



ATLAS Note

GROUP-2017-XX



25th 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	7
18	1 Introduction	7
19	II Theoretical Motivation	10
20	2 The Standard Model and the Higgs Boson	10
21	2.1 The Forces and Particles of the Standard Model	10
22	2.2 The Higgs Mechanism	13
23	2.2.1 The Higgs Field	13
24	2.2.2 Electroweak Symmetry Breaking	15
25	2.3 $t\bar{t}H$ Production	18
26	2.4 Extensions to the Standard Model	20
27	III The LHC and the ATLAS Detector	23
28	3 The LHC	23
29	4 The ATLAS Detector	26
30	4.1 Inner Detector	28
31	4.2 Calorimeters	29
32	4.3 Muon Spectrometer	30
33	4.4 Trigger System	31
34	IV Measurement of WZ + Heavy Flavor	32
35	5 Introduction	32
36	6 Data and Monte Carlo Samples	34
37	6.1 Data Samples	34
38	6.2 Monte Carlo Samples	35
39	7 Object Reconstruction	37
40	7.1 Trigger	38
41	7.2 Light leptons	38
42	7.3 Jets	40
43	7.4 B-tagged Jets	40
44	7.5 Missing transverse energy	42
45	7.6 Overlap removal	42

46	8 Event Selection and Signal Region Definitions	43
47	8.1 Event Preselection	43
48	8.2 Fit Regions	48
49	8.3 Non-Prompt Lepton Estimation	65
50	8.3.1 $t\bar{t}$ Validation	65
51	8.3.2 Z+jets Validation	71
52	9 tZ Interference Studies and Separation Multivariate Analysis	76
53	9.1 Top Mass Reconstruction	76
54	9.2 tZ BDT	78
55	10 Data and Monte Carlo Samples	82
56	10.1 Data Samples	83
57	10.2 Monte Carlo Samples	84
58	11 Object Reconstruction	86
59	11.1 Trigger	87
60	11.2 Light leptons	87
61	11.3 Jets	89
62	11.4 B-tagged Jets	89
63	11.5 Missing transverse energy	91
64	11.6 Overlap removal	91
65	12 Event Selection and Signal Region Definitions	92
66	12.1 Event Preselection	92
67	12.2 Fit Regions	97
68	12.3 Non-Prompt Lepton Estimation	114
69	12.3.1 $t\bar{t}$ Validation	114
70	12.3.2 Z+jets Validation	120
71	13 tZ Interference Studies and Separation Multivariate Analysis	125
72	13.1 Top Mass Reconstruction	125
73	13.2 tZ BDT	127
74	14 Systematic Uncertainties	131
75	15 Results	140
76	15.1 1-jet Fit Results	141
77	15.2 2-jet Fit Results	150
78	V Differential Studies of $t\bar{t}H$ Multilepton	159
79	16 Data and Monte Carlo Samples	159
80	16.1 Data Samples	160

81	16.2 Monte Carlo Samples	160
82	17 Object Reconstruction	161
83	17.1 Trigger Requirements	161
84	17.2 Light Leptons	162
85	17.3 Jets	163
86	17.4 Missing Transverse Energy	163
87	18 Higgs Momentum Reconstruction	164
88	18.1 Decay Candidate Reconstruction	165
89	18.2 b-jet Identification	166
90	18.2.1 2lSS Channel	167
91	18.2.2 3l Channel	171
92	18.3 Higgs Reconstruction	174
93	18.3.1 2lSS Channel	175
94	18.3.2 3l Semi-leptonic Channel	179
95	18.3.3 3l Fully-leptonic Channel	183
96	18.4 p_T Prediction	186
97	18.4.1 2lSS Channel	188
98	18.4.2 3l Semi-leptonic Channel	191
99	18.4.3 3l Fully-leptonic Channel	193
100	18.5 3l Decay Mode	197
101	19 Signal Region Definitions	199
102	19.1 Pre-MVA Event Selection	200
103	19.2 Event MVA	205
104	19.3 Signal Region Definitions	211
105	20 Background Rejection MVA	215
106	20.1 Pre-MVA Event Selection	216
107	20.2 Event MVA	217
108	20.3 Signal Region Definitions	223
109	20.3.1 2lSS	226
110	20.3.2 3l – Semi – leptonic	226
111	20.3.3 3l – Fully – leptonic	226
112	21 Systematic Uncertainties	226
113	22 Results	228
114	22.1 Results - 80 fb^{-1}	228
115	22.2 Projected Results - 140 fb^{-1}	231

116	VI Conclusion	233
117	Appendices	235
118	A Machine Learning Models	235
119	A.1 Higgs Reconstruction Models	235
120	A.1.1 b-jet Identification Features - 2lSS	235
121	A.1.2 b-jet Identification Features - 3l	239
122	A.1.3 Higgs Reconstruction Features - 2lSS	244
123	A.1.4 Higgs Reconstruction Features - 3lS	249
124	A.1.5 Higgs Reconstruction Features - 3lF	254
125	A.2 Background Rejection MVAs	258
126	A.2.1 Background Rejection MVA Features - 2lSS	258
127	A.2.2 Background Rejection MVA Features - 3l	262
128	A.3 Alternate b-jet Identification Algorithm	267
129	A.4 Binary Classification of the Higgs p_T	267
130	A.5 Impact of Alternative Jet Selection	269

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

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

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

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

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

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

161 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-
162 2018 provides the opportunity to make this measurement for the first time. The unprecedented
163 energy achieved by the LHC during this period greatly increase the rate at which $t\bar{t}H$ events are
164 produced, and the large amount of data collected provides the necessary statistics for a differential
165 measurement to be performed.

166 A study of $t\bar{t}H$ events with multiple leptons in the final state is performed, using 139 fb^{-1}
167 of data from proton-proton collisions at an energy $\sqrt{s} = 13 \text{ TeV}$ collected by the ATLAS detector
168 from 2015-2018. Events are separated into channels based on the number of light leptons in the

169 final state - either two same-sign leptons, or three leptons. A deep neural network is used to
170 reconstruct the momentum of the Higgs Boson in each event. This momentum spectrum is fit to
171 data for each analysis channel, the result of which is used to place limits on BSM effects.

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

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

Part II**Theoretical Motivation****2 The Standard Model and the Higgs Boson**

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

2.1 The Forces and Particles of the Standard Model

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

Standard Model of Elementary Particles

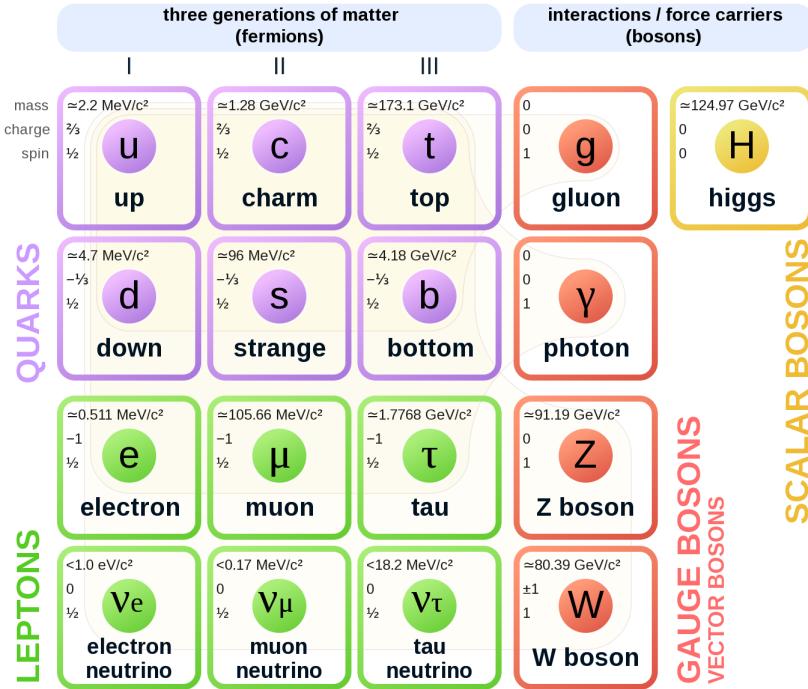


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. []

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

200 Leptons are fermions interact via the electroweak force, but not the strong force. The three
 201 generation of leptons consist of the electron and electron neutrino, the muon and muon neutrino,
 202 the tau and tau neutrino. The quarks, which do interact via the strong force - which is to say they
 203 have color charge - in addition to the electroweak force. The three generations include the up

204 and down quarks, the strange and charm quarks, and the top and bottom quarks. Each of these
 205 generations form left-handed doublets invariant under SU(2) transformations. For the leptons
 206 these doublets are:

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

207 And for the quarks:

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

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

214 Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-
 215 exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1
 216 particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0,
 217 particle - the Higgs Boson. Of the gauge bosons, the W^+ , W^- and Z bosons - which are the

²¹⁸ mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon
²¹⁹ mediates the electric force, and the gluon mediates the strong force.

²²⁰ **2.2 The Higgs Mechanism**

²²¹ A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to
²²² the Lagrangian giving mass to the the gauge bosons would violate the underlying symmetry of
²²³ the theory. This presents a clear problem with the theory: The experimental observation that the
²²⁴ W and Z bosons have mass seems to contradict the basic structure of the SM.

²²⁵ Rather than abandoning gauge invariance, an alternative way for particles to acquire mass
²²⁶ beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout
²²⁷ in 1964 []. This procedure for introducing masses for the gauge bosons while preserving local
²²⁸ gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory
²²⁹ by Weinberg in 1967 [].

²³⁰ **2.2.1 The Higgs Field**

²³¹ The Higgs mechanism introduces a complex scalar SU(2) doublet, Φ , with the form:

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

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

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

233 Where μ and λ are free parameters of the new field. This represents the most general
234 potential allowed while preserving $SU(2)_L$ invariance and renormalizability. In the case that
235 $\mu^2 < 0$, this potential takes the form shown in figure 2.2.

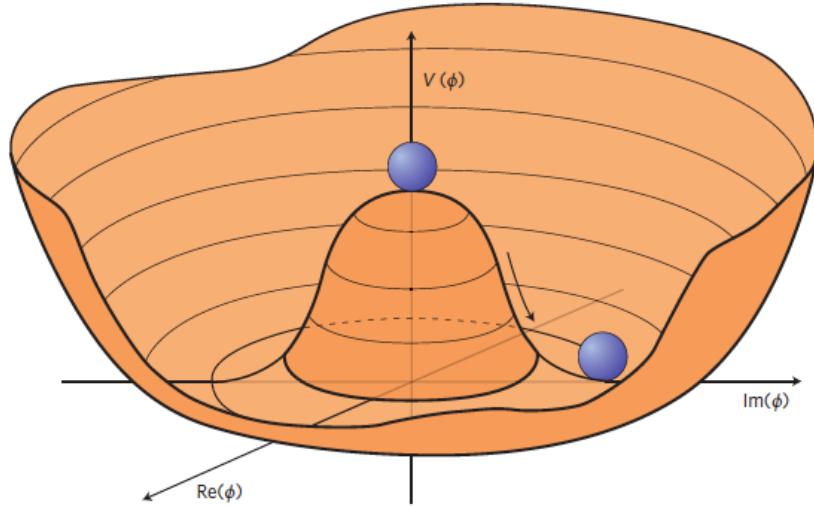


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ , for the case that $\mu^2 < 0$ [].

236 The significant feature of this potential is that its minimum does not occur for a value of
237 $\Phi = 0$. Instead, it is minimized when $|\Phi^\dagger \Phi| = -\mu^2/\lambda$. This means that in its ground state, the
238 Higgs field takes on a non-zero value - referred to as a vacuum expectation value (VEV). So while
239 the Higgs potential is globally symmetric, about the minimum this symmetry is broken. Since

²⁴⁰ the minimum is determined only by $\Phi^\dagger \Phi$, there is some ambiguity in the particular definition of

²⁴¹ the VEV, but it is generally represented as

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

²⁴² The full value of Φ can be written as

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

²⁴³ with v being the value of the VEV, and H being the real value of the scalar field.

²⁴⁴ 2.2.2 Electroweak Symmetry Breaking

²⁴⁵ The Electroweak (EWK) interaction is described in the SM by a $SU(2)_L \otimes U(1)_Y$ gauge theory.

²⁴⁶ 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,

²⁴⁷ B_μ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar

²⁴⁸ field Φ in the Lagrangian:

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

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

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

253 Expanding these terms into the mass eigenstates of the electroweak interaction yields four
 254 physical gauge bosons, two charged and two neutral, which are linear combinations of the fields
 255 W_μ^1 , W_μ^2 , W_μ^3 , and B_μ :

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

256 And the masses of these fields are given by:

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

257 This produces exactly the particles we observe - three massive gauge bosons and a single
 258 massless photon. The massless photon represents the portion of the gauge symmetry, a single
 259 $U(1)$ of the electromagnetic force, that remains unbroken by the VEV.

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

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

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

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

269 The fermion's Yukawa coupling to the VEV take the same form as the fermion's coupling
 270 to the Higgs boson - λ_ψ . Therefore, the strength of a fermion's interaction with the Higgs is
 271 directly proportional to its mass. We now have a model that predicts a Higgs boson with mass
 272 $M_H = \mu$, which interacts with the fermions with coupling strength λ_ψ . Because μ and λ_ψ are

²⁷³ free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions

²⁷⁴ must be measured experimentally.

²⁷⁵ **2.3 $t\bar{t}H$ Production**

²⁷⁶ The strength of a particles interaction with the Higgs, given by its Yukawa coupling, is propor-
²⁷⁷ tionate to its mass. The top quark - as the heaviest known particle - has the strongest interaction,
²⁷⁸ making this interaction particularly interesting to study. While several processes involve interac-
²⁷⁹ tions between the Higgs and the top, some Higgs production modes include the top interaction
²⁸⁰ only as a part of a loop diagram, such as the gluon-gluon fusion diagram shown in Figure 2.3.

²⁸¹ This process therefore only allows for an indirect probe of the Higgs-top Yukawa coupling, as
²⁸² the flavor of the quark in this diagram is not unique.

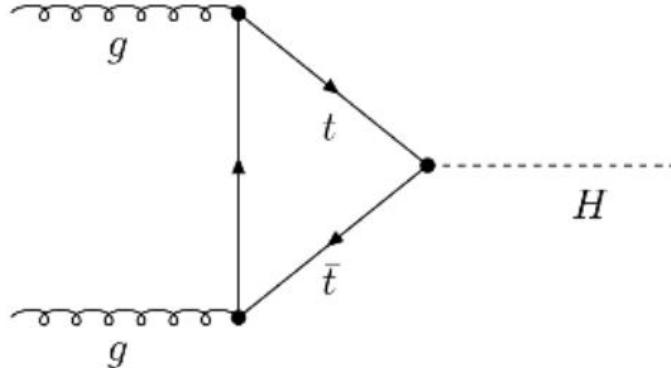


Figure 2.3: Diagram of a Higgs boson produced via gluon-gluon fusion.

²⁸³ Studying the Higgs produced in association with top quark pairs, $t\bar{t}H$, allows this interac-
²⁸⁴ tion to be measured directly. This process, as shown in Figure 2.4, involves a unique coupling

285 between the Higgs and the top, which can be identified by the top quark pair in the final state.

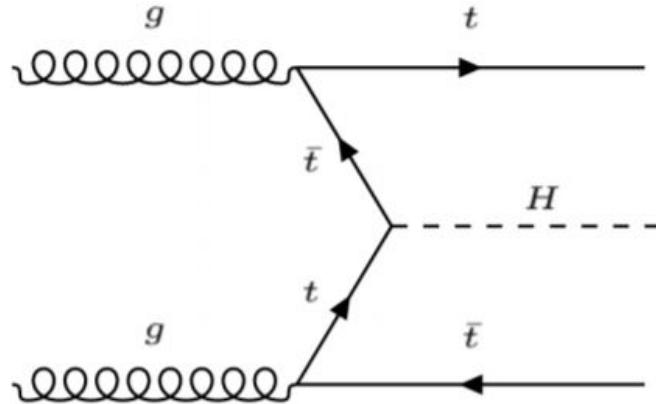


Figure 2.4: Diagram of a Higgs boson produced in association with a pair of top quarks.

286 The Higgs boson, as well as the top quarks, have very short lifetimes - on the order of
 287 10^{-22} s and 10^{-25} s respectively - meaning they can only be observed via their decay products.
 288 This thesis focuses on $t\bar{t}H$ events with multiple leptons in the final state, $t\bar{t}H - ML$. This
 289 includes $H \rightarrow W^+W^-$ events, where at least one of the W bosons decays leptonically.

290 While the branching ratio of $H \rightarrow W^+W^-$ is smaller than $H \rightarrow b\bar{b}$, it produces a clearer
 291 signal. On the other hand, $H \rightarrow \gamma\gamma$ produces the most easily identifiable signal, but has a much
 292 smaller branching ratio than $H \rightarrow W^+W^-$. Therefore, compared with other final state, the
 293 $t\bar{t}H - ML$ channel is an attractive candidate for study, as it involves a good balance between
 294 statistical power and identifiability.

295 **2.4 Extensions to the Standard Model**

296 While the SM has been tested to great precision, particularly at the LHC, it is generally accepted
297 that it is only valid up to a certain energy scale. It is assumed that above a certain energy, at the
298 scale where something like a Grand Unified Theory (GUT) or quantum gravity become relevant,
299 the SM will not be applicable. Further, there are several experimental observations that the SM
300 fails to explain. For example, the SM predicts neutrinos to be massless, despite experimental
301 observation to the contrary, and fails to explain the observation of dark matter and dark energy.

302 Another example, relevant to the Higgs sector, is known as the hierarchy problem: large
303 quantum corrections to the Higgs mass from loop diagrams, such as those shown in Figure 2.5,
304 are many orders of magnitude larger than the Higgs mass itself. The observed value of the Higgs
305 mass therefore requires extremely precise cancellation between these corrections and the bare
306 mass of the Higgs, a cancellation which seems unnatural and suggests something missing in our
307 theoretical picture.

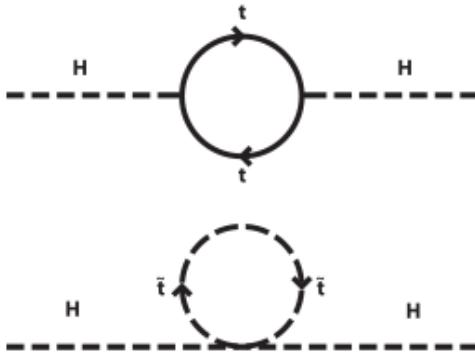


Figure 2.5: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram [].

308 Because so many of the properties of the Higgs boson have not yet been studied, its
 309 interactions are a promising place to search for new physics that could resolve some of the
 310 limitations of the SM. As explained above, the interactions of the Higgs with the top quark, as in
 311 $t\bar{t}H$ production, are particularly interesting: As the most massive particle in the Standard Model,
 312 the top quark is the most strongly interacting with the Higgs. Therefore, any new physics effects
 313 are likely to be seen most prominently in this interaction.

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

319 An Effective Field Theory approach can be used to model the low energy effects of new,

320 high energy physics, by parameterizing BSM effects as higher dimensional operators. These
 321 additional operators can then be added to the SM Lagrangian to write an effective Lagrangian
 322 that accounts for the effects of these higher energy physics. The lowest order of these that could
 323 contribute to Higgs-top couplings are dimension-six, as represented in Equation 2.12.

$$\mathcal{L}_{\text{BSM}} = \mathcal{L}_{\text{SM}} + \frac{f}{\Lambda} O^6 \quad (2.12)$$

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

327 The addition of these operators can be shown to modify the transverse momentum (p_T)
 328 spectrum of the Higgs Boson in Higg-top interactions, without effecting the overall rate of $t\bar{t}H$
 329 production []. The effect of these higher order effects on the as shown in Figure 2.6.

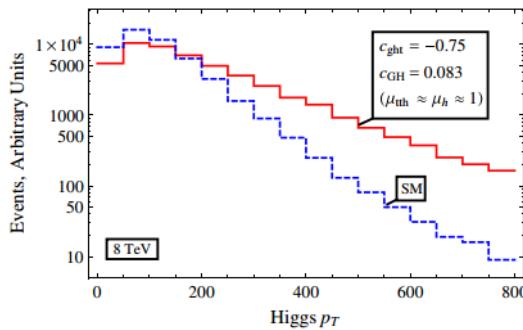


Figure 2.6: The momentum spectrum of the Higgs boson produced via top-quarks with (red) and without (blue) the presence of dimension-six operators.

330 This provides a clear, physics observable that could be used to search for BSM effect.

331 Therefore, reconstructing the momentum spectrum of the Higgs in $t\bar{t}H$ events provides a means

332 to observe new physics in the Higgs sector.

333 **Part III**

334 **The LHC and the ATLAS Detector**

335 **3 The LHC**

336 The Large Hadron Collider (LHC) is a particle accelerator consisting of a 27 km ring, designed

337 to collide protons at high energy. Located outside of Geneva, Switzerland and buried about 100

338 m underground, it consists of a ring of superconducting magnets which are used to accelerate

339 opposing beams of protons - or lead ions - which collide at the center of one of the various

340 detectors located around the LHC ring which record the result of these collisions. These

341 detectors include two general purpose detectors, ATLAS and CMS, which are designed to make

342 precision measurements of a broad range of physics phenomenon, and two more specialized

343 experiments, LHCb and ALICE, which are optimized to study b-quarks and heavy-ion physics,

344 respectively.

345 The LHC first began running in 2009 at a proton-proton center of mass energy of $\sqrt{s} = 8$

346 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during

347 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,
348 and collected data at an increased energy of $\sqrt{s} = 13$ TeV until 2018, a period referred to as Run
349 2.

350 The LHC consists of a chain of accelerators, which accelerate the protons to higher and
351 higher energies until they are injected into the main ring. This process is summarized in figure
352 3.1. Protons extracted from a tank of ionized hydrogen are fed into a linear accelerator, LINAC2,
353 where they reach an energy of 50 MeV. From there, they enter a series of three separate circular
354 accelerators, before being injected into the main accelerator ring at an energy of 450 GeV. Within
355 the main ring protons are separated into two separate beams moving in opposite directions,
356 and their energy is increased to their full collision energy. Radiofrequency cavities are used to
357 accelerate these particles and sort them into bunches. From 2015-2018, these bunches consisted
358 of around 100 billion protons each with an energy of 6.5 TeV per proton, which collided at a rate
359 of 40 MHz, or every 25 ns.

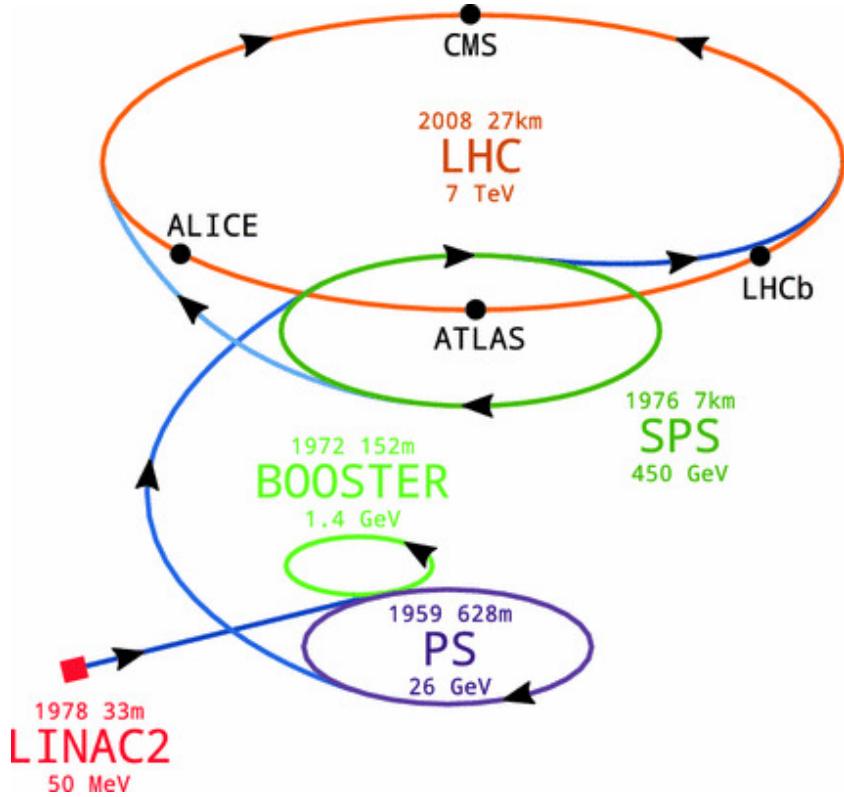


Figure 3.1: A summary of the accelerator chain used to feed protons into the LHC [].

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

364 The amount of data collected by the LHC is measured in terms of luminosity, which is the

365 ratio of the number of events detected per unit time, $\frac{dN}{dt}$, and the interaction cross-section, σ .

$$\mathcal{L} = \frac{1}{\sigma} \frac{dN}{dt} \quad (3.1)$$

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

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} dt \quad (3.2)$$

369 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around
370 140 fb^{-1} , exceeding the expected integrated luminosity of 100 fb^{-1} .

371 4 The ATLAS Detector

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

³⁷⁷ The ATLAS detector consists of multiple layers, each of which serves a different purpose
³⁷⁸ in reconstructing collisions. At the very center of the detector is the interaction point where the
³⁷⁹ proton beams of the LHC collide.

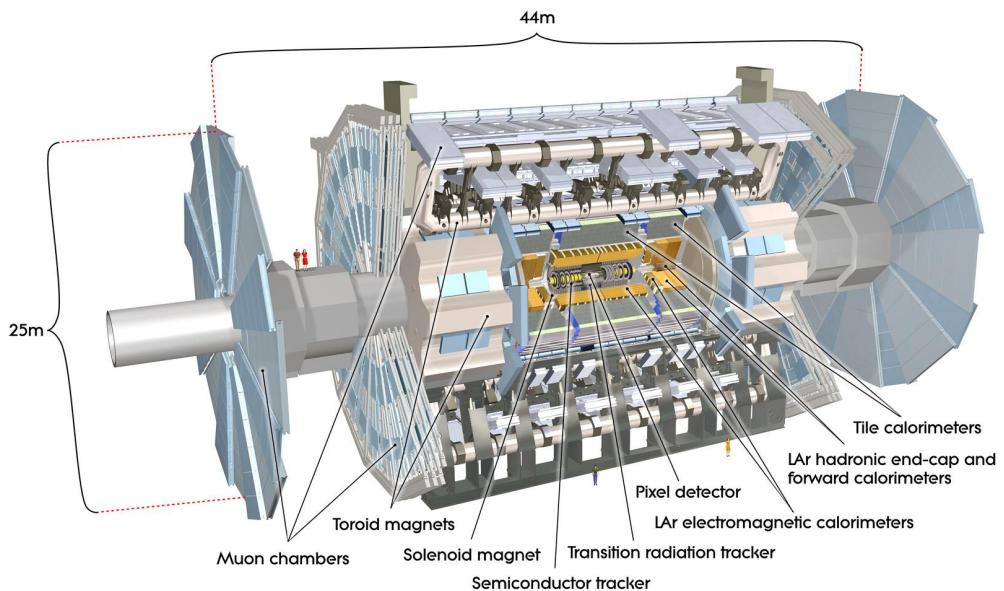


Figure 4.1: Cutaway view of the ATLAS detector, with labels of its major components [].

380 **4.1 Inner Detector**

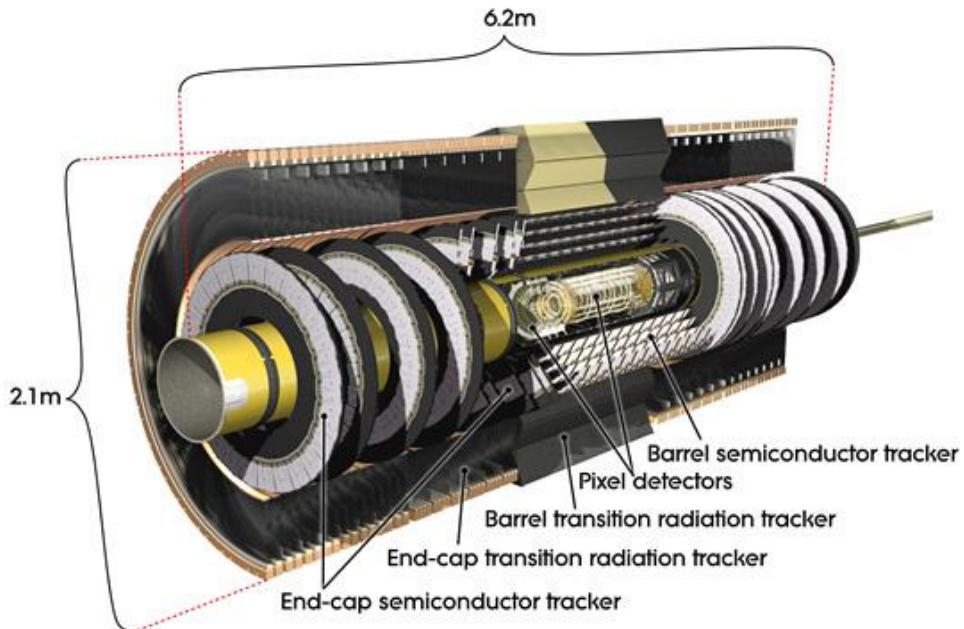


Figure 4.2: Cutaway view of the Inner Detector [].

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

387 The Inner Detector consists of three components - the Pixel Detector, the Semi-Conductor
388 Tracker (SCT), and the Transition Radiation Tracker (TRT). The Pixel Detector is the innermost

389 of these, beginning just 33.25 mm away from the beam line. It consists of three silicon layers
 390 along the barrel, as well as three endcap layers, covering a range of $|\eta| < 2.5$.

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

393 4.2 Calorimeters

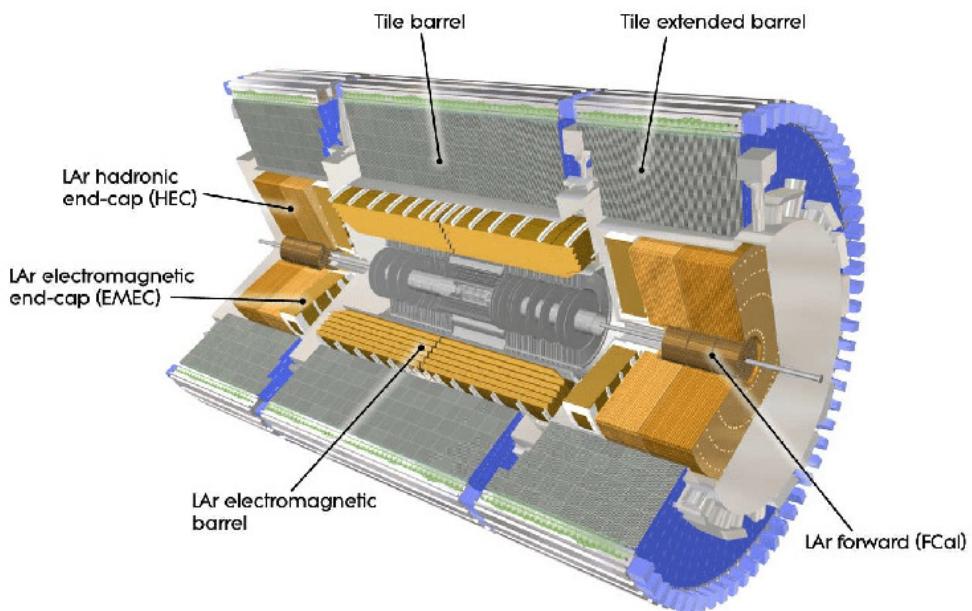


Figure 4.3: Cutaway view of the calorimeter system of the ATLAS detector [].

394 Situated outside the Inner Detector are two concentric calorimeters. The inner calorimeter uses
 395 liquid argon (LAr) to measure energy of particles that interact electromagnetically, which includes
 396 photons and any charged particle. The LAr calorimeter is made of heavy metals, primarily lead
 397 and copper, which causes electromagnetically interacting particles to shower, depositing their

398 energy in the detector. The showering of the high energy particles that pass through calorimeter
399 cause the liquid argon to ionize, and the ionized electrons are detected by electronic readouts.
400 The LAr calorimeter consists of around 180,000 readout channels.

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

406 **4.3 Muon Spectrometer**

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

415 Two large toroidal magnets within the muon system generate a large magnetic field which
416 covers an area 26 m long with a radius of 10 m. Because the area covered by this magnet system

417 is so large, a uniform magnetic field like the one produced in the Inner Detector is impractical.
418 Instead, the magnetic field that exists in the muon spectrometer ranges between 2 T and 8 T, and
419 is much less uniform. The path of the muons passing through the spectrometer is bent by this
420 field, allowing their charge to be determined.

421 1200 tracking chambers are placed in the muon system in order to precisely measure the
422 tracks of muons with high spatial resolution.

423 **4.4 Trigger System**

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

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

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

442 **Part IV**

443 **Measurement of WZ + Heavy Flavor**

444 **5 Introduction**

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

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

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

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

462 Section 10 details the data and Monte Carlo (MC) samples used in the analysis. The
463 reconstruction of various physics objects is described in Section 11. Section 12 describes the
464 event selection applied to these samples, along the definitions of the various regions used in
465 the fit. The multivariate analysis techniques used to separate the tZ background from WZ +
466 heavy flavor are described in Section 13. Section 21 describes the various sources of systematic
467 uncertainties considered in the fit. Finally, the results of the analysis are summarized in Section
468 22, followed by a brief conclusion in Section ??.

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

472 6 Data and Monte Carlo Samples

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

- 479 • at least two light leptons within a range $|\eta| < 2.6$, with leading lepton $p_T > 15$ GeV and
480 subleading lepton $p_T > 5$ GeV
- 481 • at least one light lepton with $p_T > 15$ GeV within a range $|\eta| < 2.6$, and at least two hadronic
482 taus with $p_T > 15$ GeV.

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

486 6.1 Data Samples

487 The study uses a sample of proton-proton collision data collected by the ATLAS detector from
488 2015 through 2018 at an energy of $\sqrt{s} = 13$ TeV, which represents an integrated luminosity of

489 139 fb^{-1} . This data set was collected with a bunch crossing rate of 25 ns. All data used in this
490 analysis was verified by data quality checks, having been included in the following Good Run
491 Lists:

492 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
493 _PHYS_StandardGRL_All_Good_25ns.xml

494 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
495 _PHYS_StandardGRL_All_Good_25ns.xml

496 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL
497 _All_Good_25ns_Triggerno17e33prim.xml

498 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL
499 _All_Good_25ns_Triggerno17e33prim.xml

500 Runs included from the AllYear period containers are included.

501 **6.2 Monte Carlo Samples**

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

505 about the Monte Carlo samples being used can be found in Table 28. A list of the specific samples
 506 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.

507 7 Object Reconstruction

- 508 All regions defined in this analysis share a common lepton, jet, and overall event preselection.
- 509 The selection applied to each physics object is detailed here; the event preselection, and the
- 510 selection used to define the various fit regions, is described in Section 12.

⁵¹¹ **7.1 Trigger**

⁵¹² 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.

⁵¹³ **7.2 Light leptons**

⁵¹⁴ Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that

⁵¹⁵ are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].

⁵¹⁶ Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Muon candidates

⁵¹⁷ are reconstructed by combining inner detector tracks with track segments or full tracks in the

⁵¹⁸ muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have $p_T > 10$ GeV and

519 $|\eta| < 2.5$. Candidates in the transition region between different electromagnetic calorimeter
520 components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant com-
521 bining shower shape and track information is used to distinguish real electrons from hadronic
522 showers (fake electrons). To further reduce the non-prompt electron contribution, the track is
523 required to be consistent with originating from the primary vertex; requirements are imposed
524 on the transverse impact parameter significance ($|d_0|/\sigma_{d_0} < 5$) and the longitudinal impact
525 parameter ($|\Delta z_0 \sin \theta_\ell| < 0.5 \text{ mm}$).

526 Muon candidates are reconstructed by combining inner detector tracks with track segments
527 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
528 $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$. The longitudinal impact parameter is the same for both electrons
529 and muons, while muons are required to pass a slightly tighter transverse impact parameter,
530 $|d_0|/\sigma_{d_0} < 3$.

531 All leptons are required to be isolated, as defined through the standard PLVLoose working
532 point supported by combined performance groups. Leptons are additionally required to pass a
533 non-prompt BDT selection developed by the $t\bar{t}H$ multilepton/ $t\bar{t}W$ analysis group. This BDT is
534 described in detail in [**ttW_140**]. Optimized working points and scale factors for this BDT are
535 taken from that analysis.

536 7.3 Jets

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

545 7.4 B-tagged Jets

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

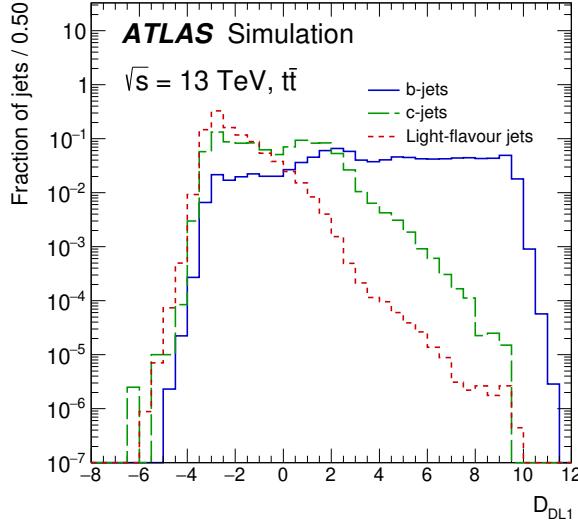


Figure 7.1: Output distribution of the DL1r algorithm for b-jets, charm jets, and light jets

551 From the output of the BDT, working points (WPs) are developed based on the efficiency
 552 of truth b-jets at particular values of the DL1r algorithm. These WPs are taken from the March
 553 2020 CDI file, 2020-21-13TeV-MC16-CDI-2020-03-11_v2.root. The working points used in
 554 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

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

559 60% WP fall into another, and looking at the full psuedo-continuous DL1r spectrum of the jets,
560 additional information can be gained. The psuedo-continuous b-tag spectrum is used in this case
561 to separate out WZ + b, WZ + charm, and WZ + light.

562 7.5 Missing transverse energy

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

570 7.6 Overlap removal

571 To avoid double counting objects and remove leptons originating from decays of hadrons, overlap
572 removal is performed in the following order: any electron candidate within $\Delta R = 0.1$ of another
573 electron candidate with higher p_T is removed; any electron candidate within $\Delta R = 0.1$ of a muon
574 candidate is removed; any jet within $\Delta R = 0.3$ of an electron candidate is removed; if a muon
575 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
576 is kept and the muon is removed.

577 This algorithm is applied to the preselected objects. The overlap removal procedure is
 578 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.

579 8 Event Selection and Signal Region Definitions

580 Event are required to pass a preselection described in Section 12.1 and summarized in Table 14.
 581 Those that pass this preselection are divided into various fit regions described in Section 12.2,
 582 based on the number of jets in the event, and the b-tag score of those jets.

583 8.1 Event Preselection

584 Events are required to include exactly three reconstructed light leptons passing the requirement
 585 described in 11.2, which have a total charge of ± 1 . As the opposite sign lepton is found to
 586 be prompt the vast majority of the time [ttH_paper], it is required to be loose and isolated,
 587 as defined though the standard PLVLoose working point supported by combined performance
 588 groups. The same sign leptons are required to be very tightly isolated, as per the recommended
 589 PLVTight.

590 The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose
 591 charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.
 592 the smallest ΔR , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton
 593 0 is required to have $p_T > 10 \text{ GeV}$, while the same sign leptons, 1 and 2, are required to have
 594 $p_T > 20 \text{ GeV}$ to reduce the contribution of non-prompt leptons.

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

599 An additional requirement is placed on the missing transverse energy, $E_T^{\text{miss}} > 20 \text{ GeV}$,
 600 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
 601 missing transverse energy, and l_{other} is the lepton not included in the Z-candidate.

602 Events are required to have one or two reconstructed jets passing the selection described
 603 in Section 11.3. Events with more than two jets are rejected in order to reduce the contribution
 604 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

605 The event yields in the preselection region for both data and Monte Carlo are summarized
 606 in Table 12.1, which shows good agreement between data and Monte Carlo, and demonstrates that
 607 this region consists primarily of WZ events. The WZ events are split into WZ + b, WZ + c, and
 608 WZ + l based on the truth flavor of the associated jet in the event. Specifically, this determination
 609 is made based on the HadronConeExclTruthLabelID of the jet, as recommended by the b-
 610 tagging working group [BtagWG]. In this ordering b-jet supersedes charm, which supersedes
 611 light. That is, WZ + l events contain no charm and no b jets at truth level, WZ + c contain at
 612 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}

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

615 Simulations are further validated by comparing the kinematic distributions of the Monte
616 Carlo with data, which are shown in Figure 12.1. Here, bins with 5% or more WZ+b are
617 blinded.

WZ Fit Region - Inclusive

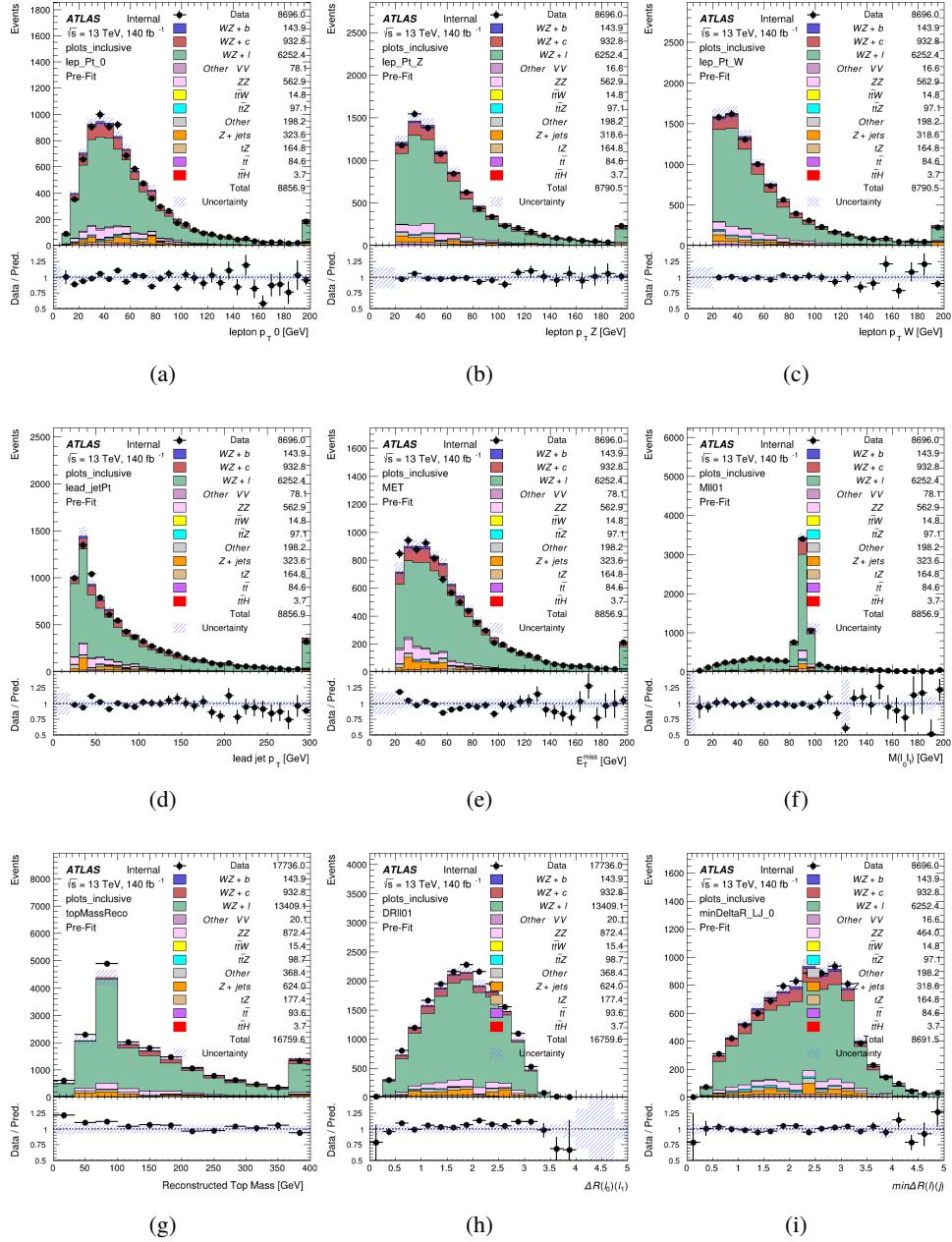


Figure 8.1: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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.

618 8.2 Fit Regions

619 Once preselection has been applied, the remaining events are categorized into one of twelve
 620 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	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} >= 1, tZ \text{ BDT} < 0.725$

621 The working points discussed in Section 11.4 are used to separate events into fit regions
 622 based on the highest working point reached by a jet in each event. Because the background
 623 composition differs significantly based on the number of b-jets, events are further subdivided
 624 into 1-jet and 2-jet regions in order to minimize the impact of background uncertainties.

625 An unfolding procedure is performed to account for differences in the number of recon-
 626 structed jets compared to the number of truth jets in each event. The `AntiKt4TruthDressedWZJets`
 627 truth jet collection is used to make this determination. In order to account for migration of WZ+1-
 628 jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal samples are
 629 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 12.2. 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 21.

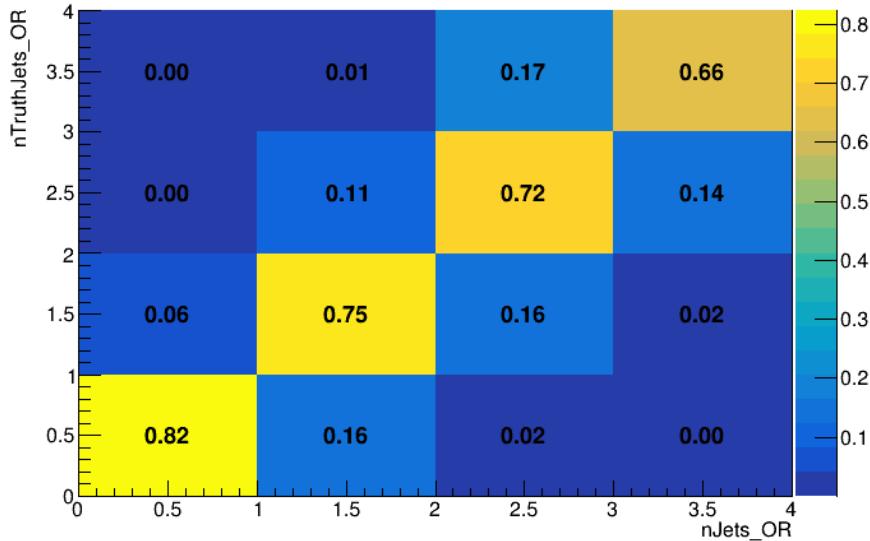


Figure 8.2: Number of truth jets compared to the number of reconstructed jets in each WZ event falling in the preselection region. Each row is normalized to unity.

An additional tZ control region is created based on the BDT described in Section 13. 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

640 various kinematic distributions. These plot are shown in Figures 12.3-12.16.

WZ Fit Region - 1j Inclusive

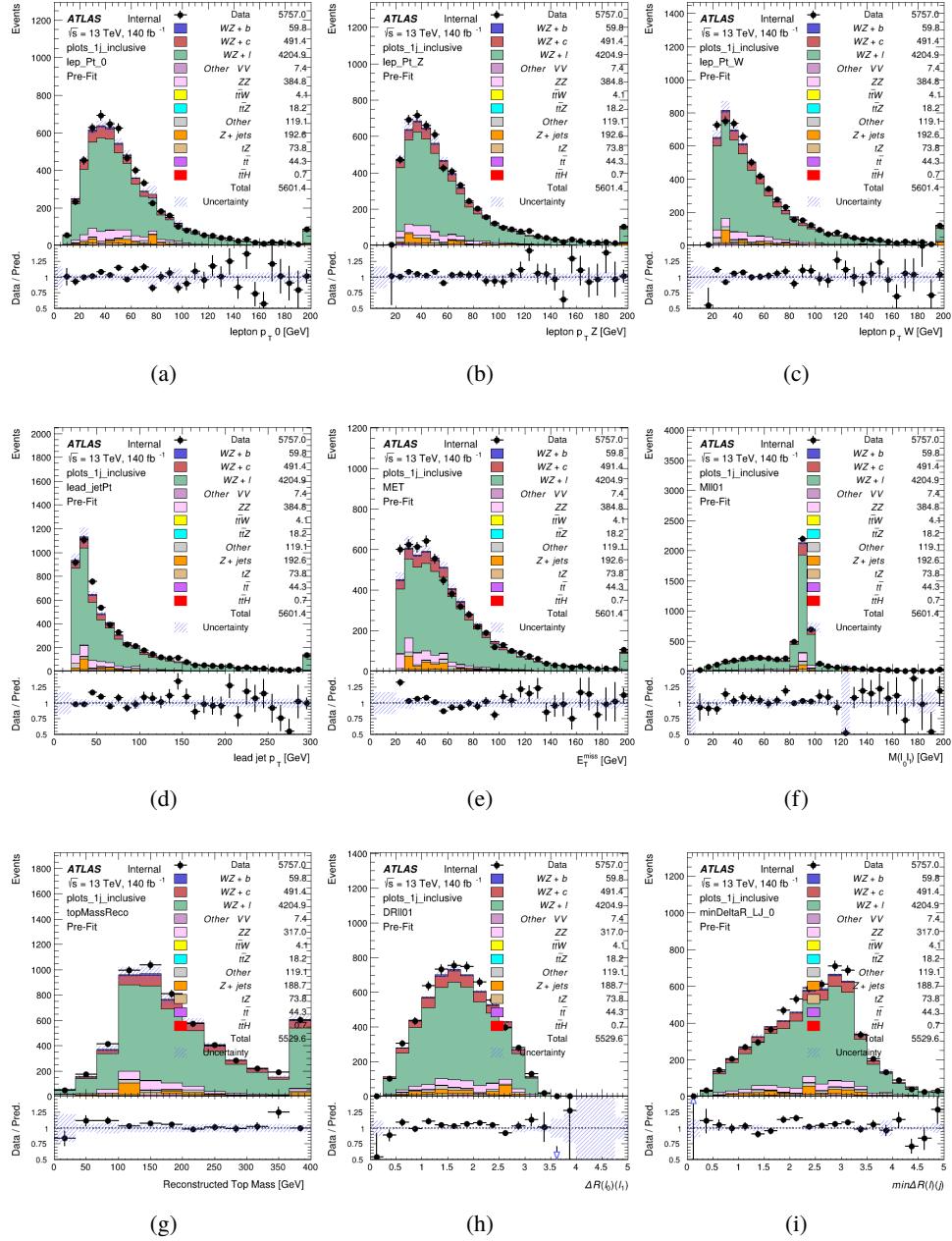


Figure 8.3: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

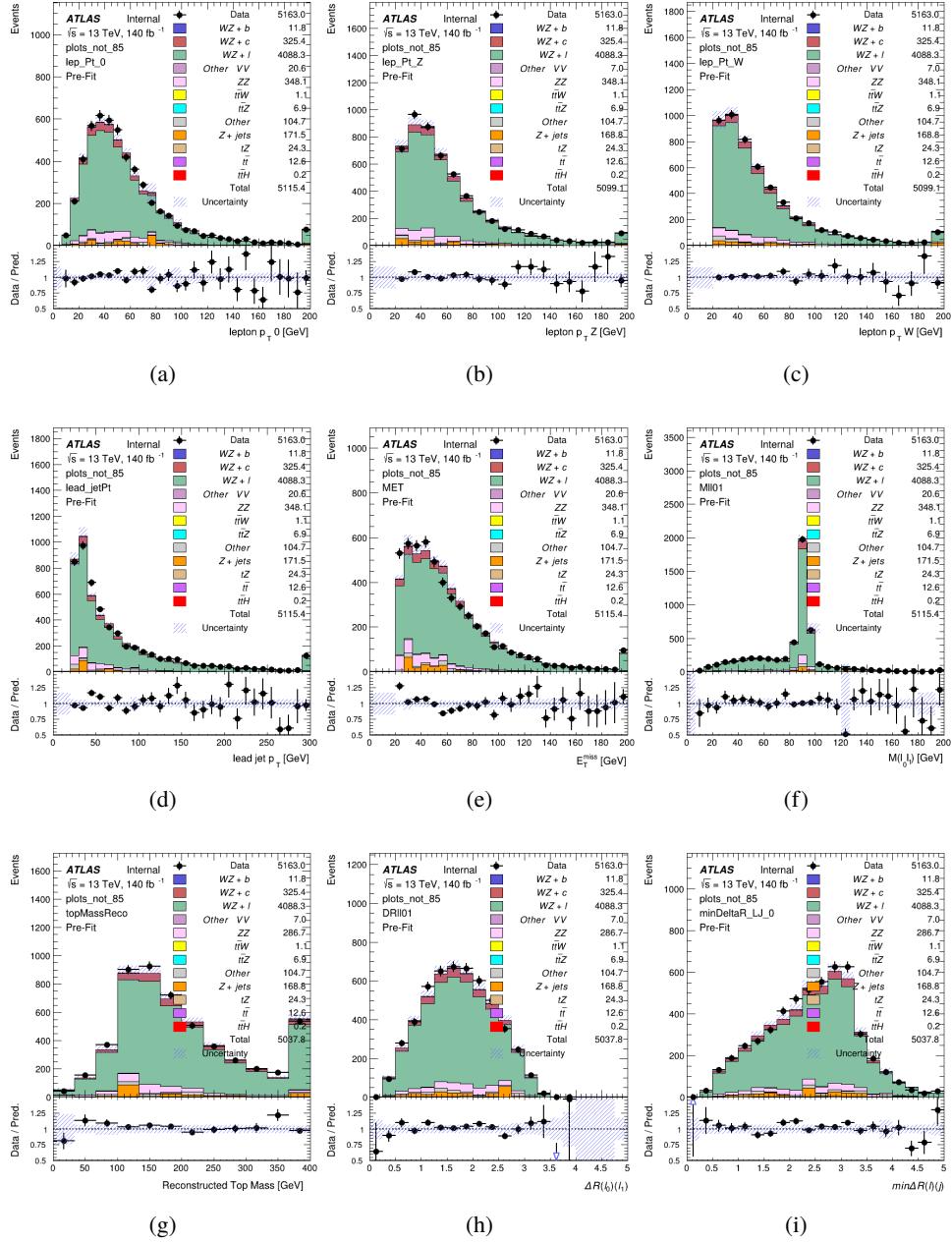


Figure 8.4: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

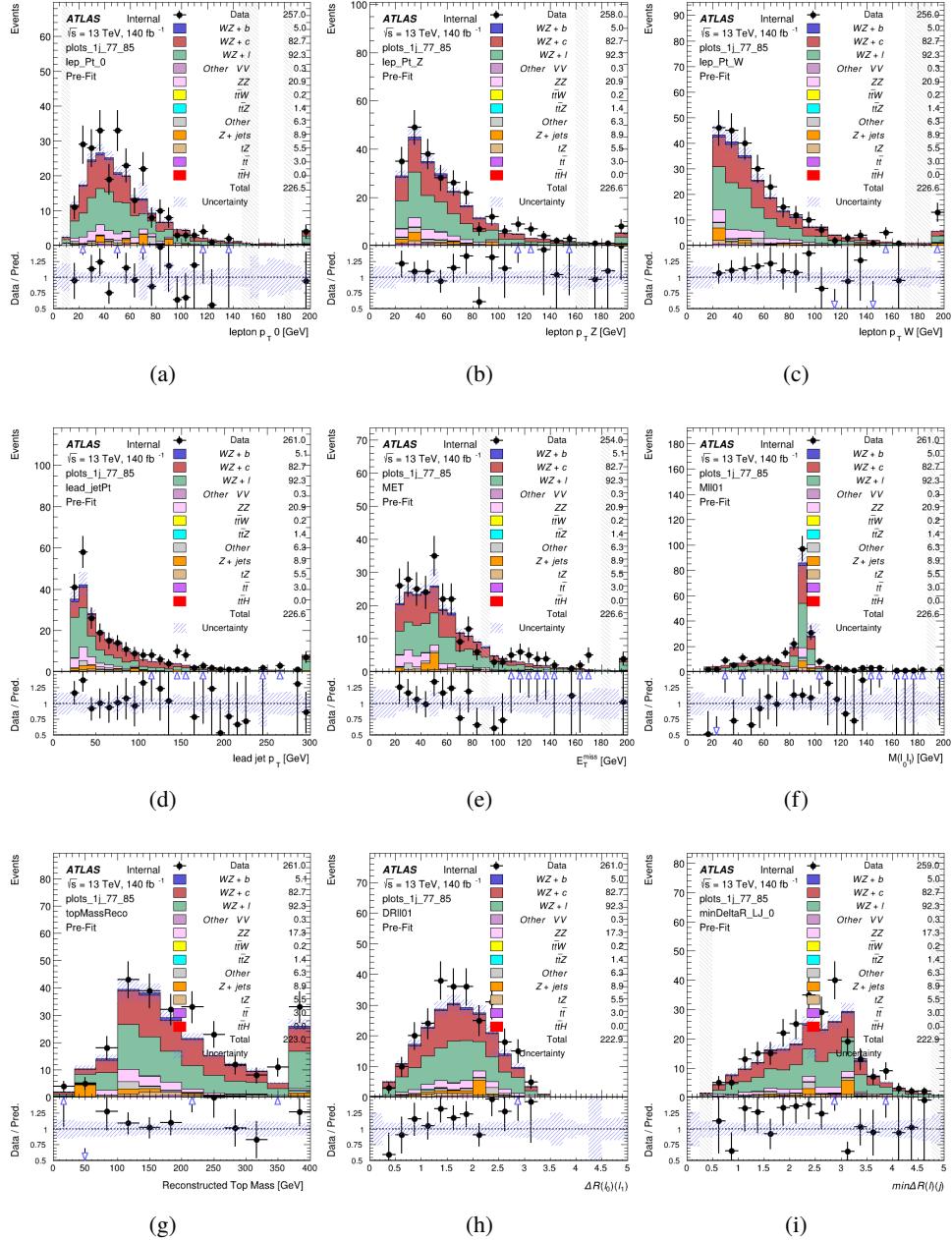


Figure 8.5: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^miss , and (f) the invariant 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

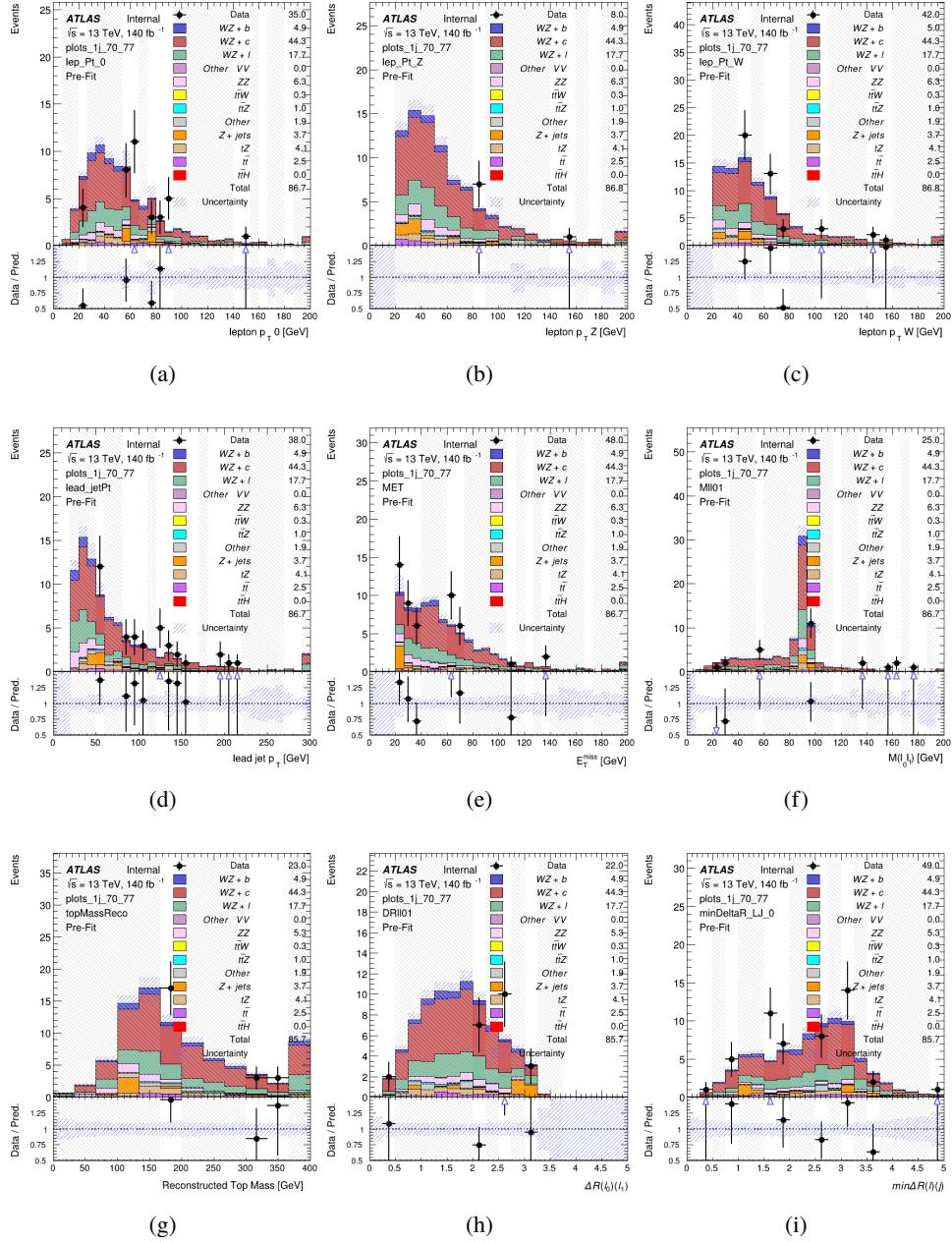


Figure 8.6: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

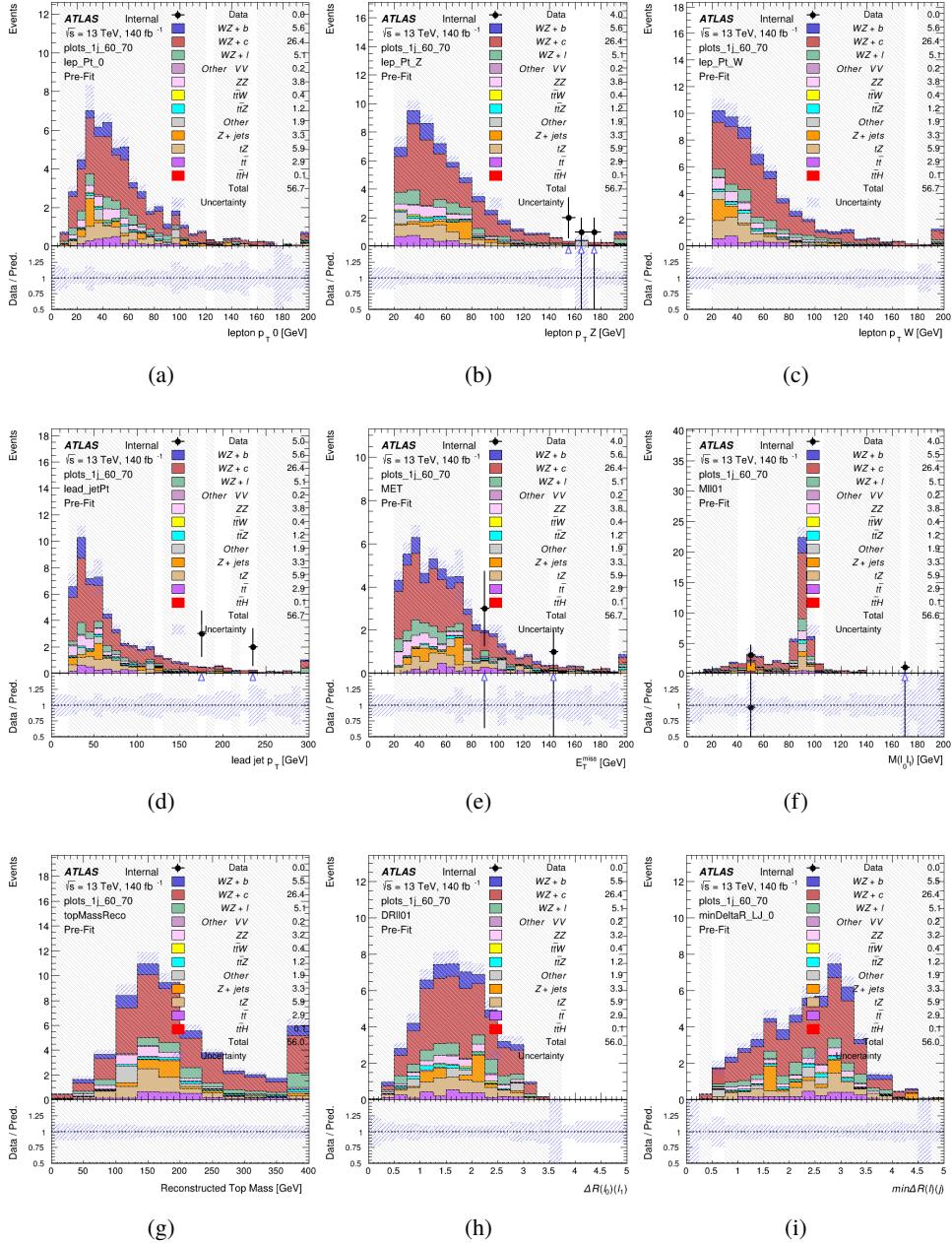


Figure 8.7: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

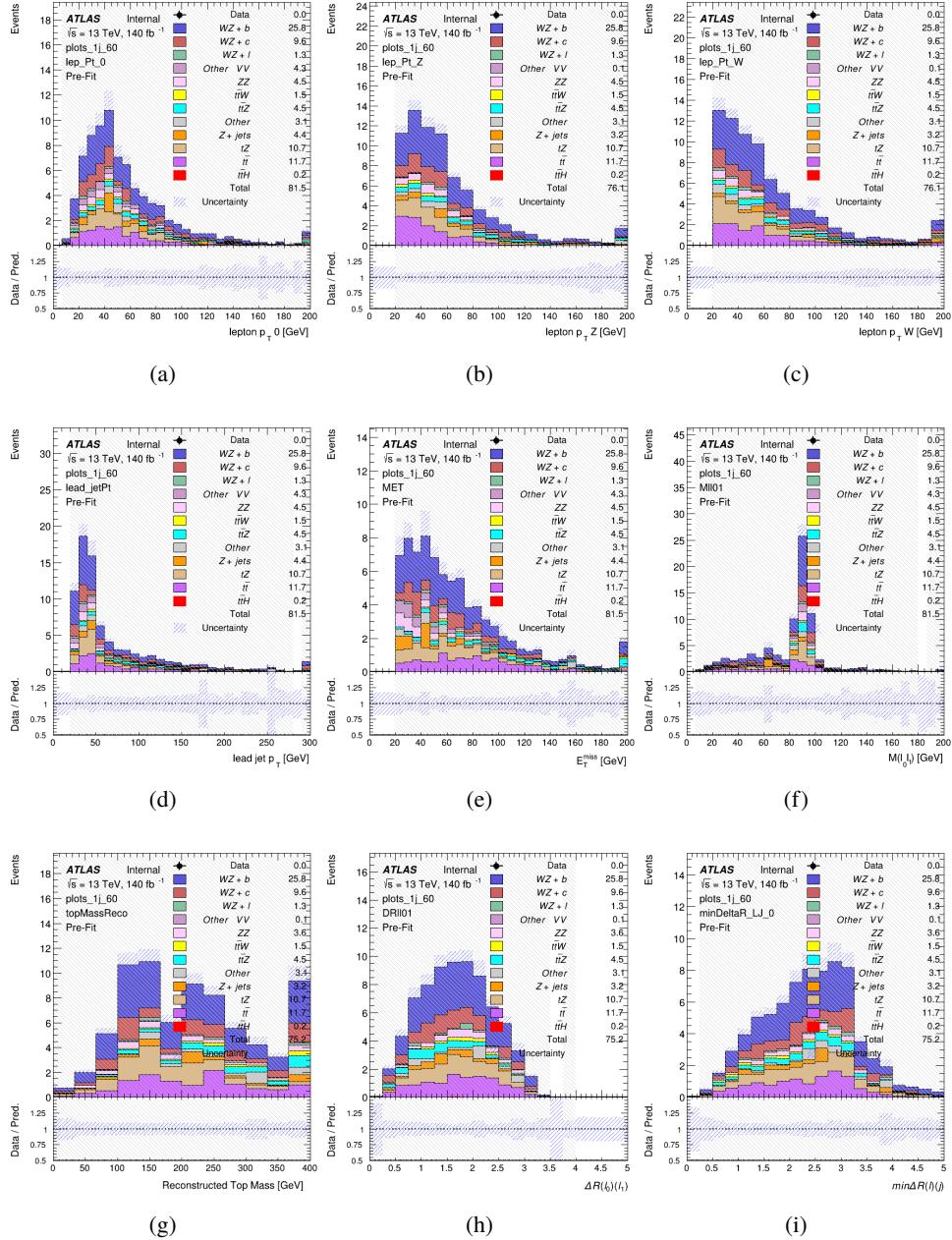


Figure 8.8: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

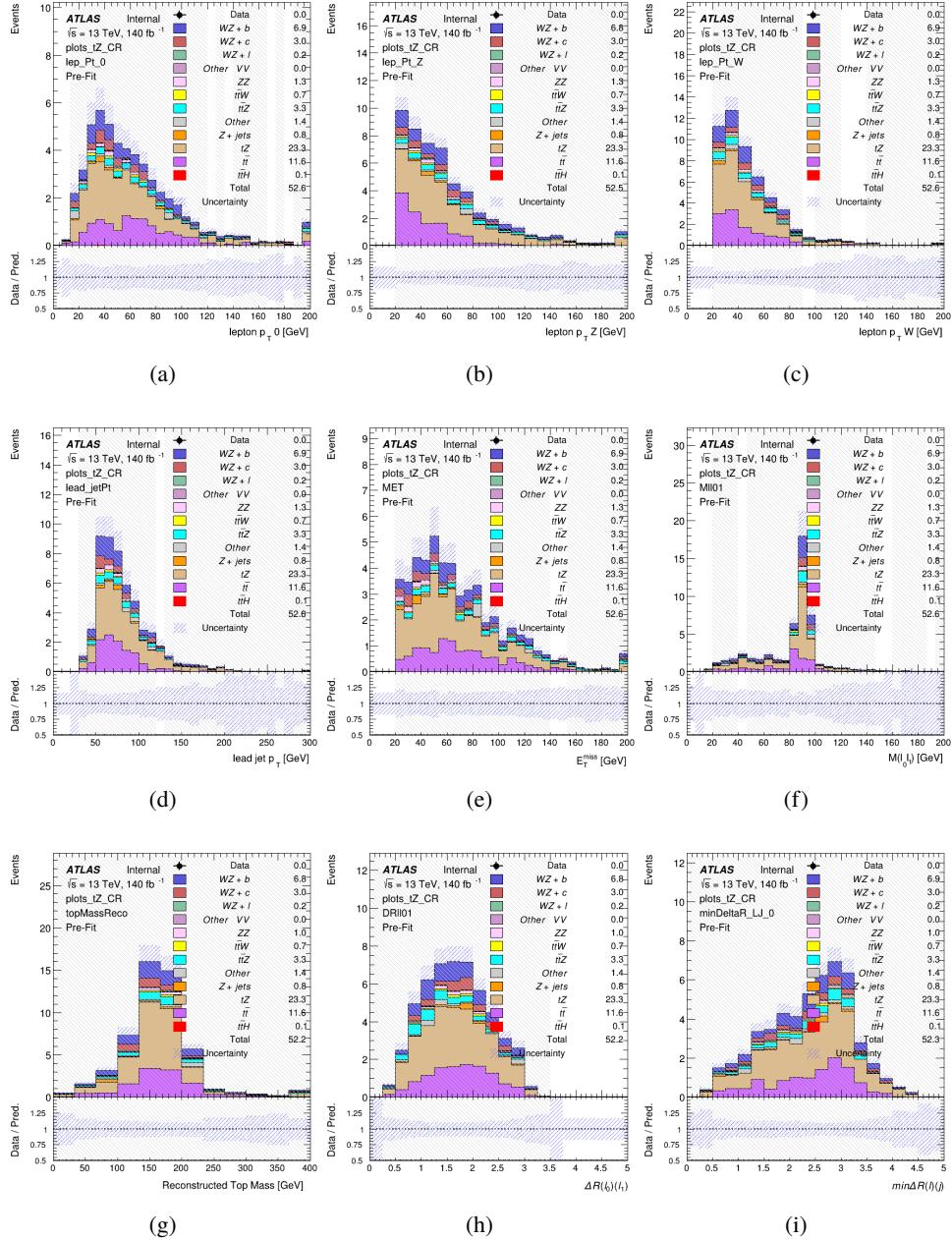


Figure 8.9: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

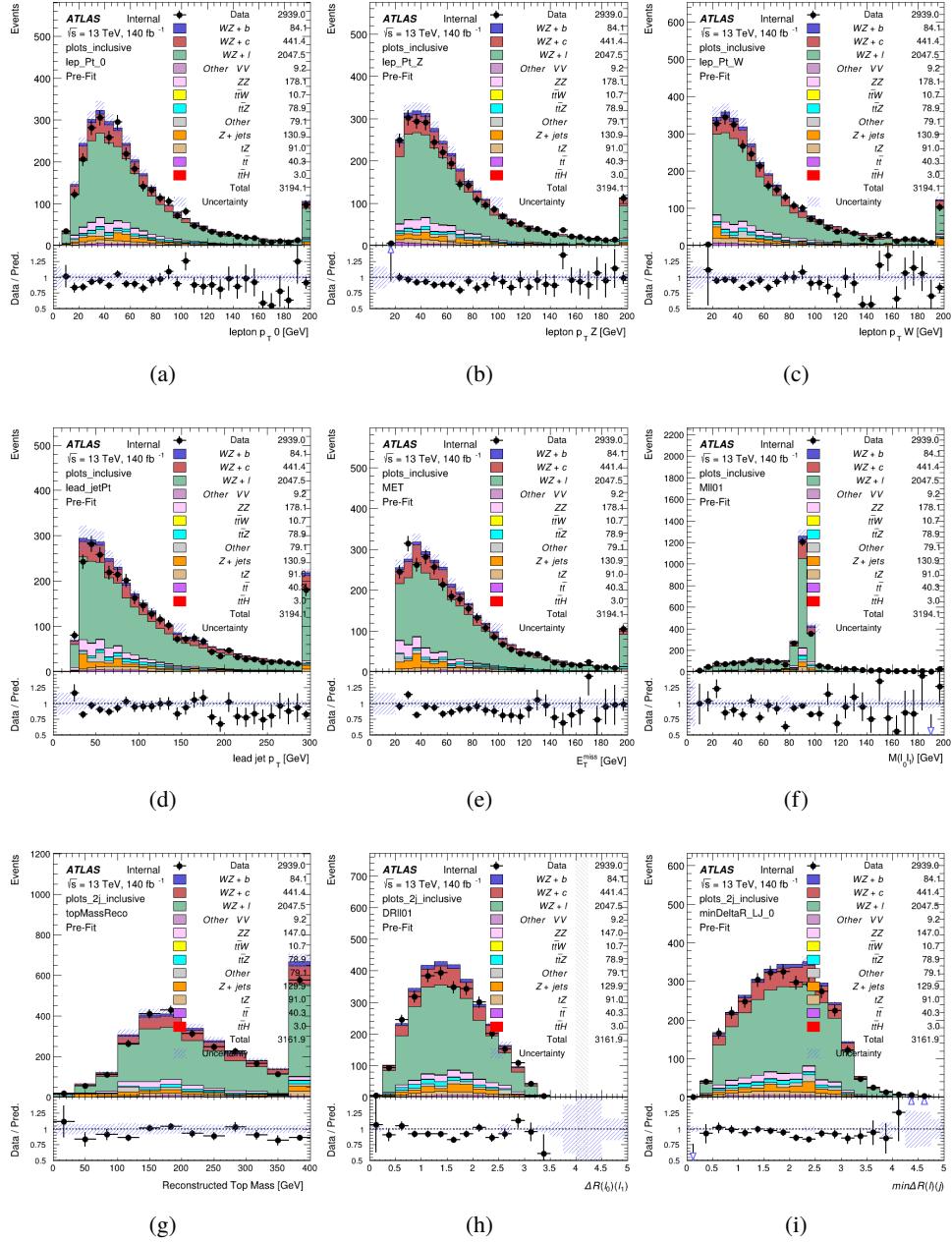


Figure 8.10: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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.

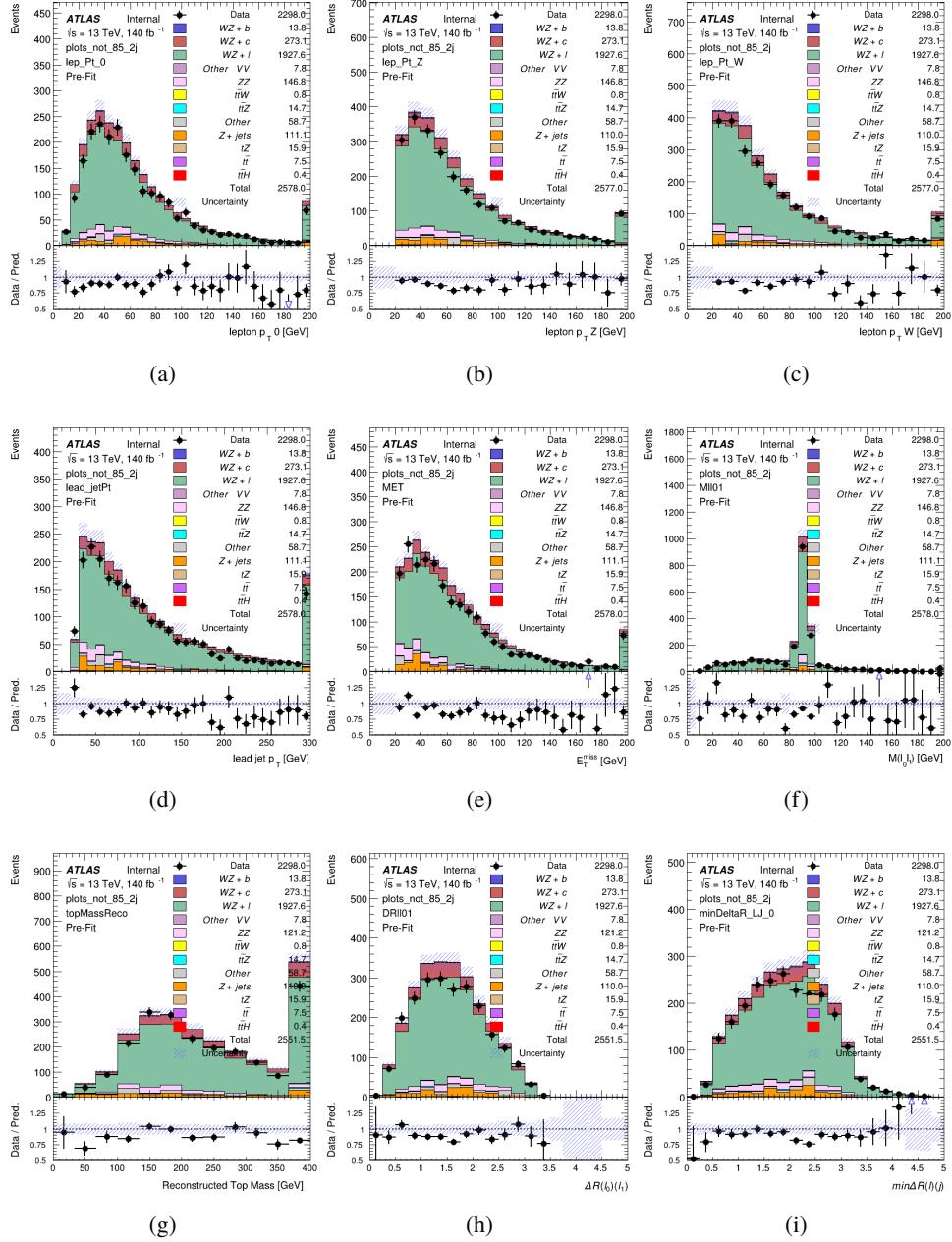
WZ Fit Region - $2j < 85\%$ WP

Figure 8.11: 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

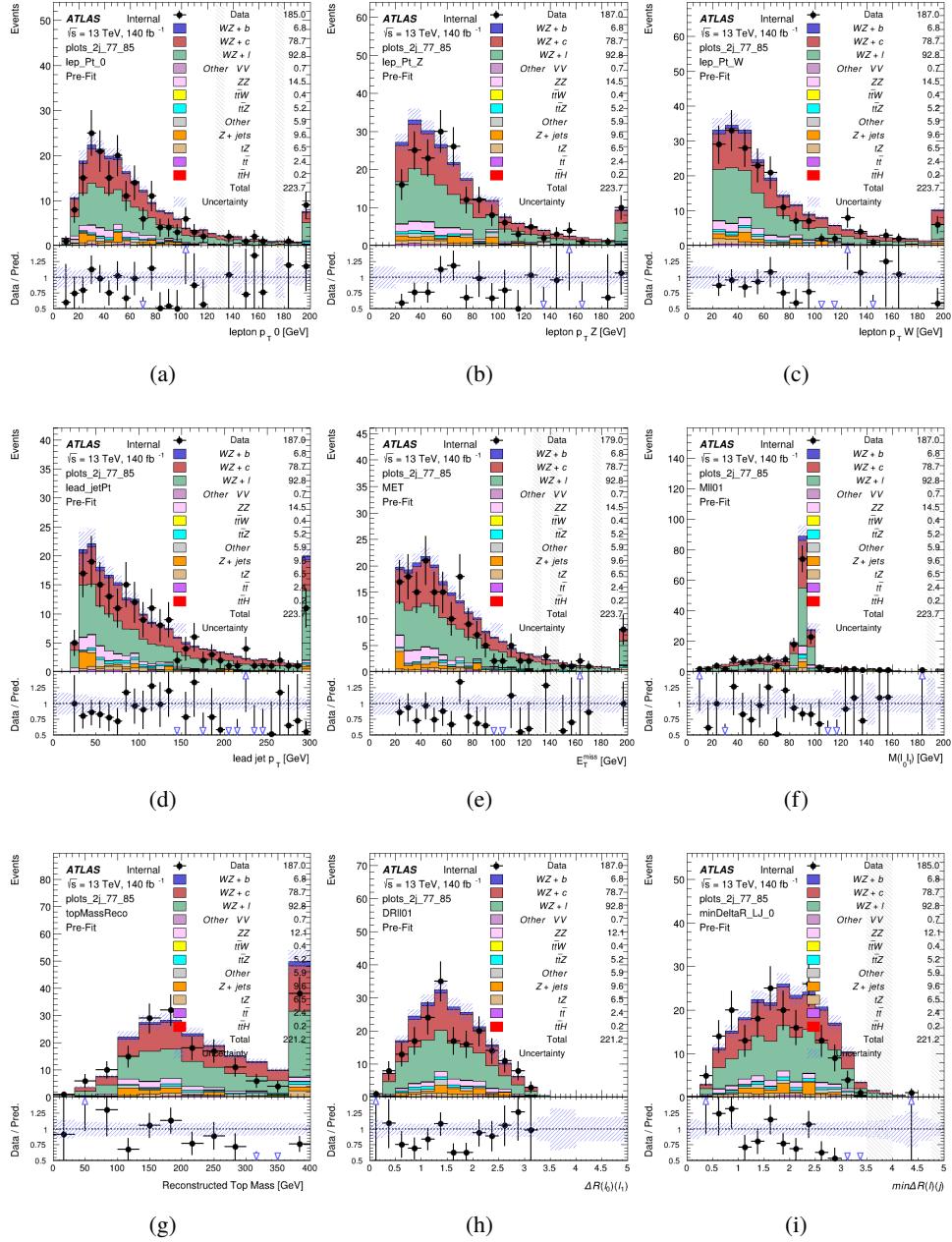


Figure 8.12: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

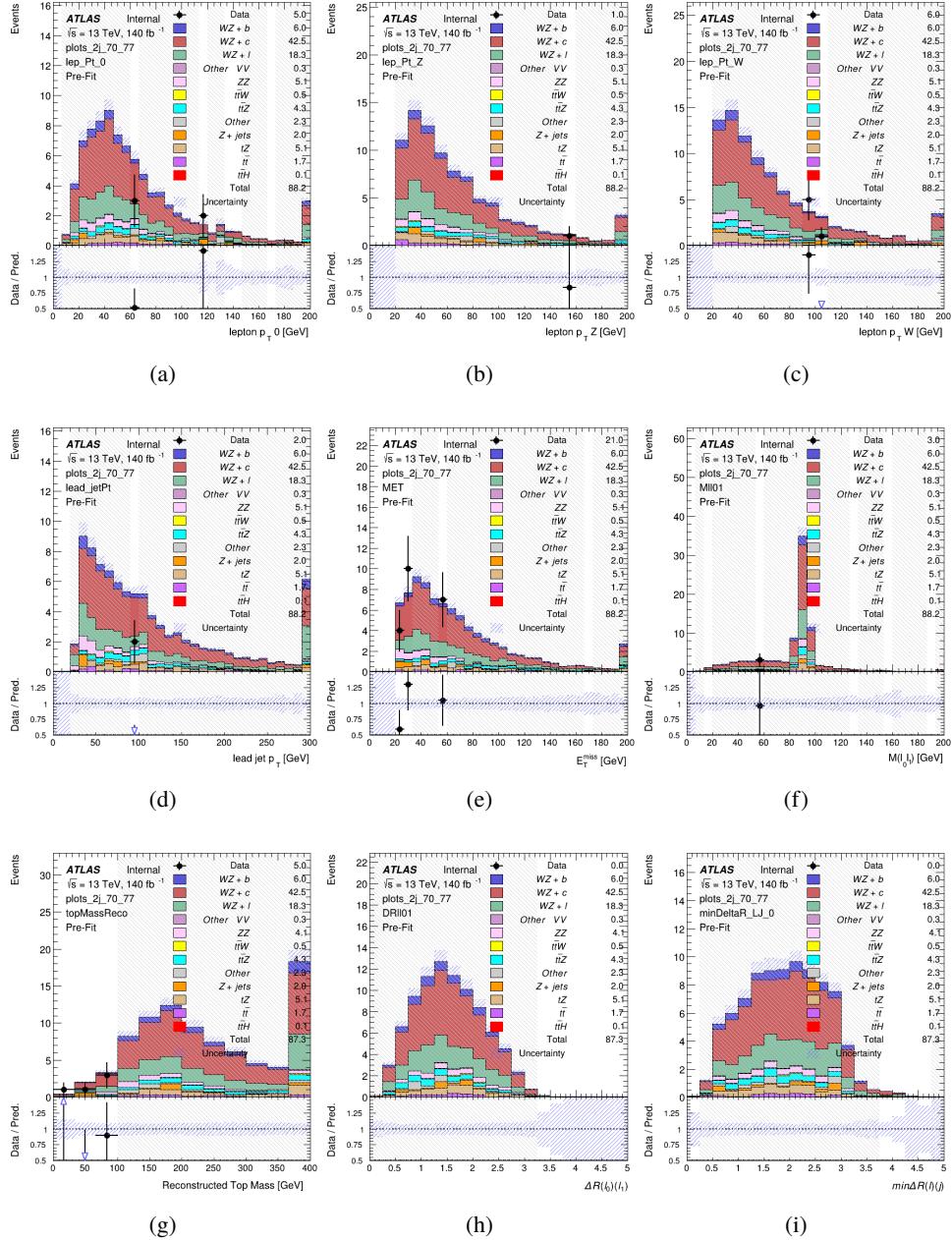


Figure 8.13: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

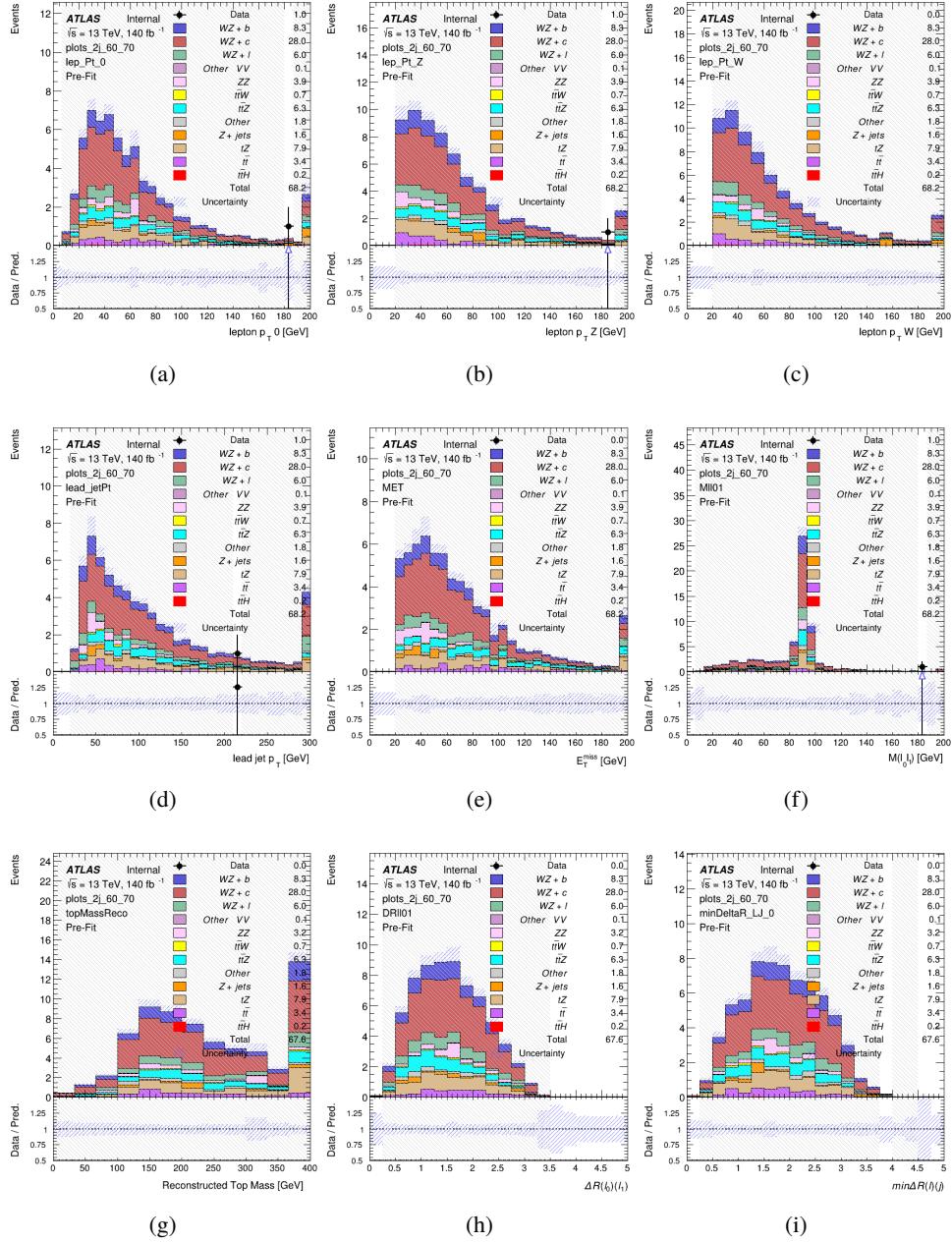


Figure 8.14: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

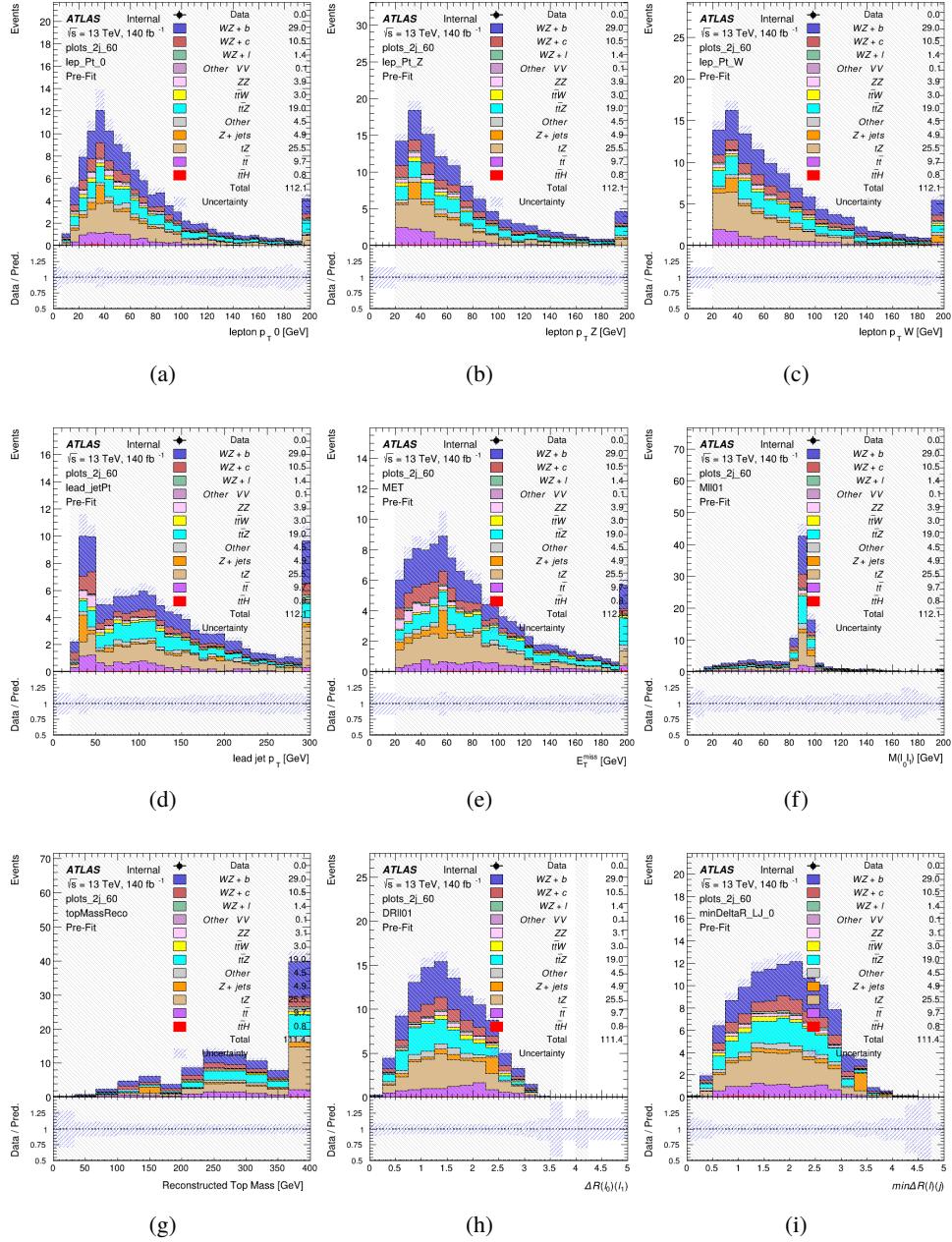


Figure 8.15: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

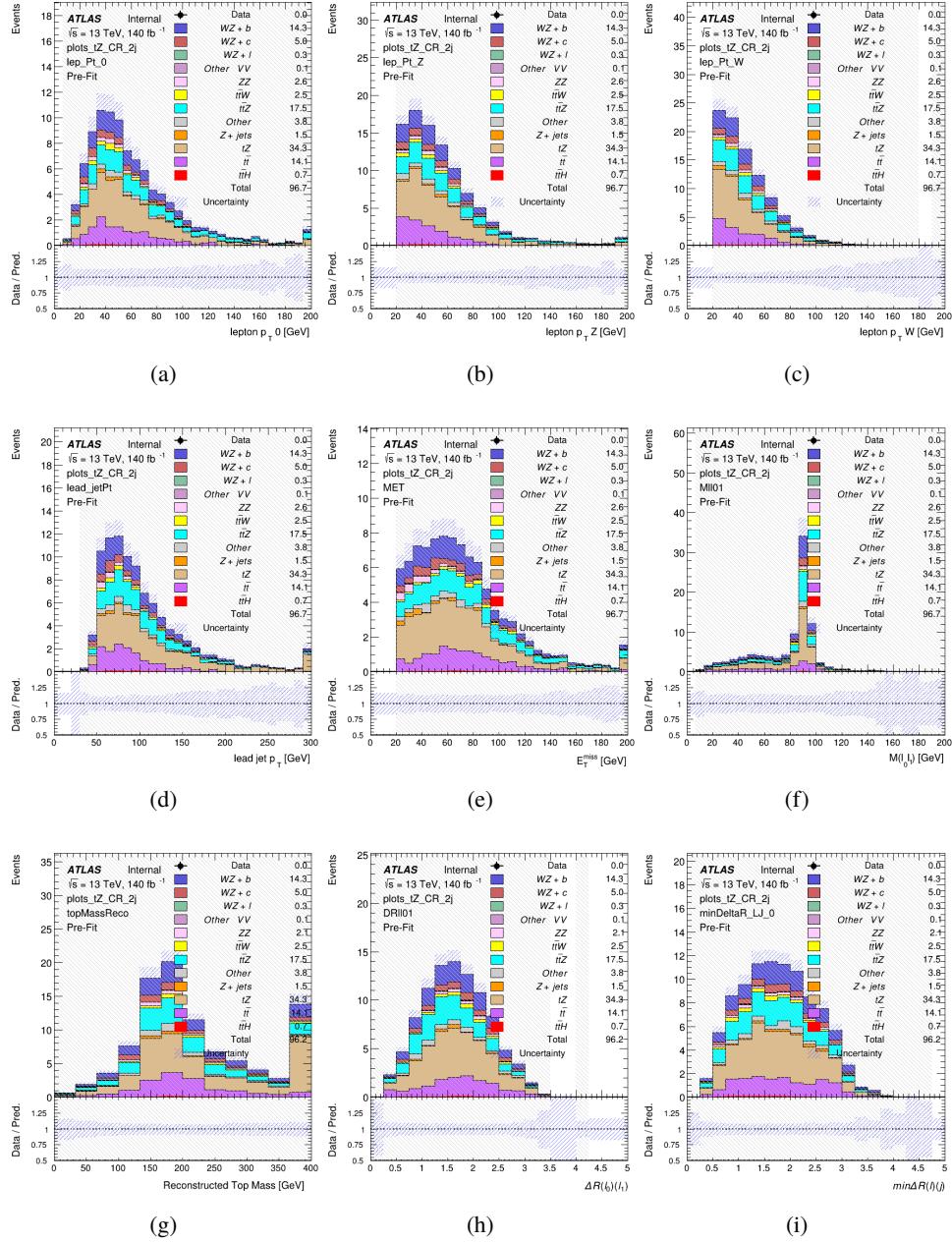


Figure 8.16: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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.

641 **8.3 Non-Prompt Lepton Estimation**

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

646 **8.3.1 $t\bar{t}$ Validation**

647 $t\bar{t}$ events can produce two prompt leptons from the decay of each of the tops. These top decays
 648 produce two b-quarks, the decay of which can produce additional non-prompt leptons, which
 649 occasionally pass the event preselection. In order to validate that the Monte Carlo accurately
 650 simulates this process accurately, the MC prediction in a non-prompt $t\bar{t}$ enriched control region
 651 is compared to data.

652 The $t\bar{t}$ control region is similar to the preselection region - three leptons meeting the
 653 criteria described in Section 12 are required, and the requirements on E_T^{miss} remain the same.
 654 However, the selection requiring a lepton pair form a Z-candidate are reversed. Events where the
 655 invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of 91.2 GeV are
 656 rejected. This ensures the $t\bar{t}$ control region is orthogonal to the preselection region.

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

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

660 Various kinematic plots of this region are shown in Figure 12.17.

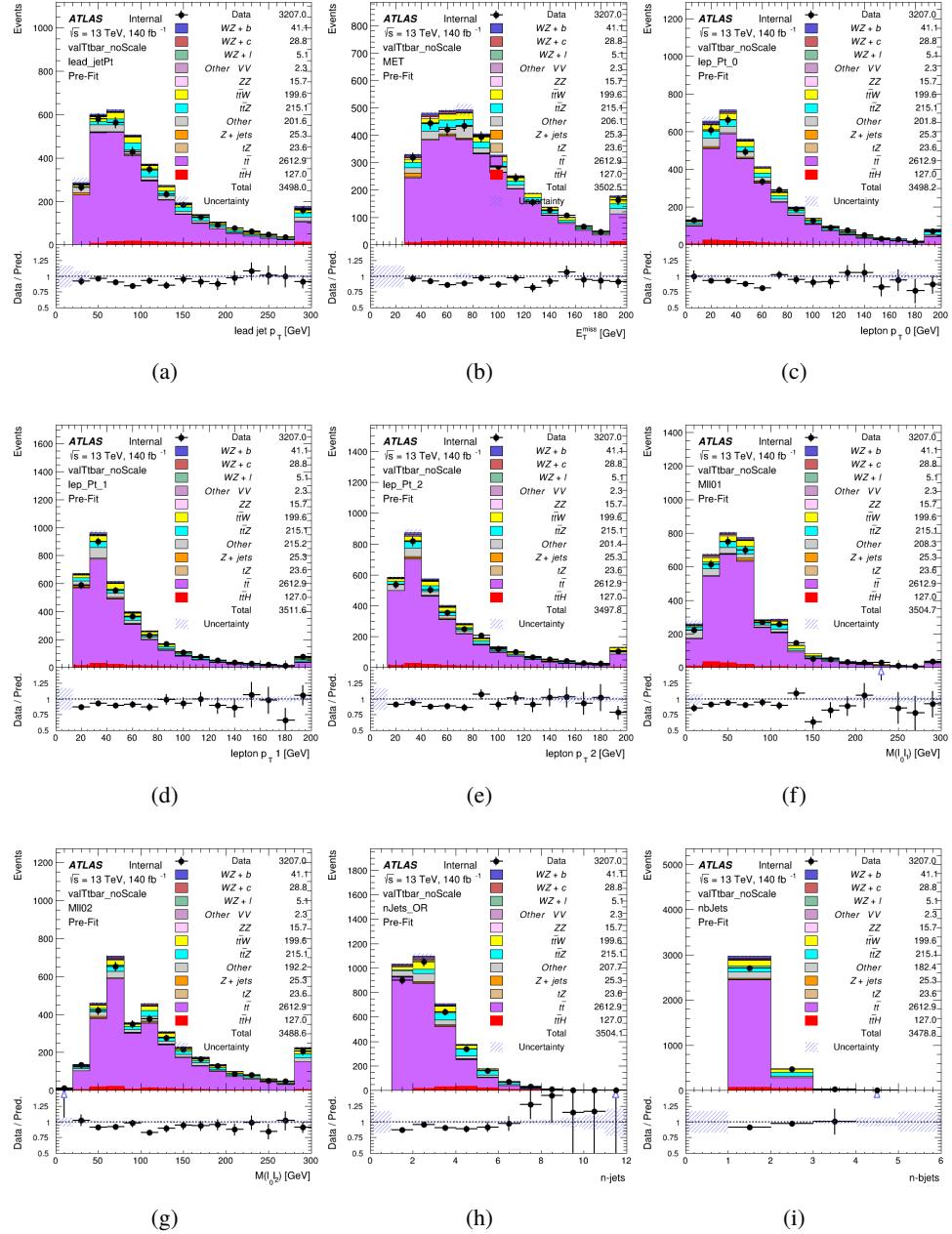


Figure 8.17: Comparisons between the data and MC distributions in the $t\bar{t}$ control region for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) the number of b-tagged jets.

661 The shape of each distribution agrees quite well between data and MC, with a constant
662 offset between the two. This is accounted for by applying a constant correction factor of 0.883
663 to the $t\bar{t}$ MC prediction. Plots showing the kinematics of the $t\bar{t}$ VR after this correction factor
664 has been applied are shown in Figure 12.18.

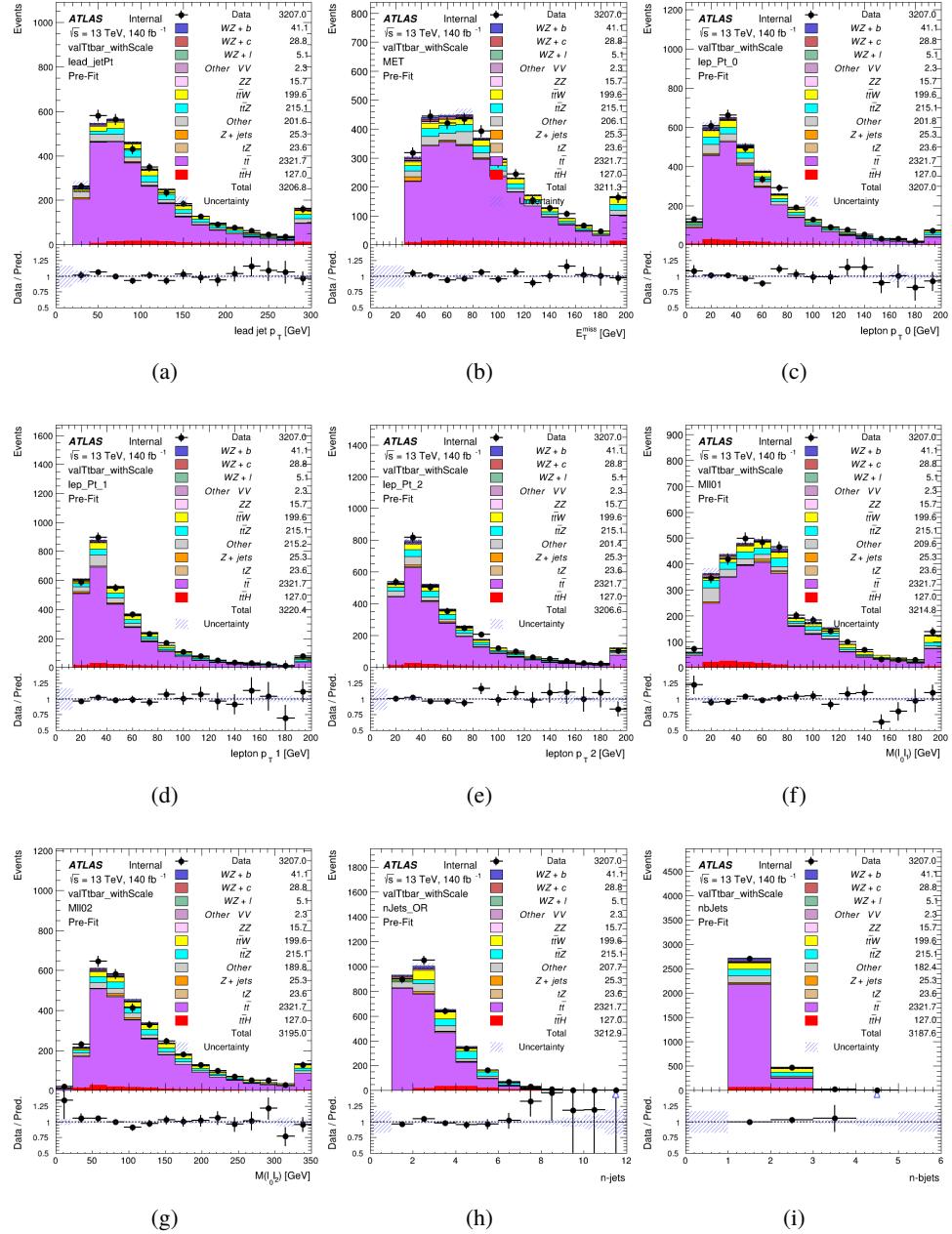


Figure 8.18: Comparisons between the data and MC distributions in the $t\bar{t}$ control region after the correction factor has been applied for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass 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.

665 The modeling is further validated by looking at the yield in the $t\bar{t}$ VR for each DL1r WP,
 666 giving a clearer correspondence to the signal regions used in the fit. Each region shown in Figure
 667 12.19 requires one or more jets pass the listed WP, with no jets passing the next highest WP.

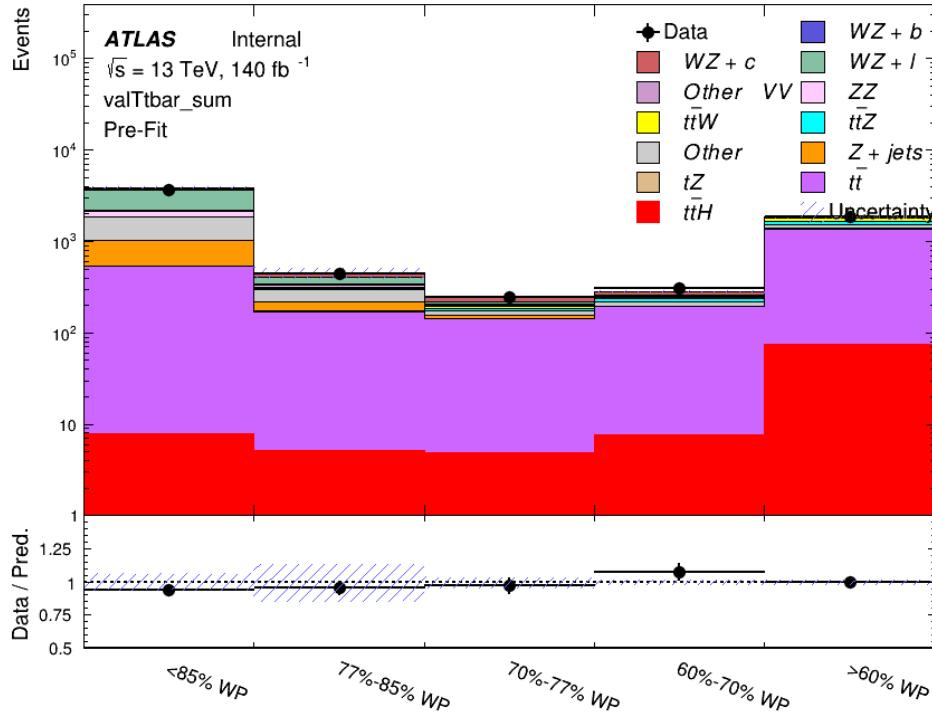


Figure 8.19: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the $t\bar{t}$ VR selection and correction factor have been applied

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

670 **8.3.2 Z+jets Validation**

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

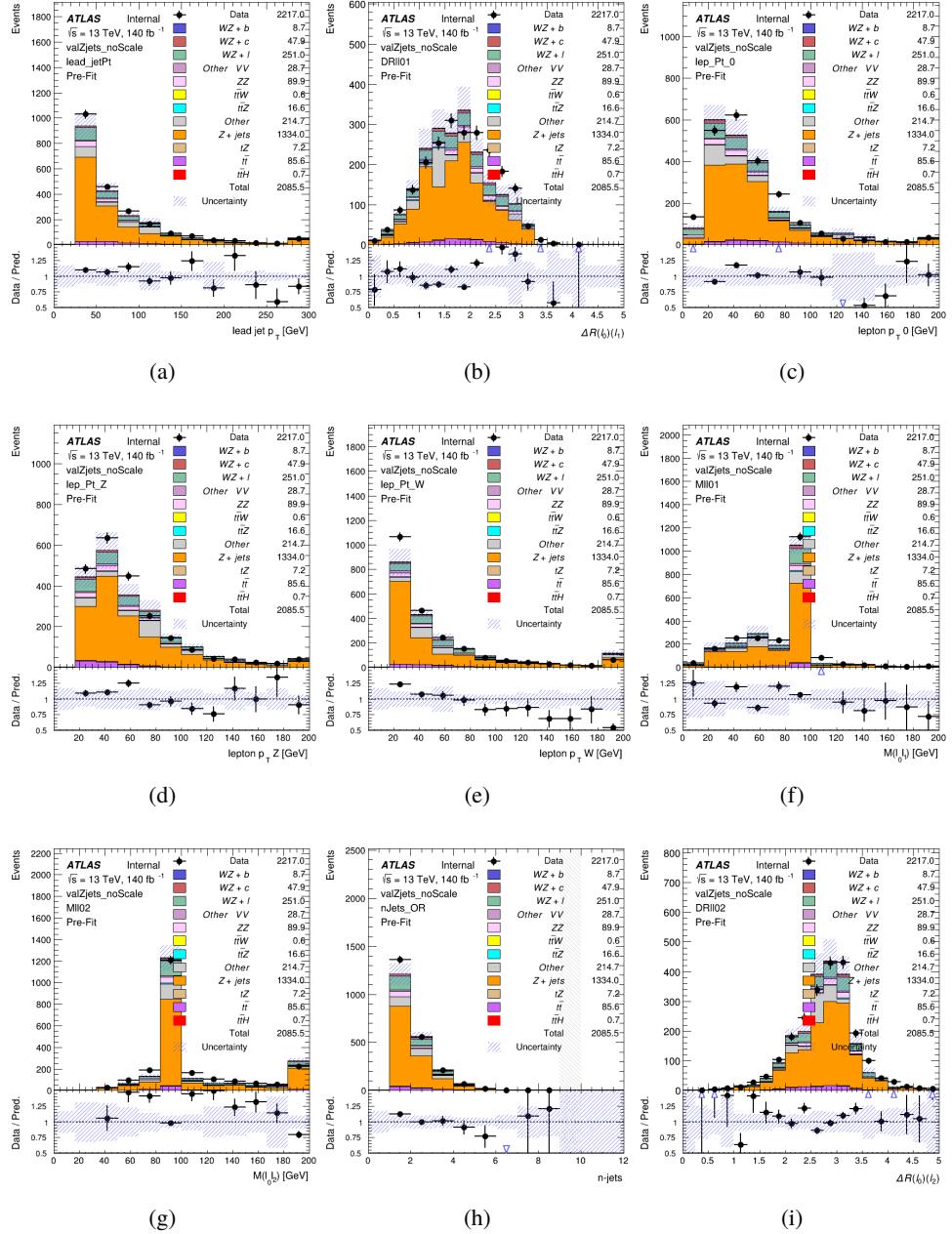


Figure 8.20: Comparisons between the data and MC distributions in the $Z + \text{jets}$ control region for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant 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

678 While there is general agreement between data and MC within statistical uncertainty, the
679 shape of the p_T spectrum of the lepton from the W is found to differ. To account for this
680 discrepancy, a variable correction factor is applied to Z+jets. χ^2 minimization of the lepton 2 p_T
681 spectrum is performed to derive a correction factor of $1.53 - 6.6 \cdot 10^{-6}(\text{lep}_\text{Pt}_W)$. Kinematic
682 plots of the Z + jets control region after this correction factor has been aplied are shown in Figure
683 [12.21](#).

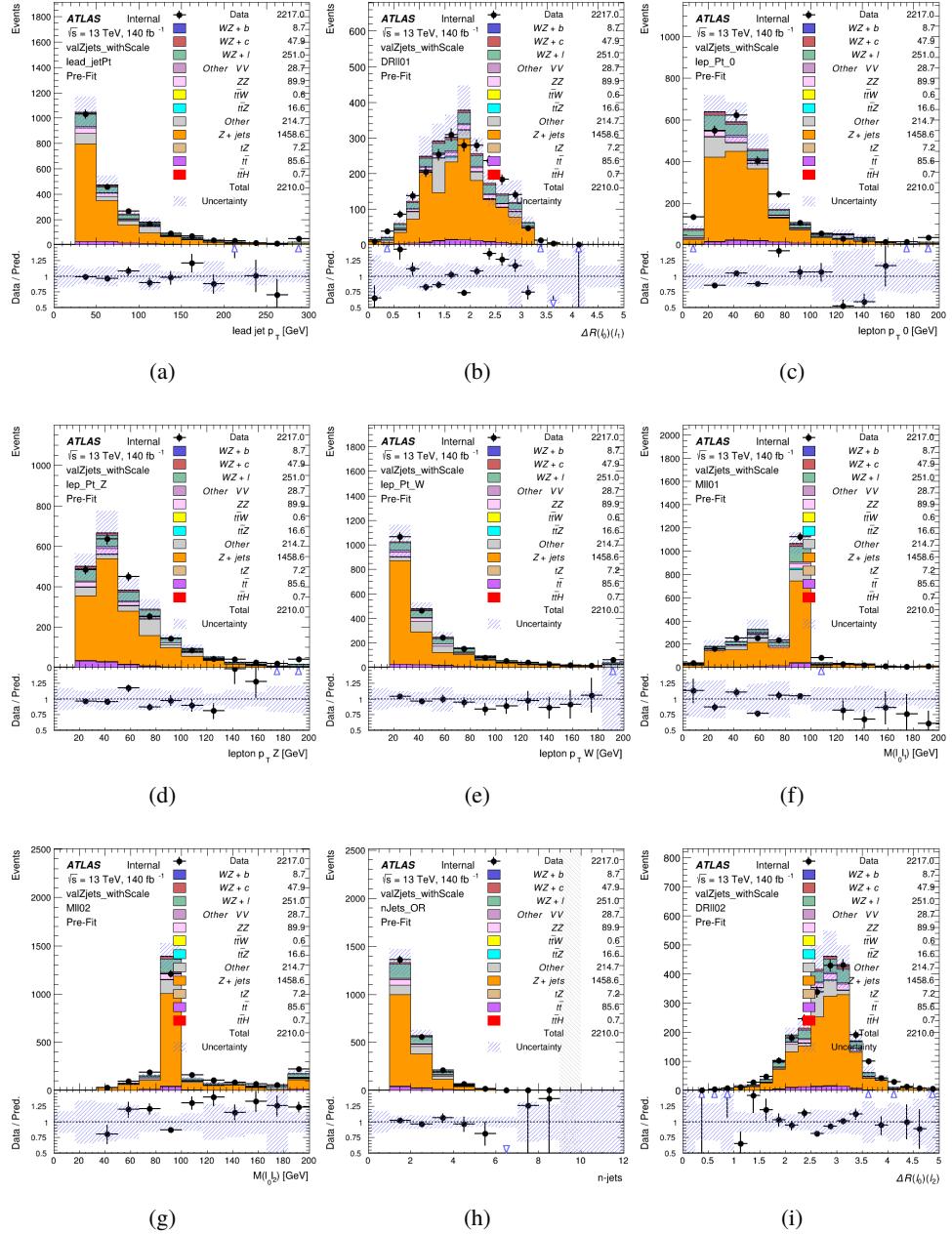


Figure 8.21: Comparisons between the data and MC distributions in the $Z + \text{jets}$ control region after the correction factor has been applied for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS 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

684 The modeling is further validated by looking at the yield in the Z+jets VR for each DL1r
 685 WP, giving a clearer correspondence to the signal regions used in the fit. Each region shown in
 686 Figure 12.22 requires one or more jets pass the listed WP, with no jets passing the next highest
 687 WP.

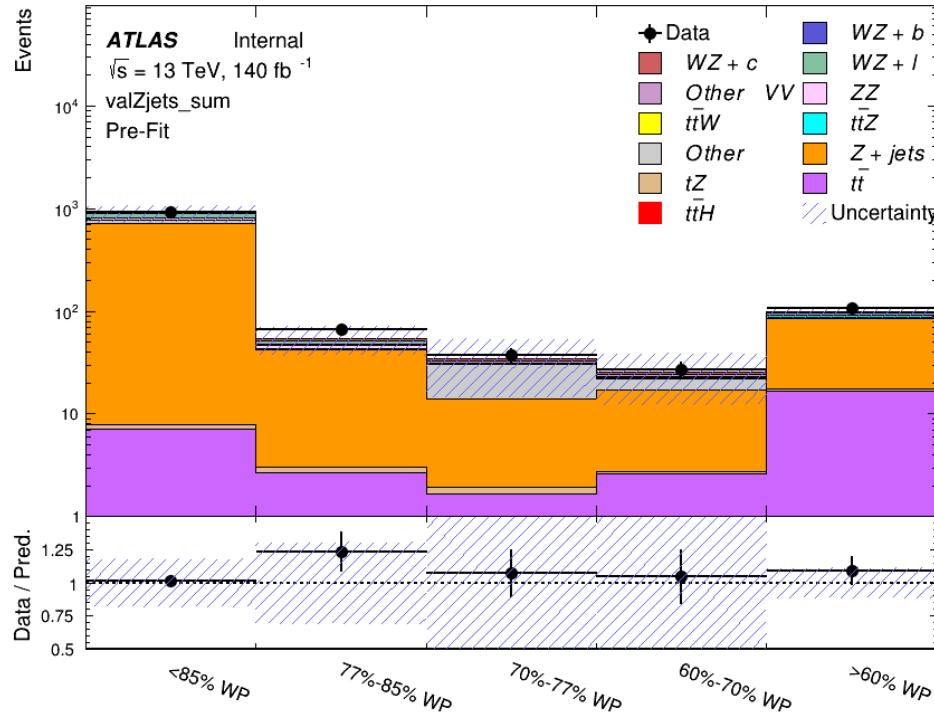


Figure 8.22: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the Z+jets VR selection and correction factor have been applied

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

691 **9 tZ Interference Studies and Separation Multivariate Analysis**

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

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

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

701 **9.1 Top Mass Reconstruction**

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

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

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

$$712 \quad (p_l + p_\nu)^2 = m_W^2$$

713 Expanding this out into components, this equation gives:

$$714 \quad \sqrt{p_{T\nu}^2 + p_{z\nu}^2} E_l = \frac{m_W^2 - m_l^2}{2} + p_{T\nu}(p_{lx}\cos\phi_\nu + p_{ly}\sin\phi_\nu) + p_{lz}p_{\nu z}$$

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

720 The reconstructed top mass distribution for tZ and WZ + b can be seen in figure 13.1.

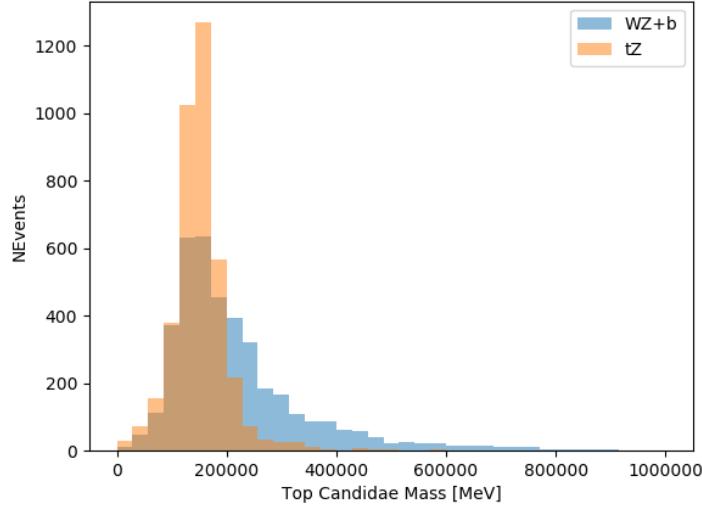


Figure 9.1: Reconstructed top mass distributions for tZ and WZ + b, measured in MeV.

721 9.2 tZ BDT

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

724 • The invariant mass of the reconstructed top candidate

725 • p_T of each of the leptons, jet

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

727 • E_T^{miss}

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

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

730 The training samples included only events meeting the requirements of the 1-jet, $>60\%$
731 region, i.e. passing all the selection described in section 12 and having exactly one jet which
732 passes the tightest (60%) DL1r working point.

733 The distributions of a few of these features for both signal and background is shown in
734 figure 13.2.

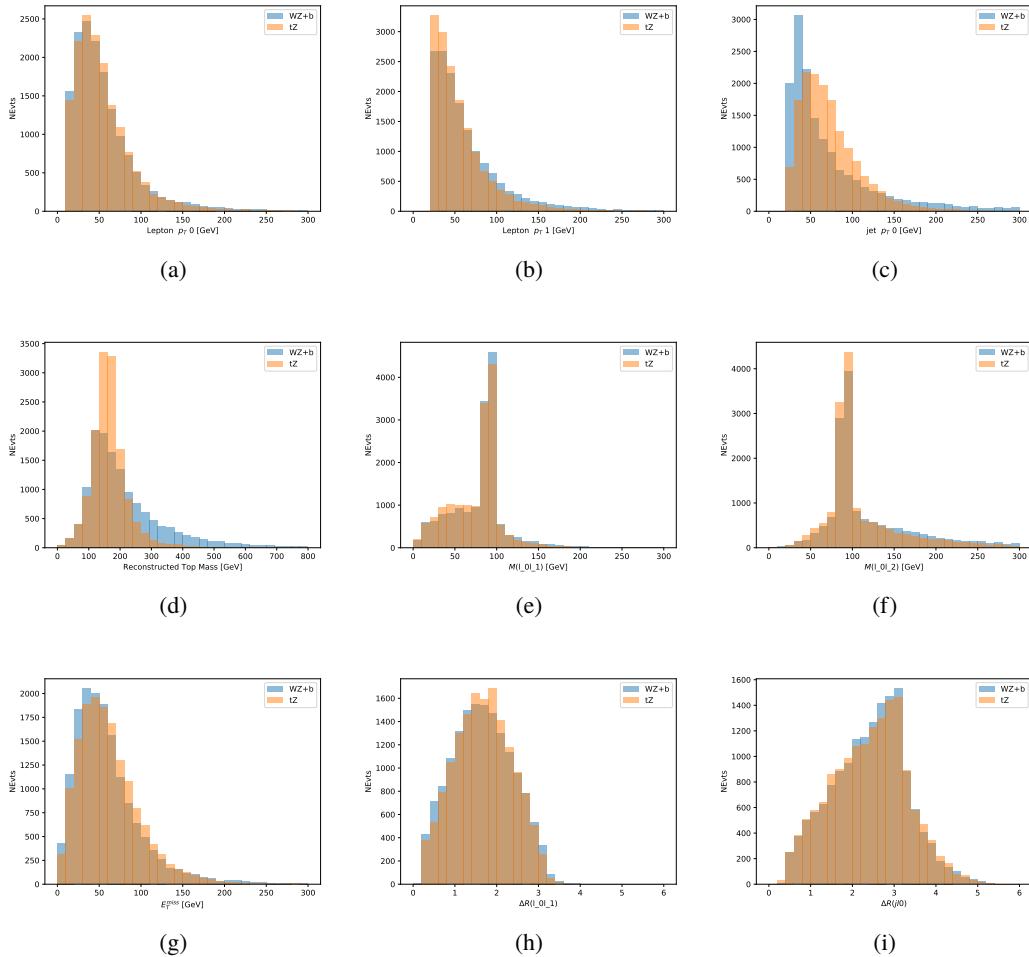


Figure 9.2: Distribution of input features of the BDT for signal (WZ) and background (tZ). Both are scaled to an equal number of events. (a), (b) and (c) show the p_T of lepton 0, lepton 1, and the jet, (d) show the reconstructed top mass, (e) and (f) show the invariant mass of leptons 0 and 1, leptons 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.

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

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

740 The results of the BDT training are shown in figure 13.3. The output scores for both signal
 741 and background events is shown on the left. The right shows the receiving operating characteristic
 742 (ROC) curve that results from the MVA. The ROC curve represents the background rejection
 743 as a function of signal efficiency, where each point on the curve represents a different response
 744 score. The ROC curve of the BDT is compared to the performance of using an optimal set of flat
 745 selections on the same set of input variables.

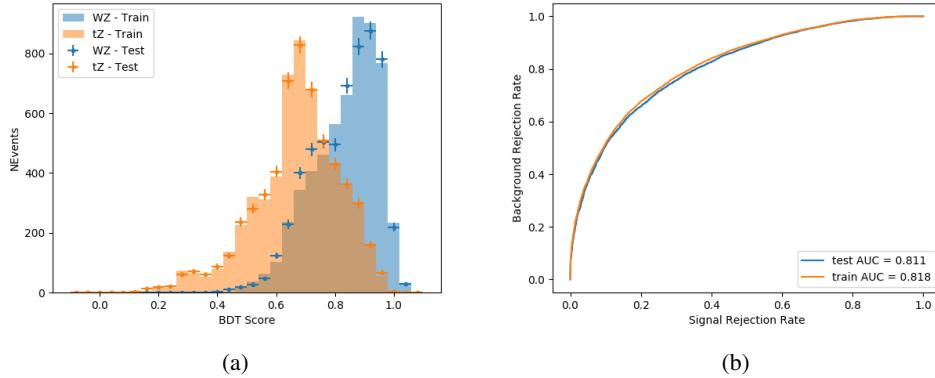


Figure 9.3: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

746 The relative important of each input feature in the model, measured by how often they
 747 appeared in the decision trees, is shown in figure 13.4.

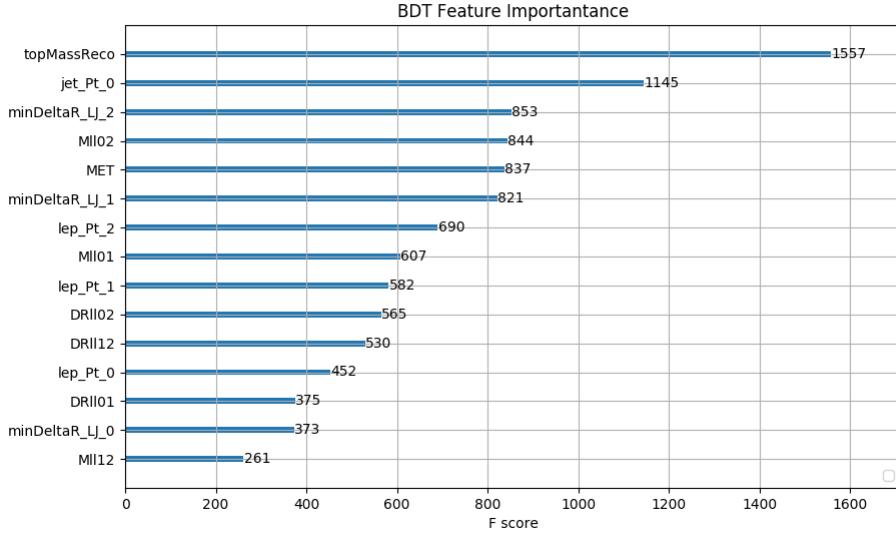


Figure 9.4: Relative importance of each input feature in the model.

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

754 10 Data and Monte Carlo Samples

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

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

- 761 • at least two light leptons within a range $|\eta| < 2.6$, with leading lepton $p_T > 15$ GeV and
762 subleading lepton $p_T > 5$ GeV
- 763 • at least one light lepton with $p_T > 15$ GeV within a range $|\eta| < 2.6$, and at least two hadronic
764 taus with $p_T > 15$ GeV.

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

768 **10.1 Data Samples**

769 The study uses a sample of proton-proton collision data collected by the ATLAS detector from
770 2015 through 2018 at an energy of $\sqrt{s} = 13$ TeV, which represents an integrated luminosity of
771 139 fb^{-1} . This data set was collected with a bunch crossing rate of 25 ns. All data used in this
772 analysis was verified by data quality checks, having been included in the following Good Run
773 Lists:

- 774 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
775 _PHYS_StandardGRL_All_Good_25ns.xml
- 776 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
777 _PHYS_StandardGRL_All_Good_25ns.xml
- 778 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL
779 _All_Good_25ns_Triggerno17e33prim.xml
- 780 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL
781 _All_Good_25ns_Triggerno17e33prim.xml

782 Runs included from the AllYear period containers are included.

783 **10.2 Monte Carlo Samples**

784 Several different generators were used to produce Monte Carlo simulations of the signal and
785 background processes. For all samples, the response of the ATLAS detector is simulated using
786 Geant4. The WZ signal samples are simulated using Sherpa 2.2.2 [[sherpa](#)]. Specific information
787 about the Monte Carlo samples being used can be found in Table 28. A list of the specific samples
788 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.

789 11 Object Reconstruction

- 790 All regions defined in this analysis share a common lepton, jet, and overall event preselection.
- 791 The selection applied to each physics object is detailed here; the event preselection, and the
- 792 selection used to define the various fit regions, is described in Section 12.

793 **11.1 Trigger**

794 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.

795 **11.2 Light leptons**

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

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

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

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

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

801 $|\eta| < 2.5$. Candidates in the transition region between different electromagnetic calorimeter
802 components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant com-
803 bining shower shape and track information is used to distinguish real electrons from hadronic
804 showers (fake electrons). To further reduce the non-prompt electron contribution, the track is
805 required to be consistent with originating from the primary vertex; requirements are imposed
806 on the transverse impact parameter significance ($|d_0|/\sigma_{d_0} < 5$) and the longitudinal impact
807 parameter ($|\Delta z_0 \sin \theta_\ell| < 0.5 \text{ mm}$).

808 Muon candidates are reconstructed by combining inner detector tracks with track segments
809 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
810 $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$. The longitudinal impact parameter is the same for both electrons
811 and muons, while muons are required to pass a slightly tighter transverse impact parameter,
812 $|d_0|/\sigma_{d_0} < 3$.

813 All leptons are required to be isolated, as defined through the standard PLVLoose working
814 point supported by combined performance groups. Leptons are additionally required to pass a
815 non-prompt BDT selection developed by the $t\bar{t}H$ multilepton/ $t\bar{t}W$ analysis group. This BDT is
816 described in detail in [**ttW_140**]. Optimized working points and scale factors for this BDT are
817 taken from that analysis.

818 11.3 Jets

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

827 11.4 B-tagged Jets

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

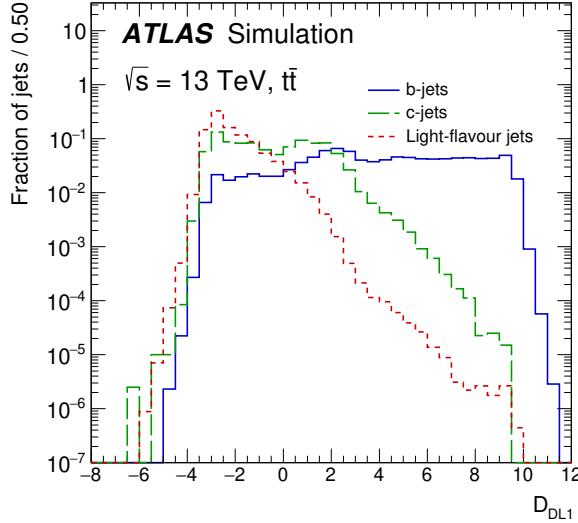


Figure 11.1: Output distribution of the DL1r algorithm for b-jets, charm jets, and light jets

833 From the output of the BDT, working points (WPs) are developed based on the efficiency
 834 of truth b-jets at particular values of the DL1r algorithm. These WPs are taken from the March
 835 2020 CDI file, 2020-21-13TeV-MC16-CDI-2020-03-11_v2.root. The working points used in
 836 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

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

841 60% WP fall into another, and looking at the full psuedo-continuous DL1r spectrum of the jets,
842 additional information can be gained. The psuedo-continuous b-tag spectrum is used in this case
843 to separate out WZ + b, WZ + charm, and WZ + light.

844 **11.5 Missing transverse energy**

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

852 **11.6 Overlap removal**

853 To avoid double counting objects and remove leptons originating from decays of hadrons, overlap
854 removal is performed in the following order: any electron candidate within $\Delta R = 0.1$ of another
855 electron candidate with higher p_T is removed; any electron candidate within $\Delta R = 0.1$ of a muon
856 candidate is removed; any jet within $\Delta R = 0.3$ of an electron candidate is removed; if a muon
857 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
858 is kept and the muon is removed.

859 This algorithm is applied to the preselected objects. The overlap removal procedure is

860 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.

861 12 Event Selection and Signal Region Definitions

862 Event are required to pass a preselection described in Section 12.1 and summarized in Table 14.

863 Those that pass this preselection are divided into various fit regions described in Section 12.2,

864 based on the number of jets in the event, and the b-tag score of those jets.

865 12.1 Event Preselection

866 Events are required to include exactly three reconstructed light leptons passing the requirement

867 described in 11.2, which have a total charge of ± 1 . As the opposite sign lepton is found to

868 be prompt the vast majority of the time [ttH_paper], it is required to be loose and isolated,

869 as defined though the standard PLVLoose working point supported by combined performance

870 groups. The same sign leptons are required to be very tightly isolated, as per the recommended

871 PLVTight.

872 The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose
 873 charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.
 874 the smallest ΔR , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton
 875 0 is required to have $p_T > 10 \text{ GeV}$, while the same sign leptons, 1 and 2, are required to have
 876 $p_T > 20 \text{ GeV}$ to reduce the contribution of non-prompt leptons.

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

881 An additional requirement is placed on the missing transverse energy, $E_T^{\text{miss}} > 20 \text{ GeV}$,
 882 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
 883 missing transverse energy, and l_{other} is the lepton not included in the Z-candidate.

884 Events are required to have one or two reconstructed jets passing the selection described
 885 in Section 11.3. Events with more than two jets are rejected in order to reduce the contribution
 886 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 14: Summary of the selection applied to events for inclusion in the fit

887 The event yields in the preselection region for both data and Monte Carlo are summarized
 888 in Table 12.1, which shows good agreement between data and Monte Carlo, and demonstrates that
 889 this region consists primarily of WZ events. The WZ events are split into WZ + b, WZ + c, and
 890 WZ + l based on the truth flavor of the associated jet in the event. Specifically, this determination
 891 is made based on the HadronConeExclTruthLabelID of the jet, as recommended by the b-
 892 tagging working group [BtagWG]. In this ordering b-jet supersedes charm, which supersedes
 893 light. That is, WZ + l events contain no charm and no b jets at truth level, WZ + c contain at
 894 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}

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

897 Simulations are further validated by comparing the kinematic distributions of the Monte
898 Carlo with data, which are shown in Figure 12.1. Here, bins with 5% or more WZ+b are
899 blinded.

WZ Fit Region - Inclusive

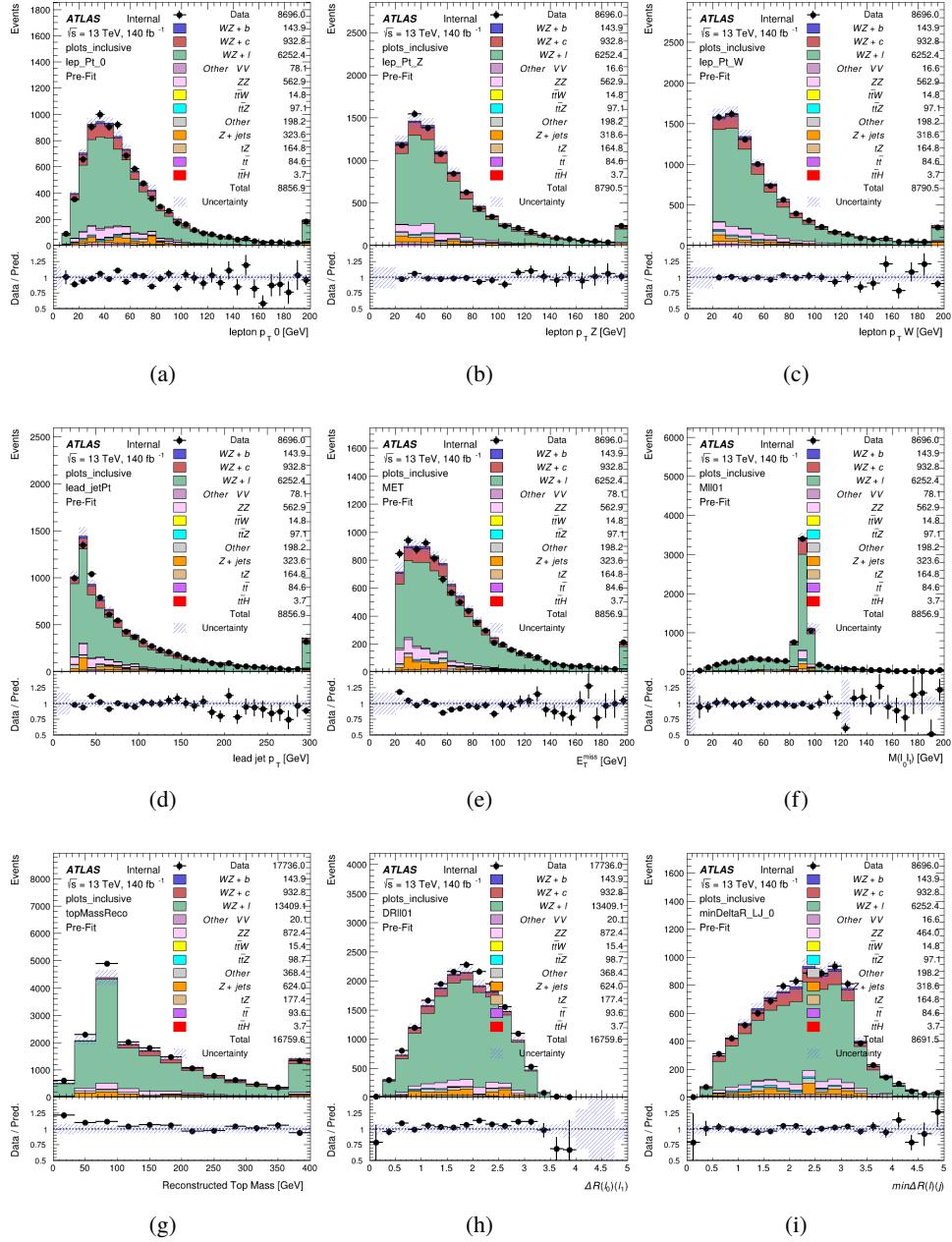


Figure 12.1: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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.

900 **12.2 Fit Regions**

901 Once preselection has been applied, the remaining events are categorized into one of twelve
 902 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%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} >= 1, n\text{Jets_DL1r_77}=0$
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	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} >= 1, tZ \text{ BDT} < 0.725$

903 The working points discussed in Section 11.4 are used to separate events into fit regions
 904 based on the highest working point reached by a jet in each event. Because the background
 905 composition differs significantly based on the number of b-jets, events are further subdivided
 906 into 1-jet and 2-jet regions in order to minimize the impact of background uncertainties.

907 An unfolding procedure is performed to account for differences in the number of recon-
 908 structed jets compared to the number of truth jets in each event. The `AntiKt4TruthDressedWZJets`
 909 truth jet collection is used to make this determination. In order to account for migration of WZ+1-
 910 jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal samples are
 911 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 12.2. 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 21.

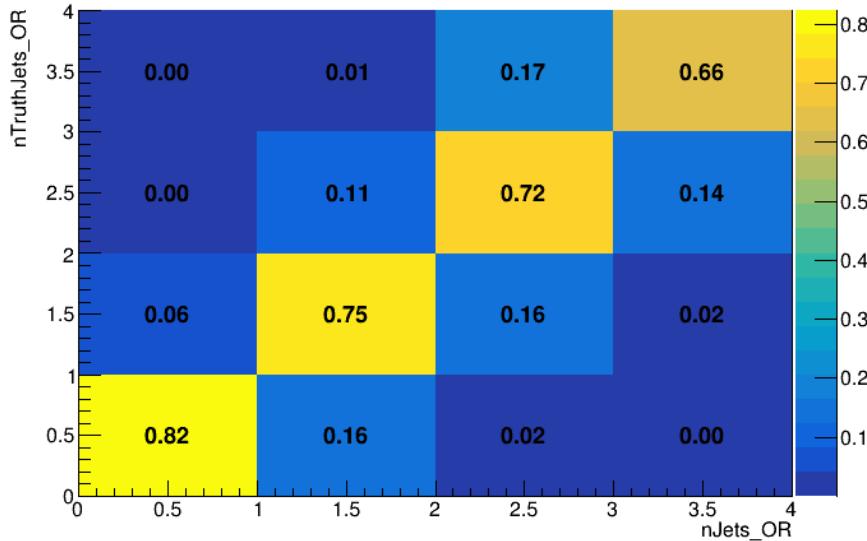


Figure 12.2: Number of truth jets compared to the number of reconstructed jets in each WZ event falling in the preselection region. Each row is normalized to unity.

An additional tZ control region is created based on the BDT described in Section 13. 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

922 various kinematic distributions. These plot are shown in Figures 12.3-12.16.

WZ Fit Region - 1j Inclusive

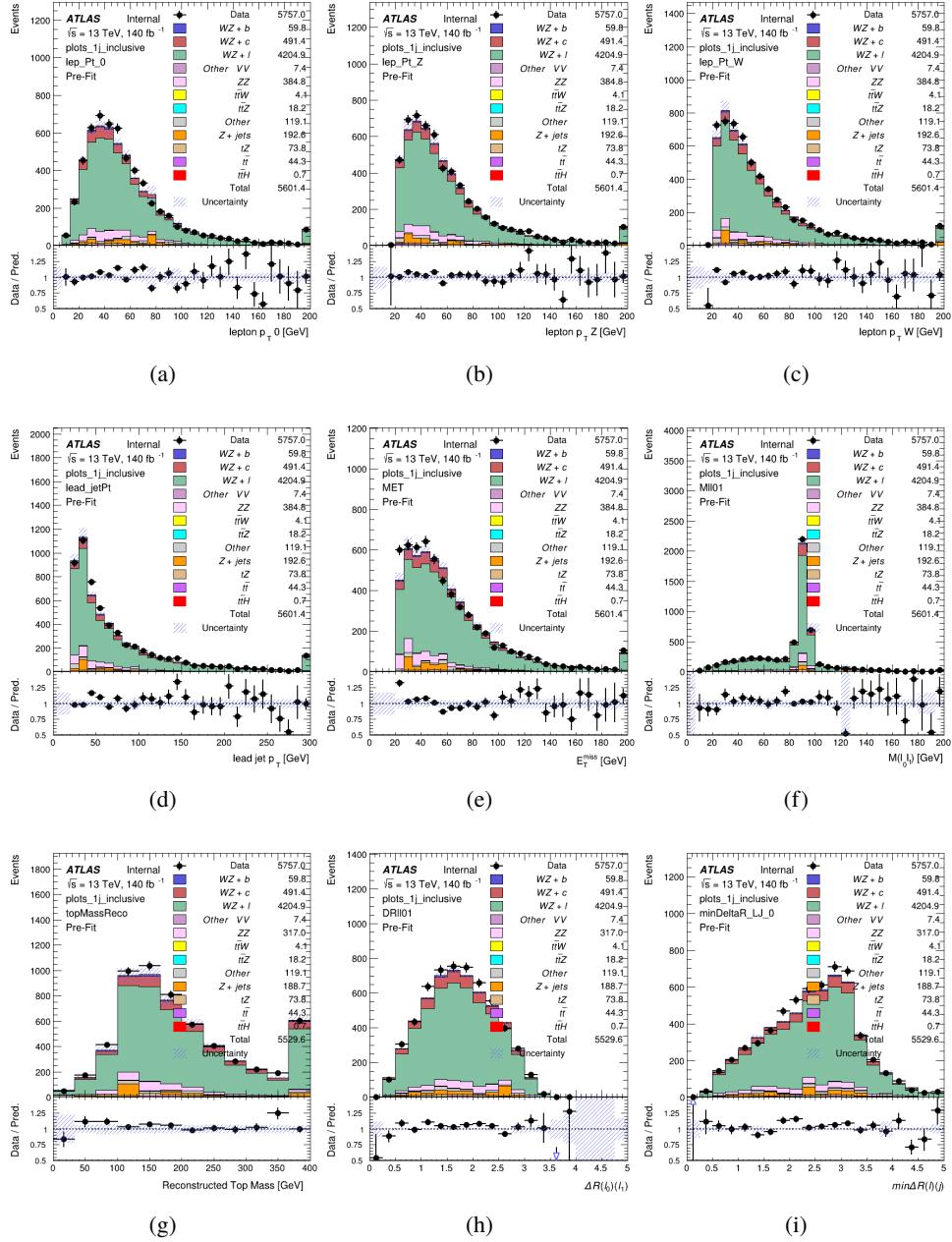


Figure 12.3: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

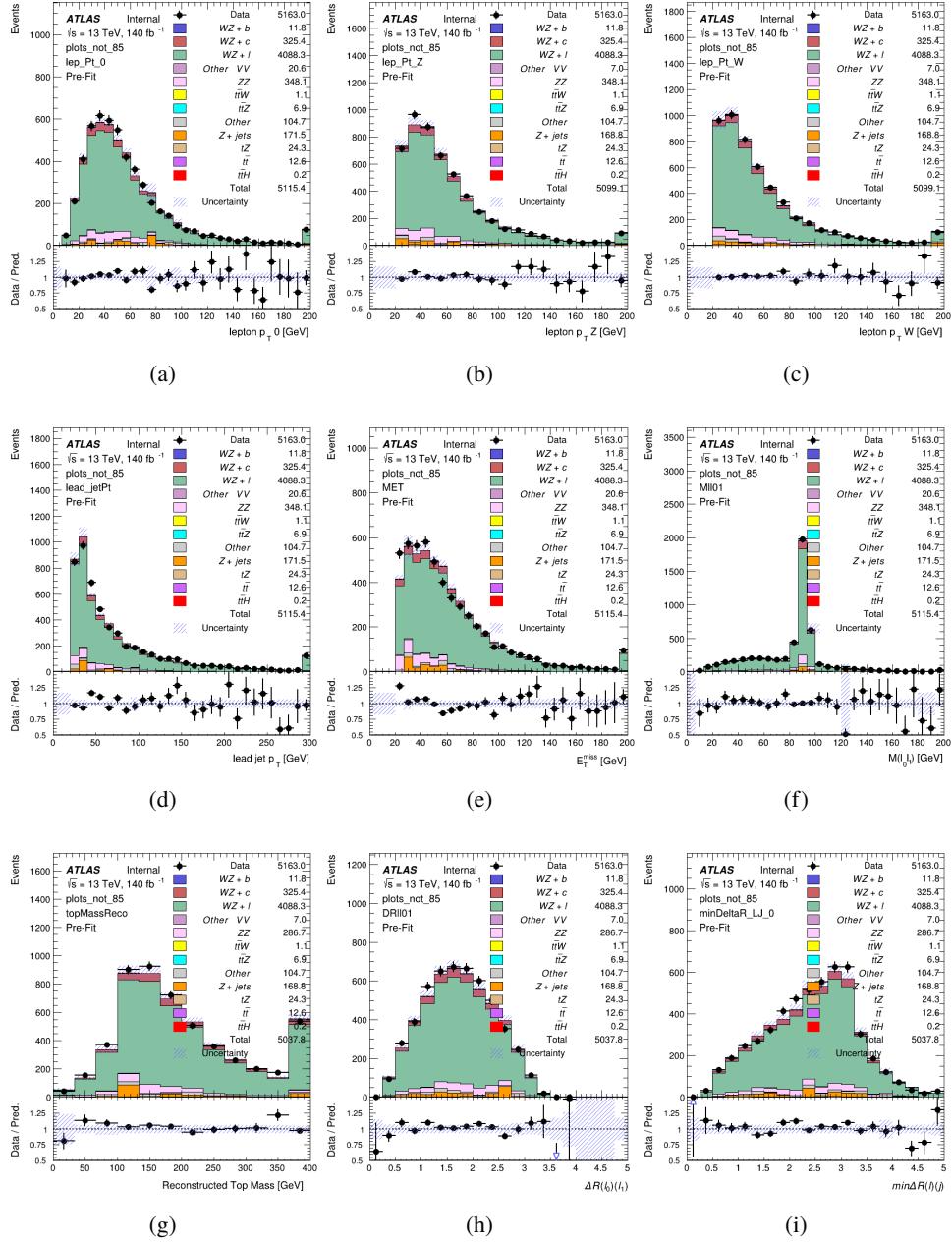


Figure 12.4: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

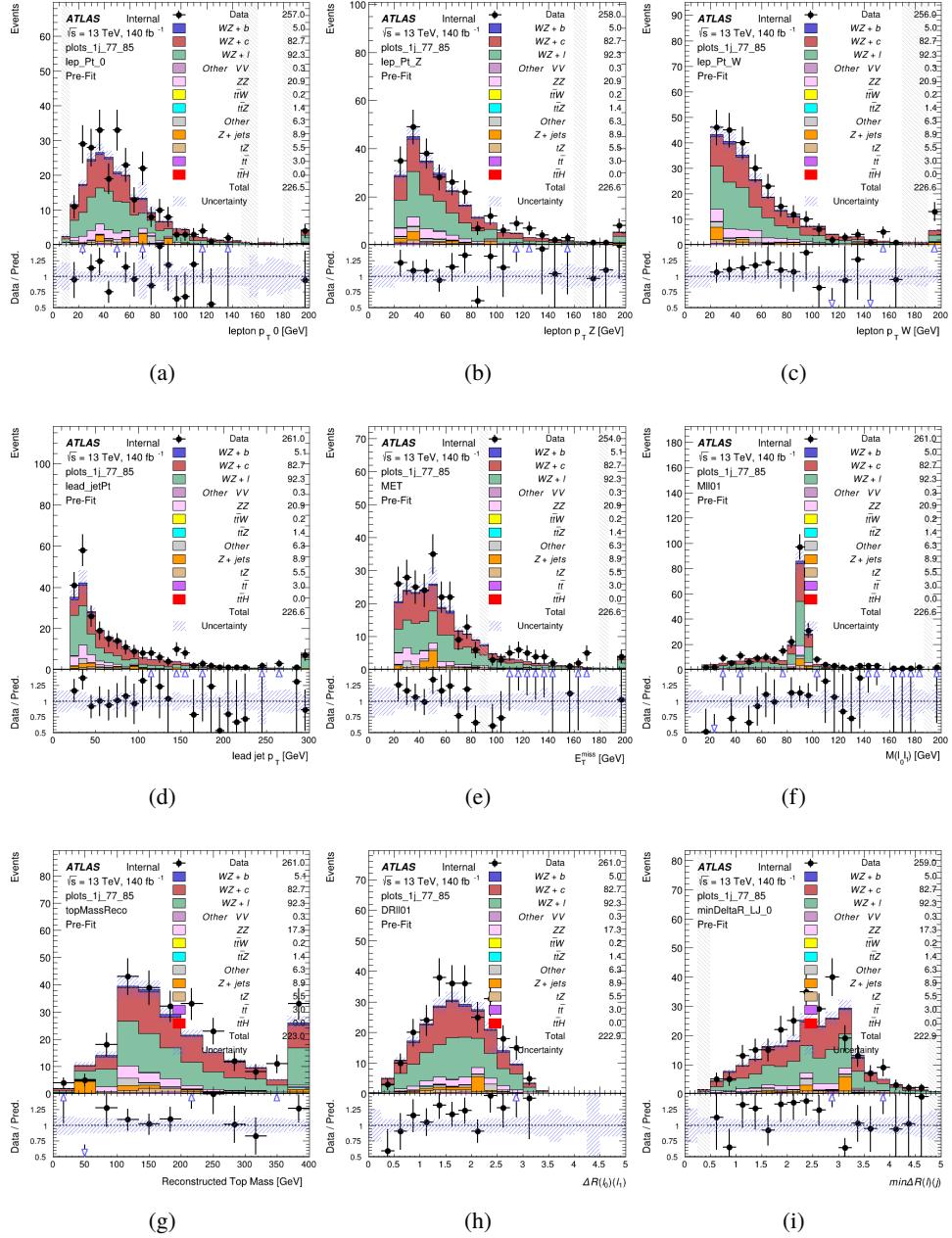


Figure 12.5: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

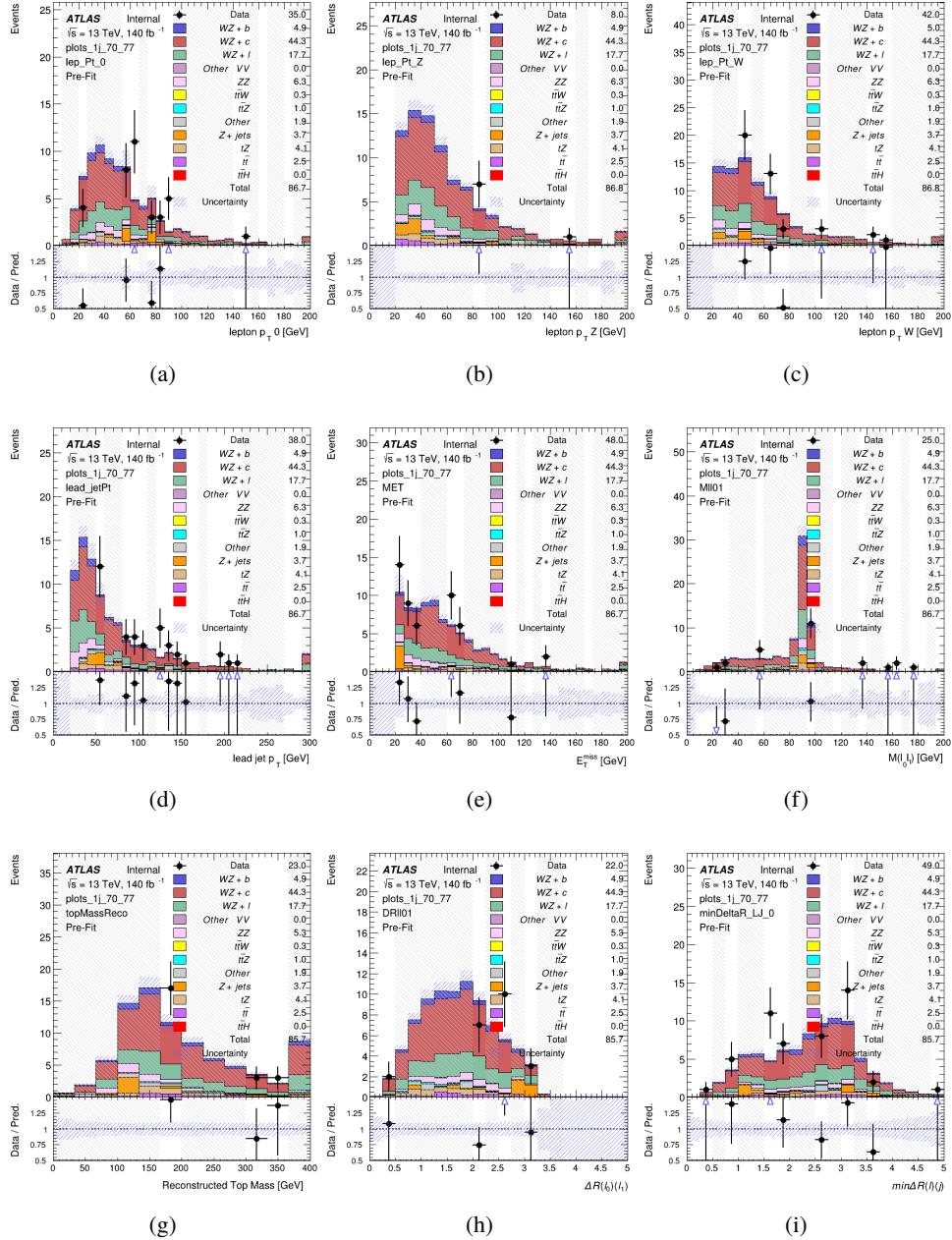


Figure 12.6: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

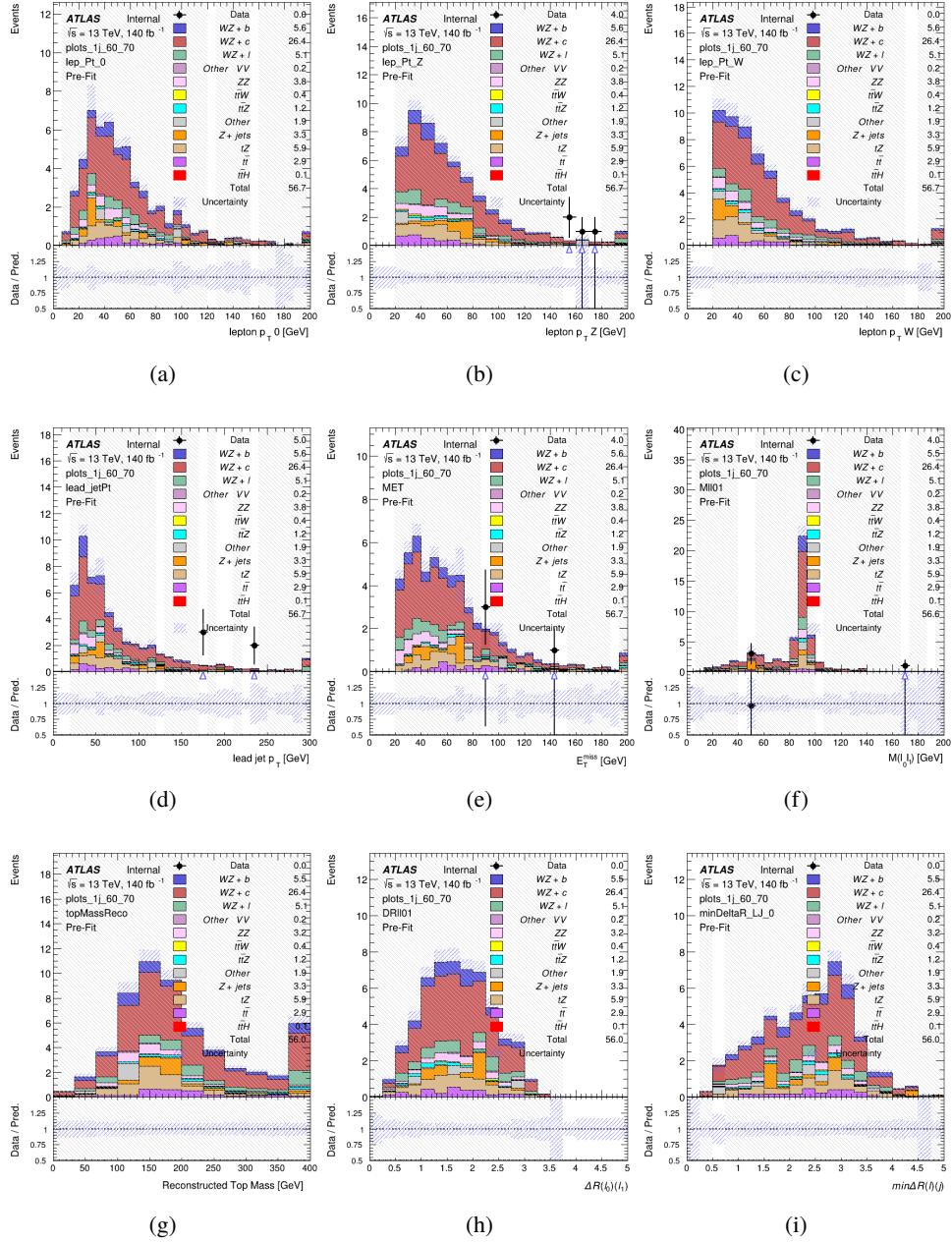


Figure 12.7: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

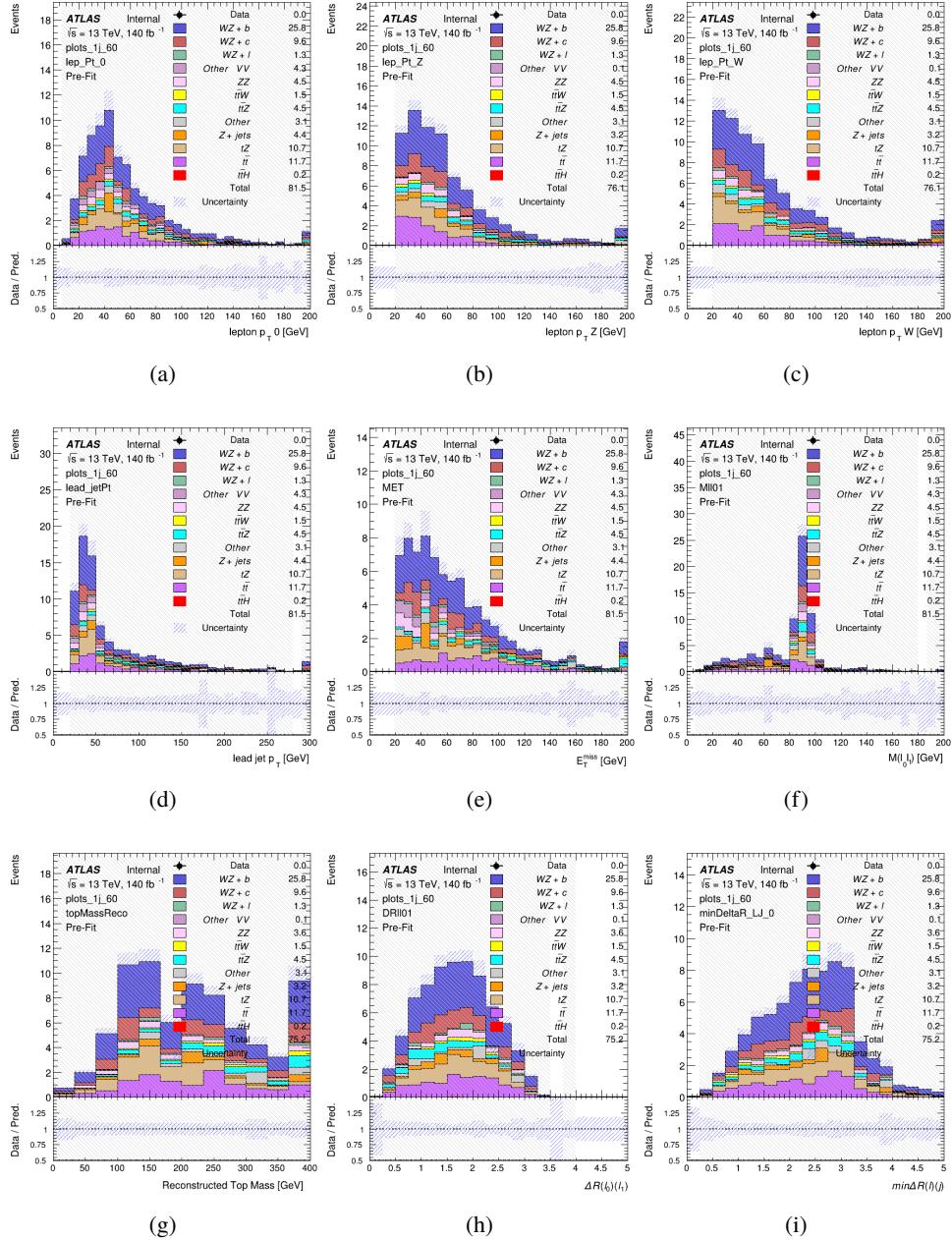


Figure 12.8: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

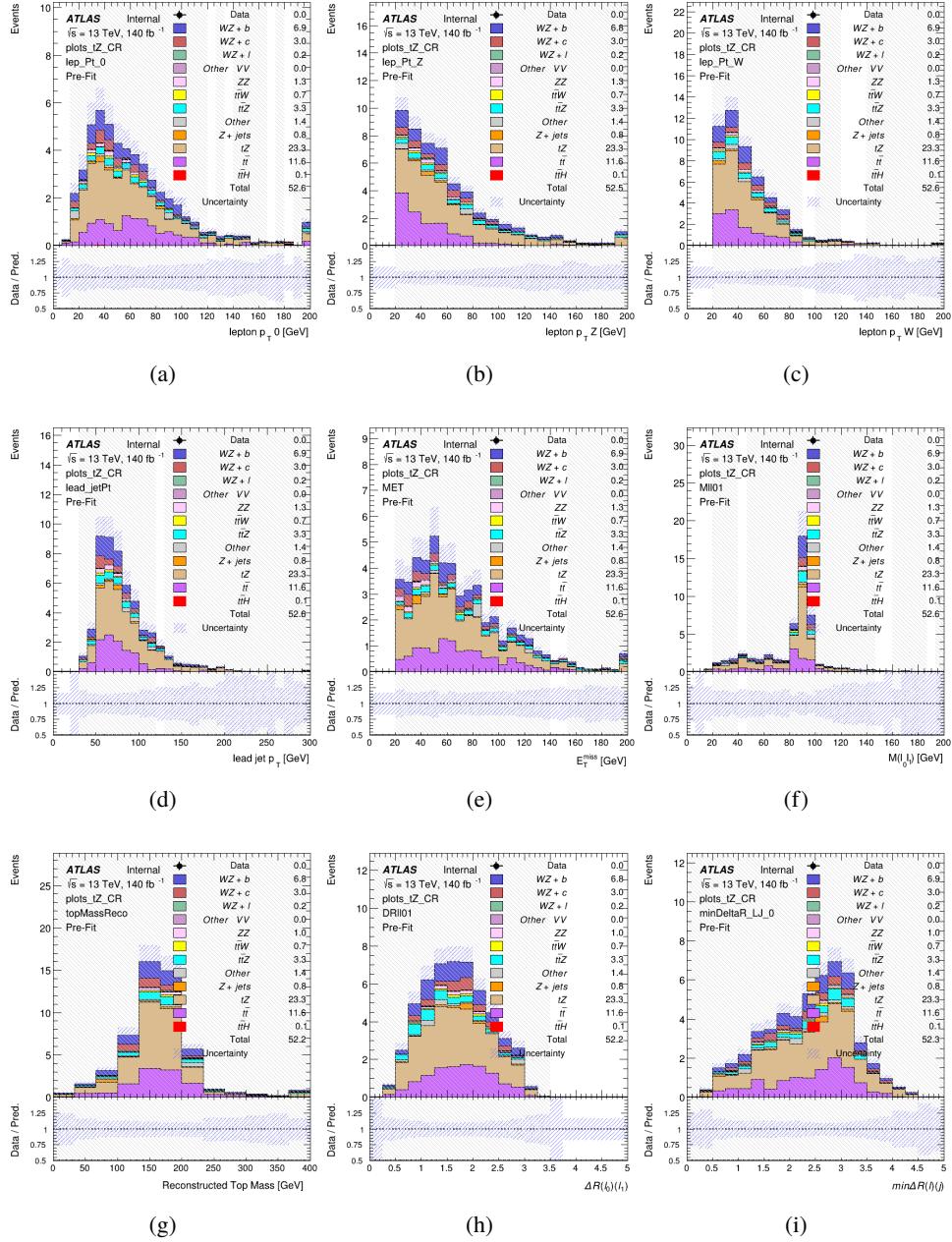


Figure 12.9: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

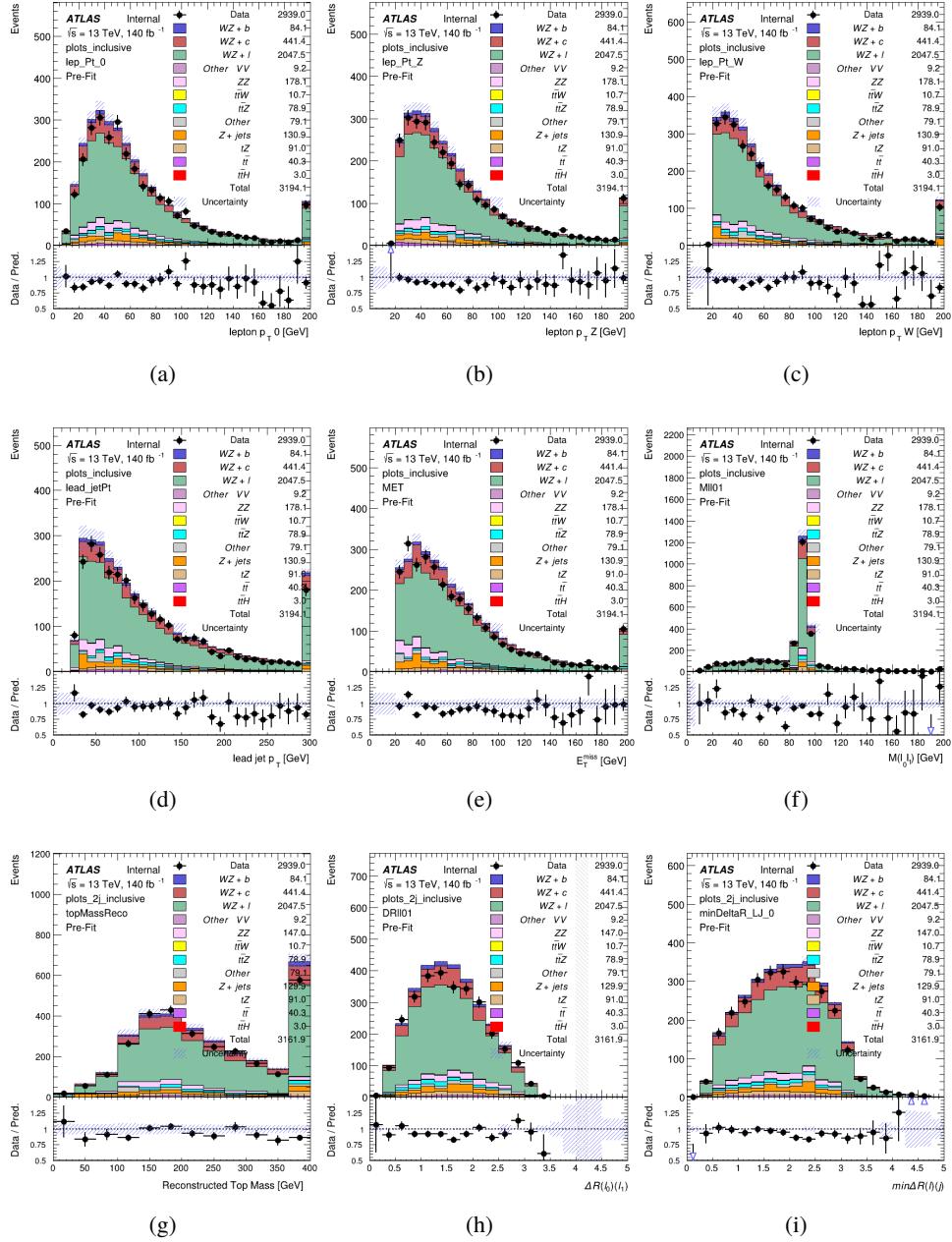


Figure 12.10: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

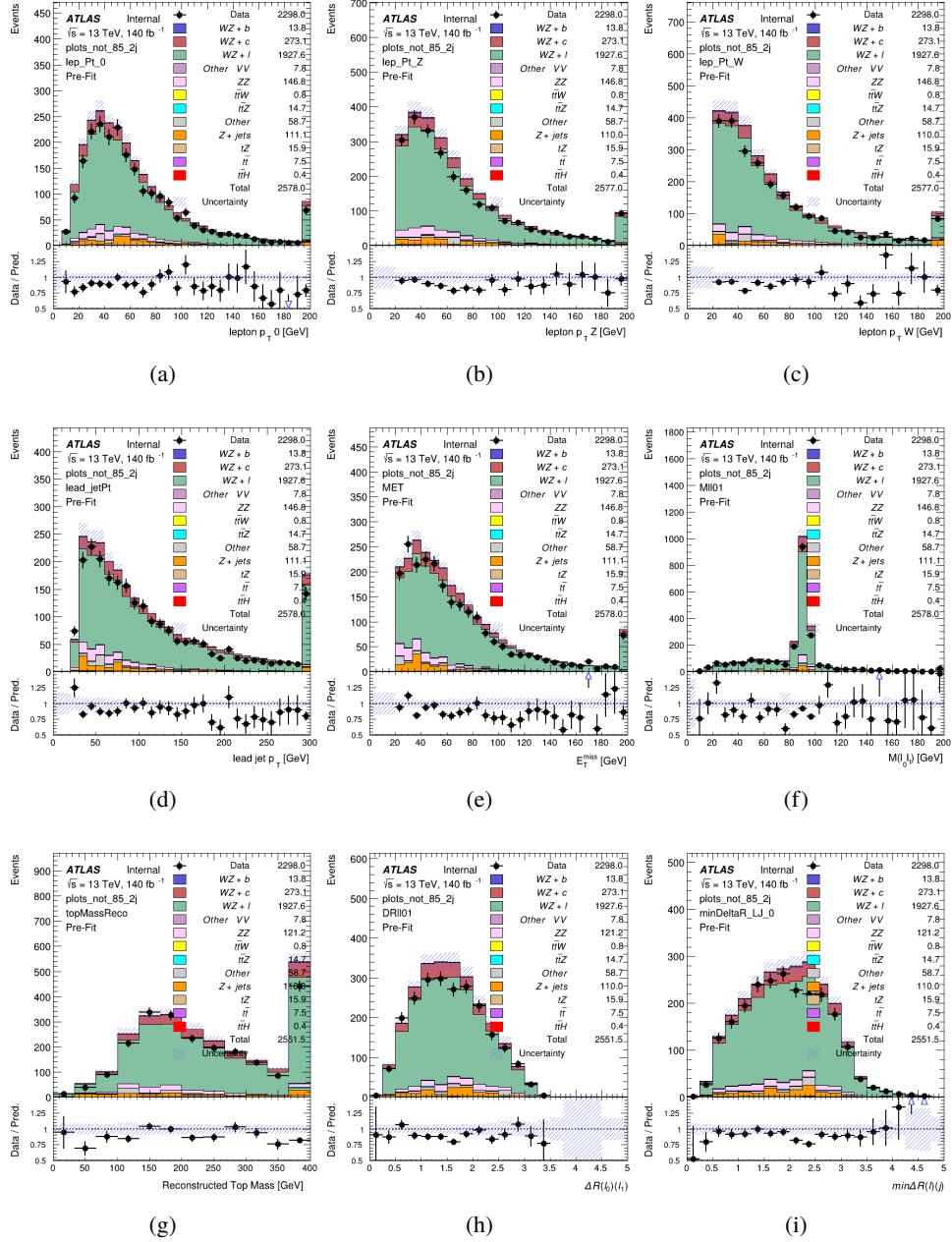


Figure 12.11: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

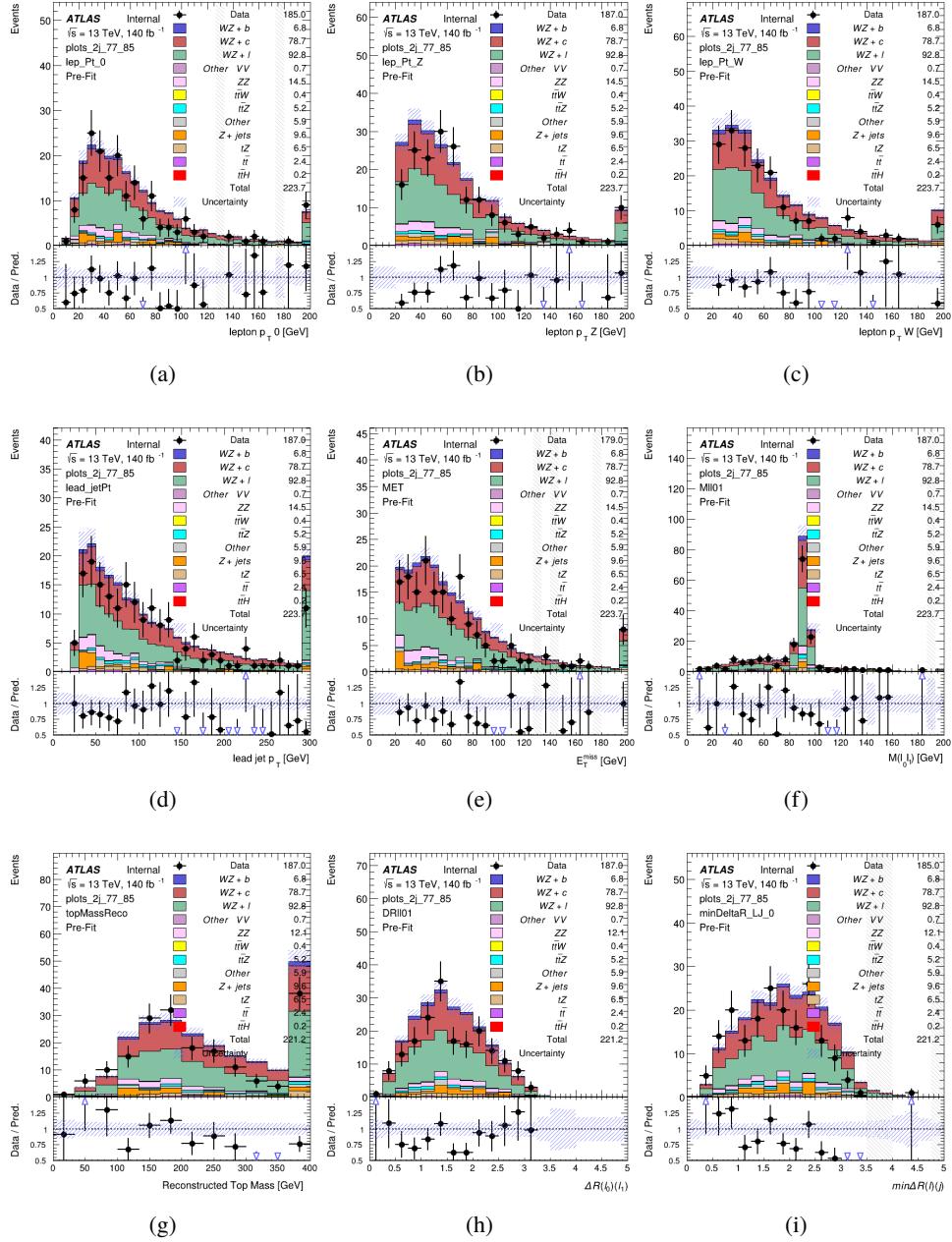


Figure 12.12: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

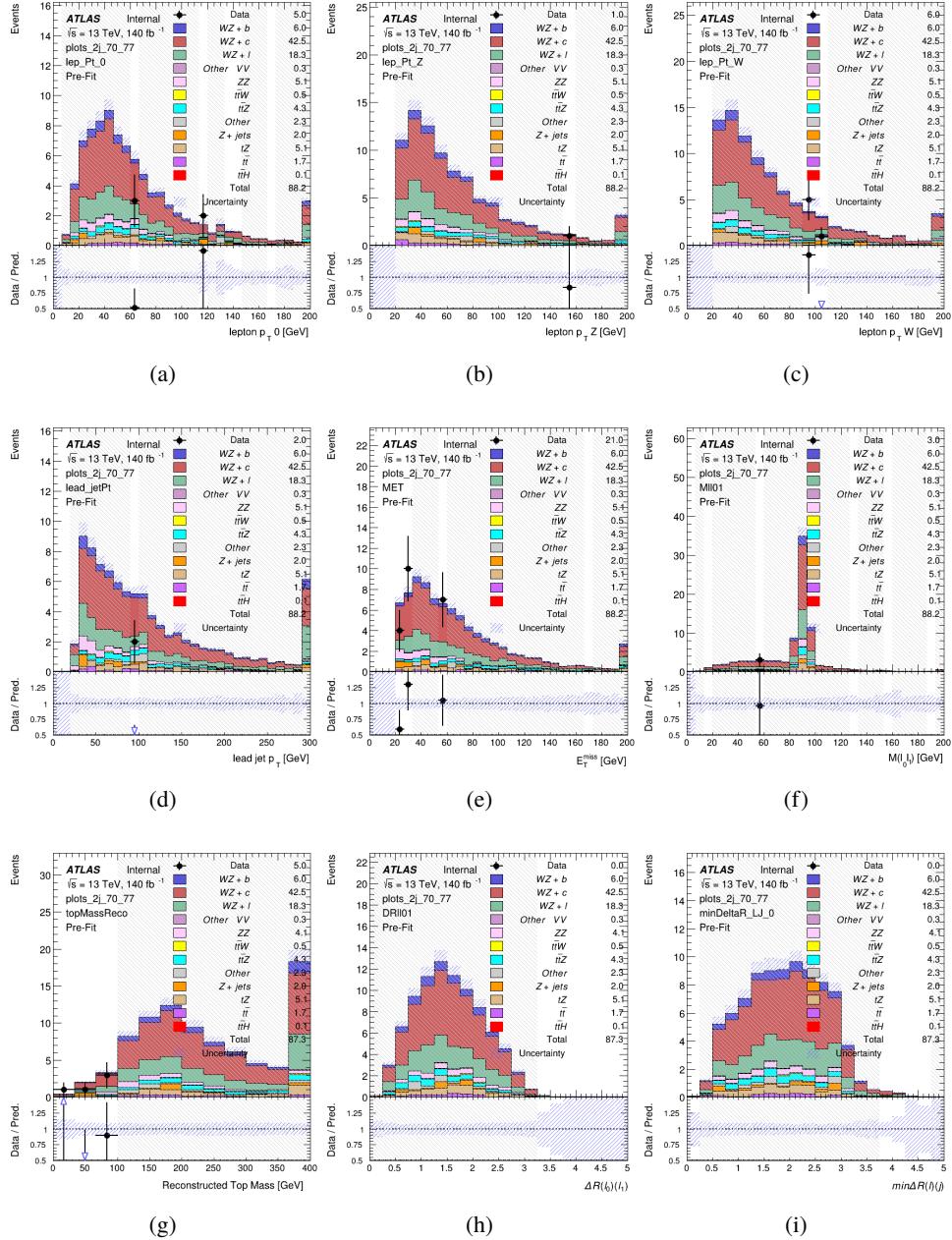


Figure 12.13: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

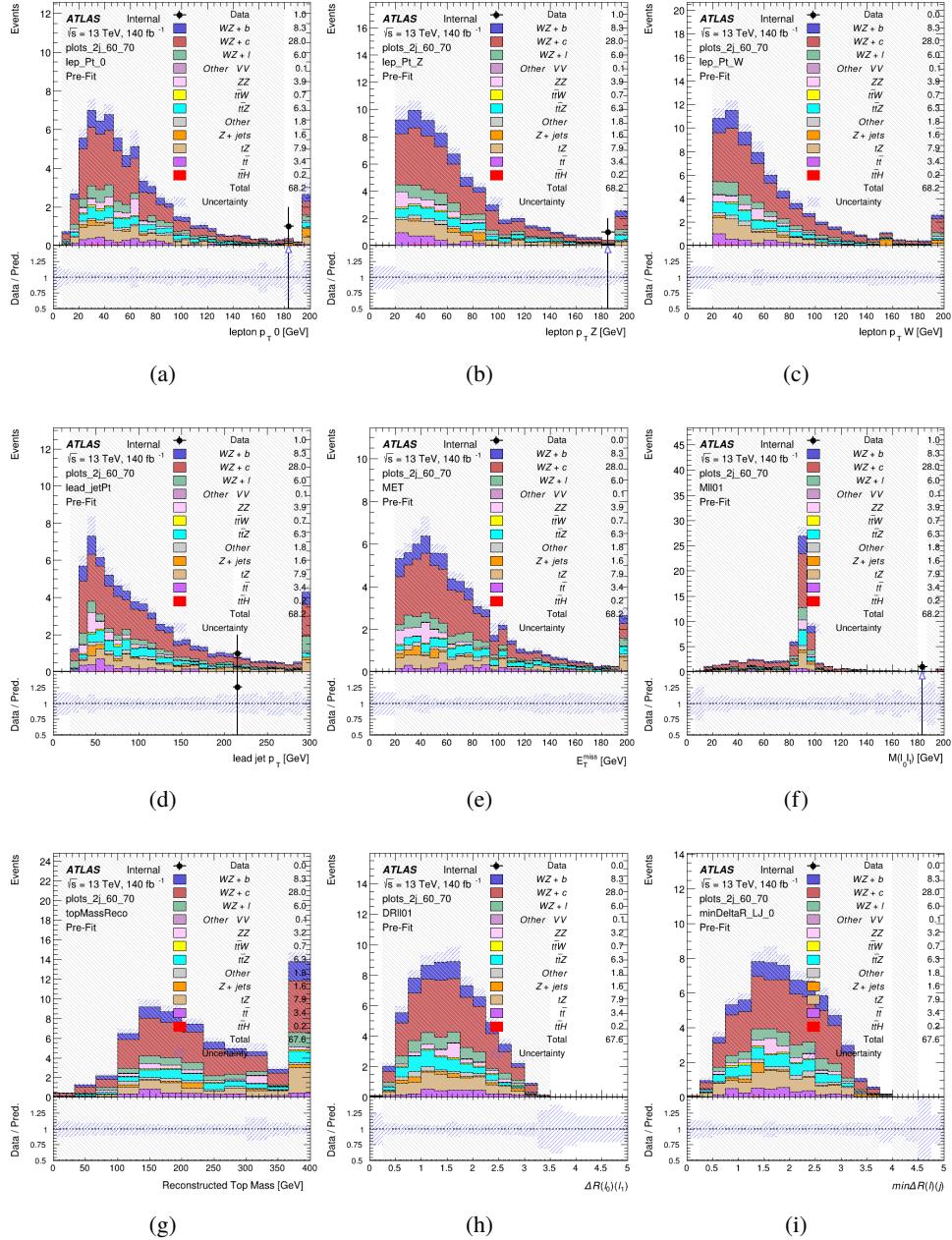


Figure 12.14: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

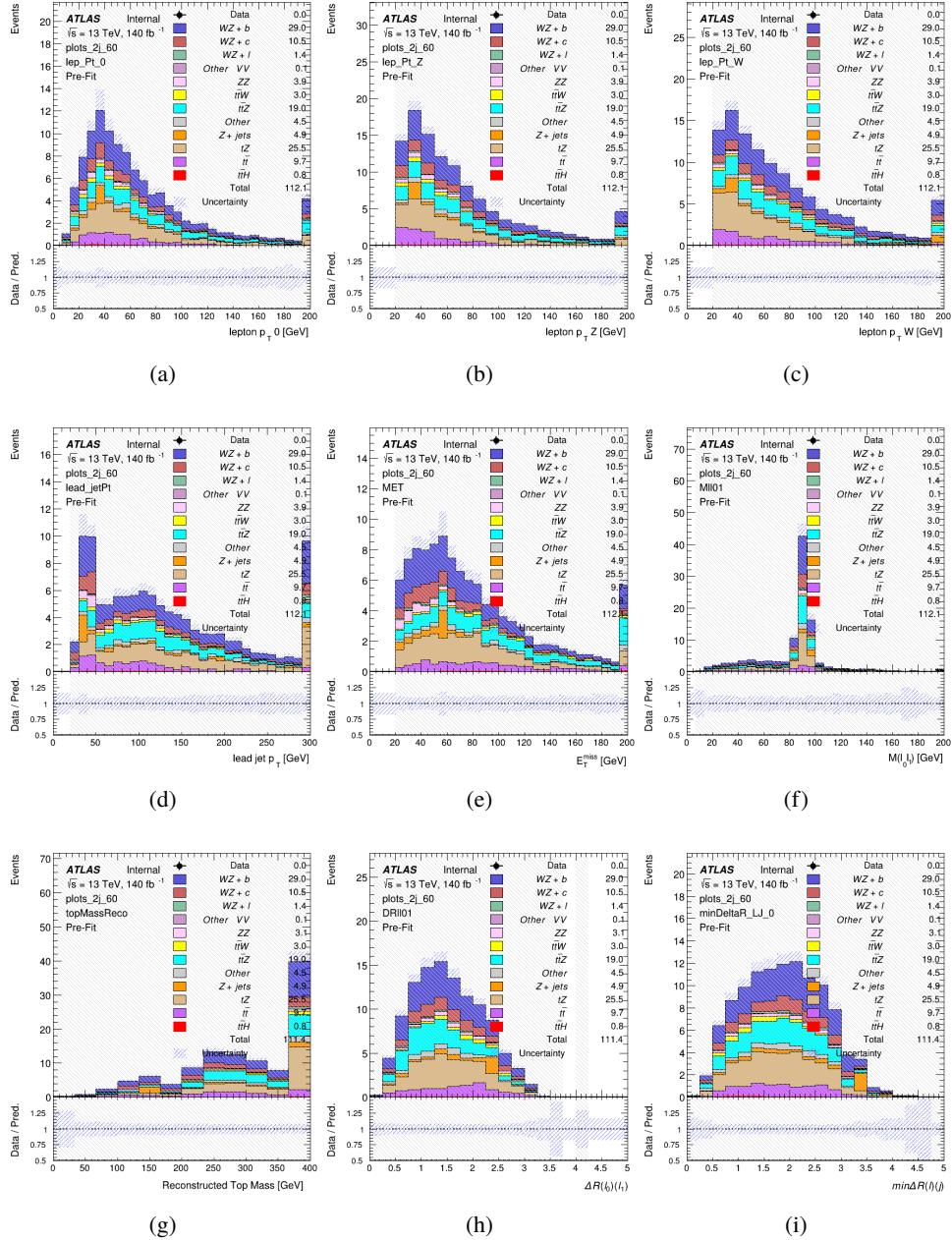


Figure 12.15: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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

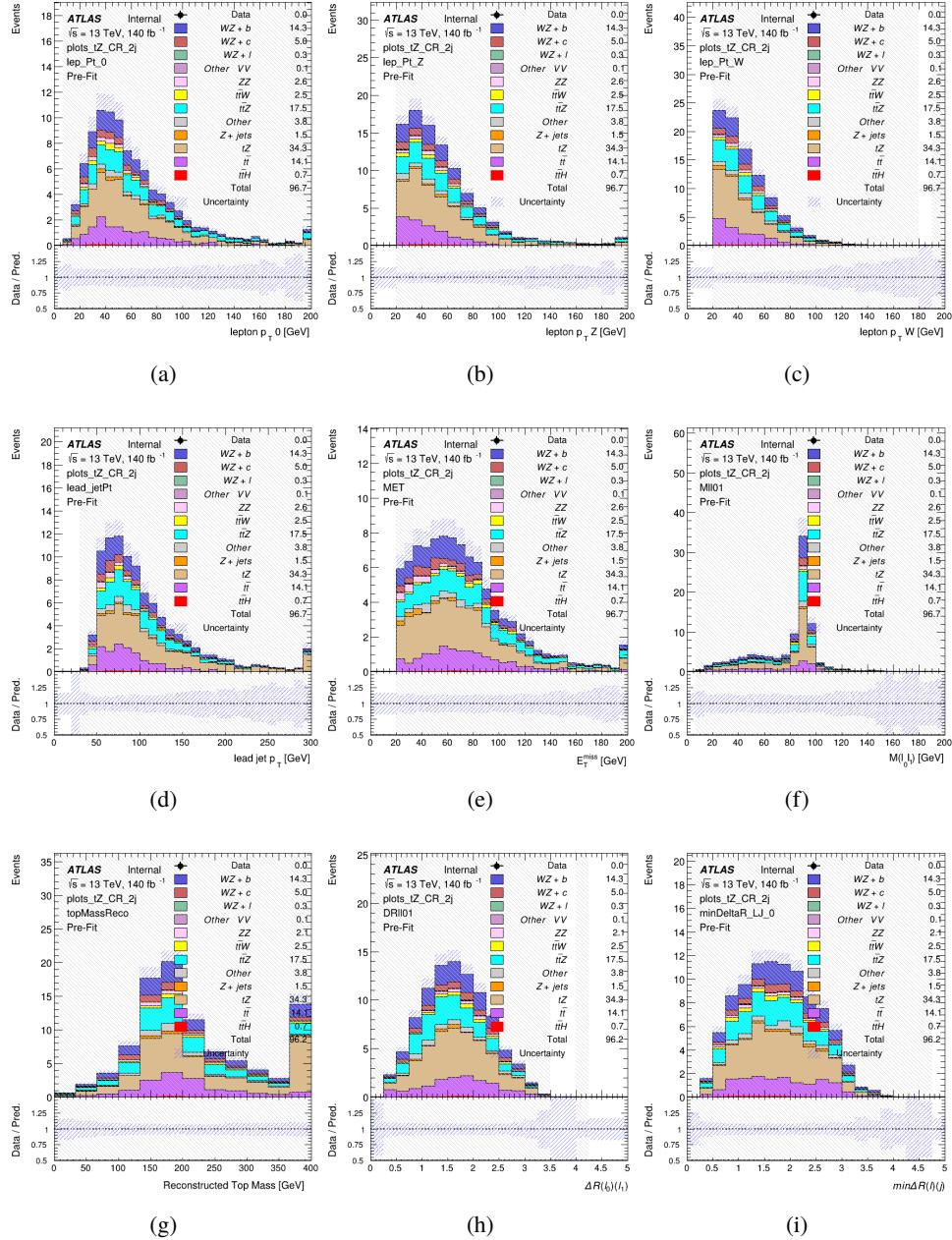


Figure 12.16: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant 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.

923 **12.3 Non-Prompt Lepton Estimation**

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

928 **12.3.1 $t\bar{t}$ Validation**

929 $t\bar{t}$ events can produce two prompt leptons from the decay of each of the tops. These top decays
 930 produce two b-quarks, the decay of which can produce additional non-prompt leptons, which
 931 occasionally pass the event preselection. In order to validate that the Monte Carlo accurately
 932 simulates this process accurately, the MC prediction in a non-prompt $t\bar{t}$ enriched control region
 933 is compared to data.

934 The $t\bar{t}$ control region is similar to the preselection region - three leptons meeting the
 935 criteria described in Section 12 are required, and the requirements on E_T^{miss} remain the same.
 936 However, the selection requiring a lepton pair form a Z-candidate are reversed. Events where the
 937 invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of 91.2 GeV are
 938 rejected. This ensures the $t\bar{t}$ control region is orthogonal to the preselection region.

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

⁹⁴¹ from top decays, at least one b-tagged jet passing the 70% DL1r WP in each event is required.

⁹⁴² Various kinematic plots of this region are shown in Figure 12.17.

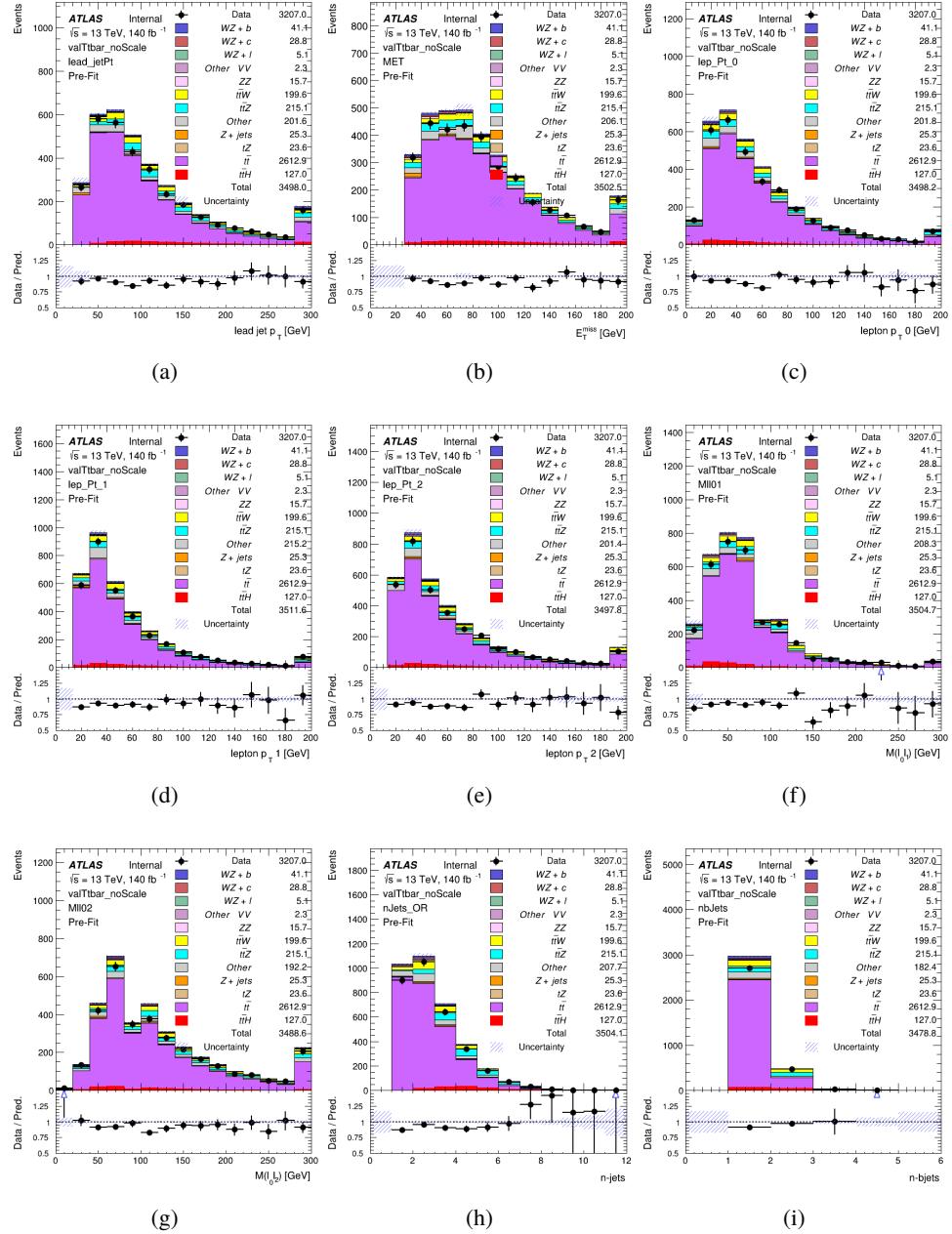


Figure 12.17: Comparisons between the data and MC distributions in the $t\bar{t}$ control region for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) the number of b-tagged jets.

943 The shape of each distribution agrees quite well between data and MC, with a constant
944 offset between the two. This is accounted for by applying a constant correction factor of 0.883
945 to the $t\bar{t}$ MC prediction. Plots showing the kinematics of the $t\bar{t}$ VR after this correction factor
946 has been applied are shown in Figure 12.18.

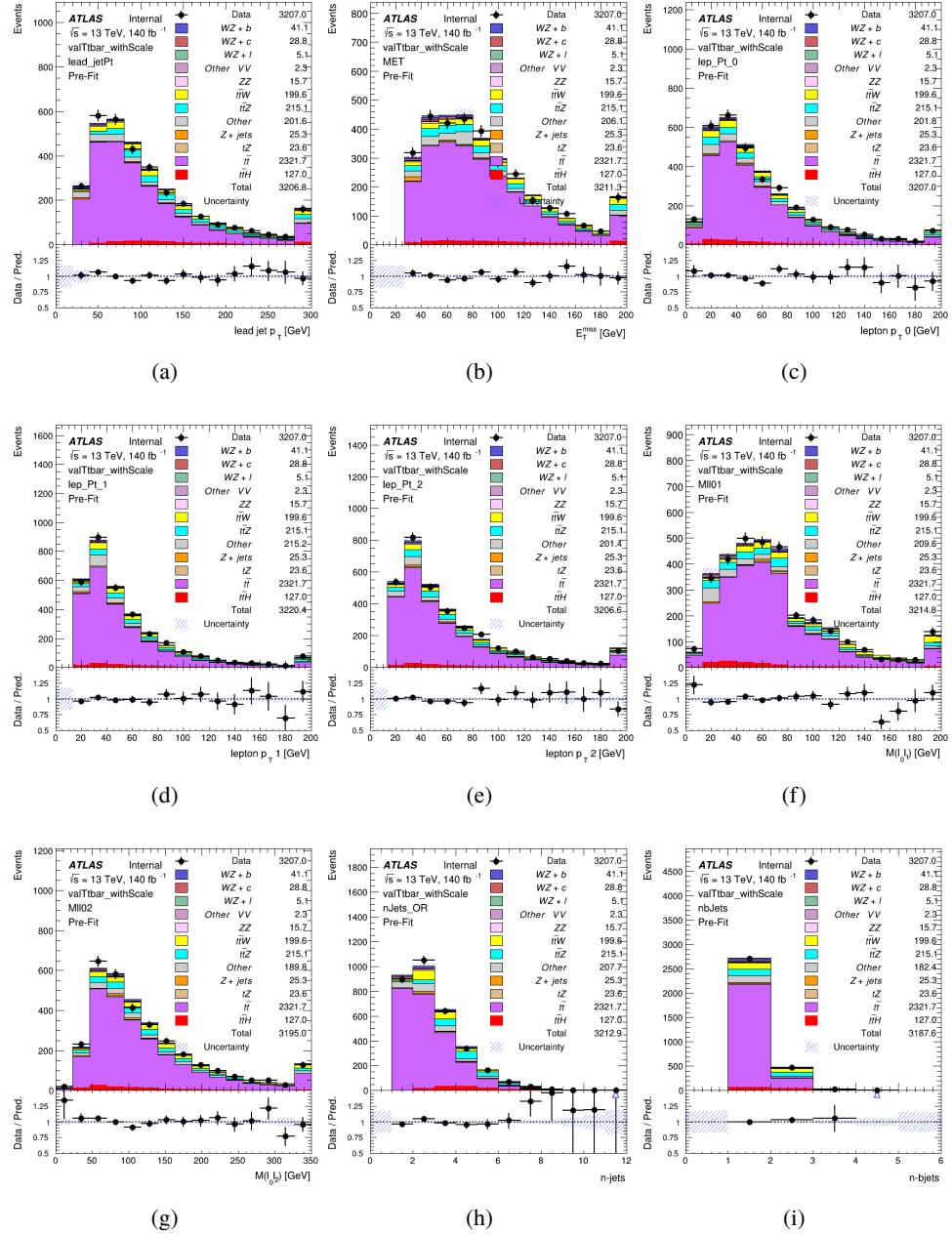


Figure 12.18: Comparisons between the data and MC distributions in the $t\bar{t}$ control region after the correction factor has been applied for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass 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.

947 The modeling is further validated by looking at the yield in the $t\bar{t}$ VR for each DL1r WP,
 948 giving a clearer correspondence to the signal regions used in the fit. Each region shown in Figure
 949 12.19 requires one or more jets pass the listed WP, with no jets passing the next highest WP.

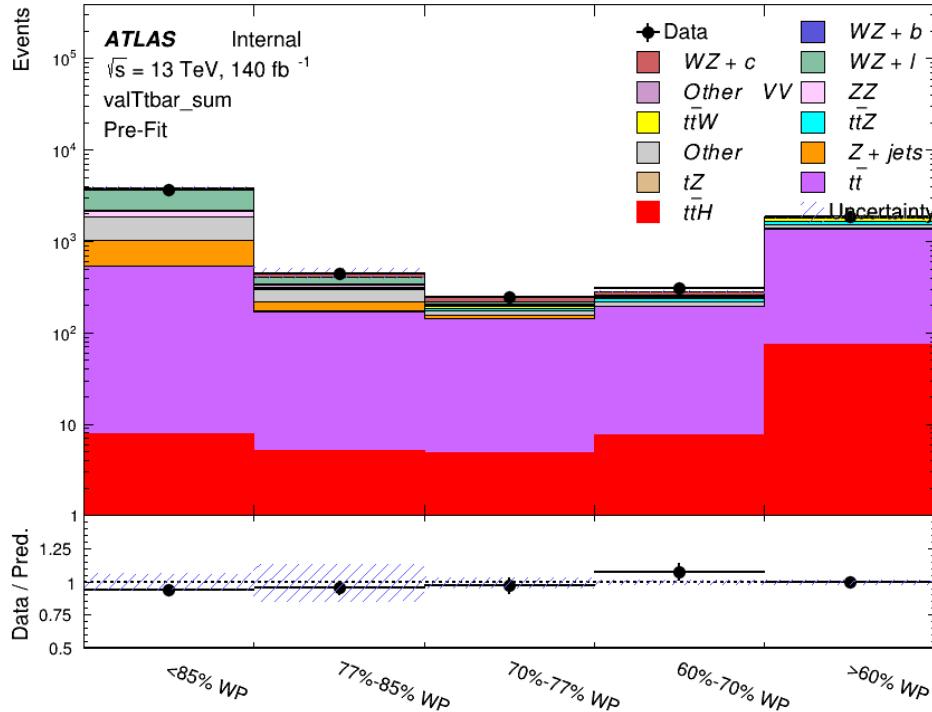


Figure 12.19: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the $t\bar{t}$ VR selection and correction factor have been applied

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

952 12.3.2 Z+jets Validation

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

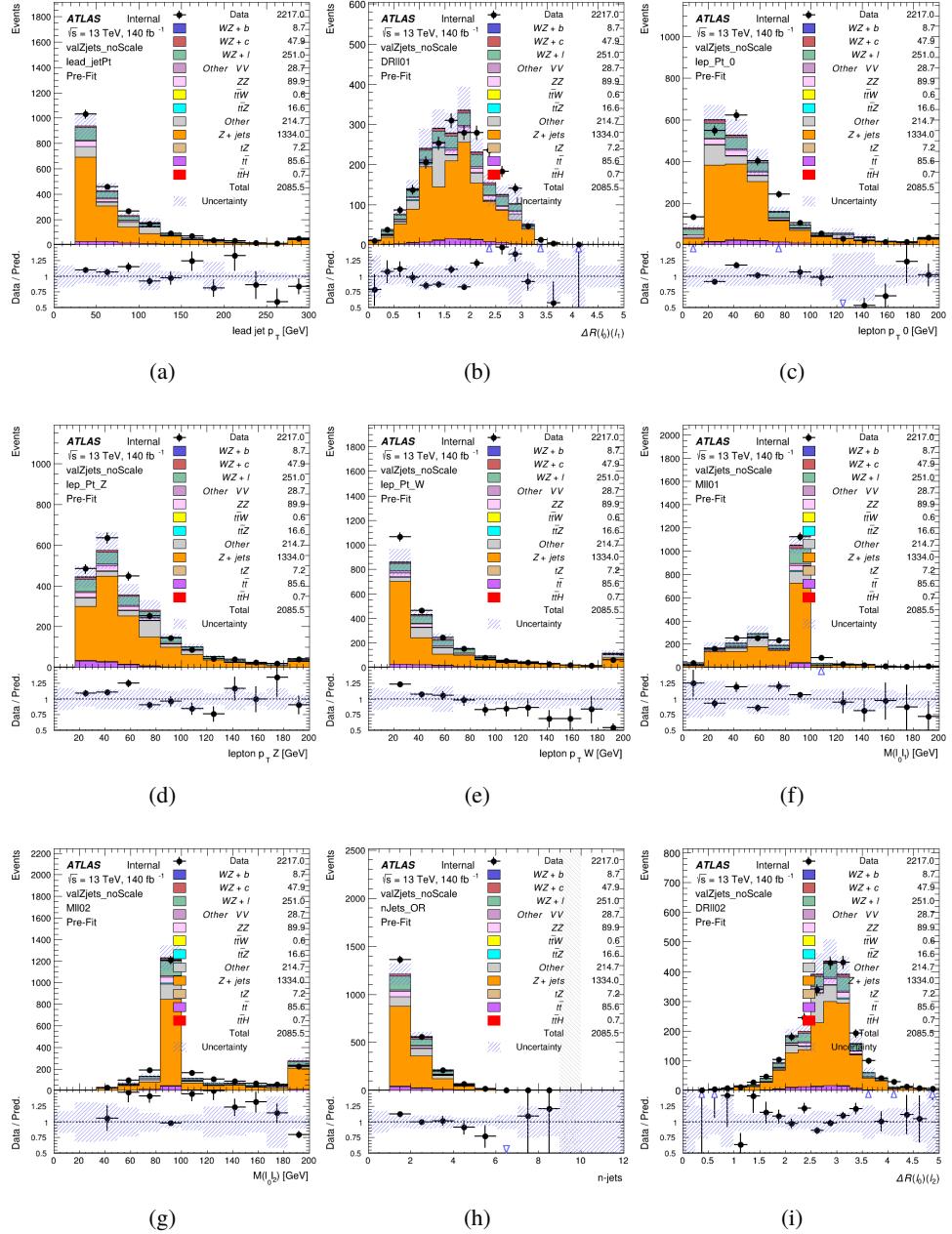


Figure 12.20: Comparisons between the data and MC distributions in the $Z + \text{jets}$ control region for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant 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

960 While there is general agreement between data and MC within statistical uncertainty, the
961 shape of the p_T spectrum of the lepton from the W is found to differ. To account for this
962 discrepancy, a variable correction factor is applied to Z+jets. χ^2 minimization of the lepton 2 p_T
963 spectrum is performed to derive a correction factor of $1.53 - 6.6 * 10^{-6}(\text{lep}_\text{Pt}_W)$. Kinematic
964 plots of the Z + jets control region after this correction factor has been aplied are shown in Figure
965 [12.21](#).

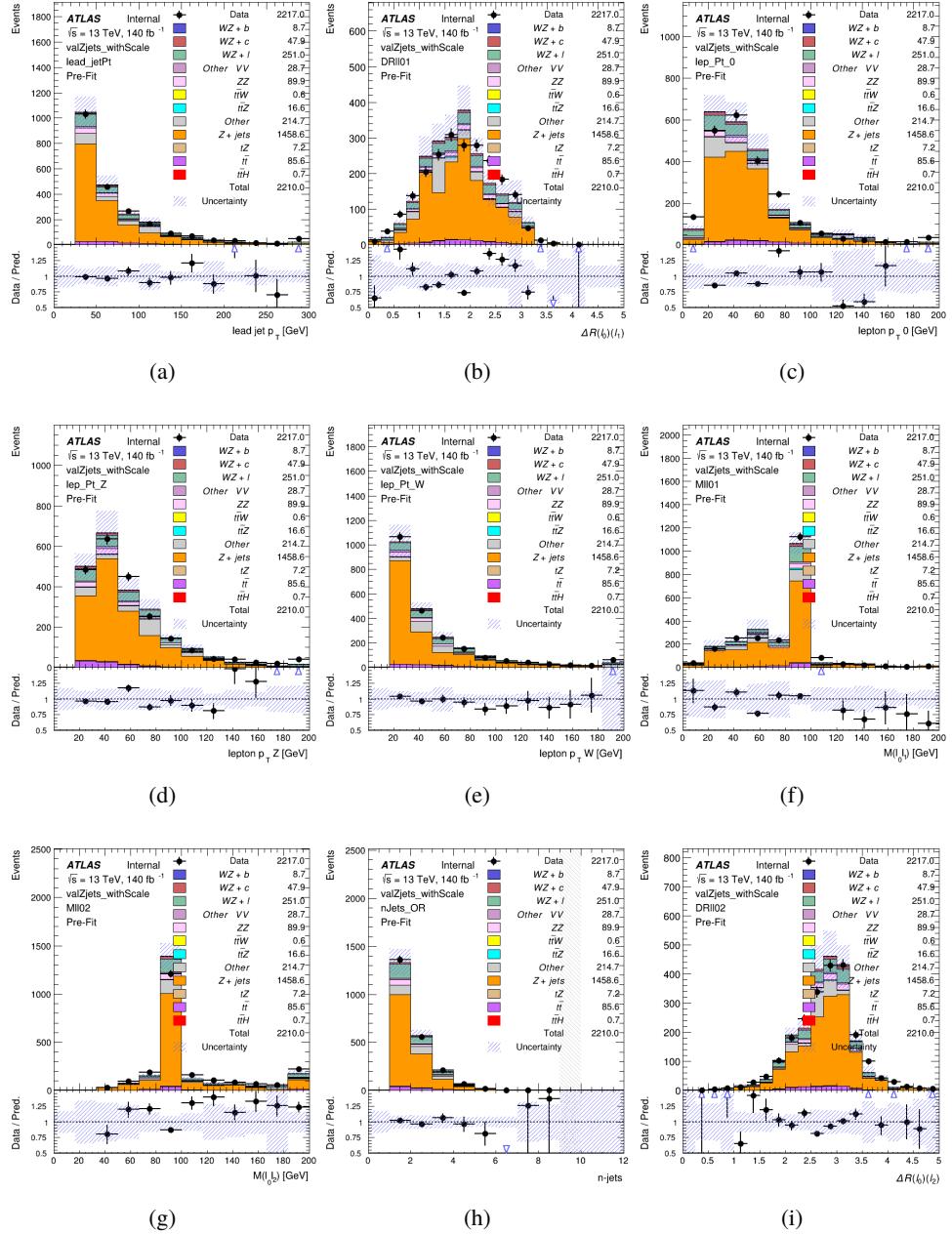


Figure 12.21: Comparisons between the data and MC distributions in the $Z + \text{jets}$ control region after the correction factor has been applied for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS 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

966 The modeling is further validated by looking at the yield in the Z+jets VR for each DL1r
 967 WP, giving a clearer correspondence to the signal regions used in the fit. Each region shown in
 968 Figure 12.22 requires one or more jets pass the listed WP, with no jets passing the next highest
 969 WP.

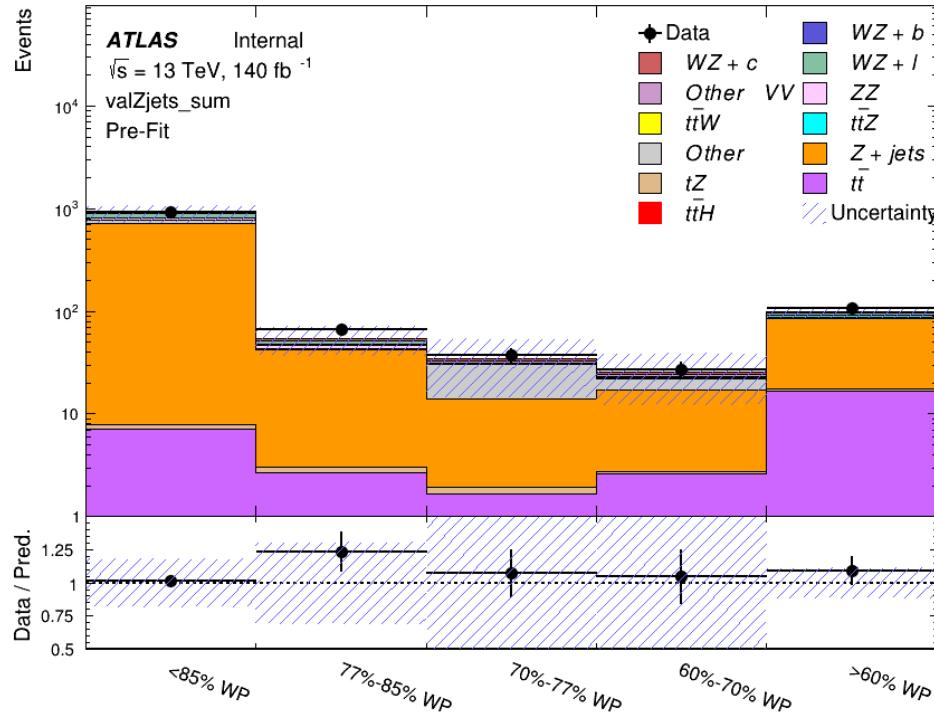


Figure 12.22: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the Z+jets VR selection and correction factor have been applied

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

973 13 tZ Interference Studies and Separation Multivariate Analysis

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

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

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

983 13.1 Top Mass Reconstruction

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

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

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

$$994 \quad (p_l + p_\nu)^2 = m_W^2$$

995 Expanding this out into components, this equation gives:

$$996 \quad \sqrt{p_{T\nu}^2 + p_{z\nu}^2} E_l = \frac{m_W^2 - m_l^2}{2} + p_{T\nu}(p_{lx}\cos\phi_\nu + p_{ly}\sin\phi_\nu) + p_{lz}p_{\nu z}$$

997 This equation gives two solutions for $p_{\nu z}$. For cases where only one of these solutions is
 998 real, that is taken as the value of $p_{\nu z}$. For instances with two real solutions, the one which is
 999 shown to be correct in the largest fraction of simulations is taken. For cases when no real solution
 1000 is found, often because of detector effects, the value of E_T^{miss} is varied in decreasing increments
 1001 of 100 MeV until a real solution is found.

1002 The reconstructed top mass distribution for tZ and WZ + b can be seen in figure 13.1.

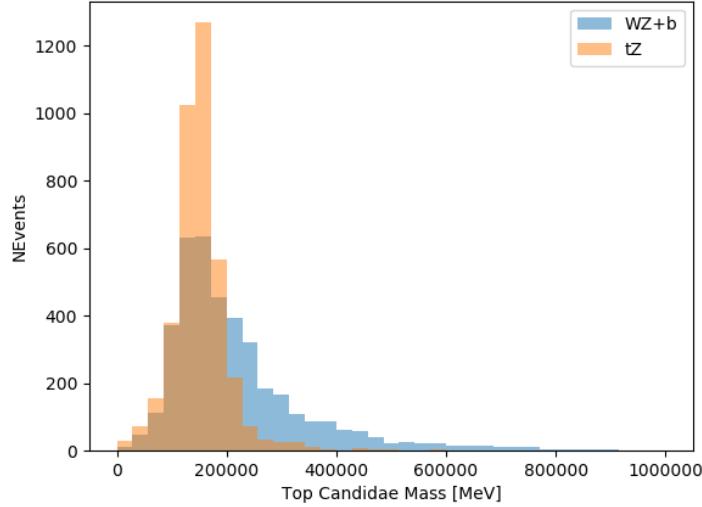


Figure 13.1: Reconstructed top mass distributions for tZ and WZ + b, measured in MeV.

13.2 tZ BDT

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

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

- 1011 • Distance between each lepton and the jet, $\Delta R(lj)$

1012 The training samples included only events meeting the requirements of the 1-jet, $>60\%$
1013 region, i.e. passing all the selection described in section 12 and having exactly one jet which
1014 passes the tightest (60%) DL1r working point.

1015 The distributions of a few of these features for both signal and background is shown in
1016 figure 13.2.

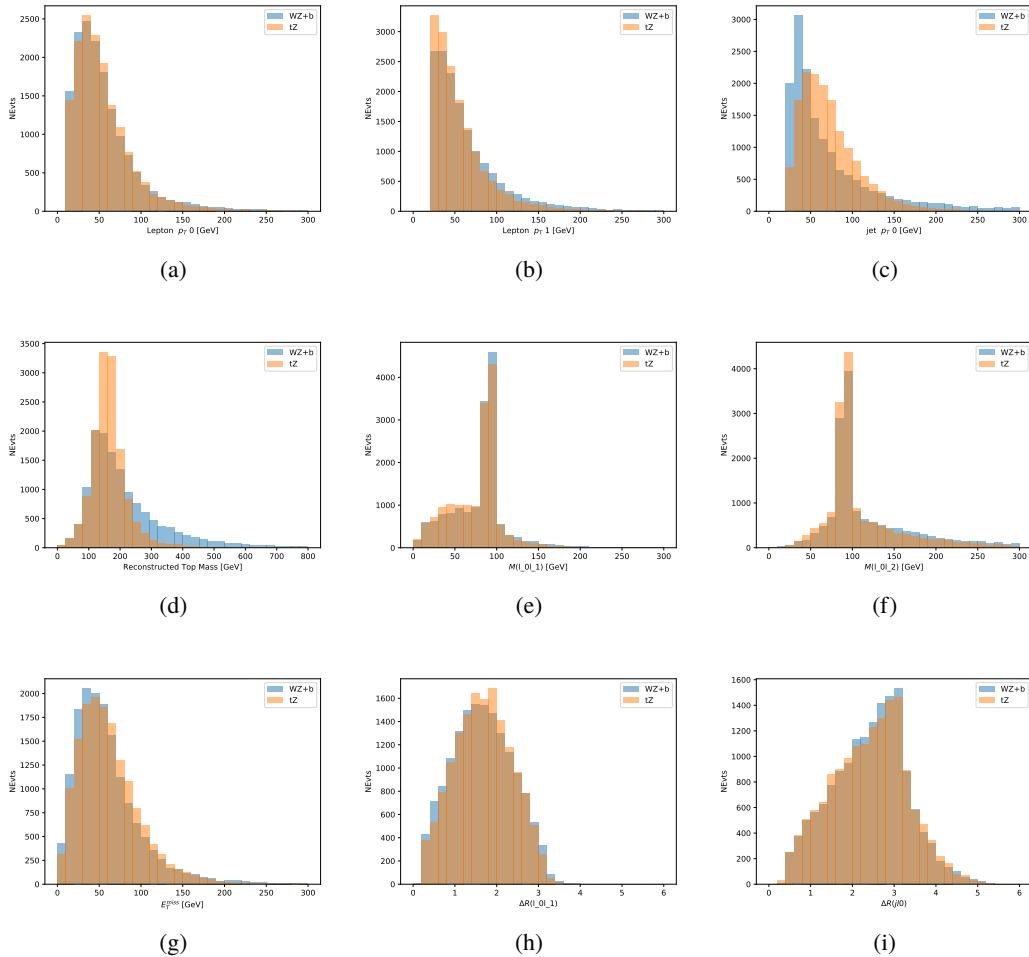


Figure 13.2: Distribution of input features of the BDT for signal (WZ) and background (tZ). Both are scaled to an equal number of events. (a), (b) and (c) show the p_T of lepton 0, lepton 1, and the jet, (d) show the reconstructed top mass, (e) and (f) show the invariant mass of leptons 0 and 1, leptons 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.

1017 A sample of 20,000 background (tZ) and signal (WZ+b) Monte Carlo events are used to
 1018 train the BDT. And additional 5,000 events are reserved for testing the model, in order to prevent
 1019 over-fitting. A total of 750 decision trees with a maximum depth of 6 branches are used to build
 1020 the model. These parameters are chosen empirically, by training several models with different

1021 parameters and selecting the one that gave the best separation for the test sample.

1022 The results of the BDT training are shown in figure 13.3. The output scores for both signal
 1023 and background events is shown on the left. The right shows the receiving operating characteristic
 1024 (ROC) curve that results from the MVA. The ROC curve represents the background rejection
 1025 as a function of signal efficiency, where each point on the curve represents a different response
 1026 score. The ROC curve of the BDT is compared to the performance of using an optimal set of flat
 1027 selections on the same set of input variables.

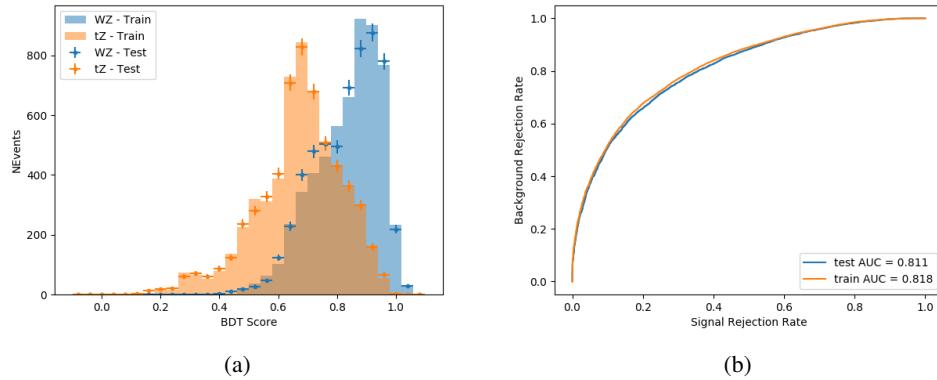


Figure 13.3: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

1028 The relative important of each input feature in the model, measured by how often they
 1029 appeared in the decision trees, is shown in figure 13.4.

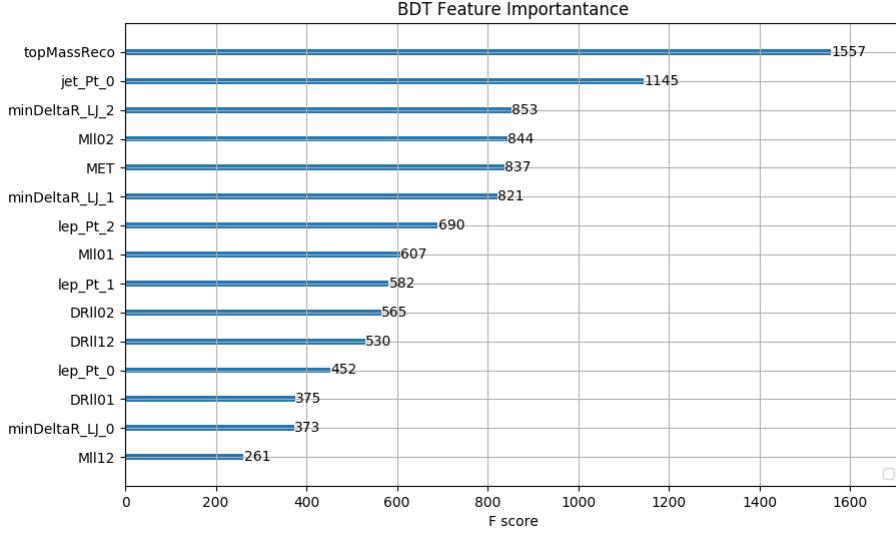


Figure 13.4: Relative importance of each input feature in the model.

1030 These results suggest that some amount of separation can be achieved between these two
 1031 processes, with a high BDT score selecting a set of events that is pure in $WZ + b$. A BDT score
 1032 of 0.725 is selected as a cutoff, where events with scores higher than this form a signal enriched
 1033 region, and events with scores lower than this form a tZ control region. This cutoff is selected by
 1034 varying the value of this cutoff in stat-only Asimov fits, and selecting the value that minimizes
 1035 the statistical uncertainty on $WZ + b$.

1036 14 Systematic Uncertainties

1037 The systematic uncertainties that are considered are summarized in Table 49. These are imple-
 1038 mented in the fit either as a normalization factors or as a shape variation or both in the signal

¹⁰³⁹ and background estimations. The numerical impact of each of these uncertainties is outlined in

¹⁰⁴⁰ Section 22.

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

¹⁰⁴¹ The uncertainty in the combined integrated luminosity is derived from a calibration of the

¹⁰⁴² luminosity scale performed for 13 TeV proton-proton collisions [**lumi**], [**LUCID2**].

¹⁰⁴³ The experimental uncertainties are related to the reconstruction and identification of light

¹⁰⁴⁴ leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The TOTAL electron ID

¹⁰⁴⁵ correlation model is used, corresponding to 1 electron ID systematic. Electron ID is found to be

1046 a subleading systematic that is unconstrained by the fit, making it an appropriate choice for this
1047 analysis.

1048 The sources which contribute to the uncertainty in the jet energy scale (JES) [**jes**] are decom-
1049 posed into uncorrelated components and treated as independent sources in the analysis. The
1050 CategoryReduction model is used to account for JES uncertainties, which decomposes the uncer-
1051 tainties into 30 nuisance parameters included in the fit. The SimpleJER model is used to account
1052 for jet energy resolution (JER) uncertainties, and 8 JER uncertainty components included as
1053 NPs in the fit.

1054 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
1055 [**btag_cal**] are also decomposed into uncorrelated components. The large number of components
1056 for b-tagging is due to the calibration of the distribution of the MVA discriminant.

1057 The full list of systematic uncertainties considered in the analysis is summarized in Tables
1058 [18](#), [19](#) and [20](#).

1059

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.

1060 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and scale
 1061 uncertainties are taken from theory calculations, with the exception of non-prompt and diboson
 1062 backgrounds. The cross-section uncertainty on tZ is taken from [tZ_paper]. Derivation of the
 1063 non-prompt background uncertainties, Z+jets and t \bar{t} , are explained in detail in Section 12.3.
 1064 These normalization uncertainties are chosen so as to account for the complete uncertainty in the
 1065 non-prompt contribution, and therefore no additional modelling uncertainties are considered for
 1066 Z+jets and t \bar{t} .

1067 The other VV + heavy flavor processes (namely VV+b and VV+charm, which primarily
 1068 consist of ZZ events) are also poorly understood, because these processes involve the same
 1069 physics as WZ + heavy flavor, and have also not been measured. Therefore, a conservative 50%

¹⁰⁷⁰ uncertainty is applied to those samples. While this uncertainty is large, it is found to have little
¹⁰⁷¹ impact on the significance of the final result.

¹⁰⁷² The theory uncertainties applied to the predominate background estimates are summarized
¹⁰⁷³ 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
t̄t Z (aMC@NLO+Pythia8)	QCD Scale: $^{+9.6}_{-11.3}$ PDF($+\alpha_S$): ± 4
t̄t W (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.

¹⁰⁷⁴ Additional signal uncertainties are estimated by comparing estimates from the nominal Sherpa
¹⁰⁷⁵ WZ samples with alternate WZ samples generated with Powheg+PYTHIA8 (DSID 361601).
¹⁰⁷⁶ MC/MC scale factors are applied to make these comparisons. The shape of the templates used
¹⁰⁷⁷ in the fit are compared between these two samples for WZ + b, WZ + charm and WZ + light, as
¹⁰⁷⁸ shown in Figures 14.1 and 14.2. Each of these plots are normalized to unity in order to capture
¹⁰⁷⁹ differences in shape.

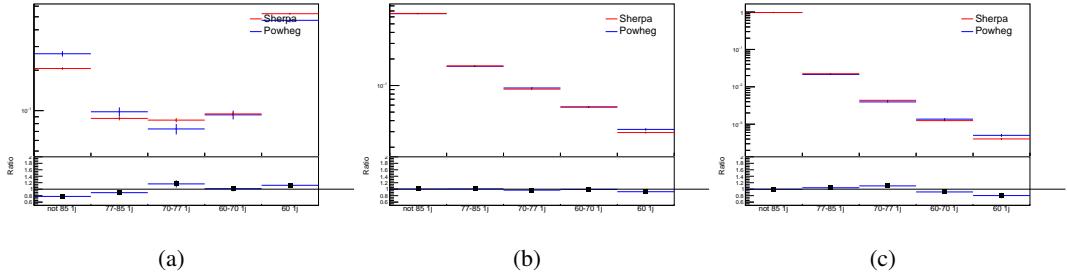


Figure 14.1: Comparison between Sherpa and Powheg predictions of the distribution of (a) WZ + b, (b) WZ + charm, and (c) WZ + light among the various b-tag WPs used in the 1-jet fit.

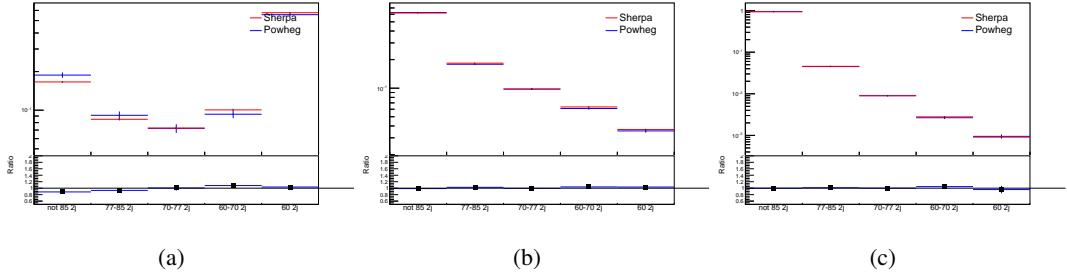


Figure 14.2: Comparison between Sherpa and Powheg predictions of the distribution of (a) WZ + b, (b) WZ + charm, and (c) WZ + light among the various b-tag WPs used in the 2-jet fit.

1080 Separate systematics are included in the fit for WZ + b, WZ + charm and WZ + light,
 1081 where the distribution among each of the fit regions is varied based on the prediction of the
 1082 Powheg sample.

1083 A similar approach is taken to account for uncertainties in migrations between the number of
 1084 reco and truth jets. The fraction of events with 1 truth jet which fall into the 1 jet bin versus the
 1085 2 jet bin at reco level is compared for Sherpa and Powheg. The same is done for events with 2
 1086 truth jets. This comparison is shown in figure 14.3.

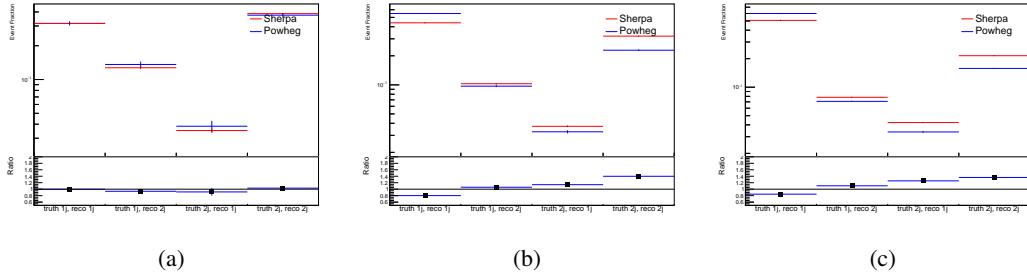


Figure 14.3: Comparison between Sherpa and Powheg predictions for truth jet migrations between the 1 and 2 jet reco bins for (a) WZ + b, (b) WZ + charm, and (c) WZ + light

1087 A systematic is included where events are shifted between the 1-jet and 2-jet regions based
 1088 on the differences between these two shapes. This is done independently for each of the WZ +
 1089 b, WZ + charm, and WZ + light templates.

1090 Additional systematics are included to account for the uncertainty in the contamination of
 1091 0 jet and 3 or more jet events (at truth level) in the 1 and 2 reco jet bins. Because these events
 1092 fall outside the scope of this measurement, these events are included as a background. As such,
 1093 a normalization, rather than a shape, uncertainty is applied for this background.

1094 The number of WZ events with 0-jets and ≥ 3 -jets in the reconstructed 1-jet and 2-jet
 1095 regions are compared for Sherpa and Powheg, as seen in figure 14.4. These differences are taken
 1096 as separate normalization systematics on the yield of WZ+0-jet and WZ+ ≥ 3 -jet events.

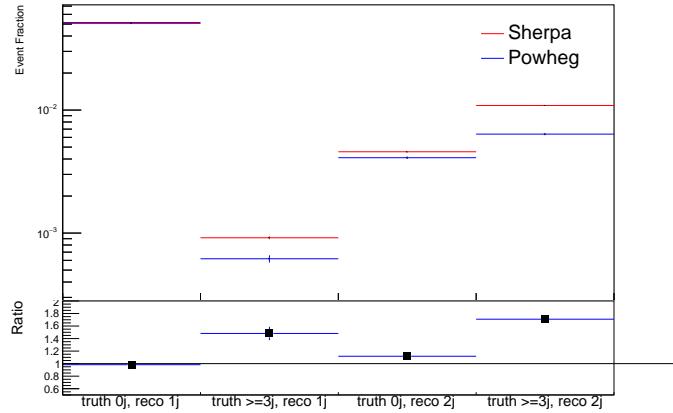


Figure 14.4: Comparison between Sherpa and Powheg predictions for 0 and ≥ 3 truth jet contributions in the 1 and 2 jet reco bins

15 Results

1097 A separate maximum-likelihood fit is performed over the 1-jet and 2-jet fit regions in order to
 1099 extract the best-fit value of the $WZ + b$ -jet and $WZ + c$ charm jet contributions. The $WZ + b$,
 1100 $WZ + c$ and $WZ + \text{light}$ contributions are allowed to float, with the remaining background
 1101 contributions are held fixed. **The current fit strategy treats the $WZ + b$ -jet contribution as**
 1102 **the parameter of interest, with the normalization of the $WZ + c$ charm and the $WZ + \text{light}$**
 1103 **contributions taken as systematic uncertainties. This could however be adjusted, depending**
 1104 **on whether it is decided the goal of the analysis should be to measure $WZ+b$ specifically or**
 1105 **$WZ + \text{heavy flavor overall}$.** The result of the fit is used to extract the cross-section of $WZ +$
 1106 heavy-flavor production.

1107 A maximum likelihood fit to data is performed simultaneously in the regions described

1108 in Section 12. The parameters μ_{WZ+b} , $\mu_{WZ+charm}$, $\mu_{WZ+light}$, where $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$, are
1109 extracted from the fit.

1110 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
1111 $WZ + b$. The fitted cross-section modifiers for $WZ + \text{charm}$ and $WZ + \text{light}$ are $1.00 \pm 0.17 \pm 0.15$
1112 and $1.00 \pm 0.04 \pm 0.07$, respectively.

1113 The expected cross-section of $WZ+b$ with 1 associated jet obtained from the fit is
1114 $1.74^{+0.65}_{-0.61}(\text{stat})^{+0.42}_{-0.37}(\text{sys}) \text{ fb}$ with an expected significance of 2.2σ . The expected cross-section
1115 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.

1116 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 $WZ +$
1117 b . The fitted cross-section modifiers for $WZ + \text{charm}$ and $WZ + \text{light}$ are $1.00 \pm 0.17 \pm 0.15$ and
1118 $1.00 \pm 0.04 \pm 0.08$, respectively.

1119 The expected $WZ + b$ cross-section in the 2-jet region is $2.46^{+0.83}_{-0.81}(\text{stat})^{+0.073}_{-0.68}(\text{sys}) \text{ fb}$
1120 with an expected significance of 2.6σ . The 2-jet expected cross-section of $WZ + \text{charm}$ is
1121 $12.7 \pm 2.2(\text{stat}) \pm 2.0(\text{sys}) \text{ fb}$, and the correlation between $WZ + \text{charm}$ and $WZ + b$ is -0.26.

1122 15.1 1-jet Fit Results

1123 The results of the fit are currently blinded.

¹¹²⁴ The pre-fit yields in each of the regions used in the fit are shown in Table 15.1, and

¹¹²⁵ summarized in Figure 15.1.

¹¹²⁶

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.

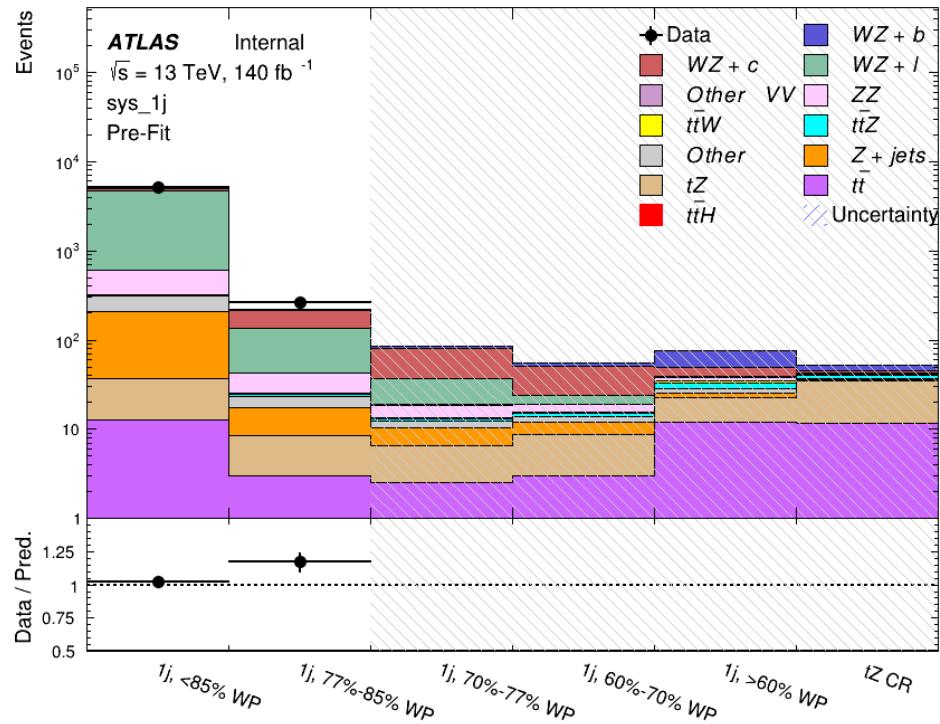


Figure 15.1: Pre-fit summary of the 1-jet fit regions.

1127

The post-fit yields in each region are summarized in Figure 15.1.

1128

	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.

¹¹²⁹

A post-fit summary plot of the 1-jet fitted regions is shown in Figure 15.2:

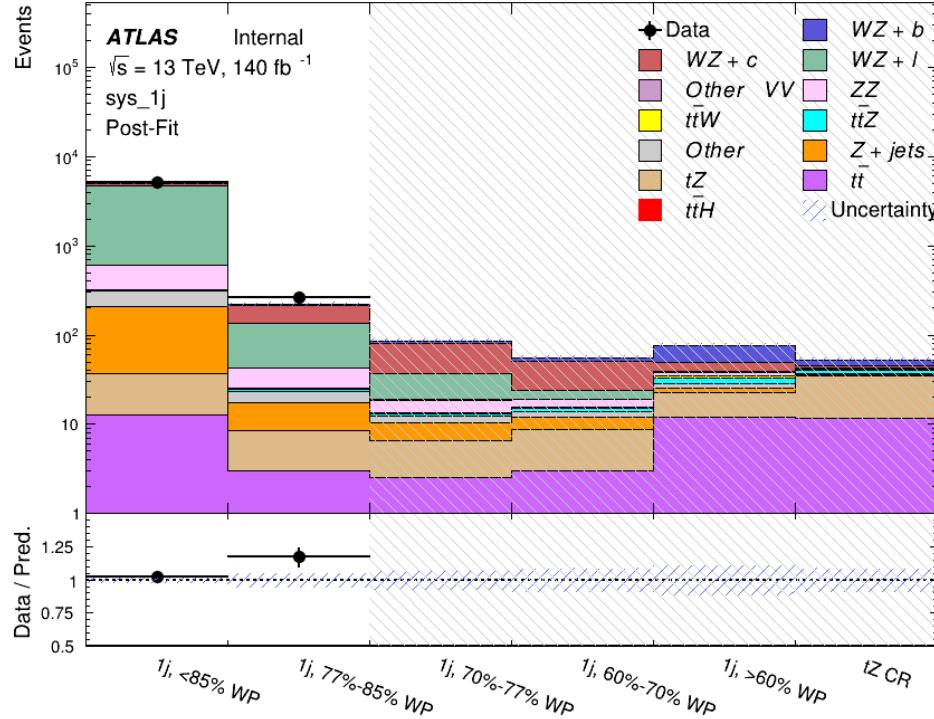


Figure 15.2: Post-fit summary of the 1-jet fit regions.

1130 As described in Section 21, there are 226 systematic uncertainties that are considered
 1131 as NPs in the fit. These NPs are constrained by Gaussian or log-normal probability density
 1132 functions. The latter are used for normalisation factors to ensure that they are always positive.
 1133 The expected number of signal and background events are functions of the likelihood. The prior
 1134 for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its
 1135 nominal value.

1136 The impact of each NP is calculated by performing the fit with the parameter of interest held
 1137 fixed, varied from its fitted value by its uncertainty, and calculating $\Delta\mu$ relative to the baseline

¹¹³⁸ fit. The impact of the most significant sources of systematic uncertainties is summarized in Table

¹¹³⁹ [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.

¹¹⁴⁰ The ranking and impact of those nuisance parameters with the largest contribution to the

¹¹⁴¹ overall uncertainty is shown in Figure [15.3](#).

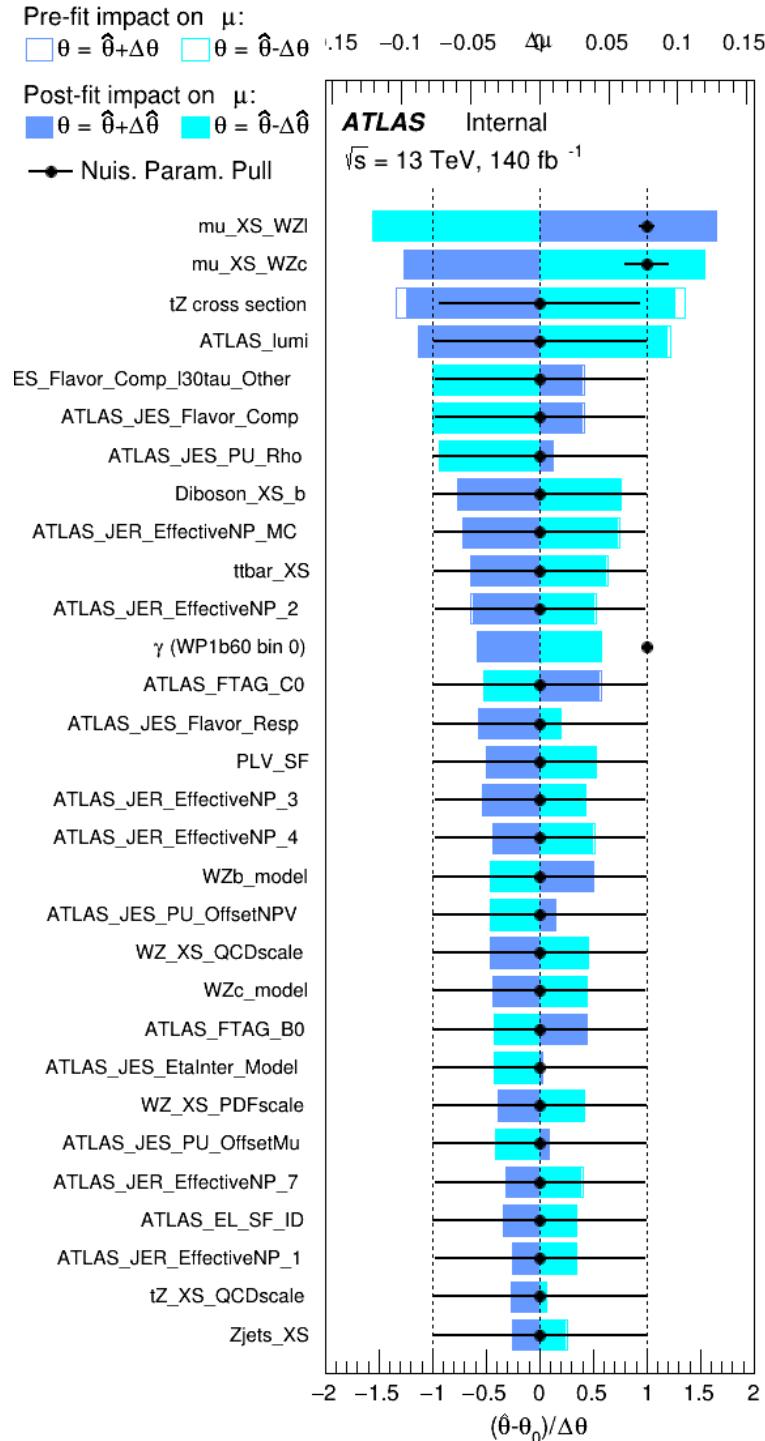


Figure 15.3: Impact of systematic uncertainties on the signal-strength of $WZ + b$ for events with exactly one jet

1142 The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the
 1143 shape of the fit regions depends heavily on the modeling of the jets. The other major sources
 1144 of uncertainty come from background modelling and cross-section uncertainty. The pie charts
 1145 in Figure 15.4 show that for the modelling uncertainties that contribute most correspond to the
 1146 most significant backgrounds.

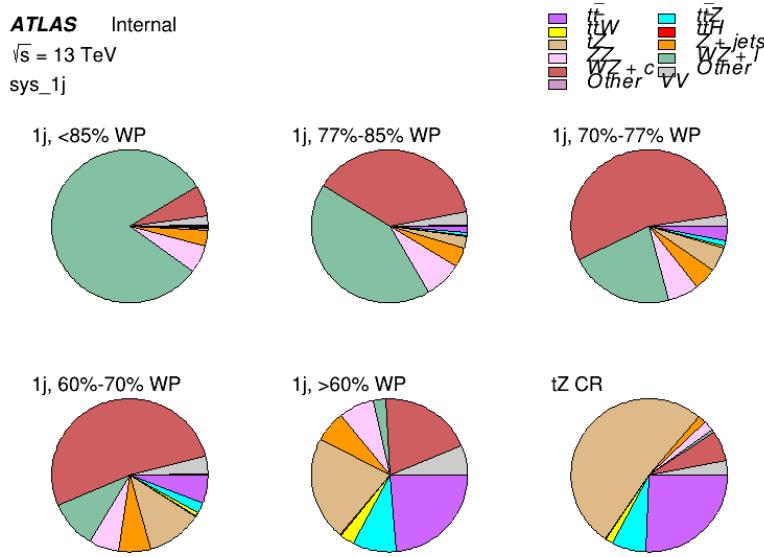


Figure 15.4: Post-fit background composition of the fit regions.

1147 The correlations between these nuisance parameters are summarized in Figure 15.5.

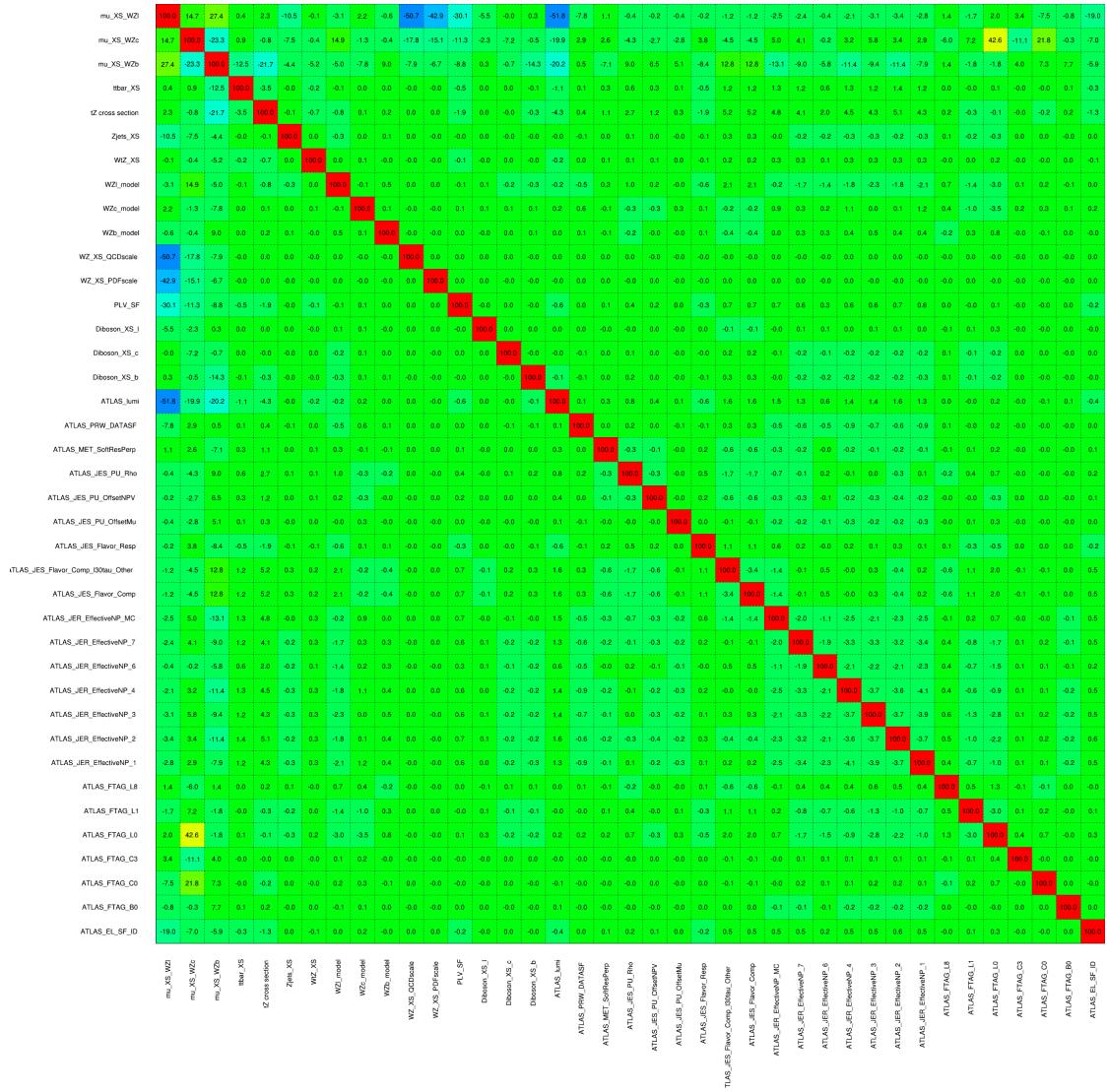


Figure 15.5: Correlations between nuisance parameters

1148

The negative correlations between $\mu_{WZ+charm}$ and μ_{WZ+b} and $\mu_{WZ+light}$ are expected:

1149

WZ + charm is present in both the WZ + b and WZ + light enriched regions, therefore increasing

1150

the fraction of charm requires increasing the fraction of WZ + b and WZ + light. This reasoning

1151

also explains the positive correlation between μ_{WZ+b} and $\mu_{WZ+light}$.

1152 Two of the major backgrounds in the region with the highest purity of $WZ + b$ are tZ and
1153 Other $VV + b$, explaining the negative correlations between μ_{WZ+b} and the tZ cross section, and
1154 the $VV + b$ cross section.

1155 The high correlation between the luminosity and $\mu_{WZ+light}$ arises from the fact that the
1156 uncertainty on $\mu_{WZ+light}$ is very low (around 4%). Small changes in luminosity cause a change
1157 in the yield of $WZ + light$ that is large compared to its uncertainty, producing a large correlation
1158 between these two parameters.

1159 **15.2 2-jet Fit Results**

1160 **The results of the fit are currently blinded.**

1161 Pre-fit yields in each of the 2-jet fit regions are shown in Figure 15.2.

	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.

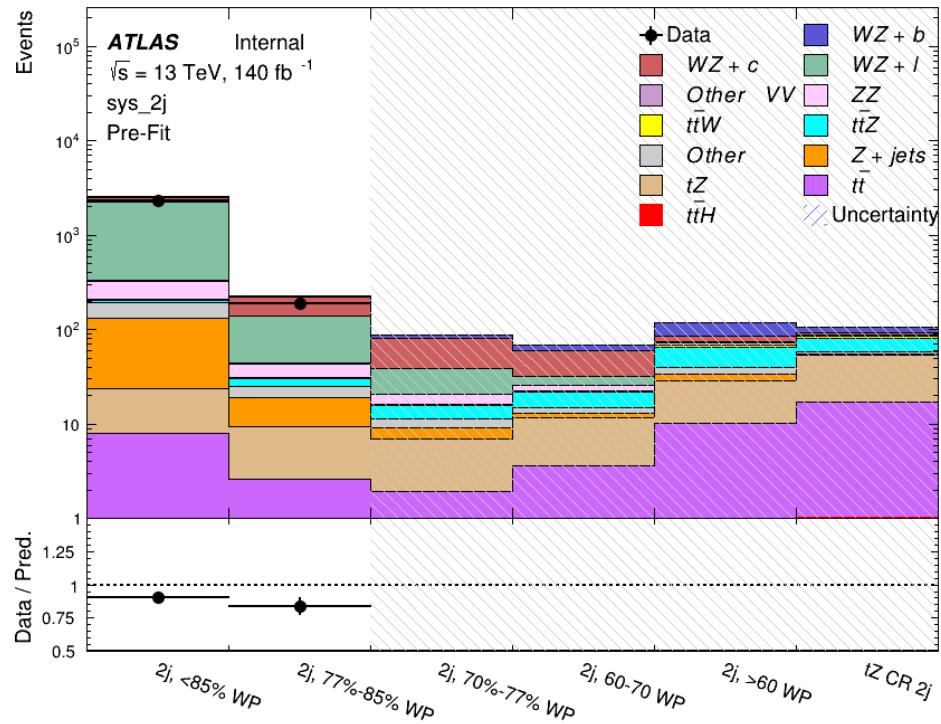


Figure 15.6: Pre-fit summary of the 2-jet fit regions.

1162

The post-fit yields in each region are summarized in Figure 15.2.

	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.

1163

A post-fit summary of the fitted regions is shown in Figure 15.7:

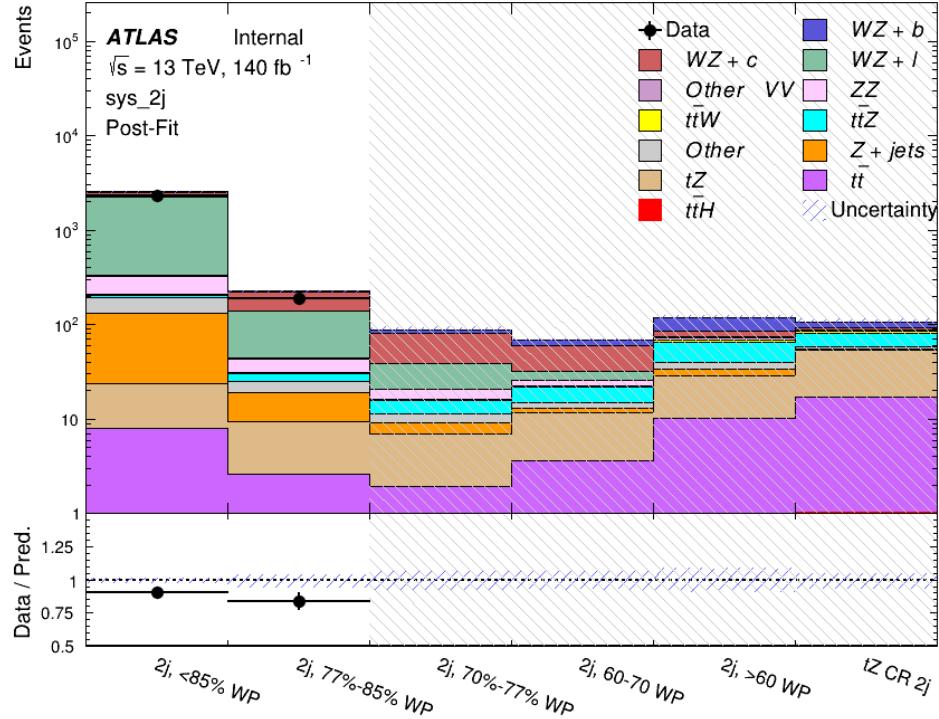


Figure 15.7: Post-fit summary of the fit over 2-jet regions.

1164 The same set of systematic uncertainties consider for the 1-jet fit are included in the 2-jet
 1165 fit as well. The impact of the most significant systematic uncertainties is summarized in Table
 1166 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.

¹¹⁶⁷ The ranking and impact of those nuisance parameters with the largest contribution to the
¹¹⁶⁸ overall uncertainty is shown in Figure 15.8.

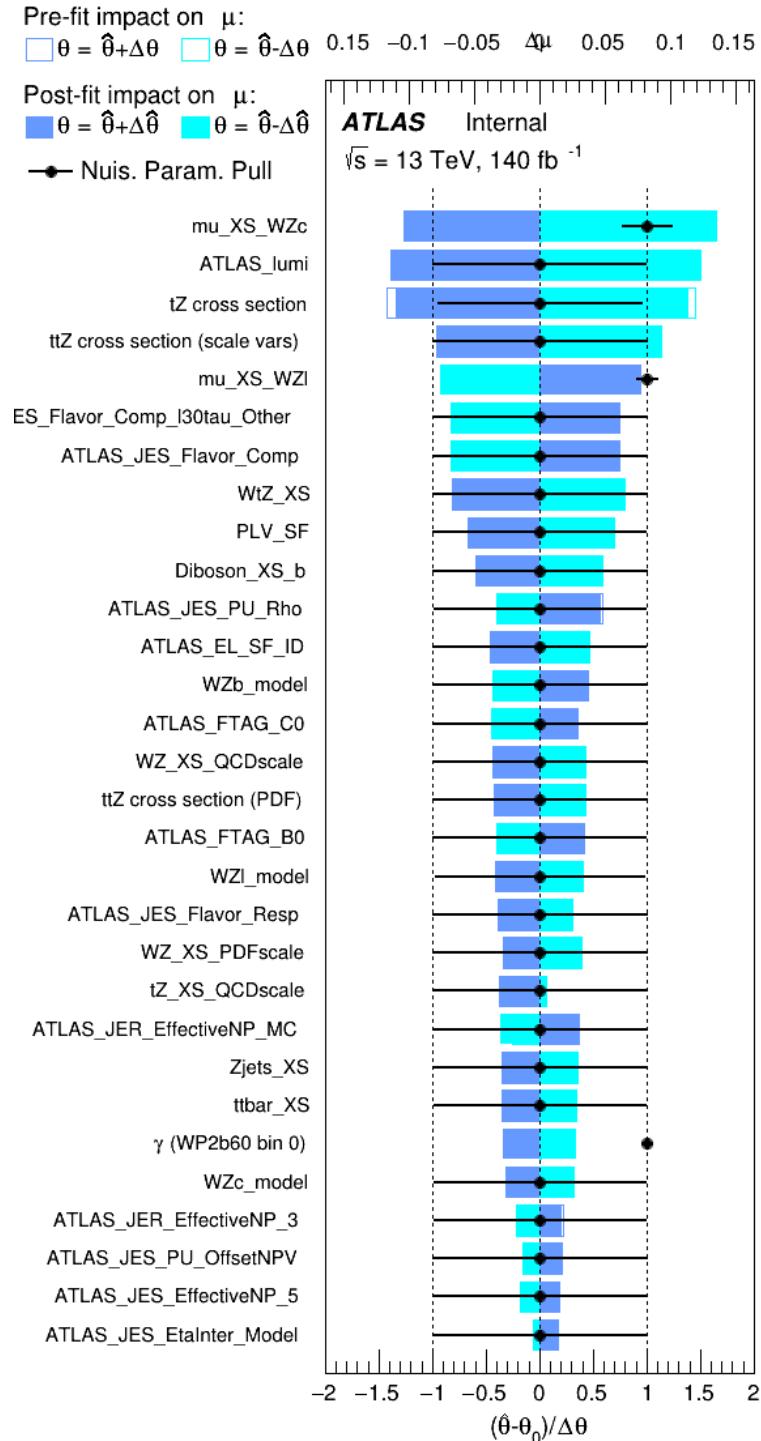


Figure 15.8: Impact of systematic uncertainties on the signal-strength of $WZ + b$ in 2-jet events.

1169 The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the
 1170 shape of the fit regions depends heavily on the modeling of the jets. The other major sources
 1171 of uncertainty come from background modelling and cross-section uncertainty. The pie charts
 1172 in Figure 15.9 show that for the modelling uncertainties that contribute most correspond to the
 1173 most significant backgrounds.

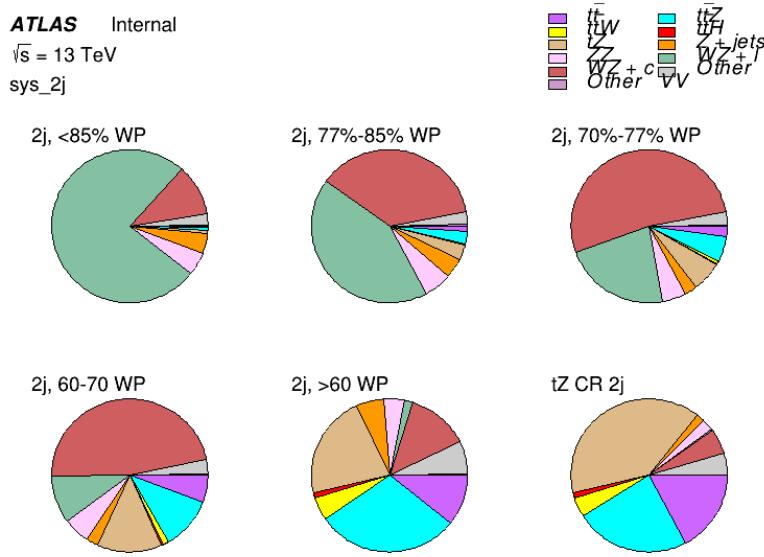


Figure 15.9: Post-fit background composition of the 2-jet fit regions.

1174 The correlations between these nuisance parameters are summarized in Figure 15.10.

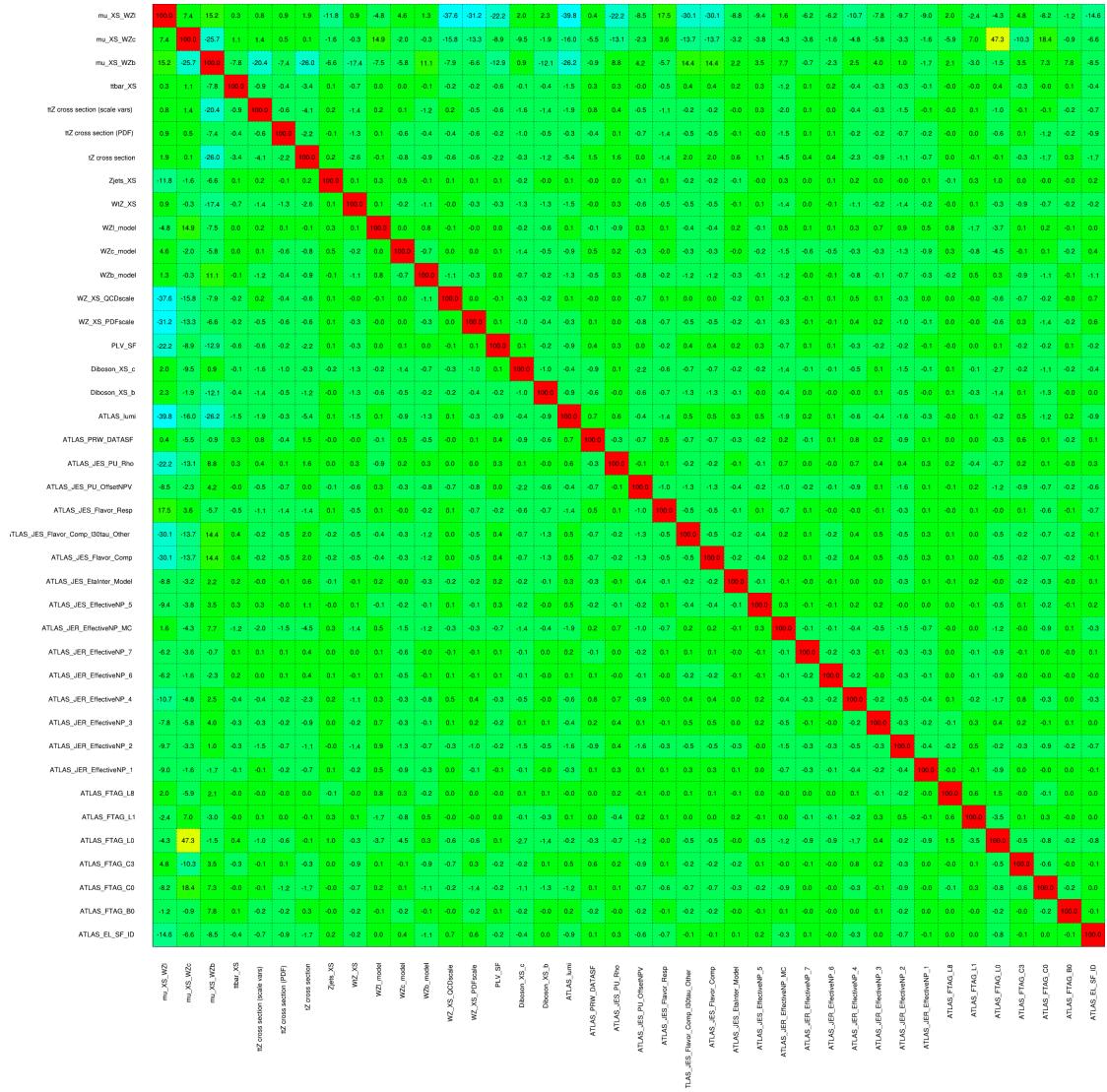


Figure 15.10: Correlations between nuisance parameters in the 2-jet fit

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

Part V**1178 Differential Studies of $t\bar{t}H$ Multilepton****1179 16 Data and Monte Carlo Samples**

1180 For both data and Monte Carlo (MC) simulations, samples were prepared in the `xAOD` format,
1181 which was used to produce a `xAOD` based on the `HIGG8D1` derivation framework. This framework
1182 was designed for the main $t\bar{t}H$ multi-lepton analysis. Because this analysis targets events with
1183 multiple light leptons, as well as tau hadrons, this framework skims the dataset of any events that
1184 do not meet at least one of the following requirements:

- 1185 • at least two light leptons within a range $|\eta| < 2.6$, with leading lepton $p_T > 15$ GeV and
1186 subleading lepton $p_T > 5$ GeV
- 1187 • at least one light lepton with $p_T > 15$ GeV within a range $|\eta| < 2.6$, and at least two hadronic
1188 taus with $p_T > 15$ GeV.

1189 Samples were then generated from these `HIGG8D1` derivations using a modified version of
1190 `AnalysisBase` version 21.2.127.

1191 16.1 Data Samples

1192 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through
1193 2018, which represents an integrated luminosity of 139 fb^{-1} and an energy of $\sqrt{s} = 13 \text{ TeV}$. All
1194 data used in this analysis was included in one of the following Good Run Lists:

- 1195 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
1196 _PHYS_StandardGRL_All_Good_25ns.xml
- 1197 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
1198 _PHYS_StandardGRL_All_Good_25ns.xml
- 1199 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL
1200 _All_Good_25ns_Triggerno17e33prim.xml
- 1201 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL
1202 _All_Good_25ns_Triggerno17e33prim.xml

1203 16.2 Monte Carlo Samples

1204 Several Monte Carlo (MC) generators were used to simulate both signal and background pro-
1205 cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific
1206 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)
t̄t(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
t̄tγ	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
t̄t̄t, t̄t̄t̄t	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

17 Object Reconstruction

¹²⁰⁷ All analysis channels considered in this note share a common object selection for leptons and jets, as well as a shared trigger selection.

17.1 Trigger Requirements

¹²¹¹ 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.

17.2 Light Leptons

1212 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that
 1213 are associated with charged particle tracks reconstructed in the inner detector [**ATLAS-CONF-2016-024**].
 1214 Electron candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the
 1215 transition region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| <$
 1216 1.52, are rejected. A multivariate likelihood discriminant combining shower shape and track
 1217 information is used to distinguish prompt electrons from nonprompt leptons, such as those
 1218 originating from hadronic showers.
 1219

1220 To further reduce the non-prompt contribution, the track of each electron is required to

1221 originate from the primary vertex; requirements are imposed on the transverse impact parameter
1222 significance ($|d_0|/\sigma_{d_0}$) and the longitudinal impact parameter ($|\Delta z_0 \sin \theta_\ell|$).

1223 Muon candidates are reconstructed by combining inner detector tracks with track segments
1224 or full tracks in the muon spectrometer [**PERF-2014-05**]. Muon candidates are required to have
1225 $p_T > 10$ GeV and $|\eta| < 2.5$. All leptons are required to be isolated, and pass a non-prompt BDT
1226 selection described in detail in [**ttH_paper**].

1227 **17.3 Jets**

1228 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
1229 calorimeters [**ATL-PHYS-PUB-2015-015**], using the anti- k_t algorithm with a radius parameter
1230 $R = 0.4$. Jets with energy contributions likely arising from noise or detector effects are removed
1231 from consideration [**ATLAS-CONF-2015-029**], and only jets satisfying $p_T > 25$ GeV and
1232 $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
1233 association algorithm is used to confirm that the jet originates from the selected primary vertex,
1234 in order to reject jets arising from pileup collisions [**PERF-2014-03**].

1235 **17.4 Missing Transverse Energy**

1236 Because all $t\bar{t}H - ML$ channels considered include multiple neutrinos, missing transverse
1237 energy (E_T^{miss}) is present in each event. The missing transverse momentum vector is defined as
1238 the inverse of the sum of the transverse momenta of all reconstructed physics objects as well

1239 as remaining unclustered energy, the latter of which is estimated from low- p_T tracks associated
1240 with the primary vertex but not assigned to a hard object [**ATL-PHYS-PUB-2015-027**].

1241 **18 Higgs Momentum Reconstruction**

1242 Reconstructing the momentum of the Higgs boson is a particular challenge for channels with
1243 leptons in the final state: Because all channels include at least two neutrinos in the final state, the
1244 Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted
1245 by a neural network when provided with the four-vectors of the Higgs Boson decay products, as
1246 shown in section 18.1. With this in mind, several layers of MVAs are used to reconstruction the
1247 Higgs momentum.

1248 The first layer is a model designed to select which jets are most likely to be the b-jets that
1249 came from the top decay, detailed in section 18.2. As described in section 18.3, the kinematics
1250 of these jets are fed into the second layer, which is designed to identify the decay products of
1251 the Higgs Boson itself. The kinematics of these particles are then fed into yet another neural-
1252 network, which predicts the momentum of the Higgs (18.4). MVAs are also used in the analysis
1253 to determine the decay of the Higgs boson in the 3l channel (18.5).

1254 Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_AMC.
1255 Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training.

1256 For all of these models, the Keras neural network framework, with Tensorflow as the
1257 backend, is used, and the number of hidden layers and nodes are determined using grid search
1258 optimization. Each neural network uses the LeakyReLU activation function, a learning rate
1259 of 0.01, and the Adam optimization algorithm, as alternatives are found to either decrease or
1260 have no impact on performance. Batch normalization is applied after each layer. For the
1261 classification algorithms (b-jet matching, Higgs reconstruction, and 3l decay identification)
1262 binary-cross entropy is used as the loss function, while the p_T reconstruction algorithm uses
1263 MSE.

1264 The specific inputs features used for each model are arrived at through a process of trial
1265 and error - features considered potentially useful are tried, and those that are found to increase
1266 performance are included. While each model includes a relatively large number of features,
1267 some using upwards of 30, this inclusive approach is found to maximize the performance of each
1268 model while decreasing the variance compared to a reduced number of inputs. Each input feature
1269 is validated by comparing MC simulations to 80 fb^{-1} of data, as shown in the sections below.

1270 **18.1 Decay Candidate Reconstruction**

1271 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson
1272 using MC simulations of $t\bar{t}H$ events. These include light leptons and jets. Reconstructed
1273 physics objects are matched to truth level particles, in order to identify the parents of these

1274 reconstructed objects. The kinematics of the decay product candidates as well as event level
1275 variables are used as inputs.

1276 Leptons considered as possible Higgs and top decay candidates are required to pass the
1277 selection described in section 17.2. For jets, however, it is found that a large fraction that originate
1278 from either the top decay or the Higgs decay fall outside the selection described in section 17.3.
1279 Specifically, jets from the Higgs decay tend to be soft, with 32% having $p_T < 25$ GeV. Therefore
1280 jets with $p_T < 15$ GeV are considered as possible candidates in the models described below. By
1281 contrast, less than 5% of the jets originating from the Higgs fall below this p_T . The jets are found
1282 to be well modeled even down to this low p_T threshold, as shown in section 20.1. The impact of
1283 using different p_T selection for the jet candidates is considered in detail in section A.5. As they
1284 are expected to originate from the primary vertex, jets are also required to pass a JVT cut.

1285 18.2 b-jet Identification

1286 Including the kinematics of the b-jets that originate from the top decay is found to improve the
1287 identification of the Higgs decay products, and improve the accuracy with which the Higgs
1288 momentum can be reconstructed. Because these b-jets are reconstructed by the detector with
1289 high efficiency (just over 90% of the time), and can be identified relatively consistently, the first
1290 step in reconstructing the Higgs is selecting the b-jets from the top decay.

1291 Exactly two b-jets are expected in the final state of $t\bar{t}H - \text{ML}$ events. However, in both
1292 the 3l and 2lSS channels, only one or more b-tagged jets are required (where the 70% DL1r b-tag

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

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

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

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

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

Table 30: Input features used in the b-jet identification algorithm for the 2lSS channel

1311 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for
 1312 training.

1313 The difference between the distributions for a few of these features for the correct(i.e.
 1314 both jets are truth b-jets), and incorrectcombinations are shown in figure 18.1. The correct and
 1315 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences
 1316 in the distributions.

1317 The modeling of these inputs is validated against data, with figure 18.2 showing good
 1318 general agreement between data and MC. Plots for the complete list of features can found in
 1319 section A.

1320 Based on the results of grid search evaluation, the optimal architecture is found to include
 1321 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as
 1322 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for

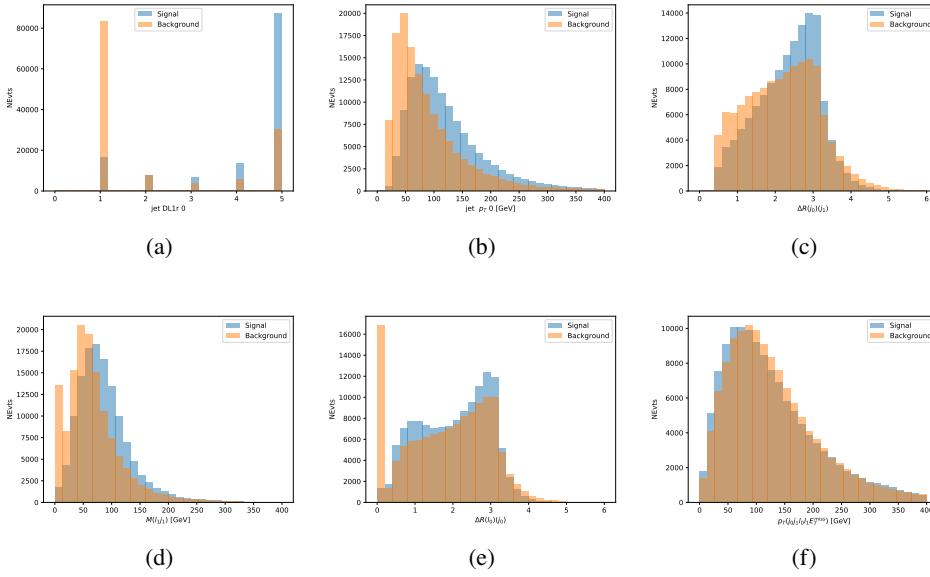


Figure 18.1: Input features for top2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T 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} .

1323 the trained model are shown in figure 18.2.1. The model is found to identify the correct pairing
 1324 of jets for 73% of 2lSS signal events on test data.

1325 For point of comparison, a naïve approach to identify b-jets is used as well: The two jets
 1326 which pass the highest DL1r b-tag working point are assumed to be the b-jets from the top decay.
 1327 In the case that multiple jets meet the same b-tag working point, the jet with higher p_T is used.
 1328 This method identifies the correct jet pair 65% of the time.

1329 The accuracy of the model for different values of n-bjets, compared to this naive approach,
 1330 is shown in table 31.

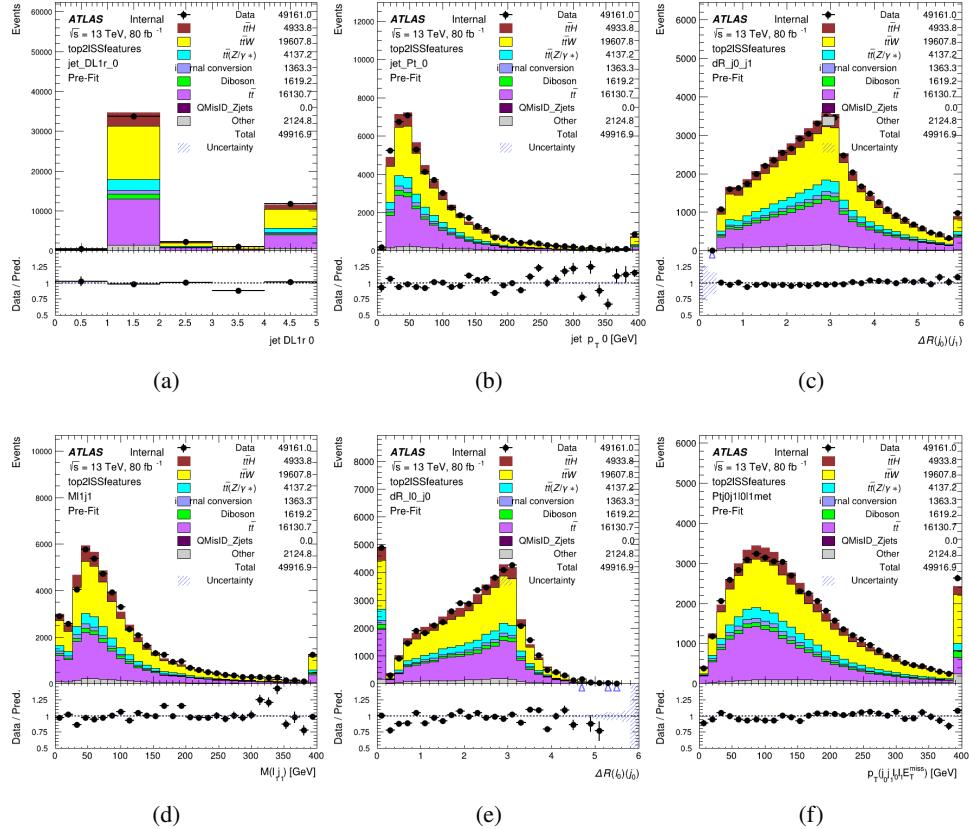


Figure 18.2: Data/MC comparisons of input features for top2ISS training for 80 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (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.

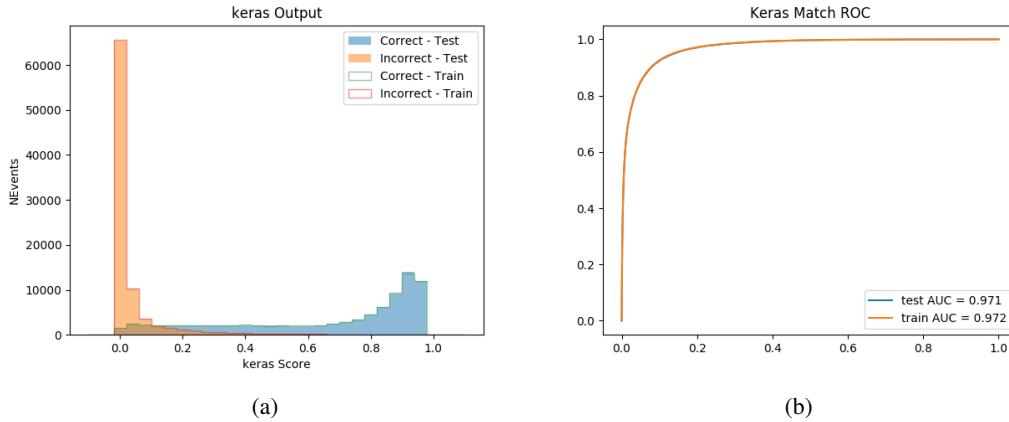


Figure 18.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

18.2.2 3l Channel

The input features used in the 3l channel are listed in table 32, with the same naming convention as the 2lSS channel.

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

A few of these features are shown in figure 18.4, comparing the distributions for correct

1335 and incorrect combinations of jets.

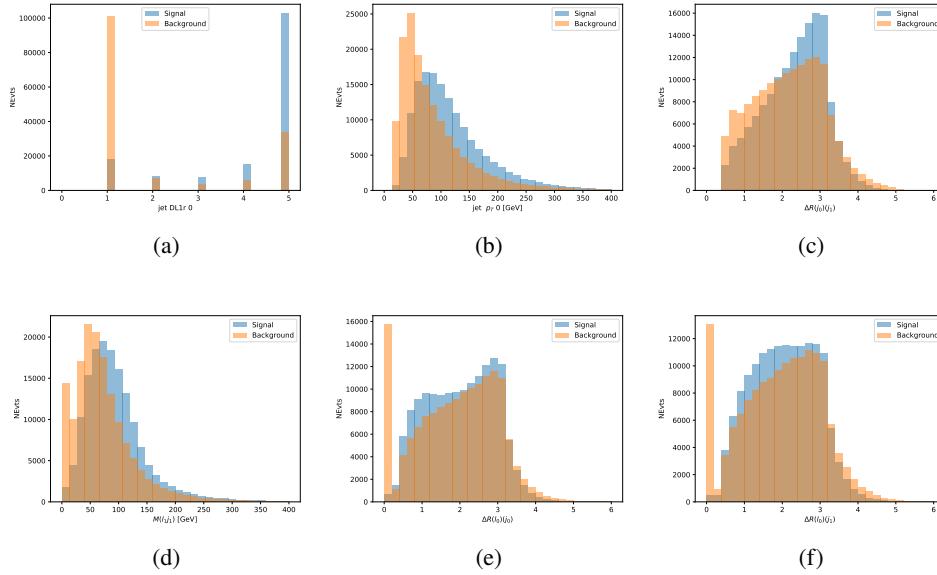


Figure 18.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and 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

1336 The modeling of these inputs is validated against data, with figure 18.5 showing good
 1337 general agreement between data and MC. Plots for the complete list of features can found in
 1338 section A.

1339 Again, the dataset is downsized to reduce the ratio of correct and incorrect combination
 1340 from 20:1, to 5:1. Around 7 million events are used for training, with 10% set aside for testing.
 1341 Based on the results of grid search evaluation, the optimal architecture is found to include 5
 1342 hidden layers with 60 nodes each. The output score distribution as well as the ROC curve for the
 1343 trained model are shown in figure 18.2.2.

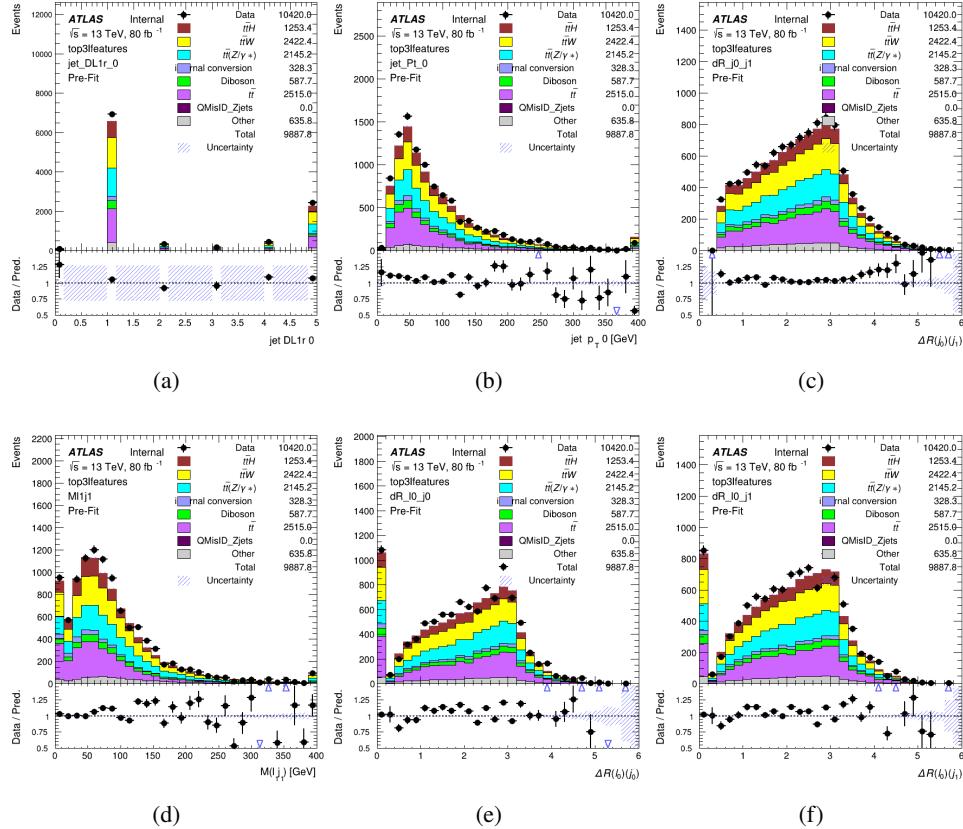


Figure 18.5: Data/MC comparisons of input features for top3l training for 80 fb^{-1} of data. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

1344 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal
1345 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%

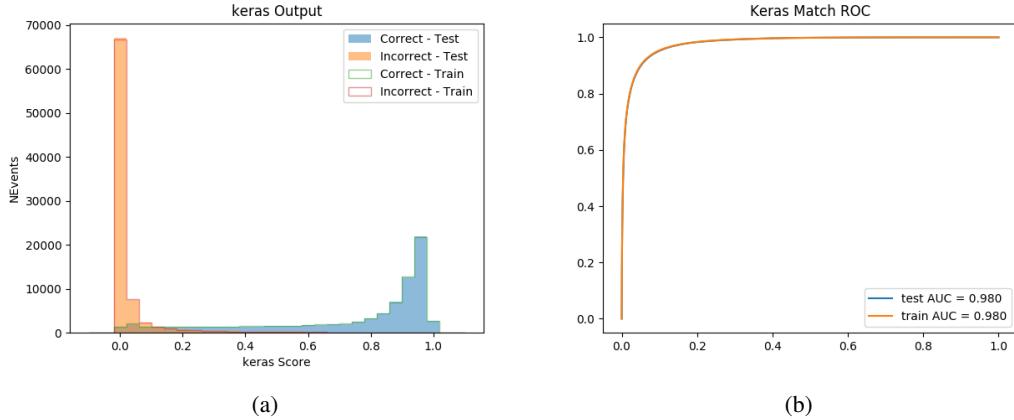


Figure 18.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

18.3 Higgs Reconstruction

1346 Techniques similar to the b-jet identification algorithms are employed to select the decay products
 1347 of the Higgs: kinematics of all possible combinations of reconstructed objects are fed into a neural
 1348 network to determine which of those is most mostly to be the decay products of the Higgs.

1350 Again separate models are used for the 2lSS and 3l channels, while the 3l channel has now
 1351 been split into two: $t\bar{t}H$ events with three leptons in the final state include both instances where
 1352 the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the Higgs
 1353 decays to two leptons.

1354 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In
 1355 the semi-leptonic case the reconstructed decay products consist of two jets and a single leptons.
 1356 For the fully-leptonic case, the decay products include 2 of the three leptons associated with the

1357 event. For training the models, events are separated into these two categories using truth level
1358 information. A separate MVA, described in section 18.5, is used to make this distinction at reco
1359 level and determine which model to use.

1360 For all channels, the models described in section 18.2 are used to identify b-jet candidates,
1361 whose kinematics are used to identify the Higgs decay products. These jets are not considered
1362 as possible candidates for the Higgs decay, justified by the fact that these models are found to
1363 misidentify jets from the Higgs decay as jets from the top decay less than 1% of the time.

1364 **18.3.1 2lSS Channel**

1365 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural
1366 network is trained on the kinematics of different combinations of leptons and jets, as well as the
1367 b-jets identified in section 18.2, 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

1368 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 1369 lepton in the event is labeled l_T , as it is assumed to have come from the decay of one of the top
 1370 quarks. b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet
 1371 Reco Score is the output of the b-jet reconstruction algorithm.

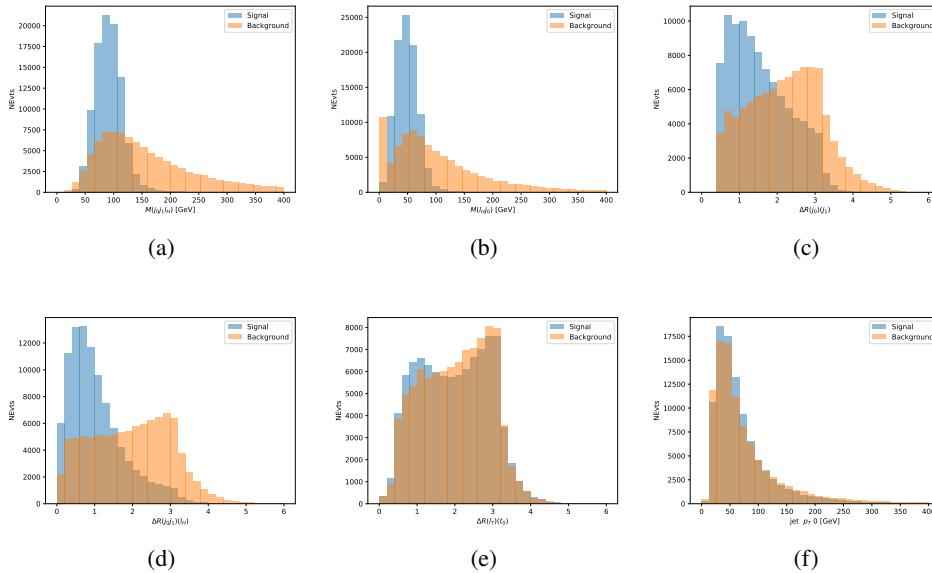


Figure 18.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (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.

1372 The modeling of these inputs is validated against data, with figure 18.2 showing good
 1373 general agreement between data and MC. Plots for the complete list of features can found in
 1374 section A.

1375 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2
 1376 million events, with an additional 200,000 reserved for testing the model. In order to compensate

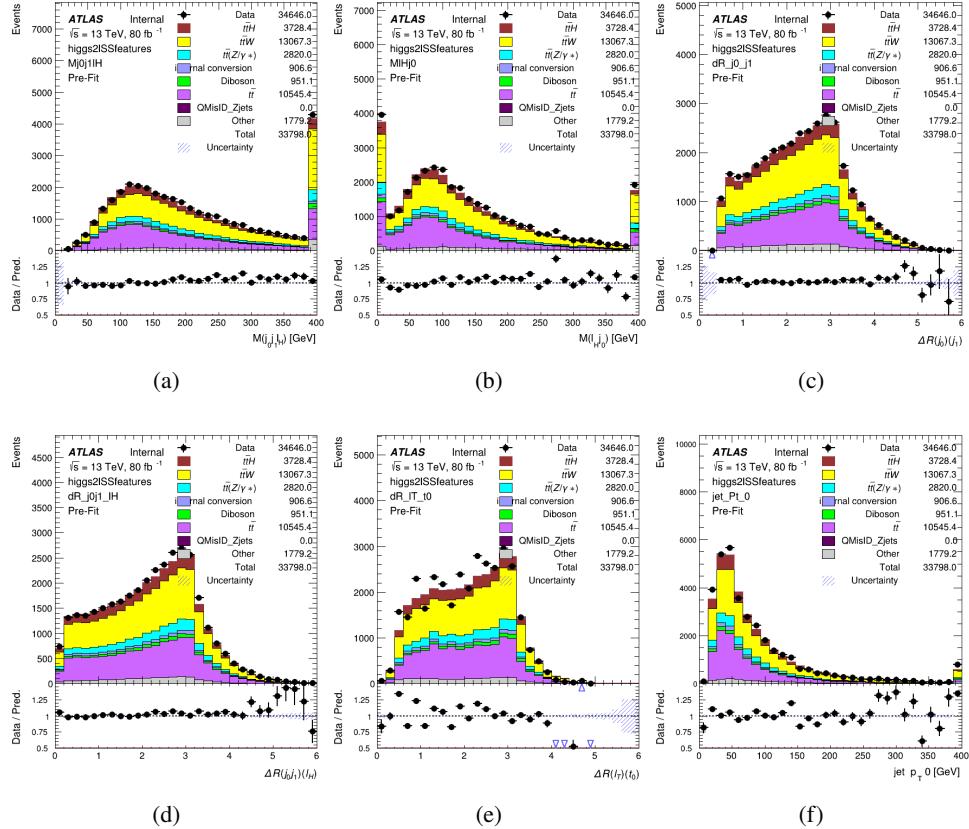


Figure 18.8: Data/MC comparisons of input features for higgs2lSS training for 80 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 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 18.3.1.

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.

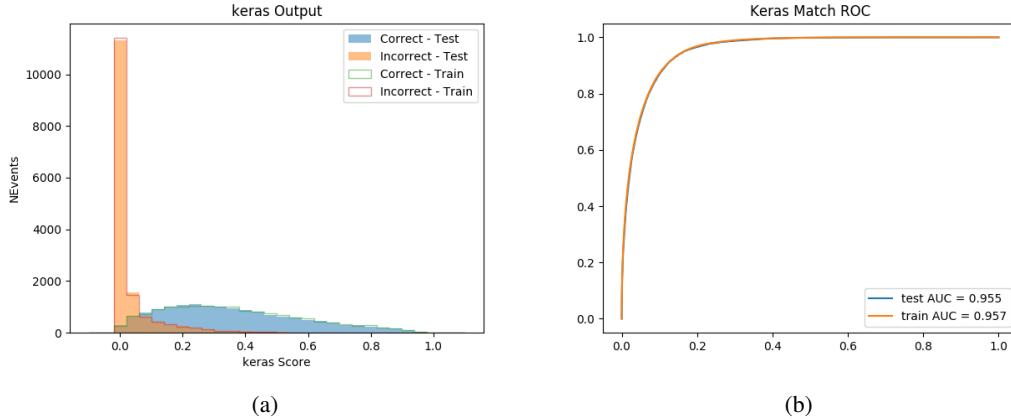


Figure 18.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

18.3.2 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

1388 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 1389 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
 1390 the b-jet identification algorithm. The b-jet Reco Score is the output of the Higgs reconstruction
 1391 algorithm.

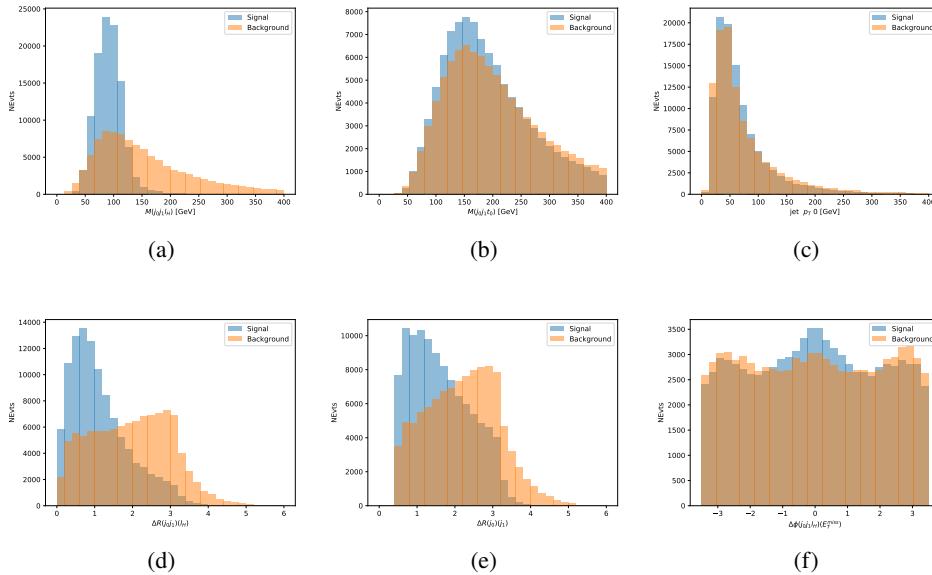


Figure 18.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1392 The modeling of these inputs is validated against data, with figure 18.11 showing good
 1393 general agreement between data and MC. Plots for the complete list of features can found in
 1394 appendix A.1.

1395 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.
 1396 Once again, incorrect combinations are downsampled, such that the correct combinations are
 1397 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN

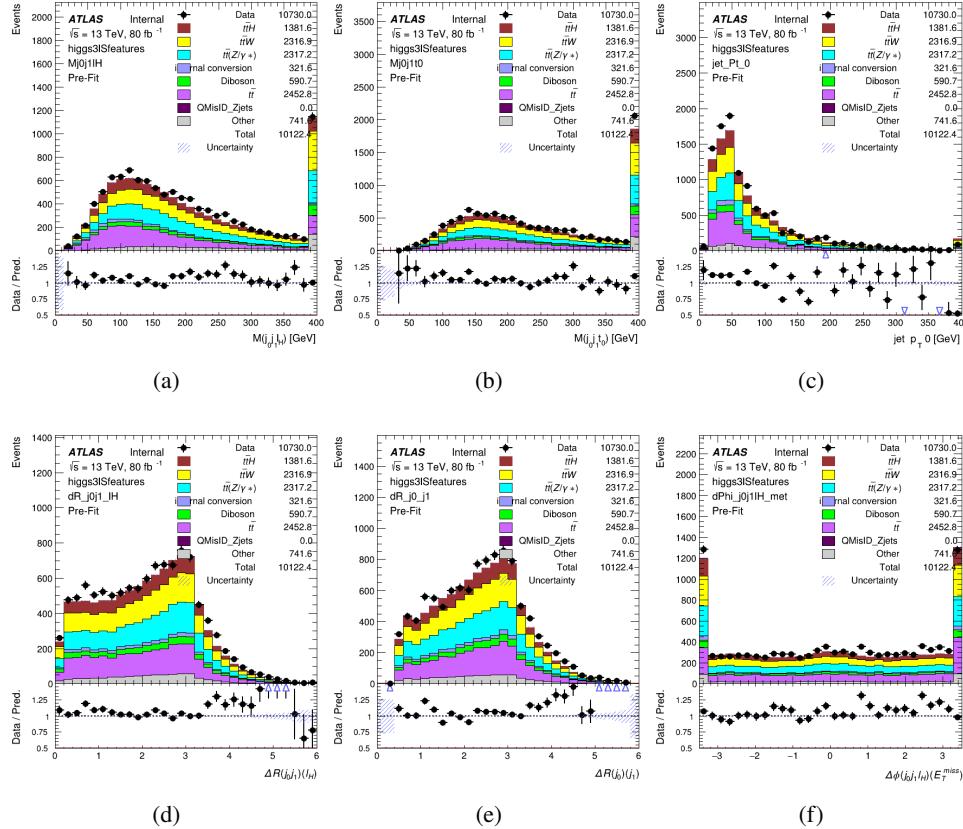


Figure 18.11: Data/MC comparisons of input features for higgs3IS training for 80 fb^{-1} of data.

1398 is summarized in figure 18.3.2.

1399 The neural network identifies the correct combination 64% of the time. It identifies the
1400 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
1401 83% of the time.

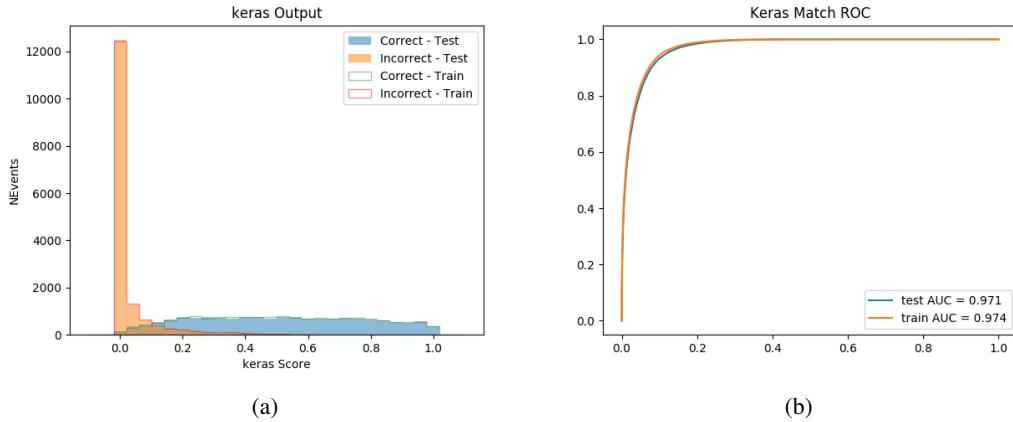


Figure 18.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1402 18.3.3 3l Fully-leptonic Channel

1403 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 1404 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to
 1405 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 1406 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 1407 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

1408 Here l_{H0} and l_{H1} are the Higgs decay candidates. The other lepton in the event is labeled
 1409 l_T . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco
 1410 Score is the output of the Higgs reconstruction algorithm.

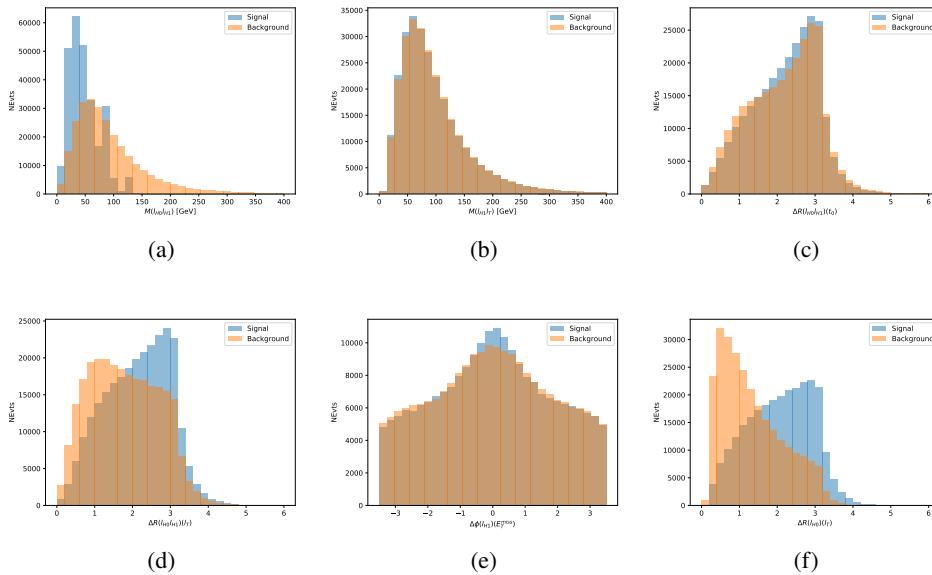


Figure 18.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1411 The modeling of these inputs is validated against data, with figure 18.14 showing good
 1412 general agreement between data and MC. Plots for the complete list of features can found in
 1413 section A.

1414 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,
 1415 with 10% of the dataset reserved for testing. The output of the model is summarized in figure
 1416 18.3.3.

1417 The correct lepton is identified by the model for 80% of events in the testing data set.

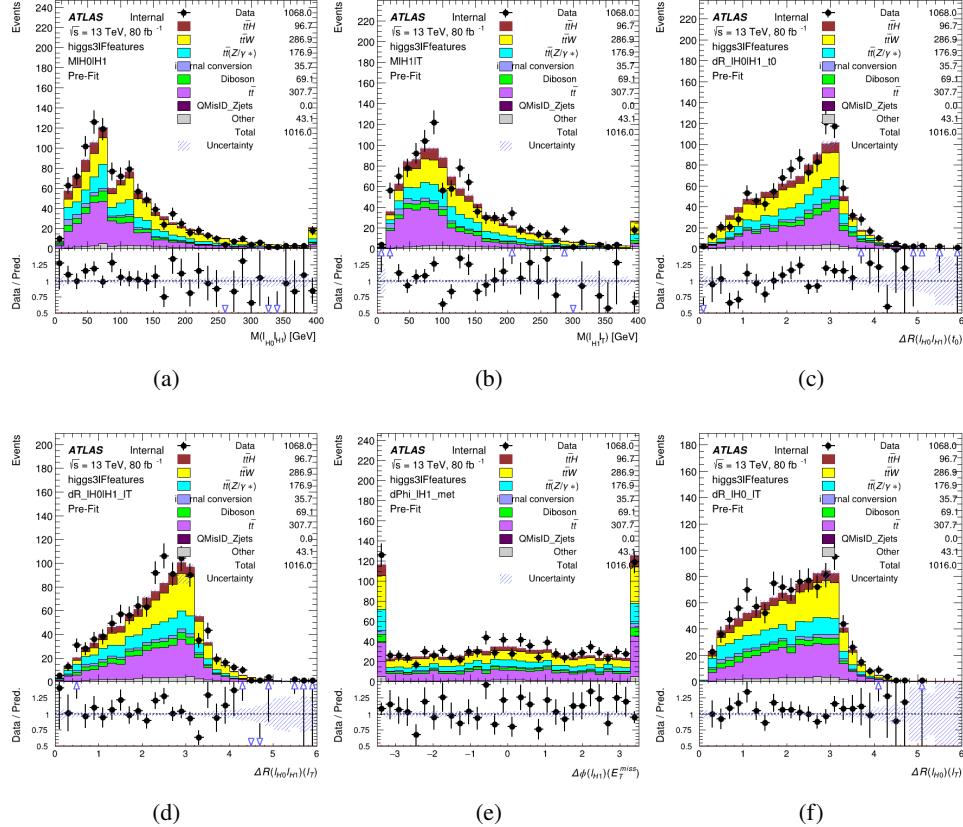


Figure 18.14: Data/MC comparisons of input features for higgs3IF training for 80 fb^{-1} of data.

1418 18.4 p_T Prediction

1419 Once the most probable decay products have been identified, their kinematics are used as inputs
 1420 to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,
 1421 a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay
 1422 are included as well, as these are found to improve performance. The truth p_T of the Higgs,
 1423 as predicted by MC, are used as labels. Separate models are built for each channel - 2lSS, 3l
 1424 Semi-leptonic and 3l Fully-leptonic.

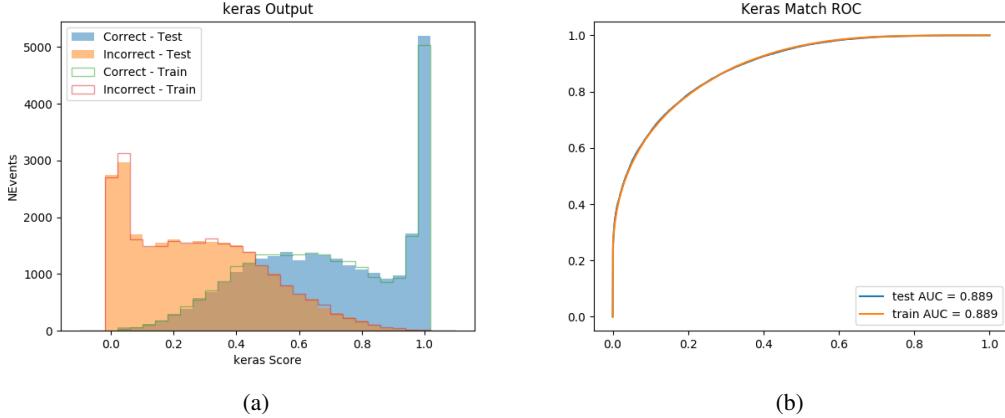


Figure 18.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

1425 As a two-bin fit is targeted for the final result, some metrics evaluating the performance
 1426 of the models aim to show how well it distinguished between "high p_T " and "low p_T " events. A
 1427 cutoff point of 150 GeV is used to define these two categories.

1428 Because the analysis uses a two bin fit of the Higgs p_T , the momentum reconstruction
 1429 could be treated as a binary classification problem, rather than a regression problem. This
 1430 approach is explored in detail in section A.4, and is found not to provide any significant increase
 1431 in sensitivity. The regression approach is used because it provides more flexibility for future
 1432 analyses, as it is independent of the cutoff between high and low p_T , as well as the number of
 1433 bins. Further, a regression allows the output of the neural network to be more clearly understood,
 1434 as it can be directly compared to a physics observable.

1435 18.4.1 2lSS Channel

1436 The input variables listed in table 37 are used to predict the Higgs p_T in the 2lSS channel. Here
1437 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
1438 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
1439 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
1440 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
1441 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

1442 The optimal neural network architecture for this channel is found to consist of 7 hidden
 1443 layers with 60 nodes each. The input data set includes 1.2 million events, 10% of which is used
 1444 for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1445 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 1446 truth Higgs p_T in figure 18.16. In order to visualize the model performance more clearly, in (a)
 1447 of that figure, the color of each point is determined by Kernel Density Estimation (KDE). The
 1448 color shown represents the logarithm of the output from KDE, to counteract the large number of
 1449 low p_T events. For that same reason, each column of the histogram shown in (b) of figure 18.16
 1450 is normalized to unity. This plot therefore demonstrates what the model predicts for each slice
 1451 of truth p_T .

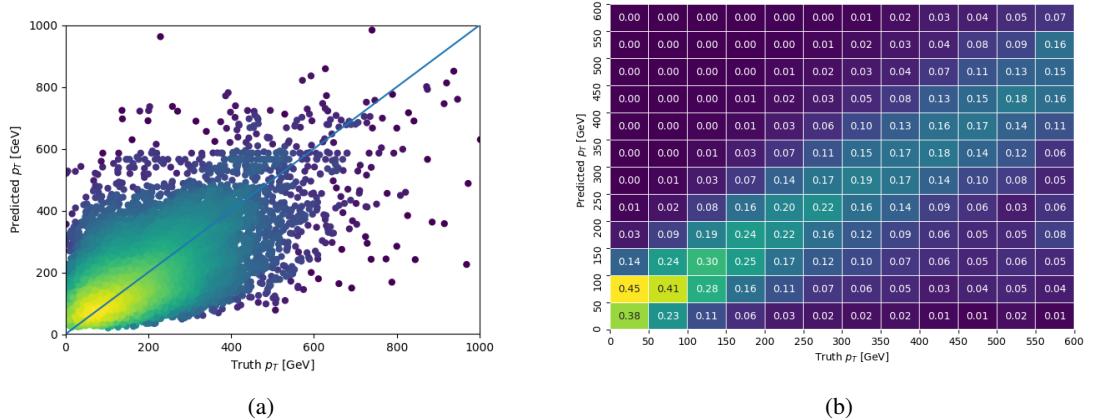


Figure 18.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1452 We are also interested in how well the model distinguishes between events with $p_T < 150$
 1453 GeV and > 150 GeV. Figure 18.17 demonstrates the NN output for high and low p_T events based

1454 on this cutoff.

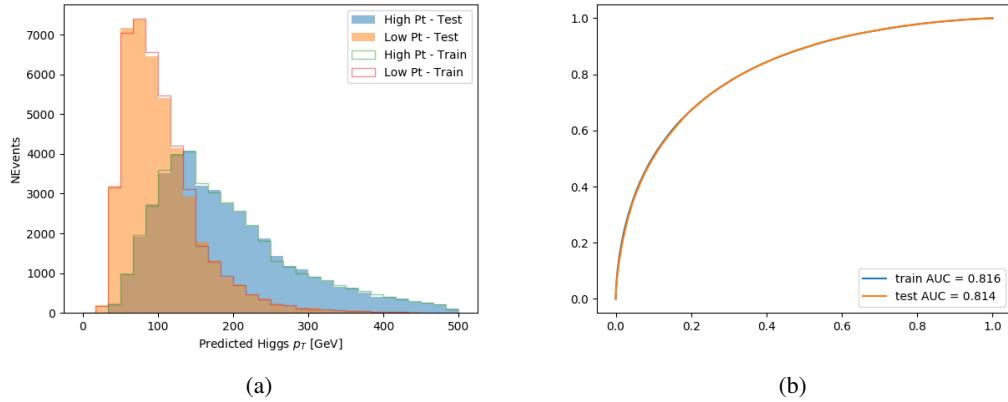


Figure 18.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

1455 **18.4.2 3l Semi-leptonic Channel**

1456 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 18.18. Once again, (a) of 18.18 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 18.19 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.

18.4.3 3l Fully-leptonic Channel

The features listed in 39 are used to construct a model for predictin the Higgs p_T for 3lF events.

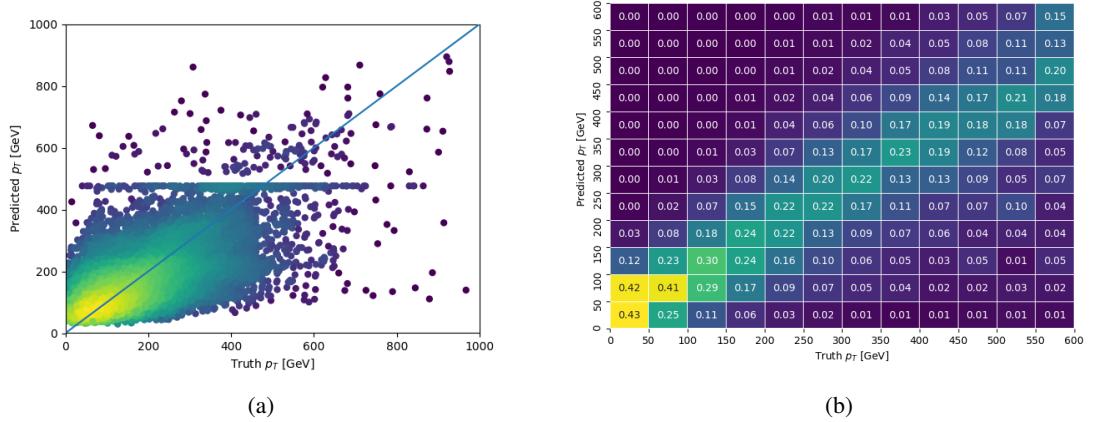


Figure 18.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

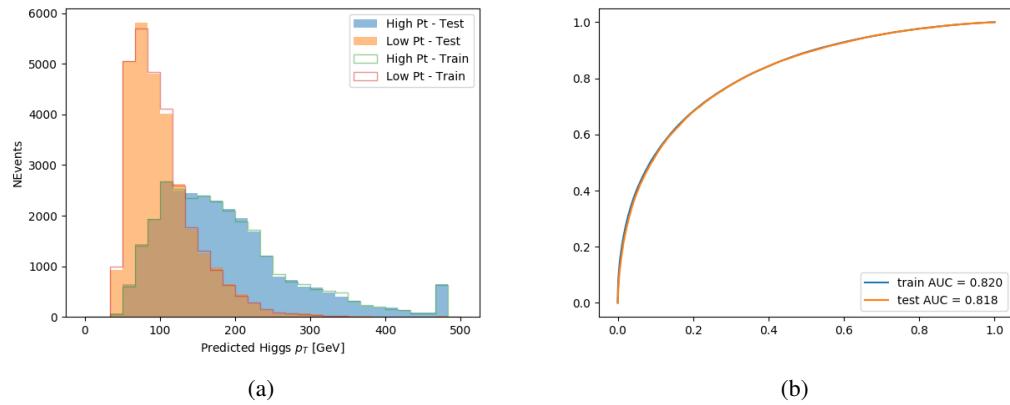


Figure 18.19: (a) shows the reconstructed Higgs p_T for 3lS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

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

1475 l_{H0} and l_{H1} represent the two leptons identified by the Higgs reconstruction model as
 1476 originating from the Higgs, while l_T is the other lepton in the event. The Higgs Reco Score and
 1477 b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification
 1478 algorithm, respectively.

1479 The optimal neural network architecture for this channel is found to consist of 5 hidden
 1480 layers with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for
 1481 testing, the other 90% for training. Training is found to converge after around 150 epochs.

1482 The predicted transverse momentum, as a function of the truth p_T , is shown in figure
 1483 [18.20](#).

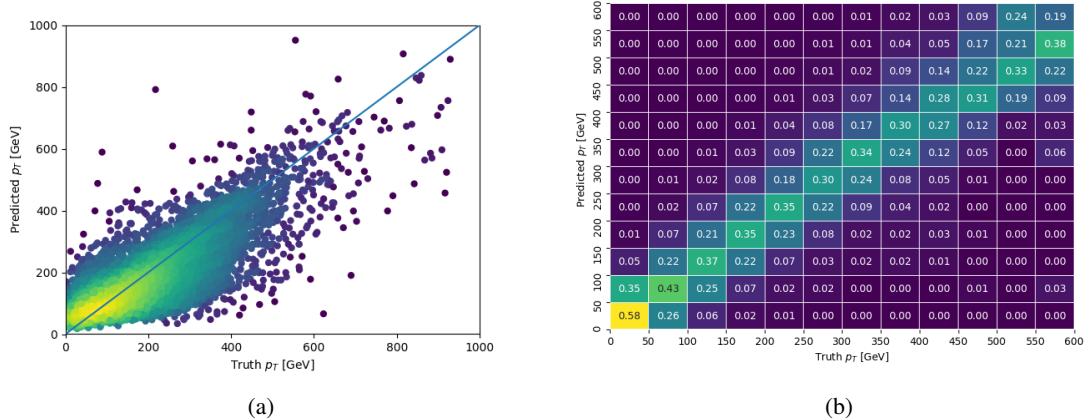


Figure 18.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1484 When split into high and low p_T , based on a cutoff of 150 GeV, the

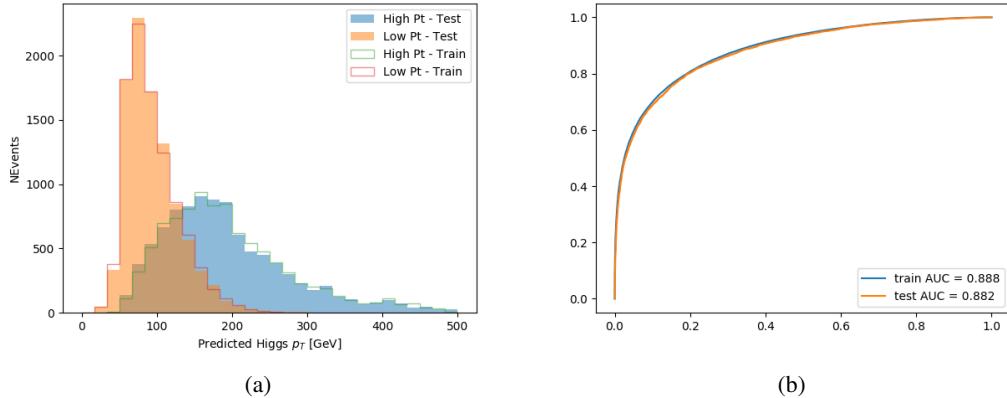


Figure 18.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

1485 18.5 3l Decay Mode

1486 In the 3l channel, there are two possible ways for the Higgs to decay, both involving intermediate
 1487 W boson pairs: Either both W bosons decay leptonically, in which case the reconstructed decay
 1488 consists of two leptons (referred as the fully-leptonic 3l channel), or one W decays leptonically
 1489 and the other hadronically, giving two jets and one lepton in the final state (referred to as the
 1490 semi-leptonic 3l channel). In order to accurately reconstruct the Higgs, it is necessary to identify
 1491 which of these decays took place for each 3l event.

1492 The kinematics of each event, along with the output scores of the Higgs and top recon-
 1493 struction algorithms, are used to distinguish these two possible decay modes. The particular
 1494 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.

1495 Here l_0 is the opposite charge lepton, l_1 and l_2 are the two SS leptons order by ΔR
 1496 from lepton 0. score3lF and score3lS are the outputs of the 3lS and 3lF Higgs reconstruction
 1497 algorithms, while topScore is the output of the b-jet identification algorithm.

1498 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these
 1499 two decay modes. The output of the model is summarized in figure 18.22.

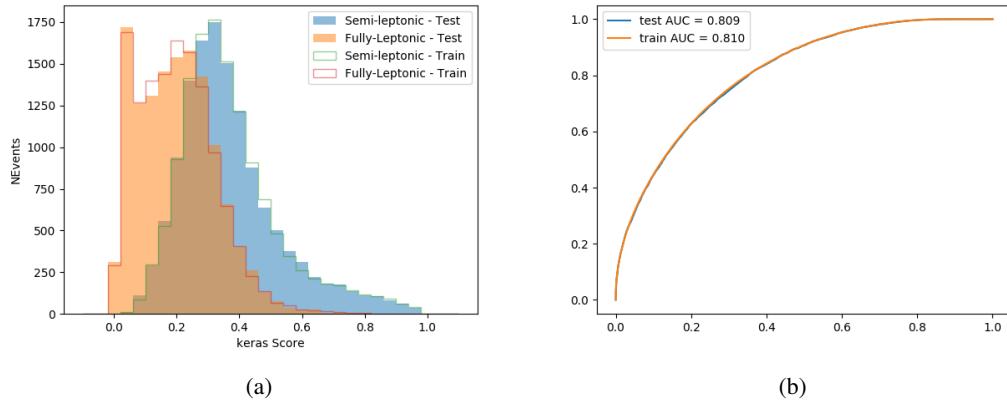


Figure 18.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

1500 A cutoff of 0.23 is determined to be optimal for separating 3lS and 3lF in the fit.

1501 19 Signal Region Definitions

1502 Events are divided into two channels based on the number of leptons in the final state: one with
 1503 two same-sign leptons, the other with three leptons. The 3l channel includes events where both
 1504 leptons originated from the Higgs boson as well as events where only one of the leptons

1505 **19.1 Pre-MVA Event Selection**

1506 A preselection is applied to define orthogonal analysis channels based on the number of leptons
1507 in each event. For the 2lSS channel, the following preselection is used:

- 1508 • Two very tight, same-charge, light leptons with $p_T > 20 \text{ GeV}$
1509 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets
1510 • No reconstructed tau candidates

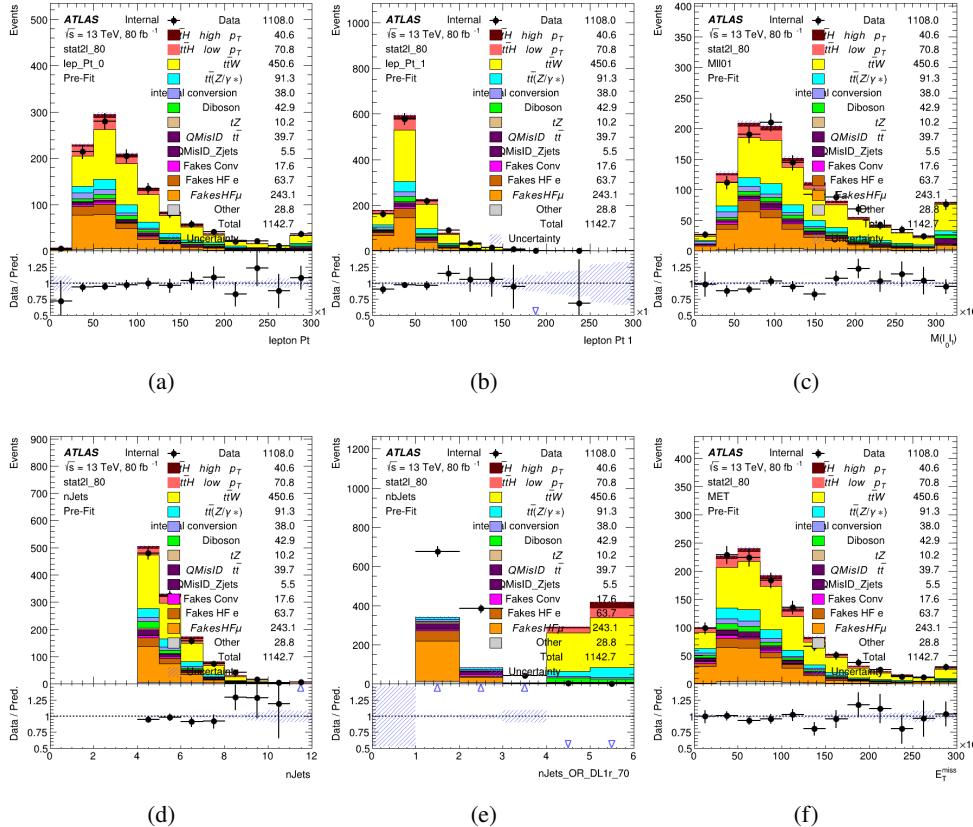
1511 The event yield after the 2lSS preselection has been applied, for MC and data at 80 fb^{-1} ,
1512 is shown in table 41.

	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.

1513

figure 20.1. Good general agreement is found.

Figure 19.1: 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.

1514

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

1515

- 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

1516

- 1518 • ≥ 2 reconstructed jets, ≥ 1 b-tagged jets
- 1519 • No reconstructed tau candidates
- 1520 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs
- 1521 The event yield after the 3l preselection has been applied, for MC and data at 80 fb^{-1} , is
- 1522 shown in table 19.1.

	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

1523

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

1524

20.2.

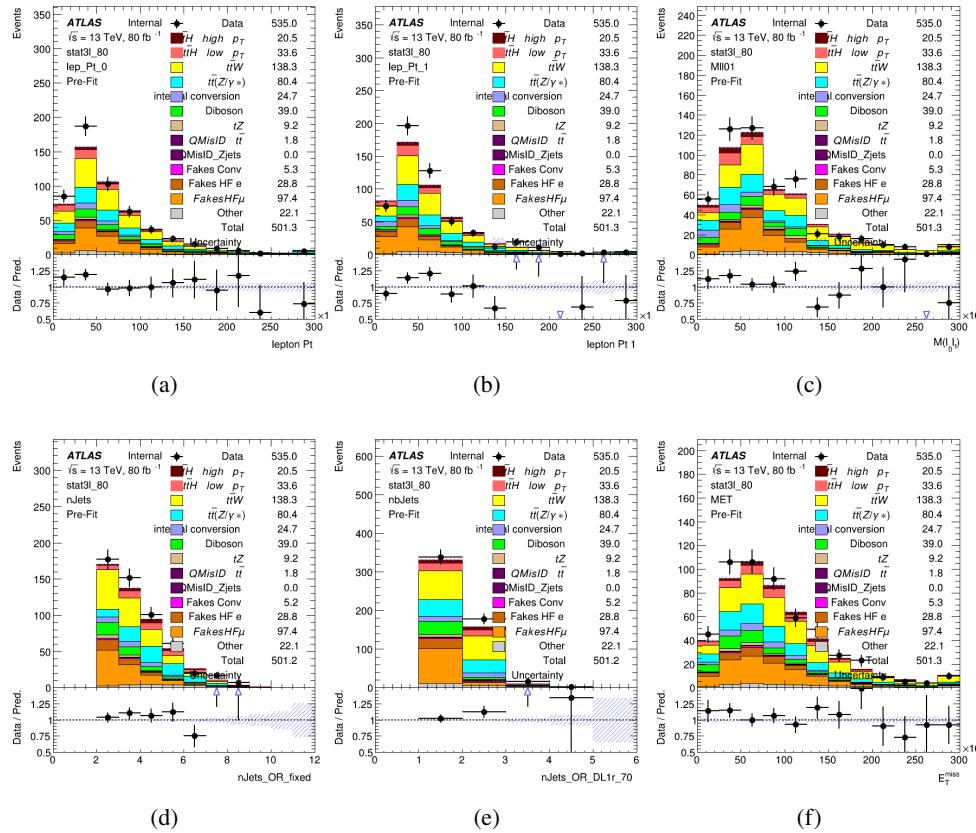


Figure 19.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1525

19.2 Event MVA

1526

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

1527

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

1528

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

1529 [xgboost] are trained using the kinematics of signal and background events derived from Monte
1530 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo
1531 event.

1532 Because the background composition differs for events with a high reconstructed Higgs p_T
1533 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and
1534 low p_T regions. This is found to provide better significance than attempting to build an inclusive
1535 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6
1536 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
1537 3lF high p_T , and 3lF low p_T .

1538 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.

1539

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.

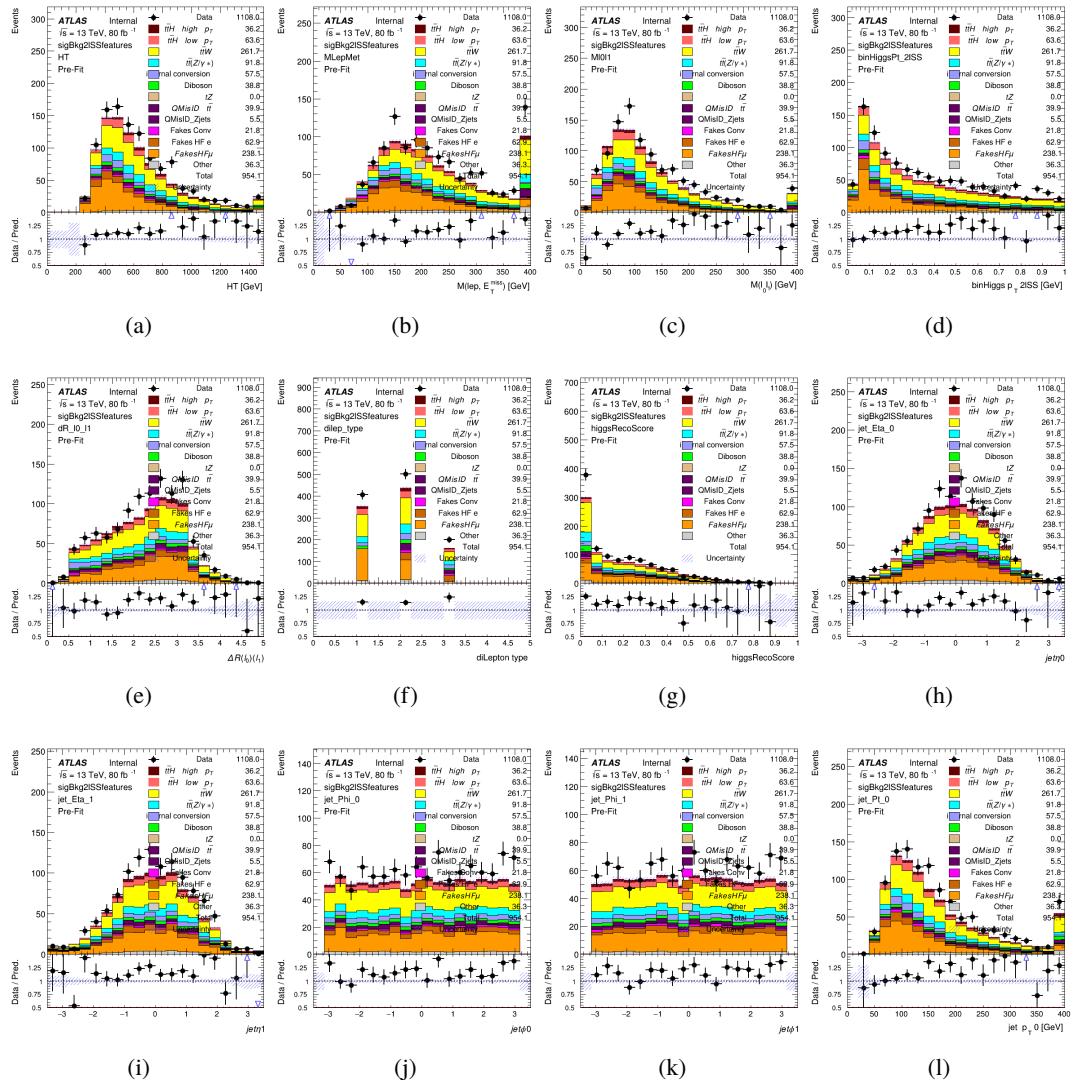


Figure 19.3:

1540 The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss

1542 Output distributions of each MVA comparing MC prediction to data at 80 fb^{-1} are shown
1543 in figures 19.7-19.2.

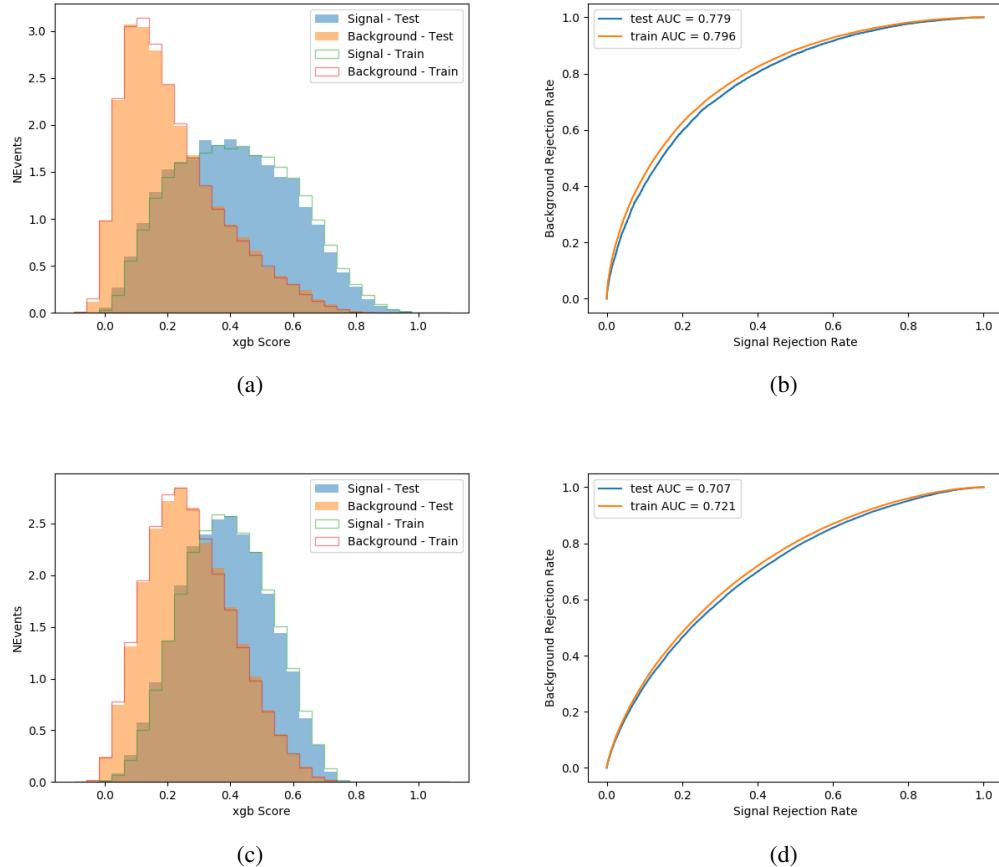


Figure 19.4:

1544 **19.3 Signal Region Definitions**

1545 Once pre-selection has been applied, channels are further refined based on the MVAs described
1546 above. The output of the model described in section 18.5 is used to separate the three channel
1547 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs
1548 boson. This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

1549 For each event, depending on the number of leptons as well as whether the p_T of the Higgs

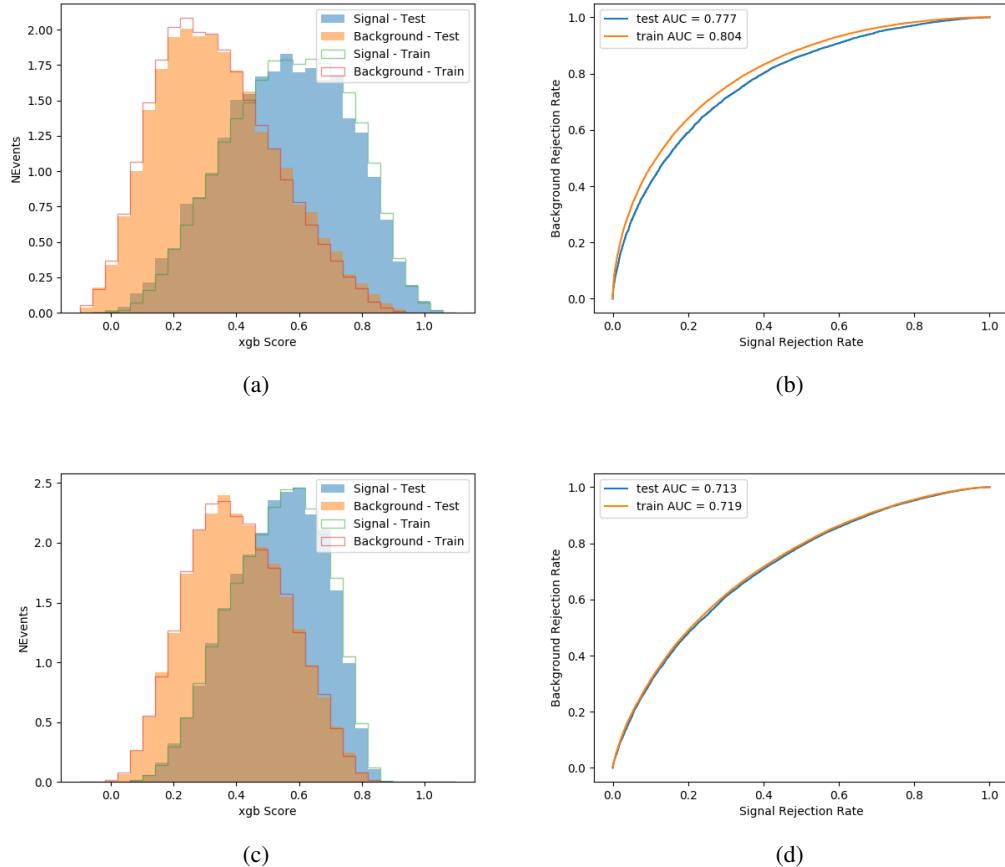


Figure 19.5:

1550 is predicted to be high (> 150 GeV) or low (< 150 GeV), a cut on the appropriate background
 1551 rejection MVA is applied. The particular cut values, listed in table 45, are determined by
 1552 maximizing S/\sqrt{B} in each region.

1553 The event preselection and MVA selection define the three signal regions. These signal
 1554 region definitions are summarized in table 46.

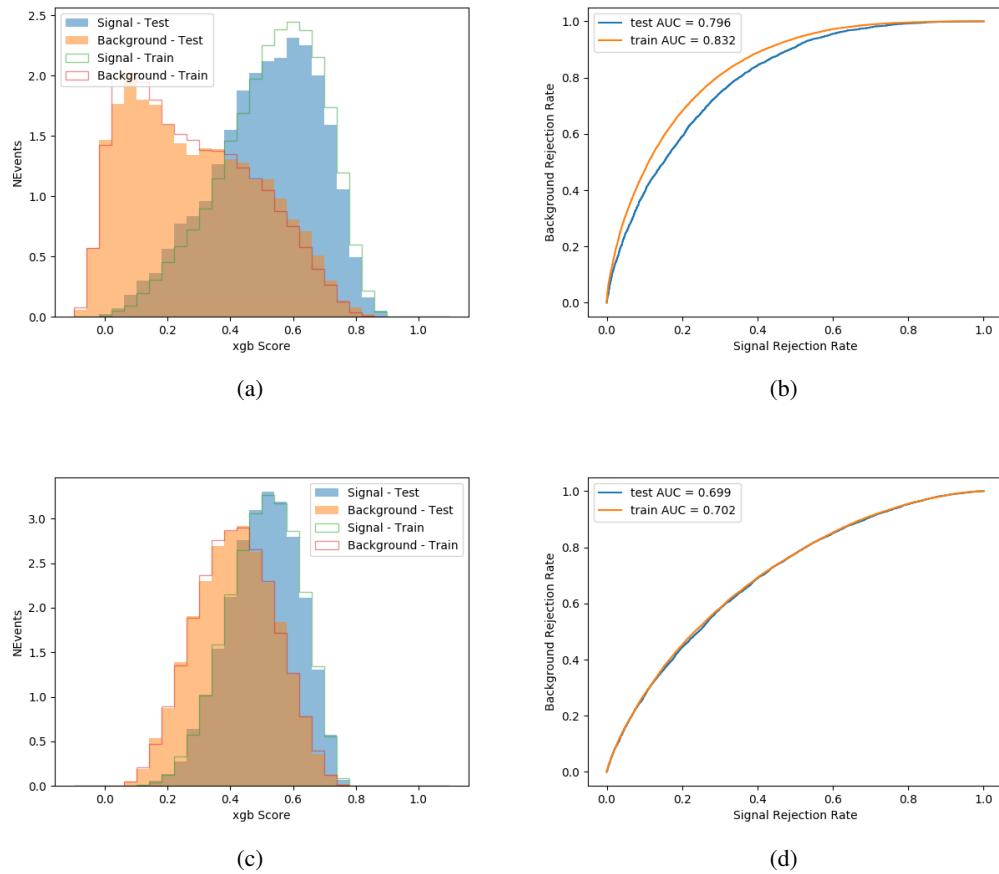
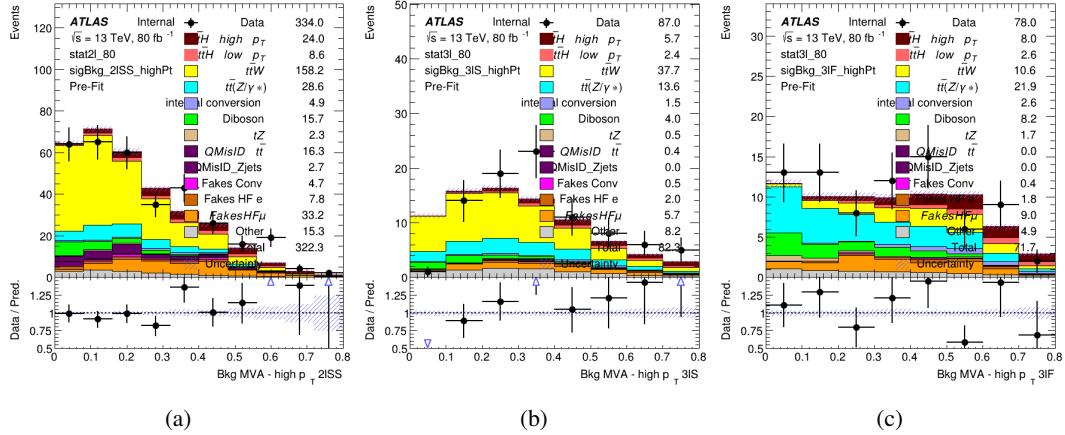
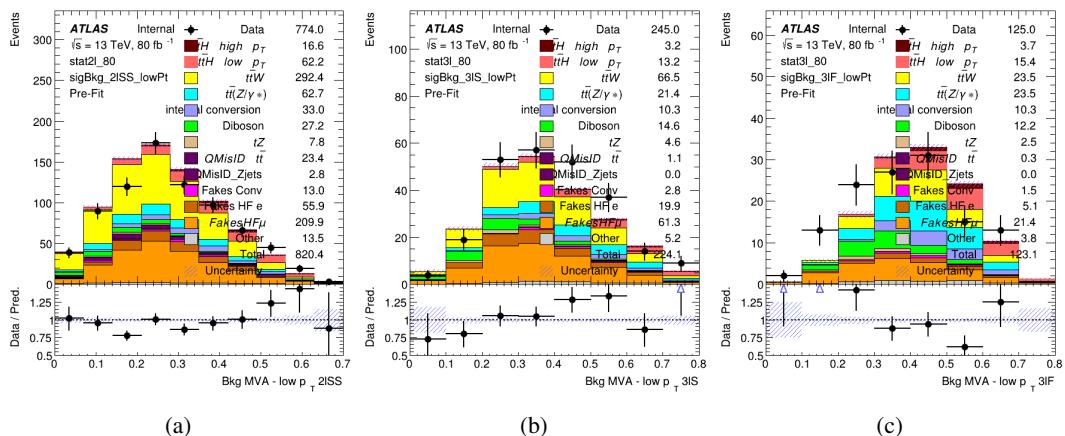


Figure 19.6:

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 19.7: Output score of the high p_T BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channelsFigure 19.8: Output score of the low p_T BDTs in the (a) 2ISS, (b) 3IS, and (c) 3IF channels

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

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

20 Background Rejection MVA

1555 Events are divided into two channels based on the number of leptons in the final state: one with
 1556 two same-sign leptons, the other with three leptons. The 3l channel includes events where both
 1557 leptons originated from the Higgs boson as well as events where only one of the leptons

1559 20.1 Pre-MVA Event Selection

1560 A preselection is applied to define orthogonal analysis channels based on the number of leptons
 1561 in each event. For the 2lSS channel, the following preselection is used:

- 1562 • Two very tight, same-charge, light leptons with $p_T > 20$ GeV

- 1563 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets

- 1564 • No reconstructed tau candidates

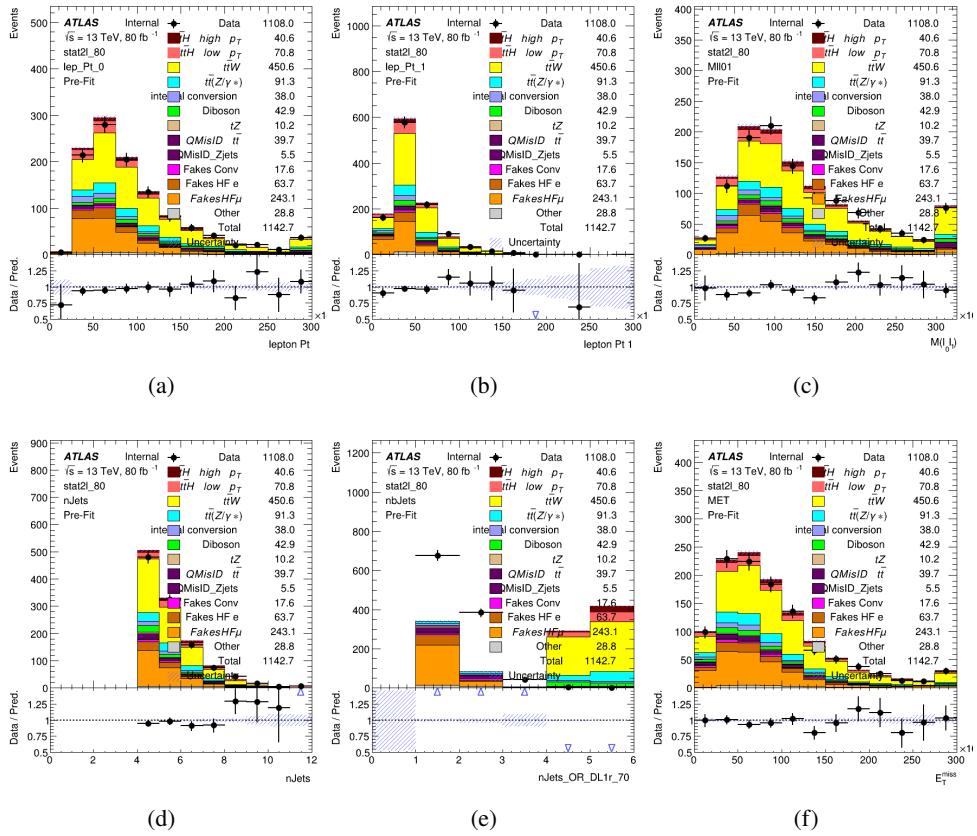


Figure 20.1:

1565 For the 3l channel, the following selection is applied:

- 1566 • Three light leptons with total charge ± 1
- 1567 • Same charge leptons are required to be very tight, with $p_T > 20 \text{ GeV}$
- 1568 • Opposite charge lepton must be loose, with $p_T > 10 \text{ GeV}$
- 1569 • $>= 2$ reconstructed jets, $>= 1$ b-tagged jets
- 1570 • No reconstructed tau candidates
- 1571 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs

1572 **20.2 Event MVA**

1573 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
1574 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
1575 (3lF). In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost
1576 [**xgboost**] are trained using the kinematics of signal and background events derived from Monte
1577 Carlo simulations. Events are weighted in the BDT training by the weight of each Monte Carlo
1578 event.

1579 Because the background composition differs for events with a high reconstructed Higgs p_T
1580 compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and

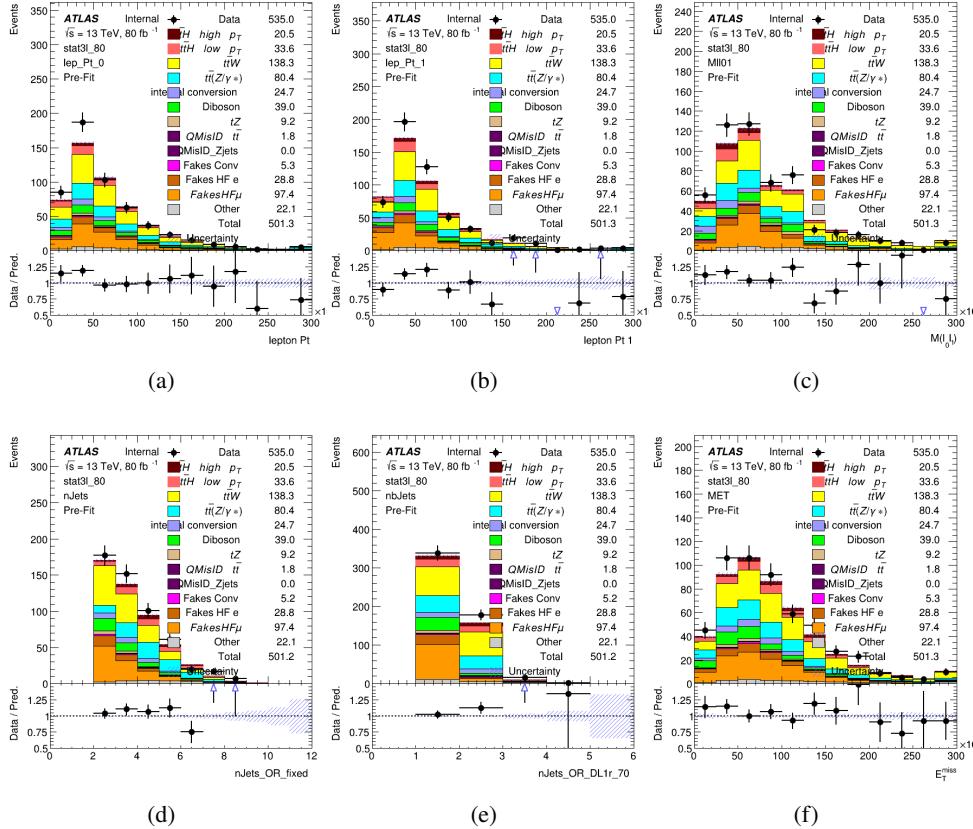


Figure 20.2:

1581 low p_T regions. This is found to provide better significance than attempting to build an inclusive
 1582 model, as demonstrated in appendix A.2. A cutoff of 150 GeV is used. This gives a total of 6
 1583 background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
 1584 3lF high p_T , and 3lF low p_T .

1585 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.

1586

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.

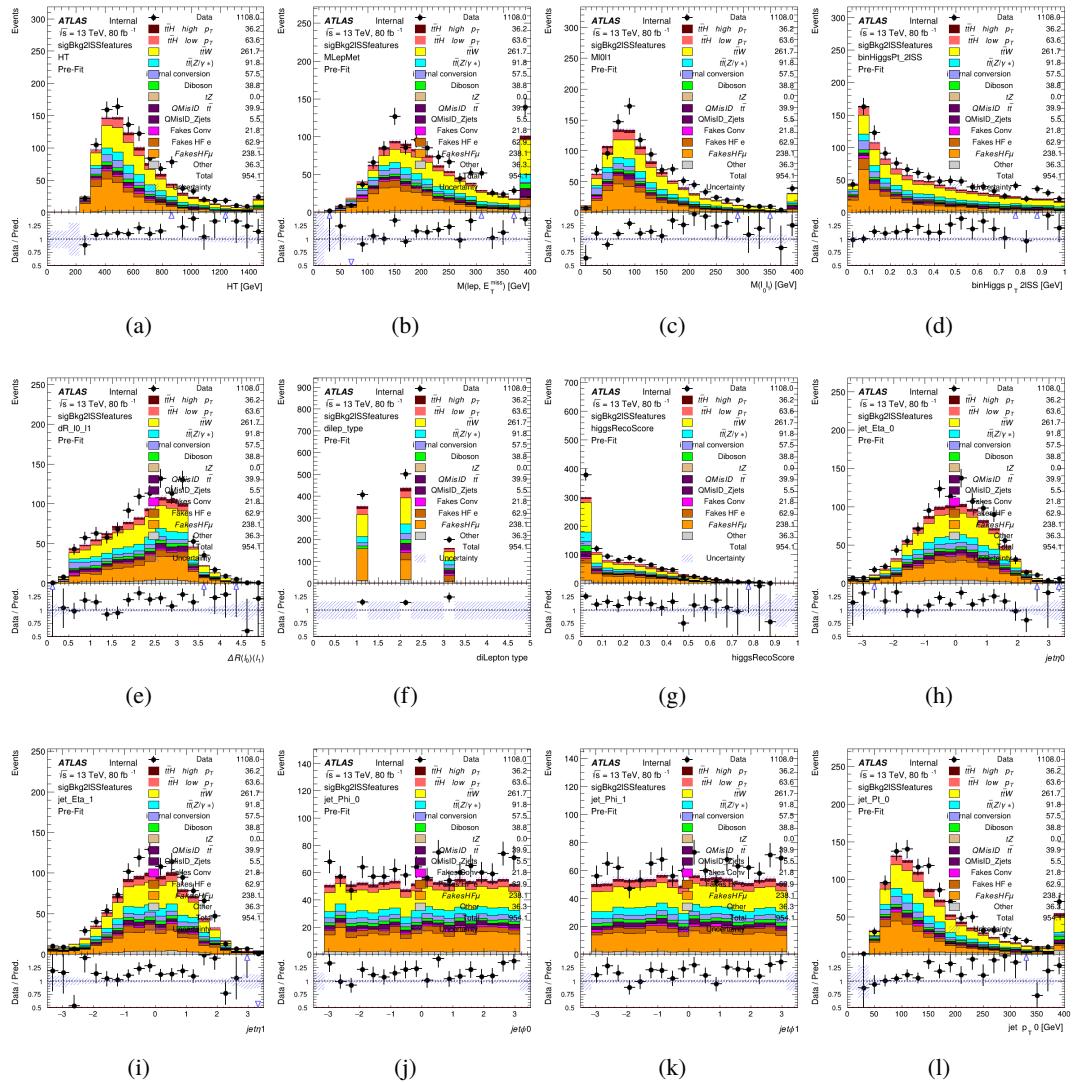


Figure 20.3:

1587 The BDTs are produced with a maximum tree depth of 6, using AUC as the target loss
 1588 function.

1589 Output distributions of each MVA comparing MC prediction to data at 80 fb^{-1} are shown
 1590 in figure 20.2.

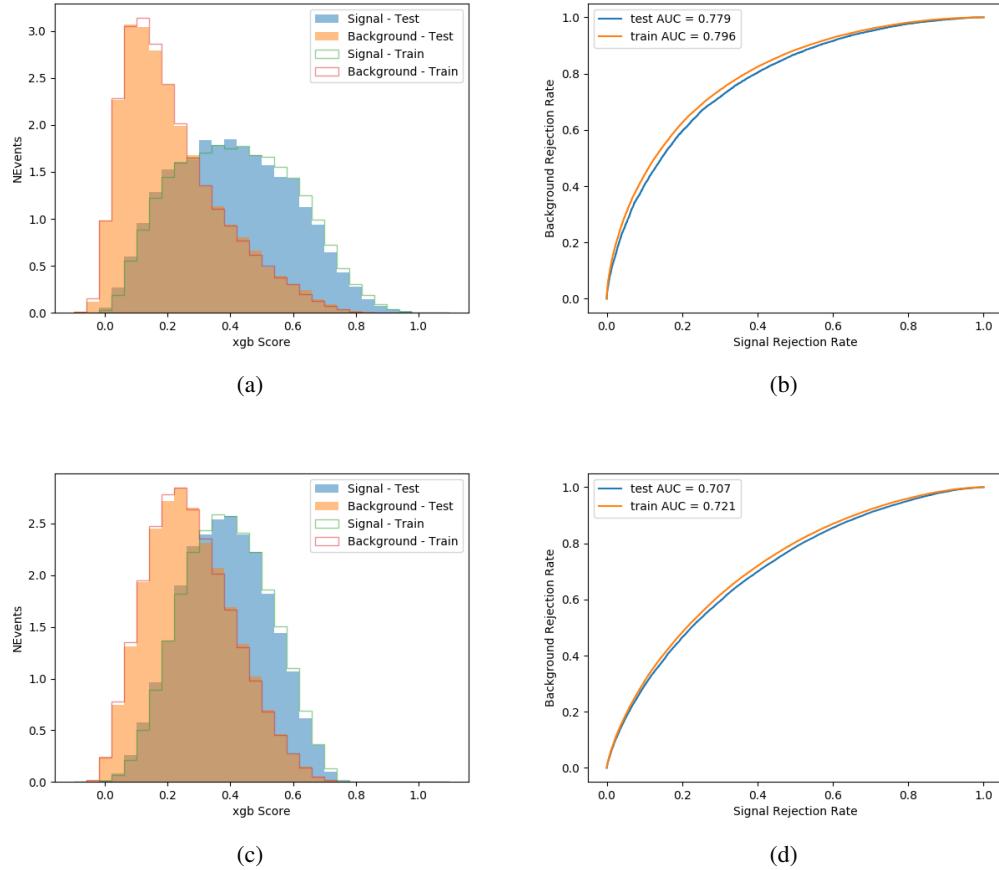


Figure 20.4:

1591 **20.3 Signal Region Definitions**

1592 Once pre-selection has been applied, channels are further refined based on the MVAs described
 1593 above. The output of the model described in section 18.5 is used to separate the three channel
 1594 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs
 1595 boson.

1596 For each event, depending on the channel as well as the predicted p_T of the Higgs derived

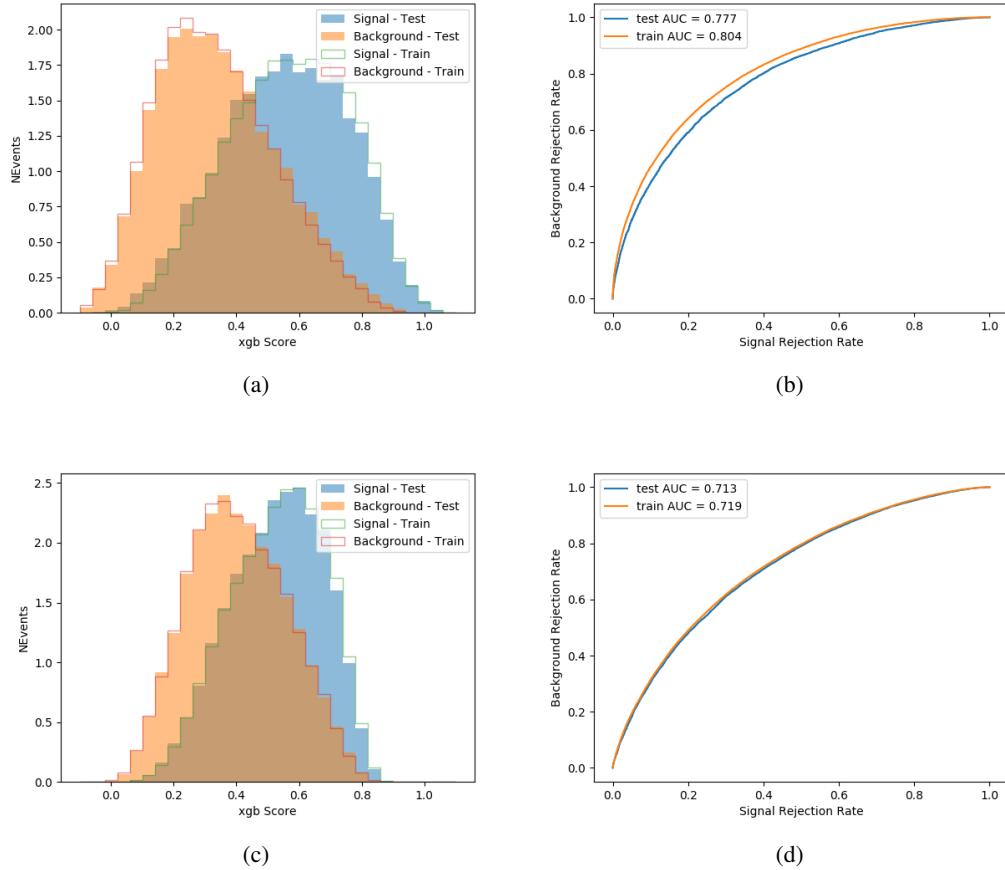


Figure 20.5:

from the algorithm described in section 18.4, a cut on the appropriate background rejection
 1597 algorithm is applied. The specific selection used, and the event yield in each channel after this
 1598 selection has been applied, is summarized below.
 1599

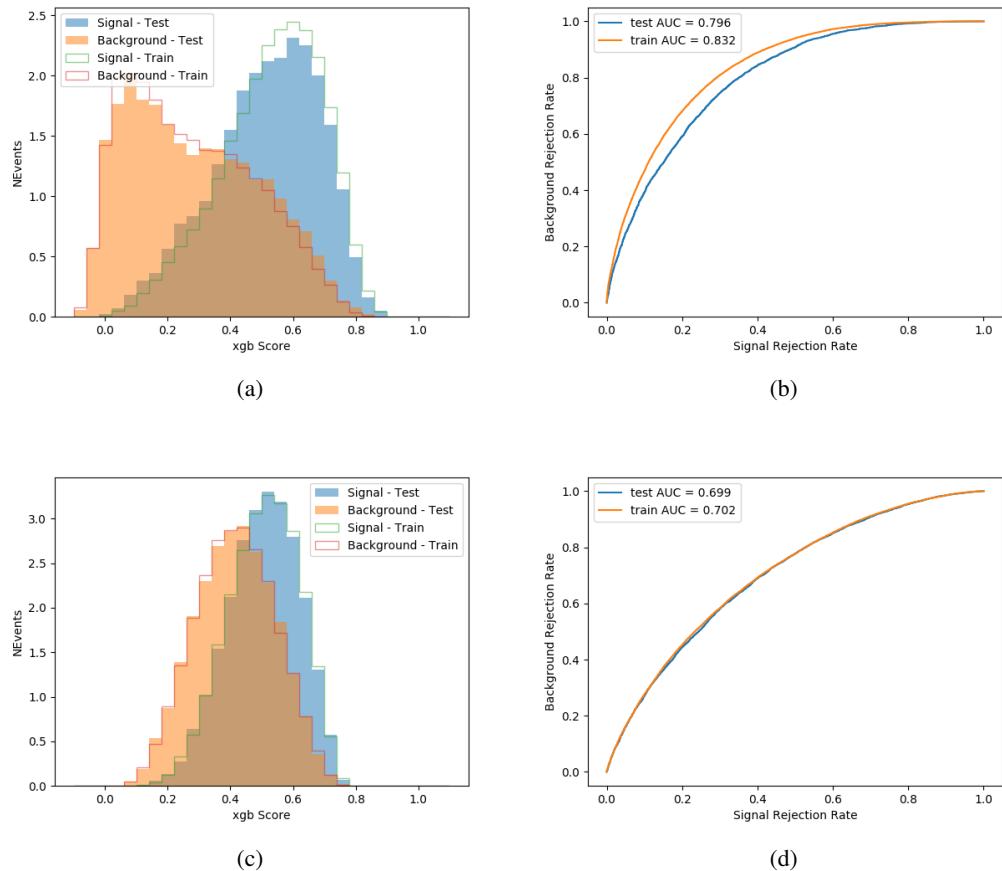


Figure 20.6:

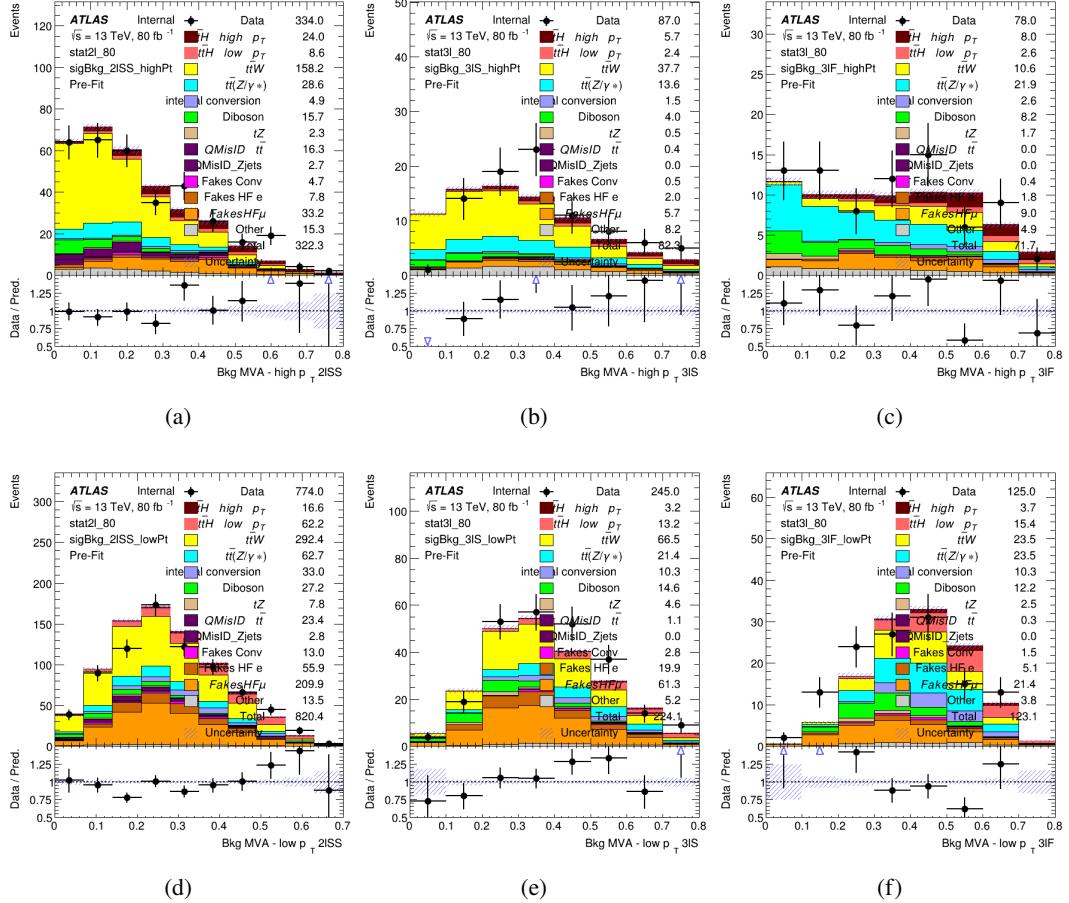


Figure 20.7: scores

1600 **20.3.1 2ISS**1601 **20.3.2 3l – Semi – leptonic**1602 **20.3.3 3l – Fully – leptonic**1603 **21 Systematic Uncertainties**

1604 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
 1605 25th March 2021 – 21:56

¹⁶⁰⁶ and background estimations. The numerical impact of each of these uncertainties is outlined in
¹⁶⁰⁷ section 22.

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

¹⁶⁰⁸ The uncertainty in the combined 2015+2016 integrated luminosity is derived from a
¹⁶⁰⁹ calibration of the luminosity scale using x-y beam-separation scans performed in August 2015
¹⁶¹⁰ and May 2016 [**lumi**].

¹⁶¹¹ The experimental uncertainties are related to the reconstruction and identification of light
¹⁶¹² leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The sources which contribute

1613 to the uncertainty in the jet energy scale [**jes**] are decomposed into uncorrelated components and
1614 treated as independent sources in the analysis.

1615 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
1616 [**btag_cal**] are also decomposed into uncorrelated components. The large number of components
1617 for b-tagging is due to the calibration of the distribution of the BDT discriminant.

1618 The systematic uncertainties associated with the signal and background processes are
1619 accounted for by varying the cross-section of each process within its uncertainty.

1620 **22 Results**

1621 Unblinded results are shown for the 80 fb^{-1} data set, as well as MC only projections of results
1622 using the full Run-2, 140 fb^{-1} dataset.

1623 **22.1 Results - 80 fb^{-1}**

1624 A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p_T spectrum
1625 in the three signal regions, 2lSS, 3lS, and 3lF, shown in figure 22.1. The $t\bar{t}H$ MC is split into
1626 high and low p_T , based on whether the truth p_T of the Higgs is above or below 150 GeV. The
1627 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.

1628 As the data collected from 2015-2017 has been unblinded for $t\bar{t}H$ – ML channels, rep-
 1629 resenting 80 fb^{-1} , those events are unblinded.

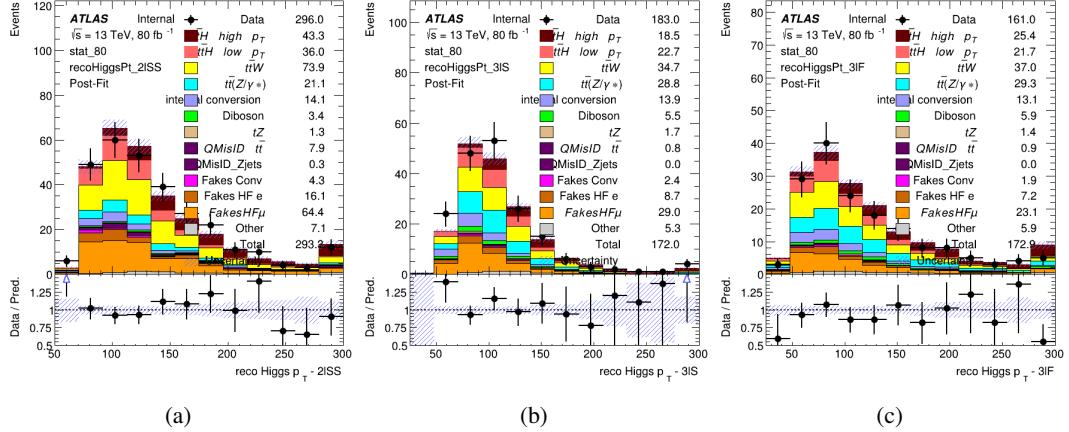


Figure 22.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 80 fb^{-1} of MC

1630 A post-fit summary of the fitted regions is shown in figure 22.2.

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 $\mu = \sigma_{\text{obs}} / \sigma_{\text{pred}}$

1632 **Need to add something about systematics here**

1633 The background composition of each of the fit regions is shown in figure 22.3.

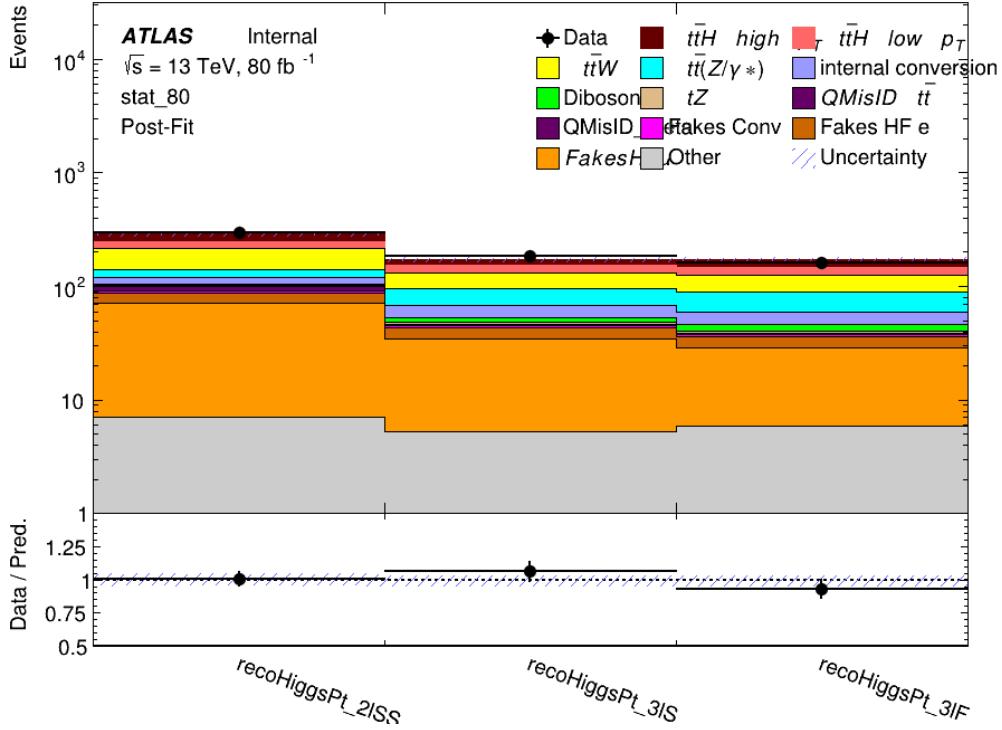


Figure 22.2: Post-fit summary of the yields in each signal region.

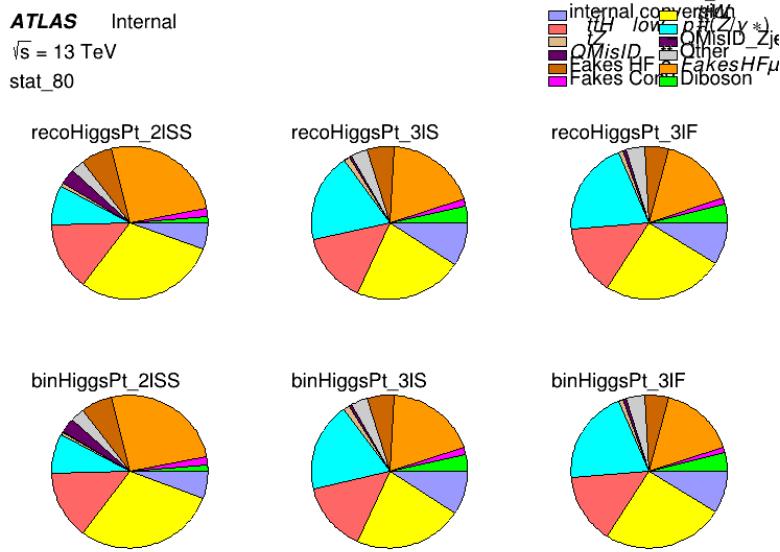


Figure 22.3: Background composition of the fit regions.

1634 22.2 Projected Results - 140 fb^{-1}

1635 As data collected in 2018 has not yet been unblinded for $t\bar{t}H - \text{ML}$ at the time of this note, data
 1636 from that year remains blinded. Instead, an Asimov fit is performed - with the MC prediction
 1637 being used both as the SM prediction as well as the data in the fit - in order to give expected
 1638 results.

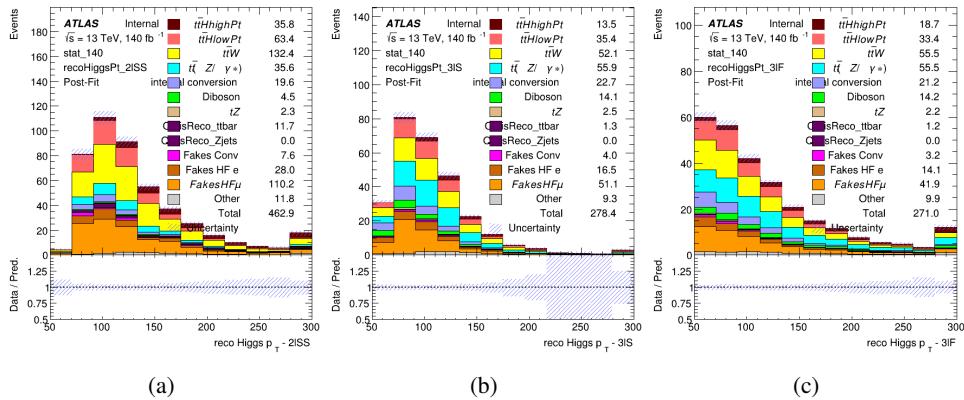


Figure 22.4: Blinded post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 140 fb^{-1} of data

1639 The μ values for high and low p_T Higgs are shown in 51.

$$\begin{aligned}\mu_{t\bar{t}Hhighp_T} &= 1.0^{+0.0}_{-0.0} (\text{stat})^{+0.0}_{-0.0} (\text{sys}) \\ \mu_{t\bar{t}Hlowp_T} &= 1.0^{+0.0}_{-0.0} (\text{stat})^{+0.0}_{-0.0} (\text{sys})\end{aligned}$$

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

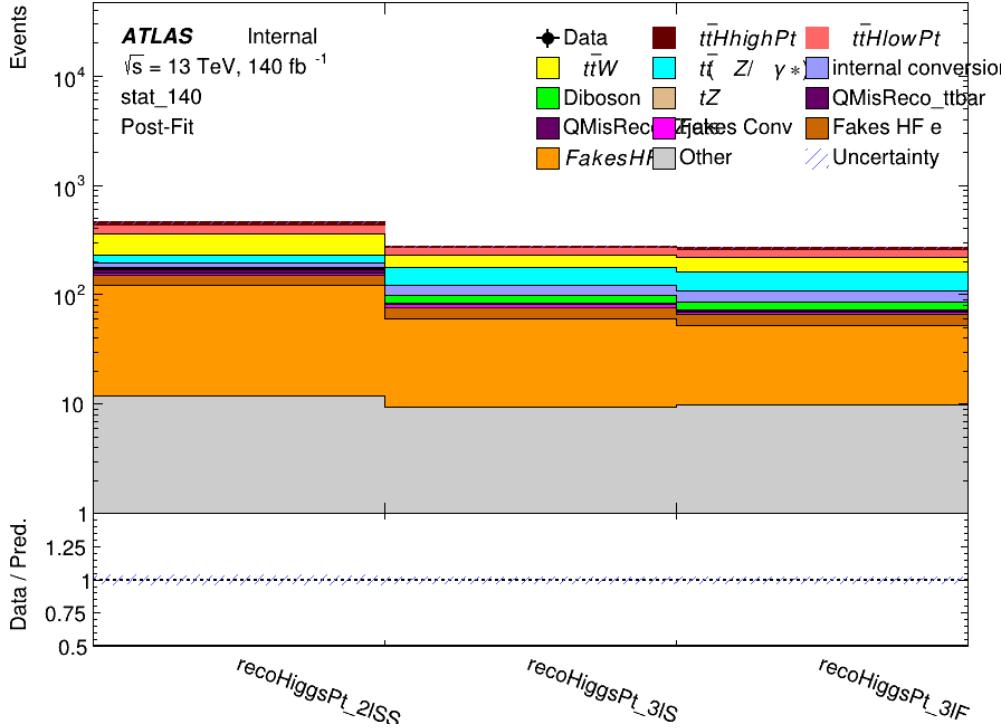


Figure 22.5: Post-fit summary of fit.

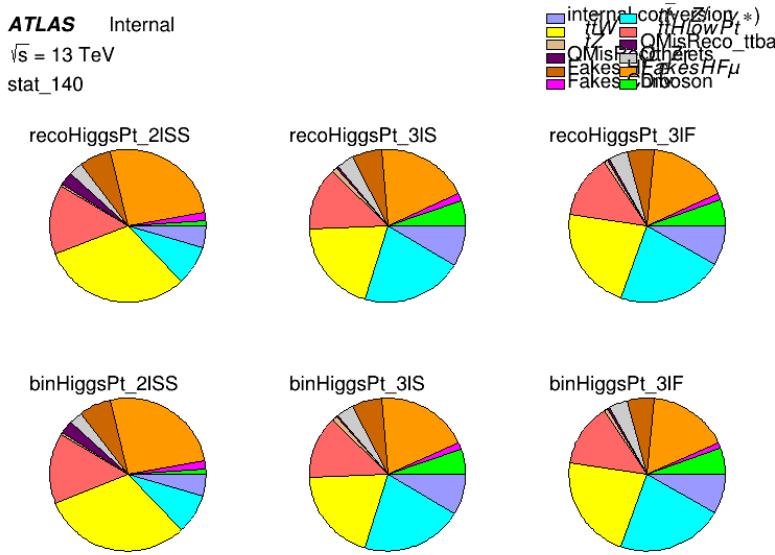


Figure 22.6: Background composition of the fit regions.

Part VI**Conclusion**

1640 As search for the effects of dimension-six operators on $t\bar{t}H$ production is performed. An effective
1641 field theory approached is used to parametrize the effects of high energy physics on the Higgs
1642 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,
1643 and the limits on dimension-six operators are limited to X.
1644

1646 List of contributions

1647

¹⁶⁴⁸ **Appendices**

¹⁶⁴⁹ **A Machine Learning Models**

¹⁶⁵⁰ The following section provides details of the various MVAs as well as a few studies performed
¹⁶⁵¹ in support of this analysis, exploring alternate decisions and strategies.

¹⁶⁵² **A.1 Higgs Reconstruction Models**

¹⁶⁵³ **A.1.1 b-jet Identification Features - 2lSS**

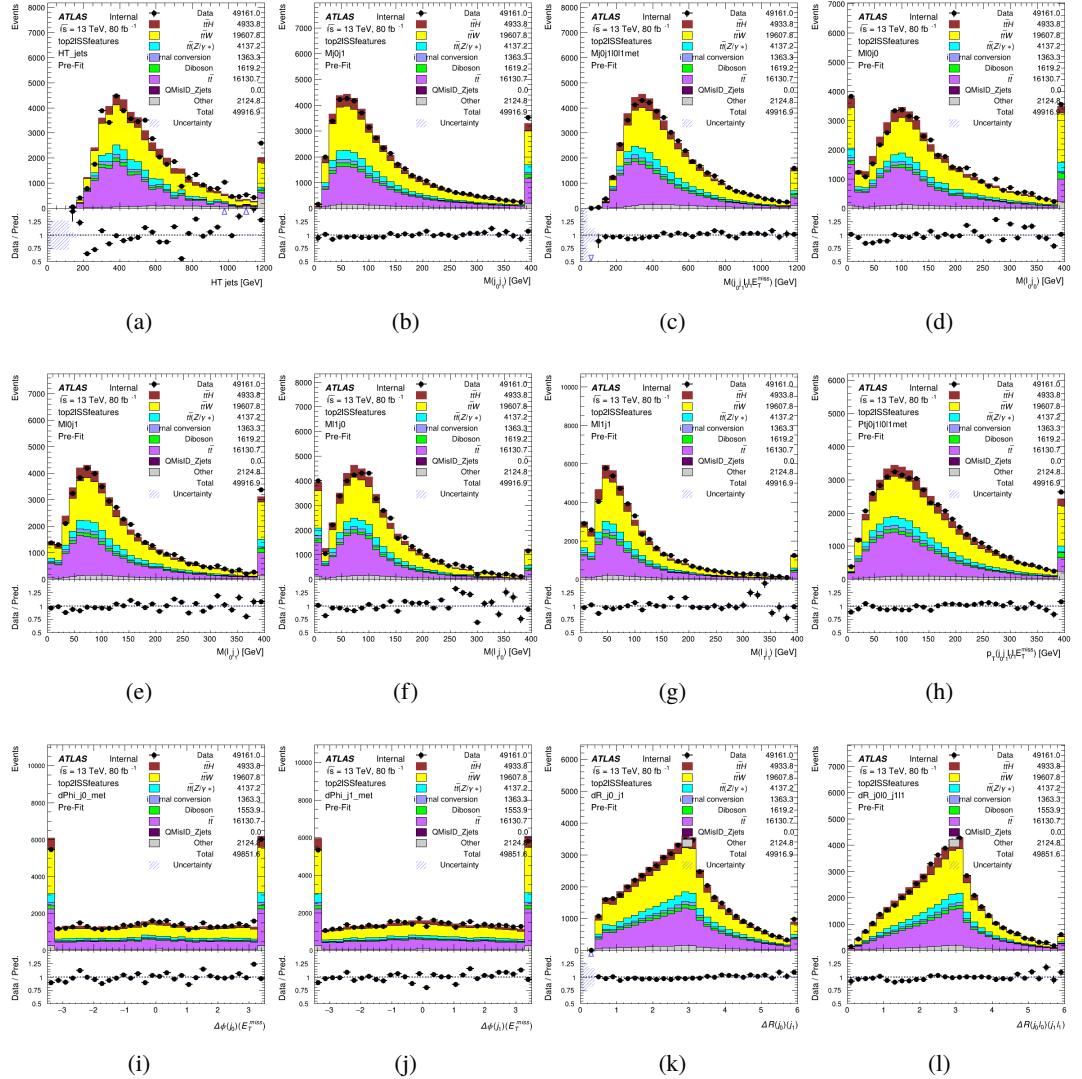


Figure A.1: Input features for top2lSS

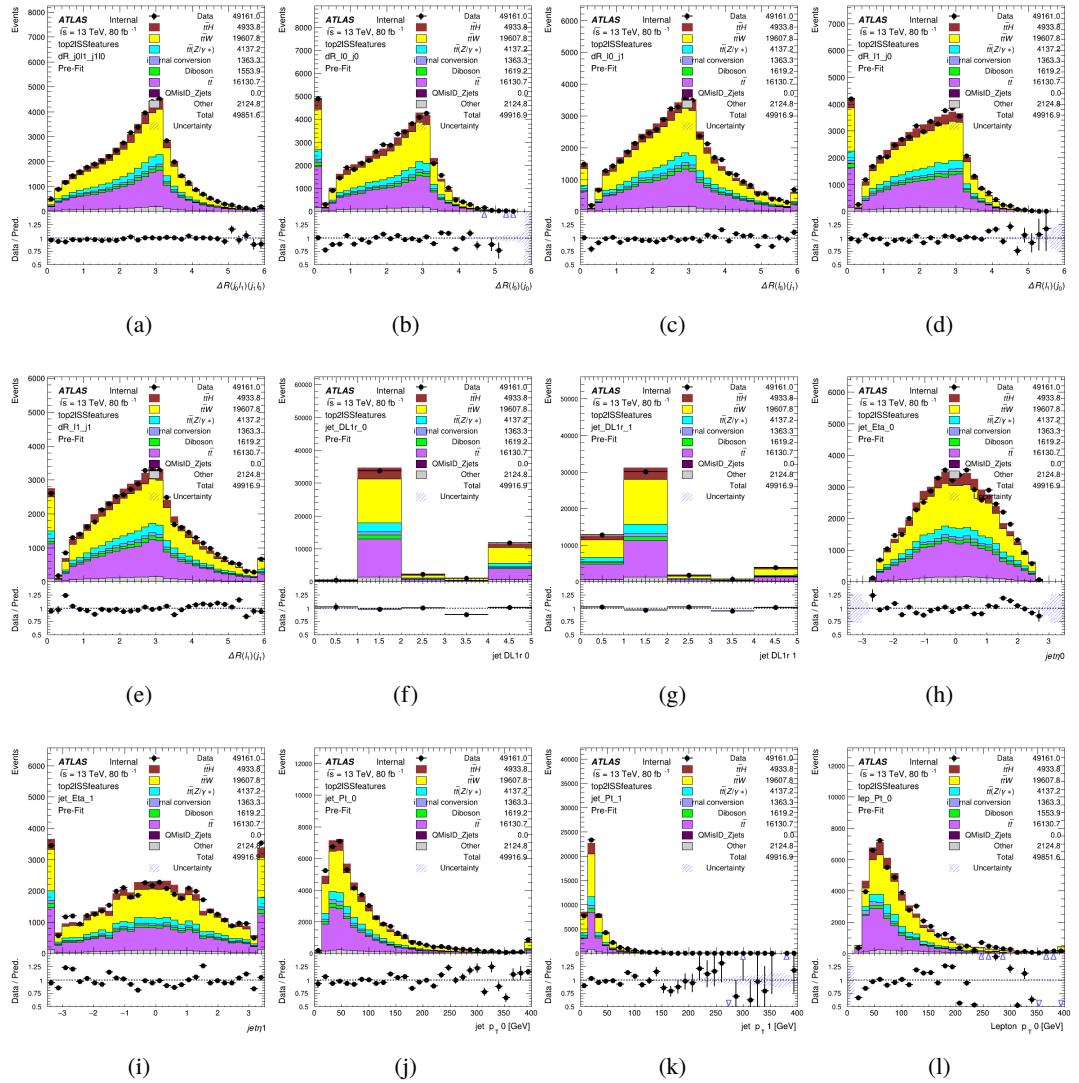


Figure A.2: Input features for top2ISS

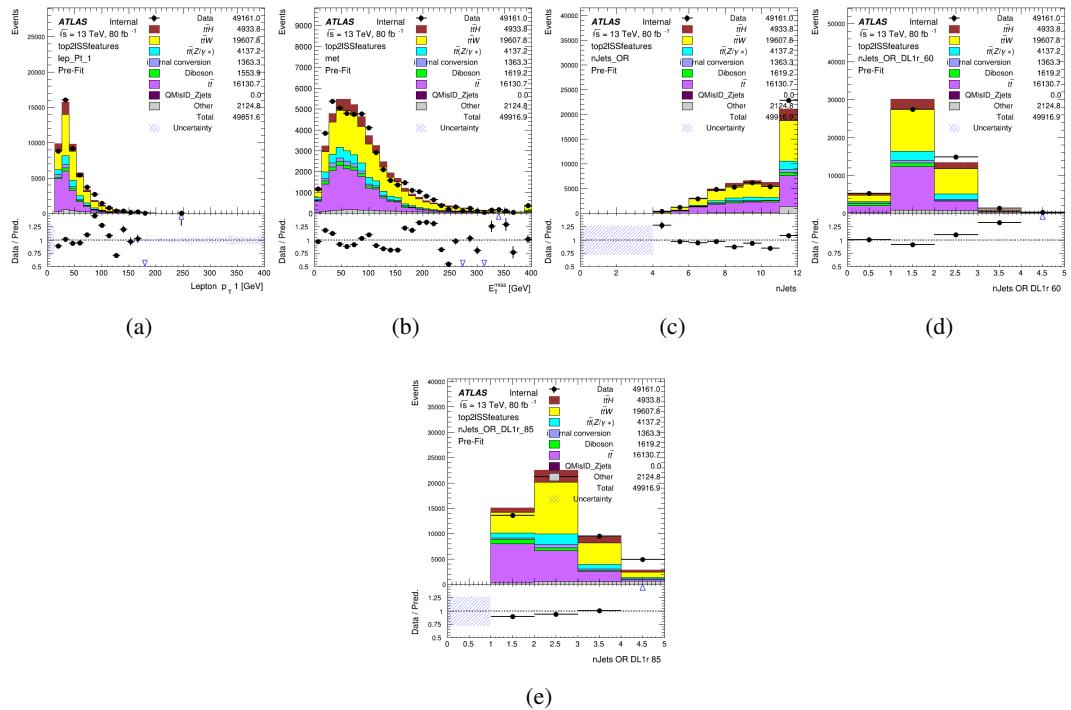


Figure A.3: Input features for top2lSS

1654 **A.1.2 b-jet Identification Features - 3l**

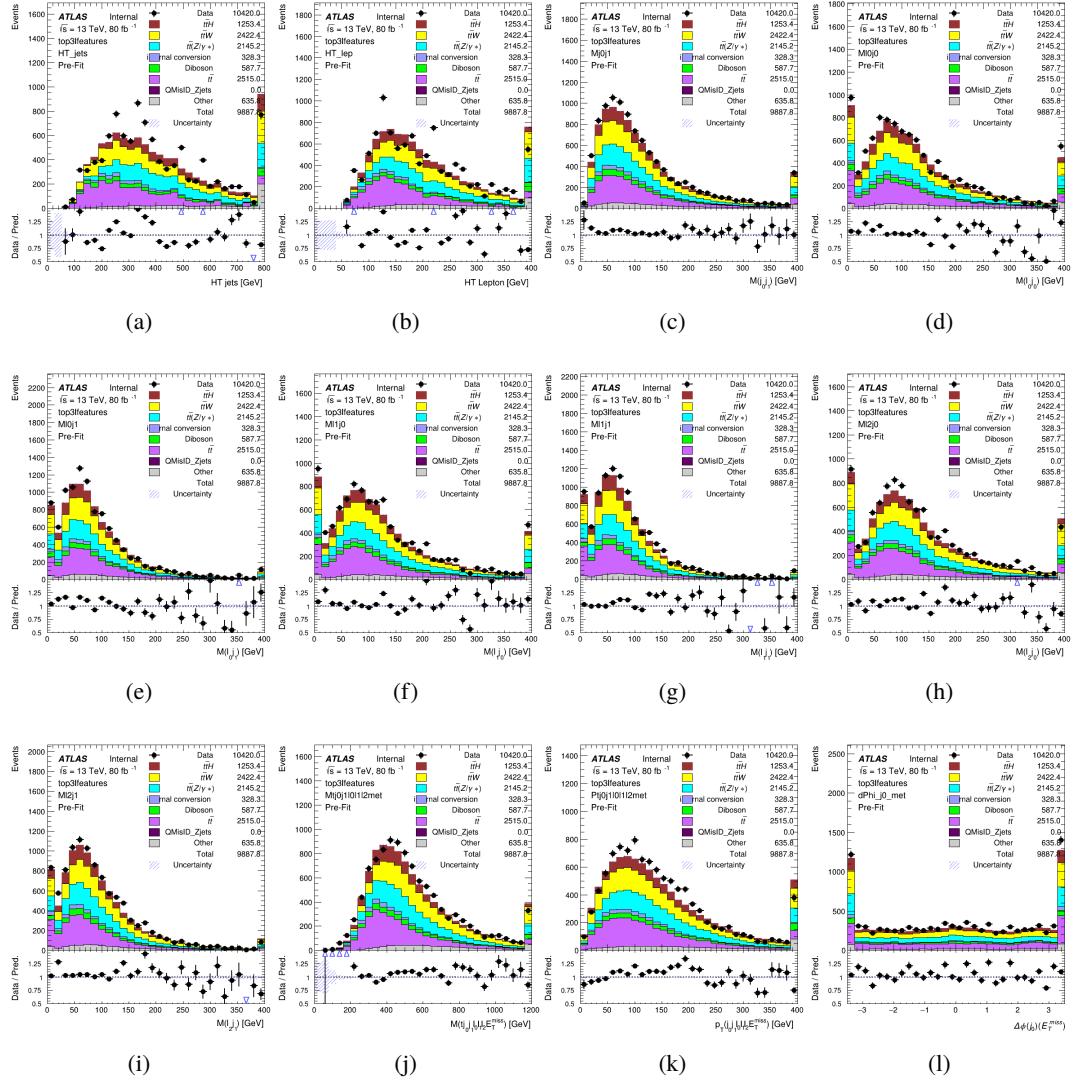


Figure A.4: Input features for top31

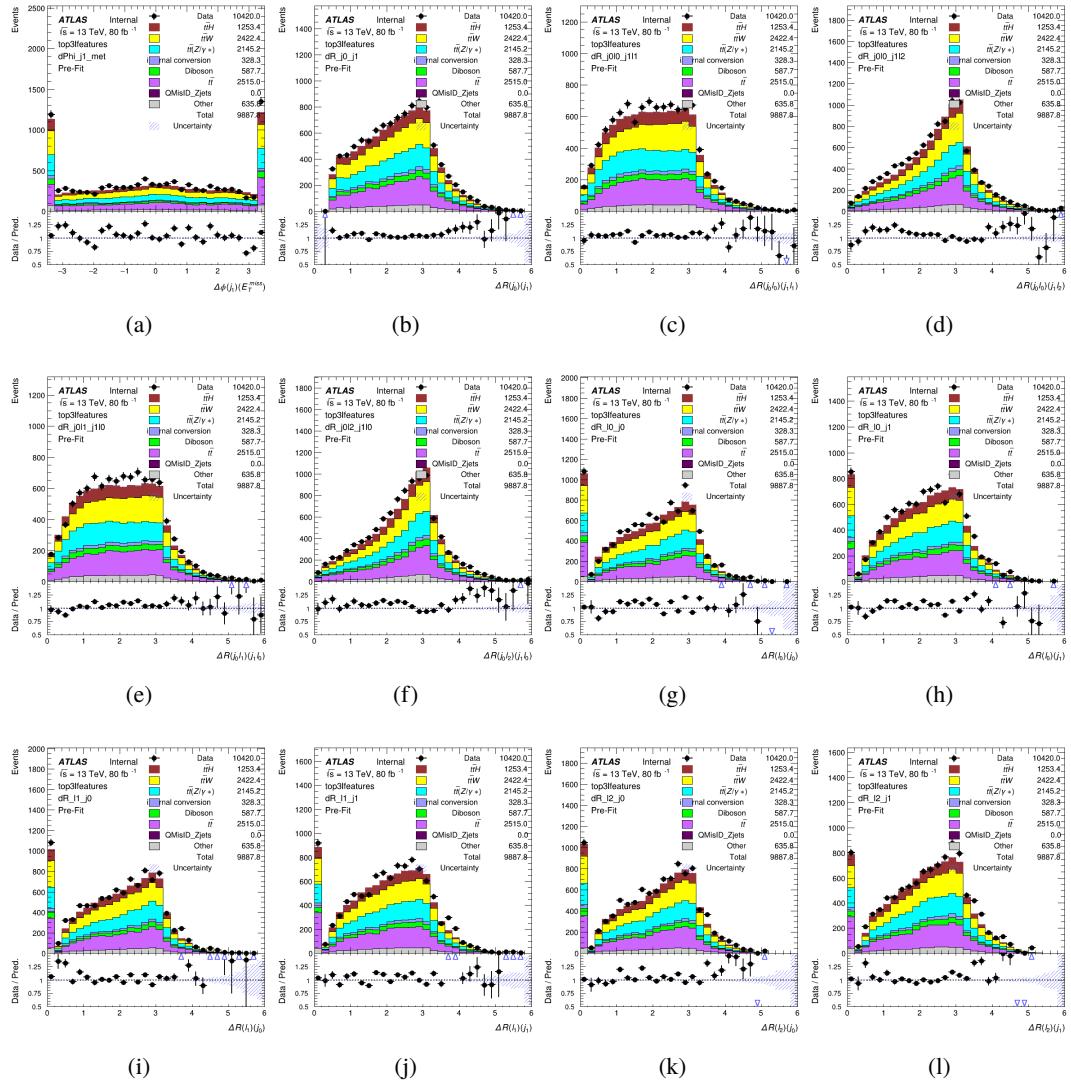


Figure A.5: Input features for top31

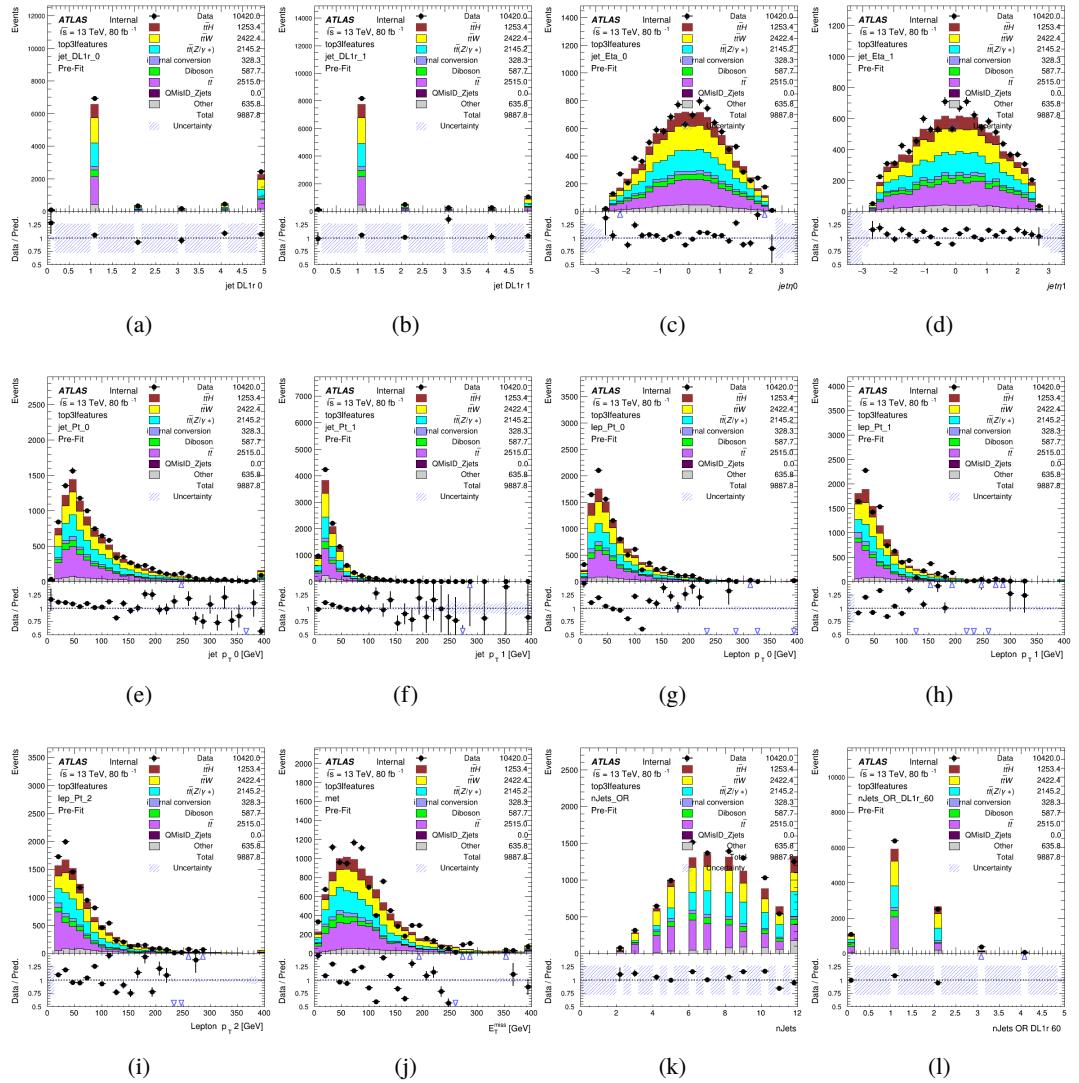
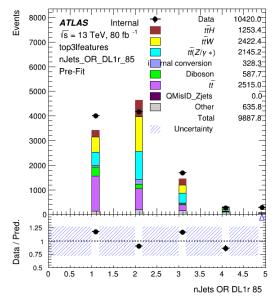


Figure A.6: Input features for top31



(a)

Figure A.7: Input features for top3l

1655 **A.1.3 Higgs Reconstruction Features - 2lSS**

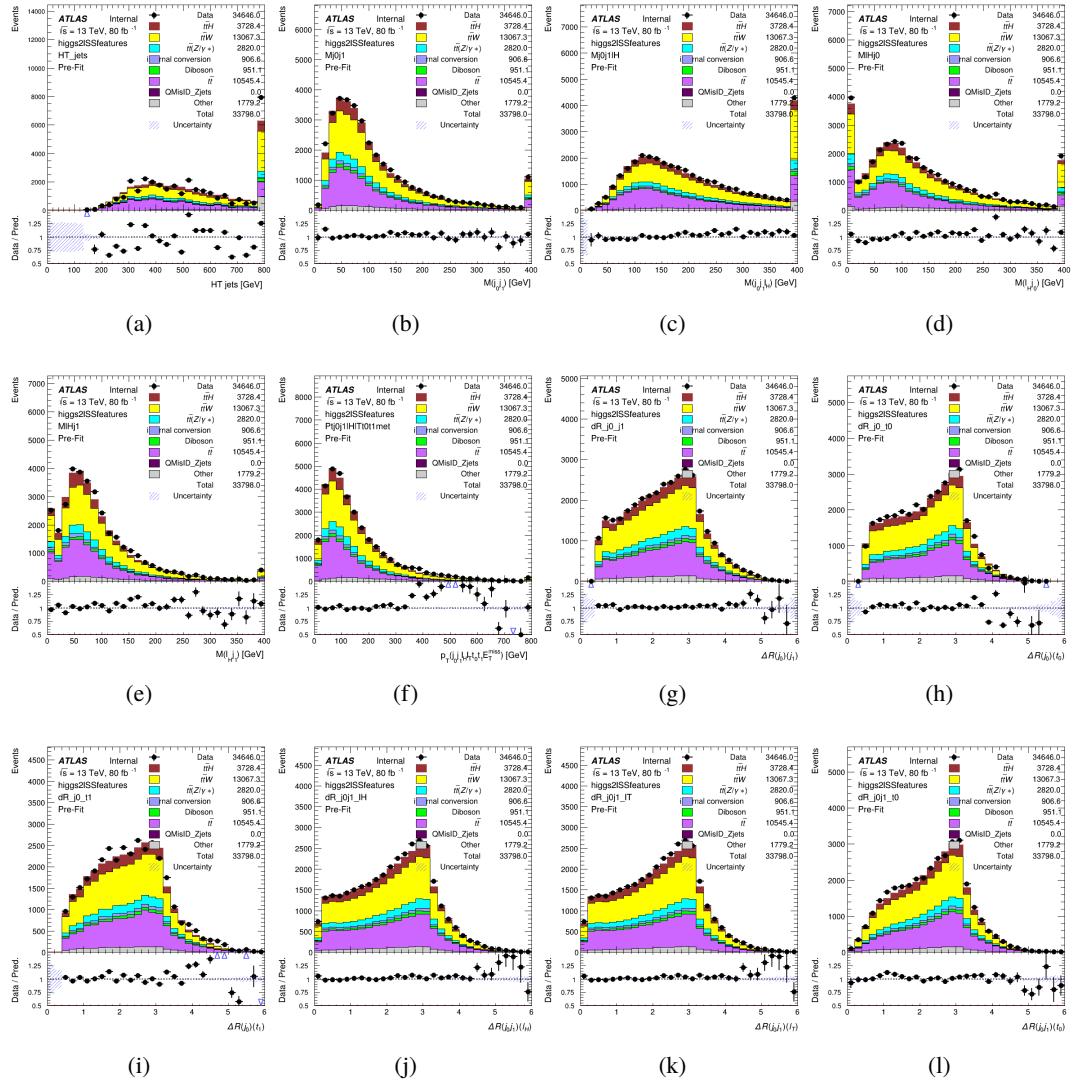


Figure A.8: Input features for higgs2lSS

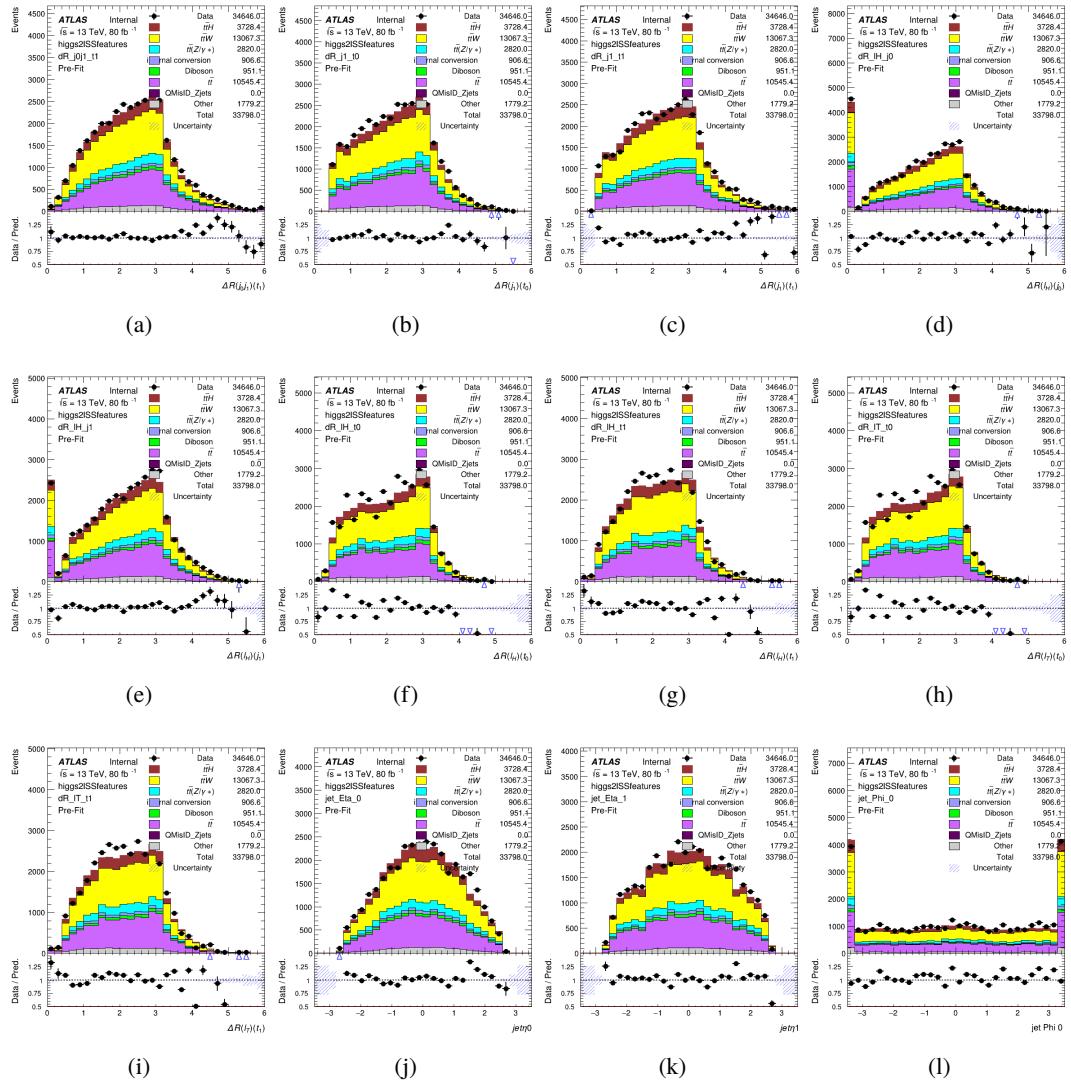


Figure A.9: Input features for higgs2lSS

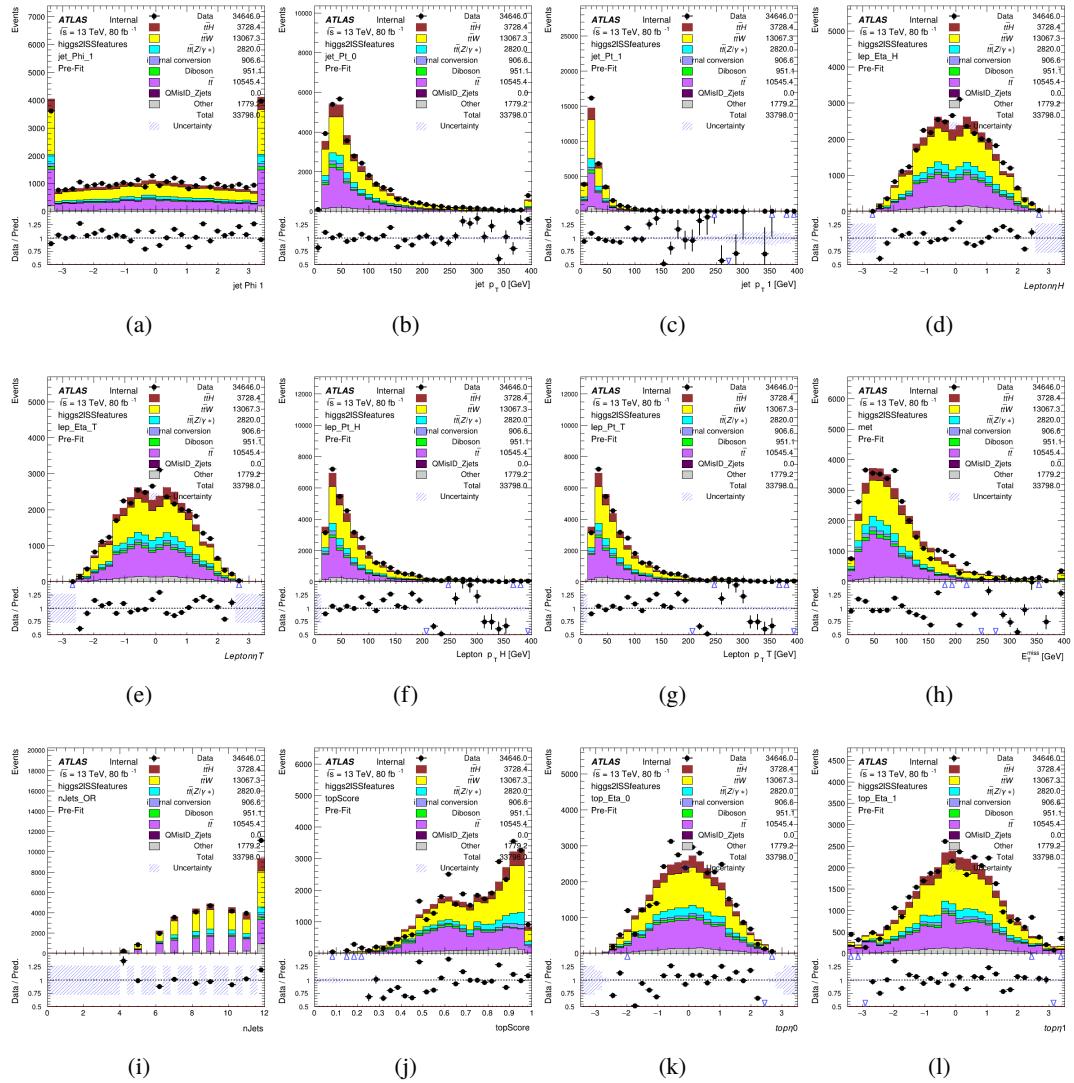


Figure A.10: Input features for higgs2ISS

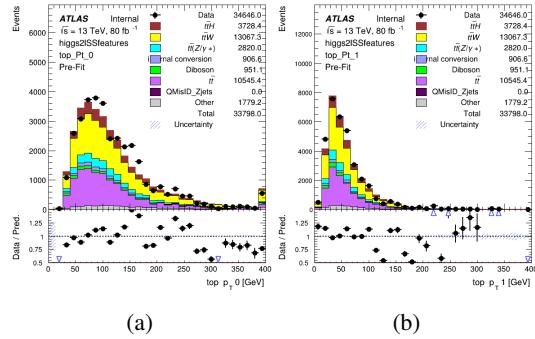


Figure A.11: Input features for higgs2ISS

1656 **A.1.4 Higgs Reconstruction Features - 3lS**

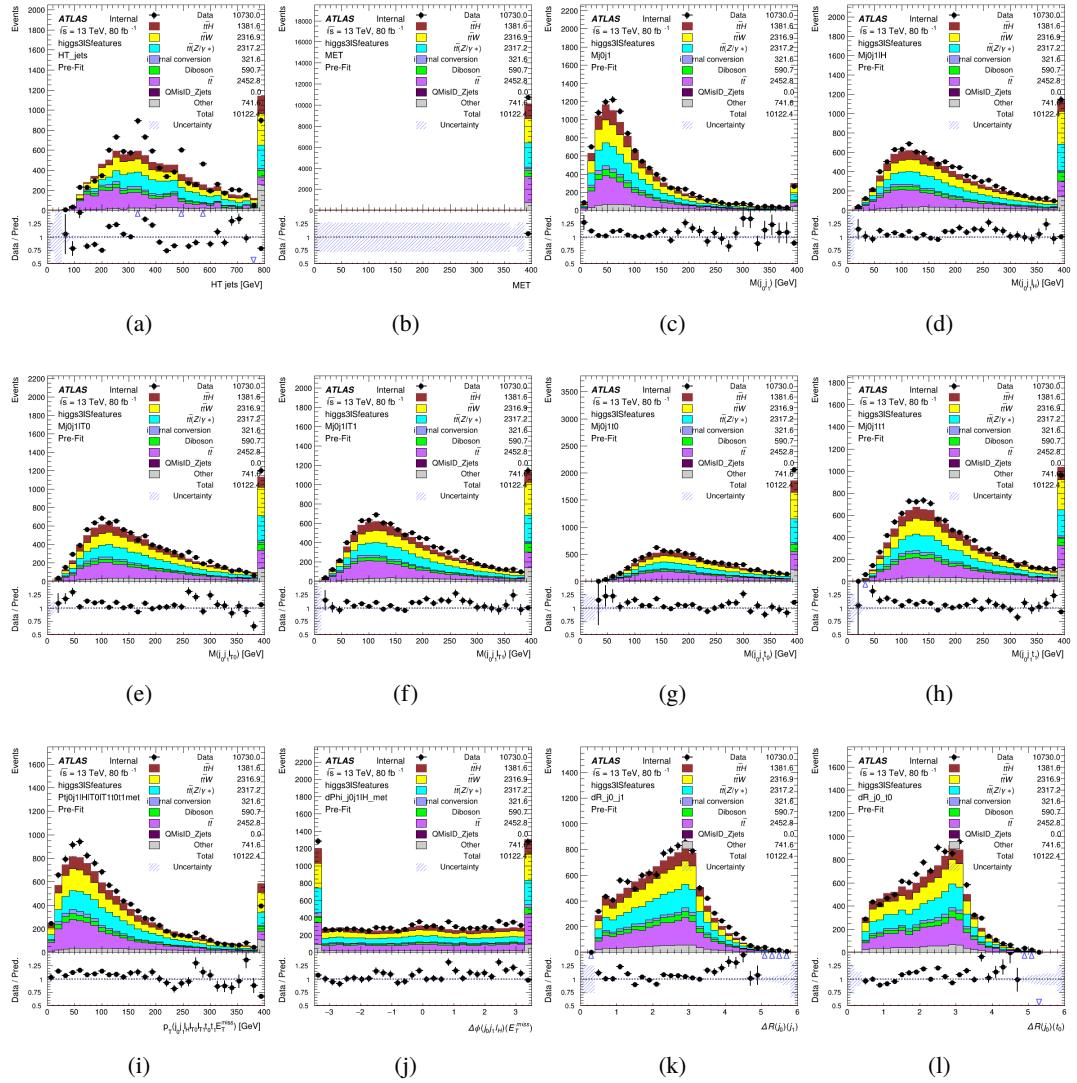


Figure A.12: Input features for higgs3IS

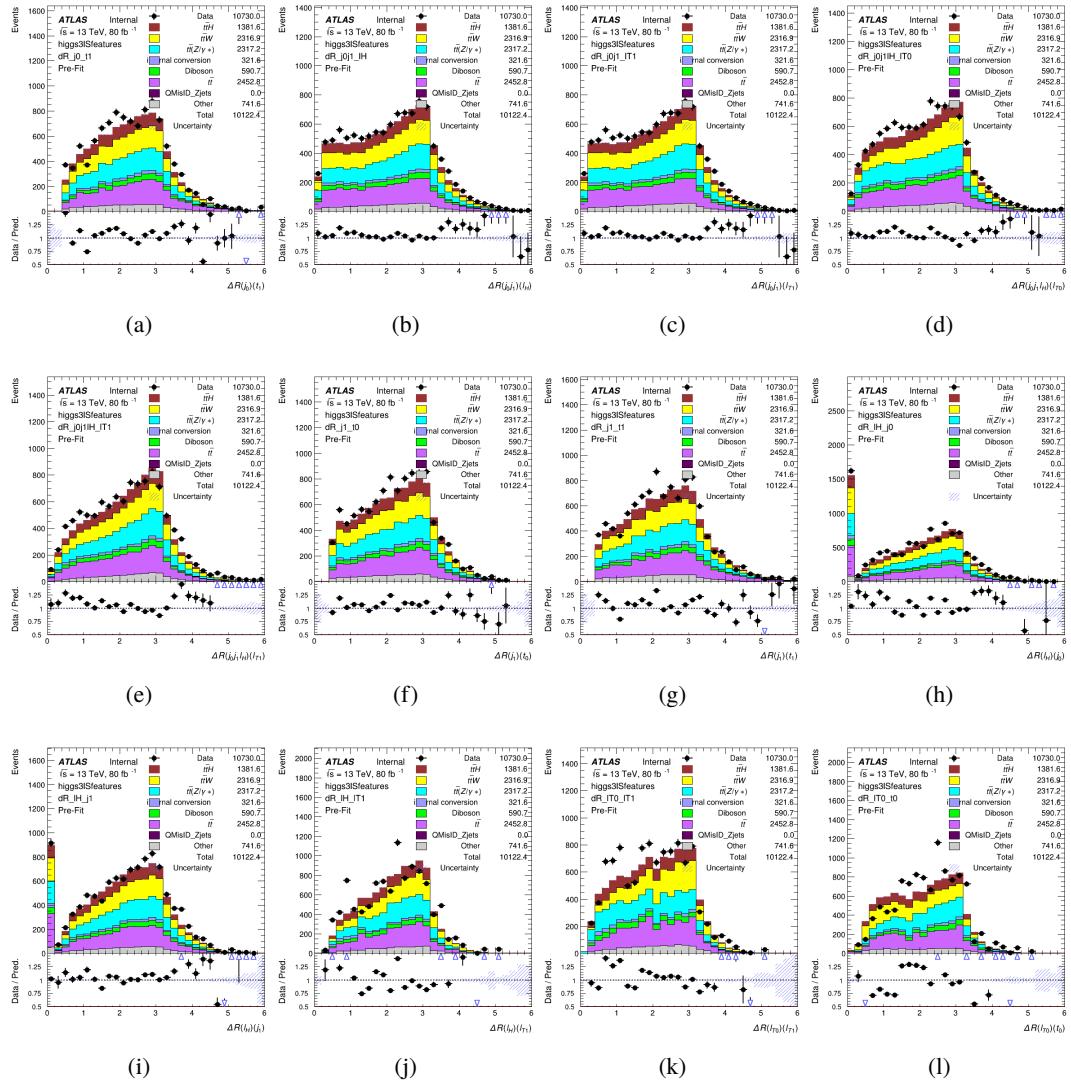


Figure A.13: Input features for higgs31S

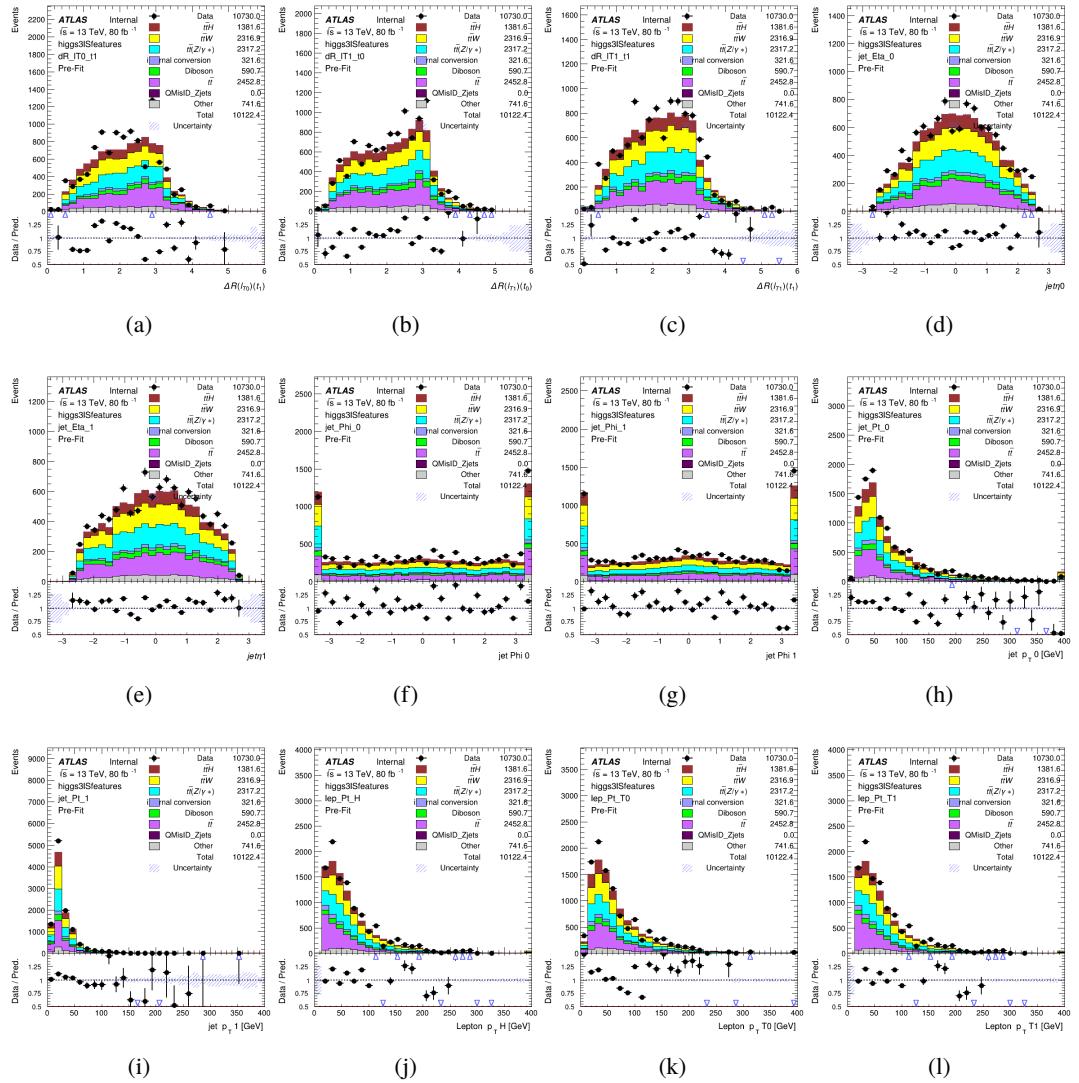


Figure A.14: Input features for higgs3lS

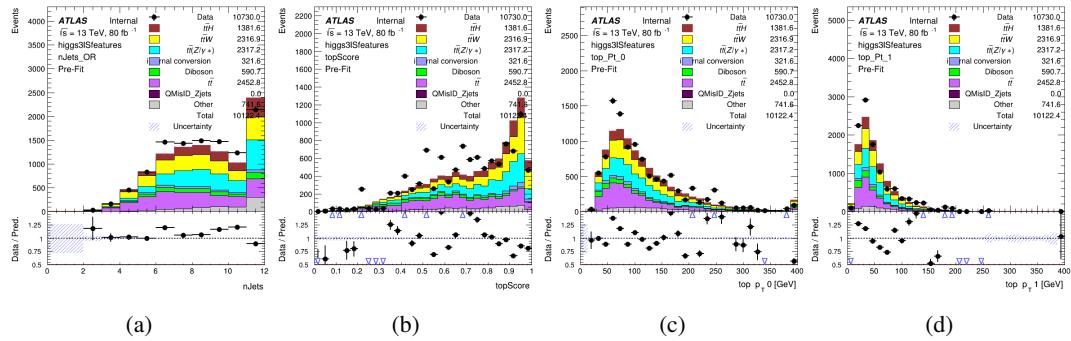


Figure A.15: Input features for higgs3lS

1657 **A.1.5 Higgs Reconstruction Features - 3lF**

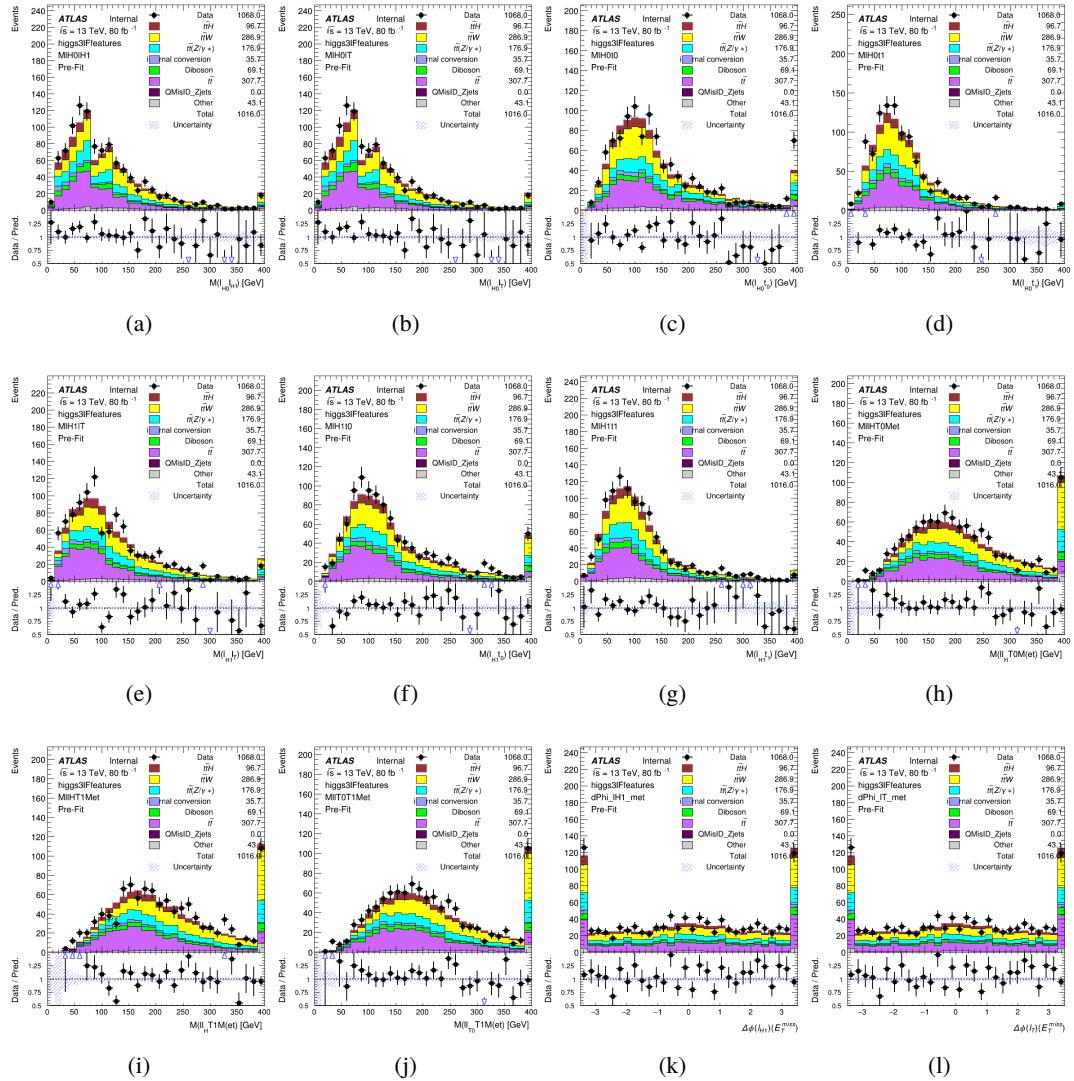


Figure A.16: Input features for higgs3IF

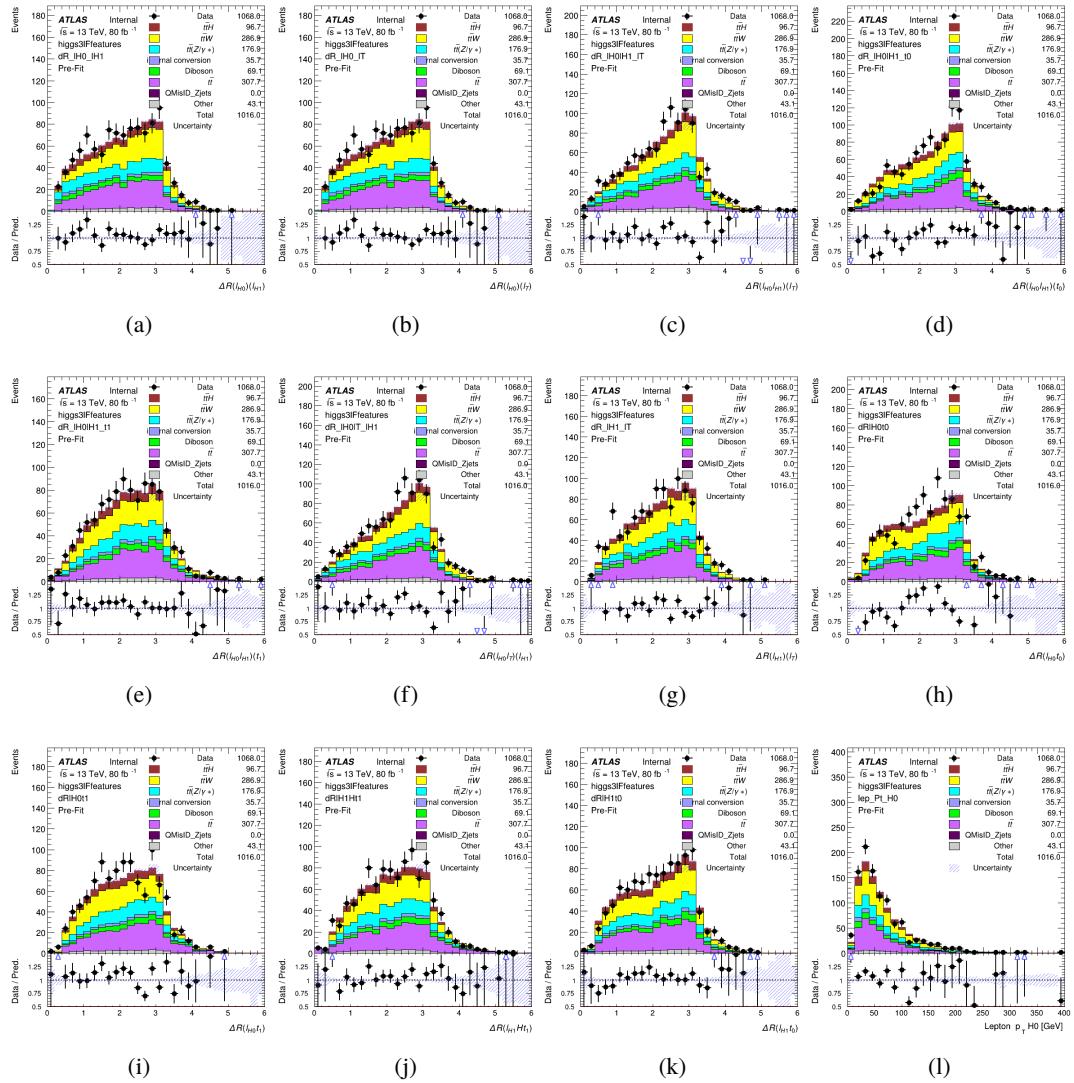


Figure A.17: Input features for higgs3lF

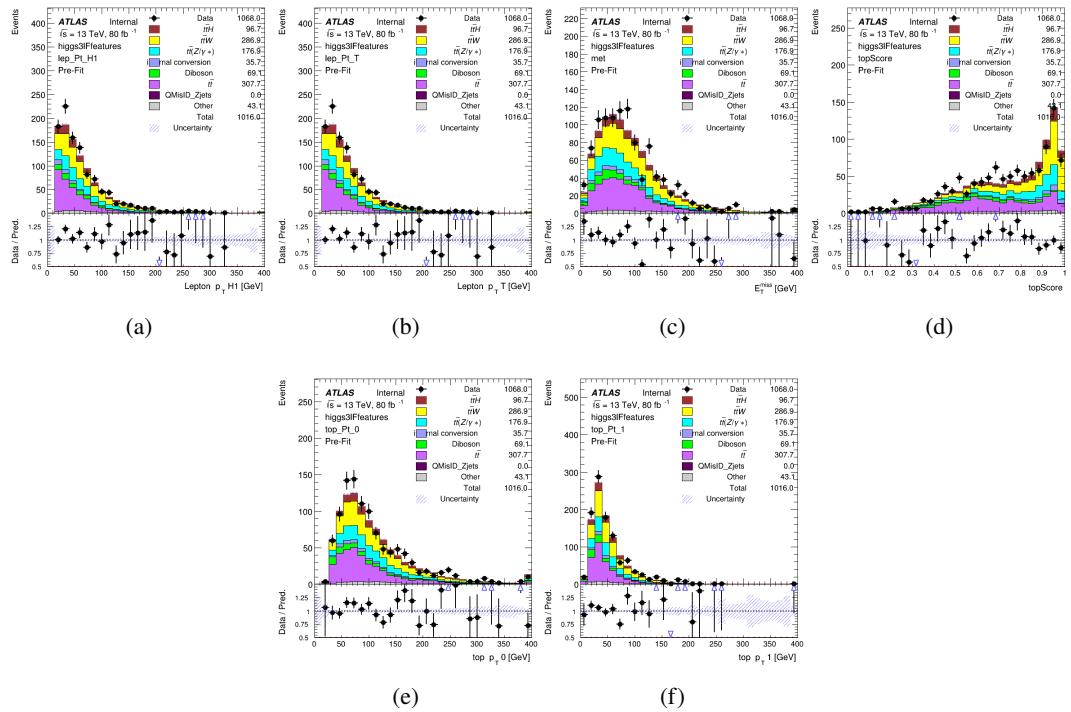


Figure A.18: Input features for higgs3IF

¹⁶⁵⁸ **A.2 Background Rejection MVAs**

¹⁶⁵⁹ **A.2.1 Background Rejection MVA Features - 2lSS**

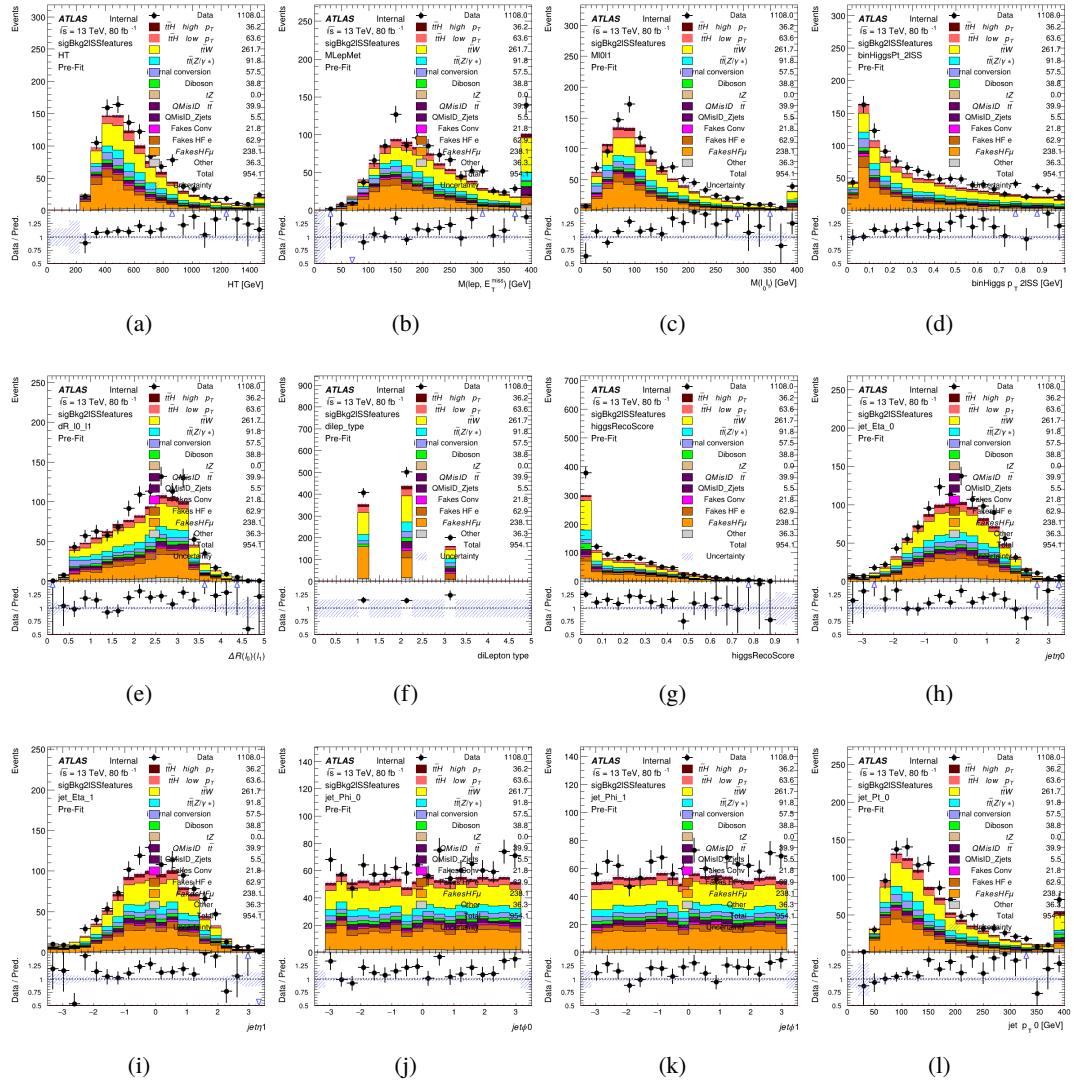


Figure A.19: Input features for sigBkg2lSS

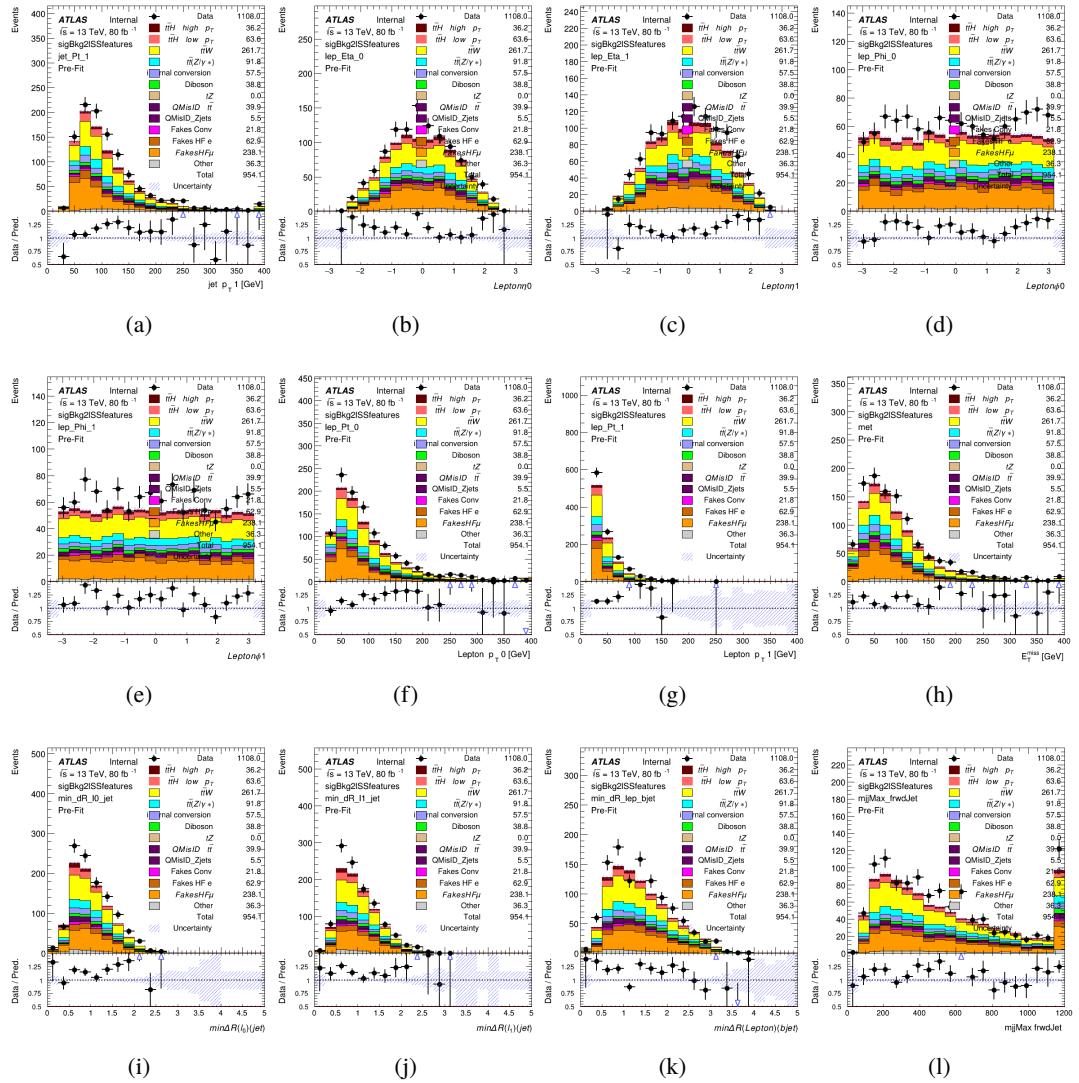


Figure A.20: Input features for sigBkg2lSS

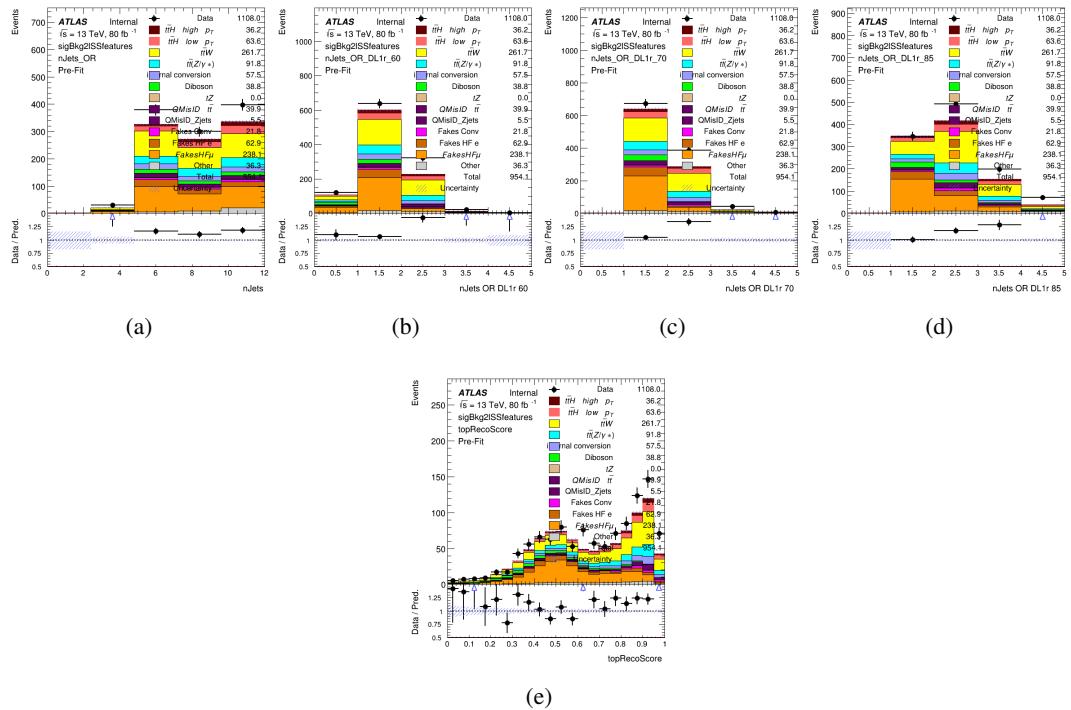


Figure A.21: Input features for sigBkg2ISS

1660 **A.2.2 Background Rejection MVA Features - 3l**

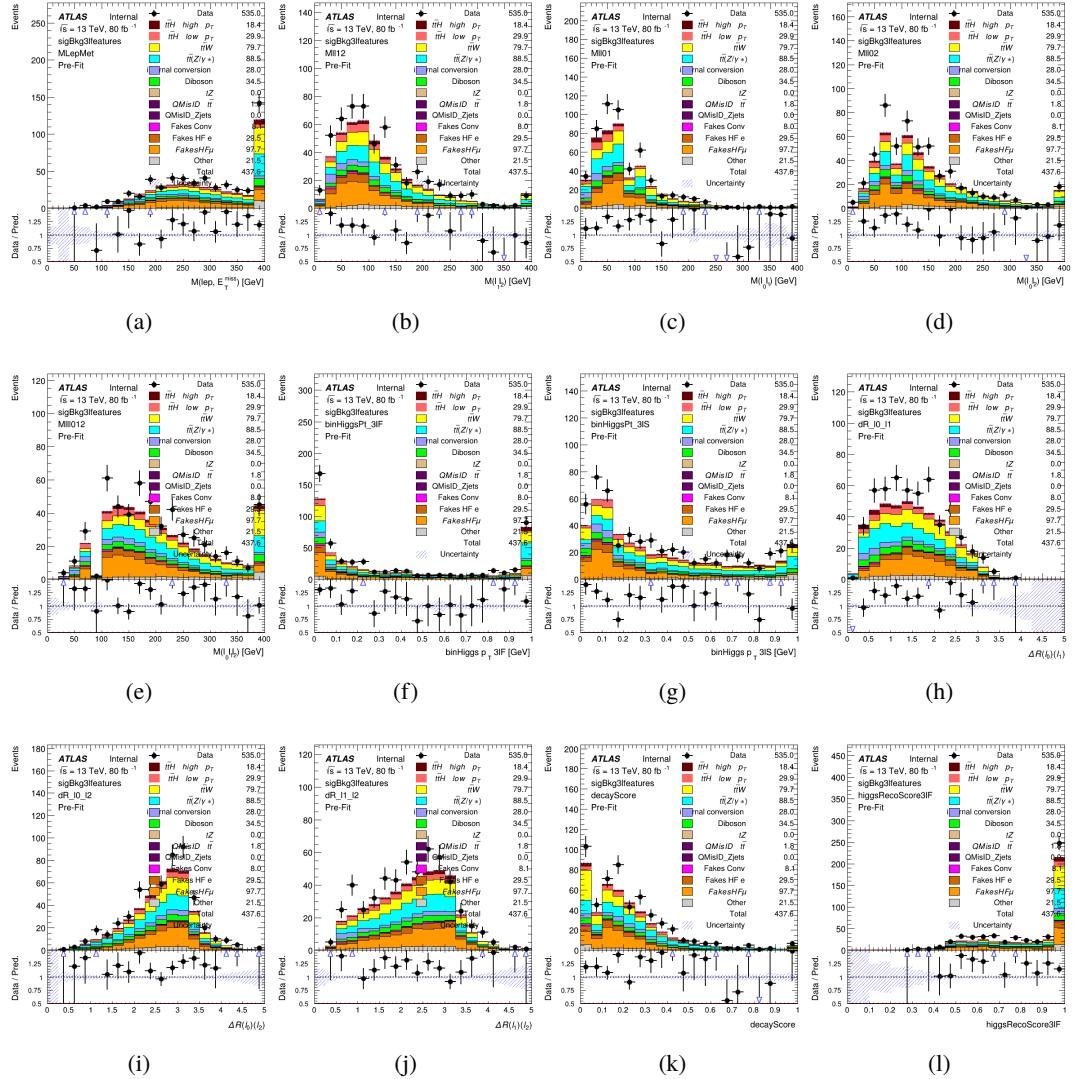


Figure A.22: Input features for sigBkg3l

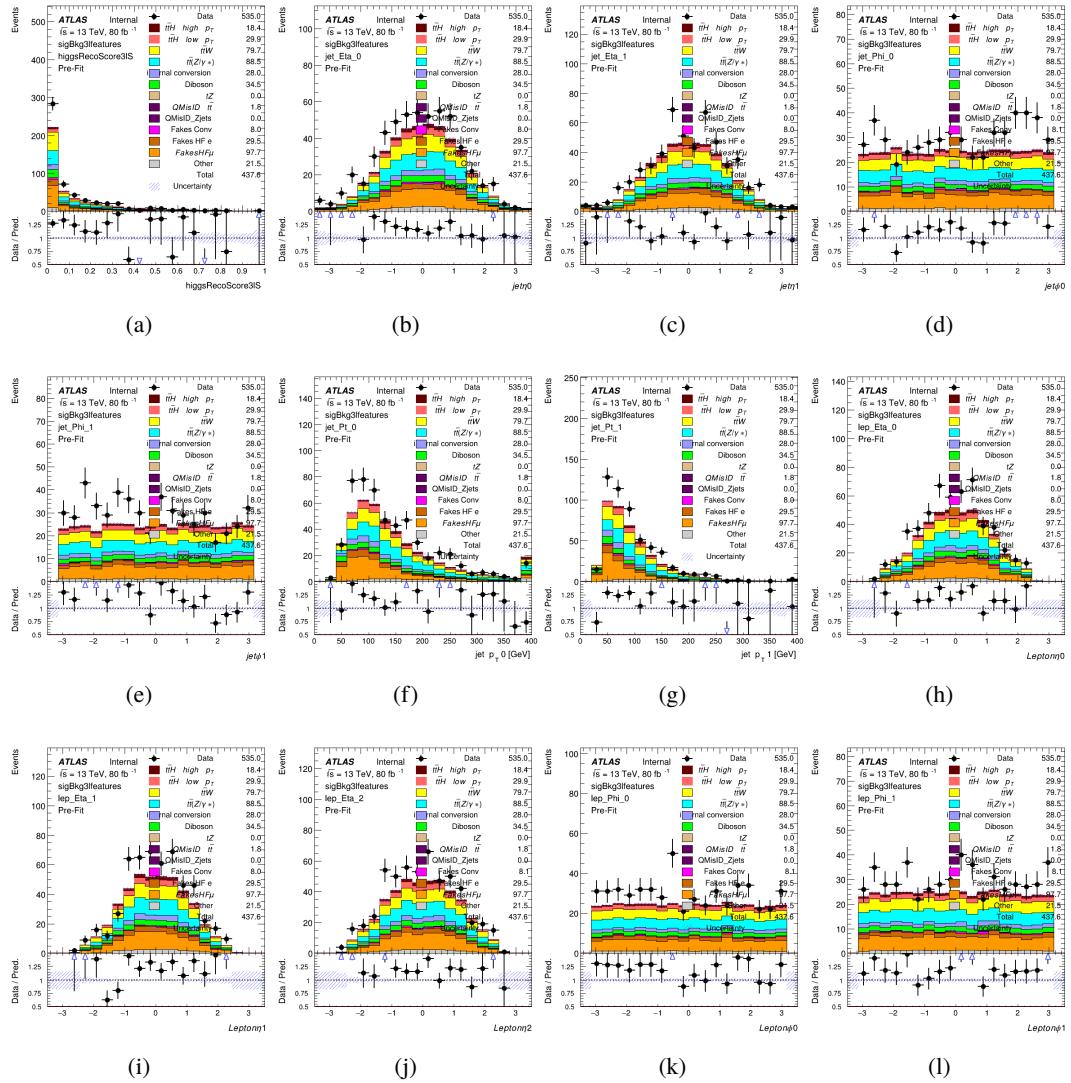


Figure A.23: Input features for sigBkg3l

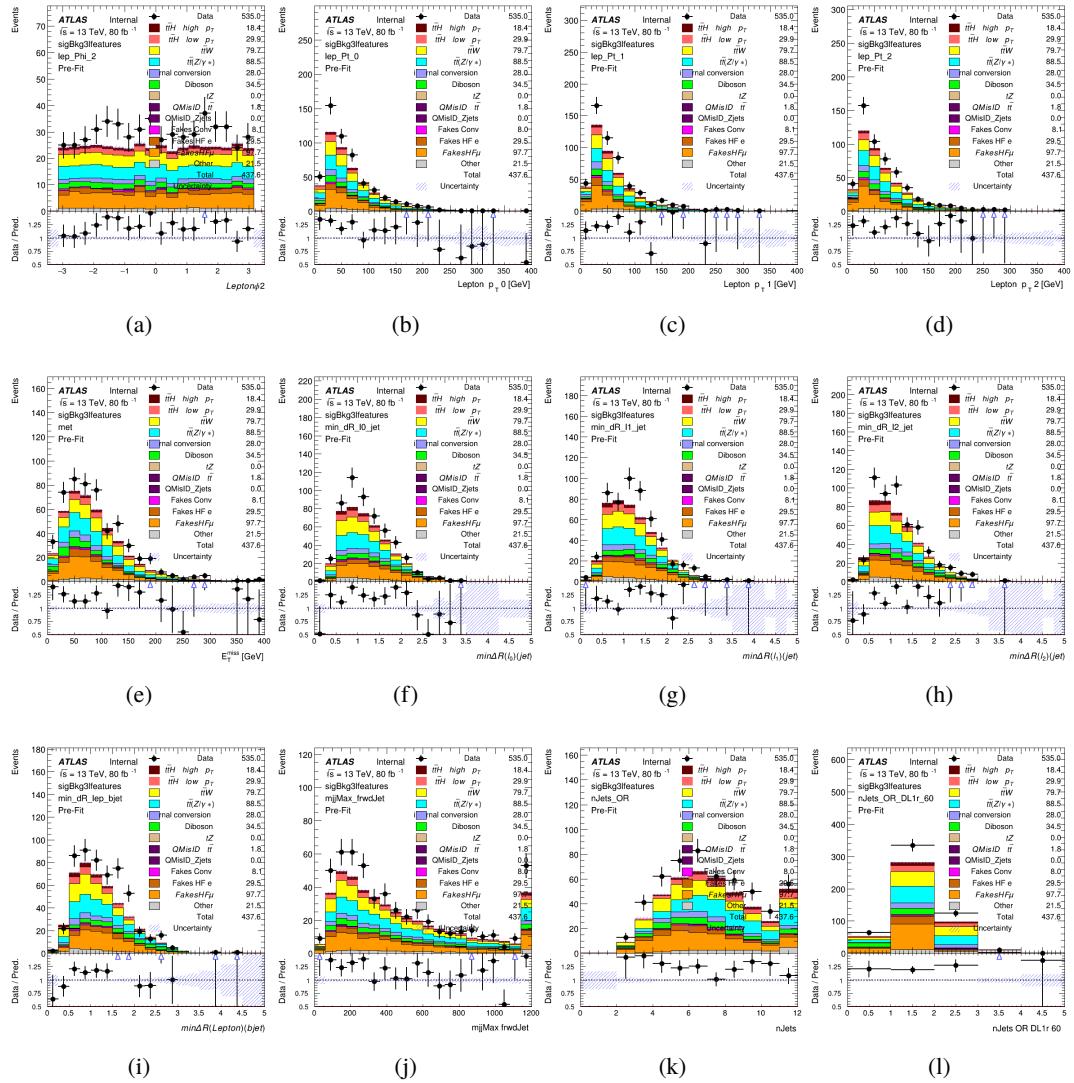


Figure A.24: Input features for sigBkg3l

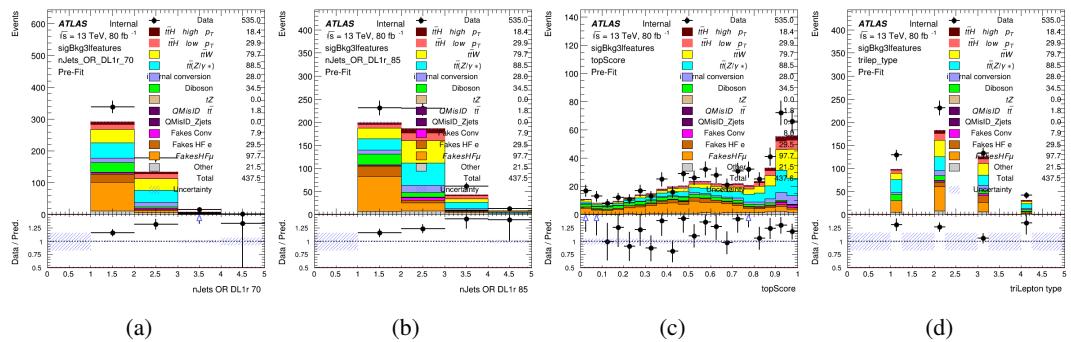


Figure A.25: Input features for sigBkg3l

1661 **A.3 Alternate b-jet Identification Algorithm**

1662 The nominal analysis reconstructs the b-jets by considering different combinations of jets, and
 1663 asking a neural network to determine whether each combination consists of b-jets from top quark
 1664 decays. An alternate approach would be to give the neural network about all of the jets in an event
 1665 at once, and train it to select which two are most likely to be the b-jets from top decay. It was
 1666 hypothesized that this could perform better than considering each combination independently, as
 1667 the neural network could consider the event as a whole. While this is not found to be the case,
 1668 these studies are documented here as a point of interest and comparison.

1669 For these studies, the kinematics of the 10 highest p_T jets in each event are used for
 1670 training. This includes the vast majority of truth b-jets. Specifically the p_T , η , ϕ , E , and DL1r
 1671 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.
 1672 The p_T , η , ϕ , and E of the leptons and E_T^{miss} are included as well. Categorical cross entropy is
 1673 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%

1674 **A.4 Binary Classification of the Higgs p_T**

1675 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer
 1676 resolution. This means separating high and low p_T events is sufficient for this analysis. As

1677 such, rather than attempting the reconstruct the full Higgs p_T spectrum, a binary classification
 1678 approach is explored.

1679 A model is built to determine whether $t\bar{t}H$ events include a high p_T (>150 GeV) or low
 1680 p_T (<150 GeV) Higgs Boson. While this is now a classification model, it uses the same input
 1681 features described in section 18.4. Binary crossentropy is used as the loss function.

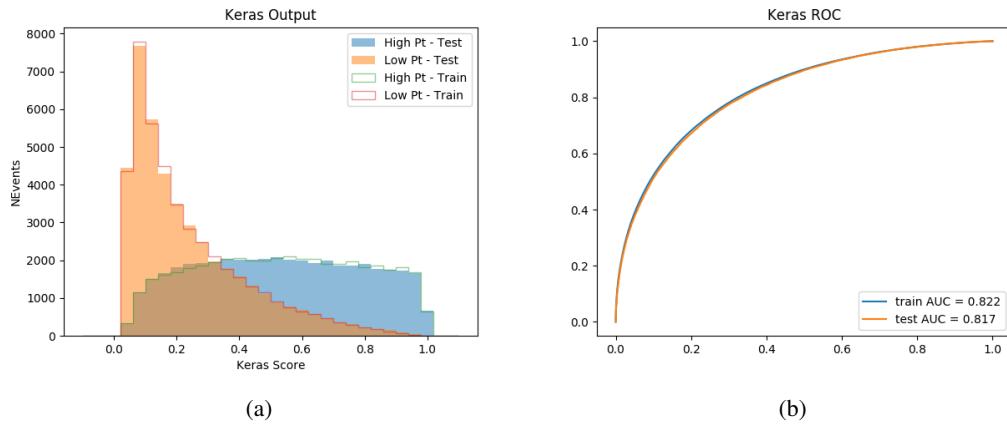


Figure A.26:

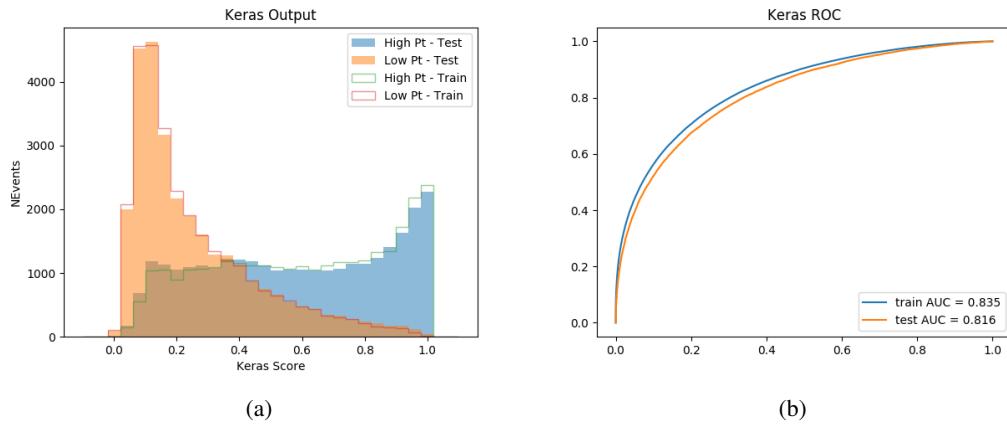


Figure A.27:

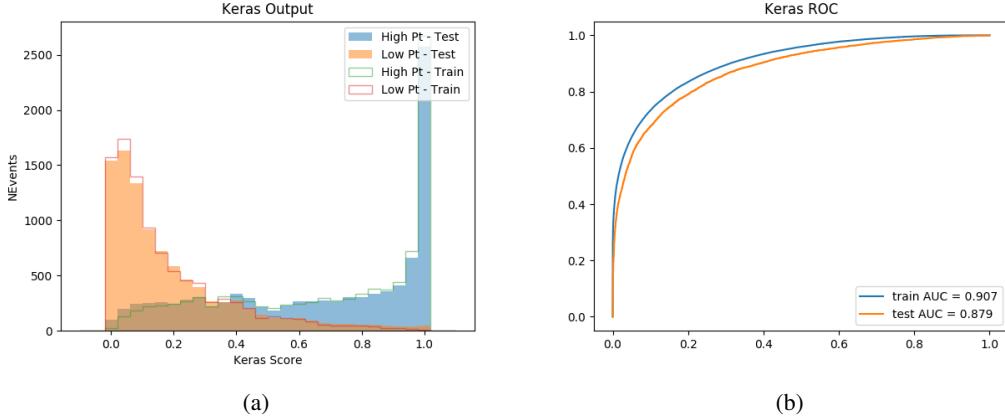


Figure A.28:

1682 A.5 Impact of Alternative Jet Selection

1683 A relatively low p_T threshold of 15 GeV is used to determine jet candidates, as the jets originating
 1684 from the Higgs decay are found to fall between 15 and 25 GeV a large fraction of the time. The
 1685 impact of different jet p_T cuts on our ability to reconstruct the Higgs p_T is explored here. The
 1686 performance of the Higgs p_T prediction models is evaluated for jet p_T cuts of 10, 15, 20, and 25
 1687 GeV.

1688 **B**

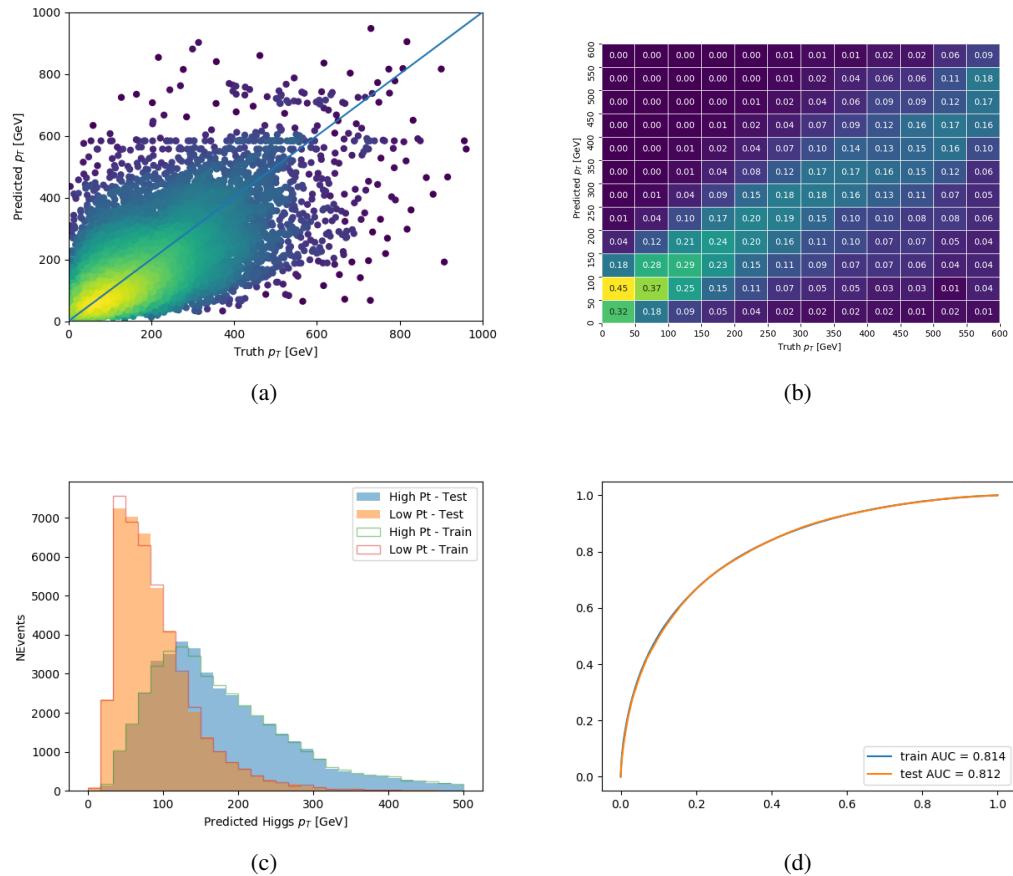


Figure A.29: