



ATLAS Note

GROUP-2017-XX



24th March 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	34
47	8.6	Overlap removal	34
48	9	Event Selection and Signal Region Definitions	35
49	9.1	Event Preselection	35
50	9.2	Fit Regions	40
51	9.3	Non-Prompt Lepton Estimation	57
52	9.3.1	$t\bar{t}$ Validation	57
53	9.3.2	Z+jets Validation	63
54	10	tZ Interference Studies and Separation Multivariate Analysis	68
55	10.1	Top Mass Reconstruction	68
56	10.2	tZ BDT	70
57	11	Data and Monte Carlo Samples	74
58	11.1	Data Samples	75
59	11.2	Monte Carlo Samples	76
60	12	Object Reconstruction	78
61	12.1	Trigger	79
62	12.2	Light leptons	79
63	12.3	Jets	80
64	12.4	B-tagged Jets	81
65	12.5	Missing transverse energy	83
66	12.6	Overlap removal	83
67	13	Event Selection and Signal Region Definitions	84
68	13.1	Event Preselection	84
69	13.2	Fit Regions	89
70	13.3	Non-Prompt Lepton Estimation	106
71	13.3.1	$t\bar{t}$ Validation	106
72	13.3.2	Z+jets Validation	112
73	14	tZ Interference Studies and Separation Multivariate Analysis	117
74	14.1	Top Mass Reconstruction	117
75	14.2	tZ BDT	119
76	15	Systematic Uncertainties	123
77	16	Results	132
78	16.1	1-jet Fit Results	133
79	16.2	2-jet Fit Results	142

80	V Differential Studies of $t\bar{t}H$ Multilepton	151
81	17 Data and Monte Carlo Samples	151
82	17.1 Data Samples	152
83	17.2 Monte Carlo Samples	152
84	18 Object Reconstruction	153
85	18.1 Trigger Requirements	153
86	18.2 Light Leptons	154
87	18.3 Jets	155
88	18.4 Missing Transverse Energy	155
89	19 Higgs Momentum Reconstruction	156
90	19.1 Decay Candidate Reconstruction	157
91	19.2 b-jet Identification	158
92	19.2.1 2lSS Channel	159
93	19.2.2 3l Channel	163
94	19.3 Higgs Reconstruction	166
95	19.3.1 2lSS Channel	167
96	19.3.2 3l Semi-leptonic Channel	171
97	19.3.3 3l Fully-leptonic Channel	175
98	19.4 p_T Prediction	178
99	19.4.1 2lSS Channel	180
100	19.4.2 3l Semi-leptonic Channel	183
101	19.4.3 3l Fully-leptonic Channel	185
102	19.5 3l Decay Mode	189
103	20 Signal Region Definitions	191
104	20.1 Pre-MVA Event Selection	192
105	20.2 Event MVA	197
106	20.3 Signal Region Definitions	203
107	21 Background Rejection MVA	207
108	21.1 Pre-MVA Event Selection	208
109	21.2 Event MVA	209
110	21.3 Signal Region Definitions	215
111	21.3.1 2lSS	218
112	21.3.2 3l – Semi – leptonic	218
113	21.3.3 3l – Fully – leptonic	218
114	22 Systematic Uncertainties	218
115	23 Results	220
116	23.1 Results - 80 fb^{-1}	220
117	23.2 Projected Results - 140 fb^{-1}	223

118	VI Conclusion	225
119	Appendices	227
120	A Machine Learning Models	227
121	A.1 Higgs Reconstruction Models	227
122	A.1.1 b-jet Identification Features - 2lSS	227
123	A.1.2 b-jet Identification Features - 3l	231
124	A.1.3 Higgs Reconstruction Features - 2lSS	236
125	A.1.4 Higgs Reconstruction Features - 3lS	241
126	A.1.5 Higgs Reconstruction Features - 3lF	246
127	A.2 Background Rejection MVAs	250
128	A.2.1 Background Rejection MVA Features - 2lSS	250
129	A.2.2 Background Rejection MVA Features - 3l	254
130	A.3 Alternate b-jet Identification Algorithm	259
131	A.4 Binary Classification of the Higgs p_T	259
132	A.5 Impact of Alternative Jet Selection	261

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

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

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

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

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

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

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

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

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

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

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

180 **Part II**

181 **Theoretical Motivation**

182 **2 The Standard Model and the Higgs Boson**

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

¹⁸⁸ theory, invariant under the Lie Group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where C refers to color
¹⁸⁹ charge, L, the helicity of the particle, and Y, the hypercharge.

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

¹⁹¹ The SM particles, summarized in figure 2.1, can be classified into two general categories based
¹⁹² on their spin: fermions, and bosons.

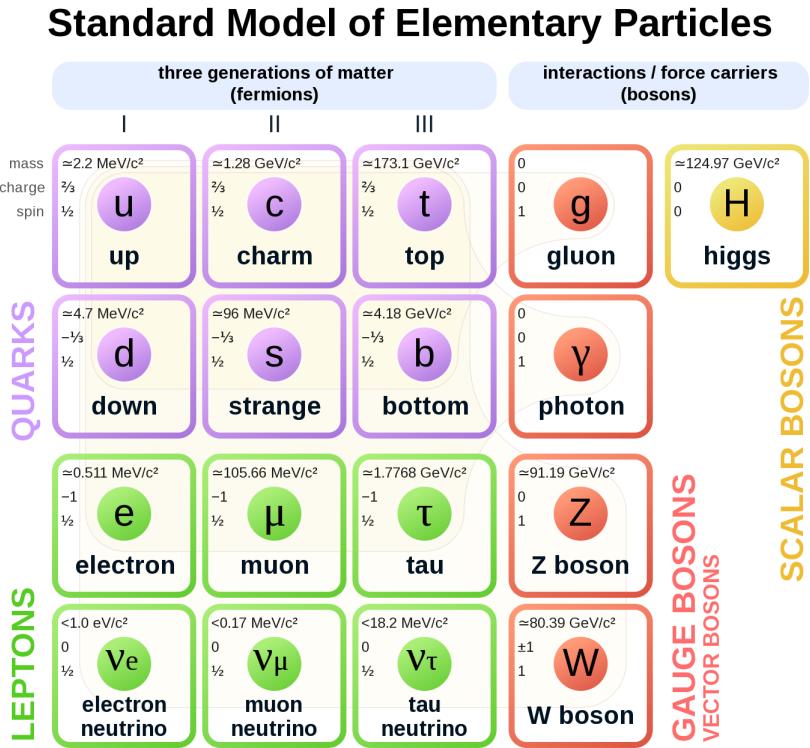


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. []

¹⁹³ Fermions are particles with $\frac{1}{2}$ -integer spin, which according to the spin-statistics theorem,
¹⁹⁴ causes them to comply with the Pauli-exclusion principle []. They can be separated into two

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

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

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

204 And for the quarks:

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

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

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

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

217 **2.2 The Higgs Mechanism**

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

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

²²⁷ **2.2.1 The Higgs Field**

²²⁸ The Higgs mechanism introduces a complex scalar $SU(2)$ doublet, Φ , with the form:

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

²²⁹ This field introduces a scalar potential to the Lagrangian of the form:

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

²³⁰ Where μ and λ are free parameters of the new field. This represents the most general
²³¹ potential allowed while preserving $SU(2)_L$ invariance and renormalizability. In the case that
²³² $\mu^2 < 0$, this potential takes the form shown in figure 2.2.

²³³ The significant feature of this potential is that its minimum does not occur for a value of
²³⁴ $\Phi = 0$. Instead, it is minimized when $|\Phi^\dagger \Phi| = -\mu^2/\lambda$. This means that in its ground state, the
²³⁵ Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while
²³⁶ the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

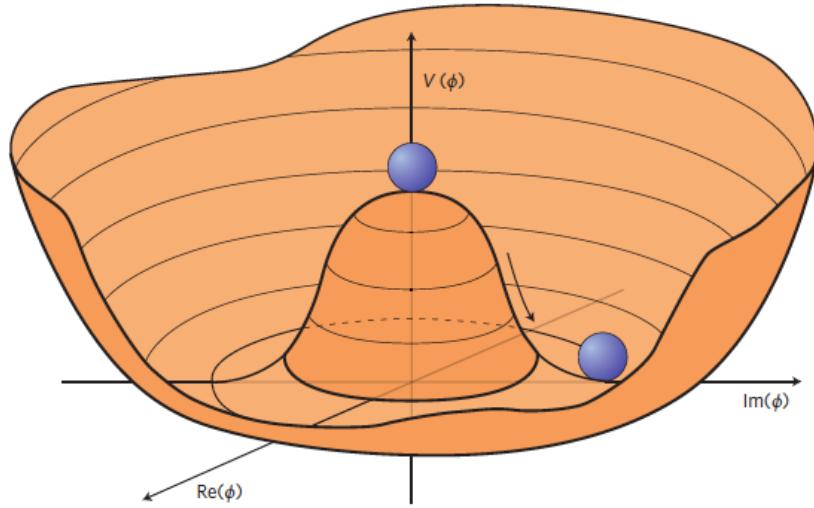


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ , for the case that $\mu^2 < 0$ [].

²³⁷ the minimum is determined only by $\Phi^\dagger \Phi$, there is some ambiguity in the particular definition of
²³⁸ the VEV, but it is generally represented as

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

²³⁹ The full value of Φ can be written as

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

²⁴⁰ with v being the value of the VEV, and H being the real value of the scalar field.

²⁴¹ **2.2.2 Electroweak Symmetry Breaking**

²⁴² The Electroweak (EWK) interaction is described in the SM by a $SU(2)_L \otimes U(1)_Y$ gauge theory.
²⁴³ 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,
²⁴⁴ B_μ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar
²⁴⁵ field Φ in the Lagrangian:

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

²⁴⁶ Here D_μ represents the covariant derivative required to preserve gauge invariance, g and
²⁴⁷ g' represent coupling constant of the gauge bosons, σ^a denotes the Pauli matrices of $SU(2)$,
²⁴⁸ and Y represents the hypercharge of $U(1)$. The terms in this interaction which contribute to the
²⁴⁹ masses of the gauge bosons can be written as:

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

²⁵⁰ Expanding these terms into the mass eigenstates of the electroweak interaction yields four
²⁵¹ physical gauge bosons, two charged and two neutral, which are linear combinations of the fields

²⁵² $W_\mu^1, W_\mu^2, W_\mu^3$, and B_μ :

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

²⁵³ And the masses of these fields are given by:

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

²⁵⁴ This produces exactly the particles we observe - three massive gauge bosons and a single
²⁵⁵ massless photon. The massless photon represents the portion of the gauge symmetry, a single
²⁵⁶ $U(1)$ of the electromagnetic force, that remains unbroken by the VEV.

²⁵⁷ Interactions with the Higgs field also lead to the generation of the fermion masses, which
²⁵⁸ in the Lagrangian take the form:

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

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

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

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

272 2.3 Limitations of the Standard Model

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

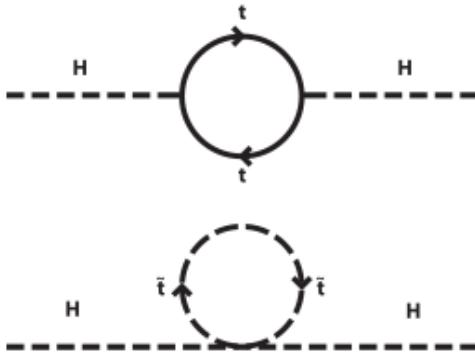


Figure 2.3: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram [6].

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

²⁷⁷ Higher dimension operators are a common way to paramaterize the effects of physics at very
²⁷⁸ high energies into

²⁷⁹ **3.1 Extensions to the Higgs Sector**

²⁸⁰ **3.2 Six Dimensional Operators**

²⁸¹ While the SM has been tested to great precision, particularly at the LHC, it is generally accepted
²⁸² that it is only valid up to a certain energy scale. It is assumed that above a certain energy, at the
²⁸³ scale where something like a Grand Unified Theory (GUT) or quantum gravity become relevant,
²⁸⁴ the SM will not be applicable.

Part III**286 The LHC and the ATLAS Detector****287 4 The LHC**

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

297 The LHC first began running in 2009 at a proton-proton center of mass energy of $\sqrt{s} = 8$
298 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during
299 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,
300 and collected data at an increased energy of $\sqrt{s} = 13$ TeV until 2018, a period referred to as Run
301 2.

302 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.

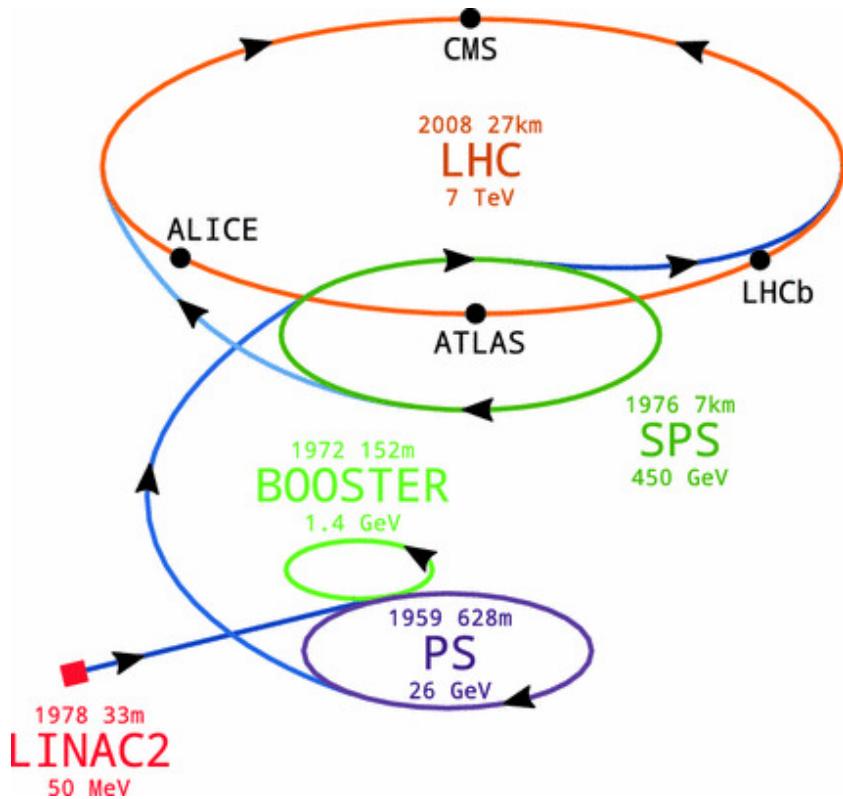


Figure 4.1: A summary of the accelerator chain used to feed protons into the LHC [].

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

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

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

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

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

321 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around
 322 140 fb^{-1} , exceeding the expected integrated luminosity of 100 fb^{-1} .

323 5 The ATLAS Detector

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

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

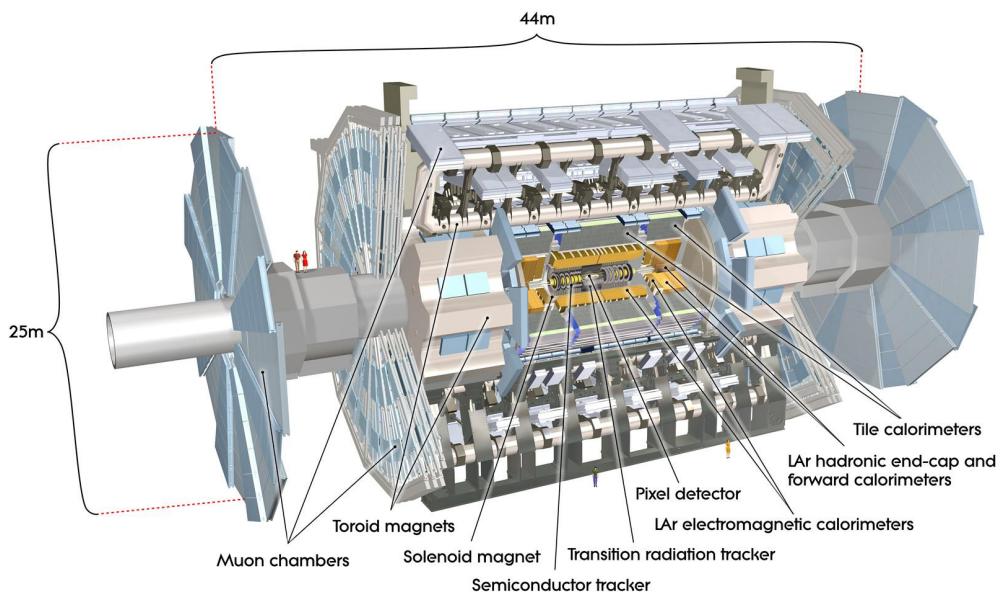


Figure 5.1: Cutaway view of the ATLAS detector, with labels of its major components [].

³³² **5.1 Inner Detector**

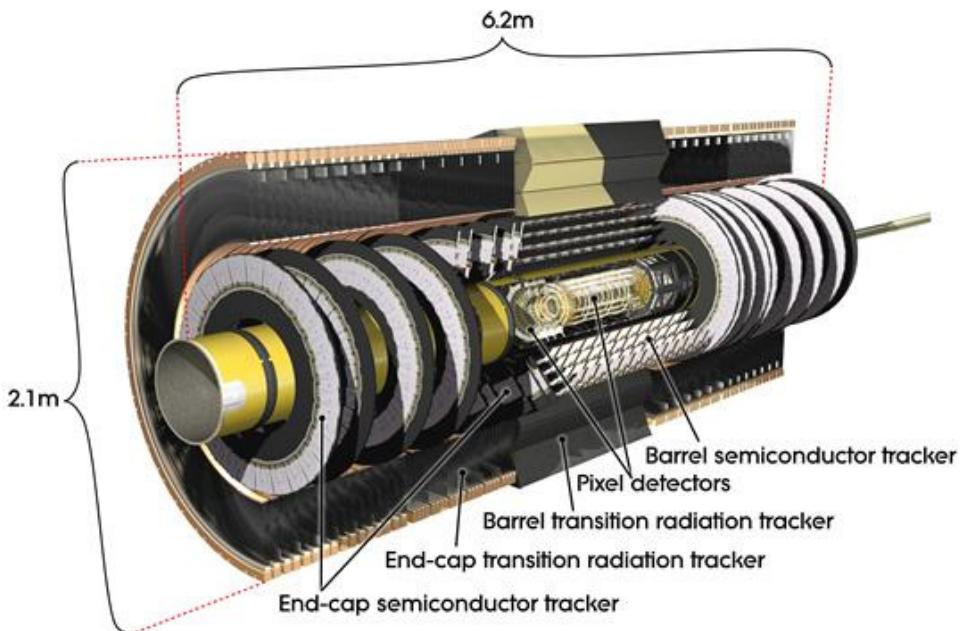


Figure 5.2: Cutaway view of the Inner Detector [].

³³³ Just surrounding the interaction point is the Inner Detector, designed to track the path
³³⁴ of charged particles moving through the detector. An inner solenoid surrounding the Innder
³³⁵ Detector is used to produces a magnetic field of 2 T. This large magnetic field causes the path
³³⁶ of charged particles moving through the Inner Detector to bend. Because this magnetic field is
³³⁷ uniform and well known, it can be used in conjunction with the curvature of a particles path to
³³⁸ measure its charge and momentum.

³³⁹ The Inner Detector consists of three components - the Pixel Detector, the Semi-Conductor
³⁴⁰ Tracker (SCT), and the Transition Radiation Tracker (TRT). The Pixel Detector is the innermost
³⁴¹ of these, beginning just 33.25 mm away from the beam line. It consists of three silicon layers

³⁴² along the barrel, as well as three endcap layers, covering a range of $|\eta| < 2.5$.

³⁴³ The Semiconductor Tracker (SCT) is similar to the Pixel detector, but uses long strips
³⁴⁴ rather than small pixel to cover a larger spatial area.

³⁴⁵ **5.2 Calorimeters**

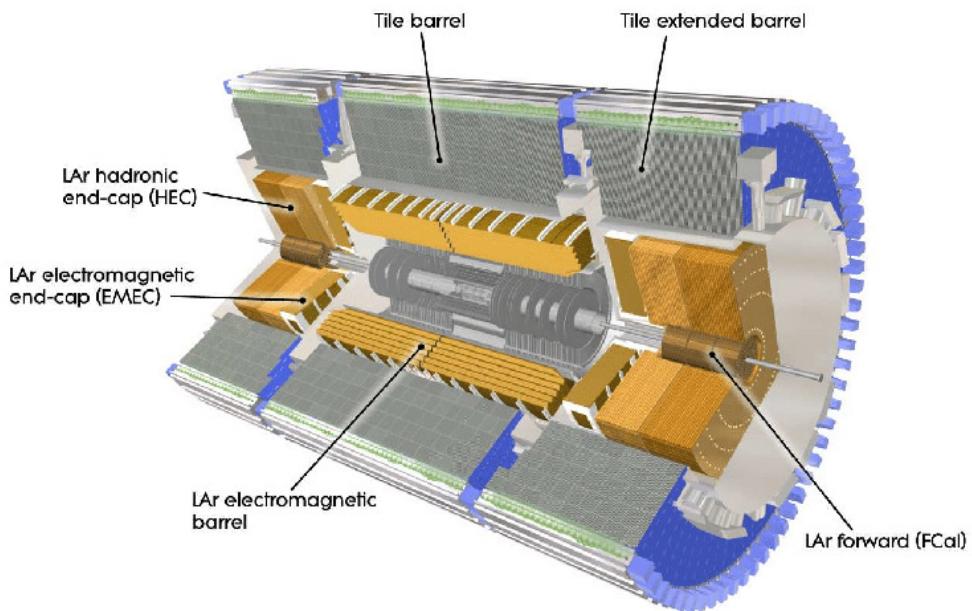


Figure 5.3: Cutaway view of the calorimeter system of the ATLAS detector [].

³⁴⁶ Situated outside the Inner Detector are two concentric calorimeters. The inner calorimeter
³⁴⁷ uses liquid argon (LAr) to measure energy of particles that interact electromagnetically, which
³⁴⁸ includes photons and any charged particle. The LAr calorimeter is made of heavy metals,
³⁴⁹ primarily lead and copper, which causes electromagnetically interacting particles to shower,
³⁵⁰ depositing their energy in the detector. The showering of the high energy particles that pass

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

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

358 **5.3 Muon Spectrometer**

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

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

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

373 1200 tracking chambers are placed in the muon system in order to precisely measure the
374 tracks of muons with high spatial resolution.

375 **5.4 Trigger System**

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

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

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

394 **Part IV**

395 **Measurement of WZ + Heavy Flavor**

396 **6 Introduction**

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

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

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

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

414 Section 11 details the data and Monte Carlo (MC) samples used in the analysis. The
415 reconstruction of various physics objects is described in Section 12. Section 13 describes the
416 event selection applied to these samples, along the definitions of the various regions used in
417 the fit. The multivariate analysis techniques used to separate the tZ background from WZ +
418 heavy flavor are described in Section 14. Section 22 describes the various sources of systematic
419 uncertainties considered in the fit. Finally, the results of the analysis are summarized in Section
420 23, followed by a brief conclusion in Section ??.

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

424 7 Data and Monte Carlo Samples

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

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

435 Samples were then generated from these HIGG8D1 derivations with p-tag of p4134 for data
 436 and p4133 for Monte Carlo using AnalysisBase version 21.2.127 modified to include custom
 437 variables.

438 7.1 Data Samples

439 The study uses a sample of proton-proton collision data collected by the ATLAS detector from
 440 2015 through 2018 at an energy of $\sqrt{s} = 13$ TeV, which represents an integrated luminosity of

441 139 fb^{-1} . This data set was collected with a bunch crossing rate of 25 ns. All data used in this
442 analysis was verified by data quality checks, having been included in the following Good Run
443 Lists:

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

445 _PHYS_StandardGRL_All_Good_25ns.xml

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

447 _PHYS_StandardGRL_All_Good_25ns.xml

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

449 _All_Good_25ns_Triggerno17e33prim.xml

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

451 _All_Good_25ns_Triggerno17e33prim.xml

452 Runs included from the AllYear period containers are included.

453 **7.2 Monte Carlo Samples**

454 Several different generators were used to produce Monte Carlo simulations of the signal and
455 background processes. For all samples, the response of the ATLAS detector is simulated using
456 Geant4. The WZ signal samples are simulated using Sherpa 2.2.2 [[sherpa](#)]. Specific information

⁴⁵⁷ about the Monte Carlo samples being used can be found in Table 28. A list of the specific samples
⁴⁵⁸ used by data set ID is shown in Table 10.

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, t̄tt̄	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄t	Powheg-BOX v2 [powhegtt]	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.

459 8 Object Reconstruction

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

463 **8.1 Trigger**

464 Events are required to be selected by dilepton triggers, as summarized in Table 29.

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.

465 **8.2 Light leptons**

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

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

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

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

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

⁴⁷¹ $|\eta| < 2.5$. Candidates in the transition region between different electromagnetic calorimeter
⁴⁷² components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant com-
⁴⁷³ bining shower shape and track information is used to distinguish real electrons from hadronic
⁴⁷⁴ showers (fake electrons). To further reduce the non-prompt electron contribution, the track is
⁴⁷⁵ required to be consistent with originating from the primary vertex; requirements are imposed
⁴⁷⁶ on the transverse impact parameter significance ($|d_0|/\sigma_{d_0} < 5$) and the longitudinal impact
⁴⁷⁷ parameter ($|\Delta z_0 \sin \theta_\ell| < 0.5 \text{ mm}$).

⁴⁷⁸ Muon candidates are reconstructed by combining inner detector tracks with track segments
⁴⁷⁹ or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
⁴⁸⁰ $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$. The longitudinal impact parameter is the same for both electrons
⁴⁸¹ and muons, while muons are required to pass a slightly tighter transverse impact parameter,
⁴⁸² $|d_0|/\sigma_{d_0} < 3$.

⁴⁸³ All leptons are required to be isolated, as defined through the standard PLVLoose working
⁴⁸⁴ point supported by combined performance groups. Leptons are additionally required to pass a
⁴⁸⁵ non-prompt BDT selection described in detail in [**ttH_paper**]. Optimized working points and
⁴⁸⁶ scale factors for this BDT are taken from that analysis.

⁴⁸⁷ 8.3 Jets

⁴⁸⁸ Jets are reconstructed from calibrated topological clusters built from energy deposits in the
⁴⁸⁹ calorimeters [**ATL-PHYS-PUB-2015-015**], using the anti- k_t algorithm with a radius para-

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

496 **8.4 B-tagged Jets**

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

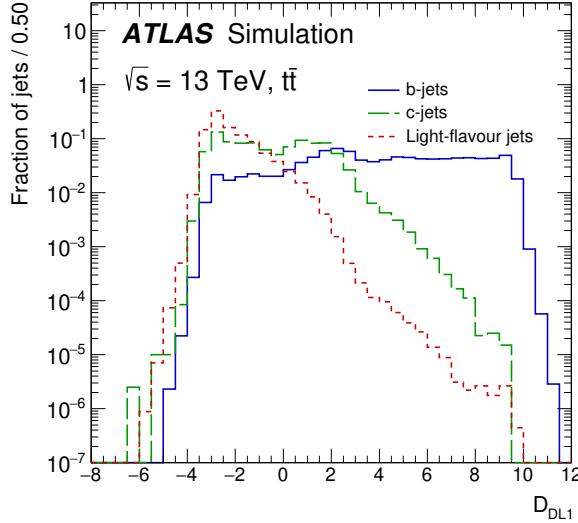


Figure 8.1: Output distribution of the DL1r algorithm for b-jets, charm jets, and light jets

502 From the output of the BDT, working points (WPs) are developed based on the efficiency
 503 of truth b-jets at particular values of the DL1r algorithm. These WPs are taken from the March
 504 2020 CDI file, 2020-21-13TeV-MC16-CDI-2020-03-11_v2.root. The working points used in
 505 this analysis are summarized in Table 12.

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

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

506 A tighter WP will accept fewer b-jets, but reject a higher fraction of charm and light jets.
 507 Generally, analyses that include b-jets will use a fixed working point, for example, requiring that
 508 a jet pass the 70% threshold. By instead treating these working point as bins, e.g. events with
 509 jets that fall between the 85% and 77% WPs fall into one bin, while events with jets passing the

510 60% WP fall into another, and looking at the full psuedo-continuous DL1r spectrum of the jets,
511 additional information can be gained. The psuedo-continuous b-tag spectrum is used in this case
512 to separate out WZ + b, WZ + charm, and WZ + light.

513 8.5 Missing transverse energy

514 Missing transverse momentum (E_T^{miss}) is used as part of the event selection. The missing
515 transverse momentum vector is defined as the inverse of the sum of the transverse momenta of
516 all reconstructed physics objects as well as remaining unclustered energy, the latter of which is
517 estimated from low- p_T tracks associated with the primary vertex but not assigned to a hard object,
518 with object definitions taken from [ATL-PHYS-PUB-2015-027]. Light leptons considered in
519 the E_T^{miss} reconstruction are required to have $p_T > 10 \text{ GeV}$, while jets are required to have $p_T > 20$
520 GeV .

521 8.6 Overlap removal

522 To avoid double counting objects and remove leptons originating from decays of hadrons, overlap
523 removal is performed in the following order: any electron candidate within $\Delta R = 0.1$ of another
524 electron candidate with higher p_T is removed; any electron candidate within $\Delta R = 0.1$ of a muon
525 candidate is removed; any jet within $\Delta R = 0.3$ of an electron candidate is removed; if a muon
526 candidate and a jet lie within $\Delta R = \min(0.4, 0.04 + 10[\text{GeV}] / p_T(\text{muon}))$ of each other, the jet
527 is kept and the muon is removed.

528 This algorithm is applied to the preselected objects. The overlap removal procedure is
 529 summarized in Table 13.

Keep	Remove	Cone size (ΔR)
electron	electron (low p_T)	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	$\min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$
electron	tau	0.2

Table 5: Summary of the overlap removal procedure between electrons, muons, and jets.

530 9 Event Selection and Signal Region Definitions

531 Event are required to pass a preselection described in Section 13.1 and summarized in Table 14.
 532 Those that pass this preselection are divided into various fit regions described in Section 13.2,
 533 based on the number of jets in the event, and the b-tag score of those jets.

534 9.1 Event Preselection

535 Events are required to include exactly three reconstructed light leptons passing the requirement
 536 described in 12.2, which have a total charge of ± 1 . As the opposite sign lepton is found to
 537 be prompt the vast majority of the time [ttH_paper], it is required to be loose and isolated,
 538 as defined though the standard PLVLoose working point supported by combined performance
 539 groups. The same sign leptons are required to be very tightly isolated, as per the recommended
 540 PLVTight.

541 The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose
 542 charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.
 543 the smallest ΔR , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton
 544 0 is required to have $p_T > 10 \text{ GeV}$, while the same sign leptons, 1 and 2, are required to have
 545 $p_T > 20 \text{ GeV}$ to reduce the contribution of non-prompt leptons.

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

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

553 Events are required to have one or two reconstructed jets passing the selection described
 554 in Section 12.3. Events with more than two jets are rejected in order to reduce the contribution
 555 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 $p_T > 20 \text{ GeV}$
 One loose Iso, medium ID opposite charge lepton with $p_T > 10 \text{ GeV}$
 $m(l^+l^-)$ within 10 GeV of 91.2 GeV
 Transverse mass of W-candidate, $m_T(E_T^{\text{miss}} + \text{lep}_{\text{other}}) > 30 \text{ GeV}$
 Missing transverse energy, $E_T^{\text{miss}} > 20 \text{ GeV}$
 One or two jets with $p_T > 25 \text{ GeV}$

Table 6: Summary of the selection applied to events for inclusion in the fit

556 The event yields in the preselection region for both data and Monte Carlo are summarized
 557 in Table 13.1, which shows good agreement between data and Monte Carlo, and demonstrates that
 558 this region consists primarily of WZ events. The WZ events are split into WZ + b, WZ + c, and
 559 WZ + l based on the truth flavor of the associated jet in the event. Specifically, this determination
 560 is made based on the HadronConeExclTruthLabelID of the jet, as recommended by the b-
 561 tagging working group [BtagWG]. In this ordering b-jet supersedes charm, which supersedes
 562 light. That is, WZ + l events contain no charm and no b jets at truth level, WZ + c contain at
 563 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
t̄tW	16.8 ± 2.3
t̄tZ	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
t̄tWW	0.23 ± 0.05
Z + jets	600 ± 260
V + γ	37 ± 54
tZ	190 ± 70
tW	5.5 ± 1.2
WtZ	25.8 ± 1.1
VVV	26.2 ± 0.9
VH	94 ± 7
t̄t	108.68 ± 8
t̄tH	4.3 ± 0.5
Total	10600 ± 530
Data	10574

Table 7: Event yields in the preselection region at 139.0 fb^{-1}

564 Here Other VV represents diboson processes other than WZ, and consists predominantly
565 of $ZZ \rightarrow llll$ events where one of the leptons is not reconstructed.

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

WZ Fit Region - Inclusive

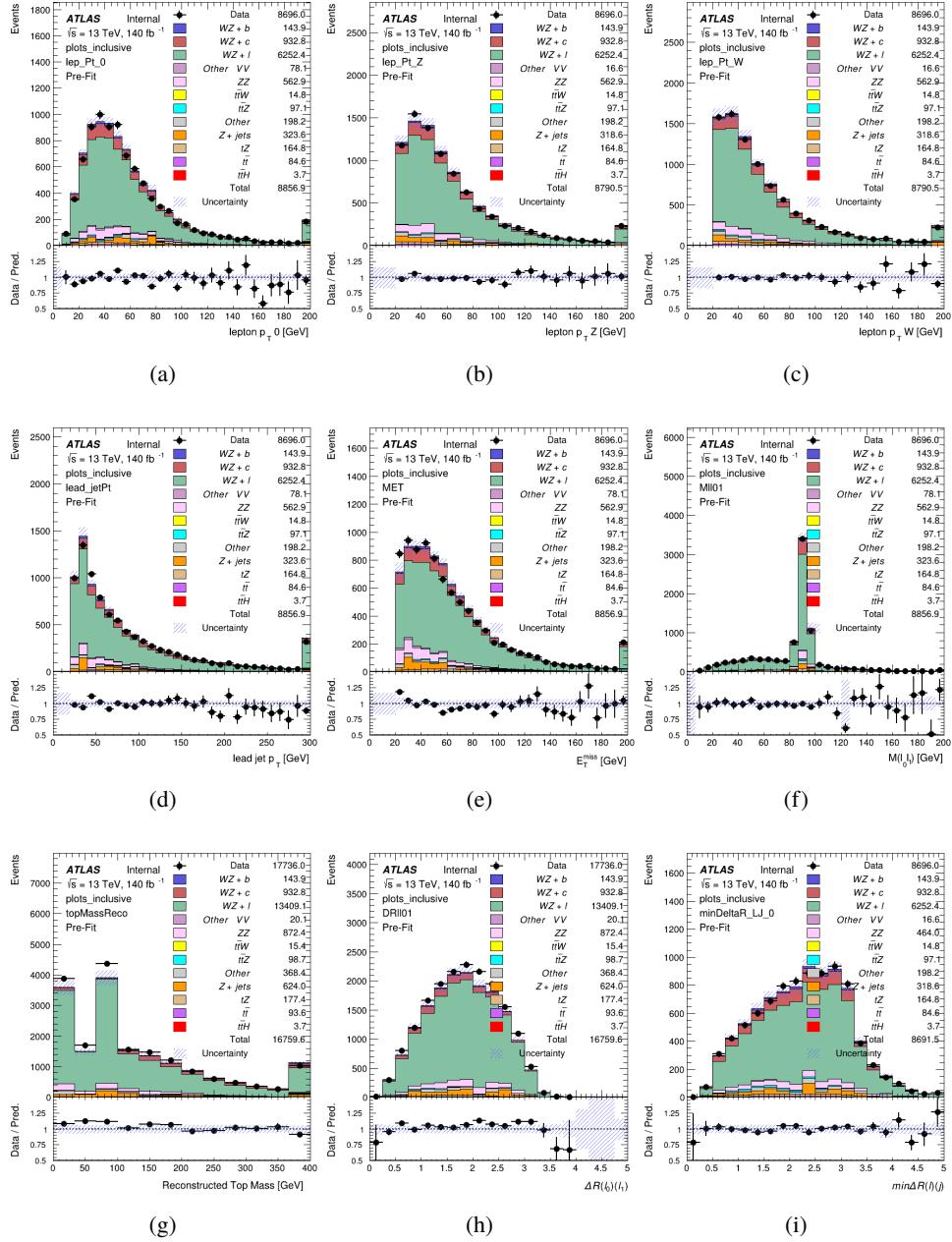


Figure 9.1: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

569 **9.2 Fit Regions**

570 Once preselection has been applied, the remaining events are categorized into one of twelve
 571 orthogonal regions. The regions used in the fit are summarized in Table 16.

Table 8: 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	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} >= 1, tZ \text{ BDT} < 0.725$

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

576 An unfolding procedure is performed to account for differences in the number of recon-
 577 structed jets compared to the number of truth jets in each event. The `AntiKt4TruthDressedWZJets`
 578 truth jet collection is used to make this determination. In order to account for migration of WZ+1-
 579 jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal samples are
 580 separated based on the number of truth jets. Events with 0 jets or more than 3 jets at truth level,

yet fall within one of the categories listed in Table 16, are categorized as WZ + other, and treated as a background. The migration matrix in the number of jets at truth level versus reco level is shown in Figure 13.2. The composition of the number of truth jets in each reco jet bin is taken from MC, with uncertainties in these estimates described in detail in Section 22.

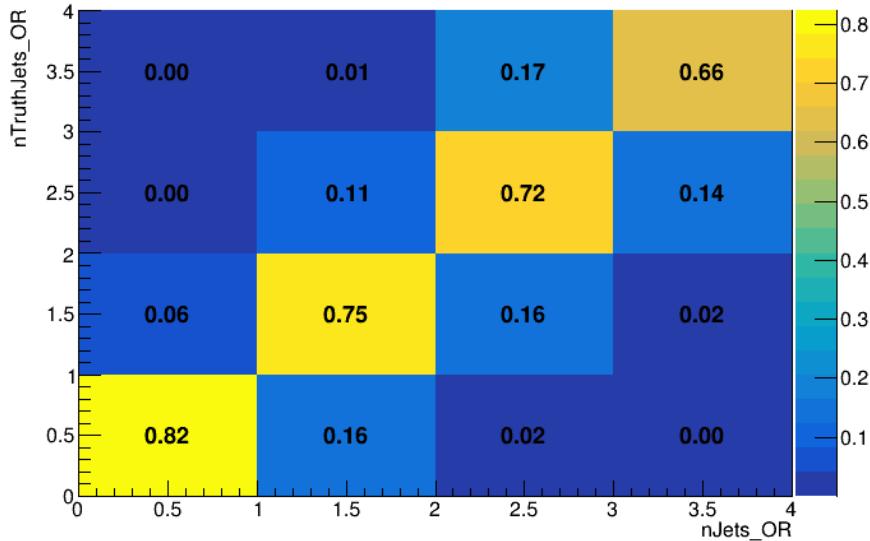


Figure 9.2: Number of truth jets compared to the number of reconstructed jets in each WZ event falling in the preselection region. Each row is normalized to unity.

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

The modeling in each region is validated by comparing data and MC predictions for

⁵⁹¹ various kinematic distributions. These plot are shown in Figures [13.3-13.16](#).

WZ Fit Region - 1j Inclusive

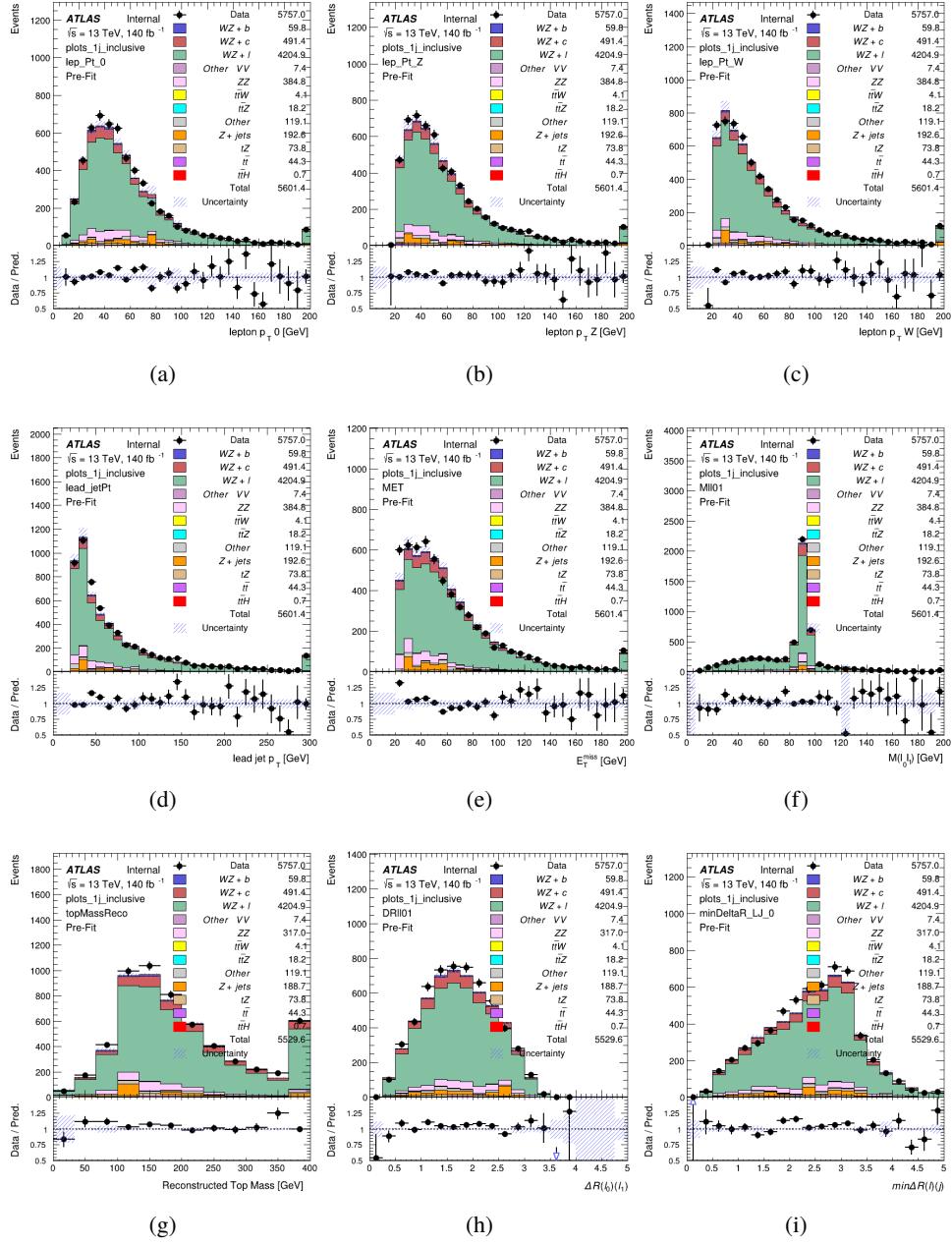


Figure 9.3: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j < 85% WP

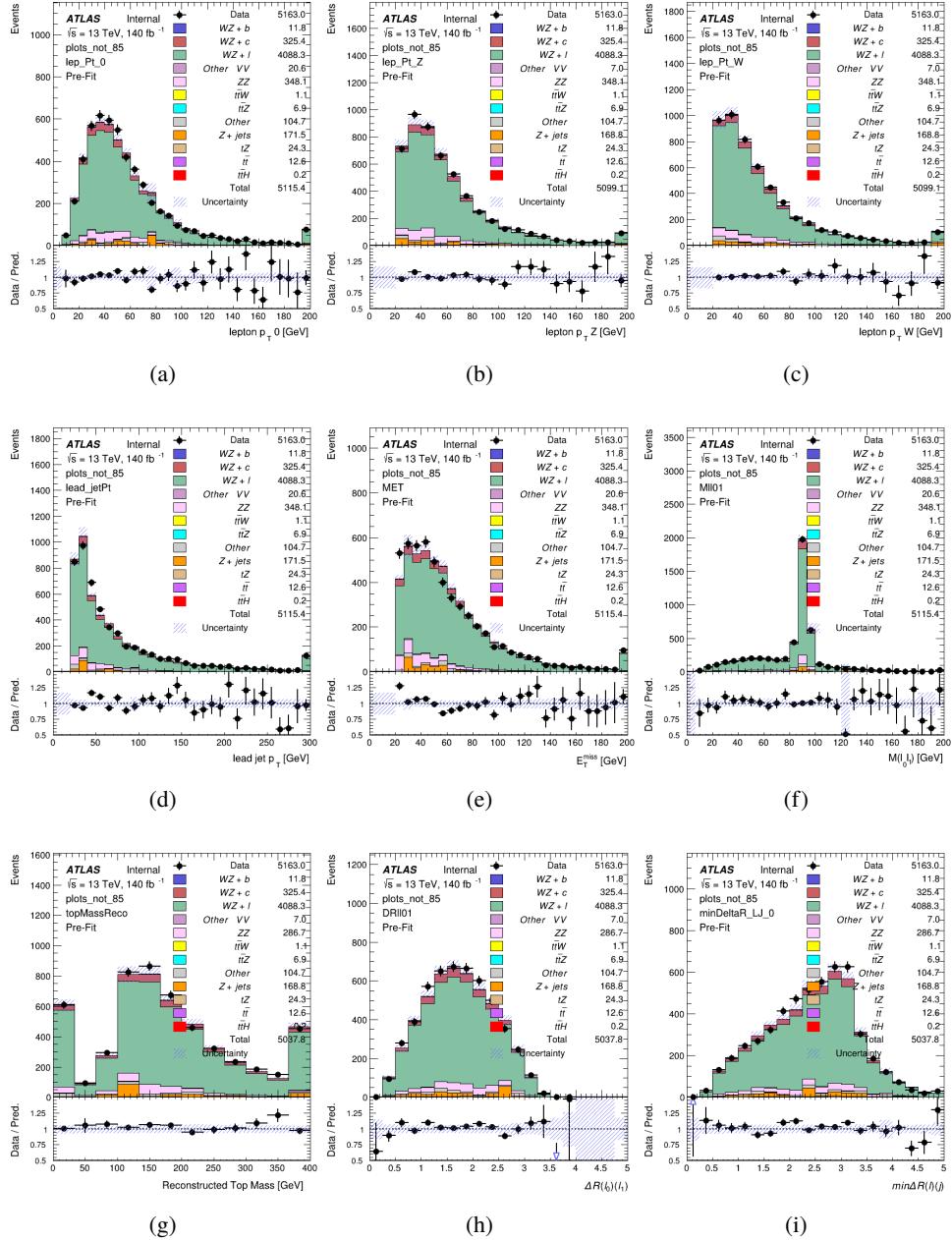


Figure 9.4: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j 77-85% WP

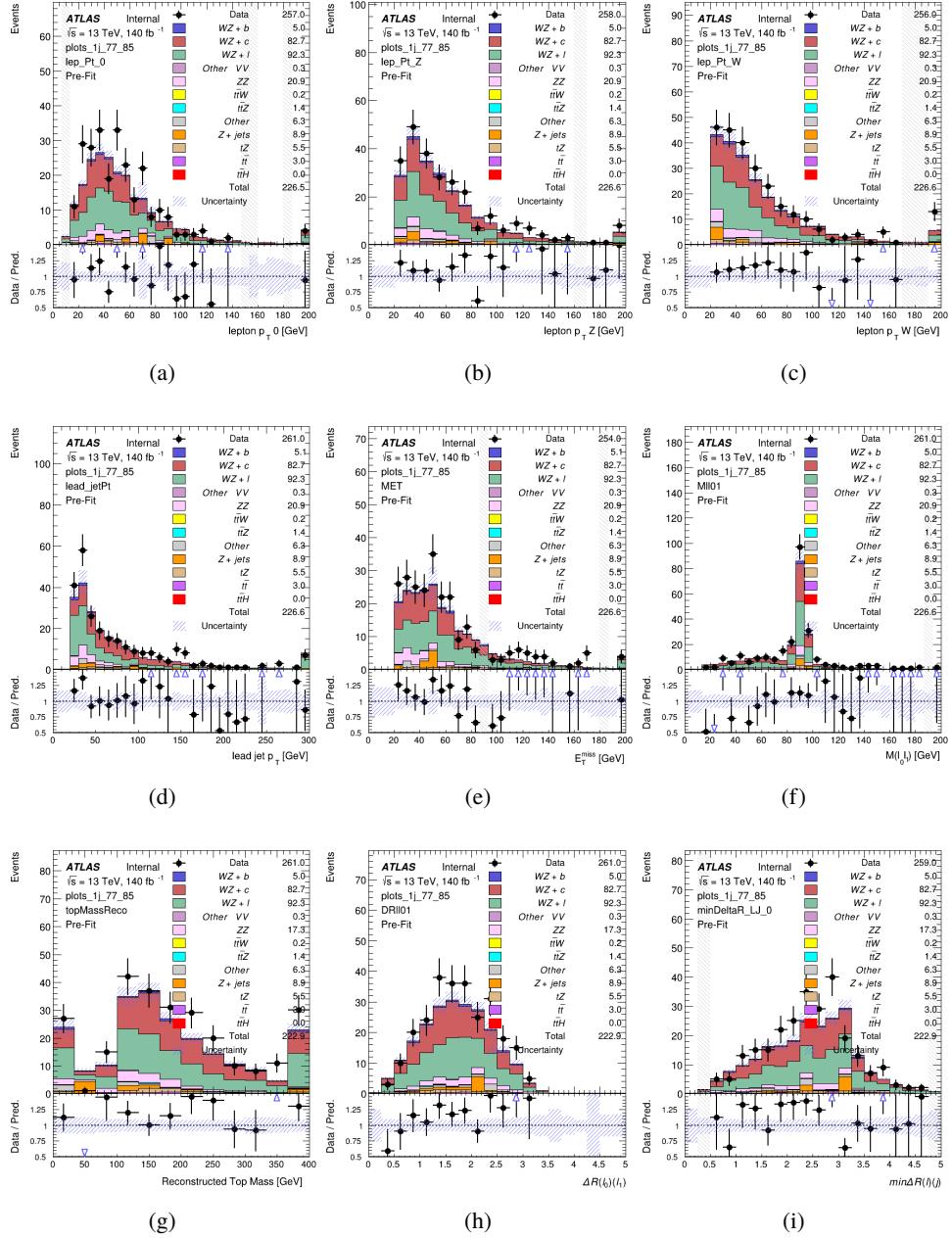


Figure 9.5: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j 70-77% WP

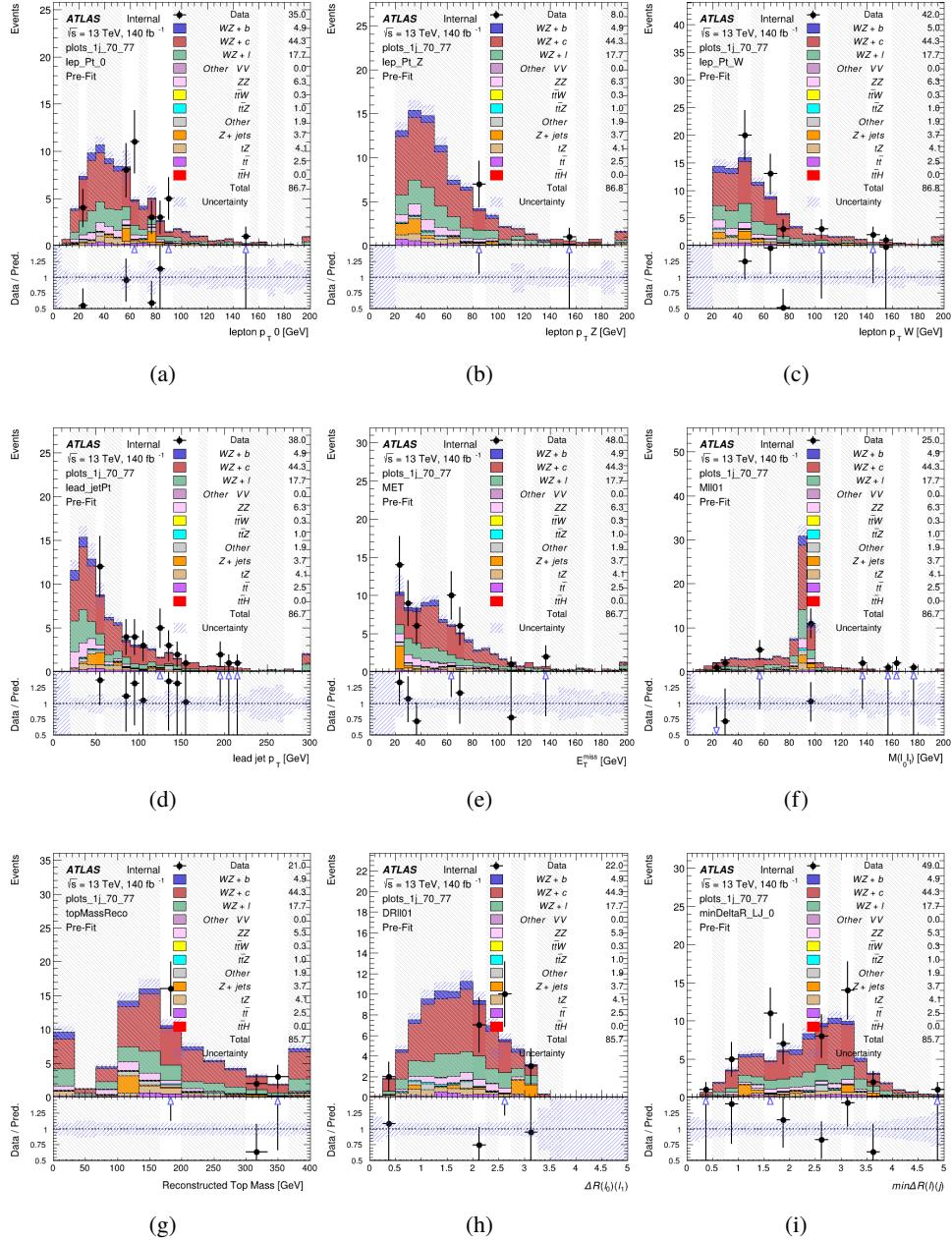


Figure 9.6: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j 60-70% WP

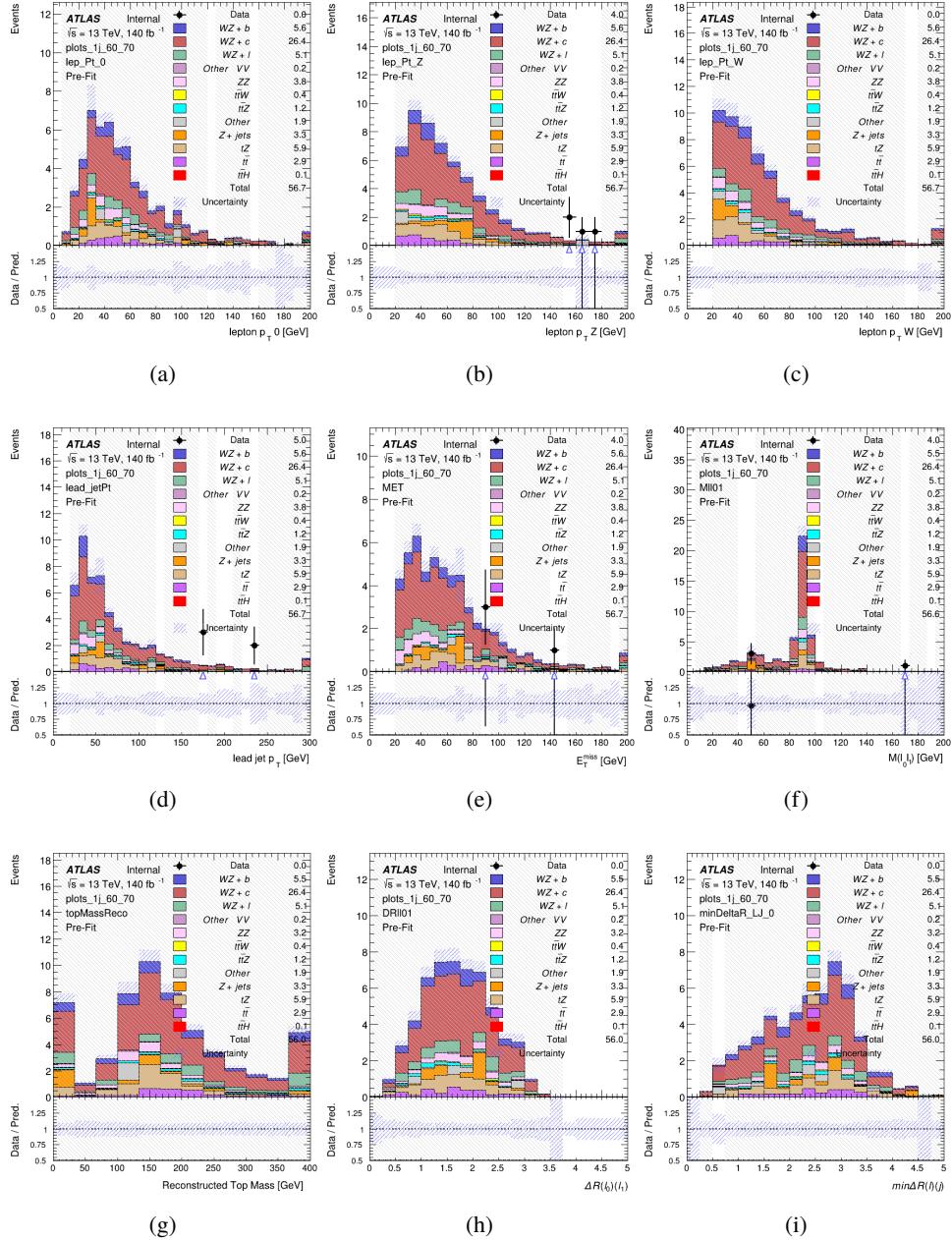


Figure 9.7: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j 60% WP

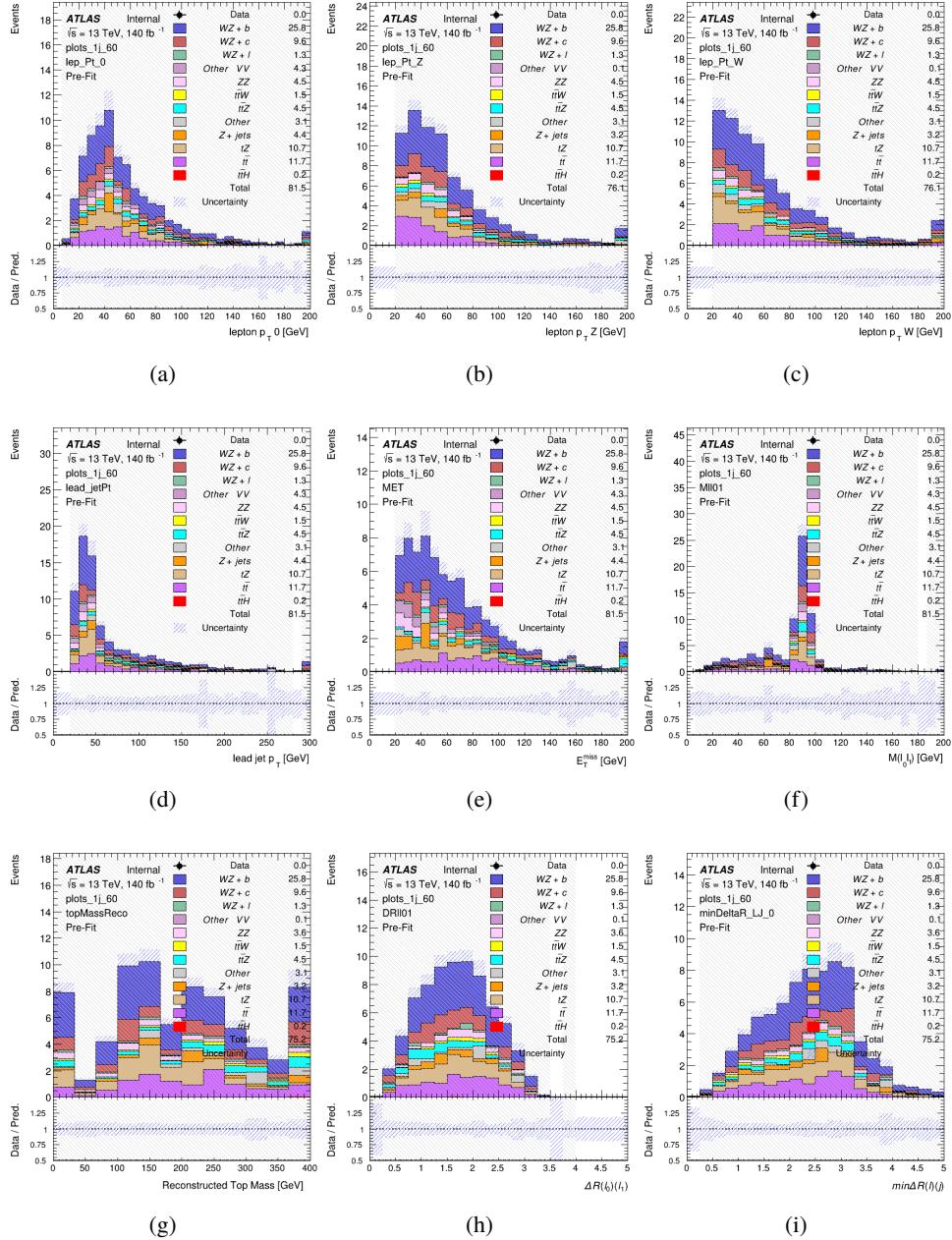


Figure 9.8: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - tZ-CR

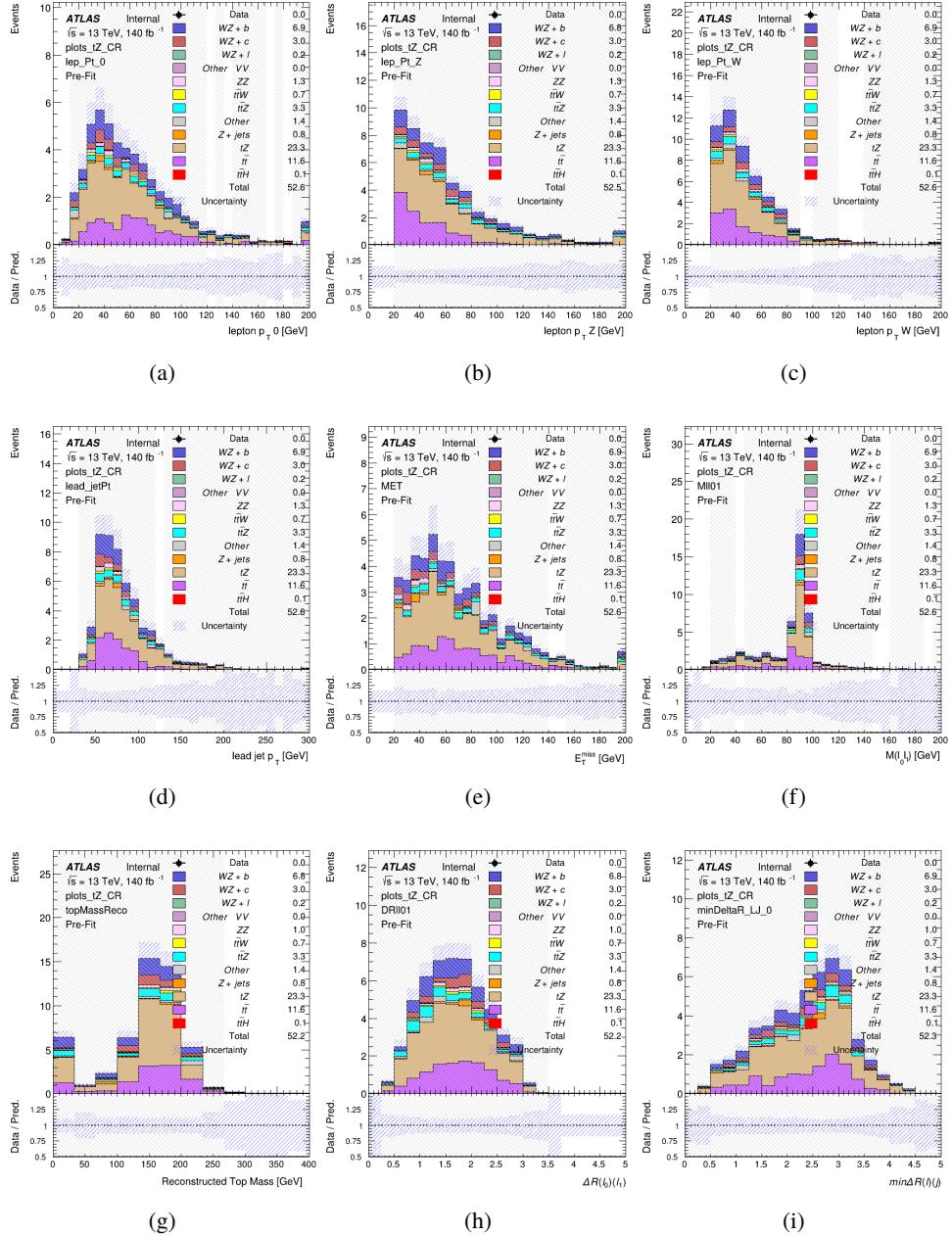


Figure 9.9: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j Inclusive

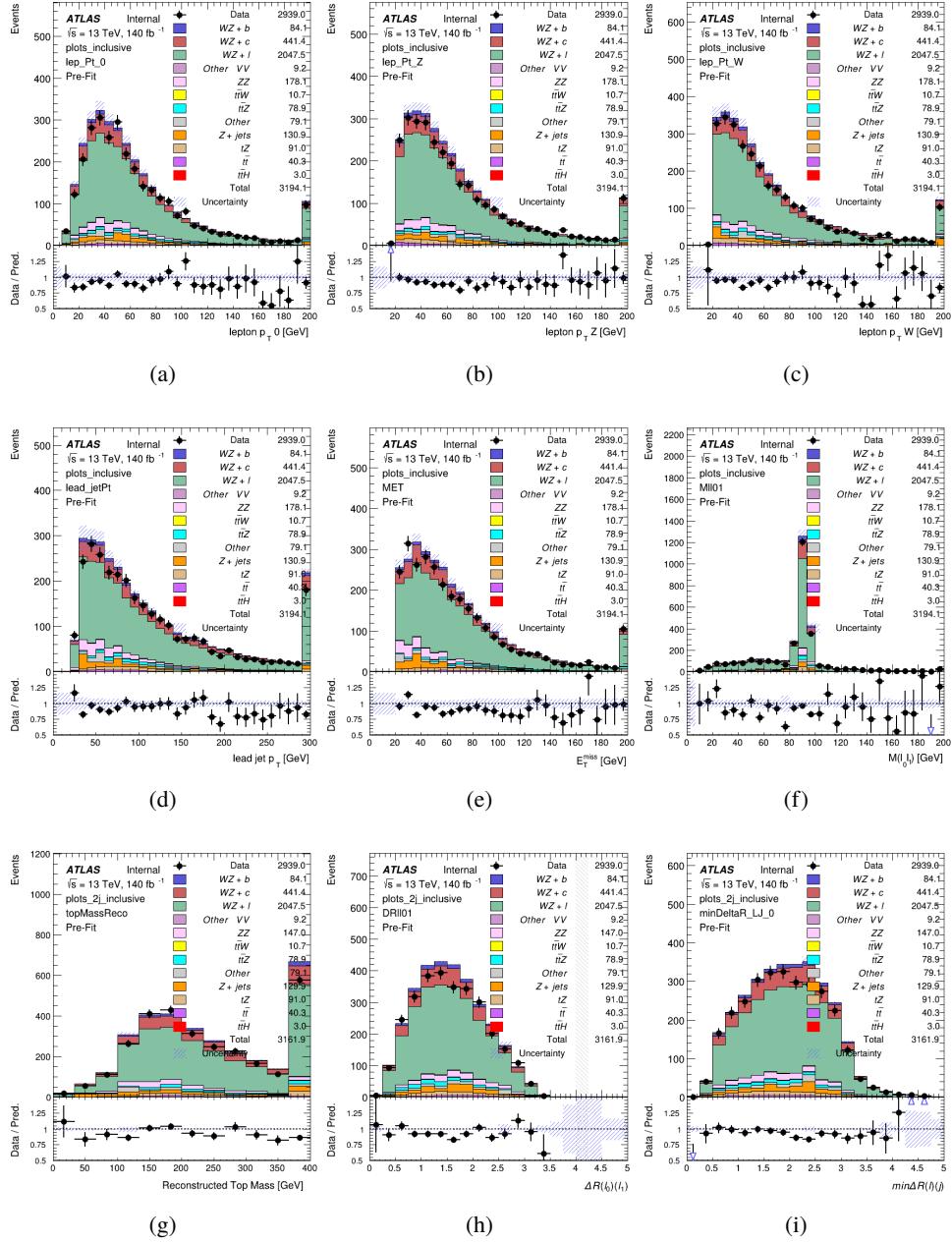


Figure 9.10: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

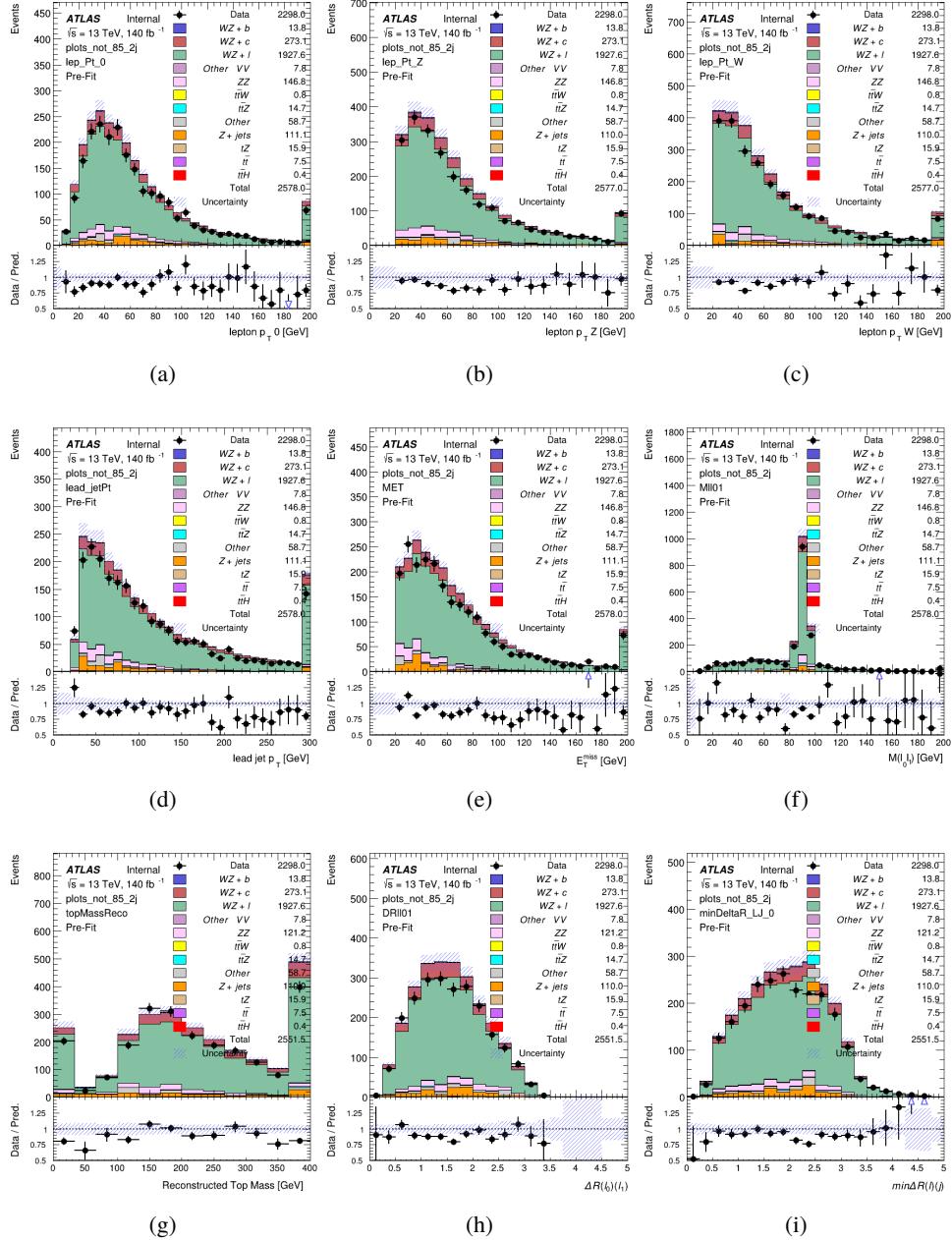
WZ Fit Region - $2j < 85\%$ WP

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 leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j 77-85% WP

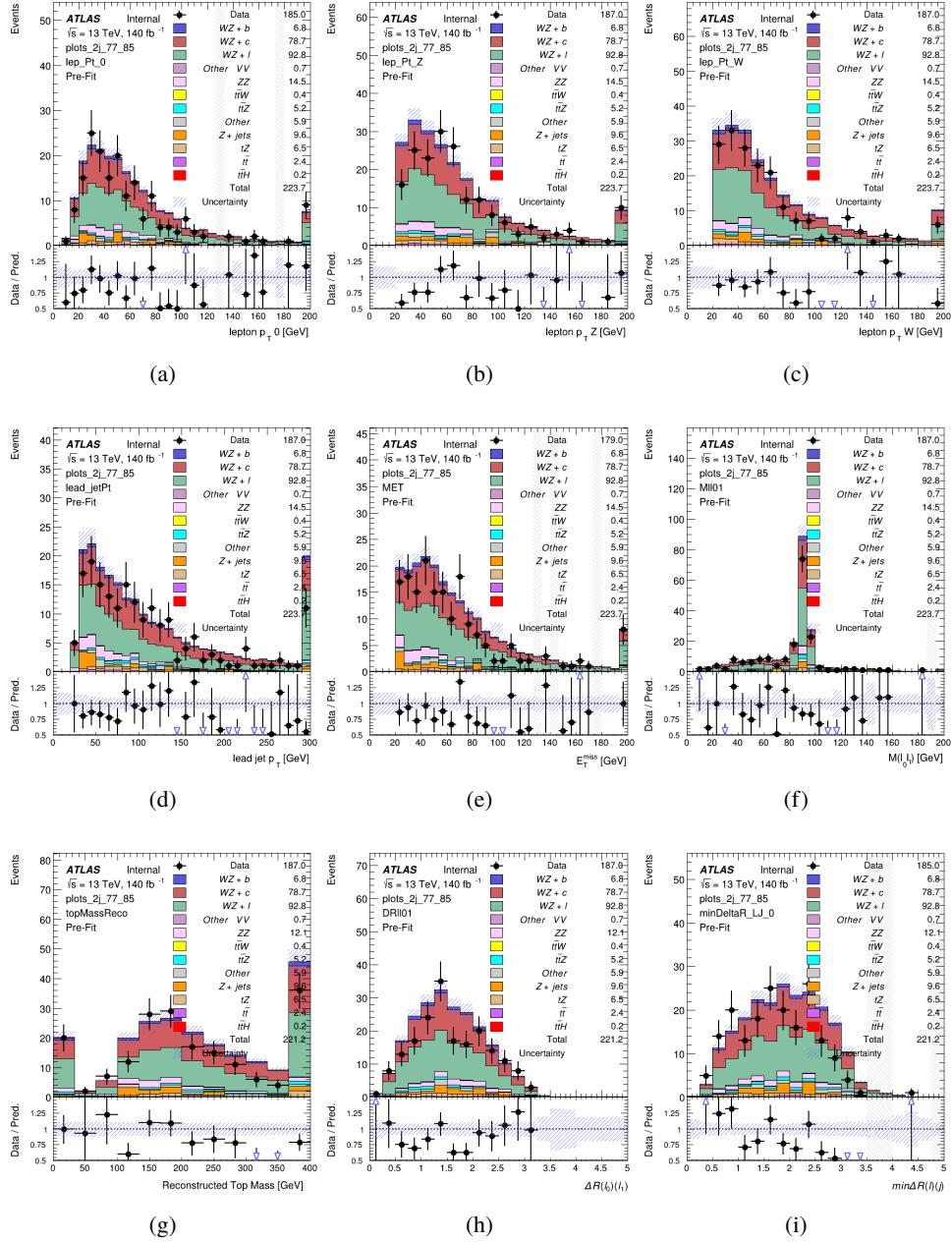


Figure 9.12: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j 70-77% WP

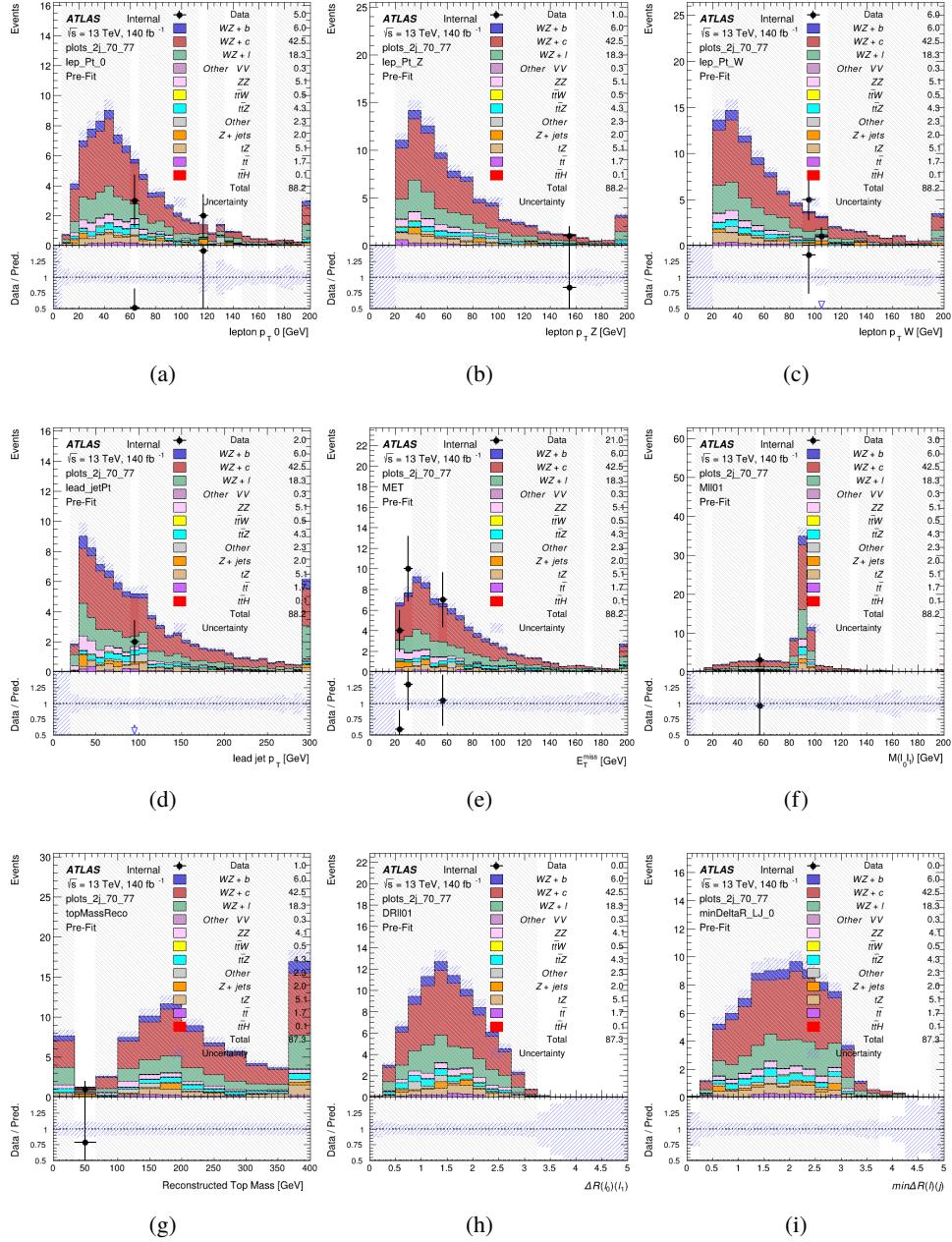


Figure 9.13: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j 60-70% WP

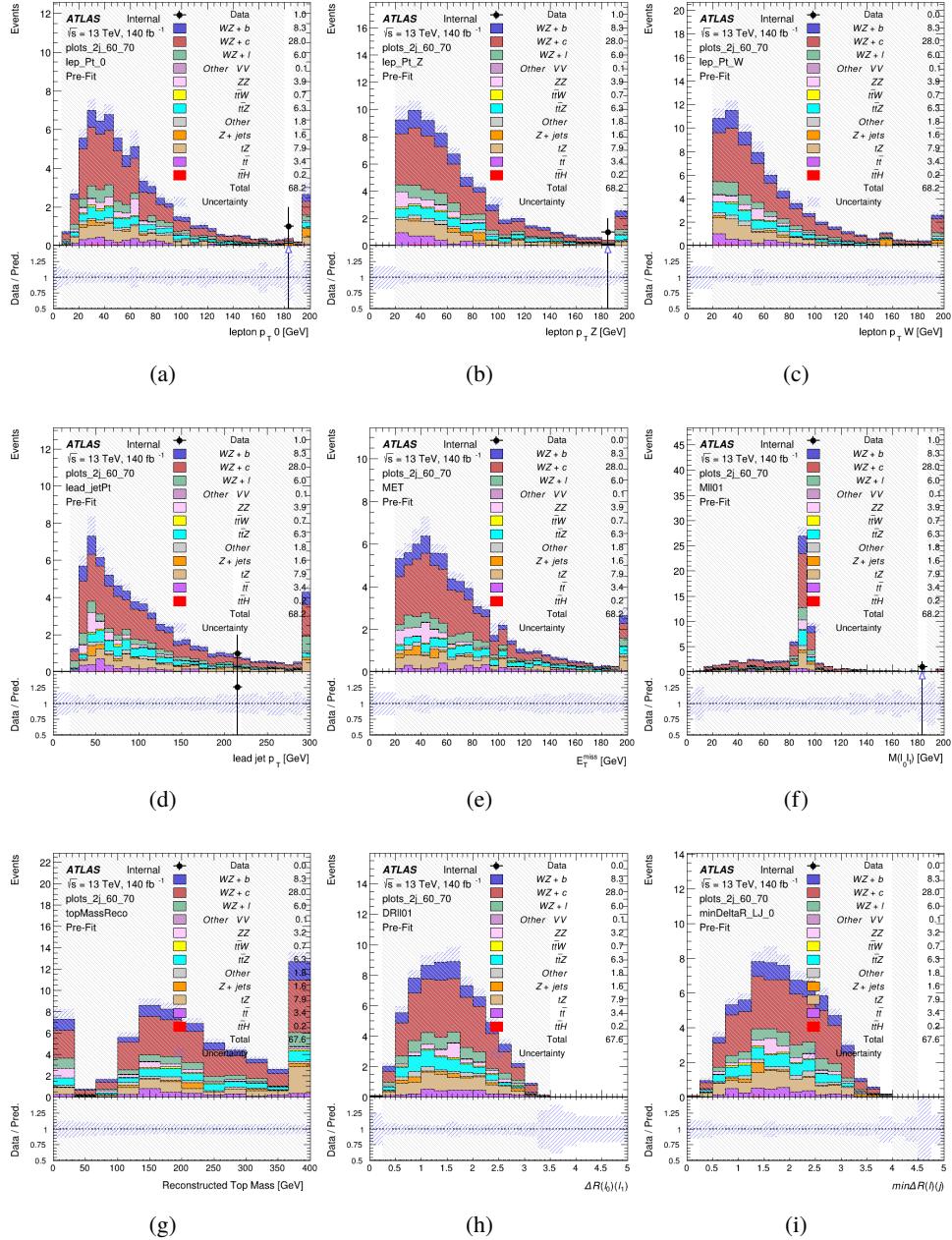


Figure 9.14: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j 60% WP

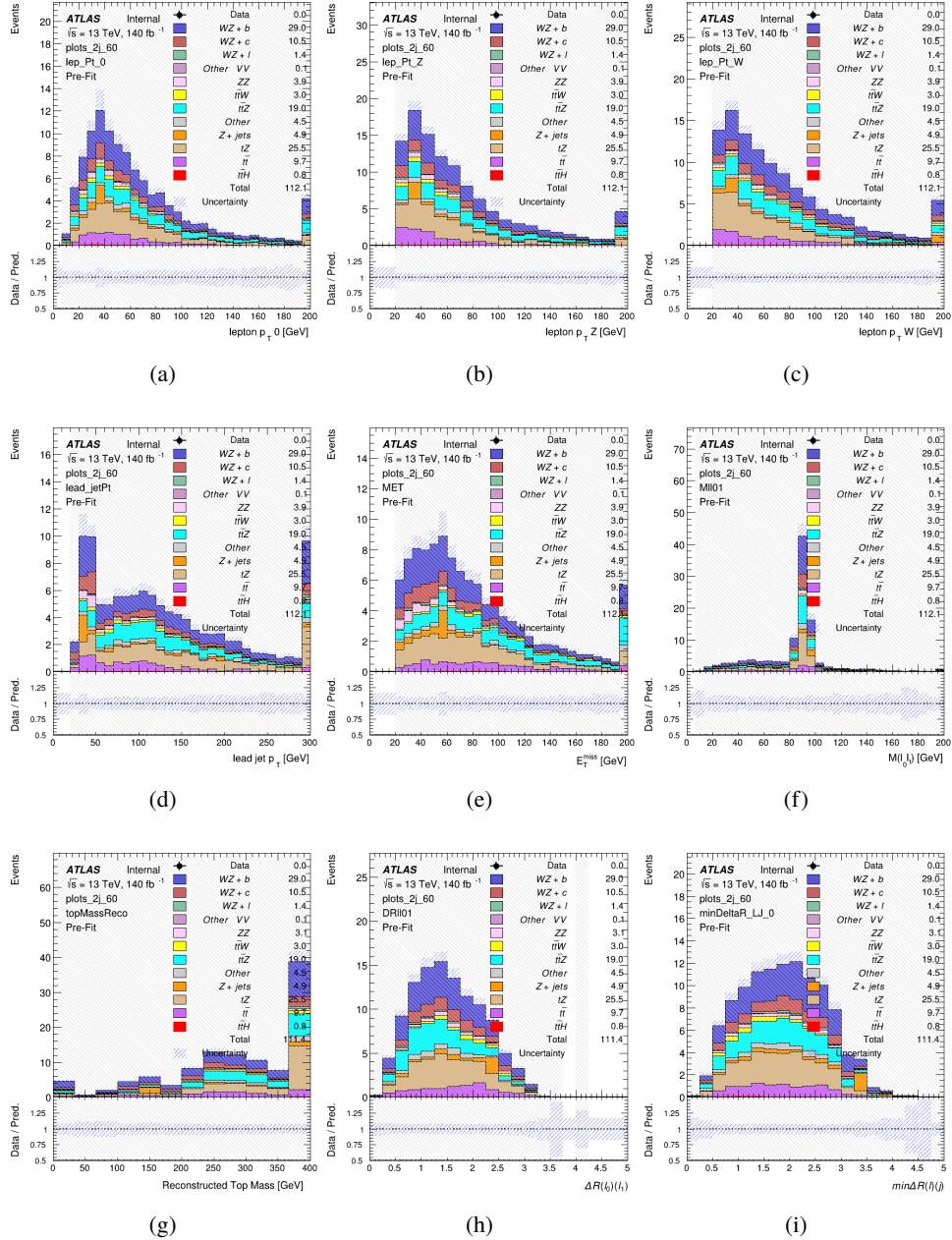


Figure 9.15: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - tZ-CR-2j

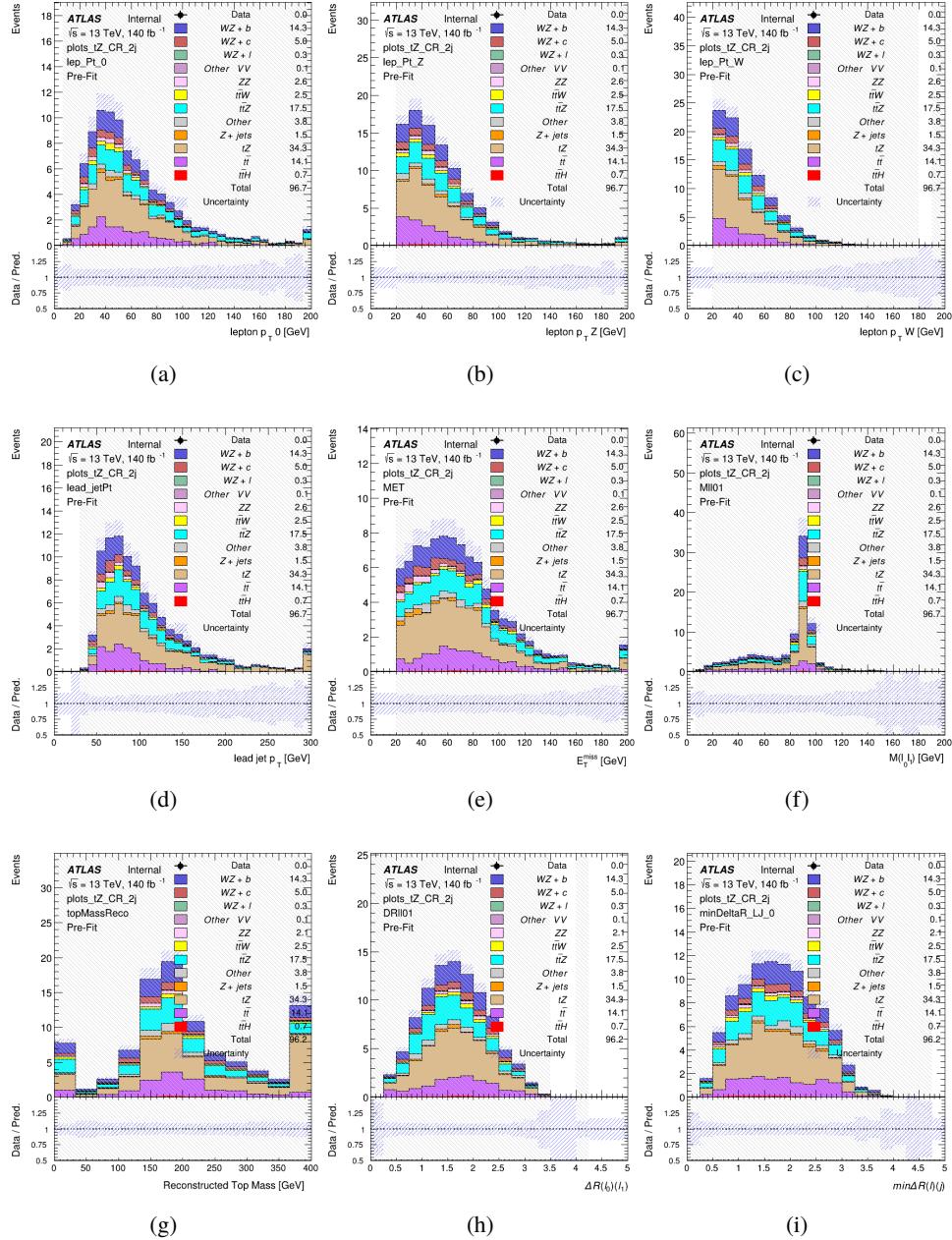


Figure 9.16: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

592 **9.3 Non-Prompt Lepton Estimation**

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

597 **9.3.1 $t\bar{t}$ Validation**

598 $t\bar{t}$ events can produce two prompt leptons from the decay of each of the tops. These top decays
 599 produce two b-quarks, the decay of which can produce additional non-prompt leptons, which
 600 occasionally pass the event preselection. In order to validate that the Monte Carlo accurately
 601 simulates this process accurately, the MC prediction in a non-prompt $t\bar{t}$ enriched control region
 602 is compared to data.

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

608 Further, because the jet multiplicity of $t\bar{t}$ events tends to be higher than WZ, the number
 609 of jets in each event is required to be greater than 1. As b-jets are almost invariably produced

⁶¹⁰ from top decays, at least one b-tagged jet passing the 70% DL1r WP in each event is required.

⁶¹¹ Various kinematic plots of this region are shown in Figure 13.17.

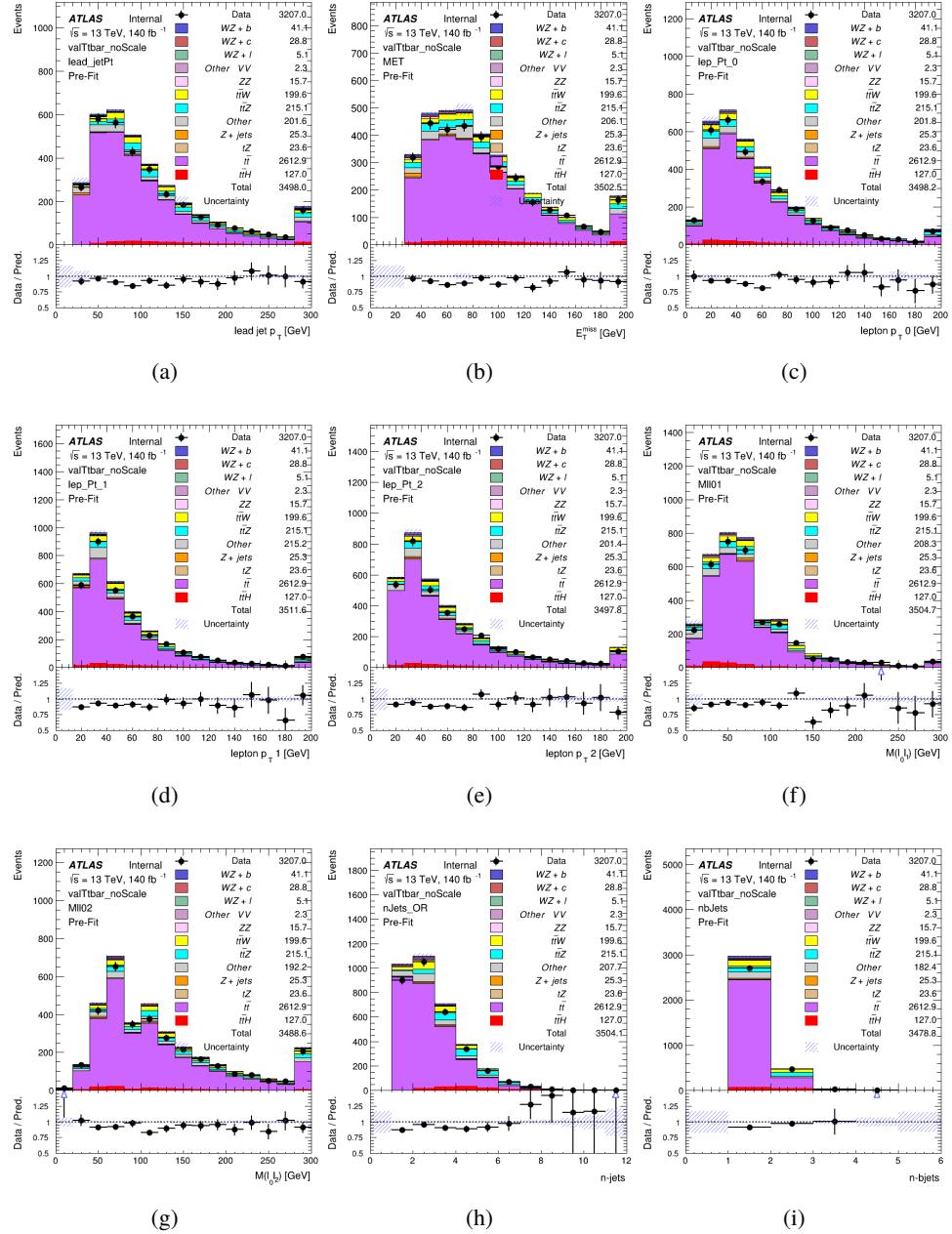


Figure 9.17: Comparisons between the data and MC distributions in the $t\bar{t}$ control region for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of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

612 The shape of each distribution agrees quite well between data and MC, with a constant
613 offset between the two. This is accounted for by applying a constant correction factor of 0.883
614 to the $t\bar{t}$ MC prediction. Plots showing the kinematics of the $t\bar{t}$ VR after this correction factor
615 has been applied are shown in Figure 13.18.

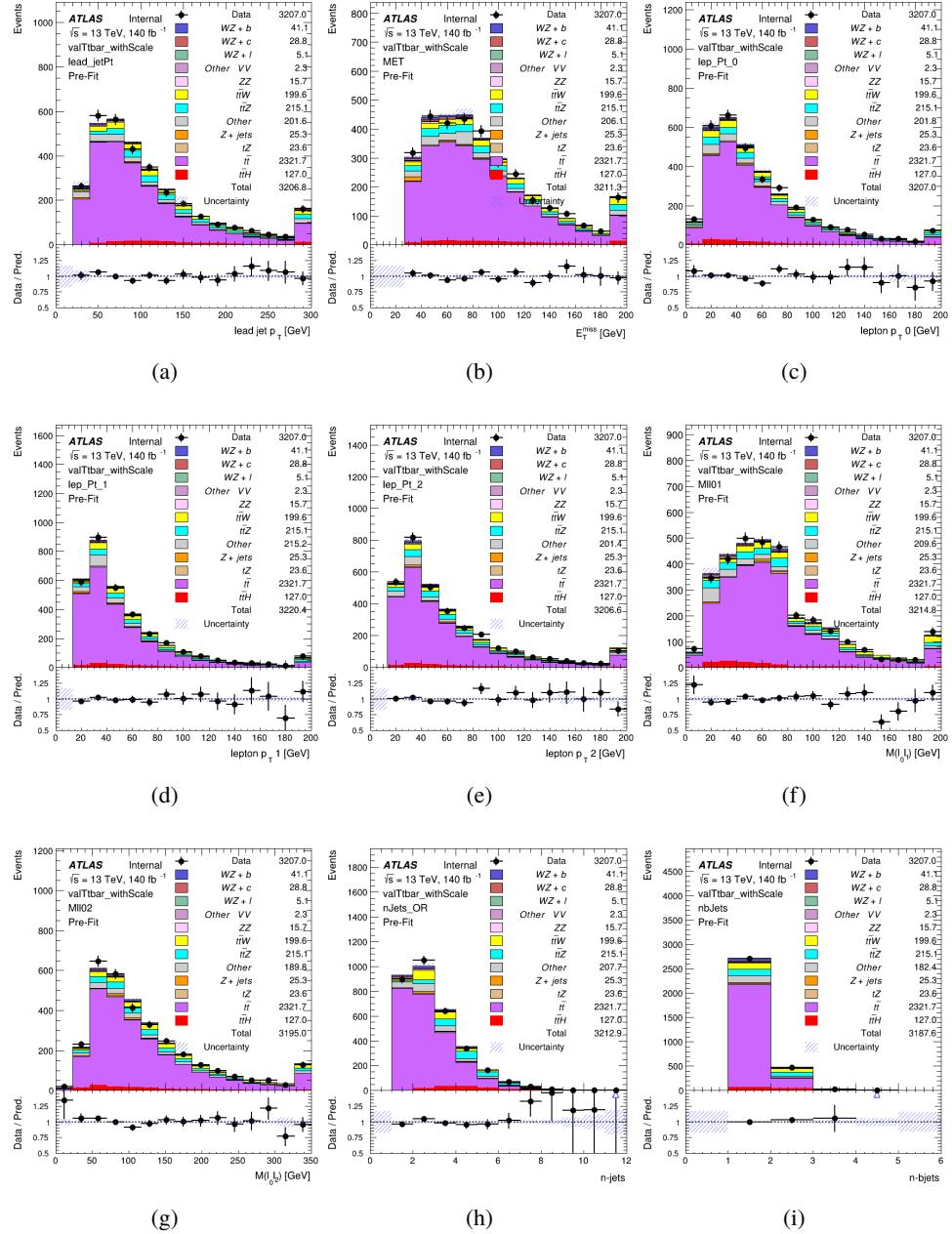


Figure 9.18: Comparisons between the data and MC distributions in the $t\bar{t}$ control region after the correction factor has been applied for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass 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.

616 The modeling is further validated by looking at the yield in the $t\bar{t}$ VR for each DL1r WP,
 617 giving a clearer correspondence to the signal regions used in the fit. Each region shown in Figure
 618 13.19 requires one or more jets pass the listed WP, with no jets passing the next highest WP.

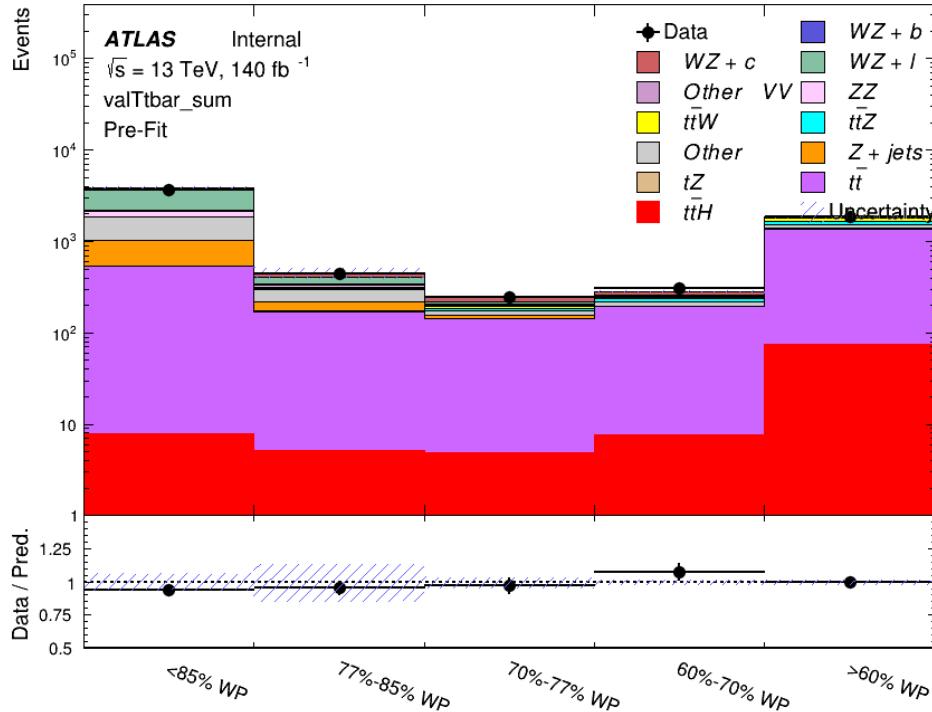


Figure 9.19: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the $t\bar{t}$ VR selection and correction factor have been applied

619 As data and MC are found to agree within 20% for each of these working points, a 20%
 620 systematic uncertainty on the $t\bar{t}$ prediction is included for the analysis.

621 **9.3.2 Z+jets Validation**

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

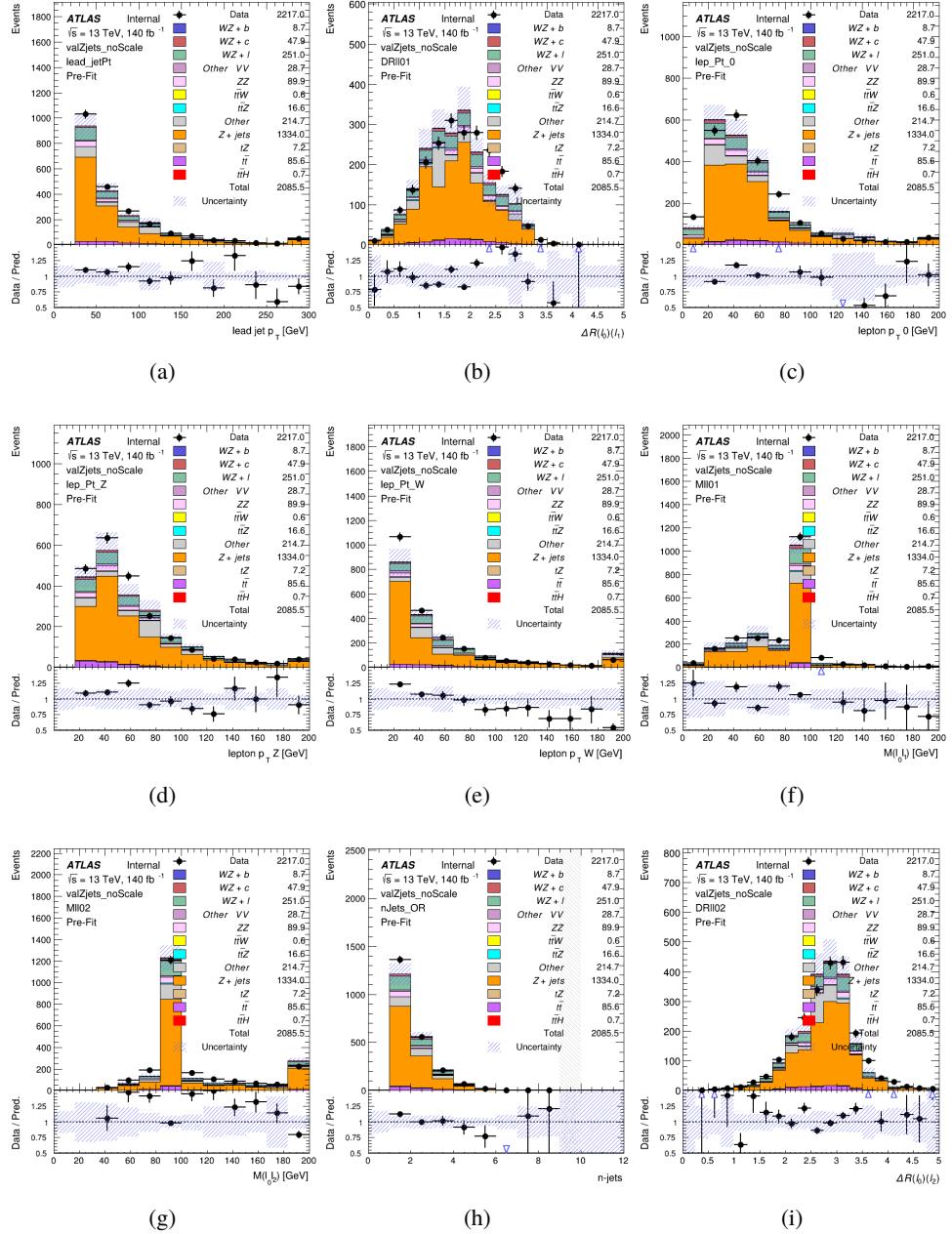


Figure 9.20: Comparisons between the data and MC distributions in the $Z + \text{jets}$ control region for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant 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

629 While there is general agreement between data and MC within statistical uncertainty, the
630 shape of the p_T spectrum of the lepton from the W is found to differ. To account for this
631 discrepancy, a variable correction factor is applied to Z+jets. χ^2 minimization of the lepton 2 p_T
632 spectrum is performed to derive a correction factor of $1.53 - 6.6 * 10^{-6}(\text{lep}_\text{Pt}_W)$. Kinematic
633 plots of the Z + jets control region after this correction factor has been aplied are shown in Figure
634 [13.21](#).

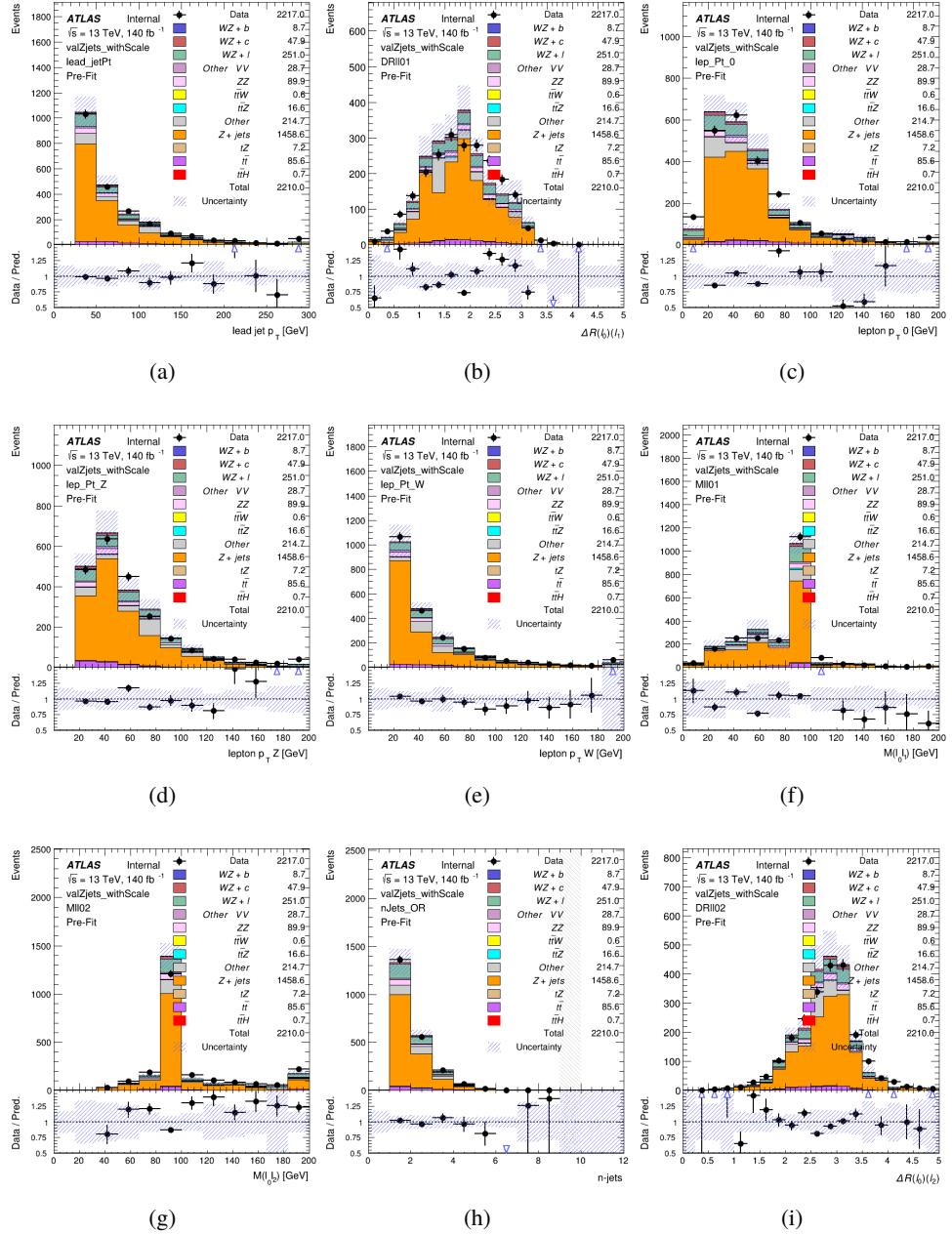


Figure 9.21: Comparisons between the data and MC distributions in the $Z + \text{jets}$ control region after the correction factor has been applied for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS 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

635 The modeling is further validated by looking at the yield in the Z+jets VR for each DL1r
 636 WP, giving a clearer correspondence to the signal regions used in the fit. Each region shown in
 637 Figure 13.22 requires one or more jets pass the listed WP, with no jets passing the next highest
 638 WP.

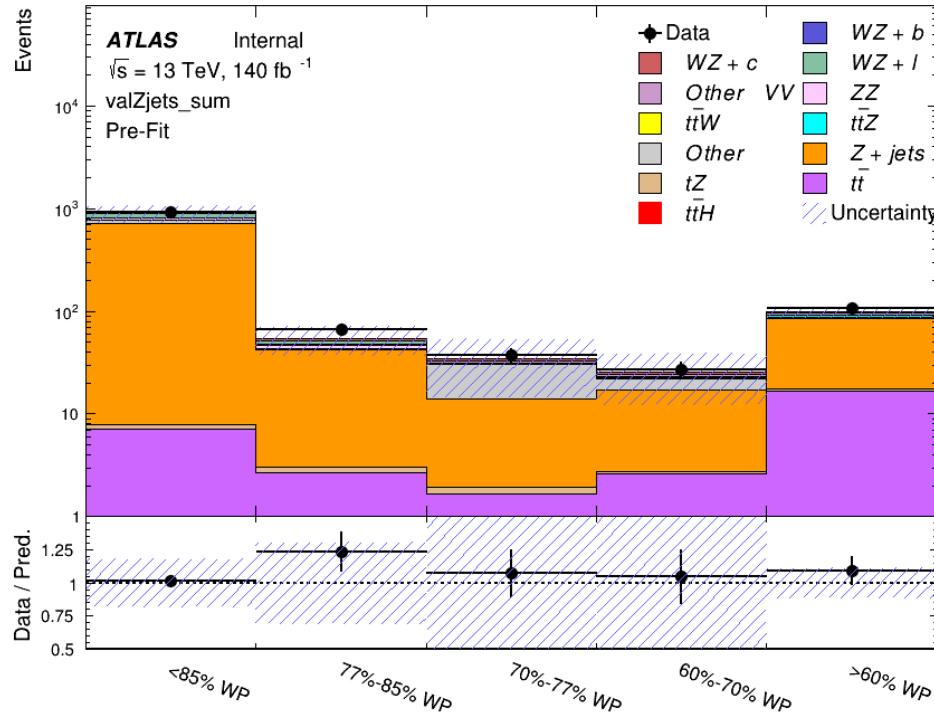


Figure 9.22: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the Z+jets VR selection and correction factor have been applied

639 For each of the b-tagging working points considered, the data falls within 25% of the MC
 640 prediction once this correction factor has been applied. Therefore, a 25% systematic uncertainty
 641 is applied to Z + jets in the analysis.

642 10 tZ Interference Studies and Separation Multivariate Analysis

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

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

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

652 10.1 Top Mass Reconstruction

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

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

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

$$663 \quad (p_l + p_\nu)^2 = m_W^2$$

664 Expanding this out into components, this equation gives:

$$665 \quad \sqrt{p_{T\nu}^2 + p_{z\nu}^2} E_l = \frac{m_W^2 - m_l^2}{2} + p_{T\nu}(p_{lx}\cos\phi_\nu + p_{ly}\sin\phi_\nu) + p_{lz}p_{\nu z}$$

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

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

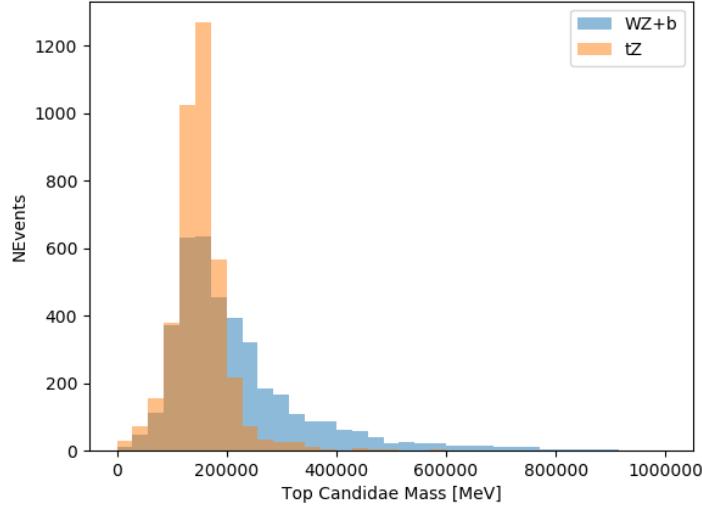


Figure 10.1: Reconstructed top mass distributions for tZ and WZ + b, measured in MeV.

672 10.2 tZ BDT

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

675 • The invariant mass of the reconstructed top candidate

676 • p_T of each of the leptons, jet

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

678 • E_T^{miss}

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

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

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

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

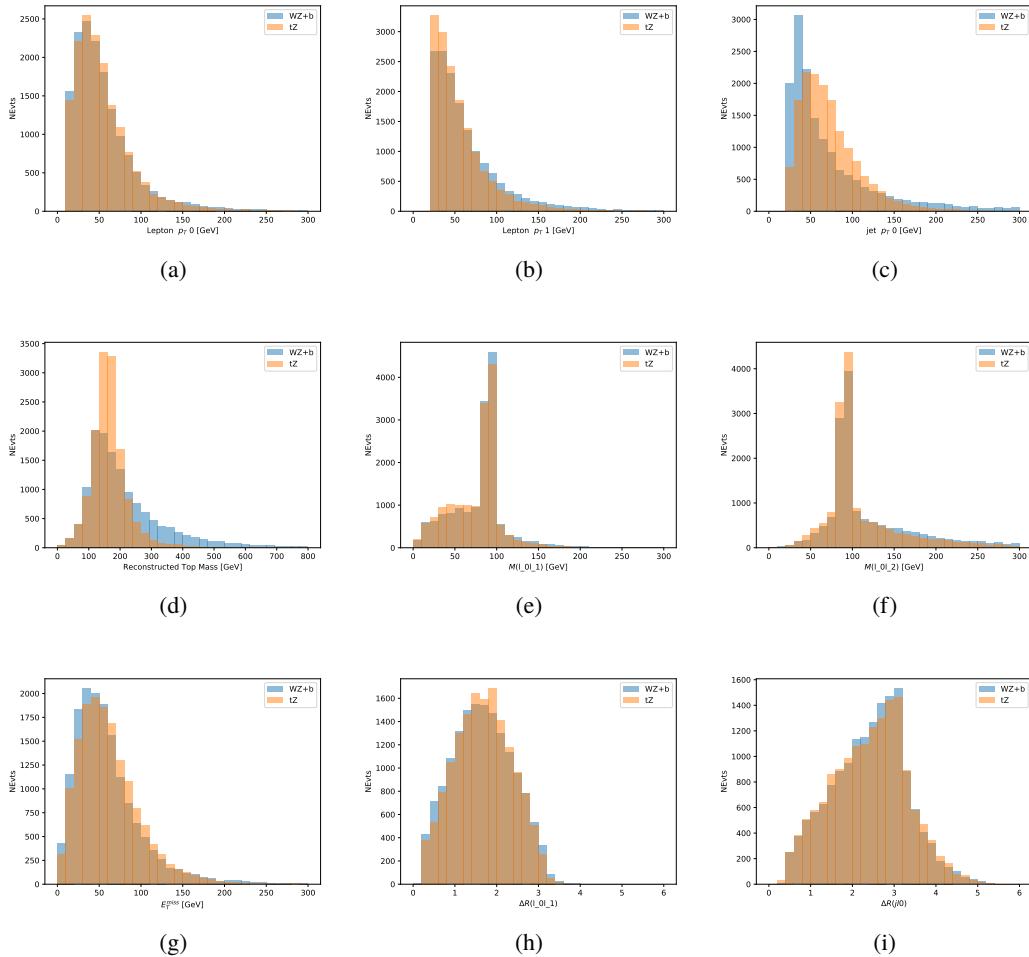


Figure 10.2: Distribution of input features of the BDT for signal (WZ) and background (tZ). Both are scaled to an equal number of events. (a), (b) and (c) show the p_T of lepton 0, lepton 1, and the jet, (d) show the reconstructed top mass, (e) and (f) show the invariant mass of leptons 0 and 1, leptons 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.

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

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

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

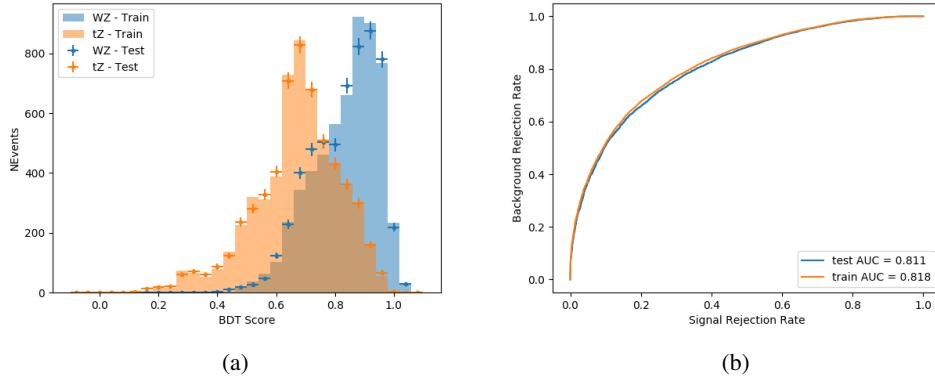


Figure 10.3: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

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

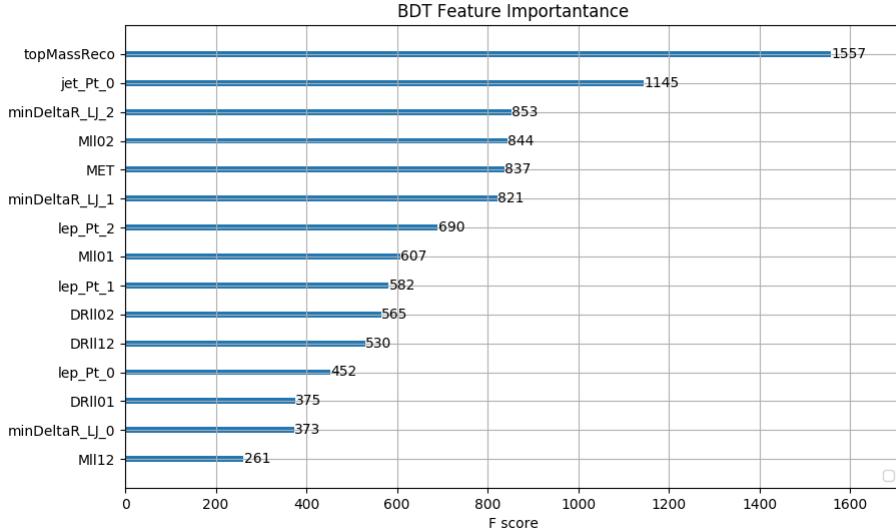


Figure 10.4: Relative importance of each input feature in the model.

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

705 11 Data and Monte Carlo Samples

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

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

- 712 • at least two light leptons within a range $|\eta| < 2.6$, with leading lepton $p_T > 15$ GeV and
713 subleading lepton $p_T > 5$ GeV
- 714 • at least one light lepton with $p_T > 15$ GeV within a range $|\eta| < 2.6$, and at least two hadronic
715 taus with $p_T > 15$ GeV.

716 Samples were then generated from these HIGG8D1 derivations with p-tag of p4134 for data
717 and p4133 for Monte Carlo using AnalysisBase version 21.2.127 modified to include custom
718 variables.

719 **11.1 Data Samples**

720 The study uses a sample of proton-proton collision data collected by the ATLAS detector from
721 2015 through 2018 at an energy of $\sqrt{s} = 13$ TeV, which represents an integrated luminosity of
722 139 fb^{-1} . This data set was collected with a bunch crossing rate of 25 ns. All data used in this
723 analysis was verified by data quality checks, having been included in the following Good Run
724 Lists:

- 725 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
726 _PHYS_StandardGRL_All_Good_25ns.xml
- 727 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
728 _PHYS_StandardGRL_All_Good_25ns.xml
- 729 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL
730 _All_Good_25ns_Triggerno17e33prim.xml
- 731 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL
732 _All_Good_25ns_Triggerno17e33prim.xml
- 733 Runs included from the AllYear period containers are included.

734 **11.2 Monte Carlo Samples**

735 Several different generators were used to produce Monte Carlo simulations of the signal and
736 background processes. For all samples, the response of the ATLAS detector is simulated using
737 Geant4. The WZ signal samples are simulated using Sherpa 2.2.2 [[sherpa](#)]. Specific information
738 about the Monte Carlo samples being used can be found in Table 28. A list of the specific samples
739 used by data set ID is shown in Table 10.

Table 9: 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
t <bar>t>W</bar>	410155
t <bar>t>Z</bar>	410156, 410157, 410218-20
low mass t <bar>t>Z</bar>	410276-8
Rare Top	410397, 410398, 410399
single Top	410658-9, 410644-5
three Top	304014
four Top	410080
t <bar>t>WW</bar>	410081
Z + jets	364100-41
low mass Z + jets	364198-215
W + jets	364156-97
V γ	364500-35
tZ	410560
tW	410013-4
WtZ	410408
VVV	364242-9
VH	342284-5
WtH	341998
t <bar>t>γ</bar>	410389
t <bar>t></bar>	410470
t <bar>t>H</bar>	345873-5, 346343-5

Table 10: List of Monte Carlo samples by data set ID used in the analysis.

740 12 Object Reconstruction

741 All regions defined in this analysis share a common lepton, jet, and overall event preselection.

742 The selection applied to each physics object is detailed here; the event preselection, and the

743 selection used to define the various fit regions, is described in Section 13.

⁷⁴⁴ **12.1 Trigger**

⁷⁴⁵ Events are required to be selected by dilepton triggers, as summarized in Table 29.

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

⁷⁴⁶ **12.2 Light leptons**

⁷⁴⁷ Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that

⁷⁴⁸ are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].

⁷⁴⁹ Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Muon candidates

⁷⁵⁰ are reconstructed by combining inner detector tracks with track segments or full tracks in the

⁷⁵¹ muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have $p_T > 10$ GeV and

752 $|\eta| < 2.5$. Candidates in the transition region between different electromagnetic calorimeter
753 components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant com-
754 bining shower shape and track information is used to distinguish real electrons from hadronic
755 showers (fake electrons). To further reduce the non-prompt electron contribution, the track is
756 required to be consistent with originating from the primary vertex; requirements are imposed
757 on the transverse impact parameter significance ($|d_0|/\sigma_{d_0} < 5$) and the longitudinal impact
758 parameter ($|\Delta z_0 \sin \theta_\ell| < 0.5 \text{ mm}$).

759 Muon candidates are reconstructed by combining inner detector tracks with track segments
760 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
761 $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$. The longitudinal impact parameter is the same for both electrons
762 and muons, while muons are required to pass a slightly tighter transverse impact parameter,
763 $|d_0|/\sigma_{d_0} < 3$.

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

768 **12.3 Jets**

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

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

777 **12.4 B-tagged Jets**

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

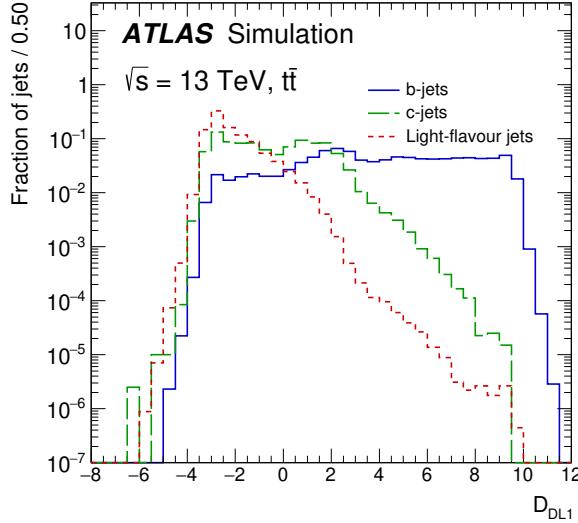


Figure 12.1: Output distribution of the DL1r algorithm for b-jets, charm jets, and light jets

783 From the output of the BDT, working points (WPs) are developed based on the efficiency
 784 of truth b-jets at particular values of the DL1r algorithm. These WPs are taken from the March
 785 2020 CDI file, 2020-21-13TeV-MC16-CDI-2020-03-11_v2.root. The working points used in
 786 this analysis are summarized in Table 12.

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

Table 12: B-tagging Working Points by tightness and b-jet efficiency

787 A tighter WP will accept fewer b-jets, but reject a higher fraction of charm and light jets.
 788 Generally, analyses that include b-jets will use a fixed working point, for example, requiring that
 789 a jet pass the 70% threshold. By instead treating these working point as bins, e.g. events with
 790 jets that fall between the 85% and 77% WPs fall into one bin, while events with jets passing the

791 60% WP fall into another, and looking at the full psuedo-continuous DL1r spectrum of the jets,
792 additional information can be gained. The psuedo-continuous b-tag spectrum is used in this case
793 to separate out WZ + b, WZ + charm, and WZ + light.

794 **12.5 Missing transverse energy**

795 Missing transverse momentum (E_T^{miss}) is used as part of the event selection. The missing
796 transverse momentum vector is defined as the inverse of the sum of the transverse momenta of
797 all reconstructed physics objects as well as remaining unclustered energy, the latter of which is
798 estimated from low- p_T tracks associated with the primary vertex but not assigned to a hard object,
799 with object definitions taken from [ATL-PHYS-PUB-2015-027]. Light leptons considered in
800 the E_T^{miss} reconstruction are required to have $p_T > 10 \text{ GeV}$, while jets are required to have $p_T > 20$
801 GeV .

802 **12.6 Overlap removal**

803 To avoid double counting objects and remove leptons originating from decays of hadrons, overlap
804 removal is performed in the following order: any electron candidate within $\Delta R = 0.1$ of another
805 electron candidate with higher p_T is removed; any electron candidate within $\Delta R = 0.1$ of a muon
806 candidate is removed; any jet within $\Delta R = 0.3$ of an electron candidate is removed; if a muon
807 candidate and a jet lie within $\Delta R = \min(0.4, 0.04 + 10[\text{GeV}] / p_T(\text{muon}))$ of each other, the jet
808 is kept and the muon is removed.

809 This algorithm is applied to the preselected objects. The overlap removal procedure is
 810 summarized in Table 13.

Keep	Remove	Cone size (ΔR)
electron	electron (low p_T)	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	$\min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$
electron	tau	0.2

Table 13: Summary of the overlap removal procedure between electrons, muons, and jets.

811 13 Event Selection and Signal Region Definitions

812 Event are required to pass a preselection described in Section 13.1 and summarized in Table 14.
 813 Those that pass this preselection are divided into various fit regions described in Section 13.2,
 814 based on the number of jets in the event, and the b-tag score of those jets.

815 13.1 Event Preselection

816 Events are required to include exactly three reconstructed light leptons passing the requirement
 817 described in 12.2, which have a total charge of ± 1 . As the opposite sign lepton is found to
 818 be prompt the vast majority of the time [ttH_paper], it is required to be loose and isolated,
 819 as defined though the standard PLVLoose working point supported by combined performance
 820 groups. The same sign leptons are required to be very tightly isolated, as per the recommended
 821 PLVTight.

822 The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose
 823 charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.
 824 the smallest ΔR , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton
 825 0 is required to have $p_T > 10$ GeV, while the same sign leptons, 1 and 2, are required to have
 826 $p_T > 20$ GeV to reduce the contribution of non-prompt leptons.

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

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

834 Events are required to have one or two reconstructed jets passing the selection described
 835 in Section 12.3. Events with more than two jets are rejected in order to reduce the contribution
 836 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 14: Summary of the selection applied to events for inclusion in the fit

837 The event yields in the preselection region for both data and Monte Carlo are summarized
 838 in Table 13.1, which shows good agreement between data and Monte Carlo, and demonstrates that
 839 this region consists primarily of WZ events. The WZ events are split into WZ + b, WZ + c, and
 840 WZ + l based on the truth flavor of the associated jet in the event. Specifically, this determination
 841 is made based on the HadronConeExclTruthLabelID of the jet, as recommended by the b-
 842 tagging working group [BtagWG]. In this ordering b-jet supersedes charm, which supersedes
 843 light. That is, WZ + l events contain no charm and no b jets at truth level, WZ + c contain at
 844 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
t̄tW	16.8 ± 2.3
t̄tZ	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
t̄tWW	0.23 ± 0.05
Z + jets	600 ± 260
V + γ	37 ± 54
tZ	190 ± 70
tW	5.5 ± 1.2
WtZ	25.8 ± 1.1
VVV	26.2 ± 0.9
VH	94 ± 7
t̄t	108.68 ± 8
t̄tH	4.3 ± 0.5
Total	10600 ± 530
Data	10574

Table 15: Event yields in the preselection region at 139.0 fb^{-1}

845 Here Other VV represents diboson processes other than WZ, and consists predominantly
846 of $ZZ \rightarrow llll$ events where one of the leptons is not reconstructed.

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

WZ Fit Region - Inclusive

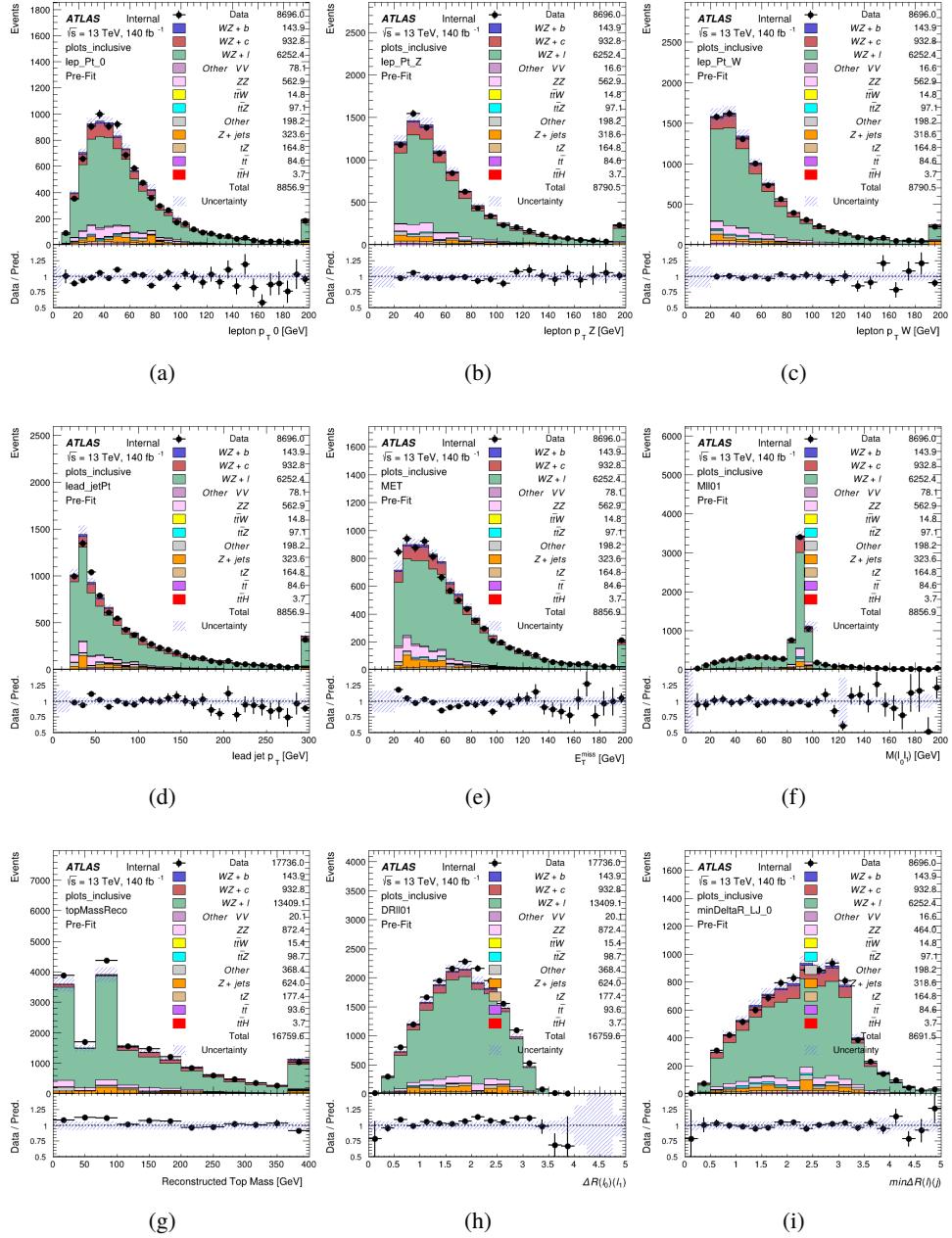


Figure 13.1: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

850 **13.2 Fit Regions**

851 Once preselection has been applied, the remaining events are categorized into one of twelve
 852 orthogonal regions. The regions used in the fit are summarized in Table 16.

Table 16: 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%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} \geq 1, n\text{Jets_DL1r_77} = 0$
2j, 77%-70%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_77} \geq 1, n\text{Jets_DL1r_70} = 0$
2j, 70%-60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_70} \geq 1, n\text{Jets_DL1r_60} = 0$
2j, >60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} \geq 1, tZ \text{ BDT} > 0.725$
2j tZ CR	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} \geq 1, tZ \text{ BDT} < 0.725$

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

857 An unfolding procedure is performed to account for differences in the number of recon-
 858 structed jets compared to the number of truth jets in each event. The `AntiKt4TruthDressedWZJets`
 859 truth jet collection is used to make this determination. In order to account for migration of WZ+1-
 860 jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal samples are
 861 separated based on the number of truth jets. Events with 0 jets or more than 3 jets at truth level,

yet fall within one of the categories listed in Table 16, are categorized as WZ + other, and treated as a background. The migration matrix in the number of jets at truth level versus reco level is shown in Figure 13.2. The composition of the number of truth jets in each reco jet bin is taken from MC, with uncertainties in these estimates described in detail in Section 22.

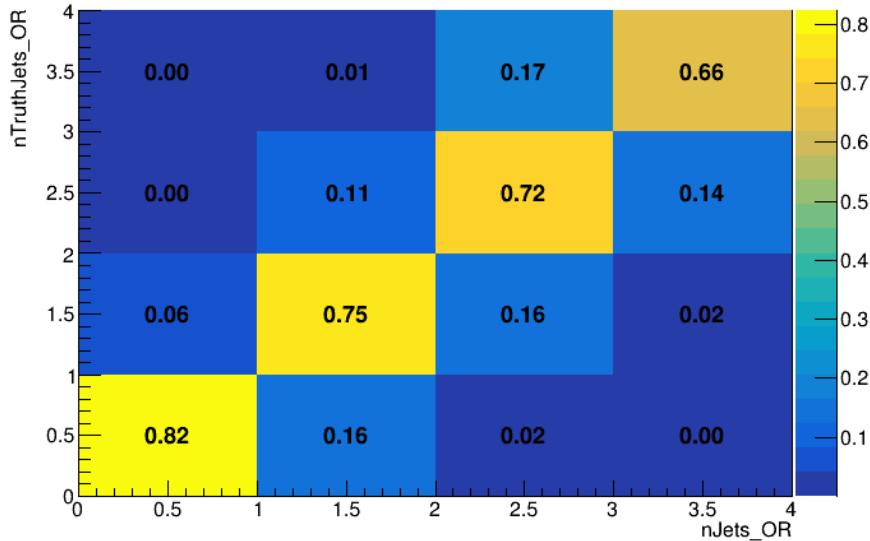


Figure 13.2: Number of truth jets compared to the number of reconstructed jets in each WZ event falling in the preselection region. Each row is normalized to unity.

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

The modeling in each region is validated by comparing data and MC predictions for

⁸⁷² various kinematic distributions. These plot are shown in Figures [13.3-13.16](#).

WZ Fit Region - 1j Inclusive

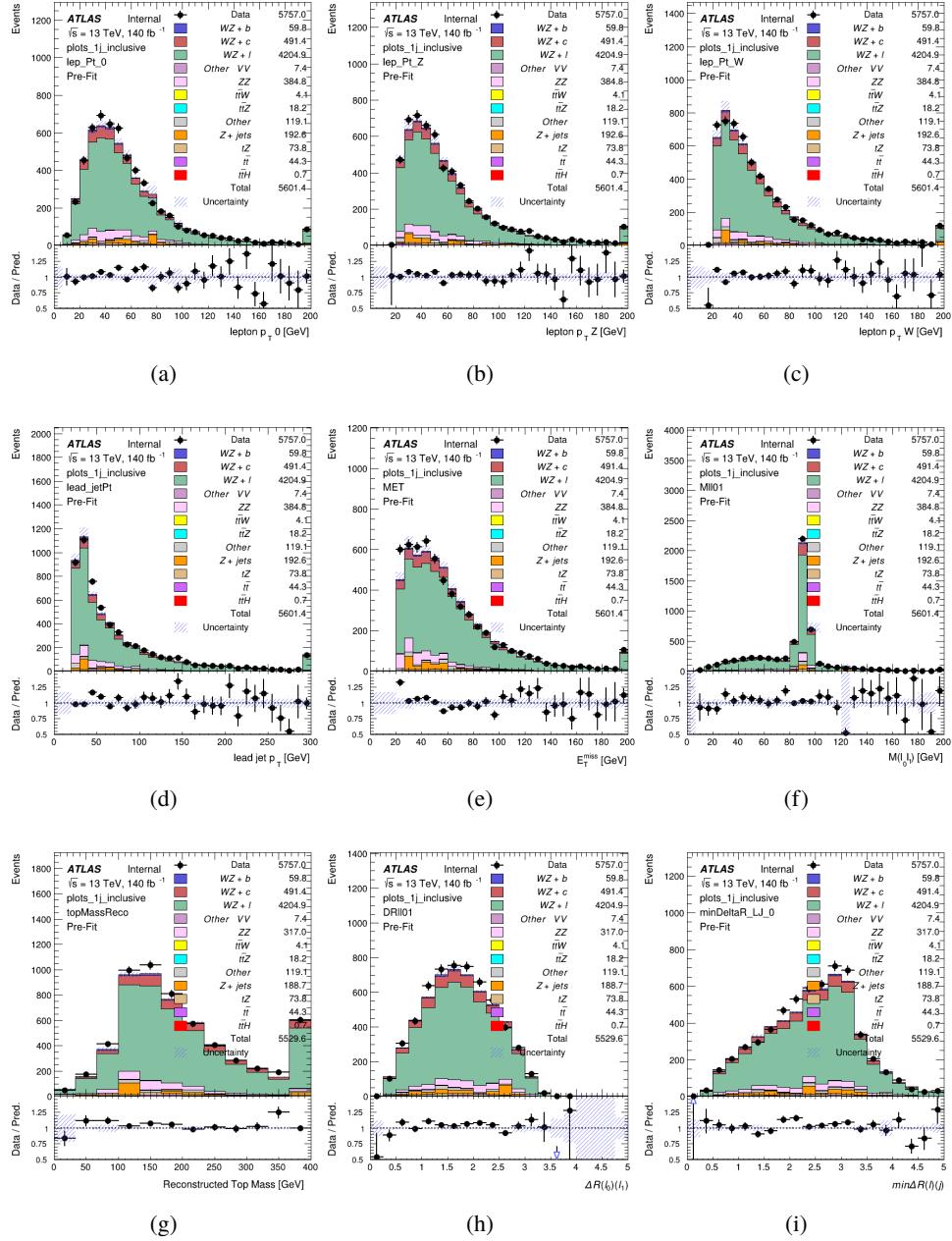


Figure 13.3: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j < 85% WP

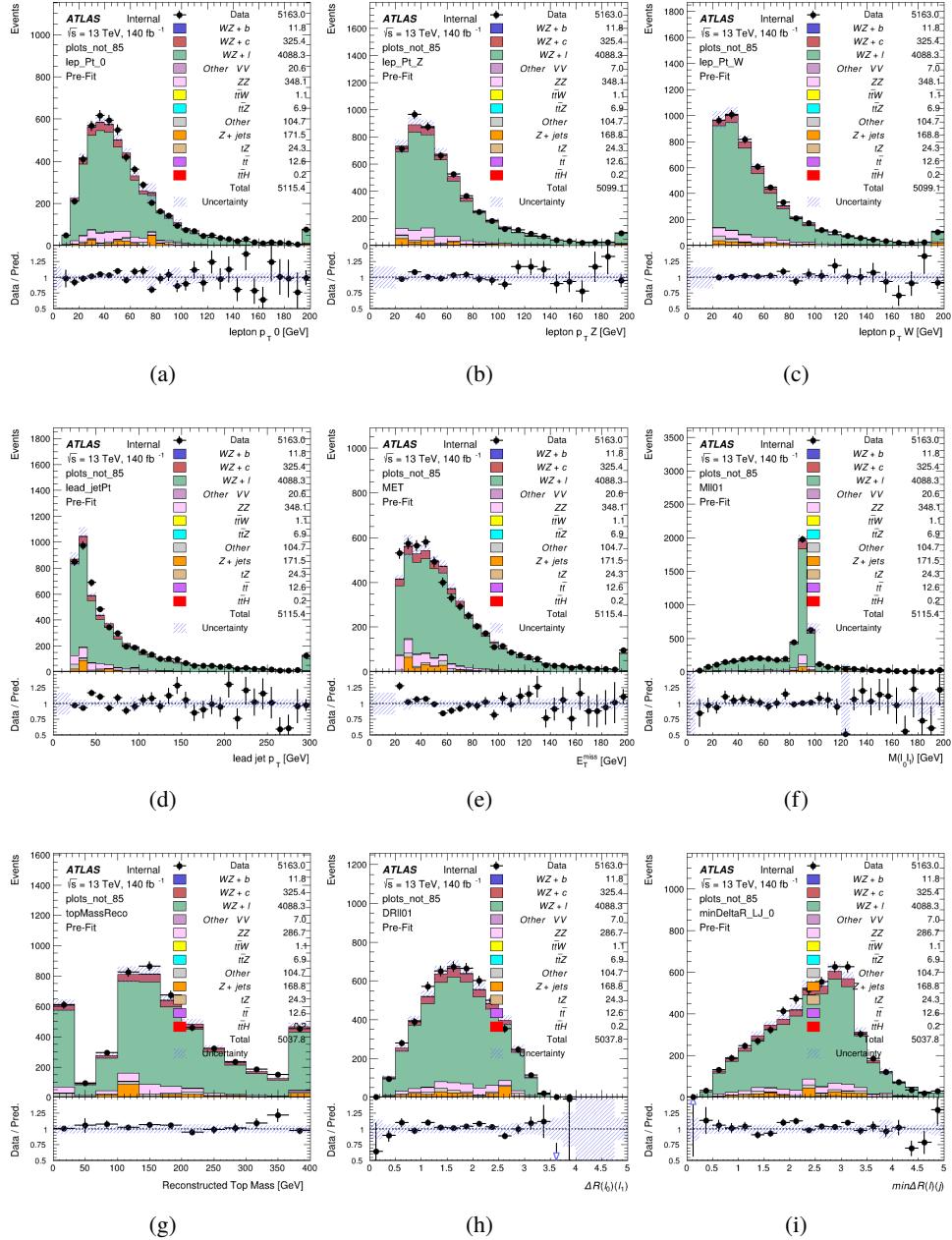


Figure 13.4: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j 77-85% WP

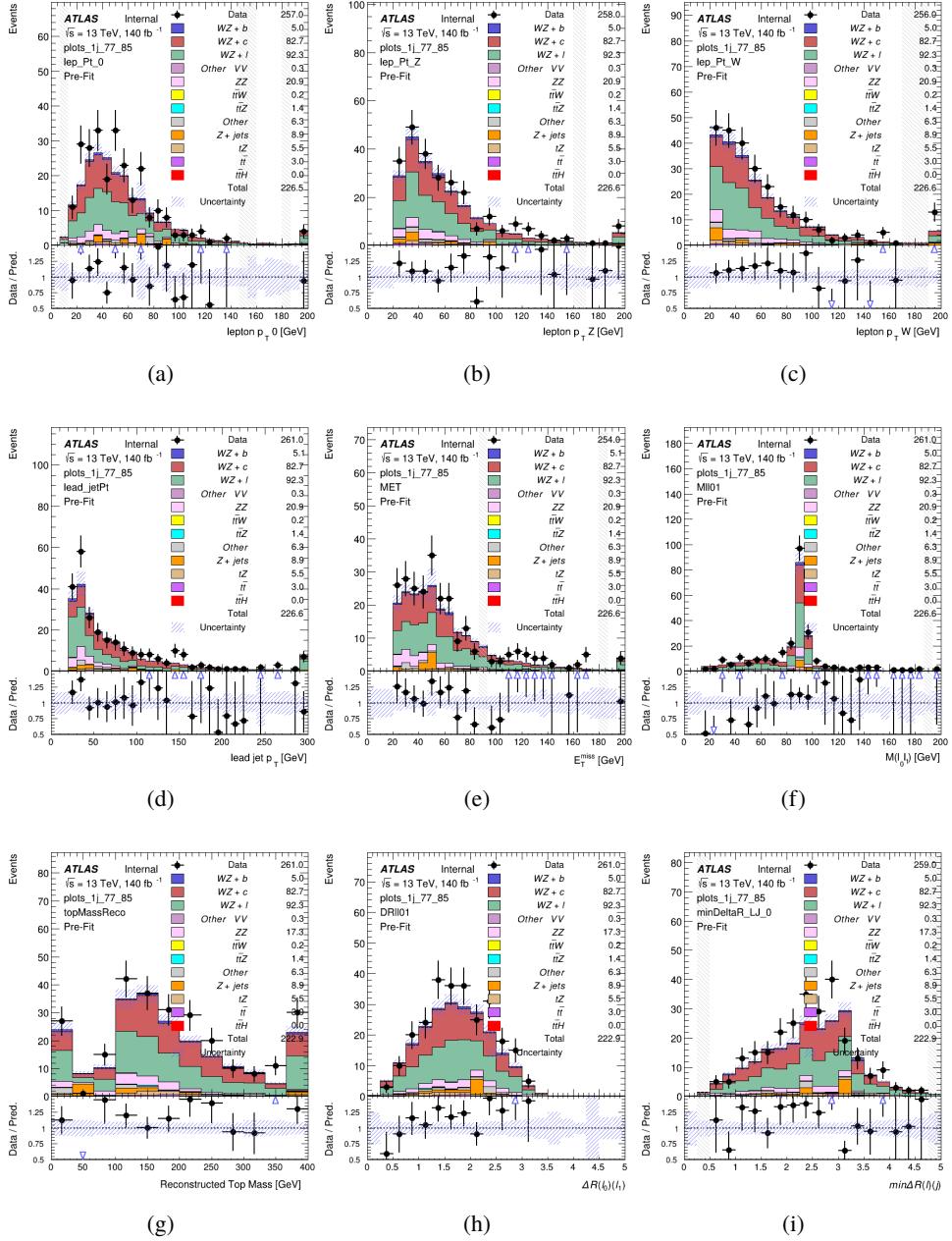


Figure 13.5: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j 70-77% WP

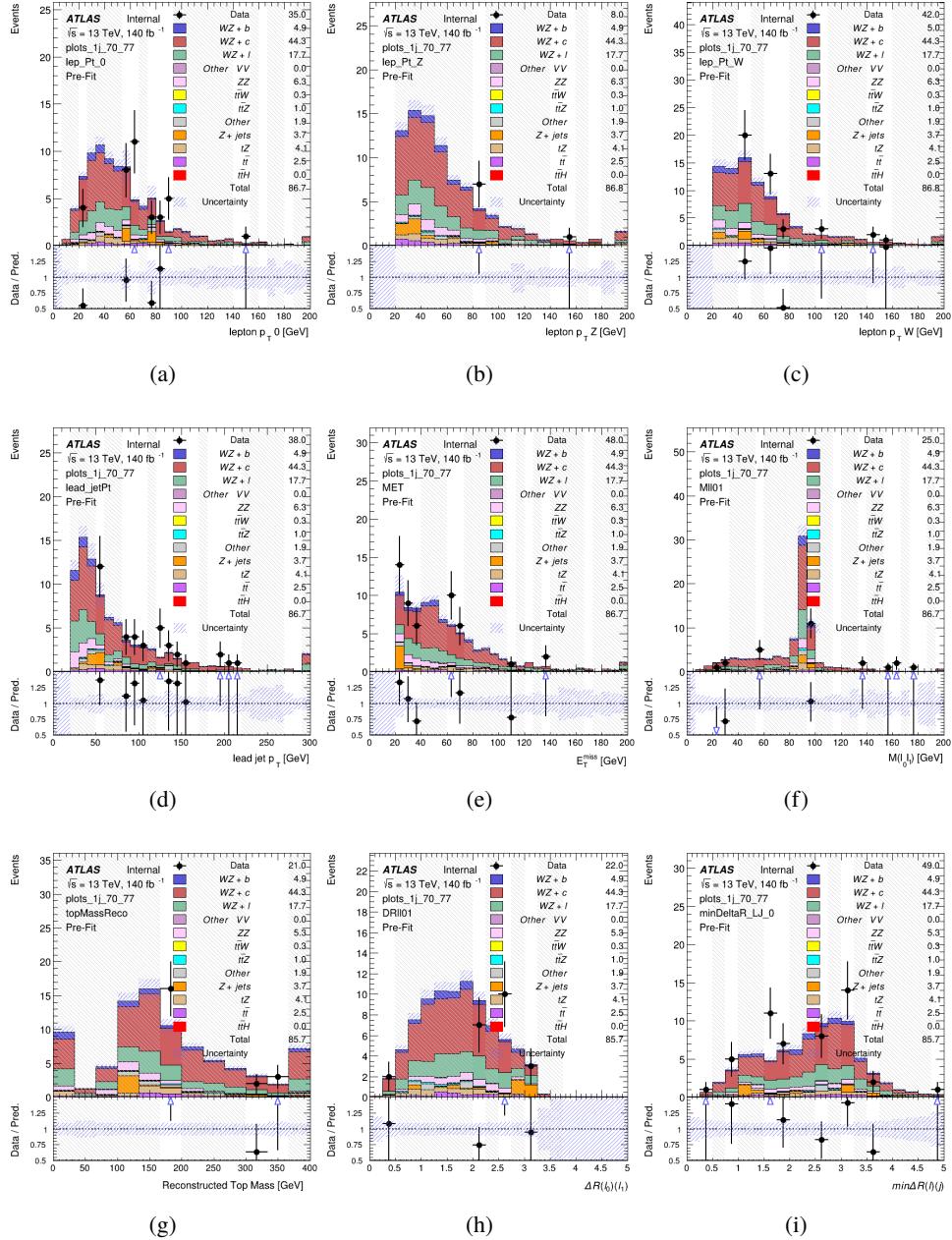


Figure 13.6: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j 60-70% WP

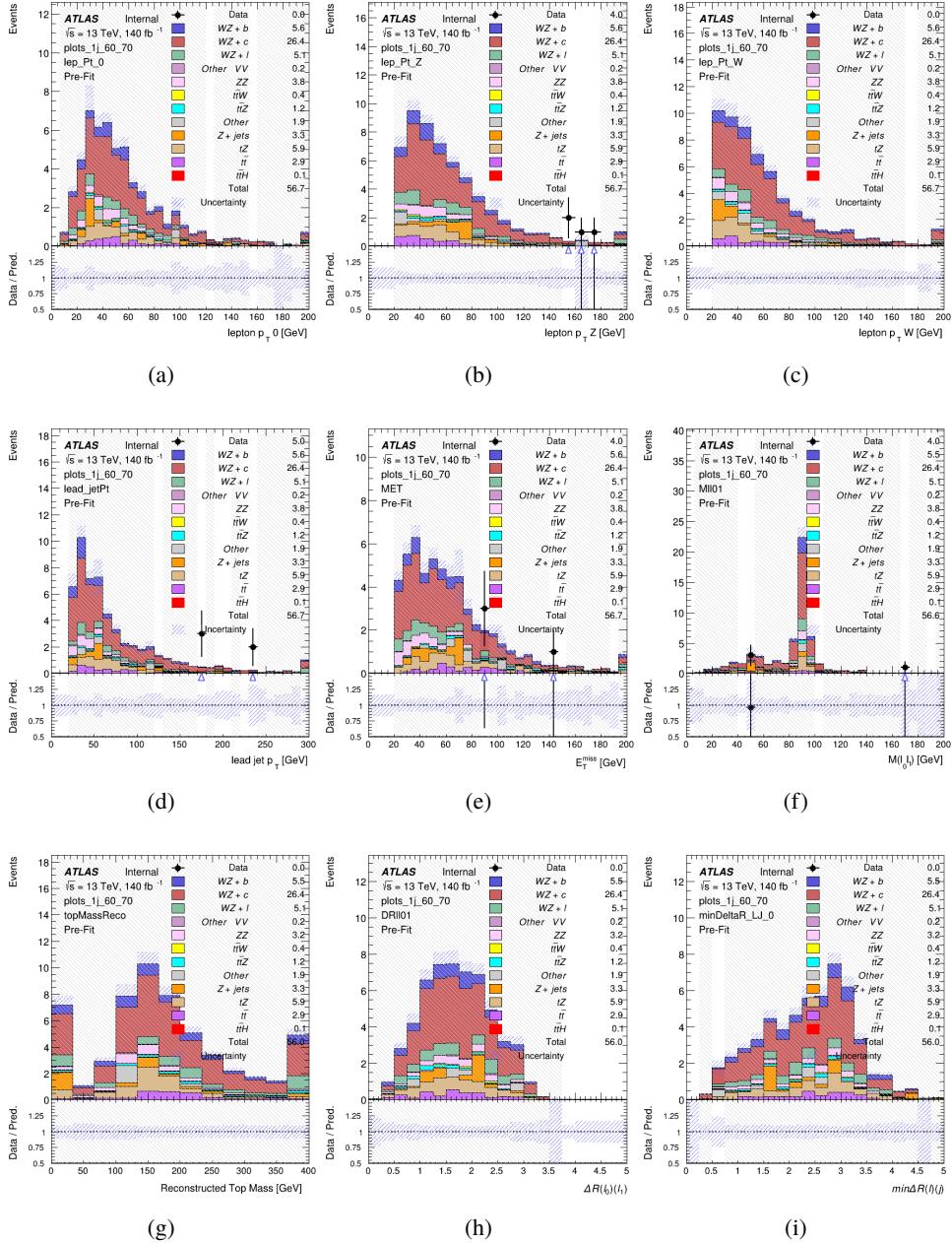


Figure 13.7: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 1j 60% WP

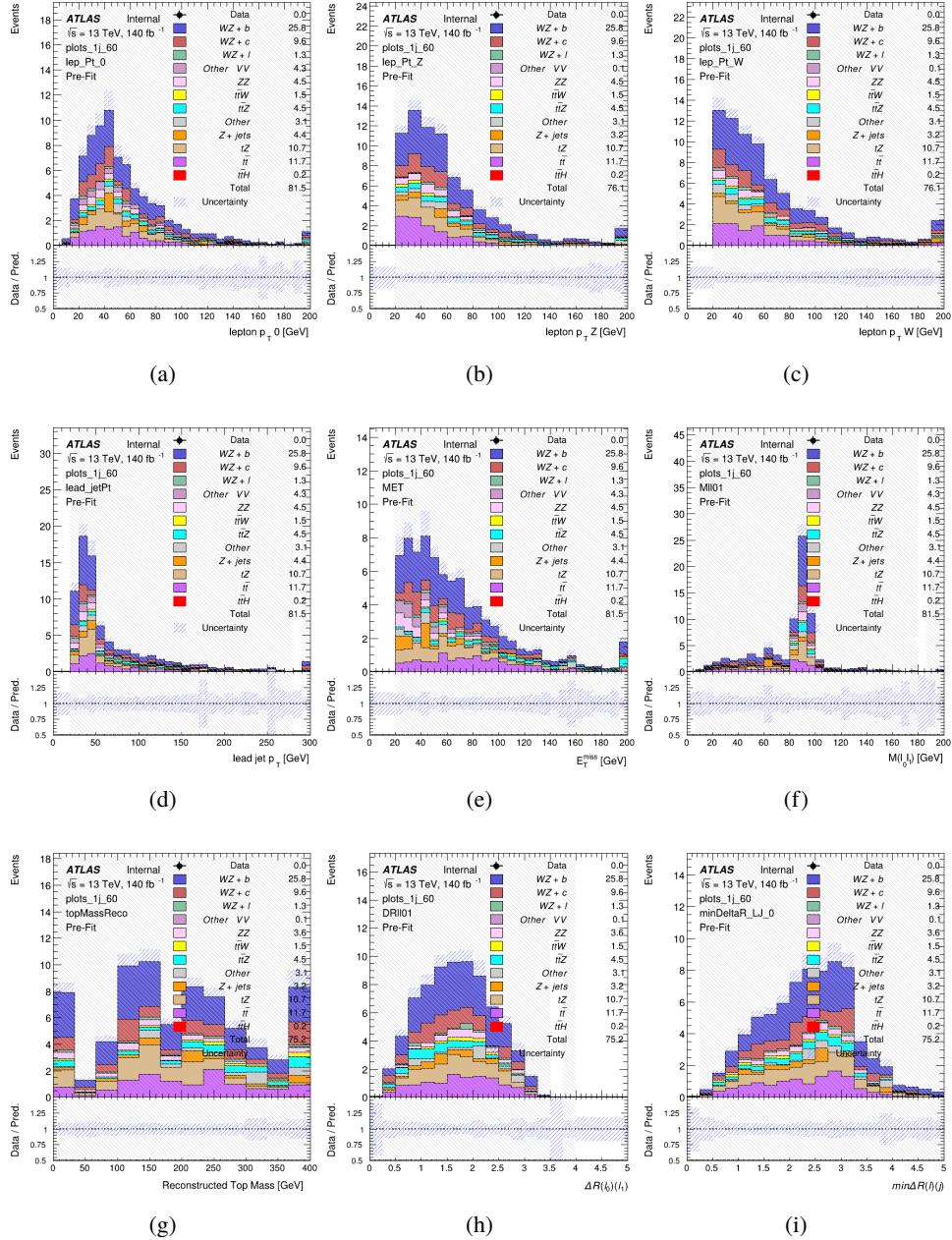


Figure 13.8: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - tZ-CR

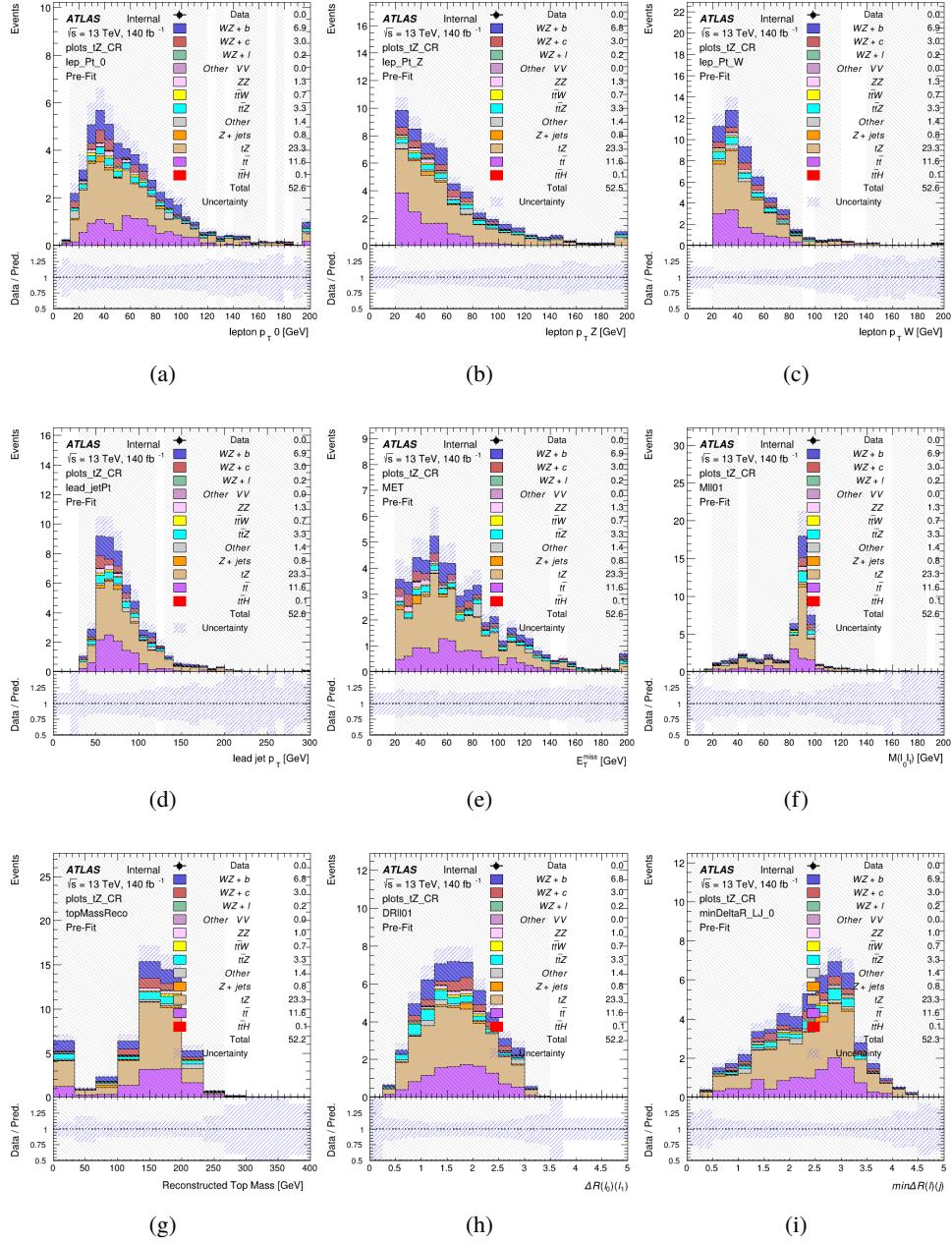


Figure 13.9: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j Inclusive

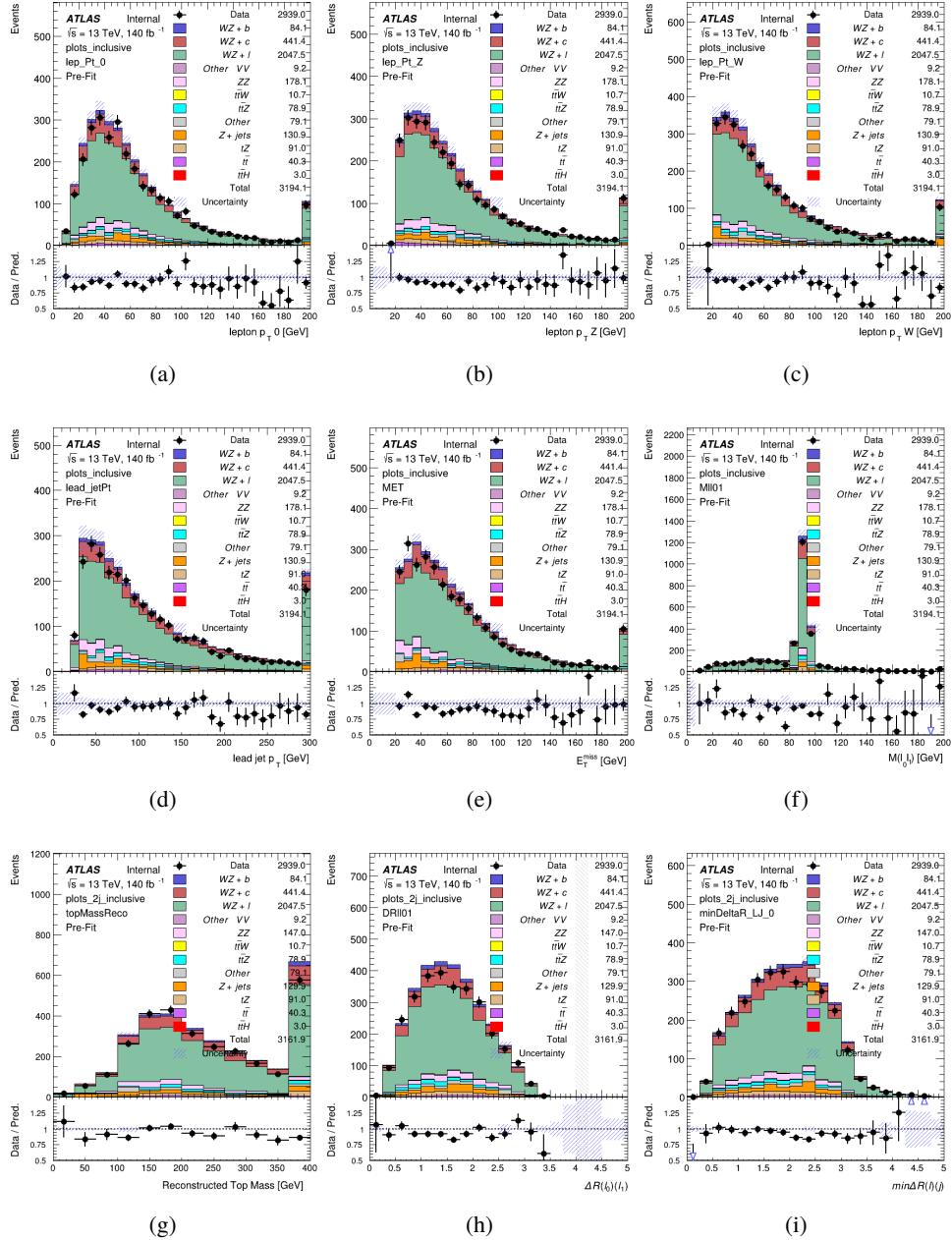


Figure 13.10: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j < 85% WP

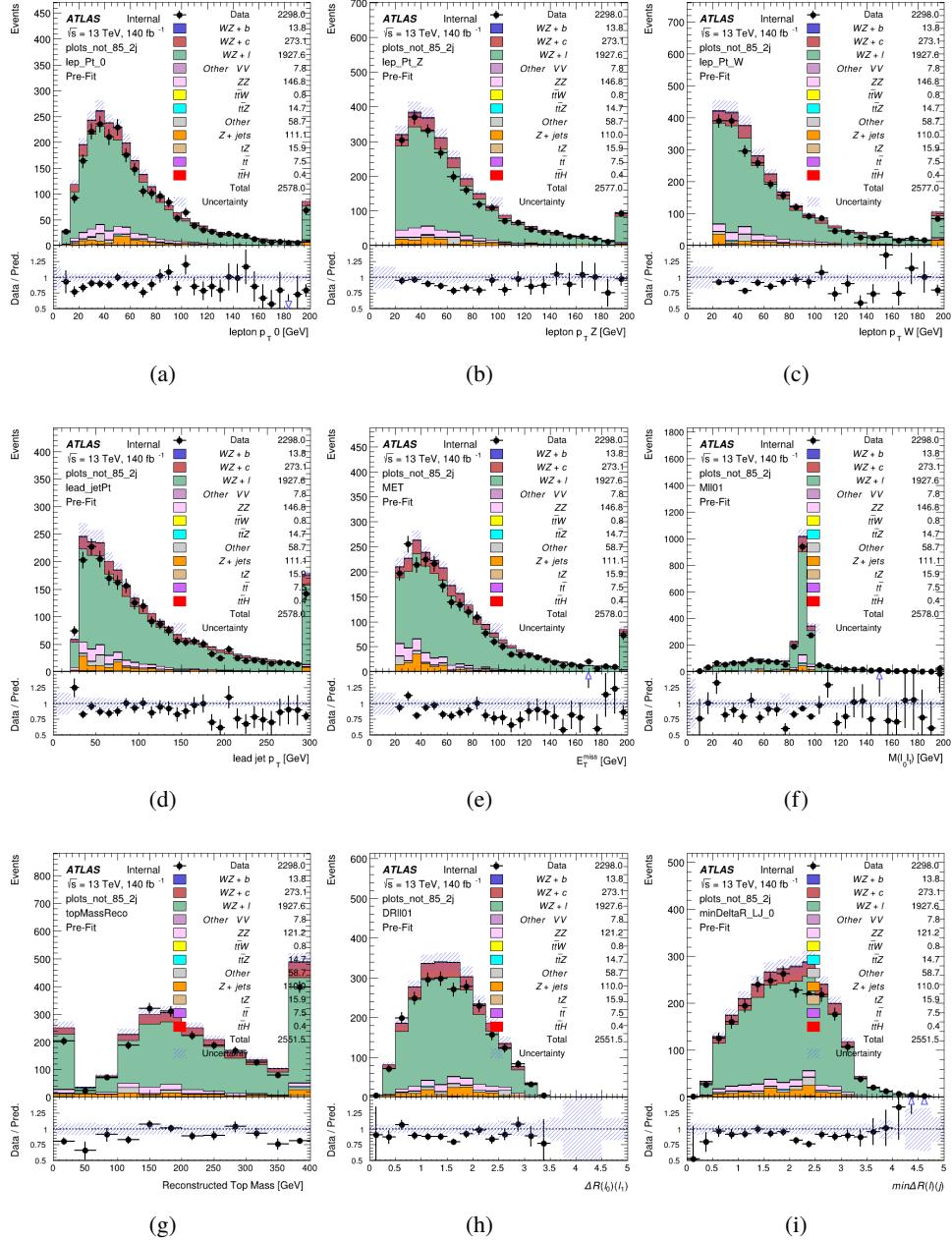


Figure 13.11: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j 77-85% WP

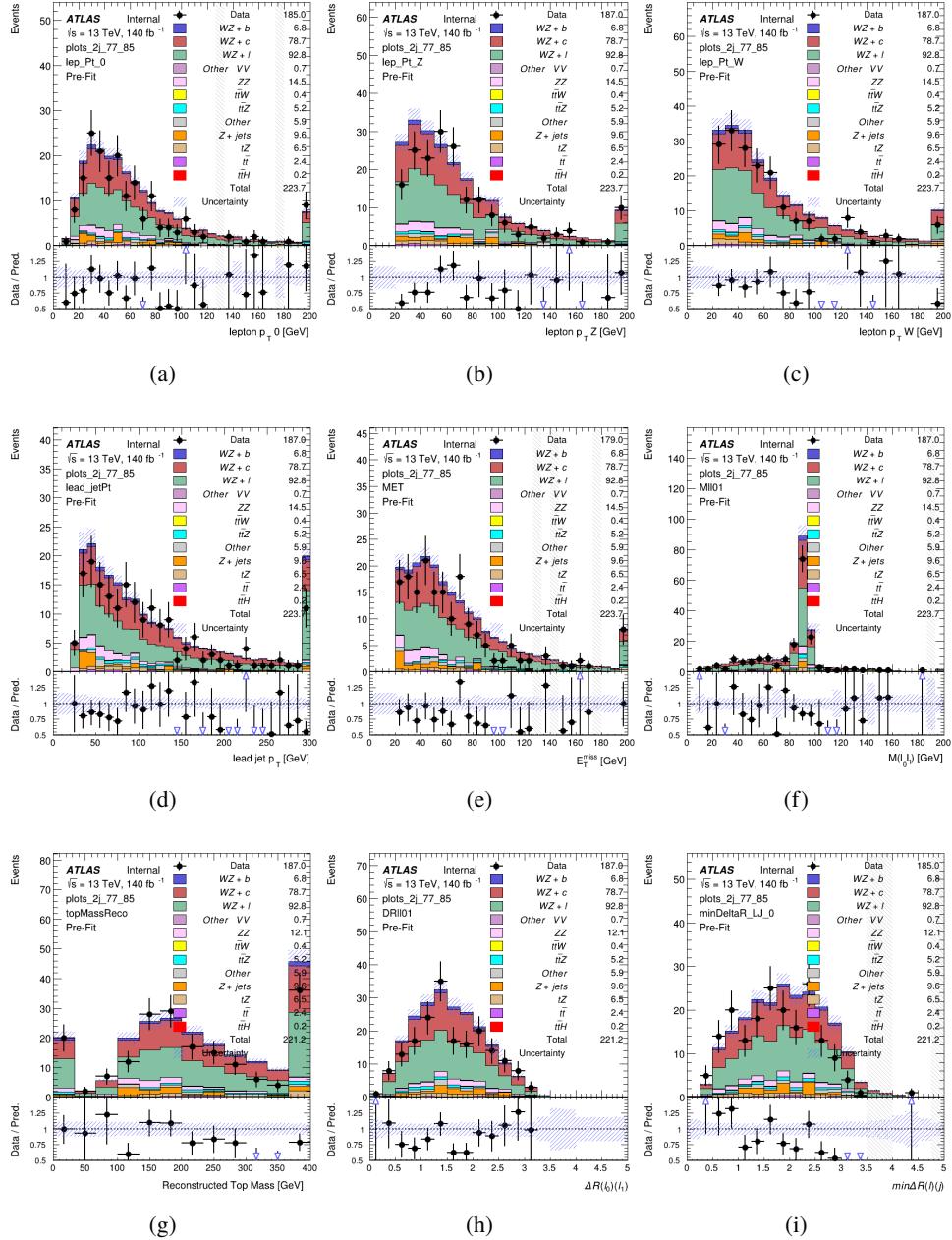


Figure 13.12: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j 70-77% WP

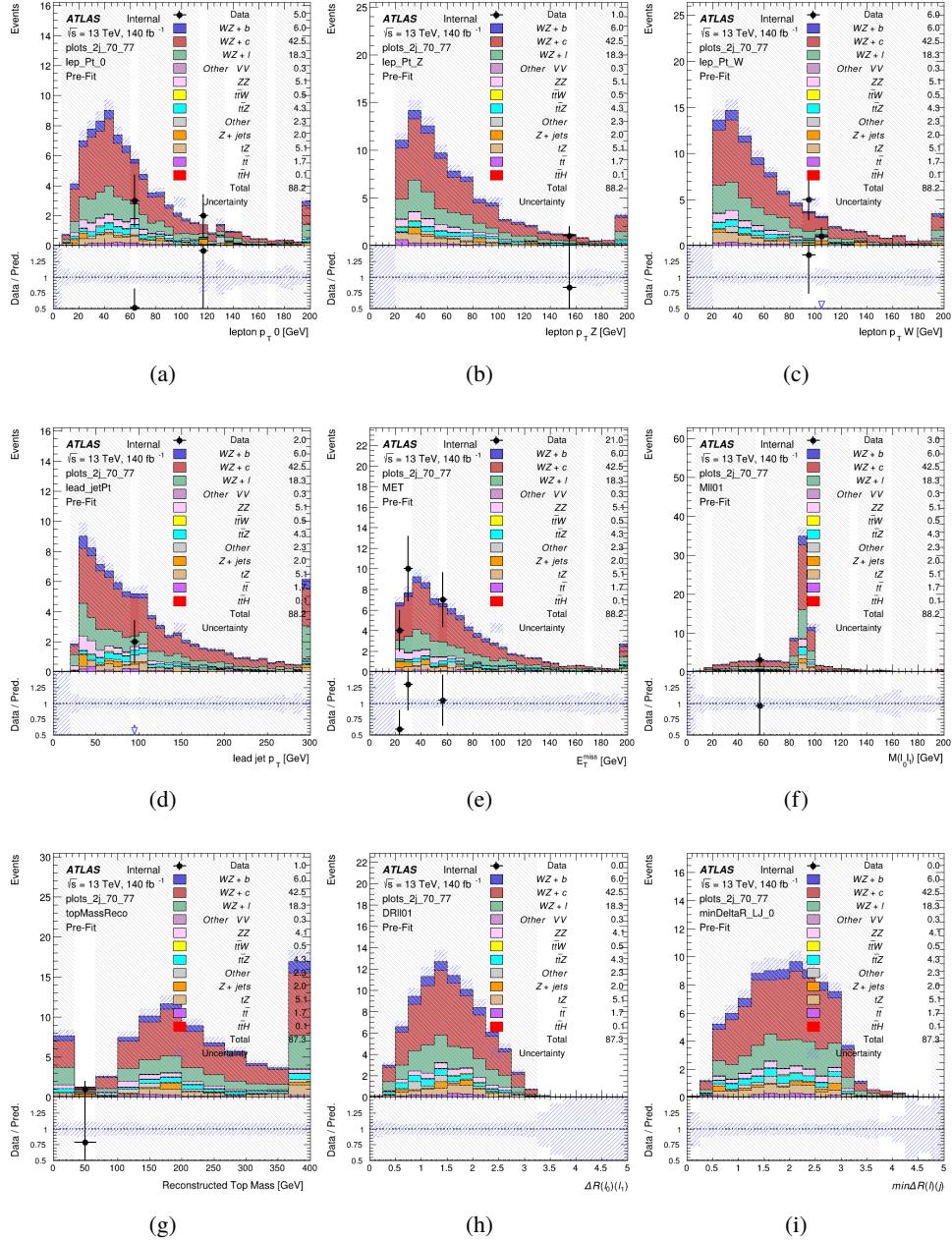


Figure 13.13: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j 60-70% WP

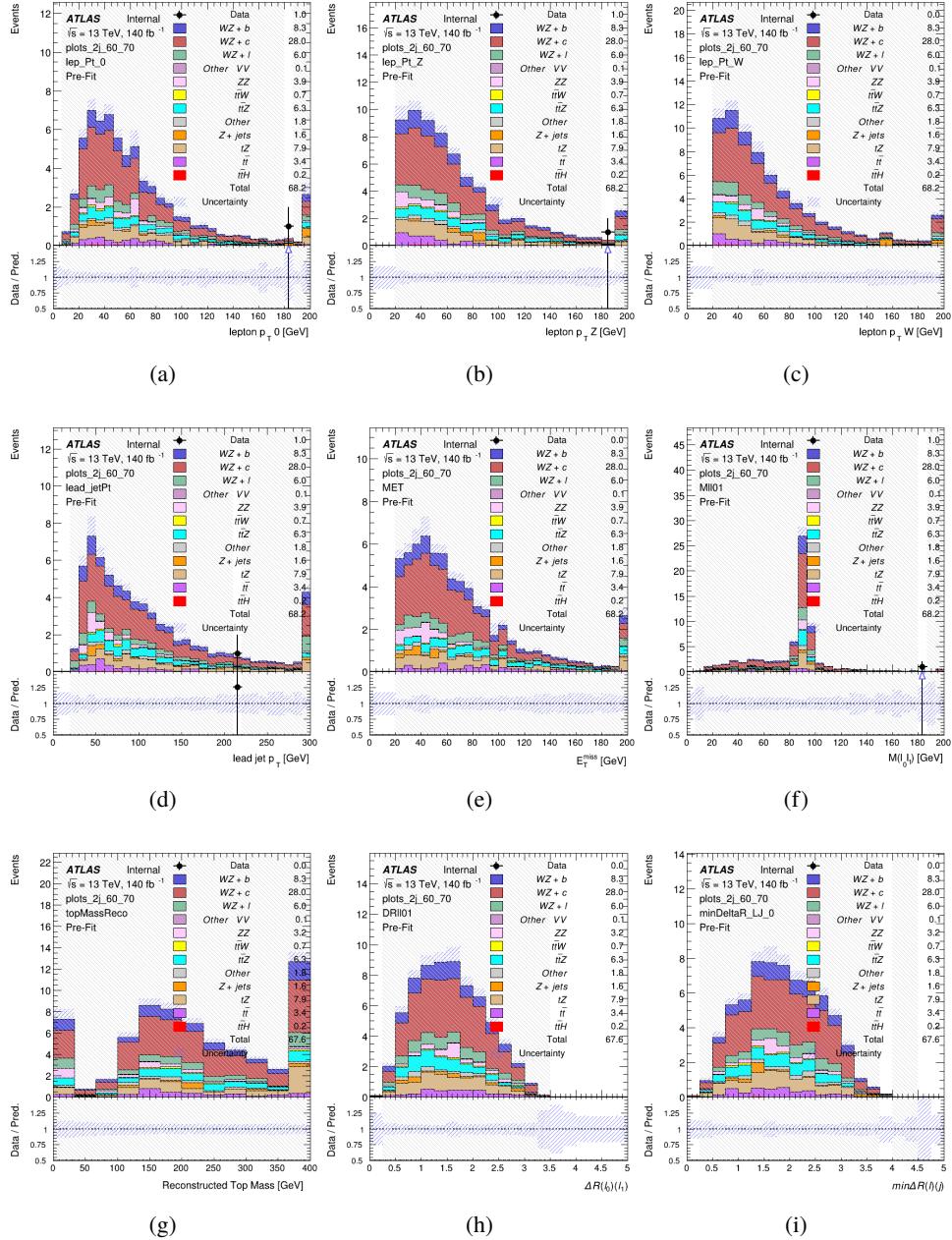


Figure 13.14: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - 2j 60% WP

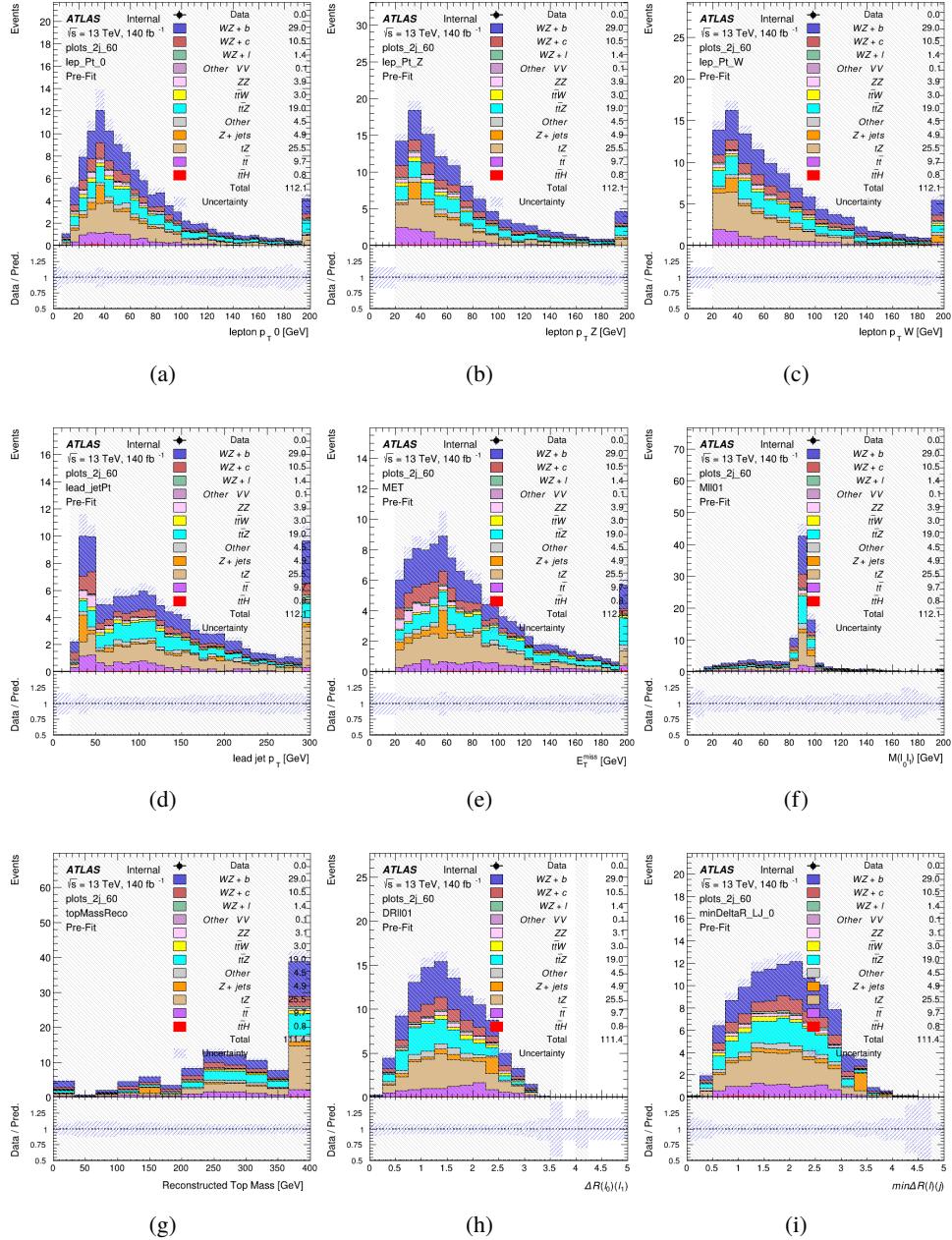


Figure 13.15: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

WZ Fit Region - tZ-CR-2j

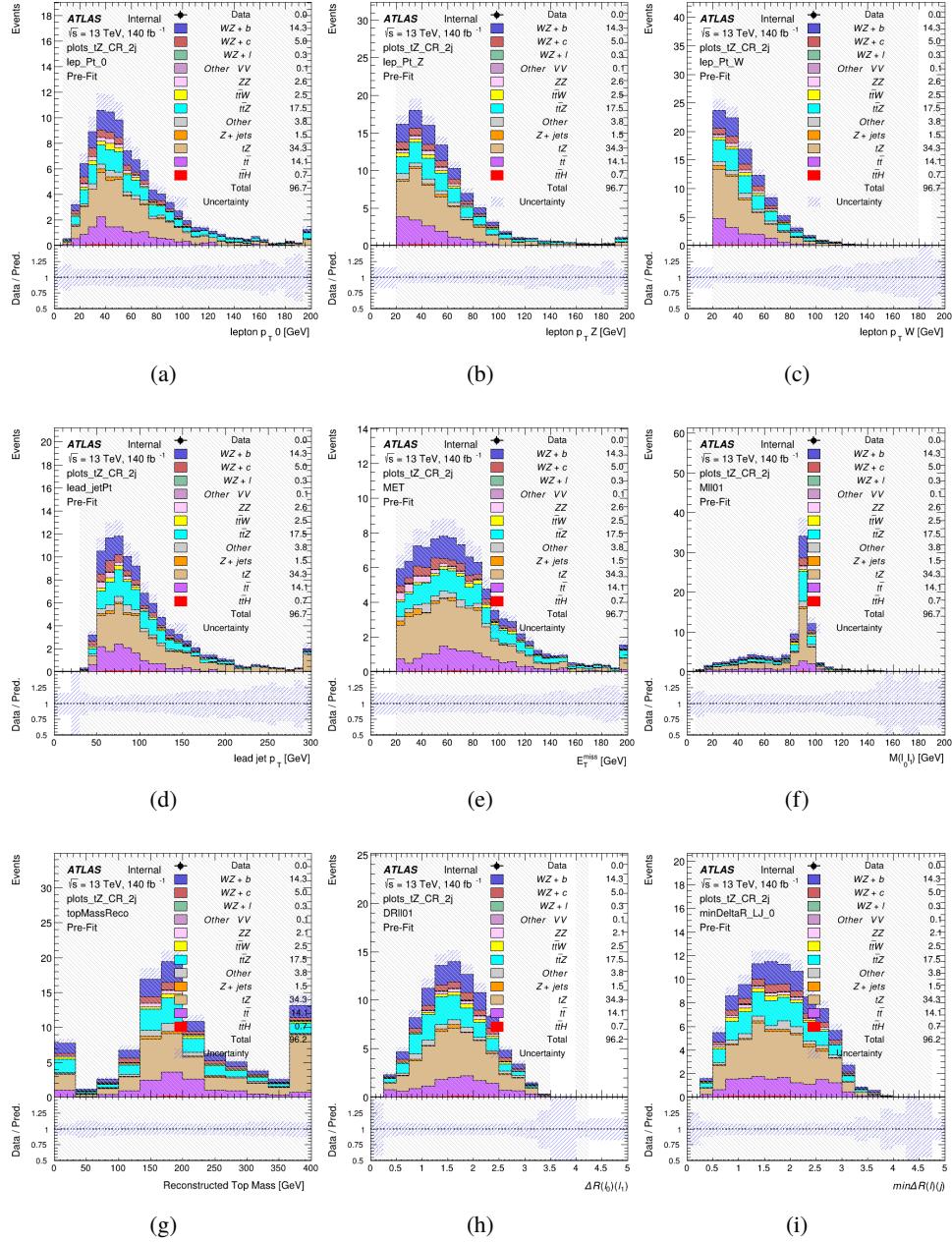


Figure 13.16: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant mass of leptons 0 and 1, (g) the reconstructed top mass, (h) $\Delta(R)$ between lepton 0 and 1, (i) the $\Delta(R)$ between the closest lepton and jet.

873 **13.3 Non-Prompt Lepton Estimation**

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

878 **13.3.1 $t\bar{t}$ Validation**

879 $t\bar{t}$ events can produce two prompt leptons from the decay of each of the tops. These top decays
 880 produce two b-quarks, the decay of which can produce additional non-prompt leptons, which
 881 occasionally pass the event preselection. In order to validate that the Monte Carlo accurately
 882 simulates this process accurately, the MC prediction in a non-prompt $t\bar{t}$ enriched control region
 883 is compared to data.

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

889 Further, because the jet multiplicity of $t\bar{t}$ events tends to be higher than WZ, the number
 890 of jets in each event is required to be greater than 1. As b-jets are almost invariably produced

⁸⁹¹ from top decays, at least one b-tagged jet passing the 70% DL1r WP in each event is required.

⁸⁹² Various kinematic plots of this region are shown in Figure 13.17.

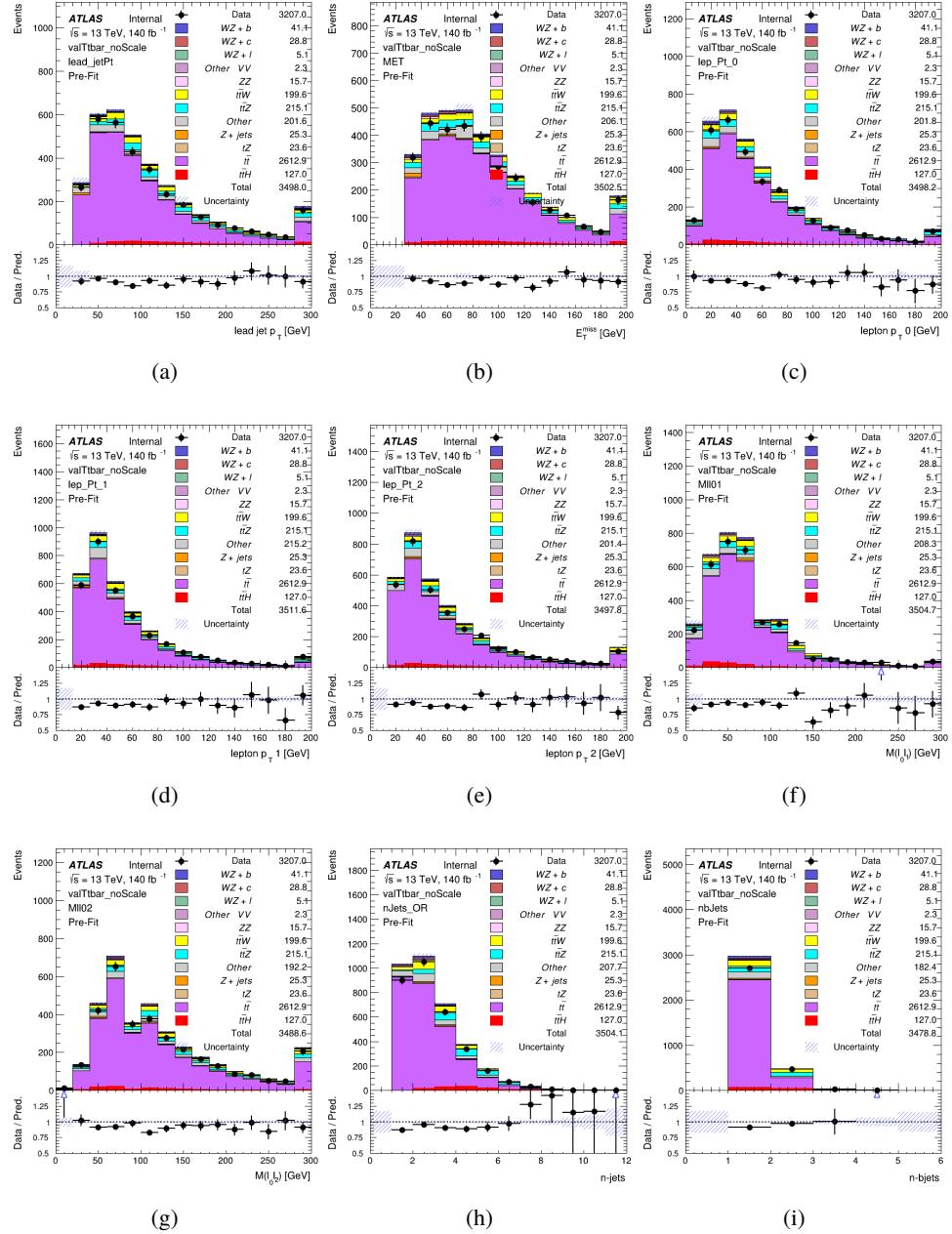


Figure 13.17: Comparisons between the data and MC distributions in the $t\bar{t}$ control region for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass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

893 The shape of each distribution agrees quite well between data and MC, with a constant
894 offset between the two. This is accounted for by applying a constant correction factor of 0.883
895 to the $t\bar{t}$ MC prediction. Plots showing the kinematics of the $t\bar{t}$ VR after this correction factor
896 has been applied are shown in Figure 13.18.

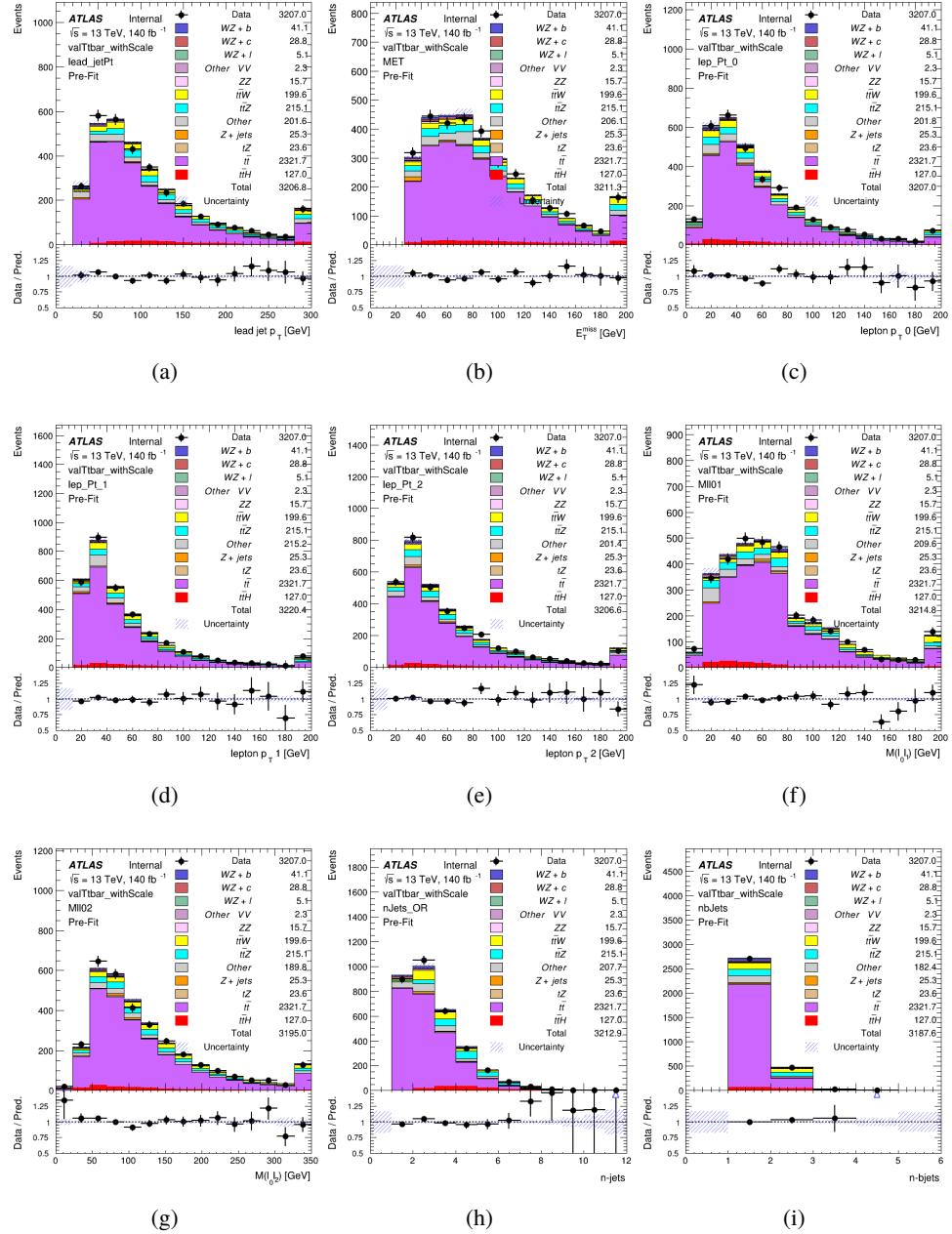


Figure 13.18: Comparisons between the data and MC distributions in the $t\bar{t}$ control region after the correction factor has been applied for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass 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.

897 The modeling is further validated by looking at the yield in the $t\bar{t}$ VR for each DL1r WP,
 898 giving a clearer correspondence to the signal regions used in the fit. Each region shown in Figure
 899 13.19 requires one or more jets pass the listed WP, with no jets passing the next highest WP.

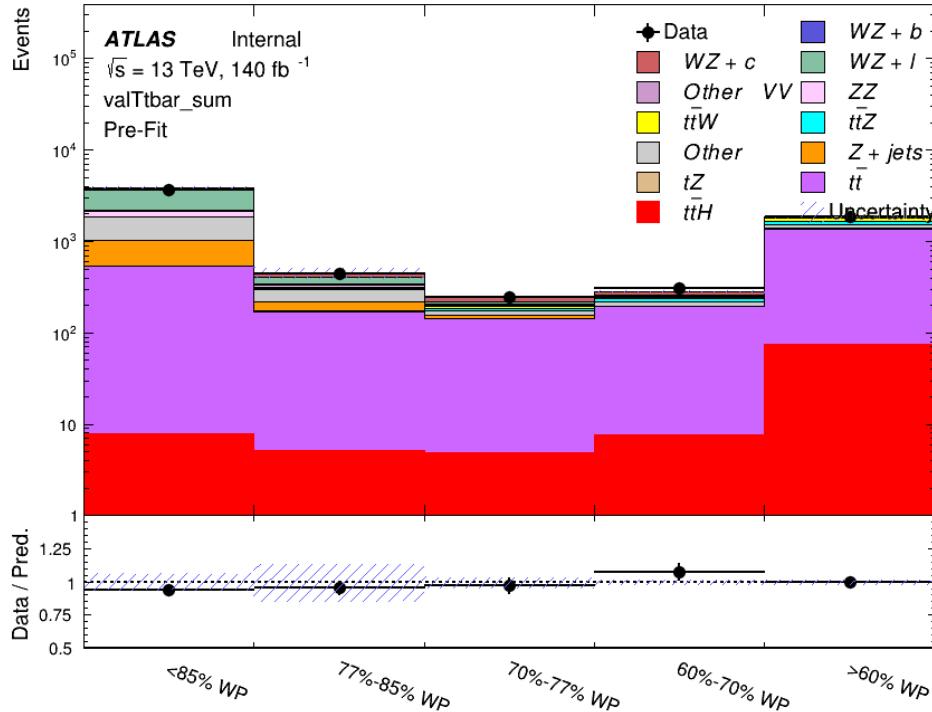


Figure 13.19: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the $t\bar{t}$ VR selection and correction factor have been applied

900 As data and MC are found to agree within 20% for each of these working points, a 20%
 901 systematic uncertainty on the $t\bar{t}$ prediction is included for the analysis.

902 13.3.2 Z+jets Validation

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

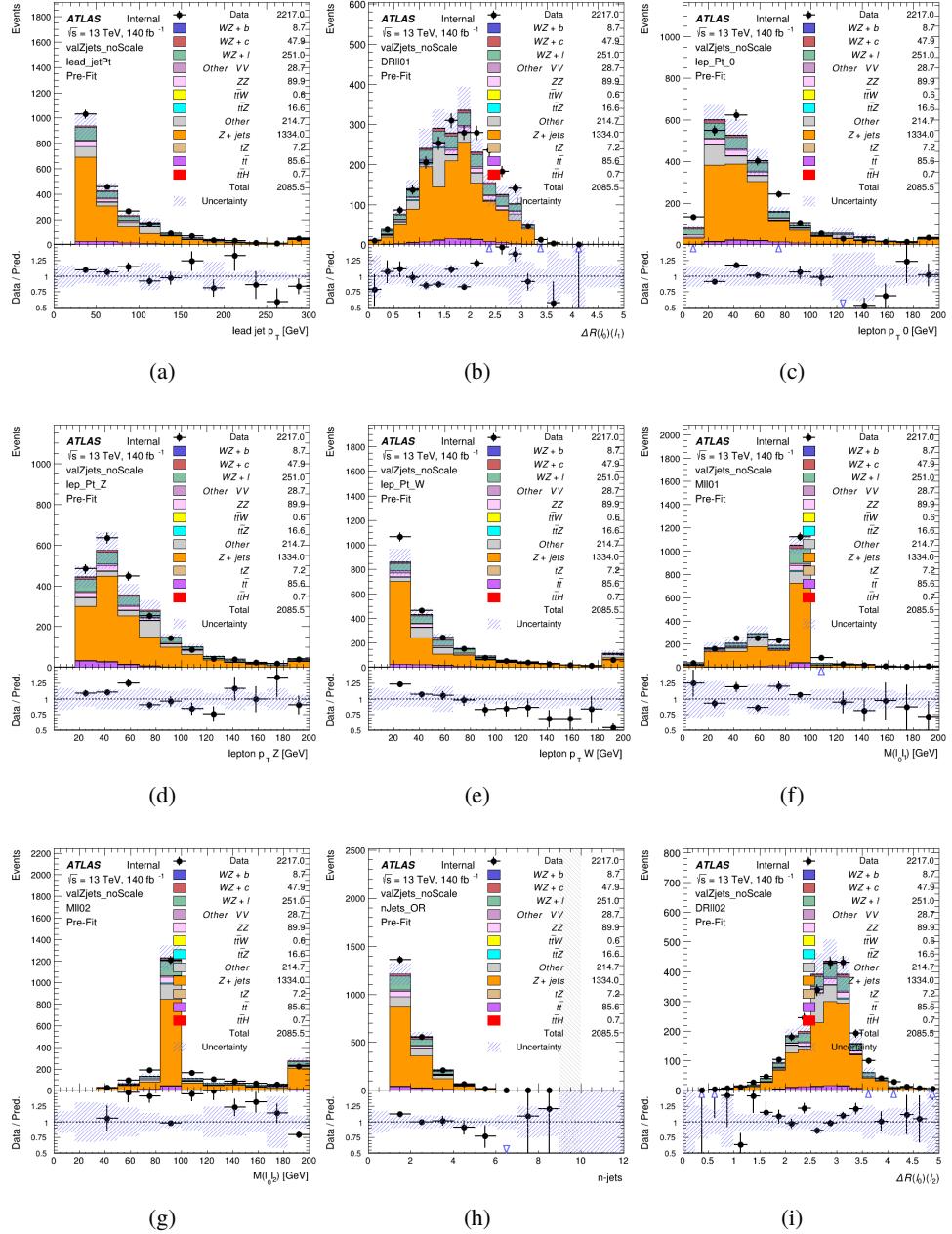


Figure 13.20: Comparisons between the data and MC distributions in the $Z + \text{jets}$ control region for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant 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

910 While there is general agreement between data and MC within statistical uncertainty, the
911 shape of the p_T spectrum of the lepton from the W is found to differ. To account for this
912 discrepancy, a variable correction factor is applied to Z+jets. χ^2 minimization of the lepton 2 p_T
913 spectrum is performed to derive a correction factor of $1.53 - 6.6 * 10^{-6}(\text{lep}_\text{Pt}_W)$. Kinematic
914 plots of the Z + jets control region after this correction factor has been aplied are shown in Figure
915 [13.21](#).

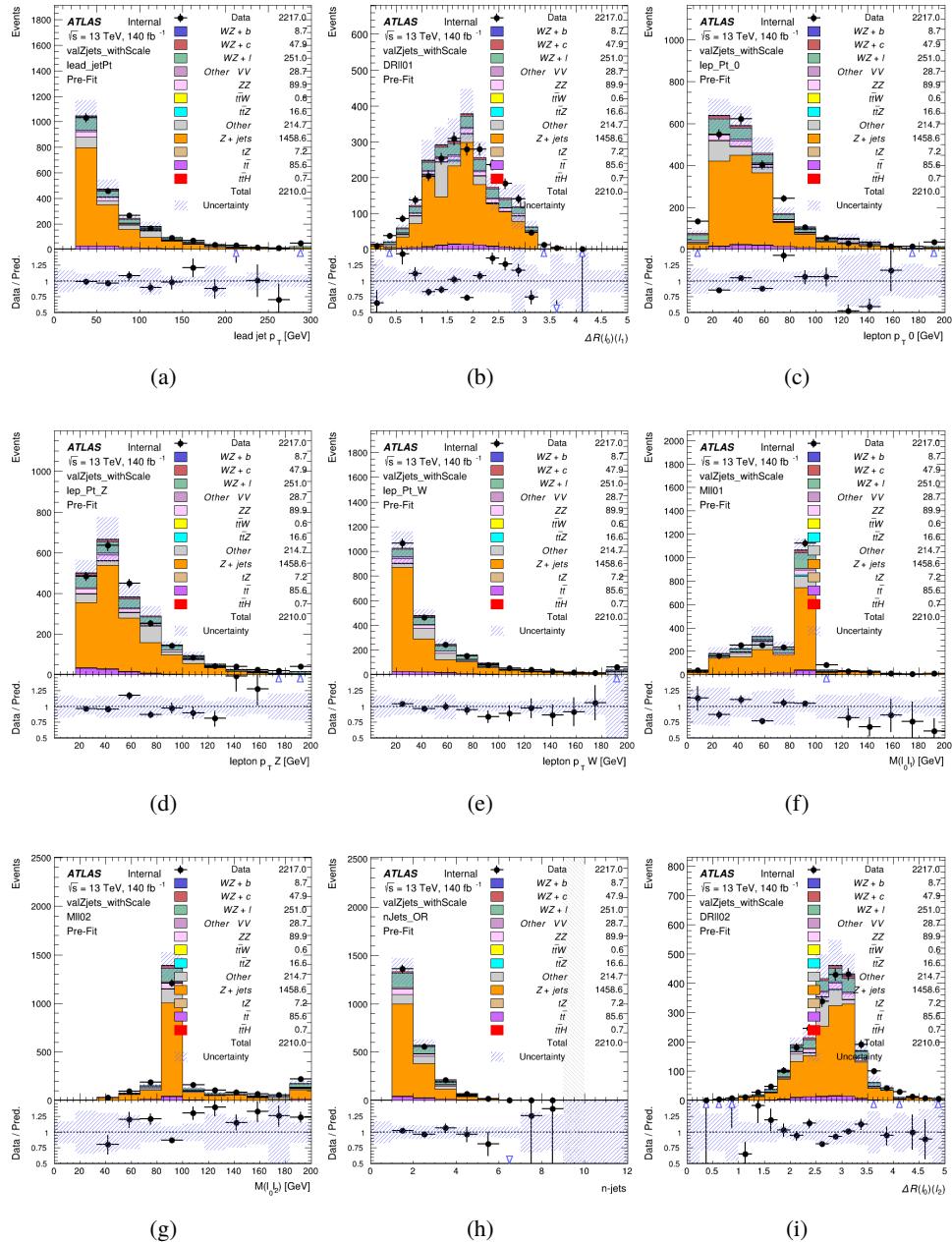


Figure 13.21: Comparisons between the data and MC distributions in the $Z + \text{jets}$ control region after the correction factor has been applied for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS 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

916 The modeling is further validated by looking at the yield in the Z+jets VR for each DL1r
 917 WP, giving a clearer correspondence to the signal regions used in the fit. Each region shown in
 918 Figure 13.22 requires one or more jets pass the listed WP, with no jets passing the next highest
 919 WP.

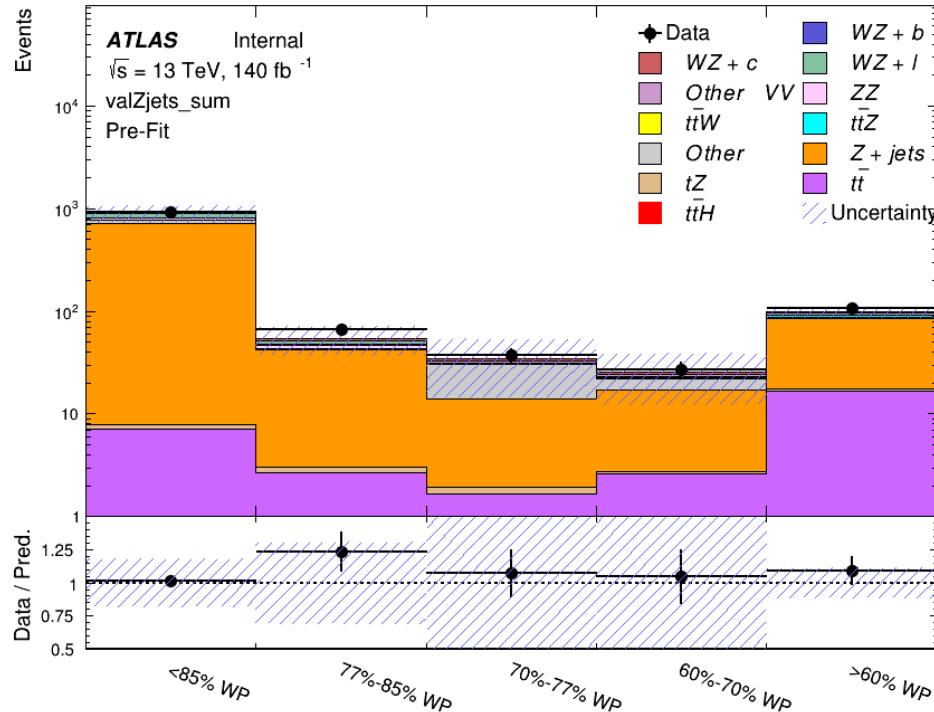


Figure 13.22: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the Z+jets VR selection and correction factor have been applied

920 For each of the b-tagging working points considered, the data falls within 25% of the MC
 921 prediction once this correction factor has been applied. Therefore, a 25% systematic uncertainty
 922 is applied to Z + jets in the analysis.

923 14 tZ Interference Studies and Separation Multivariate Analysis

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

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

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

933 14.1 Top Mass Reconstruction

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

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

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

$$944 \quad (p_l + p_\nu)^2 = m_W^2$$

945 Expanding this out into components, this equation gives:

$$946 \quad \sqrt{p_{T\nu}^2 + p_{z\nu}^2} E_l = \frac{m_W^2 - m_l^2}{2} + p_{T\nu}(p_{lx}\cos\phi_\nu + p_{ly}\sin\phi_\nu) + p_{lz}p_{\nu z}$$

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

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

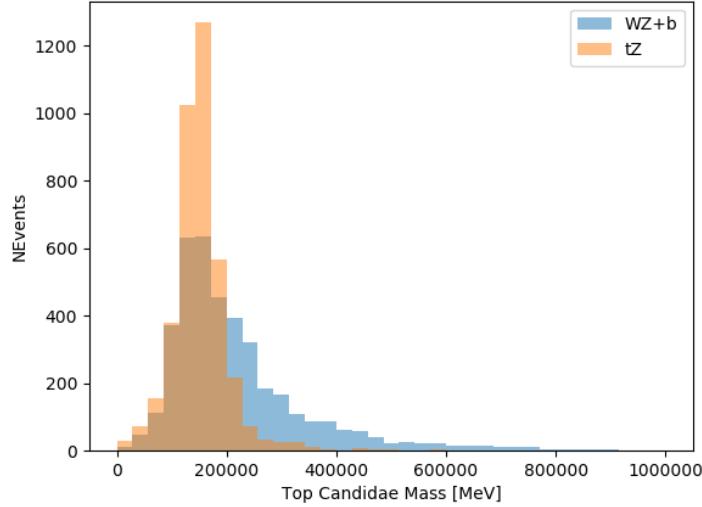


Figure 14.1: Reconstructed top mass distributions for tZ and WZ + b, measured in MeV.

953 14.2 tZ BDT

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

956 • The invariant mass of the reconstructed top candidate

957 • p_T of each of the leptons, jet

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

959 • E_T^{miss}

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

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

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

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

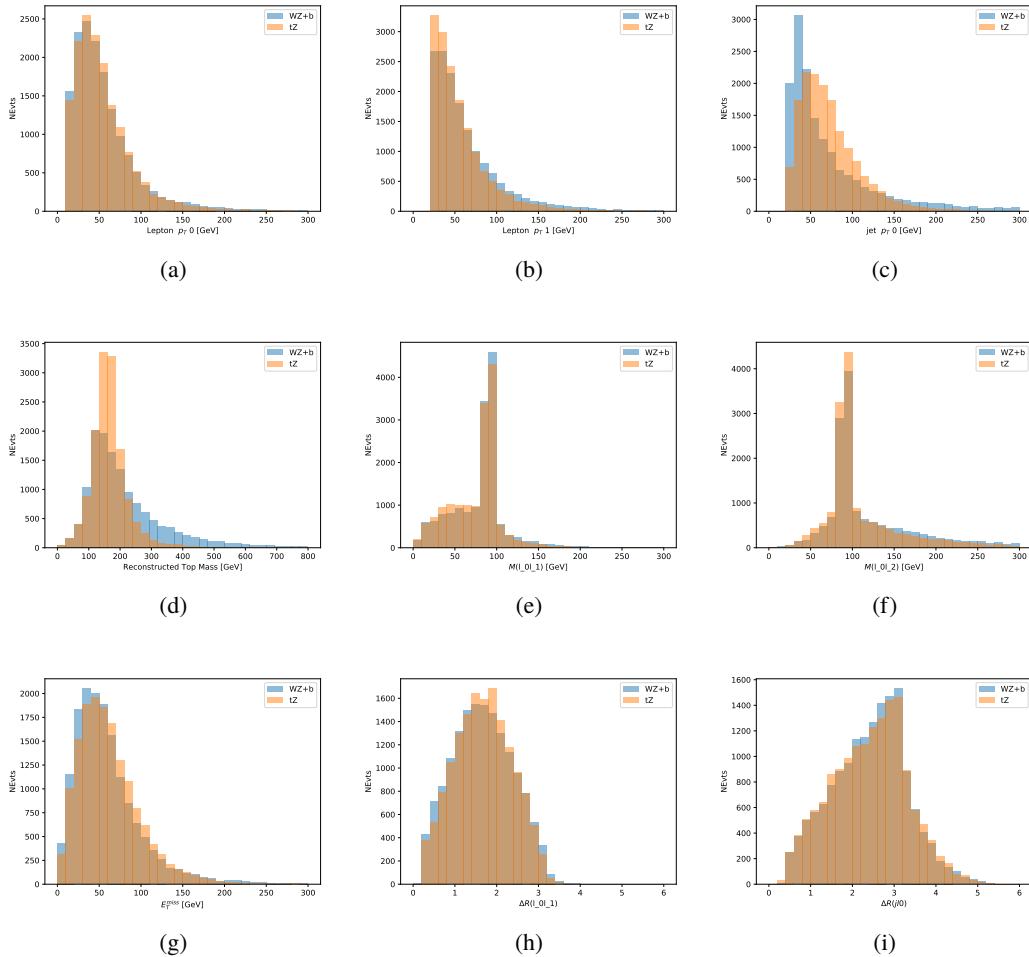


Figure 14.2: Distribution of input features of the BDT for signal (WZ) and background (tZ). Both are scaled to an equal number of events. (a), (b) and (c) show the p_T of lepton 0, lepton 1, and the jet, (d) show the reconstructed top mass, (e) and (f) show the invariant mass of leptons 0 and 1, leptons 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.

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

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

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

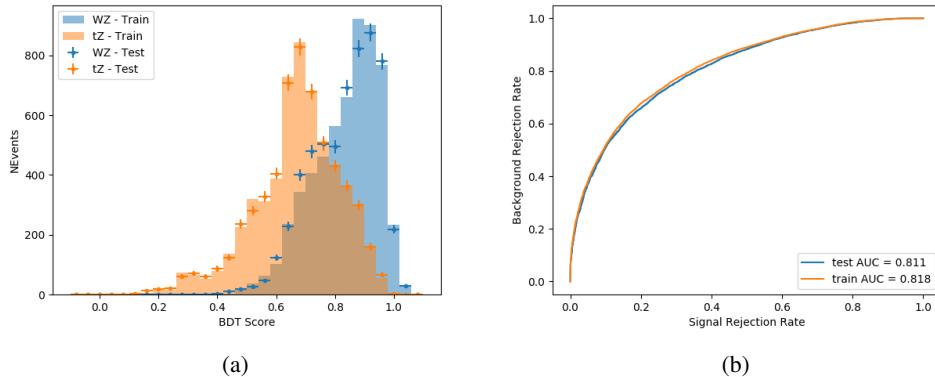


Figure 14.3: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

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

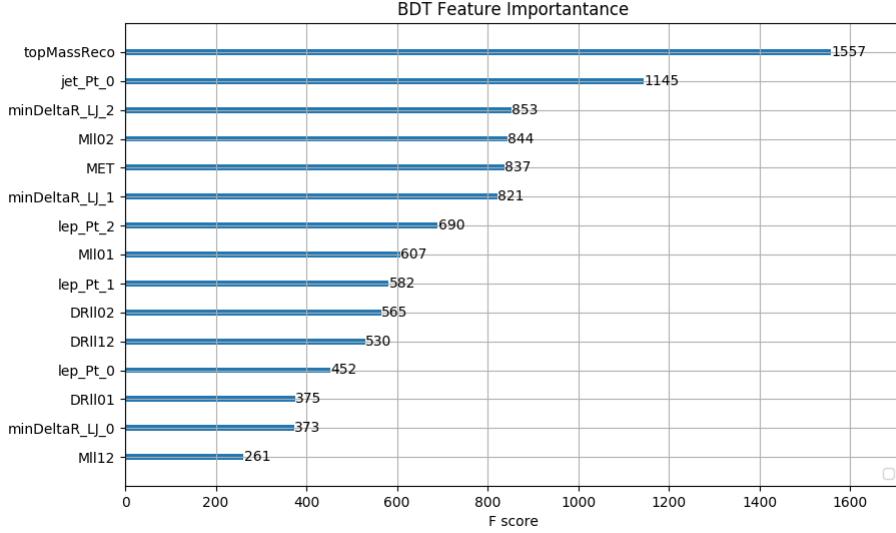


Figure 14.4: Relative importance of each input feature in the model.

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

986 15 Systematic Uncertainties

987 The systematic uncertainties that are considered are summarized in Table 49. These are imple-
 988 mented in the fit either as a normalization factors or as a shape variation or both in the signal

and background estimations. The numerical impact of each of these uncertainties is outlined in
 Section 23.

Table 17: Sources of systematic uncertainty considered in the analysis.

Systematic uncertainty	Components
Luminosity	1
Pileup reweighting	1
Physics Objects	
Electron	6
Muon	15
Jet energy scale	28
Jet energy resolution	8
Jet vertex fraction	1
Jet flavor tagging	131
E_T^{miss}	3
Total (Experimental)	194
Signal Modeling	
Shape modelling	3
Renormalization and factorization scales	5
nJet Migration	4
Background Modeling	
Cross section	15
Renormalization and factorization scales	12
Total (Signal and background modeling)	35
Total (Overall)	229

The uncertainty in the combined integrated luminosity is derived from a calibration of the
 luminosity scale performed for 13 TeV proton-proton collisions [**lumi**], [**LUCID2**].

The experimental uncertainties are related to the reconstruction and identification of light
 leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The TOTAL electron ID
 correlation model is used, corresponding to 1 electron ID systematic. Electron ID is found to be
 a subleading systematic that is unconstrained by the fit, making it an appropriate choice for this

997 analysis.

998 The sources which contribute to the uncertainty in the jet energy scale (JES) [jes] are
999 decomposed into uncorrelated components and treated as independent sources in the analysis.

1000 The CategoryReduction model is used to account for JES uncertainties, which decomposes the
1001 uncertainties into 30 nuisance parameters included in the fit. The SimpleJER model is used
1002 to account for jet energy resolution (JER) uncertainties, and 8 JER uncertainty components
1003 uncluded as NPs in the fit.

1004 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
1005 [btag_cal] are also decomposed into uncorrelated components. The large number of components
1006 for b-tagging is due to the calibration of the distribution of the MVA discriminant.

1007 The full list of systematic uncertainties considered in the analysis is summarized in Tables
1008 [18](#), [19](#) and [20](#).

1009

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 18: 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 19: 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 20: 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.

1010 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and
 1011 scale uncertainties are taken from theory calculations, with the exception of non-prompt and
 1012 diboson backgrounds. The cross-section uncertainty on tZ is taken from [[tZ_paper](#)]. Derivation
 1013 of the non-prompt background uncertainties, Z+jets and $t\bar{t}$, are explained in detail in Section
 1014 [13.3](#).

1015 The other VV + heavy flavor processes (namely VV+b and VV+charm, which primarily
 1016 consist of ZZ events) are also poorly understood, because these processes involve the same
 1017 physics as WZ + heavy flavor, and have also not been measured. Therefore, a conservative 50%
 1018 uncertainty is applied to those samples. While this uncertainty is large, it is found to have little
 1019 impact on the significance of the final result.

1020 The theory uncertainties applied to the predominate background estimates are summarized
 1021 in Table 21.

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	± 20
Z + jets	± 25
Others	± 50

Table 21: Summary of theoretical uncertainties for MC predictions in the analysis.

1022 Additional signal uncertainties are estimated by comparing estimates from the nominal
 1023 Sherpa WZ samples with alternate WZ samples generated with Powheg+PYTHIA8 (DSID
 1024 361601). MC/MC scale factors are applied to make these comparisons. The shape of the
 1025 templates used in the fit are compared between these two samples for WZ + b, WZ + charm and
 1026 WZ + light, as shown in Figures 15.1 and 15.2. Each of these plots are normalized to unity in
 1027 order to capture differences in shape.

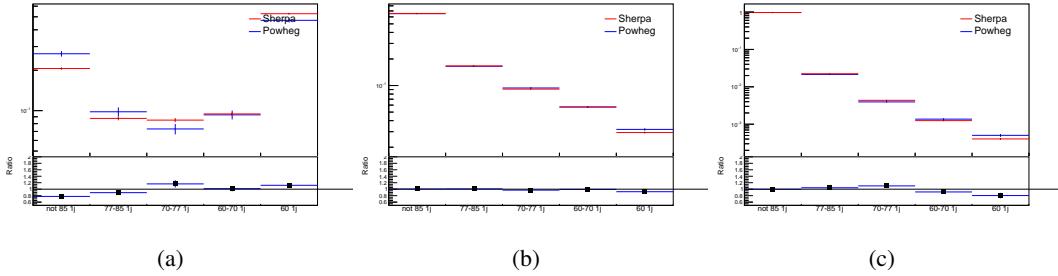


Figure 15.1: Comparison between Sherpa and Powheg predictions of the distribution of (a) WZ + b, (b) WZ + charm, and (c) WZ + light among the various b-tag WPs used in the 1-jet fit.

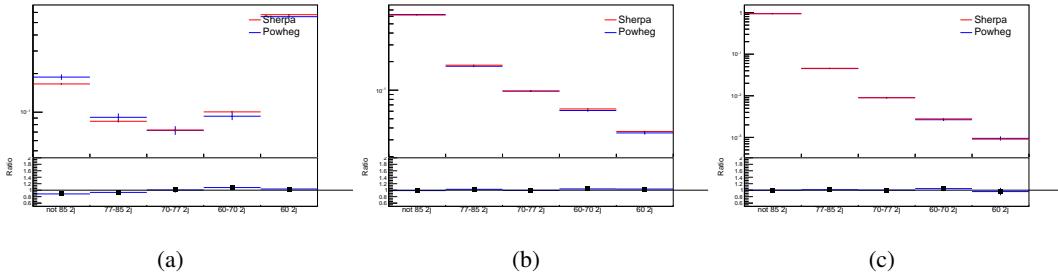


Figure 15.2: Comparison between Sherpa and Powheg predictions of the distribution of (a) WZ + b, (b) WZ + charm, and (c) WZ + light among the various b-tag WPs used in the 2-jet fit.

1028 Separate systematics are included in the fit for WZ + b, WZ + charm and WZ + light,
 1029 where the distribution among each of the fit regions is varied based on the prediction of the
 1030 Powheg sample.

1031 A similar approach is taken to account for uncertainties in migrations between the number
 1032 of reco and truth jets. The fraction of events with 1 truth jet which fall into the 1 jet bin versus
 1033 the 2 jet bin at reco level is compared for Sherpa and Powheg. The same is done for events with
 1034 2 truth jets. This comparison is shown in figure 15.3.

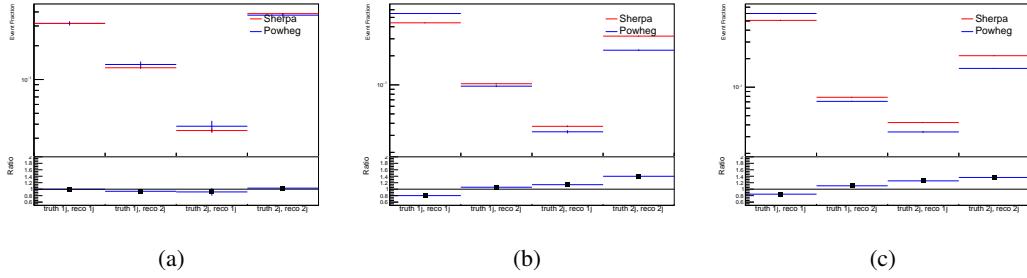


Figure 15.3: Comparison between Sherpa and Powheg predictions for truth jet migrations between the 1 and 2 jet reco bins for (a) WZ + b, (b) WZ + charm, and (c) WZ + light

1035 A systematic is included where events are shifted between the 1-jet and 2-jet regions based
 1036 on the differences between these two shapes. This is done independently for each of the WZ +
 1037 b, WZ + charm, and WZ + light templates.

1038 Additional systematics are included to account for the uncertainty in the contamination of
 1039 0 jet and 3 or more jet events (at truth level) in the 1 and 2 reco jet bins. Because these events
 1040 fall outside the scope of this measurement, these events are included as a background. As such,
 1041 a normalization, rather than a shape, uncertainty is applied for this background.

1042 The number of WZ events with 0-jets and ≥ 3 -jets in the reconstructed 1-jet and 2-jet
 1043 regions are compared for Sherpa and Powheg, as seen in figure 15.4. These differences are taken
 1044 as separate normalization systematics on the yield of WZ+0-jet and WZ+ ≥ 3 -jet events.

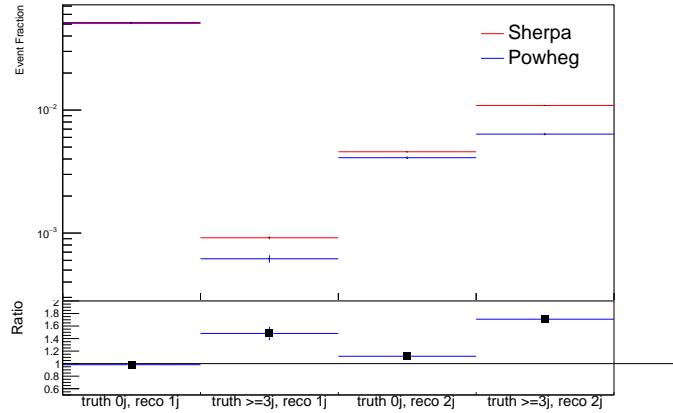


Figure 15.4: Comparison between Sherpa and Powheg predictions for 0 and ≥ 3 truth jet contributions in the 1 and 2 jet reco bins

16 Results

A separate maximum-likelihood fit is performed over the 1-jet and 2-jet fit regions in order to extract the best-fit value of the WZ + b-jet and WZ + charm jet contributions. The WZ + b, WZ + charm and WZ + light contributions are allowed to float, with the remaining background contributions are held fixed. **The current fit strategy treats the WZ + b-jet contribution as the parameter of interest, with the normalization of the WZ + charm and the WZ + light contributions taken as systematic uncertainties. This could however be adjusted, depending on whether it is decided the goal of the analysis should be to measure WZ+b specifically or WZ + heavy flavor overall.** The result of the fit is used to extract the cross-section of WZ + heavy-flavor production.

A maximum likelihood fit to data is performed simultaneously in the regions described

1056 in Section 13. The parameters μ_{WZ+b} , $\mu_{WZ+charm}$, $\mu_{WZ+light}$, where $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$, are
 1057 extracted from the fit.

1058 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
 1059 $WZ + b$. The fitted cross-section modifiers for $WZ + \text{charm}$ and $WZ + \text{light}$ are $1.00 \pm 0.17 \pm 0.15$
 1060 and $1.00 \pm 0.04 \pm 0.07$, respectively.

1061 The expected cross-section of $WZ+b$ with 1 associated jet obtained from the fit is
 1062 $1.74^{+0.65}_{-0.61}(\text{stat})^{+0.42}_{-0.37}(\text{sys}) \text{ fb}$ with an expected significance of 2.2σ . The expected cross-section
 1063 of $WZ + \text{charm}$ is measured to be $14.6 \pm 2.5(\text{stat}) \pm 2.1(\text{sys}) \text{ fb}$, with a correlation of -0.23.

1064 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 $WZ +$
 1065 b . The fitted cross-section modifiers for $WZ + \text{charm}$ and $WZ + \text{light}$ are $1.00 \pm 0.17 \pm 0.15$ and
 1066 $1.00 \pm 0.04 \pm 0.08$, respectively.

1067 The expected $WZ + b$ cross-section in the 2-jet region is $2.46^{+0.83}_{-0.81}(\text{stat})^{+0.073}_{-0.68}(\text{sys}) \text{ fb}$
 1068 with an expected significance of 2.6σ . The 2-jet expected cross-section of $WZ + \text{charm}$ is
 1069 $12.7 \pm 2.2(\text{stat}) \pm 2.0(\text{sys}) \text{ fb}$, and the correlation between $WZ + \text{charm}$ and $WZ + b$ is -0.26.

1070 16.1 1-jet Fit Results

1071 **The results of the fit are currently blinded.**

¹⁰⁷² The pre-fit yields in each of the regions used in the fit are shown in Table 16.1, and

¹⁰⁷³ summarized in Figure 16.1.

¹⁰⁷⁴

Sample	1j, <85% WP	1j, 77%-85% WP	1j, 70%-77% WP	1j, 60%-70% WP	1j, >60% WP	tZ CR
WZ + b - 1j	8.1 ± 1.6	4.7 ± 0.5	4.6 ± 0.4	5.1 ± 0.4	18.2 ± 2.4	4.8 ± 0.6
WZ + c - 1j	260 ± 22	81 ± 6	43.1 ± 3.6	25.8 ± 2.6	9.4 ± 1.8	2.9 ± 0.6
WZ + l - 1j	3090 ± 250	91 ± 13	17 ± 3	4.9 ± 1.6	1.3 ± 0.4	0.2 ± 0.1
WZ + b - 2j	1.10 ± 0.37	0.44 ± 0.11	0.39 ± 0.06	0.62 ± 0.14	2.1 ± 0.5	0.59 ± 0.14
WZ + c - 2j	21 ± 5	5.6 ± 1.2	3.0 ± 0.7	2.0 ± 0.5	0.70 ± 0.20	0.30 ± 0.08
WZ + l - 2j	250 ± 60	5.7 ± 1.6	0.73 ± 0.53	0.31 ± 0.15	0.07 ± 0.06	0.01 ± 0.01
WZ - Other	13 ± 5	1.4 ± 0.4	0.42 ± 0.08	0.2 ± 0.01	0.30 ± 0.05	0.67 ± 0.15
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
t̄tW	1.1 ± 0.2	0.2 ± 0.1	0.3 ± 0.1	0.4 ± 0.1	1.5 ± 0.3	0.7 ± 0.2
t̄tZ	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 + jets	169 ± 38	8.9 ± 1.9	3.7 ± 0.8	3.3 ± 0.7	3.2 ± 0.7	0.8 ± 0.17
V + γ	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
t̄t	12.1 ± 1.6	2.9 ± 0.6	2.5 ± 0.5	2.8 ± 0.5	11.2 ± 1.4	10.9 ± 1.5
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	5010 ± 260	227 ± 24	88 ± 12	57 ± 8	76 ± 16	53 ± 8

Table 22: Pre-fit yields in each of the 1-jet fit regions.

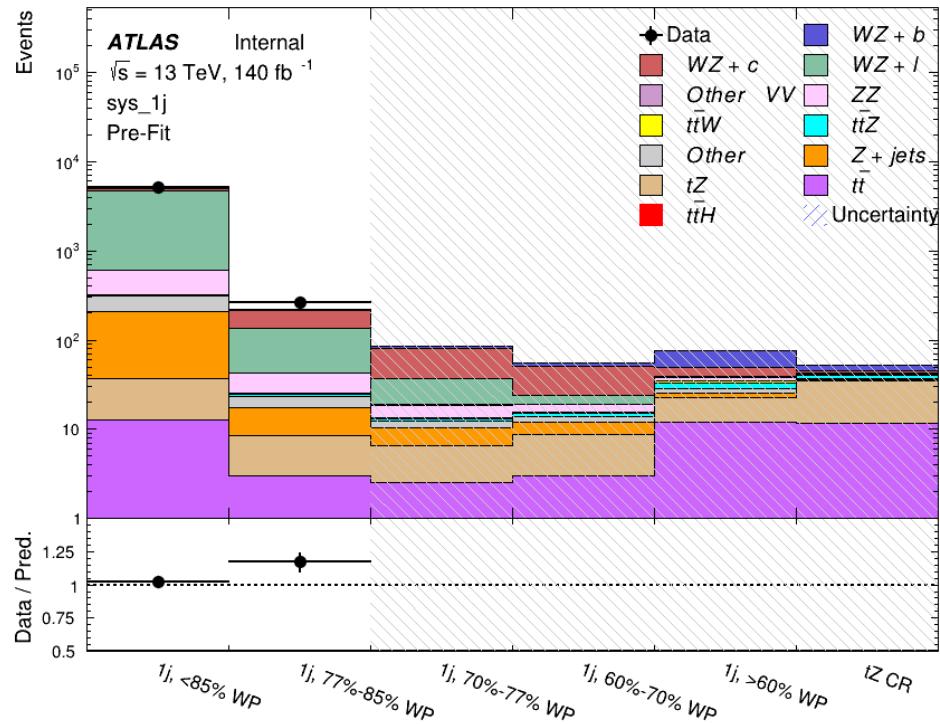


Figure 16.1: Pre-fit summary of the 1-jet fit regions.

1075

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

1076

	1j, <85% WP	1j, 77%-85% WP	1j, 70%-77% WP	1j, 60%-70% WP	1j, >60% WP	tZ CR
WZ + b	8.1 ± 4.9	4.7 ± 2.0	4.6 ± 2.0	5.1 ± 2.1	18 ± 10	5.0 ± 2.5
WZ + c	260 ± 60	80 ± 14	43 ± 7	26 ± 5	9.4 ± 2.3	2.9 ± 0.7
WZ + l	3090 ± 130	90 ± 11	17.3 ± 2.8	4.9 ± 1.6	1.3 ± 0.4	0.23 ± 0.1
WZ + b - 2j	1.10 ± 0.37	0.44 ± 0.11	0.39 ± 0.06	0.62 ± 0.14	2.1 ± 0.5	0.59 ± 0.1
WZ + c - 2j	21 ± 5	5.6 ± 1.2	3.0 ± 0.7	2.0 ± 0.5	0.70 ± 0.20	0.30 ± 0.0
WZ + l - 2j	250 ± 60	5.7 ± 1.6	0.73 ± 0.53	0.31 ± 0.15	0.07 ± 0.06	0.01 ± 0.0
WZ - Other	13 ± 5	1.4 ± 0.4	0.42 ± 0.08	0.2 ± 0.01	0.30 ± 0.05	0.67 ± 0.1
Other VV	6.2 ± 0.6	0.92 ± 0.07	0.02 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.0
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 ± 0.0
t̄tWW	0.04 ± 0.03	0.01 ± 0.02	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.02	0.01 ± 0.0
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	0.01 ± 0.0				
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 ± 0.4
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 ± 0.0
VH	40 ± 6	2.6 ± 1.4	0.9 ± 0.8	0.7 ± 0.8	0.4 ± 0.6	0.01 ± 0.0
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 ± 0.0
Total	5100 ± 110	227 ± 12	87 ± 6	56.7 ± 4.4	76 ± 9	52.5 ± 4.2

Table 23: Post-fit yields in each of the 1-jet fit regions.

1077

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

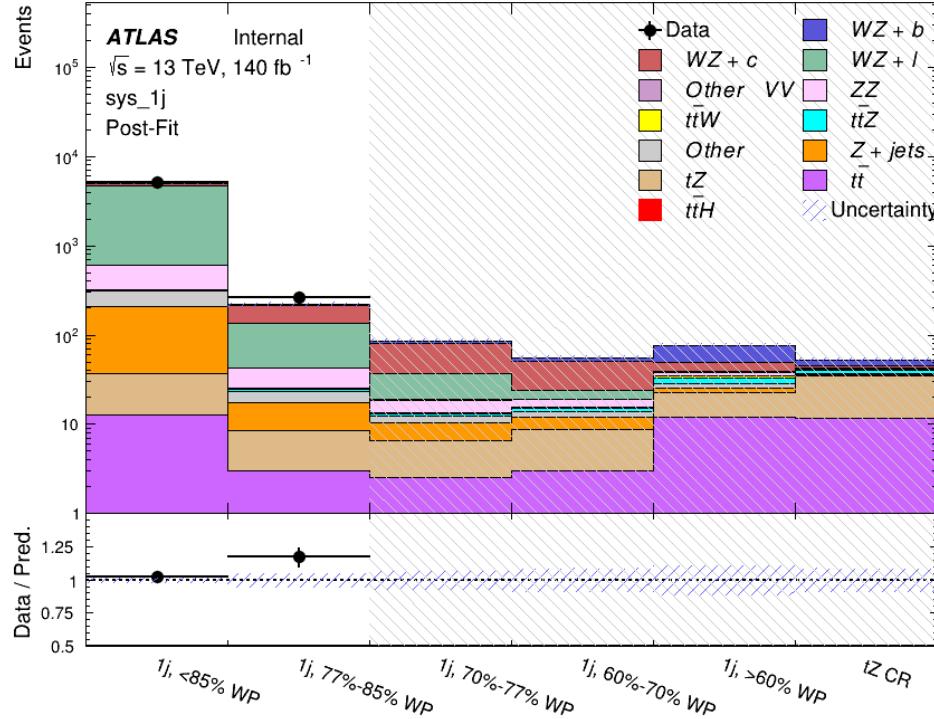


Figure 16.2: Post-fit summary of the 1-jet fit regions.

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

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

¹⁰⁸⁶ fit. The impact of the most significant sources of systematic uncertainties is summarized in Table

¹⁰⁸⁷ [24.](#)

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

Table 24: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

¹⁰⁸⁸ The ranking and impact of those nuisance parameters with the largest contribution to the

¹⁰⁸⁹ overall uncertainty is shown in Figure [16.3](#).

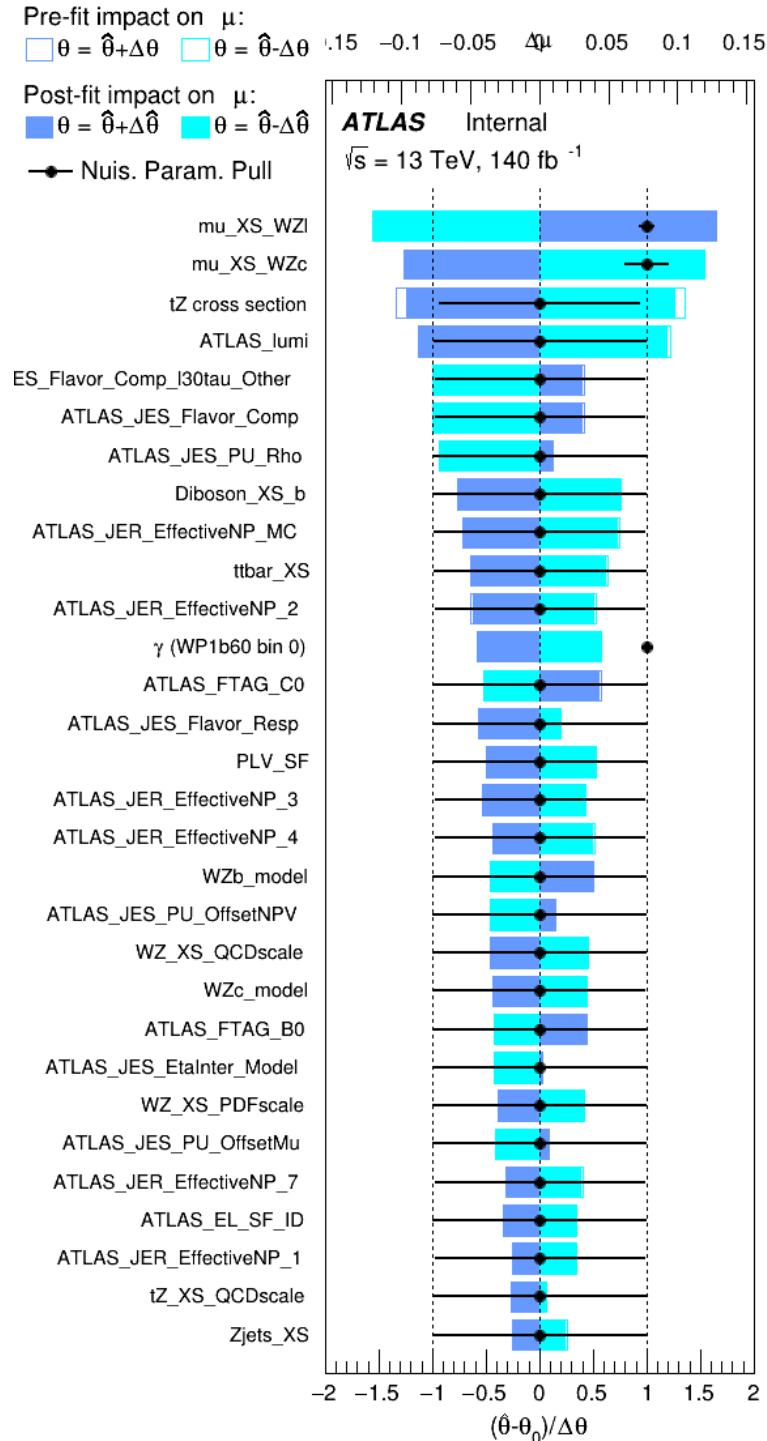


Figure 16.3: Impact of systematic uncertainties on the signal-strength of $WZ + b$ for events with exactly one jet

1090 The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the
 1091 shape of the fit regions depends heavily on the modeling of the jets. The other major sources
 1092 of uncertainty come from background modelling and cross-section uncertainty. The pie charts
 1093 in Figure 16.4 show that for the modelling uncertainties that contribute most correspond to the
 1094 most significant backgrounds.

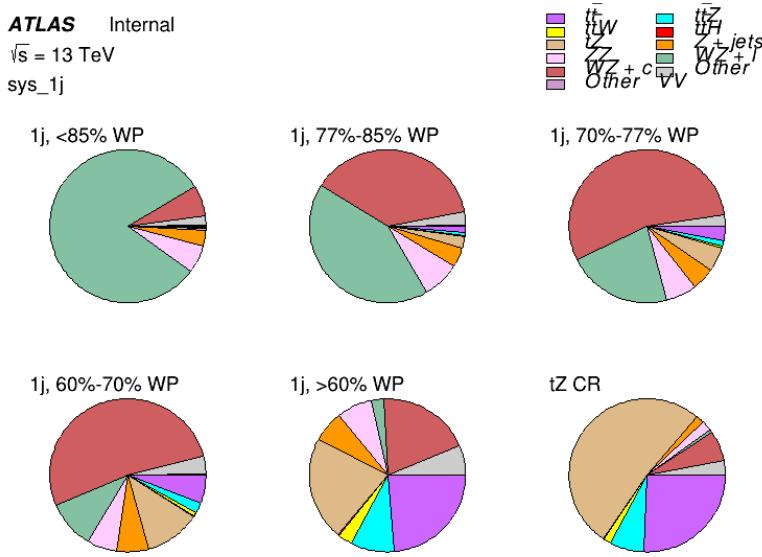


Figure 16.4: Post-fit background composition of the fit regions.

1095 The correlations between these nuisance parameters are summarized in Figure 16.5.

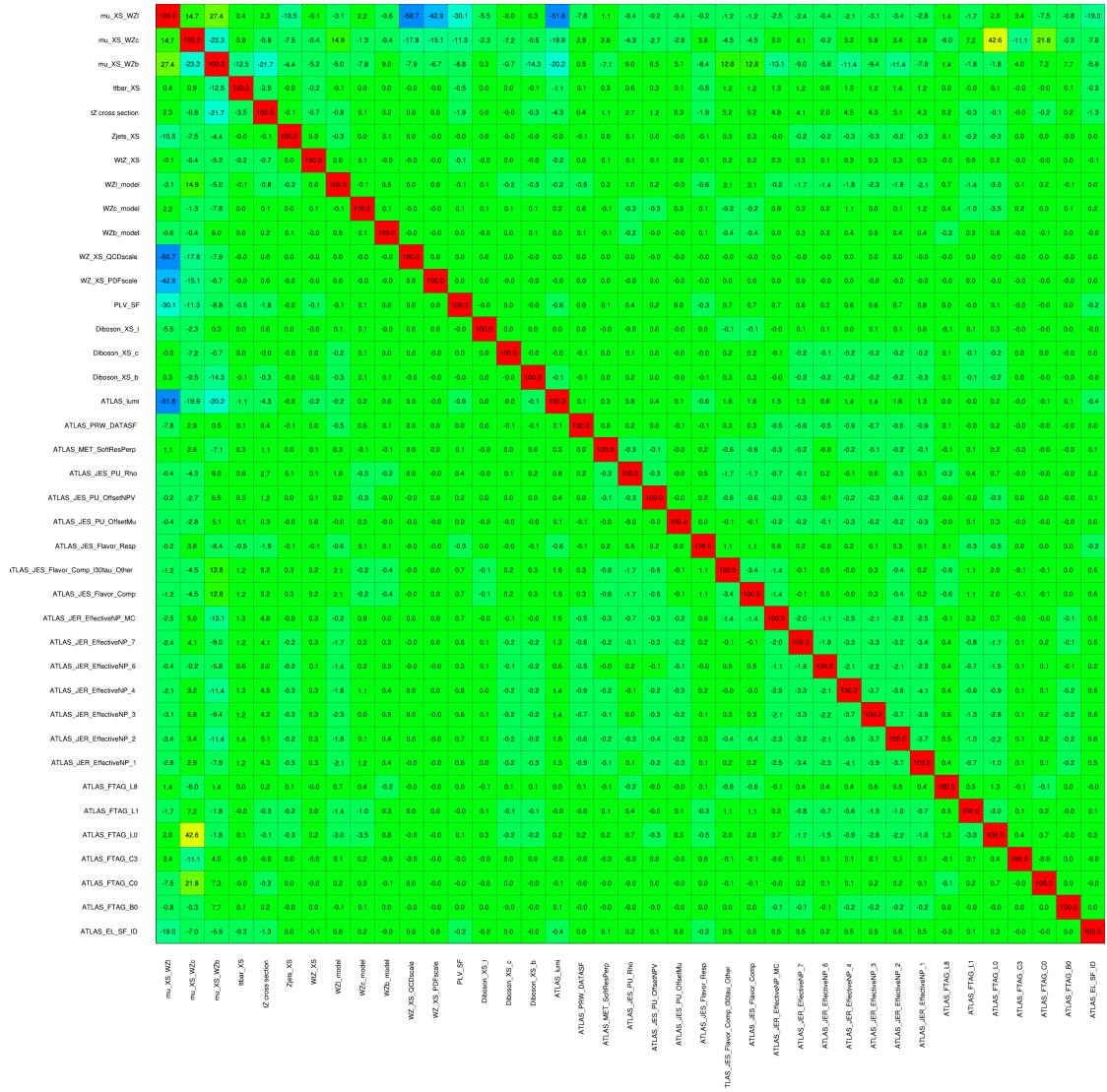


Figure 16.5: Correlations between nuisance parameters

1096 The negative correlations between $\mu_{WZ+charm}$ and μ_{WZ+b} and $\mu_{WZ+light}$ are expected:
1097 WZ + charm is present in both the WZ + b and WZ + light enriched regions, therefore increasing
1098 the fraction of charm requires increasing the fraction of WZ + b and WZ + light. This reasoning
1099 also explains the positive correlation between μ_{WZ+b} and $\mu_{WZ+light}$.

1100 Two of the major backgrounds in the region with the highest purity of $WZ + b$ are tZ and
1101 Other $VV + b$, explaining the negative correlations between μ_{WZ+b} and the tZ cross section, and
1102 the $VV + b$ cross section.

1103 The high correlation between the luminosity and $\mu_{WZ+light}$ arises from the fact that the
1104 uncertainty on $\mu_{WZ+light}$ is very low (around 4%). Small changes in luminosity cause a change
1105 in the yield of $WZ + light$ that is large compared to its uncertainty, producing a large correlation
1106 between these two parameters.

1107 **16.2 2-jet Fit Results**

1108 **The results of the fit are currently blinded.**

1109 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 - 2j	3.1 ± 1.6	6.7 ± 0.5	5.6 ± 0.4	8.0 ± 0.6	31 ± 2	14 ± 1
WZ + c - 2j	180 ± 20	54 ± 6	41 ± 3	24 ± 3	11 ± 2	4.8 ± 0.6
WZ + l - 2j	1250 ± 150	90 ± 14	18 ± 3	5.8 ± 1.4	1.4 ± 0.4	0.25 ± 0.1
WZ + b - 1j	3.4 ± 0.6	1.52 ± 0.35	1.58 ± 0.23	1.95 ± 0.39	6.7 ± 1.1	1.9 ± 0.6
WZ + c - 1j	56 ± 14	17.6 ± 4.0	8.6 ± 2.2	6.3 ± 1.8	3.0 ± 0.9	0.7 ± 0.2
WZ + l - 1j	427 ± 120	24 ± 7	4.7 ± 2.3	1.6 ± 0.7	0.3 ± 0.2	0.01 ± 0.0
WZ - Other	129 ± 29	6.1 ± 4.6	1.2 ± 0.3	0.3 ± 0.2	2.9 ± 0.5	3.6 ± 0.6
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 ± 0.0
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 ± 0.0
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 ± 0.0
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
t̄t	6.8 ± 1.0	2.4 ± 0.5	1.8 ± 0.4	3.3 ± 0.6	8.4 ± 1.2	13.6 ± 1.1
t̄tH	0.4 ± 0.1	0.2 ± 0.1	0.16 ± 0.02	0.23 ± 0.03	0.94 ± 0.11	1.03 ± 0.1
Total	2580 ± 160	229 ± 24	89 ± 13	69 ± 11	120 ± 15	108 ± 11

Table 25: Pre-fit yields in each of the 2-jet fit regions.

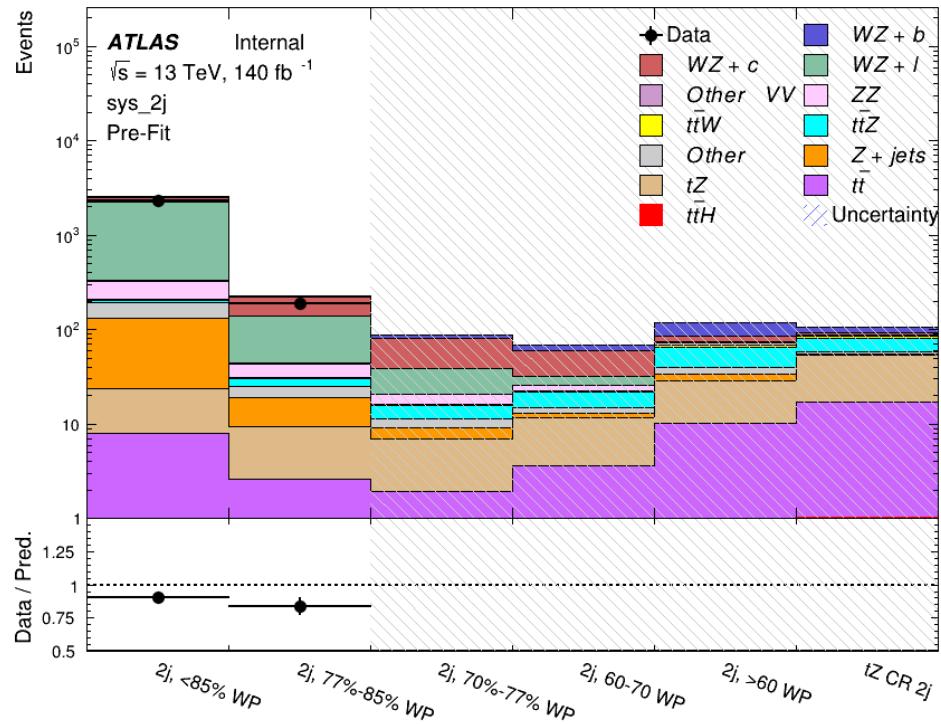


Figure 16.6: Pre-fit summary of the 2-jet fit regions.

1110

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
WZ + b - 1j	3.4 ± 0.6	1.52 ± 0.35	1.58 ± 0.23	1.95 ± 0.39	6.7 ± 1.1	1.9 ± 0.6
WZ + c - 1j	56 ± 14	17.6 ± 4.0	8.6 ± 2.2	6.3 ± 1.8	3.0 ± 0.9	0.7 ± 0.2
WZ + l - 1j	427 ± 120	24 ± 7	4.7 ± 2.3	1.6 ± 0.7	0.3 ± 0.2	0.01 ± 0.0
WZ - Other	129 ± 29	6.1 ± 4.6	1.2 ± 0.3	0.3 ± 0.2	2.9 ± 0.5	3.6 ± 0.6
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 ± 3.0
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 ± 0.0
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	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 ± 0.0
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 ± 0.0
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 ± 0.1
Total	2580 ± 60	229 ± 11	89 ± 6	69.1 ± 4.1	120 ± 10	108 ± 6

Table 26: Post-fit yields in each of the 2-jet fit regions.

1111

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

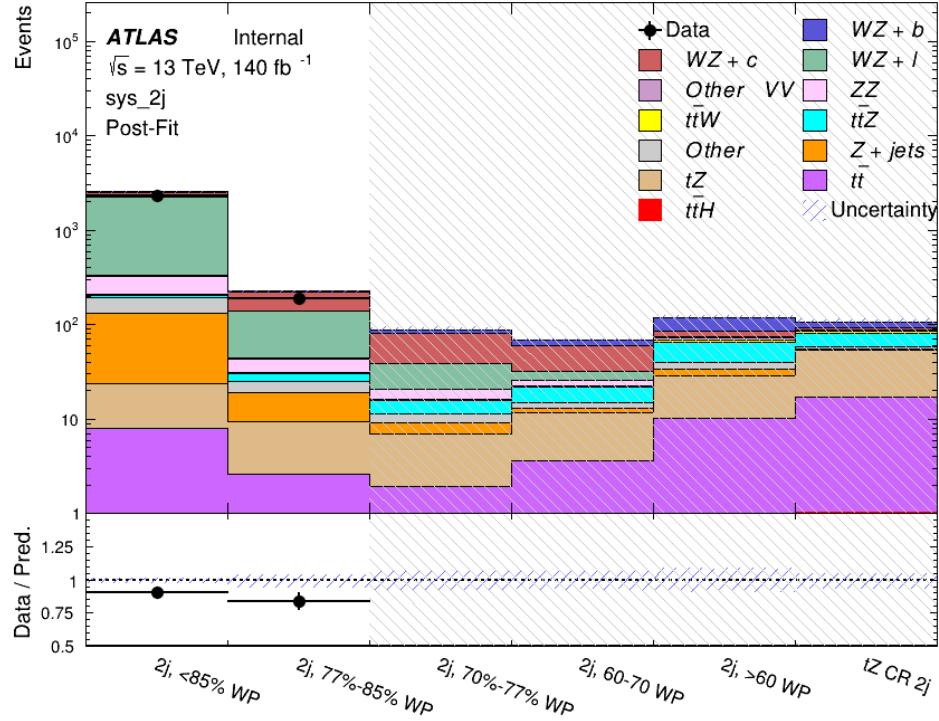


Figure 16.7: Post-fit summary of the fit over 2-jet regions.

1112 The same set of systematic uncertainties consider for the 1-jet fit are included in the 2-jet
 1113 fit as well. The impact of the most significant systematic uncertainties is summarized in Table
 1114 [27](#).

Uncertainty Source	$\Delta\mu$	
WZ + c cross-section	-0.1	0.14
Luminosity	-0.11	0.12
tZ cross-section	-0.11	0.11
Jet Energy Scale	-0.11	0.11
ttZ cross-section - QCD scale	-0.08	0.09
WZ + l cross-section	0.08	-0.07
WtZ cross-section	-0.07	0.07
Flavor tagging	0.05	0.05
Other VV + b cross-section	-0.05	0.05
Jet Energy Resolution	-0.04	0.04
Total	0.29	0.31

Table 27: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b 2-jet events.

1115 The ranking and impact of those nuisance parameters with the largest contribution to the
1116 overall uncertainty is shown in Figure 16.8.

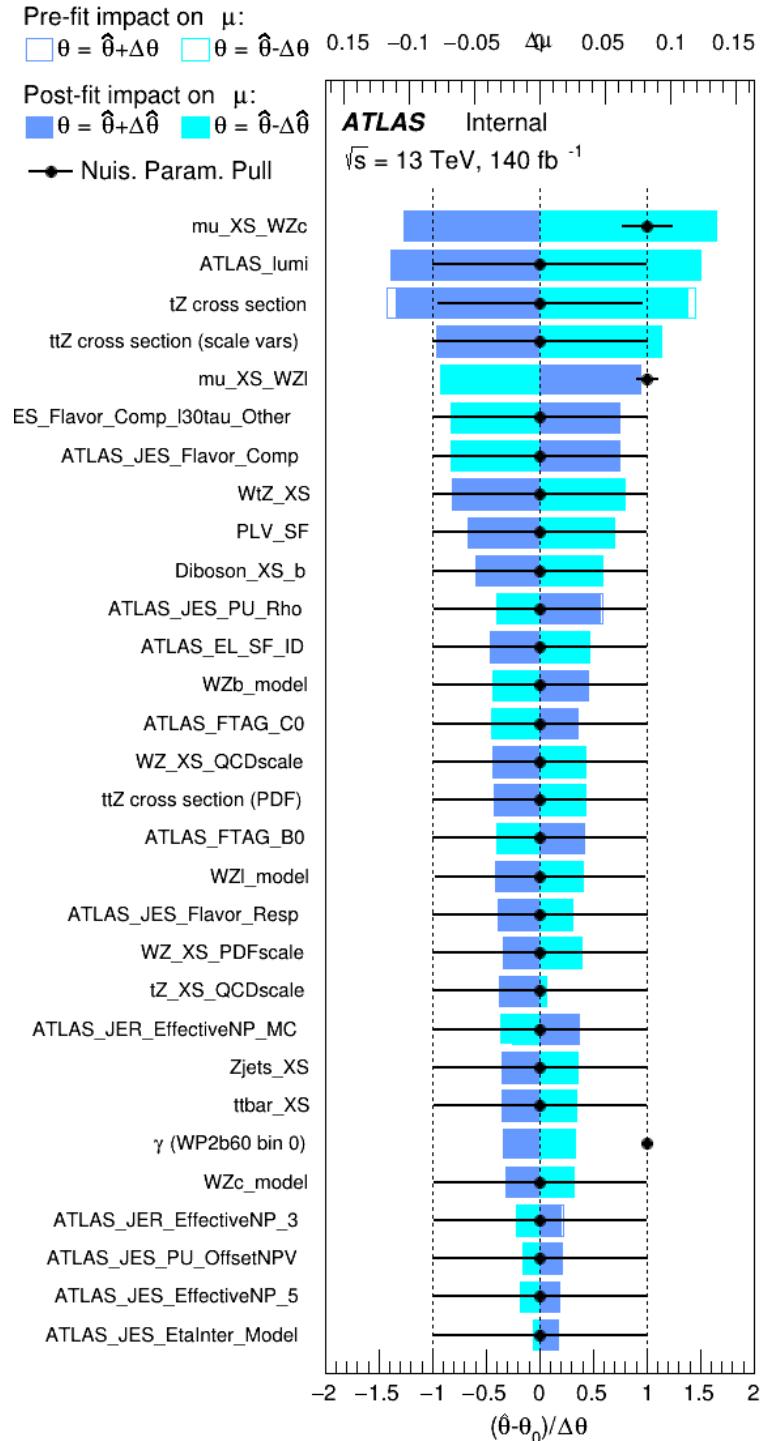


Figure 16.8: Impact of systematic uncertainties on the signal-strength of $WZ + b$ in 2-jet events.

1117 The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the
 1118 shape of the fit regions depends heavily on the modeling of the jets. The other major sources
 1119 of uncertainty come from background modelling and cross-section uncertainty. The pie charts
 1120 in Figure 16.9 show that for the modelling uncertainties that contribute most correspond to the
 1121 most significant backgrounds.

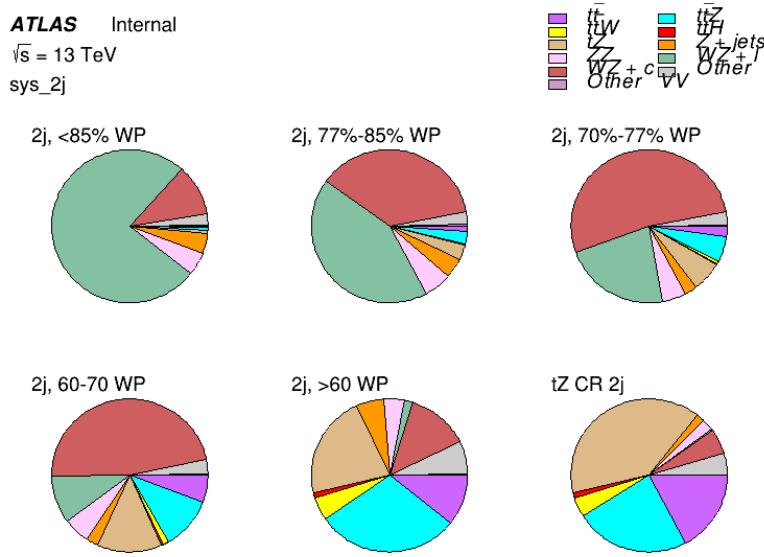


Figure 16.9: Post-fit background composition of the 2-jet fit regions.

1122 The correlations between these nuisance parameters are summarized in Figure 16.10.



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

1123

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

1124 parameters.

Part V**1126 Differential Studies of $t\bar{t}H$ Multilepton****1127 17 Data and Monte Carlo Samples**

1128 For both data and Monte Carlo (MC) simulations, samples were prepared in the `xAOD` format,
1129 which was used to produce a `xAOD` based on the `HIGG8D1` derivation framework. This framework
1130 was designed for the main $t\bar{t}H$ multi-lepton analysis. Because this analysis targets events with
1131 multiple light leptons, as well as tau hadrons, this framework skims the dataset of any events that
1132 do not meet at least one of the following requirements:

- 1133 • at least two light leptons within a range $|\eta| < 2.6$, with leading lepton $p_T > 15$ GeV and
1134 subleading lepton $p_T > 5$ GeV
- 1135 • at least one light lepton with $p_T > 15$ GeV within a range $|\eta| < 2.6$, and at least two hadronic
1136 taus with $p_T > 15$ GeV.

1137 Samples were then generated from these `HIGG8D1` derivations using a modified version of
1138 `AnalysisBase` version 21.2.127.

1139 17.1 Data Samples

1140 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through
1141 2018, which represents an integrated luminosity of 139 fb^{-1} and an energy of $\sqrt{s} = 13 \text{ TeV}$. All
1142 data used in this analysis was included in one of the following Good Run Lists:

1143 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02

1144 _PHYS_StandardGRL_All_Good_25ns.xml

1145 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04

1146 _PHYS_StandardGRL_All_Good_25ns.xml

1147 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL

1148 _All_Good_25ns_Triggerno17e33prim.xml

1149 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL

1150 _All_Good_25ns_Triggerno17e33prim.xml

1151 17.2 Monte Carlo Samples

1152 Several Monte Carlo (MC) generators were used to simulate both signal and background pro-
1153 cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific
1154 event generator used for each of these MC samples is listed in table 28.

Table 28: 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

¹¹⁵⁵ 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

¹¹⁵⁹ Events are required to be selected by dilepton triggers, as summarized in table 29.

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

1160 18.2 Light Leptons

1161 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that
 1162 are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].
 1163 Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the
 1164 transition region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| <$
 1165 1.52, are rejected. A multivariate likelihood discriminant combining shower shape and track
 1166 information is used to distinguish prompt electrons from nonprompt leptons, such as those
 1167 originating from hadronic showers.

1168 To further reduce the non-prompt contribution, the track of each electron is required to

1169 originate from the primary vertex; requirements are imposed on the transverse impact parameter
1170 significance ($|d_0|/\sigma_{d_0}$) and the longitudinal impact parameter ($|\Delta z_0 \sin \theta_\ell|$).

1171 Muon candidates are reconstructed by combining inner detector tracks with track segments
1172 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
1173 $p_T > 10$ GeV and $|\eta| < 2.5$. All leptons are required to be isolated, and pass a non-prompt BDT
1174 selection described in detail in [**ttH_paper**].

1175 **18.3 Jets**

1176 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
1177 calorimeters [**ATL-PHYS-PUB-2015-015**], using the anti- k_t algorithm with a radius parameter
1178 $R = 0.4$. Jets with energy contributions likely arising from noise or detector effects are removed
1179 from consideration [**ATLAS-CONF-2015-029**], and only jets satisfying $p_T > 25$ GeV and
1180 $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
1181 association algorithm is used to confirm that the jet originates from the selected primary vertex,
1182 in order to reject jets arising from pileup collisions [**PERF-2014-03**].

1183 **18.4 Missing Transverse Energy**

1184 Because all $t\bar{t}H - ML$ channels considered include multiple neutrinos, missing transverse
1185 energy (E_T^{miss}) is present in each event. The missing transverse momentum vector is defined as
1186 the inverse of the sum of the transverse momenta of all reconstructed physics objects as well

1187 as remaining unclustered energy, the latter of which is estimated from low- p_T tracks associated
1188 with the primary vertex but not assigned to a hard object [**ATL-PHYS-PUB-2015-027**].

1189 **19 Higgs Momentum Reconstruction**

1190 Reconstructing the momentum of the Higgs boson is a particular challenge for channels with
1191 leptons in the final state: Because all channels include at least two neutrinos in the final state, the
1192 Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted
1193 by a neural network when provided with the four-vectors of the Higgs Boson decay products, as
1194 shown in section 19.1. With this in mind, several layers of MVAs are used to reconstruction the
1195 Higgs momentum.

1196 The first layer is a model designed to select which jets are most likely to be the b-jets that
1197 came from the top decay, detailed in section 19.2. As described in section 19.3, the kinematics
1198 of these jets are fed into the second layer, which is designed to identify the decay products of
1199 the Higgs Boson itself. The kinematics of these particles are then fed into yet another neural-
1200 network, which predicts the momentum of the Higgs (19.4). MVAs are also used in the analysis
1201 to determine the decay of the Higgs boson in the 3l channel (19.5).

1202 Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_AMC.
1203 Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training.

1204 For all of these models, the Keras neural network framework, with Tensorflow as the
1205 backend, is used, and the number of hidden layers and nodes are determined using grid search
1206 optimization. Each neural network uses the LeakyReLU activation function, a learning rate
1207 of 0.01, and the Adam optimization algorithm, as alternatives are found to either decrease or
1208 have no impact on performance. Batch normalization is applied after each layer. For the
1209 classification algorithms (b-jet matching, Higgs reconstruction, and 3l decay identification)
1210 binary-cross entropy is used as the loss function, while the p_T reconstruction algorithm uses
1211 MSE.

1212 The specific inputs features used for each model are arrived at through a process of trial
1213 and error - features considered potentially useful are tried, and those that are found to increase
1214 performance are included. While each model includes a relatively large number of features,
1215 some using upwards of 30, this inclusive approach is found to maximize the performance of each
1216 model while decreasing the variance compared to a reduced number of inputs. Each input feature
1217 is validated by comparing MC simulations to 80 fb^{-1} of data, as shown in the sections below.

1218 **19.1 Decay Candidate Reconstruction**

1219 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson
1220 using MC simulations of $t\bar{t}H$ events. These include light leptons and jets. Reconstructed
1221 physics objects are matched to truth level particles, in order to identify the parents of these

1222 reconstructed objects. The kinematics of the decay product candidates as well as event level
 1223 variables are used as inputs.

1224 Leptons considered as possible Higgs and top decay candidates are required to pass the
 1225 selection described in section 18.2. For jets, however, it is found that a large fraction that originate
 1226 from either the top decay or the Higgs decay fall outside the selection described in section 18.3.
 1227 Specifically, jets from the Higgs decay tend to be soft, with 32% having $p_T < 25$ GeV. Therefore
 1228 jets with $p_T < 15$ GeV are considered as possible candidates in the models described below. By
 1229 contrast, less than 5% of the jets originating from the Higgs fall below this p_T . The jets are found
 1230 to be well modeled even down to this low p_T threshold, as shown in section 21.1. The impact of
 1231 using different p_T selection for the jet candidates is considered in detail in section A.5. As they
 1232 are expected to originate from the primary vertex, jets are also required to pass a JVT cut.

1233 19.2 b-jet Identification

1234 Including the kinematics of the b-jets that originate from the top decay is found to improve the
 1235 identification of the Higgs decay products, and improve the accuracy with which the Higgs
 1236 momentum can be reconstructed. Because these b-jets are reconstructed by the detector with
 1237 high efficiency (just over 90% of the time), and can be identified relatively consistently, the first
 1238 step in reconstructing the Higgs is selecting the b-jets from the top decay.

1239 Exactly two b-jets are expected in the final state of $t\bar{t}H - ML$ events. However, in both
 1240 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 $t\bar{t}H - \text{ML}$ 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 30 are used for training. Here j_0 and j_1 are the two jet candidates, while l_0 and l_1 are the two leptons in the event, both ordered by p_T . jet DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet, where 5 represents the tightest working point and 1 represents the loosest. The variables $n\text{Jets DL1r 60\%}$ and $n\text{Jets DL1r 85\%}$ 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 30: Input features used in the b-jet identification algorithm for the 2lSS channel

1259 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for
 1260 training.

1261 The difference between the distributions for a few of these features for the correct(i.e.
 1262 both jets are truth b-jets), and incorrectcombinations are shown in figure 19.1. The correct and
 1263 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences
 1264 in the distributions.

1265 The modeling of these inputs is validated against data, with figure 19.2 showing good
 1266 general agreement between data and MC. Plots for the complete list of features can found in
 1267 section A.

1268 Based on the results of grid search evaluation, the optimal architecture is found to include
 1269 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as
 1270 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for

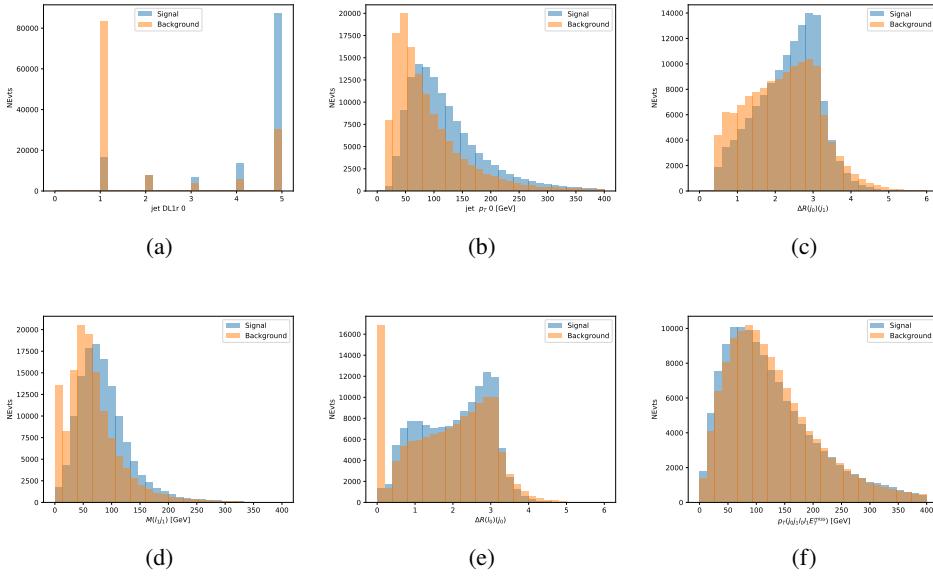


Figure 19.1: Input features for top2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

1271 the trained model are shown in figure 19.2.1. The model is found to identify the correct pairing
 1272 of jets for 73% of 2lSS signal events on test data.

1273 For point of comparison, a naïve approach to identify b-jets is used as well: The two jets
 1274 which pass the highest DL1r b-tag working point are assumed to be the b-jets from the top decay.
 1275 In the case that multiple jets meet the same b-tag working point, the jet with higher p_T is used.
 1276 This method identifies the correct jet pair 65% of the time.

1277 The accuracy of the model for different values of n-bjets, compared to this naive approach,
 1278 is shown in table 31.

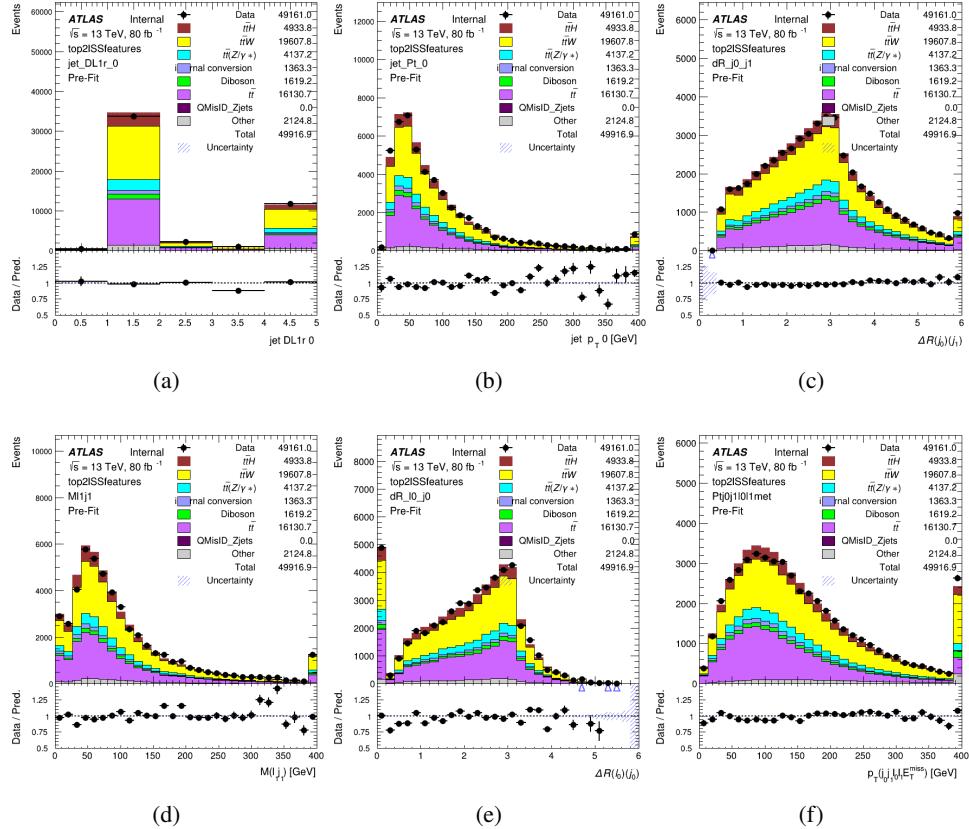


Figure 19.2: Data/MC comparisons of input features for top2ISS training for 80 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (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 31: 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.

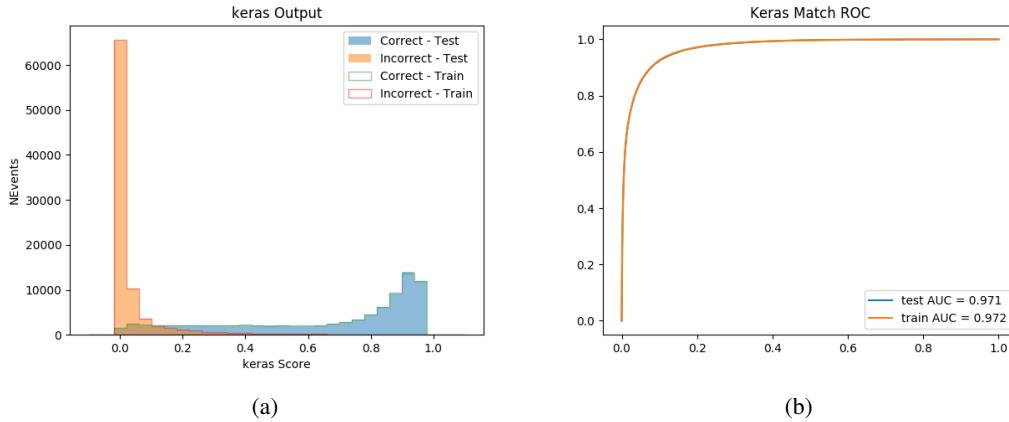


Figure 19.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1279 19.2.2 3l Channel

1280 The input features used in the 3l channel are listed in table 32, with the same naming convention
1281 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 32: Input features for the b-jet identification algorithm in the 3l channel.

1282 A few of these features are shown in figure 19.4, comparing the distributions for correct

1283 and incorrect combinations of jets.

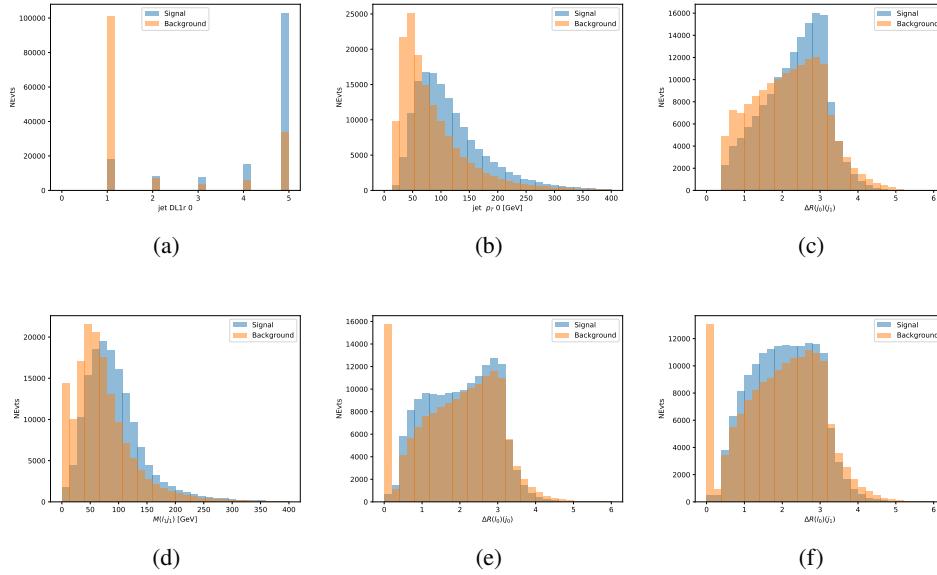


Figure 19.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and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

1284 The modeling of these inputs is validated against data, with figure 19.5 showing good
1285 general agreement between data and MC. Plots for the complete list of features can found in
1286 section A.

1287 Again, the dataset is downsized to reduce the ratio of correct and incorrect combination
1288 from 20:1, to 5:1. Around 7 million events are used for training, with 10% set aside for testing.
1289 Based on the results of grid search evaluation, the optimal architecture is found to include 5
1290 hidden layers with 60 nodes each. The output score distribution as well as the ROC curve for the
1291 trained model are shown in figure 19.2.2.

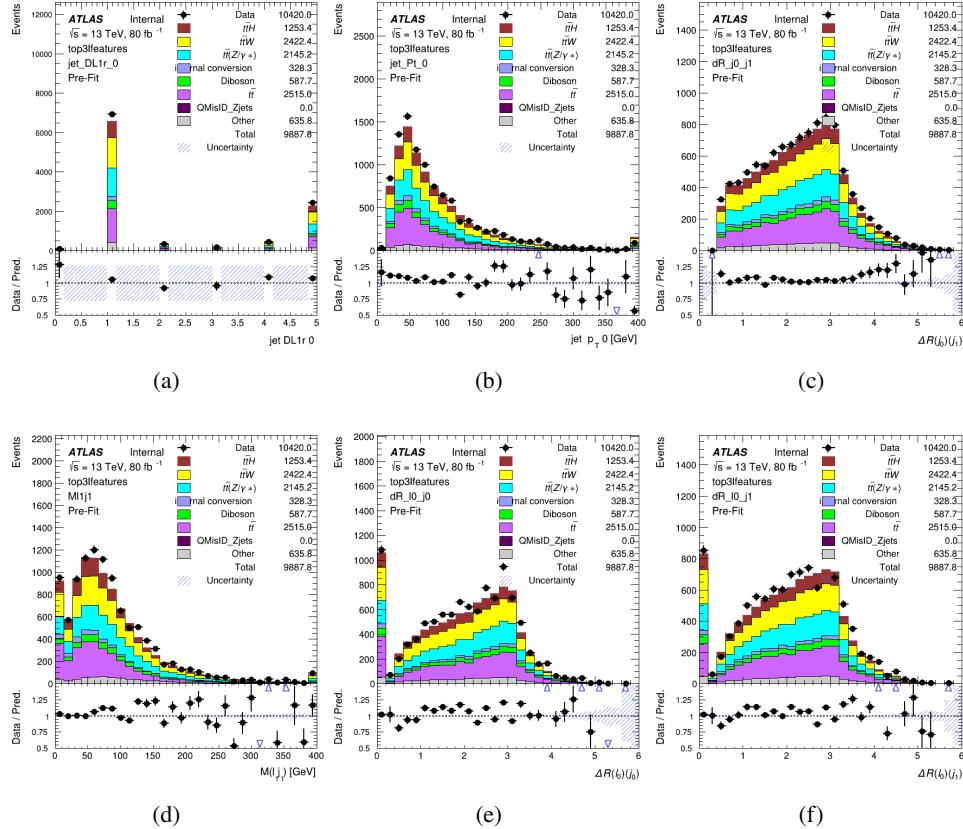


Figure 19.5: Data/MC comparisons of input features for top3l training for 80 fb^{-1} of data. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

1292 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal
 1293 events. The accuracy of the model is summarized in table 33.

Table 33: 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%

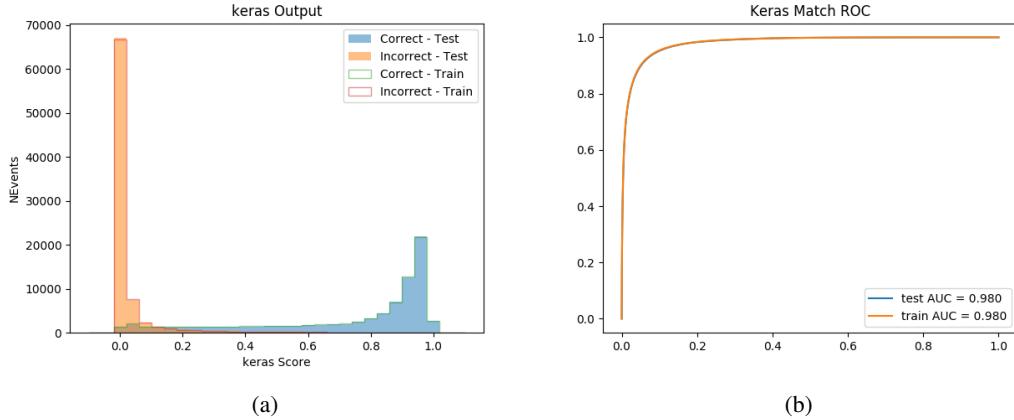


Figure 19.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

19.3 Higgs Reconstruction

1294 Techniques similar to the b-jet identification algorithms are employed to select the decay products
 1295 of the Higgs: kinematics of all possible combinations of reconstructed objects are fed into a neural
 1296 network to determine which of those is most mostly to be the decay products of the Higgs.

1298 Again separate models are used for the 2lSS and 3l channels, while the 3l channel has now
 1299 been split into two: $t\bar{t}H$ events with three leptons in the final state include both instances where
 1300 the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the Higgs
 1301 decays to two leptons.

1302 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In
 1303 the semi-leptonic case the reconstructed decay products consist of two jets and a single leptons.
 1304 For the fully-leptonic case, the decay products include 2 of the three leptons associated with the

1305 event. For training the models, events are separated into these two categories using truth level
1306 information. A separate MVA, described in section 19.5, is used to make this distinction at reco
1307 level and determine which model to use.

1308 For all channels, the models described in section 19.2 are used to identify b-jet candidates,
1309 whose kinematics are used to identify the Higgs decay products. These jets are not considered
1310 as possible candidates for the Higgs decay, justified by the fact that these models are found to
1311 misidentify jets from the Higgs decay as jets from the top decay less than 1% of the time.

1312 **19.3.1 2lSS Channel**

1313 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural
1314 network is trained on the kinematics of different combinations of leptons and jets, as well as the
1315 b-jets identified in section 19.2, with the specific input features listed in table 34.

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

1316 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 1317 lepton in the event is labeled l_T , as it is assumed to have come from the decay of one of the top
 1318 quarks. b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet
 1319 Reco Score is the output of the b-jet reconstruction algorithm.

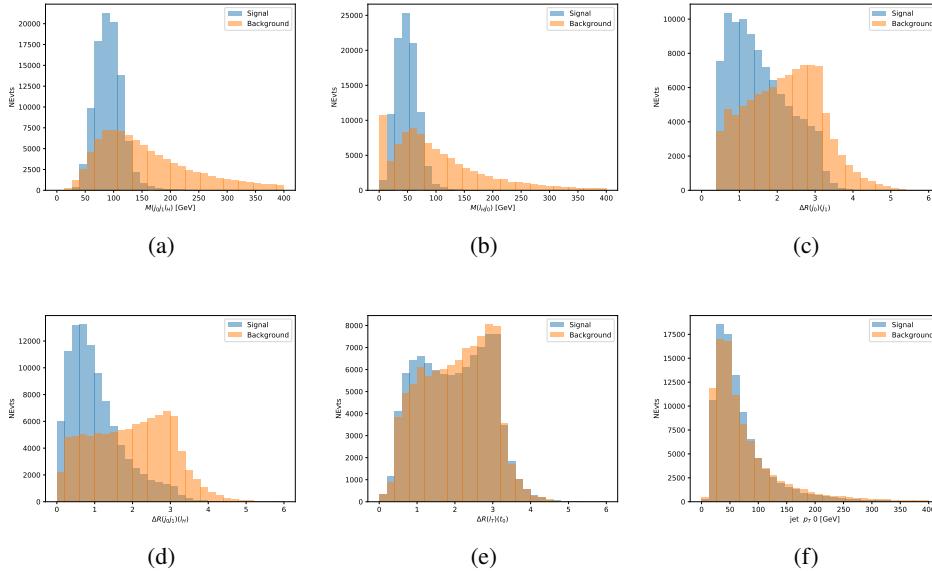


Figure 19.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate,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

1320 The modeling of these inputs is validated against data, with figure 19.2 showing good
 1321 general agreement between data and MC. Plots for the complete list of features can found in
 1322 section A.

1323 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2
 1324 million events, with an additional 200,000 reserved for testing the model. In order to compensate

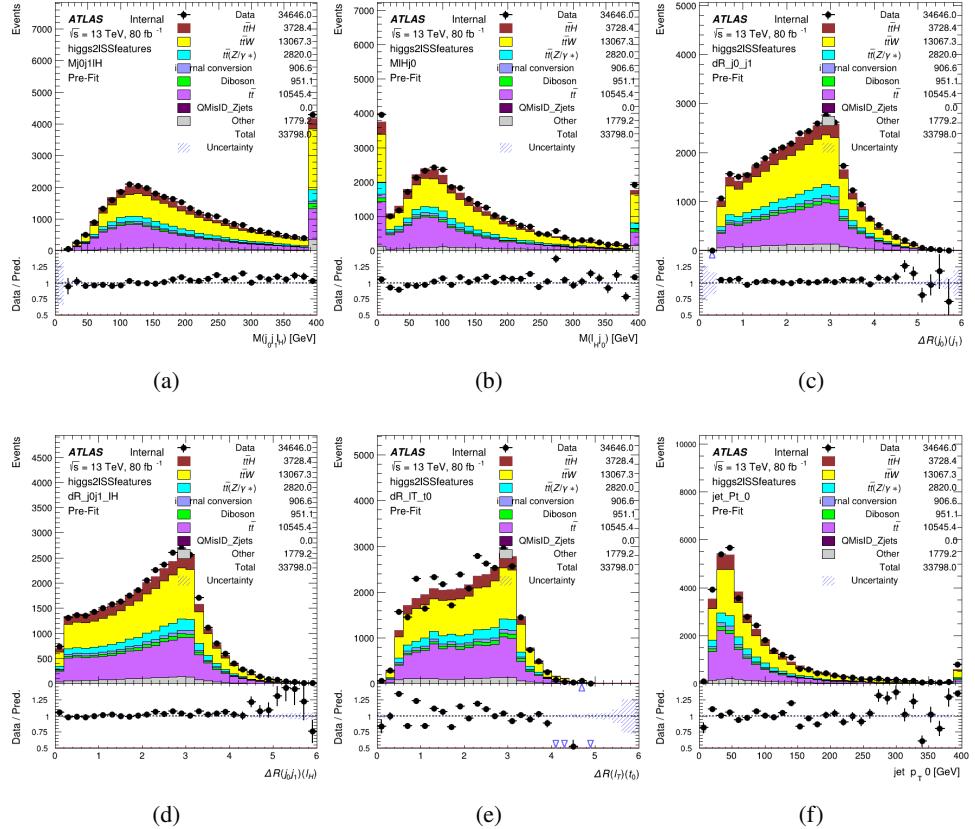


Figure 19.8: Data/MC comparisons of input features for higgs2lSS training for 80 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1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

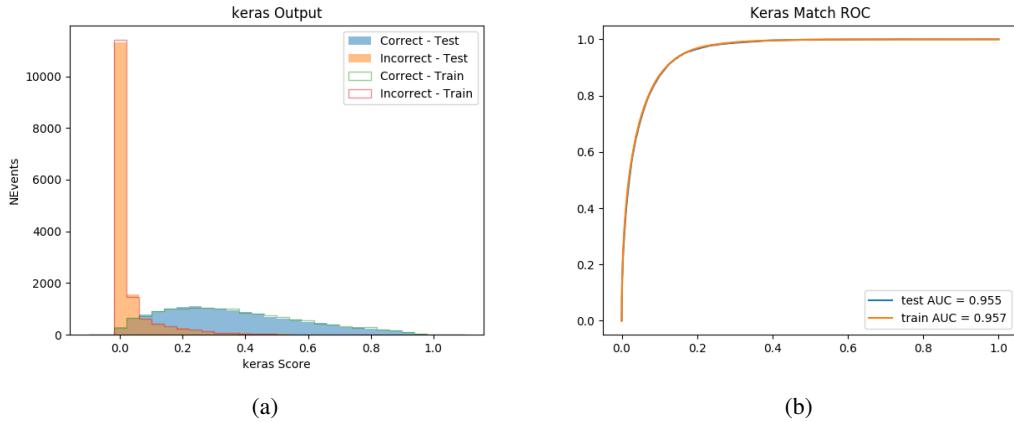


Figure 19.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

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 $\eta 0$	jet $\eta 1$
jet $\phi 0$	jet $\phi 1$	$\Delta R(j_0)(j_1)$
$M(j_0 j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0 j_1)(l_H)$	$\Delta R(j_0 j_1)(l_{T_1})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0 j_1 l_{T_0})$	$M(j_0 j_1 l_{T_1})$
$M(j_0 j_1 l_H)$	$\Delta R(j_0 j_1 l_H)(l_{T_0})$	$\Delta R(j_0 j_1 l_H)(l_{T_1})$
$\Delta\phi(j_0 j_1 l_H)(E_T^{\text{miss}})$	$p_T(j_0 j_1 l_H l_{T_0} l_{T_1} b_0 b_1 E_T^{\text{miss}})$	$M(j_0 j_1 b_0)$
$M(j_0 j_1 b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
topScore	MET	HT jets
nJets		

Table 35: Input features used to identify the Higgs decay products in 3lS events

1336 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 1337 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
 1338 the b-jet identification algorithm. The b-jet Reco Score is the output of the Higgs reconstruction
 1339 algorithm.

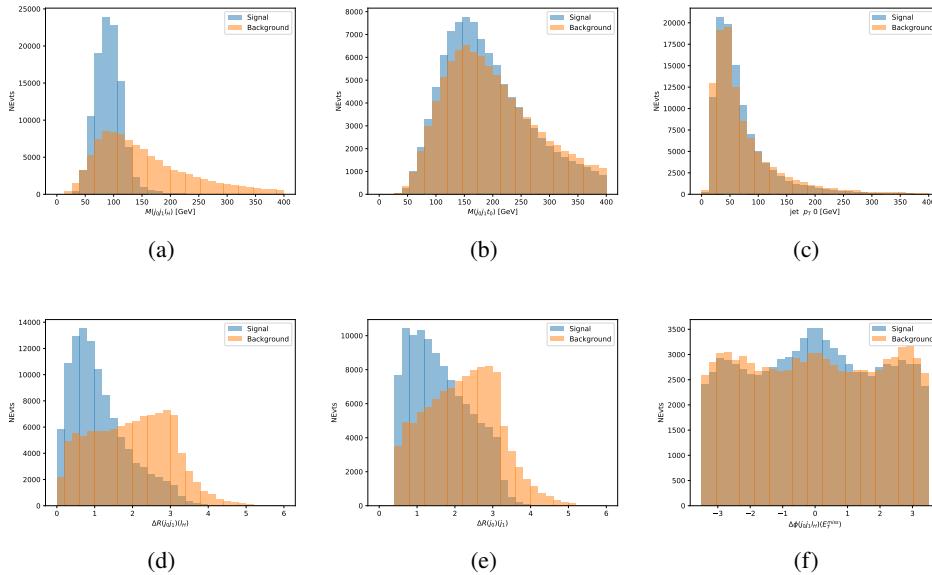


Figure 19.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1340 The modeling of these inputs is validated against data, with figure 19.11 showing good
 1341 general agreement between data and MC. Plots for the complete list of features can found in
 1342 appendix A.1.

1343 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.
 1344 Once again, incorrect combinations are downsampled, such that the correct combinations are
 1345 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN

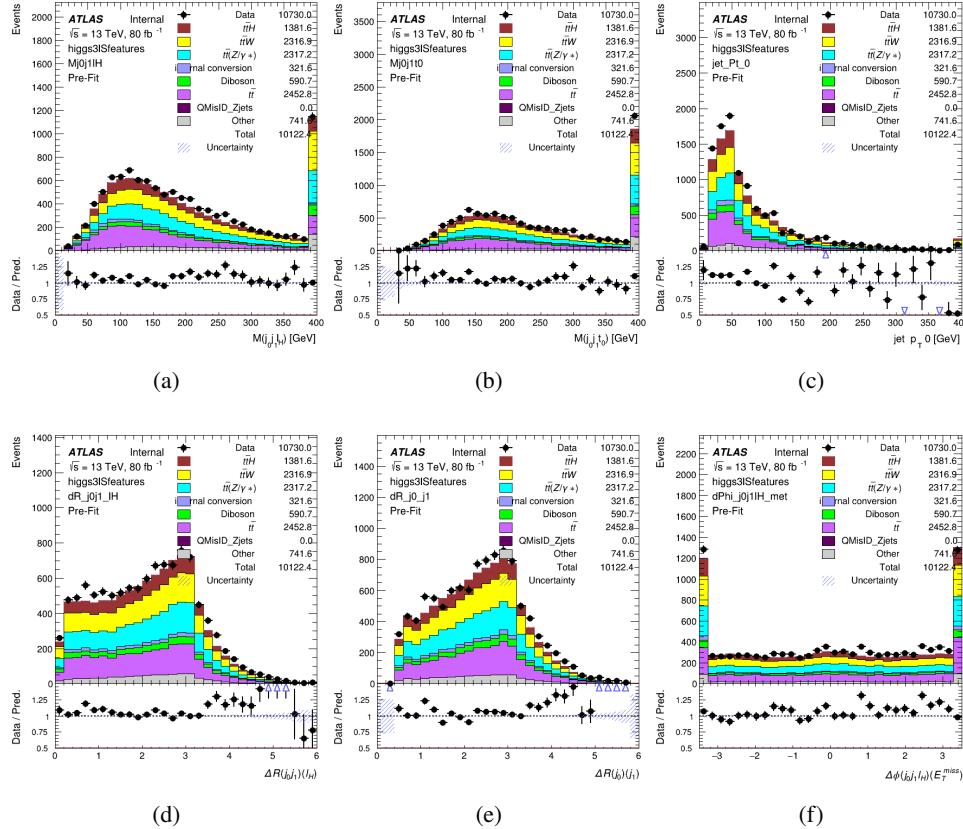


Figure 19.11: Data/MC comparisons of input features for higgs3IS training for 80 fb^{-1} of data.

1346 is summarized in figure 19.3.2.

1347 The neural network identifies the correct combination 64% of the time. It identifies the
1348 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
1349 83% of the time.

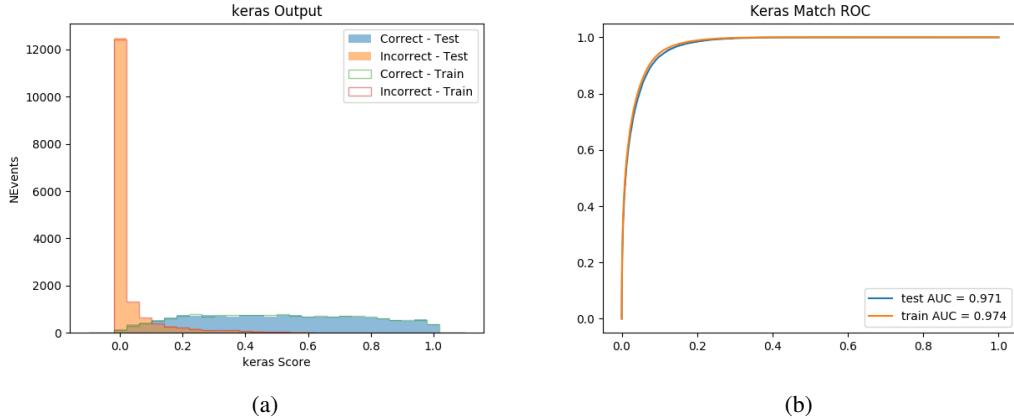


Figure 19.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1350 19.3.3 3l Fully-leptonic Channel

1351 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 1352 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to
 1353 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 1354 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 1355 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 36: Input features used to identify the Higgs decay products in 3lF events

1356 Here l_{H0} and l_{H1} are the Higgs decay candidates. The other lepton in the event is labeled
 1357 l_T . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco
 1358 Score is the output of the Higgs reconstruction algorithm.

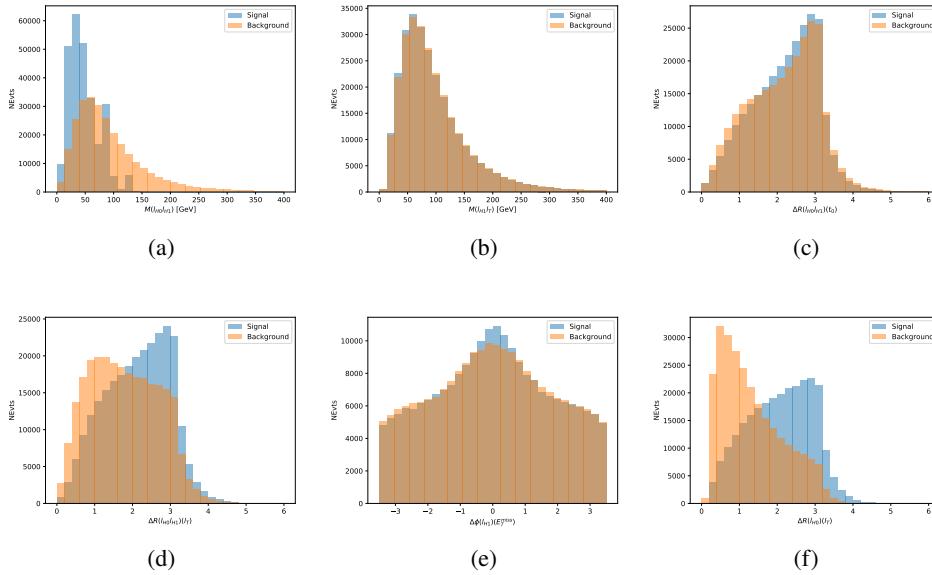


Figure 19.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1359 The modeling of these inputs is validated against data, with figure 19.14 showing good
 1360 general agreement between data and MC. Plots for the complete list of features can found in
 1361 section A.

1362 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,
 1363 with 10% of the dataset reserved for testing. The output of the model is summarized in figure
 1364 19.3.3.

1365 The correct lepton is identified by the model for 80% of events in the testing data set.

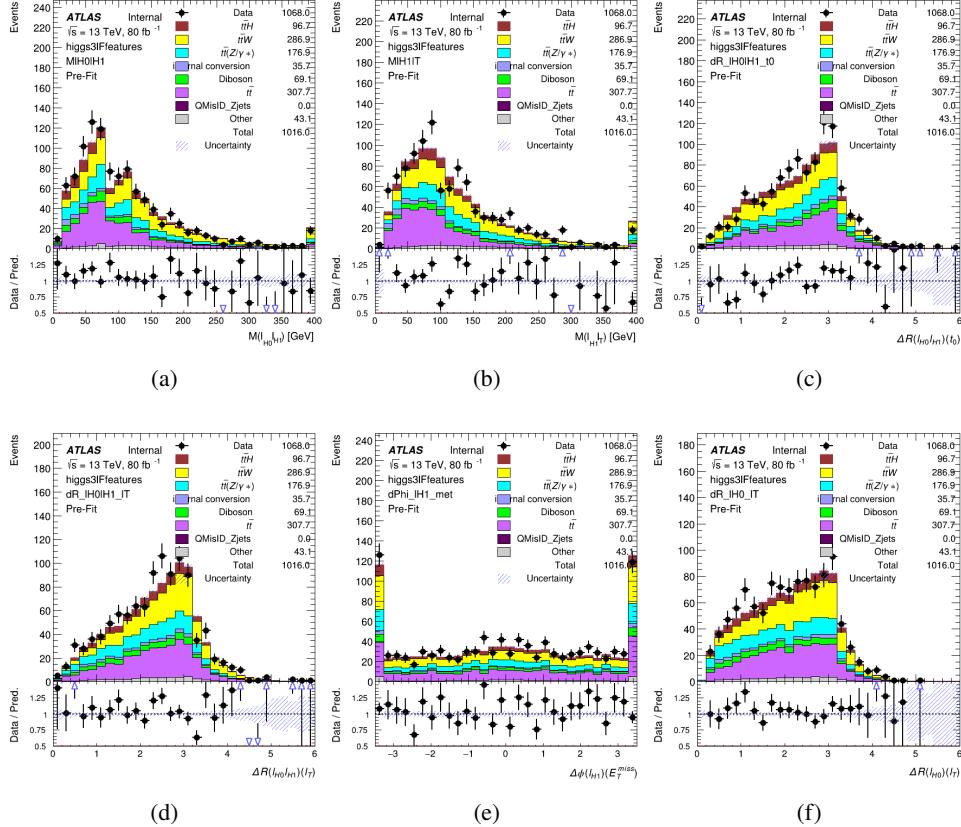


Figure 19.14: Data/MC comparisons of input features for higgs3IF training for 80 fb^{-1} of data.

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.

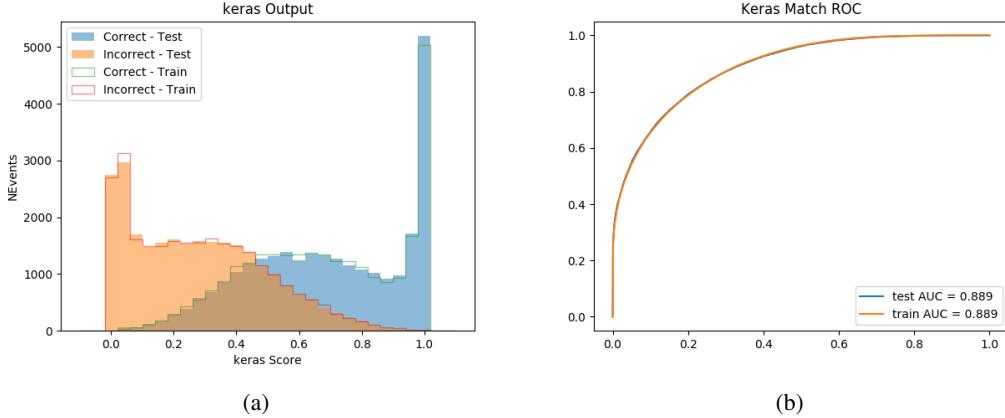


Figure 19.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

1373 As a two-bin fit is targeted for the final result, some metrics evaluating the performance
 1374 of the models aim to show how well it distinguished between "high p_T " and "low p_T " events. A
 1375 cutoff point of 150 GeV is used to define these two categories.

1376 Because the analysis uses a two bin fit of the Higgs p_T , the momentum reconstruction
 1377 could be treated as a binary classification problem, rather than a regression problem. This
 1378 approach is explored in detail in section A.4, and is found not to provide any significant increase
 1379 in sensitivity. The regression approach is used because it provides more flexibility for future
 1380 analyses, as it is independent of the cutoff between high and low p_T , as well as the number of
 1381 bins. Further, a regression allows the output of the neural network to be more clearly understood,
 1382 as it can be directly compared to a physics observable.

1383 19.4.1 2lSS Channel

1384 The input variables listed in table 37 are used to predict the Higgs p_T in the 2lSS channel. Here
1385 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
1386 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
1387 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
1388 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
1389 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 37: Input features for reconstructing the Higgs p_T spectrum for 2lSS events

1390 The optimal neural network architecture for this channel is found to consist of 7 hidden
 1391 layers with 60 nodes each. The inputdata set includes 1.2 million events, 10% of which is used
 1392 for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1393 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 1394 truth Higgs p_T in figure 19.16. In order to visualize the model performance more clearly, in (a)
 1395 of that figure, the color of each point is determined by Kernal Density Estimation (KDE). The
 1396 color shown represents the logarithm of the output from KDE, to counteract the large number of
 1397 low p_T events. For that same reason, each column of the histogram shown in (b) of figure 19.16
 1398 is normalized to unity. This plot therefore demonstrates what the model predicts for each slice
 1399 of truth p_T .

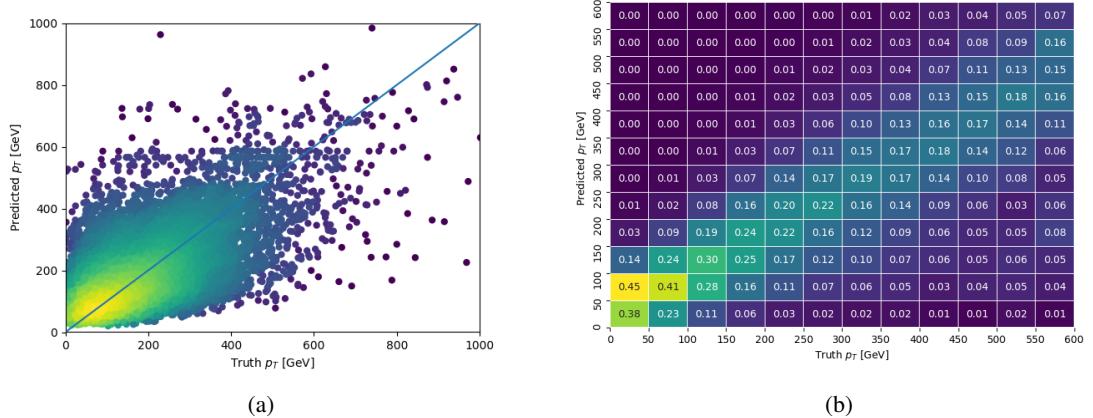


Figure 19.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernal Density Estimation, and (b) a histogram where each column has been normalized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

1400 We are also interested in how well the model distinguishes between events with $p_T < 150$
 1401 GeV and > 150 GeV. Figure 19.17 demonstrates the NN output for high and low p_T events based

1402 on this cutoff.

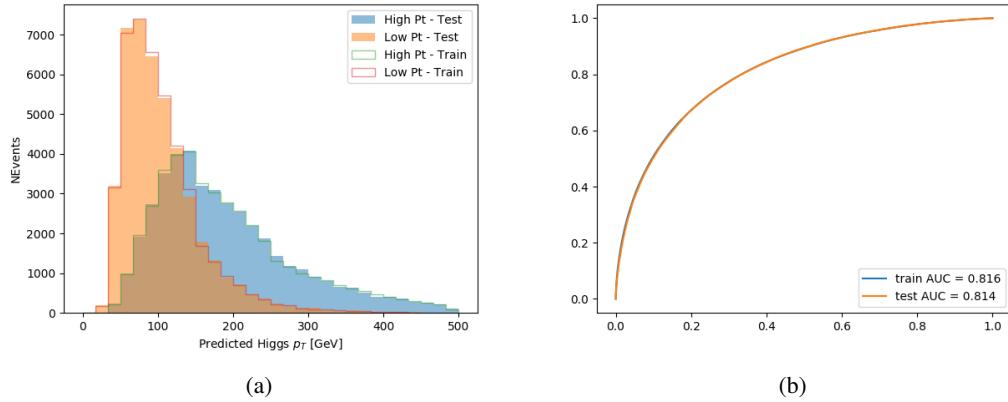


Figure 19.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

1403 **19.4.2 3l Semi-leptonic Channel**

1404 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 38: 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 39 are used to construct a model for predictin the Higgs p_T for 3lF events.

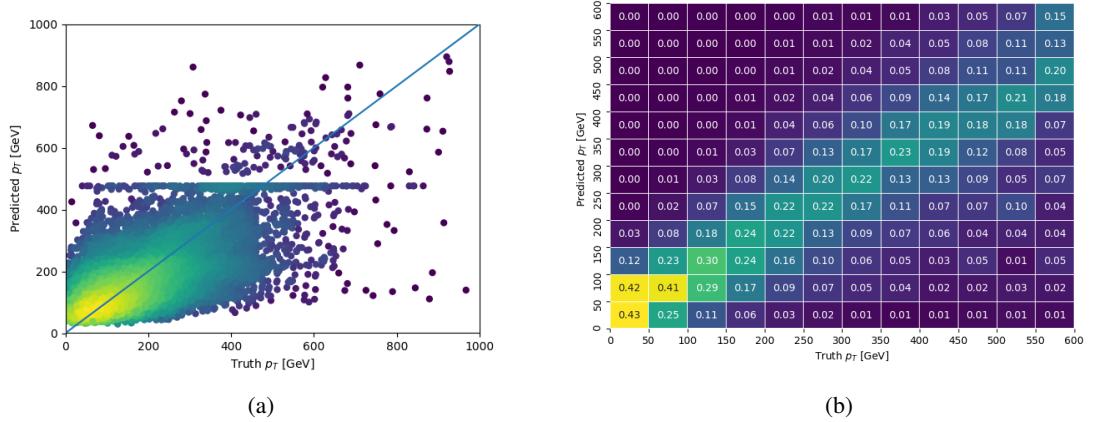


Figure 19.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

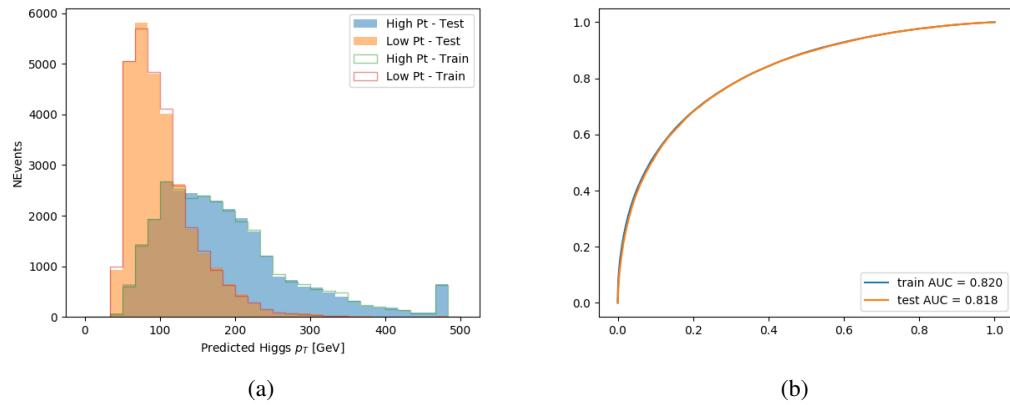


Figure 19.19: (a) shows the reconstructed Higgs p_T for 3lS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

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

1423 l_{H0} and l_{H1} represent the two leptons identified by the Higgs reconstruction model as
 1424 originating from the Higgs, while l_T is the other lepton in the event. The Higgs Reco Score and
 1425 b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification
 1426 algorithm, respectively.

1427 The optimal neural network architecture for this channel is found to consist of 5 hidden
 1428 layers with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for
 1429 testing, the other 90% for training. Training is found to converge after around 150 epochs.

1430 The predicted transverse momentum, as a function of the truth p_T , is shown in figure
 1431 [19.20.](#)

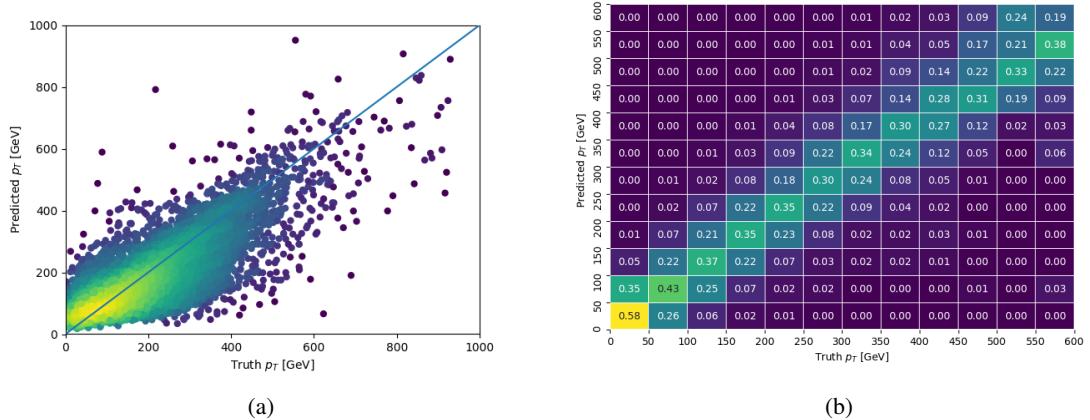


Figure 19.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized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

1432 When split into high and low p_T , based on a cutoff of 150 GeV, the

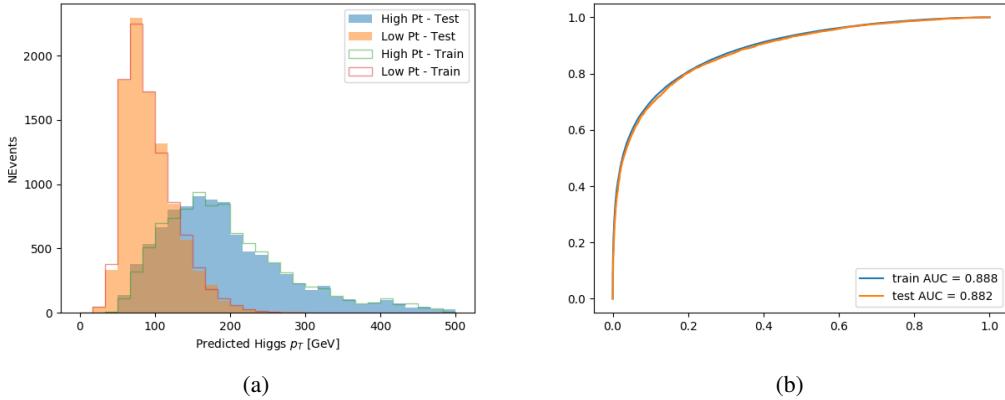


Figure 19.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

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 40.

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

1443 Here l_0 is the opposite charge lepton, l_1 and l_2 are the two SS leptons order by ΔR
 1444 from lepton 0. score3lF and score3lS are the outputs of the 3lS and 3lF Higgs reconstruction
 1445 algorithms, while topScore is the output of the b-jet identification algorithm.

1446 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these
 1447 two decay modes. The output of the model is summarized in figure 19.22.

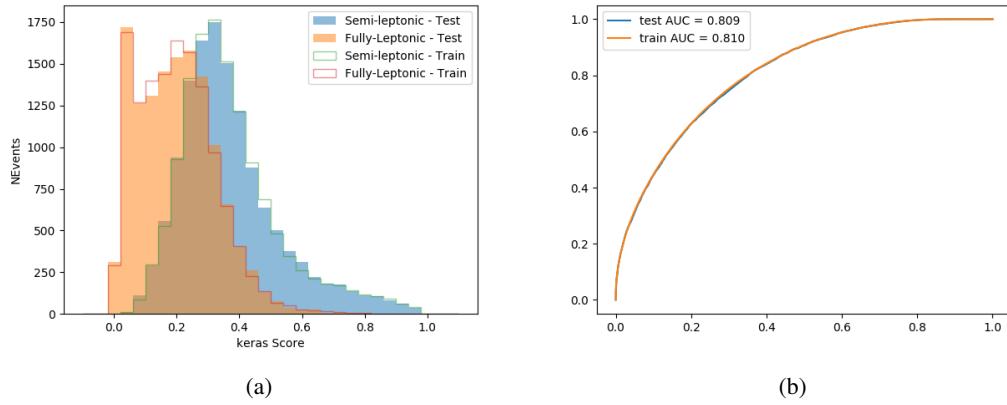


Figure 19.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

1448 A cutoff of 0.23 is determined to be optimal for separating 3lS and 3lF in the fit.

1449 20 Signal Region Definitions

1450 Events are divided into two channels based on the number of leptons in the final state: one with
 1451 two same-sign leptons, the other with three leptons. The 3l channel includes events where both
 1452 leptons originated from the Higgs boson as well as events where only one of the leptons

1453 20.1 Pre-MVA Event Selection

1454 A preselection is applied to define orthogonal analysis channels based on the number of leptons

1455 in each event. For the 2lSS channel, the following preselection is used:

- 1456
- Two very tight, same-charge, light leptons with $p_T > 20 \text{ GeV}$

- 1457
- ≥ 4 reconstructed jets, ≥ 1 b-tagged jets

- 1458
- No reconstructed tau candidates

1459 The event yield after the 2lSS preselection has been applied, for MC and data at 80 fb^{-1} ,

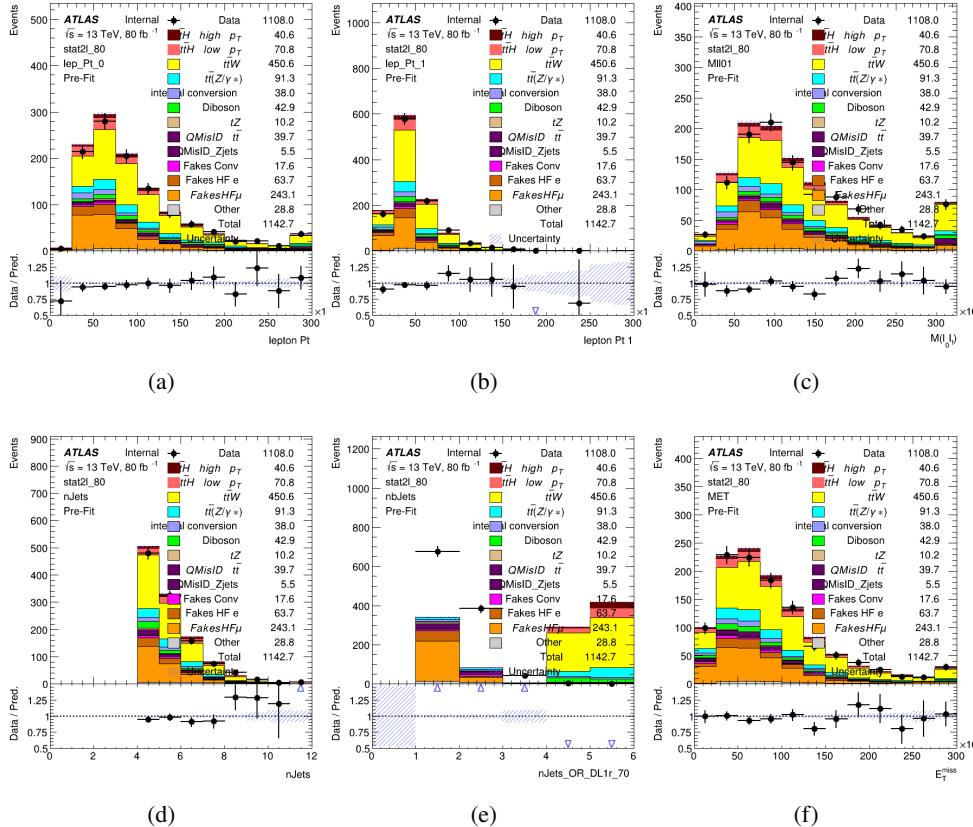
1460 is shown in table 41.

	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 41: Event yield in the 2ISS preselection region.

1461

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.

1462

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

1463

- 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

1464

1465

- 1466 • ≥ 2 reconstructed jets, ≥ 1 b-tagged jets
- 1467 • No reconstructed tau candidates
- 1468 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs
- 1469 The event yield after the 3l preselection has been applied, for MC and data at 80 fb^{-1} , is
- 1470 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 42: Yields of the analysis

1471 Comparisons of kinematic distributions for data and MC in this region are shown in figure

1472 21.2.

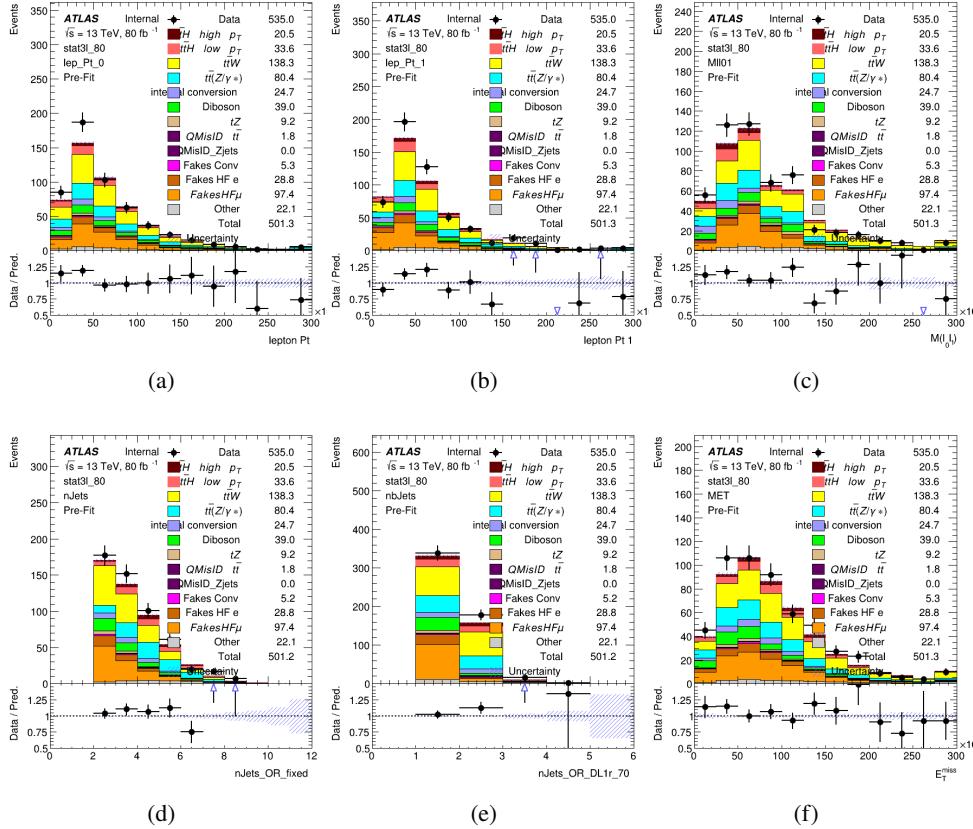


Figure 20.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1473 20.2 Event MVA

1474 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
 1475 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
 1476 (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

1477 [xgboost] are trained using the kinematics of signal and background events derived from Monte
1478 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo
1479 event.

1480 Because the background composition differs for events with a high reconstructed Higgs p_T
1481 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and
1482 low p_T regions. This is found to provide better significance than attempting to build an inclusive
1483 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6
1484 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
1485 3lF high p_T , and 3lF low p_T .

1486 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 43: Input features used to distinguish signal and background events in the 2ISS channel.

1487

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

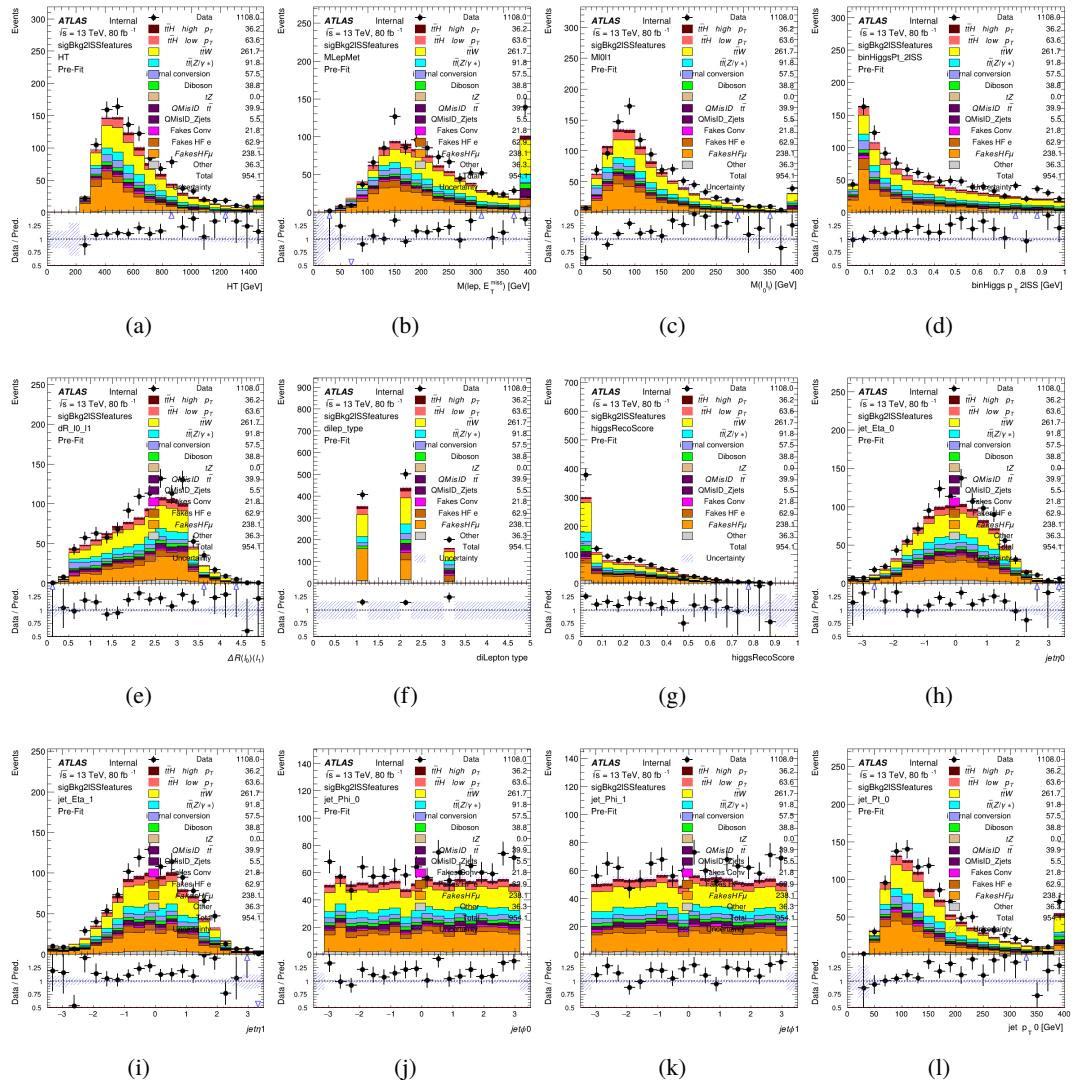


Figure 20.3:

1488 The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss
1489 function.

1490 Output distributions of each MVA comparing MC prediction to data at 80 fb^{-1} are shown
1491 in figures 20.7-20.2.

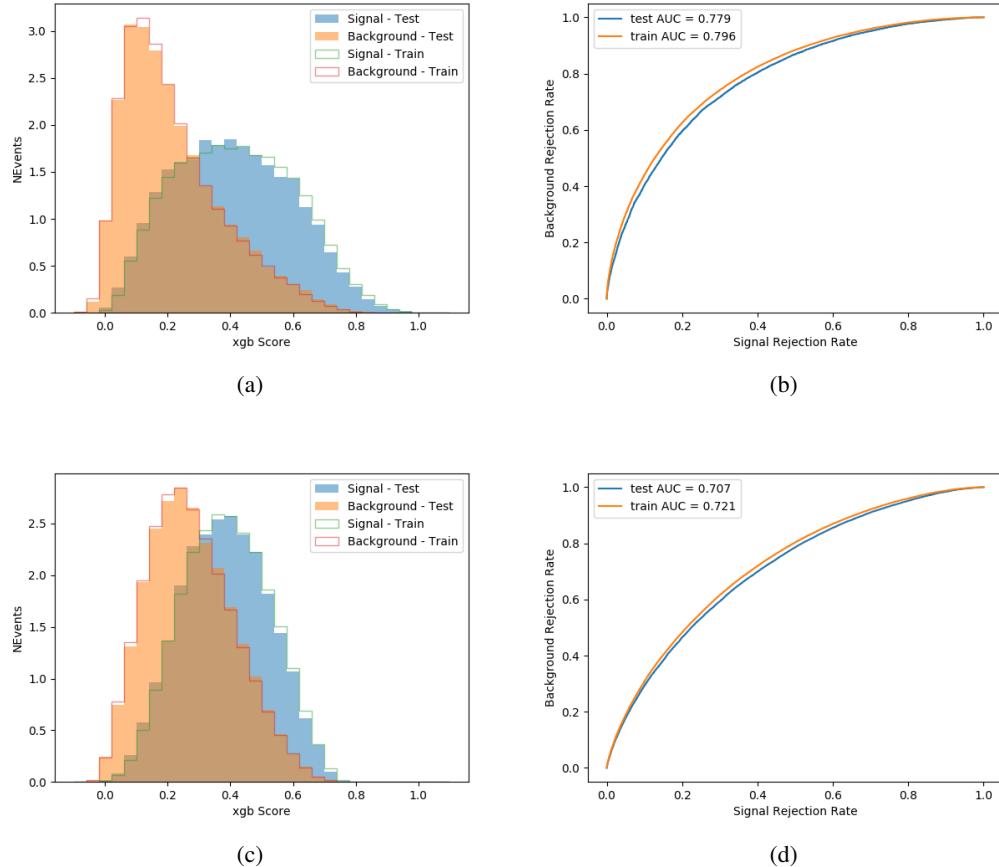


Figure 20.4:

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

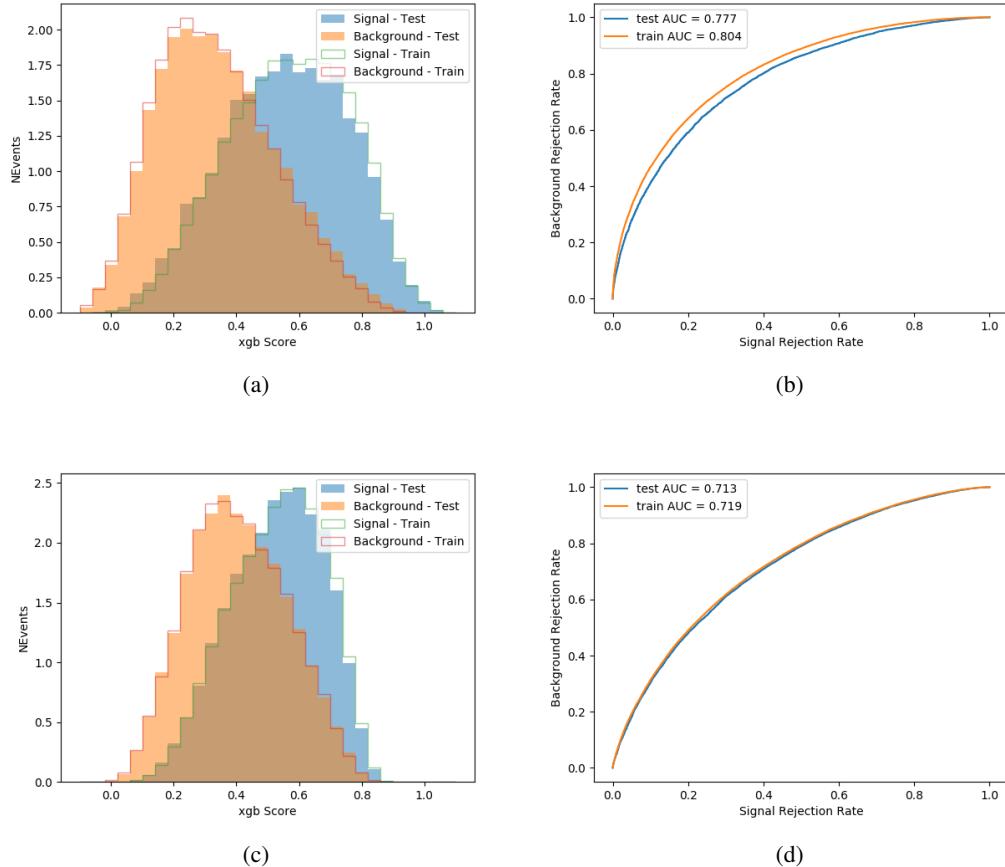


Figure 20.5:

1498 is predicted to be high (> 150 GeV) or low (< 150 GeV), a cut on the appropriate background
 1499 rejection MVA is applied. The particular cut values, listed in table 45, are determined by
 1500 maximizing S/\sqrt{B} in each region.

1501 The event preselection and MVA selection define the three signal regions. These signal
 1502 region definitions are summarized in table 46.

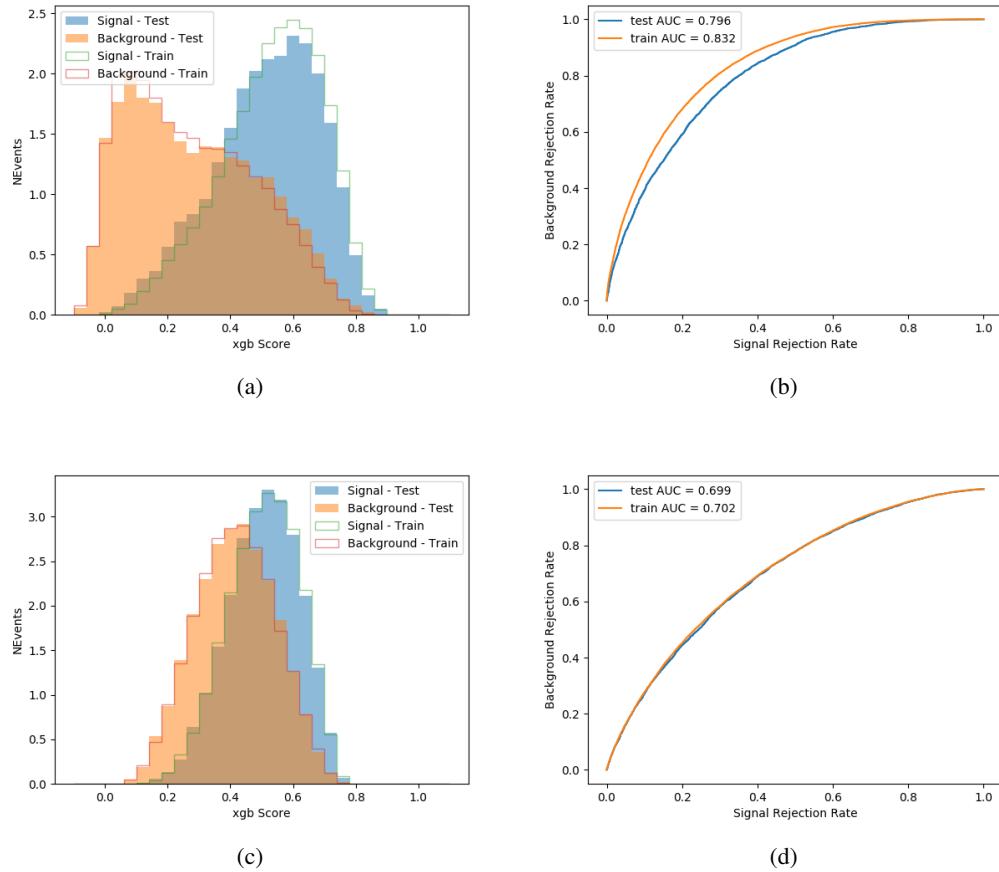
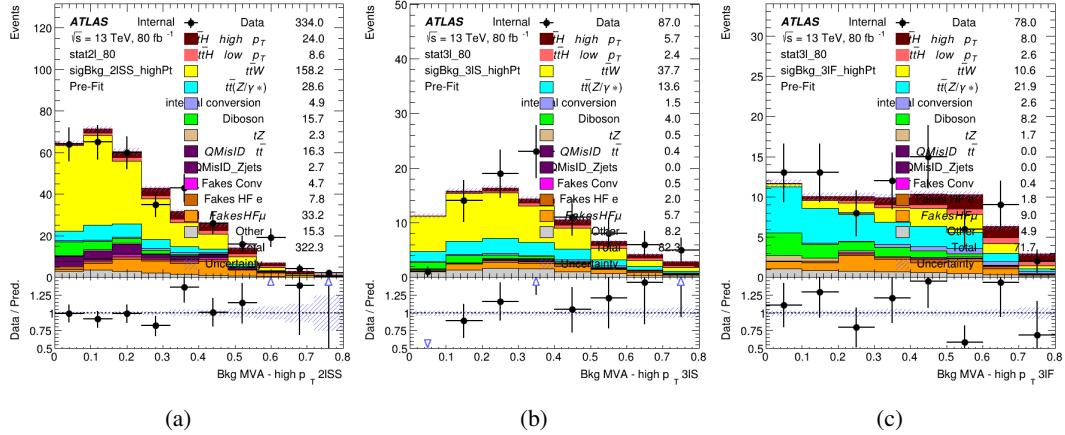
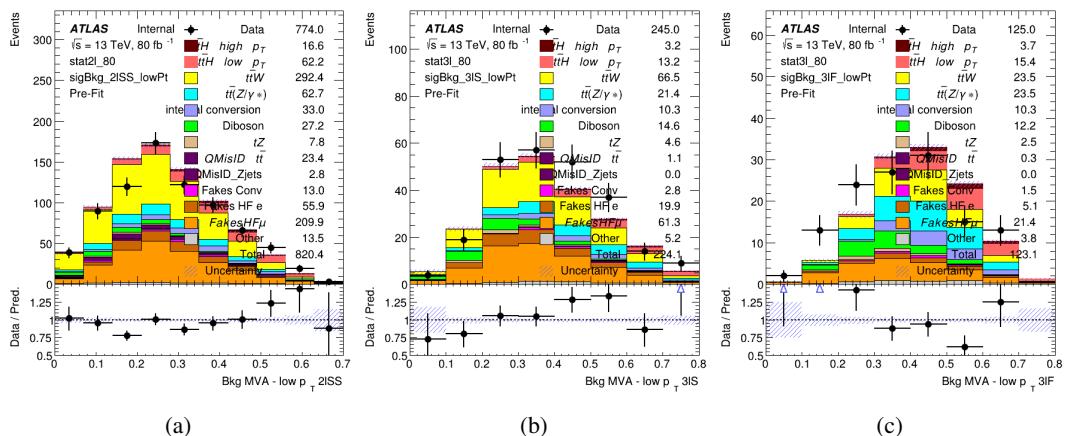


Figure 20.6:

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

Table 46: Selection applied to define the three signal regions used in the fit.

21 Background Rejection MVA

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

1507 21.1 Pre-MVA Event Selection

1508 A preselection is applied to define orthogonal analysis channels based on the number of leptons
 1509 in each event. For the 2lSS channel, the following preselection is used:

- 1510 • Two very tight, same-charge, light leptons with $p_T > 20$ GeV

- 1511 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets

- 1512 • No reconstructed tau candidates

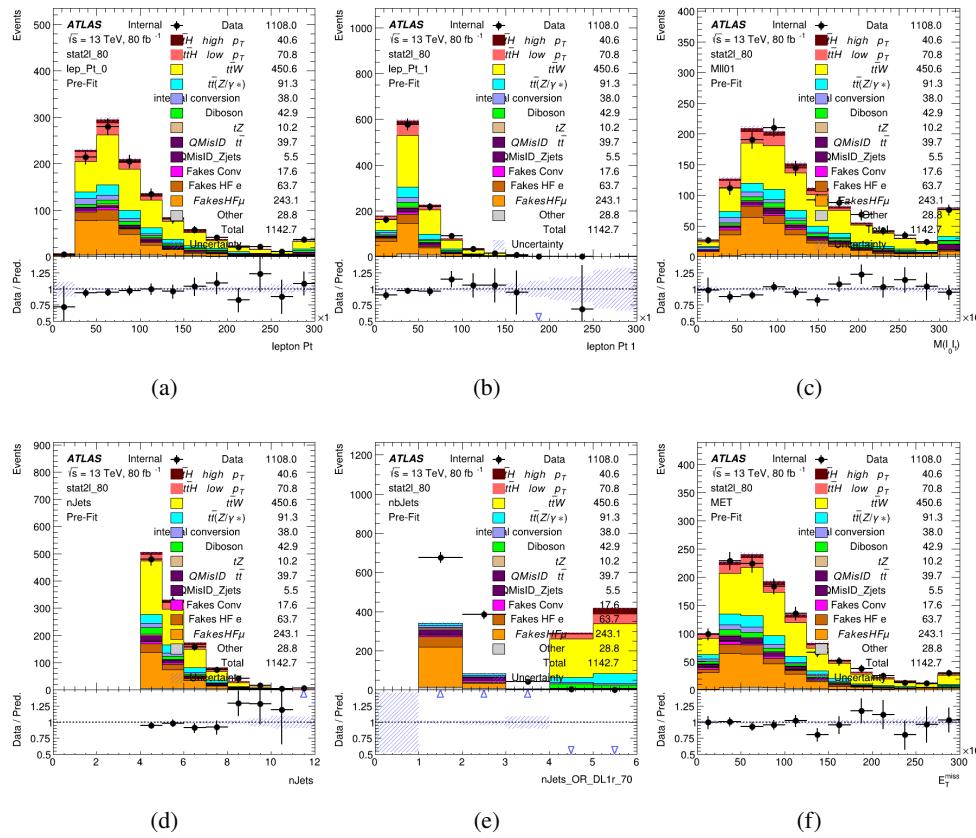


Figure 21.1:

1513 For the 3l channel, the following selection is applied:

1514 • Three light leptons with total charge ± 1

1515 • Same charge leptons are required to be very tight, with $p_T > 20 \text{ GeV}$

1516 • Opposite charge lepton must be loose, with $p_T > 10 \text{ GeV}$

1517 • $>= 2$ reconstructed jets, $>= 1$ b-tagged jets

1518 • No reconstructed tau candidates

1519 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs

1520 21.2 Event MVA

1521 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events

1522 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic

1523 (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

1524 [xgboost] are trained using the kinematics of signal and background events derived from Monte

1525 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo

1526 event.

1527 Because the background composition differs for events with a high reconstructed Higgs p_T

1528 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and

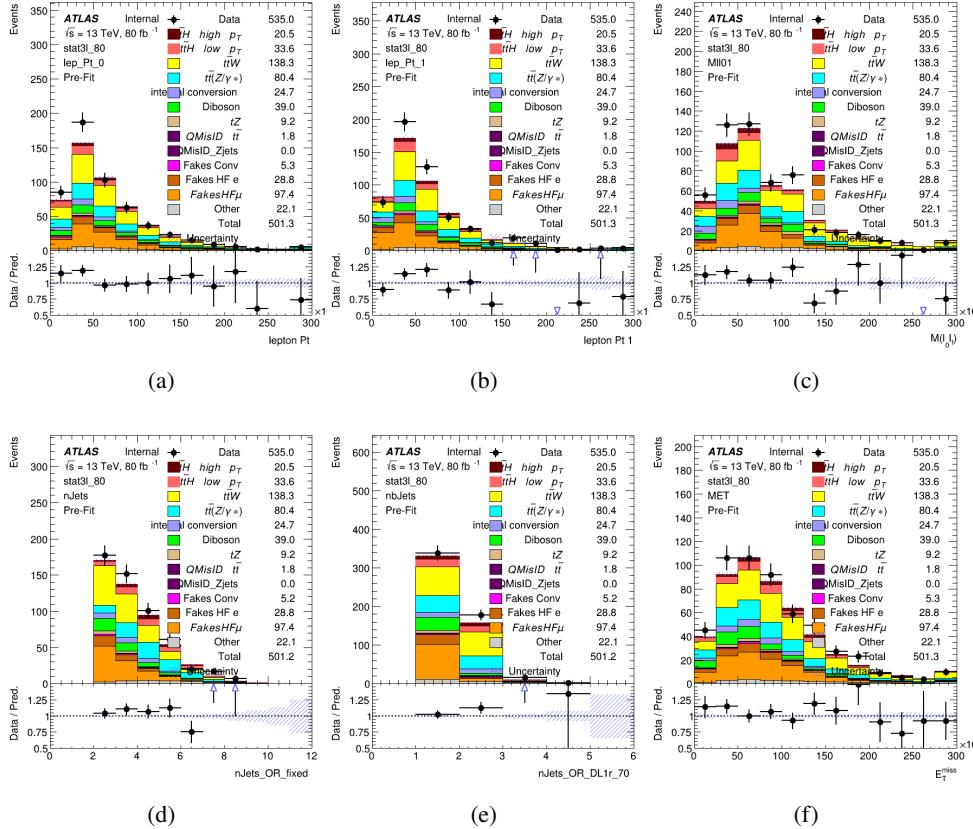


Figure 21.2:

1529 low p_T regions. This is found to provide better significance than attempting to build an inclusive
 1530 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6
 1531 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
 1532 3lF high p_T , and 3lF low p_T .

1533 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 47: Input features used to distinguish signal and background events in the 2ISS channel.

1534

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

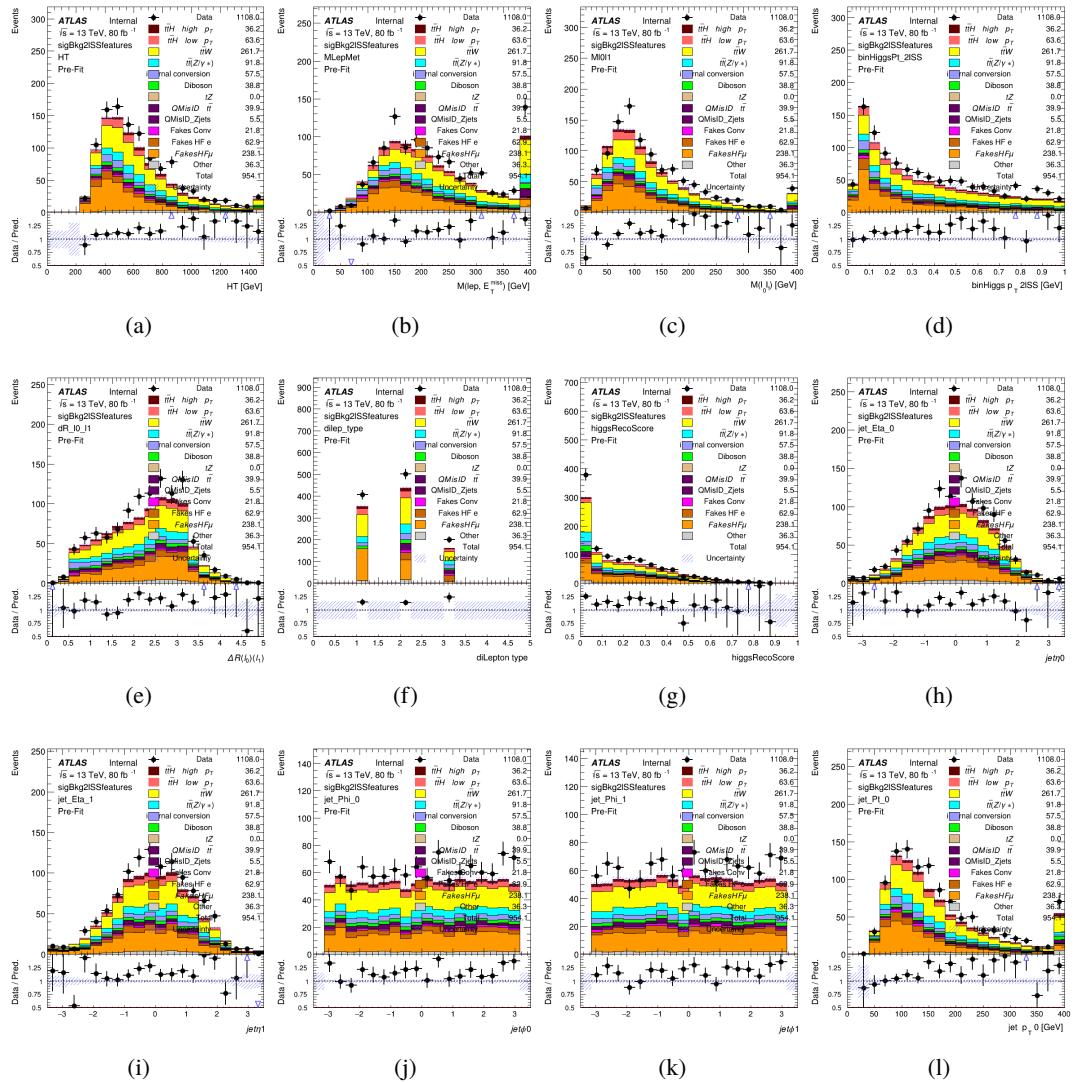


Figure 21.3:

1535 The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss
 1536 function.

1537 Output distributions of each MVA comparing MC prediction to data at 80 fb^{-1} are shown
 1538 in figure 21.2.

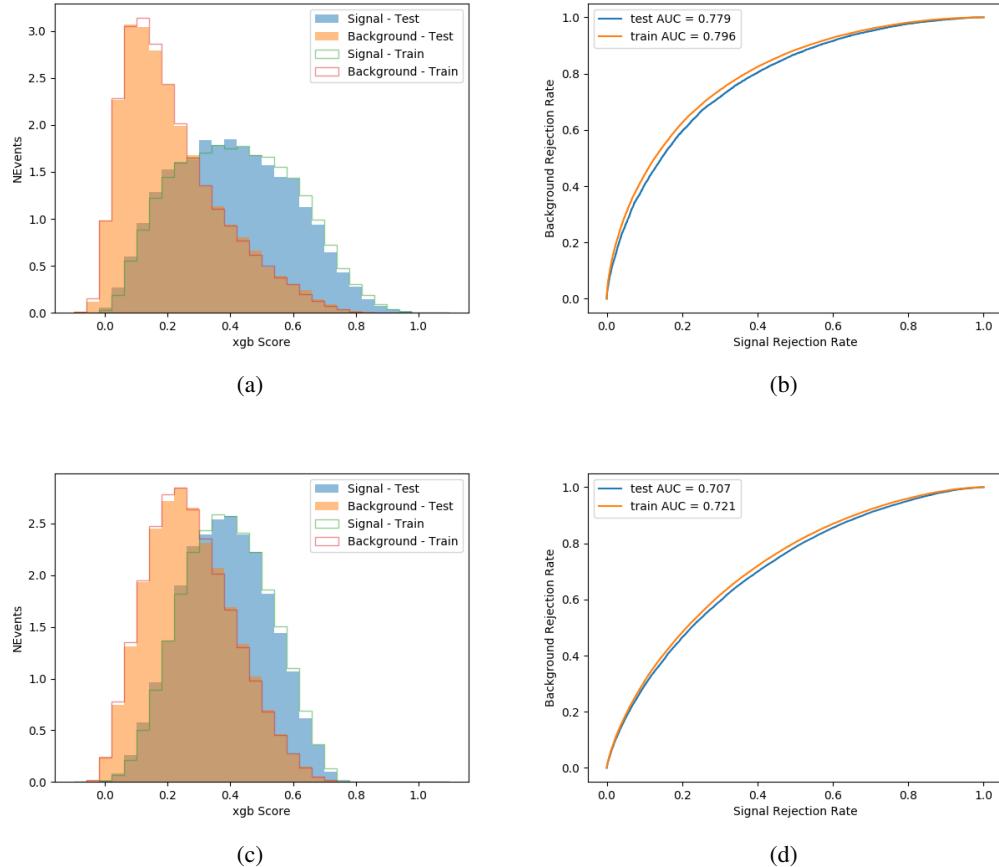


Figure 21.4:

1539 21.3 Signal Region Definitions

1540 Once pre-selection has been applied, channels are further refined based on the MVAs described
 1541 above. The output of the model described in section 19.5 is used to separate the three channel
 1542 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs
 1543 boson.

1544 For each event, depending on the channel as well as the predicted p_T of the Higgs derived

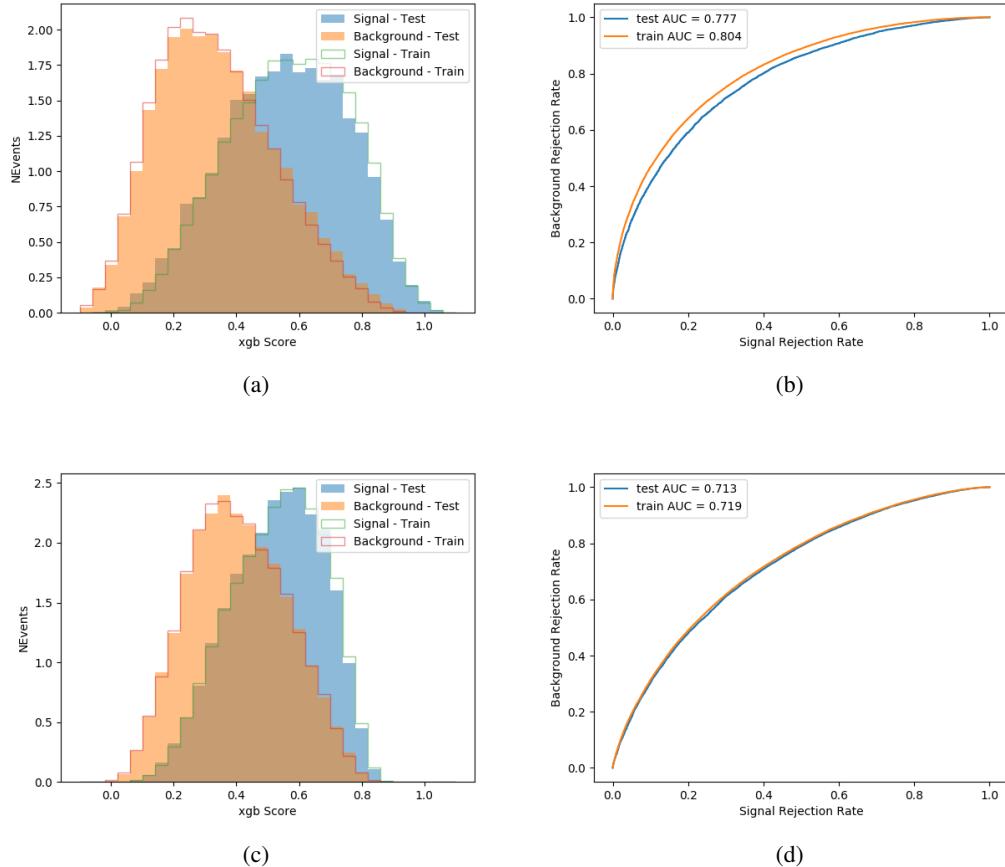


Figure 21.5:

from the algorithm described in section 19.4, a cut on the appropriate background rejection
 1545 algorithm is applied. The specific selection used, and the event yield in each channel after this
 1546 selection has been applied, is summarized below.
 1547

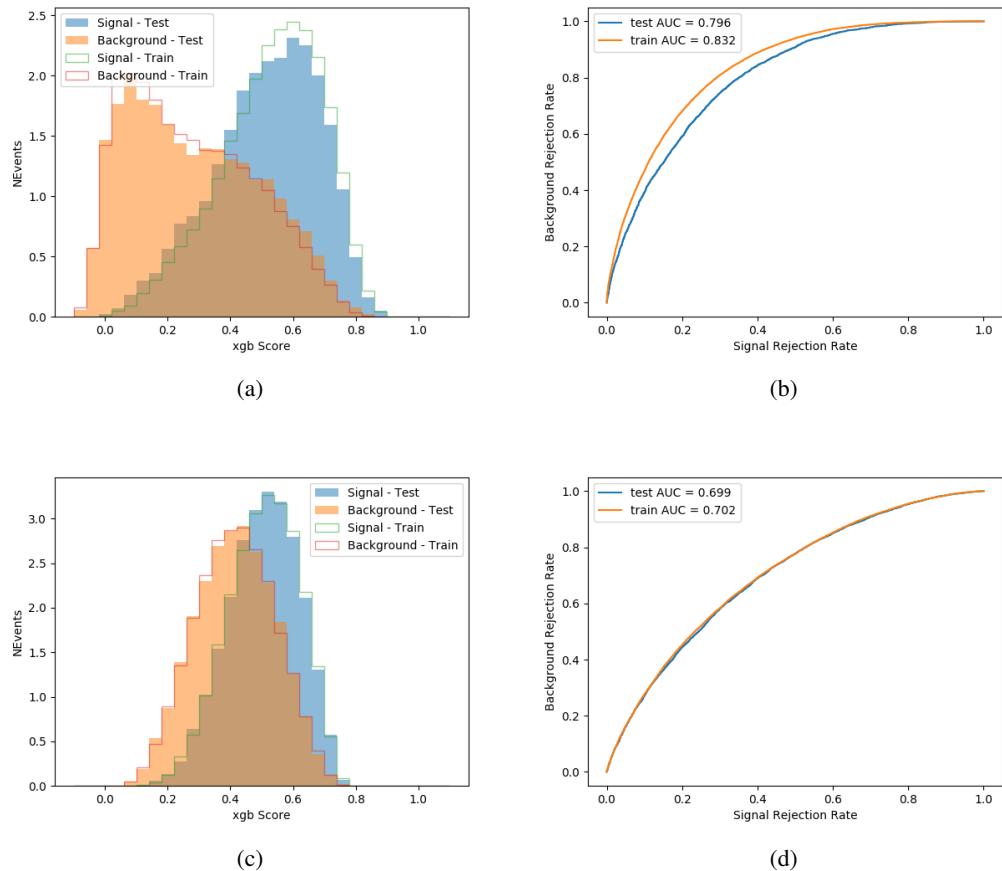


Figure 21.6:

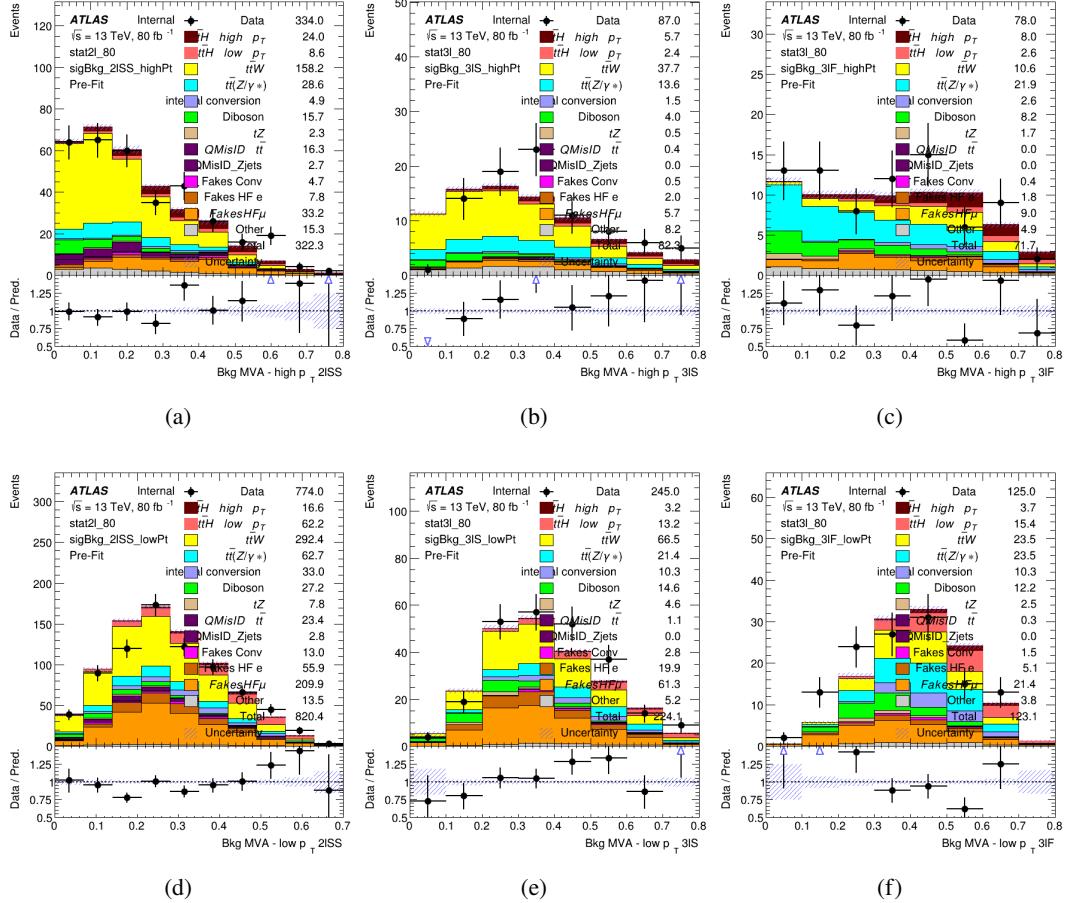


Figure 21.7: scores

1548 **21.3.1 2ISS**1549 **21.3.2 3l – Semi – leptonic**1550 **21.3.3 3l – Fully – leptonic****22 Systematic Uncertainties**

1552 The systematic uncertainties that are considered are summarized in table 49. These are implemented in the fit either as a normalization factors or as a shape variation or both in the signal
 1553 24th March 2021 – 17:56

1554 and background estimations. The numerical impact of each of these uncertainties is outlined in
1555 section 23.

Table 49: 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

1556 The uncertainty in the combined 2015+2016 integrated luminosity is derived from a
1557 calibration of the luminosity scale using x-y beam-separation scans performed in August 2015
1558 and May 2016 [**lumi**].

1559 The experimental uncertainties are related to the reconstruction and identification of light
1560 leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The sources which contribute

1561 to the uncertainty in the jet energy scale [**jes**] are decomposed into uncorrelated components and
1562 treated as independent sources in the analysis.

1563 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
1564 [**btag_cal**] are also decomposed into uncorrelated components. The large number of components
1565 for b-tagging is due to the calibration of the distribution of the BDT discriminant.

1566 The systematic uncertainties associated with the signal and background processes are
1567 accounted for by varying the cross-section of each process within its uncertainty.

1568 23 Results

1569 Unblinded results are shown for the 80 fb^{-1} data set, as well as MC only projections of results
1570 using the full Run-2, 140 fb^{-1} dataset.

1571 23.1 Results - 80 fb^{-1}

1572 A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p_T spectrum
1573 in the three signal regions, 2lSS, 3lS, and 3lF, shown in figure 23.1. The $t\bar{t}H$ MC is split into
1574 high and low p_T , based on whether the truth p_T of the Higgs is above or below 150 GeV. The
1575 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.

1576

As the data collected from 2015-2017 has been unblinded for $t\bar{t}H$ – ML channels, rep-

1577

resenting 80 fb^{-1} , those events are unblinded.

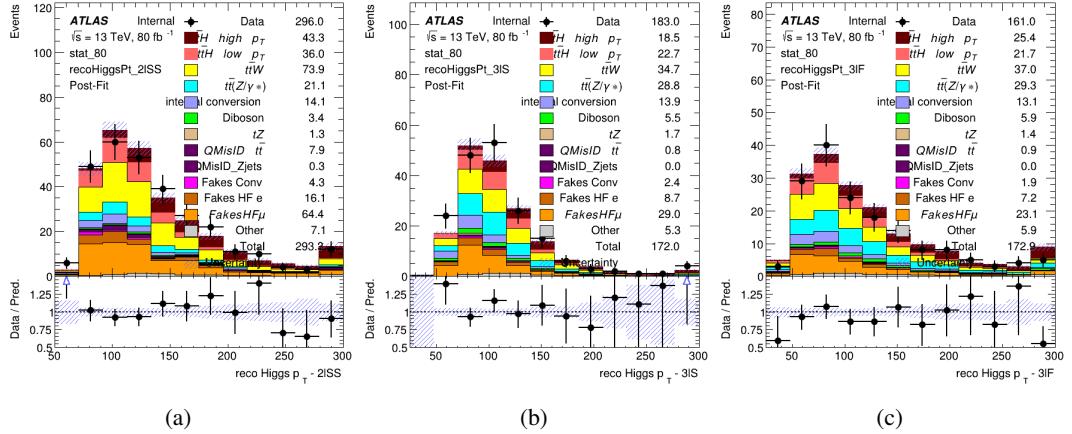


Figure 23.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 80 fb^{-1} of MC

1578

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

1579

The μ values for high and low p_T Higgs are shown in 50.

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

Table 50: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1580

Need to add something about systematics here

1581

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

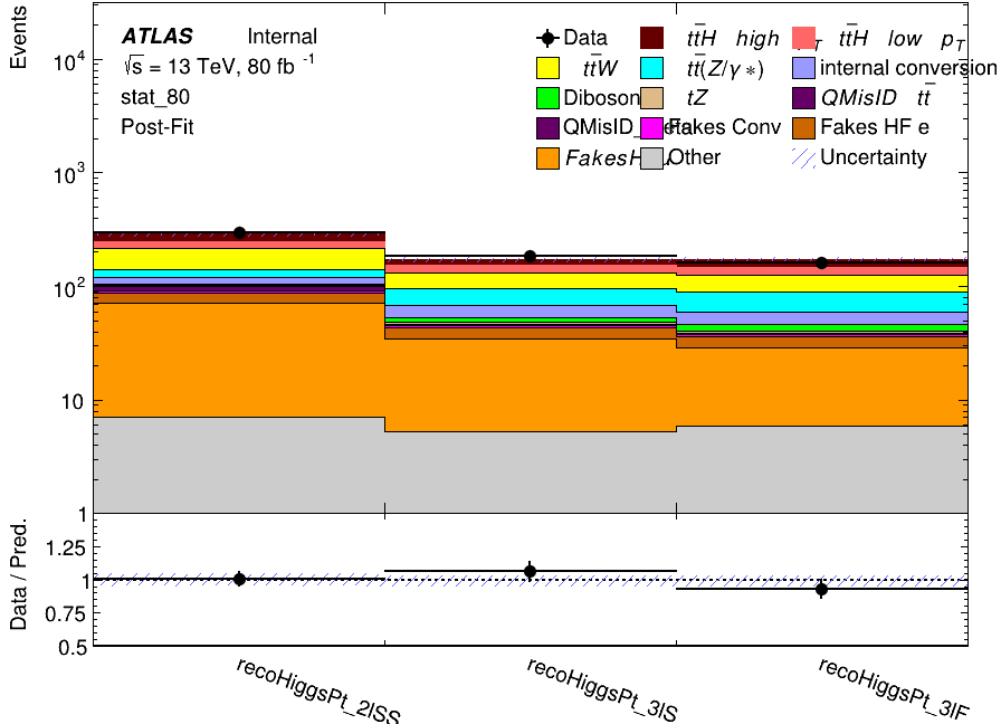


Figure 23.2: Post-fit summary of the yields in each signal region.

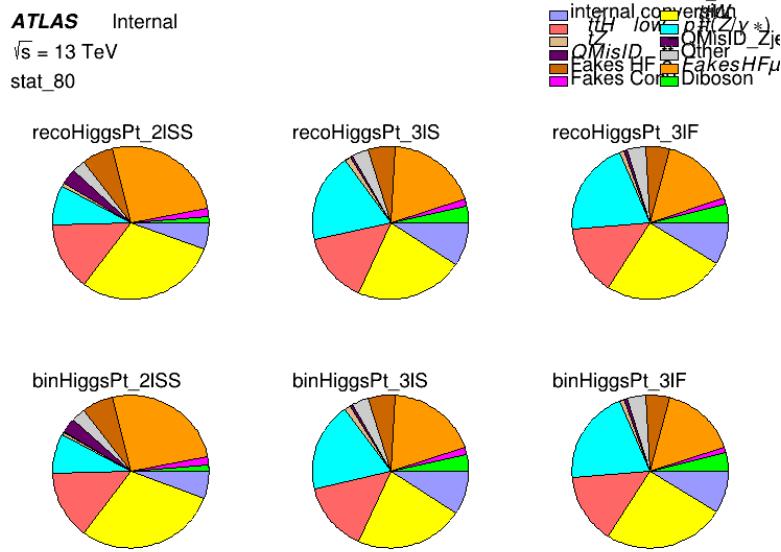


Figure 23.3: Background composition of the fit regions.

1582 **23.2 Projected Results - 140 fb^{-1}**

1583 As data collected in 2018 has not yet been unblinded for $t\bar{t}H - \text{ML}$ at the time of this note, data
 1584 from that year remains blinded. Instead, an Asimov fit is performed - with the MC prediction
 1585 being used both as the SM prediction as well as the data in the fit - in order to give expected
 1586 results.

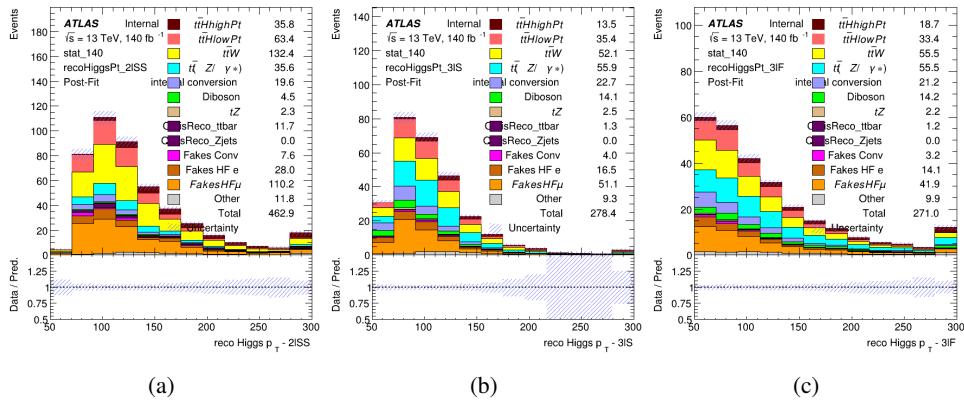


Figure 23.4: Blinded post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 140 fb^{-1} of data

1587 The μ values for high and low p_T Higgs are shown in 51.

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

Table 51: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

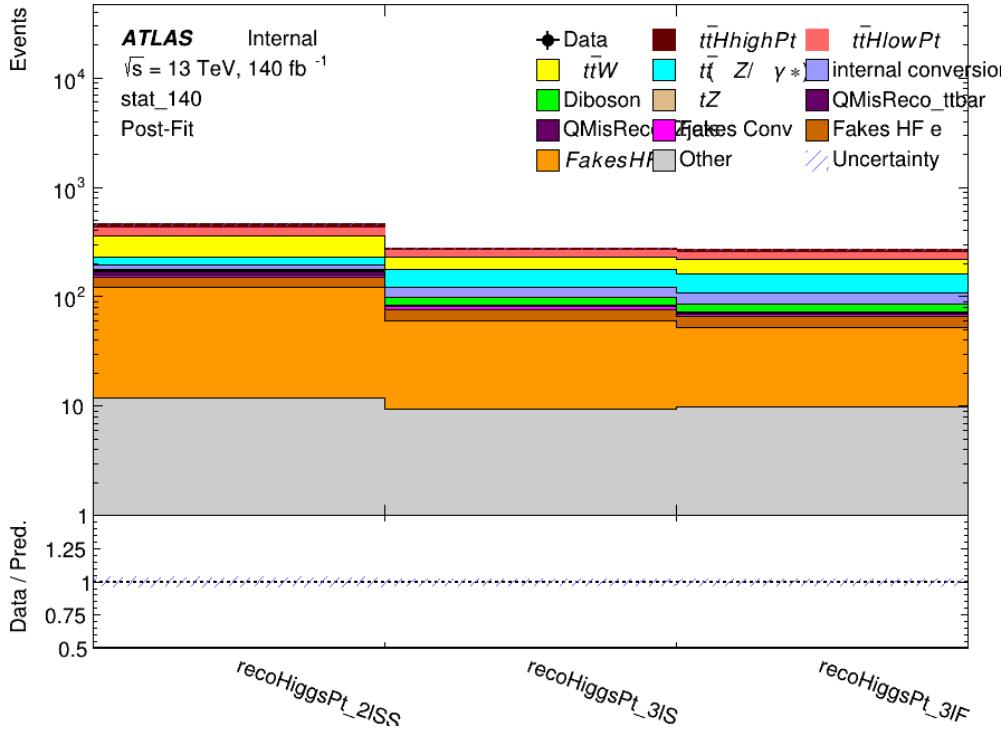


Figure 23.5: Post-fit summary of fit.

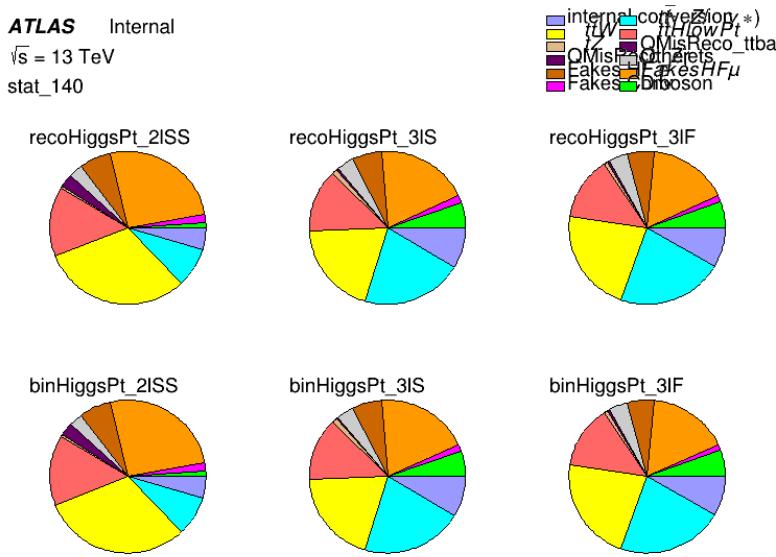


Figure 23.6: Background composition of the fit regions.

1588 Part VI**1589 Conclusion**

1590 As search for the effects of dimension-six operators on $t\bar{t}H$ production is performed. An effective
1591 field theory approached is used to parametrize the effects of high energy physics on the Higgs
1592 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,
1593 and the limits on dimension-six operators are limited to X.

List of contributions

1594

1595

1596 **Appendices**

1597 **A Machine Learning Models**

1598 The following section provides details of the various MVAs as well as a few studies performed
1599 in support of this analysis, exploring alternate decisions and strategies.

1600 **A.1 Higgs Reconstruction Models**

1601 **A.1.1 b-jet Identification Features - 2lSS**

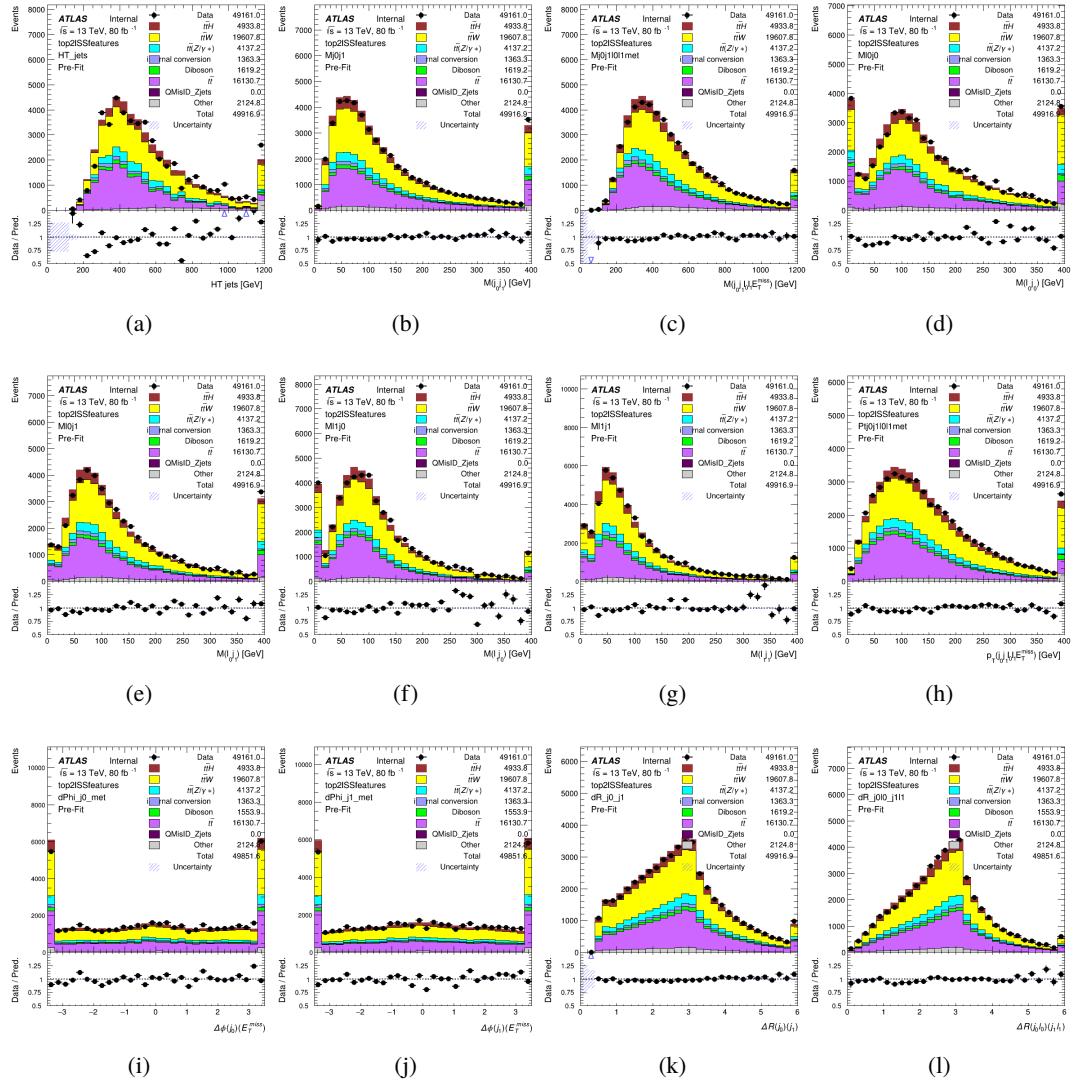


Figure A.1: Input features for top2lSS

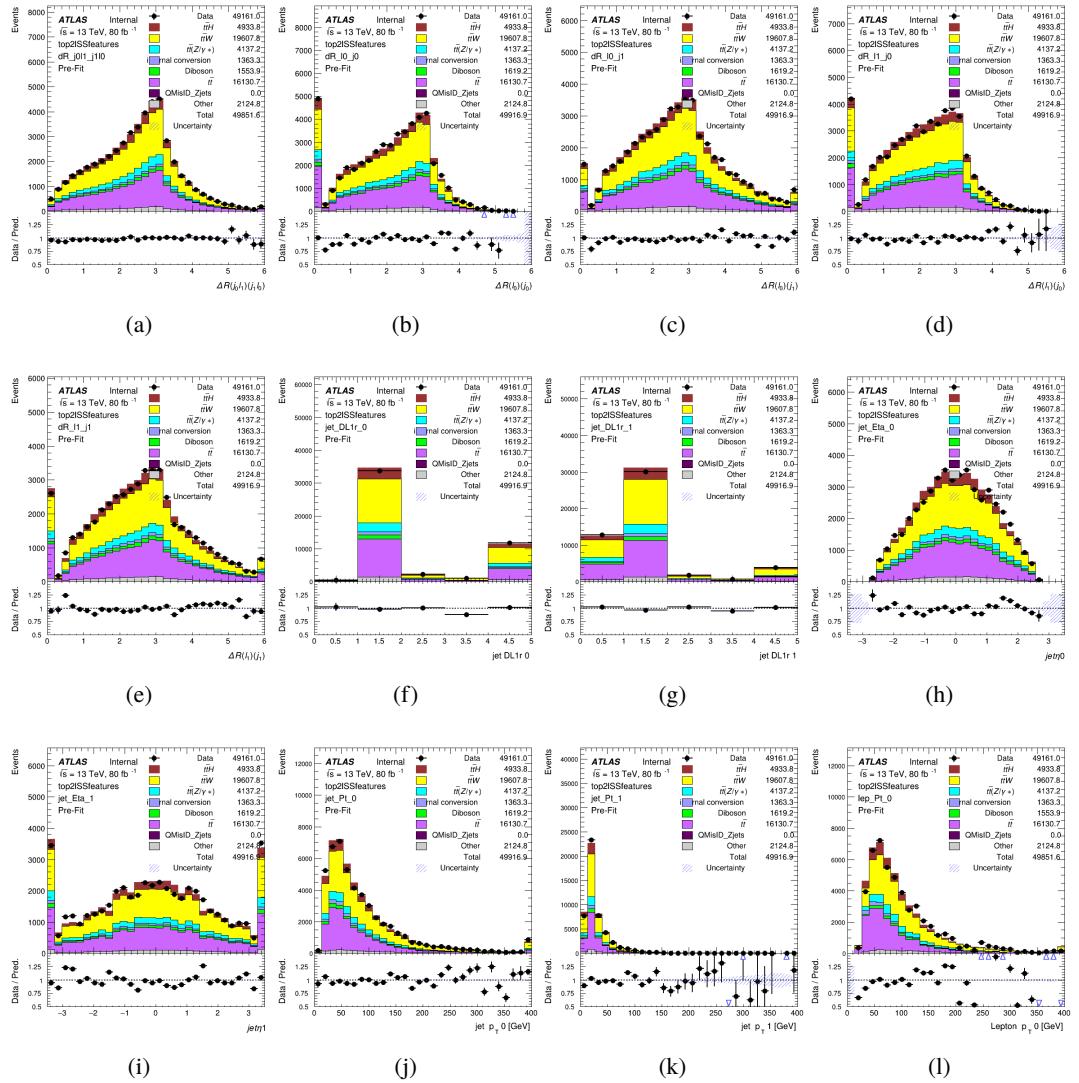


Figure A.2: Input features for top2lSS

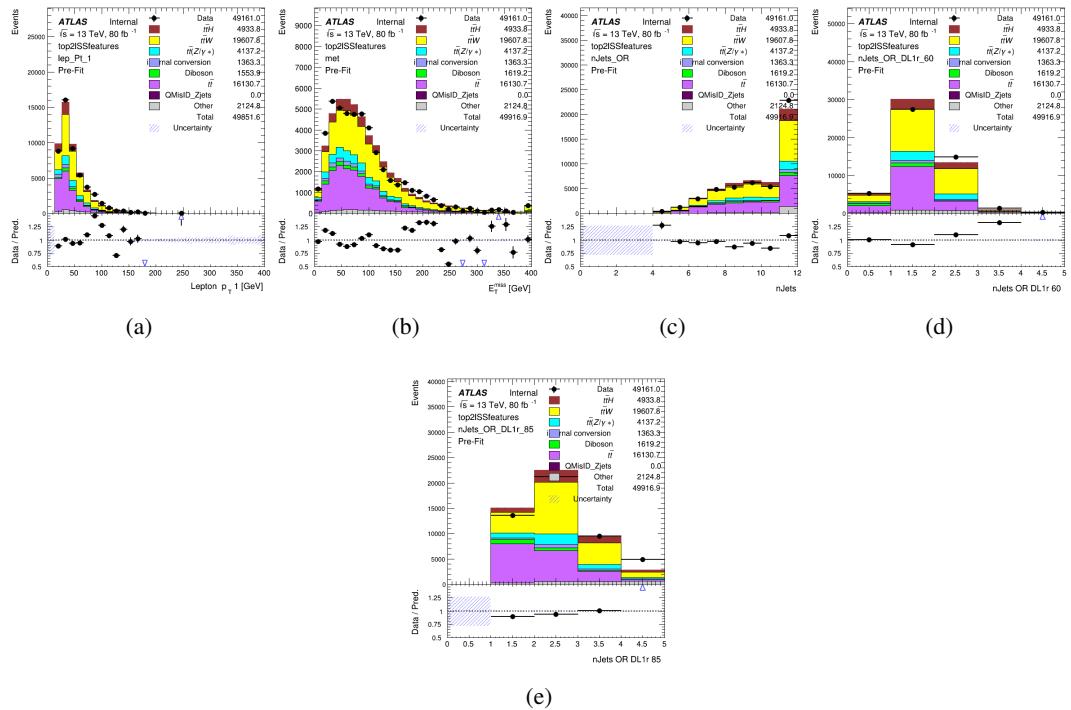


Figure A.3: Input features for top2lSS

1602 **A.1.2 b-jet Identification Features - 3l**

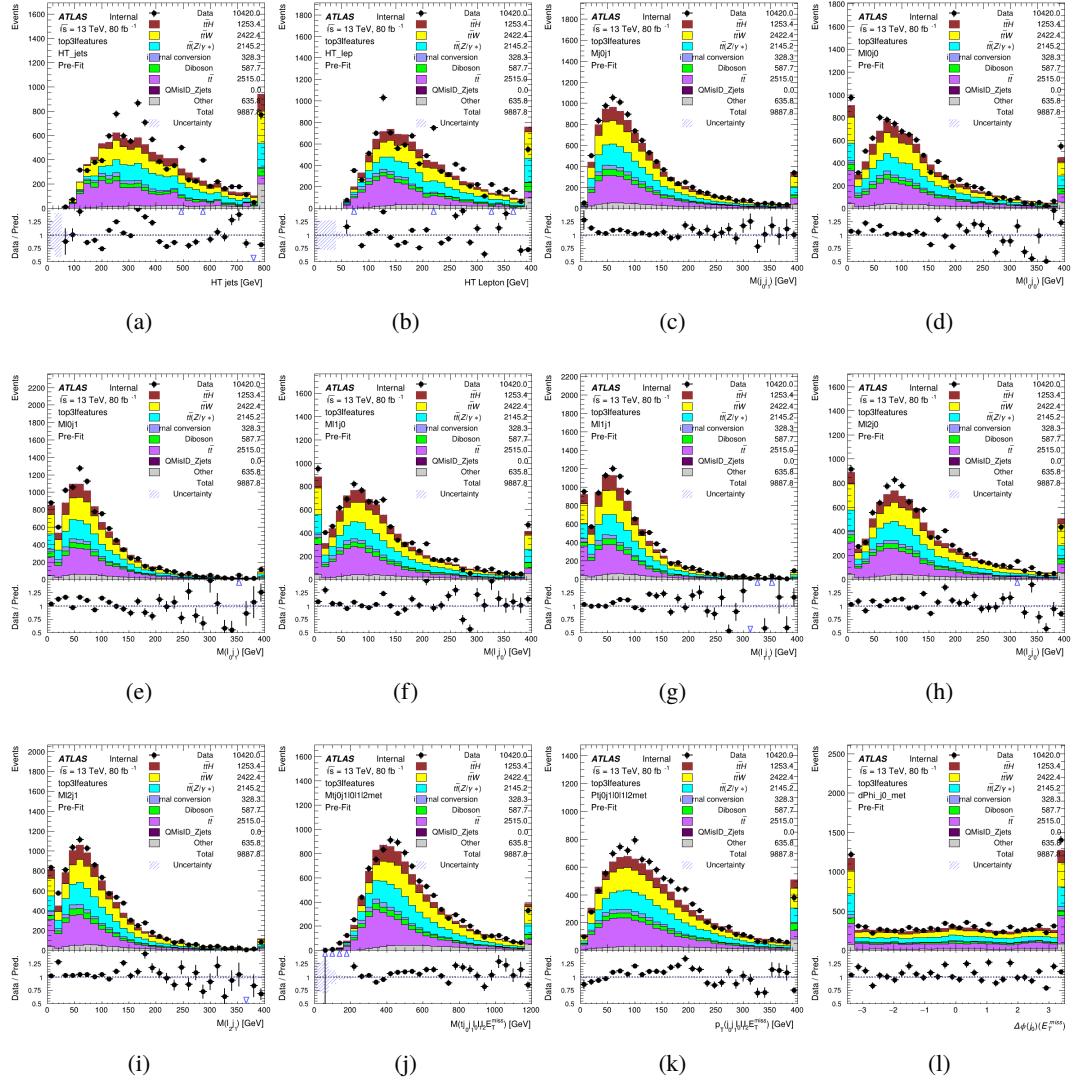


Figure A.4: Input features for top31

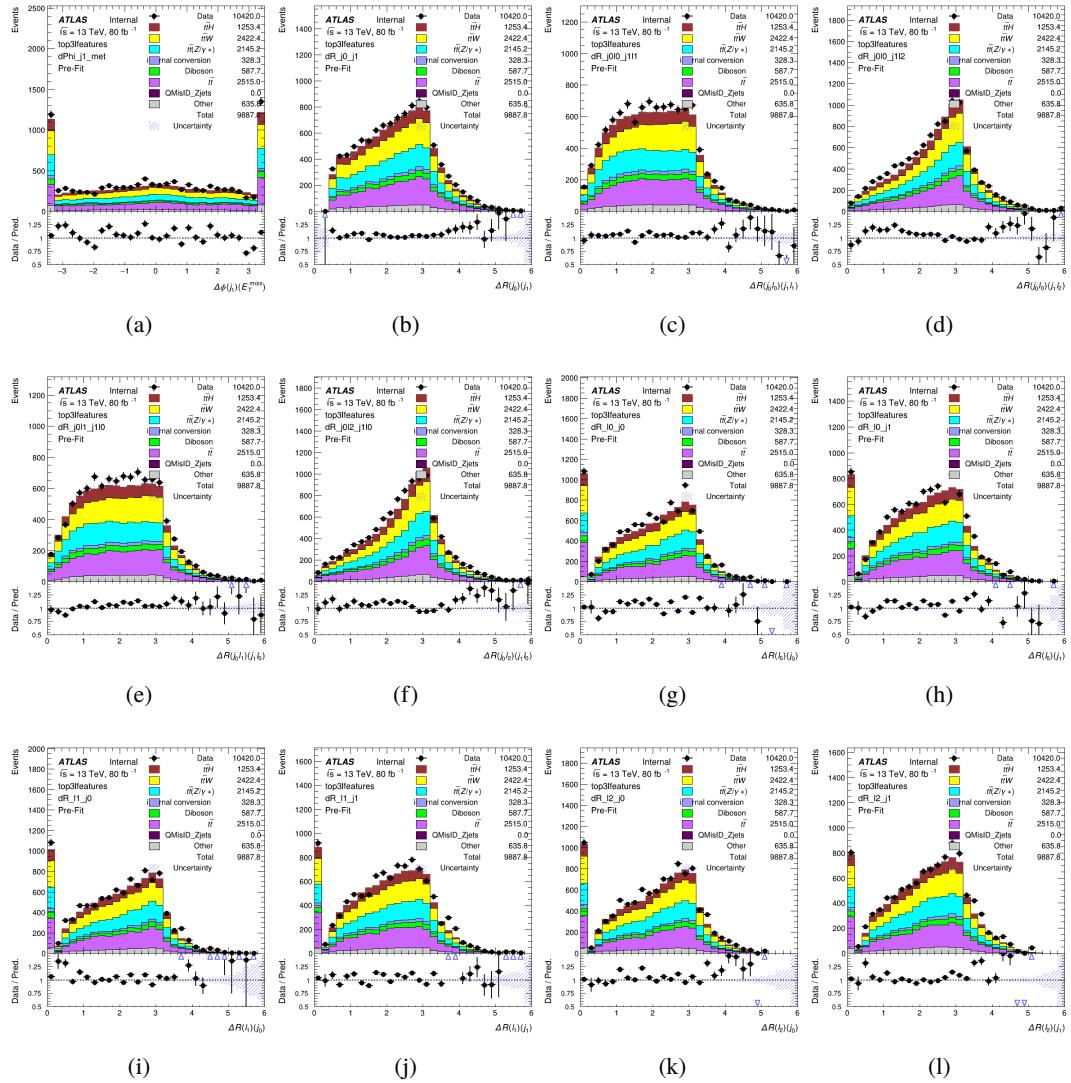


Figure A.5: Input features for top31

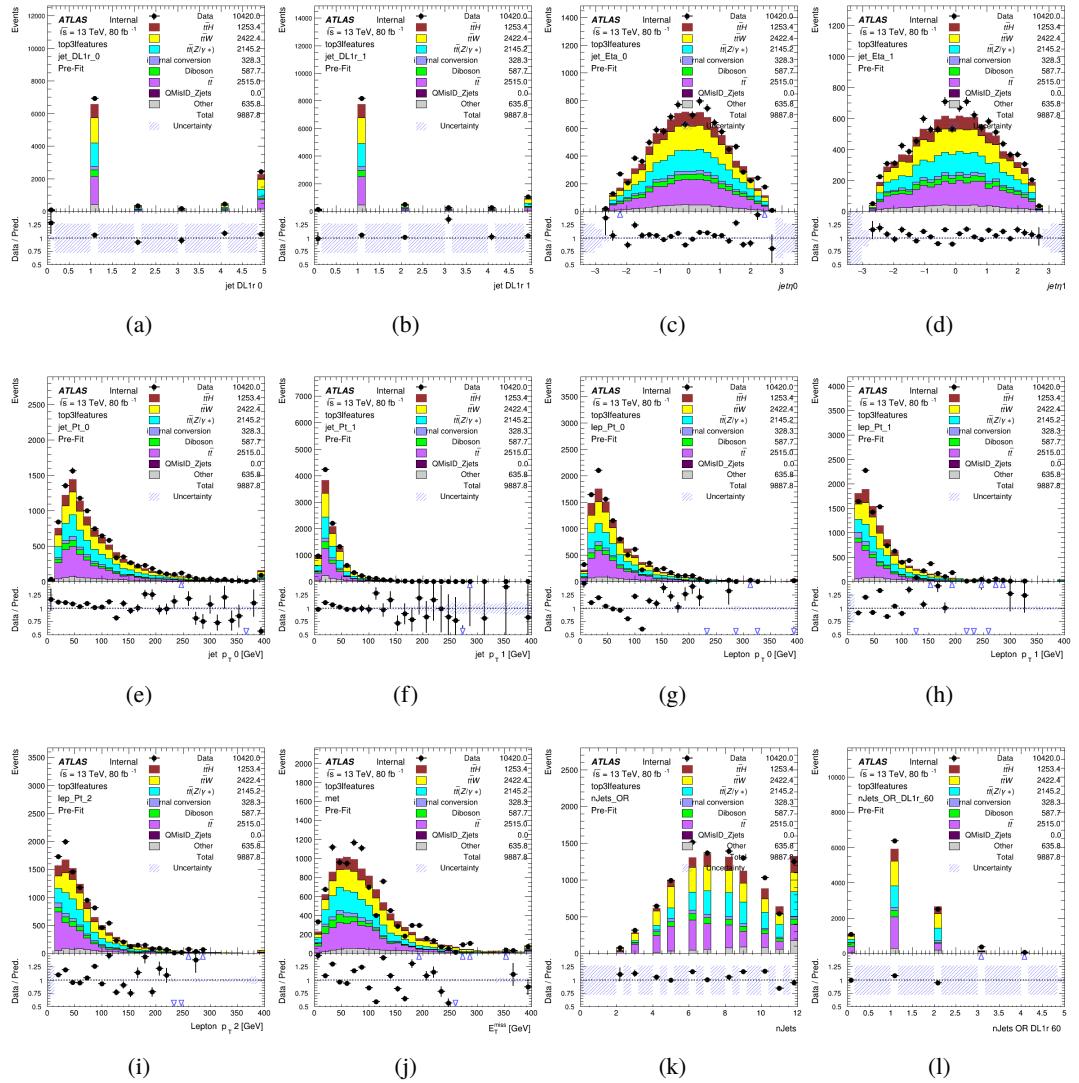
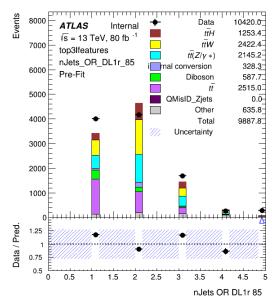


Figure A.6: Input features for top31



(a)

Figure A.7: Input features for top3l

1603 **A.1.3 Higgs Reconstruction Features - 2lSS**

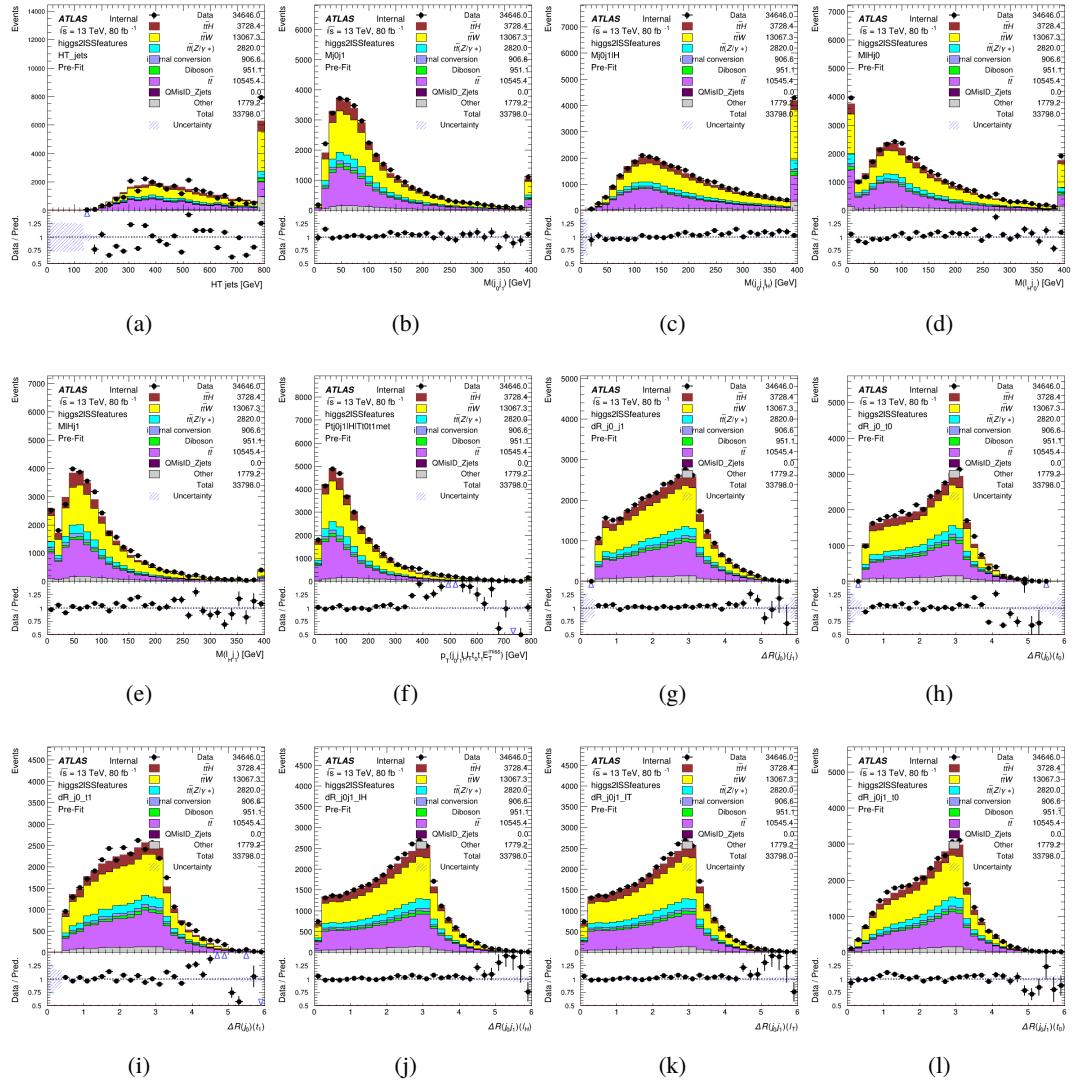


Figure A.8: Input features for higgs2lSS

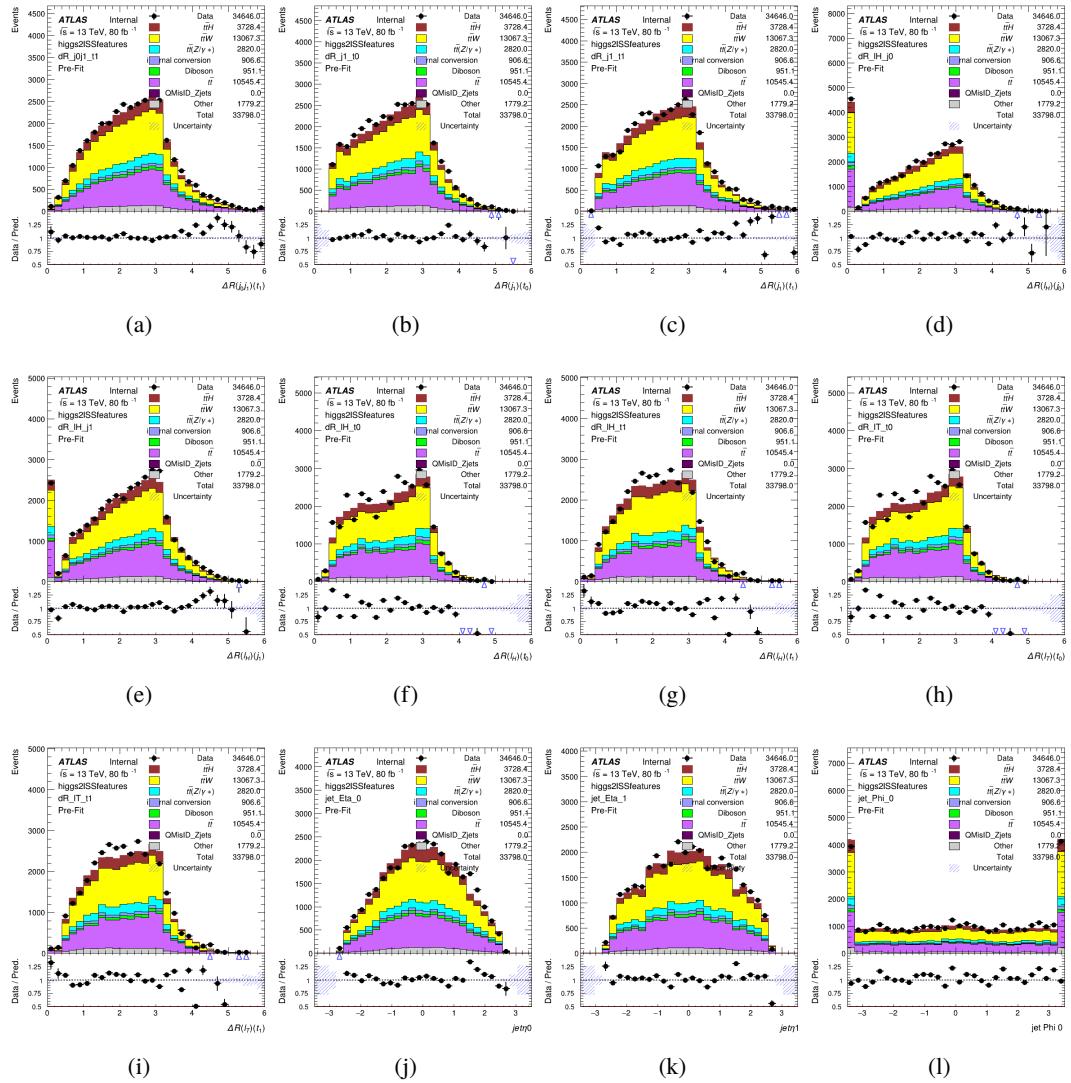


Figure A.9: Input features for higgs2lSS

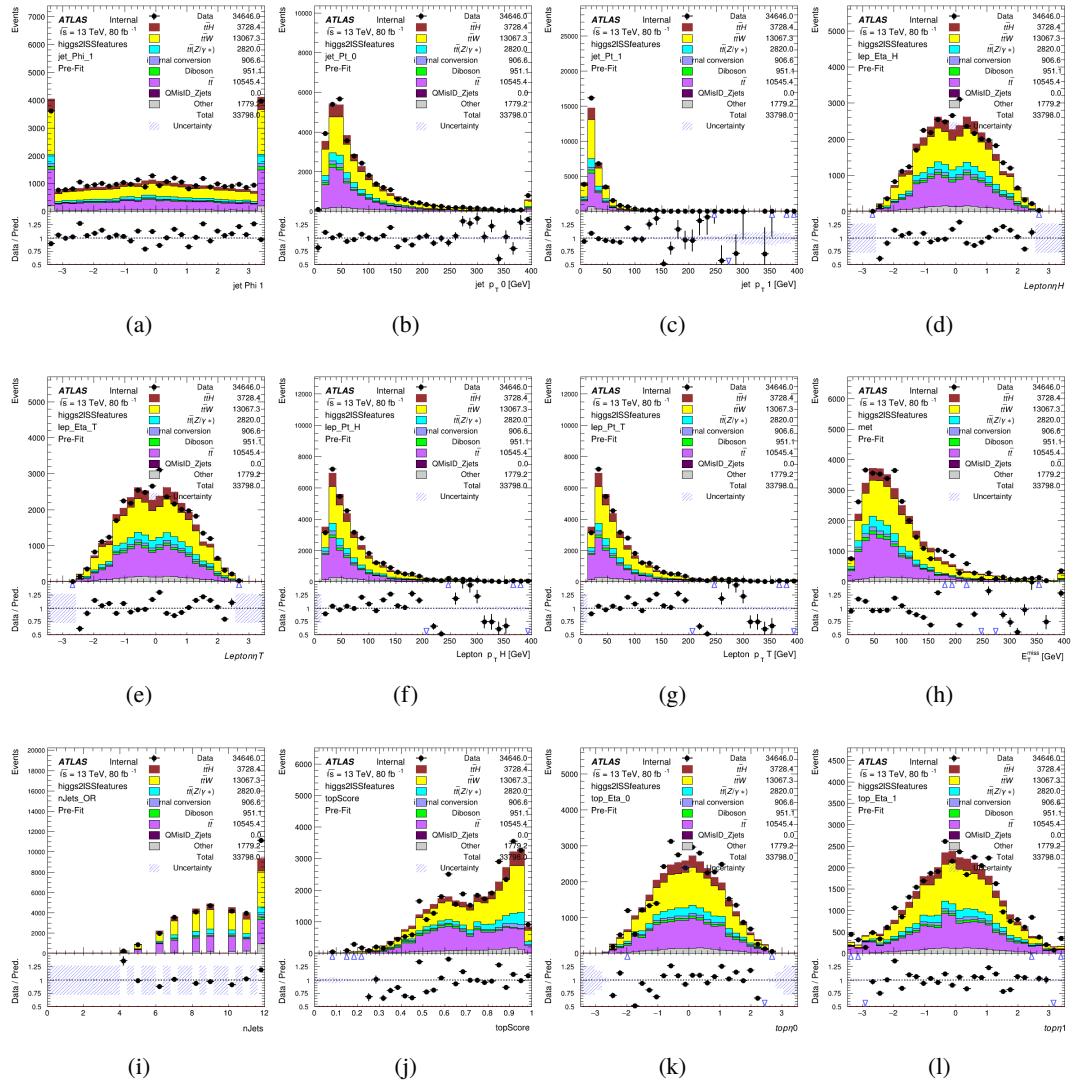


Figure A.10: Input features for higgs2ISS

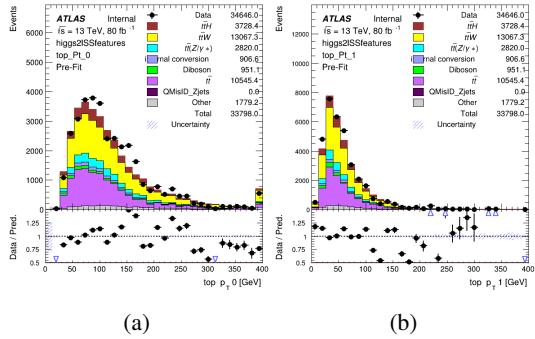


Figure A.11: Input features for higgs2ISS

1604 **A.1.4 Higgs Reconstruction Features - 3lS**

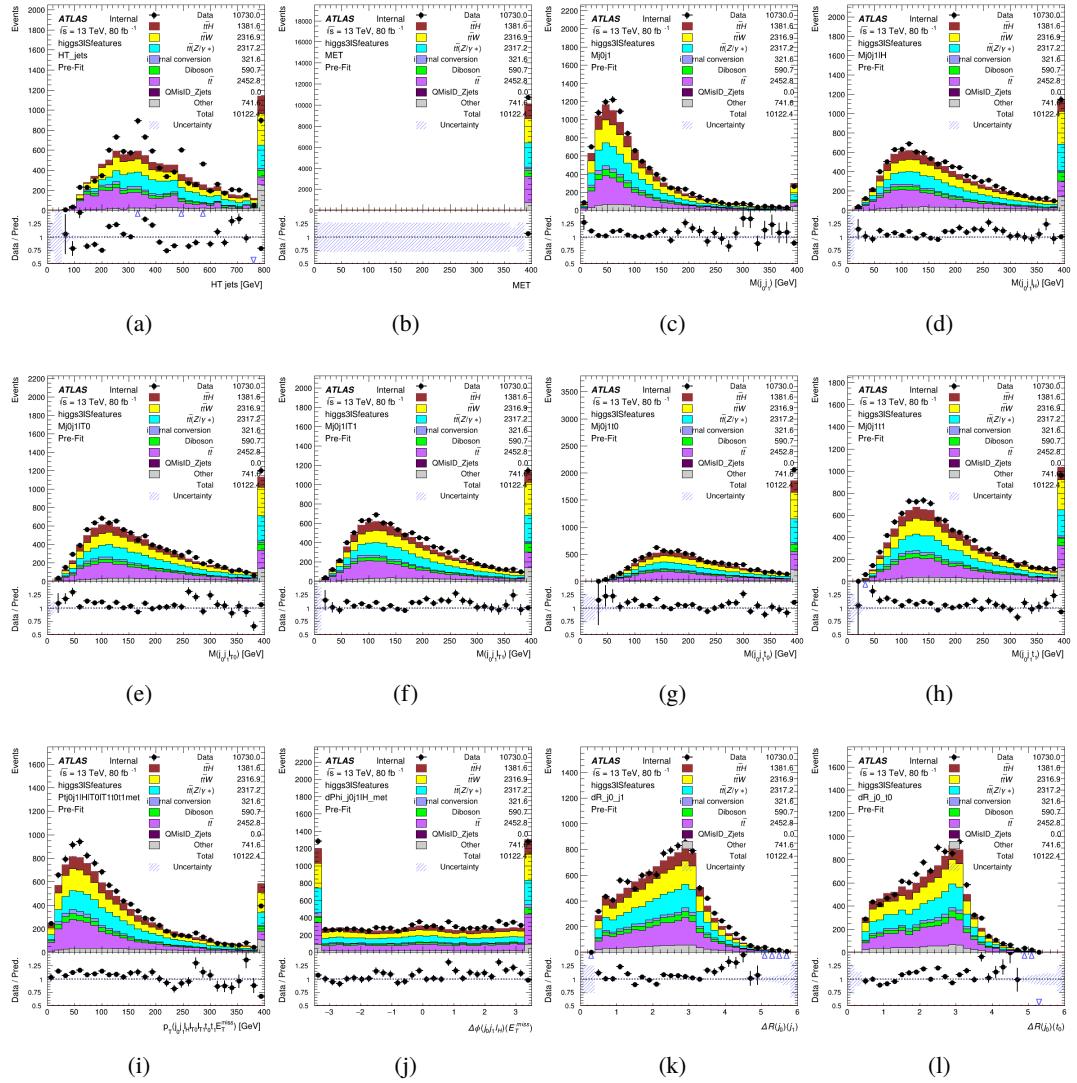


Figure A.12: Input features for higgs31S

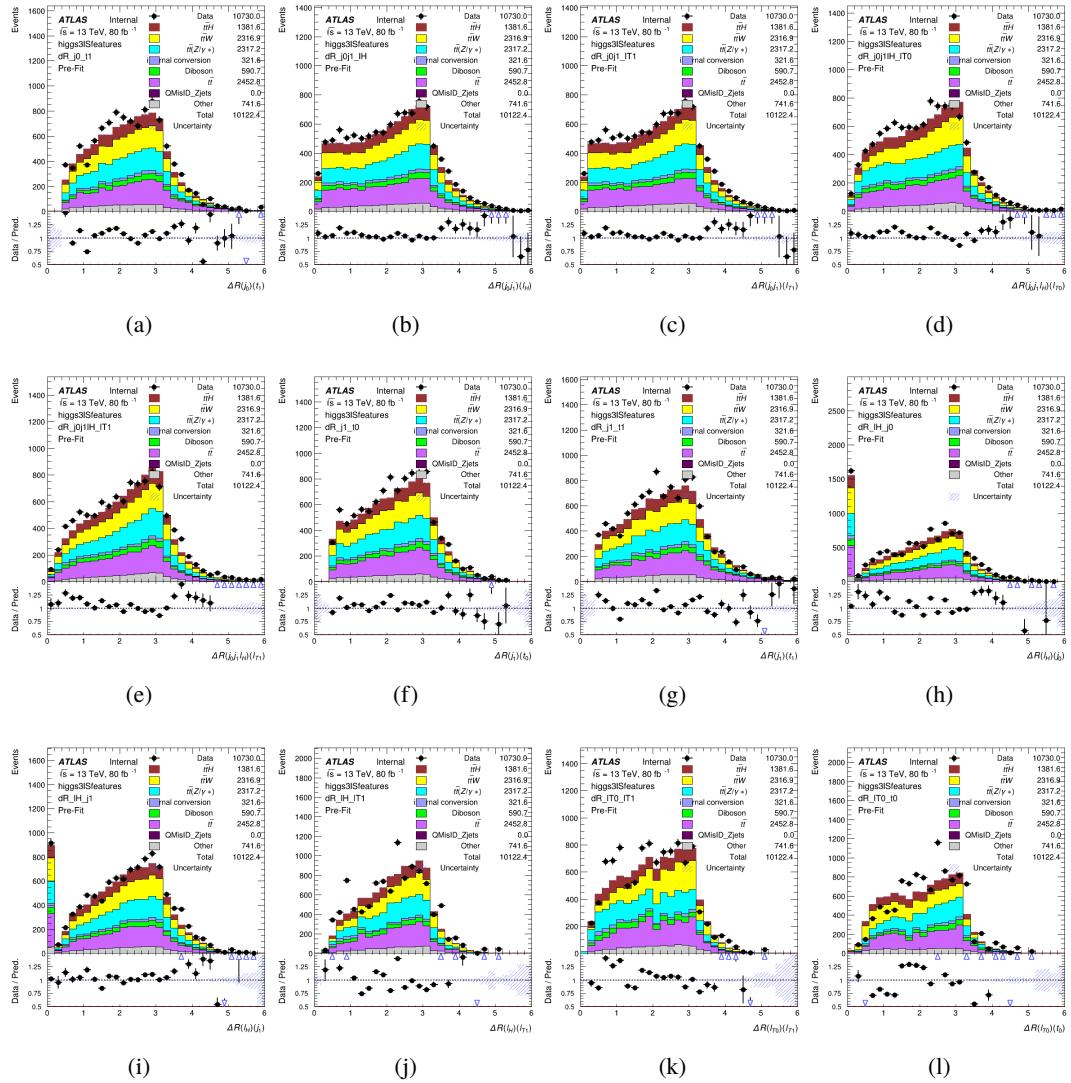


Figure A.13: Input features for higgs3lS

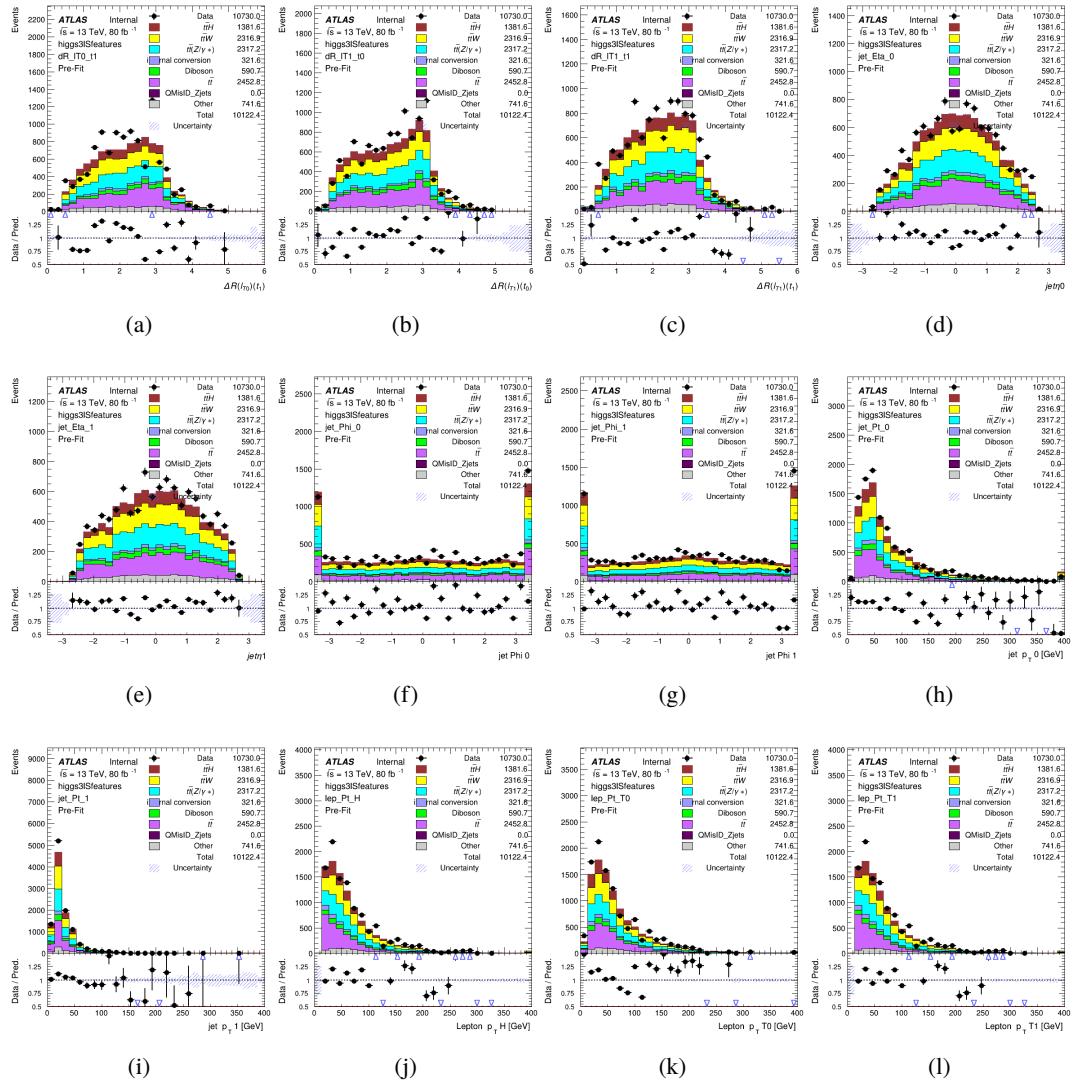


Figure A.14: Input features for higgs3lS

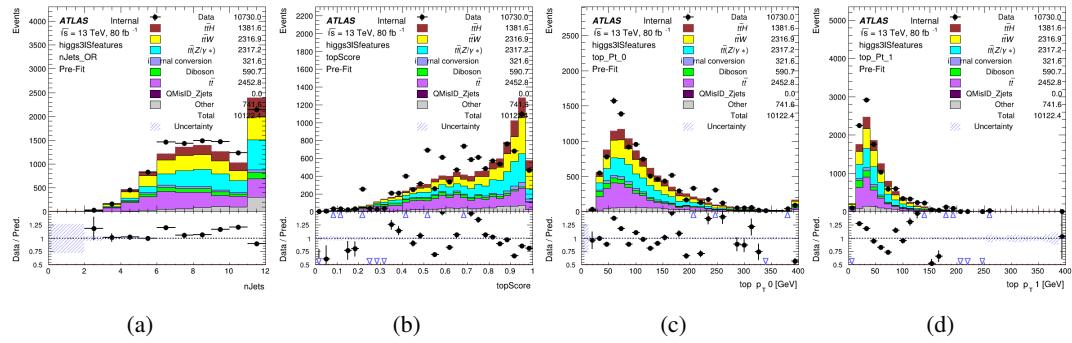


Figure A.15: Input features for higgs3lS

1605 **A.1.5 Higgs Reconstruction Features - 3lF**

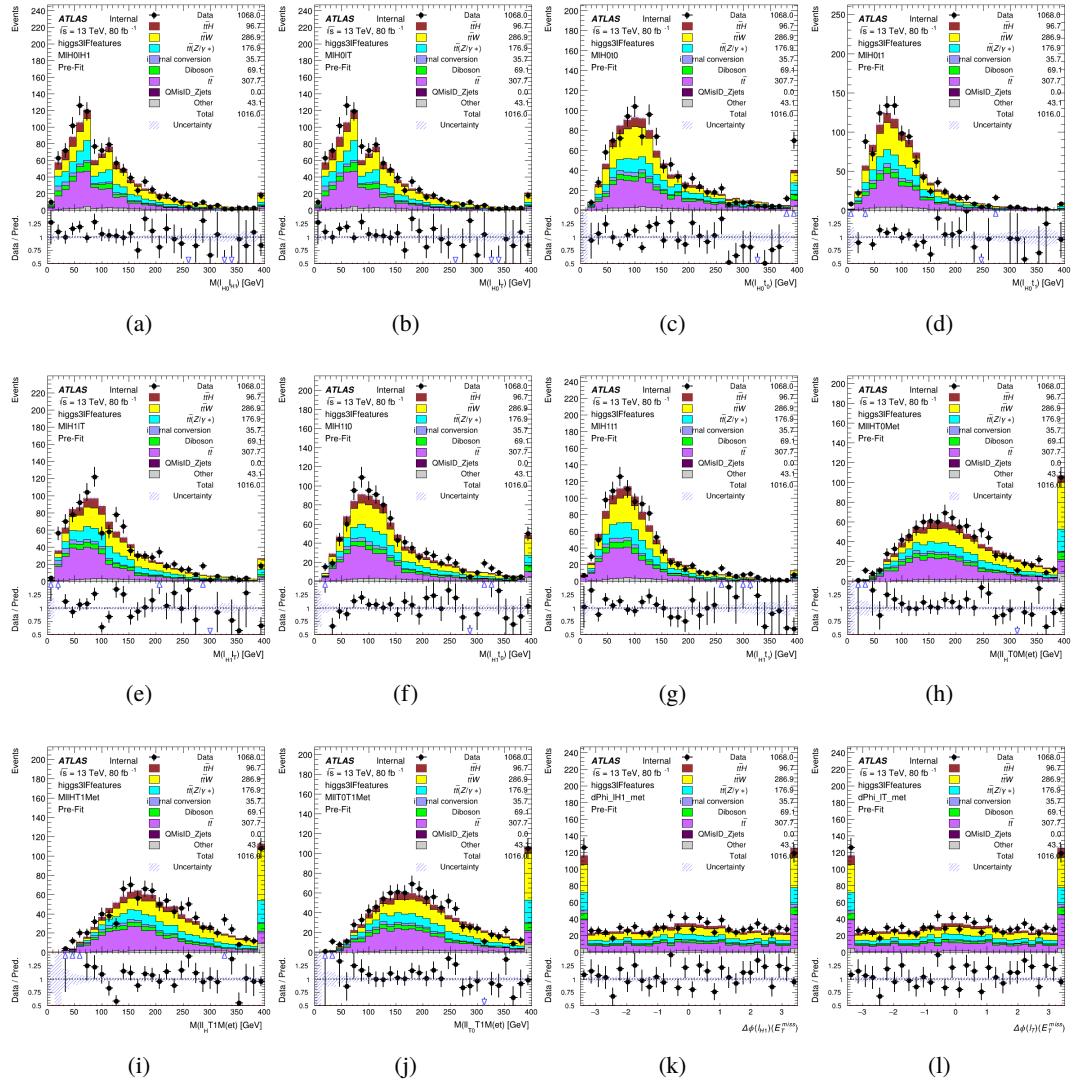


Figure A.16: Input features for higgs3lF

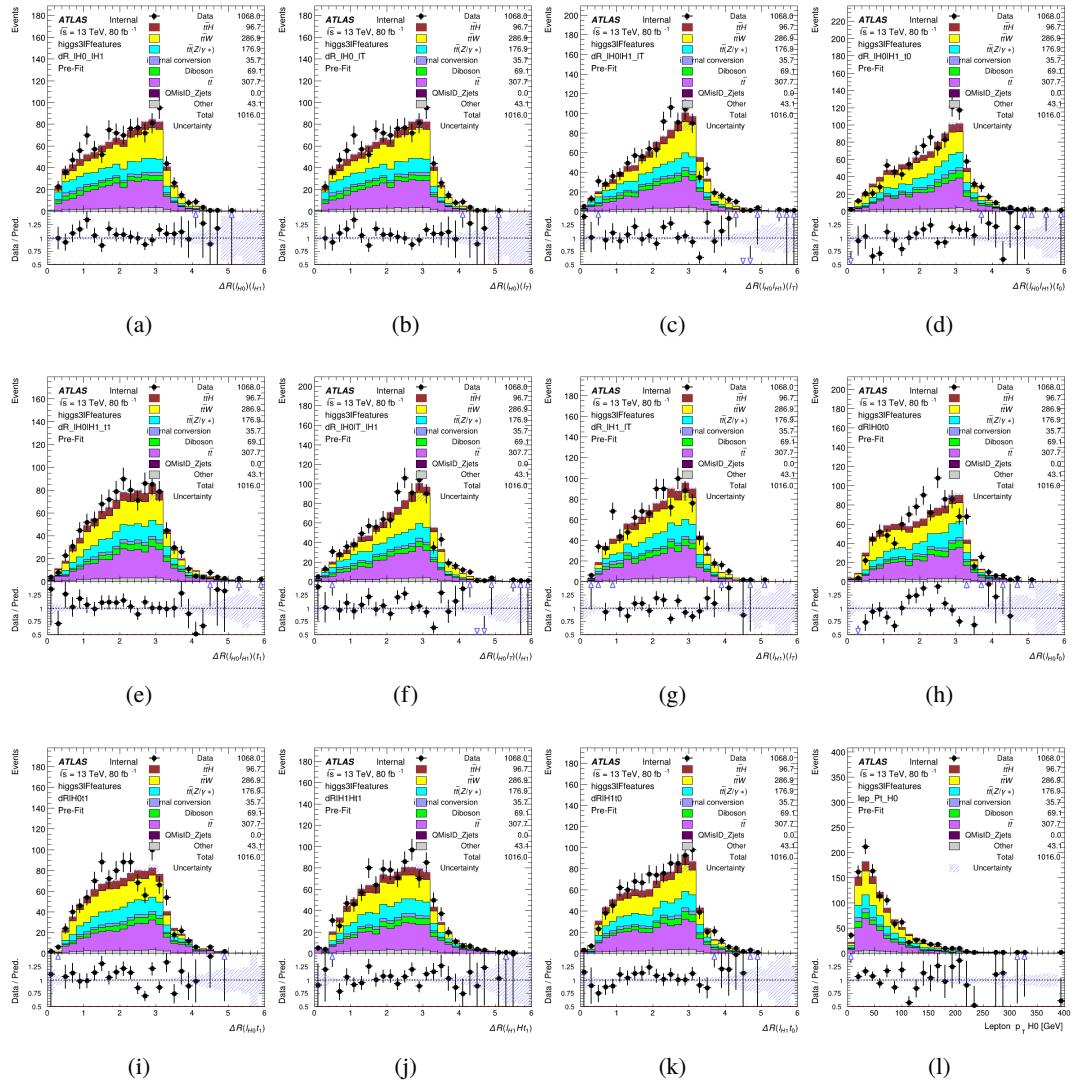


Figure A.17: Input features for higgs3lF

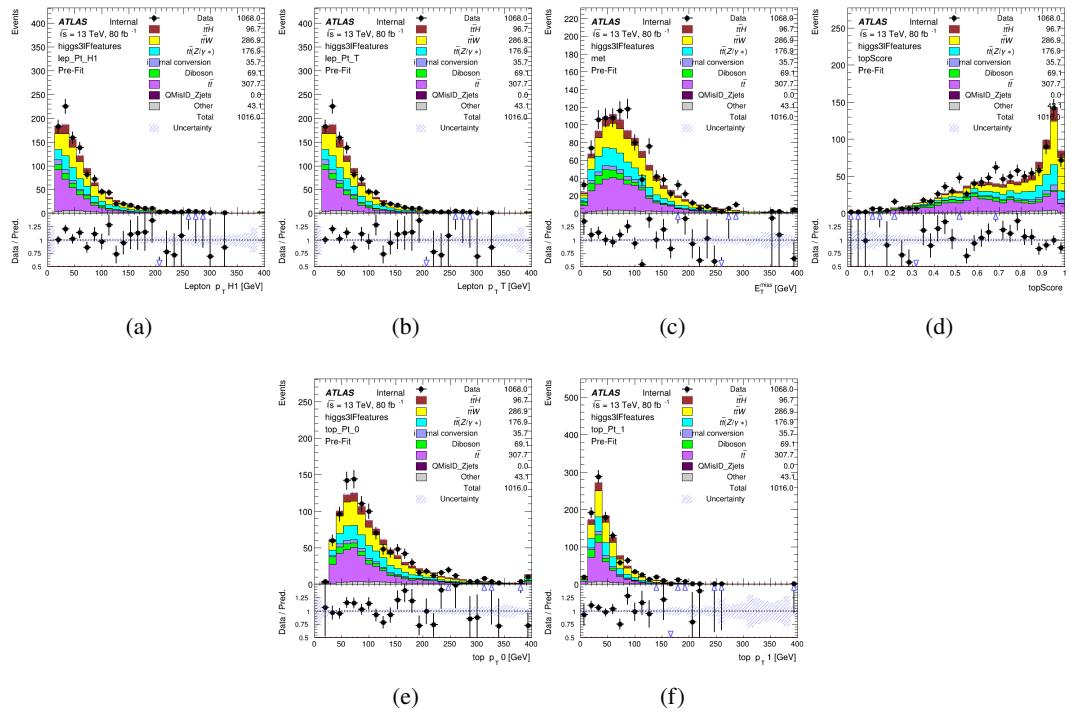


Figure A.18: Input features for higgs3IF

1606 **A.2 Background Rejection MVAs**

1607 **A.2.1 Background Rejection MVA Features - 2lSS**

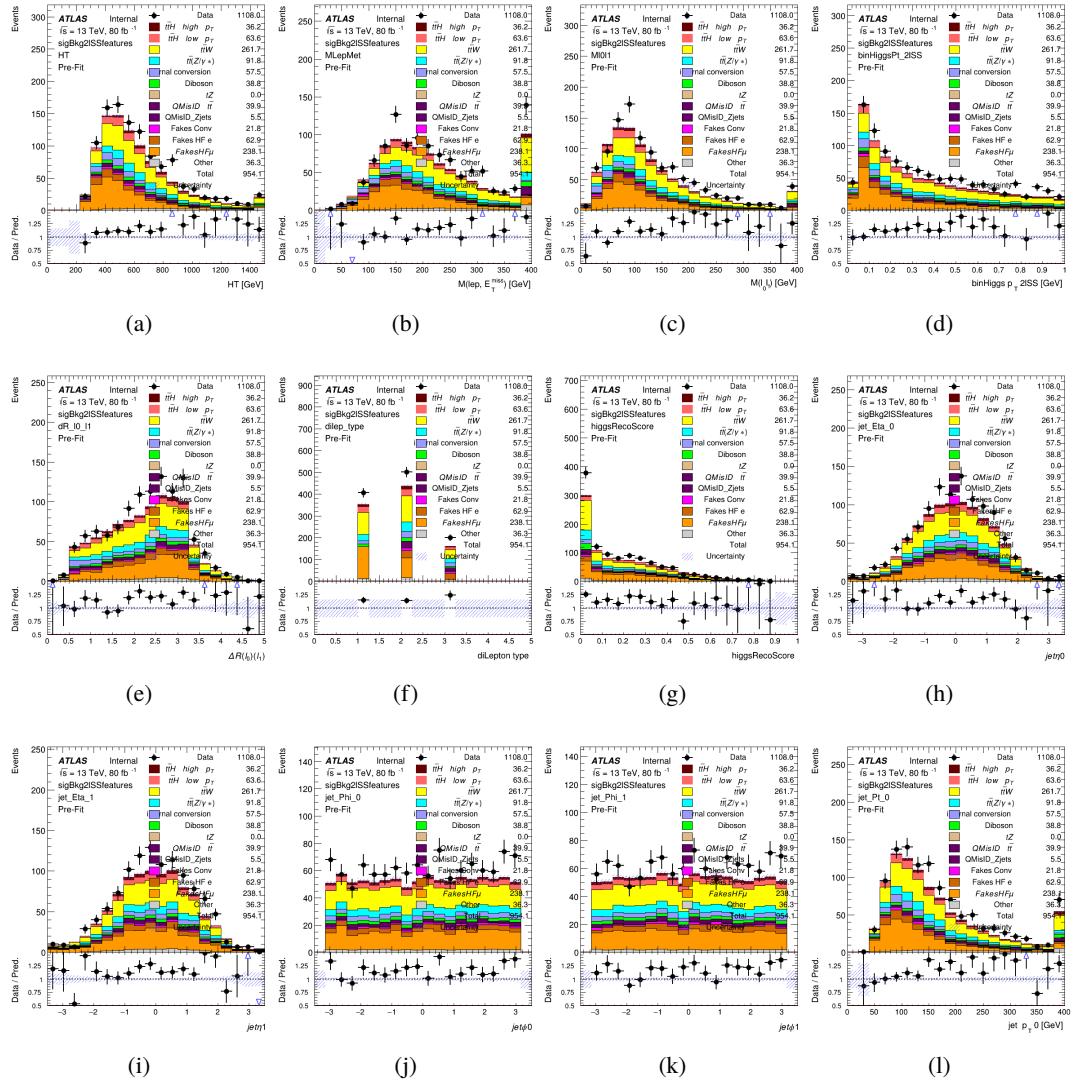


Figure A.19: Input features for sigBkg2lSS

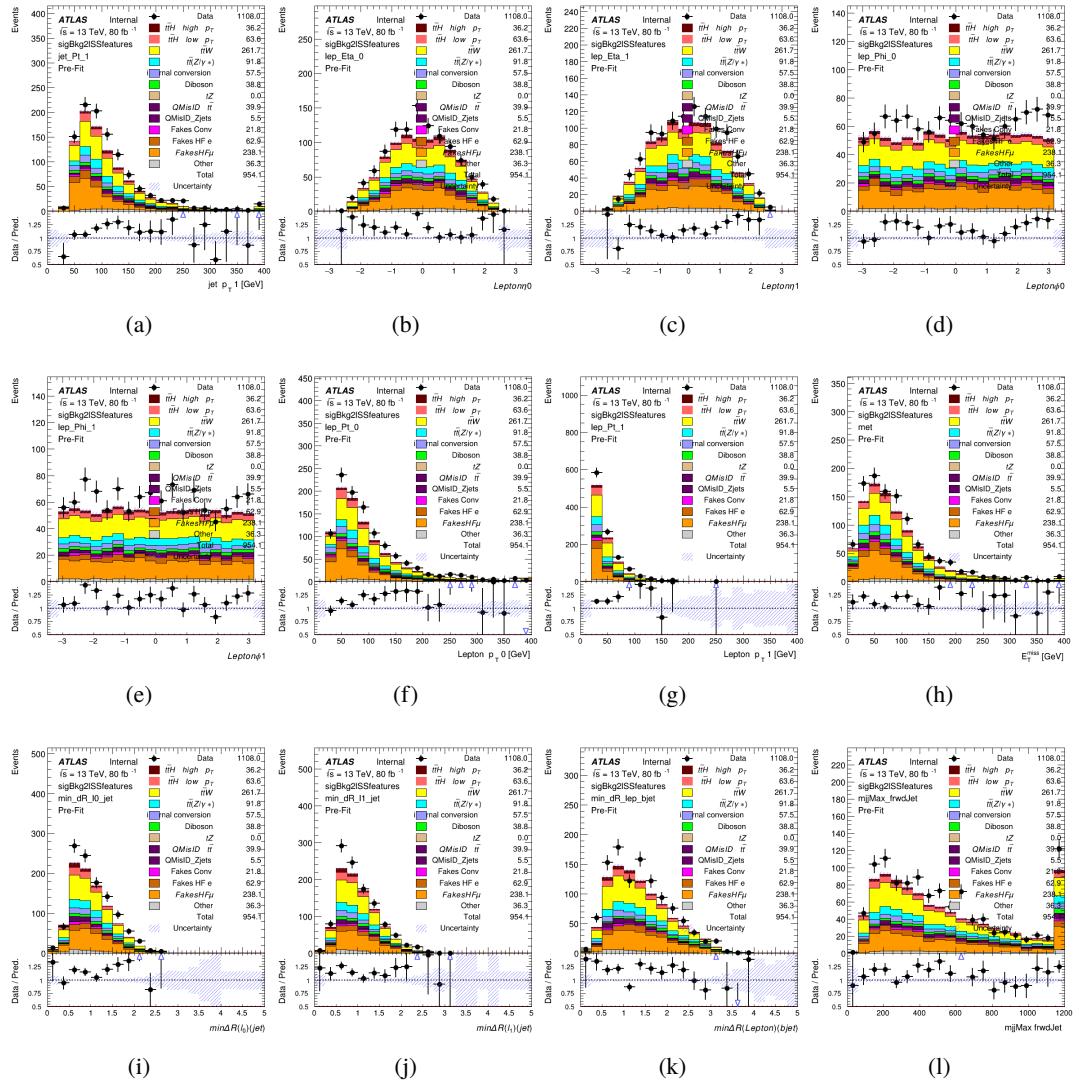


Figure A.20: Input features for sigBkg2lSS

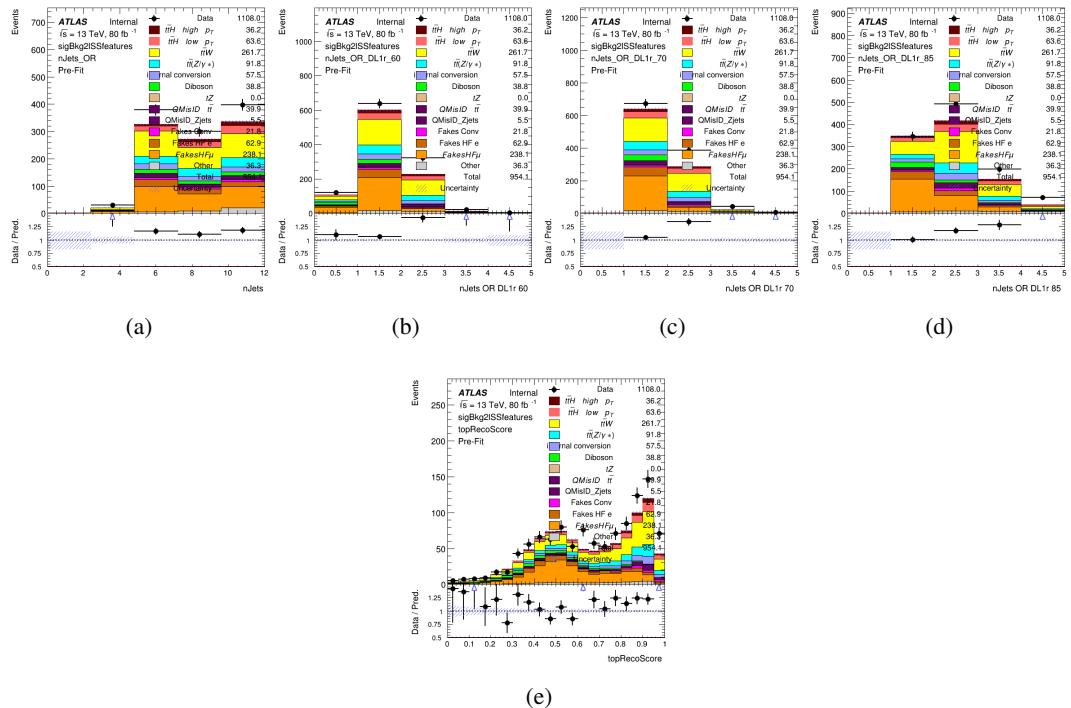
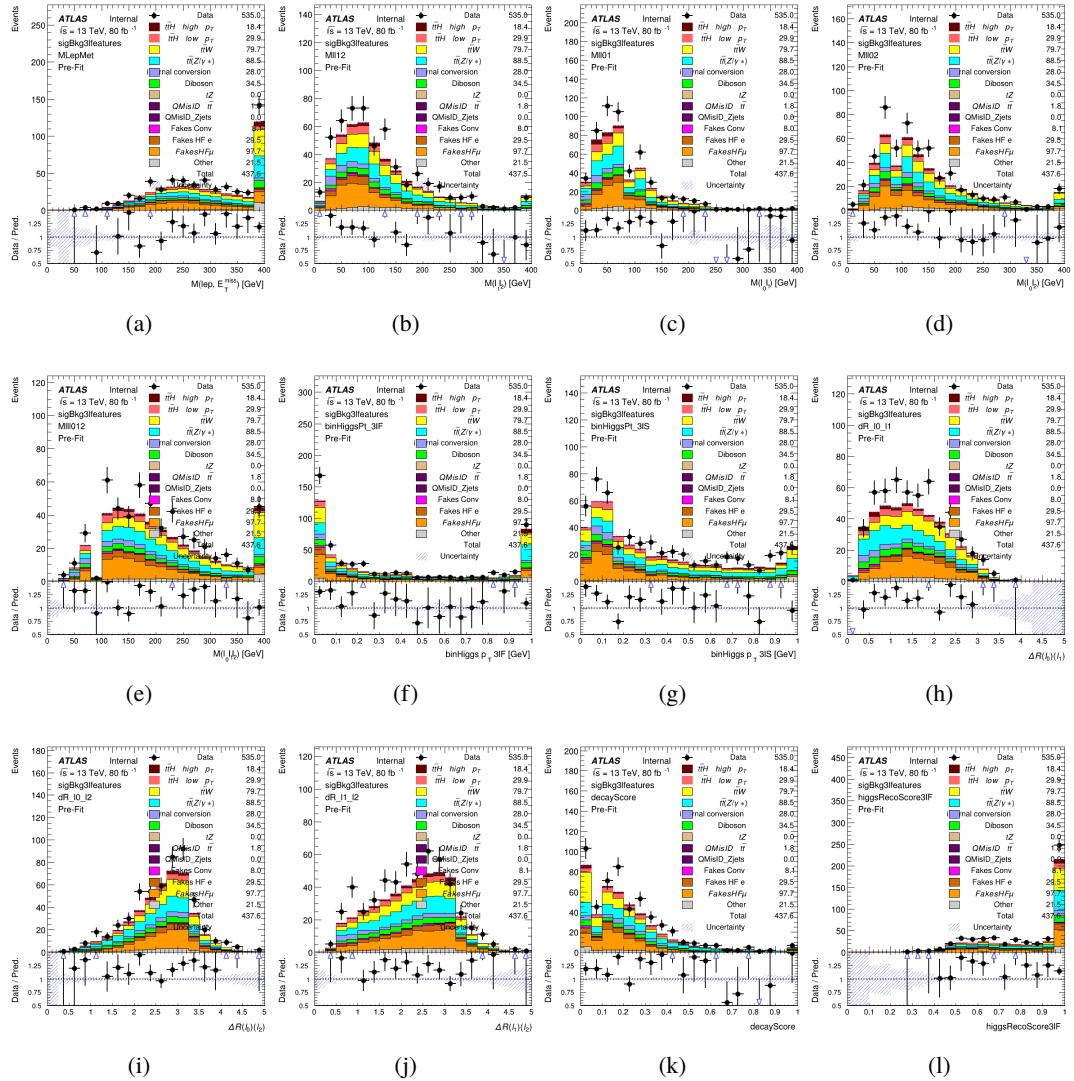


Figure A.21: Input features for sigBkg2ISS

1608 **A.2.2 Background Rejection MVA Features - 3l**

Figure A.22: Input features for sigBkg3l

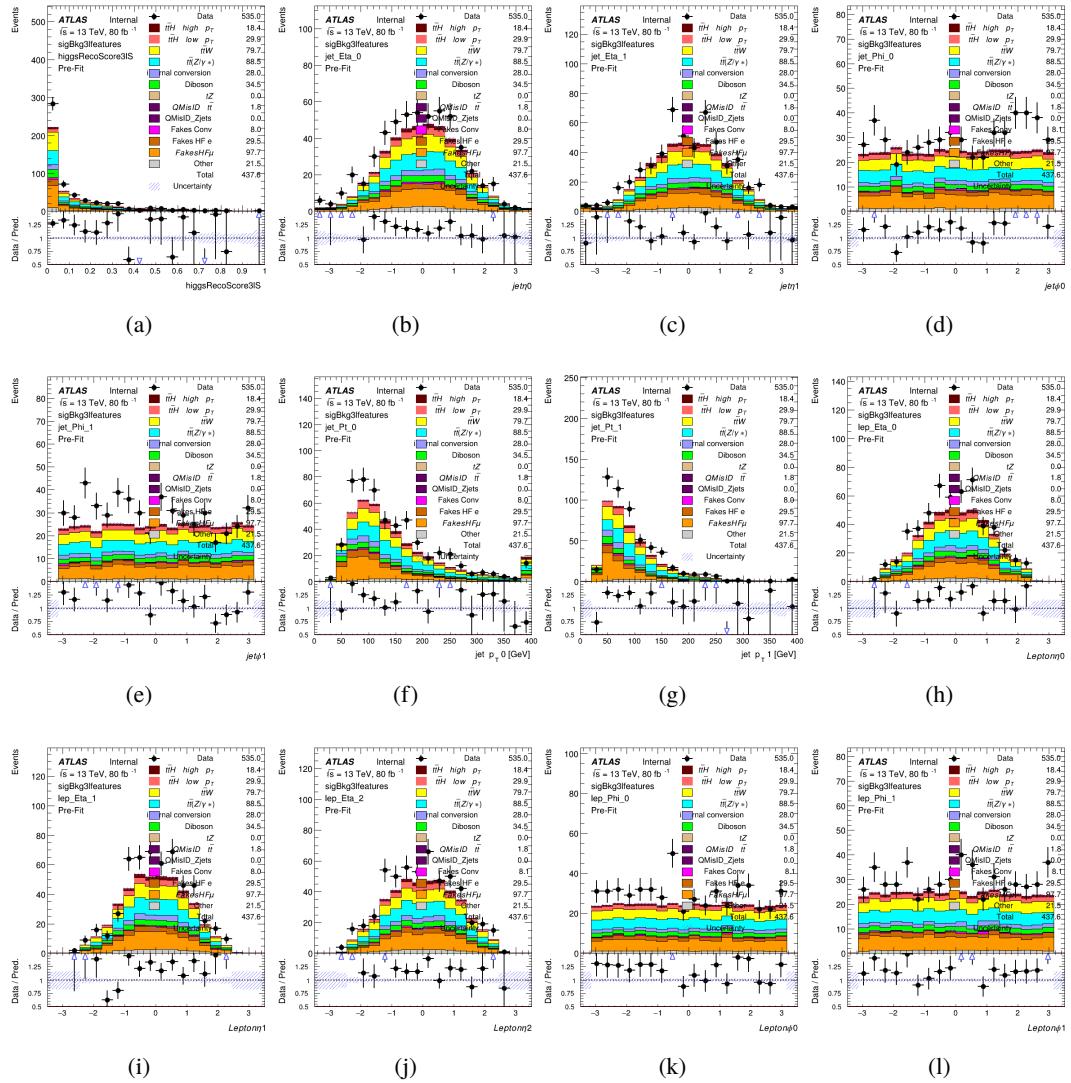


Figure A.23: Input features for sigBkg3l

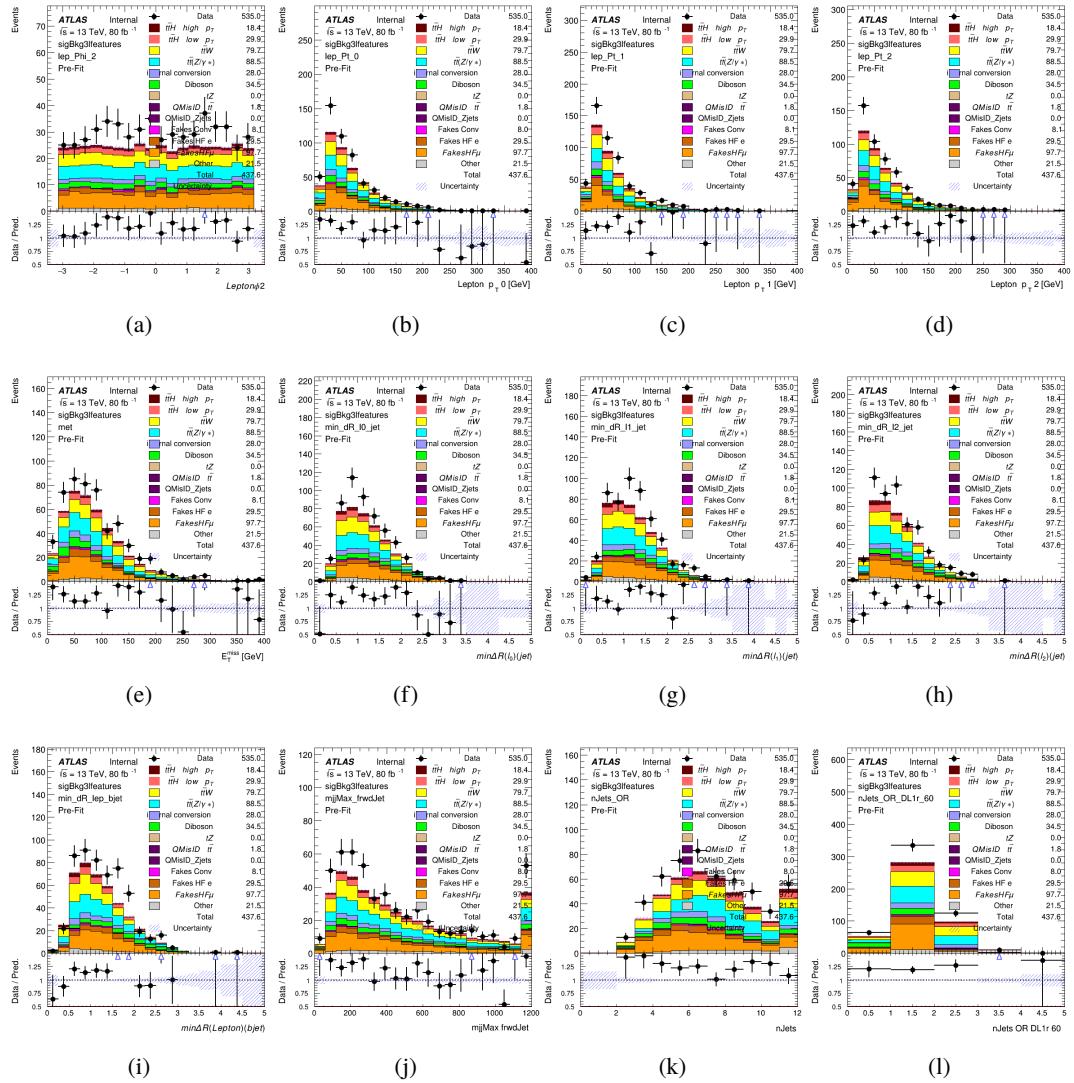


Figure A.24: Input features for sigBkg3l

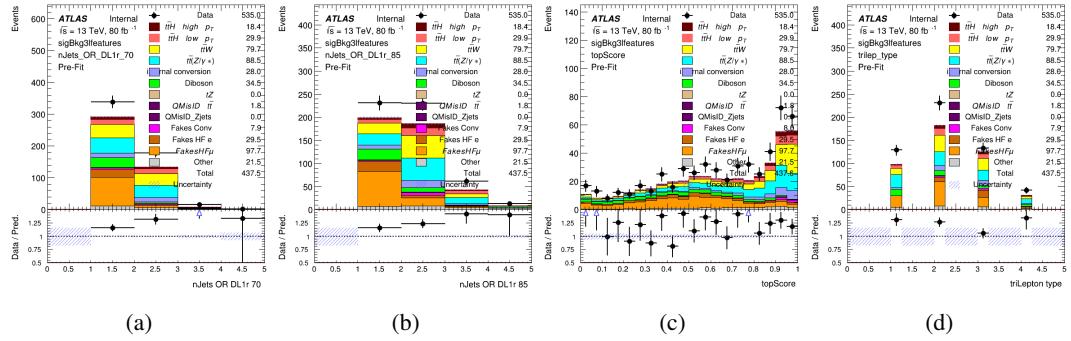


Figure A.25: Input features for sigBkg3l

1609 **A.3 Alternate b-jet Identification Algorithm**

1610 The nominal analysis reconstructs the b-jets by considering different combinations of jets, and
 1611 asking a neural network to determine whether each combination consists of b-jets from top quark
 1612 decays. An alternate approach would be to give the neural network about all of the jets in an event
 1613 at once, and train it to select which two are most likely to be the b-jets from top decay. It was
 1614 hypothesized that this could perform better than considering each combination independently, as
 1615 the neural network could consider the event as a whole. While this is not found to be the case,
 1616 these studies are documented here as a point of interest and comparison.

1617 For these studies, the kinematics of the 10 highest p_T jets in each event are used for
 1618 training. This includes the vast majority of truth b-jets. Specifically the p_T , η , ϕ , E , and DL1r
 1619 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.
 1620 The p_T , η , ϕ , and E of the leptons and E_T^{miss} are included as well. Categorical cross entropy is
 1621 used as the loss function.

Table 52: 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%

1622 **A.4 Binary Classification of the Higgs p_T**

1623 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer
 1624 resolution. This means separating high and low p_T events is sufficient for this analysis. As

1625 such, rather than attempting the reconstruct the full Higgs p_T spectrum, a binary classification
 1626 approach is explored.

1627 A model is built to determine whether $t\bar{t}H$ events include a high p_T (>150 GeV) or low
 1628 p_T (<150 GeV) Higgs Boson. While this is now a classification model, it uses the same input
 1629 features described in section 19.4. Binary crossentropy is used as the loss function.

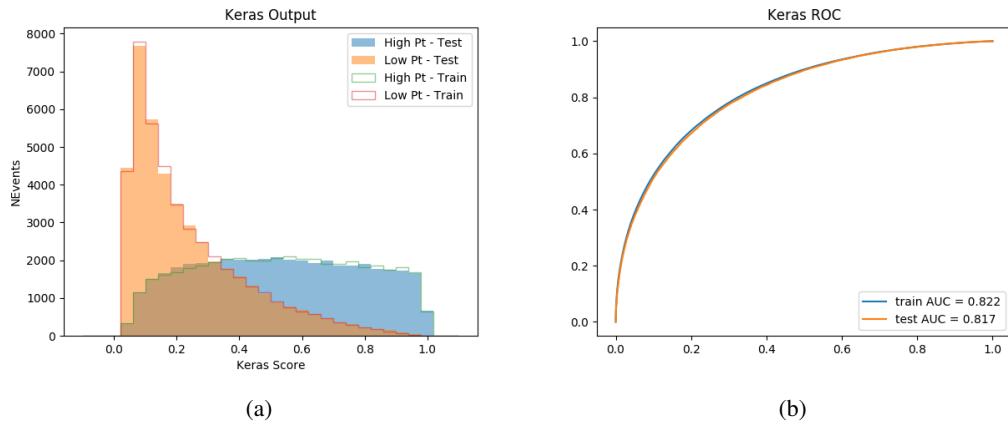


Figure A.26:

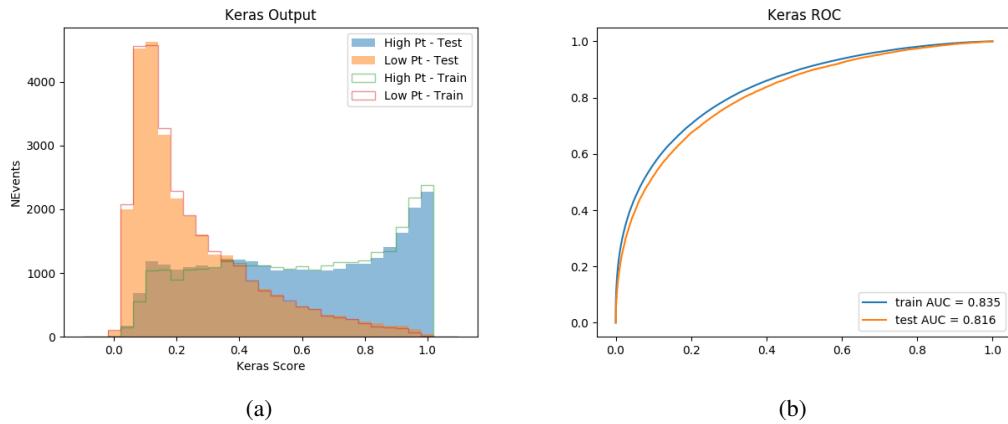


Figure A.27:

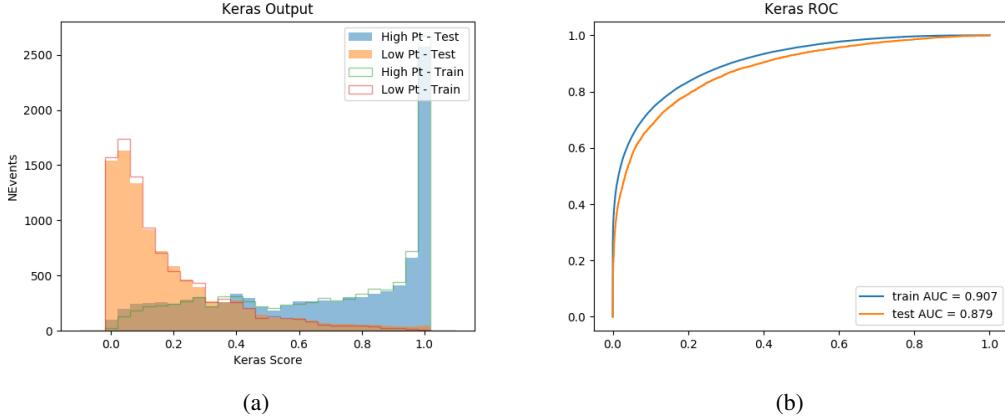


Figure A.28:

1630 A.5 Impact of Alternative Jet Selection

1631 A relatively low p_T threshold of 15 GeV is used to determine jet candidates, as the jets originating
 1632 from the Higgs decay are found to fall between 15 and 25 GeV a large fraction of the time. The
 1633 impact of different jet p_T cuts on our ability to reconstruct the Higgs p_T is explored here. The
 1634 performance of the Higgs p_T prediction models is evaluated for jet p_T cuts of 10, 15, 20, and 25
 1635 GeV.

1636 **B**

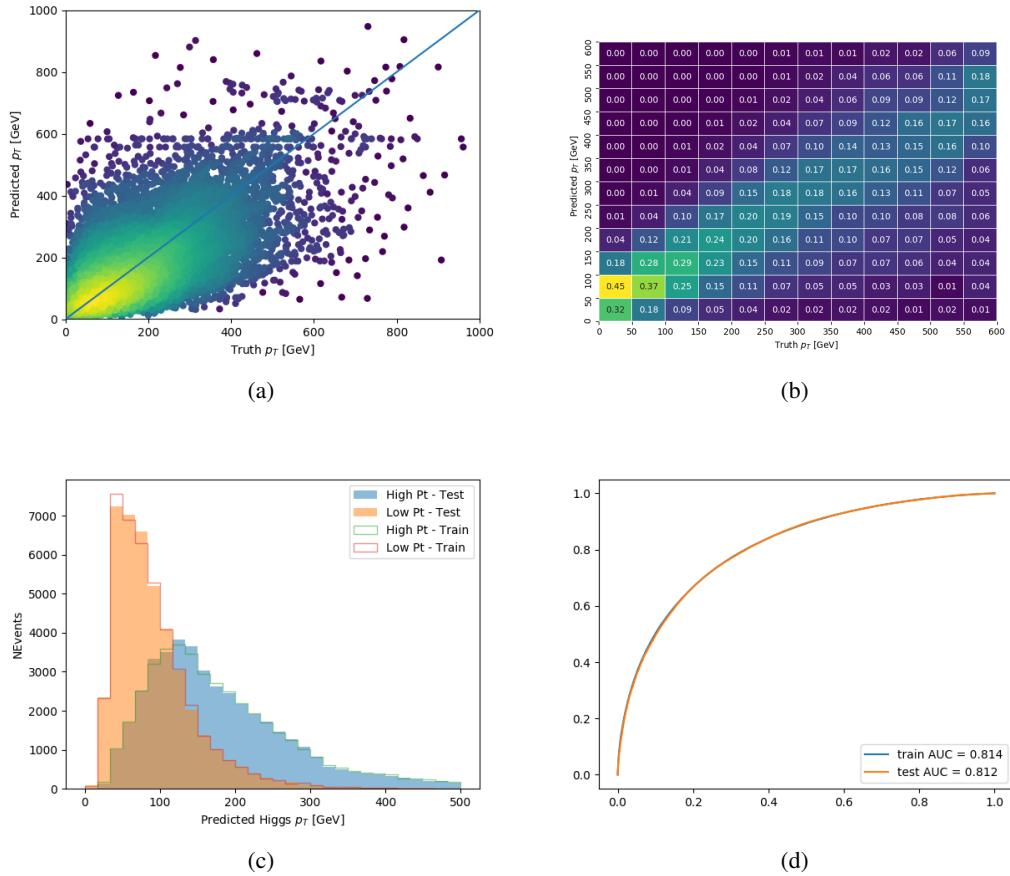


Figure A.29: