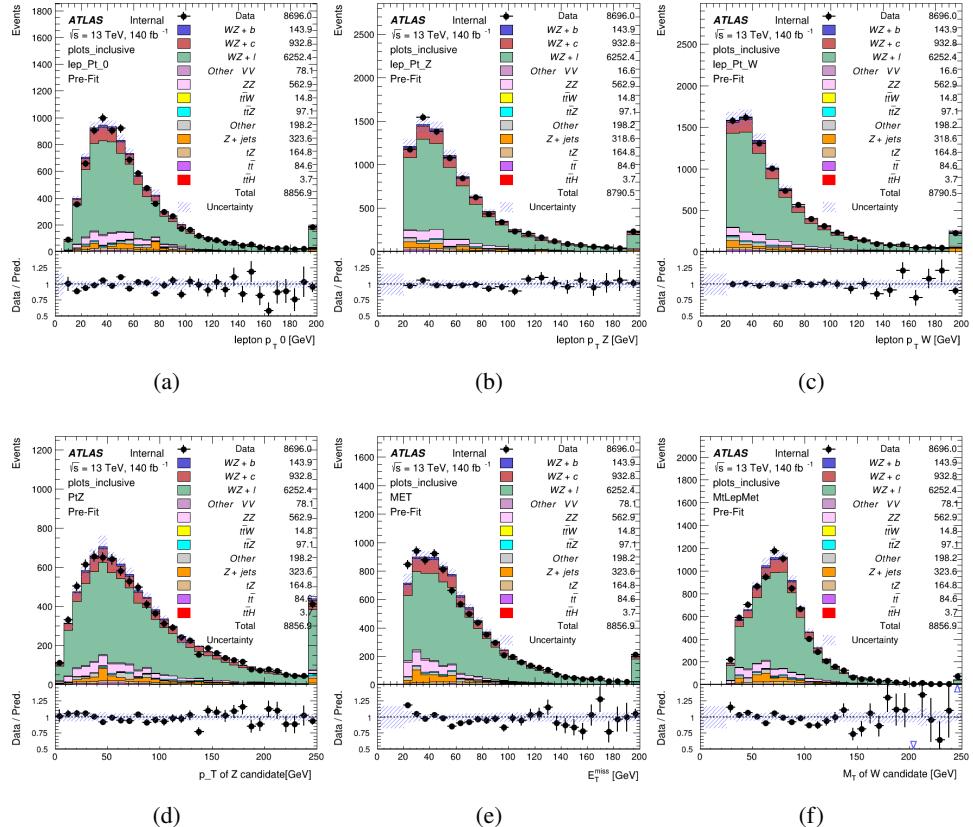


Figure 13.1: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

813 13.2 Fit Regions

814 Once preselection has been applied, the remaining events are categorized into one of twelve
 815 orthogonal regions. The regions used in the fit are summarized in table 14.

Table 14: A list of the regions used in the fit and the selection used for each.

Region	Selection
1j, <85%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_85} = 0$
1j, 85%-77%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_85} = 1, n\text{Jets_DL1r_77}=0$
1j, 77%-70%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_77} = 1, n\text{Jets_DL1r_70}=0$
1j, 70%-60%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_70} = 1, n\text{Jets_DL1r_60}=0$
1j, >60%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_60} = 1, tZ \text{ BDT} > 0.725$
1j tZ CR	$N_{\text{jets}} = 1, n\text{Jets_DL1r_60} = 1, tZ \text{ BDT} < 0.725$
2j, <85%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} = 0$
2j, 85%-77%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} >= 1, n\text{Jets_DL1r_77}=0$
2j, 77%-70%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_77} >= 1, n\text{Jets_DL1r_70}=0$
2j, 70%-60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_70} >= 1, n\text{Jets_DL1r_60}=0$
2j, >60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} >= 1, tZ \text{ BDT} > 0.725$
2j tZ CR	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} >= 1, tZ \text{ BDT} < 0.725$

816 The working points discussed in section 12.4 are used to separate events into fit regions
 817 based on the highest working point reached by a jet in each event. Because the background
 818 composition differs significantly based on the number of b-jets, events are further subdivided
 819 into 1 jet and 2 jet regions in order to minimize the impact of background uncertainties.

820 An additional tZ control region is created based on the BDT described in section 14. The
 821 region with 1-jet passing the 60% working point is split in two - a signal enriched region of
 822 events with a BDT score greater than 0.03, and a tZ control region including events with less
 823 than 0.03. This cutoff is arrived at by performing an Asimov fit with a variety of cutoffs, and
 824 selecting the value that produces the highest significance for the measurement of WZ + b.

825 The modeling in each region is validated by comparing data and MC predictions for
 826 various kinematic distributions. These plot are shown in figures 13.2-13.15.

WZ Fit Region - 1j Inclusive

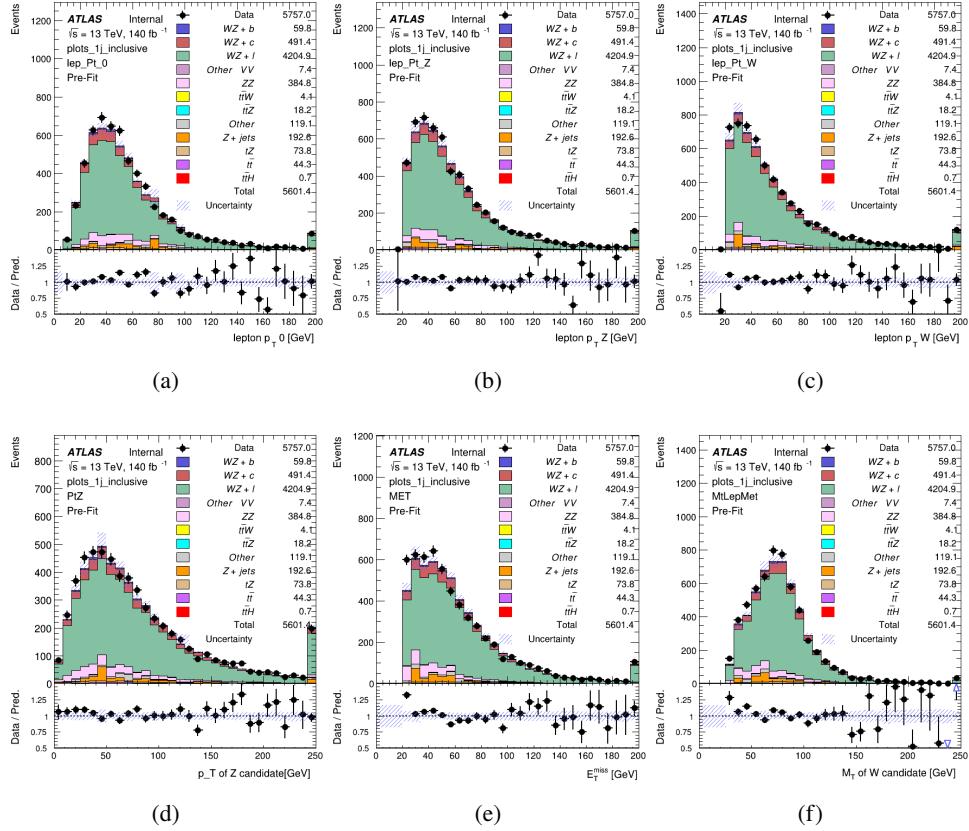


Figure 13.2: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 1j < 85% WP

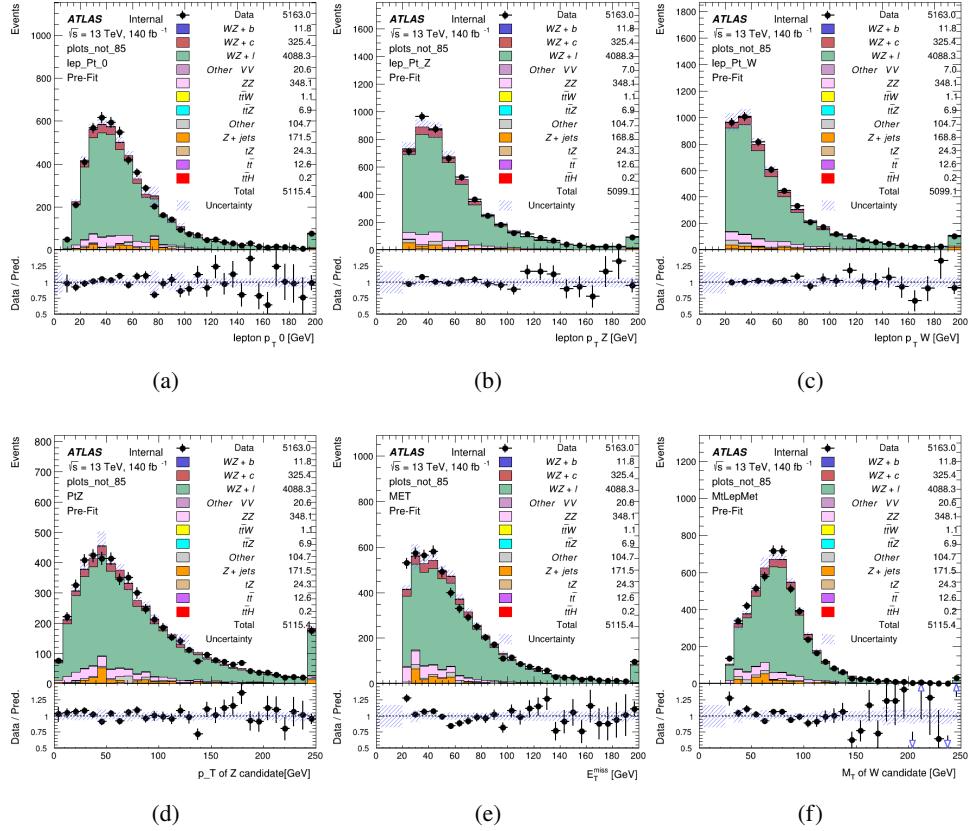


Figure 13.3: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 1j 77-85% WP

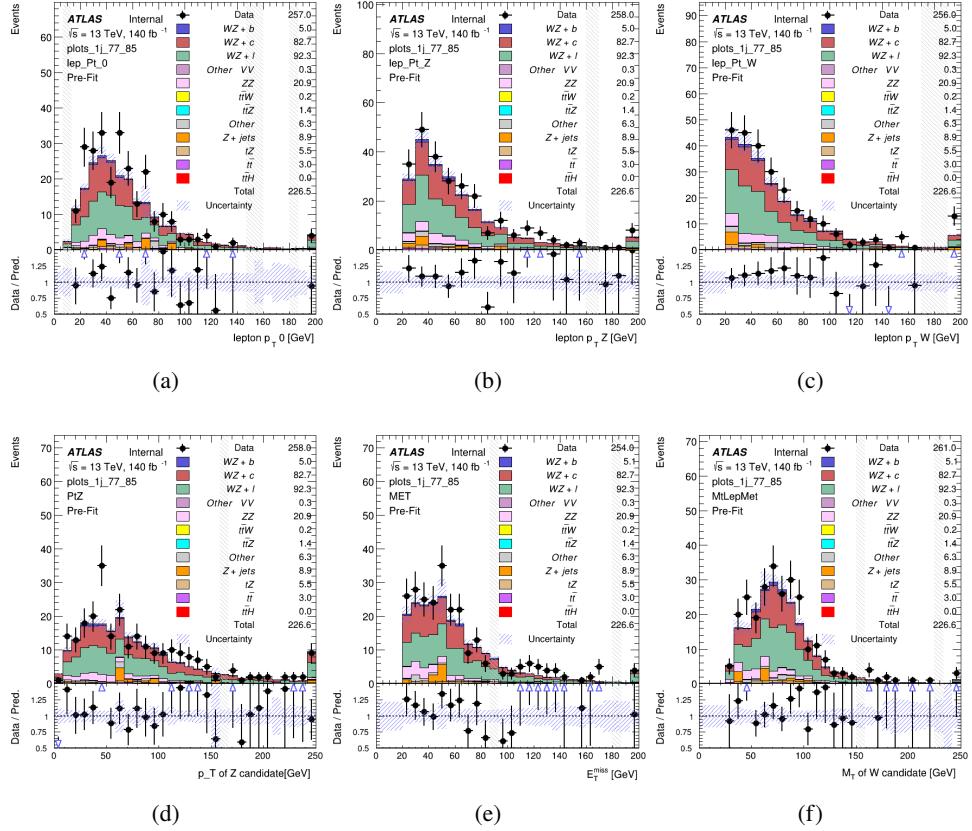


Figure 13.4: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 1j 70-77% WP

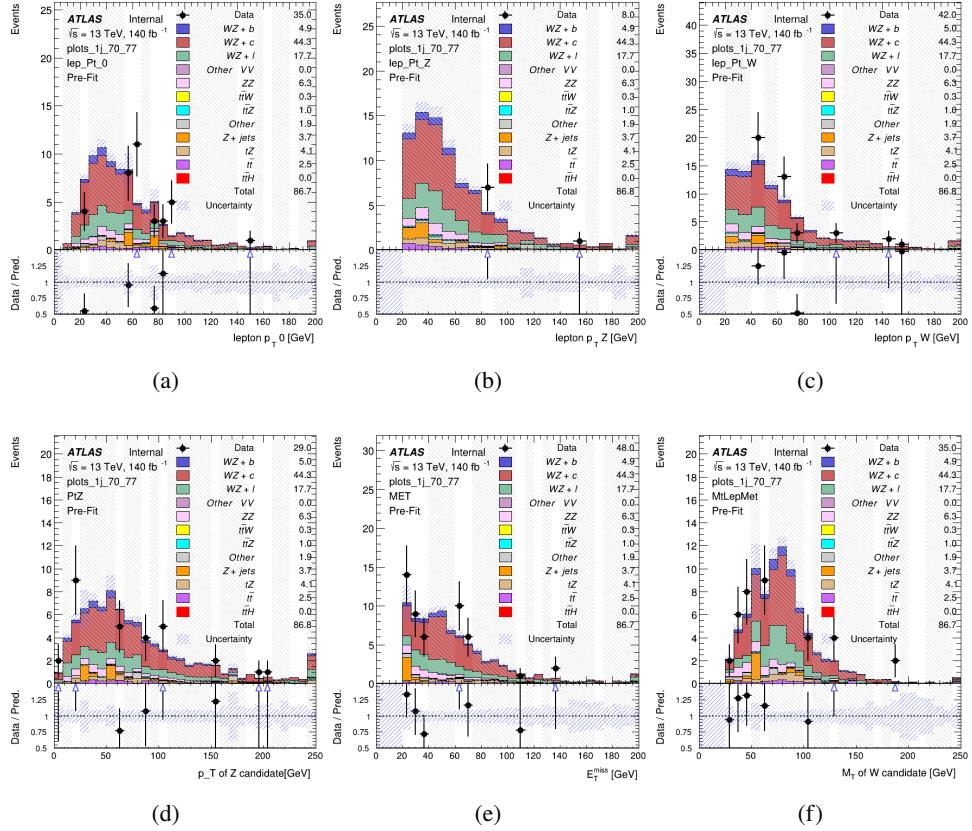


Figure 13.5: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 1j 60-70% WP

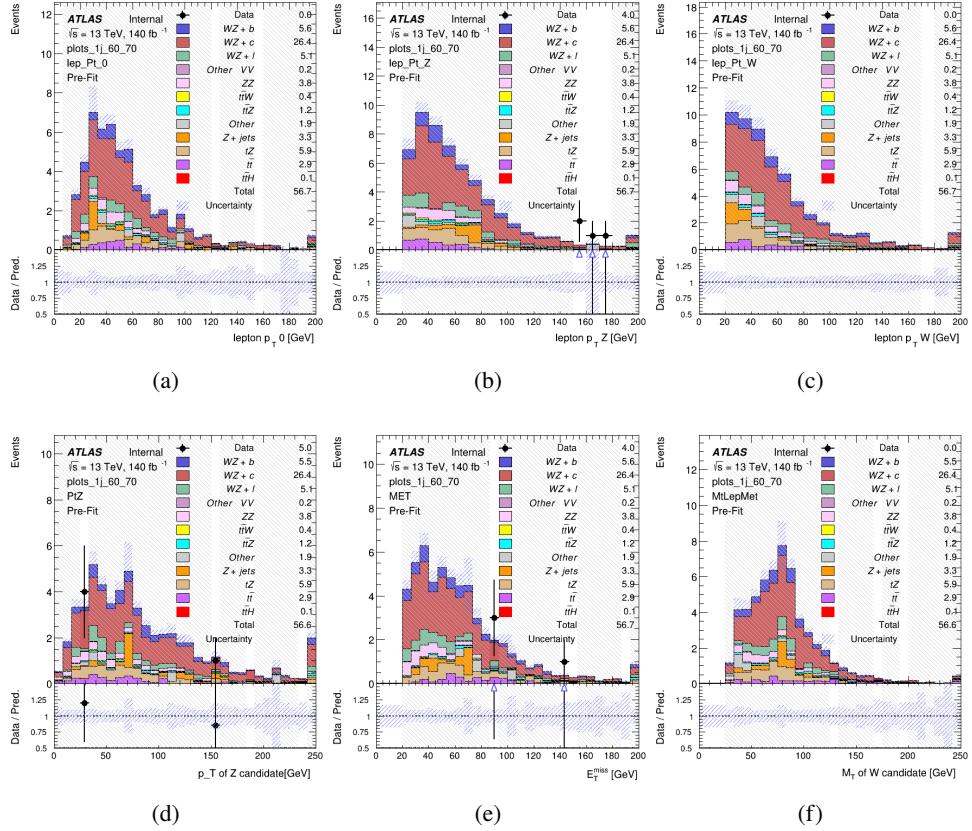


Figure 13.6: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 1j 60% WP

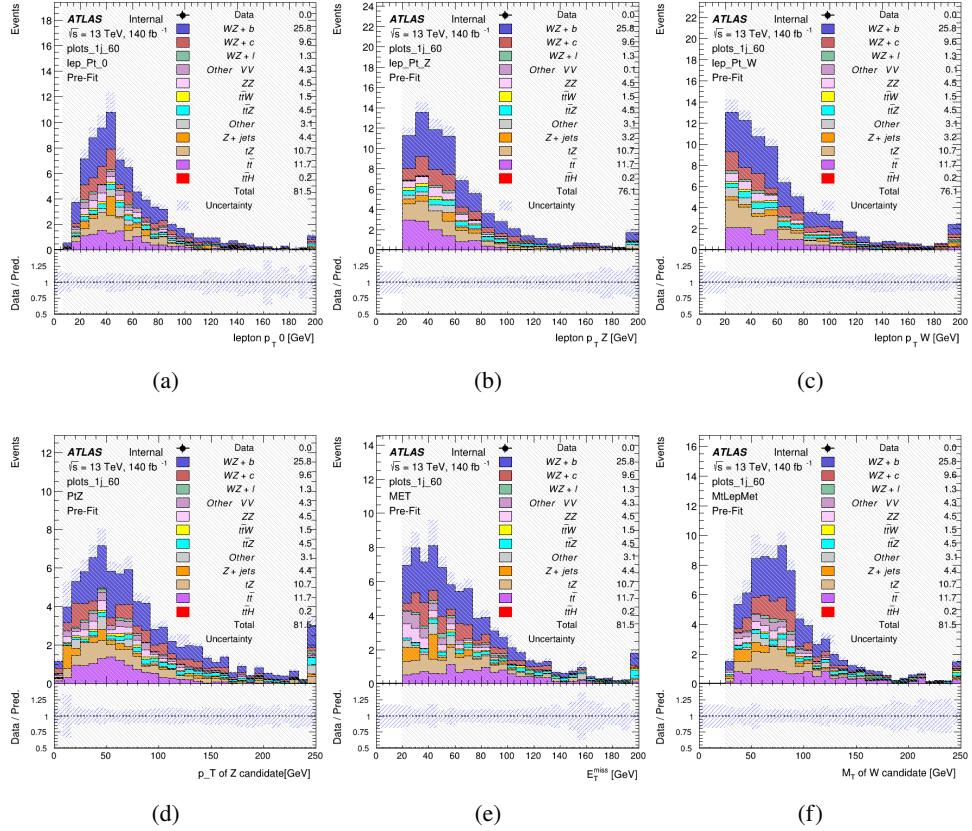


Figure 13.7: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - tZ-CR

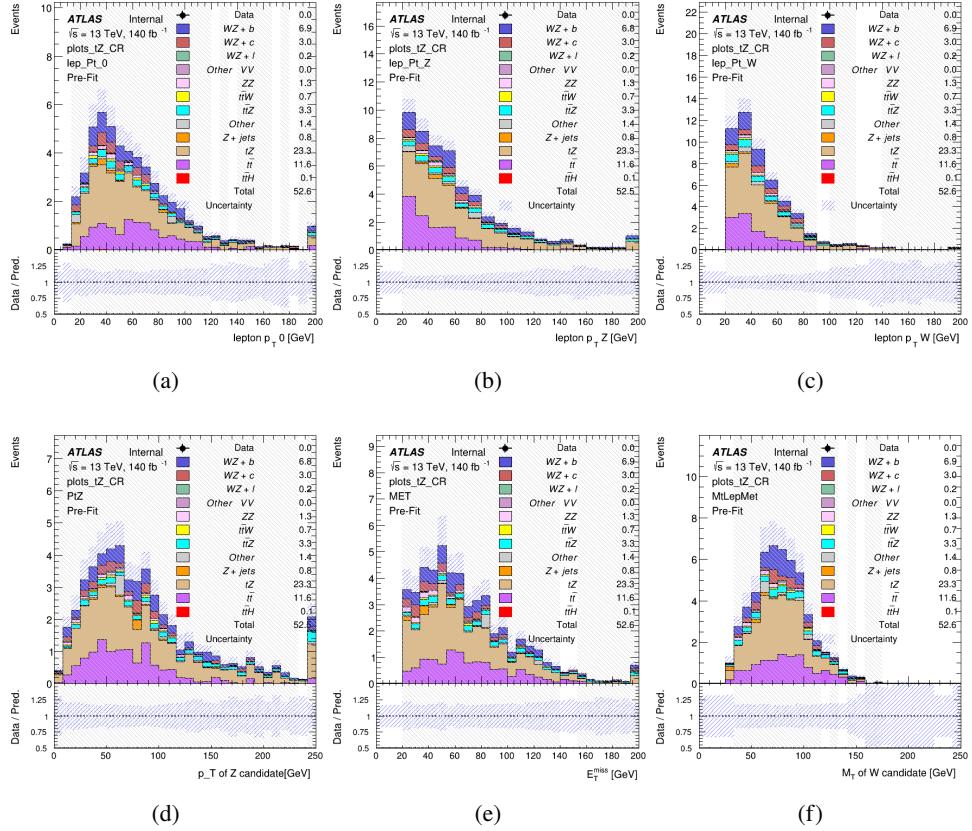


Figure 13.8: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 2j Inclusive

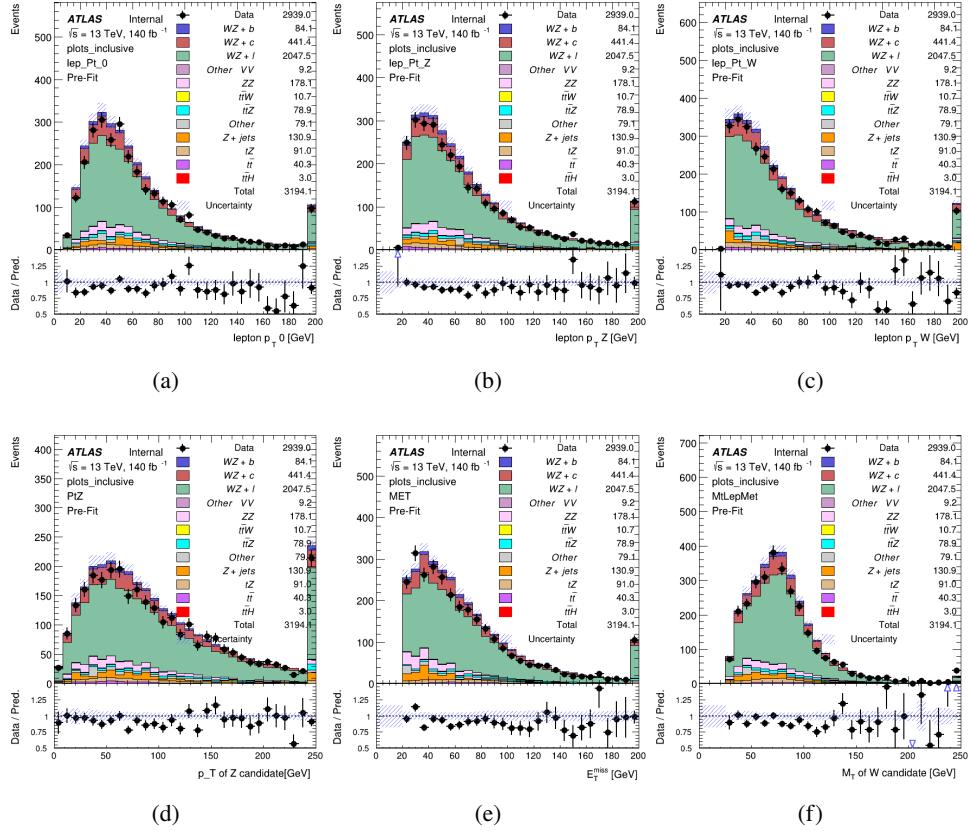


Figure 13.9: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the M_T of the W candidate.

WZ Fit Region - 2j < 85% WP

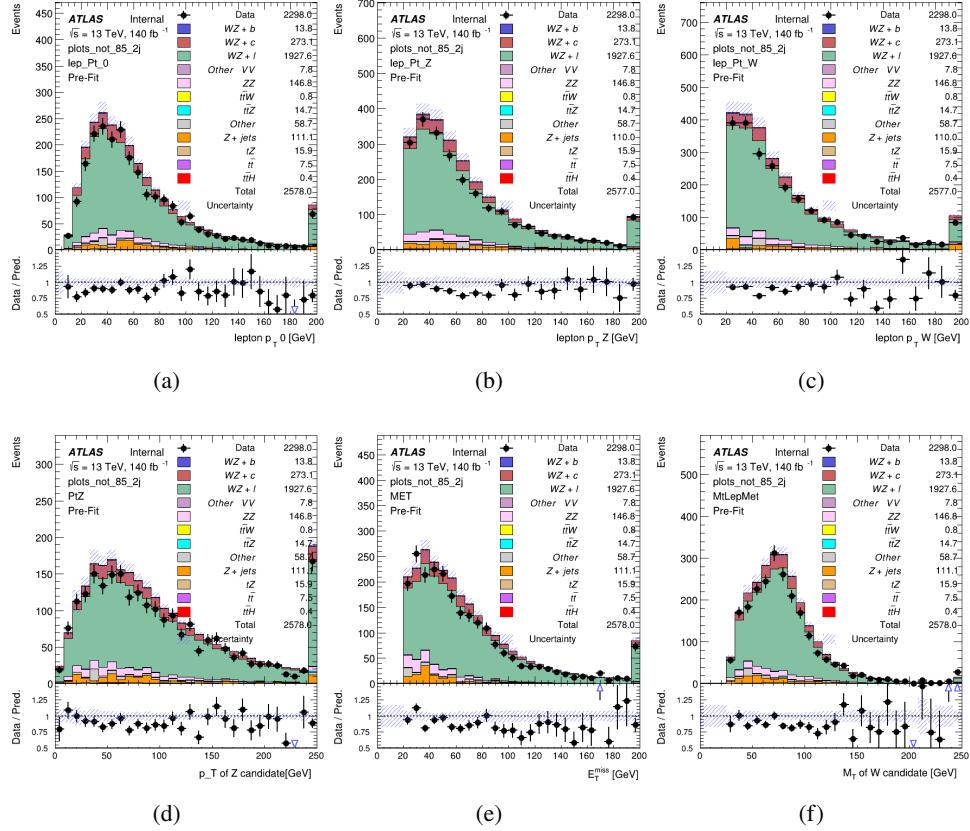


Figure 13.10: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss}, and (f) the m_T of the W candidate.

WZ Fit Region - 2j 77-85% WP

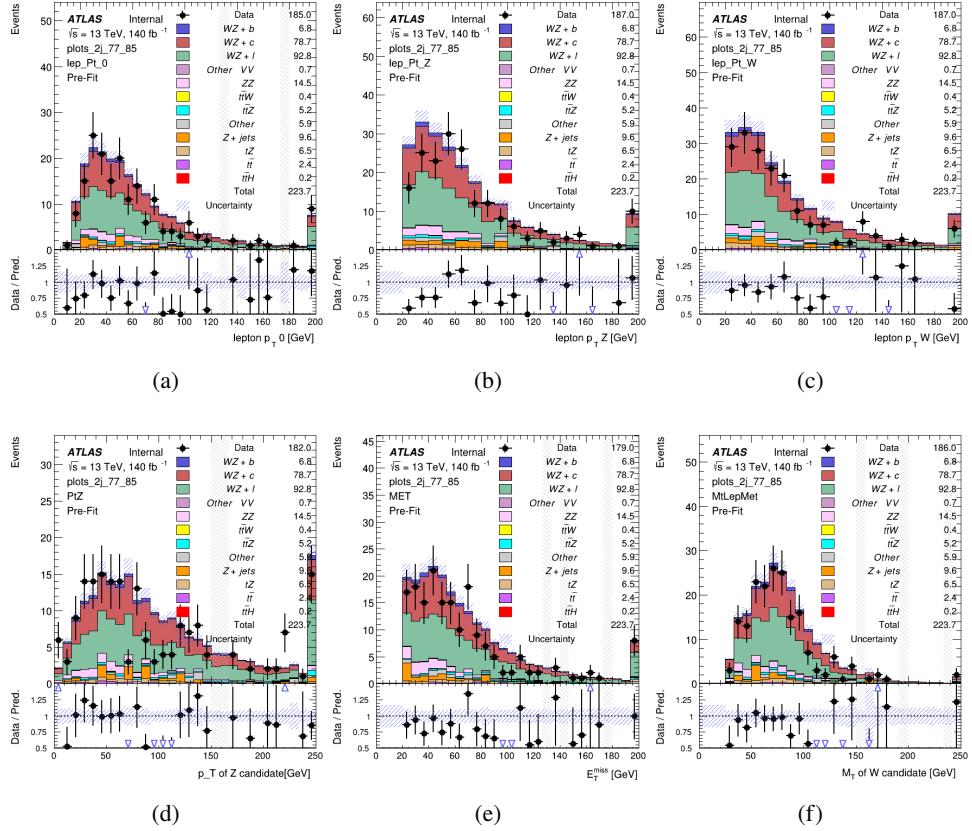


Figure 13.11: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 2j 70-77% WP

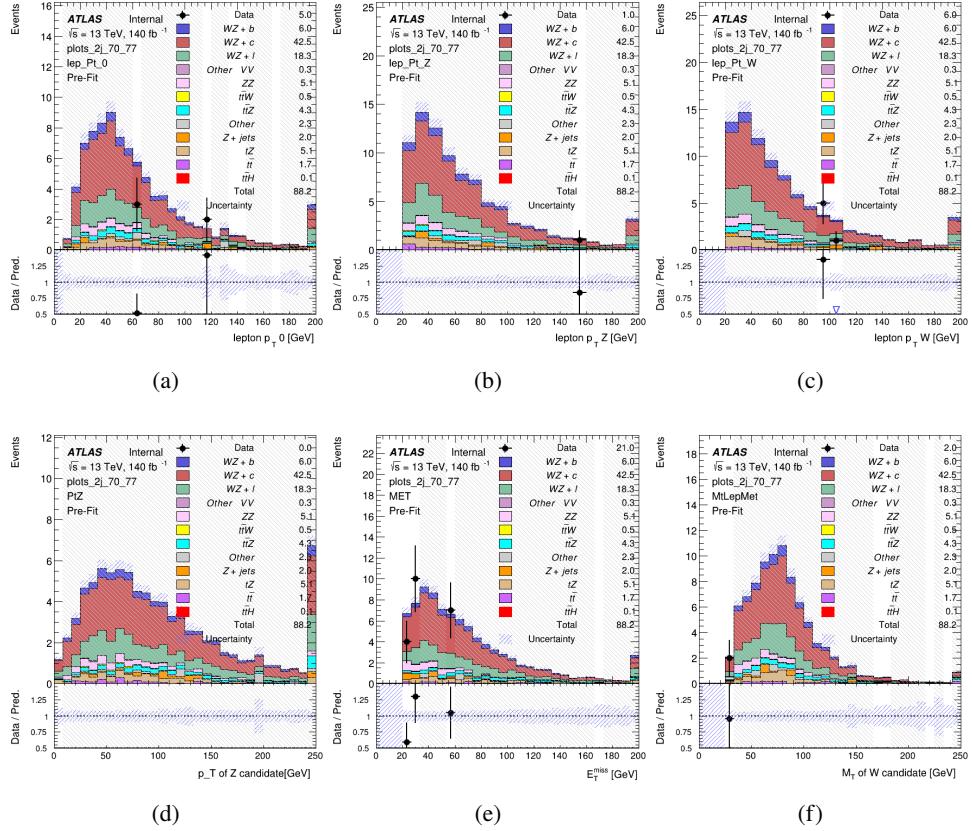


Figure 13.12: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 2j 60-70% WP

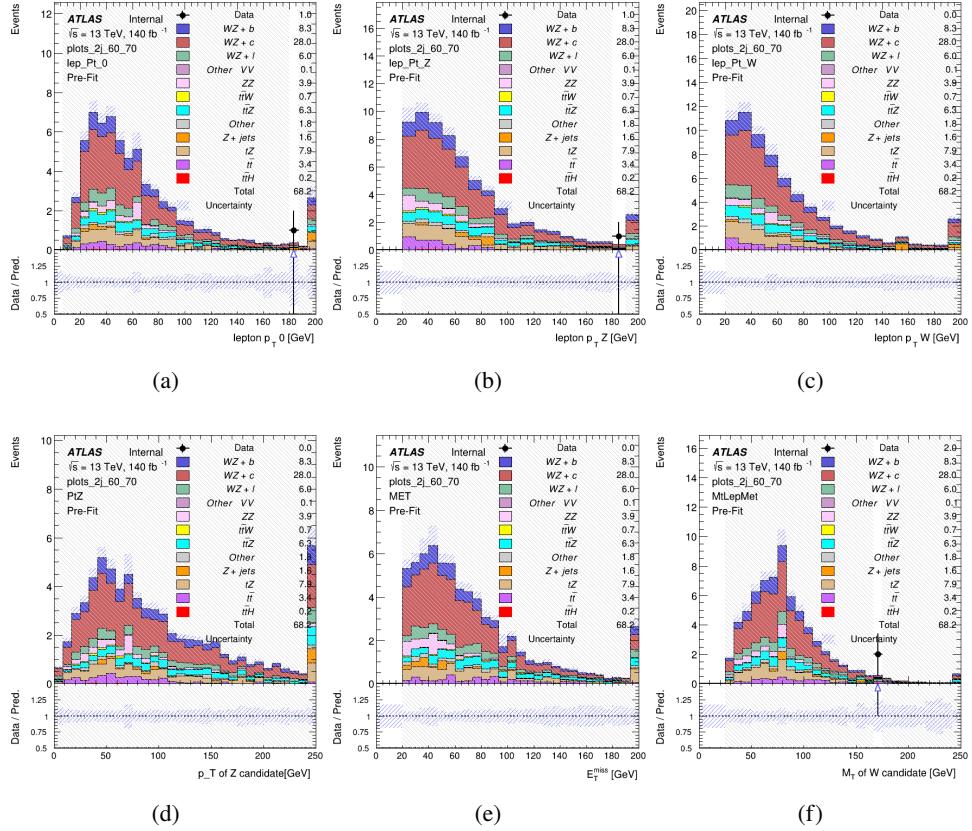


Figure 13.13: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - 2j 60% WP

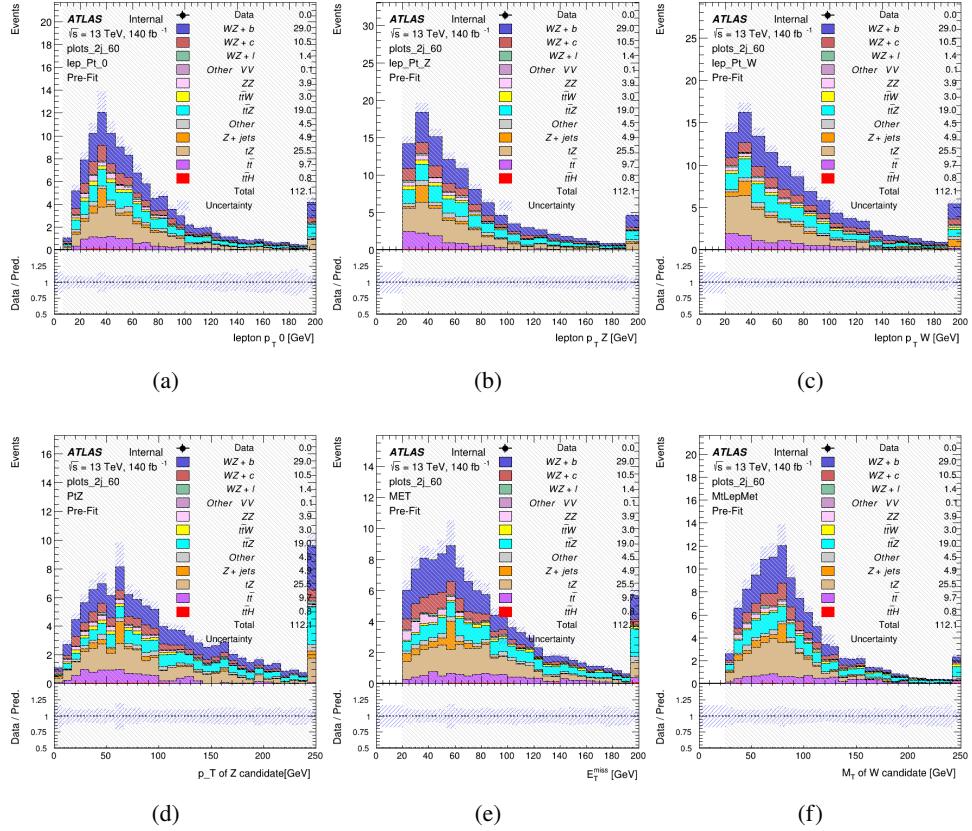


Figure 13.14: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the m_T of the W candidate.

WZ Fit Region - tZ-CR-2j

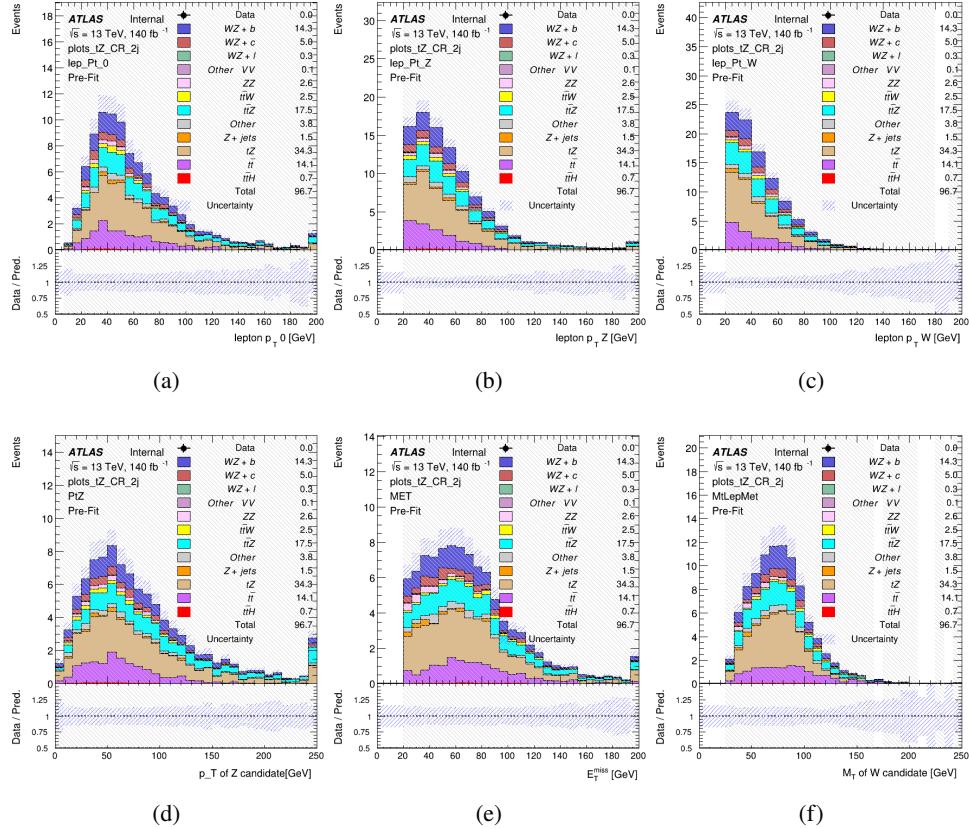


Figure 13.15: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the p_T of the Z candidate, (e) the E_T^{miss} , and (f) the M_T of the W candidate.

13.3 Non-Prompt Lepton Estimation

Two processes act as sources of non-prompt leptons appear in the analysis: $t\bar{t}$ and $Z+jet$ production both produce two prompt leptons, and each contribute to the 31 region 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 enriched validation regions.

832 13.3.1 $t\bar{t}$ Validation

833 $t\bar{t}$ events can produce two prompt leptons from the decay of each of the tops. These top decays
834 produce two b-quarks, the decay of which can produce additional non-prompt leptons, which
835 occasionally pass the event preselection. In order to validate that the Monte Carlo accurately
836 simulates this process accurately, the MC prediction in a non-prompt $t\bar{t}$ enriched validation
837 region is compared to data.

838 The $t\bar{t}$ validation region is similar to the preselection region - three leptons meeting the
839 criteria described in section 13 are required, and the requirements on E_T^{miss} remain the same.
840 However, the selection requiring a lepton pair form a Z-candidate are reversed. Events where the
841 invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of 91.2 GeV are
842 rejected. This ensures the $t\bar{t}$ validation region is orthogonal to the preselection region.

843 Further, because the jet multiplicity of $t\bar{t}$ events tends to be higher than WZ, the number
844 of jets in each event is required to be greater than 1. As b-jets are almost invariably produced
845 from top decays, at least one b-tagged jet passing the 70% DL1r WP in each event is required.
846 Various kinematic plots of this region are shown in figure 13.16.

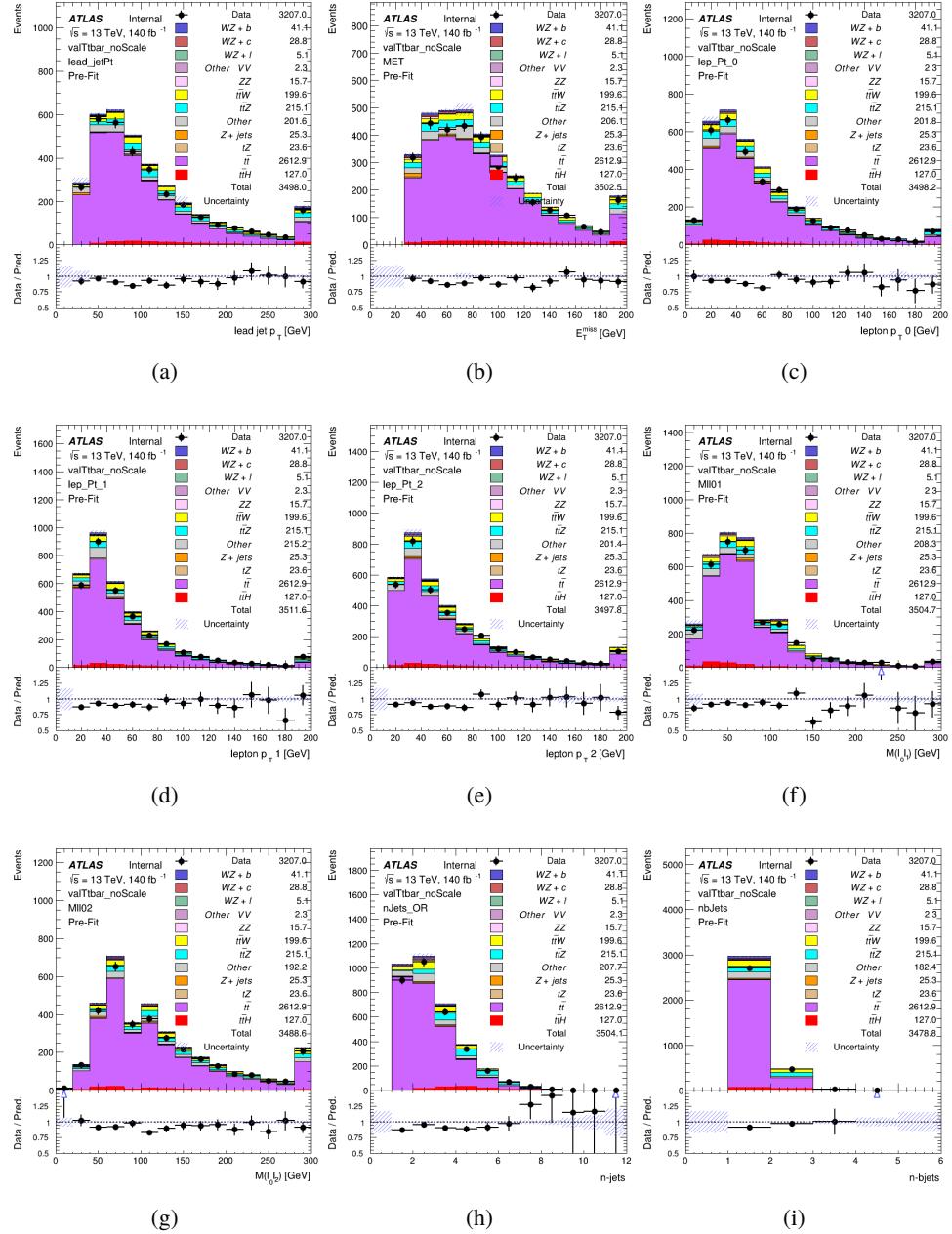


Figure 13.16: Comparisons between the data and MC distributions in the $t\bar{t}$ validation region for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) the number of b-tagged jets.

847 The shape of each distribution agrees quite well between data and MC, with a constant
848 offset between the two. This is accounted for by applying a constant correction factor of 0.883
849 to the $t\bar{t}$ MC prediction. Plots showing the kinematics of the $t\bar{t}$ VR after this correction factor
850 has been applied are shown in figure 13.17.

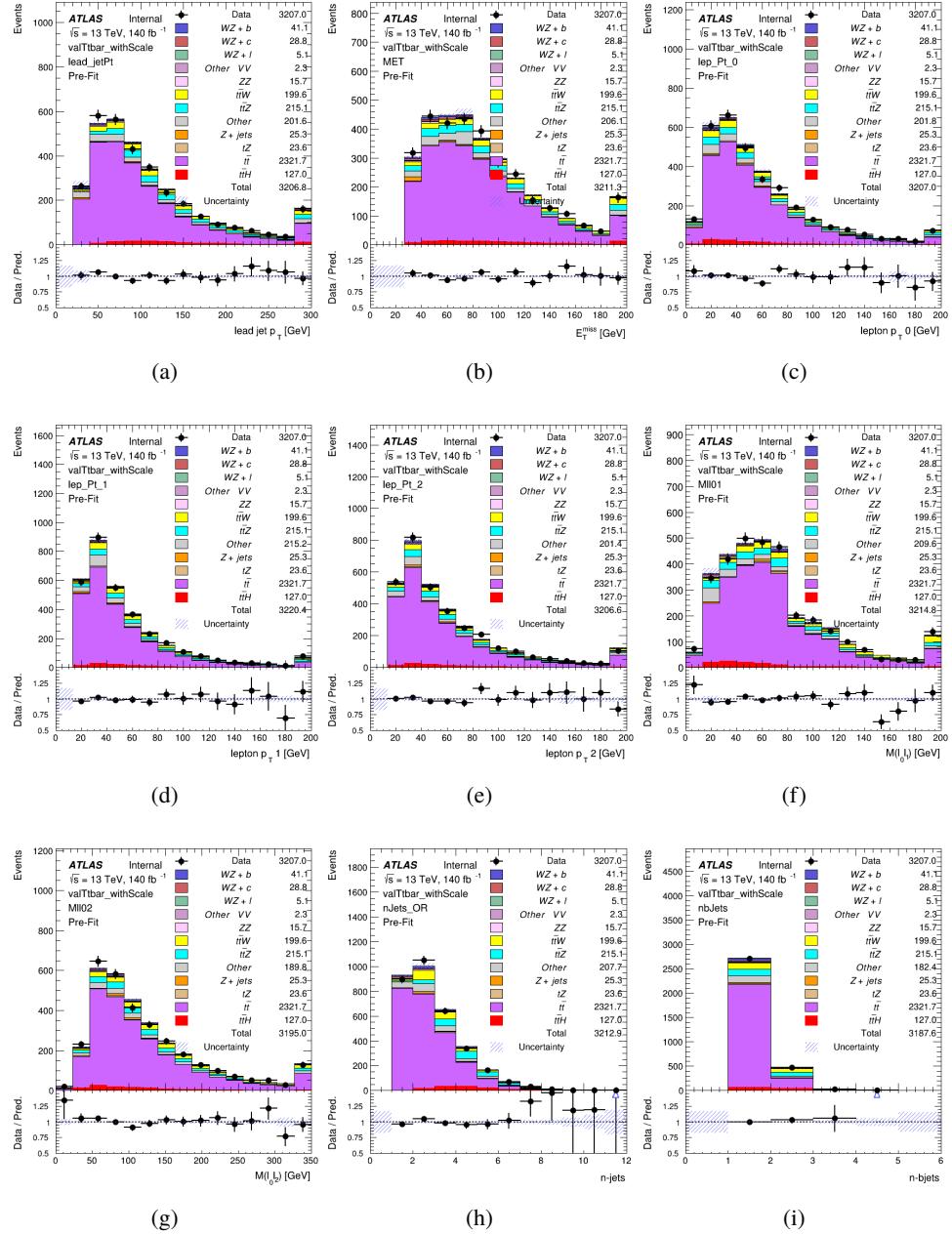


Figure 13.17: Comparisons between the data and MC distributions in the $t\bar{t}$ validation region after the correction factor has been applied for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) the number of b-tagged jets.

851 The modeling is further validated by looking at the yield in the $t\bar{t}$ VR for each DL1r WP,
 852 giving a clearer correspondence to the signal regions used in the fit. Each region shown in figure
 853 13.21 requires one or more jets pass the listed WP, with no jets passing the next highest WP.

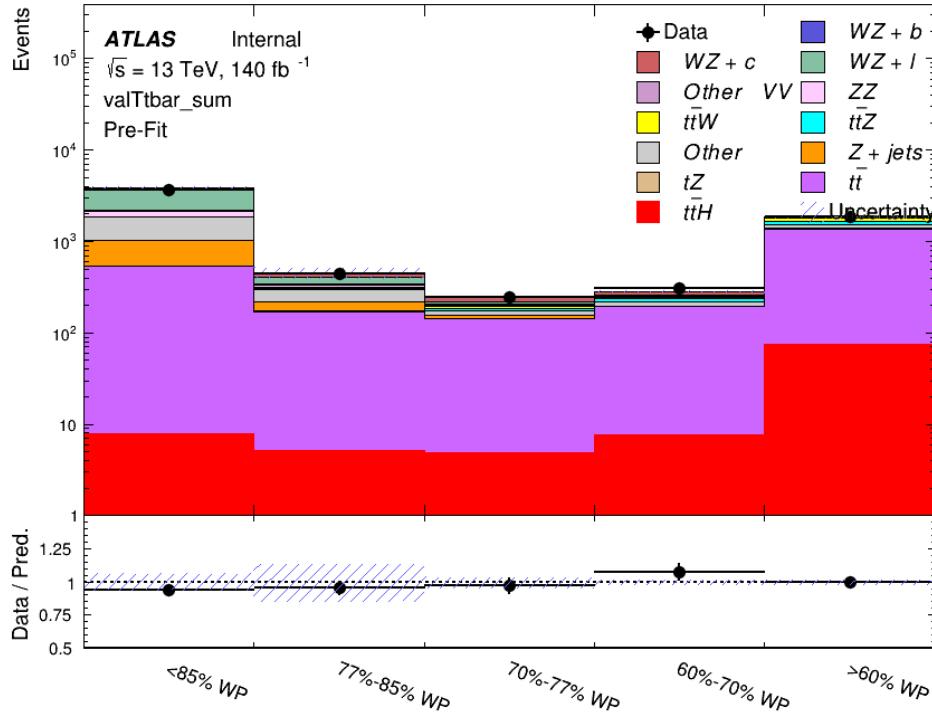


Figure 13.18: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the $t\bar{t}$ VR selection and correction factor have been applied

854 As data and MC are found to agree within 10% for each of these working points, a 10%
 855 systematic uncertainty on the $t\bar{t}$ prediction is included for the analysis.

856 13.3.2 Z+jets Validation

857 Similar to $t\bar{t}$, a non-prompt Z +jets validation region is produced in order to validate the MC
858 predictions. The lepton requirements remain the same as the preselection region. Because no
859 neutrinos are present for this process, the E_T^{miss} cut is reversed, requiring $E_T^{\text{miss}} < 30 \text{ GeV}$. This
860 also ensures this validation region is orthogonal to the preselection region. Further, the number
861 of jets in each event is required to be greater than or equal to one. Various kinematic plots of this
862 region are shown below. The general agreement between data and MC in each of these suggests
863 that the non-prompt contribution of Z +jets is well modeled by Monte Carlo.

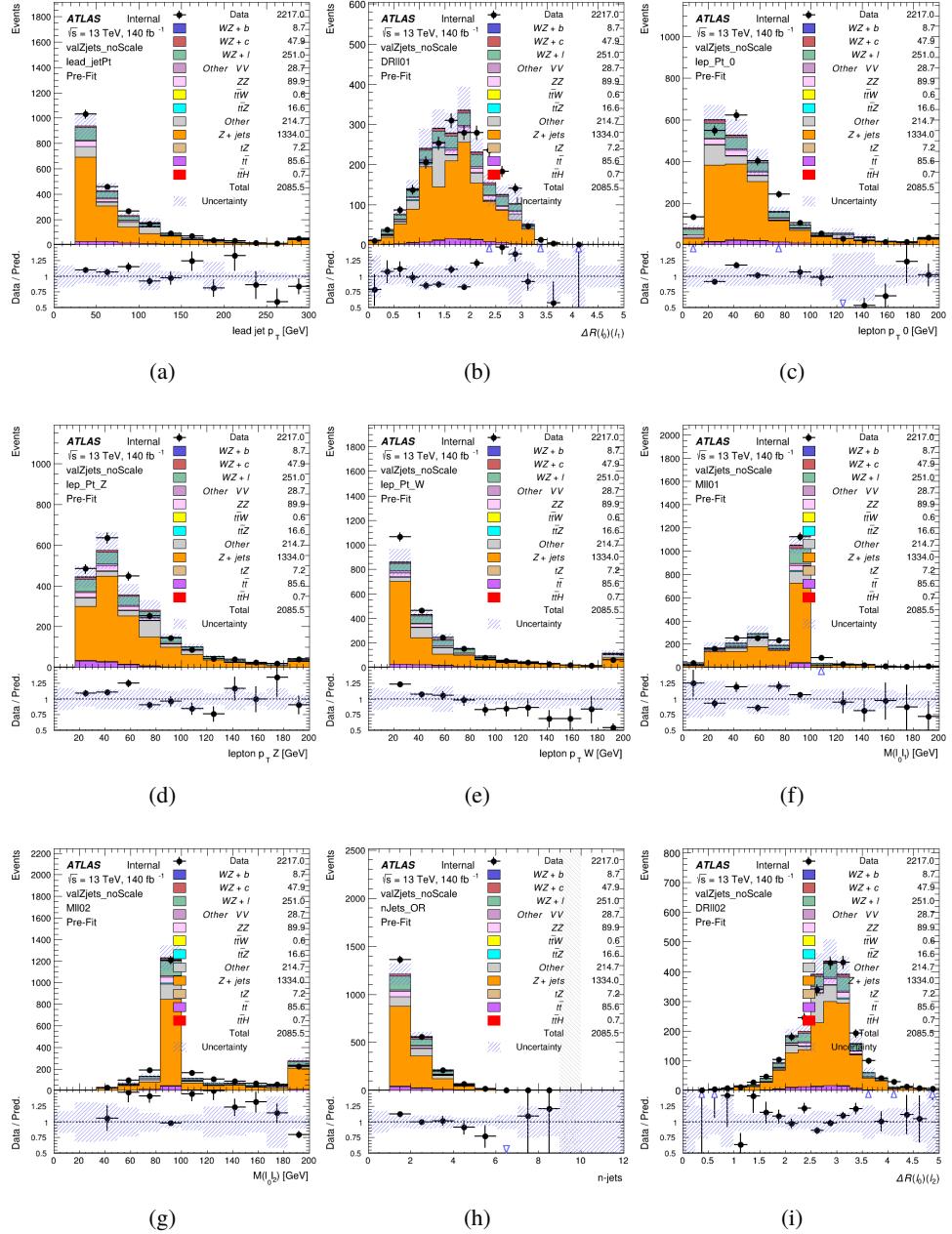


Figure 13.19: Comparisons between the data and MC distributions in the $Z+jets$ validation region for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) ΔR between leptons 0 and 2. Includes only statistical uncertainties

864 While there is general agreement between data and MC within statistical uncertainty, the
865 shape of the p_T spectrum of the lepton from the W is found to differ. To account for this
866 discrepancy, a variable correction factor is applied to Z+jets. χ^2 minimization of the lepton 2 p_T
867 spectrum is performed to derive a correction factor of $1.53 - 6.6 * 10^{-6}(\text{lep}_\text{Pt}_W)$. Kinematic
868 plots of the Z + jets validation region after this correction factor has been aplied are shown in
869 figure 13.20.

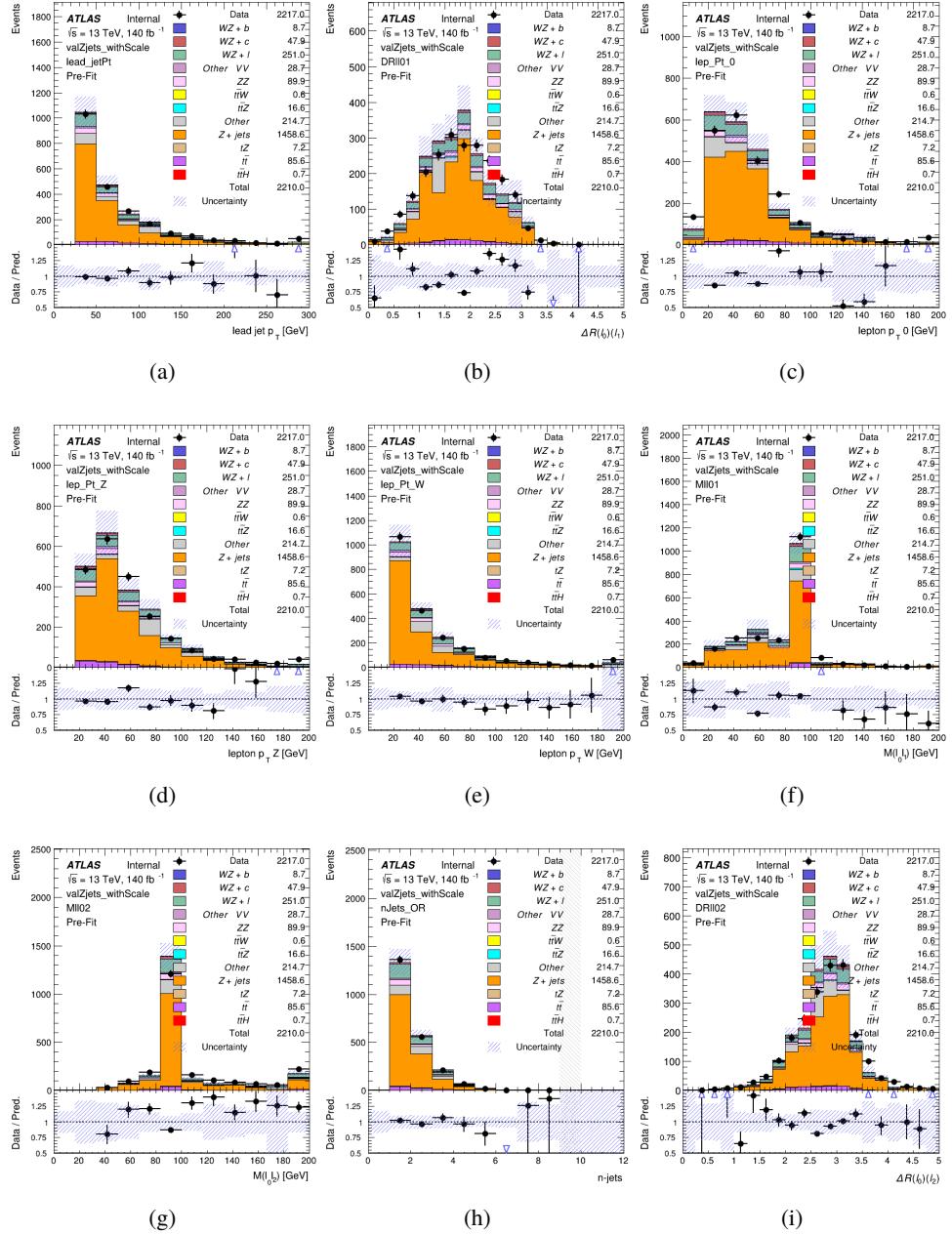


Figure 13.20: Comparisons between data and MC distributions in the Z+jets validation region after the correction factor has been applied for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) ΔR between leptons 0 and 2

870 The modeling is further validated by looking at the yield in the Z+jets VR for each DL1r
 871 WP, giving a clearer correspondence to the signal regions used in the fit. Each region shown in
 872 figure 13.21 requires one or more jets pass the listed WP, with no jets passing the next highest
 873 WP.

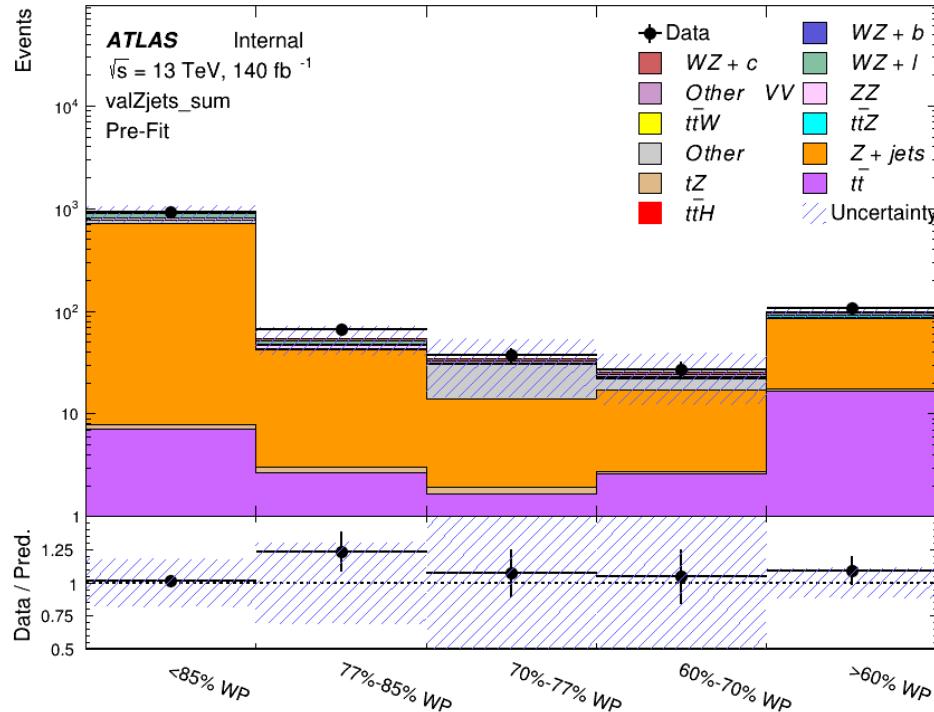


Figure 13.21: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the Z+jets VR selection and correction factor have been applied

874 For each of the b-tagging working points considered, the data falls within 20% of the MC
 875 prediction once this correction factor has been applied. Therefore, a 20% systematic uncertainty
 876 is applied to Z + jets in the analysis.

877 14 tZ Interference Studies and Separation Multivariate Analysis

878 Because tZ produces a final state identical to signal, it represents a predominant background in
879 the most signal enriched regions. That is, the region with one jet passing the 60% DL1r WP.
880 Therefore, a boosted decision tree (BDT) algorithm is trained using TMVA [TMVA_guide] to
881 separate WZ + heavy flavor from tZ.

882 Separation between tZ and WZ + heavy flavor is achieved in part by reconstructing the
883 invariant mass of the top candidate, which clusters more closely to the top mass for tZ than WZ
884 + heavy flavor.

885 The result of this BDT is used to create a tZ enriched region in the fit, reducing its impact
886 on the measurement of WZ + heavy flavor.

887 14.1 Top Mass Reconstruction

888 The reconstruction of the top mass follows the procedure described in detail in section 6.1 of
889 [ttZ_paper]. The mass of the top quark candidate is reconstructed from the jet, the lepton not
890 included in the Z-candidate, and a reconstructed neutrino. In the case that there is one jet in the
891 event, there is only possible b-jet candidate. For events with two jets, the jet with the highest
892 DL1r score is used.

893 The neutrino from the W decay is expected to be the only source of E_T^{miss} . Therefore, the
 894 E_T and ϕ of the neutrino are taken from the E_T^{miss} measurement. This leaves the z-component
 895 of the neutrino momentum, $p_{\nu z}$ as the only unknown.

896 This unknown is solved for by taking the combined invariant mass of the lepton and
 897 neutrino to give the invariant mass of the W boson:

$$898 \quad (p_l + p_\nu)^2 = m_W^2$$

899 Expanding this out into components, this equation gives:

$$900 \quad \sqrt{p_{T\nu}^2 + p_{z\nu}^2} E_l = \frac{m_W^2 - m_l^2}{2} + p_{T\nu}(p_{lx}\cos\phi_\nu + p_{ly}\sin\phi_\nu) + p_{lz}p_{\nu z}$$

901 This equation gives two solutions for $p_{\nu z}$. For cases where only one of these solutions is
 902 real, that is taken as the value of $p_{\nu z}$. For instances with two real solutions, the one which is
 903 shown to be correct in the largest fraction of simulations is taken. For cases when no real solution
 904 is found, often because of detector effects, the value of E_T^{miss} is varied in decreasing increments
 905 of 100 MeV until a real solution is found.

906 The reconstructed top mass distribution for tZ and WZ + b can be seen in figure 14.1.

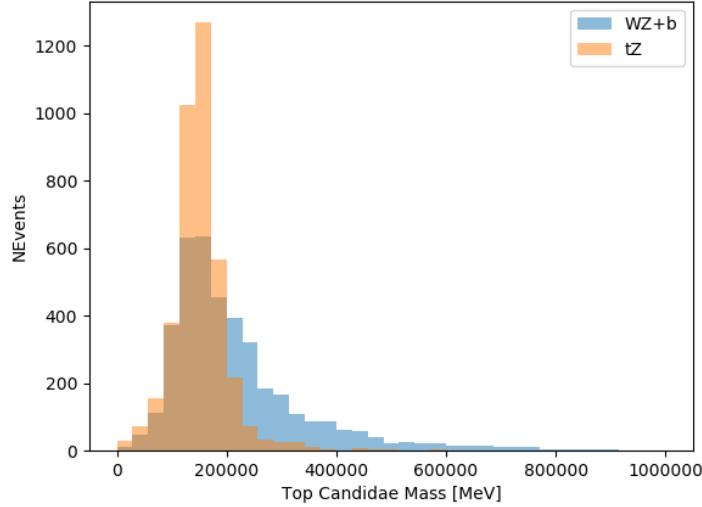


Figure 14.1: Reconstructed top mass distributions for tZ and WZ + b, measured in MeV.

907 14.2 tZ BDT

908 A Boosted Decision Tree (BDT), specifically XGBoost [xgboost_cite], is used to provide separ-
 909 ation between tZ and WZ+b. The following kinematic variables are used as inputs:

- 910 • The invariant mass of the reconstructed top candidate
- 911 • p_T of each of the leptons, jet
- 912 • The invariant mass of each combination of lepton pairs, $M(l\bar{l})$
- 913 • E_T^{miss}
- 914 • Distance between each combination of leptons, $\Delta R(l\bar{l})$

- 915 • Distance between each lepton and the jet, $\Delta R(lj)$

916 The training samples included only events meeting the requirements of the 1-jet, $>60\%$
917 region, i.e. passing all the selection described in section 13 and having exactly one jet which
918 passes the tightest (60%) DL1r working point.

919 The distributions of a few of these features for both signal and background is shown in
920 figure 14.2.

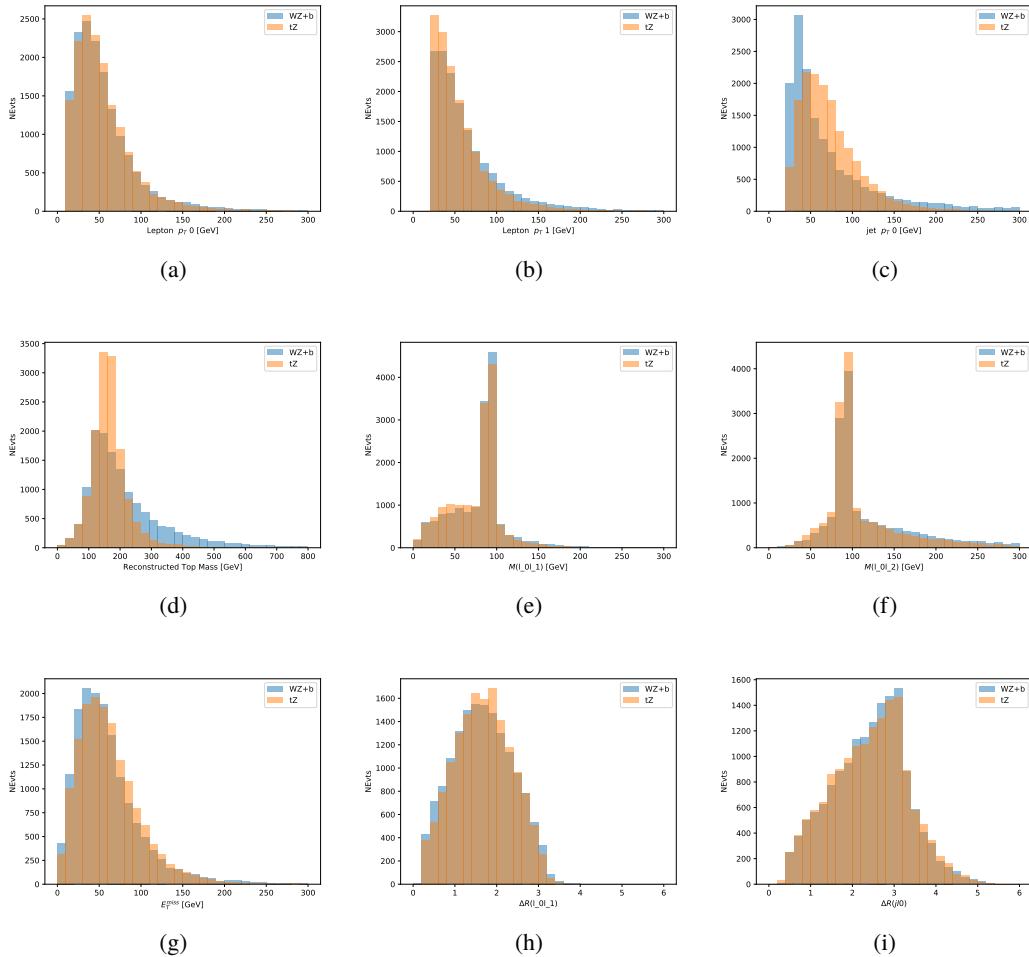


Figure 14.2: Distribution of input features of the BDT for signal (WZ) and background (tZ). Both are scaled to an equal number of events. (a), (b) and (c) show the p_T of lepton 0, lepton 1, and the jet, (d) show the reconstructed top mass, (e) and (f) show the invariant mass of leptons 0 and 1, leptons 0 and 2. (g) shows the E_T^{miss} of each event. (h) and (i) show the ΔR between lepton 0 and lepton 1, and the jet.

921 A sample of 20,000 background (tZ) and signal (WZ+b) Monte Carlo events are used to
 922 train the BDT. And additional 5,000 events are reserved for testing the model, in order to prevent
 923 over-fitting. A total of 750 decision trees with a maximum depth of 6 branches are used to build
 924 the model. These parameters are chosen empirically, by training several models with different

925 parameters and selecting the one that gave the best separation for the test sample.

926 The results of the BDT training are shown in figure 14.3. The output scores for both signal
 927 and background events is shown on the left. The right shows the receiving operating characteristic
 928 (ROC) curve that results from the MVA. The ROC curve represents the background rejection
 929 as a function of signal efficiency, where each point on the curve represents a different response
 930 score. The ROC curve of the BDT is compared to the performance of using an optimal set of flat
 931 selections on the same set of input variables.

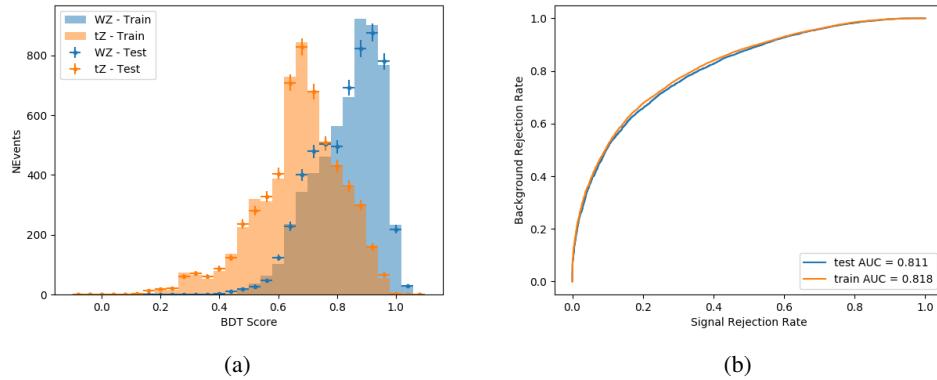


Figure 14.3: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

932 The relative important of each input feature in the model, measured by how often they
 933 appeared in the decision trees, is shown in figure 14.4.

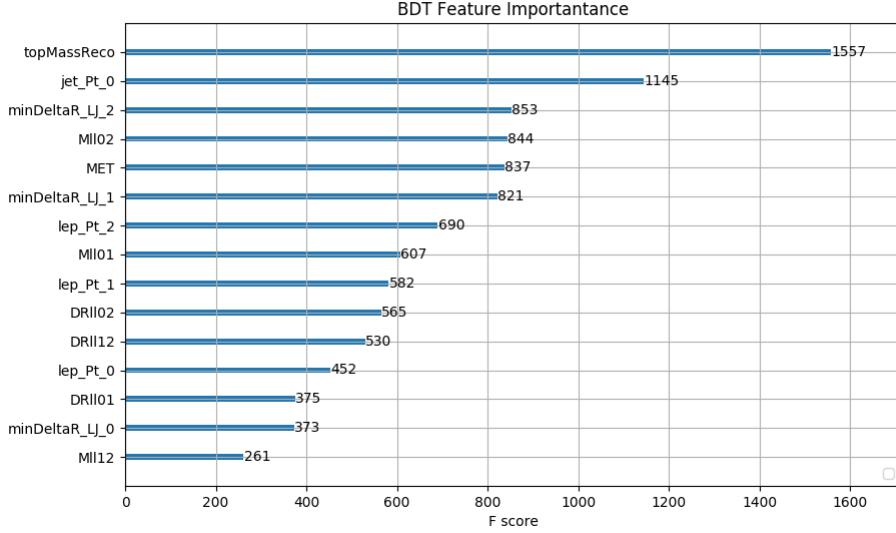


Figure 14.4: Relative importance of each input feature in the model.

934 These results suggest that some amount of separation can be achieved between these two
 935 processes, with a high BDT score selecting a set of events that is pure in $WZ + b$. A BDT score
 936 of 0.725 is selected as a cutoff, where events with scores higher than this form a signal enriched
 937 region, and events with scores lower than this form a tZ control region. This cutoff is selected by
 938 varying the value of this cutoff in stat-only Asimov fits, and selecting the value that minimizes
 939 the statistical uncertainty on $WZ + b$.

940 15 Systematic Uncertainties

941 The systematic uncertainties that are considered are summarized in table 47. These are imple-
 942 mented in the fit either as a normalization factors or as a shape variation or both in the signal

943 and background estimations. The numerical impact of each of these uncertainties is outlined in

944 section 23.

Table 15: 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
Background Modeling	
Cross section	15
Renormalization and factorization scales	12
Total (Signal and background modeling)	35
Total (Overall)	229

945 The uncertainty in the combined integrated luminosity is derived from a calibration of the

946 luminosity scale performed for 13 TeV proton-proton collisions [lumi], [LUCID2].

947 The experimental uncertainties are related to the reconstruction and identification of light

948 leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The sources which contribute to

949 the uncertainty in the jet energy scale (JES) [jes] are decomposed into uncorrelated components

950 and treated as independent sources in the analysis. The CategoryReduction model is used to

951 account for JES uncertainties, which decomposes the uncertainties into 30 nuisance parameters

952 included in the fit. The SimpleJER model is used to account for jet energy resolution (JER)
953 uncertainties, and 8 JER uncertainty components unclued as NPs in the fit.

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

957 The full list of systematic uncertainties considered in the analysis is summarized in tables
958 [16](#), [17](#) and [18](#).

959

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

Table 16: Summary of experimental systematics considered for leptons and E_T^{miss} . Includes type, description, name of systematic as used in the fit, and mode of application. The mode of application indicates the systematic evaluation, e.g. as an overall event re-weighting (Event Weight) or rescaling (p_T Correction).

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

Table 17: Jet systematics take into account effects of jets calibration method, η inter-calibration, high p_T jets, pile-up, and flavor response. They are all diagonalised into effective parameters.

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

Table 18: Summary of experimental systematics to be included for b-tagging of jets in the analysis, using the continuous DL1r tagging algorithm. All of the b-tagging related systematics are applied as event weights. From left: type, description, and the name of systematic used in the fit.

960 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and
 961 scale uncertainties are taken from theory calculations, with the exception of non-prompt and
 962 diboson backgrounds. The cross-section uncertainty on tZ is taken from [[tZ_paper](#)]. Derivation
 963 of the non-prompt background uncertainties, Z+jets and t \bar{t} , are explained in detail in section
 964 [13.3](#).

965 The other VV + heavy flavor processes (namely VV+b and VV+charm, which primarily
 966 consist of ZZ events) are also poorly understood, because these processes involve the same
 967 physics as WZ + heavy flavor, and have also not been measured. Therefore, a conservative 50%
 968 uncertainty is applied to those samples. While this uncertainty is large, it is found to have little
 969 impact on the significance of the final result.

970 The theory uncertainties applied to the predominate background estimates are summarized
 971 in table 19.

Process	X-section [%]
tZ	X-sec: ± 15.2 QCD Scale: $^{+5.2}_{-1.3}$ PDF($+\alpha_S$): ± 1.2
t̄H (aMC@NLO+Pythia8)	QCD Scale: $^{+5.8}_{-9.2}$ PDF($+\alpha_S$): ± 3.6
t̄Z (aMC@NLO+Pythia8)	QCD Scale: $^{+9.6}_{-11.3}$ PDF($+\alpha_S$): ± 4
t̄W (aMC@NLO+Pythia8)	QCD Scale: $^{+12.9}_{-11.5}$ PDF($+\alpha_S$): ± 3.4
VV + b/charm (Sherpa 2.2.1)	± 50
VV + light (Sherpa 2.2.1)	± 6
t̄t	± 10
Z + jets	± 20
Others	± 50

Table 19: Summary of theoretical uncertainties for MC predictions in the analysis.

972 Additional signal uncertainties are estimated by comparing estimates from the nominal
 973 Sherpa WZ samples with alternate WZ samples generated with Powheg+PYTHIA8 (DSID
 974 361601). The shape of the templates used in the fit are compared between these two samples for
 975 WZ + b, WZ + charm and WZ + light, as shown in figures 15.1 and 15.2. Each of these plots are
 976 normalized to unity in order to capture differences in shape.

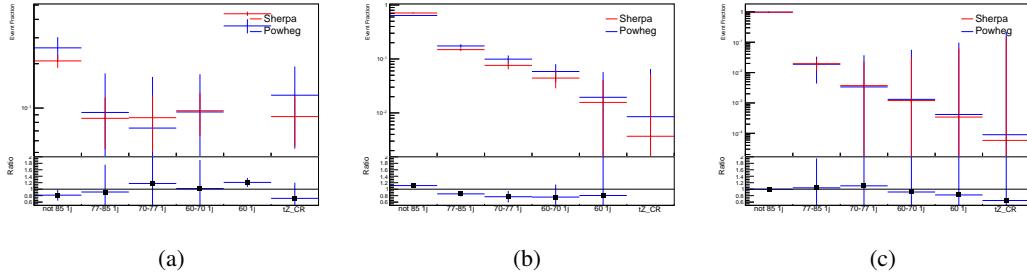


Figure 15.1: Comparison between Sherpa and Powheg predictions of the distribution of (a) $\text{WZ} + \text{b}$, (b) $\text{WZ} + \text{charm}$, and (c) $\text{WZ} + \text{light}$ among the various b-tag WPs used in the 1-jet fit.

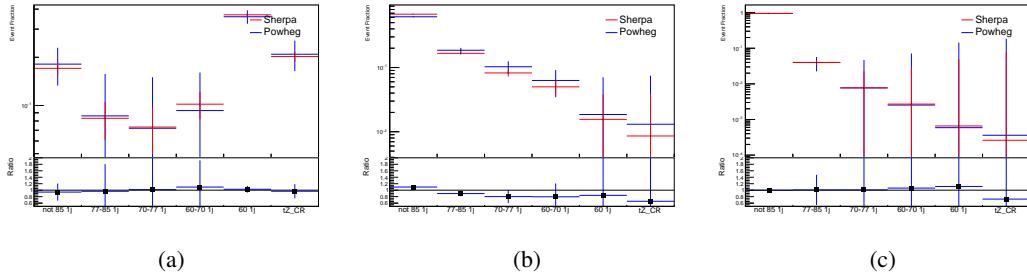


Figure 15.2: Comparison between Sherpa and Powheg predictions of the distribution of (a) $\text{WZ} + \text{b}$, (b) $\text{WZ} + \text{charm}$, and (c) $\text{WZ} + \text{light}$ among the various b-tag WPs used in the 2-jet fit.

977 Separate systematics are included in the fit for $\text{WZ} + \text{b}$, $\text{WZ} + \text{charm}$ and $\text{WZ} + \text{light}$,
 978 where the distribution among each of the fit regions is varied based on the prediction of the
 979 Powheg sample.

980 16 Results

981 A separate maximum-likelihood fit is performed over the 1-jet and 2-jet fit regions in order to
 982 extract the best-fit value of the $\text{WZ} + \text{b}$ -jet and $\text{WZ} + \text{charm}$ jet contributions. The $\text{WZ} + \text{b}$,

983 WZ + charm and WZ + light contributions are allowed to float, with the remaining background
 984 contributions are held fixed. **The current fit strategy treats the WZ + b-jet contribution as**
 985 **the parameter of interest, with the normalization of the WZ + charm and the WZ + light**
 986 **contributions taken as systematic uncertainties. This could however be adjusted, depending**
 987 **on whether it is decided the goal of the analysis should be to measure WZ+b specifically or**
 988 **WZ + heavy flavor overall.** The result of the fit is used to extract the cross-section of WZ +
 989 heavy-flavor production.

990 A maximum likelihood fit to data is performed simultaneously in the regions described
 991 in section 13. The parameters μ_{WZ+b} , $\mu_{WZ+charm}$, $\mu_{WZ+light}$, where $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$, are
 992 extracted from the fit.

993 The Asimov fit for 1-jet events gives an expected μ value of $1.00^{+0.37}_{-0.35}(\text{stat})^{+0.24}_{-0.22}(\text{sys})$ for
 994 WZ + b. The fitted cross-section modifiers for WZ + charm and WZ + light are $1.00 \pm 0.17 \pm 0.15$
 995 and $1.00 \pm 0.04 \pm 0.07$, respectively.

996 The expected cross-section of WZ+b with 1 associated jet obtained from the fit is
 997 $1.74^{+0.65}_{-0.61}(\text{stat})^{+0.42}_{-0.37}(\text{sys}) \text{ fb}$ with an expected significance of 2.2σ .

998 For 2-jet events, the fit gives an expected μ value of $1.00^{+0.34}_{-0.33}(\text{stat})^{+0.29}_{-0.28}(\text{sys})$ for WZ +
 999 b. The fitted cross-section modifiers for WZ + charm and WZ + light are $1.00 \pm 0.17 \pm 0.15$ and
 1000 $1.00 \pm 0.04 \pm 0.08$, respectively.

¹⁰⁰¹ The expected $WZ + b$ cross-section in the 2-jet region is $2.46^{+0.83}_{-0.81}(\text{stat})^{+0.073}_{-0.68}(\text{sys}) \text{ fb}$
¹⁰⁰² with an expected significance of 2.6σ .

¹⁰⁰³ **16.1 1-jet Fit Results**

¹⁰⁰⁴ **The results of the fit are currently blinded.**

¹⁰⁰⁵ The pre-fit yields in each of the regions used in the fit are shown in table 16.1, and
¹⁰⁰⁶ summarized in figure 16.1.

¹⁰⁰⁷

Sample	1j, <85% WP	1j, 77%-85% WP	1j, 70%-77% WP	1j, 60%-70% WP	1j, >60% WP	tZ CR
$WZ + b$	11 ± 2	4.7 ± 0.5	4.6 ± 0.4	5.1 ± 0.4	24 ± 2	6.0 ± 0.5
$WZ + c$	318 ± 22	81 ± 6	43.1 ± 3.6	25.8 ± 2.6	9.4 ± 1.8	2.9 ± 0.6
$WZ + l$	4020 ± 250	91 ± 13	17 ± 3	4.9 ± 1.6	1.3 ± 0.4	0.2 ± 0.1
Other VV	6.2 ± 0.6	0.2 ± 0.4	0.2 ± 0.04	0.07 ± 0.1	0.1 ± 0.1	0.1 ± 0.2
ZZ	336 ± 26	17.8 ± 2.1	4.3 ± 0.6	1.7 ± 0.5	0.36 ± 0.08	0.10 ± 0.03
$t\bar{t}W$	1.1 ± 0.2	0.2 ± 0.1	0.3 ± 0.1	0.4 ± 0.1	1.5 ± 0.3	0.7 ± 0.2
$t\bar{t}Z$	6.8 ± 1.2	1.4 ± 0.3	1.0 ± 0.2	1.2 ± 0.2	4.4 ± 0.8	3.2 ± 0.6
$Z + \text{jets}$	169 ± 38	8.9 ± 1.9	3.7 ± 0.8	3.3 ± 0.7	3.2 ± 0.7	0.8 ± 0.17
$V + \gamma$	45 ± 28	1.9 ± 2.4	0.1 ± 0.1	0.02 ± 0.01	1.0 ± 0.9	0.02 ± 0.03
tZ	24.3 ± 4.3	5.5 ± 1.1	4.1 ± 0.8	5.9 ± 1.1	10.7 ± 2.0	23 ± 4
tW	1.4 ± 0.8	0.2 ± 0.5	0.0 ± 0.2	0.7 ± 0.6	0.26 ± 0.42	0.39 ± 0.41
WtZ	2.3 ± 1.2	0.6 ± 0.3	0.3 ± 0.21	0.27 ± 0.2	1.1 ± 0.7	0.6 ± 0.5
VVV	12.4 ± 0.5	0.93 ± 0.06	0.35 ± 0.03	0.13 ± 0.02	0.14 ± 0.03	0.02 ± 0.01
VH	40 ± 6	2.6 ± 1.4	0.9 ± 0.8	0.7 ± 0.8	0.5 ± 0.6	0.0 ± 0.0
$t\bar{t}$	12.1 ± 1.6	2.9 ± 0.6	2.5 ± 0.5	2.8 ± 0.5	11.2 ± 1.4	10.9 ± 1.5
$t\bar{t}H$	0.24 ± 0.03	0.05 ± 0.01	0.04 ± 0.01	0.06 ± 0.01	0.20 ± 0.03	0.13 ± 0.02
Total	5010 ± 260	227 ± 24	88 ± 12	57 ± 8	76 ± 16	53 ± 8

Table 20: Pre-fit yields in each of the 1-jet fit regions.

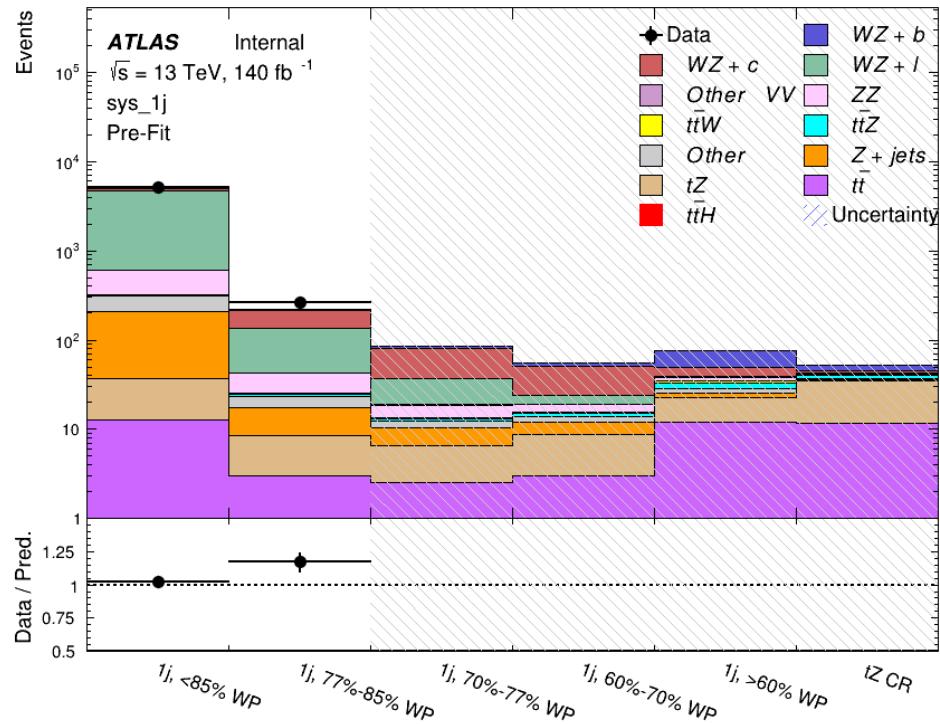


Figure 16.1: Pre-fit summary of the 1-jet fit regions.

1008

The post-fit yields in each region are summarized in figure 16.1.

1009

	1j, <85% WP	1j, 77%-85% WP	1j, 70%-77% WP	1j, 60%-70% WP	1j, >60% WP	tZ CR
WZ + b	11 ± 5	4.7 ± 2.0	4.6 ± 2.0	5.1 ± 2.1	24 ± 10	6.0 ± 2.50
WZ + c	320 ± 60	80 ± 14	43 ± 7	26 ± 5	9.4 ± 2.3	2.9 ± 0.7
WZ + l	4020 ± 130	90 ± 11	17.3 ± 2.8	4.9 ± 1.6	1.3 ± 0.4	0.23 ± 0.13
Other VV	6.2 ± 0.6	0.92 ± 0.07	0.02 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01
ZZ	346 ± 57	19 ± 5	4.3 ± 0.8	2.7 ± 0.5	2.4 ± 0.1	2.1 ± 0.6
t̄tW	1.09 ± 0.21	0.2 ± 0.1	0.1 ± 0.1	0.4 ± 0.1	1.5 ± 0.3	0.1 ± 0.2
t̄tZ	6.8 ± 1.2	1.4 ± 0.3	1.0 ± 0.2	1.2 ± 0.2	4.4 ± 0.7	3.2 ± 0.5
rare Top	0.14 ± 0.04	0.04 ± 0.02	0.04 ± 0.0	0.1 ± 0.03	0.14 ± 0.04	0.15 ± 0.05
t̄tWW	0.04 ± 0.03	0.01 ± 0.02	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.02	0.01 ± 0.01
Z + jets	169 ± 37	8.9 ± 1.9	3.7 ± 0.8	3.3 ± 0.7	3.2 ± 0.7	0.8 ± 0.2
W + jets	0.01 ± 0.01					
V + γ	46 ± 28	1.9 ± 2.4	0.1 ± 0.1	0.0 ± 0.2	1.0 ± 0.9	0.0 ± 0.0
tZ	24 ± 4	5.5 ± 1.0	4.1 ± 0.8	5.9 ± 1.1	10.7 ± 1.8	23.3 ± 3.7
tW	1.37 ± 0.82	0.18 ± 0.26	0.01 ± 0.12	0.67 ± 0.64	0.26 ± 0.42	0.39 ± 0.41
WtZ	2.3 ± 1.2	0.6 ± 0.3	0.3 ± 0.2	0.3 ± 0.2	1.1 ± 0.6	0.6 ± 0.3
VVV	12.4 ± 0.4	0.9 ± 0.1	0.4 ± 0.1	0.13 ± 0.02	0.14 ± 0.03	0.02 ± 0.01
VH	40 ± 6	2.6 ± 1.4	0.9 ± 0.8	0.7 ± 0.8	0.4 ± 0.6	0.01 ± 0.01
t̄t	12.1 ± 1.6	2.9 ± 0.6	2.5 ± 0.5	2.8 ± 0.5	11.2 ± 1.5	10.9 ± 1.4
t̄tH	0.24 ± 0.03	0.05 ± 0.01	0.04 ± 0.01	0.06 ± 0.01	0.20 ± 0.03	0.13 ± 0.02
Total	5100 ± 110	227 ± 12	87 ± 6	56.7 ± 4.4	76 ± 9	52.5 ± 4.2

Table 21: Post-fit yields in each of the 1-jet fit regions.

¹⁰¹⁰

A post-fit summary plot of the 1-jet fitted regions is shown in figure 16.2:

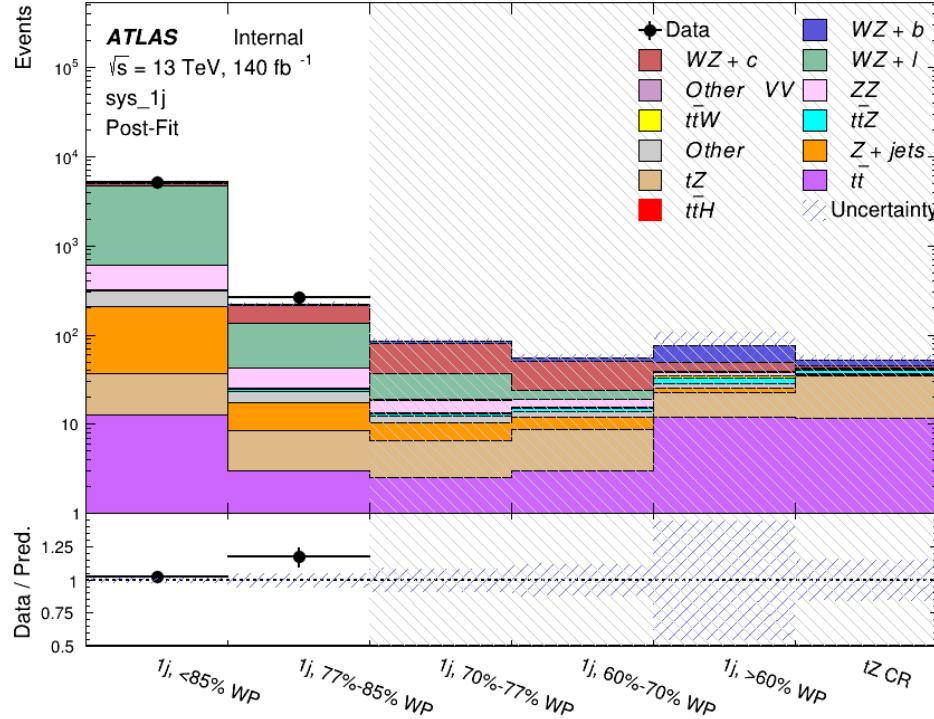


Figure 16.2: Post-fit summary of the 1-jet fit regions.

1011 As described in section 22, there are 226 systematic uncertainties that are considered
 1012 as NPs in the fit. These NPs are constrained by Gaussian or log-normal probability density
 1013 functions. The latter are used for normalisation factors to ensure that they are always positive.
 1014 The expected number of signal and background events are functions of the likelihood. The prior
 1015 for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its
 1016 nominal value.

1017 The impact of each NP is calculated by performing the fit with the parameter of interest held
 1018 fixed, varied from its fitted value by its uncertainty, and calculating $\Delta\mu$ relative to the baseline

¹⁰¹⁹ fit. The impact of the most significant sources of systematic uncertainties is summarized in table

¹⁰²⁰ [22.](#)

Uncertainty Source	$\Delta\mu$	
WZ + charm cross-section	-0.1966	0.2171
tZ cross-section	-0.1521	0.1518
WZ + light cross-section	0.1485	-0.1411
Other VV + b cross-section	-0.1115	0.1163
Flavor Tagging	0.0955	0.0957
Jet Energy Scale	0.0613	0.081
t̄t cross-section	-0.0662	0.0654
Luminosity	-0.0609	0.0655
Z + jets cross-section	-0.0284	0.0284
Other VV + charm cross-section	0.0207	-0.0202
Muon Trigger Scale Factor	0.019	0.0209
Total Systematic Uncertainty	0.3511	0.3679

Table 22: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

¹⁰²¹ The ranking and impact of those nuisance parameters with the largest contribution to the

¹⁰²² overall uncertainty is shown in figure [16.3](#).

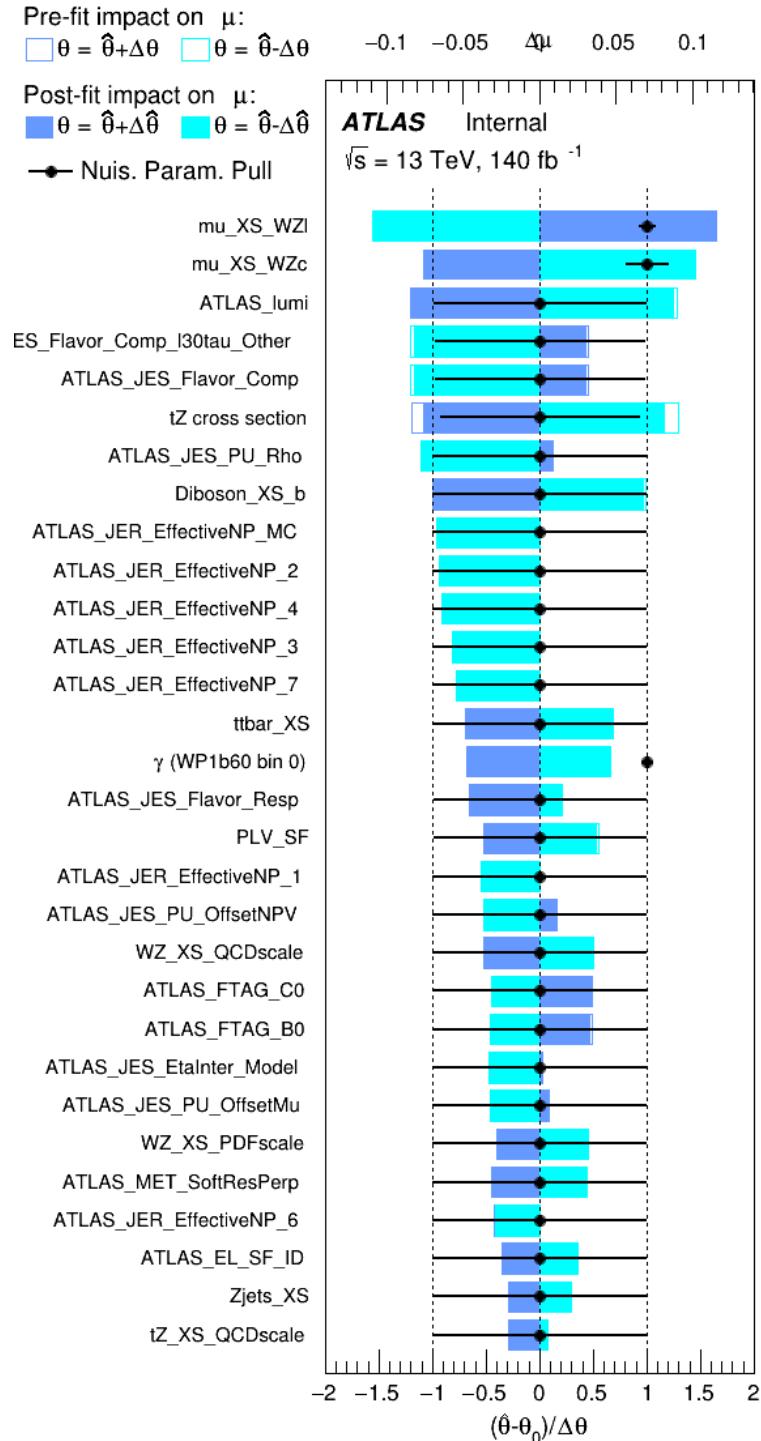


Figure 16.3: Impact of systematic uncertainties on the signal-strength of $WZ + b$ for events with exactly one jet

1023 The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the
 1024 shape of the fit regions depends heavily on the modeling of the jets. The other major sources of
 1025 uncertainty come from background modelling and cross-section uncertainty. The pie charts in
 1026 figure 16.4 show that for the modelling uncertainties that contribute most correspond to the most
 1027 significant backgrounds.

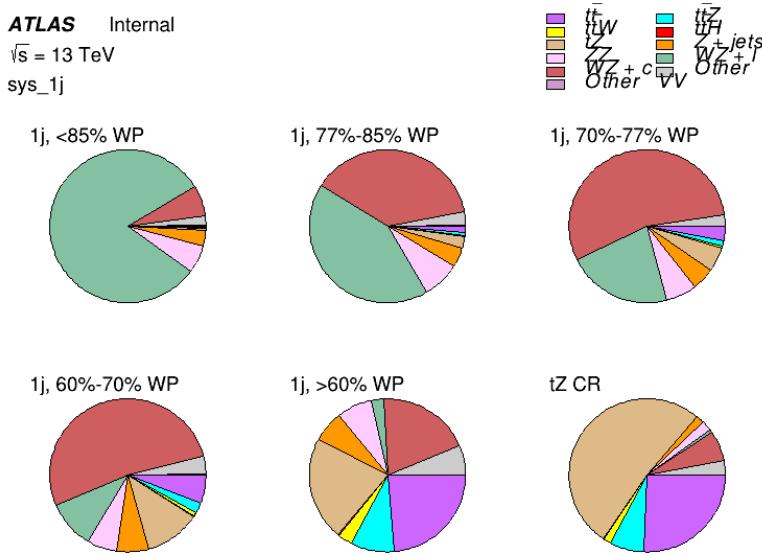


Figure 16.4: Post-fit background composition of the fit regions.

1028 The correlations between these nuisance parameters are summarized in figure 16.5.

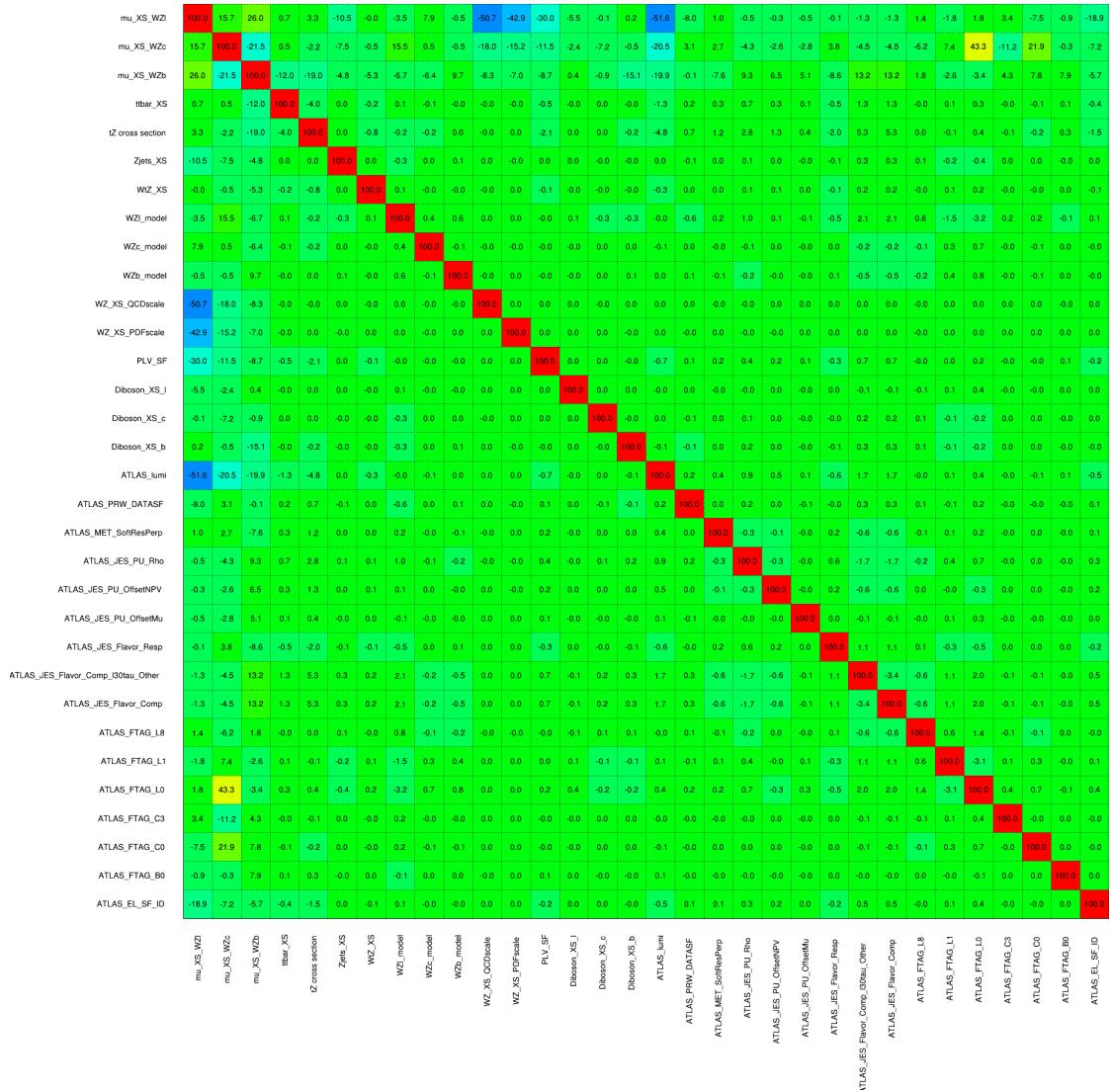


Figure 16.5: Correlations between nuisance parameters

1029 The negative correlations between $\mu_{WZ+charm}$ and μ_{WZ+b} and $\mu_{WZ+light}$ are expected:
1030 WZ + charm is present in both the WZ + b and WZ + light enriched regions, therefore increasing
1031 the fraction of charm requires increasing the fraction of WZ + b and WZ + light. This reasoning
1032 also explains the positive correlation between μ_{WZ+b} and $\mu_{WZ+light}$.

1033 Two of the major backgrounds in the region with the highest purity of $WZ + b$ are tZ and
1034 Other $VV + b$, explaining the negative correlations between μ_{WZ+b} and the tZ cross section, and
1035 the $VV + b$ cross section.

1036 The high correlation between the luminosity and $\mu_{WZ+light}$ arises from the fact that the
1037 uncertainty on $\mu_{WZ+light}$ is very low (around 4%). Small changes in luminosity cause a change
1038 in the yield of $WZ + light$ that is large compared to its uncertainty, producing a large correlation
1039 between these two parameters.

1040 **16.2 2-jet Fit Results**

1041 **The results of the fit are currently blinded.**

1042 Pre-fit yields in each of the 2-jet fit regions are shown in figure 16.2.

	2j, <85% WP	2j, 77%-85% WP	2j, 70%-77% WP	2j, 60%-70% WP	2j, >60% WP	tZ CR 2j
WZ + b	13 ± 2	6.7 ± 0.5	5.6 ± 0.4	8.0 ± 0.6	31 ± 2	14 ± 1
WZ + c	260 ± 20	77 ± 6	41 ± 3	27 ± 3	11 ± 2	4.8 ± 0.6
WZ + l	1860 ± 150	90 ± 14	18 ± 3	5.8 ± 1.4	1.4 ± 0.4	0.25 ± 0.15
Other VV	7.63 ± 0.63	0.6 ± 0.5	0.16 ± 0.03	0.01 ± 0.01	0.1 ± 0.1	0.1 ± 0.1
ZZ	135 ± 20	14.1 ± 3.2	4.7 ± 0.8	4.0 ± 0.6	4.1 ± 0.7	3.1 ± 0.5
t̄tW	0.8 ± 0.2	0.4 ± 0.1	0.5 ± 0.1	0.7 ± 0.2	4.3 ± 0.6	3.9 ± 0.6
t̄tZ	14.7 ± 2.2	5.6 ± 0.8	4.5 ± 0.7	6.5 ± 1.1	25.4 ± 4.0	21.9 ± 3.4
rare Top	0.14 ± 0.04	0.07 ± 0.03	0.03 ± 0.02	0.09 ± 0.03	0.37 ± 0.07	0.6 ± 0.1
t̄tWW	0.04 ± 0.03	0.02 ± 0.02	0.01 ± 0.01	0.01 ± 0.00	0.03 ± 0.03	0.01 ± 0.01
Z + jets	110.0 ± 22.9	9.6 ± 2.0	2.1 ± 0.50	1.6 ± 0.4	5.1 ± 1.1	1.5 ± 0.3
W + jets	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
V + γ	25 ± 18	0.5 ± 0.2	0.1 ± 0.1	0.1 ± 0.1	0.0 ± 0.02	0.05 ± 0.07
tZ	15.9 ± 2.9	6.9 ± 1.3	5.1 ± 1.0	8.0 ± 1.5	18.7 ± 3.2	36.4 ± 6.1
tW	0.9 ± 0.7	0.2 ± 0.3	0.0 ± 0.1	0.0 ± 0.0	0.8 ± 0.6	0.2 ± 0.2
WtZ	4.9 ± 2.5	1.5 ± 0.8	1.1 ± 0.6	1.3 ± 0.7	4.6 ± 2.4	3.3 ± 1.7
VVV	7.4 ± 0.3	1.0 ± 0.1	0.4 ± 0.1	0.2 ± 0.1	0.13 ± 0.03	0.04 ± 0.01
VH	19.5 ± 4.2	2.8 ± 1.6	0.7 ± 0.7	0.1 ± 0.2	0.0 ± 0.0	0.0 ± 0.0
t̄t	0.7 ± 0.4	0.1 ± 0.1	0.05 ± 0.06	0.15 ± 0.13	0.8 ± 0.5	2.3 ± 1.2
t̄t̄	6.8 ± 1.0	2.4 ± 0.5	1.8 ± 0.4	3.3 ± 0.6	8.4 ± 1.2	13.6 ± 1.7
t̄tH	0.4 ± 0.1	0.2 ± 0.1	0.16 ± 0.02	0.23 ± 0.03	0.94 ± 0.11	1.03 ± 0.12
Total	2580 ± 160	229 ± 24	89 ± 13	69 ± 11	120 ± 15	108 ± 11

Table 23: Pre-fit yields in each of the 2-jet fit regions.

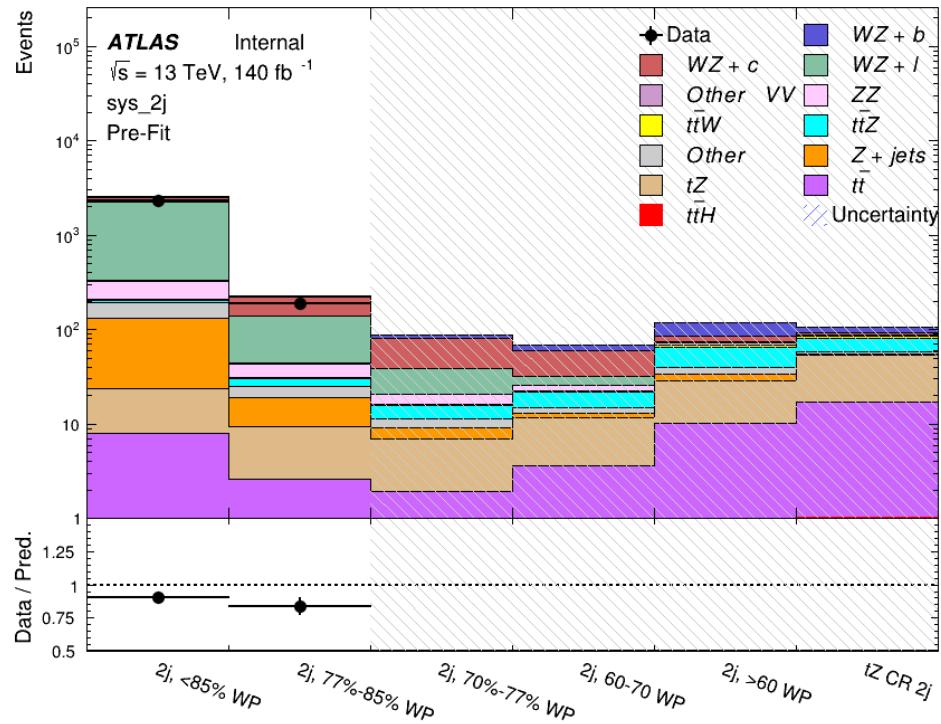


Figure 16.6: Pre-fit summary of the 2-jet fit regions.

1043

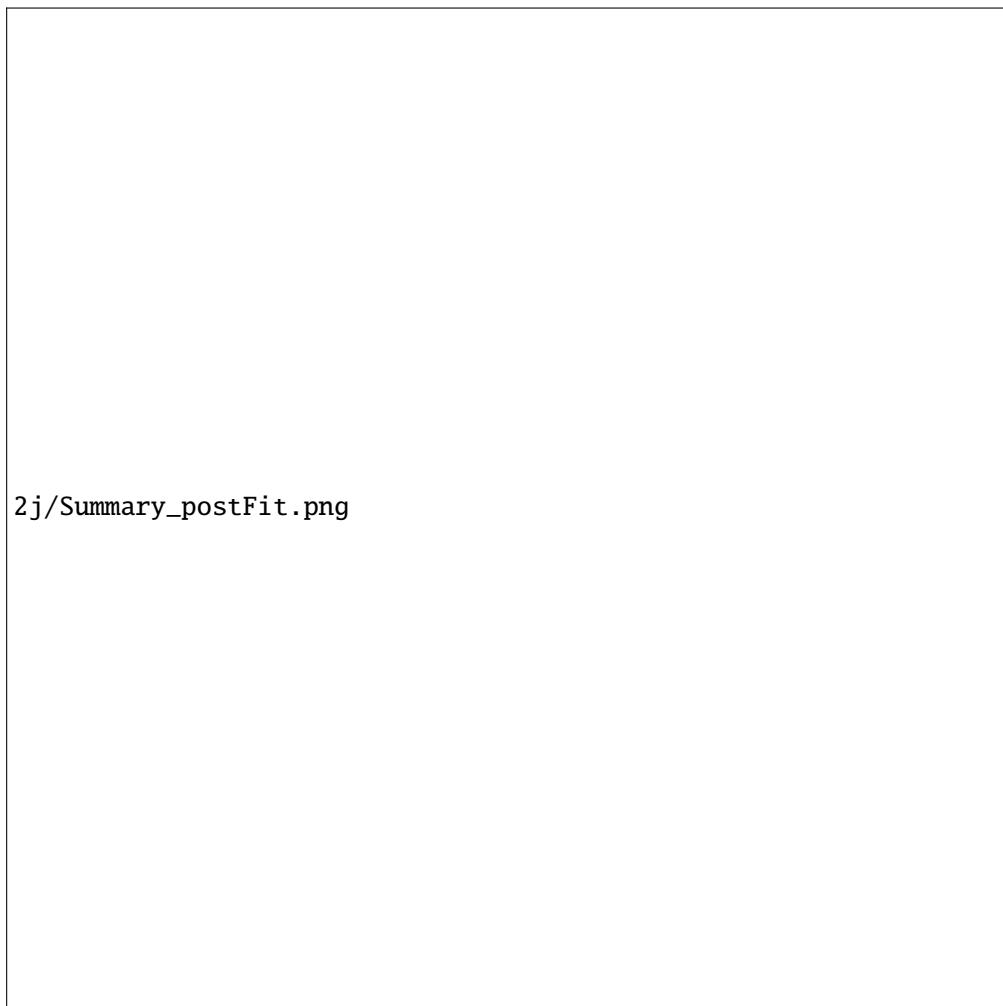
The post-fit yields in each region are summarized in figure 16.2.

	2j, <85% WP	2j, 77%-85% WP	2j, 70%-77% WP	2j, 60%-70% WP	2j, >60% WP	tZ CR 2j
WZ + b	13 ± 6	6.7 ± 2.9	5.8 ± 2.5	8.0 ± 3.5	31 ± 13	14 ± 5
WZ + c	260 ± 60	77 ± 15	41 ± 8	26 ± 5	10.9 ± 2.4	4.8 ± 1.1
WZ + l	1860 ± 90	90 ± 12	17.6 ± 2.8	5.8 ± 1.3	1.4 ± 0.4	0.3 ± 0.2
Other VV	7.6 ± 0.6	0.3 ± 0.3	0.3 ± 0.1	0.1 ± 0.06	0.03 ± 0.02	0.1 ± 0.1
ZZ	145 ± 30	11.3 ± 4.4	2.7 ± 1.6	1.0 ± 0.3	4.0 ± 0.1	2.4 ± 0.1
t̄tW	0.8 ± 0.2	0.4 ± 0.1	0.54 ± 0.12	0.74 ± 0.15	4.3 ± 0.6	3.9 ± 0.6
t̄tZ	14.7 ± 2.2	5.6 ± 0.8	4.5 ± 0.7	6.5 ± 1.0	25.4 ± 3.9	21.9 ± 3.3
rare Top	0.14 ± 0.04	0.07 ± 0.03	0.03 ± 0.02	0.09 ± 0.03	0.4 ± 0.1	0.6 ± 0.1
t̄tWW	0.04 ± 0.03	0.02 ± 0.02	0.01 ± 0.01	0.01 ± 0.01	0.03 ± 0.03	0.01 ± 0.01
Z + jets	110 ± 23	9.6 ± 2.0	2.1 ± 0.5	1.6 ± 0.4	5.1 ± 1.1	1.5 ± 0.3
W + jets	0.0 ± 0.0					
V + γ	25 ± 19	0.5 ± 0.2	0.1 ± 0.1	0.13 ± 0.14	0.0 ± 0.02	0.05 ± 0.07
tZ	15.9 ± 2.7	6.9 ± 1.2	5.1 ± 0.9	8.0 ± 1.4	18.7 ± 3.0	36 ± 6
tW	0.1 ± 0.7	0.2 ± 0.3	0.0 ± 0.1	0.0 ± 0.0	0.8 ± 0.6	0.2 ± 0.2
WtZ	4.9 ± 2.5	1.5 ± 0.8	1.1 ± 0.6	1.3 ± 0.7	4.6 ± 2.3	3.3 ± 1.7
VVV	7.4 ± 0.3	1.0 ± 0.1	0.36 ± 0.03	0.19 ± 0.03	0.13 ± 0.03	0.04 ± 0.01
VH	19 ± 4	2.8 ± 1.6	0.7 ± 0.7	0.1 ± 0.2	0.0 ± 0.0	0.0 ± 0.0
t̄t	6.8 ± 1.0	2.4 ± 0.5	1.8 ± 0.4	3.3 ± 0.6	8.4 ± 1.2	13.6 ± 1.7
t̄tH	0.40 ± 0.05	0.19 ± 0.03	0.16 ± 0.02	0.23 ± 0.03	0.94 ± 0.11	1.03 ± 0.11
Total	2580 ± 60	229 ± 11	89 ± 6	69.1 ± 4.1	120 ± 10	108 ± 6

Table 24: Post-fit yields in each of the 2-jet fit regions.

¹⁰⁴⁴

A post-fit summary of the fitted regions is shown in figure 16.7:



2j/Summary_postFit.png

Figure 16.7: Post-fit summary of the fit over 2-jet regions.

1045 The same set of systematic uncertainties consider for the 1-jet fit are included in the 2-jet
1046 fit as well. The impact of the most significant systematic uncertainties is summarized in table
1047 [25](#).

Uncertainty Source	$\Delta\mu$	
WZ + charm cross-section	-0.1966	0.2171
tZ cross-section	-0.1521	0.1518
WZ + light cross-section	0.1485	-0.1411
Other VV + b cross-section	-0.1115	0.1163
Flavor Tagging	0.0955	0.0957
Jet Energy Scale	0.0613	0.081
t <bar>t</bar>	-0.0662	0.0654
Luminosity	-0.0609	0.0655
Z + jets cross-section	-0.0284	0.0284
Other VV + charm cross-section	0.0207	-0.0202
Muon Trigger Scale Factor	0.019	0.0209
Total Systematic Uncertainty	0.3511	0.3679

Table 25: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b 2-jet events.

¹⁰⁴⁸ The ranking and impact of those nuisance parameters with the largest contribution to the
¹⁰⁴⁹ overall uncertainty is shown in figure 16.8.

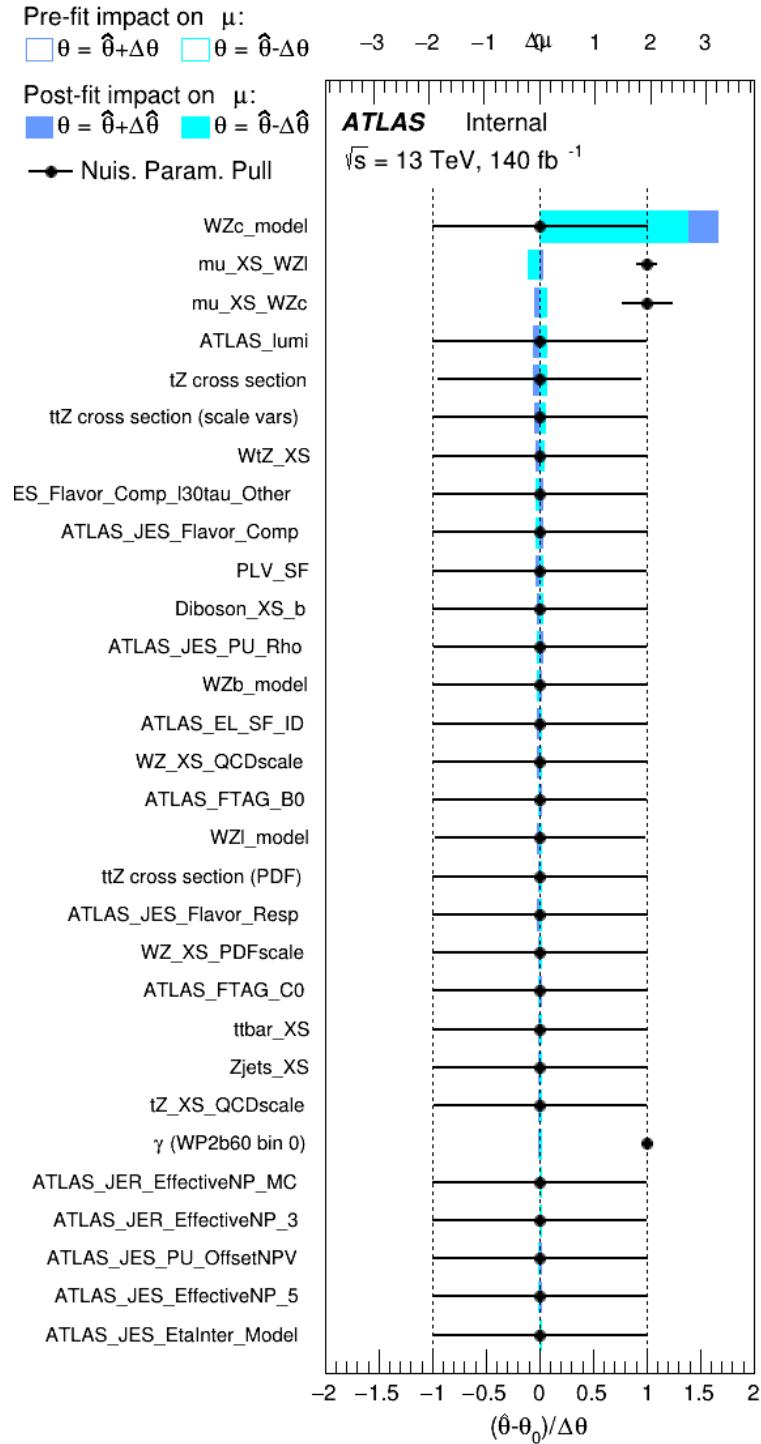


Figure 16.8: Impact of systematic uncertainties on the signal-strength of $WZ + b$ in 2-jet events.

1050 The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the
1051 shape of the fit regions depends heavily on the modeling of the jets. The other major sources of
1052 uncertainty come from background modelling and cross-section uncertainty. The pie charts in
1053 figure 16.9 show that for the modelling uncertainties that contribute most correspond to the most
1054 significant backgrounds.

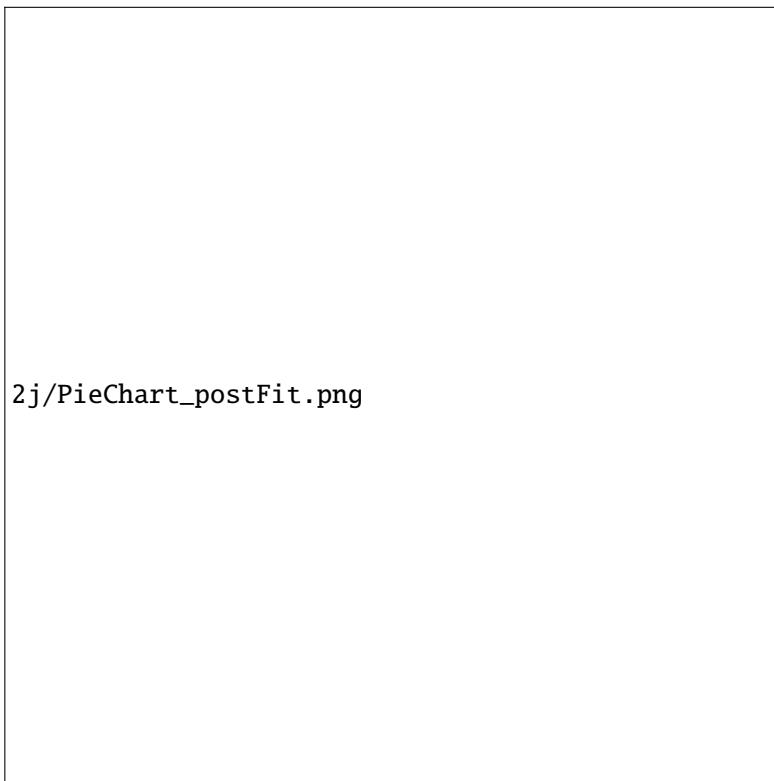


Figure 16.9: Post-fit background composition of the 2-jet fit regions.

1055 The correlations between these nuisance parameters are summarized in figure 16.10.

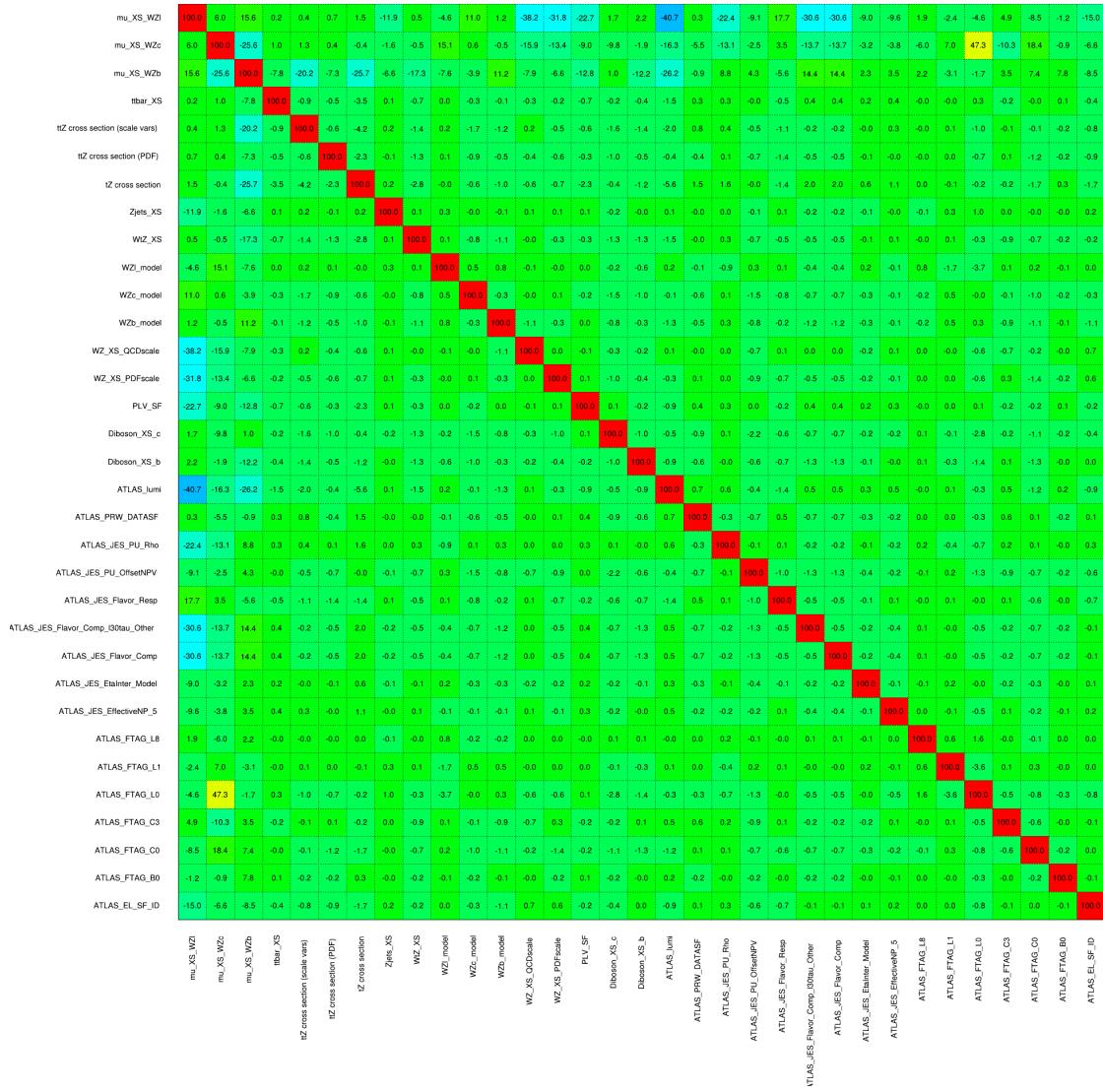


Figure 16.10: Correlations between nuisance parameters in the 2-jet fit

As in the 1-jet case, no significant, unexpected correlations are found between nuisance parameters.

1058 Part V**1059 Differential Studies of $t\bar{t}H$ Multilepton****1060 17 Data and Monte Carlo Samples**

1061 For both data and Monte Carlo (MC) simulations, samples were prepared in the `xAOD` format,
1062 which was used to produce a `xAOD` based on the `HIGG8D1` derivation framework. This framework
1063 was designed for the main $t\bar{t}H$ multi-lepton analysis. Because this analysis targets events with
1064 multiple light leptons, as well as tau hadrons, this framework skims the dataset of any events that
1065 do not meet at least one of the following requirements:

- 1066 • at least two light leptons within a range $|\eta| < 2.6$, with leading lepton $p_T > 15$ GeV and
1067 subleading lepton $p_T > 5$ GeV
- 1068 • at least one light lepton with $p_T > 15$ GeV within a range $|\eta| < 2.6$, and at least two hadronic
1069 taus with $p_T > 15$ GeV.

1070 Samples were then generated from these `HIGG8D1` derivations using a modified version of
1071 `AnalysisBase` version 21.2.127.

1072 17.1 Data Samples

1073 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through
 1074 2018, which represents an integrated luminosity of 139 fb^{-1} and an energy of $\sqrt{s} = 13 \text{ TeV}$. All
 1075 data used in this analysis was included in one of the following Good Run Lists:

- 1076 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
 _PHYs_StandardGRL_All_Good_25ns.xml
- 1078 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
 _PHYs_StandardGRL_All_Good_25ns.xml
- 1080 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYs_StandardGRL
 _All_Good_25ns_Triggerno17e33prim.xml
- 1082 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYs_StandardGRL
 _All_Good_25ns_Triggerno17e33prim.xml

1084 17.2 Monte Carlo Samples

1085 Several Monte Carlo (MC) generators were used to simulate both signal and background pro-
 1086 cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific
 1087 event generator used for each of these MC samples is listed in table 26.

Table 26: The configurations used for event generation of signal and background processes, including the event generator, matrix element (ME) order, parton shower algorithm, and parton distribution function (PDF).

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

18 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.

18.1 Trigger Requirements

Events are required to be selected by dilepton triggers, as summarized in table 27.

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

Table 27: List of lowest p_T -threshold, un-prescaled dilepton triggers used for 2015-2018 data taking.

18.2 Light Leptons

1093 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that
 1094 are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].
 1095
 1096 Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the
 1097 transition region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| <$
 1098 1.52, are rejected. A multivariate likelihood discriminant combining shower shape and track
 1099 information is used to distinguish prompt electrons from nonprompt leptons, such as those
 1100 originating from hadronic showers.

1101 To further reduce the non-prompt contribution, the track of each electron is required to

1102 originate from the primary vertex; requirements are imposed on the transverse impact parameter
1103 significance ($|d_0|/\sigma_{d_0}$) and the longitudinal impact parameter ($|\Delta z_0 \sin \theta_\ell|$).

1104 Muon candidates are reconstructed by combining inner detector tracks with track segments
1105 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
1106 $p_T > 10$ GeV and $|\eta| < 2.5$. All leptons are required to be isolated, and pass a non-prompt BDT
1107 selection described in detail in [**ttH_paper**].

1108 **18.3 Jets**

1109 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
1110 calorimeters [**ATL-PHYS-PUB-2015-015**], using the anti- k_t algorithm with a radius parameter
1111 $R = 0.4$. Jets with energy contributions likely arising from noise or detector effects are removed
1112 from consideration [**ATLAS-CONF-2015-029**], and only jets satisfying $p_T > 25$ GeV and
1113 $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
1114 association algorithm is used to confirm that the jet originates from the selected primary vertex,
1115 in order to reject jets arising from pileup collisions [**PERF-2014-03**].

1116 **18.4 Missing Transverse Energy**

1117 Because all $t\bar{t}H - ML$ channels considered include multiple neutrinos, missing transverse
1118 energy (E_T^{miss}) is present in each event. The missing transverse momentum vector is defined as
1119 the inverse of the sum of the transverse momenta of all reconstructed physics objects as well

1120 as remaining unclustered energy, the latter of which is estimated from low- p_T tracks associated
1121 with the primary vertex but not assigned to a hard object [**ATL-PHYS-PUB-2015-027**].

1122 **19 Higgs Momentum Reconstruction**

1123 Reconstructing the momentum of the Higgs boson is a particular challenge for channels with
1124 leptons in the final state: Because all channels include at least two neutrinos in the final state, the
1125 Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted
1126 by a neural network when provided with the four-vectors of the Higgs Boson decay products, as
1127 shown in section 19.1. With this in mind, several layers of MVAs are used to reconstruction the
1128 Higgs momentum.

1129 The first layer is a model designed to select which jets are most likely to be the b-jets that
1130 came from the top decay, detailed in section 19.2. As described in section 19.3, the kinematics
1131 of these jets are fed into the second layer, which is designed to identify the decay products of
1132 the Higgs Boson itself. The kinematics of these particles are then fed into yet another neural-
1133 network, which predicts the momentum of the Higgs (19.4). MVAs are also used in the analysis
1134 to determine the decay of the Higgs boson in the 3l channel (19.5).

1135 Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_AMC.
1136 Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training.

1137 For all of these models, the Keras neural network framework, with Tensorflow as the
1138 backend, is used, and the number of hidden layers and nodes are determined using grid search
1139 optimization. Each neural network uses the LeakyReLU activation function, a learning rate
1140 of 0.01, and the Adam optimization algorithm, as alternatives are found to either decrease or
1141 have no impact on performance. Batch normalization is applied after each layer. For the
1142 classification algorithms (b-jet matching, Higgs reconstruction, and 3l decay identification)
1143 binary-cross entropy is used as the loss function, while the p_T reconstruction algorithm uses
1144 MSE.

1145 The specific inputs features used for each model are arrived at through a process of trial
1146 and error - features considered potentially useful are tried, and those that are found to increase
1147 performance are included. While each model includes a relatively large number of features,
1148 some using upwards of 30, this inclusive approach is found to maximize the performance of each
1149 model while decreasing the variance compared to a reduced number of inputs. Each input feature
1150 is validated by comparing MC simulations to 80 fb^{-1} of data, as shown in the sections below.

1151 **19.1 Decay Candidate Reconstruction**

1152 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson
1153 using MC simulations of $t\bar{t}H$ events. These include light leptons and jets. Reconstructed
1154 physics objects are matched to truth level particles, in order to identify the parents of these

1155 reconstructed objects. The kinematics of the decay product candidates as well as event level
1156 variables are used as inputs.

1157 Leptons considered as possible Higgs and top decay candidates are required to pass the
1158 selection described in section 18.2. For jets, however, it is found that a large fraction that originate
1159 from either the top decay or the Higgs decay fall outside the selection described in section 18.3.
1160 Specifically, jets from the Higgs decay tend to be soft, with 32% having $p_T < 25$ GeV. Therefore
1161 jets with $p_T < 15$ GeV are considered as possible candidates in the models described below. By
1162 contrast, less than 5% of the jets originating from the Higgs fall below this p_T . The jets are found
1163 to be well modeled even down to this low p_T threshold, as shown in section 21.1. The impact of
1164 using different p_T selection for the jet candidates is considered in detail in section A.5. As they
1165 are expected to originate from the primary vertex, jets are also required to pass a JVT cut.

1166 **19.2 b-jet Identification**

1167 Including the kinematics of the b-jets that originate from the top decay is found to improve the
1168 identification of the Higgs decay products, and improve the accuracy with which the Higgs
1169 momentum can be reconstructed. Because these b-jets are reconstructed by the detector with
1170 high efficiency (just over 90% of the time), and can be identified relatively consistently, the first
1171 step in reconstructing the Higgs is selecting the b-jets from the top decay.

1172 Exactly two b-jets are expected in the final state of $t\bar{t}H - \text{ML}$ events. However, in both
1173 the 3l and 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r b-tag

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

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

19.2.1 2lSS Channel

For the 2lSS channel, the input features shown in table 28 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 . $jetDL1r$ is an integer corresponding to the calibrated b-tagging working points reached by each jet, where 5 represents the tightest working point and 1 represents the loosest. The variables $nJets$, $DL1r\ 60\%$ and $nJets\ DL1r\ 85\%$ represent the number of jets in the event passing the 60% and 85% b-tag working points, respectively.

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

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

Table 28: Input features used in the b-jet identification algorithm for the 2lSS channel

1192 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for
 1193 training.

1194 The difference between the distributions for a few of these features for the correct(i.e.
 1195 both jets are truth b-jets), and incorrectcombinations are shown in figure 19.1. The correct and
 1196 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences
 1197 in the distributions.

1198 The modeling of these inputs is validated against data, with figure 19.2 showing good
 1199 general agreement between data and MC. Plots for the complete list of features can found in
 1200 section A.

1201 Based on the results of grid search evaluation, the optimal architecture is found to include
 1202 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as
 1203 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for

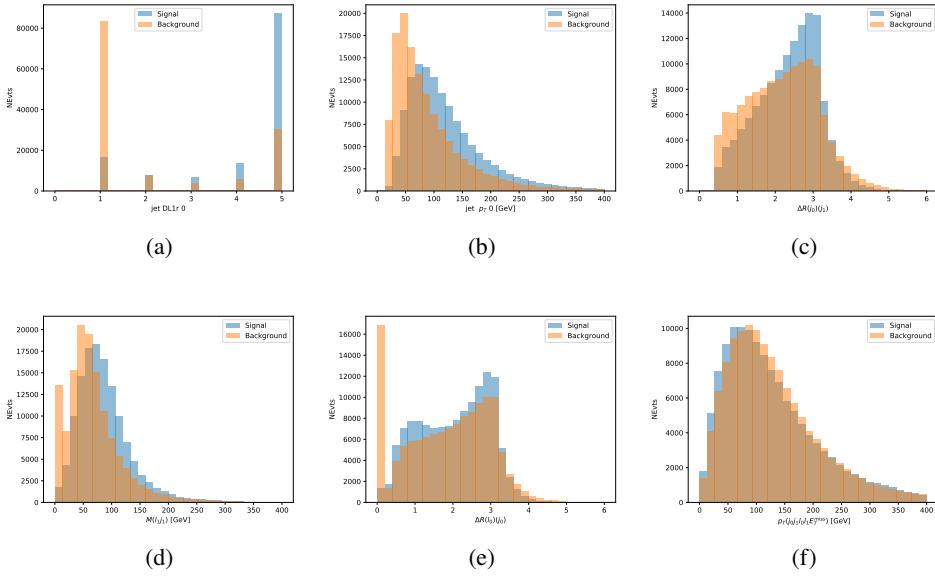


Figure 19.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

1204 the trained model are shown in figure 19.2.1. The model is found to identify the correct pairing
 1205 of jets for 73% of 2ISS signal events on test data.

1206 For point of comparison, a naïve approach to identify b-jets is used as well: The two jets
 1207 which pass the highest DL1r b-tag working point are assumed to be the b-jets from the top decay.
 1208 In the case that multiple jets meet the same b-tag working point, the jet with higher p_T is used.
 1209 This method identifies the correct jet pair 65% of the time.

1210 The accuracy of the model for different values of n-bjets, compared to this naive approach,
 1211 is shown in table 29.

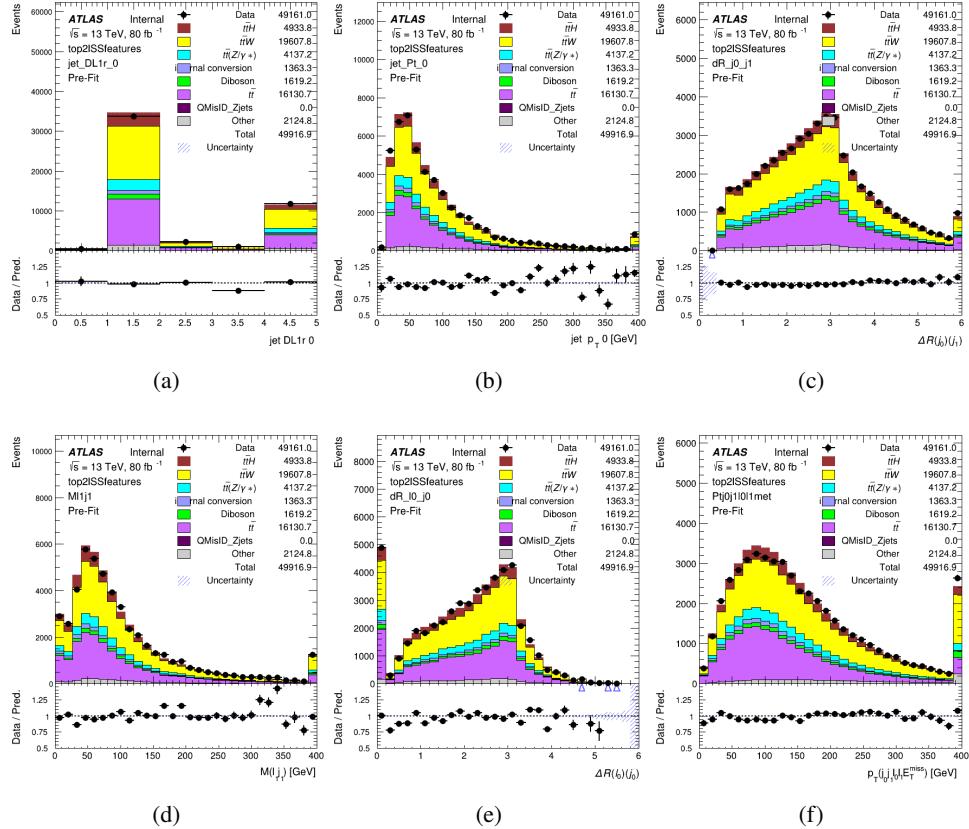


Figure 19.2: Data/MC comparisons of input features for top2ISS training for 80 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

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

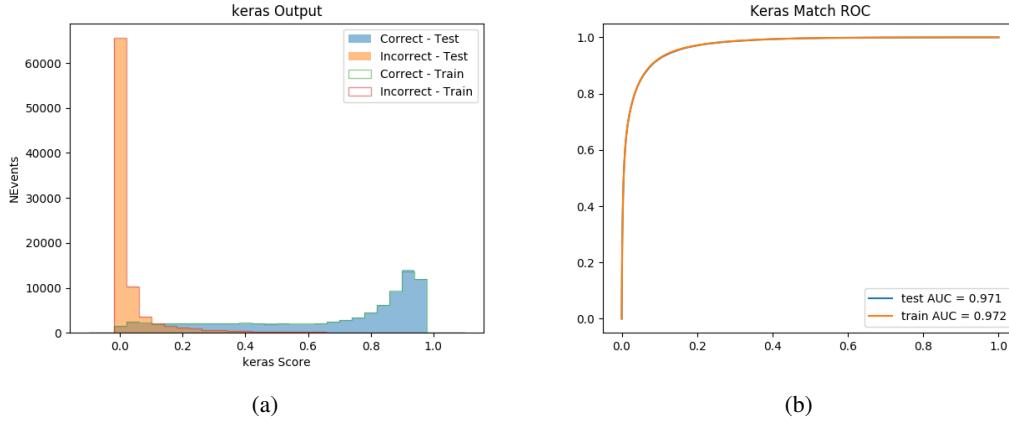


Figure 19.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

19.2.2 3l Channel

1212 The input features used in the 3l channel are listed in table 30, with the same naming convention
 1213 as the 2lSS channel.

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

Table 30: Input features for the b-jet identification algorithm in the 3l channel.

1215 A few of these features are shown in figure 19.4, comparing the distributions for correct

1216 and incorrect combinations of jets.

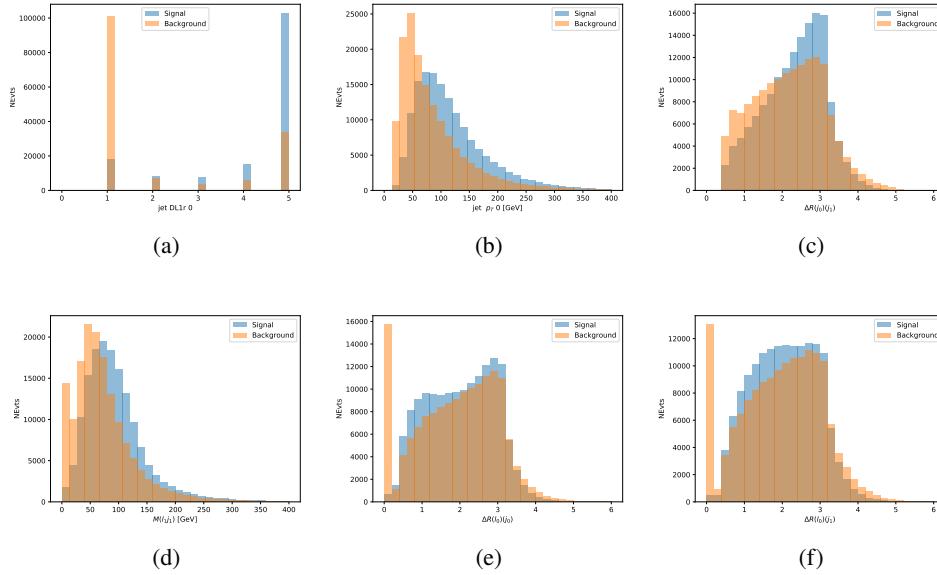


Figure 19.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and 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

1217 The modeling of these inputs is validated against data, with figure 19.5 showing good
1218 general agreement between data and MC. Plots for the complete list of features can found in
1219 section A.

1220 Again, the dataset is downsized to reduce the ratio of correct and incorrect combination
1221 from 20:1, to 5:1. Around 7 million events are used for training, with 10% set aside for testing.
1222 Based on the results of grid search evaluation, the optimal architecture is found to include 5
1223 hidden layers with 60 nodes each. The output score distribution as well as the ROC curve for the
1224 trained model are shown in figure 19.2.2.

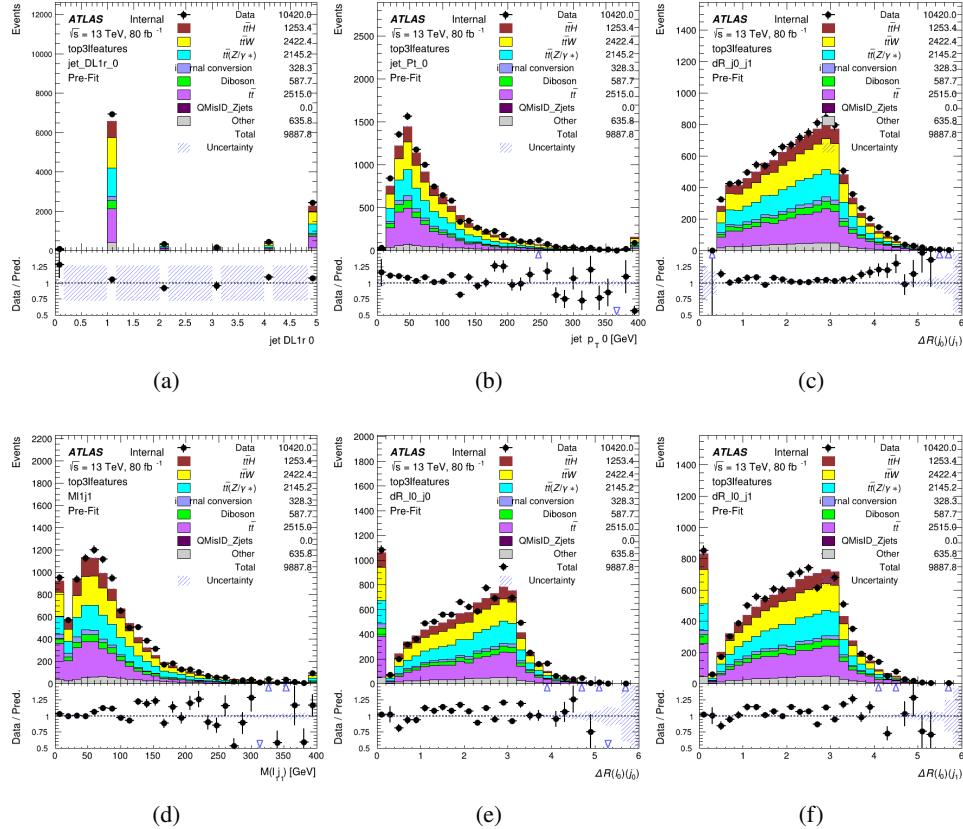


Figure 19.5: Data/MC comparisons of input features for top3l training for 80 fb^{-1} of data. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

1225 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal
1226 events. The accuracy of the model is summarized in table 31.

Table 31: Accuracy of the NN in identifying b-jets from tops, compared to the naive method of taking the highest b-tagged jets.

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

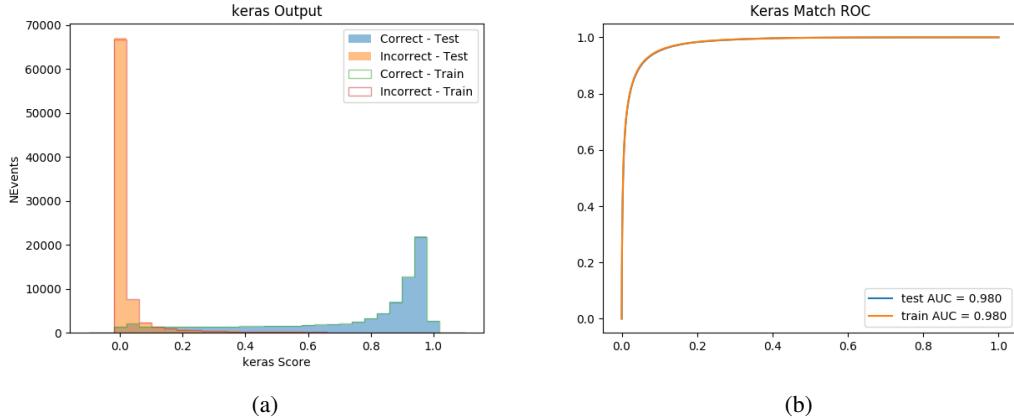


Figure 19.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

19.3 Higgs Reconstruction

1227 Techniques similar to the b-jet identification algorithms are employed to select the decay products
 1228 of the Higgs: kinematics of all possible combinations of reconstructed objects are fed into a neural
 1229 network to determine which of those is most mostly to be the decay products of the Higgs.

1231 Again separate models are used for the 2lSS and 3l channels, while the 3l channel has now
 1232 been split into two: $t\bar{t}H$ events with three leptons in the final state include both instances where
 1233 the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the Higgs
 1234 decays to two leptons.

1235 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In
 1236 the semi-leptonic case the reconstructed decay products consist of two jets and a single leptons.
 1237 For the fully-leptonic case, the decay products include 2 of the three leptons associated with the

1238 event. For training the models, events are separated into these two categories using truth level
1239 information. A separate MVA, described in section 19.5, is used to make this distinction at reco
1240 level and determine which model to use.

1241 For all channels, the models described in section 19.2 are used to identify b-jet candidates,
1242 whose kinematics are used to identify the Higgs decay products. These jets are not considered
1243 as possible candidates for the Higgs decay, justified by the fact that these models are found to
1244 misidentify jets from the Higgs decay as jets from the top decay less than 1% of the time.

1245 **19.3.1 2lSS Channel**

1246 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural
1247 network is trained on the kinematics of different combinations of leptons and jets, as well as the
1248 b-jets identified in section 19.2, with the specific input features listed in table 32.

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 η T	$\Delta R(j_0)(j_1)$
$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$	$M(j_0 j_1)$
$M(l_H j_0)$	$M(l_H j_1)$	$\Delta R(l_H)(b_0)$
$\Delta R(l_H)(b_1)$	$\Delta R(l_T)(b_0)$	$\Delta R(l_T)(b_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1)(l_T)$	$\Delta R(j_0 j_1)(b_0)$
$\Delta R(j_0 j_1)(b_1)$	$\Delta R(j_0)(b_0)$	$\Delta R(j_0)(b_1)$
$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$	$M(j_0 j_1 l_H)$
$p_T(j_0 j_1 l_H l_T b_0 b_1 E_T^{\text{miss}})$	topScore	E_T^{miss}
nJets	HT jets	

Table 32: Input features used to identify the Higgs decay products in 2ISS events

1249 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 1250 lepton in the event is labeled l_T , as it is assumed to have come from the decay of one of the top
 1251 quarks. b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet
 1252 Reco Score is the output of the b-jet reconstruction algorithm.

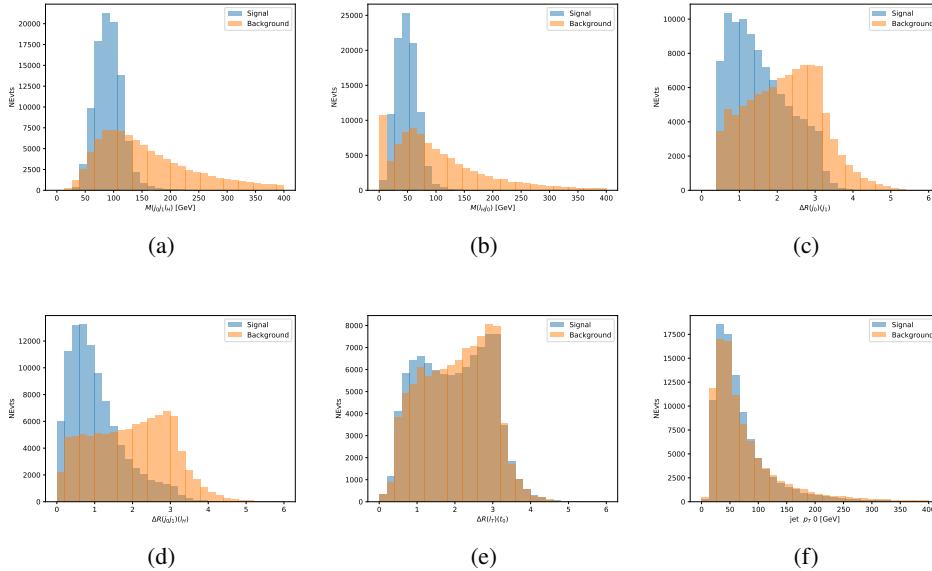


Figure 19.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (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.

1253 The modeling of these inputs is validated against data, with figure 19.2 showing good
 1254 general agreement between data and MC. Plots for the complete list of features can found in
 1255 section A.

1256 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2
 1257 million events, with an additional 200,000 reserved for testing the model. In order to compensate

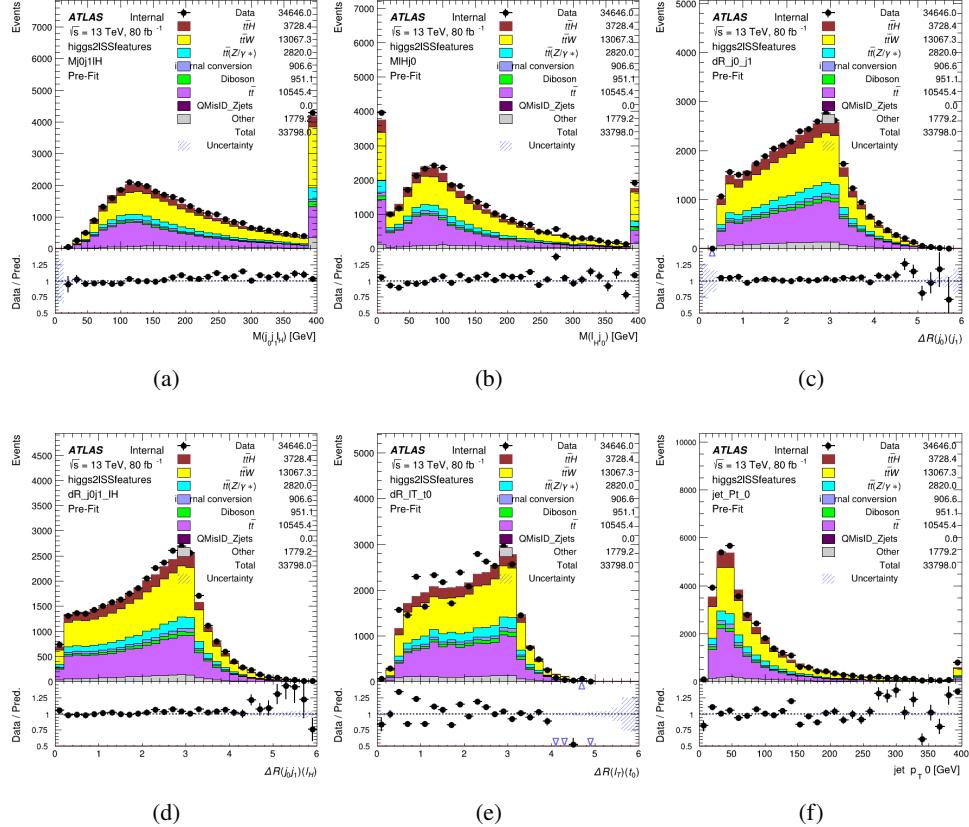


Figure 19.8: Data/MC comparisons of input features for higgs2lSS training for 80 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 and the lepton candidate, (e) ΔR between the lepton from the top and the leading b-jet, (f) the p_T of jet 0.

for 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 19.3.1.

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.

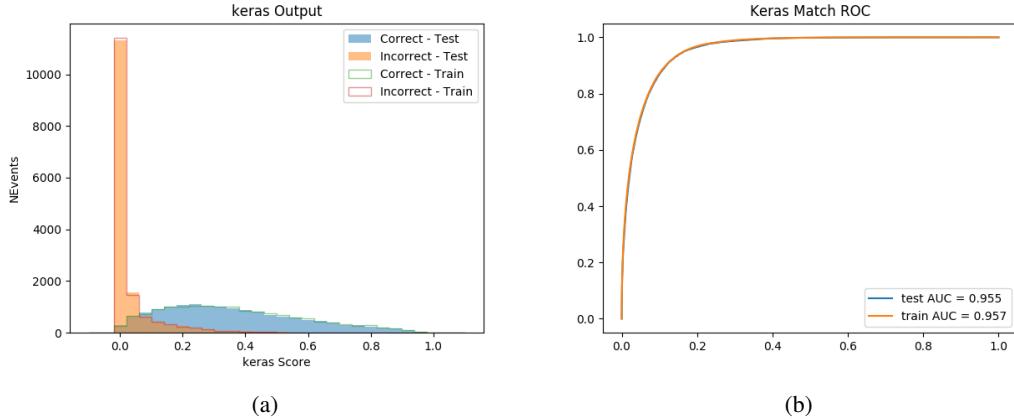


Figure 19.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

19.3.2 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 from the Higgs. This leaves 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	$\Delta R(j_0)(j_1)$
$M(j_0j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0j_1)(l_H)$	$\Delta R(j_0j_1)(l_{T_1})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0j_1l_{T_0})$	$M(j_0j_1l_{T_1})$
$M(j_0j_1l_H)$	$\Delta R(j_0j_1l_H)(l_{T_0})$	$\Delta R(j_0j_1l_H)(l_{T_1})$
$\Delta\phi(j_0j_1l_H)(E_T^{\text{miss}})$	$p_T(j_0j_1l_Hl_{T_0}l_{T_1}b_0b_1E_T^{\text{miss}})$	$M(j_0j_1b_0)$
$M(j_0j_1b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
topScore	MET	HT jets
nJets		

Table 33: Input features used to identify the Higgs decay products in 3lS events

1269 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 1270 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
 1271 the b-jet identification algorithm. The b-jet Reco Score is the output of the Higgs reconstruction
 1272 algorithm.

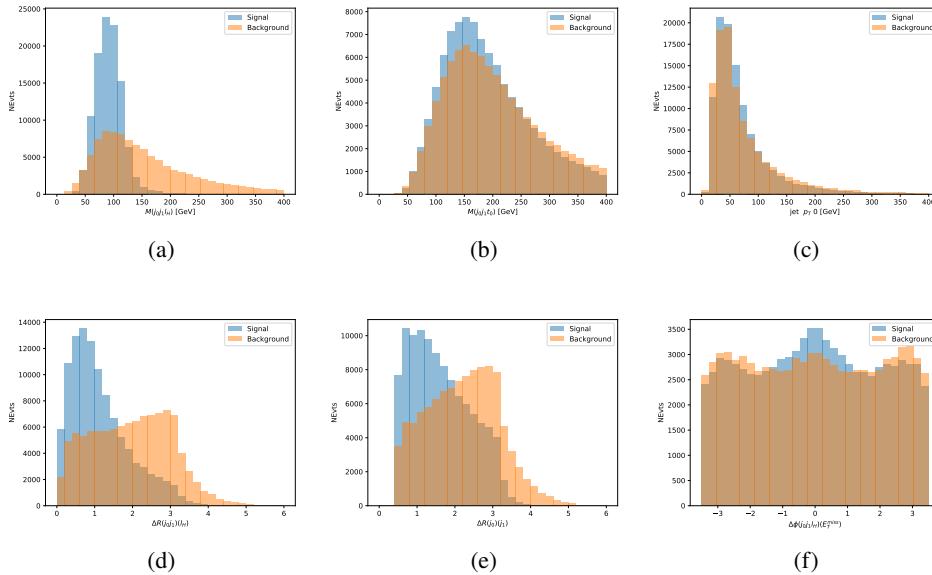


Figure 19.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1273 The modeling of these inputs is validated against data, with figure 19.11 showing good
 1274 general agreement between data and MC. Plots for the complete list of features can found in
 1275 appendix A.1.

1276 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.
 1277 Once again, incorrect combinations are downsampled, such that the correct combinations are
 1278 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN

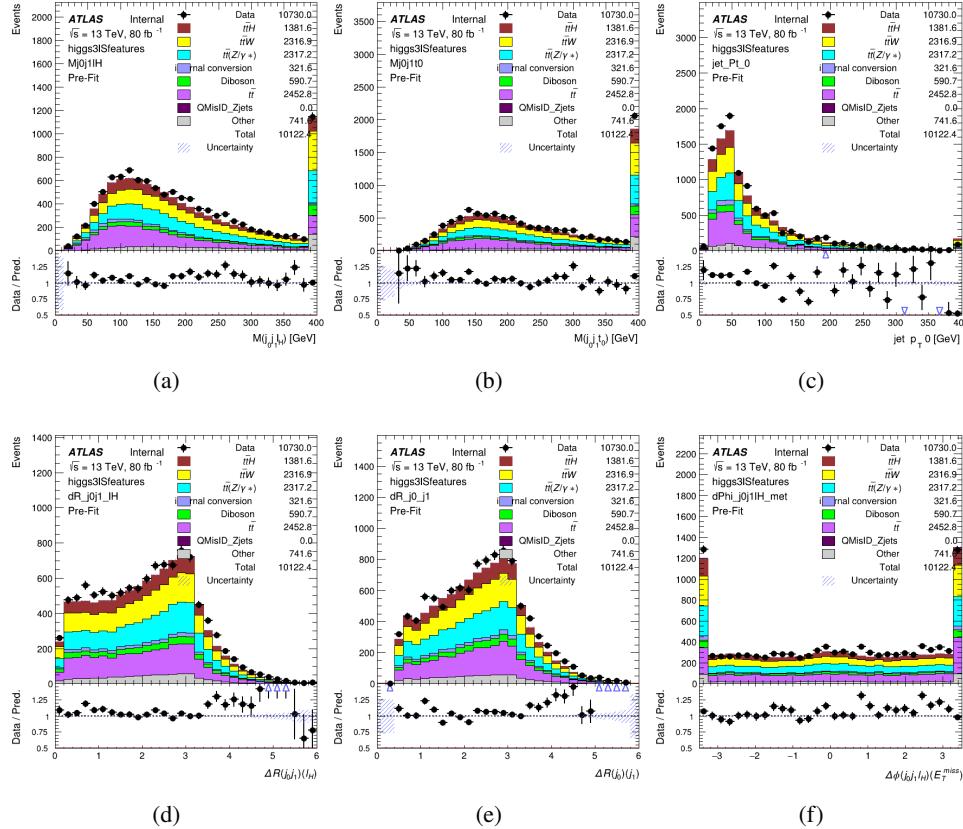


Figure 19.11: Data/MC comparisons of input features for higgs3IS training for 80 fb^{-1} of data.

1279 is summarized in figure 19.3.2.

1280 The neural network identifies the correct combination 64% of the time. It identifies the
1281 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
1282 83% of the time.

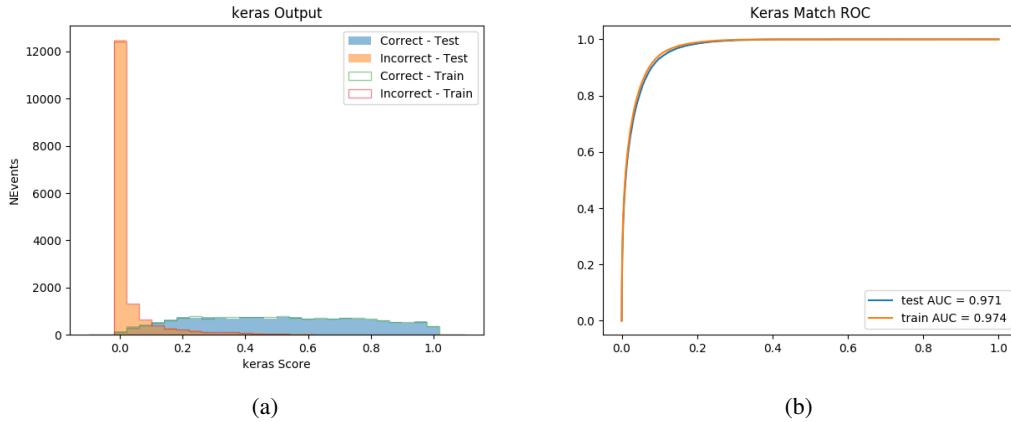


Figure 19.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1283 19.3.3 3l Fully-leptonic Channel

1284 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 1285 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to
 1286 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 1287 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 1288 and one where it is not.

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

Table 34: Input features used to identify the Higgs decay products in 3lF events

1289 Here l_{H0} and l_{H1} are the Higgs decay candidates. The other lepton in the event is labeled
 1290 l_T . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco
 1291 Score is the output of the Higgs reconstruction algorithm.

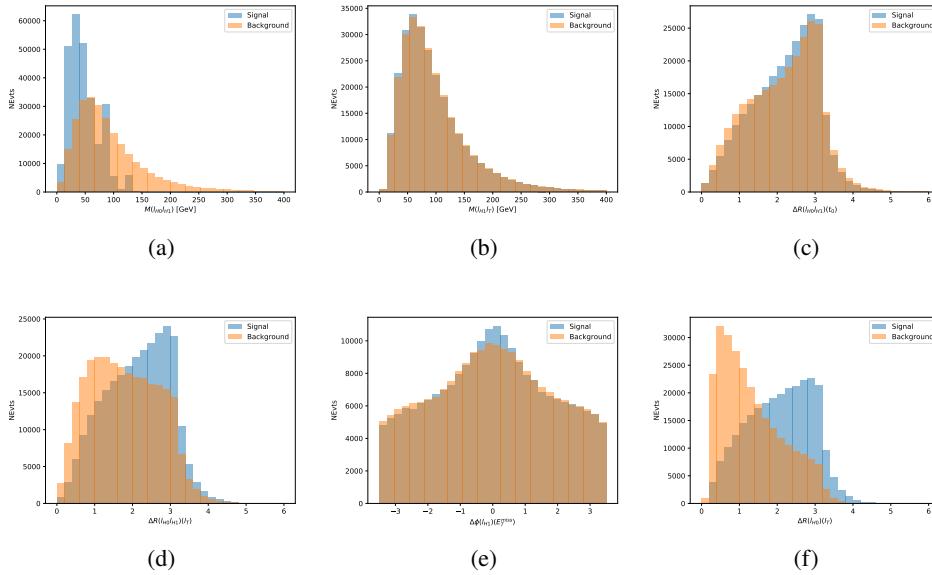


Figure 19.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1292 The modeling of these inputs is validated against data, with figure 19.14 showing good
 1293 general agreement between data and MC. Plots for the complete list of features can found in
 1294 section A.

1295 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,
 1296 with 10% of the dataset reserved for testing. The output of the model is summarized in figure
 1297 19.3.3.

1298 The correct lepton is identified by the model for 80% of events in the testing data set.

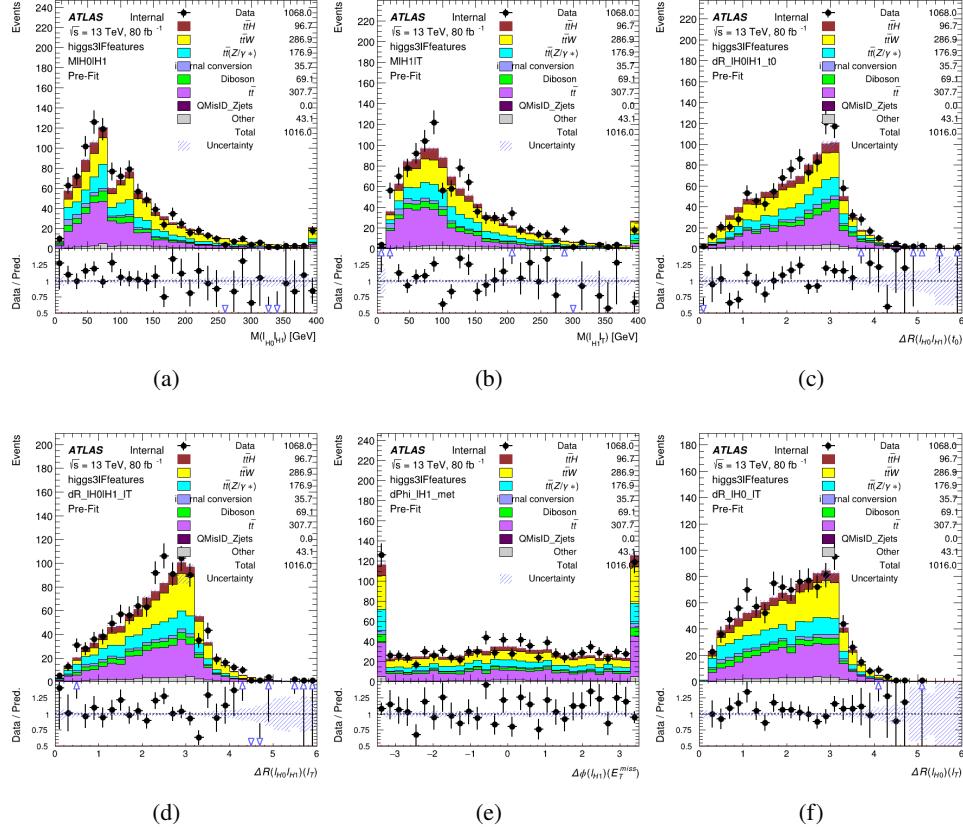


Figure 19.14: Data/MC comparisons of input features for higgs3IF training for 80 fb^{-1} of data.

19.4 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.

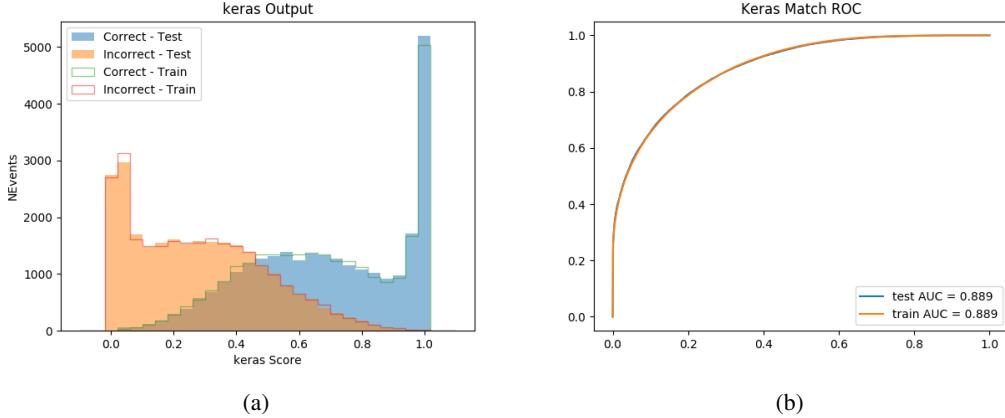


Figure 19.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

1306 As a two-bin fit is targeted for the final result, some metrics evaluating the performance
 1307 of the models aim to show how well it distinguished between "high p_T " and "low p_T " events. A
 1308 cutoff point of 150 GeV is used to define these two categories.

1309 Because the analysis uses a two bin fit of the Higgs p_T , the momentum reconstruction
 1310 could be treated as a binary classification problem, rather than a regression problem. This
 1311 approach is explored in detail in section A.4, and is found not to provide any significant increase
 1312 in sensitivity. The regression approach is used because it provides more flexibility for future
 1313 analyses, as it is independent of the cutoff between high and low p_T , as well as the number of
 1314 bins. Further, a regression allows the output of the neural network to be more clearly understood,
 1315 as it can be directly compared to a physics observable.

19.4.1 2lSS Channel

1316 **19.4.1 2lSS Channel**

1317 The input variables listed in table 35 are used to predict the Higgs p_T in the 2lSS channel. Here
1318 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
1319 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
1320 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
1321 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
1322 reconstruction algorithm, and the b-jet identification algorithm, respectively.

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

Table 35: Input features for reconstructing the Higgs p_T spectrum for 2lSS events

1323 The optimal neural network architecture for this channel is found to consist of 7 hidden
 1324 layers with 60 nodes each. The input data set includes 1.2 million events, 10% of which is used
 1325 for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1326 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 1327 truth Higgs p_T in figure 19.16. In order to visualize the model performance more clearly, in (a)
 1328 of that figure, the color of each point is determined by Kernel Density Estimation (KDE). The
 1329 color shown represents the logarithm of the output from KDE, to counteract the large number of
 1330 low p_T events. For that same reason, each column of the histogram shown in (b) of figure 19.16
 1331 is normalized to unity. This plot therefore demonstrates what the model predicts for each slice
 1332 of truth p_T .

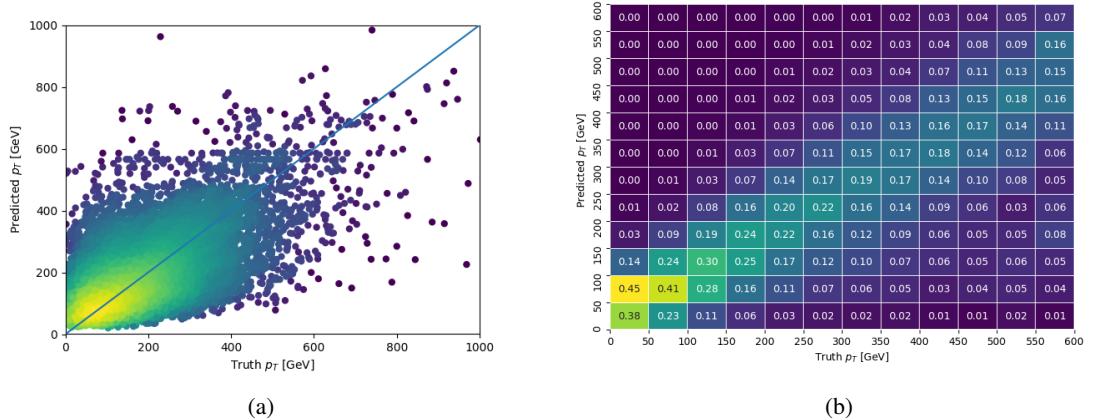


Figure 19.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1333 We are also interested in how well the model distinguishes between events with $p_T < 150$
 1334 GeV and > 150 GeV. Figure 19.17 demonstrates the NN output for high and low p_T events based

1335 on this cutoff.

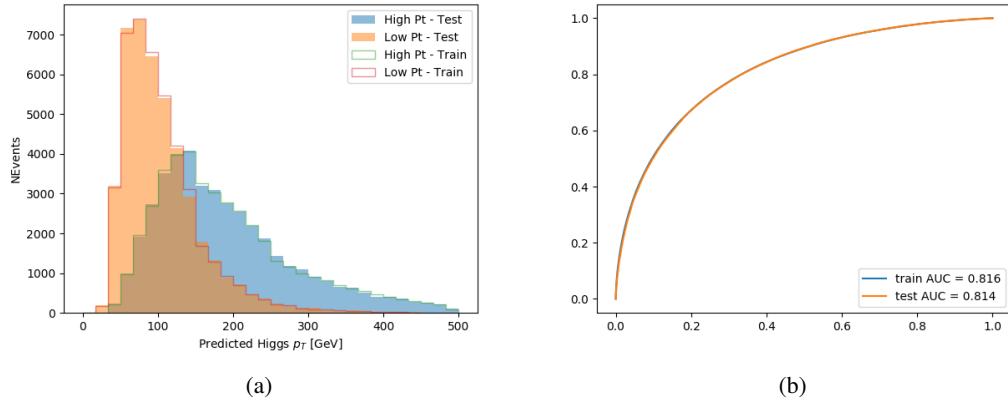


Figure 19.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

1336 **19.4.2 3l Semi-leptonic Channel**

1337 The following input features are used to predict the Higgs p_T for events in the 3lS channel:

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

Table 36: 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 19.18. Once again, (a) of 19.18 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 .

Figure 19.19 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.

19.4.3 3l Fully-leptonic Channel

The features listed in 37 are used to construct a model for predictin the Higgs p_T for 3lF events.

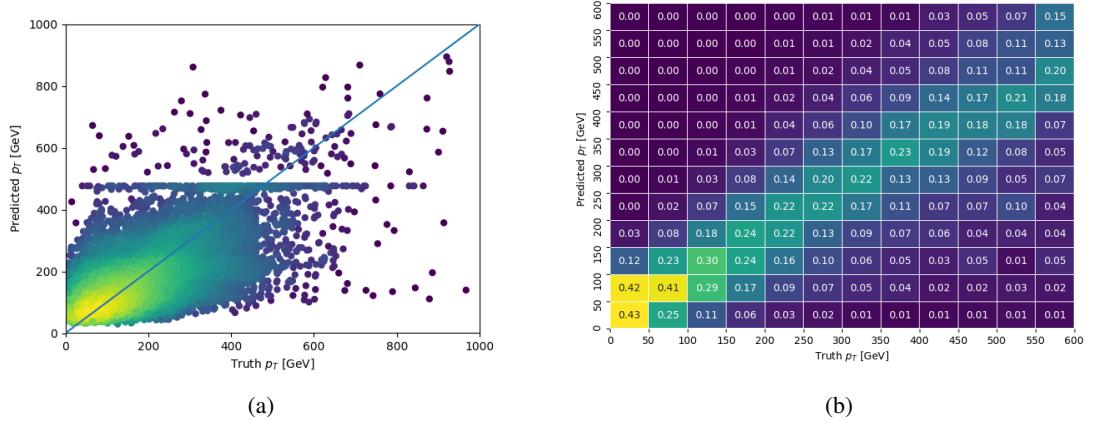


Figure 19.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

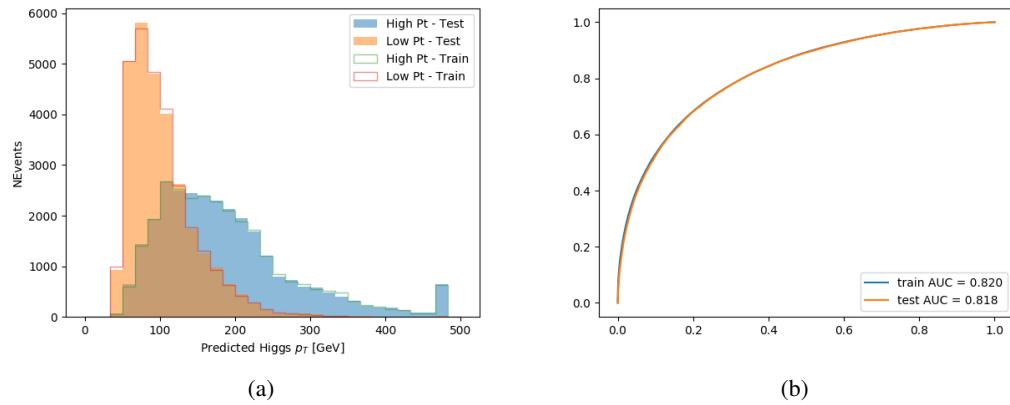


Figure 19.19: (a) shows the reconstructed Higgs p_T for 3lS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

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

Table 37: Input features for reconstructing the Higgs p_T spectrum for 3lF events

1356 l_{H0} and l_{H1} represent the two leptons identified by the Higgs reconstruction model as
 1357 originating from the Higgs, while l_T is the other lepton in the event. The Higgs Reco Score and
 1358 b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification
 1359 algorithm, respectively.

1360 The optimal neural network architecture for this channel is found to consist of 5 hidden
 1361 layers with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for
 1362 testing, the other 90% for training. Training is found to converge after around 150 epochs.

1363 The predicted transverse momentum, as a function of the truth p_T , is shown in figure
 1364 [19.20.](#)

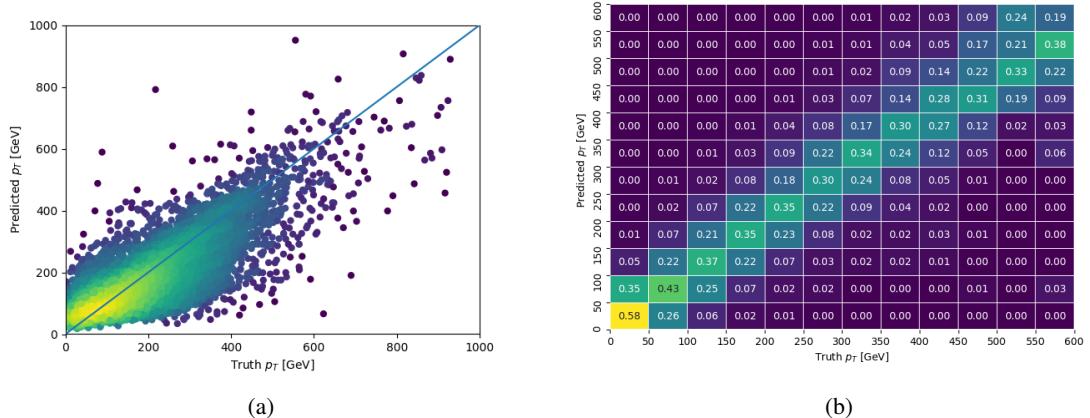


Figure 19.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1365 When split into high and low p_T , based on a cutoff of 150 GeV, the

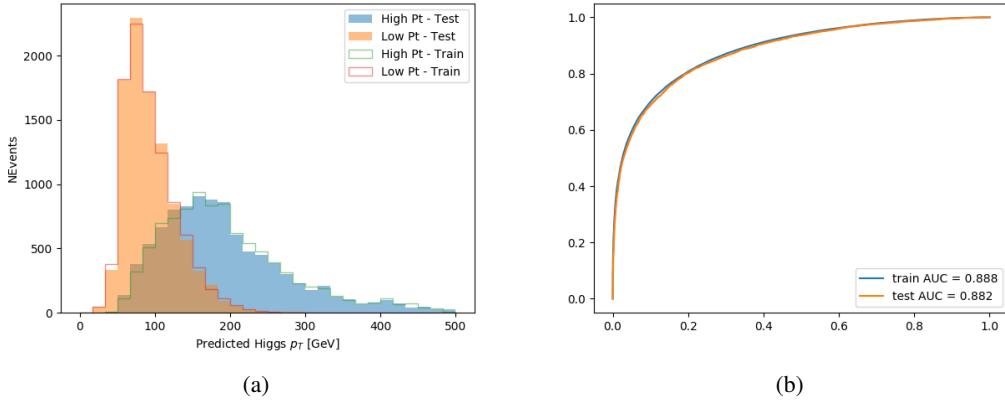


Figure 19.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

19.5 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 38.

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

Table 38: Input features used to distinguish semi-leptonic and fully-leptonic Higgs decays in the 3l channel.

1376 Here l_0 is the opposite charge lepton, l_1 and l_2 are the two SS leptons order by ΔR
 1377 from lepton 0. score3lF and score3lS are the outputs of the 3lS and 3lF Higgs reconstruction
 1378 algorithms, while topScore is the output of the b-jet identification algorithm.

1379 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these
 1380 two decay modes. The output of the model is summarized in figure 19.22.

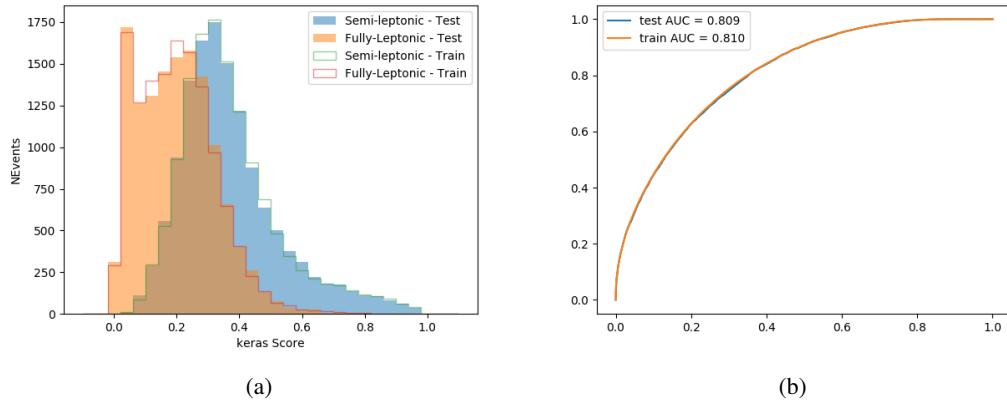


Figure 19.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

1381 A cutoff of 0.23 is determined to be optimal for separating 3lS and 3lF in the fit.

1382 20 Signal Region Definitions

1383 Events are divided into two channels based on the number of leptons in the final state: one with
 1384 two same-sign leptons, the other with three leptons. The 3l channel includes events where both
 1385 leptons originated from the Higgs boson as well as events where only one of the leptons

1386 20.1 Pre-MVA Event Selection

1387 A preselection is applied to define orthogonal analysis channels based on the number of leptons

1388 in each event. For the 2lSS channel, the following preselection is used:

- 1389
- Two very tight, same-charge, light leptons with $p_T > 20 \text{ GeV}$

- 1390
- ≥ 4 reconstructed jets, ≥ 1 b-tagged jets

- 1391
- No reconstructed tau candidates

1392 The event yield after the 2lSS preselection has been applied, for MC and data at 80 fb^{-1} ,

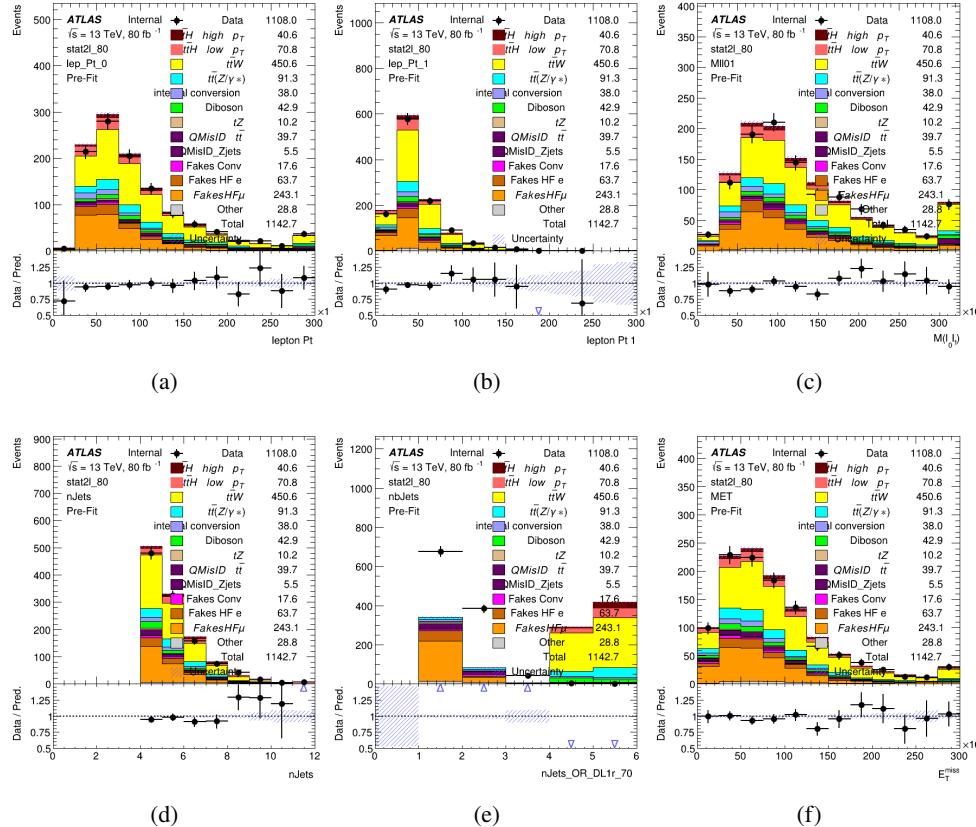
1393 is shown in table 39.

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

Table 39: Event yield in the 2ISS preselection region.

1394

figure 21.1. Good general agreement is found.

Figure 20.1: Data/MC comparisons of the 2lSS pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1395

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

1396

- 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

1398

- 1399 • ≥ 2 reconstructed jets, ≥ 1 b-tagged jets
- 1400 • No reconstructed tau candidates
- 1401 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs
- 1402 The event yield after the 3l preselection has been applied, for MC and data at 80 fb^{-1} , is
- 1403 shown in table 20.1.

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

Table 40: Yields of the analysis

1404

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

1405 21.2.

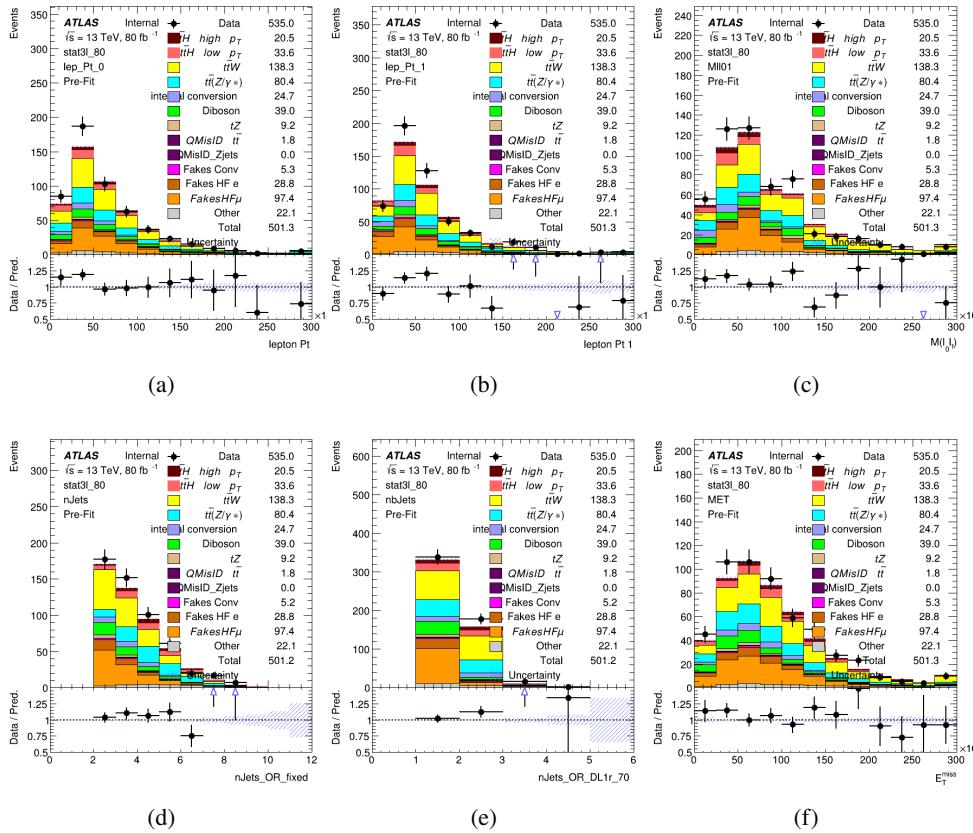


Figure 20.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1406

20.2 Event MVA

1407

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

1410 [xgboost] are trained using the kinematics of signal and background events derived from Monte
1411 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo
1412 event.

1413 Because the background composition differs for events with a high reconstructed Higgs p_T
1414 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and
1415 low p_T regions. This is found to provide better significance than attempting to build an inclusive
1416 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6
1417 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
1418 3lF high p_T , and 3lF low p_T .

1419 The following features are used in both the high and low p_T 2lSS BDTs:

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

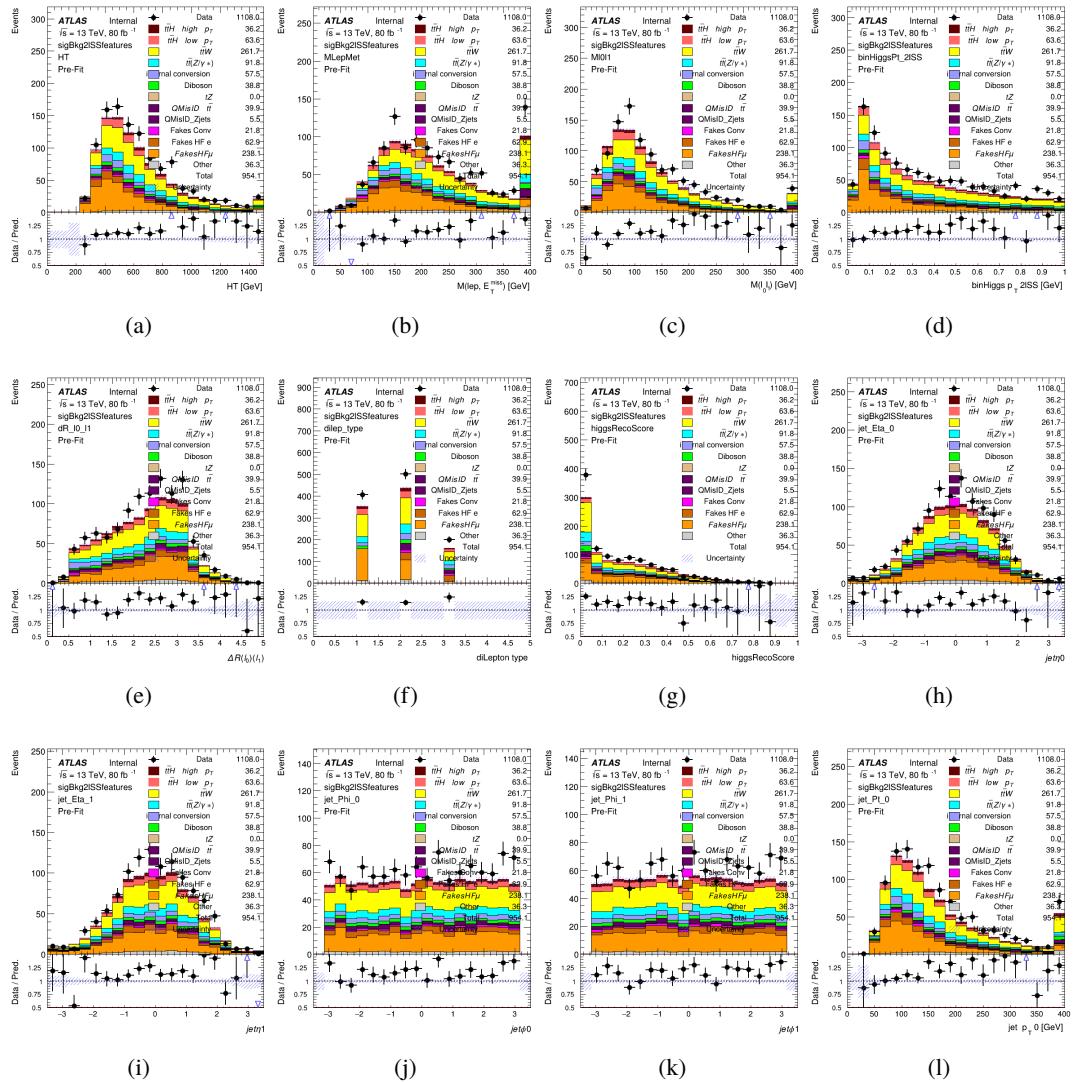
Table 41: Input features used to distinguish signal and background events in the 2ISS channel.

1420

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

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

Table 42: Input features used to distinguish signal and background events in the 3l channel.



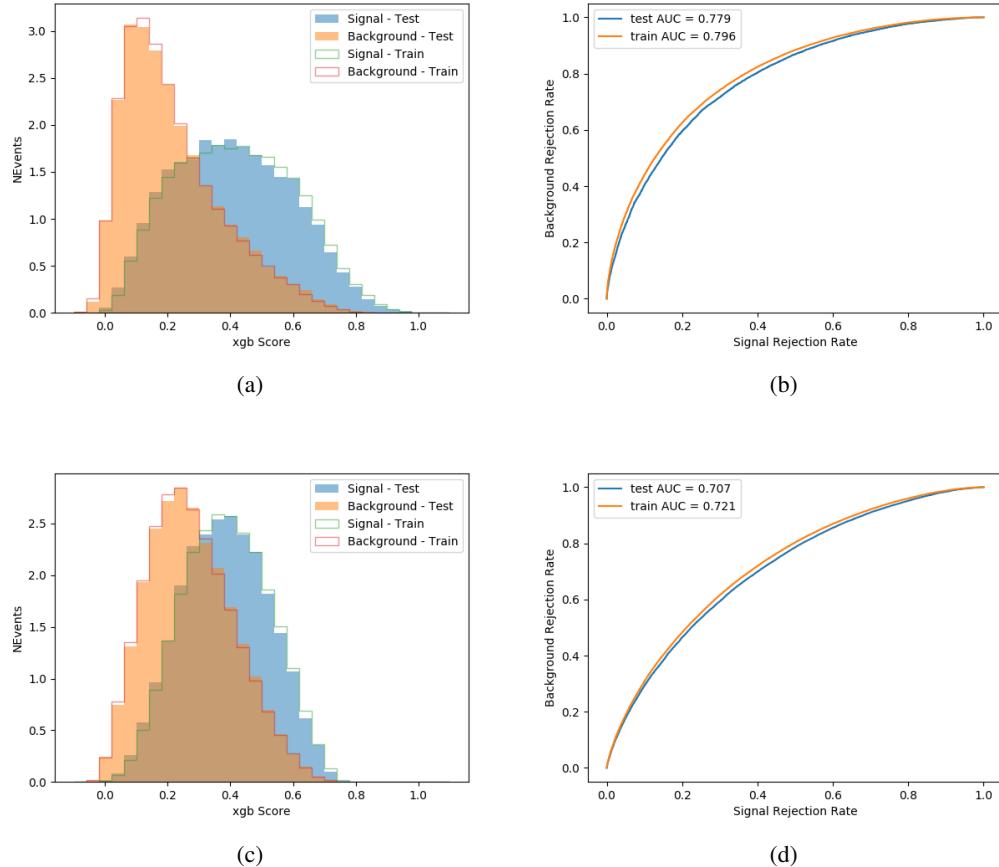


Figure 20.4:

20.3 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 19.5 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 - 2lSS, 3lS, and 3lF.

For each event, depending on the number of leptons as well as whether the p_T of the Higgs

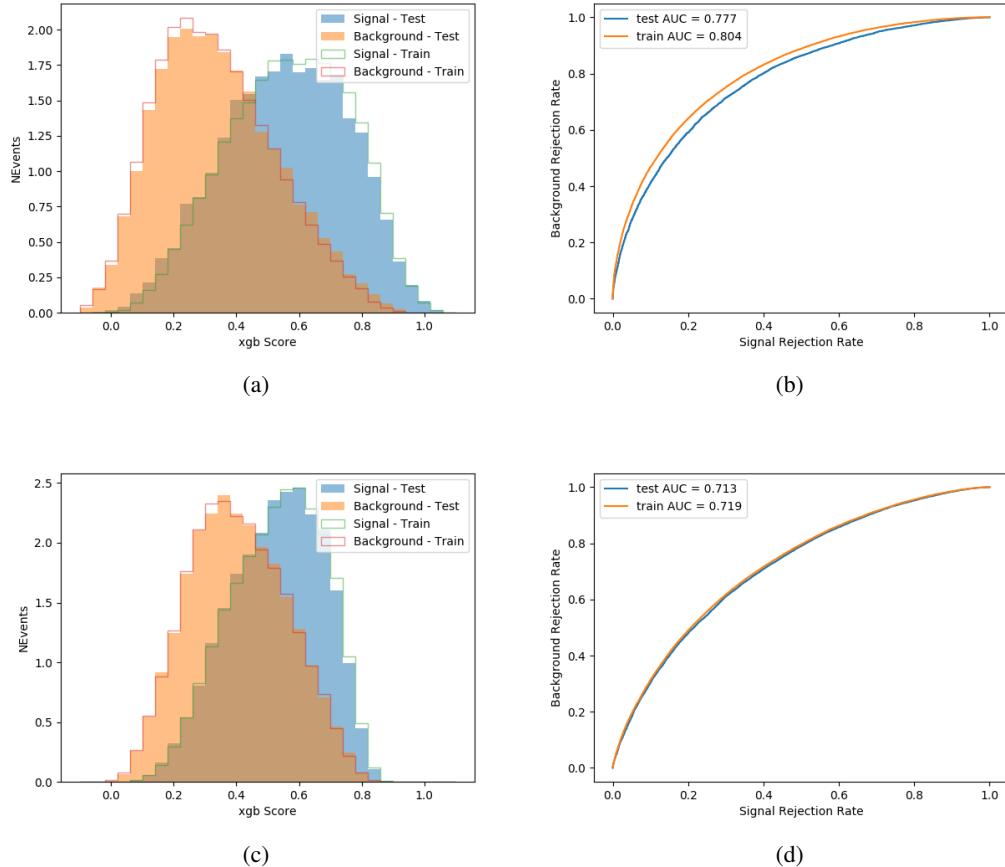


Figure 20.5:

¹⁴³¹ is predicted to be high (> 150 GeV) or low (< 150 GeV), a cut on the appropriate background
¹⁴³² rejection MVA is applied. The particular cut values, listed in table 43, are determined by
¹⁴³³ maximizing S/\sqrt{B} in each region.

¹⁴³⁴ The event preselection and MVA selection define the three signal regions. These signal
¹⁴³⁵ region definitions are summarized in table 44.

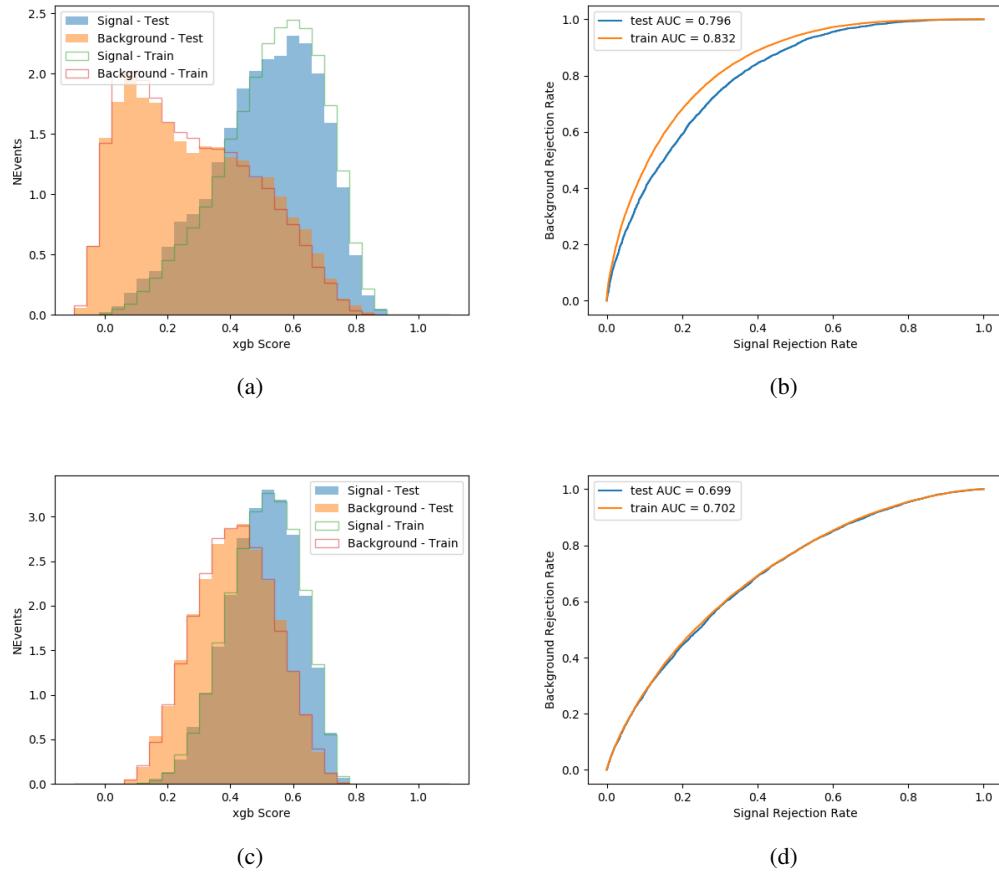


Figure 20.6:

Channel	BDT Score
2lSS high p_T	0.36
2lSS low p_T	0.34
3lS high p_T	0.51
3lS low p_T	0.43
3lF high p_T	0.33
3lF low p_T	0.41

Table 43: Cutoff values on background rejection MVA score applied to signal regions.

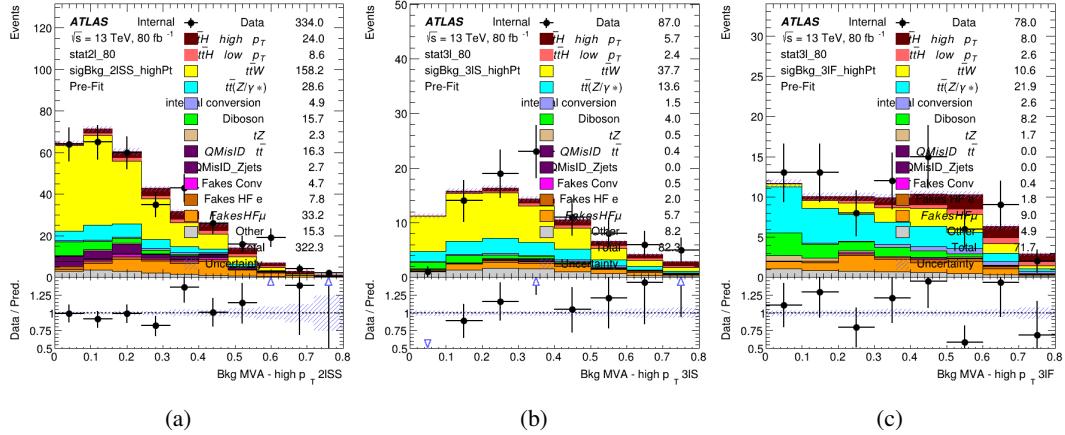


Figure 20.7: Output score of the high p_T BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channels

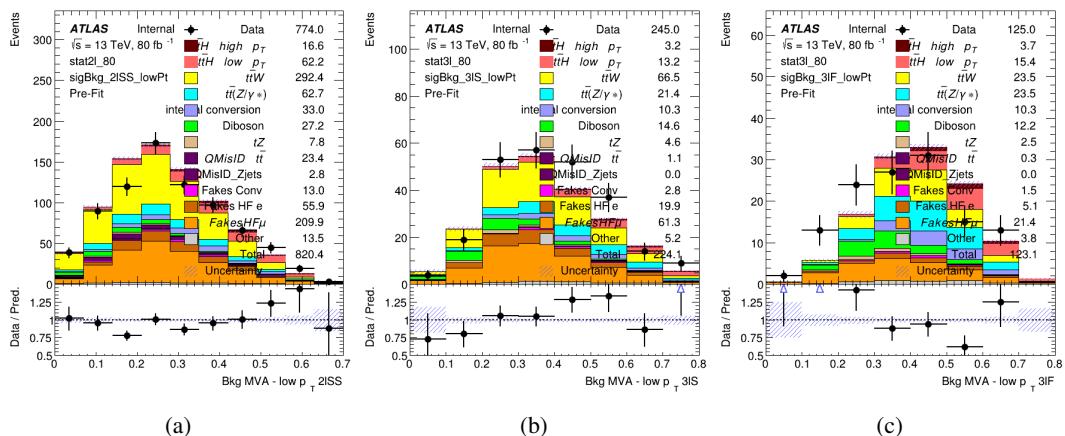


Figure 20.8: Output score of the low p_T BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channels

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

Table 44: Selection applied to define the three signal regions used in the fit.

21 Background Rejection MVA

- ¹⁴³⁶ Events are divided into two channels based on the number of leptons in the final state: one with
- ¹⁴³⁸ two same-sign leptons, the other with three leptons. The 3l channel includes events where both
- ¹⁴³⁹ leptons originated from the Higgs boson as well as events where only one of the leptons

1440 21.1 Pre-MVA Event Selection

1441 A preselection is applied to define orthogonal analysis channels based on the number of leptons
 1442 in each event. For the 2lSS channel, the following preselection is used:

- 1443 • Two very tight, same-charge, light leptons with $p_T > 20$ GeV

- 1444 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets

- 1445 • No reconstructed tau candidates

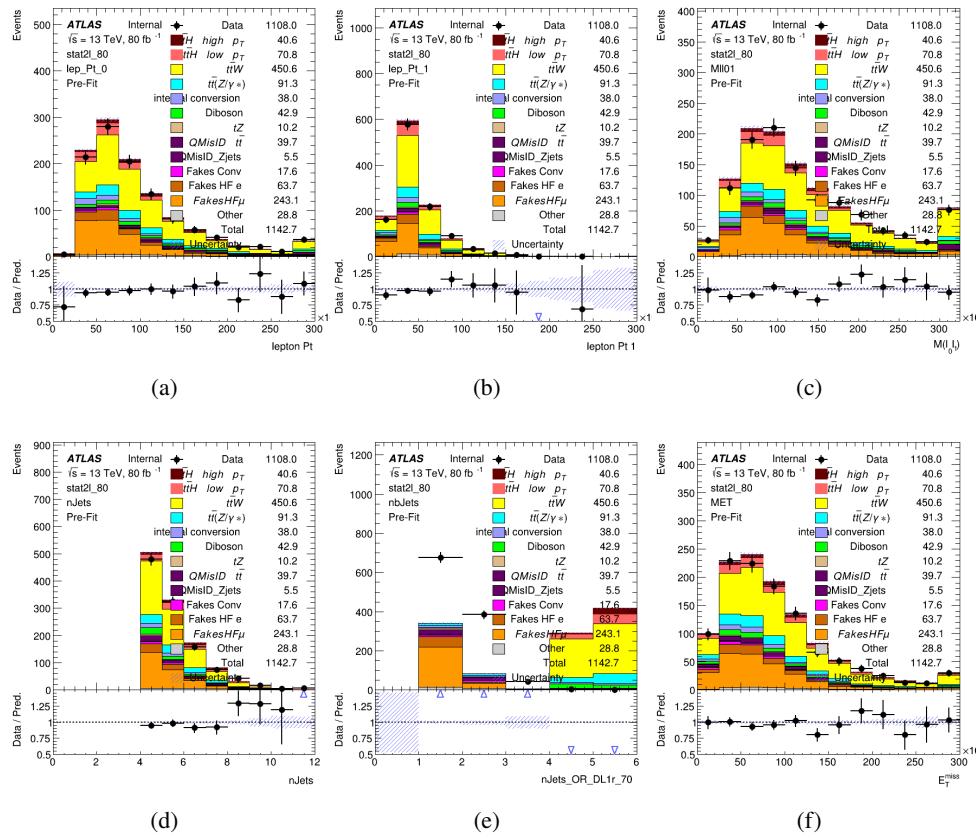


Figure 21.1:

1446 For the 3l channel, the following selection is applied:

- 1447 • Three light leptons with total charge ± 1
- 1448 • Same charge leptons are required to be very tight, with $p_T > 20 \text{ GeV}$
- 1449 • Opposite charge lepton must be loose, with $p_T > 10 \text{ GeV}$
- 1450 • $>= 2$ reconstructed jets, $>= 1$ b-tagged jets
- 1451 • No reconstructed tau candidates
- 1452 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs

1453 21.2 Event MVA

1454 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
1455 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
1456 (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost
1457 [**xgboost**] are trained using the kinematics of signal and background events derived from Monte
1458 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo
1459 event.

1460 Because the background composition differs for events with a high reconstructed Higgs p_T
1461 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and

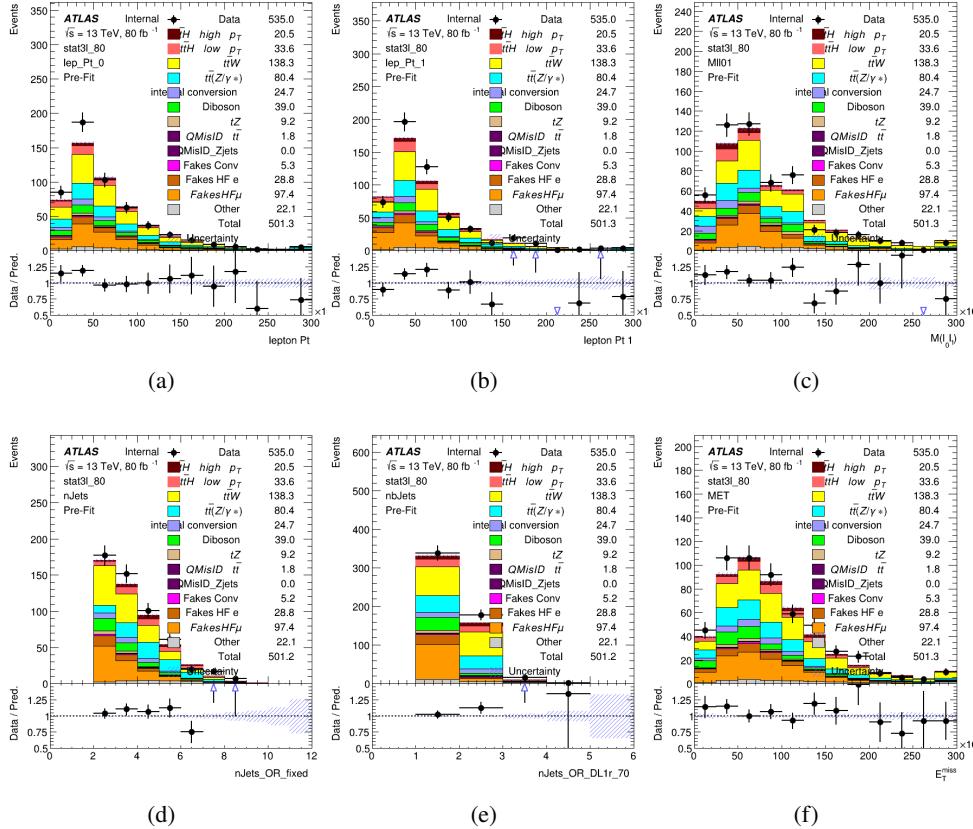


Figure 21.2:

1462 low p_T regions. This is found to provide better significance than attempting to build an inclusive
 1463 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6
 1464 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
 1465 3lF high p_T , and 3lF low p_T .

1466 The following features are used in both the high and low p_T 2lSS BDTs:

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

Table 45: Input features used to distinguish signal and background events in the 2lSS channel.

1467

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

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

Table 46: Input features used to distinguish signal and background events in the 3l channel.

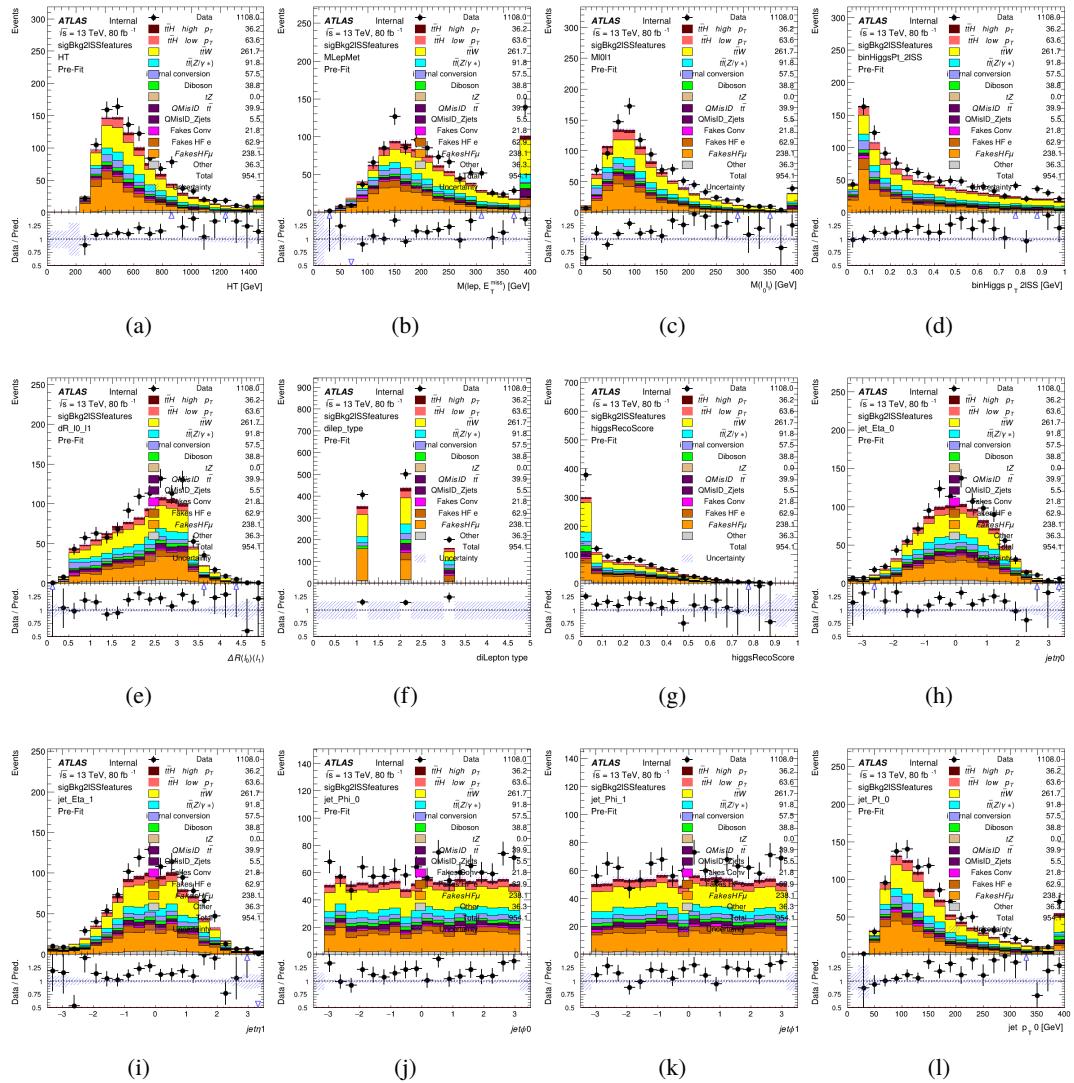


Figure 21.3:

1468 The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss

1470 Output distributions of each MVA comparing MC prediction to data at 80 fb^{-1} are shown
1471 in figure 21.2.

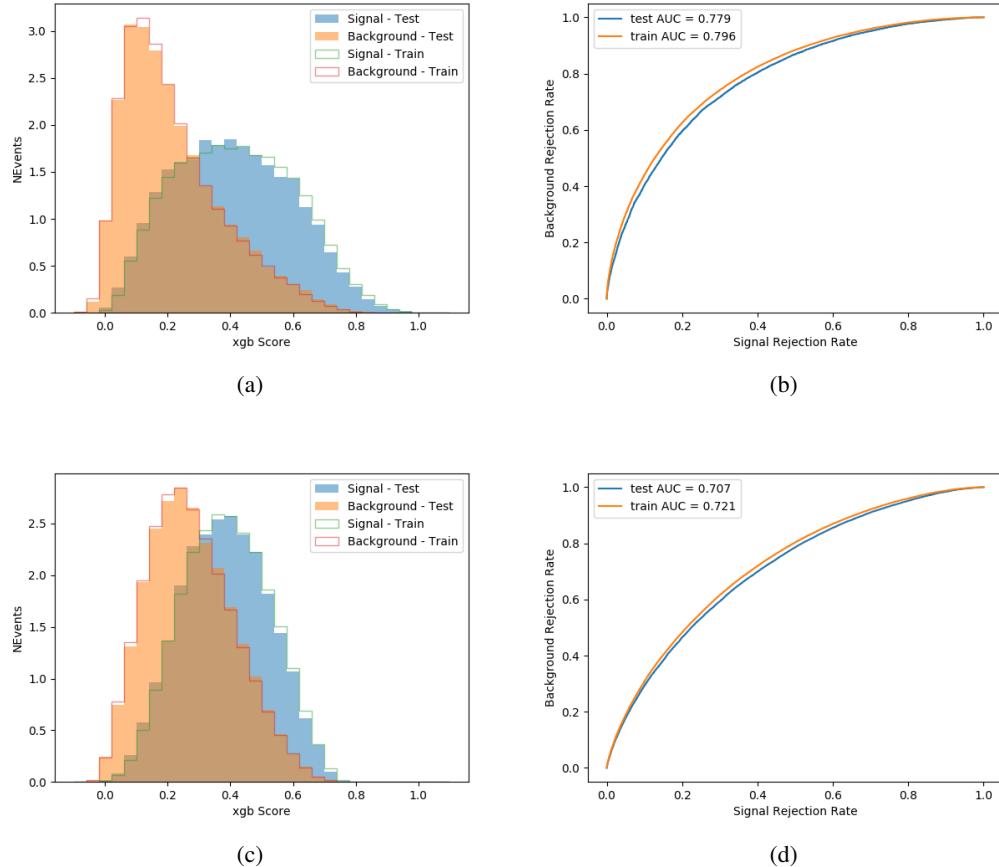


Figure 21.4:

21.3 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 19.5 is used to separate the three channel into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs boson.

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

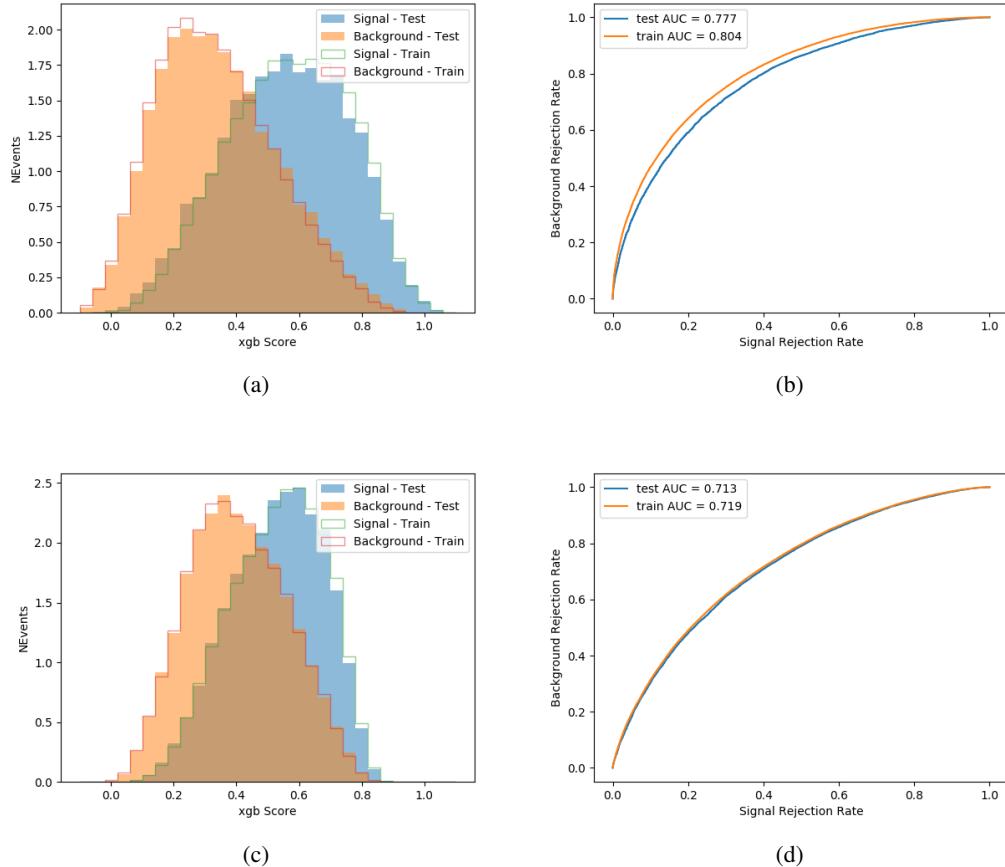


Figure 21.5:

from the algorithm described in section 19.4, a cut on the appropriate background rejection
 1478 algorithm is applied. The specific selection used, and the event yield in each channel after this
 1479 selection has been applied, is summarized below.
 1480

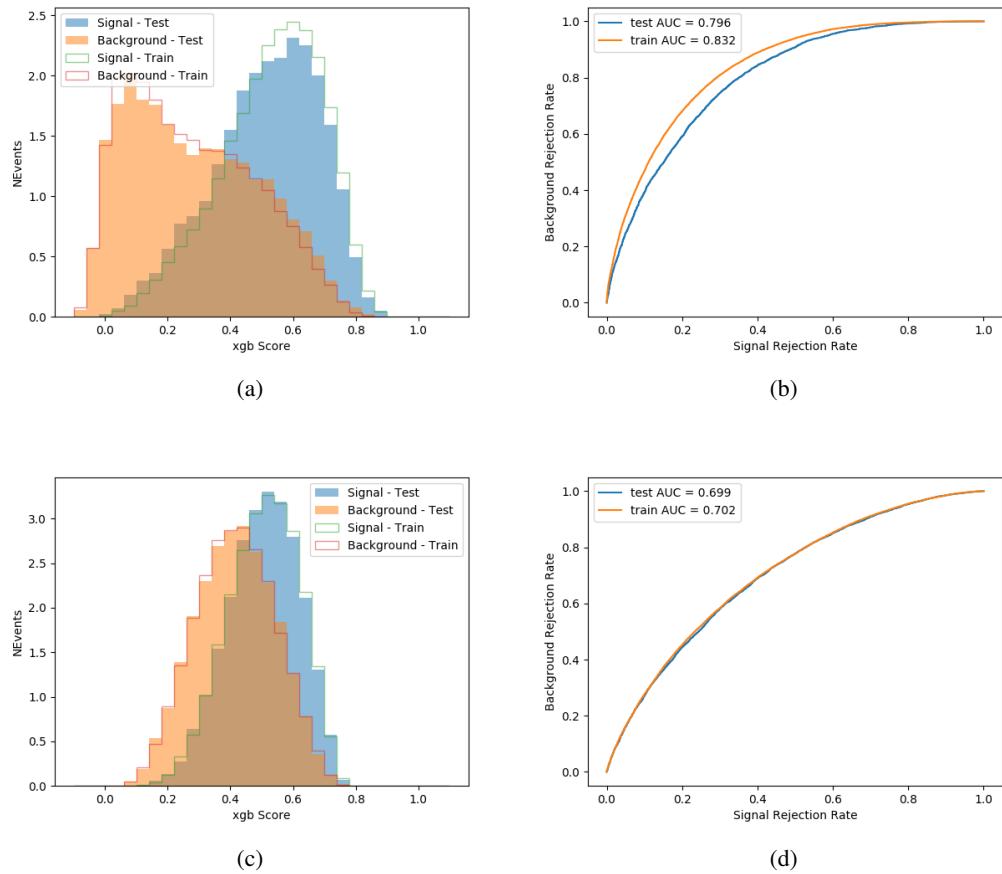


Figure 21.6:

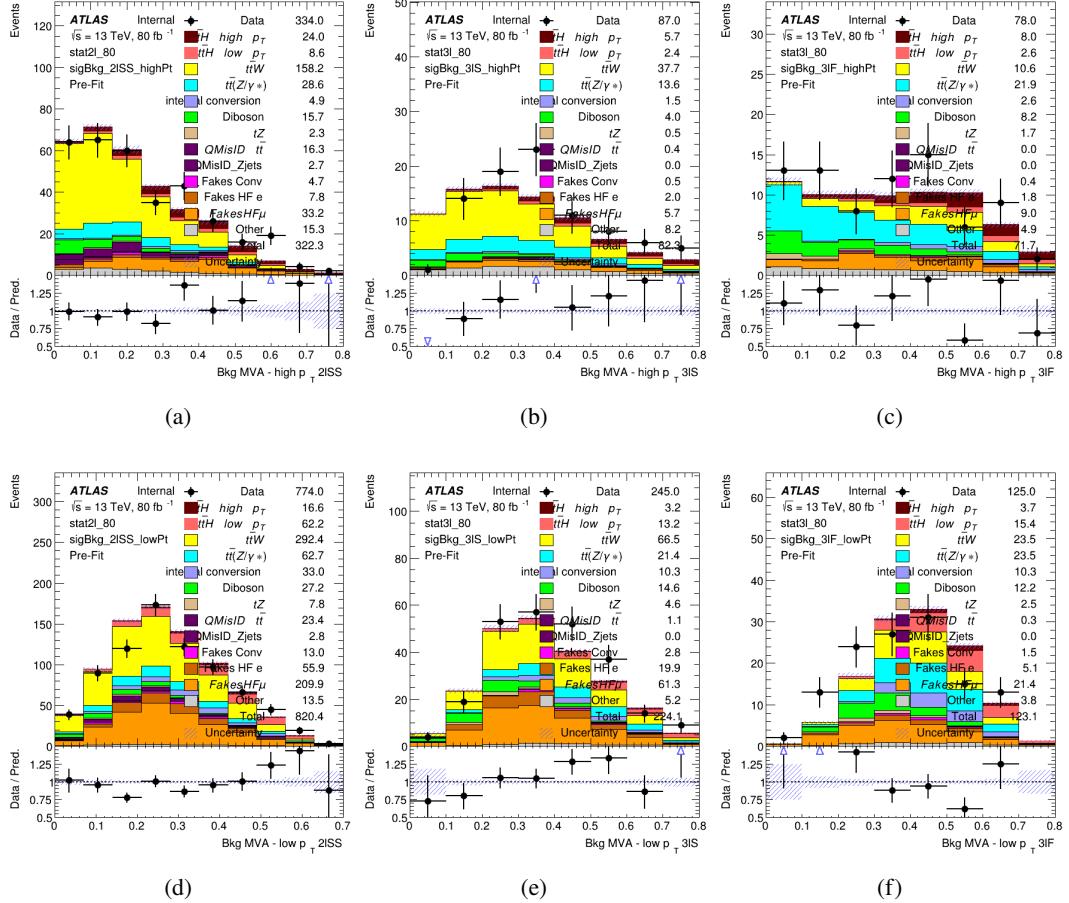


Figure 21.7: scores

1481 **21.3.1 2ISS**1482 **21.3.2 3l – Semi – leptonic**1483 **21.3.3 3l – Fully – leptonic**

1484 **22 Systematic Uncertainties**

1485 The systematic uncertainties that are considered are summarized in table 47. These are implemented in the fit either as a normalization factors or as a shape variation or both in the signal
 1486 24th February 2021 – 14:28

¹⁴⁸⁷ and background estimations. The numerical impact of each of these uncertainties is outlined in
¹⁴⁸⁸ section 23.

Table 47: 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

¹⁴⁸⁹ The uncertainty in the combined 2015+2016 integrated luminosity is derived from a
¹⁴⁹⁰ calibration of the luminosity scale using x-y beam-separation scans performed in August 2015
¹⁴⁹¹ and May 2016 [**lumi**].

¹⁴⁹² The experimental uncertainties are related to the reconstruction and identification of light
¹⁴⁹³ leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The sources which contribute

1494 to the uncertainty in the jet energy scale [**jes**] are decomposed into uncorrelated components and
 1495 treated as independent sources in the analysis.

1496 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
 1497 [**btag_cal**] are also decomposed into uncorrelated components. The large number of components
 1498 for b-tagging is due to the calibration of the distribution of the BDT discriminant.

1499 The systematic uncertainties associated with the signal and background processes are
 1500 accounted for by varying the cross-section of each process within its uncertainty.

1501 23 Results

1502 Unblinded results are shown for the 80 fb^{-1} data set, as well as MC only projections of results
 1503 using the full Run-2, 140 fb^{-1} dataset.

1504 23.1 Results - 80 fb^{-1}

1505 A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p_T spectrum
 1506 in the three signal regions, 2lSS, 3lS, and 3lF, shown in figure 23.1. The $t\bar{t}H$ MC is split into
 1507 high and low p_T , based on whether the truth p_T of the Higgs is above or below 150 GeV. The
 1508 parameters $\mu_{t\bar{t}H\text{high}p_T}$ and $\mu_{t\bar{t}H\text{low}p_T}$, where $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$, are extracted from the fit.

1509 As the data collected from 2015-2017 has been unblinded for $t\bar{t}H$ – ML channels, rep-
 1510 resenting 80 fb^{-1} , those events are unblinded.

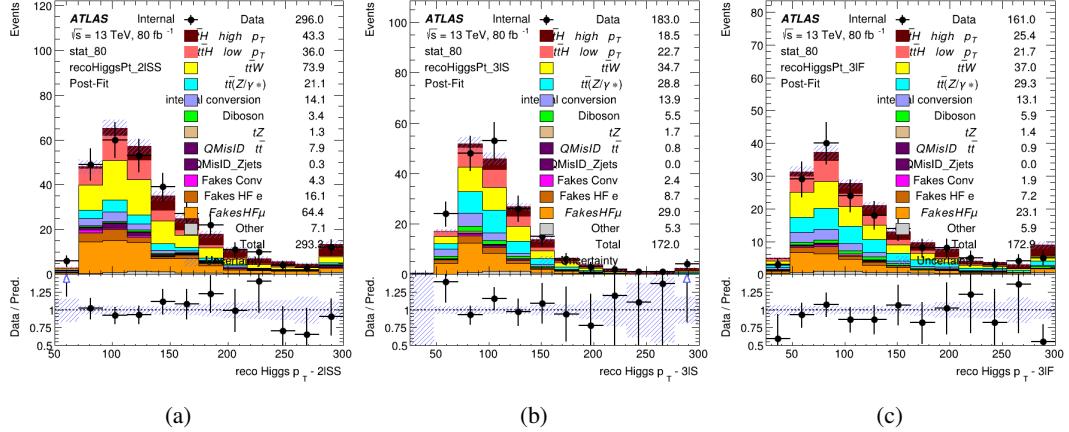


Figure 23.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 80 fb^{-1} of MC

1511 A post-fit summary of the fitted regions is shown in figure 23.2.

The μ values for high and low p_T Higgs are shown in 48.

$$\begin{aligned}\mu_{t\bar{t}H \text{ high } p_T} &= 1.0^{+0.0}_{-0.0} (\text{stat})^{+0.0}_{-0.0} (\text{sys}) \\ \mu_{t\bar{t}H \text{ low } p_T} &= 1.0^{+0.0}_{-0.0} (\text{stat})^{+0.0}_{-0.0} (\text{sys})\end{aligned}$$

Table 48: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1513 **Need to add something about systematics here**

1514 The background composition of each of the fit regions is shown in figure 23.3.

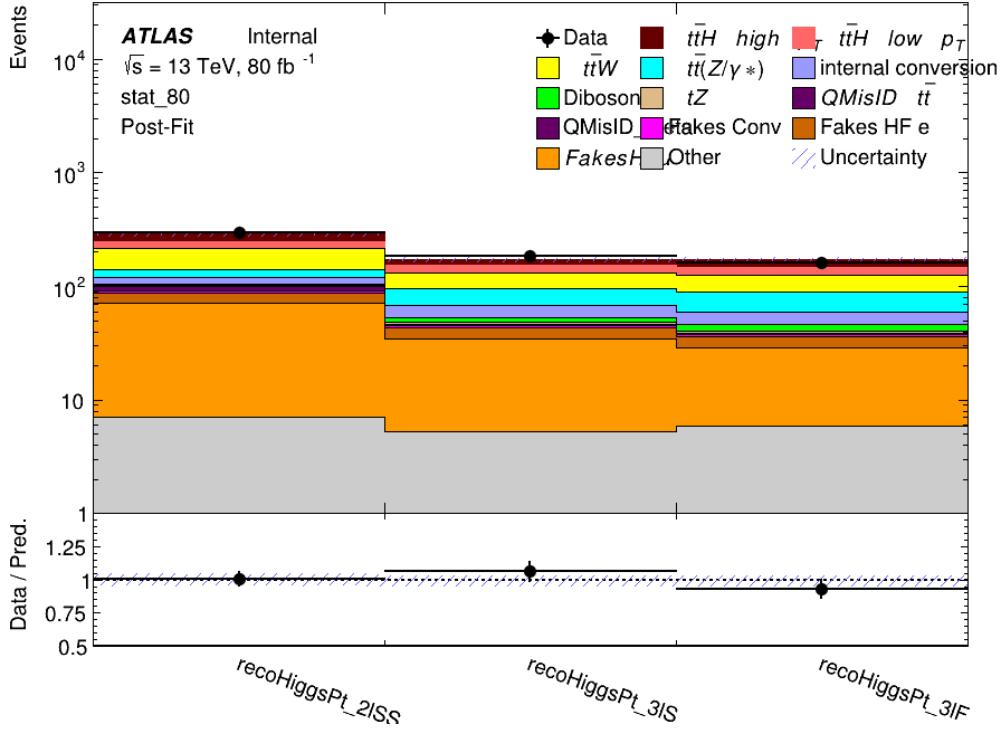


Figure 23.2: Post-fit summary of the yields in each signal region.

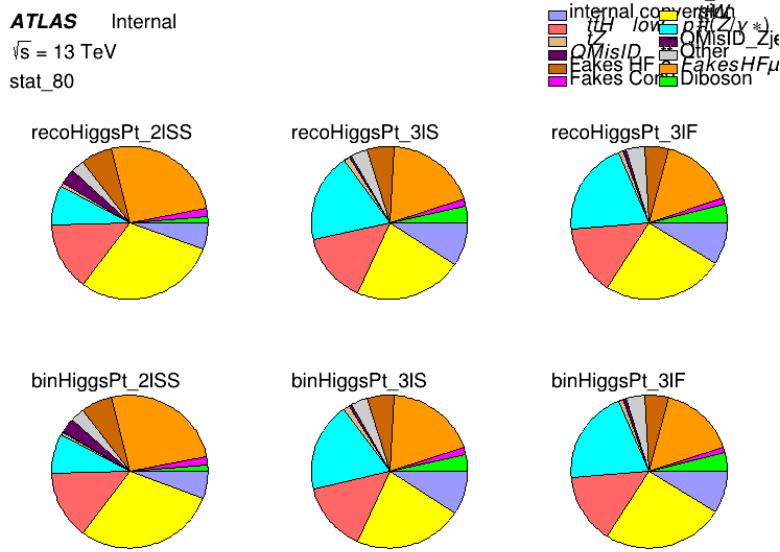


Figure 23.3: Background composition of the fit regions.

1515 **23.2 Projected Results - 140 fb^{-1}**

1516 As data collected in 2018 has not yet been unblinded for $t\bar{t}H - \text{ML}$ at the time of this note, data
 1517 from that year remains blinded. Instead, an Asimov fit is performed - with the MC prediction
 1518 being used both as the SM prediction as well as the data in the fit - in order to give expected
 1519 results.

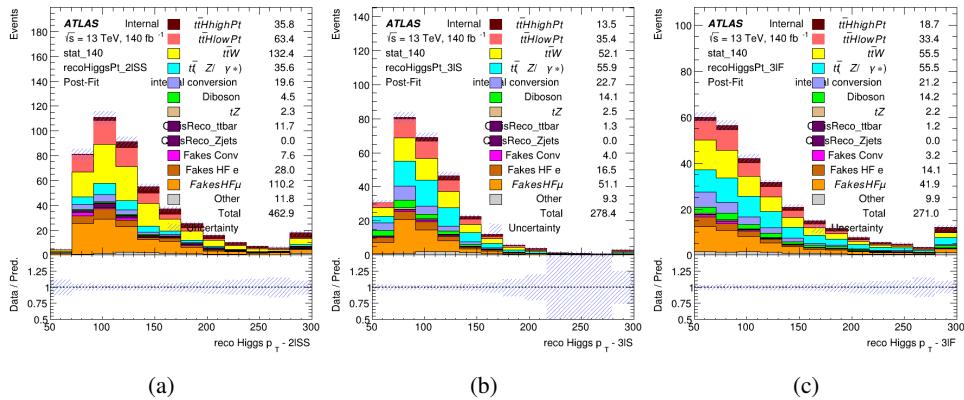


Figure 23.4: Blinded post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 140 fb^{-1} of data

1520 The μ values for high and low p_T Higgs are shown in 49.

$$\begin{aligned}\mu_{t\bar{t}H\text{high}p_T} &= 1.0^{+0.0}(\text{stat})^{+0.0}(\text{sys}) \\ \mu_{t\bar{t}H\text{low}p_T} &= 1.0^{+0.0}(\text{stat})^{+0.0}(\text{sys})\end{aligned}$$

Table 49: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

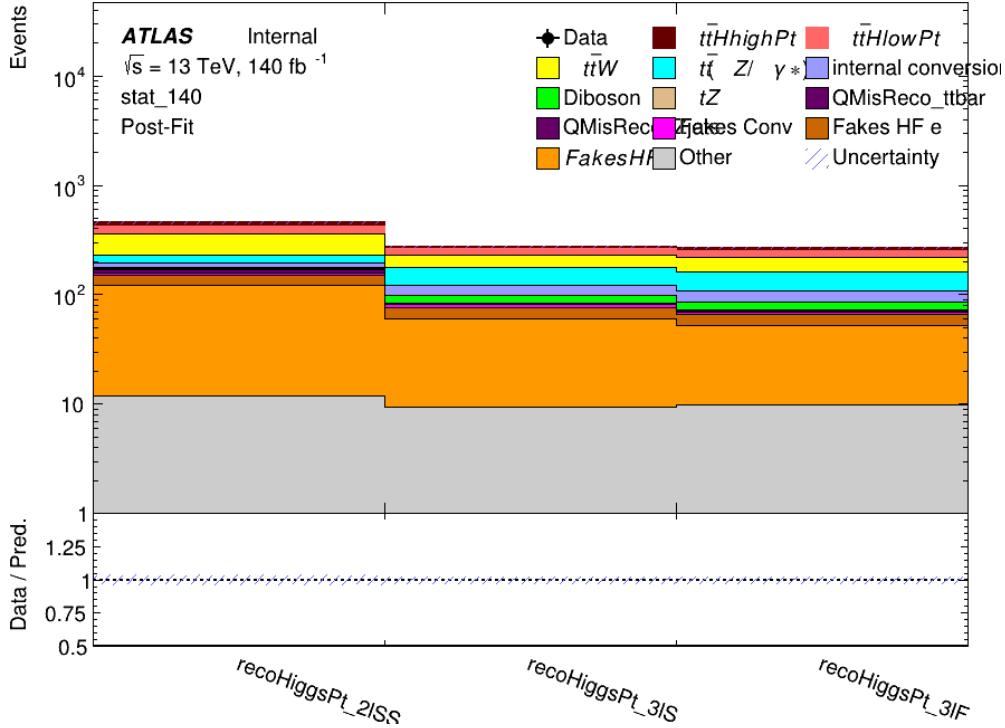


Figure 23.5: Post-fit summary of fit.

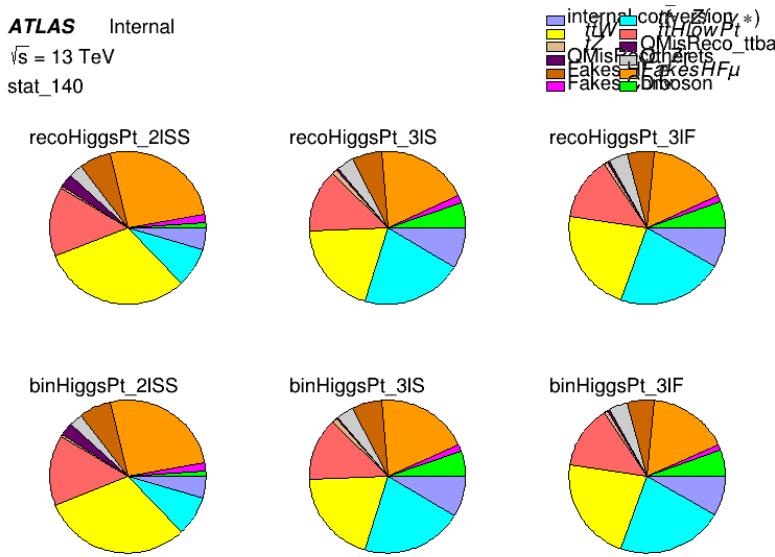


Figure 23.6: Background composition of the fit regions.

Part VI**Conclusion**

1521 As search for the effects of dimension-six operators on $t\bar{t}H$ production is performed. An effective
1522 field theory approached is used to parametrize the effects of high energy physics on the Higgs
1523 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,
1526 and the limits on dimension-six operators are limited to X.

List of contributions

1528

1529 **Appendices**

1530 **A Machine Learning Models**

1531 The following section provides details of the various MVAs as well as a few studies performed
1532 in support of this analysis, exploring alternate decisions and strategies.

1533 **A.1 Higgs Reconstruction Models**

1534 **A.1.1 b-jet Identification Features - 2lSS**

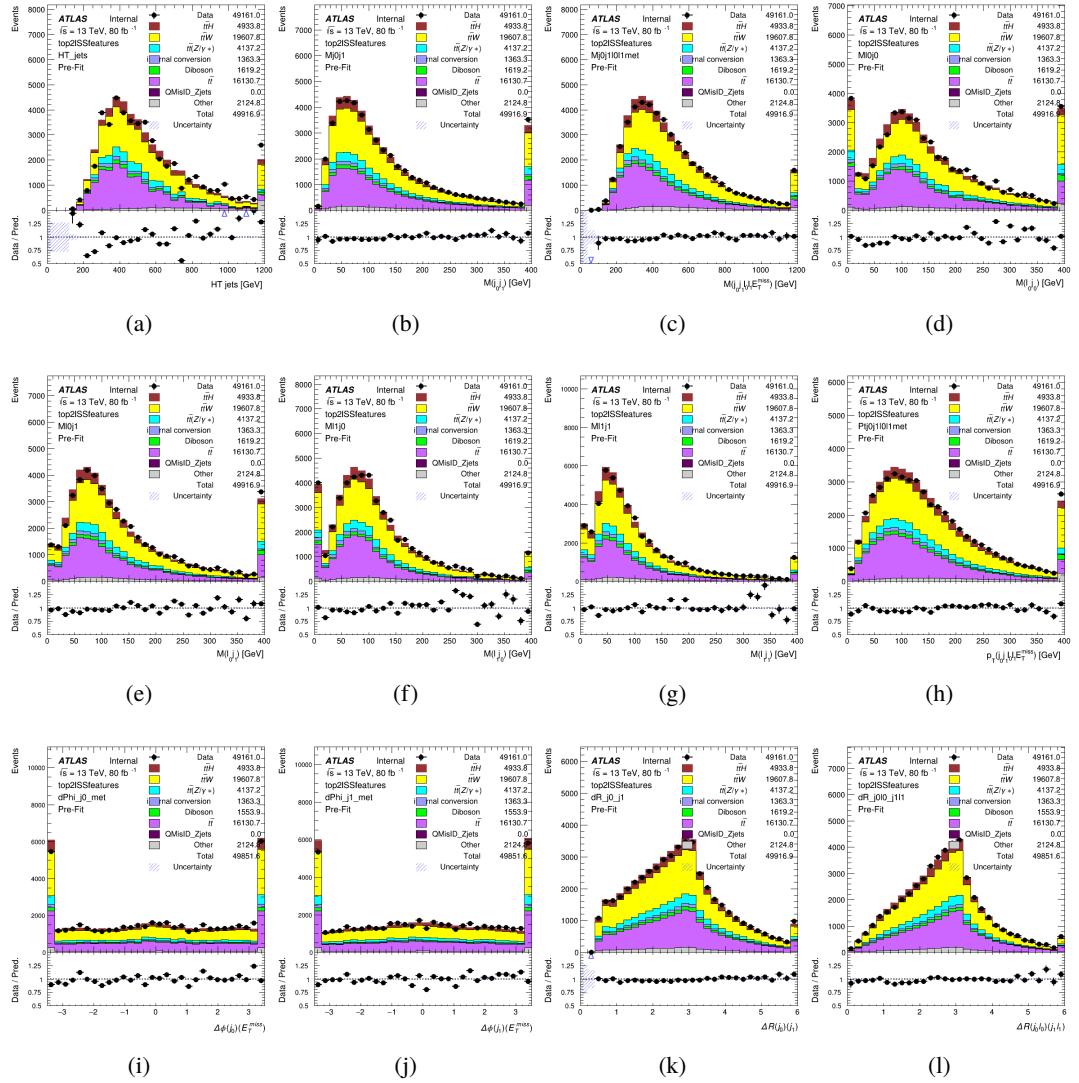


Figure A.1: Input features for top2lSS

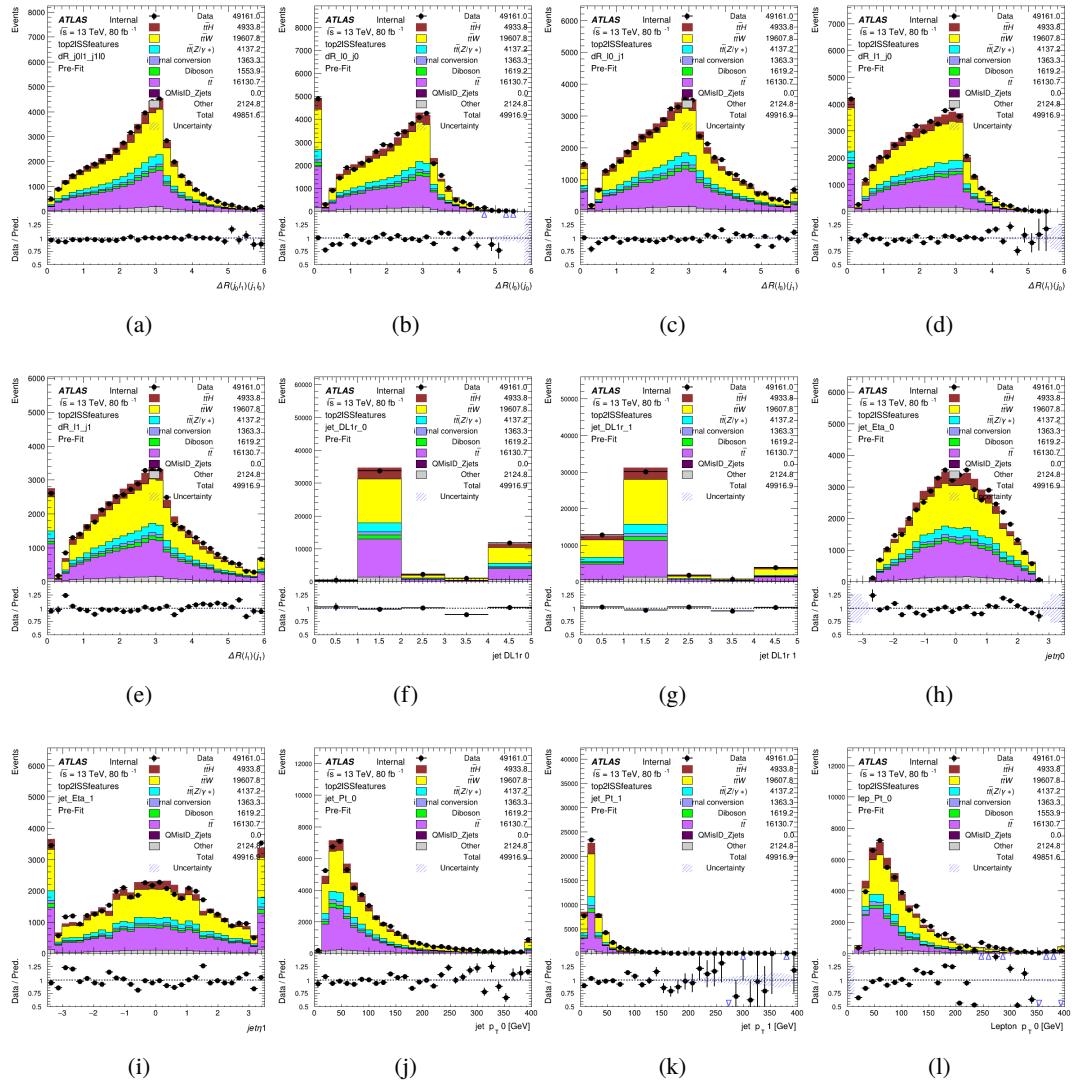


Figure A.2: Input features for top2lSS

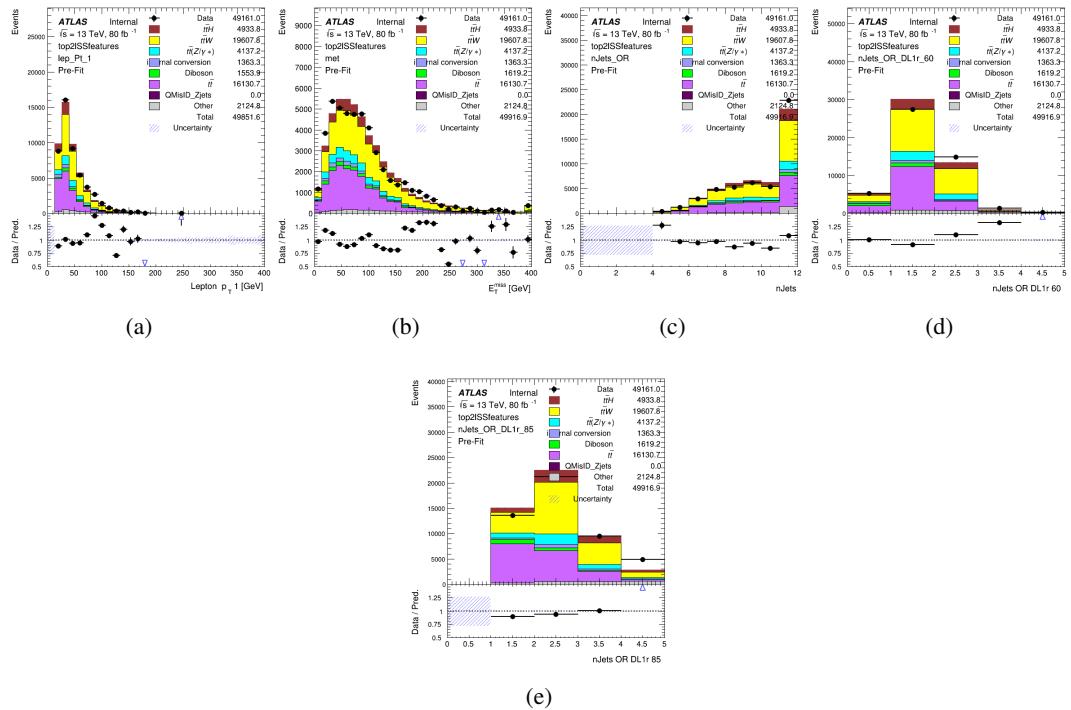


Figure A.3: Input features for top2lSS

1535 **A.1.2 b-jet Identification Features - 3l**

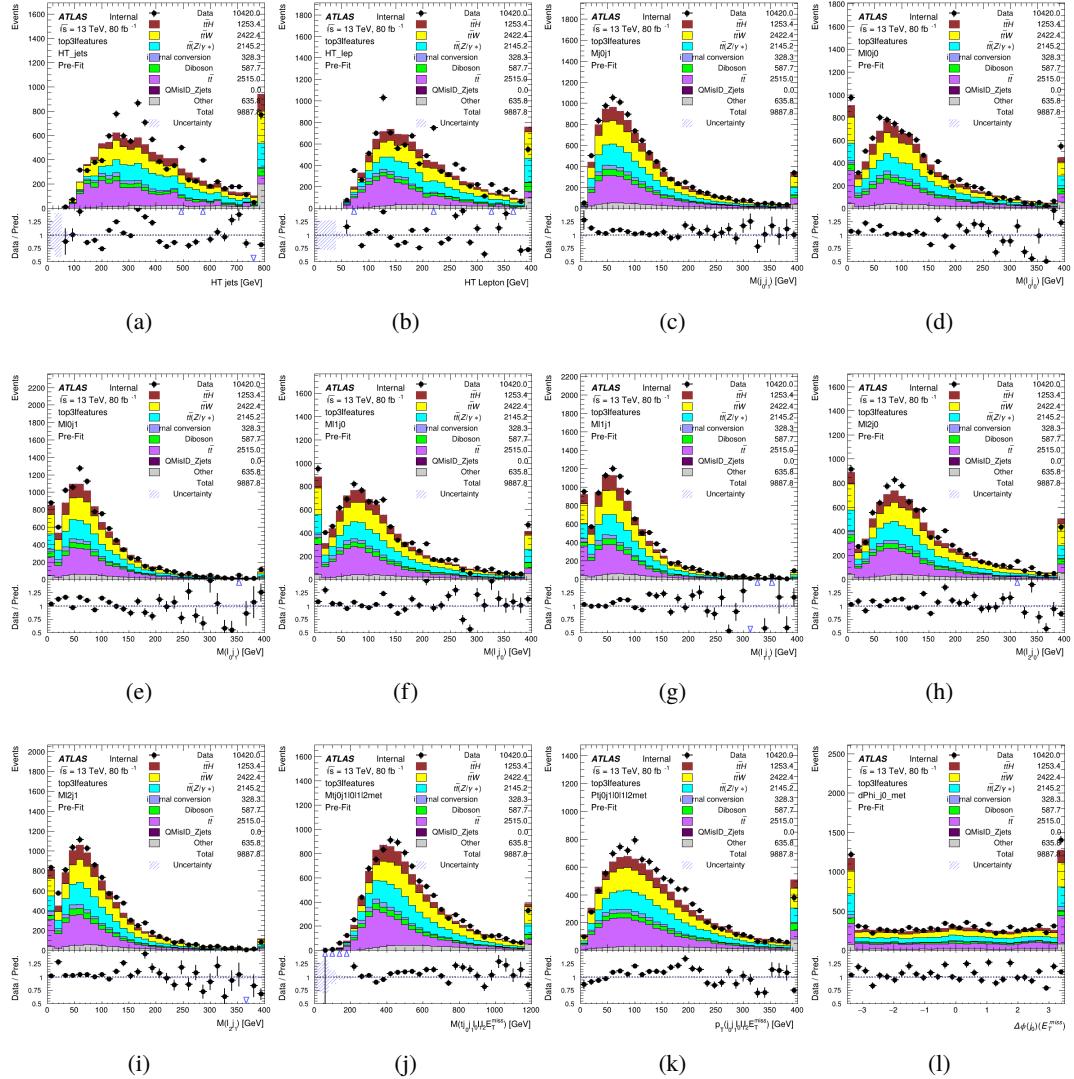


Figure A.4: Input features for top31

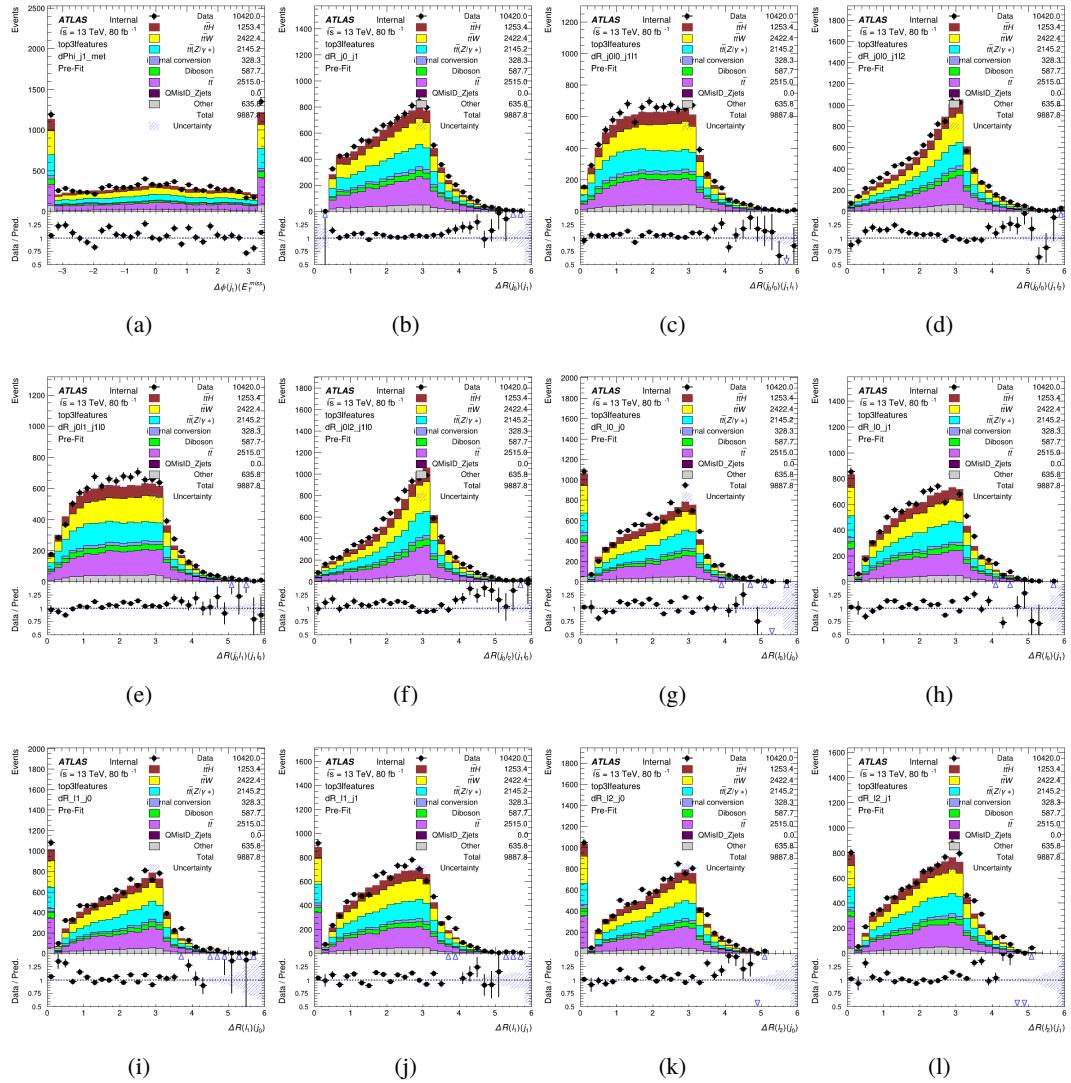


Figure A.5: Input features for top31

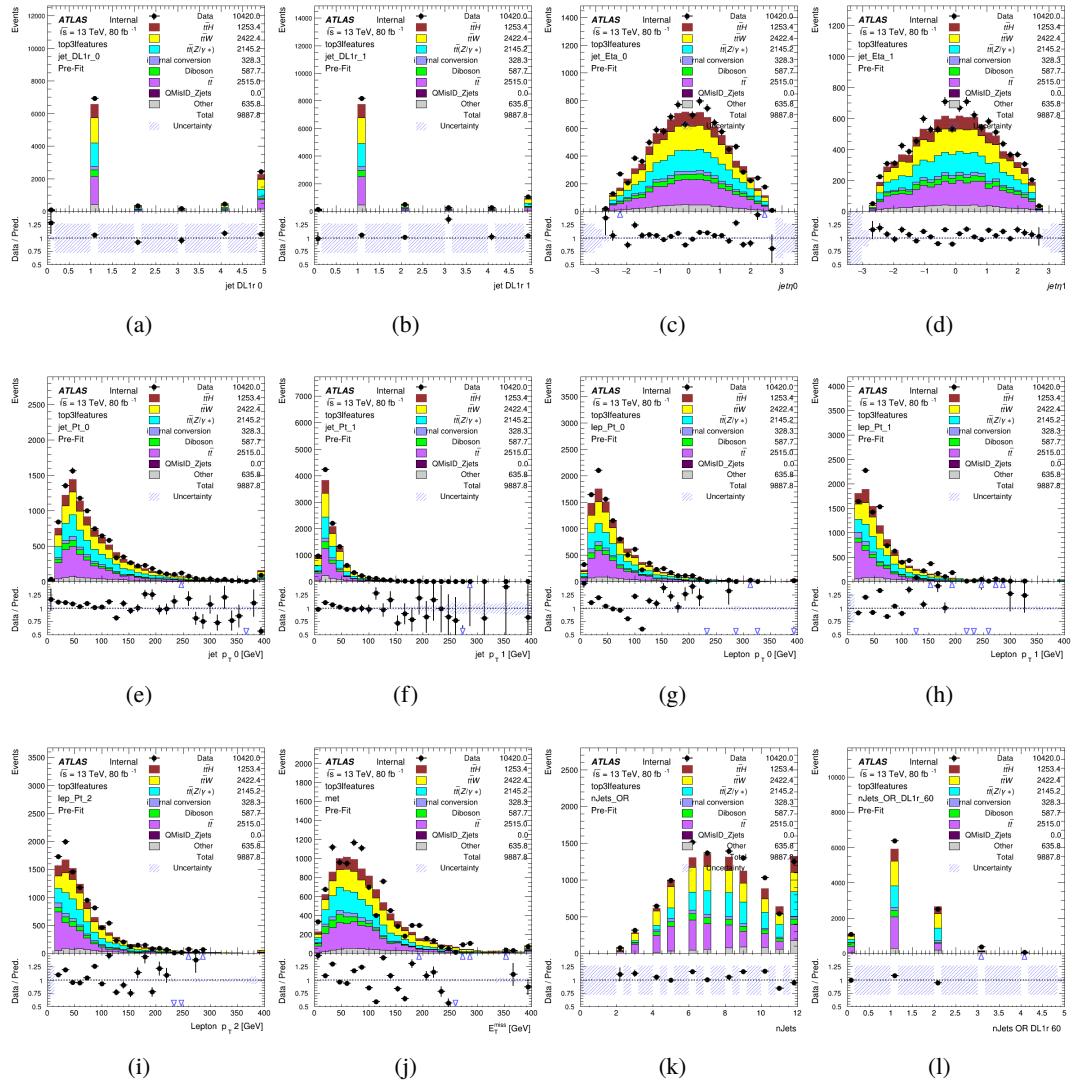
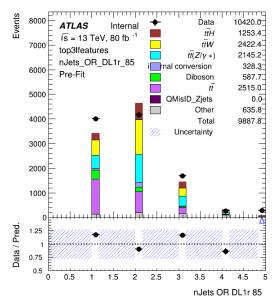


Figure A.6: Input features for top31



(a)

Figure A.7: Input features for top3l

¹⁵³⁶ **A.1.3 Higgs Reconstruction Features - 2lSS**

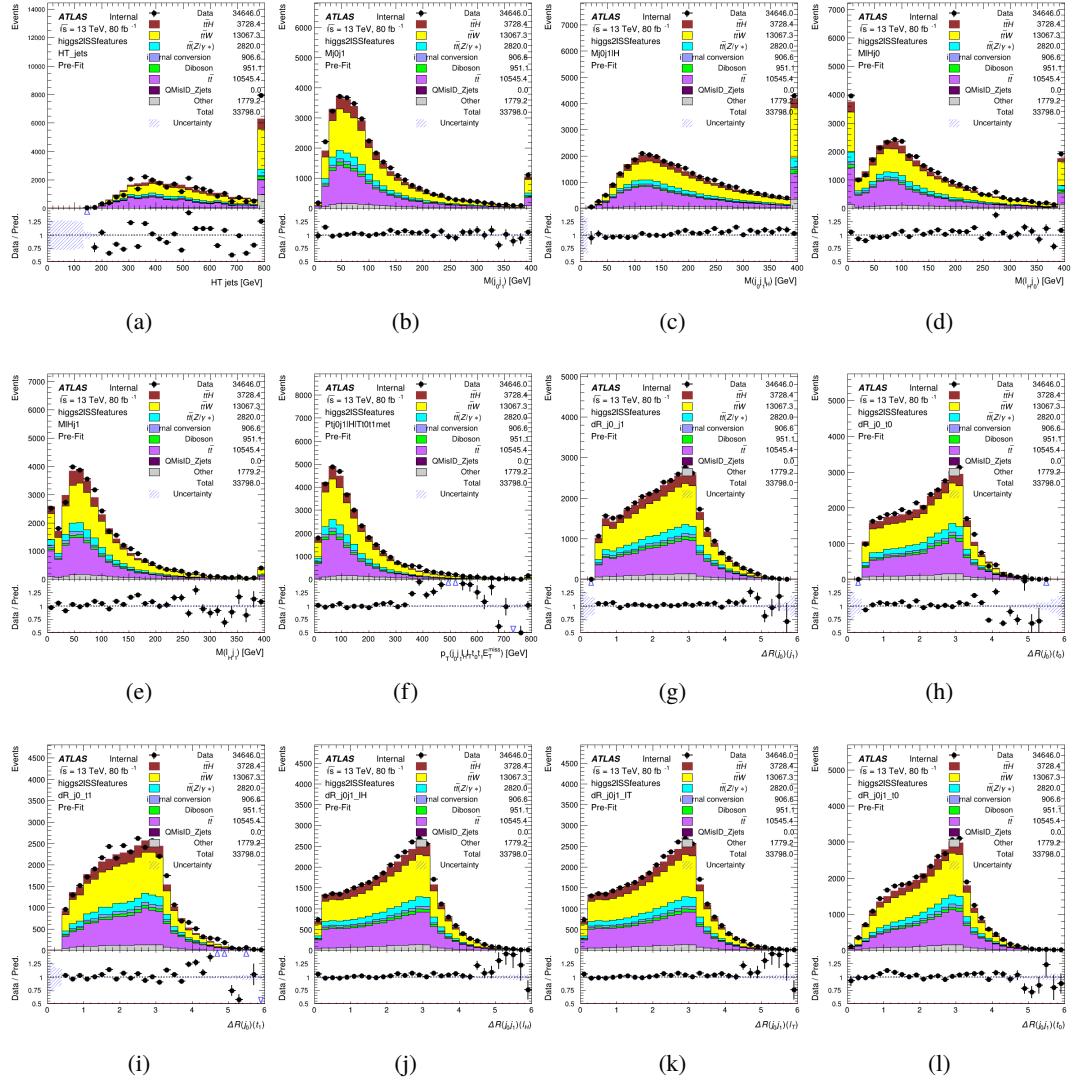


Figure A.8: Input features for higgs2lSS

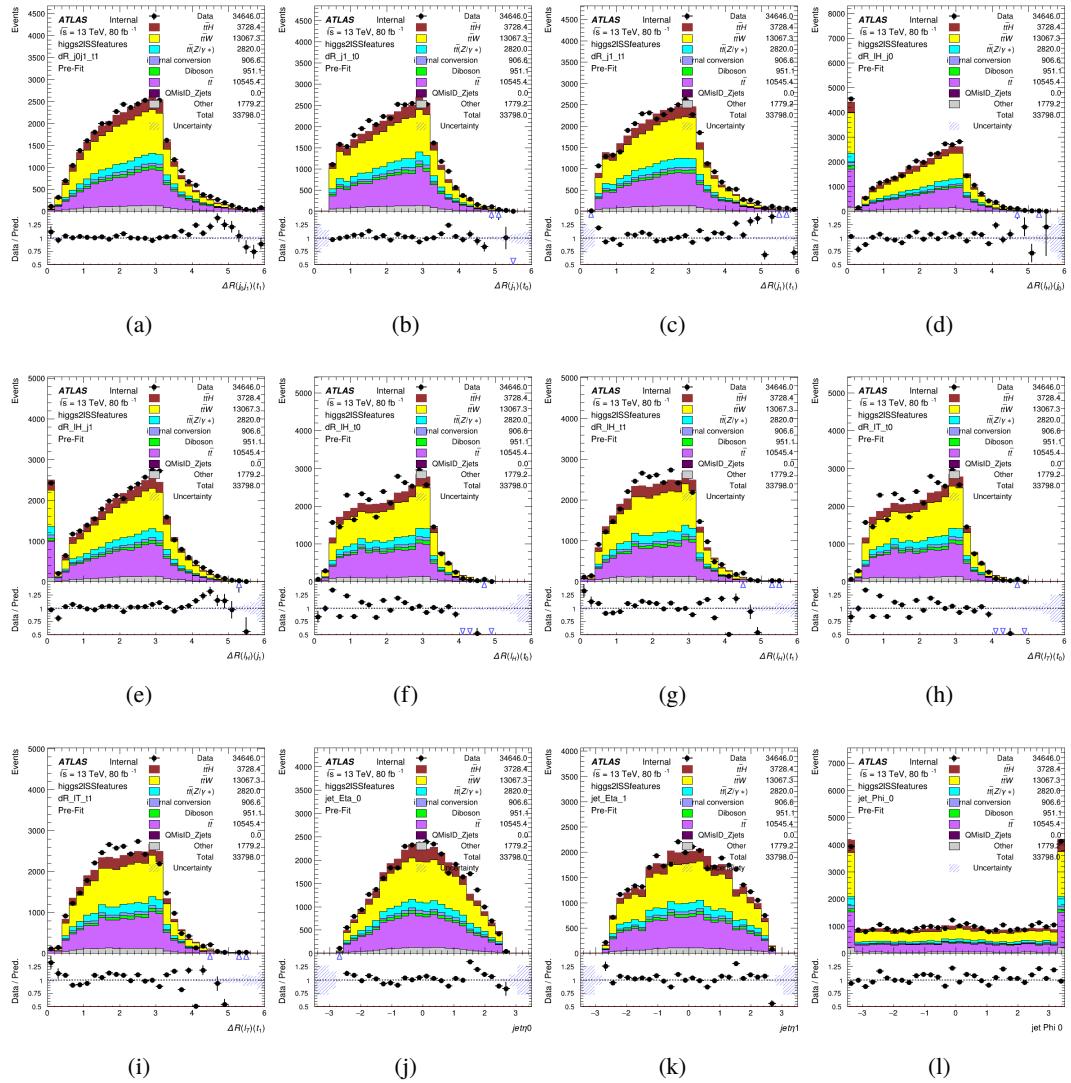


Figure A.9: Input features for higgs2lSS

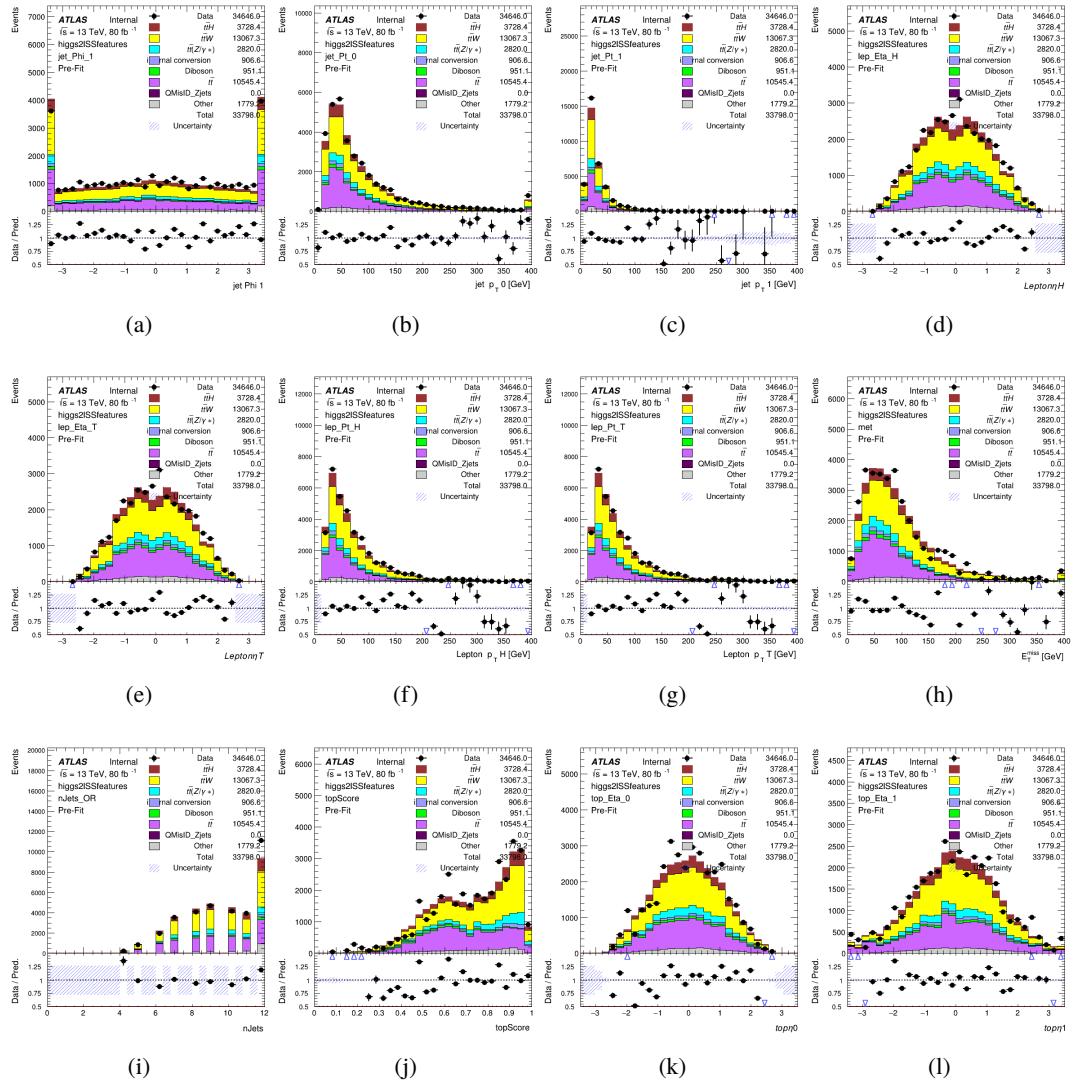


Figure A.10: Input features for higgs2ISS

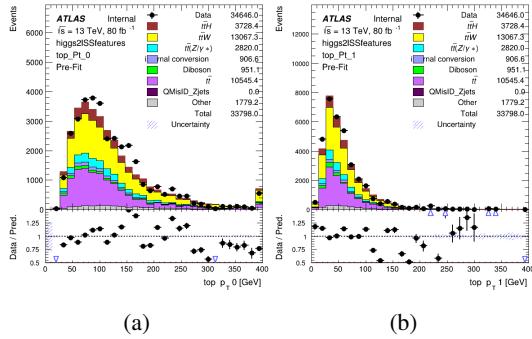


Figure A.11: Input features for higgs2ISS

1537 **A.1.4 Higgs Reconstruction Features - 3lS**

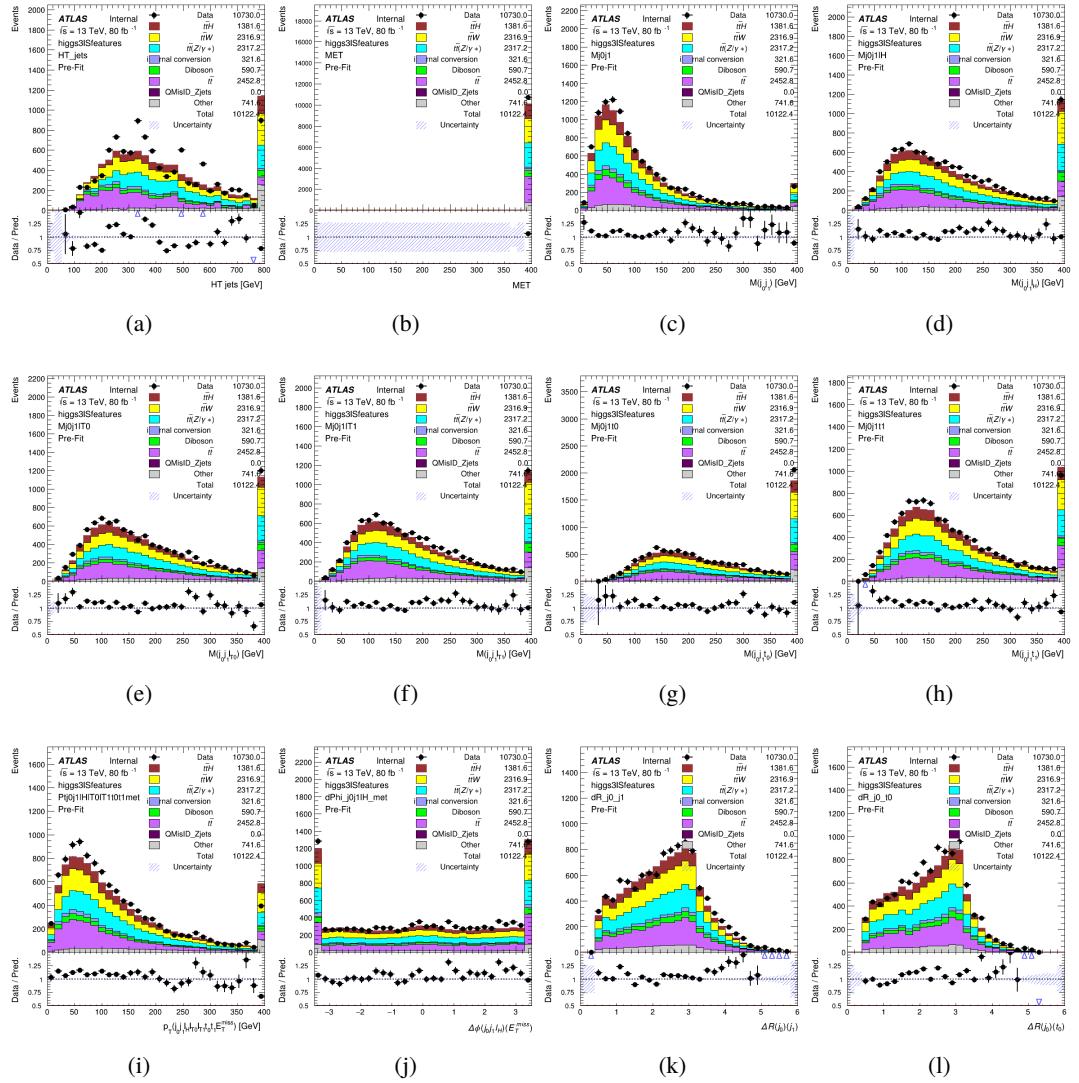


Figure A.12: Input features for higgs31S

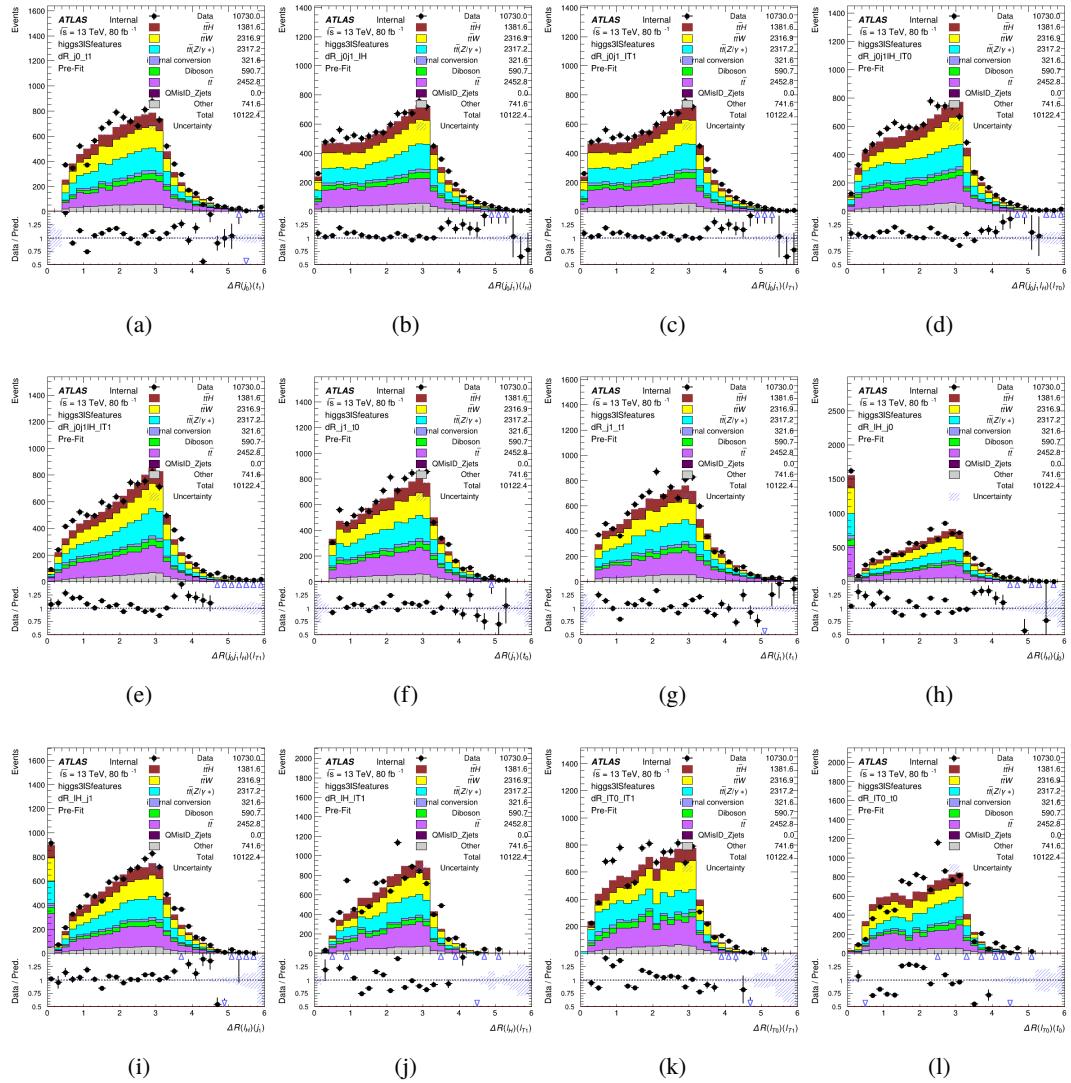


Figure A.13: Input features for higgs3lS

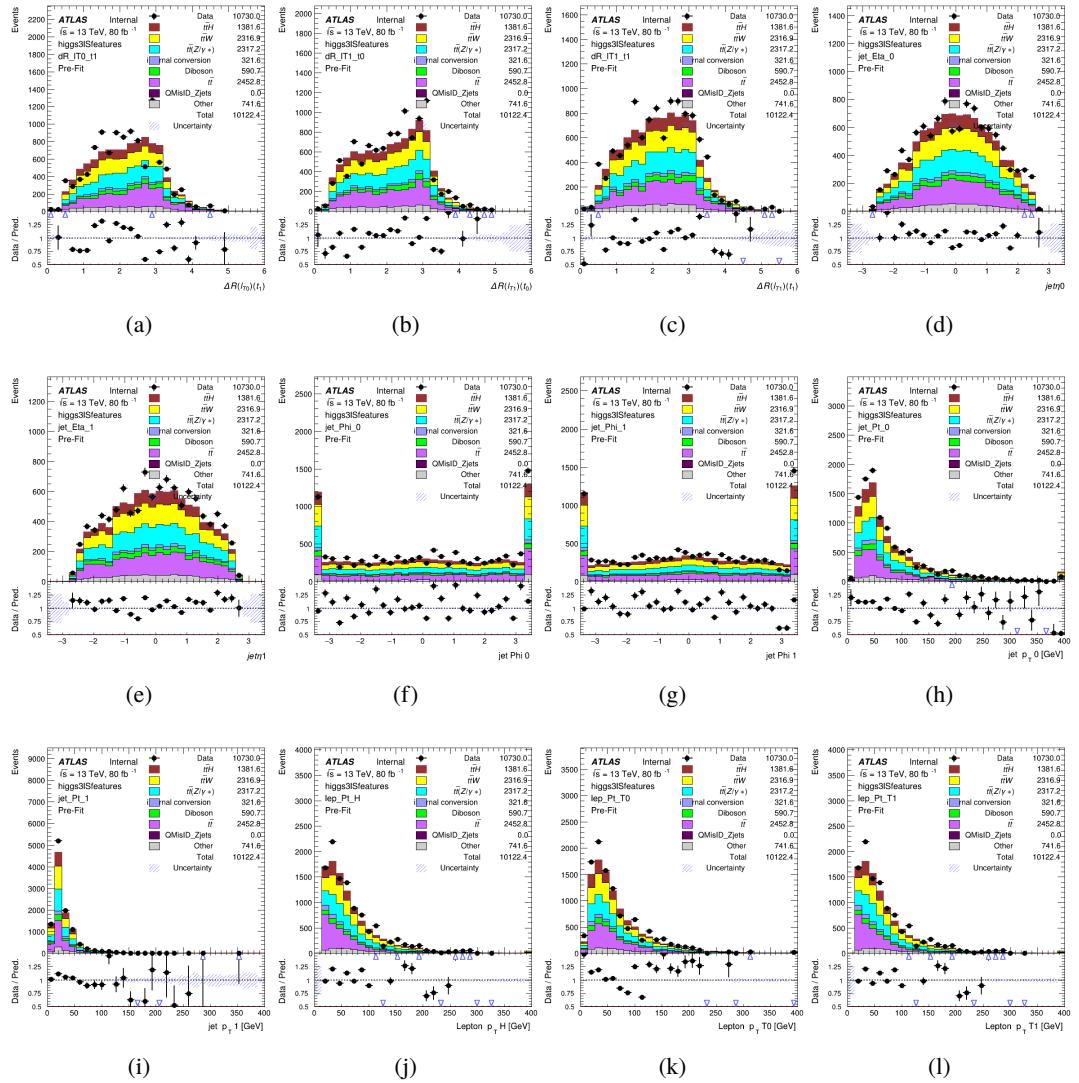


Figure A.14: Input features for higgs3lS

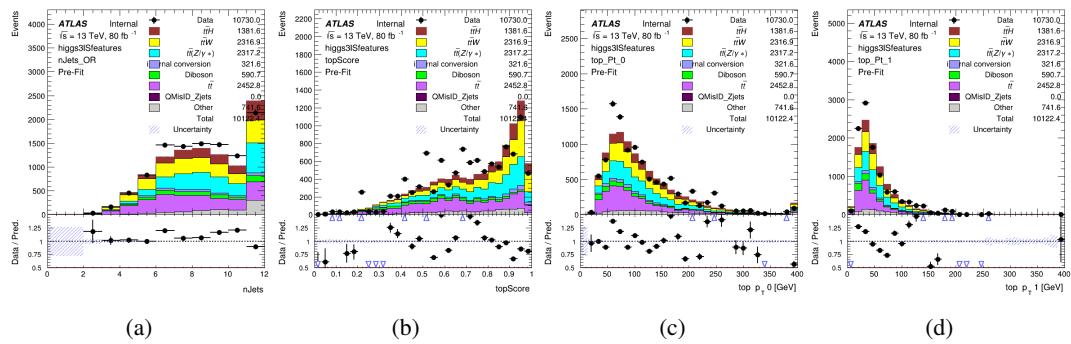


Figure A.15: Input features for higgs31S

1538 **A.1.5 Higgs Reconstruction Features - 3lF**

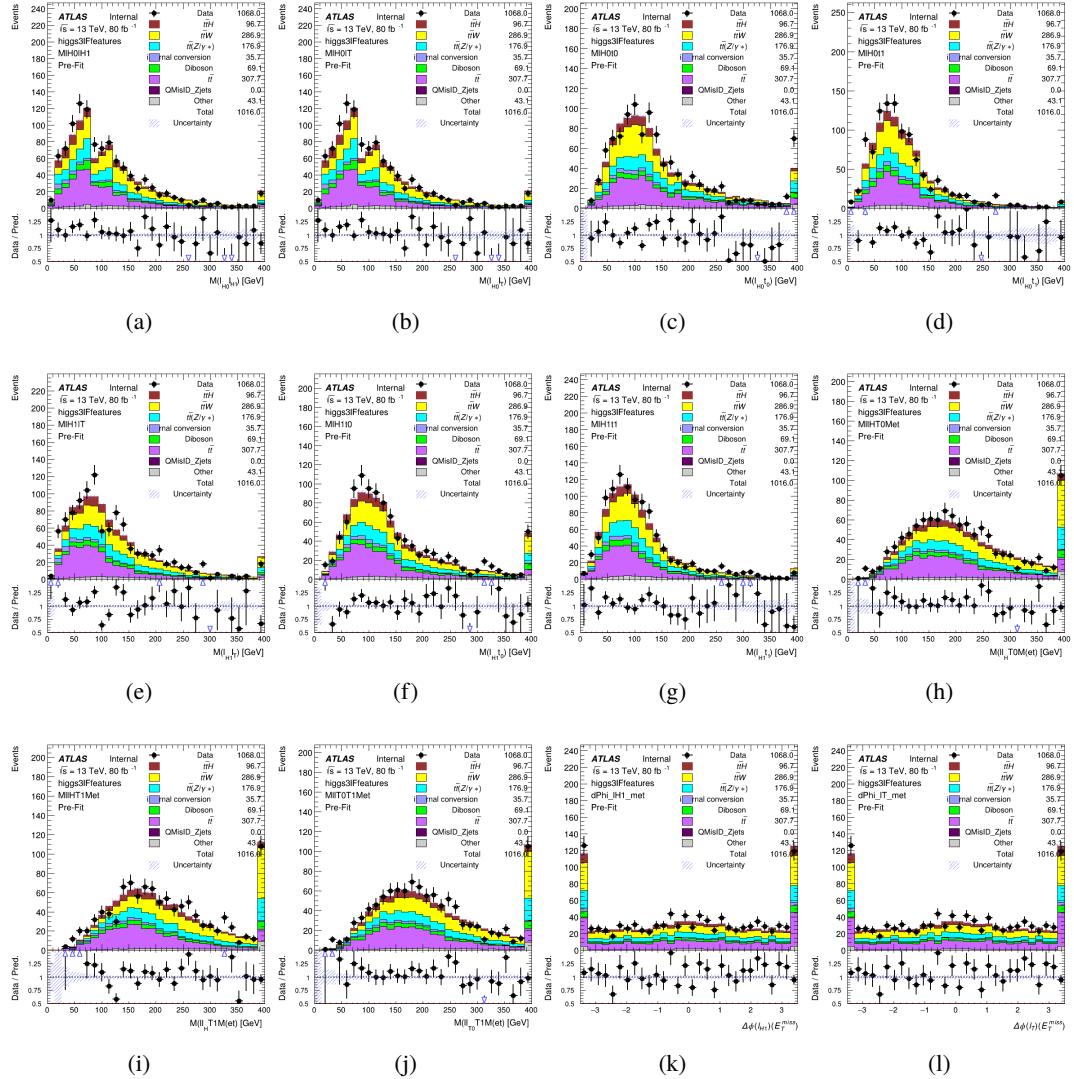


Figure A.16: Input features for higgs3IF

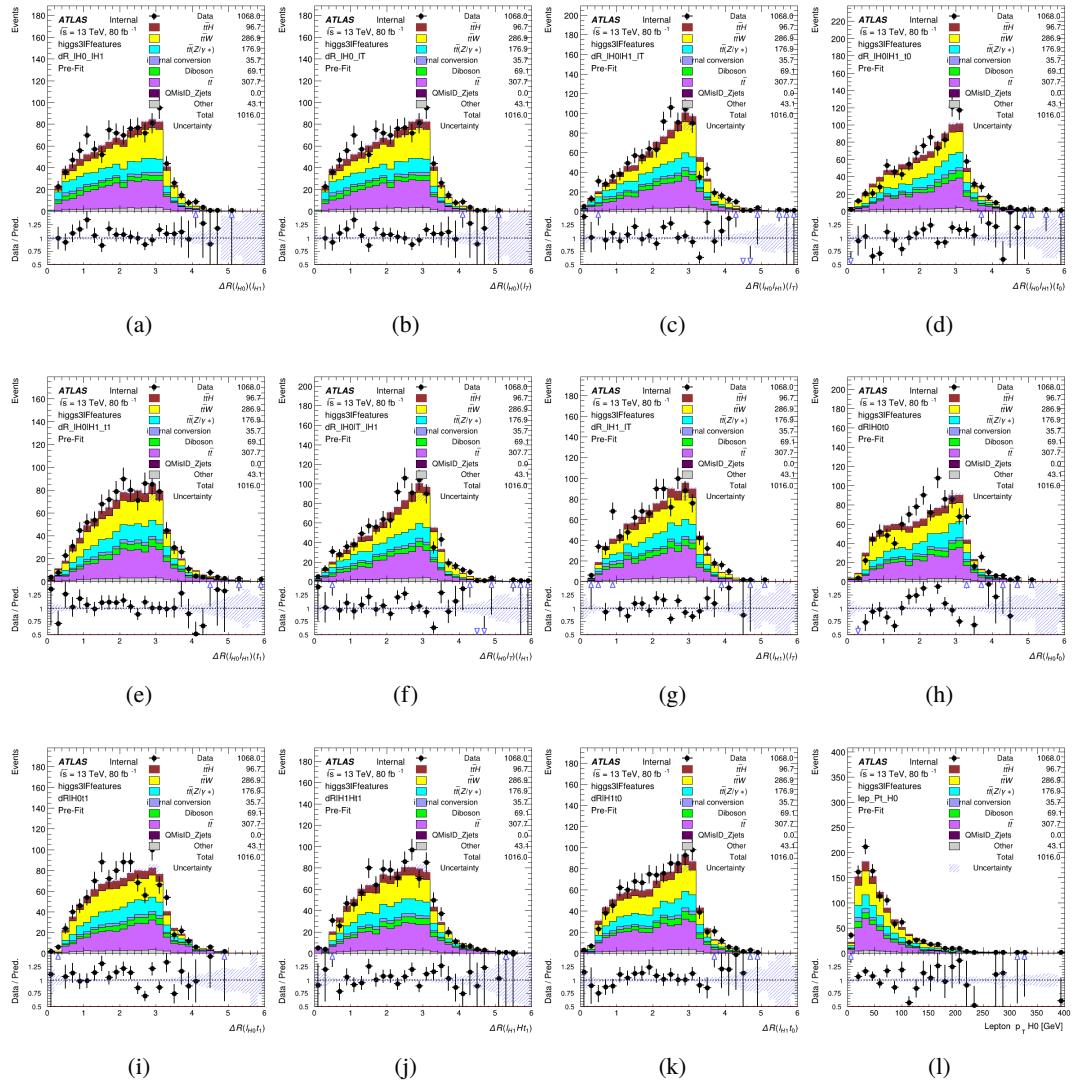


Figure A.17: Input features for higgs3lF

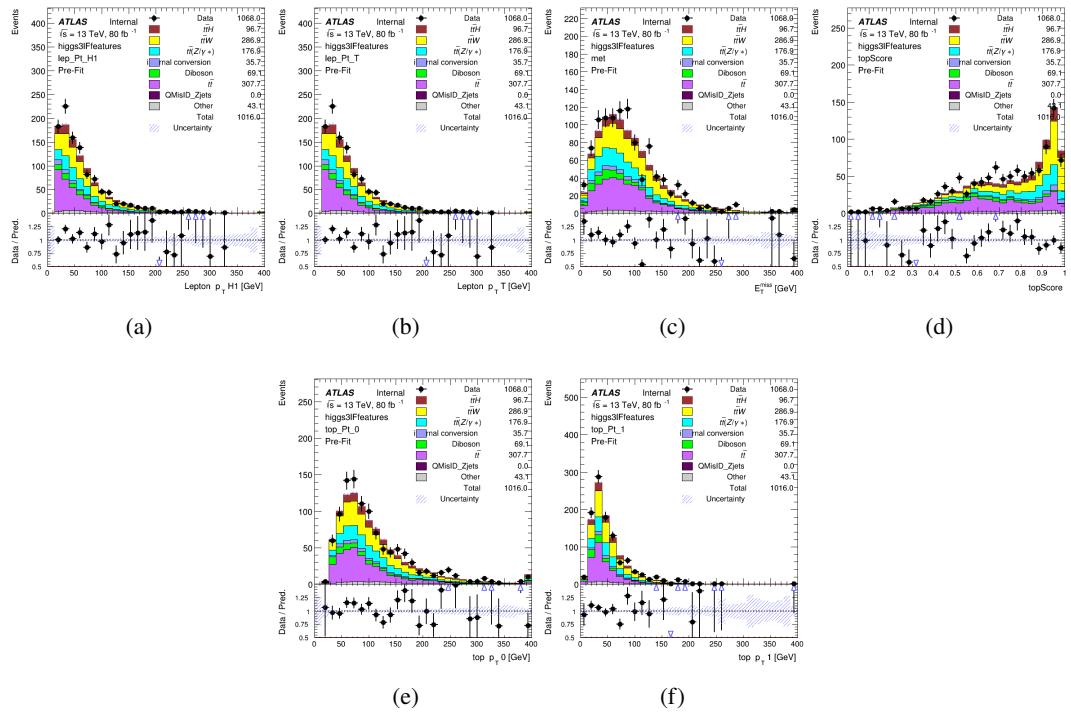
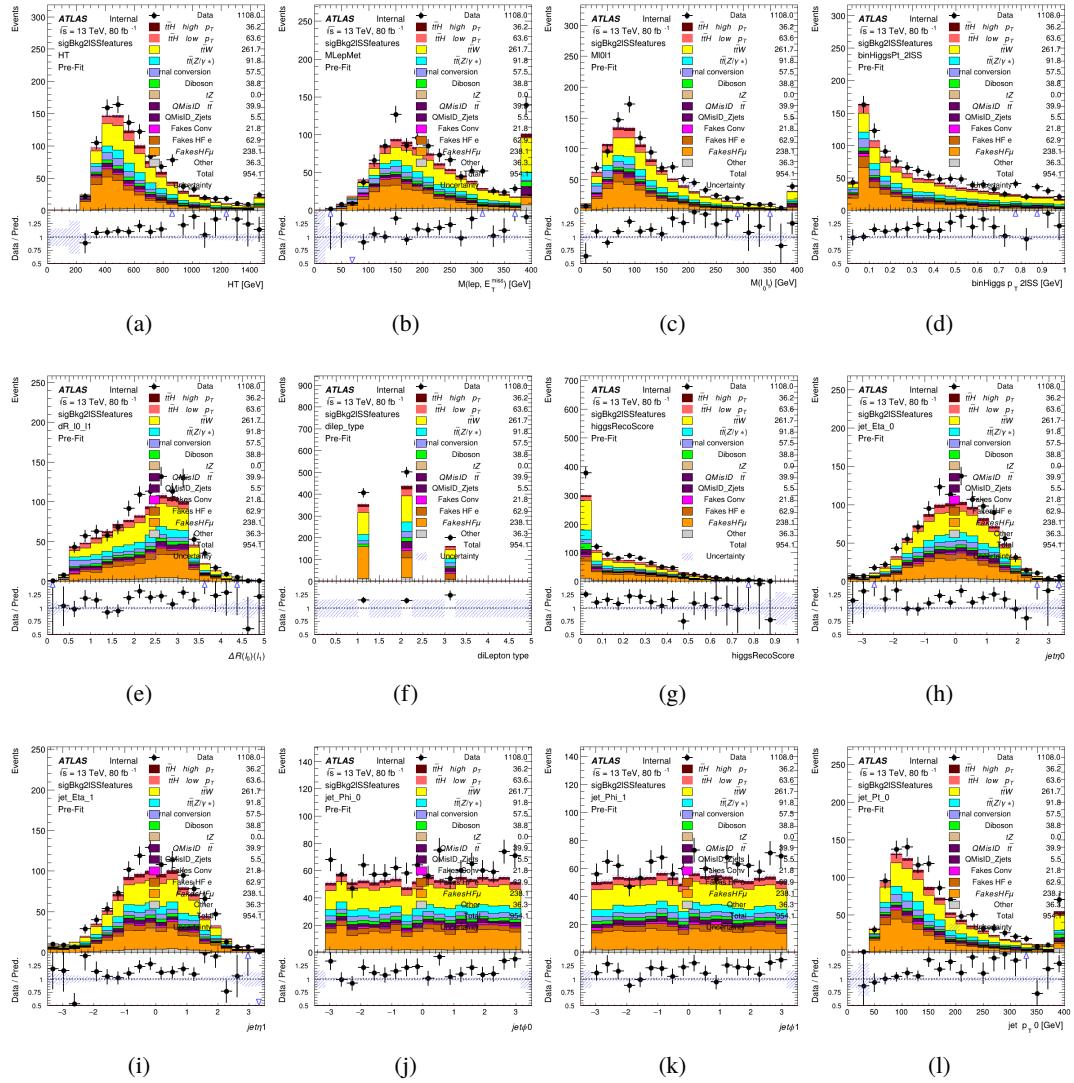


Figure A.18: Input features for higgs3IF

1539 **A.2 Background Rejection MVAs**

1540 **A.2.1 Background Rejection MVA Features - 2lSS**

Figure A.19: Input features for `sigBkg2lSS`

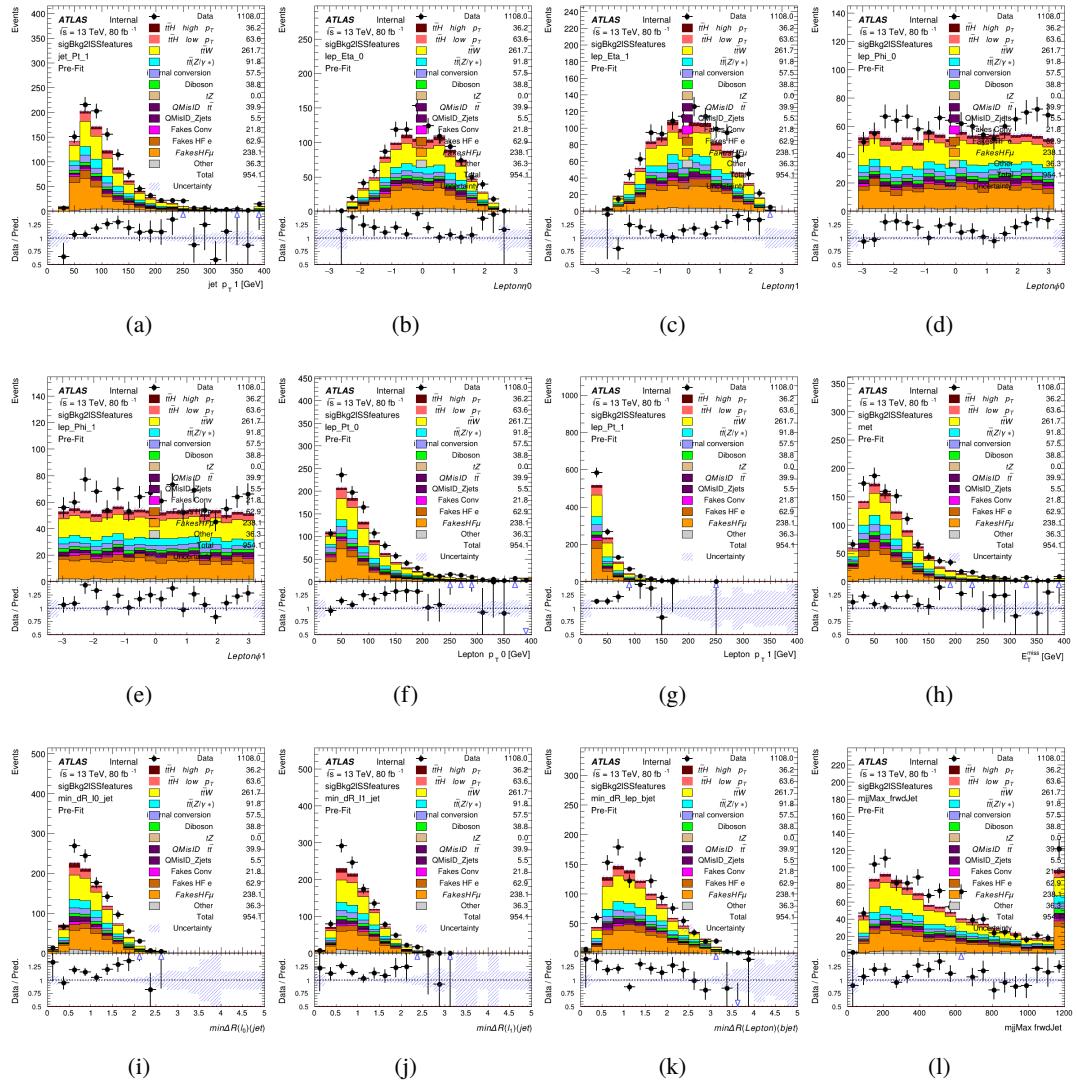


Figure A.20: Input features for sigBkg2lSS

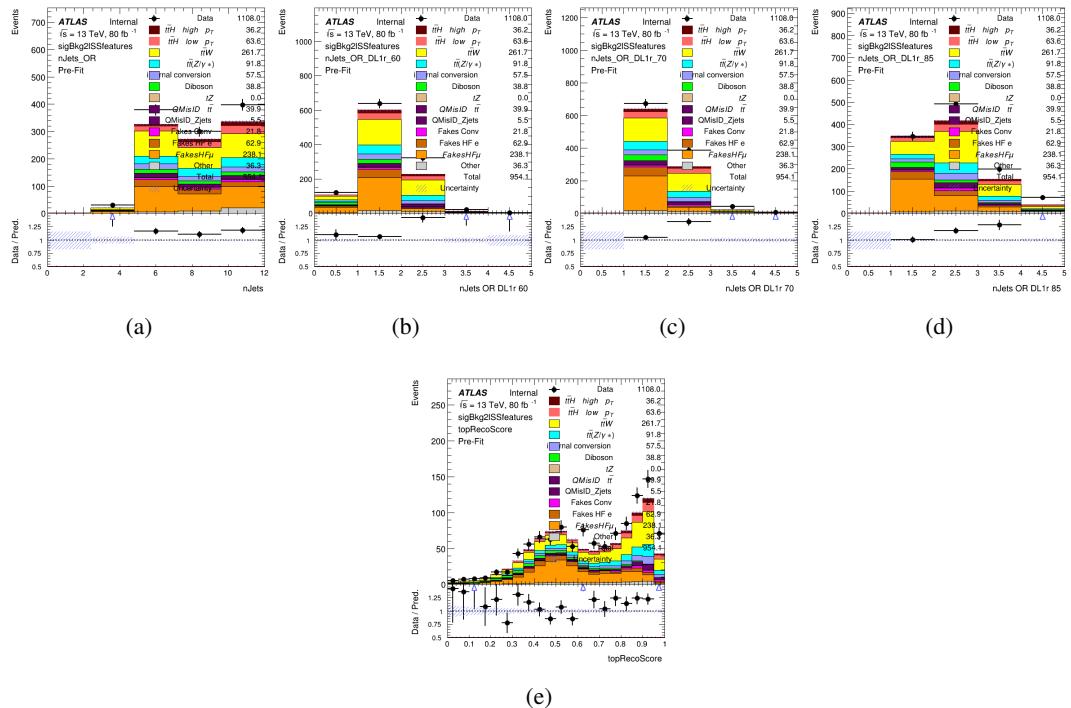


Figure A.21: Input features for sigBkg2ISS

¹⁵⁴¹ **A.2.2 Background Rejection MVA Features - 3l**

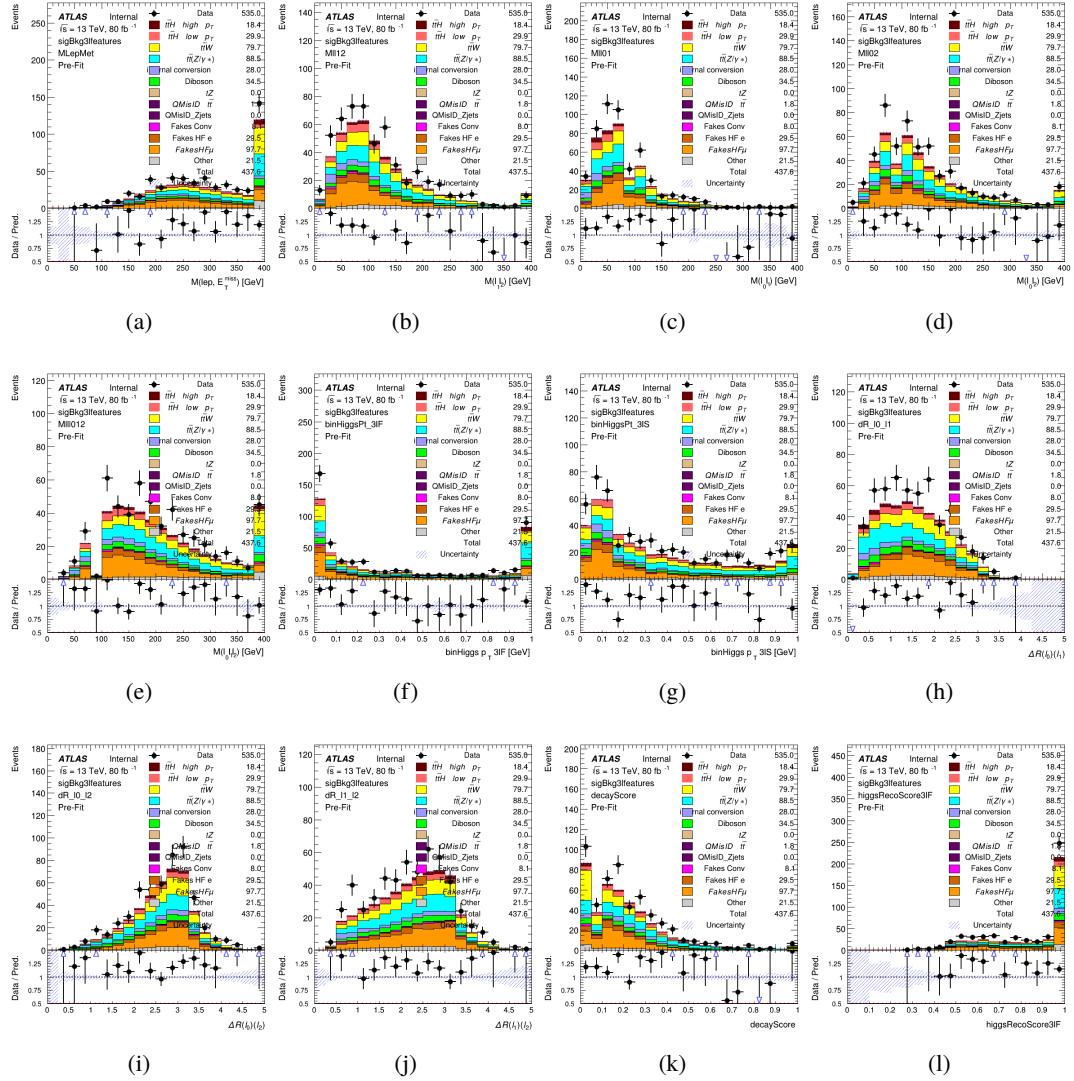


Figure A.22: Input features for sigBkg3l

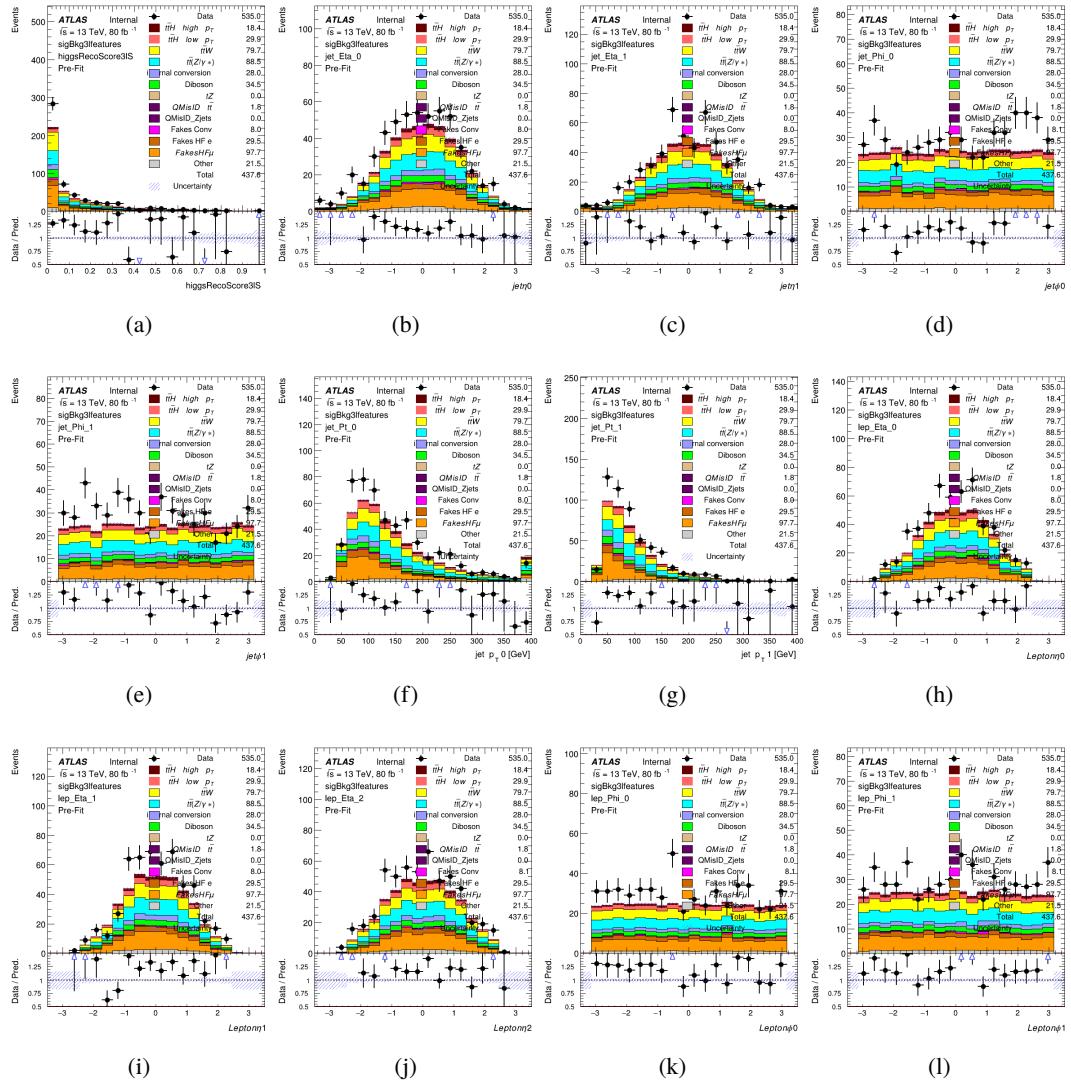


Figure A.23: Input features for sigBkg3l

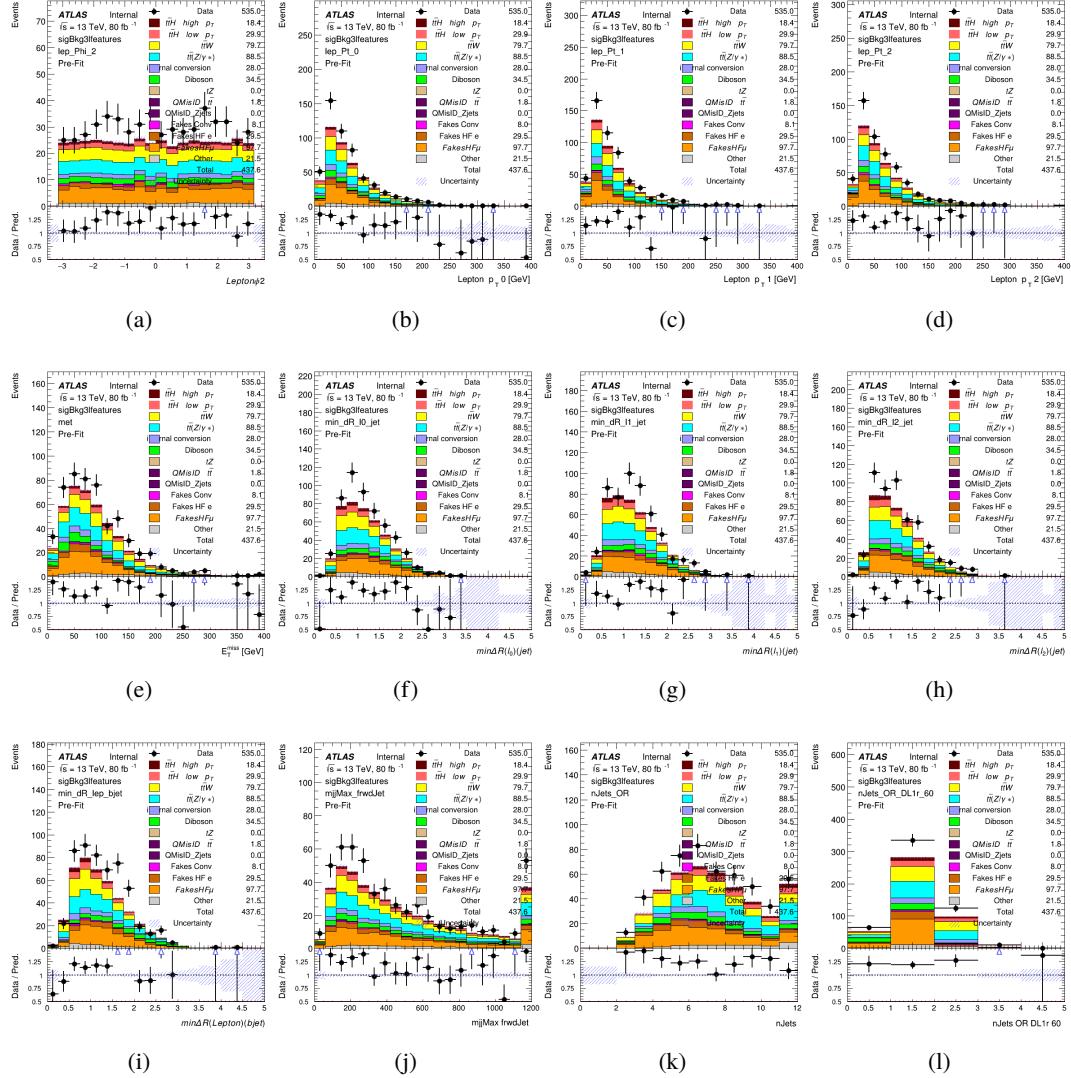


Figure A.24: Input features for sigBkg3l

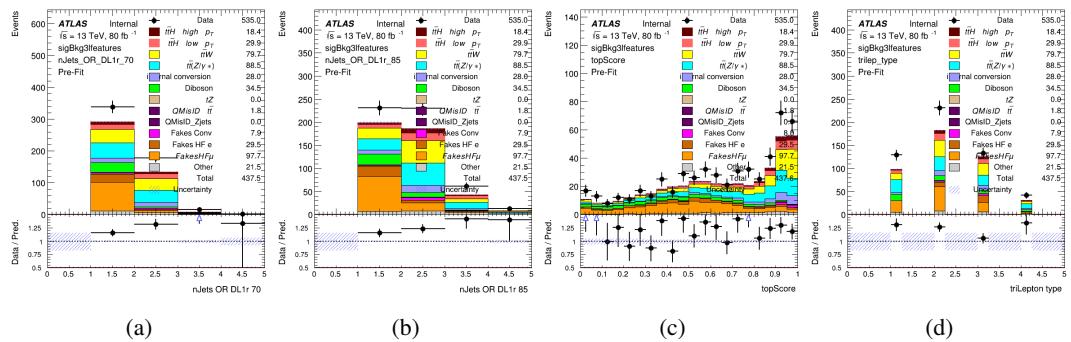


Figure A.25: Input features for sigBkg3l

1542 **A.3 Alternate b-jet Identification Algorithm**

1543 The nominal analysis reconstructs the b-jets by considering different combinations of jets, and
 1544 asking a neural network to determine whether each combination consists of b-jets from top quark
 1545 decays. An alternate approach would be to give the neural network about all of the jets in an event
 1546 at once, and train it to select which two are most likely to be the b-jets from top decay. It was
 1547 hypothesized that this could perform better than considering each combination independently, as
 1548 the neural network could consider the event as a whole. While this is not found to be the case,
 1549 these studies are documented here as a point of interest and comparison.

1550 For these studies, the kinematics of the 10 highest p_T jets in each event are used for
 1551 training. This includes the vast majority of truth b-jets. Specifically the p_T , η , ϕ , E , and DL1r
 1552 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.
 1553 The p_T , η , ϕ , and E of the leptons and E_T^{miss} are included as well. Categorical cross entropy is
 1554 used as the loss function.

Table 50: Accuracy of the NN in identifying b-jets from tops in 2lSS events for the alternate categorical method compared to the nominal method.

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

1555 **A.4 Binary Classification of the Higgs p_T**

1556 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer
 1557 resolution. This means separating high and low p_T events is sufficient for this analysis. As

1558 such, rather than attempting the reconstruct the full Higgs p_T spectrum, a binary classification
 1559 approach is explored.

1560 A model is built to determine whether $t\bar{t}H$ events include a high p_T (>150 GeV) or low
 1561 p_T (<150 GeV) Higgs Boson. While this is now a classification model, it uses the same input
 1562 features described in section 19.4. Binary crossentropy is used as the loss function.

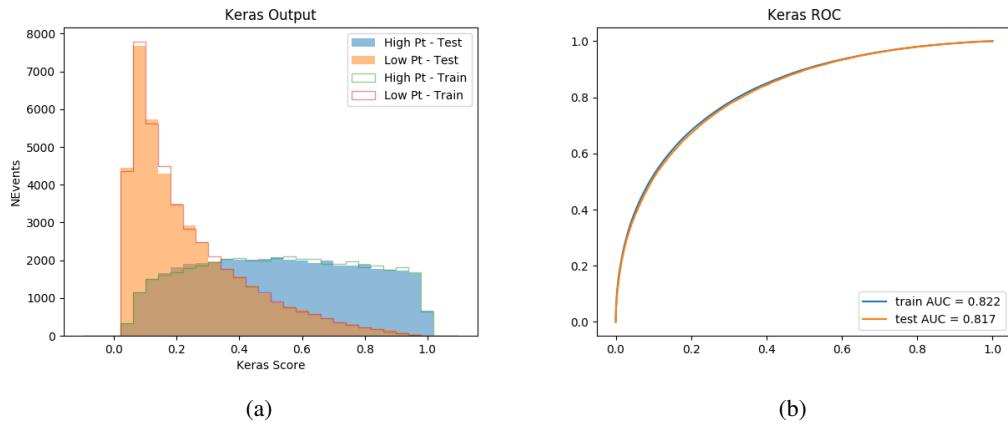


Figure A.26:

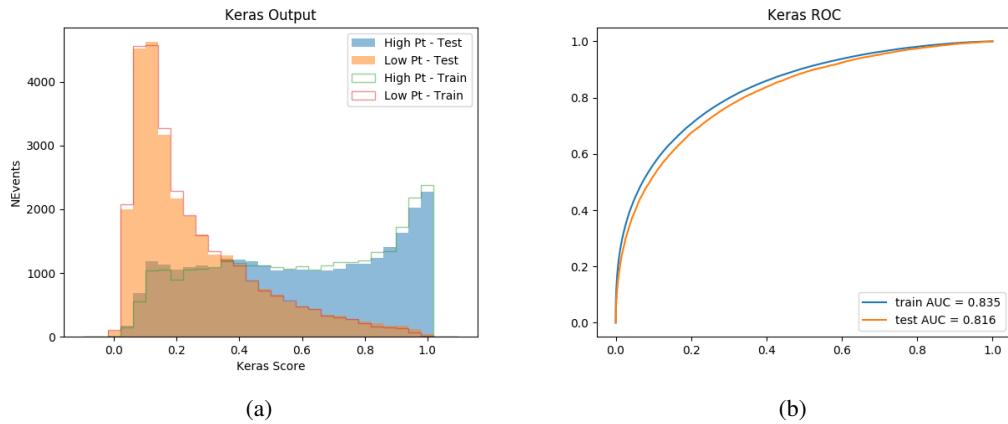


Figure A.27:

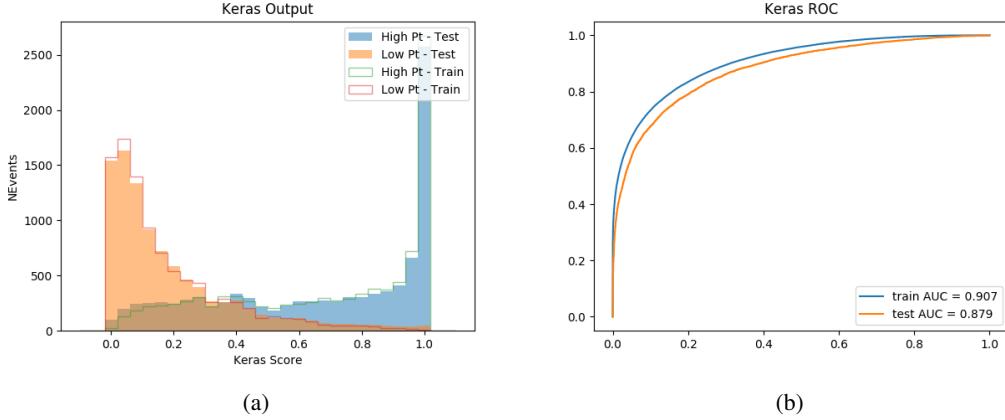


Figure A.28:

1563 A.5 Impact of Alternative Jet Selection

1564 A relatively low p_T threshold of 15 GeV is used to determine jet candidates, as the jets originating
 1565 from the Higgs decay are found to fall between 15 and 25 GeV a large fraction of the time. The
 1566 impact of different jet p_T cuts on our ability to reconstruct the Higgs p_T is explored here. The
 1567 performance of the Higgs p_T prediction models is evaluated for jet p_T cuts of 10, 15, 20, and 25
 1568 GeV.

1569 B

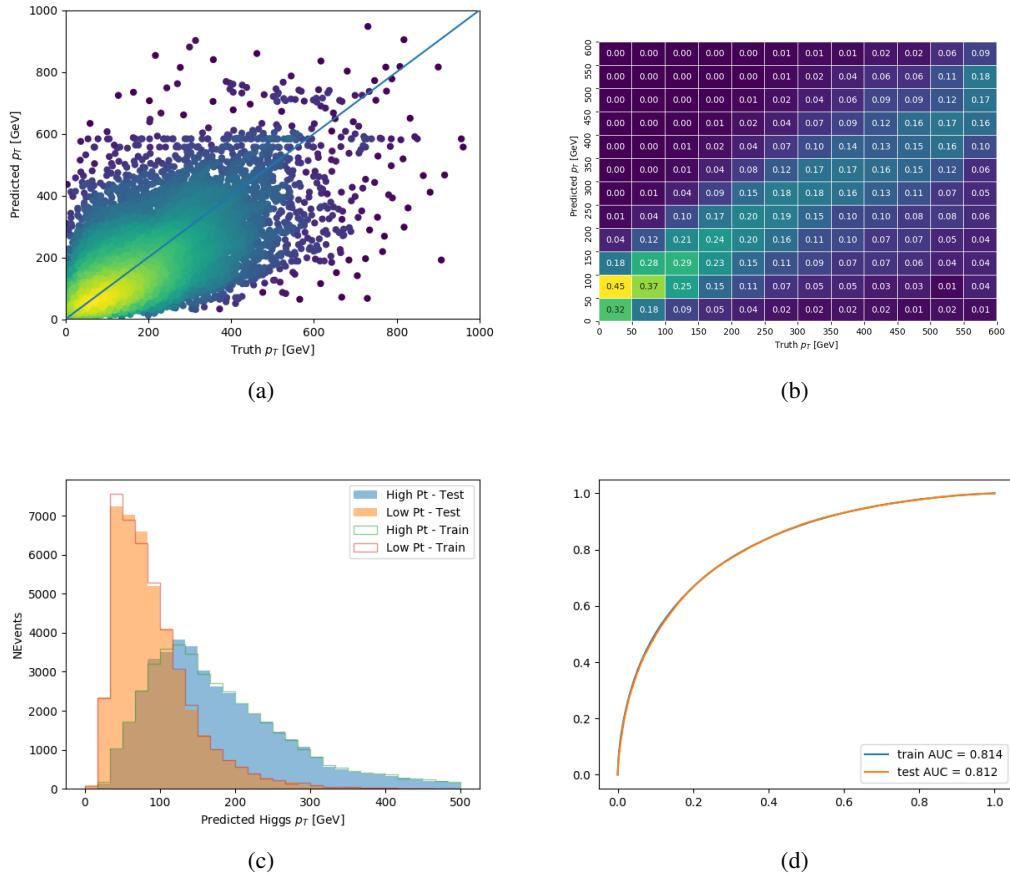


Figure A.29: